

[译] Deep Residual Learning for Image Recognition (ResNet)



zhwhong (/u/38cd2a8c425e) [+ 关注](#)

2017.01.15 17:30* 字数 9411 阅读 9264 评论 0 喜欢 20 赞赏 1

(/u/38cd2a8c425e)

题目：图像识别领域的深度残差学习

- 文章地址：《Deep Residual Learning for Image Recognition》
(<https://link.jianshu.com?t=https://arxiv.org/abs/1512.03385>) arXiv.1512.03385
- ResNet Github参考：<https://github.com/tornadomeet/ResNet>
(<https://link.jianshu.com?t=https://github.com/tornadomeet/ResNet>)

(转载请注明出处：<http://www.jianshu.com/p/f71ba99157c7>
(<https://www.jianshu.com/p/f71ba99157c7>)，谢谢！)

Abstract

摘要：更深的神经网络往往更难以训练，我们在此提出一个残差学习的框架，以减轻网络的训练负担，这是个比以往的网路要深的多的网络。我们明确地将层作为输入学习残差函数，而不是学习未知的函数。我们提供了非常全面的实验数据来证明，残差网络更容易被优化，并且可以在深度增加的情况下让精度也增加。在ImageNet的数据集上我们评测了一个深度152层（是VGG的8倍）的残差网络，但依旧拥有比VGG更低的复杂度。残差网络整体达成了3.57%的错误率，这个结果获得了ILSVRC2015的分类任务第一名，我们还用CIFAR-10数据集分析了100层和1000层的网络。

在一些计算机视觉识别方向的任务当中，深度表示往往是重点。我们极深的网络让我们得到了28%的相对提升（对COCO的对象检测数据集）。我们在深度残差网络的基础上做了提交的版本参加ILSVRC和COCO2015的比赛，我们还获得了ImageNet对象检测，Imagenet对象定位，COCO对象检测和COCO图像分割的第一名。

一.简介

深度卷积神经网络使得图像分类问题上的研究向前飞跃了一大步，深度网络自然的整合了低中高不同层次的特征，并且使用端到端的多层次分类，特征的“层次”可以靠加深网络层数来丰富。最近的研究揭示了网络深度是非常重要的关键点。有代表性的几个研究团队，在Imagnet中竞赛的都不约而同的使用了“超级深”的网络，从17到30层不等。其他一些计算机视觉的问题也受益于超级深的网络模型。



受到深度的意义的驱使，出现了这样一个问题：是不是更多的堆叠层就一定能学习出更好的网络？这个问题的一大障碍就是臭名昭著的梯度消失/爆炸问题，它从一开始就阻碍了收敛，然而梯度消失/爆炸的问题，很大程度上可以通过标准的初始化和正则化层来基本解决，确保几十层的网络能够收敛（用SGD+反向传播）。

(a
utn

然而当开始考虑更深层的网络的收敛问题时，退化问题就暴露了：随着神经网络深度的增加，精确度开始饱和（这是不足为奇的），然后会迅速的变差。出人意料的，这样一种退化，并不是过拟合导致的，并且增加更多的层匹配深度模型，会导致更多的训练误差，就像文章中说的，通过我们的实验将得到充分证实。图1展示了一个典型的例子。

Figure 1 CIFAR-10数据集，训练集误差（左），测试集误差（右）。20层/56层普通网络，越深的网络错误率越高，imagenet数据集也是这样。

训练精度的退化表明，不是所有的系统都同样容易优化。让我们考虑一个浅层架构和它的对应的增加了更多层的深层架构。存在一个解决方案来构建更深层次的模型：添加的层是自身映射，其他层是从训练好的浅模型中复制而来。这种特殊的构建方式让我们推测，深的模型应该不会比浅的模型产生更高的训练误差。但实验结果表明，我们手头上的方案都找不到解，找不到更好或者同样好的解（或者是无法在可接受的时间里做完）。

在本文中，我们通过引入一个深度残差学习框架，解决了这个退化问题。我们不期望每一层能直接吻合一个映射，我们明确的让这些层去吻合残差映射。形式上看，就是用 $H(X)$ 来表示最优解映射，但我们让堆叠的非线性层去拟合另一个映射 $F(X) := H(X) - X$ ，此时原最优解映射 $H(X)$ 就可以改写成 $F(X)+X$ ，我们假设残差映射跟原映射相比更容易被优化。极端情况下，如果一个映射是可优化的，那也会很容易将残差推至0，把残差推至0和把此映射逼近另一个非线性层相比要容易的多。

$F(X)+X$ 的公式可以通过在前馈网络中做一个“快捷连接”来实现（如图2），快捷连接跳过一个或多个层。在我们的用例中，快捷连接简单的执行自身映射，它们的输出被添加到叠加层的输出中。自身快捷连接既不会添加额外的参数也不会增加计算复杂度。整个网络依然可以用SGD+反向传播来做端到端的训练，并且可以很容易用大众框架来实现（比如Caffe）不用修改solver配置（solver是caffe中的核心solver.prototxt）



Figure 2 残差网络：一个结构块

我们目前用ImageNet的数据集做了很多综合实验，来证实退化问题和评估我们的方法。我们发现：1. 我们的超深残差网络是很容易去优化的，不过对应的普通网络（简单的堆叠层）当深度增加时，表现出更高的错误误差。2. 我们的深度残差网络可以轻松的享受深度增加带来的精度增加，产生的效果要远远优于以前的那些网络们。类似的现象在CIFAR-10数据集的实验中也一样，这表明着优化是困难的，我们提出的训练方法在这个数据集中，超过100层的网络表现很成功，还可以扩展到1000层。

在ImageNet对象分类数据集上，我们用深度残差网络获得了很棒的结果，我们152层的残差网络是ImageNet的参赛网络中最深的，然而却拥有比VGG更低的复杂度。我们最终的效果是测试集上3.57%的错误率，以此摘取了ILSVRC2015对象分类的第一名。这种超级深的表示方法在其他识别任务中也有良好的泛化能力，使我们进一步赢得了多个比赛的第一名（有ImageNet detection, Imagenet localization, COCOdetection COCOsegmentation），这般的有利的证据证明残差学习的原则是可泛化的，我们同样期望残差学习的方法能用在其他的视觉和非视觉问题上。

二. 相关工作

残差表示：在图像识别任务中，VLAD[18]是用基于词典的残差向量的来进行特征编码的，fisher向量可以看作VLAD的一个概率版本，它们在图像检索和浅层分类中都是挺不错的，对于矢量化，编码残差向量都被证明了比编码原始向量要更有效果。

在低级视觉和计算机图形学中，求解偏微分方程（PDE），通常是使用多重网格（Multigrid）法，把系统重建成多尺度的子问题，每个子问题负责求解出粗粒度和细粒度之间的残差，除此之外，另一种求解PDE的方法是级基预处理[45,46]，是基于表达两个尺度之间残差的向量进行的。在[3,44,45]中证明了这些用了残差的解法收敛速度都比不用残差的普通解法要快的多。这些研究表明，一个好的模型重构或者预处理手段是能简化优化过程的。

快捷连接：实践和理论引出了“快捷连接”这个想法，它已经被研究了很长的时间。在训练多层感知器网络（MLP）的早期实践，包括添加一个线性层（从网络的输入直连到输出[33,48]），在[43,24]中提到，少量中间层被直接连到附加的分类层解决梯度消失/爆炸问题，论文[38,37,31,46]中提出的层响应置中（centering layer responses）解决梯度和传播误差，也用到了快捷连接。在论文[43]中，一个“开始层”是由一个快捷分支和少量较深的分支构成。

和我们同期的工作也有一些，“Highway network”[41,42]提出的高速公路网络，展示了设置了门选通的快捷连接，这些门函数是数据相关的并且有参数要进行调整，对比而言，我们的自身快捷连接（恒等捷径）是没参数的。当一个门快捷连接呈关闭状态（接近0），highway network的层就代表着非残余函数，相反的，我们的方法总是学习残差方程。我们的自身快捷连接是永不关闭的，因此信息总能通过，与借此学习残差函数。此外highway network没有表现出精度随深度增加的特性（比如超过100层后）。

三.深度残差网络

3.1 残差学习

让我们考虑 $H(X)$ 是一个有若干堆叠的网络层将进行拟合的映射（不一定要整个网络）， X 表示这些层中第一层的输入。如果有一个假设：多层的非线性网络层可以逐渐逼近很复杂的函数，那么相当于可以假设它们同样能逼近残差函数。 $H(X) - X$ （假设输入和输出都有着相同的维度）。所以与其让这些层去逼近 $H(X)$ ，我们更期望让它们去逼近残差函数 $F(X) := H(X) - X$ 。对应的可以将原始的方程改成 $F(X) + X$ ，尽管这两种形式都应该可以逐步逼近目标函数（根据假设），但训练的简便程度也将大不相同。



这个重构的动机是出于对退化问题的反直觉现象（图1，左）。正如我们在介绍中讨论的，如果添加的层可以以恒等的方式被构造为自身映射，一个加深的模型的训练误差一定会不大于较浅的对应模型。退化问题表明，求解过程中在使多个非线性层逼近自身映射时有困难。而用残差的方法重构它，如果自身映射达到最佳的，则求解可能仅仅是更新多个非线性层的权值向零去接近自身映射。

在现实情况下，自身映射一开始就达到最优几乎是不可能的事，但我们的重构将有助于对此问题做预处理。如果优化的函数比起零映射更接近于自身映射的话，网络会更容易学习去确定自身映射的扰动参考，而不是将其作为一个全新的函数去学习。我们通过实验验证（图7），学习的残差函数一半都响应较小，这表明自身映射是更合理的预处理手段。

3.2用快捷连接实现自身映射

我们将残差学习的方式应用到了每一组堆叠层，一个构造块在图2所示，在本文中，我们把一个构造块定义成：

此处， \mathbf{x} 和 \mathbf{y} 分别表示构造块的输入和输出向量，函数 $\mathbf{F}(\mathbf{x}, \{\mathbf{W}_i\})$ 表示被训练的残差映射。举个例子，在图2 中有两层， $\mathbf{F}=\mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{x})$ 中的 σ ， σ 表示RELU，出于简化考虑省略了偏置项。操作 $\mathbf{F}+\mathbf{x}$ 是由一个快捷连接进行逐元素的添加得。我们在做加法后得到的模型具有二阶非线性。

公式1中介绍的这个快捷连接既没有引入额外的参数和也没有增加计算复杂性。这不仅是在应用中有吸引力，在我们对普通及残差网络的比较中也尤为重要。这样我们可以公平的比较参数个数、深度、宽度和计算代价完全一致的简单/残差网络（除了可以忽略不计的逐元素加法运算）。

公式1中 \mathbf{x} 的维度和 \mathbf{F} 必须保持一致，如果不一致（比如改变输入输出的通道数）我们可以在快捷连接上进行一个线性投影 \mathbf{W}_s 来匹配维度：

我们同样可以在公式1中用一个平方矩阵 \mathbf{W}_s ，不过我们的实验显示，自身映射足以解决退化问题，因此 \mathbf{W}_s 仅仅被用来匹配尺寸。

残差函数 \mathbf{F} 的形式是灵活的，本文的实验包括了 \mathbf{F} 为2层或3层的情况（图5），虽然更多的层也是可以的，但如果只有一个层（公式1）会等价于一个线性层， $\mathbf{y} = \mathbf{W}_1\mathbf{x} + \mathbf{x}$ ，这样一来就没有可见的优势了。

我们还注意到尽管上述的公式为了简便起见，都是关于完全连接层的，但是它们同样适用于卷积层。函数 $\mathbf{F}(\mathbf{x}, \{\mathbf{W}_i\})$ 可以代表多个卷积层。逐元素的加法运算则是两个特征图谱的加法，按照通道对应。

3.3网络结构

我们测试过非常多种普通/残差网络，并观察到一致的现象，为了提供讨论的实证，我们将在下文描述（用于ImageNet的）两个模型。

普通网络。我们的普通基准网络主要是受到VGG网络的启发，如图三左。卷积层的filter大多为3x3，遵循了两个设计原则：

1. 对于相同的尺寸的输出特征图谱，每层必须含有相同数量的过滤器。

2. 如果特征图谱的尺寸减半，则过滤器的数量必须翻倍，以保持每层的时间复杂度。

我们直接通过卷积层（stride=2）进行下采样，网络末端以全局的均值池化层结束，有1000路的全连接层（Softmax激活）。含有权重的网络层的总计为34层（见图3中）。

值得注意的是，我们的模型包含了更少的过滤器和比VGG更低的复杂度，我们的34层基本计算量为3.6亿FLOPS（包括乘法和加法），这仅仅是VGG（196亿FLOPs）的18%。

Figure 3 网络结构。左VGG19（19.6亿），中普通34层（3.6亿），右残差34层（3.6亿）。

残差网络。在简单网络的基础上，我们插入了快捷连接（图3，右），将网络转化为其对应的残差版本。当输入输出是相同尺寸的时候，自身捷径（公式（1））。当输入输出尺寸发生增加时（图3中的虚线的快捷连接），我们考虑两个策略：

- （a）快捷连接仍然使用自身映射，对于维度的增加用零来填补空缺。此策略不会引入额外的参数；
- （b）投影捷径（公式2）被用来匹配尺寸（靠1×1的卷积完成）。

对于这两种选项，当快捷连接在两个不同大小的特征图谱上出现时，用stride=2来处理。

3.4实现

我们用于ImageNet的网络是根据[21,40]来实现的，图片被根据短边等比缩放，按照[256,480]区间的尺寸随机采样进行尺度增强[40]。一个224x224的裁切是随机抽样的图像或其水平翻转，并将裁剪结果减去它的平均像素值[21]，进行了标准颜色的增强。我们把

批量正则化 (batch-normalization , BN) 用在了每个卷积层和激活层之间，我们初始化了权重按照[12]说的方法，分别从0开始训练普通/残差网络。我们使用SGD算法，mini-batch的大小为256.学习速率初始化为0.1，当到达错误率平台时就把学习速率除以10，对各模型进行了长达60万次迭代，我们用的了权重衰减，参数设了0.0001，动量参数为0.9，我们不用dropout，参考了[16]的实验结果。

测试时，对结果进行了比较研究，我们采用了标准的10-crop实验，为达到最佳效果，我们全连接卷积形式的网络就像[40,12]中说的一样，最终结果为对多个尺寸图像（图像分别被调整到短边{224,256,384,480,640}）的实验结果求平均值。

四. 实验

4.1 ImageNet 分类数据集

我们用ImageNet2012的分类数据集，有1000个分类，用这个数据集来评估我们的方法。各模型均用128万张训练图片，用来评估的验证集有5万张交叉验证图片，我们还用10万张测试图获得了一个最终结果，结果是由测试服务器报告的，我们还分别验证了第一和前5的错误率。

普通网络。我们首先评测了18层和34层的普通网络。34层普通网络（见图3中间），18层的网络是一个相似的结构。看表1来获取更细节的结构。

Table1: Imagenet的结构，块结构在了括号里（也可以看表5）几种类型的块堆积成网络架构，下采样用的stride=2的conv3_1 conv4_1, conv5_1。

结果在表2 中，证明更深的34层普通网络比18层的普通网络有更高的错误率，为了揭示此现象的原因，在图4左边我们比较了训练集和验证集的错误率（在训练过程中的），我们观察到了退化问题-----34层的普通网络在整个训练过程中都有更高的训练误差，尽管18层普通网络仅仅是34层的子集。

Table2: 最大错误率（%，10-crop测试），在imagenet的验证集上做测试。这儿残差网络和普通网络相比也没有任何额外参数。图4表现了训练的过程。

图4：Imagenet的训练，细的曲线表示训练集误差，粗的曲线表示验证集误差。左：普通网络（18/34层），右：残差网络（18/34层）

为了揭示原因，在图4（左），我们比较他们的培训过程中的训练集/验证验证集错误。在这个图中，残差网络与普通网络相比没有任何额外的参数。

我们认为，这种优化困难不太可能是梯度消失造成的。这些普通的网络是用了BN训练的[16]，这确保了前向传播时有非零的方差。我们还确认了反向传播时的梯度表现的很健康（有BN）。因此，既不是向前也不是向后的梯度消失。事实上，34层的普通网络仍然能够达到有竞争力的精度（表3），这表明在一定程度上是能工作的。我们推测，深的普通网络可能有指数级的较低的收敛速度，这会对训练误差的降低产生影响。这样优化困难的原因我们还将在未来进一步探究。

残差网络。接下来我们验证18层和34层的残差网络（ResNets）。残差网络的基本架构和上述普通网络相同，不同的是多了一些快捷连接，被添加到每对3x3的过滤器之间，如图3（右）。在第一个对比中（表2和图4右），我们对所有快捷连接使用自身映射和用零填充增加的维度（optionA），因此和普通网络相比没有任何额外的参数。

我们三个主要的观察，从表2和图4中得出。1.情况逆转：残差学习在34层上表现比18层好（2.8%）。更重要的，34层残差网络表现出了相当低的训练误差，并且同样适用于验证集。这表明在这种情况下，退化问题得到了很好的控制，即我们能在增加深度时获取更高的精度。2.相比于它基于的普通网络版本，34层残差网络降低了3.5%的最大错误率（表2），成功的降低了训练误差（图4右 vs 左）。这个对比验证了残差网络在深度学习系统中的有效性。

最后，我们也注意到，18层普通/残差网络是相对接近的（表2），但18层残差网络的收敛速度更快（图4右与左）。当网络“不深”（18层以下）的时候，目前的SGD算法仍然能在平凡网络上找到较好的解决方案。在这种情况下，残差网络能加速优化，在训练初期提供更快收敛速度。

自身捷径（identity）vs 投影(projection)捷径：我们已经证实了无参数的自身捷径对训练有增益作用。接下来我们打算调查投影捷径（Eqn2）在表三中我们比较三个选项：

- A 零填充捷径用来增加维度，所有的捷径都是没有参数的自身捷径（跟表2和图4右一样）

B 投影捷径用来增加维度，其他的捷径都是没有参数的自身捷径。

C 所有的捷径都是投影捷径

表3展示了这三个选项都远优于对应的普通网络，B稍微比A好一点儿，我们认为这是因为补零填充的维度在A中没有进行残差学习。C优于B，我们把这个成就归功于额外的参数（在一些投影捷径中的参数）但是ABC之间这么一点点的不同表明，投影捷径在解决退化问题上不是重点，所以我们在后文的其他部分就再也不用C方法了，以减少内存/时间复杂度和模型大小。自身连接是在不增加复杂度上是非常重要的，特别是针对我们下面要介绍的瓶颈结构。

深度瓶颈架构。接下来我们描述我们为ImageNet准备的更深的网络。因为太过漫长的训练时间我们负担不起，所以修改了单元块，改为一种瓶颈设计。对于每个残差函数F，我们使用3层来描述，而不是2层（图5）。这三层分别是1×1、3×3，和1×1的卷积层，其中1×1层负责先减少后增加（恢复）尺寸的，使3×3层具有较小的输入/输出尺寸瓶颈。图5显示了一个例子，两种设计都具有相似的时间复杂度。

无参数自身快捷连接在瓶颈架构中非常的重要，如果把自身连接（图5右）换成投影连接，可以看出时间复杂度和模型尺寸都会翻倍，因为该快捷连接连到了两个高维端，所以自身连接会为瓶颈设计带来更高的效率。

50层残差网络：我们把34层网络中的每一个2层的块都改换成3层的瓶颈块，在50层残差网络中的表现结果（见表1），我们用了OptionB来增加维度，这个模型的基础计算量为3.8亿Flops

101层和152层残差网络：我们建立了101层和152层的残差网络，用了更多的3层块（表1），显著的，尽管深度是在很明显的增加，152层的残差网络（1.13亿Flops）仍然比VGG16/19（15.3亿，19.6亿）有更低的复杂度。

50/101/152层的残差网络与34层相比，有着可观的精确度提升（表3,4）我们没有观察到退化问题，从而享受着深度增加带来的显著的精度增益。深度的好处从所有的评估指标中可以见证（表3,4）

与其他最先进技术的比较。在表4中，我们与之前最好的单模型结果进行了比较。我们的34层基础残差网络取得了非常有竞争力的准确度。我们的152层残差网络单模型对前5的验证错误率4.49%。这个单模型的结果优于所有以前的综合模型的记过（表5）。我们结合了六个不同深度的模型，来形成一个综合模型（在提交时只用了两个152层），达成了测试集上3.57%的top-5误差（表5）。这个结果在2015 ILSVRC获得了第一名。



4.2 CIFAR-10数据集的结果和分析

我们用CIFAR-10数据集进行了更多的研究，CIFAR-10数据集包括50K的训练图像，10K的测试图像，分10类。我们的实验用训练集训练和用测试集评估。我们关注的重点是极深网络的表现，而不是推动最先进的结果，所以我们故意使用如下的简单的架构。

普通/残差架构按照图3（中/右）的形式建立。网络的输入是32x32的图像，预先减去每一个像素的均值。第一层是3x3卷积层。然后我们对于尺寸分别为{32, 16, 8 }的特征图谱分别使用了一组包括了6n个3x3卷积层，每个尺寸的特征图谱使用2n个层，即过滤器的数量分别为{16,32,64}，降采样是通过步长为2的卷积进行的，网络以全局的均值池化终止，一个10路全连通层，softmax。以上一共6n+2个有权重的层。下方的表格总结了结构：

当快捷连接被使用时，它们分别连接到3x3的网络层对（一共3n个快捷连接）。在该数据集上，我们在各种情形下都使用自身快捷路径（即选项A），所以我们的残差模型跟普通模型有这一模一样的深度、宽度和参数个数

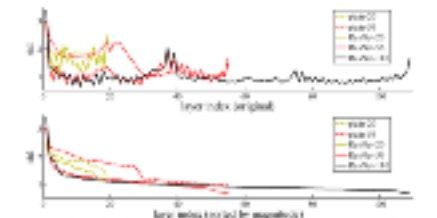
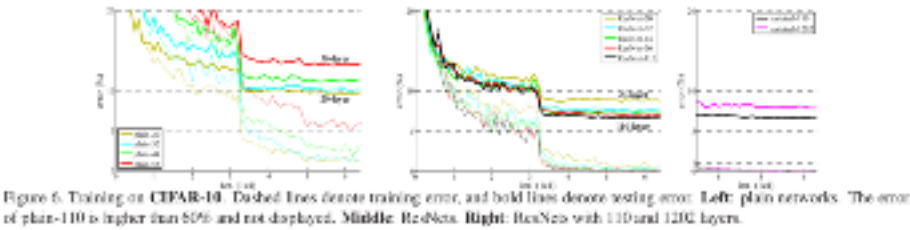
表6：在CIFAR-10数据集上的分类错误，所有方法都有数据增强，对于110层的残差网络我们运行了5次，展示了“最优值（平均+波动）”

我们用的权重衰减参数为0.0001动量为0.9，并采用论文[13]提到的权重初始化[12]，Batch-normalization[16]，不过不使用dropout。这些模型训练时用了两个gpu，mini-batch大小为128。我们初始学习率为0.1，在32k和48k次迭代时除以10，在64k次迭代时终止训练，按45k/5k的比例确定训练集/验证集。我们遵循[24]提出的简单数据增强策略进行训练：4像素被增加在各边上，32x32的切割是完全随机的，从填充后的图像或者其翻转中采样。在验证时，我们只是评估了原本的32x32的图片。

我们比较了 $n=\{3,5,7,9\}$,分别对应20,32,44,56层的网络。图6左展示了普通网络的表现。深度普通网络受深度影响,在深度增加时表现出了更高的训练误差,这个现象和在Imagenet(图4左)以及MNIST数据集[41]是如出一辙的,表明了这样的优化困难问题是一个共通的根本问题。

图6(中)展示了残差网络的表现。也和用Imagenet做实验时表现的差不多(图4右),我们残差网络就是克服了优化困难的问题,并且做到了让精度随着深度增加而增加。

我们更多的研究了 $n=18$ 的情况,这个对应了110层的残差网络,在这个情况下,我们发现初始学习速率设成0.1太大了,无法让网络开始收敛,所以们用了0.01的初始学习速率来预热训练,直到训练误差低于80%的时候(大约400次迭代),然后再回到0.1的学习速率继续训练。接下来的训练方案是跟前文提到的一模一样。这个110层的网络收敛的很好(图6中)它和其他网络(Fitnet[34],highway[41])比起来有更少的参数。且获得了很好的结果(6.43%表6)



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 appendix for better results.

metric	mAP@0.5	mAP@0.5 : 0.95
VGG-16	41.5	24.2
ResNet-101	48.4	27.2

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

网络层响应分析。图7展示了层响应的标准差(std)。响应是指的每个3x3层的输出 (Batch-normalization后,其他非线性操作前(Relu/addition))。对于残差网络,这个分析解释了残差函数的响应强度。图7展示了残差网络和普通网络比起来有着较小的响应。这些结果支撑了我们的原始动机 (Sec3.1),即残差函数可能比非残差函数更加的接近0。我们通过比较resnet-20,56,110的结果,还注意到,更深的残差网络有着更小幅度的响应,当有更多层的时候,单个的Resnet层会修改的更少。

超1000层网络的探究。我们探索了一个更深的模型,超过1000层。我们把n设成200,这对应了1202层的网络,用前文一样的方法去训练。我们的方法没有显示出优化的困难,1000层的网络可以获得训练误差<0.1%的结果(图6右),它的测试集错误率仍然相当的不错(7.93%表6)。

不过对于这种太深的网络同样有问题,1202层的网络在测试集上的结果比110层差,尽管在训练集上的错误率表现的比较相近。我们认为这是因为过拟合。1202层的网络相对于这个数据集来说实在是庞大到有点没必要。诸如maxout[9]或者dropout[13]通常用在此数据集上来解决这个问题,以获得更好的结果[9,25,24,34]。在此论文中,我们没有用Maxout/dropout,只是简单的实施了正规化来配合增大/减少网络深度架构设计,不偏离探究优化困难的主路线。然而结合更强的正规化手段进一步提升结果,是我们未来要研究的方向。

4.3 物体检测 (PASCAL和MS COCO)

我们的方法在其他的识别任务上同样有着很好的表现,表7和表8显示了在对象检测的基本结果。用的数据集是PSCAL VOC 2007,2012[5]和COCO[26],我们采用了更快的R-CNN[32]作为检测方法,这儿我们重点关注的是把VGG16替换成ResNet101后的性能提升。使用这两种检测模型的实现是类似的(见附录),所以增益只能归功于更好的网络

结构。值得注意的是，在COCO数据集上的挑战，我们获得的结果比COCO标准测试（mAP@.5,.95）提升了6.0%，相对提升量达28%，这个成就完全归功于本文描述的残差学习方法。

在深度残差网络的基础上，我们获得了更多项目的第一名：ImageNet detection, ImageNet localization COCO detection COCO segmentation等。细节见附录。

References

[1] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2):157–166, 1994.

[2] C. M. Bishop. Neural networks for pattern recognition. Oxford university press, 1995.

[3] W. L. Briggs, S. F. McCormick, et al. A Multigrid Tutorial. Siam, 2000.

[4] K. Chatfield, V. Lempitsky, A. Vedaldi, and A. Zisserman. The devil is in the details: an evaluation of recent feature encoding methods. In BMVC, 2011.

[5] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The Pascal Visual Object Classes (VOC) Challenge. IJCV, pages 303–338, 2010.

[6] R. Girshick. Fast R-CNN. In ICCV, 2015.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.

[8] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In AISTATS, 2010.

[9] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. arXiv:1302.4389, 2013.

[10] K. He and J. Sun. Convolutional neural networks at constrained time cost. In CVPR, 2015.



[11] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deepconvolutional networks for visual recognition. In ECCV, 2014.

[12] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In ICCV, 2015.

[13] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing coadaptation of feature detectors. arXiv:1207.0580, 2012.

[14] S. Hochreiter. Untersuchungen zu dynamischen neuronalen netzen. Diploma thesis, TU Munich, 1991.

[15] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

[16] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML, 2015.

[17] H. Jegou, M. Douze, and C. Schmid. Product quantization for nearest neighbor search. TPAMI, 33, 2011.

[18] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, and C. Schmid. Aggregating local image descriptors into compact codes. TPAMI, 2012.

[19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. arXiv:1408.5093, 2014.

[20] A. Krizhevsky. Learning multiple layers of features from tiny images. Tech Report, 2009.

[21] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[22] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural computation, 1989.



[23] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller. Efficient backprop. In *Neural Networks: Tricks of the Trade*, pages 9–50. Springer, 1998.

[24] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeply supervised nets. *arXiv:1409.5185*, 2014.

[25] M. Lin, Q. Chen, and S. Yan. Network in network. *arXiv:1312.4400*, 2013.

[26] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common objects in context. In *ECCV*. 2014.

[27] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.

[28] G. Montúfar, R. Pascanu, K. Cho, and Y. Bengio. On the number of linear regions of deep neural networks. In *NIPS*, 2014.

[29] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010.

[30] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *CVPR*, 2007.

[31] T. Raiko, H. Valpola, and Y. LeCun. Deep learning made easier by linear transformations in perceptrons. In *AISTATS*, 2012.

[32] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.

[33] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge university press, 1996.

[34] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio. Fitnets: Hints for thin deep nets. In *ICLR*, 2015.

[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. ImageNet large scale visual recognition challenge. *arXiv:1409.0575*, 2014.



[36] A. M. Saxe, J. L. McClelland, and S. Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. arXiv:1312.6120, 2013.

[37] N. N. Schraudolph. Accelerated gradient descent by factor-centering decomposition. Technical report, 1998.

[38] N. N. Schraudolph. Centering neural network gradient factors. In *Neural Networks: Tricks of the Trade*, pages 207–226. Springer, 1998.

[39] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *ICLR*, 2014.

[40] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015.

[41] R. K. Srivastava, K. Greff, and J. Schmidhuber. Highway networks. arXiv:1505.00387, 2015.

[42] R. K. Srivastava, K. Greff, and J. Schmidhuber. Training very deep networks. 1507.06228, 2015.

[43] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR*, 2015.

[44] R. Szeliski. Fast surface interpolation using hierarchical basis functions. *TPAMI*, 1990.

[45] R. Szeliski. Locally adapted hierarchical basis preconditioning. In *SIGGRAPH*, 2006.

[46] T. Vatanen, T. Raiko, H. Valpola, and Y. LeCun. Pushing stochastic gradient towards second-order methods—backpropagation learning with transformations in nonlinearities. In *Neural Information Processing*, 2013.

[47] A. Vedaldi and B. Fulkerson. *VLFeat: An open and portable library of computer vision algorithms*, 2008.



[48] W. Venables and B. Ripley. Modern applied statistics with s-plus.1999.

[49] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutionalneural networks. In ECCV, 2014.9778

(注：感谢您的阅读，希望本文对您有所帮助。如果觉得不错欢迎分享转载，但请先点击这里 (/link.jianshu.com?t=https://101705160006292.bqy.mobi/) 获取授权。本文由版权印 (/link.jianshu.com?t=https://www.banquanyin.com/) 提供保护，禁止任何形式的未授权违规转载，谢谢！)

哎呦，不行了，要睡觉了，赏我 e(=2.72) 元咖啡钱吧，您的支持将鼓励我继续创作！

赞赏支持



(/u/d4ed8a426833)

ML论文翻译 (/nb/8413272)

举报文章 © 著作权归作者所有



zhwong (/u/38cd2a8c425e)

写了 279267 字，被 947 人关注，获得了 1193 个喜欢 (/u/38cd2a8c425e)


+ 关注

Stay Hungry Stay Foolish 常葆求知若饥 常存虚怀若愚 Blog：http://zhwong.ml

喜欢 (/sign_in?utm_source=desktop&utm_medium=not-signed-in-like-button) | 20

更多分享

(http://cwb.assets.jianshu.io/notes/images/8471215



下载简书 App ▶

随时随地发现和创作内容



(/apps/download?utm_source=nbc)



登录后发表评论 (/sign_in?utm_source=desktop&utm_medium=not-signed-in-comment-form)

评论

智慧如你，不想发表一点想法 (/sign_in?utm_source=desktop&utm_medium=not-signed-in-nocomments-text)咩~



被以下专题收入，发现更多相似内容

- 深度学习-神经网络 (/c/71b2d8a98305?utm_source=desktop&utm_medium=notes-included-collection)
- 神经网络与深度学习 (/c/12a43f2ae156?utm_source=desktop&utm_medium=notes-included-collection)
- 机器学习与数据挖掘 (/c/9ca077f0fae8?utm_source=desktop&utm_medium=notes-included-collection)
- 深度学习-计算图 (/c/1249336e61cb?utm_source=desktop&utm_medium=notes-included-collection)
- 机器学习与计算机视觉 (/c/ee1275bb82ca?utm_source=desktop&utm_medium=notes-included-collection)
- 深度学习 (/c/13b41686e443?utm_source=desktop&utm_medium=notes-included-collection)
- 机器学习之深度学习 (/c/86d8ab021061?utm_source=desktop&utm_medium=notes-included-collection)

展开更多

推荐阅读

更多精彩内容 > (/)

推荐 | 九本不容错过的深度学习和神经网络书籍 (/p/c20917a91472?utm_campaign=maleskine&utm_content=note&utm_source=recommendation)

原文：机器之心 aiotify 针对 30 多本深度学习和神经网络书籍，我们（AI Optify 数据团队）使用不同指标（比如，在线评价、打分、所涉主题、作者影... zhwhong (/u/38cd2a8c425e?utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)

[译] Every Filter Extracts A Specific Texture In Con... (/p/20b854ffab02?utm_campaign=maleskine&utm_content=note&utm_source=recommendation)

题目：卷积神经网络中的每一个过滤器提取一个特定的特征 文章地址：《Every Filter Extracts A Specific Texture In Convolutional Neural Networks》... zhwhong (/u/38cd2a8c425e?utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)

2017 结束了。你害怕的，你都经历过了 (/p/39d407daaf7e?utm_campaign=maleskine&utm_content=note&utm_source=recommendation)

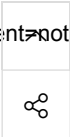
Sayings: 2017 就这样结束了。但，此刻我想问问你：你还记得这一年是怎么开始的吗？我去翻了翻年初读者们的留言，然后回访了其中几位。他们当中，... 新世相 (/u/880b482939ec?utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)

这个冬天，一定要来这里看雪 (/p/5a66ea10e3cf?utm_campaign=maleskine&utm_content=note&utm_source=recommendation)

不知道你们最近有没有被朋友圈各个地方的雪刷屏，西安、南京这些古城都下起了雪，有些地方一下雪，就像穿越了千年以前，今天我强烈跟大家推荐一个看... 流浪摄影师 (/u/df7a9d2e01e6?utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)

坚持五点半起床半年后，给我带来的蜕变 (/p/564f3ac19709?utm_campaign=maleskine&utm_content=note&utm_source=recommendation)

前不久，和几位上学时的小伙伴聊天，话题转到了我们现在的生活上。许久未联系的我们，说着每个人的近况。小陈在家乡考上了事业编，每天朝九晚五... 羊达令 (/u/ce94d617e045?utm_campaign=maleskine&utm_content=user&utm_medium=pc_all_hots&utm_source=recommendation)



(/p/f3b8141ac43b?



(/a
utn

utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
ResNet论文翻译——中英文对照 (/p/f3b8141ac43b?utm_campaign=male...

声明：作者翻译论文仅为学习，如有侵权请联系作者删除博文，谢谢！ Deep Residual Learning for Image Recognition Abstract Deeper neural networks are more difficult to train....

SnailTyan (/u/7731e83f3a4e?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/eae07a953727?



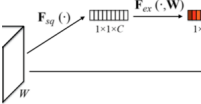
utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
ResNet论文翻译——中文版 (/p/eae07a953727?utm_campaign=maleskin...

声明：作者翻译论文仅为学习，如有侵权请联系作者删除博文，谢谢！ Deep Residual Learning for Image Recognition 摘要 更深的神经网络更难训练。我们提出了一种残差学习框架来减轻网络训练，这些网络比...

SnailTyan (/u/7731e83f3a4e?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/72a5484b9f09?



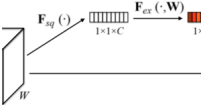
utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
Squeeze-and-Excitation Networks论文翻译——中英文对照 (/p/72a5484b...

文章作者：Tyan博客：noahsnail.com | CSDN | 简书 Squeeze-and-Excitation Networks Abstract Convolutional neural networks are built upon the convoluti...

SnailTyan (/u/7731e83f3a4e?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

(/p/608941724182?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
Squeeze-and-Excitation Networks论文翻译——中文版 (/p/60894172418...

文章作者：Tyan博客：noahsnail.com | CSDN | 简书 声明：作者翻译论文仅为学习，如有侵权请联系作者删除博文，谢谢！ Squeeze-and-Excitation Networks 摘要 卷积神经网络建立在卷积运算的基础上，通过融...

SnailTyan (/u/7731e83f3a4e?

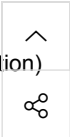
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)


(/p/9d6082068f53?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)
Very Deep Convolutional Networks for Large-Scale Image Recognition...


Very Deep Convolutional Networks for Large-Scale Image Recognition 摘要 在这项工作中，我们研究了卷积网络深度在大规模的图像识别环境下对准确性的影响。我们的主要贡献是使用非常小的（3×3）卷积滤...



 SnailTyan (/u/7731e83f3a4e?
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

使用Babel和ES7创建JavaScript模块 (/p/e1411cb9e955?utm_campaign=...

【编者按】本文主要介绍通过 ES7 与 Babel 建立 JavaScript 模块。文章系国内 ITOM 管理平台 OneAPM 工程师编译呈现，以下为正文。 去年，新版的JavaScript发布了，它有很多新的优点。其中之一就是导入导...

 OneAPM_Official (/u/0c6074c10464?
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)


(/p/3bcc66040892?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

秦腔 (/p/3bcc66040892?utm_campaign=maleskine&utm_content=note...

我是有间歇性夜跑习惯的人。说来惭愧，既然是间歇性便不能称之为习惯。这算是恬不知耻的粉饰自己热爱锻炼的体面措辞。值得庆幸的是，鄙人所居与大明宫遗址公园勉强毗邻，也算是行了极大的方便。这园子...

 悟踪 (/u/cfa2847f1957?
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)


(/p/b3c33d3439c6?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)


爱自己胜过爱爱情 (/p/b3c33d3439c6?utm_campaign=maleskine&utm_c...

送给伤着心的人 --- 作者：划过天空的青鸟 那是电影《致我们终将逝去的青春》里的一句台词：“我们都爱自己胜过爱爱情。”当时听了觉得好搞笑，好酸，好文艺。过了这么久，看到那些心酸的故事，忽然想起了...

 划过天空的青鸟 (/u/51dedcbf63c3?
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

170722晨读感悟 藏在好奇背后的真面目 (/p/9a75d621d33d?utm_campaig...

“好奇心害死猫”和“好奇的活着”两者之间的距离，就是欲望和需求的关系。欲望永无止境 人的欲望有很多，而且永远无法满足。此时的欲望，在彼时一旦得到了实现，便会冒出一个新的欲望出来，永无尽头。正如...

 简单的自洽 (/u/281aefa32a6b?
utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommendation)

