Detection, Recognition Classification, Identification & Their Index

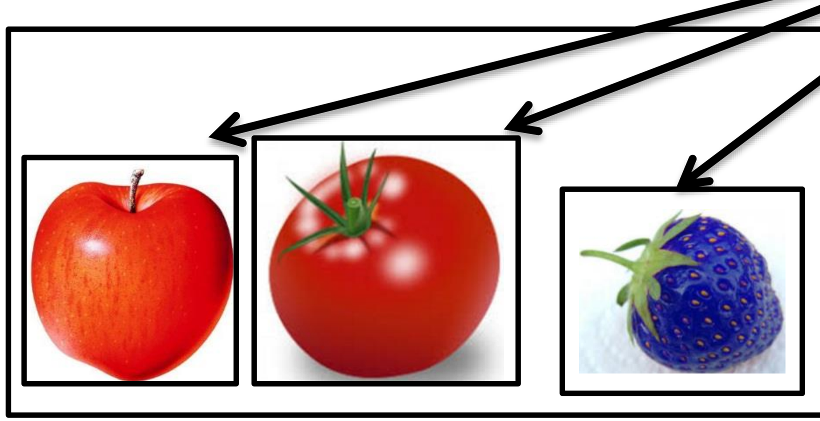
Alex Lin

Safety Sensing & Control Department
Intelligent Mobility Technology Division
Mechanical and Systems Research Laboratories
Industrial Technology Research Institute

Detection, recognition, classification & identification

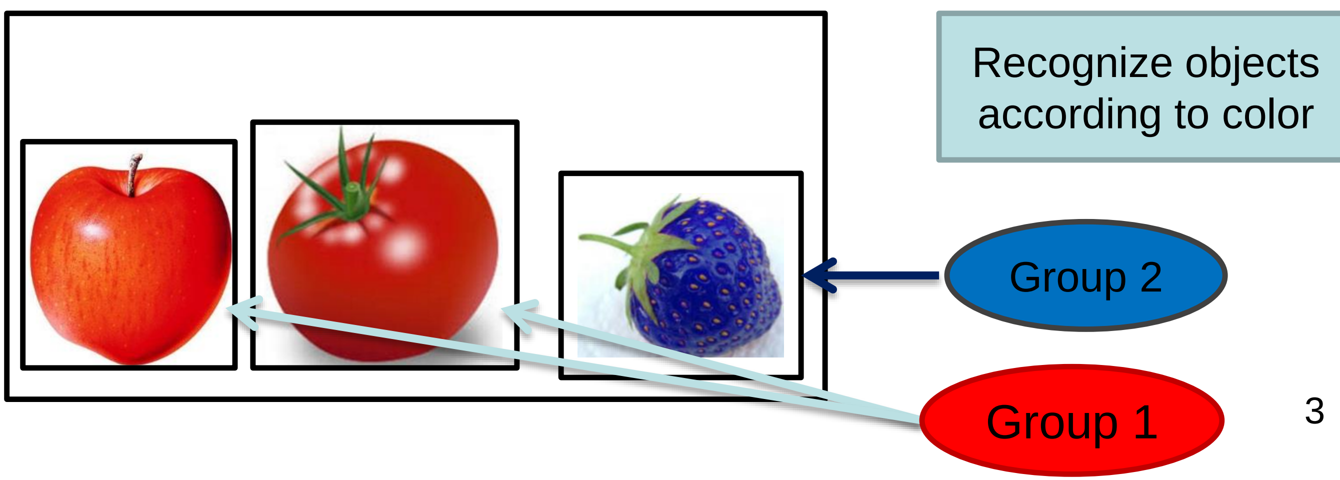
- **Detection**: Find out a particular or set of features or objects in the images.
- Recognition: form groups of similar objects based on the similar measure.
- Classification: classify the objects to certain classes.
- Identification: recognize an unknown object and name it

Detect objects



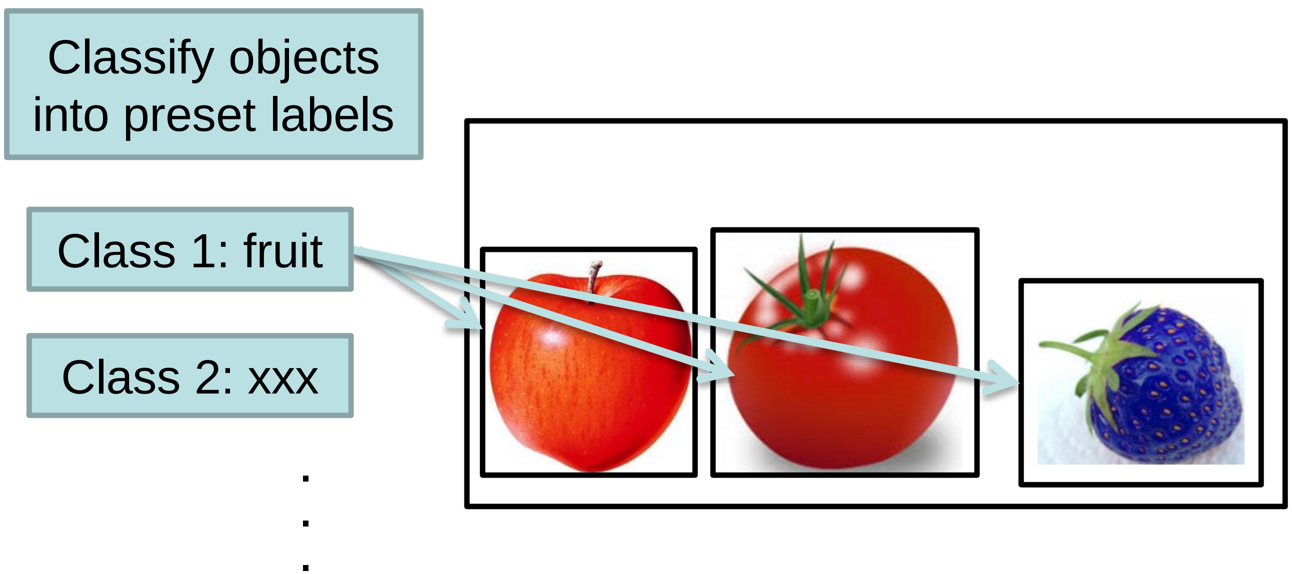
Detection, recognition, classification & identification

- Detection: Find out a particular or set of features or objects in the images.
- **Recognition**: form groups of similar objects based on the similar measure.
- Classification: classify the objects to certain classes.
- Identification: recognize an unknown object and name it



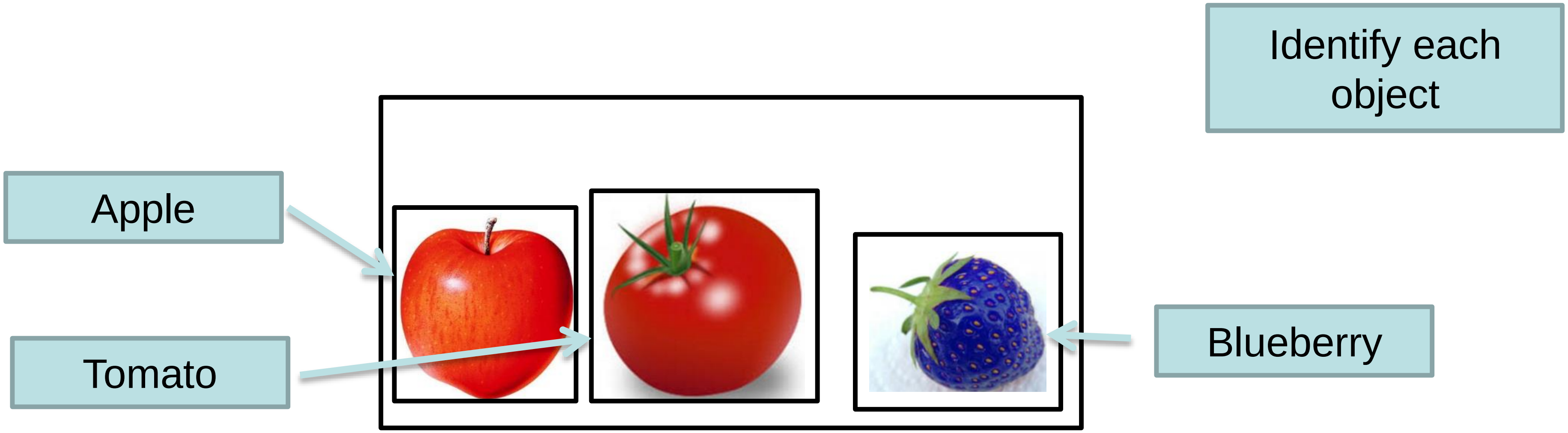
Detection, recognition, classification & identification

- Detection: Find out a particular or set of features or objects in the images.
- Recognition: form groups of similar objects based on the similar measure.
- **Classification**: classify the objects to certain classes.
- Identification: recognize an unknown object and name it



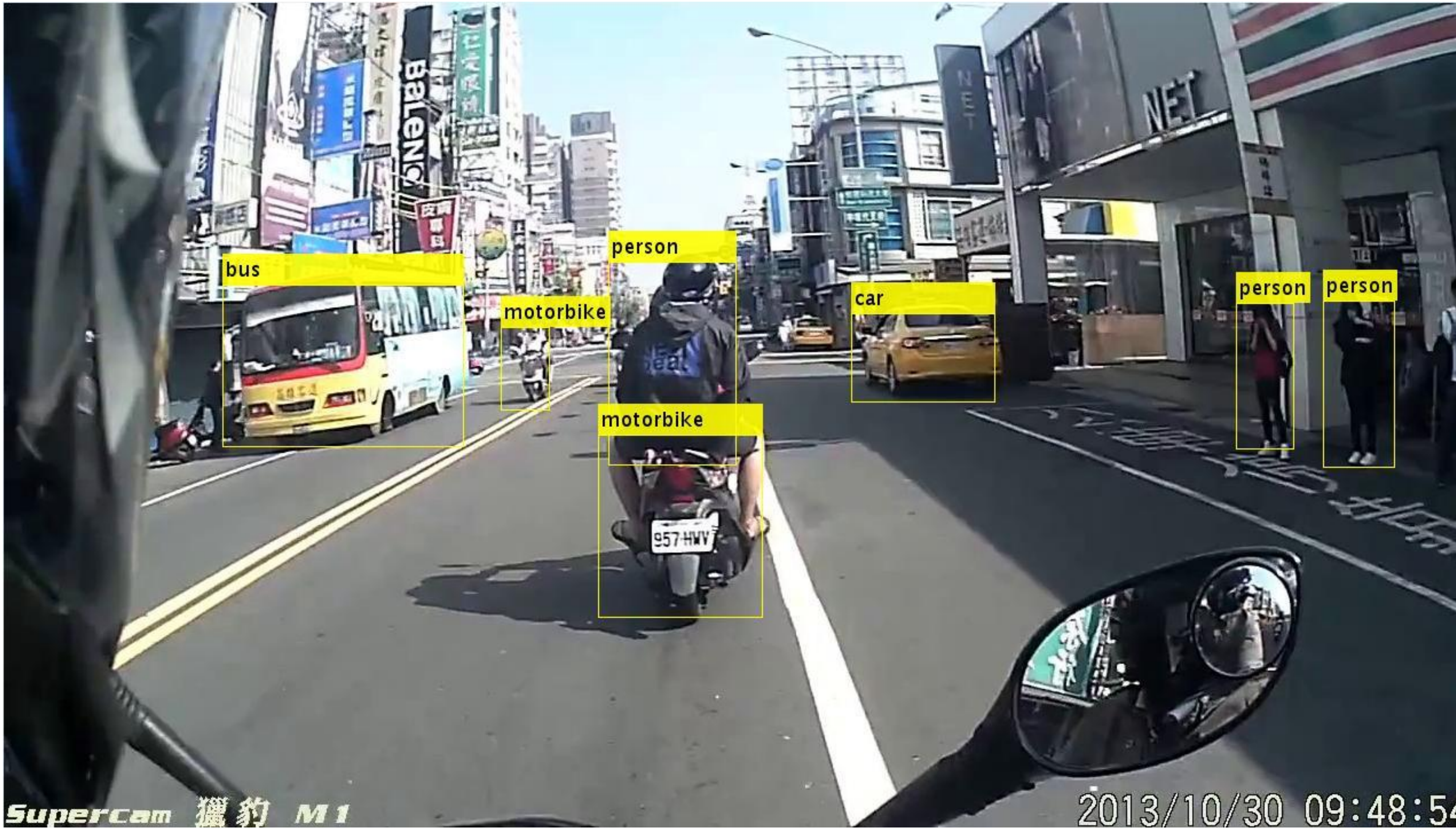
Detection, recognition, classification & identification

- Detection: Find out a particular or set of features or objects in the images.
- Recognition: form groups of similar objects based on the similar measure.
- Classification: classify the objects to certain classes.
- **Identification**: recognize an unknown object and name it



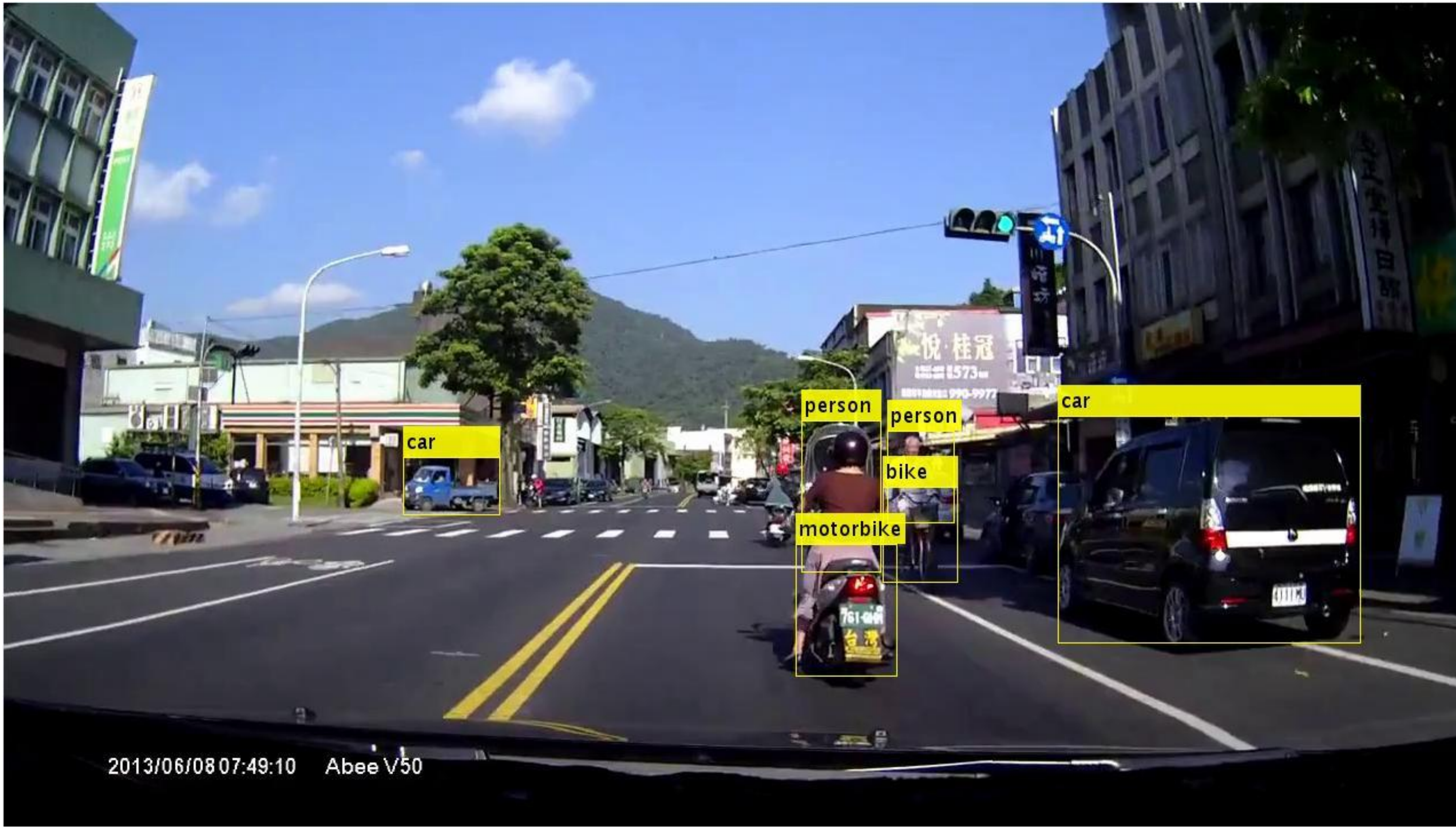
Dataset before being Labeled 1

Dataset after being Labeled 1



Dataset before being Labeled 2

Dataset after being Labeled 2



Output of label


```
<?xml version="1.0"?>
<annotation>
    <folder>all_dataset</folder>
    <filename>009967.jpg</filename>
  - <object>
        <name>car</name>
      - <bndbox>
            <xmax>1274</xmax>
            <xmin>1055</xmin>
            <ymax>590</ymax>
            <ymin>316</ymin>
        </bndbox>
        <difficult>0</difficult>
        <occluded>0</occluded>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
    </object>
  - <object>
        <name>bus</name>
      - <bndbox>
            <xmax>330</xmax>
            <xmin>289</xmin>
            <ymax>347</ymax>
            <ymin>301</ymin>
        </bndbox>
        <difficult>0</difficult>
        <occluded>0</occluded>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
    </object>
  - <object>
        <name>car</name>
      - <bndbox>
            <xmax>360</xmax>
            <xmin>332</xmin>
            <ymax>347</ymax>
            <ymin>323</ymin>
        </bndbox>
        <difficult>0</difficult>
        <occluded>0</occluded>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
    </object>
  - <object>
        <name>car</name>
      - <bndbox>
            <xmax>714</xmax>
            <xmin>630</xmin>
            <ymax>368</ymax>
            <ymin>327</ymin>
        </bndbox>
        <difficult>0</difficult>
        <occluded>0</occluded>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
    </object>
    <segmented>0</segmented>
  - <size>
        <depth>3</depth>
        <height>720</height>
        <width>1280</width>
    </size>
  - <source>
        <annotation>Id_000320</annotation>
        <database>vatic</database>
        <image>vatic</image>
    </source>
  - <owner>
        <flickrid>vatic</flickrid>
        <name>vatic</name>
    </owner>
</annotation>
```

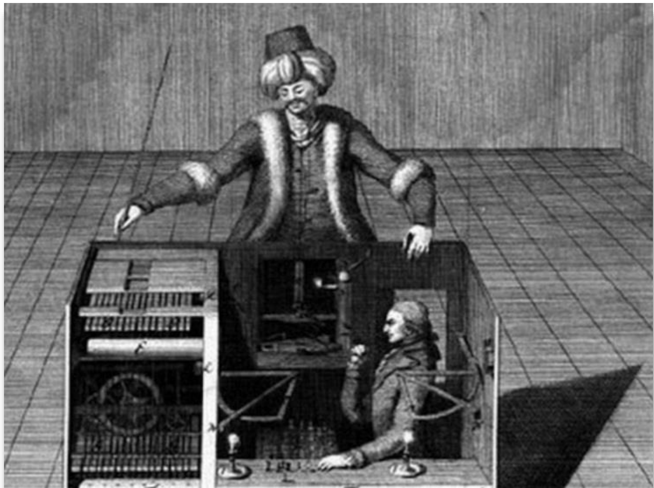
How to efficiently label?

Ref:http://carlvondrick.com/vatic/

Ground-Truth Labeling Providers

- Amazon Mechanical Turk:
  - A US company which provides crowdsourcing service and only accepts US-based requesters
- CloudFactory:
  - A US company who hired people in Kathmandu, Nepal.
- CrowdFlower:
  - A US company who provides crowdsourcing service
- Datatang:
  - A China crowdsourcing platform

数据堂
Shujutang



Quantitative Metrics

		Predicted condition	
	Total population	Predicted Condition positive	Predicted Condition negative
True condition	condition positive	**True positive**	**False Negative** (Type II error)
	condition negative	**False Positive** (Type I error)	**True negative**

Accuracy=(TP + TN)/(TP+TN+FP+FN)

Recall(True Positive Rate, Detection rate)=TP/(TP+FN)

Precision=TP/(TP+FP)

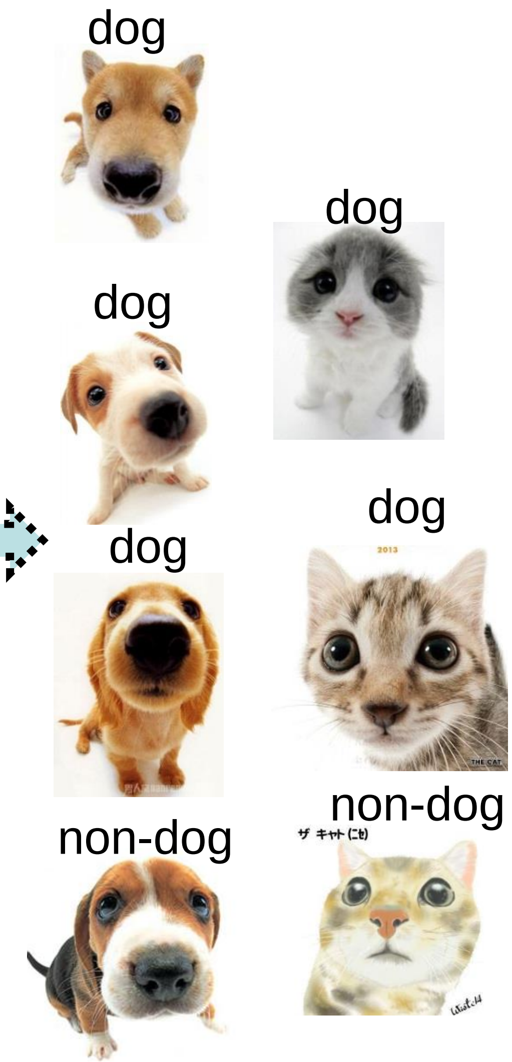
False Positive Rate=FP/(FP+TN)

Ex: 100% detection rate with 50% False Positive rate

A pure binary classification example

Input: images
Output: labels(dog, non-dog)

Classifier



Accuracy=(TP + TN)/(TP+TN+FP+FN) =(3+1) /7=0.57

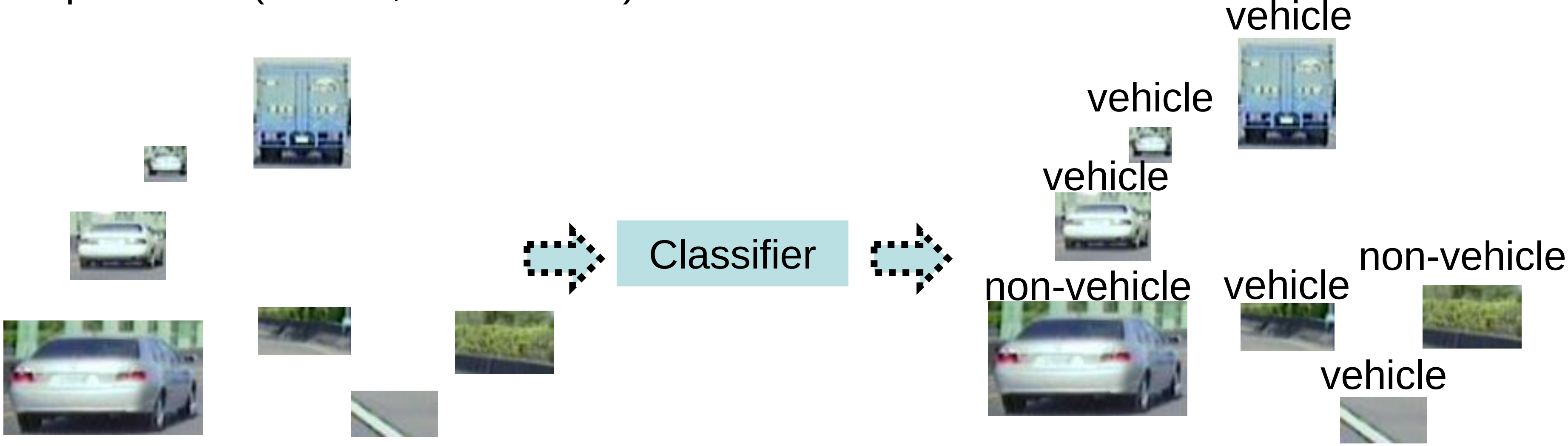
Recall(true positive rate, Detection rate) =TP/(TP+FN)=3/(3+1) =0.75

Precision=TP/(TP+FP) =3/(3+2)=0.6

False Positive Rate=FP/(FP+TN)=2/(2+1)=0.67

Another pure binary classification example

Input: images
Output: labels(vehicle, non-vehicle)



Accuracy=(TP + TN)/(TP+TN+FP+FN)
=(3+1) /7=0.57

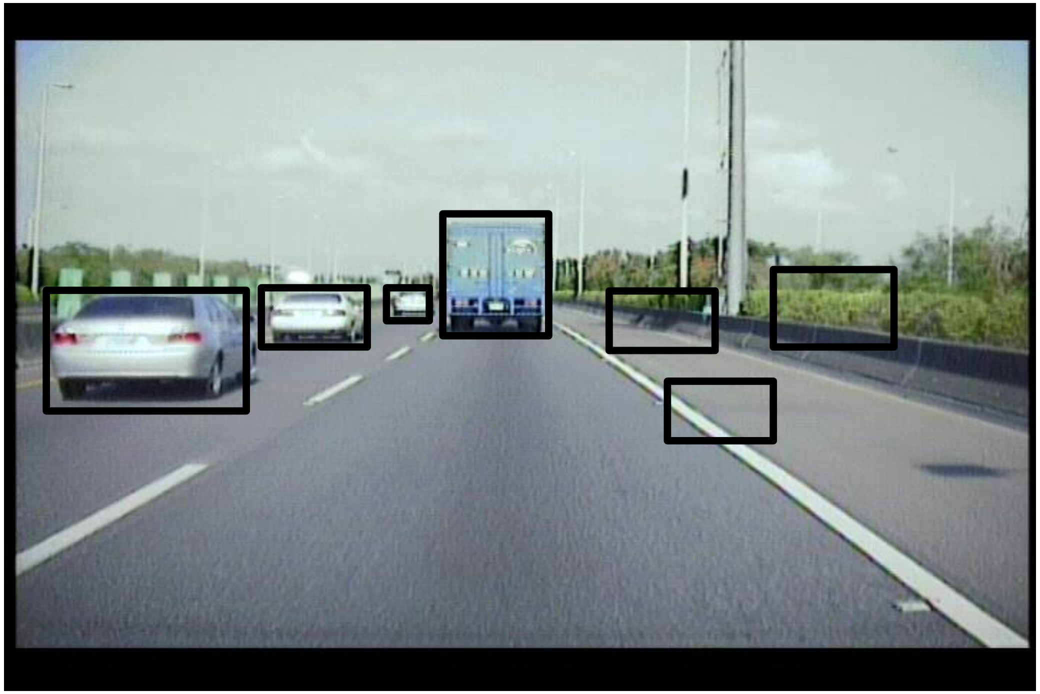
Recall(true positive rate, Detection rate)
=TP/(TP+FN)=3/(3+1)=0.75

Precision=TP/(TP+FP)=3/(3+2)=0.6

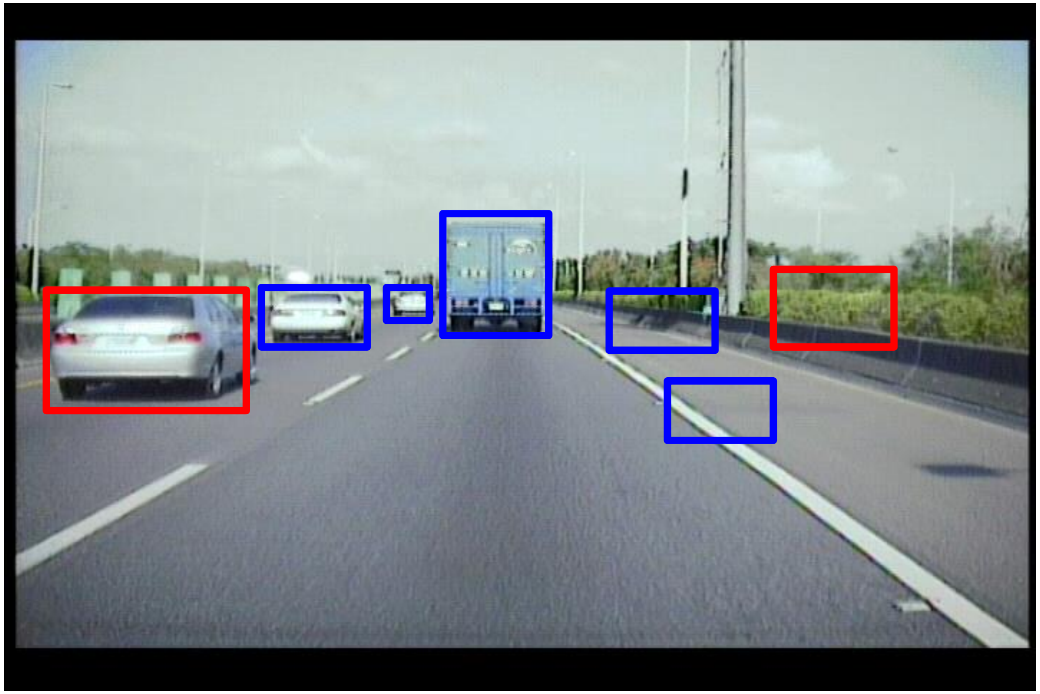
False Positive Rate=FP/(FP+TN)=2/(2+1)=0.67

How could we apply binary classification onto detection?

- The production of the following bounding boxes are called Hypothesis Generation.
- Alternatively, such bounding boxes are also called region proposals.
- HG is highly domain-specific.
- For example, vehicle candidates are not possibly in the sky.



Example for Vehicle Detection:



Accuracy=(TP + TN)/(TP+TN+FP+FN)
=(3+1) /7=0.57

Recall(true positive rate, Detection rate)
=TP/(TP+FN)=3/(3+1)=0.75

Precision=TP/(TP+FP)=3/(3+2)=0.6

False Positive
Rate=FP/(FP+TN)=2/(2+1)=0.67

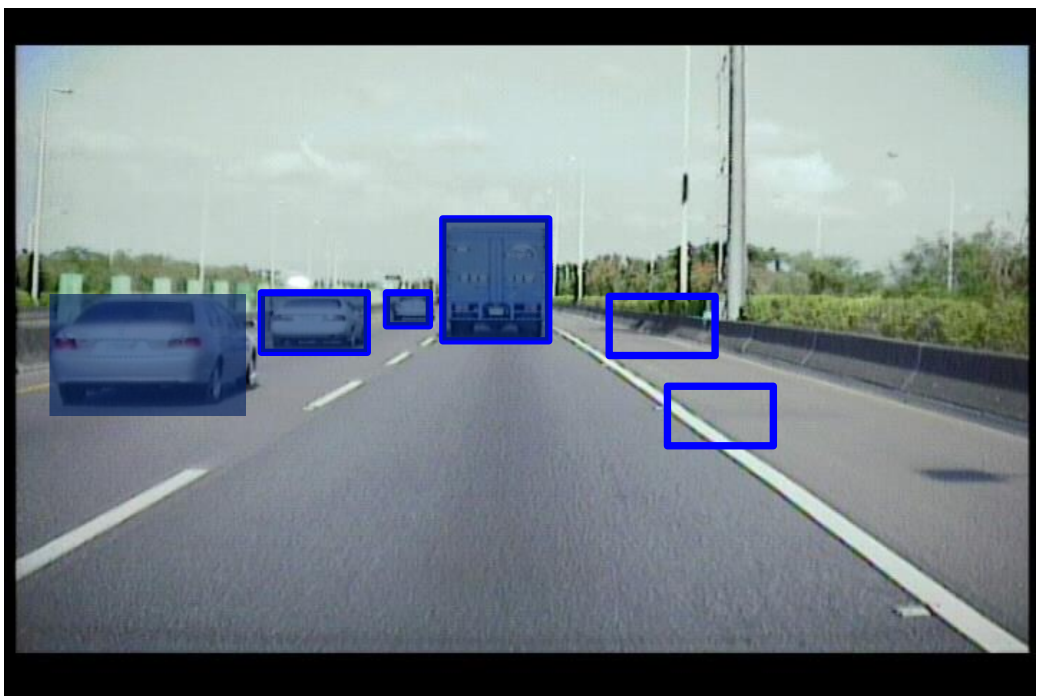
: vehicle

: non-vehicle

Such analysis is kind of impractical!
Why?
Non-vehicle (TN) is not the outcome of the final result!

Example for Vehicle Detection: In reality

- The final result only shows FP+TP because the production of region proposals is in the “black box”.
- The comparison between the detection result and the Ground-Truth helps to distinguish FP+TP from the results
- FN are those ground-truth labels not detected.



~~Accuracy=(TP + TN)/(TP+TN+FP+FN)~~
~~=(3+1) /7=0.57~~

Recall(true positive rate, Detection rate)
=TP/(TP+FN)=3/(3+1)=0.75

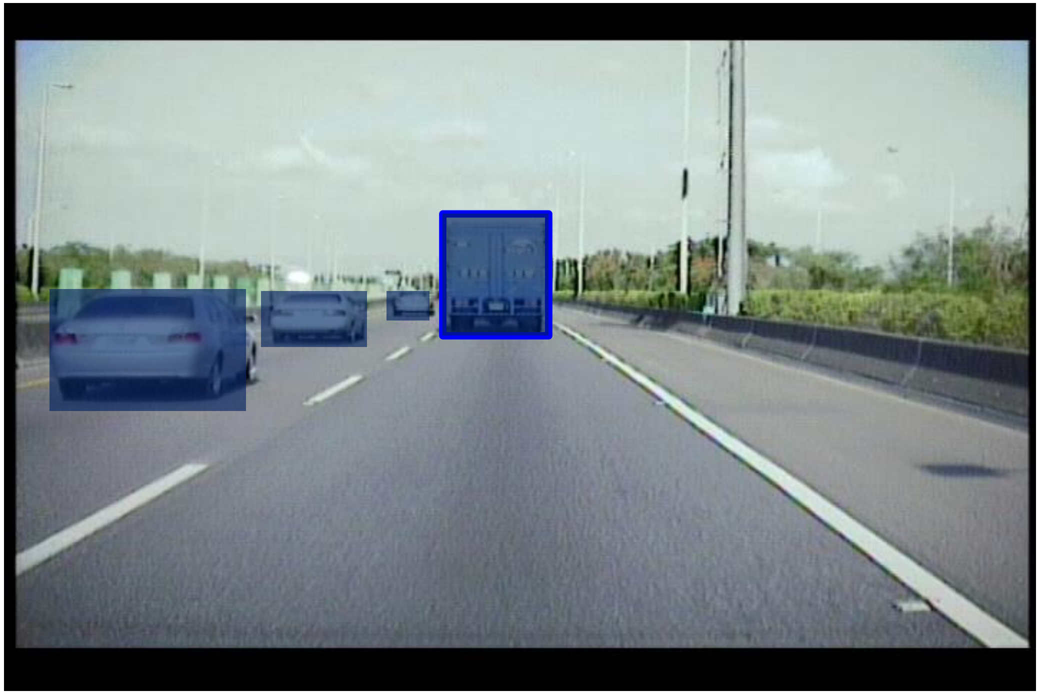
Precision=TP/(TP+FP)=3/(3+2)=0.6

~~False Positive Rate=FP/(FP+TN)=2/(2+1)=0.67~~

: ground-truth labeling

An extreme example for Vehicle Detection: high precision

- Sometimes, high precision might be misleading if recall is not calculated.



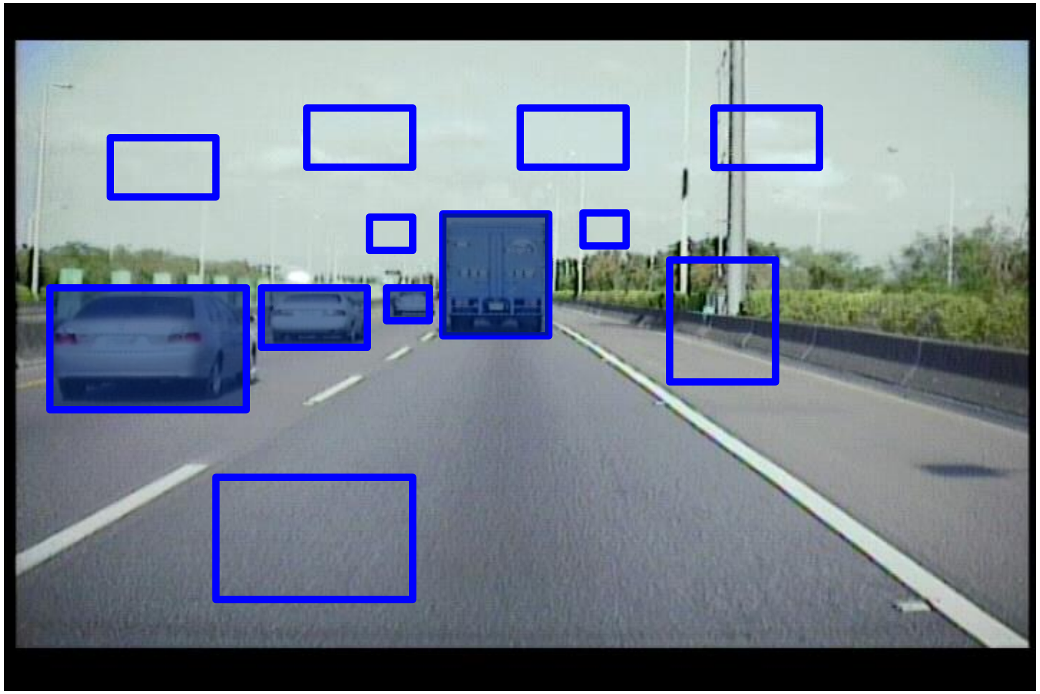
Recall(true positive rate, Detection rate)
=TP/(TP+FN)=1/(1+3)=0.25

Precision=TP/(TP+FP)=1/(1+0)=1

: ground-truth labeling

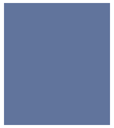
An extreme example for Vehicle Detection: high recall

- Also, high recall might be misleading if precision is not calculated.



Recall(true positive rate, Detection rate)
=TP/(TP+FN)=4/(4+0)=1

Precision=TP/(TP+FP)=4/(4+8)=0.33



: ground-truth labeling

Quantitative Result Analysis in Terms of “Confidence”

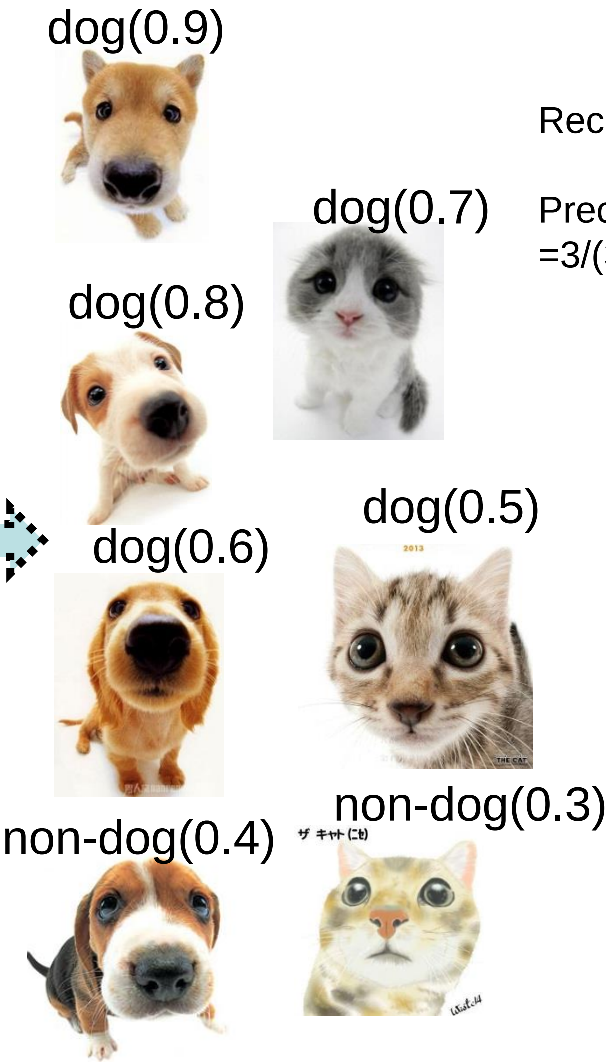
- Do objects with higher confidence tend to be more correct?
  Ans: Not exactly!
- Do we expect most detection error coming from low confidence?
  Ans: Yes!
- Do the above metrics reflect errors in high/low confidence ?
  Ans:No!
- Is there any metric to assess the detection results with confidence?
  Ans: Yes, It is called Average Precision(AP)

A Classification Example Considering Confidence: case 1

Input: images
Output: labels(dog, non-dog)

Threshold=(0.45)

Classifier



Recall=3/(3+1)=0.75

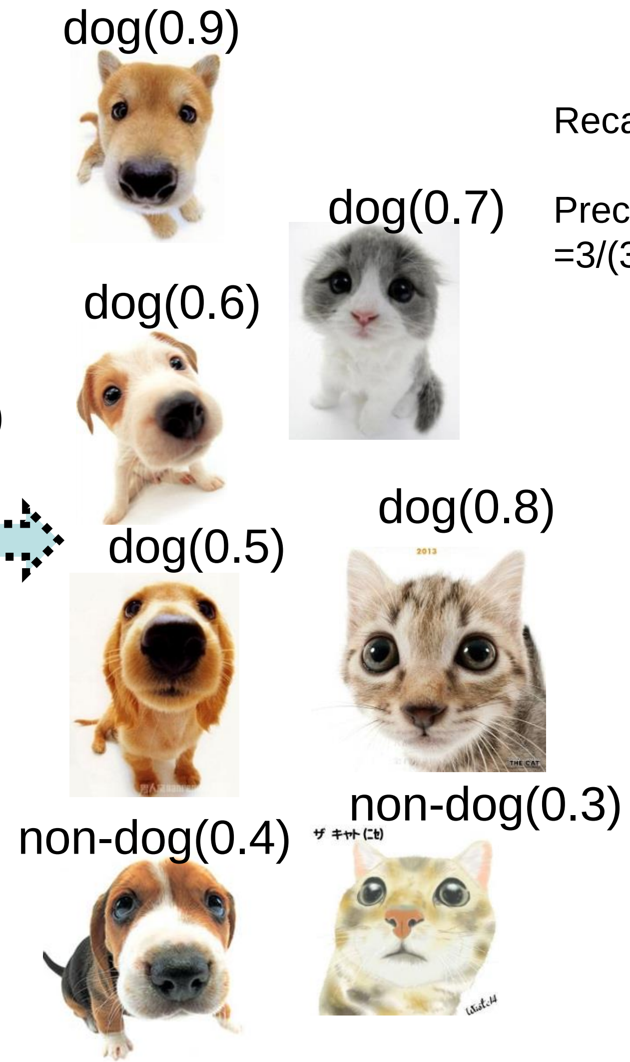
Precision=TP/(TP+FP)
=3/(3+2)=0.6

A Classification Example Considering Confidence: case 2

Input: images
Output: labels(dog, non-dog)

Threshold=(0.45)

Classifier



Recall=3/(3+1)=0.75

Precision=TP/(TP+FP)
=3/(3+2)=0.6

Comparison with case 1 & 2: which one is better?

- Case 1 is better because we expect the higher the confidence, the more correct the detection result

Case 1

dog(0.9)	dog(0.8)	dog(0.7)	dog(0.6)	dog(0.5)	Non-dog(0.4)	Non-dog(0.3)

Case 2

dog(0.9)	dog(0.8)	dog(0.7)	dog(0.6)	dog(0.5)	Non-dog(0.4)	Non-dog(0.3)

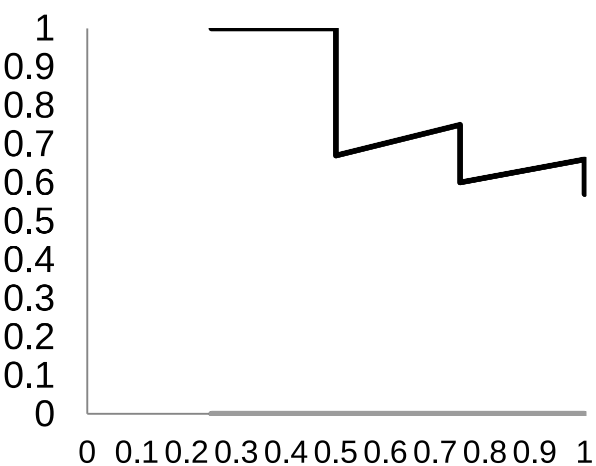
AP for case 1

dog(0.9)	dog(0.8)	dog(0.7)	dog(0.6)	dog(0.5)	Non-dog(0.4)	Non-dog(0.3)

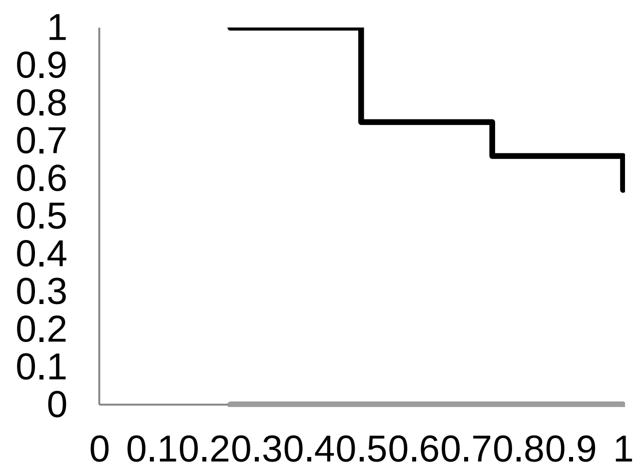
Precision	1/1	2/2	2/3	3/4	3/5	4/6	4/7
Recall	1/4	2/4	2/4	3/4	3/4	4/4	4/4

AP=0.5*1+0.25*0.75+0.25*0.66=0.85

$$\hat{P}(R) = \max_i \{P_i : R_i \geq R\}$$



For preserving monotonicity

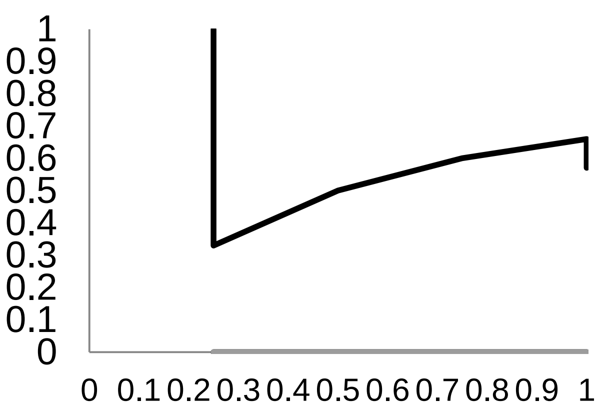


AP for case 2

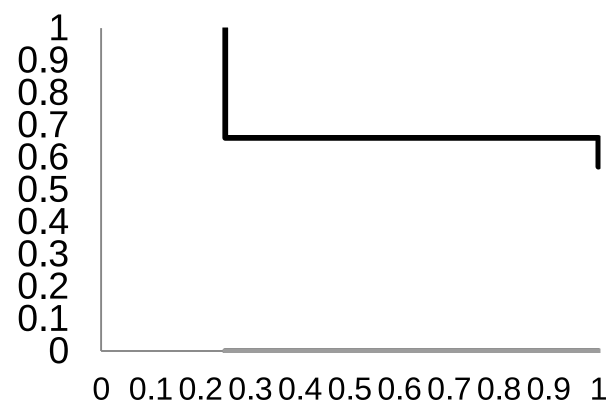
	dog(0.9)	dog(0.8)	dog(0.7)	dog(0.6)	dog(0.5)	Non-dog(0.4)	Non-dog(0.3)
							
Precision	**1/1**	**1/2**	**1/3**	**2/4**	**3/5**	**4/6**	**4/7**
Recall	**1/4**	**1/4**	**1/4**	**2/4**	**3/4**	**4/4**	**4/4**

AP=0.25*1+0.75*0.66=0.745

$$\hat{P}(R) = \max_i\{P_i : R_i \geq R\}$$

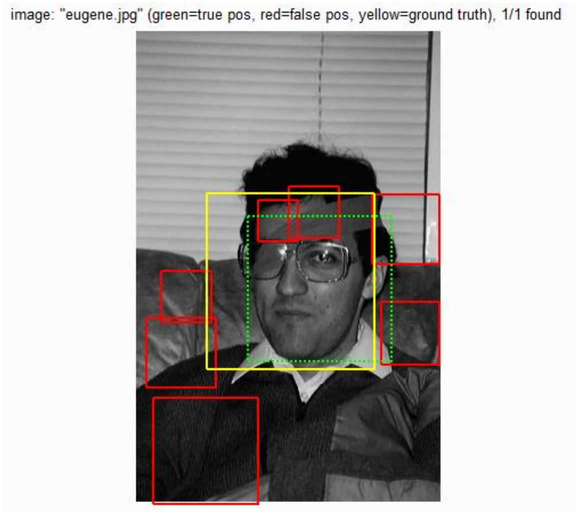


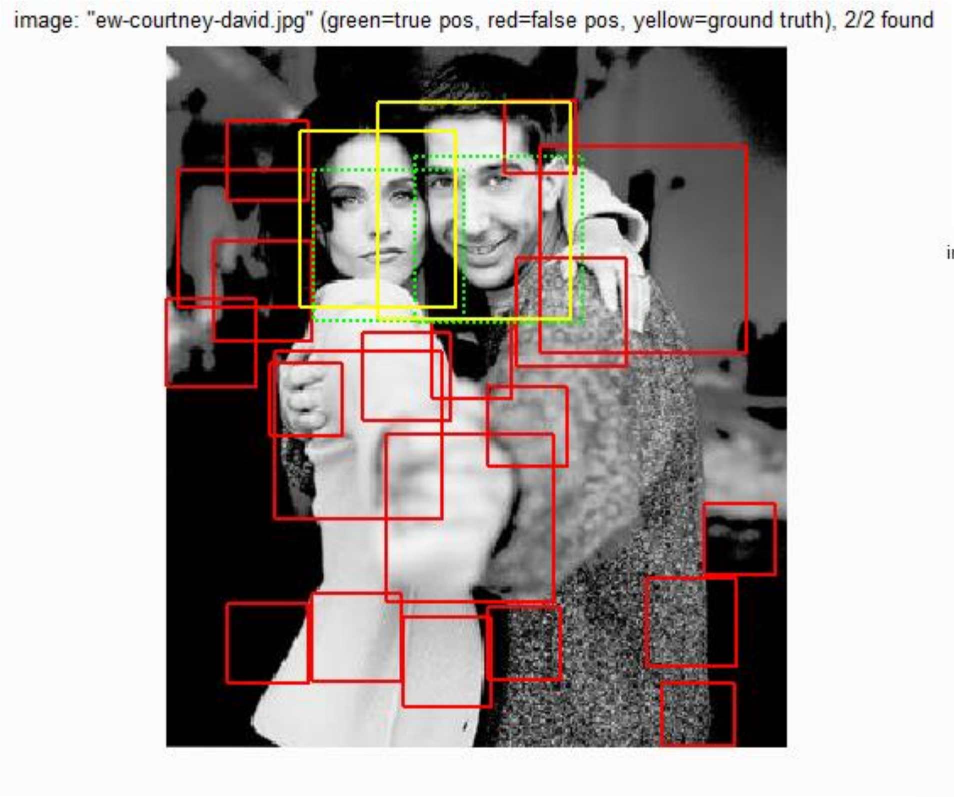
For preserving monotonicity

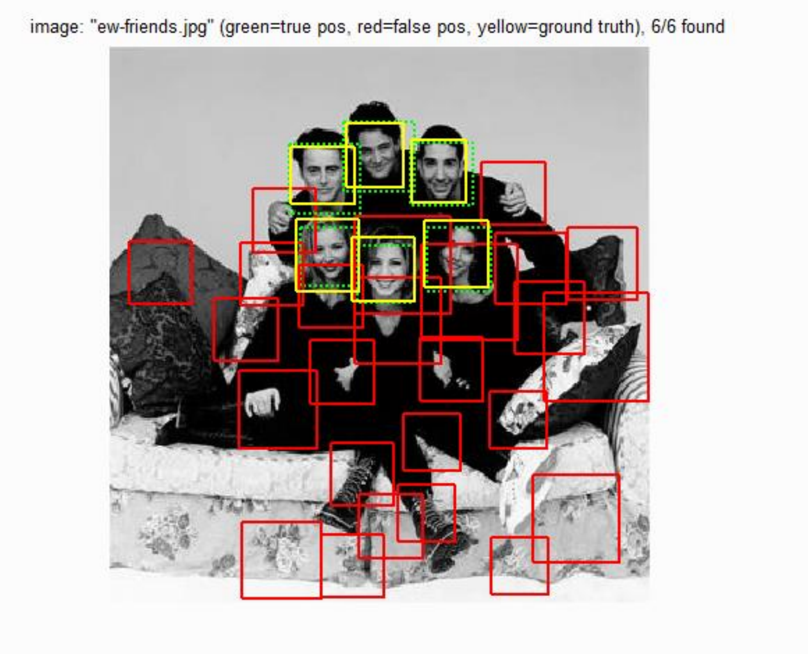


Do you really understand AP?

- How much AP do you think these 3 photos have?

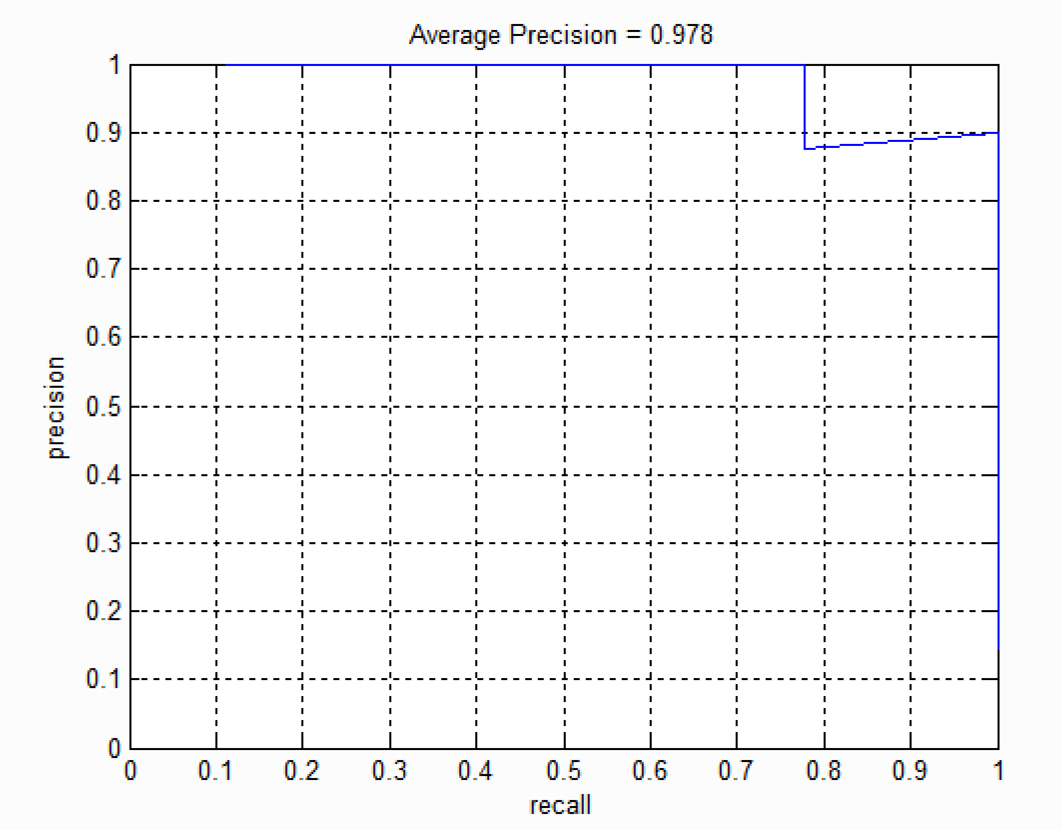






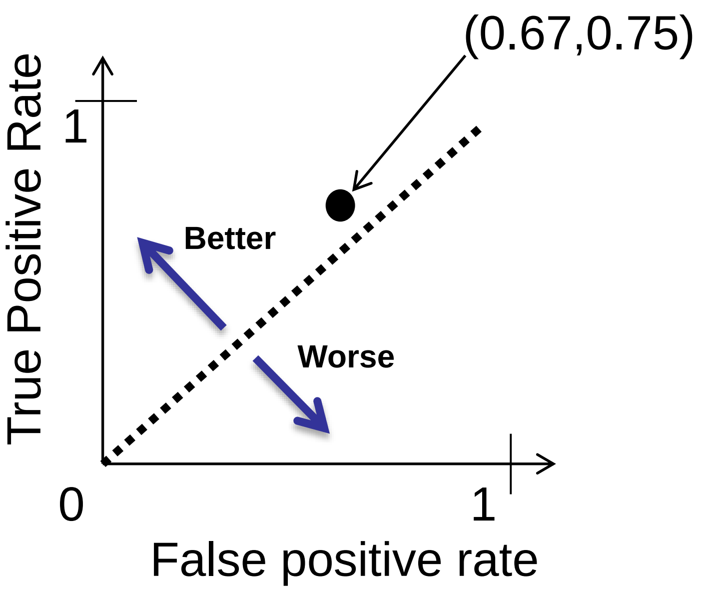
The idea of AP

- The precision-recall metric does not penalize a detector for producing false positives, as long as those false positives have lower confidence than true positives!!!



ROC curve

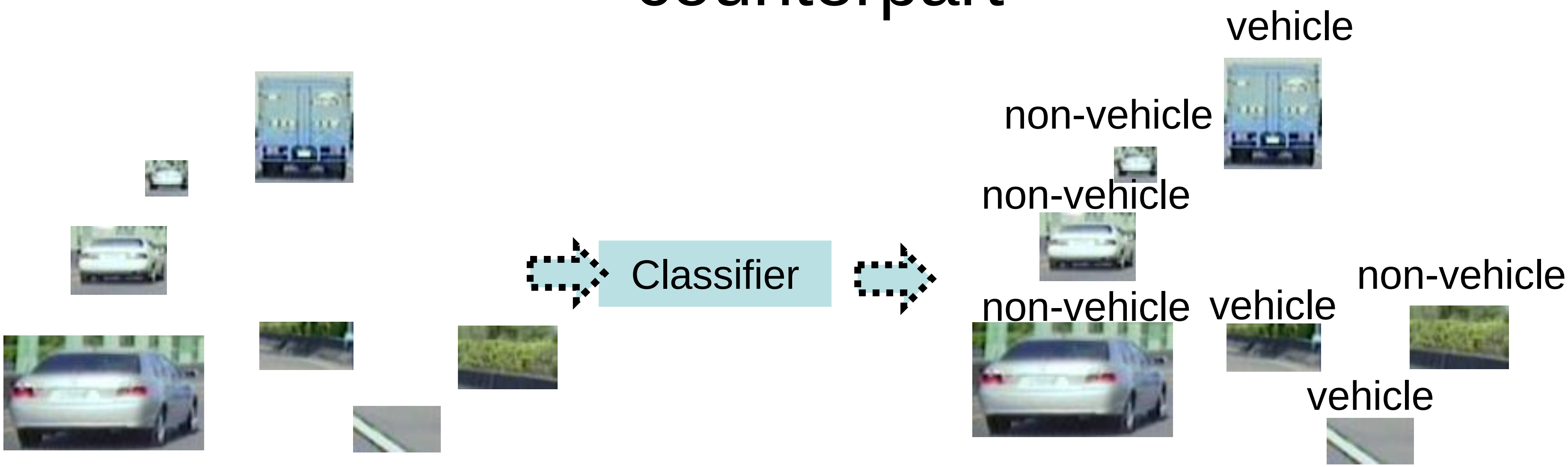
- ROC is called Receiver Operating Characteristic curve.
- n “point” in the ROC space could be obtained by classified results given
  - (1) n thresholds of a classifier or
  - (2) n classifiers
- Recursively substituting a lot of different thresholds, the precision and recall tell us would vary
- For a pure classification problem, TP, TN, FP & FN are all observable



Recall(true positive rate, Detection rate)
=TP/(TP+FN)=3/(3+1)=0.75

False Positive
Rate=FP/(FP+TN)=2/(2+1)=0.67

A poor binary classification example and its counterpart

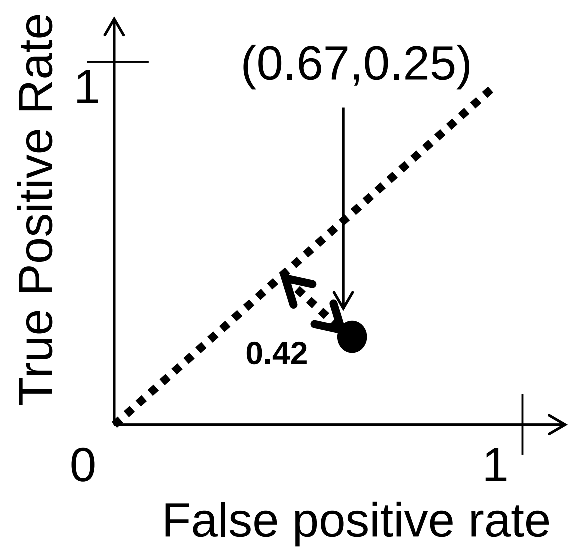


Accuracy=(TP + TN)/(TP+TN+FP+FN)
=(1+1) /7=0.29

Recall(true positive rate, Detection rate)
=TP/(TP+FN)=1/(1+3)=0.25

Precision=TP/(TP+FP)=1/(1+2)=0.33

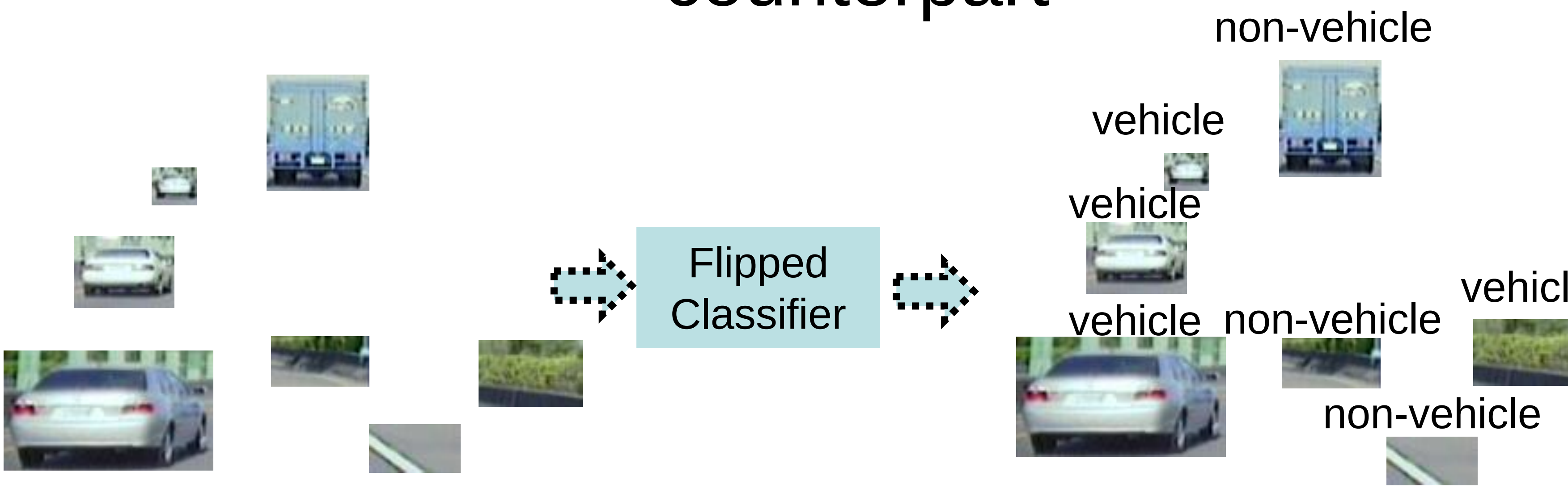
False Positive Rate=FP/(FP+TN)=2/(2+1)=0.67



Is this classifier really bad?

Not really!
We could flip the label to improve!!

A poor binary classification example and its counterpart

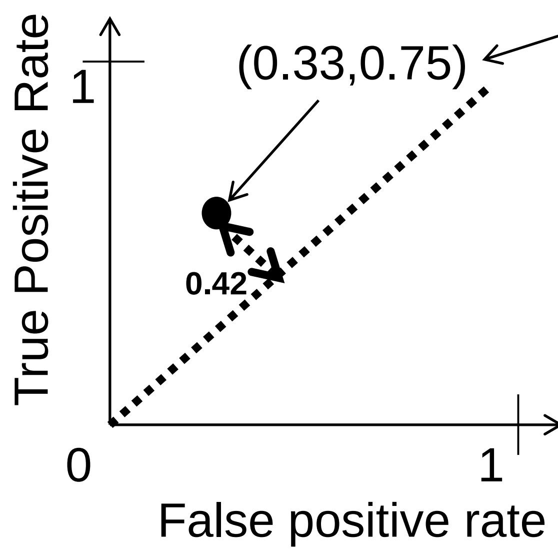


Accuracy=(TP + TN)/(TP+TN+FP+FN)
=(2+3) /7=

Recall(true positive rate, Detection rate)
=TP/(TP+FN)=3/(3+1)=0.75

Precision=TP/(TP+FP)=3/(3+1)=0.75

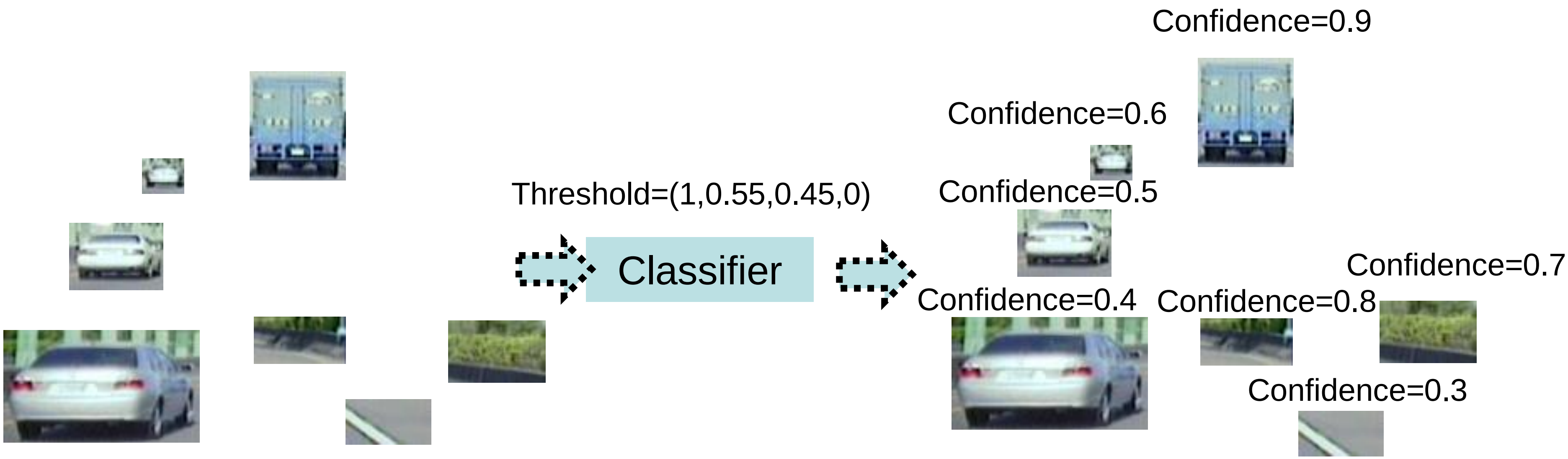
False Positive Rate=FP/(FP+TN)=1/(1+2)=0.33



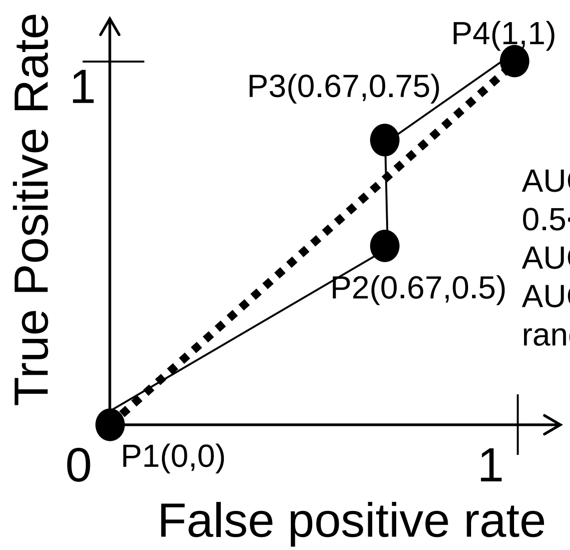
By mirroring (0.67,0.25)
with (0.5,0.5)

Now, It works better!

AUC of ROC



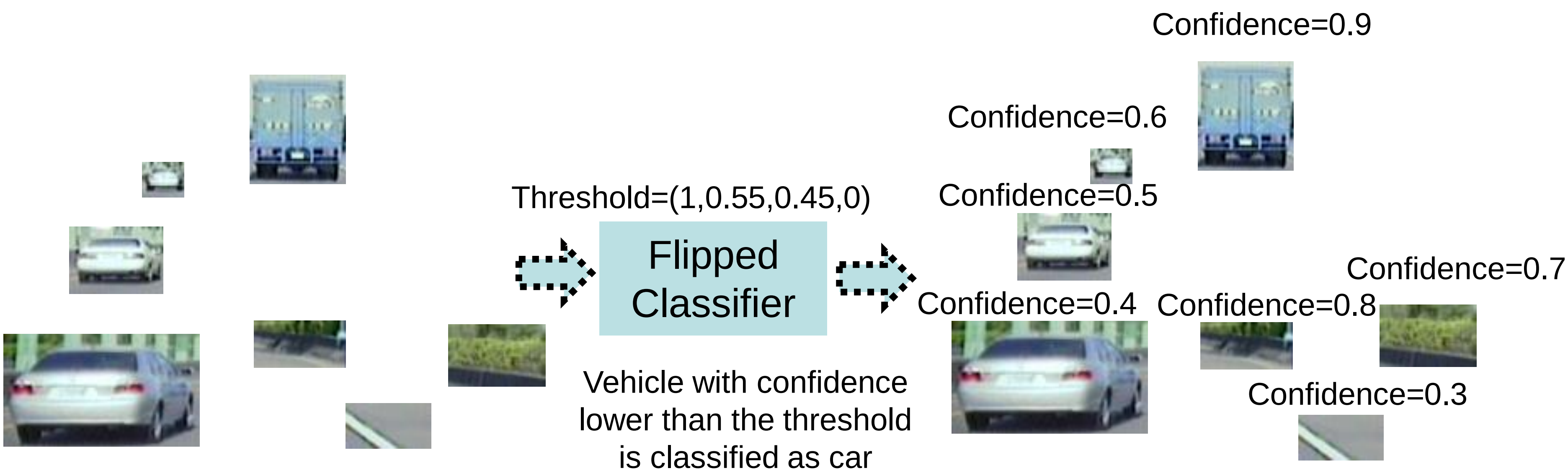
P1(Threshold=1):
False positive rate=FP/(FP+TN)=0/(0+3)=0
True positive rate=TP/(TP+FN)=0/(0+4)=0
P2(Threshold=0.55):
False positive rate=FP/(FP+TN)=2/(2+1)=0.67
True positive rate=TP/(TP+FN)=2/(2+2)=0.5
P3(Threshold=0.45):
False positive rate=FP/(FP+TN)=2/(2+1)=0.67
True positive rate=TP/(TP+FN)=3/(3+1)=0.75
P4(Threshold=1):
False positive rate=FP/(FP+TN)=3/(3+0)=1
True positive rate=TP/(TP+FN)=4/(4+0)=1



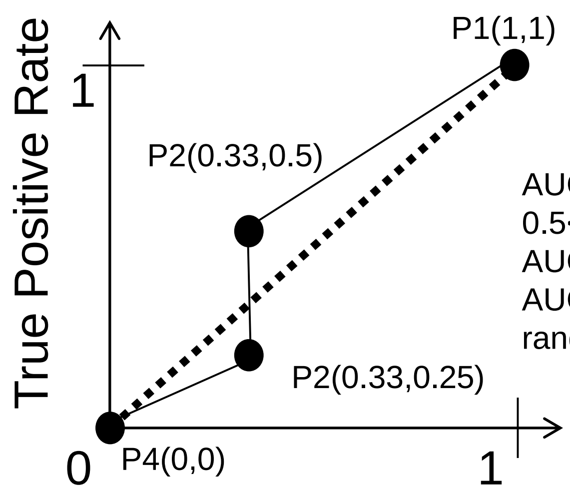
AUC=0.45625

AUC=1: perfect classifier
0.5<AUC<1: better than random guess
AUC=0.5: random guess and worthless
AUC<0.5: flip to get results better than random guess

AUC of ROC with flipped classifier



P1(Threshold=1):
False positive rate=FP/(FP+TN)=3/(3+0)=1
True positive rate=TP/(TP+FN)=4/(4+0)=1
P2(Threshold=0.55):
False positive rate=FP/(FP+TN)=1/(1+2)=0.33
True positive rate=TP/(TP+FN)=2/(2+2)=0.5
P3(Threshold=0.45):
False positive rate=FP/(FP+TN)=1/(1+2)=0.33
True positive rate=TP/(TP+FN)=1/(1+3)=0.25
P4(Threshold=1):
False positive rate=FP/(FP+TN)=0/(0+3)=0
True positive rate=TP/(TP+FN)=0/(0+4)=0



AUC=0.54375

AUC=1: perfect classifier
0.5<AUC<1: better than random guess
AUC=0.5: random guess and worthless
AUC<0.5: flip to get results better than random guess

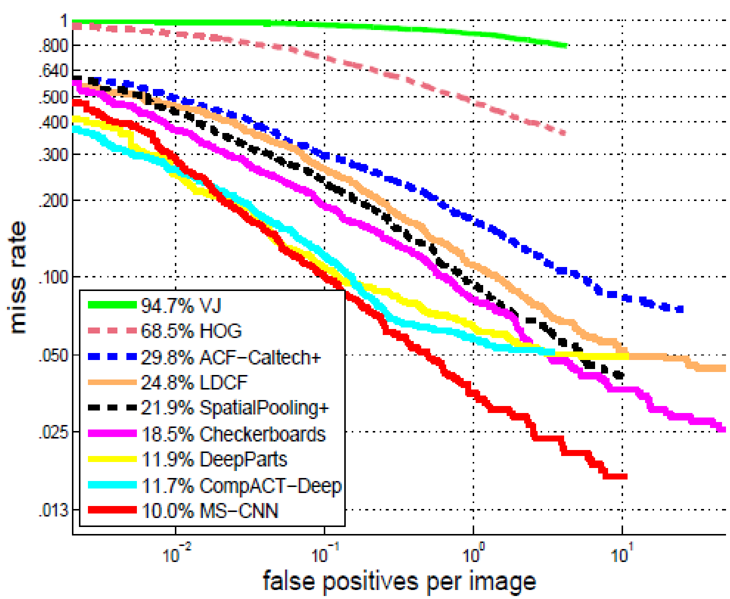
Now, It works better!

Miss rate vs FPPI curve

- FPPI stands for False Positives Per Image
- Normally, both recall and FPs increase when threshold is lowered.
- In other words, FNs decrease and FPs increase when threshold is lowered.
- Which method is the best one in the following figure?
  - Ans: MS-CNN (the lowest one!!)
- This metric is used to analyze how good this classifier is in terms of confidence.

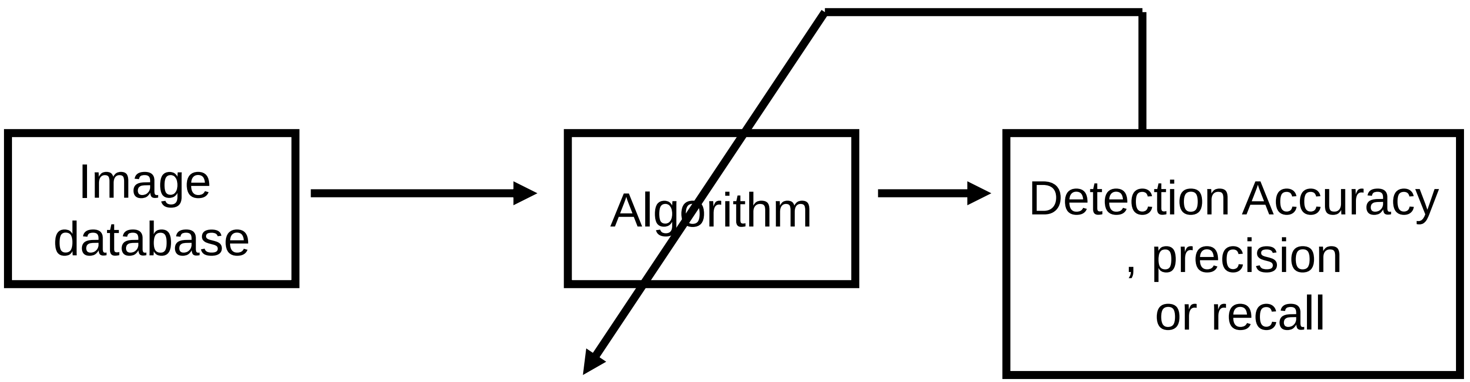
Miss Rate=FN/(TP+FN)

FPPI=$\sum_{i=1}^{N} FP_i/N$



Benefit of building Ground-Truth database

- Optimizing parameters in LDW, FCW,..etc



Typical testing database

Weather	Images	precision	recall	Average precision
Sunny day	Xxxxxx	82.15%	93.75%	?
Cloudy day	xxxx	88.33%	94.91%	?
Raining day				
Raining night				
Night				
• • •				

What's Problem of Vehicle Testing ?

- 測試所需的路型與標線如何取得?
  - Ex:歐盟, 美國, 澳洲
- 需求的交通情境如何掌控?
  - Ex:測試車周邊車輛行為
- 特定的天候與光線環境如何取得?
  - 日間 / 夜間 / 黃昏 / 雨 / 霧
- 如何重覆執行相同的測試?
  - ISO測試規範

Driving Simulation & Verification

- 藉由虛擬實境系統進行即時光影模擬，提供車輛影像安全輔助系統辨識率統計與除錯使用。
- 高精度之車輛物理模型搭配即時運算平台，提供同步之車輛動態運算結果。
- 提供車輛影像安全系統所需之車輛訊號，並接收相關訊號，進行HIL測試(Hardware-in-the-Loop)。
- 評分系統於測試過程同步提供辨識率統計結果與誤判之影\像資料。
- 可重現辨識錯誤情境提供演算法修正，確保系統於實車測試前已具備相當之系統強健性與可靠度。
- 模擬駕駛環境無法完全取代實車測試，但可以大幅降低實車測試成本與次數。

Surrounding Projection Systems

User Interface

Real car Interior

Force feed-back

Steering System

Throttle/ Brake paddle/

Gear shifter/

Functional Buttons

VR光影系統介紹

路面反光與柏油補丁

道路指示牌陰影

路燈桿陰影

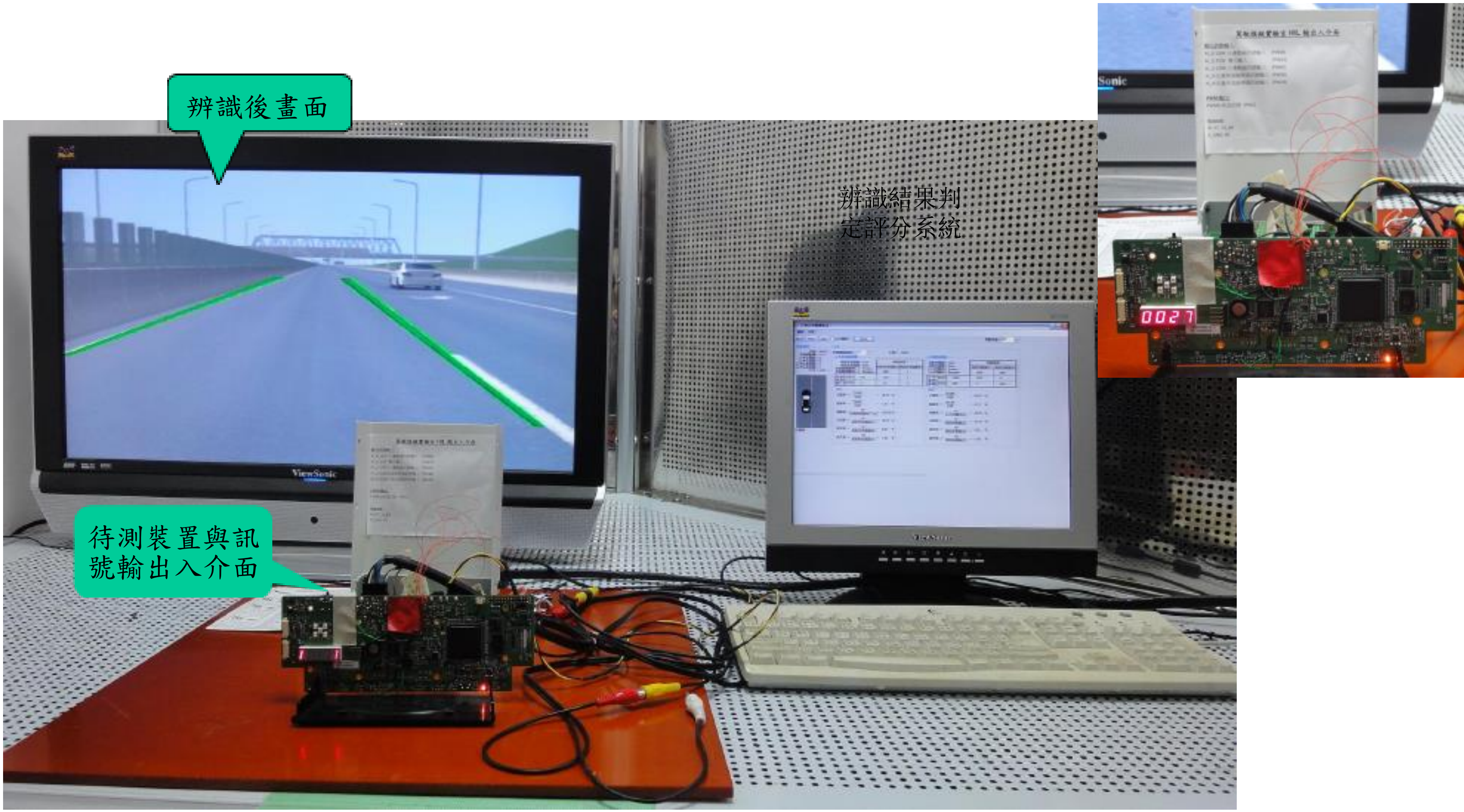
遮光版陰影

樹木陰影

橋樑陰影

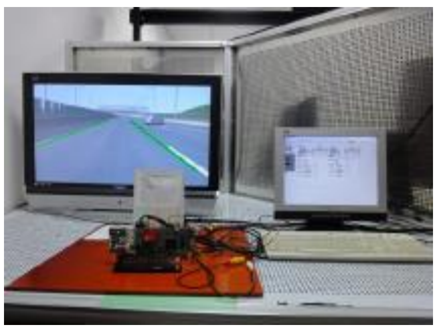
LDW/ FCW HIL Testing

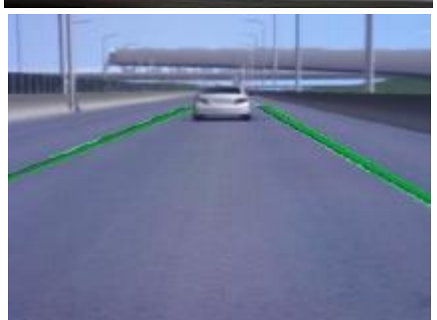
~ Virtual Lab. ~



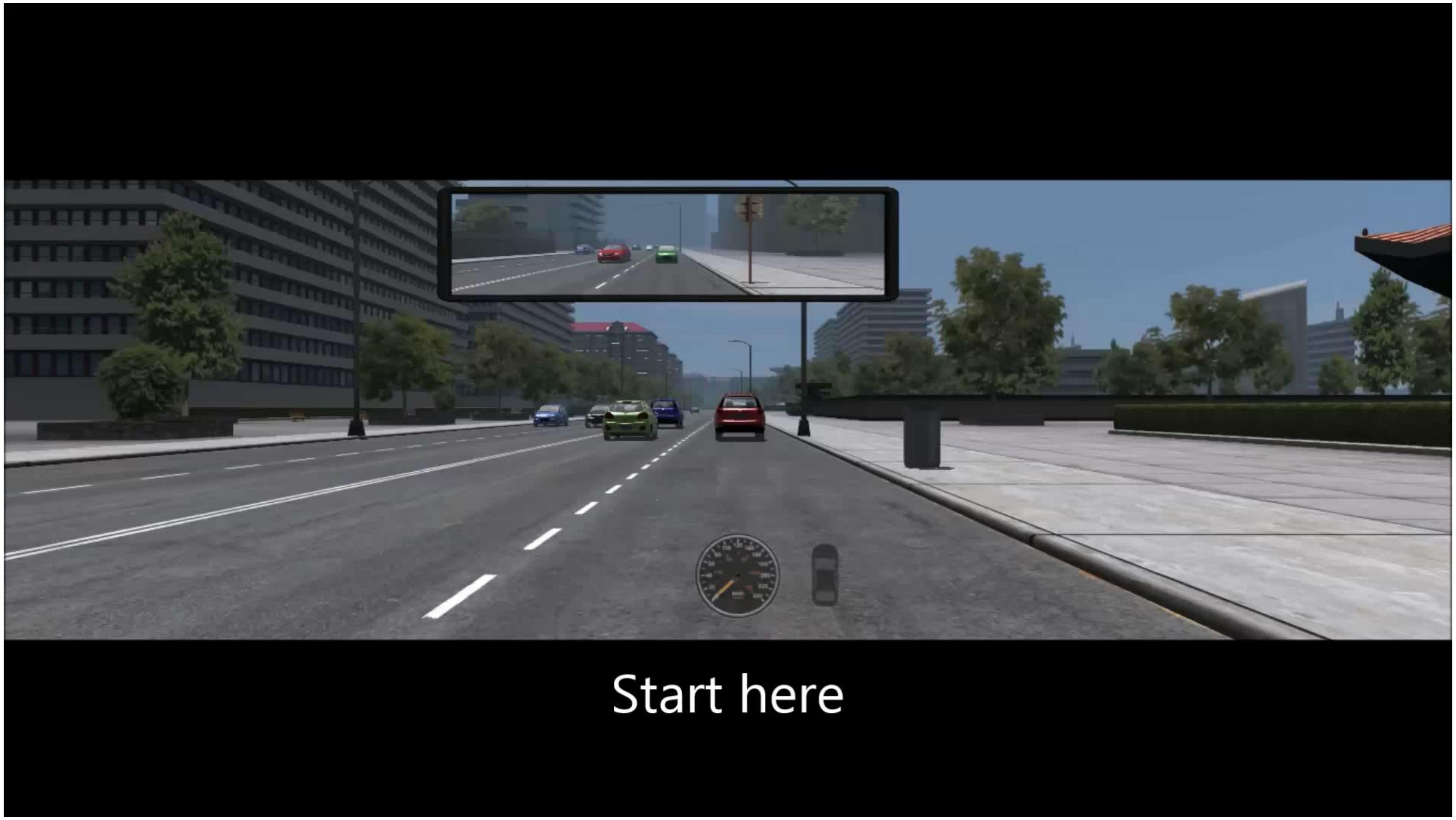
LDW/ FCW HIL Testing

- 泛用型LDW 自動評分系統
  - 藉由比對Ground Truth與LDW裝置辨識結果自動進行辨識正確率判定，同時進行完整的辨識結果統計，包含正確率、錯誤率、精確率、召回率、誤判率與遺失率等，完全不需人工辨識且可批次化自動進行，有效協助開發工程師進行系統除錯並大幅縮短驗證所需時間與人力。





How realistic does a simulated driving simulator could achieve?



A very real simulator: GTA

Observation

- Don’t directly develop on a platform which can’t use any libraries.
- The time of Big Data is coming.
- Don’t label all the data by yourself.
- GPU will be everywhere in the future.
- In-The-Loop tests are very important because no programmer could do the right thing right the first time.
- Winner takes all.