CNN模型架构学习分享

宋禹凝

2022/11/18

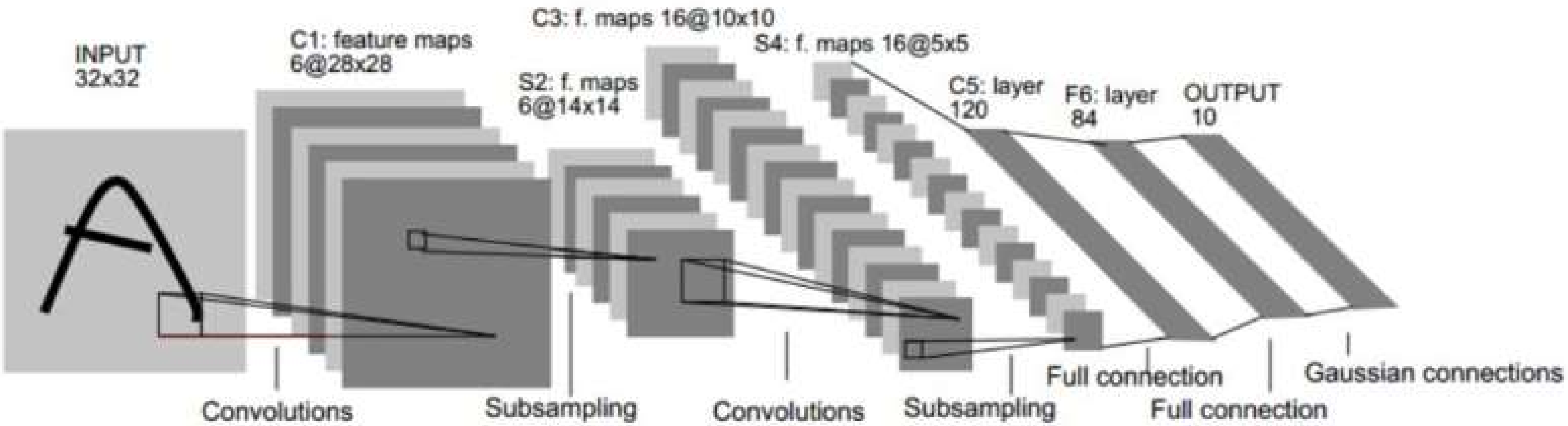
目录

/01

增加深度

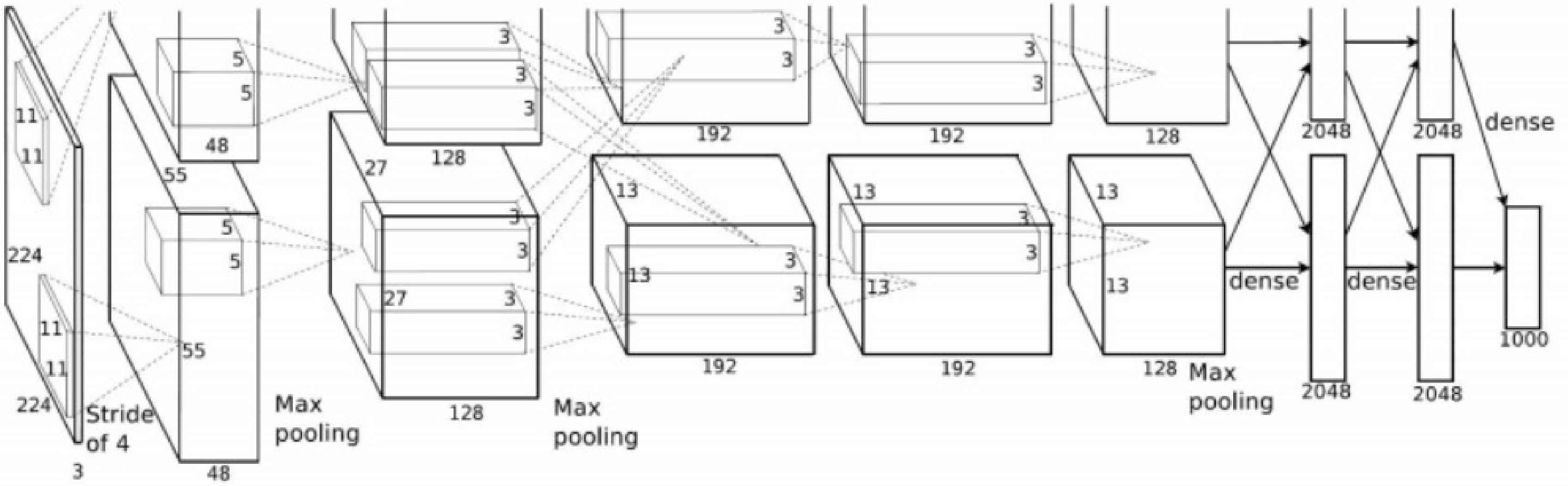
LeNet，AlexNet，VGGNet

LeNet



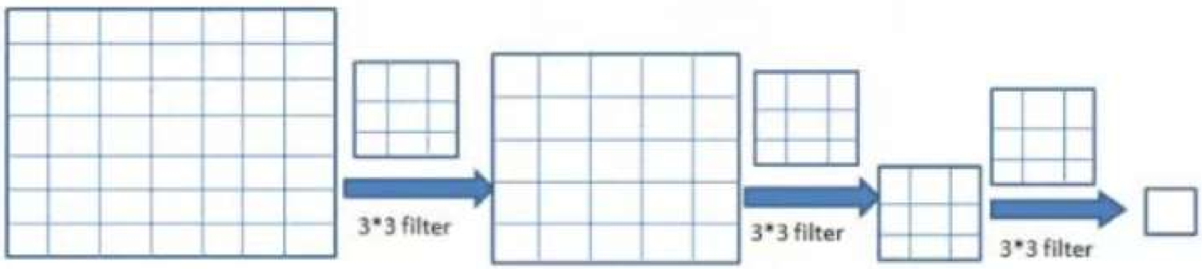
LeNet包含3个卷积层，2个池化层，1个全连接层。所有卷积层的所有卷积核都为5x5，步长为1，池化方法为Max pooling，激活函数为Sigmoid。

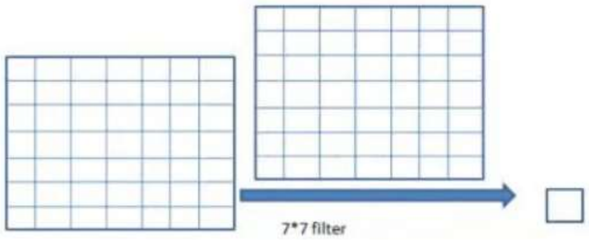
AlexNet



AlexNet是更大更深的LeNet，包括5个卷积层，3个池化层，3个全连接层，激活函数为ReLu，在训练时使用Dropout随机忽略部分神经元，避免模型过拟合。

VGGNet





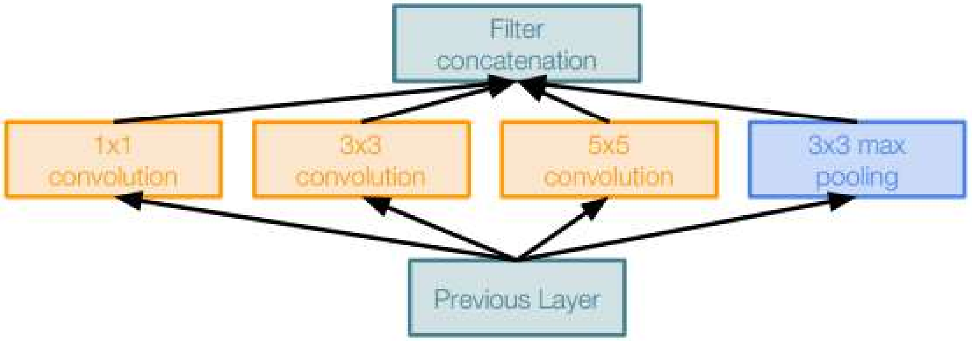
VGGNet相比于AlexNet而言，层数增多且具有更小的卷积核。VGGNet使用的卷积核是3x3的，步长为1，填充为1，池化核为2×2。
3层3×3卷积理论上的感受野和一层7×7的卷积相同，但参数更少，且层数越多，网络更深，非线性表达能力得到提高。

/02

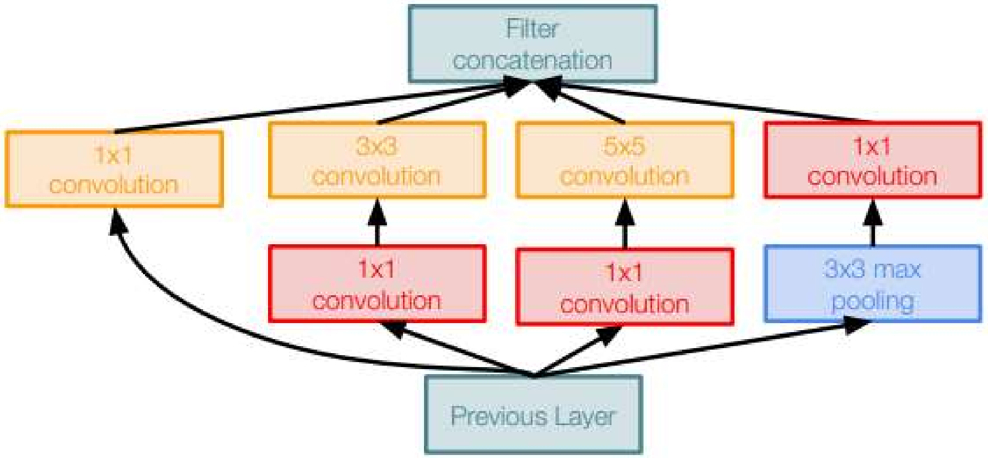
GoogLeNet

Inception， NIN

GoogLeNet



多尺度并行卷积



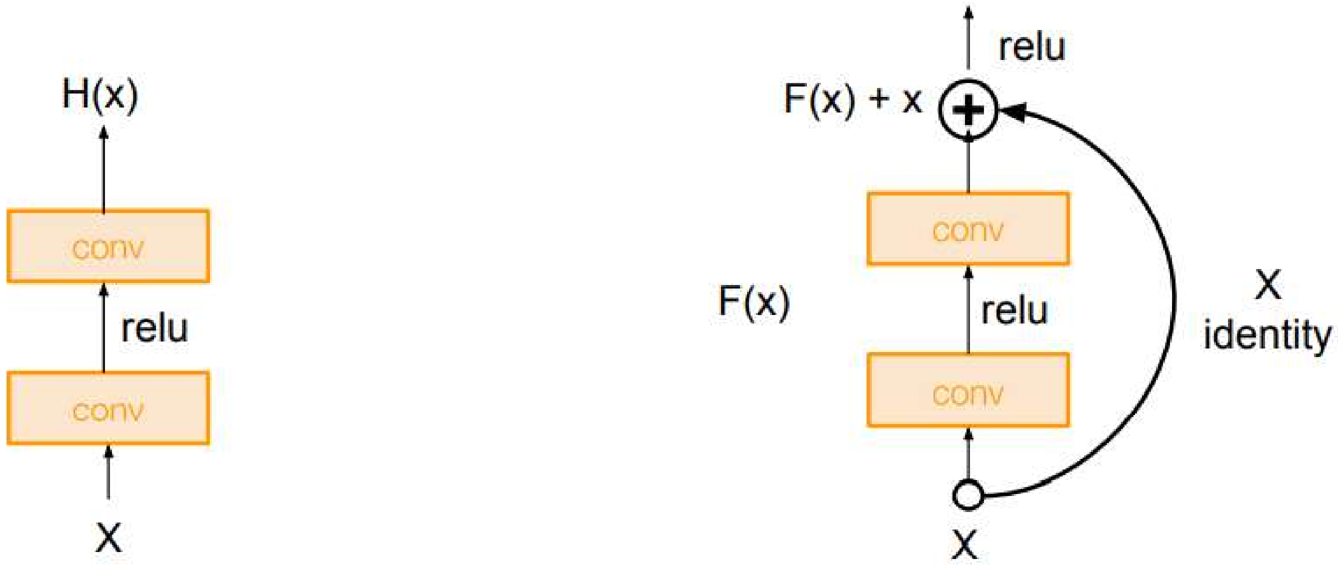
增加1×1卷积模块

/03

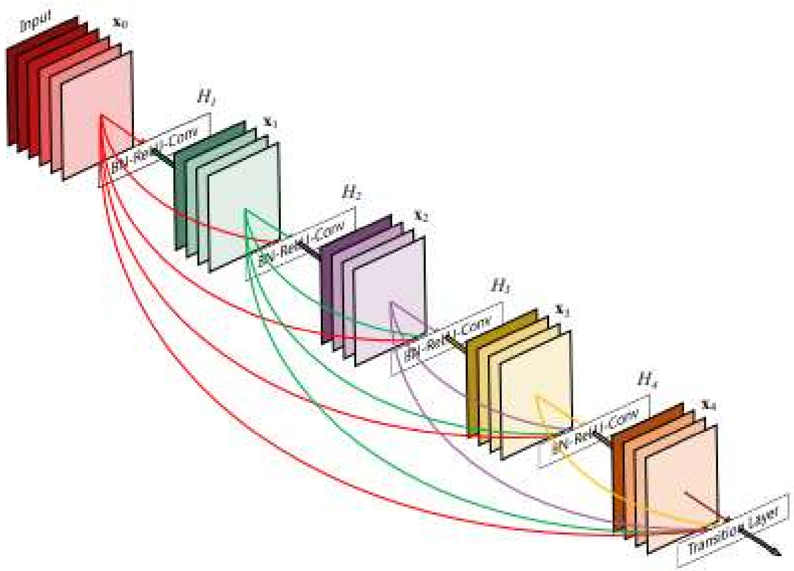
残差

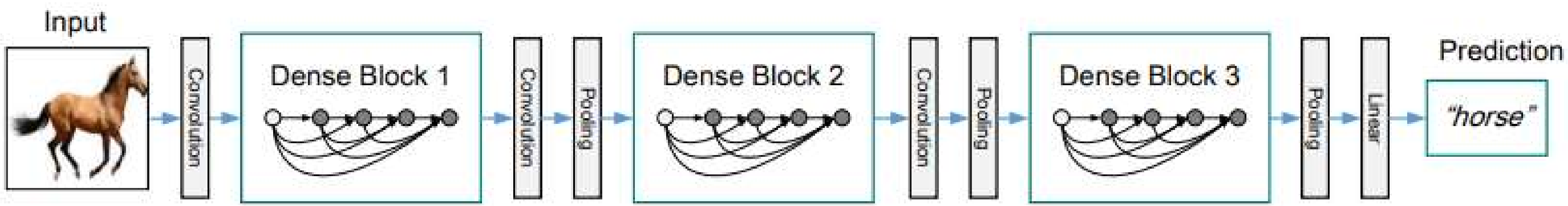
ResNet，DenseNet

残差结构



DenseNet


Input
x_0
H_1
BN-ReLU-Conv
x_1
H_2
BN-ReLU-Conv
x_2
H_3
BN-ReLU-Conv
x_3
H_4
BN-ReLU-Conv
x_4
Transition Layer

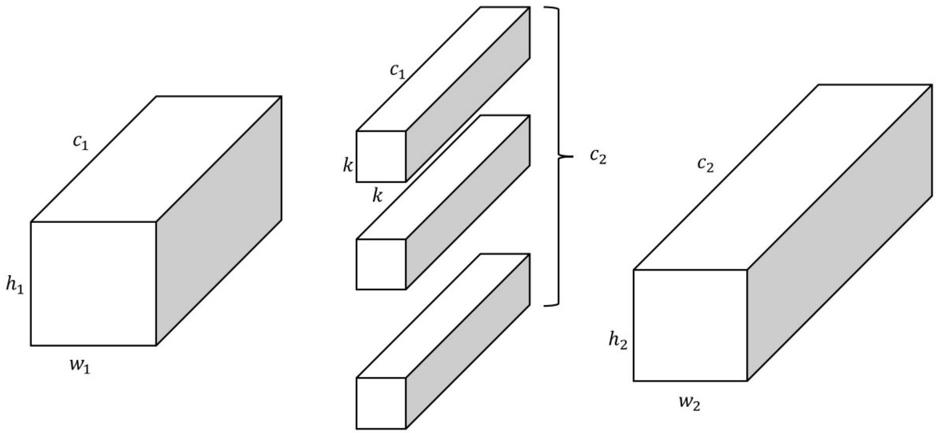

Input
Convolution
Dense Block 1
Convolution
Pooling
Dense Block 2
Convolution
Pooling
Dense Block 3
Pooling
Linear
Prediction
"horse"

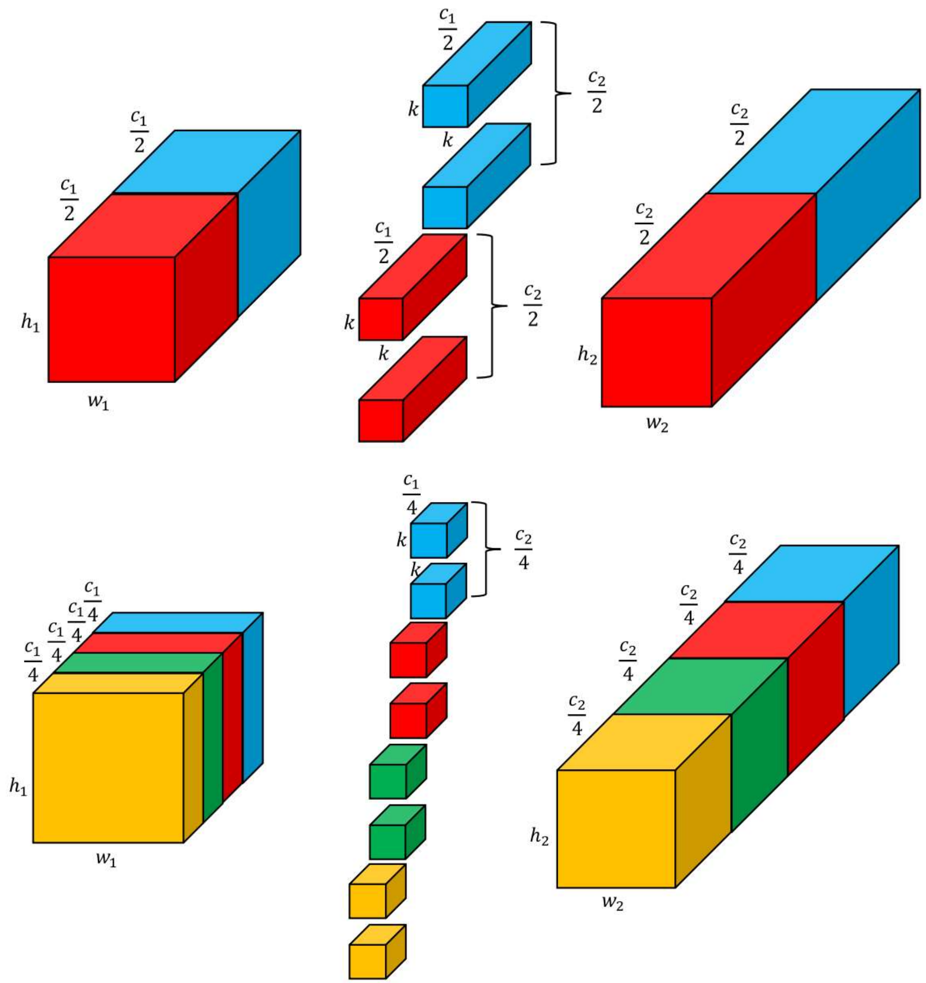
/04

轻量化

分组卷积，深度可分离卷积

分组卷积

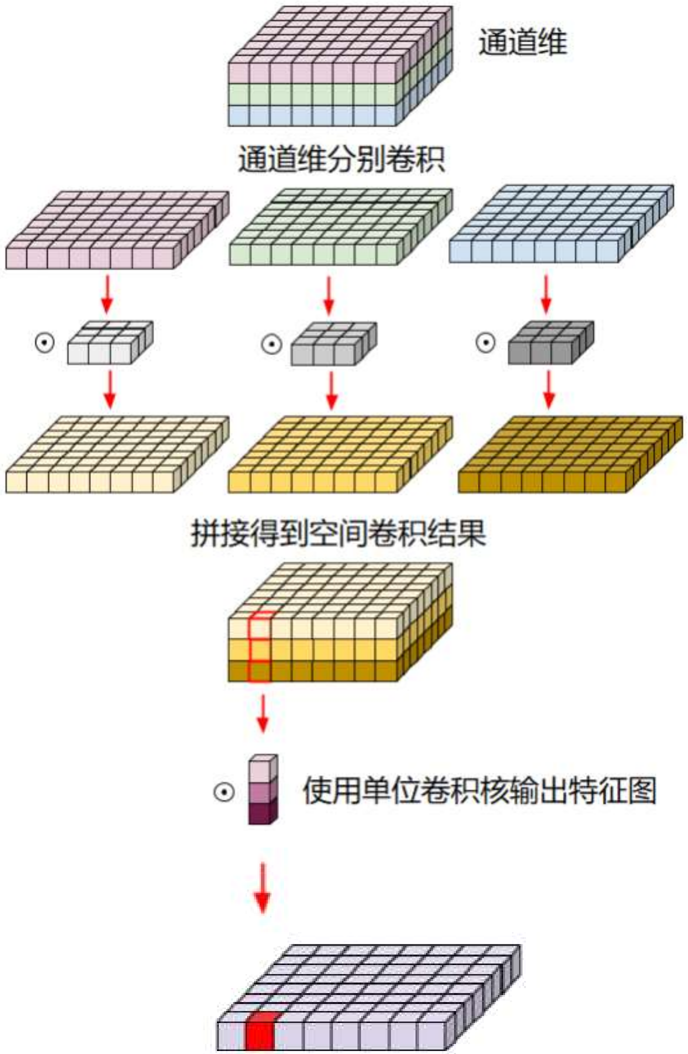




不分组时卷积的参数量： $n = k^2 c_1 c_2$

分m组时卷积的参数量： $n = mk^2 \frac{c_1}{m}\frac{c_2}{m} = \frac{k^2 c_1 c_2}{m}$

深度可分离卷积



Depth-wise Convolution

Point-wise Convolution

/05

大卷积核

RepLKNet

RepLKNet

Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs

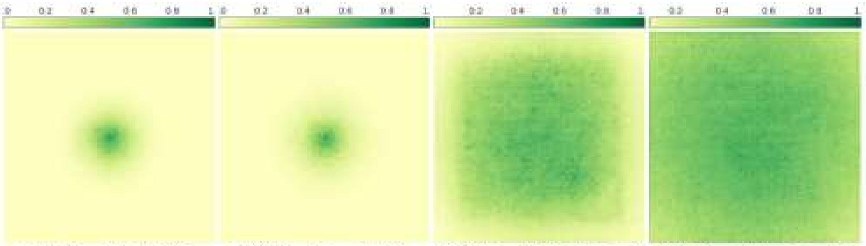
- 用少量大卷积核设计卷积神经网络，总结了5条大卷积核架构设计原则。
- 提出一种新的大卷积核神经网络架构RepLKNet，在分类、检测、分割上均显著强于传统CNN架构，取得了和主流Vision Transformers相似或更强的性能。
- 指出Self-Attention模块的大感受野是Vision Transformers取得优异性能的一个重要原因。使用大卷积核设计之后，在Shape bias等方面和ViTS的表现更加接近。

大卷积核的优势

› **提升有效感受野**

$$\sqrt{\mathrm{Var}[S_n]} = \sqrt{n}\sqrt{\sum_{m=0}^{k-1}\frac{m^2}{k} - \left(\sum_{m=0}^{k-1}\frac{m}{k}\right)^2} = \sqrt{\frac{n(k^2-1)}{12}} = O(k\sqrt{n})$$

› **回避深度增加带来的优化问题**



(A) ResNet-101 (B) ResNet-152 (C) RepLKNet-13 (D) RepLKNet-31

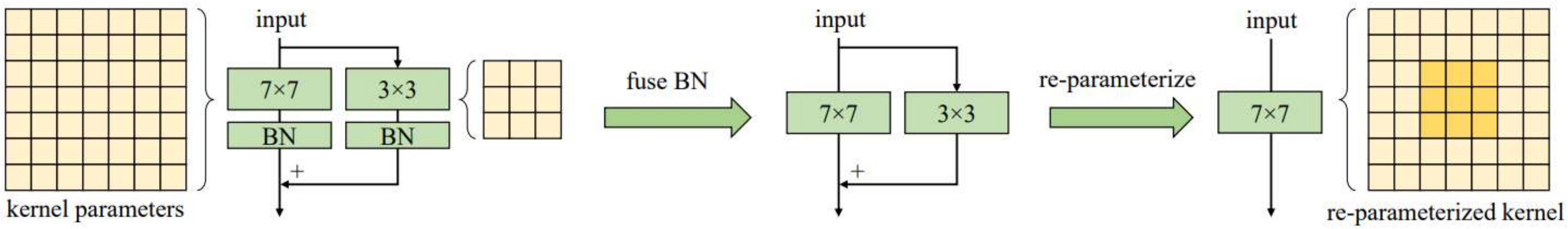
大卷积核结构改进方法

- 高效：使用DW等结构稀疏化卷积，辅以恰当的底层优化
- 残差结构

Table 2. Results of different kernel sizes in normal/shortcut-free MobileNet V2.

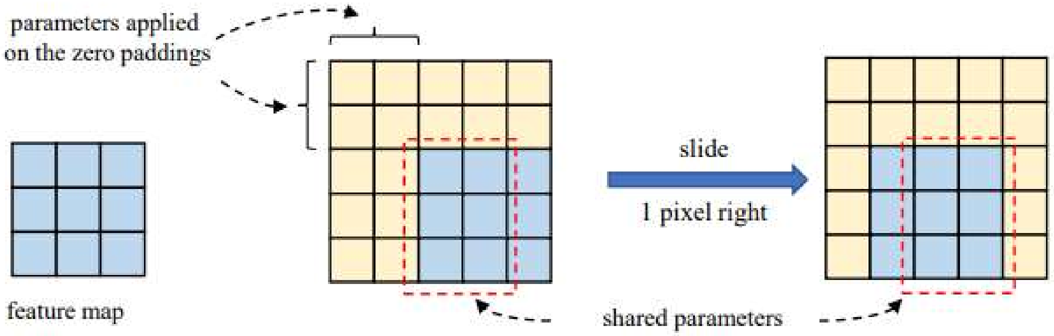
Shortcut	Kernel size	ImageNet top-1 accuracy (%)
✓	3×3	71.76
✓	13×13	**72.53**
	3×3	**68.67**
	13×13	53.98

- 用小卷积核重参数化大卷积核，避免过度平滑



大卷积核结构改进方法

› 在小特征图使用比特征图更大的卷积核

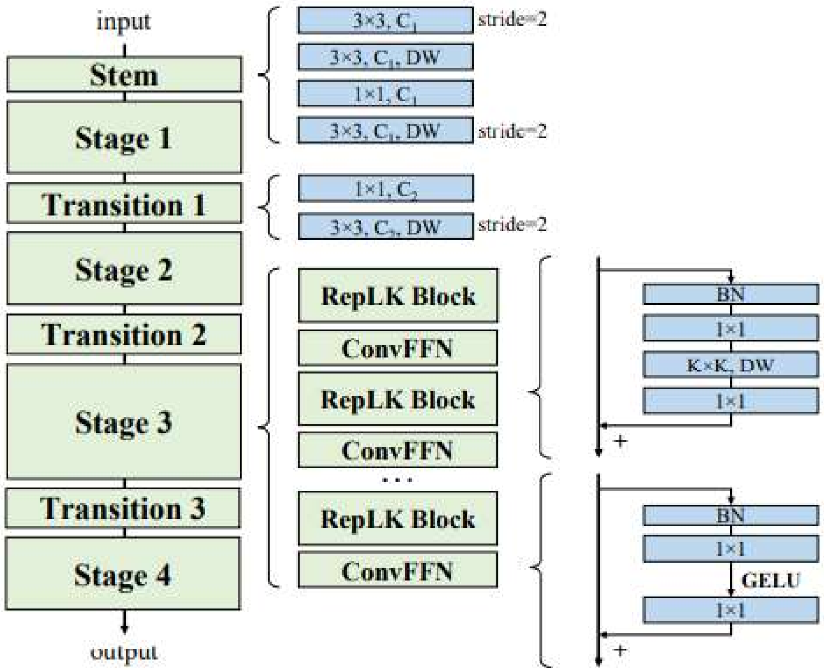


› 注意下游任务的性能

Kernel size	ImageNet acc (%)	Cityscapes mIoU (%)
3×3	71.76	72.31
7×7	**72.00**	74.30
13×13	71.97	**74.62**

RepLKNet

- 宏观架构参考Swin Transformer



实验结果

Kernel size	Architecture	ImageNet Top-1	ImageNet Params	ImageNet FLOPs	ADE20K mIoU	ADE20K Params	ADE20K FLOPs
7-7-7-7	ConvNeXt-Tiny	81.0	29M	4.5G	44.6	60M	939G
7-7-7-7	ConvNeXt-Small	82.1	50M	8.7G	45.9	82M	1027G
7-7-7-7	ConvNeXt-Base	82.8	89M	15.4G	47.2	122M	1170G
31-29-27-13	ConvNeXt-Tiny	81.6	32M	6.1G	**46.2**	64M	973G
31-29-27-13	ConvNeXt-Small	82.5	58M	11.3G	**48.2**	90M	1081G

Model	Input resolution	Top-1 acc	Params (M)	FLOPs (G)	Throu examp
RepLKNet-31B	224×224	83.5	79	15.3	295
Swin-B	224×224	83.5	88	15.4	226
RepLKNet-31B	384×384	84.8	79	45.1	97
Swin-B	384×384	84.5	88	47.0	67
RepLKNet-31B ‡	224×224	85.2	-	-	-
Swin-B ‡	224×224	85.2	-	-	-
RepLKNet-31B ‡	384×384	86.0	-	-	-
Swin-B ‡	384×384	86.4	-	-	-
RepLKNet-31L ‡	384×384	86.6	172	96.0	50
Swin-L ‡	384×384	87.3	197	103.9	36
RepLKNet-XL ◇	320×320	87.8	335	128.7	39

Backbone	Method	mIoU (ss)	mIoU (ms)	Param (M)	FLOPs (G)
RepLKNet-31B	UperNet	**49.9**	**50.6**	112	1170
ResNet-101	UperNet [102]	43.8	44.9	86	1029
ResNeSt-200 [112]	DeepLabv3 [15]	-	48.4	113	1752
Swin-B	UperNet	48.1	49.7	121	1188
Swin-B	UperNet + [38]	48.4	50.1	121	-
ViT-Hybrid	DPT-Hybrid [73]	-	49.0	90	-
ViT-L	DPT-Large	-	47.6	307	-
ViT-B	SETR-PUP [117]	46.3	47.3	97	-
ViT-B	SETR-MLA [117]	46.2	47.7	92	-
RepLKNet-31B ‡	UperNet	**51.5**	**52.3**	112	1829
Swin-B ‡	UperNet	50.0	51.6	121	1841
RepLKNet-31L ‡	UperNet	**52.4**	52.7	207	2404
Swin-L ‡	UperNet	52.1	**53.5**	234	2468
ViT-L ‡	SETR-PUP	48.6	50.1	318	-
ViT-L ‡	SETR-MLA	48.6	50.3	310	-
RepLKNet-XL ◇	UperNet	**55.2**	**56.0**	374	3431

Backbone	Method	AP^{box}	AP^{mask}	Param (M)	FLOPs (G)
RepLKNet-31B	FCOS	**47.0**	-	87	437
X101-64x4d	FCOS	42.6	-	90	439
RepLKNet-31B	Cas Mask	**52.2**	**45.2**	137	965
X101-64x4d	Cas Mask	48.3	41.7	140	972
ResNeSt-200	Cas R-CNN [9]	49.0	-	-	-
Swin-B	Cas Mask	51.9	45.0	145	982
RepLKNet-31B ‡	Cas Mask	**53.0**	**46.0**	137	965
Swin-B ‡	Cas Mask	**53.0**	45.8	145	982
RepLKNet-31L ‡	Cas Mask	**53.9**	46.5	229	1321
Swin-L ‡	Cas Mask	**53.9**	**46.7**	254	1382
RepLKNet-XL ◇	Cas Mask	**55.5**	**48.0**	392	1958

Thanks.

Thanks