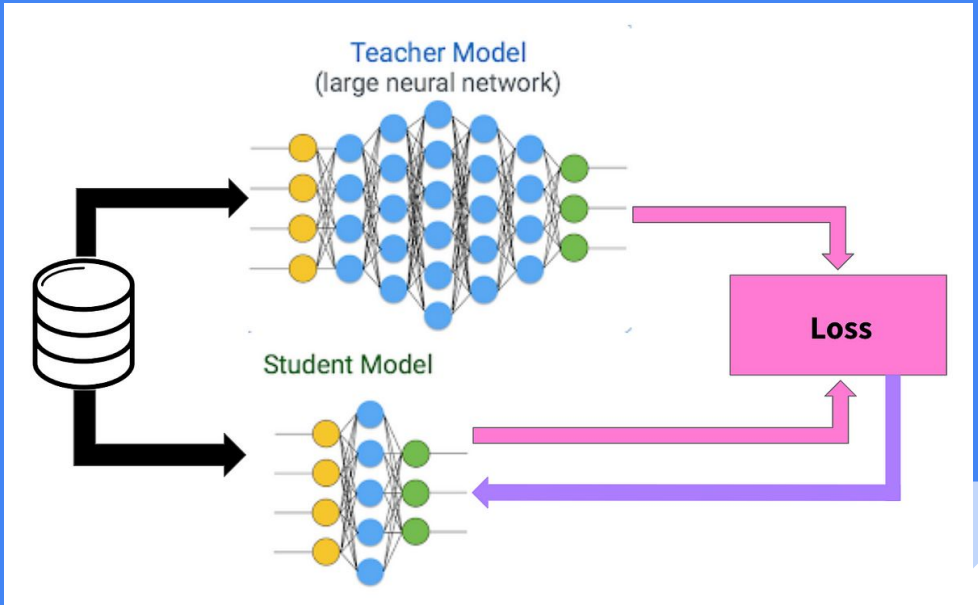
Knowledge Distillation for Radiology Dataset

Design Credits Project
(Jan-April 2023)

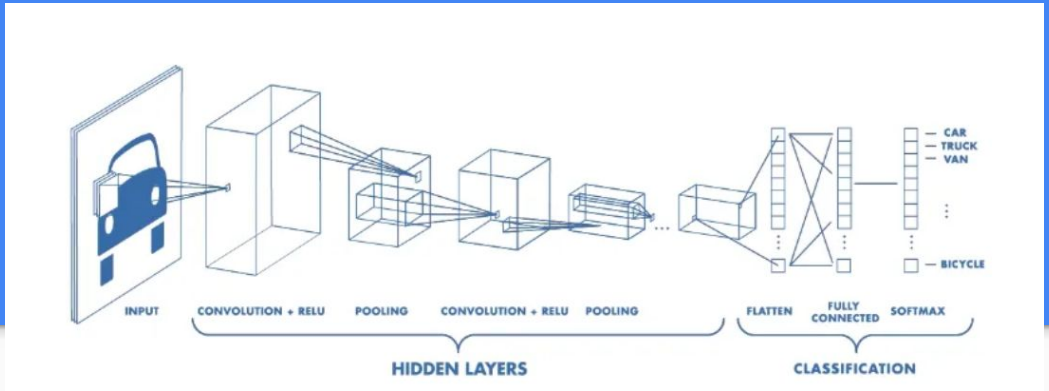
By: Vedasamhitha Challapalli (B20CS078)



Concepts Learnt in the project:

- CNNs and implementations in Pytorch
- ViT architecture
- Knowledge distillation
- Student-Teacher Model
- Implementation on CIFAR-10 dataset
- Implementation on Pneumonia Chest X-ray dataset
- Model Optimization

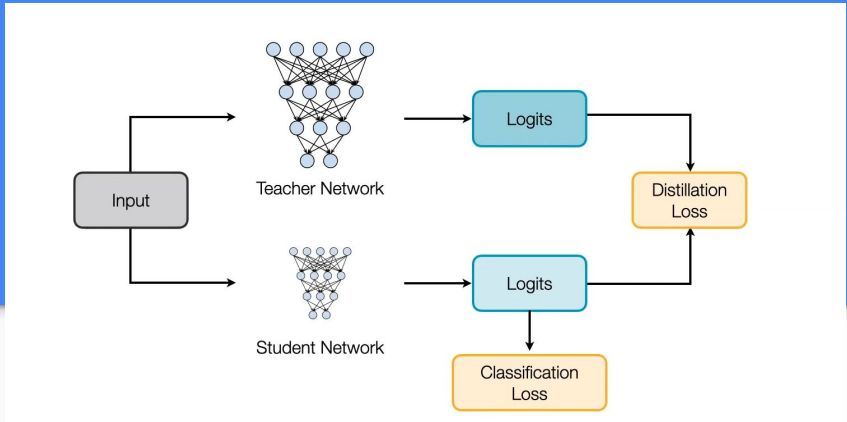
CNNs



- A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image.

- A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

Knowledge Distillation



- What is Knowledge distillation?
- What are the different ways of knowledge distillation?
- Benefits of Knowledge distillation
  - ➢ Model compression
  - ➢ Decreased computational power
  - ➢ Improved generalization
  - ➢ Better efficiency
- Usage

Student-Teacher model

- What is a Teacher model?
  - Densenet121 Model implemented
- What is a Student model?
  - MobilenetV2 Model implemented
- How does learning happen from teacher to student?
- KL Divergence loss

Implementation on CIFAR-10 dataset

- Model used:

  Teacher Network Layers = 10 conv layers and 1 pooling layer.
  Student Network = 3 conv layers + 1 pool layer
  Fully-Connected layers (in both) = 1 FC layer with 256 nodes.

- Results in the report
- Conclusion:

  Student model performs better than the teacher model in terms of accuracy and losses.

Implementation on Pneumonia Chest X-ray dataset

Conclusions:

- SGD optimizer with learning rate 0.0001 performed better on the dataset
- Weight decay of 0.01 was suitable
- Using only the pretrained models with pretrained = True doesn’t let the training happen
- Student performs almost equal or better than the teacher even though the models are overfitted individually if student trains from teacher properly
- Sometimes overfitting occurs on validation set if the validation set is too small.

Model Optimization

Concepts Learnt:

- Optimizers and hyperparameter tuning
- Weight decay
- Learning rate
- Dropout Layer
- Data Augmentation
- Temperature and Alpha

Future Work in Summer

- Learn more on model optimization techniques
- Learn about Federated Learning

THANK YOU!!