# 深层融合[[1]](#footnote-1)

赵淼译

(北京交通大学电子与信息工程学院，电子科学与技术，15650761201)

视觉识别需要丰富的表示（从低层次到高层次，从小型的到大型的，分辨率从高到低）。即使是在卷积网络的深度特征表示中，一个孤立的层并不足够，对于表示的聚集和聚合可以提高对于是什么和在哪里的推断。架构上的努力是多个维度的，包括更深的或更宽的架构，但是怎样对层级和模块进行聚合在一个网络里需要更多的注意力。虽然跳越性连接已经被合并到组合层中，但是这些连接本身就很浅，并且只通过简单的一步操作来融合。我们用更深层的聚合来增强标准体系结构，以便更好地跨层融合信息。我们的深层融合结构是交互性地和层次性地结合层次性特征使网络获得更好的准确率和更少的参数。在架构和任务上的实验显示深层聚合提升了识别和解析力相比于现有的技术。

## 1 介绍

表示学习和迁移学习已经渗透到了计算机视觉中成为了识别的引擎。复合性和可微性成为了一系列深度架构的基础。卷积神经网络作为许多视觉任务的主干，在正确的任务扩展和数据14、35、42中为不同的目的做好准备，使架构的改变成为维持进步的中心驱动力。网络的规模和范围不断扩大，现在正致力于实现（模块和连接模式）的设计模式，这些模式可以系统地组装起来。这实现了网络更深和更宽，但是如何才能连接的更紧密呢。

更多的非线性，更大的容量和更大的感受野通常会提升准确率，但是对于优化和计算是有问题的。为了解决这一问题，瓶颈模块为了维度减少，残差，门机制，连接操作对于特征和梯度传播。网络设计可以衍生出大量的模式。

尽管如此，进一步的探索应该用于如何连接这些层和模块。层次性网络从LeNet，AlexNet到ResNet序列化的堆积了一些层。层级准确率比较，迁移分析和表示可视化显示了更深的层级可以提取更多语义性的信息和更全局的信息，但是这不意味着最后一层对于任何任务都是都是最终的表示，事实上，跳跃性连接已经证明是高效的对于分类和回归和更结构化的任务。深度和宽度的聚合对于结构是关键的维度。

在这项工作中，我们调查了如何聚合各层以获得更好的语义信息和空间信息在识别和定位上。对于现在的浅的跳跃式连接，我们的聚合结构包含了更多的深度和分享性。我们引入了两种结构（DLA）：一种是交互式深层聚合（IDA）,另一种是层次化深层聚合（HAD）。这些结构通过网络架构进行解释，独立于主干的选择，对于现在和未来的网络都具有兼容性。IDA集中于融合分辨率和尺度而HDA集中于结合各模块和各通道的特征。HAD跟随着基本的层次根据阶段重新定义了分辨率和聚合尺度。HAD集合了自己的树形结构连接跨阶段和融合各阶段聚合不同层次的表示。

我们的实验评估了跨标准架构和任务的深层聚合，以扩展ResNet和ResNeXt ，用于大规模图像分类、细粒度识别、语义细分和边界检测。我们的结果显示了性能、参数计数和内存使用情况的改进，超过了基线ResNet、ResNeXT和DenseNet体系结构。DLA在分类模型中实现了最先进的结果。如果没有进一步的架构，相同的网络就会在几个细粒度的识别基准上获得最先进的结果。通过标准技术对结构化输出进行重铸，DLA在城市景观的语义划分上达到了最佳的精确度，在PASCAL边界上的边界检测是最先进的。深层聚合是对深层视觉架构的一般和有效的扩展。

## 2 相关工作

我们回顾了视觉识别的架构，突出了层次结构和金字塔尺度的关键架构，并将它们与我们对深度、规模和分辨率的深度聚合的关注联系起来。

这里的架构与最先进的层次化混合特征相关联。混合的关键是语义上的和空间上的。语义混合是跨通道的和深度的，提升了对是什么的理解。空间上的融合提升了对分辨率和范围的融合，提升了对在哪里的理解。深层融合可以看作两者的结合。

DenseNet系的网络是语义混合的主要架构。设计用来更好地传播特性和损失，通过跳过连接，将所有的层连接在一起。我们的分层深度聚合对短路径和重用的重要性有着相同的见解，并扩展了跳跃连接使用款阶段和分层次的树形连接而不是简单的串联。密集连接和深度聚合的网络实现了更高的准确性，更好的参数和内存效率。

特征金字塔网络是主流的空间融合架构，设计后通过在金字塔特征层次结构的自下而上的和横向连接实现了对分辨率和标准化语义的权衡。我们的交互式深度融合同样提升了分辨率。FPN连接是线性的和前面的层不能被更多的聚合以抵消他们语义相对弱化的缺点。金字塔和深度融合网络更好的解决了结构化输出任务的内容和位置。

## 3 深层融合

我们定义融合为同意同一网络不同层之间的整合。在这项工作中我们集中于一系列高效的深度，分辨率和尺寸融合的的架构。我们成称一种融合为深度的当这种融合是组合式的，非线性的和在更浅的网络层通过多种融合方式进行融合。由于网络可以包含许多层和连接，模块化设计有助于通过分组和重复来对抗复杂性。网络层分组到模块，然后通过特征的分辨率分组到各个阶段。我们关心的是模块和阶段的融合。

### 迭代式深度融合

迭代式深度融合通过骨骼架构之间的反复堆叠实现。我们将堆叠的模块根据特征分辨率分成阶段。更深的阶段更具有语义特性但是空间上更为粗粒度。跳跃式的连接从浅层到深层对尺度和分辨率进行了合并。然而已有的工作，FCN，U-Net，FPN是线性的，浅层信息很少被融合，作者这里建议逐渐深化的融合，从浅层开始以最浅的和最小规模开始融合，然后迭代的进行更大规模的特征融合。这种方式下浅层特征在不同阶段被重新定义。图2（c）显示了这种结构。

、

表达式为：



### 层次化深度融合

层次化的深度融合将模块和阶段结合成树形结构取保留和组合特征通道。通过浅层和深层的组合学习到了更丰富的组合，扩展了特征的层次。虽然IDA高效的组合了阶段，但是它对一个网络许多模块之间的高效融合仍然不足够，因为它是序列化的结构。深层的，分支结构的层次融合如下图：



这里进一步增加其深度和效率，将每个融合结点的输出作为下一阶段的输入，得到下图。然后继续将同一层次的融合结点合并获得统一的融合结点，如下图。



|  |  |
| --- | --- |
|  | 1 |

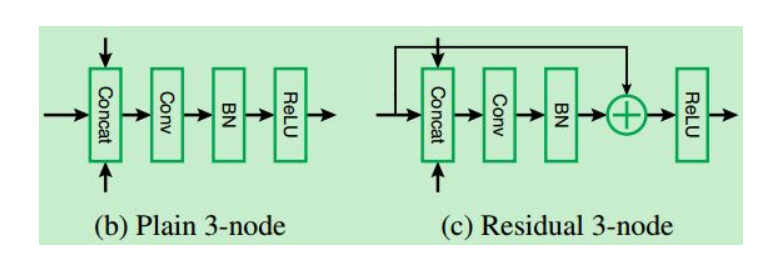
式中代表深度为n的层次化融合结点，R和L如下定义。

|  |  |
| --- | --- |
|  | 1 |
|  |  |

B代表卷积模块

### 架构元素

聚合结点：一个聚合结点的主要功能式组合和压缩输入。这个结点会选择和组合一些重要的信息，并且保持输出和输入是相同的维度。在IDA中总是两个输入，而HAD的参数数目取决于树的深度。



虽然融合结点可以基于任何模块和层，但是为了简单和效率，我们选择了一个单一的卷积层带有批归一化和非线性操作。在图像分类网络中，所有的结点使用1\*1的卷积。在语义分割上，我们增加了一个更深层次的递进式的深度聚合插入特征，使用了3\*3的卷积。

残差连接对于深度网络的聚合是重要的，这里也进行了包括，但不清楚是否必要。对于HDA来说任何模块到达主干的最短路径就是层次的深度，因此在不会出现衰减得或是爆炸性的增长。在实验中发现残差连接在层次超过4时会对HAD有帮助，在层次数更少时会有害。前面N就可以如下式定义

|  |  |
| --- | --- |
|  | 1 |
| 带残差的 |  |

**模块和阶段**：这里对于模块和阶段的内部结构没有要求。这里实验了三种类型的残差模块。基本的模块组合了堆叠的卷积和直接的跳跃性连接。瓶颈模块对卷积堆叠进行了正则化，减少了维度通过1\*1卷积。分割模块增加了特征的多样性通过对将通道组合成一些单独的路径。

## 4 应用

### 1.分类结构

我们的分类网络对于ResNet进行了增强使用IDA和HAD。这些是阶段性网络。每个阶段降低一半的分辨率。第一阶段保持输入分辨率，最后阶段31倍下采样。最后特征图使用全局平均池化进行。我们通过IDA对阶段间进行联系。和在阶段内补和阶段间使用了HAD。这种聚合简单的加入了聚合结点。在这种情况下简单的在每个阶段使用等式1和2.我们的阶段使用了最大池化，size为2，步幅为2.

早期的阶段由他们自己的结构。借鉴了DRN的方法，在阶段一和到2使用步幅卷积代替最大池化。阶段1是7\*7卷积组成。阶段2是一个基本模块。对所有其他模块，都是用了IDA和HAD的结合。

### 密集预测网络

使用插值的IDA通过投影和上采样既增加了深度又增加了分辨率。



所有的投影和上采样的参数可以在网络优化过程中联合学习。上采样通过双线性内插的方法按照论文35进行学习。我们对阶段3到6的输出投影到32个通道，内插这些阶段到相同的分辨率和阶段2.最后迭代的聚合这些阶段的学习一个高层次和深层次的深度融合。虽然有与FCN跳越连接35、超列特性15和FPN自顶向下连接30的相同目的，但我们的聚合在方法上有所不同，从浅到深到进一步细化特性。请注意，在这种情况下，我们使用了IDA两次：一次在主干网络中连接各个阶段，第二次恢复分辨率。

## 5 结果

结果更好，参数更少。

在分类，语义分割等领域取得了更好的效果。

1. 赵淼 电信研1706班 电子科学与技术 17120036 15650761201 [↑](#footnote-ref-1)