

AI+X 고급 (5주차)

# Evaluation of Object Detection

# 지난학기

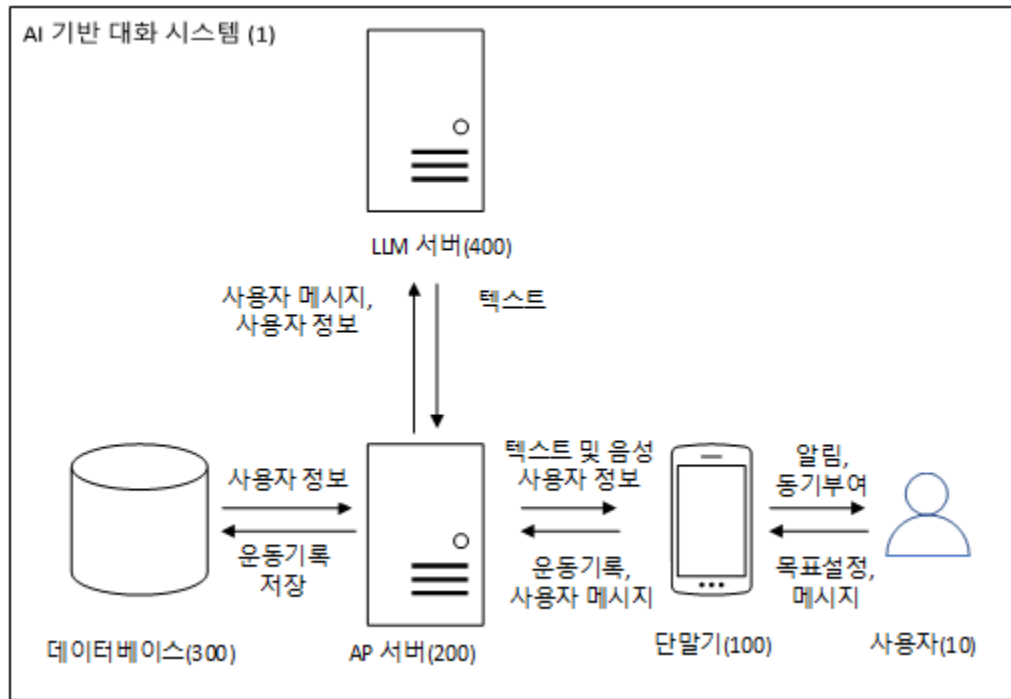
논문 2件

Student Draft 완성

“I will take it from here.”

# 지난학기

## 특허 준비 진행 중



# 이번학기 과제

# The Plan (1/2)

9/8 Intro

9/15 6 Ideas Presentation (15% 결과: Pass)  
(KT 교수님 지도)

9/22 Theme 확정, 데이터 수집 검토

※ 데이터 수집계획(案) (어떤 데이터를 어떻게?) 9/29(금)까지 제출 (제출 완료)

9/29 추석

# The Plan (2/2)

10/6 데이터 수집 및 전처리 착수, 모델링 착수

10/13 모델링 지속, 중간발표 가이드, 성능평가 강의

10/20 데이터 및 모델링 점검 (KT 교수님 지도)

10/27 모델링 개선 및 Back/Front 개발 착수

11/3 **중간고사 발표** (KT 교수님 지도) (20%)

# 중간발표 평가항목

- 컨셉, 시나리오 및 기능 정의
- USP (기존 서비스 대비) 도출 및 BM 기획
- 데이터 수집, 전처리, 모델 개발
- 모델성능 평가
- 추후 계획 (B/F 개발 포함)

# 과제현황 및 고민 (1/3)

## 1. 진행상황공유

- GPU설정을 꼭 먼저 해야 함
- 버스 정류장의 형태가 쉘터형과
- 거리계산 알고리즘 추천 논문
  - § 추가
  - § 이의철교수님 왁
  - § 고정된구도와 이미지면 거리계산 가능하지만 사용자가 찍어서 하는 방법이면 거리계산이 어려울 것이다.
  - 거리계산은 포기하는게 정신건강에 이로울 것 같다.
- 버스정류장 크롤링 진행해보았다 → 하다보니 버스정류장의 형태가 온전하지 못하게 나오는것 같다.
- 한국버스정류장 데이터가 많이 없다(AI HUB랑 로보플로우를 합쳐서 사용해야 할 듯)



# 과제현황 및 고민 (2/3)

- 지하철역 출구데이터가 문제
  - § 지하철역 이름으로 디텍션을 하는 방법으로는 안되나?
  - § 지하철역앞 보도블록
- **키크보드 2200장 돌려봤을때는 잘 탐지하는것 같다.**
- 일단 지하철역과 버스정류장 디텍션을 하고 추가로 점자블록과 횡단보도를 하는것이 베스트(특허와의 차별성)

## 2.정해야 할 일

- 역할나누기
  - § 데이터 수집 -
    - § 지하철역 (최소 100장 최대한 개별적인 역 위주로 - 기둥이나 간판이 나오게)
    - § 1호선 - 유진 1000부터
    - § 2호선 - 만서
    - § 3호선 - 진석

# 과제현황 및 고민 (3/3)

§ 4호선 - 우성 4000부터 시작

§ 7호선 - 수환

§ 6호선 - 종한

→ 금요일까지 최대한 해보기

§ 버스정류장은 → **지하철 해보고 나서**

§ 전처리 및 수집

§ 모델링

○ 분류모델을 안쓰 이유

§ 신고를 할때는 불법인 경우의 사진을 올리기 때문에?

§ 불법이 아닌경우에 대한 것 - 자동이면 필요한가?

# Project Scope

1. 컨셉 구체화 및 기능 및 서비스 정의
2. 데이터 수집 및 전처리 로드가 큼  $\pi$
3. AI 모델링
4. Backend/Frontend 개발

주어진 것은 한 학기

# Scoping

1. PM은 **킥보드로 한정**
2. 불법주차 위치가 실제로는 다양하지만  
→ 1~2개로 선정

# 개발순서 (지난주)

1. YOLO 통한 키포드 Object Detection 개발 (Model 1)

2. 합법/불법 분류 모델 (Model 2)

- **불법위치**를 인식하는게 제일 중요!
- 여러가지 불법주차 위치에서의 사진을 찍어 학습하기엔 시간이 많이 걸리므로,  
Roboflow  
**1~2개 불법위치만 선정**하고 사진 모우고 증강 및 학습
- 핵심은 1~2개가 빈번하고 자주 “불법 ” 하는 곳이어야함  
*줄인다고 아무거나 선정하면 안됨*

# 개발순서 (지난주)

## 2. 합법/불법 분류 모델 (Model 2)

- 불법위치를 인식하는게 제일 중요!
- 여러가지 불법주차 위치에서의 사진을 찍어 학습하기엔 시간이 많이 걸리므로,  
1~2개 불법위치만 선정하고 사진 모우고 증강 및 학습
- 핵심은 1~2개가 **빈번하고 자주 “불법 ” 하는 곳이어야함**
  - . 버스정류장이나 지하철입구 등 사람들이 많이 이용하는데 위치한 키포드를 찾아 사진 찍고 이미지 학습, 이미지 불법/합법 분류
  - . AND/OR 위도경도로 지하철위치, 버스정류장위치 찾아 불법 여부 확인
  - . 키포드와 불법위치가 포함된 사진을 모으는게 핵심
- CF) 키포드와 불법위치 포함된 사진에서 키포드와 불법위치 둘 다 객체탐지 되면 좋는데 쉽지 않음

# 개발순서 (지난주)

3. 사진을 업로드 → 1. 키패드 Detection 모델  
2. 합법/불법 분류 모델

4. 키패드 신고사이트 자동 연동 (서비스 기획 중)

# 피드백

- 불법주차 전체 탐지에서 제한한 것은 현실적인 의사결정
  - 지하철 및 버스정류장에서 **지하철만으로 한정**하는 것도 고려
  - 지하철 사진 수집, BBox 작업, 라벨링도 많은 시간과 노력
  - **Fallback Plan**: 분류 모델로 지하철인지 아닌지
- 현재 있어야 할 지점
  - . 어떤 데이터를, 어떤 목적으로, 어떻게 모델링하겠다 라는 안(案) 나와 있어야 함



# 다음 주

10/6 데이터 수집 및 전처리 착수, 모델링 착수

10/13 모델링 지속, 중간고사 발표 가이드

10/20 데이터 및 모델링 점검 (KT 교수님 지도)

10/27 모델링 개선 및 Back/Front 개발 착수

11/3 **중간고사 발표** (KT 교수님 지도) (20%)



# 성능지표

# Object Detection

## (객체검출)

## Performance

## Criteria

# mAP

(Mean Average  
Precision)

Common Metric for  
Evaluating Object Detection Models

# mAP

The average precision  
(AP) for all classes

# AP

The area under the  
**precision–recall curve**  
for each class

# IoU (Intersection Over Union)


How close the predicted bounding box is to the ground truth bounding box.





**GT와 모델이 예측한 값이 얼마나 겹치는가?**

**에 대한 지표**

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


The diagram illustrates the calculation of the Intersection over Union (IoU) metric. It consists of two parts. The top part shows two overlapping squares: one with a blue outline and one with a solid blue fill. The bottom part shows the union of these two squares as a single solid blue shape. The formula for IoU is given as the ratio of the Area of Overlap to the Area of Union.

실제 Box와 예측한 Box의

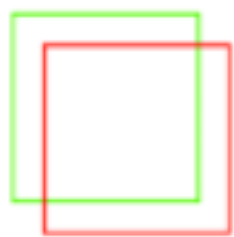
교집합(분자) / 합집합(분모)

Sample IoU scores

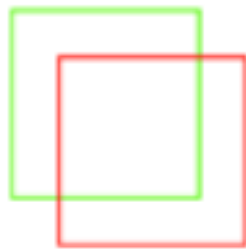
0.905



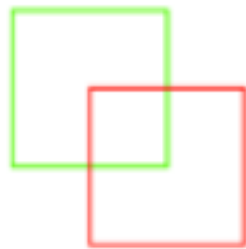
0.532



0.391



0.143



0.0

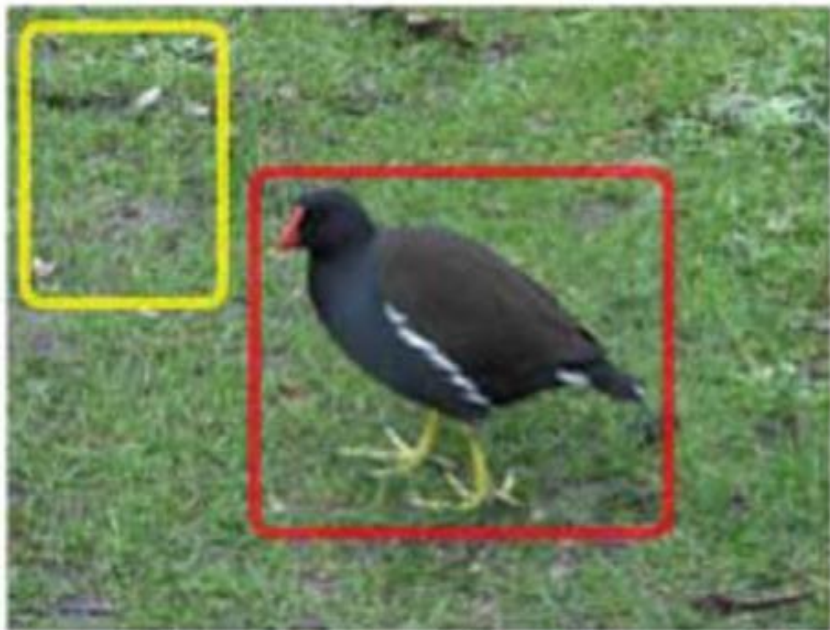


## mAP 계산과의 관계

- 다양한 수준의 IoU 역치(Thresholds)가 고려됨  
(0.5, 0.75, etc.  
0.5 to **0.95** with step size 0.05)

# CS (Confidence Score)

BBox內 물체가 있을 확률



알고리즘마다 다름

- 그냥 Object 있을 확률
- Object 있을 확률  $\times$  IoU
- 특정 Class일 확률  $\times$  IoU

# Precision vs Recall

# 어떻게 평가할 것인가?

- 회귀 모델 평가
  - 회귀 모델이 정확한 값을 예측하기는 사실상 어려움
  - 예측 값과 실제 값에 차이(=오차)가 존재할 것이라 예상함
  - 예측 값이 실제 값에 가까울 수록 좋은 모델이라 할 수 있음
  - 예측한 값과 실제 값의 차이(=오차)로 모델 성능을 평가
- 분류 모델 평가
  - 분류 모델은 0인지 1인지를 예측하는 것
  - 실제 값도 0과 1이고 예측 값도 0과 1임(이진분류의 경우)
  - 하지만 0을 1로 예측하거나 1을 0으로 예측할 수 있음
  - 예측 값이 실제 값과 많이 같을 수록 좋은 모델이라 할 수 있음
  - 정확히 예측한 비율로 모델 성능을 평가

오차를  
줄여라!

정확도를  
높여라!



# 얼마나 맞췄는지 확인

실제	예측
1	1
1	1
0	1
0	0
1	0
0	0
0	0
1	0
1	1
1	1

		예측값	
		0	1
실제값	0	0 → 0 3	0 → 1 1
	1	1 → 0 2	1 → 1 4

얼마나 정확히 맞췄을까?

- 1과 0을 정확히 예측한 비율은?
- 1이라 예측한 것 중에서 정말 1인 비율은?
- 실제 1인 것을 1이라고 예측한 비율은?

# 평가한 값에 이름을 부여

		예측값	
		0	1
실제값	0	0 → 0 3	0 → 1 1
	1	1 → 0 2	1 → 1 4

		Predicted Value	
		Negative 0	Positive 1
GT Actual Value	Negative 0	True Negative TN	False Positive FP
	Positive 1	False Negative FN	True Positive TP

- 1과 0을 정확히 예측한 비율은  $= \frac{7}{10} = \text{정확도(Accuracy)} = \frac{TN + TP}{TN + FP + FN + TP}$
- 1이라 예측한 것 중에서 정말 1인 비율은  $= \frac{4}{5} = \text{정밀도(Precision)} = \frac{TP}{FP + TP}$
- 실제 1인 것을 1이라고 예측한 비율은  $= \frac{4}{6} = \text{재현율(Recall)} = \frac{TP}{FN + TP}$

# 왜 이리 많은 평가 방법이 필요?

## 2024년 비가 올 지 예측!!

(0: 비 안 온다. 1: 비 온다)

2/15	3/20	4/20	5/5	5/18	6/7	7/15
0	0	0	0	0	0	1

8/12	9/23	10/5	11/1	11/24	12/2	12/25
1	0	0	0	0	0	0



제 예측이 몇 % 정도 맞출 것 같  
나요? 정확도가 80~90% 되지  
않을까요? 그럼 돛자리 깔까요?

분명 정확도 만으로는 신통력을 인정할 수 없을 것입니다.



True / False

Positive / Negative

암기법 (안헛갈리는 법!)

- 예측결과와의 진위(True, False) + 예측(Positive, Negative)  
“내가 Positive으로 예상을 했는데 결론적으로 맞았다.”

True Positive

- 뒤의 명사부터 볼 것

합성명사 (형용사 + 명사)

Bad Boy

False Positive (FP)

# 혼동 행렬

- Confusion Matrix(오분류표)

		Predicted Value	
		Negative 0	Positive 1
Actual Value	Negative 0	True Negative <b>TN</b>	False Positive <b>FP</b>
	Positive 1	False Negative <b>FN</b>	True Positive <b>TP</b>

결과

T: 예측이 맞음  
F: 예측이 틀림

T	N
F	P
F	N
T	P

예측

P: 양성(Positive)으로 예측  
N: 음성(Negative)으로 예측

- TN(True Negative, 진음성): 음성으로 잘 예측한 것(음성을 음성이라고 예측한 것)
- FP(False Positive, 위양성): 양성으로 잘 못 예측한 것(음성을 양성이라고 예측한 것)
- FN(False Negative, 위음성): 음성으로 잘 못 예측한 것(양성을 음성이라고 예측한 것)
- TP(True Positive, 진양성): 양성으로 잘 예측한 것(양성을 양성이라고 예측한 것)

# Accuracy

- 정확도
  - 정분류율 이라고 부르기도 함 **TRUE!**
  - 전체 중에서 Positive와 Negative 로 정확히 예측한(TN + TP) 비율
  - Negative를 Negative로 예측한 경우도 옳은 예측임을 고려하는 평가 지표
  - 가장 직관적으로 모델 성능을 확인할 수 있는 평가 지표

		Predicted Value	
		Negative 0	Positive 1
Actual Value	Negative 0	True Negative <b>TN</b>	False Positive <b>FP</b>
	Positive 1	False Negative <b>FN</b>	True Positive <b>TP</b>

$$(Accuracy) = \frac{TN + TP}{TN + FP + FN + TP}$$



# Precision

- 정밀도
  - Positive로 예측한 것(FP + TP) 중에서 실제 Positive(TP)인 비율
  - 예) 비가 내릴 것으로 예측한 날 중에서 실제 비가 내린 날의 비율
  - 예) 암이라 예측한 환자 중에서 실제 암인 환자의 비율
- 정밀도가 낮을 경우 발생하는 상황
  - 비가 오지 않는데 비가 온다고 했으니 불필요한 우산을 챙기는 사고 발생
  - 암이 아닌데 암이라 했으니 불필요한 치료 발생

		Predicted Value	
		Negative 0	Positive 1
Actual Value	Negative 0	True Negative <b>TN</b>	False Positive <b>FP</b>
	Positive 1	False Negative <b>FN</b>	True Positive <b>TP</b>

$$(Precision) = \frac{TP}{FP + TP}$$

모든 검출 결과 중 (FP + TP)

옳게 검출한 비율! (TP)

# Precision

- 정밀도
  - Positive로 예측한 것(FP + TP) 중에서 실제 Positive(TP)인 비율
  - 예) 비가 내릴 것으로 예측한 날 중에서 실제 비가 내린 날의 비율
  - 예) 암이라 예측한 환자 중에서 실제 암인 환자의 비율
- 정밀도가 낮을 경우 발생하는 상황
  - 비가 오지 않는데 비가 온다고 했으니 불필요한 우산을 챙기는 사고 발생
  - 암이 아닌데 암이라 했으니 불필요한 치료 발생

		Predicted Value	
		Negative 0	Positive 1
Actual Value	Negative 0	True Negative <b>TN</b>	False Positive <b>FP</b>
	Positive 1	False Negative <b>FN</b>	True Positive <b>TP</b>

Correct Rejection

Miss

False Alarm (*Precision*) =  $\frac{TP}{FP + TP}$

Hit

# Recall

- 재현율
  - 실제 Positive(FN + TP) 중에서 Positive로 예측한(TP) 비율
  - 민감도(Sensitivity)라고 부르는 경우가 많음
  - 예) 실제 비가 내린 날 중에서 비가 내릴 것으로 예측한 날의 비율
  - 예) 실제 암인 환자 중에서 암이라고 예측한 환자의 비율
- 재현율이 낮을 경우 발생하는 문제
  - 비가 내리는 날 내리지 않을 것이라 했으니 우산을 챙기지 않아 비를 맞음
  - 암인 사람에게 암이 아니라 했으니 심각한 결과 초래

Correct Rejection

		Predicted Value		
		Negative 0	Positive 1	
Actual Value	Negative 0	True Negative <b>TN</b>	False Positive <b>FP</b>	False Alarm
	Positive 1	False Negative <b>FN</b>	True Positive <b>TP</b>	Hit

Miss

$(Recall) = \frac{TP}{FN + TP}$

검출해내야 하는 물체 중에서

(FN + TP)

제대로 검출한 비율

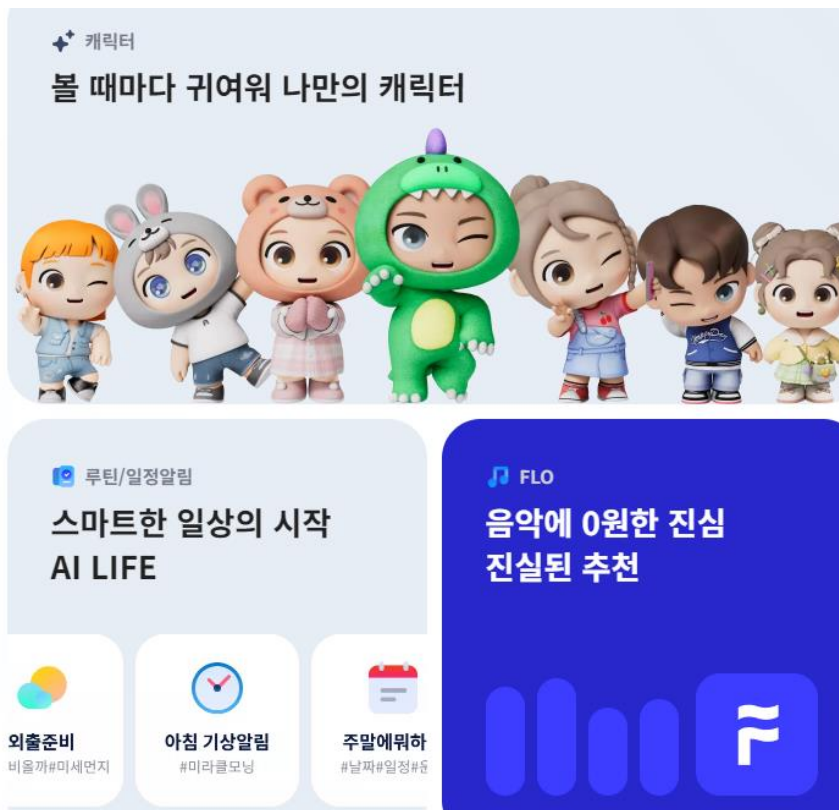
(TP)

모델 성능이 좋기 위해서는

물론 둘 다 높아야 하나,

일반적으로 Precision과

Recall은 반비례 관계



Voice Trigger 기술

# F1-Score

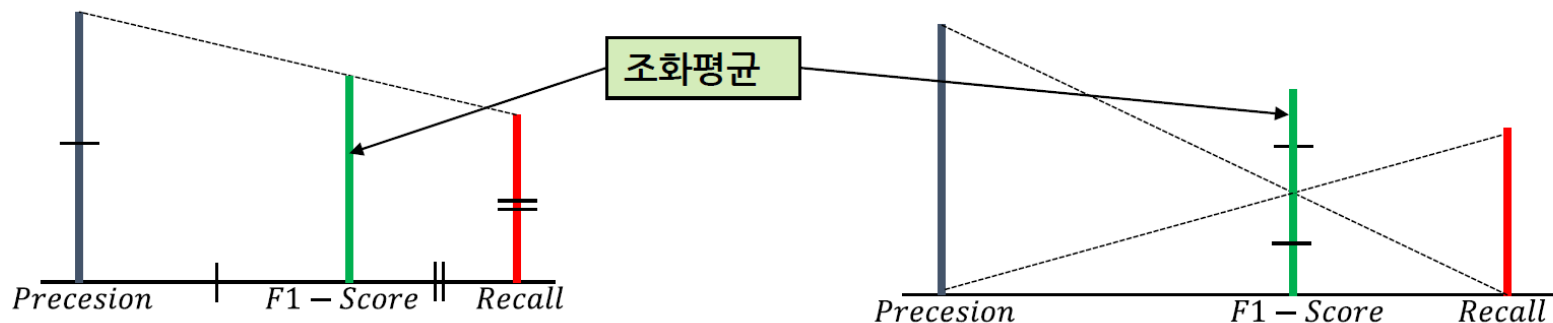
- 정밀도와 재현율의 조화평균 **Harmonic Mean**
- 분자가 같지만 분모가 다를 경우 조화평균이 정확
- 정밀도와 재현율이 적절하게 요구 될 때 사용

$$\frac{a+b}{2} \geq \sqrt{ab} \geq \frac{2ab}{a+b}$$

산술평균      기하평균      조화평균

$$(F1 - Score) = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

- 조화평균



- 작은 값 쪽으로 치우친, 작은 값과 큰 값 사이의 값을 가진 평균
- 산술평균 보다 큰 값이 끼치는 영향이 줄어듦



# 분류 평가 지표 정리

		Predicted Values	
		Negative 0	Positive 1
Actual Values	Negative 0		
	Positive 1		

$$(Precision) = \frac{\boxed{\phantom{000}}}{\boxed{\phantom{000}} + \boxed{\phantom{000}}}$$

$$(Recall) = \frac{\boxed{\phantom{000}}}{\boxed{\phantom{000}} + \boxed{\phantom{000}}}$$

$$(Accuracy) = \frac{\boxed{\phantom{000}} + \boxed{\phantom{000}}}{\boxed{\phantom{000}} + \boxed{\phantom{000}} + \boxed{\phantom{000}} + \boxed{\phantom{000}}}$$

$$(F1 - Score) = \frac{2 \times \boxed{\phantom{000000}}}{\boxed{\phantom{000000}}}$$

# 이중 분류 성능 평가 예

confusion\_matrix

[[69 23]  
[14 14]]

		Predicted Values	
		0	1
Actual Values	0	69	23
	1	14	14

classification\_report

	precision	recall	f1-score	support
0	0.83	0.75	0.79	92
1	0.38	0.50	0.43	28
accuracy			0.69	120
macro avg	0.60	0.62	0.61	120
weighted avg	0.73	0.69	0.71	120

정밀도 재현율

	Precision	Recall	F1-Score
0	69/83	69/92	
1	14/37	14/28	
Accuracy	83/120		

	Precision	Recall	F1-Score
0	0.83	0.75	0.79
1	0.38	0.50	0.43
Accuracy	0.69		

# 다중 분류 성능 평가 예

confusion\_matrix

```
[[ 7  0  0]
 [ 0 11  1]
 [ 0  2  9]]
```

		Predicted Values		
		A	B	C
Actual Values	A	7	0	0
	B	0	11	1
	C	0	2	9

	Precision	Recall	F1-Score
A	7/7	7/7	
B	11/13	11/12	
C	9/10	9/11	
Accuracy	27/30		

classification\_report

	precision	recall	f1-score	support
A	1.00	1.00	1.00	7
B	0.85	0.92	0.88	12
C	0.90	0.82	0.86	11
accuracy			0.90	30
macro avg	0.92	0.91	0.91	30
weighted avg	0.90	0.90	0.90	30

	Precision	Recall	F1-Score
A	1.00	1.00	1.00
B	0.85	0.92	0.88
C	0.90	0.82	0.86
Accuracy	0.90		

모델 성능이 좋기 위해서는

물론 둘 다 높아야 하나,

일반적으로 Precision과

Recall은 반비례 관계

Detections	confidences	TP or FP
A	57%	TP
B	78%	TP
C	43%	FP
D	85%	TP
E	91%	TP
F	13%	FP
G	45%	TP
H	68%	FP
I	95%	TP
J	81%	TP

## Case

- Image has 15 objects
- Model detected 10
- TP count : 7
- FP count : 3

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	<b>TP = 7</b>	<b>FN = 8</b>
	NEGATIVE (0)	<b>FP = 3</b>	<b>TN</b>

Confidence Threshold 0

Precision  
7/10

Recall  
7/15

역치를 0으로 잡다니

너무 하는 것 아닙니까?

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	<b>TP = 1</b>	<b>FN = 14</b>
	NEGATIVE (0)	<b>FP = 0</b>	<b>TN</b>

Confidence Threshold  
95 (그래서 이걸 서비스가  
결정함!!!!)

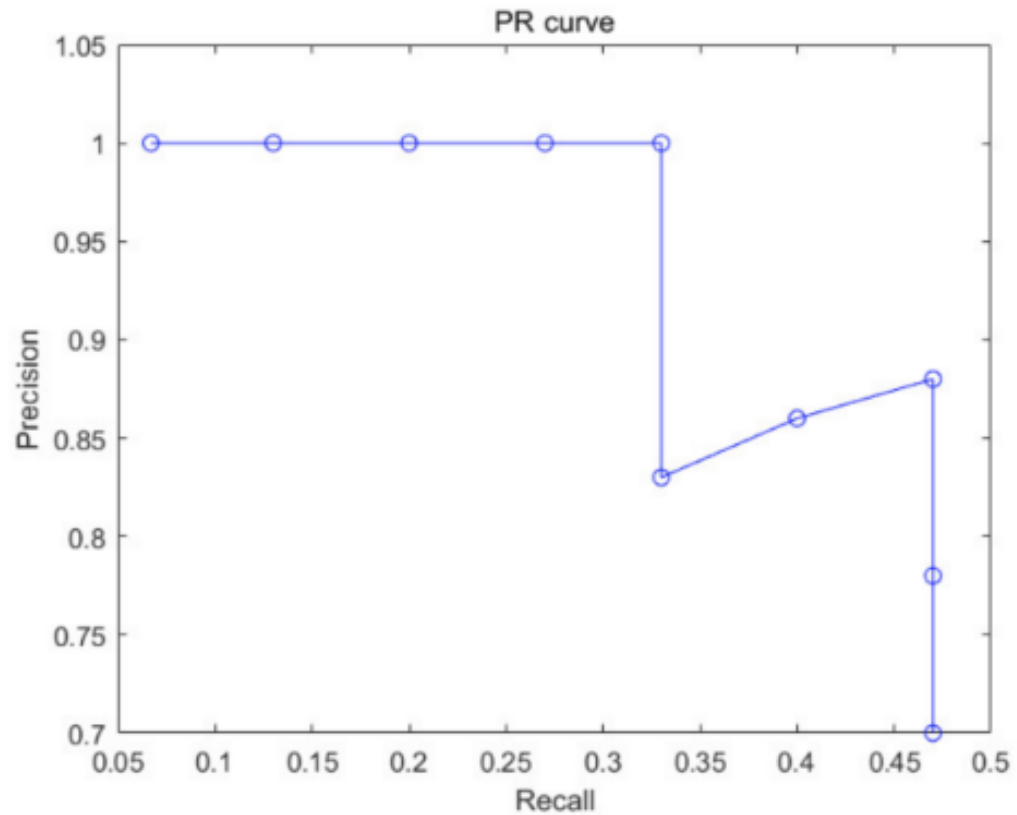
Precision  
1/1 (!!!!)

Recall  
1/15



## 내림차순 정리

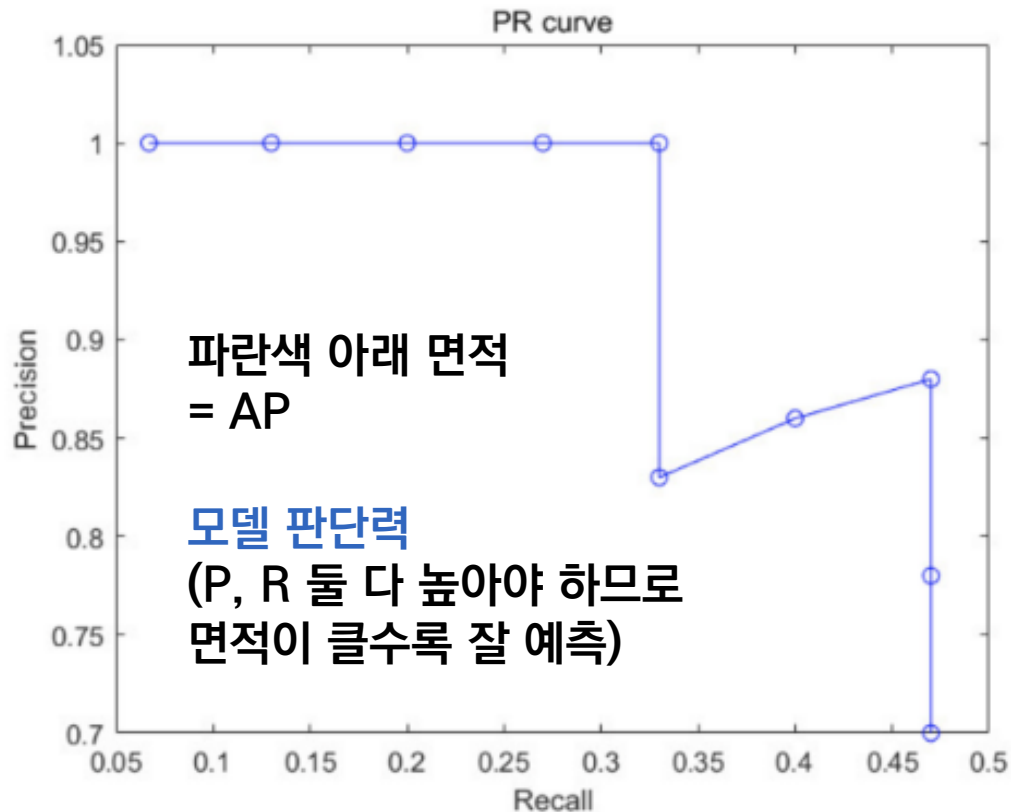
Detections	confidences	TP or FP	누적 TP	누적 FP	Precision	Recall
I	95%	TP	1	0	$1/1=1$	$1/15=0.067$
E	91%	TP	2	0	$2/2=1$	$2/15=0.13$
D	85%	TP	3	0	$3/3=1$	$3/15=0.2$
J	81%	TP	4	0	$4/4=1$	$4/15=0.27$
B	78%	TP	5	0	$5/5=1$	$5/15=0.33$
H	68%	FP	5	1	$5/6=0.83$	$5/15=0.33$
A	57%	TP	6	1	$6/7=0.86$	$6/15=0.4$
G	45%	TP	7	1	$7/8=0.88$	$7/15=0.47$
C	43%	FP	7	2	$7/9=0.78$	$7/15=0.47$
F	13%	FP	7	3	$7/10=0.7$	$7/15=0.47$



Precision – Recall Curve

# AP

The area under the  
**precision–recall curve**  
for each class



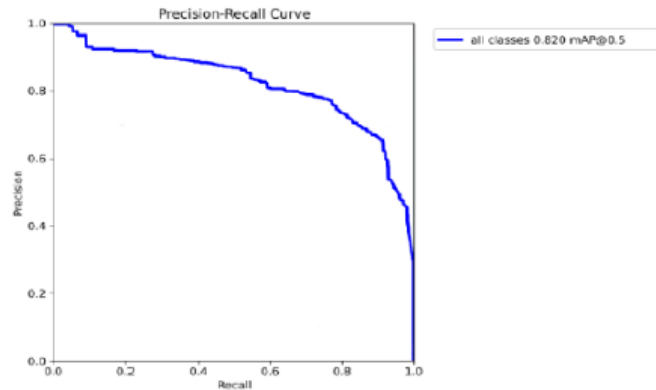
Precision – Recall Curve

# mAP

(Mean Average  
Precision)

Class마다 AP 계산  
그들의 평균

## mAP50



$$\text{mAP50} = 0.820$$

# 참고영상

mAP 설명 및 파이썬 코드

<https://www.youtube.com/watch?v=FppOzcDvaDI>