Deep Layer Aggregation
深层聚合

日期：2018-01-04

作者：Fisher Yu (/search?search_txt=Fisher Yu) 、Dequan Wang (/search?search_txt=Dequan Wang) 、Evan Shelhamer (/search?search_txt=Evan Shelhamer) 、Trevor Darrell (/search?search_txt=Trevor Darrell)

论文：http://arxiv.org/pdf/1707.06484v2.pdf (http://arxiv.org/pdf/1707.06484v2.pdf)

Abstract
摘要

Visual recognition requires rich representations that span levels from low to high, scales from small to large, and resolutions from fine to coarse. Even with the depth of features in a convolutional network, a layer in isolation is not enough: compounding and aggregating these representations improves inference of what and where. Architectural efforts are exploring many dimensions for network backbones, designing deeper or wider architectures, but how to best aggregate layers and blocks across a network deserves further attention. Although skip connections have been incorporated to combine layers, these connections have been “shallow” themselves, and only fuse by simple, one-step operations. We augment standard architectures with deeper aggregation to better fuse information across layers. Our deep layer aggregation structures iteratively and hierarchically merge the feature hierarchy to make networks with better accuracy and fewer parameters. Experiments across architectures and tasks show that deep layer aggregation improves recognition and resolution compared to existing branching and merging schemes.
视觉识别需要丰富的表现形式，涵盖从低到高的级别，从小到大的级别，以及从细到粗的分辨率。即使卷积网络具有深度的特征，一个孤立的层也是不够的：复合和聚合这些表示可以改进对什么和在哪里的推断。架构工作正在探索网络骨干的许多维度，设计更深或更宽的架构，但如何在网络中最佳地聚合层和块则值得进一步关注。虽然跳过连接已经被合并到层次结合中，但这些连接本身是“浅”的，只能通过简单的一步操作来融合。我们通过更深层次的聚合来增强标准体系结构，以更好地融合各层的信息我们的深层聚合结构迭代地和分层地合并特征层次结构以使网络具有更好的准确性和更少的参数。体系结构和任务的实验表明，与现有的分支和合并方案相比，深层聚合可以提高识别和解决方案。

1. Introduction
1.介绍

Representation learning and transfer learning now permeate computer vision as engines of recognition. The simple fundamentals of compositionality and differentiability give rise to an astonishing variety of deep architectures [23, 39, 37, 16, 47]. The rise of convolutional networks as the backbone of many visual tasks, ready for different purposes with the right task extensions and data [14, 35, 42], has made architecture search a central driver in sustaining progress. The ever-increasing size and scope of networks now directs effort into devising design patterns of modules and connectivity patterns that can be assembled systematically. This has yielded networks that are deeper and wider, but what about more closely connected?
表示学习和转移学习现在已经渗透到计算机视觉中作为识别的引擎。组合性和可微性的简单基础引起了令人惊讶的各种深层架构[23,39,37,16,47]。卷积网络作为许多视觉任务的支柱，随着正确的任务扩展和数据为不同目的而准备

[14,35,42]，使得架构搜索成为维持进步的中心动力。现在，网络规模和范围的不断扩大正在指导设计模块和连接模式的设计模式，以便系统地进行组装。这产生了更深更广的网络，但更紧密的联系呢？

More nonlinearity, greater capacity, and larger receptive fields generally improve accuracy but can be problematic for optimization and computation. To overcome these bar

更多的非线性，更大的容量和更大的接收域通常可以提高准确性，但对于优化和计算可能会产生问题。克服这些障碍

1 riers, different blocks or modules have been incorporated to balance and temper these quantities, such as bottlenecks for dimensionality reduction [29, 39, 17] or residual, gated, and concatenative connections for feature and gradient propagation [17, 38, 19]. Networks designed according to these schemes have $100+$ and even $1000+$ layers.

为了平衡和调节这些数量，已经引入了不同的模块或模块来平衡和调节这些数量，例如降维的瓶颈[29,39,17]或特征和梯度传播的残差，门控和连接连接[17,38,19] 。根据这些方案设计的网络具有$100+$甚至$1000+$层。

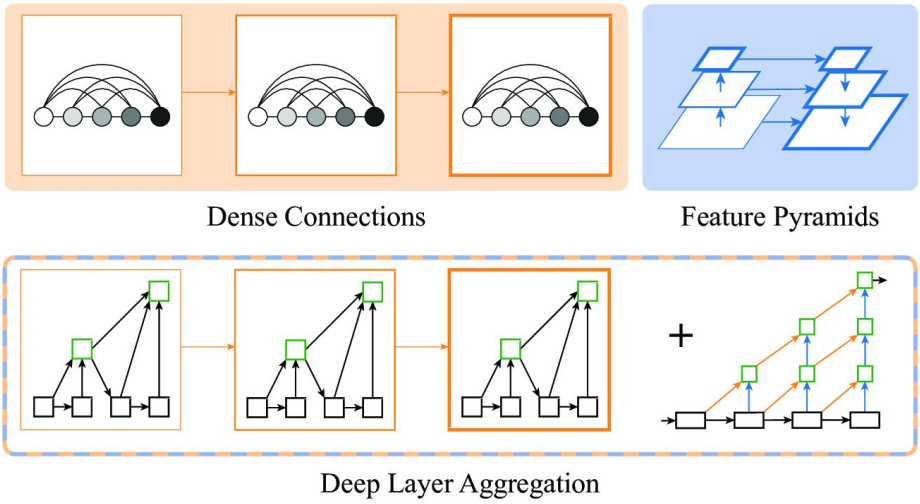


Figure 1: Deep layer aggregation unifies semantic and spatial fusion to better capture what and where. Our aggregation architectures encompass and extend densely connected networks and feature pyramid networks with hierarchical and iterative skip connections that deepen the representation and refine resolution.

图l：深层聚合统一语义和空间融合，以更好地捕捉什么和在哪里。我们的聚合体系结构包含并扩展了密集连接的网络，并通过分层和迭代跳过连接来构建金字塔网络，从而加深了表示和精确解析。

Nevertheless, further exploration is needed on how to connect these layers and modules. Layered networks from LeNet [26] through AlexNet [23] to ResNet [17] stack layers and modules in sequence. Layerwise accuracy comparisons [11, 48, 35], transferability analysis [44], and representation visualization [48, 46] show that deeper layers extract more semantic and more global features, but these signs do not prove that the last layer is the ultimate representation for any task. In fact, skip connections have proven effective for classification and regression [19, 4] and more structured tasks [15, 35, 30]. Aggregation, like depth and width, is a critical dimension of architecture.

然而，如何连接这些层和模块需要进一步的探索。分层网络从LeNet [26]到AlexNet [23]到ResNet [17]堆栈层和模块。分层精度比较[11,48,35]，可转移性分析[44]和表示可视化[48,46]表明，更深的层提取更多的语义和更全局的特征，但这些迹象并不能证明最后一层是最终的代表任何任务。事实上，跳过连接已被证明是有效的分类和回归[19,4]和更结构化的任务[15,35,30]。聚合，如深度和宽度，是架构的关键维度。

In this work, we investigate how to aggregate layers to better fuse semantic and spatial information for recognition and localization. Extending the “shallow” skip connections of current approaches, our aggregation architectures incor porate more depth and sharing. We introduce two structures for deep layer aggregation (DLA): iterative deep aggregation (IDA) and

hierarchical deep aggregation (HDA). These structures are expressed through an architectural framework, independent of the choice of backbone, for compatibility with current and future networks. IDA focuses on fusing resolutions and scales while HDA focuses on merging features from all modules and channels. IDA follows the base hierarchy to refine resolution and aggregate scale stage-bystage. HDA assembles its own hierarchy of tree-structured connections that cross and merge stages to aggregate different levels of representation. Our schemes can be combined to compound improvements.
在这项工作中，我们研究如何聚合图层以更好地融合用于识别和定位的语义和空间信息。通过扩展当前方法的“浅层”跳转连接，我们的聚合体系结构更加深入和共享。我们介绍两种用于深层聚合（DLA）的结构：迭代深度聚合（IDA）和分层深度聚合（HDA）。这些结构通过独立于骨干选择的架构框架来表达，以便与当前和未来网络兼容。IDA专注于融合解决方案和规模，而HDA专注于合并来自所有模块和渠道的功能。国际开发协会遵循基础层次结构，逐步改进解决方案和聚合规模阶段。HDA组装自己的树形结构连接层次结构，交叉和合并阶段以聚合不同级别的表示。我们的方案可以结合起来，以改进复合。

Our experiments evaluate deep layer aggregation across standard architectures and tasks to extend ResNet [16] and ResNeXt [41] for large-scale image classification, finegrained recognition, semantic segmentation, and boundary detection. Our results show improvements in performance, parameter count, and memory usage over baseline ResNet, ResNeXT, and DenseNet architectures. DLA achieve stateof-the-art results among compact models for classification. Without further architecting, the same networks obtain stateof-the-art results on several fine-grained recognition benchmarks. Recast for structured output by standard techniques, DLA achieves best-in-class accuracy on semantic segmentation of Cityscapes [8] and state-of-the-art boundary detection on PASCAL Boundaries [32]. Deep layer aggregation is a general and effective extension to deep visual architectures.
我们的实验通过标准体系结构和任务来评估深层聚合，以扩展ResNet [16]和ResNeXt [41]的大规模图像分类，融合识别，语义分割和边界检测。我们的结果显示，与基准ResNet，ResNeXT和DenseNet体系结构相比，性能，参数数量和内存使用量有所提高。DLA在紧凑模型中实现了最先进的分类结果。如果没有进一步的架构，相同的网络可以在几个细粒度的识别基准上获得最先进的结果。通过标准技术重新构造结构化输出，DLA在城市风景的语义分割[8]和PASCAL边界上的最先进的边界检测方面实现了同类最佳的准确性[32]。深层聚合是对深层视觉体系结构的一般和有效的扩展。

2. Related Work
2.相关工作

We review architectures for visual recognition, highlight key architectures for the aggregation of hierarchical features and pyramidal scales, and connect these to our focus on deep aggregation across depths, scales, and resolutions.
我们回顾了用于视觉识别的体系结构，重点介绍了用于聚合分层特征和金字塔尺度的关键架构，并将这些架构与我们关注深度，尺度和分辨率深度聚合的重点相关联。

The accuracy of AlexNet [23] for image classification on ILSVRC [34] signalled the importance of architecture for visual recognition. Deep learning diffused across vision by establishing that networks could serve as backbones, which broadcast improvements not once but with every better architecture, through transfer learning [11, 48] and metaalgorithms for object detection [14] and semantic segmentation [35] that take the base architecture as an argument. In this way GoogLeNet [39] and VGG [39] improved accuracy on a variety of visual problems. Their patterned components prefigured a more systematic approach to architecture.
AlexNet [23]对ILSVRC图像分类的准确性[34]标志着体系结构对视觉识别的重要性。通过建立网络可以作为骨干网的深度学习，通过传递学习[11,48]和用于目标检测[14]和语义分割[35]基础架构作为参数。通过这种方式，GoogLeNet [39]和VGG [39]提高了各种视觉问题的准确性。他们的图案化组件预先构建了一个更系统化的架构方法。

Systematic design has delivered deeper and wider networks such as residual networks (ResNets) [16] and highway networks [38] for depth and ResNeXT [41] and FractalNet [25] for width. While these architectures all contribute their own structural ideas, they incorporated bottlenecks and shortened paths inspired by earlier techniques. Network-innetwork [29] demonstrated channel mixing as a technique to fuse features, control dimensionality, and go deeper. The companion and auxiliary losses of deeply-supervised networks [27] and GoogLeNet [39] showed that it helps to keep learned layers and losses close. For the most part these architectures derive from innovations in connectivity: skipping, gating, branching, and aggregating.
系统设计提供了更深更广的网络，例如残余网络（ResNets）[16]和高速公路网[38]以及ResNeXT [41]和FractalNet [25]。虽然这些架构都贡献了自己的结构思想，但它们融合了先前技术所带来的瓶颈和缩短的路径。网络内网[29]展示了信道混合作为一种融合特征，控制维度并进一步深入的技术。深度监督网络[27]和GoogLeNet [39]的伴随和辅助损失表明，它有助于保持学习层次和损失的接近。这些体系结构大部分来自连接创新：跳过，门控，分支和聚合。

Our aggregation architectures are most closely related to leading approaches for fusing feature hierarchies. The key axes of fusion are semantic and spatial. Semantic fusion, or aggregating across channels and depths, improves inference of what. Spatial fusion, or aggregating across resolutions and scales, improves inference of where. Deep layer aggregation can be seen as the union of both forms of fusion.
我们的聚合体系结构与融合功能层次结构的领先方法关系最密切。融合的关键轴是语义和空间的。语义融合或跨渠道和深度聚合，可以提高对内容的推断。空间融合或跨分辨率和尺度的聚合，改善了对何处的推断。深层聚合可以被看作两种融合形式的结合。

Densely connected networks (DenseNets) [19] are the dominant family of architectures for semantic fusion, designed to better propagate features and losses through skip connections that concatenate all the layers in stages. Our hierarchical deep aggregation shares the same insight on the importance of short paths and re-use, and extends skip connections with trees that cross stages and deeper fusion than concatenation. Densely connected and deeply aggregated networks achieve more accuracy as well as better parameter and memory efficiency.
密集连接网络（DenseNets）[19]是语义融合架构的主要家族，旨在通过跳过连接更好地传播特征和损失，连接所有层级。我们的分层深度聚合共享关于短路径和重用的重要性的相同洞察，并延伸与跨阶段和更深层融合而不是级联的树的跳过连接。密集连接和深度聚合的网络实现更高的精度以及更好的参数和内存效率。

Feature pyramid networks (FPNs) [30] are the dominant family of architectures for spatial fusion, designed to equalize resolution and standardize semantics across the levels of a pyramidal feature hierarchy through top-down and lateral connections. Our iterative deep aggregation likewise raises resolution, but further deepens the representation by nonlinear and progressive fusion. FPN connections are linear and earlier levels are not aggregated more to counter their relative semantic weakness. Pyramidal and deeply aggregated networks are better able to resolve what and where for structured output tasks.
特征金字塔网络（FPNs）[30]是空间融合架构的主要家族，旨在通过自顶向下和横向连接均衡分辨率和标准化金字塔特征层级的语义。我们的迭代深度聚合同样提高了分辨率，但通过非线性和渐进式融合进一步加深了表示。FPN连接是线性的，而较早的级别不会更多地聚合以应对其相对语义上的弱点。金字塔和深度聚合的网络能够更好地解决结构化输出任务的内容和位置。

3. Deep Layer Aggregation
3.深层聚合

We define aggregation as the combination of different layers throughout a network. In this work we focus on a family of architectures for the effective aggregation of depths, resolutions, and scales. We call a group of aggregations deep if it is compositional, nonlinear, and the earliest aggregated layer passes through multiple aggregations.

我们将聚合定义为整个网络中不同层的组合。在这项工作中，我们将重点放在一系列架构上，以有效聚合深度，分辨率和尺度。如果它是合成的，非线性的，并且最早的聚合层通过多个聚合，我们称它为深度聚合。

As networks can contain many layers and connections, modular design helps counter complexity by grouping and repetition. Layers are grouped into blocks, which are then grouped into stages by their feature resolution. We are concerned with aggregating the blocks and stages.

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应对复杂性。图层被分组为块，然后按照它们的特征分辨率将其分组。我们关心的是聚合块和阶段。

3.1. Iterative Deep Aggregation

3.1。迭代深度聚合

Iterative deep aggregation follows the iterated stacking of the backbone architecture. We divide the stacked blocks of the network into stages according to feature resolution. Deeper stages are more semantic but spatially coarser. Skip connections from shallower to deeper stages merge scales and resolutions. However, the skips in existing work, e.g. FCN [35], U-Net [33], and FPN [30], are linear and aggregate the shallowest layers the least, as shown in Figure 2(b). We propose to instead progressively aggregate and deepen the representation with IDA. Aggregation begins at the shallowest, smallest scale and then iteratively merges deeper, larger scales. In this way shallow features are refined as they are propagated through different stages of aggregation. Figure 2(c) shows the structure of IDA.

迭代深度聚合遵循骨干架构的迭代堆叠。我们根据特征分辨率将网络的堆叠块分成几个阶段。更深的阶段是更多的语义，但空间更粗糙。跳过从较浅到较深阶段的连接合并比例和分辨率。然而，现有工作中的跳过，例如如图2（b）所示，FCN [35]，U-Net [33]和FPN [30]是线性的，聚集最浅的层。我们建议逐步汇总和深化与IDA的代表。聚合始于最浅，最小的尺度，然后迭代合并更深，更大的尺度。通过这种方式，浅层特征在通过不同阶段的聚合进行传播时被改进。图2（c）显示了IDA的结构。

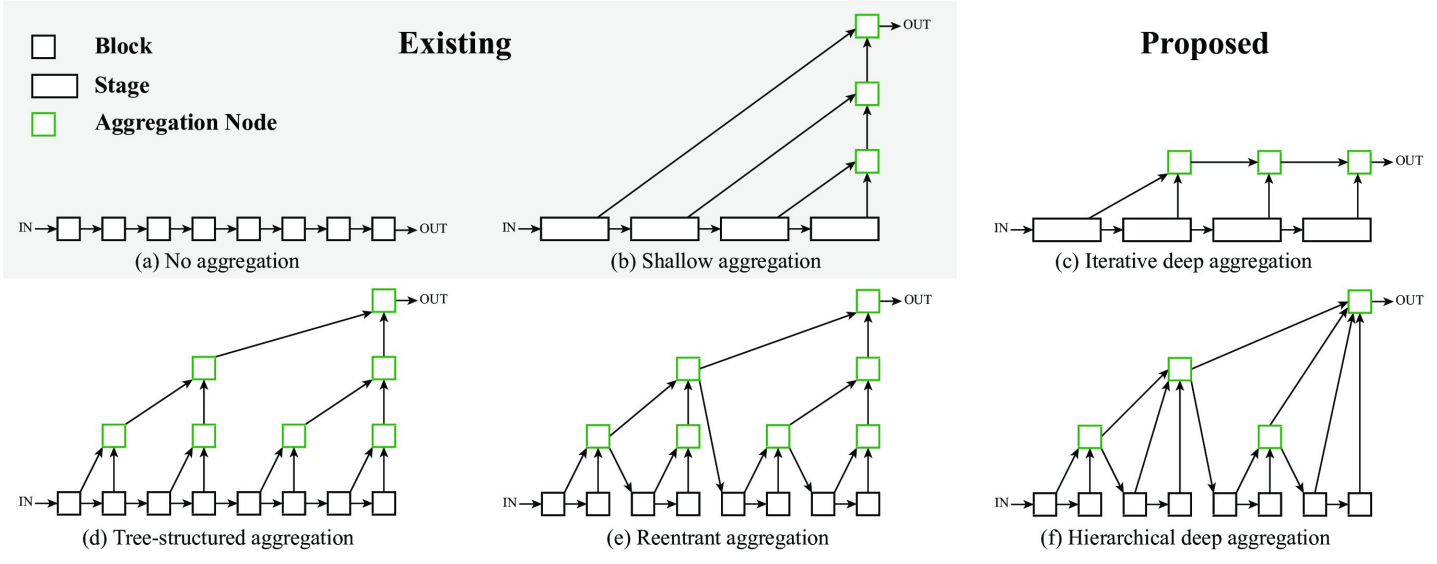


Figure 2: Different approaches to aggregation. (a) composes blocks without aggregation as is the default for classification and regression networks. (b) combines parts of the network with skip connections, as is commonly used for tasks like segmentation and detection, but does so only shallowly by merging earlier parts in a single step each. We propose two deep aggregation architectures: (c) aggregates iteratively by reordering the skip connections of (b) such that the shallowest parts are aggregated the most for further processing and (d) aggregates hierarchically through a tree structure of blocks to better span the feature hierarchy of the network across different depths. (e) and (f) are refinements of (d) that deepen aggregation by routing intermediate

aggregations back into the network and improve efficiency by merging successive aggregations at the same depth. Our experiments show the advantages of (c) and (f) for recognition and resolution.

图2：不同的聚合方法。 （a）构成没有聚合的块，这是分类和回归网络的默认值。 （b）将网络的各个部分与跳过连接结合起来，就像通常用于分割和检测这样的任务一样，但是这样做只能通过每个步骤合并较早的部分来进行。我们提出了两种深度聚合体系结构： （c）通过对（b）的跳过连接进行重新排序来迭代聚合，使得最浅部分聚集得最多以用于进一步处理，以及（d）通过块的树结构分层聚合以更好地跨越特征跨越不同深度的网络层次结构。 （e）和（f）是（d）的细节，通过将中间聚合路由回网络来加深聚合，并通过合并相同深度的连续聚合来提高效率。我们的实验显示了（c）和（f）对识别和解析的优点。

The iterative deep aggregation function I for a series of layers $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with increasingly deeper and semantic information is formulated as

针对具有越来越深和语义信息的一系列层$\mathbf{x}_1, \ldots, \mathbf{x}_n$的迭代深度聚合函数I被表述为

$$I(\mathbf{x}_1, ..., \mathbf{x}_n) = \begin{cases} \mathbf{x}_1 & \text{if } n = 1 \\ I(N(\mathbf{x}_1, \mathbf{x}_2), ..., \mathbf{x}_n) & \text{otherwise,} \end{cases} \quad (1)$$

where N is the aggregation node.

其中N是汇聚节点。

3.2. Hierarchical Deep Aggregation

3.2。分层深度聚合

Hierarchical deep aggregation merges blocks and stages in a tree to preserve and combine feature channels. With HDA shallower and deeper layers are combined to learn richer combinations that span more of the feature hierarchy. While IDA effectively combines stages, it is insufficient for fusing the many blocks of a network, as it is still only sequential. The deep, branching structure of hierarchical aggregation is shown in Figure 2(d).

分层深度聚合合并树中的块和阶段以保留和组合特征通道。借助HDA，将浅层和深层结合起来学习更丰富的组合，从而跨越更多的要素层次结构。虽然IDA有效地结合了各个阶段，但融合网络的众多模块并不足够，因为它仍然只是顺序的。图2（d）显示了分层聚合的深层分支结构。

Having established the general structure of HDA we can improve its depth and efficiency. Rather than only routing intermediate aggregations further up the tree, we instead feed the output of an aggregation node back into the backbone as the input to the next sub-tree, as shown in Figure 2(e). This propagates the aggregation of all previous blocks instead of the preceding block alone to better preserve features. For efficiency, we merge aggregation nodes of the same depth (combining the parent and left child), as shown in Figure 2(f). The hierarchical deep aggregation function T_n, with depth n, is formulated as

建立HDA的一般结构后，我们可以提高其深度和效率。我们不是只将中间聚合路由到树的更远处，而是将聚合节点的输出作为下一个子树的输入返回到主干中，如图2（e）所示。这会传播所有先前块的聚合而不是单独的前一个块，以更好地保留特征。为了提高效率，我们合并相同深度的汇聚节点（结合父节点和左侧节点），如图2（f）所示。具有深度n的分层深度聚合函数T_n被表达为

$$T_n(\mathbf{x}) = N(R^n_{n-1}(\mathbf{x}), R^n_{n-2}(\mathbf{x}), ..., R^n_1(\mathbf{x}), L^n_1(\mathbf{x}), L^n_2(\mathbf{x})), \quad (2)$$

where N is the aggregation node. R and L are defined as

其中N是汇聚节点。 R和L被定义为

$$L_2^n(\mathbf{x}) = B(L_1^n(\mathbf{x})), \quad L_1^n(\mathbf{x}) = B(R_1^n(\mathbf{x})),$$
$$R_m^n(\mathbf{x}) = \begin{cases} T_m(\mathbf{x}) & \text{if } m = n-1 \\ T_m(R_{m+1}^n(\mathbf{x})) & \text{otherwise,} \end{cases}$$

where B represents a convolutional block.

其中B代表卷积块。

3.3. Architectural Elements

3.3。建筑元素

Aggregation Nodes The main function of an aggregation node is to combine and compress their inputs. The nodes learn to select and project important information to maintain the same dimension at their output as a single input. In our architectures IDA nodes are always binary, while HDA nodes have a variable number of arguments depending on the depth of the tree.

聚合节点聚合节点的主要功能是合并和压缩其输入。节点学会选择和投影重要信息，以便在输出上保持与单个输入相同的维度。在我们的体系结构中，IDA节点始终是二进制的，而HDA节点具有可变数量的参数，具体取决于树的深度。

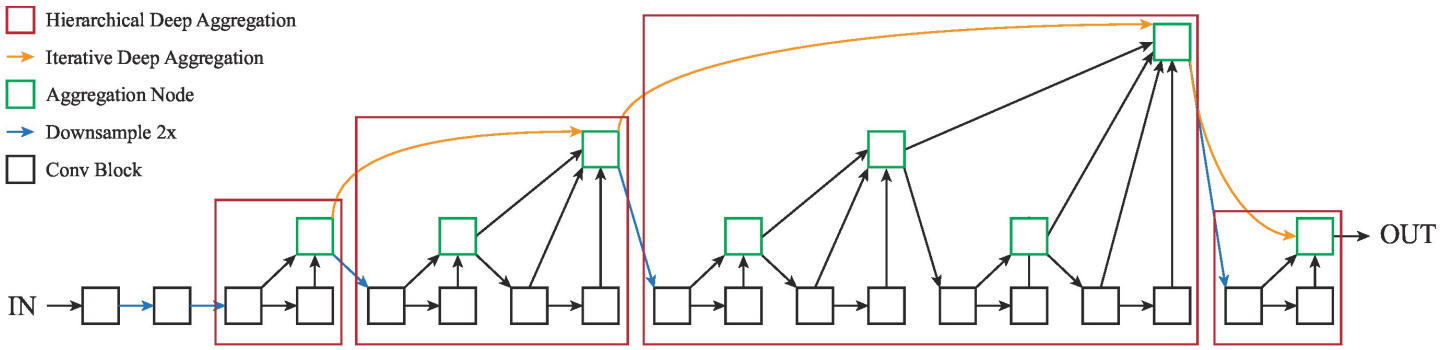


Figure 3: Deep layer aggregation learns to better extract the full spectrum of semantic and spatial information from a network. Iterative connections join neighboring stages to progressively deepen and spatially refine the representation. Hierarchical connections cross stages with trees that span the spectrum of layers to better propagate features and gradients.

图3：深层聚合学习更好地从网络中提取全部语义和空间信息。迭代连接与相邻阶段相结合，逐步加深并在空间上改善表示。分层连接跨越阶段，跨越各层的树以更好地传播特征和渐变。

Although an aggregation node can be based on any block or layer, for simplicity and efficiency we choose a single convolution followed by batch normalization and a nonlinearity. This avoids overhead for aggregation structures. In image classification networks, all the nodes use 1×1 convolution. In semantic segmentation, we add a further level of iterative deep aggregation to interpolate features, and in this case use 3×3 convolution.

虽然聚合节点可以基于任何块或层，但为了简单和有效，我们选择单卷积，然后进行批量归一化和非线性。这避免了聚合结构的开销。在图像分类网络中，所有节点都使用1×1卷积。在语义分割中，我们增加更深层次的迭代深度聚合来插入特征，在这种情况下使用3×3卷积。

As residual connections are important for assembling very deep networks, we can also include residual connections in our aggregation nodes. However, it is not immediately clear that they are necessary for aggregation. With HDA, the shortest path from any block to the root is at most the depth of the hierarchy, so diminishing or exploding gradients may not appear along the

aggregation paths. In our experiments, we find that residual connection in node can help HDA when the deepest hierarchy has 4 levels or more, while it may hurt for networks with smaller hierarchy. Our base aggregation, i.e. N in Equation 1 and 2, is defined by:

由于剩余连接对于组装非常深的网络很重要，因此我们还可以在聚合节点中包含剩余连接。不过，目前尚不清楚他们是否需要进行汇总。使用HDA时，从任何块到根的最短路径至多是层次结构的深度，因此沿着聚合路径可能不会出现递减或爆炸渐变。在我们的实验中，我们发现当最深层次结构有4级或更多级别时，节点中的剩余连接可以帮助HDA，但对于层级较小的网络可能会造成伤害。我们的基础聚合，即等式1和2中的N，定义为：

$$N(\mathbf{x}_1, ..., \mathbf{x}_n) = \sigma(\text{BatchNorm}(\sum_i W_i \mathbf{x}_i + \mathbf{b})), \quad (3)$$

where σ is the non-linear activation, and $\mathbf{w}_i$ and b are the weights in the convolution. If residual connections are added, the equation becomes

其中σ是非线性激活，$\mathbf{w}_i$和b是卷积中的权重。如果添加剩余连接，则等式变为

$$N(\mathbf{x}_1, ..., \mathbf{x}_n) = \sigma(\text{BatchNorm}(\sum_i W_i \mathbf{x}_i + \mathbf{b}) + \mathbf{x}_n). \quad (4)$$

Note that the order of arguments for N does matter and should follow Equation 2.

请注意，N的自变量顺序很重要，应该遵循方程2。

Blocks and Stages Deep layer aggregation is a general architecture family in the sense that it is compatible with different backbones. Our architectures make no requirements of the internal structure of the blocks and stages.

块和阶段深层聚合是一种通用架构系列，因为它与不同的骨干兼容。我们的架构对街区和街区的内部结构没有要求。

The networks we instantiate in our experiments make use of three types of residual blocks [17, 41]. Basic blocks combine stacked convolutions with an identity skip connection. Bottleneck blocks regularize the convolutional stack by reducing dimensionality through a 1×1 convolution. Split blocks diversify features by grouping channels into a number of separate paths (called the cardinality of the split). In this work, we reduce the ratio between the number of output and intermediate channels by half for both bottleneck and split blocks, and the cardinality of our split blocks is 32. Refer to the cited papers for the exact details of these blocks.

我们在实验中实例化的网络使用三种类型的残差块[17,41]。基本块将堆积的卷积与身份跳过连接组合在一起。瓶颈块通过1×1卷积减少维度来调整卷积栈。拆分块通过将通道分组为多个单独的路径（称为拆分的基数）来使功能多样化。在这项工作中，我们将瓶颈和拆分块的输出和中间通道数量之比减半，我们拆分块的基数为32。有关这些块的确切细节，请参阅引用的文章。

4. Applications
4.应用程序

We now design networks with deep layer aggregation for visual recognition tasks. To study the contribution of the aggregated representation, we focus on linear prediction without further machinery. Our results do without ensembles for recognition and context modeling or dilation for resolution. Aggregation of semantic and spatial information matters for classification and dense prediction alike.

我们现在设计具有深层聚合的网络来进行视觉识别任务。为了研究汇总表示的贡献，我们专注于线性预测而无需其他机制。我们的结果没有用于识别和上下文建模的合奏或用于解决的扩大。语义和空间信息汇总对于分类和密集预测都很重要。

4.1. Classification Networks
4.1。分类网络

Our classification networks augment ResNet and ResNeXT with IDA and HDA. These are staged networks, which group blocks by spatial resolution, with residual connections within each block. The end of every stage halves resolution, giving six stages in total, with the first stage maintaining the input resolution while the last stage is 32× downsampled. The final feature maps are collapsed by global average pooling then linearly scored. The classification is predicted as the softmax over the scores.
我们的分类网络使用IDA和HDA增强了ResNet和ResNeXT。这些是分段网络，按空间分辨率对块进行分组，每个块内有剩余连接。每个阶段的结束都将分辨率减半，总共有六个阶段，第一阶段保持输入分辨率，而最后阶段是32倍下采样。最终的特征地图被全球平均汇总折叠，然后进行线性评分。分类预测为分数上的softmax。

We connect across stages with IDA and within and across stages by HDA. These types of aggregation are easily combined by sharing aggregation nodes. In this case, we only need to change the root node at each hierarchy by combin ing Equation 1 and 2. Our stages are downsampled by max pooling with size 2 and stride 2.
我们通过HDA跨阶段与IDA进行连接，并且在阶段内和跨阶段进行连接。这些类型的聚合很容易通过共享聚合节点进行组合。在这种情况下，我们只需要通过组合等式1和2来更改每个层级的根节点。我们的舞台通过最大的泳池大小2和步幅2进行降采样。

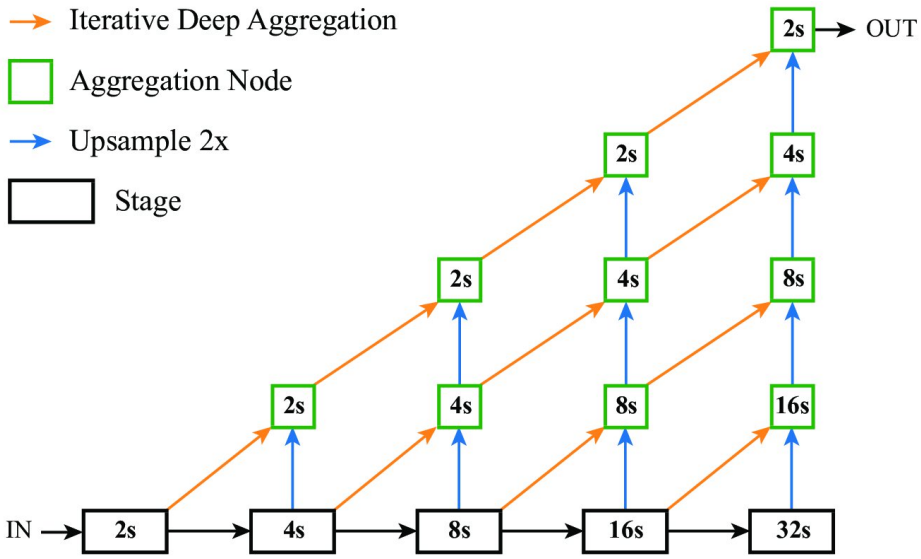


Figure 4: Interpolation by iterative deep aggregation. Stages are fused from shallow to deep to make a progressively deeper and higher resolution decoder.
图4：通过迭代深度聚合进行插值。阶段融合从浅到深，逐渐变得更深和更高分辨率的解码器。

The earliest stages have their own structure. As in DRN [46], we replace max pooling in stages 1–2 with strided convolution. The stage 1 is composed of a 7×7 convolution followed by a basic block. The stage 2 is only a basic block. For all other stages, we make use of combined IDA and HDA on the backbone blocks and stages.
最早的阶段有自己的结构。和DRN [46]一样，我们用阶梯卷积代替了阶段1-2中的最大汇集。阶段1由一个7×7卷积和一个基本块组成。阶段2只是一个基本的块。对于所有其他阶段，我们在骨干块和阶段使用组合的IDA和HDA。

For a direct comparison of layers and parameters in different networks, we build networks with a comparable number of layers as ResNet-34, ResNet-50 and ResNet-101. (The exact depth varies as to keep our novel hierarchical structure intact.) To further illustrate the efficiency of DLA for condensing the representation, we make compact networks with fewer parameters. Table 1 lists our networks and Figure 3 shows a DLA architecture with HDA and IDA.
为了直接比较不同网络中的层和参数，我们构建了与ResNet-34，ResNet-50和ResNet-101相当数量的层的网络。（确切的深度不同，以保持我们的新颖层次结构完整无缺。）为了进一步说明DLA用于表示凝聚的效率，我们使用更少的参数制作紧凑型网络。表1列出了我们的网络，图3显示了具有HDA和IDA的DLA体系结构。

4.2. Dense Prediction Networks
4.2。密集预测网络

Semantic segmentation, contour detection, and other image-to-image tasks can exploit the aggregation to fuse local and global information. The conversion from classification DLA to fully convolutional DLA is simple and no different than for other architectures. We make use of interpolation and a further augmentation with IDA to reach the necessary output resolution for a task.
语义分割，轮廓检测和其他图像到图像任务可以利用聚合来融合本地和全局信息。从分类DLA到完全卷积DLA的转换很简单，与其他体系结构没有区别。我们利用插值和IDA的进一步增强来达到任务所需的输出分辨率。

IDA for interpolation increases both depth and resolution by projection and upsampling as in Figure 4. All the projection and upsampling parameters are learned jointly during the optimization of the network. The upsampling steps are initialized to bilinear interpolation and can then be learned as in [35]. We first project the outputs of stages 3–6 to 32 channels and then interpolate the stages to the same resolution as stage 2. Finally, we iteratively aggregate these stages to learn a deep fusion of low and high level features. While having the same purpose as FCN skip connections [35], hypercolumn features [15], and FPN top-down connections [30], our aggregation differs in approach by going from shallow-todeep to further refine features. Note that we use IDA twice in this case: once to connect stages in the backbone network and again to recover resolution.
用于插值的IDA通过投影和上采样来增加深度和分辨率，如图4所示。所有投影和上采样参数在网络优化期间共同学习。上采样步骤初始化为双线性插值，然后可以在[35]中学习。我们首先投影阶段3-6到32通道的输出，然后将阶段内插到与阶段2相同的分辨率。最后，我们迭代地聚合这些阶段，以学习低层和高层特征的深层融合。虽然与FCN跳过连接[35]，超列功能[15]和FPN自顶向下连接[30]具有相同的用途，但我们的聚合方法从浅到深到进一步改进功能的方法不同。请注意，在这种情况下，我们使用两次IDA：一次连接骨干网络中的各个阶段，再次恢复分辨率。

5. Results
5.结果

We evaluate our deep layer aggregation networks on a variety of tasks: image classification on ILSVRC, several kinds of fine-grained recognition, and dense prediction for semantic segmentation and contour detection. After establishing our classification architecture, we transfer these networks to the other tasks with little to no modification. DLA improves on or rivals the results of special-purpose networks.
我们在各种任务上评估我们的深层聚合网络：ILSVRC上的图像分类，几种细粒度识别，以及用于语义分割和轮廓检测的密集预测。在建立我们的分类体系结构之后，我们将这些网络转移到其他任务几乎不需要修改的地方。DLA改善或与特定用途网络的结果相媲美。

5.1. ImageNet Classification
5.1。 ImageNet分类

We first train our networks on the ImageNet 2012 training set [34]. Similar to ResNet [16], training is performed by SGD for 120 epochs with momentum 0.9, weight decay 10^{-4} and batch size 256. We start the training with learning rate 0.1, which is reduced by 10 every 30 epochs. We use scale and aspect ratio augmentation [41], but not color perturbation. For fair comparison, we also train the ResNet models with the same training procedure. This leads to slight improvements over the original results.
我们首先在ImageNet 2012培训集上训练我们的网络[34]。与ResNet [16]类似，培训由SGD进行120个时期，动量为0.9，重量衰减10^{-4}和批量为256。我们开始学习率为0.1的训练，每30个时期减少10。我们使用比例和纵横比增加[41]，但不是颜色扰动。为了公平比较，我们还使用相同的培训程序来训练ResNet模型。这导致了对原始结果的轻微改进。

We evaluate the performance of trained models on the ImageNet 2012 validation set. The images are resized so that the shorter side has 256 pixels. Then central 224×224 crops are extracted from the images and fed into networks to measure prediction accuracy.
我们评估ImageNet 2012验证集上训练模型的性能。调整图像大小以使短边具有256个像素。然后从图像中提取中央224×224作物并馈送到网络中以测量预测精度。

DLA vs. ResNet compares DLA networks to ResNets with similar numbers of layers and the same convolutional blocks as shown in Figure 5. We find that DLA networks can achieve better performance with fewer parameters. DLA-34 and ResNet-34 both use basic blocks, but DLA-34 has about 30% fewer parameters and $\sim$ 1 point of improvement in top-1 error rate. We usually expect diminishing returns of performance of deeper networks. However, our results show that compared to ResNet-50 and ResNet-101, DLA networks can still outperform the baselines significantly with fewer parameters.
DLA与ResNet将DLA网络与ResNets进行比较，具有相同数量的层和相同的卷积块，如图5所示。我们发现，使用更少的参数，DLA网络可以实现更好的性能。DLA-34和ResNet-34都使用基本块，但DLA-34的参数减少了大约30％，而$\sim$ 1的前1个错误率有所改善。我们通常期望更深层网络的性能收益递减。然而，我们的结果显示，与ResNet-50和ResNet-101相比，DLA网络仍然可以在较少参数的情况下显着优于基线。

DLA vs. ResNeXt shows that DLA is flexible enough to use different convolutional blocks and still have advantage in accuracy and parameter efficiency as shown in Figure 5. Our models based on the split blocks have much fewer parameters but they still have similar performance with ResNeXt models. For example, DLA-X-102 has nearly the half number of parameters compared to ResNeXt-101, but the error rate difference is only 0.2%.
DLA与ResNeXt的比较表明，DLA足够灵活地使用不同的卷积块，并且仍然在精度和参数效率方面具有优势，如图5所示。我们基于拆分块的模型参数少得多，但它们与ResNeXt模型仍具有相似的性能。例如，与ResNeXt-101相比，DLA-X-102的参数数量接近一半，但错误率差异仅为0.2％。

DLA vs. DenseNet compares DLA with the dominant architecture for semantic fusion and feature re-use. DenseNets are composed of dense blocks that aggregate all of their layers by concatenation and transition blocks that reduce dimensionality for tractability. While these networks can
DLA与DenseNet比较DLA与主导架构的语义融合和特征重用。DenseNets由稠密块组成，它们通过串联和转换块来聚合所有层，从而降低易处理性的维数。虽然这些网络可以

Name Block Stage 1 Stage 2 Stage 3 Stage 4 Stage 5 Stage 6 aggressively reduce depth and parameter count by feature reuse, concatenation is a memory-intensive fusion operation. DLA achieves higher accuracy with lower memory usage because the aggregation node fan-in size is log of the total number of convolutional blocks in HDA.
名称块第1阶段第2阶段第3阶段第4阶段第5阶段第6阶段通过特性重用大量减少深度和参数数量，连接是一种内存密集型融合操作。由于聚合节点扇入大小是HDA中卷积块总数的对数，因此DLA可以实现更高的准确性并降低内存使用量。

DLA-34	Basic	16	32	1-64	2-128	2-256	1-512

DLA-48-C	Bottleneck	16	32	1-64	2-64	2-128	1-256
DLA-60	Bottleneck	16	32	1-128	2-256	3-512	1-1024
DLA-102	Bottleneck	16	32	1-128	3-256	4-512	1-1024
DLA-169	Bottleneck	16	32	2-128	3-256	5-512	1-1024
DLA-X-48-C	Split	16	32	1-64	2-64	2-128	1-256
DLA-X-60-C	Split	16	32	1-64	2-64	3-128	1-256
DLA-X-60	Split	16	32	1-128	2-256	3-512	1-1024
DLA-X-102	Split	16	32	1-128	3-256	4-512	1-1024

Table 1: Deep layer aggregation networks for classification. Stages 1 and 2 show the number of channels n while further stages show d-n where d is the aggregation depth. Models marked with "-C" are compact and only have ∼1 million parameters.
表1：用于分类的深层聚合网络。阶段1和2显示通道数目n，而进一步阶段显示d-n，其中d是聚集深度。标有"-C"的型号结构紧凑，只有约100万个参数。

Compact models have received a lot of attention due to the limited capabilities of consumer hardware for running convolutional networks. We design parameter constrained DLA networks to study how efficiently DLA can aggregate and re-use features. We compare to SqueezeNet [20], which shares a block design similar to our own. Table 2 shows that DLA is more accurate with the same number of parameters. Furthermore DLA is more computationally efficient by operation count.
由于消费者硬件运行卷积网络的能力有限，紧凑型模型受到了很多关注。我们设计参数约束的DLA网络来研究有效的DLA如何聚合和重用功能。我们比较SqueezeNet [20]，它与我们自己的模块设计相似。表2显示，使用相同数量的参数，DLA更准确。此外，DLA在操作计数方面的计算效率更高。

Top-1 Top-5 Params FMAs
Top-1前5个参数FMAs

SqueezNet-A	42.5	19.7	1.2M	1.70B
SqueezNet-B	39.6	17.5	1.2M	0.72B
DLA-46-C	36.8	15.0	1.3M	0.58B
DLA-46-C	34.0	13.7	**1.1M**	**0.53B**
DLA-X-60-C	**32.5**	**12.0**	1.3M	0.59B

Table 2: Comparison with compact models. DLA is more accurate at the same number of parameters while inference takes fewer operations (counted by fused multiply-adds).
表2：与紧凑型号的比较。在相同数量的参数下，DLA更准确，而推断需要更少的操作（通过融合乘法计算）。

5.2. Fine-grained Recognition
5.2。细粒度识别

We use the same training procedure for all of fine-grained experiments. The training is performed by SGD with a minibatch size of 64, while the learning rate starts from 0.01 and is then divided by 10 every 50 epochs, for 110 epochs in total. The other hyperparameters are fixed to their settings for ImageNet classification. In order to mitigate over-fitting, we carry out the following data augmentation: Inception-style

我们对所有细粒度实验使用相同的训练程序。培训由SGD进行，小批量大小为64，学习率从0.01开始，然后每50个时期除以10，共计110个时期。其他超参数被固定为ImageNet分类的设置。为了减轻覆盖率，我们进行了以下数据增强：启发式

#Class #Train (per class) #Test (per class) scale and aspect ratio variation [39], AlexNet-style PCA color noise[23], and the photometric distortions of [18].
#Class #Train（每班）#Test（每班）尺度和纵横比变化[39]，AlexNet式PCA色噪声[23]以及[18]的光度失真。

Bird	200	5994 (30)	5794 (29)
Car	196	8144 (42)	8041 (41)
Plane	102	6667 (67)	3333 (33)
Food	101	75750 (750)	25250 (250)
ILSVRC	1000	1,281,167 (1281)	100,000 (100)

Table 3: Statistics for fine-grained recognition datasets. Compared to generic, large-scale classification, these tasks contain more specific classes with fewer training instances.
表3：细粒度识别数据集的统计。与通用的大规模分类相比，这些任务包含更多的特定类和更少的训练实例。

We evaluate our models on various fine-grained recognition datasets: Bird (CUB) [40], Car [22], Plane [31], and Food [5]. The statistics of these datasets can be found in Table 3, while results are shown in Figure 6. For fair comparison, we follow the experimental setup of [9]: we randomly crop 224×224 in resized 256×256 images for [5] and 448×448 in resized 512×512 for the rest of datasets, while keeping 224×224 input size for original VGGNet.
我们在各种细粒度识别数据集上评估我们的模型：Bird（CUB）[40]，Car [22]，Plane [31]和Food [5]。这些数据集的统计数据可以在表3中找到，而结果如图6所示。为了公平比较，我们遵循[9]的实验设置：我们在调整大小的256×256图像中随机裁剪224×224 [5]和448×448大小为512×512的其余数据集，同时保留224×224原始VGGNet的输入大小。

Our results improve or rival the state-of-the-art without further annotations or specific modules for fine-grained recognition. In particular, we establish new state-of-the-arts results on Car, Plane, and Food datasets. Furthermore, our models are competitive while having only several million parameters. However, our results are not better than stateof-the-art on Birds, although note that this dataset has fewer instances per class so further regularization might help.
我们的结果改善了或者与最新的技术相媲美，没有进一步的细节识别模块。特别是，我们在Car，Plane和Food数据集上建立了新的最新技术成果。此外，我们的模型具有竞争力，同时只有几百万个参数。然而，我们的结果并不比鸟类的技术水平更好，但请注意，此数据集每个类的实例较少，因此进一步的正则化可能会有所帮助。

5.3. Semantic Segmentation
5.3。语义分割

We report experiments for urban scene understanding on CamVid [6] and Cityscapes [8]. Cityscapes is a largerscale, more challenging dataset for comparison with other methods while CamVid is more convenient for examining ablations. We use the standard mean intersection-over-union (IoU) score [12] as the evaluation metric for both datasets. Our networks are trained only on the training set without the usage of validation or other further data.
我们在CamVid [6]和Cityscapes [8]中报告了城市场景理解的实验。与其他方法相比，Cityscapes是一个大尺度，更具挑战性的数据集，而CamVid更便于检查消融。我们使用标准平均交叉口联合（IoU）评分[12]作为两个数据集的评估指标。我们的网络仅在训练集上进行训练，而不使用验证或其他更多数据。

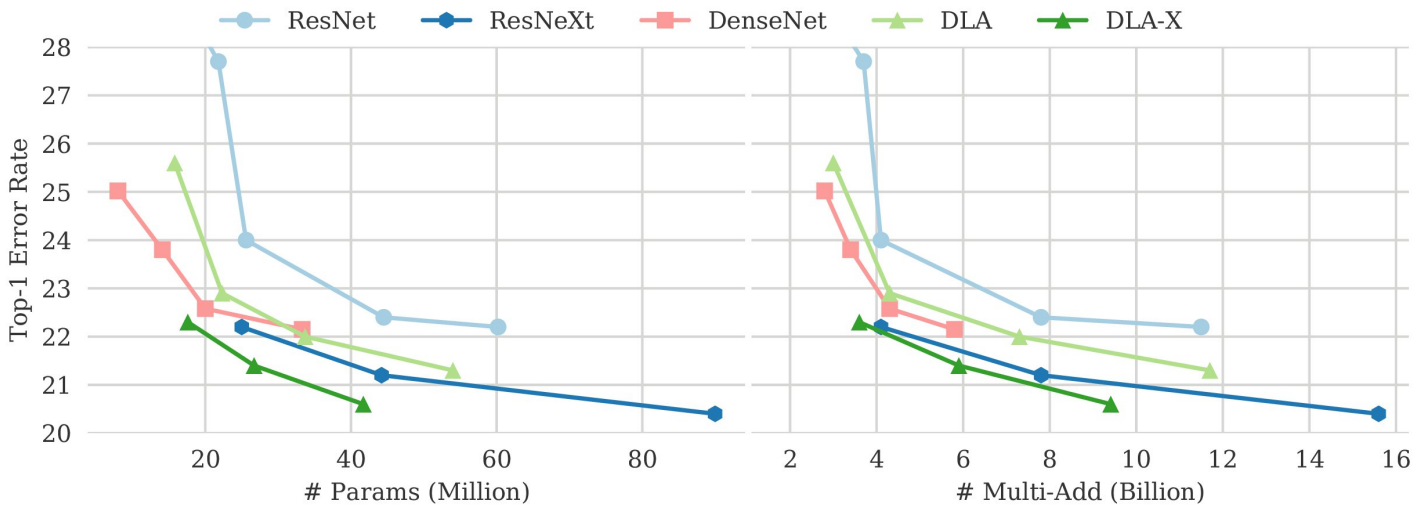


Figure 5: Evaluation of DLA on ILSVRC. DLA/DLA-X have ResNet/ResNeXT backbones respectively. DLA achieves the highest accuracies with fewer parameters and fewer computation.

图5：评估ILSVRC上的DLA。 DLA / DLA-X分别拥有ResNet / ResNeXT主干。DLA通过更少的参数和更少的计算达到最高的精度。

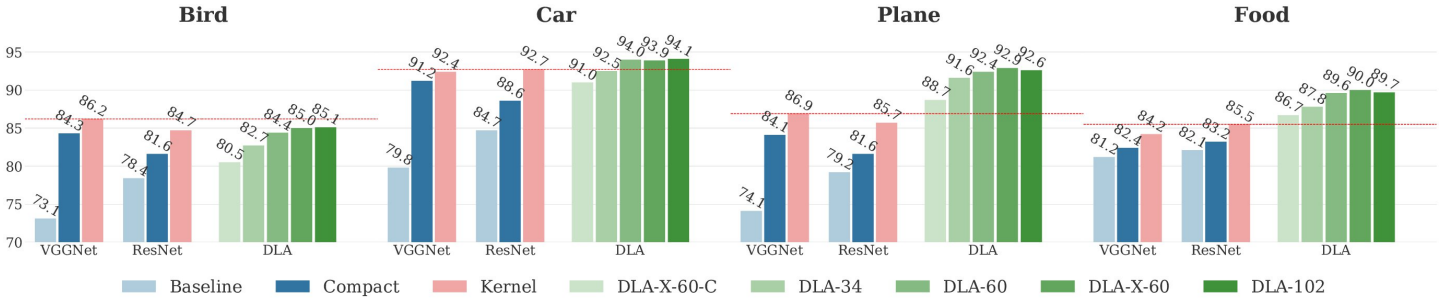


Figure 6: Comparison with state-of-the-art methods on fine-grained datasets. Image classification accuracy on Bird [40], Car [22], Plane [31], and Food [5]. Higher is better. P is the number of parameters in each model. For fair comparison, we calculate the number of parameters for 1000-way classification. V- and R- indicate the base model as VGGNet-16 and ResNet-50, respectively. The numbers of Baseline, Compact [13] and Kernel [9] are directly cited from [9].

图6：与细粒度数据集上现有技术方法的比较。Bird [40]，Car [22]，Plane [31]和Food [5]的图像分类精度。越高越好。P是每个模型中的参数数量。为了公平比较，我们计算了1000路分类的参数数量。V-和R-分别表示基本型号为VGGNet-16和ResNet-50。Baseline，Compact [13]和Kernel [9]的数目直接引自[9]。

CamVid has 367 training images, 100 validation images, and 233 test images with 11 semantic categories. We start the training with learning rate 0.01 and divide it by 10 after 800 epochs. The results are shown in Table 8. We find that models with downsampling rate 2 consistenly outperforms those downsampling by 8. We also try to augment the data by randomly rotating the images between [-10, 10] degrees and randomly scaling the images between 0.5 and 2. The final results are significantly better than prior methods.

CamVid拥有367个训练图像，100个验证图像和233个包含11个语义类别的测试图像。我们开始训练，学习率为0.01，在800个时代之后除以10。结果如表8所示。我们发现，下采样率2的模型一致性优于下采样8。我们还试图通过在[-10,10]度之间随机旋转图像并在0.5和2之间随机缩放图像来增加数据。最终的结果明显优于先前的方法。

Cityscapes has 2, 975 training images, 500 validation images, and 1, 525 test images with 19 semantic categories. Following previous works [49], we adopt the poly learning rate $\left(1 - \text{epoch} - 1 \text{ total epoch}\right)^{0.9}$ with momentum 0.9 and train the model

for 500 epochs with batch size 16. The starting learning rate is 0.01 and the crop size is chosen to be 864. We also augment the data by randomly rotating within 10 degrees and scaling between 0.5 and 2. The validation results are shown in 9. Surprisingly, DLA-34 performs very well on this dataset and it is as accurate as DLA-102. It should be noted that fine spatial details do not contribute much for this choice of metric. RefineNet [28] is the strongest network in the same class of methods without the computational costs of additional data, dilation, and graphical models. To make a fair comparison, we evaluate in the same multi-scale fashion as that approach with image scales of [0.5, 0.75, 1, 1.25, 1.5] and sum the predictions. DLA improves by $2+$ points.

城市景观拥有2,975个训练图像，500个验证图像和1 525个含19个语义类别的测试图像。继之前的工作[49]，我们采用聚合学习速率$(1-\text{时代}-1\text{总时代})^{0.9}$，动量为0.9，训练模型500个时期，批次大小为16。开始学习率为0.01，裁剪尺寸选为864。我们还通过在10度内随机旋转并在0.5和2之间缩放来增加数据。验证结果显示在9中。令人惊讶的是，DLA-34在这个数据集上表现得非常好，并且与DLA-102一样精确。应该指出，精细的空间细节对这种度量的选择没有太多的贡献。ReNetNet [28]是同类方法中最强大的网络，没有附加数据，膨胀和图形模型的计算成本。为了进行公平比较，我们以与图像尺度[0.5,0.75,1,1.25,1.5]相同的多尺度方式进行评估并对预测进行求和。 DLA改进了$2+$分。

5.4. Boundary Detection
5.4。边界检测

Boundary detection is an exacting task of localization. Although as a classification problem it is only a binary task of whether or not a boundary exists, the metrics require precise spatial accuracy. We evaluate on classic BSDS [1] with multiple human boundary annotations and PASCAL boundaries [32] with boundaries defined by instances masks of select semantic classes. The metrics are accuracies at different thresholds, the optimal dataset scale (ODS) and

边界检测是本地化的一项艰巨任务。尽管作为分类问题，它只是边界是否存在的二元任务，但这些度量需要精确的空间精度。我们用经过多种人类边界注释和PASCAL边界[32]的经典BSDS [1]进行评估，边界由实例选定语义类的掩码定义。度量标准在不同阈值下的精度，最佳数据集标度（ODS）和

Method Split mIoU Method ODS OIS AP more lenient optimal image scale (OIS), as well as average precision (AP). Results are shown in for BSDS in Table 6 and the precision-recall plot of Figure 7 and for PASCAL boundaries in Table 7.

方法拆分mIoU方法ODS OIS AP更宽松的最佳图像比例（OIS）以及平均精度（AP）。结果显示在表6中的BSDS和图7的精度 - 召回图以及表7中的PASCAL边界中。

DLA-34 8s	Val	73.4
DLA-34 2s		74.5
DLA-102 2s		74.4
FCN-8s [35]	Test	65.3
RefineNet-101 [28]		73.6
DLA-102		75.3
DLA-169		**75.9**

Table 4: Evaluation on Cityscapes to compare strides on validation and to compare against existing methods on test. DLA is the best-in-class among methods in the same setting.

表4：对城市风景进行评估以比较验证的步幅，并与现有的测试方法进行比较。DLA在相同环境中的方法中是同类中最好的。

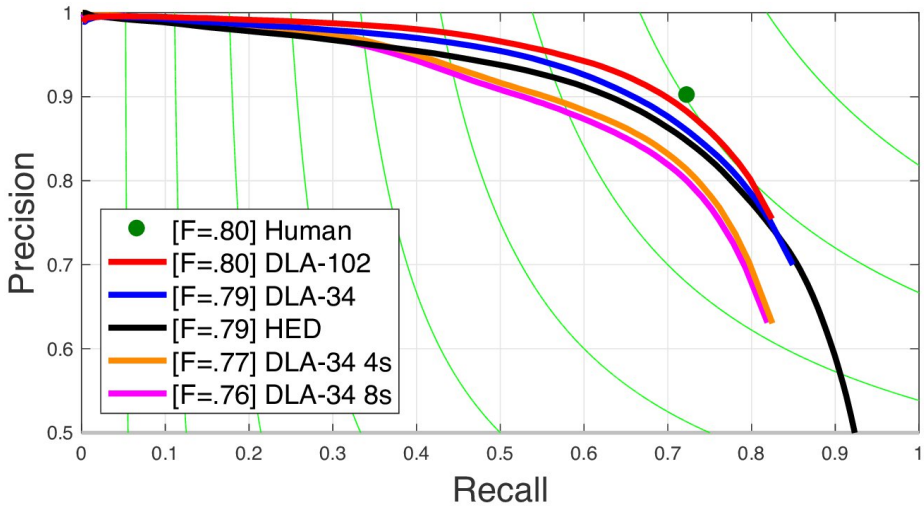


Figure 7: Precision-recall evaluation on BSDS. DLA is the closest to human performance.

图7：对BSDS的精确回忆评估。 DLA最接近人的表现。

SE [10]	0.746	0.767	0.803
DeepEdge [3]	0.753	0.772	0.807
DeepContour [36]	0.756	0.773	0.797
HED [42]	0.788	0.808	0.840
CEDN [43]†	0.788	0.804	0.821
UberNet [21] (1-Task)†	0.791	0.809	**0.849**
DLA-34 8s	0.760	0.772	0.739
DLA-34 4s	0.767	0.778	0.751
DLA-34 2s	0.794	0.808	0.787
DLA-102 2s	**0.803**	**0.813**	0.781

Table 6: Evaluation on BSDS († indicates outside data). ODS and OIS are state-of-the-art, but AP suffers due to recall. See Figure 7.

To address this task we follow the training procedure of HED [42]. In line with other deep learning methods, we

为了解决这个问题，我们遵循HED的培训程序[42]。与其他深层次的学习方法一致，我们

Method mIoU Method Train ODS OIS AP take the consensus of human annotations on BSDS and only supervise our network with boundaries that three or more annotators agree on. Following [43], we give the boundary labels 10 times weight of the others. For inference we simply run the net forward, and do not make use of ensembles or multi-scale testing. Assessing the role of resolution by comparing strides of 8, 4, and 2 we find that high output resolution is critical for accurate boundary detection. We also find that deeper networks does not continue improving the prediction performance on BSDS.

方法mIoU方法培训ODS OIS AP在BSDS上达成人类注释的共识，并且仅以三个或更多注释者达成一致的边界监督我们的网络。[43]之后，我们给边界标签10倍的其他权重。为了推断，我们只需简单地向前运行网络，而不要使用合奏或多尺度测试。通过比较8，4和2的步幅来评估分辨率的作用，我们发现高输出分辨率对于准确的边界检测至关重要。我们还发现深层网络不会继续改进BSDS的预测性能。

SegNet [2]	46.4
DeepLab-LFOV [7]	61.6
Dilation8 [45]	65.3
FSO [24]	66.1
DLA-34 8s	66.7
DLA-34 2s	68.5
DLA-102 2s	**71.0**

Table 5: Evaluation on CamVid. Higher depth and resolution help. DLA is state-of-the-art.
表5：CamVid评估。更高的深度和分辨率帮助。 DLA是最先进的。

SE [10]	BSDS	0.541	0.570	0.486
HED [43]		0.553	0.585	0.518
DLA-34 2s		0.642	0.668	0.624
DLA-102 2s		0.648	0.674	0.623
DSBD [32]	PASCAL	0.643	0.663	0.650
M-DSBD [32]		0.652	0.678	0.674
DLA-34 2s		0.743	0.757	**0.763**
DLA-102 2s		**0.754**	**0.766**	0.752

Table 7: Evaluation on PASCAL Boundaries. DLA is stateof-the-art.
表7：评估PASCAL边界。 DLA是最先进的。

On both BSDS and PASCAL boundaries we achieve state-of-the-art ODS and OIS scores. In contrast the AP on BSDS is surprisingly low, so to understand why we plot the precision-recall curve in Figure 7. Our approach has lower recall, but this is explained by the consensus ground truth not covering all of the individual, noisy boundaries. At the same time it is the closest to human performance. On the other hand we achieve state-of-the-art AP on PASCAL boundaries since it has a single, consistent notion of boundaries. When training on BSDS and transferring to PASCAL boundaries the improvement is minor, but training on PASCAL boundaries itself with $\sim 10\times$ the data delivers more than 10% relative improvement over competing methods.
在BSDS和PASCAL边界上，我们都可以实现最先进的ODS和OIS分数。相比之下，BSDS上的AP值惊人地低，因此要理解我们为什么绘制图7中的精度 - 回忆曲线。我们的方法召回率较低，但这可以通过未涵盖所有单个嘈杂边界的共识基础事实来解释。同时它最接近人类的表现。另一方面，我们在PASCAL边界上实现了最先进的AP，因为它具有单一的，一致的边界概念。在对BSDS进行培训并转移到PASCAL边界时，改进很小，但对PASCAL与$\sim 10\times$本身的界限进行培训后，与竞争方法相比，相对数据改进超过10%。

6. Conclusion
六，结论

Aggregation is a decisive aspect of architecture, and as the number of modules multiply their connectivity is made all the more important. By relating architectures for aggregating channels, scales, and resolutions we identified the need for deeper aggregation, and addressed it by iterative deep aggregation and hierarchical deep aggregation. Our deep layer aggregation networks are more accurate and make more efficient use of parameters and computation than baseline networks. Our aggregation

extensions improve on dominant architectures like residual and densely connected networks. Bridging the gaps of architecture makes better use of layers in aggregate.

聚合是体系结构的决定性方面，随着模块数量的增加，其连接性变得更加重要。通过将通道，尺度和分辨率聚合的体系结构相结合，我们确定了需要进行更深层次的聚合，并通过迭代深度聚合和层次深度聚合来解决它。我们的深层聚合网络比基线网络更准确，可以更有效地使用参数和计算。我们的聚合扩展改进了残余和密集连接网络等主流架构。缩小体系结构的差距可以更好地利用层。

References
参考

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. TPAMI, 2011. 7

[1] P. Arbelaez，M. Maire，C. Fowlkes和J. Malik。轮廓检测和分层图像分割。 TPAMI，2011。7

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561, 2015. 8

[2] V. Badrinarayanan，A.肯德尔和R.西波拉。 Segnet：用于图像分割的深度卷积编码器 - 解码器架构。 arXiv预印本arXiv：1511.00561,2015

[3] G. Bertasius, J. Shi, and L. Torresani. DeepEdge: A multiscale bifurcated deep network for top-down contour detection. In CVPR, 2015. 8

[3] G. Bertasius，J. Shi和L. Torresani。 DeepEdge：用于自顶向下轮廓检测的多尺度分叉深度网络。在CVPR，2015年。8

[4] C. M. Bishop. Pattern recognition and machine learning, page 229. Springer-Verlag New York, 2006. 1

[4] C. M. Bishop。模式识别和机器学习，第229页。Springer-Verlag New York，2006

[5] L. Bossard, M. Guillaumin, and L. Van Gool. Food-101– mining discriminative components with random forests. In ECCV, 2014. 6, 7

[5] L. Bossard，M. Guillaumin和L. Van Gool。 Food-101-挖掘具有随机森林的区分组件。在ECCV中，2014.6,7

[6] G. J. Brostow, J. Fauqueur, and R. Cipolla. Semantic object classes in video: A high-definition ground truth database. Pattern Recognition Letters, 2009. 6

[6] G. J. Brostow，J. Fauqueur和R. Cipolla。视频中的语义对象类：高定义的地面实况数据库。模式识别字母，2009年

[7] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In ICLR, 2015. 8

[7] L.-C. Chen，G.Papandreou，I.Kokkinos，K.Murphy和A.L.Yuille。深卷积网和全连接crfs的语义图像分割。 2015年ICLR。8

[8] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In CVPR, 2016. 2, 6

[8] M. Cordts，M. Omran，S. Ramos，T. Rehfeld，M. Enzweiler，R. Benenson，U. Franke，S. Roth和B. Schiele。用于城市场景语义理解的城市景观数据集。在CVPR，2016年2月6日

[9] Y. Cui, F. Zhou, J. Wang, X. Liu, Y. Lin, and S. Belongie. Kernel pooling for convolutional neural networks. In CVPR, 2017. 6, 7

[9] Y.崔，F。周，J。王，X.刘，Y.林和S. Belongie。卷积神经网络的核心池。在CVPR，2017.6,7

[10] P. Doll´ar and C. L. Zitnick. Structured forests for fast edge detection. In ICCV, 2013. 8

[10] P. Doll'ar和C. L. Zitnick。用于快速边缘检测的结构化森林。在ICCV，2013年。8

[11] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In ICML, 2014. 1, 2
[11] J. Donahue，Y. Jia，O. Vinyals，J. Hoffman，N. Zhang，E. Tzeng和T. Darrell。 Decaf：用于通用视觉识别的深层卷积激活功能。在ICML中，2014. 1，2

[12] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. IJCV, 2010. 7
[12] M. Everingham，L. Van Gool，C. K. Williams，J. Winn和A. Zisserman。 pascal可视对象类（voc）挑战。 IJCV，2010

[13] Y. Gao, O. Beijbom, N. Zhang, and T. Darrell. Compact bilinear pooling. In CVPR, 2016. 7
[13] Y. Gao，O. Beijbom，N. Zhang和T. Darrell。紧凑的双线性池。在CVPR，2016。7

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Regionbased convolutional networks for accurate object detection and segmentation. PAMI, 2015. 1, 2
[14] R. Girshick，J. Donahue，T. Darrell和J. Malik。基于区域的卷积网络，用于精确的目标检测和分割。 PAMI，2015。1，2

[15] B. Hariharan, P. Arbel´aez, R. Girshick, and J. Malik. Hypercolumns for object segmentation and fine-grained localization. In CVPR, 2015. 1, 5
[15] B. Hariharan，P.阿尔贝阿兹，R. Girshick和J.马利克。用于对象分割和细化本地化的高列。在CVPR，2015。1，5

[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016. 1, 2, 5, 11
[16] K. He，X. Zhang，S. Ren和J. Sun.图像识别的深度残留学习。在CVPR，2016年。1，2，5，11

[17] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In ECCV, 2016. 1, 4
[17] K. He，X. Zhang，S. Ren和J. Sun.深度残差网络中的身份映射。在ECCV，2016。1，4

[18] A. G. Howard. Some improvements on deep convolutional neural network based image classification. arXiv preprint arXiv:1312.5402, 2013. 6
[18] A. G. Howard。基于深度卷积神经网络图像分类的一些改进。 arXiv预印本arXiv：1312.5402,2013。6

[19] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. In CVPR, 2017. 1, 2
[19] G. Huang，Z. Liu，K. Q. Weinberger和L. van der Maaten。密集连接的卷积网络。在CVPR，2017年1月2日

[20] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. arXiv preprint arXiv:1602.07360, 2016. 6
[20] F. N. Iandola，S. Han，M. W. Moskewicz，K. Ashraf，W. J. Dally和K. Keutzer。 Squeezenet：Alexnet级精度，参数少于50x，模型大小小于0.5 mb。 arXiv预印本arXiv：1602.07360,2016

[21] I. Kokkinos. Ubernet: Training auniversal’convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. arXiv preprint arXiv:1609.02132, 2016. 8
[21] I. Kokkinos。 Ubernet：使用不同的数据集和有限的记忆，为低，中，高级视觉训练一个“通用”卷积神经网络。 arXiv预印本arXiv：1609.02132,2016

[22] J. Krause, M. Stark, J. Deng, and L. Fei-Fei. 3d object representations for fine-grained categorization. In ICCV Workshops, 2013. 6, 7

[22] J. Krause，M. Stark，J. Deng和L. Fei-Fei。用于细粒度分类的3d对象表示。在ICCV研讨会上，2013.6,7

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012. 1, 2, 6
[23] A. Krizhevsky，I. Sutskever和G. E. Hinton。 Imagenet分类与深卷积神经网络。在NIPS，2012年。1，2，6

[24] A. Kundu, V. Vineet, and V. Koltun. Feature space optimization for semantic video segmentation. In CVPR, 2016. 8
[24] A. Kundu，V. Vineet和V. Koltun。用于语义视频分割的特征空间优化。在CVPR，2016年。8

[25] G. Larsson, M. Maire, and G. Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. In ICLR, 2017. 2
[25] G. Larsson，M. Maire和G. Shakhnarovich。分形网：没有残差的超深度神经网络。 2017年ICLR。2

[26] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradientbased learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998. 1
[26] Y. LeCun，L. Bottou，Y. Bengio和P. Haffner。基于渐变的学习应用于文档识别。Proceedings of the IEEE，86（11）：2278-2324,1998

[27] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeplysupervised nets. In Artificial Intelligence and Statistics, pages 562–570, 2015. 2
[27] C.-Y. Lee，S. Xie，P. Gallagher，Z. Zhang和Z. Tu。深度监督网。人工智能与统计，第562-570页，2015年。2

[28] G. Lin, A. Milan, C. Shen, and I. Reid. Refinenet: Multipath refinement networks with identity mappings for highresolution semantic segmentation. In CVPR, 2017. 7, 8
[28] G. Lin，A. Milan，C. Shen和I. Reid。 Renetnet：具有高分辨率语义分割的身份映射的多路径优化网络。在CVPR，2017.7,8

[29] M. Lin, Q. Chen, and S. Yan. Network in network. In ICLR, 2014. 1, 2
[29] M. Lin，Q. Chen和S. Yan。网络中的网络。在ICLR，2014。1，2

[30] T.-Y. Lin, P. Doll´ar, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In CVPR, 2017. 1, 2, 3, 5
[30] T.-Y. Lin，P. Doll'ar，R. Girshick，K. He，B. Hariharan和S. Belongie。特征金字塔网络用于对象检测。在CVPR，2017年1，2，3，5

[31] S. Maji, E. Rahtu, J. Kannala, M. Blaschko, and A. Vedaldi. Fine-grained visual classification of aircraft. arXiv preprint arXiv:1306.5151, 2013. 6, 7
[31] S. Maji，E. Rahtu，J. Kannala，M. Blaschko和A. Vedaldi。飞机的细粒度视觉分类。 arXiv预印本arXiv：1306.5151,2013。6,7

[32] V. Premachandran, B. Bonev, X. Lian, and A. Yuille. Pascal boundaries: A semantic boundary dataset with a deep semantic boundary detector. In WACV, 2017. 2, 7, 8
[32] V. Premachandran，B. Bonev，X. Lian和A. Yuille。 Pascal边界：带有深层语义边界检测器的语义边界数据集。在WACV，2017.2,7,8

[33] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and ComputerAssisted Intervention, 2015. 3
[33] O. Ronneberger，P. Fischer和T. Brox。 U-Net：用于生物医学图像分割的卷积网络。2015年医学影像计算和计算机辅助干预国际会议。3

[34] 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. IJCV, 2015. 2, 5
[34] O. Russakovsky，J. Deng，H. Su，J. Krause，S. Satheesh，S. Ma，Z. Huang，A. Karpathy，A. Khosla，M.Bernstein，et al。 Imagenet大规模视觉识别挑战。 IJCV，2015.2,5

[35] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. TPAMI, 2016. 1, 2, 3, 5, 8
[35] E. Shelhamer，J. Long和T. Darrell。用于语义分割的完全卷积网络。 TPAMI，2016,1,2,3,5,8

[36] W. Shen, X. Wang, Y. Wang, X. Bai, and Z. Zhang. DeepContour: A deep convolutional feature learned by positive-sharing loss for contour detection. In CVPR, 2015. 8
[36] W. Shen，X. Wang，Y. Wang，X. Bai和Z. Zhang。 DeepContour：深度卷积特征通过正共享损失学习轮廓检测。在CVPR，2015年。8

[37] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015. 1
[37] K. Simonyan和A. Zisserman。用于大规模图像识别的非常深的卷积网络。 2015年ICLR。1

[38] R. K. Srivastava, K. Greff, and J. Schmidhuber. Highway networks. In NIPS, 2015. 1, 2
[38] R. K. Srivastava，K. Greff和J. Schmidhuber。公路网络。在NIPS，2015。1，2

[39] 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. 1, 2, 6
[39] C. Szegedy，W. Liu，Y. Jia，P. Sermanet，S. Reed，D. Anguelov，D. Erhan，V. Vanhoucke和A. Rabinovich。进一步与卷积。在CVPR，2015年。1，2，6

[40] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 6, 7
[40] C.华，S.布兰森，P. Welinder，P.佩罗纳和S. Belongie。 caltech-ucsd鸟-200-2011数据集。 2011年6月7日

[41] S. Xie, R. Girshick, P. Doll´ar, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In CVPR, 2017. 2, 4, 5, 11
[41] S.谢，R. Girshick，P. Doll'ar，Z. Tu和K. He。深度神经网络的聚合残差变换。在CVPR，2017年2月，4日，5日，11日

[42] S. Xie and Z. Tu. Holistically-nested edge detection. In ICCV, 2015. 1, 8
[42] S.谢和Z. Tu。全局嵌套边缘检测。在ICCV，2015年1月8日

[43] J. Yang, B. Price, S. Cohen, H. Lee, and M.-H. Yang. Object contour detection with a fully convolutional encoder-decoder network. In CVPR, 2016. 8
[43] J. Yang，B. Price，S. Cohen，H. Lee和M.-H.杨。具有完全卷积编码器 - 解码器网络的对象轮廓检测。在CVPR，2016年。8

[44] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In NIPS, 2014. 1
[44] J. Yosinski，J. Clune，Y. Bengio和H. Lipson。深度神经网络中的特征如何转移？在NIPS，2014年。1

[45] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. In ICLR, 2016. 8
[45] F. Yu和V. Koltun。多尺度上下文聚合扩展卷积。在ICLR，2016年。8

[46] F. Yu, V. Koltun, and T. Funkhouser. Dilated residual networks. In CVPR, 2017. 1, 5
F. Yu，V. Koltun和T. Funkhouser。膨胀的残余网络。在CVPR，2017。1，5

[47] S. Zagoruyko and N. Komodakis. Wide residual networks. arXiv preprint arXiv:1605.07146, 2016. 1

[47] S. Zagoruyko和N. Komodakis。广泛的残余网络。 arXiv预印本arXiv：1605.07146，2016

[48] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In ECCV, 2014. 1, 2
[48] M. D. Zeiler和R. Fergus。可视化和理解卷积网络。在ECCV，2014. 1，2

[49] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. arXiv preprint arXiv:1612.01105, 2016. 7
[49] H. Zhao，J. Shi，X. Qi，X. Wang和J. Jia。金字塔场景解析网络。 arXiv预印本arXiv：1612.01105,2016

A. ImageNet Classification
A. ImageNet分类

The concept of DLA framework is general since it doesn't require particular designs of convolution blocks. We also turn to simple design of aggregation nodes in our applications. Figure 8 shows the aggregation nodes for hierarchical aggregation. It is a concatenation of the input channels followed by a 1 × 1 convolution. We explore adding residual connection in the aggregation as shown in Figure 8(c). It is only used when there are more than 100 layers in the classification network. Three types of convolutional blocks are studied in this paper, as shown in Figure 9, since they are widely used in deep learning literature. Because the convolutional blocks will be combined with additional linear projection in the aggregation nodes, we reduce the ratio of bottleneck blocks from 4 to 2.
DLA框架的概念是通用的，因为它不需要卷积块的特定设计。我们也转而在我们的应用程序中设计简单的聚合节点。图8显示了分层聚合的聚合节点。它是一个1×1卷积后的输入通道的串联。我们探索在聚合中添加剩余连接，如图8（c）所示。只有在分类网络中有超过100层的情况下才会使用它。本文研究了三种卷积块，如图9所示，因为它们广泛用于深度学习文献。由于卷积块将与聚合节点中的附加线性投影相结合，我们将瓶颈块的比率从4减少到2。

We compare DLA and DLA-X to other networks in Figure 5 in the submitted paper in terms of network parameters and classification accuracy. DLA includes the networks using residual blocks in ResNet and DLA-X includes those using the block in ResNeXt. For fair comparison, we design DLA and DLA-X networks with similar depth and channels with their counterparts. The ResNet models are ResNet-34, ResNet-50, ResNet-101 and ResNet-152. The corresponding DLA model depths are 34, 60, 102, 169. The ResNeXt models are ResNeXt-50 (32x4d), ResNeXt-101 (32x4d), and ResNeXt-101 (64x4d). The corresponding DLA-X model depths are 60 and 102, while the third DLA-X model double the number of 3 × 3 bottleneck channels, similar to ResNeXt-101 (64x4d).
在网络参数和分类精度方面，我们将DLA和DLA-X与图5中的其他网络进行比较。DLA包括ResNet中使用残余块的网络，DLA-X包括那些使用ResNeXt中的块的网络。为了公平比较，我们设计了DLA和DLA-X网络，它们与对应的深度和频道相似。ResNet型号是ResNet-34，ResNet-50，ResNet-101和ResNet-152。相应的DLA模型深度为34,60,102,169。ResNeXt型号为ResNeXt-50（32x4d），ResNeXt-101（32x4d）和ResNeXt-101（64x4d）。与ResNeXt-101（64x4d）类似，相应的DLA-X型号深度为60和102，而第三个DLA-X型号则是3×3瓶颈通道数量的两倍。

B. Semantic Segmentation
B.语义分割

We report experiments for semantic segmentation on CamVid and Cityscapes. Table 8 shows a breakdown of the accuracies for the categories. We also add data agumention in the CamVid training, as shown in the third group of Table 8. It includes random rotating the images between -10 and 10 degrees and random scaling between 0.5 and 2. We find the results can be further improved by the augmentation. Table 9 shows the breakdown of categories in Cityscapes on the validation set. We also test the models on multiple scales of the images. This testing procedure is used in evaluating the models on the testing images in the previous works.

我们在CamVid和Cityscapes上报告语义分割的实验。表8显示了这些类别的精确度的细分。如表8的第三组所示，我们还在CamVid培训中添加了数据提示。它包括在-10和10度之间随机旋转图像以及0.5和2之间的随机比例。我们发现增强可以进一步改善结果。表9显示了验证集中Cityscapes中类别的细分。我们还在图像的多个比例上测试模型。该测试程序用于评估以前工作中测试图像上的模型。

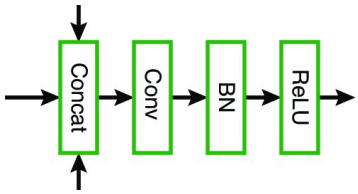


Figure 8: Illustration of aggregation node architectures.

图8：聚合节点体系结构的图示。

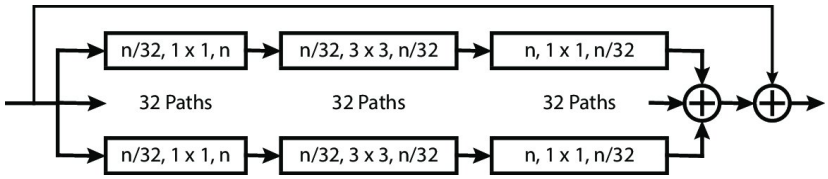


Figure 9: Convolutional blocks used in this paper. Our aggregation architecture is as general as stacking layers, so we can use the building blocks of existing networks. The layer labels indicate output channels, kernel size and input channels. (a) and (b) are derived from [16] and (c) from [41].

图9：本文中使用的卷积块。我们的聚合架构与堆叠层一样通用，所以我们可以使用现有网络的构建模块。图层标签指示输出通道，内核大小和输入通道。 （a）和（b）来自[16]的[16]和（c）[41]。

DLA-34 8s		83.2	77.2	91.2	83.6	48.8	94.3	58.6	32.0	27.8	81.1	55.4	66.7
DLA-60 8s		83.0	77.0	91.4	84.1	46.9	94.1	58.3	32.8	26.0	81.3	56.8	66.5
DLA-34	No	83.2	76.4	92.5	84.6	52.1	94.4	61.5	29.4	35.1	82.0	57.8	68.1
DLA-60		84.4	77.7	92.6	87.1	51.4	95.3	62.2	32.1	36.2	84.5	64.1	69.8
DLA-102		84.9	78.0	92.5	86.4	50.8	94.9	62.8	45.4	35.7	83.7	65.8	71.0
DLA-60		86.6	79.3	92.5	90.9	55.3	96.2	65.5	48.6	37.4	86.9	66.5	73.2
DLA-102	Yes	86.6	78.8	92.2	90.3	57.9	96.5	66.7	49.6	38.7	87.9	66.7	73.8
DLA-169		86.9	78.9	92.5	89.9	58.5	96.5	66.1	55.4	39.0	87.7	67.7	74.4

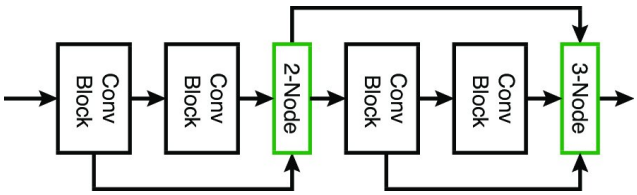
Table 8: Semantic segmentation results on the CamVid dataset.

表8：CamVid数据集上的语义分割结果。

DLA-34 8s	97.9	83.2	91.9	47.7	57.7	62.4	68.6	77.3	92.2	60.4	94.8	81.1	59.8	94.1	57.5	76.6	54.2	59.7	76.6	73.4
DLA-34	98.0	83.5	92.1	51.0	56.8	64.9	69.6	78.5	92.4	62.9	95.1	81.5	59.6	94.5	59.0	78.4	57.8	62.9	76.9	74.5
DLA-102	98.0	84.3	92.3	43.2	56.9	67.2	71.6	80.9	92.5	61.4	94.6	82.7	61.5	94.5	60.3	77.7	53.8	62.2	78.5	74.4
DLA-169	98.2	84.8	92.5	45.9	60.0	68.0	72.3	81.1	92.7	61.9	95.1	83.4	63.3	95.3	70.9	80.8	48.1	65.4	79.1	75.7
DLA-34 MS	98.2	84.7	92.5	54.3	59.5	65.9	71.1	79.5	92.7	64.1	95.3	82.6	61.8	94.7	63.3	83.7	64.6	64.2	77.6	76.3
DLA-102 MS	98.5	85.0	92.5	47.1	56.7	66.9	74.4	78.6	93.6	71.7	95.1	85.8	67.4	95.3	55.8	63.5	57.8	68.1	76.1	76.1
DLA-169 MS	98.3	85.9	92.8	48.3	61.2	69.0	73.4	82.2	92.9	63.1	95.4	84.2	65.1	95.7	76.3	82.9	49.6	68.5	80.2	77.1

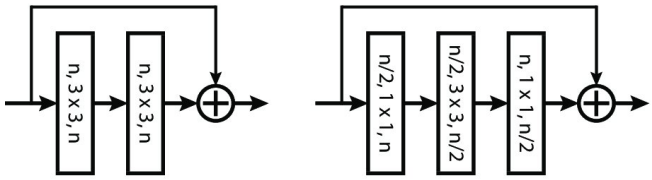
Table 9: Performance of DLA on the Cityscapes validation set. s8 indicates the input image is downsampled by 8 in the model output. It is 2 by default. Lower downsampling rate usually leads to higher accuracy. "MS" indicates the models are tested on on multiple scales of the input images.

表9：城市景观验证集的DLA表现。 s8表示输入图像在模型输出中被降采样8。默认情况下是2。较低的下采样率通常会导致更高的精度。 "MS"表示模型在输入图像的多个比例上进行测试。



(a) 3-level hierarchical aggregation (a) Basic (b) Bottleneck .

（a）三级分层聚合（a）基本（b）瓶颈。

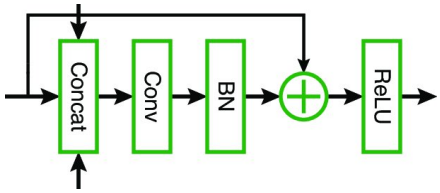


g g u n A i d a l e n t i y e r g a u k a r i D B T S C S

g g u n A i d e n n t i y e r r a u k a r i D B T S C S

n k o g l i a n t a w i e t d d t e l c e e h a l n l n d i g a l g g o u W e o e R B P L V S F S i i i (b) Plain 3-node (c) Residual 3-node (c) Split n U a k t i l s o r a i I t l w s e c d n e c e e y a a d n l d c o e e e o i i R B m P F P S e l c U y o e n c n I r l i k o r c a n o n e c y r s s i t a y r r d r u a o c a e e k u e i r r M i m T S P R C T B T B

Ñķø克升I A为N t的w i的（E T）d d T E L CËËh的L N L N d I G A L克克邻ü宽EöéR B P L V的F S I I I（B）普通纸3节点（c）中的剩余3节点（c）中拆分N个U A K T I升S 0 R A I I T L瓦特发EÇd N个EÇËE Y所涉及的dÑ升d C 0 E E E O I I R B量m P在fp发E L CüY 2 O简C N我R L I KöR C—N 2 O 2 N个EÇýřššI T—ýR R d - [R U A摄氏度A EëķüE I R R M I MŤ性s P R C T B T B



所有论文 (/all_papers/0)

添加客服微信，加入用户群