

이미지 데이터 분석 (4일차)

# Object Detection & GAN

# 이미지 분류 vs. 물체 탐지

- 이미지 분류(image classification): What?
- 위치 추정(localization): Where?
- 물체 탐지(object detection): Where and What?

# 위치 추정 localization

- 이미지에서 대상 물체의 위치를 추정하는 것
- 이미지의 유형과 함께 경계 상자(bounding box)의 좌표를 함께 출력

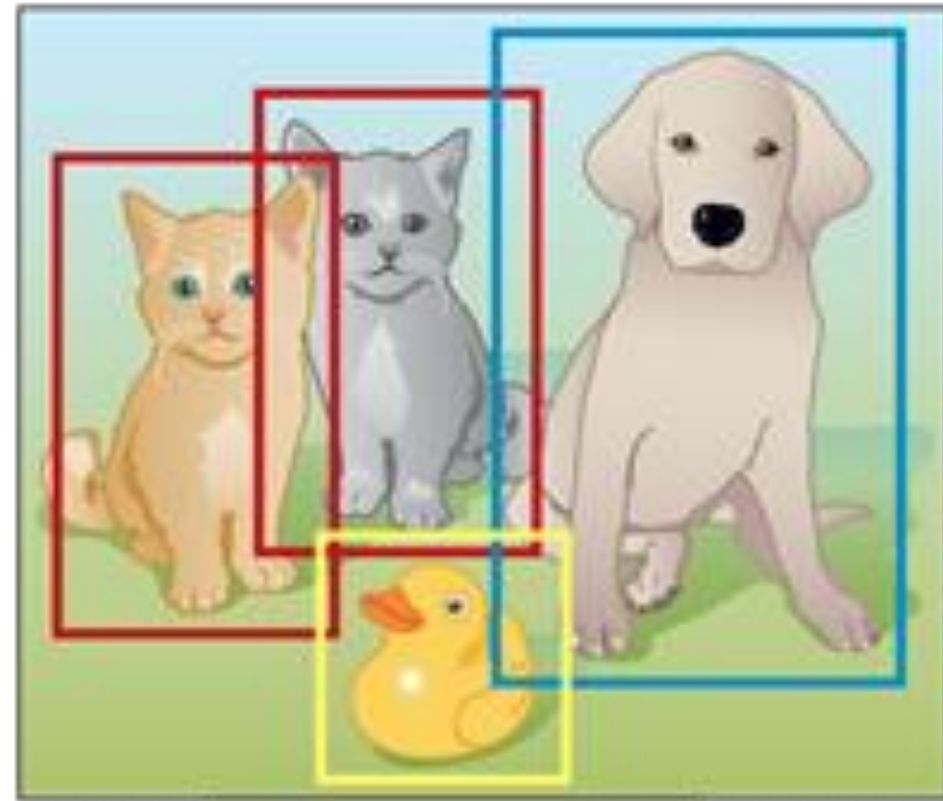
# Classification vs. Detection

Image classification



Cat

Object detection  
(classification and localization)



Cat, Cat, Duck, Dog

# 특징점 탐지 landmark detection

- 얼굴이나 자세의 경우 경계 상자보다 특징점을 탐지하는 것이 낫을 수 있음
- 특징점의 예: 왼쪽 눈꼬리의 위치, 오른쪽 팔꿈치의 위치 등

# 물체 탐지의 예측

- 경계 상자의 좌표 (연속 변수)
- 물체의 분류 (범주형 변수)
- 두 가지 예측의 손실을 모두 최소화 (multitask loss)

# 물체 탐지 모형

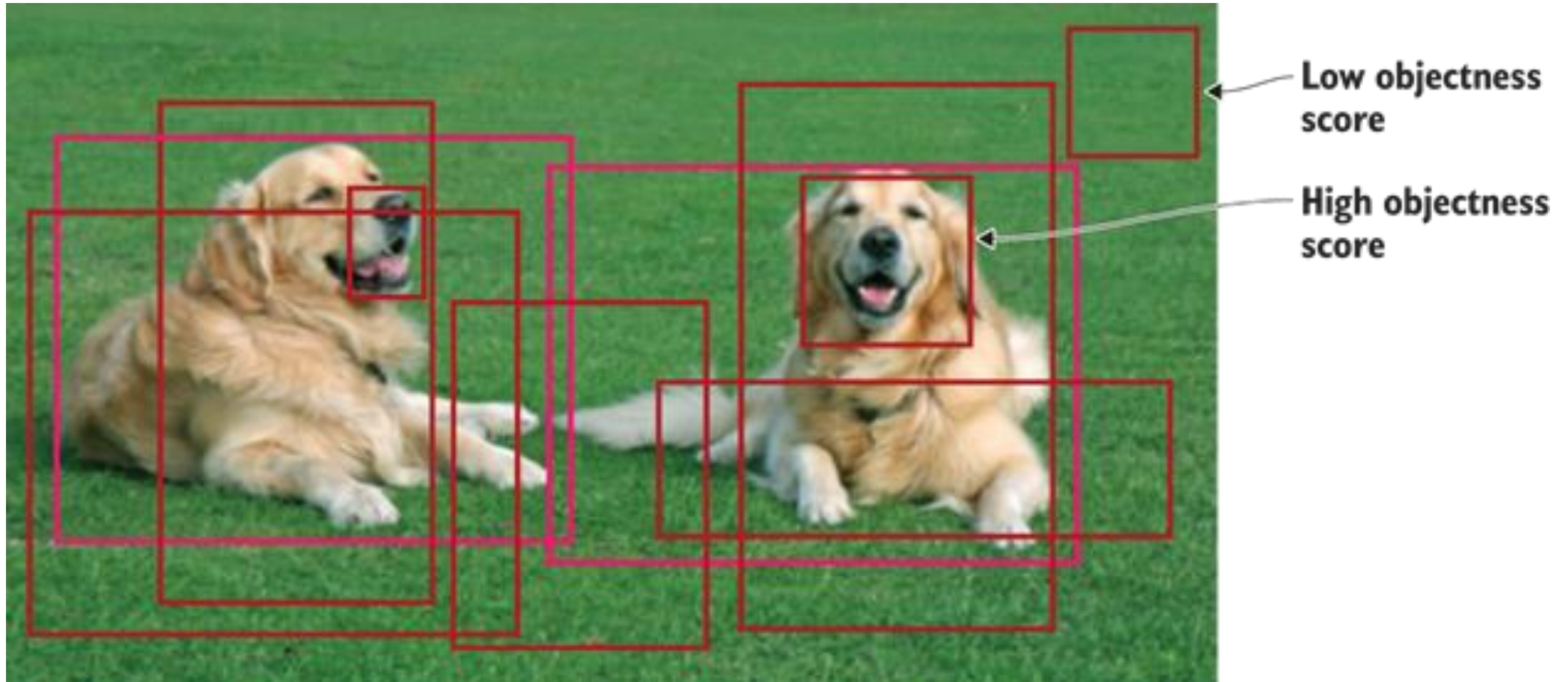
- 영역 제안과 물체 분류를 2단계로 처리:
  - R-CNN
- 1단계로 처리:
  - SSD
  - YOLO

# 물체 탐지 개요

- 1.영역 제안(Region proposal): 관심 대상인 물체가 있는 영역 (regions of interest, ROI) 을 제안
  - objectness score : 해당 영역에 물체가 있음직함의 점수
2. ROI의 물체를 분류
- 3.비-최댓값 억제(Non-maximum suppression, NMS): 동일한 물체가 여러 번 탐지될 경우, 가장 점수가 높은 영역만 남기고 나머지 영역은 억제
- 4.평가: mAP, PR curve, IoU 등



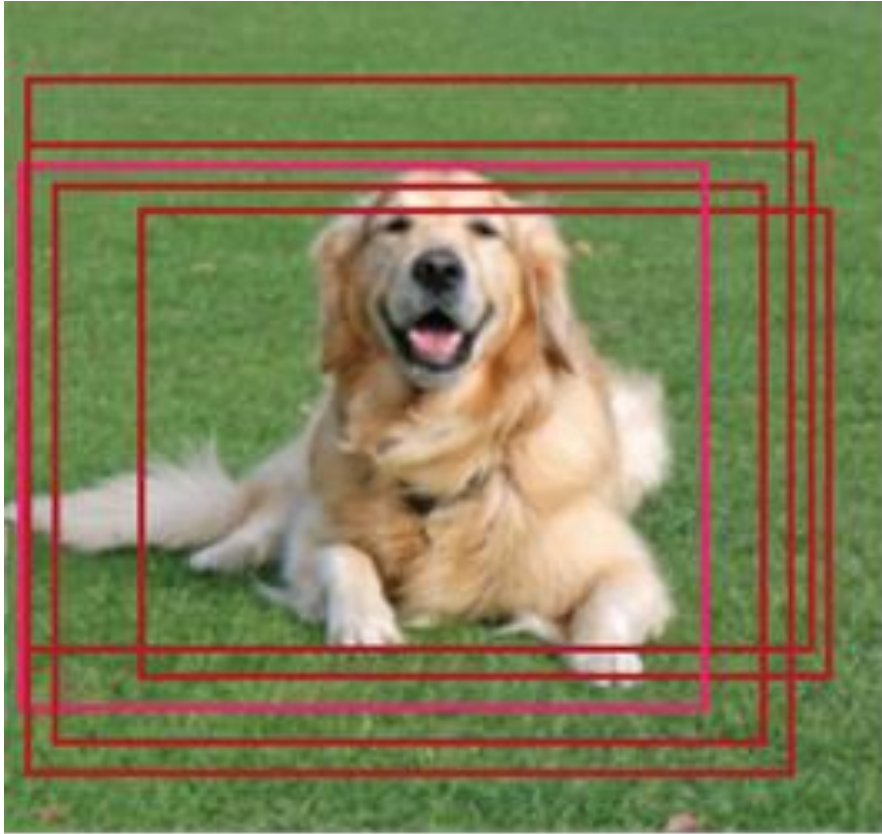
# Region Proposal



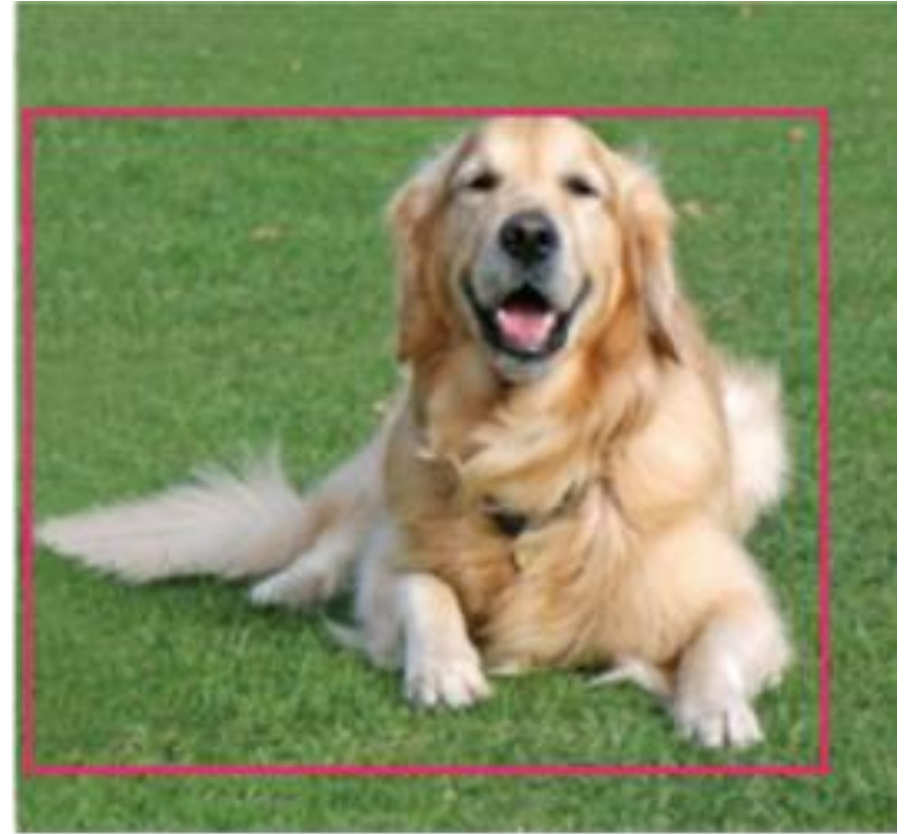
# 영역 제안 region proposal

- 슬라이딩 윈도우는 물체가 없는 위치에도 계산을 하므로 낭비
- R-CNN은 물체가 있을 법한 위치를 제안하는 모형
- 문제점: 영역 제안 자체가 계산이 오래 걸림

# Non-maximum suppression



Predictions before NMS



After applying non-maximum suppression

# 비-최댓값 억제 Non-max suppression

- 물체 탐지에서 동일 물체가 여러 번 탐지될 수 있음
- 확률이 높은 상자를 선택
- 선택된 상자과 많이 겹치는(IOU) 상자를 삭제

# Intersection over Union

- $IOU = \text{교집합} / \text{합집합}$
- 자카드 유사도(Jaccard similarity)라고도 함

# 물체 탐지 평가 지표

- 초당 프레임(frames per second, FPS): 처리 속도의 지표
- Intersection Over Union
- Precision-Recall Curve
- Mean Average Precision (mAP)

# IOU를 통한 평가



Predicted person bounding box

Ground truth person bounding box

Area of overlap

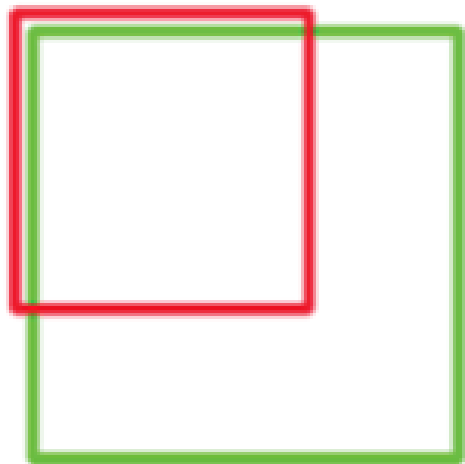
Score =

Area of Union



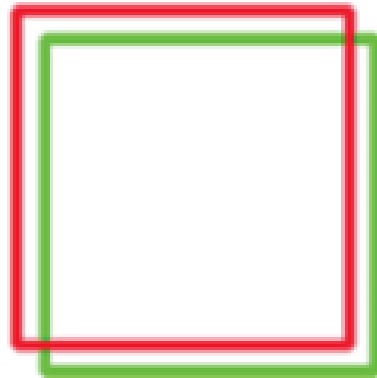
# IOU를 통한 평가

IoU: 0.4034



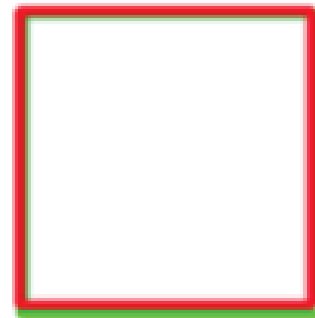
Poor

IoU: 0.7330



Good

IoU: 0.9264



Excellent

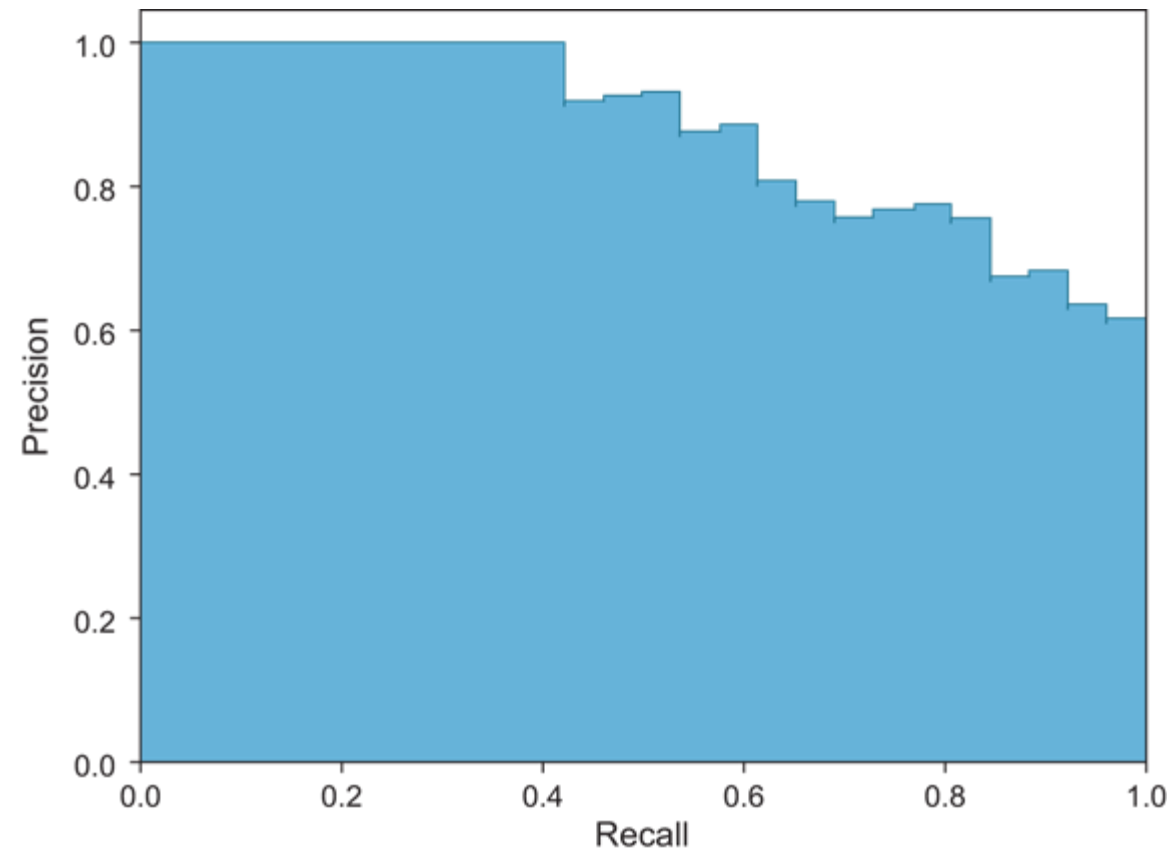


# Precision-Recall Curve

- 가로축을 Recall, 세로축을 Precision

PR 곡선의 곡선하 면적(AUC) =

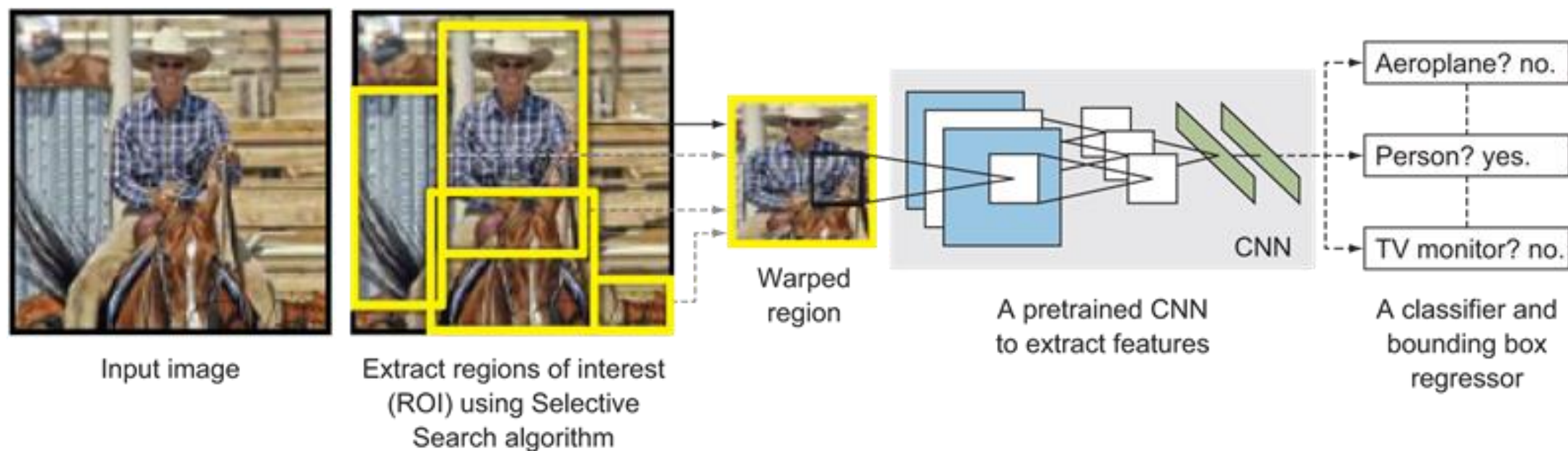
평균 Precision (Average Precision)



# mean Average Precision

- 모든 종류의 물체의 AP의 평균
- 물체 탐지에서 가장 많이 사용하는 평가 지표

# R-CNN



# R-CNN 작동 방식

- Selective Search Algorithm을 이용해 ROI를 추출
  - 후보 영역을 추출
  - 영역끼리 유사도를 계산
  - 비슷한 후보 영역을 합침
  - 유사도를 재계산, 이하 반복
- 추출된 이미지를 CNN에 입력하여 분류

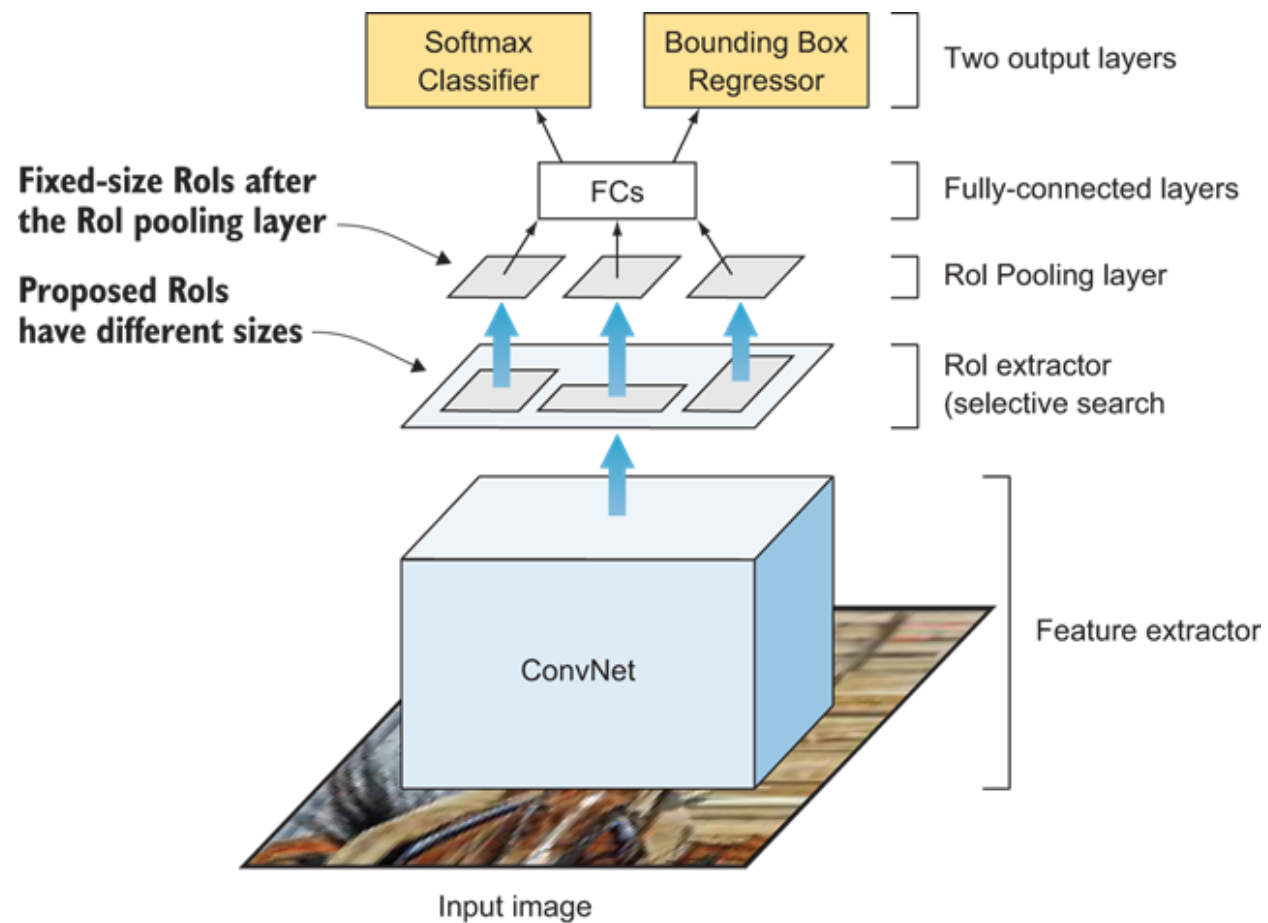
# R-CNN의 문제

- Selective Search Algorithm이 느리고 불필요한 영역을 제안
- 여러 단계로 나뉜 처리 과정
- CNN을 반복 적용 하므로 계산 비용이 높음

# Fast R-CNN

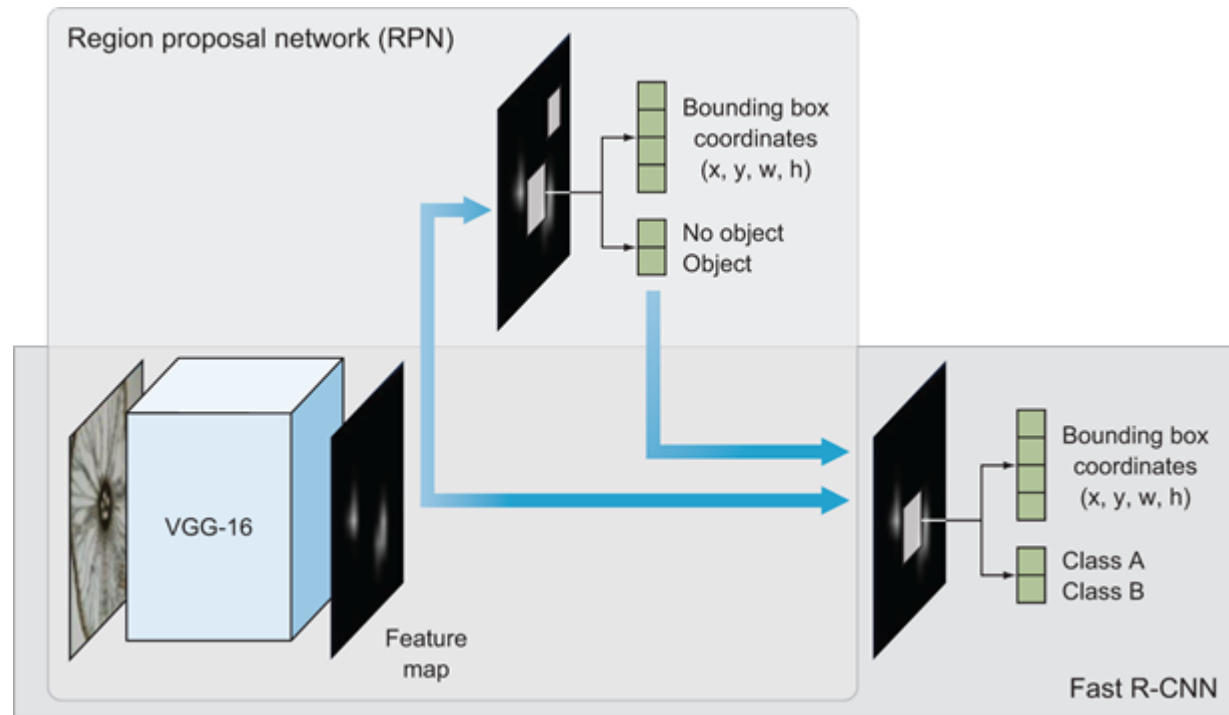
- 원본 이미지에서 영역 제안을 하는 대신, 이미지를 합성곱 신경망에 통과시킨 후 feature map에서 영역 제안
- 영역의 테두리에 대한 예측과 대상 분류를 모두 신경망으로 처리
- CNN을 반복적용하는 문제는 해결되었지만, 영역 제안에서 Selective Search의 느린 속도는 동일

# Fast R-CNN



# Faster R-CNN

- Selective Search 대신 영역 제안을 위한 신경망(RPN)을 사용





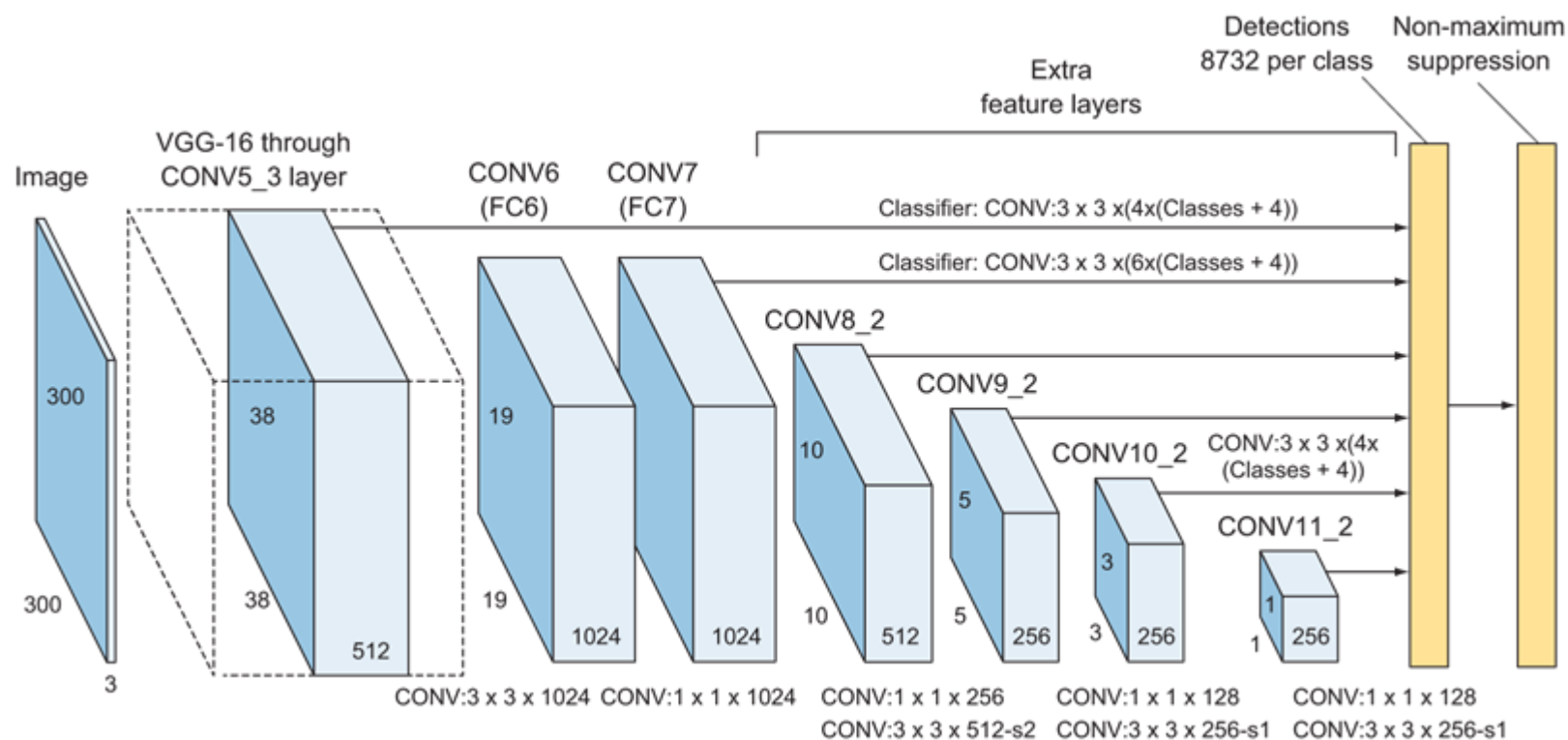
# Faster R-CNN의 한계

- R-CNN에 비해 250배 이상 속도 향상
- 2단계로 나뉘어져 있어 여전히 느리고 계산이 많이 필요
- 한 단계로 처리되는 모형들이 제안

# Single Shot Detection (SSD)

- 영역 제안과 물체 분류를 동시에 처리
- VGG16, ResNet 등의 이미지 분류 모형을 전이 학습
- 중간 feature map에서 매우 많은 경계 상자를 예측
- 큰 물체일 수록 후반부의 feature map이 예측

# Single Shot Detection (SSD)

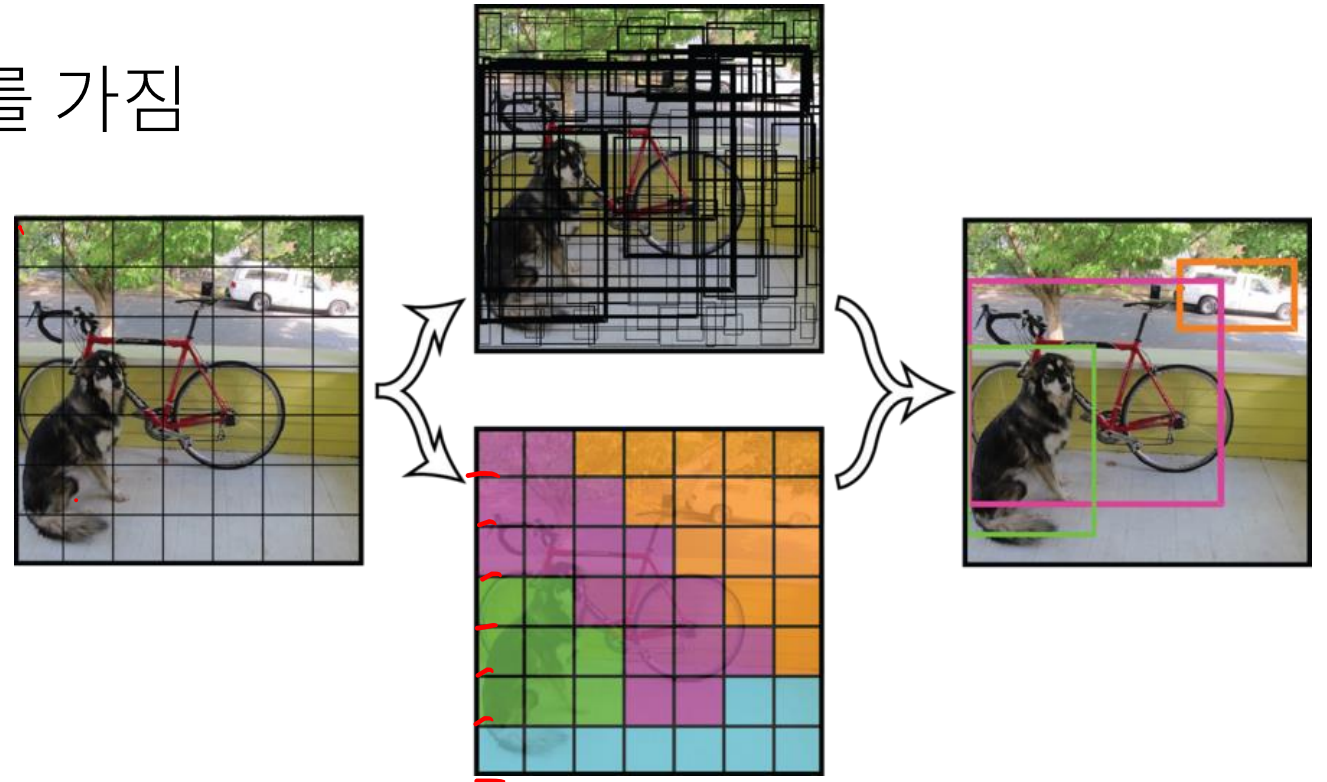


# YOLO You Only Look Once

- 처리 속도를 크게 향상
- 마지막 feature map을 하나만 사용
- 상대적으로 적은 수의 경계 상자만 만들

# YOLO You Only Look Once

- 이미지를 일정한 크기의 격자로 나눔
- 격자 한 칸이 하나의 경계 상자를 가짐
- 격자마다 물체를 분류
- 두 결과를 합쳐서 물체 탐지



# YOLO의 한계

- 한 이미지에서 탐지할 수 있는 물체의 수에 한계
- 작은 물체나 여러 물체의 무리도 잘 탐지 안됨

# Generative Adversarial Networks

- Ian Goodfellow et al. (2014)
- "최근 10년간 기계학습에서 가장 흥미로운 아이디어"(Yann LeCunn)
- 생성 모형: 이미지, 비디오, 음악, 텍스트들을 생성할 수 있음

# GAN으로 생성한 이미지



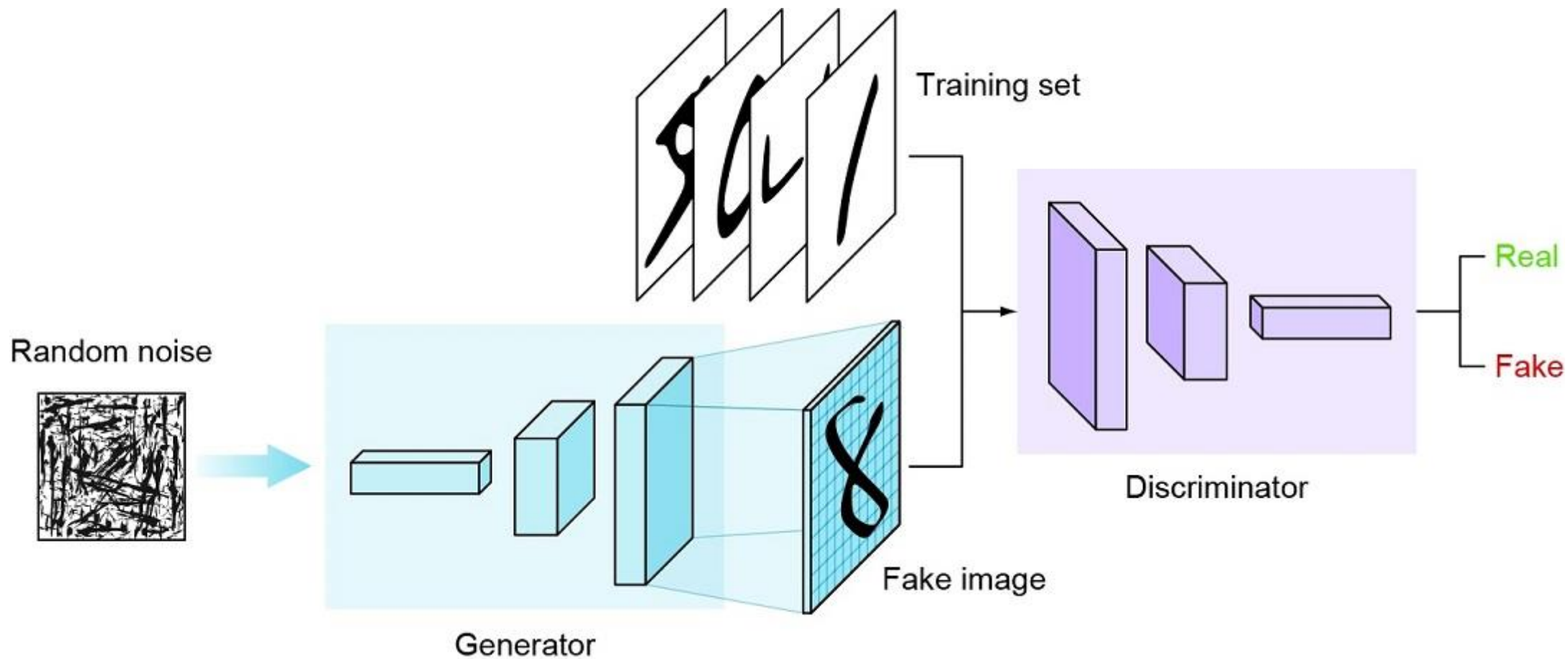
<https://generated.photos/faces>



# GAN의 기본 구조

- 두 개의 경쟁적 모형으로 구성
- 생성망(generator): 이미지를 생성
- 구분망(discriminator): 생성된 이미지와 실제 이미지를 구분
- 생성망은 구분망을 속이도록 학습

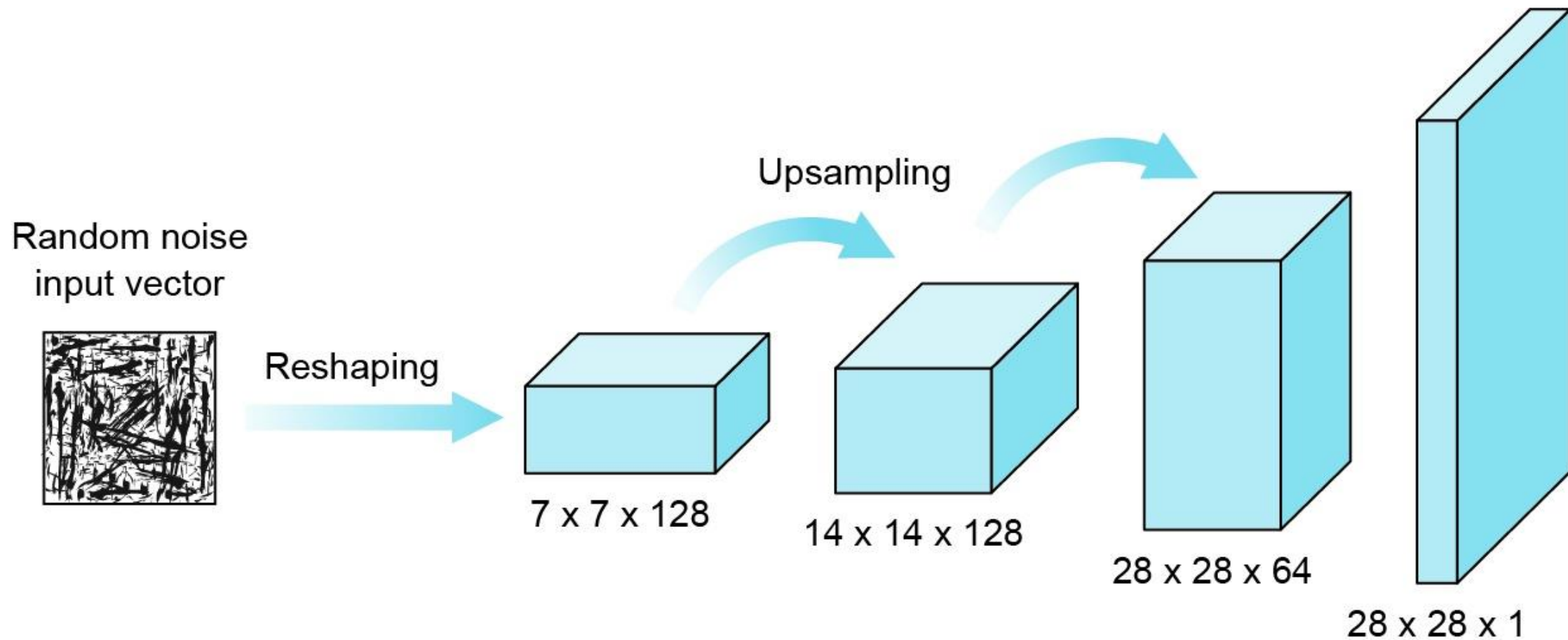
# GAN의 기본 구조



# Deep Convolutional GANs (DCGANs)

- 합성곱 레이어를 사용한 GAN
- Radford, et al. (2016) Unsupervised representation learning with deep convolutional generative adversarial networks
- 현재는 모든 GAN 모형이 합성곱 레이어를 사용

# 생성망의 기본 형태



# 다운샘플링과 업샘플링

- 다운샘플링: 이미지를 압축하는 것
- CNN에서는 이미지가 처리되면서 추상화된 정보로 다운샘플링
- 업샘플링: 다운샘플링의 역과정
- CNN에서는 추상화된 정보를 업샘플링하여 이미지화

# CNN에서 업샘플링

- 언풀링(unpooling): 풀링의 역과정
- 전치 합성곱(transpose convolution): 합성곱의 전치 형태

# 언풀링 Unpooling

1	2
3	4




# 최근접 이웃 Nearest Neighbor

1	2
3	4



1	1	2	2
1	1	2	2
3	3	4	4
3	3	4	4



# 지압판 Bed of Nails

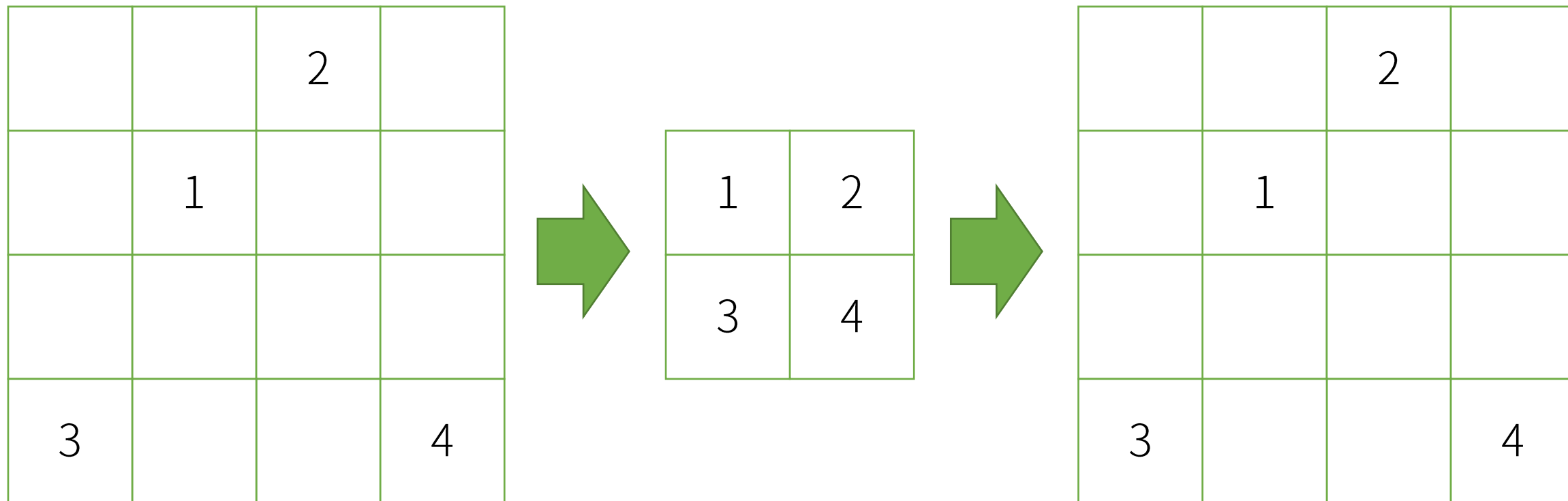
1	2
3	4



1	0	2	0
0	0	0	0
3	0	4	0
0	0	0	0

# 최대 언풀링 Max Unpooling

- 최대 풀링(max pooling)했을 때의 위치를 기억해서 똑같은 위치에 언풀링



# 전치 합성곱 Transpose Convolution

- 합성곱: 필터와 픽셀을 각각 곱한 후 더하여 하나의 값으로 만듦
- 전치 합성곱: 하나의 값을 필터에 곱하여 각각의 픽셀에 대입


# GAN의 학습

- 구분망은 생성된 이미지를 0, 실제 이미지를 1로 예측하도록 학습
- 구분망의 학습을 중단
- 생성망으로 생성한 이미지를 구분망에 입력
- 구분망의 출력을 손실함수로 생성망을 학습

# 미니맥스 게임

- 2인의 제로섬 게임. 한 쪽은 최대화, 다른 쪽은 최소화를 시도
- GAN에서 구분망은 최대한 잘 구분하는 것을, 생성망은 구분을 최소한으로 하게 하는 것이 목적

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}} [\log D(x)] + E_{z \sim p_Z(z)} [\log(1 - D(G(z)))]$$



Discriminator output  
for real data  $x$

Discriminator output  
for generated fake data  $G(z)$

# GAN의 어려움

- Mode collapse: 비슷한 이미지만 생성됨
- Diminished gradient: 구분자의 성능이 너무 빨리 좋아지면 생성자의 학습이 불가능
- 학습의 불안정: 학습이 잘 되지 않거나 설정에 따라 학습 결과에 차이가 큼

# GAN의 평가

- 사람에 의한 평가: 평가 방식, 평가자에 대한 피드백에 따라 결과가 달라짐
  - Goodfellow et al. (2016) Improved Techniques for Training GANs
- Inception score
- Fréchet Inception Distance
- 이외도 영역 특수적인 평가 방법들이 있음

# Inception Score

- ImageNet 등에 사전 학습된 모형에 생성된 이미지를 입력
- 다음 두 가지 특성을 요구:
- 각 이미지는 여러 분류 중 특정의 확률이 높음 (실제와 비슷함)
- 생성된 이미지들이 여러 분류에 고르게 분포될 것임 (다양함)



# Fréchet Inception Distance (FID)

- Heusel et al. (2017) GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium
- 사전 학습된 모형에 실제 이미지를 입력했을 때의 활성화 분포와 생성된 이미지를 입력했을 때의 활성화 분포를 프레스 거리(프레셰 거리)를 이용해 비교

# GAN의 활용

- Text-to-Photo Synthesis
- Image-to-image translation (Pix2Pix GAN)
- Image Super-Resolution GAN (SRGAN)

# Text-to-Photo Synthesis

- StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks
- 텍스트를 바탕으로 사진과 같은 이미지를 생성
- 저화질 이미지를 먼저 생성 후, 고화질 이미지로 생성

# Text-to-Photo Synthesis

This bird is white with some black on its head and wings, and has a long orange beak



(a) StackGAN  
Stage-I  
64x64  
images

This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face



This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments



(b) StackGAN  
Stage-II  
256x256  
images



# Image-to-image translation (Pix2Pix GAN)

- 여름 사진, 겨울 사진과 같이 짝을 이루는 데이터를 사용
- 생성망은 실제 이미지 중 한 쪽을 입력하면, 다른 쪽 이미지를 생성
- 구분망은 두 이미지를 입력 받아 두 번째 이미지의 실제/생성을 구분

# Image-to-image translation (Pix2Pix GAN)

BW to Color



input

output

Edges to Photo



input

output

Labels to Street Scene



input

output

Day to Night



input

output

# Image Super-Resolution GAN (SRGAN)

original



SRGAN  
(21.15dB/0.6868)

