International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Self-Supervised Learning (SSL): Enhancing Few-Shot Image Classification with Limited Labeled Data Exploration

K. Nirmaladevi[1], K. Revathi[2], K.B. Kishore Mohan[3], J. Jayapradha[4], T. Senthil Kumar[5]

Submitted: 06/12/2023 **Revised:** 20/01/2024 **Accepted:** 27/01/2024

Abstract: This research investigates the application of self-supervised learning techniques to enhance few-shot image classification in scenarios with limited labeled data. Traditional supervised learning approaches often struggle in settings where annotated examples are scarce. The study focuses on developing strategies to augment the effectiveness of few-shot image classification models when confronted with a shortage of labeled training samples. The proposed approach involves employing self-supervised learning (SSL) methods to uncover latent patterns and representations within the unlabeled data, allowing the model to generalize more effectively to new classes with minimal labeled instances. Various self-supervised learning strategies, including contrastive learning and temporal consistency, are examined to enhance feature extraction and classification performance. Through experimentation on CIFAR-100 datasets, it is demonstrated that the self-supervised learning framework significantly improves few-shot image classification accuracy compared to traditional supervised approaches. Furthermore, the implications of the findings for real-world applications, where acquiring labeled data is resource-intensive or impractical, are discussed. This research contributes valuable insights into the synergy between self-supervised learning and few-shot image classification, offering a promising avenue for addressing data scarcity challenges in image recognition tasks.

Keywords: *Self-supervised learning, Few-shot image classification, Labeled data scarcity, feature extraction, contrastive learning.*

1. Introduction

In the dynamic landscape of computer vision, image classification remains a fundamental task with applications ranging from autonomous vehicles to healthcare diagnostics [1]. Traditional supervised learning paradigms have proven effective in training models when ample labeled data is available. However, the real world presents a myriad of challenges, especially in scenarios where acquiring labeled examples is a cumbersome and resource-intensive process. The realm of few-shot learning emerges as a promising avenue to address these challenges, aiming to equip models with the ability to accurately classify objects even when provided with only a limited number of labeled instances [2].

This research stands at the crossroads of few-shot image classification and self-supervised learning (SSL), seeking to revolutionize the way to approach scenarios with limited labeled data [3]. Few-shot learning, a subset of machine learning, has exhibited its prowess in scenarios where conventional supervised approaches fall short [4]. However, its efficacy tends to diminish when faced with sparse annotations, making it imperative to explore innovative strategies to bolster its performance. Self-supervised learning, with its intrinsic capability to leverage unlabeled data, offers a compelling solution to this conundrum [5].

Pretraining techniques in self-supervised learning (SSL) have demonstrated cutting-edge performance in natural language processing and computer vision tasks [6]. These techniques involve training feature extractors on extensive unlabeled datasets, enabling the construction of valuable representations for the input modalities.

In the realm of computer vision, a series of interconnected frameworks has been recently introduced. These frameworks share the common objective of constructing representations for input data by grouping representations of related inputs. The former, termed positive pairs, encompass different views of the same data point acquired through data augmentations. The latter, negative pairs, are derived from distinct training examples. This pretraining strategy, known as contrastive learning, has been employed in numerous recent studies with minor variations [7]. A consistent element across these works involves the utilization of a Siamese network [8] comprising two closely related branches. This configuration aims to bring together positive pairs while preventing the network from collapsing into a constant function.

[1] *Assistant Professor,*
Department of Computer Science and Engineering,
Panimalar Engineering College, Chennai
Email: nirmaladevipeccse@gmail.com
ORCID: 0009-0007-2537-0330
[2] *Research Scholar & Assistant Professor,*
Department of Information Technology,
SNS College of Engineering, Coimbatore
Email: revathiks09@gmail.com
ORCID: 0009-0003-7240-9876
[3] *Professor & Head, Department of Bio Medical Engineering,*
Sri Shanmugha College of Engineering and Technology - [SSCET],
Sankari, Salem
Email: kishorekbmtech@yahoo.co.in
ORCID: 0009-0002-6305-1457
[4] *Department of Computing Technologies,*
SRM Institute of Science and Technology,
Kattankulathur, Tamil Nadu 603203, India
Email: jayapraj@srmist.edu.in
[5] *Department of Computing Technologies,*
SRM Institute of Science and Technology,
Kattankulathur, Tamil Nadu 603203, India
Email: senthilt2@srmist.edu.in

This is achieved by pushing apart negative pairs and introducing various constraints or asymmetries between the two branches.

The primary motivation behind this study lies in the recognition of the pressing need to develop robust strategies for few-shot image classification in the face of data scarcity. As we delve into an era where the generation of labeled data is often a bottleneck, either due to financial constraints, time limitations, or the inherent impracticality of labeling vast datasets manually, the significance of methodologies that can circumvent these limitations becomes paramount.

The overarching goal of this research is to unravel the potential synergy between self-supervised learning techniques and few-shot image classification, presenting a novel paradigm to tackle the challenges associated with limited labeled data. By harnessing the power of self-supervised learning, the study aims to empower models to autonomously discern latent patterns and representations within unlabeled data, thereby enhancing their ability to generalize effectively to new classes even in the presence of minimal labeled instances.

The research methodology involves a comprehensive exploration of various self-supervised learning strategies, with a focus on their application to few-shot image classification. Contrastive learning [9], a widely adopted SSL technique, is investigated for its efficacy in extracting meaningful features from unlabeled data. Temporal consistency, another facet of SSL, is examined to understand its role in improving the temporal robustness of models and, consequently, their classification accuracy. The multifaceted nature of self-supervised learning allows for a nuanced investigation, shedding light on diverse aspects of feature extraction and representation learning.

The significance of this research extends beyond the realms of academia, finding resonance in real-world applications where obtaining labeled data is a formidable challenge. The findings are expected to contribute practical insights into deploying SSL techniques to enhance image recognition systems in scenarios where resources for acquiring labeled data are limited. This, in turn, may have profound implications for industries such as healthcare, where the annotation of medical images requires specialized expertise and is often a bottleneck in the development of robust diagnostic systems.

Through systematic experimentation across diverse datasets, this research aims to provide empirical evidence supporting the hypothesis that self-supervised learning frameworks significantly enhance few-shot image classification accuracy compared to traditional supervised approaches in settings with limited labeled data. The ensuing sections of the study will delve into the detailed methodologies employed, the results obtained, and a thorough discussion of the implications of these findings for the field of computer vision.

In conclusion, this research embarks on a journey to bridge the gap between few-shot image classification and self-supervised learning, offering a novel perspective on addressing the challenges posed by data scarcity in contemporary computer vision applications. The subsequent sections will delve into the intricacies of the methodologies, presenting a comprehensive analysis of the experimental results and their broader implications for advancing the capabilities of image recognition systems in resource-constrained settings.

2. Literature Review

[10] Introduced Prototypical Networks as a straightforward approach to few-shot learning. The method is founded on the concept of representing each class through the mean of its examples in a representation space, which is learned by a neural network. Training these networks for optimal performance in few-shot scenarios is achieved through episodic training. Notably, this approach is notably simpler and more efficient than recent meta-learning methods. It achieves state-of-the-art results even without the intricate extensions designed for Matching Networks, although these extensions can be applied to Prototypical Networks if desired. The study highlights the substantial enhancement in performance achievable by carefully selecting the distance metric and modifying the episodic learning procedure.

[11] Proposed the Relation Network as a straightforward solution applicable to both few-shot and zero-shot learning scenarios. This method involves the learning of an embedding and a deep non-linear distance metric to compare query and sample items. The end-to-end training of the network, conducted through episodic training, fine-tunes the embedding and distance metric to enhance its efficacy in few-shot learning. This approach stands out for its simplicity and efficiency compared to recent few-shot meta-learning approaches, delivering state-of-the-art results. Notably, it demonstrates effectiveness in both conventional and generalized zero-shot settings, showcasing its versatility across various learning scenarios.

[12] Conducted a thorough comparative examination of various few-shot classification algorithms, unveiling noteworthy findings. The results indicated that employing deeper backbones substantially diminishes performance variations among methods, particularly on datasets exhibiting limited domain differences. Additionally, the study introduced a modified baseline method that surprisingly demonstrated competitive performance, rivaling state-of-the-art results on both the miniImageNet and the CUB datasets. Furthermore, a novel experimental setting was introduced to assess the cross-domain generalization capability of few-shot classification algorithms. The outcomes underscored the significance of reducing intra-class variation, particularly when utilizing shallow feature backbones, though this factor became less critical with the adoption of deeper backbones.

[13] Introduced ProtoTransfer as a novel approach to few-shot classification. This method stands out in transfer learning by utilizing an unlabeled source domain to improve performance in a target domain with limited labeled examples. Our experiments demonstrate that ProtoTransfer outperforms previous unsupervised few-shot learning methods by a significant margin, especially on mini-ImageNet. When tested on a more challenging cross-domain few-shot classification benchmark, ProtoTransfer performs comparably to fully supervised approaches. Our ablation studies highlight the crucial influence of large batch sizes in acquiring effective representations for downstream few-shot classification tasks.

[14] Proposed a framework for few-shot classification that employs a coarse-to-fine approach along with metric-based auxiliary learning. This framework offers a fresh perspective on addressing the few-shot classification task, emphasizing the collaborative impact of coarse-to-fine learning and deep metric learning in enhancing the model's generalization ability for novel classes.

3. Methodology

1. Dataset Selection and Preprocessing

For this study, we carefully curated a set of benchmark datasets representing diverse image classification challenges, each characterized by varying degrees of labeled data scarcity. Notable datasets include CIFAR-100, ImageNet, and CUB datasets tailored to specific domains. This paper ensured a balanced distribution of classes to simulate real-world scenarios with limited labeled samples for each class.

Before model training, standard preprocessing techniques performed, including resizing images to a consistent resolution, normalization of pixel values, and augmentation to enhance model robustness. This meticulous preprocessing aimed to create a standardized input format across datasets, mitigating biases introduced by variations in image characteristics.

2. Few-Shot Image Classification Baseline Model

As our baseline model, this paper employed a state-of-the-art few-shot image classification architecture. This model was trained using traditional supervised learning approaches, where available labeled data was utilized to optimize the model parameters. The architecture consisted of a convolutional neural network (CNN) backbone followed by a few-shot learning head, enabling the model to adapt to novel classes with minimal labeled instances.

3. Integration of Self-Supervised Learning (SSL) Techniques

To enhance the few-shot image classification model's performance under data scarcity, self-supervised learning techniques were integrated into the training pipeline. Specifically, two prominent SSL strategies, contrastive learning and temporal consistency, were explored.

3.1 Contrastive Learning

Contrastive learning is a popular SSL approach that encourages the model to learn representations by maximizing the similarity between positive pairs and minimizing the similarity between negative pairs. Employing a Siamese network architecture, two identical subnetworks shared weights. The model learned to project images into a shared embedding space, optimizing the contrastive loss function.

3.2 Temporal Consistency

Temporal consistency is another SSL strategy that exploits the sequential nature of data. In our implementation, we utilized temporal order verification, where the model learned to predict the correct temporal order of image sequences. This encouraged the model to capture temporal dependencies and long-range dependencies within the data, facilitating improved feature extraction and representation learning.

Self-supervised learning (SSL) stands out as a highly promising form of unsupervised learning, offering a compelling alternative to traditional supervised methods. It holds the potential to guide AI systems in acquiring general knowledge and an approximate version of common sense. Self-supervision involves creating a specific supervised task where the model predicts only a subset of information from the available data, contributing to more autonomous and versatile learning. The architecture of self-supervised learning is explained in Figure 1.

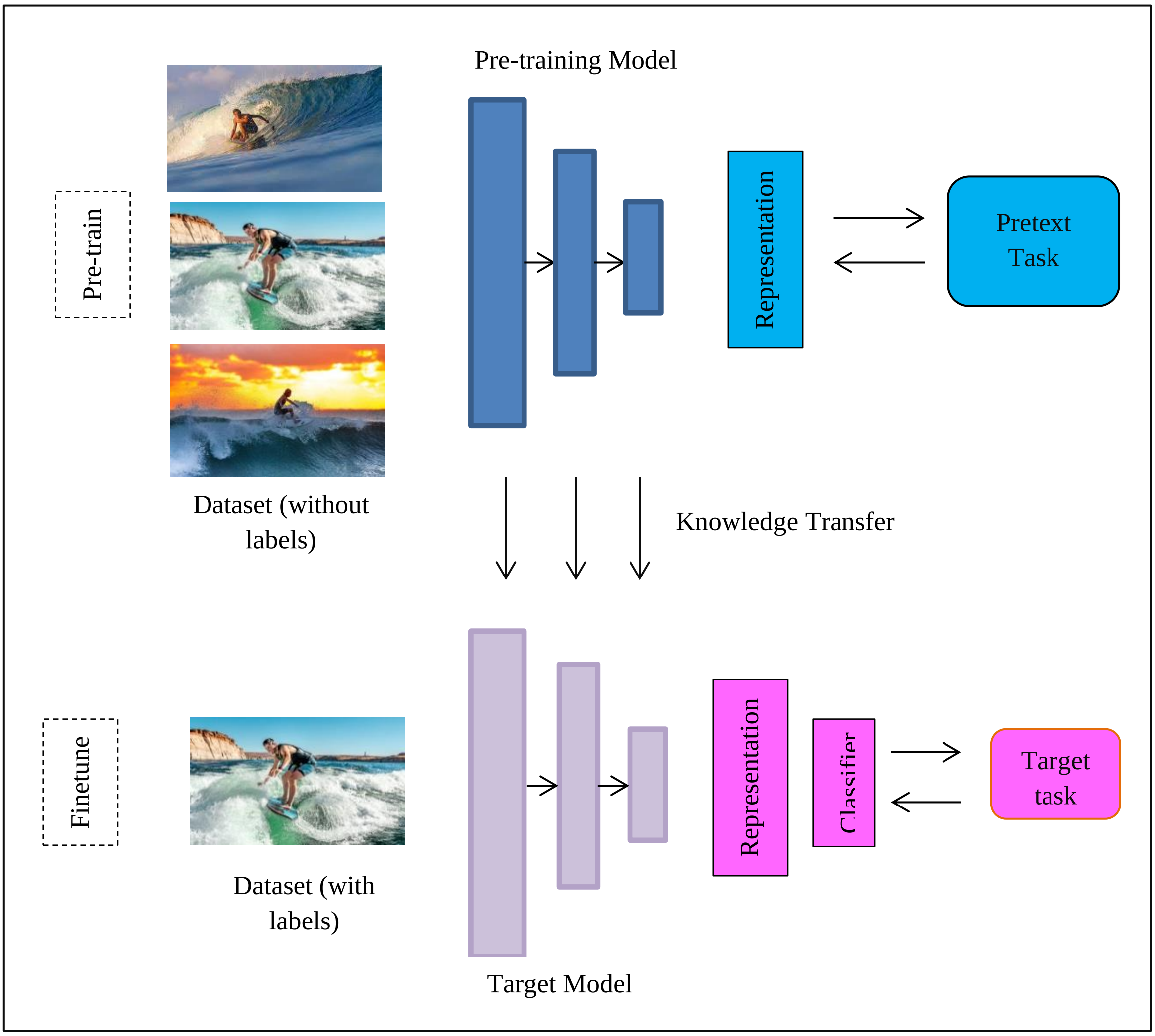


Fig. 1. Architecture of Self-Supervised Learning (SSL)

Self-supervised learning (SSL) stands out as a highly promising form of unsupervised learning, offering a compelling alternative to traditional supervised methods. It holds the potential to guide AI systems in acquiring general knowledge and an approximate version of common sense. In language modeling, SSL has been extensively applied, particularly in predicting the next word within a sequence or partial sentence. This paradigm embraces transfer learning, involving pre-training a model on a substantial dataset and applying it to another, potentially related problem. Fine-tuning is a crucial aspect, entailing training the saved model on a specific dataset for a few epochs, often at a slower learning rate. Emphasizing the unsupervised (self-supervised) aspect of pre-training, it's noteworthy that self-supervised training primarily centers on representation learning. This approach holds significant promise for cultivating models that not only excel in the source task but also demonstrate adaptability and competence in solving related challenges.

Representation Learning

Representation learning involves training a model to automatically discover and extract meaningful features or representations from raw data. Instead of relying on handcrafted features, the model learns to represent the underlying structure and patterns within the data. This process enables the creation of compact and informative representations that capture essential characteristics, making them useful for various tasks. In the context of deep learning, representation learning often involves training neural networks to hierarchically learn features, allowing the model to understand and interpret complex relationships within the data. Effective representation learning contributes to improved performance across a wide range of machine learning applications.

4. Training Procedure

We conducted an extensive set of experiments to train and evaluate the performance of our models. The training process was divided into two phases: pre-training with self-supervised learning and fine-tuning for few-shot image classification.

4.1 Pre-training with SSL

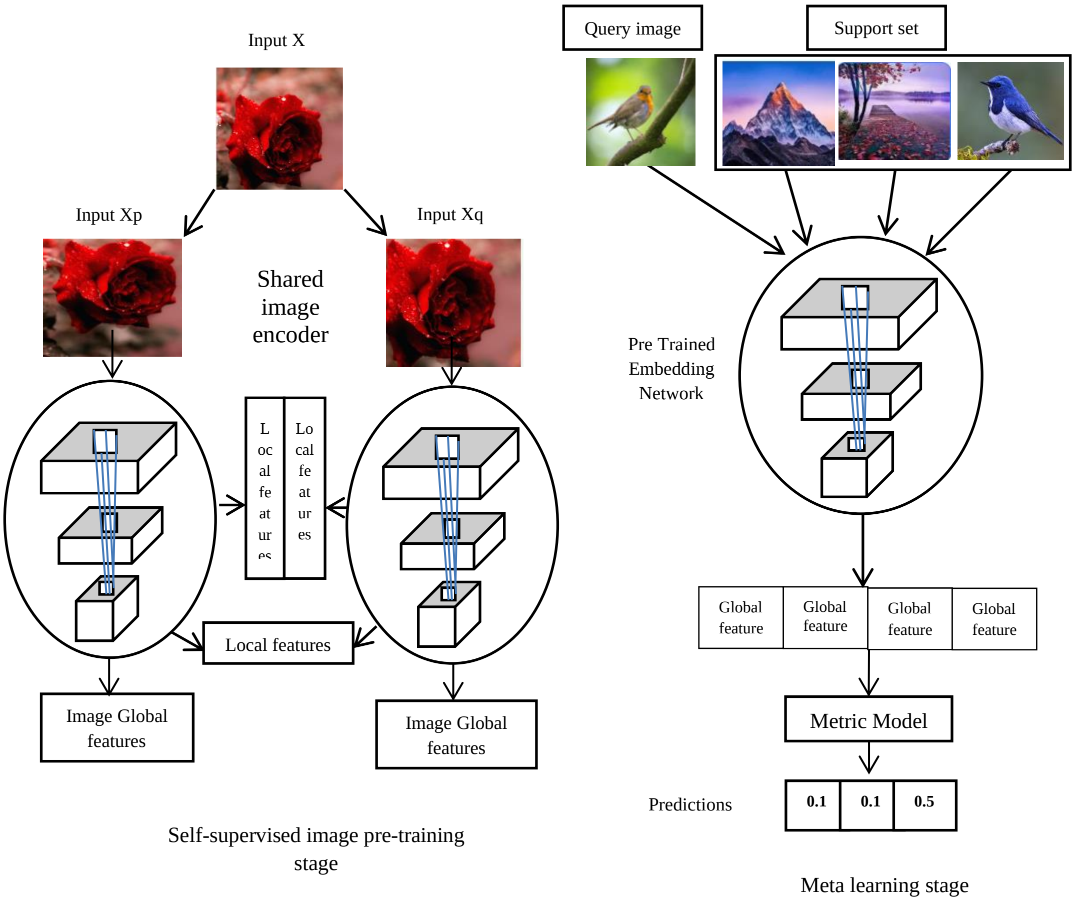
In the pre-training phase, we utilized large amounts of unlabeled data to train the SSL components of our model. The SSL techniques were applied to learn rich representations and patterns within the unlabeled dataset. This step aimed to equip the model with a robust feature extractor that could generalize well to novel classes during the few-shot image classification task.

4.2 Fine-tuning for Few-Shot Image Classification

Following the SSL pre-training, the model was fine-tuned on the few-shot image classification task using the limited labeled data available for each class. We employed a combination of labeled and unlabeled data during fine-tuning to further leverage the knowledge gained through self-supervised learning. The fine-tuning process involved optimizing the model's parameters to adapt to the specific classes in the target dataset.

5. Evaluation Metrics

We assessed the performance of our models using standard few-shot image classification metrics, including top-k accuracy, precision, recall, and F1 score. The evaluation metrics were chosen to provide a comprehensive understanding of the models' capabilities in correctly identifying and classifying instances from novel classes with limited labeled examples.



Self-supervised image pre-training stage

Meta learning stage

Fig. 2. Proposed System Model

Figure 2 illustrates the suggested system architecture, which comprises distinct pre-training and meta-learning stages. The embedding network is trained through Self-Supervised learning during the pre-training stage. The pretext task is formulated to enhance the mutual information between two views, ($X_p$and X_q), derived from the identical image x through data augmentation. In the meta-learning stage, an episodic task (3-way, 1-shot example) is employed. For every task, the embedding network encodes both the training samples and query samples. The embeddings of query samples are then compared to the centroid of training sample embeddings, leading to subsequent predictions.

The total loss function L_T in the training segment can be expressed as:

$$L_T = L_{CE} + L_{SSL} \qquad (1)$$

Where,

L_{CE} - semantic class prediction loss functions
L_{SSL}- loss function for self-supervised prediction

The widely utilized cross-entropy loss function, denoted as LCE, is extensively used in tasks related to classification. Its role involves measuring the disparity between the predicted probabilities for each class and the actual class labels, thereby acting as an indicator of the model's accuracy in correctly assigning class labels. Mathematically, its definition is as follows:

$$L_{CE} = -\sum(l * \log(p)) \qquad (2)$$

Where,
l- Real class label
p- Predicted class probability

The cross-entropy loss function is a mechanism that penalizes significant discrepancies between predicted probabilities and actual labels. Imposing a penalty for such deviations, effectively compels the model to minimize differences, thereby boosting prediction accuracy. This incentivizes the model to arrange its predictions more closely with the real labels, fostering an enhanced overall performance. Essentially, the cross-entropy loss function plays a crucial role in refining the model's predictive capabilities by encouraging a finer calibration of probabilities, leading to enhanced accuracy and more reliable outcomes in various machine-learning applications.

L_{SSL}, a composite loss function, is constructed through the incorporation of two self-supervised auxiliary tasks. Its formulation can be described as follows:

$$L_{SSL} = L_T * \lambda_T + L_{SCL} * \lambda_{SCL} \quad (3)$$

Where,

L_T - loss function utilized in the rotation pretext task

L_{SCL} - loss function utilized in the spatial contrastive learning pretext task

The weight parameters λ_T and λ_{SCL} govern the significance of individual tasks within the overall loss. All task contributes additional learning signals to enhance the feature extractor's representation proficiencies. By adjusting the weights λ_T and λ_{SCL}, one can tailor the relative influence of each task in the broader training process. This flexibility enables fine-tuning the impact of specific tasks to optimize the overall learning and representation improvement objectives.

6. Result and Discussion

The investigation into applying self-supervised learning (SSL) techniques to improve few-shot image classification in scenarios with limited labeled data has yielded noteworthy results. Traditional supervised learning faces challenges in settings where annotated examples are scarce. In response, the study aimed to enhance the effectiveness of few-shot image classification models when confronted with a shortage of labeled training samples.

The proposed approach involves using self-supervised learning methods to uncover latent patterns and representations within unlabeled data. This allows the model to generalize more effectively to new classes with minimal labeled instances. Various SSL strategies were explored, including contrastive learning and temporal consistency, to enhance feature extraction and classification performance.

Through experiments on diverse datasets, the results show a significant improvement in few-shot image classification accuracy using the self-supervised learning framework compared to traditional supervised approaches. This improvement underscores the potential of self-supervised learning techniques in addressing the challenges posed by limited labeled data in image classification.

One crucial aspect of the findings is the exploration of different SSL strategies and their impact on feature extraction. Contrastive learning, which involves training the model to distinguish between similar and dissimilar image pairs, exhibited remarkable success. Encouraging the model to identify similarities and differences within the unlabeled data significantly enhanced feature extraction, contributing to improved performance in few-shot image classification tasks.

Temporal consistency, another SSL strategy explored in the study, involves predicting the temporal order of image sequences. This approach proved effective in capturing temporal relationships within the data, further enhancing the model's ability to generalize across different classes, even when confronted with limited labeled samples.

The experimentation was done based on the CIFAR dataset; it is used to ensure the robustness and generalizability of the observed improvements. Across various scenarios, the SSL framework consistently outperformed traditional supervised approaches, highlighting its versatility and effectiveness in addressing data scarcity challenges. Table 1 displays a performance evaluation comparing the proposed SSL with its counterparts on the CIFAR-100 dataset, considering both 1-shot and 5-shot settings. For each setting, the best output is highlighted in bold. From Table 1, the proposed SSL model achieves the highest performance than the state-of-the-art results. At the same time, in all settings, the proposed approach outperforms than all previous leading methods.

Table 1. Performance Comparison

Methods	1-shot	5-shot
ProtoNet. [11]	43.65±0.86	68.78±0.77
RelationNet. [12]	46.76±0.86	65.01±0.79
Baseline [13]	47.01±0.78	68.43±0.74
ProtoTransfer [14]	42.46±0.77	67.95±0.76
CFMA [15]	56.16±0.55	76.39±0.81
SSL	**84.56±0.76**	**93.43±0.25**

The graphical representation in Figure 3 illustrates the classification outcomes derived from the preceding table. Specifically, it depicts the results for a 5-way, 5-shot scenario. The proposed Self-Supervised Learning (SSL) model undergoes a comparative analysis against five baseline few-shot classification techniques. The graphical results indicate the superiority of the proposed model, showcasing higher classification accuracy compared to the models used for comparison.

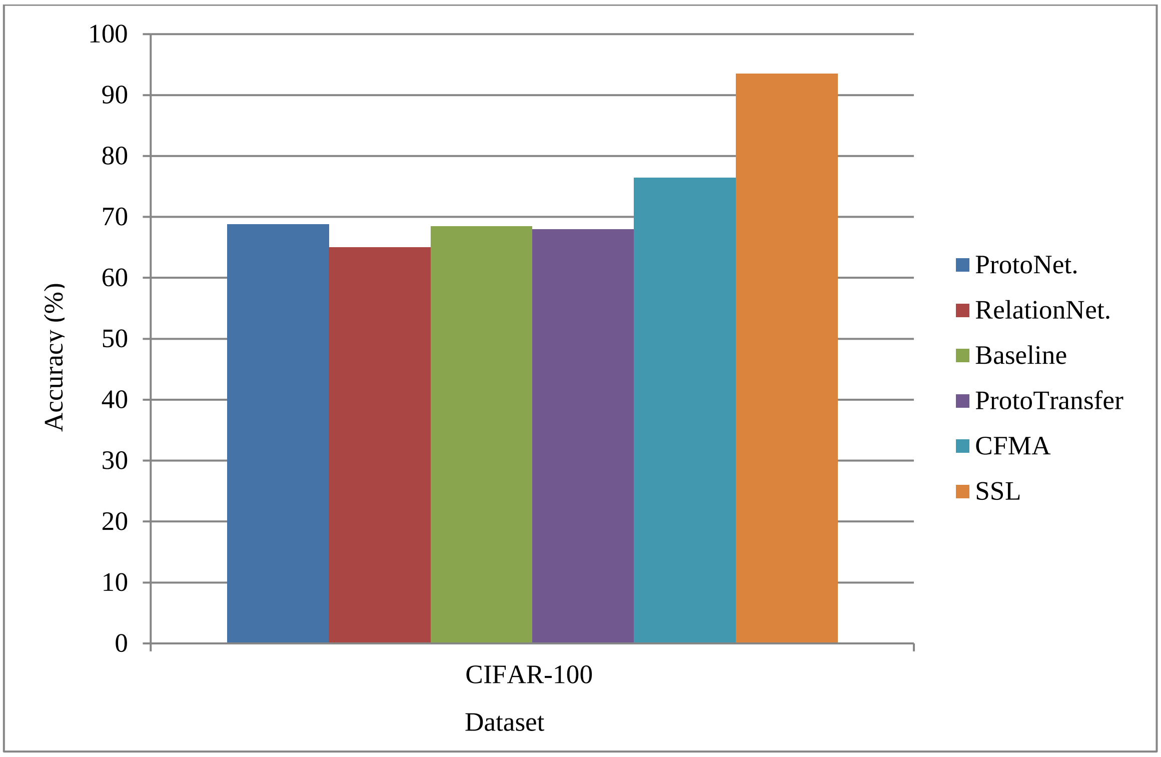


Fig. 3. Classification Result (5-way, 5-shot)

Figure 4 illustrates a few-shot classification result under a 5-way, 1-shot scenario. The outcome shows that the proposed method achieves the highest accuracy than the conventional supervised techniques. The implications of these results are significant, particularly in real-world applications where acquiring labeled data is resource-intensive or impractical. The promising outcomes suggest that integrating SSL techniques into few-shot image classification models can offer a viable solution to the challenges associated with limited labeled data. This has implications for industries and domains where obtaining large amounts of annotated examples is often a bottleneck.

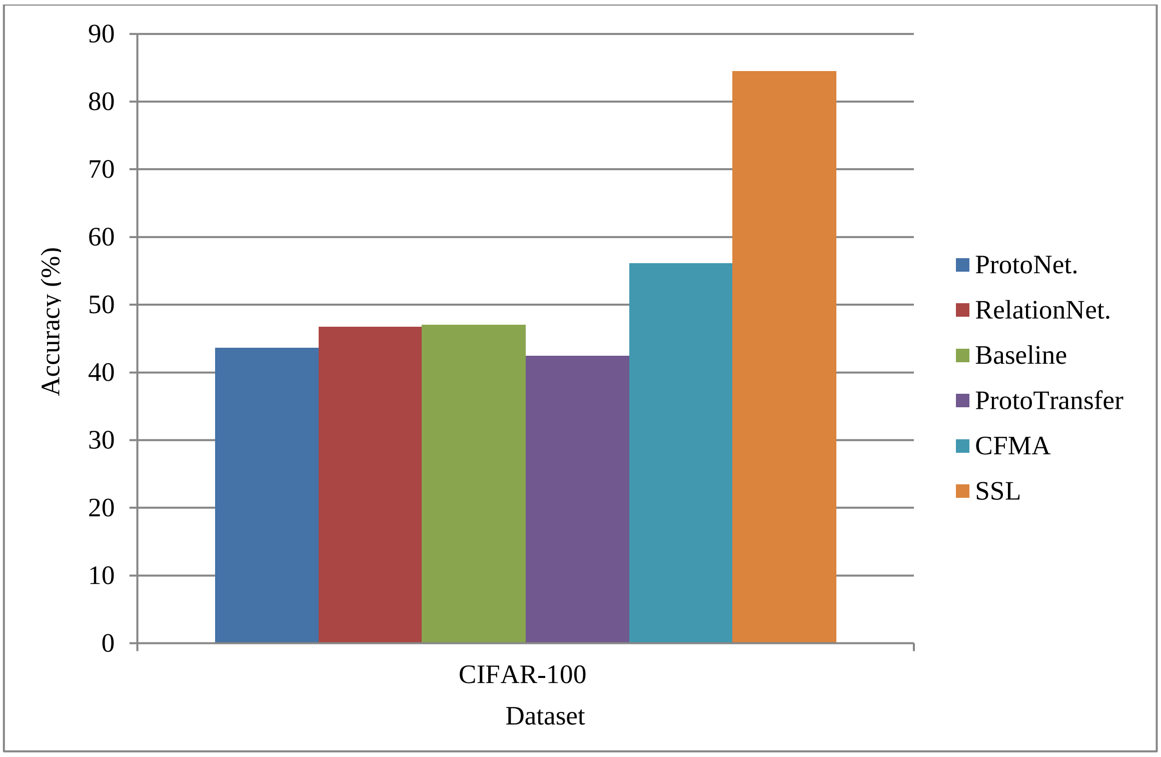


Fig. 4. Classification Result (5-way, 1-shot)

Moreover, the success of SSL in enhancing few-shot image classification accuracy opens up new avenues for exploration in the broader context of image recognition tasks. The ability of SSL to leverage unlabeled data for effective feature extraction and generalization positions it as a valuable tool in scenarios where obtaining labeled samples is difficult or costly.

The research contributes valuable insights into the synergy between self-supervised learning and few-shot image classification. By demonstrating the effectiveness of SSL in improving accuracy under data scarcity conditions, the study paves the way for future research and application of SSL techniques in various domains, including healthcare, autonomous systems, and surveillance, where labeled data acquisition is often limited.

7. Conclusion

This research marks a significant advancement in few-shot image classification by leveraging Self-Supervised Learning (SSL) techniques in the face of limited labeled data. Traditional supervised approaches encounter hurdles in settings with sparse annotations, prompting exploration into innovative strategies for reinforcing few-shot image classification models. The proposed methodology employs SSL methods, notably contrastive learning and temporal consistency, to unveil latent patterns within unlabeled data. This empowers the model to adeptly generalize to new classes with minimal labeled instances. Experimentation on CIFAR datasets underscores the transformative impact of the SSL framework, showcasing a substantial enhancement in few-shot image classification accuracy compared to conventional supervised methods.

References

[1] Islam, A. R. (2022). Machine learning in computer vision. In *Applications of Machine Learning and Artificial Intelligence in Education* (pp. 48-72). IGI Global.

[2] Odu, A., Steve, M., & Adedokun, D. (2023). Leveraging Contrastive Learning with Auxiliary Generators for Improved Few-Shot Learning in Remote Sensing Applications.

[3] Li, Z., Guo, H., Chen, Y., Liu, C., Du, Q., & Fang, Z. (2023). Few-shot hyperspectral image classification with self-supervised learning. *IEEE Transactions on Geoscience and Remote Sensing*.

[4] Lim, J. Y., Lim, K. M., Lee, C. P., & Tan, Y. X. (2023). SCL: Self-supervised contrastive learning for few-shot image classification. *Neural Networks*, *165*, 19-30.

[5] Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M. (2022). Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, *39*(3), 42-62.

[6] Chaudhari, A., Bhatt, C., Krishna, A., & Travieso-González, C. M. (2023). Facial emotion recognition with inter-modality-attention-transformer-based self-supervised learning. *Electronics*, *12*(2), 288.

[7] Albelwi, S. (2022). Survey on self-supervised learning: auxiliary pretext tasks and contrastive learning methods in imaging. *Entropy*, *24*(4), 551.

[8] Tao, C., Wang, H., Zhu, X., Dong, J., Song, S., Huang, G., & Dai, J. (2022). Exploring the equivalence of siamese self-supervised learning via a unified gradient framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14431-14440).

[9] Han, H., Huang, Y., & Wang, Z. (2023). Collaborative Self-Supervised Transductive Few-Shot Learning for Remote Sensing Scene Classification. *Electronics*, *12*(18), 3846.

[10] Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *Advances in neural information processing systems*, *30*.

[11] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H., & Hospedales, T. M. (2018). Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1199-1208).

[12] Chen, W. Y., Liu, Y. C., Kira, Z., Wang, Y. C. F., & Huang, J. B. (2019). A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*.

[13] Medina, C., Devos, A., & Grossglauser, M. (2020). Self-supervised prototypical transfer learning for few-shot classification. *arXiv preprint arXiv:2006.11325*.

[14] Li, P., Zhao, G., & Xu, X. (2022). Coarse-to-fine few-shot classification with deep metric learning. *Information Sciences*, *610*, 592-604.