Experiment Report: Impact of Label Smoothing on Self-Supervised Learning and Transfer Learning

Objective

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.

Experimental Setup

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. (Figure 1)
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. (Figure 2)
3. **Fine-tuning Randomly Initialized Model**: Initialize a new ResNet18 model and fine-tune it similarly to Stage 2. (Figure 3)
4. **Supervised Learning with Pre-trained Model**: Perform fully supervised training on the model obtained from Stage 1. (Figure 4)
5. **Supervised Learning with Randomly Initialized Model**: Train a randomly initialized ResNet18 model in a fully supervised manner. (Figure 5)

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.

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.
- **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.

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. (Figure 6)

Analysis and 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.

Conclusions and Recommendations

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:

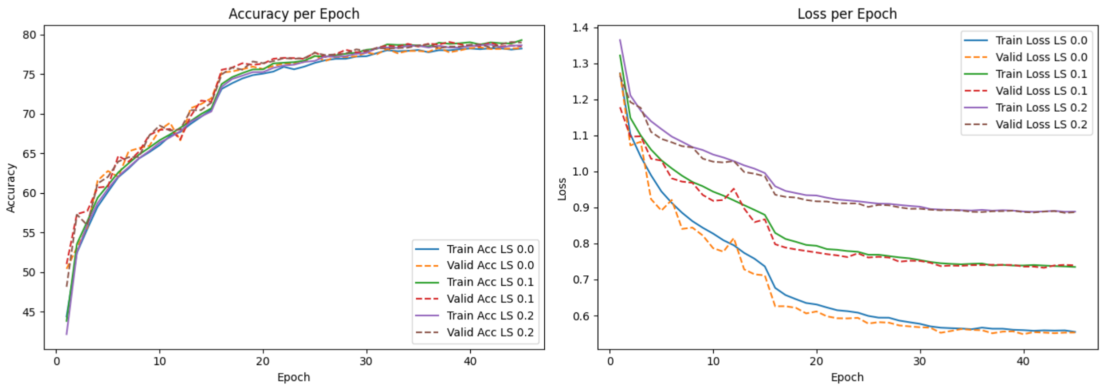
- The impact of label smoothing on different neural network architectures.
- The effects of label smoothing in larger, more complex datasets.

Author

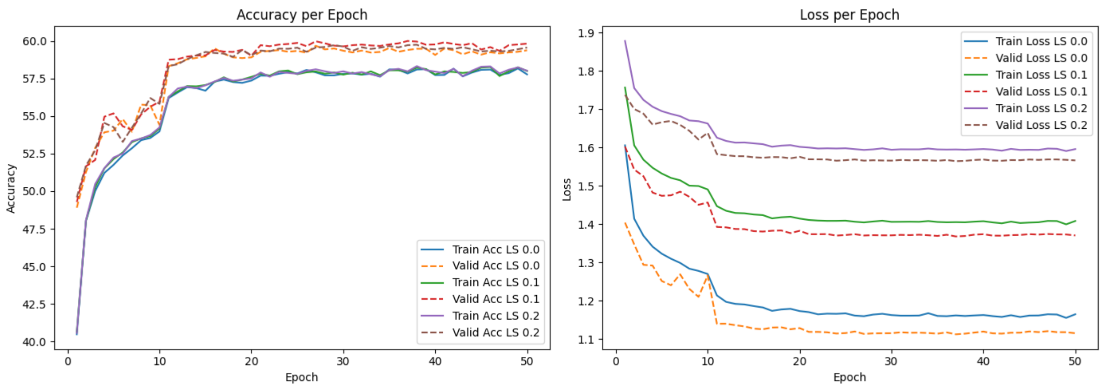
Author: Ren-Di Wu
Department: National Sun Yat-sen University Department of Information Management
Contact: whats2000mc@gmail.com

Appendices

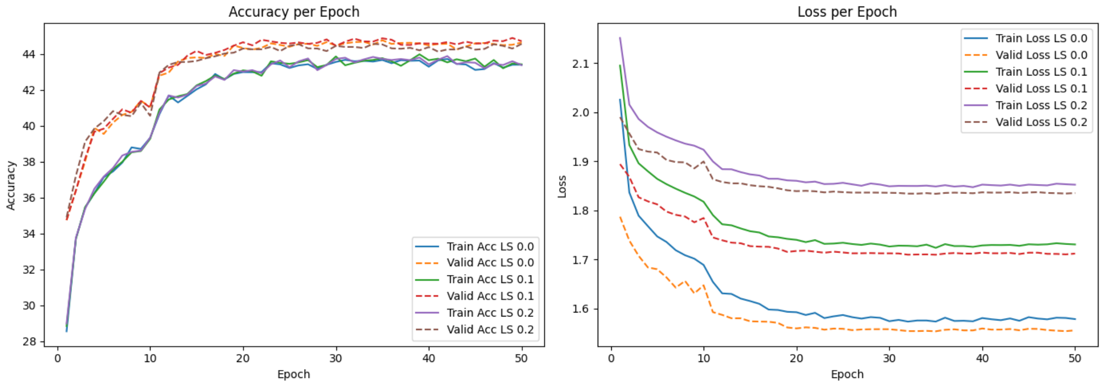
The appendices would include detailed tables and graphs of the experiment results, as well as any additional statistical analyses conducted.



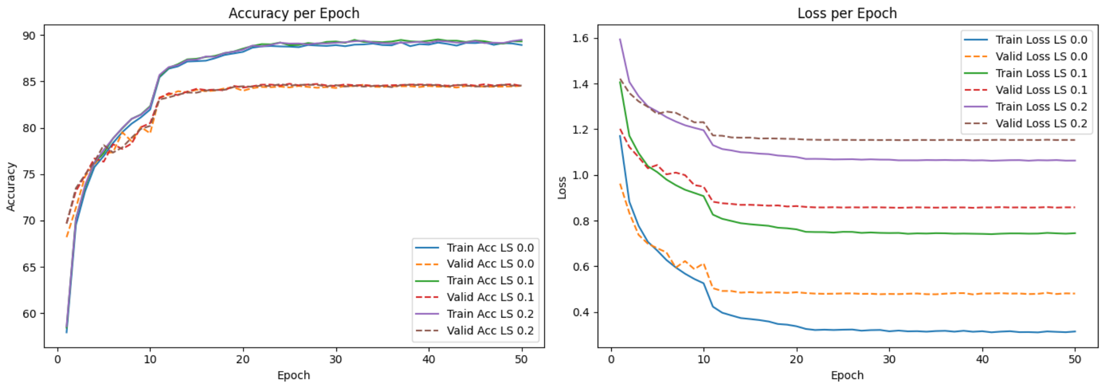
^ Figure 1: Self-Supervised Learning



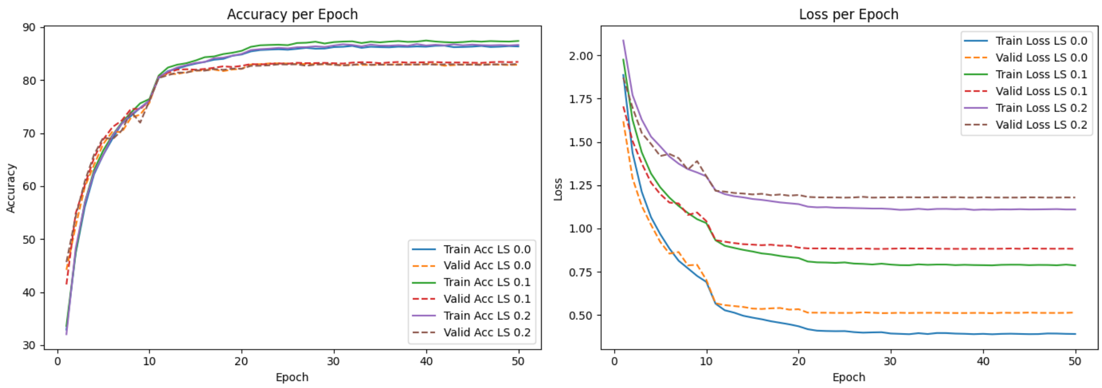
^ Figure 2: Fine-tuning Pre-trained Model



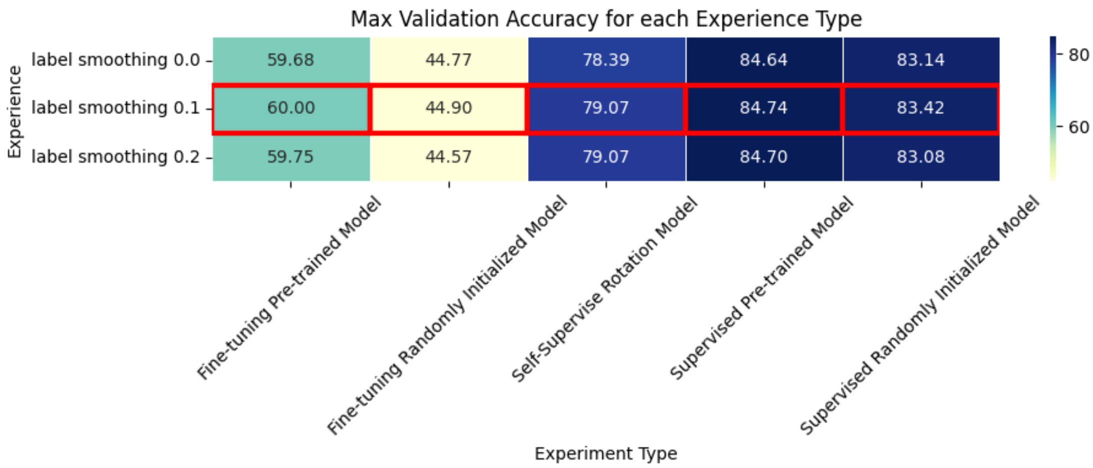
^ *Figure 3: Fine-tuning Randomly Initialized Model*



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



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



^ Figure 6: Table for validation accuracy