ConvNet Case Study I: Alexnet

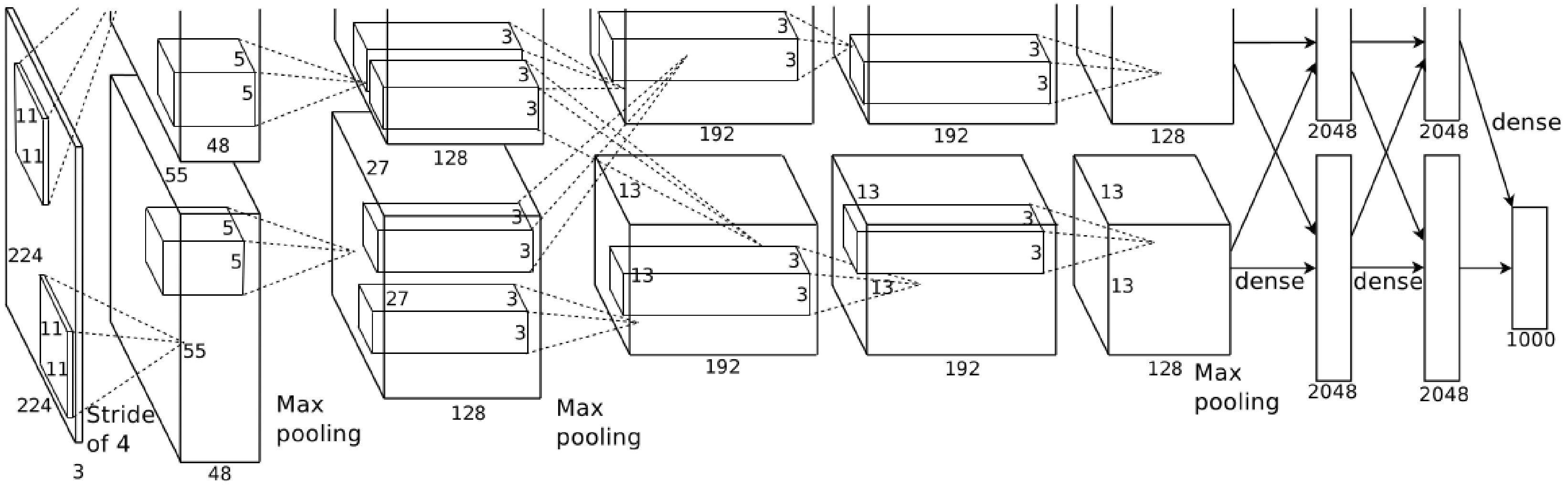
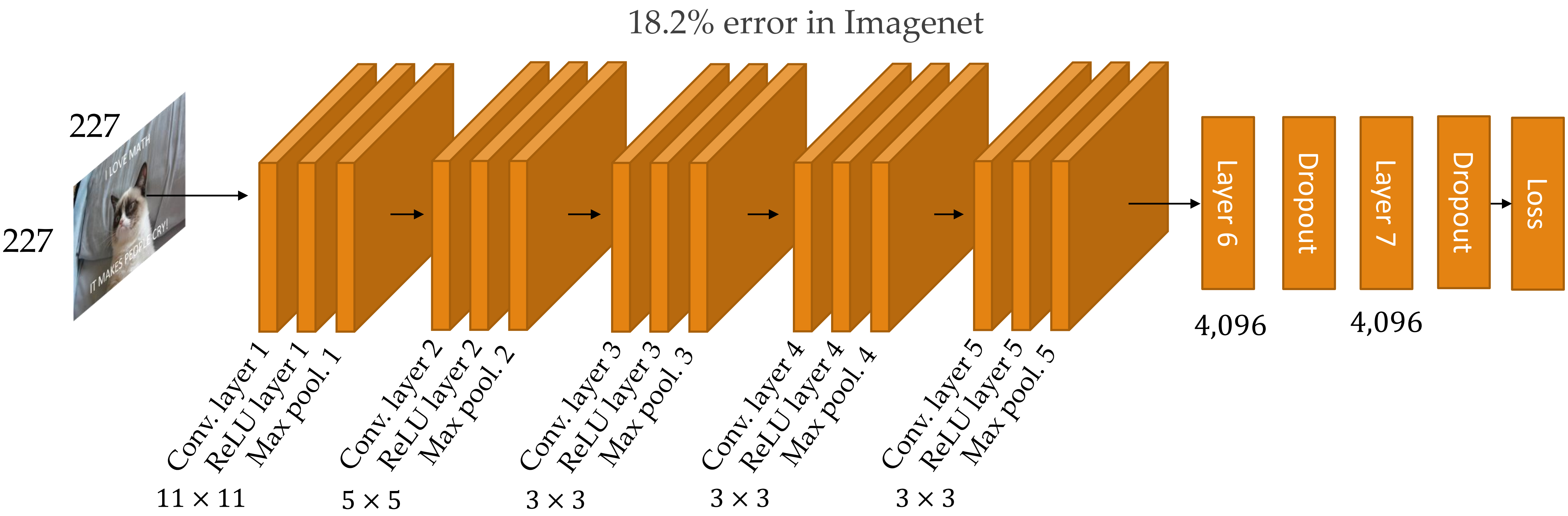
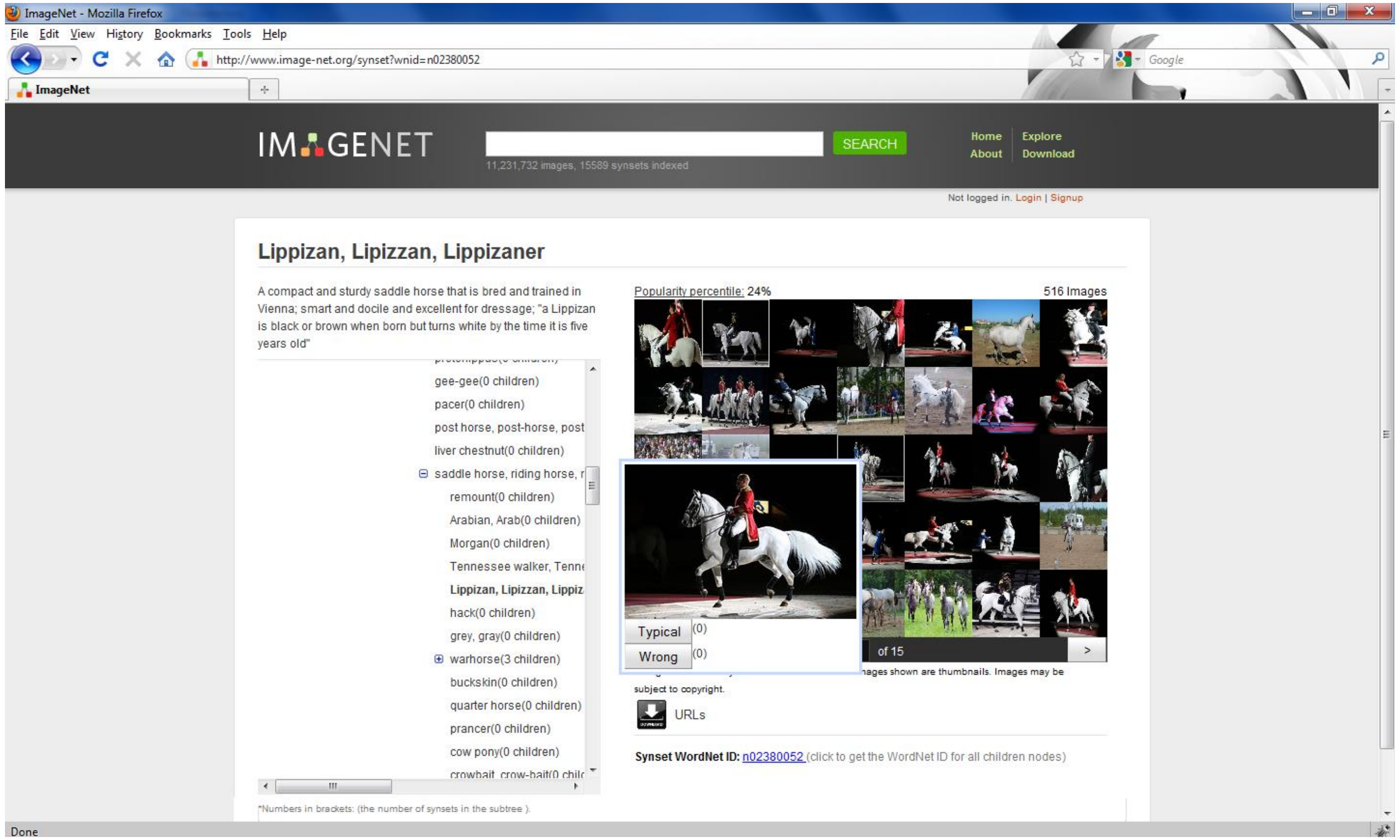


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

Architectural details



IMAGENET



http://www.image-net.org

Constructing ImageNet

Step 1:
Collect candidate images
via the Internet

Step 2:
Clean up the candidate
Images by humans

YAHOO! picsearch flickr Live Search Google Image Search

amazon mechanical turk
Artificial Artificial Intelligence

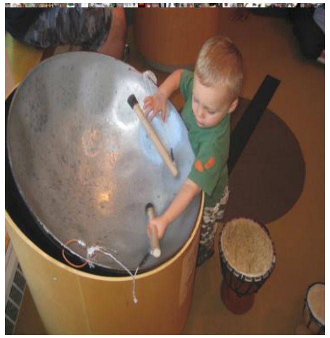
Some statistics

- July 2008: 0 images
- Dec 2008: 3 million images, 6K+ synsets
- April 2010: 11 million images, 15K+ synsets
- Currently: 14 million images, 21K synsets indexed

ImageNet Large Scale Visual Recognition Challenge

- Ran from 2010 to 2017
  - Today a Kaggle competition

- Main task: image classification
  - Automatically label 1.4M images with 1K objects
  - Measure top-5 classification error



Output		**Output**	
Scale		Scale	
T-shirt		T-shirt	
Steel drum	✓	Giant panda	✗
Drumstick		Drumstick	
Mud turtle		Mud turtle	

Deep learning at ImageNet classification challenge

CNN based, non-CNN based

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

Figures from Y. LeCun’s CVPR 2015 plenary talk

Deep learning at ImageNet classification challenge

CNN based, non-CNN based

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

Figures from Y. LeCun's CVPR 2015 plenary talk

Deep learning at ImageNet classification challenge

CNN based, non-CNN based

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

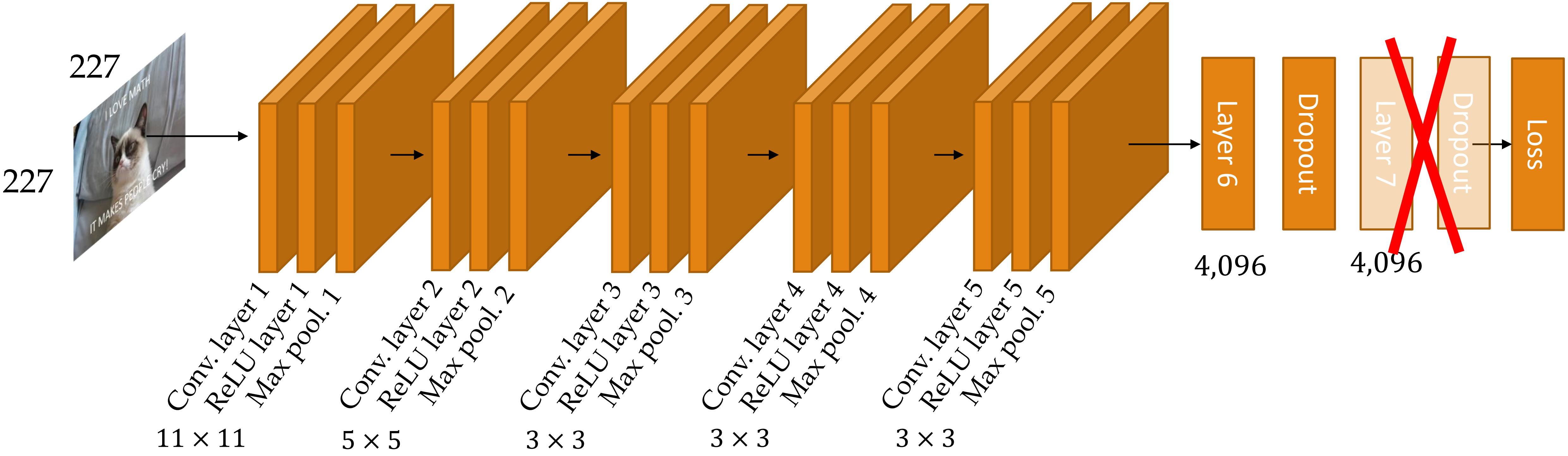
2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

2014 Teams	%error
GoogLeNet	6.6
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	12.1

Figures from Y. LeCun’s CVPR 2015 plenary talk

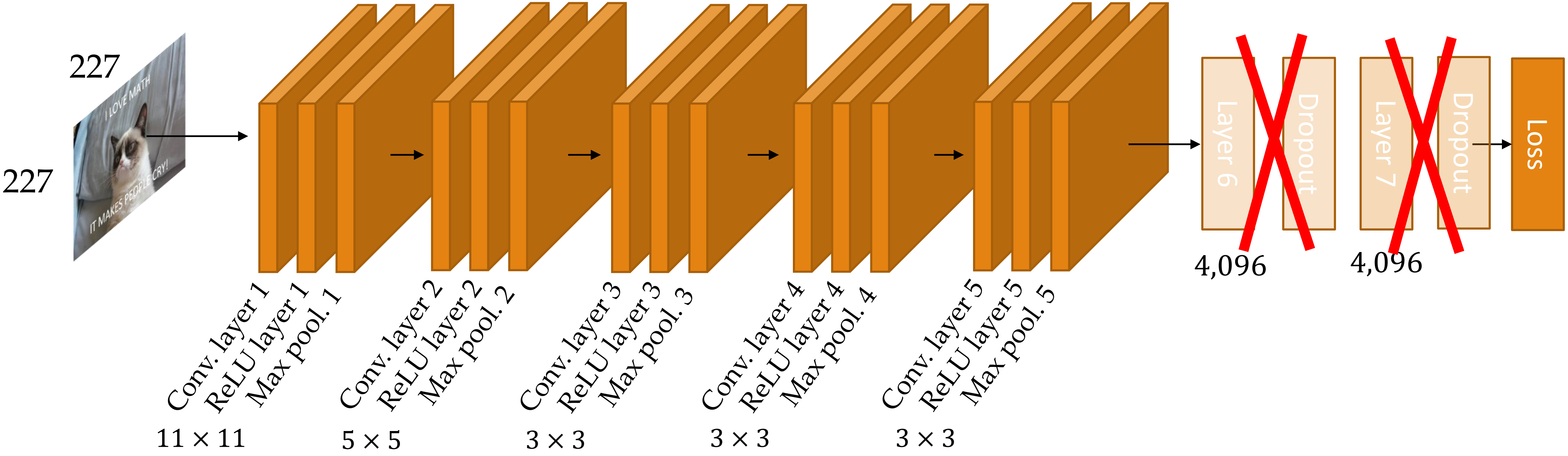
Removing layer 7

1.1% drop in performance, 16 million less parameters



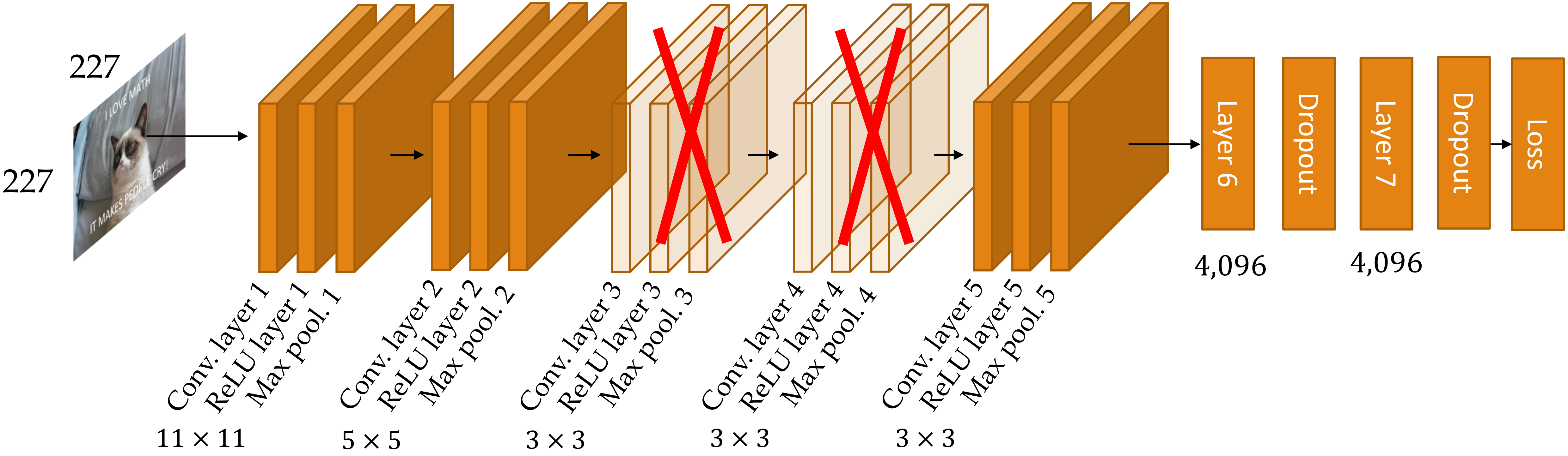
Removing layer 6, 7

5.7% drop in performance, 50 million less parameters

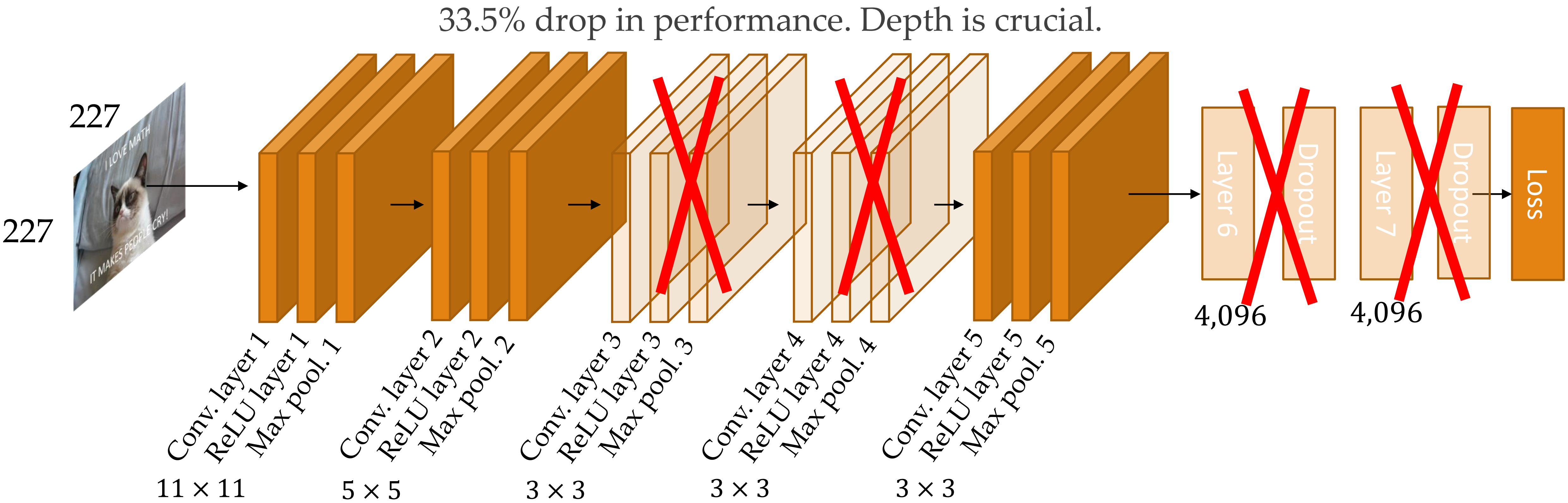


Removing layer 3, 4

3.0% drop in performance, 1 million less parameters. Why?

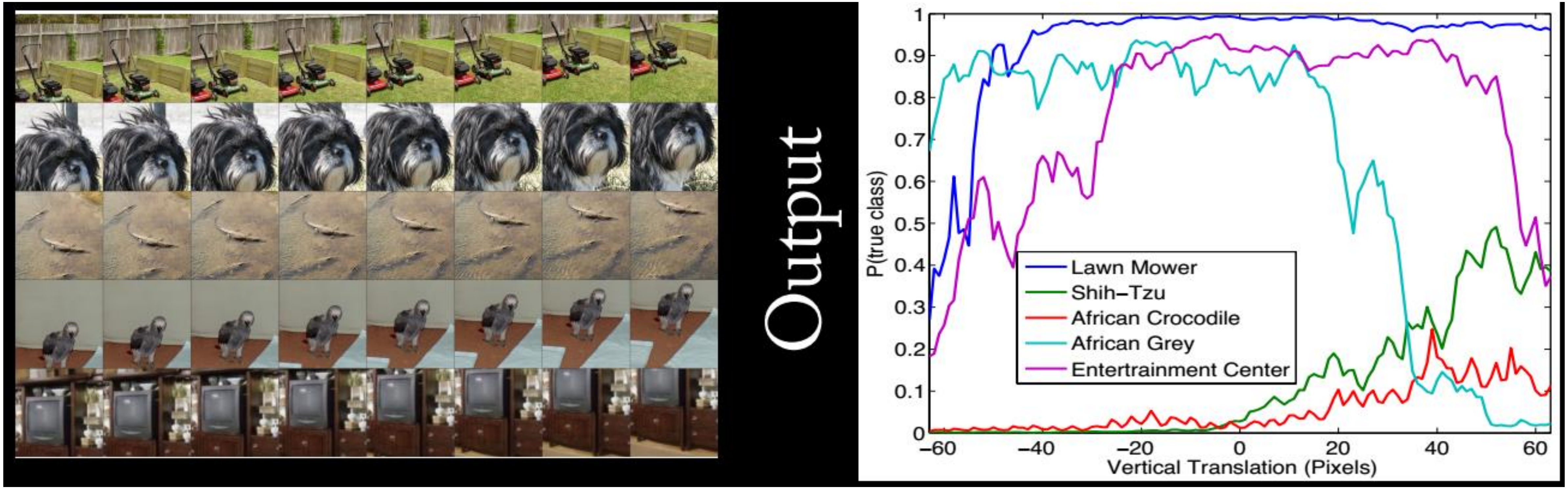


Removing layer 3, 4, 6, 7



Translation invariance

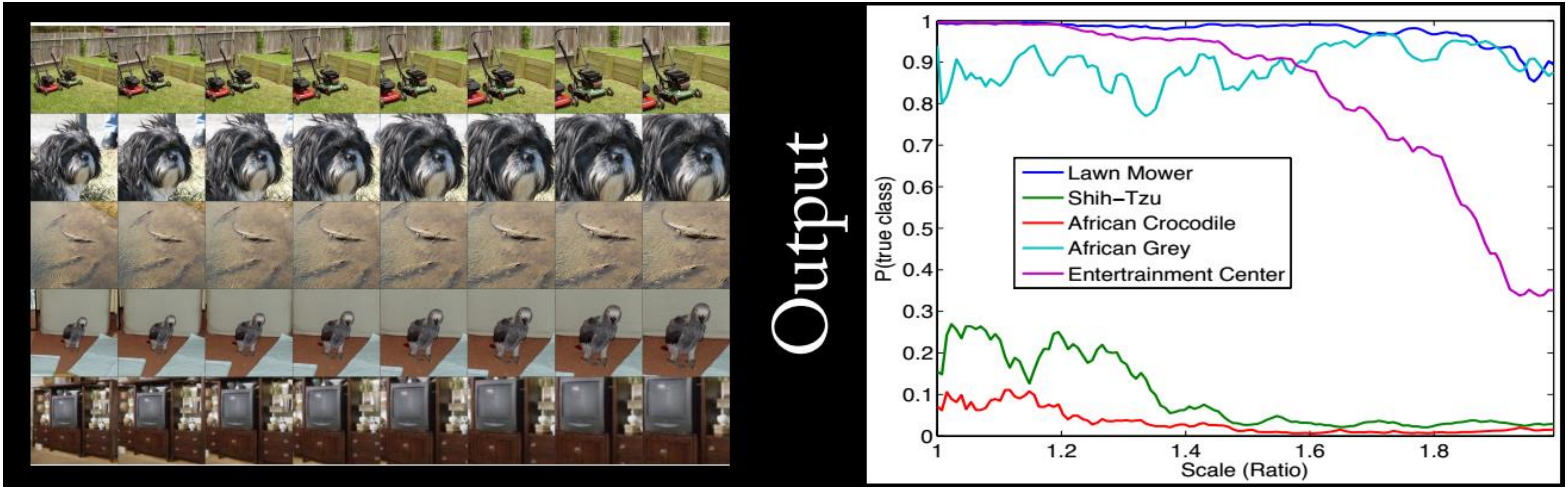
- CNNs are translation invariant



Credit: R. Fergus slides in Deep Learning Summer School 2016

Scale invariance

- CNNs are scale invariant to some degree
  - The standard convolutional filters not scale invariant
  - Scale invariance learnt depends on scale variations present in data



Rotation invariance

- CNNs are not rotation invariant
  - The standard convolutional filters not rotation invariant
  - And only few rotated examples in the training set. Augmentation can help