How to use text description?

How to do ensemble?

Steps:

1. Dataloader – augmentation (normalization, random crop, random rotation, random horizontal flips) – dataset is very nice so doing too much cropping and flipping is perhaps harmful
2. CNN – architecture (start with resnet pretrained parameters on image net)
3. Combine text data with image data
4. Fine tune (lr, regularization constant, research other things)
5. Ensemble

TODO:

1. Combine text data with image data
2. Use language process model to analyze description – first remove uncommon words, then use attention for classification.
3. Check noisyTextDescription to see if it contains any labels (should be very high accuracy).
4. For common mistakes (e.x. sandal vs shoe), train another CNN just to distinguish them whenever the predicted value is sandal or shoe.

Vayyar:

1. Run model until a good ckpt is reached
2. Apply t-SNE on linear layers
3. Convert model to output 5 classes
4. Does converting from complex number (amplitude and phase) to real number lose information?