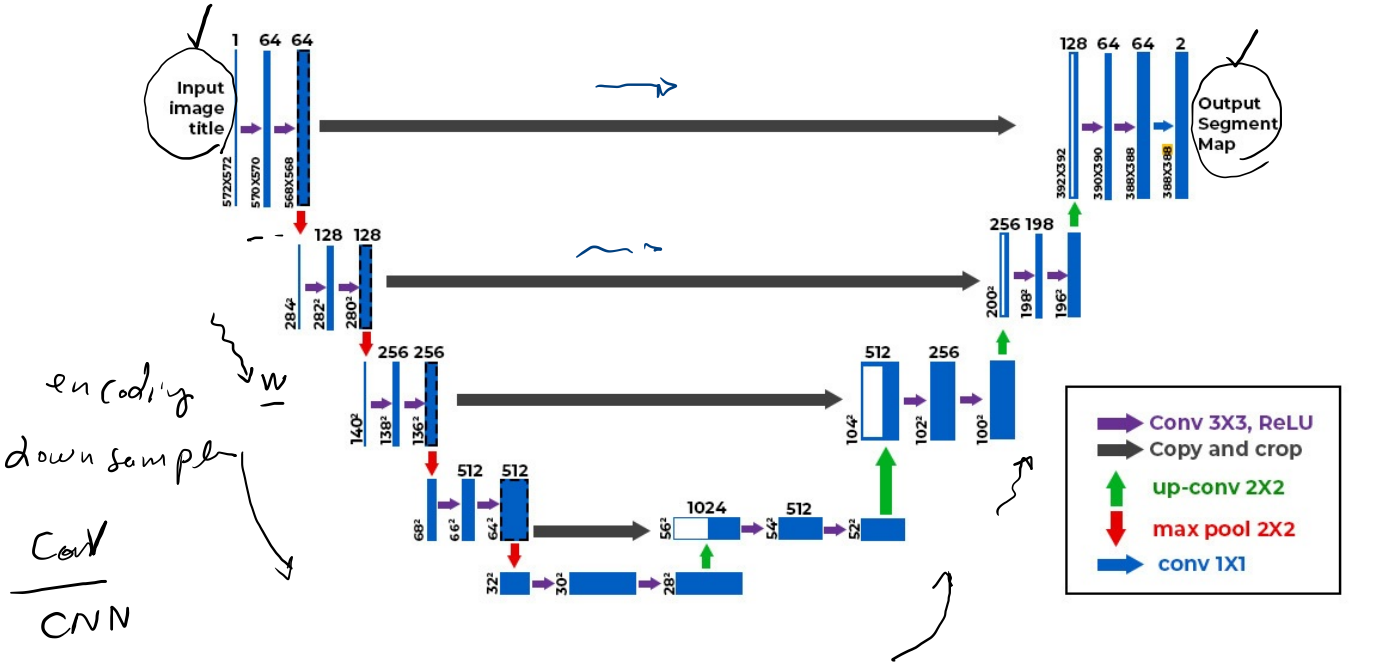
DeepLab v3



accuracy ↑

speed ↓

encoding

down sample

Conv

CNN

upsample → TConv

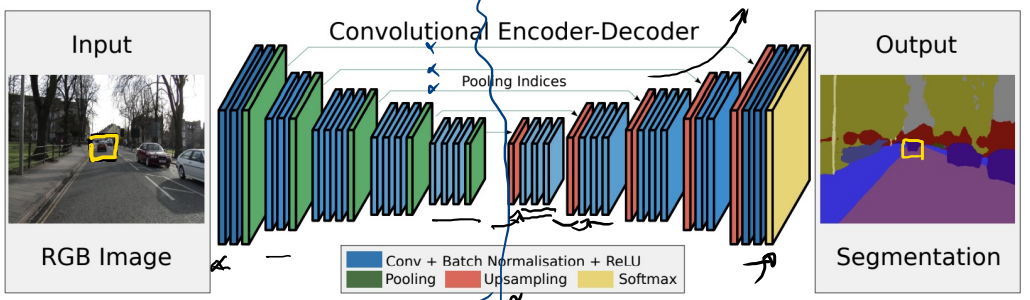


Fig. 2. An illustration of the SegNet architecture. There are no fully connected layers and hence it is only convolutional. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map(s). It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification.

پس از اینا بود یه چیزی که هم دقت بالا

در segment لبه‌های کوچیک داشته باشیم

هم سرعت بالایی داشته باشیم!

Deeplab V3

Deeplab ⇝ Feature extraction → losing info.

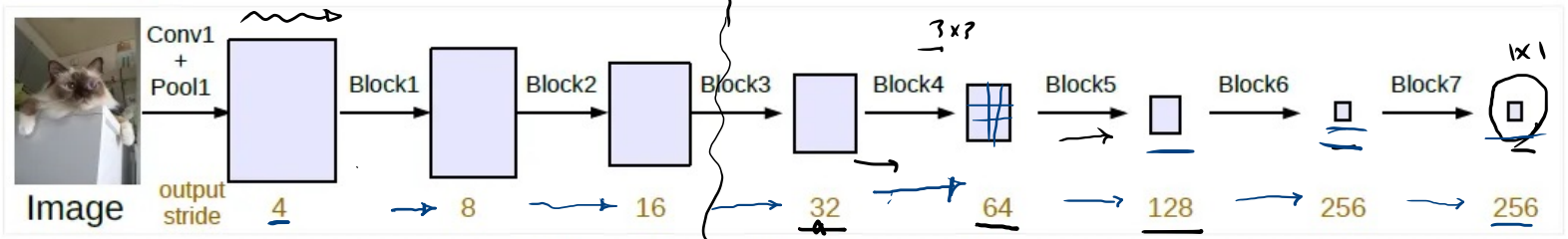
→ رزولوشن کم!

[illegible]!

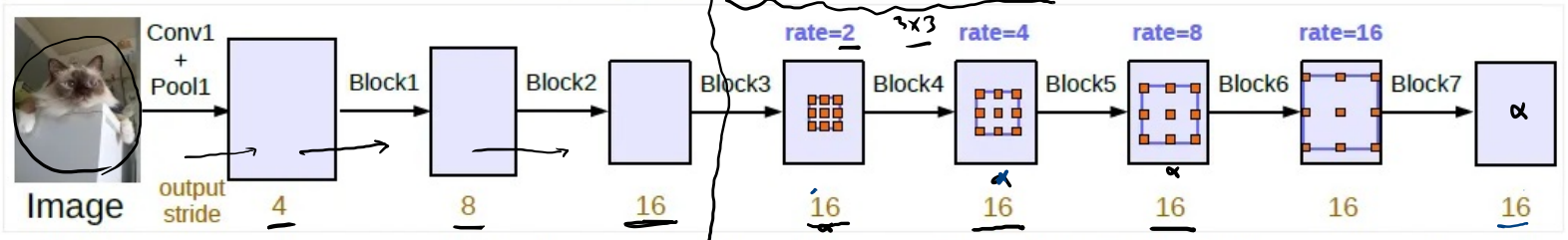
↳ { accuracy!

small objects, too

Atrous Conv.!

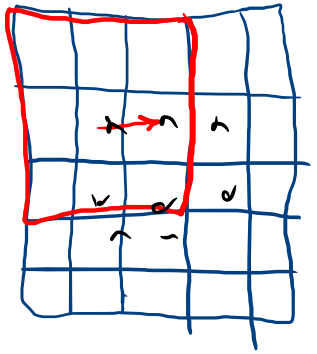


(a) Going deeper without atrous convolution.



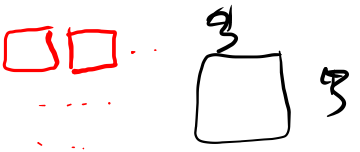
(b) Going deeper with atrous convolution. Atrous convolution with $rate > 1$ is applied after block3 when $output\_stride = 16$.

Traditional Conv

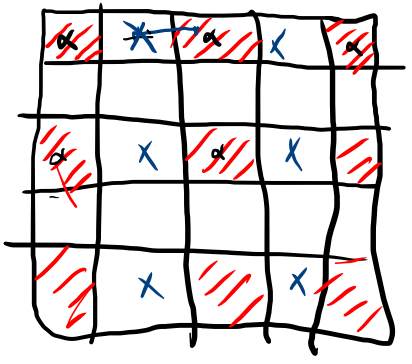


Conv

X

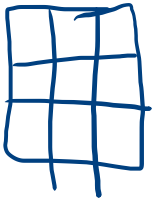


Atrous Conv

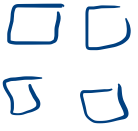


Atrous Ratio = 2

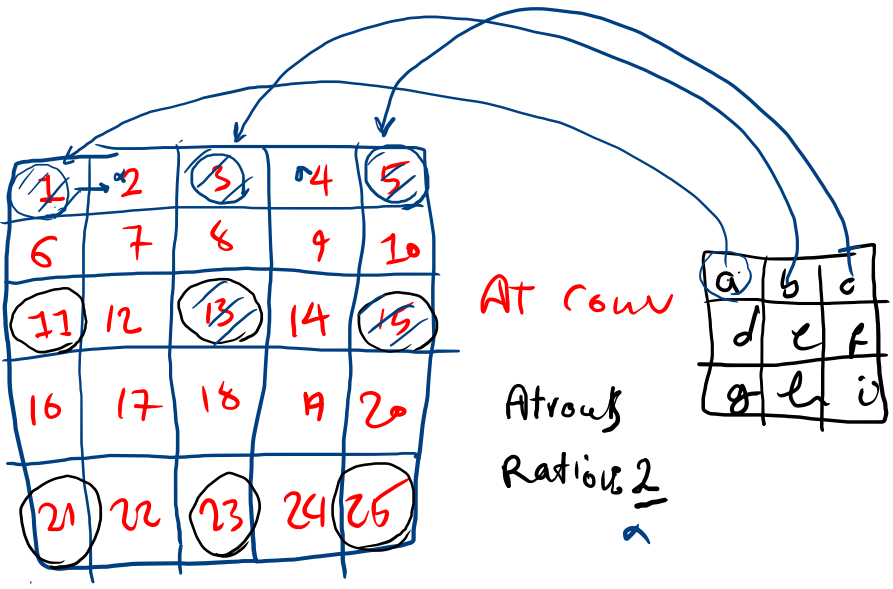
At Conv



X



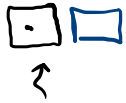
Atrous Conv



At conv

Atrous Ratio: 2

= 3x3

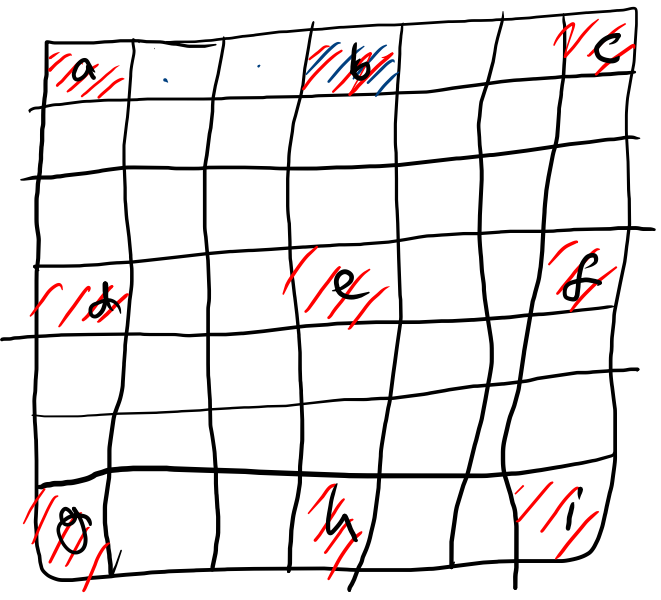


$a(1) + b(3) + c(5)$

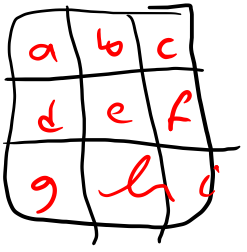
$+ d(11) + e(13) + f(15)$

$+ g(21) + h(23) + i(25) =$

atrous Rate = 3



At
Conv



input feature map { H = 5, W = 5

Atrous kernel { kh = 3, kw = 3

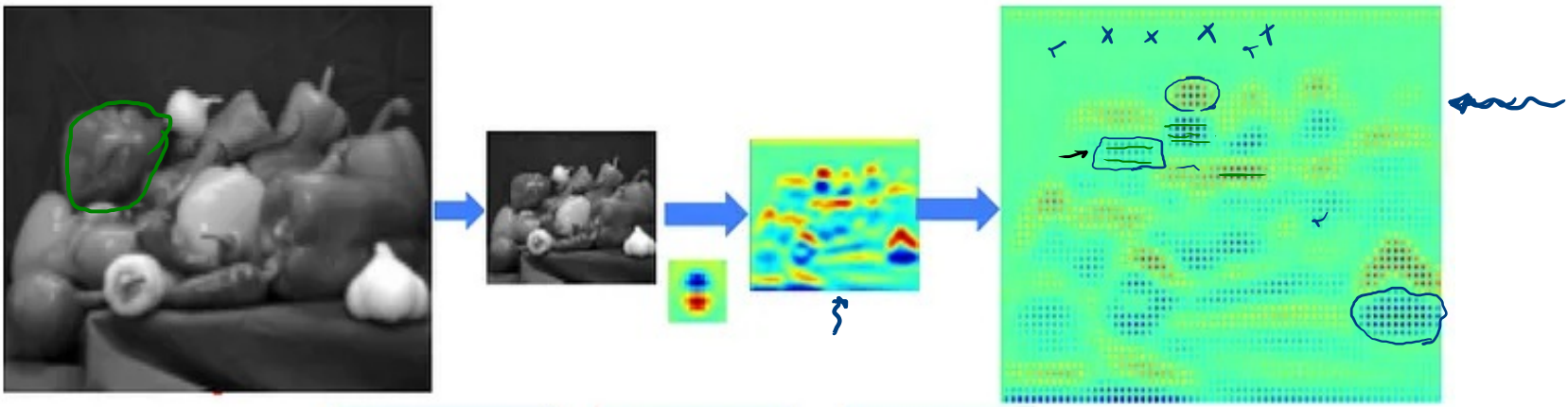
(Atrous) dilation rate { Ar = 2

Height = (H - 1) * Ar + kh

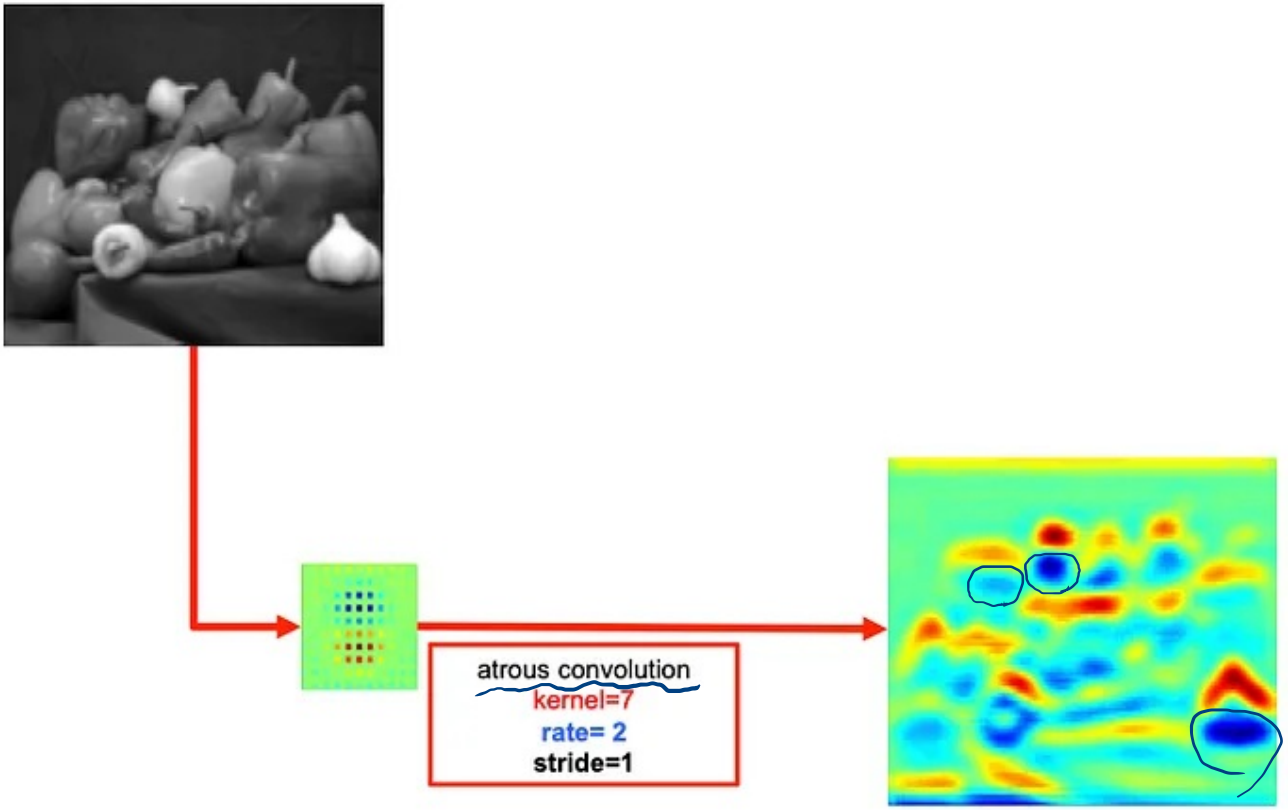
width = (W - 1) * Ar + kw

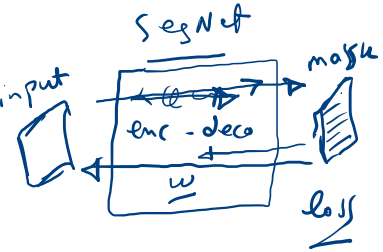
H = (5 - 1)(2) + 3 = 11

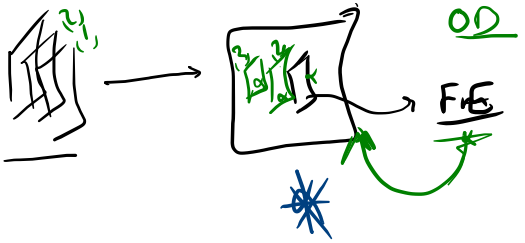
W = (5 - 1)(2) + 3 = 11


downsampling
stride= 2
convolution
kernel=7
upsampling
stride=2

Atrous conv





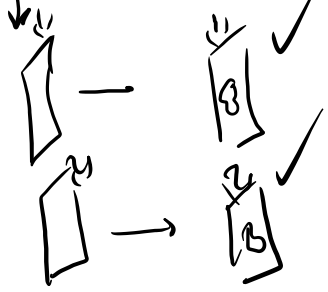


VideoCap —

object

Tracking

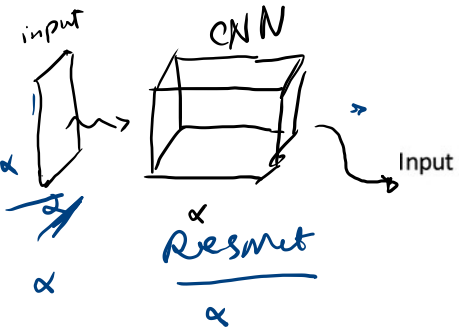
segmentation

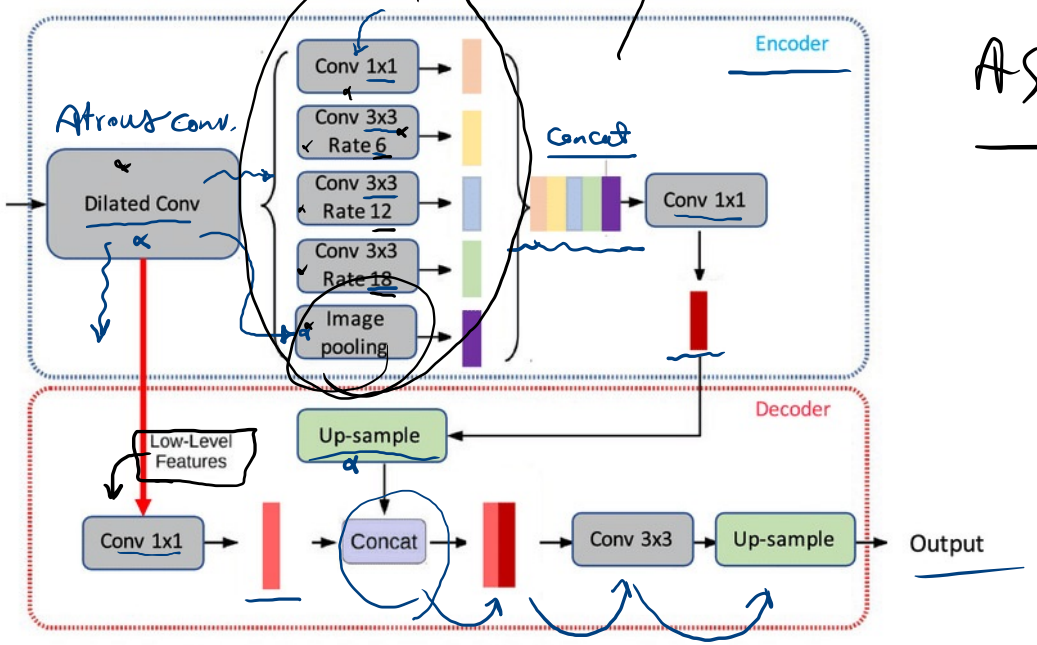


Deeplab Architecture.

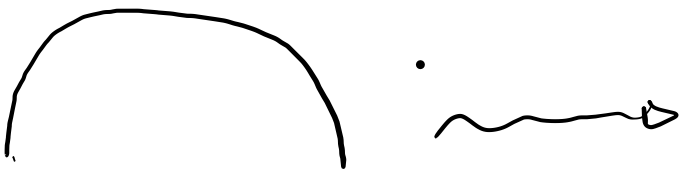
Atrous spatial pyramid pooling.

ASPP

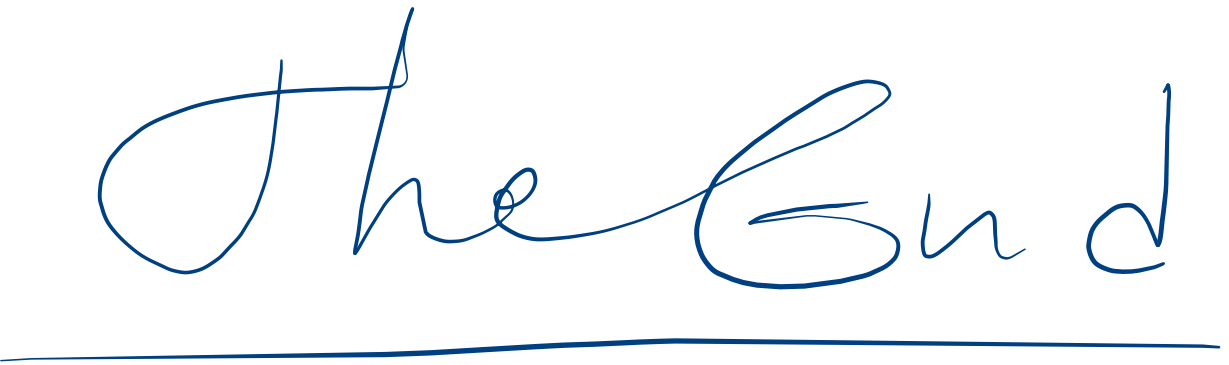




low level



high level


The End