AI61002: Deep Learning Foundations and Applications (Spring 2024)

Programming Assignment 2

In this assignment, you will be building a training pipeline along with the model for the dataset provided <u>here</u>. The dataset contains images containing paperclips. Your task is to predict the number of paperclips present in the image. [M.M : 25 Marks]

1. Implement a data loader which does the following -
   Takes the three inputs - Python list of train image ids, train.csv and path to the images folder. Based on the input image id/index, read the corresponding image from the image folder and the label from the .csv file.
   a) Convert the image to tensor and print the shape of the image. [1 Marks]
   b) Resize the image to 3x28x28. [1 Marks]
   c) If the number constituted by the last two digits of your roll number is x, apply random rotation augmentation between the range [-x, x] degrees. [1 Marks]
   d) Apply random horizontal flipping with the probability equal to x/100. [1 Marks]
   e) Normalise each channel of the image. **Hint:** Use the mean and standard deviation computed on the original training data for each channel of the image individually. [3 Marks]
   f) Convert the transformed image to a flattened 1D tensor and return it along with the image id and ground truth tensor i.e actual count of clips in the image. [1 Marks]

2. Implement a MLP model with following specifications -
   a) Since, the output of the image will be the number of clips, design the MLP in such a way that, after layer i, the tensor dimension reduces by 8^i until the final output tensor size is 1. For an instance, if the input length of the tensor is l then the 1st layer should reduce the length to $floor(\frac{l}{8})$, second layer reduces the length to $floor(\frac{l}{64})$ and so on. [4 Marks]
   b) Use the Relu activation in between the hidden layers. [1 Marks]
   c) Choose the appropriate activation function for the output layer based on the task at hand. [2 Marks]

3. Write the code for the forward pass in following way -
   a) Create the train data loader and validation data loader objects with batch size of 8. [1 Marks]
   b) Initialise the model in a way to utilise GPU for the forward pass. [1 Marks]
   c) Do the forward pass for one batch and visualise the first 5 input original images along with the predicted output and ground truths. [1 Marks]
   d) Print the summary of the model with layer wise outputs and number of parameters. [1 Marks]

4. Train the model in following configuration -
   a) Prepare the training, validation and test sets by randomly sampling 2500 images from original data and partitioning it into three disjoint subsets as follows
   Training set: 2000 images
   Validation set: 250 images
   Test set: 250 images   [2 Marks]

b) Use Adam optimiser and initialise the learning rate to 0.001 [0.5 Marks]

d) Train for 20 epochs using Mean Square Error (MSE) loss. [0.5 Marks]

e) Plot the training and validation loss vs epoch graph and report epoch with lowest MSE on the validation set. Using the model checkpoint saved at this epoch, plot the ground truth and predicted number of clips on the test set using a scatter plot. and report the average MSE on the test set [3 Marks]

Note: If your roll number is **19**AE310**03**, set the seed value to 1903 everywhere i.e. first two and last two digits combined. For training, we recommend using Google Colab for GPU access.