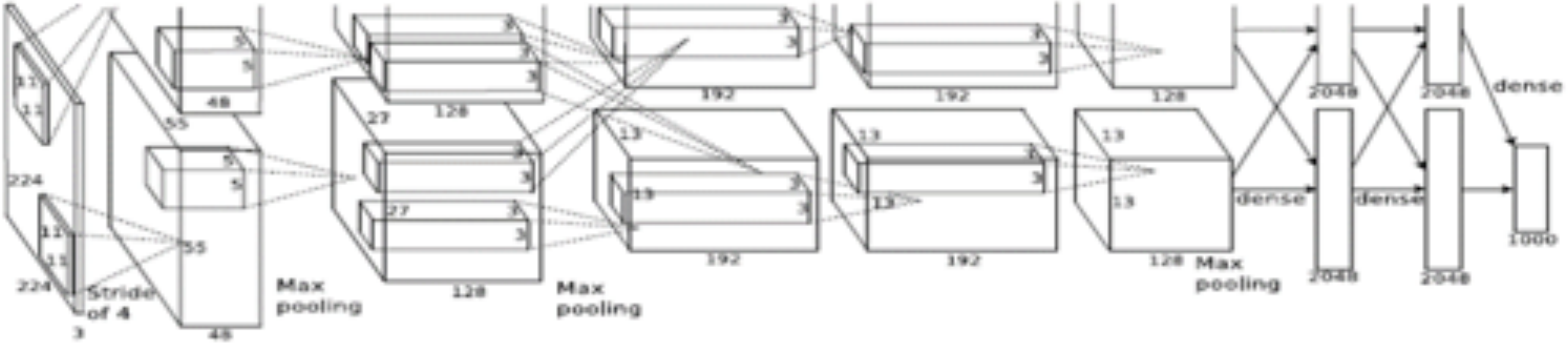
Turning Point: AlexNet



ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

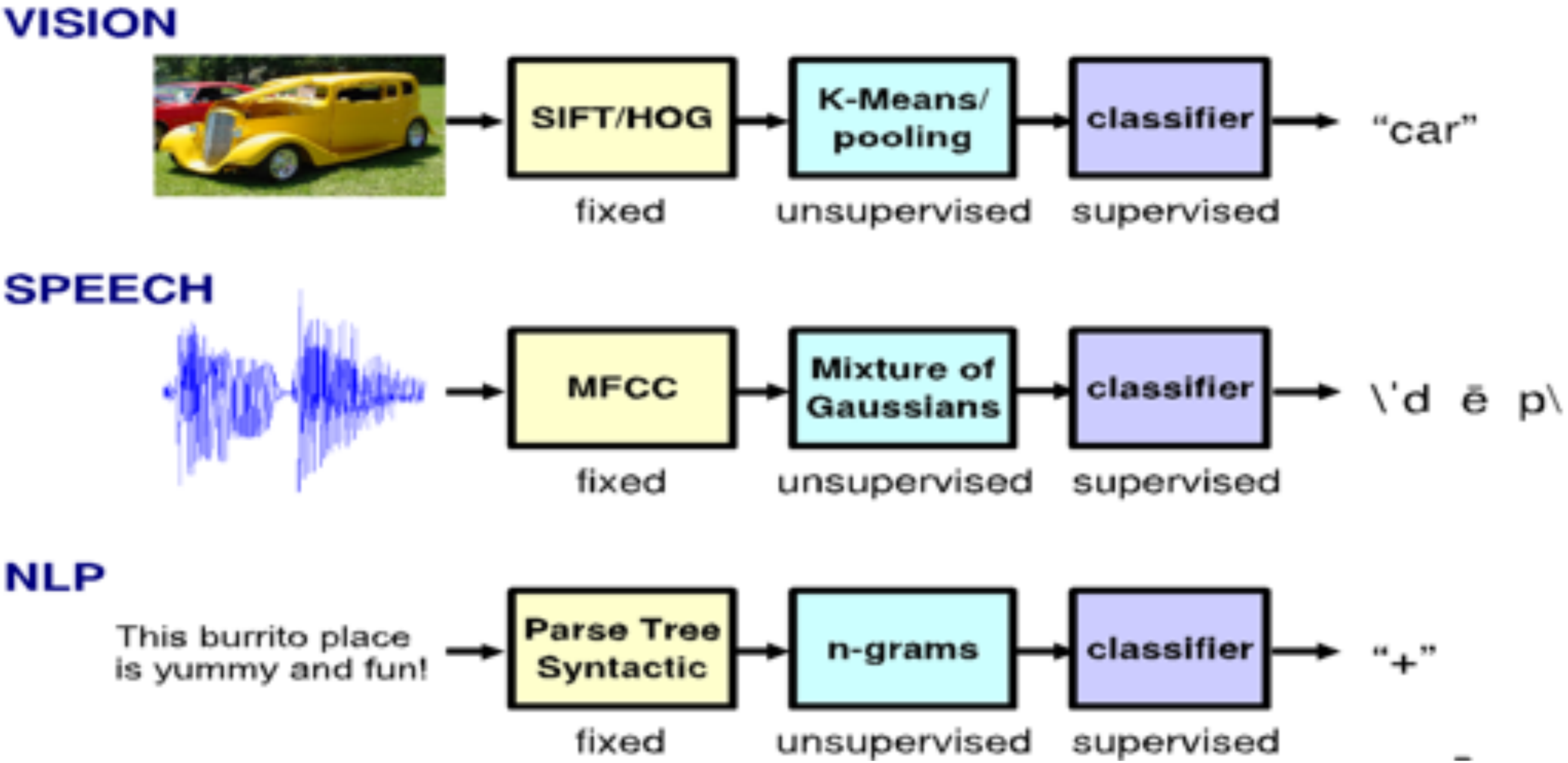
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

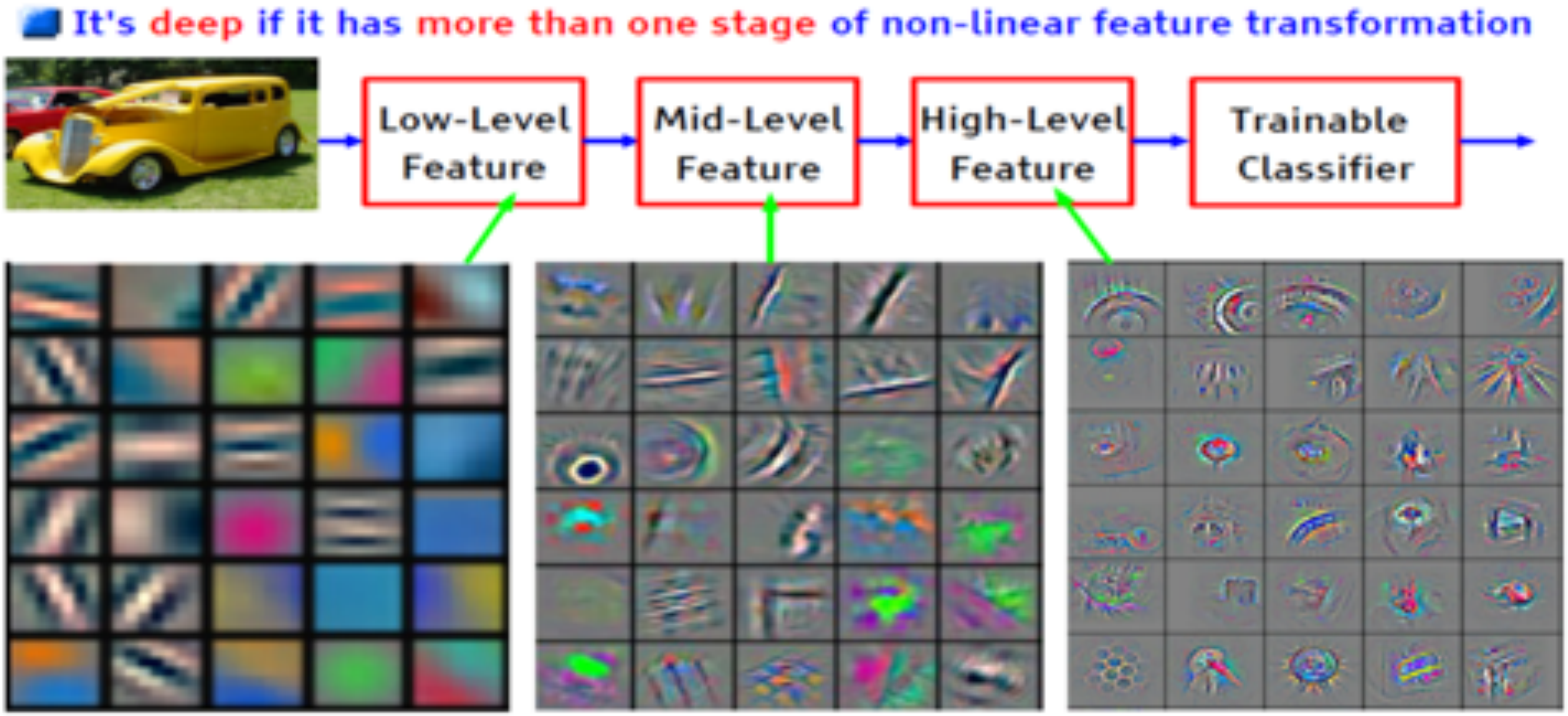
ImageNet Classification Task:

Previous Best : ~25% (CVPR-2011)
AlexNet : ~15 % (NIPS-2012)

Common Pipeline: Till Then



Deep Learnt Features



Learn the full pipeline

VISION

pixels → edge → texton → motif → part → object

SPEECH

sample → spectral band → formant → motif → phone → word

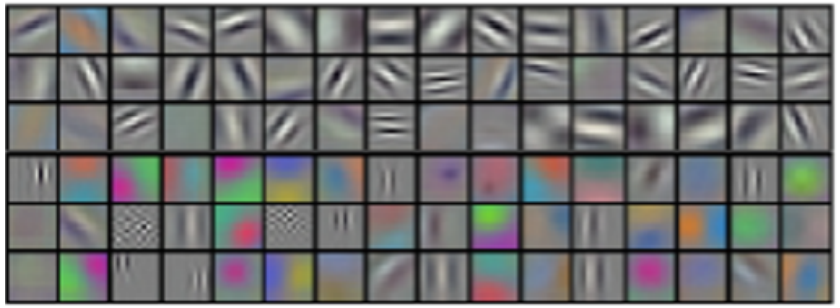
NLP

character → word → NP/VP/.. → clause → sentence → story

Visualizing CNNs

- CNNs are cool ☺️ but some of the below questions need answers before we move forward :-
- How do I interpret the learned filters?
- What is it that stimulates/excites a neuron?
- How do I decide the architecture or improve existing ones?

Source: Krizhevsky et.al. NIPS’12



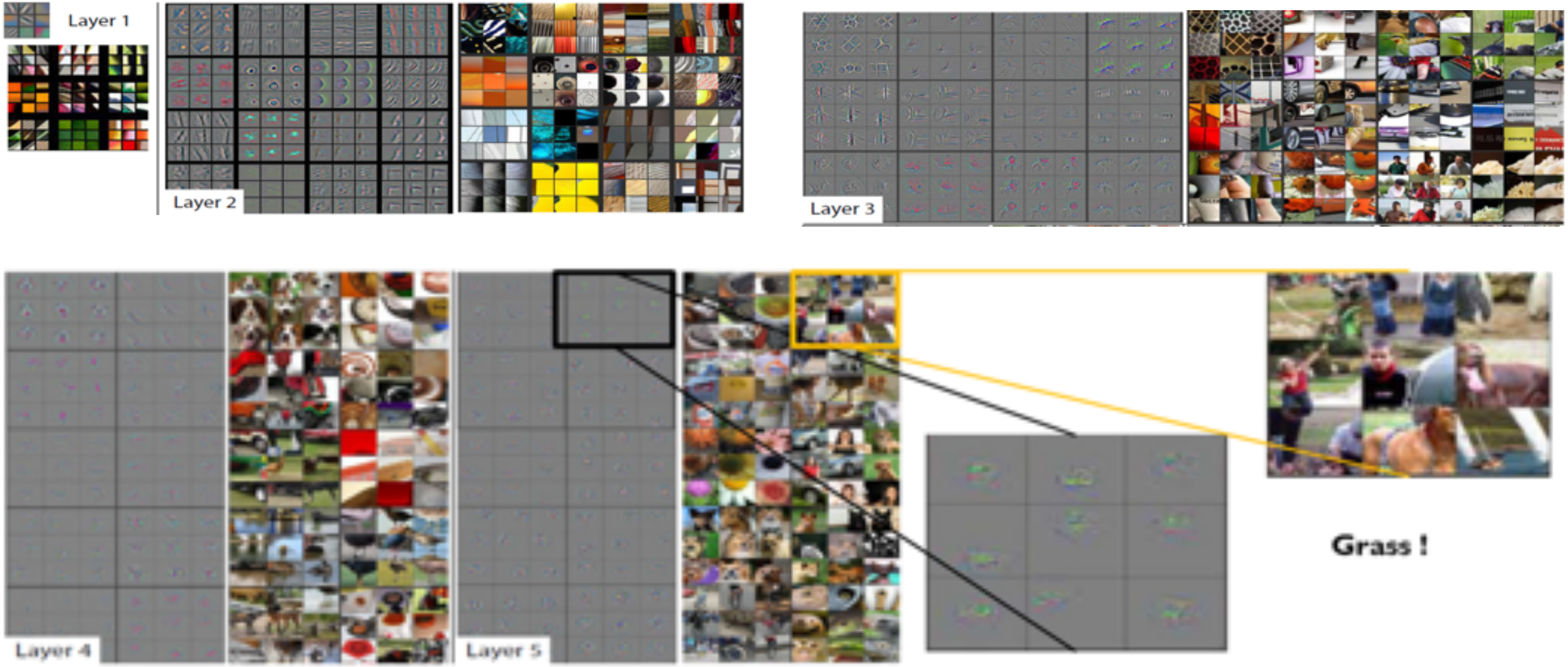
Visualizing the first conv. layer is possible but how about the later layers.

?

Zeiler and Fergus , Visualizing and Understanding Convolutional Networks,. ECCV 2014

Visualizing CNNs

A. How do I interpret the learned filters?



Source: Zeiler e.t. al. ECCV'14

Zeiler and Fergus , Visualizing and Understanding Convolutional Networks,. ECCV 2014

Early Layers Converge Faster

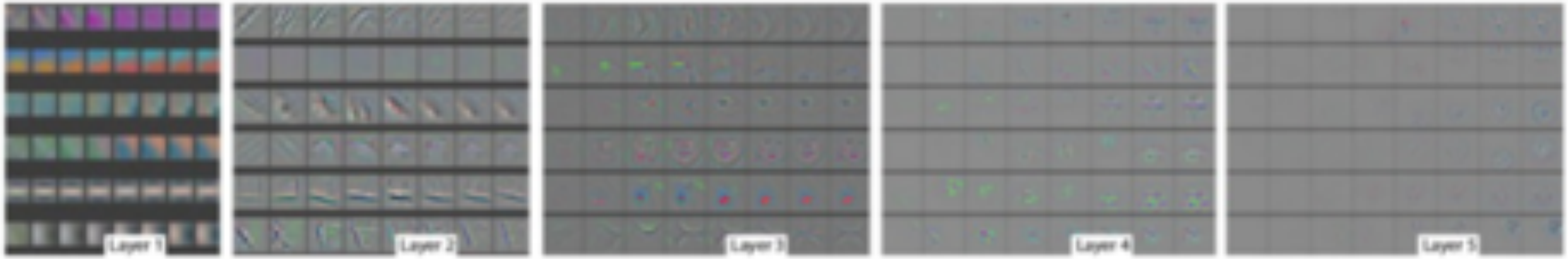
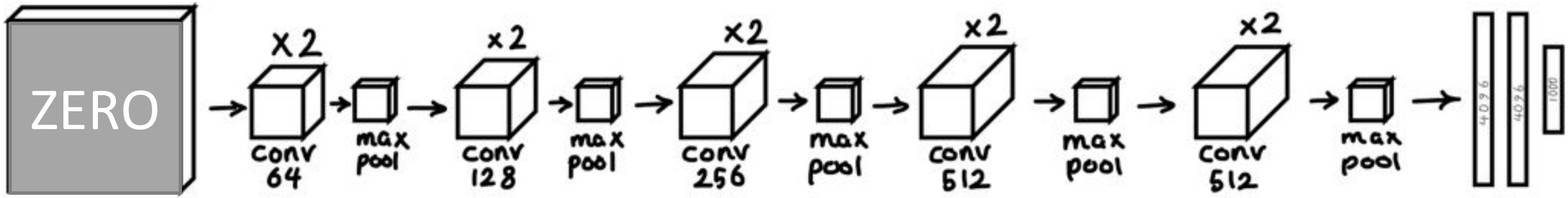
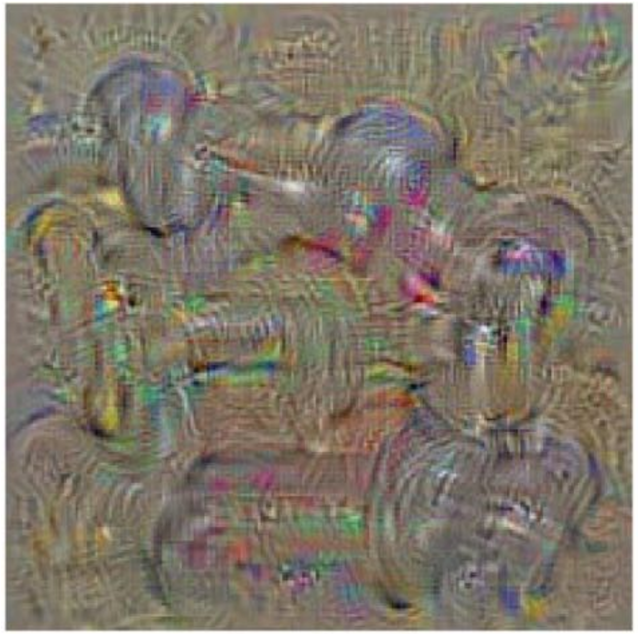


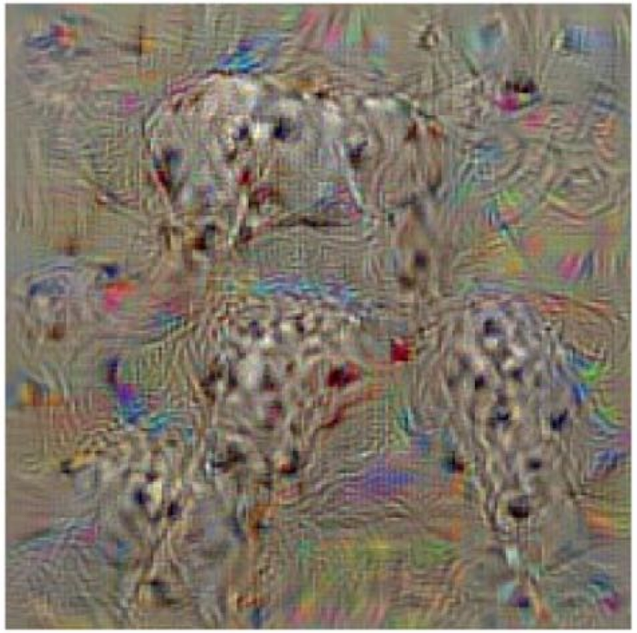
Figure: Evolution of randomly chosen subset of model features generated using deconvnet through training at epoch 1, 2, 5, 10, 20, 30, 40, 64.

Image optimization

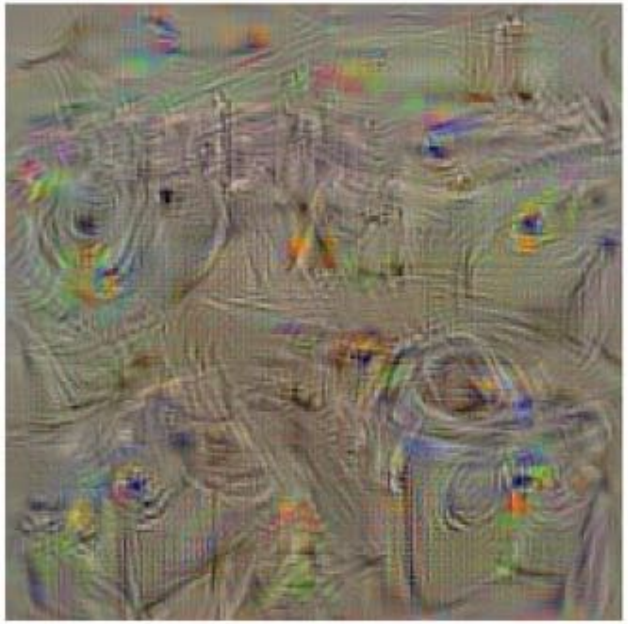




dumbbell



dalmatian

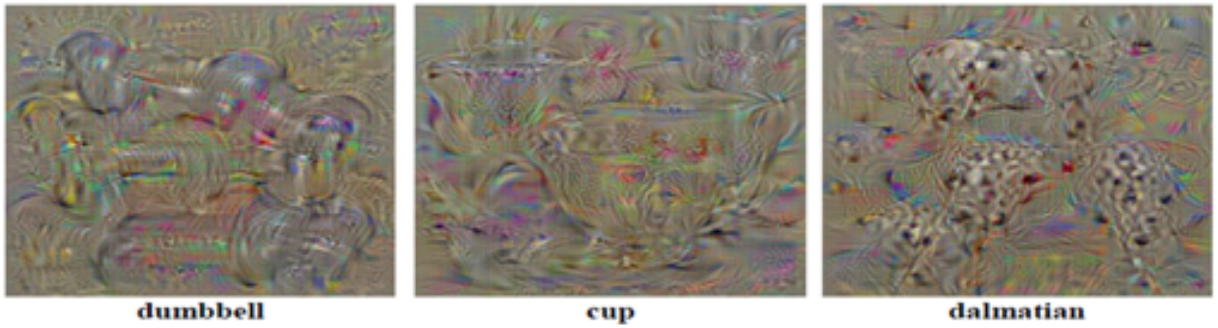


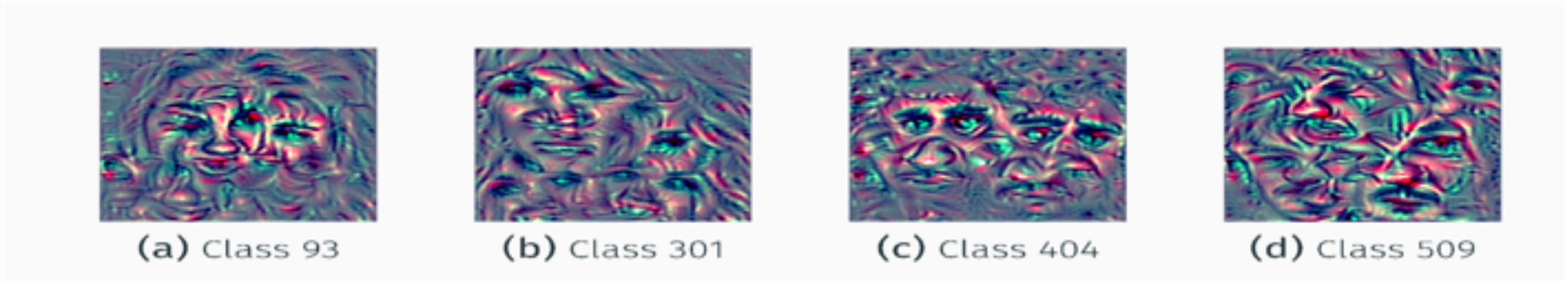
washing machine

Visualizing CNNs

- Class Model Visualization
  – Find an L_2 normalized image which maximizes the C_l class score
  – Initialize with mean image.
  – Back-propagate.

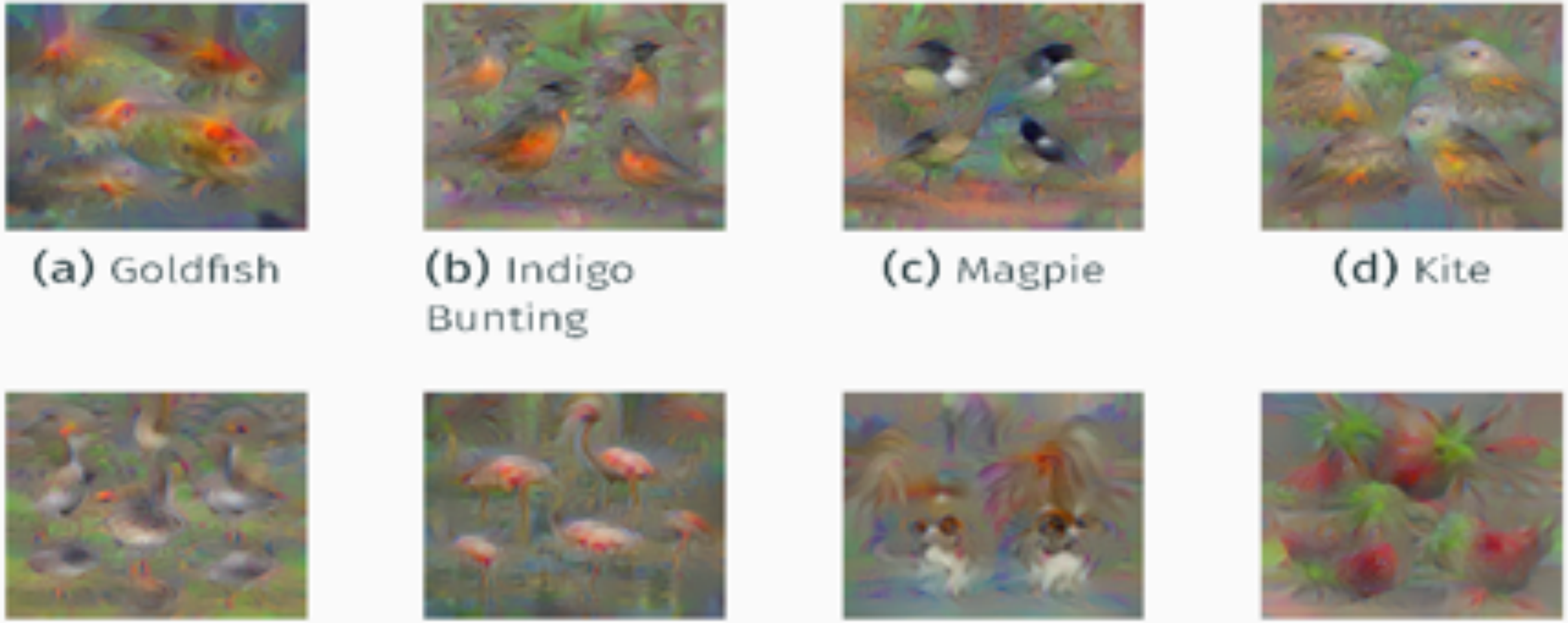
$$\arg\max_I S_c(I) - \lambda \|I\|_2^2,$$





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

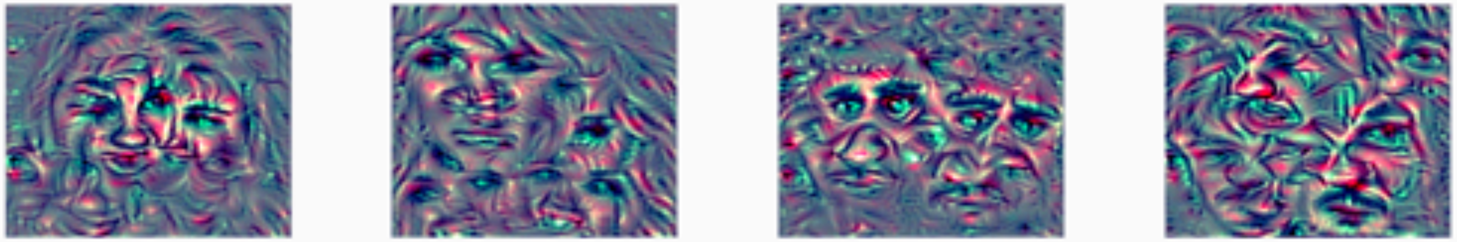
Better Visualizations



(a) Goldfish (b) Indigo Bunting (c) Magpie (d) Kite

(e) Goose (f) Flamingo (g) Japanese Spaniel (h) Strawberry

Better Visualizations

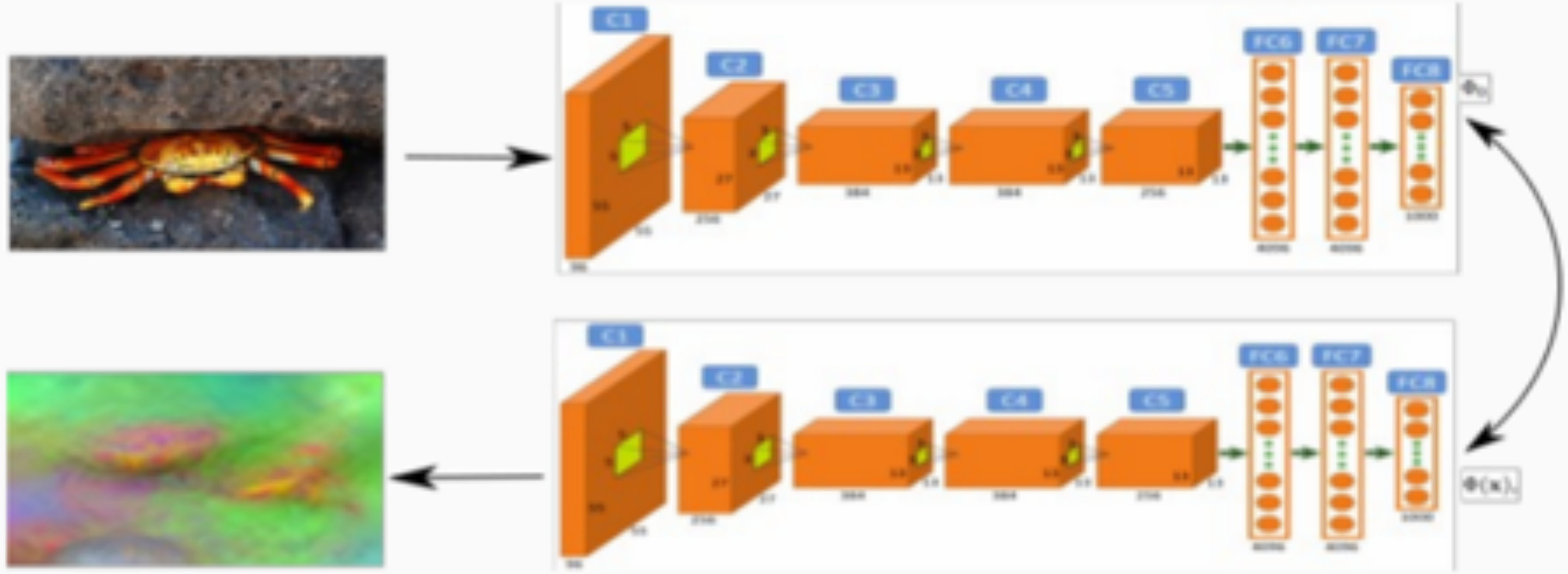


(a) Class 93 (b) Class 301 (c) Class 404 (d) Class 509

Simonyan et al. NIPS 2014,
Mahendran et al. CVPR 2015

Deep dream

Inverting Specific Representation

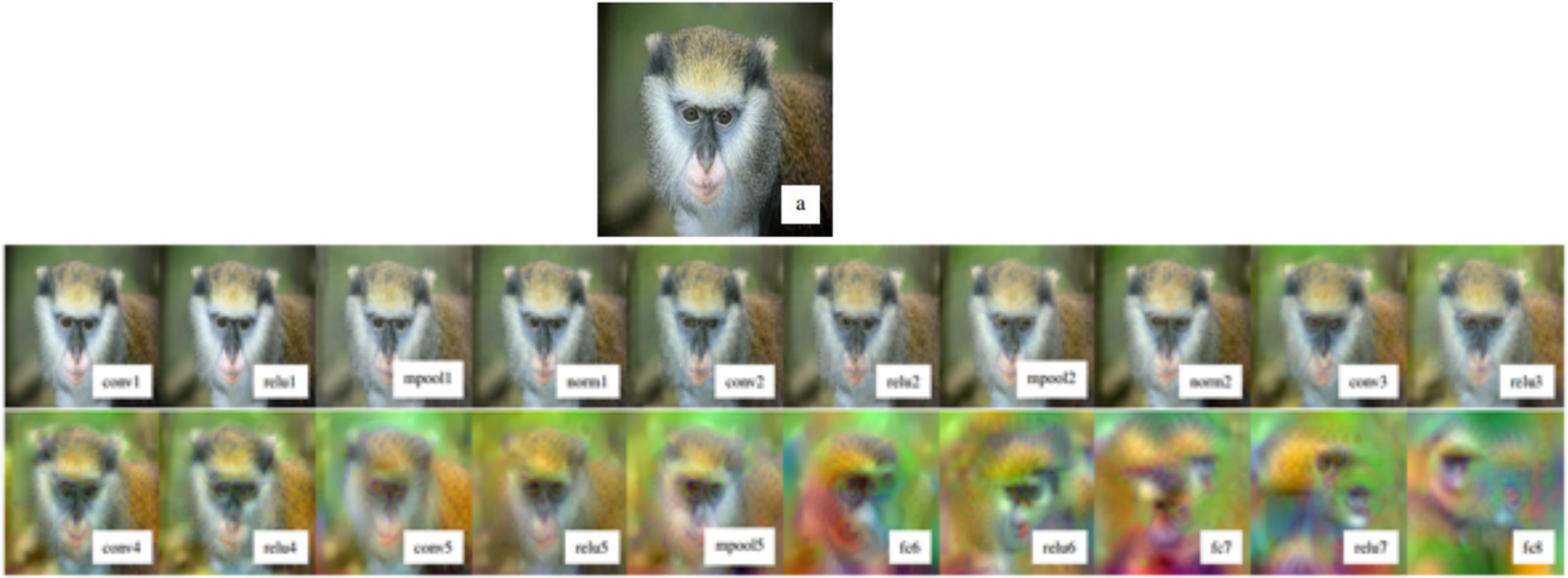


$$x^* = \underset{x \in \mathbb{R}^{H \times W \times C}}{\operatorname{argmin}} L(\varphi(x), \varphi_0) + \lambda R(x) \qquad (2)$$

Aravindh Mahendran and Andrea Vedaldi, Understanding Deep Image Representations by Inverting Them, CVPR’15

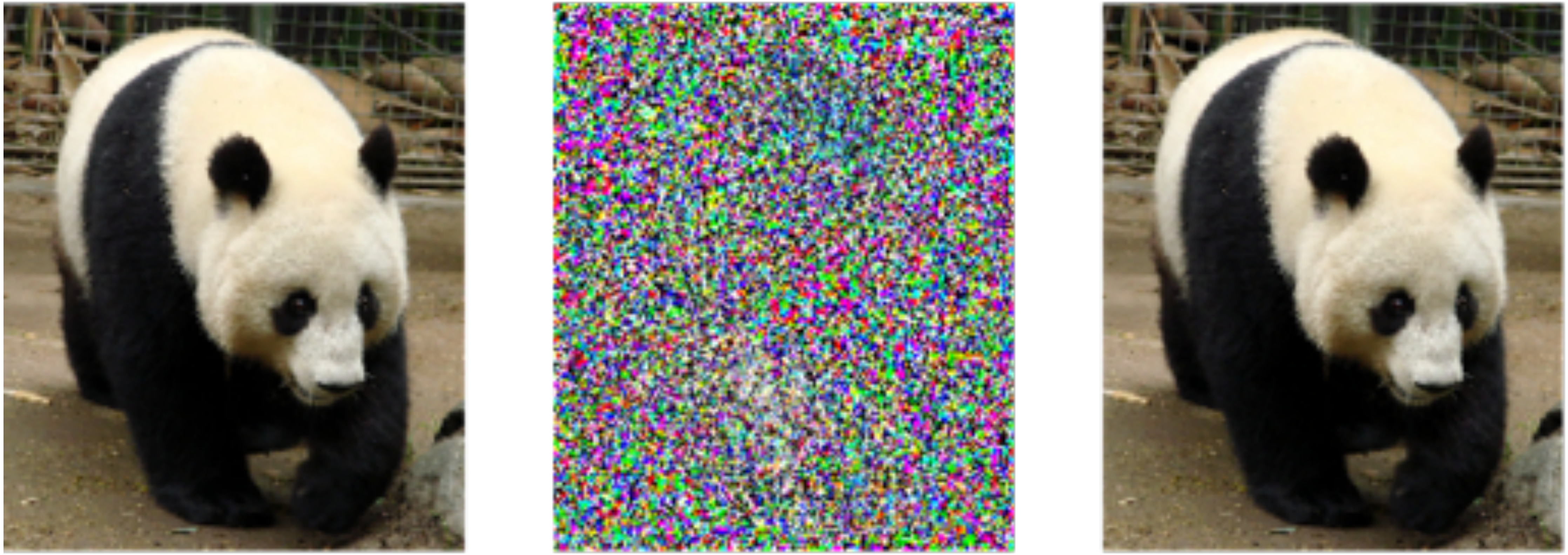
Inverting at Different Stages

Inverting at Different Stages



Reconstructions from intermediate layers

Fooling CNN's

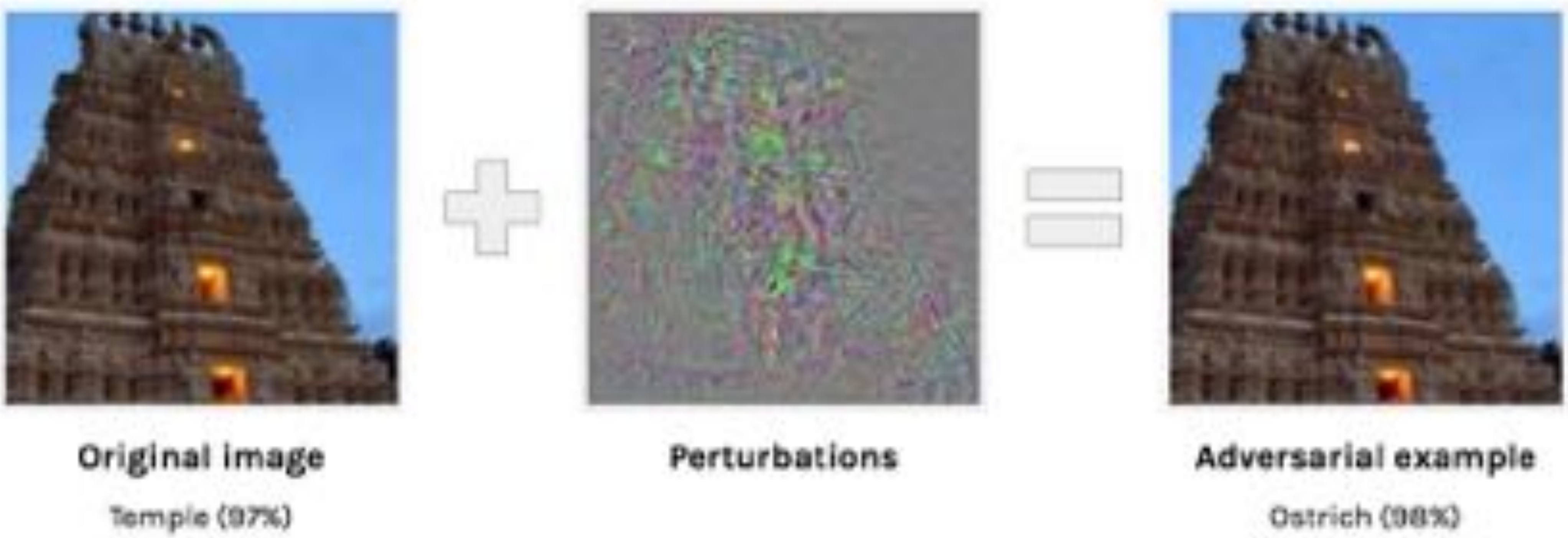


original image
prediction: giant_panda

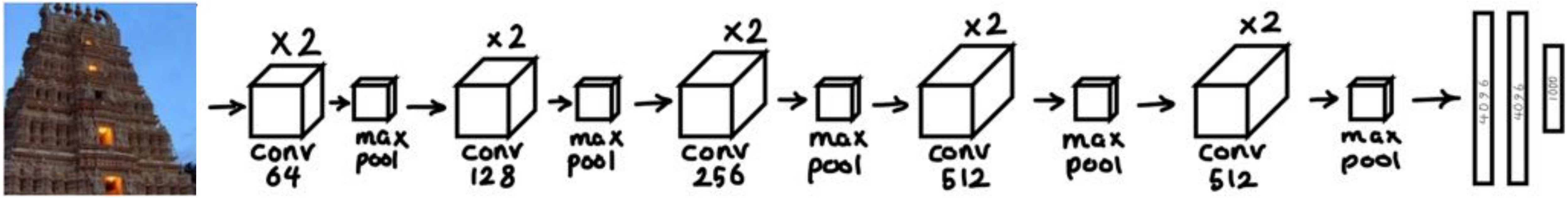
the perturbation,
enhanced 127 times

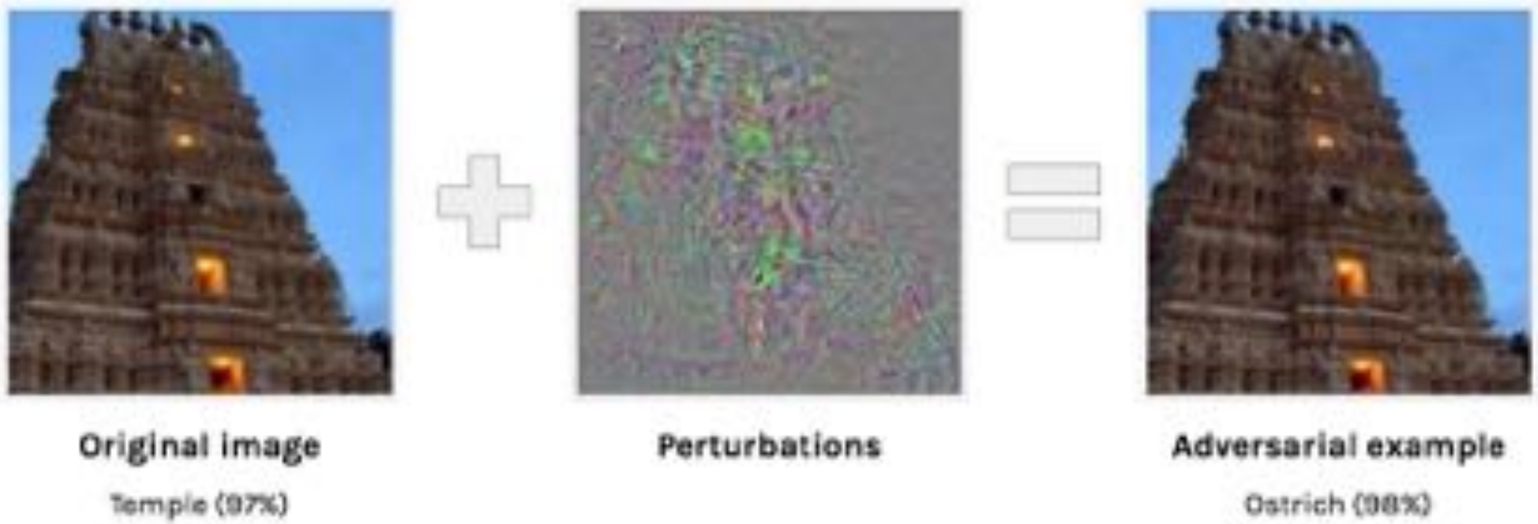
perturbed image
prediction: bucket

Fooling CNN's

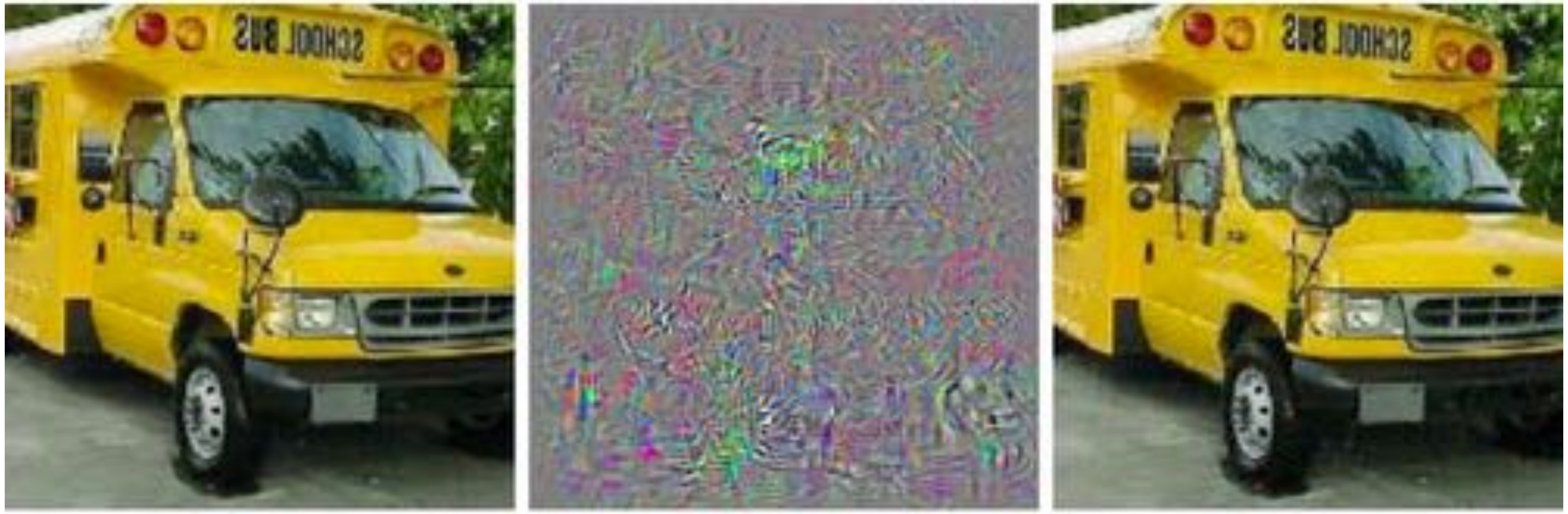


Fooling CNN's



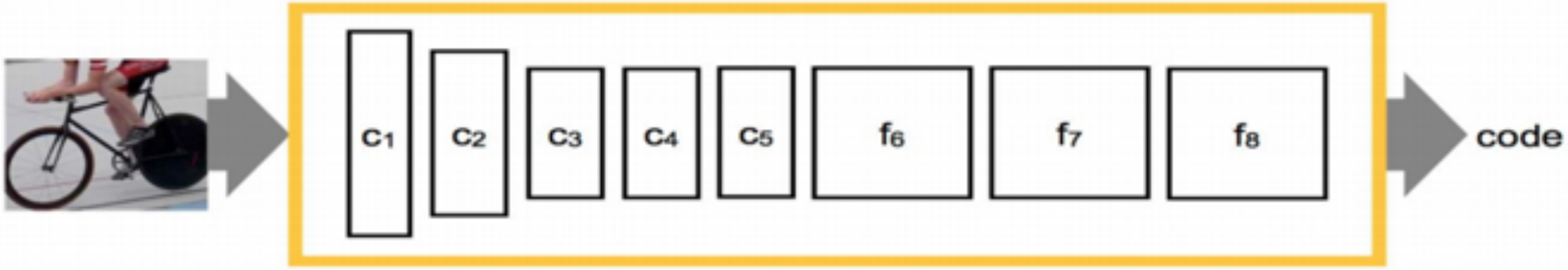


Fooling CNN’s



“bus” adversarial noise “ostrich”

Learned Representations



Learned Representations

CNN Features can be used for wider applications:

- Train the CNN (deep network) on a very large database such as imageNet.

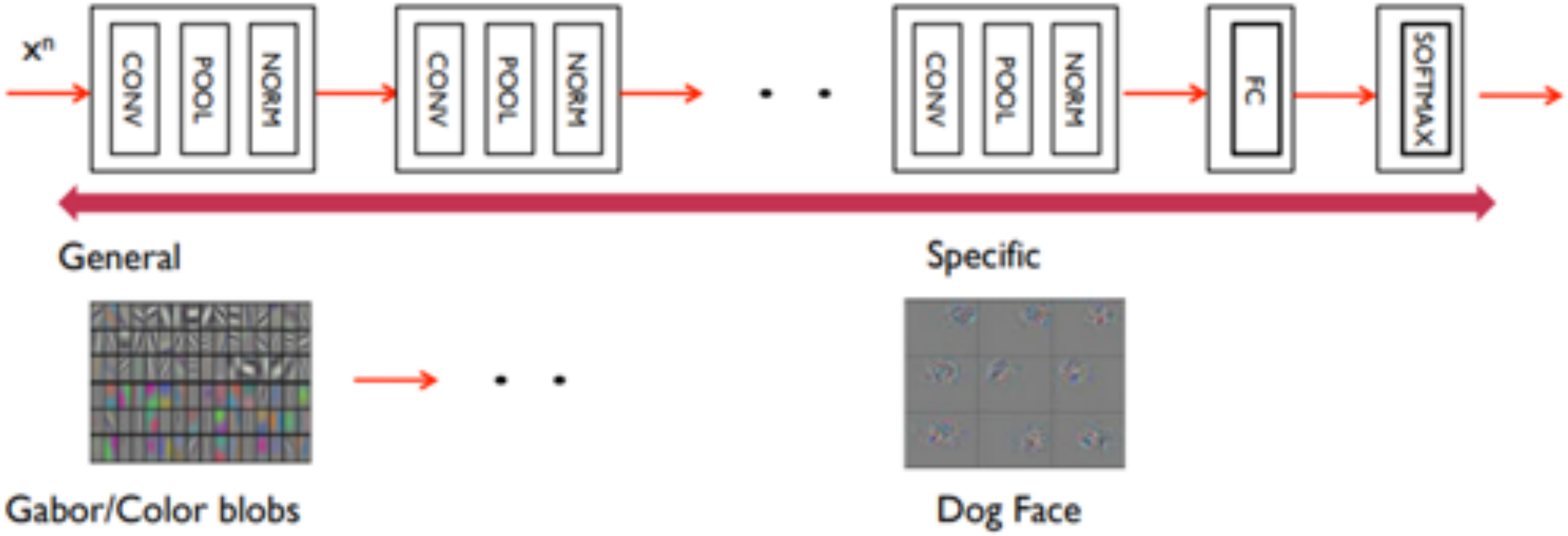
- Re-use CNN to solve smaller problems
  – Remove the last layer (classification layer)
  – Output is the code/feature representation

New Settings

- Extend to more classes
  - Extend from 1000 classes (say people) to another new 100
- Extend to new tasks
  - Extend from object classification to scene classification
- Extend to new data sets
  - Extend from imageNet to PASCAL (SLR to webcams)
- When we have a lesser amount of data.

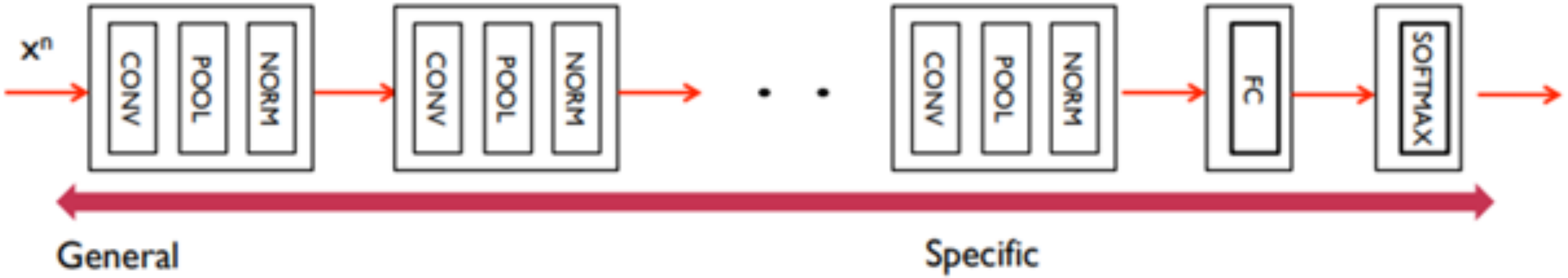
Transfer Learning

- **A key observation that we noticed in visualization:-**



Transfer Learning

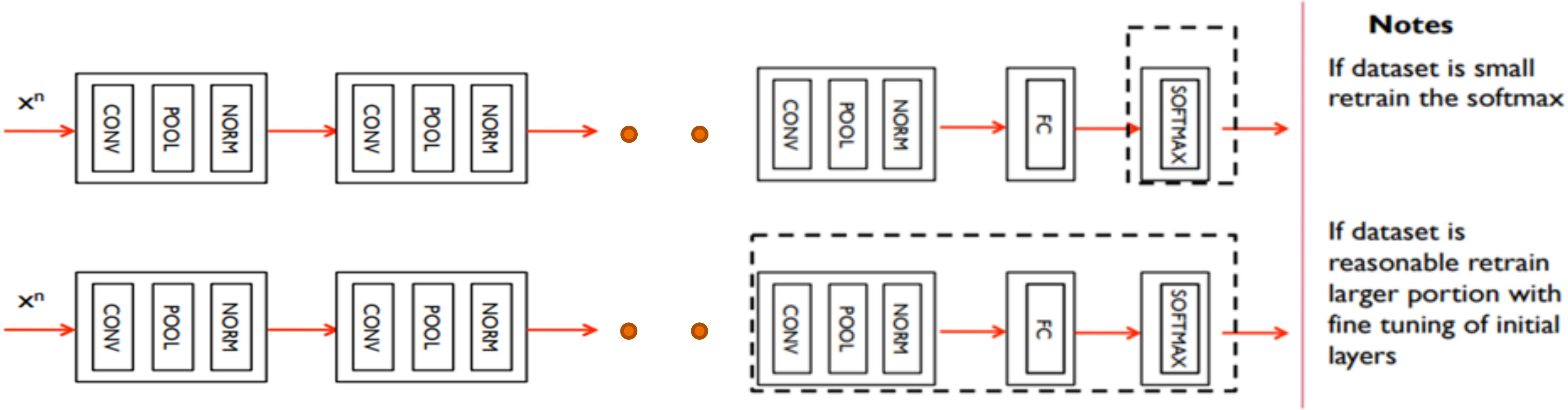
- **A key observation that we noticed in visualization:-**



- Further questions?
  – Can we quantify the layer generality/specificity ?
  – Where does the transition occur?
  – Is the transition sudden or spread over layers?

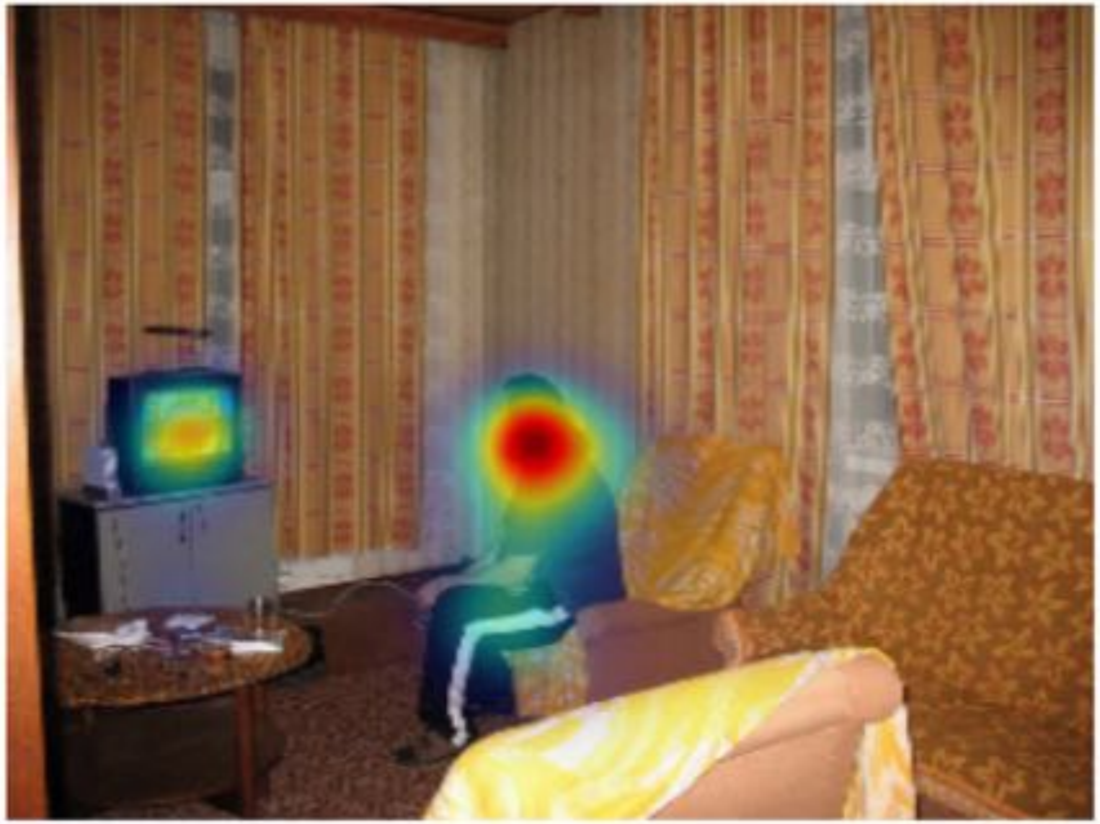
Transfer Learning

- Take away message



- Initializing a network with transferred features almost always gives better generalization

Transfer Learning - examples



Transfer Learning - examples