Data Preprocessing:

* Will only use Training Data: 234, 842 images
  + 164,382 Training (70%)
  + 23,483 Validation (10%)
  + 46,977 Testing (20%)
* Hyper Parameters
  + Embedding size - 1024
  + learning\_rate = .001
  + number of layers = 1
  + hidden size = 512
  + weight decay -