PROJECT SUMMARY
PROF. ANGSHUMAN PAUL
VEDASAMHITHA CHALLAPALLI (B20CS078)

Meeting 1:

Points Covered:

- 2 semesters of work has to be done.
- Deep knowledge of Convolutional Neural Networks is required.
- In depth knowledge of Deep learning is required
- Implementation of Deep Learning in PyTorch is required.
- Weekly updates must be provided regarding the work done

Work to do for next Thursday:

- Revise machine learning basics and go through the Convolution Neural Network playlist: https://www.youtube.com/playlist?list=PLblh5JKOoLUIxGDQs4LFFD--41Vzf-ME1

Meeting 2:

Points Covered:

- Searching is done by matching the search string with the key and the output is called the value
- Transformers are used in deep learning which adopt self attention technique, can be used in NLP and vision.
- An image, say 200 by 200 image is divided into non overlapping patches of size, say 20 by 20 and the corresponding 100 images are passed into a neural network and self attention is done in transformers.
- Positional information must be passed along with the patch because context of the image is important.

Work for next thursday:

- Attention and transformer networks: https://www.youtube.com/watch?v=OyFJWRnt_AY

- Vision transformers playlist: https://www.youtube.com/watch?v=DVoHvmww2lQ&list=PLpZBeKTZRGPMddKHcsJAOIghV8MwzwQV6

- Paper which deals with multi head self attention: (An image is worth 16 * 16 Words) https://arxiv.org/abs/2010.11929

Meeting 3:

Points covered:

- Doubts in vision transformer architecture asked

Work to be done:

- To code the vision transformer end to end

Meeting 4:

- Introduction to Knowledge distillation and the problem statement of the project (Knowledge distillation using vision transformers)
- Big CNN models need a lot of computational power, similar results can be achieved using models of less computational power.
- Student model learns the required necessary parameters from the teacher model and the student model is trained.

Work to be done:

- Go through any youtube resources regarding Knowledge distillation

Meeting 5:

- Doubts regarding knowledge distillation asked

Work to be done:

- Explore more on knowledge distillation

Meeting 6:

- Downloaded one of the github codes of knowledge distillation and tried to understand the code
- Asked doubts regarding the loss function
- Asked doubts regarding KL divergence

Work to be done:

- Change the dataset for the code given
- Use Fashion MNIST dataset or CIFAR dataset instead
- Divide the data set into 70% training data, 20% testing data, 10 % validation data
- Plot the losses for each of the 10 epochs for training and testing data

Meeting 7:

- Plotted the accuracy curves for student and teacher models for training and validation datasets
- Evaluated the performance on the testing dataset.

Work to be done:

- Use densenet model for the teacher model and mobilenet for the student model and plot the graphs and test the performance for Kaggle Pneumonia dataset (Chest X rays)

Meeting 8:

- Hyper parameter tuning to be done
- Model Optimization needed

Meeting 9:

- Use techniques like L2 regularization or early stopping to reduce overfitting

Meeting 10:

- More optimization needs to be done