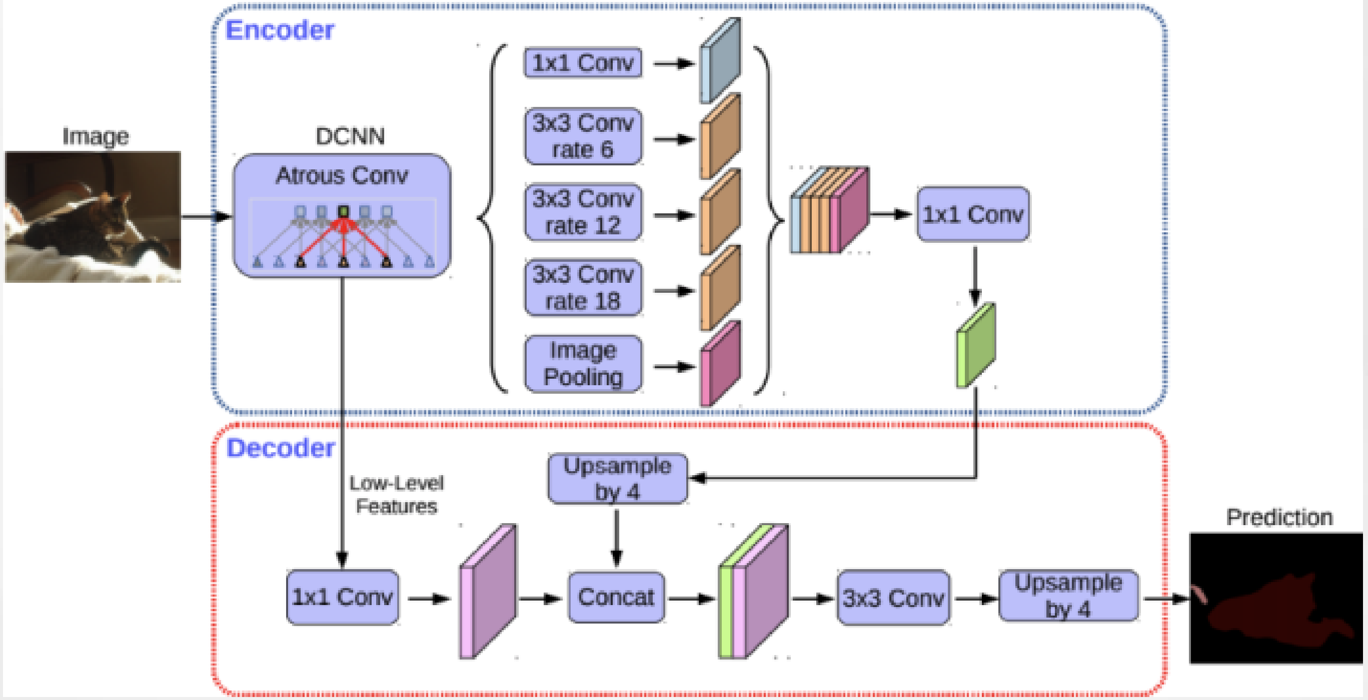
Chapter 03. 이미지 처리 분야 딥러닝 모델 (Image Segmentation)

DeepLab v3+



Deep Lab

1. DeepLabv1 (2015) : Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs

2. DeepLabv2 (2017) : DeepLab : Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs

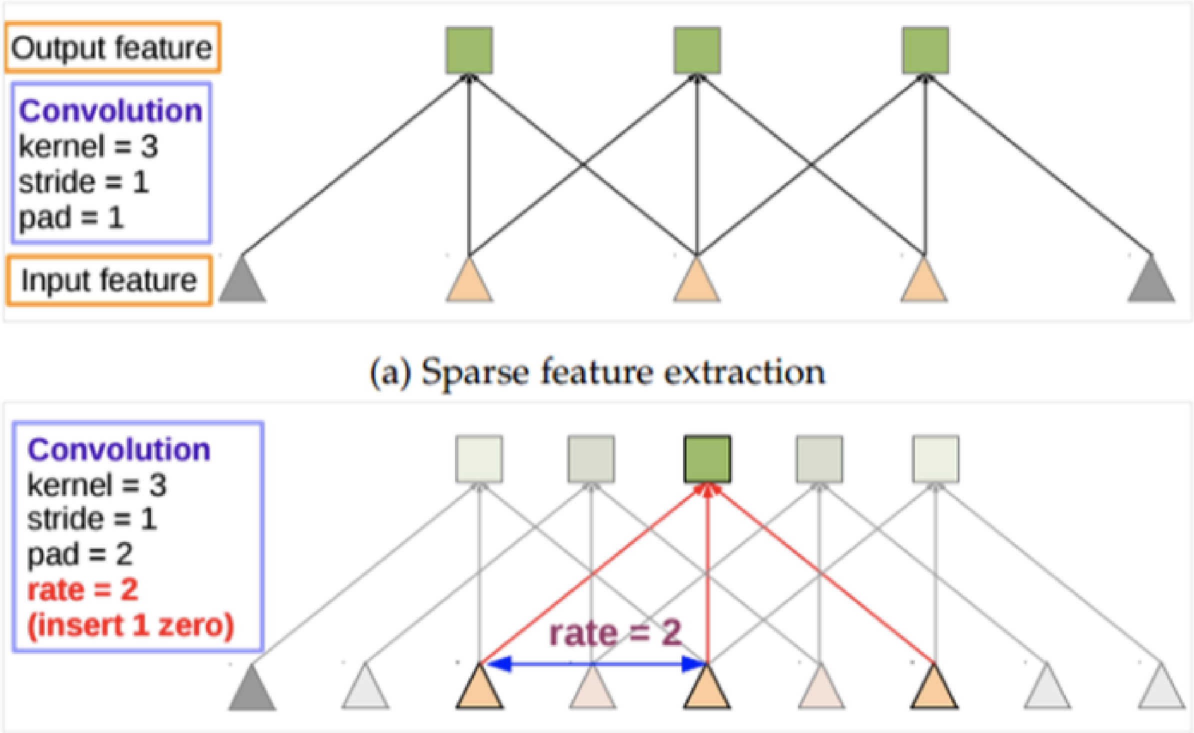
3. DeepLabv3 (2017) : Rethinking Atrous Convolution for Semantic Image Segmentation

4 DeepLabv3+ (2018) : Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation

https://laonple.blog.me/221000648527

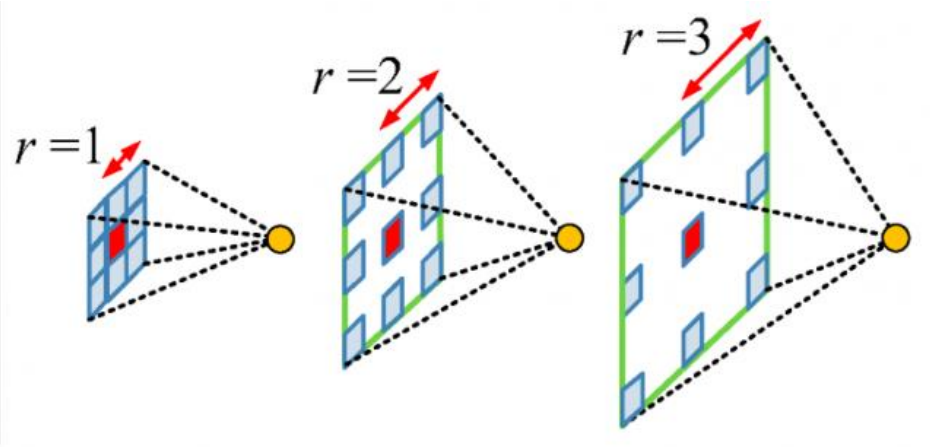
Astrous Convolution

Wavelet 신호 분석에서 주로 사용
Expanding Receptive Field with hole convolution



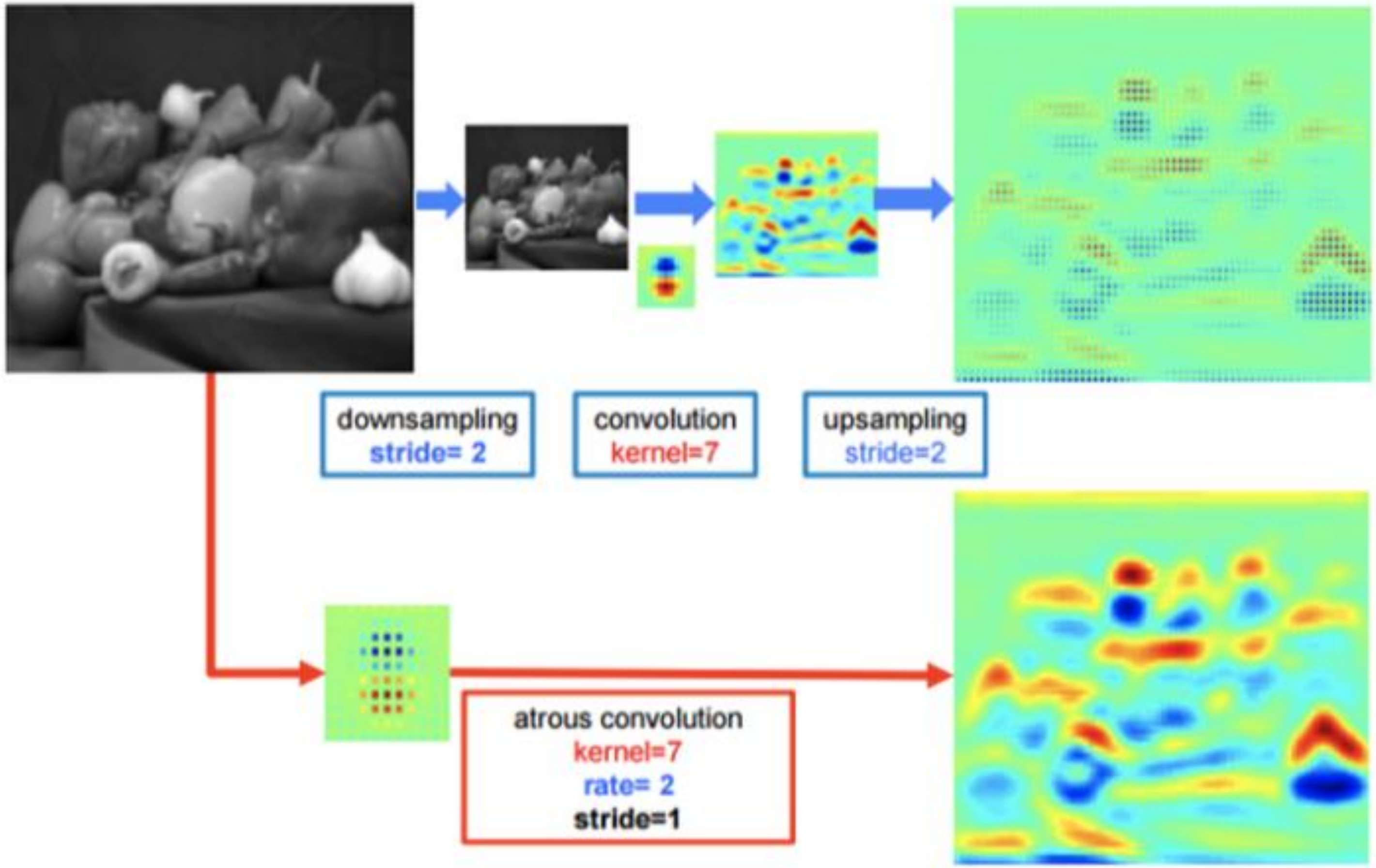
(a) Sparse feature extraction

(b) Dense feature extraction

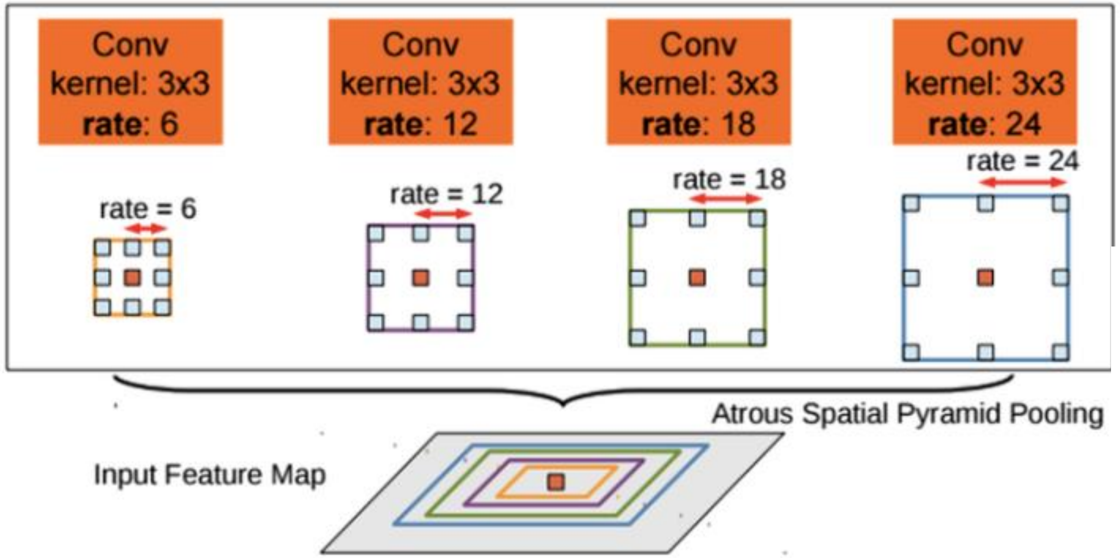


https://laonple.blog.me/221000648527

Astrous Convolution

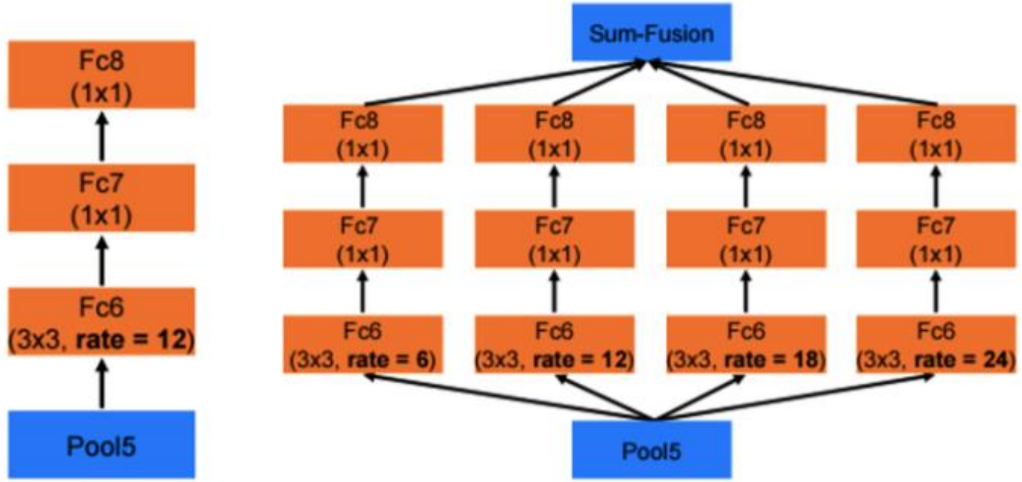


Spatial Pyramid Pooling



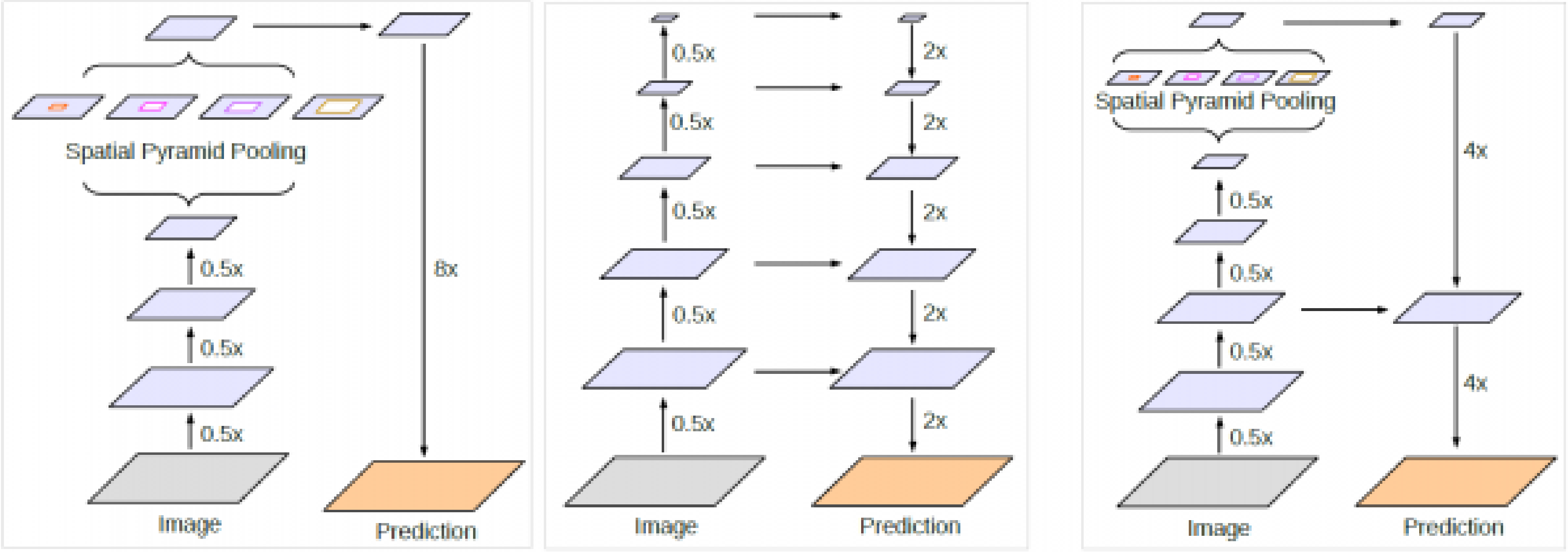
DeepLab v2

다양한 Receptive Field 에 대응



(a) DeepLab-LargeFOV (b) DeepLab-ASPP

Spatial Pyramid Pooling

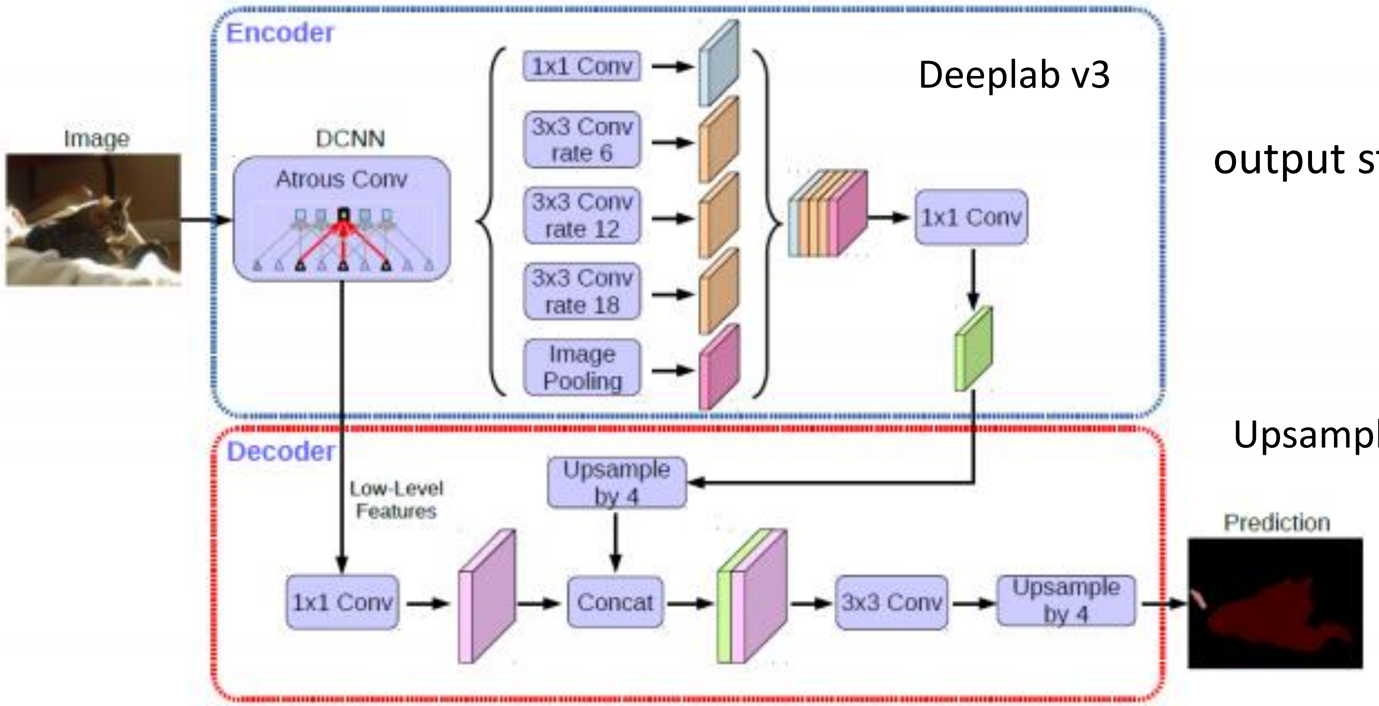


(a) Spatial Pyramid Pooling (b) Encoder-Decoder (c) Encoder-Decoder with Atrous Conv

Fig. 1. We improve DeepLabv3, which employs the spatial pyramid pooling module (a), with the encoder-decoder structure (b). The proposed model, DeepLabv3+, contains rich semantic information from the encoder module, while the detailed object boundaries are recovered by the simple yet effective decoder module. The encoder module allows us to extract features at an arbitrary resolution by applying atrous convolution.

fast campus

Deep Lab v3+



output stride

Upsampling : Unet(concat)

Fig. 2. Our proposed DeepLabv3+ extends DeepLabv3 by employing a encoder-decoder structure. The encoder module encodes multi-scale contextual information by applying atrous convolution at multiple scales, while the simple yet effective decoder module refines the segmentation results along object boundaries.

fast campus

https://pytorchhair.gitbook.io/project/undefined/deeplab-v3+

Changed Xception Backbone

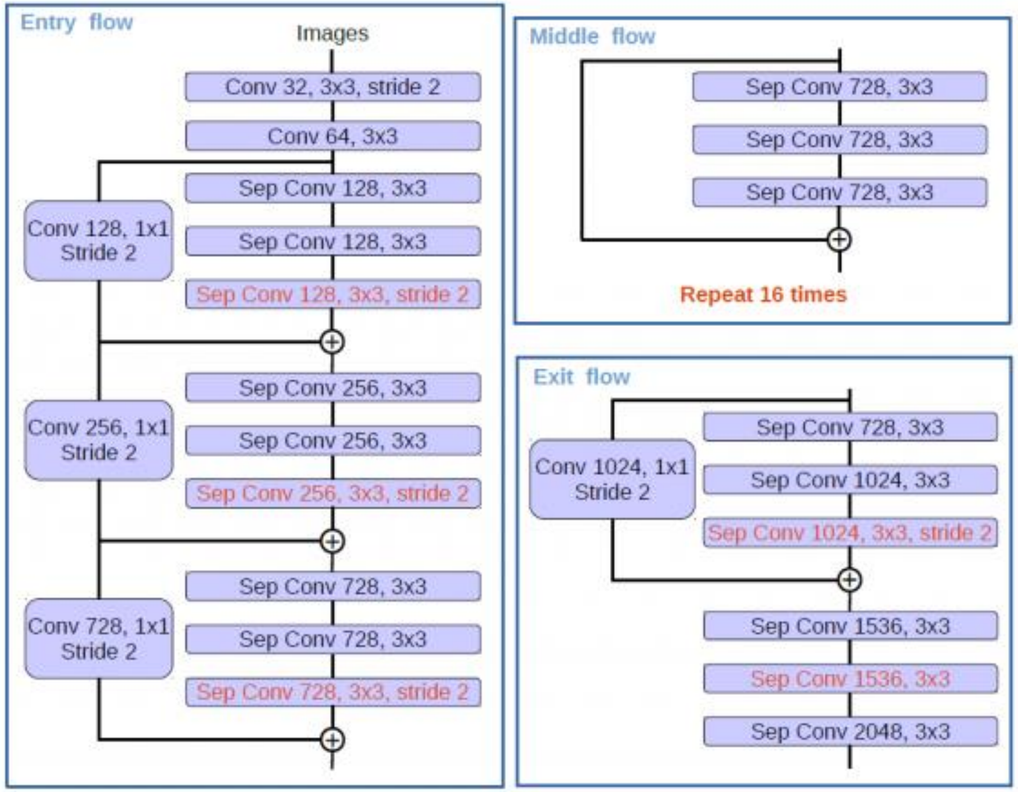


Fig. 4. We modify the Xception as follows: (1) more layers (same as MSRA's modification except the changes in Entry flow), (2) all the max pooling operations are replaced by depthwise separable convolutions with striding, and (3) extra batch normalization and ReLU are added after each 3×3 depthwise convolution, similar to MobileNet.

1) Atrous separable convolution을 적용하기 위해 모든 pooling operation을 depthwise separable convolution으로 대체.

1) 각각의 3 x 3 depthwise convolution 이후에 추가적으로 bath normalization과 ReLU 활성화 함수를 추가.

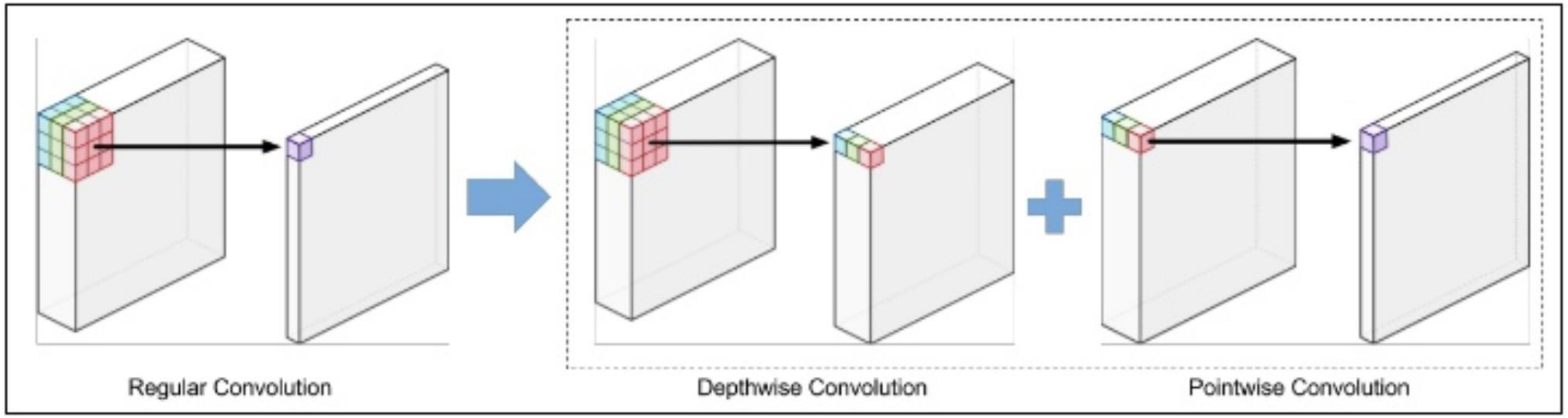
Model	Top-1 Error	Top-5 Error
Reproduced ResNet-101	22.40%	6.02%
Modified Xception	20.19%	5.17%

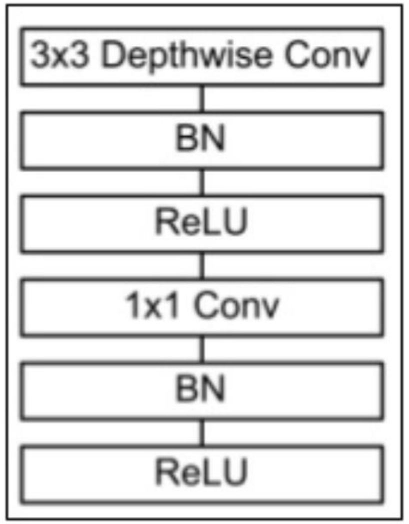
Table 4. *Single-model* error rates on ImageNet-1K validation set.

Depthwise Seperable Conv

https://www.slideshare.net/NaverEngineering/designing-more-efficient-convolution-neural-network-122869307

	Baseline (Original Convolution)	Proposed (Depth-wise Separable Convolution)
# of params.	$N(CK^2+1)$ $= NCK^2 + N$	$C(K^2+1) + N(C+1)$ $= CK^2 + NC + C + N$
# of operations.	$CHWK^2N$	$CHWK^2 + CHWN$ $= CHW(K^2+N)$





Decoder effect

Encoder train OS	Encoder eval OS	Decoder	MS	Flip	SC	COCO	JFT	mIOU	Multiply-Adds
16	16							79.17%	68.00B
16	16		✓					80.57%	601.74B
16	16		✓	✓				80.79%	1203.34B
16	8							79.64%	240.85B
16	8		✓					81.15%	2149.91B
16	8		✓	✓				81.34%	4299.68B
16	16	✓						79.93%	89.76B
16	16	✓	✓					81.38%	790.12B
16	16	✓	✓	✓				81.44%	1580.10B
16	8	✓						80.22%	262.59B
16	8	✓	✓					81.60%	2338.15B
16	8	✓	✓	✓				81.63%	4676.16B

Image w/ BU w/ Decoder

mIoU(mean Intersection over union)

Jaccard Overlap = IoU = $\frac{\text{Area of Overlap}}{\text{Area of Union}}$

Thank You