

Deep Residual Learning for Image Recognition

导师：余老师



Deep Residual Learning for Image Recognition

图像识别中的深度残差学习网络

作者: Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

单位: MSRA

发表会议及时间: CVPR 2016



前期知识储备

Pre-knowledge reserve



机器学习

了解机器学习基本原理及概念，如数据集划分，损失函数，优化方法等

神经网络

了解神经网络基本知识，特别是卷积神经网络的工作原理等

图像分类

了解了解图像分类的概念，掌握图像分类的流程

PyTorch

了解PyTorch基本使用方法，如数据读取处理，模型构建，损失优化等



学习目标

Learning objectives



课程安排

The schedule of course



第一课：论文导读

The first lesson: the paper guide

目 录

1/ 论文研究背景、成果及意义

2/ 论文泛读

3/ 本课回顾及下节预告

论文研究背景、成果及意义

Background、Results and Meanings

背景、成果和意义

Background、 Results and Meanings



研究背景

Research background

ILSVRC: 大规模图像识别挑战赛

ImageNet Large Scale Visual Recognition Challenge 是李飞飞等人于2010年创办的图像识别挑战赛，自2010起连续举办8年，极大地**推动计算机视觉发展**

比赛项目涵盖：图像分类(Classification)、目标定位(Object localization)、目标检测(Object detection)、视频目标检测(Object detection from video)、场景分类(Scene classification)、场景解析(Scene parsing)

竞赛中脱颖而出大量经典模型：**alexnet, vgg, googlenet, resnet, densenet**等

<http://www.image-net.org>

研究背景

Research background

相关研究

Highway Network

首个成功训练成百上千层（100层及900层）的卷积神经网络

思路：借鉴LSTM，引入门控单元，将传统前向传播增加一条计算路径，变成公式（3）形式

增加了额外训练参数 W_T

Highway Networks: <https://arxiv.org/abs/1505.00387>
Training Very Deep Networks: <https://arxiv.org/abs/1507.06228>

$$y = H(x, W_H). \quad (1)$$

H is usually an affine transform followed by a non-linear activation function, but in general it may take other forms, possibly convolutional or recurrent. For a highway network, we additionally define two non-linear transforms $T(x, W_T)$ and $C(x, W_C)$ such that

$$y = H(x, W_H) \cdot T(x, W_T) + x \cdot C(x, W_C). \quad (2)$$

We refer to T as the *transform gate* and C as the *carry gate*, since they express how much of the output is produced by transforming the input and carrying it, respectively. For simplicity, in this paper we set $C = 1 - T$, giving

$$y = H(x, W_H) \cdot T(x, W_T) + x \cdot (1 - T(x, W_T)). \quad (3)$$

The dimensionality of x , y , $H(x, W_H)$ and $T(x, W_T)$ must be the same for Equation 3 to be valid. Note that this layer transformation is much more flexible than Equation 1. In particular, observe that for particular values of T ,

$$y = \begin{cases} x, & \text{if } T(x, W_T) = 0, \\ H(x, W_H), & \text{if } T(x, W_T) = 1. \end{cases} \quad (4)$$

Similarly, for the Jacobian of the layer transform,

研究背景

Research background

相关研究

Highway Network

首个成功训练成百“上千层”（100层及900层）的卷积神经网络

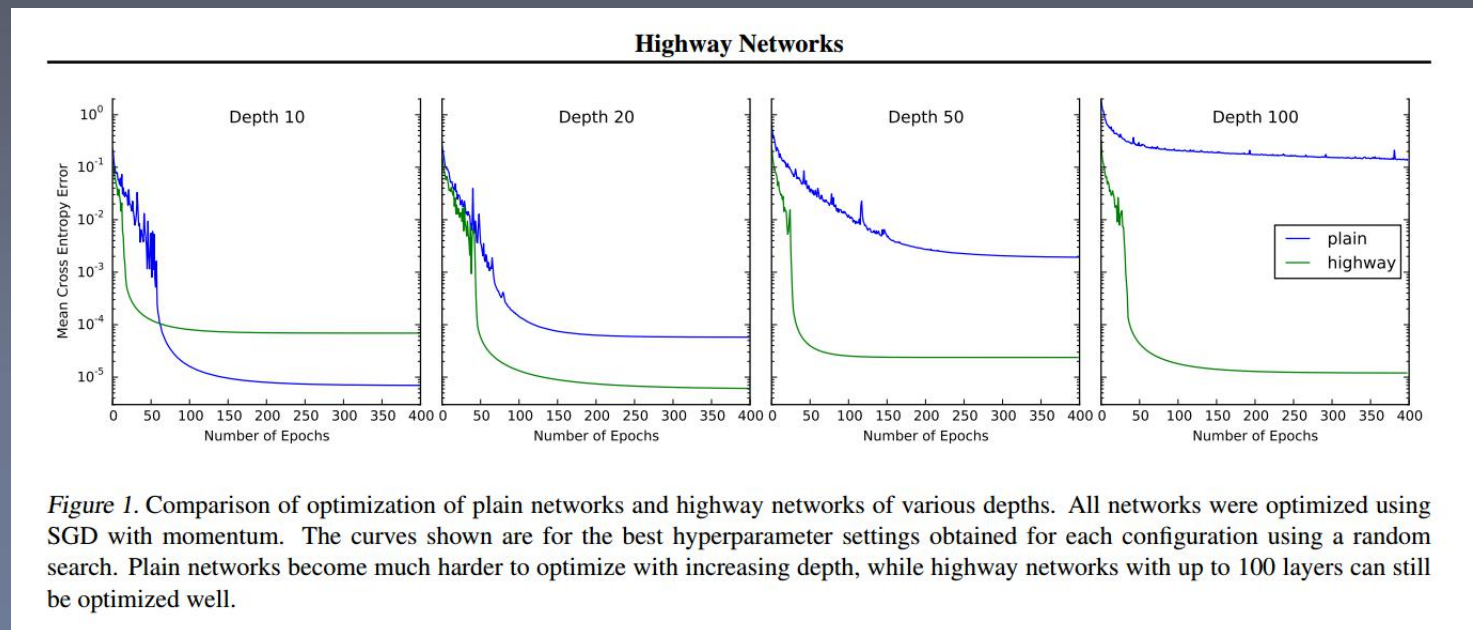
名词解释：

Plain network：普通结构网络

Building block：组件，数个网络层构成的固定系列操作

Parameter-free Identity

shortcut：不带参数的跳层连接



研究成果

Research Results

霸榜各大比赛

仅采用ResNet结构，
无额外的技巧，夺得五个冠军，且与第二名拉开差距

ImageNet 分类、定位、检测

COCO 检测、分割

“Features matter.” (quote [Girshick et al. 2014], the R-CNN paper)

task	2nd-place winner	ResNets	margin (relative)
ImageNet Localization (top-5 error)	12.0	9.0	27%
ImageNet Detection (mAP@.5)	53.6	62.1	16%
COCO Detection (mAP@.5:.95)	33.5	37.3	11%
COCO Segmentation (mAP@.5:.95)	25.1	28.2	12%

- Our results are all based on **ResNet-101**
- Our features are **well transferrable**

研究意义

Research Meaning

研究意义

1. 简洁高效的ResNet受到工业界宠爱，自提出以来已经成为**工业界最受欢迎**的卷积神经网络结构
2. 近代卷积神经网络发展史的又一**里程碑**，突破千层网络，跳层连接成为标配

Deep residual learning for image recognition

[K He](#), [X Zhang](#), [S Ren](#), [J Sun](#) - ... and pattern recognition, 2016 - openaccess.thecvf.com

Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer ...

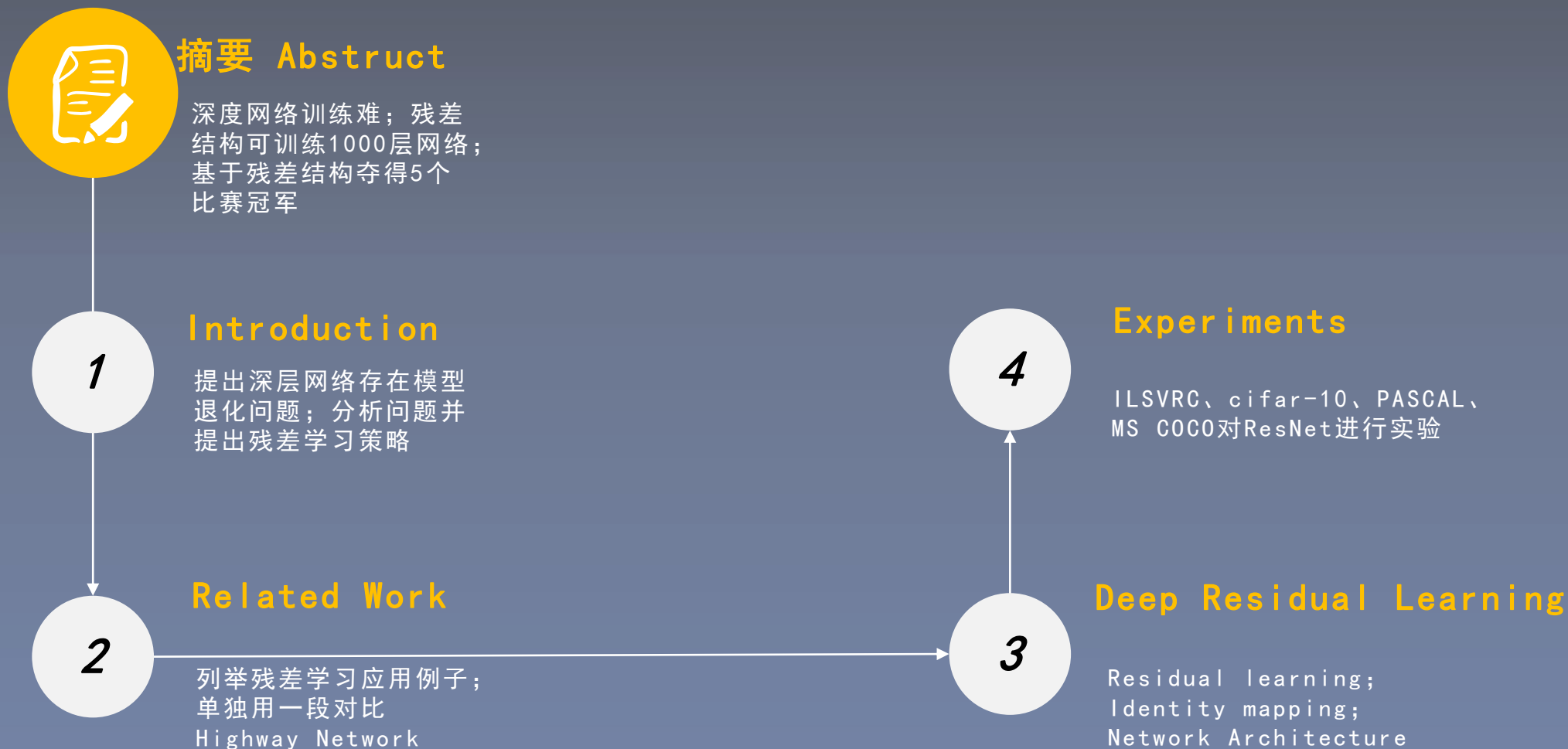
☆ 被引用次数: **47722** 相关文章 所有 64 个版本

论文泛读

Strcuture of Paper

论文结构

Structure of Papers



摘要

abstract

摘要核心

1. 提出问题：深度卷积网络难训练
2. 本文方法：残差学习框架可以让深层网络更容易训练
3. 本文优点：ResNet易优化，并随着层数增加精度也能提升
4. 本文成果：ResNet比VGG深8倍，但是计算复杂度更低，在ILSVRC-2015获得3.57%的top-error
5. 本文其它工作：CIFAR-10上训练1000层的ResNet
6. 本文其它成果：在coco目标检测任务中提升28%的精度，并基于ResNet夺得ILSVRC的检测、定位，COCO的检测和分割四大任务的冠军



论文小标题

Paper title

1. Introduction

2. Related Work

3. Deep Residual Learning

3.1 Residual Learning

3.2 Identity Mapping by Shortcuts

3.3 Network Architectures

3.4 Implementation

4. Experiment

4.1 ImageNet Classification

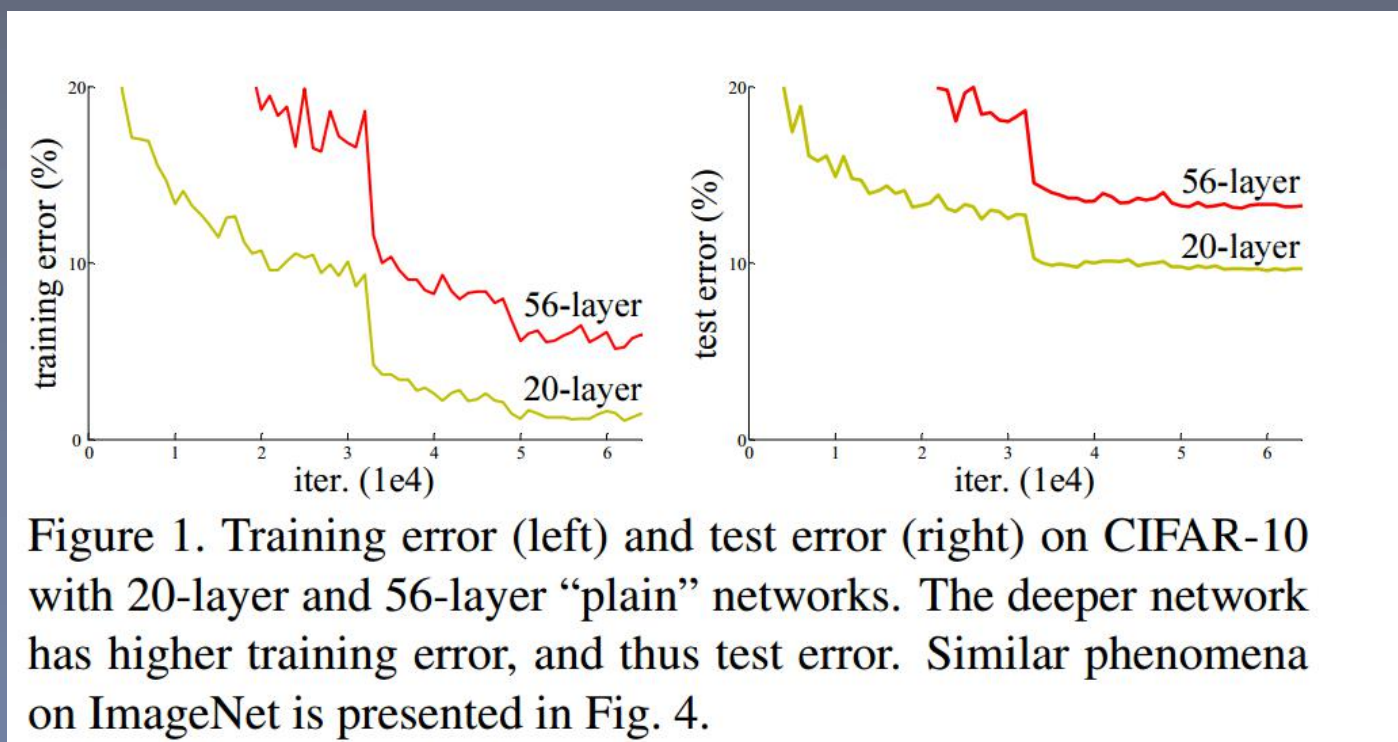
4.2 CIFAR-10 and Analysis

4.3. Object Detection on PASCAL and MS
COCO

论文图表

Figure & Table

图1. CIFAR-10上对比浅层网络和深层网络的精度(plain network)

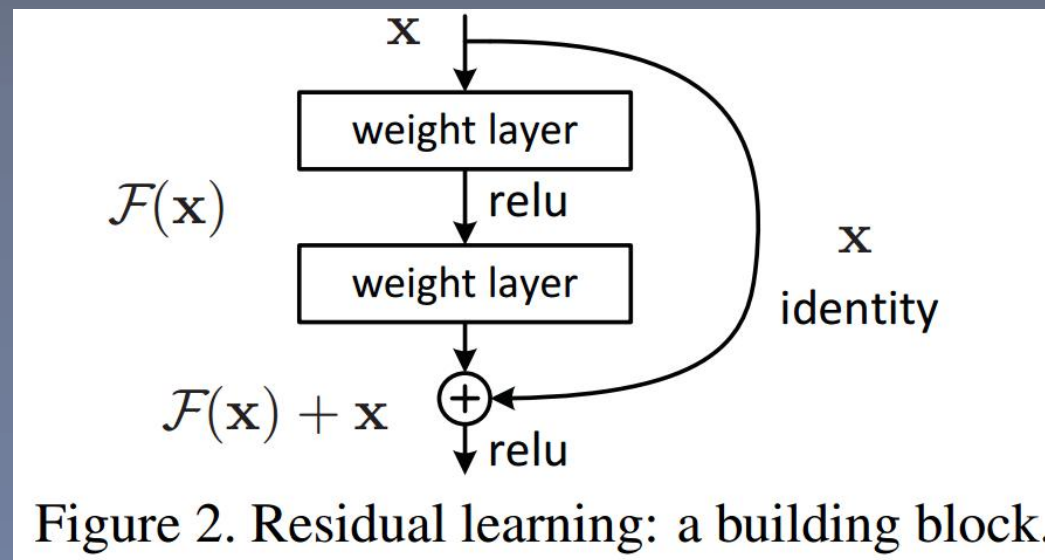


论文图表

Figure & Table

图1. CIFAR-10上对比浅层网络和深层网络的精度 (plain network)

图2. 残差学习模块





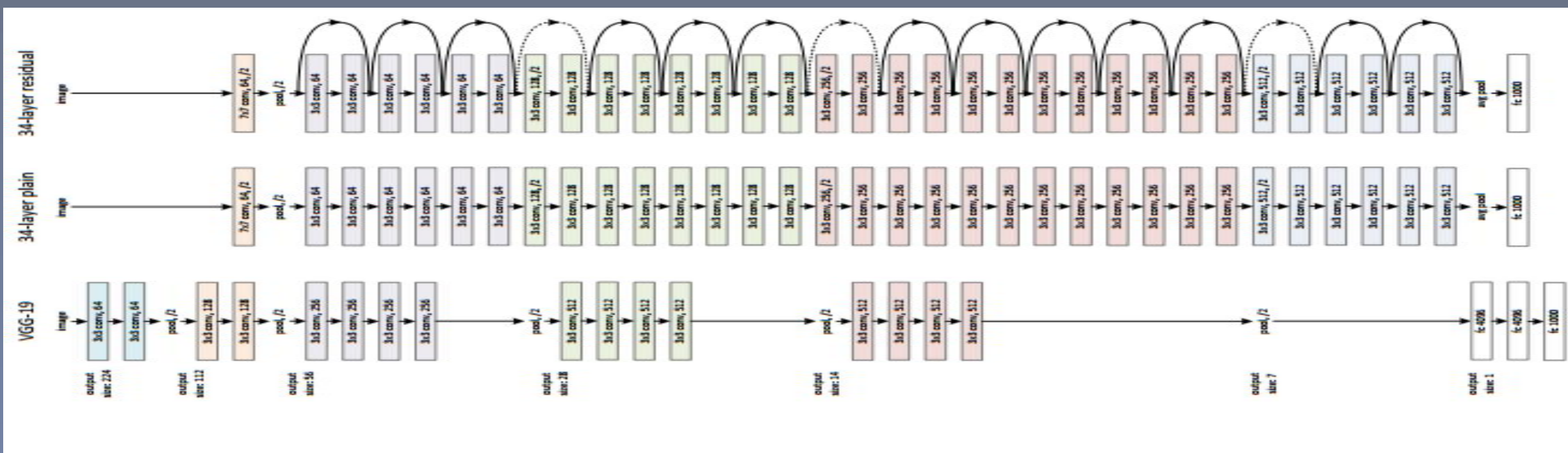
论文图表

Figure & Table

图1. CIFAR-10上对比浅层网络和深层网络的精度 (plain network)

图2. 残差学习模块

图3. VGG-19与ResNet-34的对比, 为说明ResNet更深但参数少



论文图表

Figure & Table

图1. CIFAR-10上对比浅层网络和深层网络的精度 (plain network)

图2. 残差学习模块

图3. VGG-19与ResNet-34的对比, 为说明ResNet更深但参数少

图4. plain network与resnet的对比, resnet完美解决plain存在的问题, 即深层网路性能比浅层网络差

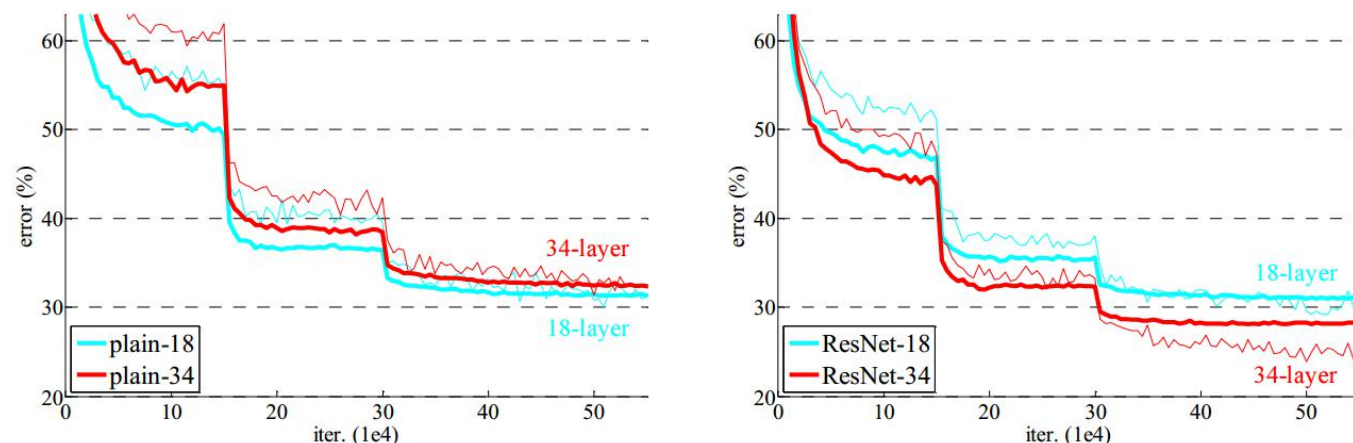


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.

论文图表

Figure & Table

图1. CIFAR-10上对比浅层网络和深层网络的精度(plain network)

图2. 残差学习模块

图3. VGG-19与ResNet-34的对比, 为说明ResNet更深但参数少

图4. plain network与resnet的对比, resnet完美解决plain存在的问题, 即深层网路性能比浅层网络差

图5. 深层的ResNet, 其residual block进行相应改变

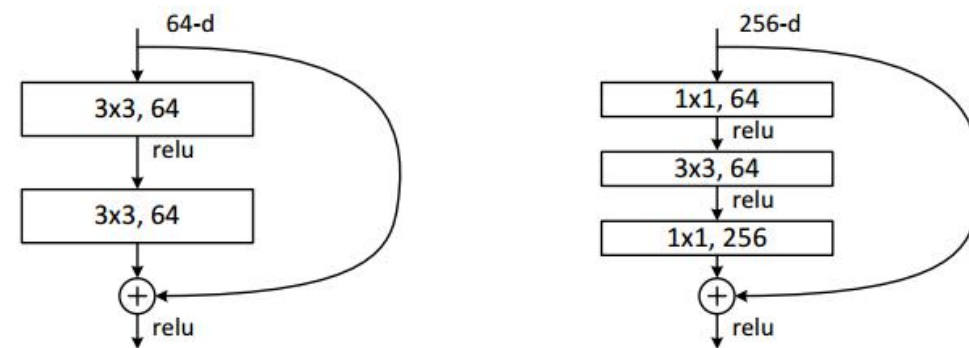


Figure 5. A deeper residual function \mathcal{F} for ImageNet. Left: a building block (on 56×56 feature maps) as in Fig. 3 for ResNet-34. Right: a “bottleneck” building block for ResNet-50/101/152.

论文图表

Figure & Table

图1. CIFAR-10上对比浅层网络和深层网络的精度(plain network)

图2. 残差学习模块

图3. VGG-19与ResNet-34的对比, 为说明ResNet更深但参数少

图4. plain network与resnet的对比, resnet完美解决plain存在的问题, 即深层网路性能比浅层网络差

图5. 深层的ResNet, 其residual block进行相应改变

图6. CIFAR-10上plain
和resnet随着层数加深
, error的变化曲线

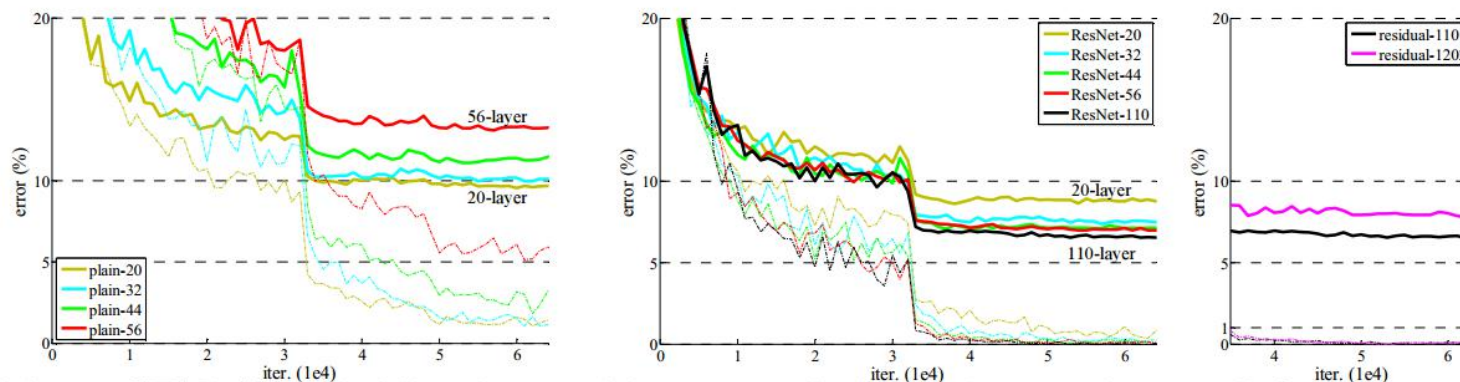


Figure 6. Training on **CIFAR-10**. Dashed lines denote training error, and bold lines denote testing error. **Left:** plain networks. The error of plain-110 is higher than 60% and not displayed. **Middle:** ResNets. **Right:** ResNets with 110 and 1202 layers.

论文图表

Figure & Table

图1. CIFAR-10上对比浅层网络和深层网络的精度(plain network)

图2. 残差学习模块

图3. VGG-19与ResNet-34的对比, 为说明ResNet更深但参数少

图4. plain network与resnet的对比

图5. 深层的ResNet, 其residual block进行相应改变

图6. CIFAR-10上plain和resnet随着层数加深, error的变化

图7. 神经元响应值的标准差变化, 对比plain和resnet, 发现resnet的响应值更低, 印证3.1小节提出的残差结构的输出比普通结构更倾向于0

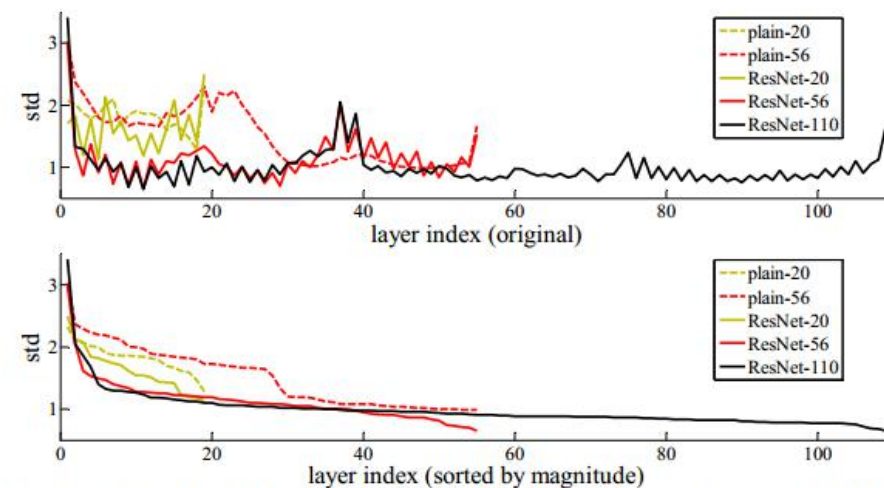


Figure 7. Standard deviations (std) of layer responses on CIFAR-10. The responses are the outputs of each 3×3 layer, after BN and before nonlinearity. **Top:** the layers are shown in their original order. **Bottom:** the responses are ranked in descending order.

论文图表

Figure & Table



深度之眼
deepshare.net

表1. resnet-18/34/50/101/152网络结构示意图

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				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	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

Table 1. Architectures for ImageNet. Building blocks are shown in brackets (see also Fig. 5), with the numbers of blocks stacked. Down-sampling is performed by conv3_1, conv4_1, and conv5_1 with a stride of 2.

论文图表

Figure & Table

表1. resnet-18/34/50/101/152网络结构示意

表2. ILSVRC top1 error对比

	plain	ResNet
18 layers	27.94	27.88
34 layers	28.54	25.03

Table 2. Top-1 error (% , 10-crop testing) on ImageNet validation. Here the ResNets have no extra parameter compared to their plain counterparts. Fig. 4 shows the training procedures.

论文图表

Figure & Table

表1. resnet-18/34/50/101/152网络结构示意图

表2. ILSVRC top1 error对比

表3.横向比较各个模型，以及ABC三种策略，10-crop的指标

model	top-1 err.	top-5 err.
VGG-16 [41]	28.07	9.33
GoogLeNet [44]	-	9.15
PReLU-net [13]	24.27	7.38
plain-34	28.54	10.02
ResNet-34 A	25.03	7.76
ResNet-34 B	24.52	7.46
ResNet-34 C	24.19	7.40
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	21.43	5.71

Table 3. Error rates (% , **10-crop** testing) on ImageNet validation. VGG-16 is based on our test. ResNet-50/101/152 are of option B that only uses projections for increasing dimensions.

论文图表

Figure & Table

表1. resnet-18/34/50/101/152网络结构示意图

表2. ILSVRC top1 error对比

表3. 横向比较各个模型，以及ABC三种策略，10-crop的指标

表4. 单模型横向对比

表5. 多模型融合的横向对比

method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43 [†]
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PreLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	19.38	4.49

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except [†] reported on the test set).

method	top-5 err. (test)
VGG [41] (ILSVRC'14)	7.32
GoogLeNet [44] (ILSVRC'14)	6.66
VGG [41] (v5)	6.8
PreLU-net [13]	4.94
BN-inception [16]	4.82
ResNet (ILSVRC'15)	3.57

Table 5. Error rates (%) of **ensembles**. The top-5 error is on the test set of ImageNet and reported by the test server.

论文图表

Figure & Table

表1. resnet-18/34/50/101/152网络结构示意图

表2. ILSVRC top1 error对比

表3. 横向比较各个模型，以及ABC三种策略，10-crop的指标

表4. 单模型横向对比

表5. 多模型融合的横向对比

表6. CIFAR-10上的实验对比

method			error (%)
Maxout [10]			9.38
NIN [25]			8.81
DSN [24]			8.22
	# layers	# params	
FitNet [35]	19	2.5M	8.39
Highway [42, 43]	19	2.3M	7.54 (7.72±0.16)
Highway [42, 43]	32	1.25M	8.80
ResNet	20	0.27M	8.75
ResNet	32	0.46M	7.51
ResNet	44	0.66M	7.17
ResNet	56	0.85M	6.97
ResNet	110	1.7M	6.43 (6.61±0.16)
ResNet	1202	19.4M	7.93

Table 6. Classification error on the **CIFAR-10** test set. All methods are with data augmentation. For ResNet-110, we run it 5 times and show “best (mean±std)” as in [43].

论文图表

Figure & Table

表1. resnet-18/34/50/101/152网络结构示意图

表2. ILSVRC top1 error对比

表3. 横向比较各个模型，以及ABC三种策略，10-crop的指标

表4. 单模型横向对比

表5. 多模型融合的横向对比

表6. CIFAR-10上的实验对比

表7/8. 利用resnet-101替换Faster RCNN中的vgg-16的结果对比

training data	07+12	07++12
test data	VOC 07 test	VOC 12 test
VGG-16	73.2	70.4
ResNet-101	76.4	73.8

Table 7. Object detection mAP (%) on the PASCAL VOC 2007/2012 test sets using **baseline** Faster R-CNN. See also Table 10 and 11 for better results.

metric	mAP@.5	mAP@[.5, .95]
VGG-16	41.5	21.2
ResNet-101	48.4	27.2

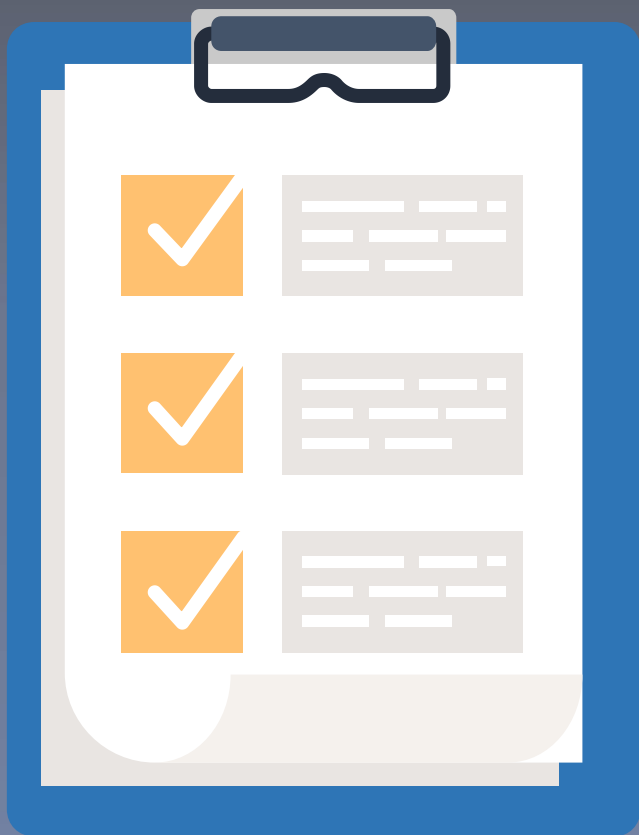
Table 8. Object detection mAP (%) on the COCO validation set using **baseline** Faster R-CNN. See also Table 9 for better results.

本课回顾及下节预告

Review in the lesson and Preview of next lesson

本课回顾

Review in the lesson



01 ResNet研究背景

ILSVRC-2015, Highway Networks, VGG

02 VGG研究成果及意义

ILSVRC-2015 分类、定位、目标检测冠军, COCO 分割、目标检测冠军
加快CNN落地, 成为工业界应用最广的CNN, 近代CNN发展史里程碑

03 论文摘要

提出问题, 本文方法, 本文优点, 本文工作, 本文成果

04 论文图表

对比plain结构和ResNet结构, 精度随网络层数变化情况
横向、纵向比较ResNet系列与其它模型的精度

下节预告

Preview of next lesson



01 残差结构

引入模型退化问题，提出Residual Learning的方式有效解决
残差结构有三种探讨，A、B和C

02 ResNet网络结构

ResNet-18, 34, 50, 101, 152等，设计原则借鉴VGG, 3*3卷积堆叠，
building block堆叠

03 实验结果分析

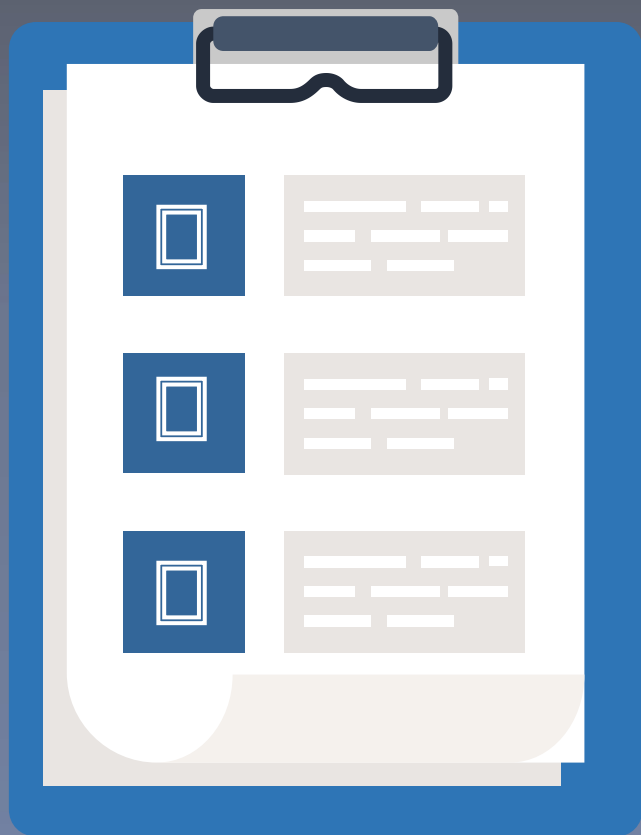
训练预热，采用较小学习率预热模型；横纵对比ResNet-
34/50/101/152, 表明ResNet-152效果最佳，优于现有所有模型

04 论文总结

总结论文中创新点、关键点及启发点

下节课前准备

Preview of next lesson



- 下载论文
- 泛读论文
- 筛选出自己不懂的部分，带着问题进入下一课时

——结 语——

读书有三到，谓心到，眼到，口到
朱熹





深度之眼
deepshare.net

联系我们：

电话：18001992849

邮箱：service@deepshare.net

Q Q： 2677693114



公众号



客服微信

