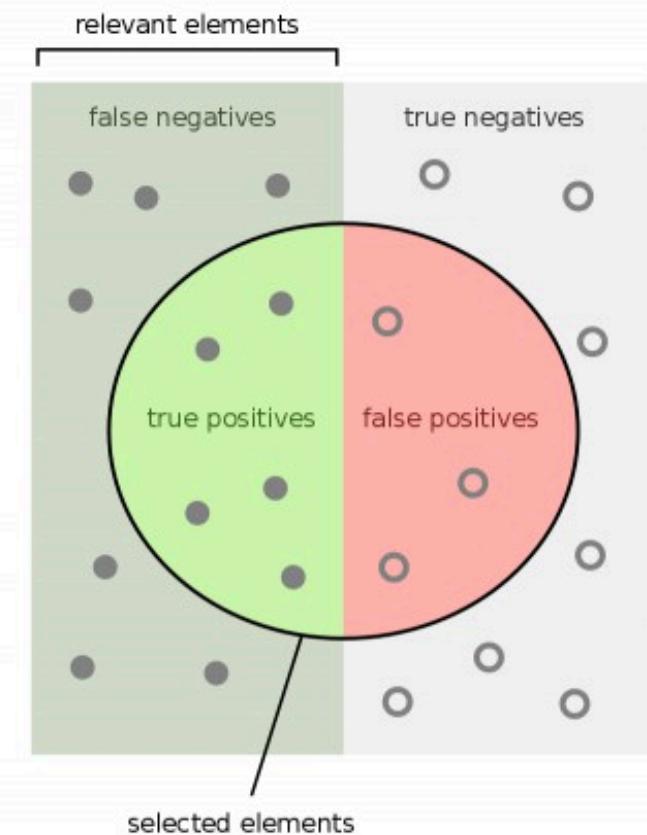


Confusion Matrix

2

		Actual = Yes	Actual = No
Predicted = Yes	TP	FP	
	FN	TN	
		Predicted	
Actual	True	True Positives	False Negatives
	False	False Positives	True Negatives



How many relevant items are selected?
e.g. How many sick people are correctly identified as having the condition.

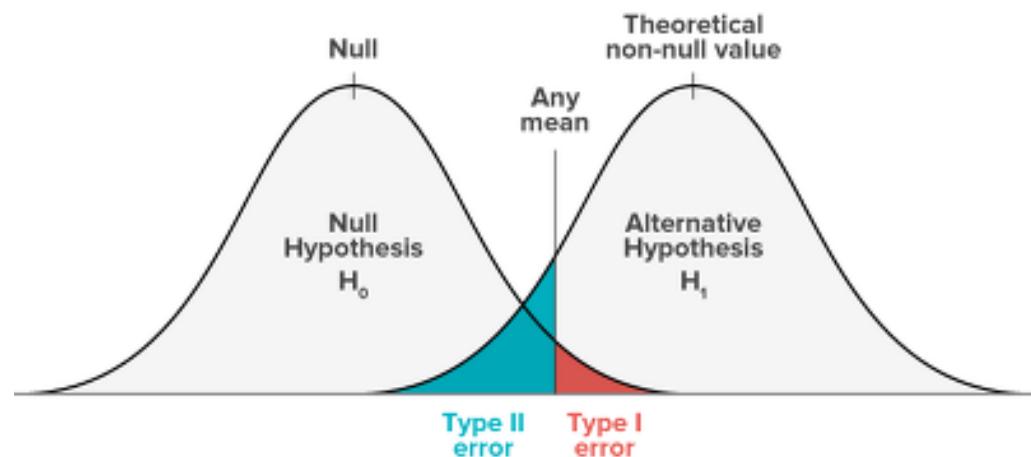
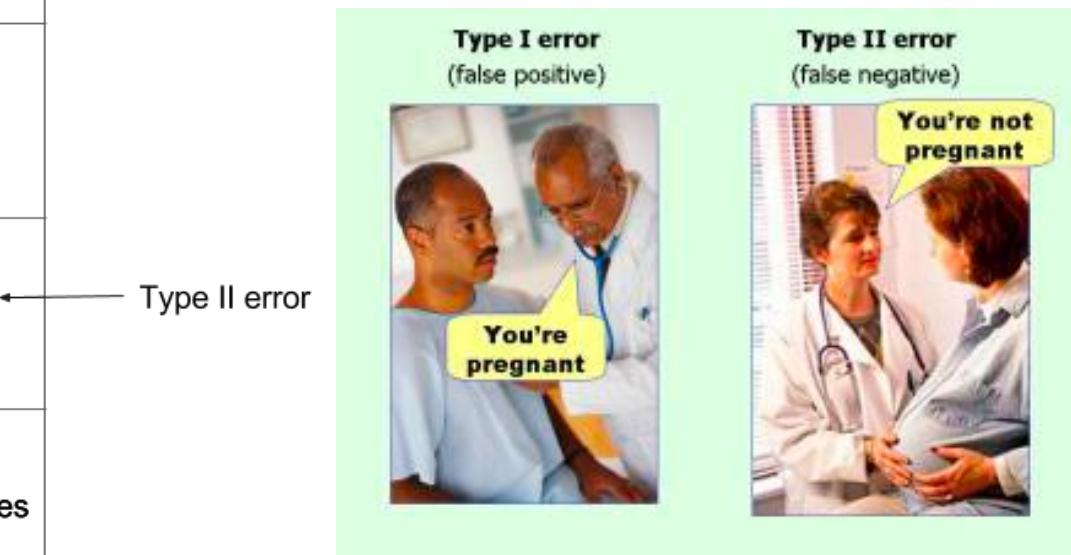
$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

How many negative selected elements are truly negative?
e.g. How many healthy people are identified as not having the condition.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

		predicted	
		p'	n'
actual	p	true positives	false negatives
	n	false positives	true negatives

↑
Type I error



TP(true positive): true를 true로 잘 예측한 것

TN(true negative): false를 false로 잘 예측한 것

FP(false positive): false를 true로 잘못 예측한 것

FN(false negative): true를 false로 잘못 예측한 것

		predictions (output)			
		A	B	C	D
actual class (input)	A	9	1	0	0
	B	1	15	3	1
	C	5	0	24	1
	D	0	4	1	15

TP(True Positive)

A	B	C	D	
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

TN_A(True Negative for A)

A	B	C	D	
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

TN_D(True Negative for D)

	A	B	C	D
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

FP_A(False Positive for A)

A	B	C	D	
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

FP_B(False Positive for B)

A	B	C	D	
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

FN_A(False Negative for A)

A	B	C	D	
A	9	1	0	0
B	1	15	3	1
C	5	0	24	1
D	0	4	1	15

FN_D(False Negative for D)

■ Recall or Sensitivity or TPR (True Positive Rate)

- Number of items correctly identified as positive out of total true positives- $TP/(TP+FN)$

■ Specificity or TNR (True Negative Rate)

- Number of items correctly identified as negative out of total negatives- $TN/(TN+FP)$

■ Precision

- Number of items correctly identified as positive out of total items identified as positive- $TP/(TP+FP)$

■ False Positive Rate or Type I Error or Fall-out

- Number of items wrongly identified as positive out of total true negatives- $FP/(FP+ TN)$

■ False Negative Rate or Type II Error

- Number of items wrongly identified as negative out of total true positives- $FN/(FN+TP)$

■ F1 Score

- It is a harmonic mean of precision and recall given by- $F1 = 2*Precision*Recall/(Precision + Recall)$

■ Accuracy

- Percentage of total items classified correctly- $(TP+TN)/(N+P)$

Balanced Data

Model 1

predictions
(output) →

		A	B	C	D
actual class (input)	A	10	0	0	0
	B	0	5	3	2
	C	0	1	8	1
	D	0	1	0	9

Accuracy = sum(TP) / total data set #
 $(10+5+8+9) / 40 = 0.8$

Model 2

predictions
(output) →

		A	B	C	D
actual class (input)	A	8	2	0	0
	B	1	7	0	2
	C	0	0	9	1
	D	2	3	0	5

Accuracy = sum(TP) / total data set #
 $(8+7+9+5) / 40 = 0.725$

Imbalanced Data

Model 1

		predictions (output)			
		A	B	C	D
actual class (input)	A	100	80	10	10
	B	0	9	0	1
	C	0	1	8	1
	D	0	1	0	9

Recall

$$\text{Accuracy} = \text{sum}(TP) / \text{Total Dataset \#}$$

$$\text{Precision} = TP / (TP+FP)$$

$$(\text{macro})\text{average precision} = \text{sum}(\text{precision}) / (\text{the number of classes})$$

$$\text{Recall} = TP / (TP+FN)$$

$$(\text{macro})\text{average recall} = \text{sum}(\text{recall}) / (\text{the number of classes})$$

$$\text{F1 Score} = 2 * ((\text{precision} * \text{recall}) / (\text{precision}+\text{recall}))$$

$$\text{Accuracy} = (100+9+8+9) / 230 = 0.547$$

- Precision A = $100 / (100+0) = 1$
- Precision B = $9 / (9+81) = 9/91$
- Precision C = $8 / (8+10) = 8/18$
- Precision D = $9 / (9+12) = 9/21$

- recall A = $100 / (100+100) = 100/200$
- recall B = $9 / (9+1) = 9/10$
- recall C = $8 / (8+2) = 8/10$
- recall D = $9 / (9+1) = 9/10$

$$\text{average precision} = 0.492$$

$$\text{average recall} = 0.775$$

$$\text{F1 Score} = 2 * ((0.492*0.775) / (0.492+0.775)) = 0.601$$

Model 2

		predictions (output)			
		A	B	C	D
actual class (input)	A	198	2	0	0
	B	7	1	0	2
	C	0	8	1	1
	D	2	3	4	1

$$\text{Accuracy} = (198+1+1+1) / 230 = 0.87$$

- Precision A = $198 / (198+9) = 198/207$
- Precision B = $1 / (1+13) = 1/14$
- Precision C = $1 / (1+4) = 1/5$
- Precision D = $1 / (1+3) = 1/4$

- recall A = $198 / (198+2) = 198/200$
- recall B = $1 / (1+9) = 1/10$
- recall C = $1 / (1+9) = 1/10$
- recall D = $1 / (1+9) = 1/10$

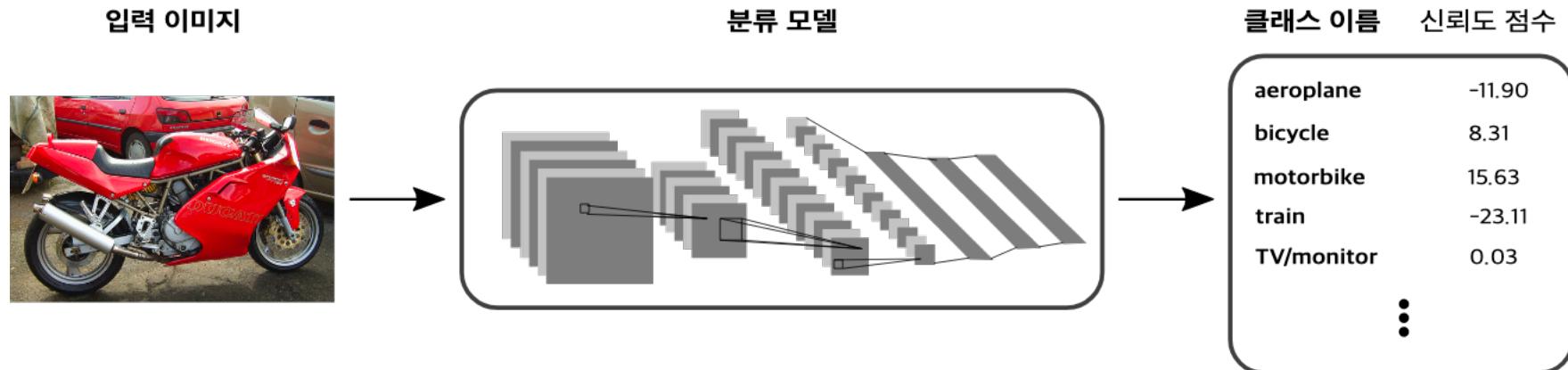
$$\text{average precision} = 0.369$$

$$\text{average precision} = 0.323$$

$$\text{F1 Score} = 2 * ((0.369*0.323) / (0.369+0.323)) = 0.344$$

Accuracy에 의해 모델을 선택하면 Model2가 더 좋으나 A에만 좋은 모델 B,C,D에 대한 예측은 떨어지는 문제 발생,
Imbalanced data인 경우에는 F1 Score를 평가 지표로 이용

confidence score



- 주어진 이미지 안에 특정 클래스의 사물이 존재할 '가능성'
 - 해당 결과물에 대한 사후적인 해석의 여지
 - threshold**을 미리 설정해 놓고, 주어진 이미지의 각 클래스 별 신뢰도 점수가 문턱값보다 큰 경우에 한하여 '주어진 이미지 안에 해당 클래스가 포함되어 있을 것이다'고 결론

$$\text{Accuracy} = \frac{\text{올바르게 분류한 이미지 수}}{\text{전체 이미지 수}}$$

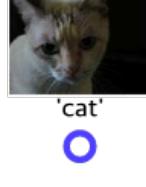
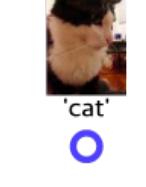
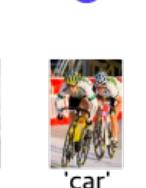
예측 클래스	실제 클래스	정답 여부					
	'aeroplane'	'aeroplane'	○		'bus'	'bus'	○
	'car'	'bicycle'	✗		'bird'	'bird'	○
	'horse'	'horse'	○		'cow'	'cow'	○
	'train'	'TV/monitor'	✗		'dog'	'dog'	○
	'potted plant'	'potted plant'	○		'bicycle'	'motorbike'	✗

$$\text{Precision}_c = \frac{\text{올바르게 분류한 클래스 } c \text{ 이미지 수}}{\text{클래스 } c\text{일 것으로 예측한 이미지 수}}$$

$$\begin{aligned} \text{Recall}_c &= \frac{\text{올바르게 분류한 클래스 } c \text{ 이미지 수}}{\text{전체 클래스 } c \text{ 이미지 수}} \\ \text{Sensitivity, hit rate} \end{aligned}$$

예측 클래스	실제 클래스, 정답 여부					정밀도
'car'						0.4
'cat'						0.6
'bicycle'						0.4
⋮						

평균 정밀도
 $(0.4+0.6+0.4)/3=0.47(47\%)$

실제 클래스	예측 클래스, 정답 여부					재현율
'car'						
						0.6
						1.0
'cat'						0.8
						0.8
⋮						

평균 재현율은 $(0.6+1.0+0.8)/3=0.8(80\%)$

클래스 이름 신뢰도 점수



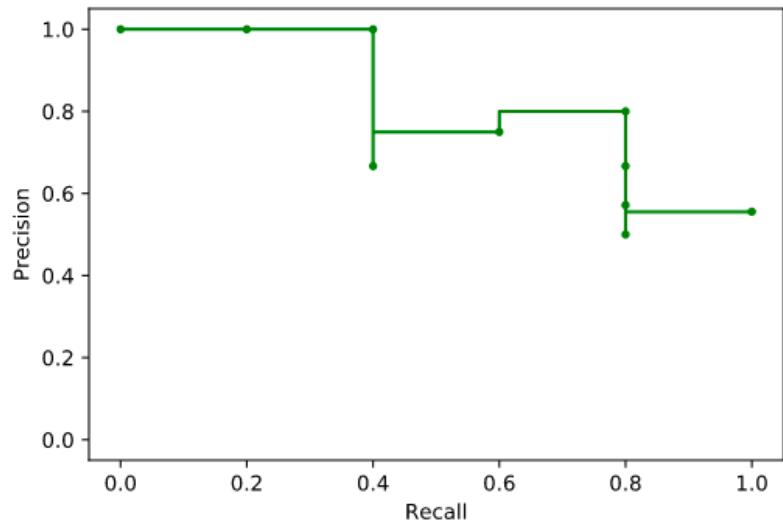
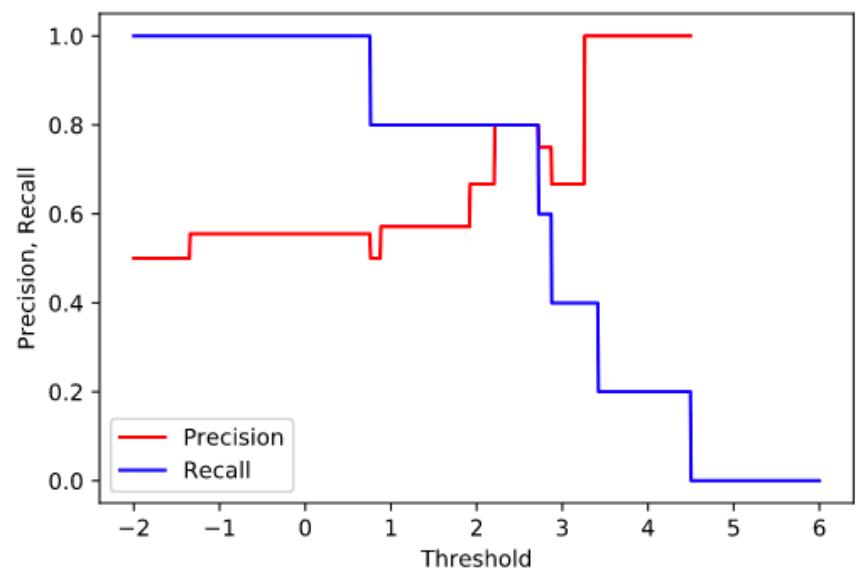
aeroplane	-10.22	dining table	-5.98
bicycle	-1.92	dog	29.99
bird	-21.56	horse	8.89
boat	-18.77	motorbike	-20.01
bottle	0.56	person	7.67
bus	-1.09	potted plant	-1.23
car	1.11	sheep	15.23
cat	37.26	sofa	2.89
chair	5.60	train	-23.11
cow	7.27	TV/monitor	0.03

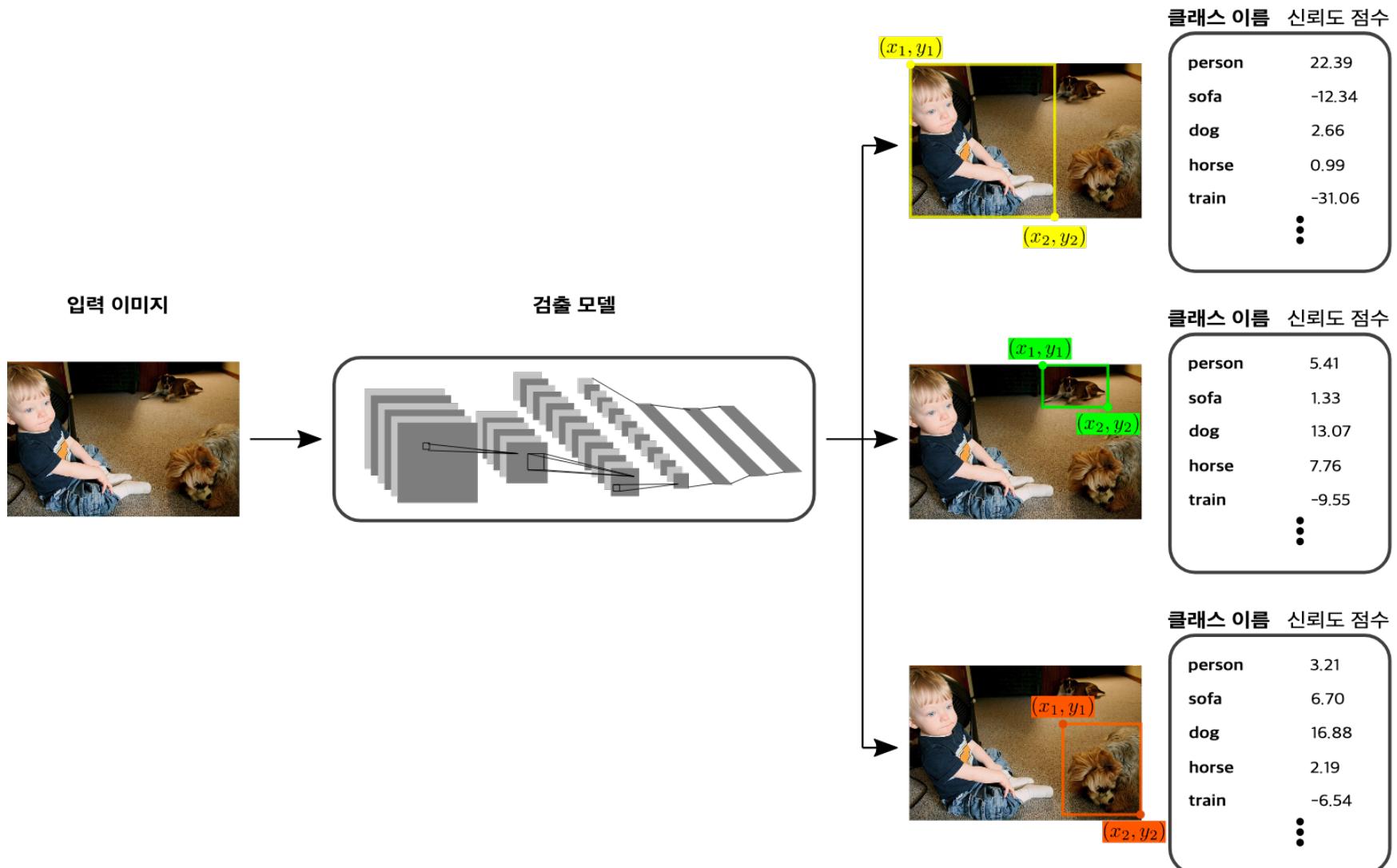
클래스 이름(문턱값) 신뢰도 점수



aeroplane(13.72)	-1.99	dining table(7.76)	-10.11
bicycle(10.66)	2.21	dog(13.21)	-6.54
bird(17.58)	-3.14	horse(14.44)	13.21
boat(16.70)	-33.09	motorbike(8.97)	4.44
bottle(8.11)	-12.13	person(8.01)	11.33
bus(10.11)	5.72	potted plant(21.31)	-10.10
car(15.61)	9.22	sheep(17.17)	16.66
cat(15.09)	0.98	sofa(12.34)	-1.99
chair(11.22)	-8.90	train(16.06)	2.21
cow(9.98)	27.38	TV/monitor(11.99)	-29.34

실제 클래스		'car'	'bus'	'bus'	'car'	'car'	'car'	'car'	'bicycle'	'bicycle'	'cat'	정밀도	재현율
이미지													
'car' 신뢰도 점수		2.72	3.26	0.88	0.76	3.42	2.87	4.50	1.92	2.21	-1.35		
문턱값에 따른 예측 클래스 및 정답 여부		not 'car' X	not 'car' O	not 'car' O	not 'car' X	not 'car' X	not 'car' X	not 'car' X	not 'car' O	not 'car' O	not 'car' O	n/a	0.0
		not 'car' X	not 'car' O	not 'car' O	not 'car' X	not 'car' X	not 'car' X	'car' O	not 'car' O	not 'car' O	not 'car' O	1.0	0.2
		not 'car' X	'car' X	not 'car' O	not 'car' X	'car' O	not 'car' X	'car' O	not 'car' O	not 'car' O	not 'car' O	0.67	0.4
		'car' O	'car' X	not 'car' O	not 'car' X	'car' O	'car' O	'car' O	'car' O	'car' X	not 'car' O	0.67	0.8
		'car' O	'car' X	not 'car' O	not 'car' X	'car' O	'car' O	'car' O	'car' X	'car' X	not 'car' O	0.57	0.8
		'car' O	'car' X	'car' X	'car' O	'car' O	'car' O	'car' O	'car' X	'car' X	not 'car' O	0.56	1.0



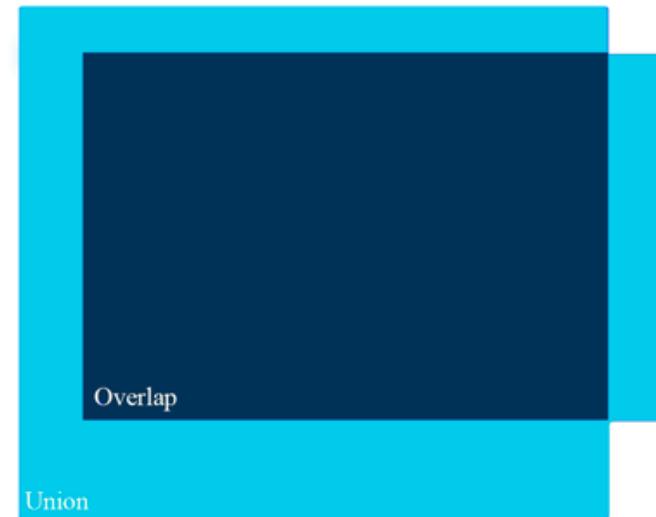


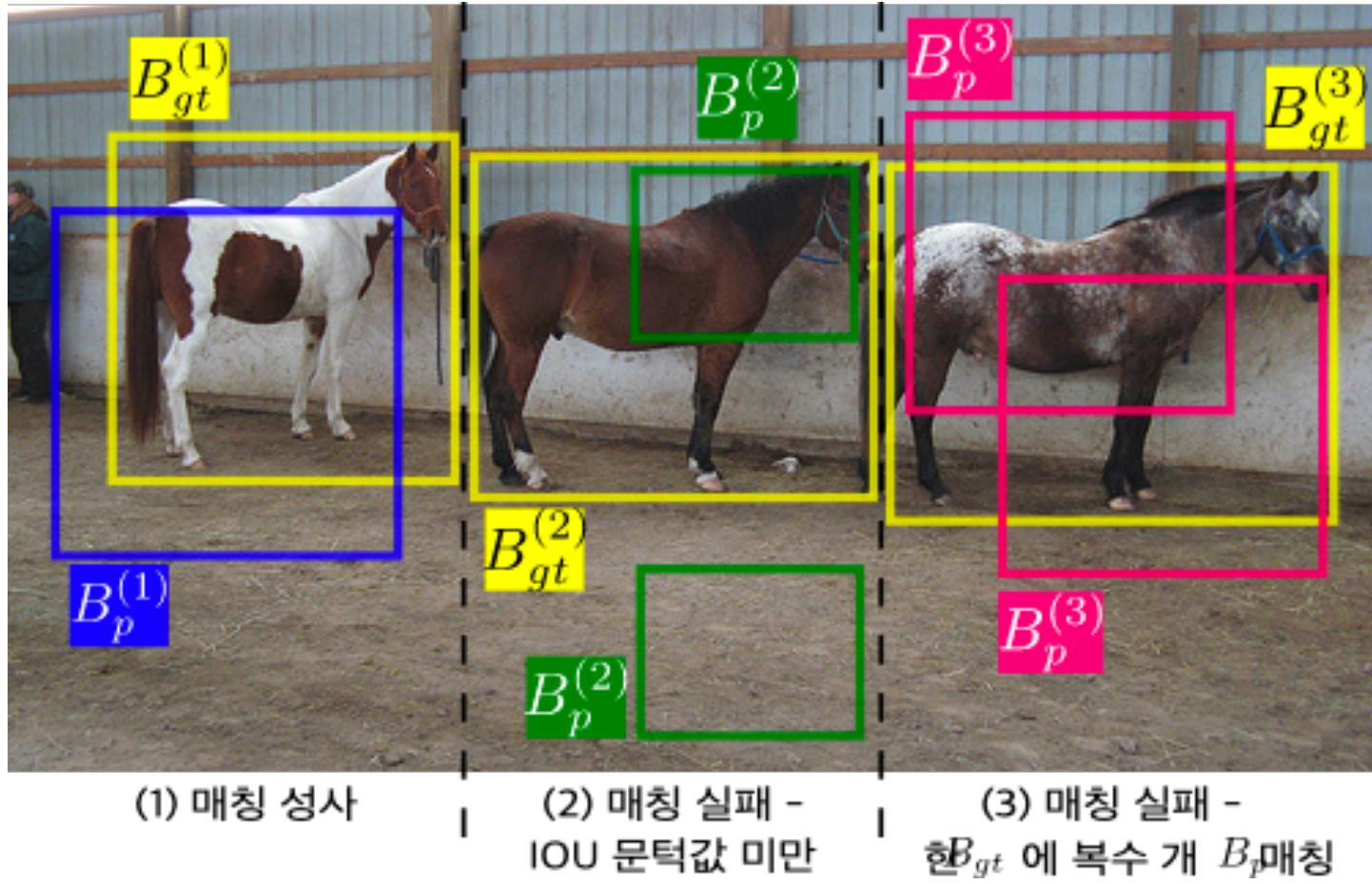
Intersection over unit



- █ Ground truth
- █ Prediction

$$IoU = \frac{\text{area of overlap}}{\text{area of union}}$$





 = relevant documents for query 1

Ranking #1



Recall 0.2 0.2 0.4 0.4 0.4 0.6 0.6 0.6 0.8 1.0

Precision 1.0 0.5 0.67 0.5 0.4 0.5 0.43 0.38 0.44 0.5



= relevant documents for query 2

Ranking #2



Recall 0.0 0.33 0.33 0.33 0.67 0.67 1.0 1.0 1.0 1.0

Precision 0.0 0.5 0.33 0.25 0.4 0.33 0.43 0.38 0.33 0.3

$$\text{average precision query 1} = (1.0 + 0.67 + 0.5 + 0.44 + 0.5) / 5 = 0.62$$

$$\text{average precision query 2} = (0.5 + 0.4 + 0.43) / 3 = 0.44$$

$$\text{mean average precision} = (0.62 + 0.44) / 2 = 0.53$$