Lecture 5: Deep Neural Networks

Wei Qi Yan

Auckland University of Technology

March 29, 2023

Table of Contents

Awarded Work on Deep Learning (CVPR)

- X. Chen, K. He. Exploring Simple Siamese Representation Learning, 2021 (Best Paper Honorable Mention)
- T. Karras, S. Laine and T. Aila. A style-based generator architecture for generative adversarial networks, 2019 (Best Paper Honorable Mention)
- A. Zamir, et al. Taskonomy: Disentangling task transfer learning, 2018 (Best Paper)
- A. Zanfir and C. Sminchisescu. Deep learning of graph matching, 2018 (Best Paper Honorable Mention)
- G. Huang, et al. Densely connected convolutional networks, 2017 (Best Paper)
- A. Shrivastava, et al. Learning from simulated and unsupervised images through adversarial training, 2017 (Best Paper)
- J. Redmon, A. Farhadi. YOLO9000: Better, faster, stronger, 2017 (Best Paper Honorable Mention)
- L. Castrejon, K. Kundu, R. Urtasun, S. Fidler. Annotating object instances with a Polygon-RNN, 2017 (Best Paper Honorable Mention)
- K. He, et al. Deep residual learning for image recognition, 2016 (Best Paper)
- A. Jain, et al. Structural-RNN: Deep learning on spatio-temporal graphs, 2016 (Best Student Paper)
- J. Long, et al. Fully convolutional networks for semantic segmentation, 2015 (Best Paper Honorable Mention)

Awarded Work on Deep Learning

Awarded Work on Deep Learning (Marr Prize)

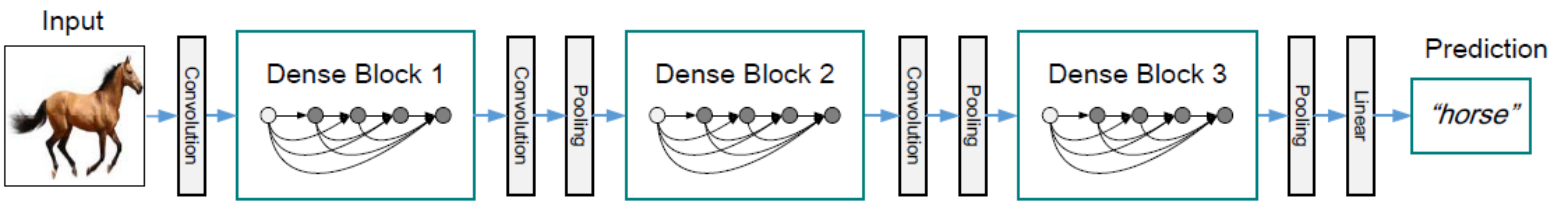
The Marr Prize is a biennial conference award in computer vision given by the ICCV. The prize is one of the top honors for a computer vision researcher.

- ...
- Z. Liu, Y. Lin, Y. Cao, H. Hu, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, 2021.
- T. Shaham, T. Michaeli, T. Dekel. SinGAN: Learning a generative model from a single natural image, 2019
- K. He, et al. Mask R-CNN, 2017
- P. Kontschieder, et al. Deep neural decision forests, 2015
- ...

Questions?

DenseNets

- CNNs can be substantially deeper, accurate, and efficient to train if they contain shorter connections between layers.
- To preserve the feedforward nature, DenseNets use direct connections from any layer to all subsequent layers.
- DenseNets alleviate the vanishing gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters.

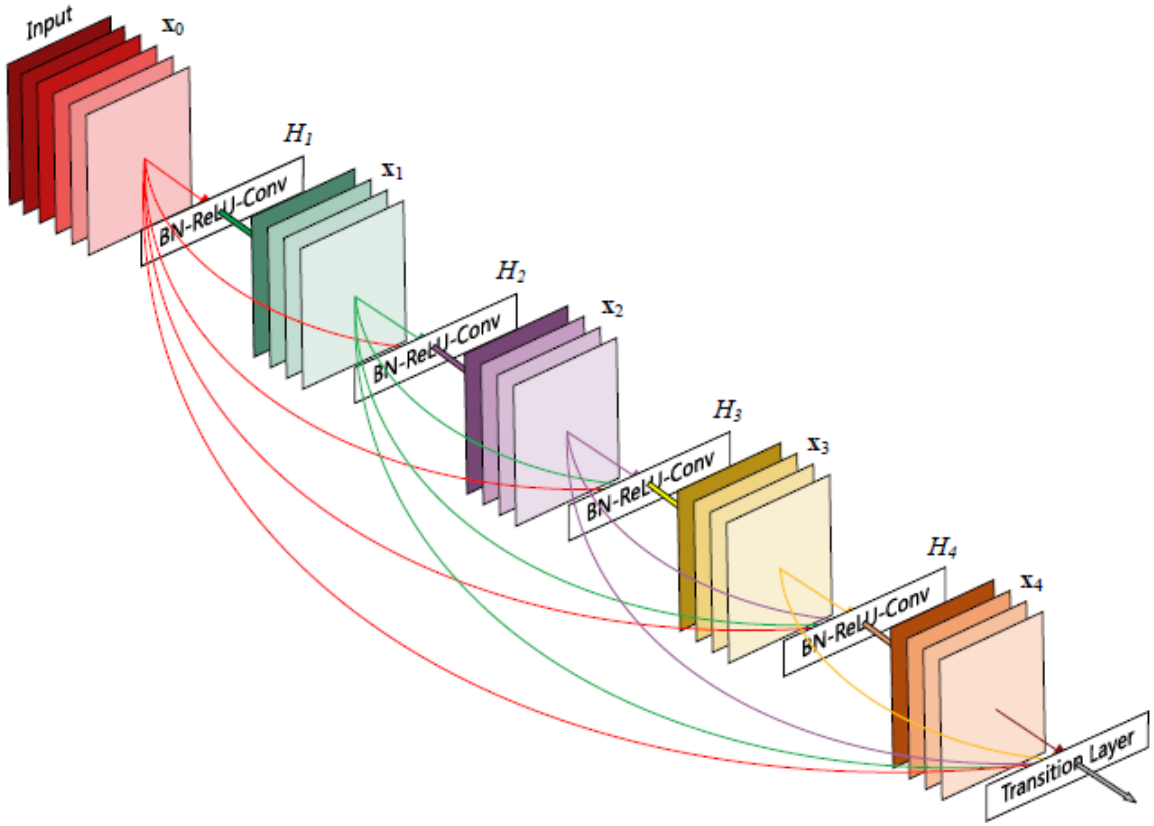


A deep DenseNet with three dense blocks.

G. Huang, et al. Densely Connected Convolutional Networks. IEEE CVPR'17.

DenseNets: Blocks

Each layer takes all preceding feature maps as input.



G. Huang, et al. Densely Connected Convolutional Networks. IEEE CVPR'17.

DenseNets: Architecture

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv; 3 × 3 conv] × 24	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 48	[1 × 1 conv; 3 × 3 conv] × 64
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv; 3 × 3 conv] × 16	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 48
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

DenseNet architectures for ImageNet.

G. Huang, et al. Densely Connected Convolutional Networks. IEEE CVPR'17.

DenseNets: Summery

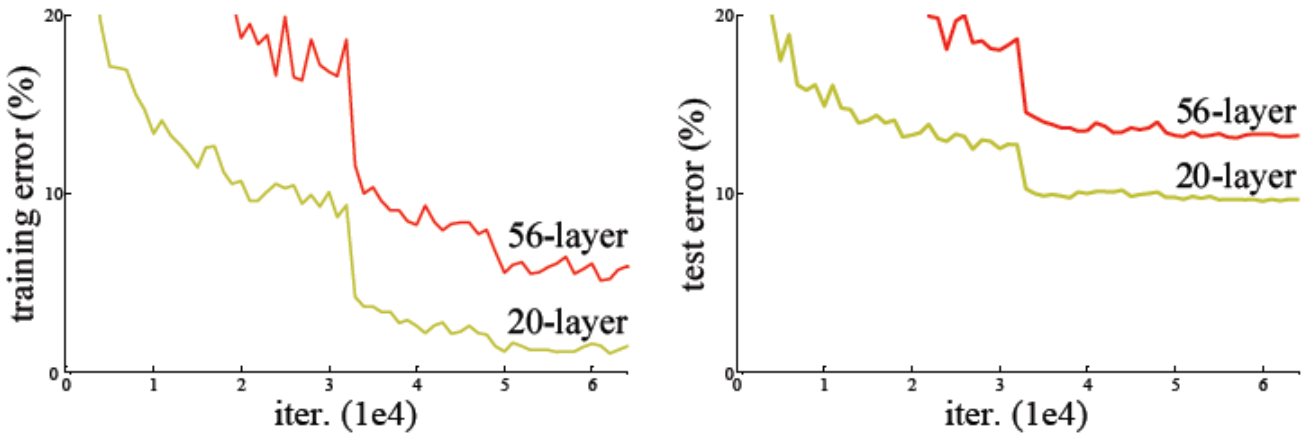
- DenseNet introduces direct connections between any two layers with the same feature-map size.
- DenseNets scale naturally to hundreds of layers, while exhibiting no optimization difficulties.
- DenseNets require substantially fewer parameters and less computation to achieve the state-of-the-art performances.
- Accuracy of DenseNets may be obtained by more detailed tuning of hyperparameters and learning rates.
- DenseNets allow feature reuse throughout the networks and can consequently learn more compact and more accurate models.

G. Huang, et al. Densely Connected Convolutional Networks. IEEE CVPR'17.

ResNets

The degradation problem: With the network depth increasing, accuracy gets *saturated*.

- ResNets are easy to be optimized.
- ResNets can easily enjoy accuracy gains from greatly increased depth.



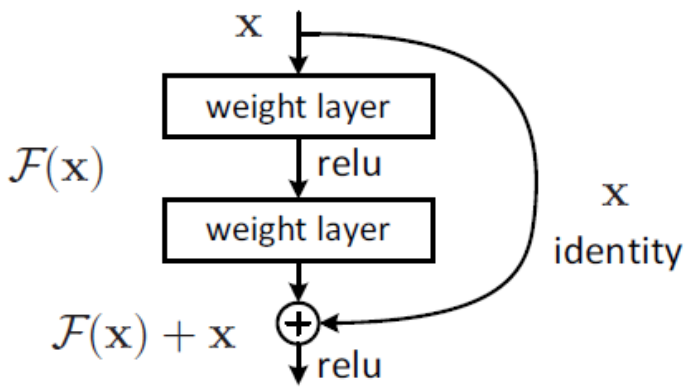
The deeper network has higher training error, and thus test error.

K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

ResNets

The degradation problem: With the network depth increasing, accuracy gets saturated.

- ResNets are easy to be optimized
- ResNets can easily enjoy accuracy gains from greatly increased depth.

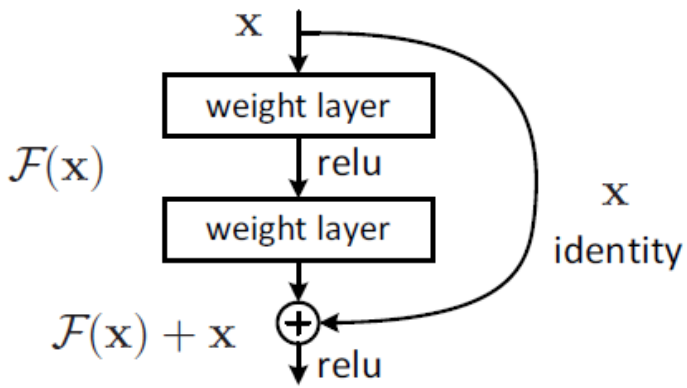


Residual learning: a building block.

K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

ResNets

- Insert shortcut connections to convert a plain network to ResNet.
- The identity shortcuts can be directly used when the input and output have the same dimensions.



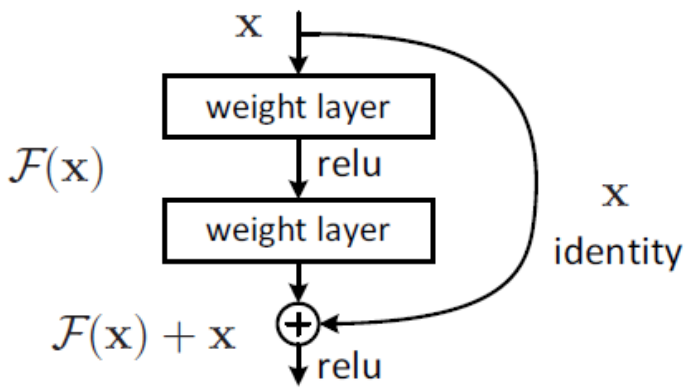
Residual learning: a building block.

K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

ResNets

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}$$

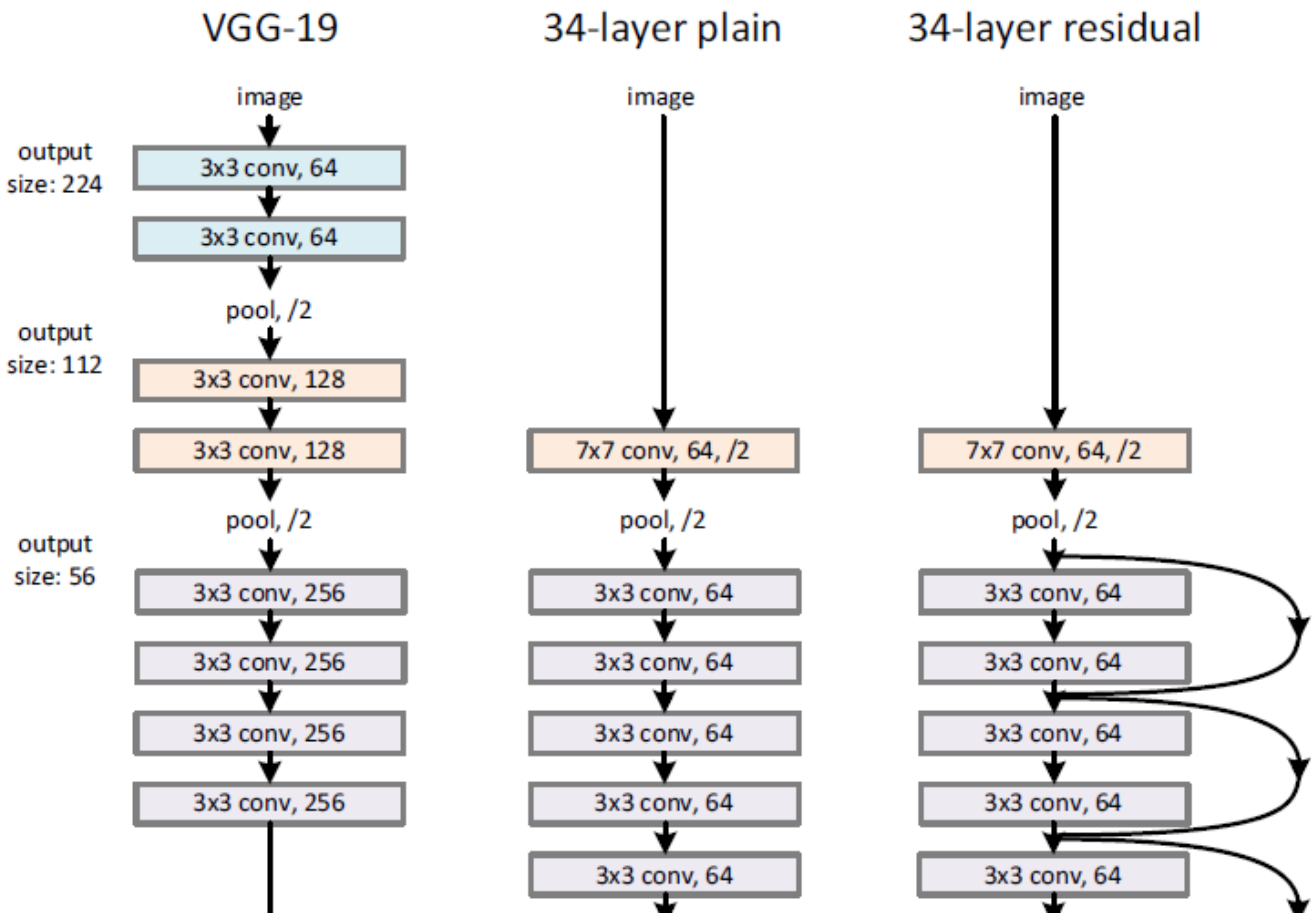
where $\mathbf{x}$ and $\mathbf{y}$ are the input and output vectors of the layers, $\mathcal{F}(\cdot)$ is the residual mapping function, e.g. $\mathcal{F} = W_{2\sigma}(W_{1x})$, $\sigma(\cdot)$ is the ReLU activation function.



Residual learning: a building block.

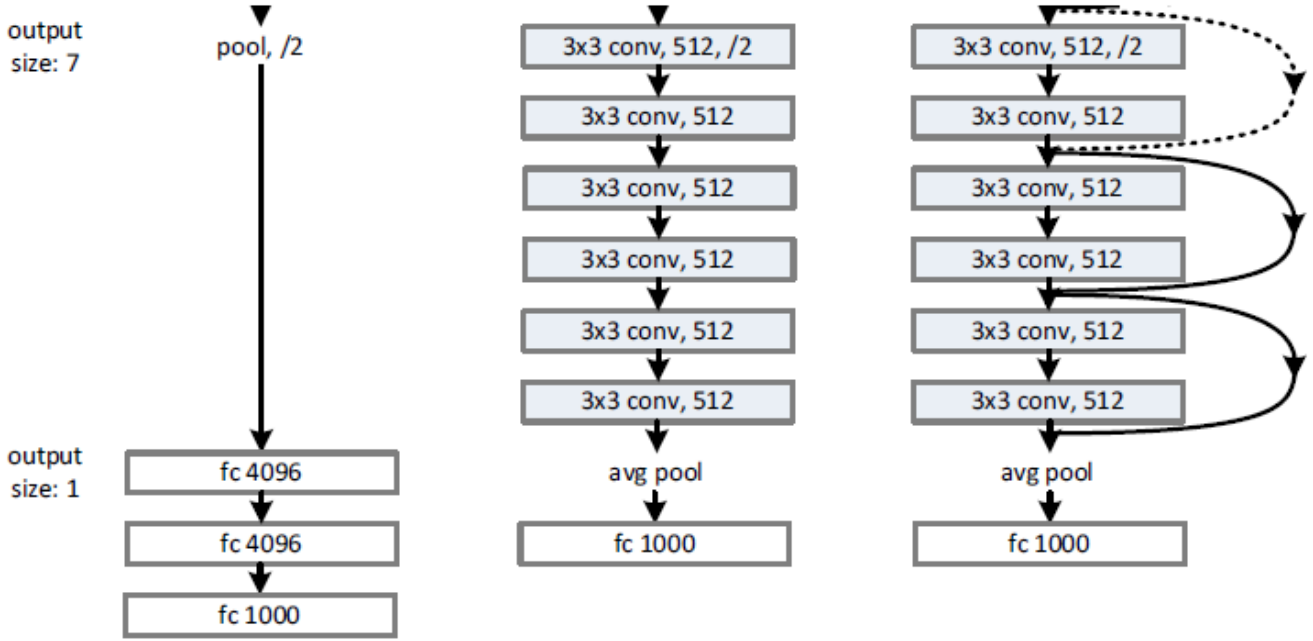
K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

ResNets



K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

ResNets



K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

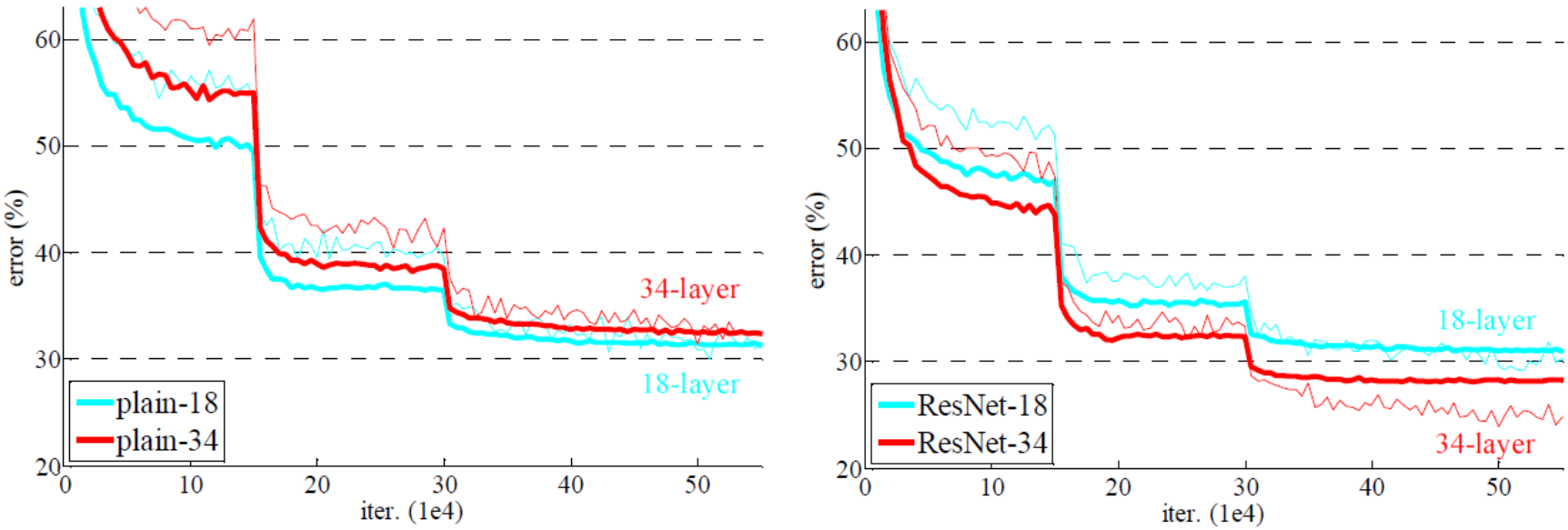
ResNets: Architecture

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Architectures for ImageNet.

K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

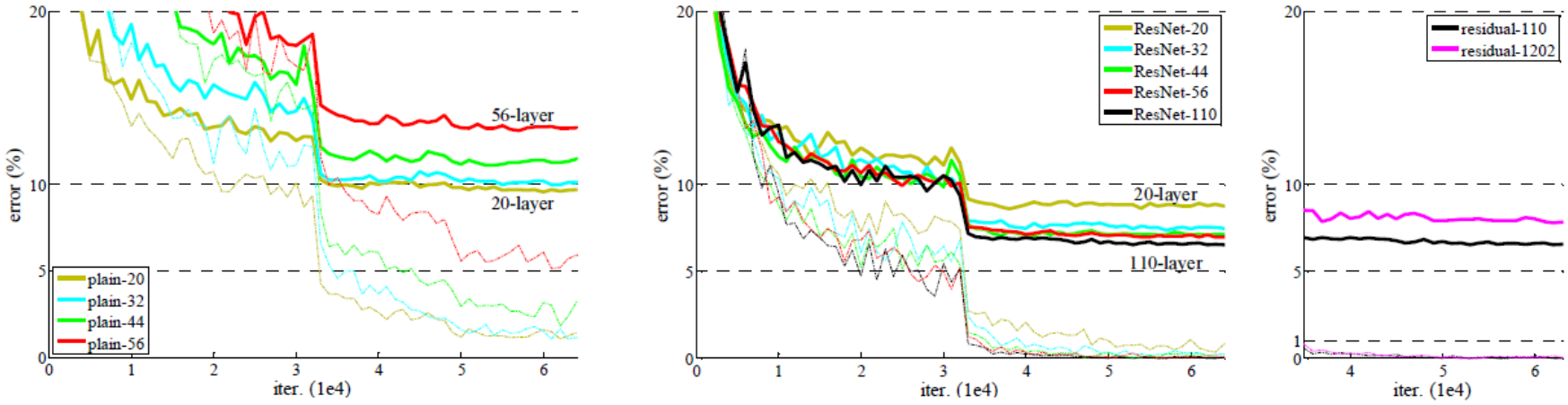
ResNets: Training



Training on **ImageNet**.

K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

ResNets: Results



Training on **CIFAR-10**. Dashed lines denote training error, and bold lines denote testing error.

K. He, et al. Deep Residual Learning for Image Recognition, CVPR'16.

Questions?

MATLAB AlexNet

- AlexNet is a convolutional neural network that is trained based on more than a million images from the ImageNet.
- AlexNet is eight layers deep and can classify images into 1,000 object classes, such as keyboard, mouse, pencil, and many animals.
- AlexNet has learned pretty rich features from a wide range of images.
- AlexNet has the image input size 227×227.

Web: https://au.mathworks.com/help/deeplearning/ref/alexnet.html

MATLAB AlexNet

AlexNet classifies an image following the steps:

- Load a pretrained MATLAB AlexNet model;
- Read a test image;
- Crop or resize the image to the input size of the deep net;
- Classify the image using the trained classifier;
- Show the image and classification result together.

Web: https://au.mathworks.com/help/deeplearning/ref/alexnet.html

MATLAB AlexNet Result

MATLAB VGG-19

- VGG-19 is a convolutional neural network that is trained based on more than a million images from the ImageNet.
- VGG-19 is 19 layers deep and can classify images into 1,000 object classes, such as keyboard, mouse, pencil, and many animals.
- VGG-19 has learned features from a wide range of images.
- VGG-19 has the image input size 224×224.

Web:https://au.mathworks.com/help/deeplearning/ref/vgg19.html

MATLAB GoogLeNet

- GoogLeNet is a pretrained convolutional neural network that is 22 layers deep.
- GoogLeNet is trained based on either the ImageNet or Places365 datasets.
- GoogLeNet is trained based on ImageNet and classifies images into 1,000 object classes.
- GoogLeNet classifies images into 365 different classes, such as field, park, runway, and lobby, etc.
- GoogLeNet has learned from different features of a wide range of images.
- GoogLeNet has the image input size 224×224.

Web:https://au.mathworks.com/help/deeplearning/ref/googlenet.html

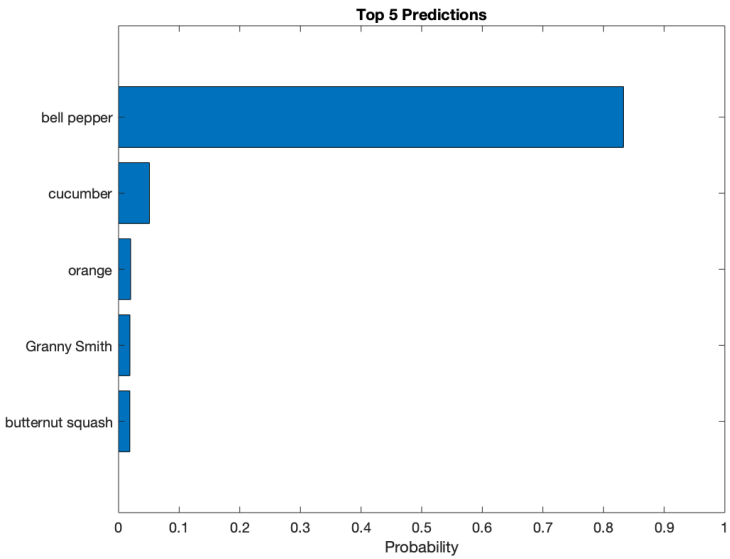
MATLAB GooLeNet Result

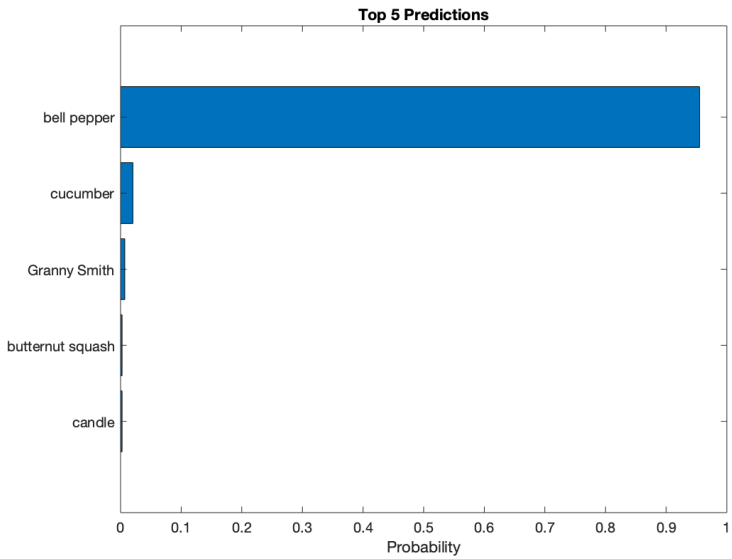
MATLAB Inception-v3

- Inception-v3 is a CNN network that is trained based on more than a million images from the ImageNet.
- The network has 48 layers and can classify images into 1,000 classes.
- The network has the image input size 299×299.

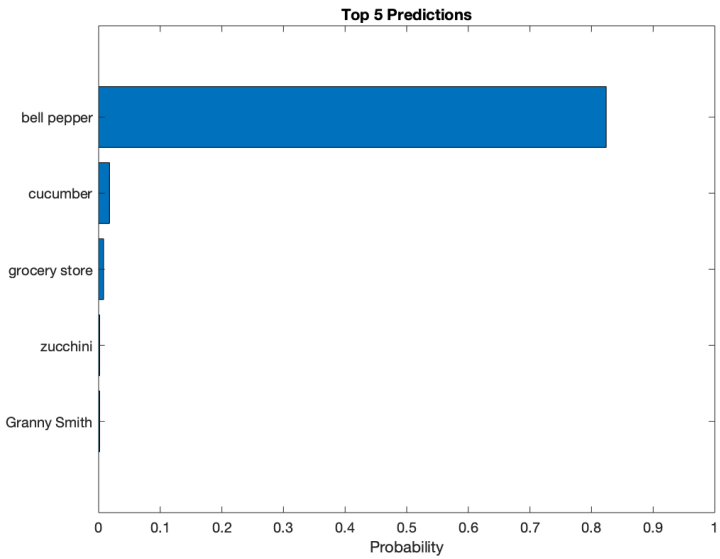
Web:https://au.mathworks.com/help/deeplearning/ref/inceptionv3.html

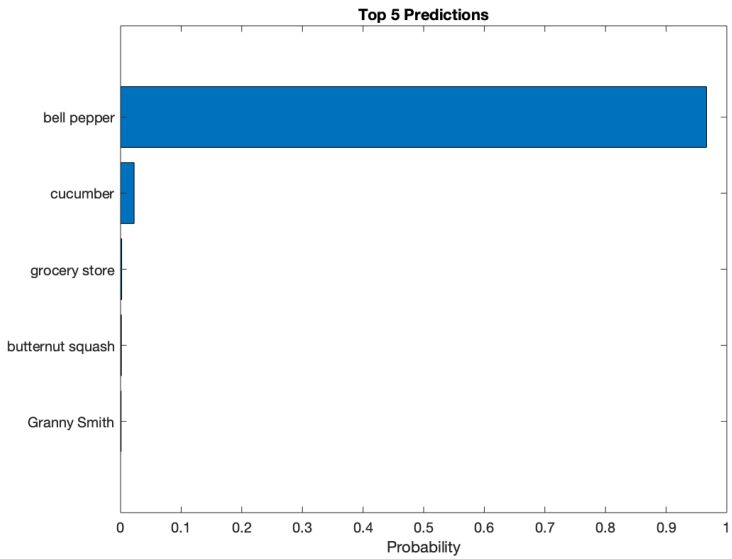
Comparisons of MATLAB AlexNet and GoogLeNet





Comparisons of MATLAB Inception-v3 and VGG-19





Questions?

Deep Neural Networks

Learning Objectives

- Design and analyse algorithms of deep neural networks.
- Demonstrate advanced understanding of the state-of-the-art in the practice of deep learning.