SegNet

Presenter: Khakim Akhunov
22.10.2019, NTNU

SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, Senior Member, IEEE

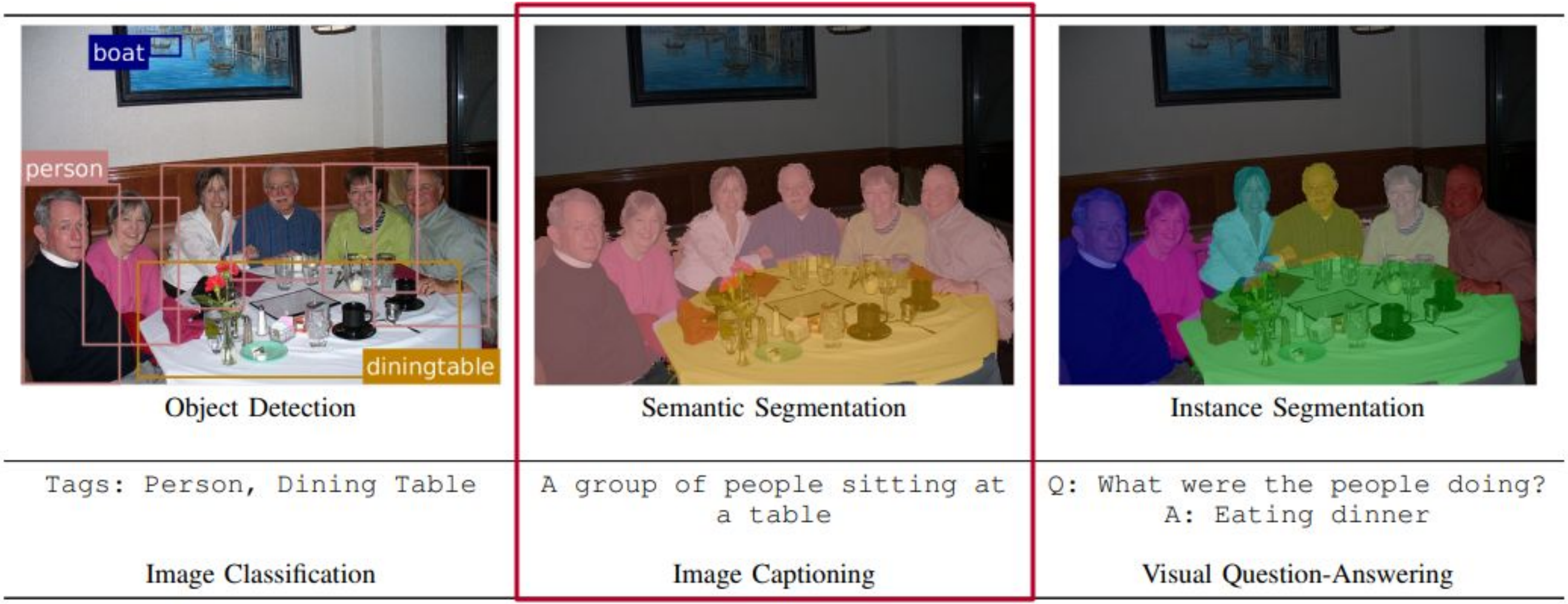
10 Oct 2016

What Semantic Segmentation is?

The process of assigning a label to every pixel in the image.

Semantic segmentation treats multiple objects of the same class as a single entity.

Example of various Scene Understanding tasks



Source: http://www.robots.ox.ac.uk/~tvg/publications/2017/CRFMeetCNN4SemanticSegmentation.pdf

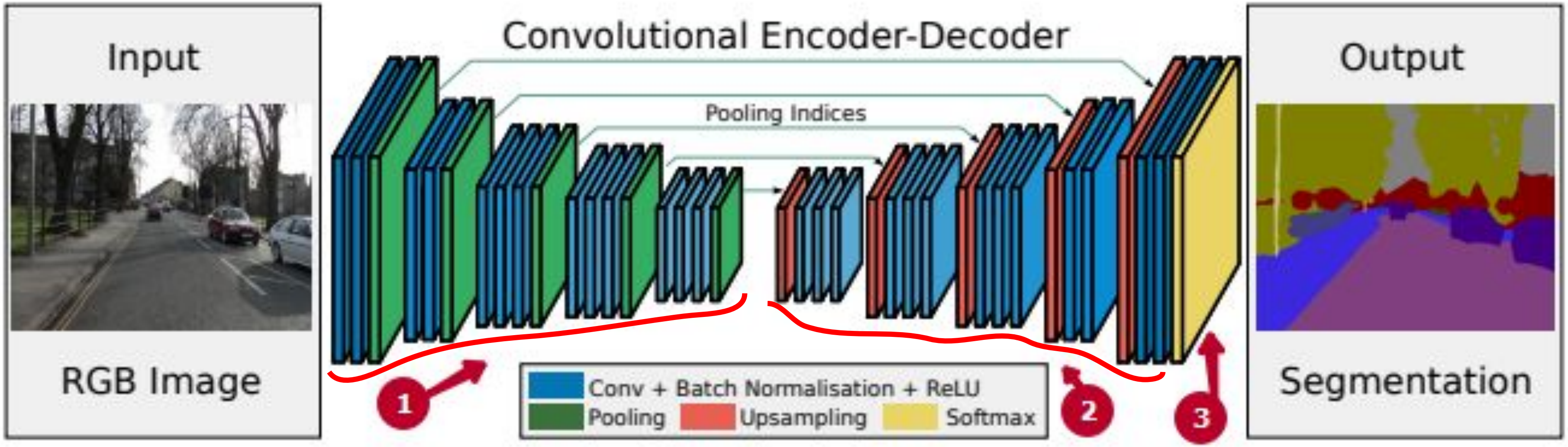
What SegNet is?

Novel and practical deep fully convolutional neural network architecture for semantic pixel-wise segmentation.

SegNet = (Encoder + Decoder) + Pixel-Wise Classification layer

It is primarily motivated by road scene understanding applications which require the ability to model appearance (road, building), shape (cars, pedestrians) and understand the spatial-relationship (context) between different classes such as road and side-walk.

SegNet architecture



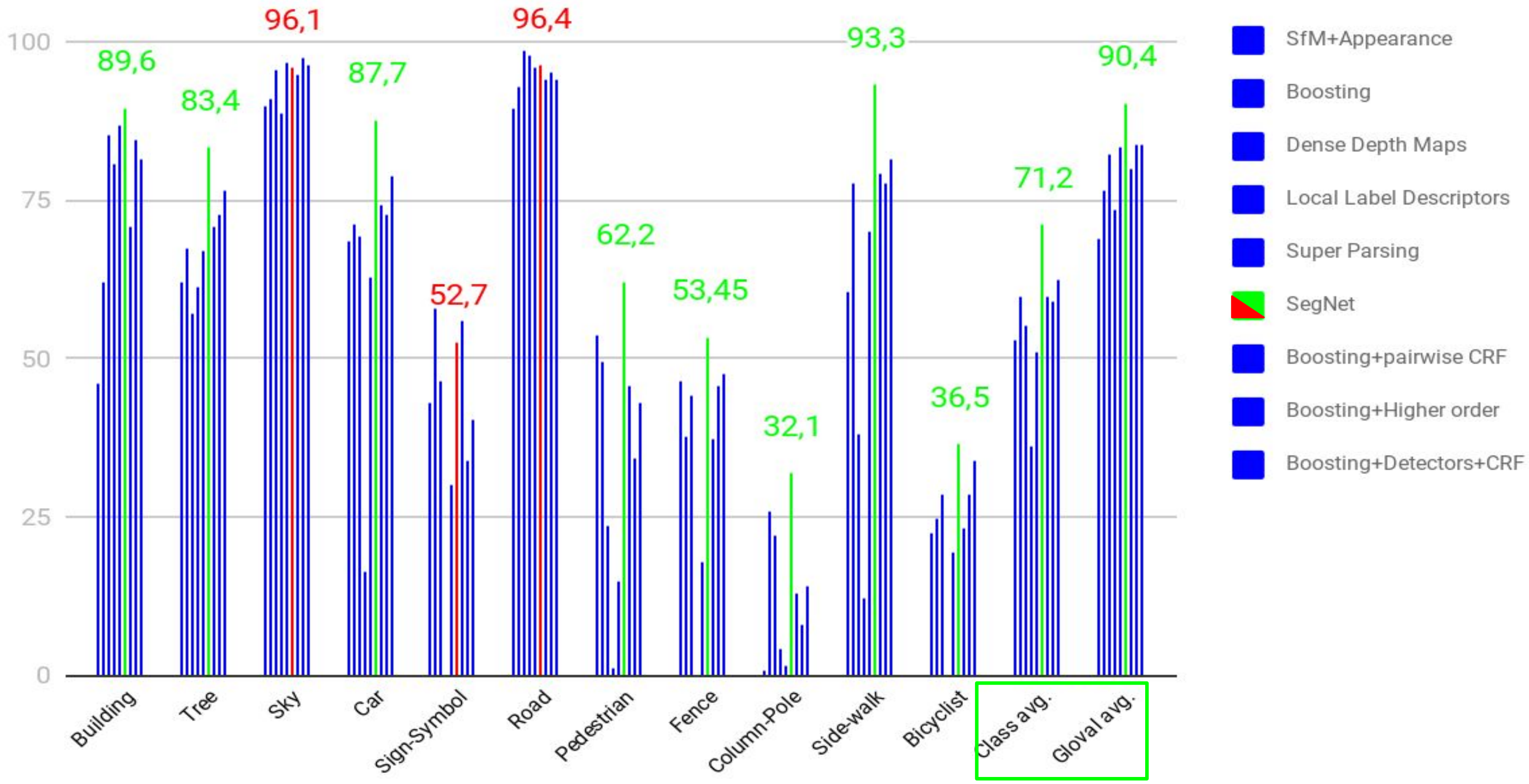
1 - encoder network
2 - decoder network
3 - pixel-wise classification layer

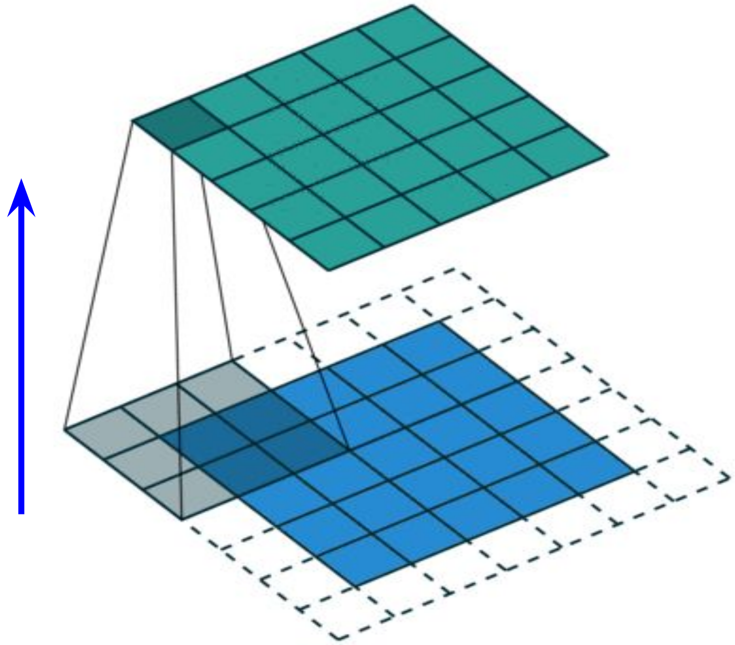
Main motivations behind SegNet

- Retain boundary information in the extracted image representation
- Efficient in terms of both memory and computation time
- Able to train end-to-end using efficient weight update technique

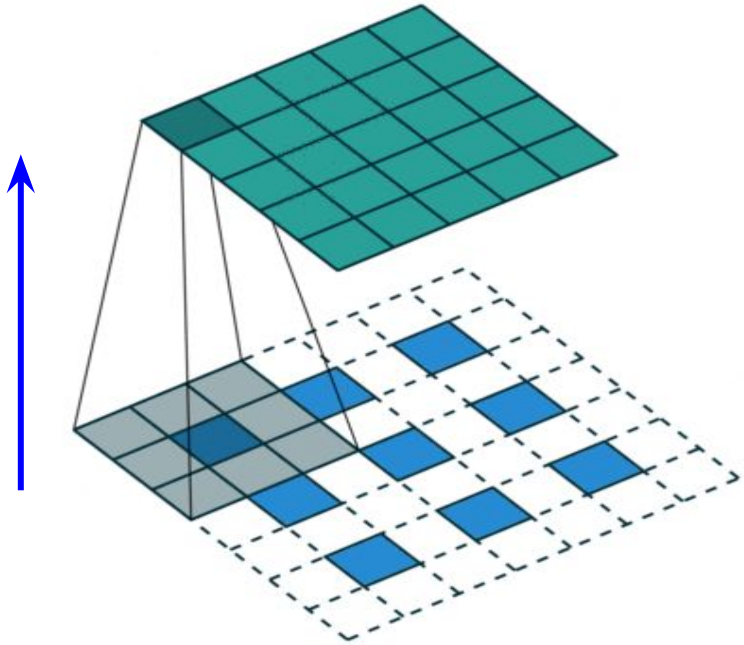
SegNet predictions on road scenes and indoor scenes

Quantitative comparisons of SegNet with traditional methods on the CamVid 11 road class segmentation problem



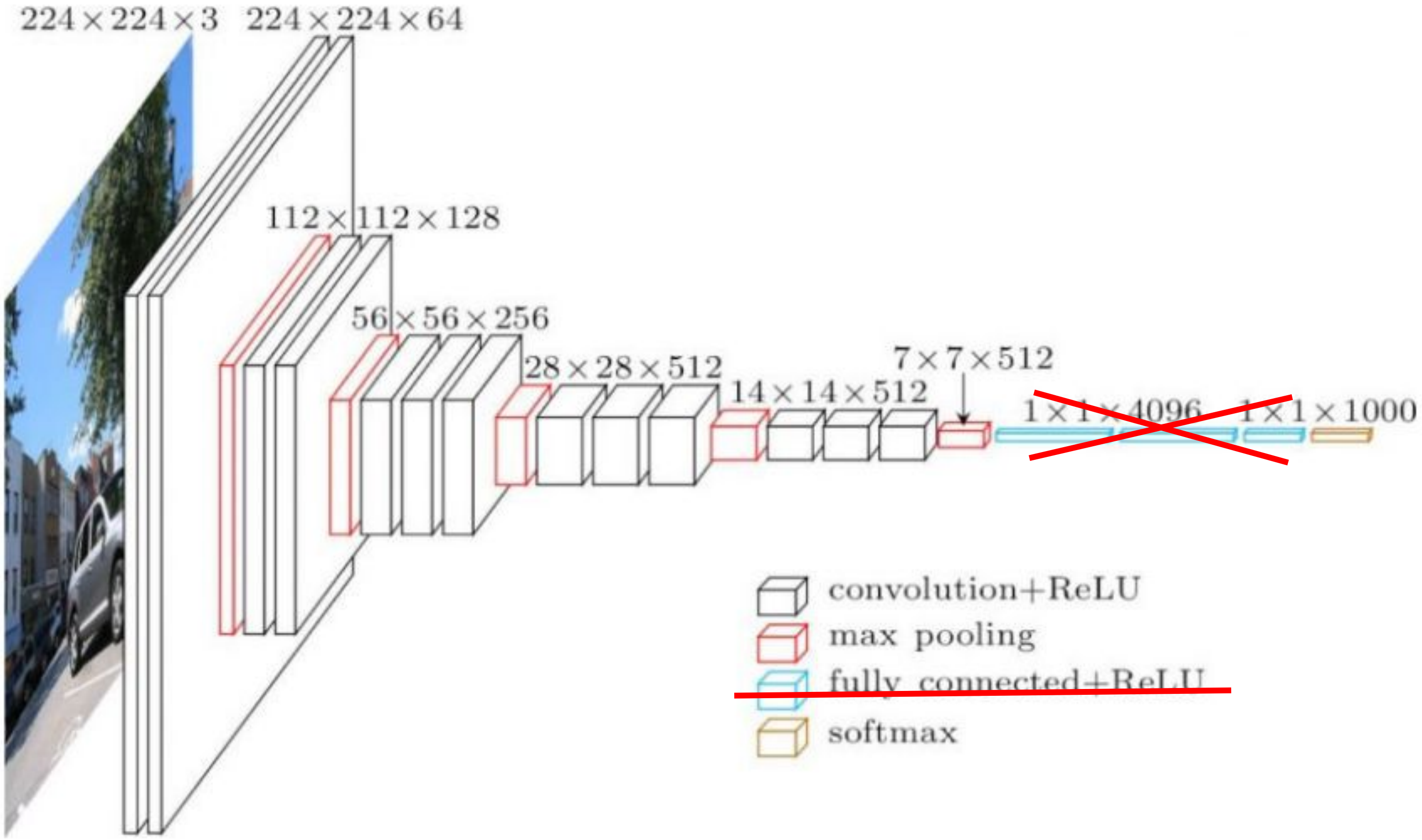


Encoding

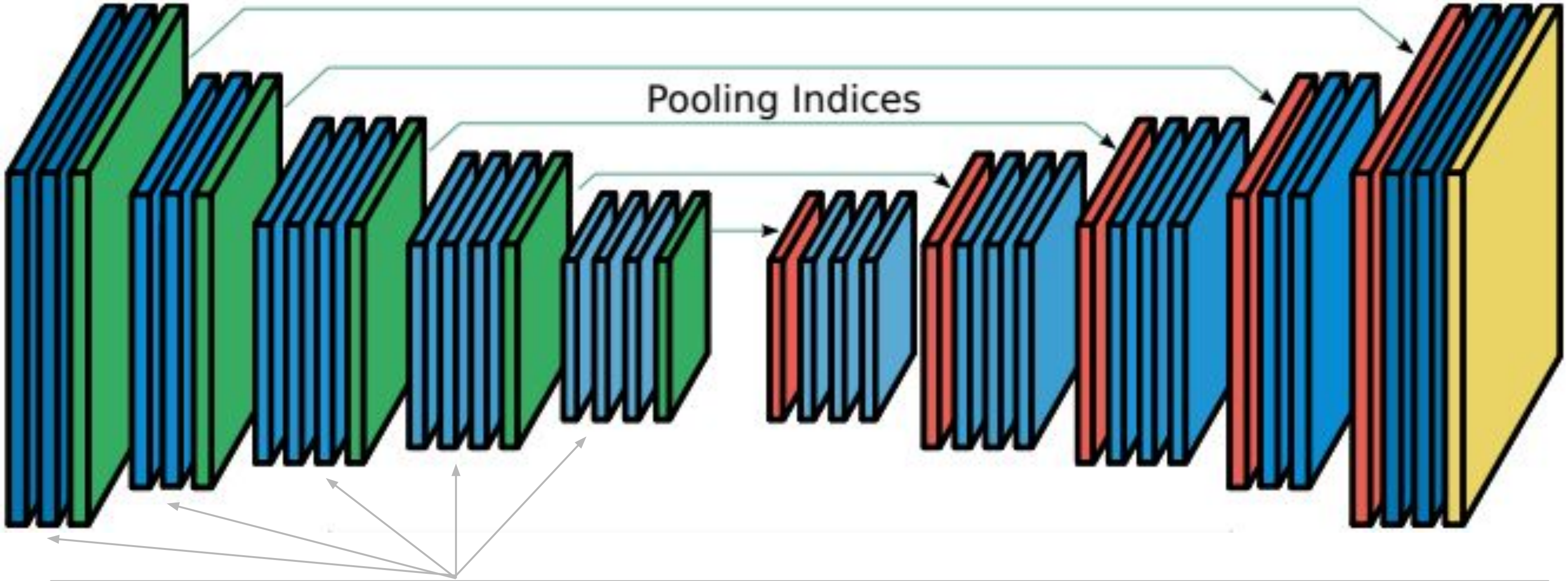


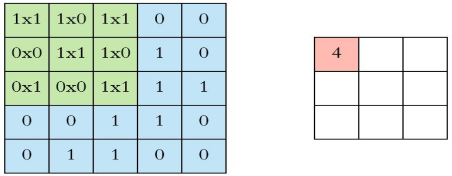
Decoding

VGG16 architecture

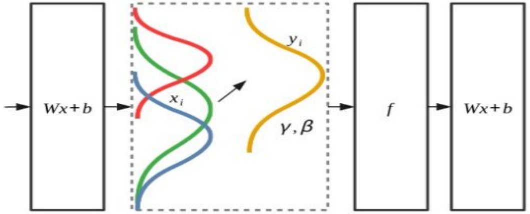


Convolutional Encoder-Decoder

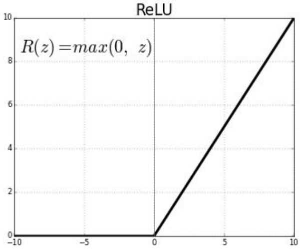




3x3conv

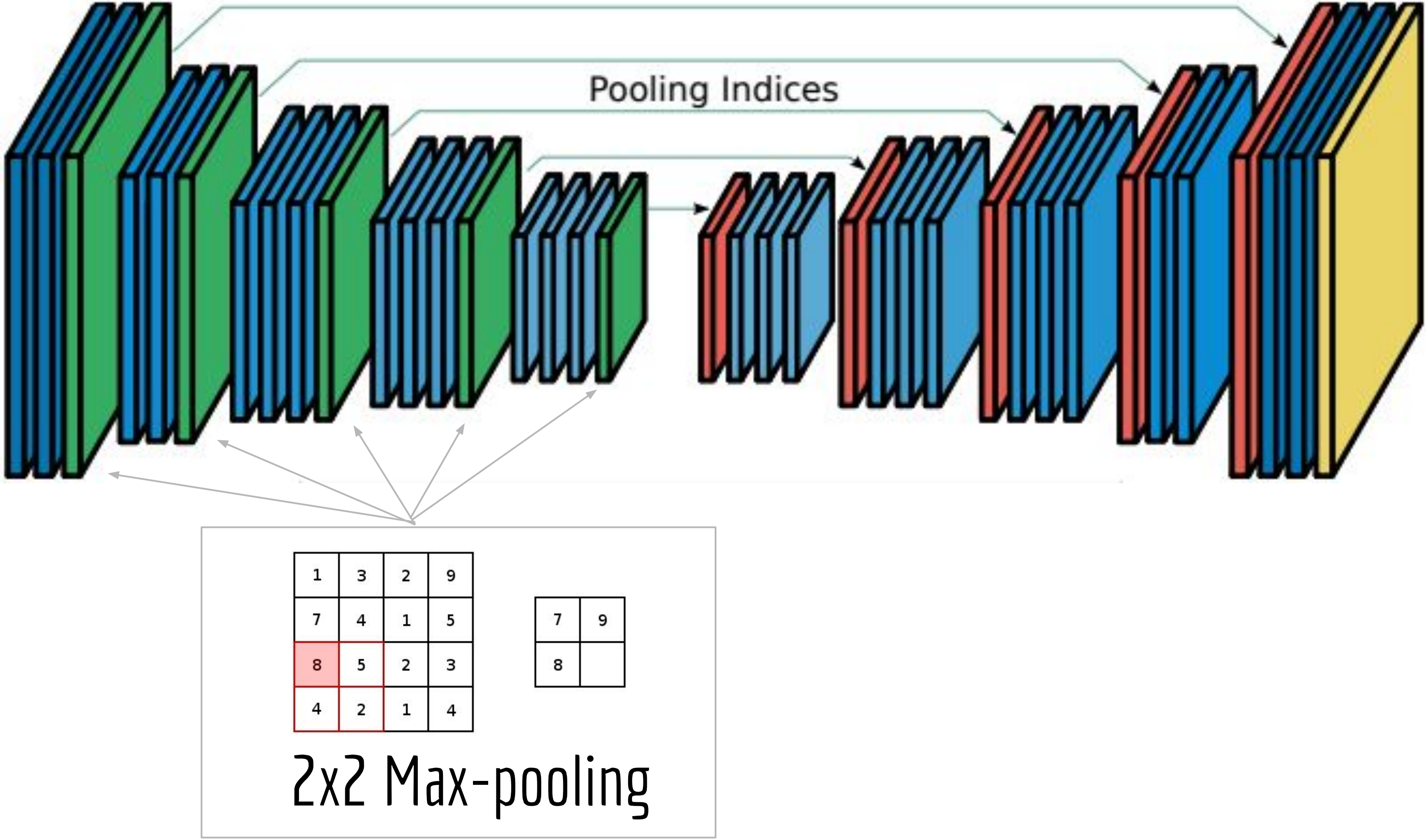


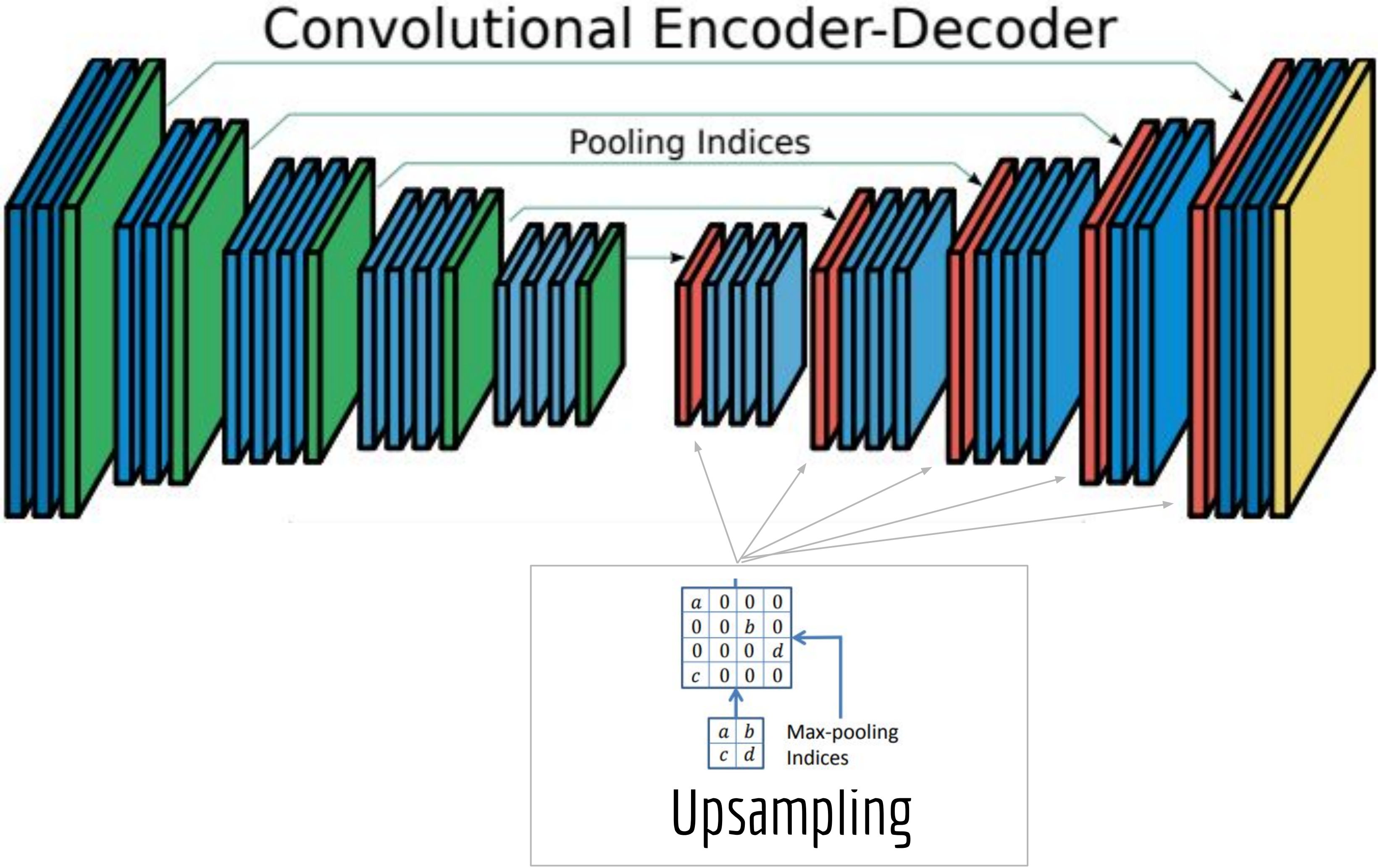
Batch Normalization



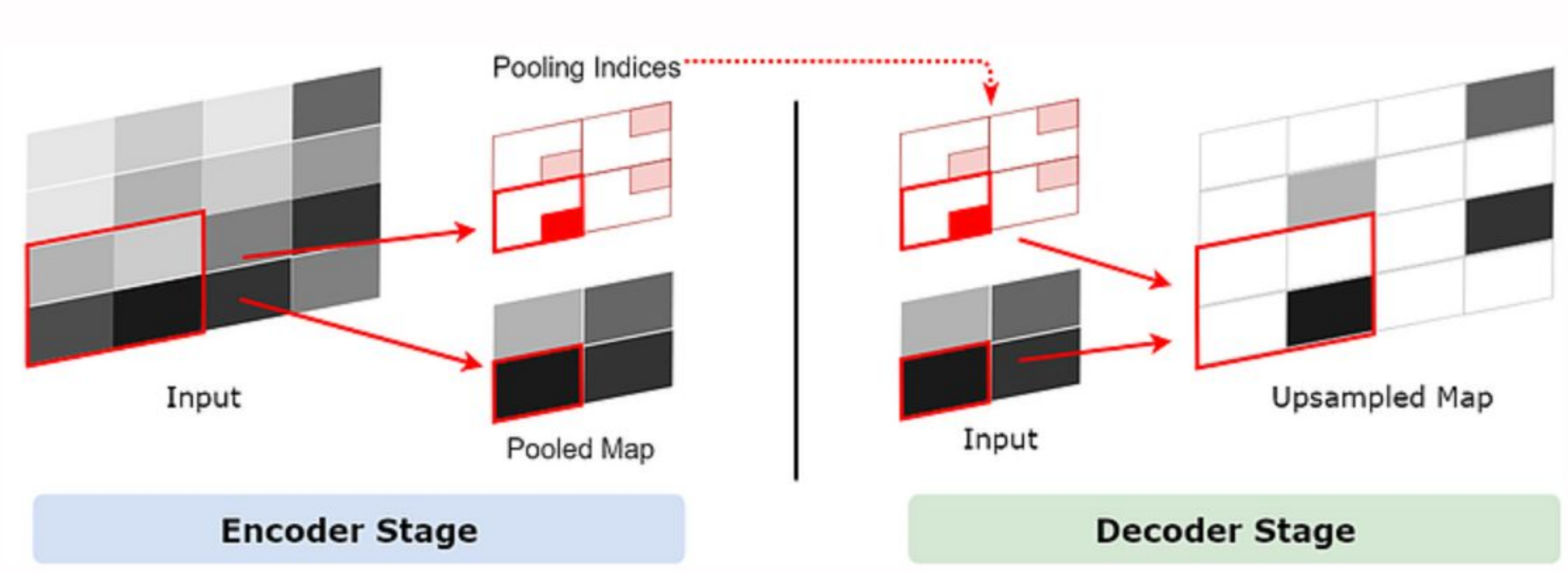
ReLU

Convolutional Encoder-Decoder



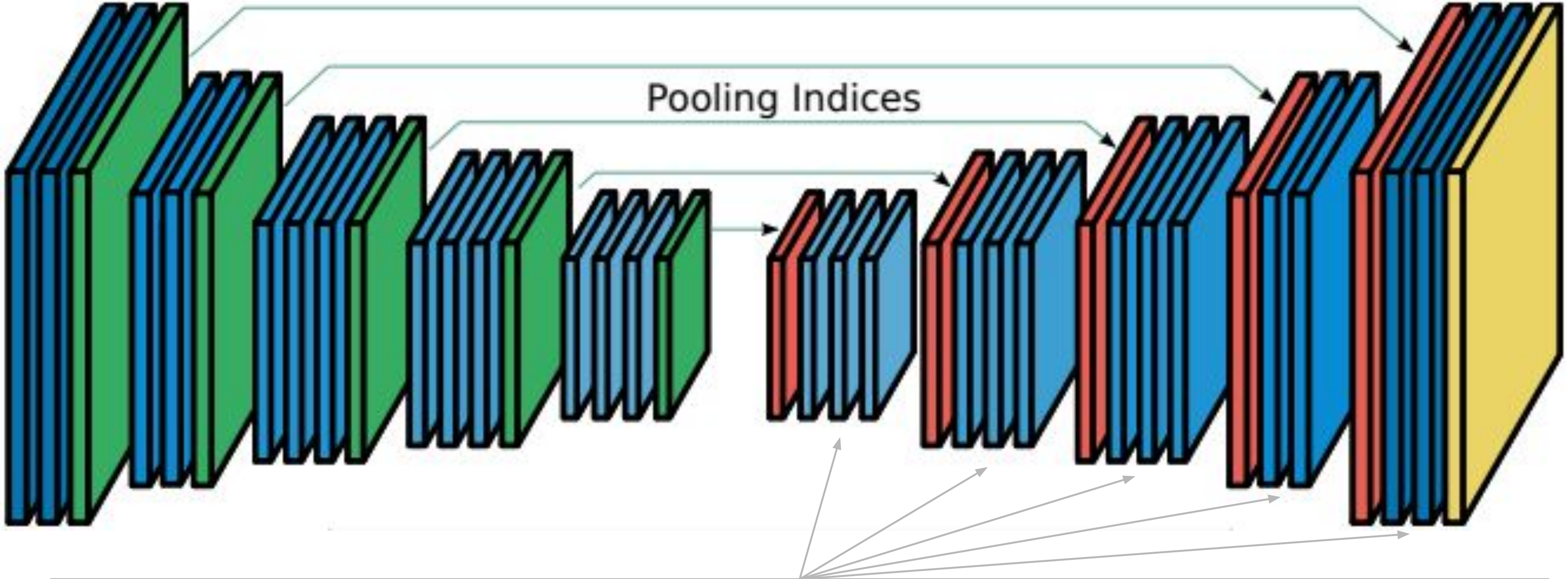

Convolutional Encoder-Decoder
Pooling Indices
a 0 0 0
0 0 b 0
0 0 0 d
c 0 0 0
a b
c d
Max-pooling
Indices
Upsampling

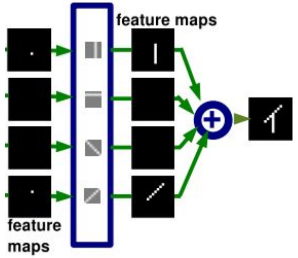
Using pooling indices for upsampling



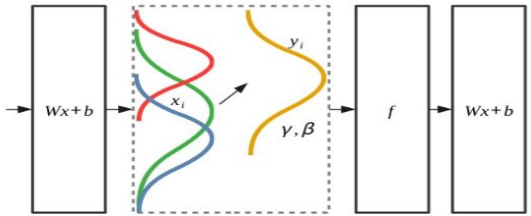
Source: https://www.cyberailab.com/home/segnet-an-image-segmentation-neural-network

Convolutional Encoder-Decoder

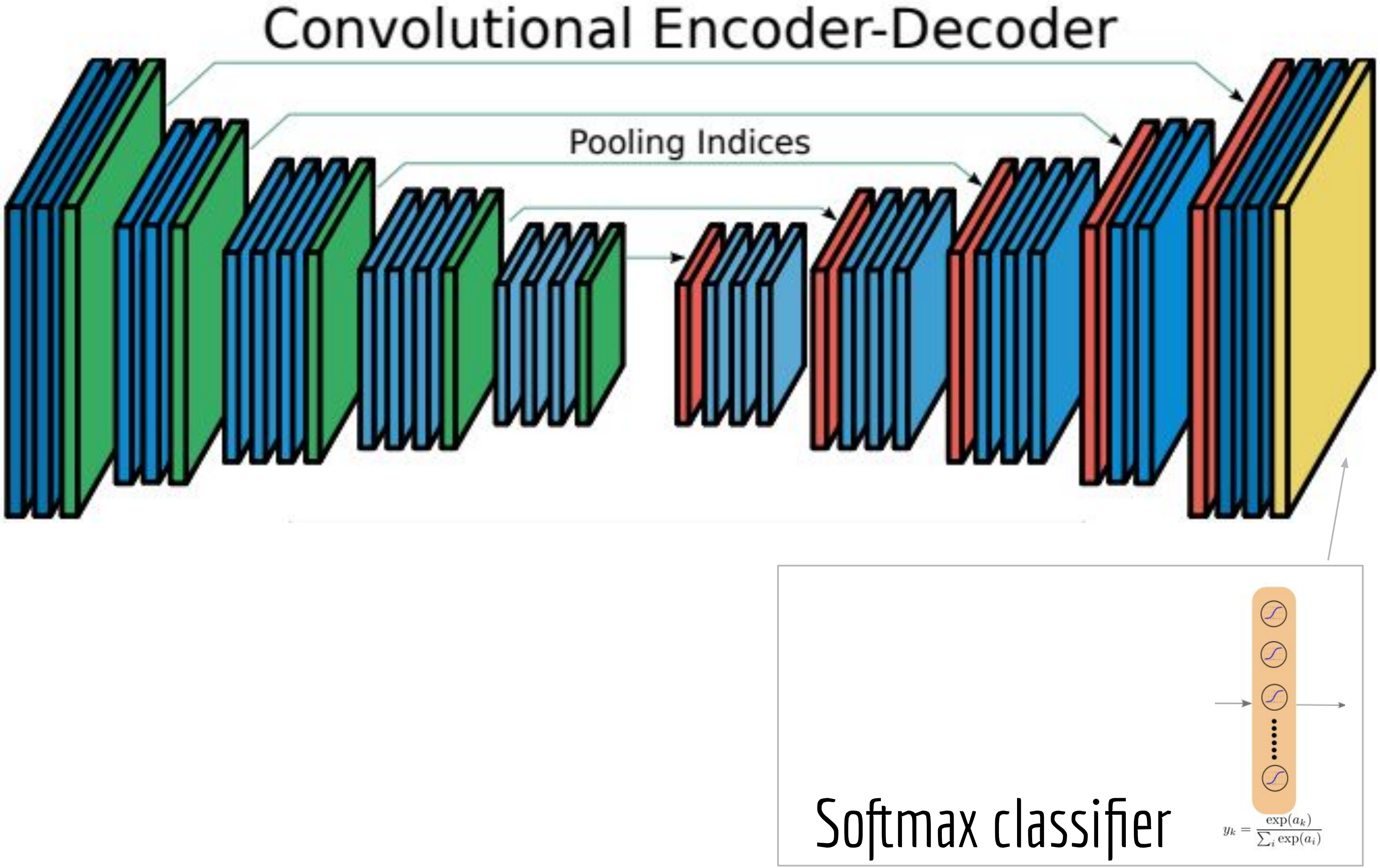




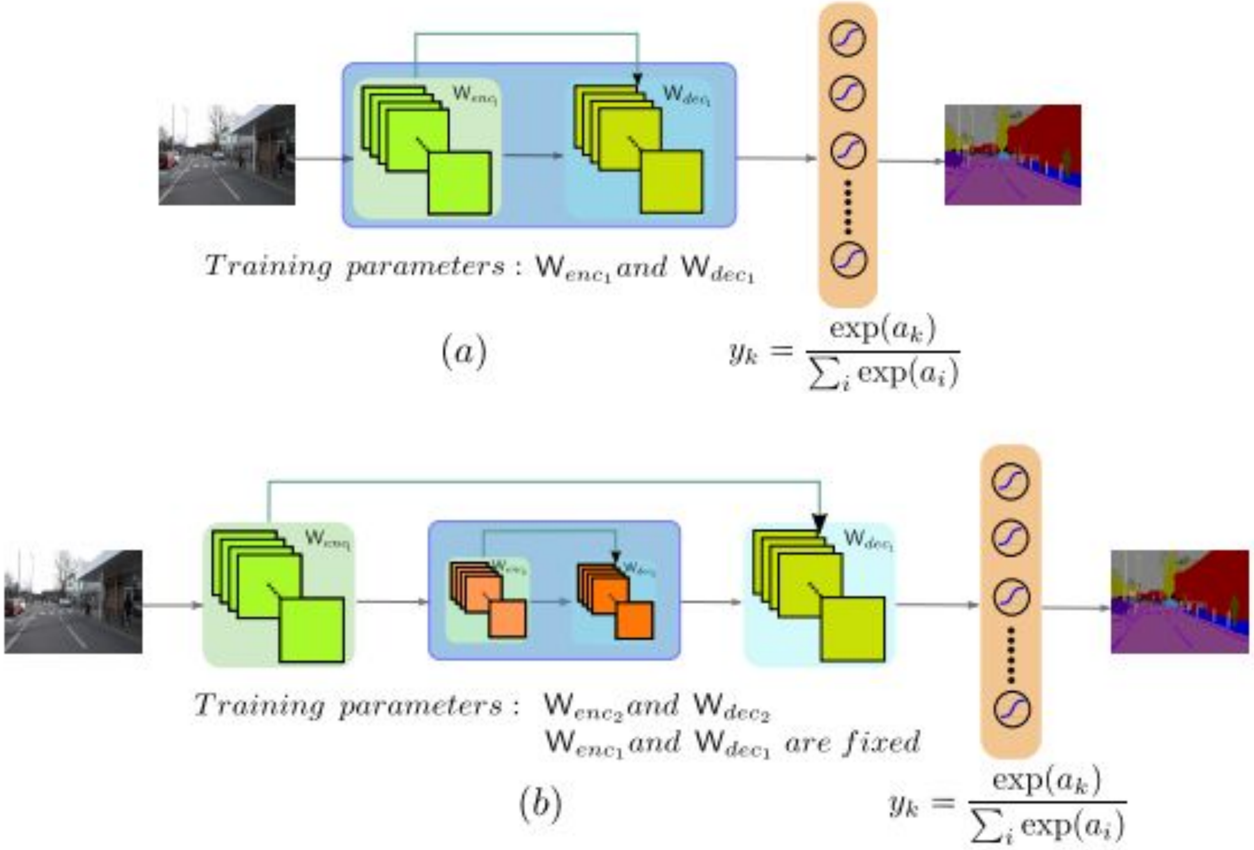
Decoder Filter Bank



Batch Normalization


Convolutional Encoder-Decoder
Pooling Indices
Softmax classifier
$y_k = \frac{\exp(a_k)}{\sum_i \exp(a_i)}$

Training the SegNet



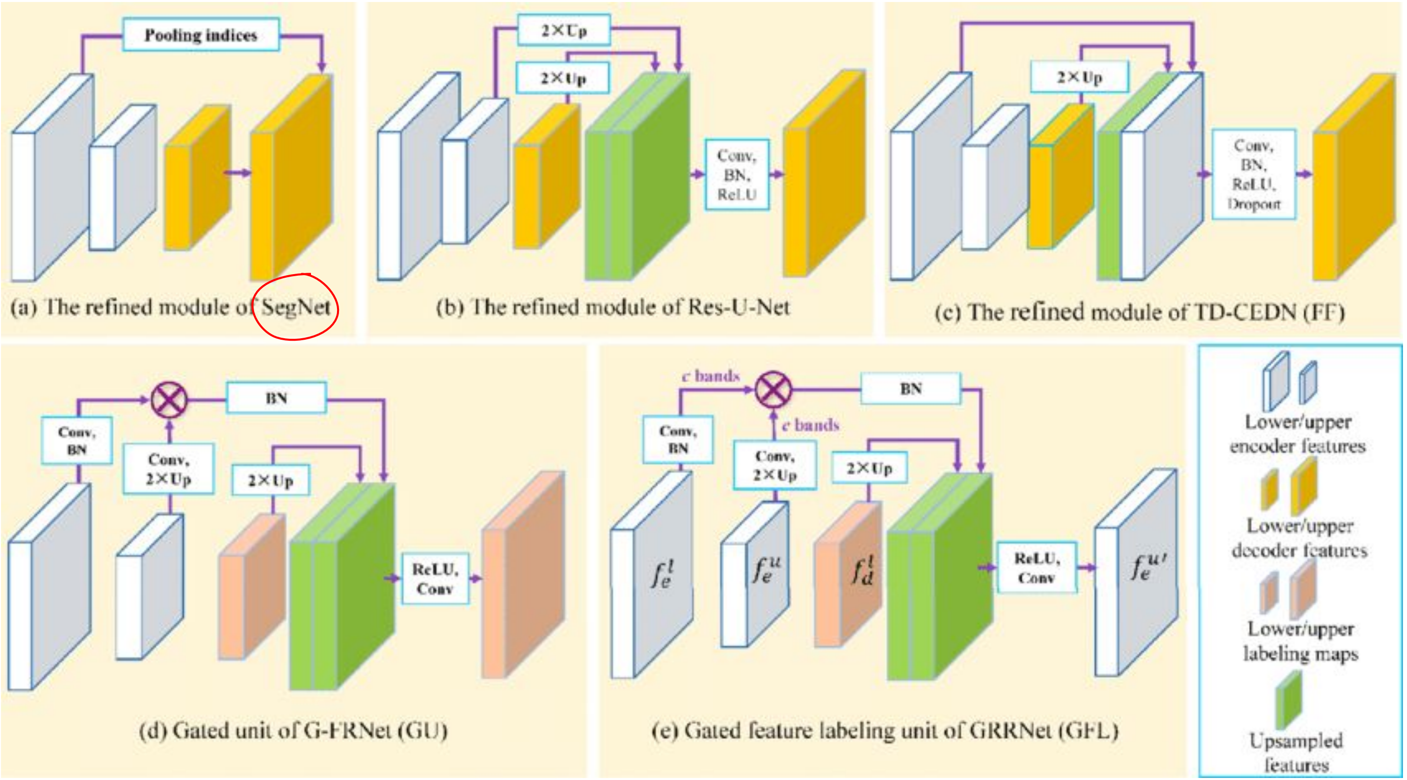
Comparison of decoder variants

Variant	Params (M)	Storage multiplier	Inference time (ms)
Fixed upsampling			
Bilinear-Interpolation	0.625	0	24.2
Upsampling using max-pooling indices			
SegNeg-Basic	1.425	1	52.6
SegNeg-Basic-Encoder Addition	1.425	64	53.0
SegNeg-Basic-SingleChannelDecoder	0.625	1	33.1
Learning to upsample			
FCN-Basic	0.65	11	24.2
FCN-Basic-NoAddition	0.65	n/a	23.8
FCN-Basic-NoDimReduction	1.625	64	44.8
FCN-Basic-NoAddition-NoDimReduction	1.625	0	43.9

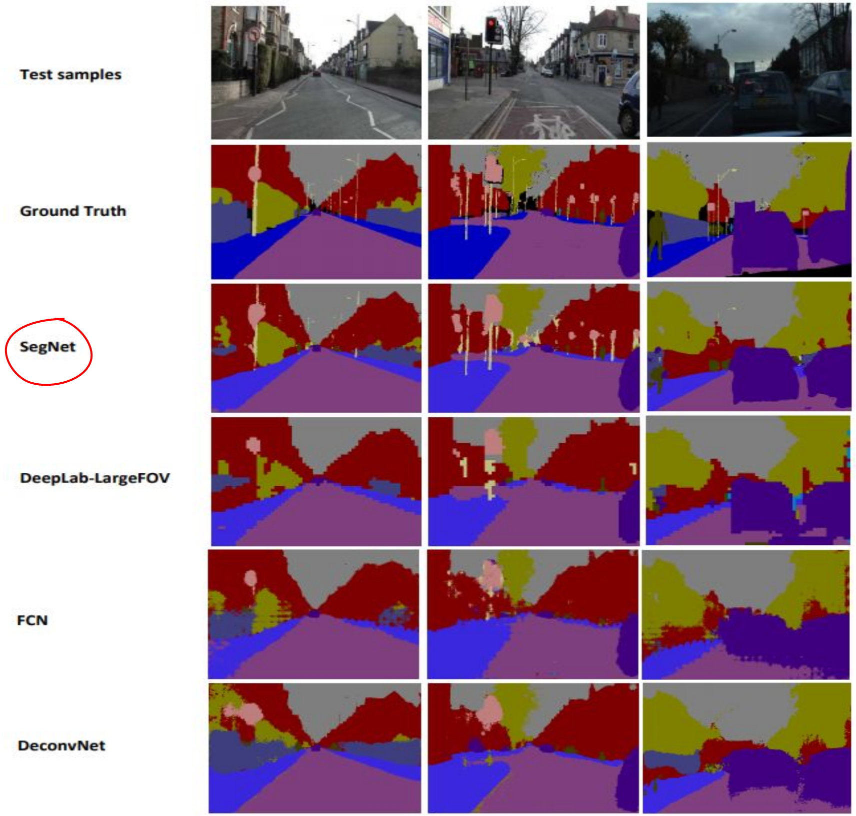
Summary of different decoders analysis

- The best performance is achieved when encoder feature maps are stored in full
- Compressed forms of encoder feature maps can be stored and used for decoding to meet memory constraints
- Larger decoders increase performance for a given encoder netrowork

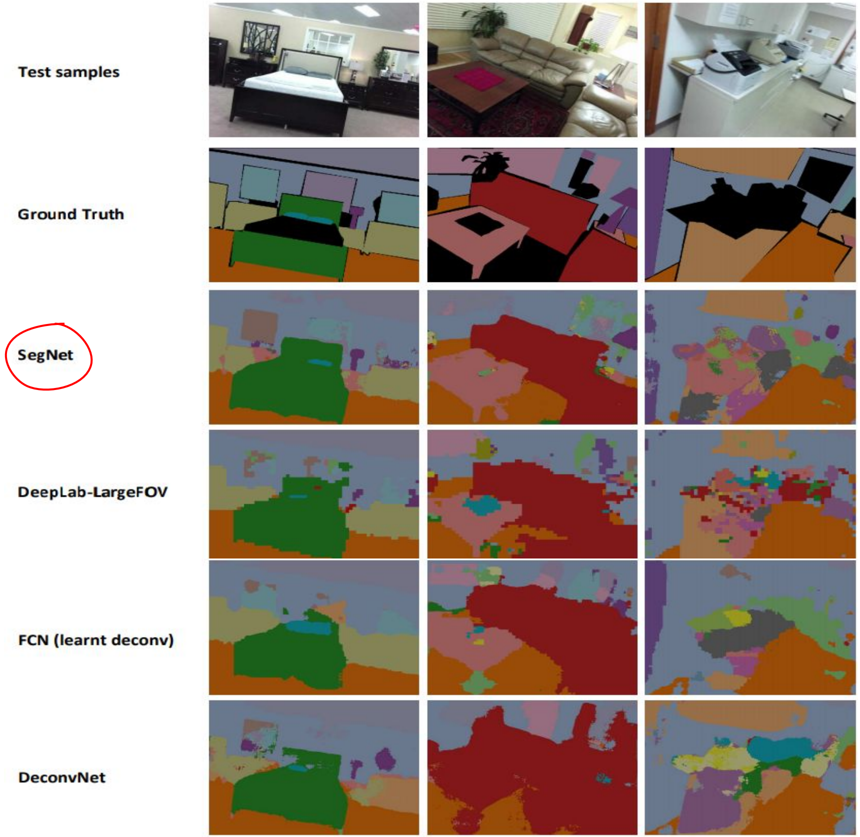
Schematic representation of different architectures



Road scene segmentation



Indoor scene segmentation



Conclusion

- SegNet is more efficient compared to other architectures since it only stores the max-pooling indices of the feature maps and uses them in its decoder network to achieve good performance
- On large and well known datasets SegNet performs competitively, achieving high scores for road scene understanding
- End-to-end learning of deep segmentation architectures is a harder challenge

References

1) http://mi.eng.cam.ac.uk/projects/segnet/
2) https://www.cyberailab.com/home/segnet-an-image-segmentation-neural-network
3) http://www.robots.ox.ac.uk/~tvg/publications/2017/CRFMeetCNN4SemanticSegmentation.pdf
4) https://neurohive.io/en/popular-networks/vgg16/

Questions?