Visualizing and Understanding Convolutional Networks

SMAI Major Project - November 2021

2021900003 - Kambalapally Vamsheedar Reddy

2021701024 - Kandru Siri Venkata Pavan Kumar

210364211 - Harsha Nookala

2020801004 - Prakash Tekchandani

TA: Yashas Samaga

Introduction

Convolutional Neural networks are the staple backbones of many states of the art model architectures used to solve problems with image modality. Deeper CNNs usually learn more complex feature representations that are compositional and capture complex invariances. Unfortunately, this depth of the network precludes us from understanding the model's internal working and its behavior.

The main goal of this study is to implement visualization strategy proposed by Zeiler et al., 2013 in their paper Visualizing and Understanding Convolutional Networks. We also seek to perform various feature analysis, correspondence analysis and occlusion sensitivity analysis at different layer depth. Finally, we aim to compare Reset50 and AlexNet using these methods.

Proposal

1. Implement Alexnet and ResNet50

AlexNet and ResNet50 are CNN based Neural Network architectures with different number of layers and Convolution Kernel sizes. They are primarily used for image classification. Figure 1 briefly summarizes the architectures of these networks.

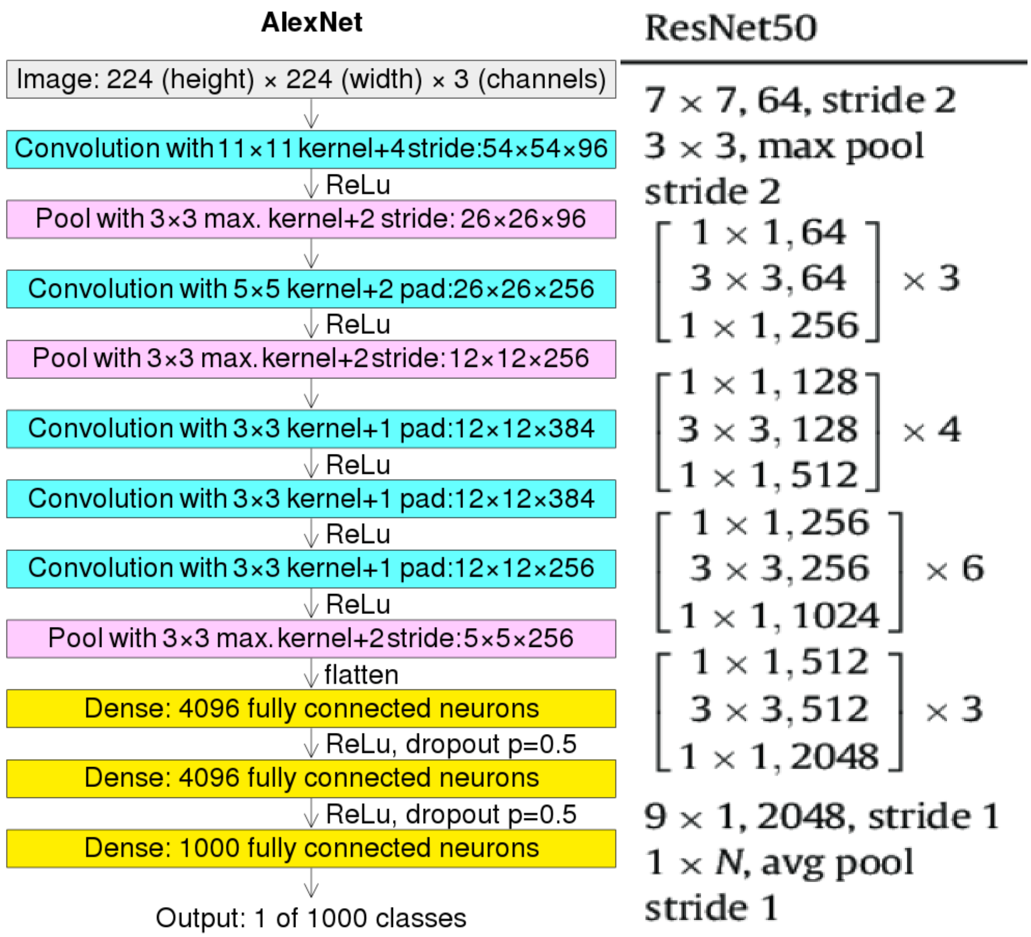

Figure 1. Architectures of AlexNet and ResNet

The topmost layer in both networks is a SoftMax layer with as many neurons as the number of classes. AlexNet contains 5 convolution layers,3 Pooling Layers and 3 fully connected layers while ResNet 50 contains 48 convolution layers ,2 pooling layers and 1 fully connected layer. ResNet uses residual connections which makes it possible to train networks with such large depths. Figure 2 shows an example of Residual connections.

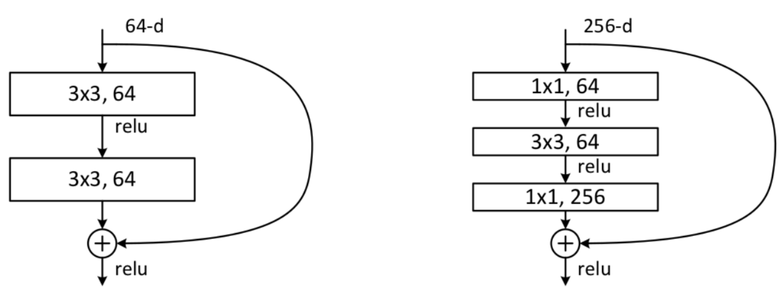

Figure 2. Residual Connections

2. Implement Transpose Convolutional Network

Deconvolutional networks are first proposed to perform unsupervised learning thereby capturing feature representations that can be used for image representation and synthesis. However, the application of transpose convolution for visualization does not warrant for learning of the parameters it. Instead, we use the parameters of the convnet but in reverse(transposed) to map the features back into easily interpretable input pixel space. This is done by applying similar transformations to convnet but in reverse direction. The following are the steps to perform deconvolution.

- Perform a forward pass of the network with a sample image and save all the intermediate feature maps.
- Select the feature maps to visualize at any convolution layer. This can be done by taking feature maps with highest activations.
- Set all other feature maps to zero and pass the set of feature maps to the following steps.
- Unpool the feature maps and generate a larger size image. This operation is not trivial because of the non-invertible nature of pooling. This can be solved by creating switches that store additional information during forward pass through the network and help in unpooling.
- ReLU is applied on the image
- Convolution filters are flipped vertically and horizontally and used to map the feature map into pixel space.

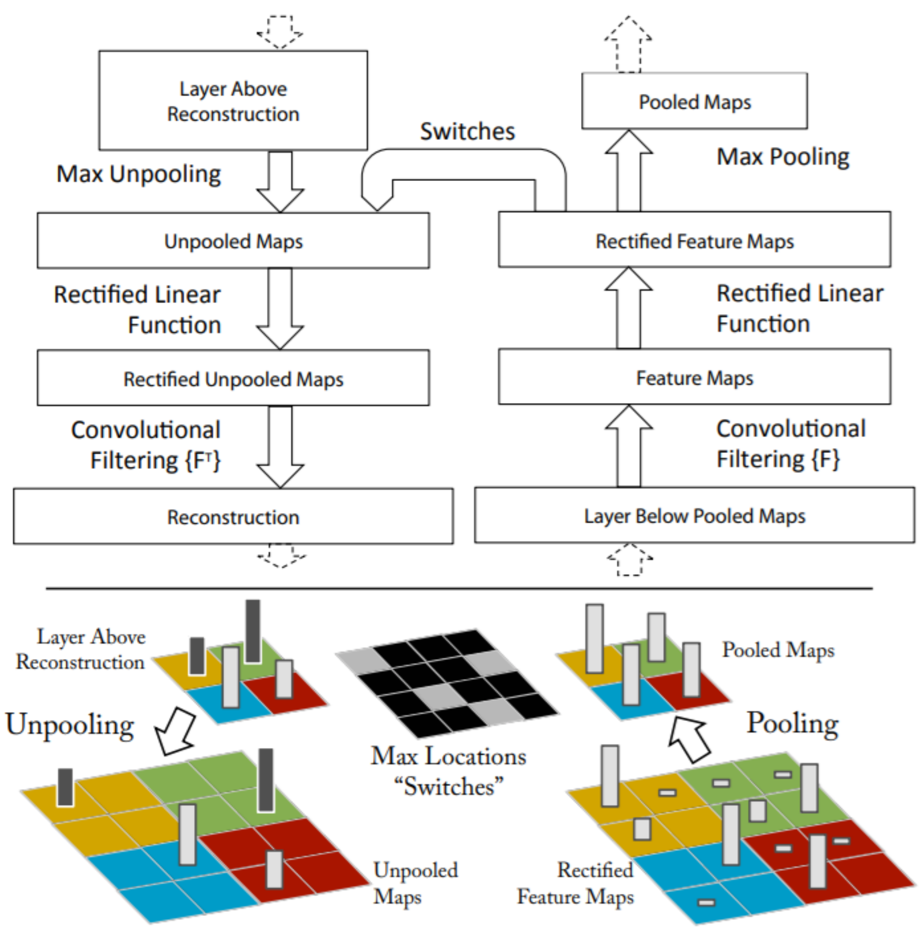

Figure 3. Convolution and Deconvolution Block

3. Visualize top activations using Transposed Convolutions

Above mentioned deconvolution network is used to visualize top 9 feature maps each layer for different images. Qualitative Analysis is performed on these Visualizations.

4. Perform Invariance Analysis

Translational, Rotational and Scale Variance is introduced into image samples and corresponding visualizations are compared to check for invariances.

5. Feature Evolution Analysis

Strongest activations of randomly chosen set of feature maps are visualized at different epochs of training. These visualizations help us understand the evolution of feature maps as the training progresses.

6. Perform Occlusion Sensitivity Analysis

Some parts of the input images are occluded, and visualizations are analyzed to observe the effect of occlusion.

7. Perform Correspondence Analysis

Similar objects in different images are occluded and visualizations of features are compared. This may give us insight into the importance or lack thereof of certain objects in the image.

8. Compare Alexnet and LeNet architectures

Prior Analysis tasks are performed on both Alexnet and LeNet. The Feature maps are compared to see the effects of kernel size, stride etc., on the feature maps.

9. Evaluate Model performance on other Datasets

Best Model found through above process is used for evaluation of other datasets like Caltech-101, Caltech-256, PASCAL VOC etc. Only the last SoftMax layer is trained on the new datasets while all the lower layers are frozen. This helps us understand the generalizability of learned feature representations.

Timeline

November 13th – CNN and Transposed CNN are to be implemented; Visualizations of different layers is done.

First Deliverable: Code and Visualizations on November 17th.

November 20th – Invariance, Evolution, Occlusion and Correspondence Analysis of features

November 28th – Comparison of AlexNet and ResNet50. Generalizability to other data sets.

November 30th – Preparation of Report and presentation

Final Deliverable: Submission of code and report on December 1st.

Work Distribution

Pavan – Implementation of CNNs and generating visualizations for analysis

Prakash – Analysis of Visualizations for different properties.

Vamsheedar – Generating visualizations for analysis and Evaluating CNNS on other datasets

Harsha – Analysis of Visualizations for different properties and comparing different CNNs.

References

Visualizing and Understanding Convolutional Networks
https://arxiv.org/pdf/1311.2901.pdf

Deconvolutional Networks

https://www.matthewzeiler.com/mattzeiler/deconvolutionalnetworks.pdf

Gradient Based Learning Applied to Document Recognition
http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf

ImageNet Classification with Deep Convolutional Neural Networks

https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html