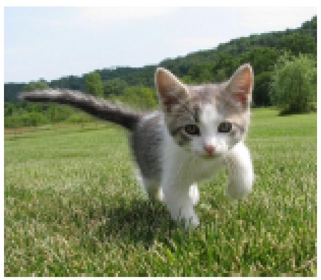
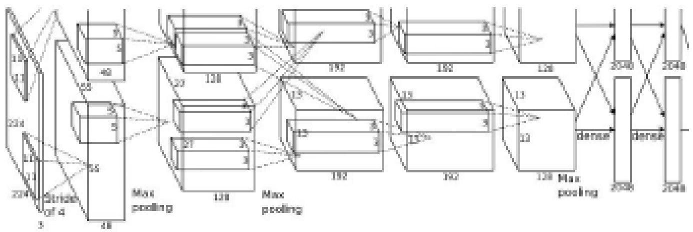
Semantic Segmentation

So far, Image Classification





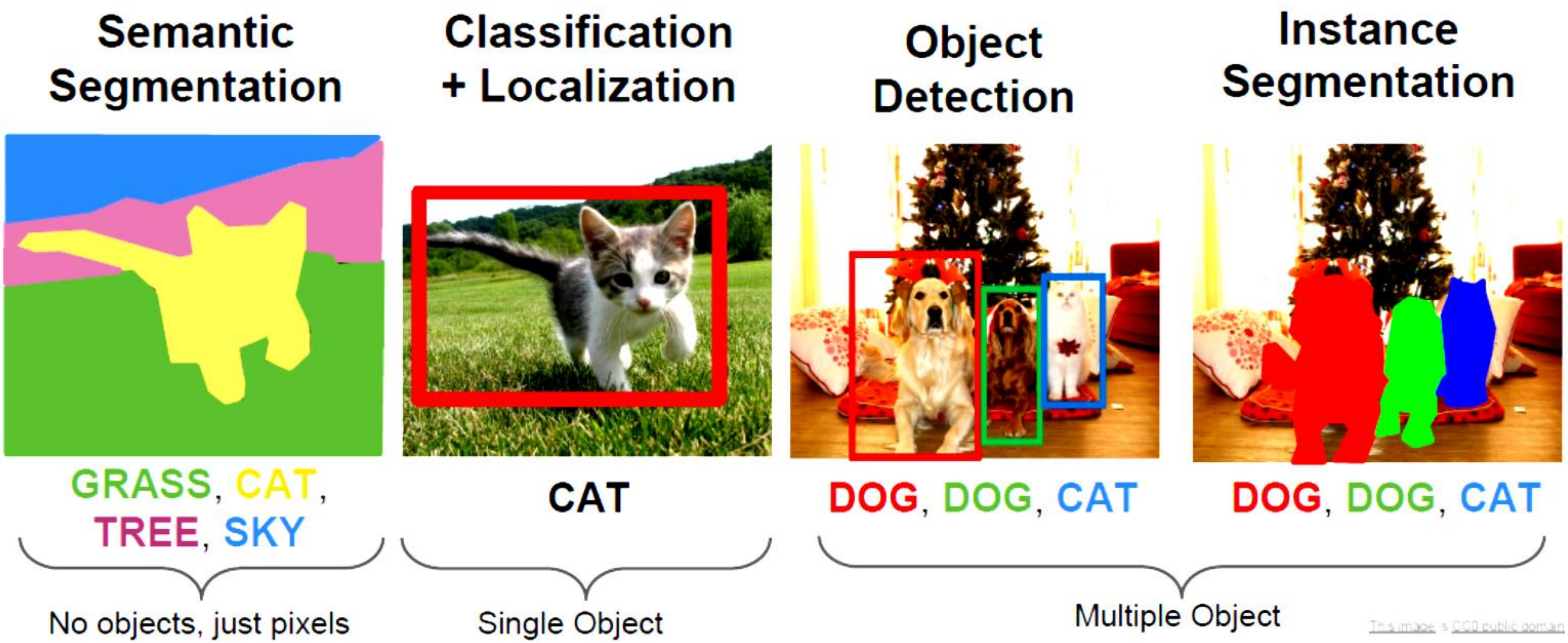
Vector:
4096

Fully-Connected:
4096 to 1000

Class Scores
Cat: 0.9
Dog: 0.05
Car: 0.01
...

Image Source:
http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Scenarios



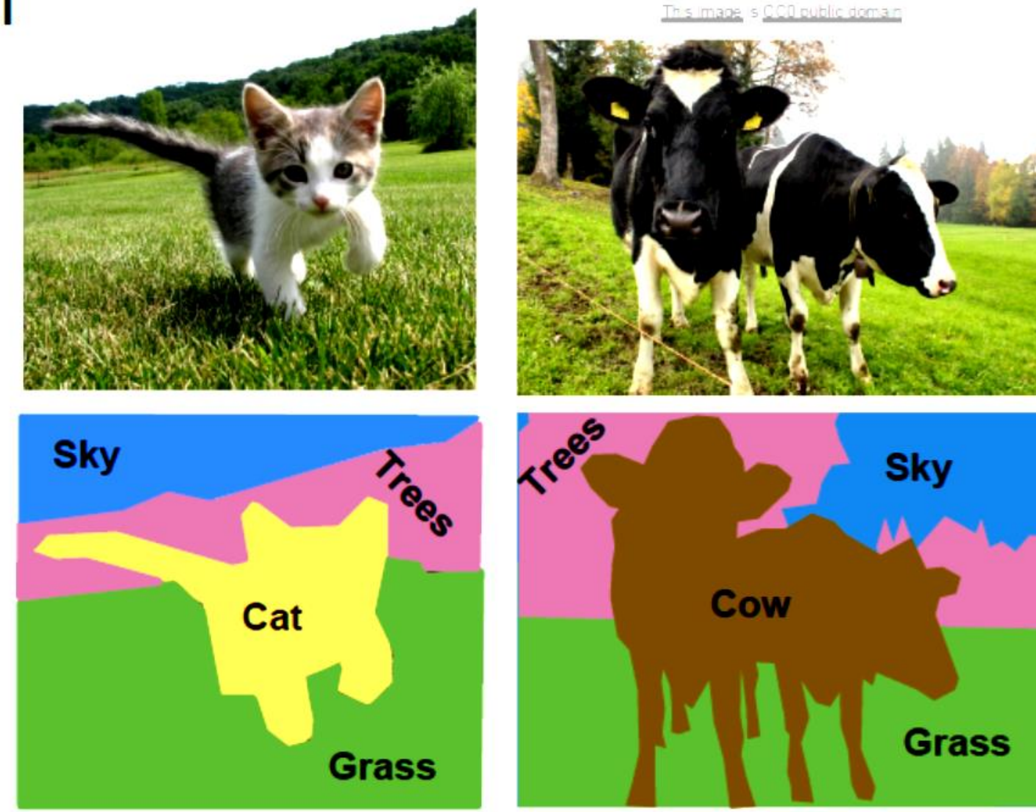
http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

Semantic Segmentation

Label each pixel in the image with a category label

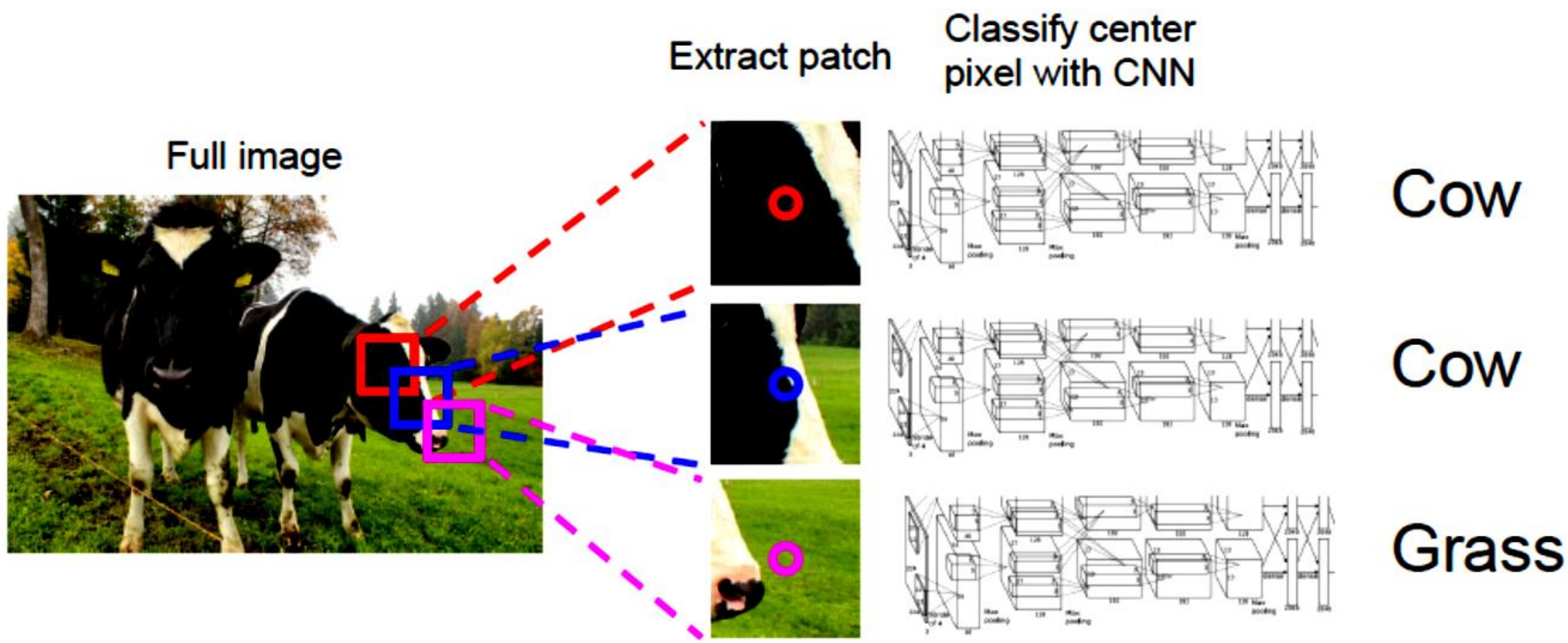
Don't differentiate instances, only care about pixels



http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

Semantic Segmentation Idea: Sliding Window

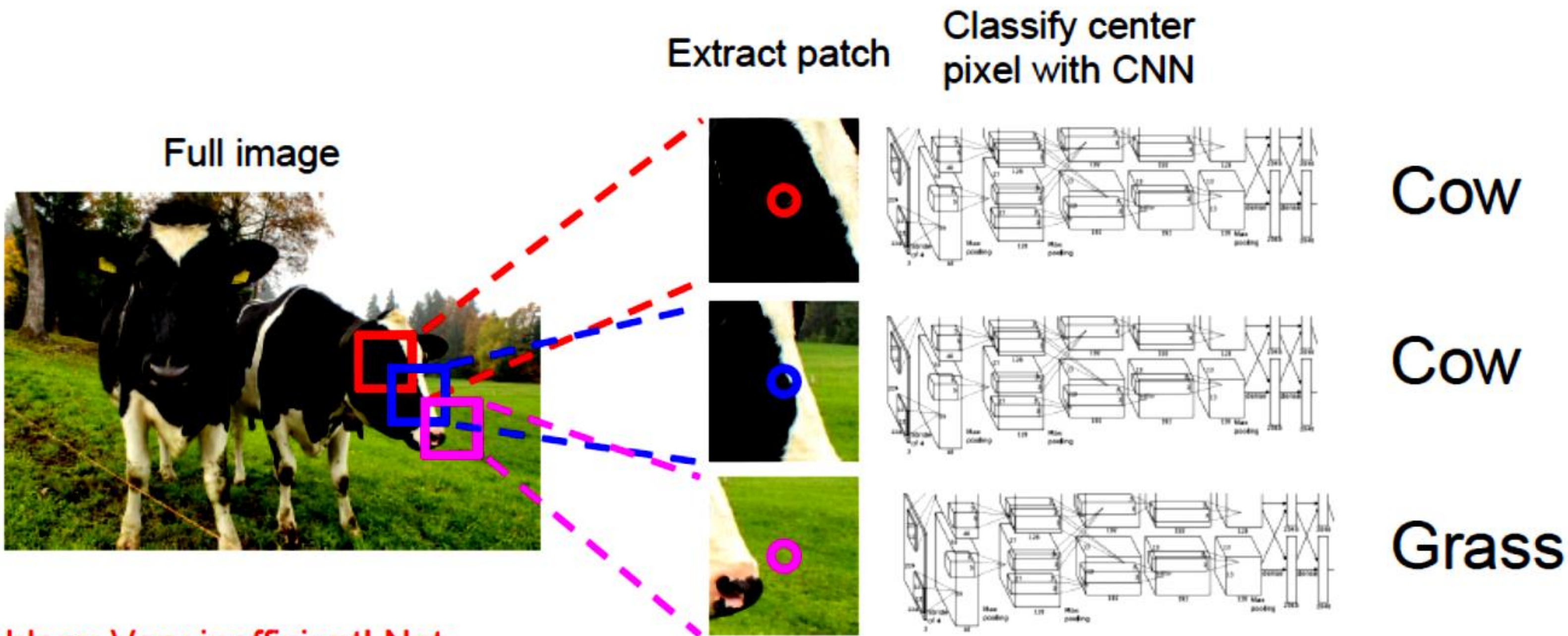


Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

Semantic Segmentation Idea: Sliding Window



Problem: Very inefficient! Not reusing shared features between overlapping patches

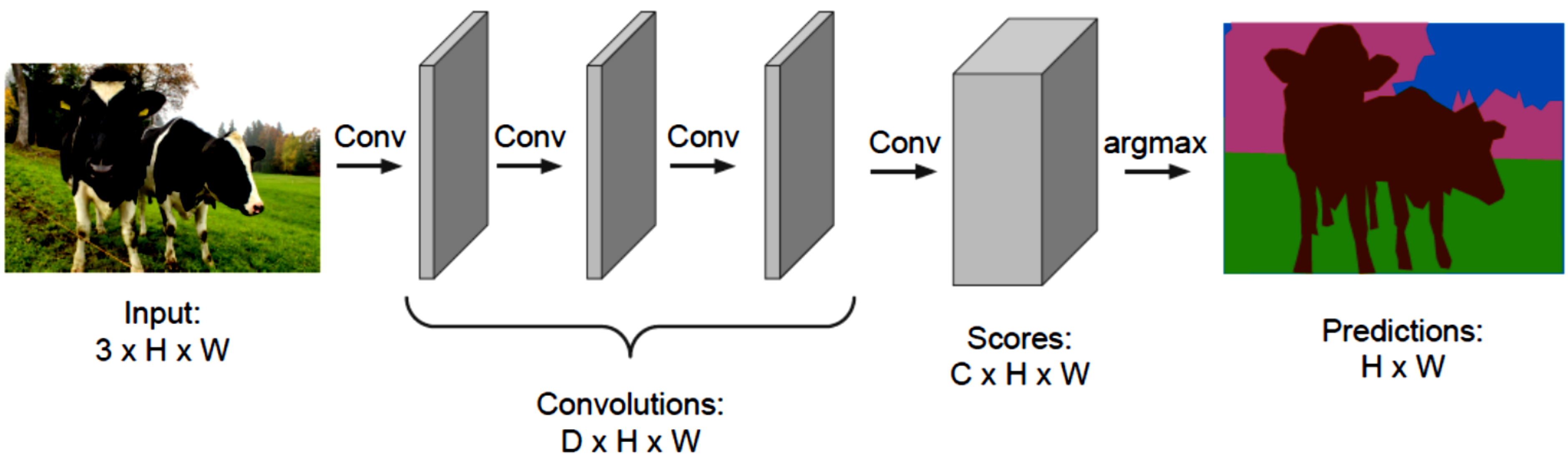
Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

Semantic Segmentation Idea: Fully Convolutional

Design a network as a bunch of convolutional layers
to make predictions for pixels all at once!

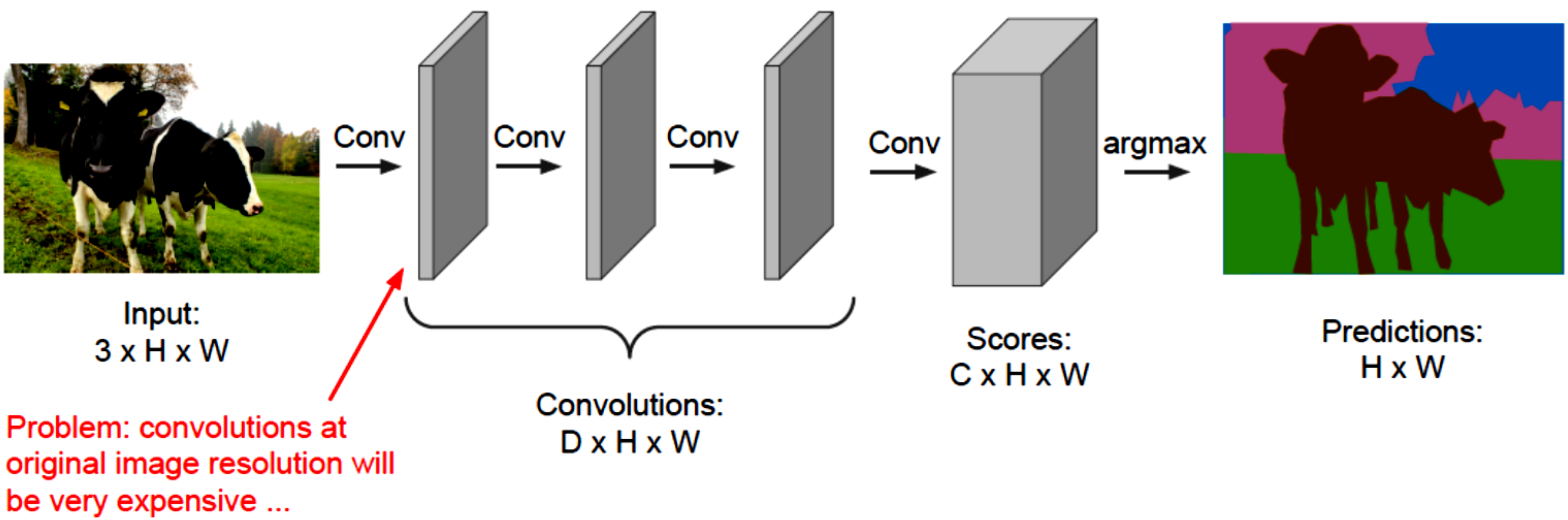


http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

Semantic Segmentation Idea: Fully Convolutional

Design a network as a bunch of convolutional layers
to make predictions for pixels all at once!

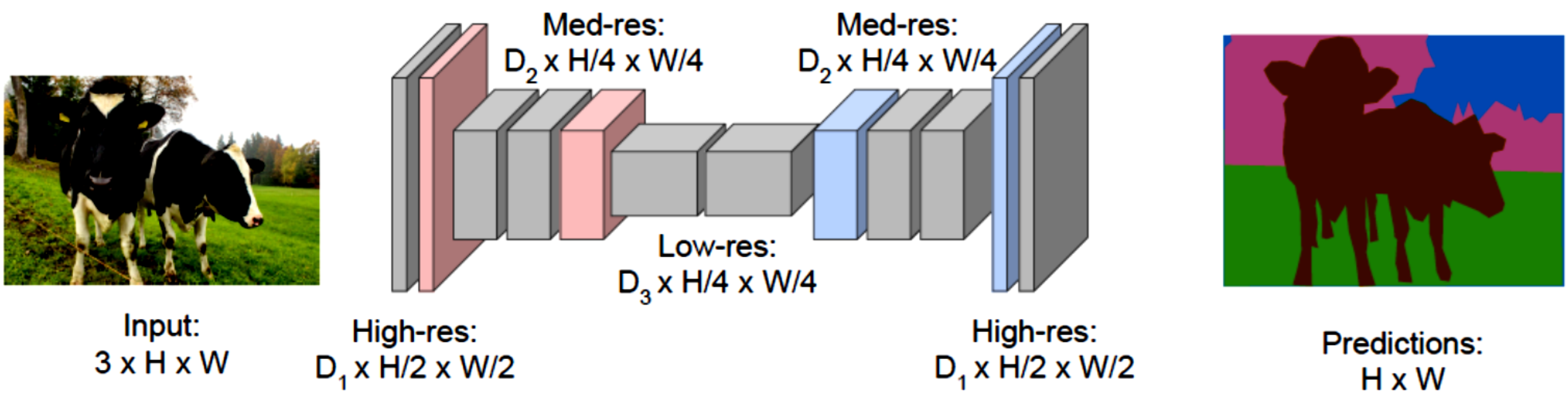


http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

Semantic Segmentation Idea: Fully Convolutional

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!



Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

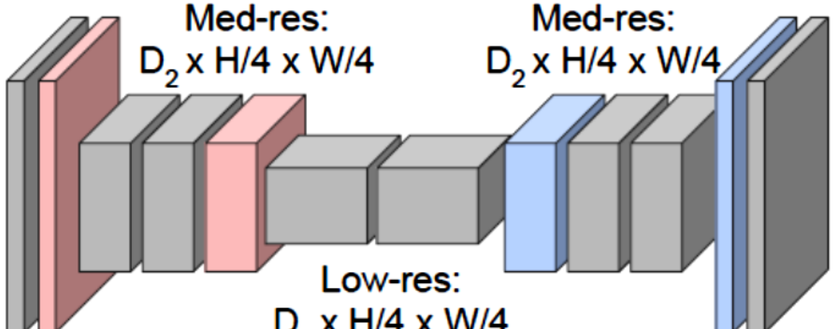
Semantic Segmentation Idea: Fully Convolutional

Downsampling: Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling: ???

Input: 3 x H x W



High-res: D_1 x H/2 x W/2

Med-res: D_2 x H/4 x W/4

Low-res: D_3 x H/4 x W/4

Med-res: D_2 x H/4 x W/4

High-res: D_1 x H/2 x W/2

Predictions: H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

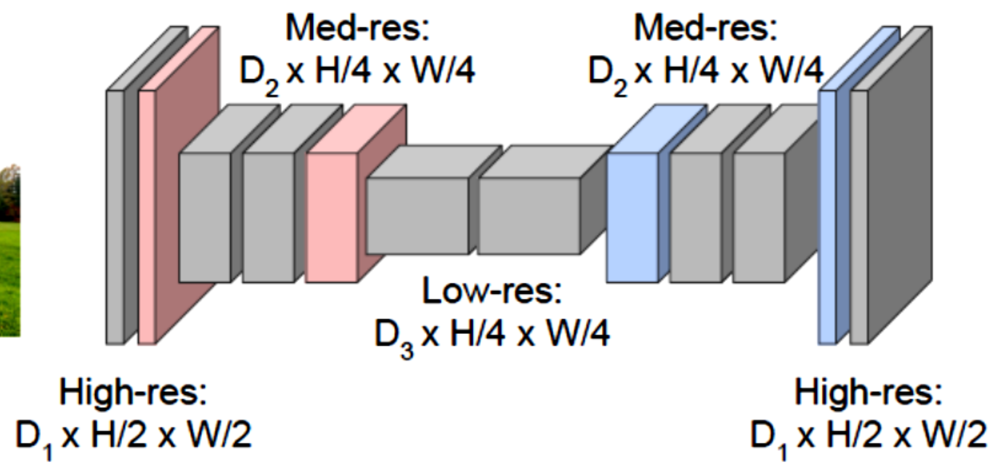
Semantic Segmentation Idea: Fully Convolutional

Downsampling: Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling: Unpooling or strided transpose convolution

Input:
3 x H x W



Predictions:
H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture11.pdf

Semantic Segmentation

UNET [1]

Objective:

- **There is large consent that successful training of deep networks requires many thousand annotated training samples.**

UNET [1]

Objective:

- **There is large consent that successful training of deep networks requires many thousand annotated training samples.**

- In UNET paper, authors presented a network and training strategy that relied on the strong **use** of **data augmentation to use the available annotated samples more efficiently**.

UNET [1]

Objective:

- **There is large consent that successful training of deep networks requires many thousand annotated training samples.**

- In UNET paper, authors presented a network and training strategy that relied on the strong **use** of **data augmentation to use the available annotated samples more efficiently**.

- The architecture consisted of a contracting path to capture context and a symmetric expanding path that enabled precise localization.

UNET [1]

Objective:

- **There is large consent that successful training of deep networks requires many thousand annotated training samples.**

- In UNET paper, authors presented a network and training strategy that relied on the strong **use** of **data augmentation to use the available annotated samples more efficiently**.

- The architecture consisted of a contracting path to capture context and a symmetric expanding path that enabled precise localization.

- **They showed that such a network could be trained end-to-end from very few images** and outperformed the prior best method on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks.

UNET [1]

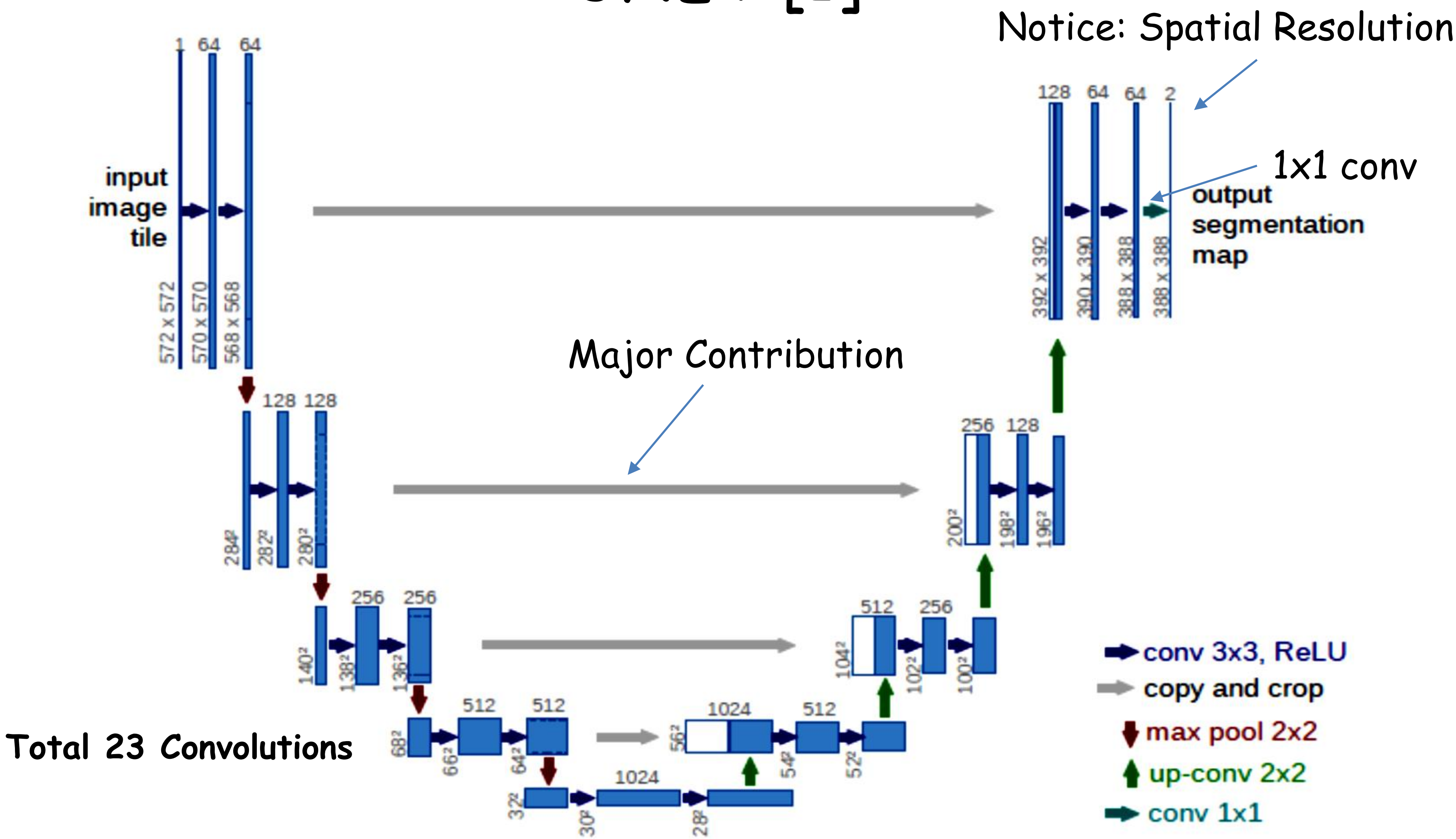


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

UNET [1]

To allow a seamless tiling of the output segmentation map (see Figure 2), it is important to select the input tile size such that all 2x2 max-pooling operations are applied to a layer with an even x- and y-size.

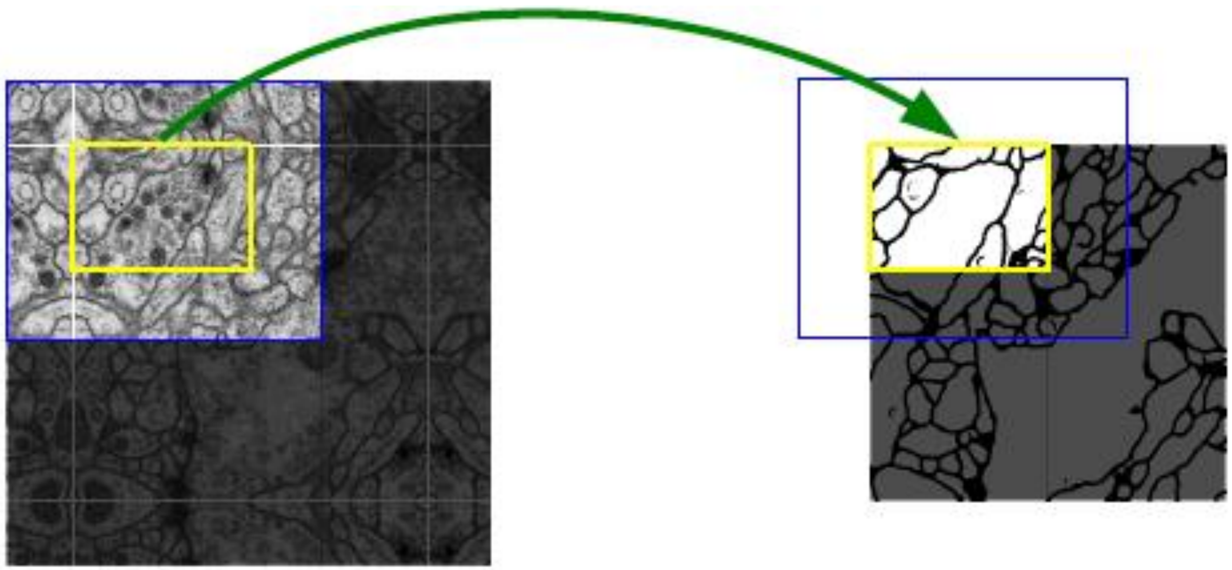


Fig. 2. Overlap-tile strategy for seamless segmentation of arbitrary large images (here segmentation of neuronal structures in EM stacks). Prediction of the segmentation in the yellow area, requires image data within the blue area as input. Missing input data is extrapolated by mirroring

UNET [1]

Major Contributions:
In order to localize, high resolution features from the contracting path are combined with the upsampled output. A successive convolution layer can then learn to assemble a more precise output based on this information.

UNET [1]

Major Contributions:

As for their tasks there is very little training data available, they used excessive data augmentation by applying elastic deformations to the available training images.

UNET [1]

Major Contributions:

As for their tasks there is very little training data available, they used excessive data augmentation by applying elastic deformations to the available training images.

This allowed the network to learn invariance to such deformations, without the need to see these transformations in the annotated image corpus.

UNET [1]

Major Contributions:

As for their tasks there is very little training data available, they used excessive data augmentation by applying elastic deformations to the available training images.

This allowed the network to learn invariance to such deformations, without the need to see these transformations in the annotated image corpus.

This is particularly important in biomedical segmentation, since deformation used to be the most common variation in tissue and realistic deformations can be simulated efficiently.

UNET – Training [1]

- The input images and their corresponding segmentation maps were used to train the network with the **stochastic gradient descent with momentum.**

- Due to the unpadded convolutions, the output image was smaller than the input by a constant border width.

- They used a high **momentum (0.99)** such that a large number of the previously seen training samples determine the update in the current optimization step.

Intersection over Union

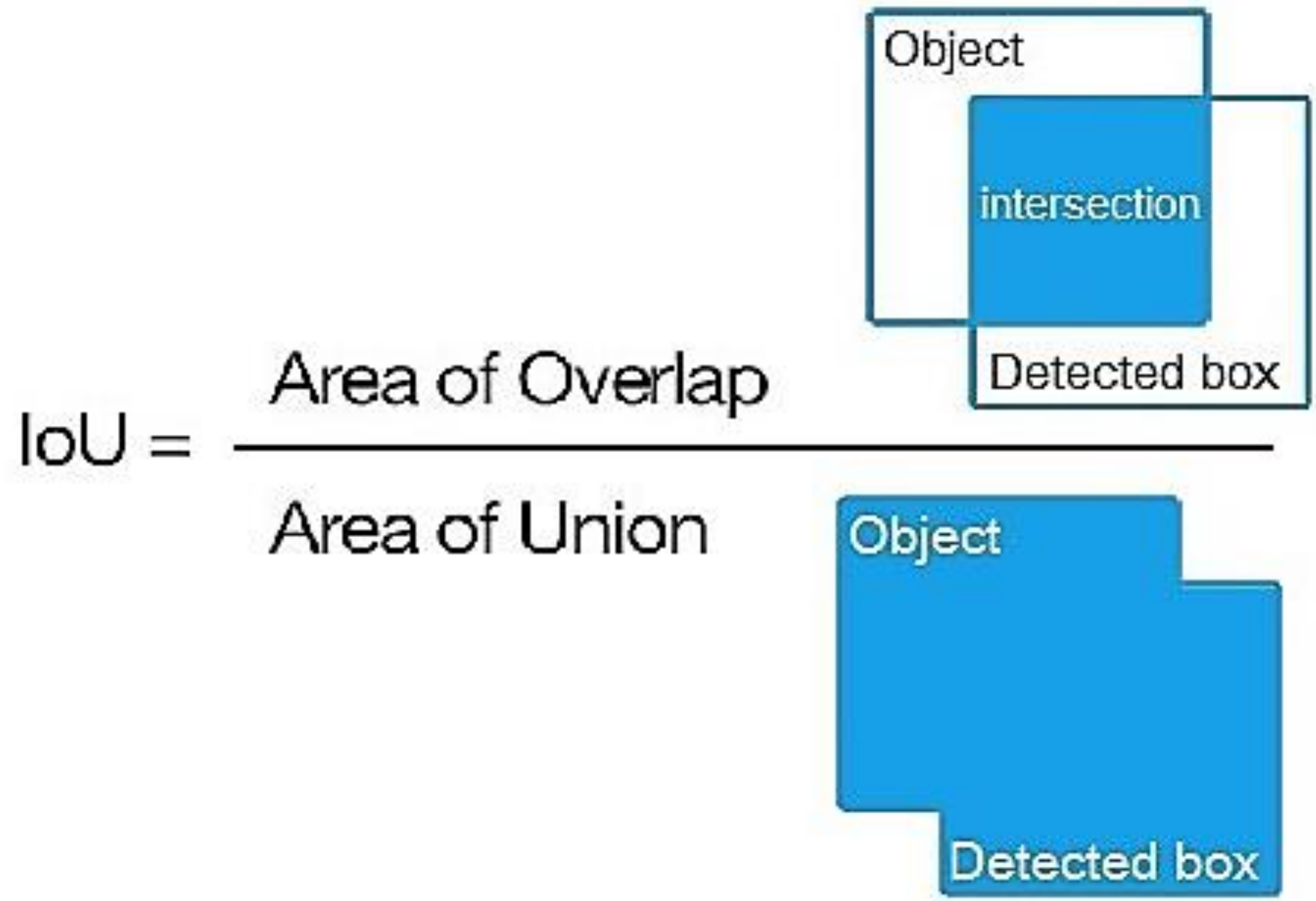


Image Source: https://i.stack.imgur.com/JlHnn.jpg

Intersection over Union

$$IoU = \frac{\text{area of overlap}}{\text{area of union}}$$

Image Source: https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173

UNET – Results [1]

Table 2. Segmentation results (IOU) on the ISBI cell tracking challenge 2015.

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

References

1. https://arxiv.org/pdf/1505.04597.pdf

Disclaimer

- These slides are not original and have been prepared from various sources for teaching purpose.