Mini-Project (15 Marks) :

Image Captioning:

Design a CNN-LSTM system (preferably in pytroch) that can perform image captioning under following conditions:

- The total model size should not exceed **100 MB**. Considering that each parameter weight is a float value (4 bytes), your CNN-LSTM model should have approximately **25 Million parameters** or less.
- Use the Flickr8K data for training and testing the model. The data is available at

  https://drive.google.com/drive/folders/1RQ5qHm0aVFqWDG9VBiSnXlNPl5T15Wf_?usp=sharing

  (Check Readme.txt for details of txt files)
- Use any pre-trained CNN weights that suits the memory requirement. LSTM weights should be trained from scratch.
- Under the memory constraints, try to achieve **maximum BLEU score** on the **test data**.
- Produce subjective results on the images shared in “subjective_img.zip”
- Share the training code, testing code and a technical report as a zip file *Groupcode*.zip (e.g. AVJS.zip)
- Include all team member details and a unique group code in the start of your technical report.
- Technical Report should contain architecture details of model with memory details , training methodology, training level experiment details, & test data evaluation details with BLEU score
- Technical Report should also contain subjective results achieved on the shared images

Note These Points:

1. To save on training time, run all images in training data to obtain CNN image embeddings for once and save it to drive. Use file read to store the image embeddings to local variable before the start of training. Same applies for validation and test data.
2. Remember that only constraint is on # of parameters/Memory. You are free to design your system respecting only the memory constraint to achieve the best BLEU score.