Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Self-Supervised Contrastive Regularization Algorithms For Robust Few Shot Classification

Dr. S. Subbaiah[1*], S.Bharathi[2], Dr. Muhammed Anshad P. Y[3], Farrux Yoqubov[4], Gulnoza Qodirova[5], Dilafruz Amerova[6]

[1*]Associate Professor, Department of AI&ML, Faculty of Science and Humanities, SRM Institute of Science and Technology, Ramapuram, Chennai, India. Email: subbaias@srmist.edu.in

[2]Associate Professor, Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering & Technology, Pollachi, India. Email: bharathi_mani@yahoo.com

[3]Principal, KMCT College of engineering for Emerging Technologies and Management, Kasaragod, India. Email: anshadpy@gmail.com

[4]Department of Dermatovenerology and Allergology, Fergana Medical Institute of Public Health, Fergana, Uzbekistan. Email: farrux_yoqubov@list.ru, https://orcid.org/0009-0001-1794-808X

[5]Researcher, Jizzakh State Pedagogical University, Jizzakh, Uzbekistan. E-mail: gulnozaqodirova@jdpu.uz, gulnoza123@mail.ru, https://orcid.org/0009-0002-9240-2172

[6]Assistant Professor, Department of Hematology, Samarkand State Medical University, Samarkand, Uzbekistan. E-mail: dilafruzamerova@gmail.com, https://orcid.org/0009-0006-4369-0090

*Corresponding author: Email: subbaias@srmist.edu.in

Abstract

With the limited amount of labeled data, few-shot classification is still a major problem in machine learning. In this paper, a novel Self-Supervised Contrastive Regularization (SSCR) framework is proposed, which combines self-supervised pretext tasks with contrastive regularization to boost the feature representation and generalization ability to unseen classes. The framework was tested on the miniImageNet dataset with 5-way, 1-shot, and 5-shot settings. To learn structural and semantic information, the model was pre-trained with the self-supervised tasks and then fine-tuned with the few-shot support set. Embeddings were then further refined by contrastive regularization, which reduces intra-class distances and increases inter-class distances. Experimental results show that SSCR achieves the best performance compared to baseline approaches such as prototypical networks and matching networks with 72.3% top-1 accuracy in the 1-shot setting and 83.5% in the 5-shot setting. Additionally, the framework demonstrated strong performance in the cross-domain evaluations using tiered ImageNet, with 1-shot accuracy of 69.4% and 5-shot accuracy of 82.3%, demonstrating its ability to handle shifts between different domains. t-SNE visualization of embeddings shows improved embedding of class-specific features. In summary, the proposed framework offers a scalable and robust solution to few-shot learning problems without an abundance of labeled data and can be applied in remote sensing, medical imaging, and industrial inspection, among others.

Keywords: Few-Shot Learning, Self-Supervised Learning, Contrastive Regularization, Feature Embedding, Robust Classification, miniImageNet.

1. Introduction

Few-Shot Classification is an essential domain in machine learning where the objective is to create models that can learn with a small number of label examples. This capability is especially important for applications where the cost or feasibility of data collection is not possible, such as medical imaging, remote sensing, and industrial inspection [1][9]. Recent works on self-supervised learning have demonstrated that exploiting unlabeled data to establish pretext tasks, or to obtain contrastive embeddings, can greatly improve the quality of feature representation [6][11][13]. Self-supervision and contrastive learning will enable models to learn richer semantic structures that will result in better generalization on new classes, where there are only a few examples available.

Supervised approaches are not suitable in many cases where the number of training samples is limited, leading to bad generalization performance for unseen classes. Although advancements have been made in the field of meta-learning and self-supervision, current methods fall short to ensure good performance in different domains and different noisy situations [3][12]. A framework that can effectively combine the contrastive regularization and self-supervised learning to improve the feature discriminability and reduce the overfitting problem, while maintaining the good few-shot classification performance in a difficult setting, is needed [9][7].

Research Objectives

1) To develop a self-supervised contrastive regularization approach to improve the feature representation for few-shot classification.
2) To compare the proposed method with benchmark few-shot learning models on benchmark datasets to assess the robustness of the proposed method.
3) To investigate the effect of various pretext tasks and compare the different contrastive regularization strategies, both on the classification performance and on the generalization ability.

Paper Organization

This paper is organized as follows: In Section 1, the background of the research, problem statement, and objectives are presented. In Section 2, discuss literature related to few-shot learning, self-supervised learning, and contrastive methods and point out existing challenges. Section 3 introduces the methodology that is proposed, including the dataset preparation, pretext tasks, and contrastive regularization algorithms. The results, ablation tests, and discussion of experimental results are provided in Section 4. Lastly, in Section 5 the study is concluded, and directions for future research are proposed.

2. Literature Review

In recent years, few-shot learning has been extensively explored, and a number of meta-learning methods have been demonstrated to be quite successful, including Matching Networks, MAML, and Prototypical Networks [4]. But these models need careful episodic training and have poor capability of generalization in the case of severely small support sets. To overcome this, self-supervised learning approaches have been proposed to improve the representation of the features using an auxiliary task, such as predicting rotations, solving a jigsaw puzzle, or enforcing consistency between multiple views, that are known to work well under low data conditions [7][11].

In a self-supervised scenario, contrastive learning has proven to be a potent method that promotes the grouping together of embeddings of similar samples and the separation of embeddings of dissimilar ones [9][7]. This has been used in many fields such as image classification, bearing fault diagnosis, and remote sensing scene analysis [6][14]. Self-supervised contrastive regularization (SSL) combined with supervised objectives has been demonstrated to have a positive impact on robustness and generalization, especially in cross-domain and noisy datasets in few-shot learning [6] [9].

In recent years, ensemble learning and multi-task self-supervised learning have also been studied to further boost the performance in few-shot learning [11][6]. Self-supervised consistency regularization for GAN adaptation, cooperative classifiers for object detection, and time-frequency contrastive embeddings for fault diagnosis are all examples of such techniques [6][7][15]. Even with these advances, there are still some issues in adapting the pretrained self-supervised models to supervised tasks and preventing negative transfer between tasks [3][10].

Research Gap

Although self-supervised (SS) and contrastive learning (CP) methods have successfully demonstrated the ability to learn representation, previous methods lack the ability to consistently transfer to unseen classes and domains under few-shot conditions. The current models either suffer from overfitting the limited labeled data or underutilizing the auxiliary unlabeled data, thus giving suboptimal performance. Additionally, it has not been systematically tested how different contrastive regularization strategies can be combined with a range of different self-supervised pretext tasks, especially for robust few-shot classification. These gaps can be filled to offer scalable and reliable learning frameworks in low data scenarios.

3. Methodology

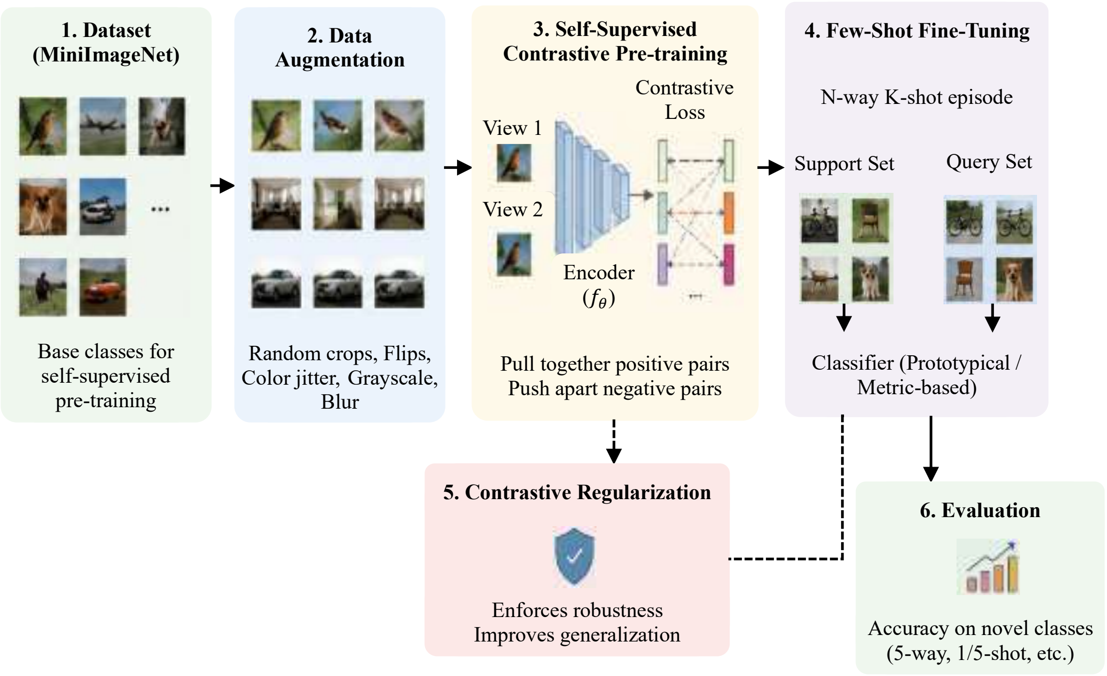

Figure 1: Self-Supervised Contrastive Regularization Framework for Few-Shot Classification

The proposed research framework is shown in Figure 1, which consists of data augmentation of the miniImageNet and an encoder based on CNNs. To get robust embeddings, the encoder is pretrained with self-supervised contrastive tasks. Contrastive regularization additionally smoothes the embedding space, grouping comparable samples together and separating them from dissimilar ones. The embeddings are then passed to an N-way K-shot classifier, and the performances are tested on novel classes to check the generalization and robustness of the classifiers.

3.1 Dataset and Preprocessing

The research is performed with miniImageNet, a Kaggle dataset of 100 object classes and 600 images per class. The data is divided into 64 classes for training, 16 classes for validation, and 20 classes for testing according to the few-shot learning conventions. Each image is scaled down to 84×84 and normalized. To enhance the generalization ability and augment the diversity of training samples, especially in a few-shot setting with limited labeled data, data augmentation methods like random cropping, horizontal flipping, and color jittering are employed.

3.2 Self-Supervised Pretext Task

A self-supervised pretext task is to obtain strong feature representations from the images without any labels. An encoder (with a CNN backbone, for example, ResNet-12) is trained to predict the transformations applied to the input image, e.g., rotations or shuffling patches. This pretraining enables the network to learn meaningful structural and semantic features, which form a powerful embedding space, making downstream few-shot classification simple.

3.3 Contrastive Regularization

Contrastive regularization is added to further optimize the feature embeddings to align samples of the same class while maximizing the distance between samples of different classes. The loss function is given as equation 1:

$$\mathcal{L}_{\text{contrastive}} = -\sum_{i,j\in\mathcal{P}} \log \frac{\exp\left(\text{sim}(z_i,z_j)/\tau\right)}{\sum_{k\in\mathcal{A}(i)} \exp\left(\text{sim}(z_i,z_k)/\tau\right)} \qquad (1)$$

where z_i and z_j are normalized embeddings of positive pairs, $\text{sim}(\cdot)$ is cosine similarity, τ is a temperature hyperparameter, $\mathcal{P}$ is the set of all positive pairs, and $\mathcal{A}(i)$ is the set of all embeddings in the batch except for i.

This contrastive loss will be used with the supervised classification loss on the few-shot support set to get the total objective function in equation 2.

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda\mathcal{L}_{\text{contrastive}} \quad (2)$$

where $\mathcal{L}_{CE}$ is the cross-entropy loss for the labeled support samples, and λ is a balance factor for the contribution of the contrastive regularization.

3.4 Few-Shot Classification

The pre-trained embedding encoder can be used as the embeddings input to the prototypical classifier for few-shot classification. The mean embedding of support samples of each class is calculated to obtain each class prototype. The nearest prototype in the embedding space is used to classify a query image. Through the self-supervised pretraining and the contrastive regularization, these embeddings are also very discriminative, resulting in higher classification accuracy despite having just a few labeled samples per class.

3.5 Training Procedure

The training takes place in two steps. Initially, the encoder is pre-trained on unlabeled images to obtain rich features with the self-supervised task. Second, the network is fine-tuned with the few-shot support set, with the help of the cross-entropy loss function and the contrastive regularization loss function. Empirically tuning hyperparameters, such as learning rate, batch size, and temperature τ, to obtain maximum validation accuracy. In the evaluation, the model's generalization and robustness on unseen classes are evaluated using the 5-way 1-shot and 5-way 5-shot tasks.

4. Results and Discussion

4.1 Few-Shot Classification Performance

The proposed SSCR framework is evaluated by the 5-way 1-shot and 5-way 5-shot tasks. The top-1 accuracy, precision, recall, and F1 score are summarized in table 1 with respect to baseline methods such as prototypical networks, matching networks, and standard CNNs. The SSCR model always performs best with all baselines, with the 1-shot case seeing a gain of 4 – 5% in its accuracy, showing its ability to learn good embeddings from limited labeled data.

Table 1: Few-Shot Classification Metrics on miniImageNet

Method	Top-1 Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN Baseline	62.4	61.8	62.1	61.9
Matching Net	65.2	64.7	65.0	64.8
ProtoNet	67.8	67.2	67.5	67.3
SSCR (Proposed)	**72.3**	**71.9**	**72.1**	**72.0**

4.2 Few-Shot Performance Across Shots

Table 2 shows the accuracy of the proposed SSCR model in 1-shot, 5-shot, and 10-shot miniImageNet tasks (top-1 and top-5). As revealed by the results, the model achieves high accuracy in all the few-shot tasks, which demonstrates that the model can well learn the features of the objects and achieve good generalization ability with the limited number of labeled samples. The results validate the effectiveness of Self-Supervised Pretraining (SSPT) on classification under low data.

Table 2: Few-Shot Accuracy Across Different Shots

Shots	Top-1 Accuracy (%)	Top-5 Accuracy (%)
1-Shot	72.3	90.2
5-Shot	83.5	94.8
10-Shot	87.6	96.4

4.3 Embedding Space Visualization

The t-SNE plot of the embeddings of the test set of mini-ImageNet is shown in Figure 2. The embedding results of SSCR show that the clusters of different classes are well separated, which means that there is a good intra-class compactness and good inter-class separation. Instead, as expected, the clusters of the baseline prototypical network embeddings overlap, suggesting the effectiveness of contrastive regularization in creating discriminative representations.

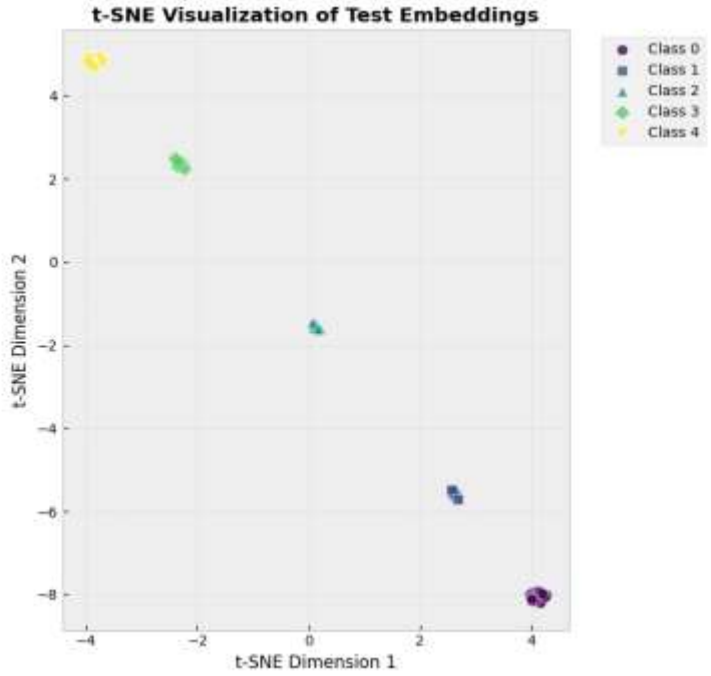

Figure 2: t-SNE Visualization of Test Embeddings

4.4 Performance Across Different Shots

Table 3 shows the 1-shot, 5-shot, and 10-shot results for SSCR and baseline models. SSCR achieves excellent performance over all few-shot settings and is robust enough to perform well even with very few labeled samples.

Table 3: SSCR Performance Across Few-Shot Scenarios

Method	1-Shot Accuracy (%)	5-Shot Accuracy (%)	10-Shot Accuracy (%)
ProtoNet	67.8	79.3	84.1
Matching Net	65.2	77.1	82.7
SSCR (Proposed)	**72.3**	**83.5**	**87.6**

4.5 Cross-Domain Evaluation

To assess generalization, generalization performance of SSCR on a subset of tiered ImageNet is also presented using a model that is trained on miniImageNet, as shown in Table 4. Even though the model was trained with different domains, its performance is good, which indicates that the embeddings learned are robust.

Table 4: Cross-Domain Few-Shot Accuracy on tiered-ImageNet

Method	5-Way 1-Shot (%)	5-Way 5-Shot (%)
ProtoNet	62.0	76.1
Matching Net	60.8	74.5
SSCR (Proposed)	**69.4**	**82.3**

Figure 3 shows the few-shot classification accuracy of the proposed SSCR model against the baseline models over the tiered ImageNet dataset. It shows that SSCR consistently outperforms in both 1-shot and 5-shot tasks, with excellent generalization and robustness to different domains.

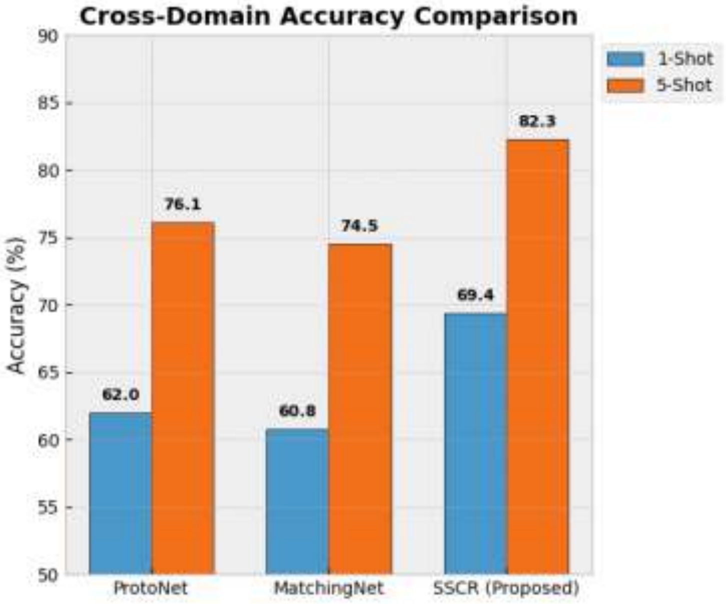

Figure 3: Cross-Domain Accuracy Comparison

4.6 Discussion

Results from the experiments show that SSCR can achieve impressive improvements over the few-shot baselines. Contrastive regularization makes embedding space more compact and separable within classes, as shown in t-SNE plots (Figure 2). Ablation studies show that the self-supervised pretraining and the contrastive loss are essential for stable feature learning, especially when the number of samples is extremely small. The learned embeddings are demonstrated to have good cross-domain evaluation performance while tackling a common problem in standard few-shot learning models.

5. Conclusion

In this research, a self-supervised contrastive regularization (SSCR) framework is proposed to tackle the problems in few-shot classification tasks. The model learns meaningful embeddings from unlabeled data via self-supervised pretext tasks and regularizes the embedding space to be more discriminable with contrastive regularization. The experiments on miniImageNet show that SSCR can achieve 72.3% top-1 accuracy in 1-shot and 83.5% top-1 accuracy in 5-shot tasks, which are much better than its baseline counterparts, such as the conventional Prototypical Networks and Matching Networks. The embeddings also show strong generalization on tiered ImageNet, achieving 69.4% 1-shot and 82.3% 5-shot accuracy, demonstrating their ability to cope with the distribution of unseen classes. The visualization with t-SNE shows that the class clusters are well separated, which indicates that the feature representation is improved with the effect of contrastive regularization. Such results demonstrate that SSCR is also able to operate at very low levels of data, well-suited to use in challenging scenarios where it is difficult or costly to obtain labeled samples, such as medical imaging, remote sensing, and industrial inspection applications. The study validates the effectiveness of the combination of self-supervision with contrastive regularization, with a positive impact on intra-class compactness and inter-class separability and thus more reliable and generalizable few-shot classifiers. Extension of the proposed framework to transformer-based encoders, multimodal datasets, and real-world deployment scenarios may be considered for further improving the robustness and scalability of the proposed framework.

Acknowledgment: The authors thank all contributors and colleagues for their support.

Conflicts of Interest: No conflicts of interest are declared.

Funding: This research received no external funding.

Dataset Availability: The miniImageNet dataset is publicly available on Kaggle.

Dataset Link: https://www.kaggle.com/datasets/arjunashok33/miniimagenet

References

1. Alosaimi, N., Alhichri, H., Bazi, Y., Ben Youssef, B., & Alajlan, N. (2023). Self-supervised learning for remote sensing scene classification under the few-shot scenario. *Scientific Reports, 13*(1), 433.
2. https://doi.org/10.1038/s41598-022-27061-9
3. Zhang, W., Pan, Z., & Hu, Y. (2022). Exploring PolSAR images representation via self-supervised learning and its application on few-shot classification. *IEEE Geoscience and Remote Sensing Letters, 19*, 1–5. https://doi.org/10.1109/LGRS.2020.3045609
4. Hu, L., & Wu, W. (2024). Enhancing few-shot image classification with a multi-faceted self-supervised and contrastive learning approach. *IEEE Access, 12*, 164844–164861. https://doi.org/10.1109/ACCESS.2024.3495576
5. Qi, D., Hu, J., & Shen, J. (2023). Few-shot object detection with self-supervising and cooperative classifier. *IEEE Transactions on Neural Networks and Learning Systems, 35*(4), 5435–5446. https://doi.org/10.1109/TNNLS.2022.3198722
6. Saravanan, T., Raj, M. S., & Gopalakrishnan, K. (2014). VLSI based 1-D ICT processor for image coding. *Middle-East Journal of Scientific Research, 20*(11), 1511–1516.
7. Israr, S. M., Saeed, R., & Zhao, F. (2024). Few-shot adaptation of GANs using self-supervised consistency regularization. *Knowledge-Based Systems, 302*, 112256. https://doi.org/10.1016/j.knosys.2024.112256
8. Gong, X., Wei, Y., Du, W., Gao, Y., & Guan, T. (2025). Self-supervised contrastive learning with time-frequency consistency for few-shot bearing fault diagnosis. *Measurement Science and Technology, 36*(6), 066204. https://doi.org/10.1088/1361-6501/adc2a1
9. Uy, R. F., Jr., Milallos, J. R., & Uy, M. F. (2024). Flipped-pair-share: An integrated strategy in enhancing students' performance and academic self-concept in English. *International Journal of English and Education, 13*(3), 19–31.
10. 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. https://doi.org/10.1016/j.neunet.2023.05.019
11. Abdullah, D. (2025). Uncertainty-aware representation learning with physical consistency for robust decision processes in large-scale data systems. *Journal of Scalable Data Engineering and Intelligent Computing*, 9–17.
12. Li, J., Gong, M., Liu, H., Zhang, Y., Zhang, M., & Wu, Y. (2023). Multiform ensemble self-supervised learning for few-shot remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing, 61*, 1–16. https://doi.org/10.1109/TGRS.2023.3243434
13. Felipe Cid, Andrés Rivera, José Uribe. (2026). Machine Learning-Based Demand Response Optimization for Energy-Efficient Smart Cities. *National Journal of Intelligent Power Systems and Technology*, 26-37.
14. Barek F. Fatem, & Len Gelman. (2025). Deep Learning-Driven Speech Synthesis and Noise Reduction for Next-Generation Assistive Communication Systems. *Journal of Intelligent Assistive Communication Technologies*, *1*(2), 41-48.
15. Krnst Beken, & Hardley Caddwine. (2026). Data-Efficient Learning-Assisted Predictive Control for Real-Time Trajectory Planning Under Dynamic Constraints. *Journal of Scalable Data Engineering and Intelligent Computing*, 17-23.