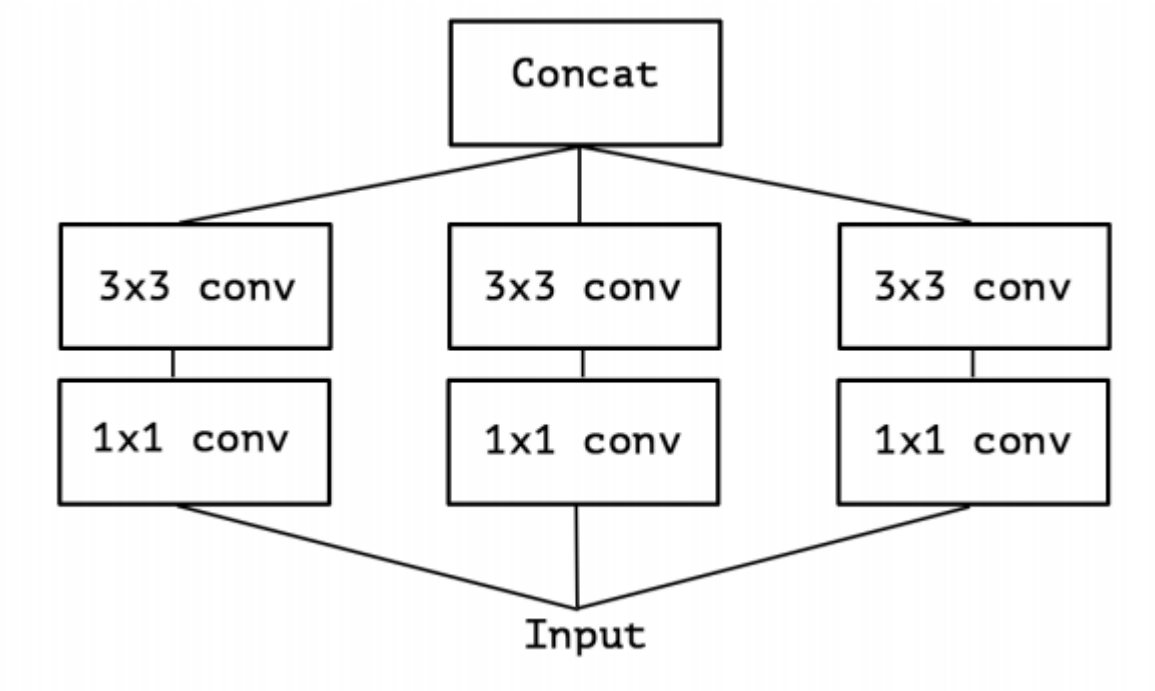
Xception:

Deep Learning with Depthwise Seaprable Convolutionns

사공진

목차

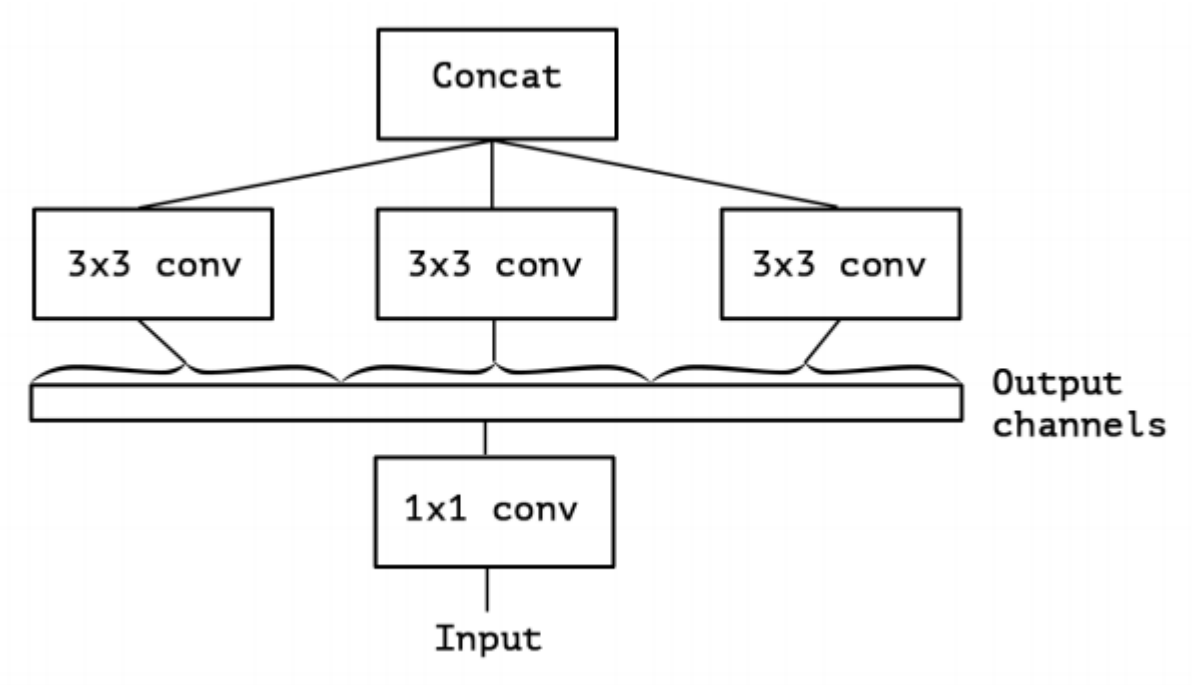
A simplified Inception module



1x1 conv에서 cross-channel correlation 계산 이후, 3x3 conv에서 spatial correlation 계산.

이때, 서로 독립적으로 수행

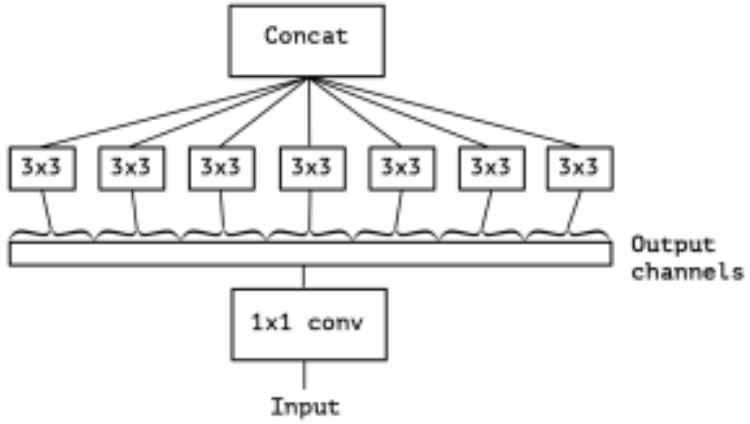
A strictly equivalent reformulation of the simplified Inception module



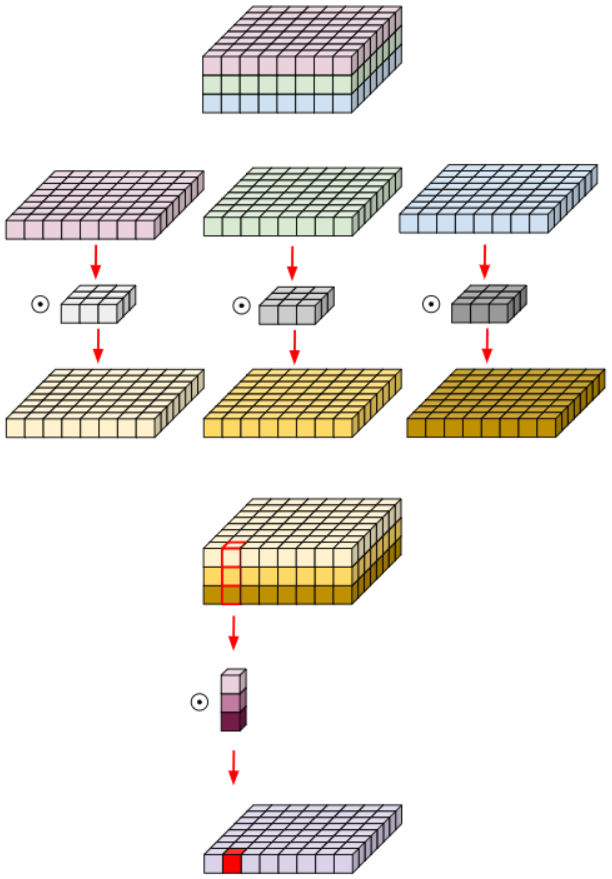
An ‘extreme’ version of our Inception module

what is the effect of the number of segments in the partition (and their size)?

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



Depthwise Separable Convolution



기존 convolution에 비해 Parameter를 줄임으로, 연산량을 줄이고, 속도를 높이는 기법

Extreme ver of Inception **VS** Depthwise Separable Convolution

1.순서

Xception: pointwise -> depthwise

DS conv: depthwise -> pointwise

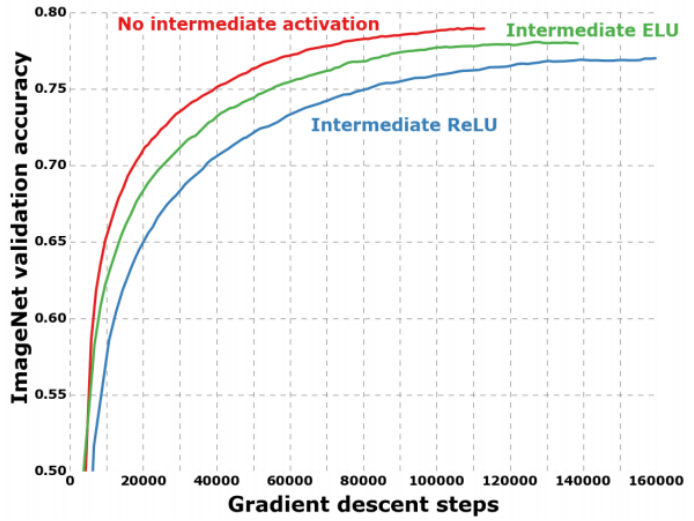
But, 스택 설정에서 사용되기에 중요한 사항 x

2.비선형성

기존 Inception: 첫 연산 후 RELu가 있다.

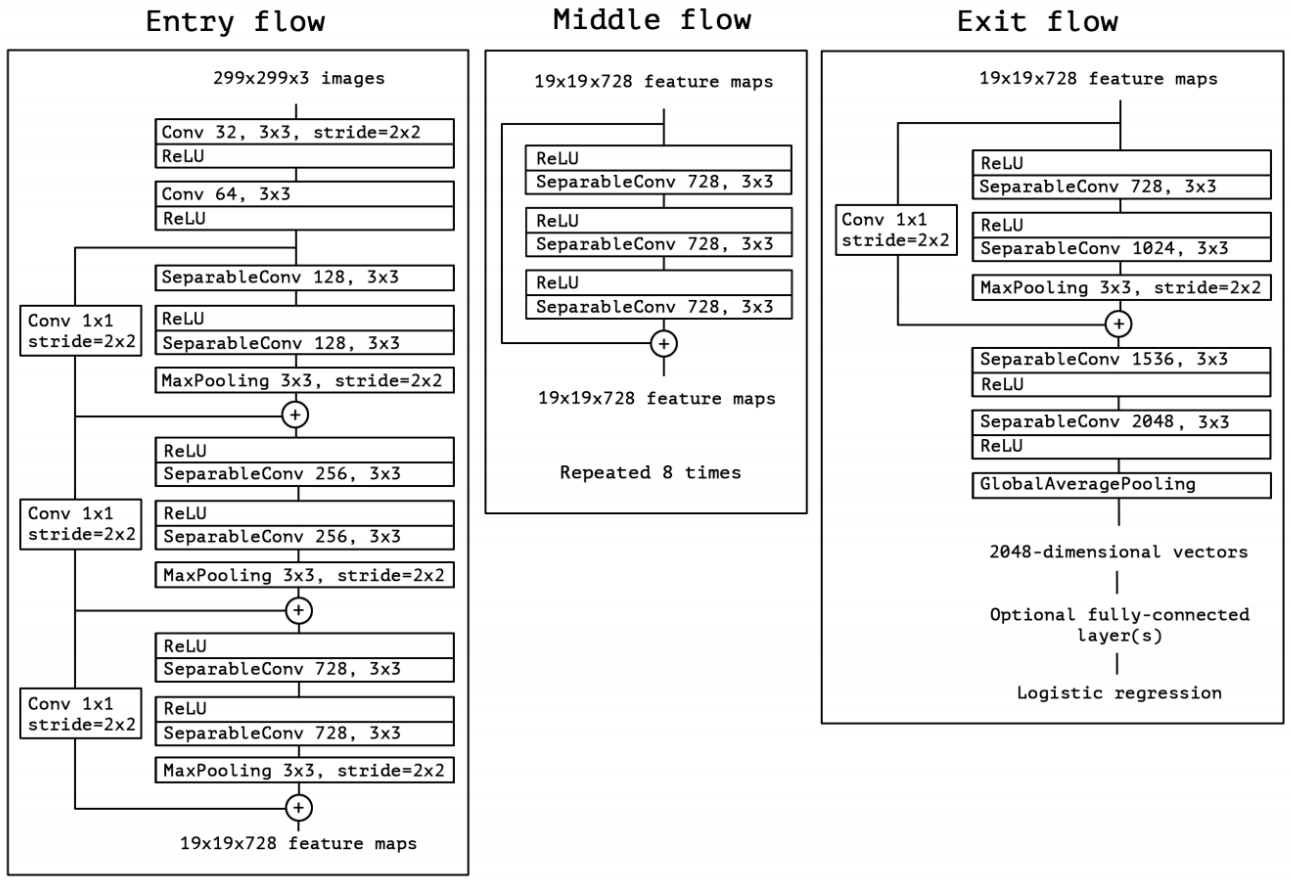
DS conv: 중간에 ReLu를 적용 X

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



Xception architecture

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).



Experiment result

1.ImageNet

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 3. Size and training speed comparison.

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

다른 모델에 비해 높은 accuracy

\+ 연산량 감소

Experiment result

1.JFT dataset

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**

ImageNet에 비해 더 큰 개선

Conclusion

높은 성능과 연산량 감소의 장점 때문에 CNN의 설계의 기초가 될 것으로 기대

참고사항

- https://www.youtube.com/watch?v=V0dLhyg5_Dw