Semantic Segmentation

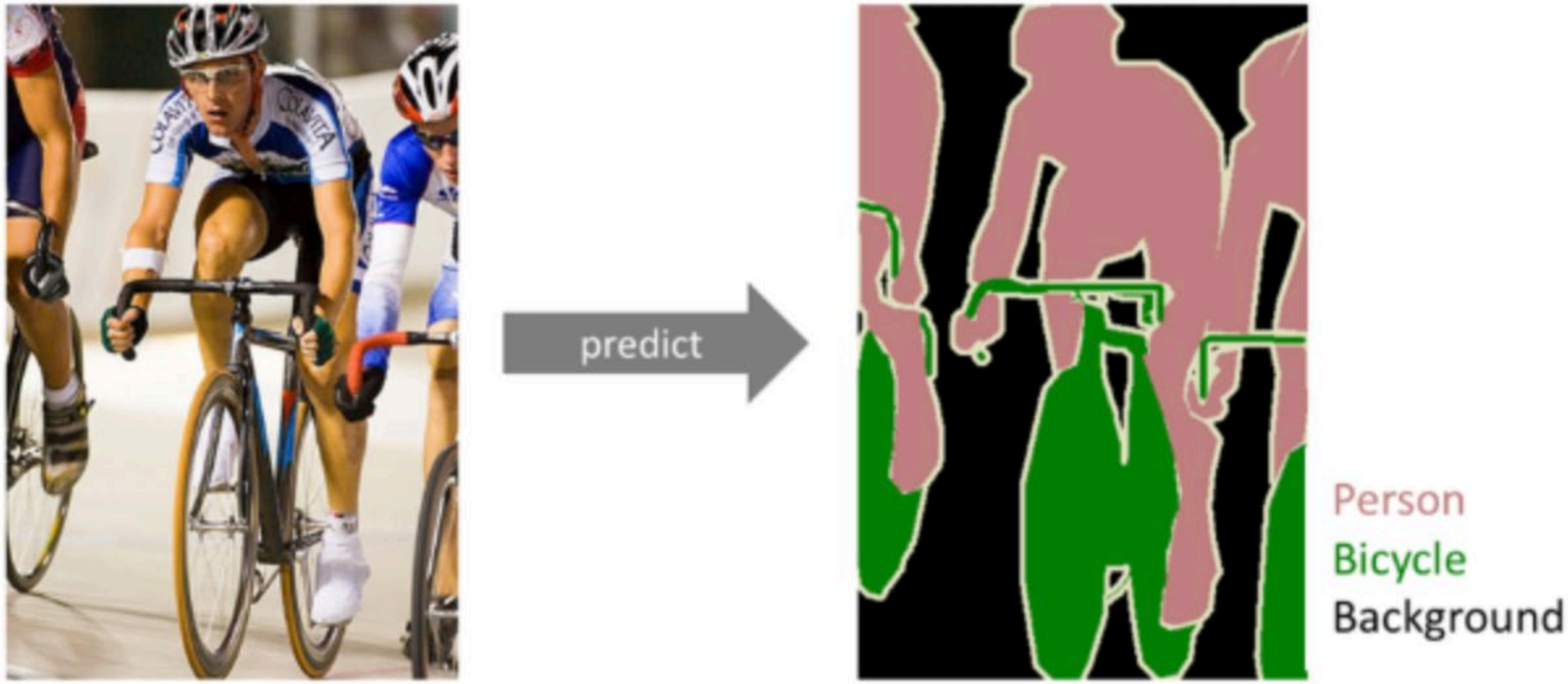
POSTECH MIP Lab.

TA: Jaeyoon Sim, Hayoung Ahn, Sungwoo Hur

Overview

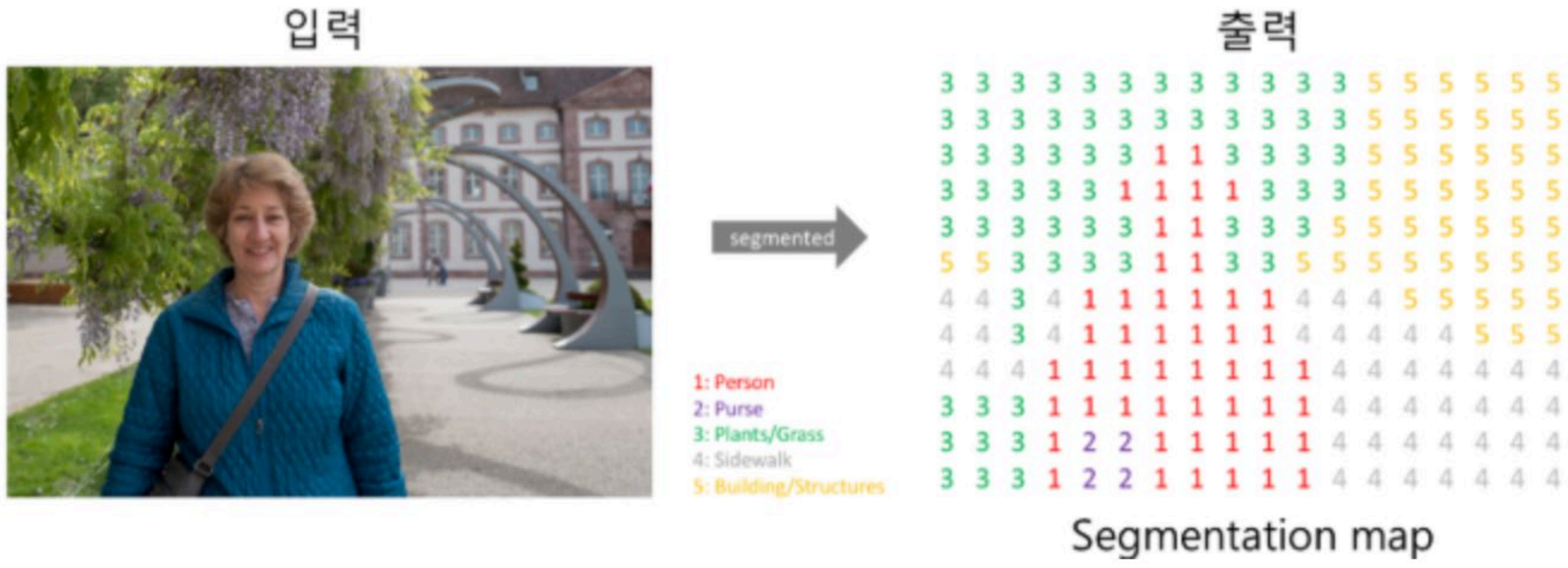
Semantic Segmentation

- A task to classify **segments with same semantic meanings /information**.
- A task to **classify each pixel** in the object.



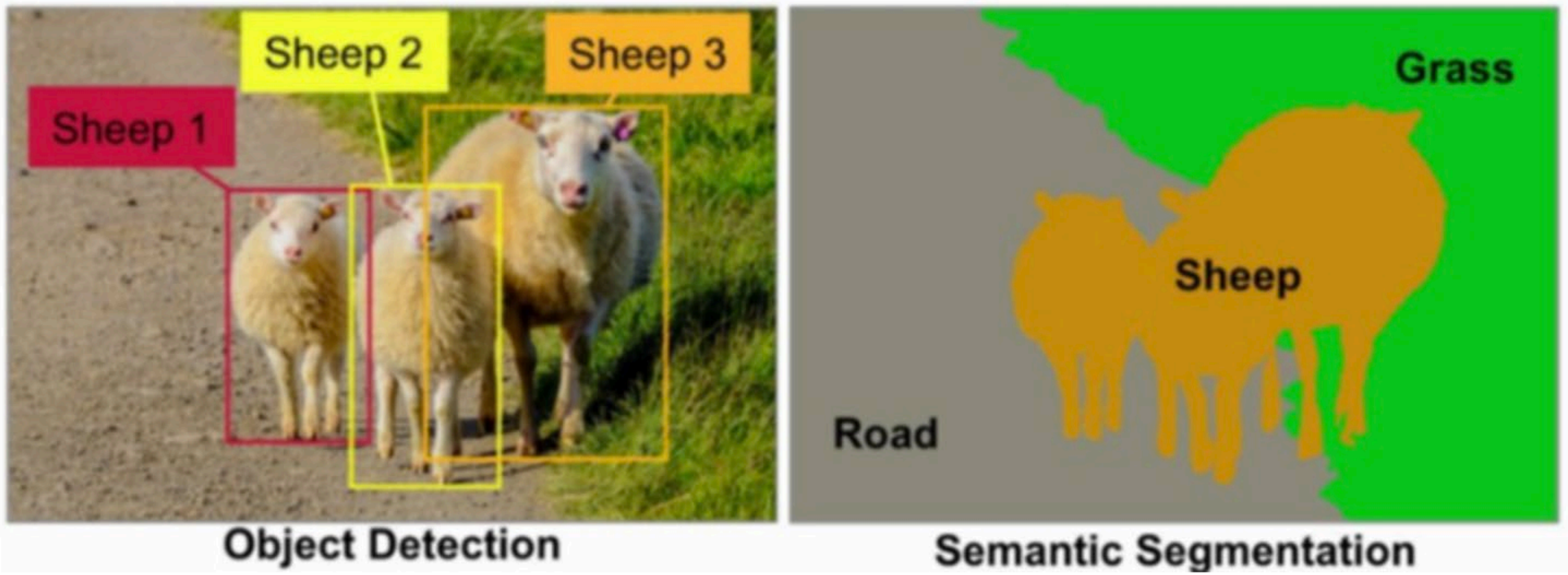
Overview

- Semantic Segmentation → Pixel-level Classification



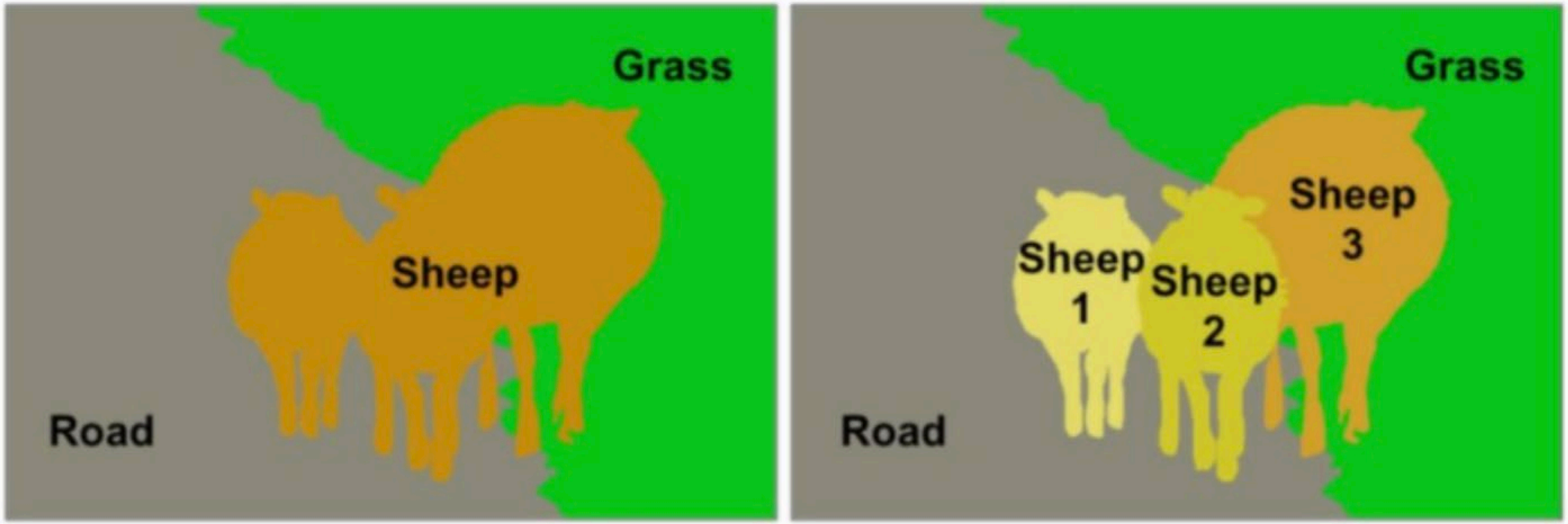
Overview

- Object Detection vs. Semantic Segmentation



Overview

- Semantic Segmentation vs. Instance Segmentation



Overview

- Classification → Detection → Semantic segmentation

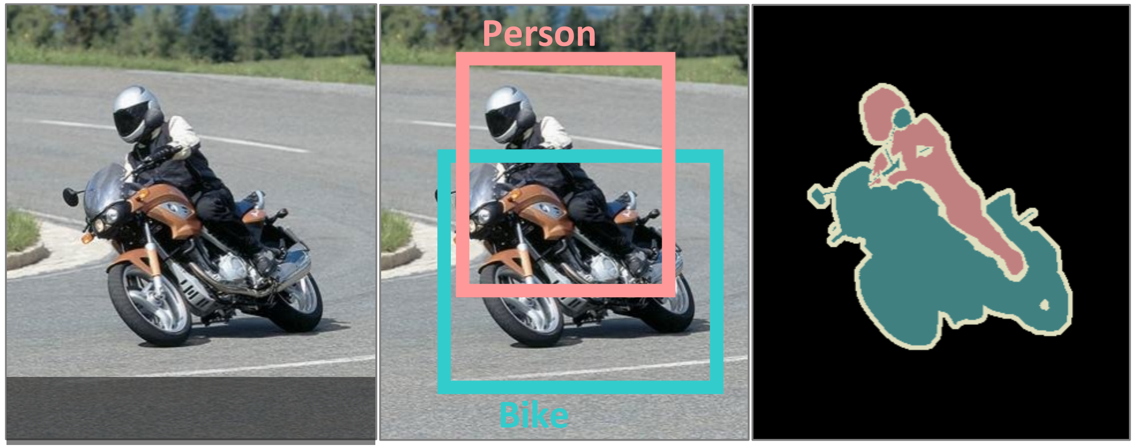


Image Classification | Object Detection | Semantic Segmentation

Higher supervision
Expensive labeling

- Semantic segmentation based on deep learning
  - FCN, DeepLab, DeconvNet, Pyramid Scene Parsing Network, ...

Image vs Semantic

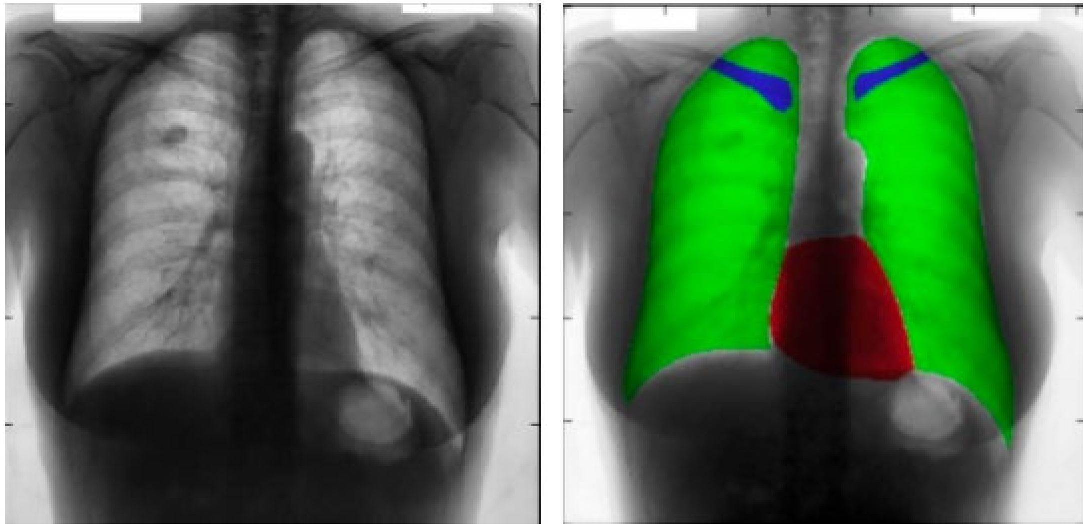
- **Classification** - determine label of image
  - Find function: Image → number of label
  - e.g. 32x32x3 → 10x1

- **Semantic segmentation** - determine label of each pixel
  - Find function: Image → number of label x Image width x Image height
  - e.g. 32x32x3 → 10x32x32, harder :<
  - But maybe not 32x32 times harder problem because locality :>

- What is the difference of two task?

Applications

- Medical images
- Autonomous driving
- Computational photography
- ...

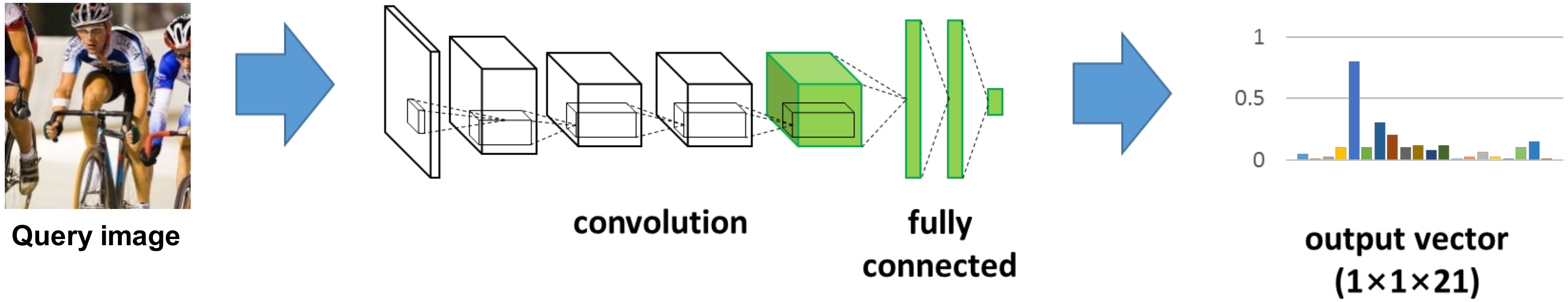


Fully Convolutional Network

Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

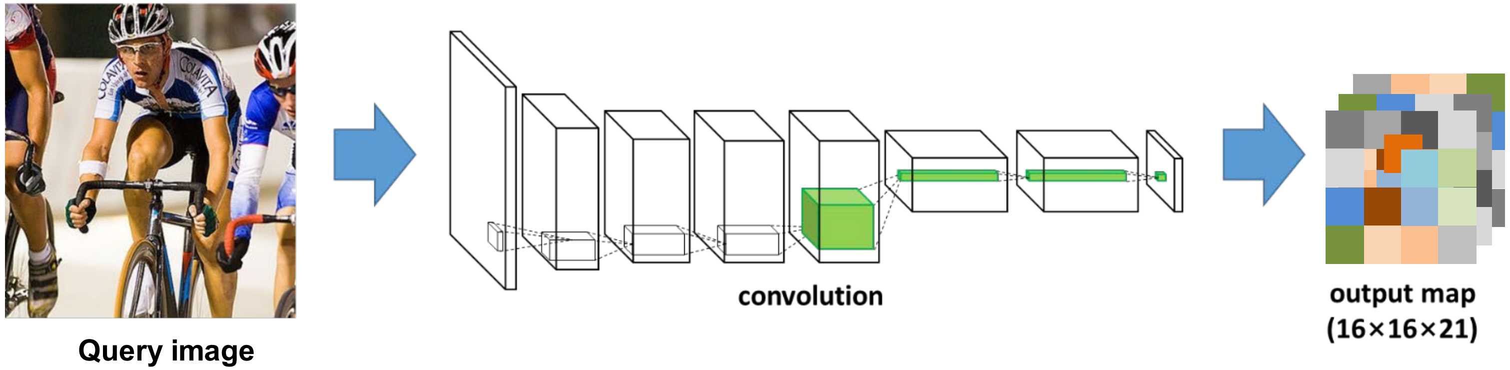
Fully Convolutional Network

Image classification



Semantic segmentation

- Given an input image, obtain pixel-wise segmentation mask
- using a deep Convolutional Neural Network (CNN)



Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

- 기존 classification model은 분류를 위해 마지막에 항상 FC layer를 붙인다.
- FC layer는 segmentation에는 적합하지 않다.
- 고정된 사이즈의 image만 받을 수 있다.
- FC layer를 거치고 나면 2차원 위치 정보가 사라진다.
- 이는 Pixel-wise classification을 하는 segmentation task에 치명적인 문제이다.

Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Fully Convolutional Network (FCN)
- 마지막 FC layer들을 모두 convolutional layer로 대체

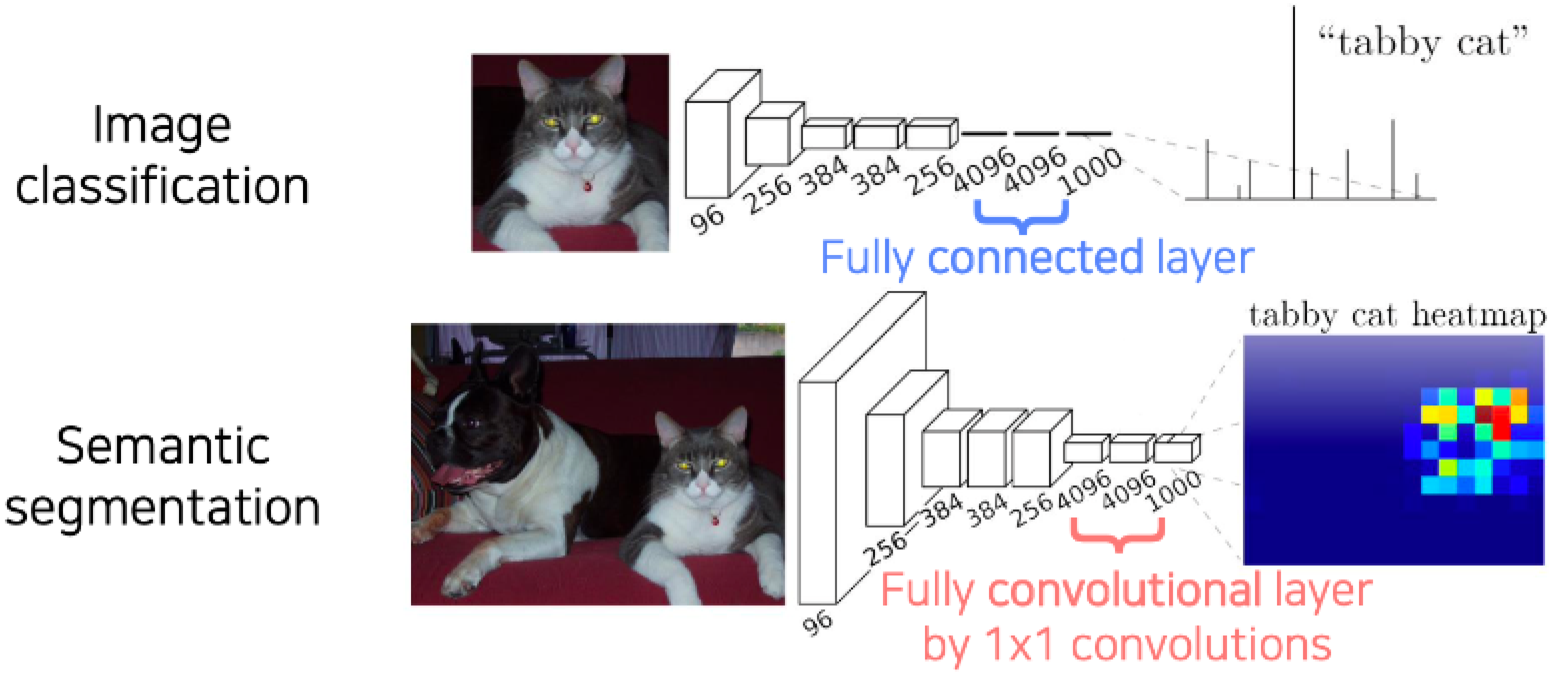
장점
- 2차원 위치 정보를 유지
- FC layer를 쓰지 않기 때문에 어떠한 input이 오더라도 모델이 수용 가능

Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Fully connected vs. Fully convolutional

- Fully connected layer: Output a fixed dimensional vector and discard spatial coordinates
- Fully convolutional layer: Output a classification map which has spatial coordinates

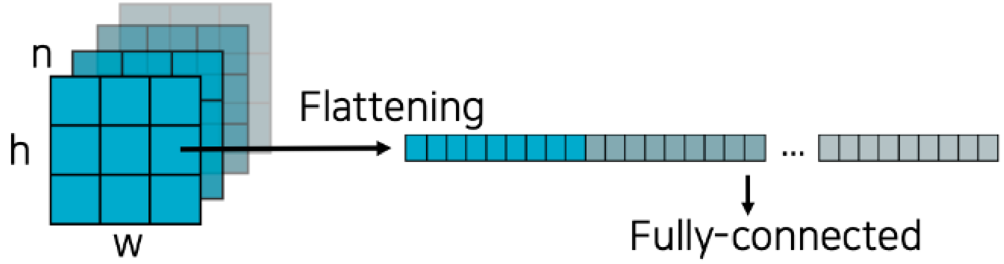


Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Interpreting fully connected layers as 1x1 convolutions

- A fully connected layer classifies a single vector
- A 1x1 convolution layer classifies every feature vector of the convolutional feature map

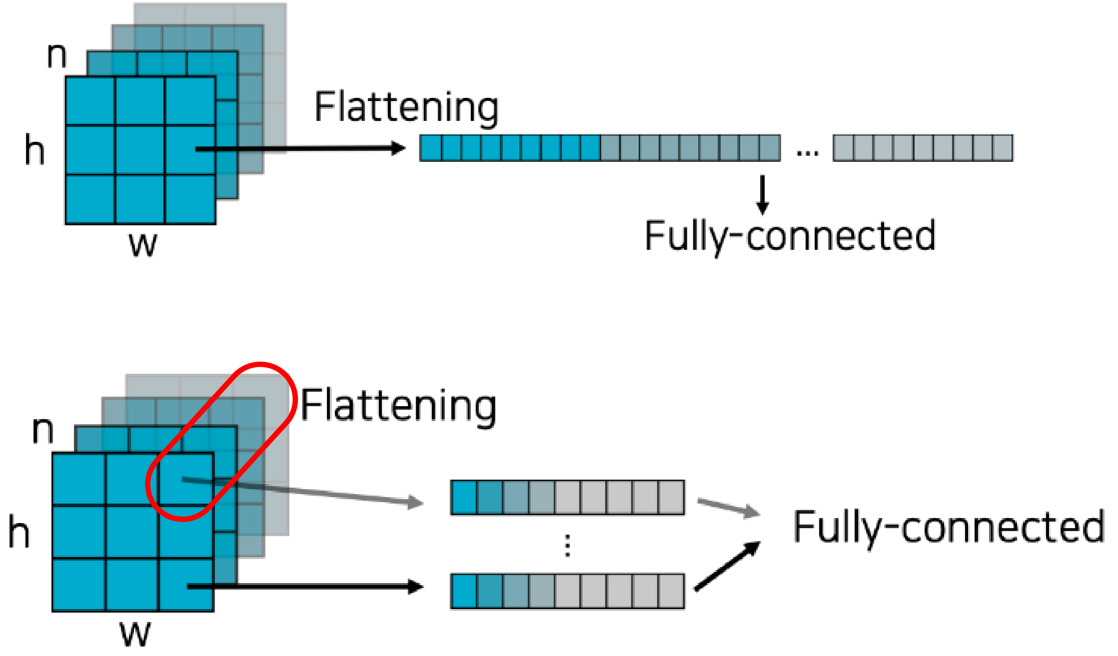


Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Interpreting fully connected layers as 1x1 convolutions

- A fully connected layer classifies a single vector
- A 1x1 convolution layer classifies every feature vector of the convolutional feature map

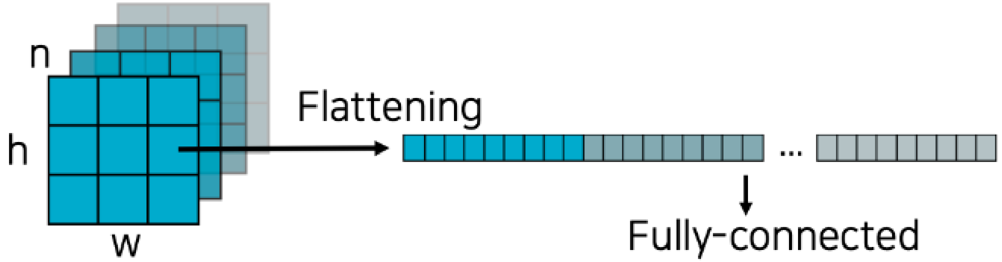


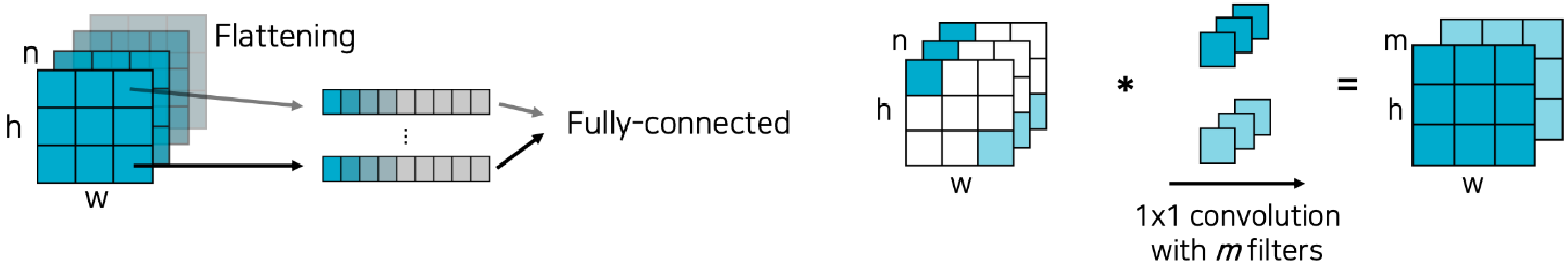
Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Interpreting fully connected layers as 1x1 convolutions

- A fully connected layer classifies a single vector
- A 1x1 convolution layer classifies every feature vector of the convolutional feature map





Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

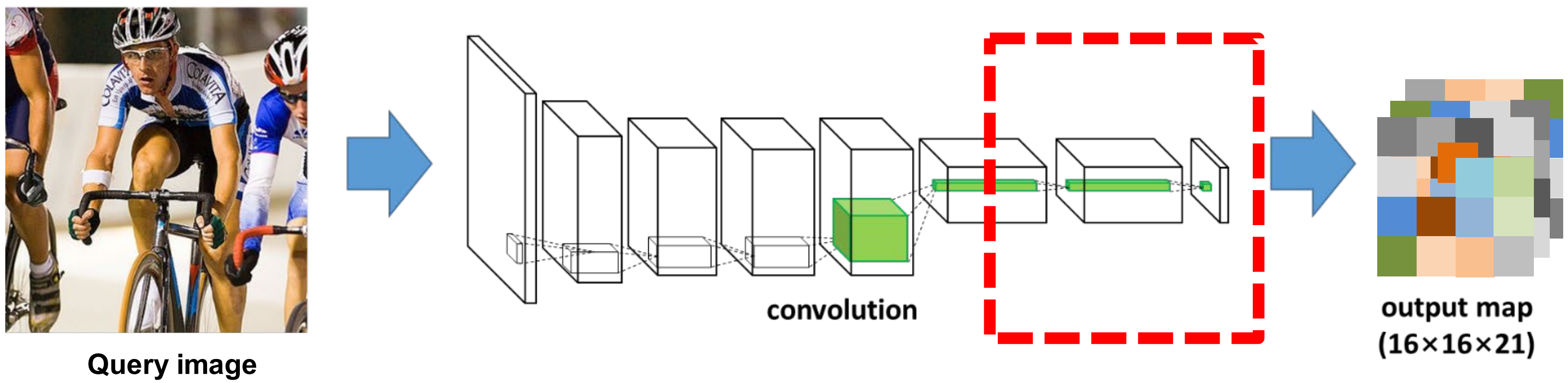
Fully Convolutional Network

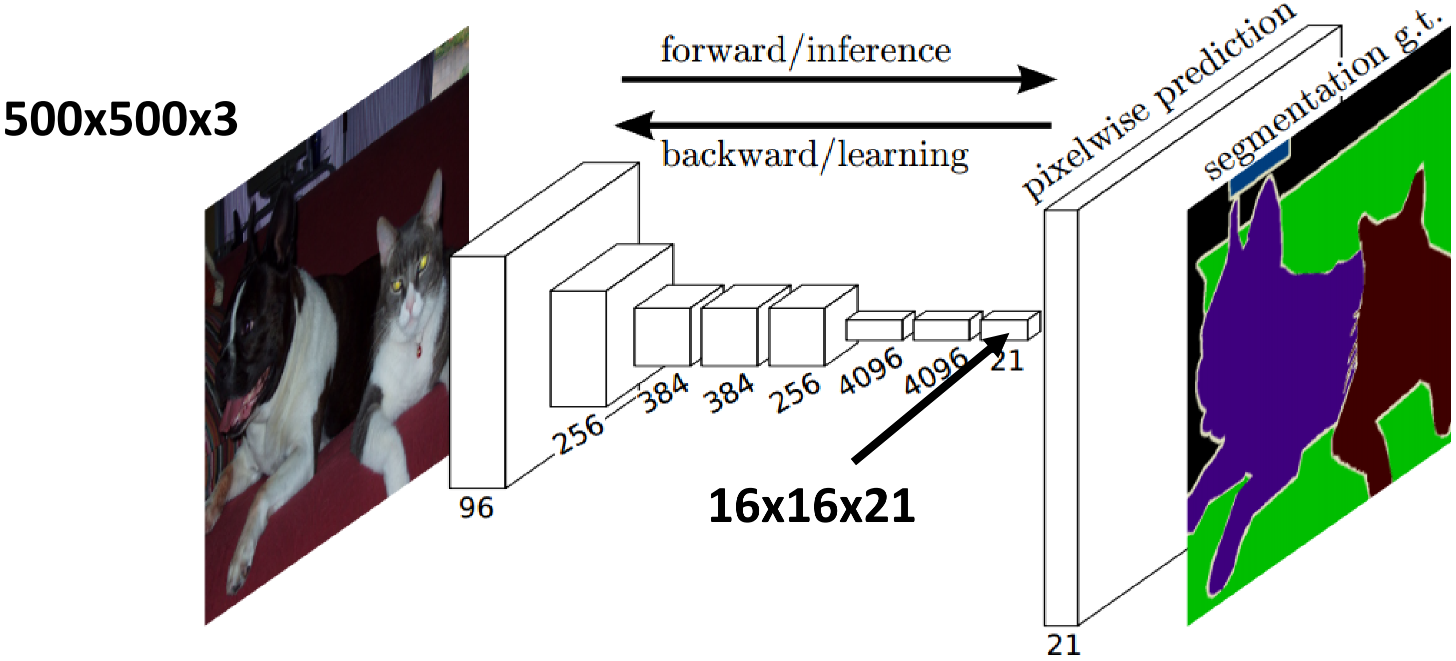
- Convolution을 통과한 마지막 feature 맵은 H * W * Class size를 가지도록 한다.
- 즉, 각 channel이 하나의 클래스에 대한 정보를 가지고 있는 것.
- 하지만 마지막 feature map은 conv와 pooling 연산을 거치면서 spatial dimension이 input에 비해 작아져 있음.
- 이것을 다시 input size에 맞게 키워주는 것이 필요.

Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

- Recall:





ic segmentation, *CVPR* 2015.

Fully Convolutional Network

애초에 Encoding 부분에서 안 줄여주면 되지 않나요?
ex) Apply padding, No pooling, …

Pooling을 하지 않거나 pooling의 stride를 줄임으로써 Feature map의 크기가 작아지는 것을 처음부터 피할 수 있음.

- 이 경우 receptive field가 줄어들어 이미지의 context를 놓치게 됨.
- Pooling이 없으면 학습 파라미터 수가 급격히 증가, 연산이 많아짐, 메모리 사용량 증가

→ 따라서 coarse feature map을 dense map으로 upsampling하는 방법 고려!

Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Feature map 사이즈를 키워주기 위한 구조 제안 (Upsampling)

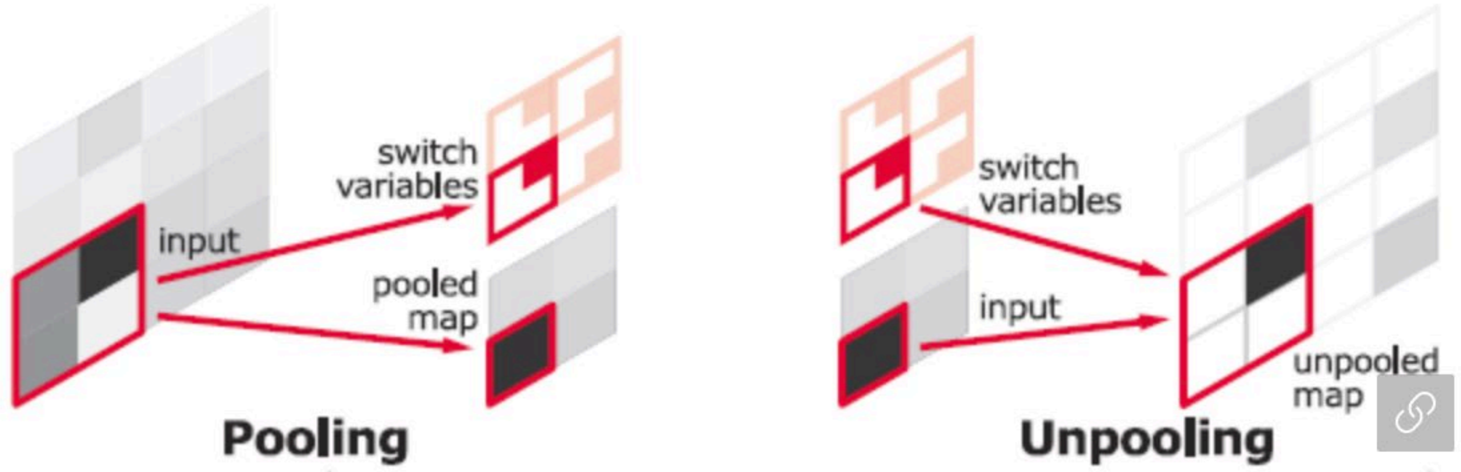
- Unpooling
- Transposed Convolution
- Skip Combining

Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Feature map 사이즈를 키워주기 위한 구조 제안

- **Unpooling**
- Transposed Convolution
- Skip Combining

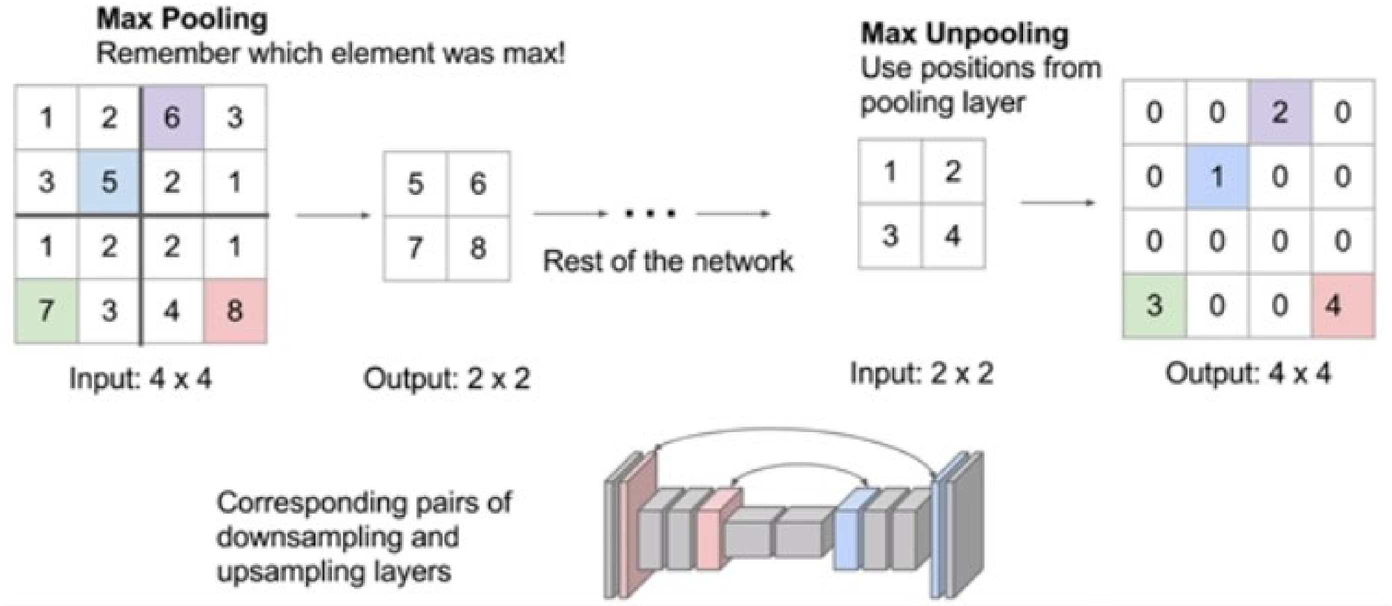


Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Feature map 사이즈를 키워주기 위한 구조 제안

- **Unpooling**
- Transposed Convolution
- Skip Combining

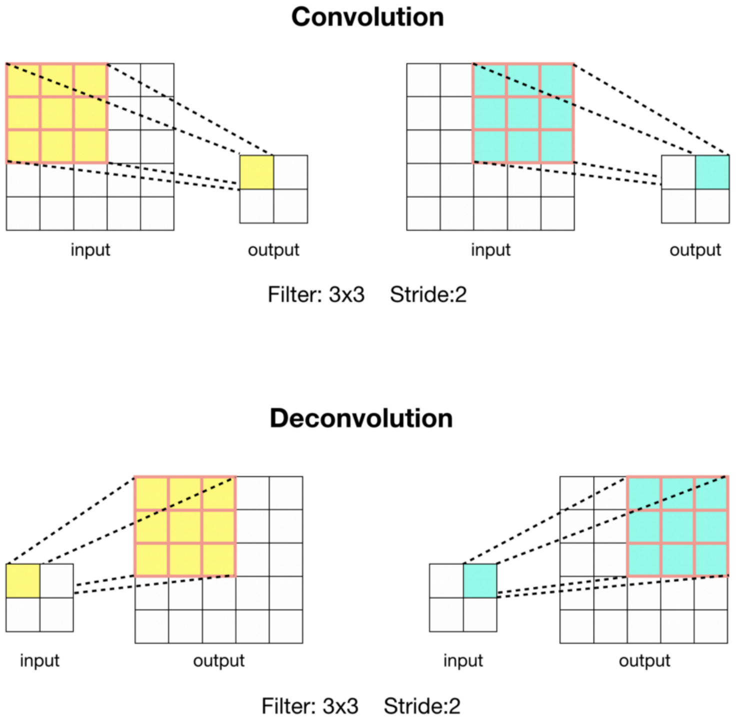


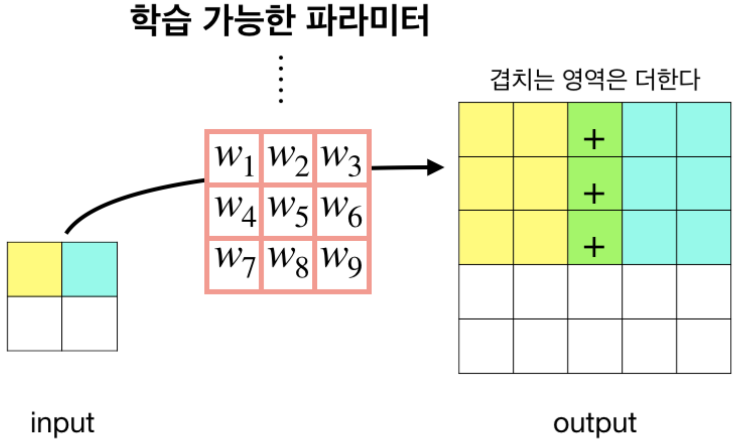
Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Feature map 사이즈를 키워주기 위한 구조 제안

- Unpooling
- **Transposed Convolution**
- Skip Combining



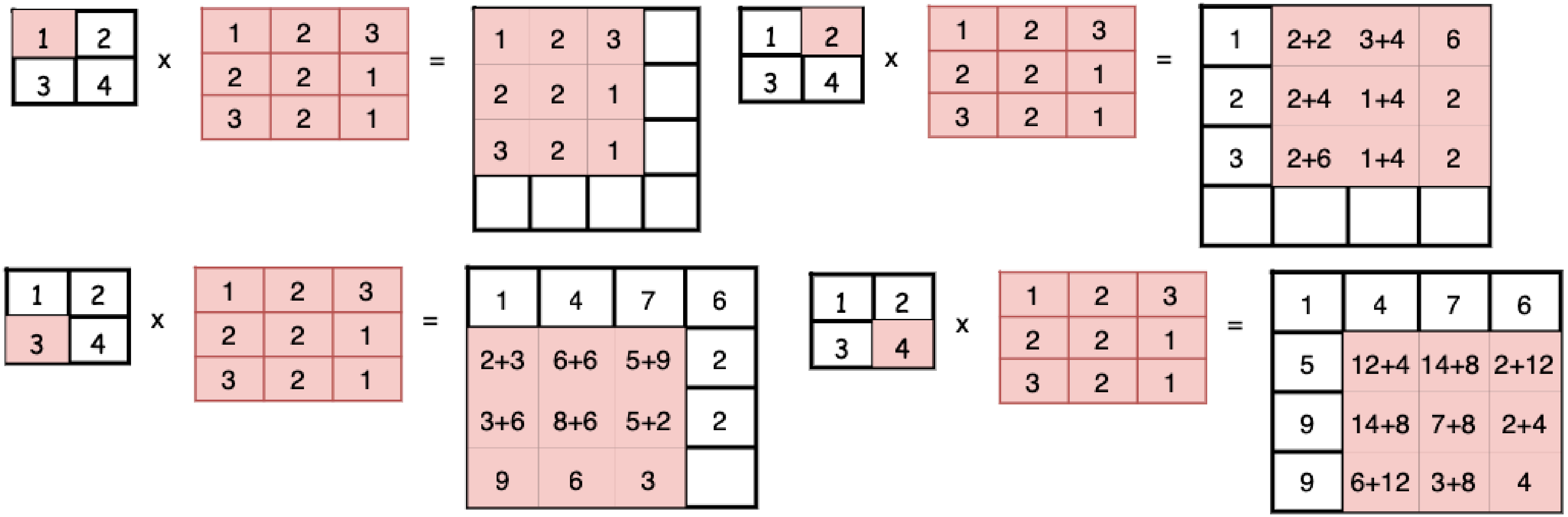


Backwards strided convolution
= Upsampling
= Deconvolution

Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Feature map 사이즈를 키워주기 위한 구조 제안

- Unpooling
- **Transposed Convolution**
- Skip Combining

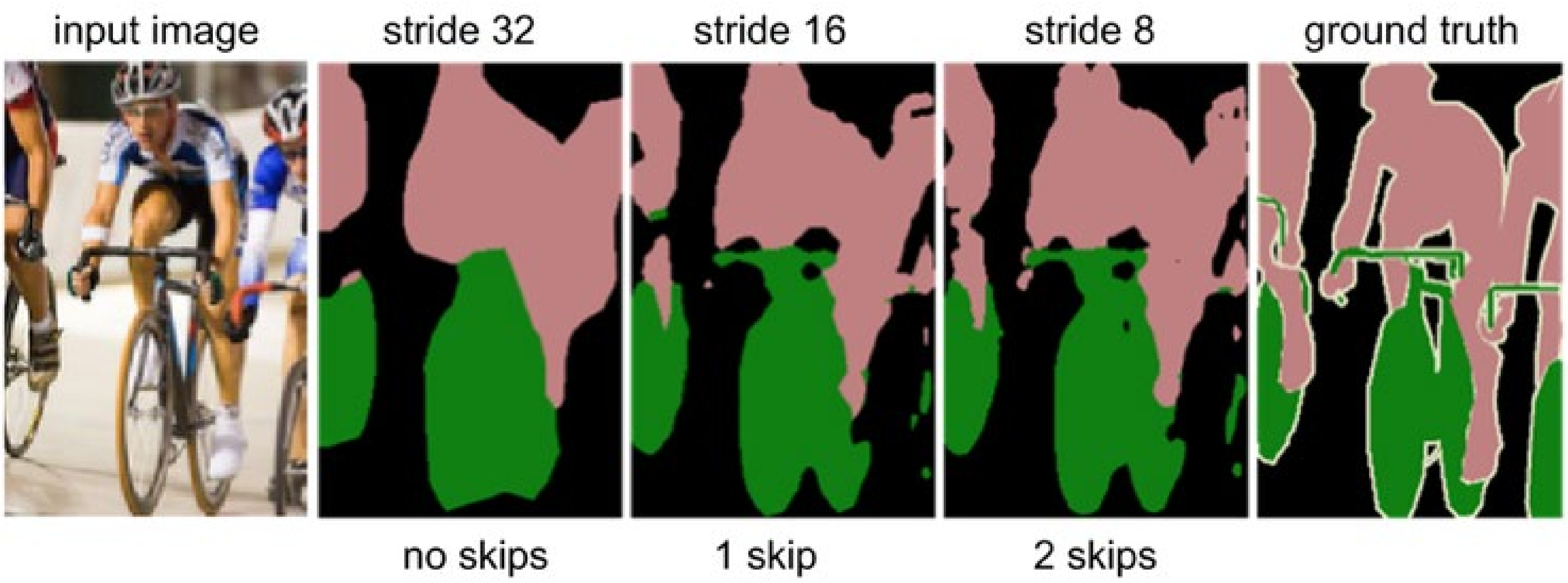


Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Feature map 사이즈를 키워주기 위한 구조 제안

- Unpooling
- Transposed Convolution
- **Skip Combining**

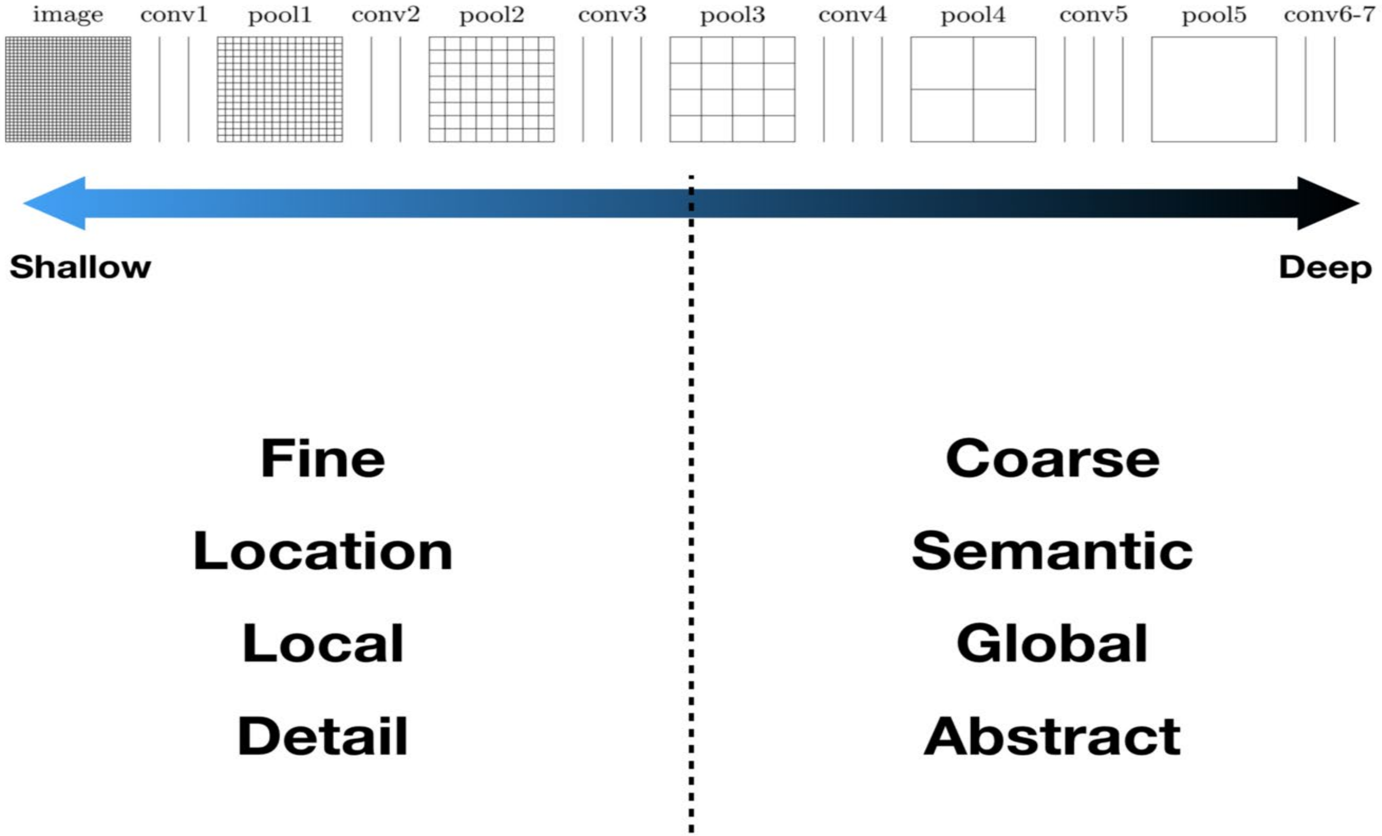


Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Feature map 사이즈를 키워주기 위한 구조 제안

- Unpooling
- Transposed Convolution
- **Skip Combining**

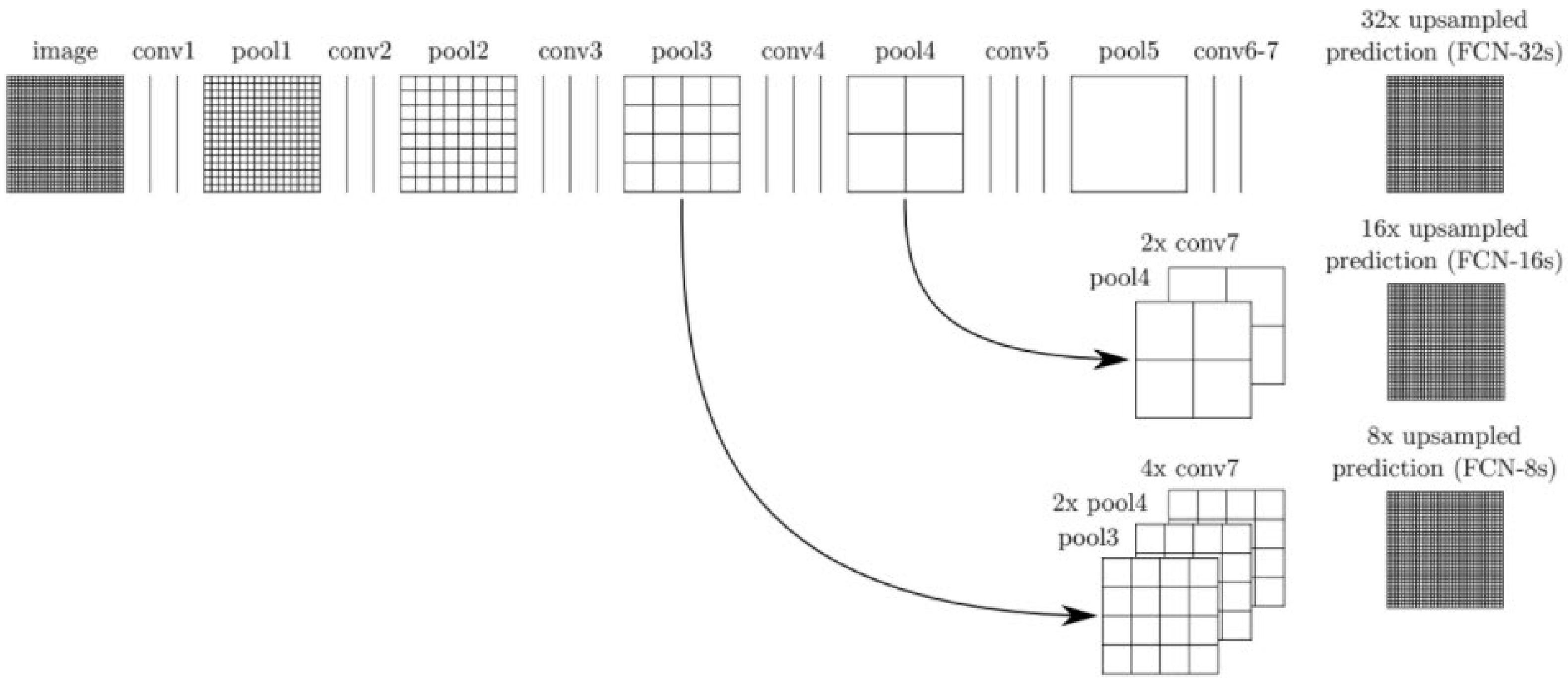


Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

Feature map 사이즈를 키워주기 위한 구조 제안

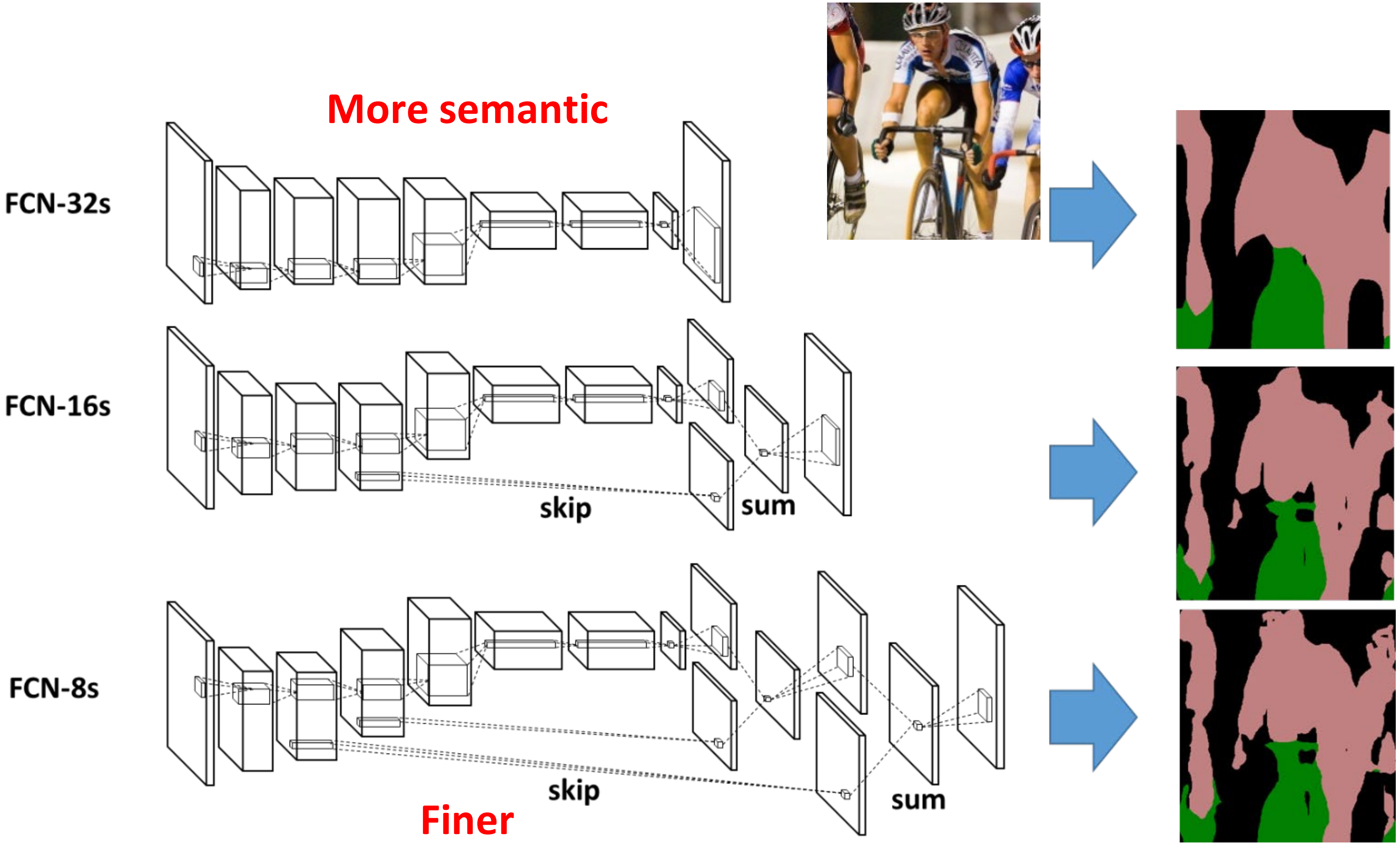
- Unpooling
- Transposed Convolution
- **Skip Combining**



Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

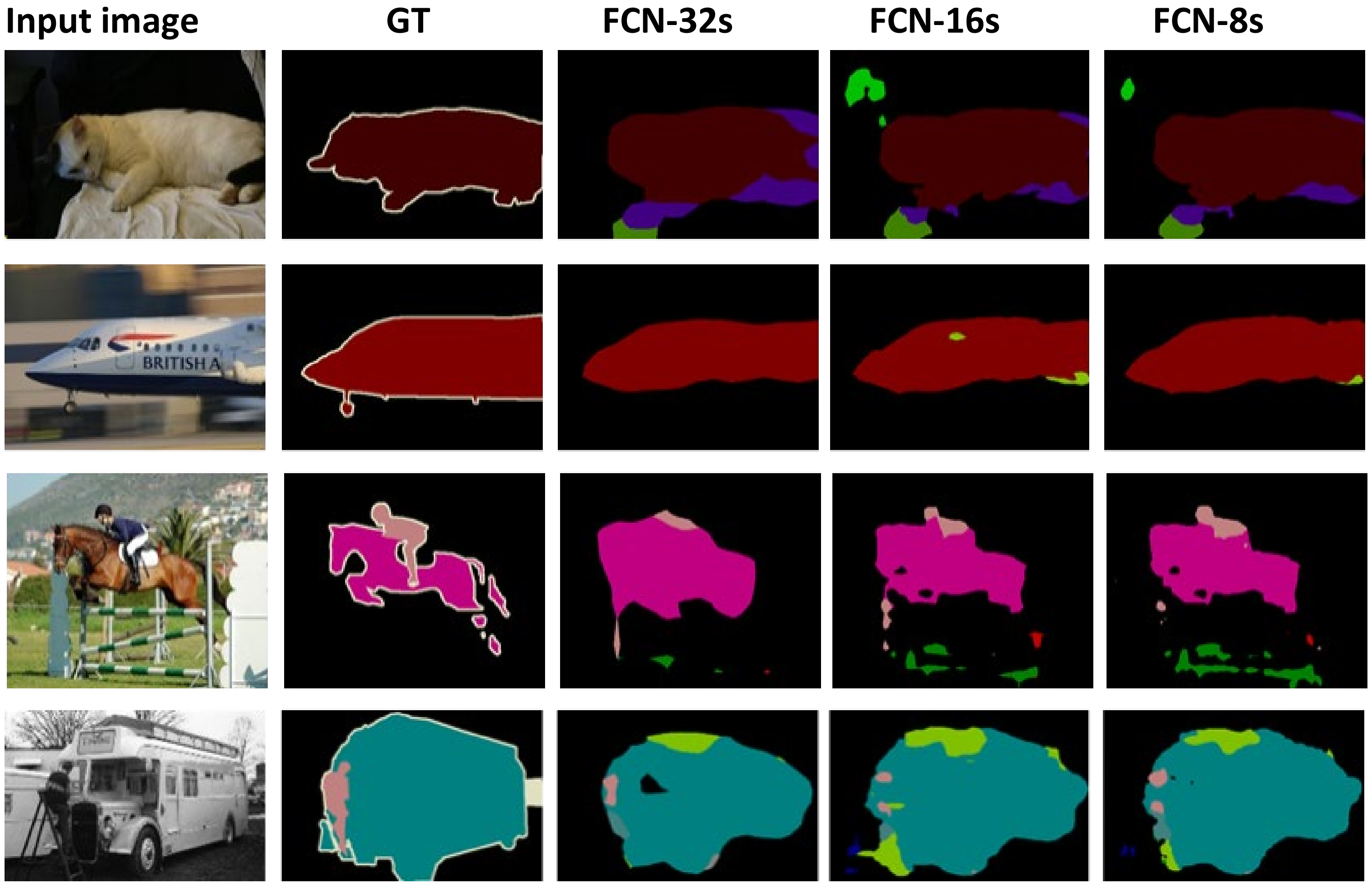
Fully Convolutional Network

• Skip architecture - Ensemble of three different scales



Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network



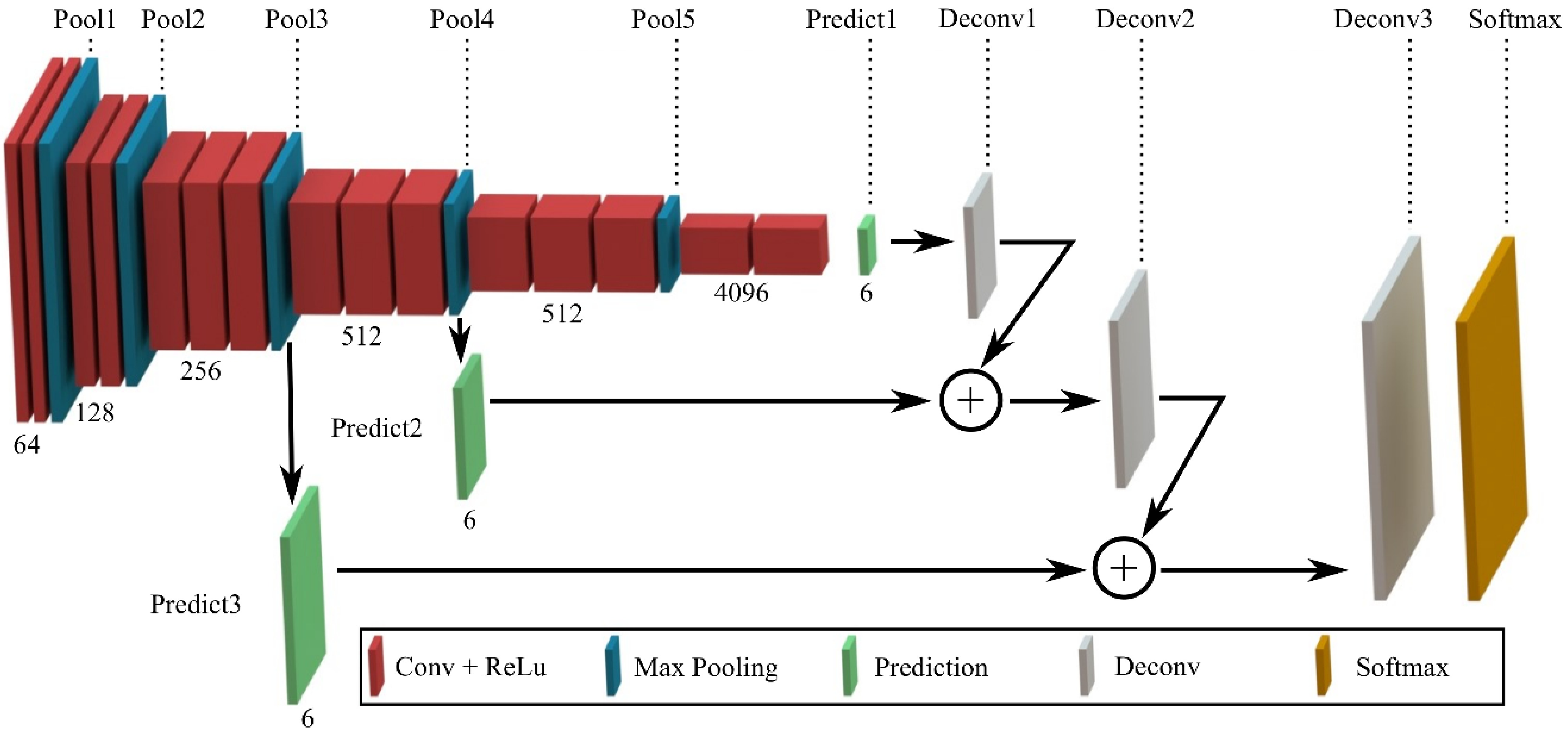
Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network

- Limitation of FCN-based semantic segmentation
  - Coarse output score map
    - A single bilinear filter should handle the variations in all kinds of object classes.
    - Difficult to capture detailed structure of objects in image
  - Fixed size receptive field
    - Unable to handle multiple scales
    - Difficult to delineate too small or large objects compared to the size of receptive field
  - Noisy predictions due to skip architecture
    - Trade off between details and noises
    - Minor quantitative performance improvement

Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

Fully Convolutional Network



Jonathan et al., **Fully convolutional networks for semantic segmentation**, *CVPR* 2015.

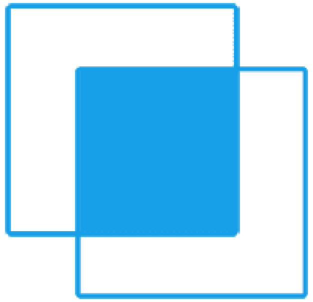
Evaluation metric for segmentation

There are several metrics for semantic segmentation

Most popular one is **Intersection over Union (IoU)**

- IoU measures the number of pixels common between the target and prediction masks divided by the total number of pixels present across both masks
- Mean IoU (mIoU)

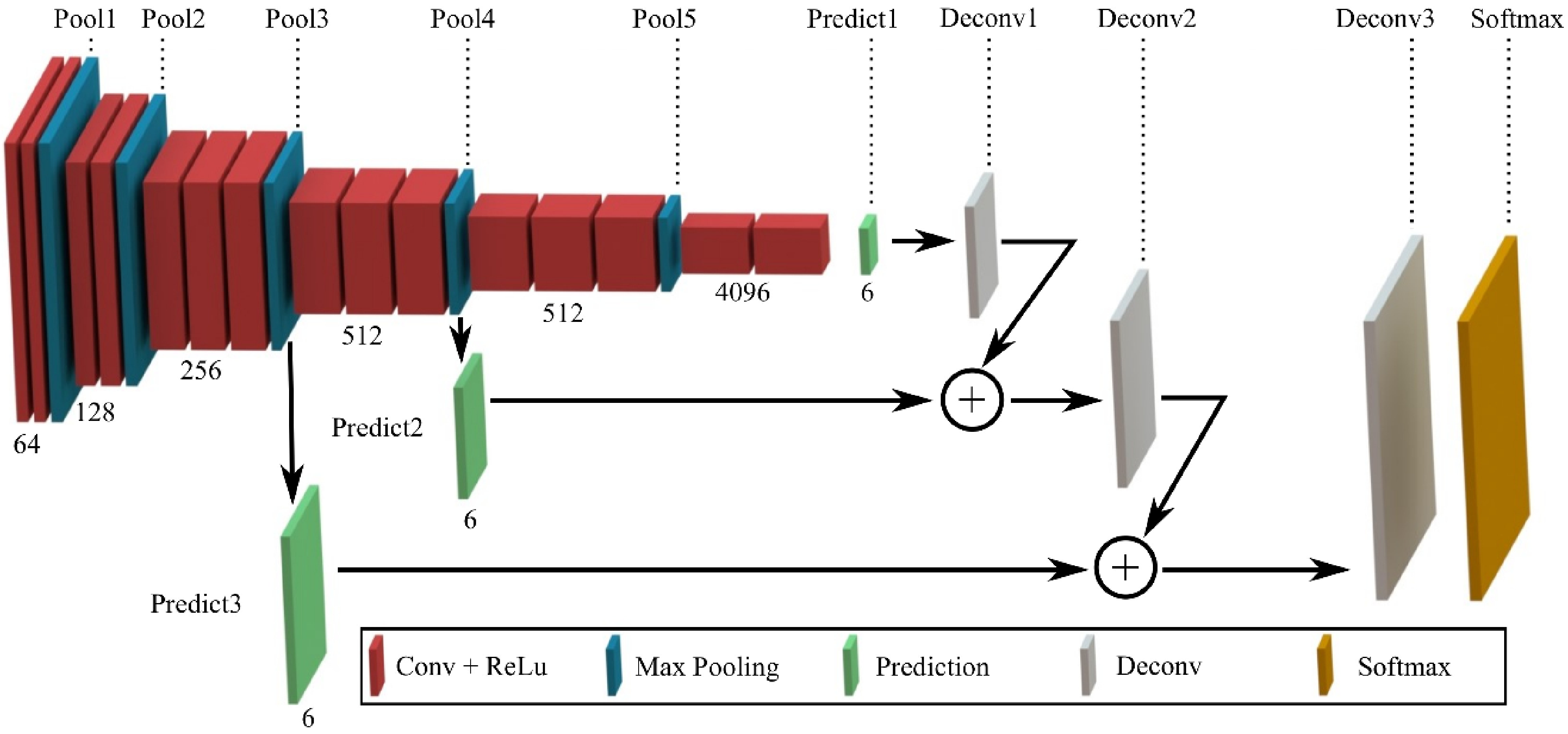
$$IoU = \frac{target \cap prediction}{target \cup prediction}$$



$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Exercise 1. FCN implementation

- Implement FCN with the following structure



Exercise 1. FCN implementation

- Exercise1. FCN implementation section in 24_Segmentation_practice.ipynb
- torch summary is shown on the right

- Things to consider
  - Batch size should be 1 → Why?
  - How to combine feature maps into one feature map to calculate the IoU
  - How to set the number of output feature

```
   | Name        | Type             | Params
-----------------------------------------------
0  | loss        | CrossEntropyLoss | 0
1  | features1   | Sequential       | 38.7 K
2  | features2   | Sequential       | 221 K
3  | features3   | Sequential       | 1.5 M
4  | features4   | Sequential       | 5.9 M
5  | features5   | Sequential       | 7.1 M
6  | maxpool     | MaxPool2d        | 0
7  | classifier  | Sequential       | 119 M
8  | upscore2    | ConvTranspose2d  | 64
9  | upscore4    | ConvTranspose2d  | 64
10 | upscore8    | ConvTranspose2d  | 1.0 K
11 | score_pool4 | Conv2d           | 1.0 K
12 | score_pool3 | Conv2d           | 514
13 | softmax     | Softmax2d        | 0
-----------------------------------------------
134 M     Trainable params
0         Non-trainable params
134 M     Total params
537.086   Total estimated model params size (MB)
```

Exercise 1. FCN implementation

- **Q. Why pad the input?**
- A. The 100 pixel input padding guarantees that the network output can be aligned to the input for any input size in the given datasets, for instance PASCAL VOC. The alignment is handled automatically by net specification and the crop layer. It is possible, though less convenient, to calculate the exact offsets necessary and do away with this amount of padding.

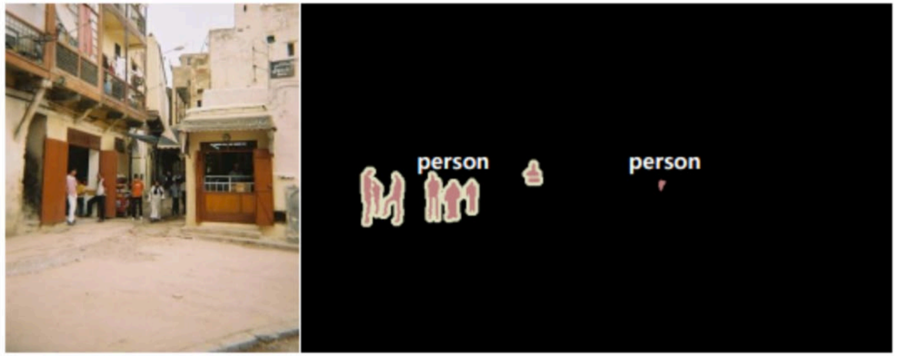
- **Q. Why is the batch size 1?**
- A. The size of the images are different in the dataset. Although the network can be trained regardless of the input size, the images in the same batch should be the same.

U-Net

Limitations of FCN

- Fixed-size receptive field: 신경망이 오직 하나의 scale이미지만 다룰 수 있음
- Deconvolution is too simple: bilinear interpolation is not good enough

(a) Inconsistent labels due to large object size

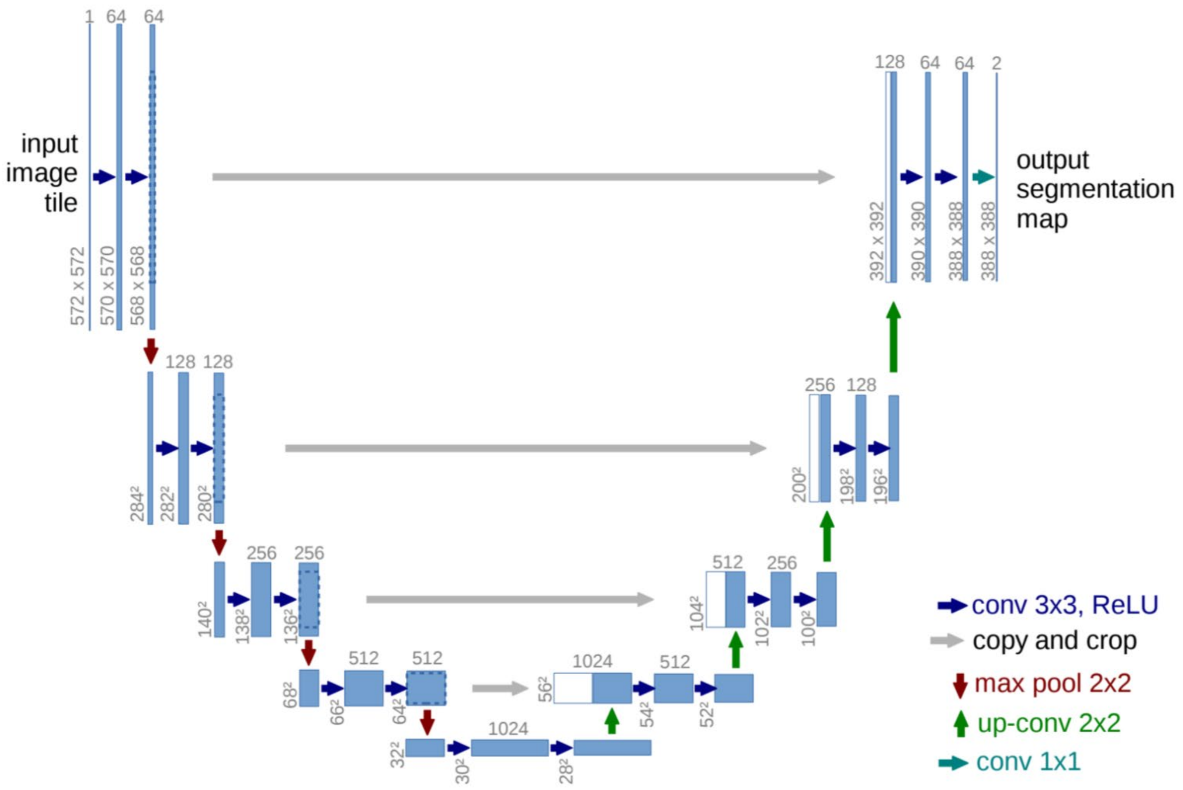


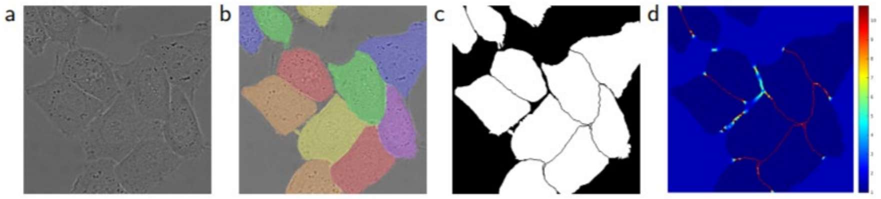
(b) Missing labels due to small object size

*Noh et al., **Learning Deconvolution Network for Semantic Segmentation**, ICCV 2015

U-Net

- U-Net: Convolutional Networks for Biomedical Image Segmentation



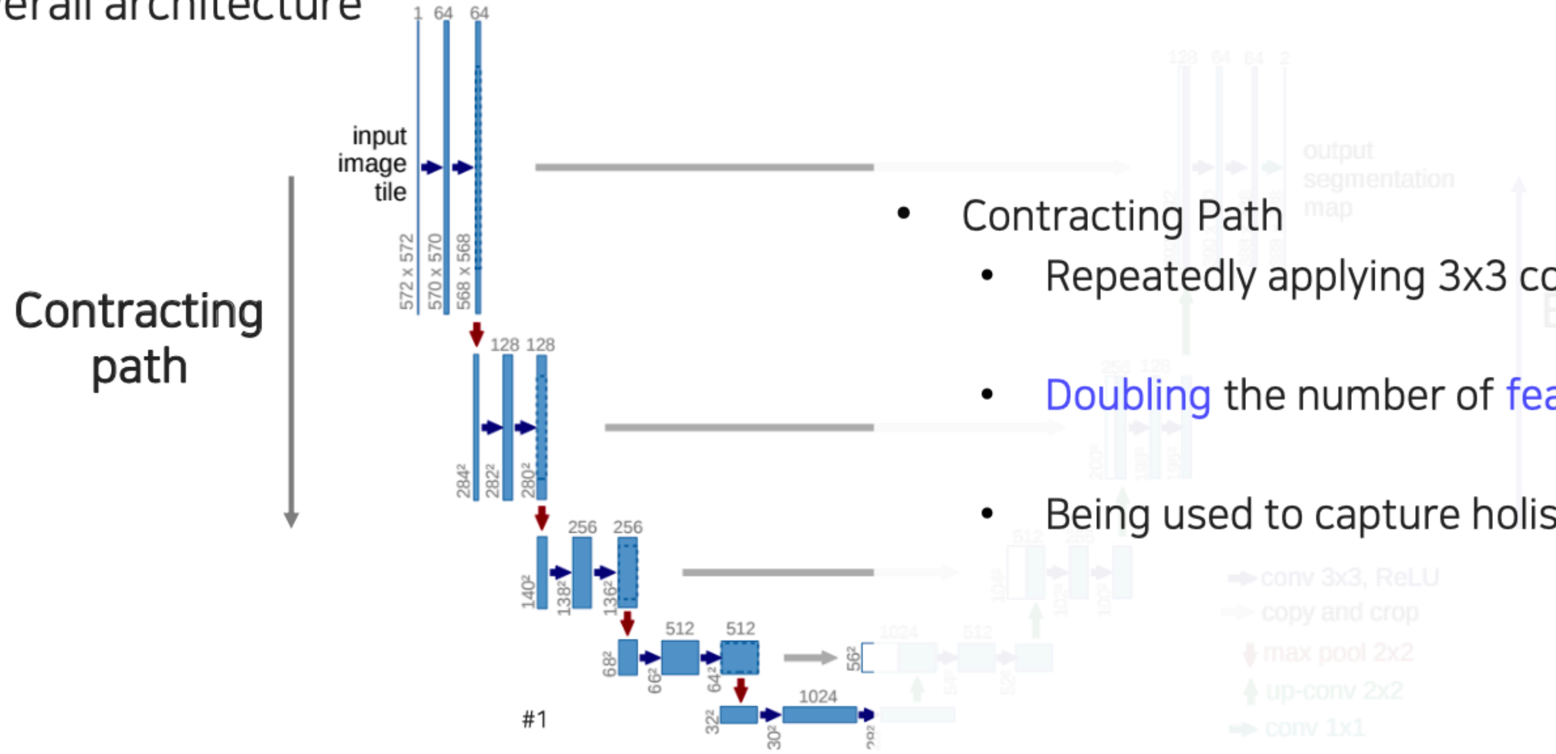


*Ronneberger et al, U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015

U-Net

- U-Net: Convolutional Networks for Biomedical Image Segmentation

◉ Overall architecture



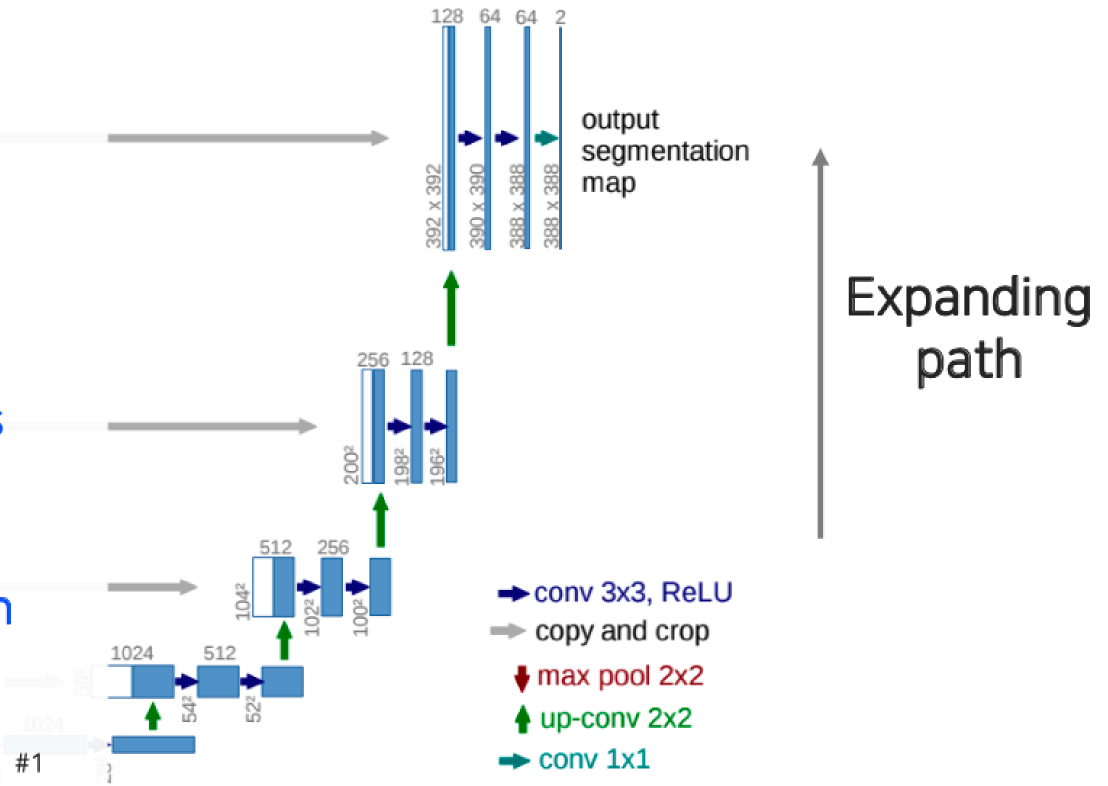
- Contracting Path
  - Repeatedly applying 3x3 convolutions
  - Doubling the number of feature channels
  - Being used to capture holistic context

*Ronneberger et al, U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015

U-Net

- U-Net: Convolutional Networks for Biomedical Image Segmentation

- Expanding Path
  - Repeatedly applying 2x2 convolutions
  - Halving the number of feature channels
  - Concatenating the corresponding feature maps from the contracting path



*Ronneberger et al, U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015

Exercise 2. U-Net implementation

- Exercise 2. U-Net implementation section in 24_Segmentation_practice.ipynb
- Think how to implement skip-connection