Image Classification, Representation Learning, and Fine-Tuning

Danna Gurari
University of Colorado Boulder
Spring 2024

https://dannagurari.colorado.edu/course/neural-networks-and-deep-learning-spring-2024/

Review

- Last lecture:
  - Computer vision
  - Era of dataset challenges
  - MNIST challenge winner: LeNet
  - Pioneering ImageNet challenge winner: AlexNet (deeper learning)
  - Programming tutorial

- Assignments (Canvas)
  - Lab assignment 2 due in 1 week

- Questions?

Today's Topics

- ImageNet challenge winners: going deeper
- Representation learning
- Fine-tuning
- Programming tutorial

Today's Topics

- ImageNet challenge winners: going deeper
- Representation learning
- Fine-tuning
- Programming tutorial

Recall: Rise of “Deep Learning”

Progress of models on ImageNet (Top 5 Error)

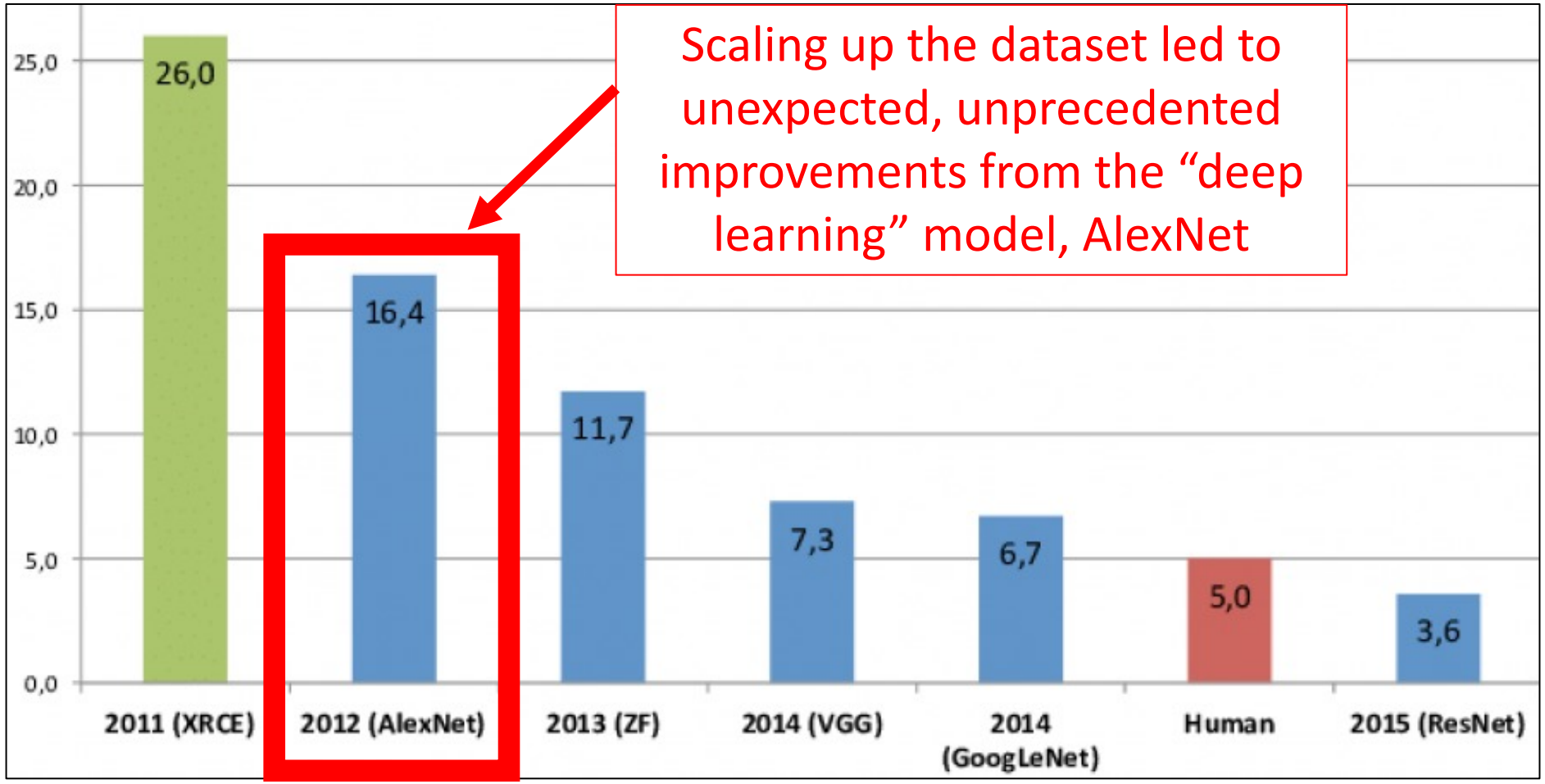
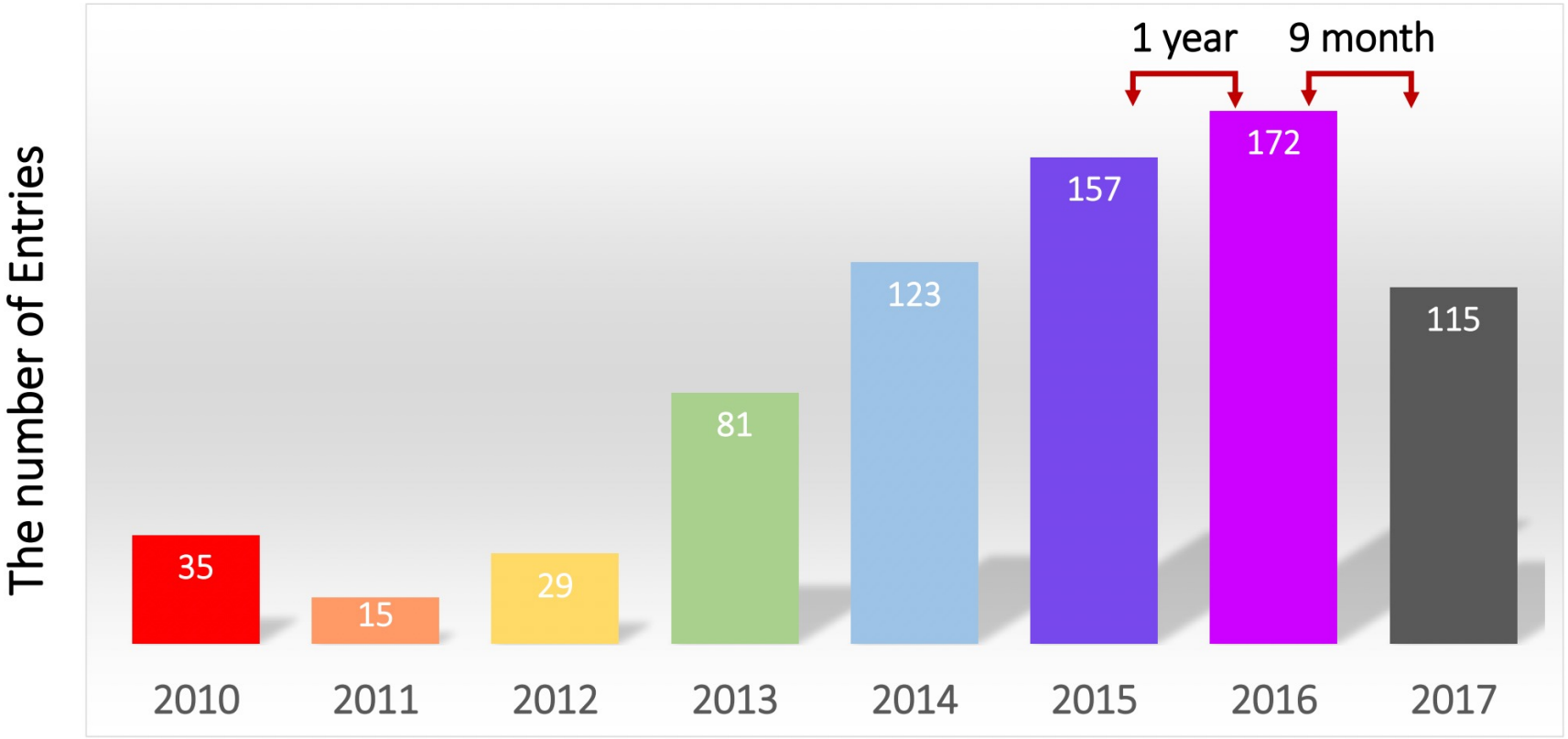


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

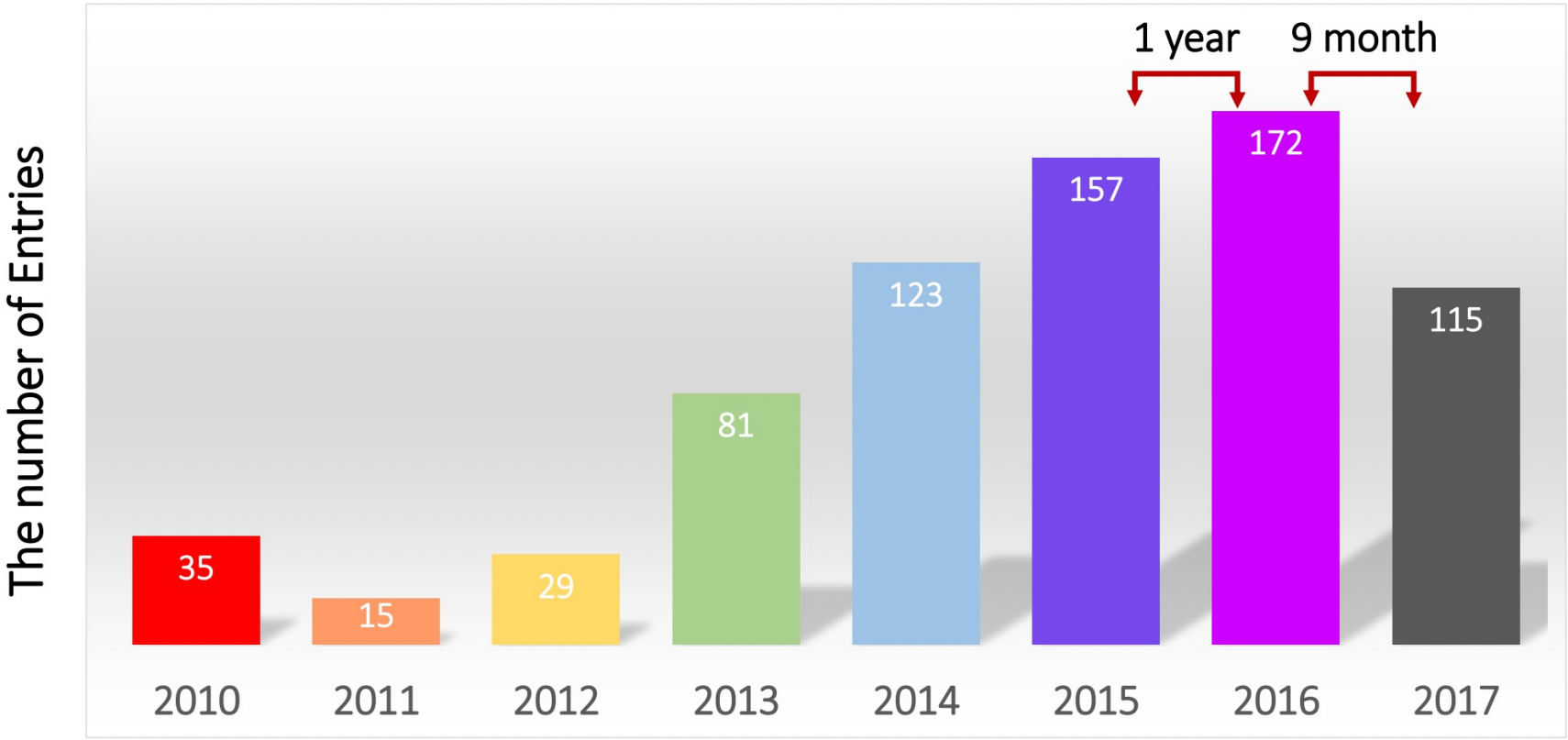
Rise of “Deep Learning” Following AlexNet



Inspired by AlexNet, many more researchers in the computer vision community proposed neural networks and showed how to make further progress over the years!

Source: https://image-net.org/static_files/files/ILSVRC2017_overview.pdf

Rise of “Deep Learning” Following AlexNet



- 727 entries (plus an entry that famously was kicked out in 2015 for cheating from Baidu)
- Labor cost ~$110 million: assuming 3 people contribute to each entry and $50k cost per person

Source: https://image-net.org/static_files/files/ILSVRC2017_overview.pdf

Secret Sauce for State-of-Art: Deeper CNNs

Progress of models on ImageNet (Top 5 Error)

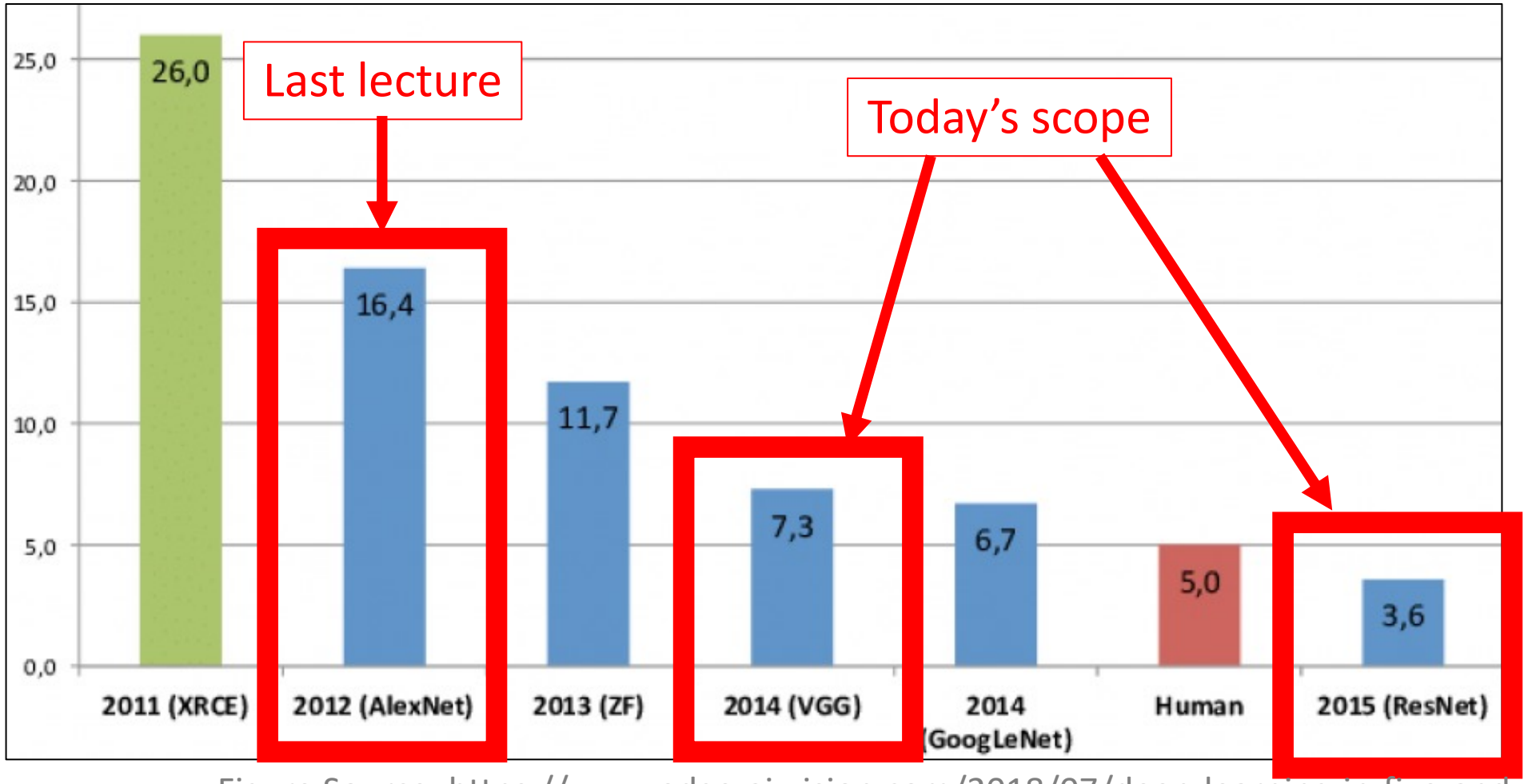


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

VGG: A Deeper CNN

Progress of models on ImageNet (Top 5 Error)

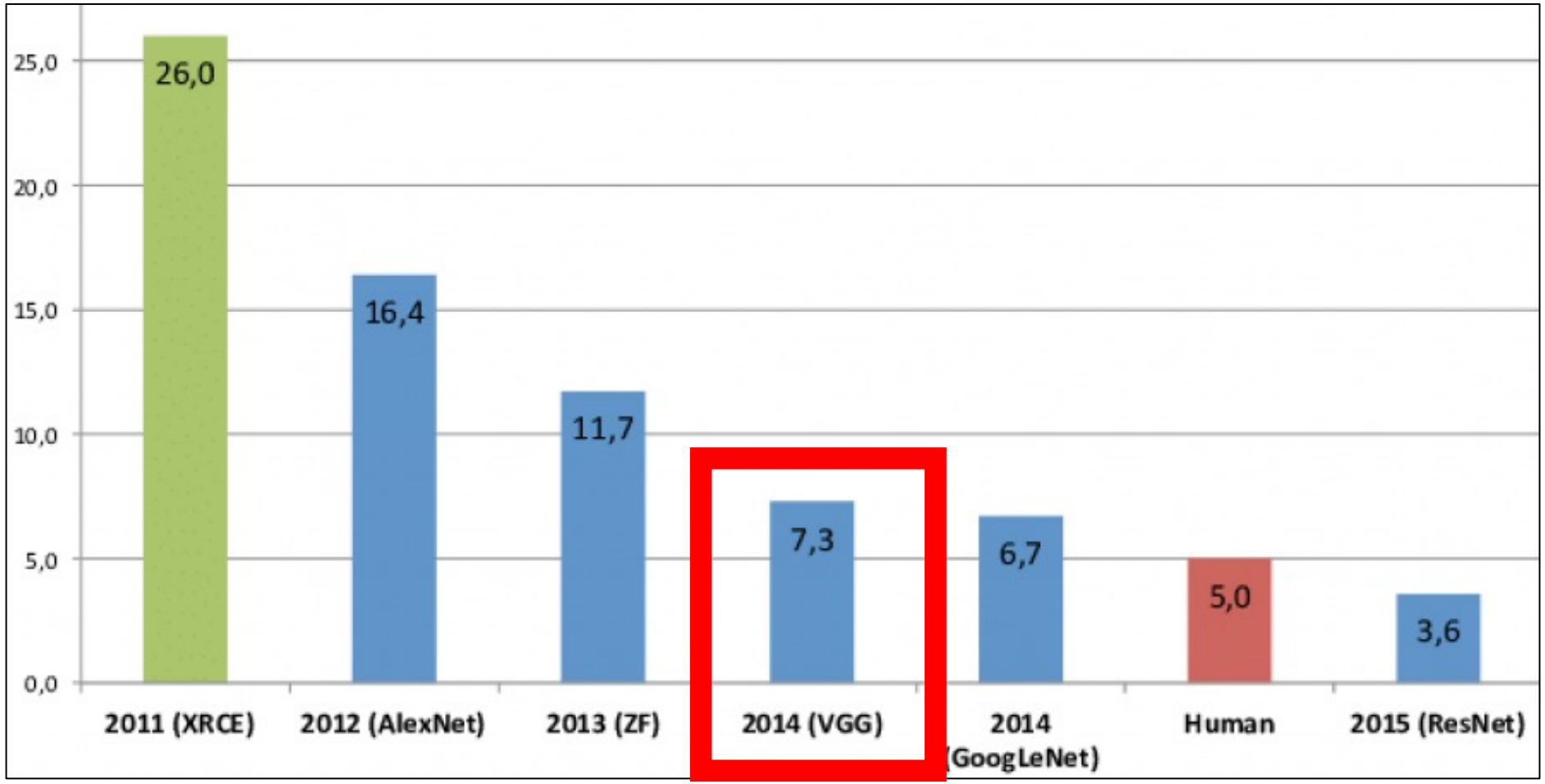
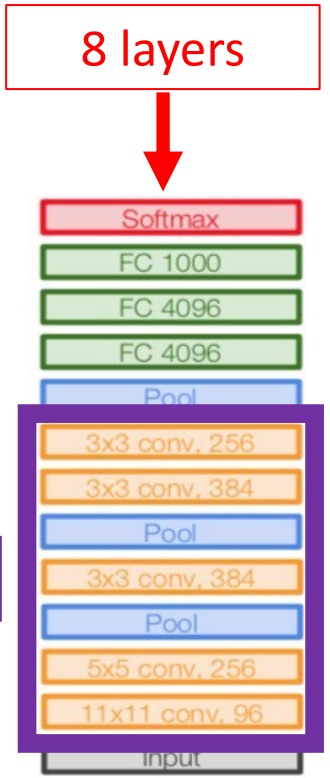


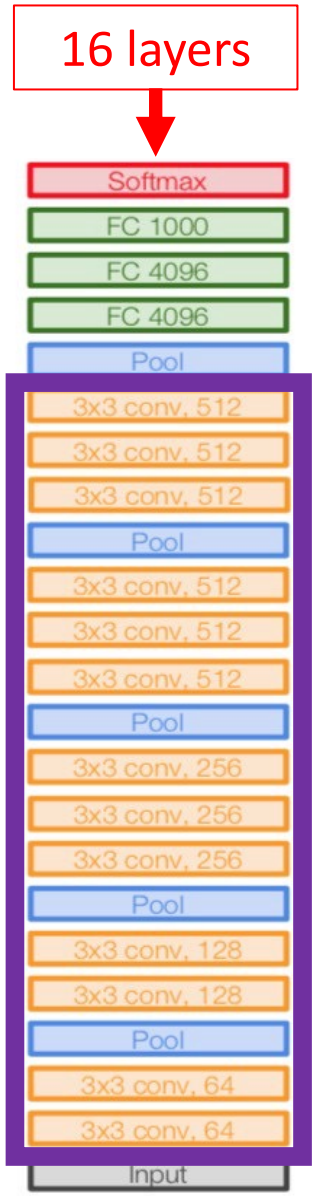
Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

Key Novelty: **Deeper** Does Better

** Number of layers with learnable model parameters between input and output layer (i.e., exclude pooling layers)*



Layers with differences



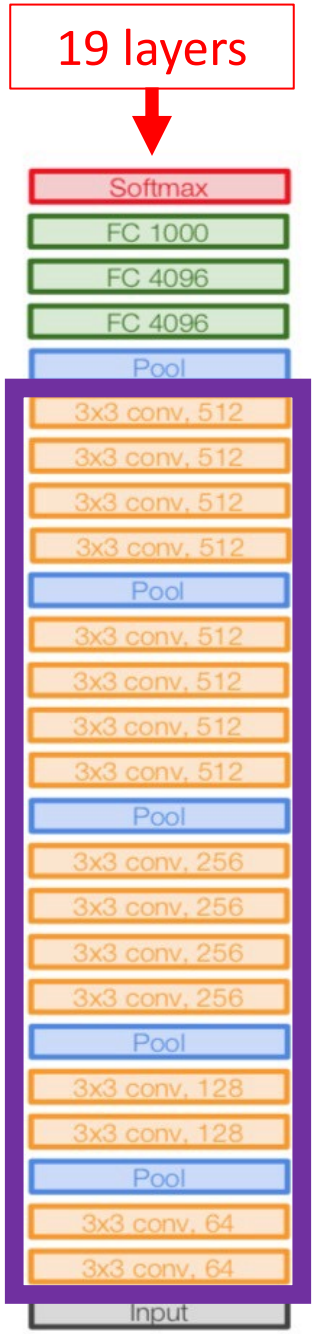


Figure Source (edited to fix mistakes): https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

Key Idea: Smaller Convolutional Filters

- Replace larger filter with stack of smaller filters

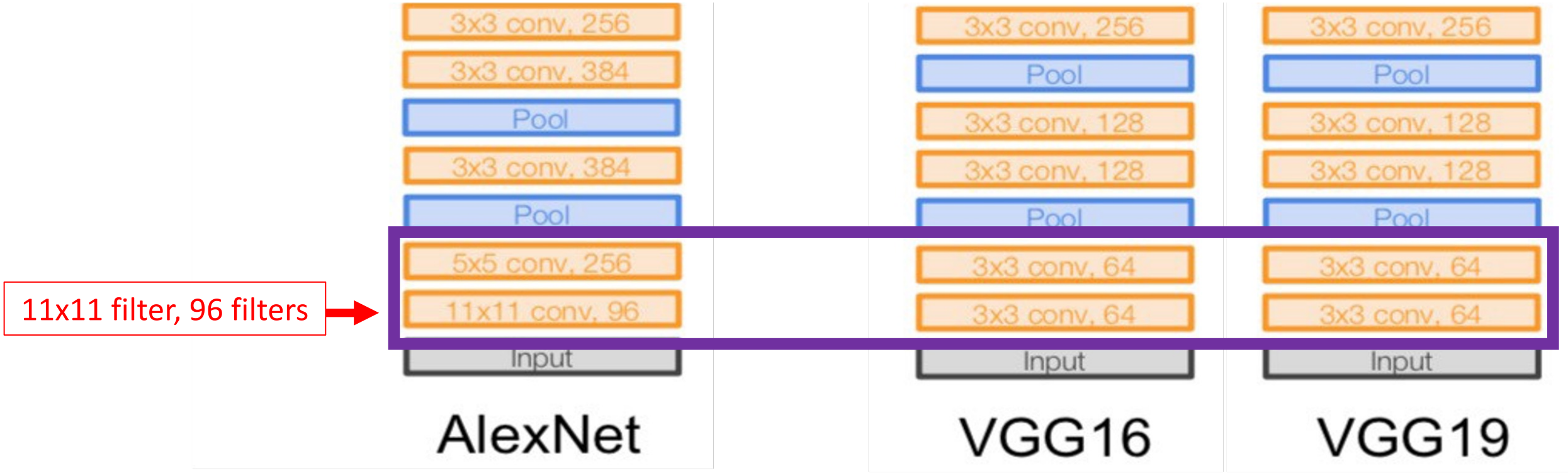
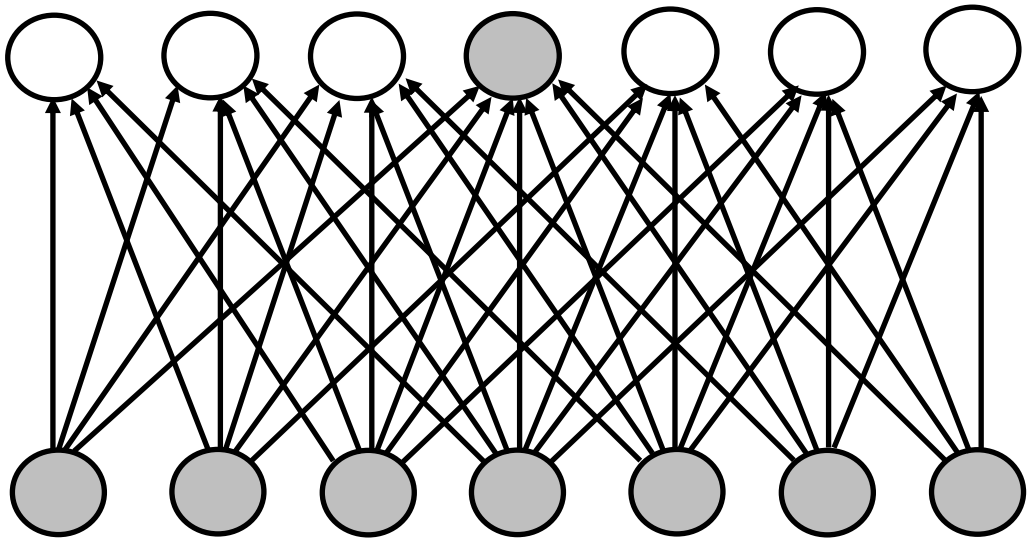
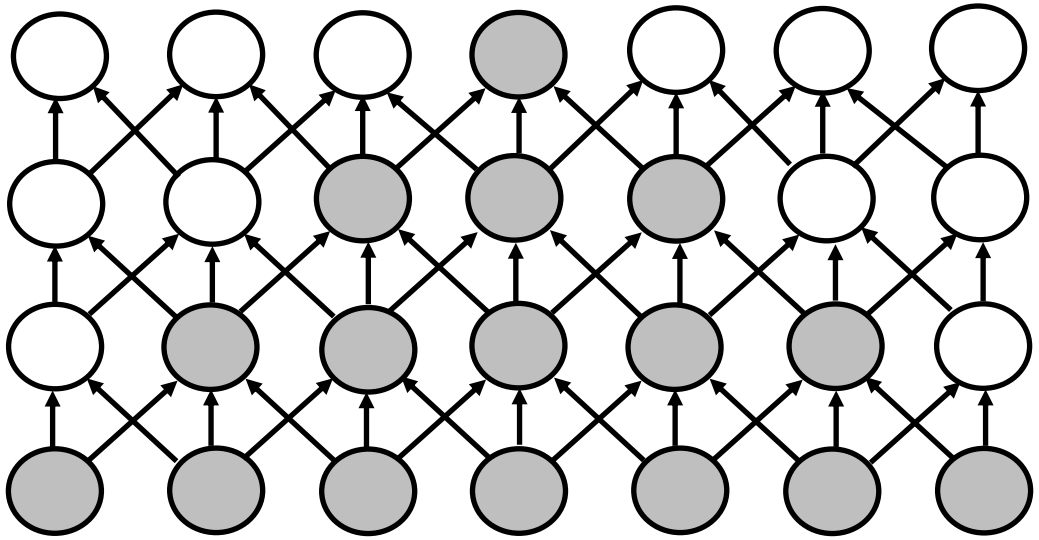


Figure Source (edited to fix mistakes): https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

Key Idea: Smaller Convolutional Filters

- Replace larger filter with stack of smaller filters; e.g., replace 7x7 with three 3x3s

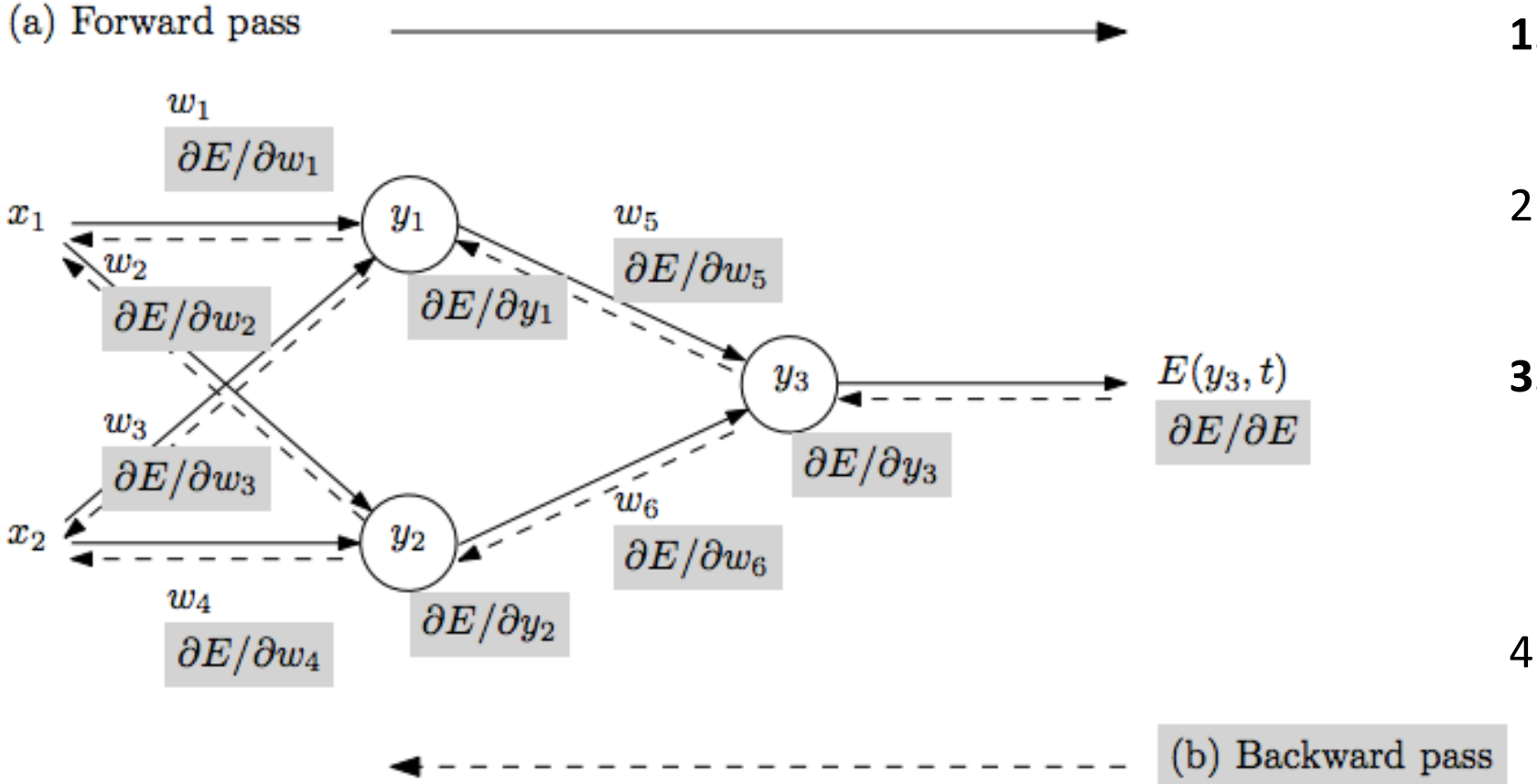




- Benefits:
  - More discriminative classifier since more non-linear rectifications: 3 vs 1
  - Reduces # of parameters: multiple of 27 (3×3^2) parameters vs 49 (7×7) parameters

https://www.deeplearningbook.org/contents/convnets.html

VGG Training (follows AlexNet): 74 Epochs



- Repeat until stopping criterion met:
  1. **Forward pass**: propagate training data through model to make prediction
  2. Quantify the dissatisfaction with a model's results on the training data
  3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each model parameter
  4. Update each parameter using calculated gradients

$$W_x = W_x - a\left(\frac{\partial Error}{\partial W_x}\right)$$

Figure from: Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, Jeffrey Mark Siskind; Automatic Differentiation in Machine Learning: a Survey; 2018

VGG Limitation: Models Are Large!

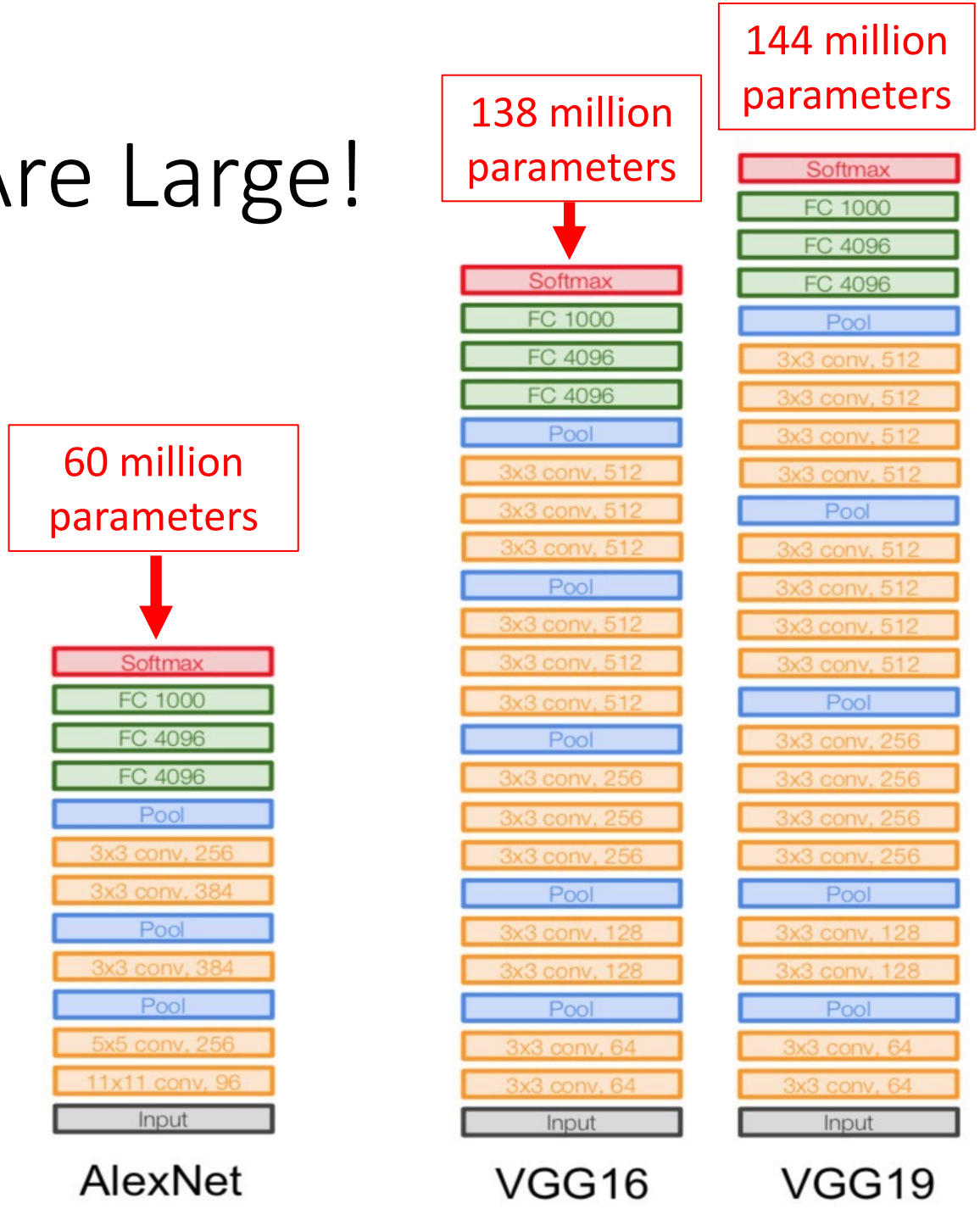
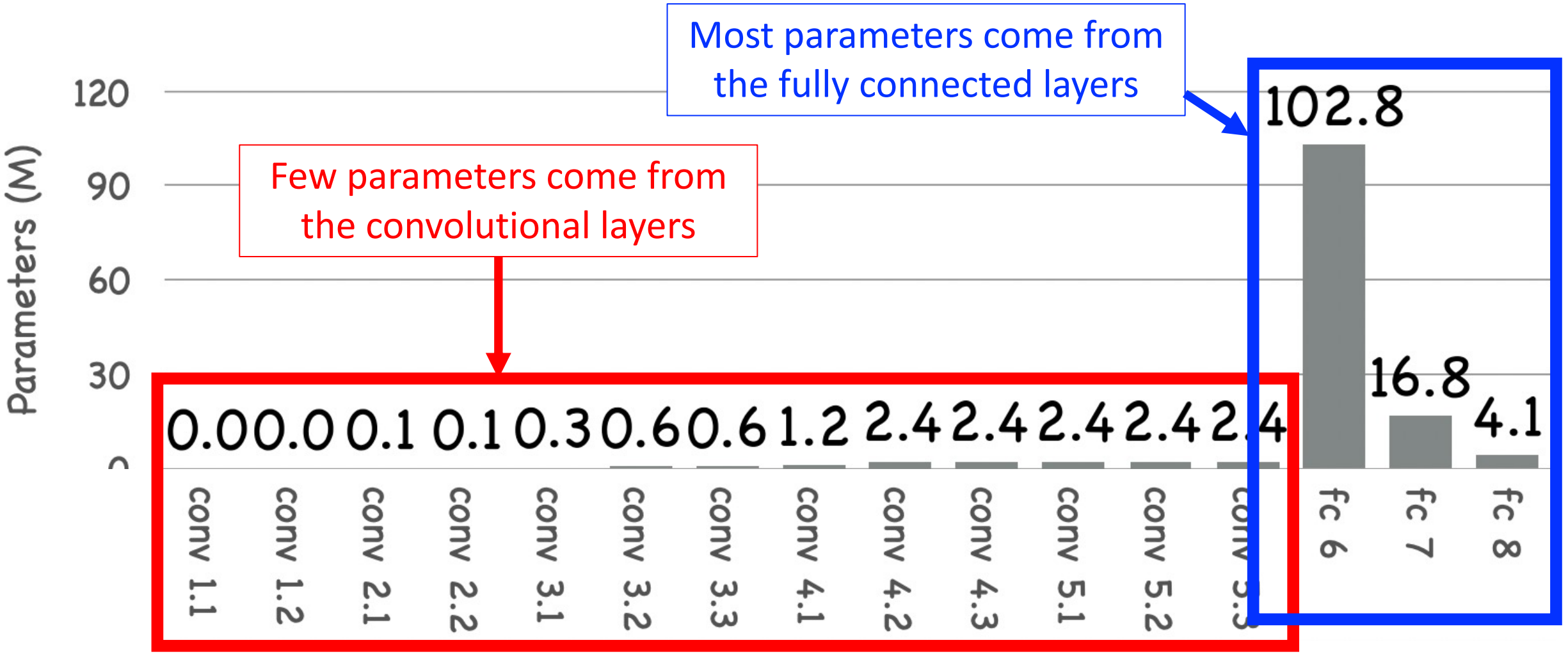


Figure Source (edited to fix mistakes): https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

VGG Limitation: Models Are Large (e.g., VGG16)



Source: http://www.philkr.net/cs342/lectures/computer_vision/03.pdf

VGG: Key Tricks for Going Deeper

- 3x3 filters instead of larger filters

- Regularization techniques
  1. Data augmentation
  2. Dropout in fully connected layers
  3. L2 parameter norm penalty

- Trained across multiple GPUs

ResNet: A Deeper CNN

Progress of models on ImageNet (Top 5 Error)

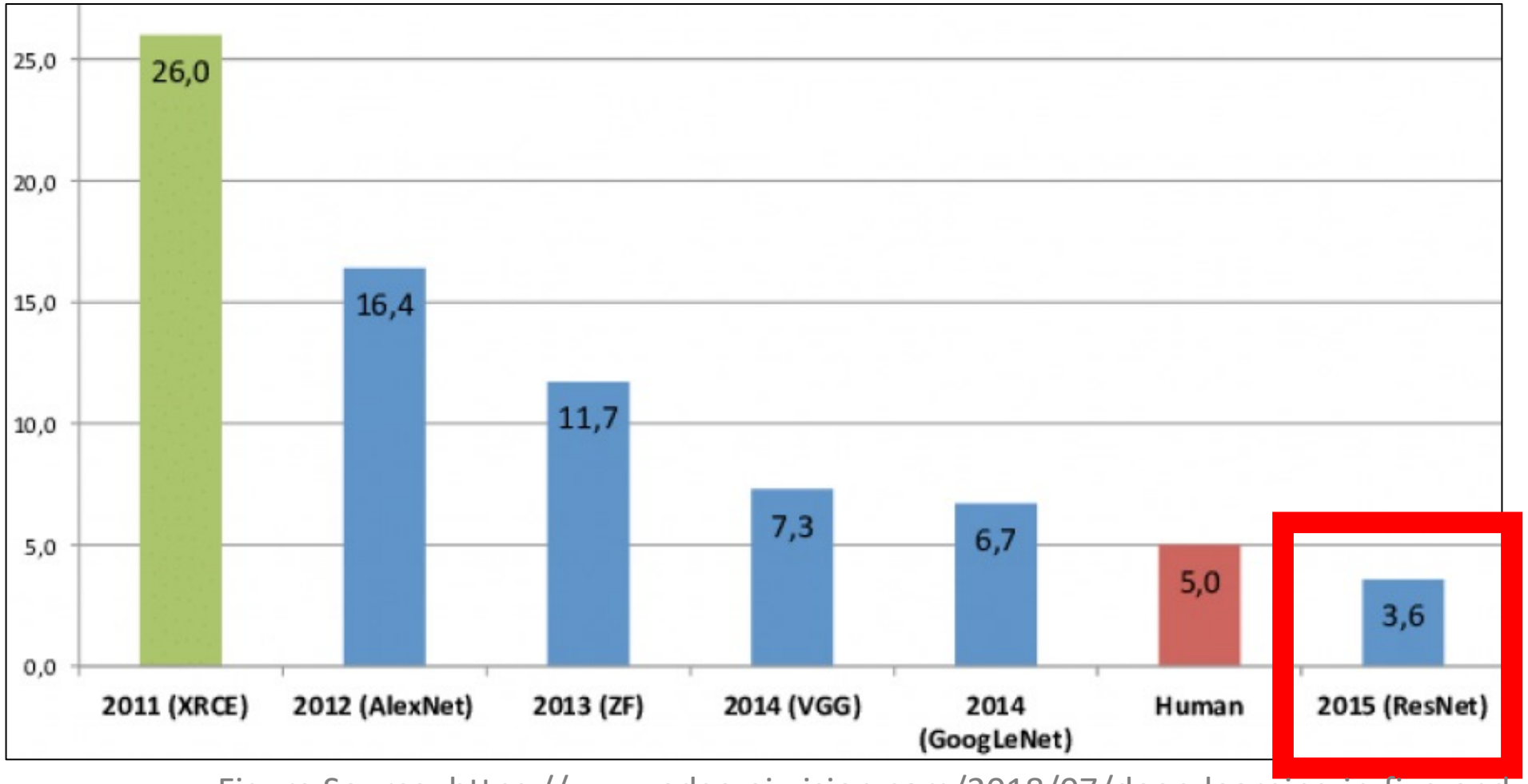
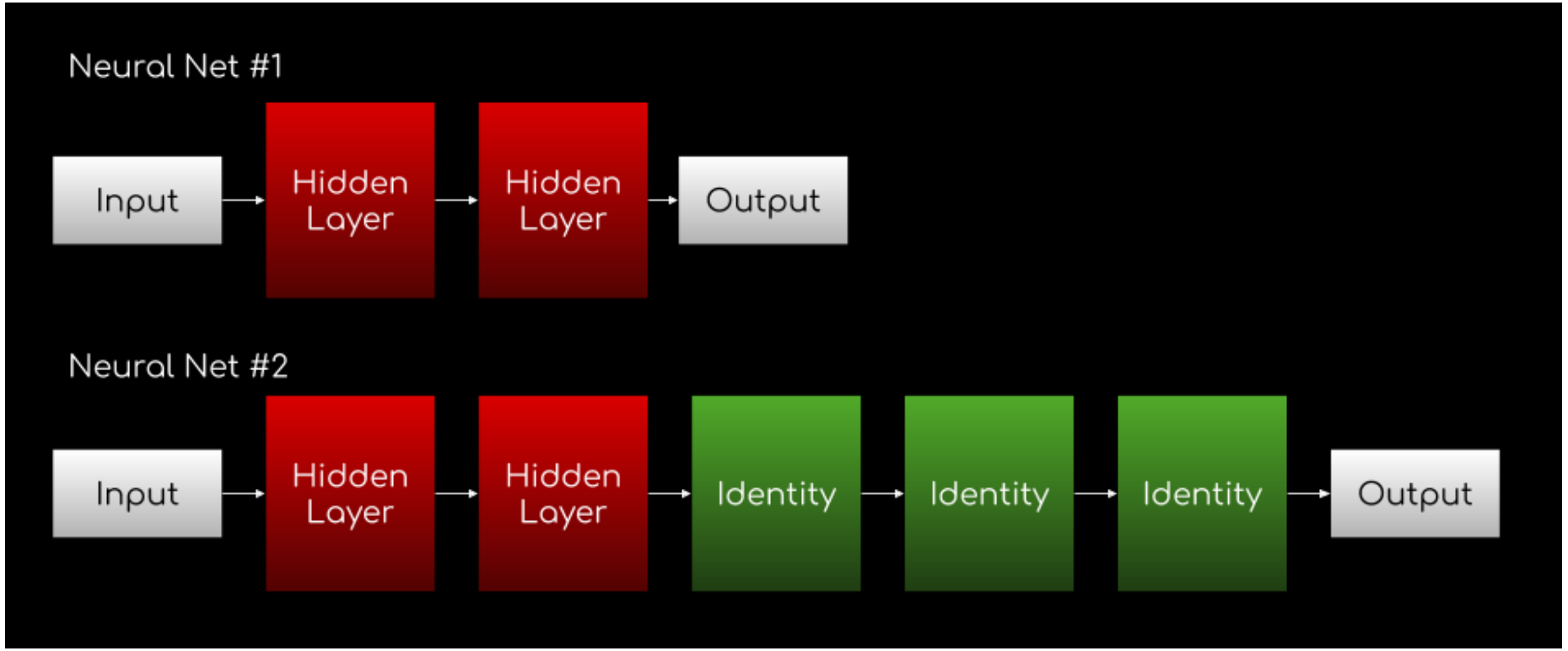


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

Motivating Observation

Idea: a deeper network should perform as good if not better than shallower networks since they can learn the shallower function by simply learning “identity” functions for later layers
Observation: adding more layers leads to WORSE results!
Is the problem overfitting?



Source: https://medium.com/@realmichaelye/intuition-for-resnet-deep-residual-learning-for-image-recognition-39d24d173e78

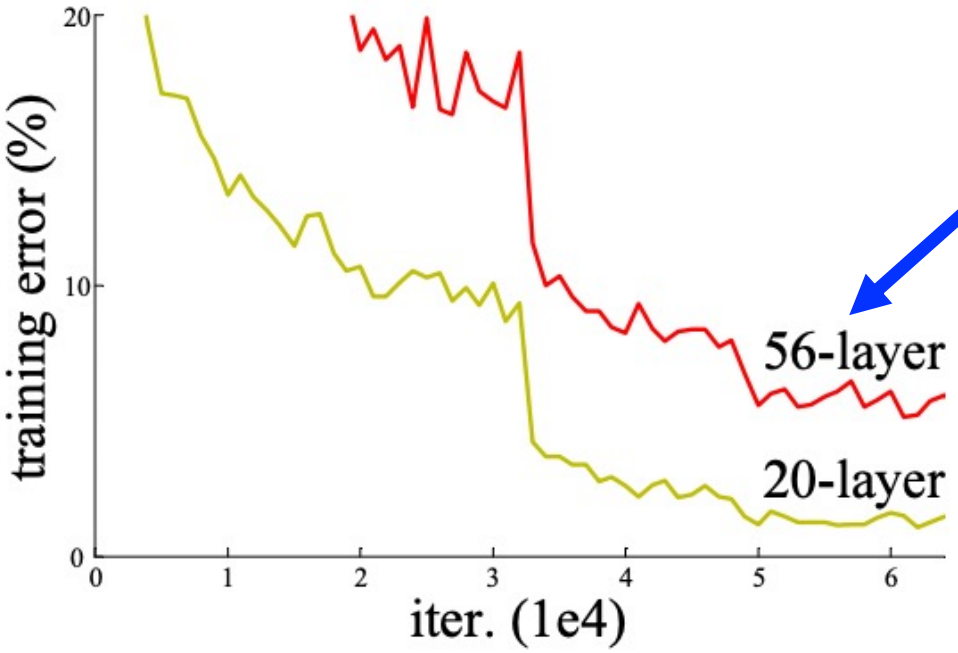
Motivating Observation

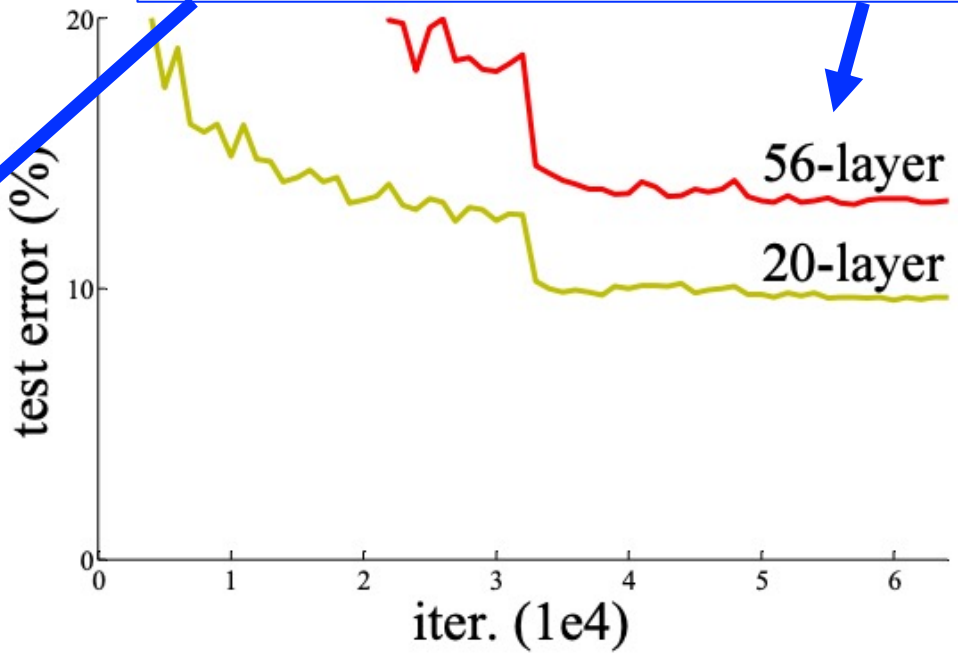
Idea: a deeper network should perform as good if not better than shallower networks since they can learn the shallower function by simply learning "identity" functions for later layers

Observation: adding more layers leads to WORSE results!

Is the problem overfitting? NO

Training data error (and test error) is greater with more layers





Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." CVPR, 2016.

Motivating Observation

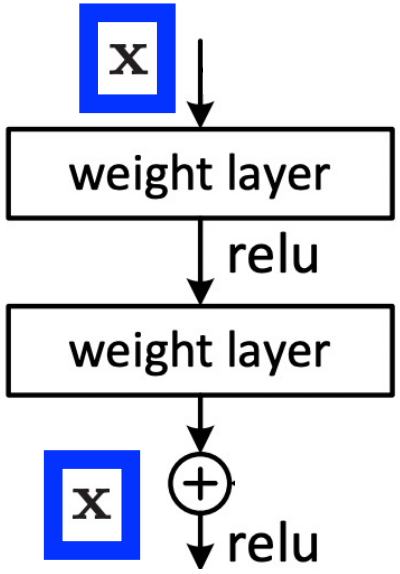
Idea: a deeper network should perform as good if not better than shallower networks since they can learn the shallower function by simply learning “identity” functions for later layers
Observation: adding more layers leads to WORSE results!
Is the problem overfitting? NO

Problem: It is difficult for the algorithm to learn layers of identity mappings

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep Residual Learning for Image Recognition." CVPR, 2016.

Problem: Difficult to Perform Identity Mapping

e.g.,

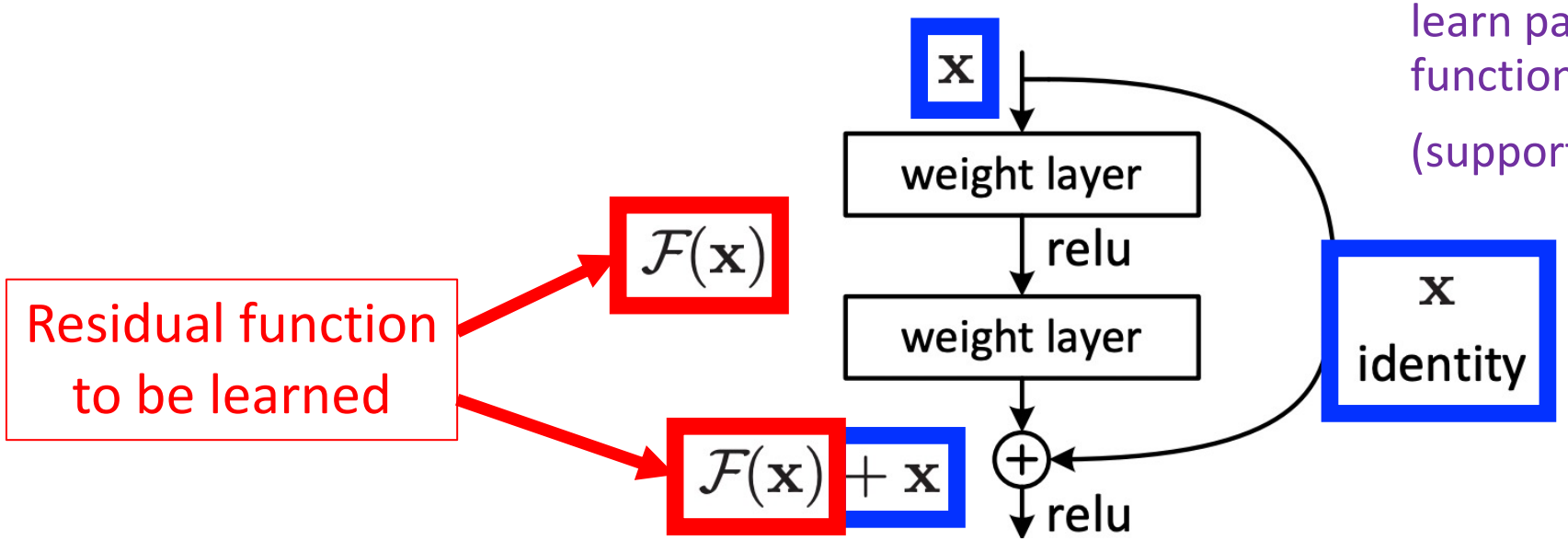


Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep Residual Learning for Image Recognition." CVPR, 2016.

Key Idea: Skip Connections that Perform Identity Mapping

Assumption: Easier for a network to learn parameters close to 0 than a function that performs identity mapping

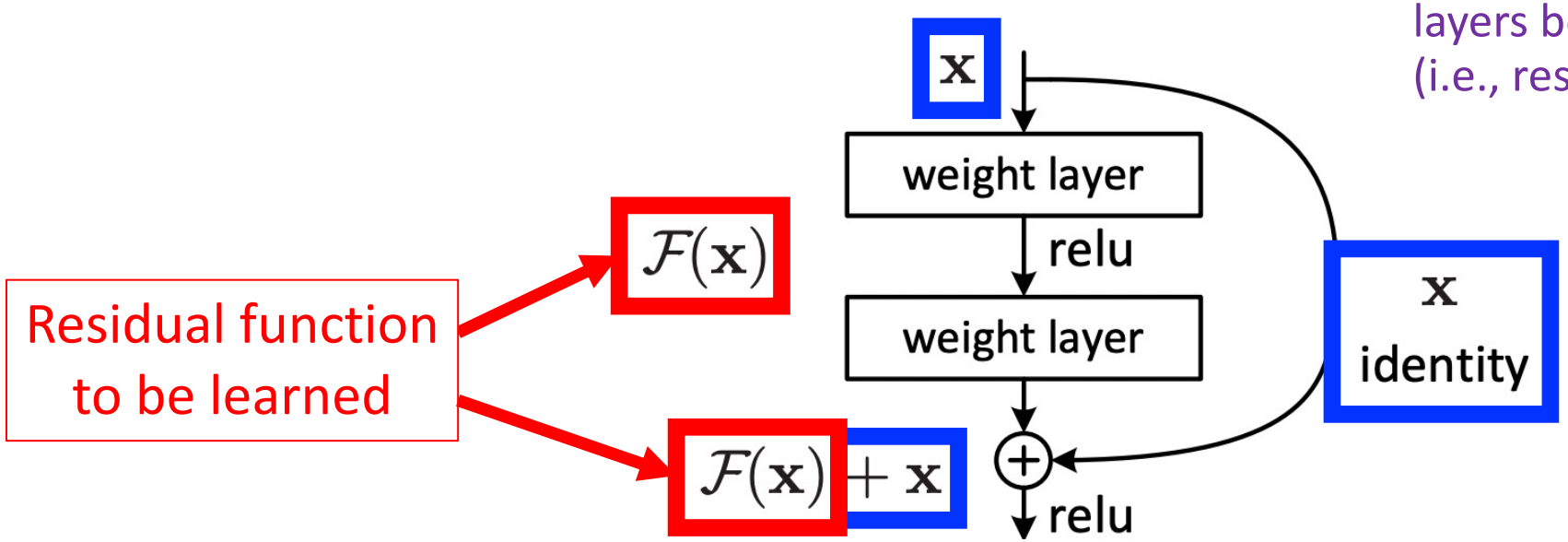
(supported by experiments in the paper)



Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep Residual Learning for Image Recognition." CVPR, 2016.

Key Idea: Skip Connections that Perform Identity Mapping

Paper explores when number of layers between skip connections (i.e., residual block) is 2 or 3



Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep Residual Learning for Image Recognition." CVPR, 2016.

Key Idea: Skip Connections that Perform Identity Mapping

How many model parameters need to be learned when adding a skip connection?

- 0

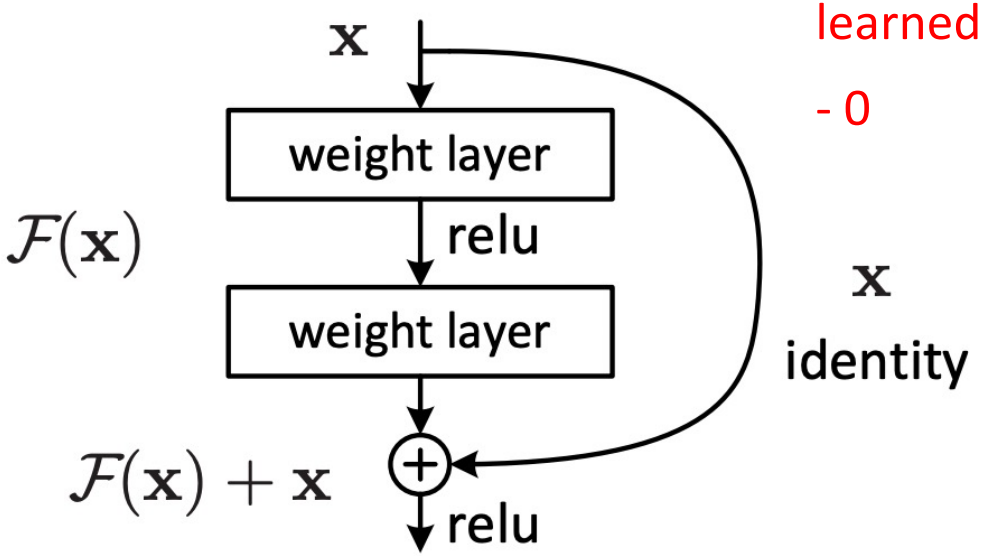
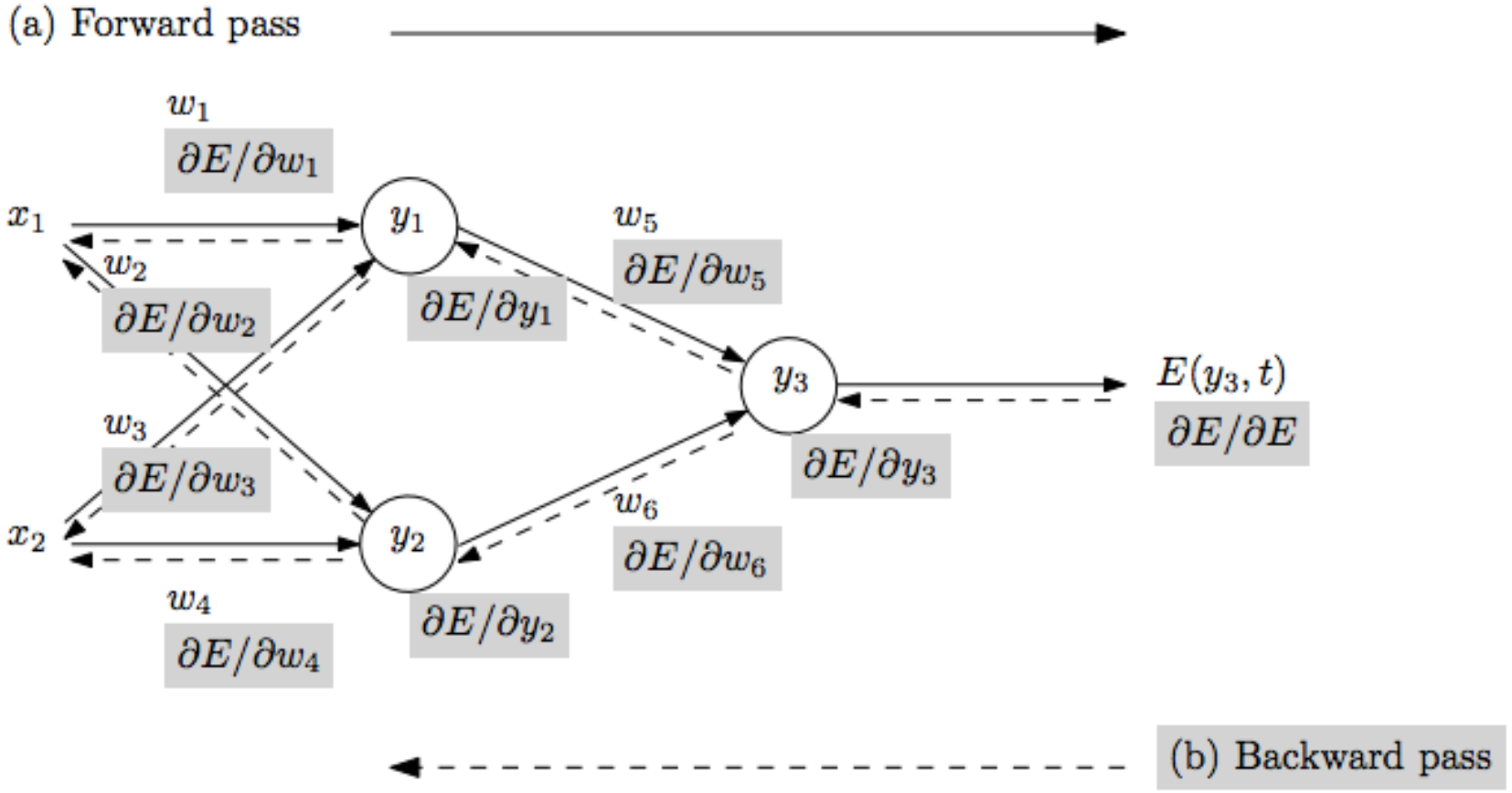


Figure 2. Residual learning: a building block.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep Residual Learning for Image Recognition." CVPR, 2016.

ResNet Training (follows AlexNet)

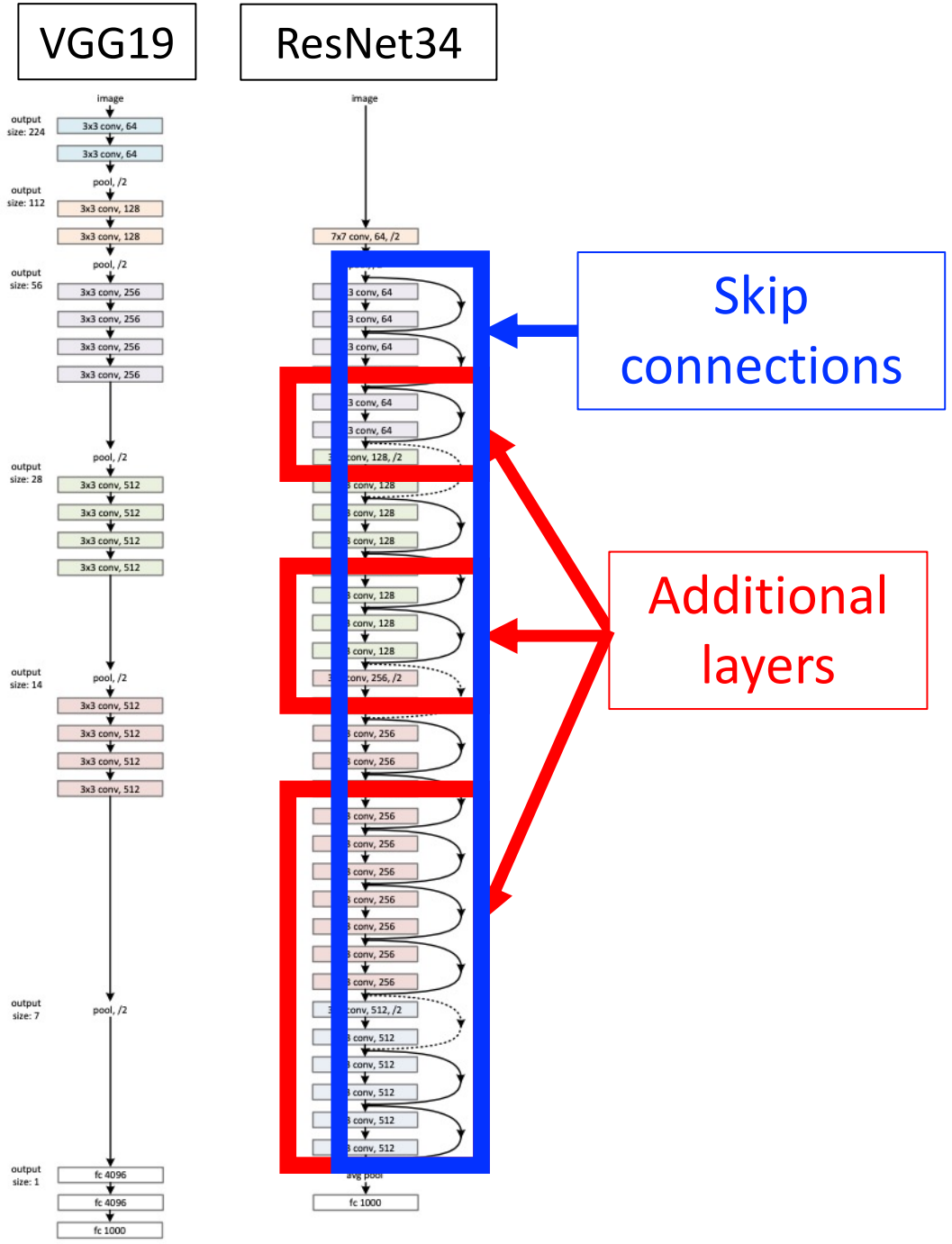


- Repeat until stopping criterion met:
  1. **Forward pass**: propagate training data through model to make prediction
  2. Quantify the dissatisfaction with a model's results on the training data
  3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each model parameter
  4. Update each parameter using calculated gradients

Figure from: Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, Jeffrey Mark Siskind; Automatic Differentiation in Machine Learning: a Survey; 2018

Experimental Results

Deep residual learning framework using skip connections obtains state-of-art performance for the ImageNet object recognition challenge and other challenges by learning deeper models than prior work (18, 34, 50, 101, & 152 layers!)



Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep Residual Learning for Image Recognition." CVPR, 2016.

Experimental Results on Validation Set

model	top-1 err.	top-5 err.
VGG-16 [40]	28.07	9.33
GoogLeNet [43]	-	9.15
PReLU-net [12]	24.27	7.38
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	**21.43**	**5.71**

Performance improves with more layers

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep Residual Learning for Image Recognition." CVPR, 2016.

ResNet Exceeded Human Performance!

Progress of models on ImageNet (Top 5 Error)

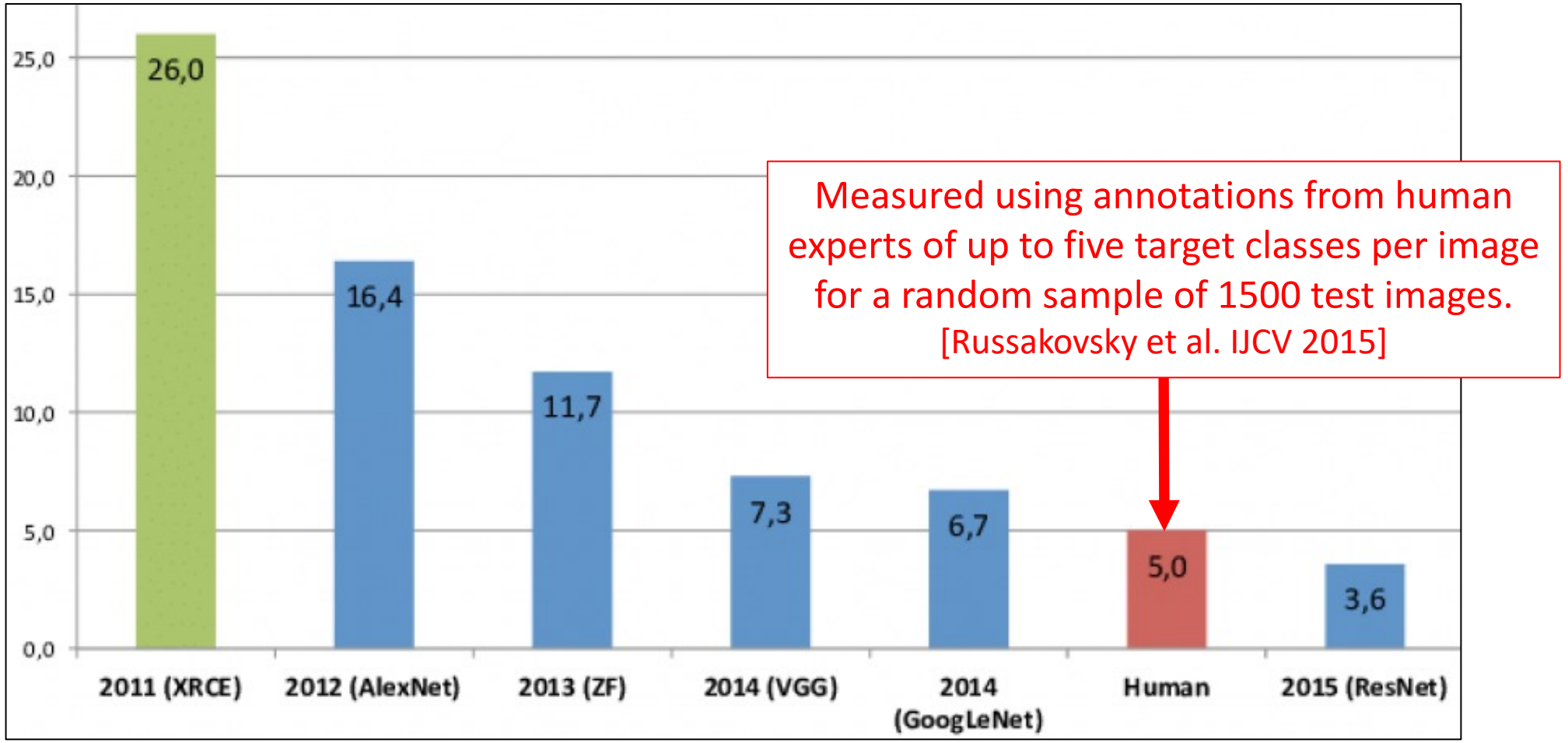


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

ResNet: Key Trick for Going Deeper

- Skip connections with residual learning

Summary: “Deeper” Models Perform Better

Progress of models on ImageNet (Top 5 Error)

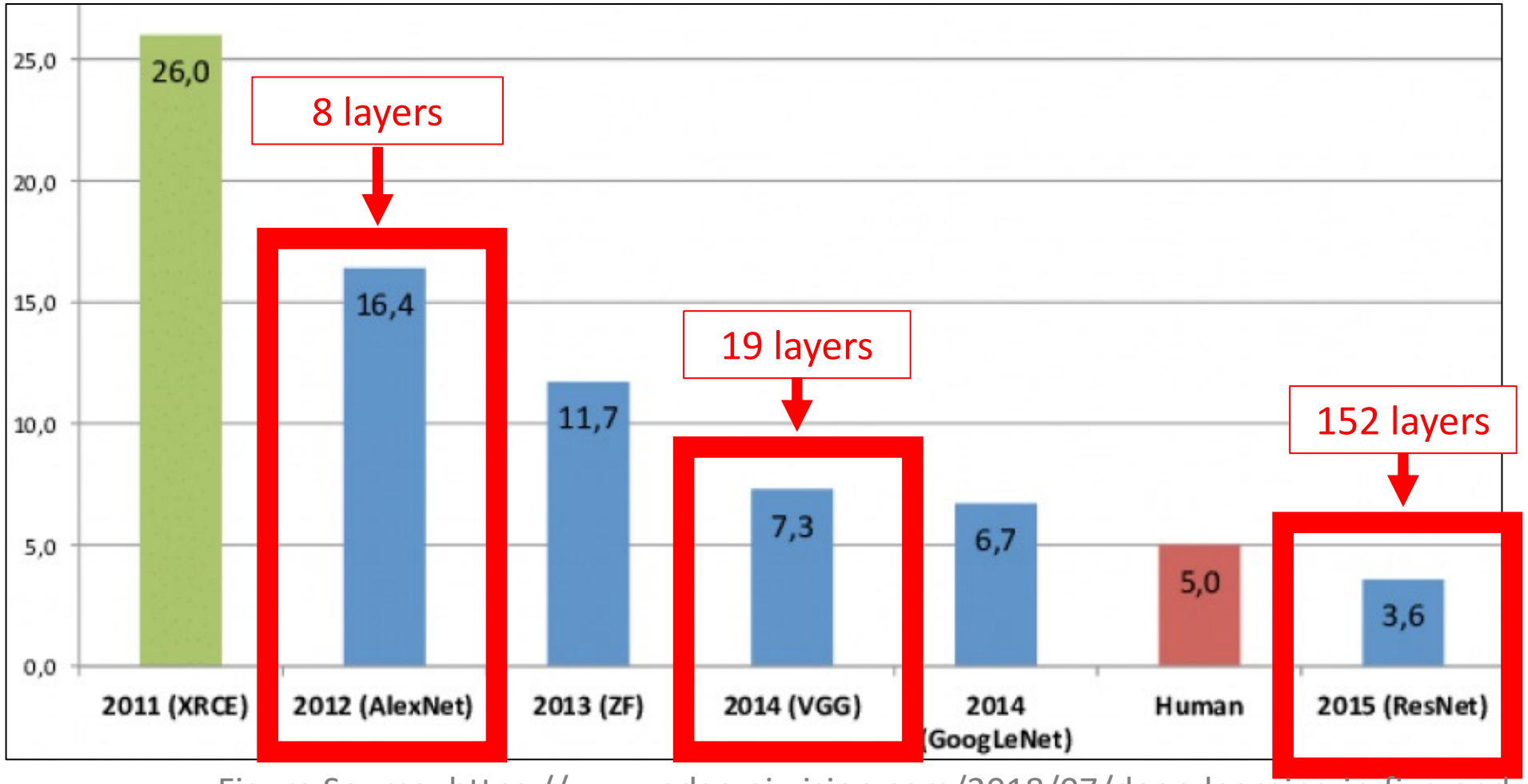
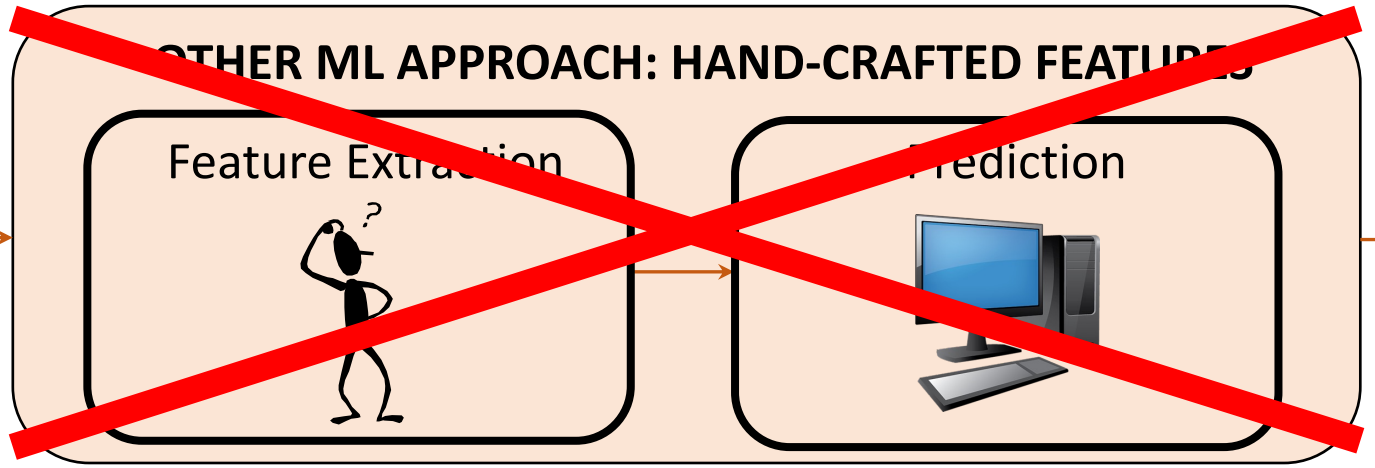


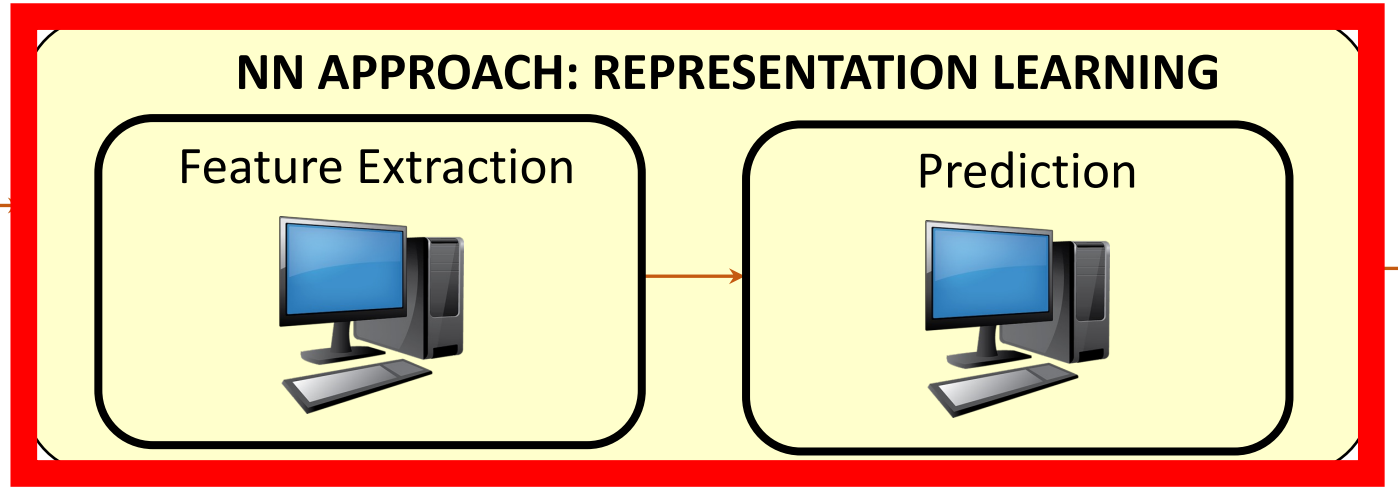
Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

Today's Topics

- ImageNet challenge winners: going deeper
- Representation learning
- Fine-tuning
- Programming tutorial

Recall: Motivation for Neural Networks (NNs) Over Other Machine Learning (ML) Approaches





What Neural Networks Learn

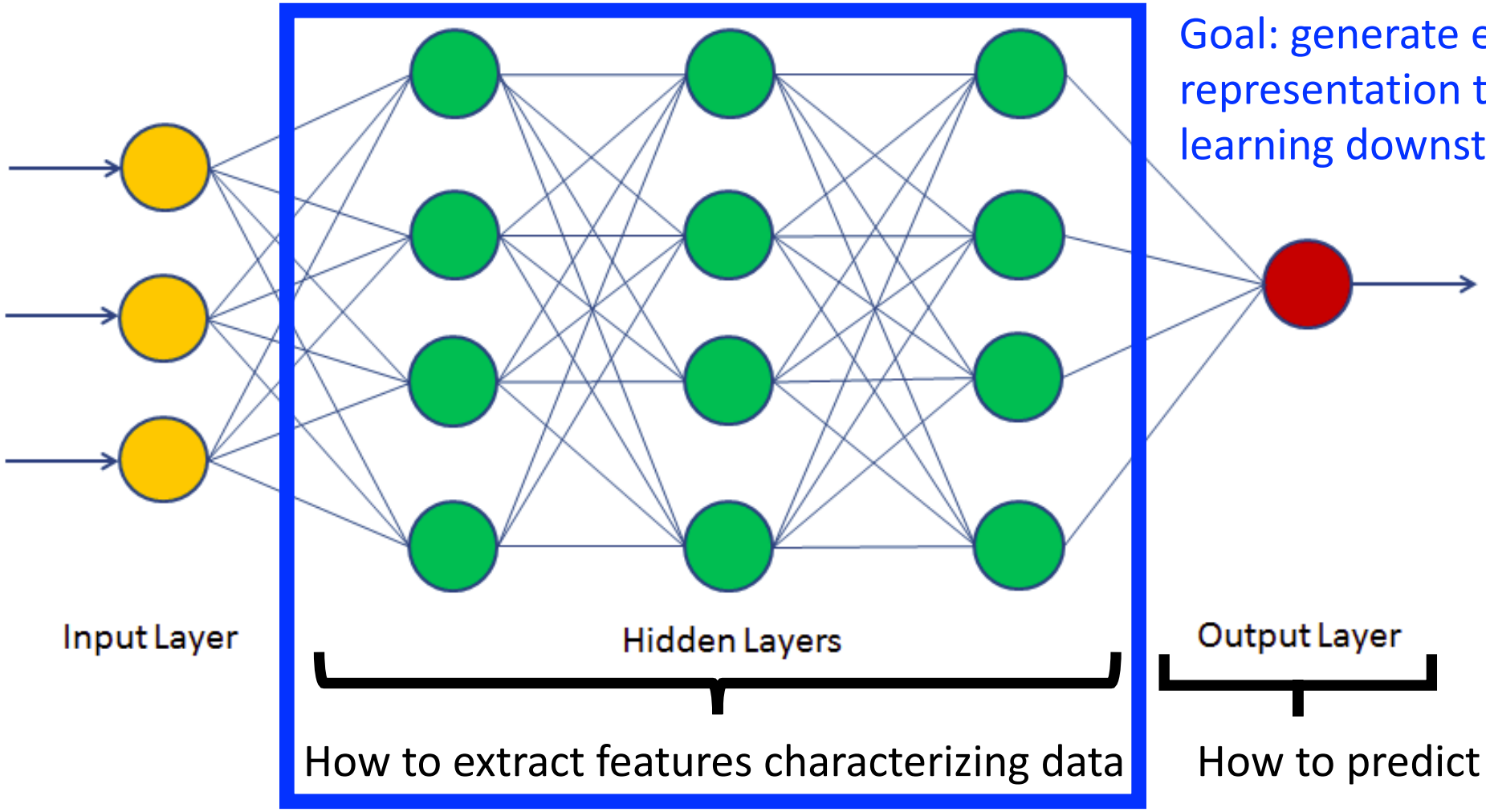


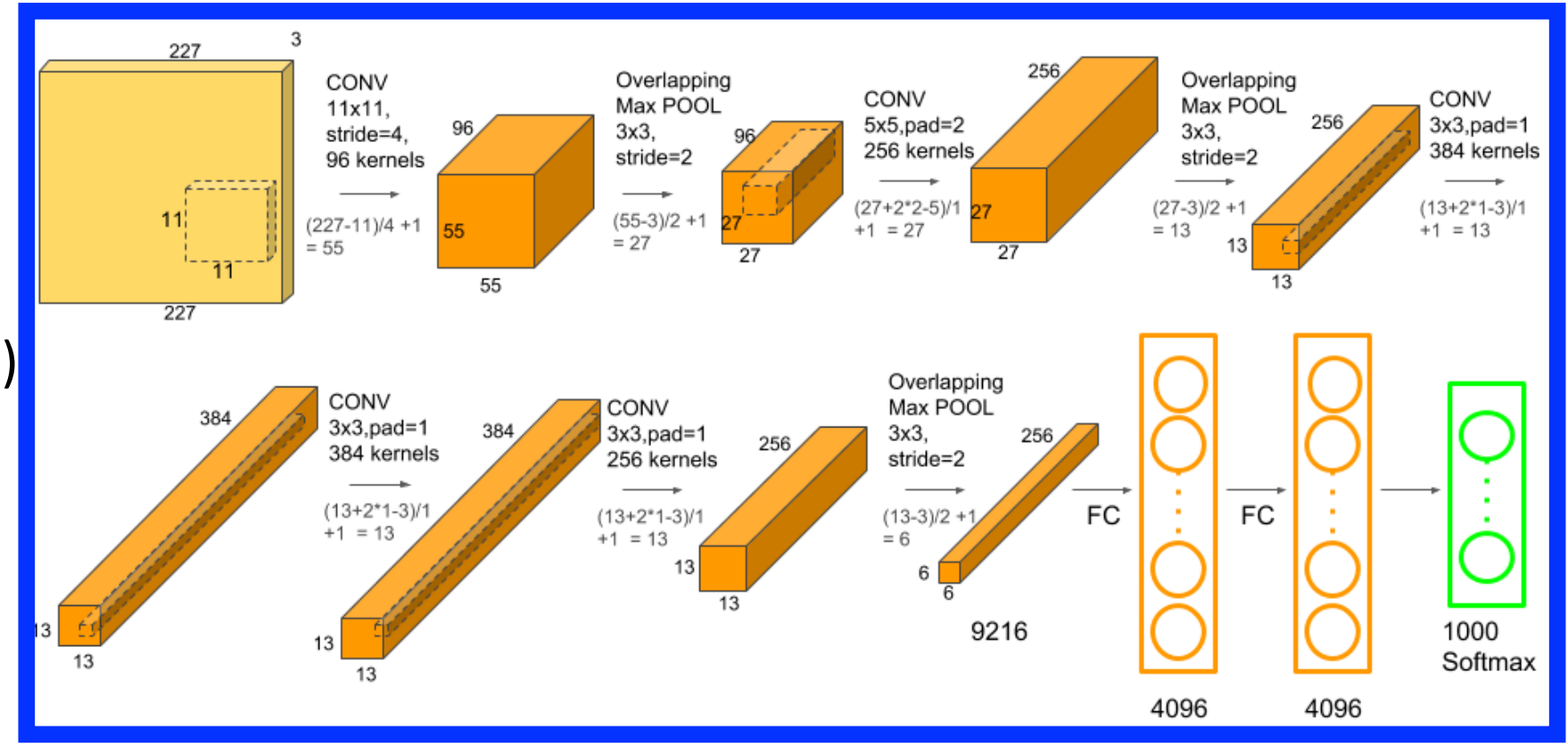
Figure Source: https://www.datacamp.com/community/tutorials/neural-network-models-r

What representations are CNNs learning?

e.g., AlexNet Pretrained Features

Can choose any intermediate layer to represent the image

A representation of the data extracted inside a network (rather than the input or predicted output)



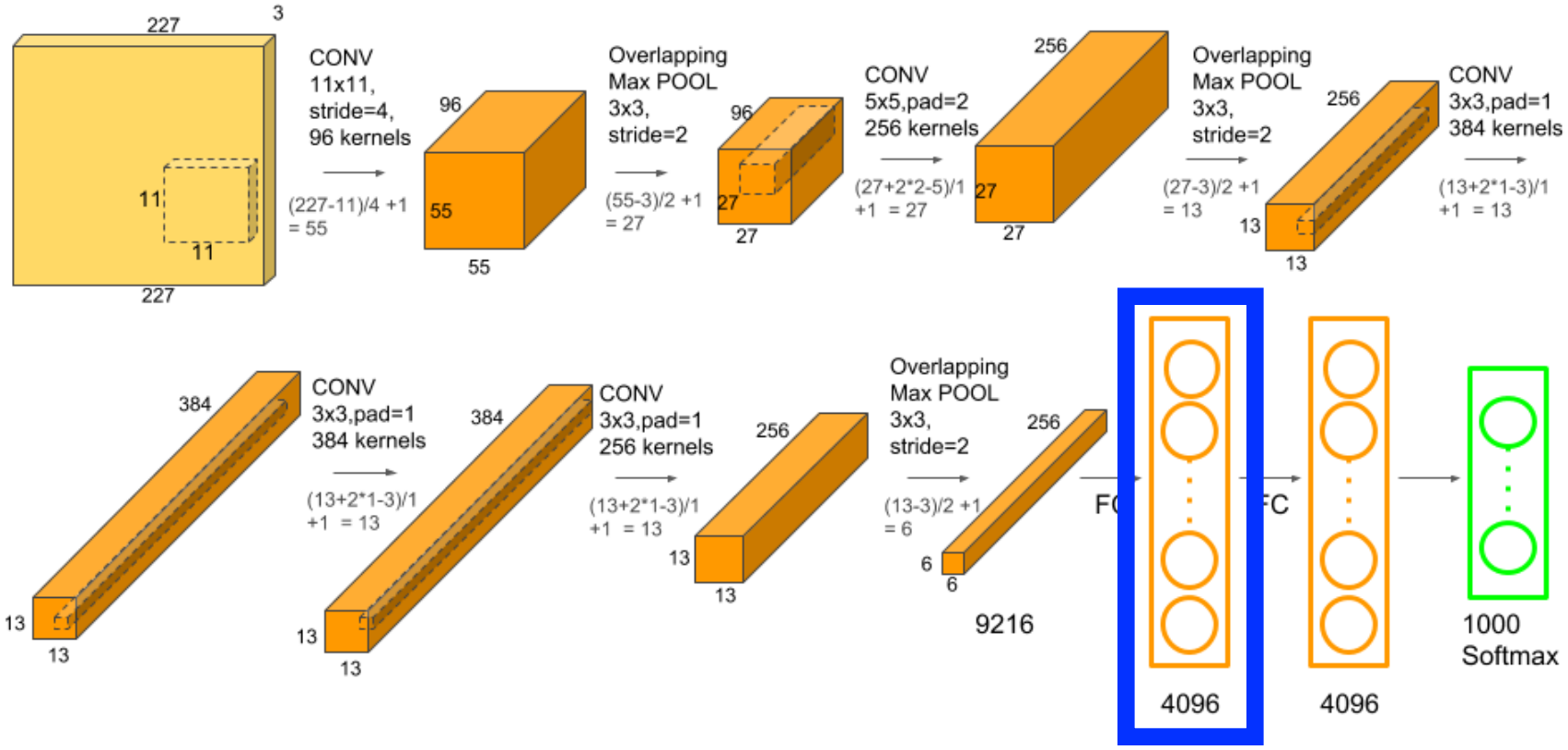
Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

e.g., AlexNet Pretrained Features

What is the dimensionality of the FC6 layer?

A representation of the data extracted inside a network (rather than the input or predicted output)

- a 227 x 227 x 3 image can now be represented with 4096 values



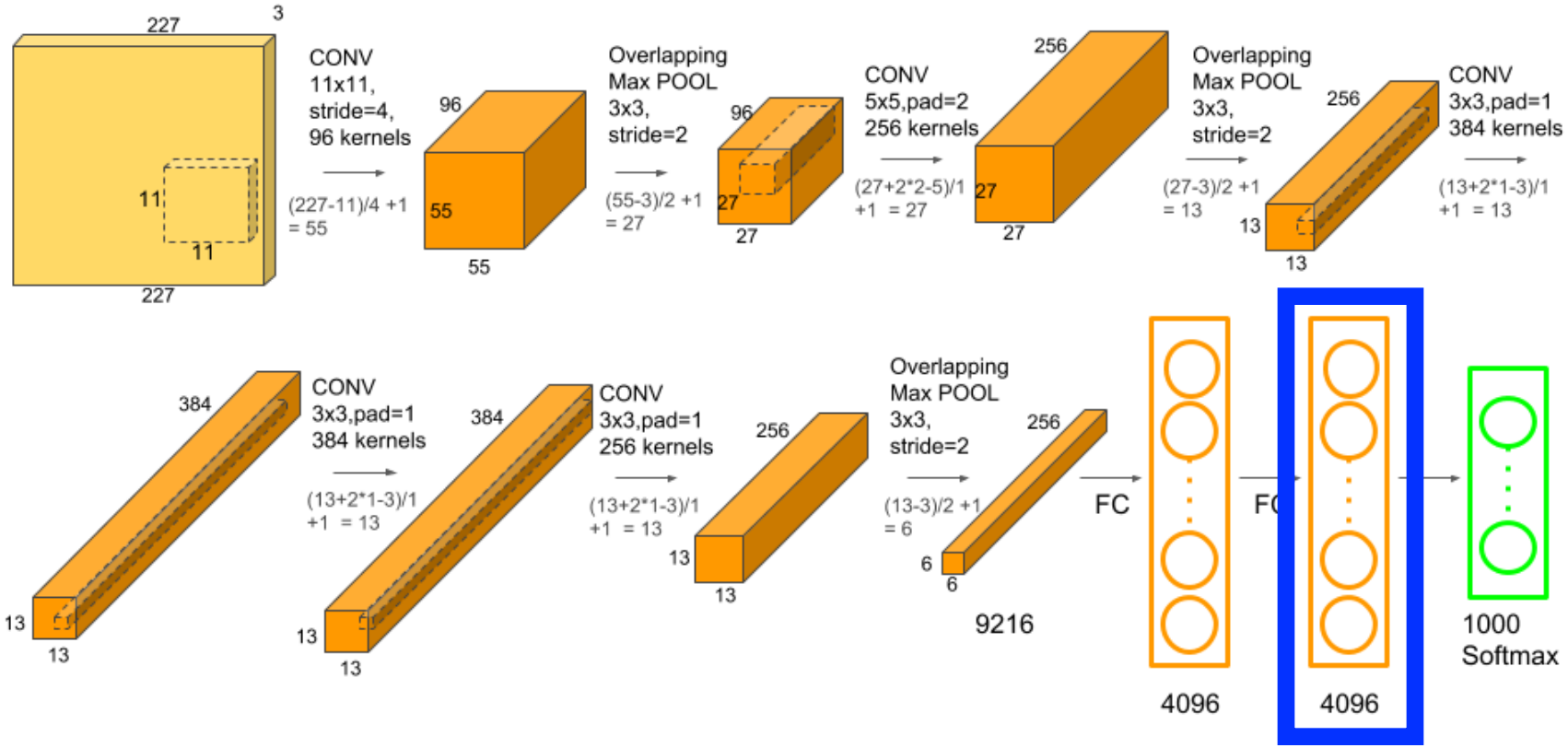
Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

e.g., AlexNet Pretrained Features

What is the dimensionality of the FC7 layer?

A representation of the data extracted inside a network (rather than the input or predicted output)

- a 227 x 227 x 3 image can now be represented with 4096 values



Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

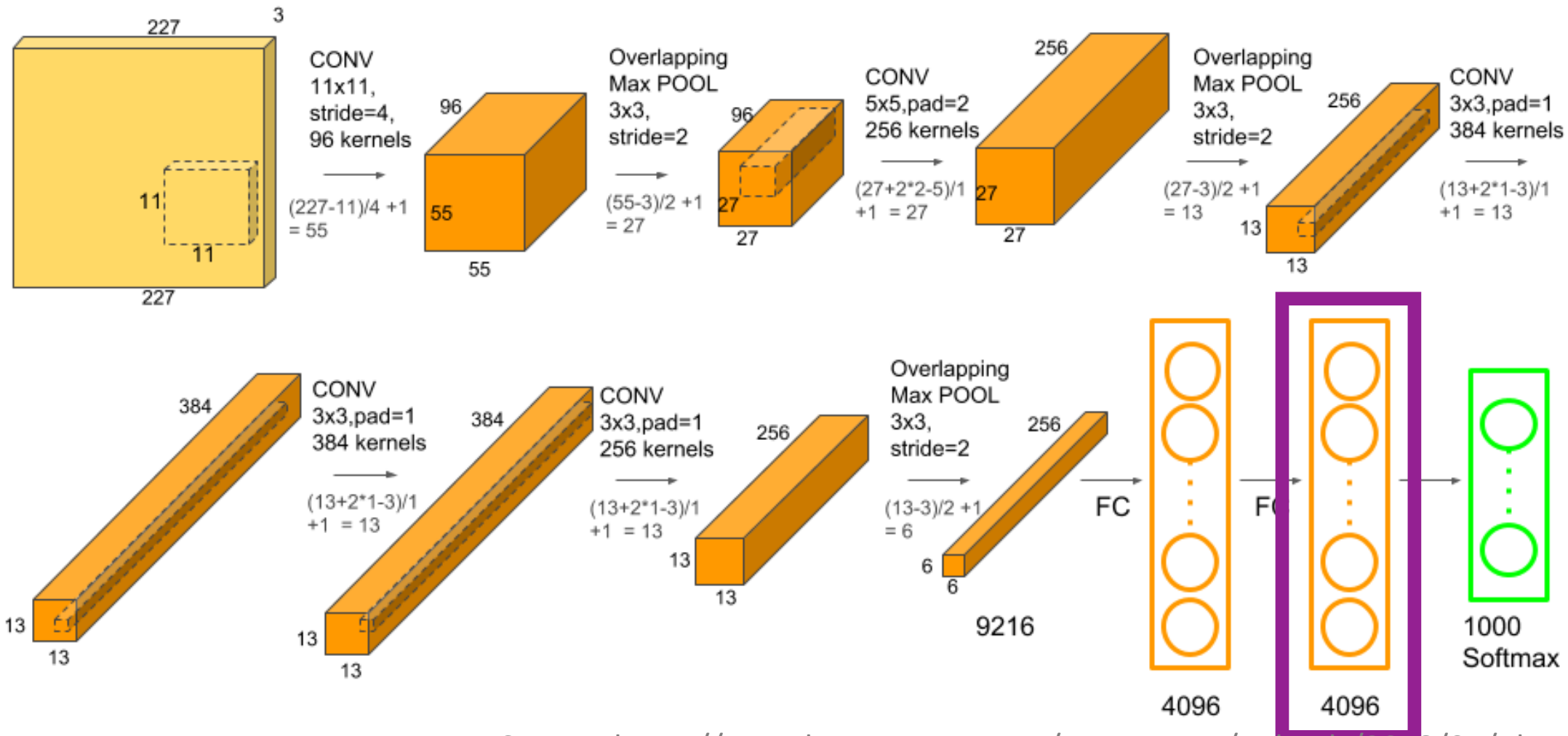
Tricks for Interpreting Representations

- Analyze images that maximally activate units in a network
- Visualize filters
- Visualize resulting activation maps
- Visualize images with most activated images by a neuron
- And many more…

Tricks for Interpreting Representations

- Analyze images that maximally activate units in a network
- Visualize filters
- Visualize resulting activation maps
- Visualize images with most activated images by a neuron
- And many more…

CNN: Pretrained Features Interpretation

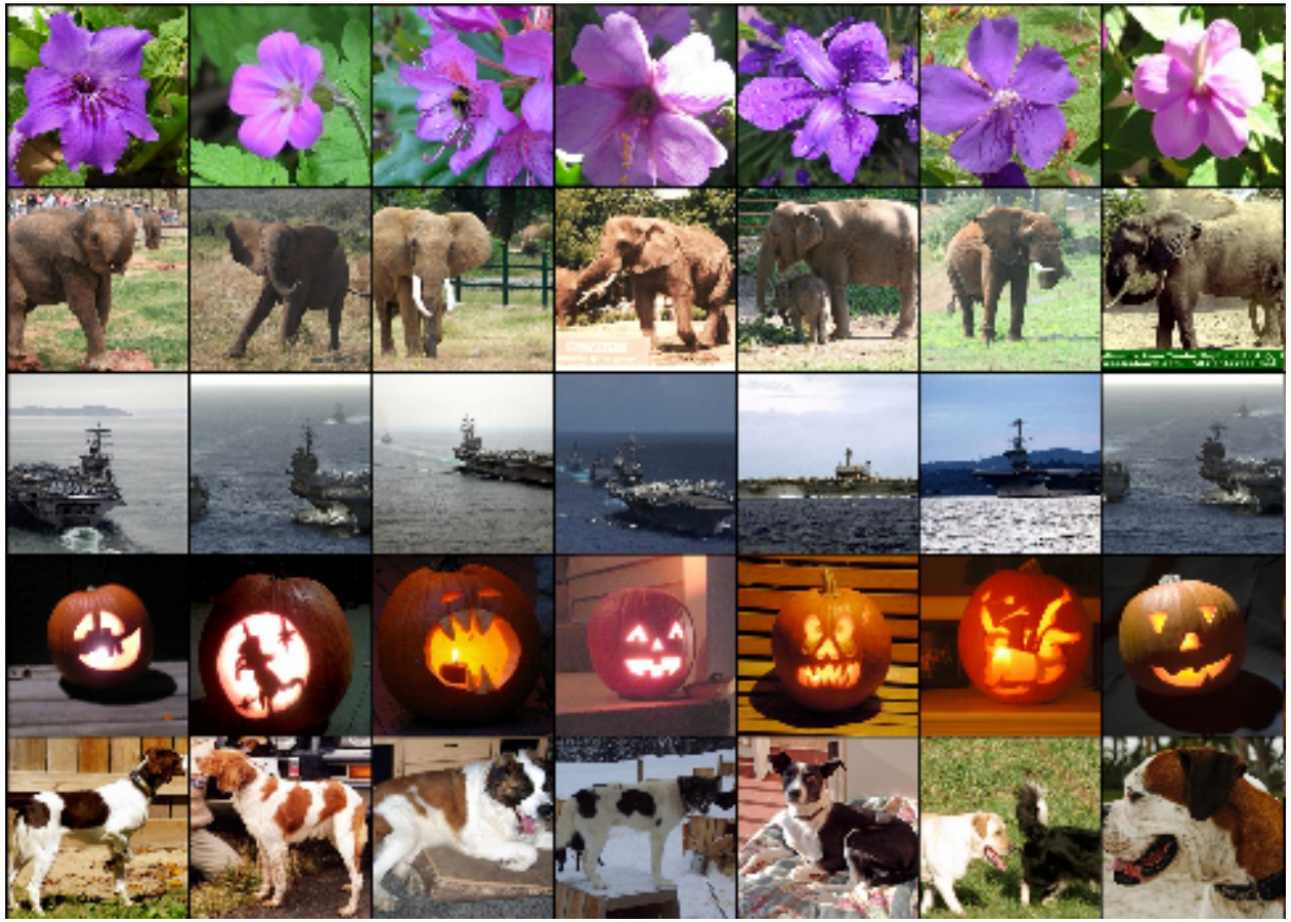


Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

CNN: Pretrained Features Interpretation

ImageNet test images

ImageNet training images with smallest Euclidean distance between its FC7 feature activation and that of the test image



What can you infer about what the FC7 feature represents?

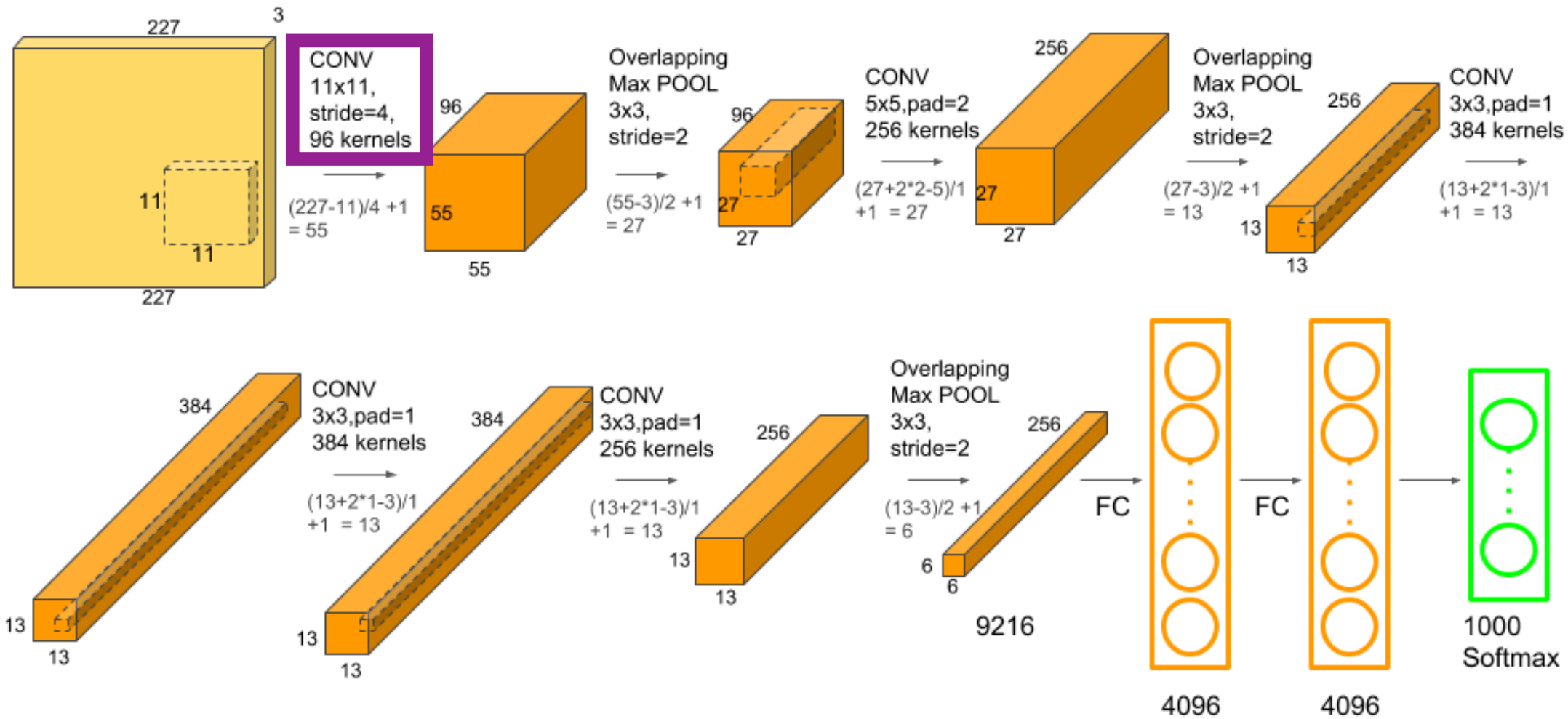
- Image semantics regardless of illumination and object pose

Krizhevsky, Sutskever, and Hinton. ImageNet Classification with Deep Convolutional Neural Networks. NeurIPS 2012.

Tricks for Interpreting Representations

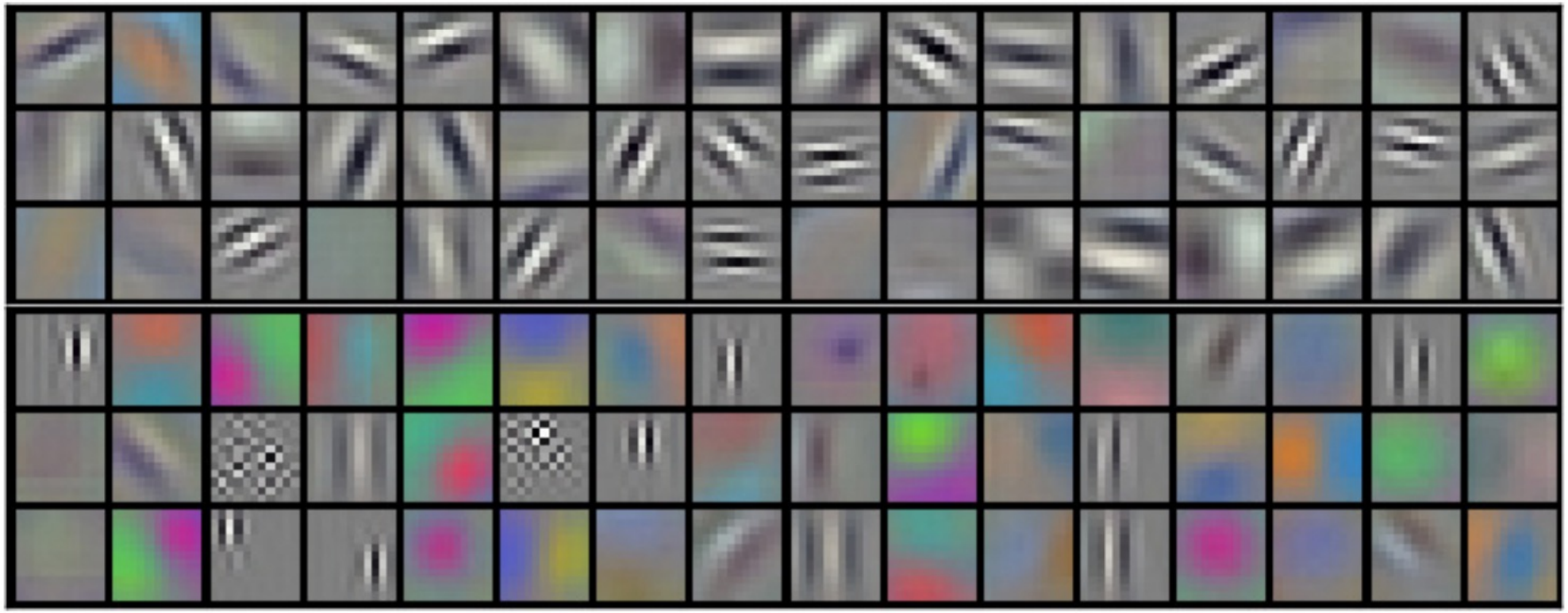
- Analyze images that maximally activate units in a network
- Visualize filters
- Visualize resulting activation maps
- Visualize images with most activated images by a neuron
- And many more…

Inspecting What Was Learned: AlexNet



Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

AlexNet: 96 Filters in Convolutional Layer 1



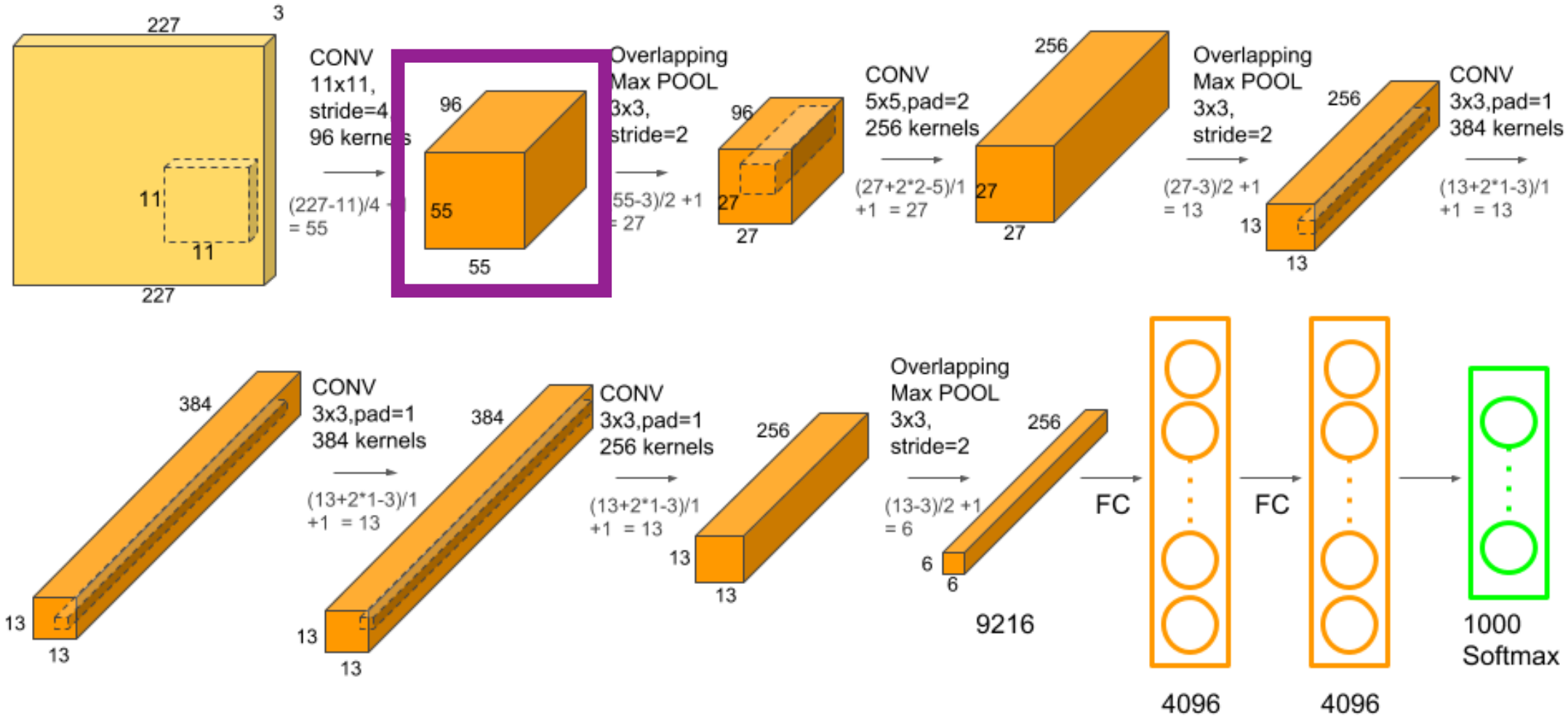
Filters for detecting different frequencies, orientations, and colors!

Krizhevsky, Sutskever, and Hinton. ImageNet Classification with Deep Convolutional Neural Networks. NeurIPS 2012.

Tricks for Interpreting Representations

- Analyze images that maximally activate units in a network
- Visualize filters
- Visualize resulting activation maps
- Visualize images with most activated images by a neuron
- And many more…

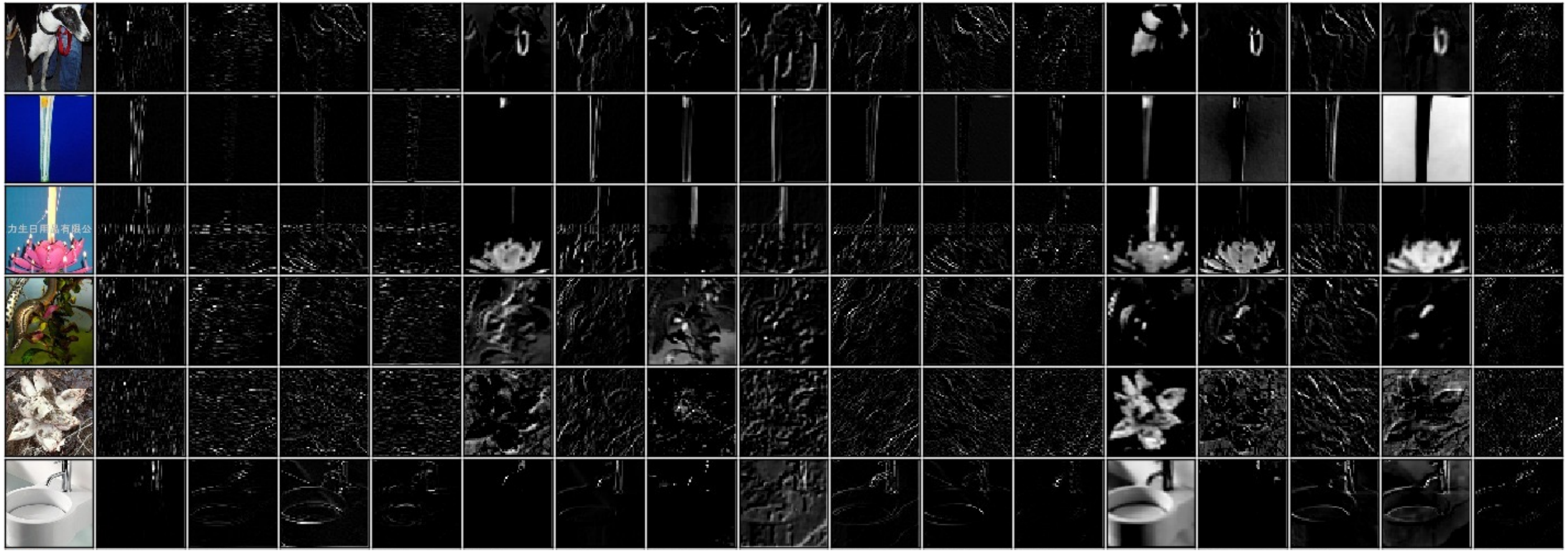
AlexNet: Example Activation Maps



Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png

AlexNet: Example Activation Maps (Recall Each Map Results from One Filter)

Images



Frequencies, orientations, and colors are detected

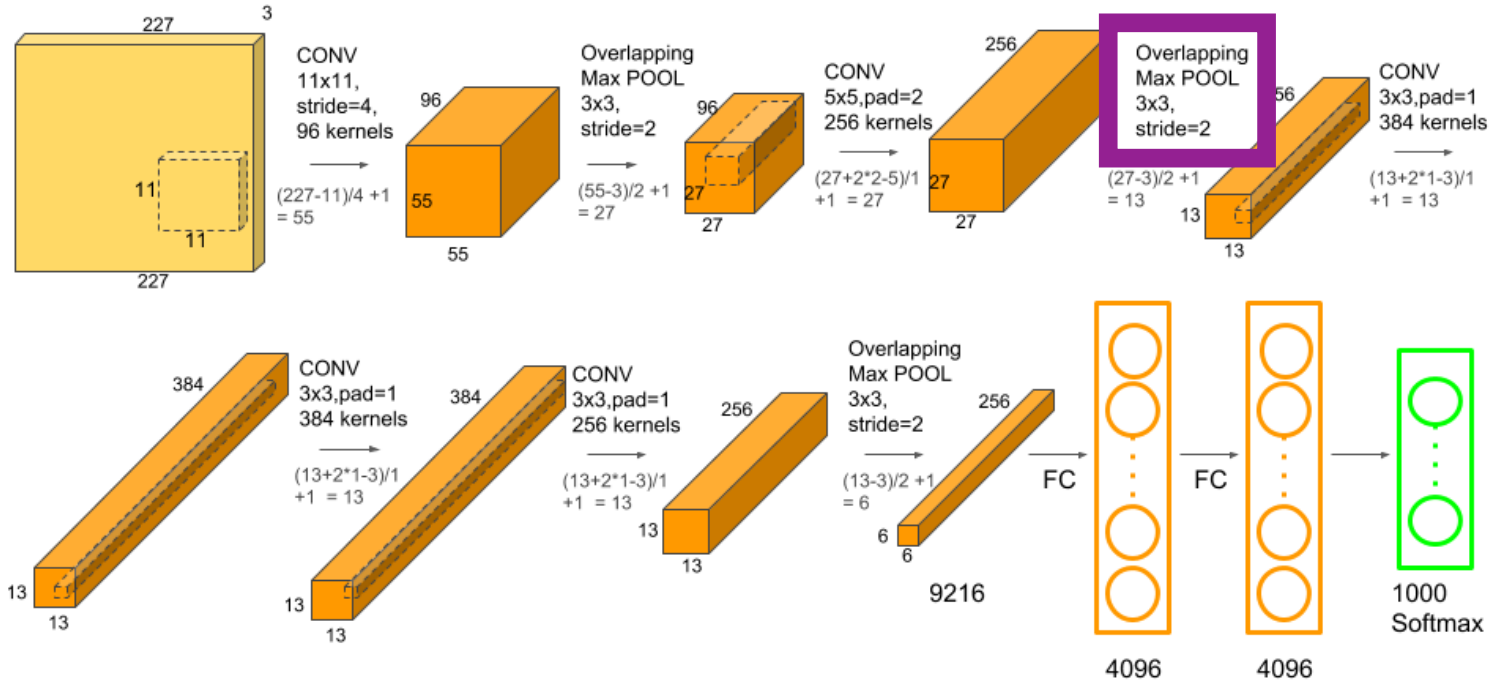
Krizhevsky, Sutskever, and Hinton. ImageNet Classification with Deep Convolutional Neural Networks. NeurIPS 2012.

Tricks for Interpreting Representations

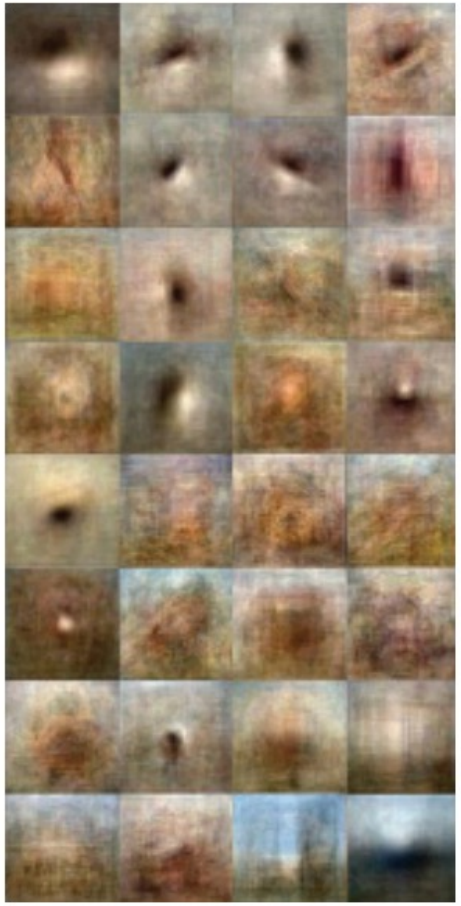
- Analyze images that maximally activate units in a network
- Visualize filters
- Visualize resulting activation maps
- Visualize images with most activated images by a neuron
- And many more…

AlexNet: Images that Lead to Maximal Activations

Mean images from the 100 test images for each unit in each layer that fire the most (i.e., highest activation scores); e.g.,



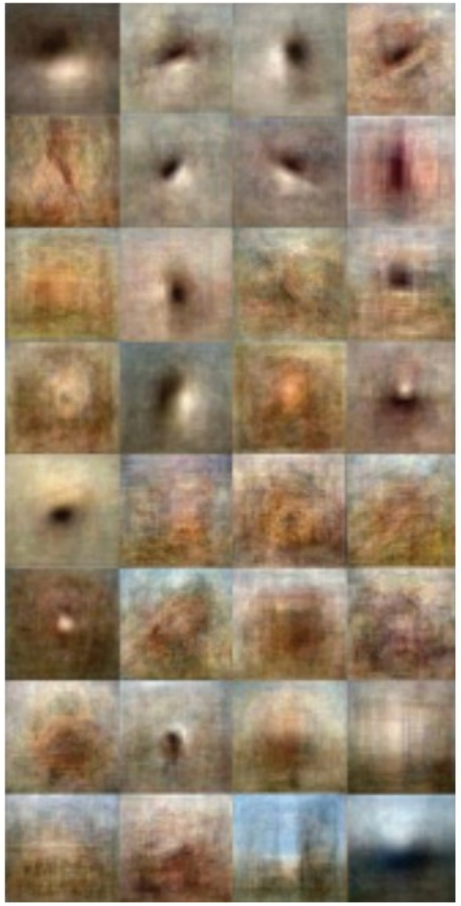
Source: https://www.learnopencv.com/wp-content/uploads/2018/05/AlexNet-1.png



Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014.

AlexNet: Images that Lead to Maximal Activations

Mean images from the 100 test images for each unit in each layer that fire the most (i.e., highest activation scores); e.g.,



What type of features does the model appear to detect?

Bolei Zhou et al. Learning Deep Features for Scene Recognition using Places Database. NIPS 2014.

Tricks for Interpreting Representations

- Analyze images that maximally activate units in a network
- Visualize filters
- Visualize resulting activation maps
- Visualize images with most activated images by a neuron
- And many more…

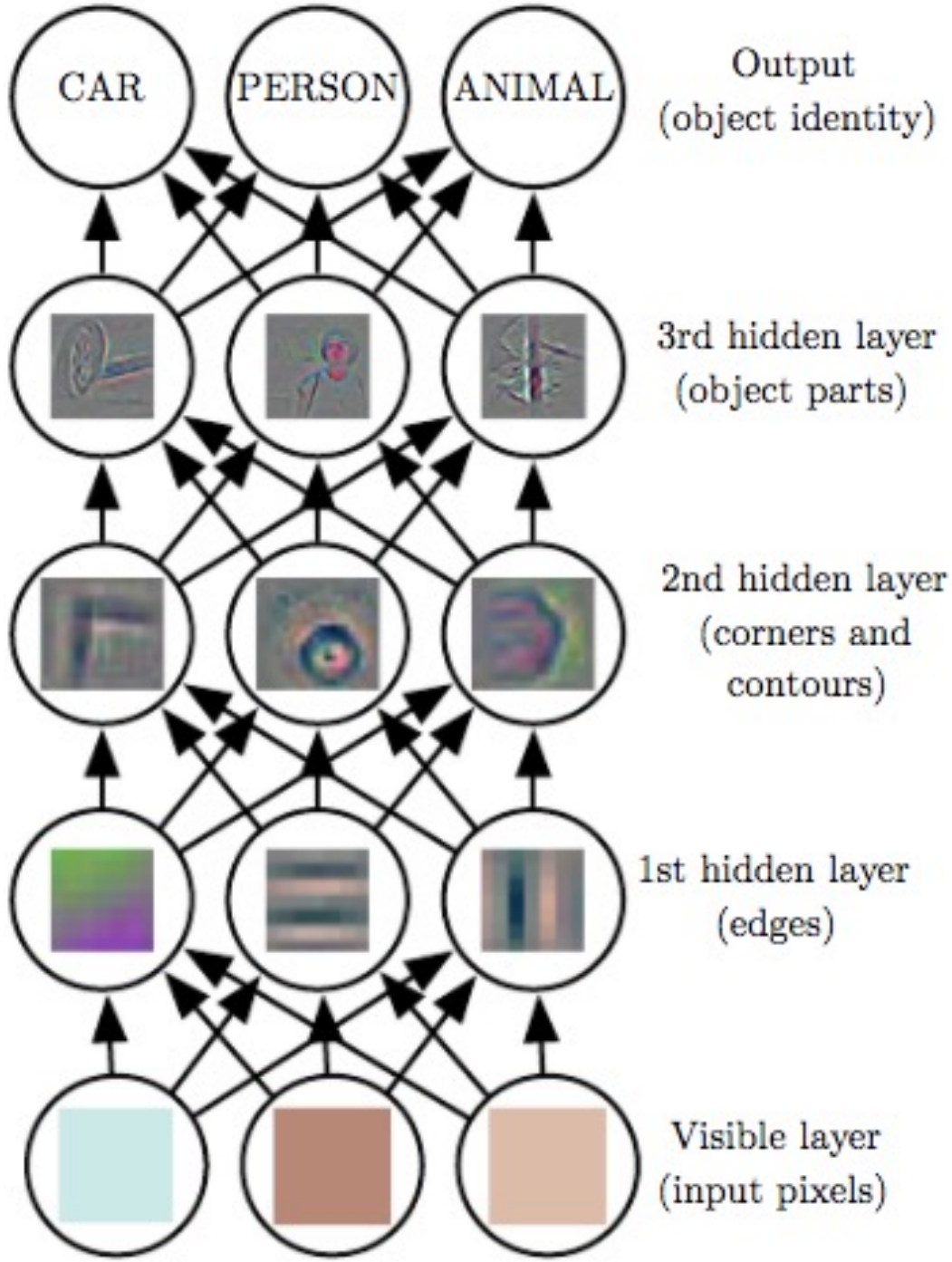
CNN: Common Representations

High-level feature maps

Mid-level feature maps

Low-level feature maps

Note: learned representations are heavily dependent on the type of training data



Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville

Today's Topics

- ImageNet challenge winners: going deeper
- Representation learning
- Fine-tuning
- Programming tutorial

What Neural Networks Learn

A pretrained network can be "fine-tuned" for a different dataset and/or task

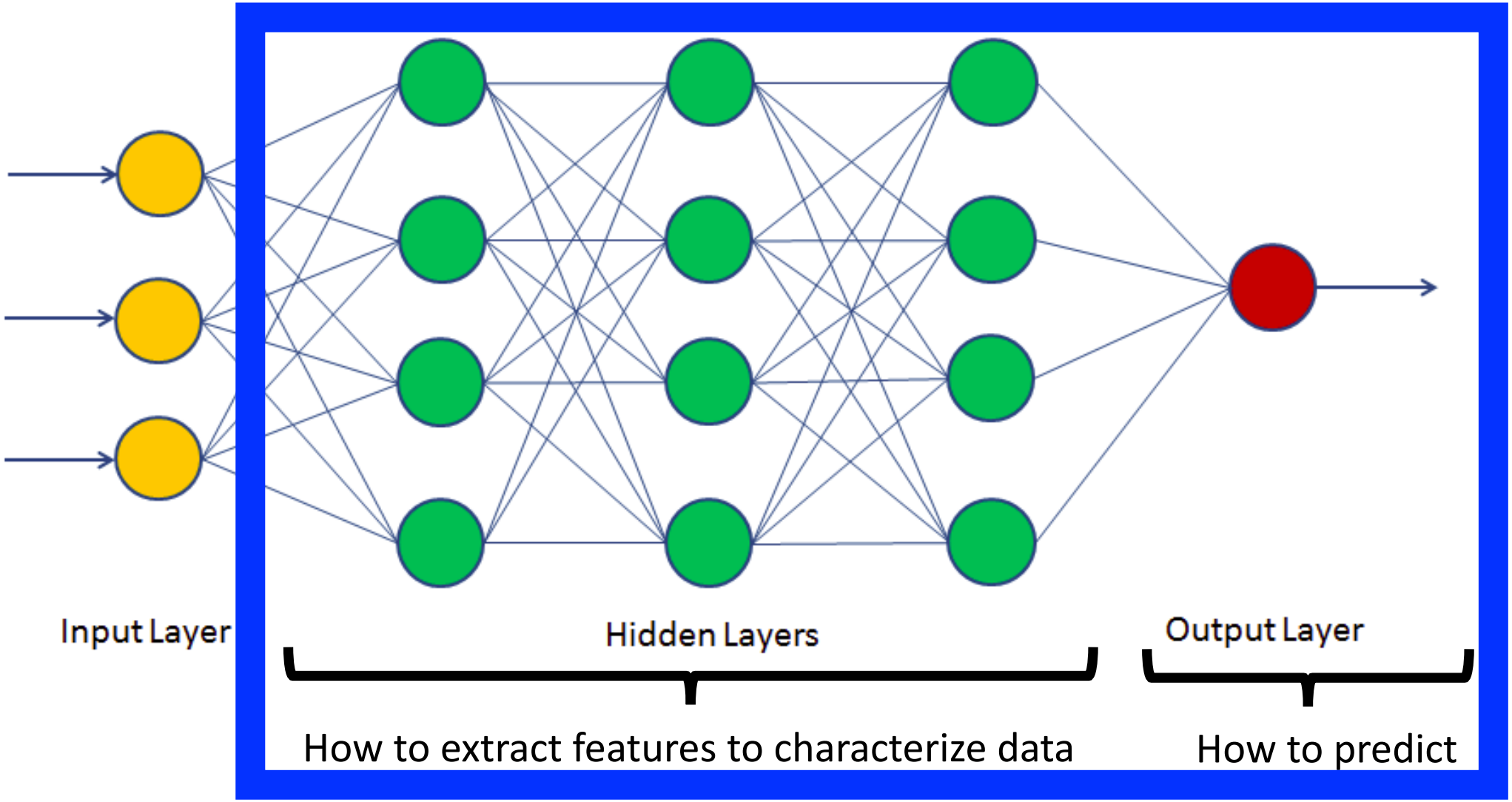


Figure Source: https://www.datacamp.com/community/tutorials/neural-network-models-r

Fine-Tuning (aka, Transfer Learning)

Use pretrained network as a starting point to train for a different dataset and/or task; e.g.,

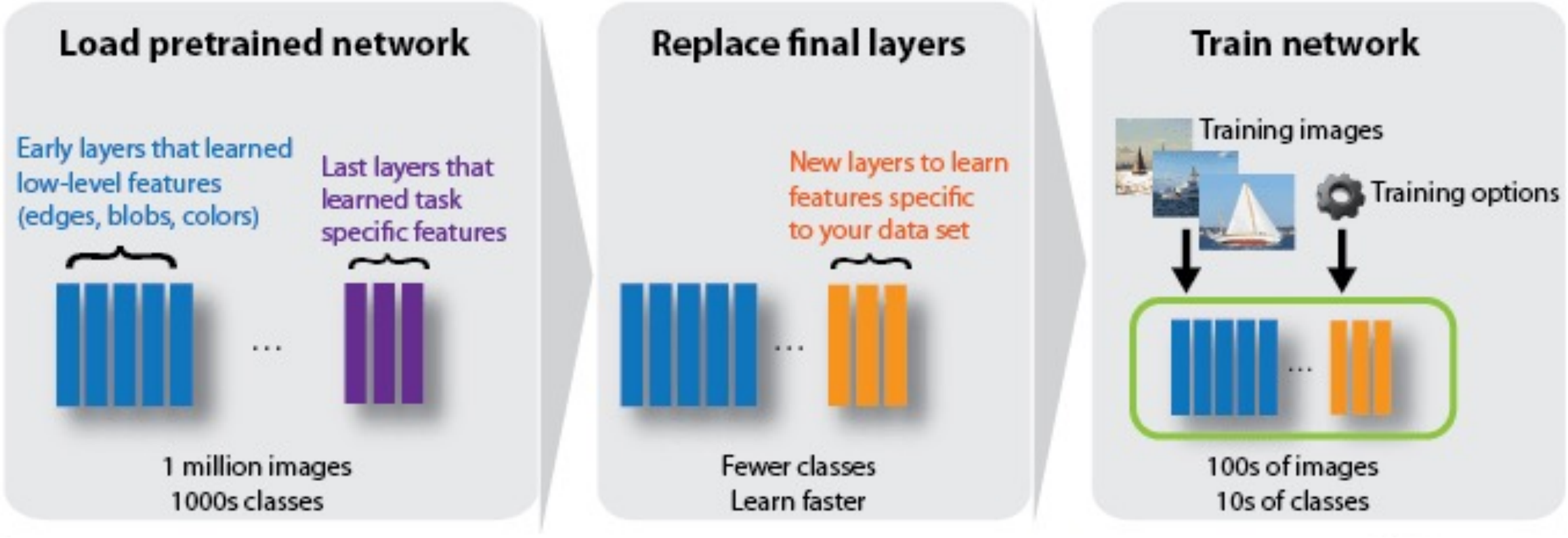


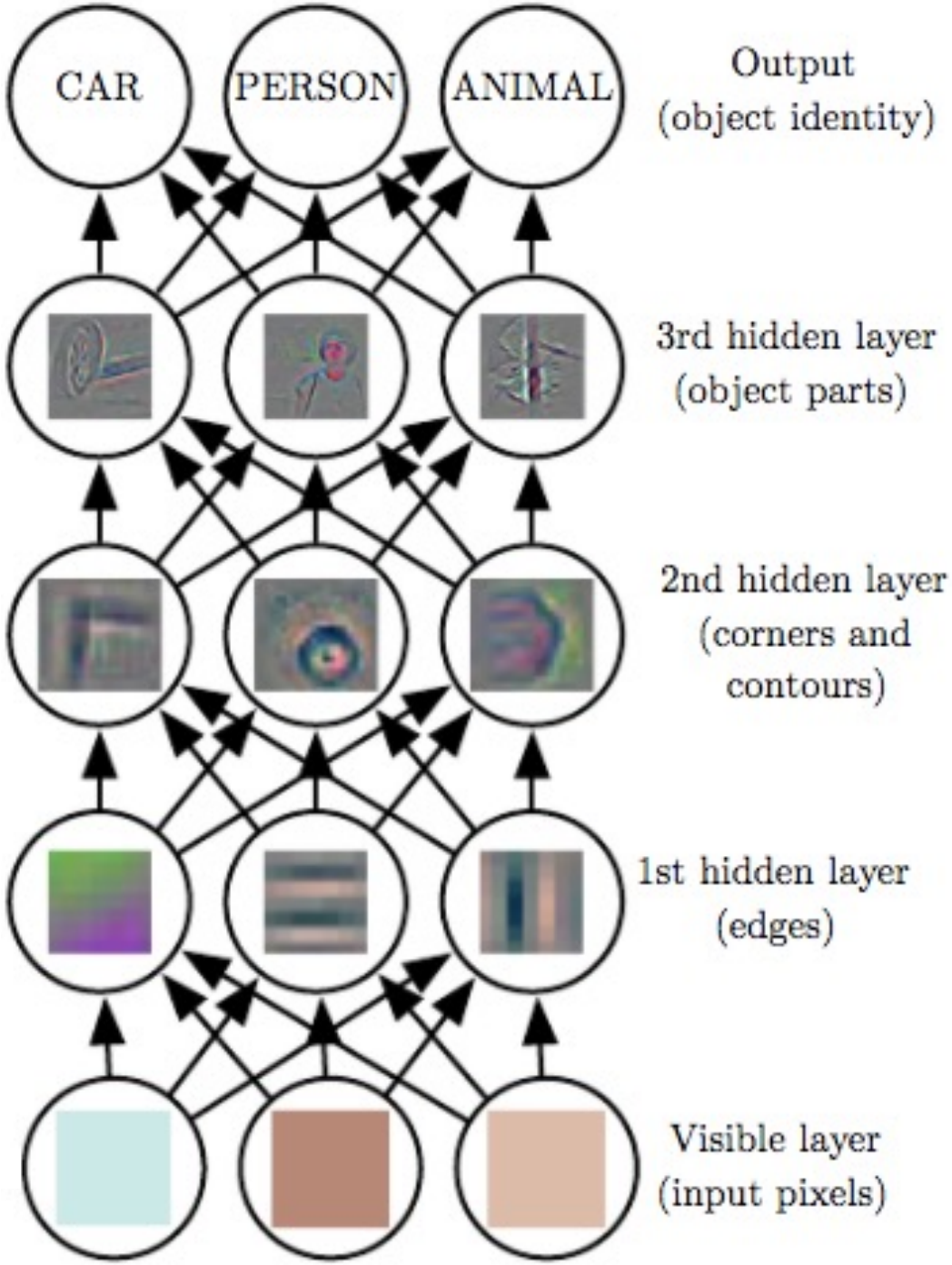
Image Source: https://www.mathworks.com/help/deeplearning/ug/transfer-learning-using-alexnet.html

Key Choice: Freeze vs Fine-Tune Layers?

High-level features

Mid-level features

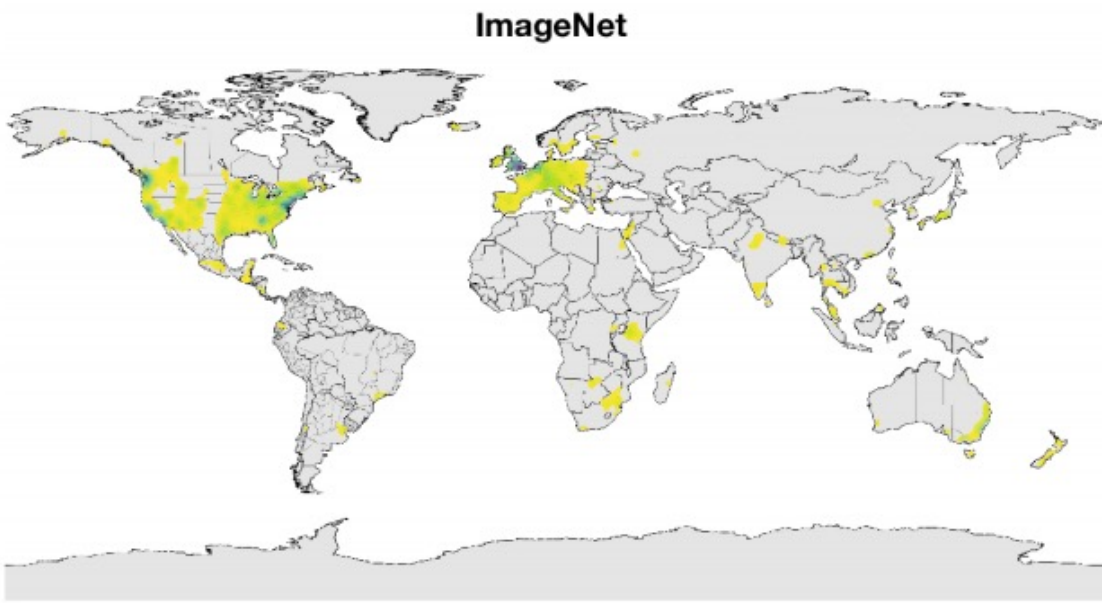
Low-level features



Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville

Common Questions through an Example

- Assume you need to develop a classifier that recognizes common items in countries with low house incomes
  - If you fine-tuned AlexNet pretrained on ImageNet, which layers would you remove and/or freeze? Why?



Zhao et al. Men also like shopping: Reducing gender bias amplification using corpus-level constraints. 2017.

Ground truth: Soap **Nepal, 288 $/month**

Azure: food, cheese, bread, cake, sandwich
Clarifai: food, wood, cooking, delicious, healthy
Google: food, dish, cuisine, comfort food, spam
Amazon: food, confectionary, sweets, burger
Watson: food, food product, turmeric, seasoning
Tencent: food, dish, matter, fast food, nutriment

Ground truth: Soap **UK, 1890 $/month**

Azure: toilet, design, art, sink
Clarifai: people, faucet, healthcare, lavatory, wash closet
Google: product, liquid, water, fluid, bathroom accessory
Amazon: sink, indoors, bottle, sink faucet
Watson: gas tank, storage tank, toiletry, dispenser, soap dispenser
Tencent: lotion, toiletry, soap dispenser, dispenser, after shave

Ground truth: Spices **Phillipines, 262 $/month**

Azure: bottle, beer, counter, drink, open
Clarifai: container, food, bottle, drink, stock
Google: product, yellow, drink, bottle, plastic bottle
Amazon: beverage, beer, alcohol, drink, bottle
Watson: food, larder food supply, pantry, condiment, food seasoning
Tencent: condiment, sauce, flavorer, catsup, hot sauce

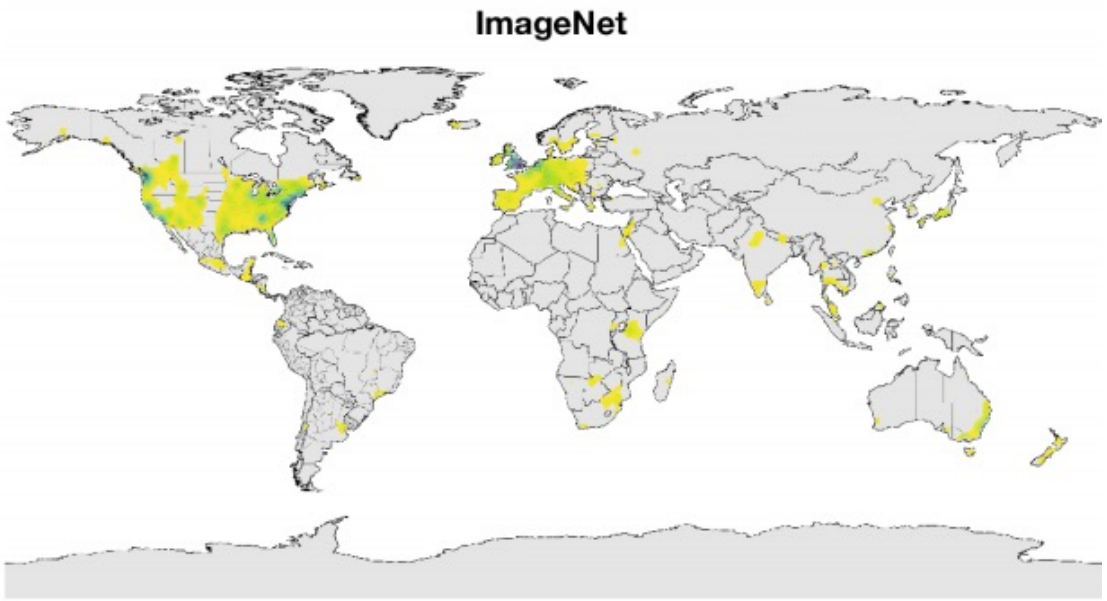
Ground truth: Spices **USA, 4559 $/month**

Azure: bottle, wall, counter, food
Clarifai: container, food, can, medicine, stock
Google: seasoning, seasoned salt, ingredient, spice, spice rack
Amazon: shelf, tin, pantry, furniture, aluminium
Watson: tin, food, pantry, paint, can
Tencent: spice rack, chili sauce, condiment, canned food, rack

DeVries et al. Does object recognition work for everyone? CVPR workshops, 2019.

Common Questions through an Example

- Assume you need to develop a classifier that recognizes common items in countries with low house incomes
  - If a large-scale dataset of low household income items was available, would you train AlexNet from scratch or fine-tune an ImageNet pretrained model? Why?



Zhao et al. Men also like shopping: Reducing gender bias amplification using corpus-level constraints. 2017.

DeVries et al. Does object recognition work for everyone? CVPR workshops, 2019.

Today's Topics

- ImageNet challenge winners: going deeper
- Representation learning
- Fine-tuning
- Programming tutorial

Today's Topics

- ImageNet challenge winners: going deeper
- Representation learning
- Fine-tuning
- Programming tutorial

Today’s Topics

- ImageNet challenge winners: going deeper
- Representation learning
- Fine-tuning
- Programming tutorial

The End