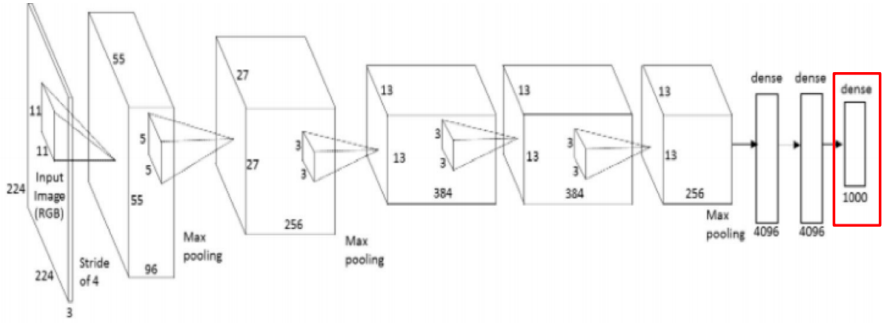
Deep Learning for Computer Vision

Explaining CNNs: Early Methods

Vineeth N Balasubramanian

Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

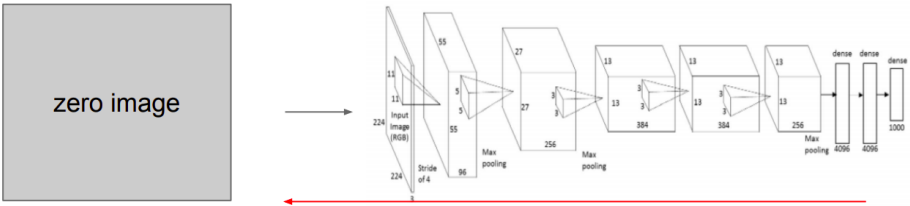
Backpropagation to Image



Question: Can we find an image that maximizes some class score?

Backpropagation to Image[1]

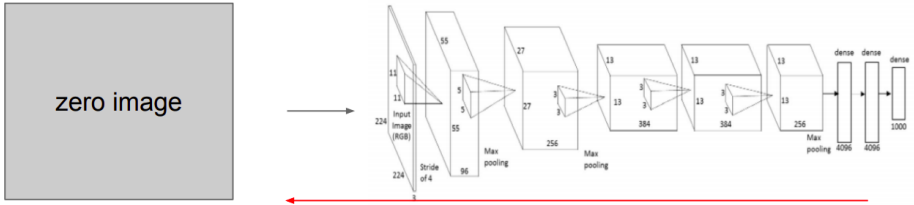
1. Feed zeros as input.



2. Set the gradient of the scores vector to be [0, 0 , .. , 1 , ..., 0]. Then backprop to image.
3. Do a small "**image update**"
4. Forward pass the image through the network.
5. Go back to step 2.

[1]Simonyan, Vedaldi, and Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop 2014

Backpropagation to Image[2]



- Formally, this optimization can be written as:

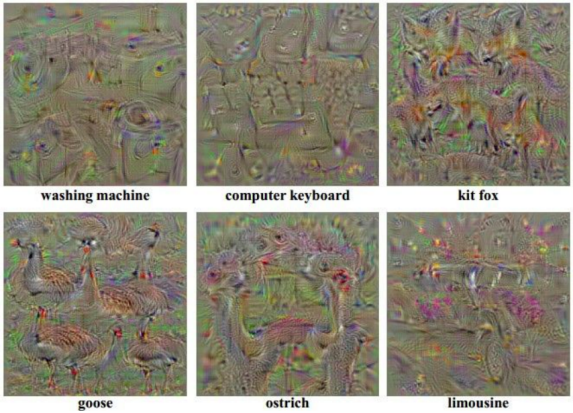
$$\arg\max_{I} S_c(I) - \lambda \|I\|_2$$

- Here, S_c is the scores vector for class c, before applying softmax.

[2]Simonyan, Vedaldi, and Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop 2014

Backpropagation to Image[3]

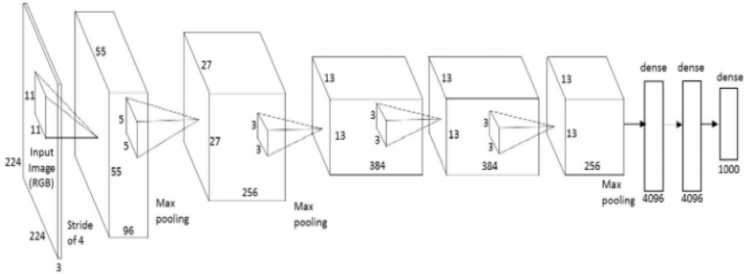
- Finding images that maximize some class score:



[3]Simonyan, Vedaldi, and Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop 2014

Backpropagation to Image[4]

- Such optimization can in fact be done for arbitrary neurons in the network



- Repeat:
  1. Forward an image
  2. Set activations in a layer of interest to all zero, except 1.0 for a neuron of interest
  3. Backprop to image
  4. Do an "image update"

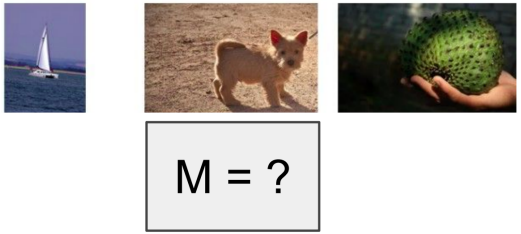
[4] Simonyan, Vedaldi, and Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop 2014

Visualizing the Data Gradient[5]

- Since the gradient on image data has three channels, visualise M such that:

$$M_{ij} = \max_c |\nabla_I S_c(I)|_{(i,j,c)}$$

- At each pixel, take absolute value and pick maximum across channels



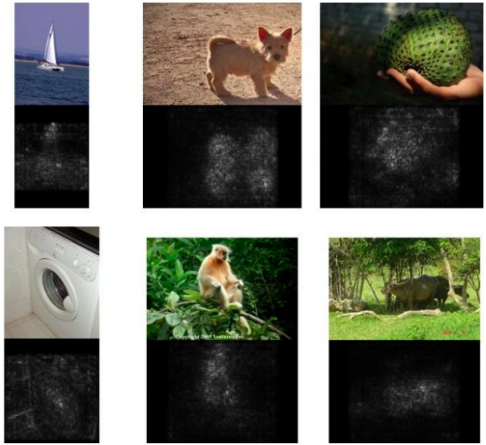
[5]Simonyan, Vedaldi, and Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop 2014

Visualizing the Data Gradient[6]

- Since the gradient on image data has three channels, visualise M such that:

$$M_{ij} = \max_c |\nabla_I S_c(I)|_{(i,j,c)}$$

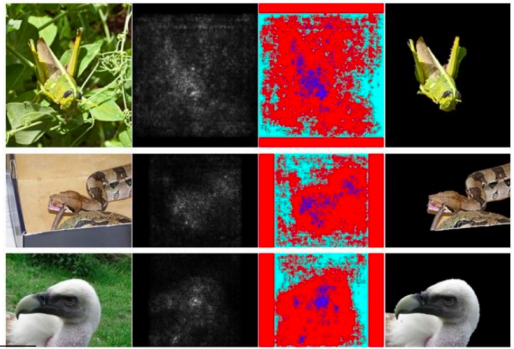
- At each pixel, take absolute value and pick maximum across channels



[6] Simonyan, Vedaldi, and Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop 2014

Visualizing the Data Gradient[8]

- **GrabCut**[7], a segmentation method, can be applied to obtain the object mask from the data gradient
  - Recall Graph-Cut segmentation we saw earlier - GrabCut is an extension/adaptation



[7] https://docs.opencv.org/3.4/d8/d83/tutorial_py_grabcut.html

[8] Simonyan, Vedaldi, and Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop 2014

Image Reconstruction from Latent Representation

Given a CNN code (latent representation from a layer, say, FC7), is it possible to reconstruct the original image?

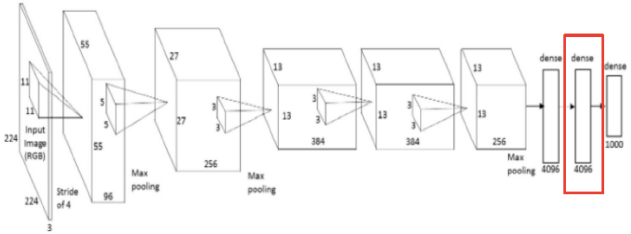
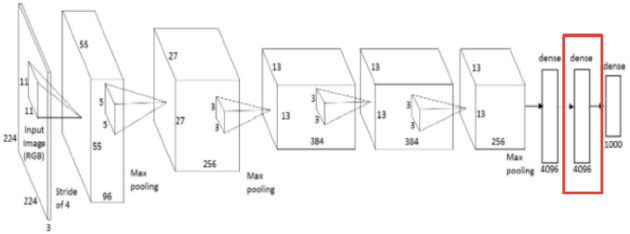


Image Reconstruction from Latent Representation

Given a CNN code (latent representation from a layer, say, FC7), is it possible to reconstruct the original image?



Yes, solve an optimization problem such that:

- The image's code is similar to a given code
- It "looks natural" (image prior regularization)

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^{H \times W \times C}}{\arg\min} \|\mathbf{\Phi}(x) - \mathbf{\Phi}_0\|^2 + \lambda \mathcal{R}(\mathbf{x})$$

Image Reconstruction from Latent Representation

On AlexNet model

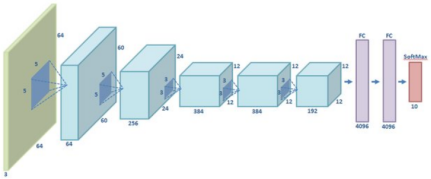
original image

reconstructions from the 1000 log probabilities for ImageNet (ILSVRC) classes

Image Reconstruction from Latent Representation

Reconstructions from representations after last pooling layer (before first FC layer) in AlexNet

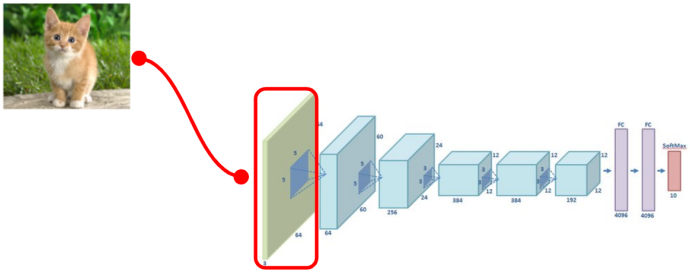
Guided Backpropagation (also known as Deconvolution method)[9]



[9]Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

Guided Backpropagation (also known as Deconvolution method)[9]

a) Feed image into net.

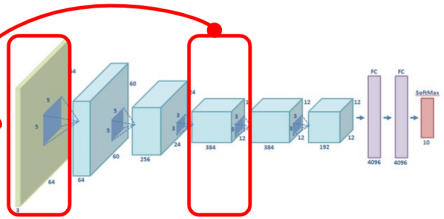


[9]Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

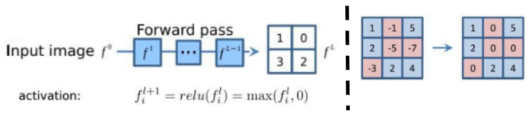
Guided Backpropagation (also known as Deconvolution method)[9]

a) Feed image into net.

b) Pick a layer, set the gradient there to zero except for the neuron of interest.



b)



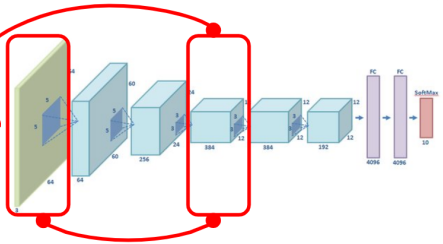
[9] Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

Guided Backpropagation (also known as Deconvolution method)[9]

a) Feed image into net.

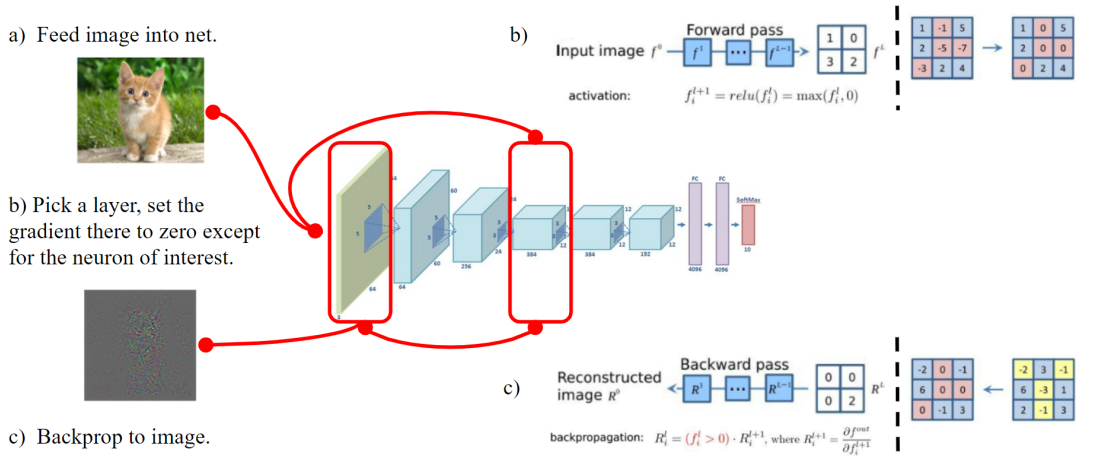
b) Pick a layer, set the gradient there to zero except for the neuron of interest.

c) Backprop to image.



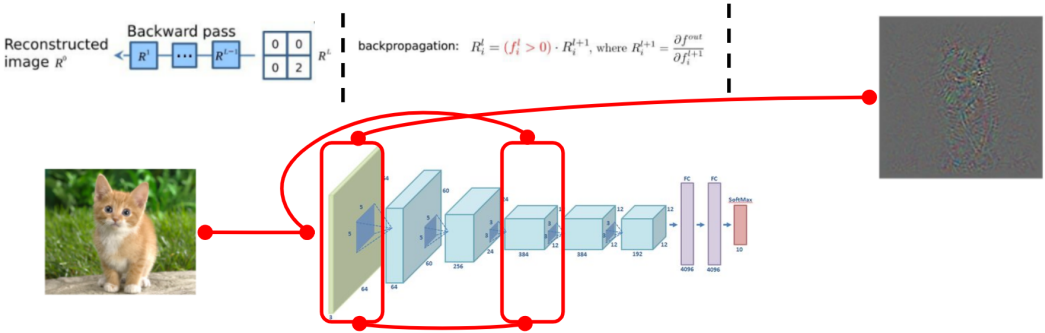
[9] Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

Guided Backpropagation (also known as Deconvolution method)[9]



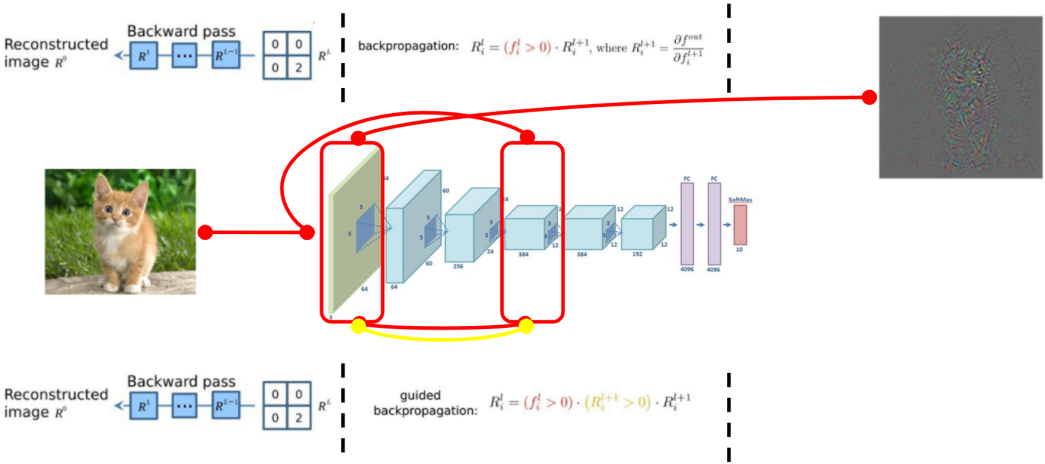
[9] Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

Guided Backpropagation (also known as Deconvolution method)[9]



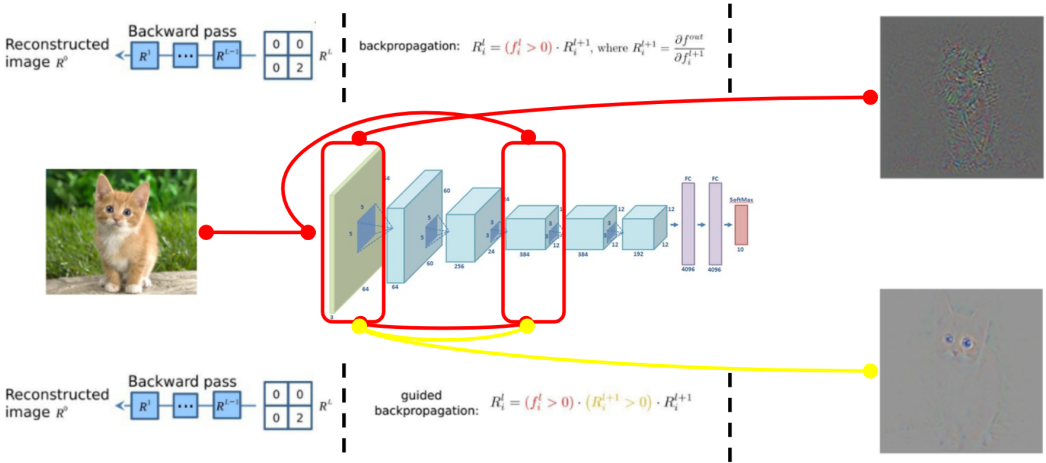
[9] Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

Guided Backpropagation (also known as Deconvolution method)[9]



[9] Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

Guided Backpropagation (also known as Deconvolution method)[9]



[9] Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

Readings

Summary of Visualizing CNNs

- Lecture Notes of CS231n, Stanford

Other Recommended Readings/References

- Simonyan, Vedaldi, and Zisserman, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, ICLR Workshop 2014
- Zeiler and Fergus, Visualizing and Understanding Convolutional Networks, ECCV 2014
- Springenberg et al, Striving for Simplicity: The All Convolutional Net, ICLR Workshop 2015

Exercises

- Understand GrabCut and how it can be used to generate masks from data gradients