**Method**

Transfer Learning:

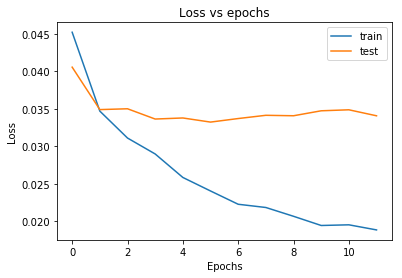
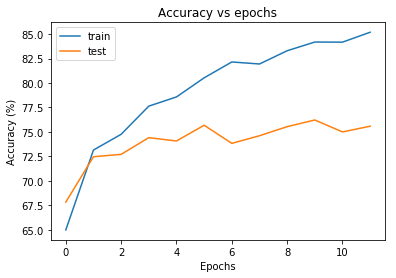
For this part, we used transfer learning with a VGG16 (with batch normalization) network with weights pretrained on ImageNet data (as vgg16\_bn in pytorch). The dataset used was Caltech256, which has 256 classes. We froze the weights of all layers except for the last fully connected layer with softmax. The last layer, originally with 1000 output units, was changed to 256 units and only this layer was trained. We used 32 images per class for training and 8 images per class for testing.

To further explore the features learned by VGG16 network, we used an example image of a dog and passed it through the trained network. The activation maps of the first and last convolutional layers were visualized and analyzed. The filters from the first convolutional layer were also plotted.

**Results**

The plots of accuracy and loss against number of epochs are shown below. After 2 epochs, the loss of training set continues to decrease while the test set maintains a consistent value with small oscillations. The train and test plots diverge after then, indicating overfitting on the training set. Eventually the final accuracy of train set is 86% and accuracy of test set is 75%.

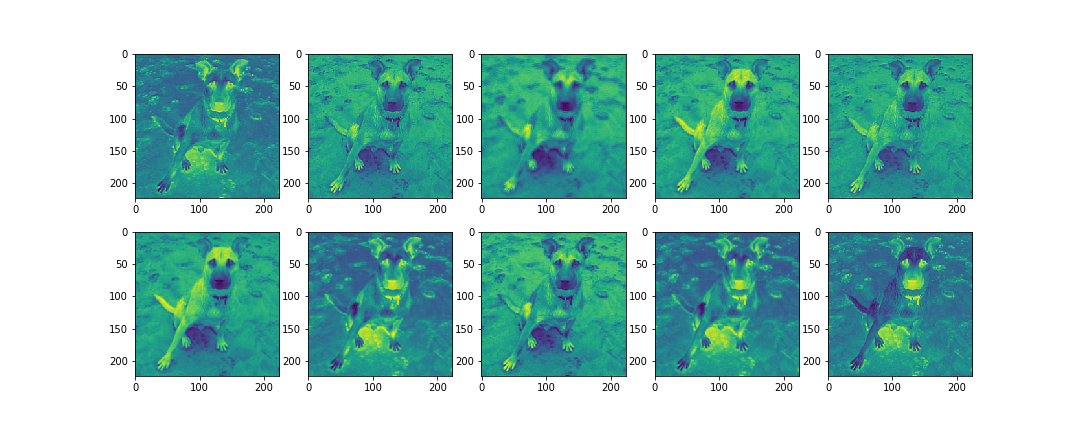
This is a good result considering only 32 training images were used per class. The good performance indicates that the weights trained from ImageNet are helpful in classifying a new dataset that has similar objects as ImageNet.



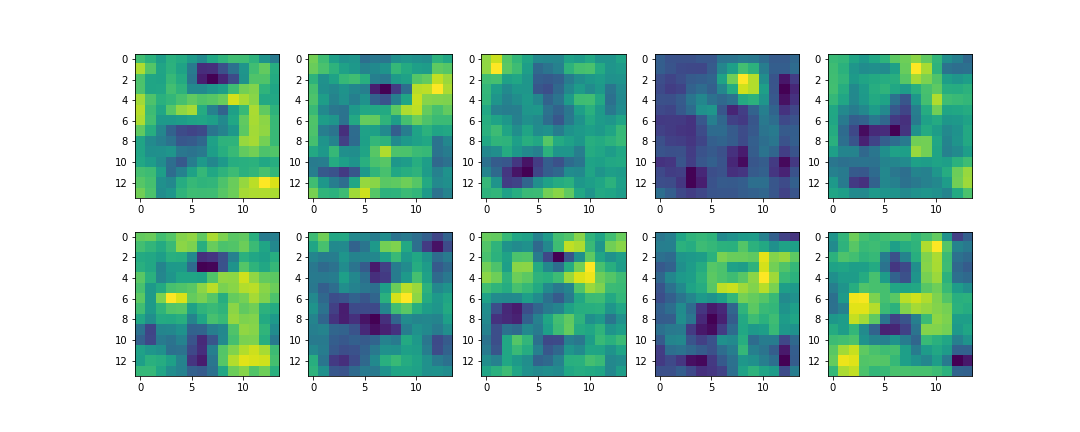
A dog image is passed through the trained network. The selected activation maps of the first and last convolutional layers (10 each) are shown below. The activation maps of the first convolutional layer look very similar to the original image, but with different regions of activations (the bright pixels). This is because each activation map corresponds to a different 2-D filter of basic structures (edges, lines and so on). On the other hand, the activation maps of the last convolutional layer do not look like a dog at all. This result is exactly as expected. The filters of the last layer should contain high level structures of objects (head, body, and so on), and the activation maps simply indicate whether these structures are found in the image.



Original image



Activation maps of first conv layer



Activation maps of the last conv layer (up)

Some filters of the first convolutional layer are shown below. As expected, they represent lines and edges.

Filters of first conv layer