Outline

- Introduction
- Keras Tutorial Discussion
- Assignment One Discussion
- Pytorch Assignment Discussion
- Project Selection
About Me

School: University of Wisconsin - Madison

Major: Biomedical Engineering and Computer Science

Math Taken: Calc I,II,III, Differential Equations, Linear Algebra, Matrix Methods in Machine Learning, Discrete Mathematics

CS Taken: Programming 1-3, Python, Java, C++, Neural Networks, General AI Class

CV Experience: Neural Networks, ResNet Work, Image Classification

CS Experiences: Python, Keras, PyTorch, NumPy, Anaconda, Google Colab, CNN
**Figure 1:** Average training accuracy & loss / Average validation accuracy & loss using the MNIST Dataset
Figure 2: Confusion Matrix using the MNIST Dataset
Assignment One - Model Parameters

Batch Size
Number of Epochs
Optimizer
Learning Rate
Weight Initialization

Momentum, Decay, Bias
Loss Function
Kernel/Filter Size, Stride, Padding
Activation Functions
Data
Preprocessing/Augmentation
Assignment One - Lightweight

Figure 3: Lightweight Model Visualization (2.7 Million Parameters)
Assignment One - VGG16 Mimic

Figure 4: VGG16 Model Visualization (4.7 Million Parameters)
Assignment One - ResNet Mimic Without Residual

Figure 5: ResNet Mimic Without Residual Model Visualization (3.7 Million Parameters)
Assignment One - ResNet Mimic With Residual

Figure 6: ResNet Mimic With Residual Model Visualization (1.5 Million Parameters)
<table>
<thead>
<tr>
<th></th>
<th>General Output</th>
<th>Parameters</th>
<th>General Notes</th>
</tr>
</thead>
</table>
| 1 | Lightweight Model | 35-40% | 2.7 Mil | • Few convolutional blocks  
• Computations were fast  
• Batch normalization and dropout layers |
| 2 | VGG16 Mimic | 50% | 4.7 Mil | • Batch normalization and dropout layers  
• More convolutional blocks compared to Lightweight |
| 3 | ResNet Without Residual | 55-60% | 3.8 Mil | • 1x1, 3x3, 1x1 Convolution Block  
• Four convolutional Blocks  
• Batch normalization  
• Significantly more convolutional blocks compared to the previous two |
| 4 | ResNet With Residual | 25% | 1.5 Mil | • 1x1, 3x3, 1x1 Convolution Block  
• Four convolutional blocks  
• Passing the outputs of the previous block into the current block |
Model Size Graphs

**Figure 7:** Error Metrics vs. Model Size. The figure above shows the relationship between the model size and the error metrics calculated.

**Figure 8:** Time Taken for Five Epochs vs. Model Size. The figure above shows the relationship between the model size and the time taken to perform five epochs.
Model Size Graphs

**Figure 9: Error Metrics vs. Number of Iterations (epochs).** This figure shows the relationship between the error metrics and the number of epochs the model is subjected to.

**Figure 10: Error Metrics vs. Dataset Size.** This figure shows the relationship between the error metrics and the dataset size.
Pytorch Assignment - CIFAR10

**Figure 9: Epochs vs. Accuracy.** This figure shows the relationship between the testing and training accuracy and the number of epochs the model is subjected to.

**Figure 10: Epochs vs Loss.** This figure shows the relationship between the training and testing loss and the number of epochs the model is subjected to.
Pytorch Assignment - CIFAR100

Figure 9: Epochs vs. Accuracy. This figure shows the relationship between the testing and training accuracy and the number of epochs the model is subjected to.

Figure 10: Epochs vs Loss. This figure shows the relationship between the training and testing loss and the number of epochs the model is subjected to.
Project Selection - Multi-modal
step-by-step distillation

- Read the currently literature about the models and approach
- Download llava pretrain ckpts, dataset, and codebase
- Generate rationales and labels from pretrained Llava models as training signals for distillation
- Use Chain of Thought to generate rationales
- Prompt the model for labels
- Apply supervised learning with cross-entropy loss to distill rationales and labels from large pretrained model to small student model