Impact of Pre-Training and Self-Supervised on Model Performance

Ren-Di Wu

whats2000mc@gmail.com

1. Introduction

This report examines the effectiveness of self-supervised learning as a pre-training step for a ResNet18 model, with the rotation task as the central theme.

We analyze how learning to predict image rotations as an unsupervised task can facilitate feature learning that is beneficial for subsequent supervised tasks.

2. Methodology

The experimental setup consisted of several distinct stages designed to explore the utility of pre-training:

1. **Self-Supervised Rotation Model**: The ResNet18 model was initially trained on a rotation prediction task to learn generalizable features without labeled data.

2. **Fine-tuning Pre-trained Model**: Utilizing the self-supervised learned weights, the model was fine-tuned on the CIFAR10 classification task, focusing only on the final convolutional and fully connected layers.

3. **Fine-tuning Randomly Initialized Model**: Parallel to the pre-trained model, a randomly initialized model was fine-tuned under the same conditions for direct comparison.

4. **Fully Supervised Pre-trained Model**: The pre-trained model was subjected to full supervised training on the CIFAR10 dataset.

5. **Supervised Randomly Initialized Model**: A new model, without pre-trained weights, was trained on the CIFAR10 dataset in a fully supervised manner.

Each model's performance was meticulously recorded to assess the influence of pre-training.

3. Results

The results were organized based on the sequence of training stages:

- **Self-Supervised Rotation Model**: The model achieved 79.07% accuracy on the test set. It laid a strong foundation for feature extraction that was instrumental in downstream tasks. (Figure 1)

- **Fine-tuning Pre-trained vs. Randomly Initialized Models**: The pre-trained model achieved 60.00% accuracy, while the randomly initialized model achieved 44.90% accuracy on the test set. A comparison of their performances is presented, with insights into the impact of pre-training on fine-tuning. (Figure 2, Figure 3)

- **Full Supervision on Pre-trained vs. Randomly Initialized Models**: Under full supervision, the pre-trained model recorded 84.74% accuracy, whereas the randomly initialized model showed 83.42% accuracy. This segment delves into the comparative analysis of both models' performances. (Figure 4, Figure 5)

4. Analysis Interpretation

My analysis revealed that pre-training on the self-supervised rotation task significantly boosts the ResNet18 model's ability to learn generalizable features. This foundational learning proved instrumental in enhancing the model's performance on subsequent supervised tasks.

By comparing a fine-tuned model, which utilized the weights learned from the self-supervised task, against a model with randomly initialized weights, I observed noticeable differences in performance.

The pre-trained model consistently outperformed the randomly initialized model on the CIFAR10 classification task. These findings underscore the advantages of transferring learned representations from an unsupervised context to a supervised one.

Furthermore, the study explored the role of label smoothing in model training. It was found that label smoothing at a parameter of 0.1 contributes to the robustness and generalization of the model, preventing overfitting by softening the confidence on label assignments.

A more detailed comparison of the effects of label smoothing across different parameters will be provided in the second experience report, which will focus on how the label smoothing affect the performance.

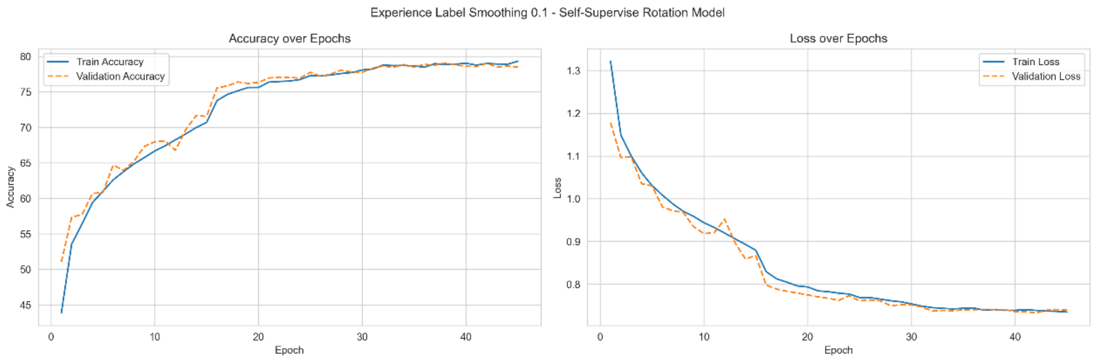
5. Conclusion

The study confirms the benefits of using pre-trained models for downstream tasks, emphasizing the effectiveness of self-supervised learning in improving generalization and performance.

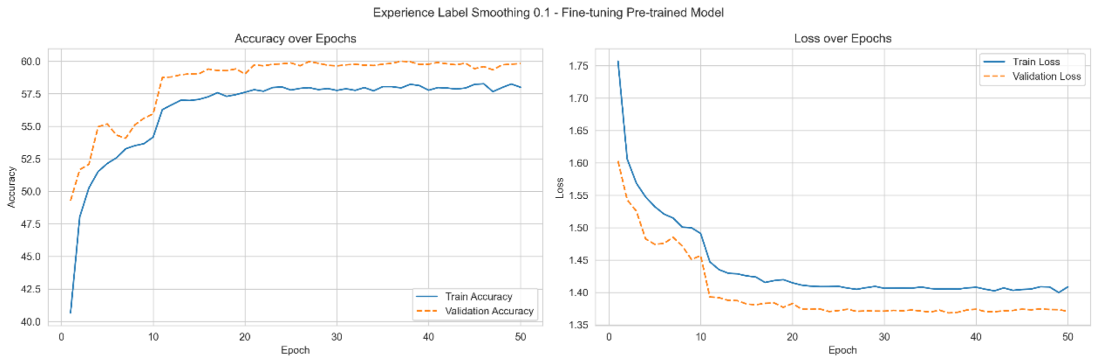
These findings underscore the value of pre-training, particularly in scenarios where labeled data is scarce or the task is complex. (Figure 6)

6. Appendices

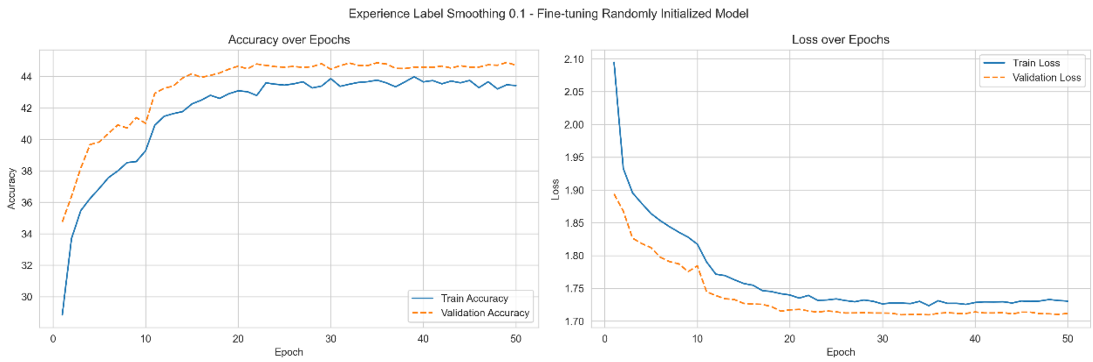
The appendices would include detailed tables and graphs of the experiment results, as well as any additional statistical analyses conducted.



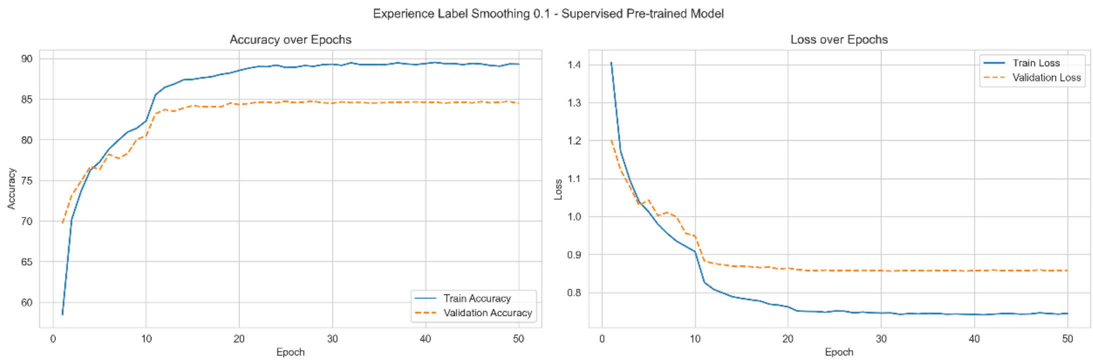
^ *Figure 1: Self-Supervised Learning*



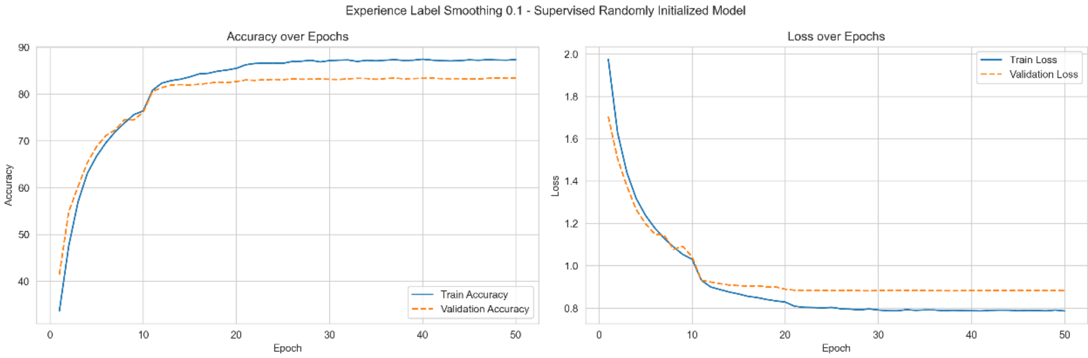
^ *Figure 2: Fine-tuning Pre-trained Model*



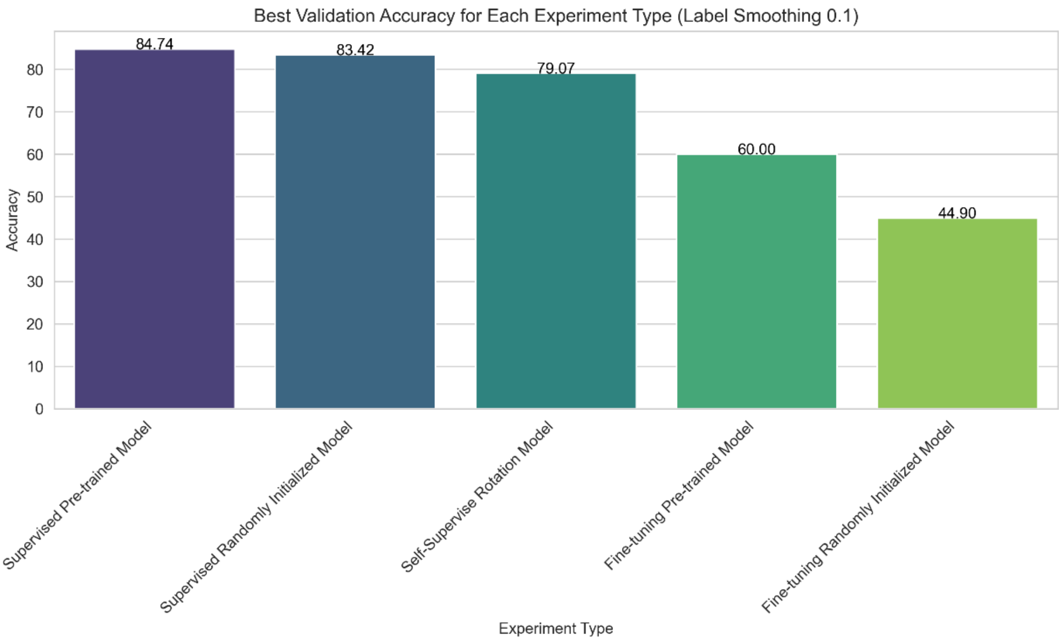
^ *Figure 3: Fine-tuning Randomly Initialized Model*



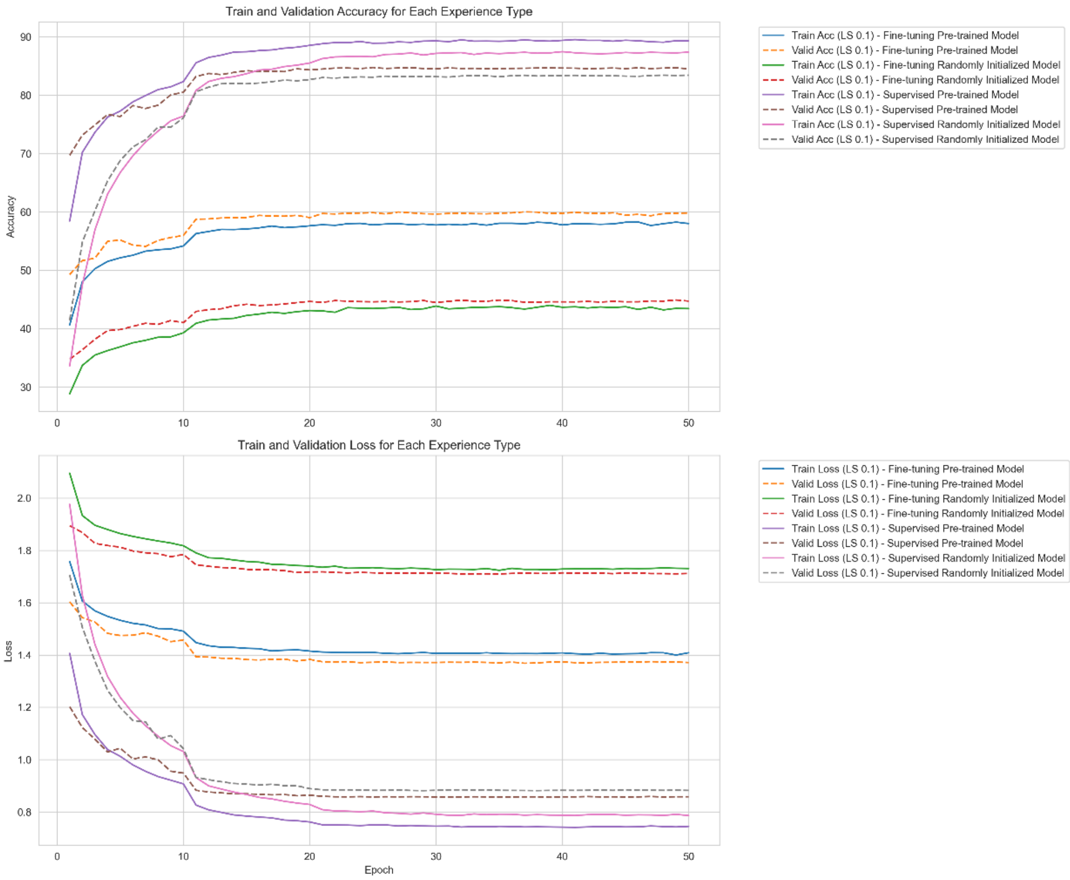
^ *Figure 4: Supervised Learning with Pre-trained Model*



^ *Figure 5: Supervised Learning with Randomly Initialized Model*



^ *Figure 6a: Comparation Between the Models*



^ *Figure 6b: Comparation Between the Models*

Impact of Label Smoothing on Self-Supervised Learning and Transfer Learning

1. Introduction

The purpose of this series of experiments is to evaluate the effect of label smoothing with values of 0, 0.1, and 0.2 on the performance of self-supervised learning and transfer learning approaches using a ResNet18 model.

2. Methodology

The experimental framework is divided into five stages:

1. **Self-Supervised Learning**: Train a ResNet18 model using a rotation task to learn feature representations without labeled data.

2. **Fine-tuning Pre-trained Model**: Fine-tune the model trained in Stage 1 by unfreezing the last fully connected layer and the final convolution layer.

3. **Fine-tuning Randomly Initialized Model**: Initialize a new ResNet18 model and fine-tune it similarly to Stage 2.

4. **Supervised Learning with Pre-trained Model**: Perform fully supervised training on the model obtained from Stage 1.

5. **Supervised Learning with Randomly Initialized Model**: Train a randomly initialized ResNet18 model in a fully supervised manner.

In stages 2 and 3, the models are fine-tuned, while stages 4 and 5 involve full training sessions. The pre-trained models for supervised learning stages are derived from the self-supervised learning in Stage 1.

3. Results

The comprehensive experiments underscore the significance of label smoothing in model training across various scenarios:

- **Label Smoothing 0.1**: Consistently outperforming other values, label smoothing at 0.1 emerged as the optimal choice, striking a delicate balance between precise label representation and the ability to generalize, leading to the highest accuracy gains in both pre-trained and randomly initialized models across all experiment stages. (Figure 6a)

- **Label Smoothing 0.2 vs. 0**: While label smoothing at 0.2 did demonstrate slight improvements over the absence of smoothing (0) in models initialized with random weights, it could not achieve the performance heights of the 0.1 setting. This outcome indicates that a moderate degree of label smoothing is generally beneficial, enhancing performance in models lacking pre-trained advantages. (Figure 6b)

The findings accentuate the pivotal role of label smoothing in bolstering model performance.

A smoothing parameter set to 0.1 not only facilitated superior accuracy in self-supervised learning contexts but also proved to be more effective during the transfer of knowledge to new tasks, thereby reinforcing its efficacy as a tool for model regularization and performance optimization.

4. Analysis Interpretation

Label smoothing serves as a regularization technique, potentially preventing overfitting by discouraging the model from becoming too confident about its predictions.

The optimal smoothing value of 0.1 could be providing a sweet spot where the model is regularized enough to generalize well without being too penalized for confident predictions.

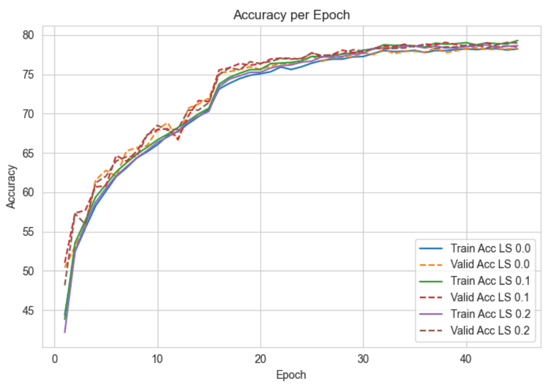
5. Conclusions

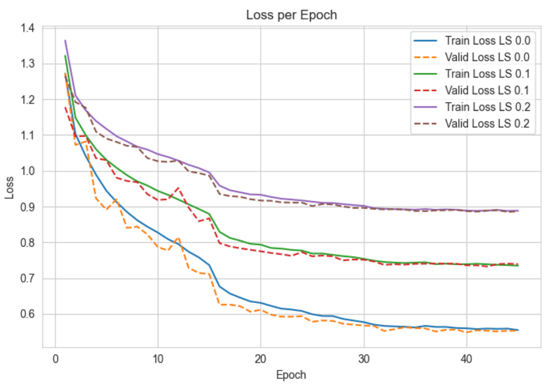
Label smoothing has proven to be a beneficial technique in enhancing the performance of self-supervised and transfer learning tasks. A smoothing value of 0.1 is recommended for similar tasks and architectures. Future experiments might explore:

- The impact of label smoothing on different neural network architectures.
- The effects of label smoothing in larger, more complex datasets.

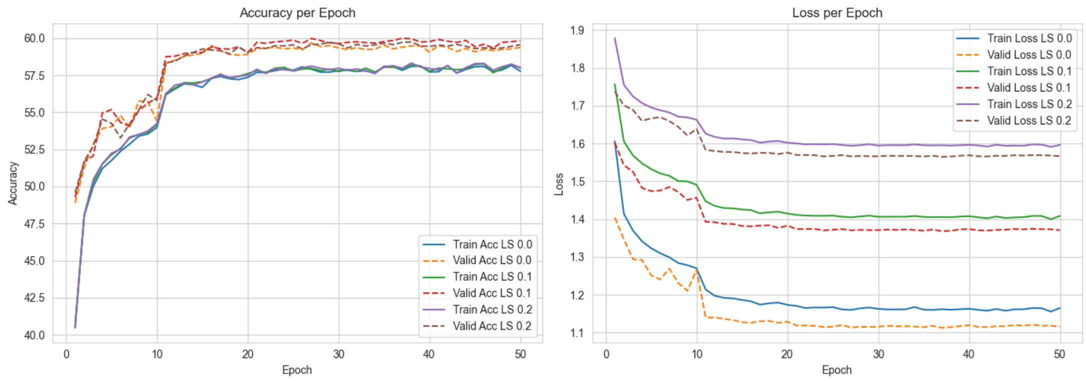
6. Appendices

The appendices would include detailed tables and graphs of the experiment results, as well as any additional statistical analyses conducted.

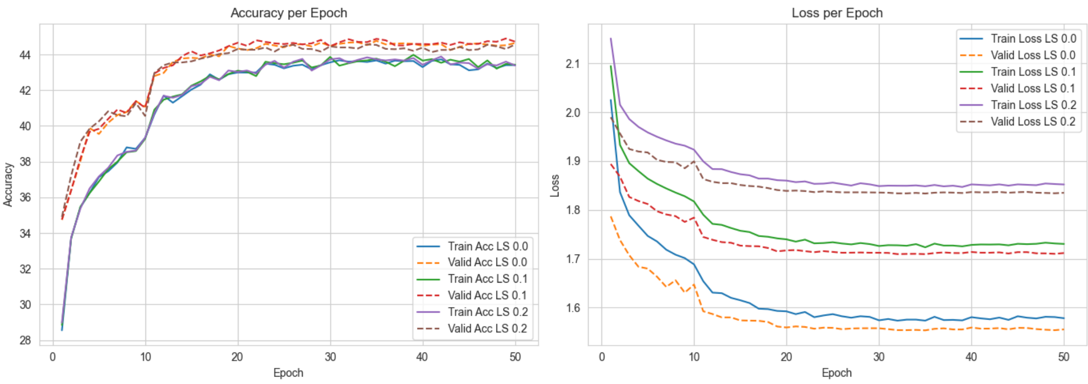




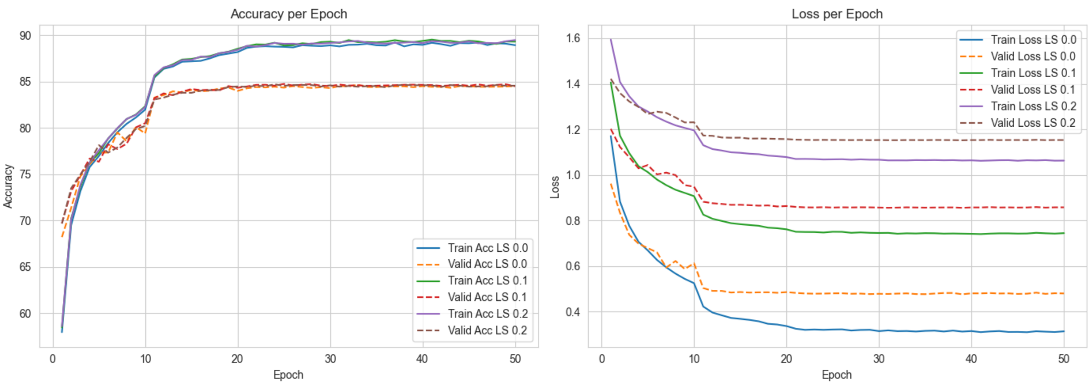
^ *Figure 1: Self-Supervised Learning*



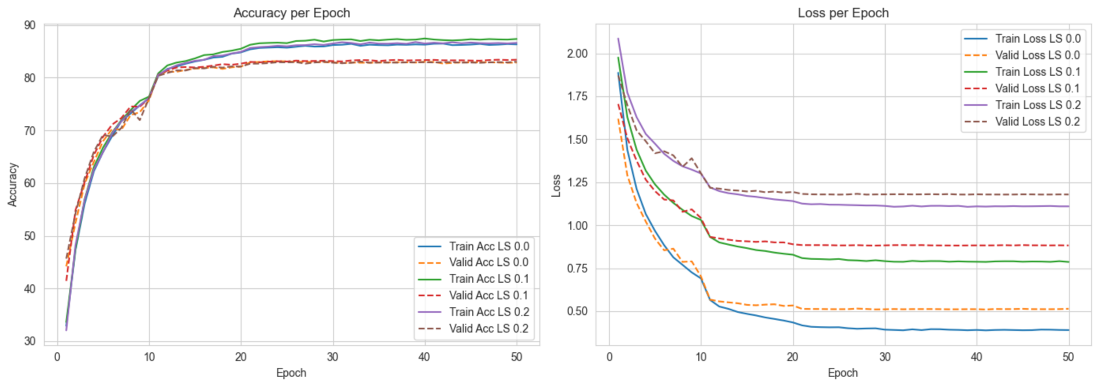
^ *Figure 2: Fine-tuning Pre-trained Model*



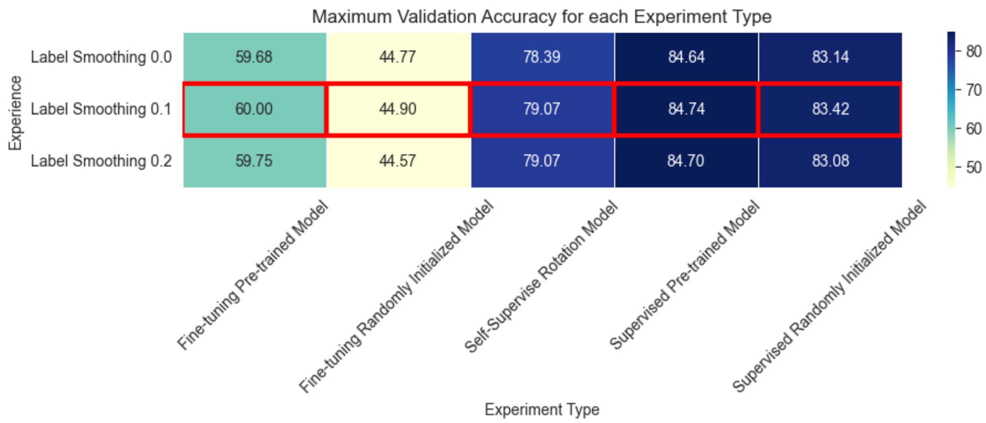
^ *Figure 3: Fine-tuning Randomly Initialized Model*



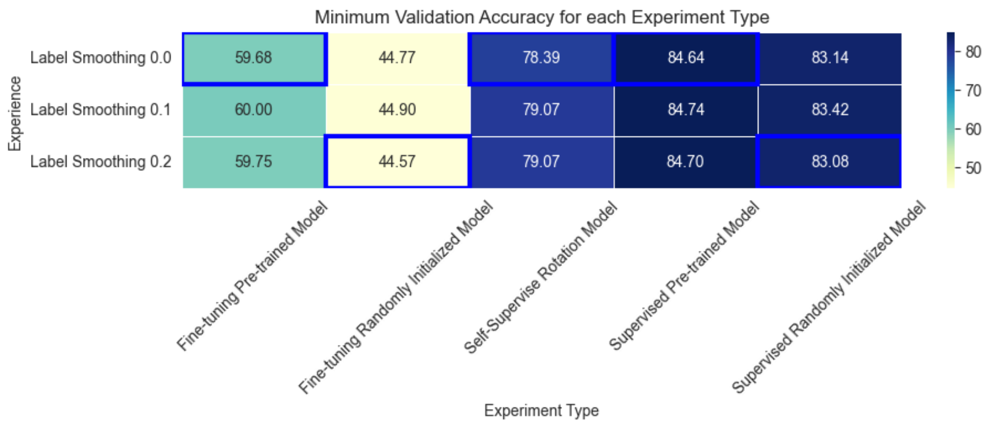
^ *Figure 4: Supervised Learning with Pre-trained Model*



^ Figure 5: Supervised Learning with Randomly Initialized Model



^ Figure 6a: Table for validation accuracy



^ Figure 6b: Table for validation accuracy

Impact of Rotation Counts in Self-Supervised Pre-Training on Downstream Task Performance

1. Introduction

This report presents an analysis of self-supervised learning with a focus on the impact of different rotation counts used during the pre-training phase.

The objective is to evaluate how the granularity of self-supervised tasks affects the performance of a neural network model when fine-tuning for a downstream task.

The rotations considered were 4, 6, and 8 directions, providing a gradient of complexity in self-supervised learning.

2. Methodology

The methodology involved pre-training a ResNet18 model on a self-supervised rotation prediction task with three different rotation granularities: 4, 6, and 8 directions.

Following the self-supervised pre-training, two downstream tasks were performed:

- Fine-tuning the pre-trained model on the CIFAR10 classification task.
- Training the same architecture from scratch (random initialization) on CIFAR10 for comparison.

Each model's performance was evaluated based on test accuracy, providing insights into the efficacy of self-supervised learning with varying rotation counts.

3. Results

The test accuracy results for each experiment type and rotation count were as follows:

- **Self-Supervised Rotation Model**: The highest accuracy was achieved with 6 rotations (85.17%), followed by 4 rotations (79.18%) and 8 rotations (79.87%). (Figure 1)
- **Fine-tuning Pre-trained Model**: The 4 rotations scenario led to the highest accuracy (63.38%), with 8 rotations trailing closely behind (60.46%), and 6 rotations showing a noticeable drop (55.77%). (Figure 2)
- **Fine-tuning Randomly Initialized Model**: All three rotation counts resulted in similar accuracies, hovering around 46%. (Figure 3)
- **Supervised Pre-trained Model**: The models pre-trained with 4 and 8 rotations achieved similar accuracies (84.64% and 84.58%, respectively), while 6 rotations resulted in a slightly lower accuracy

(84.05%). (Figure 4)

- **Supervised Randomly Initialized Model**: All three rotation counts resulted in similar accuracies, hovering around 83%. (Figure 5)

4. Analysis Interpretation

The varying rotation counts in the self-supervised pre-training phase exhibited distinct impacts on the model's performance during fine-tuning.

A higher rotation count did not necessarily translate to better performance, as seen with the 6 rotations count, which, despite having the highest accuracy in the self-supervised task, did not perform as well in the fine-tuning stage.

The results suggest that the complexity of the self-supervised task needs to be carefully balanced to ensure that the learned features are beneficial and generalizable to the downstream task.

5. Conclusions

The investigation into the effects of rotation count in self-supervised learning revealed that more complex self-supervised tasks do not always yield better downstream performance.

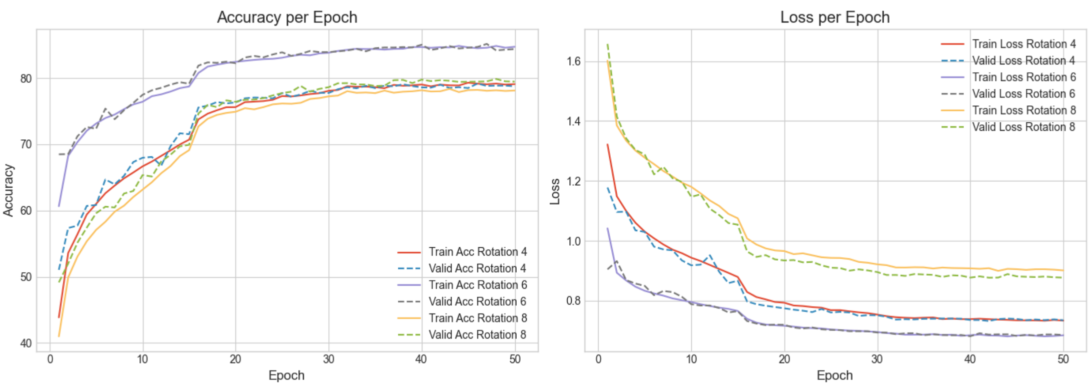
Specifically, the 4 rotations count provided the most effective transfer learning for fine-tuning, indicating that an optimal level of task difficulty might exist for self-supervised pre-training.

These findings underscore the importance of task design in self-supervised learning and its subsequent influence on supervised tasks.

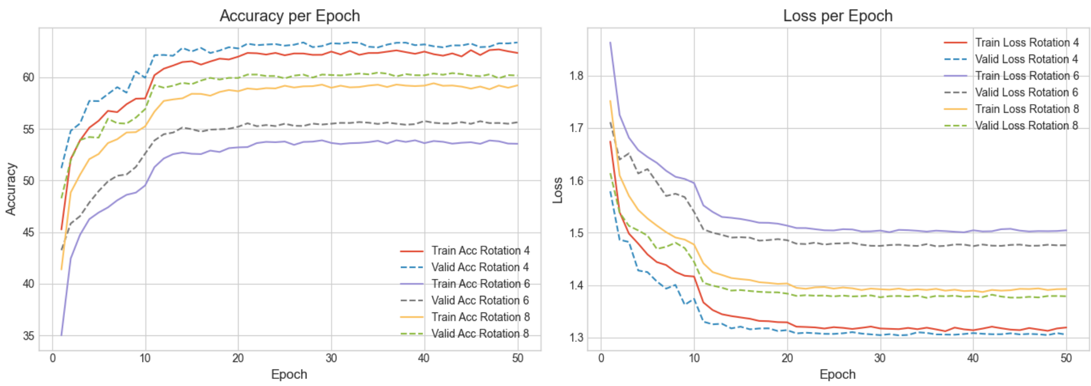
6. Appendices

The appendices would include detailed figures and tables depicting the accuracy results for each experiment type and rotation count.

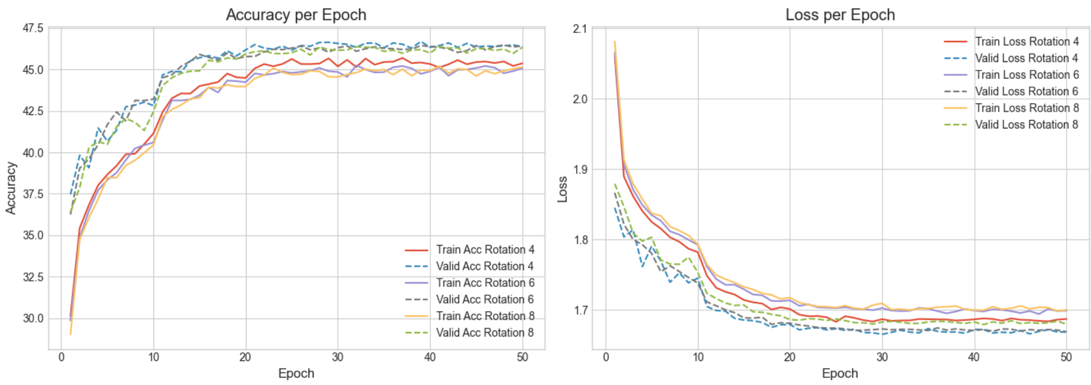
Additional analyses, such as the per-class accuracies, would provide deeper insights into the model's performance.



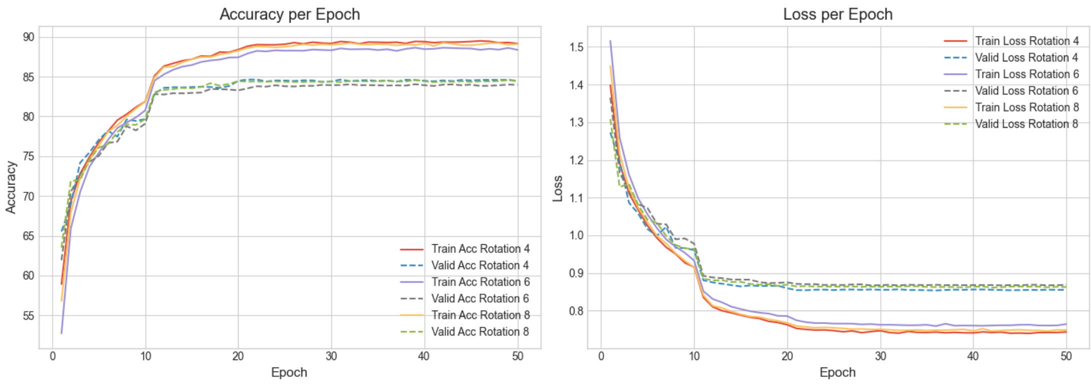
^ *Figure 1: Self-Supervised Learning*



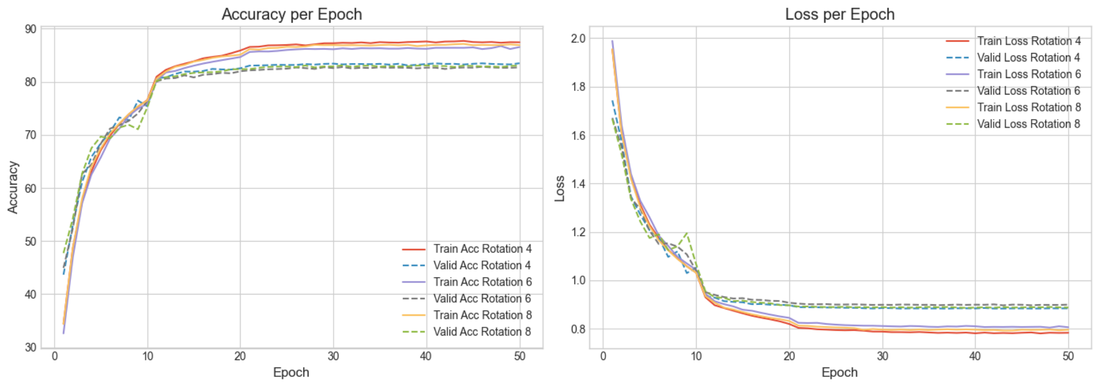
^ Figure 2: Fine-tuning Pre-trained Model



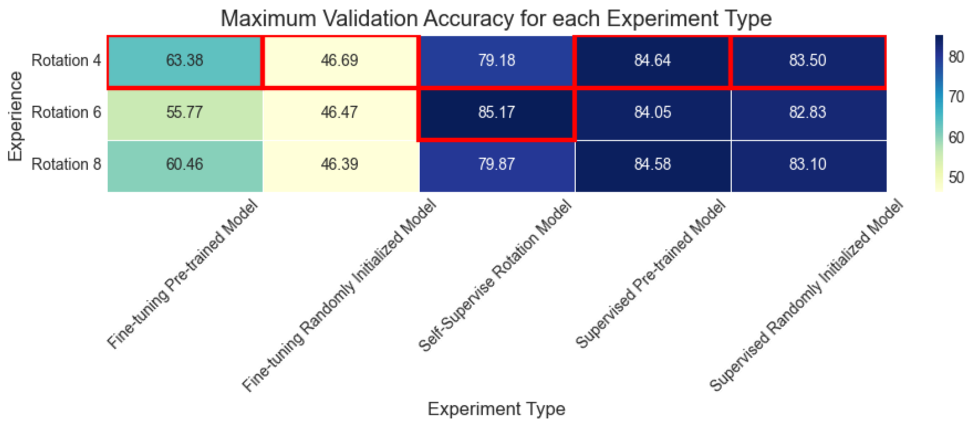
^ Figure 3: Fine-tuning Randomly Initialized Model



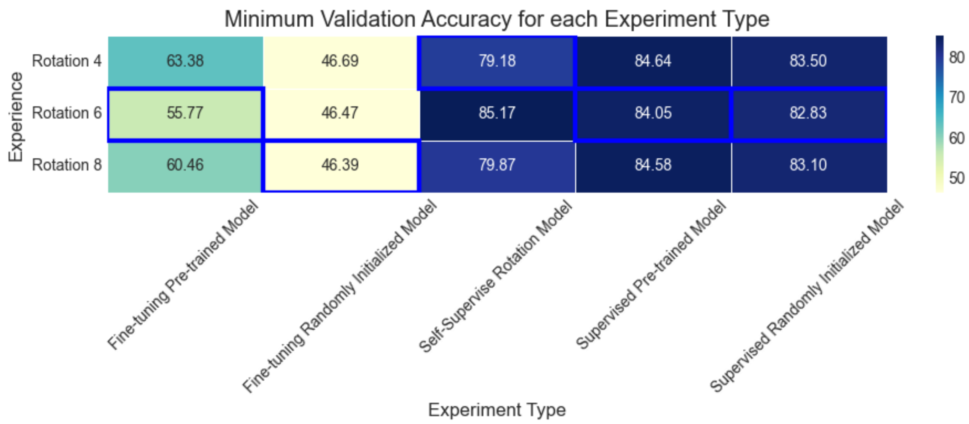
^ Figure 4: Supervised Learning with Pre-trained Model



^ *Figure 5: Supervised Learning with Randomly Initialized Model*



^ *Figure 6a: Table for validation accuracy*



^ *Figure 6b: Table for validation accuracy*

Impact of Classification with Limited Labelled Data

1. Introduction

This report presents a comparative analysis of CIFAR10 classification performance using a self-supervised learning model with fine-tuning (specifically, fine-tuned fully connected and layer 4 convolution blocks) versus a fully supervised model initialized randomly. The central focus of this investigation is to understand the efficacy of self-supervised pre-training in scenarios with varying amounts of labeled data.

2. Methodology

Two distinct training approaches were employed:

- **Self-Supervised Pre-training with Fine-Tuning**: A model pre-trained on a rotation prediction task (RotNet) was fine-tuned on CIFAR10 with only the fully connected (fc) and layer 4 convolution blocks being updated.

- **Fully Supervised Training from Random Initialization**: A model was trained from scratch on CIFAR10 with all layers being updated during training.

For both approaches, the models were trained with different numbers of labeled examples per class: 20, 50, 100, 400, 700 and 1000.

3. Results

The performance of both models was evaluated in terms of validation accuracy. The results were plotted to show the relationship between the number of labeled examples per class and the achieved validation accuracy and validation loss.

Samples Per Class	Fine-Tuned Max Accuracy	Fine-Tuned Min Loss	Random Init Max Accuracy	Random Init Min Loss
20	29.23	2.1380	25.58	2.4882
50	38.74	1.8387	30.34	2.0282
100	45.10	1.7096	39.54	1.8203
250	51.04	1.5691	48.73	1.6271
400	53.50	1.5198	57.61	1.4580
700	55.61	1.4670	66.59	1.2734
1000	56.60	1.4408	66.49	1.2566

Table 1: Maximum Validation Accuracy and Minimum Validation Loss

4. Analysis Interpretation

The fine-tuned model demonstrated higher validation accuracy and lower loss at lower sample sizes (20, 50, 100). However, when the sample size was increased to 400 and above, the fully supervised model started to outperform the fine-tuned model, suggesting that with adequate labeled data, the benefits of self-supervised pre-training become less pronounced. (Figure 1, Figure 2, Figure 3)

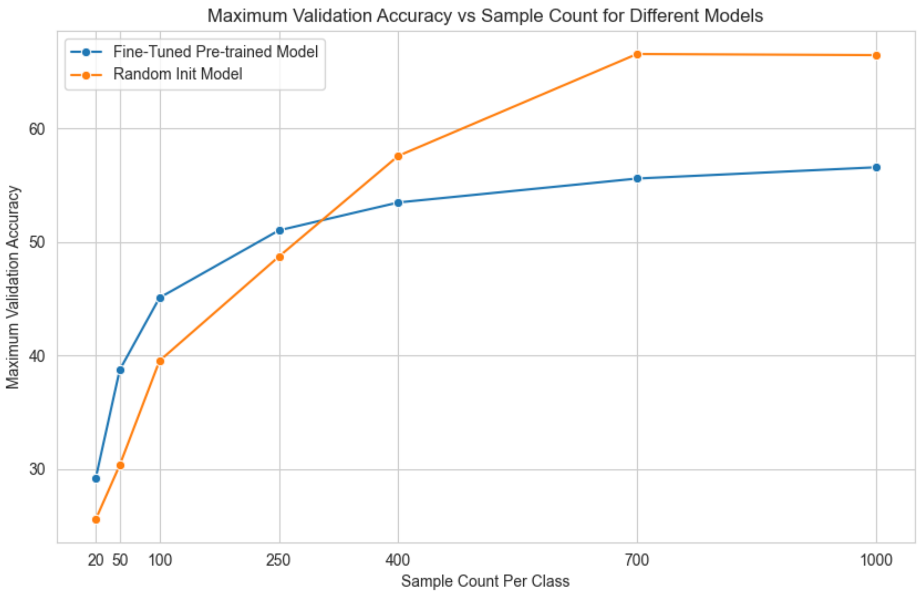
5. Conclusions

The study's findings suggest that self-supervised pre-training with fine-tuning is particularly effective when the availability of labeled data is limited. Notably, in scenarios where labeled data may be scarce or inadvertently leaked, such an approach can be leveraged to enhance model performance. In contrast, with ample labeled data, the advantages of self-supervised pre-training become less pronounced, and a fully supervised approach is shown to be more beneficial.

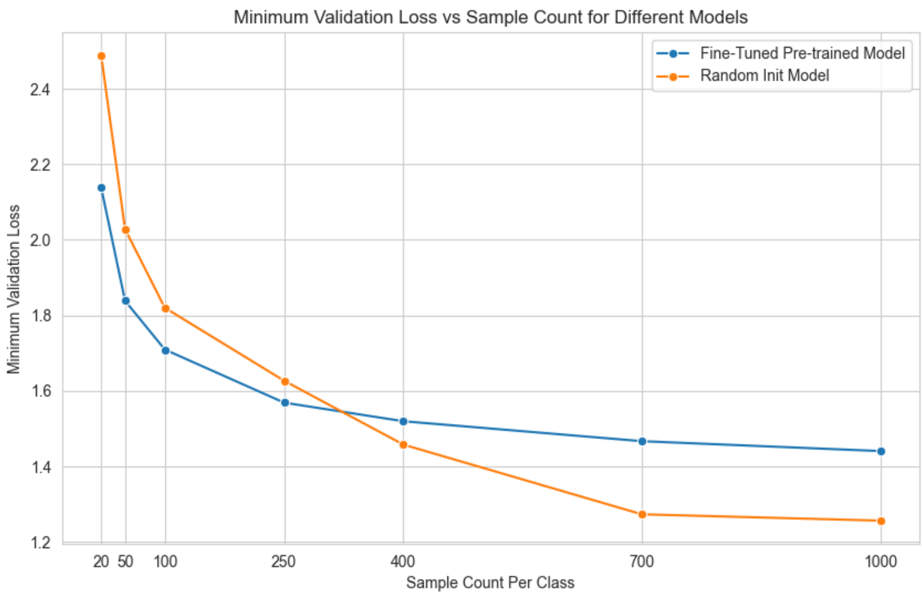
This insight is pivotal for situations where the conservation of labeled data is critical or where data labeling presents a significant cost. Implementing a self-supervised learning paradigm under these circumstances can serve as a strategic method to boost model performance and utilize available data more efficiently.

6. Appendices

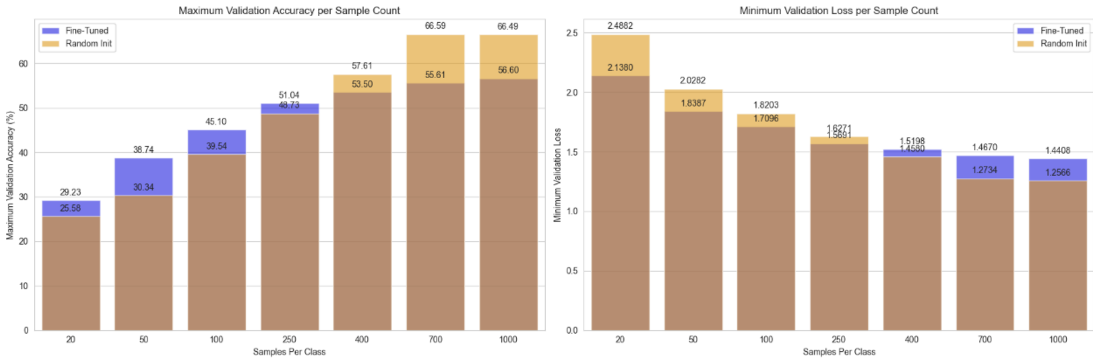
The appendices would include the generated plots that supports the findings of this report.



^ *Figure 1: Test Accuracy Plot*



^ Figure 2: Test Lost Plot



^ Figure 3: Annotated Accuracy Loss Comparison

References

[1] OpenAI. (2023). ChatGPT [Large language model]. https://chat.openai.com

[2] Feng, Z., Xu, C., & Tao, D. (2019). Self-Supervised Representation Learning by Rotation Feature Decoupling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 10364-10373) Retrieved from paper

[3] Müller, R., Kornblith, S., & Hinton, G. E. (2019). When Does Label Smoothing Help? In Advances in Neural Information Processing Systems 32 (NeurIPS 2019). Retrieved from paper