Computer Vision: Image Classification

Deep Learning
CS 435/635

Course Instructor: Chandresh
AI Lab Coordinator @IIT Indore

Course Content

Module I: History of Deep Learning, Sigmoid Neurons, Perceptrons, and learning algorithms. Multilayer Perceptrons (MLPs), Representation Power of MLPs,.

Module II: Feedforward Neural Networks. Backpropagation. first and second-order training methods. NN Training tricks, Regularization

Module III: Introduction to Autoencoders and their characteristics, relation to PCA, Regularization in autoencoders, and Types of autoencoders, Variational Autoencoder

Module IV: Architecture of Convolutional Neural Networks (CNN), types of CNNs. **Image classification**.

Module V: Architecture of Recurrent Neural Networks (RNN), Backpropagation through time.
Encoder-Decoder Models, Attention Mechanism. Advanced Topics: Transformers and BERT.
Nodule VI: Gen AI- Deep generative models: VAE, GAN,

Acknowledgement

- Russ Salakhutdinov and Hugo Larochelle's class on Neural Networks
- Neural Networks and Deep Learning course by Danna Gurari University of Colorado Boulder

Today's Topics

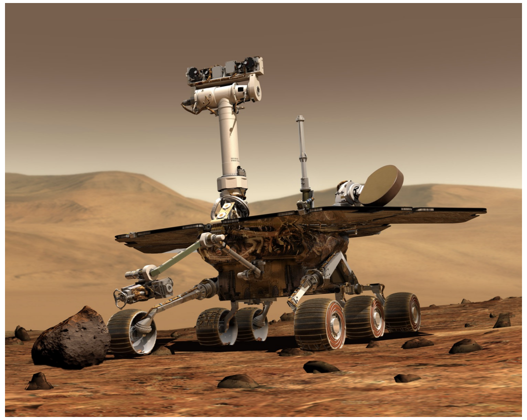
- Computer vision
- Era of dataset challenges
- MNIST challenge winner: LeNet
- ImageNet challenge winners: deeper learning (AlexNet, VGG, ResNet)
- Programming tutorial

Today’s Topics

- Computer vision
- Era of dataset challenges
- MNIST challenge winner: LeNet
- ImageNet challenge winners: deeper learning (AlexNet, VGG, ResNet)
- Programming tutorial

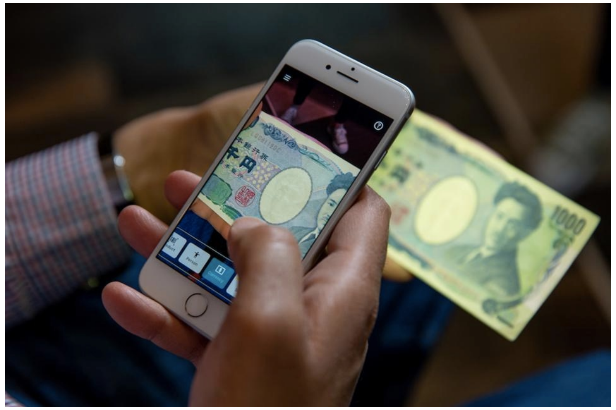
Computer Vision: Computers that "See"

Self-driving cars

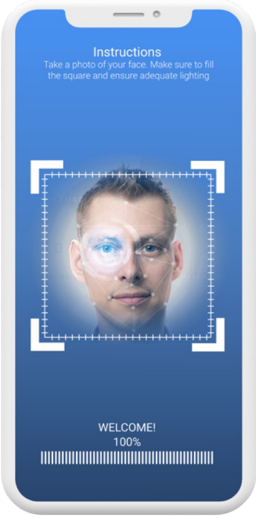


Exploration on Mars

Guided surgery



Visual assistance for people who are blind

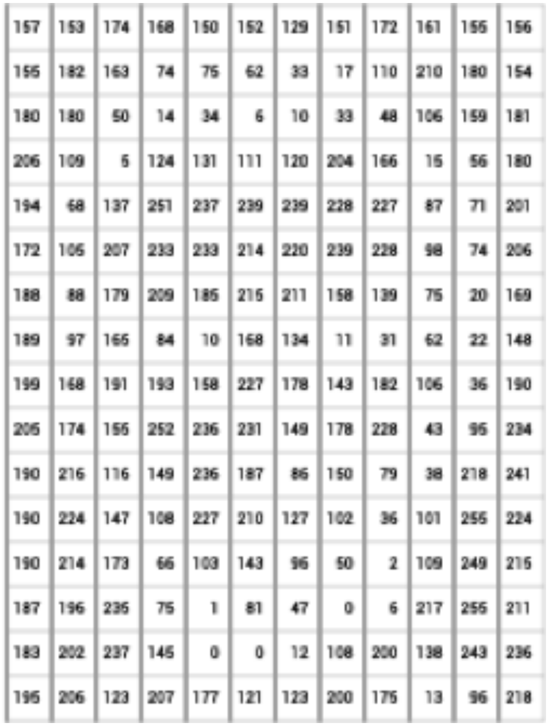


Security

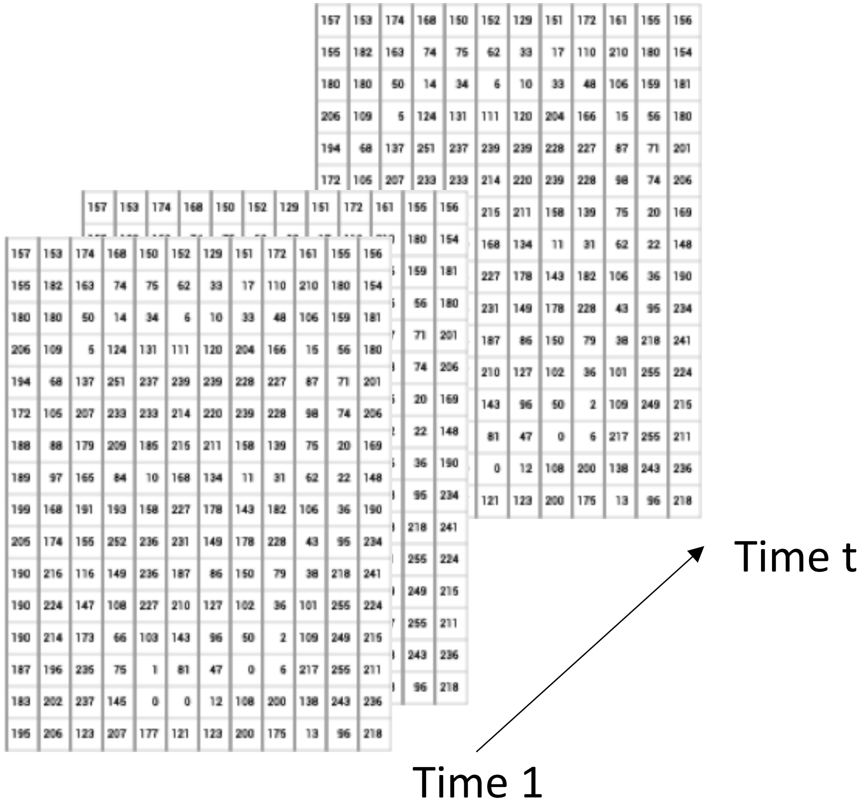
Why Discuss Computer Vision With CNNs?

- CNNs have a strong track record for vision problems
- Visual data's representation (i.e., spatial data) is naturally suited for CNNs

Image:



Video:



Historical Context: Origins of Computer Vision

1847 — Gradient descent

1945 — First programmable machine

1950 — Turing test

1956 — AI

1957 — Perceptron

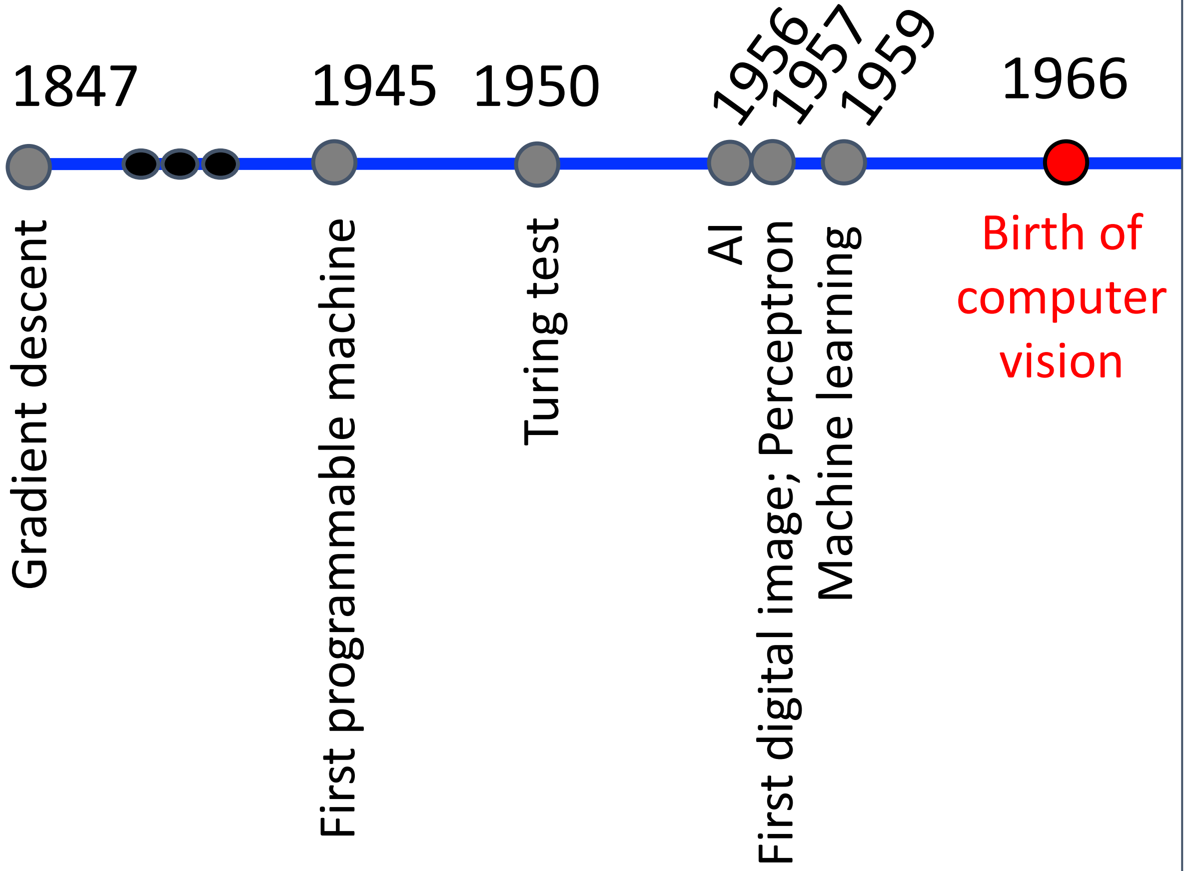
1959 — Machine learning

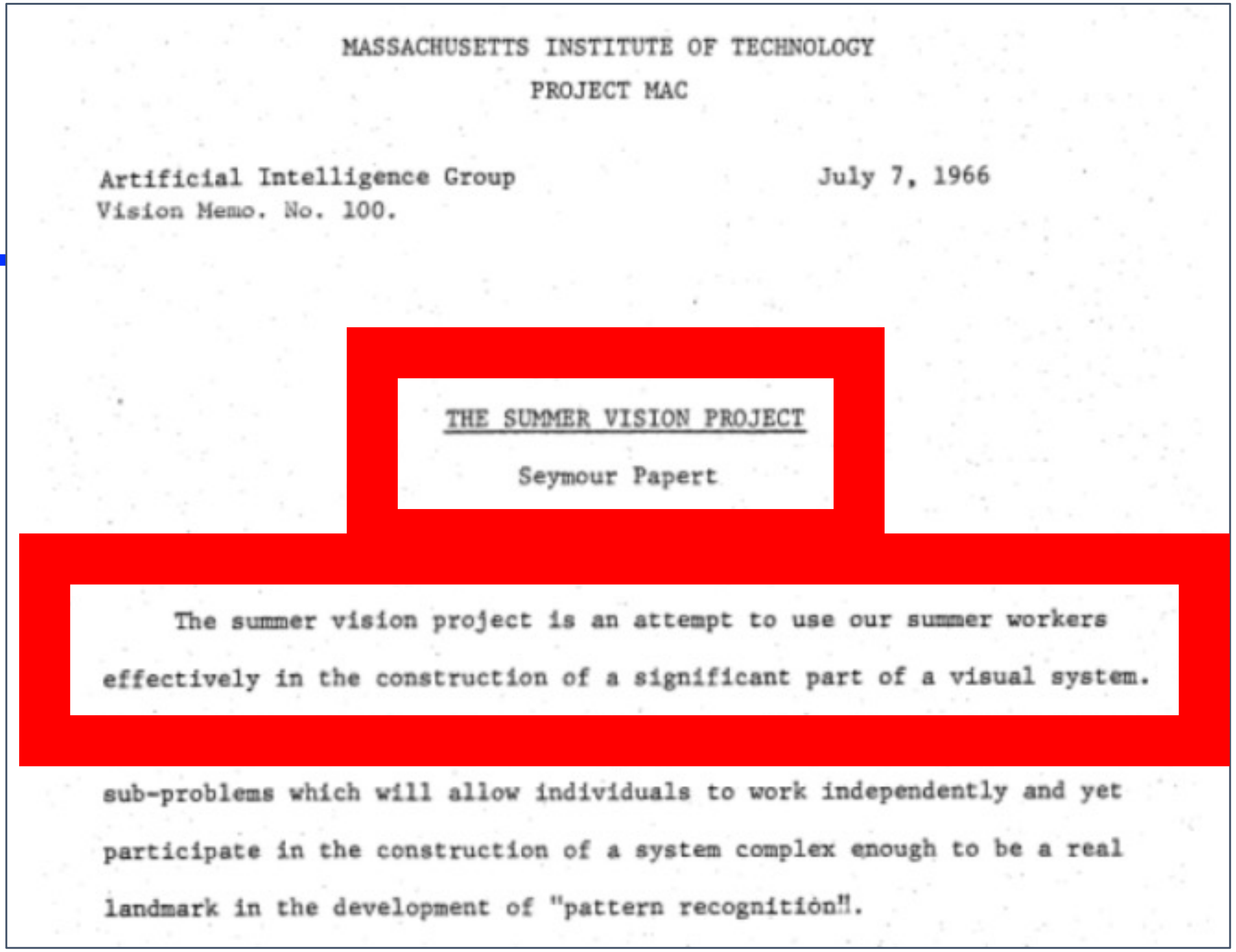
First digital image:

176 x 176 pixels

2012 — Wave 3: rise of "deep learning"

Historical Context: Origins of Computer Vision





MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

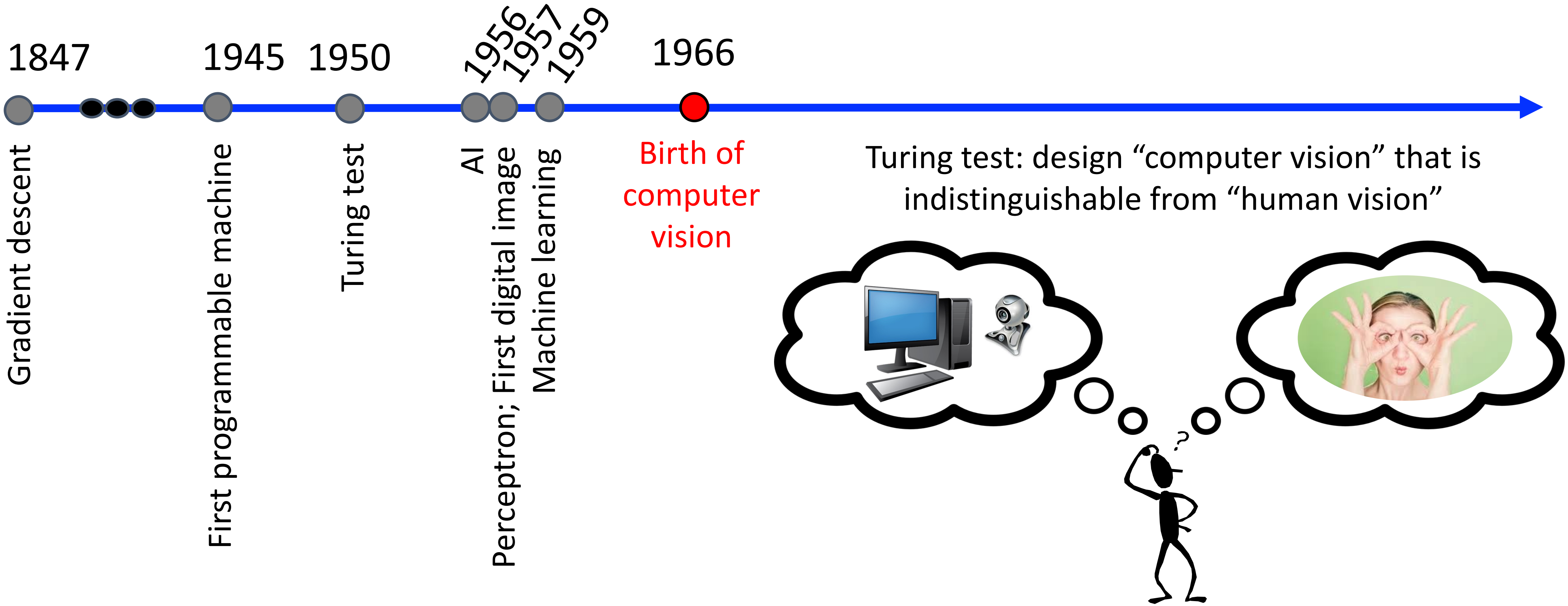
THE SUMMER VISION PROJECT

Seymour Papert

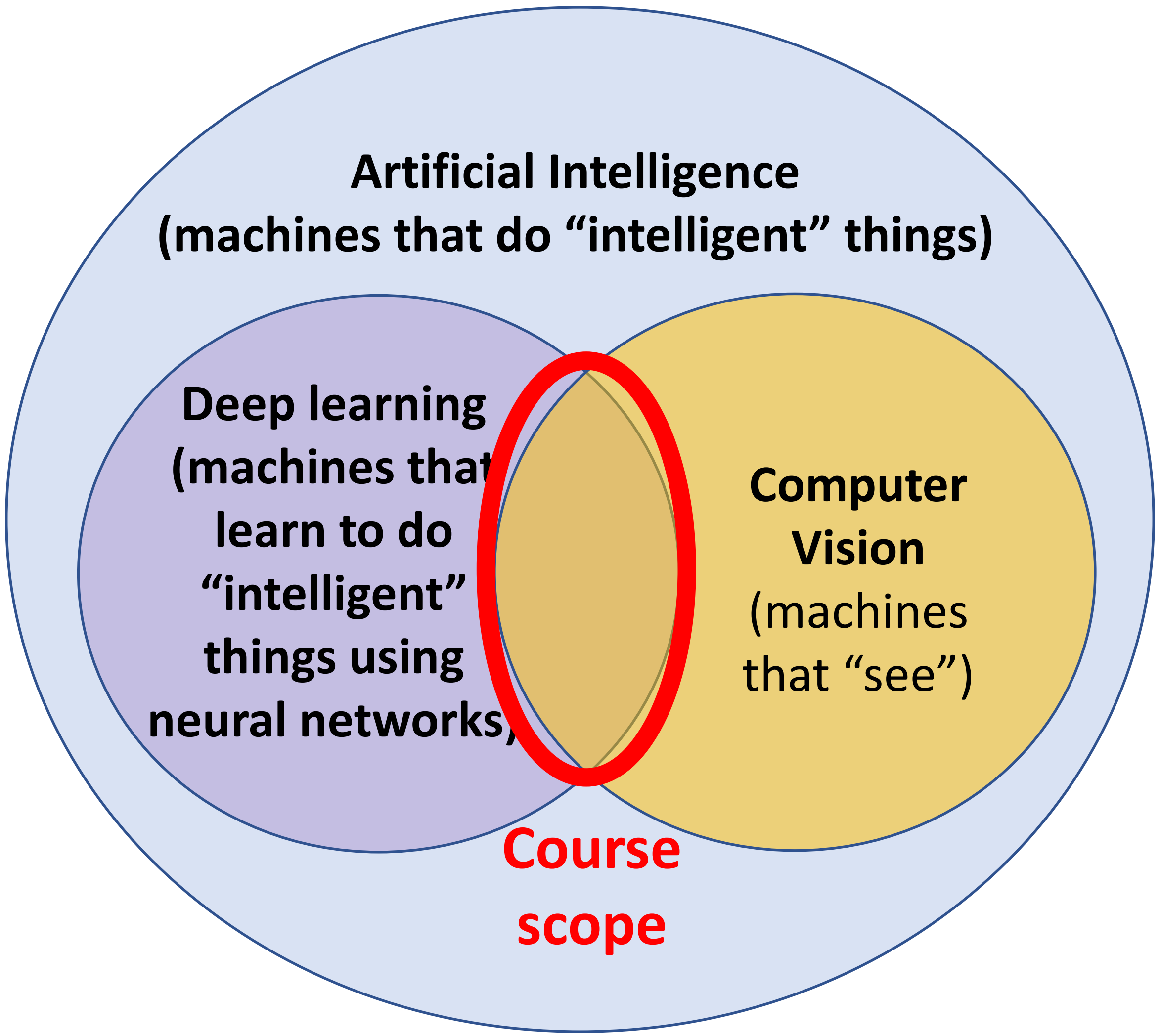
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system.

sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Historical Context: Origins of Computer Vision



Computer Vision in Context



Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

- Object recognition
- Object detection
- Segmentation
- Image captioning
- Visual question answering
- Object tracking
- Subjective problems
- And more…

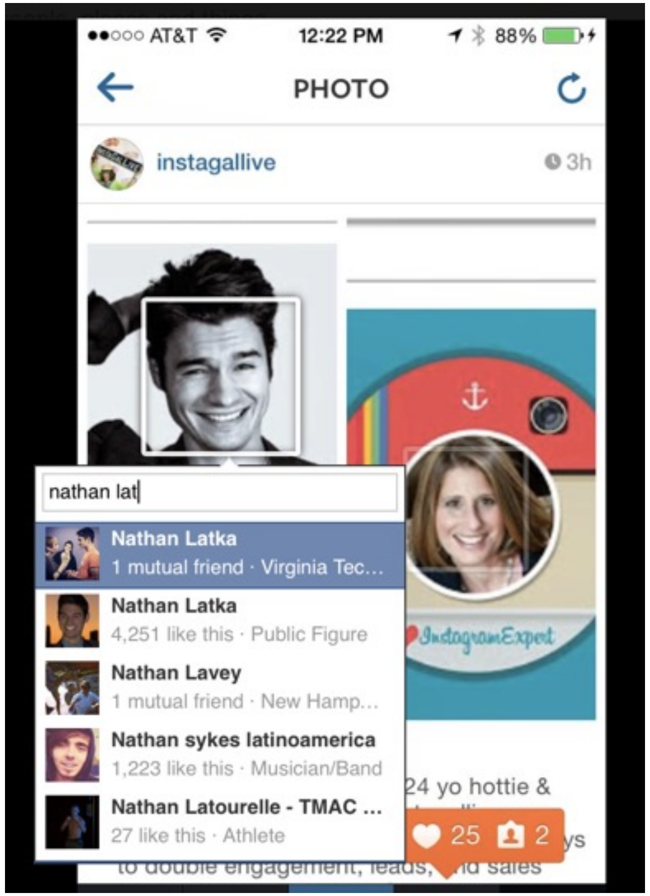
Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

- Object recognition
- Object detection
- Segmentation
- Image captioning
- Visual question answering
- Object tracking
- Subjective problems
- And more…

e.g., take a picture of an object and find where to buy it

Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

- Object recognition
- **Object detection**
- Segmentation
- Image captioning
- Visual question answering
- Object tracking
- Subjective problems
- And more…



e.g., detect faces to tag

Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

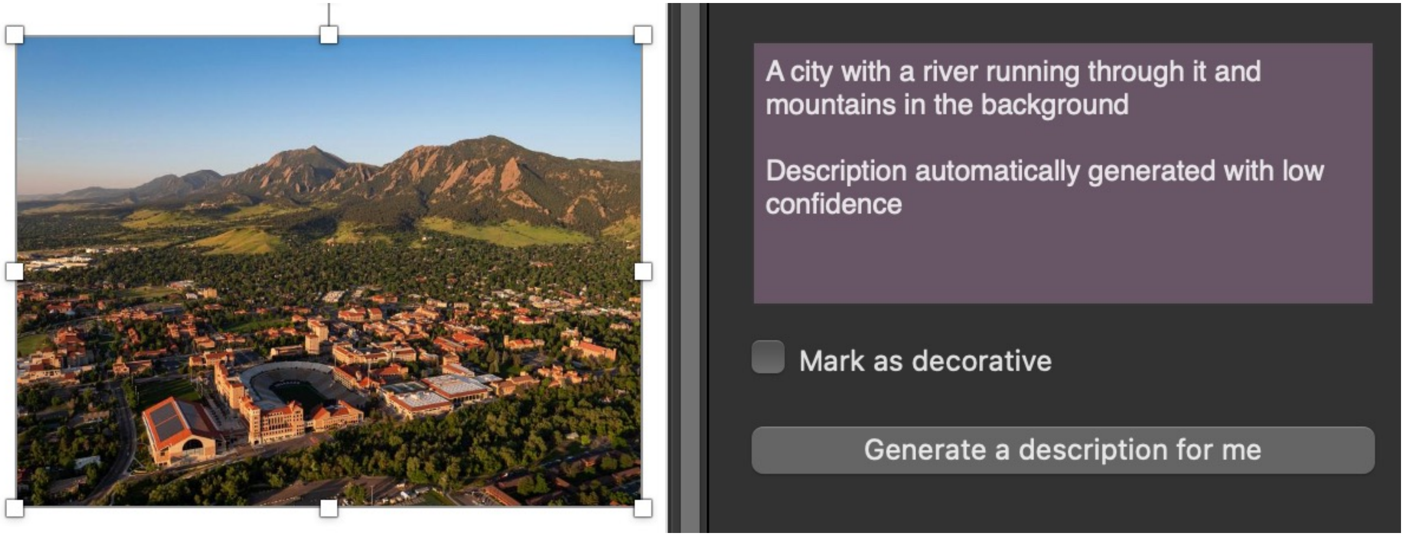
- Object recognition
- Object detection
- **Segmentation**
- Image captioning
- Visual question answering
- Object tracking
- Subjective problems
- And more…

e.g., rotoscoping

https://www.starnow.co.uk/ahmedmohammed1/photos/4650871/before-and-after-rotoscopinggreen-screening

Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

- Object recognition
- Object detection
- Segmentation
- **Image captioning**
- Visual question answering
- Object tracking
- Subjective problems
- And more…



e.g., Microsoft Power Point

Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

- Object recognition
- Object detection
- Segmentation
- Image captioning
- Visual question answering
- Object tracking
- Subjective problems
- And more…

e.g., BeSpecular

https://www.lionessesofafrica.com/blog/2015/2/15/the-startup-story-of-stephanie-cowper

Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

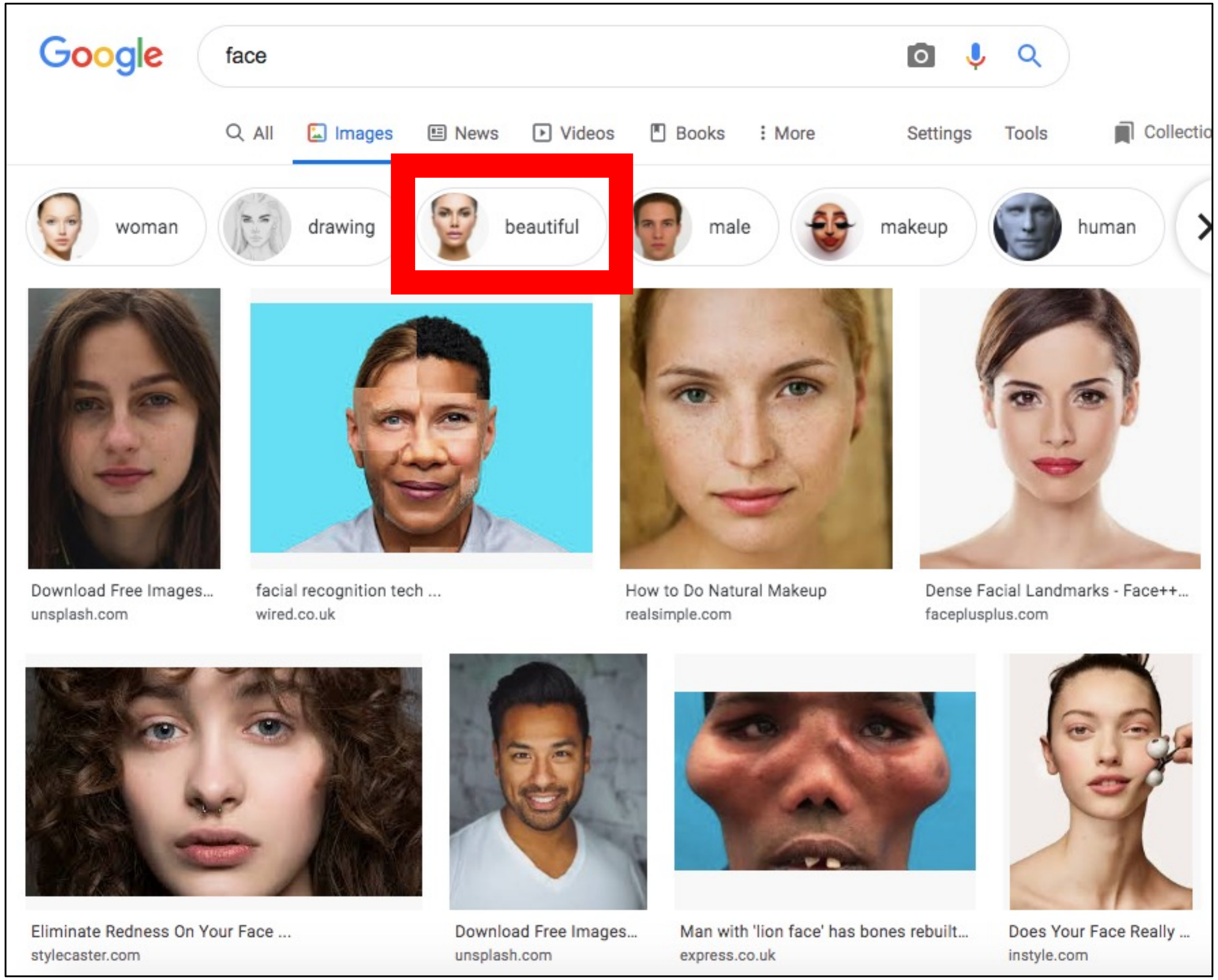
- Object recognition
- Object detection
- Segmentation
- Image captioning
- Visual question answering
- **Object tracking**
- Subjective problems
- And more…

e.g., track bowling ball path

e.g., calculate bat speed

Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

- Object recognition
- Object detection
- Segmentation
- Image captioning
- Visual question answering
- Object tracking
- Subjective problems
- And more…



Key Challenge: Replicate Human Vision for So Much Variation for **So Many Tasks**!

- Object recognition
- Object detection
- Segmentation
- Image captioning
- Visual question answering
- Object tracking
- Subjective problems
- And more…

Key Challenge: Replicate Human Vision for **So Much Variation** for So Many Tasks!


Illumination


Object pose


Clutter


Occlusions


Intra-class appearance


Viewpoint

Slide by Kristen Grauman

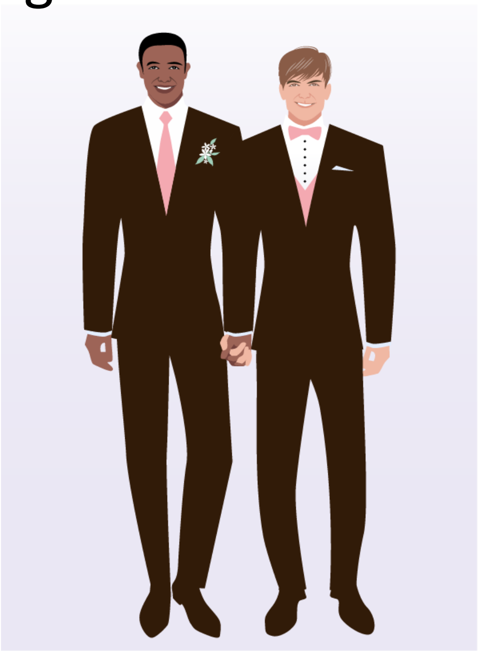
Today’s Topics

- Computer vision
- Era of dataset challenges
- MNIST challenge winner: LeNet
- ImageNet challenge winners: deeper learning (AlexNet, VGG, ResNet)
- Programming tutorial

Through 1990s, Common Approach to Developing Computer Vision Models:

Algorithm Dataset

Algorithm Dataset



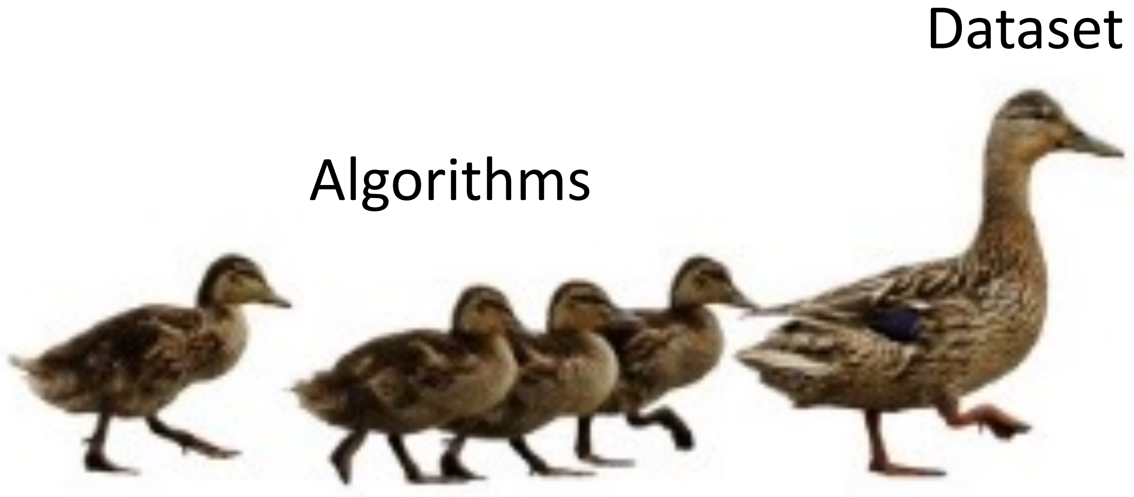
Algorithm Dataset

Algorithm

Dataset

Datasets tended to be relatively small (e.g., 10s or 100s of examples)

Since 1990s, Common Approach to Developing Computer Vision Models:

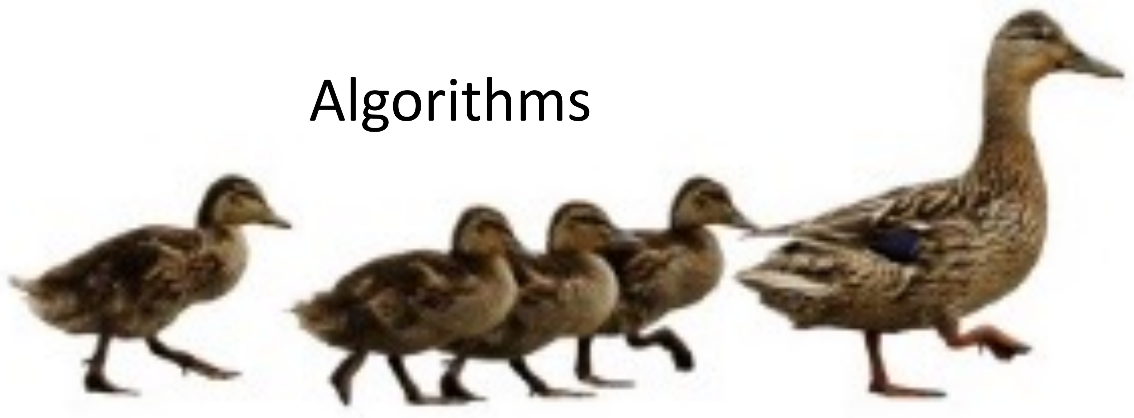


Datasets tend to be large (i.e., thousands to millions of examples)

Image Source: http://larryzitnick.org/Talks/CVPR15_Dataset.pptx

Since 1990s, Common Approach to Developing Computer Vision Models:

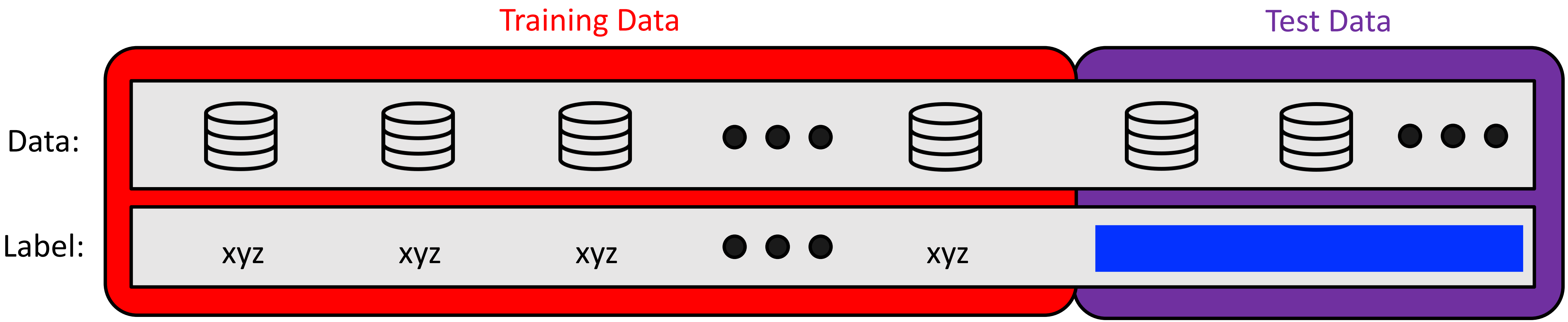
What do you think prompted this shift to large-scale datasets?



Datasets tend to be large (i.e., thousands to millions of examples)

Image Source: http://larryzitnick.org/Talks/CVPR15_Dataset.pptx

Progress Charted by Progress on Community Shared Dataset Challenges: How It Works



1. Dataset split into a "training set" and "test set" with the labels for the "test set" hidden
2. Teams design a model and submit its predictions on the test set to an evaluation server
3. A public leaderboard shows the ranking of performance for all teams' submitted models

Progress Charted by Progress on Community Shared Dataset Challenges: Why Challenges?

- Provide “fair” comparison between models
- Create a community around a shared goal

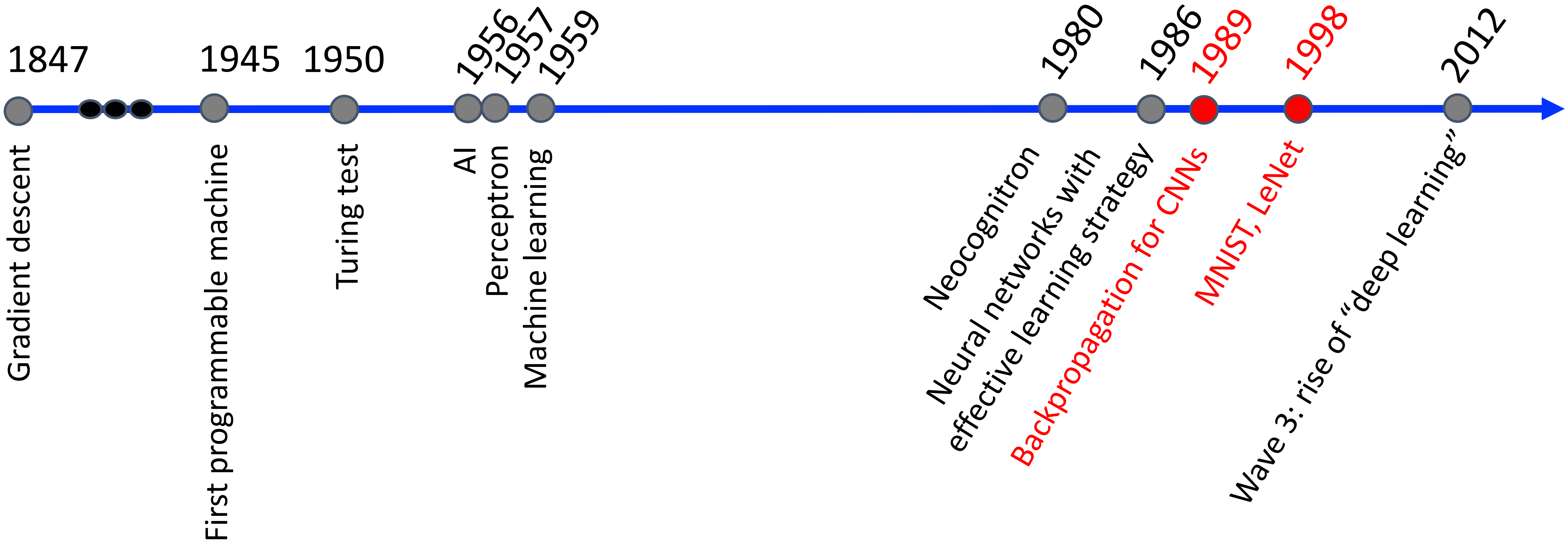
Many Public Dataset Challenges Available; e.g.,

- Google Dataset Search
- Amazon’s AWS datasets
- Kaggle datasets
- Wikipedia’s list
- UC Irvine Machine Learning Repository
- Quora.com
- Reddit
- Dataportals.org
- Opendatamonitor.eu
- Quandl.com

Today’s Topics

- Computer vision
- Era of dataset challenges
- MNIST challenge winner: LeNet
- ImageNet challenge winners: deeper learning (AlexNet, VGG, ResNet)
- Programming tutorial

Historical Context: Inspiration



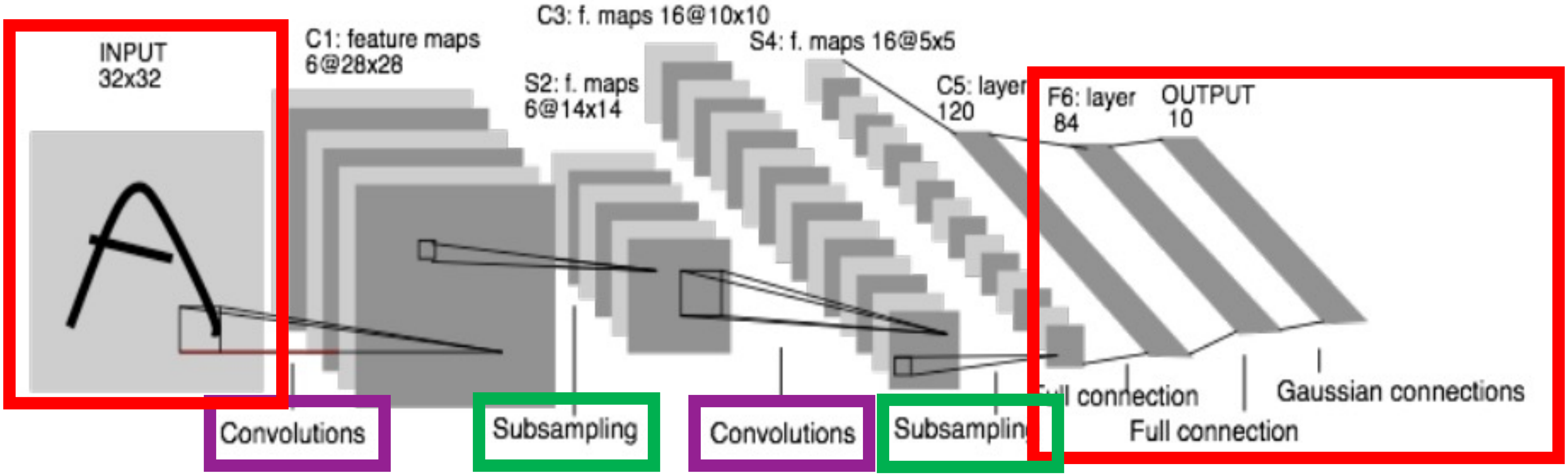
Key contribution: showing how to perform backpropagation for CNNs to enable learning thereby eliminating the need for hand-crafted filters

MNIST Dataset Challenge

- Goal: classify digit as 0, 1, …, or 9
- Evaluation metric: accuracy (% correct)
- Dataset: 60,000 training and 10,000 test examples, pre-processed to be centered and same dimension; writers were different in the two sets
- Source: images collected by NIST from a total of 500 Census Bureau employees and high school students

Dataset location: http://yann.lecun.com/exdb/mnist/
NIST dataset: https://www.nist.gov/srd/nist-special-database-19
Figure source: https://commons.wikimedia.org/w/index.php?curid=64810040

LeNet: Architecture (like Neocognitron, has convolutional layers and pooling layers)

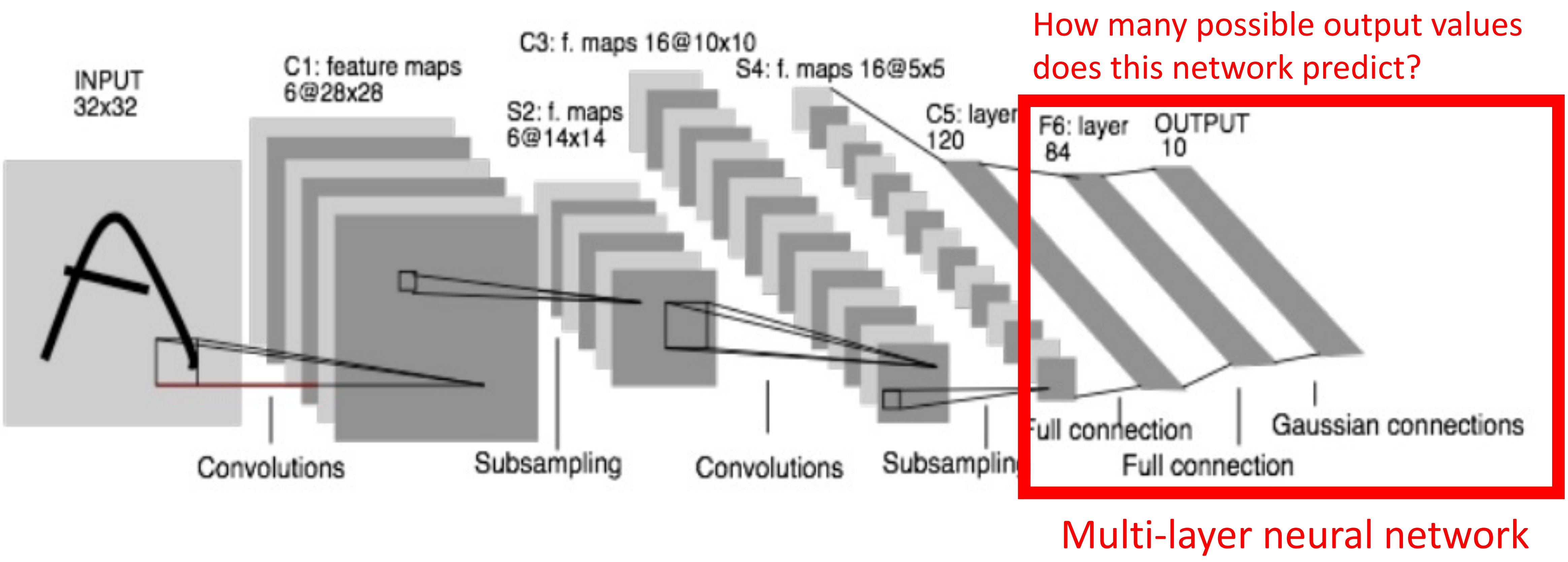


tanh is used as the activation function

Multi-layer neural network

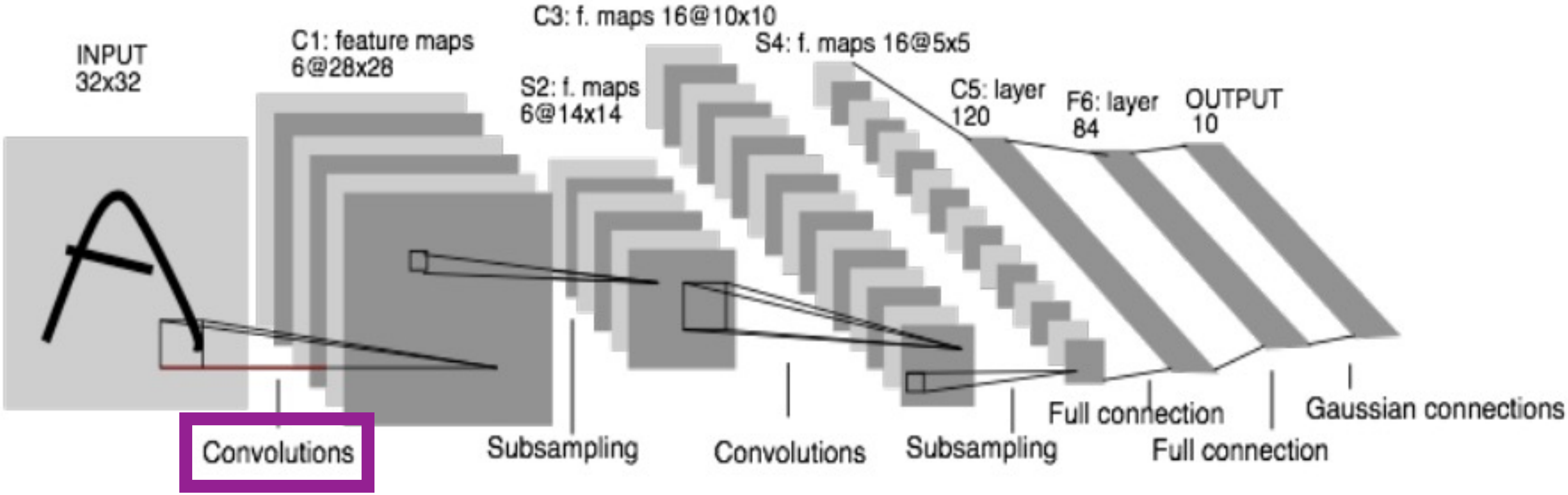
Y. Lecun ; L. Bottou ; Y. Bengio ; P. Haffner; Gradient-based learning applied to document recognition; 1998

LeNet: Architecture (like Neocognitron, has convolutional layers and pooling layers)



Y. Lecun ; L. Bottou ; Y. Bengio ; P. Haffner; Gradient-based learning applied to document recognition; 1998

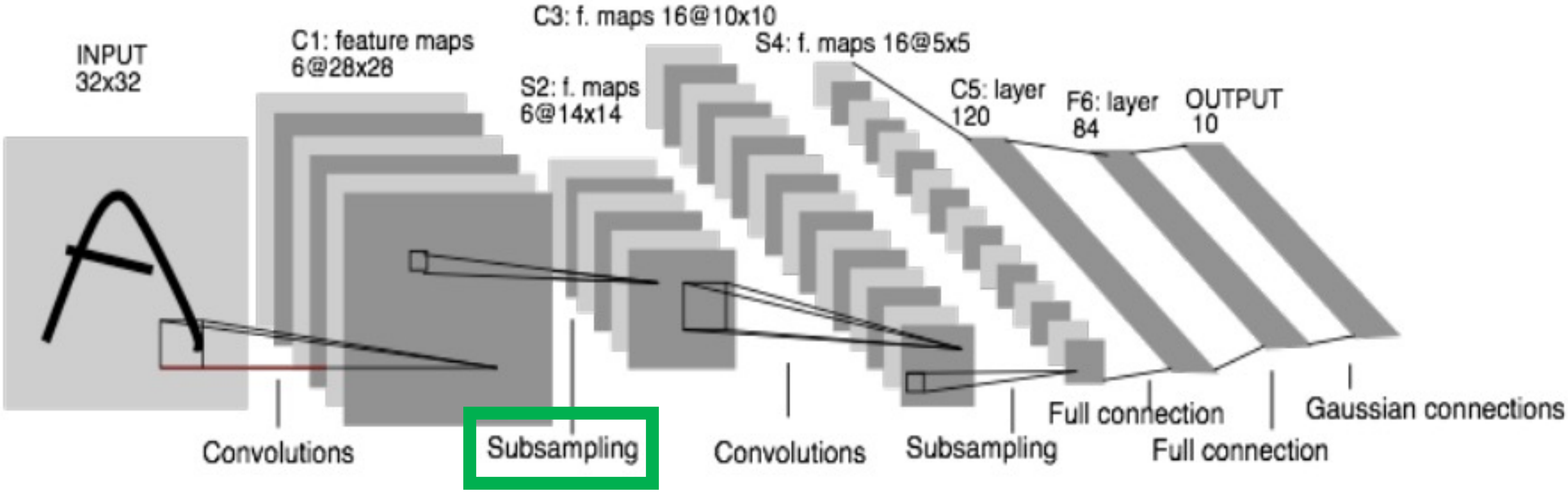
LeNet: Architecture (like Neocognitron, has convolutional layers and pooling layers)



How many filters are between the input and hidden layer 1?

Y. Lecun ; L. Bottou ; Y. Bengio ; P. Haffner; Gradient-based learning applied to document recognition; 1998

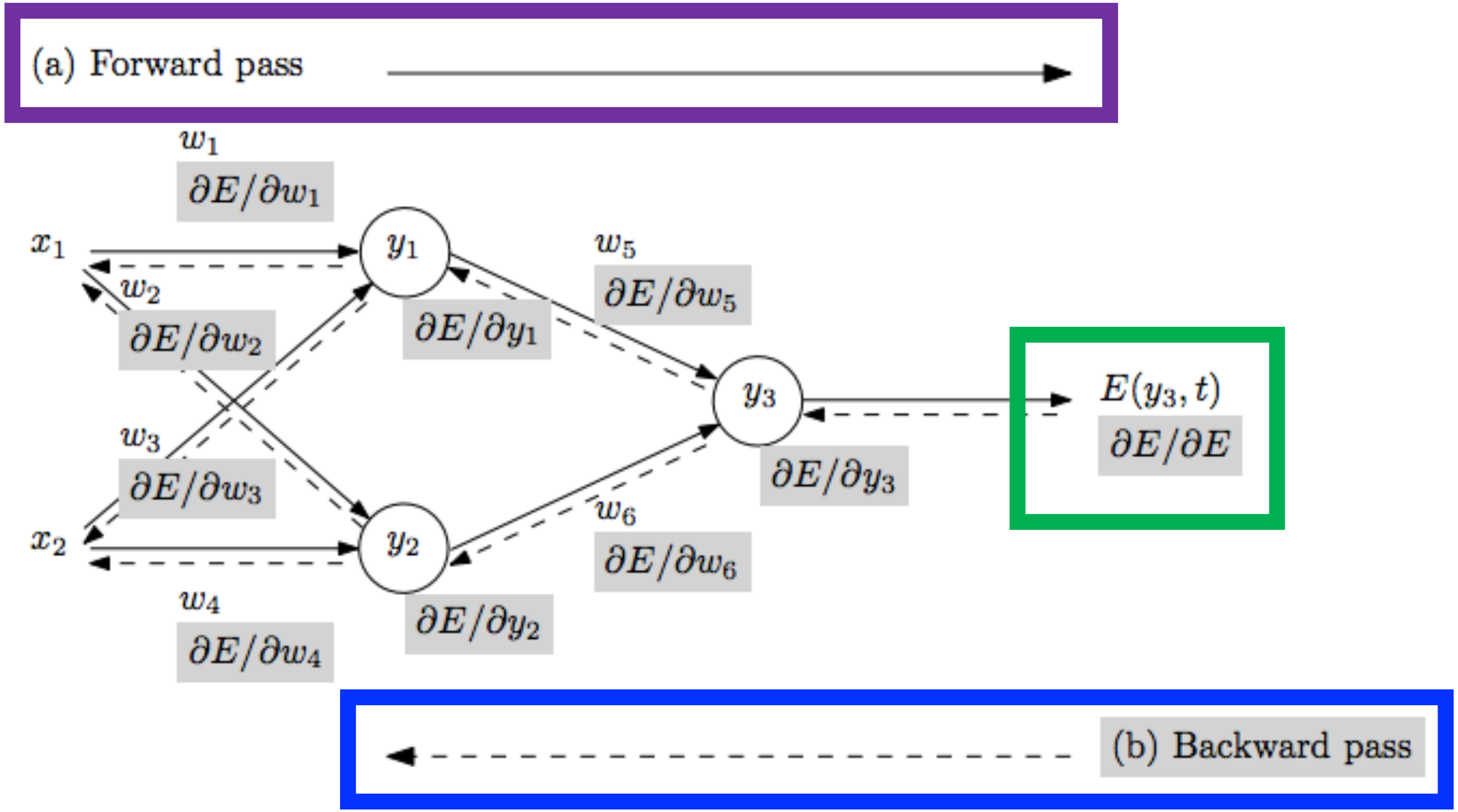
LeNet: Architecture (like Neocognitron, has convolutional layers and pooling layers)



What size of a neighborhood is used for this pooling layer?

Y. Lecun ; L. Bottou ; Y. Bengio ; P. Haffner; Gradient-based learning applied to document recognition; 1998

Training Procedure Approach (Key Novelty)

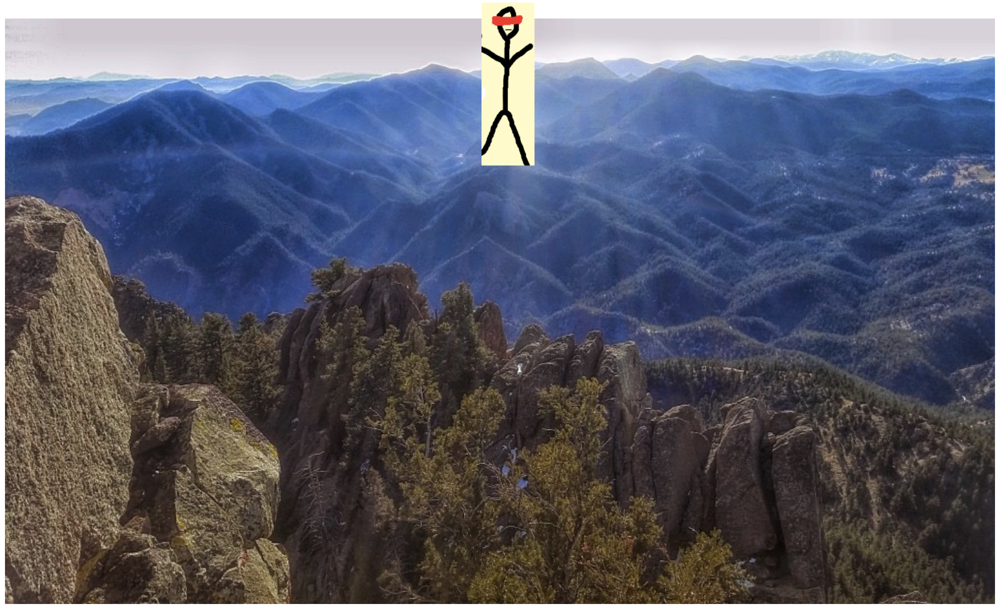


- Repeat until stopping criterion met:
  1. **Forward pass**: propagate training data through model to make prediction
  2. Quantify the dissatisfaction with a model's results on the training data
  3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each model parameter
  4. Account for weight sharing by using average of all connections for a parameter
  5. 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

Training Procedure Approach (Key Novelty)

Still obtain an error surface, E, based on the chosen objective function (e.g., using mean squared error, 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. Account for weight sharing by using average of all connections for a parameter
  5. Update each parameter using calculated gradients

Yann Lecun. Generalization and network design strategies. Technical Report, 1989

Training Procedure Approach (Key Novelty)

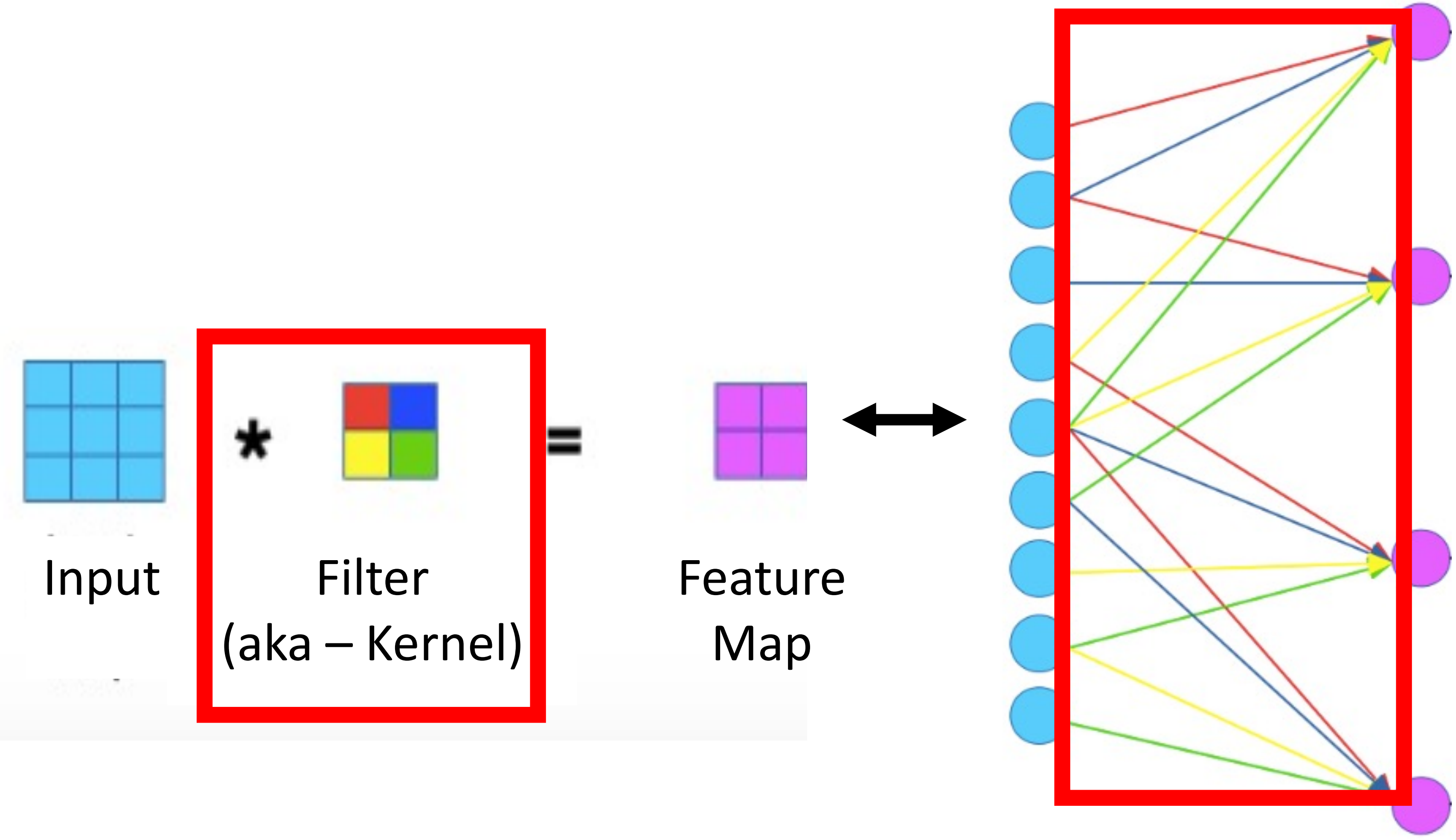
Still decide how to adjust model parameters (weights, biases) to push the predictions closer to the corresponding ground truth; a different gradient derivation used to tweak each value in each convolutional filter

(covered in Section 6.3 of Kamath book and https://www.jefkine.com/general/2016/09/05/backpropagation-in-convolutional-neural-networks/)

- 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. Account for weight sharing by using average of all connections for a parameter
  5. Update each parameter using calculated gradients

Yann Lecun. Generalization and network design strategies. Technical Report, 1989

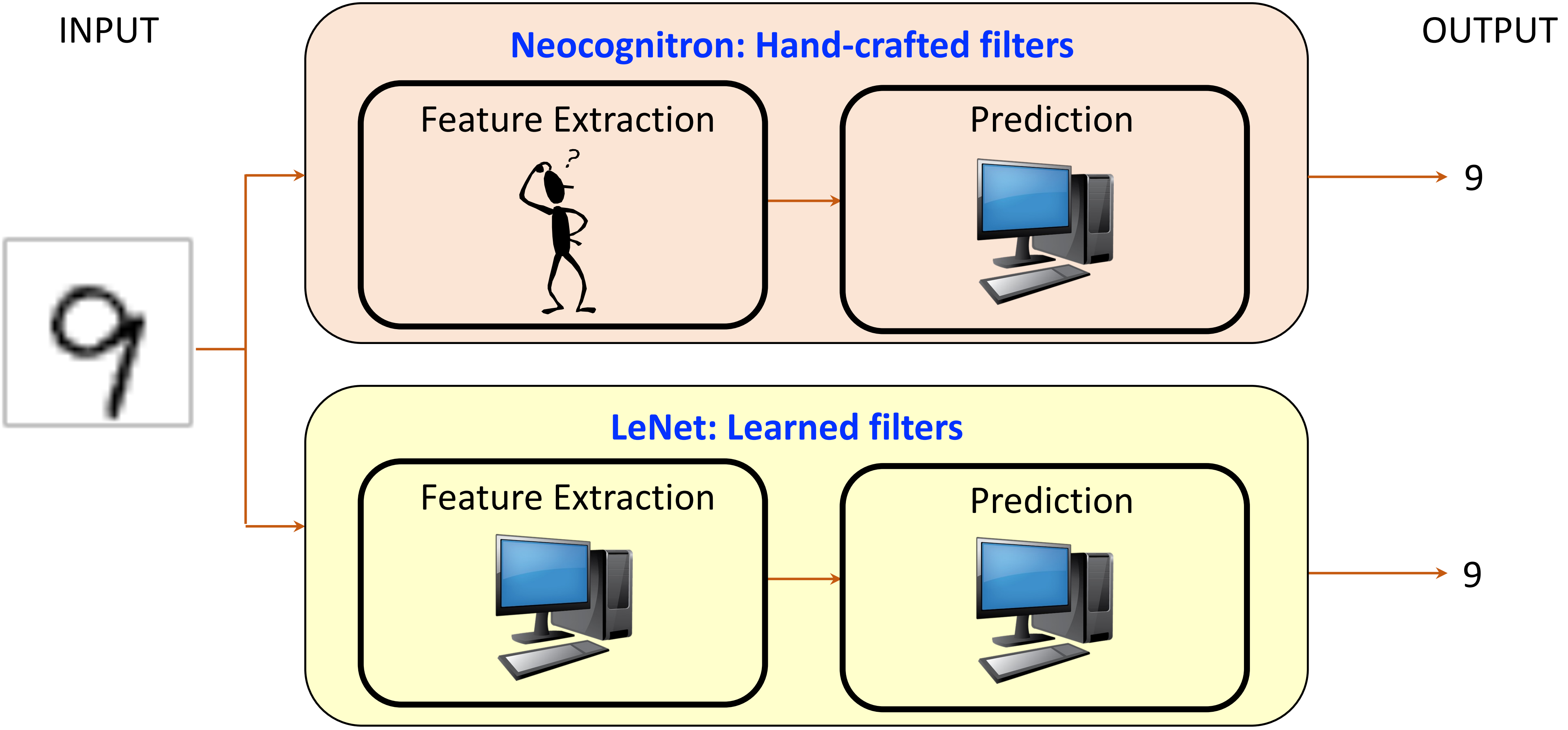
Training Procedure Approach (Key Novelty)



- 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. Account for weight sharing by using average of all connections for a parameter
    5. Update each parameter using calculated gradients

https://www.jefkine.com/general/2016/09/05/backpropagation-in-convolutional-neural-networks/

LeNet vs Neocognitron

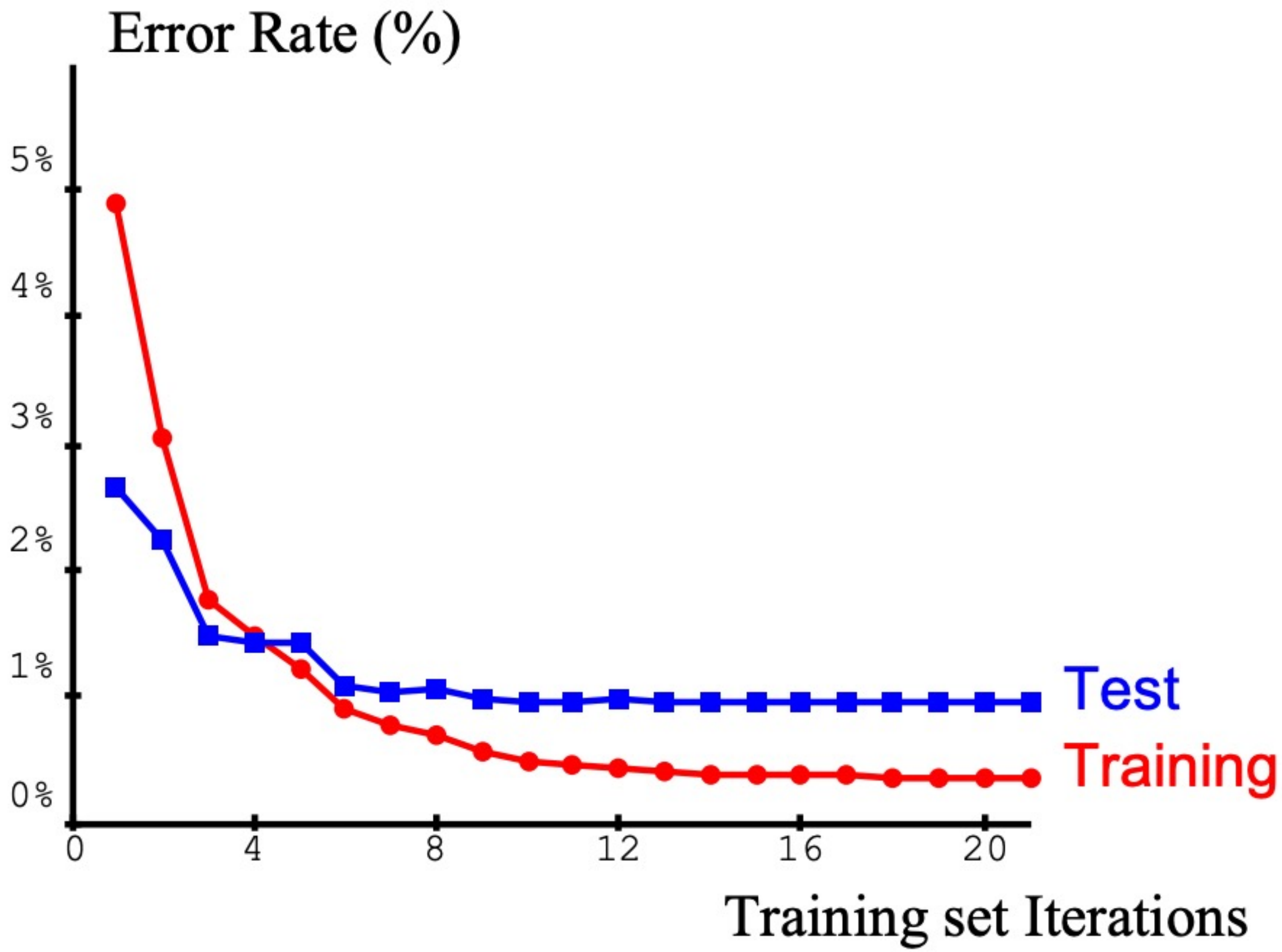


LeNet Analysis

How many epochs are needed for training to converge?

Why might overfitting not arise with more training?

- Learning rate was too large for the model to settle in a local minimum but rather oscillated randomly



Y. Lecun ; L. Bottou ; Y. Bengio ; P. Haffner; Gradient-based learning applied to document recognition; 1998

LeNet Analysis

All 82 mislabeled examples (correct answer on left, predicted answer on right):

Why might the model be making mistakes?

- Insufficient representation in the training data
- Ambiguity

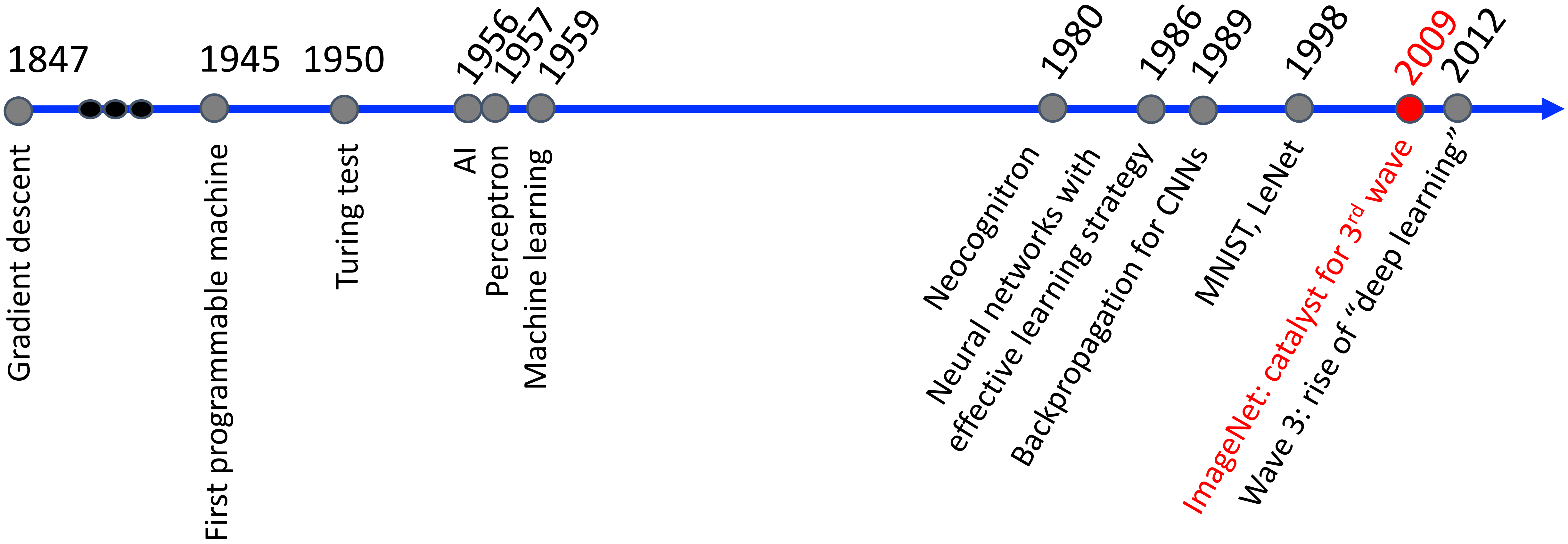
Y. Lecun ; L. Bottou ; Y. Bengio ; P. Haffner; Gradient-based learning applied to document recognition; 1998

LeNet, designed on the MNIST Challenge, was used to read over 10% of checks in North America in the 1990s, reading millions of checks every month

Today’s Topics

- Computer vision
- Era of dataset challenges
- MNIST challenge winner: LeNet
- ImageNet challenge winners: deeper learning (AlexNet, VGG, ResNet)
- Programming tutorial

Historical Context



ImageNet: Predict Category from 1000 Options

- Evaluation metric: % correct (top-1 and top-5 predictions)
- Dataset: ~1.5 million images
- Source: images scraped from search engines, such as Flickr, and labeled by crowdworkers

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

ImageNet vs MNIST

- 3D objects in natural backgrounds
- Many more categories

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

Rise of "Deep Learning"

Progress of models on ImageNet (Top 5 Error)

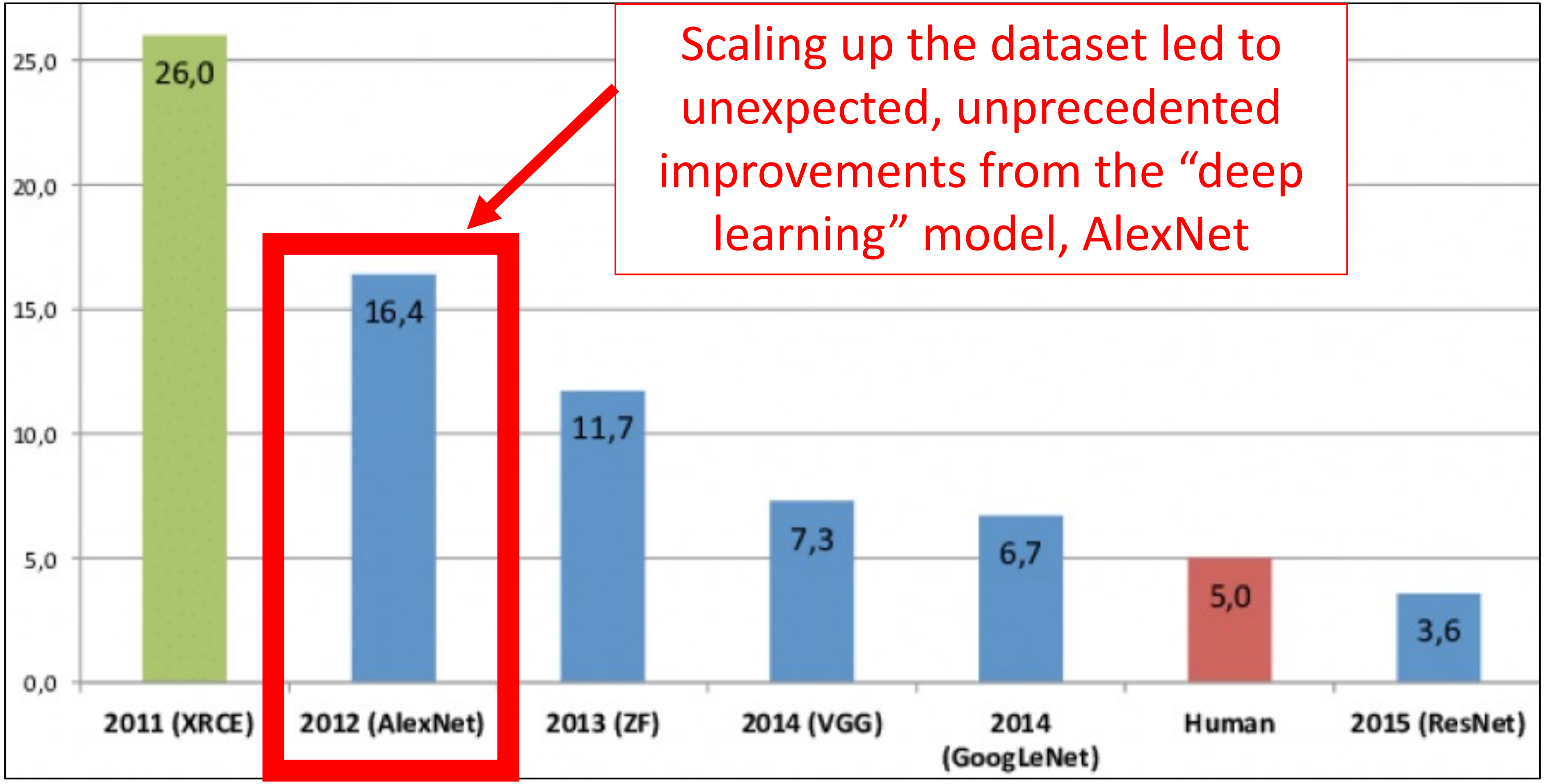
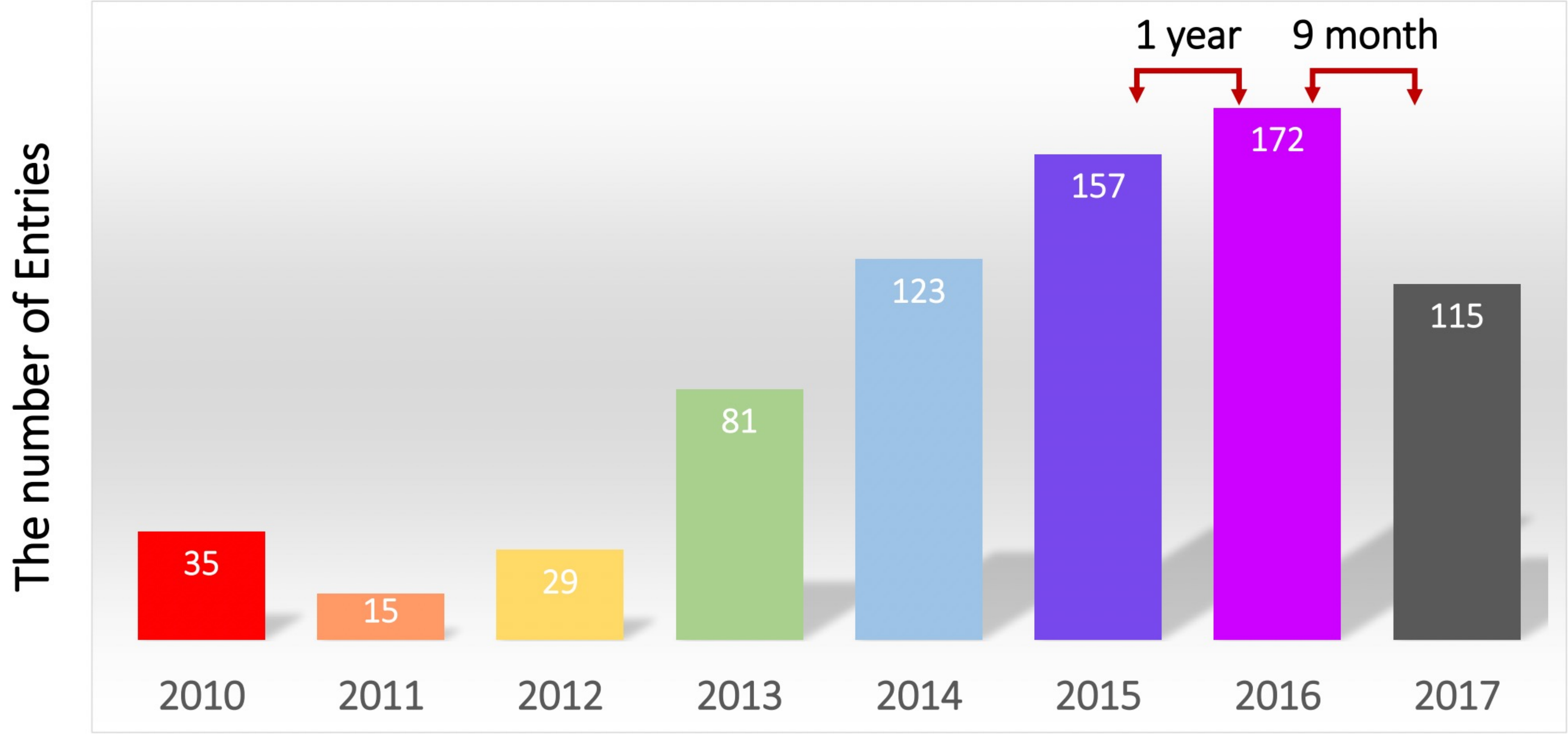


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

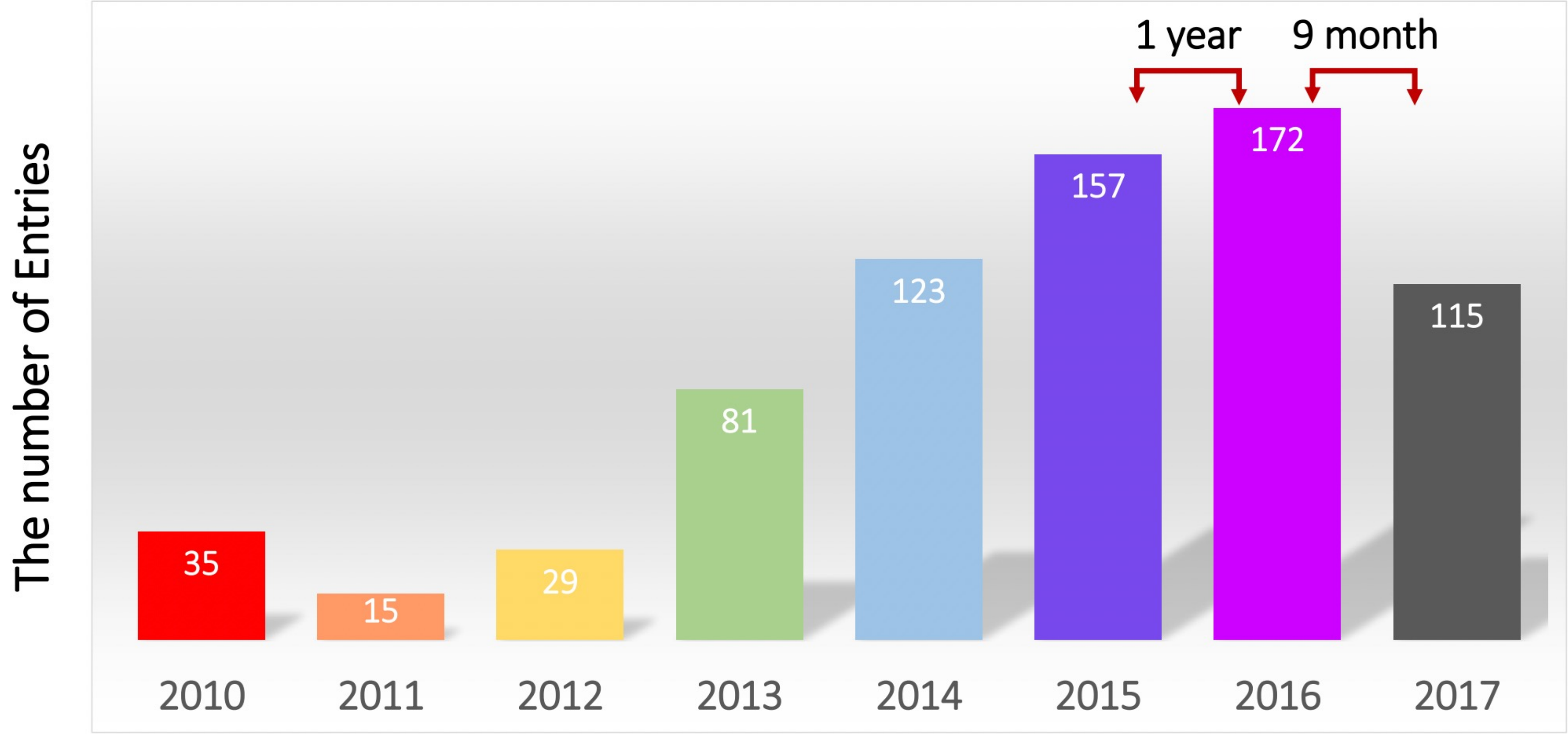
Rise of “Deep Learning” Following AlexNet



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)

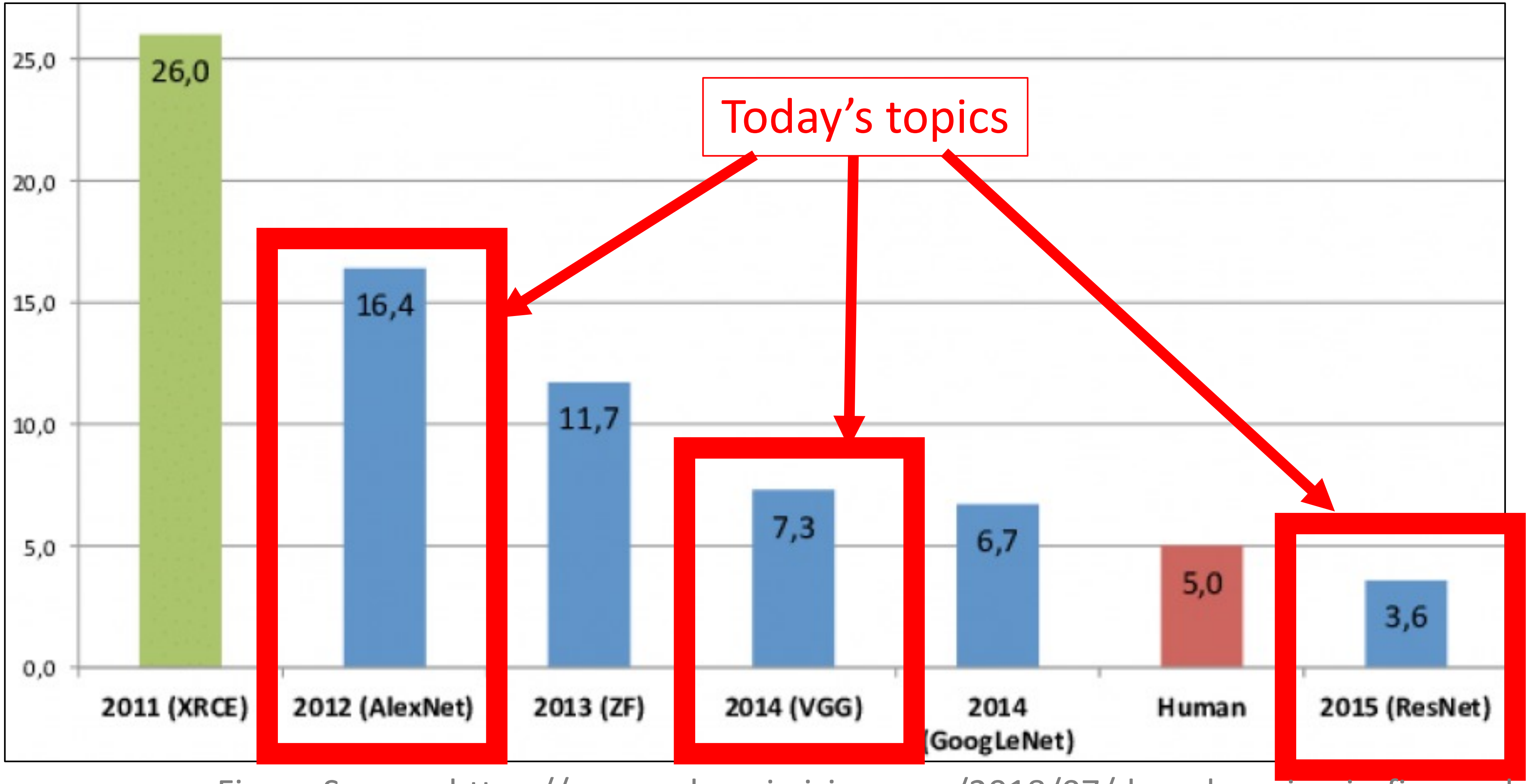
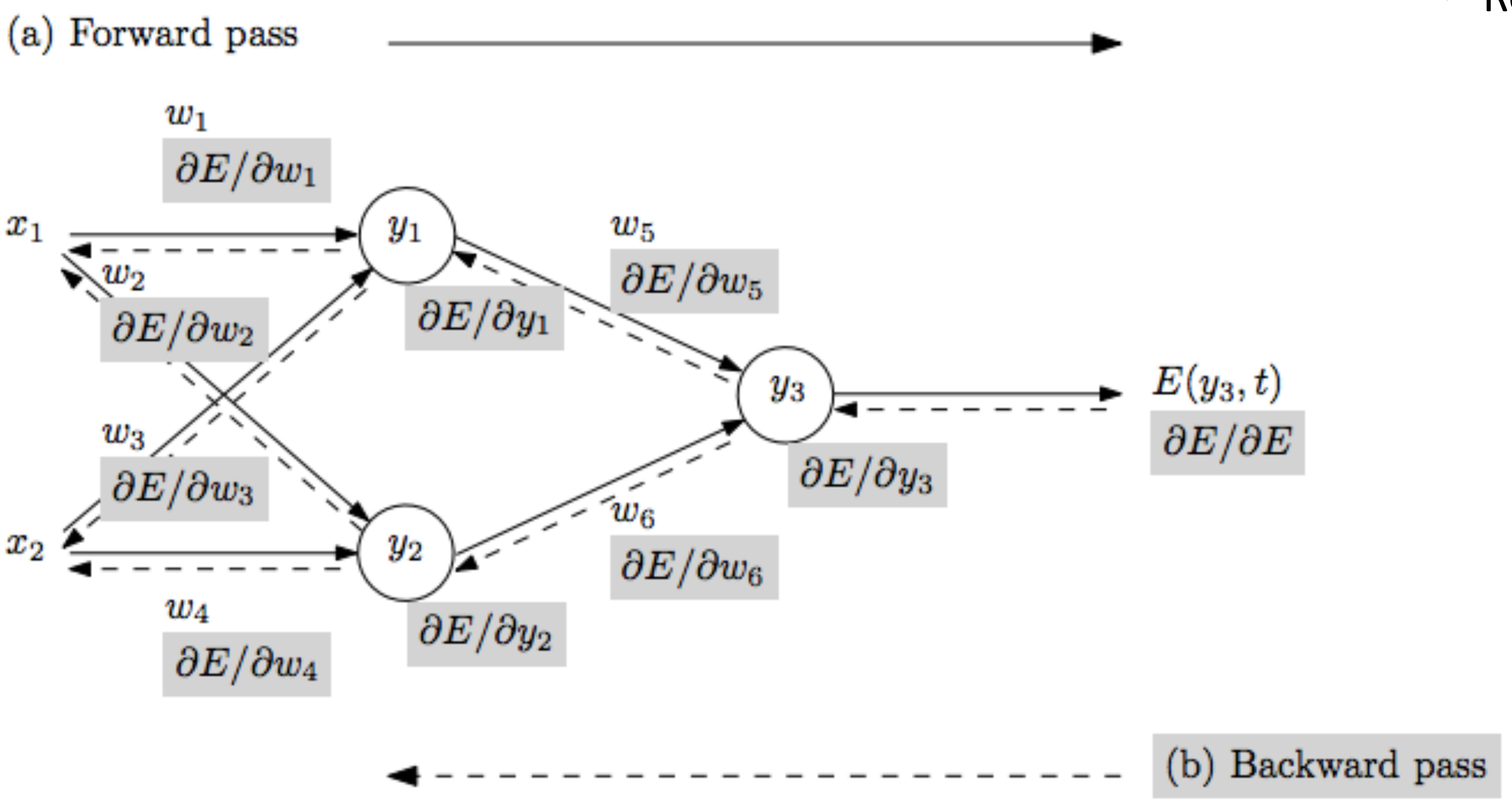


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

Why It Is Difficult to Achieve Better Performance with CNNs That Are Deeper: Vanishing Gradients



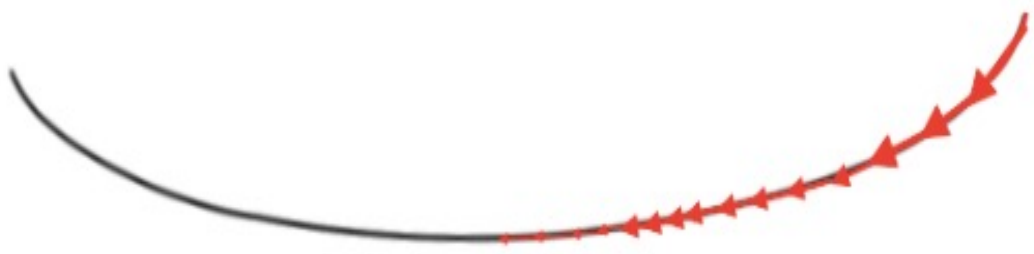
- Repeat until stopping criterion met:
  1. **Forward pass**: propagate training data through model to make prediction
  2. Quantify the dissatisfaction with a model's results on the training data
  3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each model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

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

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

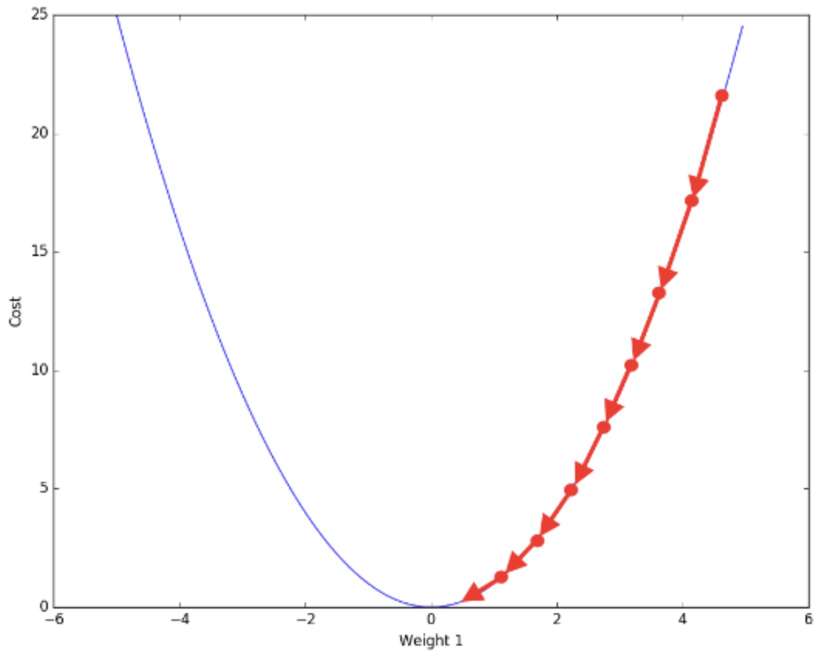
Why It Is Difficult to Achieve Better Performance with CNNs That Are Deeper: Vanishing Gradients

Cost Function 1



Small gradients so weights change slowly

Cost Function 2



Large gradients so weights change quickly

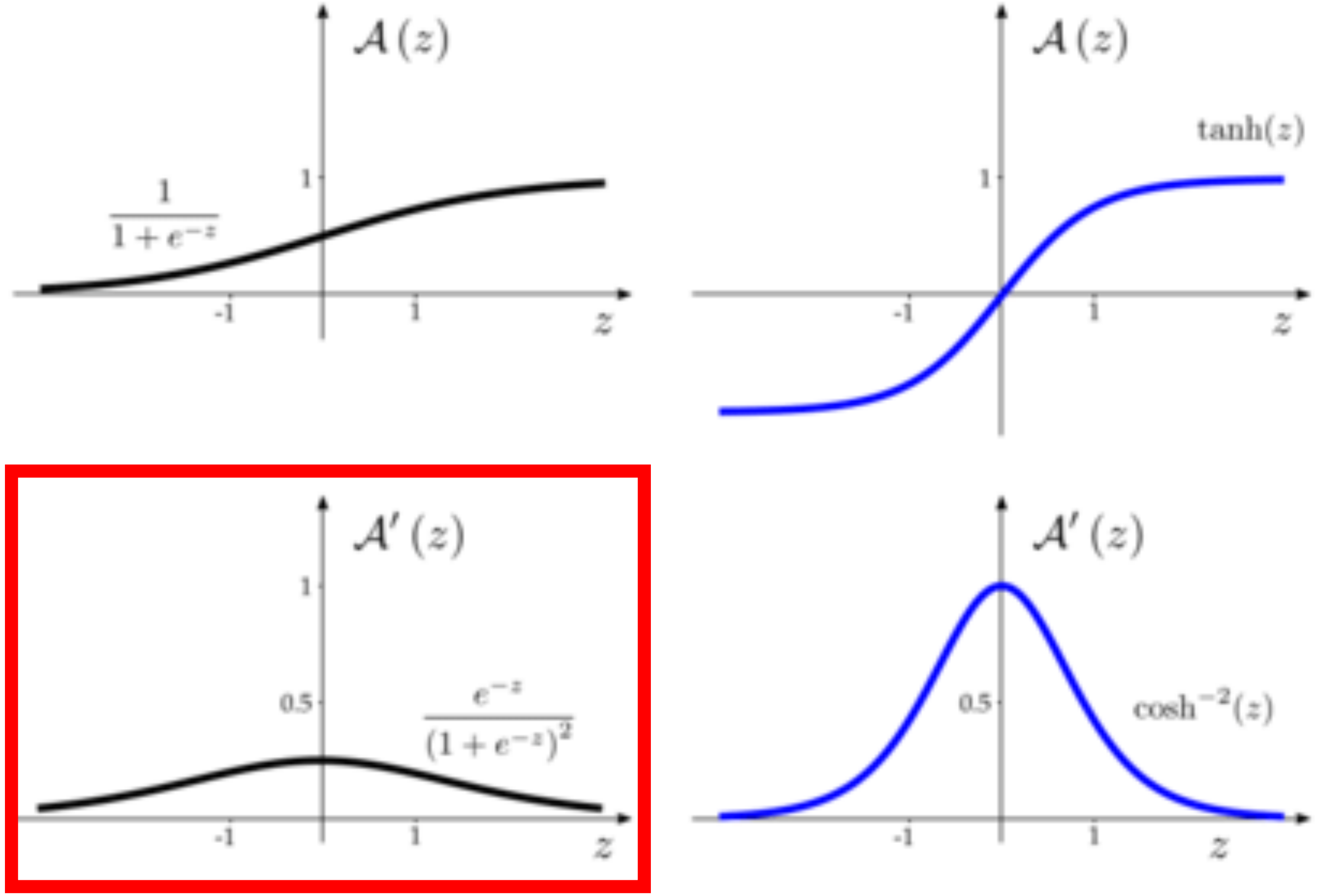
https://ayearofai.com/rohan-4-the-vanishing-gradient-problem-ec68f76ffb9b

Why It Is Difficult to Achieve Better Performance with CNNs That Are Deeper: Vanishing Gradients

Recall activation functions and their derivatives:

Sigmoid

Tanh

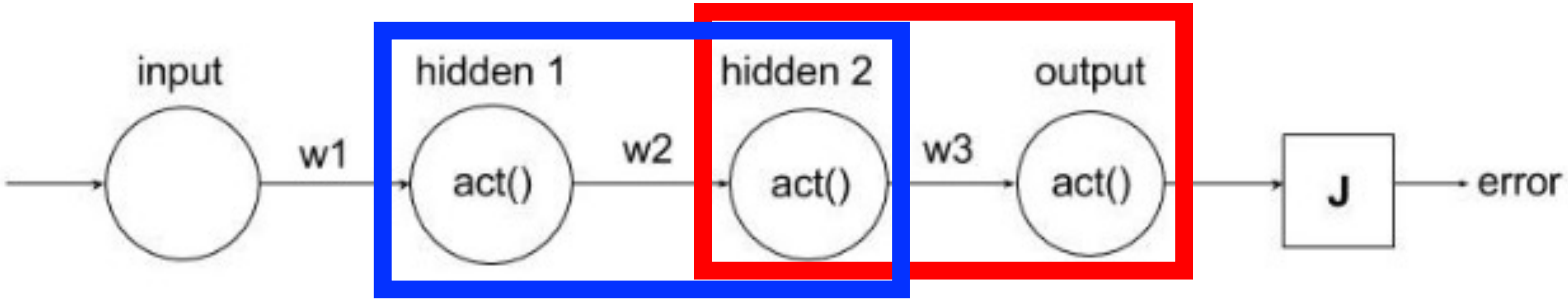


Ranges from 0 to 0.25

Masi et al. Journal of the Mechanics and Physics of Solids. 2021

Vanishing Gradient Problem (e.g., sigmoid)

- Toy example:



- Error Derivative with respect to weight w1:

$$\frac{\partial error}{\partial w1} = \frac{\partial error}{\partial output} * \boxed{\frac{\partial output}{\partial hidden2}} * \boxed{\frac{\partial hidden2}{\partial hidden1}} * \frac{\partial hidden1}{\partial w1}$$

Derivative of sigmoid activation function: (0 to 1/4]

Derivative of sigmoid activation function: (0 to 1/4]

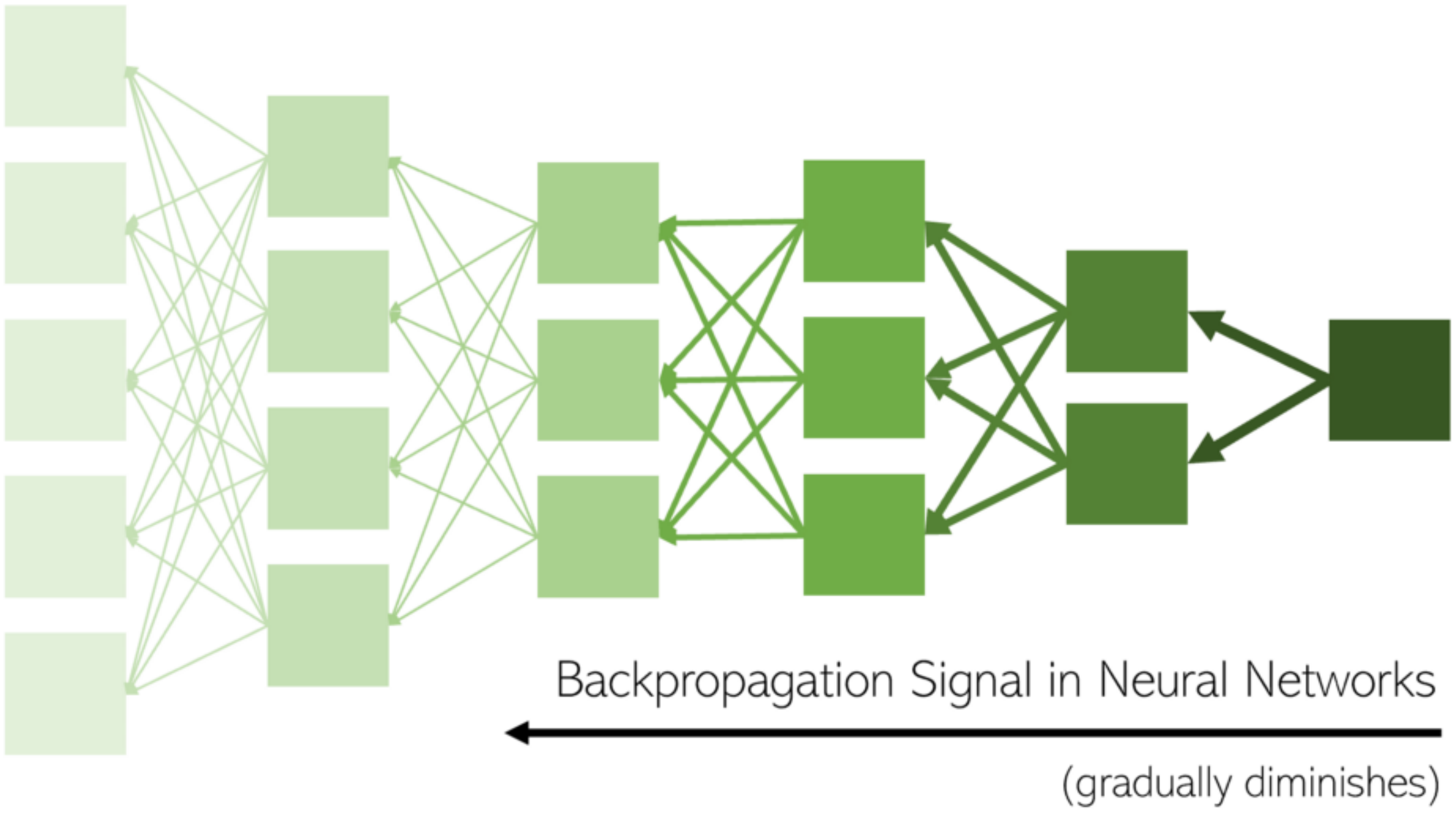
Problem: What happens as you multiply more numbers smaller than 1?

Gradient decreases as further from the last layer… and so weights barely change at training!

https://ayearofai.com/rohan-4-the-vanishing-gradient-problem-ec68f76ffb9b

Vanishing Gradient Problem (e.g., sigmoid)

Smallest gradients at **earliest layers make them slowest to train**, yet later layers depend on those earlier layers to do something useful; consequently, NNs struggle with garbage in means garbage out



https://towardsdatascience.com/batch-normalization-the-greatest-breakthrough-in-deep-learning-77e64909d81d

How can we avoid the vanishing gradient problem?

AlexNet: A Deeper CNN

Progress of models on ImageNet (Top 5 Error)

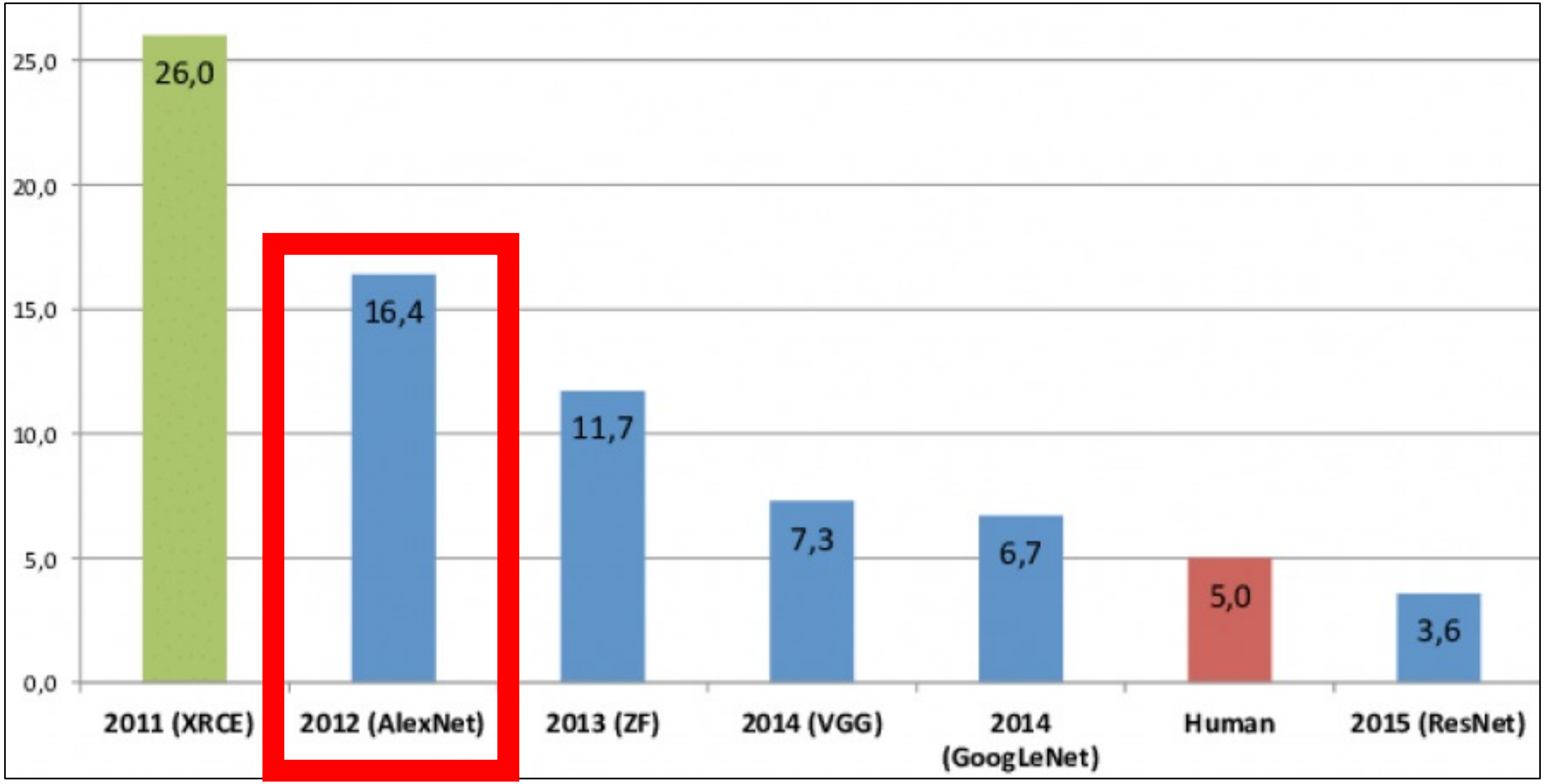
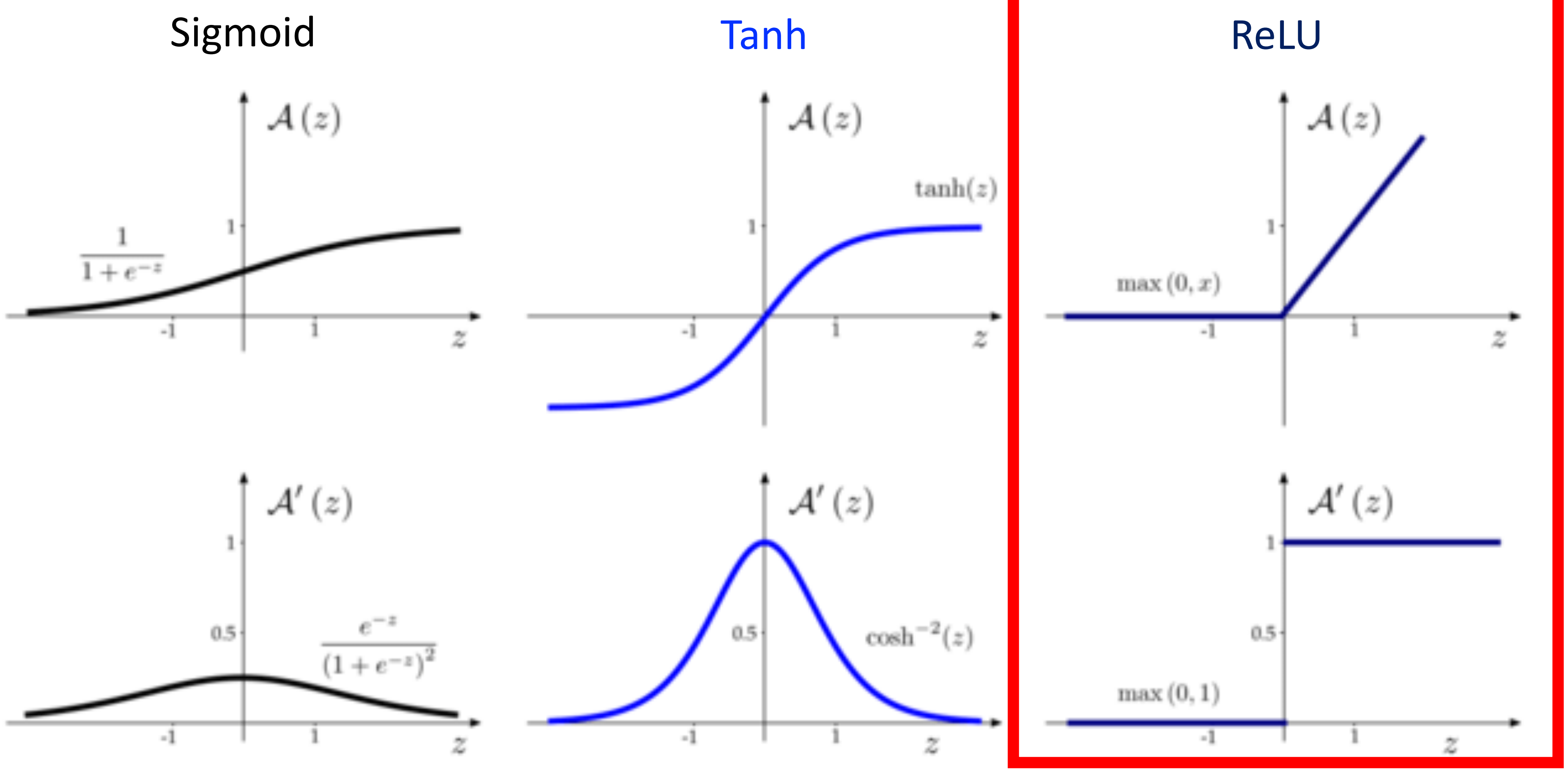


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

Key Idea: Non-Saturating Activation Functions

Use activation functions with derivative value equal to 1 (i.e., 1x1x1… doesn't vanish)



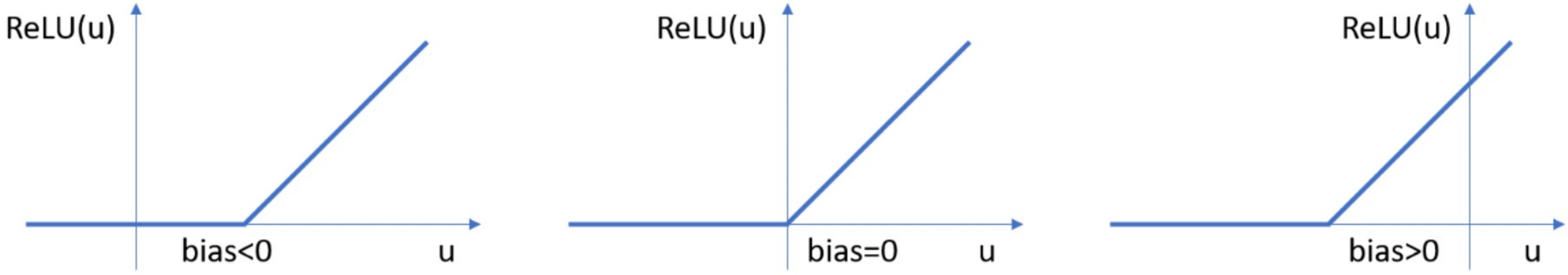
Benefits:

- Can preserve gradient
- Fast to compute
- "Dying neurons" contribute to network sparsity and so reduced model complexity

Masi et al. Journal of the Mechanics and Physics of Solids. 2021

Key Idea: Non-Saturating Activation Functions

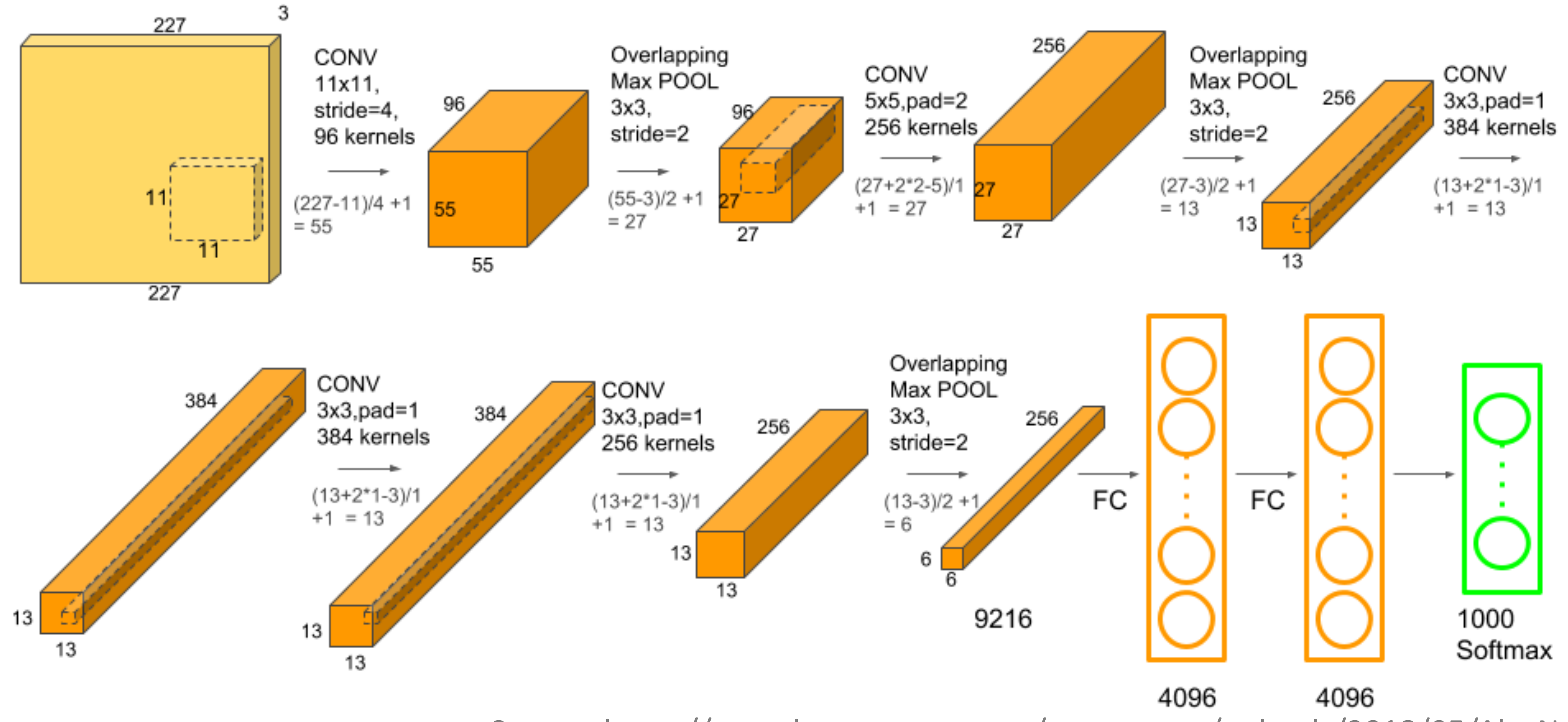
- Influence of bias term with ReLU



- What is the impact of a positive bias value?
- What is the impact of a negative bias value?

https://towardsdatascience.com/visualizing-the-fundamentals-of-convolutional-neural-networks-6021e5b07f69

AlexNet Architecture: Similar to LeNet But With More Convolutional and Pooling Layers



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

AlexNet Architecture

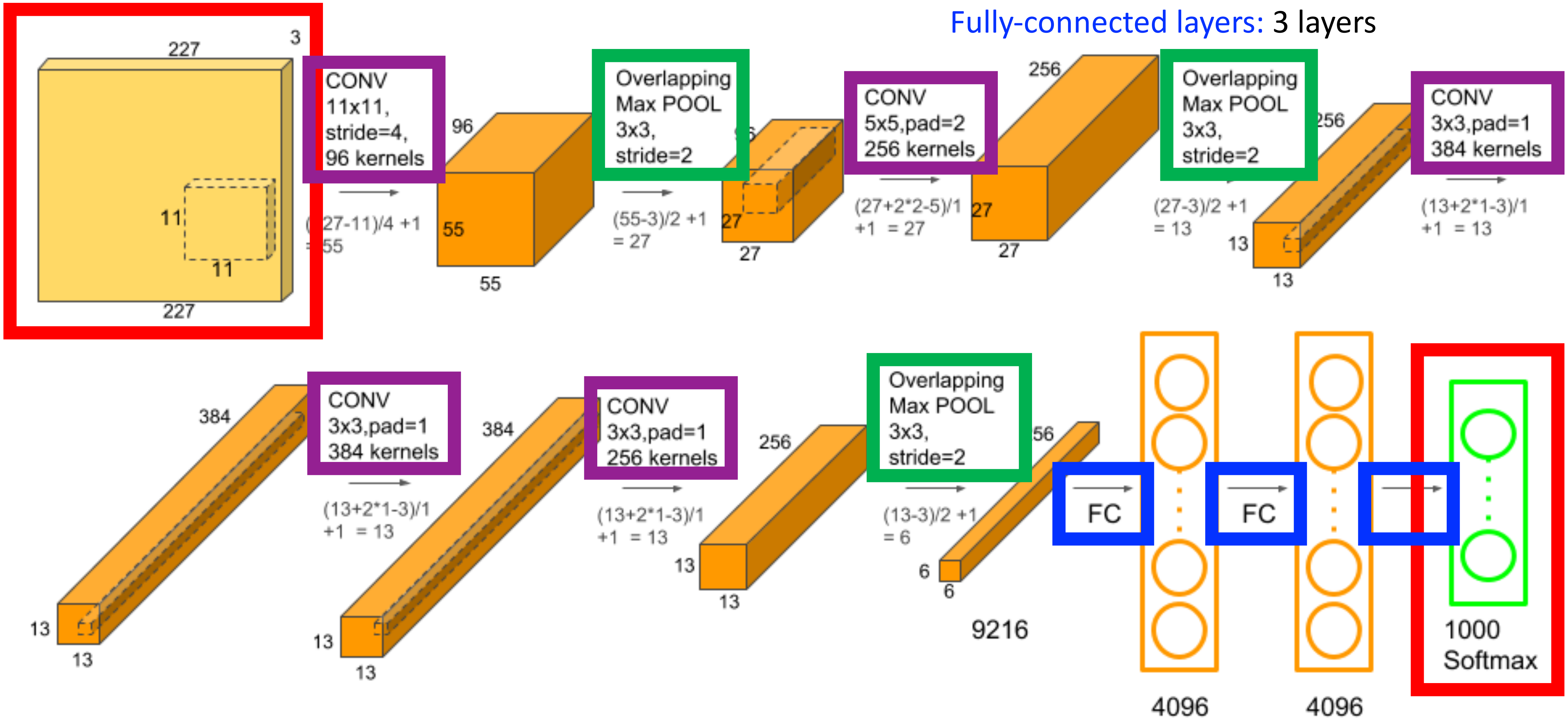
Input: RGB image resized to fixed input size

Output: 1000 class probabilities (sums to 1)

Convolutional layers: 5 layers

Pooling Layers: 3 layers

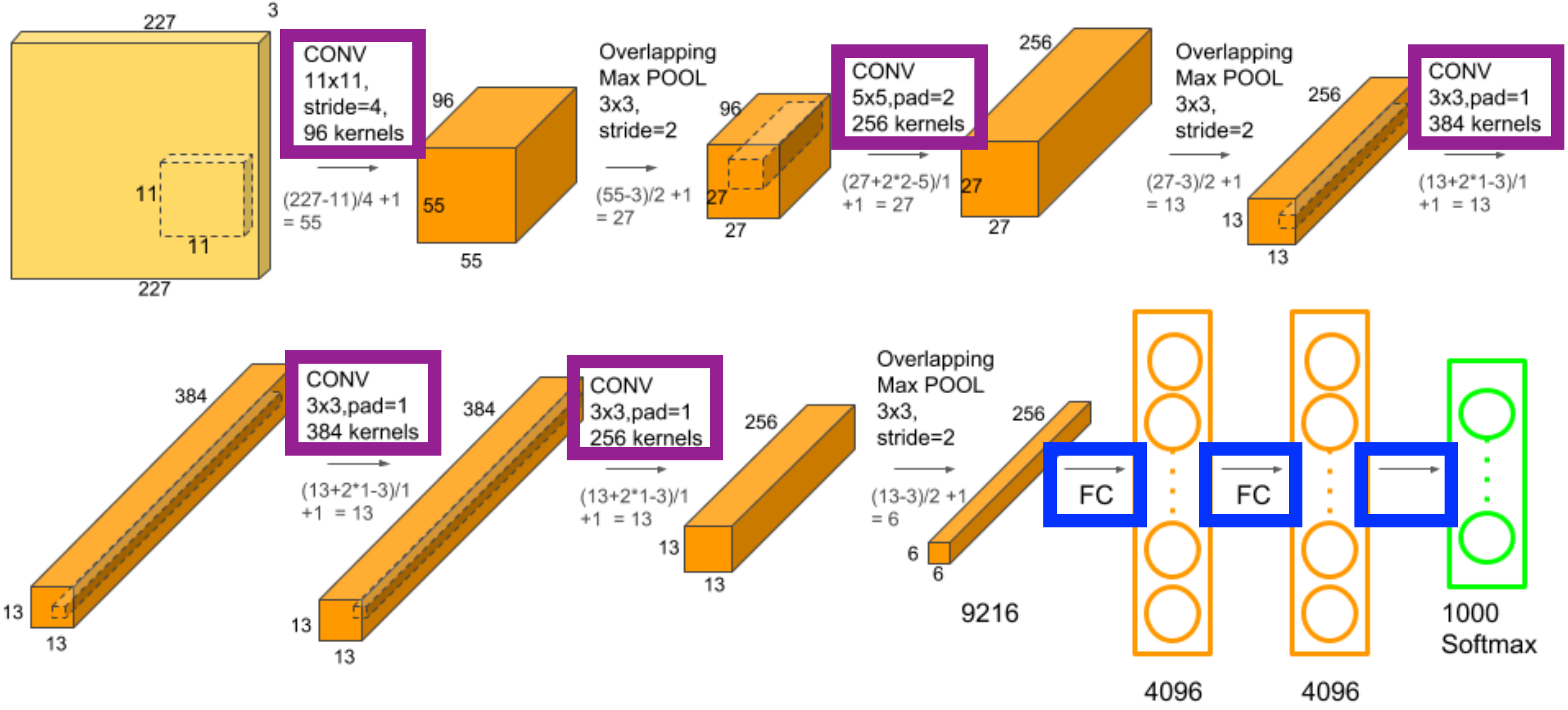
Fully-connected layers: 3 layers



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

AlexNet Architecture

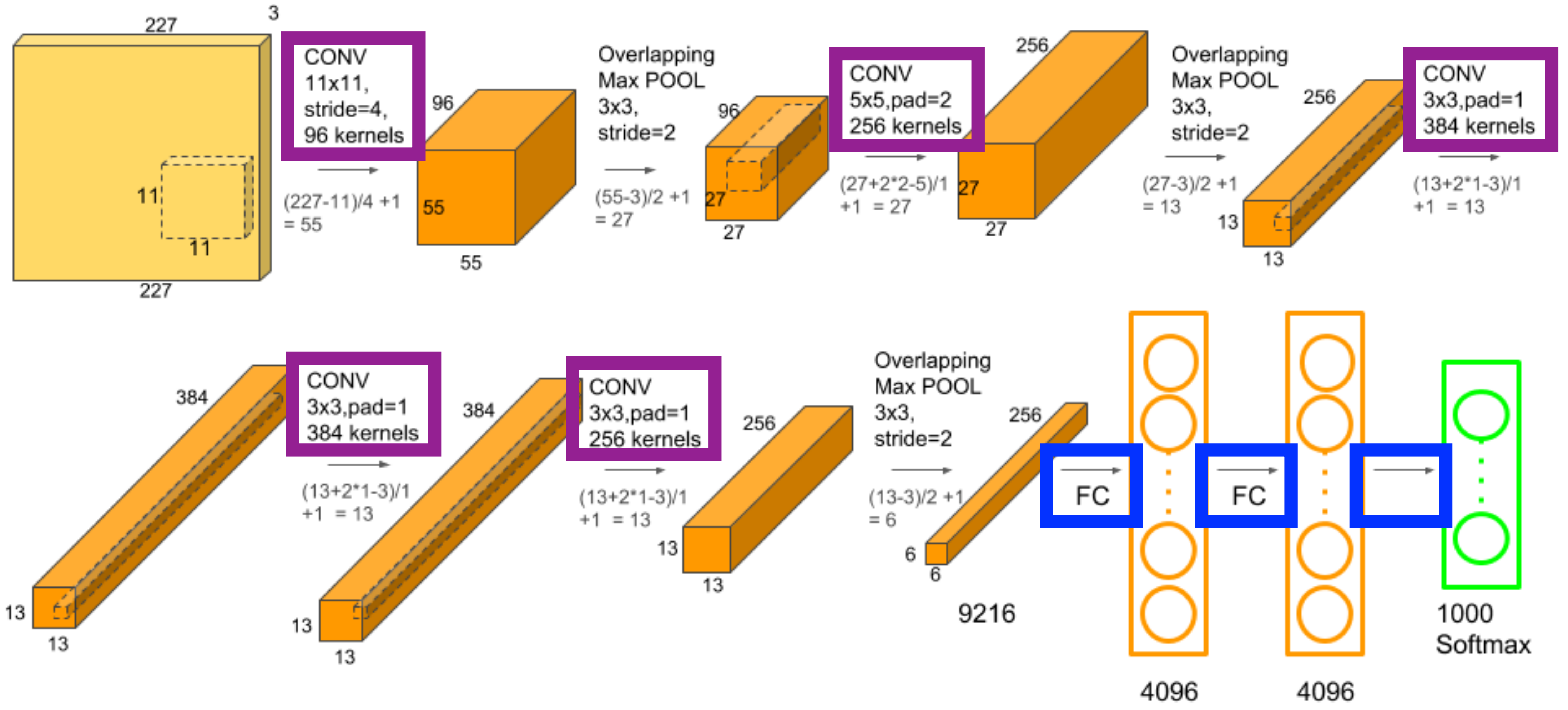
How many layers have model parameters that need to be learned?



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

AlexNet Architecture

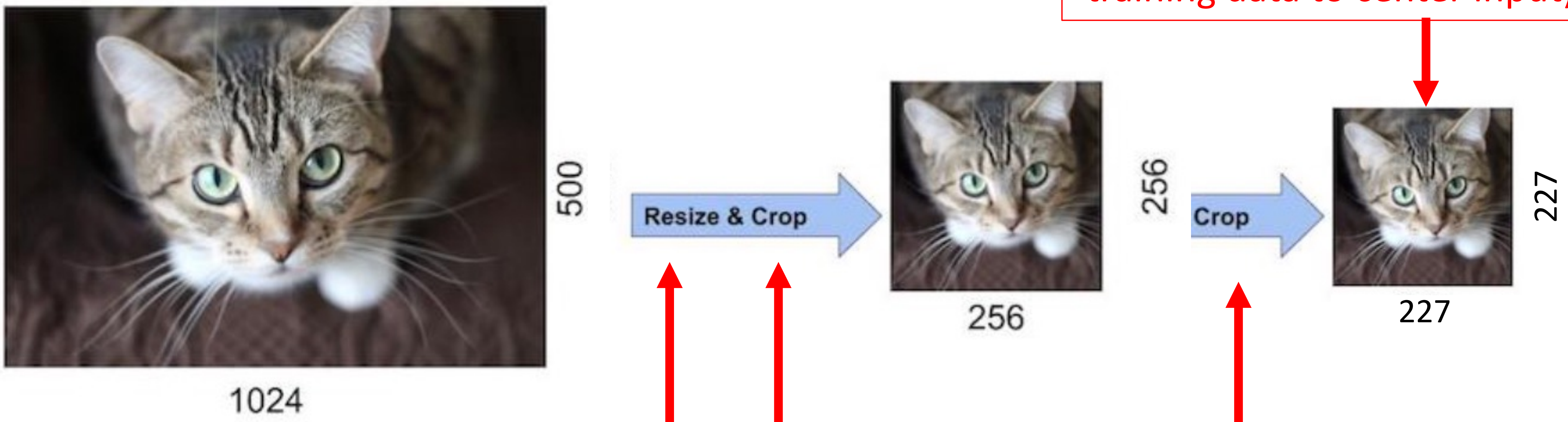
Altogether, 60 million model parameters must be learned!



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

Implementation Detail: Input Preprocessing

(subtract mean image from training data to center input)



(side with smaller dimension)

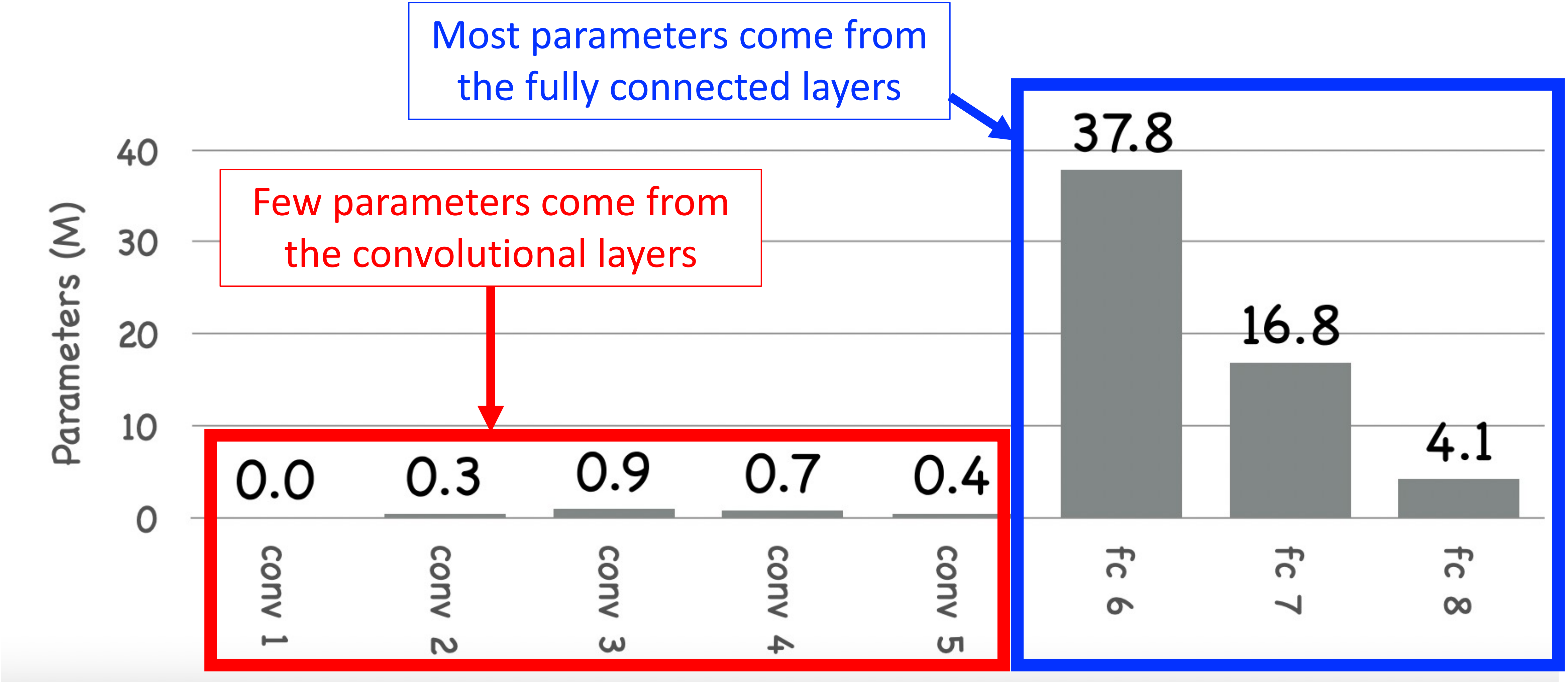
(retain center patch)

(random location)

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

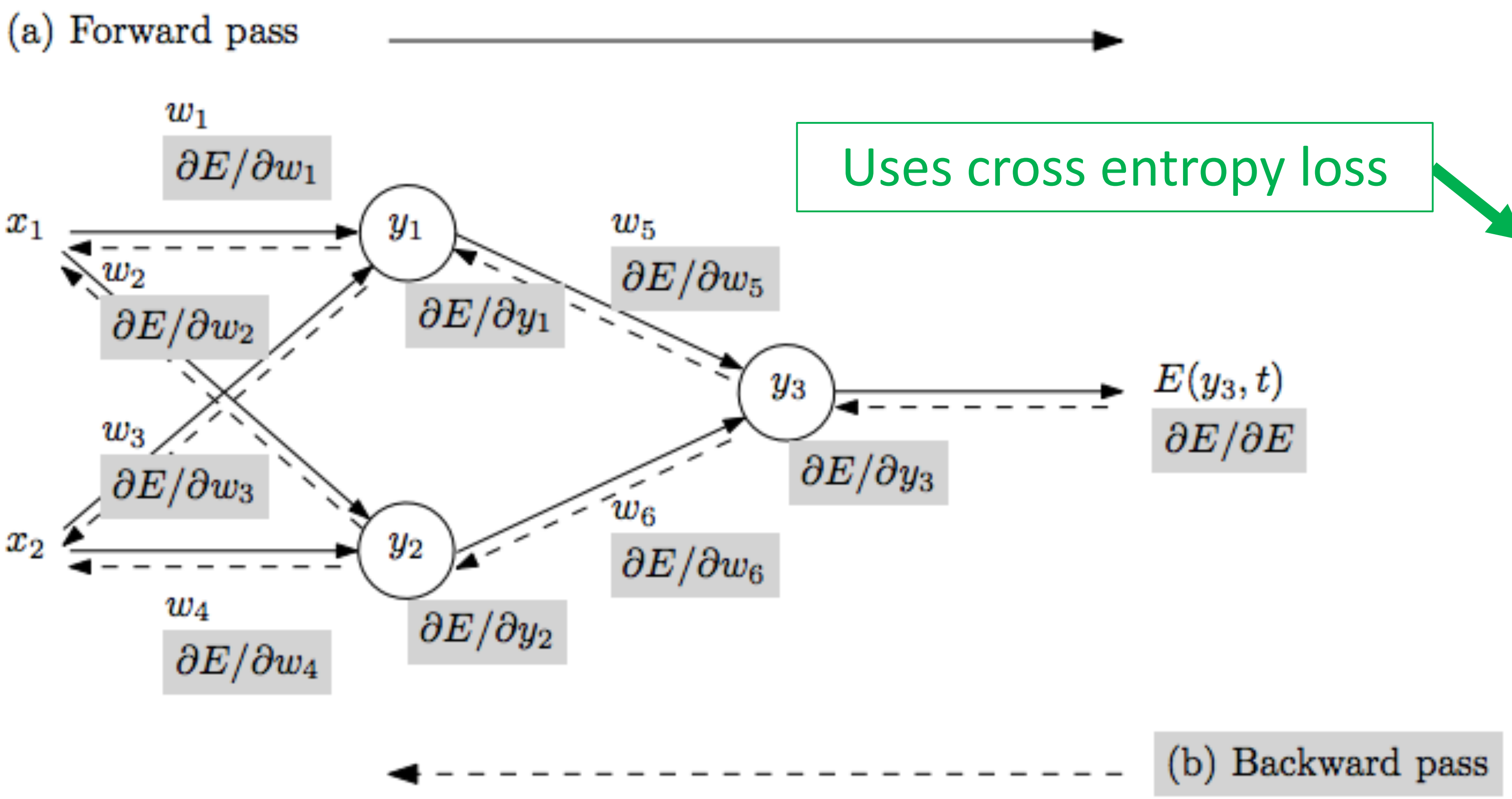
AlexNet Architecture

Altogether, 60 million model parameters must be learned!



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

AlexNet Training: 90 Epochs



Uses 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

AlexNet: Key Tricks for Going Deeper

- ReLU instead of sigmoid or tanh activation functions

- Regularization techniques: to be covered next lecture
    1. Data augmentation
    2. Dropout in fully connected layers
    3. L2 parameter norm penalty

- Trained across two GPUs

AlexNet Analysis

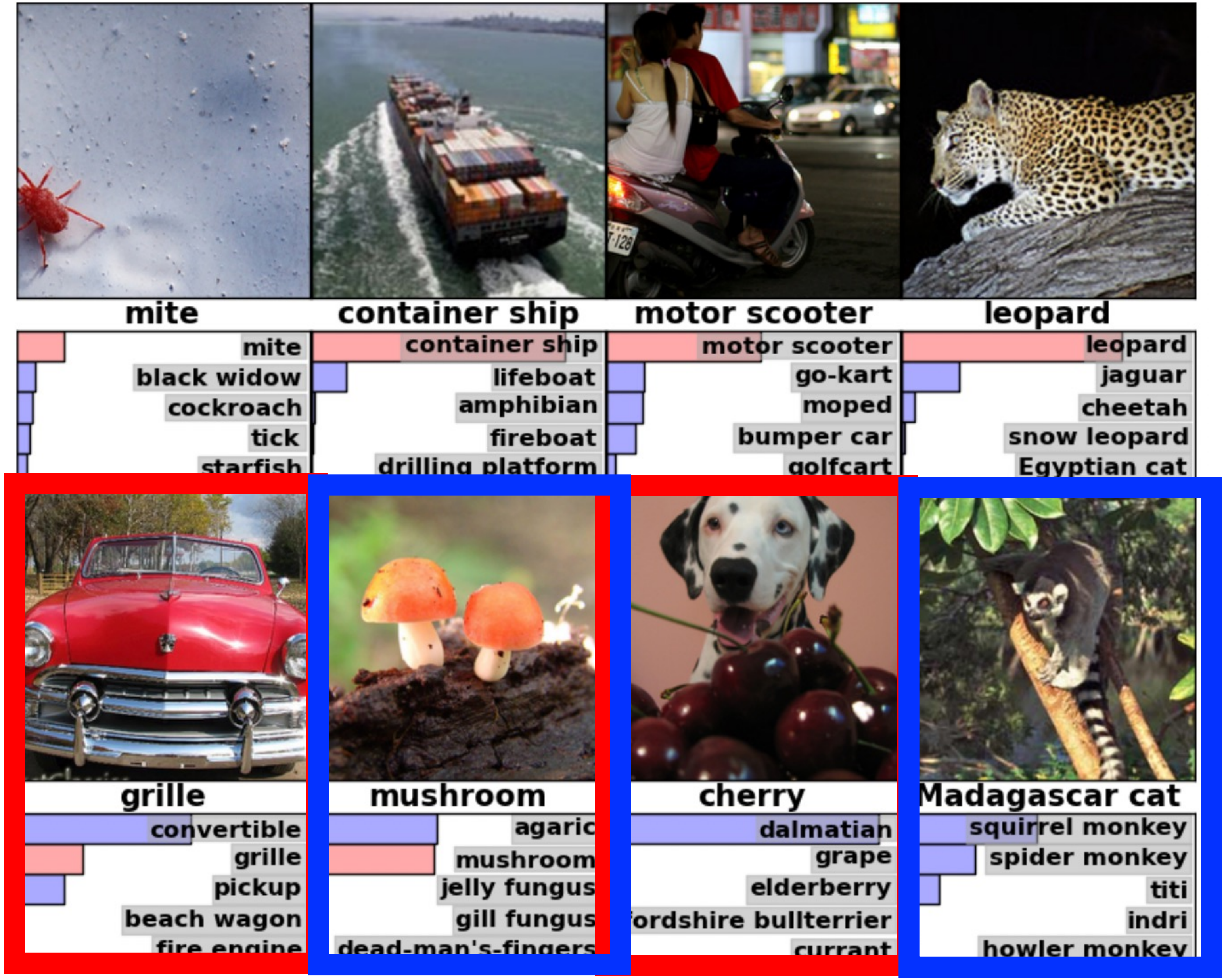
8 examples of predictions, correct and incorrect

When/why might the model succeed?

- Single well-defined object (even if off-centered)

When/why might the model fail?

- Ambiguity
- Similar categories



Krizhevsky et al. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, 2017

VGG: A Deeper CNN

Progress of models on ImageNet (Top 5 Error)

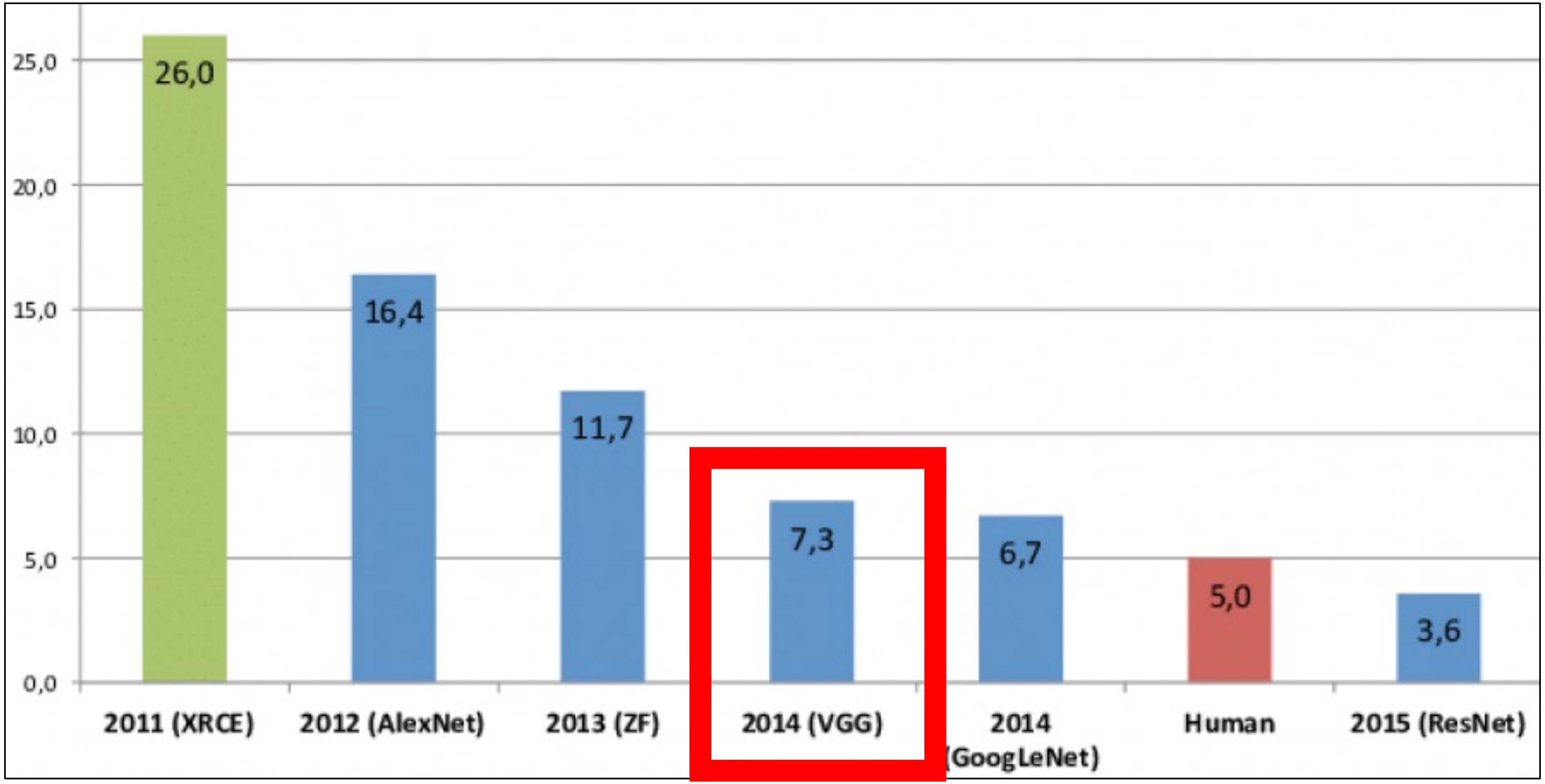


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

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

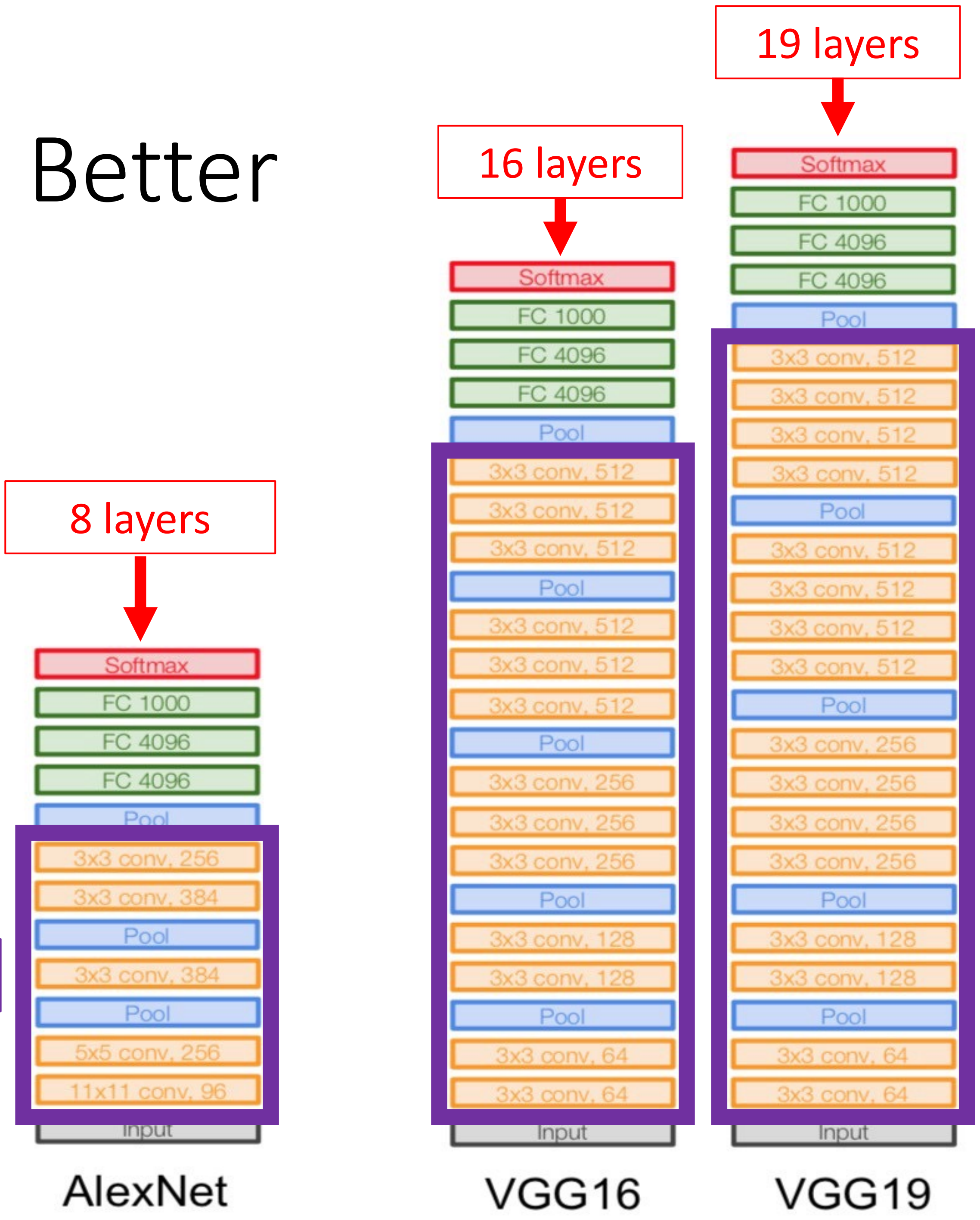


Figure Source (edited to fix mistakes): https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

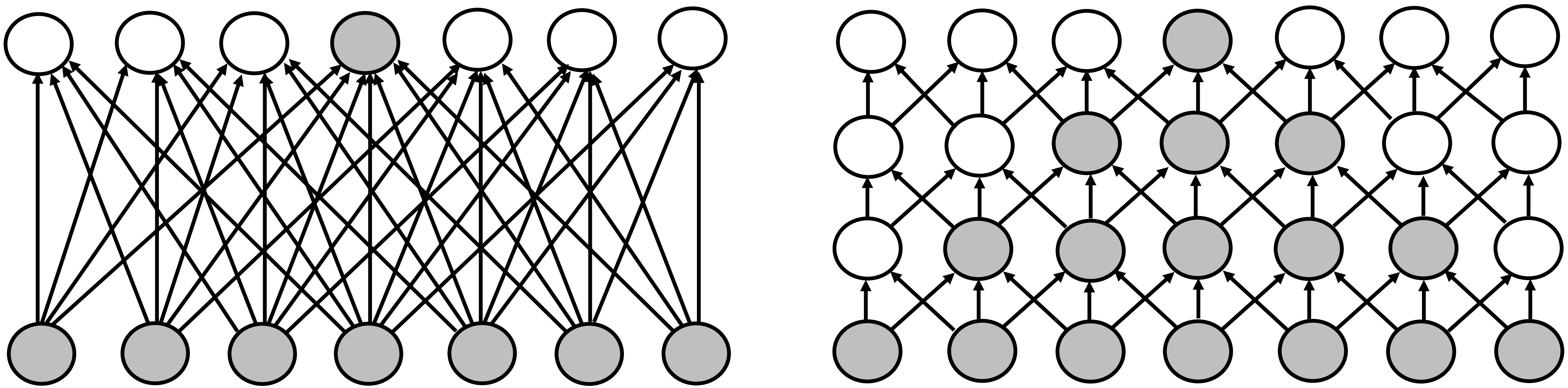
Key Idea: Smaller Convolutional Filters

- 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

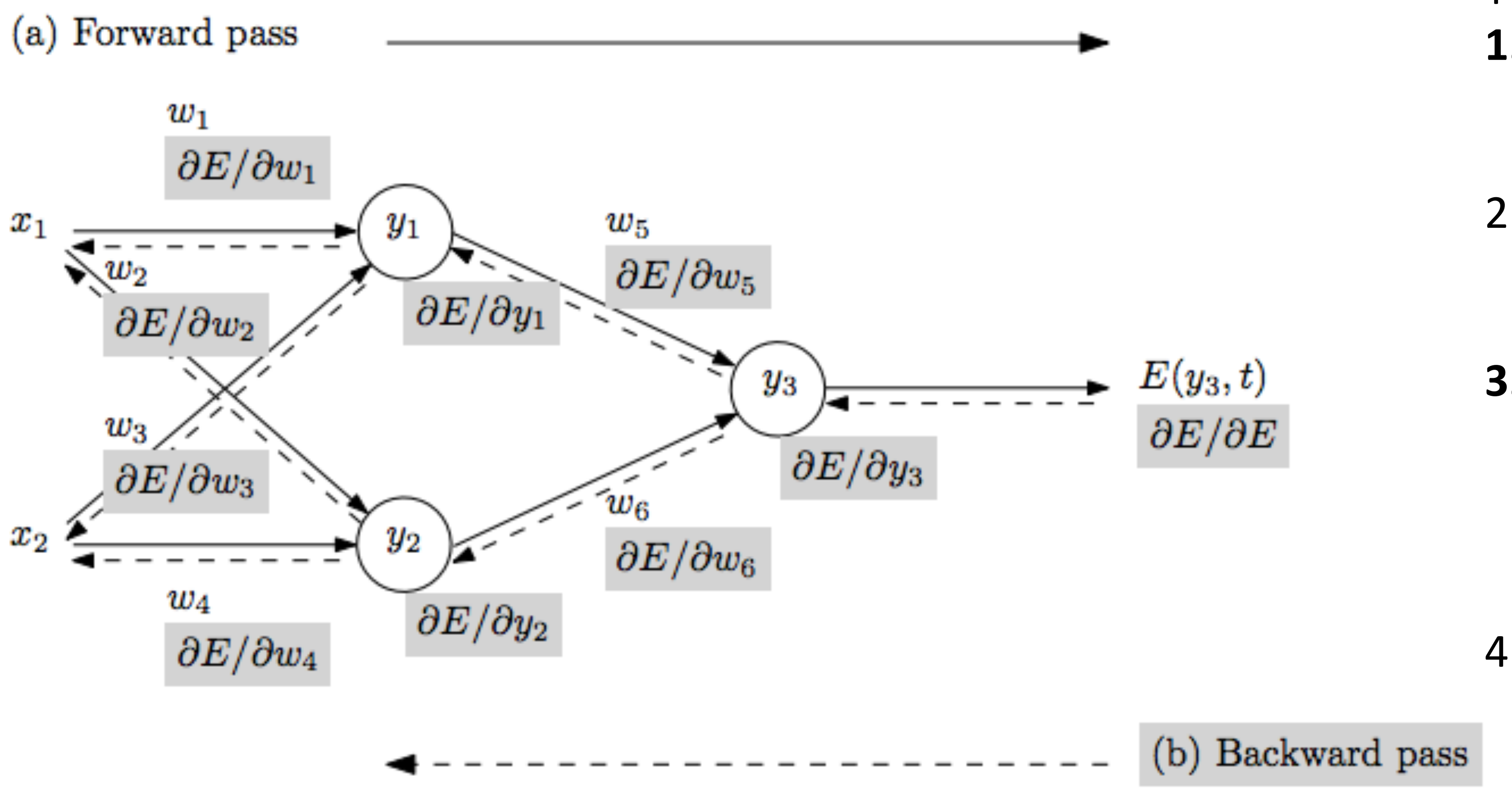
- Replace larger filter with stack of smaller filters; e.g., replace 7x7 with three 3x3s



- 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!

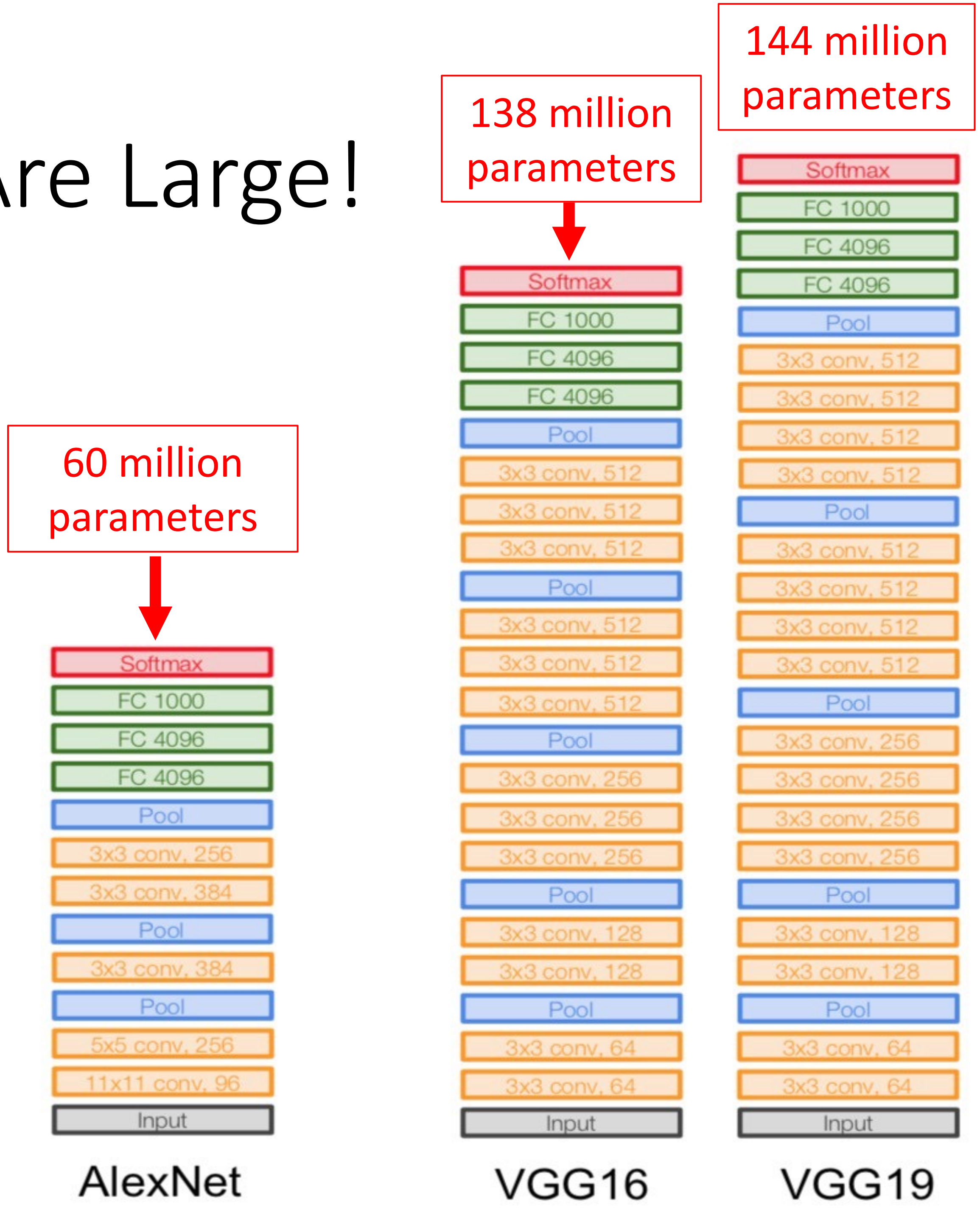
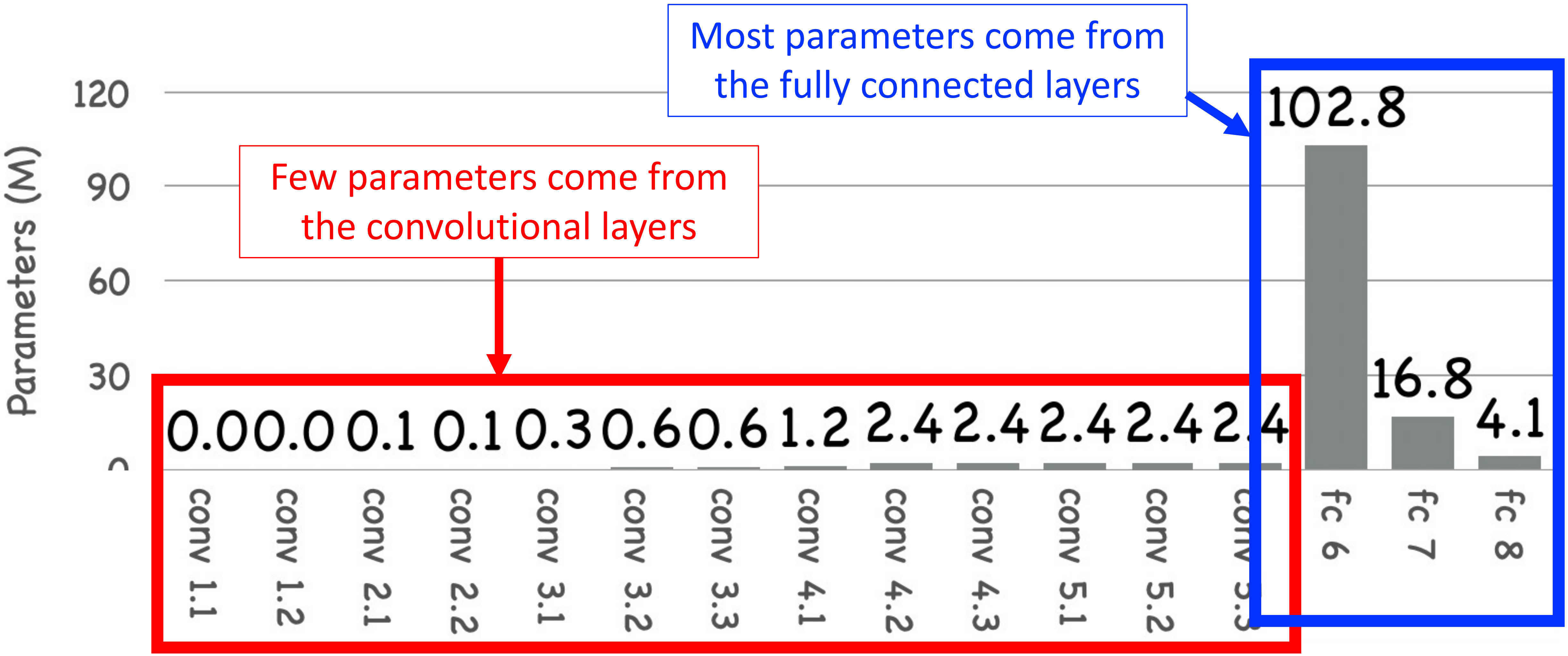


Figure Source (edited to fix mistakes): https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

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: to be covered next lecture
  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)

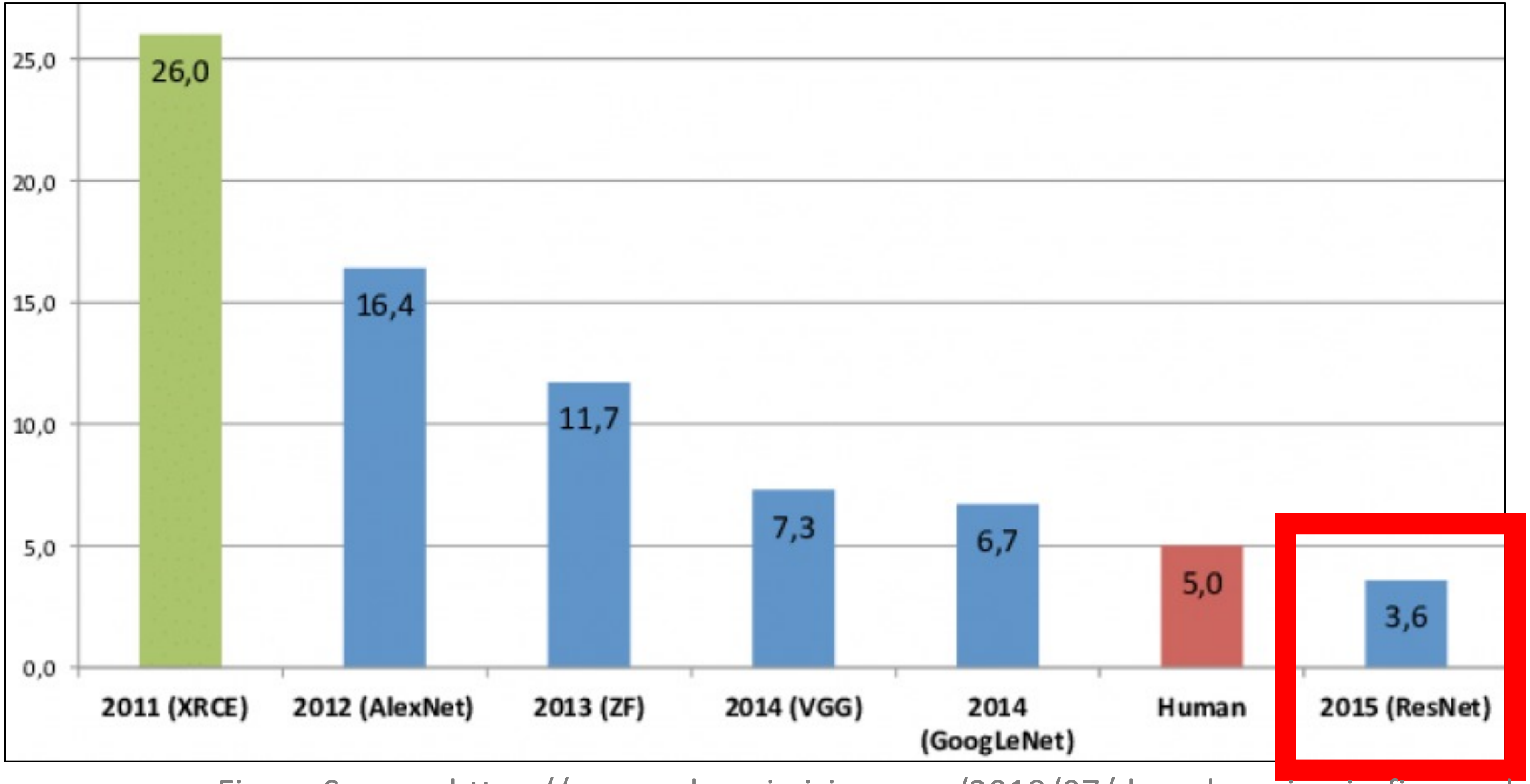
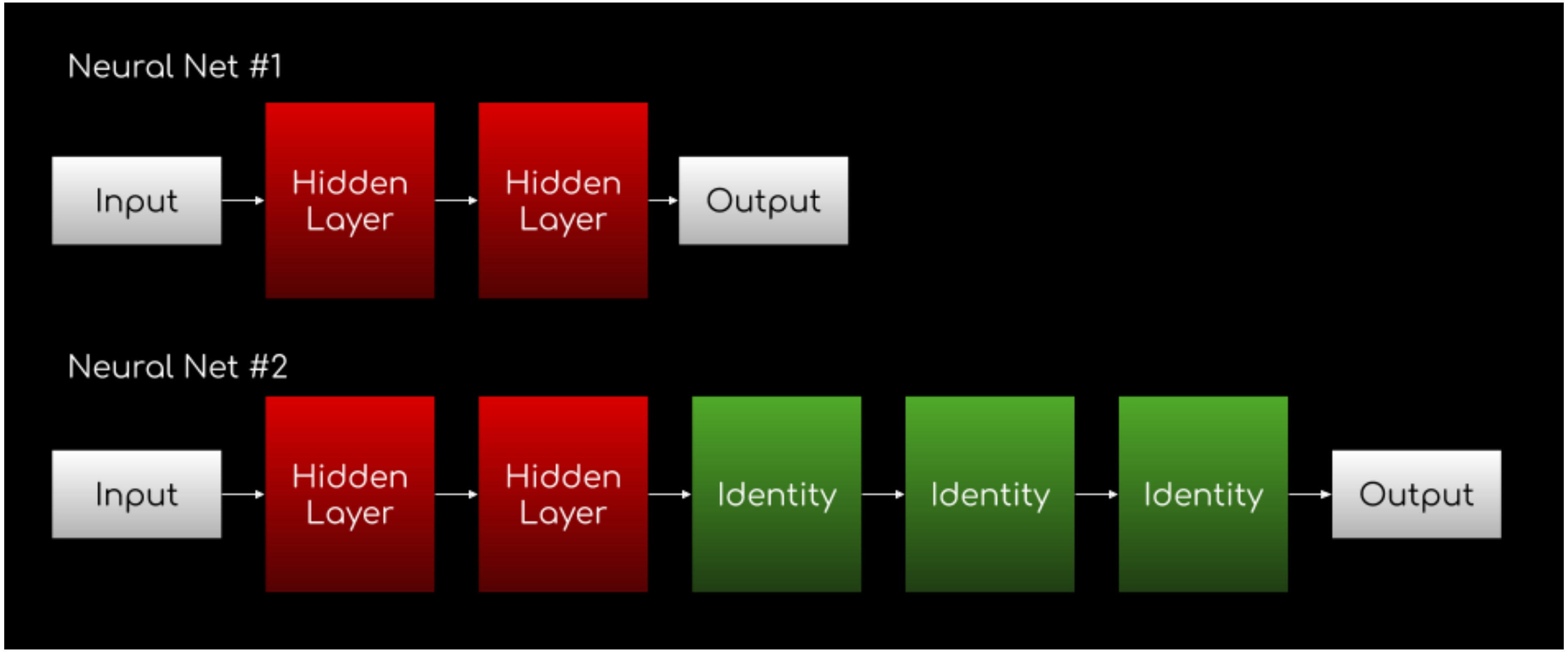


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

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?

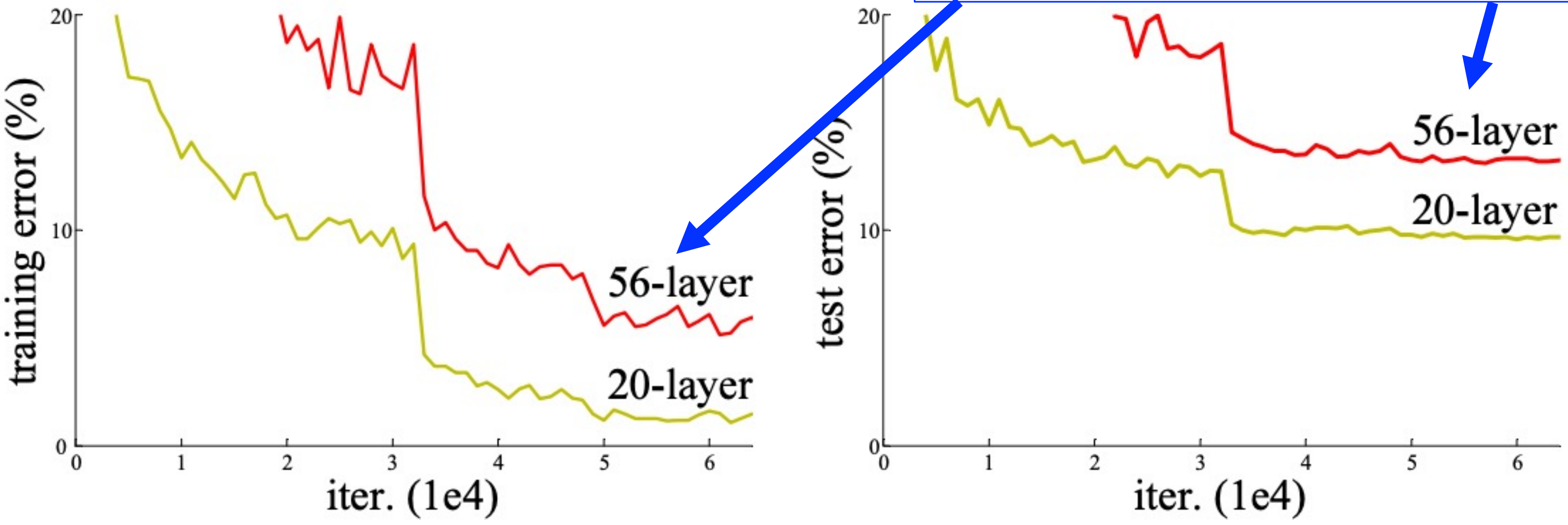


Source: https://medium.com/@realmichaelye/intuition-for-resnet-deep-residual-learning-for-image-recognition-39d24d173e78

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

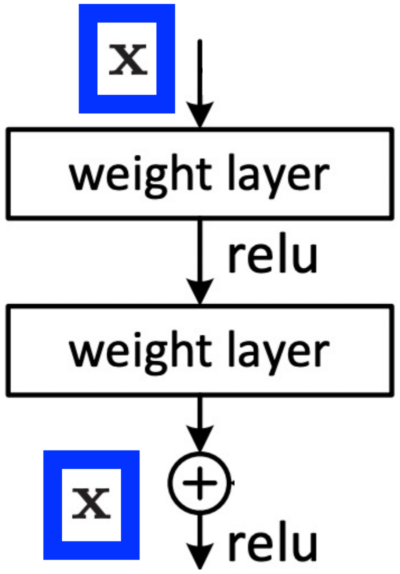
Idea: a deeper network should perform as good if not better than shallower networks since they can learn the shallower function by simply learning “identity” functions for later layers
Observation: adding more layers leads to WORSE results!
Is the problem overfitting? NO

Problem: It is difficult 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.

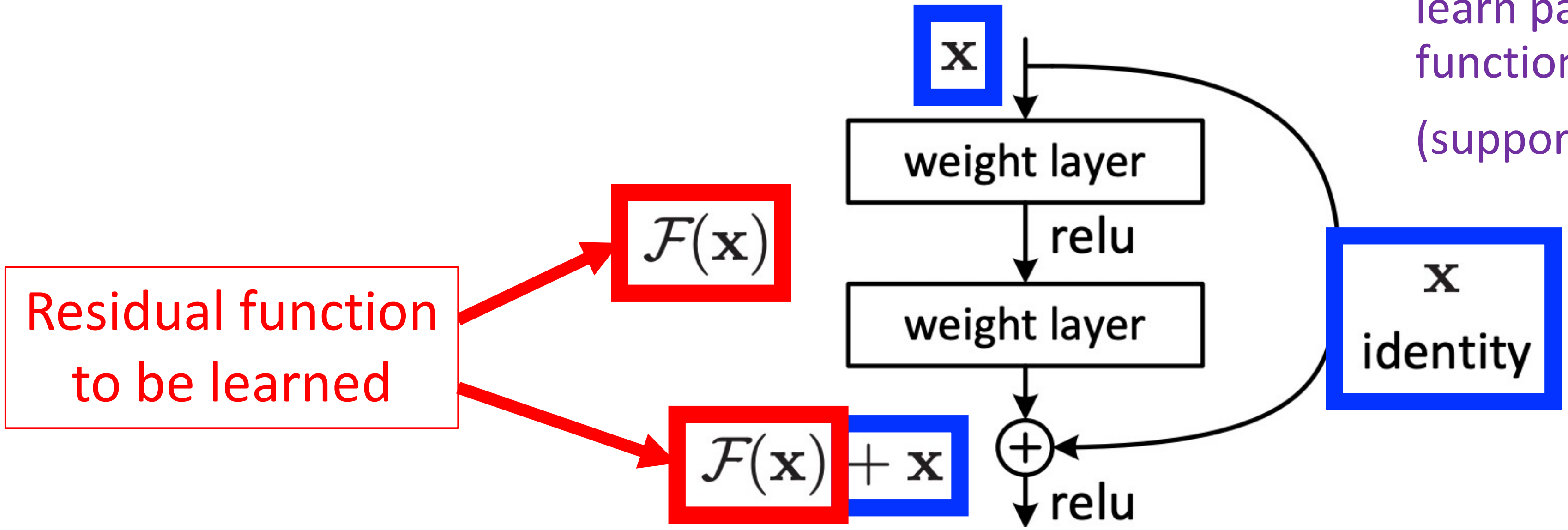
Problem: Difficult to Perform Identity Mapping

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



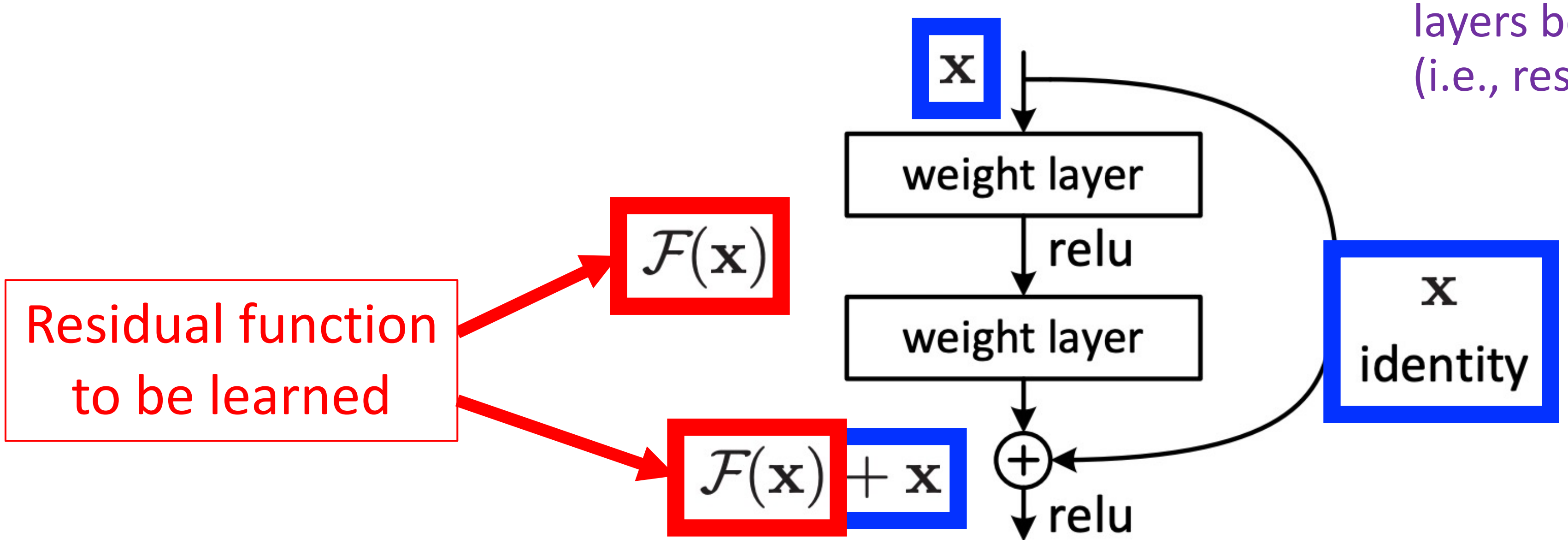
Assumption: Easier for a network to learn parameters close to 0 than a function that performs identity mapping

(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 (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

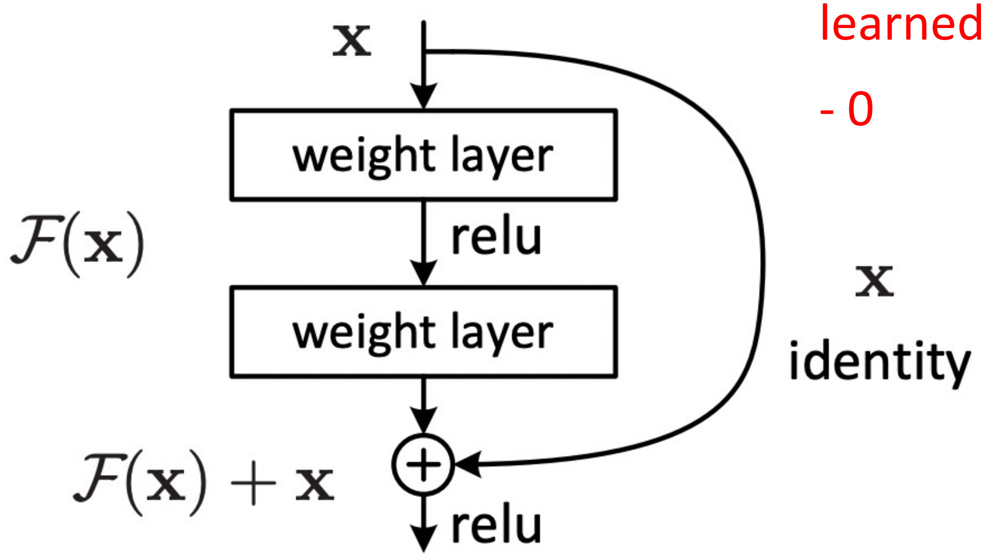
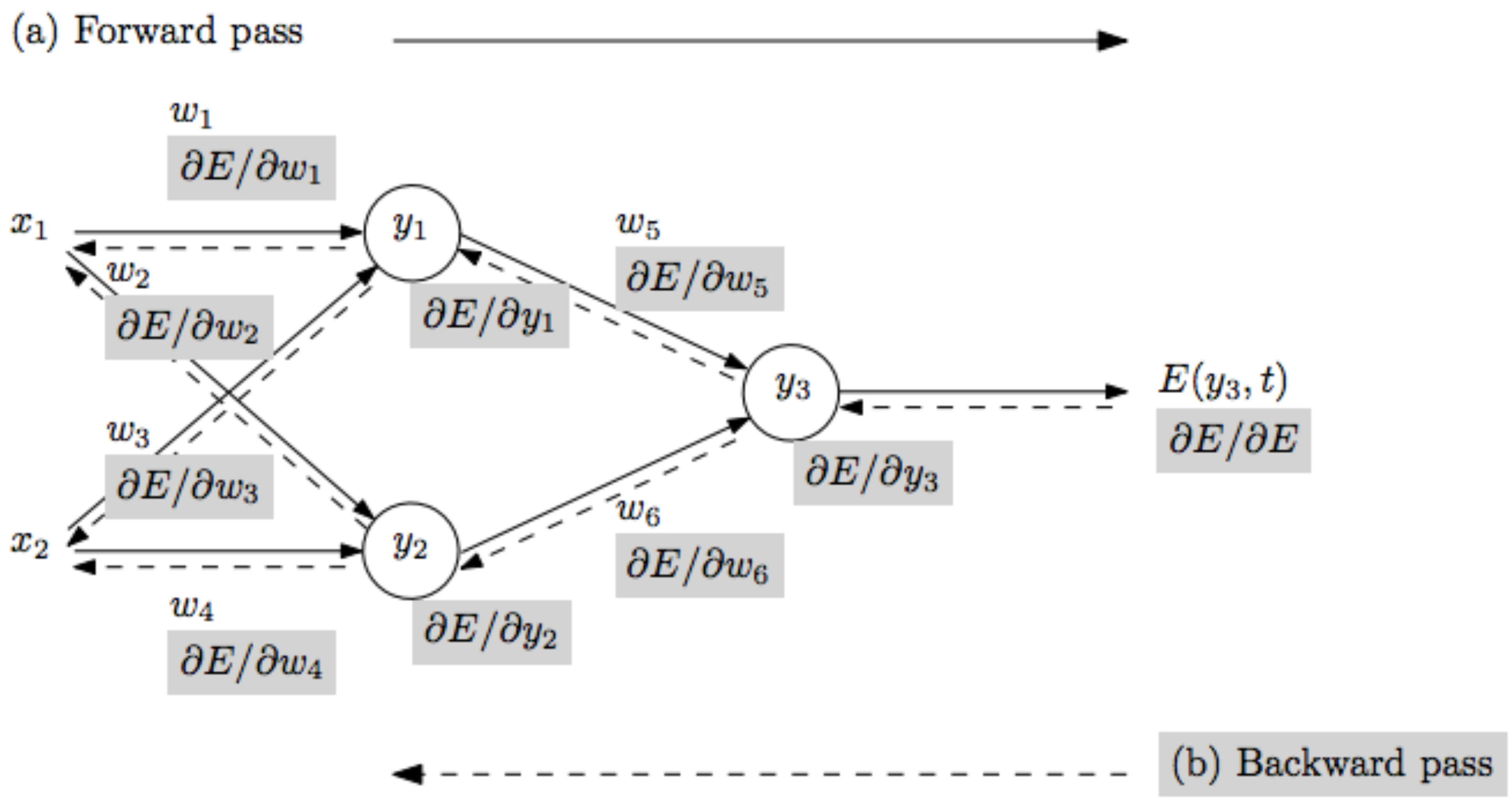


Figure 2. Residual learning: a building block.

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

ResNet Training (follows AlexNet)

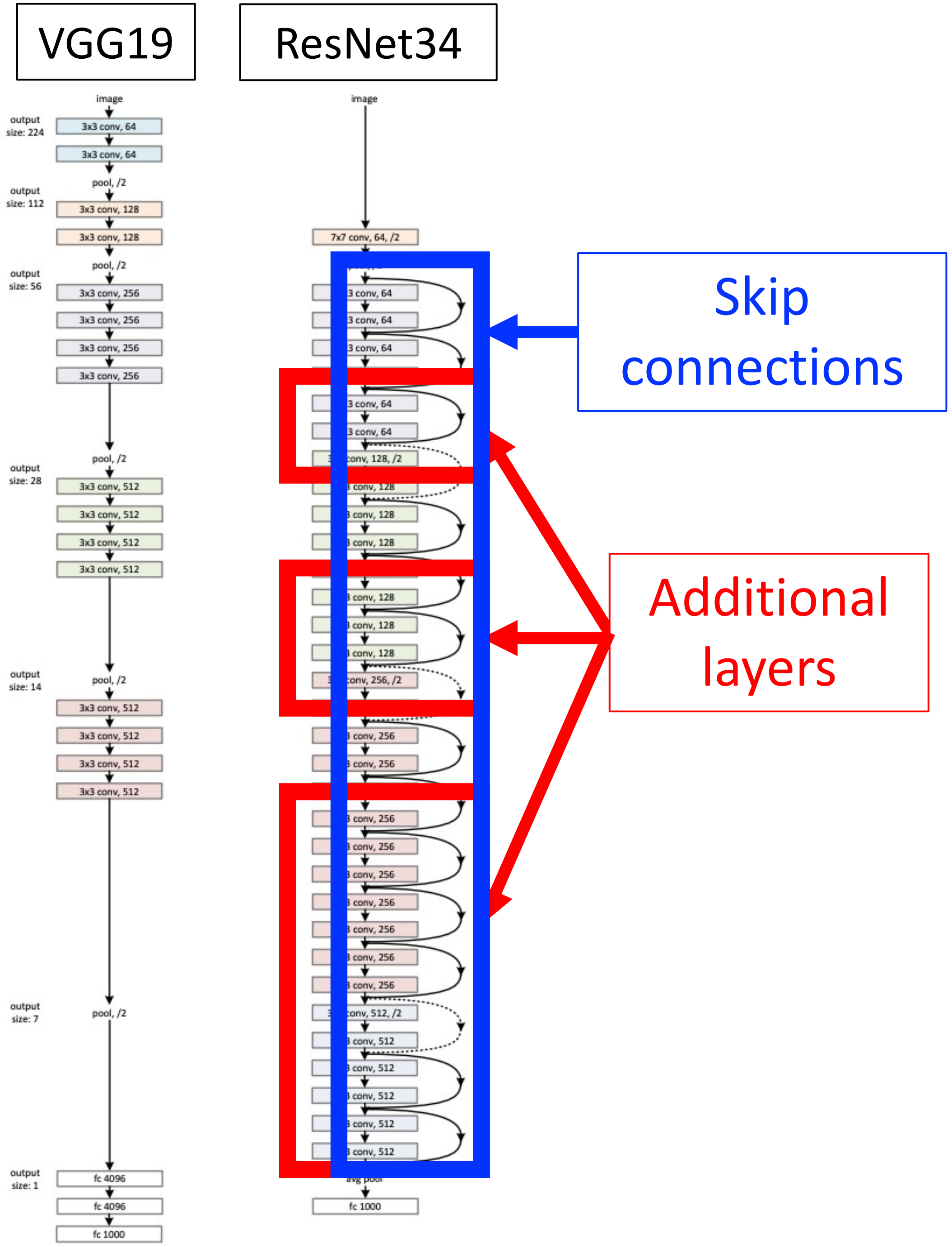


- Repeat until stopping criterion met:
    1. **Forward pass**: propagate training data through model to make prediction
    2. Quantify the dissatisfaction with a model's results on the training data
    3. **Backward pass**: using predicted output, calculate gradients backward to assign blame to each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

Progress of models on ImageNet (Top 5 Error)

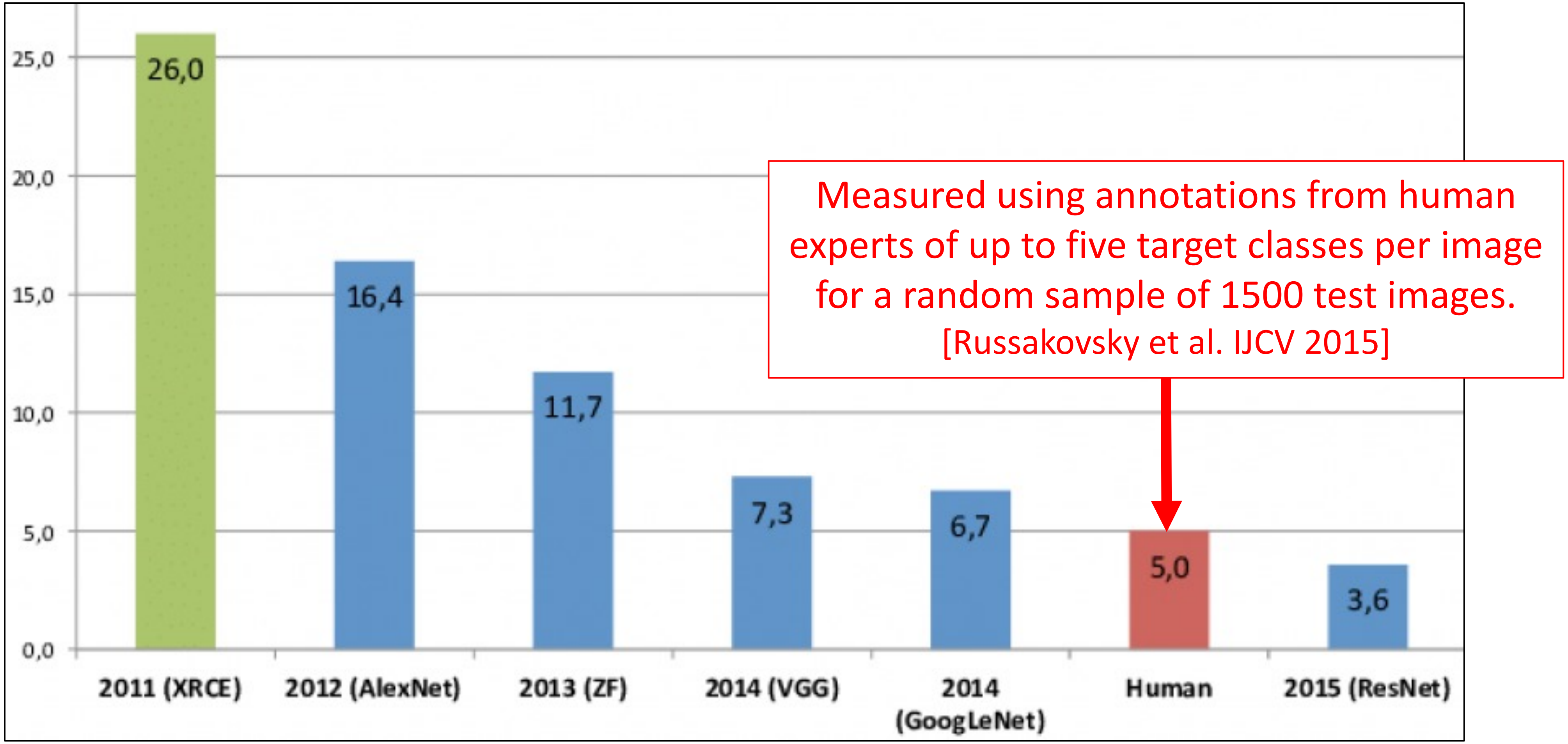


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

ResNet: Key Tricks for Going Deeper

- Skip connections with residual learning

“Deeper” Models Perform Better

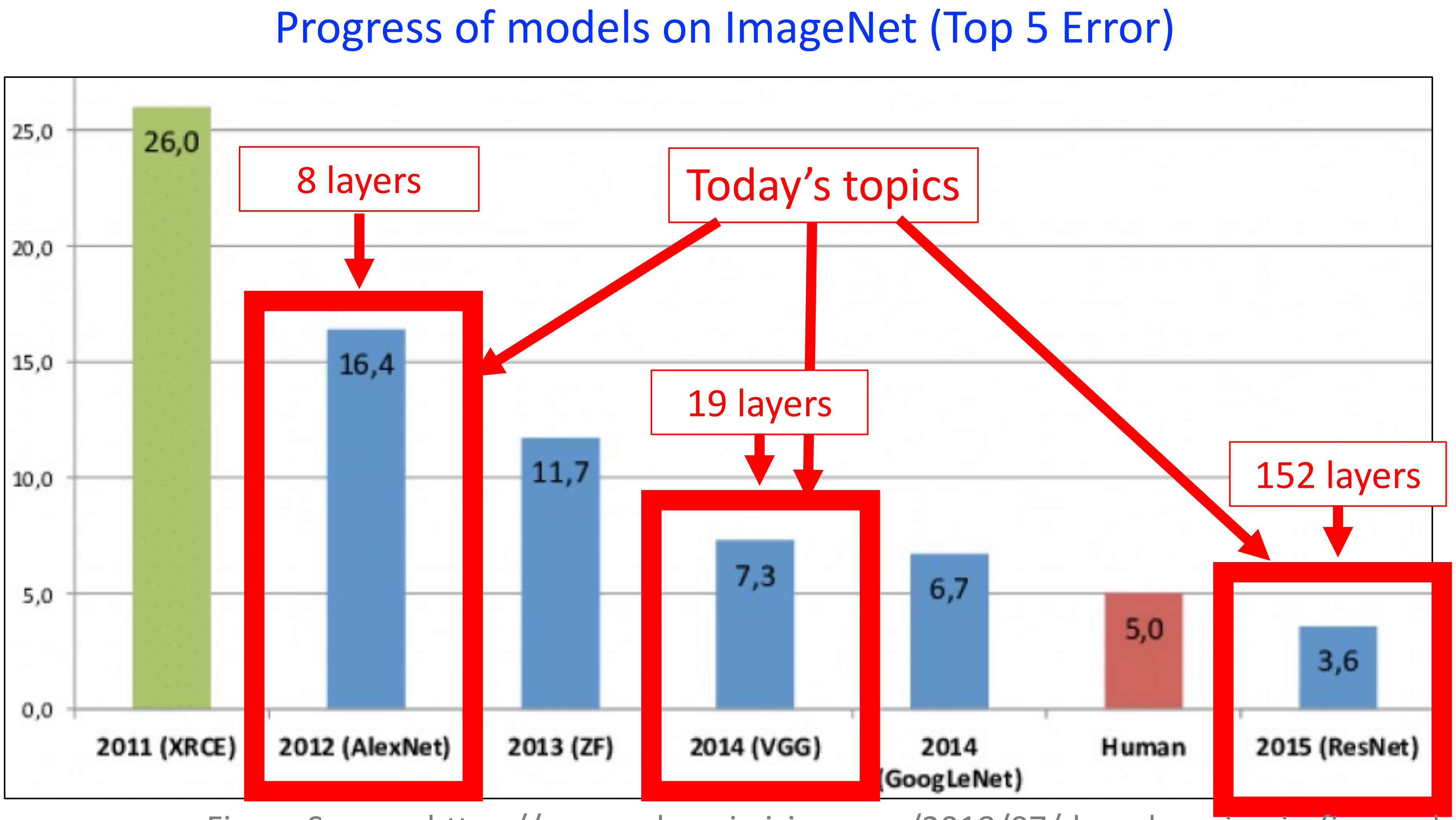


Figure Source: https://www.edge-ai-vision.com/2018/07/deep-learning-in-five-and-a-half-minutes/

ImageNet Impact Recognized in 2019

PAMI Longuet-Higgins Prize

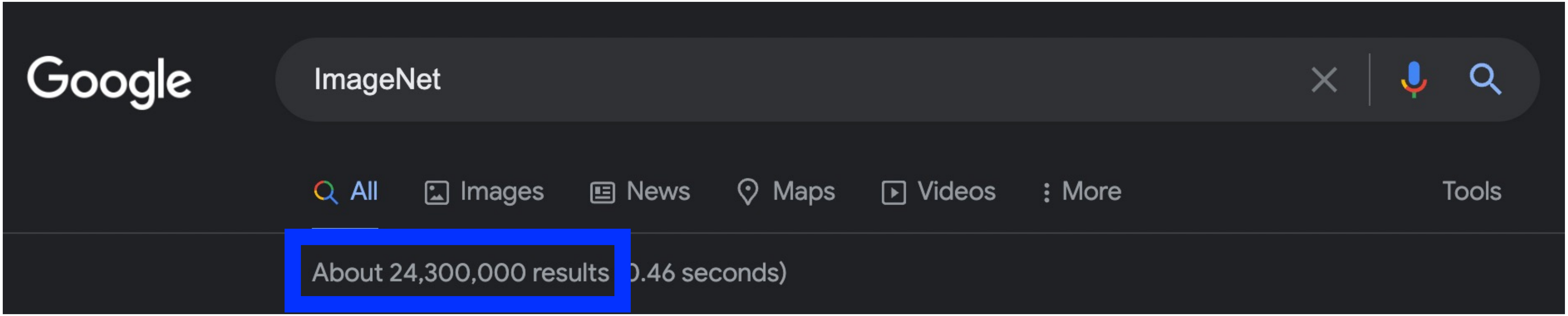
Retrospective Most Impactful Paper from CVPR 2009

ImageNet: A large-scale hierarchical image database

Jia Deng, Wei Dong, Richard Socher,
Li-Jia Li, Kai Li, and Li Fei-Fei

https://syncedreview.com/2019/06/18/cvpr-2019-attracts-9k-attendees-best-papers-announced-imagenet-honoured-10-years-later/

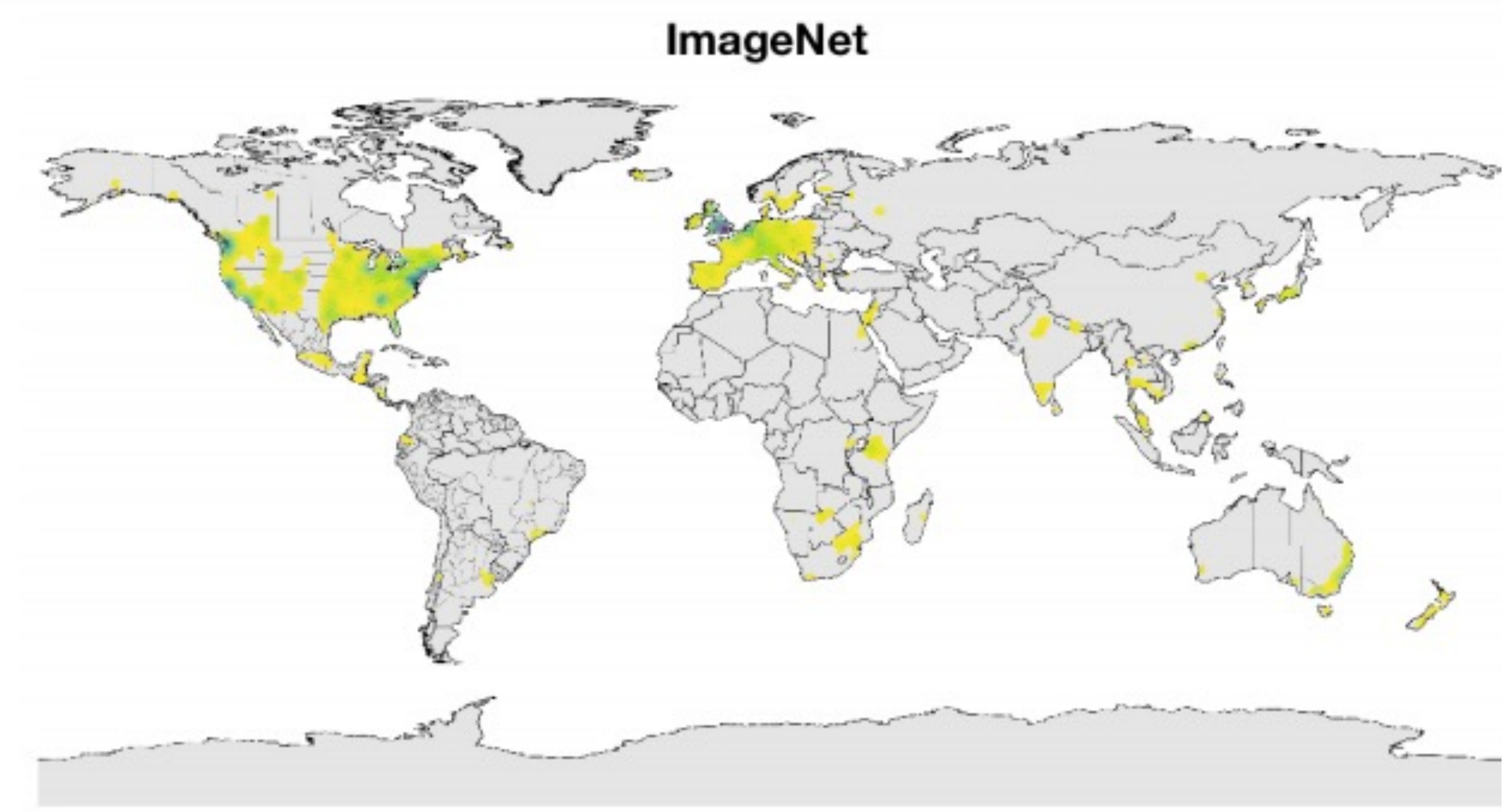
ImageNet Impact Recognized



Page count retrieved Sep 13, 2022

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.

Today’s Topics

- Computer vision
- Era of dataset challenges
- MNIST challenge winner: LeNet
- ImageNet challenge winners: deeper learning (AlexNet, VGG, ResNet)
- Programming tutorial

Today’s Topics

- Computer vision
- Era of dataset challenges
- MNIST challenge winner: LeNet
- ImageNet challenge winners: deeper learning (AlexNet, VGG, ResNet)
- Programming tutorial


The End