ResNet最强改进版来了！ResNeSt：Split-Attention Networks


ResNeSt: Split-Attention Networks

Hang Zhang, Chongruo Wu*, Zhongyue Zhang, Yi Zhu, Haibin Lin, Zhi Zhang, Yue Sun, Tong He, Jonas Mueller, R. Manmatha, Mu Li, and Alexander Smola

Amazon, University of California, Davis*
{hzaws,chongrwu,zhongyue,yzaws,haibilin,zhiz,ysunmzu, htong,jonasmue,manmatha,mli,smola}@amazon.com

《ResNeSt: Split-Attention Networks》

作者团队：亚马逊（

@张航 和 @李沐 等）&UC Davis

代码（提供PyTorch和MXNet双版本）：github.com/zhanghang198

论文：hangzhang.org/files/res

前言

开头先致敬一下 ResNet！Amusi 于2020年4月17日在谷歌学术上查看ResNet的引用量，发现已高达 43413！请注意，这还只是ResNet发表短短4年多的引用量。

这里吐槽一句，现在出现很多基于NAS的新网络（趋势），暴力出奇迹，比如MobileNetV3、EfficientNet等，但论应用场景，还是ResNet给力。实际上，很多下游工作（目标检测、图像分割等）仍然在使用ResNet或其变体，主要是因为结构简洁通用。

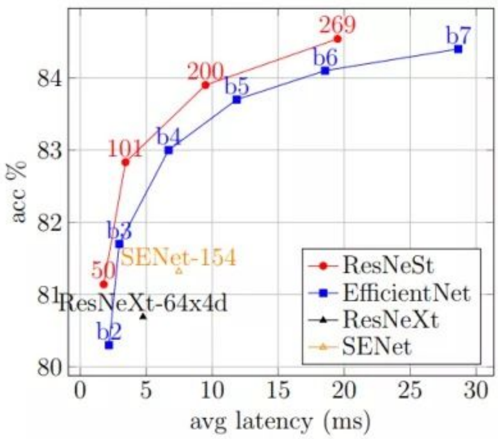
本文要介绍的是ResNet 的新变体：**ResNeSt**。继续将ResNet"发扬光大"，值得点赞。

Amusi 将标题注明了最强，很多人肯定会质疑是不是标题党？究竟有多强？往下看，你就知道了！

先说几组数据：

ResNeSt-50 在 ImageNet 上实现了81.13% top-1 准确率
简单地用ResNeSt-50替换ResNet-50，可以将MS-COCO上的Faster R-CNN的mAP从39.25%提高到42.33%！
简单地用ResNeSt-50替换ResNet-50，可以将ADE20K上的DeeplabV3的mIoU从42.1%提高到45.1%！

性能显著提升，参数量并没有显著增加，部分实验结果如下图所示。轻松超越ResNeXt、SENet等前辈（巨人）们。



	Crop	#P	Acc%
ResNeSt-50 (ours)	224	27.5M	81.1
ResNeSt-101 (ours)	256	48.3M	82.8
ResNeSt-200 (ours)	320	70.2M	83.9
ResNeSt-269 (ours)	416	111M	84.5

	Backbone	#Params	Score%
FasterRCNN [46]	ResNet-50 [57]	34.9M	39.25
	ResNeSt-50 (ours)	36.8M	42.33
DeeplabV3 [7]	ResNet-50 [57]	42.2M	42.10
	ResNeSt-50 (ours)	44.0M	45.10

Table 1: (Left) Accuracy and latency trade-off on GPU using official code implementation (details in Section 5). (Right-Top) Top-1 accuracy on ImageNet using ResNeSt. (Right-Bottom) Transfer learning results: object detection mAP on MS-COCO [42] and semantic segmentation mIoU on ADE20K [71].

ResNeSt

ResNeSt 的全称是：**Split-Attention Networks**，也就是特别引入了Split-Attention模块。如果没有猜错，ResNe**S**t 的 **S** 应该就是 **Split**。

这里要说一下，ResNeSt 实际上是站在巨人们上的"集大成者"，特别借鉴了：**Multi-path 和 Feature-map Attention思想**。

其中：

GoogleNet 采用了**Multi-path**机制，其中每个网络块均由不同的卷积kernels组成。

ResNeXt在ResNet bottle模块中采用**组卷积**，将multi-path结构转换为统一操作。
SE-Net 通过自适应地重新校准通道特征响应来引入**通道注意力（channel-attention）机制。**
SK-Net 通过两个网络分支引入**特征图注意力（feature-map attention）**。

ResNeSt 和 SE-Net、SK-Net 的对应图示如下：

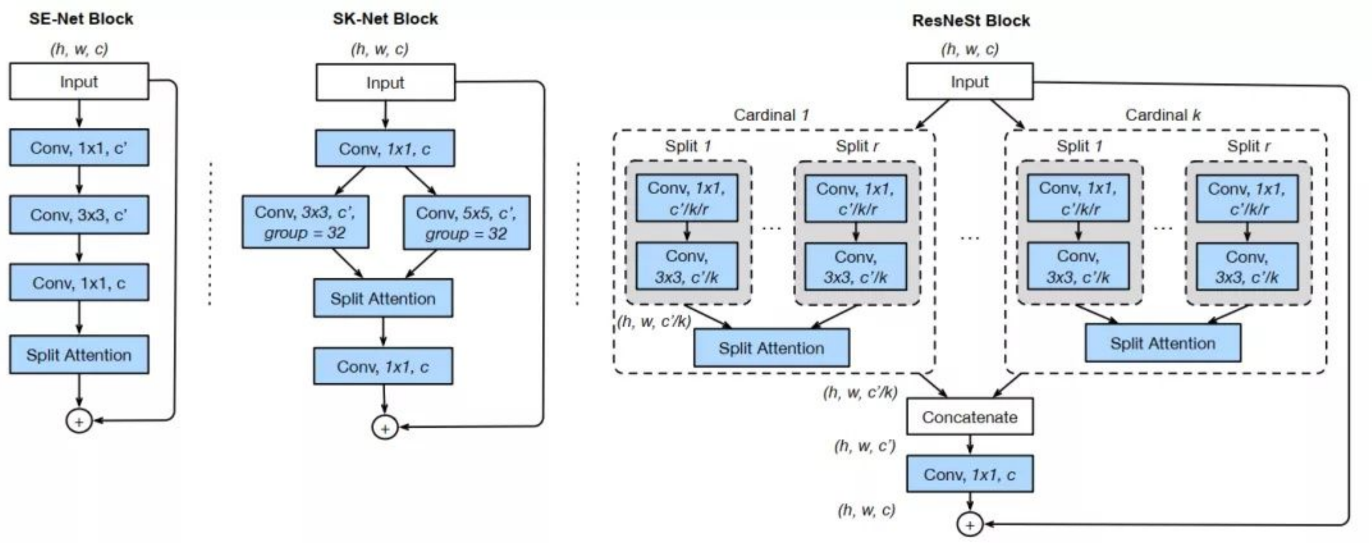


Fig. 1: Comparing our ResNeSt block with SE-Net [30] and SK-Net [38]. A detailed view of Split-Attention unit is shown in Figure 2. For simplicity, we show ResNeSt block in cardinality-major view (the featuremap groups with same cardinal group index reside next to each other). We use radix-major in the real implementation, which can be modularized and accelerated by group convolution and standard CNN layers (see supplementary material).

其中上图中都包含的 Split Attention模块如下图所示：

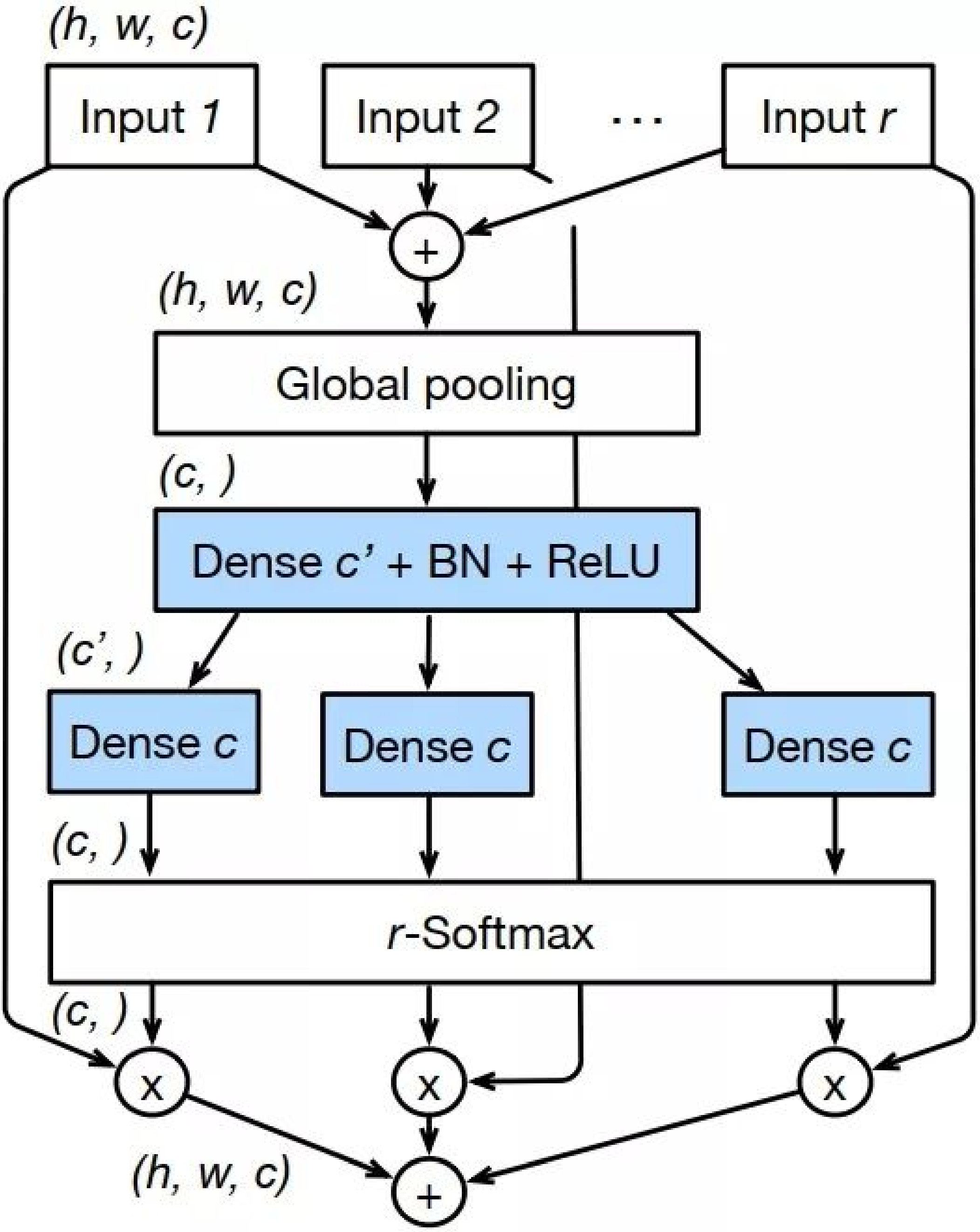


Fig. 2: Split-Attention within a cardinal group. For easy visualization in the figure, we use $c = C/K$ in this figure.

从图1和图2可知，都有split的影子。比如图1中的 **K(k)** 和图2中的 **R(r)** 都是超参数，也就是共计 **G = K*R** 组。

限于篇幅问题，本文旨在论文速递。完整理解Split Attention模块需要涉及部分公式，这里建议大家结合原文和代码进行理解。目前代码已经提供PyTorch和MXNet两个版本。

github.com/zhanghang198

同时论文还介绍了训练策略，这个对大家目前的工作应该具有很大的参考价值（涨点tricks）。

Large Mini-batch Distributed Training
Label Smoothing
Auto Augmentation
Mixup Training
Large Crop Size
Regularization

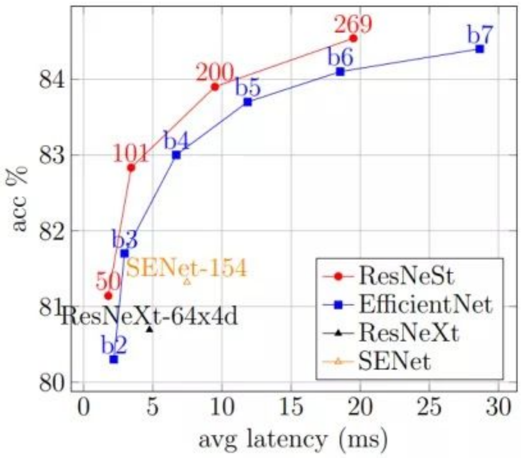
实验结果

ResNeSt 在ImageNet 图像分类性能如下，轻松超越SKNet、SENet、ResNetXt和ResNet。

	#P	GFLOPs	top-1 acc (%)	
			224×	320×
ResNet-50 [23]	25.5M	4.14	76.15	76.86
ResNeXt-50 [60]	25.0M	4.24	77.77	78.95
SENet-50 [29]	27.7M	4.25	78.88	80.29
ResNetD-50 [26]	25.6M	4.34	79.15	79.70
SKNet-50 [38]	27.5M	4.47	79.21	80.68
ResNeSt-50-fast(ours)	27.5M	4.34	**80.64**	**81.43**
ResNeSt-50(ours)	27.5M	5.39	81.13	81.82
ResNet-101 [23]	44.5M	7.87	77.37	78.17
ResNeXt-101 [60]	44.3M	7.99	78.89	80.14
SENet-101 [29]	49.2M	8.00	79.42	81.39
ResNetD-101 [26]	44.6M	8.06	80.54	81.26
SKNet-101 [38]	48.9M	8.46	79.81	81.60
ResNeSt-101-fast(ours)	48.2M	8.07	**81.97**	**82.76**
ResNeSt-101(ours)	48.3M	10.2	82.27	83.00

Table 3: Image classification results on ImageNet, comparing our proposed ResNeSt with other ResNet variants of similar complexity in 50-layer and 101-layer configurations. We report top-1 accuracy using crop sizes 224 and 320.

ResNeSt 和其他SoTA的CNN模型进行性能比较（特别是NAS阵营）



	#P	crop	img/sec	acc(%)
ResNeSt-101(ours)	48M	256	**291.3**	**83.0**
EfficientNet-B4 [55]	19M	380	149.3	83.0
SENet-154 [29]	146M	320	133.8	82.7
NASNet-A [74]	89M	331	103.3	82.7
AmoebaNet-A [45]	87M	299	-	82.8
ResNeSt-200 (ours)	70M	320	**105.3**	**83.9**
EfficientNet-B5 [55]	30M	456	84.3	83.7
AmoebaNet-C [45]	155M	299	-	83.5
ResNeSt-269 (ours)	111M	416	**51.2**	**84.5**
GPipe	557M	-	-	84.3
EfficientNet-B7 [55]	66M	600	34.9	84.4

Table 4: Accuracy vs. Latency for SoTA CNN models on ImageNet with large crop sizes. Our ResNeSt model displays the best trade-off (additional details/results in Appendix). EfficientNet variants b2-b7 are described in [55]. ResNeSt variants use a different number of layers listed in red. Average Inference latency is measured on a NVIDIA V100 GPU using the original code implementation of each model with a mini-batch of size 16.

ResNeSt 在MS-COCO 目标检测和实例分割任务上的表现性能如下，涨点太恐怖！

	Method	Backbone	mAP%
Prior Work	Faster-RCNN [46]	ResNet101 [22]	37.3
		ResNeXt101 [5][60]	40.1
		SE-ResNet101 [29]	41.9
	Faster-RCNN+DCN [12]	ResNet101 [5]	42.1
	Cascade-RCNN [2]	ResNet101	42.8
Our Results	Faster-RCNN [46]	ResNet50 [57]	39.25
		ResNet101 [57]	41.37
		ResNeSt50 (ours)	42.33
		ResNeSt101 (ours)	**44.72**
	Cascade-RCNN [2]	ResNet50 [57]	42.52
		ResNet101 [57]	44.03
		ResNeSt50 (ours)	45.41
		ResNeSt101 (ours)	**47.50**
	Cascade-RCNN [2]	ResNeSt200 (ours)	49.03

Table 5: Object detection results on the MS-COCO validation set. Both Faster-RCNN and Cascade-RCNN are significantly improved by our ResNeSt backbone.

	Method	Backbone	box mAP%	mask mAP%
Prior Work	DCV-V2 [72]	ResNet50	42.7	37.0
	HTC [4]	ResNet50	43.2	38.0
	Mask-RCNN [22]	ResNet101 [5]	39.9	36.1
	Cascade-RCNN [3]	ResNet101	44.8	38.0
Our Results	Mask-RCNN [22]	ResNet50 [57]	39.97	36.05
		ResNet101 [57]	41.78	37.51
		ResNeSt50 (ours)	42.81	38.14
		ResNeSt101 (ours)	**45.75**	**40.65**
	Cascade-RCNN [2]	ResNet50 [57]	43.06	37.19
		ResNet101 [57]	44.79	38.52
		ResNeSt50 (ours)	46.19	39.55
		ResNeSt101 (ours)	**48.30**	**41.56**

Table 6: Instance Segmentation results on the MS-COCO validation set. Both Mask-RCNN and Cascade-RCNN models are improved by our ResNeSt backbone. Models with our ResNeSt-101 outperform all prior work using ResNet-101.

ResNeSt 在ADE20K 语义分割任务上的表现性能如下：

	Method	Backbone	pixAcc%	mIoU%
Prior Work	UperNet [59]	ResNet101	81.01	42.66
	PSPNet [69]	ResNet101	81.39	43.29
	EncNet [65]	ResNet101	81.69	44.65
	CFNet [66]	ResNet101	81.57	44.89
	OCNet [63]	ResNet101	-	45.45
	ACNet [17]	ResNet101	81.96	45.90
Ours	DeeplabV3 [7]	ResNet50 [21]	80.39	42.1
		ResNet101 [21]	81.11	44.14
		ResNeSt-50 (ours)	81.17	45.12
		ResNeSt-101 (ours)	**82.07**	**46.91**

	Method	Backbone	mIoU%
Prior Work	DANet [16]	ResNet101	77.6
	PSANet [70]	ResNet101	77.9
	PSPNet [69]	ResNet101	78.4
	AAF [33]	ResNet101	79.2
	DeeplabV3 [7]	ResNet101	79.3
	OCNet [63]	ResNet101	80.1
Ours	DeeplabV3 [7]	ResNet50 [21]	78.72
		ResNet101 [21]	79.42
		ResNeSt-50 (ours)	79.87
		ResNeSt-101 (ours)	**80.42**

Table 7: Semantic segmentation results on validation set of: ADE20K (Left), Citscapes (Right). Models are trained without coarse labels or extra data.

欢迎各位CVers点赞支持！也推荐大家关注 计算机视觉论文速递 知乎专栏和 CVer 微信公众号，可以快速了解到最新优质的CV论文。