

Fuzzy Set Theory in Computer Vision: Example 3

Derek T. Anderson and James M. Keller

FUZZ-IEEE, July 2017



Overview

- ▶ Purpose of these slides are to make you aware of a few of the different CNN architectures out there



Overview

- ▶ Purpose of these slides are to make you aware of a few of the different CNN architectures out there
- ▶ NOT comprehensive

Overview

- ▶ Purpose of these slides are to make you aware of a few of the different CNN architectures out there
- ▶ NOT comprehensive
- ▶ Many folks download these networks and use them



Overview

- ▶ Purpose of these slides are to make you aware of a few of the different CNN architectures out there
- ▶ NOT comprehensive
- ▶ Many folks download these networks and use them
- ▶ MatConvNet (and other libraries) support such functionality

GoogleNet

- ▶ Going deeper with convolutions, published in 2014

GoogleNet

- ▶ Going deeper with convolutions, published in 2014
- ▶ Input data: ILSVRC 2014 challenge data set, it contains 1,000 categories. 1.2 million images for training, 50,000 for validation and 100,000 for testing



GoogleNet

- ▶ Going deeper with convolutions, published in 2014
- ▶ Input data: ILSVRC 2014 challenge data set, it contains 1,000 categories. 1.2 million images for training, 50,000 for validation and 100,000 for testing
- ▶ Input image is $224 \times 224 \times 3$



GoogleNet

- ▶ Going deeper with convolutions, published in 2014
- ▶ Input data: ILSVRC 2014 challenge data set, it contains 1,000 categories. 1.2 million images for training, 50,000 for validation and 100,000 for testing
- ▶ Input image is $224 \times 224 \times 3$
- ▶ Zero mean

GoogleNet: inception module

- ▶ The goal of the inception module is to act as a “multi-level feature extractor” by computing 11, 33, and 55 convolutions within the same module of the network—the output of these filters are then stacked along the channel dimension and before being fed into the next layer in the network

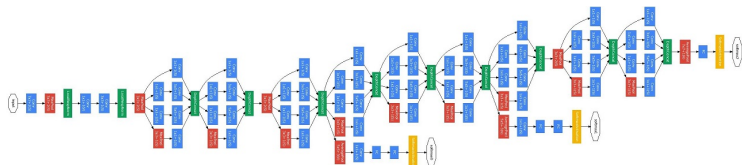
GoogleNet: inception module

- ▶ The goal of the inception module is to act as a “multi-level feature extractor” by computing 11, 33, and 55 convolutions within the same module of the network—the output of these filters are then stacked along the channel dimension and before being fed into the next layer in the network
- ▶ The original incarnation of this architecture was called GoogLeNet, but subsequent manifestations have simply been called Inception vN where N refers to the version number put out by Google



GoogleNet: inception module

- ▶ The goal of the inception module is to act as a “multi-level feature extractor” by computing 11, 33, and 55 convolutions within the same module of the network—the output of these filters are then stacked along the channel dimension and before being fed into the next layer in the network
- ▶ The original incarnation of this architecture was called GoogLeNet, but subsequent manifestations have simply been called Inception vN where N refers to the version number put out by Google



GoogleNet

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

AlexNet

- ▶ ImageNet Classification with Deep Convolutional Neural Networks, published on 2012

AlexNet

- ▶ ImageNet Classification with Deep Convolutional Neural Networks, published on 2012
- ▶ ILSVRC-2010 uses subset of ImageNet data set with roughly 1,000 RGB images in each of 1,000 categories. Roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images. Test set labels are available

AlexNet

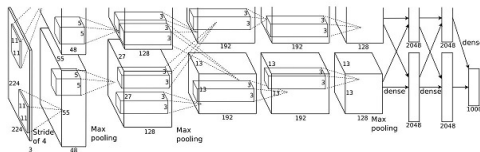
- ▶ ImageNet Classification with Deep Convolutional Neural Networks, published on 2012
- ▶ ILSVRC-2010 uses subset of ImageNet data set with roughly 1,000 RGB images in each of 1,000 categories. Roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images. Test set labels are available
- ▶ The images were down sampled to 256x256 pixels

AlexNet

- ▶ ImageNet Classification with Deep Convolutional Neural Networks, published on 2012
- ▶ ILSVRC-2010 uses subset of ImageNet data set with roughly 1,000 RGB images in each of 1,000 categories. Roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images. Test set labels are available
- ▶ The images were down sampled to 256x256 pixels
- ▶ Zero mean

AlexNet

- ▶ ImageNet Classification with Deep Convolutional Neural Networks, published on 2012
- ▶ ILSVRC-2010 uses subset of ImageNet data set with roughly 1,000 RGB images in each of 1,000 categories. Roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images. Test set labels are available
- ▶ The images were down sampled to 256x256 pixels
- ▶ Zero mean



VGG network

- ▶ VGG network architecture was introduced by Simonyan and Zisserman in their 2014 paper, Very Deep Convolutional Networks for Large Scale Image Recognition

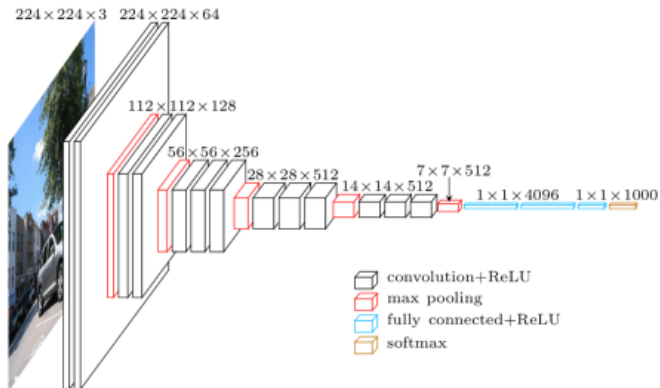
VGG network

- ▶ VGG network architecture was introduced by Simonyan and Zisserman in their 2014 paper, Very Deep Convolutional Networks for Large Scale Image Recognition
- ▶ Network is characterized by its simplicity, using only 33 convolutional layers stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier

VGG network

- ▶ VGG network architecture was introduced by Simonyan and Zisserman in their 2014 paper, Very Deep Convolutional Networks for Large Scale Image Recognition
- ▶ Network is characterized by its simplicity, using only 33 convolutional layers stacked on top of each other in increasing depth. Reducing volume size is handled by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier
- ▶ The “16” and “19” stand for the number of weight layers in the network

VGG network



VGG network

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

ResNet

- ▶ He et al. in their 2015 paper, Deep Residual Learning for Image Recognition

ResNet

- ▶ He et al. in their 2015 paper, Deep Residual Learning for Image Recognition
- ▶ ResNet architecture has become a seminal work, demonstrating that extremely deep networks can be trained using standard SGD (and a reasonable initialization function) through the use of residual modules

ResNet

- ▶ He et al. in their 2015 paper, Deep Residual Learning for Image Recognition
- ▶ ResNet architecture has become a seminal work, demonstrating that extremely deep networks can be trained using standard SGD (and a reasonable initialization function) through the use of residual modules
- ▶ Unlike traditional sequential network architectures (AlexNet, OverFeat and VGG), ResNet relies on micro-architecture modules (“network-in-network architectures”)

ResNet

- ▶ He et al. in their 2015 paper, Deep Residual Learning for Image Recognition
- ▶ ResNet architecture has become a seminal work, demonstrating that extremely deep networks can be trained using standard SGD (and a reasonable initialization function) through the use of residual modules
- ▶ Unlike traditional sequential network architectures (AlexNet, OverFeat and VGG), ResNet relies on micro-architecture modules (“network-in-network architectures”)
- ▶ Micro-architecture refers to the set of “building blocks” used to construct the network

ResNet

- ▶ He et al. in their 2015 paper, Deep Residual Learning for Image Recognition
- ▶ ResNet architecture has become a seminal work, demonstrating that extremely deep networks can be trained using standard SGD (and a reasonable initialization function) through the use of residual modules
- ▶ Unlike traditional sequential network architectures (AlexNet, OverFeat and VGG), ResNet relies on micro-architecture modules (“network-in-network architectures”)
- ▶ Micro-architecture refers to the set of “building blocks” used to construct the network
- ▶ A collection of micro-architecture building blocks (along with your standard CONV, POOL, etc. layers) leads to the macro-architecture (i.e., the end network itself)

ResNet

- ▶ Further accuracy can be obtained by updating the residual module to use identity mappings, as demonstrated in their 2016 followup publication, Identity Mappings in Deep Residual Networks

ResNet

- ▶ Further accuracy can be obtained by updating the residual module to use identity mappings, as demonstrated in their 2016 followup publication, Identity Mappings in Deep Residual Networks
- ▶ Even though ResNet is much deeper than VGG16 and VGG19, the model size is actually substantially smaller due to the usage of global average pooling rather than fully-connected layers this reduces the model size down to 102MB for ResNet50

Summary ...

- ▶ Field is changing daily almost

Summary ...

- ▶ Field is changing daily almost
- ▶ NOT a solved problem

Summary ...

- ▶ Field is changing daily almost
- ▶ NOT a solved problem
- ▶ DL has a “high cost for entry”

Summary ...

- ▶ Field is changing daily almost
- ▶ NOT a solved problem
- ▶ DL has a “high cost for entry”
- ▶ Start reading