

# Computer Vision

A – CV - object detection & segmentation 이규리

# CONTENTS

J1 Computer vision

J2 EfficientNet

J3 YOLO

J4 Mask R-CNN

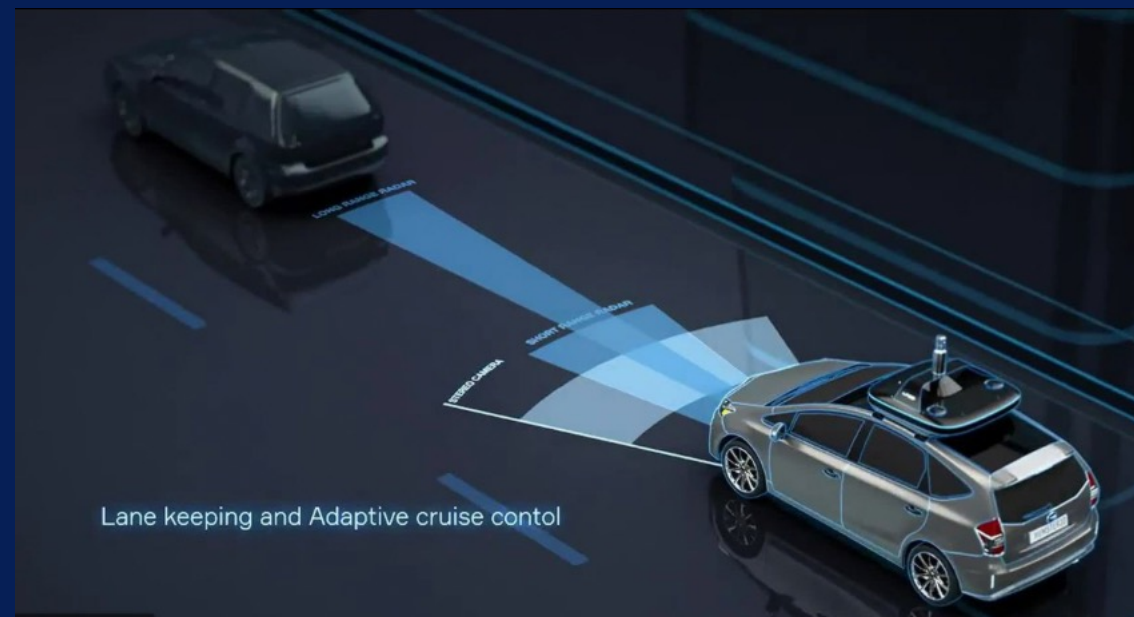
01

---

Computer Vision

# Computer Vision

시각적으로 세계를 해석하고 이해하도록 컴퓨터를 학습시키는  
인공지능 분야이다.



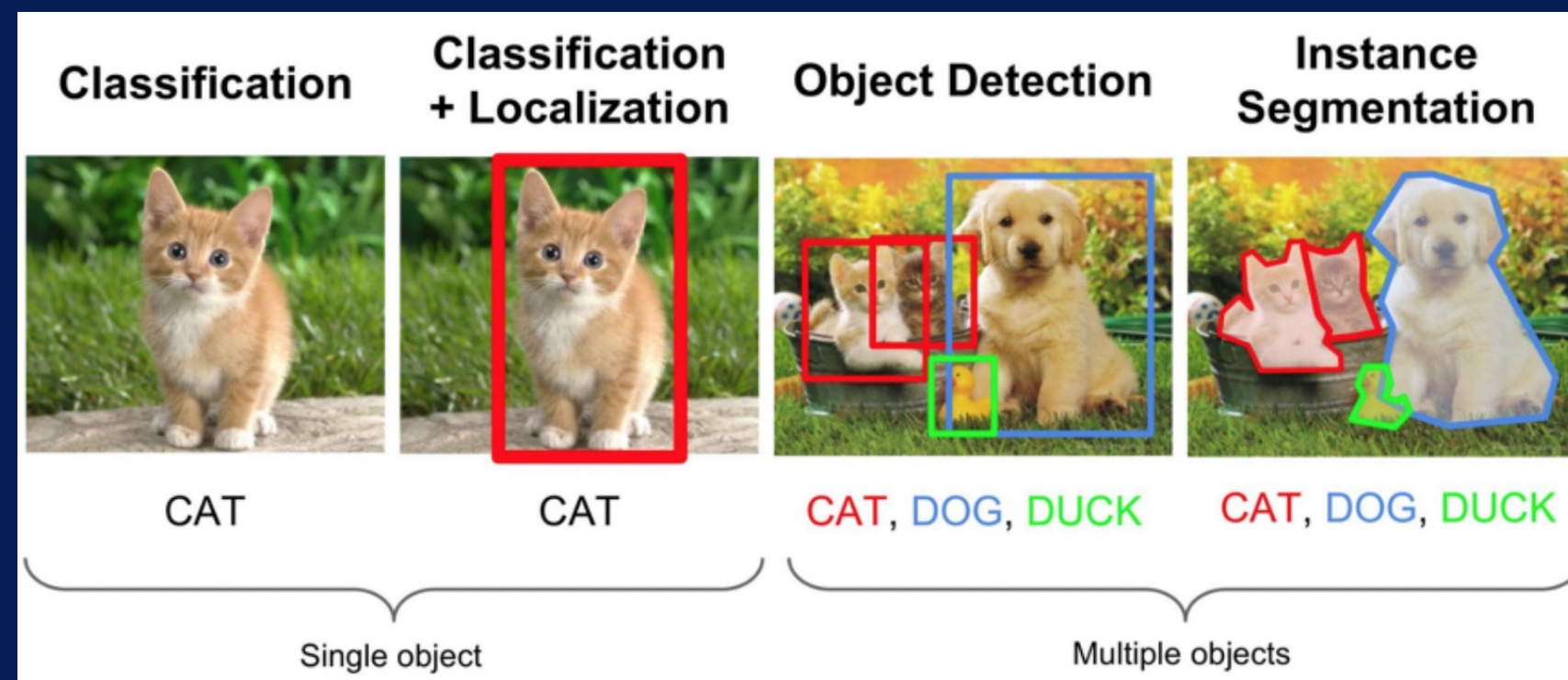
자율주행 자동차



얼굴인식

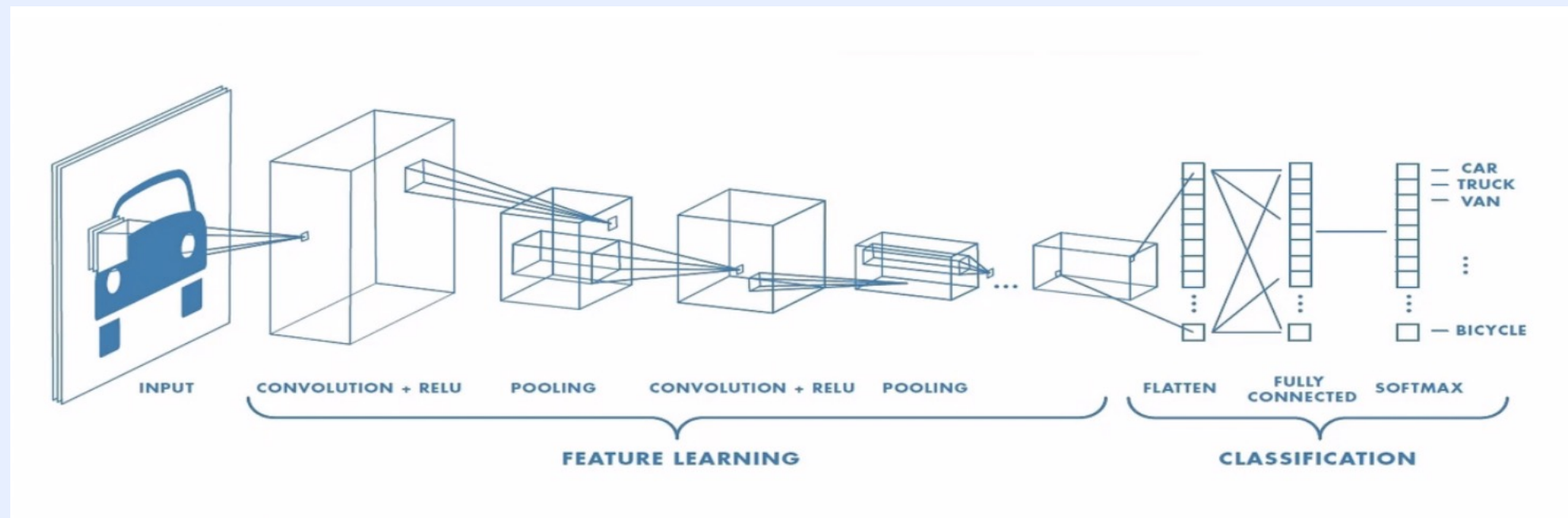
# Computer Vision의 주요 과제

1. Classification
  2. Object Detection / Localization
  3. Segmentation
- + GAN



# Convolution Neural Network(합성곱 신경망)

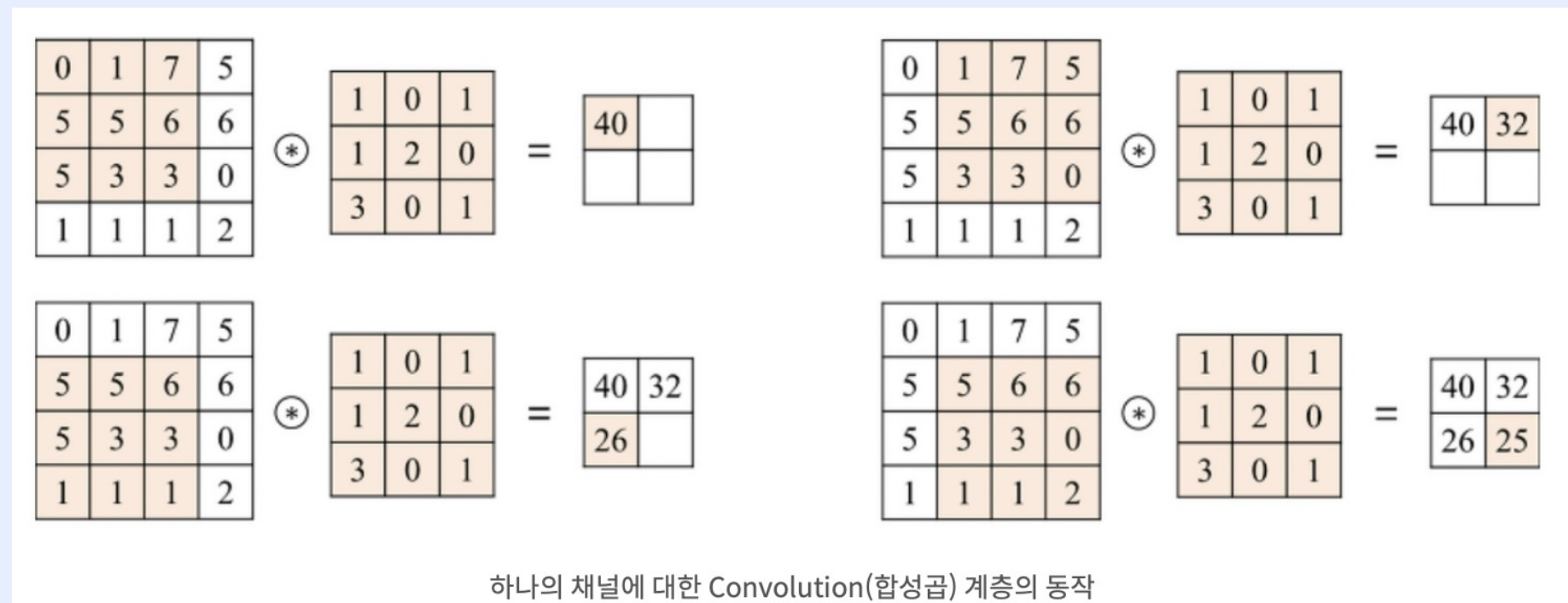
시각적 이미지를 분석하는데 사용되는 깊은 인공신경망(DNN)의 한 종류입니다.  
공간 정보를 유지한 상태로 학습이 가능한 모델입니다.



- 변환불변성 특징에 기초하여 이미지 및 비디오 인식, 이미지 분류 등에 이용된다.
- 다른 이미지 분류 알고리즘에 비해 상대적으로 데이터 전처리를 거의 사용하지 않는다.

# Convolution Neural Network(합성곱 신경망)

Filter = kernel (합성곱)



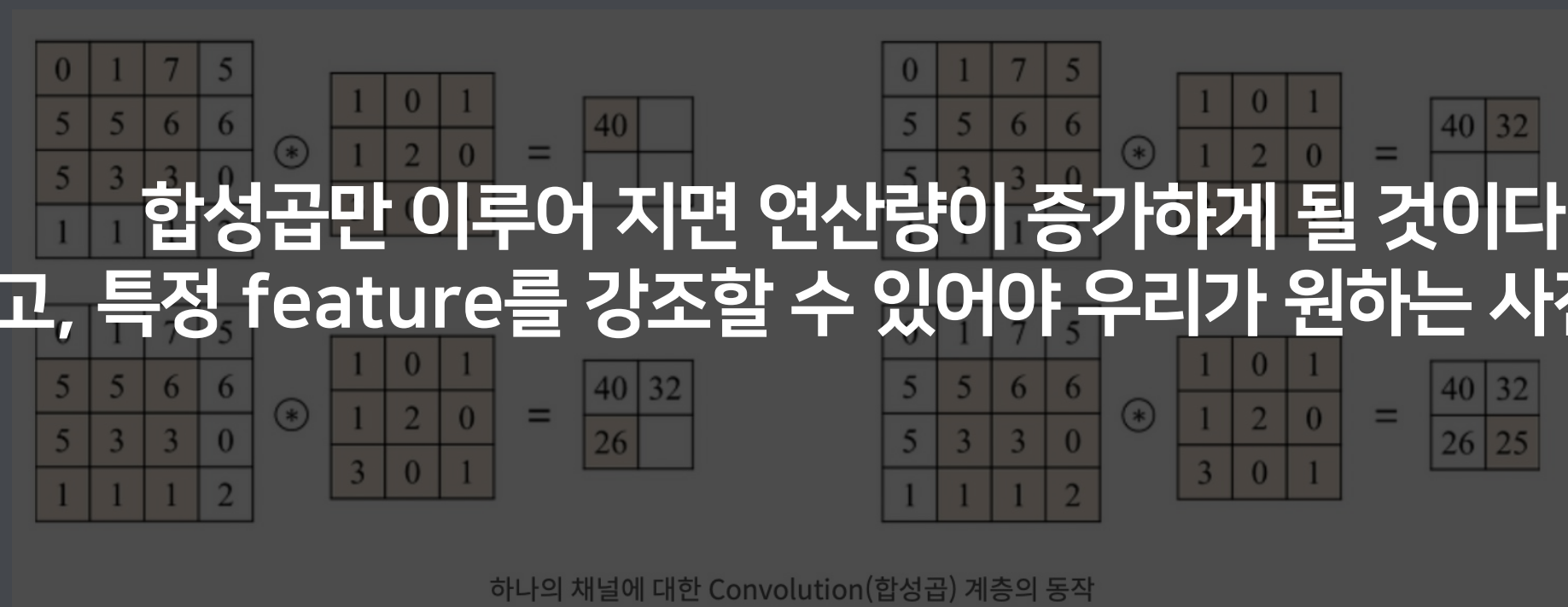
하나의 합성곱 계층에는 입력되는 이미지 개수 만큼 필터가 존재합니다.  
각 채널에 할당된 필터를 적용함으로써 합성곱 계층의 출력 이미지가 생성됩니다.



# Convolution Neural Network(합성곱 신경망)

Filter = kernel

## Filter 적용의 문제점!!!!



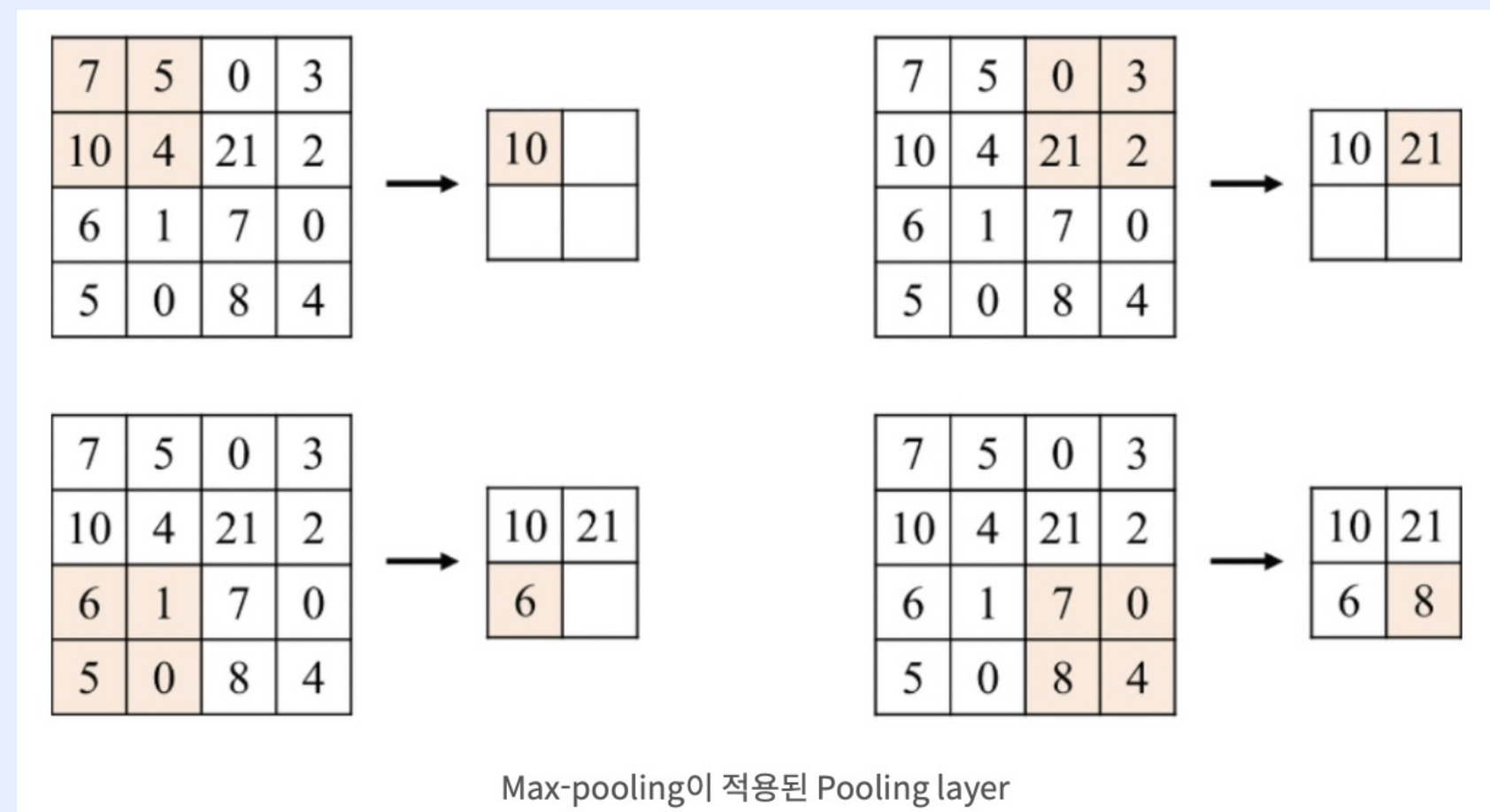
하나의 합성곱 계층에는 입력되는 이미지 개수 만큼 필터가 존재합니다.  
각 채널에 할당된 필터를 적용함으로써 합성곱 계층의 출력 이미지가 생성됩니다.



# Convolution Neural Network(합성곱 신경망)

## Pooling Layer

Feature map를 resize 하는 layer



CNN에서는 주로 Max-pooling을 사용하며, 이는 뉴런이 가장 큰 신호에 반응하는 것과 유사합니다.

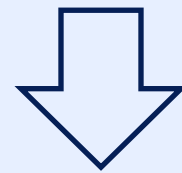
02

---

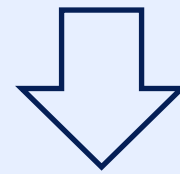
Efficient Net

# CNN의 확장

점점 연구가 활발해지면서 ConvNet이 확장된 모델이 대거 등장



하지만 이론으로 확장한 것이 아닌 그냥 실험에 입각해 ConvNet을 확장하는 모델만 존재



더 나은 정확성과 효율성을 달성할 수 있는 이론에 입각한 ConvNets을 확장하는 방법은 없을까?

# Efficient Net의 등장

## ConvNet 확장 방법

- depth (layer를 많이 쌓는다)
- width (convolution layer의 filter를 조절한다)
- image resolution (image의 높이, 너비를 조절한다.)

⇒ 기존에는 이 3가지 중 하나만 확장하는 것이 흔한 방법이었다.

# Efficient Net의 등장

## ConvNet 확장 방법

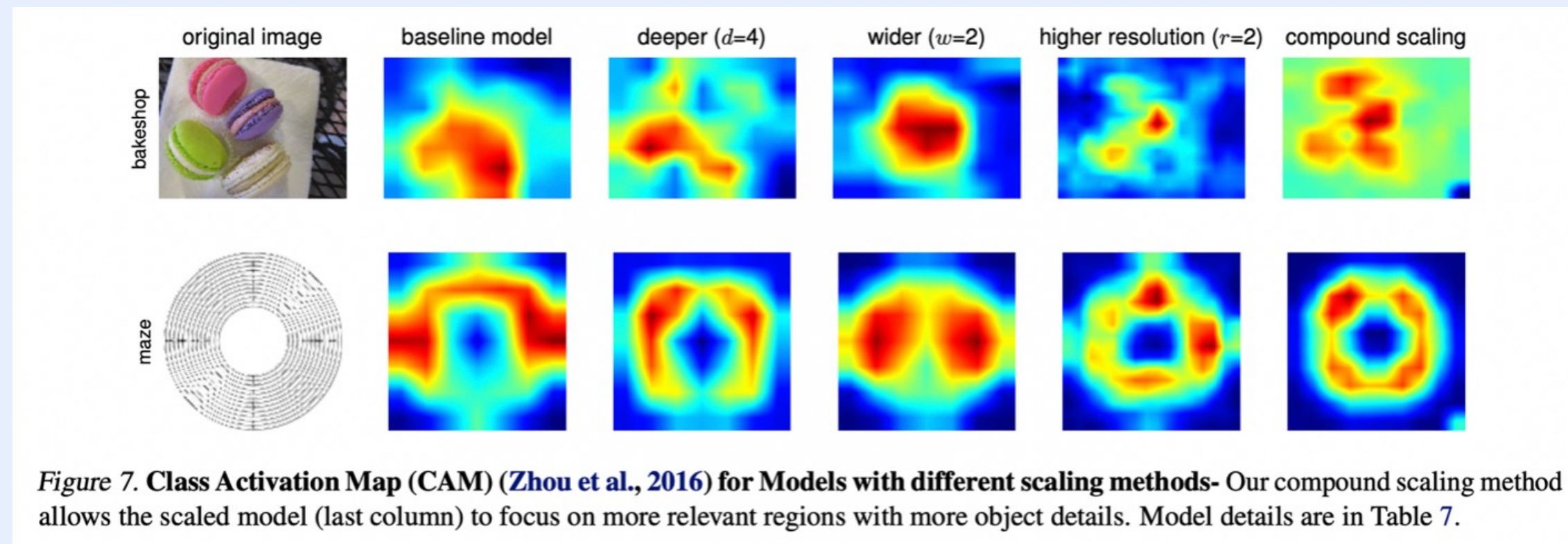
- depth (layer를 많이 쌓는다)
- width (convolution layer의 filter를 조절한다)
- image resolution (image의 높이, 너비를 조절한다.)

⇒ 기존에는 이 3가지 중 하나만 확장하는 것이 흔한 방법이었다.

💡 Network의 depth, width, resolution 사이의 균형을 맞추는 것은 성능에 있어 중요하며, 이들 간의 균형은 간단한 상수비로 구할 수 있다.

# Efficient Net

## <Class Activation Map>



Class Activation Map을 뽑아보면 depth, width, resolution을 각각 고려했을 때보다 동시에 고려했을 때 더 정교한 Class Activation Map을 얻을 수 있는 것을 확인할 수 있습니다.

03

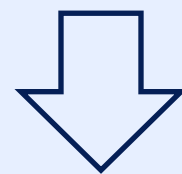
---

YOLO

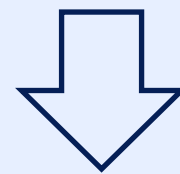


# YOLO 등장 배경

물체를 인식하는 정확도에 치우친 모델 대거 등장

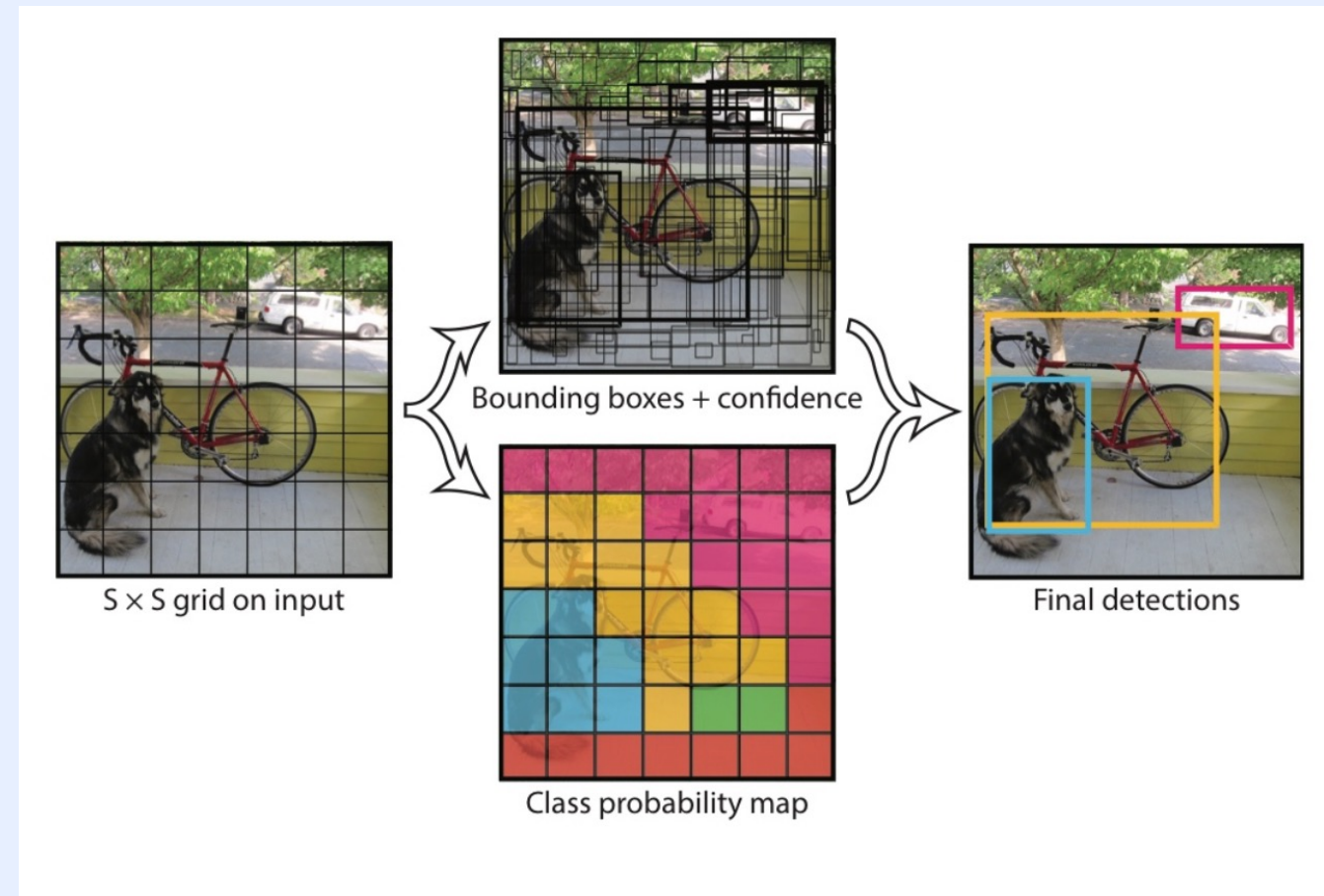


Yolo 등장 이전 딥러닝 모델은 물체를 인식하는데 시간이 오래 걸리는 문제점 존재



기존 모델에 근접한 정확도를 가지면서 같은 시간 내 더 많은 양의 이미지를 처리할 수 있는 실시간 객체 탐지 모델을 만들 수는 없을까?

# You Only Look Once (YOLO)



1. 입력 이미지를  $S \times S$  grid로 분할합니다.
2. 객체의 중심이 grid cell에 맞아 떨어지면 그 grid cell은 객체를 탐지했다고 표기합니다.
3. 각 grid cell은 각 추측된 바운딩 박스에 대한 confidence score를 예측합니다.
4. 예측한 바운딩 박스와 정답 박스간의 오차를 계산합니다.
5. 각 grid cell은 추가로 class 확률도 예측한다.
6. non - max suppression을 거쳐서 최종 바운딩 박스를 선정한다.

# You Only Look Once (YOLO)

## 장점

- 다른 기존 모델과 비교해 매우 빠른 속도를 가지고 있다.
- background error(배경 이미지에 객체가 존재한다고 잘못 탐지한 경우)가 적다
- 객체의 일반화된 representations을 학습하여 다른 도메인에서도 좋은 성능을 보인다.

## 단점 및 한계

- **작은 물체**에 대한 인식률이 떨어진다.  
(작은 물체에 대한 특성을 충분히 추출하지 못해서이다.)

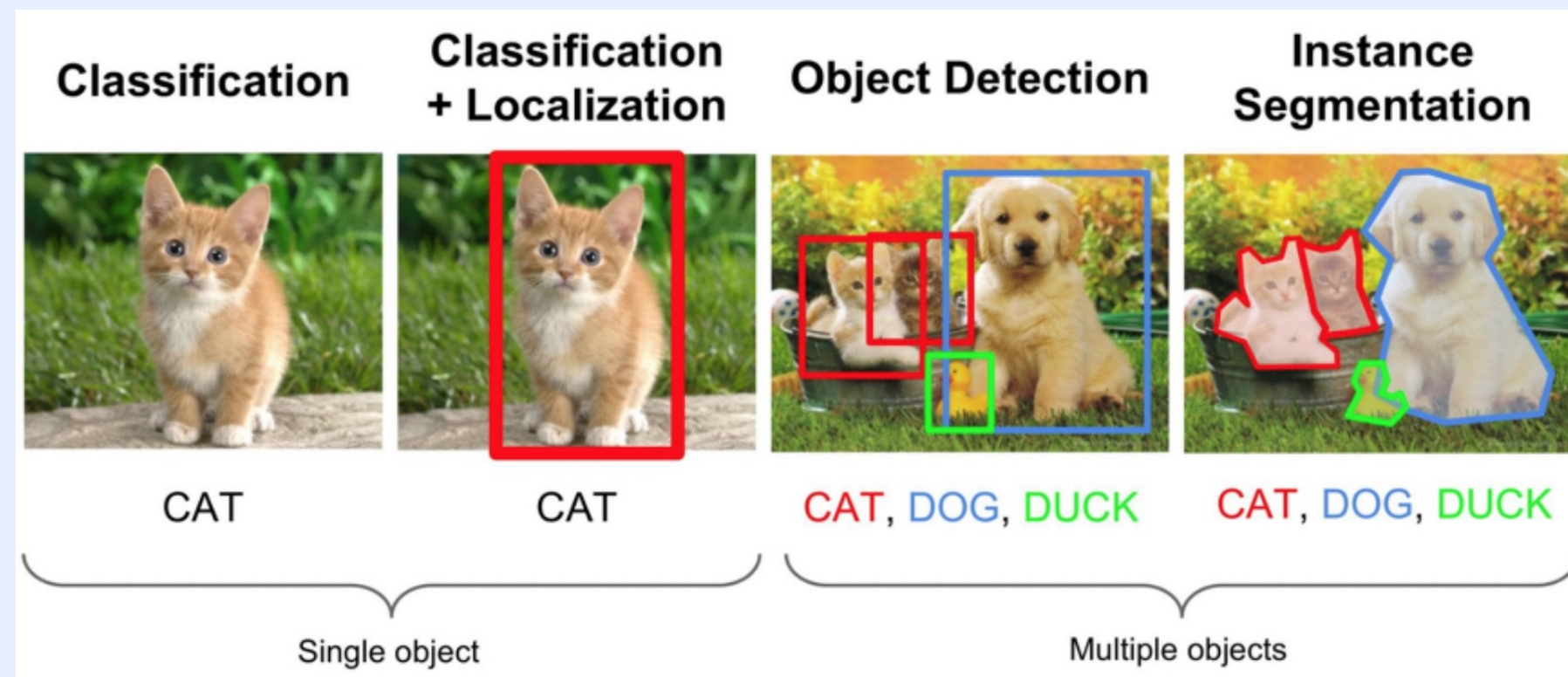
04

---

Mask R-CNN

# Computer Vision의 주요 과제

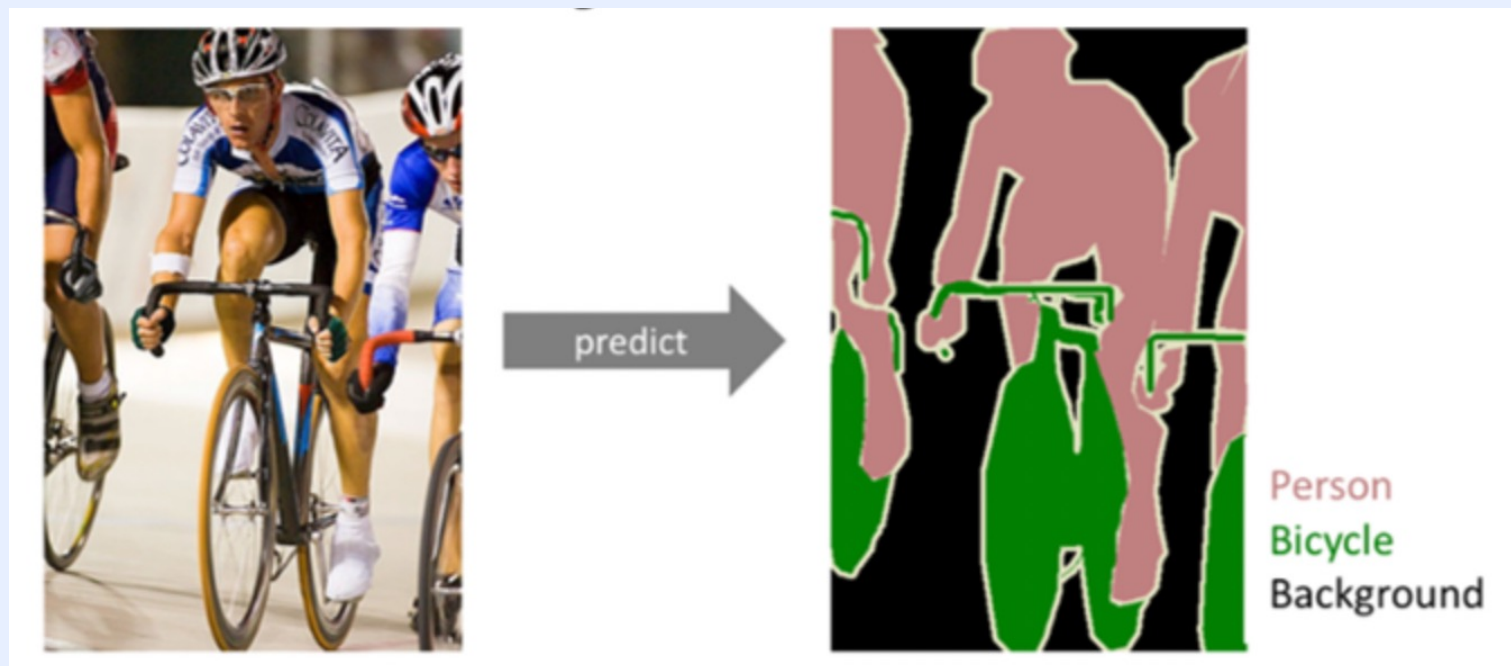
1. Classification
2. Object Detection / Localization
3. Segmentation



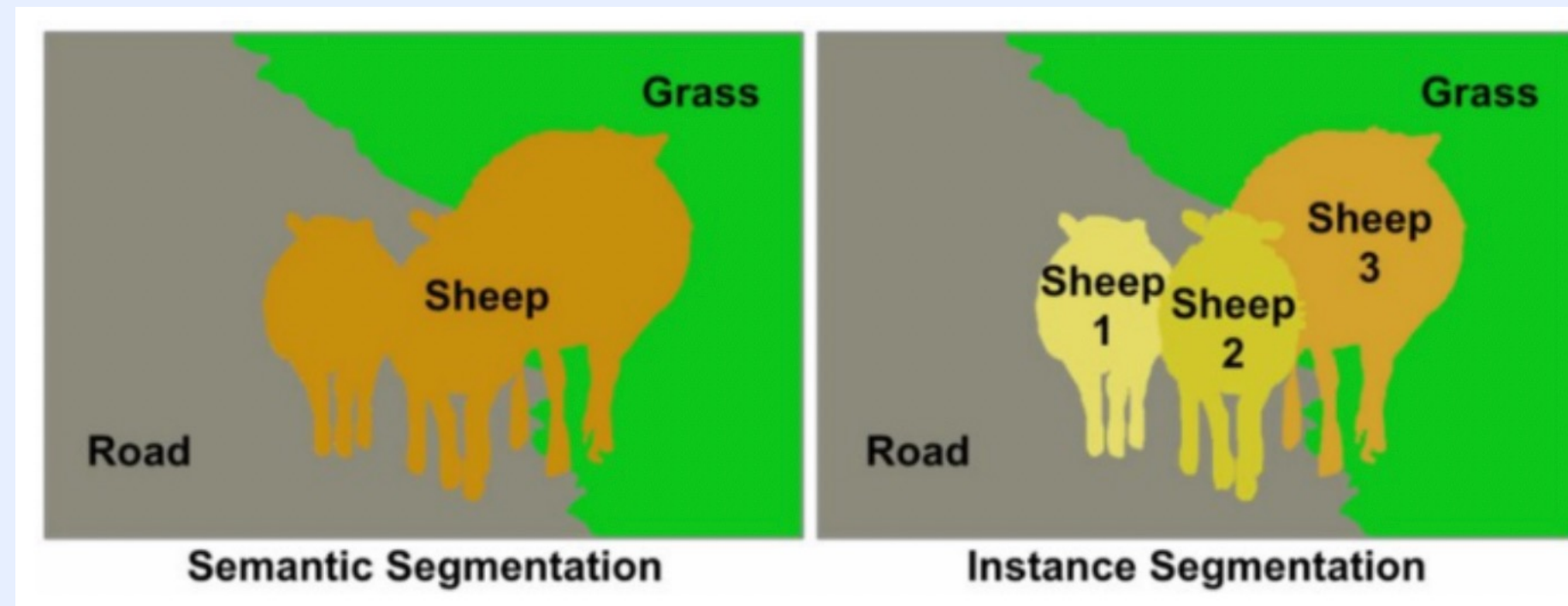


# Segmentation 이란?

Segmentation은 이미지에서 pixel 단위로 객체를 추출하는 방법입니다.



# Segmentation



## Semantic Segmentation

사진에 있는 모든 pixel을 해당하는 class로 분류하는 것입니다. semantic segmentation은 동일한 class의 객체는 따로 분류하지 않고, 그 pixel 자체가 어떤 class에 속하는 지만 분류합니다. 그래서 동일한 object는 같은 색으로 표현되며, 한 번에 Masking을 수행합니다

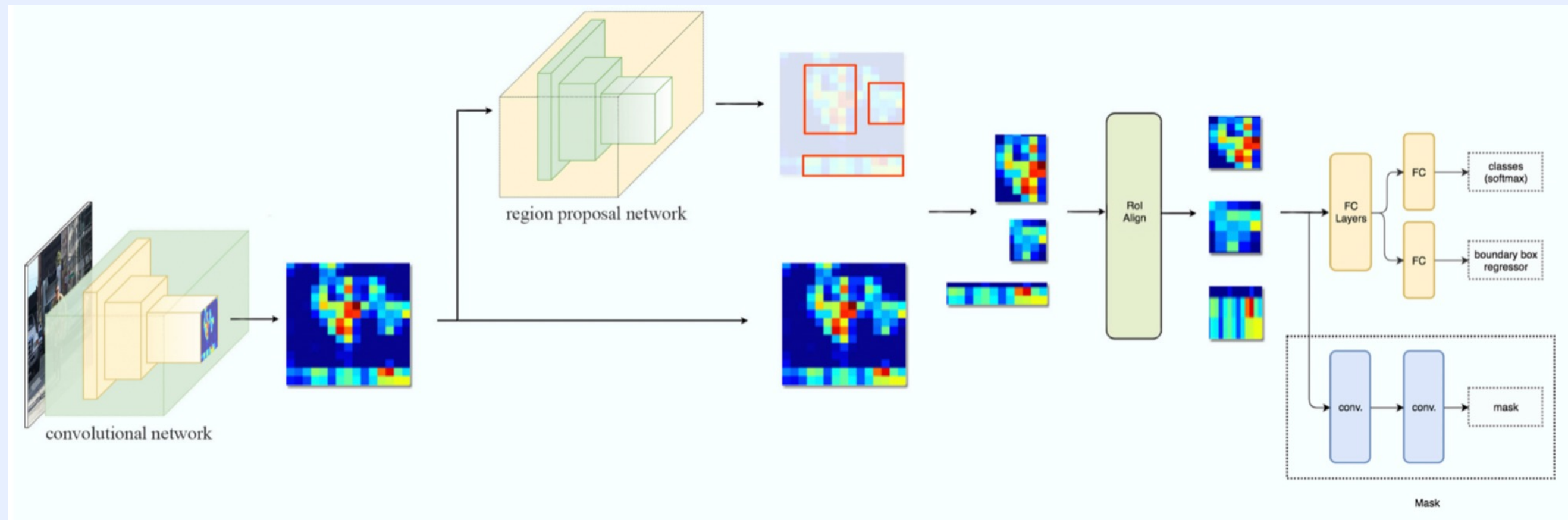
## Instance Segmentation

Semantic segmentation과 목적은 유사하지만, Semantic과 비교했을 때 각각의 같은 class의 object를 다른 label로 취급해 줍니다. 그래서 각자 객체별로 Masking을 수행하게 됩니다.



# Mask R-CNN

- Stage 1 : RPN (Region Proposal Network)
- Stage 2 : *in Parallel* predicting the **class and box offset** and **a binary mask for each RoI(Region of Interest)**



- 기존 Faster R-CNN 모델에 segmentation task를 예측하는 mask branch가 평행하게 추가된 구조입니다.
- Conv layer를 사용해 fully connected layer에 의해 output vector로 공간 정보가 붕괴되는 것을 막아 이미지 내 object에 대한 정보를 효과적으로 encoding 합니다.

# Mask R-CNN

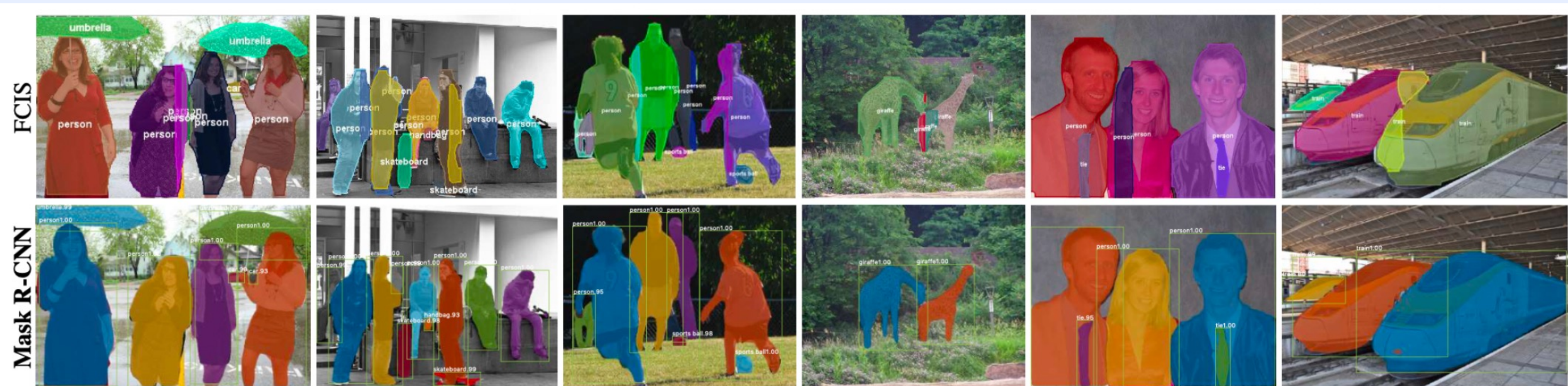


Figure 5. FCIS+++ [21] (top) vs. Mask R-CNN (bottom, ResNet-101-FPN). FCIS exhibits systematic artifacts on overlapping objects.

- 같은 사람이어도 서로 다른 class로 구별해서 인식하는 것을 확인 할 수 있다.
- 픽셀 단위로 구별하는 모델이기 때문에 작은 물체에 대한 인식이 잘 이루어 지는 것을 확인할 수 있다.





Q&A



# Thank you

A – CV - object detection & segmentation 이규리