A Preliminary Study of Few-shot Learning for Layout Analysis of Music Scores

Presentation for WORMS 2023

Authors:
Francisco J. Castellanos[1]
Antonio Javier Gallego[1]
Ichiro Fujinaga[2]

[1] *University Institute for Computing Research*
University of Alicante, Alicante, Spain

[2] *Schulich School of Music*
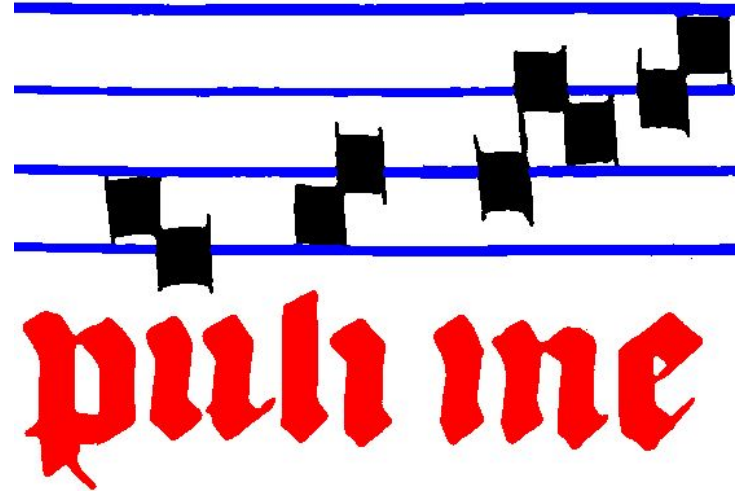*McGill University, Montreal, Canada*

November 2023

Layout Analysis (LA)

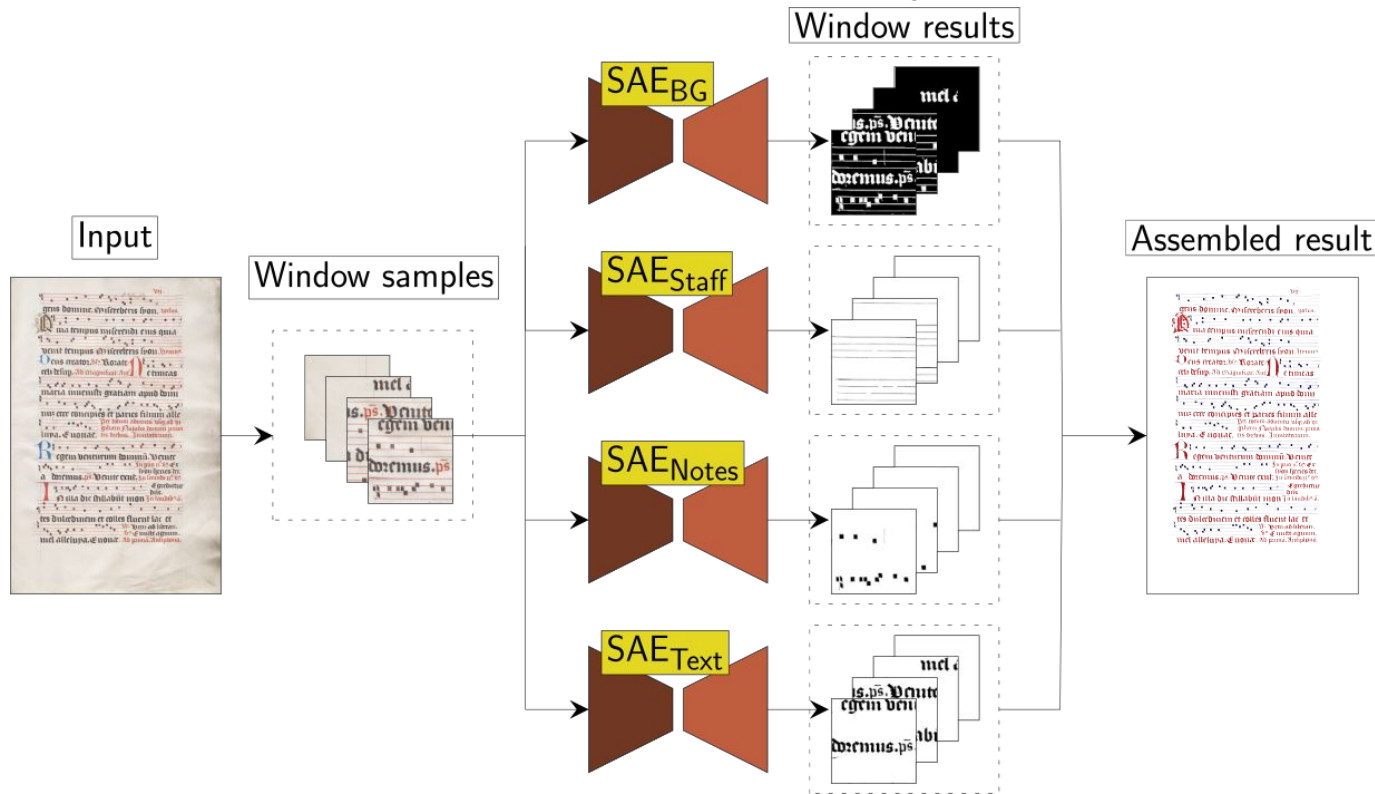- Document analysis process to segment the image.

Layout Analysis for Optical Music Recognition (OMR)

Previous work:

- A Selectional Auto-encoder (SAE) for each layout of information.

Motivation

- High detail and density.
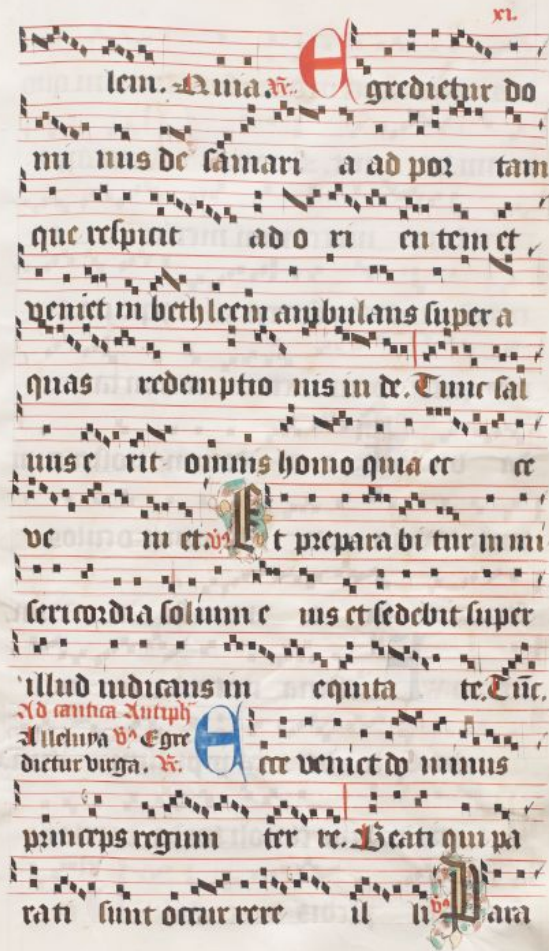- Full-page pixel-wise annotations.
- High resolution.

Resolution: 5,896 × 3,839 px

To be annotated:

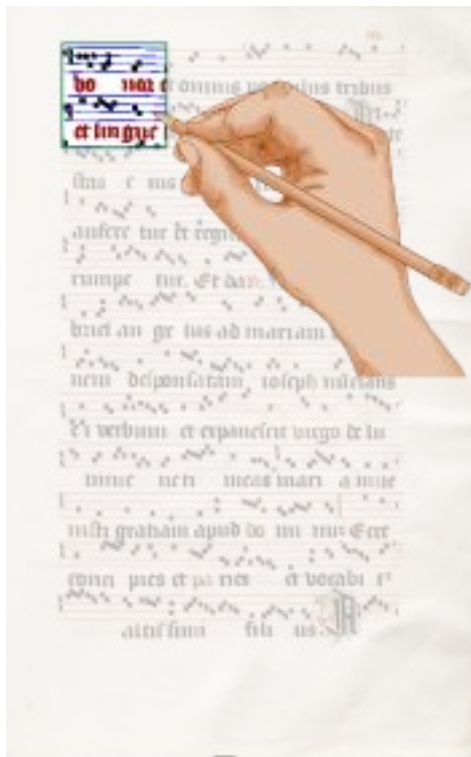|	
Symbols	682,674 px.
Staff:	557,169 px.
Text:	1,572,082 px.

Proposal: Few-shot Layout Analysis

- Reduce the number of annotations per page.

- Reduce the number of pages to be annotated.

Proposal: Few-shot Layout Analysis (1/3)

- **Step 1:** manually annotating some patch samples.
- Simple selection:
  - Sequential without overlap.
  - With 2.5% of annotations.
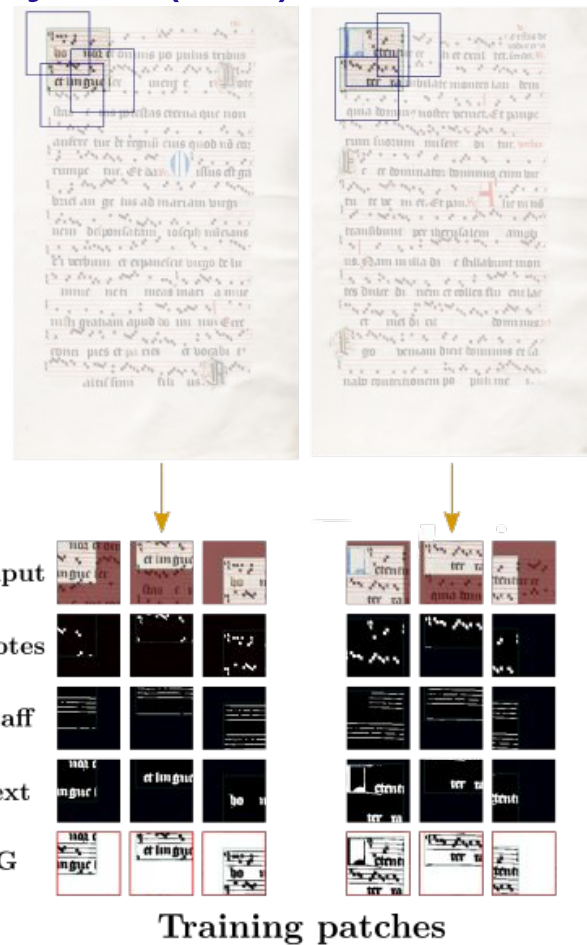 - Window size: 256 × 256 px.

P_1

P_2

Proposal: Few-shot Layout Analysis (2/3)

- **Step 2:** sample extraction.

  - Random windows around the annotated patches.

  - Number of samples to be studied.

  - Window size: 256 × 256 px.

Training patches

Proposal: Few-shot Layout Analysis (3/3)

- **Step 3:** training the models.
  - Selectional Auto-encoder models, as the SOTA.
  - One model per layer.
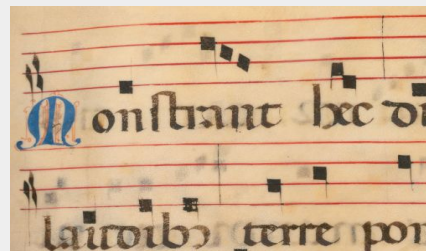 - A masking layer ignores the non-annotated pixels.
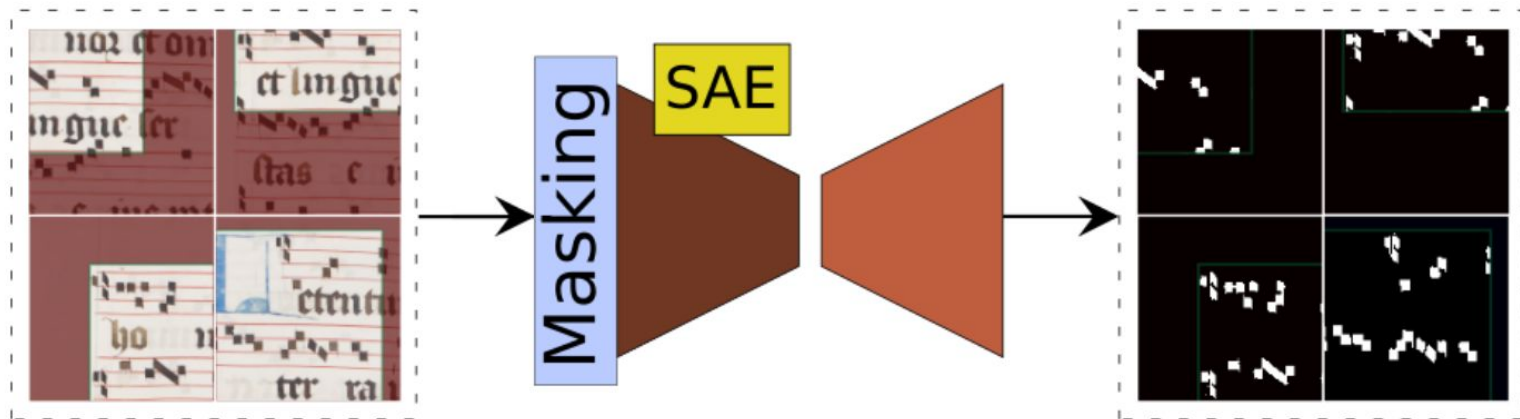
Corpora

| Salzinnes | Einsiedeln | MS73 | Capitan |

Metrics

- **F-score (F_1)**

$$\mathrm{F}_1 = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

- **Macro F-score (F_1^m)**

$$\mathrm{F}_1^m = \frac{\sum_{l=1}^{|\mathcal{L}|} \mathrm{F}_1^l}{|\mathcal{L}|}$$

where:

- TP: True positives

- FP: False positives

- FN: False negatives

- F_1^l is the F_1 for the layer l.

- $\mathcal{L}$ is the set of layers.

Preliminary results

- To study the number of random patch samples to be extracted.

- Our proposal does not require data augmentation.

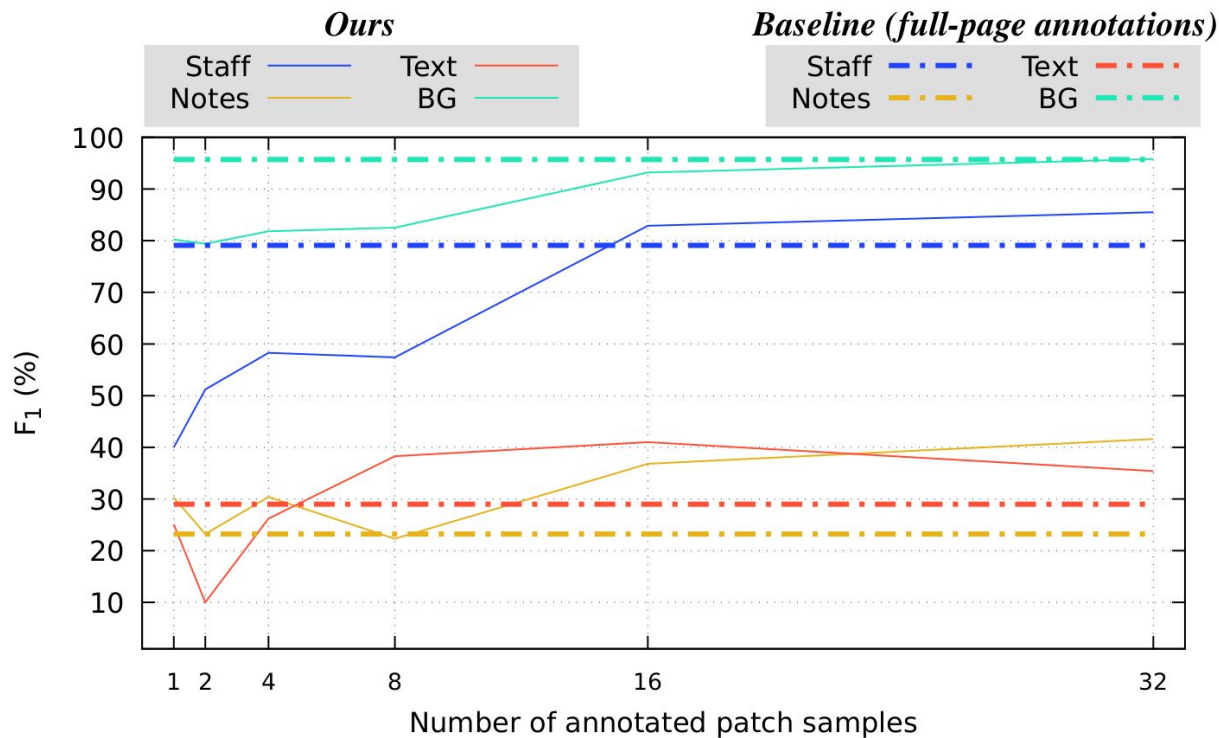- Number of random samples: **512.**
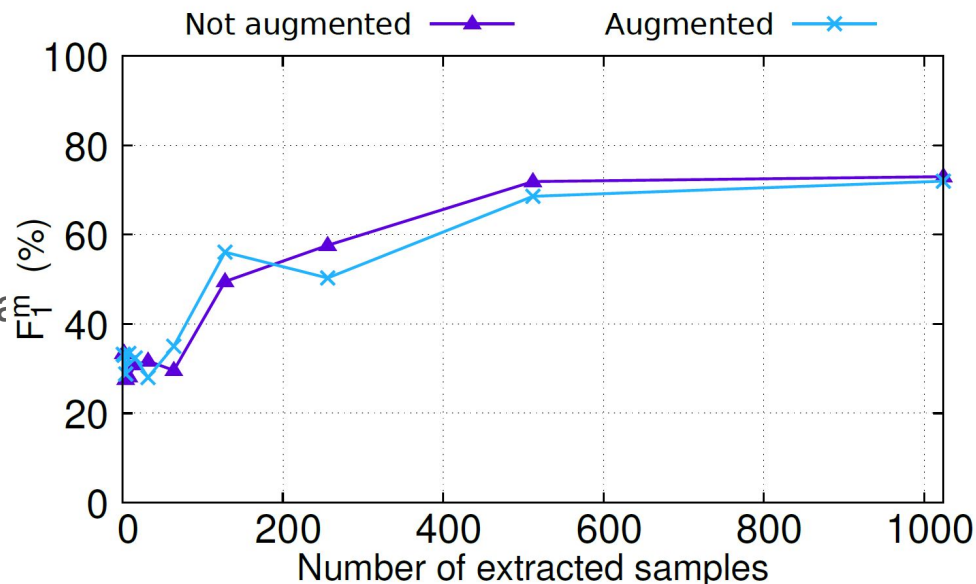
Experiments

- **Case Study I:** only 1 page available for training.

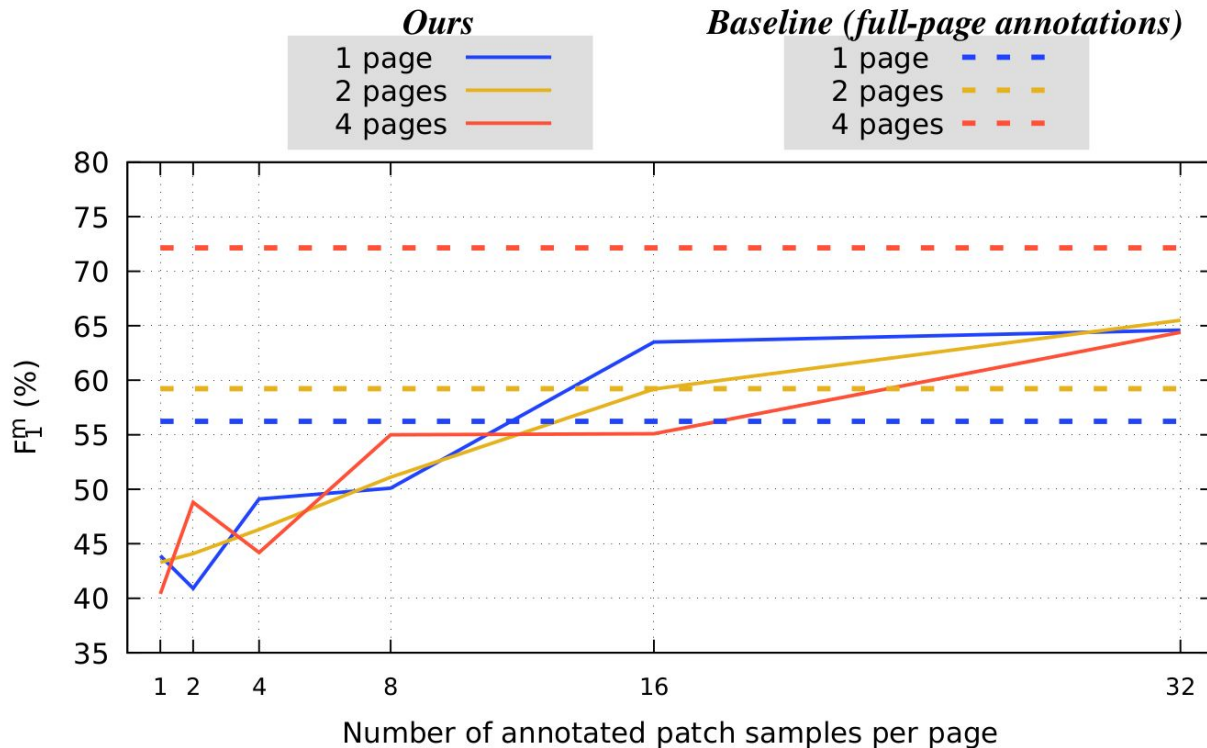- **Case Study II:** variable number of pages for training.

Results. Case Study I: 1 page

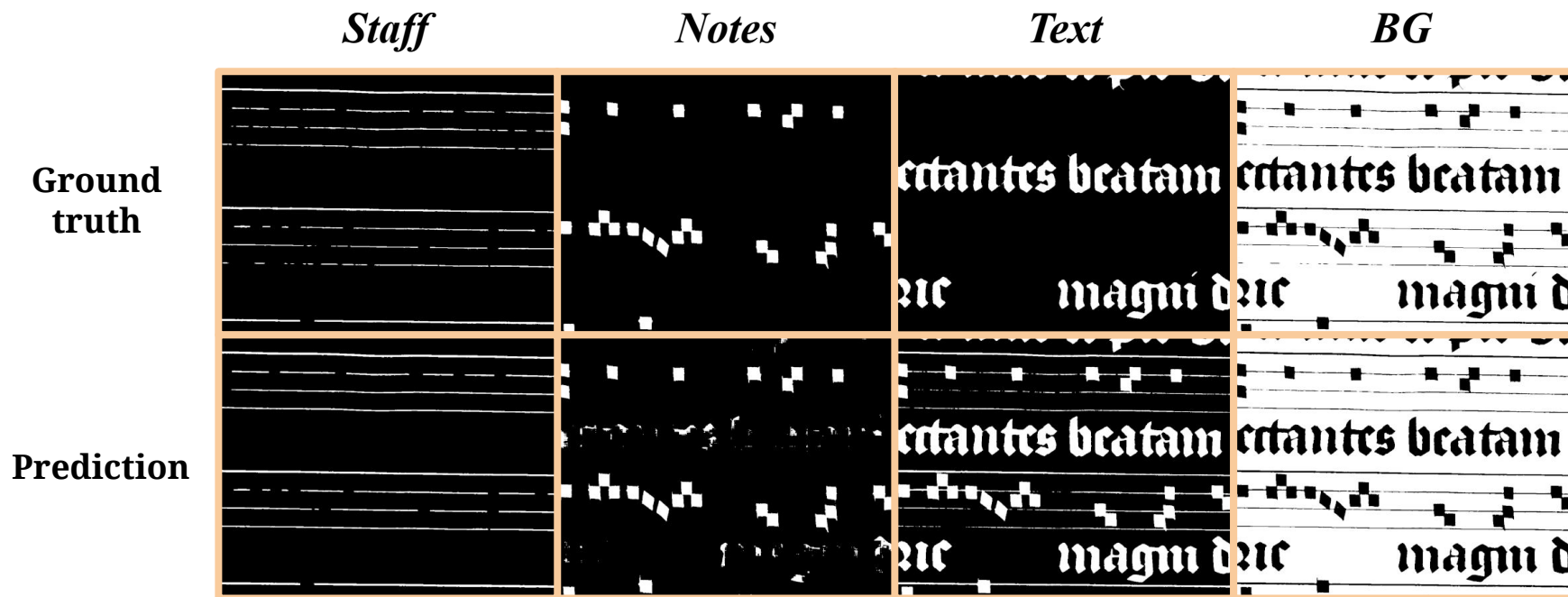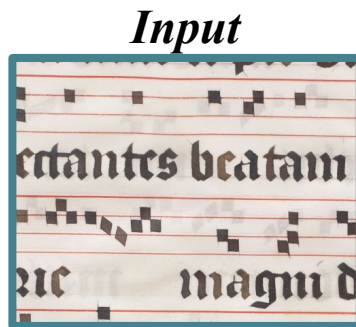- Scenario with a limited training set (only 1 page)

Results. Case Study II: multiple pages

- Scenario with variable number of pages.

Results. Qualitative result

Input

- 32 annotated patches in one page

	Staff	*Notes*	*Text*	*BG*
Ground truth				
Prediction				

Conclusions

- One page is sufficient to obtain competitive performance.

    - Annotations are required for 32 patch samples of one page.

    - Our performance (65.5%) is near to the baseline (72%) with an important reduction in ground truth..

- Our method improves the baseline with less than 4 pages.

- Room for improvement when compared to training with 4 full pages.

Future work

- We plan to investigate the combination of our proposal with:

    - Domain adaptation.

    - Incremental and active learning.

A Preliminary Study of Few-shot Learning for Layout Analysis of Music Scores

Presentation for WORMS 2023

November 2023