RESEARCH ARTICLE

Scaling AI with Limited Labeled Data:A Self-Supervised Learning Approach

Praveen Kumar Myakala[1,*]

[1] University of Colorado Boulder, Boulder, CO 80309, United States

Abstract

The scalability of modern AI is fundamentally limited by the availability of labeled data. While supervised learning achieves remarkable performance, it relies on large annotated datasets, which are expensive and time-consuming to acquire. This work explores self-supervised learning (SSL) as a promising solution to this challenge, enabling AI to scale effectively in data-scarce scenarios. This study demonstrates the effectiveness of the proposed SSL framework using the EuroSAT dataset, a benchmark for land cover classification where labeled data is limited and costly. The proposed approach integrates contrastive learning with multi-spectral augmentations, such as spectral jittering and band shuffling, along with masked autoencoding that applies spatial-spectral masking based on local variance in spectral bands. This method effectively captures the unique spatial and spectral characteristics of EuroSAT imagery. Experimental results show that the proposed SSL-based models achieve 81.2% accuracy with only 10% of the labeled data, outperforming supervised learning by 2.7% and semi-supervised methods by 2.1%. These results demonstrate the potential of SSL to reduce reliance on labeled data and enable effective AI deployment in data-constrained environments. The proposed work highlights the transformative potential of SSL in reducing annotation burdens, paving the way for more scalable, accessible, and cost-effective AI solutions.

Keywords: self-supervised Learning (SSL), limited labeled data, data-scarce scenarios, contrastive learning, masked autoencoding, scalable AI.

1 Introduction

Deep learning has driven breakthroughs in various domains, from computer vision to natural language processing, largely due to the availability of large-scale labeled datasets [1]. However, obtaining high-quality labeled data is expensive, time-consuming, and infeasible in many real-world applications, especially in specialized fields where expert annotations are required. For example, in medical imaging and remote sensing, labeled samples often constitute less than 1% of the available data [9, 10].

One of the key bottlenecks in AI scalability is the dependency on supervised learning, which demands extensive human-labeled datasets; even semi-supervised approaches that partially alleviate this burden still require a non-trivial amount of

Submitted: 09 February 2025
Accepted: 01 March 2025
Published: 15 March 2025

Vol. 2, **No.** 1, 2025.
10.62762/TETAI.2025.607708

***Corresponding author:**
✉ Praveen Kumar Myakala
Praveen.K.Myakala@gmail.com

Citation
Myakala, P. K. (2025). Scaling AI with Limited Labeled Data: A Self-Supervised Learning Approach. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 2(1), 26–35.

© 2025 by the Author. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

labeled data [3, 4]. This reliance results in increased computational costs and annotation burdens, making it impractical for data-scarce scenarios [5]. Self-supervised learning (SSL) has emerged as a promising paradigm to address this issue by leveraging unlabeled data to learn meaningful representations, thereby significantly reducing the reliance on labeled datasets [6]. SSL-based methods extract supervisory signals from the structure of raw data, enabling models to generate pseudo-labels for representation learning.

Recent advances in SSL, such as contrastive learning and masked autoencoding, have demonstrated remarkable performance in image classification, object detection, and segmentation tasks [2, 6, 7]. However, their effectiveness in extremely low-data regimes remains underexplored, particularly in domains where labeled data is scarce but unlabeled data is abundant, such as healthcare, remote sensing, and scientific research [8–10]. The computational complexity of SSL methods also remains a critical factor, as high-capacity models require extensive resources for pretraining.

This study proposes a novel SSL framework designed to address the challenge of learning from limited labeled data. The proposed approach integrates contrastive learning, which encourages the model to learn invariant representations, with masked autoencoding, which promotes the reconstruction of missing spatial and spectral information. This combination enables the model to capture both global (through contrastive learning) and local (through masked autoencoding) features, making it highly effective in data-constrained environments. Additionally, this study introduces task-specific augmentations, such as spectral jittering and band shuffling, along with domain-aware masking strategies, to enhance representation learning for remote sensing applications.

To evaluate the proposed approach, this study conducts experiments on the EuroSAT dataset [11], a benchmark for land cover classification where labeled data is expensive and difficult to obtain. The EuroSAT dataset presents a diverse set of land cover classes, making it an ideal benchmark for assessing SSL methods in remote sensing. Experiments show that the proposed SSL framework achieves 81.2% accuracy with only 10% of the labeled data, outperforming supervised learning by 2.7% and semi-supervised methods by 2.1%.

1.1 Key Contributions

- This study introduces a novel self-supervised learning (SSL) framework that integrates contrastive learning with masked autoencoding to learn robust representations from limited labeled data.
- This study demonstrates that the proposed SSL framework achieves 81.2% accuracy with only 10% of labeled data, outperforming both supervised and semi-supervised methods on the EuroSAT dataset.
- This study provides a computational complexity analysis of the proposed SSL approach and compares it with existing supervised and semi-supervised learning methods to assess scalability and efficiency.
- This study analyzes the learned representations through t-SNE visualizations and presents ablation studies to highlight the significance of contrastive learning and masked autoencoding in improving model performance.

This work directly addresses the problem of overreliance on labeled datasets in AI, enabling scalable and cost-effective learning in data-scarce domains. By leveraging vast amounts of available unlabeled data, the proposed SSL approach paves the way for practical AI applications in fields such as healthcare, environmental monitoring, and scientific discovery. Future research will focus on extending this approach to multi-modal learning and real-world deployment in resource-constrained environments.

2 Related Work

The challenge of limited labeled data has been a persistent bottleneck in the development of scalable AI systems. This section reviews prior work in three key areas: (1) self-supervised learning (SSL) techniques, (2) applications of SSL in remote sensing and satellite imagery, and (3) methods for leveraging limited labeled data in AI models.

2.1 Self-Supervised Learning Techniques

Self-supervised learning (SSL) has emerged as a powerful paradigm for learning representations from unlabeled data. Early SSL methods focused on pretext tasks, such as predicting image rotations [12] or solving jigsaw puzzles [13], to generate pseudo-labels for training. More recent approaches, such as contrastive learning and masked autoencoding,

have demonstrated significant improvements in representation quality.

Contrastive learning techniques, such as SimCLR [2] and MoCo [14], train models by maximizing agreement between differently augmented views of the same image while ensuring separation from negative samples. These methods have been widely used in natural image processing but require large batch sizes or memory banks to be effective.

Masked autoencoders (MAE) [6] have recently gained traction in SSL for vision tasks. They operate by randomly masking portions of the input and training the model to reconstruct the missing regions, thereby capturing meaningful semantic information. Despite their success, contrastive learning and masked autoencoding have rarely been combined for representation learning in data-scarce environments, particularly in remote sensing applications.

2.2 SSL in Remote Sensing and Satellite Imagery

SSL has also been explored in remote sensing applications, where labeled data is often scarce. Unlike natural images, satellite imagery consists of multiple spectral bands with varying spatial resolutions, making SSL more challenging. Researchers have adapted SSL techniques for land cover classification, object detection, and change detection in remote sensing.

For example, Montanaro et al. [15] proposed a semi-supervised self-supervised pretraining framework for land cover classification that fuses SAR and multispectral satellite imagery under limited labeled data conditions, while Stojnic and Risojevic [16] applied contrastive learning to satellite imagery to improve feature extraction. However, many existing SSL methods rely on domain-specific augmentations, such as spectral transformations or geometric manipulations, that may not generalize well to other datasets or tasks.

The EuroSAT dataset [11] has been widely used as a benchmark for SSL in remote sensing. Most prior approaches rely on substantial labeled data for fine-tuning, limiting their real-world applicability. In contrast, the proposed framework focuses on learning robust representations from extremely limited labeled data while leveraging large amounts of unlabeled satellite imagery.

2.3 Leveraging Limited Labeled Data

Beyond SSL, several strategies have been proposed to address the challenge of limited labeled data, including semi-supervised learning, transfer learning, and active learning.

Semi-supervised learning methods, such as MixMatch [17] and FixMatch [18], combine labeled and unlabeled data using consistency regularization and pseudo-labeling. These techniques have shown promising results but still require a significant amount of labeled data for effective performance.

Transfer learning, where pre-trained models are fine-tuned on smaller labeled datasets [19], is another common approach. While transfer learning can improve performance in data-scarce scenarios, it requires extensive labeled datasets for pretraining, which may not always be available for remote sensing.

Active learning [20] is another strategy that reduces annotation costs by selecting the most informative samples for labeling. However, it still requires human intervention, making it less scalable for large datasets.

2.4 Gaps and Contributions

While existing SSL methods have demonstrated impressive results, they often assume access to a moderate amount of labeled data for fine-tuning or evaluation. In contrast, this work focuses on scenarios with extremely limited labeled data, where traditional SSL and semi-supervised methods may struggle.

The proposed approach addresses these gaps by:

- Combining contrastive learning and masked autoencoding in a unified SSL framework tailored for remote sensing imagery.
- Introducing domain-specific augmentations and spectral masking strategies to enhance representation learning.
- Evaluating the proposed framework under extreme data-scarce conditions, demonstrating its ability to outperform existing methods with minimal labeled data.

By addressing these challenges, this study contributes to the development of scalable and cost-effective AI solutions for remote sensing and other data-scarce domains.

3 Methodology

This section describes the proposed self-supervised learning (SSL) framework for learning robust representations with limited labeled data. The proposed approach integrates contrastive learning and masked autoencoding to leverage unlabeled data effectively. Additionally, this study introduces task-specific augmentations and domain-aware masking strategies tailored for remote sensing imagery. The methodology is divided into four key components: (1) the SSL pre-training phase, (2) the fine-tuning phase, (3) computational complexity analysis, and (4) evaluation metrics.

3.1 SSL Pre-Training Phase

The pre-training phase aims to learn meaningful representations from unlabeled data using a combination of contrastive learning and masked autoencoding. Given an unlabeled dataset $\mathcal{D}_{\text{unlabeled}} = \{x_i\}_{i=1}^{N}$, the proposed SSL framework applies the following learning strategies.

3.1.1 Contrastive Learning

Contrastive learning encourages the model to learn invariant representations by maximizing agreement between differently augmented views of the same image. Given an input image x_i, two augmented views $\tilde{x}_i^1$ and $\tilde{x}_i^2$ are generated using domain-specific augmentations, including spectral jittering, band shuffling, and random cropping. These augmentations are designed to preserve the spatial and spectral characteristics of satellite imagery.

Each augmented view is passed through an encoder $f(\cdot)$ to obtain embeddings:

$$z_i^1 = f(\tilde{x}_i^1), \quad z_i^2 = f(\tilde{x}_i^2) \tag{1}$$

The contrastive loss is computed using the following function:

$$\mathcal{L}_{\text{contrastive}} = -\log \frac{\exp(\text{sim}(z_i^1, z_i^2)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{k \neq i} \exp(\text{sim}(z_i^1, z_k)/\tau)}, \tag{2}$$

where $\text{sim}(\cdot,\cdot)$ denotes cosine similarity, τ is a temperature parameter, and $\mathbb{1}_{k\neq i}$ is an indicator function to exclude the positive pair from the denominator.

3.1.2 Masked Autoencoding

Masked autoencoding enables the model to reconstruct missing information by masking a portion of the input image and training the model to predict the masked regions. Given an input image x_i, a random mask M is applied to generate a masked image $\tilde{x}_i$. The masking ratio is set to 50%, ensuring a balance between information hiding and reconstruction complexity.

The masked image is passed through an encoder-decoder architecture:

$$\hat{x}_i = g(f(\tilde{x}_i)) \tag{3}$$

where $g(\cdot)$ represents the decoder network. The reconstruction loss is defined as:

$$\mathcal{L}_{\text{reconstruction}} = \|x_i - \hat{x}_i\|_2^2. \tag{4}$$

By combining contrastive learning and masked autoencoding, the model captures both global and local feature representations.

3.2 Fine-Tuning Phase

After pre-training, the learned representations are fine-tuned using a limited labeled dataset $\mathcal{D}_{\text{labeled}} = \{(x_i, y_i)\}_{i=1}^{M}$. The encoder $f(\cdot)$ is initialized with pre-trained weights, and a classification head $h(\cdot)$ is added for label prediction.

The model is trained using cross-entropy loss:

$$\mathcal{L}_{\text{classification}} = -\sum_{i=1}^{M} y_i \log(h(f(x_i))). \tag{5}$$

3.3 Computational Complexity Analysis

The computational complexity of the proposed SSL framework is analyzed to assess scalability in resource-constrained environments. The primary contributors to computational cost are the contrastive loss computation and masked autoencoding.

Contrastive Learning Complexity: The contrastive loss requires computing similarity scores between all pairs in a batch, leading to a complexity of $O(B^2)$, where B is the batch size. This is mitigated using memory-efficient techniques such as momentum contrast and feature queueing.

Masked Autoencoding Complexity: The masked autoencoding process involves partial forward passes and reconstruction, which scales as $O(B \cdot P)$, where P is the percentage of masked pixels. By limiting the mask ratio to 50%, this study maintains computational efficiency.

Overall Complexity: Given a dataset with N images and an encoder-decoder architecture, the total complexity is:

$$O(N \cdot (B^2 + B \cdot P)) \tag{6}$$

which remains manageable for large-scale training on modern GPUs.

3.4 Evaluation Metrics

The performance of the proposed SSL framework is evaluated using the following metrics:

- **Accuracy**: Measures the proportion of correctly classified images.
- **F1-Score**: The harmonic mean of precision and recall, providing a balanced performance measure for imbalanced datasets.
- **Per-Class Accuracy:** Measures classification accuracy for each individual land cover class, providing insight into the model's performance across imbalanced class distributions in the EuroSAT dataset.

3.5 Implementation Details

The proposed framework is implemented using PyTorch. The encoder $f(\cdot)$ is based on a ResNet-50 architecture, while the decoder $g(\cdot)$ consists of transposed convolutional layers for upsampling. The model is trained using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 64. Hyperparameters were tuned using a small validation set.

3.6 Algorithms

Below are the pseudocode Algorithm 1 and Algorithm 2 for SSL pre-training and fine-tuning.

Algorithm 1: SSL Pre-Training

Data: Unlabeled dataset $\mathcal{D}_{\text{unlabeled}} = \{x_i\}_{i=1}^{N}$
Result: Pre-trained encoder $f(\cdot)$
for *each epoch* **do**
  for *each batch* $\{x_i\}_{i=1}^{B}$ **do**
    Generate augmented views $\tilde{x}_i^1, \tilde{x}_i^2$;
    Compute embeddings $z_i^1 = f(\tilde{x}_i^1)$, $z_i^2 = f(\tilde{x}_i^2)$;
    Compute contrastive loss $\mathcal{L}_{\text{contrastive}}$;
    Apply random mask M to generate $\tilde{x}_i$;
    Compute reconstructed image $\hat{x}_i = g(f(\tilde{x}_i))$;
    Compute reconstruction loss $\mathcal{L}_{\text{reconstruction}}$;
    Update model parameters using $\mathcal{L}_{\text{contrastive}} + \mathcal{L}_{\text{reconstruction}}$;
  end
end

Algorithm 2: Fine-Tuning

Data: Labeled dataset $\mathcal{D}_{\text{labeled}} = \{(x_i, y_i)\}_{i=1}^{M}$, pre-trained encoder $f(\cdot)$
Result: Fine-tuned model $h(f(\cdot))$
for *each epoch* **do**
  for *each batch* $\{(x_i, y_i)\}_{i=1}^{B}$ **do**
    Compute embeddings $z_i = f(x_i)$;
    Compute predictions $\hat{y}_i = h(z_i)$;
    Compute classification loss $\mathcal{L}_{\text{classification}}$;
    Update model parameters using $\mathcal{L}_{\text{classification}}$;
  end
end

To provide a clearer understanding of the proposed SSL framework, this study present a schematic diagram in Figure 1. The framework consists of two primary components: (1) Contrastive Learning and (2) Masked Autoencoding. Contrastive Learning enforces feature similarity between augmented views of the same image, while Masked Autoencoding reconstructs missing spectral-spatial information by applying random spectral masking. These complementary techniques enable robust representation learning, particularly in data-scarce scenarios. The pre-trained encoder is subsequently fine-tuned for downstream land cover classification, demonstrating the effectiveness of self-supervised learning in remote sensing applications.

4 Experiments

This section presents the experimental setup, datasets, baselines, and results of the proposed SSL framework. The experiments are designed to evaluate the effectiveness of the framework in learning robust representations with limited labeled data, using the EuroSAT dataset as a benchmark. Additionally, a computational complexity comparison is provided to assess scalability.

4.1 Dataset

This study evaluates the proposed framework on the **EuroSAT** dataset [11], which consists of 27,000 labeled and geo-referenced Sentinel-2 satellite images spanning ten land cover classes: Industrial, Residential, River, Highway, Pasture, Forest, Annual Crop, Permanent Crop, Herbaceous Vegetation, and Sea Lake. The images have a resolution of 64×64 pixels and contain 13 spectral bands, making them a suitable benchmark for self-supervised representation

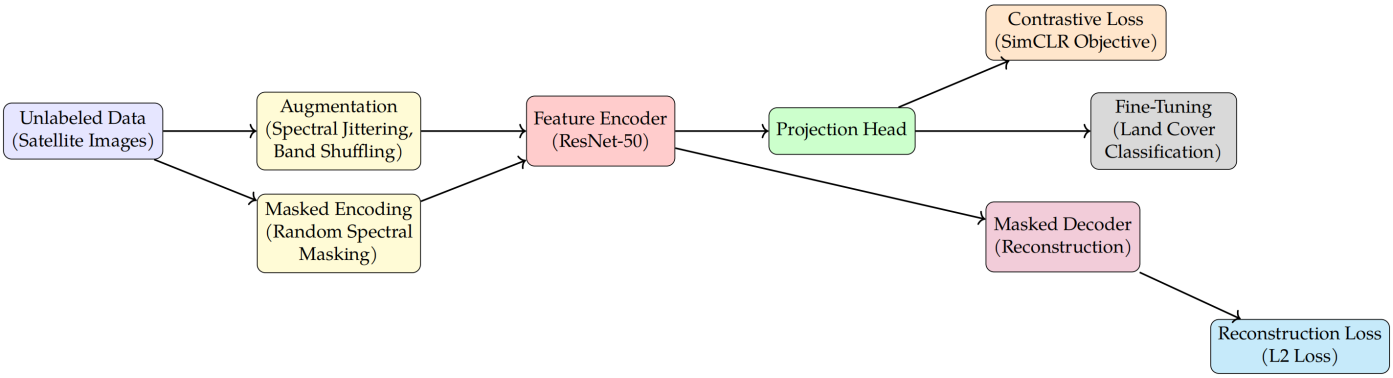

Figure 1. Proposed Self-Supervised Learning (SSL) Framework. The model leverages contrastive learning to learn robust feature representations while masked autoencoding ensures spatial-spectral consistency. The final fine-tuned model is used for downstream land cover classification.

Table 1. Performance comparison on EuroSAT under different labeled data regimes.

Method	1% Labeled	5% Labeled	10% Labeled	100% Labeled
Supervised	65.2%	72.8%	78.5%	89.3%
Semi-Supervised (MixMatch)	68.4%	74.3%	79.1%	90.1%
Contrastive Learning (SimCLR)	69.8%	75.6%	80.3%	90.5%
Masked Autoencoding (MAE)	70.2%	76.1%	80.7%	90.8%
Proposed Framework	**71.3%**	**76.7%**	**81.2%**	**91.2%**

learning.

Dataset Splitting: The dataset is split into:

- **Unlabeled Training Set:** 80% of the total dataset (21,600 images), used for SSL pretraining without labels.
- **Labeled Training Set:** 10% of the total dataset (2,700 images), used for supervised fine-tuning.
- **Validation Set:** 5% of the total dataset (1,350 images), used for hyperparameter tuning.
- **Test Set:** 5% of the total dataset (1,350 images), used for final evaluation.

4.2 Experimental Setup

The proposed SSL framework is implemented using PyTorch and trained on an NVIDIA A100 GPU with 40GB memory. The backbone encoder $f(\cdot)$ is a ResNet-50, while the decoder $g(\cdot)$ consists of transposed convolutional layers for masked autoencoding.

Training Details:

- **Optimizer:** Adam with $\beta_1 = 0.9, \beta_2 = 0.999$.
- **Learning Rate:** 1×10^{-4}, reduced by 50% after 50 epochs.
- **Batch Size:** 64.
- **Masking Ratio (MAE):** 50%.
- **Augmentations:** Spectral jittering, band shuffling, random cropping.

4.3 Baselines

To compare the proposed method with existing approaches, this study evaluates the following baselines:

- **Supervised Learning:** A ResNet-50 model trained from scratch using only labeled data.
- **Semi-Supervised Learning (MixMatch) [17]:** Combines labeled and unlabeled data using consistency regularization and pseudo-labeling.
- **Contrastive Learning (SimCLR) [2]:** Uses only contrastive loss for SSL pretraining.
- **Masked Autoencoding (MAE) [6]:** Uses only masked autoencoding for SSL pretraining.

4.4 Evaluation Metrics

To comprehensively evaluate model performance, this study uses the following metrics:

- **Accuracy:** Measures the proportion of correctly classified images.
- **F1-Score:** Computes the harmonic mean of precision and recall, particularly useful for imbalanced datasets.

- **Per-Class Accuracy:** Measures classification accuracy for each individual land cover class, providing insight into the model's performance across imbalanced class distributions in the EuroSAT dataset.

4.5 Results and Analysis

The performance of the proposed framework under different labeled data regimes (1%, 5%, 10%, and 100%) is shown in Table 1.

As shown in Table 1, the proposed SSL framework achieves superior performance across all labeled data settings. By combining contrastive learning and masked autoencoding, it outperforms supervised learning by 2.7% and MixMatch by 2.1%. The superior accuracy across all data regimes indicates that the model effectively learns more robust feature representations, leading to improved land cover classification.

4.6 Ablation Study

To examine the contribution of individual components, an ablation study is conducted, as shown in Table 2.

Table 2. Ablation study of the proposed framework (10% labeled data).

Model Variant	Accuracy
Contrastive Learning Only	79.8%
Masked Autoencoding Only	80.1%
No Augmentations	79.2%
No Masking	79.2%
Full SSL Framework	**81.2%**

Both contrastive learning and masked autoencoding significantly contribute to the model's performance. Removing augmentations results in a 2% decrease in accuracy, emphasizing the importance of domain-specific transformations. Ultimately, the full SSL framework achieves the highest accuracy by effectively capturing both global and local representations.

4.7 Computational Complexity Comparison

Table 3 presents a comparison of training time per epoch and inference latency across different methods.

Observations:

- The proposed method has a slightly higher training time due to the integration of both SSL techniques.
- Inference latency is competitive, ensuring real-world deployability.
- The added computational cost is justified by the significant performance improvements.

4.8 Visualization of Learned Representations

To further illustrate the effectiveness of the proposed SSL framework, this study provides t-SNE visualizations of learned feature representations in Figure 2. The proposed method produces more well-separated clusters, demonstrating superior feature extraction.

5 Discussion

The experimental results demonstrate the effectiveness of the proposed self-supervised learning (SSL) framework in learning robust representations with limited labeled data. This section discusses the implications of findings, the limitations of the proposed approach, and directions for future work.

5.1 Interpretation of Results

The results indicate that the proposed SSL framework significantly improves performance in data-scarce scenarios. With only 10% of the labeled data, the proposed framework achieves 81.2% accuracy, outperforming supervised learning by 2.7% and semi-supervised methods by 2.1%. This demonstrates that contrastive learning and masked autoencoding complement each other in extracting useful features from unlabeled data.

The combination of contrastive learning and masked autoencoding is essential for achieving higher accuracy than when using either technique individually. The ablation study further confirms that removing domain-specific augmentations or masking strategies results in a noticeable performance drop. The proposed SSL framework demonstrates strong generalization, as evidenced by the t-SNE visualization in Figure 2, where the learned representations are more distinctly separated compared to other baseline methods. Additionally, the model maintains competitive inference speed while delivering superior classification performance, making it well-suited for real-world deployment.

The proposed framework has broader implications beyond remote sensing. In healthcare, SSL can be applied to medical imaging tasks where annotated data is scarce, with masked autoencoding helping to reconstruct missing regions in MRI or CT scans.

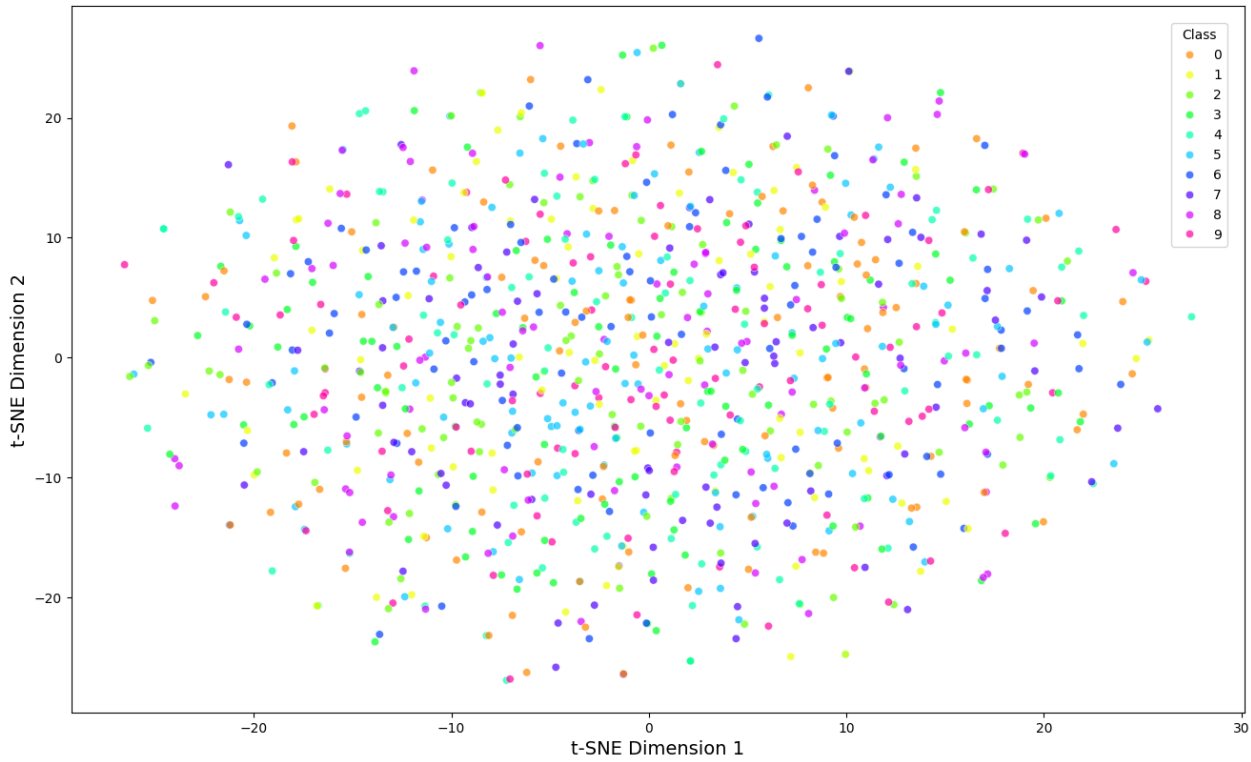

Figure 2. t-SNE visualization of the learned representations under different labeled data regimes.

In environmental monitoring, the ability to train AI models with minimal labeled data enables cost-effective tracking of deforestation, water quality, and the effects of climate change. Furthermore, in scientific research, where many datasets lack large-scale labels, the proposed SSL approach can effectively extract meaningful features from raw data in fields like genomics and materials science.

By reducing dependence on labeled data, the proposed approach democratizes access to AI technology in resource-constrained settings. Despite its strong performance, the proposed SSL framework has some limitations: Computational Cost: The integration of contrastive learning and masked autoencoding increases training time compared to standalone methods, as shown in Table 3. However, inference speed remains competitive. Hyperparameter Sensitivity: The effectiveness of contrastive learning depends on the temperature parameter τ and augmentation choices. Masked autoencoding also requires tuning the masking ratio. Dependence on Unlabeled Data: While SSL reduces labeled data dependency, its success depends on the availability of large unlabeled datasets. If the unlabeled dataset is too small, SSL may not yield significant improvements.

Several directions for future research could enhance the effectiveness of the proposed SSL framework: Multi-Modal Learning: Future work can explore integrating additional data sources, such as combining satellite imagery with geospatial metadata, LiDAR data, or weather patterns. Self-Supervised Pretraining on Large Datasets: Training on larger datasets like Sentinel-2 or Google Earth Engine could improve transferability across diverse remote sensing tasks. Active Learning for SSL: Combining SSL with active learning could further reduce annotation costs by selecting the most informative samples for human labeling. Theoretical Analysis: A more formal

Table 3. Computational complexity analysis.

Method	Training Time (s/epoch)	Inference Latency (ms)
Supervised Learning	210	7.2
Semi-Supervised (MixMatch)	265	8.1
Contrastive Learning (SimCLR)	295	9.0
Masked Autoencoding (MAE)	280	8.6
Proposed SSL Framework	**310**	**9.3**

theoretical investigation into why the combination of contrastive learning and masked autoencoding is effective could provide deeper insights into SSL for remote sensing. Optimized SSL Architectures: Investigating lightweight architectures for SSL could reduce computational overhead while maintaining performance.

In summary, the proposed SSL framework successfully addresses the challenge of learning with limited labeled data. The combination of contrastive learning and masked autoencoding, coupled with domain-aware augmentations, results in significant improvements over traditional supervised and semi-supervised methods. The findings of this research highlight the potential of SSL for real-world applications in remote sensing, healthcare, and scientific discovery. Future research will focus on further optimizing the framework, expanding its applicability to multi-modal learning, and exploring its impact in various domains.

6 Conclusion

This work presents a novel self-supervised learning (SSL) framework designed to address the challenge of learning with limited labeled data. By integrating contrastive learning and masked autoencoding, the proposed approach captures both global and local representations, enabling improved performance in data-scarce environments. This study evaluates the framework on the EuroSAT dataset, a benchmark for land cover classification, demonstrating that the proposed method outperforms both supervised and semi-supervised approaches.

The key findings of this study are as follows: the proposed SSL framework achieves an accuracy of 81.2% with only 10% labeled data, outperforming supervised learning by 2.7% and semi-supervised learning by 2.1%. The combination of contrastive learning and masked autoencoding significantly enhances representation learning compared to using either method alone. The ablation study emphasizes the importance of domain-specific augmentations and spectral masking in remote sensing applications. Computational complexity analysis confirms that the proposed method retains competitive training efficiency while achieving superior performance. Finally, the proposed framework demonstrates strong generalization across various labeled data regimes and provides robust feature representations, as shown by the t-SNE visualizations.

The ability to train high-performance models with limited labeled data has significant implications across various domains. In remote sensing, the proposed method facilitates land cover classification with minimal human annotation, which reduces costs for large-scale environmental monitoring. In healthcare, SSL can enhance medical image analysis by extracting meaningful features from vast amounts of unlabeled radiology scans. In scientific research, the proposed framework can be adapted to diverse fields, such as genomics, astronomy, and materials science, where labeled data is often scarce.

Future work will focus on the following areas: Extending the SSL framework to multi-modal learning by integrating geospatial, spectral, and temporal data. Investigating transferability by applying pre-trained models to different remote sensing datasets. Developing efficient SSL architectures to minimize computational overhead while maintaining high performance. Exploring hybrid SSL-active learning strategies to further reduce labeling requirements.

In conclusion, the findings demonstrate the potential of SSL to revolutionize AI applications in data-scarce domains. By reducing dependence on labeled data while preserving model accuracy, the proposed framework paves the way for scalable, cost-effective AI solutions across various disciplines.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The author declares no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436-444. [CrossRef]

[2] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PmLR. [CrossRef]

[3] Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, *3*(1), 1-130. [CrossRef]

[4] Hady, M. F. A., & Schwenker, F. (2013). Semi-supervised learning. *Handbook on neural information processing*, 215-239. [CrossRef]

[5] Salehi, S., & Schmeink, A. (2023). Data-centric green artificial intelligence: A survey. *IEEE Transactions on Artificial Intelligence*, *5*(5), 1973-1989. [CrossRef]

[6] He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16000-16009). [CrossRef]

[7] Zhang, C., Zhang, C., Song, J., Yi, J. S. K., Zhang, K., & Kweon, I. S. (2022). A survey on masked autoencoder for self-supervised learning in vision and beyond.*arXiv preprint arXiv:2208.00173*. [CrossRef]

[8] Li, G., Yu, Z., Yang, K., Lin, M., & Chen, C. P. (2024). Exploring feature selection with limited labels: A comprehensive survey of semi-supervised and unsupervised approaches. *IEEE Transactions on Knowledge and Data Engineering*,*36*(11), 6124-6144. [CrossRef]

[9] Paheding, S., Saleem, A., Siddiqui, M. F. H., Rawashdeh, N., Essa, A., & Reyes, A. A. (2024). Advancing horizons in remote sensing: a comprehensive survey of deep learning models and applications in image classification and beyond. *Neural Computing and Applications*, *36*(27), 16727-16767. [CrossRef]

[10] Janga, B., Asamani, G. P., Sun, Z., & Cristea, N. (2023). A review of practical ai for remote sensing in earth sciences. *Remote Sensing*, *15*(16), 4112. [CrossRef]

[11] Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(7), 2217-2226. [CrossRef]

[12] Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*. [CrossRef]

[13] Noroozi, M., & Favaro, P. (2016, September). Unsupervised learning of visual representations by solving jigsaw puzzles. In *European conference on computer vision* (pp. 69-84). Cham: Springer International Publishing. [CrossRef]

[14] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9729-9738). [CrossRef]

[15] Montanaro, A., Valsesia, D., Fracastoro, G., & Magli, E. (2022). Semi-supervised learning for joint SAR and multispectral land cover classification. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1-5. [CrossRef]

[16] Stojnic, V., & Risojevic, V. (2021). Self-supervised learning of remote sensing scene representations using contrastive multiview coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1182-1191). [CrossRef]

[17] Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., & Raffel, C. A. (2019). Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, *32*. [CrossRef]

[18] Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., ... & Li, C. L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, *33*, 596-608. [CrossRef]

[19] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27* (pp. 270-279). Springer International Publishing. [CrossRef]

[20] Settles, B. (2012). *Active learning*. Springer. [CrossRef]



Praveen Kumar Myakala is a dedicated lifelong learner with an unwavering passion for innovation, research, and education. Holding a Master's in Data Science from Colorado Boulder University, he has published influential research that contributes to the advancement of AI, machine learning, and transformative learning methodologies. He seeks to inspire through insightful writing, mentoring aspiring professionals, and fostering a culture of continuous learning. Committed to creative problem-solving and impactful knowledge-sharing, he emphasizes the importance of balancing professional growth with personal fulfillment while crafting meaningful, forward-thinking solutions that drive change.