深度之眼
deepshare.net

Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning

导师：余老师

Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning

Inception-v4, Inception-ResNet，残差连接对模型训练的影响

作者：Christian Szegedy , Sergey Ioffe，Vincent Vanhoucke

单位：Google Inc.

发表会议及时间：AAAI 2017

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

GooLeNet系列

了解GoogLeNet-V1，V2，V3的发展以及ResNet

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

- 学习目标
  - Inception-v4
    - 六大模块
      - Stem
      - Inception-A\B\C
      - Reduction-A\B
  - Inception-ResNet
    - V1 六大模块
      - Stem
      - Inception-A\B\C
      - Reduction-A\B
    - V2 六大模块
      - Stem
      - Inception-A\B\C
      - Reduction-A\B
  - Scaling of the Residuals
    - 缩小激活值，以稳定模型的训练

课程安排

The schedule of course

深度之眼
deepshare.net

01 **Part 1**
背景成果介绍

02 **Part2**
论文泛读

03 **Part 3**
Inception-v4
Inception-ResNet-V1、V2

04 **Part 4**
激活值缩放

05 **Part 5**
论文总结

06 **Part 6**
代码实现分析

深度之眼
deepshare.net

第一课：论文导读

The first lesson: the paper guide

深度之眼
deepshare.net

目 录

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、成果和意义**
  - **背景**
    - ILSVRC
    - 相关研究
      - Residual Connection
      - GoogLeNet-v1\v2\v3
  - **成果**
    - ILSVRC 超越ResNet-152
    - 将Inception和ResNet结合
  - **意义**
    - 提供模型创新思路，强强联手A+B=AB

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

深度之眼
deepshare.net

相关研究

1. GoogLeNet-V1 （Inception-V1）
2. GoogLeNet-V2 （Batch Normalization）
3. GoogLeNet-V3（Inception-V2 and Inception-V3）
4. ResNet

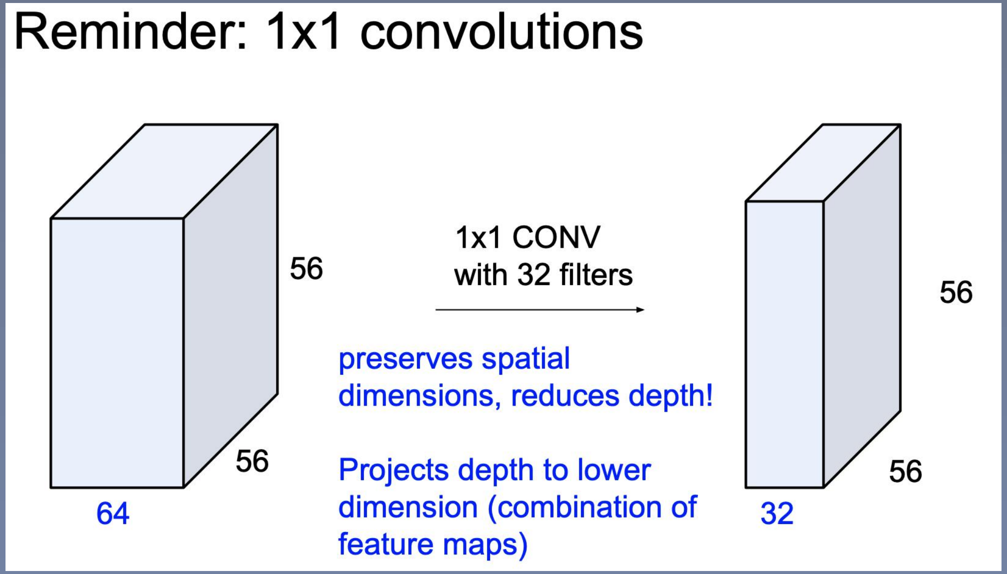
> duced in [14] and was called GoogLeNet or Inception-v1 in our exposition. Later the Inception architecture was refined in various ways, first by the introduction of batch normalization [6] (Inception-v2) by Ioffe et al. Later the architecture was improved by additional factorization ideas in the third iteration [15] which will be referred to as Inception-v3 in this report.

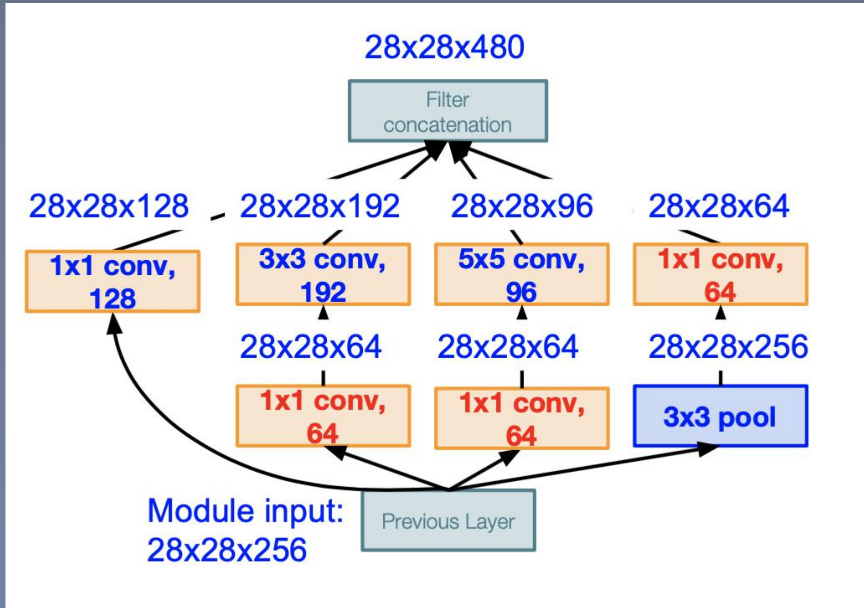
研究背景

Research background

相关研究

1. GoogLeNet-V1 （Inception-V1）：1*1卷积，Inception模块





Inception module with dimension reduction

研究背景

Research background

相关研究

2. GoogLeNet-V2（Batch Normalization）：

- 激活函数前加入BN
- 5*5卷积替换为2个3*3卷积
- 第一个Inception模块增加一个Inception结构
- 增多“5*5”卷积核
- 尺寸变化采用stride=2的卷积
- 增加9层（10-1层）到 31层
- （10表示inception数量）

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1\ldots m}\}$;
Parameters to be learned: γ, β

Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch.

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	double #3×3 reduce	double #3×3	Pool +proj
convolution*	7×7/2	112×112×64	1						
max pool	3×3/2	56×56×64	0						
convolution	3×3/1	56×56×192	1		64	192			
max pool	3×3/2	28×28×192	0						
inception (3a)		28×28×256	3	64	64	64	64	96	avg + 32
inception (3b)		28×28×320	3	64	64	96	64	96	avg + 64
inception (3c)	stride 2	28×28×576	3	0	128	160	64	96	max + pass through
inception (4a)		14×14×576	3	224	64	96	96	128	avg + 128
inception (4b)		14×14×576	3	192	96	128	96	128	avg + 128
inception (4c)		14×14×576	3	160	128	160	128	160	avg + 128
inception (4d)		14×14×576	3	96	128	192	160	192	avg + 128
inception (4e)	stride 2	14×14×1024	3	0	128	192	192	256	max + pass through
inception (5a)		7×7×1024	3	352	192	320	160	224	avg + 128
inception (5b)		7×7×1024	3	352	192	320	192	224	max + 128
avg pool	7×7/1	1×1×1024	0						

Figure 5: Inception architecture

研究背景

Research background

相关研究

3. GoogLeNet-V3（Inception-V2 and Inception-V3）

- 四个模型设计准则
- 两种卷积分解方式
- 特征图下降策略

Inception-V2结构示意→

type	patch size/stride or remarks	input size
conv	3×3/2	299×299×3
conv	3×3/1	149×149×32
conv padded	3×3/1	147×147×32
pool	3×3/2	147×147×64
conv	3×3/1	73×73×64
conv	3×3/2	71×71×80
conv	3×3/1	35×35×192
3×Inception	As in figure 5	35×35×288
5×Inception	As in figure 6	17×17×768
2×Inception	As in figure 7	8×8×1280
pool	8 × 8	8 × 8 × 2048
linear	logits	1 × 1 × 2048
softmax	classifier	1 × 1 × 1000

Network	Top-1 Error	Top-5 Error	Cost Bn Ops
GoogLeNet [20]	29%	9.2%	1.5
BN-GoogLeNet	26.8%	-	**1.5**
BN-Inception [7]	25.2%	7.8	2.0
Inception-v2	23.4%	-	3.8
Inception-v2 RMSProp	23.1%	6.3	3.8
Inception-v2 Label Smoothing	22.8%	6.1	3.8
Inception-v2 Factorized 7 × 7	21.6%	5.8	4.8
Inception-v2 BN-auxiliary	**21.2%**	**5.6%**	4.8

Table 3. Single crop experimental results comparing the cumulative effects on the various contributing factors. We compare our numbers with the best published single-crop inference for Ioffe at al [7]. For the "Inception-v2" lines, the changes are cumulative and each subsequent line includes the new change in addition to the previous ones. The last line is referring to all the changes is what we refer to as "Inception-v3" below. Unfortunately, He et al [6] reports the only 10-crop evaluation results, but not single crop results, which is reported in the Table 4 below.

研究背景

Research background

相关研究

4. ResNet

引入残差学习，成功训练超千层卷积神经网络，成为工业界标杆模型

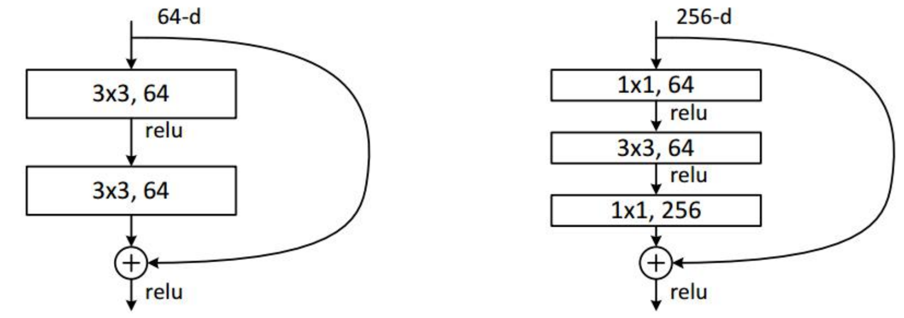


Figure 5. A deeper residual function $\mathcal{F}$ for ImageNet. Left: a building block (on 56×56 feature maps) as in Fig. 3 for ResNet-34. Right: a "bottleneck" building block for ResNet-50/101/152.

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				
		[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×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

研究成果

Research Results

深度之眼
deepshare.net

ILSVRC 3.1% 的top-5 error

模型	时间	top-5
AlexNet	2012	15.3%
ZFNet	2013	13.5%
VGG	2014	7.3%
GoogLeNet	2014	6.6%
GoogLeNet-V2	2015	4.9%
GoogLeNet-V3	2015	3.6%
ResNet-152	2015	3.6%
GoogLeNet-V4	2016	3.1%

研究意义

Research Meaning

研究意义

1. 将当下（2015年-2016年）优秀的技术结合到Inception系列中，获得更优模型，其借鉴新技术的思想值得学习，即 A+B = AB的创新思路

论文泛读

Structure of Paper

论文结构

Structure of Papers

摘要 Abstract

简介ResNet和Inception系列很强，考虑将两者进行集合，获得更优的模型Inception-ResNet

1 Introduction

残差学习有助于深度网络的训练，inception是很深的网络，适合把残差学习技术加进来

2 Related Work

简介resnet思路，并提出质疑resnet，同时给出结论：没有残差结构一样可以训练深度网络

3 Architectural Choices

Inception-V4
Inception-ResNet-V1/V2
Scaling of the Residuals

4 Training Methodology

训练配置

5 Experimental Results

实验结果对比，ILSVRC上比较ResNet，Inception-3，Inception-v4， Inception-ResNet

6 Conclusions

总结本文提出的三个网络模型的内容：Inception-V4
Inception-ResNet-V1/V2

摘要

abstract

摘要核心

1. 研究背景1：近年，深度卷积神经网络给图像识别带来巨大提升，例如Inception（王婆卖瓜）

2. 研究背景2：最近，残差连接的使用使卷积神经网络得到了巨大提升，如ILSVRC-2015冠军——ResNet，它（ResNet）与Inception-v3的精度差不多（继续卖瓜）

3. 提出问题：是否可以将Inception 与 残差连接结合起来，提高卷积神经网络呢？

4. 本文成果1：从实验经验得出，残差连接很大程度的加速了Inception的训练；提出了新的网络模型结构——streamlined architectures

5. 本文成果2：对于很宽的residual Incaption网络，提出激活值缩放策略，以使网络训练稳定

6. 本文成果3：模型融合在ILSVRC上获得3.08%的top-5 error

原文讲解

论文小标题

Paper title

1. Introduction
2. Related Work
3. Architectural Choices
   3.1. Pure Inception blocks
   3.2. Residual Inception Blocks
   3.3. Scaling of the Residuals
4. Training Methodology
5. Experimental Results
6. Conclusions

论文图表

Figure & Table

图1和图2. residual connection对比，提出1*1卷积用于较少参数量的策略（resent-50/101/152就已经用了1*1卷积了呢，只不过redisual blocks中采用3个卷积，为什么这里不提呢？）

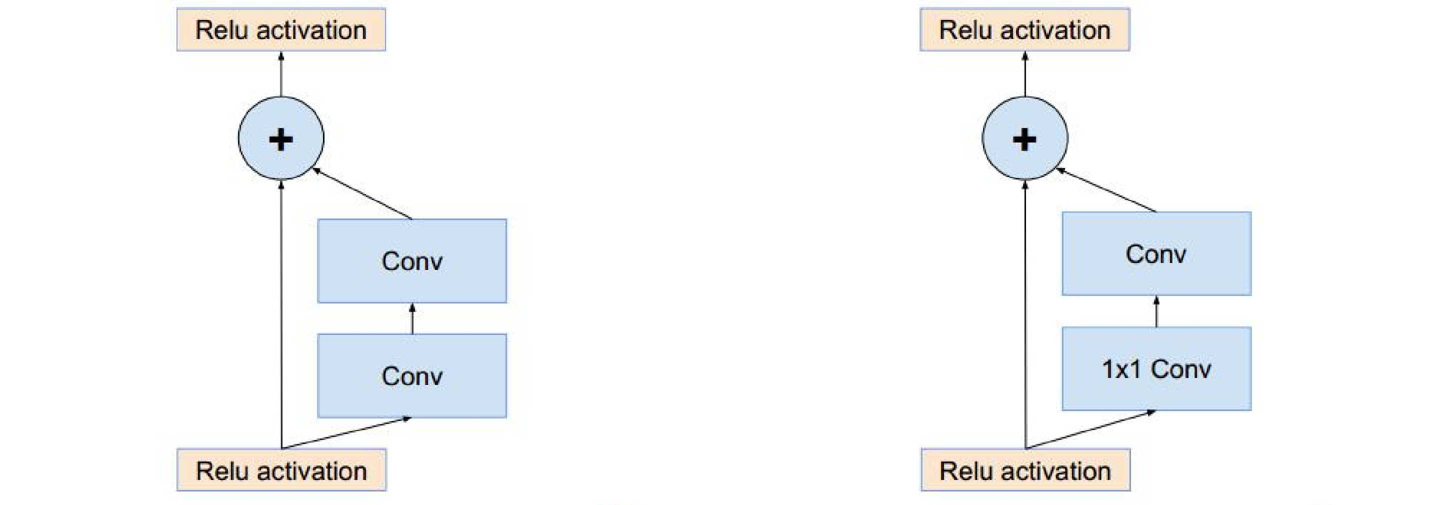


Figure 1. Residual connections as introduced in He et al. [5].

Figure 2. Optimized version of ResNet connections by [5] to shield computation.

论文图表

Figure & Table

图9. Inception-V4框架图

主要有六大模块:

1. Stem
2. inceptionA
3. inceptionB
4. inceptionC
5. reduction-A
6. reduction-B

分别对应图3-8

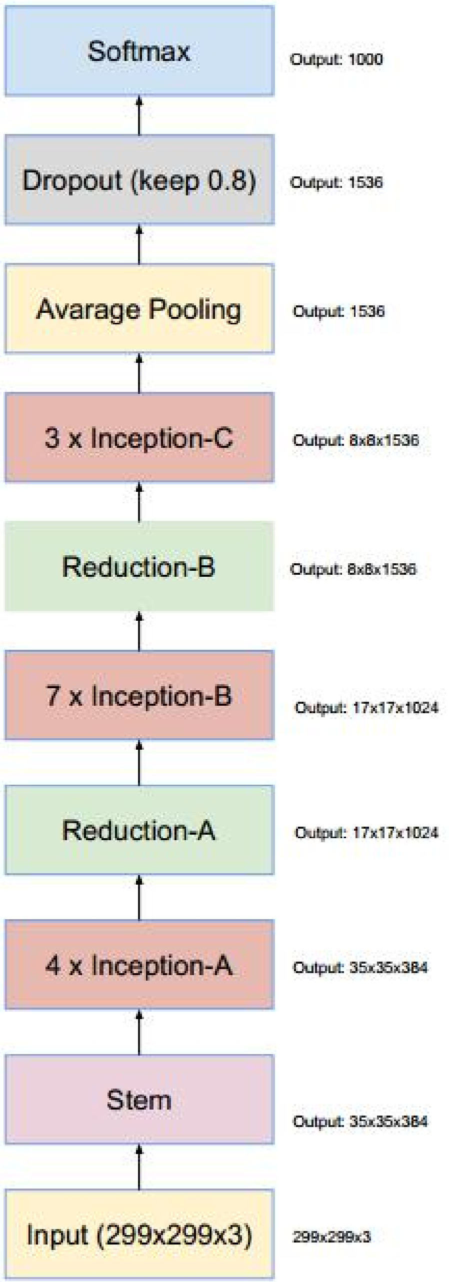


Figure 9. The overall schema of the Inception-v4 network. For the detailed modules, please refer to Figures 3, 4, 5, 6, 7 and 8 for the detailed structure of the various components.

论文图表

Figure & Table

图15. Inception-ResNet-V1、V2框架图

Inception-ResNet-V1的六大模块分别对应：

图14， 图10，图11， 图13，图7（与Inception-V4同模块），图12

Inception-ResNet-V2的六大模块分别对应：

图3（与Inception-v4同模块），图16， 图17， 图19，图7（与Inception-V4同模块），图18

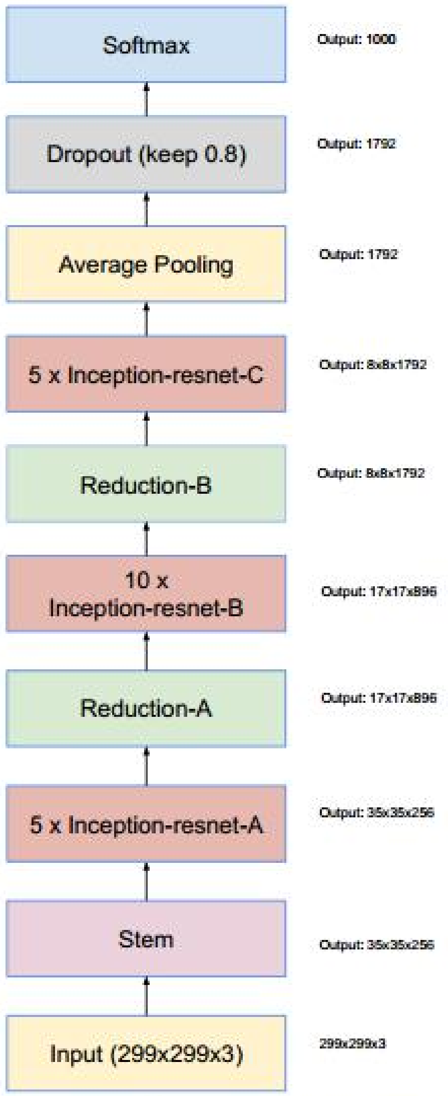


Figure 15. Schema for Inception-ResNet-v1 and Inception-ResNet-v2 networks. This schema applies to both networks but the underlying components differ. Inception-ResNet-v1 uses the blocks as described in Figures 14, 10, 7, 11, 12 and 13. Inception-ResNet-v2 uses the blocks as described in Figures 3, 16, 7,17, 18 and 19. The output sizes in the diagram refer to the activation vector tensor shapes of Inception-ResNet-v1.

论文图表

Figure & Table

图20. Scaling activation示意图

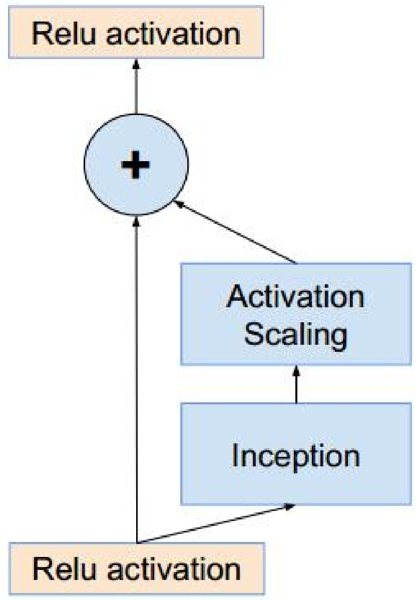
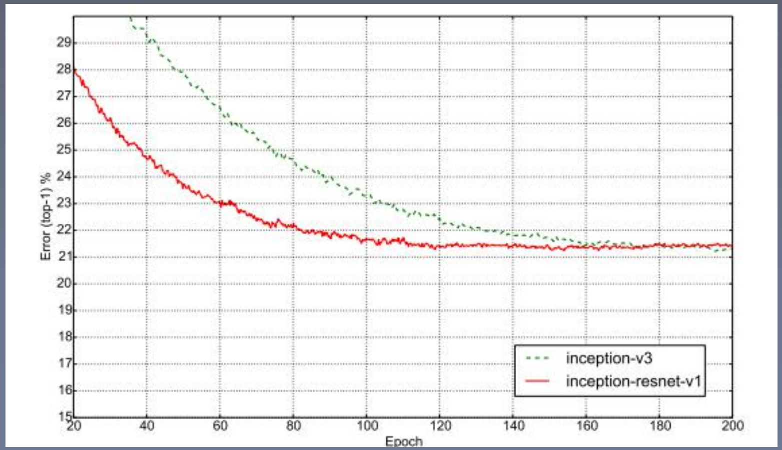


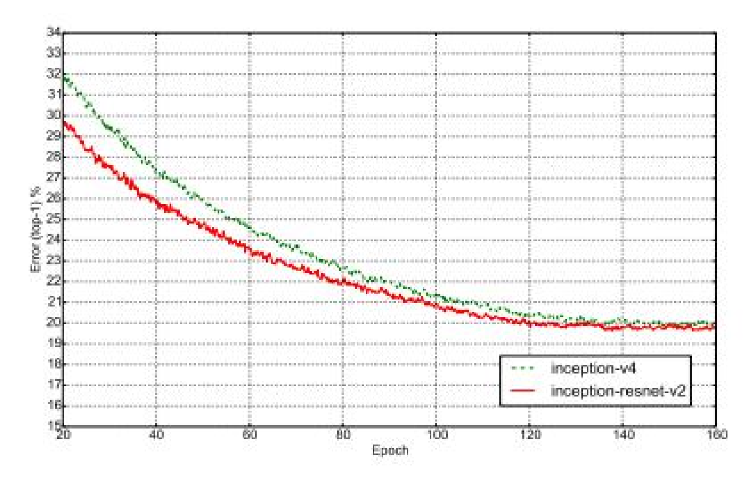
Figure 20. The general schema for scaling combined Inception-resnet moduels. We expect that the same idea is useful in the general resnet case, where instead of the Inception block an arbitrary subnetwork is used. The scaling block just scales the last linear activations by a suitable constant, typically around 0.1.

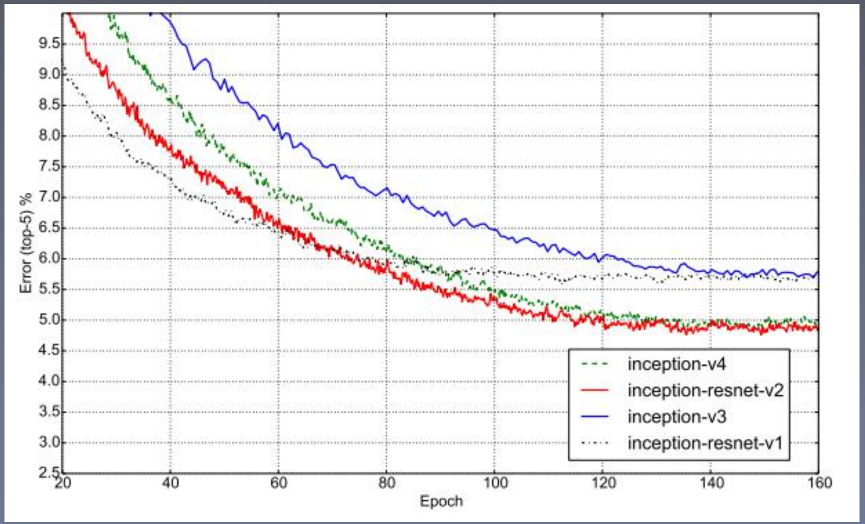
论文图表

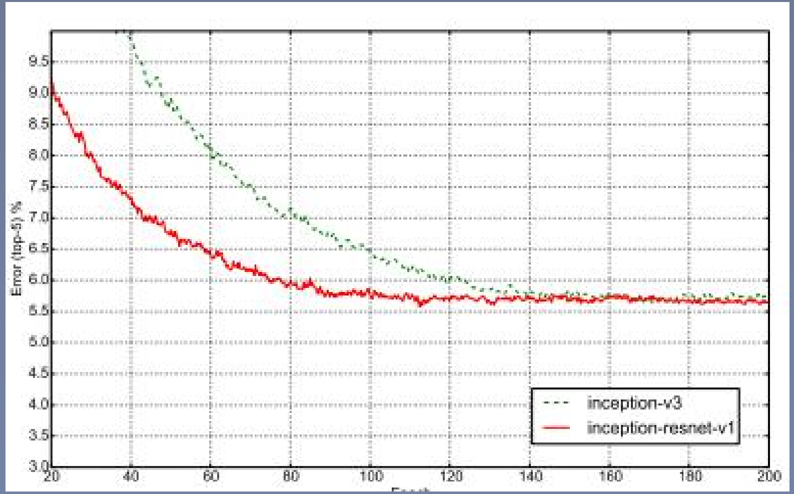
Figure & Table

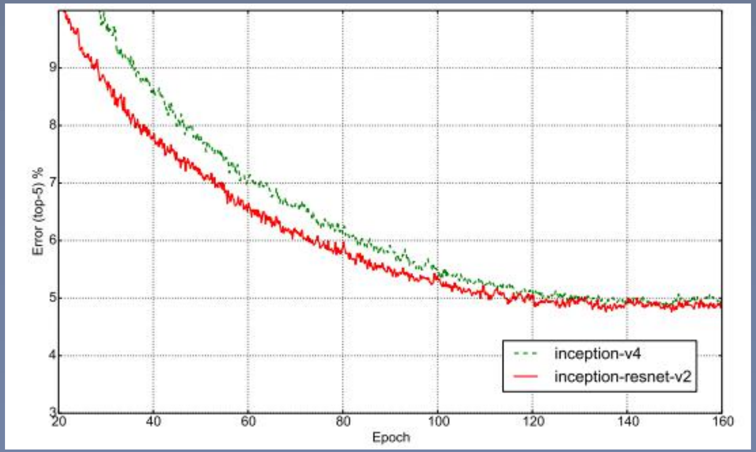
图21-26. 四个模型的单模型single crop下的曲线对比

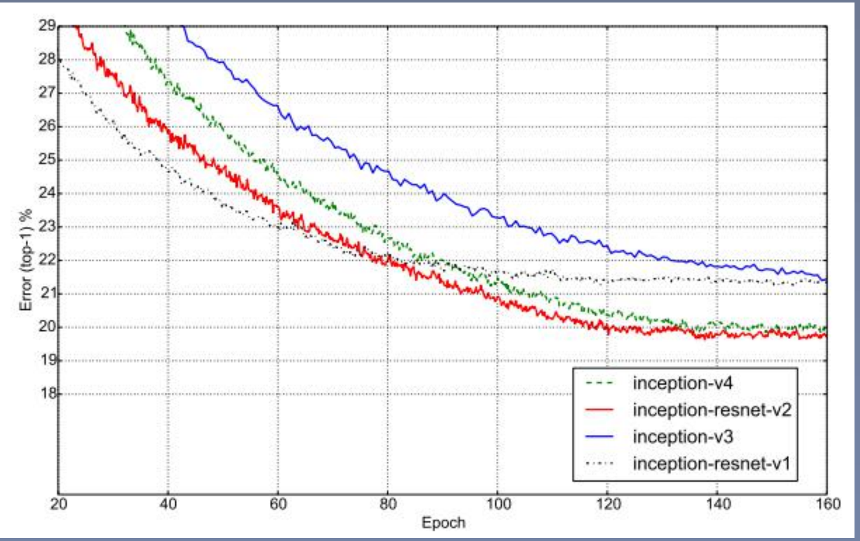












论文图表

Figure & Table

表1. 三个模型的Reduction-A模块的参数设置

Network	k	l	m	n
Inception-v4	192	224	256	384
Inception-ResNet-v1	192	192	256	384
Inception-ResNet-v2	256	256	384	384

Table 1. The number of filters of the Reduction-A module for the three Inception variants presented in this paper. The four numbers in the colums of the paper parametrize the four convolutions of Figure 7

论文图表

Figure & Table

表1. 三个模型的Reduction-A模块的参数设置

表3,4,5. 四个模型不同测试方法的精度对比

Network	Crops	Top-1 Error	Top-5 Error
ResNet-151 [5]	10	21.4%	5.7%
Inception-v3 [15]	12	19.8%	4.6%
Inception-ResNet-v1	12	19.8%	4.6%
Inception-v4	12	18.7%	4.2%
Inception-ResNet-v2	12	18.7%	4.1%

Table 3. 10/12 crops evaluations - single model experimental results. Reported on the all 50000 images of the validation set of ILSVRC 2012.

Network	Crops	Top-1 Error	Top-5 Error
ResNet-151 [5]	dense	19.4%	4.5%
Inception-v3 [15]	144	18.9%	4.3%
Inception-ResNet-v1	144	18.8%	4.3%
Inception-v4	144	17.7%	3.8%
Inception-ResNet-v2	144	17.8%	3.7%

Table 4. 144 crops evaluations - single model experimental results. Reported on the all 50000 images of the validation set of ILSVRC 2012.

Network	Models	Top-1 Error	Top-5 Error
ResNet-151 [5]	6	–	3.6%
Inception-v3 [15]	4	17.3%	3.6%
Inception-v4 + 3× Inception-ResNet-v2	4	16.5%	3.1%

Table 5. Ensemble results with 144 crops/dense evaluation. Reported on the all 50000 images of the validation set of ILSVRC 2012. For Inception-v4(+Residual), the ensemble consists of one pure Inception-v4 and three Inception-ResNet-v2 models and were evaluated both on the validation and on the test-set. The test-set performance was 3.08% top-5 error verifying that we don't overfit on the validation set.

深度之眼
deepshare.net

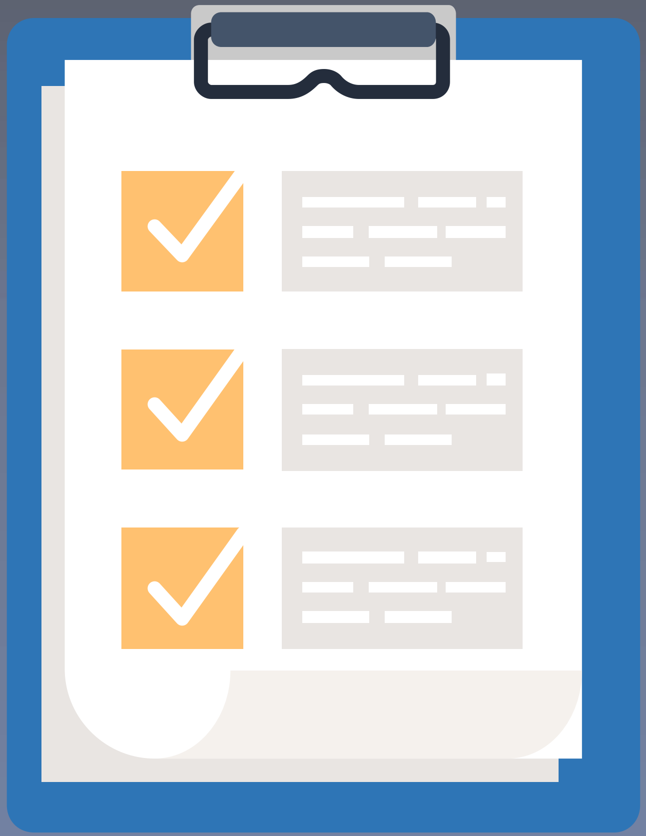
本课回顾及下节预告

Review in the lesson and Preview of next lesson

深度之眼
deepshare.net

本课回顾

Review in the lesson



01 Inception-v4 研究背景

Inception-v1、Inception-v2、Inception-v3和ResNet

02 GoogLenet-v4研究成果及意义

超越ResNet，Inception-V3，top-5 error达3.1%

03 论文摘要

研究背景1，研究背景2，提出问题，本文成果1，本文成果2，本文成果3

04 论文图表

18张图展示3个模型具体的结构，4个模型（Inception-v3、ResNet，Inception-v4和Inception-ResNet）的精度对比

深度之眼
deepshare.net

下节预告

Preview of next lesson

01 Inception-v4

Stem、Inception-A\B\C、Reduction-A\B 六大模块，
9+3*4+5*7+4*3+3+3+1 = 75层

02 Inception-ResNet网络结构

Inception-ResNet-V1共 7+5*4+3+10*4+3+5*4+1=94层

Inception-ResNet-V2共 9+5*4+3+10*4+3+5*4+1=96层

03 实验结果分析

Single crop、10-crop、144-crop及模型融合方式的top1及top5 error对比

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

读书之法,在循序而渐进,熟读而精思

——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