背景

RepLKNet 的作者和 RepVGG 的一作是同一个人。受 ViT 的启发，作者使用几个大卷积核去代替堆叠多个小卷积核，并将结构化重参数化、深度卷积融合，使得超大卷积更强更快，基于这些操作提出了一种大量采用超大卷积核的 RepLKNet 网络。

尽管已经有很多研究使用了大卷积核，但是这些文章并没有交代清楚如何将大卷积核应用到卷积神经网络中，因此也不清楚应用大卷积核能否弥补纯卷积神经网络与 Tranformers 之间的差距。

所以这篇文章旨在弥补纯卷积网路和 ViT 之间的性能差。为了实现这一目的，作者将网络的设计聚焦于实现网络深度增加时，其有效的感受野也同步扩大，对此，作者提出了两个问题“使用几个大卷积核去代替堆叠多个小卷积核会怎样？”，“通过搭建网络实现感受野逐步扩大是关键点么？”

思考

VGG 主要贡献就是使用多个小卷积核代替较大的卷积核，比如，两个 3×3 卷积核的堆叠相当于 5×5 卷积核的感受野，3 个 3×3 卷积核的堆叠相当于 7×7 卷积核的感受野。既减少了参数，又增加了更多的非线性映射，提高了拟合能力，但是这篇文章中为什么又认为用大卷积核替代多个小卷积堆叠效果更好呢？

首先，有效感受野核 $O(K\sqrt{L})$ 成正比，其中 K 是卷积核尺寸，L 是网络的层数，可以看出有效感受野随着卷积核尺寸增大而线性增长，随着网络深度次线性增长；

其次，增加网络深度使得网络优化成问题，ResNet 看似解决了深度网络的优化问题，但是其本质是多个浅层网络的映射，从下图的红色框可以看出，其有效的感受野并没有随着网络深度的增加而增大。反观本文的思想，从下图可以看出随着网络深度的增加，感受野也在同步扩大。

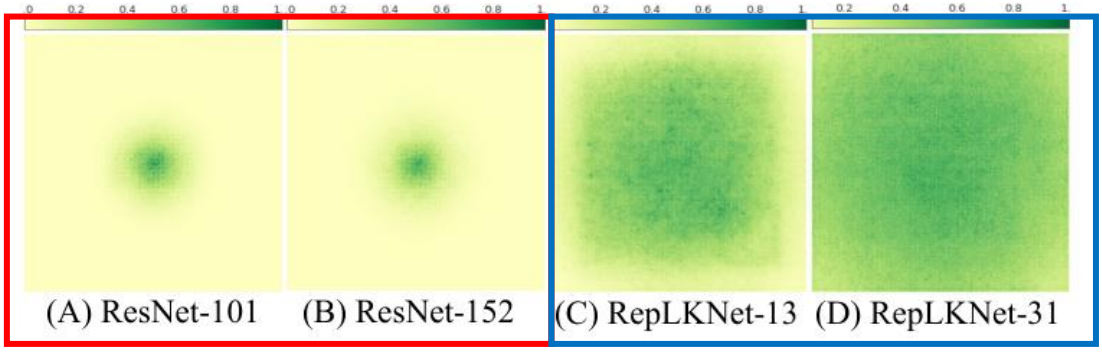

图 1

最后，大卷积核可以学到更多的形状信息，这一点更符合人类认知。

方法

对于计算机视觉中的一些下游任务来说，卷积核大小对模型的性能影响很大，但这一影响因素在设计网络模型时常被忽略，这篇文章就应用大卷积核提出了 5 条设计准则。

①大的深度卷积核可以更高效

Resolution R	Impl	Latency (ms) @ Kernel size									
		3	5	7	9	13	17	21	27	29	31
16×16	Pytorch	5.6	11.0	14.4	17.6	36.0	57.2	83.4	133.5	150.7	171.4
	Ours	5.6	6.5	6.4	6.9	7.5	8.4	8.4	8.4	8.3	8.4
32×32	Pytorch	21.9	34.1	54.8	76.1	141.2	230.5	342.3	557.8	638.6	734.8
	Ours	21.9	28.7	34.6	40.6	52.5	64.5	73.9	87.9	92.7	96.7
64×64	Pytorch	69.6	141.2	228.6	319.8	600.0	977.7	1454.4	2371.1	2698.4	3090.4
	Ours	69.6	112.6	130.7	152.6	199.7	251.5	301.0	378.2	406.0	431.7

图 2

毋庸置疑，增大卷积核带来的必然是计算量和参数量的增加，从 MobileNetV1 中知道通过应用深度可分离卷积中的 Depth-wise 卷积可以大幅减少参数量和计算量。因此将大卷积核和 Depth-wise 卷积融合。为了解决大卷积核的计算优化问题，设计了一种高效的算法，从图 2 中可以看出达到了低延迟的效果。

②可以通过使用跳跃连接弥补大卷积核导致的过渡平滑

大卷积难以捕获局部特征信息，容易导致过渡平滑，为了弥补这一细节缺失，使用跳跃连接。下图是以 MobileNetV2 作为基准点，通过增加卷积核尺寸在

ImageNet 数据集上训练 100 轮得到的准确率对比。

绿色两框对比后可以看出，使用跳跃连接后其准确率有明显提升，从红色的框中可以看出使用了跳跃连接后，增加卷积核尺寸可以提高准确率。

因此在增加卷积核尺寸的同时使用跳跃连接，既可以扩大有效感受野，又保证局部信息的获取能力。

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

图 3

③结构重参数化将小卷积融合到大卷积中

Kernel	3×3 re-param	ImageNet top-1 acc (%)
3×3	N/A	71.76
9×9		72.67
9×9	✓	**73.09**
13×13		72.53
13×13	✓	**73.24**

图 4

从两个红框中的准确率对比可以看出，使用结构重参数化将 3×3 卷积融合到 9×9 卷积核中可以提到准确率。在没有使用结构重参数化的前提下，单纯的将卷积核尺寸从 9×9 增加到 13×13 其准确率有轻微下降，但是在使用了结构重参数化后可以看出准确率从 73.09%提高到 73.24%。因此对小卷积核使用结构重参数化是有效的办法。

④大卷积核在下游任务中表现更加优越

人类识别物体一般是通过物体的形状特征信息，而不是纹理特征，除此之外，ViT 能有如此高的性能也和其强形状偏置假设有关，而传统的卷积神经网络更倾向于提取纹理信息，因此，增强纯卷积神经网络的形状偏置可以弥补纯卷积神经

网络与 Tranformers 之间的差距。随后发现简单地增大卷积核可以有效地提高形状偏置。

⑤将大卷积核应用到小特征图上更有效

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

图 5

从上图的 MobileNetV2 的结构框图可以看出最后一个 bottleneck 输入的特征图大小为 7×7，将其卷积核大小从原来的 3×3 换为 7×7 以及 13×13，从下图所示结果看出，即使所使用的卷积核比特征图大，其准确率也有所提升。

Table 4. Results of various kernel sizes in the *last stage* of MobileNet V2. Kernel sizes in previous stages remain to be 3×3.

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

图 6

总结

这篇文章摒弃了传统的通过堆叠多个小卷积核替代大卷积的思路，认为之前的思路并没有增加有效的感受野。作者通过结构重参数化以及 Depth-wise 卷积解决大卷积核带来的参数和计算量的问题，从而使用大卷积核直接实现感受野的增加。