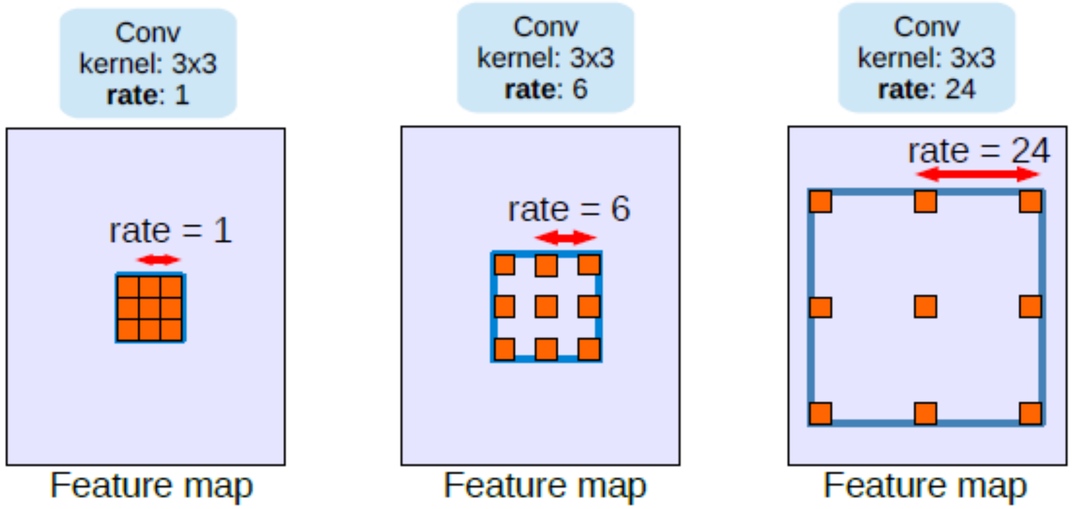
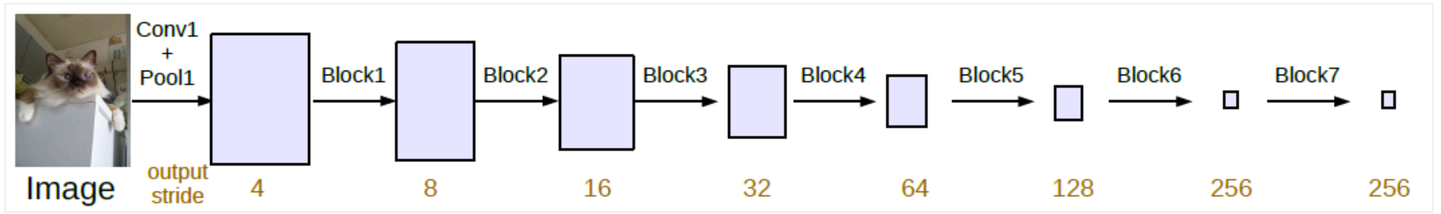
DeepLabv3+

Dilation, Depthwise, U-Net

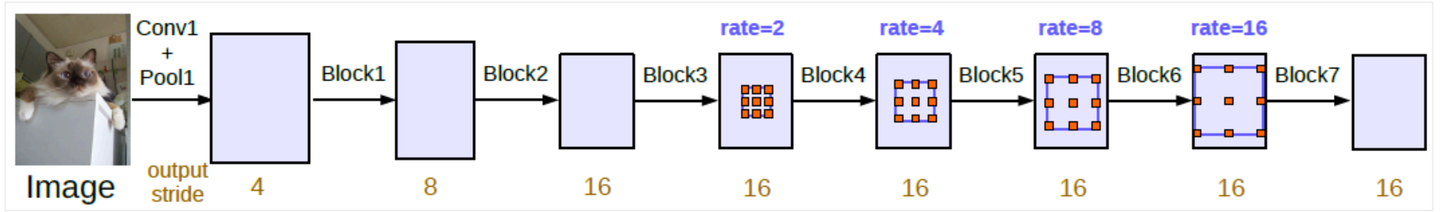
Atrous (Dilated) Convolution



Increase Rate→ Increase Field of View



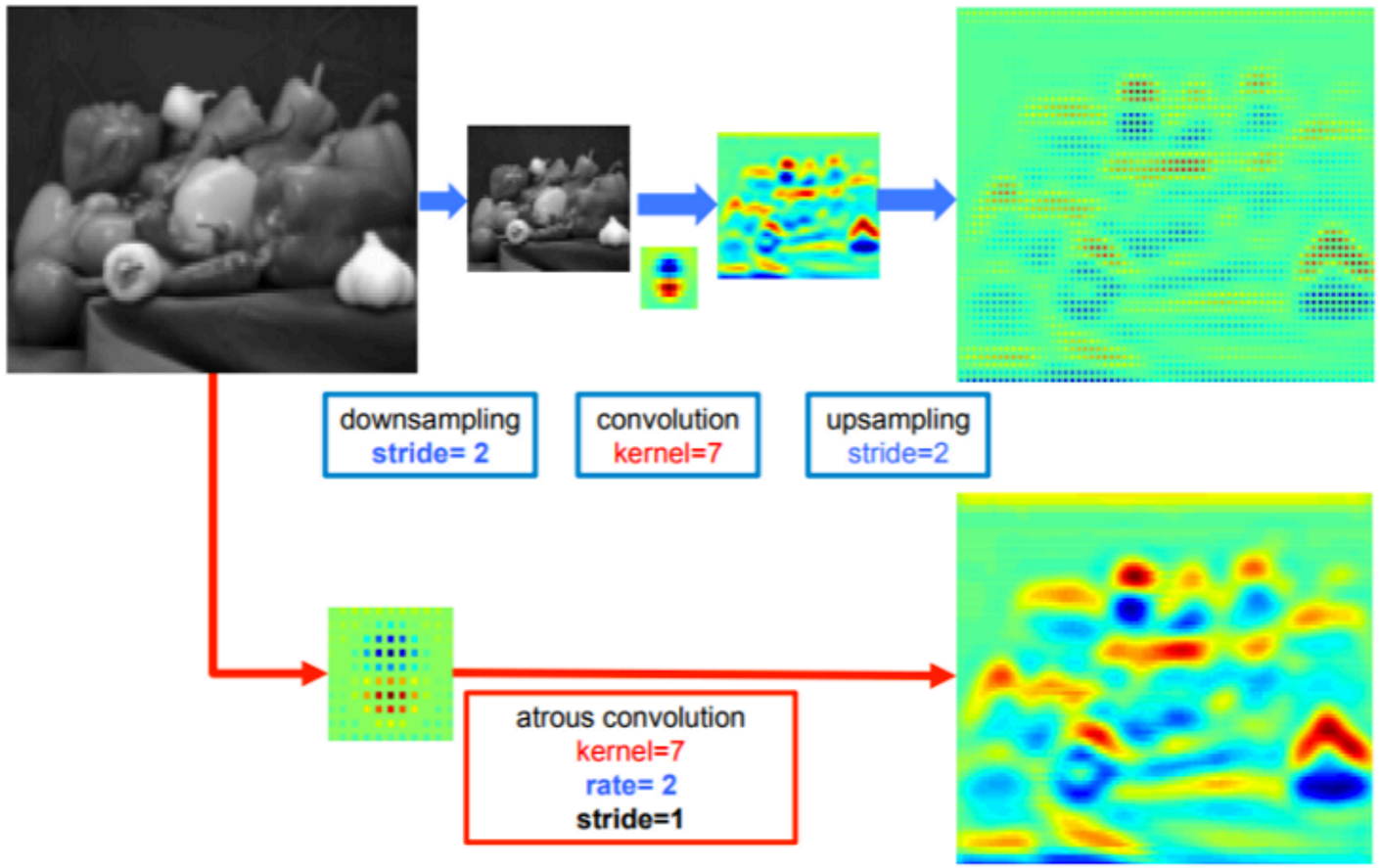
(a) Going deeper without atrous convolution.



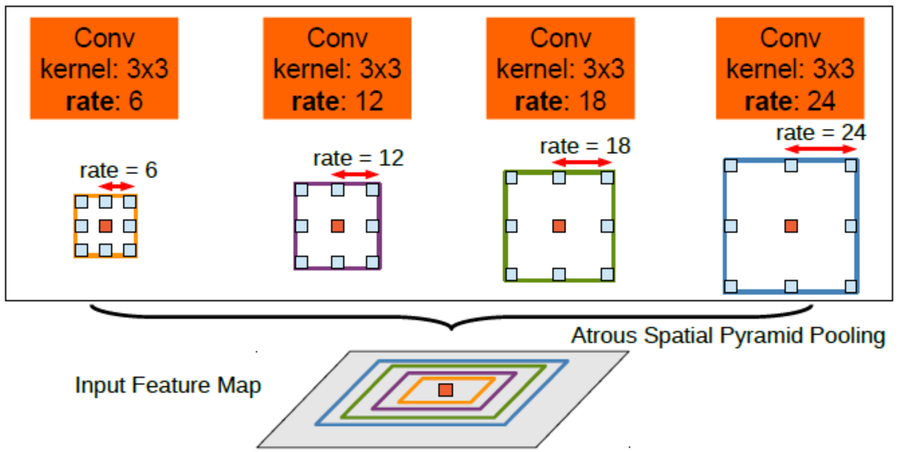
(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

Advantage over downsample:
Resolution is maintained,
but large-scale features
are still learned

Atrous vs Downsampling

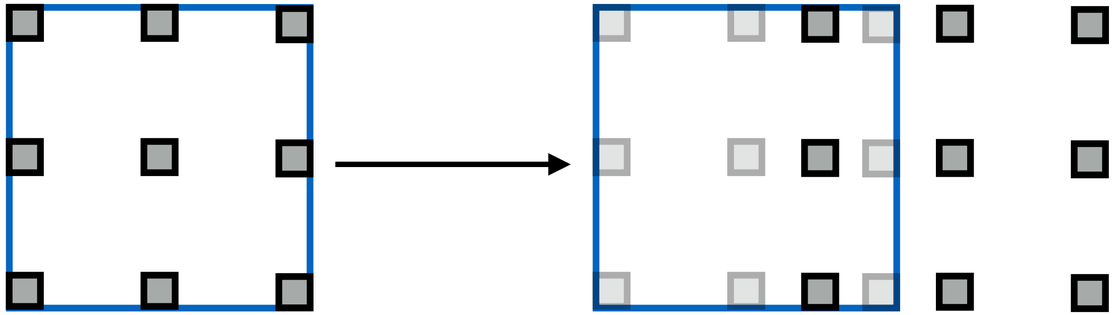


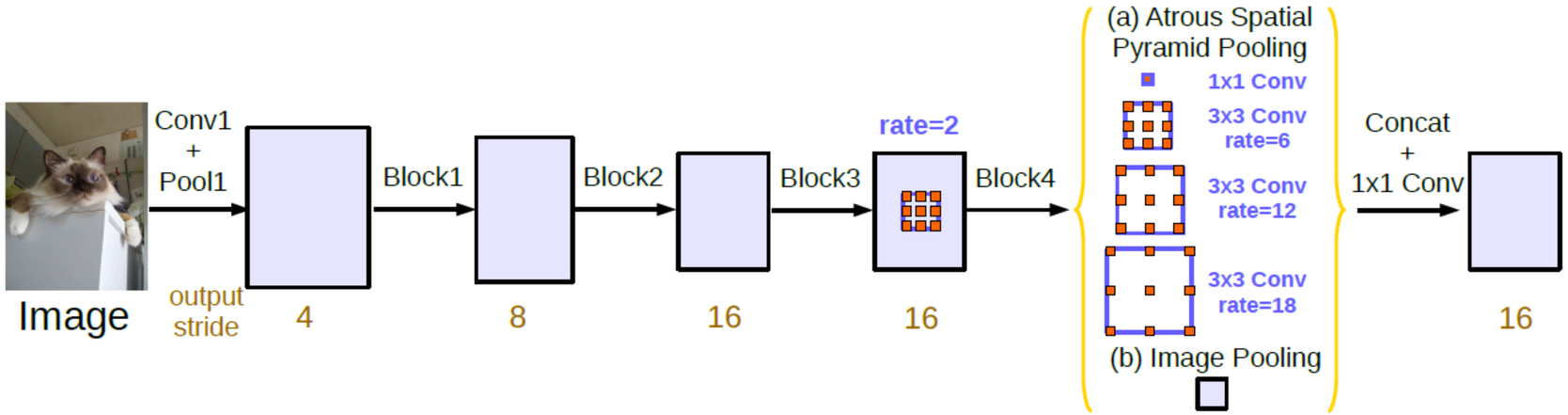
Atrous Spatial Pyramid Pooling



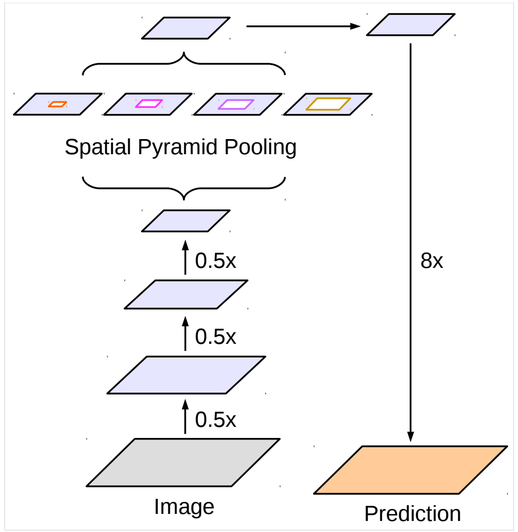
Different atrous rates concated, then pooled.

This minimizes sparse filter issues with large atrous-rate filters.



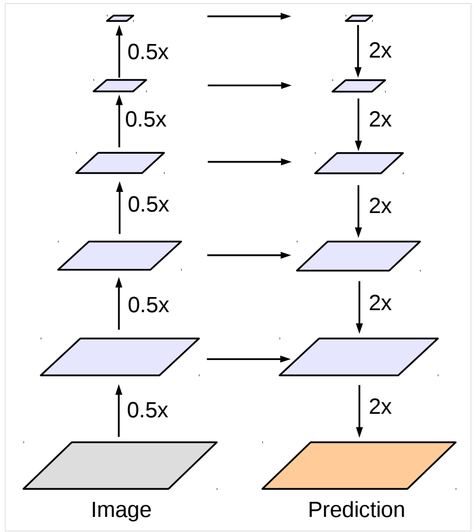


U-Net Encoder-Decoder



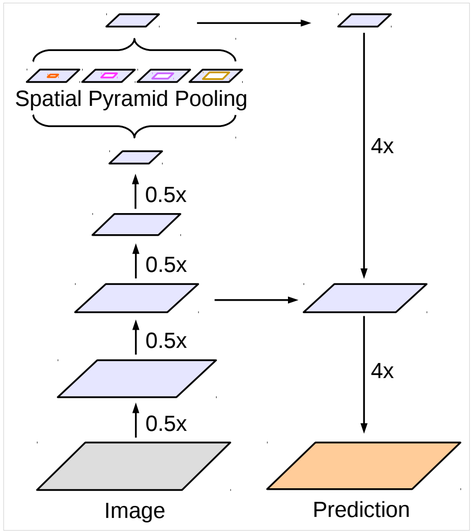
(a) Spatial Pyramid Pooling

DeepLabv3 (not +)



(b) Encoder-Decoder

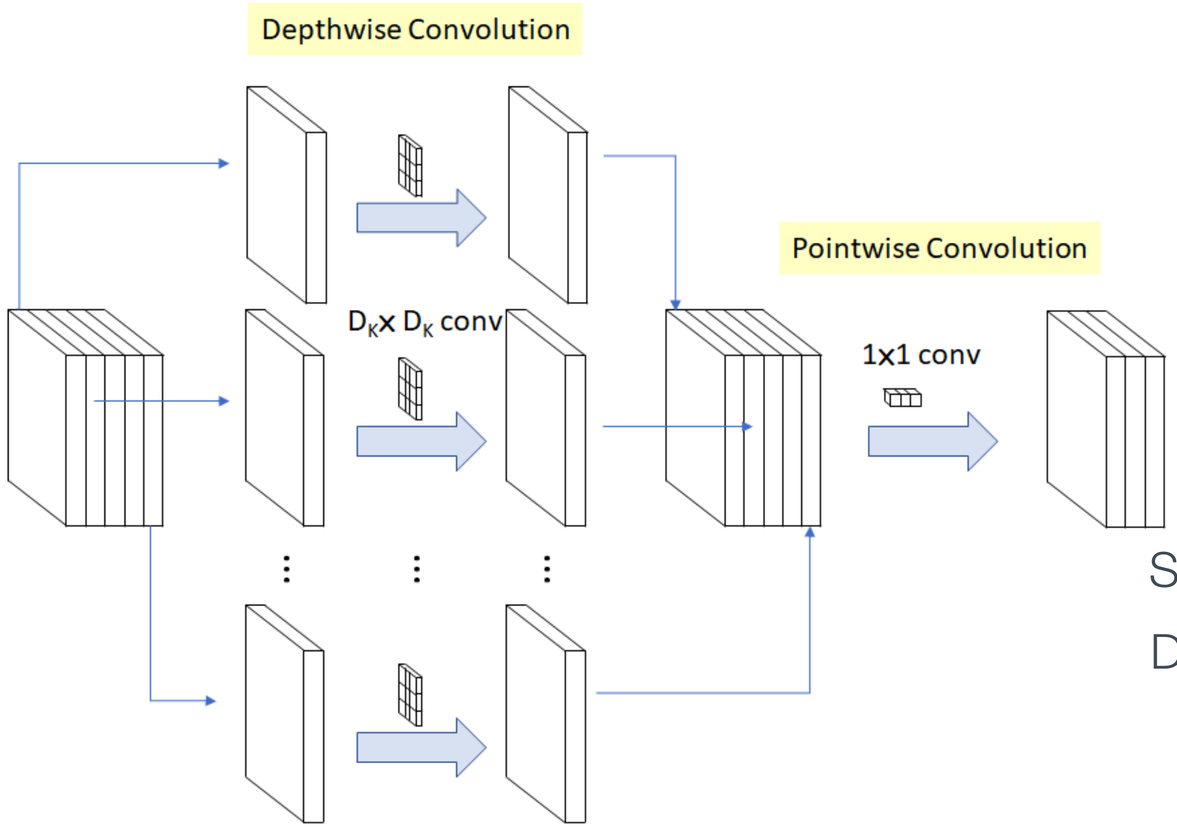
U-Net (and others)



(c) Encoder-Decoder with Atrous Conv

DeepLabv3+

Depthwise Separable Convolution



D_k: Kernel Dimension

D_f: Feature Dimension

M: Input Channels

N: Output Channels

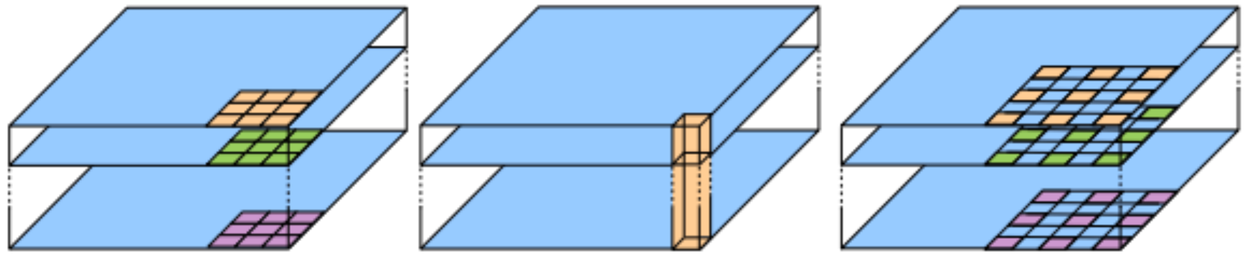
Standard Conv Cost: $D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F$

Depthwise Sep Cost: $D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$

Cost reduction: $$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2}$$

Performance is similar to standard conv in most cases (± ε)

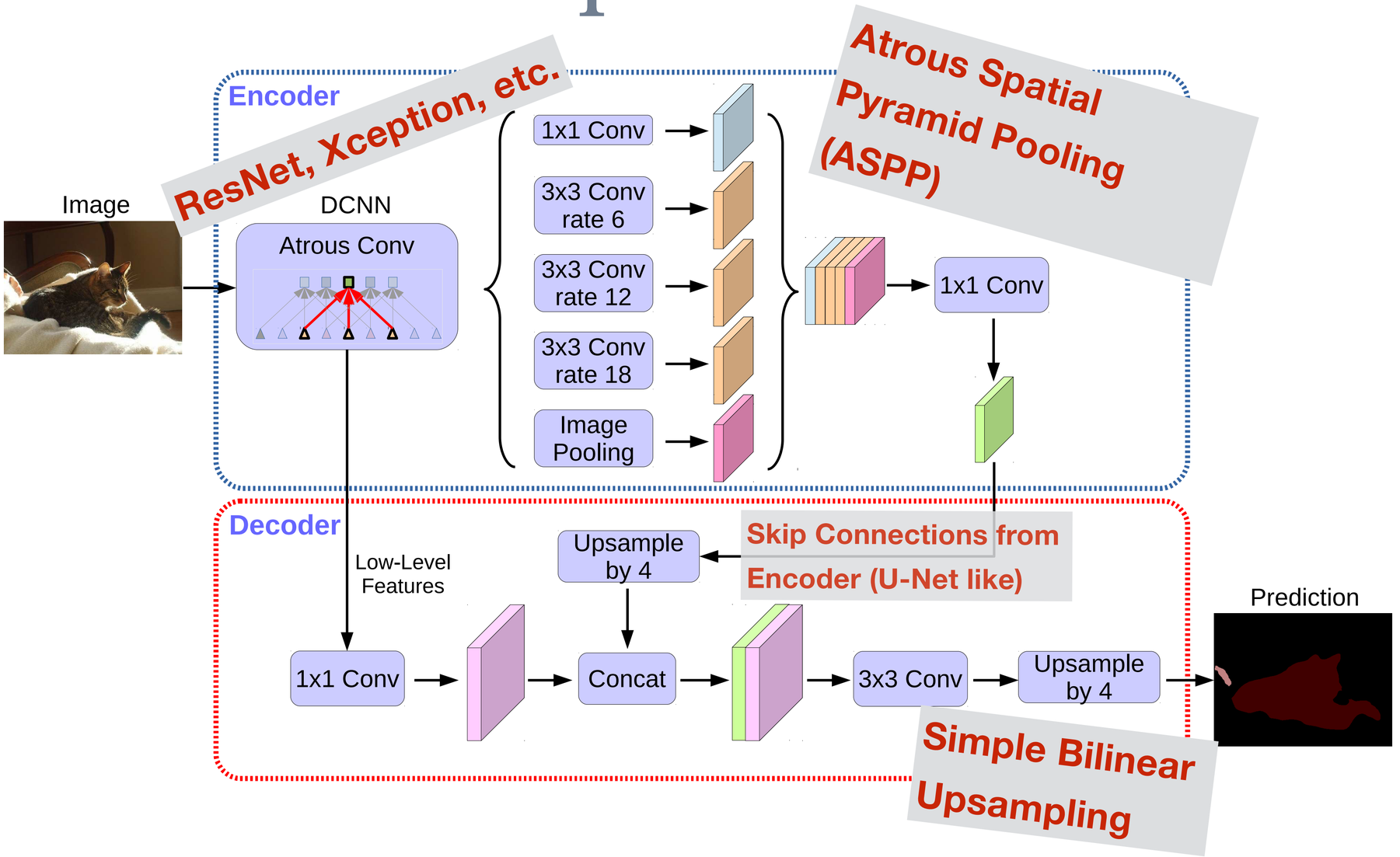
Atrous Separable Convolution



(a) Depthwise conv. (b) Pointwise conv. (c) Atrous depthwise conv.

Simple Idea, apply depthwise-separable to atrous conv.

DeepLabv3+

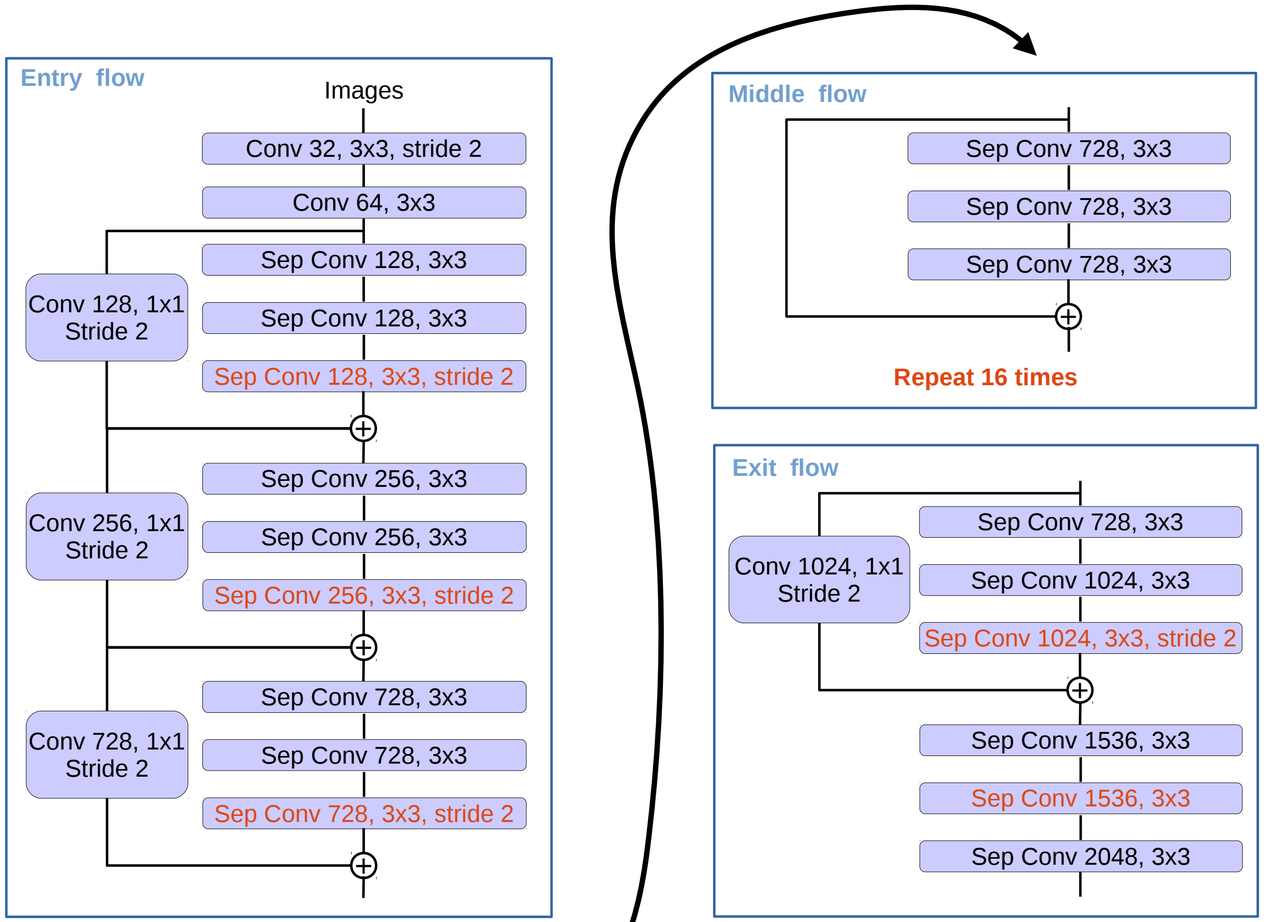

Encoder
ResNet, Xception, etc.
Atrous Spatial Pyramid Pooling (ASPP)
Image
DCNN
Atrous Conv
1x1 Conv
3x3 Conv rate 6
3x3 Conv rate 12
3x3 Conv rate 18
Image Pooling
1x1 Conv
Decoder
Low-Level Features
Upsample by 4
Skip Connections from Encoder (U-Net like)
1x1 Conv
Concat
3x3 Conv
Upsample by 4
Prediction
Simple Bilinear Upsampling

Backbone: Xception

-Deep ResNet-like Encoder

-Uses Depthwise Sep Conv

-Slightly modified (more layers, more BN, more Relu) for DeepLabv3+

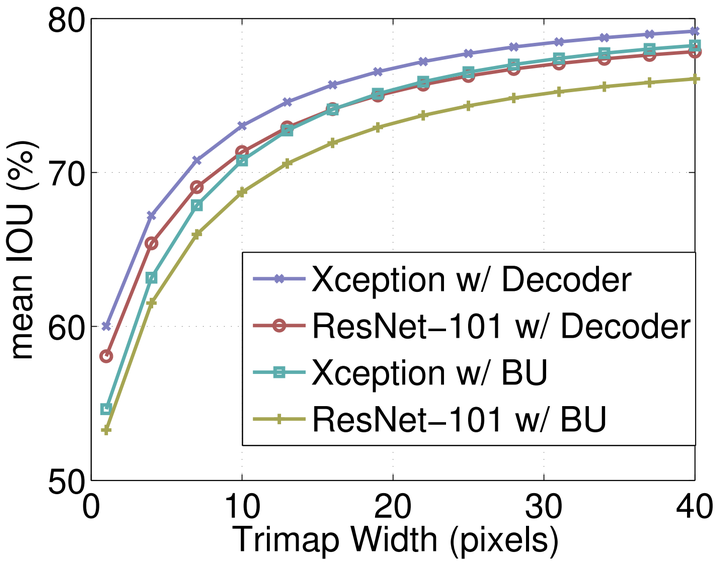


Skip connect to decoder

Performance

PASCAL VOC 2012 Dataset

Method	mIOU
Deep Layer Cascade (LC) [82]	82.7
TuSimple [77]	83.1
Large_Kernel_Matters [60]	83.6
Multipath-RefineNet [58]	84.2
ResNet-38_MS_COCO [83]	84.9
PSPNet [24]	85.4
IDW-CNN [84]	86.3
CASIA_IVA_SDN [63]	86.6
DIS [85]	86.8
DeepLabv3 [23]	85.7
DeepLabv3-JFT [23]	86.9
DeepLabv3+ (Xception)	87.8
DeepLabv3+ (Xception-JFT)	89.0



(a) mIOU *vs*. Trimap width

(b) Decoder effect

Trimap width: Side of object boundary

Depthwise Efficiency

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
16	16	✓			✓			79.79%	54.17B
16	16	✓	✓	✓	✓			81.21%	928.81B
16	8	✓			✓			80.02%	177.10B
16	8	✓	✓	✓	✓			81.39%	3055.35B
16	16	✓			✓	✓		82.20%	54.17B
16	16	✓	✓	✓	✓	✓		83.34%	928.81B
16	8	✓			✓	✓		82.45%	177.10B
16	8	✓	✓	✓	✓	✓		83.58%	3055.35B
16	16	✓			✓	✓	✓	83.03%	54.17B
16	16	✓	✓	✓	✓	✓	✓	84.22%	928.81B
16	8	✓			✓	✓	✓	83.39%	177.10B
16	8	✓	✓	✓	✓	✓	✓	84.56%	3055.35B

Table 5. Inference strategy on the PASCAL VOC 2012 *val* set when using modified *Xception*. **train OS**: The *output stride* used during training. **eval OS**: The *output stride* used during evaluation. **Decoder**: Employing the proposed decoder structure. **MS**: Multi-scale inputs during evaluation. **Flip**: Adding left-right flipped inputs. **SC**: Adopting depthwise separable convolution for both ASPP and decoder modules. **COCO**: Models pretrained on MS-COCO. **JFT**: Models pretrained on JFT.

Adding Depthwise Separable Conv greatly reduces Multiply-Adds while maintaining (slightly reducing) overall predictive performance.

Application

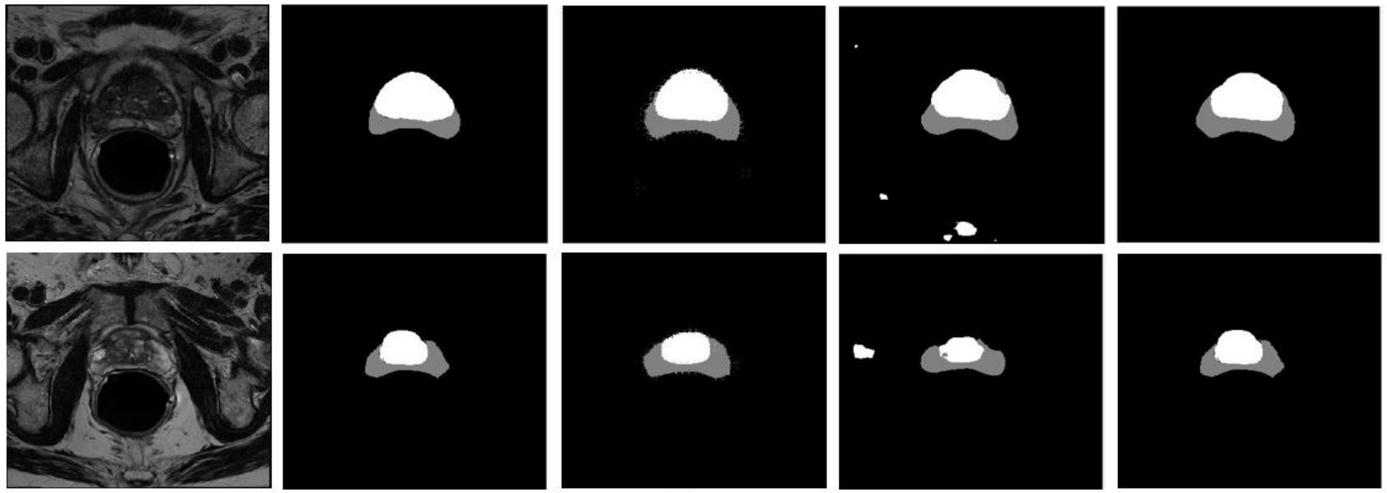


TABLE I
PERCENTAGE OF AVERAGE DICE SIMILARITY COEFFICIENT (DSC) FOR ZONAL SEGMENTATION, PERIPHERAL ZONE (PZ) AND CENTRAL ZONE (CG) OF PROSTATE MRI BY FCN [14], PSNET [16], AND DEEPLABV3.

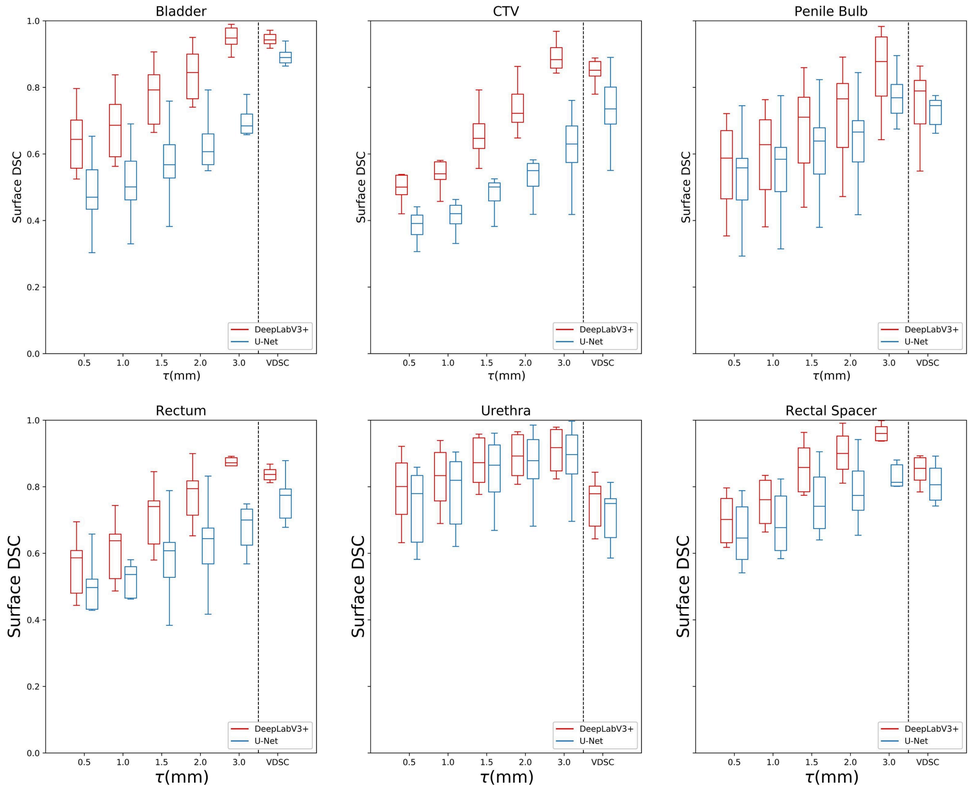
Model	Zone	3T	1.5T	1.5T+3T
FCN	PZ	63.5	63.2	67.1
	CG	71.3	70.2	75.5
PSNet	PZ	64.5	65.1	65.3
	CG	81.5	82.2	84.1
DeepLabV3+	PZ	**70.2**	**67.5**	**73.2**
	CG	**88.4**	**86.5**	**89.2**

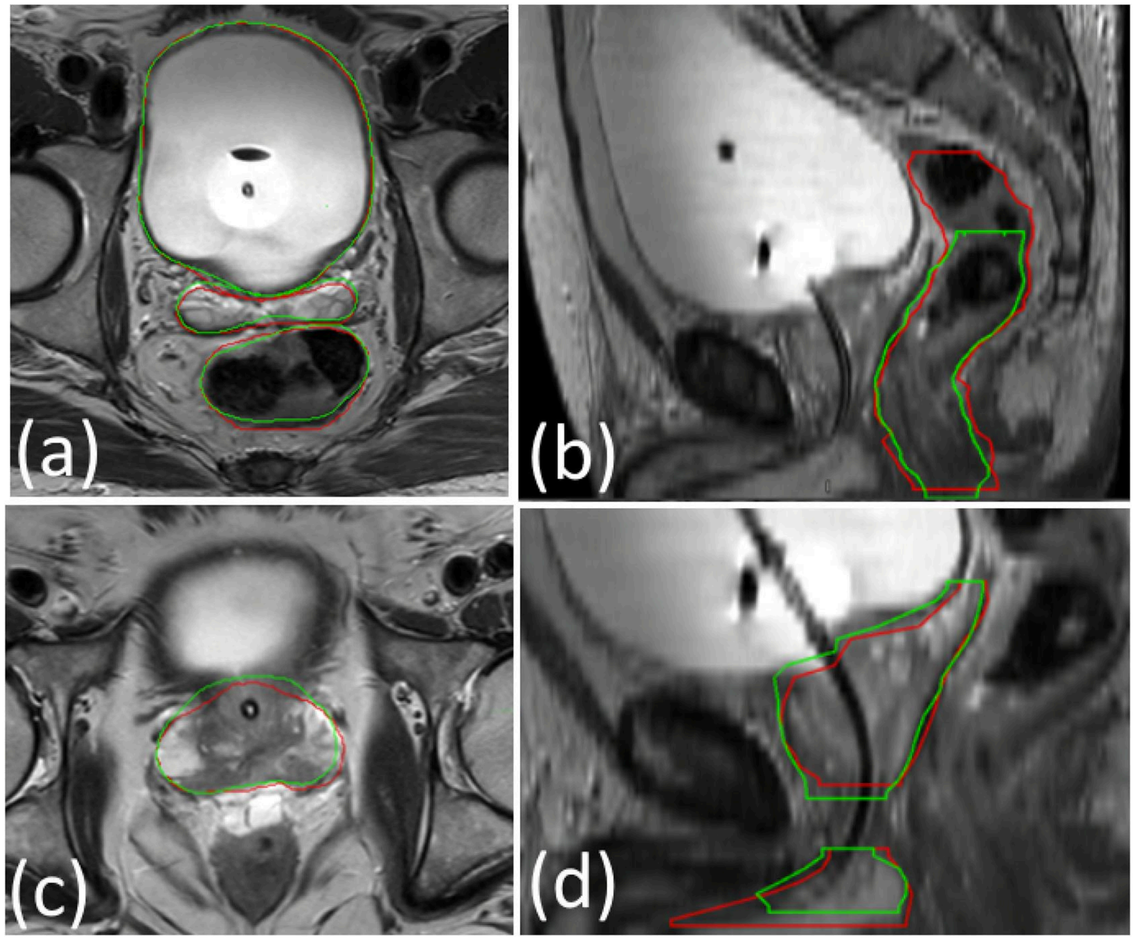
Zonal Segmentation of Prostate T2W-MRI using Atrous Convolutional Neural Network
2019 IEEE Student Conference on Research and Development (SCOReD)

Application

DeepLabV3+

U-Net





Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy.
Physics and Imaging in Radiation Oncology