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

Universitat d'Alacant
Universidad de Alicante

McGill UNIVERSITY

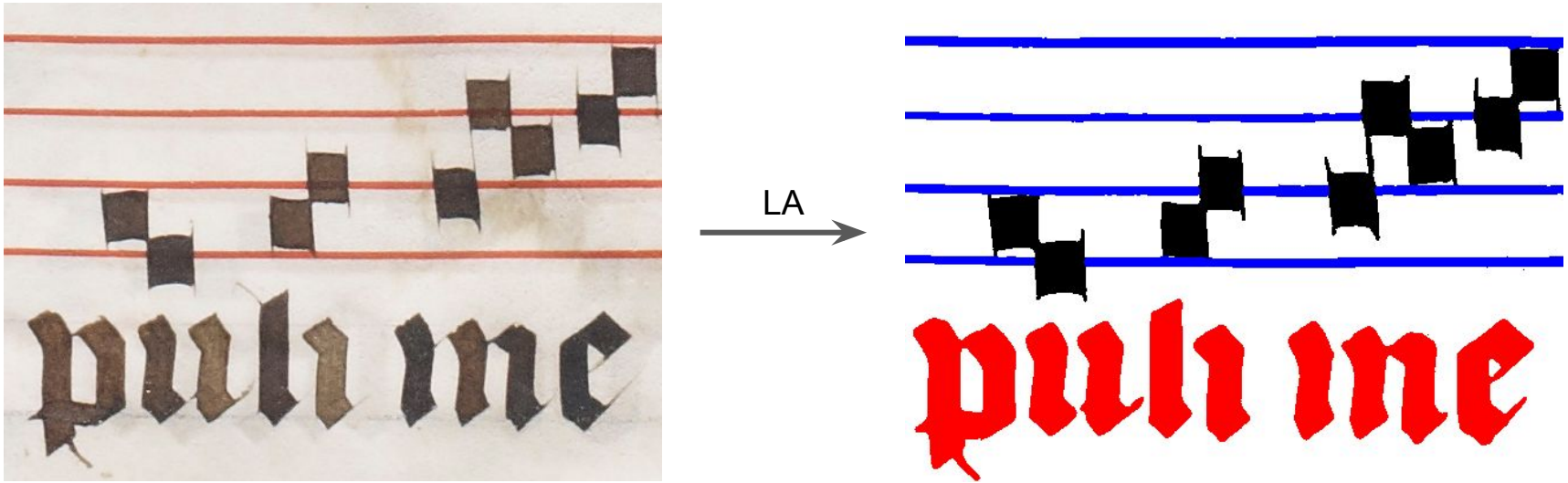
Plan de Recuperación, Transformación y Resiliencia

Social Sciences and Humanities Research Council of Canada

Fonds de recherche Société et culture Québec

Layout Analysis (LA) for Optical Music Recognition (OMR)

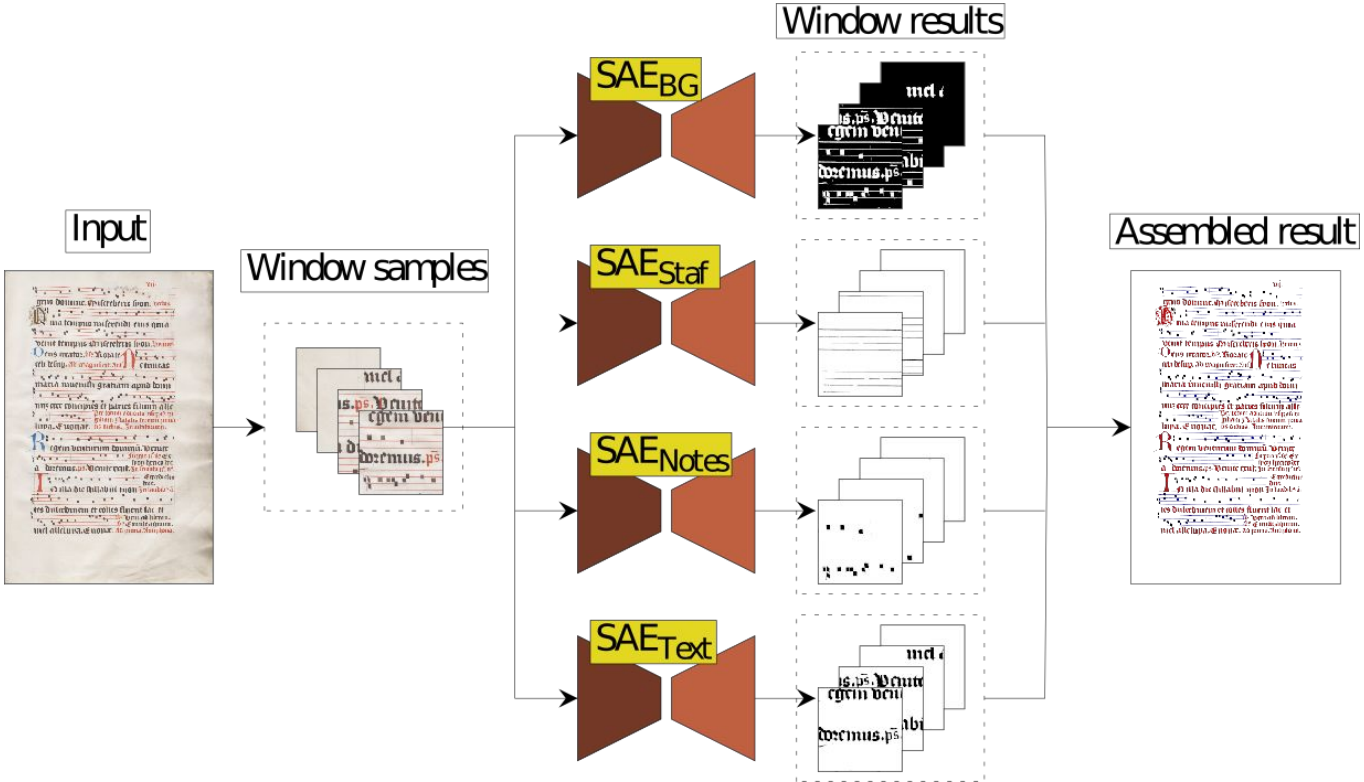
- Document analysis process to segment the image.

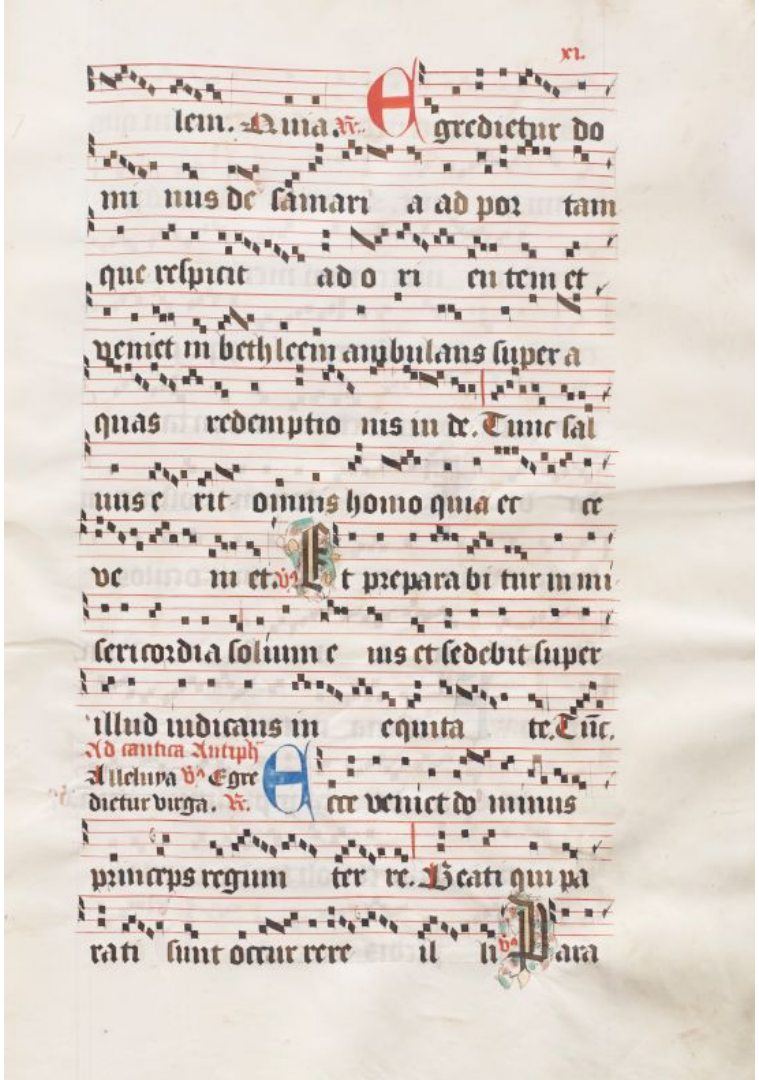


Layout Analysis (LA) for Optical Music Recognition

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

- Given partial annotations, obtaining random patches around them.
- Masking layer to ignore the non-annotated pixels.
  - Non-annotated pixels are replaced by -1.
- Only a small portion of data is required to be annotated.

Proposal: Few-shot Layout Analysis

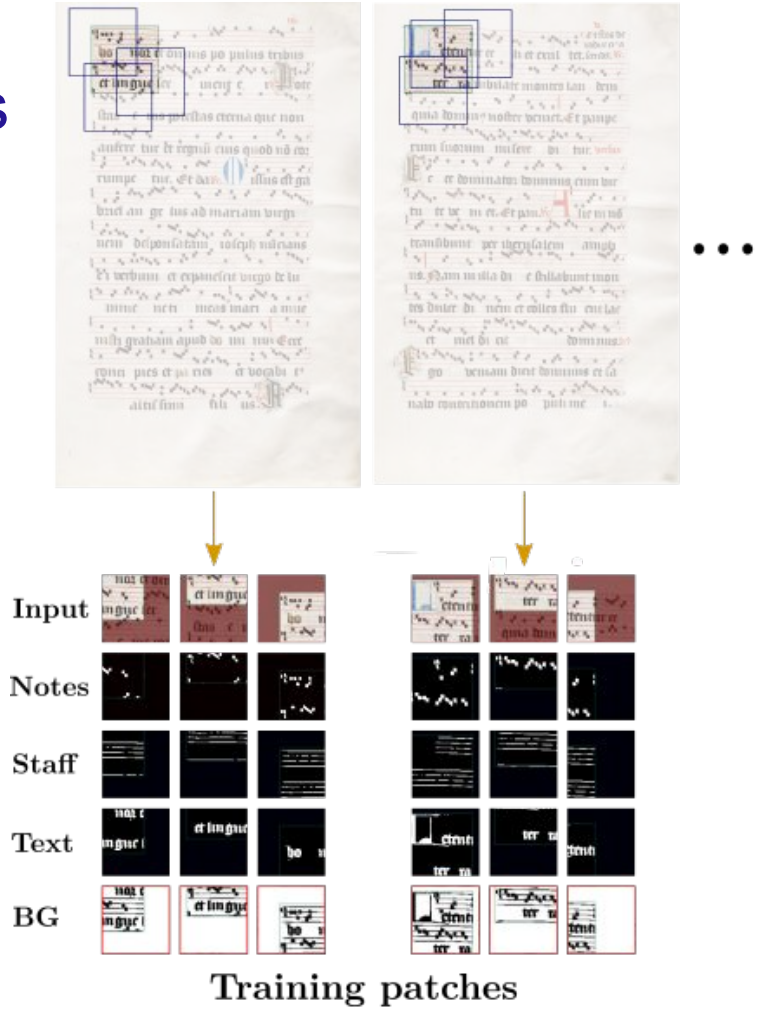
- **Step 1:** manually annotating some patch samples.
- Simple selection:
  - Sequential without overlap.
  - With 2.5% of annotations.

P_1

P_2

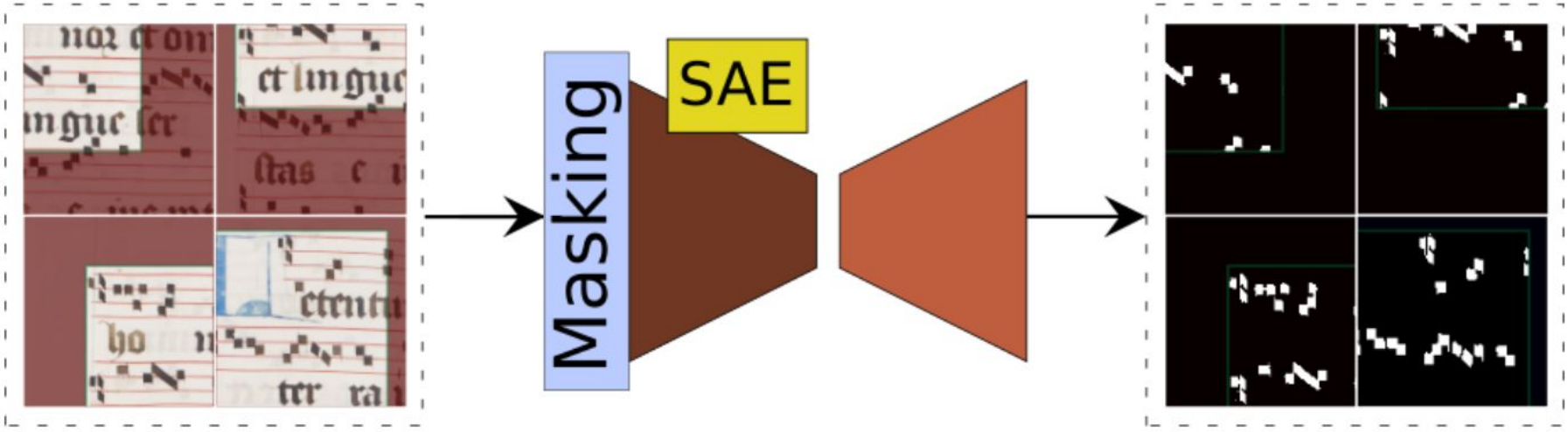
Proposal: Few-shot Layout Analysis

- **Step 2:** sample extraction.
  - Random windows around the annotated patches.
  - Number of samples to be studied.
  - Sample size: 256 × 256 px.

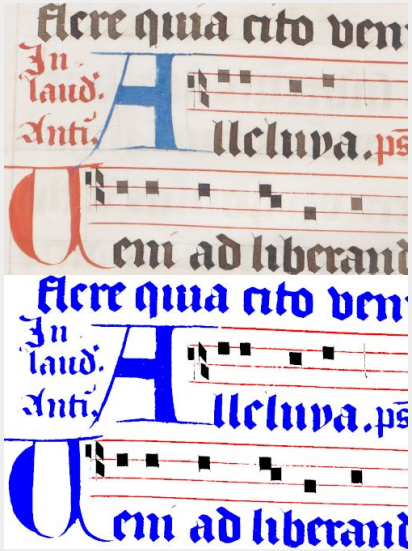


Proposal: Few-shot Layout Analysis

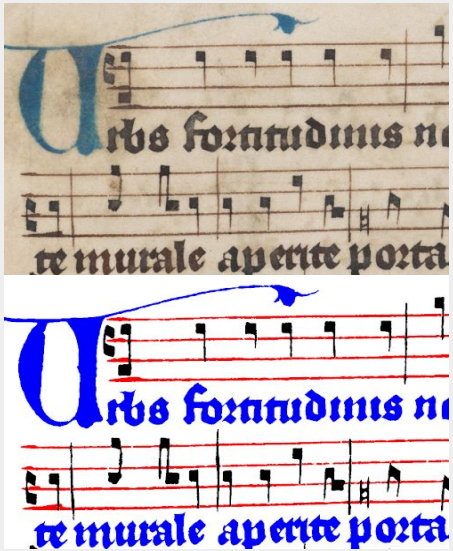
- **Step 3:** training the models.
  - Selectional Auto-encoder models, as the SOTA.
  - One model per layer.
  - A masking layer ignores the non-annotated pixels.



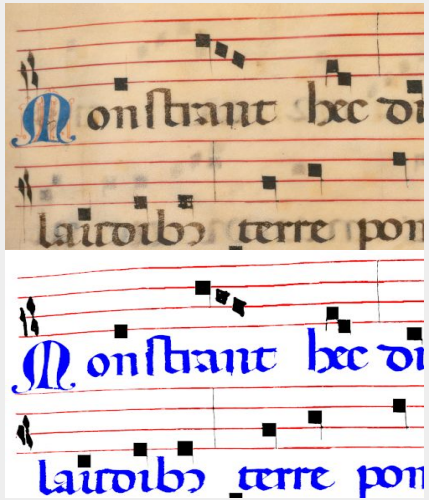
Corpora



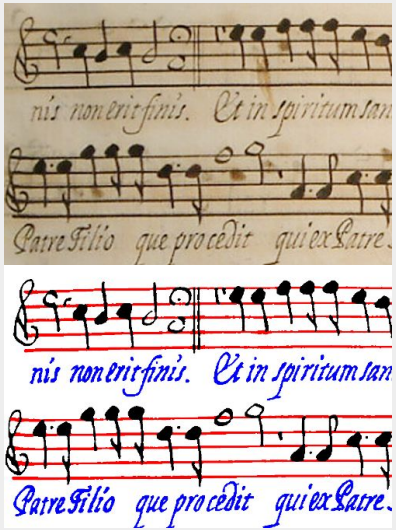
Salzinnes



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 ($\mathrm{F}_1^{\mathrm{m}}$)**

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

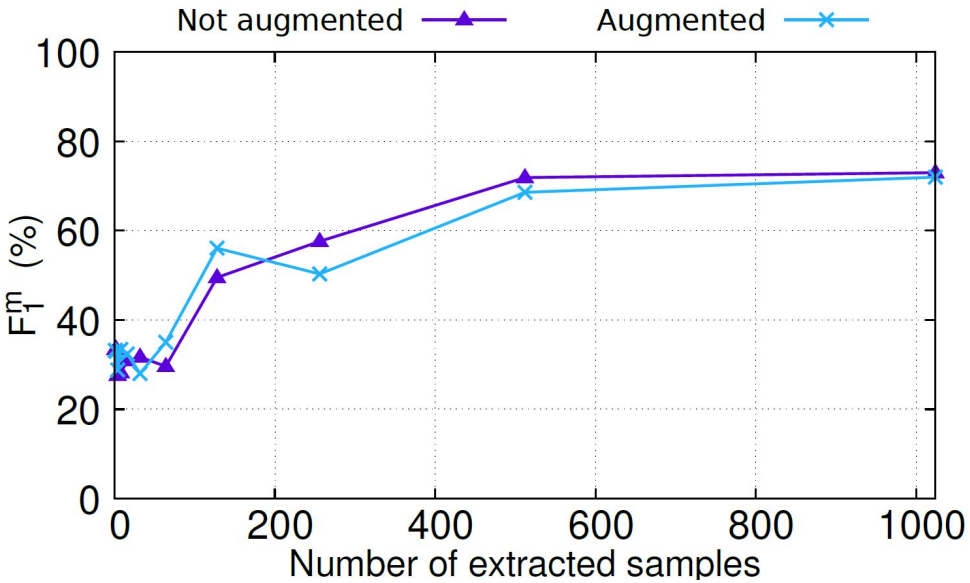
where:

- 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.**

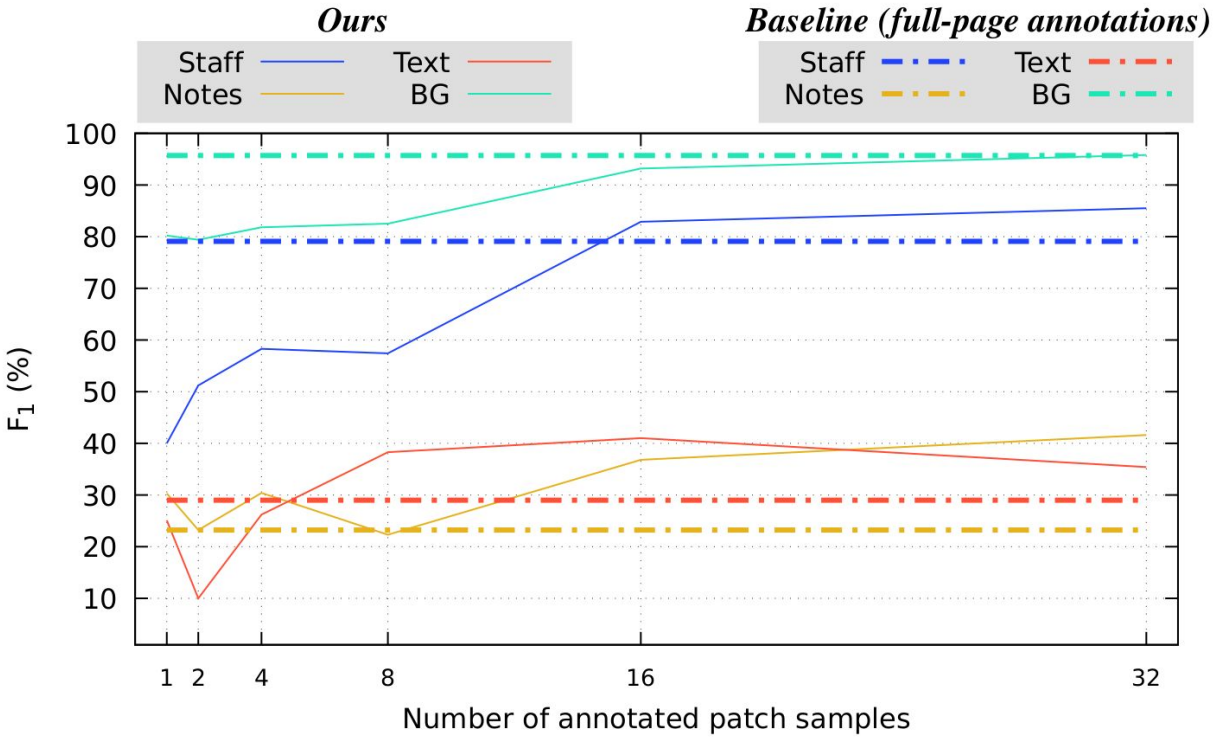


Experiments

- **Study case I:** only 1 page available for training.
- **Study case II:** variable number of pages for training.

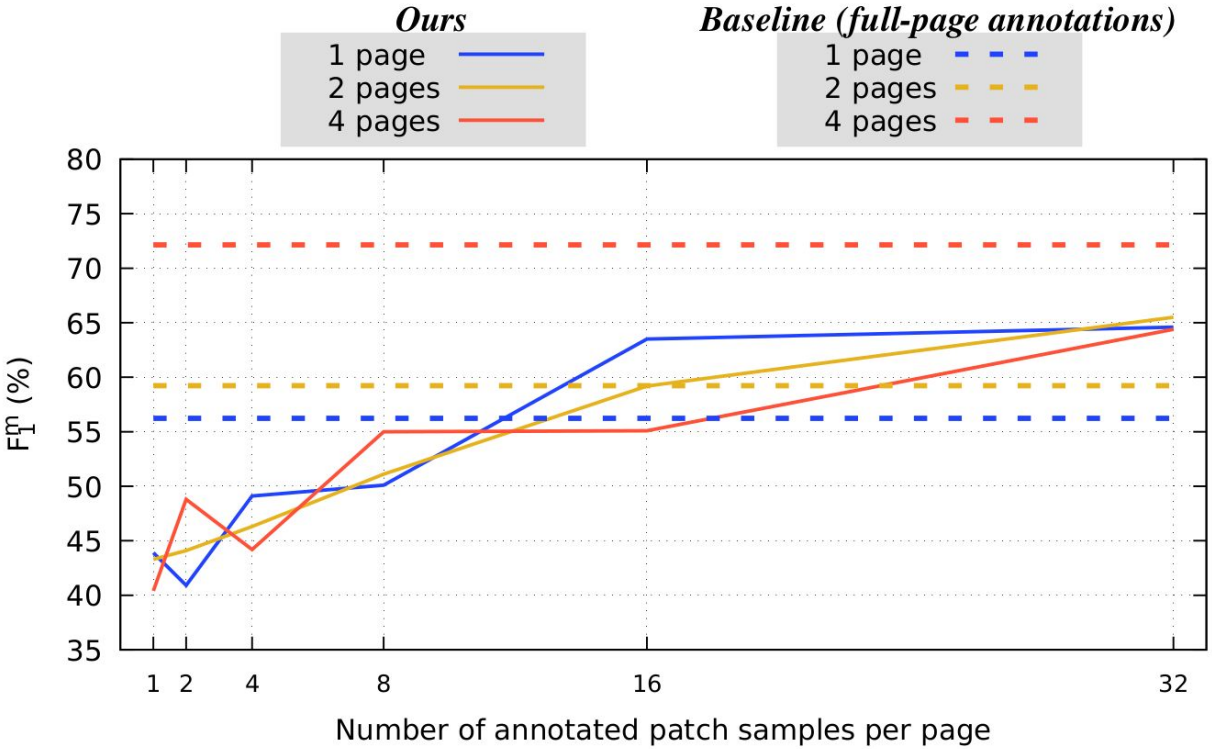
Results. Study case I: 1 page

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



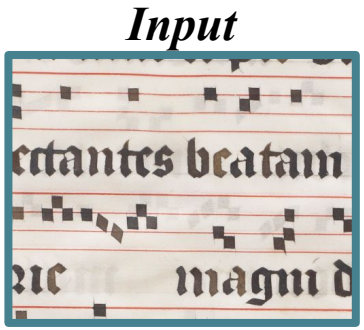
Results. Study case II: multiple pages

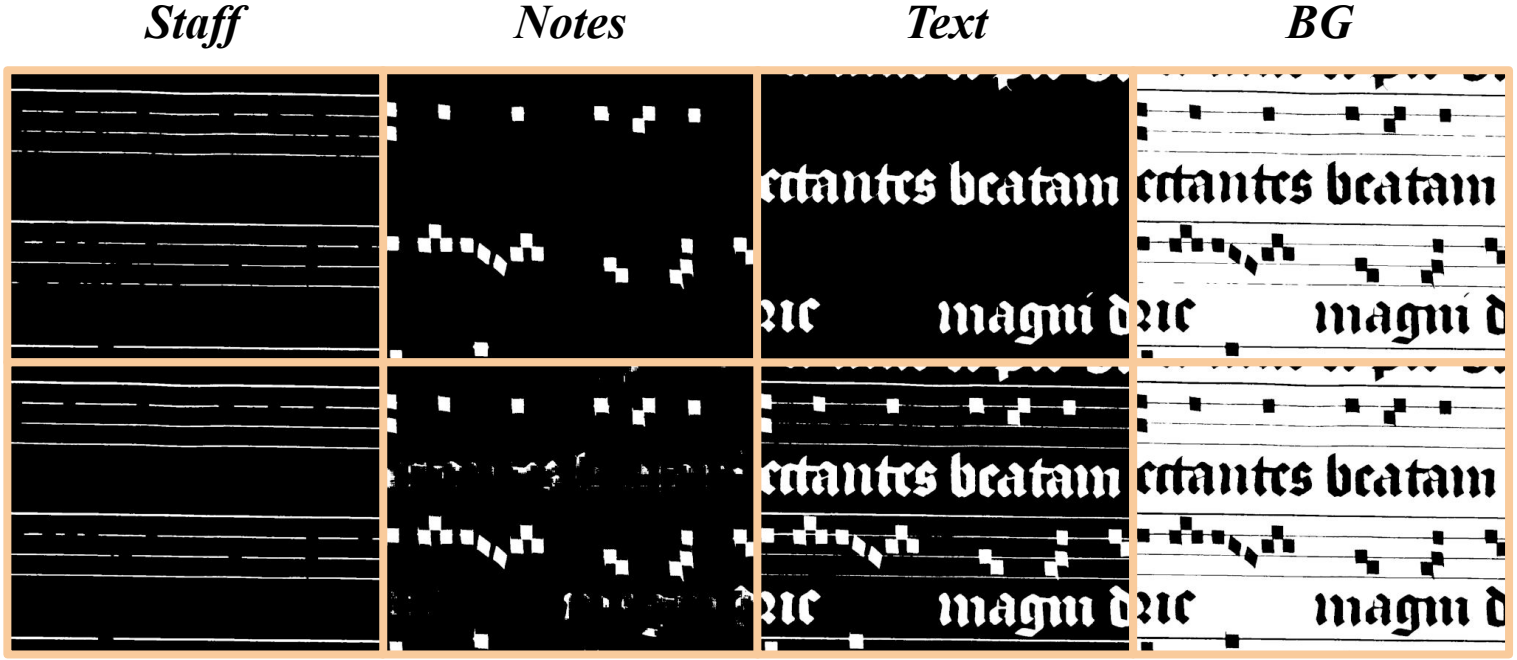
- Scenario with variable number of pages.



Results. Qualitative result

- 32 annotated patches in one page





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

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

Universitat d'Alacant
Universidad de Alicante

McGill UNIVERSITY

Plan de Recuperación, Transformación y Resiliencia

Social Sciences and Humanities Research Council of Canada

Fonds de recherche Société et culture Québec