1. Dead ReLU is when ReLU always outputs zero. This is a problem because then the neuron will not learn anything and since the gradient of ReLU when it = 0 is 0, it also passes the gradient of 0 backwards during backpropagation and the weights are not updated. Leaky ReLU addresses this by adding a small negative slope when the input is < 0. Therefore, ReLU cannot output 0 (unless the input is exactly equal to 0), and it will pass very small gradient during backpropagation allowing at least some updating to take place.
2. The role of the parameters in batch normalization is to “shift” the mean to a different value than 0, and “scale” the variance to a different value than 1. Since regular normalization requires 0 mean and a variance of 1, this is not necessarily optimal for the hidden layers within. Therefore, the net can learn the most optimal values of gamma and beta to maximize the accuracy.