第三十一届AAAI人工智能会议论文集(AAAI-17)

**Inception-v4, Inception-ResNet和残留连接对学习的影响**

**Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A. Alemi谷歌Inc。**

1600剧场百汇

加州山景城

**摘要**

近年来，深度卷积网络在图像识别性能的最大进步中发挥了核心作用。一个例子是Inception体系结构，它以相对较低的计算成本实现了非常好的性能。最近，在2015年ILSVRC挑战赛中，残余连接与更传统的建筑相结合，取得了最先进的性能;其性能与最新一代的“盗梦空间v3”网络相似。这就提出了一个问题:将先启体系结构与剩余的连接结合起来有什么好处吗?在这里，我们给出了明确的经验证据，即使用剩余连接的训练显著加速了初始网络的训练。也有一些证据表明，在没有剩余连接的情况下，剩余的初始网络的性能比同样昂贵的初始网络要好得多。我们还为剩余和非剩余初始网络提出了几个新的流线型体系结构。这些变化显著提高了ILSVRC 2012分类任务的单帧识别性能。我们进一步演示了如何适当的激活缩放稳定训练非常广泛的残余Inception网络。在ImageNet分类(CLS)挑战的测试集上，我们集成了3个残差和1个cepi -v4网络，获得了3.08%的前5位错误。

**介绍**

物体识别是计算机视觉和人工智能的核心任务。强大的视觉模型是能够处理视觉输入的人工智能系统的关键组成部分。他们的应用包括计算机用户界面(手势识别)、网络搜索、OCR系统、自动交通、医学成像、区域成像、机器人和图像处理。在2012年之前，每个特定的应用领域都需要专门的解决方案。从那时起，深度卷积神经网络成为解决这些问题的主流。卷积神经网络可以追溯到20世纪80年代(Fukushima 1980)和(LeCun et al. 1989)，但最近在大规模ImageNet图像识别基准ILSVRC (Russakovsky et al. 2014)上取得的良好结果(Krizhevsky, Sutskever和Hinton 2012)使人们对其使用重新产生了兴趣。的

版权所有?c 2017，人工智能促进协会(www.aaai.org)。保留所有权利。

同样的神经网络架构“AlexNet”(Krizhevsky, Sutskever, and Hinton 2012)已经应用于大量的应用领域，并取得了良好的效果。

相同的架构,适合对象检测可以成功地应用于各种计算机视觉任务,包括目标检测(Girshick et al . 2014),分割(长,Shelhamer,达雷尔2015),人体姿态估计(Toshev和Szegedy 2014)、视频分类(Karpathy et al . 2014),对象跟踪(王、杨2013)和超分辨率(董et al . 2014年)。这些例子只是深度卷积网络从那时起被非常成功地应用到众多应用中的一小部分。此外，已经证明，目标识别性能的卷积网络的架构改进往往转化为其他任务的性能收益。

这种普适性促使我们将重点放在广泛使用的ILSVRC12对象识别基准(Russakovsky et al. 2014)的识别模型上，该基准的任务是将图像分成一千种不同类别中的一种(或五种)。该数据集由120万张训练图像、5万张验证图像和10万张测试图像组成。所有这些都在1000个班级中平均分配。自2010年以来，该基准测试一直是衡量对象识别解决方案质量的一个非常流行的任务。

在这项工作中，我们研究了两个最新的想法的组合:残余连接(He et al. 2015)和盗梦空间架构的最新修订版本(Szegedy et al. 2015b)。在(He et al. 2015)中，有人认为剩余连接对于训练非常深的架构具有内在的重要性。由于初始网络往往非常深，所以很自然地用剩余的连接替换初始体系结构的过滤器串联阶段。这将允许Inception在保留其计算效率的同时获得剩余方法的好处。

除了简单的集成之外，我们还研究了没有剩余连接的Inception是否可以通过使其更深入和更广泛来提高效率。为此，我们评估了一个名为Inception- v4的新版本，它比Inception-v3有更统一的简化架构和更多的Inception模块。从历史上看，《盗梦空间3》继承了许多早期化身的包袱。技术约束主要来自

4278

需要使用DistBelief对模型进行划分以进行分布式培训(Dean等人，2012年)。现在，在将我们的训练设置迁移到TensorFlow (Abadi等人，2015)之后，这些约束已经被解除，这使得我们能够显著地简化架构。这个简化的体系结构的细节在从第2页开始的体系结构选择部分中描述。

在本文中，我们将比较两个纯Inception版本，Inception-v3和v4，以及同样昂贵的混合版本Inception- resnet。诚然，这些模型的选择有些特别，主要的限制是模型的参数和计算复杂性应该与非残差模型的成本有点相似。我们已经测试了更大、更广泛的“盗梦空间”变体，它们在ImageNet分类挑战(Russakovsky等人2014年)数据集上的“盗梦空间”- resnet v2级别上运行。

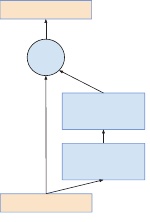
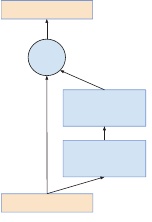
最后，我们报告对所描述的所有模型的综合评估。显然，cepep -v4和cepep - resnet -v2的性能都很好，每一个都在Im- ageNet验证数据集上超过了最先进的单帧性能，我们想看看组合是如何在这个研究得很好的数据集上推动最先进的技术。令人惊讶的是，我们发现单帧性能上的增益并不转化为装配性能上的类似的大增益。尽管如此，据我们所知，它仍然允许我们报告验证集上3.1%的前5位错误，并集成了4个模型，设置了一个新的最先进水平。

**相关工作**

在(Krizhevsky, Sutskever, and Hinton 2012)之后，卷积网络在大规模图像识别任务中变得流行。接下来的一些重要里程碑是网络中的网络(Lin, Chen, and Yan 2013)， VG- GNet (Simonyan和Zisserman 2014)和google网(启态v1) (Szegedy et al. 2015a)。

在(He et al. 2015)中引入了残差连接，他们为利用信号相加合并进行图像识别，特别是目标检测提供了令人信服的理论和实践证据。作者认为，对于训练非常深的卷积模型来说，剩余连接本质上是必要的。我们的发现似乎不支持这一观点，至少在图像识别方面。然而，这可能需要对更深层次的网络进行更多的实验，才能充分理解残留连接的真正好处。在实验部分，我们证明了在不利用剩余连接的情况下训练非常深度的竞争网络并不困难。然而，残余连接的使用似乎大大提高了训练速度，这是单独使用它们的一个重要论点。

Inception深度卷积架构在(Szegedy et al. 2015a)中被引入google et，此处命名为Inception-v1。后来，Inception体系结构以各种方式进行了改进，首先引入了批标准化(Ioffe和Szegedy 2015) (Inception-v2)。随后在第三次迭代中通过附加的因子分解思想(Szegedy et



5 hox ? DFWLYDWLRQ

**吗?**



5 hox ? DFWLYDWLRQ

5 hox ? DFWLYDWLRQ

**吗?**

&RQY



5 hox ? DFWLYDWLRQ

图1:中引入的剩余连接(He et al. 2015)。左边是原始的剩余连接。右边是一个优化版本，通过使用1 × 1卷积降低了计算成本。

al. 2015b)，将在本报告中称为“盗梦空间v3”。

**体系结构的选择**

**纯粹的《盗梦空间》块**

我们较老的Inception模型曾经以分区的方式进行训练，每个副本被划分为多个子网络，以便能够在内存中适应整个模型。然而，初始体系结构是高度可调的，这意味着对不同层中的过滤器数量有很多可能的更改，这些更改不会影响完全训练过的网络的质量。为了优化训练速度，我们仔细调整了层的大小，以平衡各个模型子网络之间的计算。相比之下，通过引入TensorFlow (Abadi等人，2015)，我们最新的模型可以在不分割副本的情况下进行训练。这在一定程度上是通过最近的内存反向传播优化实现的，通过仔细考虑哪些张量需要进行梯度计算和结构化计算来减少此类张量的数量。在历史上，我们在改变架构方面相对保守，并且在保持网络其他部分稳定的同时，将实验限制在不同的孤立网络组件上。如果不简化早期的选择，就会导致网络看起来比实际需要的更复杂。在我们最新的实验中，对于Inception-v4，我们决定摆脱这个不必要的包袱，并为每个网格大小的Inception块做出统一的选择。

Inception-v4网络的完整配置在图2中概述，其中包括总体架构和主干配置，图3详细介绍了内部模块的构造。

**剩余初始块**

对于Inception网络的剩余版本，我们使用比原始的Inception更便宜的Inception块。每个初始块后面是滤波器扩展层(1 × 1卷积，没有激活)，用于在残差相加之前扩大滤波器组的维度，以匹配输入的深度。这是必要的

4279

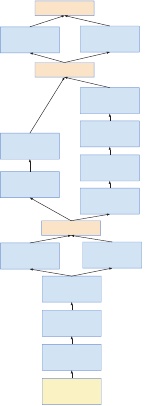
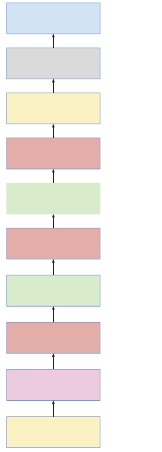
补偿由Inception块引起的维度降低。

我们尝试了Inception剩余版本的几个版本。这里只有两份详细资料。第一个“盗梦网-v1”的计算成本大致相当于“盗梦网-v3”，而“盗梦网-v2”的原始成本则相当于新推出的“盗梦网-v4”。然而，在实践中，Inception-v4的步长时间被证明是明显较慢的，这可能是由于更多的层数。

我们的残留和非残留初始变量之间的另一个小技术差异是，在我们的Inception- ResNet实验中，我们只在传统层之上使用批标准化，而不是在残留求和之上。合理的期望,彻底使用batch-normalization应该是有利的,但实施batch-normalization TensorFlow消耗大量的内存,我们需要减少整体层数,如果batch-normalization使用无处不在。

Inception-Resnet-v1网络的完整配置如图6所示，其中包含总体架构和主干配置，图4包含内部模块的详细配置。

Inception-Resnet-v2网络的完整配置使用图6中的模式、图2中的主干和图5中的模块。



6 riwpd [

2 xwsxw ? ?

“URSRXW ?NHHS ? ?

2 xwsxw ? ?

3 rrolqj YDUDJH美元?

2 xwsxw ? ?

(? ? ?, QFHSWLRQ ? &

2 xwsxw ???[?[????

5 hgxfwlrq ? %

2 xwsxw ???[?[????

(? ? ?, QFHSWLRQ ? %

2 xwsxw ????[??[????

美元5 hgxfwlrq ?

2 xwsxw ????[??[????

(? ? ?, QFHSWLRQ ?美元

2 xwsxw ????[??[???

6你流泪

2 xwsxw ????[??[???

, QSXW ?(? ? ?

(? ? ?

? ? (? ? ? ?

? ? (? ? ? ?

(? ? &RQY ?

9 ? ?

(? ? &RQY ?

? ?

? ? (? ? ? ?

???[???[??

???[???[??

???[???[??

(? ? ?

) LOWHU ? FRQFDW

(? ?&RQY 0 d [3 rro ? ?9 9 VWULGH ? ?

) LOWHU ? FRQFDW

(? ? &RQY ?

9 ? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

) LOWHU ? FRQFDW

(? ?0 d [3 rro ?[??&RQY VWULGH ? ?9 ??? 9 VWULGH ? ?

(? ? &RQY ?

? ?

(? ?9 &RQY ? ?

(? ?&RQY ? ? VWULGH ? ? 9

, QSXW ?(? ? ?

图2:左边是纯Inception-v4网络的整体模式。右边是茎的详细组成。请注意，图5、6中的盗梦空间- resnet -v2网络轮廓也使用了这种主干配置。V表示使用' Valid '填充，否则使用' Same '填充。每一层边上的大小汇总了该层输出的形状。

**残差的缩放**

我们发现，如果过滤器的数量超过1000个，残差变异开始表现出不稳定性，网络在训练初期就“死亡”了，这意味着平均池化前的最后一层在几万次迭代后开始只产生零。这是无法阻止的，可以通过降低学习速率，或者通过向这一层添加额外的批标准化。

我们发现，在将残差添加到前一层激活之前，缩小残差似乎可以稳定训练。一般来说，我们选择了0.1到0.3之间的一些缩放因子来缩放残差，然后将它们添加到累积层激活中(参见图7)。

在非常深的残差网络中也观察到了类似的不稳定性(He et al. 2015)，他们建议采用两阶段训练，即第一个“热身”阶段以非常低的学习率完成，然后是第二个学习率很高的阶段。我们发现，如果过滤器的数量非常高，那么即使非常低的(0.00001)学习率也不足以应对不稳定性，高学习率的训练有机会破坏其影响。我们发现计算残差更可靠。

即使在没有严格必要的地方，它似乎从来没有损害最终的准确性，但它有助于稳定训练。

**培训方法**

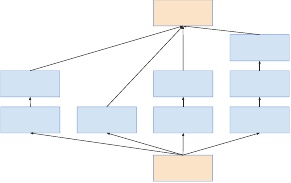
我们使用随机梯度下降法训练网络，使用TensorFlow (Abadi等人，2015)分布式机器学习系统，使用20个副本，每个副本运行NVidia Kepler GPU。我们早期的实验使用动量(Sutskever et al. 2013)，衰减为0.9，而我们的最佳模型使用RMSProp (Tieleman and Hinton)，衰减为0.9和?= 1.0。我们的学习速率是0.045，每隔两个纪元衰减一次，指数速率是0.94。此外，梯度裁剪(Pas- canu, Mikolov，和Bengio 2012)被发现对稳定训练是有用的。模型评估是使用随时间计算的参数的运行平均值来执行的。

**实验结果**

首先，我们观察了四种变量的前1和前5验证错误在训练过程中的演变。实验结束后，我们发现我们的持续评估是在验证集的一个子集上进行的，该子集由于边界框不佳而遗漏了约1700个黑名单实体。结果表明，应该只在CLSLOC基准测试中执行省略，但是与其他报告(包括我们团队早期的一些报告)相比，会产生一些无法比较的(更乐观的)数字。top-1错误的差异约为0.3%，top-5错误的差异约为0.15%。然而，由于差异是一致的，我们认为曲线之间的比较是公平的。

另一方面，我们在由50,000张图像组成的完整验证集上重新运行了我们的多茬和集成结果。最后的集成结果也在测试集上执行，并发送到ILSVRC测试服务器

4280



) LOWHU ? FRQFDW

(? ? &RQY ?

? ?3 rrolqj YJ美元?

(? ? &RQY ?? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

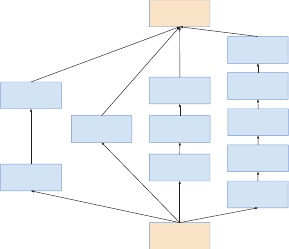
(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

) LOWHU ? FRQFDW



) LOWHU ? FRQFDW

(? ?&RQY ? ?

(? ? &RQY ?? ?

3 rrolqj YJ美元?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

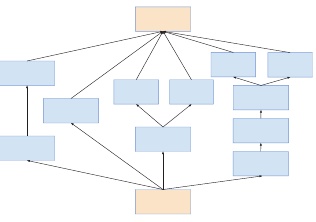
(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

) LOWHU ? FRQFDW



) LOWHU ? FRQFDW

(? ?&RQY ? ?

(? ?&RQY ? ?

(? ?&RQY ? ?

(? ?&RQY ? ?

(? ?&RQY ? ?

(? ? &RQY ?? ?

(? ?&RQY ? ?

3 rrolqj YJ美元?

(? ?&RQY ? ?

(? ? &RQY ?

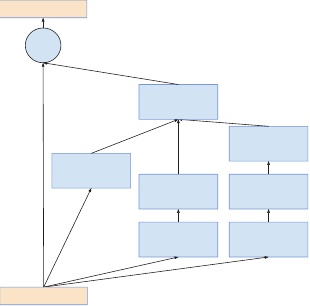
? ?

(? ? &RQY ?

? ?

) LOWHU ? FRQFDW

图3:纯净的Inception-v4网络内部网格模块的架构。35 × 35、17 × 17和8 × 8的网格模块从左到右绘制。这些是图2中的盗梦空间- a、盗梦空间- b和盗梦空间- c。



5 hox ?DFWLYDWLRQ吗?

(? ?&RQY ? ? / LQHDU

(? ? &RQY ?? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

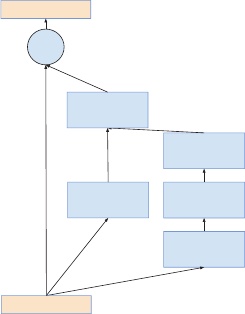
(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

5 hox ? DFWLYDWLRQ



5 hox ? DFWLYDWLRQ

**吗?**

(? ?&RQY ? ? / LQHDU

(? ?&RQY ? ?

(? ? &RQY ?

? ?

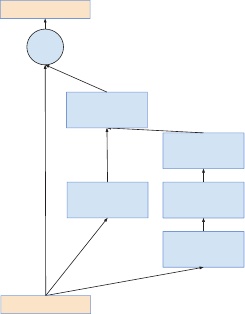
(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

5 hox ? DFWLYDWLRQ



5 hox ? DFWLYDWLRQ

**吗?**

(? ?&RQY ? ? / LQHDU

(? ?&RQY ? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

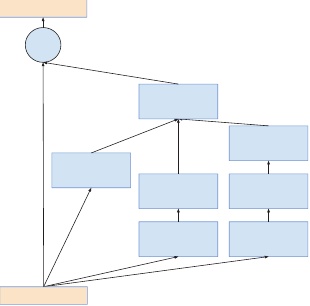
? ?

(? ? &RQY ?

? ?

5 hox ? DFWLYDWLRQ

图4:Inception-ResNet-v1网络内部网格模块的架构。35 × 35、17 × 17和8 × 8的网格模块从左到右绘制。这些是图6左侧的Inception-ResNet-v1网络架构的Inception-A、Inception-B和Inception-C块。



5 hox ?DFWLYDWLRQ吗?

(? ?&RQY ? ? / LQHDU

(? ? &RQY ?? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

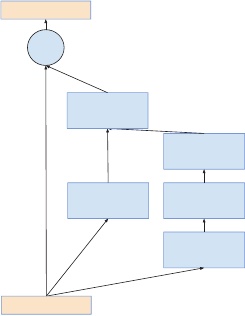
(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

5 hox ? DFWLYDWLRQ



5 hox ? DFWLYDWLRQ

**吗?**

(? ?&RQY ? ? / LQHDU

(? ?&RQY ? ?

(? ? &RQY ?

? ?

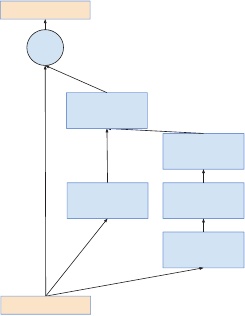
(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

5 hox ? DFWLYDWLRQ



5 hox ? DFWLYDWLRQ

**吗?**

(? ?&RQY ? ? / LQHDU

(? ?&RQY ? ?

(? ? &RQY ?

? ?

(? ? &RQY ?

? ?

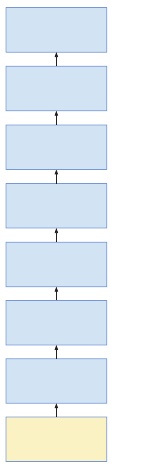
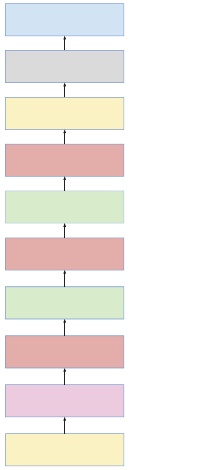
(? ? &RQY ?

? ?

5 hox ? DFWLYDWLRQ

图5:Inception-ResNet-v2网络内部网格模块架构。35 × 35、17 × 17和8 × 8的网格模块从左到右绘制。这些是图6左边的盗梦空间- resnet -v2网络的盗梦空间- a、盗梦空间- b和盗梦空间- c模块。

4281



6 riwpd [

2 xwsxw ? ?

“URSRXW ?NHHS ? ?

2 xwsxw ? ?

3 rrolqj YDUDJH美元?

2 xwsxw ? ?

(? ? ?, QFHSWLRQ ? UHVQHW ? ? &

2 xwsxw ???[?[????

5 hgxfwlrq ? %

(? ? ?, QFHSWLRQ ? UHVQHW ? ?%

2 xwsxw ???[?[????

2 xwsxw ????[??[???

美元5 hgxfwlrq ?

2 xwsxw ????[??[???

(? ? ?, QFHSWLRQ ? UHVQHW ? ?美元

2 xwsxw ????[??[???

6你流泪

2 xwsxw ????[??[???

, QSXW ?(? ? ?

(? ? ?

(? ?&RQY ? ? VWULGH ? ? 9

? ? (? ? ? ?

(? ? &RQY ?

? ?9 ? ? ? &RQY

? ?(? ?0 d [3 rro VWULGH ? ?9 ? ? ? &RQY

? ?(? ? &RQY ?

? ?9 ? (? ?&RQY ? ? VWULGH ? ? 9

, QSXW ?(? ? ?

? ? (? ? ? ?

? ? (? ? ? ?

? ? (? ? ? ?

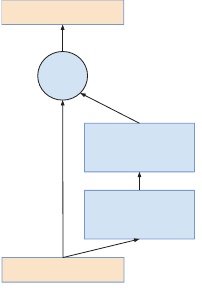
???[???[??

???[???[??

???[???[??

(? ? ?

图6:左边是Inception- Resnet-v1和Inception- resnet -v2网络的总体架构。虽然这两个网络的模式是相同的，但主干和内部模块的组成不同。Inception- Resnet-v1的主干如图2右边所示，而Inception- Resnet-v2的主干与纯Inception-v4网络相同，如图2右边所示。内部模块分别在图4和图5中表示。V表示使用' Valid '填充，否则使用' Same '填充。每一层边上的大小汇总了该层输出的形状。



5 hox ? DFWLYDWLRQ

**吗?**

FWLYDWLRQ美元?6 fdolqj

, QFHSWLRQ

5 hox ? DFWLYDWLRQ

图7:扩展合并的Inception-ResNet模块的一般模式。我们希望相同的思想在一般的ResNet情况下是有用的，在这种情况下，使用一个任意的子网而不是Inception块。缩放块只是用一个合适的常数来缩放最后的线性激活，通常在0.1左右，但我们发现，更深的网络需要更低的常数。

为了验证我们的调谐没有导致过配合。我们想强调,最后验证了只有一次,我们去年已经提交我们的结果只有两次:一次BN-Inception纸,后来在ilsvr - 2015年CLSLOC竞争,所以我们相信测试集数据构成一个真正的估计模型的泛化能力。

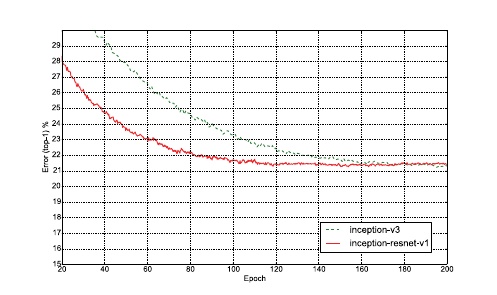


图8:纯Inception-v3与类似计算代价的残差网络训练过程中的Top-1错误演化。该评估是在ILSVRC-2012验证集的非黑名单图像上对单一作物进行测量。残差模型的训练速度快得多，但最终精度略低于传统的《盗梦空间3》。

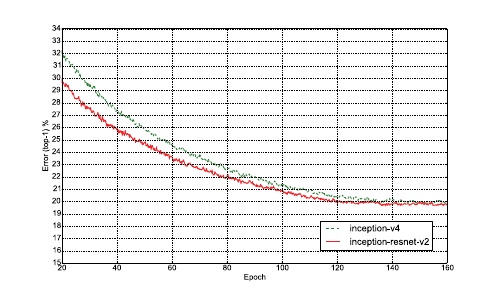


图9:纯启始v3训练期间的Top-1错误演变与类似计算代价的剩余启始。该评估是在ILSVRC-2012验证集的非黑名单图像上的单一作物上进行的。残差版本的训练速度快得多，最终精度略高于传统的Inception- v4。

最后，我们比较了不同版本的Inception和Inception- resnet。模型“盗梦空间v3”和“盗梦空间v4”是深度卷积网络，没有利用剩余连接，而“盗梦空间- ResNet-v1”和“盗梦空间- resnet -v2”是“盗梦空间”类型的网络，利用剩余连接而不是过滤器级联。

4282

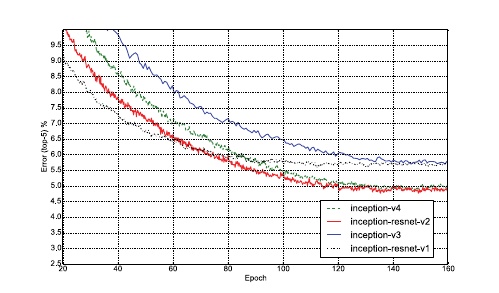


图10:所有四种模型(单模型、单作物)的前5位误差演化。这显示了由于模型尺寸变大所带来的改进。虽然残差版本收敛更快，但最终的精度似乎主要取决于模型的大小。

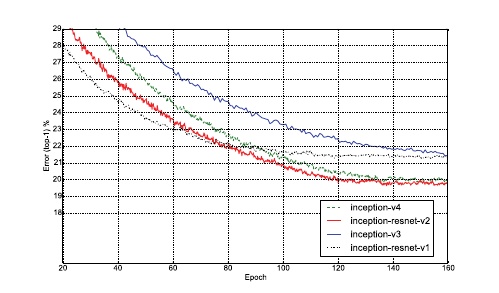


图11:所有四种模型(单一模型、单一裁剪)的Top-1误差演化。这与图10中的前5个评估类似。

|  |  |  |
| --- | --- | --- |
| **网络** | **(错误** | **五大错误** |
| BN-Inception (Ioffe and Szegedy 2015) | 25.2% | 7.8% |
| 《盗梦空间》v3 (Szegedy et al. 2015b) | 21.2% | 5.6% |
| Inception-ResNet-v1 | 21.3% | 5.5% |
| Inception-v4 | 20.0% | 5.0% |
| Inception-ResNet-v2 | 19.9% | 4.9% |

表1:单作物-单模型试验结果。报告了ILSVRC 2012验证集的非黑名单子集。

|  |  |  |  |
| --- | --- | --- | --- |
| **网络** | 作物 | **(错误** | **五大错误** |
| ResNet(他等人，2015) | 10 | 25.2% | 7.8% |
| 《盗梦空间》v3 (Szegedy et al. 2015b) | 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% |

表2:10/12作物评价-单模型试验结果。报告ILSVRC 2012验证集的5万张图片。

|  |  |  |  |
| --- | --- | --- | --- |
| **网络** | 作物 | **(错误** | **五大错误** |
| 《盗梦空间》v3 (Szegedy et al. 2015b) | 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% |

表3:144作物评价-单模型试验结果。报告ILSVRC 2012验证集的5万张图片。相比之下(He et al. 2015) ResNet- 151的Top-1 Error为19.4%，Top-5 Error为4.5%，但采用了不同的“密集”评估策略。

|  |  |  |  |
| --- | --- | --- | --- |
| **网络** | N | **(错误** | **五大错误** |
| ResNet(他等人，2015) | 6 | N/A | 3.6% |
| 《盗梦空间》v3 (Szegedy et al. 2015b) | 4 | 17.3% | 3.6% |
| Inception-v4(+剩余) | 4 | 16.4% | 3.1% |

表4:144茬/密集评价的集成结果。报告ILSVRC 2012验证集的5万张图片。第二列(N)表示集成了多少个模型。对于Inception- v4(+Residual)，集成由一个纯粹的Inception- v4和三个Inception- resnet -v2模型组成，并在验证和测试集上进行了评估。测试集的性能是3.08%的前5位错误验证，我们没有在验证集上过拟合。

表1显示了验证集上各种体系结构的单模型、单作物top-1和top-5误差。

表2显示了具有少量作物的各种模型的性能:ResNet有10种作物(He et al. 2015)，对于Inception的变体，我们使用了12种作物的评估，如(Szegedy et al. 2015a)所述。

表3显示了使用144种作物评价的各种模型的单模型性能。与报告的19.4%的前1名和4.5%的前5名错误率相比(He et al. 2015)，这些是有竞争力的，尽管“密集”评估策略略有不同。对于最初的网络，如(Szegedy et al. 2015a)所述，使用了144种作物策略。

表4比较了集成结果。对于纯残差网络，6模型密集评价结果来自(He et al. 2015)。对于初始网络，使用144种作物策略集成了4个模型，如(Szegedy et al. 2015a)所述。

本文中“盗传- resnet -v2”和“盗传-v4”模型的开源实现以及预先训练的权重可以在TensorFlow models github页面上找到:github.com/tensorflow/models。

**结论**

我们详细介绍了三种新的网络架构:

*•盗梦空间- resnet -v1:一个混合的盗梦空间版本，其计算成本与(Szegedy et al. 2015b)的盗梦空间-v3相似。*

*•盗梦空间- resnet -v2:一个更昂贵的盗梦空间混合版本*

4283

显著提高了识别性能。*•盗梦空间v4:一个纯盗梦空间变体，没有残留连接，识别性能与盗梦空间resnet -v2大致相同。*

我们研究了剩余连接的引入如何极大地提高了先启体系结构的训练速度。我们最新的模型(包括和不包括剩余连接)的性能优于所有之前的网络，这仅仅是由于模型规模的增加，同时与其他竞争方法相比，保持了参数的总数和计算成本。

**参考文献**

Abadi m;阿加瓦尔,a;巴勒p;Brevdo大肠;陈,z;雪铁龙,c;柯拉g s;戴维斯,a;院长,j .;Devin, m;格玛沃特,美国;格拉汉姆·古德费勒,即;竖琴,a;欧文·g·;Isard m;贾,y;Jozefowicz r;凯撒,l;Kudlur m;Levenberg, j .;男人´e、d;《艋舺》的r;摩尔,美国;穆雷,d;Olah c;舒斯特尔,m;Shlens, j .;施泰纳,b;Sutskever i;三、k;塔克,p;Vanhoucke诉;Vasudevan诉;f . Vi´起飞;Vinyals o .;监狱长,p;瓦滕伯格,m;Wicke m;Yu, y;郑晓。2015。TensorFlow:异构系统上的大规模机器学习。软件可从tensorflow.org获得。

院长,j .;柯拉g;《艋舺》的r;陈,k;Devin, m;毛,m;高级,a;塔克,p;杨,k;勒,问:诉;et al . 2012。大规模分布式深层网络。神经信息处理系统进展，1223-1231。

盾,c;阿来,c . c;他,k;唐晓。2014。学习一种用于图像超分辨率的深度卷积网络。在计算机视觉- eccv 2014。施普林格。184 - 199。

福岛,k . 1980。新认知:一种不受位置变化影响的模式识别机制的自组织神经网络模型。*生物控制论36(4):193 - 202。*

Girshick r;多纳休,j .;达雷尔t;Malik, J. 2014。丰富的特征层次精确的对象检测和语义分割。计算机视觉与模式识别(CVPR)会议论文集。他,k;张,x;任,美国;孙洁2015。用于图像识别的深度残差学习。*arXiv预印本arXiv: 1512.03385。*

Ioffe, S.和Szegedy, C. 2015。批归一化:通过减少内部协变量偏移来加速深度网络训练。第32届机器学习国际会议论文集，448-456。

Karpathy, a;Toderici g;Shetty,美国;梁、t;Suk - thankar r;fei fei, L. 2014。基于卷积神经网络的大规模视频分类。计算机视觉与模式识别(CVPR)IEEE。

Krizhevsky, a;Sutskever i;Hinton, g.e.， 2012。基于深度卷积神经网络的Imagenet分类。神经信息处理系统进展，1097-1105。

LeCun (y;波沙,b;德克,j·s·;亨德森,d;霍华德,r . e .;哈伯德,w;和杰克，L. D. 1989。反向传播在手写邮政编码识别中的应用。*神经计算1(4):541 - 551。*

林,m;问:陈;阎s 2013。网络在网络。*arXiv预印本arXiv: 1312.4400。*

长,j .;Shelhamer大肠;和达雷尔，T. 2015。完全卷积网络语义分割。IEEE计算机视觉与模式识别会议论文集，3431-3440。

Pascanu r;Mikolov t;本吉奥，2012年。关于循环神经网络训练的难度。*arXiv预印本arXiv: 1211.5063。*

Russakovsky o .;邓,j .;苏,h;Krause, j .;Satheesh,美国;妈,美国;黄,z;Karpathy, a;科斯拉,a;伯恩斯坦,m;et al . 2014。Imagenet大规模视觉识别挑战。

Simonyan, K.和Zisserman, A. 2014。用于大规模图像识别的深度卷积网络。*arXiv预印本arXiv: 1409.1556。*

Sutskever i;Martens j .;达尔,g;Hinton, G. 2013。关于初始化和动量在深度学习中的重要性。第30届机器学习国际会议论文集(ICML-13)，第28卷，1139 - 1147。JMLR研讨会和会议记录。

Szegedy c;刘,w;贾,y;Sermanet p;里德,美国;Anguelov d;Erhan d;Vanhoucke诉;和Rabinovich, A. 2015a。深入研究卷积。IEEE计算机视觉与模式识别会议论文集，1-9。

Szegedy c;Vanhoucke诉;约飞,美国;Shlens, j .;和Wojna, Z. 2015b。重新思考计算机视觉的初始架构。*arXiv预印本arXiv: 1512.00567。*

Tieleman, T.和Hinton, G.将梯度除以它最近大小的运行平均值。COURSERA:机器学习的神经网络，2012年4月。访问:2015 - 11-05。

Toshev, A.和Szegedy, C. 2014。深度姿态:通过深度神经网络估计人体姿态。计算机视觉与模式识别(CVPR)IEEE。

王，N，和杨，d - y。2013.学习用于视觉跟踪的深度压缩图像表示。神经信息处理系统进展，809-817。

4284