

Long-tailed representation learning algorithm based on adaptive prototypes and semantic awareness

LI Tiantian, XUE Zhen*, ZHANG Liangliang, LIAN Xu

School of Mathematics, North University of China, Taiyuan 030051, China

***Corresponding author:** XUE Zhen (xuezhen@nuc.edu.cn)

Received: December 30, 2025

Revised: January 20, 2026

Accepted: March 3, 2026

Abstract: Label scarcity and long-tailed distribution imbalance are significant challenges in industrial equipment monitoring. Currently, self-supervised learning methods are affected by sample quantity bias and semantic confusion under complex operating conditions, which limits their ability to represent sparse critical states. To address these issues, we propose a co-evolutionary prototypical contrastive learning (EPCL) framework. Through progressive learning from coarse-grained semantic discovery to fine-grained discriminative enhancement, this framework enables an in-depth analysis of the intrinsic structure of long-tailed data. Specifically, an adaptive prototype-based clustering algorithm based on optimal transport theory is introduced, thereby achieving unbiased representation learning through data-driven dynamic priors. Furthermore, a semantic-aware and hierarchical negative sample weighting scheme is designed to optimize discriminative boundaries while mitigating class imbalance by enforcing prototype consistency constraints and employing an adaptive weighting strategy. Extensive experiments were conducted on several public long-tailed visual benchmarks, including CIFAR10-LT, CIFAR100-LT, and ImageNet-100-LT, as well as the industrial fault diagnosis dataset. The results demonstrated that the EPCL achieved better performance than fifteen mainstream self-supervised methods (e.g., SimCLR and SwAV) in both linear evaluation and few-shot classification tasks. On the CIFAR100-LT dataset, the EPCL improved the tail-class accuracy by 4.56% compared to SimCLR. Ablation studies and visualization results verified the effectiveness and generalization ability of the framework. This work offers a promising insight and practical solution for representation learning from unlabeled long-tailed measurement data.

Key words: long-tailed distribution; self-supervised learning; contrastive learning; semantic awareness; adaptive prototype clustering; representation learning

0 Introduction

In the condition monitoring and fault diagnosis of complex industrial systems, high precision and reliability depend on the in-depth interpretation of multi-source sensor data. However, data collected from actual operating environments usually follow a long-tailed distribution. Most of the data generated during the life cycle of high-value equipment, such as satellite attitude control subsystems, wind turbine gearboxes, and aircraft engines, corresponds to normal operating conditions. However, samples representing various potential fault states such as initial pitting in bearings, structural cracks in blades, and intermittent circuit anomalies are extremely scarce^[1]. The significant disparity in sample size between head and tail classes results in an inherently imbalanced data distribution. Traditional diagnostic models, developed under the assumption of class balance, often fail to effectively

extract discriminative features from the limited tail-class samples under such distribution characteristics.

Current diagnostic methods predominantly rely on supervised learning. Relevant research focuses on addressing long-tailed distribution issues through data resampling, class-sensitive learning, and class-balance optimization in training strategies^[2-12]. Data-level resampling strategies alleviate class imbalance by modifying the distribution of training samples^[2-4]. At the algorithm level, reweighting methods operate by designing weight allocation mechanisms for loss functions to guide the optimization process^[5-9]. However, these methods usually require a large amount of high-quality and precisely labeled fault samples, and this requirement is difficult to meet in actual industrial applications.

To reduce reliance on manual annotation, self-supervised learning (SSL) designs pre-training tasks to learn generalizable representations from unlabeled data^[13-18,20]. Contrastive learning, the current mainstream

SSL paradigm, learns representations through the construction of positive-negative sample pairs^[17-18]. These methods incur high computational costs because of their dependence on large-batch training and a large number of negative samples. To address the reliance on negative samples, non-contrastive or consistency-based approaches have been developed. Grill et al.^[21] proposed BYOL, that uses an asymmetric teacher-student network architecture to learn effectively by focusing solely on positive pairs, thus avoiding model collapse. Chen et al.^[22] further presented SimSiam and demonstrated via a simple twin-network architecture, the critical role of stop-gradient operations in preventing model collapse. In terms of theoretical progress, Wang et al.^[23] advanced contrastive learning theory by systematically analyzing the alignment and uniformity properties, thereby establishing a theoretical framework for understanding its internal mechanisms.

The application of self-supervised learning to long-tailed data has emerged as a novel and challenging research direction, which aims to learn unbiased and discriminative representations from inherently imbalanced data without relying on labeled data. The pioneering work by Kang et al.^[24] demonstrated that long-tailed recognition problems can be decomposed into two relatively independent stages, namely representation learning and classifier learning. Chuang et al.^[19] carried out the first systematic analysis of bias problems in contrastive learning. They proposed a biased contrastive loss that rectifies bias by estimating the expected value of positive samples from negative samples, thereby laying the theoretical groundwork for contrastive learning in long-tailed settings. Subsequently, Wei et al.^[25] explored simplicity bias in long-tailed image recognition and found that tail classes are more vulnerable to this type of bias. To capture the long-tailed distribution, Lin et al.^[26] proposed the frequency-aware self-supervised learning (FASSL), a method that explicitly learns frequency-aware prototypes. It uses these prototypes to reweight samples, thus enhancing the learning of tail classes within a teacher-student framework. From a data perspective, Zhou et al.^[27] proposed boosted contrastive learning (BCL) to enhance long-tailed learning with missing labels. Zheng et al.^[28] developed a joint framework integrating clustering and self-supervised representation learning. In the field of graph neural networks, Zhao et al.^[29] proposed the long-tail augmented graph contrastive learning method (LAGCL). Subsequent research incorporated graph diffusion techniques, leading to the graph propagation-based diffusion (GPD) model^[30]. At the industrial application level, to address the data imbalance challenge in high-speed motor fault diagnosis, the two-stage self-supervised

learning framework (SMDN)^[31] combines data augmentation with contrastive learning to extract robust features from unlabeled data. The studies summarized the development of long-tailed self-supervised learning through both theoretical innovation and cross-domain application, driving substantial improvements in feature representation quality, learning efficiency for tail classes, and practical adaptability.

However, the direct application of mainstream self-supervised contrastive learning methods to real-world data with long-tailed distribution properties substantially degrades their performance. Such methods commonly rely on the implicit assumption of a uniform intra-batch class distribution. The severe class imbalance in real-world monitoring data directly violates this assumption, leading to two distinct challenges. The first challenge is the quantity bias issue. In contrastive learning, samples from numerically dominant head classes dictate the model's optimization direction, leading the learning process to overemphasize representations of majority classes. The second challenge is the semantic confusion issue. Negatives generated via random sampling may include semantically similar instances to the anchor, known as false negatives. Forcing the separation of semantically similar samples in the feature space distorts its semantic structure and impairs the model's capacity to distinguish critical minority-class patterns.

To address these issues, existing research has explored a variety of technical approaches. Specifically, a key strategy for bias reduction is to employ loss-function weighting to mitigate the optimization bias induced by data imbalance^[19]. In contrast, prototype-based methods focus on modeling the global semantic structure of data via cluster centers, thereby alleviating the degradation of feature representation resulting from long-tailed distributions^[14]. Although these approaches have achieved notable progress, most of them focus primarily on independent enhancements to individual technical modules. They typically treat global structure modeling and local bias correction as two separate components. This isolated design paradigm impedes dynamic and in-depth interaction between the two components during training.

We propose a co-evolutionary prototypical contrastive learning (EPCL) framework. The core of the framework lies in a bidirectionally driven, closed-loop learning system with an evolutionary dynamic, operating within the feature space (Fig.1). Through a dynamic and calibrated collaboration between prototype clustering and contrastive learning, EPCL effectively harmonizes information from global structural modeling and local bias correction. Under

this mechanism, the hierarchical semantic priors identified through adaptive prototypical clustering are deeply integrated with contrastive learning signals modulated by semantic awareness, forming a mutually corrective co-evolving feedback loop. This approach enables the model to guide representation learning toward a systematic analysis of the intrinsic topological distribution of long-tailed data under fully unsupervised conditions. On the one hand, clustering centers maintain stability and mitigate bias

under the discriminative constraints of contrastive learning. On the other hand, the contrastive learning process leverages the structured prior from clustering to alleviate negative sample conflicts and delineate semantic boundaries more clearly. Through such systematic synergy, the model simultaneously enhances feature discriminability and clustering structural validity, providing a novel approach for direct learning of robust features from complex imbalanced measurement data.

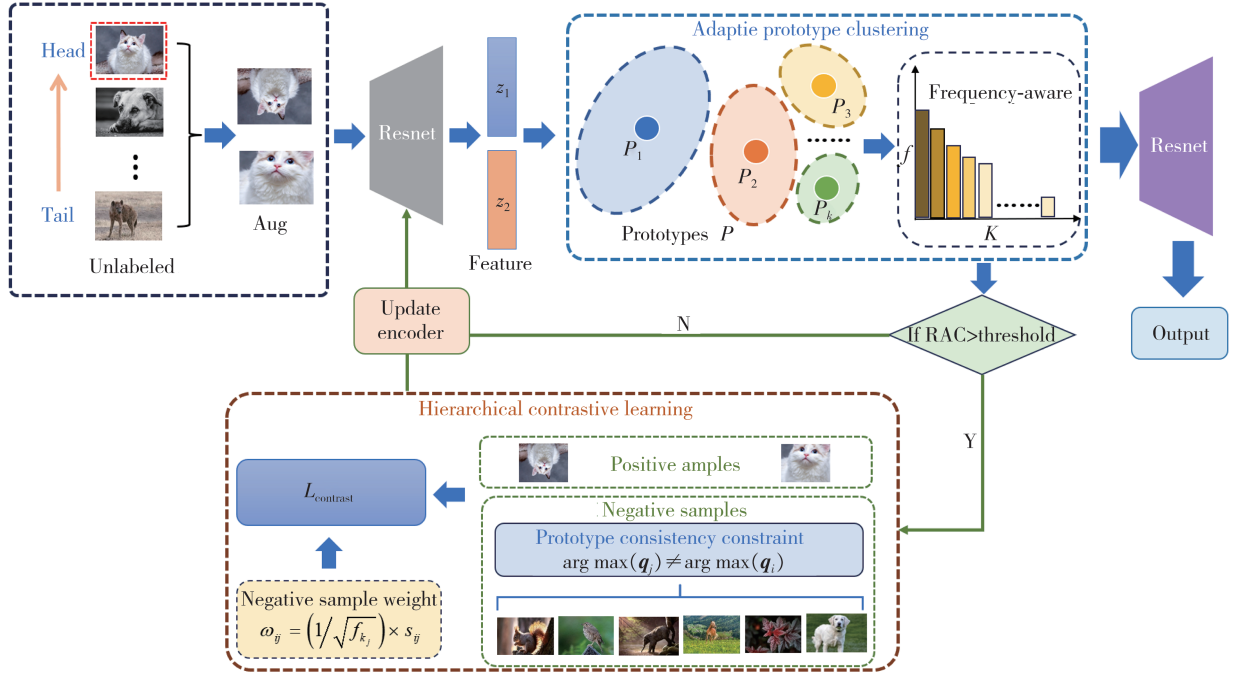


Fig. 1 Flowchart of EPCL algorithm

1 Methodology

1.1 Problem description

Let the input space be $\chi \subset \mathbf{R}^{H \times W \times C}$, and the data follows a long-tailed distribution $p(x) \propto x^{-\alpha}$ with long-tailed factor $\alpha > 0$. Given an unlabeled dataset $D = \{x_i\}_{i=1}^N \sim p(x)$. The objective of self-supervised

contrastive learning is to learn an encoder $f_\theta: \chi \rightarrow \mathbf{R}^d$ that maps input to a d -dimensional unit hypersphere space, such that semantically similar samples are close together in the embedding space, while semantically dissimilar samples are sufficiently separated.

Traditional self-supervised contrastive learning employs the InfoNCE loss function, i. e.,

$$L_{\text{NCE}} = -\mathbb{E} \left[\log \frac{\exp\left(\frac{\text{sim}(\mathbf{z}_i^{(1)}, \mathbf{z}_i^{(2)})}{\tau}\right)}{\exp\left(\frac{\text{sim}(\mathbf{z}_i^{(1)}, \mathbf{z}_i^{(2)})}{\tau}\right) + \sum_{j=1}^M \exp\left(\frac{\text{sim}(\mathbf{z}_i^{(1)}, \mathbf{z}_j)}{\tau}\right)} \right], \quad (1)$$

where $\mathbf{z}_i^{(1)}$ and $\mathbf{z}_i^{(2)}$ are different enhanced views of the same anchor image x_i , serving as positive samples; $\{\mathbf{z}_j\}_{j=1}^M$ is the negative sample set, typically comprising all other samples in the batch except the positive samples; τ denotes the temperature hyperparameter; $\text{sim}(\cdot, \cdot)$ represents the cosine similarity function, and $\mathbb{E}$ denotes the expectation operator.

Under long-tailed distributions, Eq.(1) encounters the

following fundamental challenges. The negative sample set includes numerous semantically similar yet mislabeled pseudo-negatives, and the forced separation of these samples undermines the integrity of semantic structures. The numerical superiority of head-class samples allows them to dictate the optimization direction during gradient accumulation. This dominance results in representations skewed toward head classes at the expense of tail classes.

Our solution was a progressive training strategy that prioritized learning structural information over detailed discrimination. It first employed adaptive prototype clustering to reveal the hierarchical semantic structure of long-tailed data in an unsupervised setting. When the prototype space reached a stable state, the strategy incorporated semantic-aware contrastive learning to improve feature discriminability. A bidirectional feedback mechanism was established between the prototype loss and the contrastive loss, which supported unbiased representation learning for long-tailed data in unlabeled scenarios.

1.2 Adaptive prototype clustering

1.2.1 Non-uniform optimal transmission allocation

To overcome the implicit prior assumption of approximately balanced sample counts per cluster in conventional clustering methods, we constructed an adaptive prototype clustering mechanism based on optimal transport theory. This approach allows prototype allocation to be entirely determined by the data's inherent long-tailed structure. Let $Z = \{\mathbf{z}_i\}_{i=1}^B \in \mathbf{R}^d$ denote the batch-normalized feature set and $P = \{\mathbf{p}_k\}_{k=1}^K \in \mathbf{R}^d$ denote the learnable prototype vector, with K typically being greater than the number of classes to capture fine-grained semantic information. Prototypes are initially generated by K-means++ to mitigate the risk of local optima. The cost matrix C is defined based on the cosine similarity between features and prototypes, and its elements are

$$C_{ik} = 1 - \frac{\mathbf{z}_i^T \mathbf{p}_k}{\|\mathbf{z}_i\| \|\mathbf{p}_k\|}, \quad (2)$$

where C_{ik} represents the cost of assigning sample $\mathbf{z}_i$ to prototype $\mathbf{p}_k$, and a smaller value indicates greater similarity between the sample and the prototype.

Let $\mathbf{Q}$ denote the soft assignment matrix from samples to prototypes, and its element Q_{ik} represents the quality of assigning sample $\mathbf{z}_i$ to prototype $\mathbf{p}_k$. Matrix $\mathbf{Q}$ must simultaneously satisfy row and column marginal constraints: rows are uniformly distributed, and columns correspond to the prototype prior frequency vector $\mathbf{f} \in \mathbf{R}^K$. The feasible domain is

$$\partial = \left\{ \mathbf{Q} \in \mathbf{R}_+^{B \times K} \mid \mathbf{Q} \mathbf{1}_K = \frac{1}{B} \mathbf{1}_B, \mathbf{Q}^T \mathbf{1}_B = \mathbf{f} \right\}, \quad (3)$$

where $\mathbf{1}_K$ and $\mathbf{1}_B$ represents a K -dimensional and B -dimensional all-ones vector, respectively.

Based on this, formulate the optimal transmission problem with a band-entropy regularization term.

$$\mathbf{Q}^* = \arg \min_{\mathbf{Q} \in \partial} \left(\langle C, \mathbf{Q} \rangle_F - \epsilon \cdot H(\mathbf{Q}) \right), \quad (4)$$

where $\langle \cdot, \cdot \rangle_F$ denotes the Frobenius inner product, $H(\mathbf{Q}) = - \sum_{i,k} Q_{ik} \log Q_{ik}$ represents the entropy regularization term, and ϵ controls the smoothness of the distribution. Entropy regularization ensures the optimization problem can be efficiently solved using the Sinkhorn-Knopp algorithm.

Let the nuclear matrix be $\mathbf{K}_{OT} = \exp(-M/\epsilon)$, with the exponential operation being performed element-wise. Then the solution to Eq. (4) is of the form $\mathbf{Q}^* = \text{diag}(\mathbf{u}) \mathbf{K}_{OT} \text{diag}(\mathbf{v})$, where $\mathbf{u}$ and $\mathbf{v}$ can be obtained via Sinkhorn-Knopp iteration shown as

$$\mathbf{v} \leftarrow \mathbf{f} \oslash (\mathbf{K}_{OT}^T \mathbf{u}), \mathbf{u} \leftarrow \frac{1}{B} \mathbf{1}_B \oslash (\mathbf{K}_{OT} \mathbf{v}), \quad (5)$$

where the symbol $\oslash$ denotes the operation of element-wise division.

Unlike methods such as SwAV that rely on column-wise and approximately uniform constraints, our approach derives the $\mathbf{f}$ value based on the estimation of prototype frequency. Consequently, the clustering process naturally allows head prototypes to absorb more samples and allocates only a few semantically relevant ones to tail prototypes, thereby avoiding structural distortions from forced balancing.

After obtaining the soft assignment matrix, to constrain the consistency of different augmented views from the same sample, we define the prototype consistency clustering loss as

$$L_{\text{cluster}} = \frac{1}{B} \sum_{i=1}^B \left(\frac{1}{2} \left[\text{KL}(\mathbf{q}_i^{(1)} \parallel \mathbf{q}_i^{(2)}) + \text{KL}(\mathbf{q}_i^{(2)} \parallel \mathbf{q}_i^{(1)}) \right] \right), \quad (6)$$

where $\mathbf{q}_i^{(1)}$ and $\mathbf{q}_i^{(2)}$ represent the soft classification vectors corresponding to the two augmented views of the input sample.

Minimizing Eq. (6) is equivalent to enforcing consistent membership of the same sample in the prototype space across different views. It achieves semantic clustering pre-training under completely unlabeled conditions, providing a stable and structurally sound prototype space for subsequent contrastive learning.

1.2.2 Prototype stabilization update

Non-uniform distribution relies on accurate estimation of prototype frequency. The symbol $\mathbf{f}$ is a probability vector satisfying $\sum_{k=1}^K f_k = 1$ and $f_k \in [0, 1]$. Considering that features are unstable during early training stages, directly using data-driven priors may lead to individual

prototypes monopolizing a large number of samples over time due to initialization-induced dominance. Therefore, we introduce a linear mixture strategy combining uniform priors and data-driven priors. The prototype prior employed in the t -th training iteration is

$$f_k^{(t)} = (1 - \omega_t) \frac{1}{K} + \omega_t \cdot \hat{f}_k^{(t)}, \quad (7)$$

where $\omega_t = t/T$ is entirely determined by the training progress, and T denotes the total number of training iterations. In the early stages, ω_t is small, EPCL primarily relies on a uniform prior distribution to ensure exploratory diversity. As training progresses, ω_t gradually increases, enabling prototype allocation to more faithfully reflect the true distribution of long-tailed data.

Using the soft assignment matrix $\mathbf{Q}^{(t)}$ for the current batch, the prototype activation vector for that batch can be obtained by

$$a_k^{(t)} = \frac{1}{B} \sum_{i=1}^B Q_{ik}^{(t)}. \quad (8)$$

It depicts the probability of each prototype being accessed in the current batch. Then, it updates the global frequency estimate using an exponential moving average (EMA).

$$\mathbf{f} \leftarrow \eta \mathbf{f} + (1 - \eta) \frac{\mathbf{a}}{\|\mathbf{a}\|_1}, \quad (9)$$

where η is the EMA update rate, a common momentum hyperparameter. This update ensures that $\mathbf{f}$ is a smooth, stable probability distribution.

To further enhance the stability of prototype updates for long-tailed distributions, we propose a frequency-aware adaptive momentum for each prototype. Intuitively, tail prototypes exhibit sparse samples and are more susceptible to noise interference, thus requiring higher momentum to ensure smoother updates. Head prototypes possess abundant samples and can appropriately reduce momentum to improve adaptability. For each prototype, its update momentum is negatively correlated with its frequency.

$$\gamma_k = \gamma_{\text{base}} + \alpha(1 - \sqrt{f_k}), \quad (10)$$

where γ_{base} is the base momentum, and α is an infinitesimal scalar, both of which are commonly fixed settings. The updated prototype is

$$\mathbf{p}_k \leftarrow \gamma_k \mathbf{p}_k + (1 - \gamma_k) \frac{\sum_{j=1}^B Q_{jk} \mathbf{z}_j}{\sum_{j=1}^B Q_{jk}}, \quad (11)$$

where $\sum_{j=1}^B Q_{jk} \mathbf{z}_j$ represents the weighted sum of all sample features in the current batch, weighted by their soft assignment probabilities to prototypes.

During training, both prototype priors and prototype vectors gradually converge with the data distribution, thereby constructing a semantic prototype space that matches the long-tailed structure and is robust to noise. Introducing contrastive learning too early will increase the risk of obtaining false negative results, so we use the value of the relative assignment consistency (RAC) metric to assess whether the prototype space has stabilized.

$$\text{RAC}^{(t)} = \frac{1}{B} \sum_{i=1}^B \langle \mathbf{q}_i^{(t)}, \mathbf{q}_i^{(t-1)} \rangle, \quad (12)$$

where $\mathbf{q}_i^{(t)}$ represents a probability distribution, $\text{RAC}^{(t)} \in [0, 1]$. From an information geometry perspective, $\mathbf{Q}^{(t)}$ can be regarded as a coordinate on the statistical manifold in the prototype space, while $\text{RAC}^{(t)}$ reflects the rate of change between adjacent stages on this manifold. When the prototype structure is still undergoing rapid transformation, $\text{RAC}^{(t)}$ is low. When prototype clustering has largely stabilized, $\text{RAC}^{(t)}$ approaches 1. In practice, when $\text{RAC}^{(t)}$ remains consistently above a preset threshold for several consecutive iterations, the prototype space can be considered sufficiently stable, and training proceeds to the next stage of contrastive learning.

1.3 Negative sample regulation in semantic perception

Once the prototype space stabilizes, relying solely on the clustering consistency loss proves insufficient for learning highly discriminative feature representations. To overcome this limitation, EPCL introduces a semantic-aware negative sample regulation mechanism atop the stabilized prototypes, which effectively addresses both false negatives and numerical imbalance in long-tailed scenarios.

First, define the hard prototype assignment $c_i = \arg \max_k q_{ik}$ for a sample based on the soft assignment vector $\mathbf{q}_i$. Using prototype assignment, coarse-grained semantic purification can be performed on the original negative sample set $\mathcal{C}$: only samples with different prototypes are retained as valid negative samples. For anchor samples, the new negative sample set is defined as

$$\mathcal{N}_j = \{j | c_j \neq c_i\}. \quad (13)$$

Although the number of false negatives is significantly reduced after filtering via Eq. (13), a severe numerical bias persists among the remaining negatives. The number of negatives corresponding to head prototypes far exceeds that of tail prototypes. To prevent gradients from remaining dominated by head prototypes, we design a dual-factor weighting function that integrates prior distribution and semantic similarity to finely regulate the contribution of

negatives. For the anchor sample and its negative sample, we first define a frequency-based factor as

$$\phi_{ij} = \frac{1}{\sqrt{f_{k_j}}}, \quad (14)$$

where f_{k_j} denotes the global frequency estimate for prototypes associated with negative samples. The square root transformation acts as a mild power transformation. It moderately boosts the weights of tail prototypes while reducing those of head prototypes, and it achieves a balance between training stability and class balance.

Second, introduce a semantic similarity factor based on prototype distributions. Let the soft assignment vectors for the anchor and negative sample be denoted as $\mathbf{q}_i$, $\mathbf{q}_j$, respectively. An intermediate distribution is constructed, that is

$$\mathbf{V}_{ij} = \frac{1}{2} (\mathbf{q}_i + \mathbf{q}_j). \quad (15)$$

The Jensen-Shannon (JS) divergence is defined as

$$JS_{ij} = \frac{1}{2} [\text{KL}(\mathbf{q}_i \| \mathbf{V}_{ij}) + \text{KL}(\mathbf{q}_j \| \mathbf{V}_{ij})]. \quad (16)$$

$$L_{\text{wNCE}} = -\mathbb{E} \left[\log \frac{\exp\left(\frac{\text{sim}(\mathbf{z}_i^{(1)}, \mathbf{z}_i^{(2)})}{\tau}\right)}{\exp\left(\frac{\text{sim}(\mathbf{z}_i^{(1)}, \mathbf{z}_i^{(2)})}{\tau}\right) + \sum_{j \in \mathcal{N}(j)} \omega_{ij} \exp\left(\frac{\text{sim}(\mathbf{z}_i^{(1)}, \mathbf{z}_j)}{\tau}\right)} \right]. \quad (20)$$

Compared with Eq.(1), Eq.(20) exhibits two fundamental differences. Most false negatives have been removed from $\mathcal{N}(j)$; By incorporating ω_{ij} , the focus of optimization shifts from numerically dominant easy negatives to numerically scarce yet information-rich hard negatives, significantly enhancing the classification capability for tail classes.

Ultimately, the total loss of EPCL is formed by a linear combination of prototype consistency loss and weighted contrast loss.

$$L_{\text{total}} = L_{\text{cluster}} + \beta L_{\text{wNCE}}, \quad (21)$$

where β is the weighting coefficient balancing the two types of loss. Eq.(21) embodies the synergistic relationship between structural modeling and discriminative enhancement. L_{cluster} stabilizes the semantic prototype structure, while L_{wNCE} further refining the discriminative boundaries upon it, forming a closed-loop feedback mechanism.

1.4 Time complexity and algorithm pseudocode

In the EPCL framework, the adaptive prototype clustering module employs the Sinkhorn-Knopp algorithm to solve an optimal transport problem. The

Its value range is $[0, \ln 2]$. Based on this, a normalized similarity factor is constructed as

$$s_{ij} = 1 - \frac{JS_{ij}}{\ln 2}. \quad (17)$$

The larger the value of s_{ij} , the more similar the two samples are in the prototype space. When two samples are not assigned to the same prototype but have a large s_{ij} value, it indicates they are near the decision boundary and represent hard negative samples, contributing more significantly to the learning of the decision boundary.

By combining frequency factors and similarity measures, we can assign adaptive weights to negative samples.

$$\tilde{\omega}_{ij} = \phi_{ij} \cdot s_{ij}. \quad (18)$$

To ensure the comparability of gradient scales across different negative samples, the weights of each negative sample within a batch are normalized as

$$\omega_{ij} = \frac{\tilde{\omega}_{ij}}{\sum_{j \in \mathcal{N}(j)} \tilde{\omega}_{ij}}. \quad (19)$$

Introducing the weighted mechanism, the contrastive loss naturally generalizes to weighted InfoNCE.

per-iteration computational complexity is $O(B \times K)$, where B is the batch size and K denotes the number of prototypes. During experiments, we set the iterations to 3, which ensures convergence of the assignment matrix. Under the same hardware setup, EPCL requires an average of about 45 s per epoch on the CIFAR100-LT dataset, which is only about 7% higher than that of SimCLR (approximately 42 s per epoch). This result confirms that our approach effectively improves representation learning on long-tailed data without introducing significant computational overhead.

The pseudocode of EPCL is as follows.

Algorithm 1: EPCL

Input: Dataset D , encoder f_θ , prototypes P , hyperparameters

Output: Train f_θ

for epoch = 1 to T **do**

for batch $\{x_i\}_{i=1}^B$ in D **do**

$x_i^{(1)}, x_i^{(2)} \leftarrow \text{Augment}(x_i)$

$\mathbf{z}_i^{(1)}, \mathbf{z}_i^{(2)} \leftarrow f_\theta(x_i^{(1)}), f_\theta(x_i^{(2)})$

$\mathbf{q}_i^{(1)}, \mathbf{q}_i^{(2)} \leftarrow \text{Sinkhorn}(\mathbf{z}_i, P, f)$

 Update f & P by Eq. (9), Eq. (11)

$L_{\text{cluster}} \leftarrow L_{\text{cluster}}(\mathbf{q}_i^{(1)}, \mathbf{q}_i^{(2)})$

 Calculate RAC by Eq. (12)

if RAC > threshold:

 Select the negative samples by Eq. (13)

```

Calculate  $\omega_{ij}$  by Eq. (19)
 $L_{\text{wNCE}} \leftarrow L_{\text{wNCE}}(\mathbf{z}, \omega_{ij}, \tau)$ 
 $L_{\text{total}} \leftarrow L_{\text{cluster}} + \beta \cdot L_{\text{wNCE}}$ 
else
 $L_{\text{total}} \leftarrow L_{\text{cluster}}$ 
end if
Update  $f_{\theta}$  using  $\nabla L_{\text{total}}$ 
end for
end for

```

2 Results and discussion

2.1 Experimental setup

The proposed method was comprehensively evaluated on multiple long-tailed visual benchmark datasets. CIFAR10-LT and CIFAR100-LT were chosen as the primary evaluation benchmarks. The original CIFAR datasets consisted of 60 000 images each with a 32×32 resolution. They covered 10 and 100 semantic classes, respectively. Long-tailed versions of the datasets were constructed with an exponential decay sampling strategy to simulate real-world long-tailed distributions. Three imbalance ratios of 10, 50, and 100 were set to assess the method's robustness. In the ImageNet-100-LT dataset, the number of samples per class ranges from 1 280 to 5. To further evaluate the generalization capability of our method, we conducted additional experiments on an industrial dataset: the Case Western Reserve University (CWRU)^[34] bearing fault dataset. This dataset comprises 10 classes (one normal state and nine fault states), covering nine distinct fault conditions corresponding to three different fault diameters at the inner raceway (IR), outer raceway (OR), and ball (B), with an imbalance ratio of 100. All models were evaluated on the original balanced test set to guarantee fairness and comparability.

We employed two standard evaluation protocols to systematically evaluate the quality of the learned representations. In the linear evaluation protocol, the weights of the pre-trained encoder were frozen. A linear classifier was then trained on these frozen representations using the full labeled dataset. The few-shot evaluation protocol simulated scenarios with extreme label scarcity. It kept the encoder frozen and trained the classifier on only 1% of the labeled samples per class.

For the CIFAR datasets, ResNet-18 was employed as the encoder. For the ImageNet-100-LT, ResNet-50 was used. For the CWRU dataset, ResNet-34-RSBU-CW^[39] was adopted as the encoder. The optimizer utilized SGD with an initial learning rate of 0.01 and adopted a cosine annealing scheduler. Critical hyperparameter settings (such as contrastive learning temperature) were aligned with

mainstream research to ensure experimental comparability. The momentum update parameters were set to $\eta=0.99$ and $\gamma_{\text{base}}=0.9$, both commonly used in experiments. The training epochs for the downstream linear classifier followed standard practices in the literature: 30 epochs for full-data training and 100 epochs for few-shot training. All experiments were implemented using PyTorch on a NVIDIA RTX 4060 GPU.

2.2 Evaluation metric

To rigorously evaluate the generalization capability of the proposed method, this study adhered to a stringent assessment framework. The selected evaluation metric was top-1 accuracy, which was primarily used to measure the precision of matching between the model's top-ranked predicted class and the actual ground truth class. The top-1 accuracy is calculated by

$$\text{ACC}_{\text{Top-1}} = \frac{n_{\text{correct}}}{N}, \quad (22)$$

where n_{correct} denotes the number of samples correctly predicted as top-1 in the test set; N represents the total number of samples in the test set; and $\text{ACC}_{\text{Top-1}}$ ranges from 0 to 1.

2.3 Comparison of results

To evaluate the overall performance of EPCL on mainstream long-tailed benchmark datasets, we compared it against multiple representative baseline methods. This comparison aimed to comprehensively examine the generalizability and transferability of the representations learned by EPCL. Linear evaluation directly reflects the quality of pre-trained features. We also computed the standard deviation (Std) of accuracy to evaluate the model's stability and quantify the performance disparity across different classes. The results are presented in Table 1, and optimal results are highlighted in bold.

As shown in Table 1, EPCL achieves leading performance across multiple long-tailed benchmarks. On the CIFAR100-LT dataset, EPCL records an overall accuracy of 54.17%, surpassing all comparative methods and the supervised baseline, which reaches 54.06%. Its accuracy on rare classes reaches 52.14%, surpassing supervised learning by 0.75%. It also achieves the smallest performance standard deviation of 2.01, indicating optimal prediction stability across classes. On the CIFAR10-LT dataset, EPCL attains an overall accuracy of 80.05%, closely matching the 80.76% accuracy of supervised learning. Its accuracy in rare classes reaches 77.62%,

representing an improvement of 2.19 percentage points compared with supervised learning, which demonstrates robust recognition capabilities for rare samples. On the more challenging ImageNet-100-LT dataset, EPCL achieves an overall accuracy of 68.58%, outperforming the comparative methods SDCLR and SimCLR by 2.10 percentage points and 3.12 percentage points, respectively. Its accuracy in rare classes reaches 65.81%, which significantly outperforms SimCLR's accuracy (59.69%). This performance confirms EPCL's

effectiveness in alleviating class bias under long-tailed distributions. These experimental results show that EPCL not only learns highly discriminative visual representations but also significantly improves the model's recognition accuracy in rare classes while preserving performance on frequent classes.

Furthermore, on the CWRU dataset, EPCL achieves an overall accuracy of 76.31% while maintaining stable performance, thereby validating its strong transfer capability.

Table 1 Linear evaluation results using all labeled data

Dataset	Method	All/%	Rare/%	Medium/%	Frequent/%	Std
ImageNet-100-LT	SimCLR ^[18]	65.46	59.69	63.71	69.54	4.04
	SDCLR ^[32]	66.48	60.92	65.04	70.10	3.75
	EPCL	68.58	65.81	68.46	71.47	2.31
CIFAR100-LT (imbalance ratio of 100)	LDAM-DRW+SSP ^[13]	43.43	—	—	—	—
	SwAV ^[14]	47.00	44.00	47.03	49.97	2.44
	DR+H2T ^[35]	47.73	—	—	—	—
	SimCLR ^[18]	49.76	47.58	50.36	51.35	1.60
	DnC ^[27]	48.78	45.65	46.68	54.00	4.55
	BYOL ^[21]	48.86	45.55	47.42	53.62	3.45
	BCL-I ^[27]	52.22	48.27	53.03	55.35	2.95
	BCL+SGHAF ^[38]	52.29	—	—	—	—
	LPA ^[33]	53.3	—	—	—	—
	EPCL	54.17	52.14	53.46	56.91	2.01
	Supervised	54.06	51.39	54.18	56.62	2.13
CIFAR10-LT (imbalance ratio of 100)	CE+MASW ^[36]	75.34	—	—	—	—
	SimCLR ^[18]	75.37	69.33	73.33	83.45	5.94
	BYOL ^[21]	75.66	75.43	69.83	81.70	4.85
	SwAV ^[14]	76.60	73.30	71.10	85.40	6.29
	LDAM-DRW+SSP ^[13]	77.83	—	—	—	—
	TSC ^[37]	79.70	—	—	—	—
	EPCL	80.05	77.62	76.30	86.23	4.40
	Supervised	80.76	75.43	80.93	85.93	4.95
CWRU (imbalance ratio of 100)	SimCLR ^[18]	71.36	—	—	—	—
	KCL ^[39]	73.52	—	—	—	—
	ECL ^[39]	74.48	—	—	—	—
	EPCL	76.31	71.22	74.69	83.02	4.95

“—” indicates that the original literature did not report this type of result. All/Rare/Medium/Frequent denote the top-1 linear evaluation accuracy (%) on all, rare, medium, and frequent classes, respectively.

This work further considered the challenging few-shot setting. Linear classifiers were trained and evaluated with only 1% of the labeled data to assess representation robustness and avoid overfitting. The results are shown

in Table 2.

EPCL demonstrates strong generalization capabilities under the highly challenging 1% few-shot learning setting.

Table 2 Linear evaluation results using 1% labeled data (imbalance ratio of 100)

Dataset	Method	All/%	Rare/%	Medium/%	Frequent/%	Std
CIFAR100-LT	SwAV ^[14]	20.14	13.91	20.06	26.44	5.12
	BYOL ^[21]	20.62	14.03	19.61	28.24	5.84
	SimCLR ^[18]	22.51	15.45	22.48	29.59	5.77
	SDCLR ^[32]	25.39	19.91	25.52	30.74	4.42
	EPCL	26.18	21.06	26.38	31.10	4.10
	Supervised	28.97	13.88	32.03	41.00	11.28
CIFAR10-LT	BYOL ^[21]	62.33	50.40	60.53	76.05	10.55
	SimCLR ^[18]	62.98	56.67	62.97	69.30	5.16
	SwAV ^[14]	65.16	52.90	61.97	80.60	11.53
	SDCLR ^[32]	67.23	56.63	61.87	83.20	11.49
	EPCL	68.49	59.67	61.77	84.03	11.02
	Supervised	69.82	65.07	66.47	77.93	5.76

On the CIFAR100-LT dataset, EPCL achieves an overall accuracy of 26.18%, surpassing the second-best SDCLR method by 0.79 percentage points. Its rare class accuracy reaches 21.06%, exceeding SDCLR by 1.15 percentage points. The standard deviation of EPCL is 4.10, the lowest among all self-supervised methods. This result demonstrates more stable cross-class performance. On the CIFAR10-LT dataset, EPCL leads with an overall accuracy of 68.49%. While slightly below SDCLR in medium classes, EPCL achieves the best performance on rare classes at 59.67%, surpassing SDCLR by 3.04%. Under conditions of labeled data scarcity, the performance gap between EPCL and supervised learning methods is significantly narrowed. Notably, EPCL even outperforms supervised learning on the rare classes of CIFAR100-LT, highlighting its practical value in few-shot scenarios.

2.4 Ablation experiment

To gain a deeper understanding of the contributions of each component within the EPCL framework and its internal mechanisms, multiple sets of diagnostic experiments were conducted. These experiments aimed to validate the rationality of the framework design and its effective adaptability to long-tailed distributions by analyzing key hyperparameters and component functions.

2.4.1 Prototype quantity sensitivity analysis

On the CIFAR100-LT dataset (with imbalance ratios of 10, 50, and 100), we systematically adjusted the number of prototypes K from 50 to 500 to observe changes in the model's top-1 accuracy under the linear evaluation protocol. The corresponding results are shown in Fig.2.

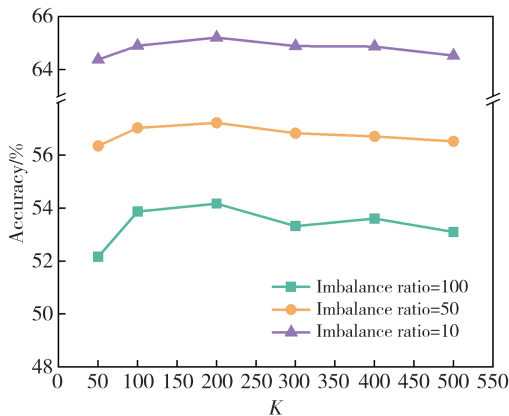


Fig. 2 Effect of prototype quantity K on linear evaluation accuracy of EPCL

As shown in Fig.2, the model achieves optimal performance across three imbalance ratios at $K=200$, indicating that a moderate level of prototype overfitting

best captures the fine-grained semantic structures in long-tailed data. When K is below 200, performance is constrained by too few prototypes to model the data's complex distribution adequately. Once K exceeds 200, performance declines, with a sharp drop at the 100:1 imbalance ratio, suggesting that excessive prototypes introduce noise and cause overfitting. Notably, all configurations consistently surpass the baseline performance. Furthermore, the model's sensitivity to K variations increases with the imbalance ratio, underscoring the critical necessity of careful prototype selection in severe long-tailed scenarios. Cross-dataset validation showed that setting the number of prototypes to be 2–3 times the number of true classes is reasonable.

2.4.2 Loss weight sensitivity analysis

To systematically evaluate the impact of the loss weight β in Eq. (21), we conducted a sensitivity analysis on the CIFAR100-LT dataset (imbalance ratio=100). Initial experiments indicate that model performance drops markedly when β is beyond the range $[0.5, 1.5]$. Therefore, we tested β across this interval in increments of 0.1. The corresponding results are shown in Fig.3.

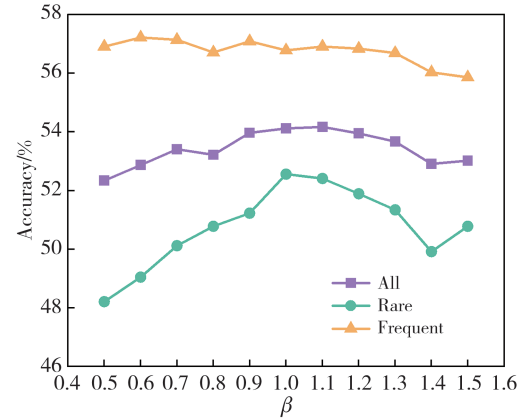


Fig. 3 Effect of loss weight β on linear evaluation accuracy of EPCL

As shown in Fig.3, overall accuracy remains stable and reaches its optimum (53.95%–54.17%) within the range $\beta \in [0.9, 1.2]$, demonstrating the model's robustness to this hyperparameter. Tail classes (Rare) are more sensitive to β , achieving the highest accuracy of 52.56% at $\beta=1.0$. In contrast, the accuracy for head classes (Frequent) varies less than 0.5% across $\beta \in [0.6, 1.2]$, indicating that representations of majority classes are largely unaffected by the loss weighting. Thus, the interval $\beta \in [0.9, 1.2]$ provides the best trade-off among overall accuracy, tail-class recognition, and stability. Although tail accuracy peaked at $\beta=1.0$, we set $\beta=1.1$ as the final configuration, as it achieves the highest overall accuracy (54.17%) while

maintaining strong performance on tail classes. This choice has also performed well in other long-tailed benchmarks.

2.4.3 Ablation experiment on OT allocating prior

To verify whether the “non-uniform prior constraints” in EPCL better accommodate long-tailed structures, we conducted ablation experiments on CIFAR100-LT (imbalance ratio of 100) under the linear evaluation protocol. All experiments employed the same prototype count $K=200$, identical training rounds, and hyperparameter settings. The corresponding results are shown in Table 3.

Table 3 Results of ablation experiment on OT allocating prior

Method	All/%	Rare/%	Medium/%	Frequent/%	Std
EPCL (Uniform OT)	50.66	45.12	50.49	56.37	4.59
EPCL	54.17	52.14	53.46	56.91	2.01

Comparative analysis was performed by constructing the following variant models.

EPCL (Uniform OT): In this setting, we only replaced the Sinkhorn column with a uniform distribution (effectively forcing each prototype to absorb approximately the same number of samples).

EPCL: Sinkhorn’s columns were given by the prototype frequency prior.

As shown in Table 3, the configuration of columns and constraints in the Sinkhorn allocation significantly influences long-tailed unsupervised representation learning. Following the introduction of non-uniform columns and prior distributions, the model shows universal performance improvements in linear evaluation on CIFAR100-LT. The overall accuracy improves from 50.66% to 54.17%, with substantial increases in rare classes (+7.02 percentage points) and medium classes (+2.97 percentage points), compared to only a marginal gain in frequent classes (+0.54 percentage points).

This asymmetric improvement indicates that non-uniform columns and constraints can more effectively adapt to the true semantic capacity distribution of long-tailed data, thereby alleviating the compression or misclustering of rare classes during the structural modeling phase. Under uniform column and constraint settings, the standard deviation reaches 4.59, while non-uniform columns reduces it significantly to 2.01. This result demonstrates that forcing prototypes to absorb an identical number of samples exacerbates performance divergence across frequency bands, leading to significant degradation in rare classes. Conversely, allowing prototype capacity to adjust adaptively based on the data distribution facilitates learning a more balanced and consistent representation space.

2.4.4 Ablation experiment on negative sample regulation mechanism

We further conducted ablation studies via a negative sample regulation mechanism to analyze the specific contributions of each discriminative enhancement component to the reliable prototype structure. The corresponding results are shown in Table 4.

Table 4 Results of ablation experiment on negative sample regulation mechanism

Method	All/%	Rare/%	Medium/%	Frequent/%	Std
EPCL w/o SC	51.23	48.15	50.89	54.65	2.66
EPCL w/o FW	52.67	50.32	52.41	55.28	2.03
EPCL w/o SW	53.12	51.21	52.54	55.61	1.84
EPCL	54.17	52.14	53.46	56.91	2.01

Variant models were constructed on the CIFAR100-LT dataset (with an imbalance ratio of 100) for comparative analysis.

EPCL w/o SC: In this setting, we removed the semantic consistency constraint for prototype attribution. Consequently, the negative sample set included all non-positive samples in the batch without semantic purification. This setting was designed to validate the critical role of this constraint in mitigating the false negative sample problem.

EPCL w/o FW: In this setting, we removed the frequency weighting factor (FW) from the dual-factor weighting to evaluate the effectiveness of balancing class quantity deviations.

EPCL w/o SW: In this ablation setting, we removed the similar weighting (SW) factor from the dual-factor weighting to evaluate its contribution to enhancing feature discriminability by focusing on difficult negative samples.

EPCL: The proposed complete model encompasses all core components.

As shown in Table 4, each component plays a critical role in model performance but yields distinct effects. Removing the semantic consistency constraint leads to a 2.94 percentage points drop in overall accuracy, with a more pronounced impact on rare classes (3.99 percentage points). This indicates that the constraint was crucial for mitigating false negatives and protecting rare classes. Removing the frequency weighting factor reduces accuracy on rare classes by 1.82 percentage points, while having minimal impact on frequent class performance. This confirmed the effectiveness of EPCL in balancing class contributions and alleviating numerical bias. The similarity weighting factor has the least impact, resulting in only 1.05 percentage points decrease in overall accuracy. This result suggested its role as an auxiliary optimizer that enhanced discriminative power

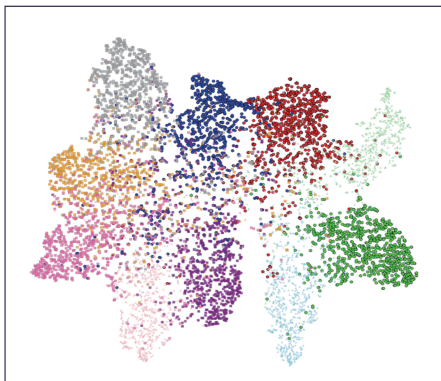
by identifying challenging samples. In summary, the semantic consistency constraint was the most critical component for protecting rare classes. The frequency weighting factor primarily improved class balance, while the similarity weighting factor provided supplementary discriminative enhancement. The synergistic interaction of these three components enabled EPCL to achieve optimal performance.

As revealed in Tables 3 and 4, the non-uniform OT distribution provides a robust foundation for EPCL. Building on this foundation, the semantic consistency constraints, frequency weighting, and similarity weighting worked progressively to mitigate false negatives and numerical bias, thereby enhancing discriminative performance.

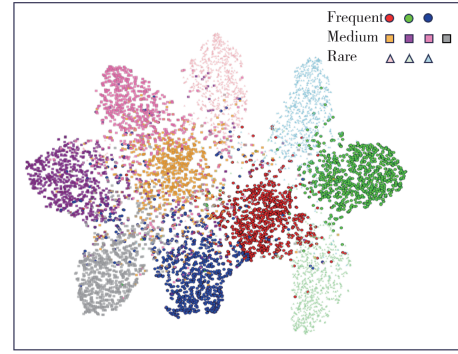
2.5 Result visualization

To further validate the adaptability of EPCL to long-tailed data and the discriminative capability of its learned features, we performed t-SNE visualization on the long-tailed CIFAR10-LT dataset (with an imbalance ratio of 100). Fig.4 compares the spatial distributions of features learned by EPCL and the baseline SimCLR. Each point represents a test sample, colored according to its class and distinguished by shape (indicating frequent, medium, or rare classes) to reflect its position in the long-tailed distribution. High-saturation colors correspond to head classes, while low-saturation colors correspond to tail classes.

As shown in Fig.4, SimCLR exhibits significant class confusion in its learned features. Tail-class samples are scattered, and the boundaries with head classes are blurred, indicating its difficulty in learning discriminative features under long-tailed distributions. In contrast, the features learned by EPCL possess a clear semantic structure. Distinct classes form compact and well-separated clusters, while tail classes also maintain strong intra-class cohesion.



(a) SimCLR



(b) EPCL

Fig. 4 Visual comparison of SimCLR and EPCL methods

Fig.4 visually validates the effectiveness of our co-evolution mechanism. It demonstrated that EPCL could simultaneously preserve feature discriminability and semantic integrity under unsupervised conditions, further enhancing the representation capability for tail classes.

3 Conclusions

This study proposes an algorithm named EPCL to effectively mitigate quantity bias and semantic confusion in unsupervised representation learning. It employs adaptive prototype clustering and semantically aware negative sample regulation to improve the representation quality of tail classes. Systematic experiments were conducted on long-tailed benchmarks, including CIFAR10-LT, CIFAR100-LT, ImageNet-100-LT, and CWRU. The results showed that EPCL outperformed current mainstream self-supervised contrastive methods in multiple tasks such as linear evaluation and few-shot classification. Specifically, on the CIFAR100-LT dataset, EPCL achieved an overall accuracy of 54.17% in linear evaluation, with tail class accuracy reaching 52.14%. Under the more challenging 1% few-shot setting, it maintained a tail class accuracy of 21.06%, demonstrating strong generalization capability. Ablation studies confirmed that each component of the framework contributed positively, and that the semantic consistency constraint was particularly crucial for improving long-tailed class performance. EPCL offers an effective solution for unsupervised representation learning on long-tailed data, with potential for transfer to broader industrial vision diagnostics tasks. Future work will explore adaptive mechanisms for determining prototype counts to further enhance the method's practicality in open environments.

Acknowledgement

The work was supported by the National Natural Science Foundation of China (No. 12401703) and the

Fundamental Research Program of Shanxi Province (Nos.202203021211088, 202403021221109, 202403021212256).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

- [1] CHEN Z, CHEN J, FENG Y, et al. Imbalance fault diagnosis under long-tailed distribution: challenges, solutions and prospects. *Knowledge-Based Systems*, 2022, 258: 110008.
- [2] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 2002, 16(1): 321-357.
- [3] SÁEZ J A, KRAWCZYK B, WOŹNIAK M. Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets. *Pattern Recognition*, 2016, 57: 164-178.
- [4] LIN C A, TSAI C F, LIN W C. Towards hybrid over- and under-sampling combination methods for class imbalanced datasets: an experimental study. *Artificial Intelligence Review*, 2023, 56(2): 845-863.
- [5] CUI Y, JIA M L, LIN T Y, et al. Class-balanced loss based on effective number of samples//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019. LongBeach, CA, USA. New York: IEEE, 2019: 9260-9269.
- [6] PARK S, LIM J, JEON Y, et al. Influence-balanced loss for imbalanced visual classification//2021 IEEE/CVF International Conference on Computer Vision, October 10-17, 2021. Montreal, QC, Canada. New York: IEEE, 2021: 715-724.
- [7] ZHU J G, WANG Z, CHEN J J, et al. Balanced contrastive learning for long-tailed visual recognition//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-24, 2022, New Orleans, LA, USA. New York: IEEE, 2022: 6898-6907.
- [8] CAO K D, WEI C L, GAIDON A, et al. Learning imbalanced datasets with label-distribution-aware margin loss//The 33rd Conference on Neural Information Processing Systems, December 8-14, 2019, Vancouver, Canada. New York: Curran Associates, 2019: 1565-1576.
- [9] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2999-3007.
- [10] LI J, TAN Z C, WAN J, et al. Nested collaborative learning for long-tailed visual recognition//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-24, 2022, New Orleans, LA, USA. New York: IEEE, 2022: 6939-6948.
- [11] LI T H, CAO P, YUAN Y, et al. Targeted supervised contrastive learning for long-tailed recognition//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-24, 2022, New Orleans, LA, USA. New York: IEEE, 2022: 6908-6918.
- [12] FU S, CHU H, HE X, et al. Meta-prototype decoupled training for long-tailed learning//The Asian Conference on Computer Vision, December 4-8, 2022, Macao, China. Cham: Springer, 2022: 569-585.
- [13] YANG Y Z, XU Z. Rethinking the value of labels for improving class-imbalanced learning//The 34th International Conference on Neural Information Processing Systems, December 6-12, 2020, Vancouver, BC, Canada. New York: ACM, 2020: 19290-19301.
- [14] CARON M, MISRA I, MAIRAL J, et al. Unsupervised learning of visual features by contrasting cluster assignments. 2020: arXiv: 2006.09882.
- [15] VAN DEN OORD A, LI Y Z, VINYALS O. Representation learning with contrastive predictive coding. 2018: arXiv: 1807.03748.
- [16] DOERSCH C, ZISSERMAN A. Multi-task self-supervised visual learning//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2070-2079.
- [17] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13-19, 2020. Seattle, WA, USA. New York: IEEE, 2020: 9729-9738.
- [18] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations. 2020: arXiv: 2002.05709.
- [19] CHUANG C Y, ROBINSON J, LIN Y C, et al. Debaised contrastive learning. *Advances in Neural Information Processing Systems*, 2020, 33: 8765-8775.
- [20] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised representation learning by predicting image rotations. 2018: arXiv: 1803.07728.
- [21] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent a new approach to self-supervised learning//The 34th International Conference on Neural Information Processing Systems, December 6-12, 2020, Vancouver, BC, Canada. New York: ACM, 2020: 21271-21284.
- [22] CHEN X L, HE K M. Exploring simple Siamese representation learning//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 20-25, 2021. Nashville, TN, USA. New York: IEEE, 2021: 15745-15753.
- [23] WANG T Z, ISOLA P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere//The 37th International Conference on Machine Learning, New York: ACM, 2020: 9929-9939.
- [24] KANG B Y, XIE S N, ROHRBACH M, et al. Decoupling representation and classifier for long-tailed recognition. 2019: arXiv: 1910.09217.

- [25] WEI X S, SUN X H, SHEN Y, et al. Delving deep into simplicity bias for long-tailed image recognition. *International Journal of Computer Vision*, 2025, 133(6): 3349-3366.
- [26] LIN C S, CHEN M H, WANG Y F. Frequency-aware self-supervised long-tailed learning//2023 IEEE/CVF International Conference on Computer Vision Workshops, October 2-6, 2023, Paris, France. New York: IEEE, 2023: 963-972.
- [27] ZHOU Z, YAO J, WANG Y F, et al. Contrastive learning with boosted memorization//The 39th International Conference on Machine Learning, July 17-23, 2022, Baltimore, MD, USA. New York: PMLR, 2022: 27367-27377.
- [28] ZHENG S F, NAM J, BAUR S, et al. Joint image clustering and self-supervised representation learning through debiased contrastive loss//Medical Imaging 2025: Image Processing, February 16-21, 2025. DiegoSan, USA. Bellingham: SPIE, 2025: 39.
- [29] ZHAO Q, WU Z W, ZHANG Z Q, et al. Long-tail augmented graph contrastive learning for recommendation//Machine Learning and Knowledge Discovery in Databases: Research Track. Cham: Springer Nature Switzerland, 2023: 387-403.
- [30] XIA Z Y, JIAN M, LIU Z H, et al. Mitigating long-tail bias in recommendations via graph diffusion. *Multimedia Systems*, 2025, 31(6): 468.
- [31] SUN F H, LI G F, HE J L, et al. Intelligent diagnosis of high-speed motors under data imbalance scenarios: self-supervised feature extraction and classification optimization. *The International Journal of Advanced Manufacturing Technology*, 2025, 140(5): 3265-3278.
- [32] JIANG Z, CHEN T, MORTAZAVI B J, et al. Self-damaging contrastive learning//The 38th International Conference on Machine Learning, July 18-24, 2021, Virtual. New York: PMLR, 2021, 139: 4927-4939.
- [33] KIM D J, KE T W, YU S X. Local pseudo-attributes for long-tailed recognition. *Pattern Recognition Letters*, 2023, 172: 51-57.
- [34] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. *Mechanical Systems and Signal Processing*, 2015, 64: 100-131.
- [35] LI M K, HU Z K, LU Y, et al. Feature fusion from head to tail for long-tailed visual recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, 38(12): 13581-13589.
- [36] ZENG W, XIAO Z Y. MinoritySalMix and adaptive semantic weight compensation for long-tailed classification. *Image and Vision Computing*, 2024, 152: 105307.
- [37] PAN H L, GUO Y, YU M J, et al. Enhanced long-tailed recognition with contrastive CutMix augmentation. *IEEE Transactions on Image Processing*, 2024, 33: 4215-4230.
- [38] ZENG W, LI M. Leveraging multi-strategy labels for long-tailed classification. *Engineering Applications of Artificial Intelligence*, 2026, 166: 113563.
- [39] ZHONG J, MAO H, MO F, et al. Uniform bullseye-guided contrastive learning with time-frequency strong augmentation for long-tailed fault diagnosis. *Mechanical Systems and Signal Processing*, 2025, 230: 112639.

基于自适应原型与语义感知的长尾表征学习算法

李甜甜, 薛震*, 张亮亮, 连旭

中北大学 数学学院, 山西 太原 030051

摘要: 标注稀缺与长尾分布不均衡是工业设备监测领域所面临的关键技术挑战。现有自监督学习方法在复杂工况下易受样本数量偏差与语义混淆影响, 导致对稀疏关键状态的表征能力受限。本文提出一种协同进化的原型对比学习框架(Co-evolutionary prototypical contrastive learning, EPCL), 通过从粗粒度语义发现到细粒度判别增强的递进式学习, 实现对长尾数据内在结构的深度解析。该框架设计了基于最优传输理论的自适应原型聚类算法, 以数据驱动的动态先验实现无偏建模。而后提出语义感知的负样本层次化调控机制, 采用原型一致性约束与自适应加权策略, 实现判别边界优化的同时缓解类别不均衡。在多个公开长尾视觉基准(CIFAR10-LT、CIFAR100-LT、ImageNet-100-LT)以及工业故障诊断数据集CWRU上的实验结果表明, EPCL相较于SimCLR、SwAV等15种主流自监督方法, 在线性评估与少样本分类任务中均取得显著提升, 其在CIFAR100-LT上尾部类别准确率比SimCLR提高4.56个百分点。消融实验与可视化结果验证了所提机制的有效性与泛化能力。本研究为无监督条件下长尾测量数据的表征学习提供了具有潜力的新思路及实施方法。

关键词: 长尾分布; 自监督学习; 对比学习; 语义感知; 自适应原型聚类; 表征学习

引用格式: LI Tiantian, XUE Zhen, ZHANG Liangliang, et al. Long-tailed representation learning algorithm based on adaptive prototypes and semantic awareness. *Journal of Measurement Science and Instrumentation*, 2026, 17(2): 331-343. DOI: 10.62756/jmsi.1674-8042.2026028