Image Segmentation with Neural Networks
Computer Vision and Machine Learning Track

Gabriel Leivas Oliveira, Tonmoy Saikia

13, November 2018

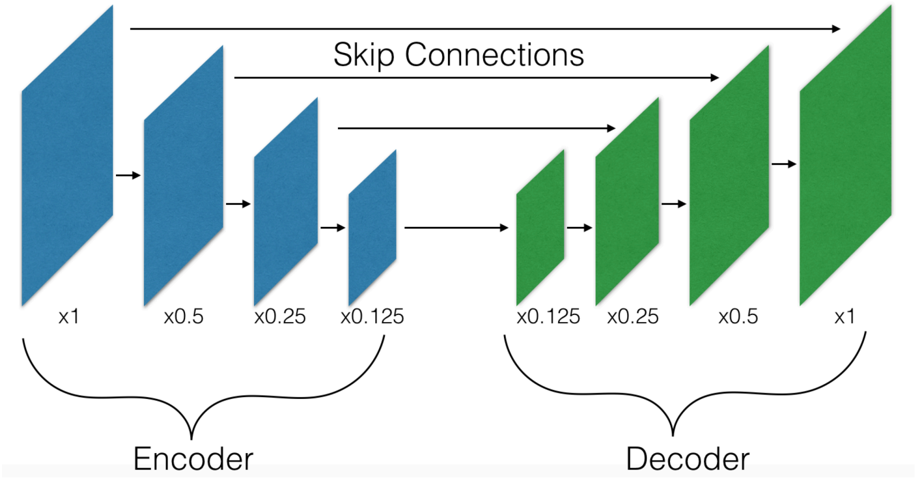
What is semantic segmentation?

Input image

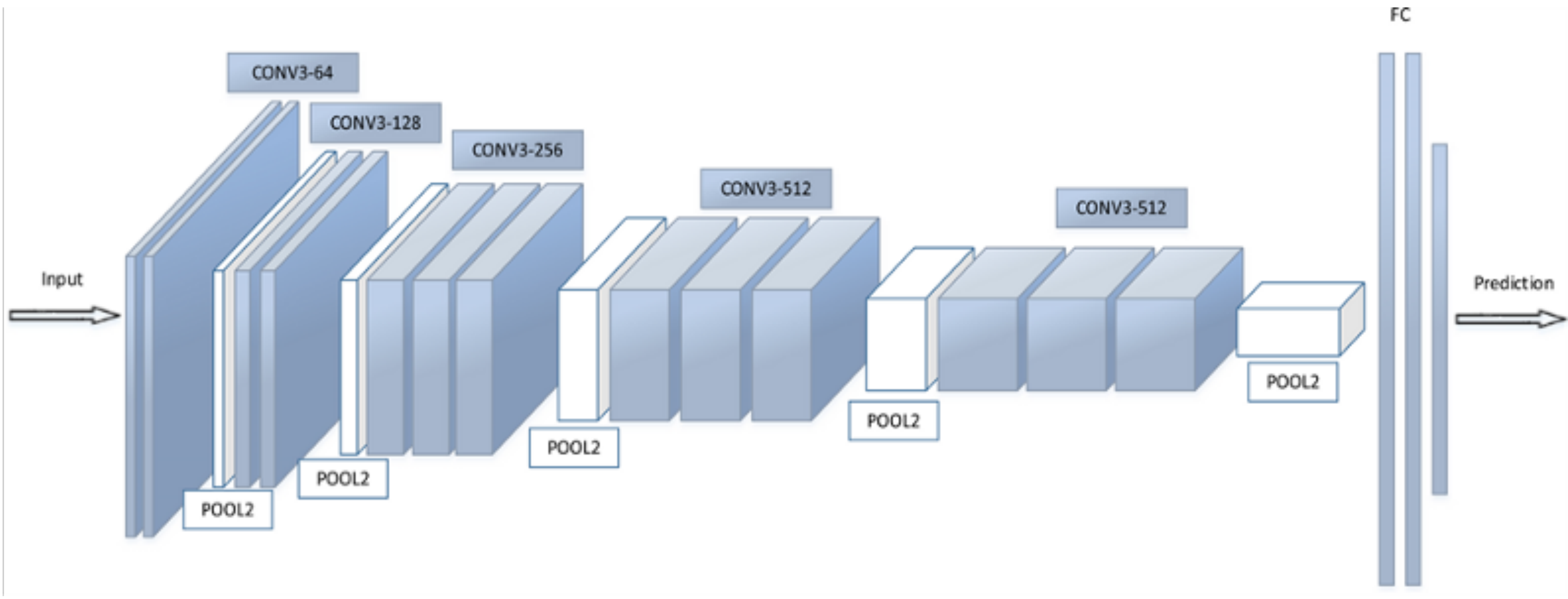
Segmentation mask

Encoder-Decoder Architectures

- Encoders:
  - Takes an input image and generates a high dimensionality feature vector
  - Aggregate features at multiple levels
- Decoders:
  - Takes a high dimensionality feature vector and generates semantic masks
  - Upsample features aggregated by encoders at multiple levels

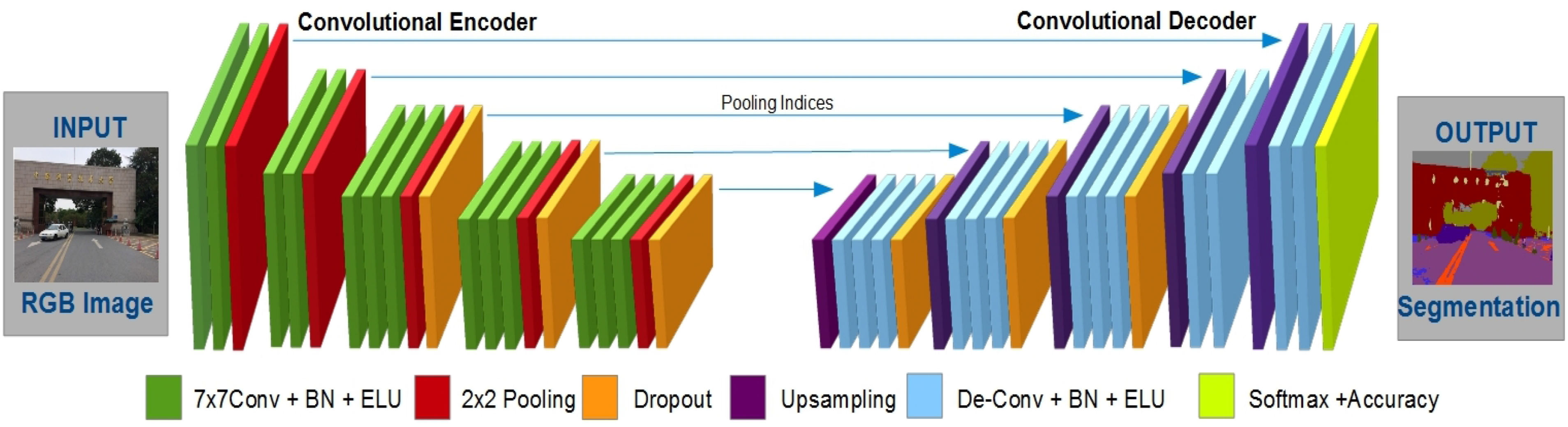


Encoder Networks



[Simonyan et al. 2014]

Encoder-decoder networks

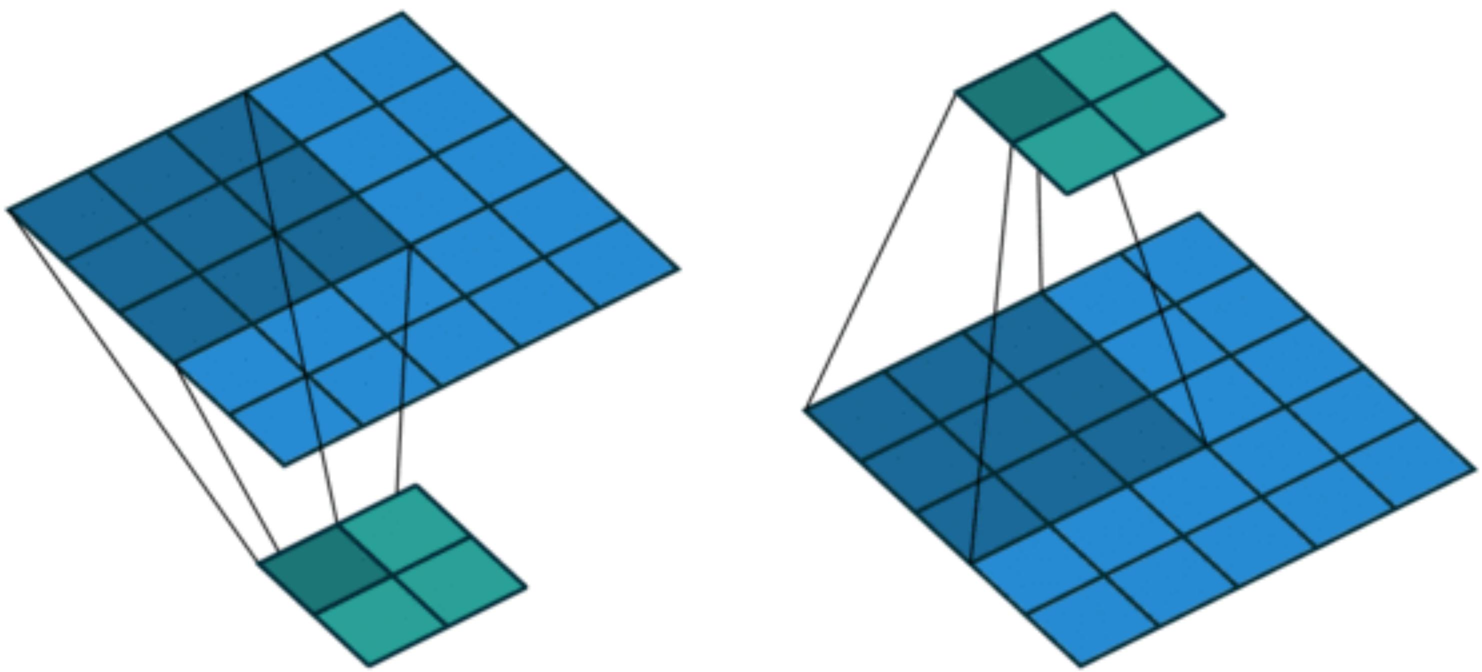


[Yasrab et al. 2017]

- Upsampling through transposed convolutions
- Refinement stages

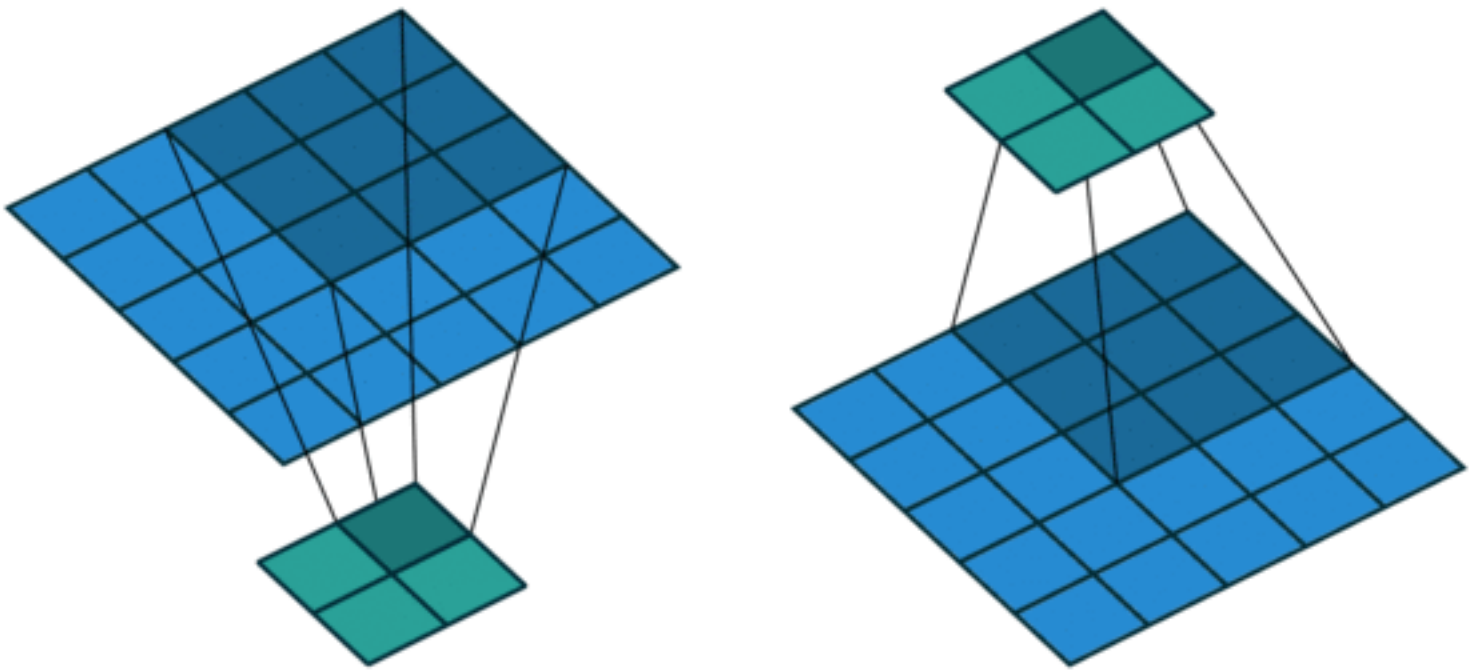
Transposed Convolutions

- Transposed convolutions are used to upsample the features.
- Example with lter kernel size [3, 3] and stride 2:



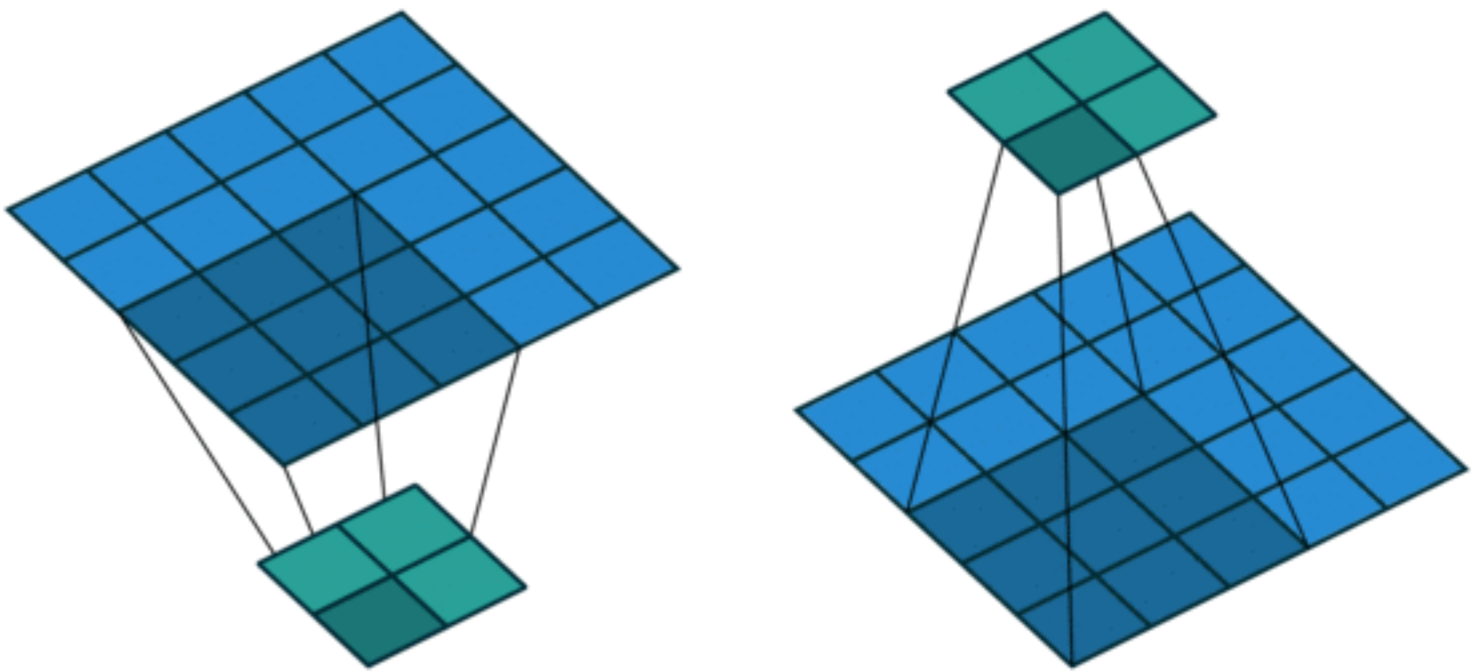
Transposed Convolutions

- Transposed convolutions are used to upsample the features.
- Example with lter kernel size [3, 3] and stride 2:



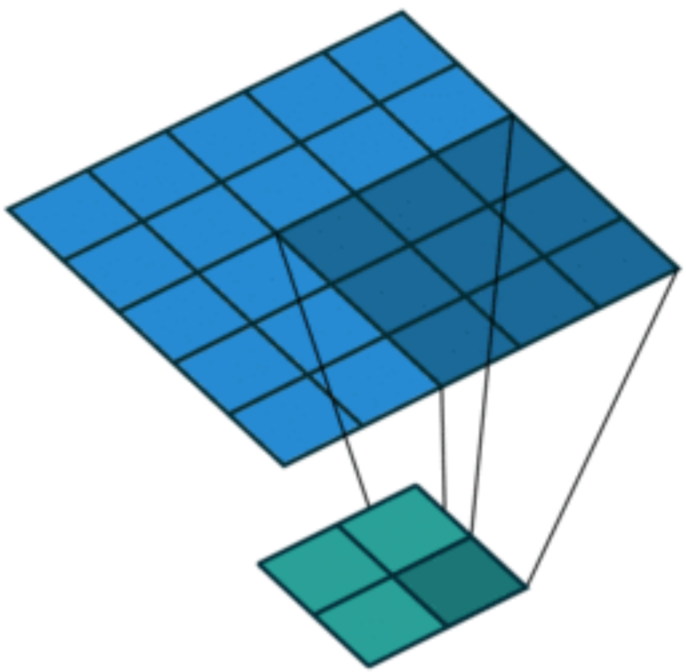
Transposed Convolutions

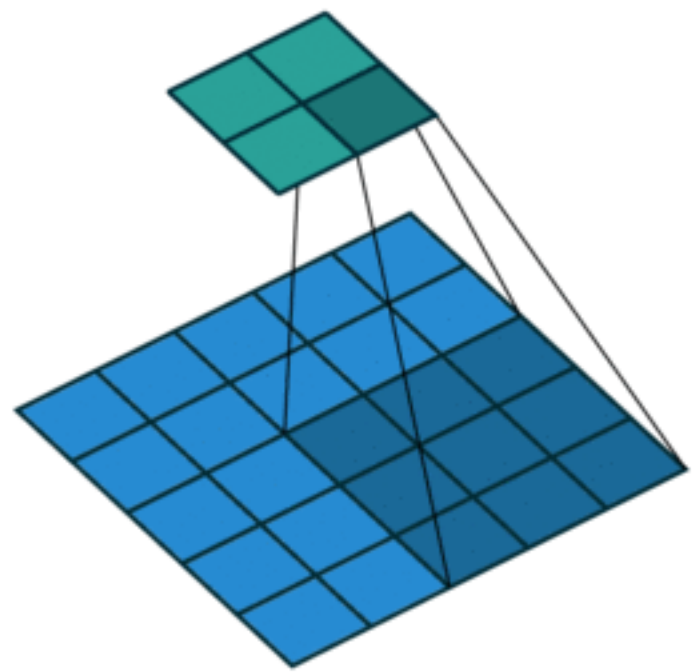
- Transposed convolutions are used to upsample the features.
- Example with lter kernel size [3, 3] and stride 2:



Transposed Convolutions

- Transposed convolutions are used to upsample the features.
- Example with lter kernel size [3, 3] and stride 2:





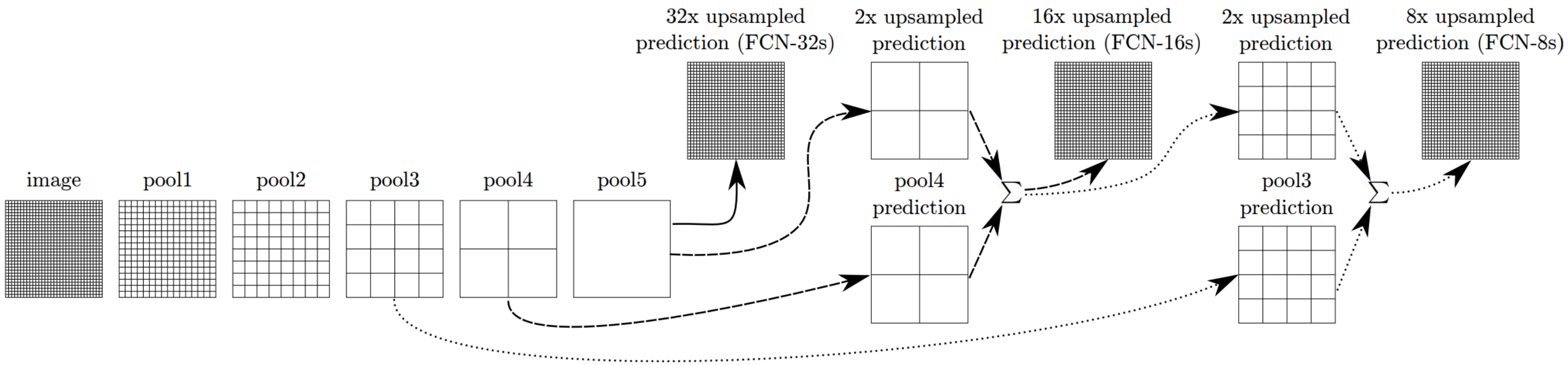
Transposed Convolutions

```
slim.conv2d_transpose(
    inputs ,
    filters ,
    kernel_size ,
    stride,
    scope =’layer_name'
)
```

- inputs: input tensor
- filters: amount of output features
- kernel_size: size of the kernel in each dimension
- stride: upsampling rate
- scope: name/id of the layer

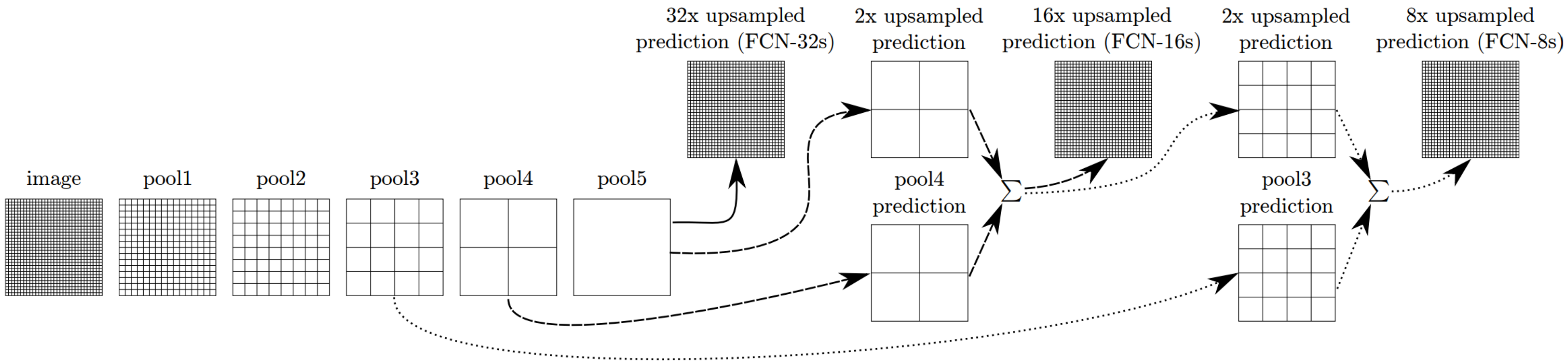
Refinement Stages

- Slowly upsampling or stage upsampling adds 'skip connections' from encoder layers to the decoder by fusing feautures which pass through less downsampling operations.



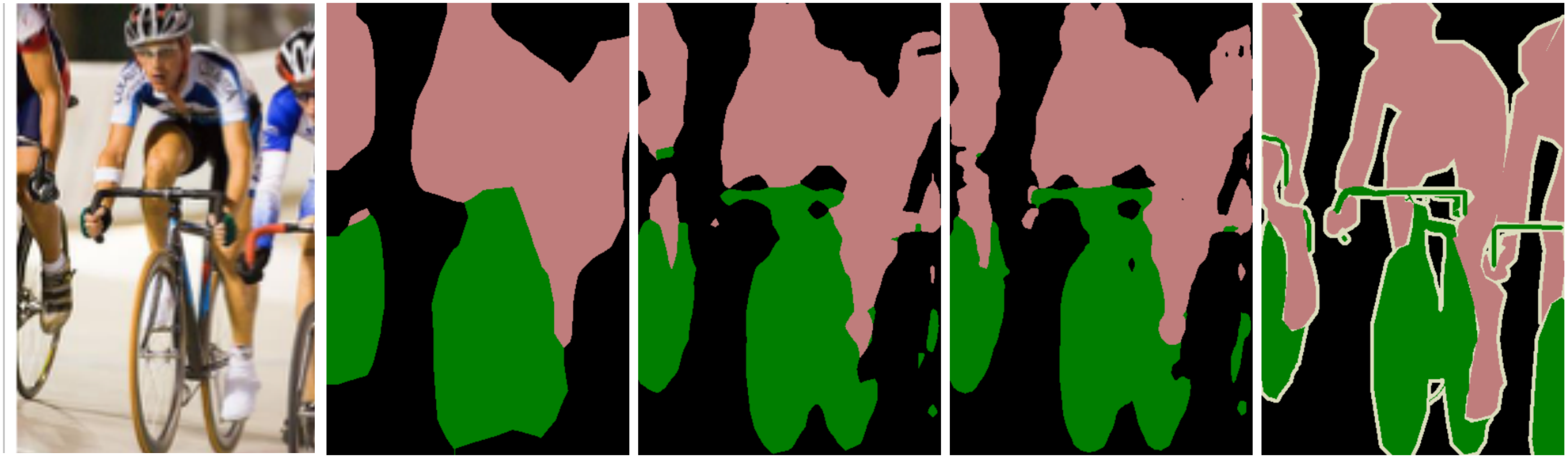
Refinement Stages

- Skip connections provide the necessary details in order to reconstruct accurate shapes for segmentation boundaries. Fine-grain segmentation masks are obtained with a multi-stage upsampling approach with skip connections



Refinement Stages

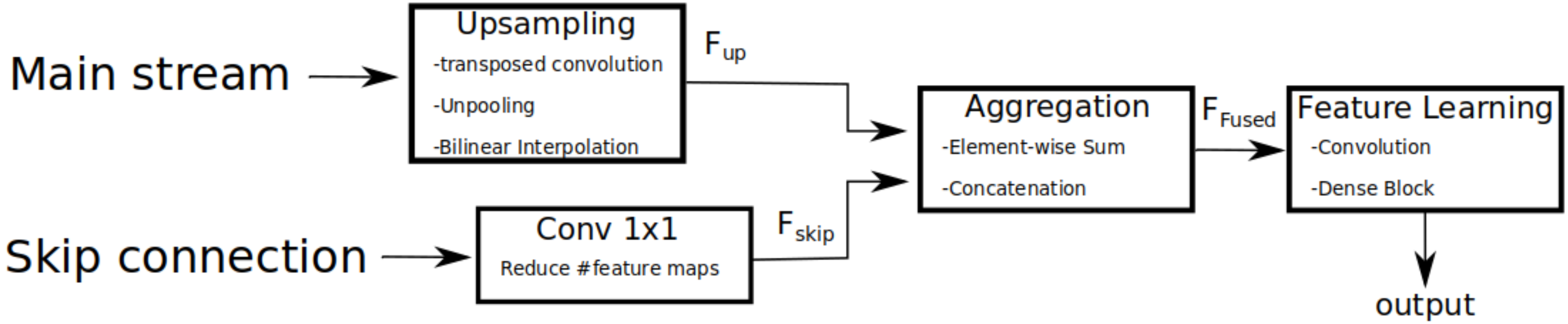
Stride 32 Stride 16 Stride 8



No Skip connection one Skip Two skip

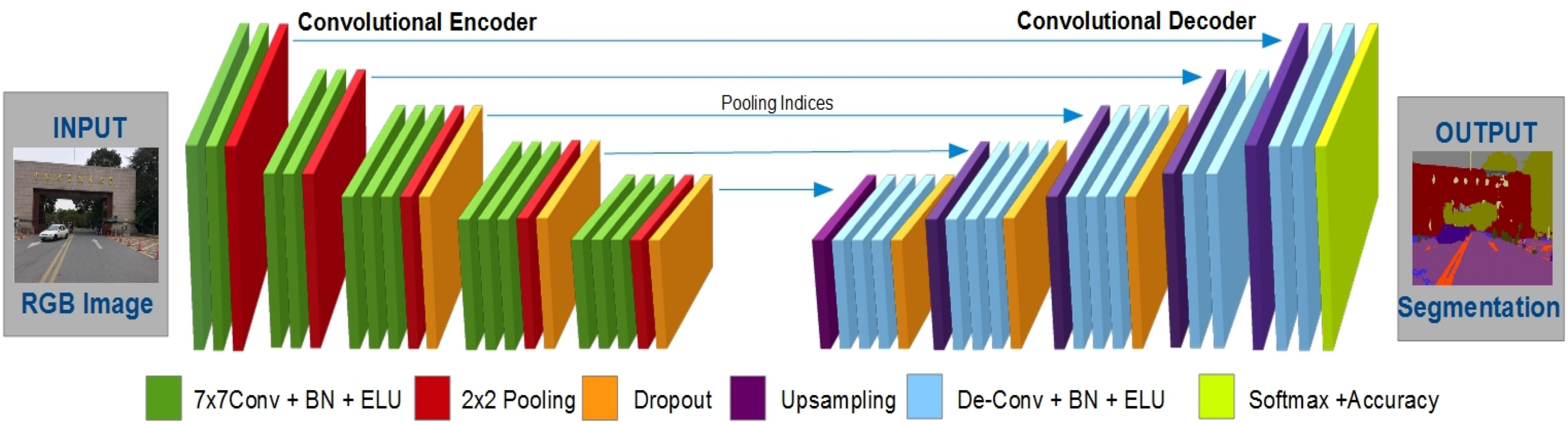
[Long et al. 2015]

Refinement block



Efficiency aspect of FCNs

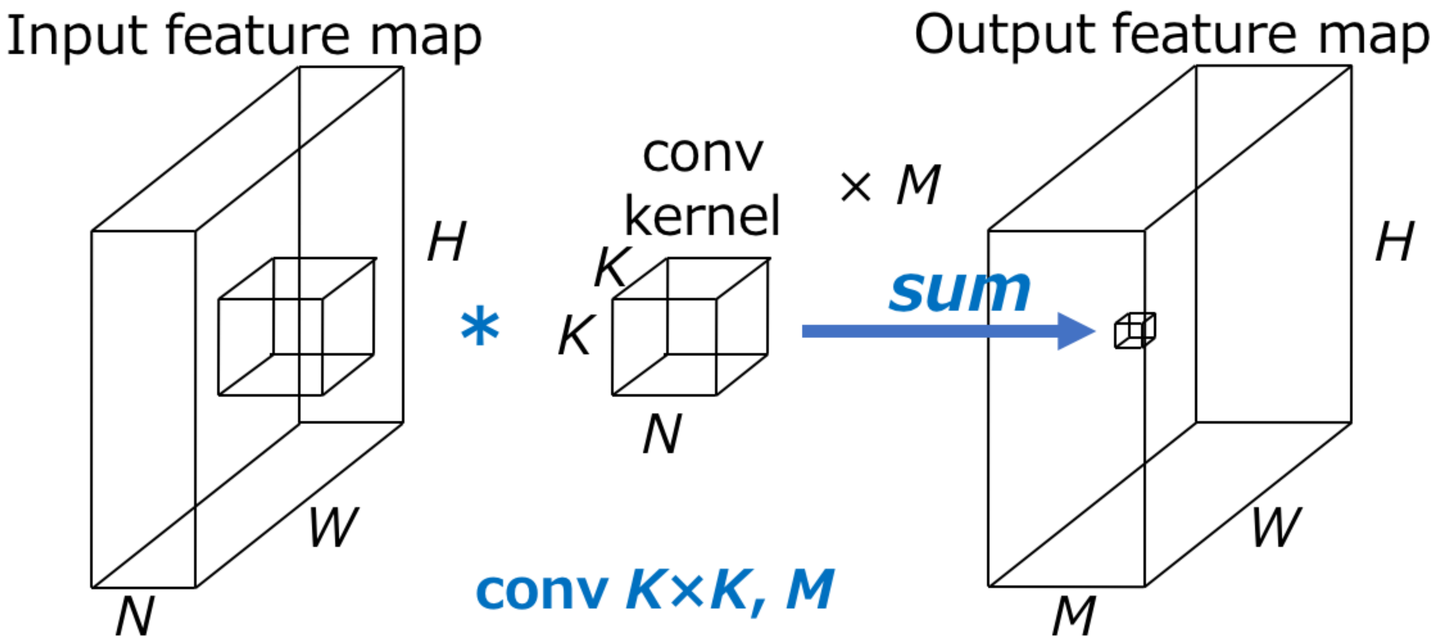
Networks parameter balancing



[Yasrab et al. 2017]

- Higher resolution layers will hold most of the computational requirements

Convolution Factorization

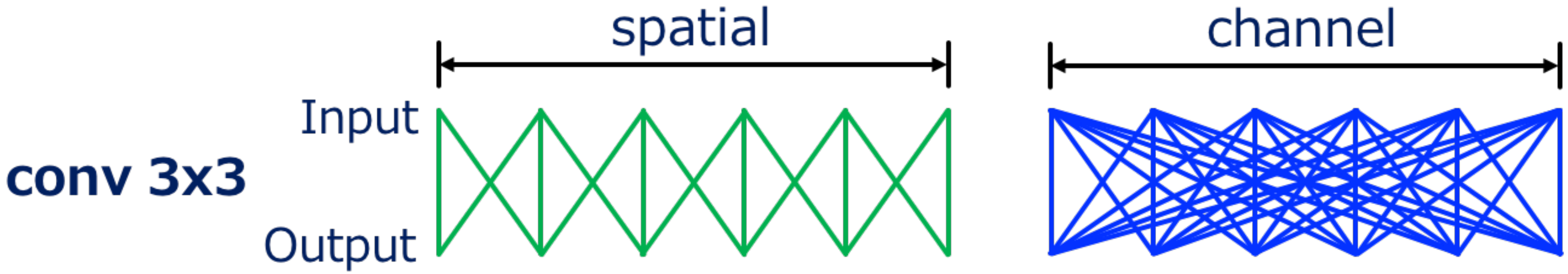


$HWNK^2M$ operations

Where N is the number of input channels and M the number of output channels

Spatial and Channel Domain

- Spatial have a neighborhood pattern
- Channel is fully connected



[Uchida, 2018]

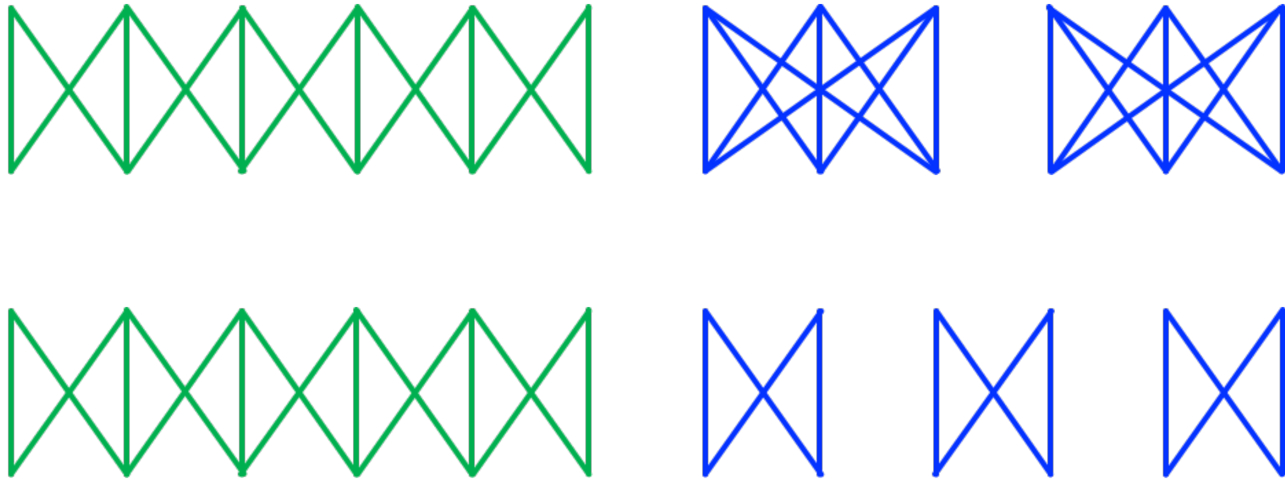
Group Convolution

- Input features are grouped and convolution is performed indepedently for each group

$KWNK^2M/G$

gconv 3x3 (g=2)

gconv 3x3 (g=3)

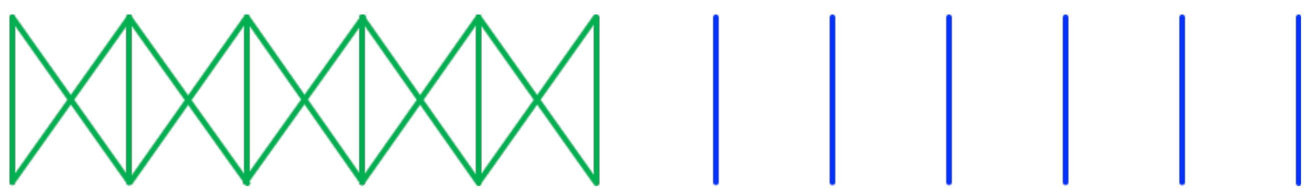


Depthwise Convolution

- Convolutions are computed for each input channel (M=N=G)

$KWNK^2$

depthwise conv



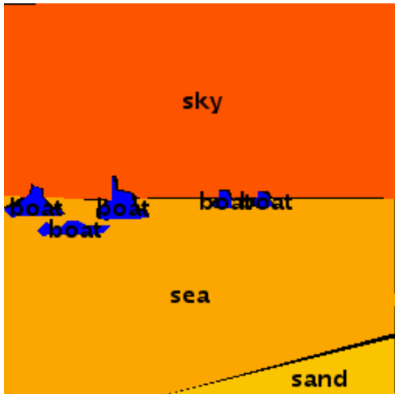
Metrics for segmentation

Overall pixel accuracy (OP)

- Accuracy measures the proportion of correctly labelled pixels
- One significant limitation of this measure is its bias in the presence of very imbalanced classes.

$$OP = \frac{\sum_{i=1}^{L} \mathbf{C}_{ii}}{\sum_{i=1}^{L} \mathbf{G}_{i}},$$

Class Balancing Problem



- Median Frequency

$$\alpha c = median\_freq/freq(c)$$

- Focal Loss

$$F_L(p_t) = -\alpha_t \left(1 - p_t\right)^{\gamma} \log\left(p_t\right)$$

Per class accuracy (PC)

- Per class accuracy measures the proportion of correctly labelled pixels for each class and then averages over the classes

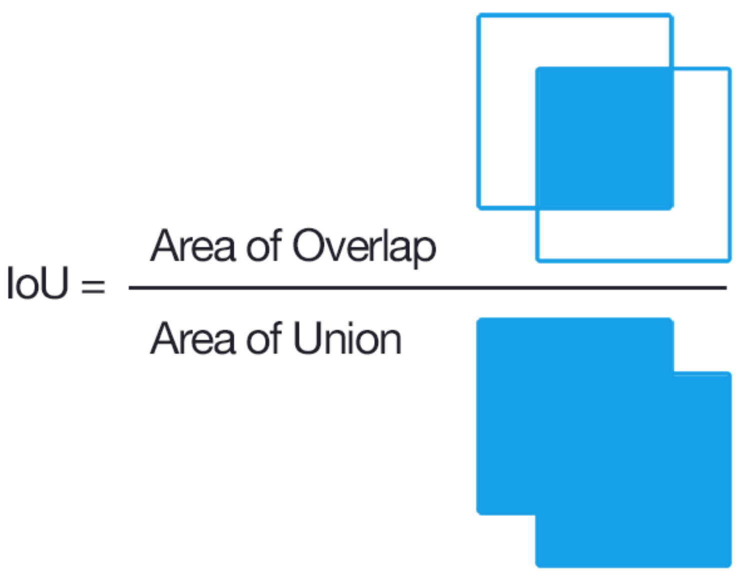
- Suitable for datasets with no background class

$$PC = \frac{1}{L} \sum_{i=1}^{L} \frac{\mathbf{C}_{ii}}{\mathbf{G}_i}$$

Intersection over Union (IoU)

- Measures the intersection over the union of the labelled segments for each class and reports the average

- IoU takes into account both the false alarms and the missed values for each class



Now assignment 3!