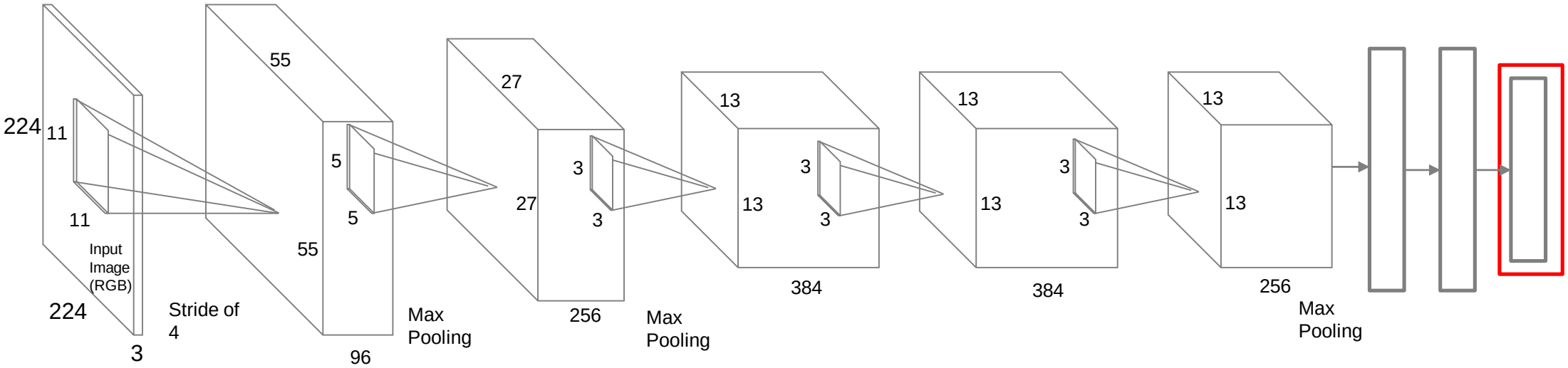
Deep Learning (for Computer Vision)

Arjun Jain

Computer Vision: Visualizing and Understanding ConvNets

Optimization to Image

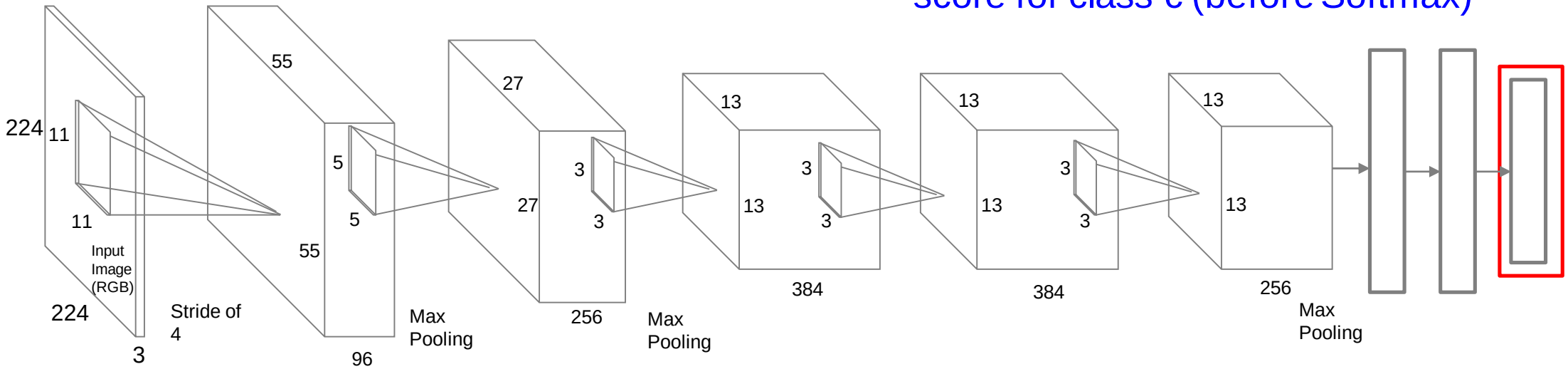


Q: can we find an image that maximizes some class score?

Optimization to Image

$$\arg\max_{I} \boxed{S_c(I)} - \lambda \|I\|_2^2$$

score for class c (before Softmax)

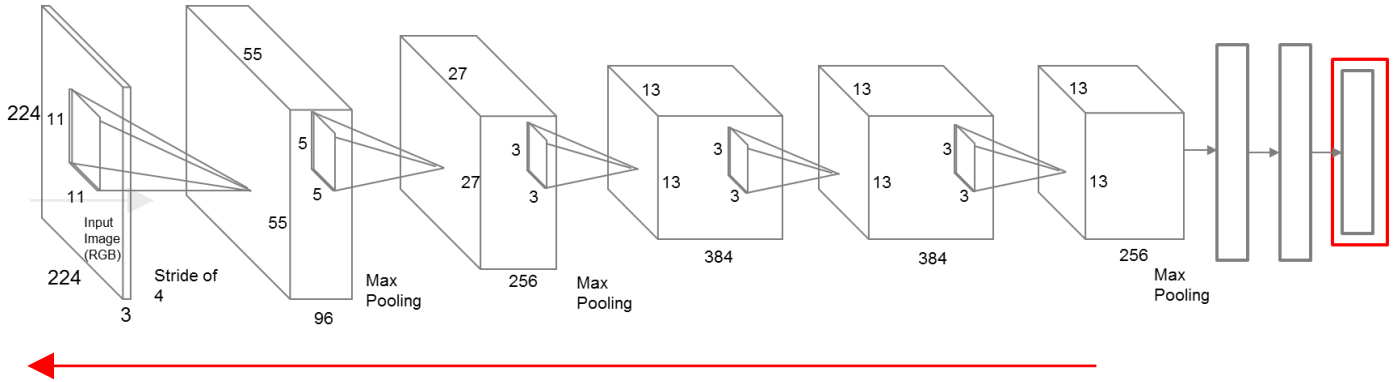


Q: can we find an image that maximizes some class score?

Optimization to Image

1. feed in zeros.

zero image

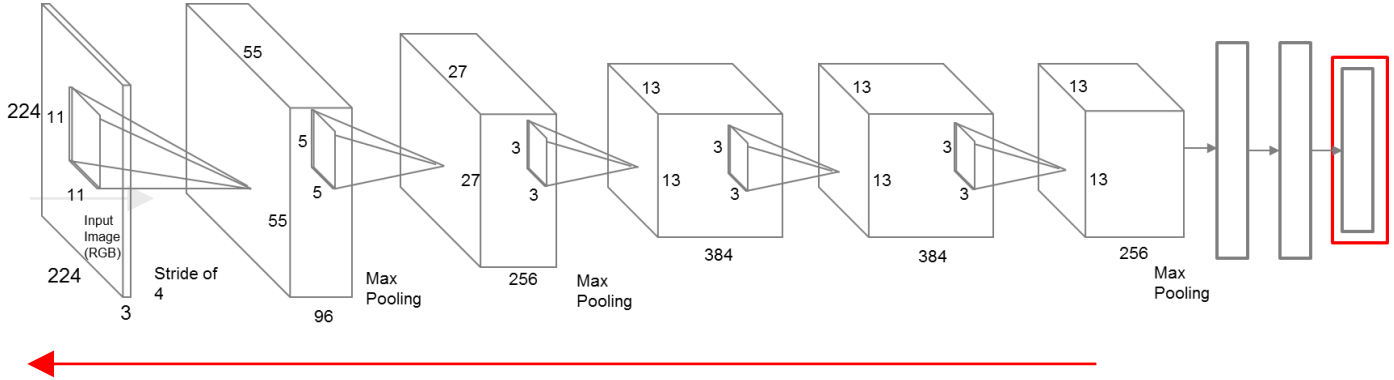


2. set the gradient of the scores vector to be [0,0,....1,....,0], then backprop to image

Optimization to Image

1. feed in zeros.

zero image



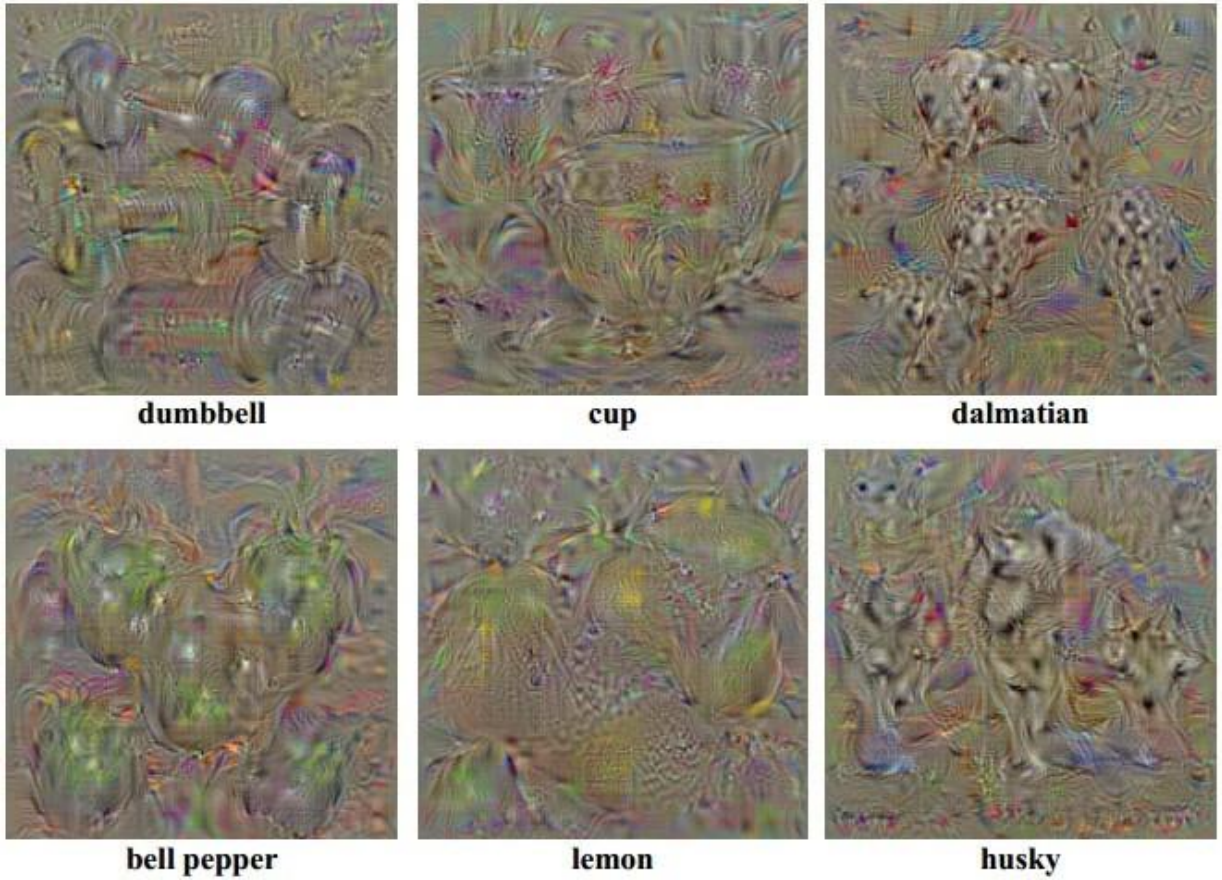
2. set the gradient of the scores vector to be [0,0,....1,....,0], then backprop to image
3. do a small "image update"
4. forward the image through the network.
5. go back to 2.

$$\arg\max_{I} \boxed{S_c(I)} - \lambda \|I\|_2^2$$

score for class c (before Softmax)

Optimization to Image

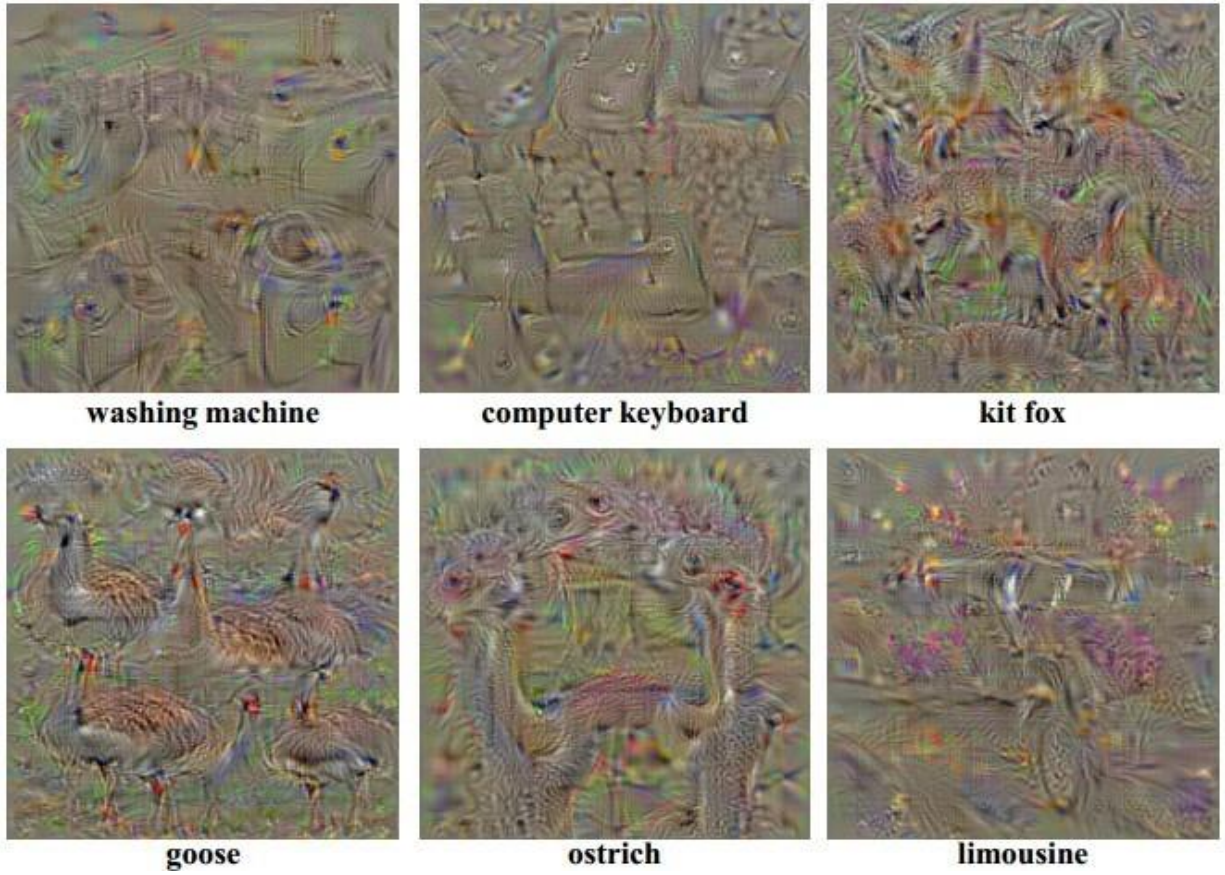
1. Find images that maximizesome class score:



Source: *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps Karen Simonyan, Andrea Vedaldi, Andrew Zisserman, 2014*

Optimization to Image

1. Find images that maximizesome class score:



Source: *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps Karen Simonyan, Andrea Vedaldi, Andrew Zisserman, 2014*

Optimization to Image

Yosinksi proposed a different form of regularizing the image

$$\arg\max_{I} S_c(I) - \boxed{\lambda \|I\|_2^2}$$

More explicit scheme:

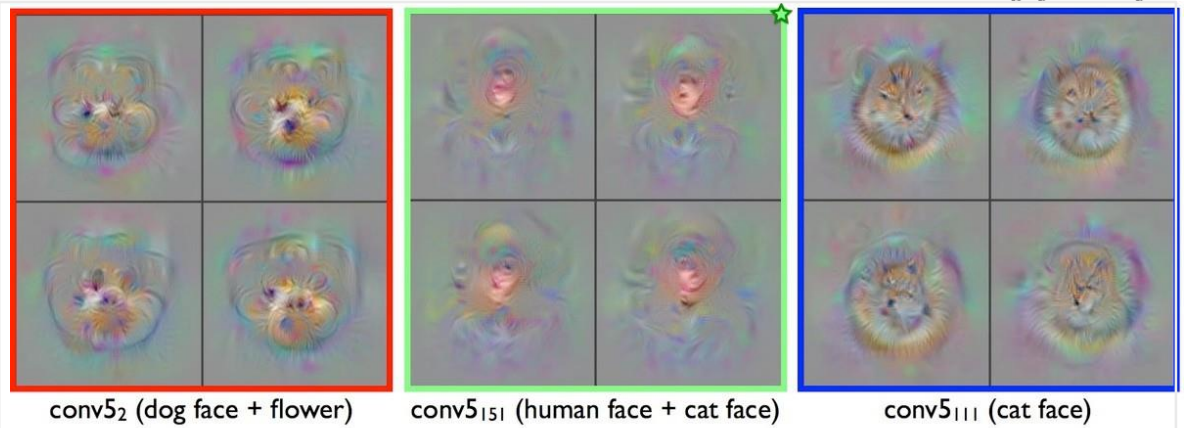
Repeat:

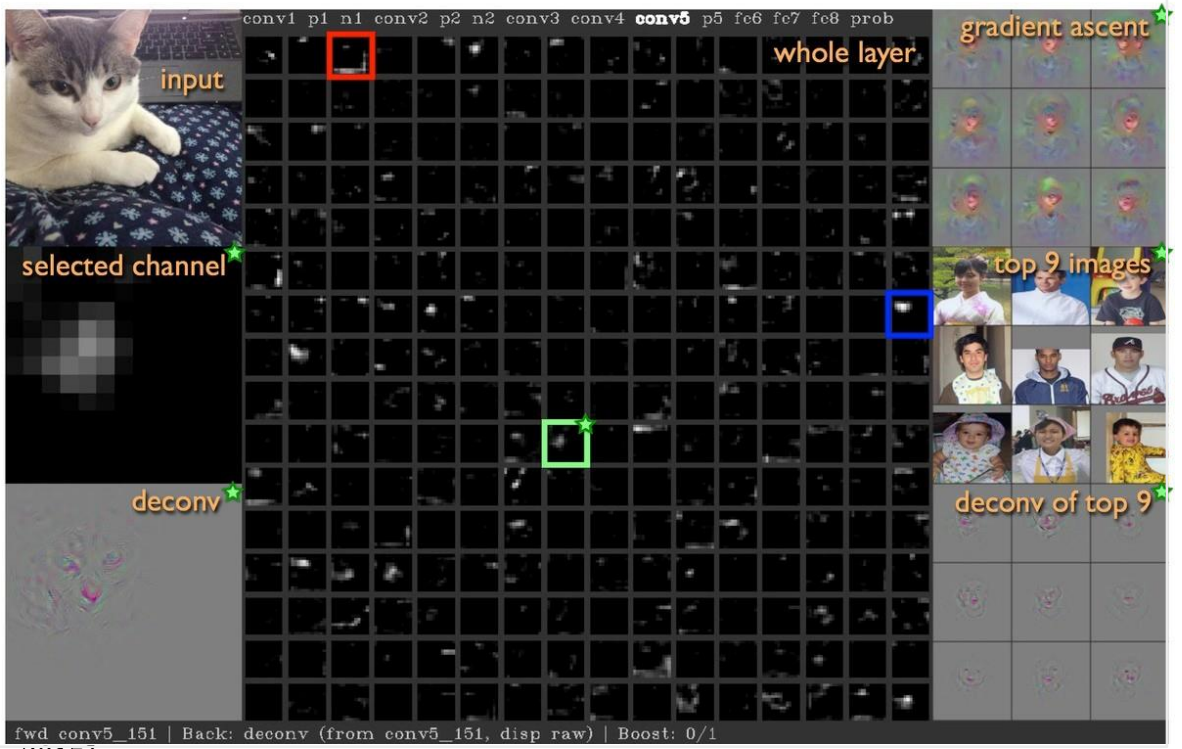
- Update the image **x** with gradient from some unit of interest
- Blur x a bit
- Take any pixel with small norm to zero (to encourage sparsity)

Source: *[Understanding Neural Networks Through Deep Visualization, Yosinski et al. , 2015]*

Optimization to Image

http://yosinski.com/deepvis





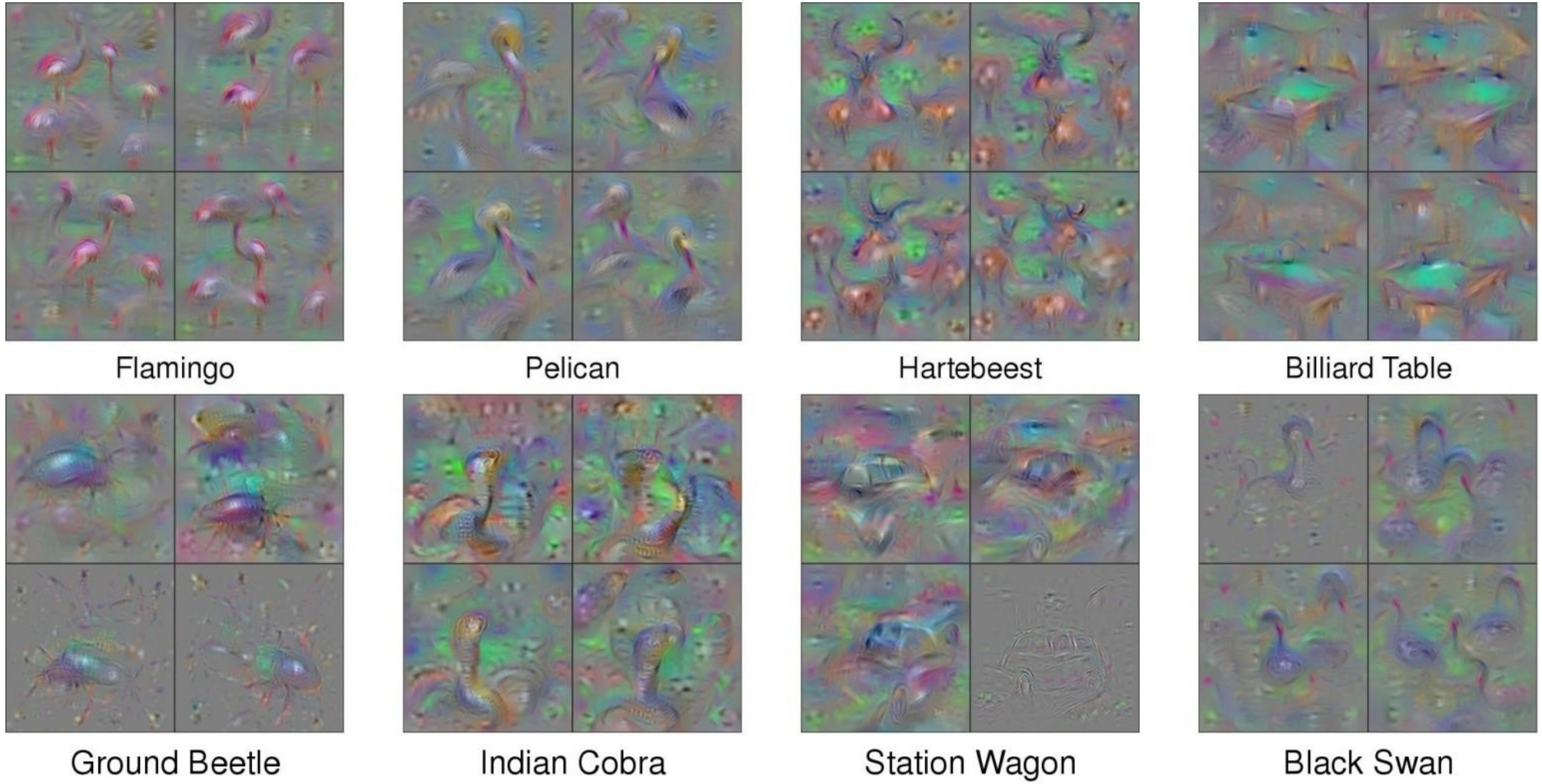
YouTube video

https://www.youtube.com/watch?v=AgkfIQ4IGaM

(4min)

Source: [Understanding Neural Networks Through Deep Visualization, Yosinski et al. , 2015]

Optimization to Image



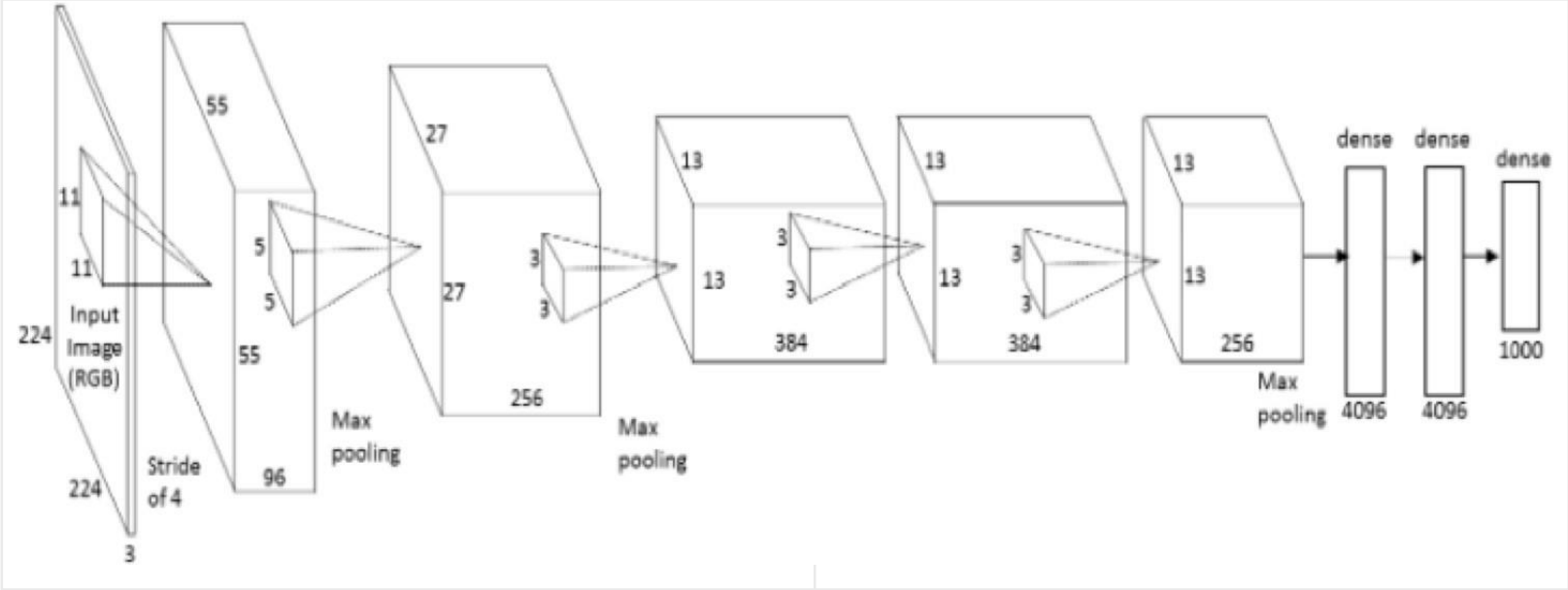
http://yosinski.com/deepvis

4 images = 4 differentinitializations

Source: *[Understanding Neural Networks Through Deep Visualization, Yosinski et al. , 2015]*

Optimization to Image

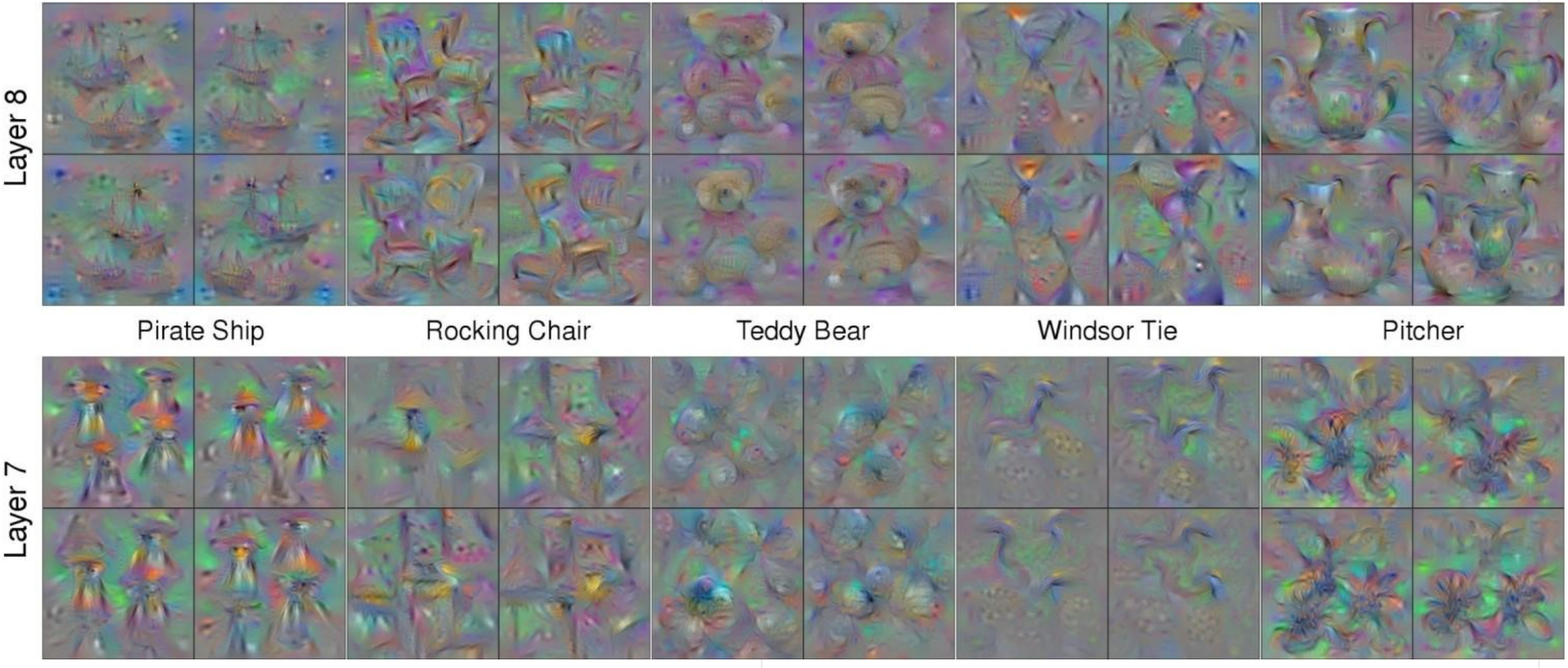
We can in fact do this for arbitrary neurons along the ConvNet



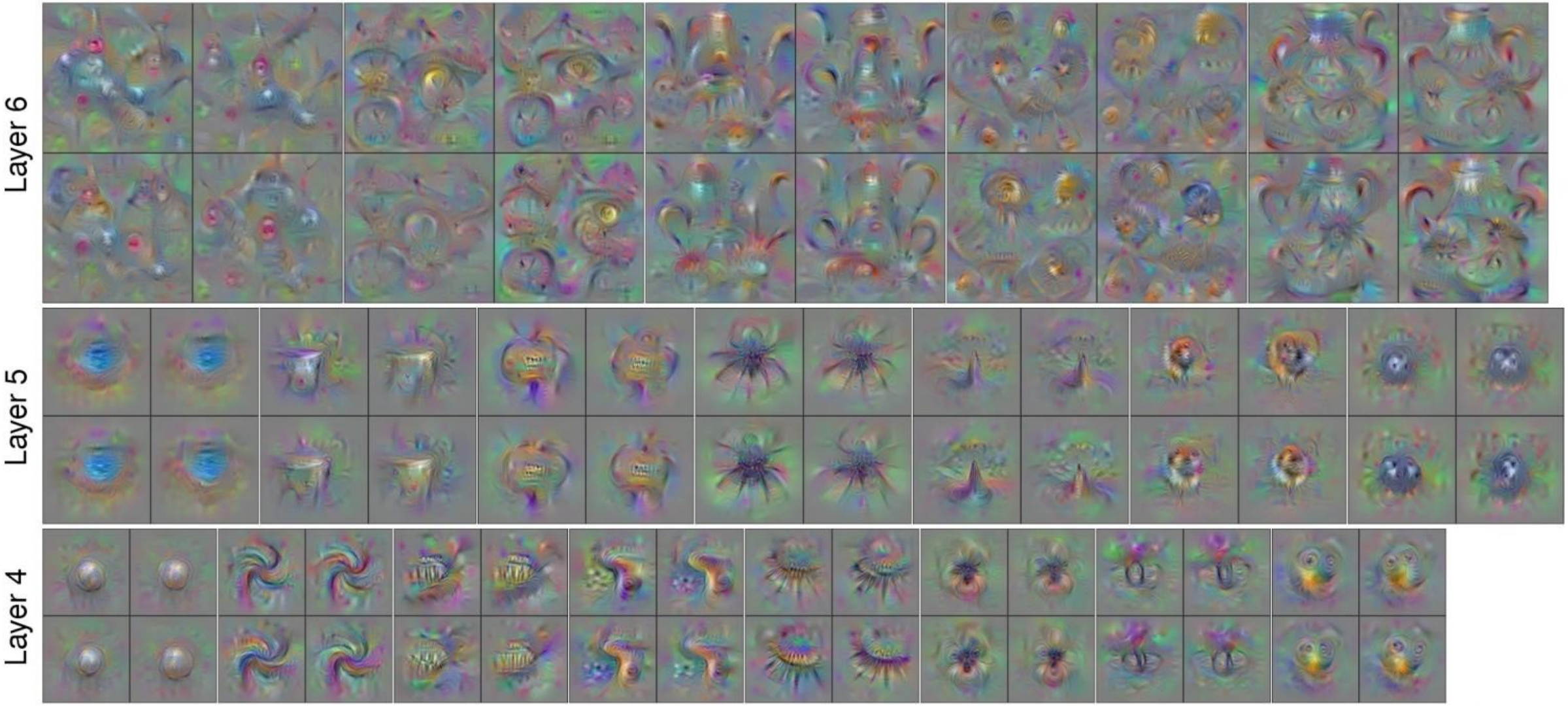
Repeat:

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

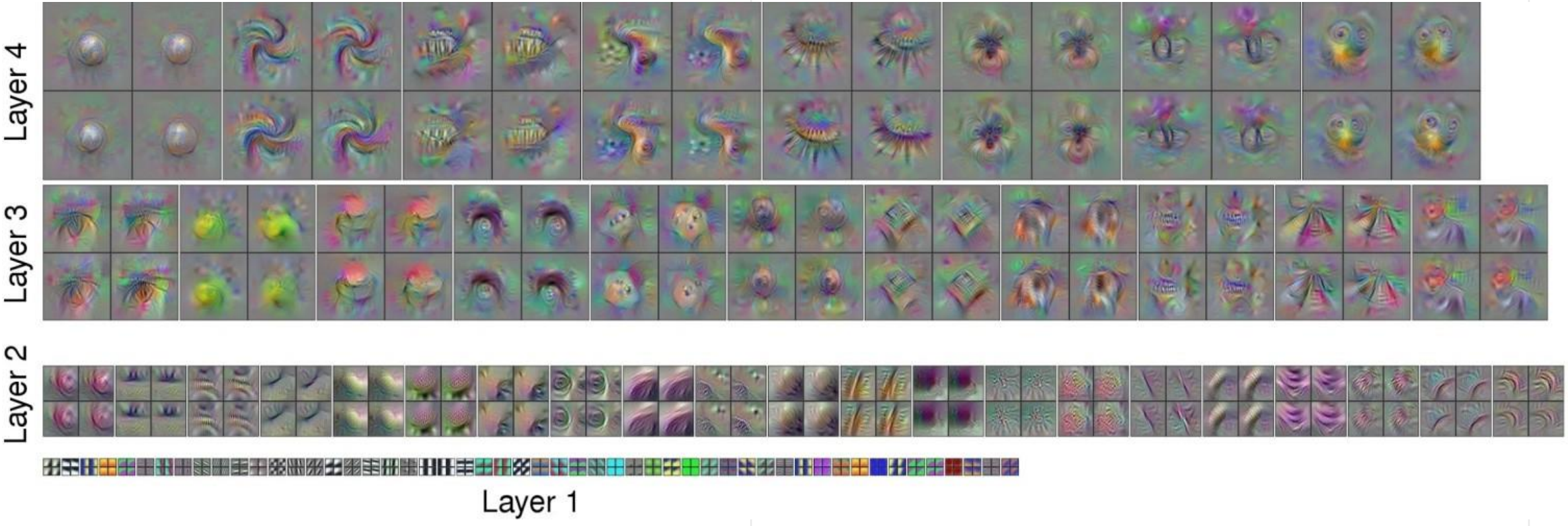
Optimization to Image



Optimization to Image



Optimization to Image



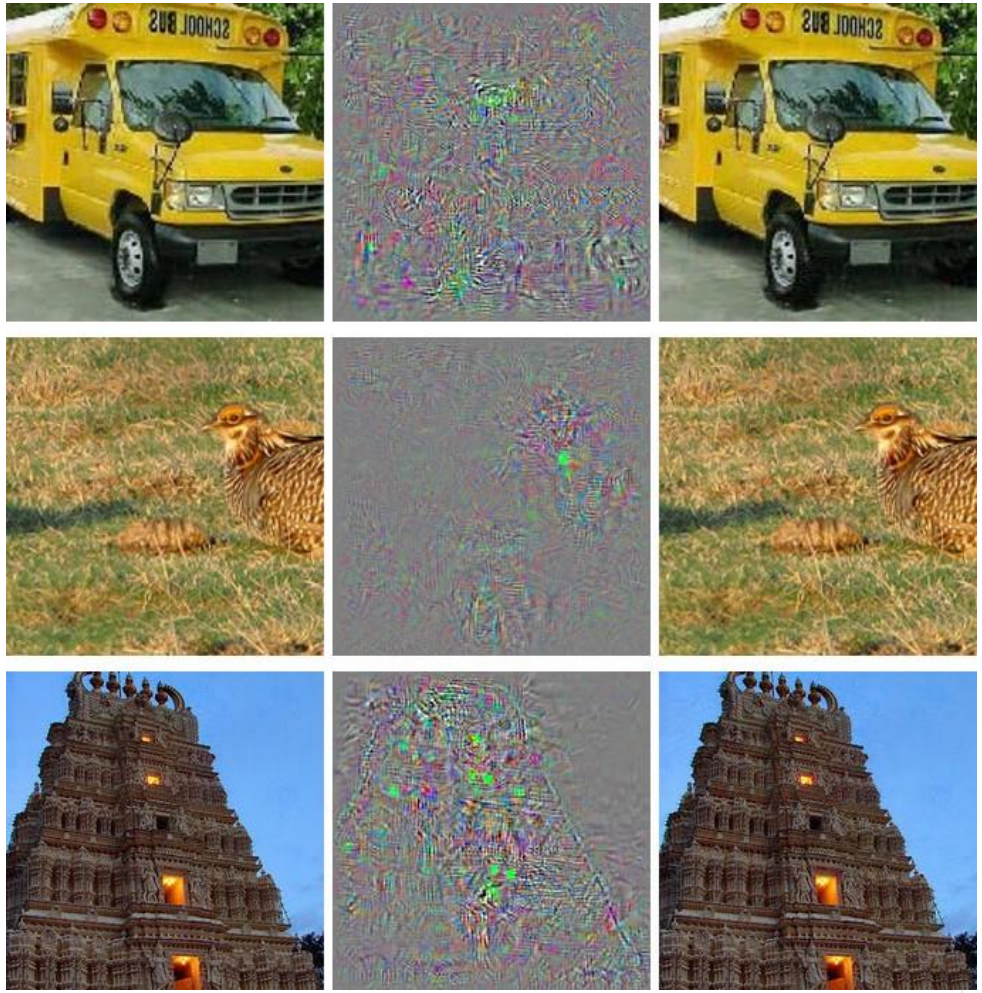
Optimization to Image

We can pose an optimization over the input image to maximize any class score.
That seems useful.

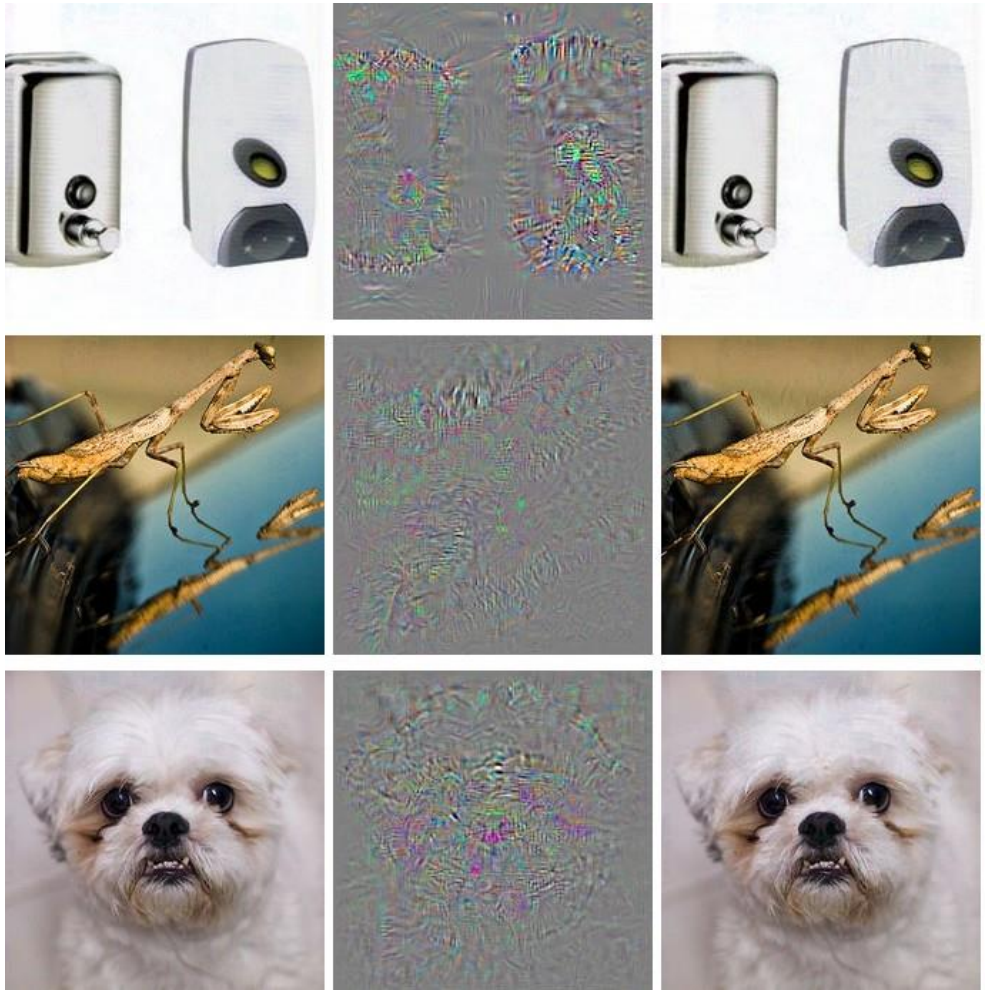
Question: Can we use this to “fool” ConvNets?

spoiler alert: yeah

Optimization to Image



correct +distort ostrich

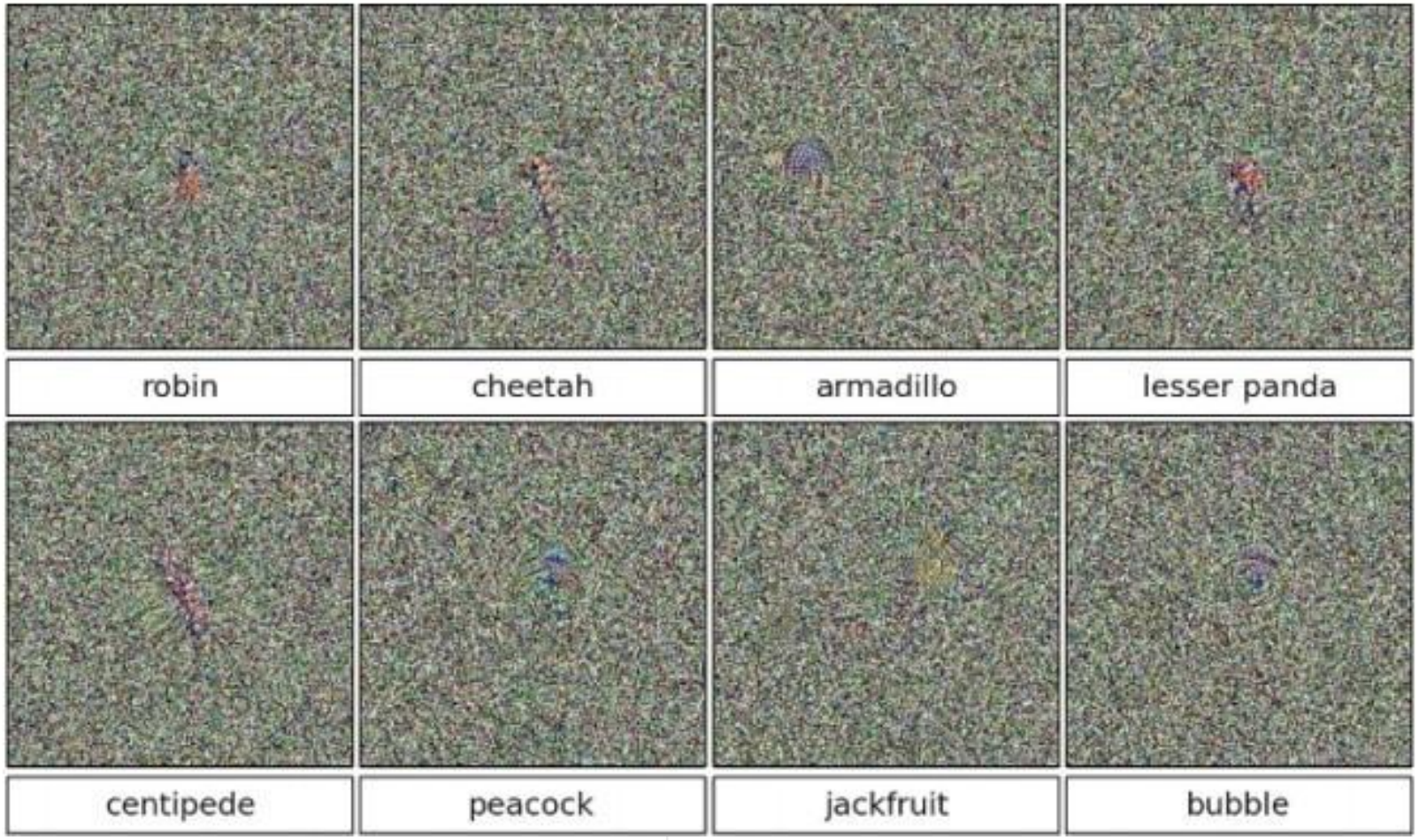


correct +distort ostrich

Source: [Intriguing properties of neural networks, Szegedy et al., 2013]

Optimization to Image

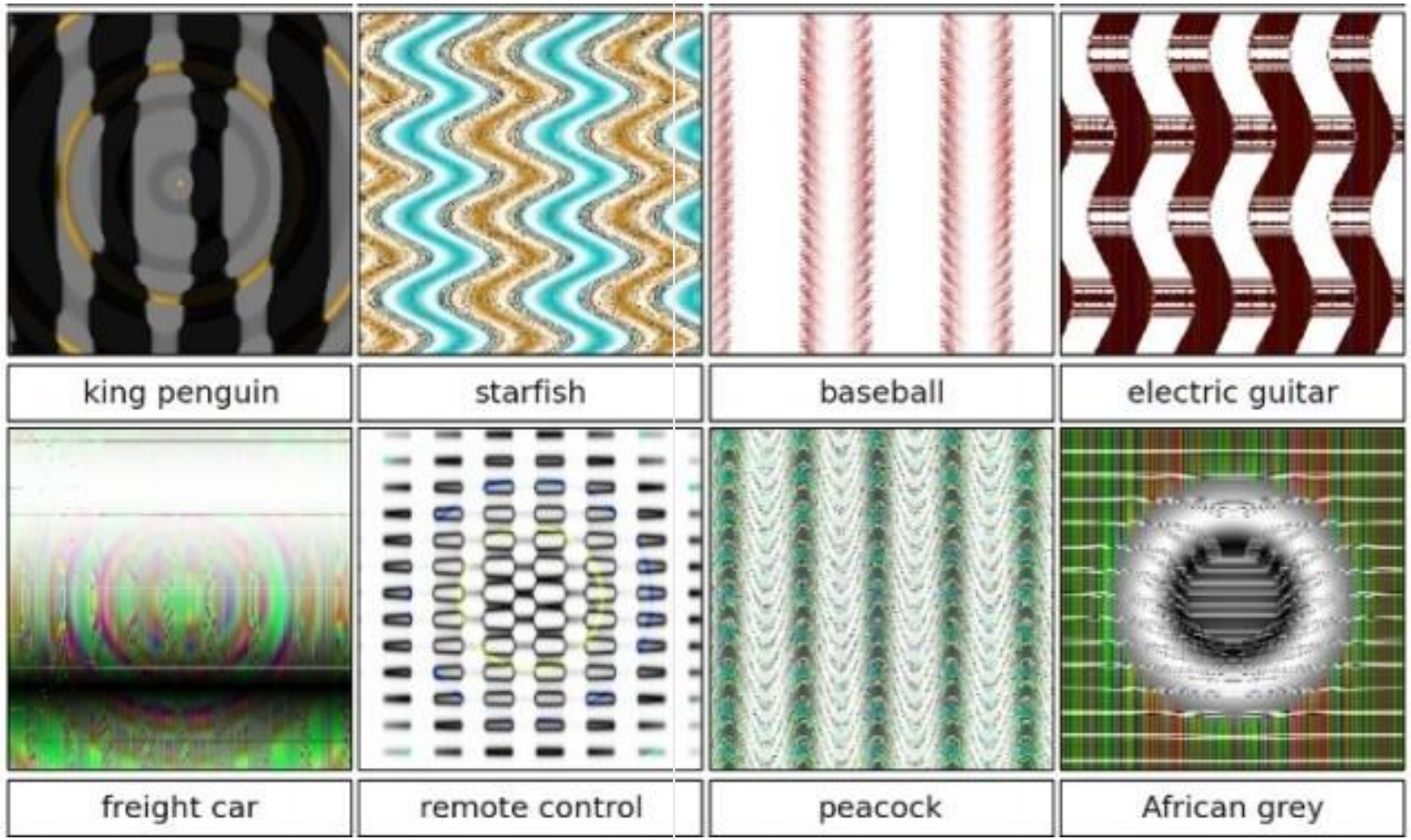
>99.6% confidences



Source: *[Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images Nguyen, Yosinski, Clune, 2014]*

Optimization to Image

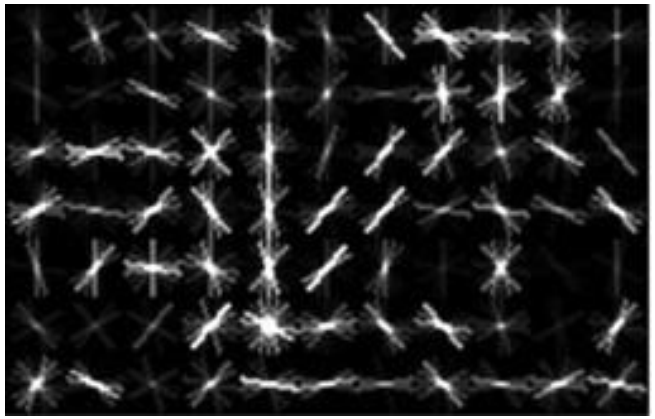
>99.6% confidences



Source: *[Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images Nguyen, Yosinski, Clune, 2014]*

Optimization to Image

These kinds of results were around even before ConvNets…



Identical HOG represention

Source: [Exploring the Representation Capabilities of the HOG Descriptor, Tatu et al., 2011]

Thank you!