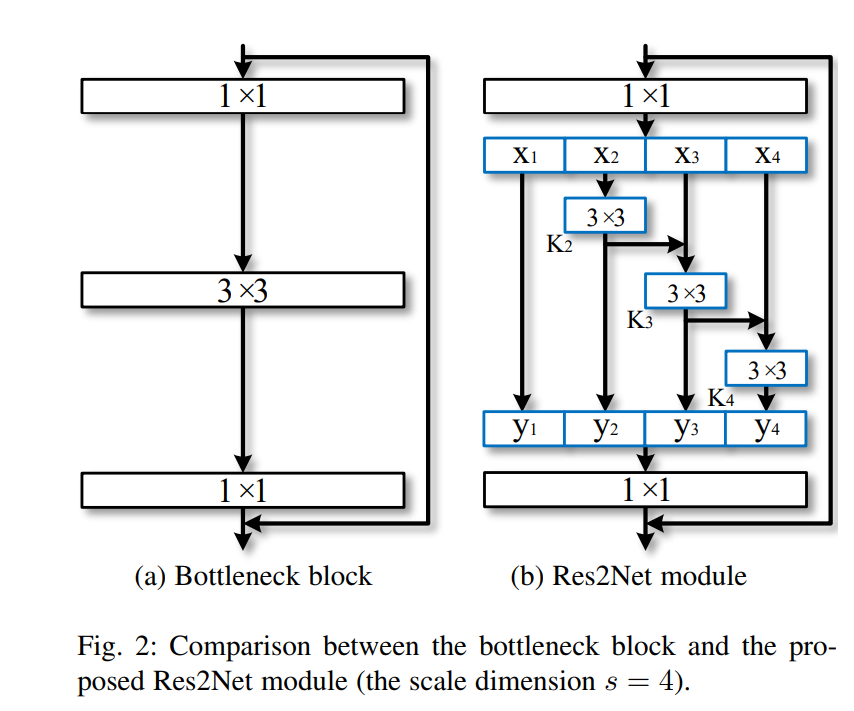
**Res2Net A New Multi-scale Backbone Architecture**

对于多数视觉任务多层次表示特征是重要的。最近在主干CNNs网络上更注重多层次表示能力，在多个应用上取得较大进展。但是，现有大多数方法以分层的形式表示多层特征。在本文中，我们提出一种新的CNNs构建块，称为Res2Net，通过在单一残差块中构建多层次类残差连接。Res2Net在细粒度层表示多尺度特征，同时增加每个网络层的感知范围。我们提出的Res2Net块可以加入到目前主流的多种CNNs模型中，如ResNet，ResNeXt和DLA等。我们评估添加Res2Net模块的所有模型，可看到其在多数常用测试集例如CIFAR-100、ImageNet等，表现优于基线模型。在更多表示性视觉任务上如目标检测、分类匹配、盐物体检测等，Res2Net优于目前最流行的基线算法。

**1 introduction**

视觉模式可多尺度出现在自然风景中如图1所示。首先，在同一图像中目标会以不同大小出现如沙发和茶杯大小不同。第二，同一物体的重要纹理信息可能比物体本身所占据的区域更大。第三，通过不同尺度的感知信息来理解物体各部分是很重要的，如细粒度分类、语义分割等。因此，对于视觉认知任务，设计好的特征进行多尺度刺激是很重要的。

在本文中，我们提出一种简单高效的多尺度处理方法。不像现有大多方法通过增加层数来提升CNNs的多层次表现能力，我们在更细粒度层提升多层次表示能力。为实现这个目标，我们用w通路更小的过滤器集合来代替原先n通道的3\*3滤波器（无损失惩罚我们使用n = s x w）。如图2所示，这些更小的滤波器组用多层残差的方式连接可增加输出特征的多层次表示。特别地，我们将输入特征匹配到许多组。先前组的输出特征和其他组的输入特征被仅送到接下来组的滤波器。这个过程重复多次直到所有输入特征图都被处理。最后，从所有组的特征图被连接然后送到另外一组的1X1滤波器来共同混合信息。



注意提出方法在更细粒度层上的多尺度，和现有利用多层次操作的方法有明显区别。因此，提出的构建模块，称为Res2Net块，可以很容易插入到大多主流CNN架构。实验结果表明Res2net模块可以明显提升现有CNNs性能，如ResNet、ResNeXt和DLA等。

**2 Related Work**

2.1 主干网络

最近几年出现了大量的主干网络，在许多视觉任务上用更强的多尺度表示实现了当前最佳表现。正如所设计的，CNNs由于输入信息从粗到细因此其皆具有基本的多尺度特征表示能力。AlexNet通过堆叠过滤序列已在视觉任务上实现性能明显优于传统方法。但是，由于网络深度和过滤器核大小的限制，Alexnet只有相对小的感知野。VGGNet增加网络深度并使用更小的核。一个更深的结构可以扩大感受野，这有利于从大尺度提取特征，通过堆叠更多层比使用更大的核更容易扩大感知域。因此，VGGNet用更少的参数提供比AlexNet更强的多尺度表示模型。但是，AlexNet和VGGNet在线性拓扑堆叠过滤器，其意味着这些网络仅有相对灵活的感觉域且更易于处理小尺度的目标。

GoogLeNet利用不同大小核的并行过滤器来提升多尺度表示能力。但是由于计算资源的限制，GoogLeNet的核不易扩展。因此，GoogLeNet的多尺度表示仍不能覆盖大范围的感受域。Inception Nets在并行的每个通道上堆叠更多的过滤器来扩展感受域。另一方面，ResNet引入一种神经网络间的短连接，在缓解梯度消失的同时增加网络结构深度。在特征提取阶段，短连接可以让不同卷积操作间相互结合。相似的，DenseNet通过更密集的连接使网络能够在多尺度处理目标。DPN将DenseNet的特征扩展能力和ResNet的特征重利用结合在一起。最近提出的DLA算法在树状结构结合层。多层次的树状结构使网络能够有更强的多层次多尺度表示能力。

2.2 视觉任务的多尺度表示

CNNs的多尺度视觉表示对很多视觉任务都是很重要的包括目标检测、盐目标检测和语义分割，能有效促进这些领域模型性能。

2.2.1 目标检测

有效的CNN模型需定位同一场景不同尺度的目标。前期工作如R-CNN主要依赖主干网络如VGGNet从多个尺度提取特征。He等提出SPP-Net方法其在主干网络后利用空间金字塔平滑来提升多尺度能力。Faster R-CNN方法提出区域候选块网络来产生多尺度的边界。基于Faster R-CNN，FPN引入特征金字塔来从单幅图像中提取不同尺度特征。SSD方法利用特征匹配从不同阶段来处理不同尺度的视觉信息。

2.2.2 语义分割

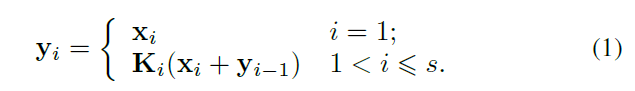
2.2.3 盐物体检测

**3 RES2NET**

**3.1 Res2Net模块**

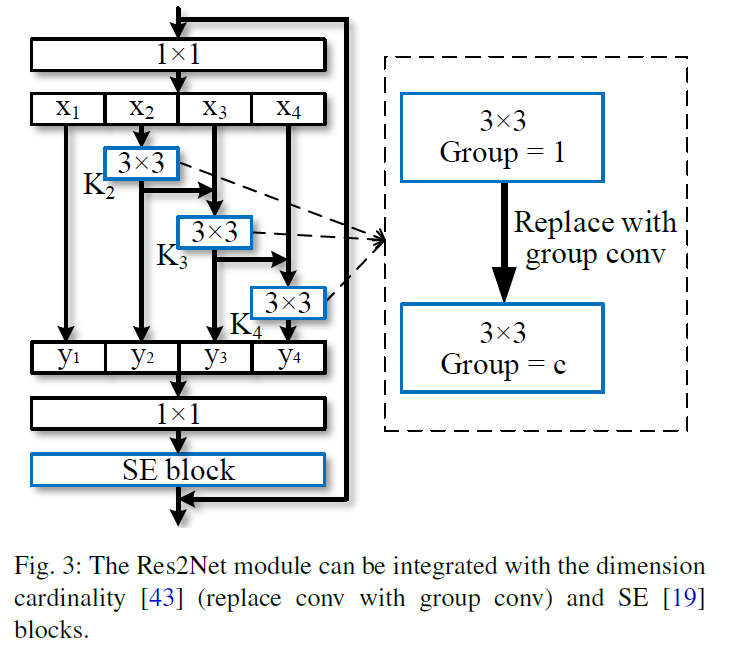
图2(a)显示的瓶颈结构是目前很多主流CNNs的主要架构，例如ResNet等。相对于在瓶颈块用一组3\*3的过滤器提取特征，我们寻求用多层次特征提取能力的可选择性特征，同时保留相似计算能力。特别地，我们用更小的滤波器组来代替3\*3的滤波器，同时连接在多层残差的不同滤波器组。由于我们提出的神经网络模块涉及单个残差块中的类残差连接，因此我们称之为Res2Net。

图2展示了瓶颈块和我们提出Res2Net模块的区别。在1\*1卷积后，我们分割特征图到s个特征图子集，定义Xi，i<{1,2,....s}。每个特征子集Xi有相同空间的大小但是相比较于输入特征图仅有1/s数量的通道。特别对x1，每个Xi有一个对应的3X3卷积，用Ki()定义。我们用yi定义Ki()的输出。特征子集xi会增加Ki-1()的输出，一起传入Ki()。为了在增加s的同时减少参数，我们删掉了x1的 3x3卷积。因此，yi可以这样写：



注意到每个3 x 3卷积操作Ki()能够从所有特征分组{xj，j < i}中接收特征信息。每次特征分组xj经过一个3 x 3卷积操作后，输出结果会有比xj更大的感知域。由于组合爆炸的影响，Res2Net模块的输出包含不同数量以及不同感知域大小及尺度的组合。

在Res2Net模块，分组用多尺度方式处理，其有助于提取全局及局部信息。为了更好地混淆不同尺度信息，我们将所有的分组通过1 X 1卷积。分组和连接策略可以加强卷积处理特征的有效性。为减少参数的数量，我们删掉第一组的卷积，其可以作为一种特征重用的形式。



在这项工作中，我们用s作为尺度维度的控制参数。更大的s对应更强多尺度能力，通过连接提前引入的计算/内存开销可以忽略。

**3.2 与现有模块集成**

近些年来，涌现出大量神经网络模块，包括Xie等引入的基数维度和Hu等提出的挤压与激励SE模块。加入尺度维度的Res2Net和这些提升呈正交状态。如图3所示，我们可以很容易的在Res2Ner模块中加入基数维度和SE模块。

3.2.1 维度基数

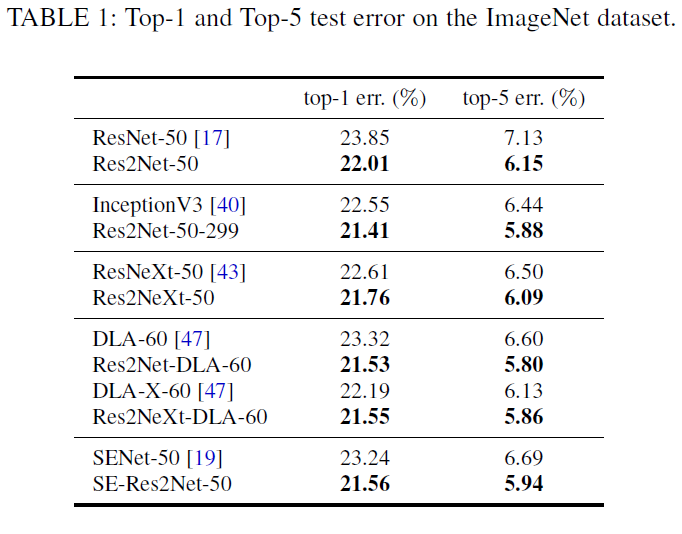
维度基数指一个过滤器中组的数量。维度将过滤器从单一分支改变为多分支，同时提升CNN模型的表示能力。在我们的设计中，我们可以用3 X 3的卷积组代替3 X 3的卷积，c表示组的数量。基数和尺度维度的对比实验在4.2和4.4 部分给出。

3.2.2 SE块

SE模块通过明确通道间依赖性来自适应地重新校准信道特征响应[19]。同[19]相似，我们在Res2Net残差模块连接之前增加SE模块。我们的Res2Net模块能够从同SE模块的结合收益，如我们在4.2和4.3部分的实验所示。

**3.3 集成模块**

由于提出的Res2Net模块对整个的网络结构没有特殊的要求，同时Res2Net模块的多尺度表示能力同CNNs模型的分层特征表示正交，我们可以很容易地将Res2Net模块结合到目前最流行的方法，如ResNet，ResNeXt,DLA等。对应模型分别称为Res2Net, Res2NeXt和Res2Net-DLA。



本文提出的尺度维度与先前工作的基数维度以及宽度维度成正交。因此，在尺度确定后，我们调整基数值以及宽度来确保整个模型的复杂度和它对应部分相似。在本工作中，我们并不聚焦于减小模型大小，因为其需要更细致的设计如深度可分离的卷积[32]，模型修剪[13]以及模型压缩。

在ImageNet数据集上的实验，由于计算资源限制，我们主要使用ResNet-50 [17], ResNeXt-50 [43] and DLA-60 [47]作为我们的基准模型。我们提出模型的计算复杂度和基准模型基本相同，参数近25M，对一幅224 X 224的图像FLOPs对50层的网络为4.2G。在CIFAR数据集上的实验，我们使用ResNeXt-29, 8c64w [43]作为我们的基准模型。经验评估和提出模型复杂度的讨论在4.4部分给出。

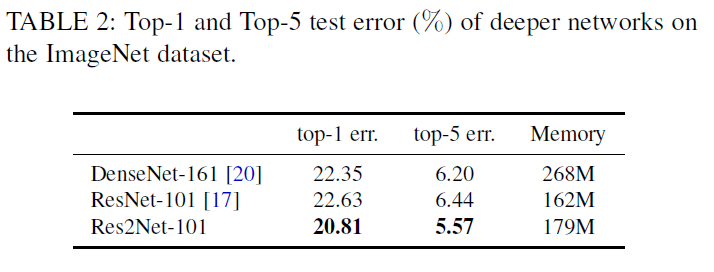
**4 Experiments**

**4.1 实现细节**

我们用Pytorch框架实现本文提出的模型。为了公平对比，我们使用Pytorch实现ResNet [17], ResNext [43] 和 DLA [47]，同时用Res2Net模块仅更新原始瓶颈部分。与之前工作相似，在ImageNet数据集上，224X224的图像是任意从既定大小图像裁剪出的。我们使用[17]的数据扩增策略。同[17]相似，我们用权重衰减为0.0001、动量为0.9、最小块为256的SGD在4块Titan Xp GPUs上。学习率初始设置为0.1，并在30轮之后除以10.

ImageNet的所有模型，包括基线和提出的模型，用相同的方法和数据扩增策略训练100轮。为了测试，我们使用[17]中的裁剪方法。在CIFAR数据集上，我们使用ResNeXt-29来实现无其余更新。对于其他任务，我们使用基线模型的原始实现，仅仅用Res2Net替代瓶颈模块。

**4.2 ImageNet**



我们在ImageNet上进行实验，其包含1000类的128万训练图像和50k验证图像。由于计算资源限制，我们构建大概50层网络模型。

4.2.1 性能增益

表1给出了ImageNet数据上top-1和top-5的测试错误率。为了简化，表1中所有的Res2Net模型尺度皆为4。Res2Net-50相比ResNet-50在top-1错误率上提升1.84%。