



## **ROBIN**

### **Présentation des métriques d'évaluation**

**Emmanuel D'Angelo (DGA)**

**Stéphane Herbin (ONERA)**

**Matthieu Ratiéville (DGA)**

# Plan

## ➤ Rappels

- Formalisme
- Objectifs

## ➤ Détection

- Définition des métriques
- Gestion des cas multiples

## ➤ Catégorisation

- Définition des métriques
- Mécanismes de rejet

# Recherche d'objets dans une image

Image à interpréter

$$\mathbf{B}, I \rightarrow \{Y_i, Z_i\}_{i=1}^{N(I)}$$

Ensemble d'images  
décrivant le ou les objets à  
rechercher

= *Base d'apprentissage*

Liste de  $N(I)$  régions  
 $Z_i$  associées à une  
catégorie  $Y_i$

La base  $\mathbf{B}$  contient une « vérité terrain » :  $I \rightarrow \{Z_k^*, Y_k^*\}_k$

# Deux fonctions

## Détection

- « Marginale »
- Contrôle de la fonction:  $\lambda$
- Associe une liste de régions à une image:

$$\mathbf{B}, I, \lambda \rightarrow \{Z_1, Z_2, \dots, Z_{N(I,\lambda)}\}$$

## Catégorisation

- « Conditionnelle »
- Contrôle de la fonction:  $\mu$
- Associe une étiquette à une région dans une image:

$$\mathbf{B}, I, Z, \mu \rightarrow Y$$

# Évaluation

- **Métrique = mesure d'écart à une exigence**
- **Exigences fonctionnelles**
  - Deux fonctions: détection et catégorisation
  - Mesures quantitatives sur bases d'apprentissage et de test
- **Exigences opérationnelles**
  - Temps de calcul, robustesse, flexibilité, interactivité...
  - Auto-évaluation qualitative (questionnaire)

# Points de fonctionnement

## ➤ **Problématique**

- Contrôle d'un traitement pour respecter au mieux un ensemble d'exigences

## ➤ **Forme du contrôle**

- Paramètres (modèle, poids, hyperplan...)
- Seuil de décision

## ➤ **Évaluation**

- Un point de fonctionnement précis (= une exigence)
- Un ensemble de points (= une variété d'exigences)

# Métriques d'évaluation de la détection - I

## Par image $I$ et contrôle $\lambda$ ...

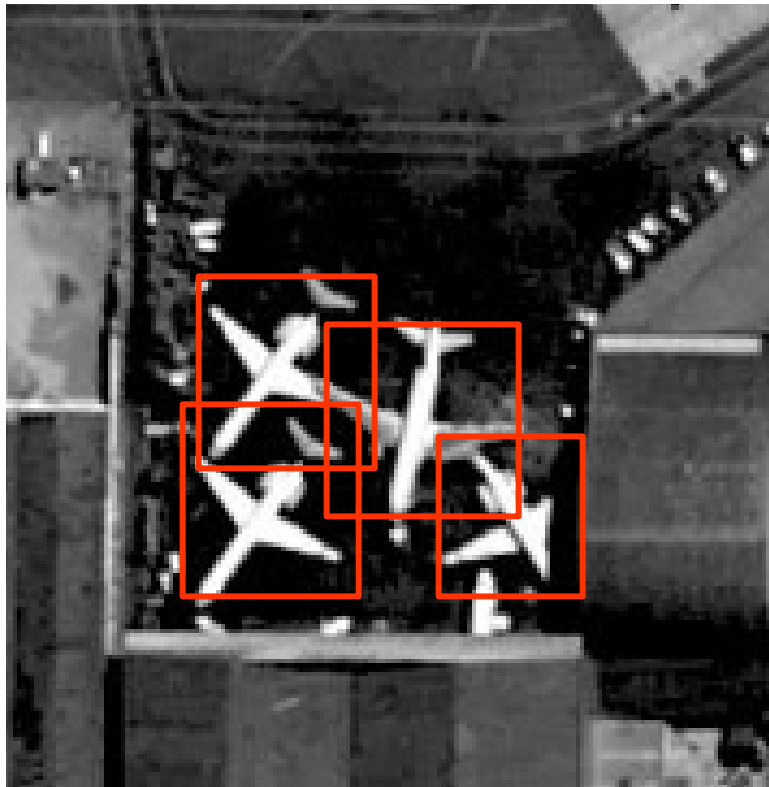
### ➤ Deux listes

- Vérité terrain :  $\{Z^*_1, Z^*_2, \dots, Z^*_{N^*(I)}\}$
- Sortie de l'algorithme :  $\{Z_1, Z_2, \dots, Z_{N(I,\lambda)}\}$

### ➤ Trois nombres

- $TP(I, \lambda)$  : Nombre de bonnes détections
- $N(I, \lambda)$  : Nombre total de détections
- $N^*(I)$  : Nombre d'objets à détecter

# Comptage des bonnes détections



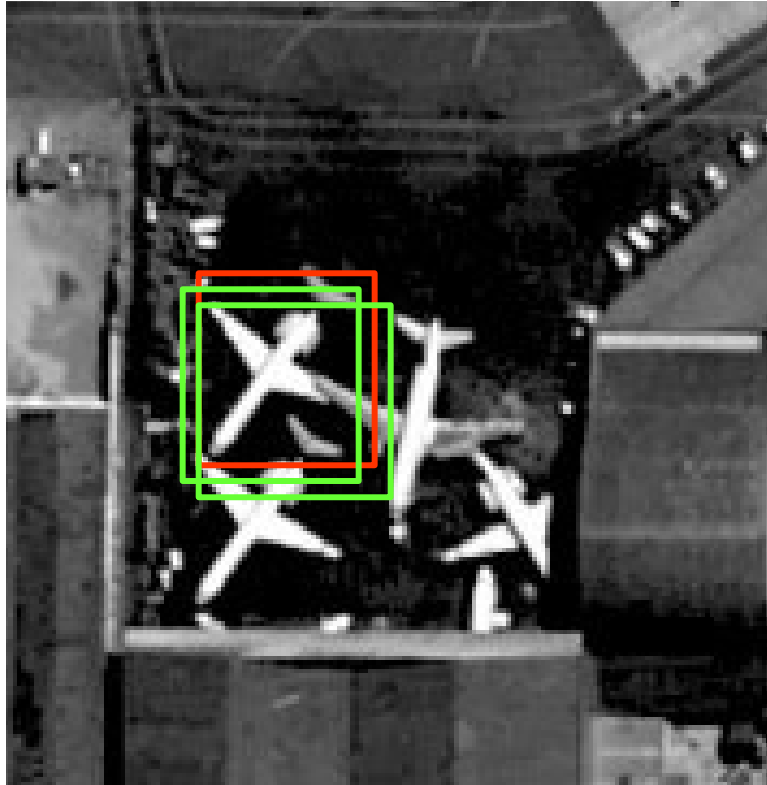
Vérité Terrain



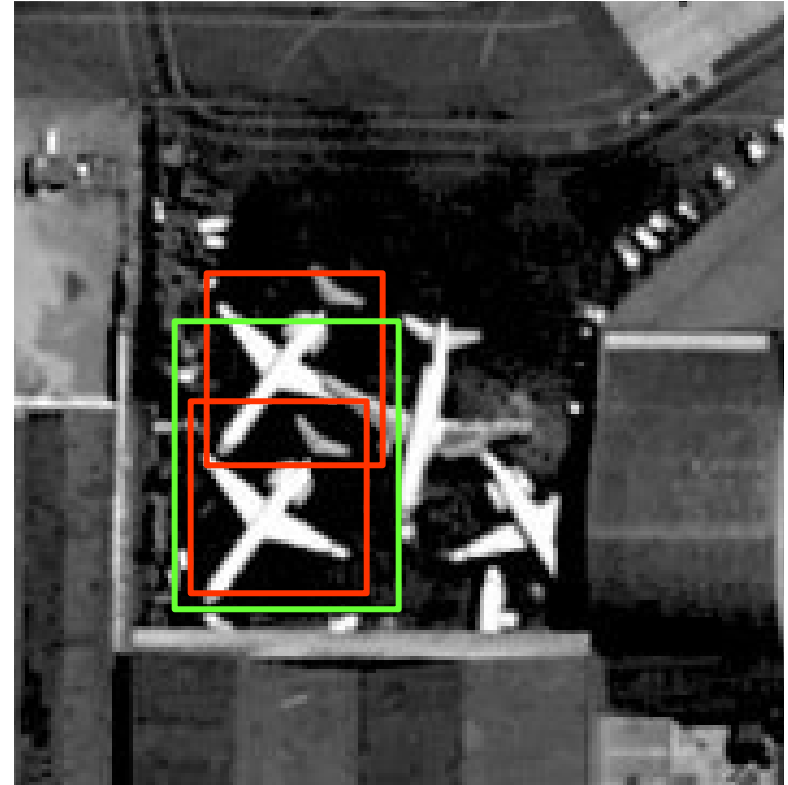
Détections



# Gestion des cas multiples



*Détections multiples*



*Appariements multiples*

# Définition d'une bonne détection

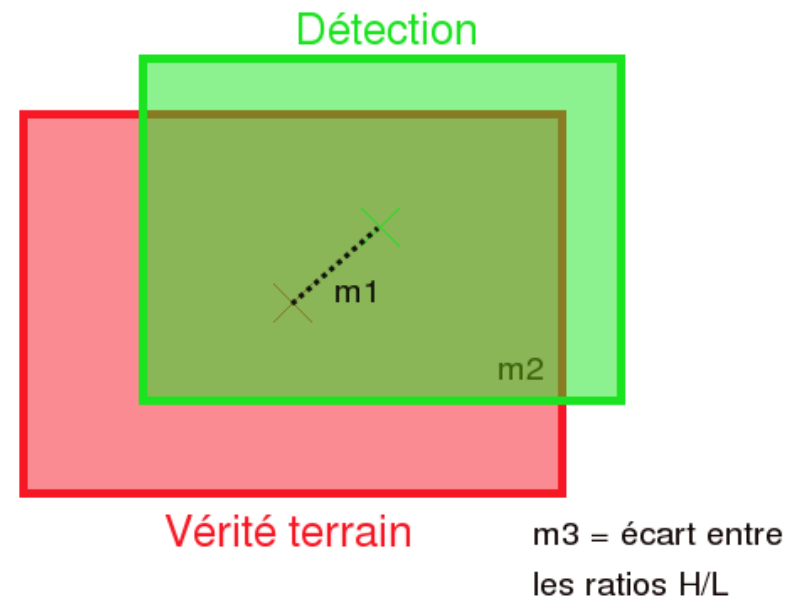
➤ Comparaison géométrique de  $Z$  et  $Z^*$

➤  $Z$  et  $Z^*$  = boîtes englobantes

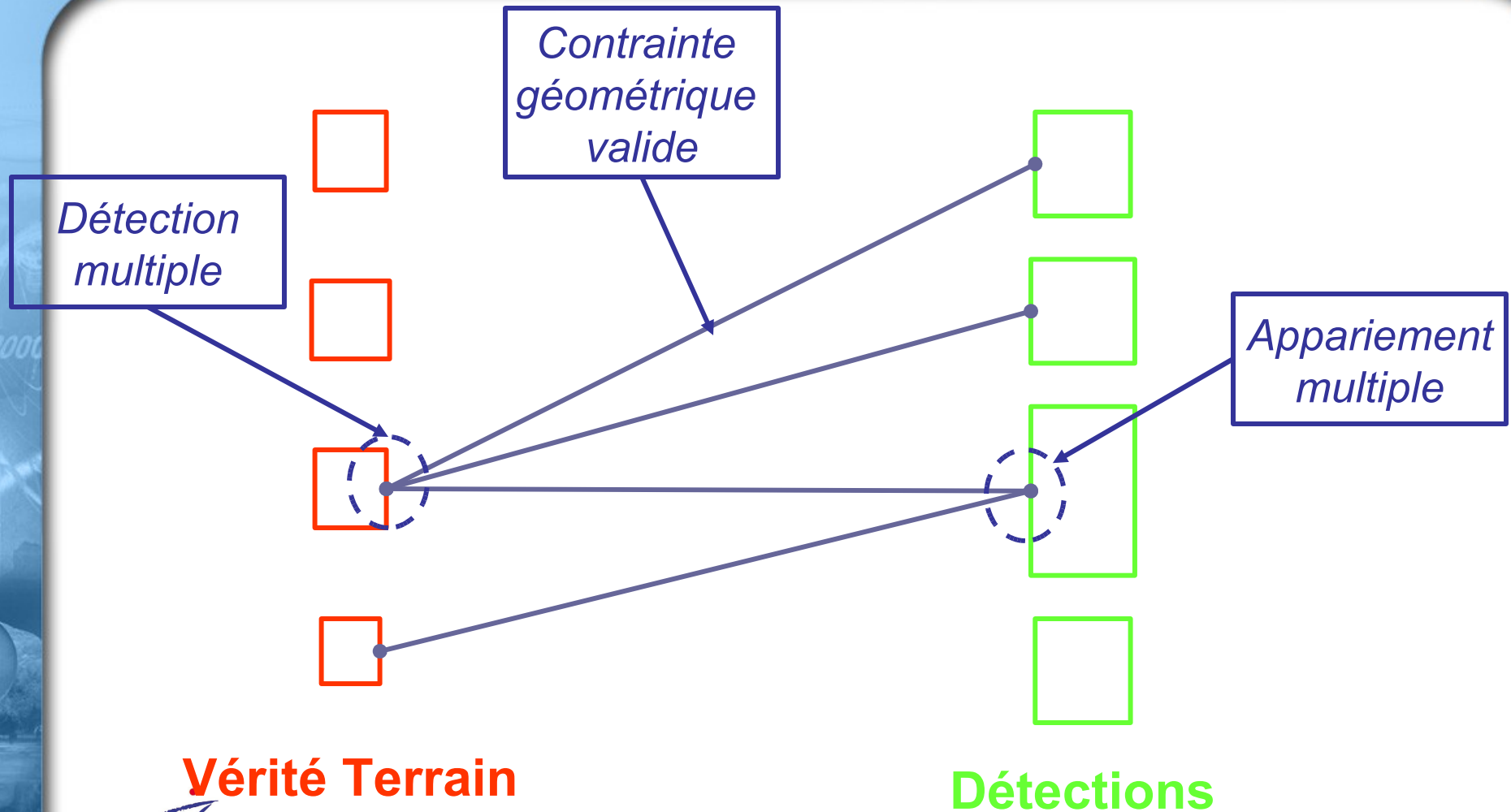
1. **Localisation** (comparaison des centres des boîtes)
2. **Complétude** (comparaison des surfaces communes)
3. **Forme** (comparaison des rapports hauteur/largeur)

➤  $Z$  = un point,  $Z^*$  = une boîte englobante

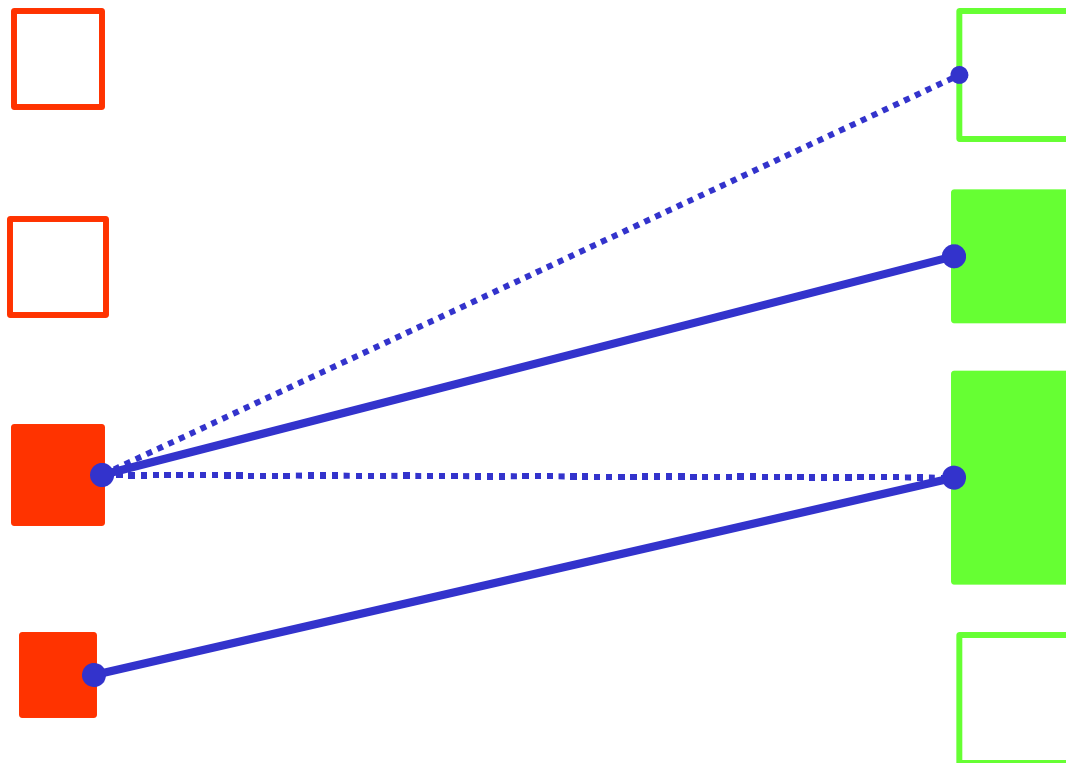
- **Localisation** (comparaison du point et du centre de la boîte)



# Bonne détection = bon appariement



# Couplage maximal N°1



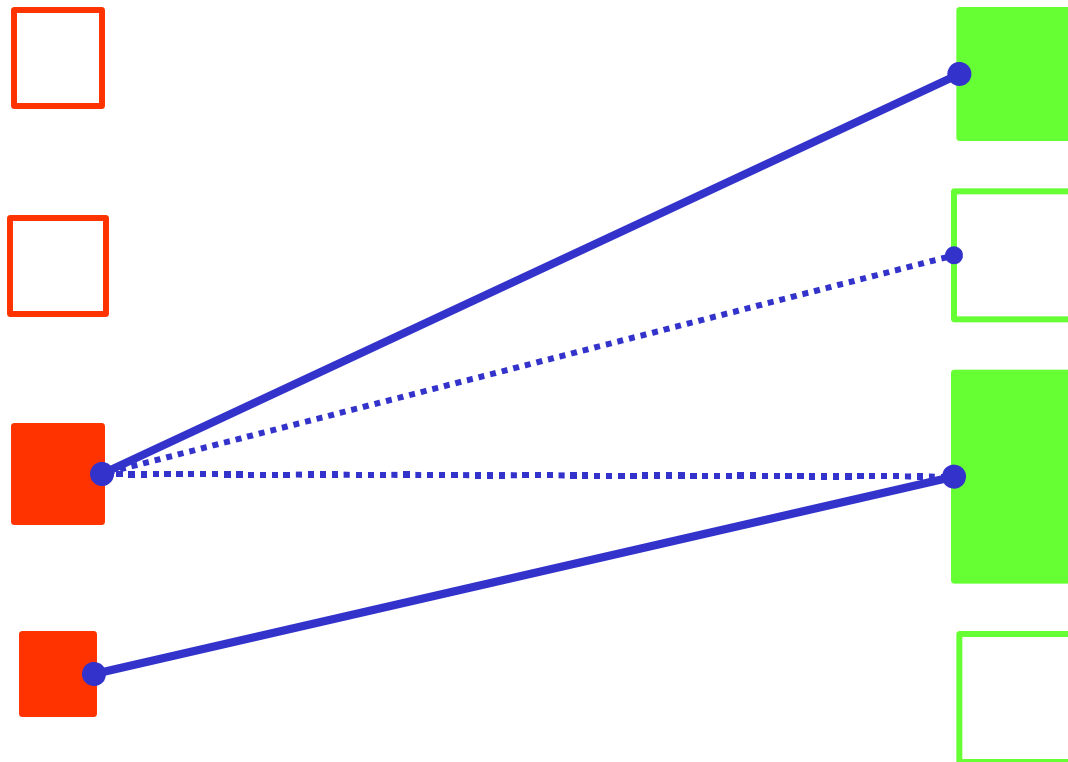
$$TP = 2$$

$$N = 4$$

$$N^* = 4$$

$TP$  = degré du couplage maximal

# Couplage maximal N°2



$$TP = 2$$

$$N = 4$$

$$N^* = 4$$

$TP$  = degré du couplage maximal

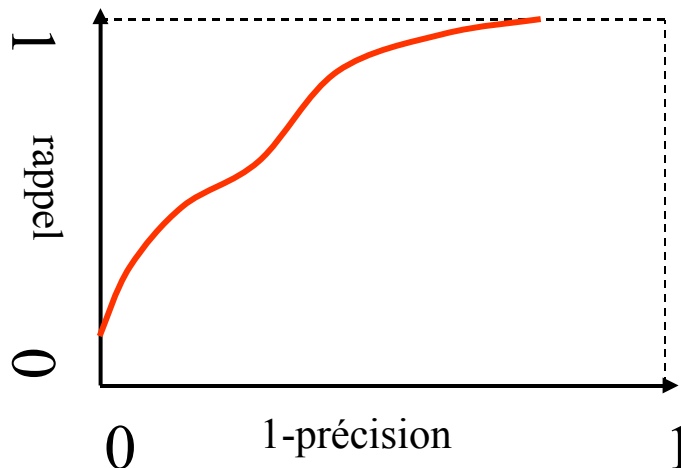
# Métriques d'évaluation de la détection - II

- Deux grandeurs paramétriques moyennes

$$\text{Rappel}(\lambda) = \frac{\sum_I TP(I, \lambda)}{\sum_I N^*(I)}$$

$$\text{Precision}(\lambda) = \frac{\sum_I TP(I, \lambda)}{\sum_I N(I, \lambda)}$$

- Un ensemble de points Rappel / Précision



*Un point  
=  
Une valeur de contrôle*

# Remarques sur diagramme Rappel / Précision

- Propriétés globales
  - Ni convexe, ni concave
  - Non monotone
- Valeurs extrémales
  - En général,  $R_{\min} \geq 0$ , mais  $R_{\max} < 1$
  - $P_{\min} \geq 0$ ,  $P_{\max} \leq 1$

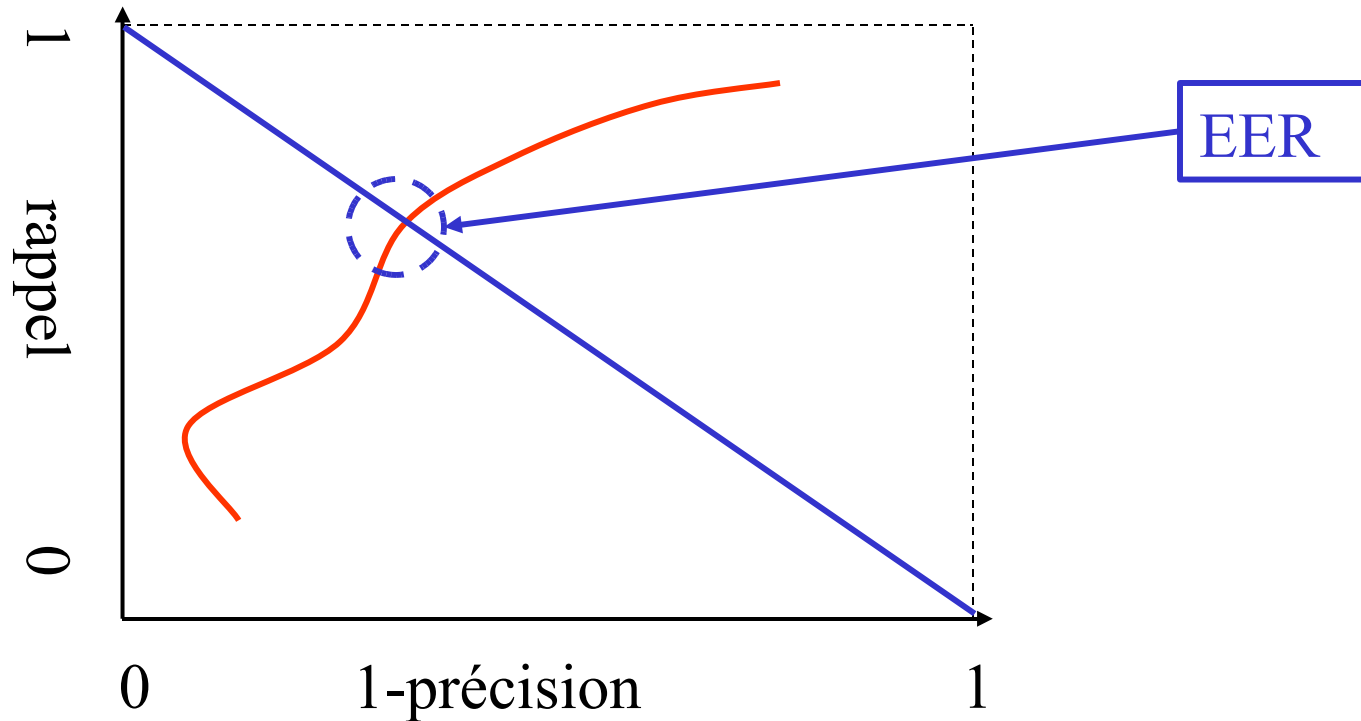
# Détection par seuillage d'indice de confiance

- Les régions  $Z_i$  peuvent être accompagnées d'un indice de confiance  $\alpha_i$  a priori
  - l'algorithme ne fournit pas de décision complète
  - Le contrôle  $\lambda$  est un seuil de décision ( $\alpha_i > \lambda$ )
- Génération des Précision/Rappel par variation du seuil
  - Un point de la courbe = Une valeur de seuil



# Point de fonctionnement : EER

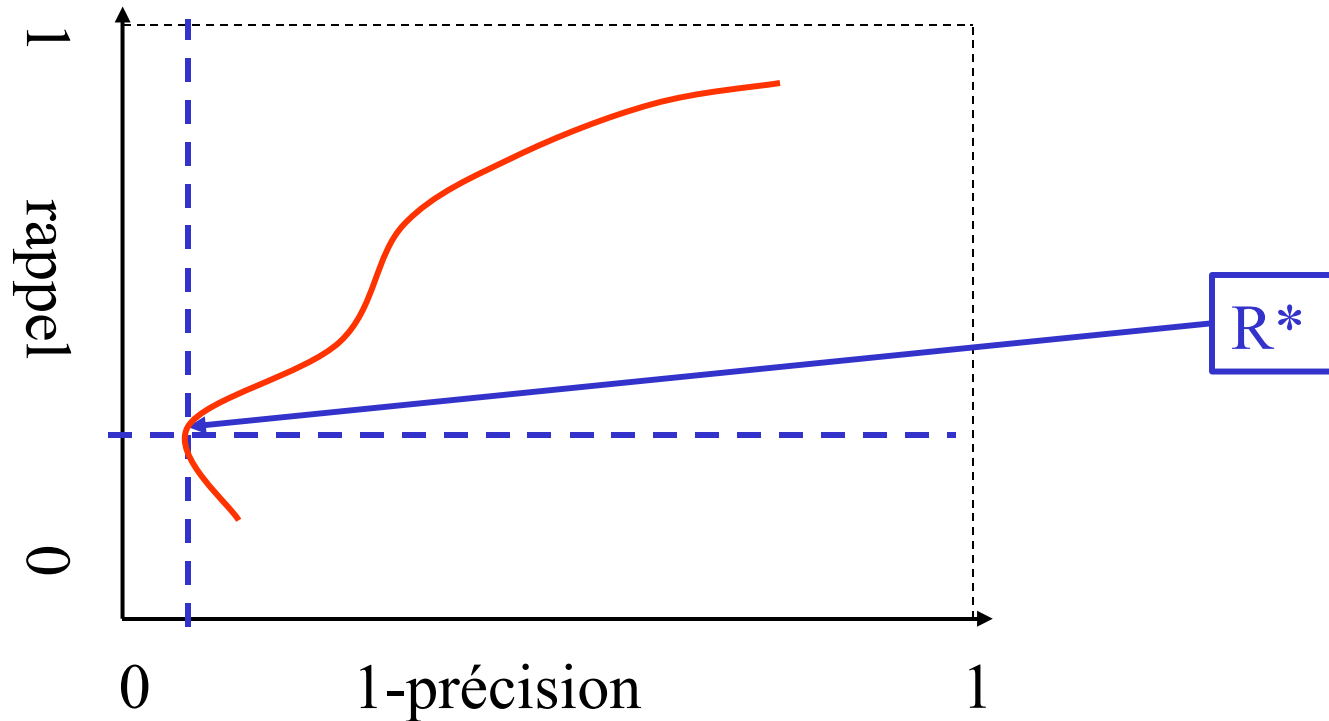
*Rappel à égale précision*



« *Bon compromis* »

# Point de fonctionnement : $R^*$

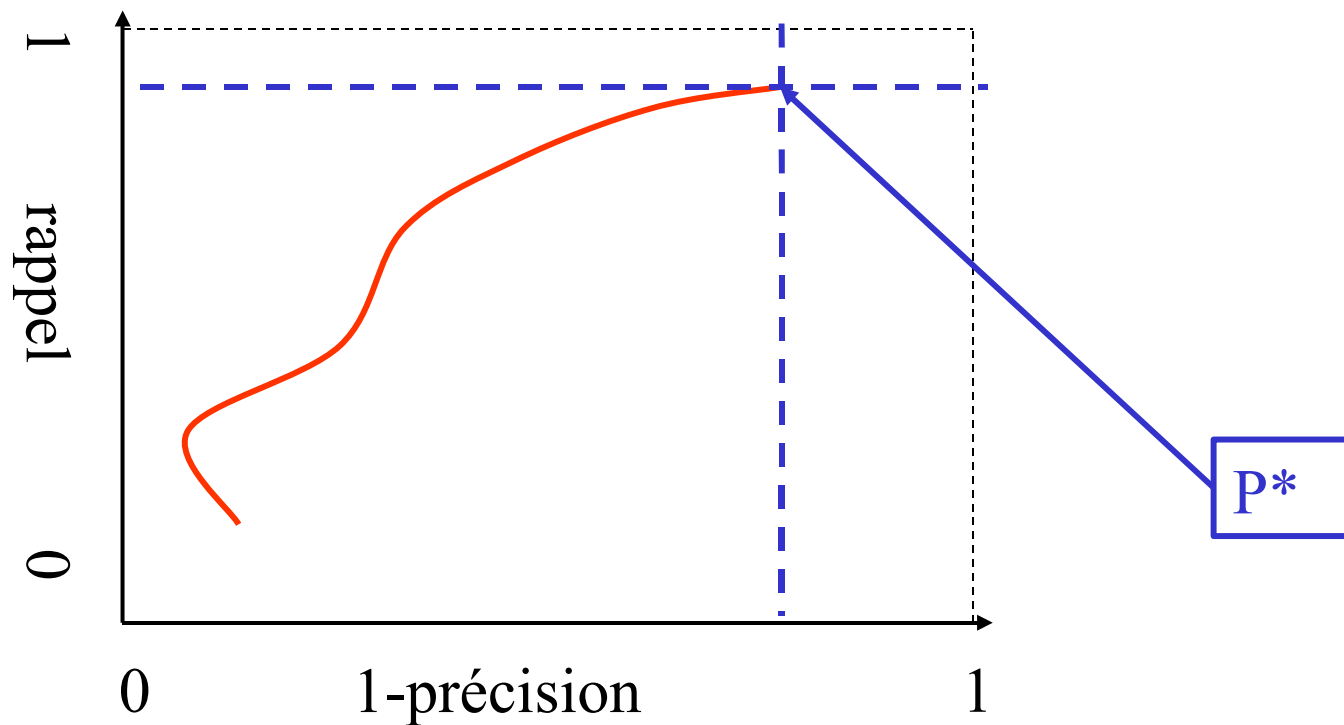
*Rappel à précision maximale*



« *Pas de mauvaise décision* »

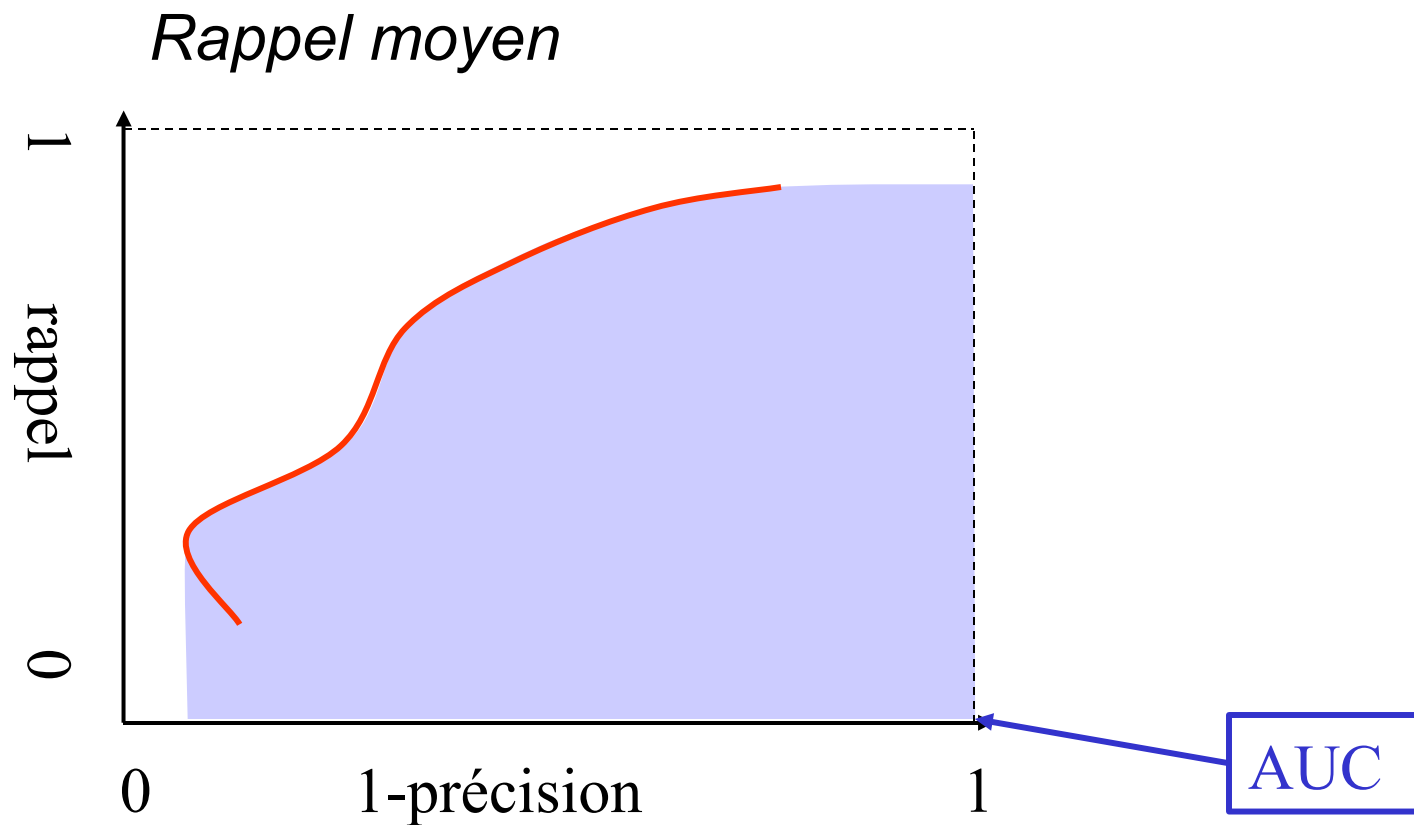
# Point de fonctionnement : $P^*$

*Précision à rappel maximal*



« *Aucun oubli* »

# Points de fonctionnement : AUC ou AMP



« *Comportement moyen* »

# Métriques d'évaluation de la détection - II

- Rappel à précision maximale :  $R^*$
- Précision à rappel maximal :  $P^*$
- Rappel à égale précision :  $EER$
- Rappel moyen :  $AUC$

# Métriques d'évaluation de la catégorisation - I

- Vérité terrain :  $Y_n^*$  associée à catégorie :  $c^*$
- Sortie de l'algorithme :  $Y_n$ 
  - Décision de catégorie :  $c$
  - Décision d'ambiguïté : « A »
  - Décision d'altérité : « O »
- Outil de base : matrices de confusion
  - Comparaison de la valeur des couples  $\{Y_n, Y_n^*\}$
  - Paramètre de contrôle  $\mu$
  - Nombre d'éléments par catégories :  $N(c^*)$  dans base  $\mathbf{T}$

$$\eta(c, c^*, \mu) = \frac{1}{N(c^*)} \sum_{n \in \mathbf{T}} \mathbf{1}_c(Y_n) \cdot \mathbf{1}_{c^*}(Y_n^*)$$

# Métriques d'évaluation de la catégorisation - II

➤ **Taux moyen de bonne discrimination**

$$D(\boldsymbol{\mu}) = \sum_{c^*} \pi(c^*) \eta(c^*, c^*, \boldsymbol{\mu})$$

➤ **Taux moyen d'incertitude**

$$U(\boldsymbol{\mu}) = \sum_{c^*} \pi(c^*) \eta(A, c^*, \boldsymbol{\mu})$$

➤ **Taux de rejet**

$$R(\boldsymbol{\mu}) = \eta(O, O, \boldsymbol{\mu})$$

# Métriques d'évaluation de la catégorisation - III

- Matrice de confusion à taux d'incertitude maximal :  $\eta^*(c, c^*)$
- Discrimination pour taux d'incertitude minimal :  $D^*$
- Incertitude pour taux de discrimination maximal :  $U^*$
- Discrimination à égale incertitude :  $EDU$
- Rejet à égale discrimination :  $ERR$



# Catégorisation : cas particuliers

- **Seuillage d'indice de confiance**
  - Sortie = catégorie associée à un indice de confiance
  - Seuil sur indice = décision de rejet
  - Pas de décision « Autre » possible
- **Catégorisation comme détections multiples**
  - Pas de gestion d'un problème multi-classe
  - Définition des cas négatifs ?
  - Ne sera pas évalué comme catégorisation

# ***DISCUSSION***