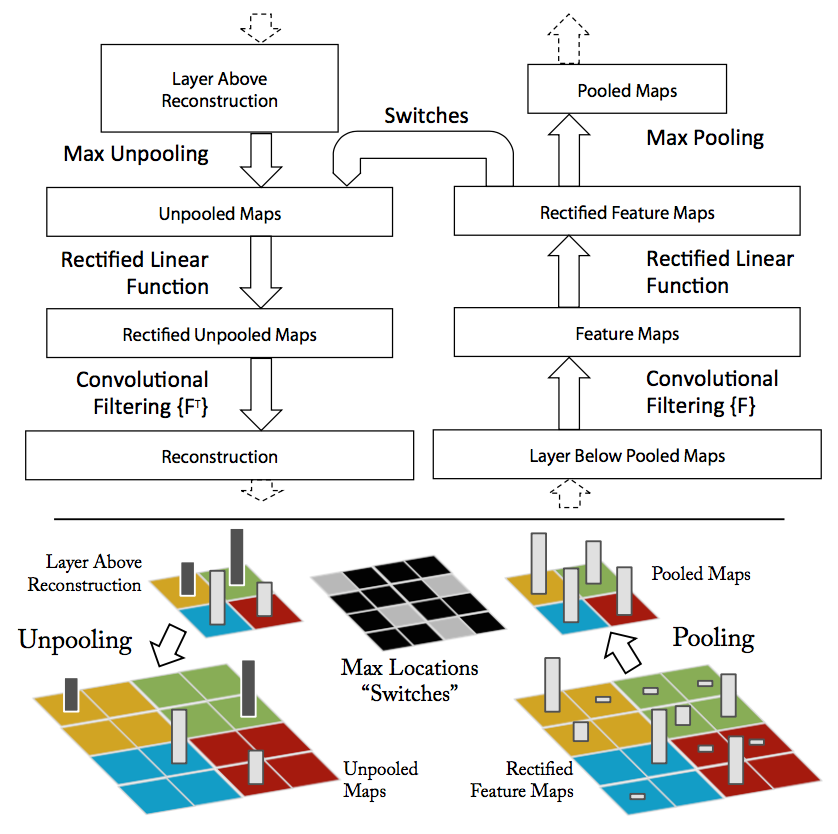
Visualizing and Understanding Convolutional Networks

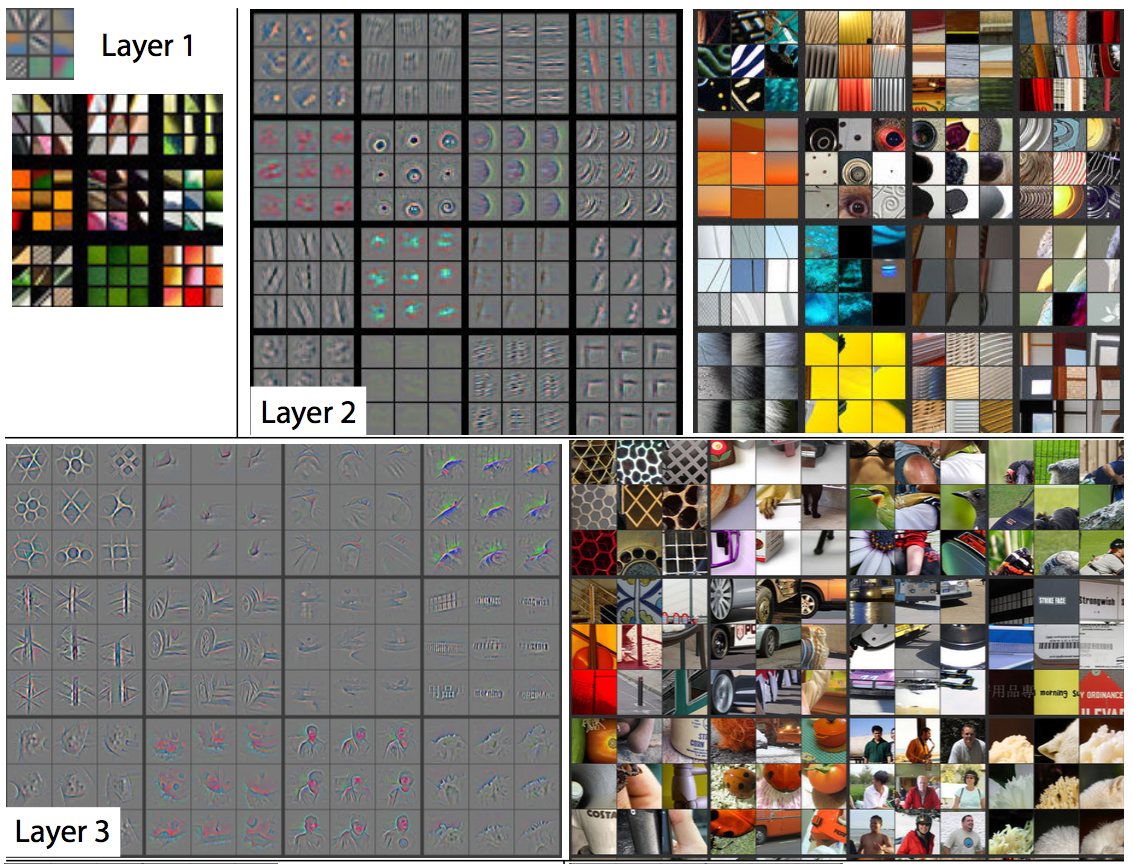
Jo MinKi

1. Introduction   
   CNN performs so well for image classification but the reason why it work does not explained clearly yet. And this research also couldn’t figure out the exact reason about that. But in this paper, they described some understanding about the CNN and how it can be improved. This work maps the feature activation to input pixel space by using Deconvolution Network that suggested by Zeiler, so it figure out the important part from the image.
2. Approach  
   They used almost same model that suggested by Alex Krizhevsky. And I don’t understand about the detail setting for dataset.   
   1. Visualization with a Deconvnet  
      CNN perform the classification by interpreting the feature inside of the image. And this research suggest a novel way to map above activities back to the input pixel space, so it shows the input pattern that cause a given activation in the feature maps. For this reverse mapping process, they used Deconvolutional Network (deconvnet). Following figure provide a continuous path back to image pixels.

This task successively perform unpooling, rectify and filter to reconstruct the activity in the layer.

* Unpooling   
  memorize the location of the pivots at the original images and store it on the ‘switch’. In the deconvnet, unpooling operation uses this switch information to reconstruct the original filter.
* Rectification   
  ReLU make sure that feature maps have always positive value. (Why did they do this?)
* Filtering   
  Use the inverted filter that used in the convolution. Use the switch information.

1. Training Detail  
   Basic structure of CNN is based on the Alex Net, but there is a difference that sparse connection on the 3rd 4th 5th layer are replaced with dense connection. Other differences relating to layer 1 and 2 are described in section 4.1. Image Net 2012 was used for the data. Stochastic gradient descent with mini batch was used. They annealed the learning rate throughout the training manually when the validation error plateaus. Dropout is used in fully connected layer with a rate of 0.5.
2. Convnet Visualization  
   Following feagure shows the visualization of features (If you want to see the rest of the image which is for layer 4,5, look the original paper) remaining figure



Feature Visualization: (I’m afraid that I can’t explain this clearly, because this is quit complicated concept and on top of that, I’m not sure that I’m fully understanding this. : ( ) Above figure shows the top 9 different activated part by having the on the left hand side picture as a input. They select the top 9 activations for a given feature map instead of selecting the single strongest activation. (??)

Projecting each separately down to pixel space reveals the different structures that excite a given feature map, hence showing its invariance to input deformations. Even though the responses on the feature map seems so similar, you can check that input image looks quite different. In other word, similar feature visualiztion will be represented when it takes similar type of inputs.

And you can check the project Do you hear the people sing singing the song of angry man this is the music of the people who will not be slaves again when the biting of your hart echoes the beating of the drum there is a life about to start tomorrow comes will you join in our crusade who will be strong and stand with me beyond the barricade is there a world you long to see than join in the fight that will give you the right to be freee2222



Question.

1. At the unpooling from the 2.1, how did they recover or process remaining spaces?
2. What does the ‘Feature extracor’ do?
3. At the section 4, what does ‘top 9 activation’ mean? As long as I understand, top 9 activation means 9 input picture that activated same region(same filter), and the feature map means the set of filter..?