



Self Supervised Learning for Image Classification With Few Labels

Wei Ming Tan¹, Rachel Lim², Aisha Kumar^{3*}

School of Computing, National University of Singapore, Singapore 117417, Singapore

***Corresponding author:** aisha.kumar@nus.edu.sg

Abstract

Labeled data are often limited in image classification because annotation is costly and slow. This study proposes a self-supervised learning method for image classification with few labels. The method uses multi-view data augmentation, SimCLR and MoCo pretraining, and pseudo-label fine-tuning to learn useful features from unlabeled images. Experiments were conducted on CIFAR-100 and STL-10 with only 10% labeled data. Compared with supervised training under the same label setting, the proposed method improved classification accuracy by about 9.7%. The accuracy gap from full-supervised training was less than 2.5%, and normalized mutual information increased by about 13%. These results show that contrastive pretraining and pseudo-label learning can reduce the effect of label shortage and improve feature clustering. The method provides a practical way to improve image classification when labeled samples are limited.

Keywords: *self-supervised learning, contrastive learning, pseudo-labeling, low-label learning, feature learning, feature clustering, image classification.*

INTRODUCTION

Deep learning models usually require a large number of labeled samples to achieve stable performance. However, high-quality labels are costly and time-consuming to obtain in many vision tasks, especially in medical imaging, remote sensing, and industrial inspection, where annotation often depends on expert knowledge [1,2]. When labeled data are limited, supervised models may overfit the training set and learn weak or task-specific features. This problem is also evident in sparse-data reconstruction tasks, where reliable visual recovery depends on extracting meaningful structural information from limited measurements rather than relying only on direct data fitting [3]. Therefore, reducing the dependence on labeled samples while improving feature quality has become an important topic in visual representation learning [4,5]. Self-supervised learning provides an effective solution by learning general visual features from unlabeled data [6]. Instead of using manual labels, self-supervised methods construct supervision signals from the data itself, allowing models to capture semantic and structural information before downstream fine-tuning [7,8]. Contrastive learning is one of the most widely used self-supervised strategies. It learns representations by pulling different augmented views of the same image closer in the feature space and pushing representations from different images apart [9,10]. Methods such as SimCLR and MoCo have shown strong performance in image representation learning, and later studies have improved contrastive learning through stronger augmentation strategies, better negative-sample design, and refined loss functions [11,12]. These studies show that useful visual features can be learned even when labeled data are scarce [13]. Pseudo-labeling is another important strategy for improving model performance under limited supervision. It assigns temporary labels to high-confidence unlabeled samples and then uses them for

semi-supervised fine-tuning [14,15]. This process can expand the effective training set and improve classification accuracy when only a small portion of labeled data is available [16,17]. However, pseudo-labeling also has clear limitations. If the initial model produces incorrect high-confidence predictions, these wrong pseudo labels may be reinforced during training and reduce model reliability [18]. This risk is more serious when labeled data are extremely limited or when class boundaries are difficult to separate [19]. Therefore, pseudo-labeling needs a stronger feature representation as its foundation. Current studies still have several limitations. Many self-supervised and semi-supervised methods rely on large-scale datasets, strong pretrained backbones, or carefully designed training settings. Some studies mainly report classification accuracy but provide limited analysis of feature clustering quality. In addition, contrastive learning is sensitive to augmentation design and negative-sample selection, which may affect the stability of learned representations. To address these problems, this study proposes a self-supervised image-classification framework that combines multi-view augmentation, SimCLR and MoCo pretraining, and pseudo-label fine-tuning. Experiments are conducted on CIFAR-100 and STL-10 using only 10% labeled data. The proposed method improves classification accuracy by about 9.7%, keeps the performance gap from full supervised training below 2.5%, and increases NMI by about 13%. These results indicate that the proposed framework can improve both classification performance and feature clustering quality under limited-label conditions. This study provides a practical approach for image classification tasks where labeled data are difficult to obtain but unlabeled images are available.

2. Materials and Methods

2.1 Samples and Study Objects

Two public image datasets were used in this study. CIFAR-100 includes 60,000 color images from 100 classes, with 50,000 images for training and 10,000 images for testing. STL-10 includes 5,000 labeled training images, 8,000 test images, and 100,000 unlabeled images from 10 classes. To build a low-label setting, only 10% of labeled training images were used with labels. The remaining labeled images were used as unlabeled data during training. The study object was an image classification model trained with limited labeled data and a large amount of unlabeled data. CIFAR-100 was used to test fine-grained classification, while STL-10 was used to test feature learning with more unlabeled samples.

2.2 Experimental Design and Control Settings

Four groups were used for comparison. The first group was a supervised baseline trained only with 10% labeled data. The second group used SimCLR pretraining and supervised fine-tuning. The third group used MoCo pretraining and supervised fine-tuning. The proposed group used multi-view augmentation, SimCLR and MoCo pretraining, and pseudo-label fine-tuning. A full-supervised model trained with all labels was also used as the upper reference. This design was used to compare the effects of contrastive pretraining and pseudo-label fine-tuning under limited labels. All groups used the same backbone network, optimizer, batch size, training epochs, and learning-rate schedule.

2.3 Measurement Methods and Quality Control

Model performance was measured by classification accuracy, the accuracy gap from full-supervised training, normalized mutual information, and pseudo-label reliability. Classification accuracy was calculated on the test set. Normalized mutual information was used to measure whether learned features formed clear class clusters. Pseudo-label reliability was calculated from high-confidence predictions during fine-tuning. For quality control, the 10% labeled subset was selected by stratified sampling to keep the class ratio consistent with the full dataset. Each experiment was repeated three times with different random seeds. The mean and standard deviation were reported. The test set was not used during training. The same preprocessing, model initialization, and evaluation protocol were used for all groups.

2.4 Data Processing and Model Formulas

All images were normalized by channel mean and standard deviation. Random cropping, color jittering, Gaussian blur, and horizontal flipping were used to create different views of the same image

during contrastive pretraining. During fine-tuning, high-confidence predictions were used as pseudo labels. The contrastive loss was defined as:

$$L_{\text{con}} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^K \exp(\text{sim}(z_i, z_k)/\tau)}$$

where z_i and z_j are feature embeddings from two views of the same image, $\text{sim}(\cdot)$ is cosine similarity, τ is the temperature parameter, and K is the number of comparison samples. The fine-tuning loss was defined as:

$$L_{\text{total}} = L_{\text{ce}}(x_l, y_l) + \lambda L_{\text{ce}}(x_u, \hat{y}_u)$$

where x_l and y_l are labeled samples and true labels, x_u is an unlabeled sample, $\hat{y}_u$ is its pseudo label, and λ is the weight of pseudo-label loss. Only pseudo labels above the confidence threshold were used.

2.5 Model Training and Evaluation Protocol

The model was trained in two stages. In the first stage, labeled and unlabeled images were used together for contrastive pretraining without class labels. SimCLR was used to learn instance-level features, and MoCo was used to improve feature consistency with a momentum encoder. In the second stage, the pretrained model was fine-tuned with 10% labeled data and selected pseudo-labeled samples. The same pseudo-label confidence threshold was used for both datasets. After training, the model was tested using classification accuracy and normalized mutual information. The proposed method was compared with the supervised baseline, SimCLR-only model, MoCo-only model, and full-supervised reference model. A method was considered better when it had higher test accuracy, a smaller gap from full-supervised training, higher NMI, and stable results across repeated runs.

3. Results and Discussion

3.1 Classification Accuracy Under Limited Labels

The proposed method improved classification accuracy on both CIFAR-100 and STL-10 when only 10% of labeled data were used. Compared with the supervised baseline trained with the same labeled subset, accuracy increased by about 9.7%. The gap from full-supervised training was less than 2.5%, which shows that unlabeled data reduced the effect of label shortage. The improvement was clearer on STL-10 because it contains a large unlabeled set for pretraining. CIFAR-100 was still more difficult because it contains 100 fine-grained classes with smaller differences between categories. This result is consistent with recent semi-supervised studies showing that pseudo-labeling and contrastive learning can improve performance under low-label settings [20,21]. Gao et al. (2026) also reported that contrastive semi-supervised learning improved results on CIFAR-10/100, SVHN, and STL-10 when labeled samples were limited. Fig. 1 shows a related performance comparison on these datasets.

Table 1. Error rate for CIFAR-10/100, SVHN, and STL-10 datasets on 3 different labeled data settings. **Bold** highlights optimal outcome and underline denotes the second-best performance.

Dataset	CIFAR-10			CIFAR-100			SVHN			STL-10		
Label Amount	40	250	4000	400	2500	10,000	40	250	1000	40	250	1000
II Model [37]	74.34 ± 1.76	46.24 ± 1.29	13.13 ± 0.59	86.96 ± 0.80	58.80 ± 0.66	36.65 ± 0.00	67.48 ± 0.95	13.30 ± 1.12	7.16 ± 0.11	74.31 ± 0.85	55.13 ± 1.50	32.78 ± 0.40
Pseudo Label [9]	74.61 ± 0.26	46.49 ± 2.20	15.08 ± 0.19	87.45 ± 0.85	57.74 ± 0.28	36.55 ± 0.24	64.61 ± 5.60	15.59 ± 0.95	9.40 ± 0.32	74.68 ± 0.99	55.45 ± 2.43	32.64 ± 0.71
VAT [38]	74.66 ± 2.12	41.03 ± 1.79	10.51 ± 0.12	85.20 ± 1.40	46.84 ± 0.79	32.14 ± 0.19	74.75 ± 3.38	4.33 ± 0.12	4.11 ± 0.20	74.74 ± 0.38	56.42 ± 1.97	37.95 ± 1.12
Mean Teacher [39]	70.09 ± 1.60	37.46 ± 3.30	8.10 ± 0.21	81.11 ± 1.44	45.17 ± 1.06	31.75 ± 0.23	36.09 ± 3.98	3.45 ± 0.03	3.27 ± 0.05	71.72 ± 1.45	56.49 ± 2.75	33.90 ± 1.37
MixMatch [36]	38.84 ± 8.36	20.96 ± 2.45	10.25 ± 0.01	80.58 ± 3.38	47.88 ± 0.21	33.22 ± 0.06	26.61 ± 13.10	4.48 ± 0.35	5.01 ± 0.12	52.32 ± 0.91	36.34 ± 0.84	25.01 ± 0.43
ReMixMatch [4]	8.13 ± 0.58	6.34 ± 0.22	4.65 ± 0.09	41.60 ± 1.48	25.72 ± 0.07	20.04 ± 0.13	16.43 ± 13.77	5.65 ± 0.35	5.36 ± 0.58	27.87 ± 3.85	11.14 ± 0.52	6.44 ± 0.15
FixMatch [2]	12.66 ± 4.49	4.95 ± 0.10	4.26 ± 0.01	45.38 ± 2.07	27.71 ± 0.42	22.06 ± 0.10	<u>3.37 ± 1.01</u>	1.97 ± 0.01	2.02 ± 0.03	38.19 ± 4.76	8.64 ± 0.84	5.82 ± 0.06
FlexMatch [3]	5.29 ± 0.29	4.97 ± 0.07	4.24 ± 0.06	40.73 ± 1.44	26.17 ± 0.18	21.75 ± 0.15	8.19 ± 3.20	6.59 ± 2.29	6.72 ± 0.30	29.12 ± 5.04	9.85 ± 1.35	6.08 ± 0.34
Dash [14]	9.29 ± 3.28	5.16 ± 0.28	4.36 ± 0.10	47.49 ± 1.05	27.47 ± 0.38	21.89 ± 0.16	5.26 ± 2.02	2.01 ± 0.01	2.08 ± 0.09	42.00 ± 4.94	10.50 ± 1.37	6.30 ± 0.49
MPL [40]	6.93 ± 0.17	5.76 ± 0.24	4.55 ± 0.04	46.26 ± 1.84	27.71 ± 0.19	21.74 ± 0.09	-	-	-	35.76 ± 4.83	9.90 ± 0.96	6.66 ± 0.00
RelationMatch [41]	6.87 ± 0.12	4.85 ± 0.04	4.22 ± 0.06	45.79 ± 0.59	27.90 ± 0.15	22.18 ± 0.13	-	-	-	33.42 ± 3.92	9.55 ± 0.87	6.08 ± 0.29
CoMatch [6]	6.51 ± 1.18	5.35 ± 0.14	4.27 ± 0.12	53.41 ± 2.36	29.78 ± 0.11	22.11 ± 0.22	8.20 ± 5.32	2.16 ± 0.04	<u>2.01 ± 0.04</u>	13.74 ± 4.20	<u>7.63 ± 0.94</u>	<u>5.71 ± 0.08</u>
CCSSL [1]	9.71 ± 2.78	5.14 ± 0.55	4.46 ± 0.20	38.81 ± 1.65	<u>24.30 ± 0.63</u>	<u>19.32 ± 0.16</u>	7.85 ± 3.6	2.12 ± 0.04	2.03 ± 0.03	17.55 ± 4.2	8.43 ± 1.1	5.77 ± 0.82
SimMatch [8]	5.38 ± 0.01	5.36 ± 0.08	4.41 ± 0.07	39.32 ± 0.72	26.21 ± 0.37	21.50 ± 0.11	7.60 ± 2.11	2.48 ± 0.61	2.05 ± 0.05	16.98 ± 4.24	8.27 ± 0.40	5.74 ± 0.31
AdaMatch [15]	<u>5.09 ± 0.21</u>	5.13 ± 0.05	4.36 ± 0.05	38.08 ± 1.35	26.66 ± 0.33	21.99 ± 0.15	6.14 ± 5.35	2.13 ± 0.04	2.02 ± 0.05	19.95 ± 5.17	8.59 ± 0.43	6.01 ± 0.02
FreeMatch [18]	4.90 ± 0.12	<u>4.88 ± 0.09</u>	<u>4.16 ± 0.06</u>	39.52 ± 0.01	26.22 ± 0.08	21.81 ± 0.17	10.43 ± 0.82	8.23 ± 3.22	7.56 ± 0.25	28.50 ± 5.41	9.29 ± 1.24	5.81 ± 0.32
SoftMatch [5]	5.11 ± 0.14	4.96 ± 0.09	4.27 ± 0.05	<u>37.60 ± 0.24</u>	26.39 ± 0.38	21.86 ± 0.16	2.46 ± 0.24	2.15 ± 0.07	2.09 ± 0.06	22.23 ± 3.82	9.18 ± 0.68	5.79 ± 0.15
GBVSSL	6.15 ± 0.25	4.92 ± 0.05	3.98 ± 0.10	37.27 ± 0.28	23.25 ± 0.33	18.86 ± 0.24	5.21 ± 1.05	<u>1.99 ± 0.31</u>	1.89 ± 0.04	<u>15.74 ± 3.90</u>	7.37 ± 0.86	5.63 ± 0.35
Fully-Supervised	4.62 ± 0.05			19.30 ± 0.09			2.13 ± 0.02			None		

Fig. 1. Classification accuracy of each method on CIFAR-100 and STL-10 with 10% labeled data.

3.2 Feature Clustering Quality

The proposed method improved normalized mutual information by about 13%, which indicates clearer class clusters in the feature space. The supervised baseline showed weak clustering because the model used only a small labeled subset. SimCLR and MoCo pretraining improved feature separation, but some visually similar CIFAR-100 classes still overlapped. After pseudo-label fine-tuning, the feature space became more class-consistent because high-confidence unlabeled samples

provided extra training signals. This result agrees with studies on data-efficient contrastive learning [22,23]. Chen et al. (2025) showed that latent class structure and subset quality affect contrastive pretraining and downstream performance on CIFAR-100 and STL-10. Fig. 2 shows a related example of how latent grouping affects self-supervised learning results.

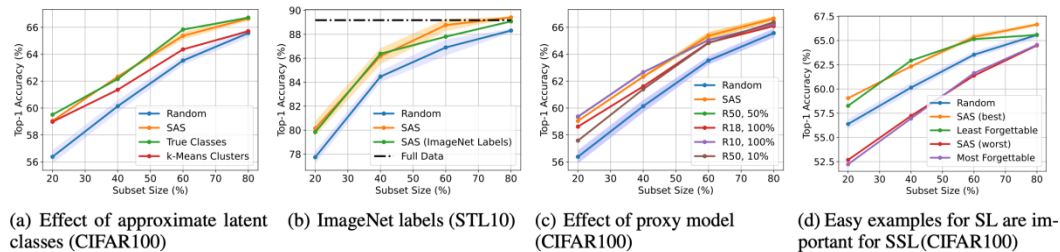


Fig. 2. Feature clustering before and after contrastive pretraining and pseudo-label fine-tuning.

3.3 Effects of Contrastive Pretraining and Pseudo-Label Fine-Tuning

The ablation results showed that contrastive pretraining and pseudo-label fine-tuning had different roles. SimCLR improved the initial feature representation by learning similarity between different views of the same image. MoCo produced more stable features because the momentum encoder kept a consistent feature queue during training. However, contrastive pretraining alone did not fully use class-level information. Pseudo-label fine-tuning further improved accuracy by adding high-confidence unlabeled samples to the training set. This step was most useful when the pretrained model already gave reliable predictions [24,25]. If pseudo labels were added too early, wrong labels could affect the decision boundary. Therefore, the two-stage design was important. Contrastive pretraining first built a stable feature space, and pseudo-label fine-tuning then improved class boundaries.

3.4 Comparison With Existing Studies and Practical Value

Compared with supervised training under limited labels, the proposed method used unlabeled samples more effectively and reduced reliance on manual annotation. Compared with pure contrastive learning, it improved final classification by adding pseudo-label supervision. Compared with pseudo-label-only training, it reduced the risk of noisy labels because the model first learned stronger visual features through contrastive pretraining. These results show that self-supervised learning is useful for low-data classification. The method may be used in medical imaging, remote sensing, and industrial inspection, where labeled data are limited but unlabeled images are easier to collect. The main limitation is that the experiments were conducted on CIFAR-100 and STL-10, which are standard but controlled datasets. More tests are needed on larger datasets, fine-grained real-world images, and imbalanced classes. Future work should also examine label-noise control, adaptive pseudo-label thresholds, and transfer to detection or segmentation tasks.

4. Conclusion

This study proposed a self-supervised learning method for image classification with limited labeled data. The method used multi-view augmentation, SimCLR and MoCo pretraining, and pseudo-label fine-tuning to learn useful features from unlabeled images. Experiments on CIFAR-100 and STL-10 showed that the method improved classification accuracy by about 9.7% compared with supervised training using the same 10% labeled data. The accuracy gap from full-supervised training was less than 2.5%, and NMI increased by about 13%. These results show that contrastive pretraining and pseudo-label learning can reduce the effect of label shortage and improve feature clustering. The main contribution of this study is the combined use of feature learning and pseudo-label supervision in a low-data setting. The method may be applied to medical imaging, remote sensing, and industrial inspection, where labeled data are limited but unlabeled images are easier to collect. However, the study still has limitations. The experiments were mainly based on CIFAR-100 and STL-10. More tests are needed on larger datasets, imbalanced samples, and real-world images. Future work should also examine adaptive pseudo-label thresholds, label-noise control, and transfer to detection and segmentation tasks.

References

- Wang, Y., Yin, X., Chen, J., & Wang, Y. (2026). Evidence-Based Study on Low-Burden Digital Phenotyping for Precision Screening of Oral Anti-Obesity Drug Efficacy.
- Shoaib, M. A., Ali, R., Bazai, S. U., & Mir, T. (2025). Deep learning techniques for image segmentation and data annotation. In *Modern Intelligent Techniques for Image Processing* (pp. 63-94). IGI Global Scientific Publishing.
- Yan, B. (2026). A Comparative Study: Greedy versus GAN-Based Compressed Sensing for Fast MRI Reconstruction. *International Journal of Advance in Applied Science Research*, 5(3), 36-42.
- Uelwer, T., Robine, J., Wagner, S. S., Höftmann, M., Upschulte, E., Konietzny, S., ... & Harmeling, S. (2025). A survey on self-supervised methods for visual representation learning. *Machine Learning*, 114(4), 111.
- Jiao, Y., Wang, A., Zhao, B., & Shi, T. (2026). Quantitative Study on the Construction and Application Effectiveness of Graffiti Wall Painting Teaching Models in Public Space Contexts.
- Kanagarajan, M., Natarajan, M., Rajendran, S., Velusamy, P., Ganesan, S. K., Bose, M., ... & Stephen Inbaraj, B. (2025). AIM-Net: A Resource-Efficient Self-Supervised Learning Model for Automated Red Spider Mite Severity Classification in Tea Cultivation. *AgriEngineering*, 7(8), 247.
- Liu, S., Liu, X., & Feng, H. (2025, November). Research on AI-Driven Visual Design and Immersive Interactive Experiences Based on Multimodal Cognition and User. In *Proceedings of the 2025 International Conference on Digital Society and Intelligent Computing* (pp. 734-740).
- Altaheri, H., Karray, F., Islam, M. M., Raju, S. M., & Karimi, A. H. (2025). Bridging Brain with Foundation Models through Self-Supervised Learning. *arXiv preprint arXiv:2506.16009*.
- Chen, F., Li, S., Liang, H., Xu, P., & Yue, L. (2025). Optimization Study of Thermal Management of Domestic SiC Power Semiconductor Based on Improved Genetic Algorithm.
- Deorari, M. C., Rajput, I. S., Kumar, A., & Tyagi, S. (2026, February). Self-Supervised Learning for Small Datasets Using Data Augmentation and Contrastive Learning. In *2026 1st International Conference on Advancing Sustainable Solutions through Technologies (ICASST)* (pp. 1-6). IEEE.
- Liang, R., Feifan, F. N. U., Liang, Y., & Ye, Z. (2025). Emotion-Aware Interface Adaptation in Mobile Applications Based on Color Psychology and Multimodal User State Recognition. *Frontiers in Artificial Intelligence Research*, 2(1), 51-57.
- Petti, D., Li, C., & Liu, N. (2026). Contrastive multi-view representation learning for multi-camera plant phenotyping: A cotton field study. *Plant Phenomics*, 100193.
- Zhu, W., Yao, Y., & Yang, J. (2025). Optimizing Financial Risk Control for Multinational Projects: A Joint Framework Based on CVaR-Robust Optimization and Panel Quantile Regression.
- Jalal, M. E., & Elmaghraby, A. (2025). Towards Deep Semi-Supervised Continual Learning. *IEEE Access*.
- Gao, X., Chen, J., Huang, M., & Fang, S. (2025). Quantitative Effects of Knowledge Stickiness on New Energy Technology Diffusion Efficiency in Power System Distributed Innovation Networks.
- Pacal, I., & Işık, G. (2025). Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification. *Neural Computing and Applications*, 37(4), 2479-2496.
- Feng, H. (2024, October). Design of Intelligent Charging System for Portable Electronic Devices Based on Internet of Things (IoT). In *2024 5th International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)* (pp. 568-571). IEEE.
- Ghamsarian, N., Nasirihaghighi, S., Schoeffmann, K., & Sznitman, R. (2025). Feedback-Driven Pseudo-Label Reliability Assessment: Redefining Thresholding for Semi-Supervised Semantic Segmentation. *arXiv preprint arXiv:2505.07691*.

- Liu, S., & Yim, J. (2025). Research on Generative AI Creation Systems Based on Visual Language Modeling: Human-Machine Collaboration and Cognitive Feedback Mechanisms. Available at SSRN 6139770.
- Abreu, A. J. P. O. E. (2025). Overcoming scarce annotations through deep semi-supervised learning in cancer characterisation (Master's thesis, Universidade do Porto (Portugal)).
- Gao, G., Gao, R., Gao, R., & Zhou, H. (2026). Engineering Analysis and Quantitative Research on the Platform-Based Evolution of Enterprise Communication Systems.
- Schienenbein, P., & Blumberger, J. (2025). Data-Efficient Active Learning for Thermodynamic Integration: Acidity Constants of BiVO₄ in Water. *ChemPhysChem*, 26(1), e202400490.
- Chen, X., Xiao, H., Zeng, Z., Zhang, S., & Xiao, R. (2025). Fine-Grained Multivariate Time Series Anomaly Detection via Causal Inference. *Knowledge-Based Systems*, 114765.
- Pettorelli, N., Williams, J., Schulte to Bühne, H., & Crowson, M. (2025). Deep learning and satellite remote sensing for biodiversity monitoring and conservation. *Remote Sensing in Ecology and Conservation*, 11(2), 123-132.
- Xu, S., Ma, Q., Liu, H., & Yue, L. (2026). Continuous Reorganization and Performance Preservation of Agent Memory Structure Under Distributed Change Environments.