Object Recognition – Part 2

Danna Gurari
University of Colorado Boulder
Fall 2021

Review

- Last lecture:
    - Object recognition problem
    - Object recognition applications
    - Object recognition datasets
    - Object recognition evaluation metric
    - Typical solution: convolutional neural network
- Assignments (Canvas)
    - Reading assignment was due today
    - New reading assignment out later today that is due next week
- Questions?

Object Recognition: Today’s Topics

- ImageNet Challenge Top Performers
- Baseline Model: AlexNet
- VGG
- ResNet
- Discussion

Object Recognition: Today’s Topics

- ImageNet Challenge Top Performers
- Baseline Model: AlexNet
- VGG
- ResNet
- Discussion

Recall: Object Recognition Problem

- What object is in the image?

Recall: Catalyst for Computer Vision Revolution

1945 1957 1966 1983 1987 1990 1998 2009

Computer Vision

CVPR ICCV ECCV MNIST ImageNet

Large-scale dataset for recognizing objects contained in 3,200,000 images

J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. 2009.

Recall: Catalyst for Computer Vision Revolution

Progress of models on ImageNet (Top 5 Error)

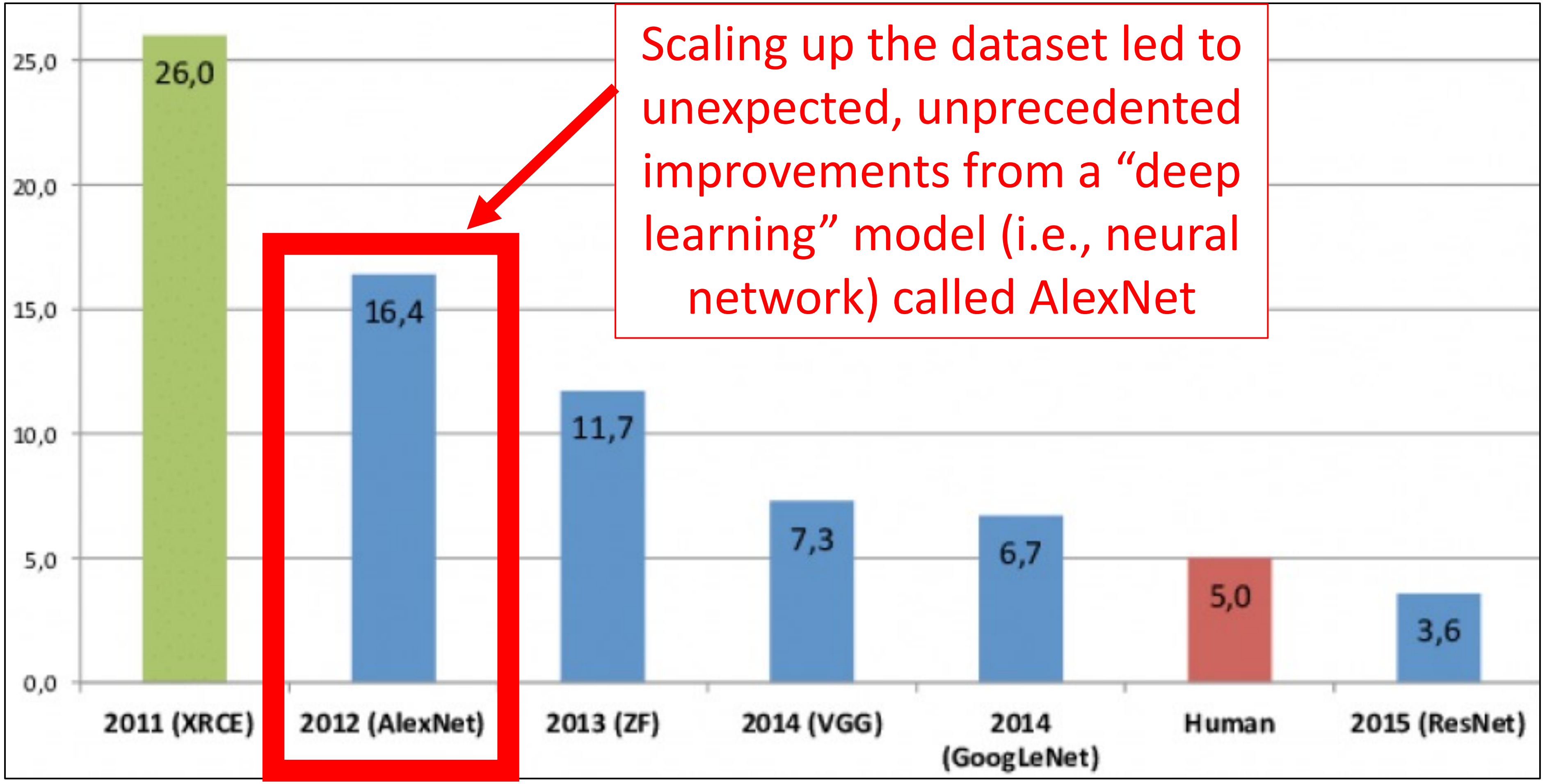
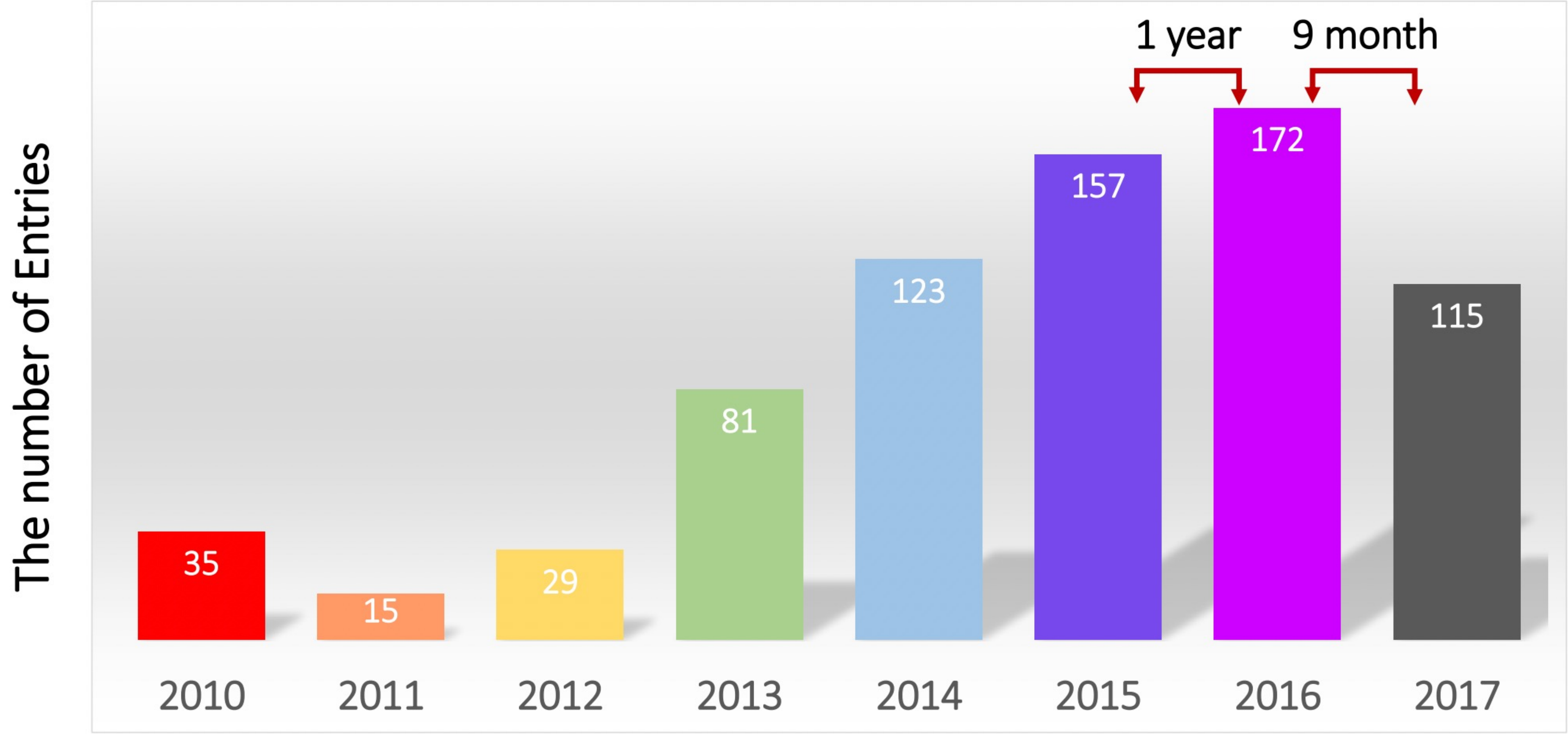


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

Recall: ImageNet Challenge Submissions



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

What Was The Secret Sauce To Be State-of-Art?

Progress of models on ImageNet (Top 5 Error)

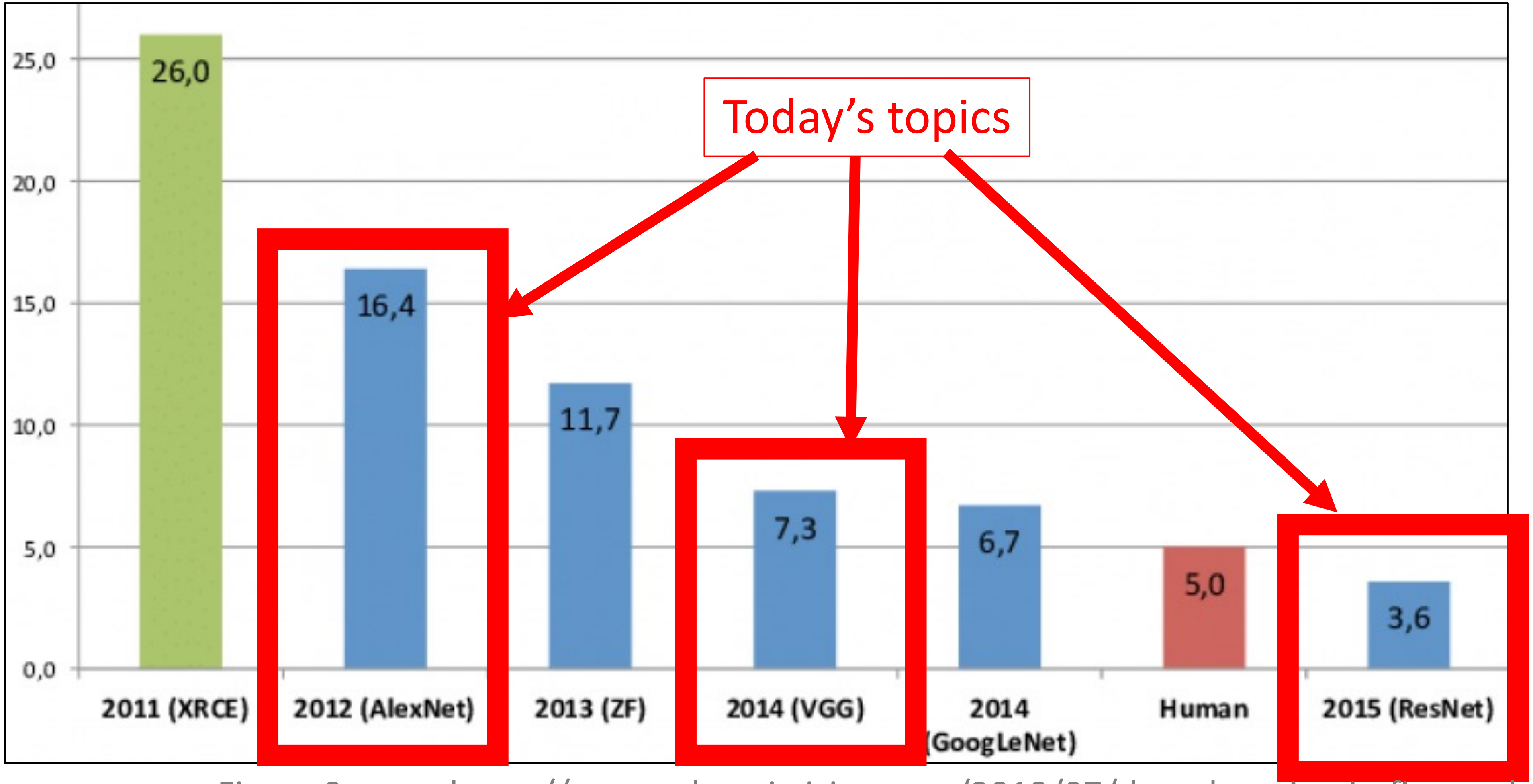


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

Object Recognition: Today’s Topics

- ImageNet Challenge Top Performers
- Baseline Model: AlexNet
- VGG
- ResNet
- Discussion

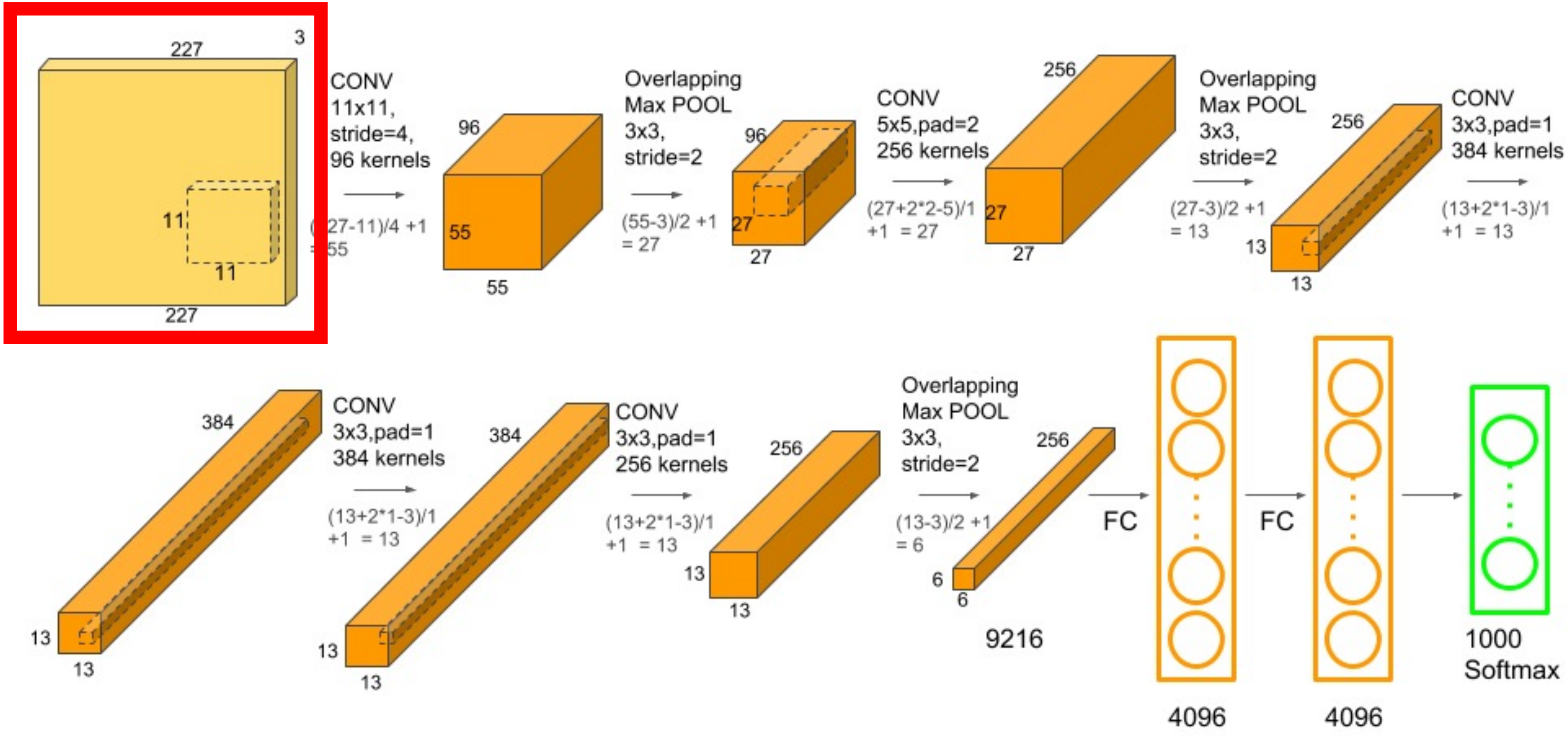
Why AlexNet?

Alex is the name of the paper's author ☺

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems (2012).*

Input: RGB image resized to fixed input size

Architecture



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

Architecture: Input Resizing

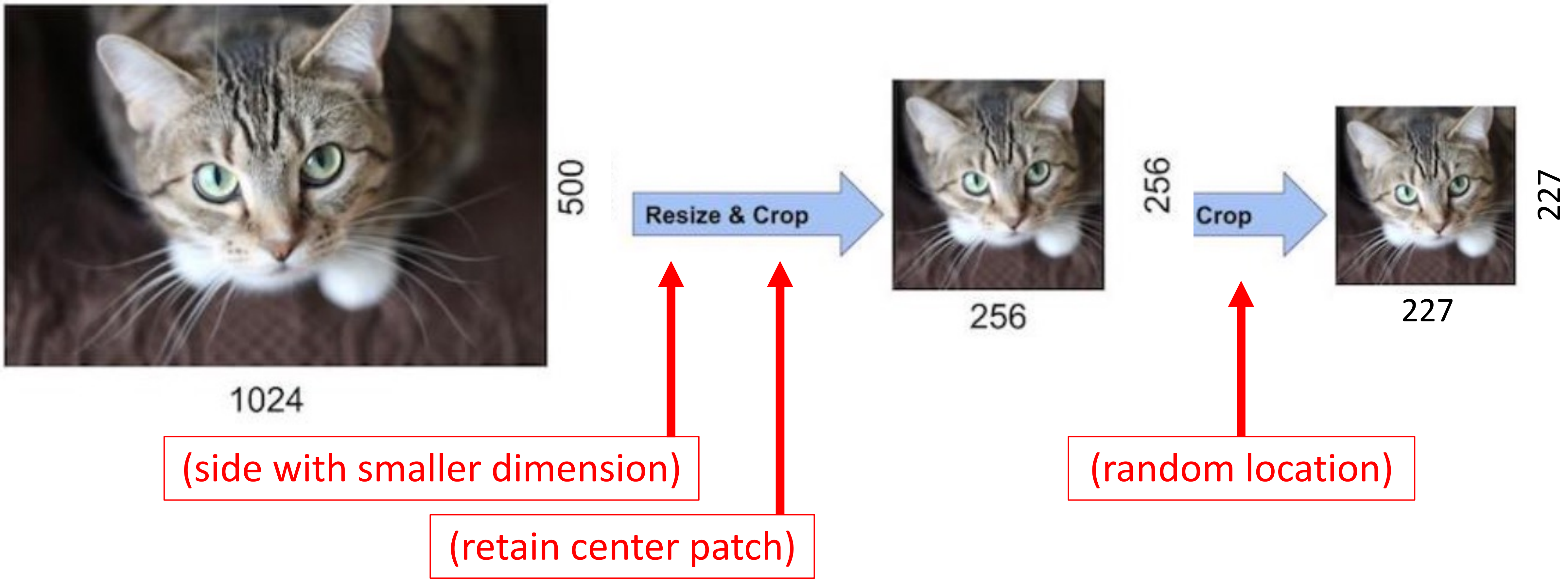
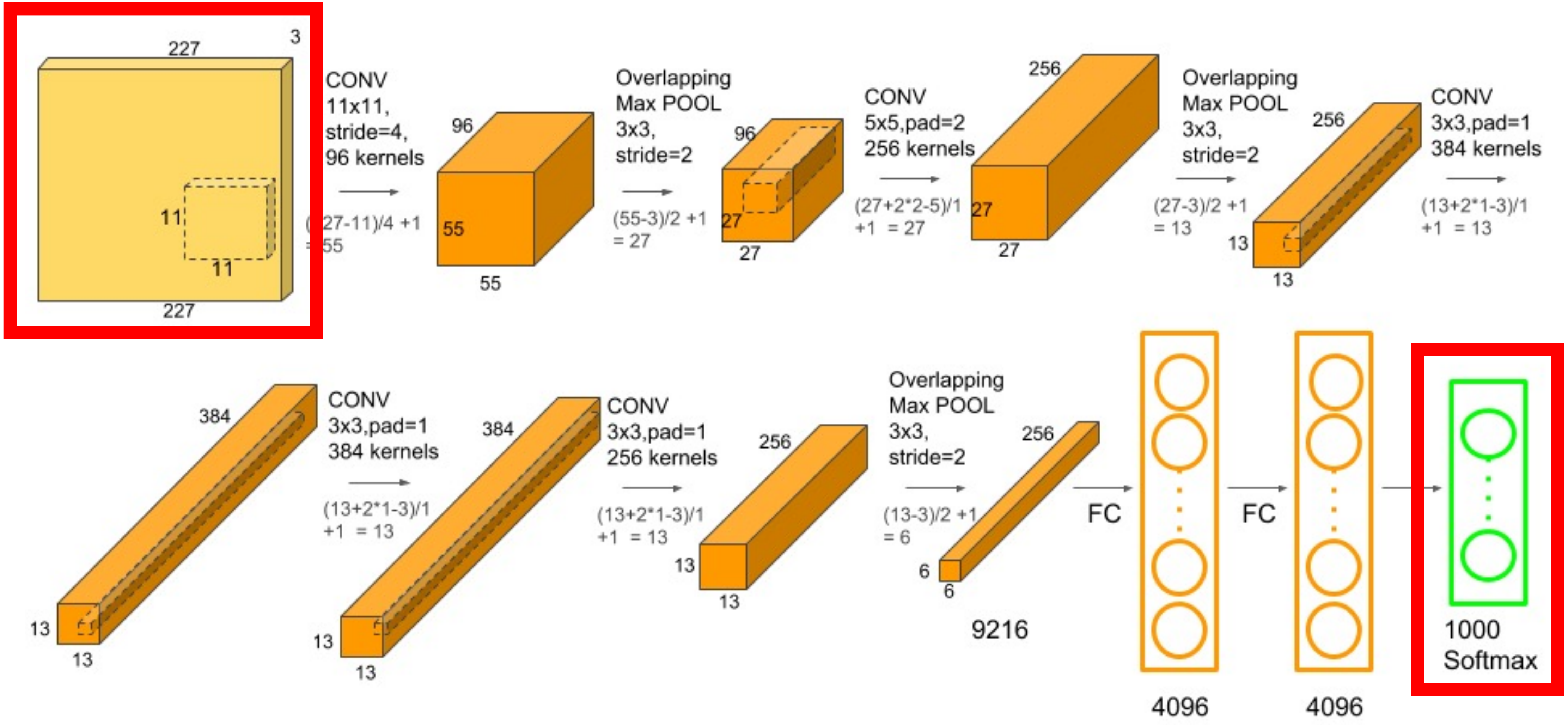


Image Source: https://learnopencv.com/understanding-alexnet/

Input: RGB image resized to fixed input size

Output: 1000 class probabilities (sums to 1)

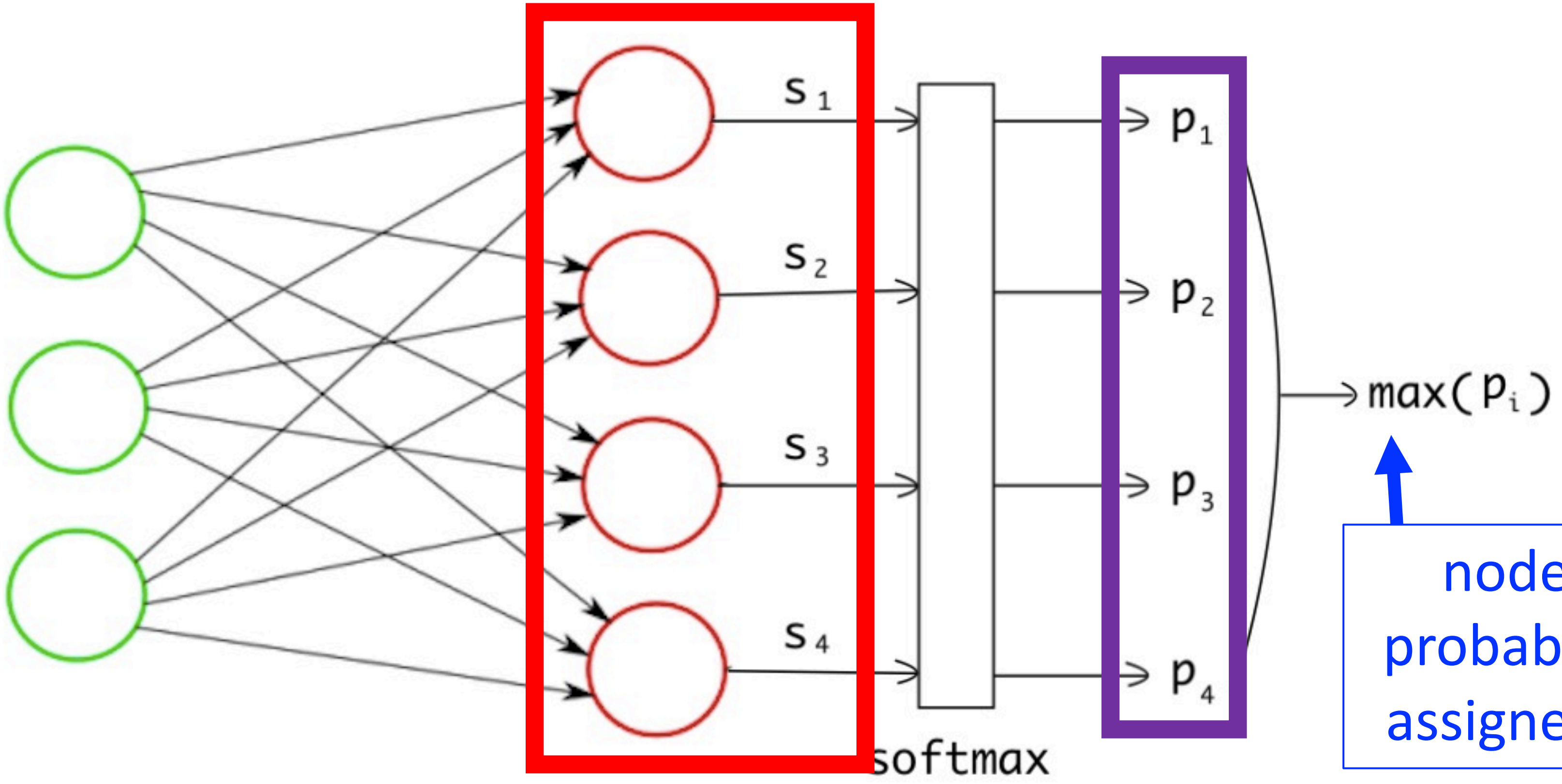
Architecture



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

Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1; e.g.,



node with largest probability is the class assigned to the image

Figure Source: https://towardsdatascience.com/multi-label-image-classification-with-neural-network-keras-ddc1ab1afede

Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1; e.g.,

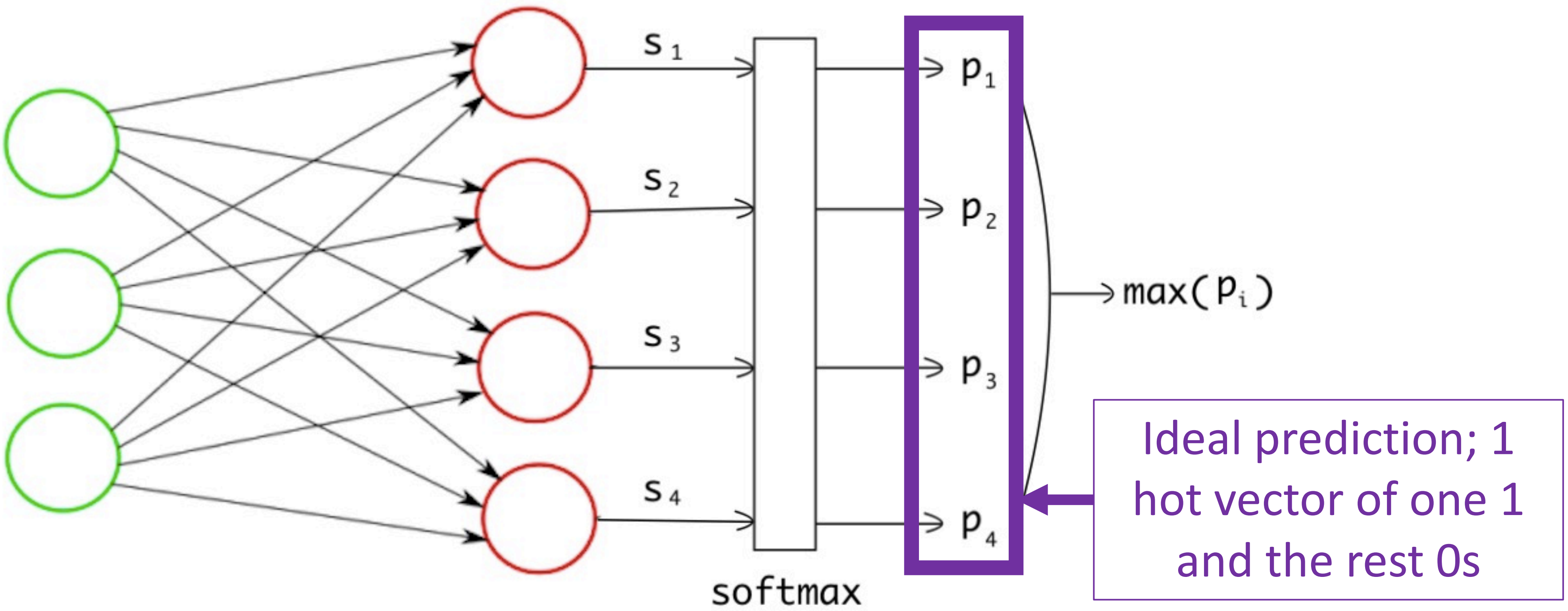


Figure Source: https://towardsdatascience.com/multi-label-image-classification-with-neural-network-keras-ddc1ab1afede

Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1

To do so, must get rid of negative values while preserving original order of scores; e causes negative scores to become slightly larger than 0 while positive values grow exponentially; choosing *e* rather than another exponent base simplifies subsequent math

$$\sigma(\mathbf{z})_i = \frac{\boxed{e}^{\boxed{z_i}}}{\boxed{\sum_{j=1}^{K} \boxed{e}^{z_j}}}$$

i = 1, …, K

Number of classes (i.e., 1000)

Want to divide each node's score by sum of all entries to make them sum to 1 (normalization)

Useful tutorial: https://towardsdatascience.com/exploring-the-softmax-function-578c8b0fb15

Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1; e.g.,

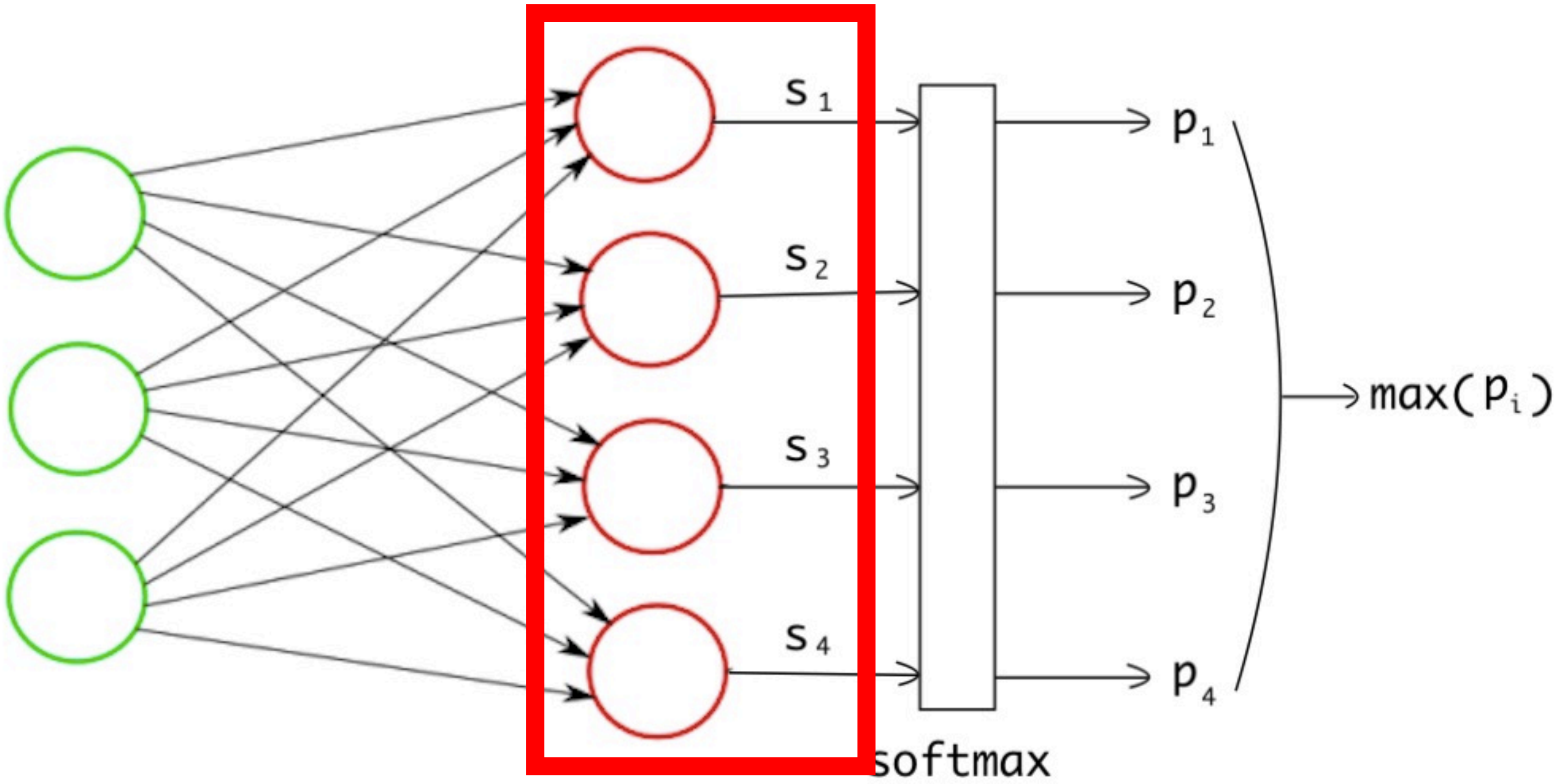


Figure Source: https://towardsdatascience.com/multi-label-image-classification-with-neural-network-keras-ddc1ab1afede

Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1; e.g.,

	Scoring Function
Dog	-3.44
Cat	1.16
Boat	-0.81
Airplane	3.91

Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1; e.g.,

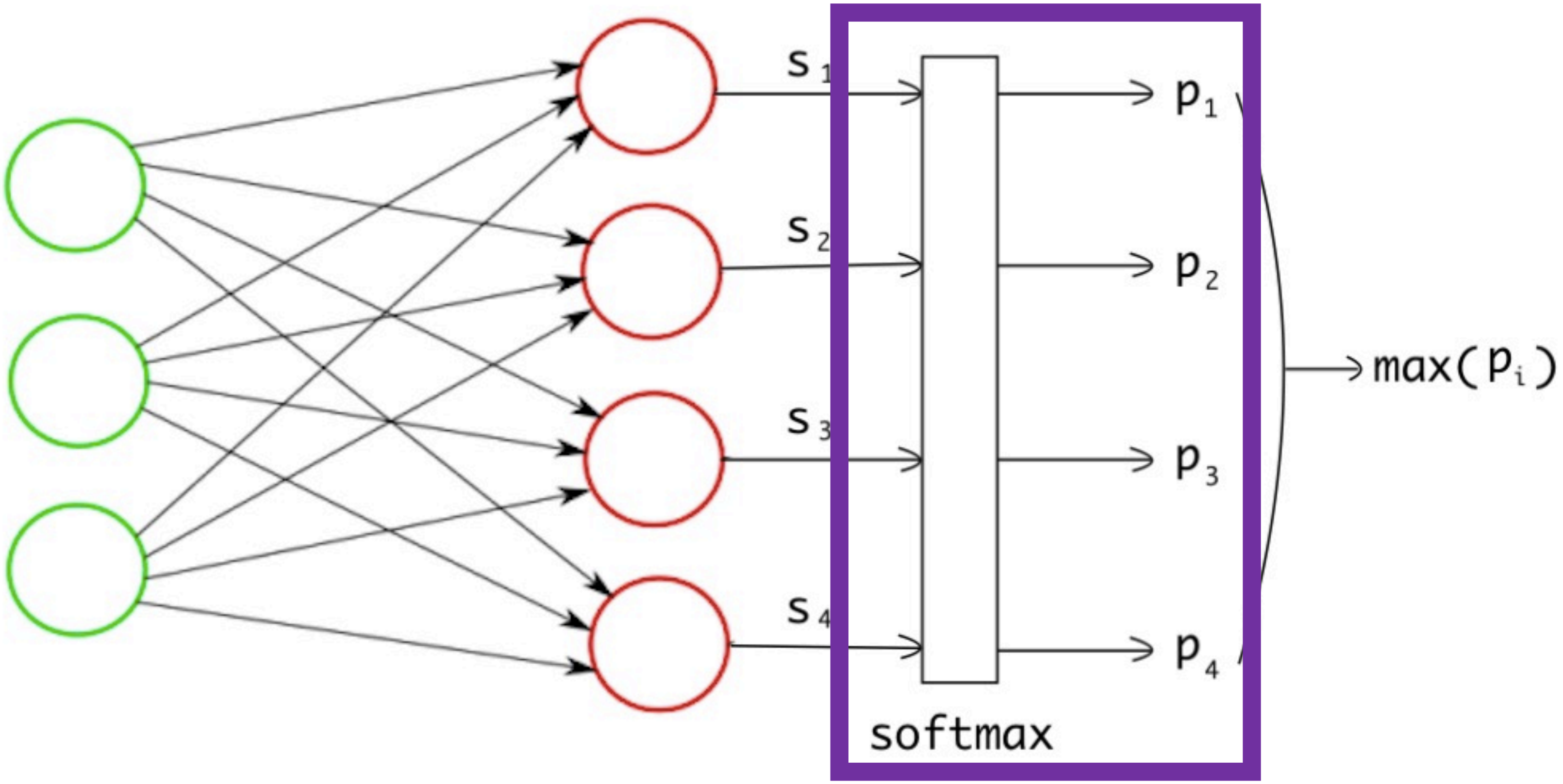


Figure Source: https://towardsdatascience.com/multi-label-image-classification-with-neural-network-keras-ddc1ab1afede

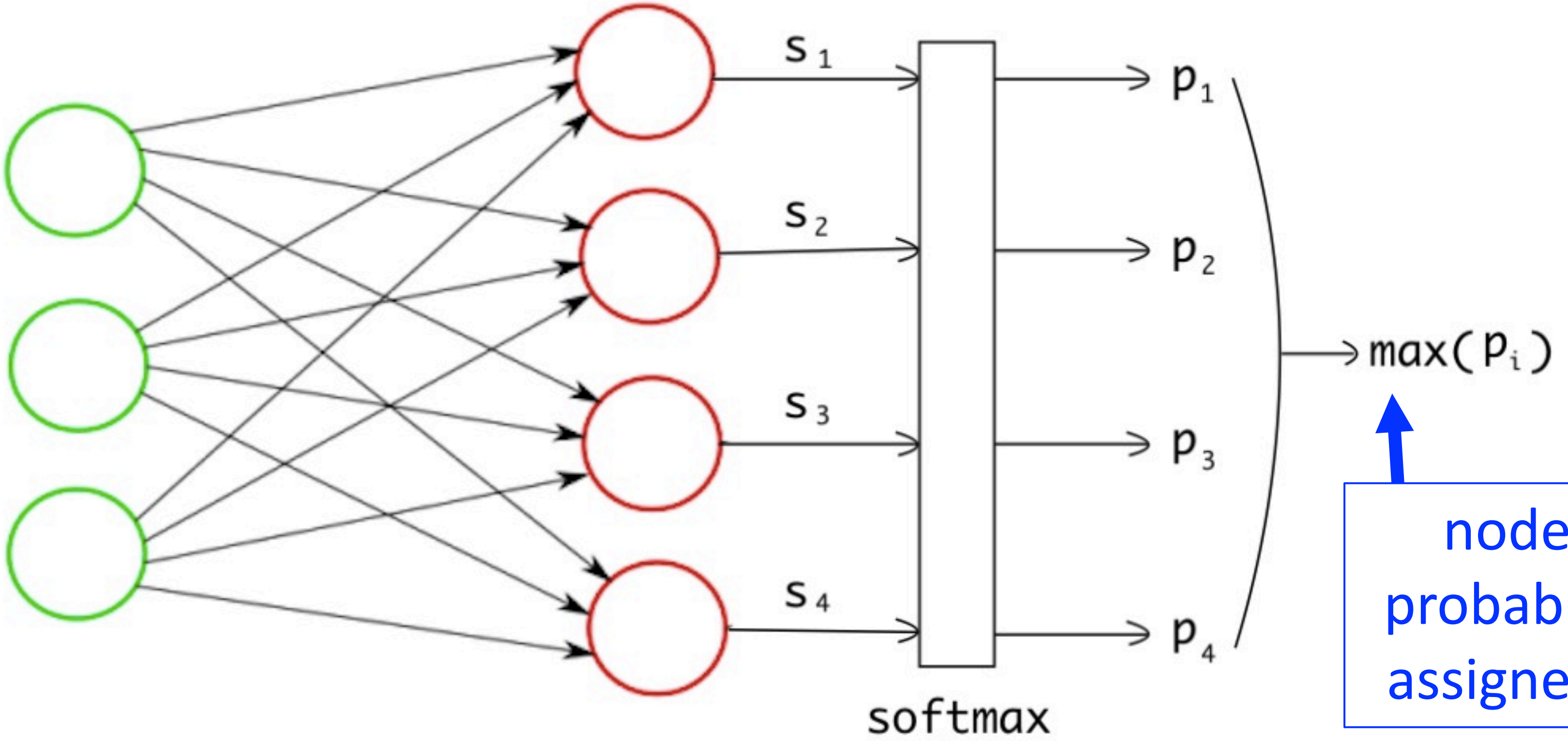
Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1; e.g.,

	Scoring Function	Unnormalized Probabilities e^{z_i}	Normalized Probabilities $\frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$
Dog	-3.44	0.0321	0.0006
Cat	1.16	3.1899	0.0596
Boat	-0.81	0.4449	0.0083
Airplane	3.91	49.8990	0.9315

Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1; e.g.,



node with largest probability is the class assigned to the image

Figure Source: https://towardsdatascience.com/multi-label-image-classification-with-neural-network-keras-ddc1ab1afede

Architecture: Output Softmax Layer (Multiclass Classification)

Softmax: converts vector of scores into a probability distribution that sums to 1; e.g.,

	Scoring Function	Unnormalized Probabilities	Normalized Probabilities
Dog	-3.44	0.0321	0.0006
Cat	1.16	3.1899	0.0596
Boat	-0.81	0.4449	0.0083
Airplane	3.91	49.8990	0.9315

Architecture

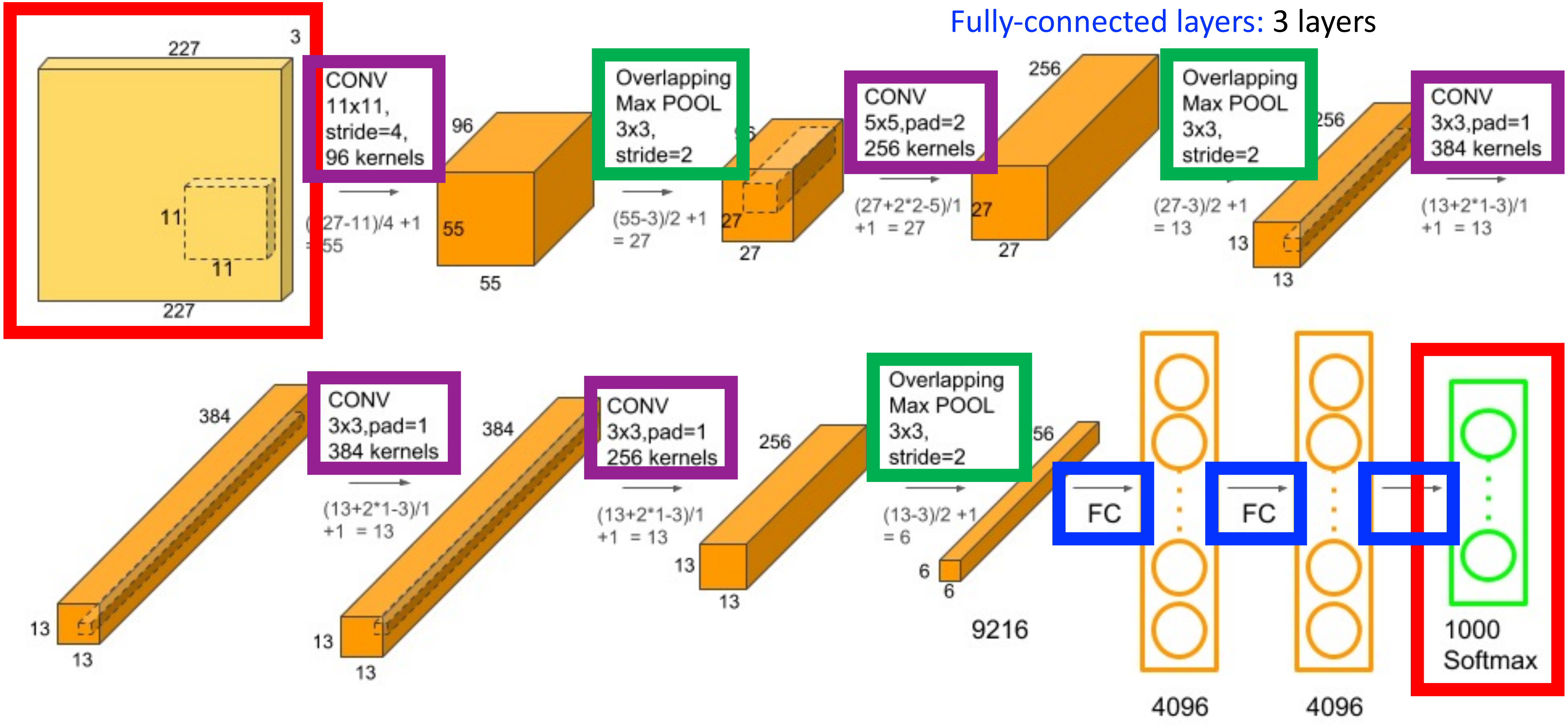
Input: RGB image resized to fixed input size

Output: 1000 class probabilities (sums to 1)

Convolutional layers: 5 layers

Pooling Layers: 3 layers

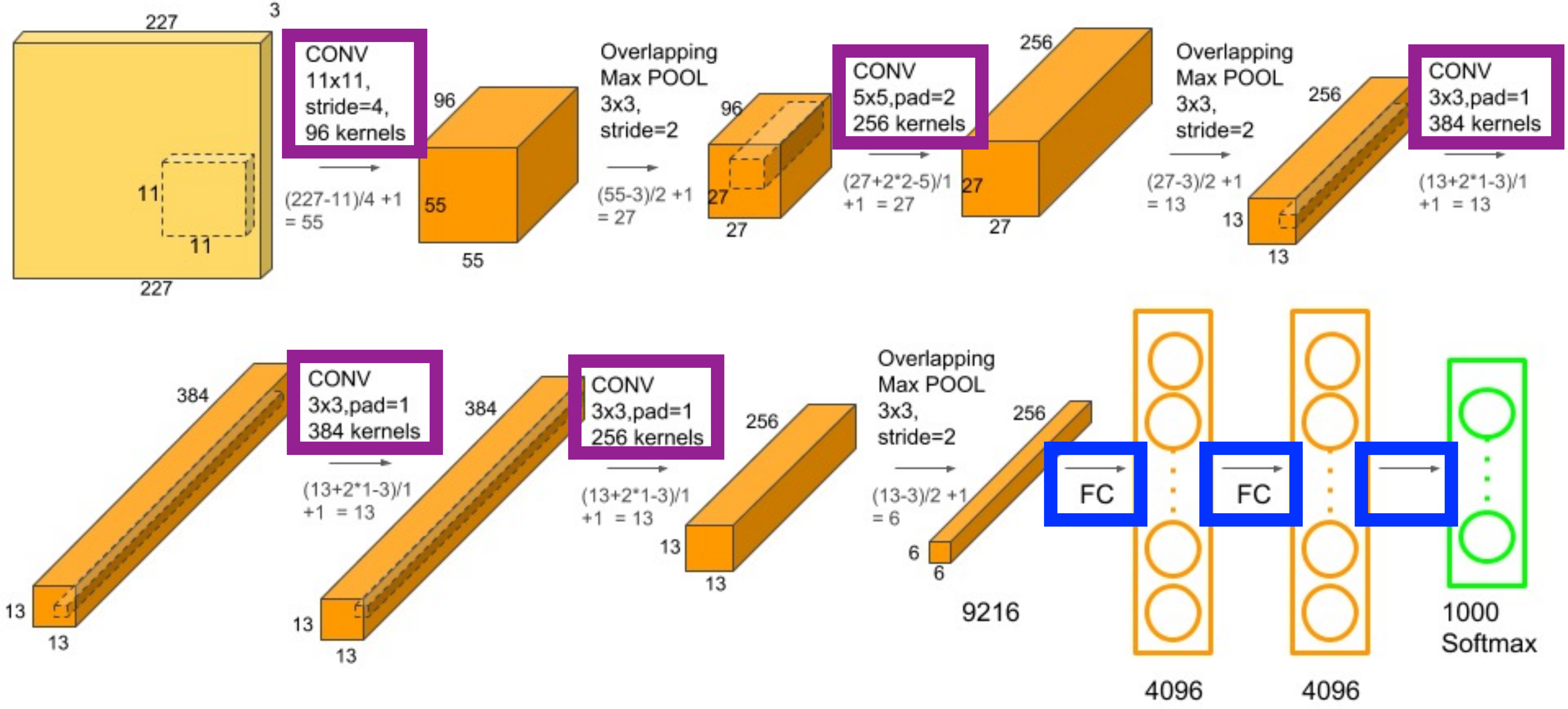
Fully-connected layers: 3 layers



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

Architecture

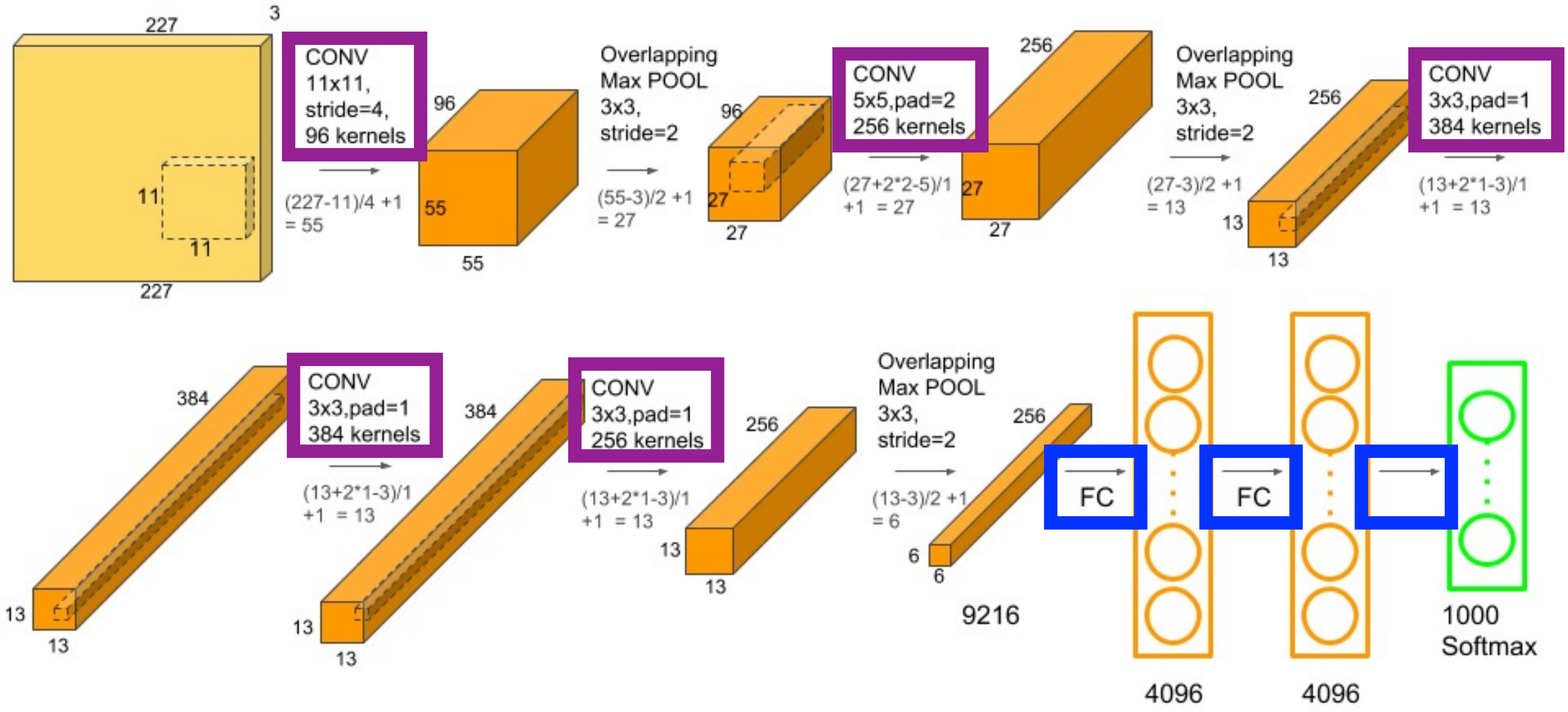
How many layers have model parameters that need to be learned?



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

Architecture

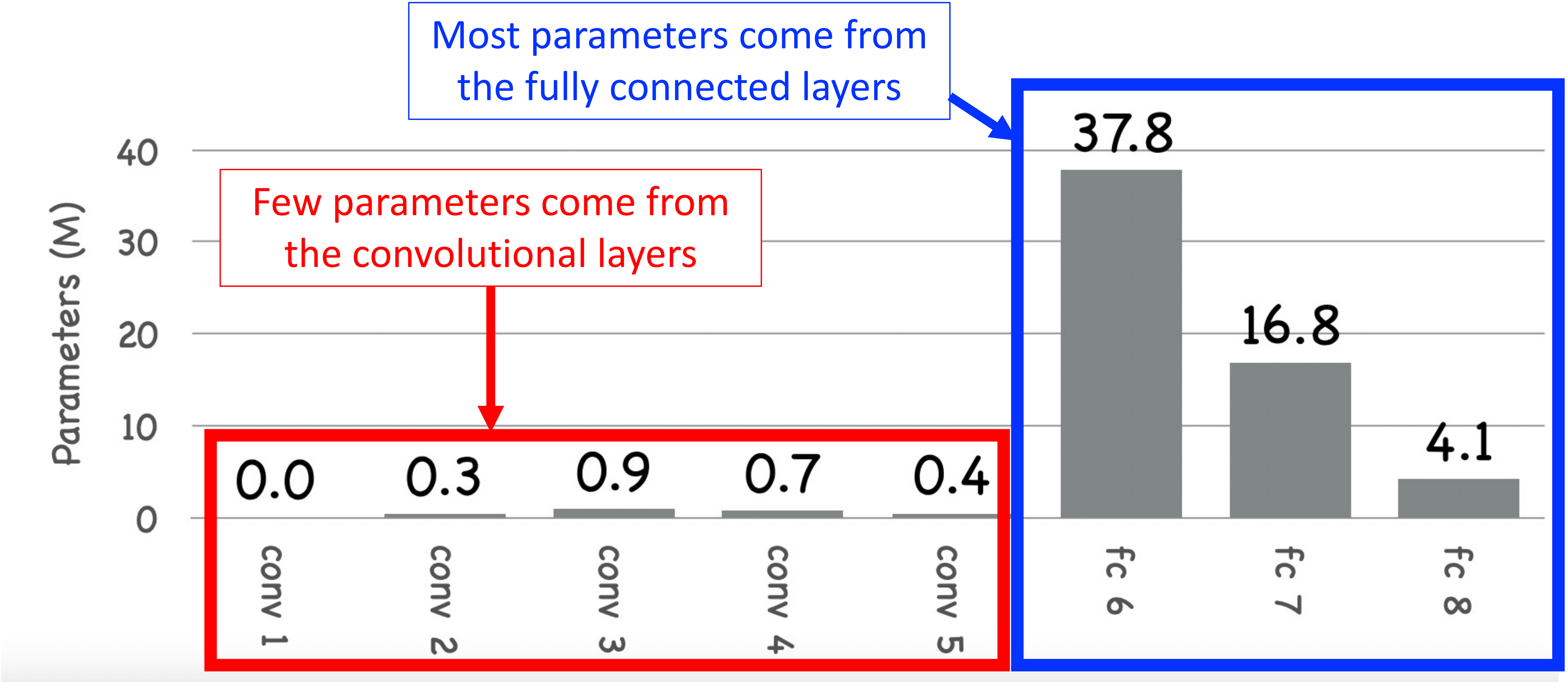
Altogether, 60 million model parameters must be learned!



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

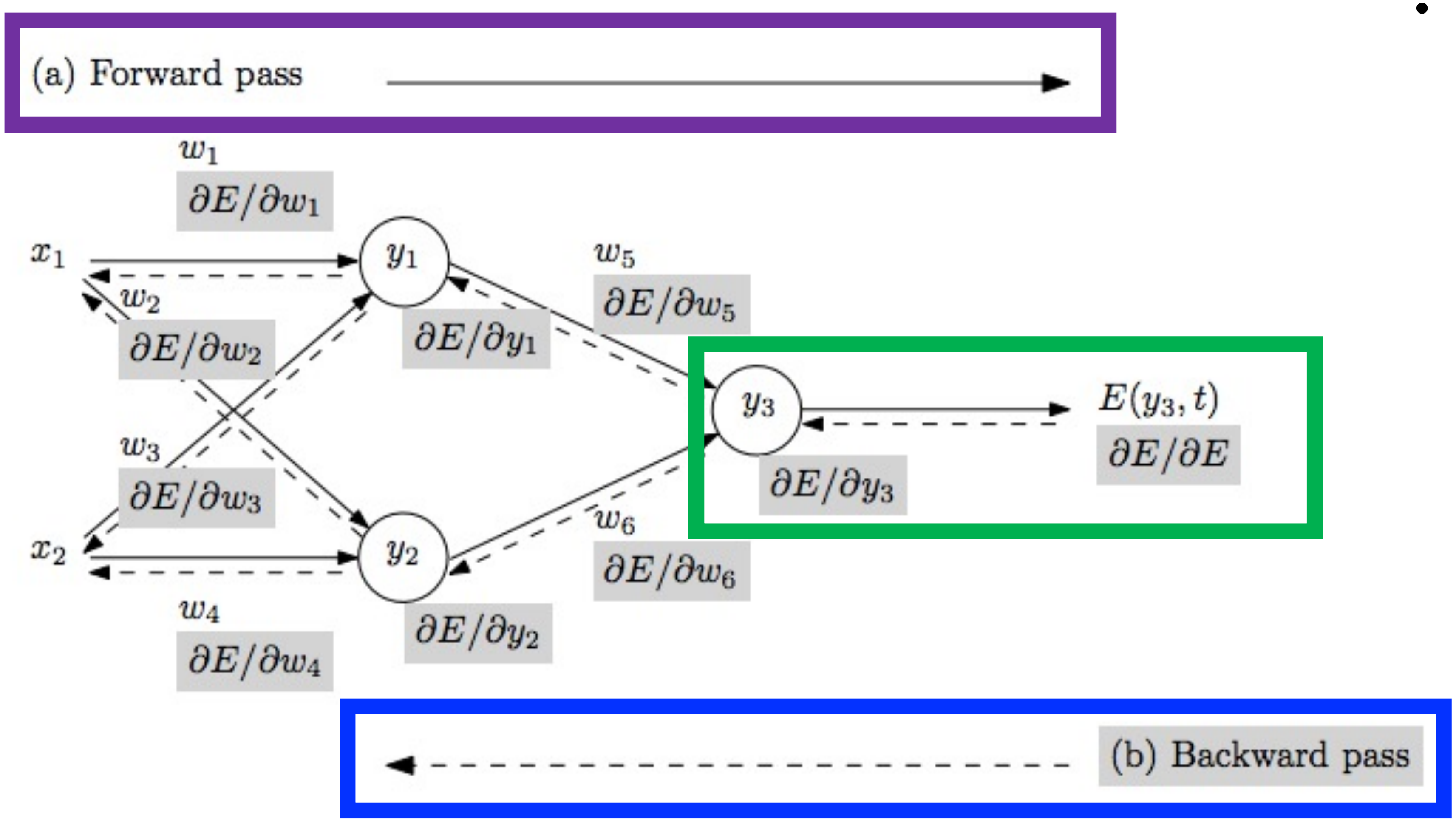
Architecture

Altogether, 60 million model parameters must be learned!



Source: http://www.philkr.net/cs342/lectures/computer_vision/02.pdf

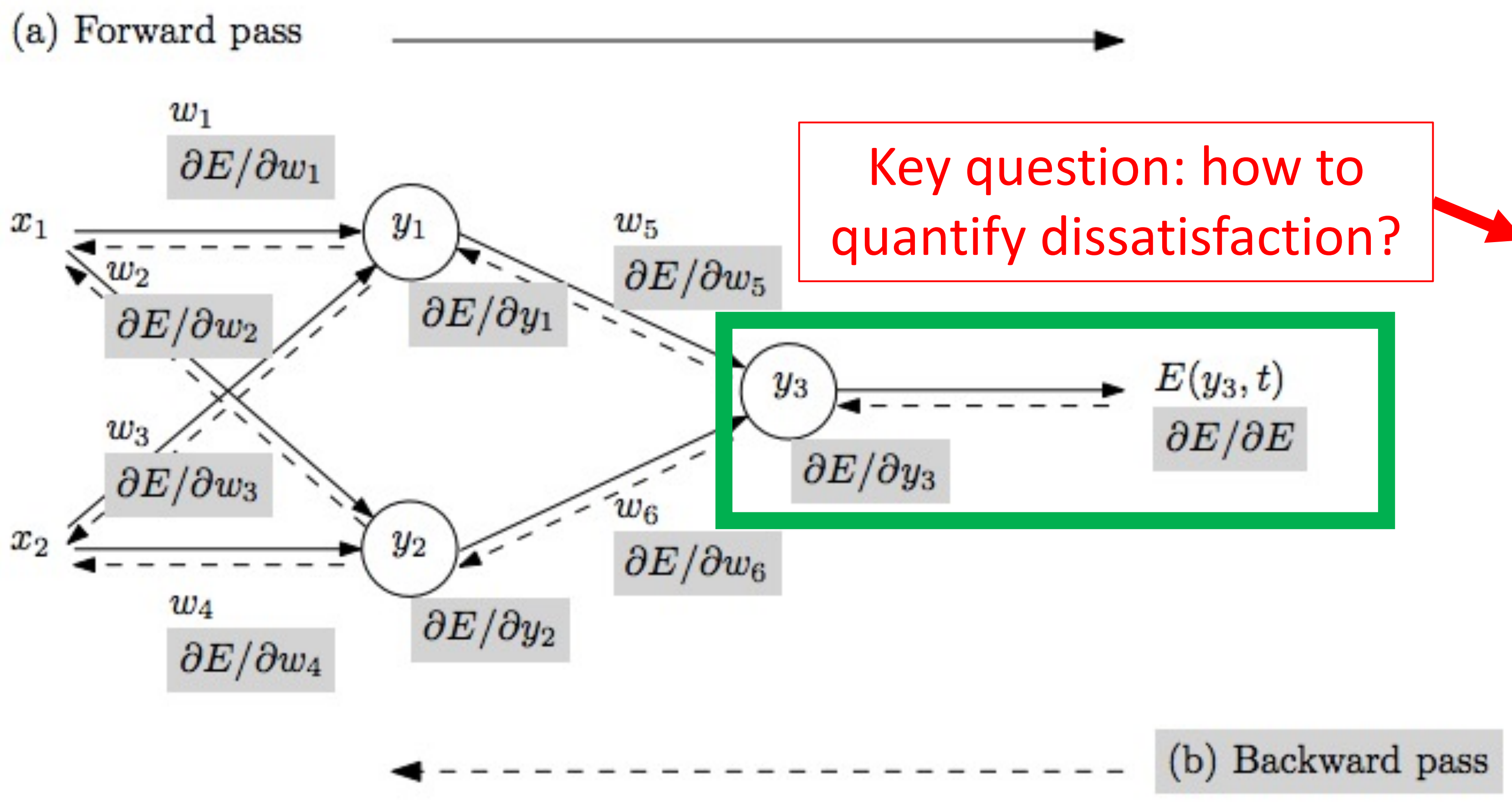
Algorithm Training: Recall How NNs Learn



- 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

Algorithm Training



Key question: how to quantify dissatisfaction?

- 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

Algorithm Training: Measure Cross Entropy Loss

Measure distance between predicted and true class distribution for each example

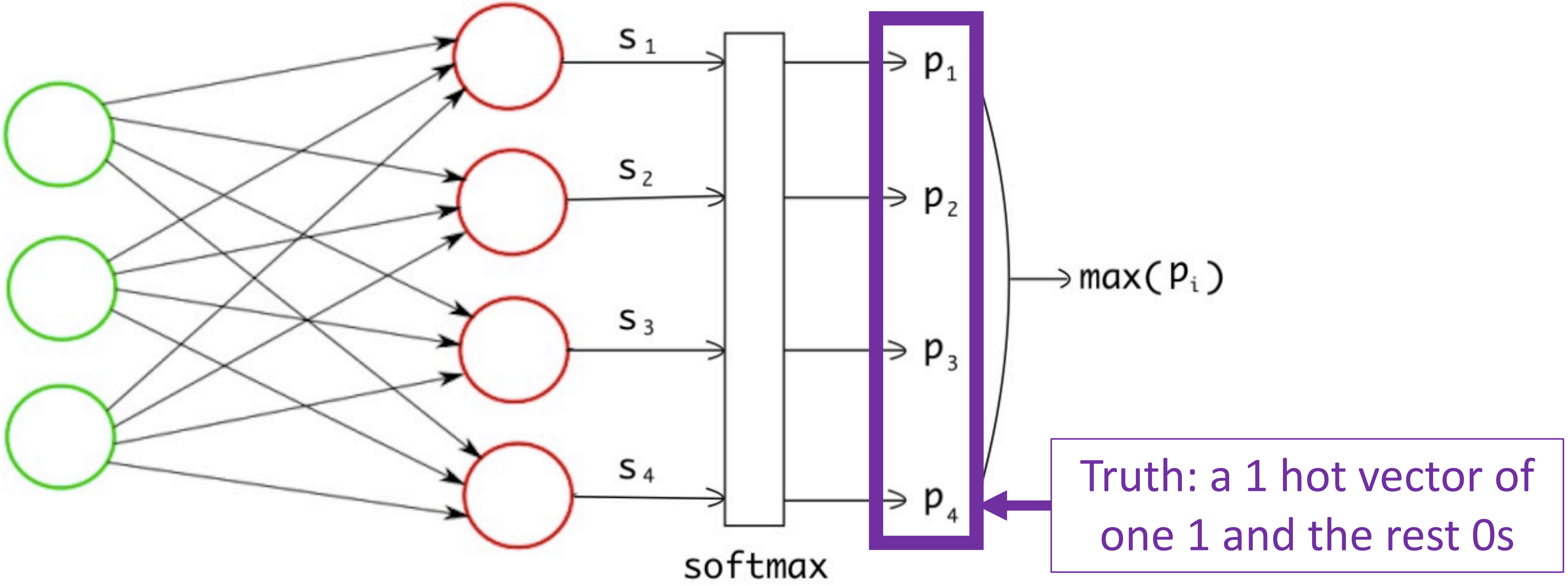


Figure Source: https://towardsdatascience.com/multi-label-image-classification-with-neural-network-keras-ddc1ab1afede

Algorithm Training: Measure Cross Entropy Loss

Probability distribution of true class

Probability distribution of predicted class

Number of classes?

Recall, truth is set to 1 for one class and 0 otherwise

Observed features

$$
\begin{aligned}
L_{\mathrm{CE}}(\hat{y}, y) &= -\sum_{k=1}^{K} y_k \log \hat{y}_k \\
&= -\sum_{k=1}^{K} y_k \log \hat{p}(y=k|x) \\
&= -\log \hat{y}_k, \quad \text{(where } k \text{ is the correct class)} \\
&= -\log \frac{\exp(w_k \cdot x + b_k)}{\sum_{j=1}^{K} \exp(w_j \cdot x + b_j)}
\end{aligned}
$$

Simplifies to the log of the predicted probability for the correct class (i.e., negative log likelihood loss)

Excellent background: https://web.stanford.edu/~jurafsky/slp3/5.pdf

Algorithm Training: Measure Cross Entropy Loss

Probability distribution of predicted class

Probability distribution of true class

Number of classes?

Recall, truth is set to 1 for one class and 0 otherwise

Observed features

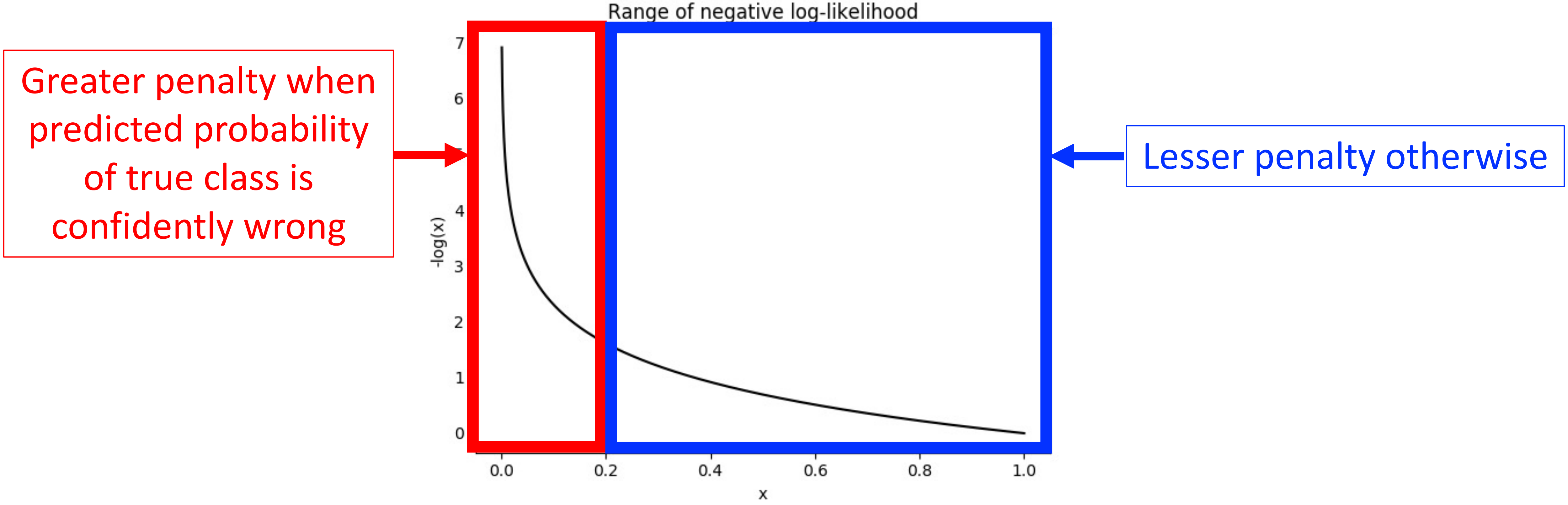
$$
\begin{aligned}
L_{\mathrm{CE}}(\hat{y}, y) &= -\sum_{k=1}^{K} y_k \log \hat{y}_k \\
&= -\sum_{k=1}^{K} y_k \log \hat{p}(y = k|x) \\
&= -\log \hat{y}_k, \quad \text{(where } k \text{ is the correct class)} \\
&= -\log \frac{\exp(w_k \cdot x + b_k)}{\sum_{j=1}^{K} \exp(w_j \cdot x + b_j)}
\end{aligned}
$$

What is the range of possible values?

- Minimum: 0 (negative log of 1)
- Maximum: Infinity (negative log of 0)

Excellent background: https://web.stanford.edu/~jurafsky/slp3/5.pdf

Algorithm Training: Measure Cross Entropy Loss



What is the range of possible values?

- Minimum: 0 (negative log of 1)
- Maximum: Infinity (negative log of 0)

$$= -\log \frac{\exp\left(w_k \cdot x + b_k\right)}{\sum_{j=1}^{K} \exp\left(w_j \cdot x + b_j\right)}$$

Figure source: https://ljvmiranda921.github.io/notebook/2017/08/13/softmax-and-the-negative-log-likelihood/

Algorithm Training: Measure Cross Entropy Loss

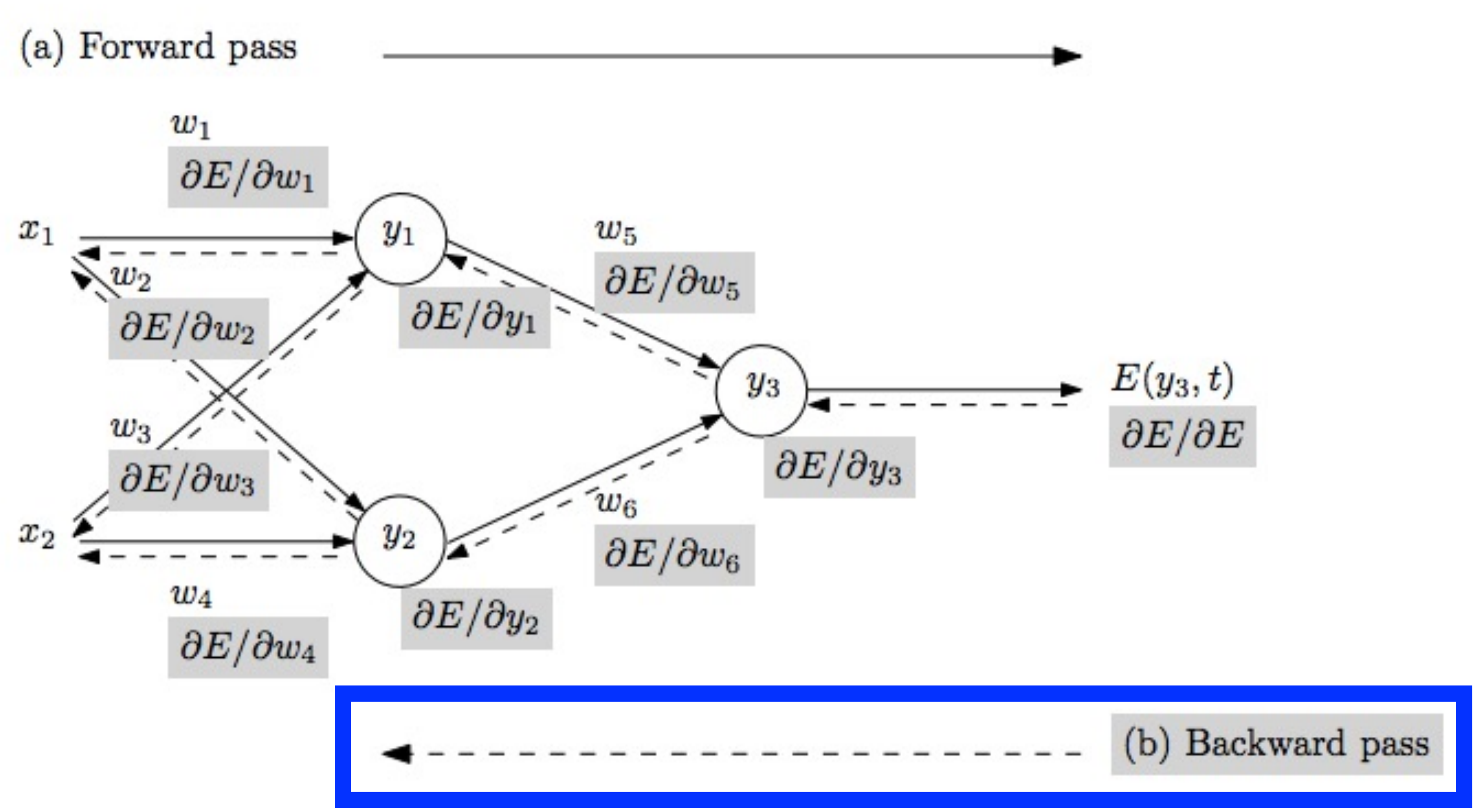
e.g., What would be the loss for this example if the true class label is cat?

$$= -\log \hat{y}_k, \quad \text{(where } k \text{ is the correct class)}$$

$$= -\log (0.0596) = 2.82$$

	Scoring Function	Unnormalized Probabilities	Normalized Probabilities
Dog	-3.44	0.0321	0.0006
Cat	1.16	3.1899	0.0596
Boat	-0.81	0.4449	0.0083
Airplane	3.91	49.8990	0.9315

Algorithm Training: Challenge



- 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

Algorithm Training: Challenge Is Overfitting

- Idea: which is a better model to separate blue from red: the green or **black** line?

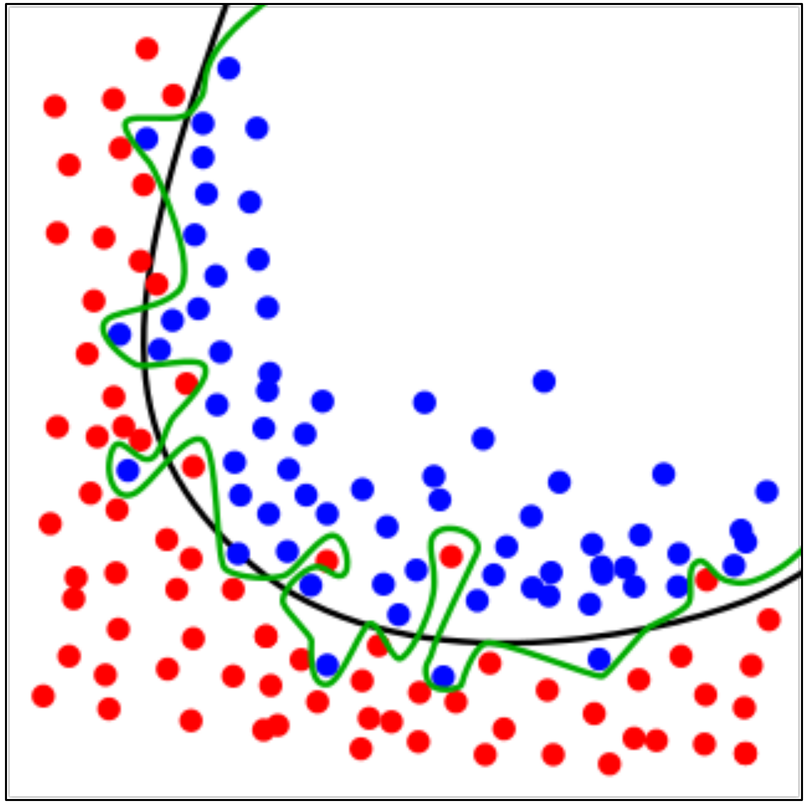
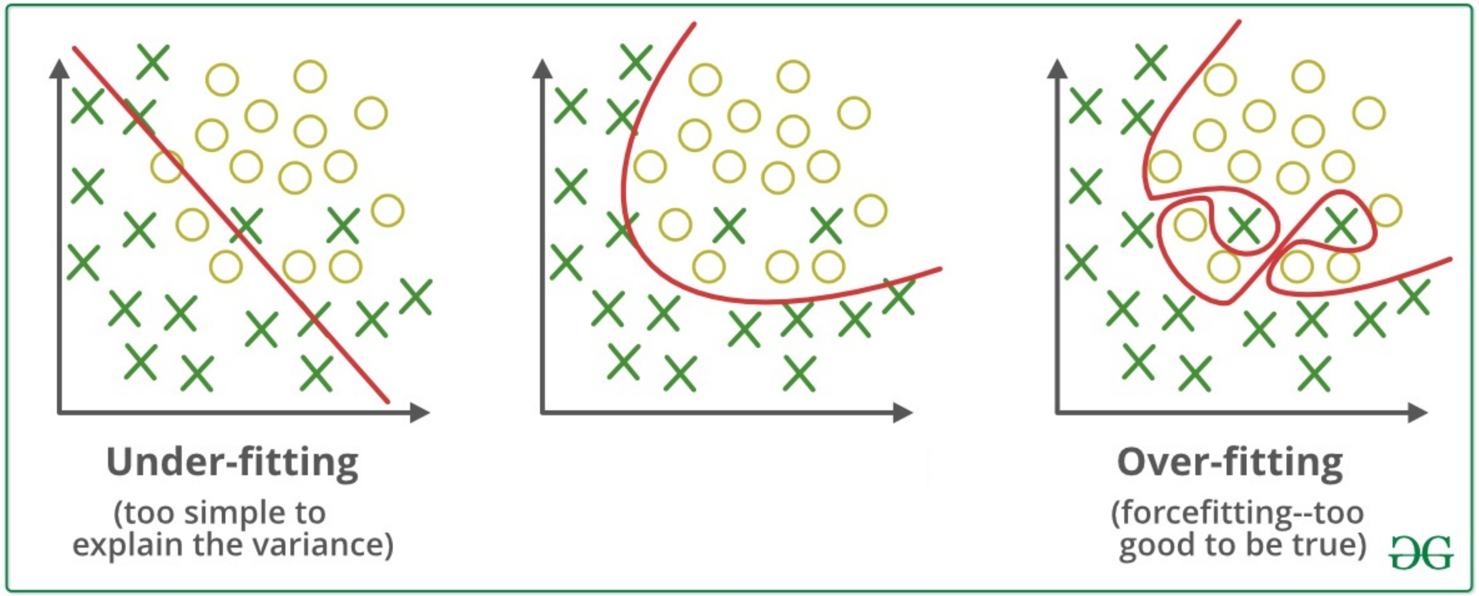


Figure source: https://towardsdatascience.com/underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6fe4a8a49dbf

Algorithm Training: Challenge Is Overfitting

- Overfitting is risk for models with larger representational capacity (i.e., # of parameters); AlexNet has 60 million parameters!



- Model learns to model **noise!** What would cause noise in a dataset?

Figure source: https://towardsdatascience.com/underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6fe4a8a49dbf

Algorithm Training: Challenge Is Overfitting

- Overfitting is risk for models with larger representational capacity (i.e., # of parameters); AlexNet has 60 million parameters!
- How to detect overfitting: plot error/loss for models tested on training data and test data
  - What happens to training data error as number of training steps increases?
    - Error shrinks
  - What happens to test data error as number of training steps increases?
    - Error shrinks and then grows
  - Why does train error ***shrink*** and test error ***grow***?
    - The model is learning to model ***noise*** in the training data (i.e., "overfit")! Models capturing noise perform well on training data while generalizing poorly to new test data

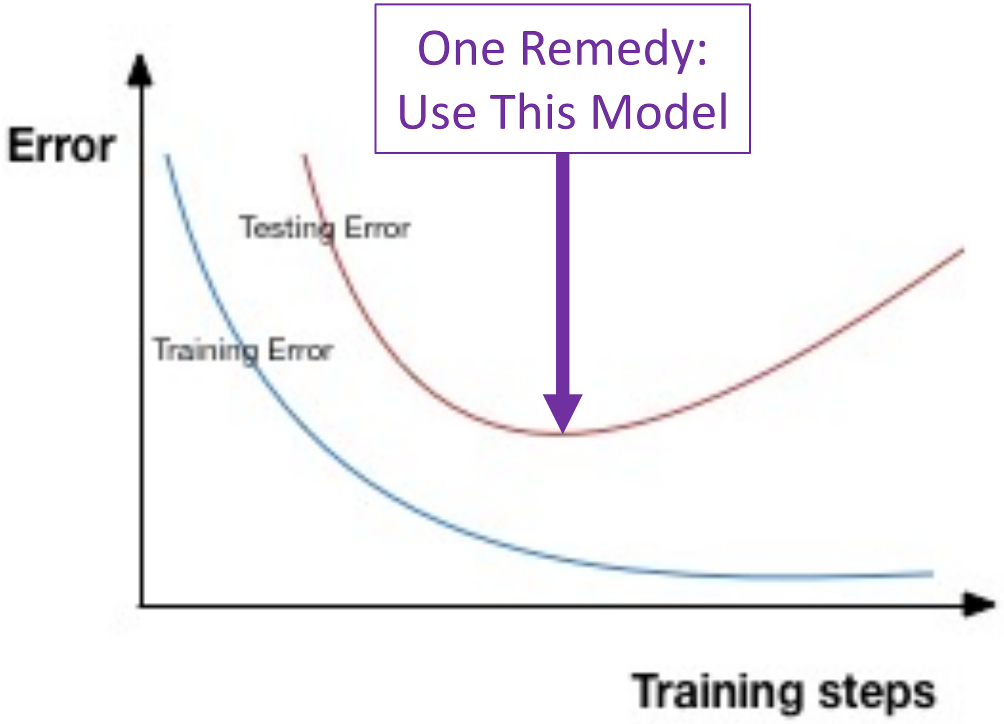
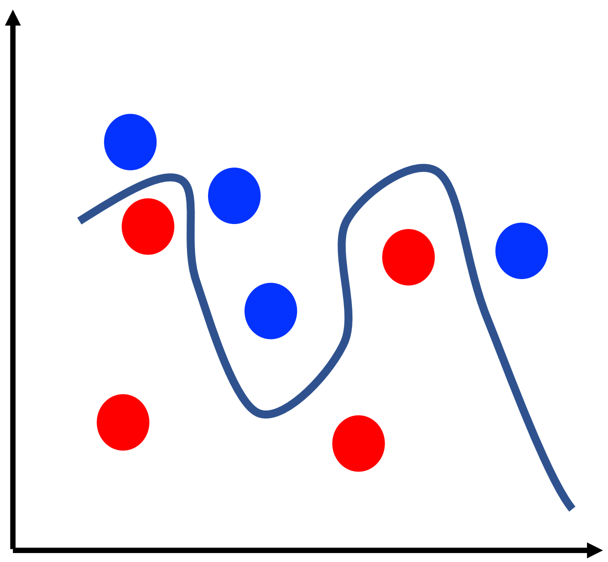
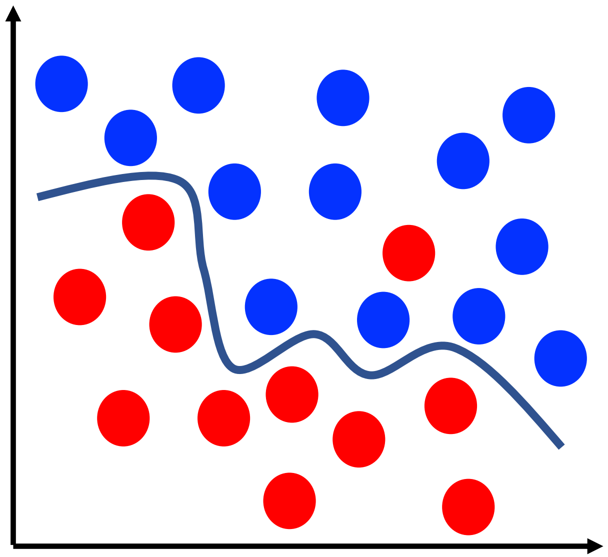


Image Source: https://chatbotslife.com/regularization-in-deep-learning-f649a45d6e0

AlexNet Remedies for Overfitting

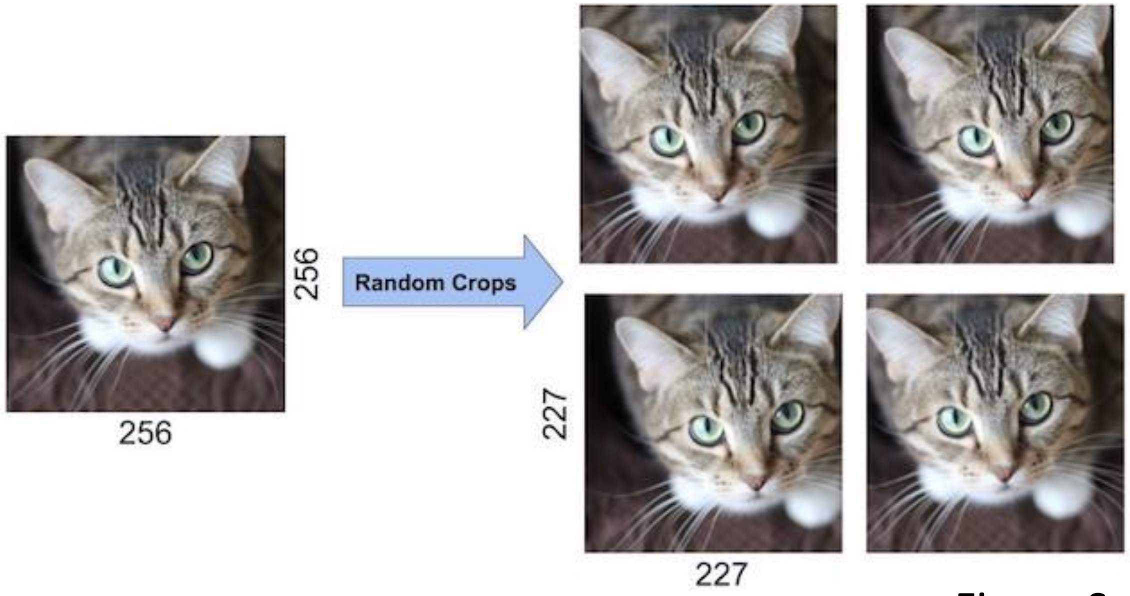
- Overfitting is risk for models with larger representational capacity (i.e., # of parameters); AlexNet has 60 million parameters!
    1. Data augmentation: add more training data; e.g., intuitively,





AlexNet Remedies for Overfitting

- Overfitting is risk for models with larger representational capacity (i.e., # of parameters); AlexNet has 60 million parameters!
  1. Data augmentation
     1. Random patches and their mirror images (2048x more data)
     2. Adjust RGB channels (using PCA to add multiples of principal components)



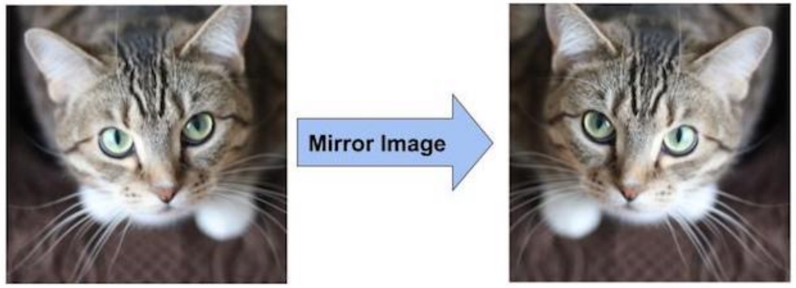
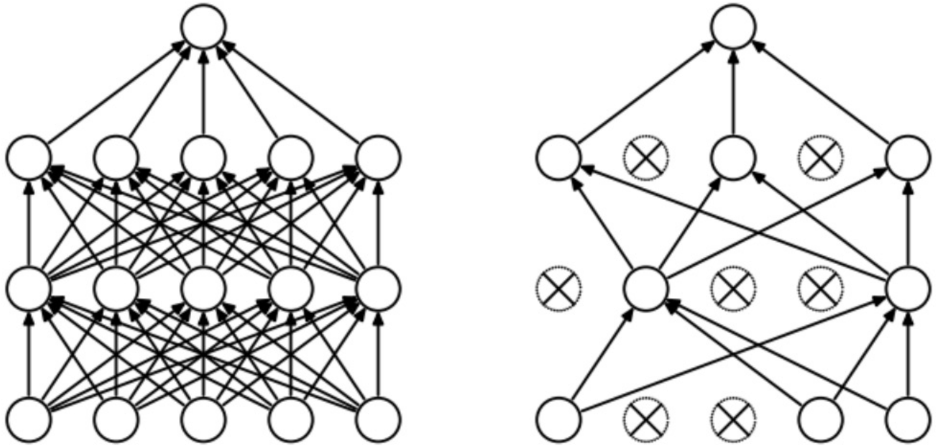


Figure Source: https://learnopencv.com/understanding-alexnet/

AlexNet Remedies for Overfitting

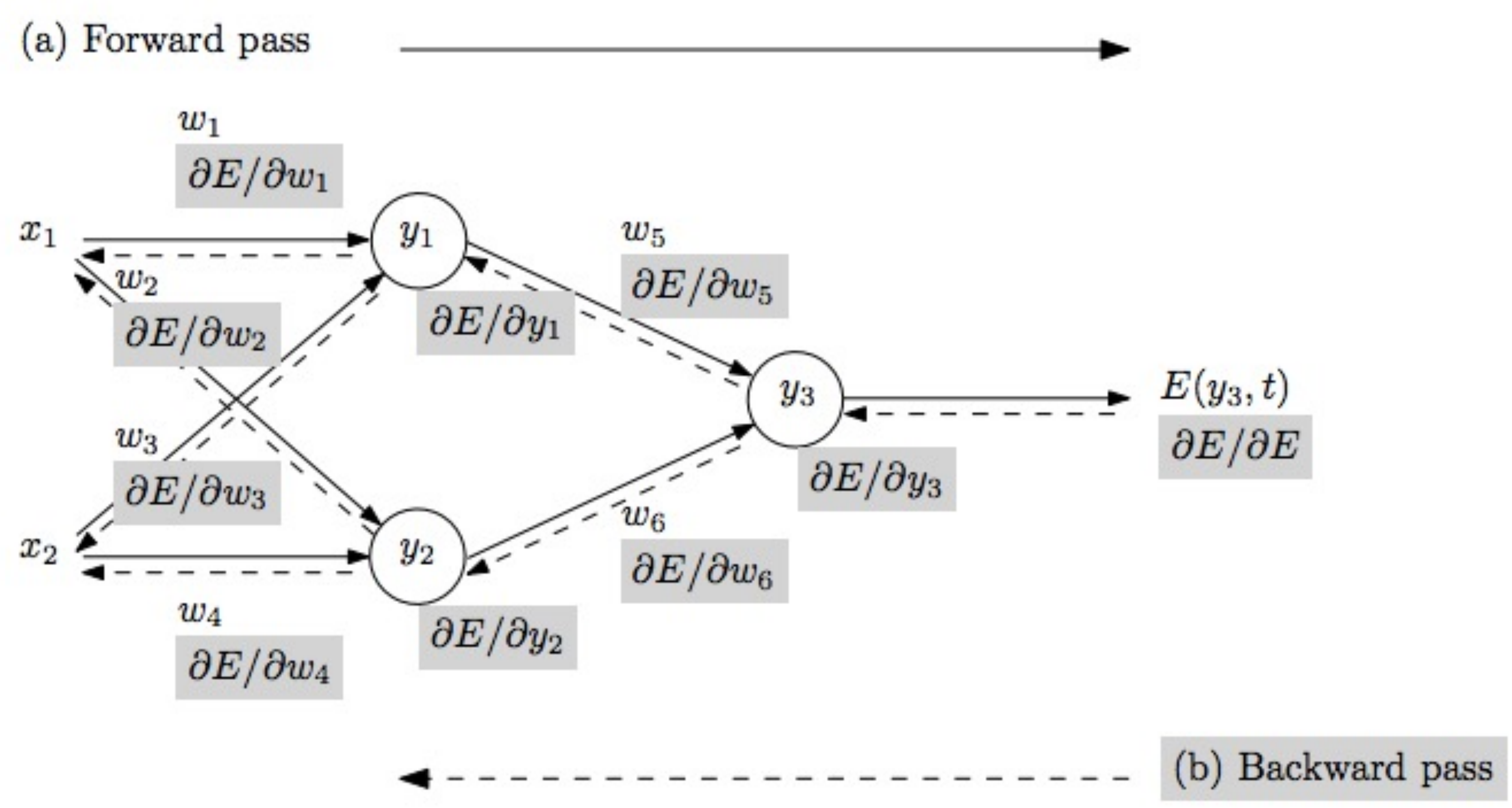
- Overfitting is risk for models with larger representational capacity (i.e., # of parameters); AlexNet has 60 million parameters!
  1. Data augmentation
     1. Random patches and their mirror images (2048x more data)
     2. Adjust RGB channels (using PCA to add multiples of principal components)
  2. Dropout (50% of nodes for first two fully connected layers); mimics ensembles by learning to solve same problem with different subnetworks



(a) Standard Neural Net (b) After applying dropout.

Figure Source: Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." JMLR, 2014.

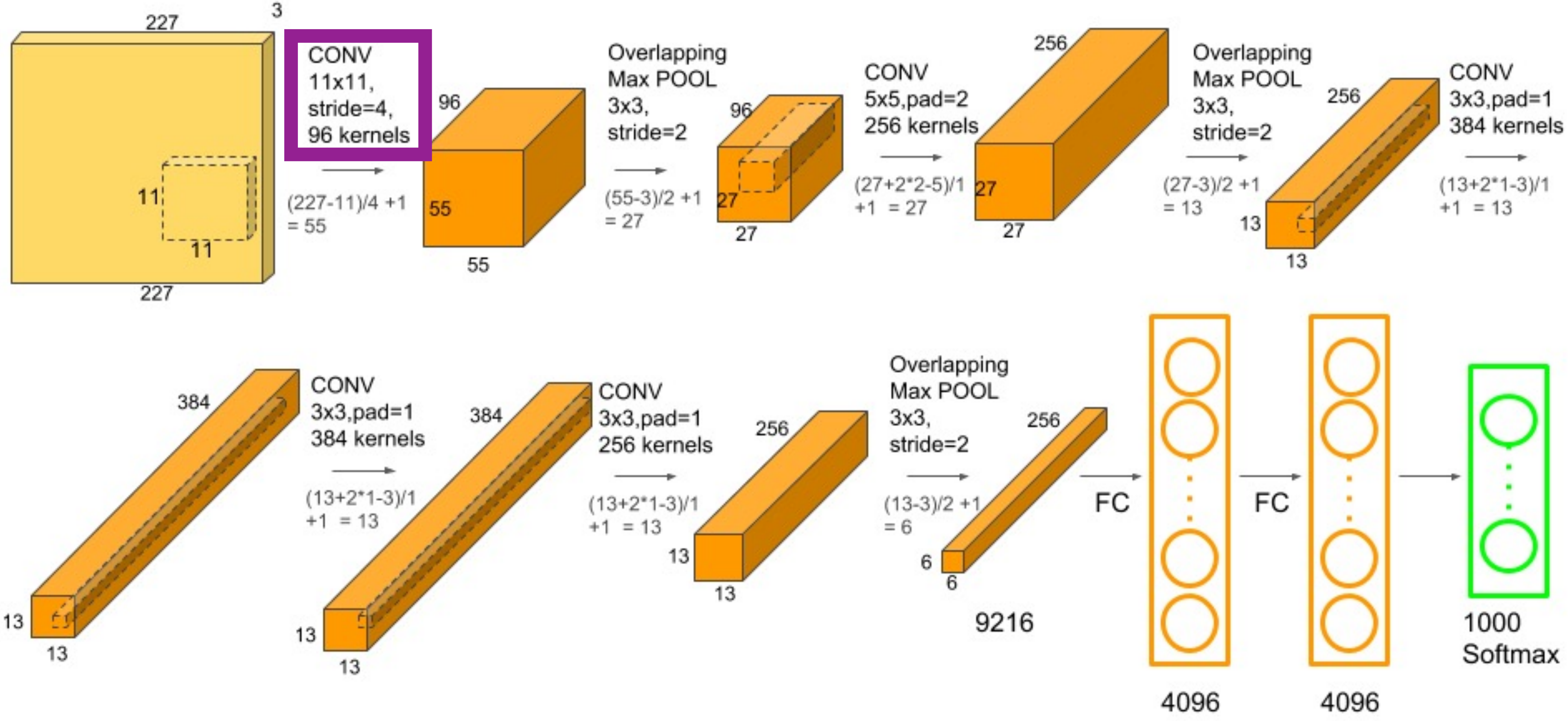
Algorithm Training: 90 Epochs on ImageNet



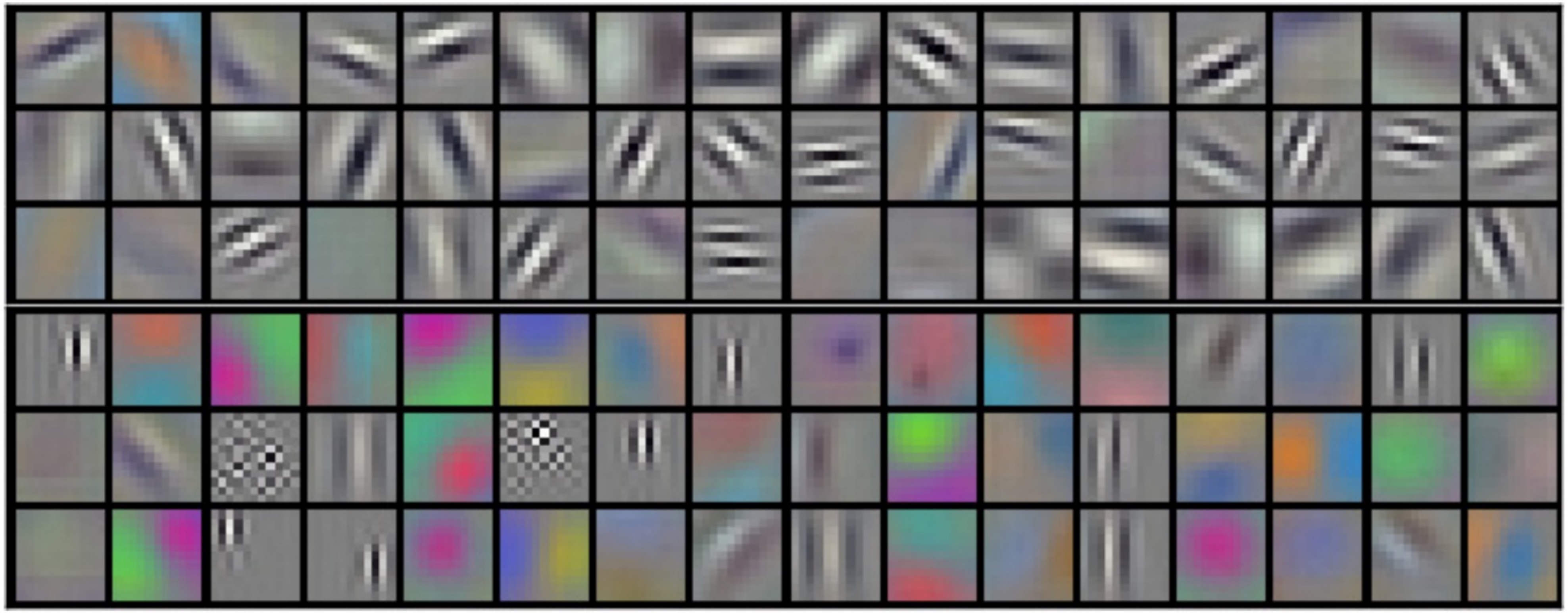
- Repeat until stopping criterion met:
  1. **Forward pass**: propagate training data through model to make prediction
  2. Quantify the dissatisfaction with a model's results on the training data
  3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each model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

AlexNet: Inspecting What It Learned



AlexNet: Inspecting What It Learned (96 Filters)



Model learned filters that select based on frequency, orientation, and color!

Object Recognition: Today’s Topics

- ImageNet Challenge Top Performers
- Baseline Model: AlexNet
- VGG
- ResNet
- Discussion

Why VGGNet?

VGG stands for the **Visual Geometry Group (VGG)** at University of Oxford where the authors were based ☺

Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." International Conference on Learning Representations (ICLR), 2015.

Key Novelty: **Deeper** Does Better

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

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

Key Novelty: **Deeper** Does Better

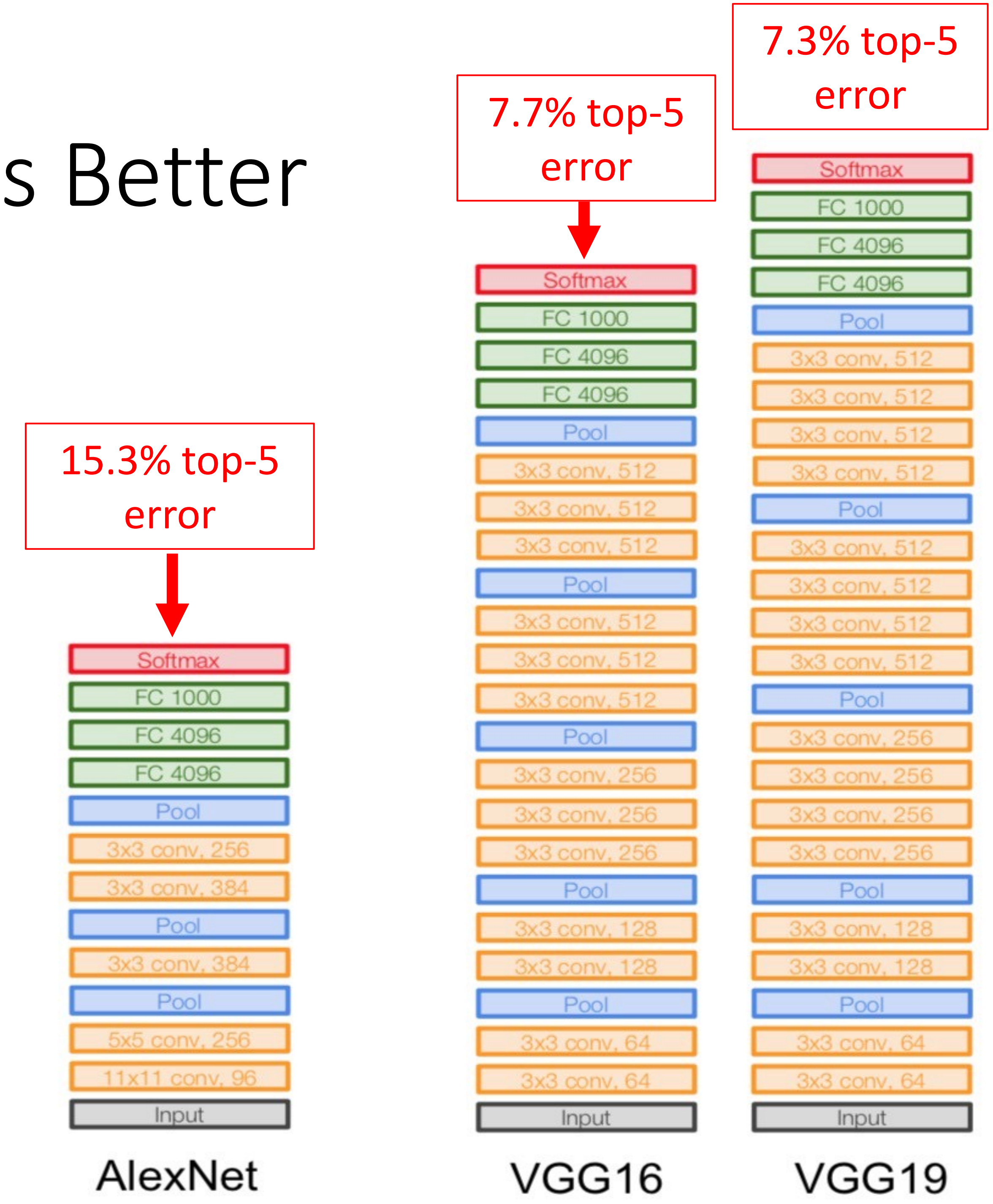


Figure Source (edited to fix mistakes): https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

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

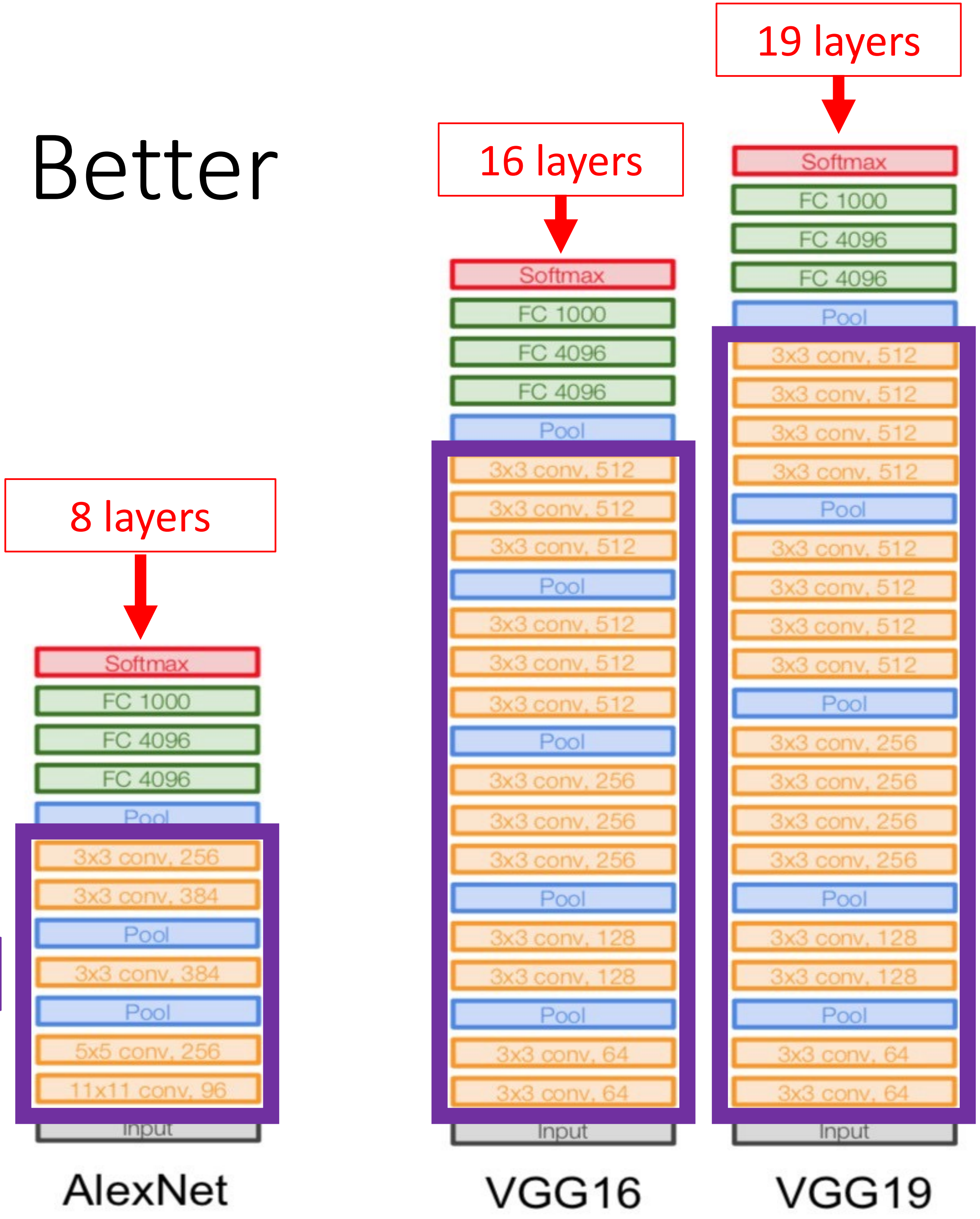


Figure Source (edited to fix mistakes): https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

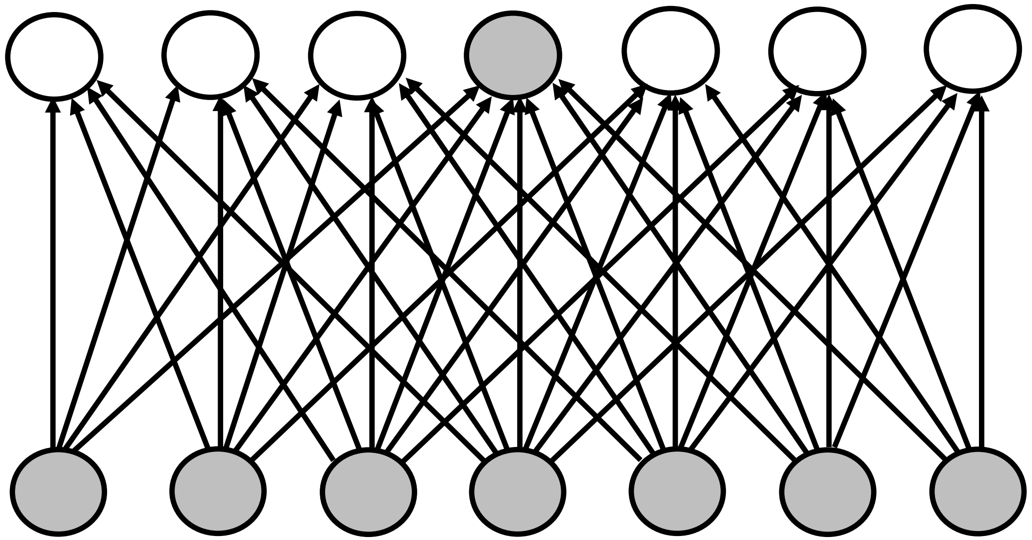
Key Idea: Smaller Convolutional Filters

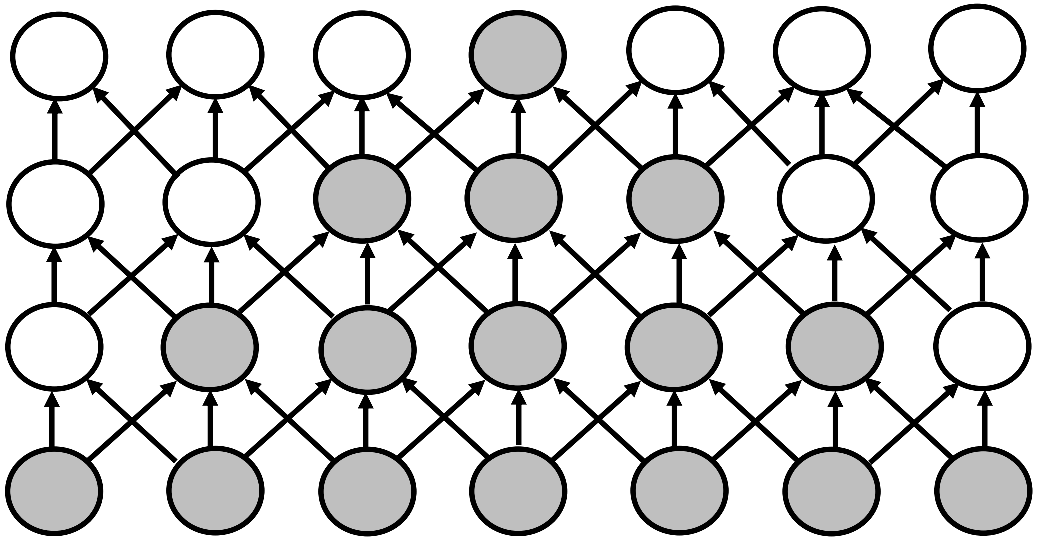
- Replace larger filter with stack of smaller filters

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

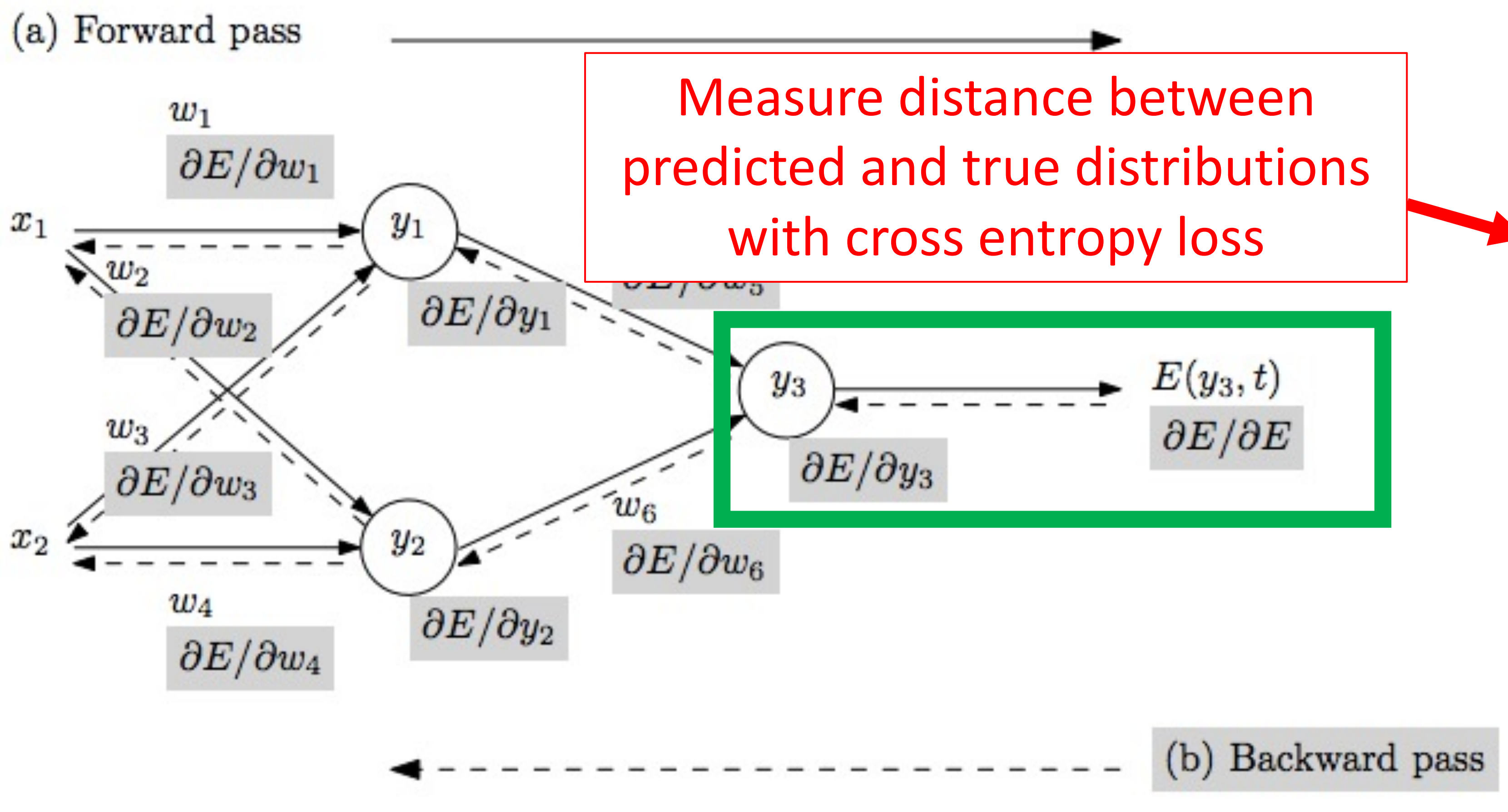




- Benefits:
    - More discriminative classifier since more non-linear rectifications: 3 vs 1
    - Reduces # of parameters: multiple of 27 (3×3^2) parameters vs 49 (7×7) parameters

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

Algorithm Training (follows AlexNet)



Measure distance between predicted and true distributions with cross entropy loss

- 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

Algorithm Training (follows AlexNet)

- Strategies to mitigate overfitting
  1. Data augmentation
     1. Random patches and their mirror images (2048x more data)
     2. Adjust RGB channels (using PCA to add multiples of principal components)
  2. Dropout (50% of nodes for first two fully connected layers); mimics ensembles by learning to solve same problem with different subnetworks

VGG Limitation: Models Are Large!

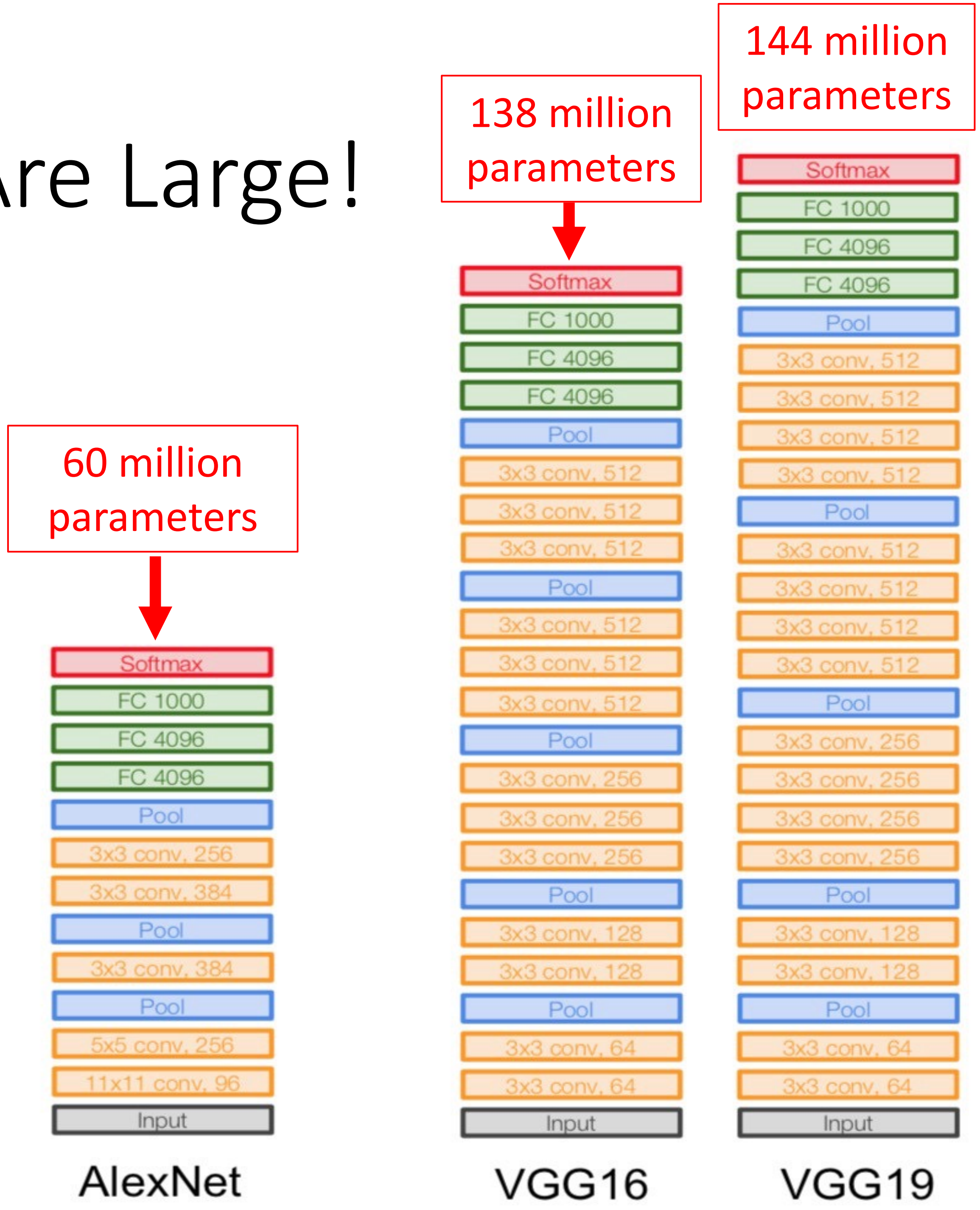
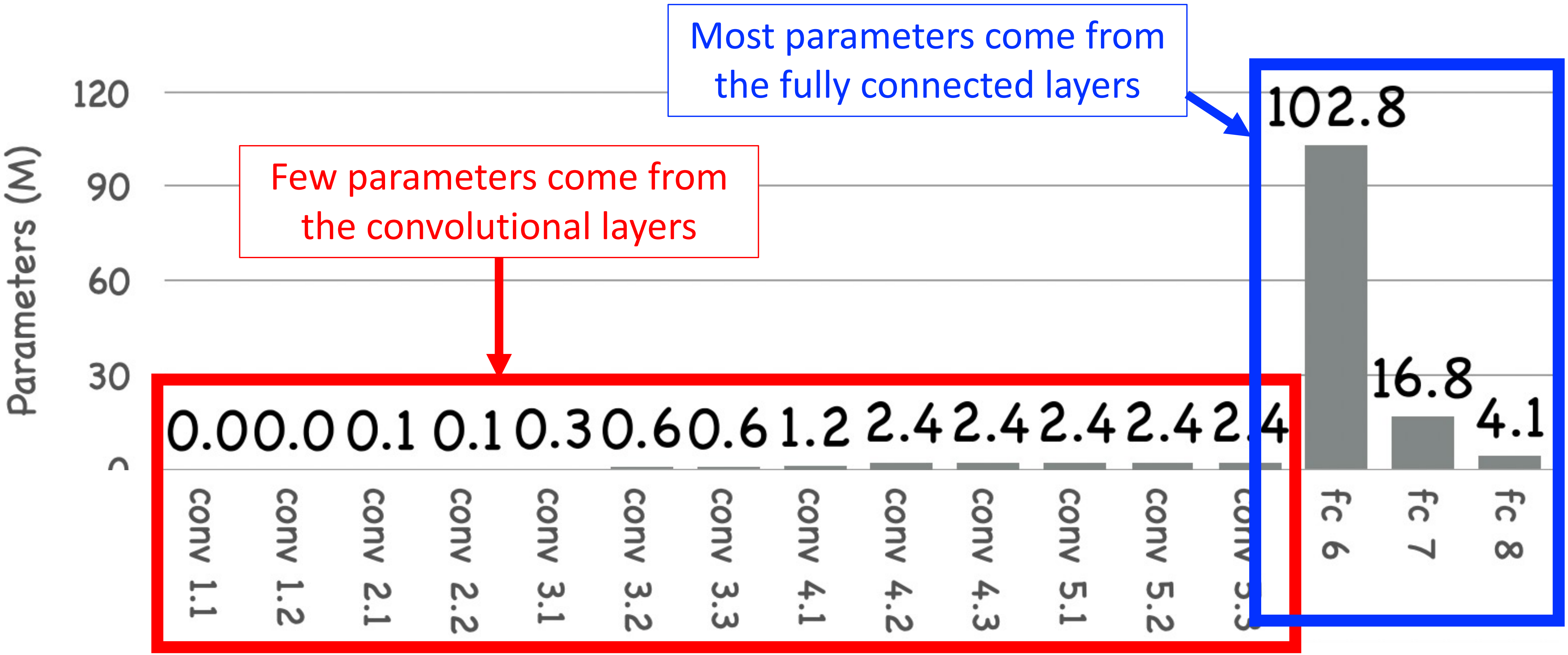


Figure Source (edited to fix mistakes): https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

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



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

Object Recognition: Today’s Topics

- ImageNet Challenge Top Performers
- Baseline Model: AlexNet
- VGG
- ResNet
- Discussion

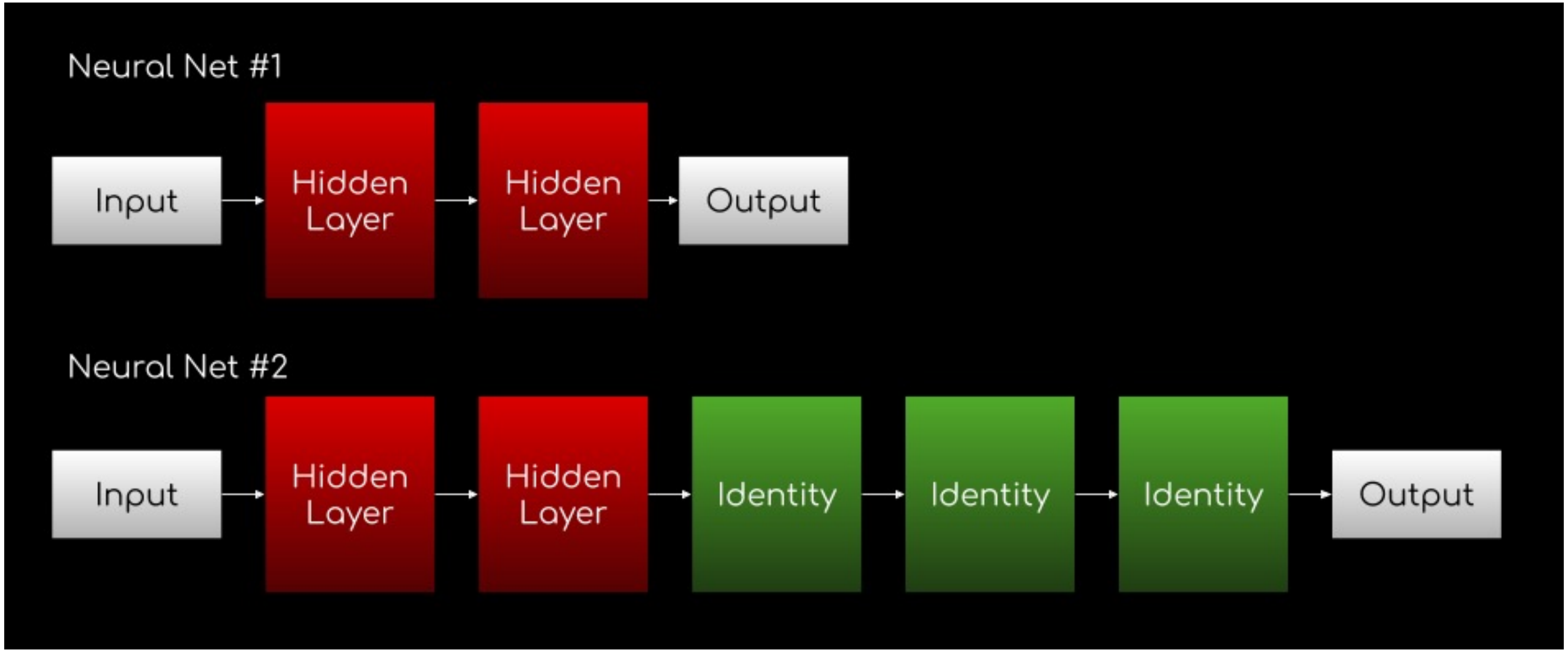
Why ResNet?

“Res” stands for residuals, which is the key novel idea in the proposed algorithm.

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?



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

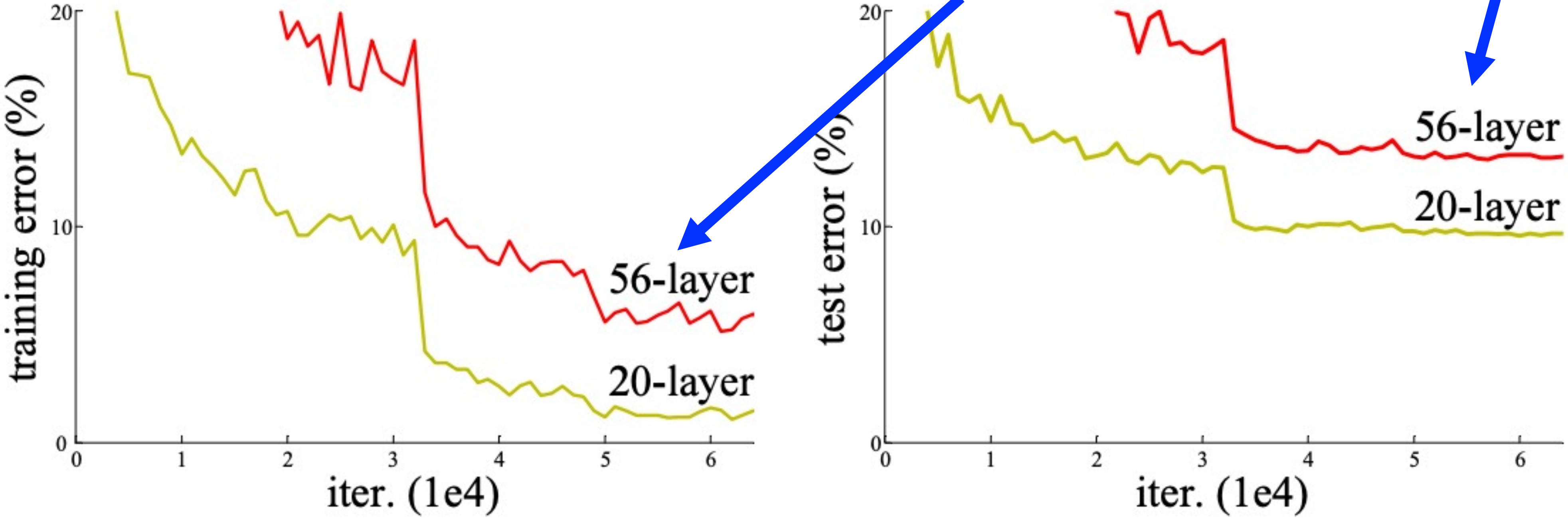
Motivating Observation

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!

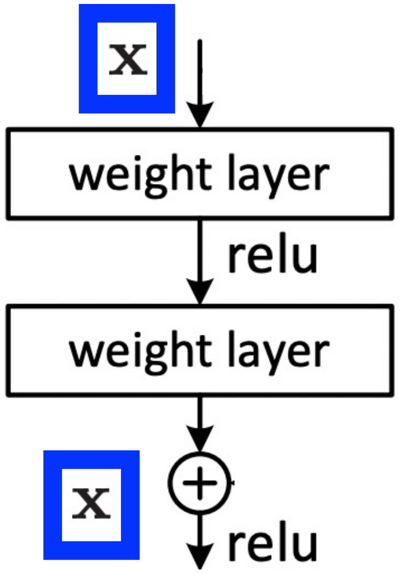
Is the problem overfitting? NO

Problem: It is difficult to learn 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.

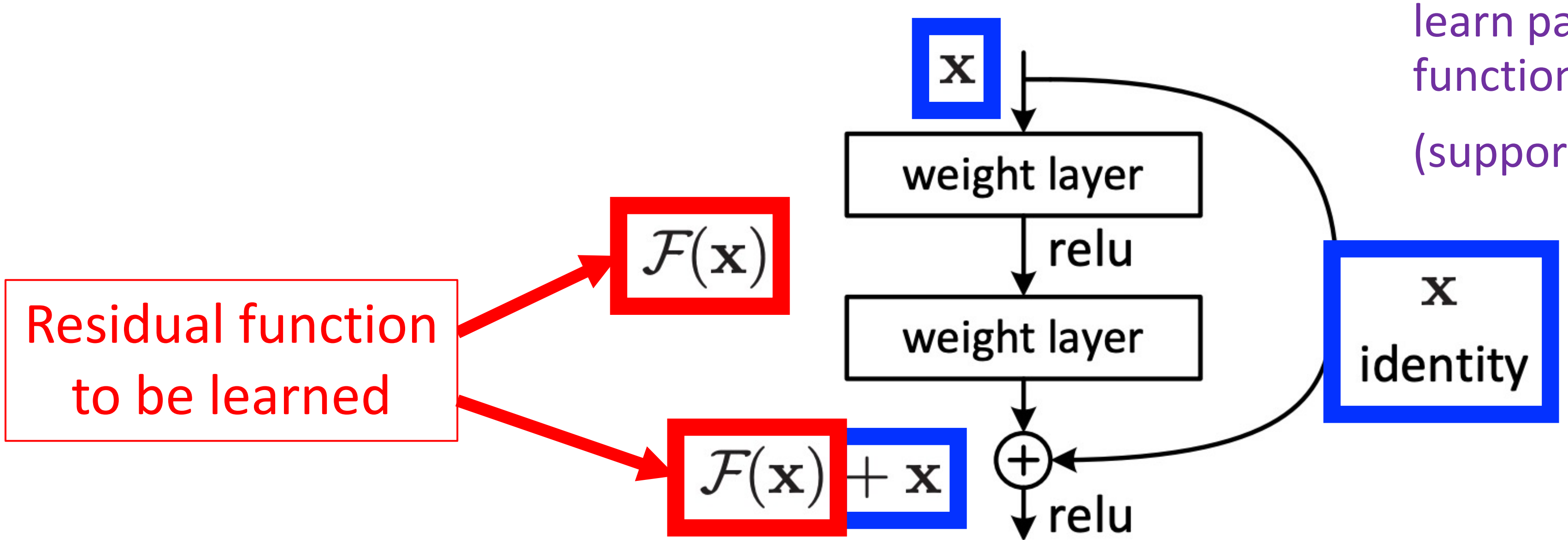
Problem: Difficult to Perform Identity Mapping

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



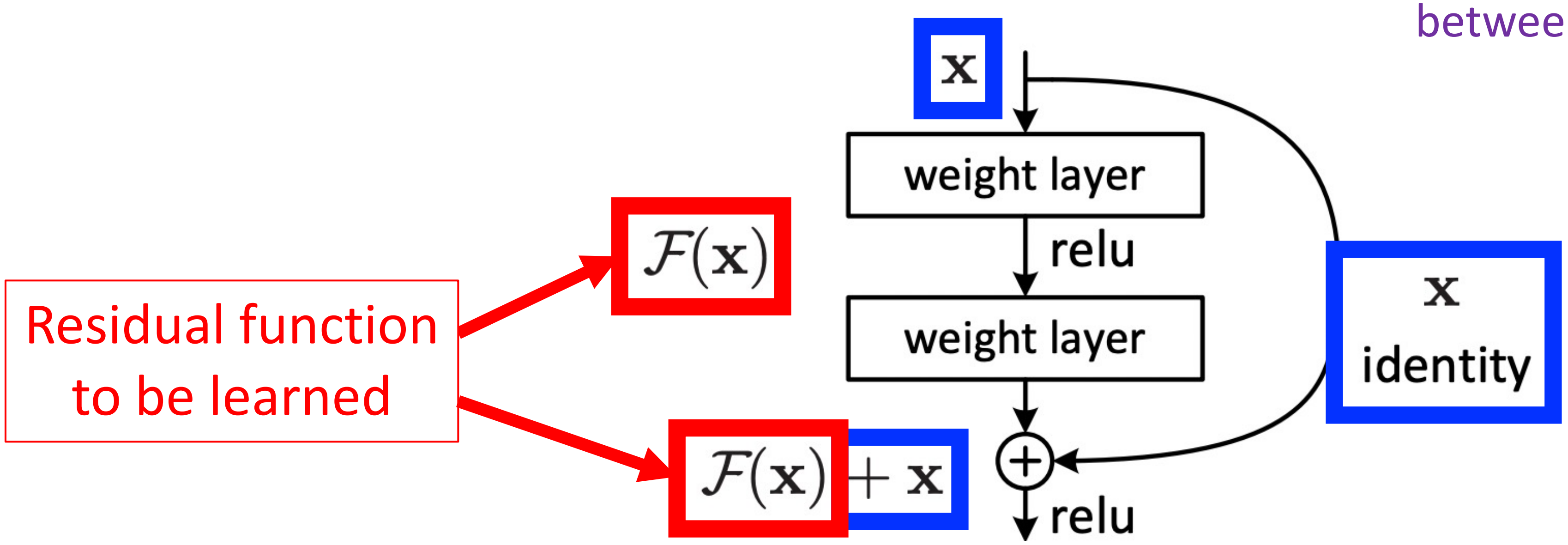
Assumption: Easier for a network to learn parameters close to 0 than a function that performs identity mapping

(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 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

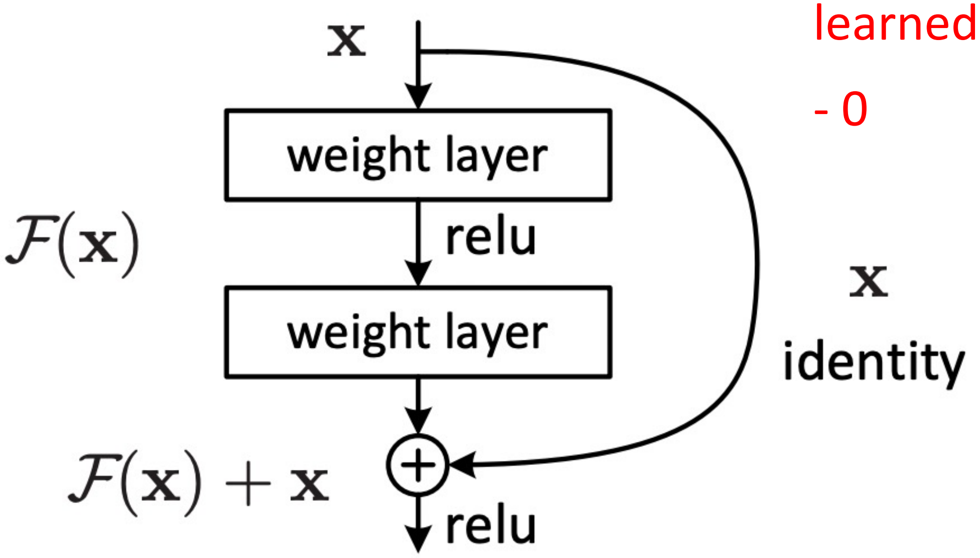
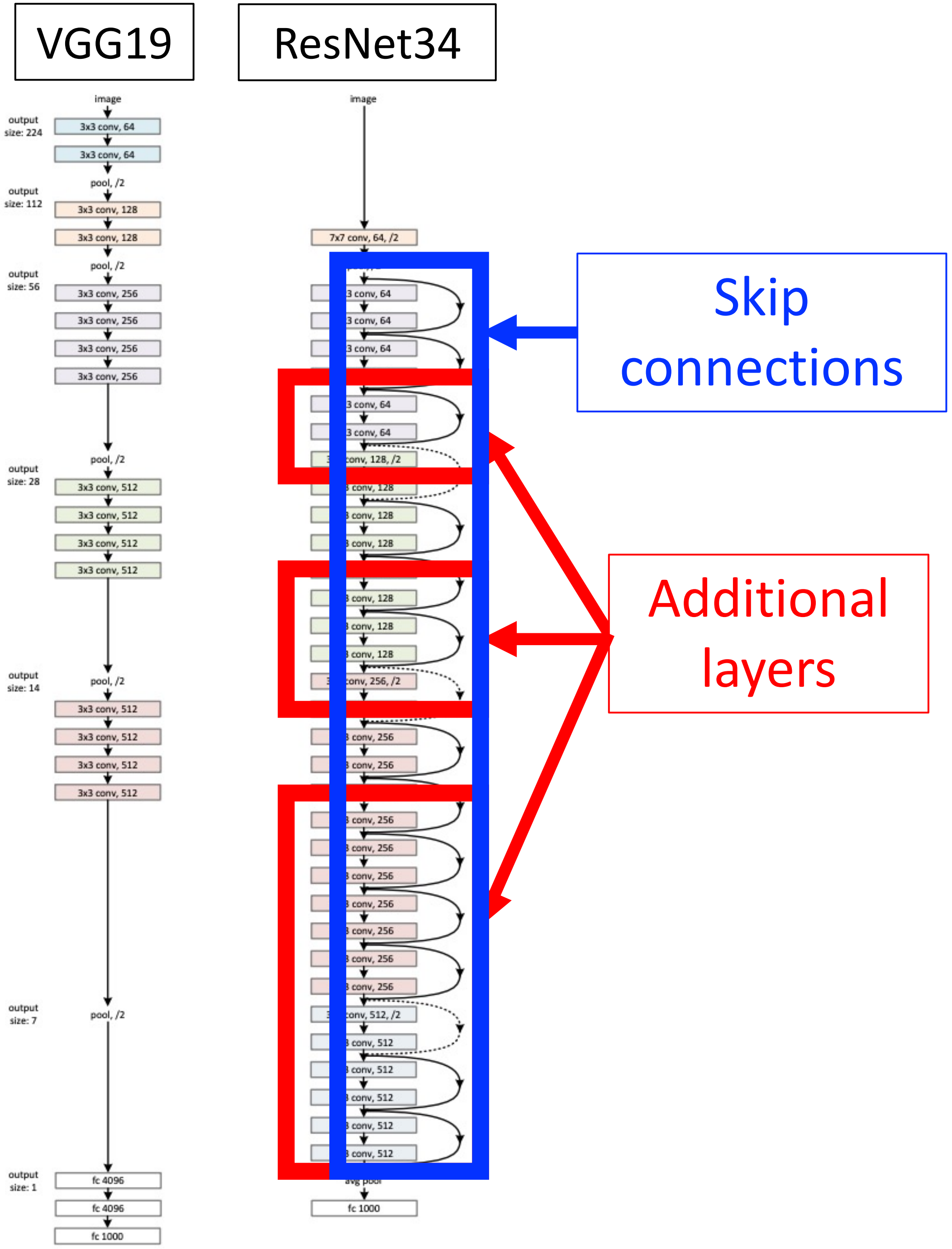


Figure 2. Residual learning: a building block.

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

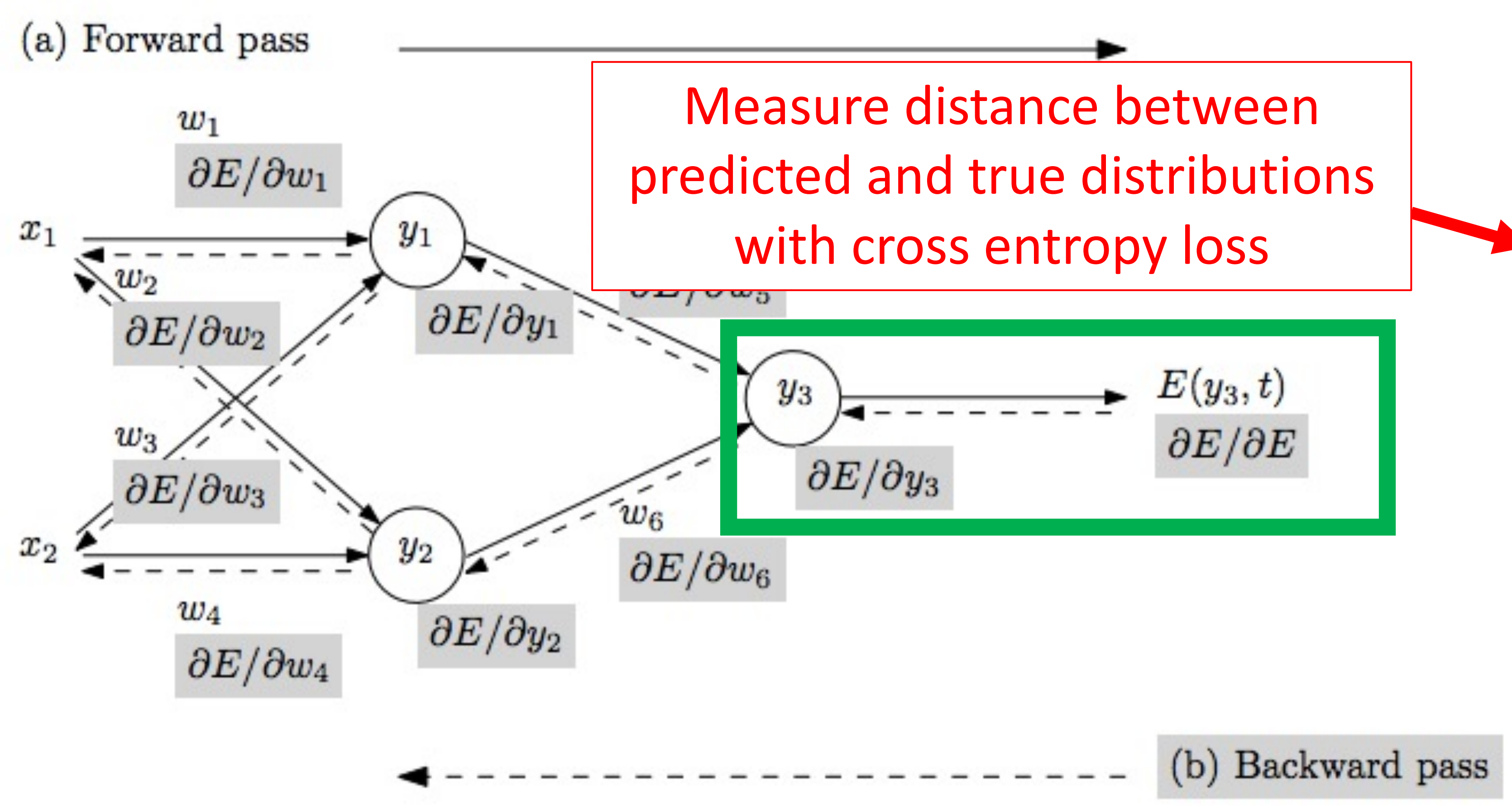
Key Contribution

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.

Algorithm Training (follows AlexNet)



Measure distance between predicted and true distributions with cross entropy loss

- 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

Algorithm Training (follows AlexNet)

- Strategy to mitigate overfitting
    1. Data augmentation
        1. Random patches and their mirror images (2048x more data)
        2. Adjust RGB channels (using PCA to add multiples of principal components)

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

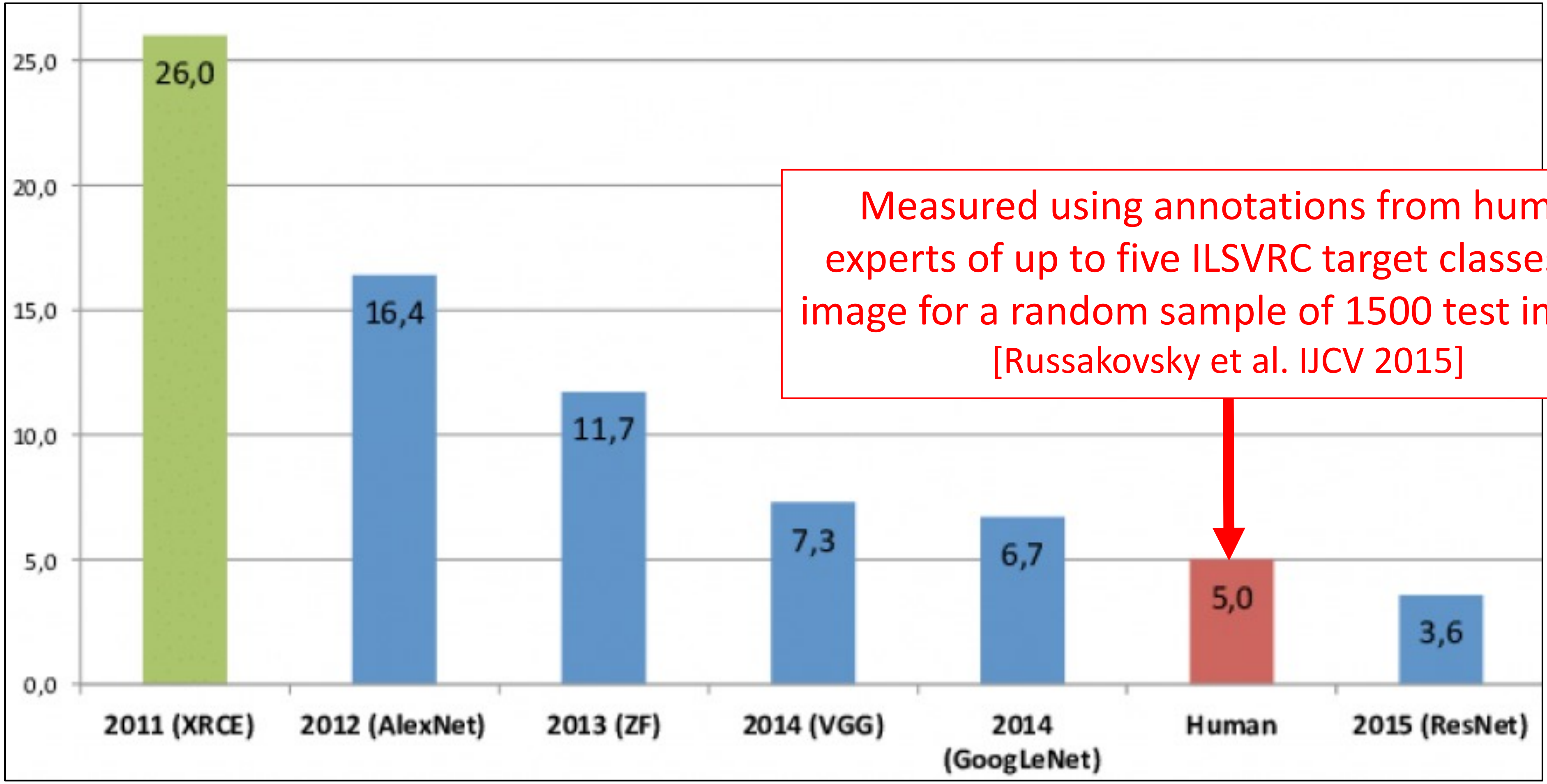
ResNet models outperform prior state-of-art models!

Object Recognition: Today's Topics

- ImageNet Challenge Top Performers
- Baseline Model: AlexNet
- VGG
- ResNet
- Discussion

State-of-Art Model Exceeds Human Performance!

Progress of models on ImageNet (Top 5 Error)



Measured using annotations from human experts of up to five ILSVRC target classes per image for a random sample of 1500 test images. [Russakovsky et al. IJCV 2015]

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

State-of-Art? Design Models That Go "Deeper"

Progress of models on ImageNet (Top 5 Error)

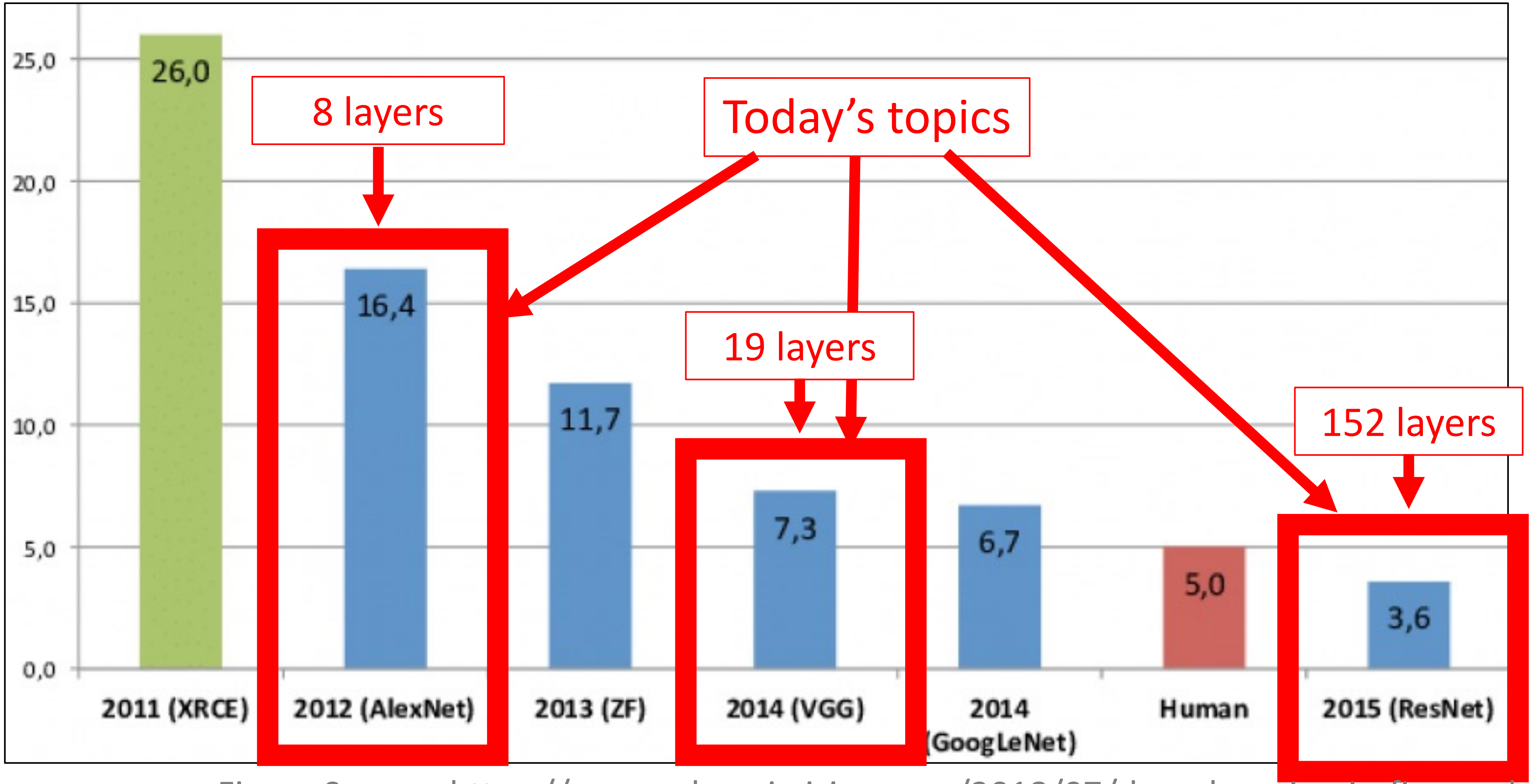
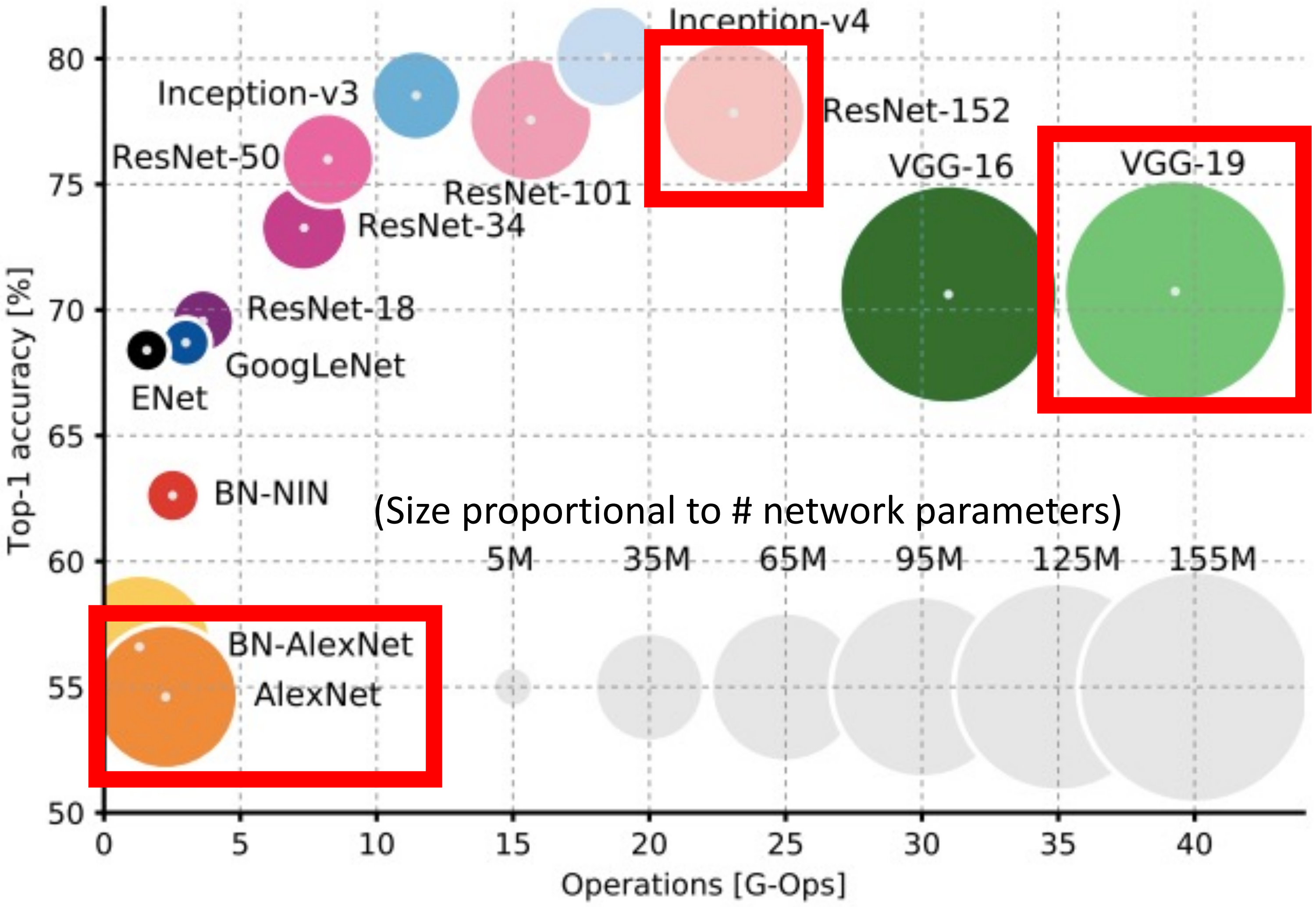


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

CNN Architectures: Great Start…

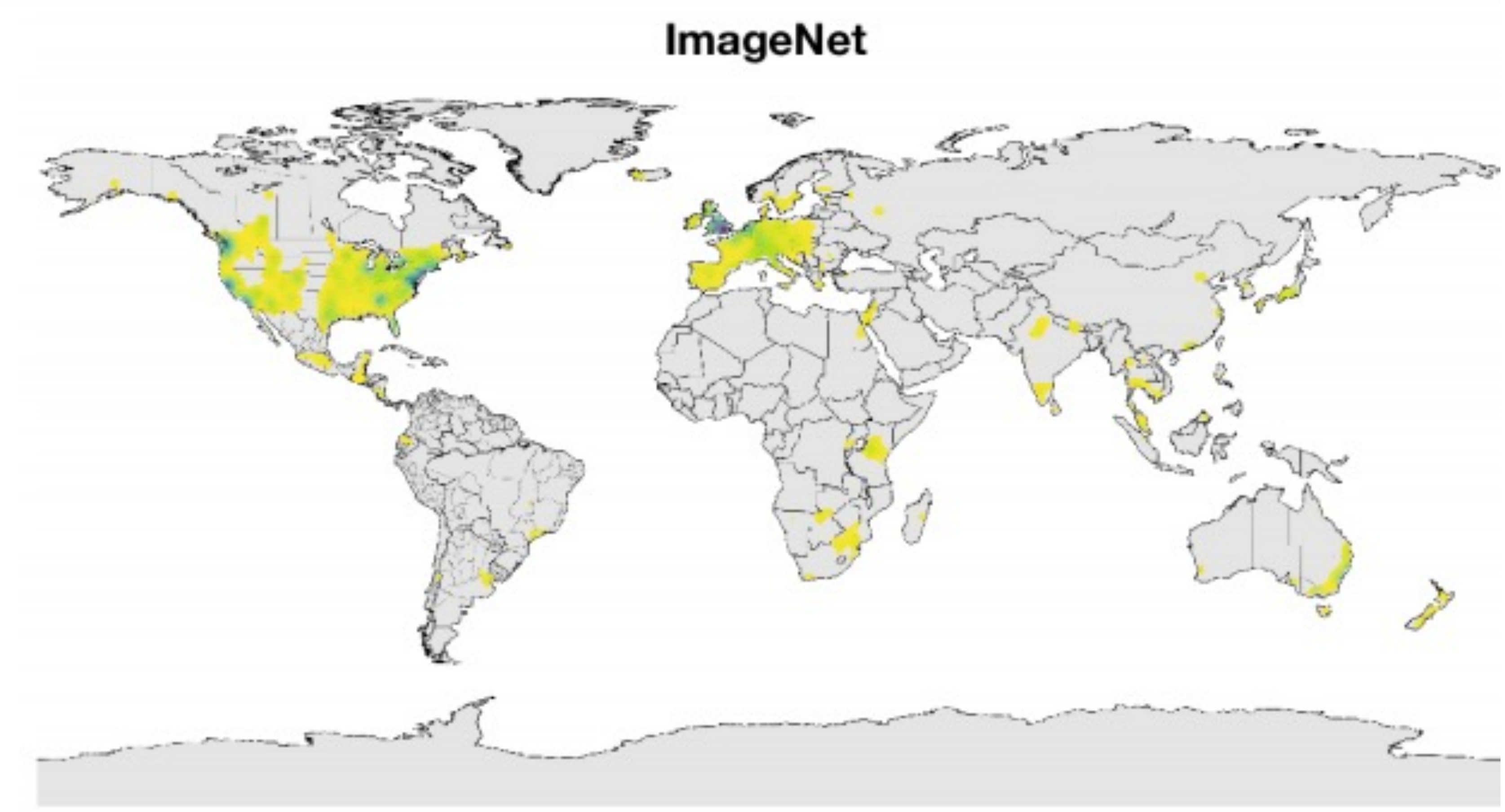


(required for a single forward pass)

https://arxiv.org/pdf/1605.07678.pdf

ImageNet: Great Start…

Geographical distribution of images in the ImageNet using Flickr metadata:



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

Group Discussion

- Vote for today’s topics in the Google form

Object Recognition: Today's Topics

- ImageNet Challenge Top Performers
- Baseline Model: AlexNet
- VGG
- ResNet
- ResNext
- Discussion


The End