

# 论文译文 ResNeSt- Split-Attention Networks

论文译文 ResNeSt: Split-Attention Networks

## Abstract

众所周知，特征图注意和多路径表征对于视觉识别非常重要。在本文中，我们提出了一种模块化架构，它将通道式注意力应用于不同的网络分支，以充分利用它们在捕捉跨特征交互和学习多样化表征方面的成功经验。我们的设计产生了一个简单而统一的计算模块，只需使用几个变量就能对其进行参数化。我们的模型名为 ResNeSt，在图像分类的准确性和延迟权衡方面优于 EfficientNet。此外，ResNeSt 还在作为骨干的几个公共基准上取得了优异的迁移学习结果，并被 COCO-LVIS 挑战赛的获奖作品所采用。整套系统和预训练模型的源代码均已公开。

## 1 Introduction

深度卷积神经网络( CNNs )已经成为计算机视觉中图像分类和其他迁移学习任务的基本方法。作为卷积神经网络的关键组成部分，卷积层学习一组滤波器，这些滤波器通过空间和通道连接来聚合邻域信息。该操作适用于捕获与每个输入通道紧密相连的输出通道的相关特征。Inception模型[ 53、54 ]探索多路径表示来学习独立的特征，其中输入被分割成几个低维嵌入，通过不同的卷积滤波器组进行变换，然后通过级联进行合并。该策略通过解耦输入通道连接来鼓励特征探索[ 63 ]。

在过去的几十年中，视觉皮层的神经元连接启发了CNNs的发展[ 30 ]。视觉表征学习的主要任务是发现给定任务的显著特征[ 74 ]。之前的工作已经对空间和通道依赖[ 2, 27, 43]进行了建模，并引入了注意力机制[ 27、36、58 ]。SE-like channelwise attention [ 27 ]使用全局池化来压缩通道统计信息，并预测一组注意力因子与原始特征图进行通道相乘。该机制对特征图通道间的依赖关系进行建模，利用全局上下文信息选择性地突出或淡化特征[ 27、36 ]。这种注意机制类似于人类初级视觉皮层的注意选择阶段[ 73 ]，即寻找识别物体的信息部分。人/动物利用不同区域的大脑皮层感知不同的视觉模式，以响应不同和特殊的视觉特征[ 45 ]。这种策略使

得在神经感知系统中容易识别出相似物体细微但显性的差异。同样，如果我们能够针对不同的视觉特征构建一个CNN架构来捕获个体的显著属性，那么我们将改进网络表示来进行图像分类。

在本文中，我们提出了一种简单的架构，将通道级注意力策略与多路径网络布局相结合。我们的方法捕获了跨通道的特征相关性，同时在元结构中保留了独立的表示。我们的网络的一个模块对低维嵌入执行一组变换，并将它们的输出串联成多路径网络。每个变换都结合了通道注意力策略来捕获特征图的相互依赖关系。我们进一步简化架构，使每个变换共享相同的拓扑( e.g.图2 (右) )。我们可以只用几个变量来参数化网络架构。此外，这样的设置还允许我们使用与统一的CNN运算符相同的实现来加速训练。我们将这样的计算块称为Split - Attention Block。将多个Split - Attention模块以ResNet的方式堆叠，创建了一个新的ResNet变体，我们称之为Split - Attention Network ( ResNeSt )。

我们在ImageNet数据集上对所提出的ResNeSt网络的性能进行了测试[ 14 ]。本文提出的ResNeSt在速度和精度之间取得了比通过神经架构搜索产生的最先进的CNN模型[ 56 ]更好的权衡，如表2所示。此外，我们还研究了迁移学习在目标检测、实例分割和语义分割上的结果。所提出的ResNeSt在作为骨干网时在几个金标准基准上取得了优越的性能。例如，我们的Cascade - RCNN [ 5 ]模型以ResNeSt - 101为主干，在MS - COCO实例分割上取得了48.3 %的box m AP和41.56 %的mask m AP。我们的DeepLabV3 [ 9 ]模型同样使用ResNeSt - 101主干，在ADE20K场景解析验证集上达到了46.9 %的mIoU，比之前最好的结果提高了1 %以上的mIoU。此外，ResNeSt已被2020年COCO - LVIS挑战赛[ 21、55、57 ]的获奖作品所采用。

## 2 Related Work

.....

## 3 Split-Attention Networks

我们现在引入Split - Attention块，它在3.1节中实现了跨不同特征图组的特征图注意力。随后，我们在3.2节描述了我们的网络实例化以及如何通过标准CNN操作符来加速该架构。

# 1. Split-Attention Block

Split - Attention块是一个计算单元，由特征图组和分裂注意力操作组成。图2 (右)描述了一个分割注意力块的概况。

Featuremap Group.正如在ResNeXt块[ 64 ]中，特征可以被分成若干组，特征图组的数量由一个基数超参数K给出。我们将得到的特征映射组称为基数组。在本文中，我们引入了一个新的基超参数R，它表示基数组内的分裂数，因此特征组的总数 $G = KR$ 。我们可以应用一系列变换 $\{ F_1, F_2, \dots, F_G \}$ 到每个个体群体，则每个群体的中间表示为 $U_i = F_i ( X )$ ，对于 $i \in \{ 1, 2, \dots, G \}$ 。

Split Attention in Cardinal Groups.在[ 28、36 ]之后，每个基数组的组合表示可以通过跨多个分裂的元素求和并进行融合得到。k阶基数群的表示为 $( U_k = \sum R_{kj} = R ( k-1 ) + 1 U_j$ ，其中 $( U_k \in \mathbb{R}^{H \times W \times C / K}, k \in 1, 2, \dots )$ 。K、H、W和C是分块输出特征图的大小。嵌入通道统计的全局上下文信息可以通过跨空间维度 $s_k \in \mathbb{R}^{C / K}$  [ 27、36]的全局平均池化来收集。这里第c个分量的计算公式为：

基数组表示 $V_k \in \mathbb{R}^{H \times W \times C / K}$ 的加权融合使用通道软注意力聚合，其中每个特征图通道使用拆分加权组合生成。则第c通道计算为：

其中 $a_{ki} ( c )$ 表示给定的(软)分配权重：

映射 $G_{ci}$ 基于全局上下文表示 $s_k$ 为第c个通道确定每个拆分的权重。

ResNet Block.然后沿通道维度串联基数组表示： $V = \text{Concat} \{ V_1, V_2, \dots, V_K \}$ 。与标准残差块一样，如果输入和输出的特征图具有相同的形状，我们的Split - Attention块的最终输出Y是通过捷径连接产生的： $Y = V + X$ 。对于有跨度的图块，将对快捷连接应用适当的变换 T 来对齐输出图形： $Y = V + T ( X )$ 。例如，T 可以是分段卷积或带池化的组合卷积。

Instantiation and Computational Costs.图2 (右)展示了我们的Split - Attention块的一个实例化，其中组变换 $F_i$ 是一个 $1 \times 1$ 卷积后接一个 $3 \times 3$ 卷积，注意力权重函数G使用两个全连接层和ReLU激活进行参数化。Split - Attention块的参数个数和FLOPS与标准残差块[ 23、63]大致相同，具有相同的基数和通道数。

Relation to Existing Attention Methods.首先在SE - Net [ 27 ]中引入了压缩注意力(在原文中称为激励)的思想，即利用全局上下文预测通道注意力因子。当基数= 1时，我们的Split - Attention块对每个基数组应用一个压缩和注意力操作，而SENet则在整个块的顶部操作，而不考虑多个组。SK-Net [36] 在两个网络流之间引入了特征注意。设置 radix = 2 时，SplitAttention 模块将类似 SK 的注意力应用于每个心形组。我们的方法将先前的特征图注意[27, 36]工作推广到了心形组设置[63]中，而且其实现仍然具有很高的计算效率。图 2 显示了与 SE-Net 和 SK-Net 区块的整体比较。

## 2. Efficient Radix-major Implementation

我们将上一节中描述的布局称为 "cardinality-major 实现"，其中具有相同 cardinal 索引的特征图组彼此相邻（图 2（右））。心数-主实施简单直观，但很难使用标准 CNN 运算符进行模块化和加速。为此，我们引入了等效的弧度-主实现。

图4给出了基-大布局中Split - Attention块的概况。首先将输入的特征图分成RK组，每组有一个基数索引和基数索引。在这种布局中，具有相同基指数的组彼此相邻。然后，我们可以对不同的分裂进行求和，从而将具有相同基数索引但不同基数索引的特征图组融合在一起。一个全局池化层聚集在空间维度上，同时保持通道维度分离，这与对每个基数组进行全局池化并将结果进行拼接是一样的。然后，在池化层之后增加两个连续的全连接(FC)层，其组数等于基数，以预测每个分割的注意力权重。分组FC层的使用使得在每个基数组上分别应用每对FC是相同的。

通过这种实现方式，前 $1 \times 1$ 卷积层可以统一为一层， $3 \times 3$ 卷积层可以使用RK的组数进行单次分组卷积实现。因此，Split - Attention模块采用标准的CNN算子进行模块化处理。

## 4 Network and Training

我们现在描述实验中使用的网络设计和训练策略。首先，我们详述了一些进一步提高性能的调整，其中一些已经在[ 25 ]中得到了经验验证。