MATH 6380O Project 2:

HAN Ruijian, HU Wei, YE Rougang, ZHAO Yuqi {rhanab, whuae, rye, yzhaobk}@connect.ust.hk
Department of Mathematics, HKUST
April 2018

1. Summary

We reproduced experiments of the paper *Understanding deep learning requires rethinking generalization* and also explored the performance of transfer learning using trained model of our experiments. Results are as follows:
- Using ResNet-28, training error of CIFAR-10 can achieve ZERO in different levels of randomization .
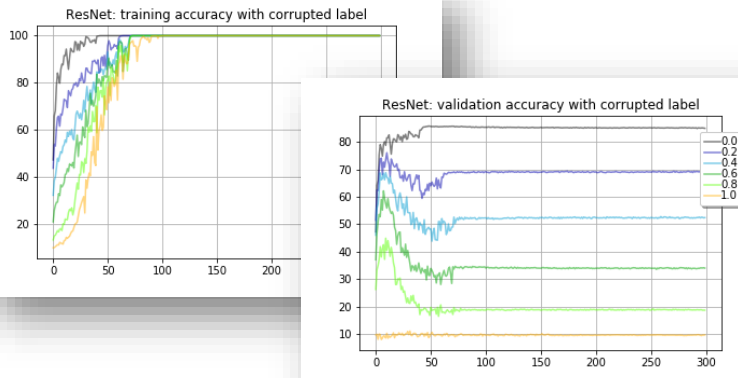- Gaussian noise helps improving the performance of transfer learning.

2. Model Architectures and Randomization Settings
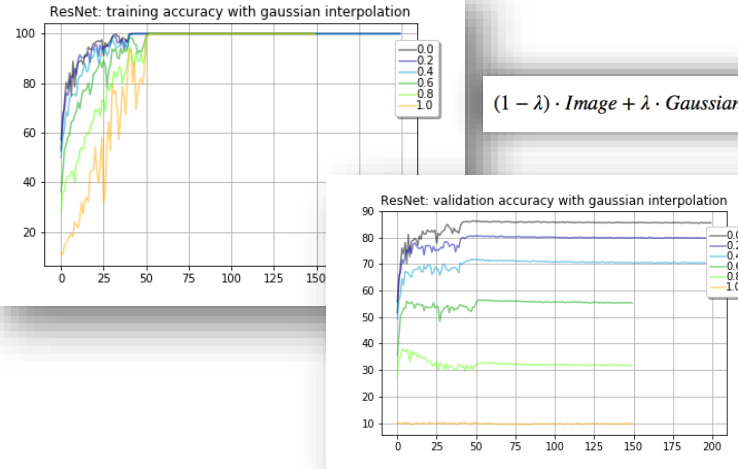
Model Architectures: Residual Network
Randomization Settings:
➤ Randomly corrupted labels;
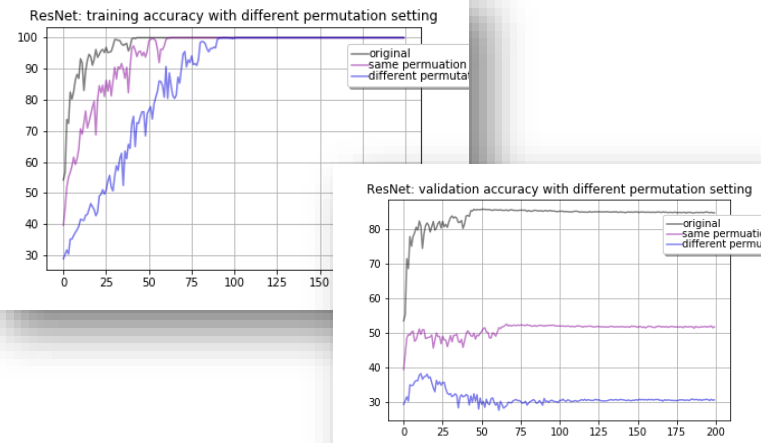➤ Randomly shuffled pixels;
➤ Gaussian interpolation.

3. Experiment Reproduction

Training/Validation Accuracy versus Epochs
when Labels are Corrupted with Different Levels

$$(1 - \lambda) \cdot Image + \lambda \cdot Gaussian$$

Training/Validation Accuracy versus Epochs
when Pixels are Interpolated with Different Levels

Training/Validation Accuracy versus Epochs
when Pixels are Shuffled with the Same/Different permutations

4. Transfer Learning

We transferred the trained ResNet model with different randomizations of CIFAR-10 dataset to Fashion MNIST dataset. We retained all layers and trained only on the last FC layer. The results showed that ResNet trained on Gaussian interpolated data can slightly improve the performance of transfer learning while ResNet trained with random labels cannot. It may illustrate that training the original data with some Gaussian noises can improve the quality of transfer learning. It's also observed that early stopping of the training on original dataset helps improving performance of transfer learning. This phenomenon will be further verified by extensive experiments.
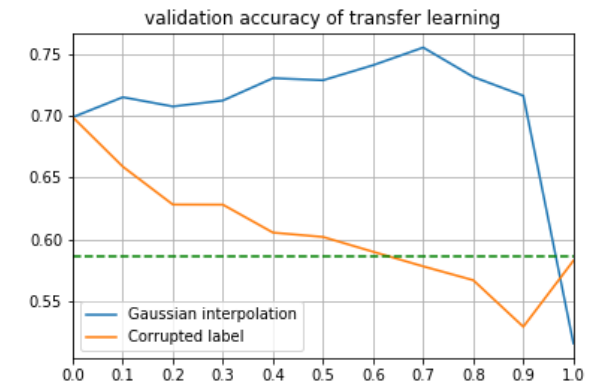
Validation Accuracy versus Levels
The green dash line represent the result of model initialized with random weights

5. Conclusions

Over-parametric ResNet has the ability to overfit any datasets.
Though the generalization error of ResNet on CIFAR-10 is relatively low, we believe overfitting also exists. Low generalization error may come from similarity or invariance (induced by architecture) between training and validation datasets. (demonstrated by transfer learning)
In transfer learning, we see that just like early stopping, interpolation with Gaussian noise prevent overfitting in some sense.