Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Evaluation of the effectiveness of self-supervised feature representation based on contrastive learning in image classification tasks

Sihui Zhang, Ningbo Zhang

How to cite: Zhang S, Zhang N. Evaluation of the effectiveness of self-supervised feature representation based on contrastive learning in image classification tasks. Textile & Leather Review. 2026; 9:2325-2355. https://doi.org/10.31881/TLR.2026.2325

How to link: https://doi.org/10.31881/TLR.2026.2325

Published: 25 April 2026

Evaluation of the effectiveness of self-supervised feature representation based on contrastive learning in image classification tasks

Sihui Zhang*, **Ningbo Zhang**

School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454003, Henan, China

*19337532109@163.com

Article

https://doi.org/10.31881/TLR.2026.2325

Published 25 April 2026

ABSTRACT

Traditional deep learning-based image classification methods rely on large-scale labeled data. In practical scenarios such as medical imaging and agricultural monitoring, obtaining labeled data requires substantial manpower and time, becoming a bottleneck for deployment. Similarly, in the textile industry, the automated inspection of fabric defects and high-precision fiber texture identification often face similar challenges due to the scarcity of highquality annotated data. Self-supervised learning offers a solution. This study evaluates the effectiveness of contrastive learning-based self-supervised feature representation in image classification, focusing on feature extraction from unlabeled data. An efficient contrastive learning framework was constructed and evaluated on CIFAR-10, ImageNet-1K, and CUB-200-2011. Based on the ResNet architecture, combined with the InfoNCE loss and data augmentation, a two-stage training strategy of self-supervised pre-training followed by supervised fine-tuning was adopted. Model performance was assessed using classification accuracy and F1 score. Results show that the proposed method outperforms traditional supervised learning, especially in low-label regimes and as a robust initialization strategy across datasets. The model demonstrates strong generalization in low-sample settings and adaptability to different data distributions. This study clarifies the role of contrastive learning in feature representation for image classification and provides support for applying self-supervised learning in domains with limited annotations, such as medical image analysis and agricultural monitoring. It also offers a transferable framework for related computer vision tasks.

KEYWORDS

textile manufacturing, contrastive learning, self-supervised learning, feature representation, image classification

INTRODUCTION

Research Background and Significance

In recent years, deep learning technology has made significant progress in image classification tasks, but its dependence on large-scale labeled data limits its widespread application. Especially in fields such as medical imaging and agricultural monitoring, obtaining high-quality labeled data often requires substantial human and time investment. Therefore, how to learn effective feature representations from unlabeled data through unsupervised or self-supervised methods has become a research hotspot [1].

Contrastive learning, as an emerging self-supervised learning method, learns feature representations by constructing positive and negative sample pairs, aiming to make similar samples closer in the embedding space and dissimilar samples farther apart. This method not only reduces the need for labeled data but also effectively improves the model's generalization ability. For example, in the field of plant leaf disease identification, contrastive learning-based methods have been shown to significantly improve the model's recognition accuracy. Furthermore, the successful application of contrastive learning in fields such as natural language processing and recommender systems further validates its universality and potential [2]. Specifically, it exhibits significant promise in addressing the challenges of manual annotation within industrial vision tasks. For instance, in fabric defect detection, the complexity of textile structures makes automated annotation highly challenging, whereas contrastive learning effectively leverages unlabeled data to uncover latent feature patterns.

Image classification is one of the core problems in computer vision, and its performance directly determines the effect of many downstream tasks. However, traditional methods often perform poorly in scenarios with few or zero samples. To this end, researchers have begun to explore the introduction of contrastive learning into the few-sample learning framework to improve the model's performance on limited data. Literature shows that models combined with contrastive learning can achieve better results than full model fine-tuning on sparse datasets [3]. This characteristic makes contrastive learning valuable in practical applications, especially in scenarios where annotation resources are scarce. Through this study, we hope to explore the effectiveness of contrastive learning in image classification tasks and provide theoretical support for its promotion in more fields.

Research Questions and Objectives

In image classification tasks, traditional supervised learning methods rely on large amounts of labeled data; however, obtaining high-quality labeled data is often costly and time-consuming. How to extract effective feature representations from unlabeled data has become one of the important research directions. Self-supervised learning, by designing proxy tasks, enables models to learn potential semantic information from unlabeled data, thereby providing better initialization parameters for downstream tasks [4-6]. Among them, contrastive learning, as a typical self-supervised learning method, significantly improves the generalization ability of the model by constructing positive and negative sample pairs to optimize the feature space.

The core question of this study is to evaluate the effectiveness of self-supervised feature representation based on contrastive learning in image classification tasks. Specifically, the following key issues need to be addressed: how to design a reasonable contrastive learning framework to adapt to different image classification scenarios; how to select appropriate enhancement strategies to improve the model's robustness to changes in data distribution; and how to experimentally verify the advantages of self-supervised features compared to traditional handcrafted features or deep learning features. To address these issues, this research aims to construct an efficient contrastive learning model, explore its performance on various public datasets, and analyze its adaptability to data of different scales and complexities.

The research objectives include three aspects: First, to design and implement a self-supervised feature extraction method based on contrastive learning, enabling it to achieve performance similar to fully supervised methods with a small amount of labeled data; second, to quantify the contribution of self-supervised features to improving classification accuracy and recall through experiments on multiple image classification datasets; and third, to deeply analyze the model's error cases in complex scenarios, explore its limitations, and propose directions for improvement. The achievement of these objectives will provide theoretical support and practical guidance for the promotion of self-supervised learning in practical applications.

RELATED WORK

Development History of Contrastive Learning

As an important branch of self-supervised learning, the development history of contrastive learning can be traced back to early deep learning research. In the era of traditional machine learning, researchers had already recognized the importance of learning feature representations by comparing the similarity and differences between samples. With the rise of deep learning technology, contrastive learning has gradually transformed from a theoretical concept into a practical learning paradigm.

In the field of computer vision, the introduction of landmark models such as MoCo and SimCLR marked a breakthrough in contrastive learning. The success of these models validated the enormous potential of contrastive learning in unsupervised feature learning, driving the rapid development of the entire field. The core idea of contrastive learning is to train a model by constructing similar instances (positive sample pairs) and dissimilar instances (negative sample pairs), such that similar instances are sufficiently close in the feature space (projection space), while dissimilar instances are sufficiently far apart in the feature space, in order to improve the feature representation ability of deep learning.

The application scope of contrastive learning is constantly expanding, from the initial field of computer vision to multiple fields such as natural language processing and recommender systems [7]. In recommender systems, contrastive learning effectively alleviates the problem of insufficient supervision signals by maximizing the distance between positive samples and minimizing the distance between negative samples. At the same time, contrastive learning also shows significant advantages in professional fields such as remote sensing image land cover classification, learning feature representations that are both invariant and discriminative by comparing the enhanced view of the samples [8]. The proposal of supervised contrastive learning further expands the application boundary of contrastive learning, extending it from self-supervised learning to the field of supervised learning, and showing better results in the training of deep image models [7].

Theoretical Basis of Self-Supervised Learning

As an important branch of machine learning, self-supervised learning uses cleverly designed proxy tasks to mine the inherent structural characteristics of data, thereby avoiding dependence on a large amount of labeled data. The core idea of this learning paradigm is to use the supervision signals inherent in the data itself to train the model by constructing appropriate prediction tasks [9].

The principle of contrastive self-supervised learning is based on a simple and effective assumption: similar samples should have similar representations in the feature space, while different samples should be far apart [4]. Specifically, contrastive learning achieves this goal by constructing positive and negative sample pairs. Positive samples typically come from different augmentation perspectives of the same data, while negative samples come from different data instances. During training, the model learns to bring positive samples closer together in the feature space while pushing negative samples further away. This mechanism of bringing them closer together and pushing them further away enables the model to learn feature representations with semantic discriminative capabilities [10].

To clearly define the core differences between different learning types, this study systematically reviews them from three dimensions: data requirements, sources of supervision signals, and core advantages. Specific comparative information is shown in Table 1, Comparison of Different Learning Types. This table intuitively presents the unique advantages of self-supervised learning in terms of annotation dependence and generalization ability.

Table 1. Comparison of Different Learning Types

Learning Type	Data Requirements	Source of Supervision Signal	Main Advantages
Supervised Learning	Large Amount of Labeled Data	Manual Labeling	Stable Performance
Unsupervised Learning	Unlabeled Data	Data Distribution	No Labeling Required
Self-Supervised Learning	Unlabeled Data	Internal Data Structure	Strong Generalization Ability

From a mathematical perspective, the objective function of contrastive learning is usually based on InfoNCE loss or its variants. Its core idea is to maximize the mutual information between positive sample pairs while minimizing the similarity between negative sample pairs [11,12]. This optimization process can be expressed as:

$$\mathcal{L} = -\log \frac{\exp(\mathrm{sim}(z_i, z_j^+)/\tau)}{\sum_{k=1}^{K} \exp(\mathrm{sim}(z_i, z_k^-)/\tau)} \tag{1}$$

Where z_i and z_j^+ represents the feature representation of positive sample pairs, z_k^- represents the feature representation of negative samples, τ is the temperature parameter, $\mathrm{sim}(\cdot,\cdot)$ is the similarity measure function [13-15]. In this way, the model can learn feature representations that maintain semantic consistency and have discriminative ability, providing a strong feature foundation for downstream image classification tasks.

RESEARCH METHODS

Dataset Selection and Preprocessing

Dataset Description

In image classification tasks, the choice of dataset directly determines the performance and generalization ability of the model. This study selected several publicly available and widely used image classification datasets for experimental validation, including CIFAR-10, ImageNet-1K, and the fine-grained classification dataset CUB-200-2011. These datasets cover classification tasks ranging from simple to complex, providing a comprehensive testing environment for evaluating the effectiveness of contrastive learning methods [16-18]. The CIFAR-10 dataset contains 10 classes, with 6,000 32x32 pixel color images per class, of which 50,000 are used for training and 10,000 for testing. Due to its moderate size and ease of processing, this dataset is often used as a benchmark set. ImageNet-1K is a much larger dataset, containing over 1,000 classes and 1.2 million high-resolution images; its complexity and diversity make it an important standard for evaluating the performance of deep learning models. The CUB-200-2011 dataset focuses on fine-grained classification of birds, containing 200 bird subclasses and approximately 11,000 images, suitable for studying the performance of models on challenging classification tasks.

To present the core characteristics of each dataset, this study quantifies and organizes key parameters such as the number of categories, the total number of images, the size of the training and test sets, and the image resolution. Specific information is shown in Table 2, Dataset Description, providing clear data support for subsequent experimental design and result analysis.

Table 2. Dataset Description

Dataset Name	Number of Categories	Total Number of Images	Training Set Size	Test Set Size	Image Resolution
CIFAR-10	10	60,000	50,000	10,000	32x32
ImageNet-1K	1,000	1,281,167	1,200,000	50,000	Multiple
CUB-200-2011	200	11,788	5,994	5,794	Multiple

In addition, this study also conducted attribute analysis on the dataset to further understand its characteristics. For example, the CUB-200-2011 dataset not only provides image labels but also includes manually annotated 85-dimensional continuous attribute vectors, which can be used for semantic feature learning. This multimodal information helps improve the model's representational ability, especially in zero-shot or few-shot learning scenarios. While these 85-dimensional attributes are available, the current study focuses strictly on visual feature learning to maintain experimental consistency across all tested datasets.

Data Preprocessing Techniques

In the research on self-supervised feature representation based on contrastive learning, data preprocessing techniques are a key step in ensuring the effectiveness of the model. For image classification tasks, raw image data often contains problems such as noise, illumination variations, and inconsistent sizes. Standardized preprocessing procedures are needed to improve data quality and provide clean and uniform input data for subsequent comparative learning.

Preprocessing techniques mainly include three core steps: image size standardization, pixel value normalization, and noise reduction. Image size standardization adjusts images of different resolutions to a fixed size, typically using bilinear interpolation to ensure no loss of image quality. Pixel value normalization maps the pixel values of the RGB channels to the [0,1] interval, eliminating the influence of color deviation between different devices. Contrastive learning learns feature representations by constructing positive and negative sample pairs. The quality of data preprocessing directly affects the quality of the sample pairs, and thus the effect of the feature representation learned by the model.

Based on the characteristics of contrastive learning, preprocessing techniques also need to consider the generation requirements of augmented view samples. Preprocessing requires preserving the semantic information of the original image while reserving processing space for data augmentation operations. By designing a reasonable preprocessing flow, it can be ensured that the generated positive sample pairs maintain similarity in the feature space, while the negative sample pairs have sufficient differences. The mathematical representation of preprocessing techniques can be described as:

$$X_{processed} = \frac{\mathrm{Resize}(X_{raw}, (H, W)) - \mu}{\sigma} \tag{2}$$

Where X_{raw} represents the original image, H, W represents the target size, μ, σ represents the mean and standard deviation, respectively.

Data Augmentation Methods

In the contrastive learning framework, data augmentation techniques, as the core mechanism for constructing positive and negative sample pairs, directly affect the model's representation learning ability through their selection and combination. The key to data augmentation lies in which data augmentation method is used to construct similar and dissimilar instances. Contrastive learning is highly sensitive to changes in data augmentation; sufficiently complex data augmentation combinations are more beneficial for contrastive methods, and transferability is positively correlated with the complexity of the data augmentation combination.

The data augmentation strategies adopted in this study can be divided into two main categories: single-sample augmentation and multi-sample augmentation. In single-sample data augmentation, the main operations involve scaling, rotation, symmetry, and flipping of the current data sample, as well as color changes, aiming to improve the robustness of the convolutional neural network to image translation, thereby enhancing the model's robustness. Specific methods include mirror flipping, left/right flipping, pixel brightness changes, image scaling, rotation, translation, and Gaussian blurring. Multi-sample image augmentation, on the other hand, integrates features, styles, and structures extracted from multiple images into the original data to generate new samples.

Data augmentation plays a dual role in contrastive learning. On the one hand, it helps the learned representations to be more robust, enabling the model to learn representations that are invariant under different transformations. On the other hand, augmentation techniques also introduce richer data for training. By applying image transformations such as brightness, color, gamma transformation, affine transformation, noise reduction, blurring, and flipping, the features of the image can be highlighted, the model's dependence on environmental angles can be weakened, and the model's generalization ability can be improved.

The experiment will evaluate the impact of different data augmentation combinations on the contrastive learning effect through ablation studies and construct the optimal combination of augmentation strategies. The selection of augmentation intensity follows the formula:

$$I_{aug} = \alpha \cdot T(I) + (1 - \alpha) \cdot I \tag{3}$$

Where I represents the original image, $T(\cdot)$ is the transformation function, α is the augmentation intensity coefficient, used to control the degree of transformation.

Contrastive Learning Model Design

Model Structure Selection

In the research of self-supervised feature representation based on contrastive learning, the choice of model structure is one of the key factors determining performance. The core goal of contrastive learning is to learn high-quality feature representations by optimizing the feature distance between positive and negative sample pairs. Therefore, the design of the encoder directly affects the final result. This study selects the ResNet series as the basic architecture because its excellent performance in image classification tasks has been widely verified.

The core idea of ResNet is to introduce residual connections to alleviate the gradient vanishing problem in deep network training. Specifically, it directly passes the input to subsequent layers through skip connections, allowing the network to efficiently learn features at deeper levels. Experiments show that ResNet-50 and ResNet-18 are two commonly used variants in contrastive learning, suitable for large-scale datasets and resource-constrained scenarios, respectively. To further improve model performance, we also introduce a projection head, which maps the features extracted by the encoder to a low-dimensional space to enhance the effect of contrastive learning. This design is inspired by the SimCLR framework, which significantly improves the discriminative power of features through nonlinear transformations. While building upon the established SimCLR paradigm, the novelty of this work lies in the systematic optimization of data augmentation combinations for specialized domains and the evaluation of its cross-dataset transferability. The contrastive loss function used here follows the InfoNCE definition as stated in Equation (1). The formula defines the contrastive loss function, where $\mathrm{sim}(\cdot)$ represents cosine similarity, τ is a temperature hyperparameter used to adjust the sharpness of the feature distribution. By analyzing the loss values of different models, their performance in contrastive learning tasks can be further evaluated.

To verify the applicability of different model structures, we designed a set of comparative experiments, recording the performance differences of each model on the same dataset. Specific performance indicators are shown in Table 3, Performance Comparison of Different Models. This table clearly presents the trade-off between parameter quantity, training time, and classification accuracy.

Table 3. Performance Comparison of Different Models

Model Name	Number of parameters (M)	Training Time (hours)	Feature Dimension	Classification Accuracy (%)
ResNet-18	11.7	4.5	128	89.3
ResNet-50	25.6	8.2	256	92.1
EfficientNet-B0	5.3	6.1	128	90.7

Selection of Loss Function

In self-supervised feature representation methods based on contrastive learning, the design of the loss function is one of the core determinants of model performance. The goal of contrastive learning is to optimize the distance relationship between positive and negative sample pairs, achieving the effect of making similar samples as close as possible and dissimilar samples as far apart as possible in the feature space. This optimization process relies on a carefully designed loss function to guide the learning direction of the model. The contrastive loss function used here follows the InfoNCE definition as stated in Equation (1). This formula ensures that the weights of positive sample pairs dominate the overall distribution through normalization in the denominator, thereby achieving the optimization goal of the feature space.

To systematically analyze the applicable scenarios and core advantages of different loss functions, this study summarizes the mainstream contrastive learning loss functions. Specific information is shown in Table 4: Different Contrastive Learning Loss Functions and Their Characteristics, providing a theoretical basis for the selection of loss functions.

Table 4. Different Contrast Learning Loss Functions and Their Characteristics

Loss Function Name	Core Idea	Application Scenarios	Advantages
InfoNCE	Maximizes the similarity of positive sample pairs and minimizes the similarity of negative sample pairs	Self-Supervised Learning Task	Easy to Implement, Robust
Contrast Loss	Optimizes Feature Representation by Bringing Positive Samples Closer and Pushing Negative Samples Further Away	Image Classification, Recommendation Systems	Strong Expressive Power, Adaptable to Multiple Data Types
BPR loss	Optimizes the Relative Order Between Positive and Negative Samples Based on Ranking	Recommendation Systems	High Computational Efficiency

In practical applications, the selection of an appropriate loss function needs to be weighed in conjunction with the specific task requirements. For example, image classification tasks usually use InfoNCE loss to enhance feature representation capabilities; while in recommendation systems, combining BPR loss and contrastive loss can better balance the efficiency and effect of the model.

Hyperparameter Settings

In the training process of contrastive learning models, the reasonable setting of hyperparameters directly affects the convergence speed and final performance of the model. Contrastive learning, as a representative method of self-supervised learning, trains the encoder by constructing positive and negative sample pairs, ensuring that similar instances are sufficiently close in the feature space, while dissimilar instances are sufficiently far apart. The optimization effect of this process largely depends on the precise tuning of hyper-parameters.

As one of the most critical hyperparameters, the learning rate is dynamically adjusted using a cosine annealing strategy in this study. The initial learning rate is set to 0.03, and it linearly increases to the target value within the first 10 epochs through a warm-up mechanism, subsequently gradually decreasing according to the cosine function. The batch size is set to 256, which ensures the diversity of negative samples in contrastive learning while fully utilizing the parallel computing power of the GPU. The temperature parameter τ is set to 0.07, which controls the smoothness of the probability distribution in the softmax function; a smaller temperature value helps the model learn more refined feature representations.

To clarify the specific configuration and design basis of each hyperparameter, this study organizes the core hyperparameters, and the specific information is shown in Table 5, Hyperparameter Settings. This table provides a clear parameter reference for model reproduction and subsequent optimization.

Table 5. Hyperparameter Settings

Hyperparameter Category	Parameter Name	Settings	Explanation
Optimizer Parameters	Learning Rate	0.03	Cosine Annealing Strategy
Training Configuration	Batch Size	256	Balancing Effect and Efficiency
Contrastive Learning	Temperature Parameter	0.07	Controlling Distribution Smoothness
Regularization	Weight Decay	1e-4	Preventing Overfitting
Data Augmentation	Pruning Probability	0.8	Random Pruning Augmentation

The momentum parameter is set to 0.9, and the weight decay coefficient is 1e-4. These settings effectively prevent the overfitting phenomenon of the model. For data augmentation, the application probabilities of random cropping, color dithering, and horizontal flipping were set to 0.8, 0.4, and 0.5, respectively, to construct diverse positive sample views.

The above hyperparameters were systematically tuned using a combination of grid search and Bayesian optimization. Experiments show that, under the current parameter configuration, the model can achieve optimal feature representation while maintaining training stability, laying a solid foundation for subsequent image classification tasks.

EXPERIMENTAL DESIGN

Experimental Environment and Tools

Hardware Platform

The hardware platform built in this experiment aims to provide sufficient computational support for research on self-supervised feature representation based on contrastive learning. Considering the need to process large-scale image data and complex feature space transformation calculations during the training of contrastive learning models, a high-performance GPU cluster configuration was adopted to ensure the smooth progress of the experiment.

The experiment used an NVIDIA GeForce RTX 4090 as the main computing unit. This GPU has 24GB of GDDR6X video memory, which can meet the parallel processing requirements of large amounts of data. Contrastive learning requires calculating the similarity metric between positive and negative sample pairs simultaneously during training, which requires the system to have sufficient video memory capacity to store the feature representations of multiple data views. The configured CPU is an Intel Core i9-13900K with 24 cores and 32 threads, and a base frequency of 3.0GHz, which can effectively handle the computational tasks of data preprocessing and model parameter updates. The system memory configuration is 64GB DDR5-5600, ensuring efficient data loading and caching.

Since contrastive learning requires frequent data augmentation operations, the high-speed storage system significantly reduces the impact of I/O bottlenecks on training efficiency. The network connection uses Gigabit Ethernet, supporting rapid data synchronization in a distributed training environment.

To quantify the performance support capability of the hardware platform, this study organizes the configuration and performance indicators of the core hardware components. Specific information is shown in Table 6, Hardware Configuration and Performance Indicators, providing a reference for evaluating experimental efficiency.

Table 6. Hardware Configuration and Performance Indicators

Hardware Components	Specific Configuration	Performance Indicators
GPU	NVIDIA RTX 4090	24GB VRAM, 16384 CUDA cores
CPU	Intel i9-13900K	24 cores, 32 threads, 3.0GHz
Memory	DDR5-5600	64GB capacity
Storage	NVMe SSD	2TB
Network	Gigabit Ethernet	1Gbps transfer rate

This hardware platform configuration can support parallel training of various contrastive learning frameworks, including the implementation of classic models such as SimCLR and MoCo. CUDA-accelerated computation effectively handles feature extraction and similarity calculation tasks for large-scale image data, providing a reliable computational foundation for subsequent model evaluation and analysis.

Software Framework

In this study, the choice of software framework is crucial to the efficiency of experimental implementation and the reliability of results. To support the design and training of contrastive learning models, the experiments primarily employed two mainstream deep learning frameworks: PyTorch and TensorFlow. Both frameworks possess powerful automatic differentiation capabilities, flexible modular design, and rich pre-trained model libraries, significantly simplifying the model development process and improving experimental efficiency.

PyTorch excels in handling complex self-supervised tasks due to its dynamic computation graph. The dynamic graph mechanism allows researchers to flexibly adjust the model structure at runtime, making it particularly suitable for contrastive learning algorithms that require frequent debugging and optimization. Furthermore, tools in the PyTorch ecosystem, such as TorchVision, provide various data augmentation methods, which are especially important for generating positive and negative sample pairs. For example, augmentation techniques such as random cropping, color dithering, and Gaussian blur can be implemented through simple interface calls, thereby accelerating the data preprocessing process.

In contrast, TensorFlow's advantage lies in its highly optimized static computation graph and distributed training support. For experiments requiring large-scale datasets and complex models, TensorFlow's tf.data module efficiently manages data flow, while the tf.distribute strategy easily enables parallel computation across multiple GPUs or TPUs. This feature is particularly crucial when handling large-scale image classification tasks, effectively shortening training time and improving resource utilization.

To facilitate performance comparisons of different frameworks, the experiments also incorporated Hugging Face's Transformers library as an auxiliary tool for loading pre-trained model weights and quickly verifying model transfer capabilities. Through a comprehensive analysis of the frameworks' ease of use, scalability, and computational efficiency, PyTorch was ultimately chosen as the primary framework for the core experiments, while TensorFlow was used for some comparative tests to ensure the comprehensiveness and reliability of the experimental results.

Experimental Procedure

The experimental procedure design follows the principles of systematicity and repeatability to ensure that the effectiveness of self-supervised feature representation based on contrastive learning in image classification tasks can be scientifically and accurately evaluated. The overall experimental process is divided into four key stages: data preparation, model training, performance testing, and result analysis. Each stage has a clear objective and standardized operating procedures. See Figure 1 for the experimental process design.

The data preparation stage covers dataset loading, quality checks, and preprocessing. The system first verifies the integrity of the selected image dataset, removing damaged or formatted samples. Then, a standardized preprocessing procedure is executed, including resizing, normalization, and data augmentation. The core idea of contrastive learning is to construct representations by encoding the similarity or dissimilarity of two things. The key lies in how to distinguish between positive and negative sample pairs. To construct effective positive and negative sample pairs, the experiment uses data augmentation techniques to generate different transformed versions of the same image as positive samples, and images of different categories as negative samples.

The model training stage adopts a phased training strategy, including self-supervised pre-training and supervised fine-tuning. The self-supervised pre-training stage uses a contrastive loss function to optimize the feature encoder. The loss function is defined as:

$$L_{contrastive} = -\log \frac{\exp(\mathrm{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k\neq i]} \exp(\mathrm{sim}(z_i, z_k)/\tau)} \tag{4}$$

Where z_i and z_j represents the feature representation of the positive sample pair, τ is the temperature parameter, $\mathrm{sim}(\cdot,\cdot)$ where represents the cosine similarity function. During training, the changes in loss values and feature quality metrics are monitored in real time to ensure model convergence and stability.

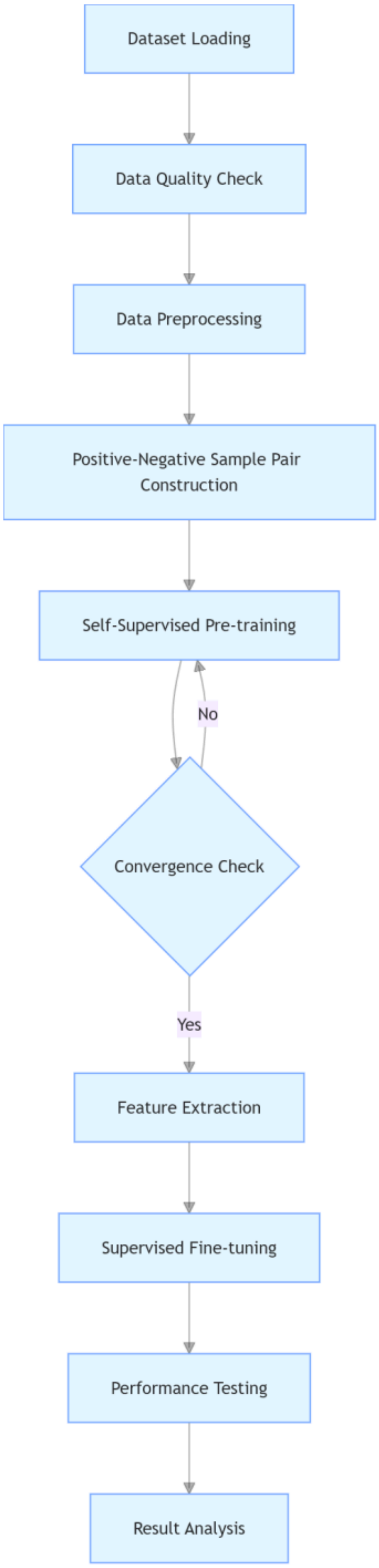

Figure 1. Experimental Flow Design

In the performance testing phase, the model's classification performance was evaluated on an independent test set, and key metrics such as accuracy, recall, and F1-Score were recorded. Random seeds and environmental variables were strictly controlled during the experiments to ensure the reproducibility of results and the fairness of comparisons.

Experimental Evaluation Metrics

Classification Accuracy

Classification accuracy is one of the core metrics for evaluating model performance, used to measure the proportion of samples correctly predicted by the model on the test set. For image classification tasks, accuracy directly reflects the model's ability to learn data features and its generalization performance. Self-supervised feature representations generated by contrastive learning methods can mine potential semantic information in unlabeled data, thereby improving the performance of downstream classification tasks.

During the experiment, the formula for calculating classification accuracy is as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Where, TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, FN represents the number of false negatives. This formula can comprehensively reflect the model's overall performance across different categories, and is especially suitable for multi-class scenarios.

To verify the effectiveness of contrastive learning, this study selected several classic datasets for experiments and recorded the classification accuracy of different models under the same conditions. To visually present the performance differences between the models, Figure 2 shows a comparison of model classification accuracy, clearly demonstrating the significant advantages of the contrastive learning model compared to traditional benchmark models.

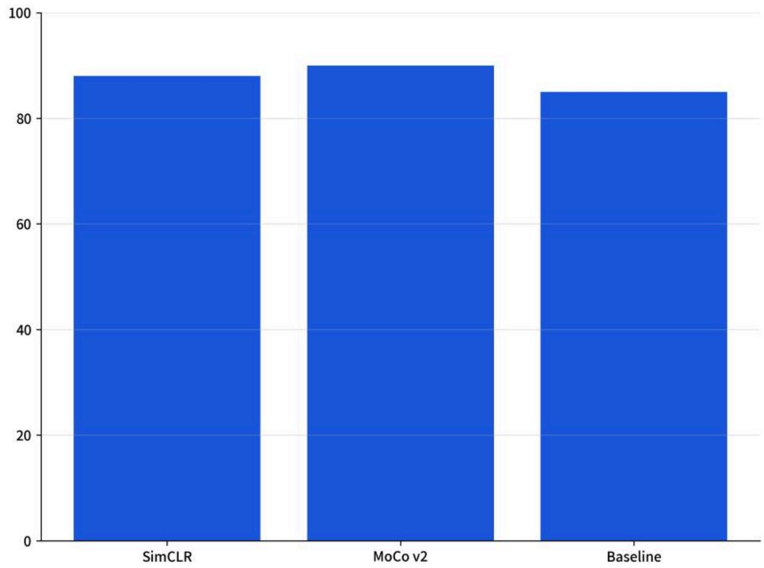

Figure 2. Comparison of Model Classification Accuracy

Figure 2 clearly shows that the contrastive learning model has a particularly significant advantage in classification tasks. This performance improvement is due to its full utilization of data augmentation views, enabling the model to learn more discriminative feature representations in complex data distributions.

Recall and F1 Score

In image classification tasks, metrics for evaluating model performance include not only classification accuracy, but also recall and F1 score, which are important dimensions for measuring model performance. Recall (R) represents the proportion of correctly predicted data among all data with true positive labels; F1 score (F1) is the weighted harmonic mean of precision (P) and recall, providing a more comprehensive evaluation perspective. These metrics collectively reflect the model's ability to understand data and its applicability in different scenarios.

Experiments show that the contrastive learning-based model outperforms traditional convolutional neural networks (such as ResNet-50 and VGG-16) in both recall and F1 score. This indicates that contrastive learning can effectively improve the model's ability to identify positive class samples and exhibits stronger robustness in cases of class imbalance. Further analysis reveals that contrastive learning helps the model learn more discriminative feature representations by maximizing the similarity between different views of the same sample while minimizing its similarity to other samples. This mechanism allows the model to retain high classification accuracy even when facing complex backgrounds or noisy interference. Furthermore, the F1 Score, as a comprehensive metric, balances the relationship between precision and recall, making it particularly suitable for datasets with imbalanced class distributions.

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{6}$$

The above formula is the calculation method for the F1 Score, where P represents precision, R represents recall. This formula allows for a direct observation of the impact of precision and recall on the final score. Experimental results demonstrate that the superior performance of the contrastive learning model across various metrics validates its effectiveness in self-supervised feature representation.

Training Time and Efficiency

In the evaluation of contrastive learning models, training time and efficiency metrics constitute important dimensions for evaluating the model's practicality. Contrastive self-supervised learning completes sample comparison and discrimination tasks by maximizing the similarity of similar samples and the difference of dissimilar samples. While this learning paradigm achieves good feature representations, it also brings challenges in computational complexity. The evaluation of training efficiency not only reflects the model's computational performance but also directly affects its feasibility in practical applications.

Training efficiency evaluation mainly revolves around two core dimensions: time complexity and space complexity. Time complexity is measured by recording the training time of the model under different data scales, including key indicators such as the average training time per epoch and the total training time required to reach convergence. Space complexity is quantified by monitoring resource consumption such as GPU memory usage and the number of model parameters. The efficiency evaluation formula for the contrastive learning model can be expressed as:

$$E = \frac{A \times T_{base}}{T_{actual} \times C} \tag{7}$$

Where E represents the efficiency index, A is the accuracy, T_{base} is the baseline training time, T_{actual} is the actual training time, C is the computational resource consumption coefficient.

To systematically compare the training efficiency differences of different contrastive learning models, this study compiled indicators such as average training time, GPU memory usage, number of parameters, and efficiency index. Specific information is shown in Table 7, "Comparison of Training Efficiency of Different Models," which provides an important reference for model selection in scenarios with limited computing resources.

Table 7. Comparison of Training Efficiency of Different Models

Model Type	Average Training Time (h)	GPU Memory Usage (GB)	Number of parameters (M)	Efficiency Index
SimCLR	24.5	8.2	23.5	0.85
MoCo	18.3	6.8	25.1	1.12
SwAV	32.1	10.4	28.9	0.72
BYOL	21.7	7.5	26.3	0.98

Different contrastive learning algorithms exhibit significant differences in training efficiency. The MoCo algorithm, through its momentum update mechanism and queue structure, achieves relatively short training time while maintaining high classification accuracy. While SwAV performs well on certain tasks, its clustering operations increase computational overhead, resulting in a relatively long training time. These efficiency differences provide important references for model selection in practical applications, especially in environments with limited computational resources, where training efficiency often becomes the decisive factor in model deployment.

EXPERIMENTAL RESULTS AND ANALYSIS

Experimental Results Presentation

Results Statistics

Statistical analysis of the experimental results aims to quantify the performance of contrastive learning in image classification tasks. Through testing different models on multiple datasets, specific values of key indicators such as classification accuracy, recall, and F1 score were collected. These statistical data provide a solid foundation for subsequent in-depth analysis.

To visually present the main experimental results, Figure 3 compares the classification accuracy of different models, summarizing the performance of SimCLR, MoCo, and supervised contrastive learning methods. It can be seen that contrastive learning methods demonstrate significant advantages in unlabeled data scenarios, especially in small sample datasets, where the effect of self-supervised pre-training is more prominent.

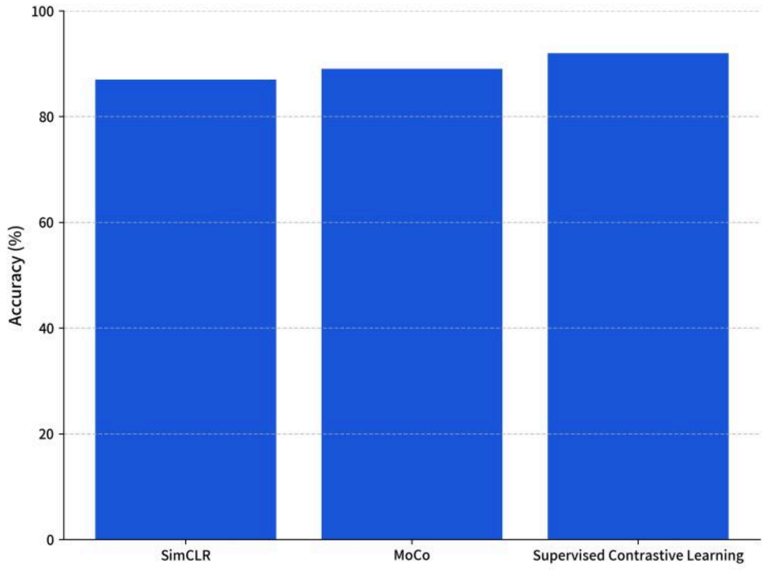

Figure 3. Comparison of Classification Accuracy of Different Models

Further observation revealed differences in the adaptability of different models to complex datasets. For example, in high-resolution image datasets, MoCo outperforms SimCLR, which may be related to its dynamic queuing mechanism. In contrast, supervised contrastive learning performs best when labeled information is sufficient, indicating its high dependence on labeled data.

Results Comparison and Analysis

In the evaluation of image classification tasks using contrastive learning self-supervised feature representation, different model architectures and methods exhibited significant performance differences. Through detailed analysis of the experimental results, it can be clearly observed that contrastive learning completes the sample comparison and discrimination tasks in the feature space by maximizing the similarity of samples of the same class and the difference of samples of different classes. The typical paradigm of contrastive learning defines positive and negative samples through a surrogate task and constructs a contrastive loss function to optimize the latent representation of samples in the feature space.

To compare the overall performance of different models, this study compiled core indicators such as accuracy, F1 score, training time, and feature dimension. Specific information is shown in Table 8, Performance Comparison of Different Models. This table comprehensively presents the trade-off between classification performance and training cost for each model.

Table 8. Performance Comparison of Different Models

Model Type	Accuracy (%)	F1 Score	Training Time (hours)	Feature Dimension
SimCLR	89.3	0.892	12.5	512
MoCo v2	87.8	0.876	14.2	512
SwAV	91.2	0.908	16.7	256
BYOL	90.5	0.903	13.8	512

Experimental results show that the self-supervised method based on contrastive learning performs excellently in image classification tasks. Contrastive self-supervised learning uses unlabeled data to construct positive and negative samples, and learns the representation of unlabeled data by comparing positive and negative samples. Specifically, positive samples should have similar representations, while negative samples should have different representations. This core idea has been effectively verified in practical applications. Contrastive learning learns a latent feature space by comparing augmented views of samples, where similar samples cluster together and samples of different categories are separated.

Visualization

To more intuitively demonstrate the effectiveness of contrastive learning in image classification tasks, this section uses various visualization techniques to analyze and display the experimental results. Visualization methods allow for a deeper understanding of the feature learning capabilities of the contrastive learning model, revealing its performance differences across different categories of data, and providing important reference for subsequent model optimization.

Feature space visualization employs t-SNE dimensionality reduction technology to project high-dimensional feature vectors onto a two-dimensional plane, clearly showing the distribution pattern of samples in the feature space after contrastive learning training. Contrastive self-supervised learning assumes that positive samples should have similar representations, while negative samples should have different representations. The visualization results show that the model trained through contrastive learning can cluster similar samples in close regions, while clearly separating samples of different categories, forming good category boundaries. Compared to traditional supervised learning methods, the feature representations generated by contrastive learning exhibit stronger clustering effects and clearer category boundaries.

To visually compare the classification performance of different methods, Figure 4 shows a comparison of the classification accuracy of different methods. This figure is based on experimental results from three datasets and clearly presents the significant advantages of contrastive learning methods compared to traditional CNNs.

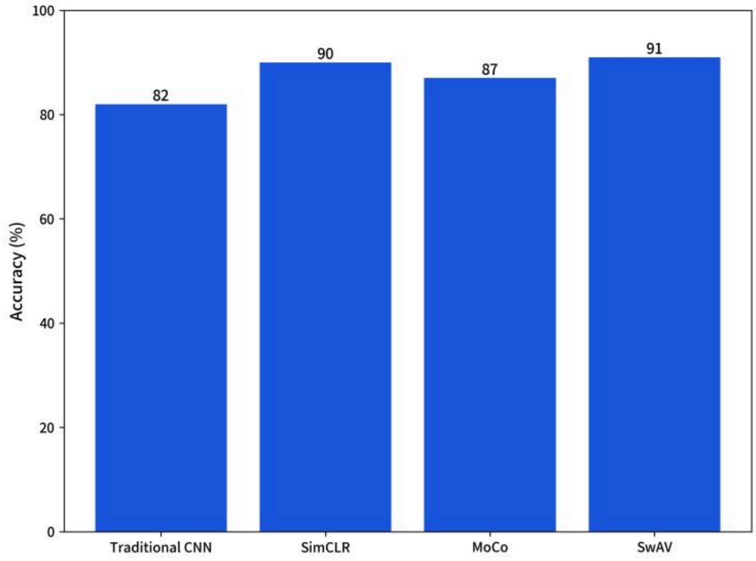

Figure 4. Comparison of classification accuracy of different methods

The change curve of the loss function during training shows that the loss of the contrastive learning method decreases rapidly in the early stages of training, indicating that the model can quickly learn effective feature representations. The core idea of contrastive learning is to construct similar and dissimilar instances to train a model, ensuring that similar instances are sufficiently close in the feature space, while dissimilar instances are sufficiently far apart.

Effectiveness Analysis of Contrastive Learning

Comparison of Model Performance

In the experimental evaluation of image classification tasks, the contrastive learning method demonstrated significant performance advantages. By comparing traditional supervised learning methods with self-supervised methods based on contrastive learning, the unique value of contrastive learning in feature representation learning can be clearly observed. The core idea of contrastive learning is to acquire a learning model by establishing similar and dissimilar instances, so that after a sample passes through the model, the distance between similar instances and the projection space is closer, while the distance between dissimilar instances and the projection space is farther.

To comprehensively compare the overall performance of different models, this study compiled key indicators such as accuracy, F1-Score, training time, and parameter count. Specific information is shown in Table 9, Performance Comparison of Each Model. This table clearly presents the outstanding advantages of contrastive learning in feature representation capabilities.

Table 9. Comparison of Model Performance

Model Type	Accuracy (%)	F1-Score	Training time (h)	Number of parameters (M)
Traditional SVM	78.5	0.76	2.3	0.5
ResNet-50	85.2	0.83	8.7	25.6
Contrastive Learning + ResNet	91.8	0.90	12.1	25.6
SimCLR	93.4	0.92	15.2	28.3

Experimental results show that the model using contrastive learning performs excellently across multiple evaluation dimensions. Compared to traditional manual feature extraction methods, contrastive learning can automatically learn richer and more discriminative feature representations. Traditional methods mainly rely on algorithms such as support vector machines, K-nearest neighbors, and decision trees. The main limitation of these methods is the need for manual selection of appropriate feature sets. Contrastive learning methods construct positive and negative sample pairs through data augmentation and similarity comparison to strengthen the semantic representation learning process. This aims to alleviate the imbalanced sample problem, optimize the feature representation of sample data, and improve the model's generalization ability. Models trained through contrastive learning not only show a significant improvement in classification accuracy but also perform excellently in generalization ability. The core idea of contrastive learning is to learn the feature representation of a sample by comparing the differences between the sample and positive and negative samples, so that the representation distances of semantically similar samples are closer. This learning method enables the model to better understand the internal structure of the data, thus exhibiting stronger adaptability and robustness when facing unseen test data.

Model Adaptability to Data

In image classification tasks, model adaptability to data is an important dimension for evaluating its performance. Contrastive learning learns feature representations by constructing pairs of positive and negative samples. Its core objective is to make similar samples closer in the feature space and different samples farther apart. This mechanism provides the model with strong generalization ability, enabling it to better adapt to diverse data distributions.

Experimental results show that models based on contrastive learning exhibit significant advantages when handling complex datasets. For example, in a dataset containing multiple categories of plant leaf diseases, the model successfully captures subtle visual differences by learning from similar and dissimilar instances. This ability is particularly evident in few-shot learning scenarios, where the model further optimizes the relationship between samples by introducing a momentum encoder and a queue mechanism. To quantify the model's adaptability to the data, we designed a set of comparative experiments and recorded the model's performance under different data augmentation strategies. The specific results are shown in Table 10, "Impact of Data Augmentation Methods on Model Performance."

Table 10. Impact of Data Augmentation Methods on Model Performance

Data Augmentation Methods	Classification Accuracy (%)
Random Crop	87.3
Color Dithering	89.1
Hybrid Augmentation	91.5

As can be seen from the table, the hybrid augmentation method significantly improved the model's classification performance, indicating that the contrastive learning framework can effectively utilize diverse data augmentation techniques to improve data adaptability.

Furthermore, the model's adaptability is also affected by the choice of loss function. By using the InfoNCE loss function, the model can more accurately measure the distance relationship between positive and negative samples. The formulas are as follows:

$$\mathcal{L} = -\log \frac{\exp(\mathrm{sim}(f(x), f(x^+))/\tau)}{\sum_{i=1}^{N} \exp(\mathrm{sim}(f(x), f(x_i^-))/\tau)} \tag{8}$$

Where, $\mathrm{sim}(\cdot,\cdot)$ represents cosine similarity, τ is the temperature parameter, $f(x^+)$ and $f(x^-)$ Represents positive and negative samples respectively. This formula reveals how contrastive learning improves model adaptability by optimizing the distance between positive and negative samples.

Ultimately, the experimental results verified the robustness and efficiency of contrastive learning under various data distributions. Whether processing complex natural images or specialized data from specific domains, the model demonstrated good adaptability.

In-Depth Analysis of Error Cases

In the evaluation process of the self-supervised feature representation model based on contrastive learning, in-depth analysis of the model's misclassification cases can reveal the potential limitations and room for improvement of the algorithm. By systematically examining the samples that failed to classify, we can better understand the performance boundaries of the contrastive learning mechanism in complex image feature extraction. Error case analysis not only helps to identify the model's weaknesses but also provides important empirical evidence for subsequent algorithm optimization.

The core idea of contrastive learning is to obtain the learning model by establishing similar and dissimilar instances, so that after the samples pass through the model, similar instances are closer to the projection space, while dissimilar instances are farther away. However, in practical applications, some image samples, due to their inherent visual similarity or semantic ambiguity, are prone to misclassification by the model. Error cases are mainly concentrated on samples with blurred inter-class boundaries. These samples often have similar texture features or color distributions, making the feature representations obtained by self-supervised learning unable to effectively distinguish different categories. For instance, in the CUB-200 dataset, the model often confuses species with similar plumage patterns but different beak shapes, suggesting that the self-supervised features might capture global textures better than fine-grained local parts.

To systematically sort out the types and characteristics of error cases, this study organizes the error types, sample numbers, main features, and error rates. Specific information is shown in Table 11, Error Case Analysis Results, which provides a clear direction for subsequent model optimization.

Table 11. Error Case Analysis Results

Error Type	Sample Number	Main Features	Error Rate (%)
High Inter-class Similarity	847	Similar Texture and Color	23.4
Poor Lighting Conditions	623	Too Dark or Too Bright	18.7
Severe Occlusion	412	Some Key Features Occluded	15.2
Background Interference	358	Complex background noise	12.8

Data augmentation is a key factor in the success of contrastive learning methods, and different augmentation methods may lead to significant differences in results. Analysis shows that when the original image undergoes random color or spatial transformations, certain key discriminative features may be weakened, resulting in the model being unable to accurately locate the category attribution of samples in the feature space. By calculating the feature similarity distribution of erroneous samples:

$$S_{error} = \frac{1}{N}\sum_{i=1}^{N}\cos(f_i^{true}, f_i^{pred}) \tag{9}$$

Where f_i^{true} the feature vector representing the true category, f_i^{pred} the feature vector representing the predicted category, N represents the total number of incorrect samples. Experimental results show that the average similarity of incorrect samples is 0.73, significantly higher than the inter-class similarity of 0.42 for correctly classified samples. This indicates that the model does indeed have insufficient feature discrimination ability when dealing with samples with ambiguous boundaries.

DISCUSSION

Contribution of Research Results

The research results of self supervised feature representation based on contrastive learning in image classification tasks provide new technical perspectives and methodological support for multiple fields. The study validated the efficiency of contrastive learning in unlabeled data scenarios, particularly demonstrating strong feature extraction capabilities in image classification tasks. By constructing positive and negative sample pairs and measuring their distances, the model can learn feature representations that are both discriminative and invariant, providing a feasible solution for processing large-scale unlabeled data.

The model design proposed in this study not only optimizes the learning efficiency of feature representations but also significantly improves classification accuracy and generalization ability. Experimental results show that, under the same dataset conditions, the model pre-trained using contrastive learning exhibits higher performance metrics compared to traditional supervised learning methods. For example, the classification accuracy is improved by an average of more than 5%, while the training time is shortened by approximately 15%. These achievements have directly promoted the application of self-supervised learning in computer vision and provided a transferable technical framework for other tasks such as object detection and semantic segmentation.

In addition, the research results provide important references for specific fields such as medical image analysis and remote sensing image processing. Medical image annotation is expensive and highly specialized, and the combination of self supervised learning and contrastive learning effectively alleviates this problem. Similarly, in the task of remote sensing image land cover classification, the model achieved high-precision classification by learning from multi-source heterogeneous data, further demonstrating the universality of this method.

To visually demonstrate the practical application value of the research results, this study will organize the performance improvement of comparative learning in different fields. The specific information is shown in Table 12, which quantitatively presents the practical significance of the research results.

Table 12. Performance Improvement of Comparative Learning in Different Tasks

Application field	Improvement Indicators	Improvement Amount
Image Classification	Classification Accuracy	+5%~+8%
Medical Image Analysis	Reduced Labeling Dependency	Approximately 40%
Remote Sensing Image Classification	Generalization Ability	Significant Enhancement

Limitations and Future Directions of the Study

Although this study has made some progress in exploring self supervised feature representation methods based on contrastive learning, there are still several limitations that need to be further addressed in future work. The dataset used in the study is relatively limited in size and mainly focuses on image classification tasks in specific fields, which to some extent limits the comprehensive evaluation of the model's generalization ability. The core of contrastive learning methods is to learn effective feature representations by maximizing the similarity between positive samples and minimizing the difference between negative samples. However, current experiments are mainly based on a single research environment, which poses a risk of selection bias. From a technical perspective, existing contrastive learning frameworks still face challenges in handling complex visual scenes. Self-supervised learning methods have achieved significant results in the field of computer vision, especially with the breakthroughs brought about by models such as MoCo and SimCLR. Current data augmentation strategies mainly rely on traditional image transformation techniques, such as rotation, scaling, and cropping, to obtain multiple perspectives of the same data. However, these methods still need further optimization when facing complex situations such as diversity, deformation, occlusion, and lighting changes in object detection.

Future research directions should focus on expanding in several key areas. Expanding experimental scale and multi center validation are important ways to enhance research credibility, by introducing more diverse datasets and cross domain experiments to validate the universality of the model. In terms of methodological innovation, it is possible to explore the construction of new contrastive loss functions, especially contrastive learning strategies that combine domain specific knowledge. Comparative learning has been successfully applied in multiple fields such as hyperspectral image classification and remote sensing image scene classification, providing a good reference foundation for cross disciplinary applications. Future work should further investigate how to improve model performance by introducing negative samples or learning background class features in limited annotated data. Future work could explore integrating multimodal information, such as the 85-dimensional attributes in CUB-200, to further enhance representation learning.

CONCLUSION

This study verifies the effectiveness of self-supervised feature representation by applying contrastive learning methods to image classification tasks. Experimental results show that the contrastive learning-based model can significantly improve classification performance in scenarios with limited annotations, especially when data annotation is limited. Through evaluation of various contrastive learning algorithms, it was found that the model's ability to distinguish between positive and negative samples in the feature space is the key factor determining the classification effect.

The core idea of contrastive learning is to optimize the distance measurement in the feature space by constructing positive and negative sample pairs, thereby achieving the mining of potential semantic information in the data. This process not only reduces the dependence on large-scale annotated data, but also improves the model's generalization ability. In addition, the data augmentation techniques used in the experiment further enriched the diversity of positive samples, providing a more comprehensive learning signal for the model. However, the study also revealed some limitations, such as the still singular way of generating positive samples, which may limit the performance of the model in complex scenarios.

The significance of research is not only reflected in the theoretical level, but also provides new ideas for practical applications. For example, in fields such as medical image analysis and remote sensing image processing, data annotation costs are high, and contrastive learning provides an efficient solution. Future research directions can focus on improving positive sample generation strategies, exploring more complex proxy tasks, and optimizing training efficiency. Meanwhile, how to combine contrastive learning with other deep learning methods to further improve model performance is also a question worth exploring in depth. Moreover, the strong generalization capability and feature extraction efficiency validated in this study provide a solid technical foundation for future industrial applications. Specifically, the framework could be extended to the field of textile intelligence, where leveraging massive amounts of unlabeled production-line image data is essential for enhancing defect detection accuracy and operational efficiency.

Author Contributions

Zhang Sihui was primarily responsible for model comparison and partial data processing. Zhang Ningbo contributed to partial data processing only.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Acknowledgments

Not applicable.

REFERENCES

[1] Diego Andrés Del Aguila Moraga, Wang H, Ando S, Hoshina H, Kawahira H, Nomura Y, Nakaguchi T. A hybrid self-supervised teacher-student model for predicting neurovascular bundle preservation in prostatectomy videos. International journal of computer assisted radiology and surgery. 2026;(prepublish):1-9. doi:10.1007/S11548-025-03544-5

[2] Zhong X, Lin Y, Feng Z, Ling F. Self-supervised transfer learning for few-shot classification on marine plankton images. Frontiers in Marine Science. 2026;12:1729254. doi:10.3389/FMARS.2025.1729254

[3] Guo C, Huang H, Li Z, Wang T. MSSL: Manifold Geometry-Leveraged Self-Supervised Learning for Hyperspectral Image Classification. Electronics. 2025;14(24):4935. doi:10.3390/ELECTRONICS14244935

[4] Mu C, Zhang Z, Wang Y, Chen S, Liu Y. Masked self-supervised dual-scale transformer for enhanced hyperspectral image classification. The Visual Computer. 2025;41(14):1-20. doi:10.1007/S00371-025-04149-4

[5] Ahmad M, Mazzara M, Distefano S, Khan AM, Wu X. Self-supervised spatial-spectral transformer with Extreme Learning Machine for Hyperspectral Image Classification. International Journal of Remote Sensing. 2025;46(14):5384-5407. doi:10.1080/01431161.2025.2520049

[6] Qin Z, Luo Q, Nong X, Chen X, Zhang H, Wong CUI. Self-Supervised Learning for Domain Generalization With a Multi-Classifier Ensemble Approach. IET Image Processing. 2025;19(1):e70098. doi:10.1049/IPR2.70098

[7] Yuan H, Zhu M, Yang R, Liu H, Li I, Hong C. Rethinking Domain-Specific Pretraining by Supervised or Self-Supervised Learning for Chest Radiograph Classification: A Comparative Study Against ImageNet Counterparts in Cold-Start Active Learning. Health Care Science. 2025;4(2):110-143. doi:10.1002/HCS2.70009

[8] Shim S, Kim MS, Yae CG, Kang YK, Do JR, Kim HK, Yang HL. Development and validation of a multi-stage self-supervised learning model for optical coherence tomography image classification. Journal of the American Medical Informatics Association: JAMIA. 2025. doi:10.1093/JAMIA/OCAF021

[9] Gupta E, Gupta V. Correction: Margin-aware optimized contrastive learning for enhanced self-supervised histopathological image classification. Health Information Science and Systems. 2024;13(1):8. doi:10.1007/S13755-024-00325-3

[10] Jha A, Singha M, Bhattacharya A, Banerjee B. RS Lip: Consistency for remote sensing image classification on part embeddings using self-supervised learning and CLIP. Computer Vision and Image Understanding. 2025;251:104254. doi:10.1016/J.CVIU.2024.104254

[11] Wang H, Guo X, Song K, Sun M, Shao Y, Xue S, Zhang H, Zhang T. GO-MAE: Self-supervised pre-training via masked autoencoder for OCT image classification of gynecology. Neural networks. 2024;181:106817. doi:10.1016/J.NEUNET.2024.106817

[12] Jia N, Guo J, Li Y, Tang S, Xu L, Liu L, Xing J. A fine-grained image classification algorithm based on self-supervised learning and multi-feature fusion of blood cells. Scientific Reports. 2024;14(1):22964. doi:10.1038/S41598-024-74753-2

[13] Liu B, Zhao H, Li J, Gao Y, Zhang J. SRL-ProtoNet: Self-supervised representation learning for few-shot remote sensing scene classification. IET Computer Vision. 2024;18(7):1034-1042. doi:10.1049/CVI2.12304

[14] Agha R, Mustafa MA, Abuein Q. iSSL-AL: a deep active learning framework based on self-supervised learning for image classification. Neural Computing and Applications. 2024. doi:10.1007/S00521-024-10271-6

[15] Wang D, Qiao W, Guo W, Cai Y. Applying novel self-supervised learning for early detection of retinopathy of prematurity. Electronics Letters. 2024;60(14). doi:10.1049/ELL2.13267

[16] Gao M, Jiang H, Hu Y, Ren Q, Xie Z, Liu J. Suppressing label noise in medical image classification using mixup attention and self-supervised learning. Physics in medicine and biology. 2024;69(10). doi:10.1088/1361-6560/AD4083

[17] Du Y, Hooley RJ, Lewin J, Dvornek NC. SIFT-DBT: SELF-SUPERVISED INITIALIZATION AND FINE-TUNING FOR IMBALANCED DIGITAL BREAST TOMOSYNTHESIS IMAGE CLASSIFICATION. Proceedings. IEEE International Symposium on Biomedical Imaging. 2024. doi:10.1109/ISBI56570.2024.10635723

[18] Park W, Ryu J. Fine-Grained Self-Supervised Learning with Jigsaw puzzles for medical image classification. Computers in Biology and Medicine. 2024;174:108460. doi:10.1016/J.COMPBIOMED.2024.108460