

Dropout: During training, per minibatch, in some layer, delete some fraction of the activations (i.e. make them 0). This helps with generalization.

\*\*\* Average activations?

\*\*\* Dropout on conv layers or fully connected / linear layers?

\*\*\* Linear and fully connected are the same thing?