法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

- 微信订阅号：小象
- 新浪微博：ChinaHadoop

卷积神经网络—高级篇

主讲人：李伟

纽约城市大学博士

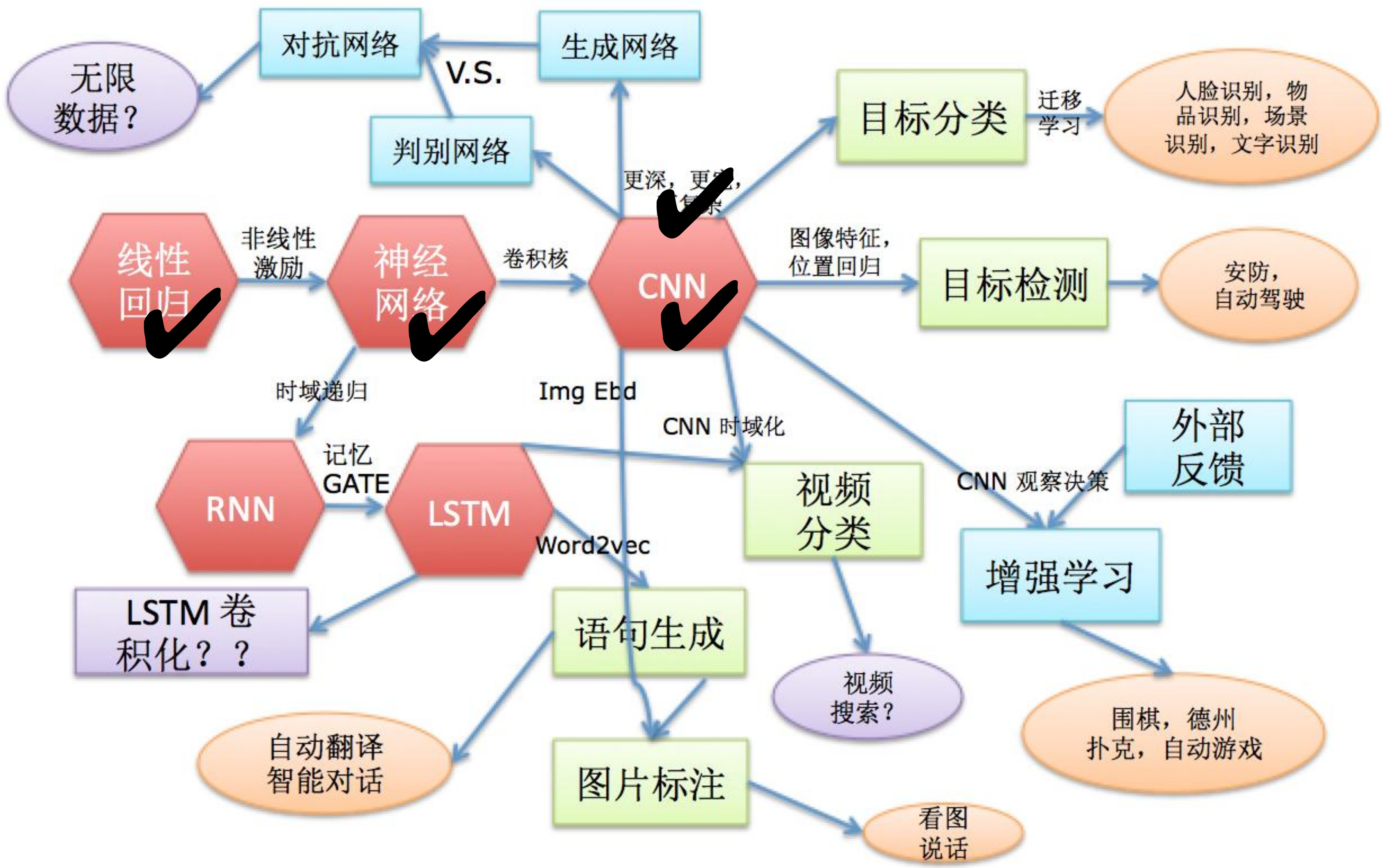
主要研究深度学习，计算机视觉，人脸计算

多篇重要研究文章作者，重要会议期刊审稿人

微博ID：weightlee03（相关资料分享）

GitHub ID：wiibrew（课程代码发布）

结构



对抗网络
生成网络
V.S.
无限
数据？
判别网络
更深，更
目标分类
迁移
学习
人脸识别，物
品识别，场景
识别，文字识别
线性
回归
非线性
激励
神经
网络
卷积核
CNN
图像特征，
位置回归
目标检测
安防，
自动驾驶
时域递归
Img Ebd
CNN 时域化
外部
反馈
记忆
GATE
RNN
LSTM
视频
分类
CNN 观察决策
Word2vec
增强学习
LSTM 卷
积化？？
语句生成
视频
搜索？
围棋，德州
扑克，自动游戏
自动翻译
智能对话
图片标注
看图
说话

提纲

- 1.AlexNet：现代神经网络起源
- 2. VGG：AlexNet增强版
- 3. GoogLeNet：多维度识别
- 4. ResNet：机器超越人类识别
- 5. DeepFace：结构化图片的特殊处理
- 6. U – Net：图片生成网络
- 7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测

期待目标

- 1. 掌握AlexNet结构特点，神经网络各层之间特征传导关系，模型参数总数计算
- 2. 了解VGG，GoogLeNet，ResNet等复杂ImageNet模型的结构特点，简单设计思想
- 3. 针对特殊数据，特殊任务设计的神经网络结构
- 4. 深度剖析VGG tf代码，学会对已有模型进行参数读取，目标预测，特征提取。

提纲

- **1.AlexNet：现代神经网络起源**
- 2. VGG：AlexNet增强版
- 3. GoogLeNet：多维度识别
- 4. ResNet：机器超越人类识别
- 5. DeepFace：结构化图片的特殊处理
- 6. U－Net：图片生成网络
- 7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测

AlexNet：现代神经网络起源

背景介绍

ImageNet Challenge: 1000类物体，每类1000张图片

传统方法思路：

1. 图片特征提取
2. 机器学习分类

AlexNet：现代神经网络起源

背景介绍

2010年冠军

NEC Empowered by Innovation

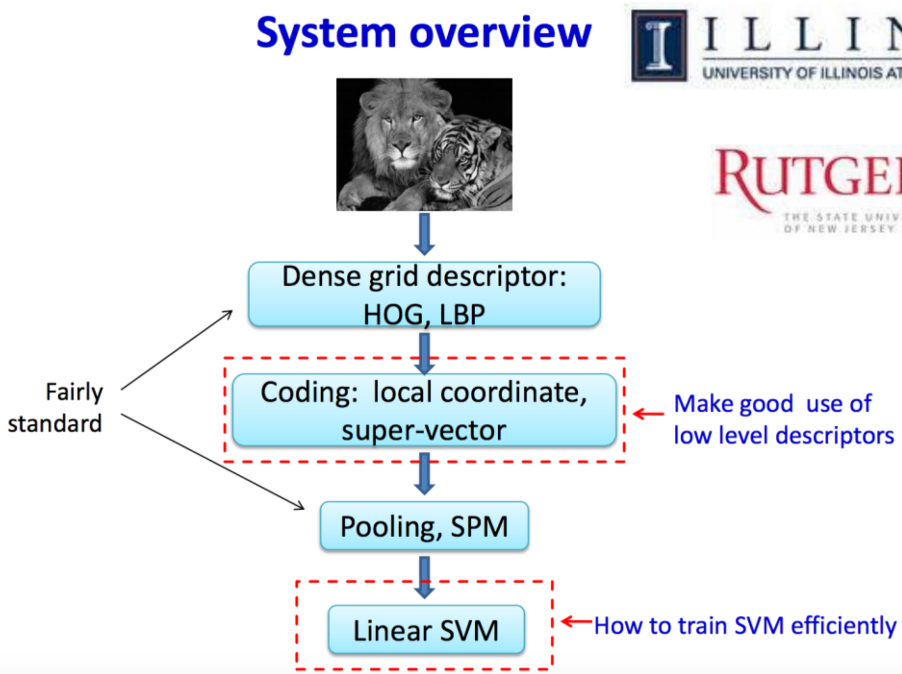
Yuanqing Lin, *Fengjun Lv, Shenghuo Zhu, Ming Yang, Timothee Cour, Kai Yu*

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

LiangLiang Cao, Zhen Li, Min-Hsuan Tsai, Xi Zhou, Thomas Huang

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

Tong Zhang



AlexNet：现代神经网络起源

背景介绍

2011年冠军：Xerox Lab

1.特征提取

Low-level feature extraction ≈ 10k patches per image

- SIFT: 128-dim
- color: 96-dim

reduced to 64-dim with PCA

2.Fisher压缩

FV extraction and compression:

- N=1,024 Gaussians, R=4 regions ⇨ 520K dim x 2
- compression: G=8, b=1 bit per dimension

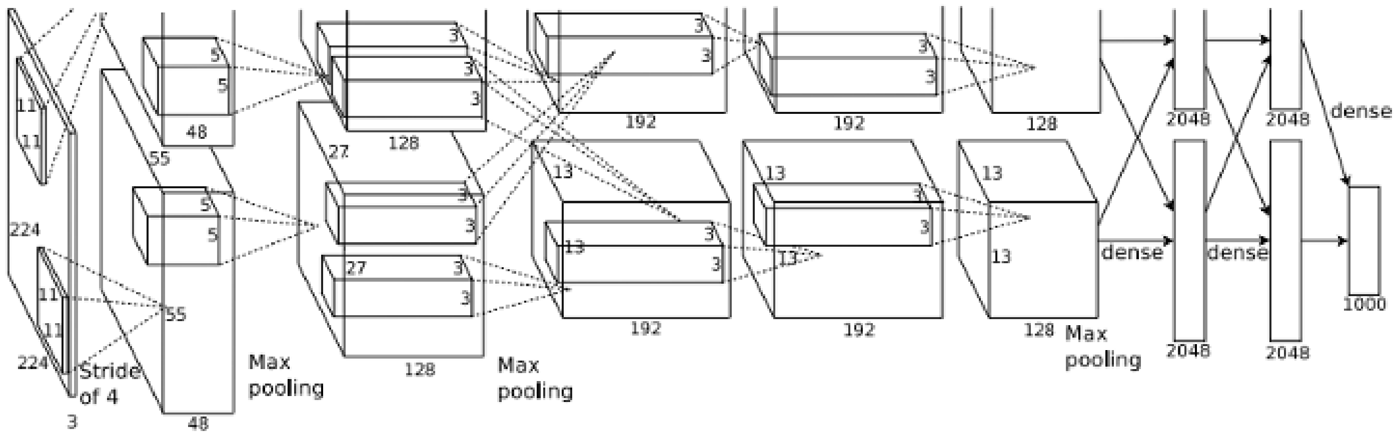
3.SVM分类

One-vs-all SVM learning with SGD

Late fusion of SIFT and color systems

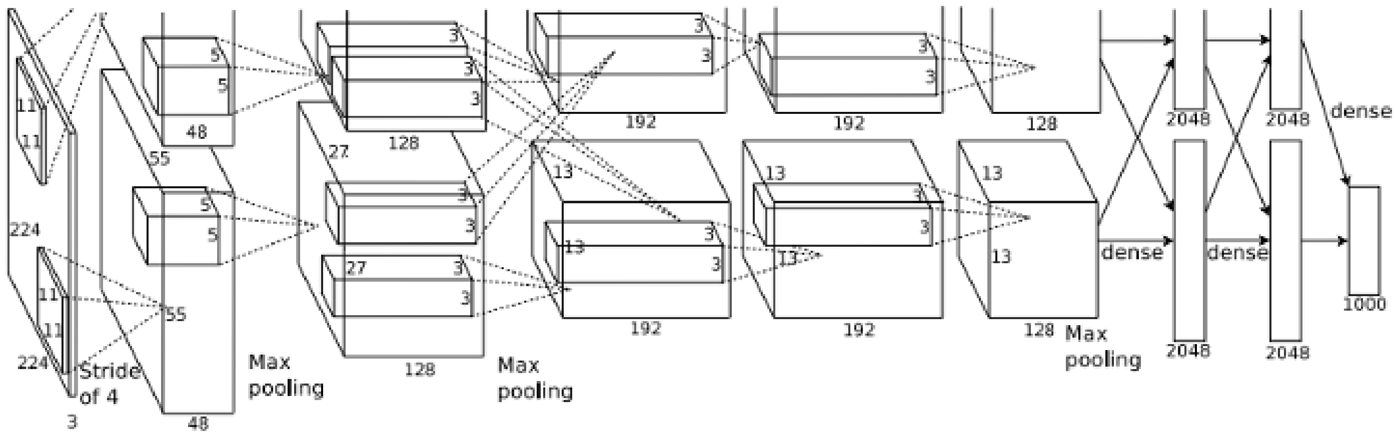
AlexNet：现代神经网络起源

- AlextNet结构

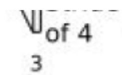


AlexNet：现代神经网络起源

☐ AlextNet结构



[227x227x3] INPUT



[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0

[27x27x96] MAX POOL1: 3x3 filters at stride 2

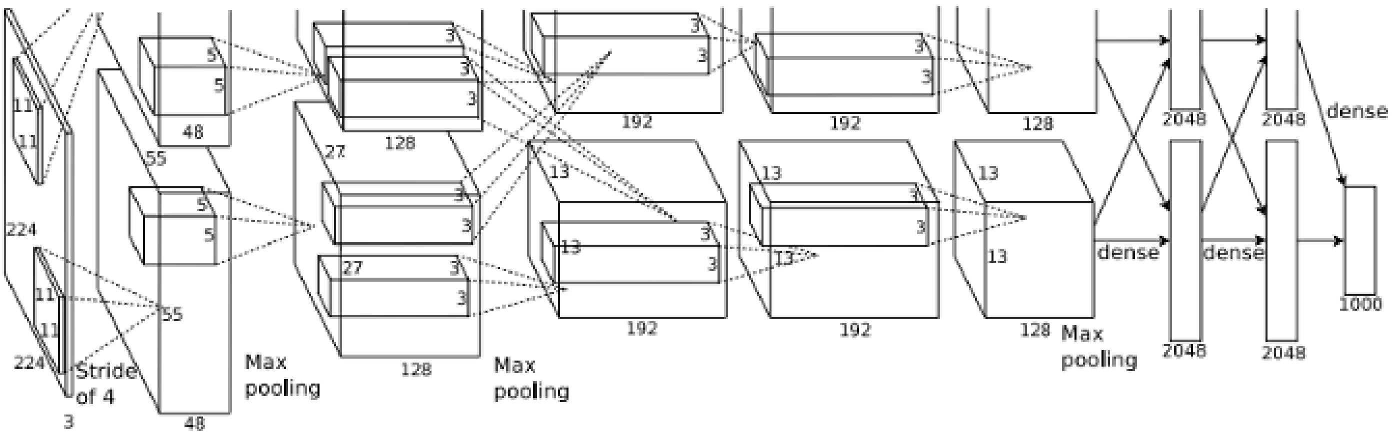
[27x27x96] NORM1: Normalization layer

[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2

[13x13x256] MAX POOL2: 3x3 filters at stride 2

AlexNet：现代神经网络起源

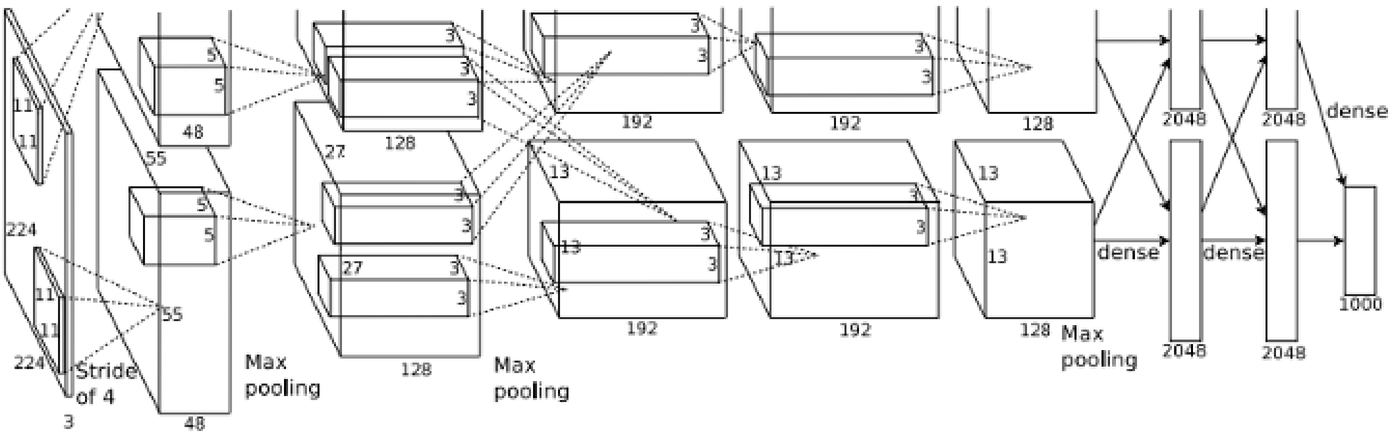
- AlextNet结构



[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2

AlexNet：现代神经网络起源

- AlextNet结构



[4096] FC6: 4096 neurons

[4096] FC7: 4096 neurons

[1000] FC8: 1000 neurons (class scores)

AlexNet：现代神经网络起源

- 参数计算

- MAX Pool3: 6x6x256
- FC1: 4096->4096x36x256=37,748,736
- FC2: 4096 ->4096x4096=16,777,216
- Final: 1000->1000x4096=4,096,000
- 大约6千万参数

AlexNet：现代神经网络起源

影响

- 深度学习开始标志

Imagenet classification with deep convolutional neural networks
A Krizhevsky, I Sutskever, GE Hinton - Advances in neural …, 2012 - papers.nips.cc
Abstract We trained a large, deep convolutional neural network to classify the 1.3 million high-resolution images in the LSVRC-2010 ImageNet training set into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 39.7\% and 18.9\% which is considerably better than the previous state-of-the-art results. The neural network, which has 60 million parameters and 500,000 neurons, consists of five convolutional ...
Cited by 10149 Related articles All 97 versions Cite Save

- 卷积神经网络的基本构成

卷积层＋池化层＋全连接层

- 第一个base model

花朵种类，鸟类种类识别

提纲

- 1.AlexNet：现代神经网络起源
- **2. VGG：AlexNet增强版**
- 3. GoogLeNet：多维度识别
- 4. ResNet：机器超越人类识别
- 5. DeepFace：结构化图片的特殊处理
- 6. U – Net：图片生成网络
- 7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测

VGG：AlexNet增强版

- VGG: Visual Geometry Group
  Department of Engineering Science, University of Oxford

VGG-AlexNet对比

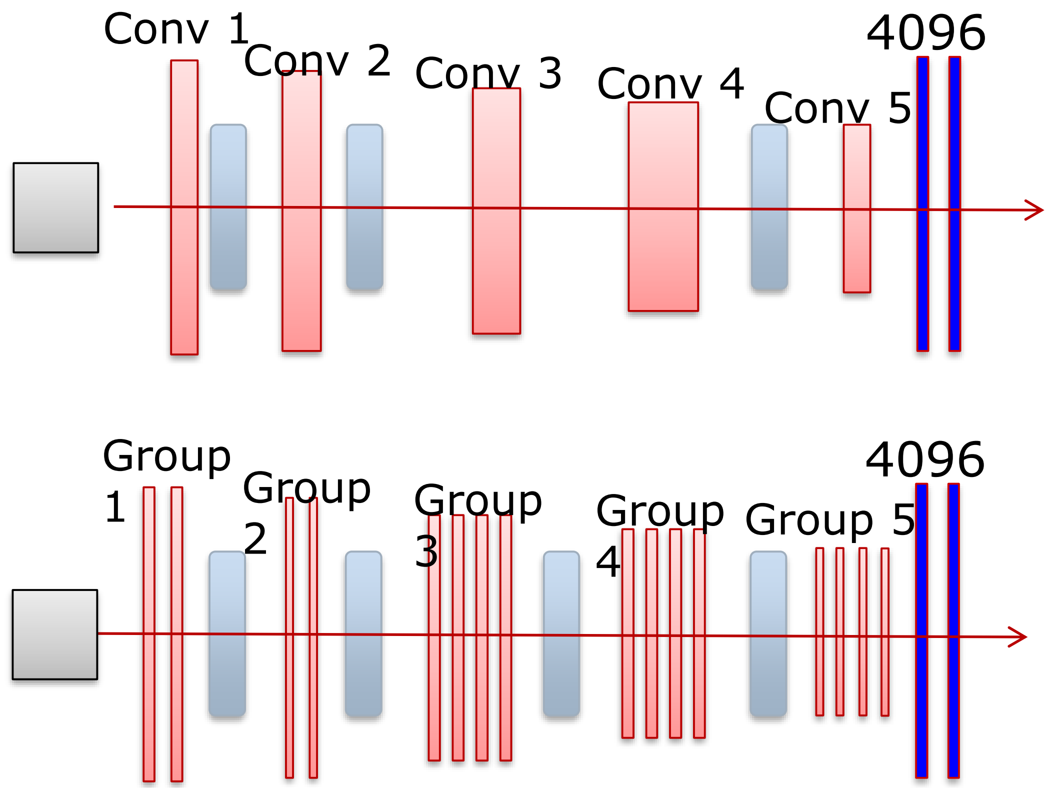
卷积层 – 卷积群

参数个数：

138m – 60m

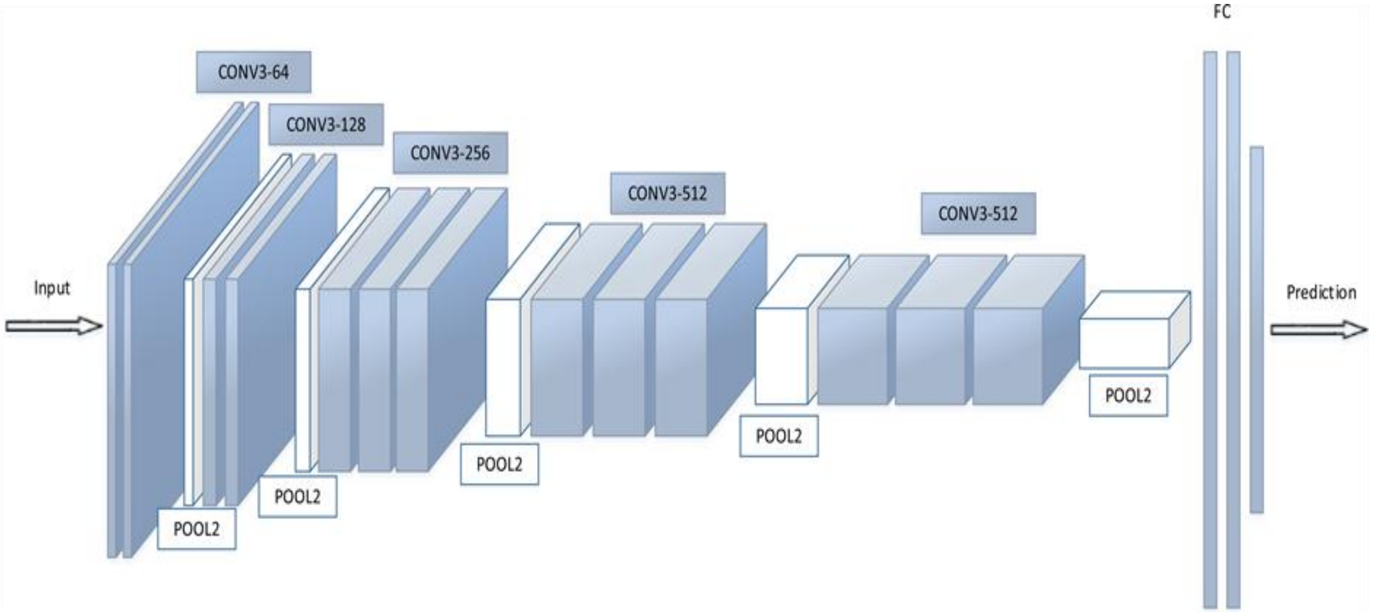
识别率（top5）：

7.3% – 15.3%



VGG：AlexNet增强版

- VGG结构



INPUT: [224x224x3]
CONV3-64: [224x224x64]
CONV3-64: [224x224x64]
POOL2: [112x112x64]
CONV3-128: [112x112x128]
CONV3-128: [112x112x128]
POOL2: [56x56x128]
CONV3-256: [56x56x256]
CONV3-256: [56x56x256]
CONV3-256: [56x56x256]
POOL2: [28x28x256]
CONV3-512: [28x28x512]
CONV3-512: [28x28x512]
CONV3-512: [28x28x512]
POOL2: [14x14x512]
CONV3-512: [14x14x512]
CONV3-512: [14x14x512]
CONV3-512: [14x14x512]
POOL2: [7x7x512]
FC: [1x1x4096]
FC: [1x1x4096]
FC: [1x1x1000]

VGG：AlexNet增强版

- VGG参数计算

(not counting biases)

INPUT: [224x224x3] memory: 224*224*3=150K params: 0
CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*3)*64 = 1,728
CONV3-64: [224x224x64] memory: 224*224*64=3.2M params: (3*3*64)*64 = 36,864
POOL2: [112x112x64] memory: 112*112*64=800K params: 0
CONV3-128: [112x112x128] memory: 112*112*128=1.6M params: (3*3*64)*128 = 73,728
CONV3-128: [112x112x128] memory: 112*112*128=1.6M params: (3*3*128)*128 = 147,456
POOL2: [56x56x128] memory: 56*56*128=400K params: 0
CONV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*128)*256 = 294,912
CONV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*256)*256 = 589,824
CONV3-256: [56x56x256] memory: 56*56*256=800K params: (3*3*256)*256 = 589,824
POOL2: [28x28x256] memory: 28*28*256=200K params: 0
CONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*256)*512 = 1,179,648
CONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*512)*512 = 2,359,296
CONV3-512: [28x28x512] memory: 28*28*512=400K params: (3*3*512)*512 = 2,359,296
POOL2: [14x14x512] memory: 14*14*512=100K params: 0
CONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296
CONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296
CONV3-512: [14x14x512] memory: 14*14*512=100K params: (3*3*512)*512 = 2,359,296
POOL2: [7x7x512] memory: 7*7*512=25K params: 0
FC: [1x1x4096] memory: 4096 params: 7*7*512*4096 = 102,760,448
FC: [1x1x4096] memory: 4096 params: 4096*4096 = 16,777,216
FC: [1x1x1000] memory: 1000 params: 4096*1000 = 4,096,000

TOTAL memory: 24M * 4 bytes ~= 93MB / image (only forward! ~*2 for bwd)
TOTAL params: 138M parameters

VGG：AlexNet增强版

- VGG作用

结构简单：同AlexNet结构类似，均为卷积层，池化层，全连接层的组合

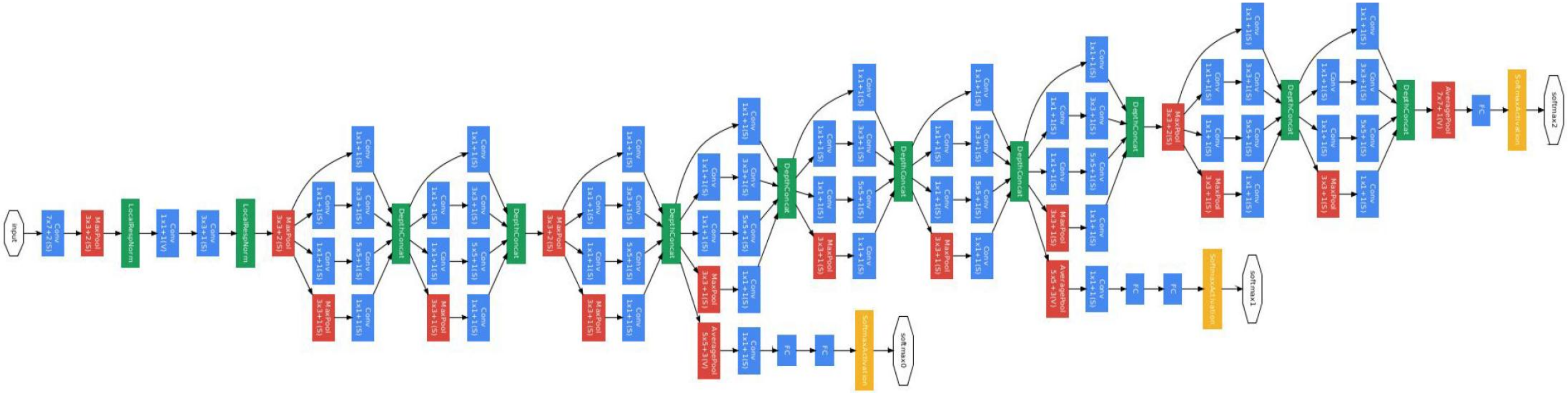
性能优异：同Alexnet提升明显，同GoogleNet，ResNet相比，表现接近

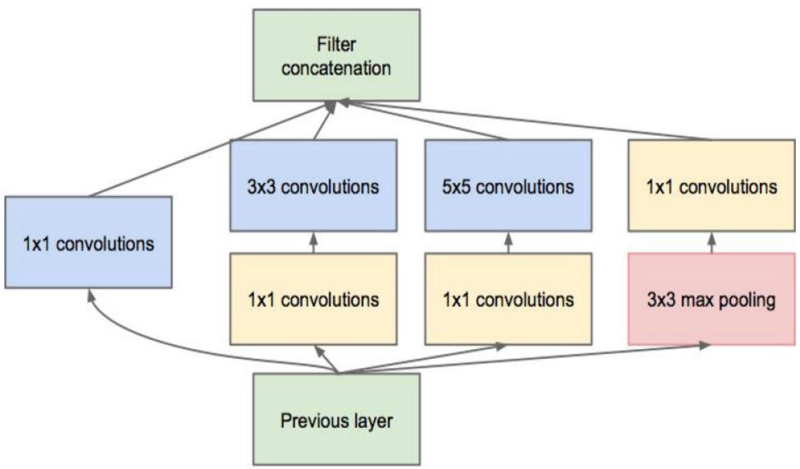
选择最多的基本模型：方便进行结构的优化，设计，SSD，RCNN，等其他任务的基本模型（base model）

提纲

- 1.AlexNet：现代神经网络起源
- 2. VGG：AlexNet增强版
- **3. GoogLeNet：多分辨率识别**
- 4. ResNet：机器超越人类识别
- 5. DeepFace：结构化图片的特殊处理
- 6. U－Net：图片生成网络
- 7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测

GoogLeNet：多分辨率融合





Inception module

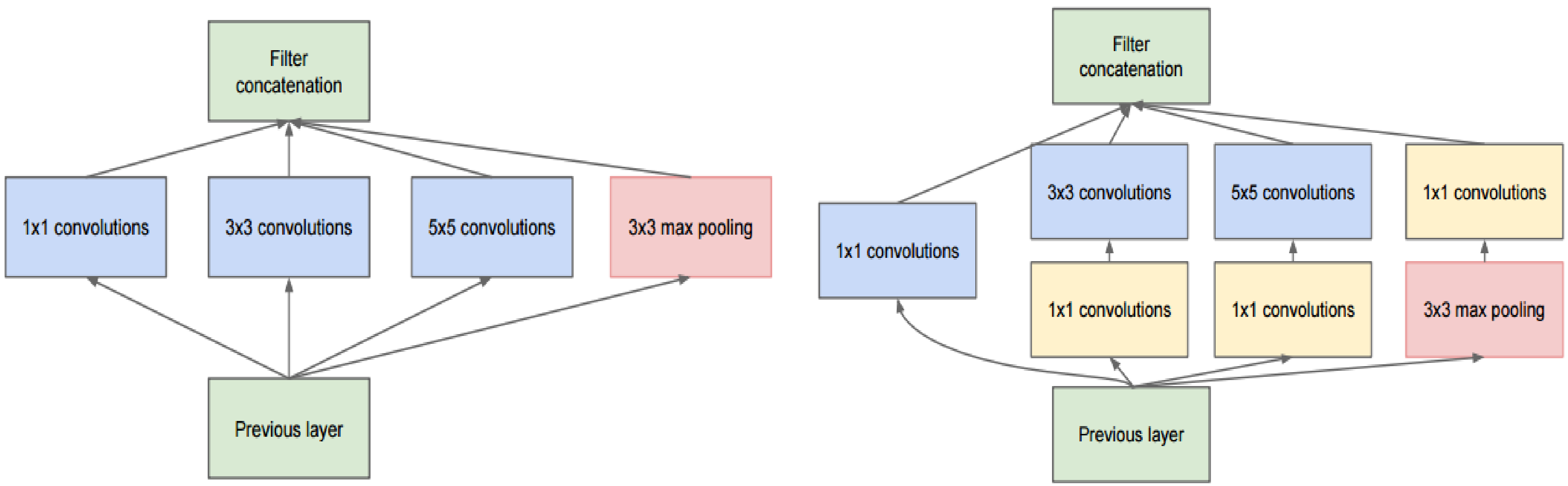
ILSVRC 2014 winner (6.7% top 5 error)

[From Stanford cs231n]

GoogLeNet：多分辨率融合

□ Inception 结构发展

All we need is to find the optimal local construction and to repeat it spatially.

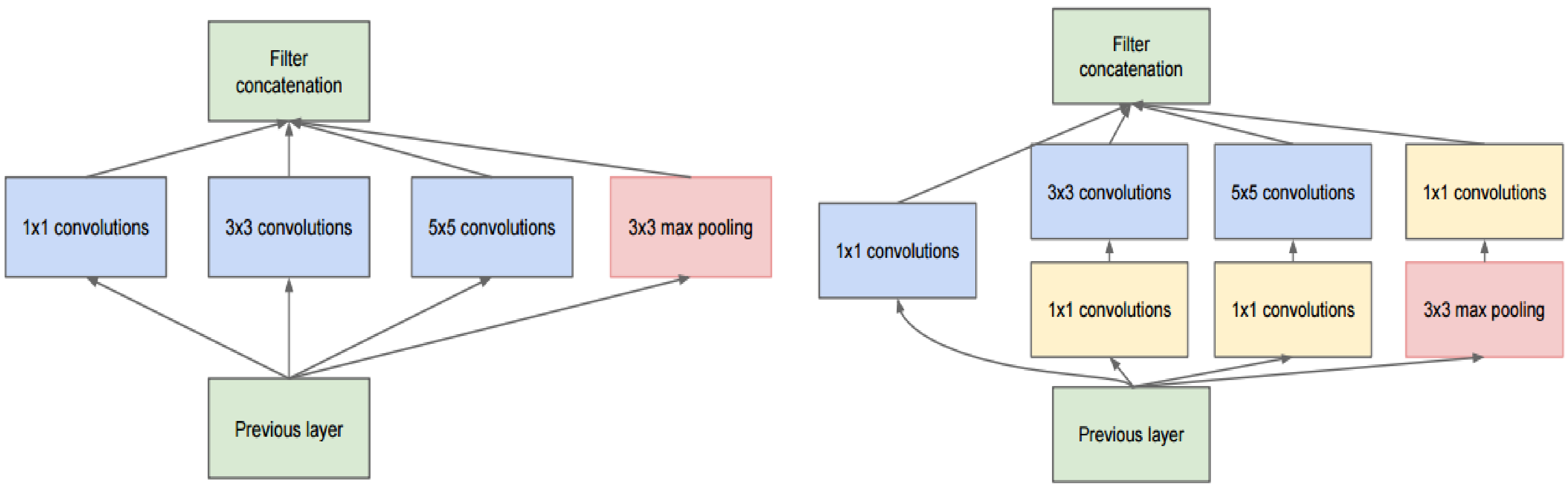


结构问题是什么？　　　　1x1 卷积的好处？

GoogLeNet：多分辨率融合

☐ Inception 结构发展

All we need is to find the optimal local construction and to repeat it spatially.



结构问题是什么？参数暴增　　1x1 卷积的好处？减少参数

GoogLeNet：多分辨率融合

❑ 结构细节

❑ 特点

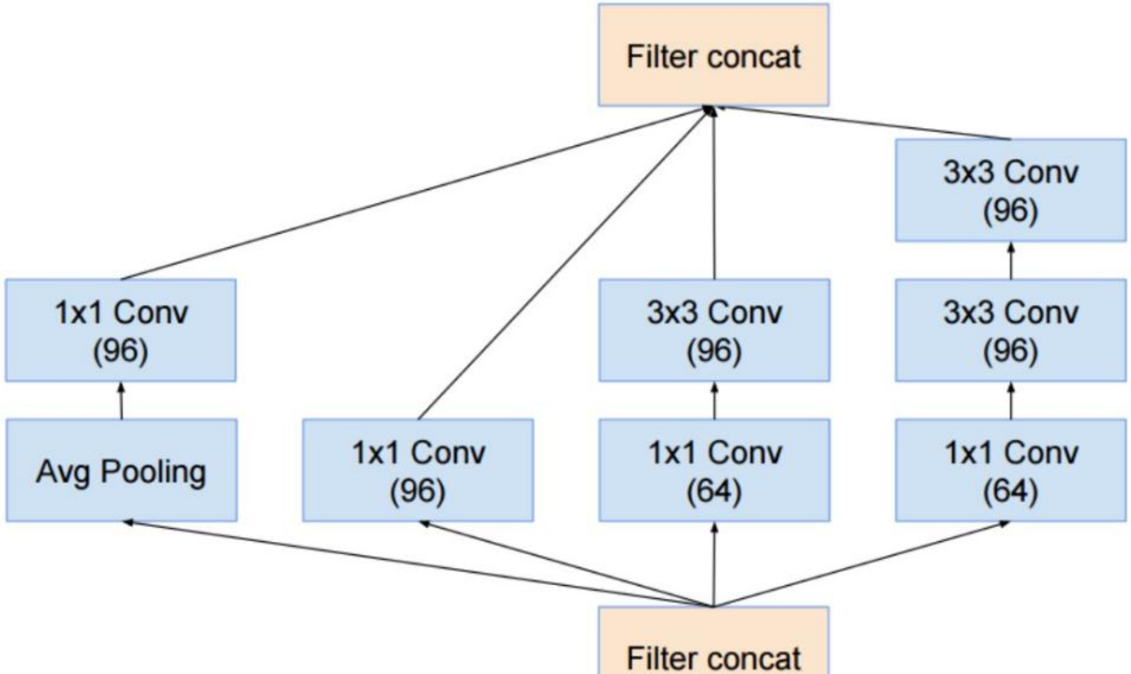
参数总数，5m

没有全连接

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								

GoogLeNet：多分辨率融合

- Inception 结构发展 – Inception v4



反超了ResNet 3.1% error on top 5.

GoogLeNet：多分辨率融合

□ 全卷积结构（FCN）

一般的神经网络：卷积层(CNN) + 全连接层(FC)

全卷积网络：没有全连接层

特点：

1. 输入图片大小无限制

2. 空间信息有丢失

3. 参数更少，表达力更强

［内容参考 https://www.quora.com/What-are-the-advantages-of-Fully-Convolutional-Networks-over-CNNs］

提纲

- 1.AlexNet：现代神经网络起源
- 2. VGG：AlexNet增强版
- 3. GoogLeNet：多维度识别
- **4. ResNet：机器超越人类识别**
- 5. DeepFace：结构化图片的特殊处理
- 6. U – Net：图片生成网络
- 7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测

ResNet：机器超越人类识别

Microsoft Research

MSRA @ ILSVRC & COCO 2015 Competitions

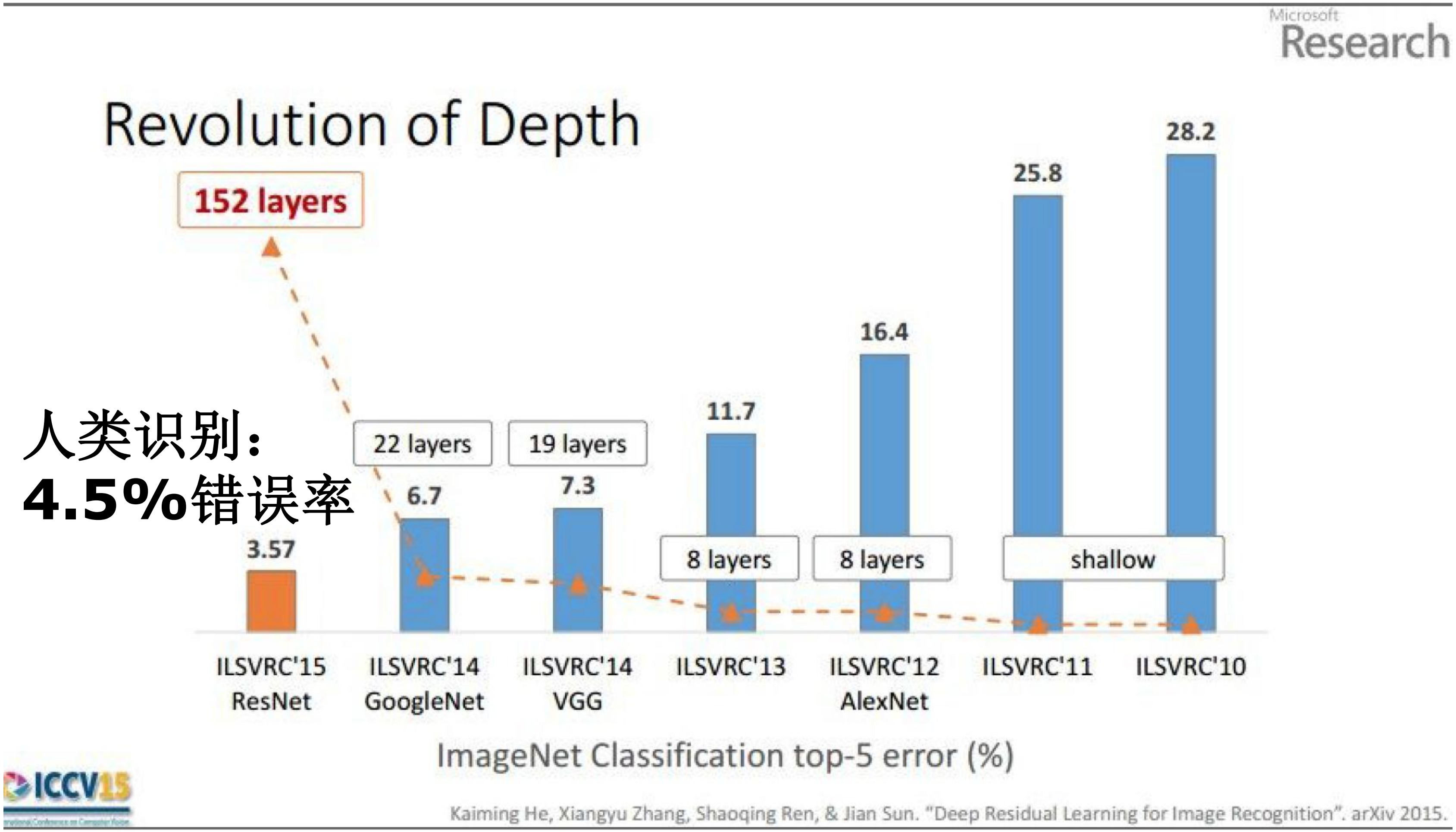
- **1st places in all five main tracks**
  - ImageNet Classification: "*Ultra-deep*" (quote Yann) 152-layer nets
  - ImageNet Detection: 16% better than 2nd
  - ImageNet Localization: 27% better than 2nd
  - COCO Detection: 11% better than 2nd
  - COCO Segmentation: 12% better than 2nd

*improvements are relative numbers

ICCV15

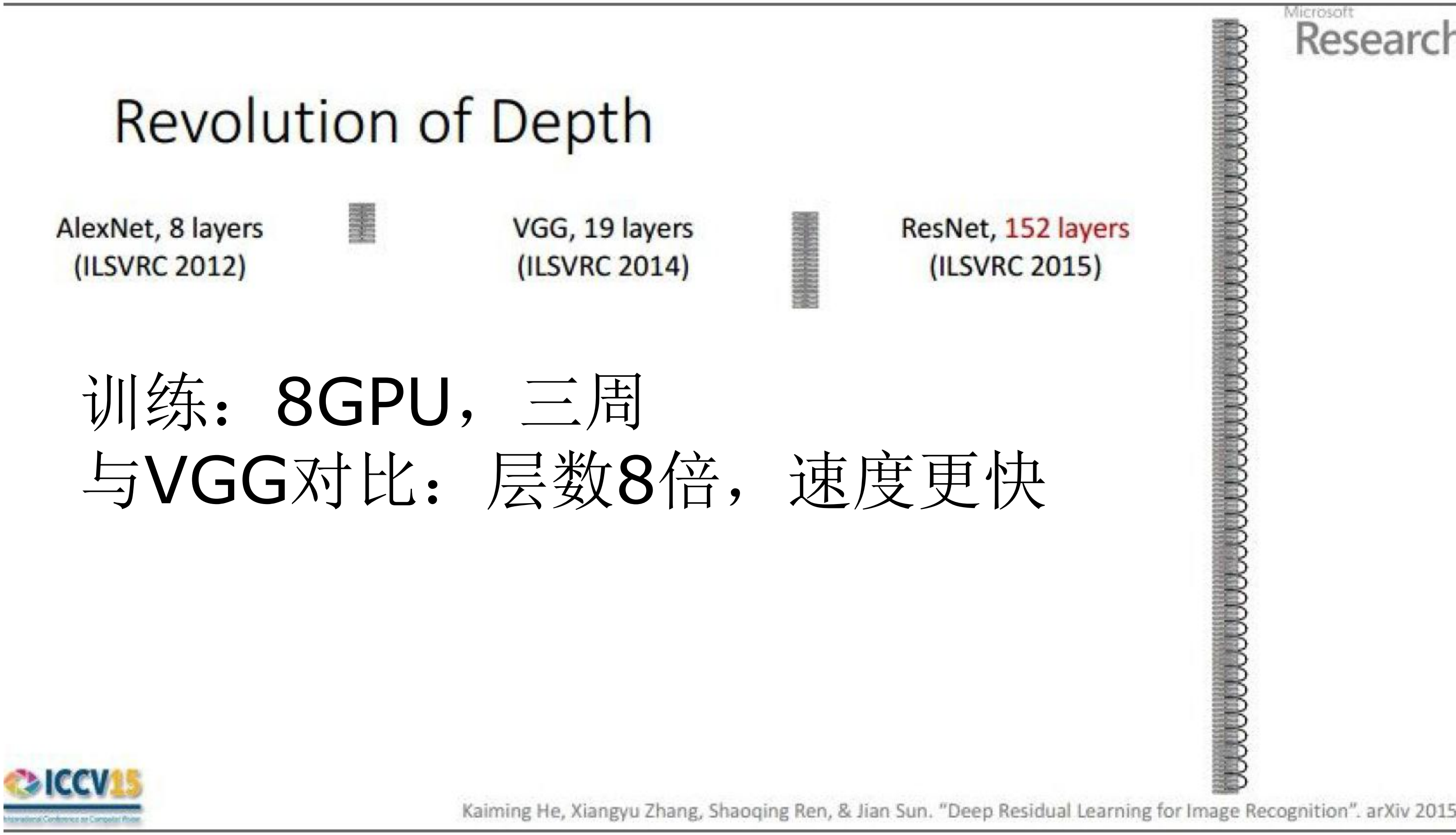
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

ResNet：机器超越人类识别



人类识别：
4.5%错误率

ResNet：机器超越人类识别

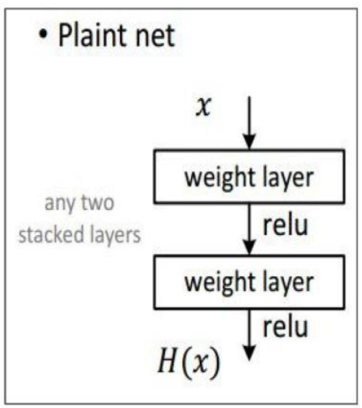


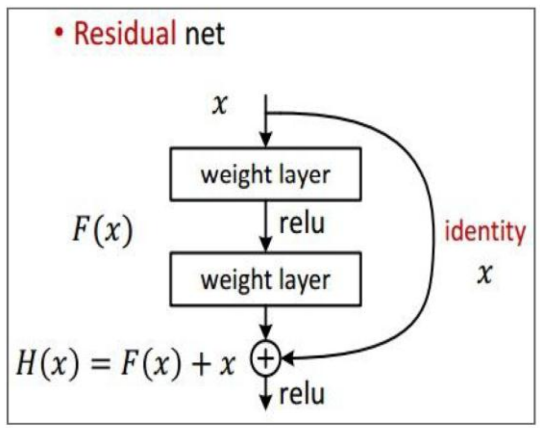
训练：8GPU，三周

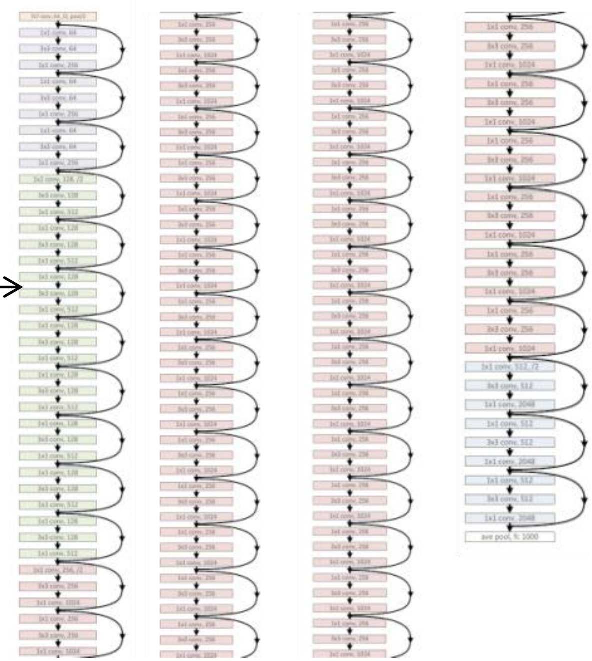
与VGG对比：层数8倍，速度更快

ResNet：机器超越人类识别

□ 结构特性







互联网新技术在线教育领航者

小象学院 ChinaHadoop.cn

ResNet：机器超越人类识别

- 为什么ResNet有效？

ResNet：机器超越人类识别

- 为什么ResNet有效？

- 1.前向计算：低层卷积网络高层卷积网络信息融合；层数越深，模型的表现力越强 [1]

- 2.反向计算：导数传递更直接，越过模型，直达各层

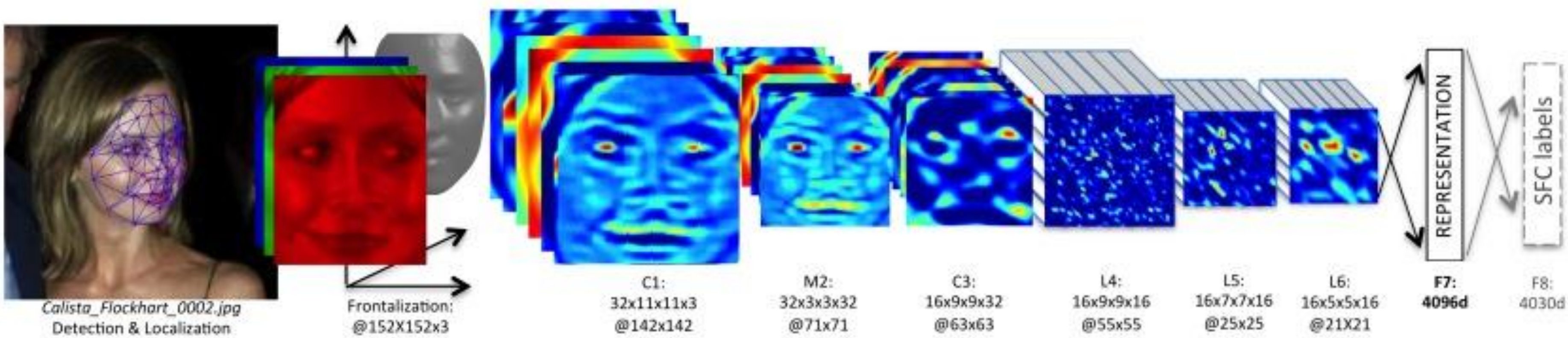
［1］Benefits of depth in neural networks by Matus Telgarsky.

提纲

- 1.AlexNet：现代神经网络起源
- 2. VGG：AlexNet增强版
- 3. GoogLeNet：多维度识别
- 4. ResNet：机器超越人类识别
- **5. DeepFace：结构化图片的特殊处理**
- 6. U – Net：图片生成网络
- 7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测

DeepFace：结构化图片的特殊处理

- 人脸识别：通过观察人脸确定对应身份，在应用中更多的是确认(verification)。



DeepFace：结构化图片的特殊处理

- 人脸识别数据特点：
- 结构化：所有人脸，组成相似，理论上能够实现对齐
- 差异化：相同位置，形貌(appearance)不同

DeepFace：结构化图片的特殊处理

- 人脸识别数据特点：

1. 结构化：所有人脸，组成相似，理论上能够实现对齐
2. 差异化：相同位置，形貌(appearance)不同

- 一般神经网络处理人脸识别的问题：

1. 卷积核同整张图片卷积运算，卷积核参数共享，不同局部特性对参数影响相互削弱
2. 解决方法：不同区域，不同参数

DeepFace：结构化图片的特殊处理

- 不同局部，不同参数

1. 人脸对准

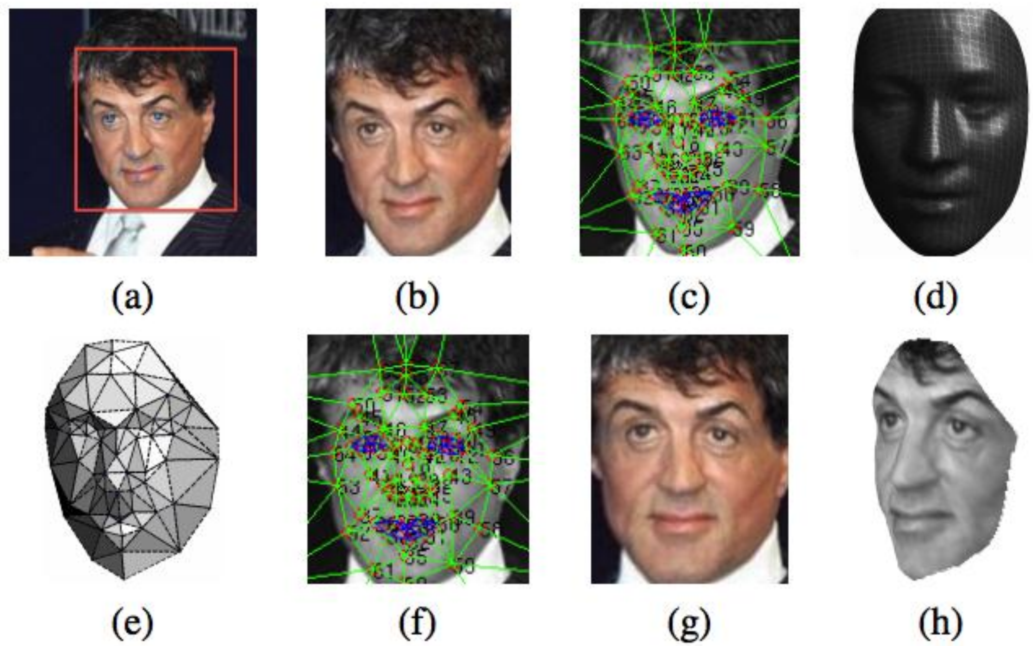
二维对准：

二维矩阵(R,T)运算

三维对准：

三维标准模版映射

三维投影二维

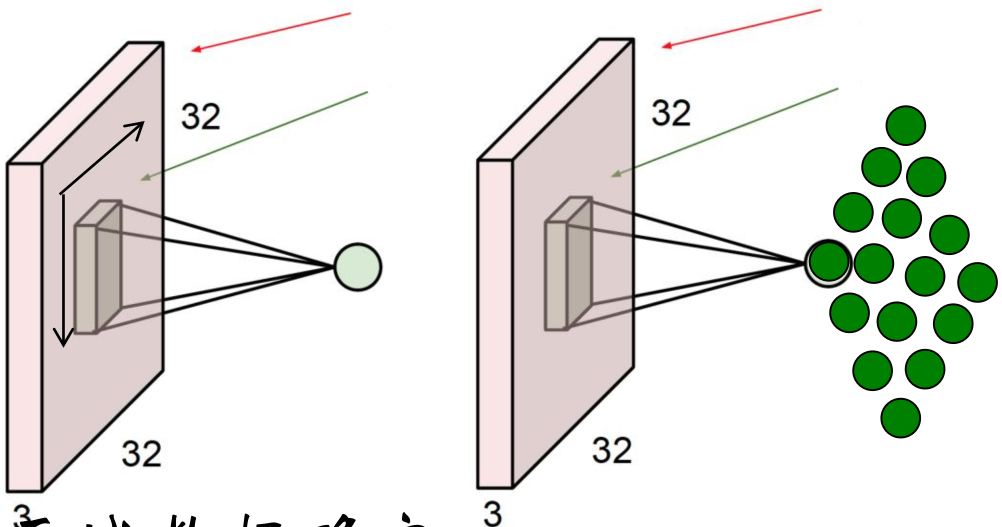


DeepFace：结构化图片的特殊处理

- 不同局部，不同参数

2. 局部卷积

- 每个卷积核固定某一区域，不移动
- 不同区域之间不共享卷积核
- 卷积核参数由固定区域数据确定



DeepFace：结构化图片的特殊处理

识别效果

Method	Accuracy ± SE	Protocol
Joint Bayesian [6]	0.9242 ±0.0108	restricted
Tom-vs-Pete [4]	0.9330 ±0.0128	restricted
High-dim LBP [7]	0.9517 ±0.0113	restricted
TL Joint Bayesian [5]	0.9633 ±0.0108	restricted
DeepFace-single	**0.9592** ±0.0029	unsupervised
DeepFace-single	**0.9700** ±0.0028	restricted
DeepFace-ensemble	**0.9715** ±0.0027	restricted
DeepFace-ensemble	**0.9735** ±0.0025	unrestricted
Human, cropped	0.9753	

DeepFace：结构化图片的特殊处理

- 全局部卷积连接的缺陷

1. 预处理：大量对准，对对准要求高，原始信息可能丢失
2. 卷积参数数量很大，模型收敛难度大，需要大量数据（Facebook 数据不公开）
3. 模型可扩展性差，基本限于人脸计算

提纲

- 1.AlexNet：现代神经网络起源
- 2. VGG：AlexNet增强版
- 3. GoogLeNet：多维度识别
- 4. ResNet：机器超越人类识别
- 5. DeepFace：结构化图片的特殊处理
- **6. U – Net：图片生成网络**
- 7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测

U—Net：图片生成网络

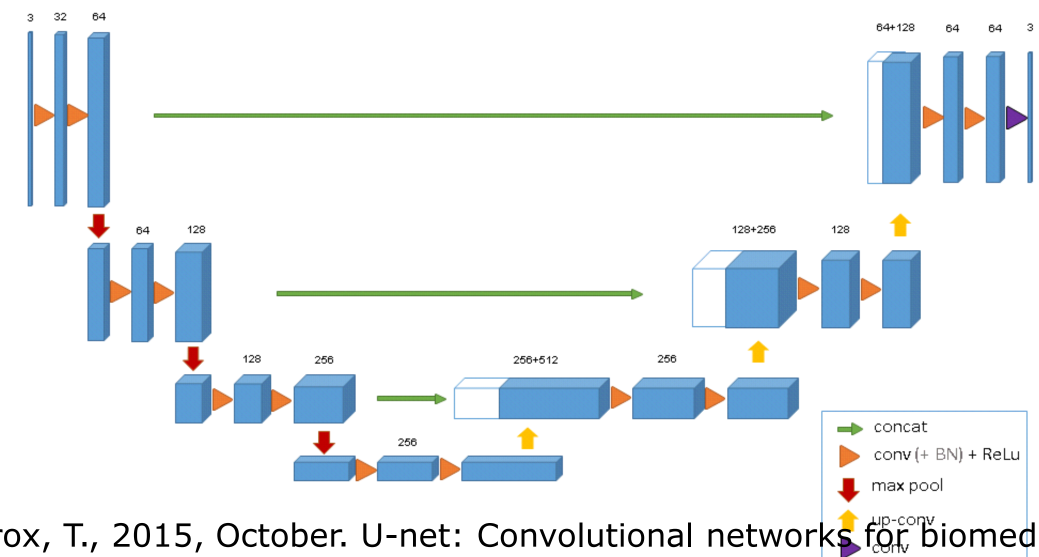
- 图片生成网络

通过卷积神经网络生成特殊类型的图片

图片所有pixel需要生成，多目标回归

- U-Net

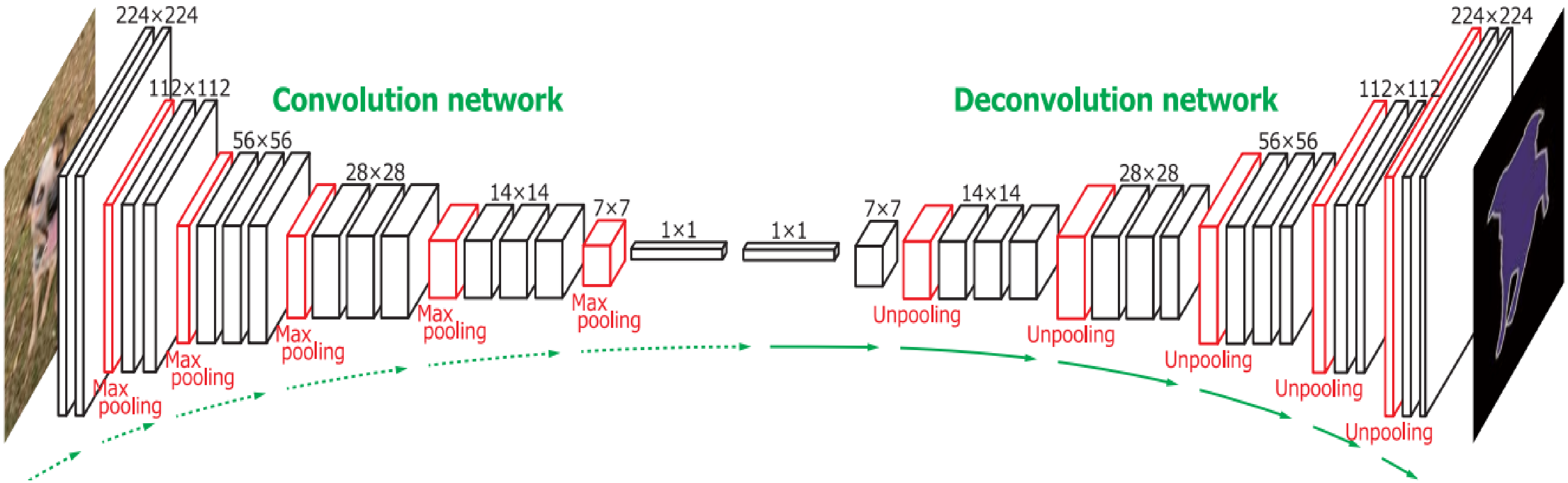
Conv-Fc-Conv



Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241). Springer International Publishing.

U－Net：图片生成网络

- VGG U-Net



[Noh, H., Hong, S. and Han, B., 2015. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1520-1528).]

U—Net：图片生成网络

- 卷积－逆卷积；池化－反池化（增维）

Convolution-Deconvolution; Pooling-Unpooling

- 反池化：

记住原有位置，

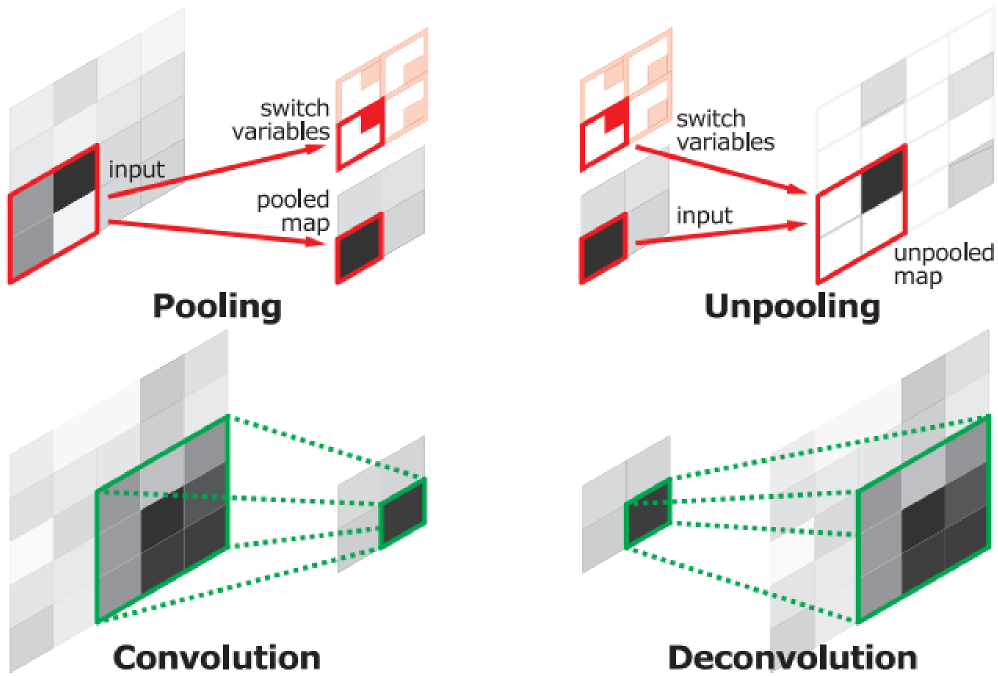
不是resize

- 逆卷积

实质：有学习能力

的上采样

名字疑问？



U—Net：图片生成网络

- 卷积－逆卷积(带参数的上采样)

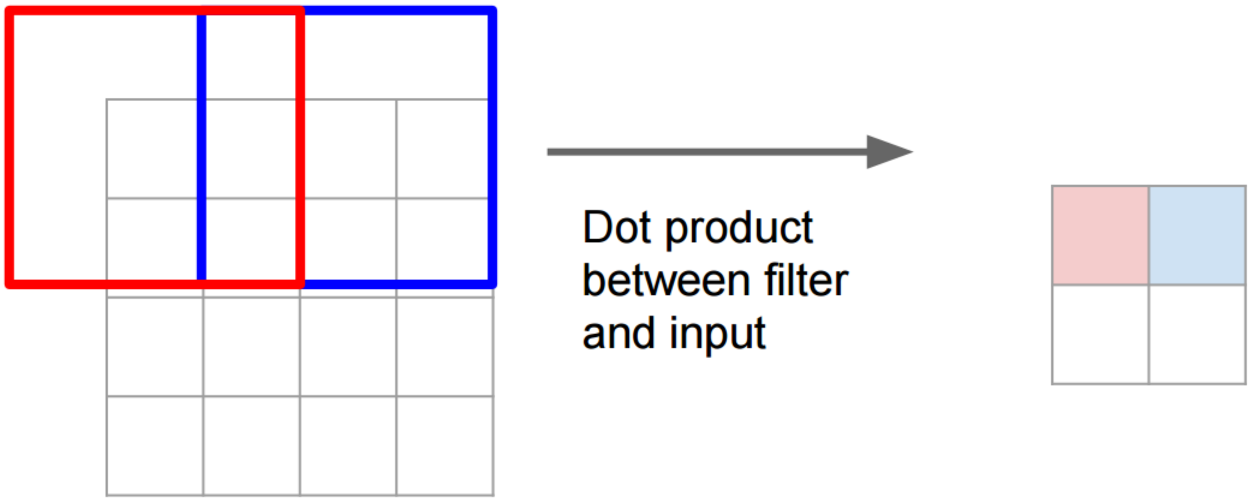
Convolution-Deconvolution (Convolution transpose)

正常卷积：

步长：2

卷积核：3x3

输出：2x2



U—Net：图片生成网络

- 卷积－逆卷积（带参数的上采样）

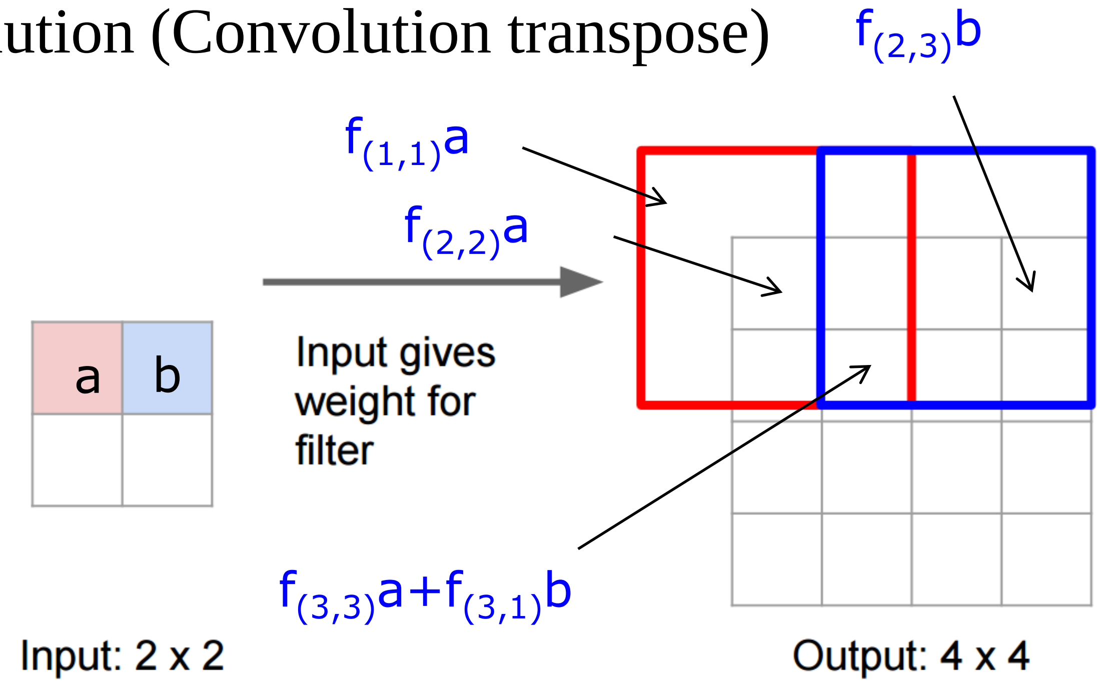
Convolution-Deconvolution (Convolution transpose)

逆卷积：

步长：2

卷积核：3x3

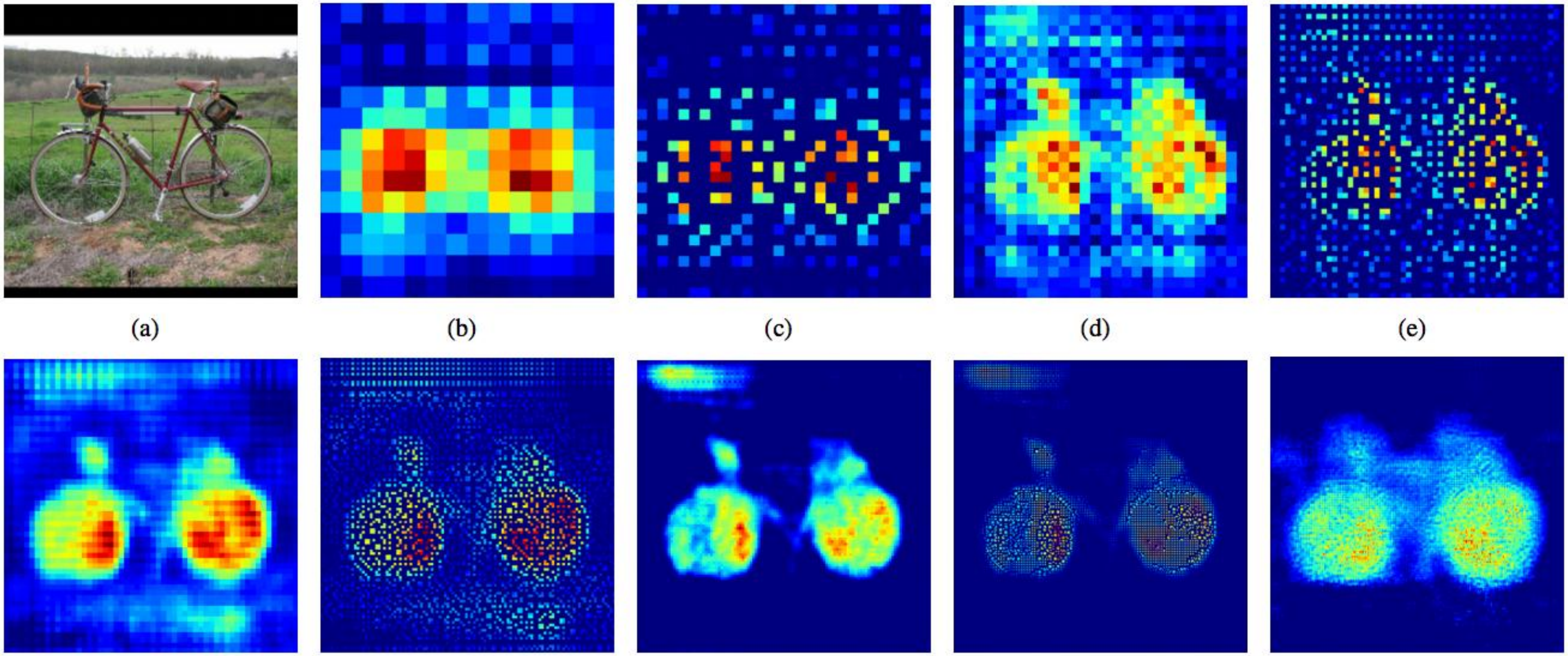
输出：4x4



有学习能力上采样，好处？生成图片更好的连贯性，更好的空间表达能力。

U—Net：图片生成网络

- 图片分割图生成

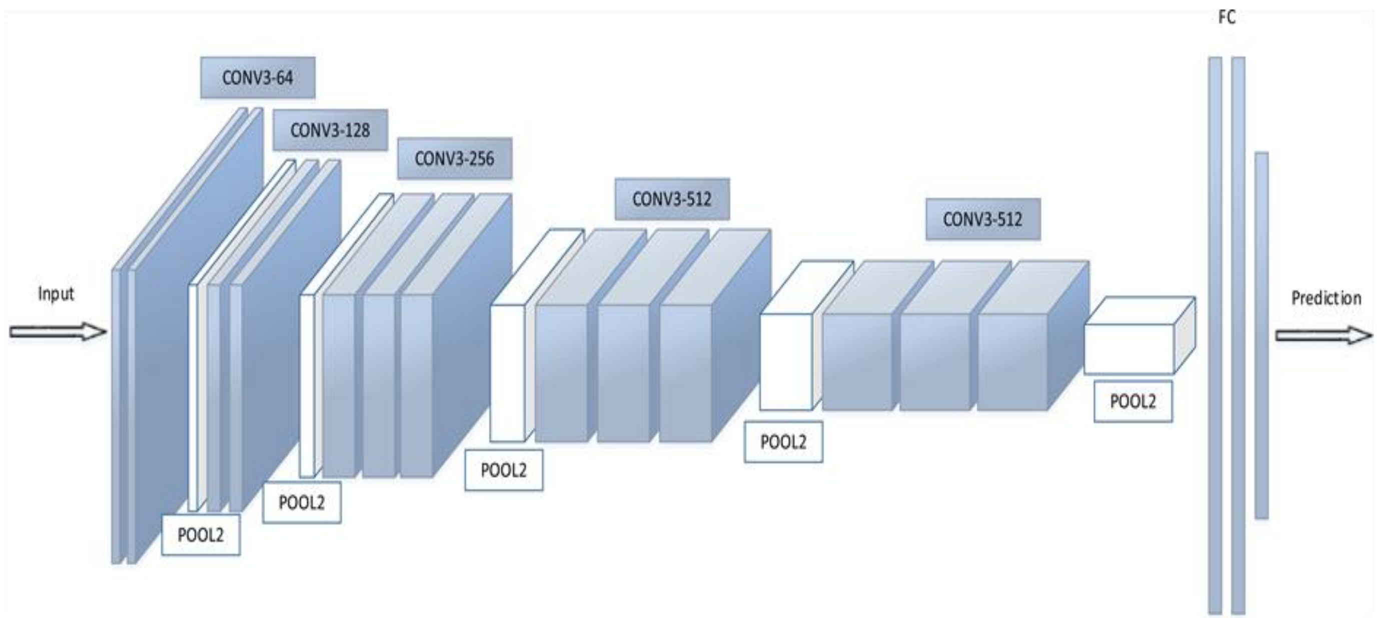


提纲

- 1.AlexNet：现代神经网络起源
- 2. VGG：AlexNet增强版
- 3. GoogLeNet：多维度识别
- 4. ResNet：机器超越人类识别
- 5. DeepFace：结构化图片的特殊处理
- 6. U – Net：图片生成网络
- **7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测**

实例运行—解剖VGG

- 1.观察模型参数
- 2.观察图片中间层（hidden layers）特征图
- 3.运用模型进行预测



实例运行—解剖VGG

- 课程中提到Tflearn地址
- https://github.com/wiibrew/tflearn

总结

- 1.AlexNet：现代神经网络起源
- 2. VGG：AlexNet增强版
- 3. GoogLeNet：多维度识别
- 4. ResNet：机器超越人类识别
- 5. DeepFace：结构化图片的特殊处理
- 6. U – Net：图片生成网络
- 7. 实例：解剖VGG，用模型进行模型参数可视化，特征提取，目标预测

下节预告

- 1. 自主设计神经网络
- 2. Fine-tunning 现有模型
- 3. 基于VGG模型，网络采集图片数据，进行相应分类器的训练

总结

- 有问题请到课后交流区

- 问题答疑： http://www.xxwenda.com/
  - 可邀请老师或者其他人回答问题

- 讲师微博：weightlee03，每周不定期分享DL资料
- GitHub ID：wiibrew （课程代码发布）