Wait, what'd ya say? - Noise Aware DNN Training

An Examination of Generalizability, Robustness, and Quantization

Ryleigh Byrne, Wyatt Focht

Introduction & Motivation

Current literature surrounding deep-neural-networks (DNNs) lacks exploration into the effect that post-training, model-weight perturbation has on the performance. We seek to analyze this effect by evaluating the impact that weight-perturbation has on model generalizability and quantization.

We also seek to employ various preventative methods to attempt to mitigate impact that weight-perturbation has on a model's performance. The preventative methods are

- **Naive noise-aware training**
- **Sharpness-Aware Minimization (SAM)**
- **SAM with multi-step weight perturbation training**

Our Contributions

- Model-resistance to varying strengths of weight perturbation attacks
- Evaluation of generalizability to new data on naive-noise, SAM, multi-step SAM models
- Evaluation of quantization on naive-noise, SAM, multi-step SAM models
- Evaluation of training performance of naive-noise, SAM, multi-step SAM models

Methodology

Model Creation

- Each model trained and validated on CIFAR-10 Dataset (10 classes; 50,000 training images; 10,000 testing images
- Each model configured in Resnet-20 architecture

- **resnetStandard**: model generated from standard training
- **resnetNaiveNoiseAware**: model generated from noise-aware training (Gaussian noise perturbations to model weights at every epoch)
- **resnetSAM**: model generated utilizing SAM during training
- **resnetMultiStepSAM**: model generated using multi-step SAM during training (increase SAM neighborhood size throughout training)

Weight Perturbation with Gaussian Noise

- Iterate over layers of each model
- Generate random Gaussian noise for layer
- Multiply generated noise by perturbation strength factor (= 0.1)
- Add result to weights of current layer

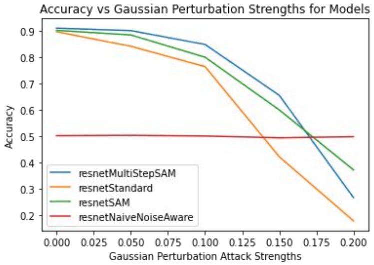
Experimental Evaluations

Overall Generalizability

Model Type	Testing Accuracy
Standard	0.8974
Naive-Noise	0.5031
SAM	0.9026
Multi-Step SAM	0.9109

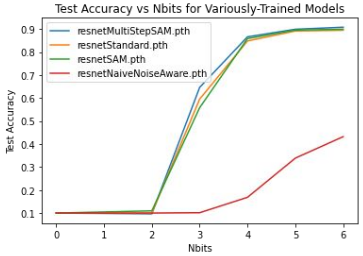
Vanilla Testing Accuracy on Unseen Testing Data

Resistance to Gaussian Weight Perturbation



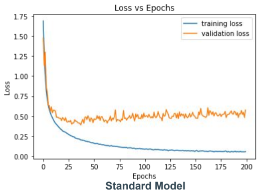
Testing Accuracy for Weight-Perturbed Models

Robustness to Quantization

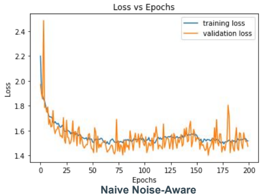


Testing Accuracy for Variously-Quantized Models

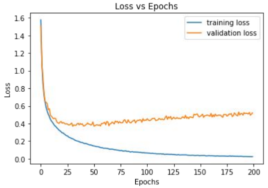
Further Discussion



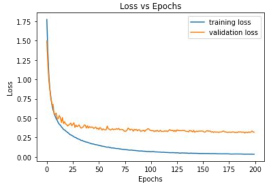
Standard Model



Naive Noise-Aware



Standard SAM



Multi-Step SAM

Model Type	Training Time (sec)	Ratio to Standard Time
Standard	4923	1
Naive-Noise	5374	1.09
SAM	6241	1.27
Multi-Step SAM	5862	1.19

Training Times for Various Models

Conclusions

Training

- High similarity in training & validation loss for Naive Noise, but low accuracy overall
- SAM & Multi-Step SAM have high training time (require two forward passes per epoch)

Generalizability

- Multi-Step SAM achieves highest level of generalizability, as expected[1]

Weight Perturbation

- Naive Noise has highest robustness to weight perturbation
- Tradeoff between accuracy and robustness to varying attack strengths

Quantization

- No models showed exceptional robustness to quantization
- Naive Noise performed exceptionally worse, due to low accuracy overall

1. Foret, Kleiner, Mobahi, & Neyshabur. (n.d.). SHARPNESS-AWARE MINIMIZATION FOR EFFICIENTLY IMPROVING GENERALIZATION. *T ICLR 2021*. https://arxiv.org/pdf/2010.01412.pdf