README

**Running the Pre-trained model:**

The saved models are named model\_good\_1.h5 for challenge dataset and model-ep004-loss3.516-val\_loss3.849.h5 for flicker8K dataset.

**For VizWiz Challenge Dataset:**

In Train\_Model.ipynb file, run the last cell of the notebook after setting the path variable in the code to be the path of the image for which you need to generate the caption.

You must have installed all the libraries required by the code in that cell.

The images present in the Sample\_test Folder are some sample images from the validation dataset that can be tried on the code by providing correct path variable.

**For Flickr8K Dataset:**

In Train\_Model\_Flicker.ipynb file, run the last cell of the notebook after setting the path variable in the code to be the path of the image for which you need to generate the caption.

You must have installed all the libraries required by the code in that cell.

The images example.jpg, example1.jpg, example2.jpg, example3.jpg, example4.jpg are some sample images from Flicker8K dataset that were used in ppt so you can use them for result verification by providing the correct path variable in the last cell code.

**Training Approach:**

To train on the images for the final captioning, first the data is preprocessed using the **VGG16 mode**l for feature extraction on **ImageNet** dataset.

The **Bottleneck Features** are extracted by removing the last layer of the model and processing all images on it. The results are stored in **features.pkl** for challenge dataset and **features\_flickr.pkl** for flickr dataset.

After compiling the model, we use an **‘adam’** optimizer for learning with the learning rate set as default i.e. 0.001 in the beginning.

We train with a batch size of 6 and learning rate = 0.001 for the first 2 epochs.

For the subsequent training, **for every 2 epochs**, keep increasing the **batch size twofold** and decreasing the **learning rate to half**.

Decreasing the learning rate is important as the model is towards convergence in the later stages of training and a high learning rate can lead to **overfitting**.

Increasing the batch size is also important as it makes the **gradient updates more powerful**.