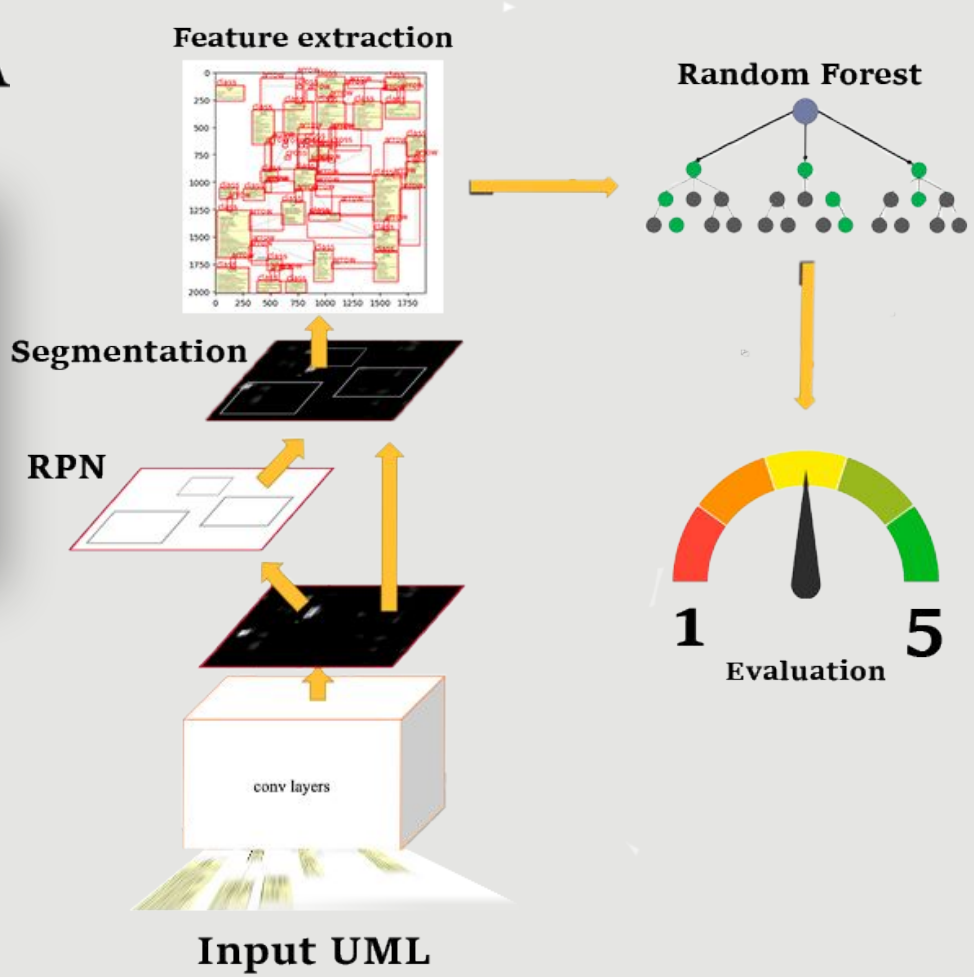
UNIVERSITÀ DI PISA

Using machine learning for automatic classification of the layout quality of UML class diagrams

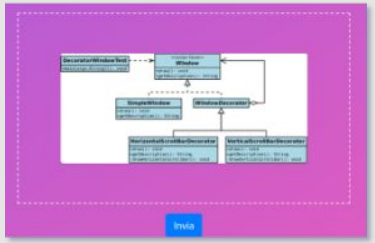
Domenico Armillotta

Supervisor: Prof. Cimino Mario G.C.A.
Alfeo Antonio Luca
Fruzzetti Chiara

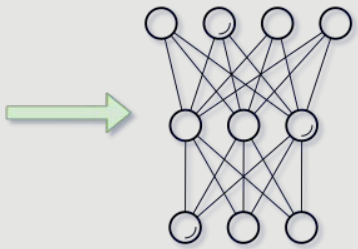


Goal of the Project :

The goal of the project is to create software that given an input image of a UML graph , gives an evaluation on the quality of the schema layout from 0 to 5 and feedback to the designer



Upload schema



Deep learning

Score

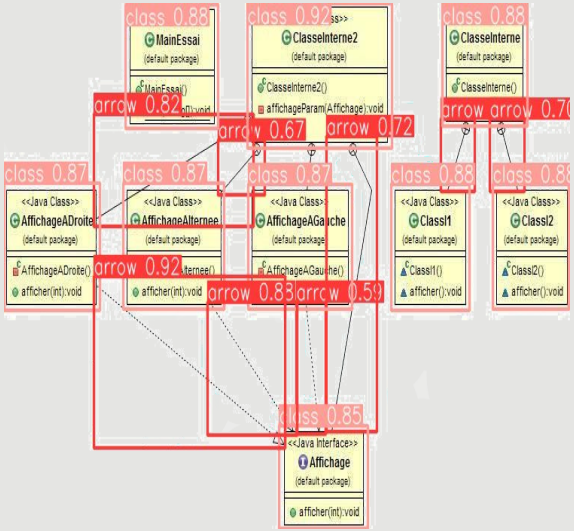
+

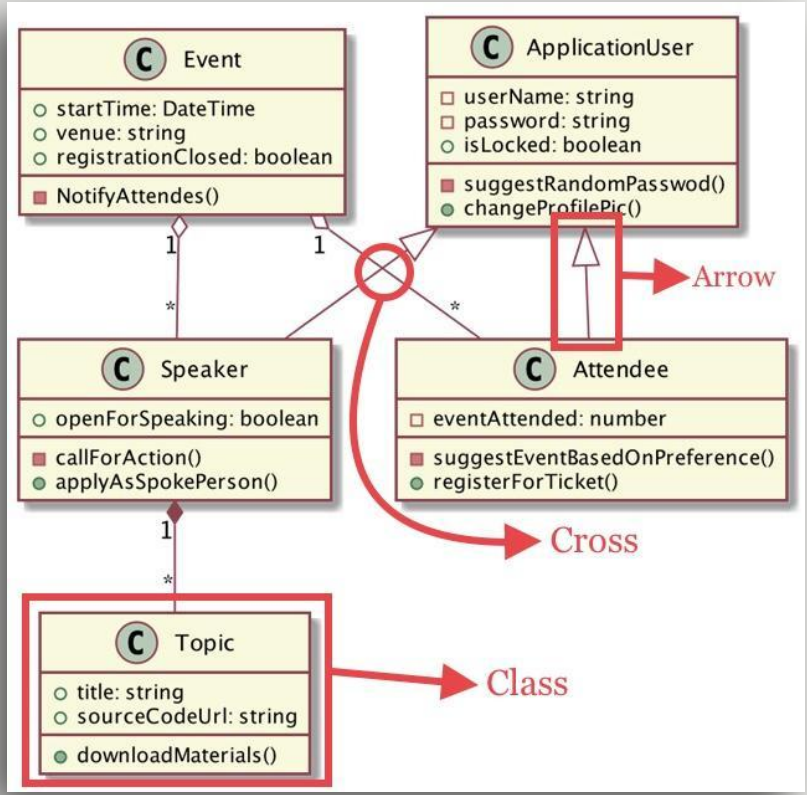
Feedback

What to segment

Segmented item :

- Class
- Arrow
- Cross

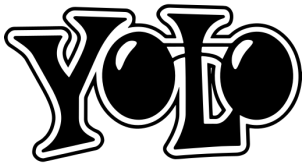




Model trained for segmentation

YOLO V8

Less accurate but slightly faster than Detecto



1

FASTER R-CNN

Very accurate but issue with complex arrows

2

MIXED OPENCV + FASTER R-CNN

Accurate and fast

+

3

Picked the best : Mixed Approach

Classes and crosses : Faster R-CNN

Arrow : Line Detection

Precision : 94%

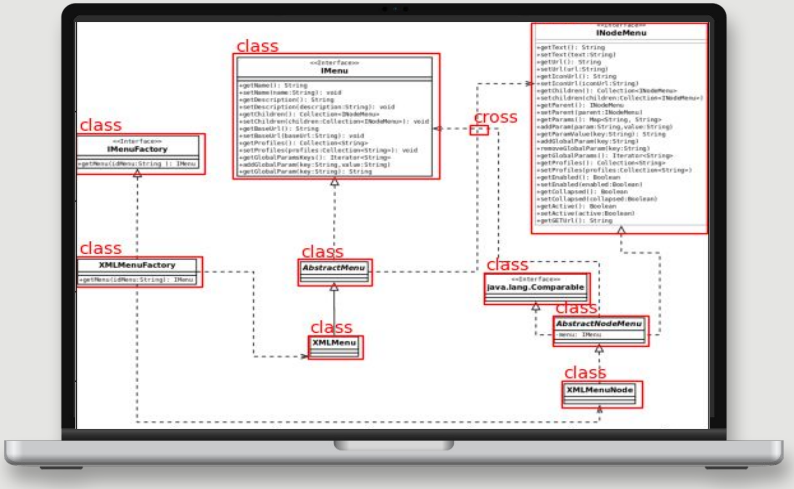
Discarded approach :

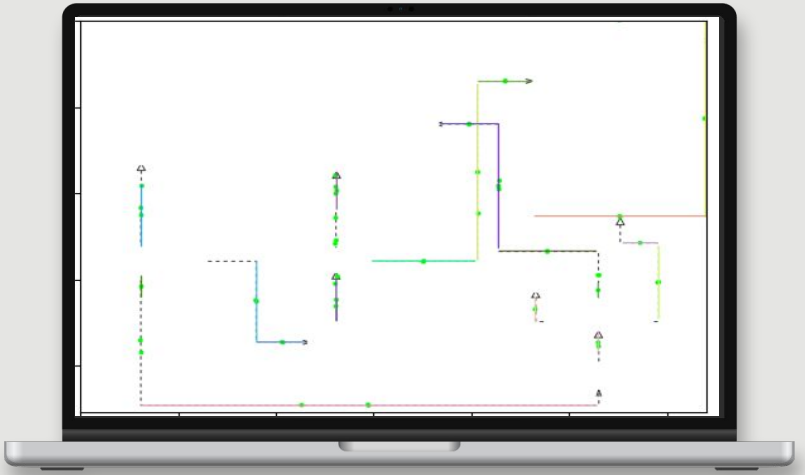
❌ YOLO : lower performance

Accuracy 0.91

❌ Faster R-CNN :

Not precise in the case of complex group of lines





Classifier Trained :

Classifier	Precision	Recall	F1	Deviation
KNN	0.48	0.44	0.46	1.02
Random Forest	**0.65**	**0.59**	**0.62**	**0.75**
Decision Tree	0.47	0.52	0.49	1.00
Naive Bayes	0.38	0.38	0.38	0.92
Neural Network	0.22	0.30	0.25	1.22
Ensemble : SVM , Gradient Boosting & Random Forest	0.50	0.38	0.42	0.87

Performance achieved after dataset balancing with SMOTE , feature importance analysis, and hyperparameter tuning with grid search on n_estimators,max_depth,min_samples_split,min_samples_leaf,max_features (in the case of Random Forest)

Web deploy with Flask

Flask allowed deployment and the latency time for each image is less than 4 seconds

Or textual

In this case is possible to load multiple schema and receive for each the rating and the feedback.

```
############ Image 0  ----> RATING : 5. ############
The Rectangle proximity is already optimal for image 0.
The Number of rectangles is already optimal for image 0.
The Line bends is already optimal for image 0.
The Number of lines is already optimal for image 0.
The Longest line is already optimal for image 0.
The Average line length is already optimal for image 0.
############ Image 1  ----> RATING : 1. ############
The Rectangle proximity should be decreased for image 1.
The Number of rectangles should be decreased for image 1.
The Line bends should be decreased for image 1.
The Number of lines should be decreased for image 1.
The Longest line should be decreased for image 1.
The Average line length should be decreased for image 1.
############ Image 2  ----> RATING : 2. ############
The Rectangle proximity should be decreased for image 2.
The Number of rectangles should be decreased for image 2.
The Line bends should be decreased for image 2.
The Number of lines should be decreased for image 2.
The Longest line should be decreased for image 2.
The Average line length should be decreased for image 2.
```