**Environment Setup for CALTEXT Code**

To run the CALTEXT code, the following dependencies and versions are required:

* **Python**: Version 3.8 or higher (python –version)
* **TensorFlow**: Version 2.17.0 (pip install tensorflow==2.17.0)
* **Matplotlib**: Version 3.4.0 or higher (pip install matplotlib==3.4.0)
* **Scikit-Image**: Version 0.18.0 or higher (pip install scikit-image==0.18.0)
* **NumPy**: Version 1.20.0 or higher (pip install numpy==1.20.0)

**Folder Structure for CALTEXT Code**

The folder structure required to run the CALTEXT code is as follows:

**Main Folder**

This is the root folder containing all necessary subfolders.

**Subfolders**

1. **PUCIT\_OHUL**

- This folder contains the ground truth labels and vocabulary files.

- Files in this folder:

1. train\_labels\_v3.xlsx
2. test\_labels\_v3.xlsx
3. vocabulary.txt

2. **data**

- This folder contains all the pickle files required for training, validation, and testing.

- Files in this folder:

1. train\_images.pkl
2. test\_images.pkl
3. valid\_images.pkl
4. train\_labels.pkl
5. test\_labels.pkl
6. valid\_labels.pkl
7. vocabulary.pkl

3. **checkpoints**

- This folder is used to save model checkpoints during training.

4. **model**

- This folder is used to save best validation model.

Now just run the file cell by cell.