



**HAL**  
open science

## Formats d'annotation, détecter dans les images

Chahan Vidal-Gorène

► **To cite this version:**

Chahan Vidal-Gorène. Formats d'annotation, détecter dans les images. École d'été de DISTAM, 2024.  
hal-04672675

**HAL Id: hal-04672675**

**<https://enc.hal.science/hal-04672675v1>**

Submitted on 19 Aug 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



# Formats d'annotation, détecter dans les images

**Chahan Vidal-Gorène**  
École nationale des chartes-PSL  
Calfa



- Découvrir les différentes approches en **Vision par ordinateur** et identifier ses besoins
- Se familiariser avec les **formats d'annotation** adaptés à la tâche
- Prendre en main deux outils *opensource* **LabelStudio** et **YOLO** pour l'annotation générique et la détection d'objets



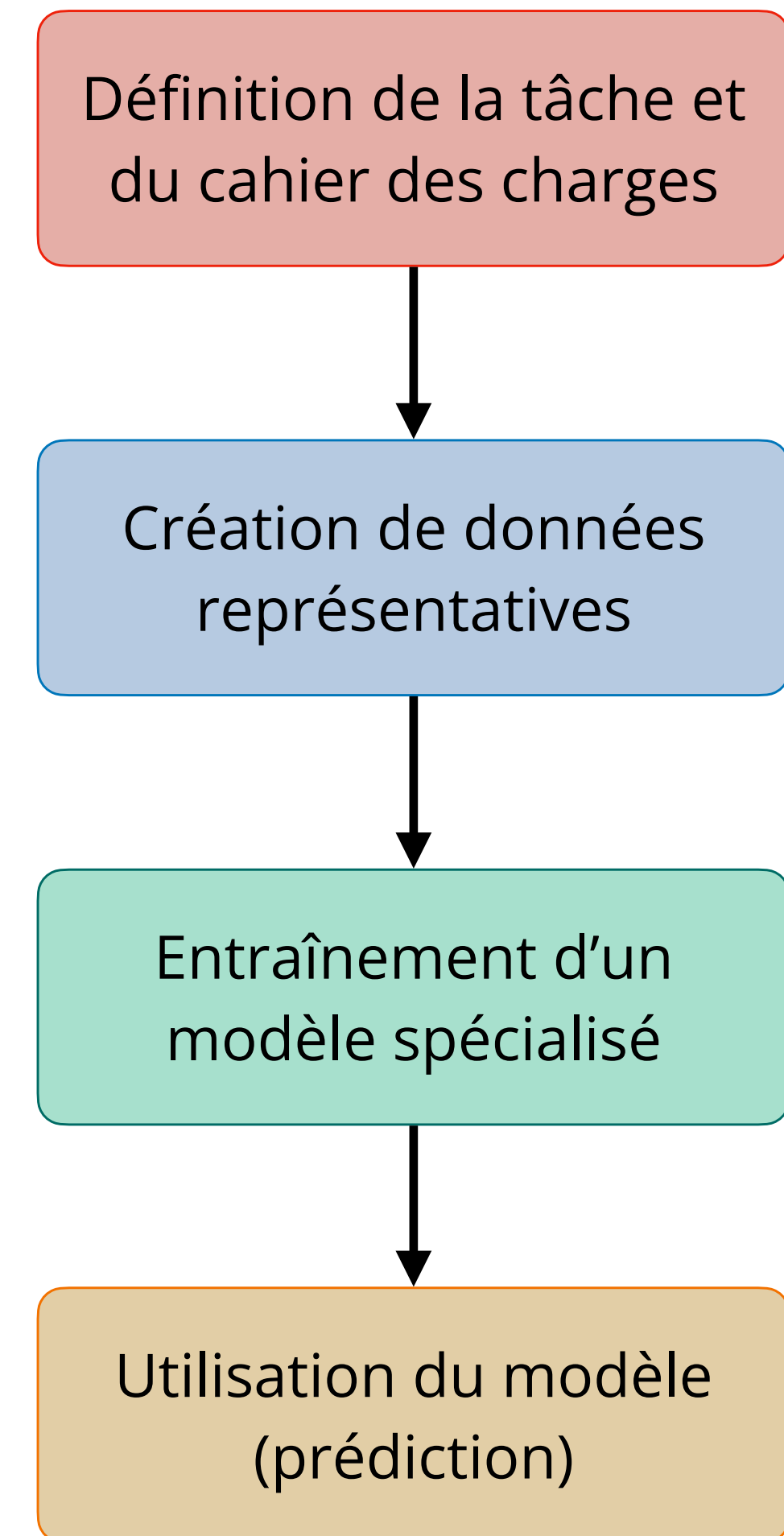
**Données** : deux datasets sont proposés à titre d'exemple : manuscrits enluminés, pommes de terre. Datasets libres autorisés.

Quelques idées, non exhaustives :

- Dégrossir l'analyse d'un (gros, très gros) corpus d'images ou de vidéos en **identifiant** les vues pertinentes ou **quantifiant** la présence d'informations
- Proposer une **version augmentée** d'un corpus en enrichissant la navigation (exemple IIF)
- **Automatiser** une tâche

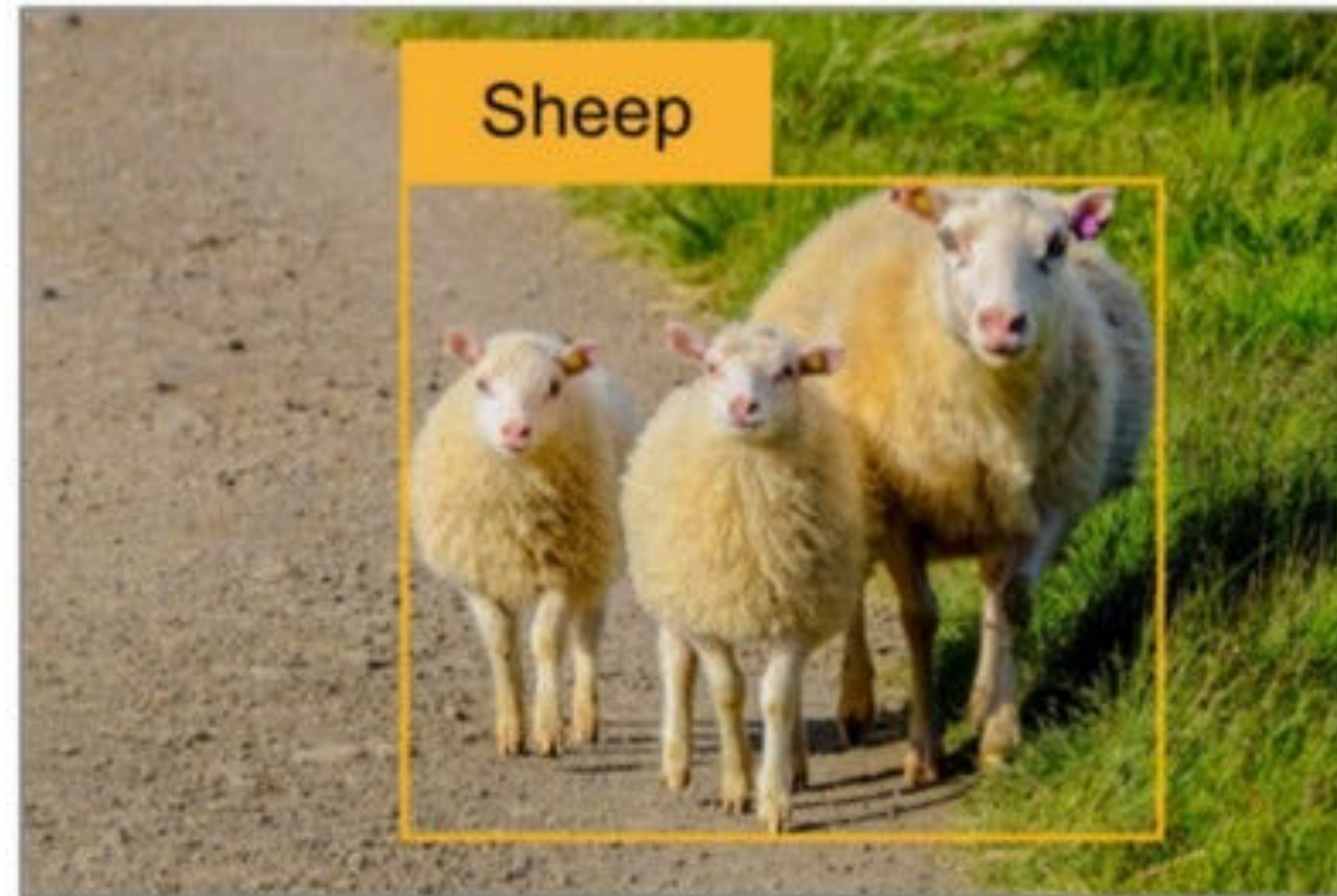
L'entraînement et l'utilisation de modèles est de plus en plus simple aujourd'hui. L'enjeu est l'**identification de ses besoins**, la **définition d'un cahier des charges** réaliste, et la **création des données** (annotation).

C'est avant tout une réflexion thématique et non technique.

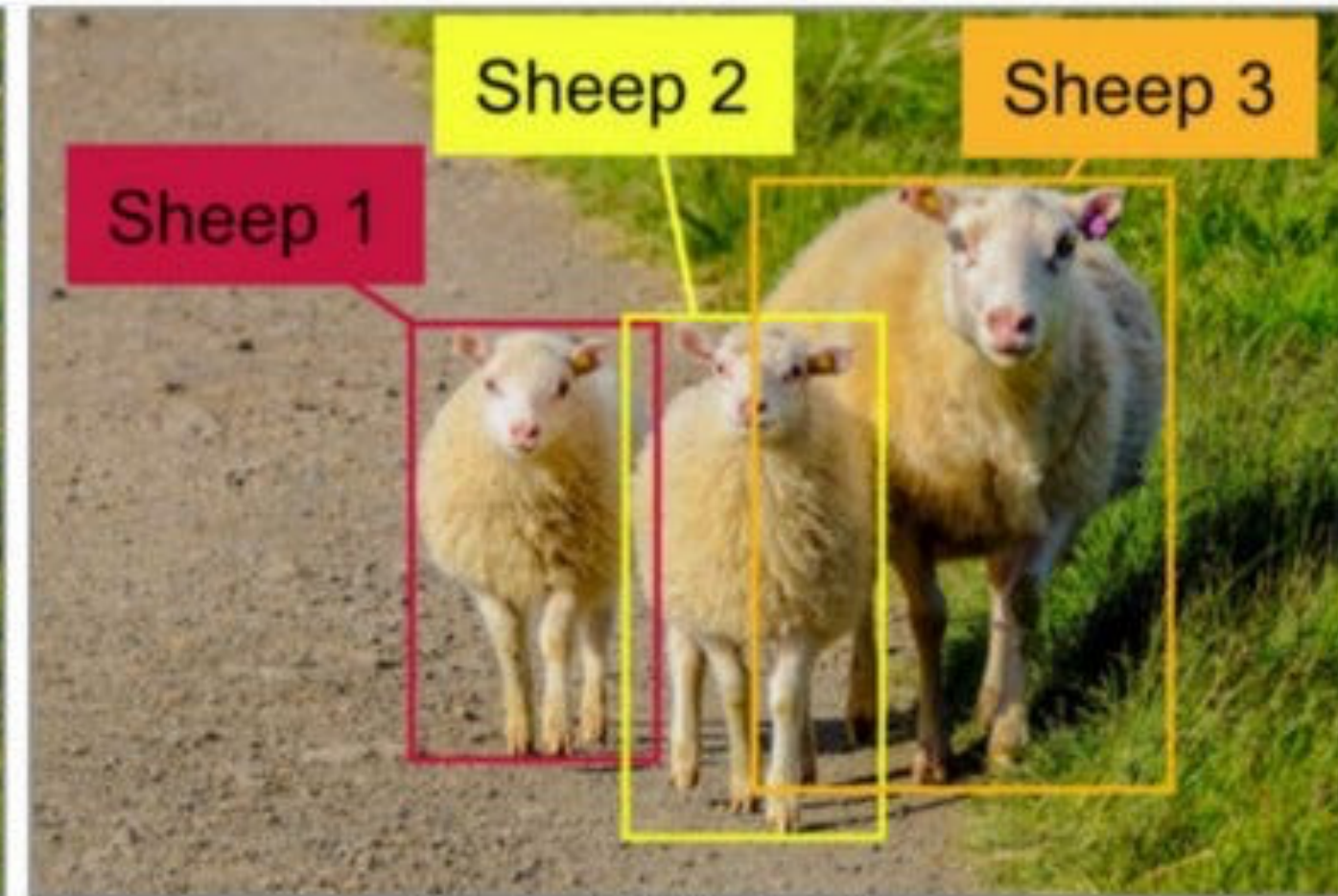




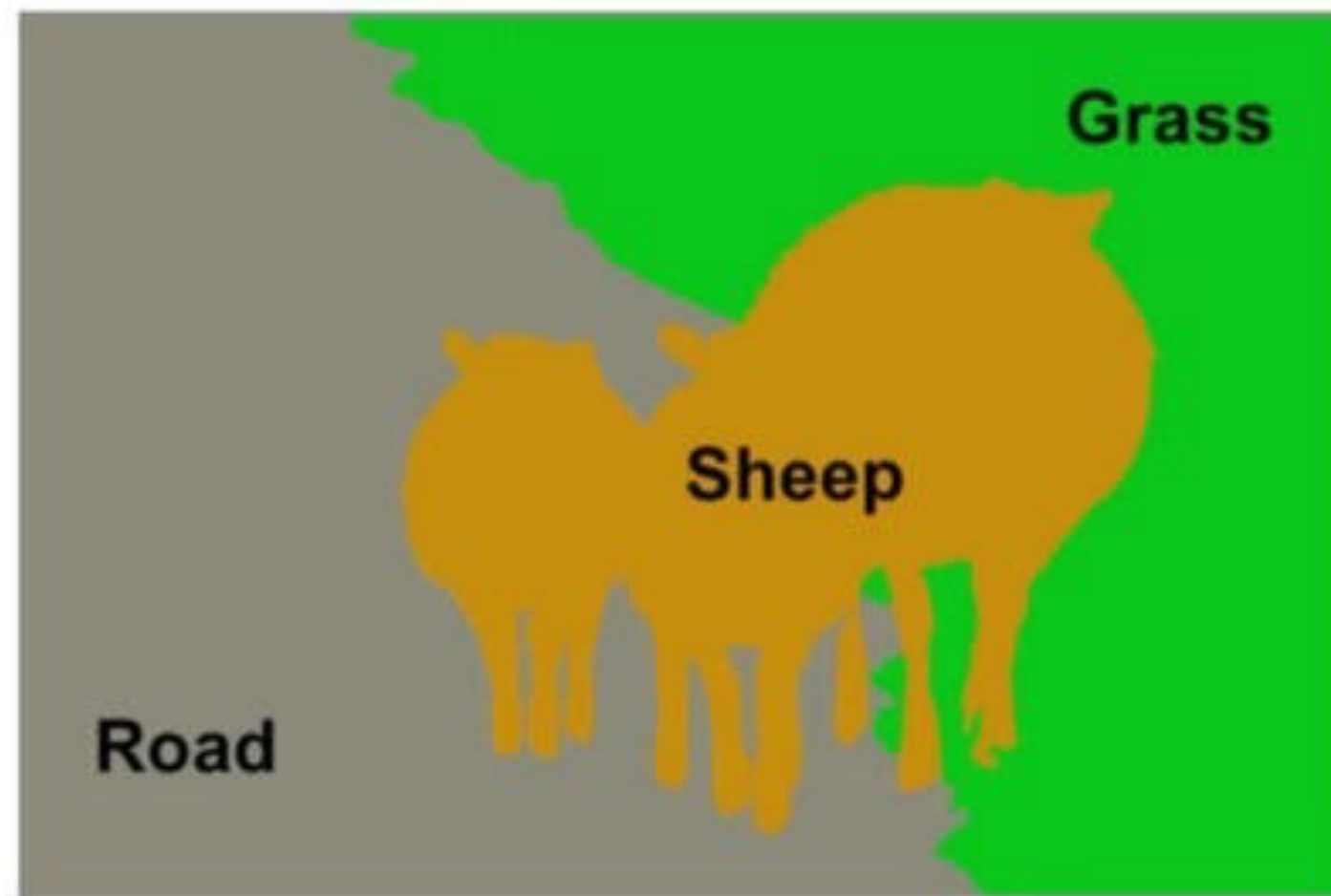
Source image: <https://nirmalamurali.medium.com/image-classification-vs-instance-segmentation-vs-semantic-segmentation-n-625c33a08d50>



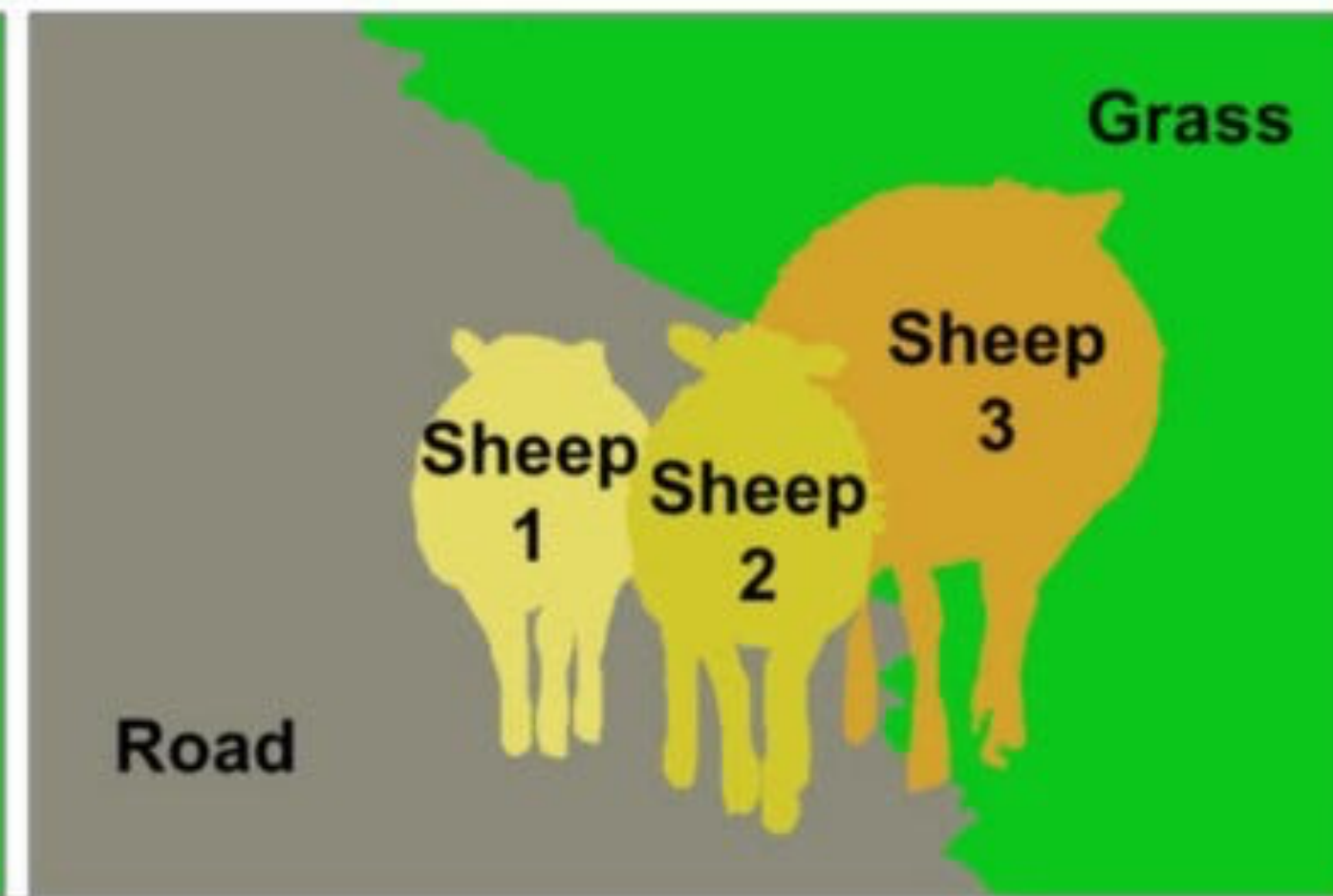
**Classification + Localization**



**Object Detection**



**Semantic Segmentation**



**Instance Segmentation**

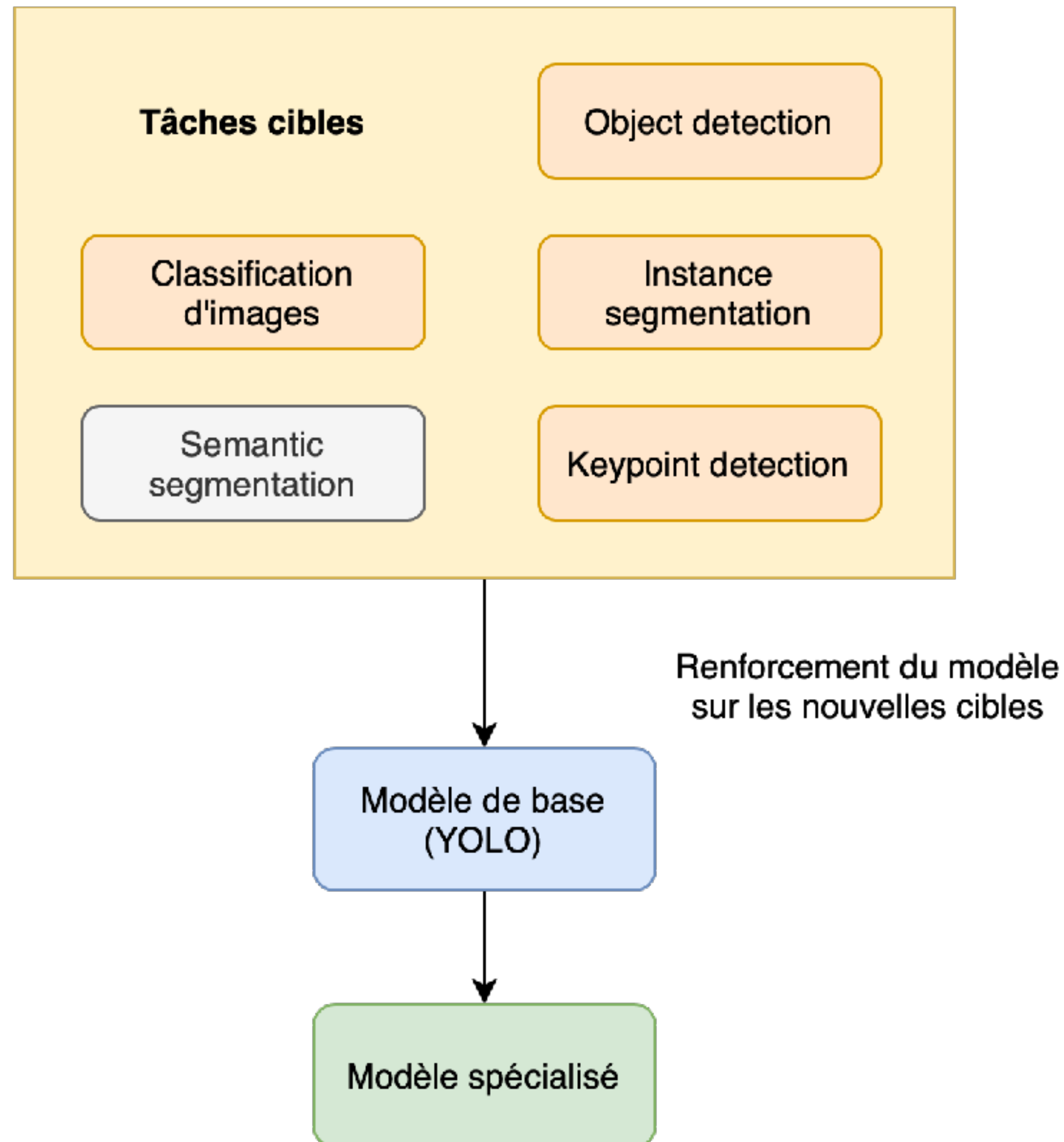
4 types d'approches + 1

Utilisation de la librairie YOLO, très facile d'installation et d'utilisation, adaptée au temps réel.

Installation et utilisation possible sur ordinateurs personnels et / ou serveurs (nécessite `python` et des environnements)



## Création de données avec LabelStudio

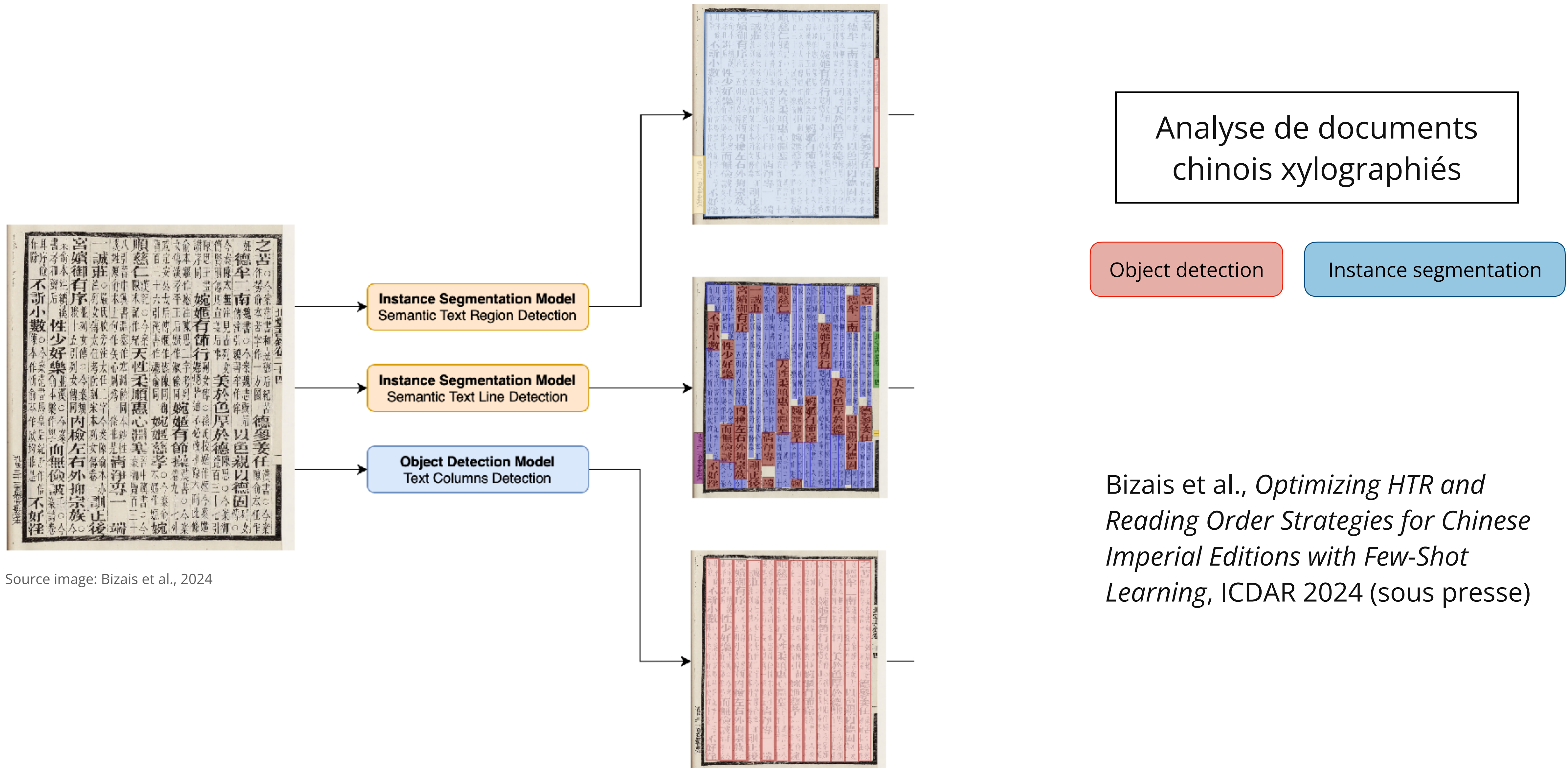


**Stratégie d'entraînement** : aujourd'hui, on ne part **jamais** de zéro (*from scratch*), on part **toujours** d'un modèle existant qu'on va spécialiser sur notre tâche (*fine-tuning*), y compris si le modèle de base n'est pas coutumier de notre tâche (stratégie de *transfer-learning*)

## Utilisations classiques :

- **classification** : catégorisation d'informations, similarité
- **object det.** : vidéo surveillance, Transkribus, dénombrement d'objets
- **semantic seg.** : imagerie médicale, eScriptorium, identification de zones et de clusters
- **instance seg.** : imagerie médicale, voitures autonomes, robotique, Calfa, dénombrement et séparation
- **keypoint det.** : vidéo surveillance





Analyse de documents chinois xylographiés

Object detection

Instance segmentation

Bizais et al., *Optimizing HTR and Reading Order Strategies for Chinese Imperial Editions with Few-Shot Learning*, ICDAR 2024 (sous presse)

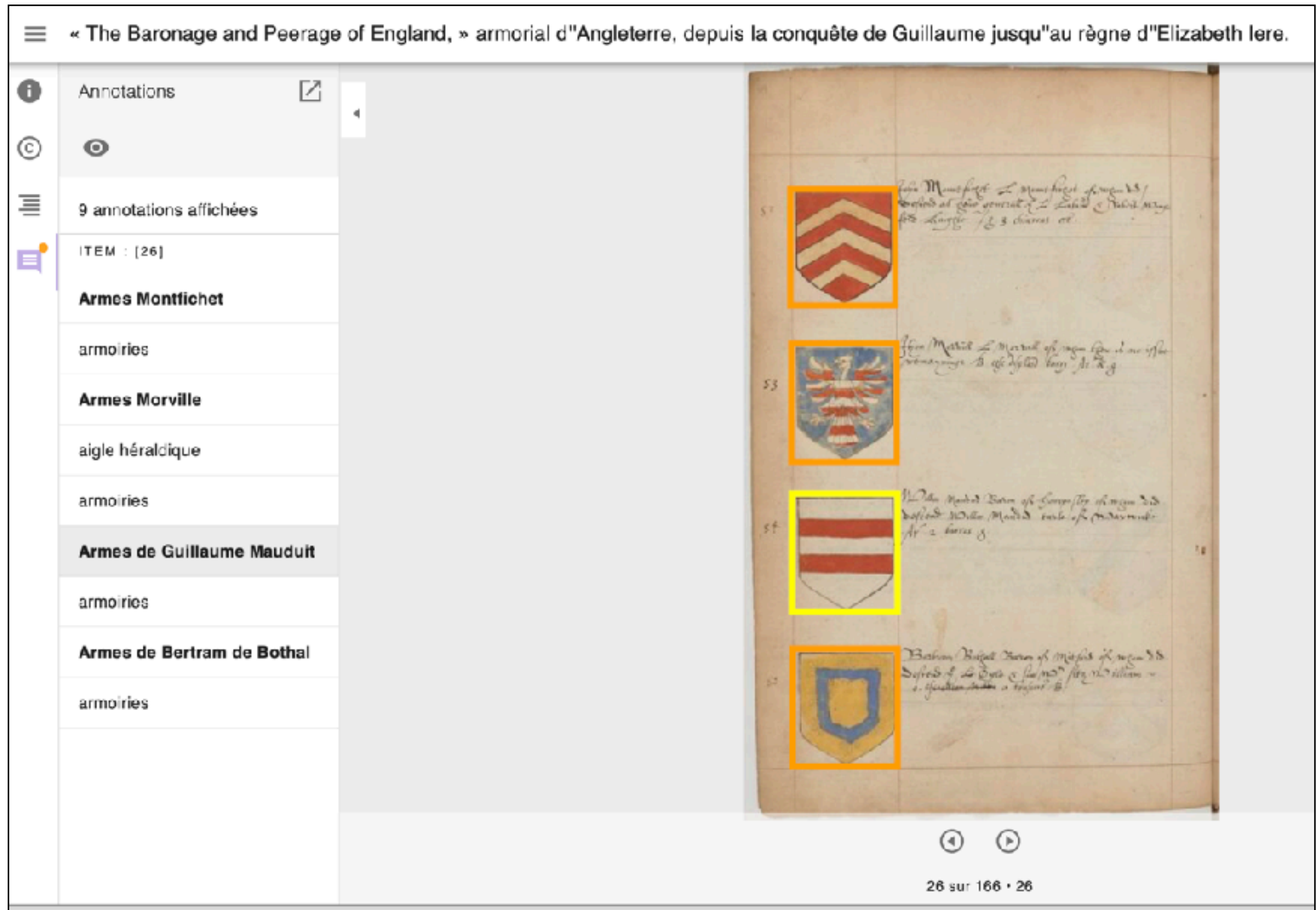
Source image: Bizais et al., 2024



Détection et identification de motifs dans une bibliothèque numérique IIF

Object detection

Classification



Source image: <https://mandragore.bnf.fr/>



Source image: <https://staatsbibliothek-berlin.de/>, Califa

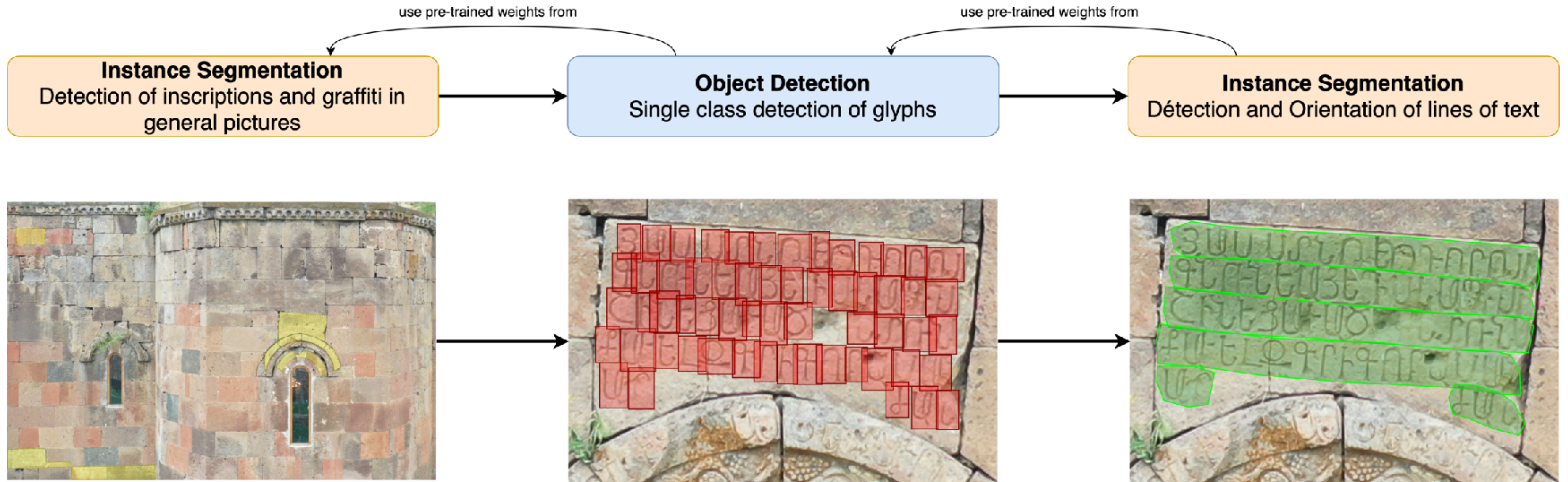


## Analyse d'inscriptions médiévales arméniennes

Object detection

Instance segmentation

Vidal-Gorène et al., *Detecting and Deciphering Damaged Medieval Armenian Inscriptions using YOLO and Vision Transformers*, ICDAR 2024 (sous presse)



Source image: Vidal-Gorène et al., 2024



Motifs iconographiques et  
ornementaux sur des  
édifices religieux

Object detection

Instance segmentation

**Projet MoNumEd** en cours,  
Sipana Tchakerian (INHA) et Chahan  
Vidal-Gorène (Enc, Calfa)



Source image: Projet MoNumEd, Wikimedia Commons, 2024



## Analyse de papyrus endommagés

Object detection

Classification

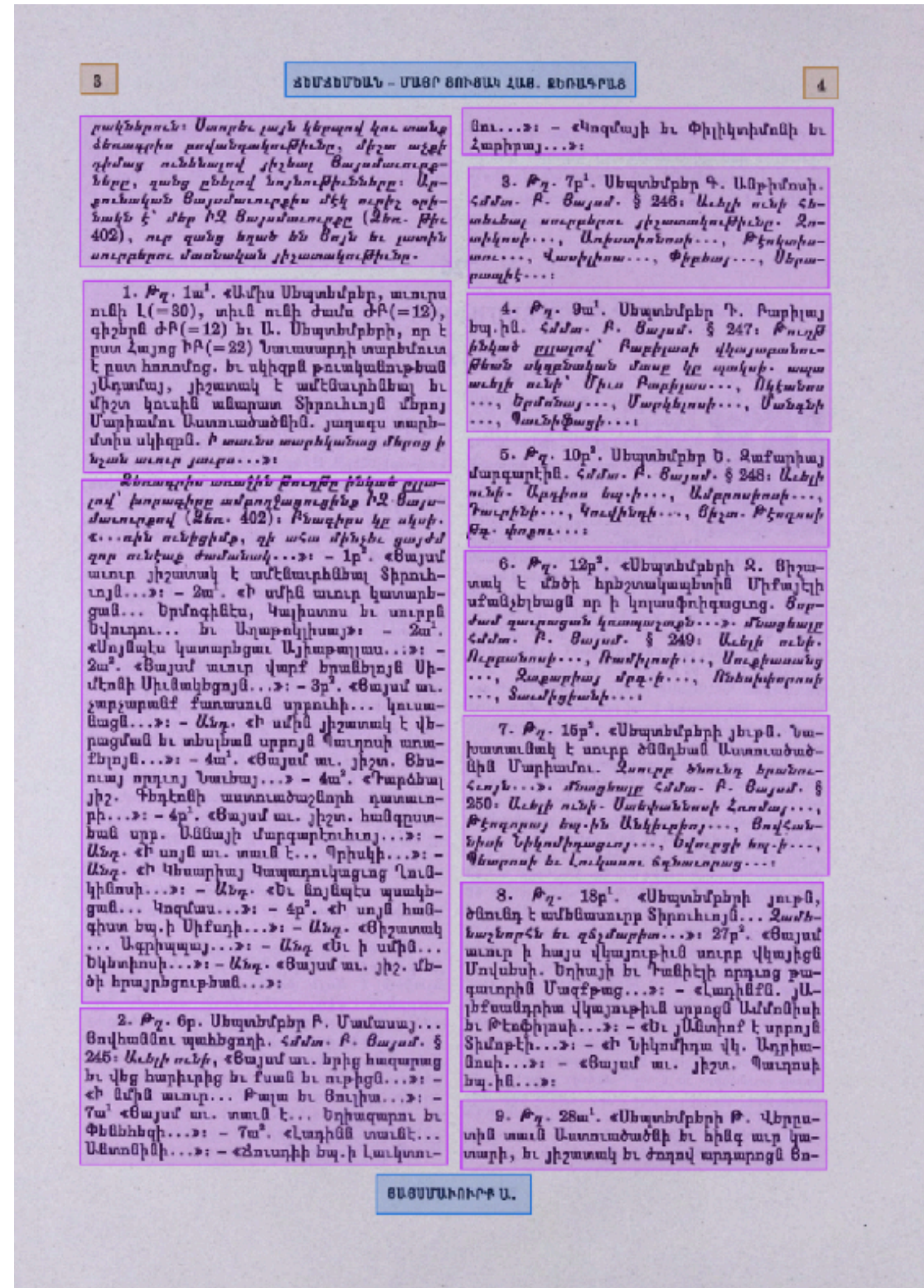
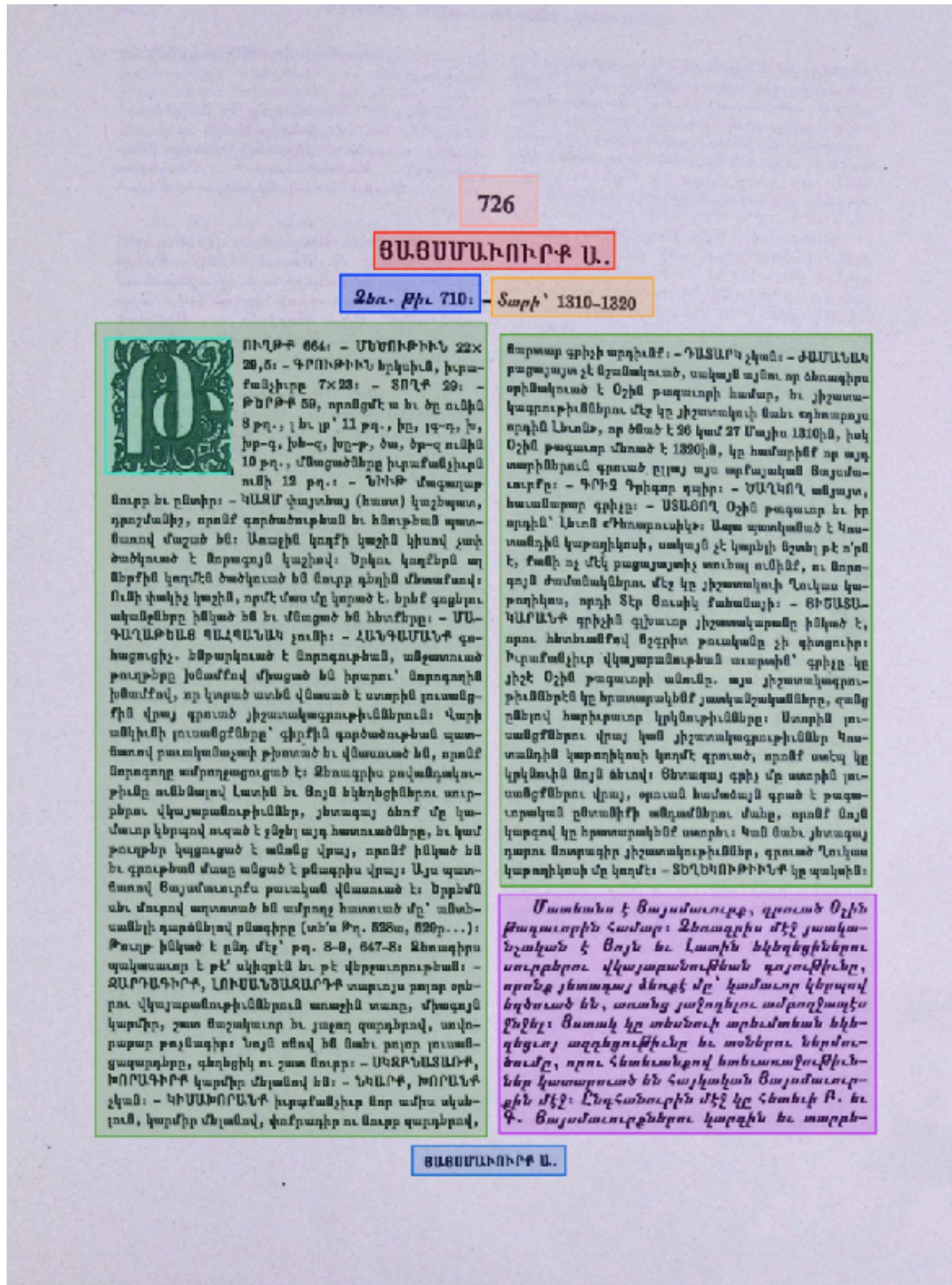
Résultats **ENCyclops** (Vidal-Gorène et al., Master Humanités Numériques, dans Seuret et al. "ICDAR 2023 Competition on Detection and Recognition of Greek Letters on Papyri." International Conference on Document Analysis and Recognition. Cham: Springer Nature Switzerland, 2023.



Source image: ENCyclops, 2023

Source image: Projet MoNumEd, Wikimedia Commons, 2024





Détection de la mise en page de catalogue

Object detection

Calfa, Le catalogue des Mékhitaristes de Venise désormais accessible en ligne, 2024, <https://calfa.fr/blog/46>

Source image: Calfa, 2024



Pour entraîner un modèle YOLO (ou autre) de vision par ordinateur, il faut créer de la donnée.

 **Label Studio**



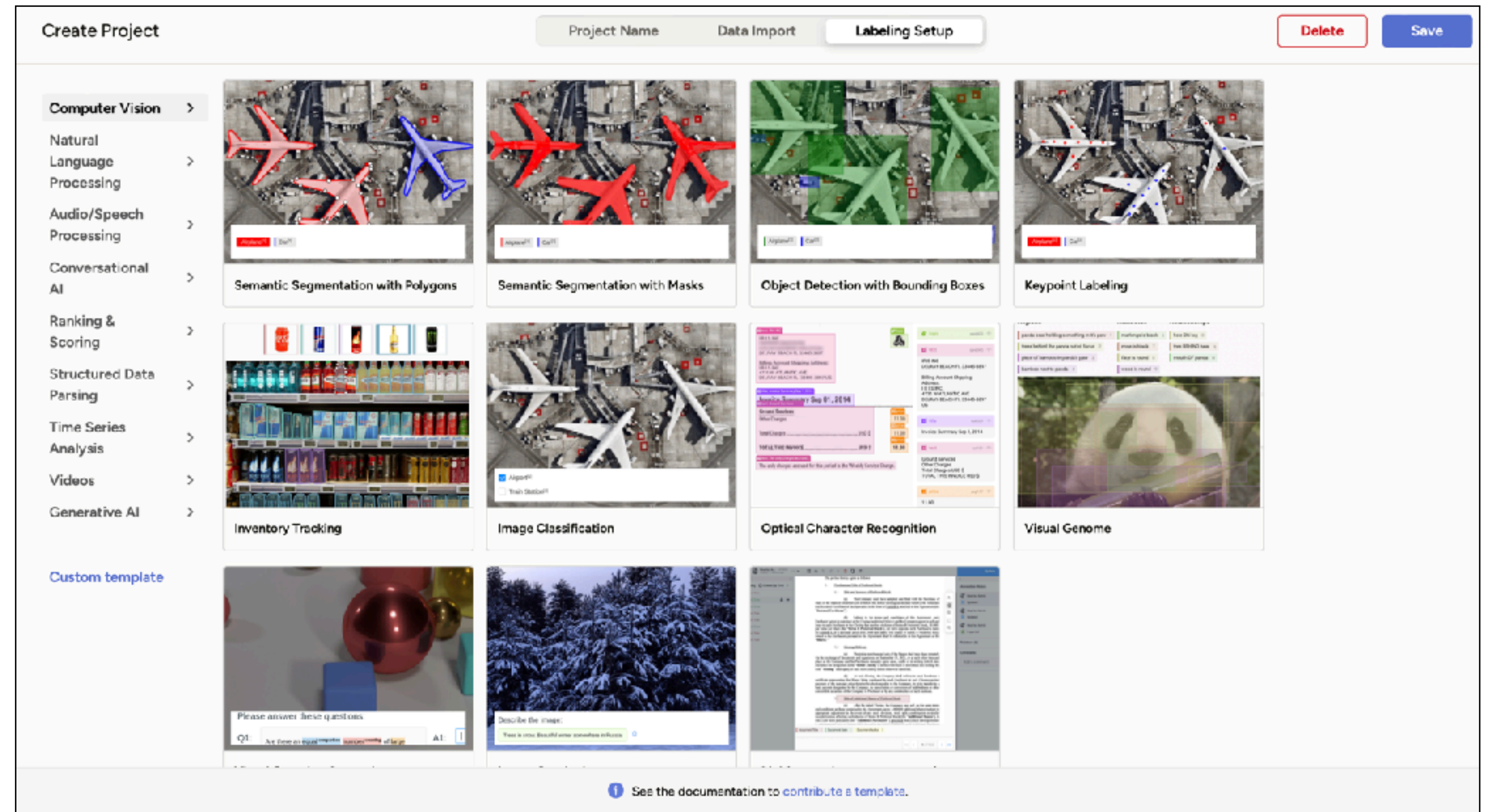
**Créer de la donnée** = montrer par l'exemple la tâche à surmonter  
= faire (généralement) ce travail à la main

**Oui MAIS** = ne peut fonctionner qu'avec un cahier des charges clair, homogène  
= ne peut fonctionner que si déjà en tant qu'annotateur humain on arrive à décrire la tâche et annoter

**LabelStudio** : outil d'annotation générique pour tout type de tâche (images, sons, textes, etc).  
Remplace la plupart des outils existants sur des tâches génériques

**Rappel** : L'entraînement et l'utilisation de modèles est de plus en plus simple aujourd'hui. L'enjeu est **l'identification de ses besoins, la définition d'un cahier des charges réaliste, et la création des données** (annotation).

C'est avant tout une réflexion thématique et non technique.

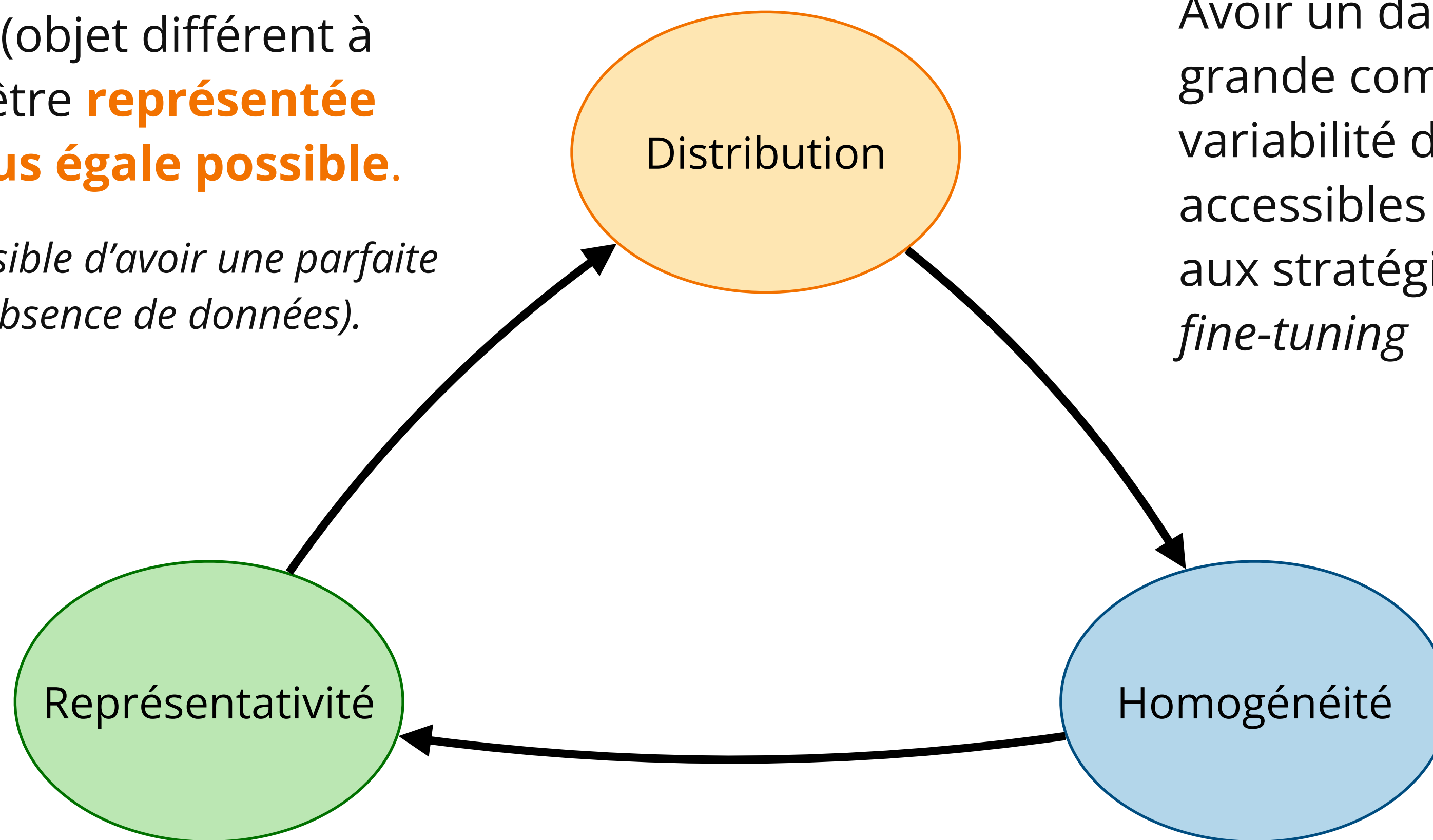


Source image: Interface de création de projets, LabelStudio, 2024



Chaque classe (objet différent à détecter) doit être **représentée de façon la plus égale possible.**

*Il n'est jamais possible d'avoir une parfaite représentativité (absence de données).*



Avoir un dataset suffisamment grand si grande complexité (variété des détections, variabilité des motifs), sinon résultats accessibles dès 50 images annotées grâce aux stratégies de *transfer-learning* et au *fine-tuning*

**Représentativité** des images (couverture de toutes les difficultés / spécificités / variations)

**Homogénéité** et non ambiguïté des annotations : importance du cahier des charges d'annotations

**Problématique** : Je souhaite détecter des motifs dans des enluminures.

**Dataset** : Je récupère via l'API BnF des pages enluminées de manuscrit.

**Composition** : Le dataset contient 150 images de manuscrits latins de la BnF, comprenant :

- 50 enluminures d'oiseaux
- 30 de chiens
- 10 d'arbres
- 3 de colibris
- 2 de corbeaux
- 1 de tulipe
- 1 d'humain
- 1 de cheval
- 1 de coléoptère

**Problème : Dataset déséquilibré**, concentré sur un type unique de documents (manuscrits latins et un cahier des charges de numérisation unique), ambigu (classe oiseau vs classes colibri et corbeau), un motif d'enluminure en particulier (les oiseaux)

**conséquence 1** : le modèle risque de ne pas être performant sur d'autres traditions manuscrites (mauvaise généralisation du modèle)

**conséquence 2** : certaines classes sont trop peu représentées ou ambiguës, il vaut mieux soit les supprimer (elles ne seront jamais détectées de toutes façons), soit les regrouper (annoter les espèces d'oiseaux en *oiseau*) soit ré-équilibrer le dataset avec de nouvelles annotations.

**Problématique** : Je souhaite détecter des motifs dans des enluminures.

**Dataset** : Je récupère via l'API BnF des pages enluminées de manuscrit.

**Composition** : Le dataset contient 150 images de manuscrits latins de la BnF, comprenant :

- 50 enluminures d'oiseaux
- 30 de chiens
- 10 d'arbres
- 3 de colibris
- 2 de corbeaux
- 1 de tulipe
- 1 d'humain
- 1 de cheval
- 1 de coléoptère

**Problème : Dataset déséquilibré**, concentré sur un type unique de documents (manuscrits latins et un cahier des charges de numérisation unique), ambigu (classe oiseau vs classes colibri et corbeau), un motif d'enluminure en particulier (les oiseaux)

Une solution possible pour surmonter ce problème est la mise en place d'une stratégie d'**active learning**, à savoir un entraînement régulier et une annotation des résultats les plus mauvais, pour renforcer le modèle sur les tâches sur lesquelles il échoue. C'est la stratégie retenue et implémentée dans LabelStudio.

Trois installations possibles :

- Installation en **local** sur son ordinateur : nécessite une gestion d'environnements type Conda (Anaconda ou miniconda)
- Installation sur un **serveur partagé** (p. ex. Institutionnel): nécessite un serveur connecté à internet
- Utilisation de **services web** (démon ou payant)

confidentialité, immédiateté, **non collaboratif**, pérennité de la sauvegarde à réaliser

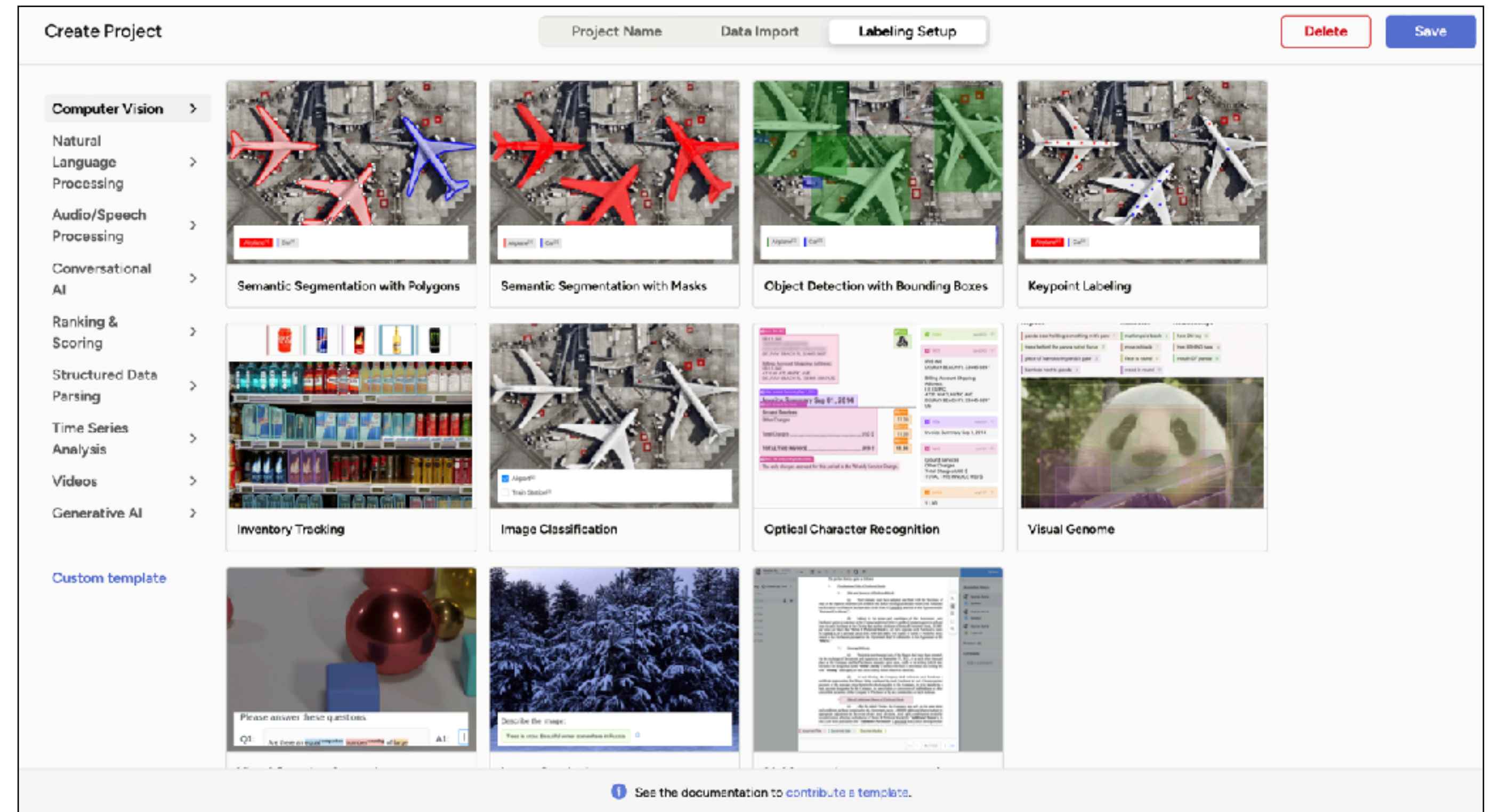
confidentialité, collaboratif, sauvegarde pérenne, potentielle latence, **maintenance à réaliser**

pas d'installation, collaboratif, tâches automatisées, **gestion de la confidentialité**



## Méthodologie :

1. Définition d'une problématique
2. Identification de la tâche (object detection, etc)
3. Téléchargement d'un jeu de données (images) représentatif de la variété et du champ d'application de notre tâche
4. Définition d'un cahier des charges d'annotation et création du projet sur LabelStudio
5. Annotation
6. Bonus : Entraînement d'un modèle
7. Bonus : active learning pour identifier les cas où le modèle n'est pas bon et ré-annotation de ces cas uniquement



Source image: Interface de création de projets, LabelStudio, 2024



**Librairie Ultralytics (YOLO)** : <https://docs.ultralytics.com/>

**Label Studio** : <https://labelstud.io/>

**API et Jeux de données de la BnF** : <https://api.bnf.fr/>

**Wikimedia Commons** : <https://commons.wikimedia.org/wiki/Accueil>

**Quelques exemples et jeux de données utilisés durant l'atelier :**

- Presse arménienne : <https://github.com/CVidalG/workshop-TUMO2022>
- Partitions de musique : <https://github.com/CVidalG/workshop-TUMO2024>
- Papyrus grecs endommagés : <https://github.com/CVidalG/competition-ICDAR2023-papyri>
- Catalogue de manuscrits : <https://calfa.fr/blog/46>

**Exemple de définition de cahier des charges pour l'annotation** : <https://segmonto.github.io/>