



Segmentation

진행자 : 멘토 현시은

무슨 뜻이지?

• 무슨 뜻이지?

Semantic
Segmentation
Transfer
Learning
Performance
Measure

Segmentation : 이미지를 픽셀 레벨로 이해하는 접근이다.

- 픽셀 하나하나에 대해서 분류 또는 탐지를 하는 task를 의미한다.
- 동일한 label을 가진 모든 픽셀이 동일한 object 또는 class에 속하도록 이미지의 각 픽셀에 label을 할당하는 프로세스를 의미한다.
- 이것은 image level로 예측을 하는 image classification보다 훨씬 어려움
- 자율주행, 로봇공학, 의료영상분석 등 다양한 분야에 응용되는 컴퓨터 비전의 기본 과제이다.

• 무슨 뜻이지?

Semantic
Segmentation
Transfer
Learning
Performance
Measure

Segmentation : 이미지를 픽셀 레벨로 이해하는 접근이다.

- Input : RGB 또는 흑백 이미지
- Output : 각 픽셀이 어느 class 에 속하는 지 나타내는 Map



Input



- 1: Person
- 2: Purse
- 3: Plants/Grass
- 4: Sidewalk
- 5: Building/Structures



Semantic Labels

사진 출처 : https://www.jeremyjordan.me/semantic-segmentation/#dilated_convolutions

- 무슨 뜻이지?

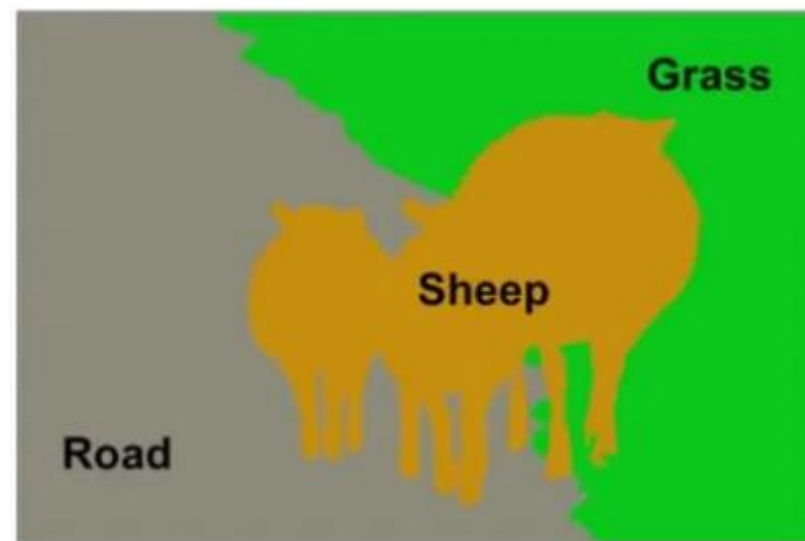
Semantic
Segmentation
Transfer
Learning
Performance
Measure

Segmentation의 종류

1. Semantic segmentation

같은 class의 instance를 구별하지 않는다.

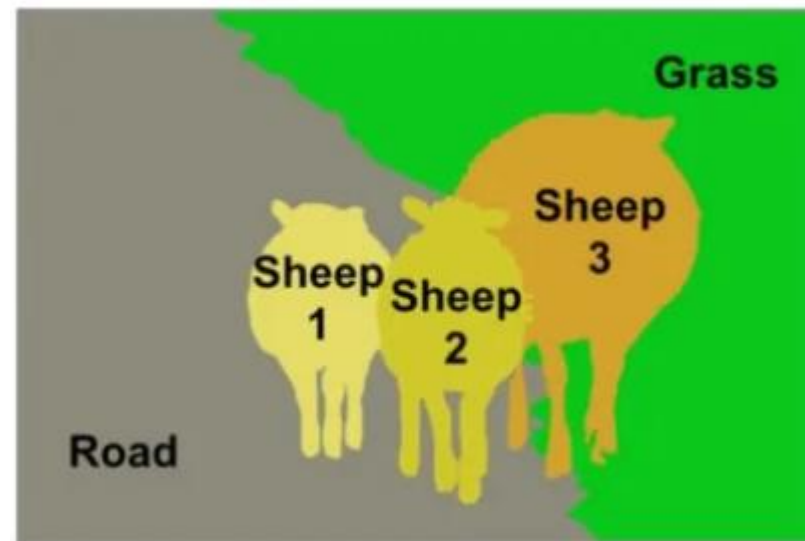
- instance란, 하나의 같은 class 안의 다른 개체를 의미한다.



Semantic Segmentation

2. Instance segmentation

같은 class 안에서도 다른 instance를 구분한다.

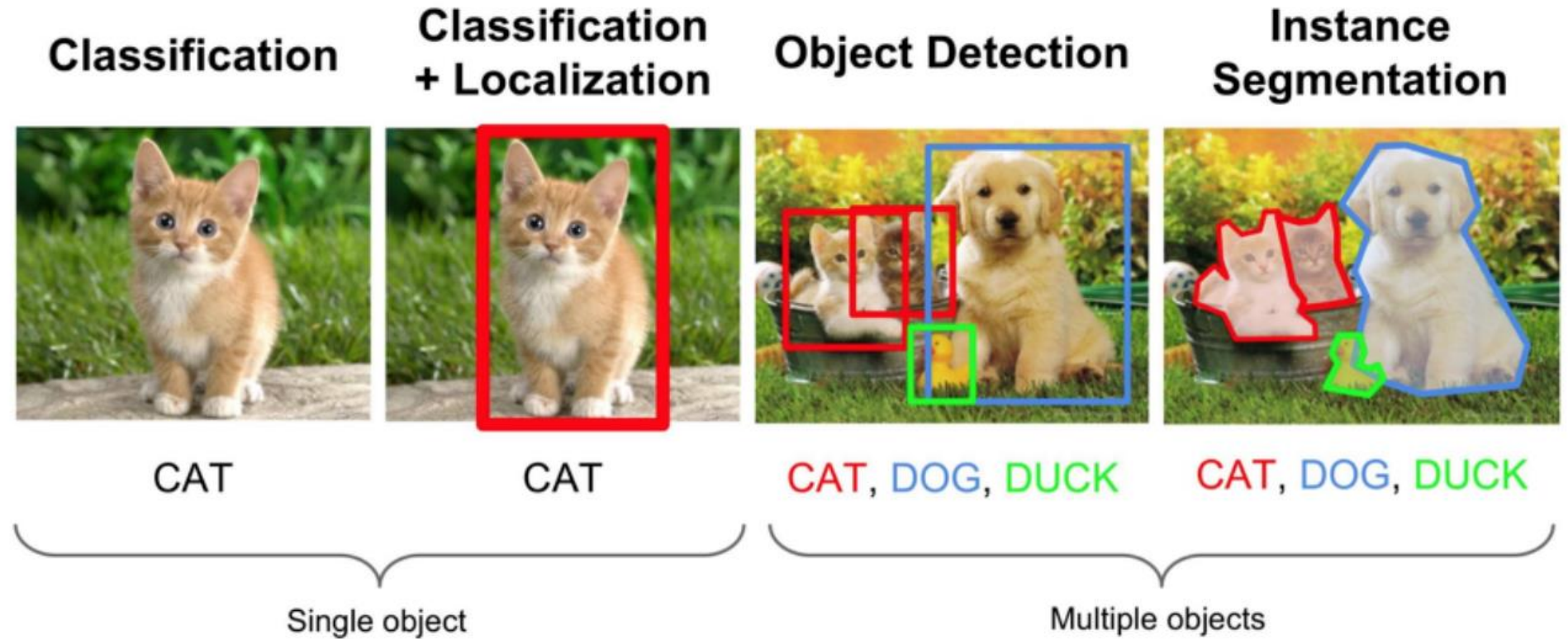


Instance Segmentation

사진 출처 : <https://towardsdatascience.com/detection-and-segmentation-through-convnets-47aa42de27ea>

- 무슨 뜻이지?

Semantic
Segmentation
Transfer
Learning
Performance
Measure



- 무슨 뜻이지?

Semantic
Segmentation
Transfer
Learning
Performance
Measure

활용되는 예시

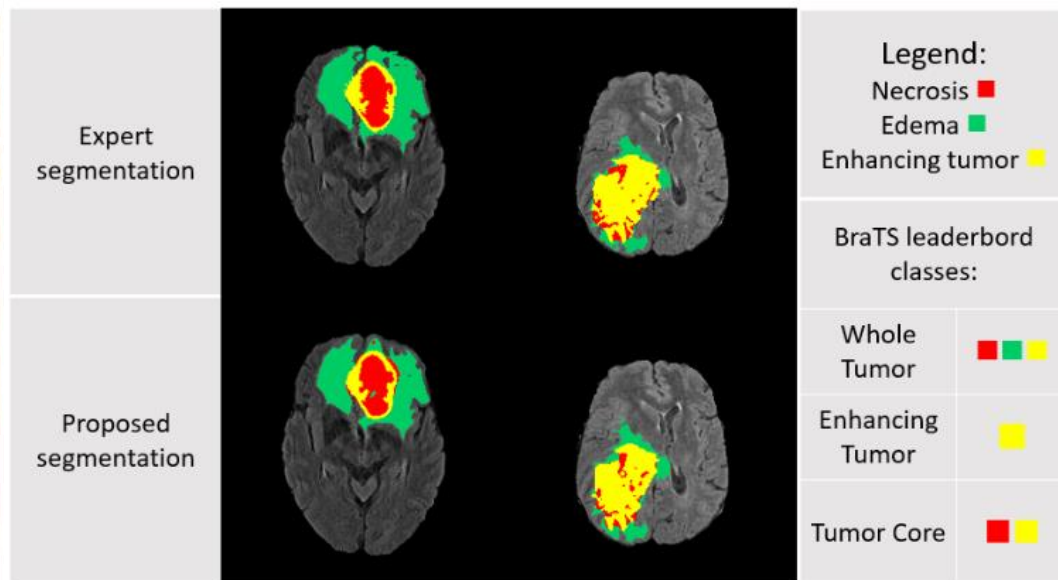


사진 출처 :

<https://devmesh.intel.com/projects/brats-2017-glioma-segmentation-using-fully-convolutional-neural-networks>

<https://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html>



Semantic Segmantation

무슨 뜻이지?

- **Semantic Segmentation**
Transfer Learning
Performance Measure

1. 픽셀 하나하나마다 classification

- 가장 단순하게 생각할 수 있는 방법
- 학습이 효과적으로 성공만 된다면, 가장 정확하고 강력한 모델일 것이다.
- 그러나, 굉장히 비효율적이고 학습이 상당히 어렵다.
: 만약 이미지 사이즈가 256×256 이라고 가정해도, 한 장의 이미지를 학습하는데 총 픽셀의 개수인 65,536번의 연산이 시행되어야 한다.

무슨 뜻이지?

- **Semantic Segmentation**
Transfer Learning
Performance Measure

2. Bilinear Interpolation

- Interpolation이란, 알려진 지점의 값 사이에 위치한 값을 알려진 값으로부터 추정하는 것을 말한다.
- 즉, 모든 픽셀을 전부 학습하지 않고 몇 개 만을 학습한 후 학습하지 않은 중간 중간의 픽셀은 interpolation으로 추정한다.
- 정확도는 다소 떨어질 수 있어도, 계산량과 학습시간을 현저히 줄일 수 있다.

무슨 뜻이지?

- Semantic Segmentation

Transfer Learning

Performance Measure

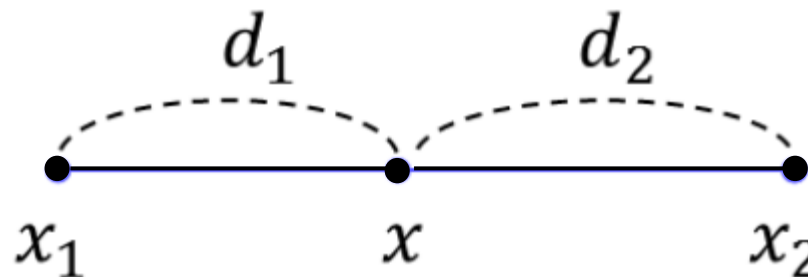
2. Bilinear Interpolation

1-D Linear Interpolation

- 즉, 두 지점과의 직선 거리에 따라 선형적으로 결정하는 방법
- 두 지점 x_1, x_2 에서의 데이터 값이 $f(x_1), f(x_2)$ 일 때, x_1 과 x_2 사이의 점 x 의 데이터값은 다음과 같은 식으로 계산된다.

$$f(x) = \frac{d_2}{d_1 + d_2} f(x_1) + \frac{d_1}{d_1 + d_2} f(x_2)$$

- 이 때, d_1 과 d_2 의 관계는 다음과 같다.



무슨 뜻이지?

- **Semantic Segmentation**

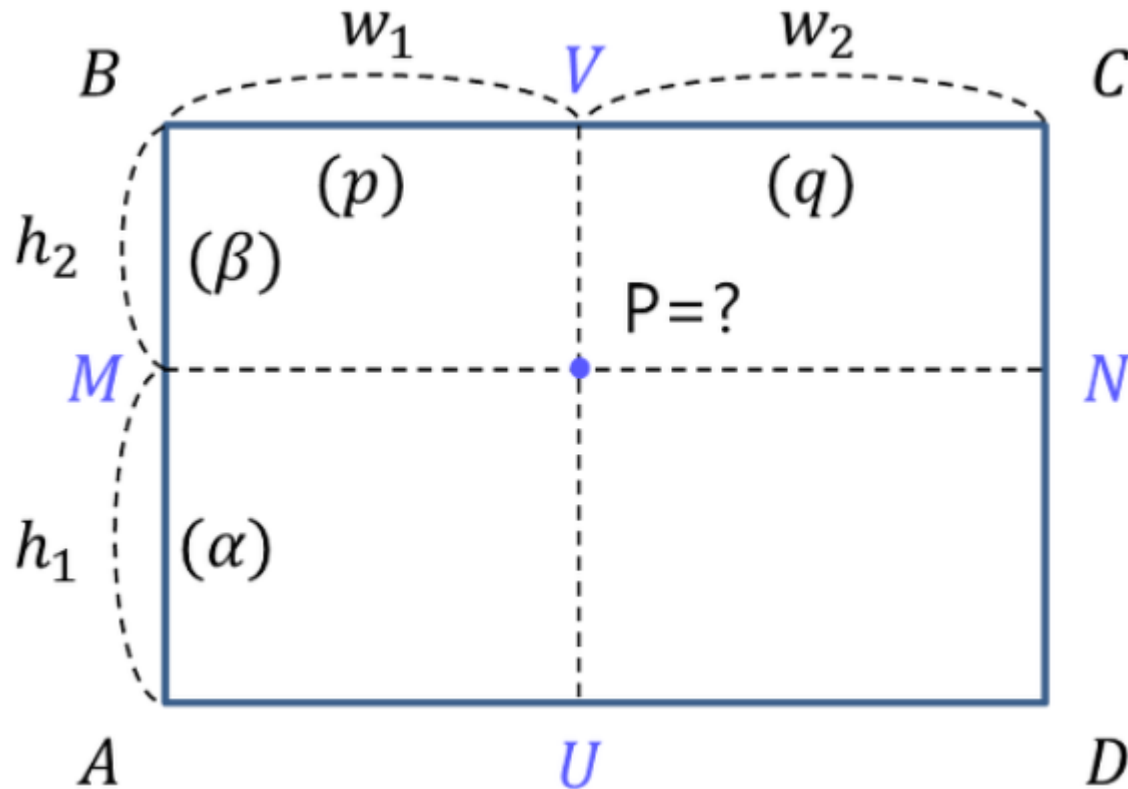
Transfer Learning

Performance Measure

2. Bilinear Interpolation

2-D Linear Interpolation (= Bilinear Interpolation)

- 1-D Linear Interpolation을 2차원으로 확장한 것이다.



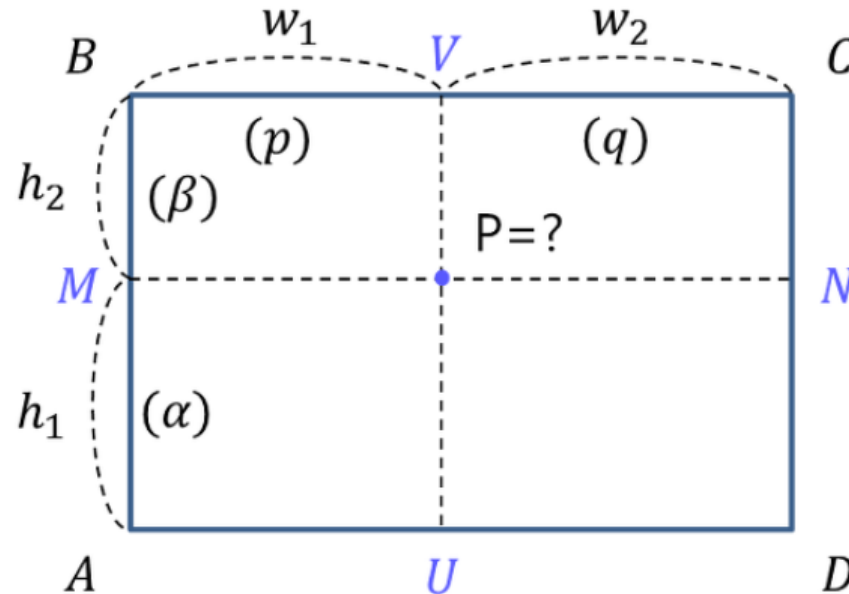
무슨 뜻이지?

- Semantic Segmentation

Transfer Learning

Performance Measure

2. Bilinear Interpolation



- 1-D Linear Interpolation을 2차원으로 확장한 것이다.

- $\alpha = \frac{h_1}{h_1+h_2}$, $\beta = \frac{h_2}{h_1+h_2}$, $p = \frac{w_1}{w_1+w_2}$, $q = \frac{w_2}{w_1+w_2}$ 일 때,

$$P = q(\beta A + \alpha B) + p(\beta D + \alpha C)$$

무슨 뜻이지?

- **Semantic Segmentation**
Transfer Learning
Performance Measure

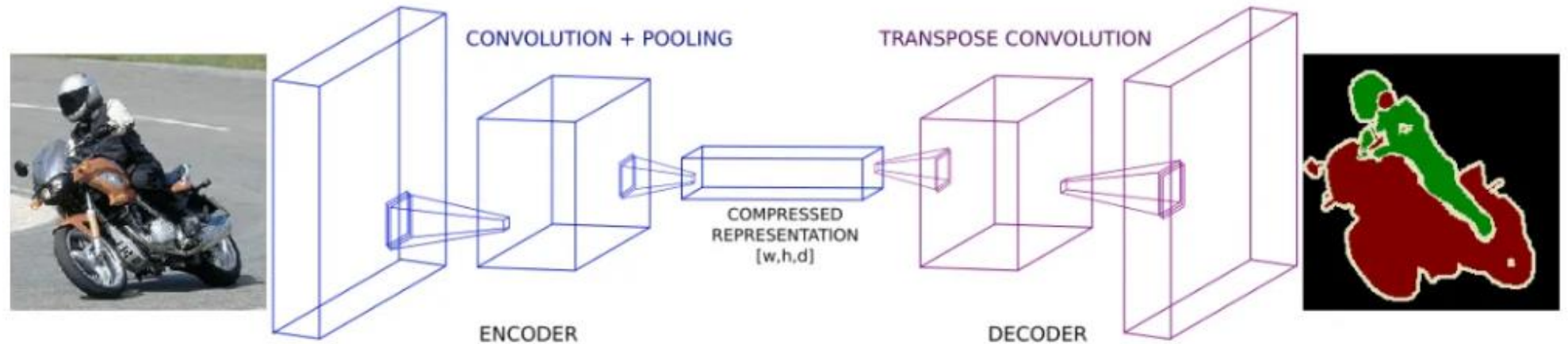
3. Convolution 기반 모델들

- VGG, Resnet 등 classification 에서 쓰이는 모델들은 Semantic Segmentation 에서는 적합하지 않다. 층이 깊어지고, 마지막 층의 Fully Connected Layer 를 통과하면서 많은 파라미터와 차원이 줄어들면서 Object 의 위치정보를 잃기 때문
- 위치 정보를 잃지 않기 위해 Pooling과 FCN 등을 없애고 stride 와 padding 을 일정하게 한다고 하면, 학습량이 너무 많아져서 비효율적이다.
- 따라서 Downsampling 과 Upsampling 을 활용한다!

무슨 뜻이지?

- Semantic Segmentation
- Transfer Learning
Performance Measure

3. Convolution 기반 모델들



- 논문에서는 보통 Downsampling을 인코더, Upsampling을 디코더라고 부른다.
- 인코더를 통해 입력 받은 이미지의 정보를 벡터로 압축시키고, 디코더를 통해서 원하는 결과물로 다시 확대한다.

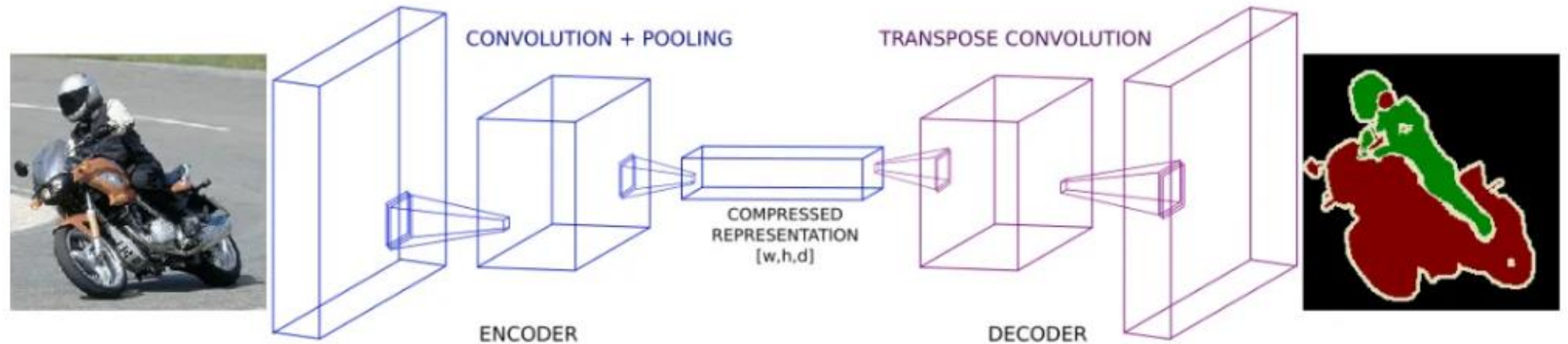
사진 출처 :

<https://medium.com/free-code-camp/diving-into-deep-convolutional-semantic-segmentation-networks-and-deeplab-v3-4f094fa387df>

무슨 뜻이지?

- Semantic Segmentation
- Transfer Learning
- Performance Measure

3. Convolution 기반 모델들



- 인코더 : 차원을 줄여서 적은 메모리로 깊은 Convolution 을 할 수 있게 한다. 주로 Convolution과 Pooling들을 이용한다.
- 디코더 : 인코더를 통과한 결과의 차원을 다시 늘려서 input과 같은 차원으로 만들어준다. 주로 Transpose Convolution을 이용한다.

사진 출처 :

<https://medium.com/free-code-camp/diving-into-deep-convolutional-semantic-segmentation-networks-and-deeplab-v3-4f094fa387df>

무슨 뜻이지?

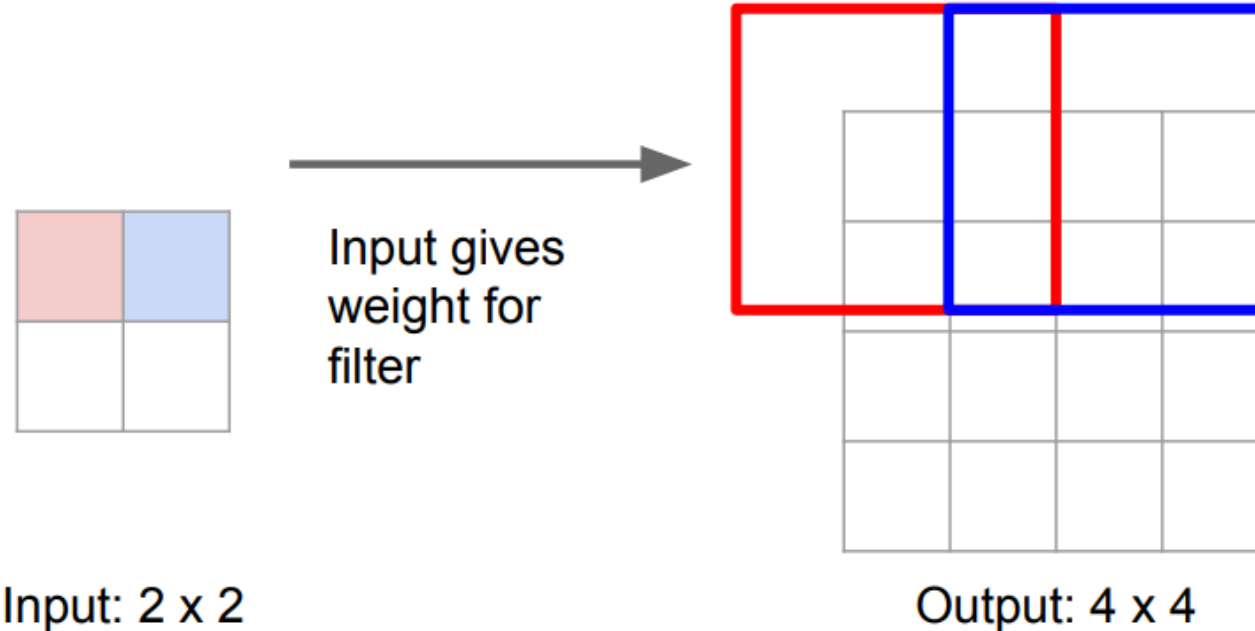
- **Semantic Segmentation**

Transfer Learning

Performance Measure

Transpose Convolution

3 x 3 transposed convolution, stride 2 pad 1



- 이것의 연산은 간단히 설명만 하고 생략
 - 자세한 내용은 다음의 링크를 참조
- 1) http://cs231n.stanford.edu/slides/2023/lecture_11.pdf
 - 2) <https://towardsdatascience.com/what-is-transposed-convolutional-layer-40e5e6e31c11>

사진 출처 : http://cs231n.stanford.edu/slides/2023/lecture_11.pdf

무슨 뜻이지?

- **Semantic Segmentation**

Transfer Learning

Performance Measure

3. Convolution 기반 모델들

- 이러한 인코더 - 디코더 구조를 가지는 대표적인 모델은 다음과 같다.
: FCN, DeepLab, U-Net 등 (우리 실습에서는 FCN을 다뤄볼 것이다.)
- 이들은 단지 Network 의 구조를 어떻게 변경했느냐의 차이이다.
(특히, Shortcut connection을 어떻게 하였느냐가 중요)

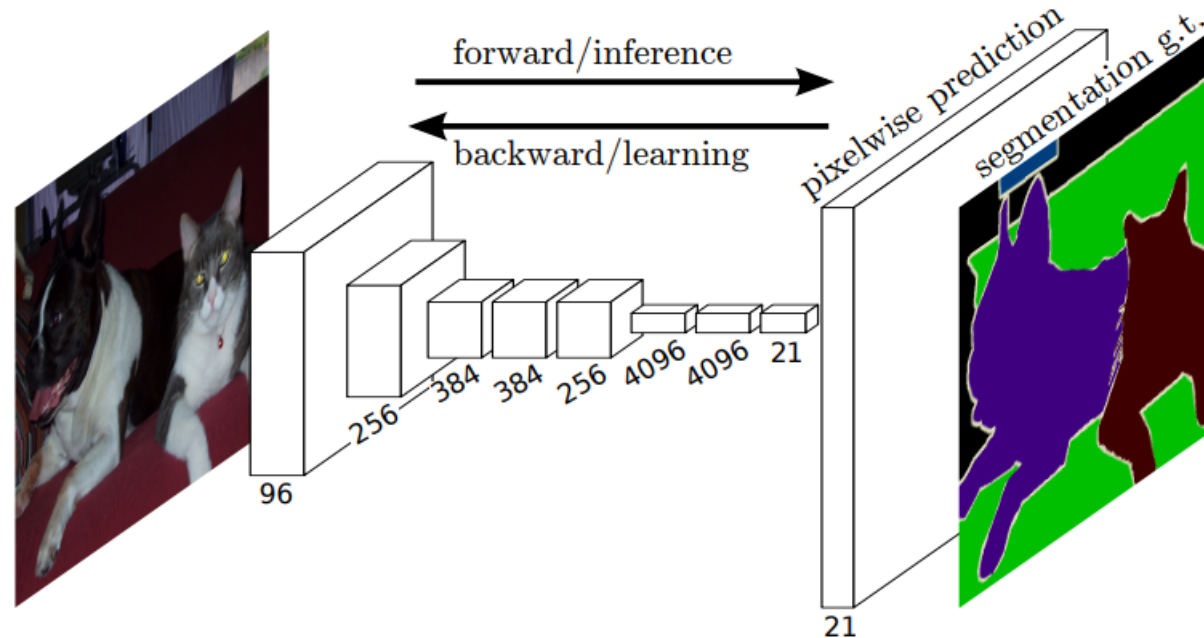
무슨 뜻이지?

- Semantic Segmentation

Transfer Learning

Performance Measure

Fully Convolutional Network for Semantic Segmentation (FCN)



- FCN은 Semantic Segmentation을 위해 Image Classification 에서 우수한 성능을 보인 CNN 기반 모델(AlexNet, VGG16 등)을 목적에 맞춰 변형시킨 것이다.
- 기존의 Image Classification 에서 사용한 fc-layer 의 문제점인 '위치 정보의 소실' 문제를 방지하기 위해 모든 fc-layer를 convolution layer 로 대체했다.
- 또한, Transfer learning 을 이용했다.



Transfer Learning

무슨 뜻이지?

Semantic
Segmentation

- **Transfer Learning**

Performance
Measure

Transfer Learning

Daffodils



Pansies



- Motivation : 초기 weight를 좋은 weight로 설정하자!!
- 데이터가 다르더라도, 이미지를 활용한 task에서는 공통된 feature 들이 존재할 것이다.
- 따라서, 공통된 feature 를 활용하여 학습 성능을 높이자.

사진 출처 : <https://www.robots.ox.ac.uk/~vgg/research/flowers/>

무슨 뜻이지?

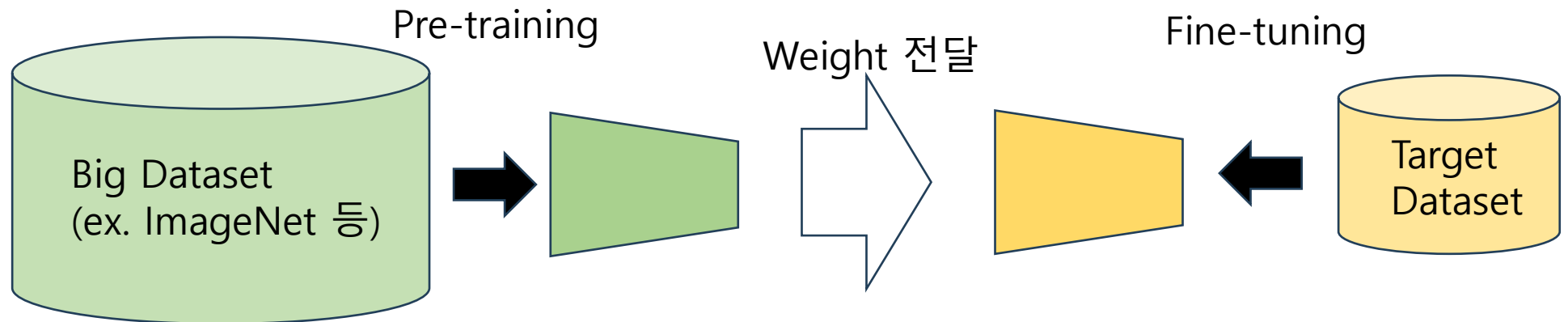
Semantic
Segmentation

• Transfer Learning

Performance
Measure

Transfer Learning

- 정의 : Transfer learning (TL) is a technique in machine learning (ML) in which knowledge learned from a task is re-used in order to boost performance on a related task. (출처 : [Transfer learning - Wikipedia](#))



무슨 뜻이지?

Semantic
Segmentation

- **Transfer
Learning**

Performance
Measure

다시 모델 설명으로 돌아갑시다!



Semantic Segmentation

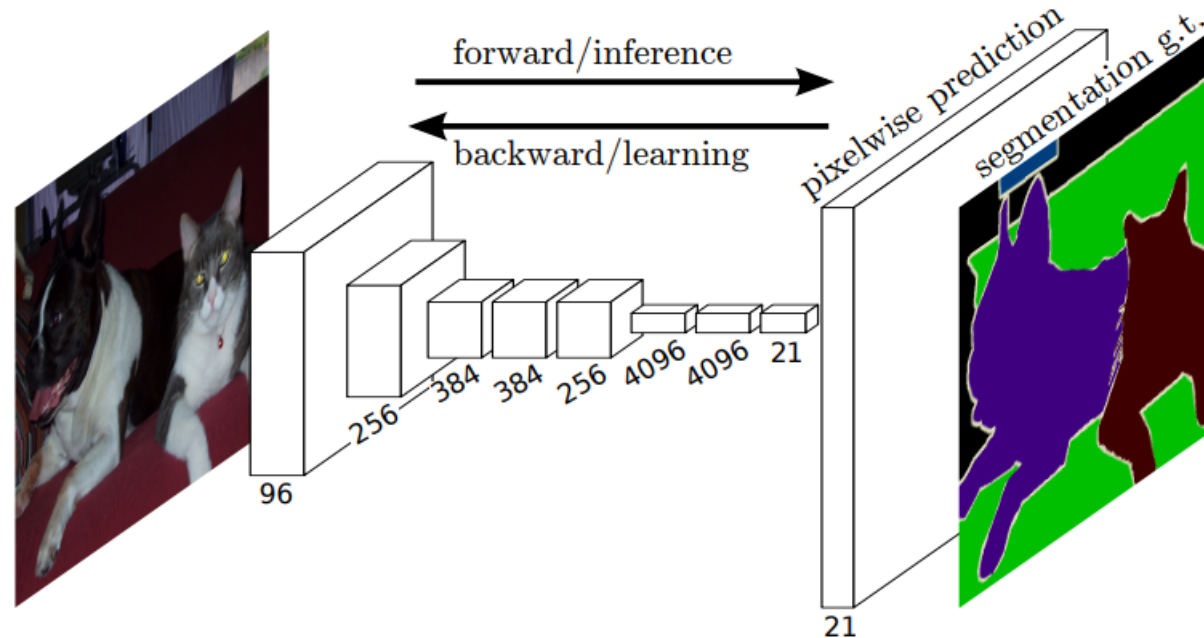
무슨 뜻이지?

- Semantic Segmentation

Transfer Learning

Performance Measure

Fully Convolutional Network for Semantic Segmentation (FCN)



- VGG16으로부터 Transfer Learning을 사용한다.
- VGG16의 마지막 layer인 fully connected layer를 1x1의 convolution layer로 바꾼다. (이 때문에 이름이 'Fully Convolutional Network'인 것이다.)
- 이후 Transposed Convolution 을 통해 압축 되어있는 heat map을 입력 이미지와 동일한 사이즈로 만든다.

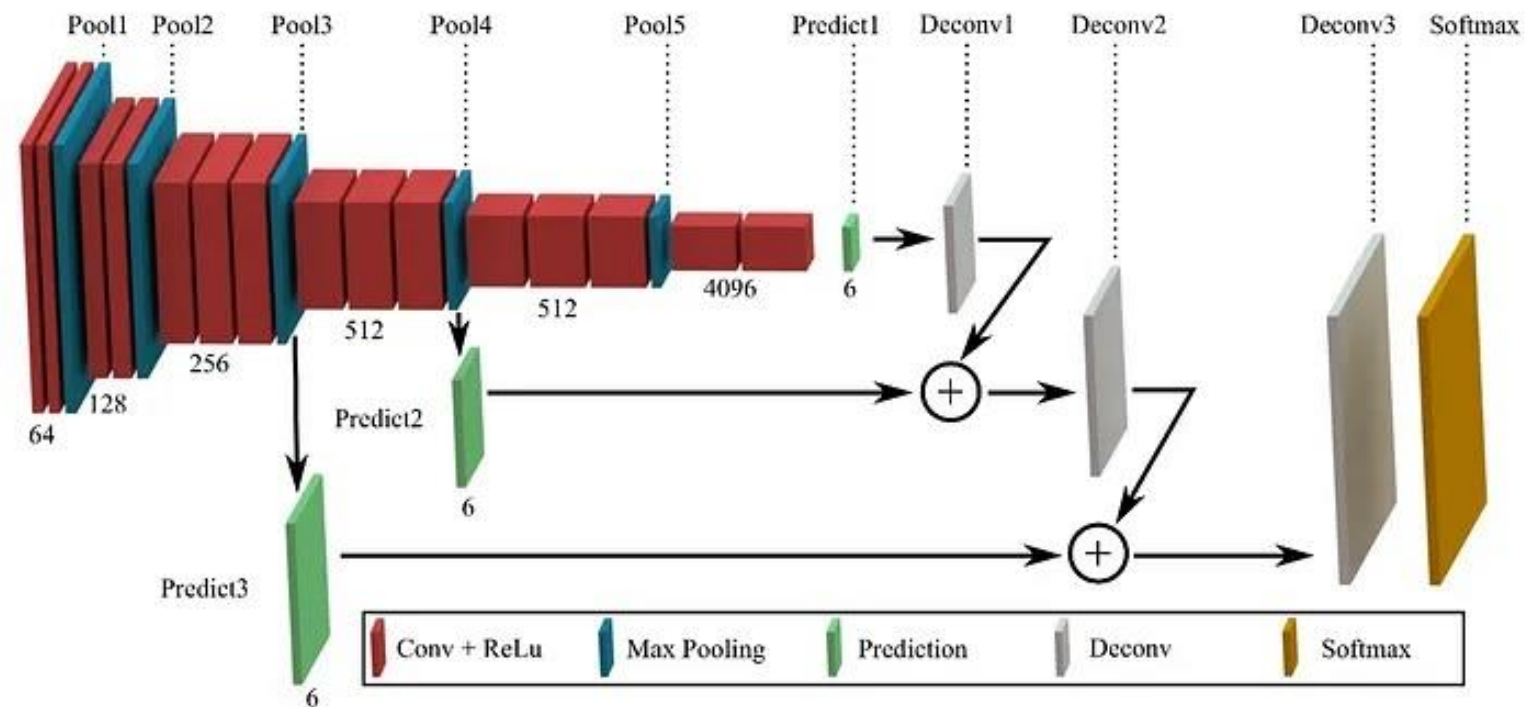
무슨 뜻이지?

- **Semantic Segmentation**

Transfer Learning
Performance Measure

Fully Convolutional Network for Semantic Segmentation (FCN)

- 또한, FCN은 'Skip Connection' 라는 구조를 적용하였다.
- Upsampling (디코딩)을 하면서 VGG16의 낮은 layer의 특징 맵을 더한다.
- Pooling을 통해 압축되며 소실되는 중요한 정보를 이 skill 을 통해 얻을 수 있다.



무슨 뜻이지?

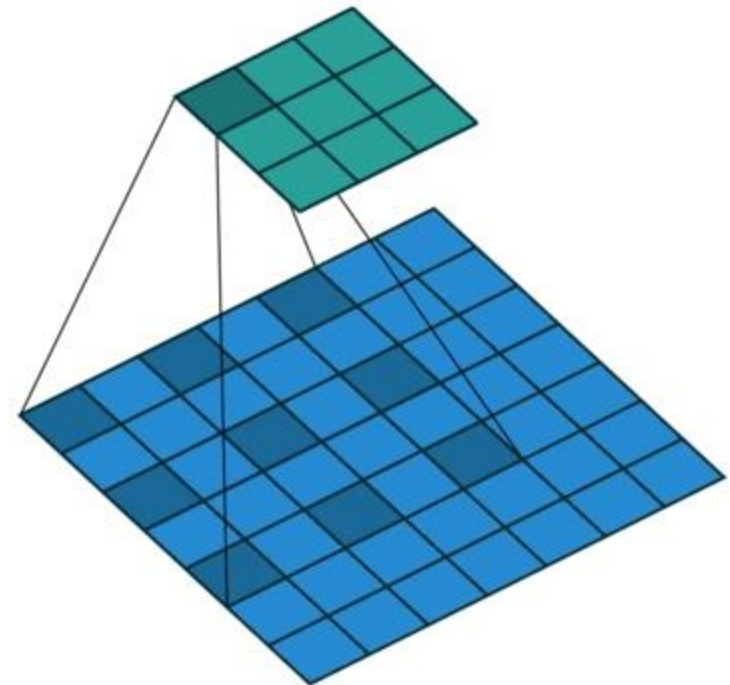
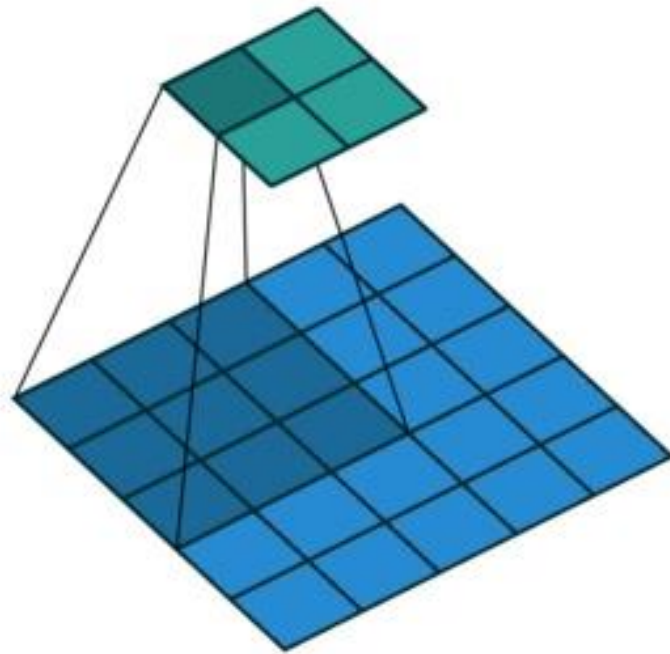
- **Semantic Segmentation**

Transfer
Learning

Performance
Measure

Deeplab

- FCN의 문제는, Receptive Field의 크기가 일정하게 정해져 있어서 작은 물체를 잘 인식하지 못했고, 일정하지 않은 결과가 나오기도 하였다.
- 따라서 Deeplab이라는 모델에서 Dilated Convolution을 이용하였다.

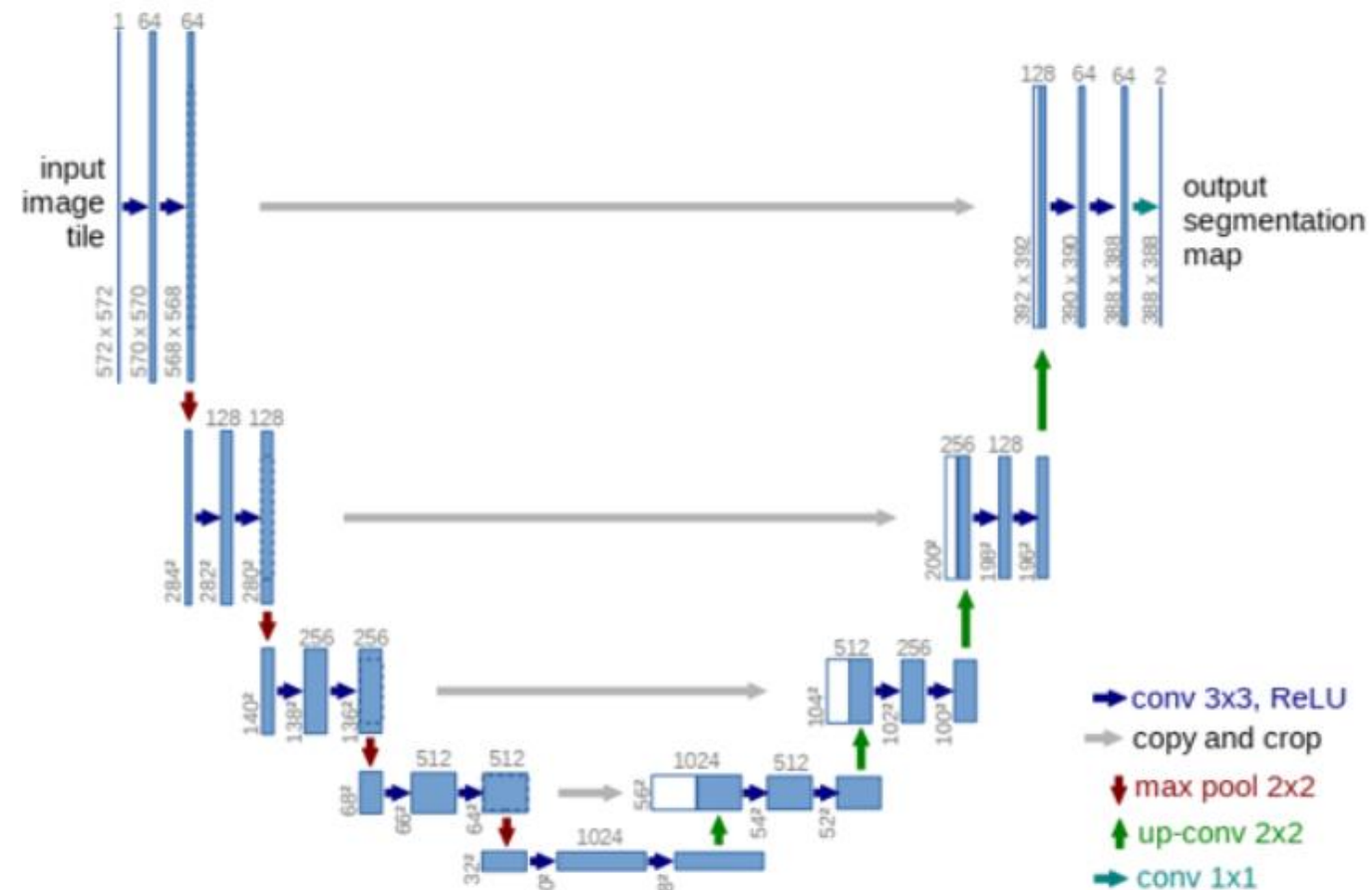


무슨 뜻이지?

- **Semantic Segmentation**
- Transfer Learning
- Performance Measure

U-Net

- 신경망 구조를 스킵 연결을 평행하게 두고 가운데를 기준으로 좌우가 대칭이 되도록 레이어를 배치하여, 이름 그대로 U자 형으로 만들었다.





Performance
measure

무슨 뜻이지?

Semantic
Segmentation

Transfer
Learning

- Performance
measure

Performance measure

- IoU

$$IoU = 100 \times \frac{\text{true positive}}{\text{true positive} + \text{false positive} + \text{false negative}}$$
$$= \frac{\text{Overlapping region}}{\text{combined region}}$$

		실제 정답	
		Positive	Negative
실험 결과	Positive	True positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

무슨 뜻이지?

Semantic
Segmentation
Transfer
Learning

- Performance
measure

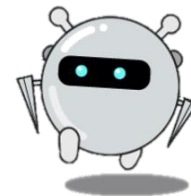
Performance measure

- PQ

$$PQ = \underbrace{\frac{\sum_{(p,g) \in TP} \text{IoU}(p, g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}} = \frac{\sum_{(p,g) \in TP} \text{IoU}(p, g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$



감사합니다



Reference

Fully Convolutional Networks for Semantic Segmentation

MULTI-SCALE CONTEXT AGGREGATION BY DILATED CONVOLUTIONS