Encoder-Decoder Networks for Semantic Segmentation

Sachin Mehta

Outline

> Overview of Semantic Segmentation
> Encoder-Decoder Networks
> Results

What is Semantic Segmentation?

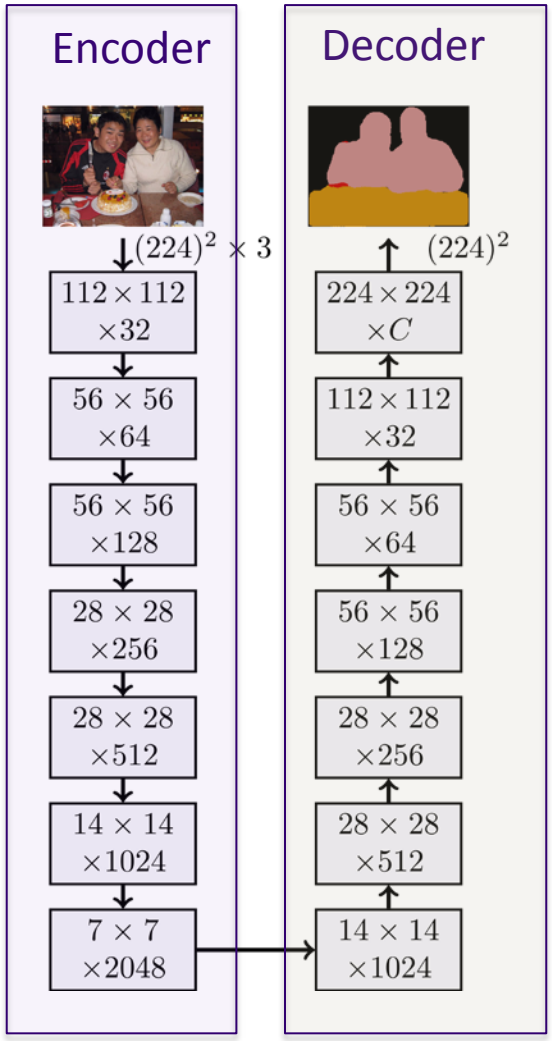
Input: RGB Image

Output: A segmentation Mask

Encoder-Decoder Networks

Encoder

- Takes an input image and generates a high-dimensional feature vector
- Aggregate features at multiple levels



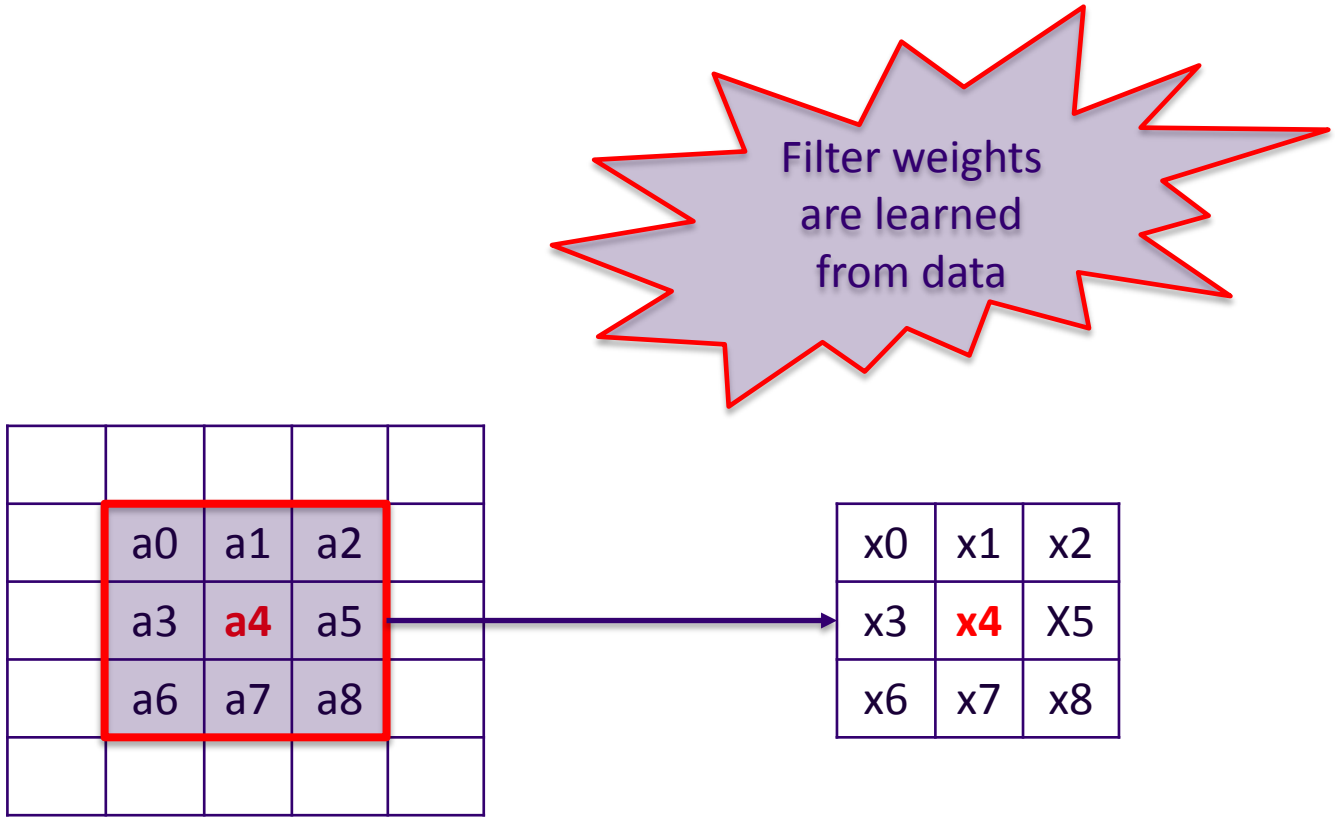
Decoder

- Takes a high-dimensional feature vector and generates a semantic segmentation mask
- Decode features aggregated by encoder at multiple levels

Building Blocks of CNNs

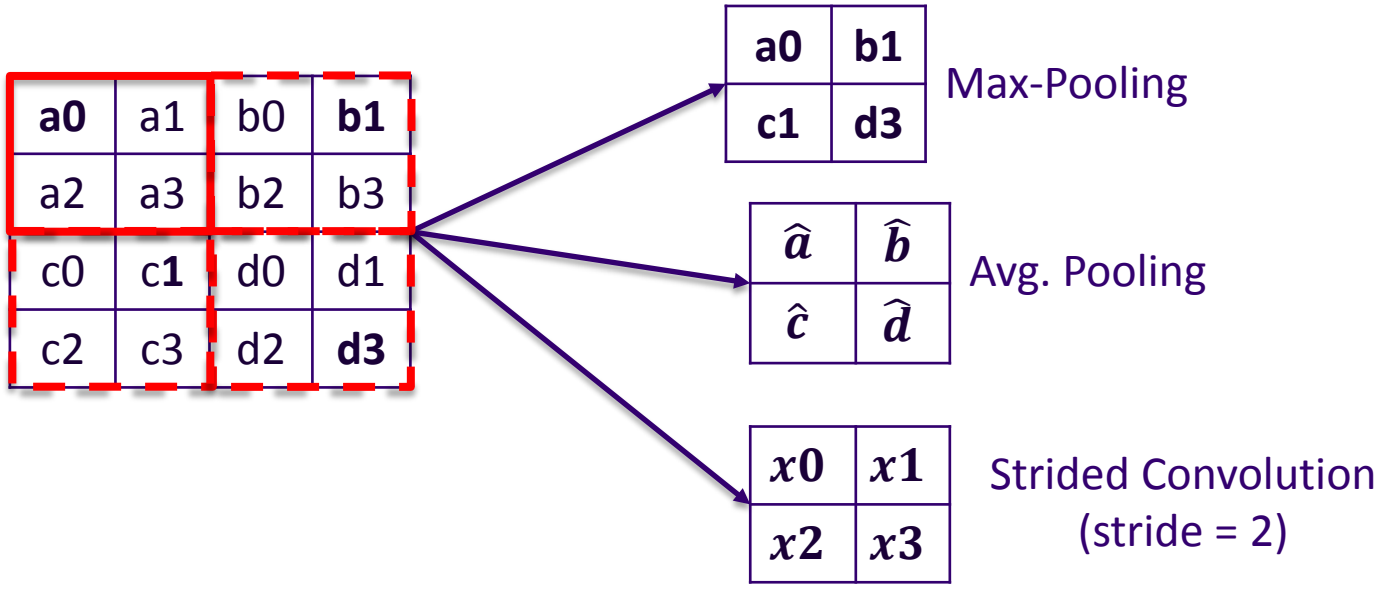
- Convolution
- Down-Sampling
- Up-Sampling

Convolution



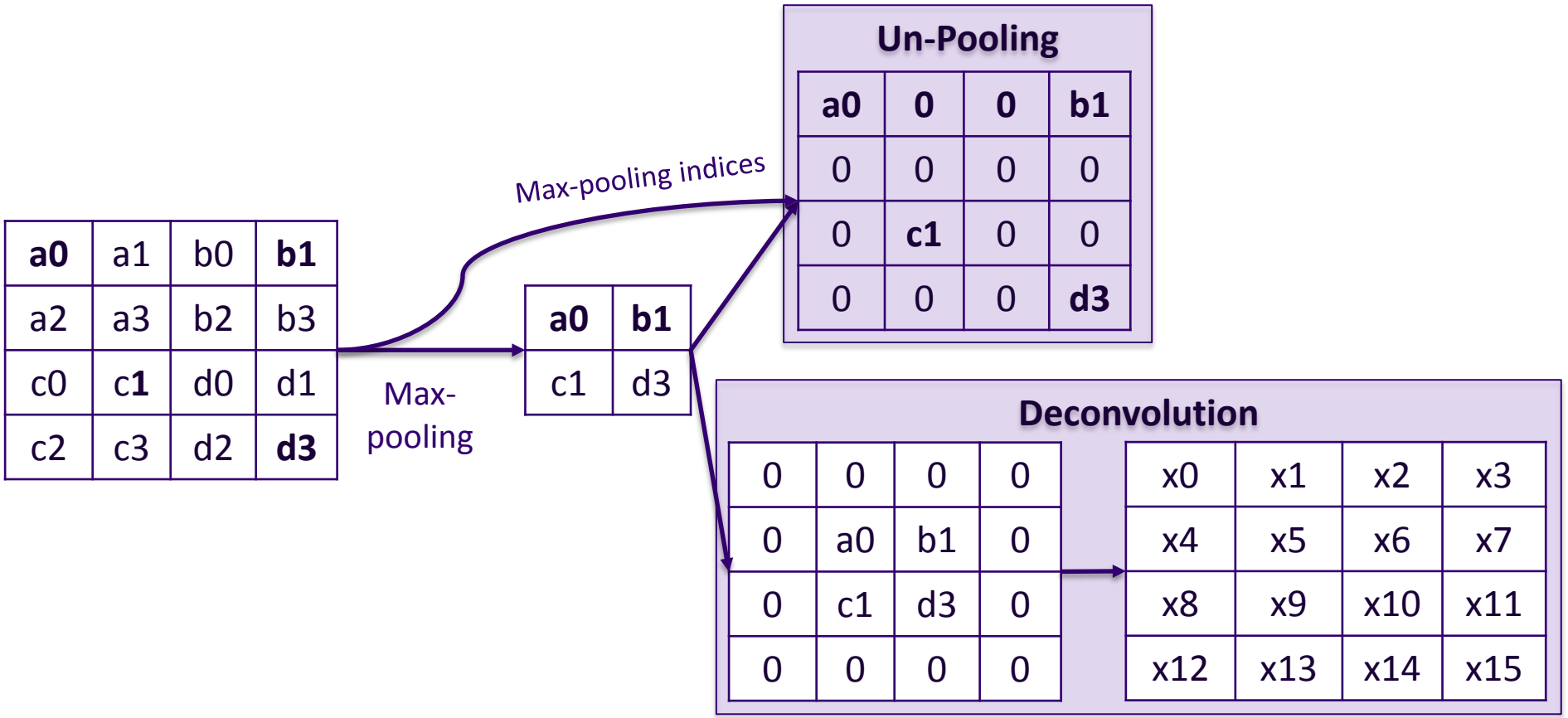
Down-Sampling

- Max-pooling
- Average Pooling
- Strided Convolution

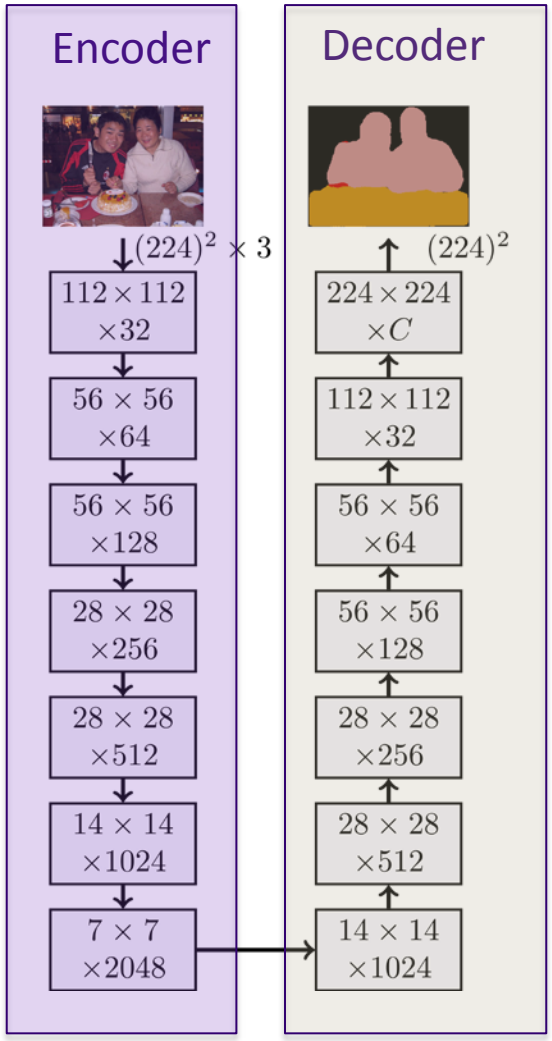


Up-Sampling

> Un-pooling
> Deconvolution



Encoder-Decoder Networks



Encoder-Decoder Networks

Different Encoding Block Types

$(224)^2 \times 3$

112×112
$\times 32$

56×56
$\times 64$

56×56
$\times 128$

28×28
$\times 256$

28×28
$\times 512$

14×14
$\times 1024$

7×7
$\times 2048$

- VGG
- Inception
- ResNet

Encoder-Decoder Networks

Different Encoding Block Types

$(224)^2 \times 3$

112×112
$\times 32$

56×56
$\times 64$

56×56
$\times 128$

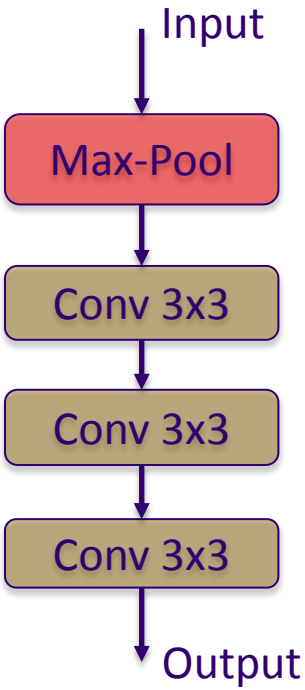
28×28
$\times 256$

28×28
$\times 512$

14×14
$\times 1024$

7×7
$\times 2048$

- VGG



Encoder-Decoder Networks

Different Encoding Block Types

$(224)^2 \times 3$

112×112
$\times 32$

56×56
$\times 64$

56×56
$\times 128$

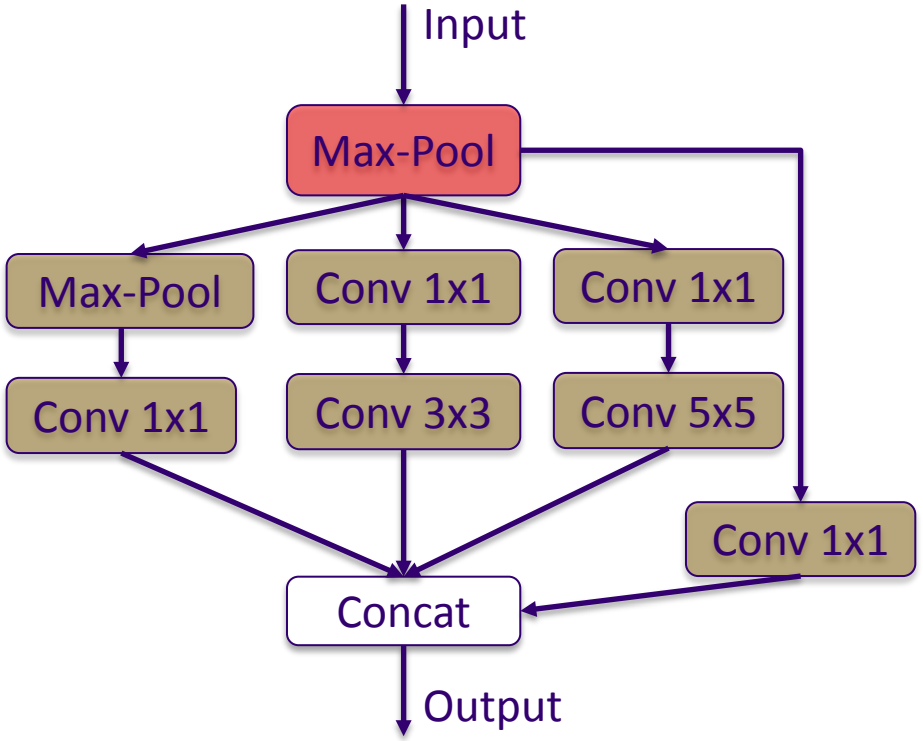
28×28
$\times 256$

28×28
$\times 512$

14×14
$\times 1024$

7×7
$\times 2048$

- Inception



Encoder-Decoder Networks

Different Encoding Block Types

$(224)^2 \times 3$

112×112
$\times 32$

56×56
$\times 64$

56×56
$\times 128$

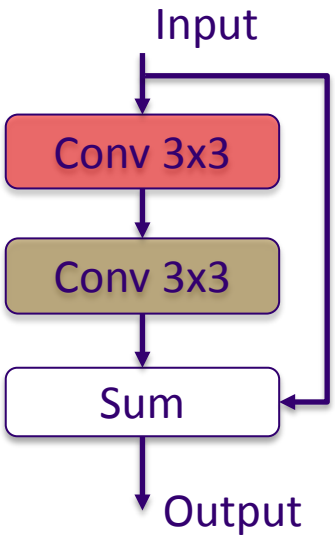
28×28
$\times 256$

28×28
$\times 512$

14×14
$\times 1024$

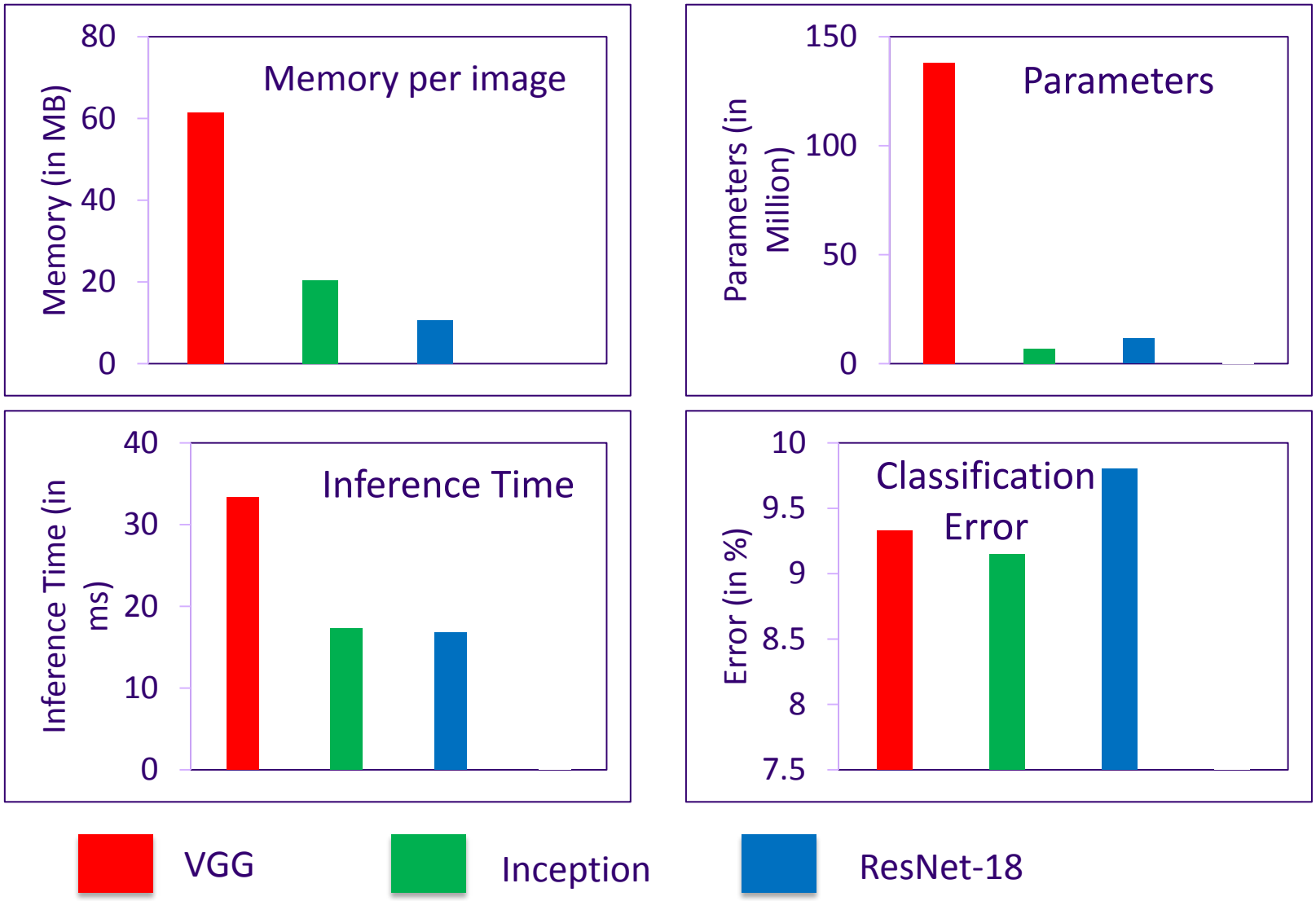
7×7
$\times 2048$

- ResNet

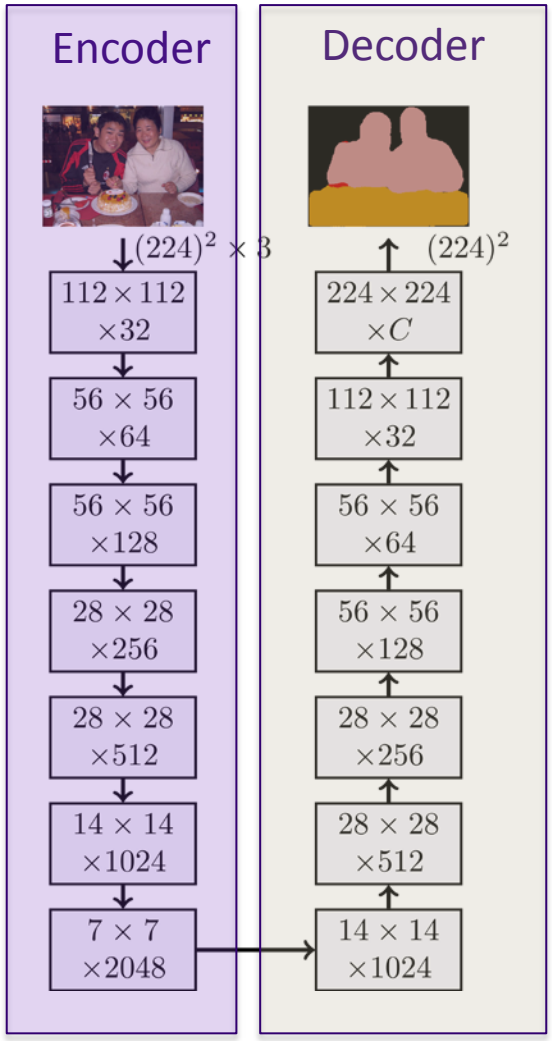


Different Encoding Block Types

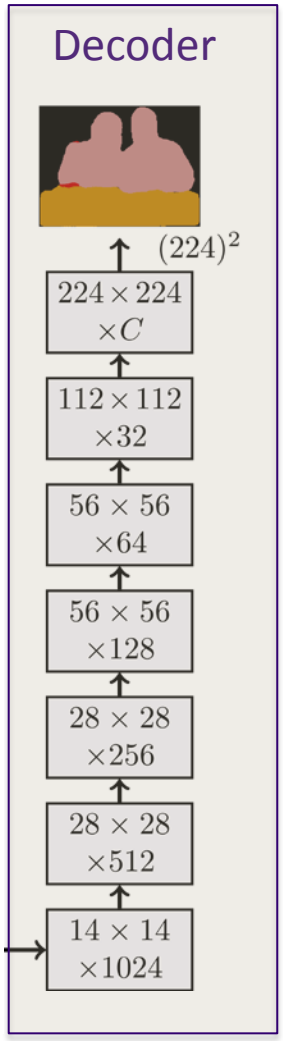
Performance on the ImageNet 2012 Validation Dataset



Encoder-Decoder Networks

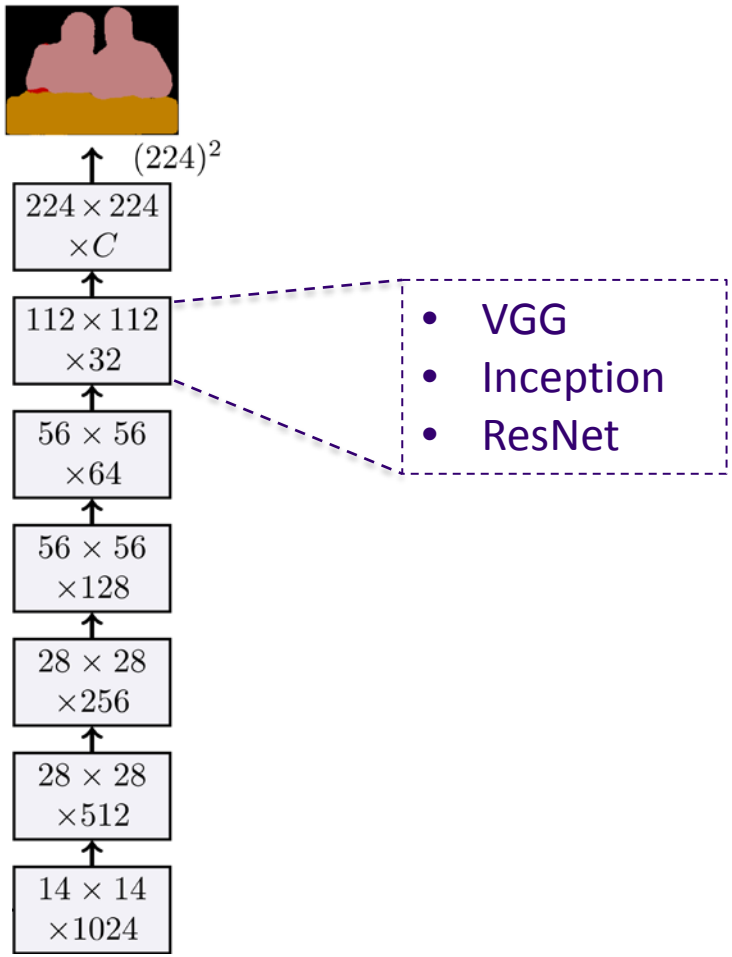


Encoder-Decoder Networks



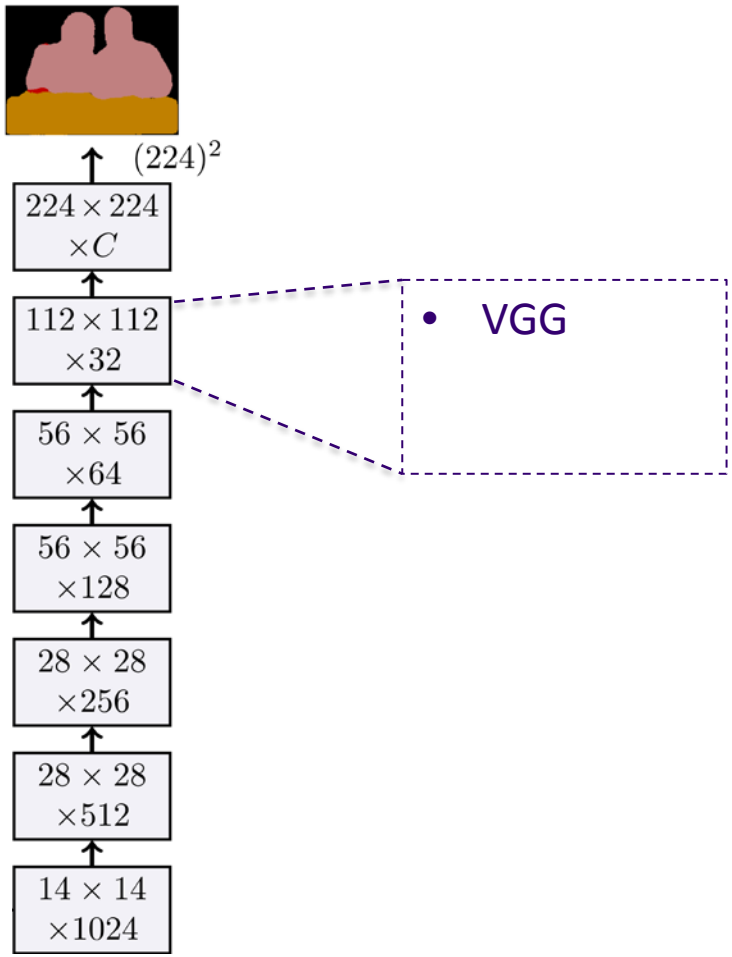
Encoder-Decoder Networks

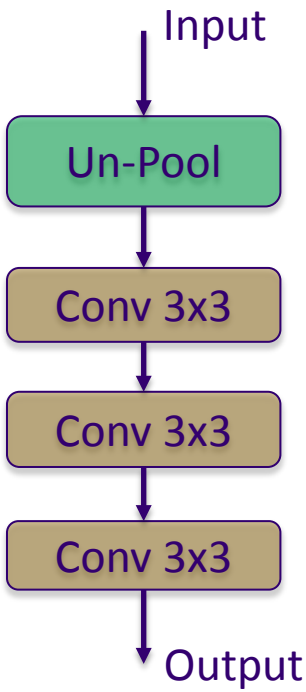
Different Decoding Block Types



Encoder-Decoder Networks

Different Decoding Block Types





Encoder-Decoder Networks

Different Decoding Block Types

$(224)^2$

224×224
$\times C$

112×112
$\times 32$

56×56
$\times 64$

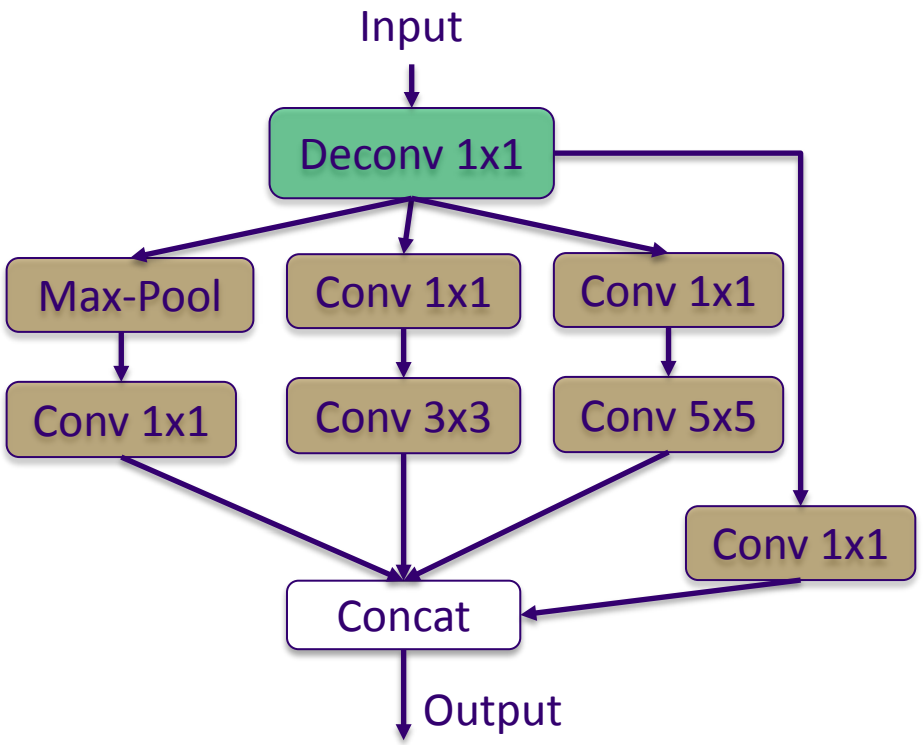
56×56
$\times 128$

28×28
$\times 256$

28×28
$\times 512$

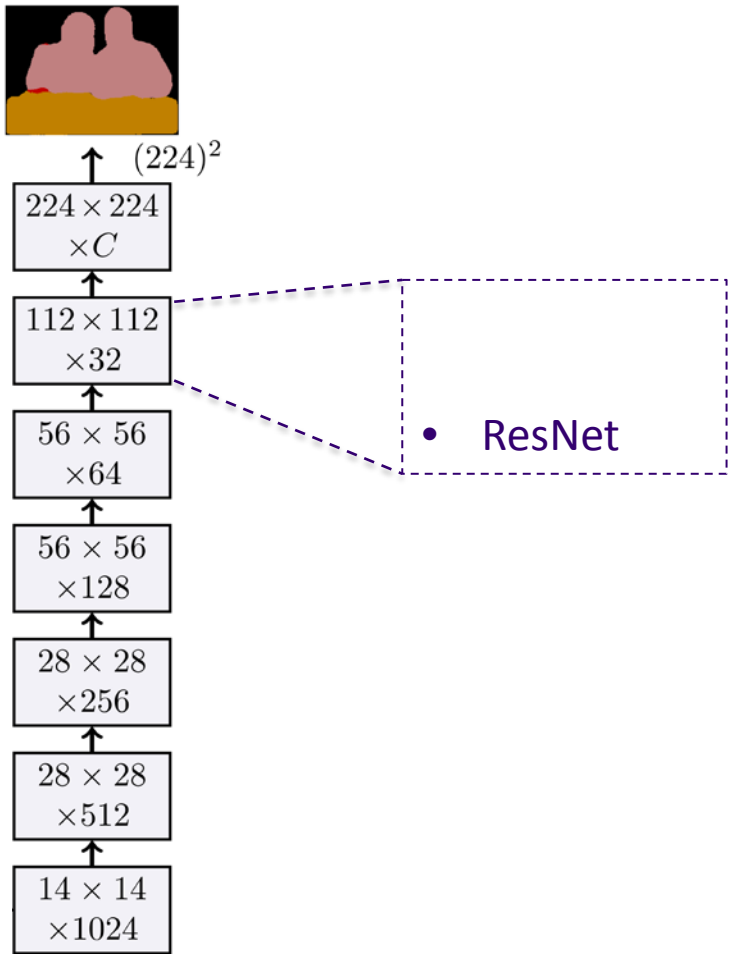
14×14
$\times 1024$

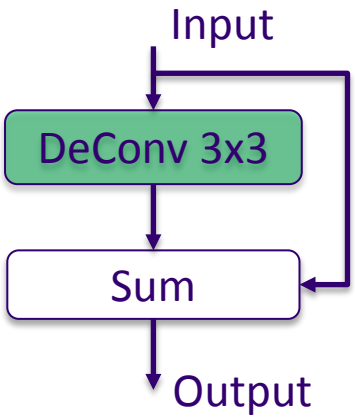
- Inception



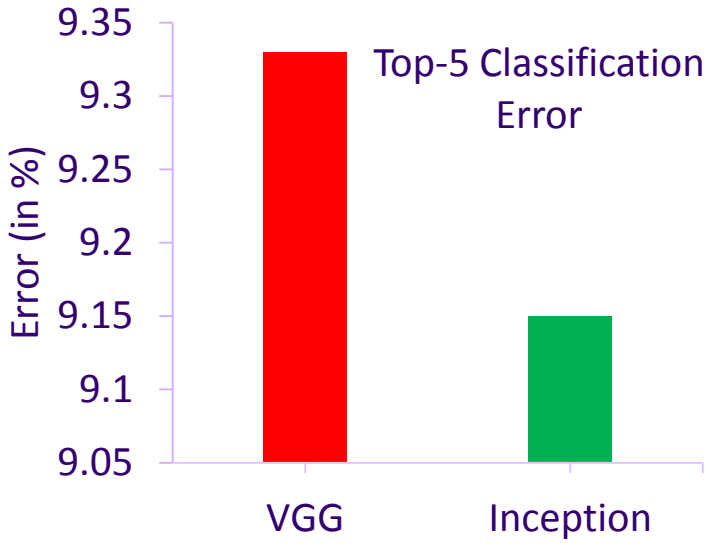
Encoder-Decoder Networks

Different Decoding Block Types

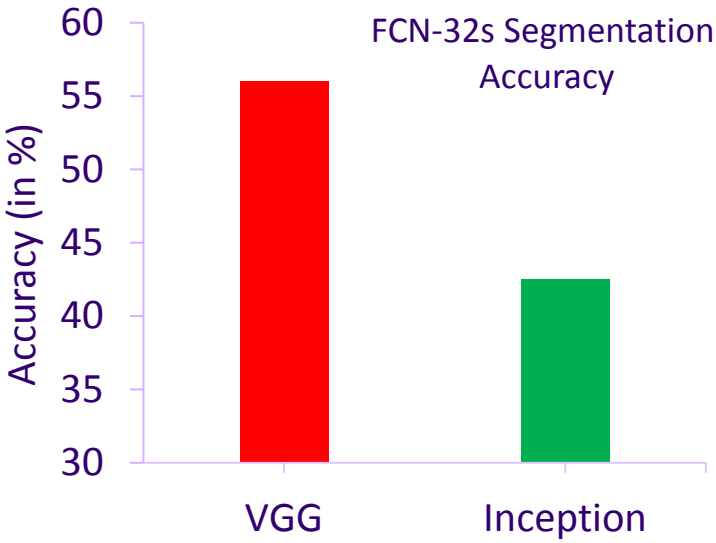




Classification vs Segmentation



(a) ImageNet Classificaiton Validation Set



(b) PASCAL VOC 2011 Validation Set

VGG

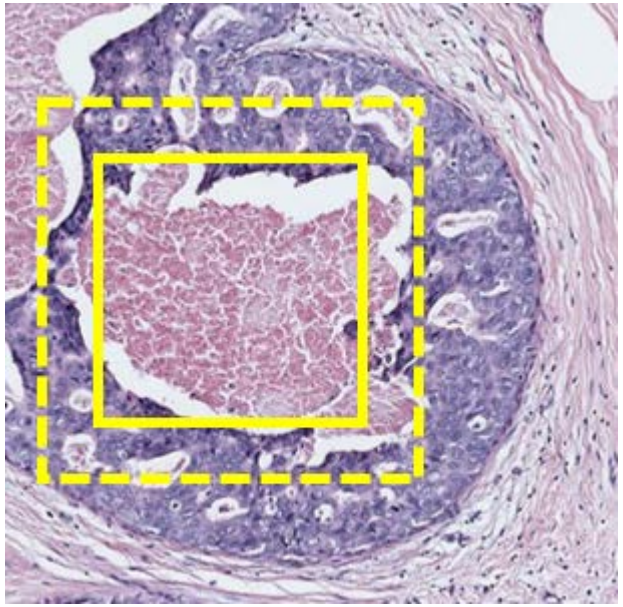
Inception

Source: Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *TPAMI*. 2016.

Our Work on Segmenting GigaPixel Breast Biopsy Images

Challenges with the dataset

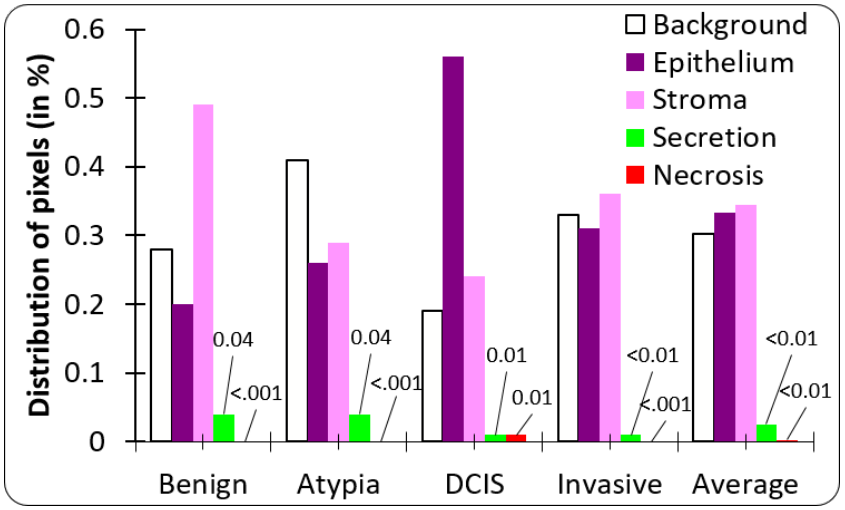
- Limited computational resources
- Sliding window approach is promising but
  - Size of patch determines the context
  - Some biological structures may cover several patches



Challenges with the dataset

- Some biological structures are rare
  - Necrosis and Secretion have less than 1% of all the pixels

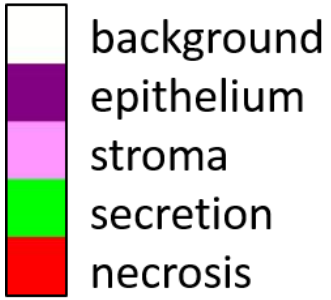
Diagnosis Category	#ROIs (total)	#ROIs (train)	#ROIs (test)	Avg. ROI size
Benign	9	4	5	$9K \times 9K$
Atypia	22	11	11	$6K \times 7K$
DCIS	22	12	10	$8K \times 10K$
Invasive	5	3	2	$38K \times 44K$
total	**58**	**30**	**28**	$10K \times 12K$

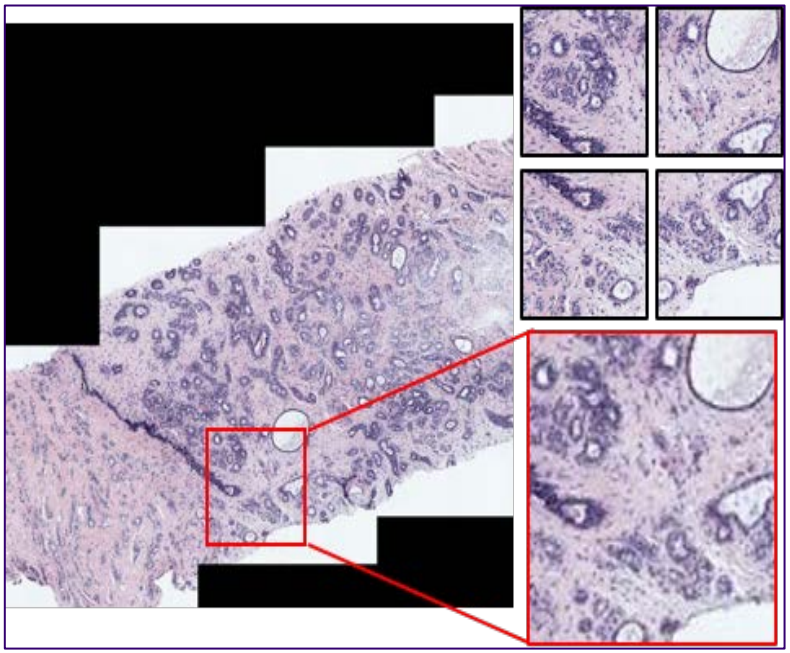


Training details

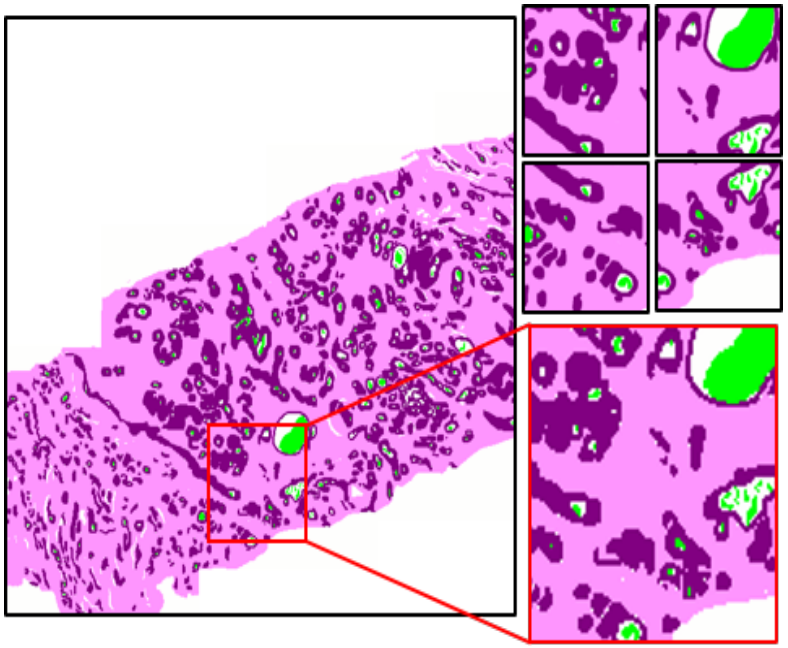
- Training Set: 30 ROIs
  - 25,992 patches of size 256x256 with augmentation
  - Split into training and validation set using 90:10 ratio
- Test Set: 28 ROIs
- Stochastic Gradient Descent for optimization
- Implemented in Torch
  - http://torch.ch/

Segmentation Results



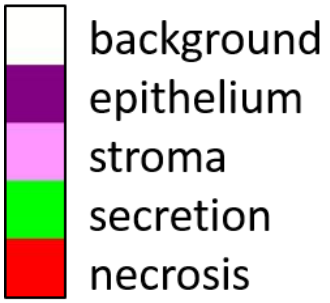


RGB Image

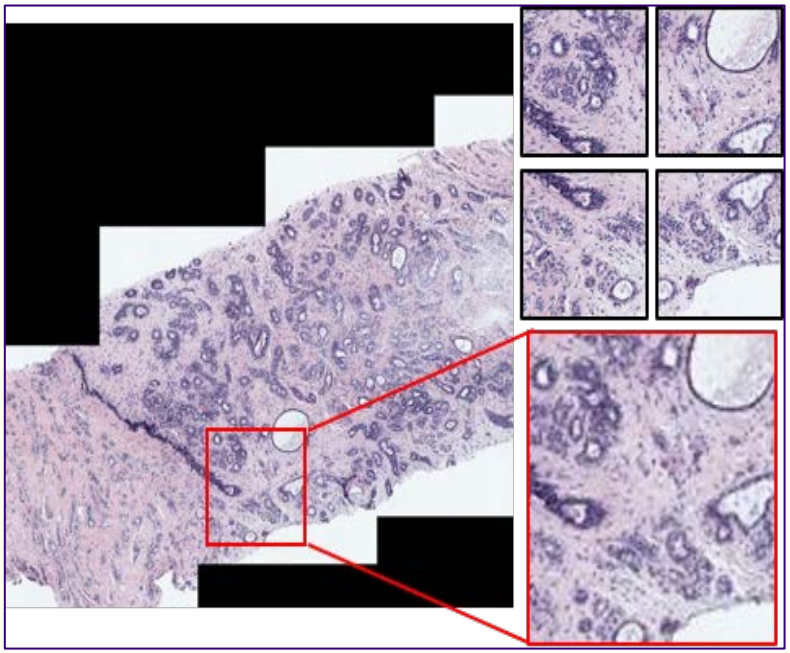


Ground Truth Label

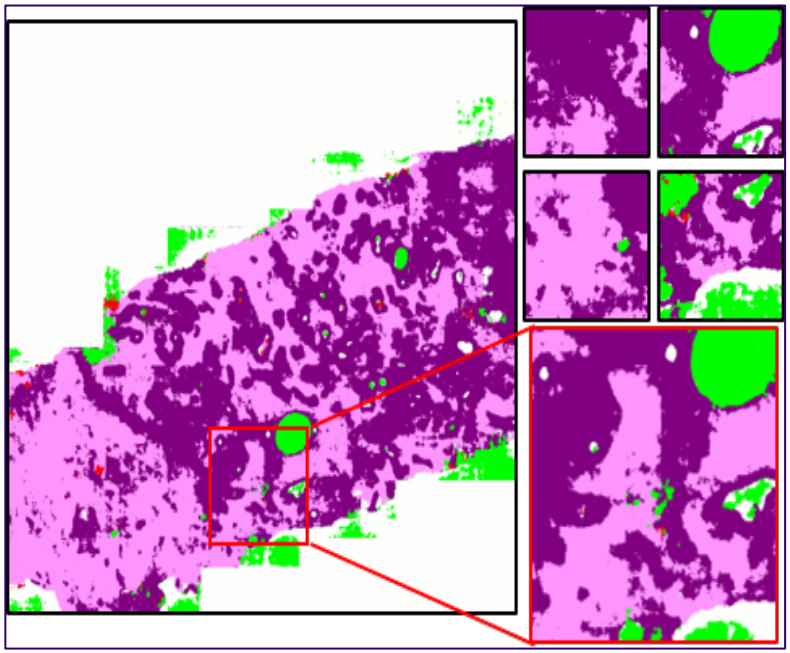
Segmentation Results



Encoder-Decoder Network with skip connection



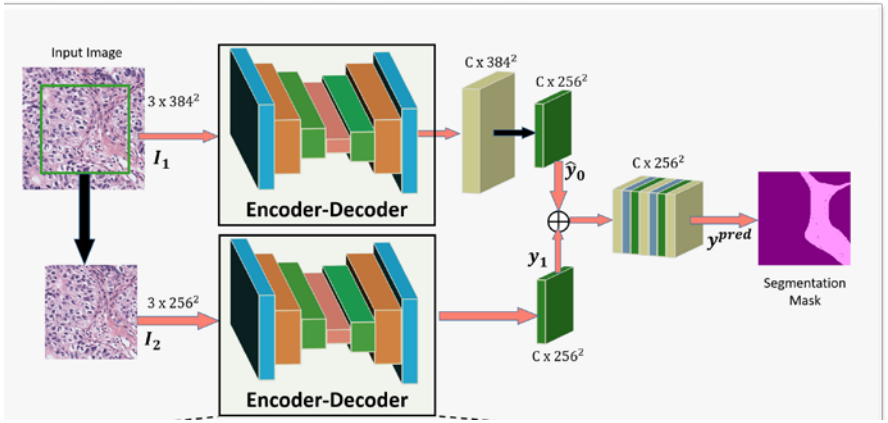
RGB Image

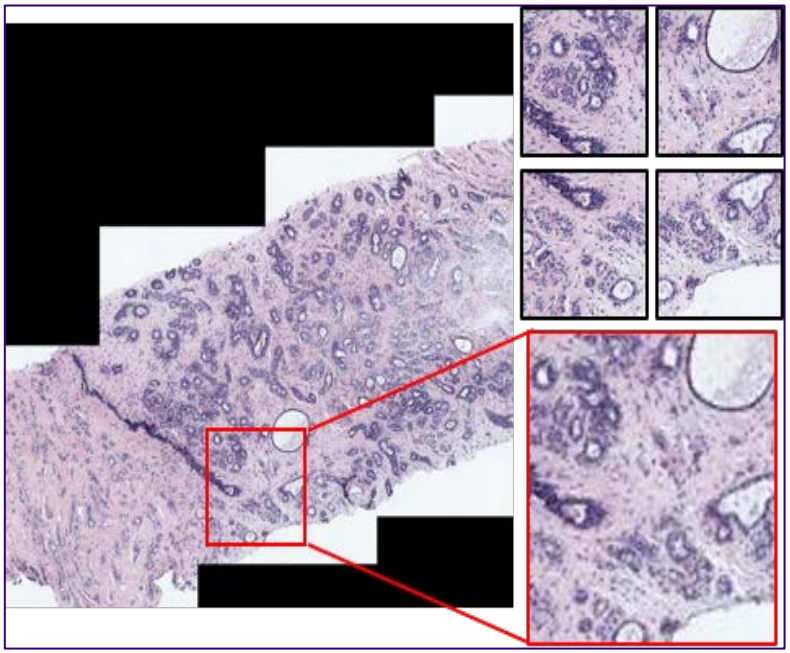


Prediction

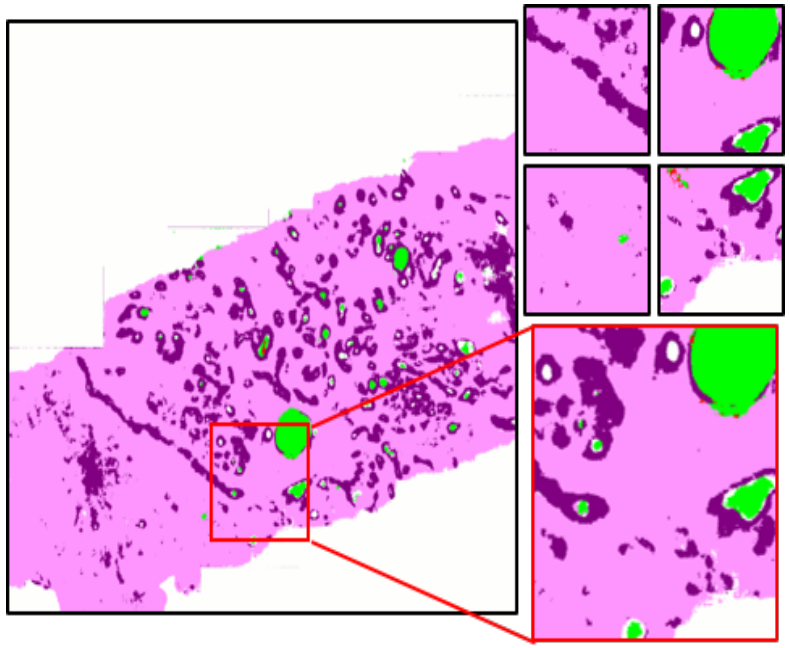
Segmentation Results

Multi-Resolution Encoder-Decoder Network



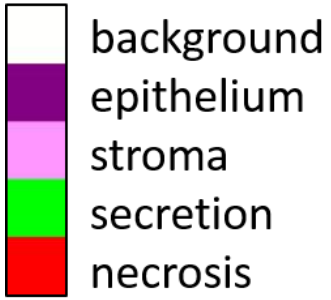


RGB Image

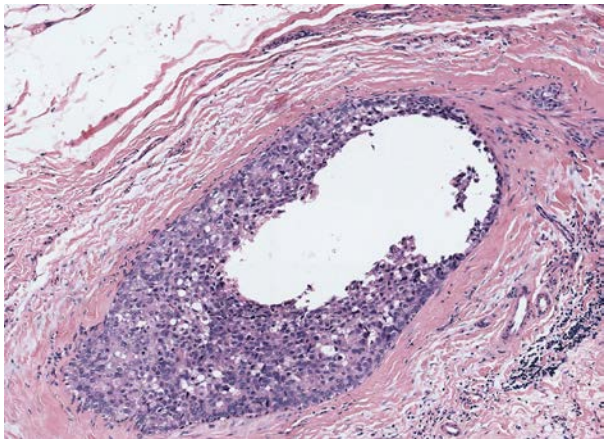


Prediction

Segmentation Results

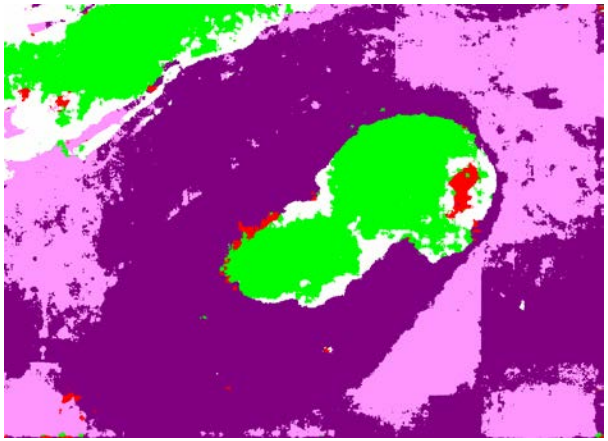


RGB Image

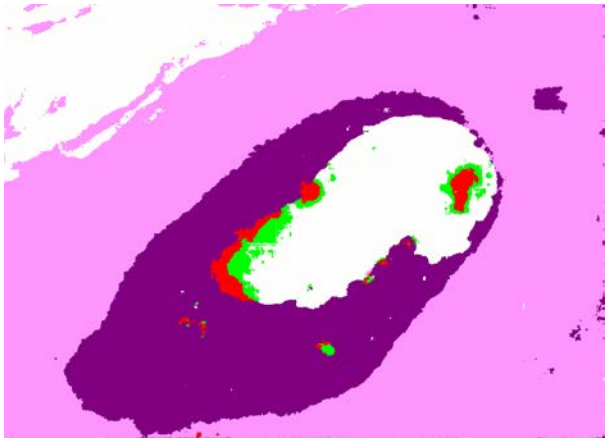


Ground Truth

Plain

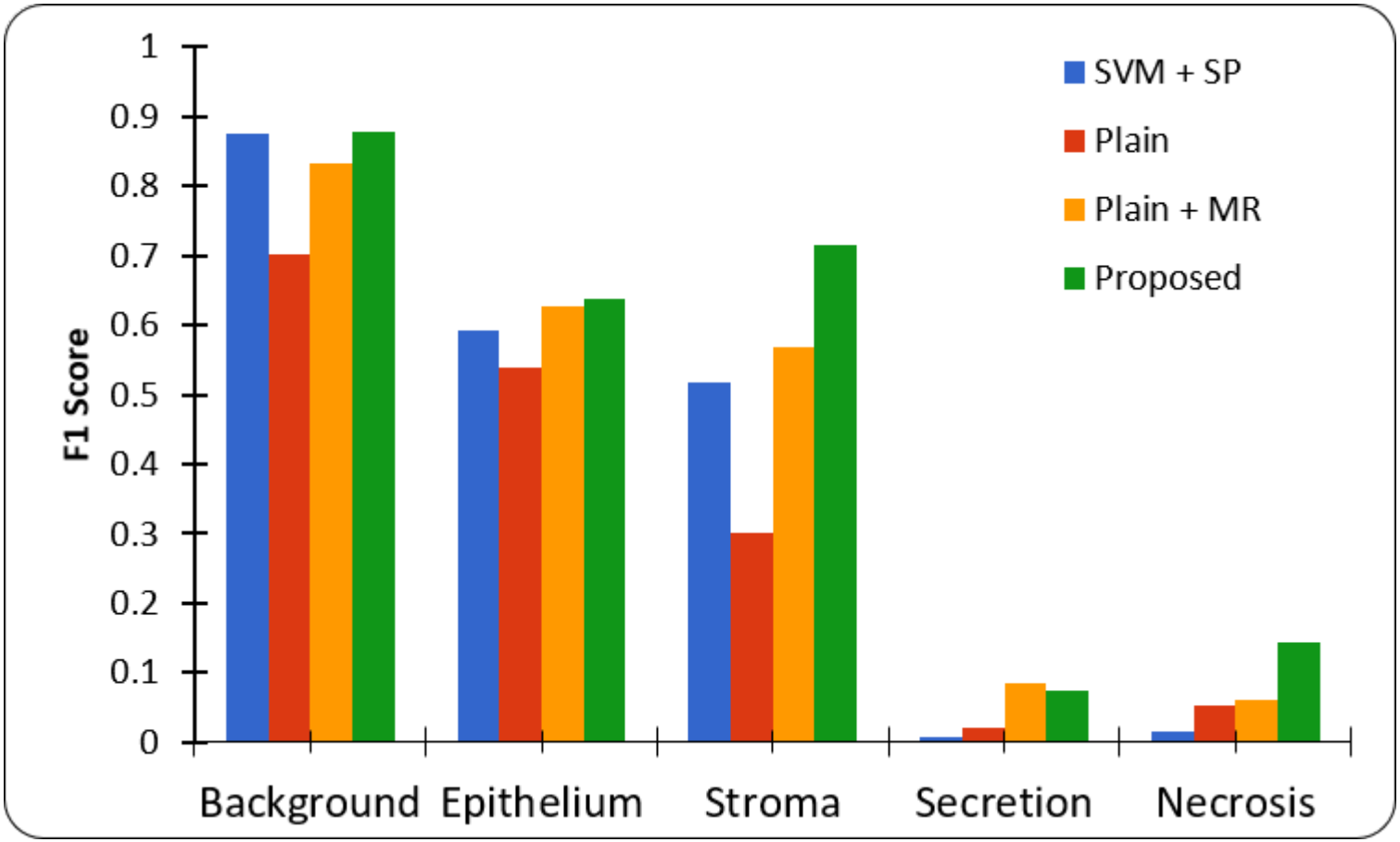


Multi-Resolution



Segmentation Results

F1-Score



Why Segmentation?

Results on Diagnosis

- Segmented whole dataset (428 ROIs) with the model trained on 30 ROIs
- Extracted histograms from segmentation masks and then trained different classifiers
- Weak classifiers are as good as strong classifiers

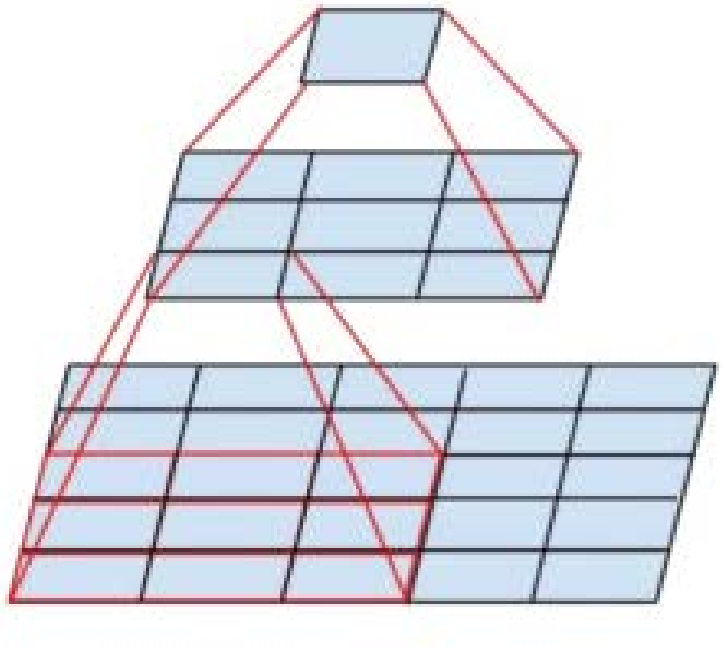
classification task	Multilayer Perceptron	SVM with RBF kernel	Logistic Regression	Random Forest
Invasive vs others	.72	.78	.75	.80
Benign vs others	.65	.54	.60	.62
Atypia vs DCIS	.75	.73	.77	.75

Thank You!!

References

1. Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *TPAMI*. 2016. (**FCN-8s**)
2. Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han. "Learning deconvolution network for semantic segmentation." *ICCV*. 2015. (**DeConvNet**)
3. V. Badrinarayanan; A. Kendall; R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation," TPAMI, 2017 (**SegNet**)
4. Yu, Fisher, and Vladlen Koltun. "Multi-scale context aggregation by dilated convolutions.“, ICLR, 2016 (**Dilation**)
5. Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." *arXiv preprint arXiv:1606.00915* (2016). (**DeepLab**)
6. Zheng, Shuai, et al. "Conditional random fields as recurrent neural networks." *ICCV*. 2015. (**CRFasRNN**)
7. Hariharan, Bharath, et al. "Hypercolumns for object segmentation and fine-grained localization." *CVPR*. 2015. (**HyperColumn**)

Two 3x3 filters are same as one 5x5 filter



Source: **Rethinking the Inception Architecture for Computer Vision** by Szegedy et al.