Xception:2017

Deep Learning with Depthwise Separable Convolutions

论文出发点或背景

卷积的方式在计算机视觉领域已经成为了主流算法，Google的Inception系列在各大数据集上面表现都不错

一个单一的卷积核的任务是同时映射跨通道相关性和空间相关性

Inception模块背后的想法是，通过将其明确地分解为一系列可独立查看跨通道相关性和空间相关性的操作，从而使此过程更轻松、更高效

Inception背后的基本假设是，跨通道相关性和空间相关性被充分解耦，最好不要联合映射它们

Figure 1. A canonical Inception module (Inception V3).

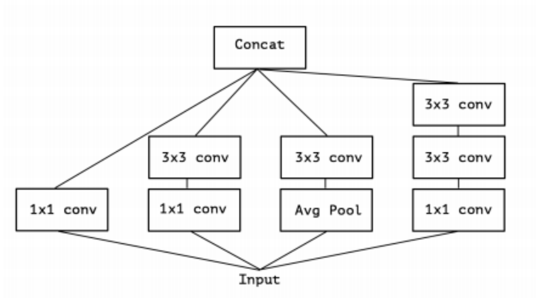
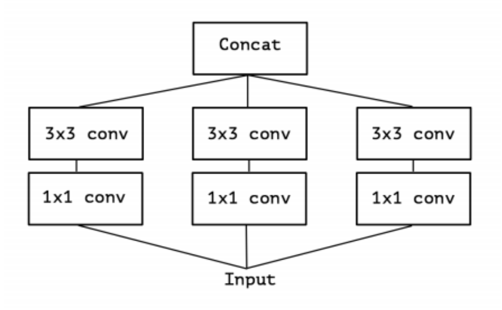
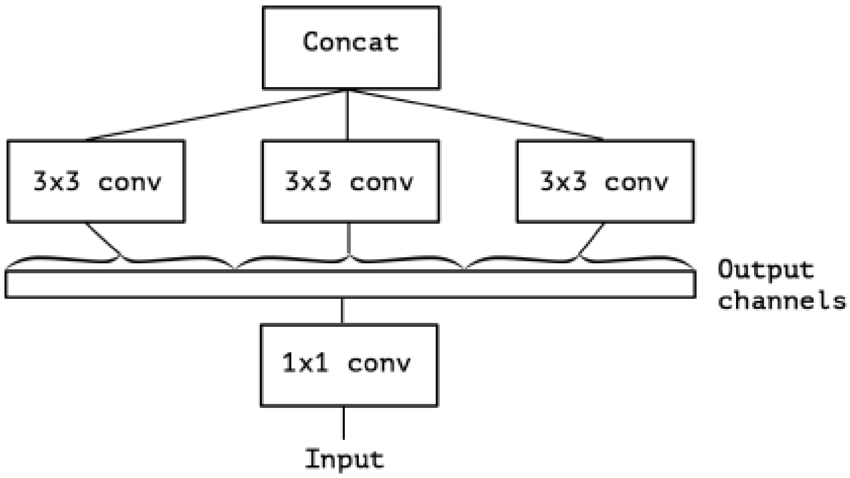


Figure 2. A simplified Inception module.



可以将该Inception模块重新构造为大的1x1卷积，然后再进行空间卷积，这些卷积将在输出通道的非重叠段上进行操作

Figure 3. A strictly equivalent reformulation of the simplified Inception module.

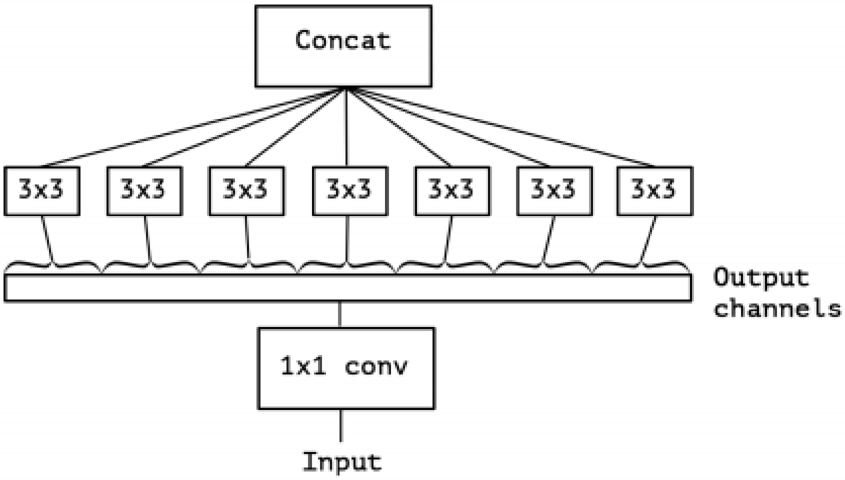


论文创新思路

分区中的分组数（及其大小）会产生什么影响？ 做出比Inception假设强得多的假设，并假设跨通道相关性和空间相关性可以完全分开映射，是否合理？

inception简化版本的极端模式：首先使用1x1卷积来映射跨通道相关性，然后分别映射每个输出通道的空间相关性

Figure 4. An “extreme” version of our Inception module, with one spatial convolution per output channel of the 1x1 convolution.



Inception模块的“极端”版本与深度可分离卷积之间的两个小区别是：

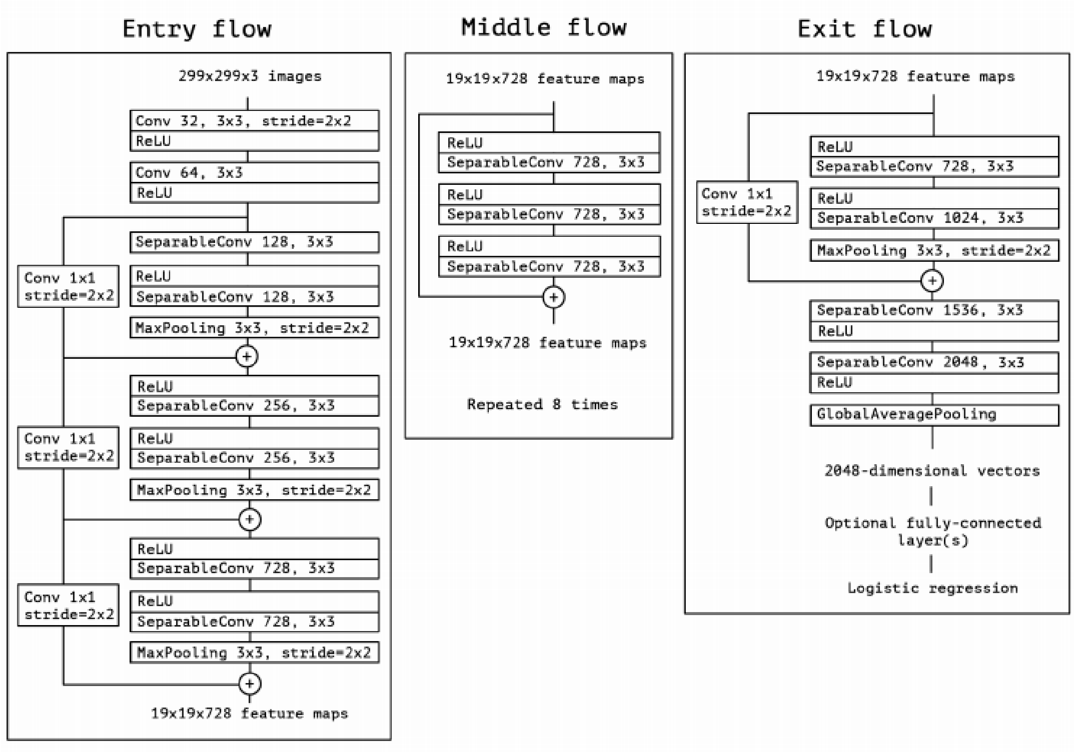
1）操作顺序：通常实现的深度可分离卷积首先执行通道空间卷积（深度卷积），然后执行1x1卷积，而Inception首先执行1x1卷积。

2）第一次操作后是否存在非线性。在Inception中，两个操作都跟随ReLU非线性，但是通常在没有非线性的情况下实现深度可分离卷积

论文方法介绍

通过用深度可分离卷积代替Inception模块，即通过构建将深度可分离卷积堆叠的模型，可以改进Inception系列体系结构

Figure 5. The Xception architecture: the data first goes through the entry flow, then through the middle flow which is repeated eight times, and finally through the exit flow. Note that all Convolution and SeparableConvolution layers are followed by batch normalization [7] (not included in the diagram). All SeparableConvolution layers use a depth multiplier of 1 (no depth expansion).



实际效果

Xception架构和Inception v3具有相同数量的参数，因此性能的提升并不是因为网络容量的增加，而是由于更有效地使用模型的参数

Table 1. Classification performance comparison on ImageNet (single crop, single model). VGG-16 and ResNet-152 numbers are only included as a reminder. The version of Inception V3 being benchmarked does not include the auxiliary tower.

	Top-1 accuracy	Top-5 accuracy
VGG-16	0.715	0.901
ResNet-152	0.770	0.933
Inception V3	0.782	0.941
Xception	**0.790**	**0.945**

Table 2. Classification performance comparison on JFT (single crop, single model).

	FastEval14k MAP@100
Inception V3 - no FC layers	6.36
Xception - no FC layers	6.70
Inception V3 with FC layers	6.50
Xception with FC layers	**6.78**

Figure 6. Training profile on ImageNet

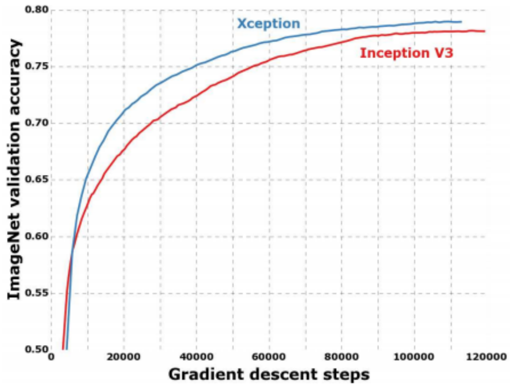


Figure 7. Training profile on JFT, without fully-connected layers

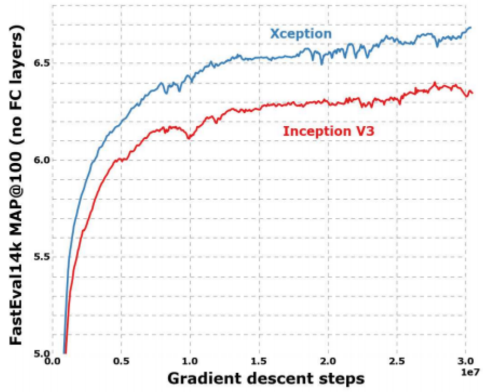


Table 3. Size and training speed comparison.

	Parameter count	Steps/second
Inception V3	23,626,728	31
Xception	22,855,952	28

Figure 8. Training profile on JFT, with fully-connected layers

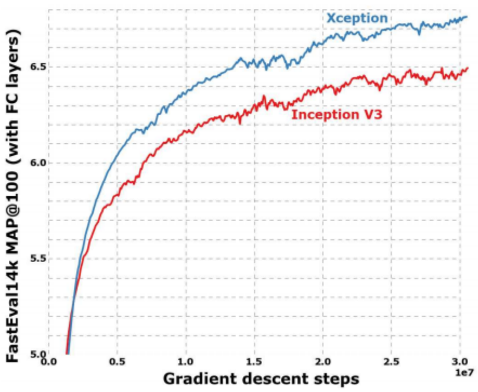


Figure 9. Training profile with and without residual connections.

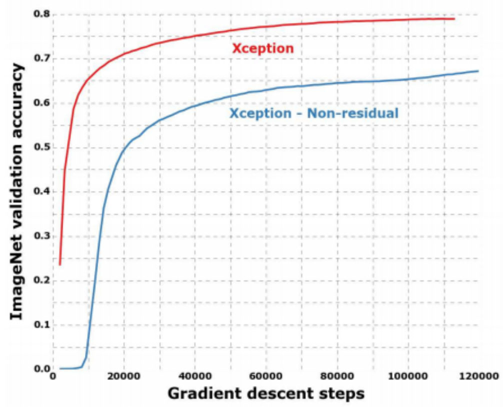
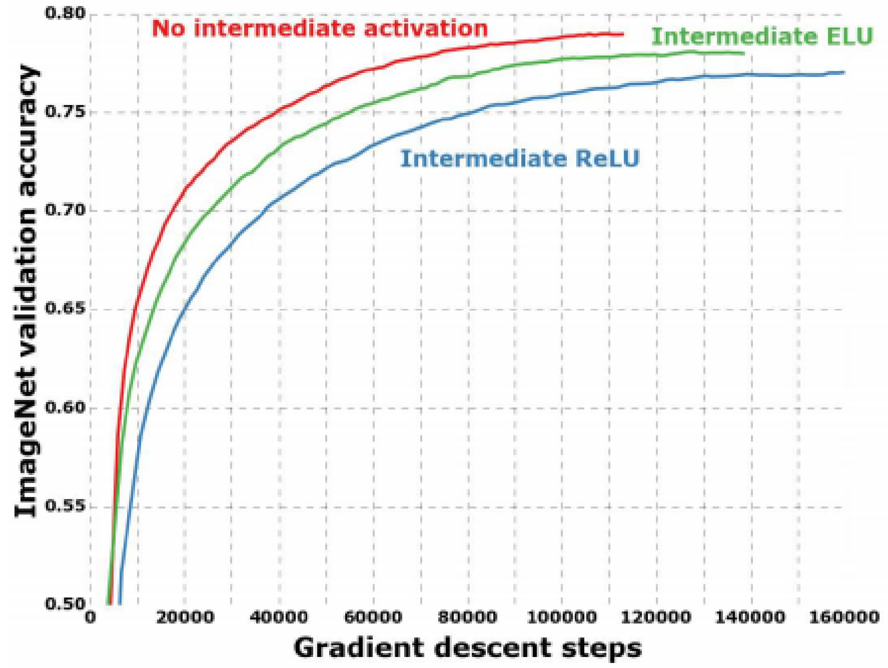


Figure 10. Training profile with different activations between the depthwise and pointwise operations of the separable convolution layers.



个人理解

该文章的话主要就是在InceptionNet的基础上引入深度可分离卷积，不增加网络复杂度和参数量的前提之下，提高模型效果，感觉没什么特别亮的创新点，给的最大启发就是改进的时候可以去思考一个网络的本质，把一个网络按照这种本质思想去做到极致，然后基于这种极端情况进行修改，再就是可以通过对激活函数的有无等一些方面做分析，有点像数学建模里面的敏感性分析，然后扩充文章内容。