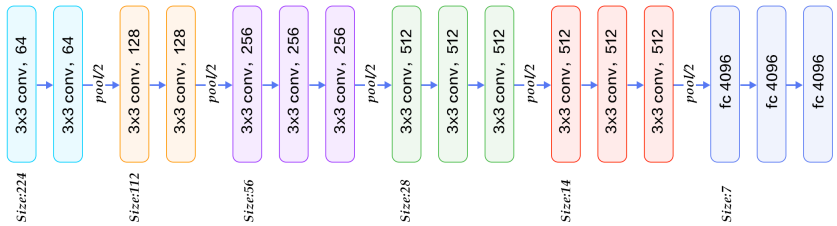
101 Example (LeNet for Dogs vs Cats)
Train LeNet on the Dogs vs Cats dataset. Is LeNet over or underfitting the data?

---

102 Example (VGG16 Network)
The VGG16 convolutional network has a particularly simple structure:



Use VGG16 (with ImageNet weight) and transfer learning on the Dogs vs Cats dataset.

---

103 Definition (Dropout)
Dropout is an efficient regularization method for neural networks. Weights in a layer are randomly set to zero with some fixed probability. The probability is an additional hyper-parameter of the network. Dropout is normally applied after activation layers and is commonly used in the final layers of a neural network, which typically are fully connected layers. For additional details, see Srivastava et al.[12]

---

104 Example (Dropout on Dogs vs Cats)
Apply dropout to the fully-connected (dense) layers for the transfer learning of Dogs vs Cats classification. What effect does the dropout probability have on the rate of learning? Note that it is generally believed that slow learners are better at generalizing than fast learners.

---

105 Example (Data Augmentation on Dogs vs Cats)
Apply data augmentation to the Dogs vs Cats data set. Which transformations make sense for the Dogs vs Cats data set?

---

106 Definition (Batch Normalization)
We have seen that standardizing (or normalizing in some fashion) the input to a neural network can dramatically accelerate training. The basic idea of batch normalization is to normalize the outputs of individual layers in a neural network. (Normalization is done before activation layers.) The word batch in batch normalization refers to the fact that minibatches are normalized. According to Ioffe and Szegedy[13] batch normalization:

- speeds up training.
- allows higher learning rates to be used.
- regularizes the network.

---

[12] Srivastava, Hinton, Krizhevsky, Sutskever and Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, 2014.

[13] Ioffe and Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015.