Impact of Pre-Training on Model Performance with Label Smoothing 0.1

Ren-Di Wu

whats2000mc@gmail.com

1. Introduction

This comprehensive report investigates the role of pre-training on the performance of a ResNet18 model, with an emphasis on a label smoothing parameter set to 0.1. The scope of analysis includes a self-supervised rotation model and its subsequent influence on fine-tuning and supervised learning stages.

2. Methodology

The experimental framework was meticulously designed to assess the benefits of pre-training across several training phases:

1. **Self-Supervised Rotation Model**: Initially, the model was self-trained on a rotation prediction task, setting the foundation for feature extraction capabilities without the need for labeled data.

2. **Fine-tuning Pre-trained Model**: Leveraging the self-supervised weights, the model underwent fine-tuning focusing on the final fully connected and last convolution layers.

3. **Fine-tuning Randomly Initialized Model**: For comparison, a model with randomly initialized weights was fine-tuned under identical conditions to the pre-trained model.

4. **Fully Supervised Pre-trained Model**: The self-pre-trained model was then fully trained using supervised learning on the target dataset.

5. **Supervised Randomly Initialized Model**: A fresh model, without the self-supervision advantage, was subjected to full supervised training from scratch.

Label smoothing of 0.1 was consistently applied to all models to explore its combined effect with pre-training and self-supervised learning.

3. Results

The results are segmented into the impact of self-supervised learning and the subsequent stages:

- **Self-Supervised Rotation Model**: Served as a robust pre-training step, enhancing feature representation learning, which proved beneficial in downstream tasks. (Figure 1)

- **Fine-tuning Pre-trained Model**: Exhibited superior gains in performance, validating the importance of pre-training for fine-tuning on specific tasks. (Figure 2)

- **Fine-tuning Randomly Initialized Model**: While improvements were observed, the absence of pre-training was evident in its relatively lower performance. (Figure 3)

- **Fully Supervised Pre-trained Model**: The compounded effects of self-supervision and label smoothing culminated in high accuracy, suggesting a significant transfer of learned features. (Figure 4)

- **Supervised Randomly Initialized Model**: It benefited from label smoothing but was outperformed by the pre-trained models, reinforcing the value of a self-supervised foundation. (Figure 5)

4. Analysis Interpretation

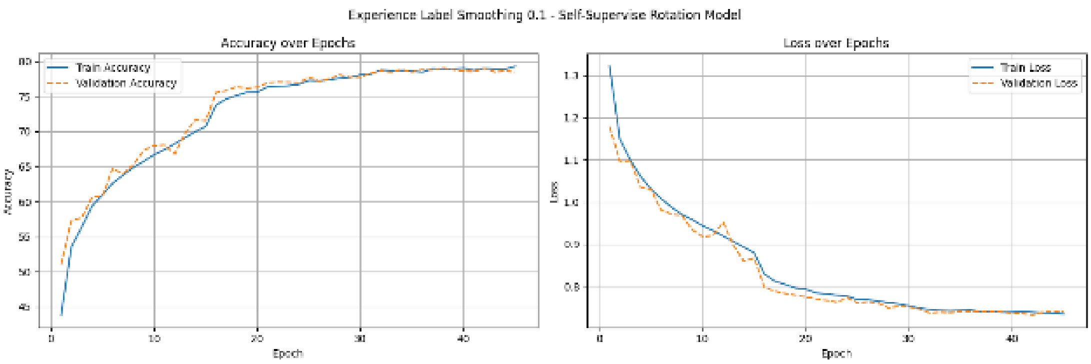
The analysis delineates the clear benefits of a self-supervised pre-training phase. Not only does it bolster the model's capacity for feature generalization, but it also establishes a conducive learning trajectory for subsequent fine-tuning and fully supervised training phases. The integration of label smoothing further aids in mitigating overconfidence and promoting a more distributed learning of classes, which is especially valuable in the presence of noisy or limited labels. (Figure 6)

5. Conclusion

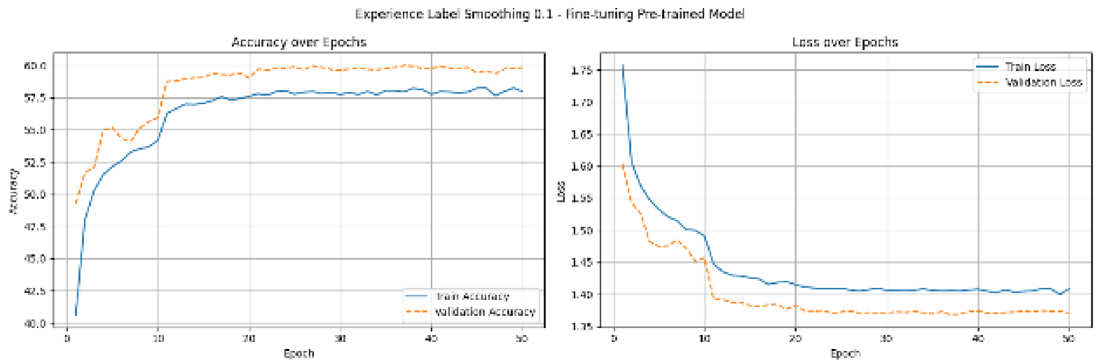
The incorporation of pre-training, specifically through self-supervised learning, alongside label smoothing of 0.1, is a potent combination for enhancing neural network generalization and performance. This empirical study reaffirms the significance of self-supervised pre-training as an indispensable element in the neural network training arsenal, particularly for complex tasks where the model benefits from robust feature extraction and regularization techniques. (Figure 6)

6. Appendices

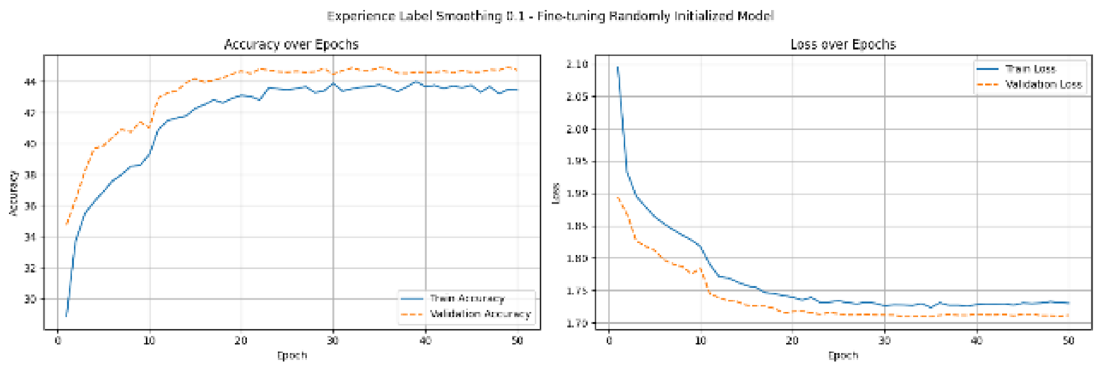
The appendices would include detailed tables and graphs of the experiment results, as well as any additional statistical analyses conducted.



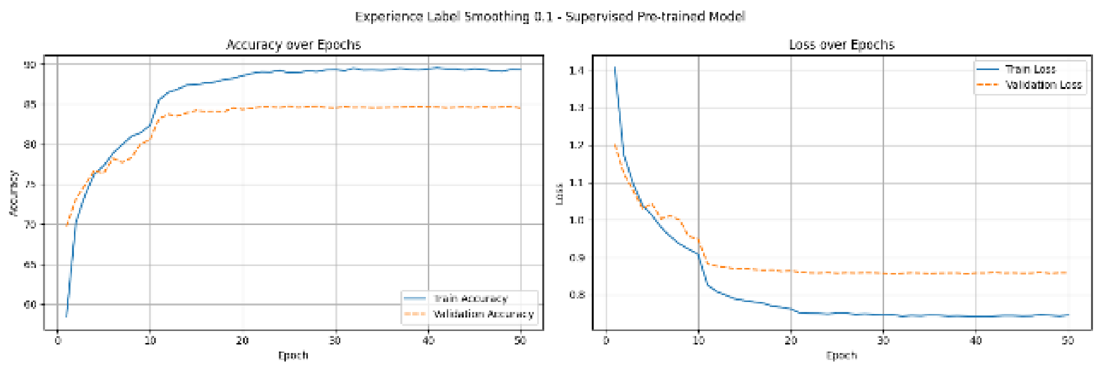
^ *Figure 1: Self-Supervised Learning*



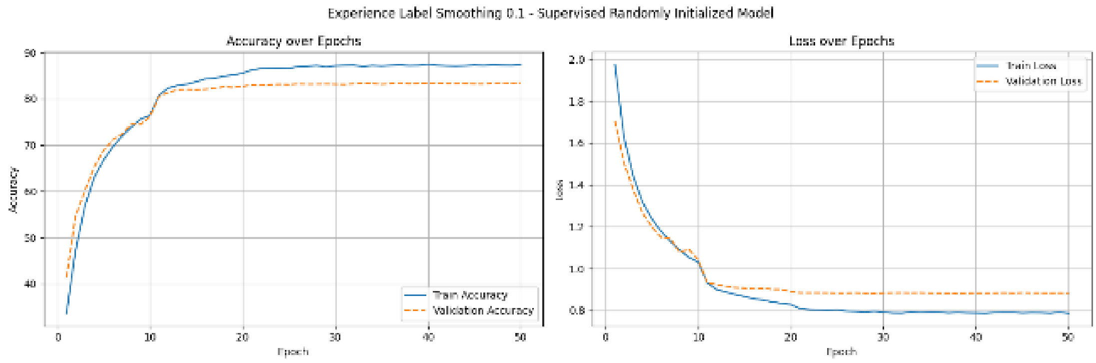
^ *Figure 2: Fine-tuning Pre-trained Model*



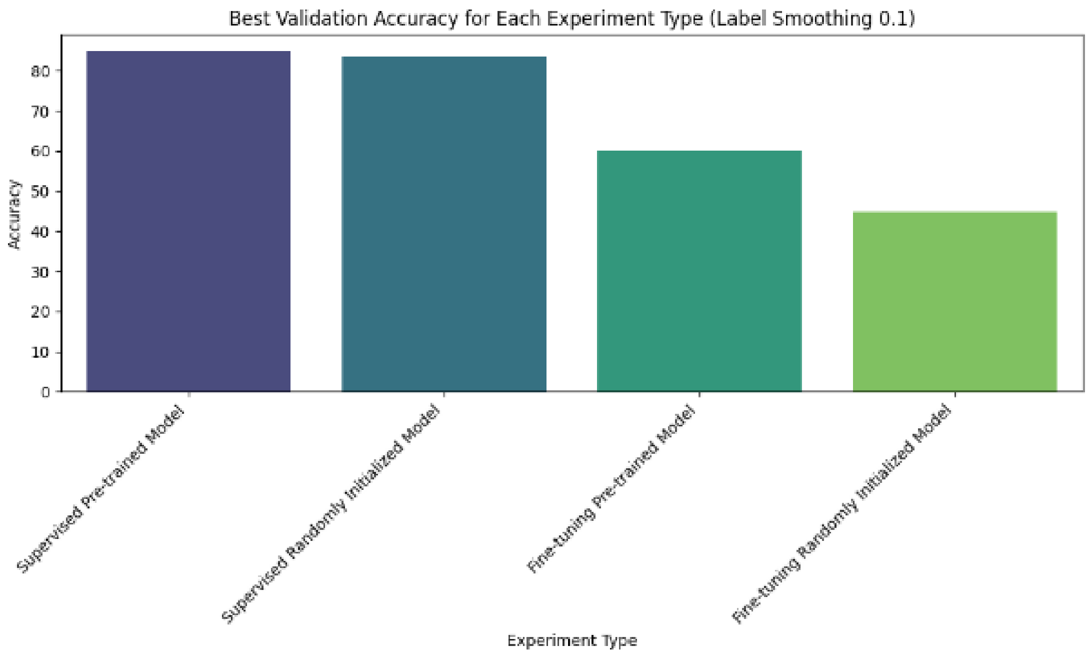
^ *Figure 3: Fine-tuning Randomly Initialized Model*



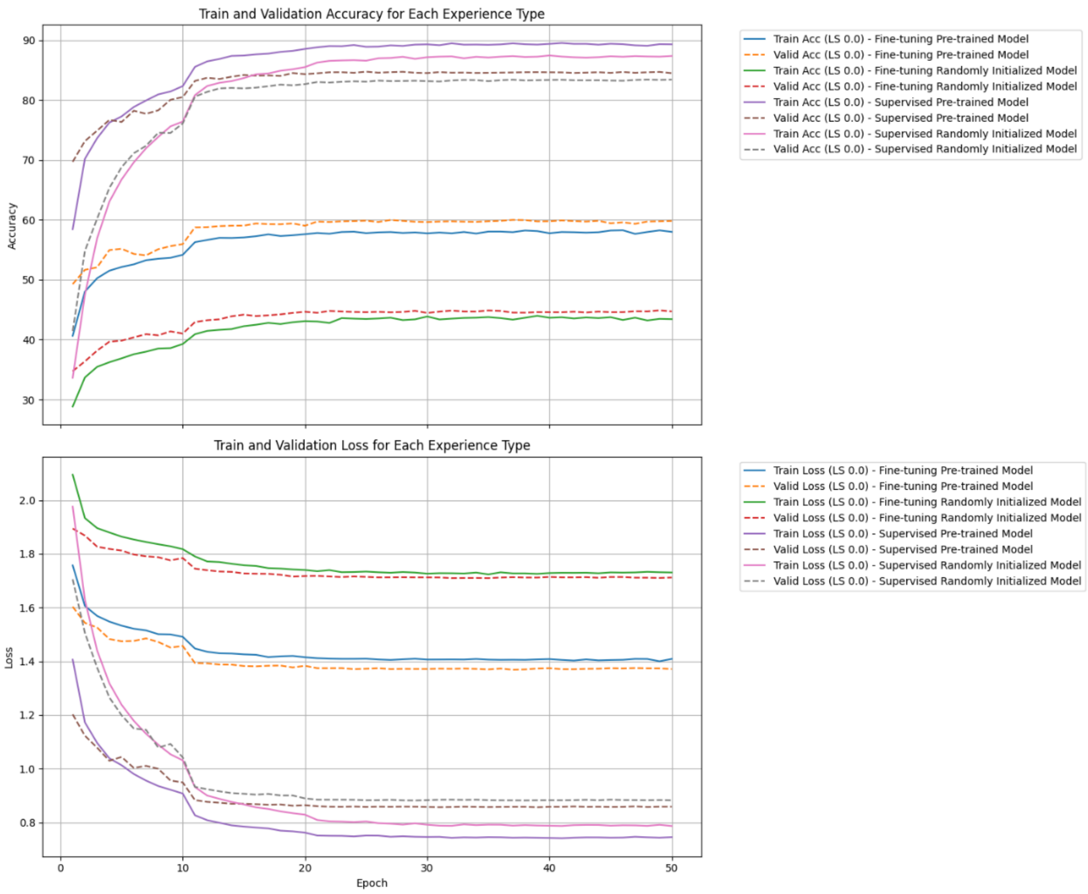
^ *Figure 4: Supervised Learning with Pre-trained Model*



^ *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 experiments revealed that:

- **Label Smoothing 0.1**: This level of label smoothing achieved the best results across all stages of the experiments, indicating a favorable balance between label precision and generalization. (Figure 6)
- **Label Smoothing 0.2 vs. 0**: Although label smoothing with a factor of 0.2 did not outperform the 0.1 setting, it won in most cases against the no smoothing condition (0), suggesting that some degree of label smoothing generally aids in model performance. (Figure 6)

The experiments suggest that label smoothing improves model performance in both self-supervised learning scenarios and when transferring learned features to new tasks. Specifically, a label smoothing value of 0.1 consistently enhanced model accuracy across different stages and learning paradigms.

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.

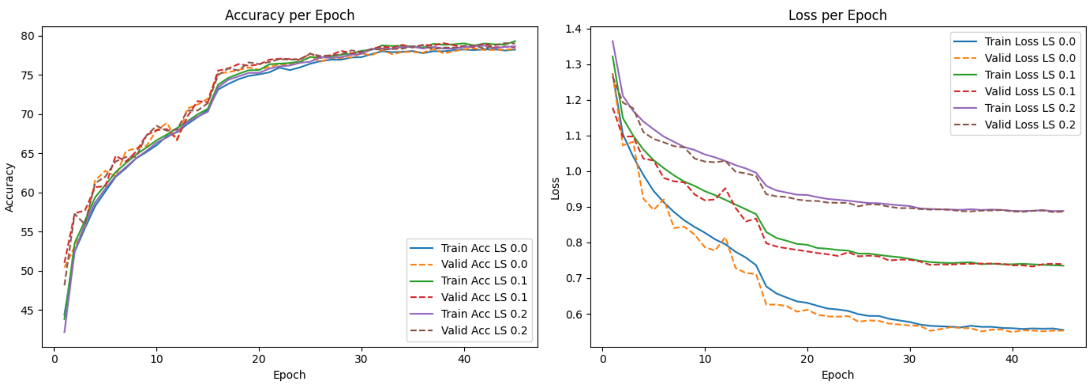
5. Conclusions

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:

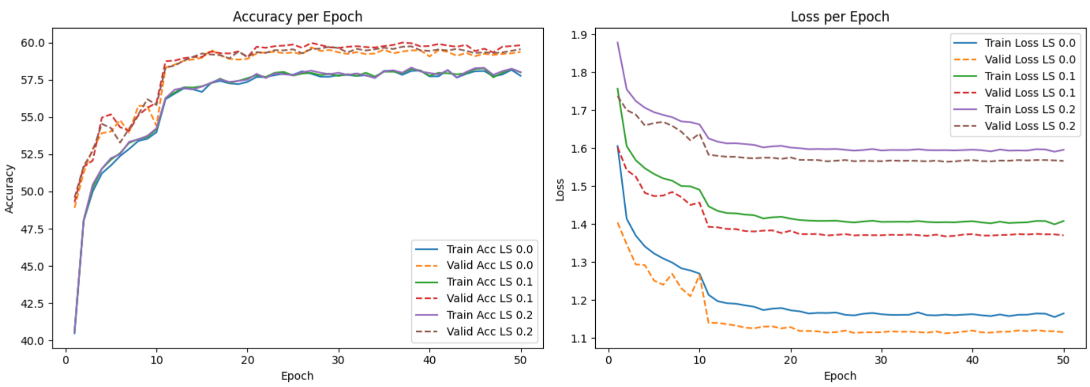
- The impact of label smoothing on different neural network architectures.
- The effects of label smoothing in larger, more complex datasets.

6. Appendices

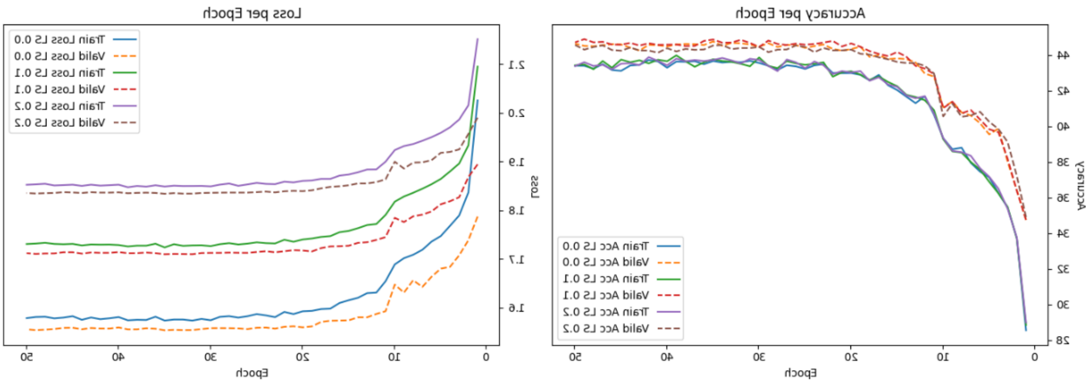
The appendices would include detailed tables and graphs of the experiment results, as well as any additional statistical analyses conducted.



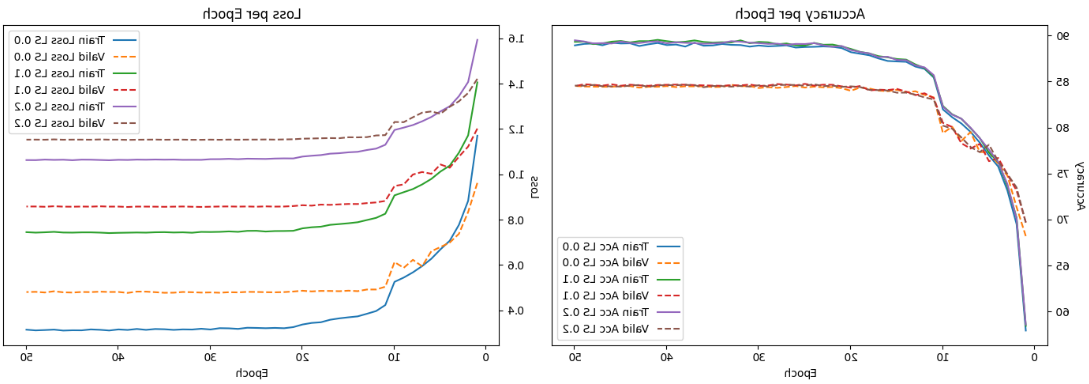
^ *Figure 1: Self-Supervised Learning*



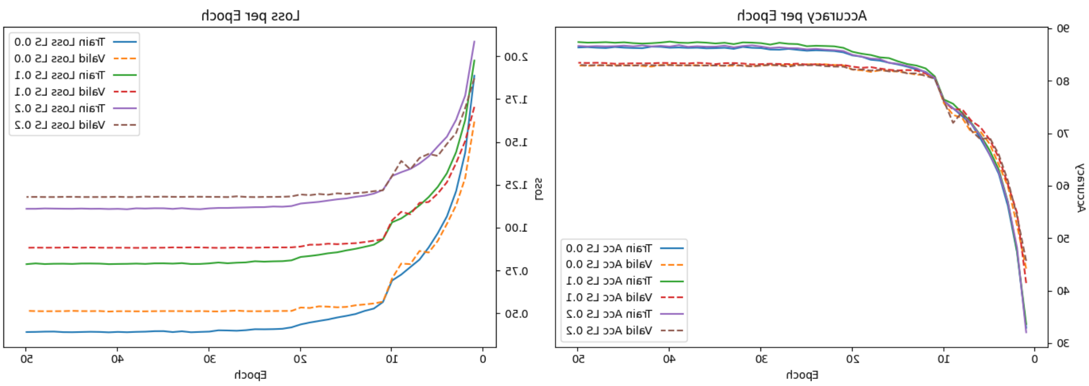
^ *Figure 2: Fine-tuning Pre-trained Model*



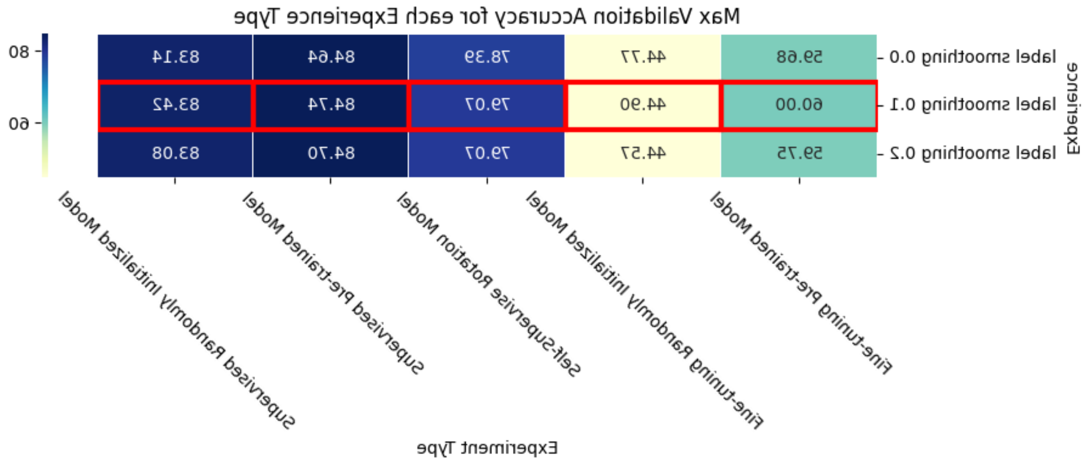
^ *Figure 3: Fine-tuning Randomly Initialized Model*



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



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



^ *Figure 6: Table for validation accuracy*

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)