Training : Computer Vision

Hackathon CentraleSupélec-ESSEC

14/02/2022

eleven
strategy consultants
London • Paris • Madrid

AGENDA

1. Computer vision approaches
2. An overview of the main Deep Learning algorithms
3. The “open” dataset : our gold mine
4. Measure the performance to improve your algorithm

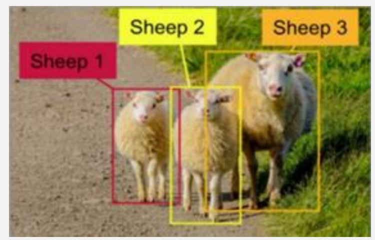
The amount of data available is a critical element to consider in an image-based application

Little data

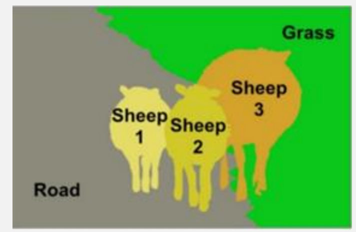
Lots of data

Image recognition

Object detection



Object segmentation



1. More hard engineering (“hacks”)
2. More ad-hoc algorithm
3. Transfer learning

1. More standard algorithms
2. Less hand engineering

Sources of knowledge

1. Labeled data (e.g. image → dog & cat)
2. Hand engineered features, network architecture, etc.

Computer vision can be broken down into a multitude of blocks that can be combined as needed (1/2)

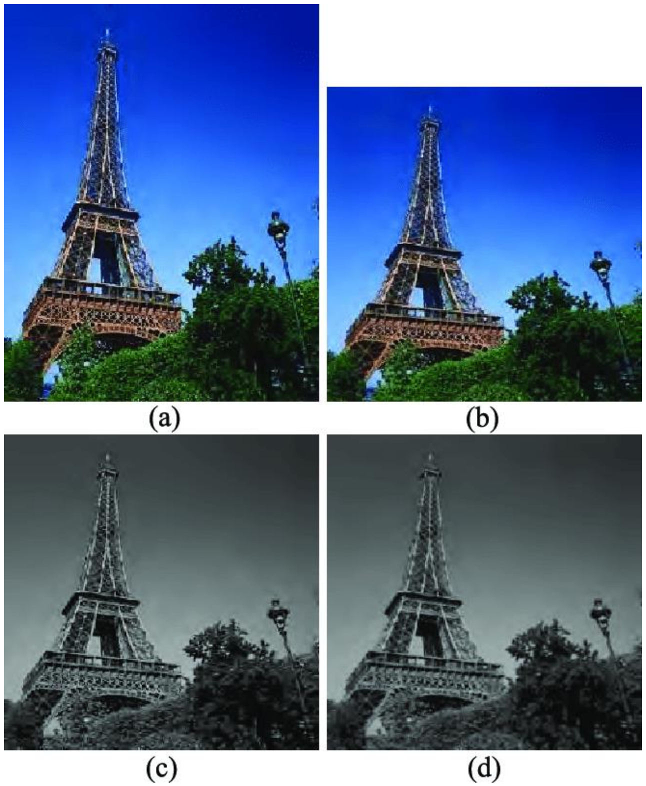
Image Pre-processing

Image scaling

Contrast enhancement

Cropping and padding

Noise reduction



(a)Original image

(b)Resized

(c)Grayed

(d)Blured

Computer vision can be broken down into a multitude of blocks that can be combined as needed (1/2)

Image Pre-processing

Image scaling

Contrast enhancement

Cropping and padding

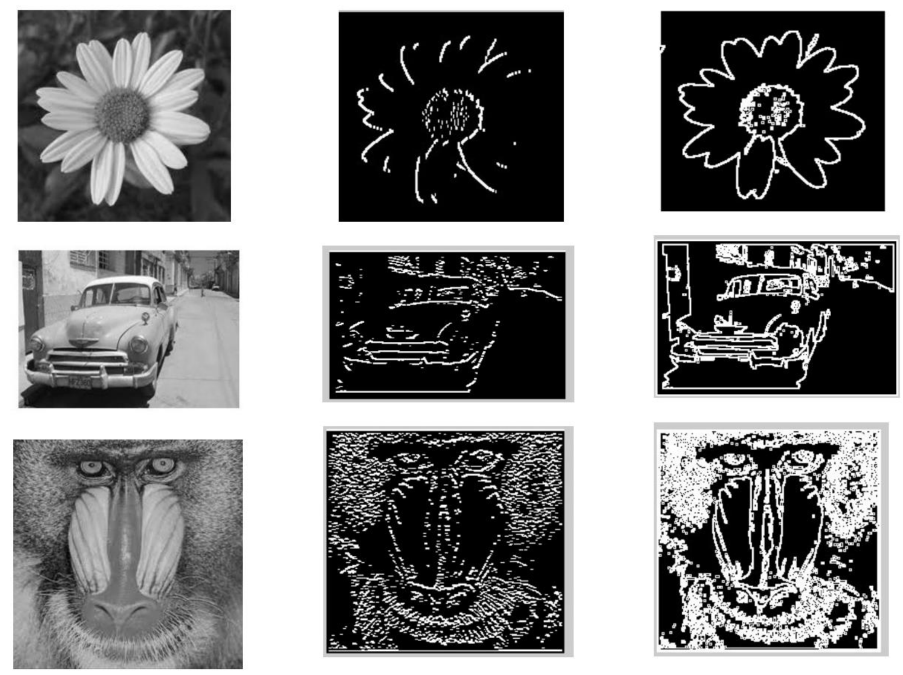
Noise reduction

Feature extraction

Lines/Edge detection

Corners/Blobs/Points

(SIFT/SURF, PCA, Watershed)



Computer vision can be broken down into a multitude of blocks that can be combined as needed (1/2)

Image Pre-processing

Image scaling

Contrast enhancement

Cropping and padding

Noise reduction

Feature extraction

Lines/Edge detection

Corners/Blobs/Points

(SIFT/SURF, PCA, Watershed)

Classification algorithm

CNN

Cat

Computer vision can be broken down into a multitude of blocks that can be combined as needed (1/2)

Image Pre-processing

Image scaling

Contrast enhancement

Cropping and padding

Noise reduction

Feature extraction

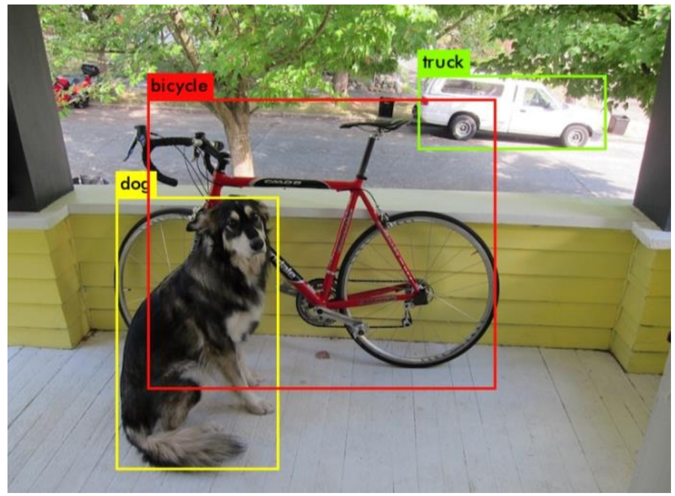
Lines/Edge detection

Corners/Bl

(SIFT/SU

Water

Classification algorithm



Object detection and/or Identification

YOLO

Computer vision can be broken down into a multitude of blocks that can be combined as needed (1/2)

Image Pre-processing

Image scaling

Contrast enhancement

Cropping and padding

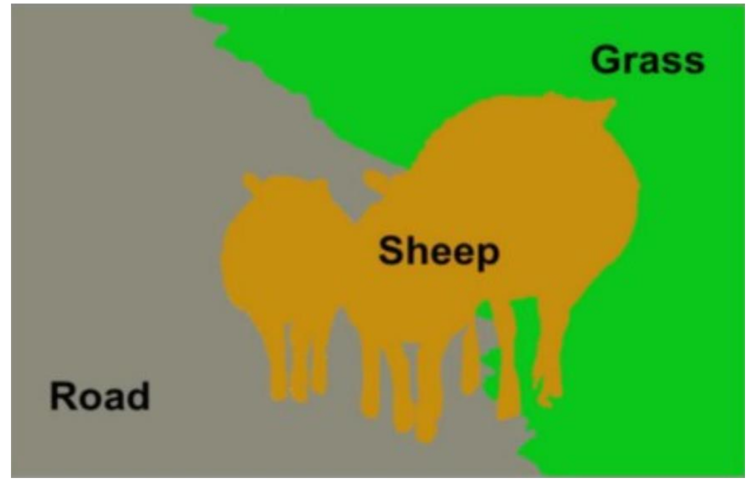
Noise reduction

Feature extraction

Lines/Edge detection

Corners/Blobs/Points

(SIFT/SURF, PCA, Watershed)



Segmentation

Faster R-CNN

Computer vision can be broken down into a multitude of blocks that can be combined as needed (2/2)

Image Pre-processing

Image scaling

Contrast enhancement

Cropping and padding

Noise reduction

Feature extraction

Lines/Edge detection

Corners/Blobs/Points

(SIFT/SURF, PCA, Watershed)

Classification algorithm

CNN

Object detection and/or Identification

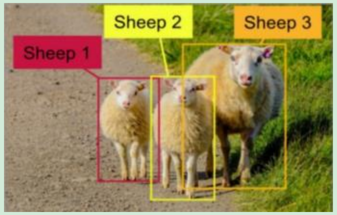
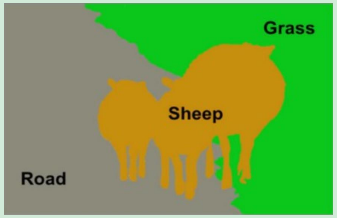
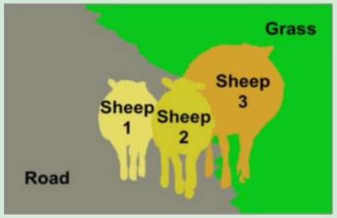
YOLO

- We can start our process by applying image pre-processing (for instance if we have few data)
- Later we can use detection algorithm in order to find the object of interest
- Finally we can leverage feature extraction to post process the results to make them more precise of follow some field rules

AGENDA

1. Computer vision approaches
2. An overview of the main Deep Learning algorithms
3. The “open” dataset : our gold mine
4. Measure the performance to improve your algorithm

The 3 main approaches to machine learning vision are Classification, Detection and Segmentation

Approach	Objective	Algorithms	Examples
Classification	Associate a class, defining its content, to each image	▪ CNN	
Detection	Spot objects in an image according to predefined categories	▪ YOLO ▪ SSD ▪ Faster R-CNN ▪ RFCN	
Segmentation	**Semantic:** Link a category to each pixel in an image. Recognize a set of pixels that form distinct categories	▪ Unet ▪ PSPNet ▪ ICNet ▪ DeepLab	
	Per instance: Associate each pixel with the instance to which it belongs.	▪ Mask R-CNN ▪ ResNet	

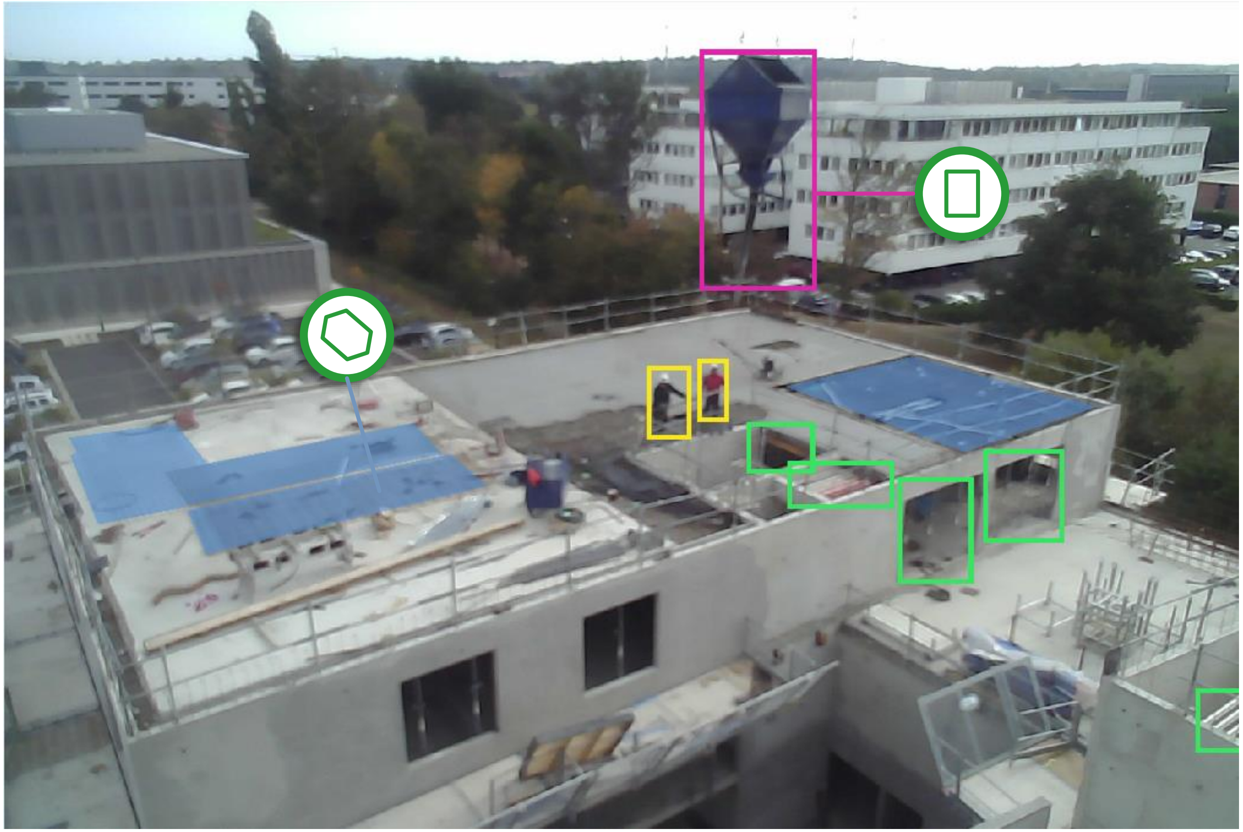
These different approaches are based on two different types of annotation

Polygon

- Polygon annotations select all pixels belonging to an instance. It is used in segmentation

Bounding box

- Bounding box annotations give position, size and type of object contained in a rectangle parallel to the axes: used in detection



AGENDA

1. Computer vision approaches
2. An overview of the main Deep Learning algorithms
3. **The "open" dataset : our gold mine**
4. Measure the performance to improve your algorithm

To train these models several datasets are available online

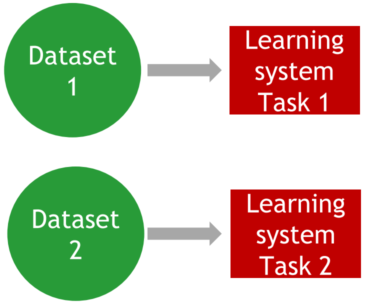
Name	Type	Annotation
COCO	Natural landscapes	Instance segmentation
ADE20K		
SUN		
IMAGE NET		Bounding boxes detection
PASCAL VOC2012		
CITYSCAPES	City landscapes	Instance segmentation

- Is the context of the dataset close to context of my images?
- Does the format of the dataset compatible with my algorithm?
- What classes are labeled in the dataset?
- Is the dataset balanced between classes?

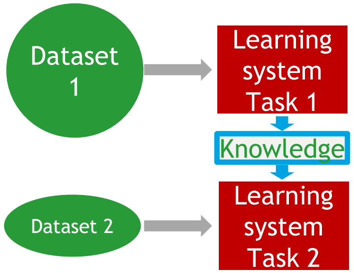
Transfer learning uses knowledge acquired for one task to solve related ones

General overview

- **Isolated**, single task learning:
  - Knowledge is not retained or accumulated
  - Learning is performed without considering past learned knowledge in other tasks



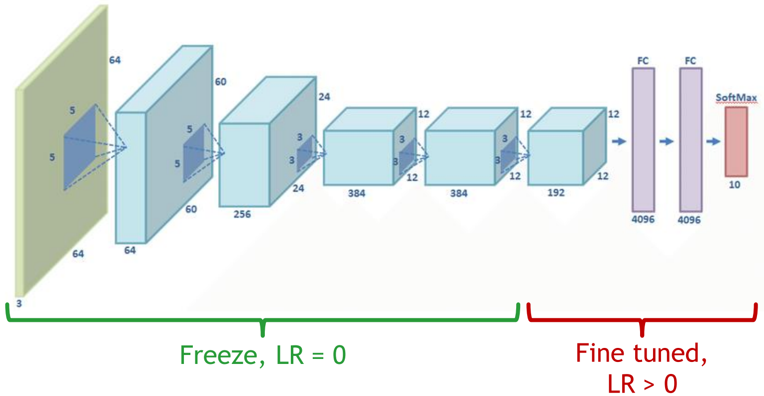
- **Learning** of a new task relies on the **previous learned tasks**:
  - Learning process can be faster, more accurate and/or need less training data



- There are two **variations** of transfer learning:
  - Same domain but different tasks
  - Same task, but different domains

Transfer learning in deep learning

- Use **pre-trained model** and **freeze** most layers of the neural network
  - Freeze layer imply that **weights are not updated** during the backpropagation, the learning rate (LR) is equal to 0
- Apply **fine tuning to lasts layers** in order to adapt the model to the specific task
  - The **weights** of fine-tuned layers are **updated** during the backpropagation



- **Keep** simple and **general features** from shallow layers
- **Retrain complex features** from deep layers

AGENDA

1. Computer vision approaches
2. An overview of the main Deep Learning algorithms
3. The “open” dataset : our gold mine
4. **Measure the performance to improve your algorithm**

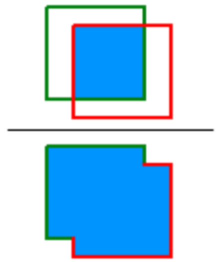
Several metrics can be used in computer vision depending on the objectives to reach

Local Metrics

- **Intersect over Union (IoU)**

Measuring the common cover on an object can be formalized as:

$$\boldsymbol{IoU} = \frac{\mathbf{aire}(\mathbf{Box}_{\mathbf{True}} \cap \mathbf{Box}_{\mathbf{Predicted}})}{\mathbf{aire}(\mathbf{Box}_{\mathbf{True}} \cup \mathbf{Box}_{\mathbf{Predicted}})}$$



We set a threshold $\varepsilon \in [0,1]$

Prediction are considered true if $\boldsymbol{IoU} \geq \boldsymbol{\varepsilon}$.

Model Metrics

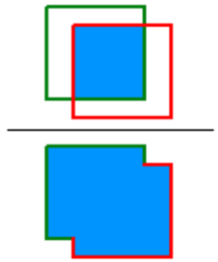
Several metrics can be used in computer vision depending on the objectives to reach

Local Metrics

- **Intersect over Union (IoU)**

Measuring the common cover on an object can be formalized as:

$$\boldsymbol{IoU} = \frac{\mathbf{aire}(\mathbf{Box}_{\mathbf{True}} \cap \mathbf{Box}_{\mathbf{Predicted}})}{\mathbf{aire}(\mathbf{Box}_{\mathbf{True}} \cup \mathbf{Box}_{\mathbf{Predicted}})}$$



We set a threshold $\varepsilon \in [0,1]$

Prediction are considered true if $\boldsymbol{IoU \geq \varepsilon}$.

Model Metrics

- **Mean Intersect over Union (mIoU)**

For a **fixed minimum confidence** level, the IoU is calculated for each annotation and then averaged over a class or globally.

- No indication of the type of error: on/under-prediction?
- Depends on a confidence threshold set

- Very "visual" metric, easy to interpret

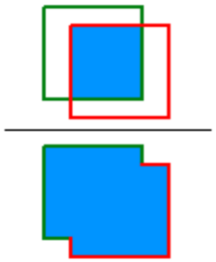
Several metrics can be used in computer vision depending on the objectives to reach

Local Metrics

- **Intersect over Union (IoU)**

Measuring the common cover on an object can be formalized as:

$$IoU = \frac{\text{aire}(\text{Box}_{\text{True}} \cap \text{Box}_{\text{Predicted}})}{\text{aire}(\text{Box}_{\text{True}} \cup \text{Box}_{\text{Predicted}})}$$



We set a threshold $\varepsilon \in [0,1]$

Prediction are considered true if $IoU \geq \varepsilon$.

- **Precision/Recall :**

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

Precision answers the question: **What share of the detected objects were the right ones?**

Recall answers the question: **What is the share of the objects that have been detected?**

Model Metrics

- **Mean Intersect over Union (mIoU)**

For a **fixed minimum confidence** level, the IoU is calculated for each annotation and then averaged over a class or globally.

- No indication of the type of error: on/under-prediction?
- Depends on a confidence threshold set

- Very "visual" metric, easy to interpret

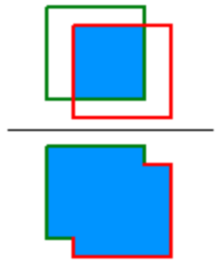
Several metrics can be used in computer vision depending on the objectives to reach

Local Metrics

- ### Intersect over Union (IoU)

Measuring the common cover on an object can be formalized as:

$$IoU = \frac{\text{aire}(\mathbf{Box}_{\mathbf{True}} \cap \mathbf{Box}_{\mathbf{Predicted}})}{\text{aire}(\mathbf{Box}_{\mathbf{True}} \cup \mathbf{Box}_{\mathbf{Predicted}})}$$



We set a threshold $\varepsilon \in [0,1]$

Prediction are considered true if $\boldsymbol{IoU} \geq \boldsymbol{\varepsilon}$.

- ### Precision/Recall :

$$\text{precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP}+\boldsymbol{FP}}$$

Precision answers the question: **What share of the detected objects were the right ones?**

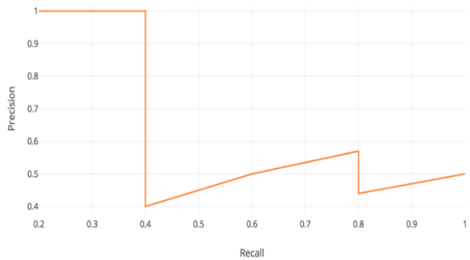
$$\text{recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP}+\boldsymbol{FN}}$$

Recall answers the question: **What is the share of the objects that have been detected?**

- ### Precision-recall curve :

Precision and recall are calculated for different confidence levels (proba of the class model) in order to obtain the curve:

$$precision = f(recall)$$



Precision-recall curve

Model Metrics

- ### Mean Intersect over Union (mIoU)

For a **fixed minimum confidence** level, the IoU is calculated for each annotation and then averaged over a class or globally.

- No indication of the type of error: over/under-prediction?
- Depends on a confidence threshold set

- Very "visual" metric, easy to interpret

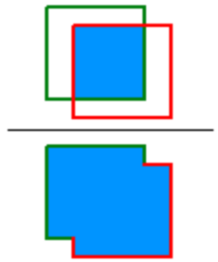
Several metrics can be used in computer vision depending on the objectives to reach

Local Metrics

- **Intersect over Union (IoU)**

Measuring the common cover on an object can be formalized as:

$$IoU = \frac{\text{aire}(\text{Box}_{\text{True}} \cap \text{Box}_{\text{Predicted}})}{\text{aire}(\text{Box}_{\text{True}} \cup \text{Box}_{\text{Predicted}})}$$



We set a threshold $\varepsilon \in [0,1]$

Prediction are considered true if $\boldsymbol{IoU \geq \varepsilon}$.

- **Precision/Recall :**

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

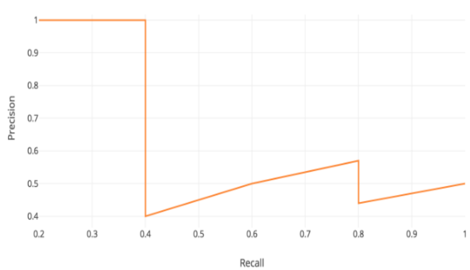
Precision answers the question: **What share of the detected objects were the right ones?**

Recall answers the question: **What is the share of the objects that have been detected?**

- **Precision-recall curve :**

Precision and recall are calculated for different confidence levels (proba of the class model) in order to obtain the curve:

$$precision = f(recall)$$



Precision-recall curve

Model Metrics

- **Mean Intersect over Union (mIoU)**

For a **fixed minimum confidence** level, the IoU is calculated for each annotation and then averaged over a class or globally.

- No indication of the type of error: on/under-prediction?
- Depends on a confidence threshold set

- Very "visual" metric, easy to interpret

- **Average Precision (AP)**

Once the IoU threshold is set, the AP of a class can be calculated as a measure of the area below the precision-recall curve.

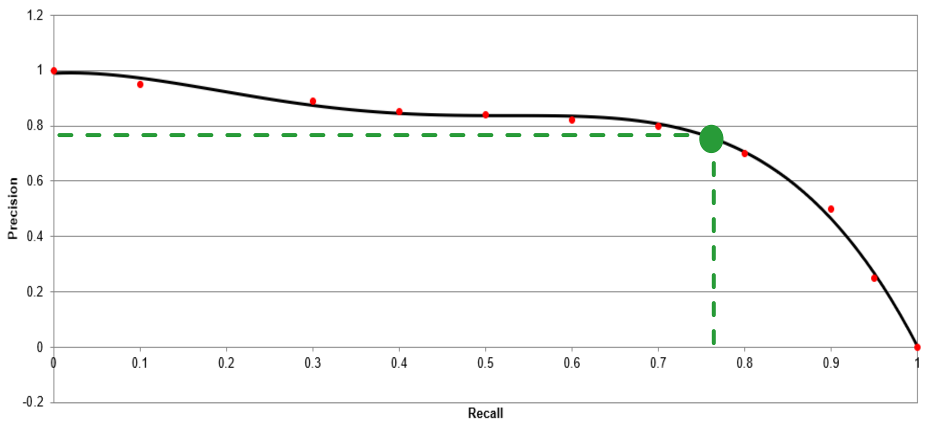
- Difficult to interpret
- Depends on a fixed IoU threshold

- Independent of confidence level
- Gives an indication of the precision/recall trade-off

➢ **The mean Average Precision (mAP)** is the **AP** averaged over all classes

The AP reflects the ability to combine precision and recall for a model

Example n°1



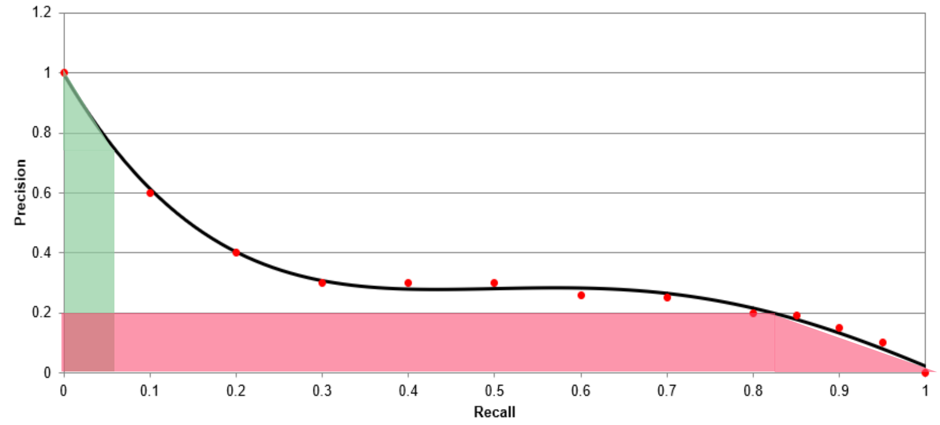
$AP \approx 0.74$

Interpretation :

- We see that we can maintain a good trade-off between the level of precision and the level of recall
- The green dot guarantees an accuracy of about 80% for a 75% recall: most objects are spotted without too many errors

➢ An important AP ensures a certain efficiency of the model that will combine precision and recall

Example n°2



$AP \approx 0.28$

Interpretation :

We see here that we cannot guarantee precision and recall: we must choose to favour one or the other

- The green zone guarantees a precision > 75% accuracy but also a recall $< 20\%$: we don't make much mistake but we spot very few objects
- The pink zone guarantees a recall > 80% but a precision < 5% : We spot almost all the object but most of the objects identified are not the one we want

➢ A low AP means it is difficult to reconcile accuracy and recall: one should be favored

Resources: all classical Machine Learning might come handy and a good understanding of Computer Vision libraries will be helpful

Computer Vision libraries

PyTorch

Standard library for deep learning

OpenCV

Standard library classical computer vision

TensorFlow

An alternative to Pytorch

ML + Viz libraries

scikit learn

To develop a wide range of ML models

pandas

To exploite model's output and aggregate them

Streamlit

To efficiently develop a dashboard / front-end