1/24 meeting

* Ideas:
  + Two types of losses for representation learning: distance based (min distance) and mutual information (max MI). What features are being found by models trained using these two loss functions?
  + Use cGAN to explore feature space of a classifier. Only modulate the pathological features.
  + Use radiological images to generate text reports (like image captioning). Then explore the feature space of this model.
  + In image-text models, find the sentence-image feature pairings that are the most *relevant* (what measure could you use here?)
  + If you have a way of exploring latent space, then compare the latent space of two encoders: one that is trained only on images and another that is jointly trained on images and text

TODO:

* Look at repo: <https://github.com/RayRuizhiLiao/joint_chestxray>
* Train the model in unsupervised fashion (only change save dir, you do not need to change data dir)

1/28

* Use image-text pairings (correct pairings = positive and random pairings = negative) and contrastive learning for localized pre-training of a model for downstream task. You could have one large pre-training dataset with various categories. You can create positive and negative pairs based on your downstream task.
  + Create a method that would show why joint image-text pairing is a better pre-training protocol than image only pre-training.
  + More localized representations could mean we could do novelty detection
* Text to image generation (no pleural effusion and pleural effusion in the left lower lobe)
* Create an API to for medical cxr classifier explanation
* Joint modelling paper end tsne graphs
  + For more severe disease, the joint representations were as good as image only representations
* Unsupervised learning for joint image-text modelling (can create positive and negative pairs). Then gets into hard negative mining
* Accuracies of joint models are just slightly higher than individual models in downstream classification tasks. Why are the joint representations better? How are they more useful? How can we know?

2/2

* Explore the embedding spaces (styleGAN, clustering). Find out the important “factors” in the space. Need to give some structure to the space before. See how the metadata (age, sex, etc) relates to these factors.
  + Reach out to Philip isola

TODO:

* Take random 224x224 image patches from each large image (think about getting the correct bounding boxes as well when you sample patches)