International Conference on Document Analysis and Recognition
2024, Athens, Greece

University of Alicante, Spain
Pattern Recognition and Artificial Intelligence Group

A region-based approach for layout analysis of music score images in scarce data scenarios

Francisco J. Castellanos, Juan P. Martinez-Esteso, Alejandro Galán-Cuenca, Antonio Javier Gallego
fcastellanos@dlsi.ua.es - juan.martinez11@ua.es - a.galan@ua.es - jgallego@dlsi.ua.es

- **Optical Music Recognition (OMR)** enables **preservation** and **accessibility** of historical manuscripts.
- Neural network-based methods require a **large amount of labeled data**, which must be obtained at a high cost.
- This paper focuses on detecting bounding boxes of staves, a common step in OMR so-called **Layout Analysis (LA)**.
- The main goal is to train a LA method using scarce labeled training sets.

FSAE: Few-shot Selectional Auto-Encoder

Our proposal (FSAE) is based on the adaptation of an existing approach (Selectional Auto-Encoder) to work with partial annotations for LA. It consists in several steps:

1. Manually **annotating** bounding boxes.
2. Adaptation of **image scale** according to the half of a window height ($\sigma = 0.5$).
3. **Extracting** λ random patch samples around the annotation, as many as needed (*oversampling*).
4. **Training** the model that includes a masking layer to ignore non-annotated pixels while training.

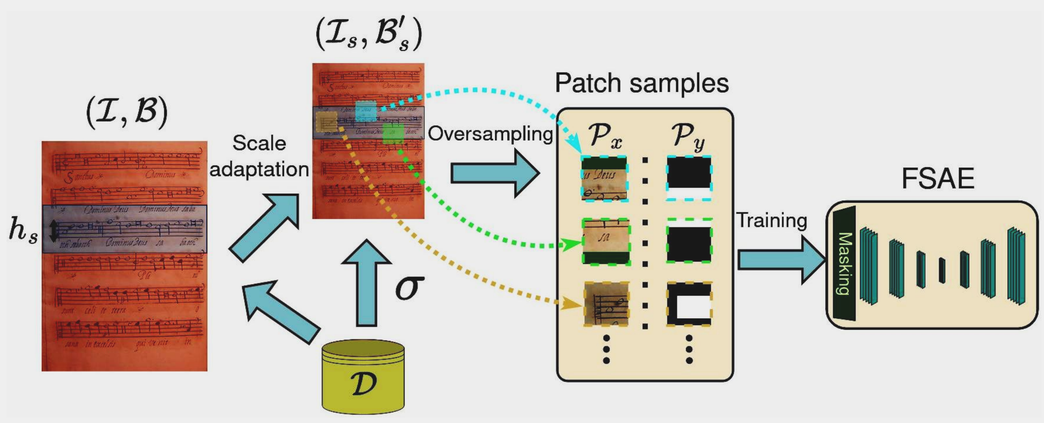
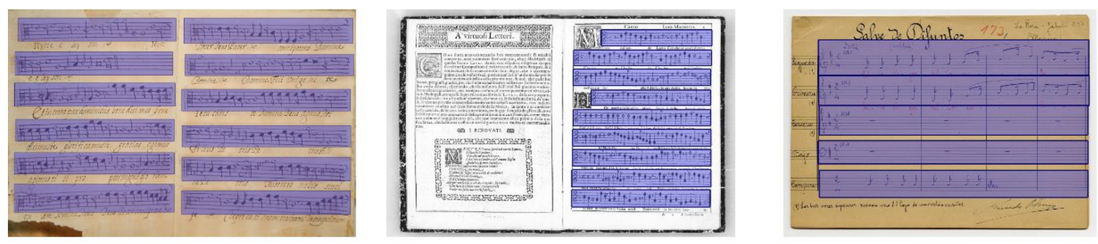


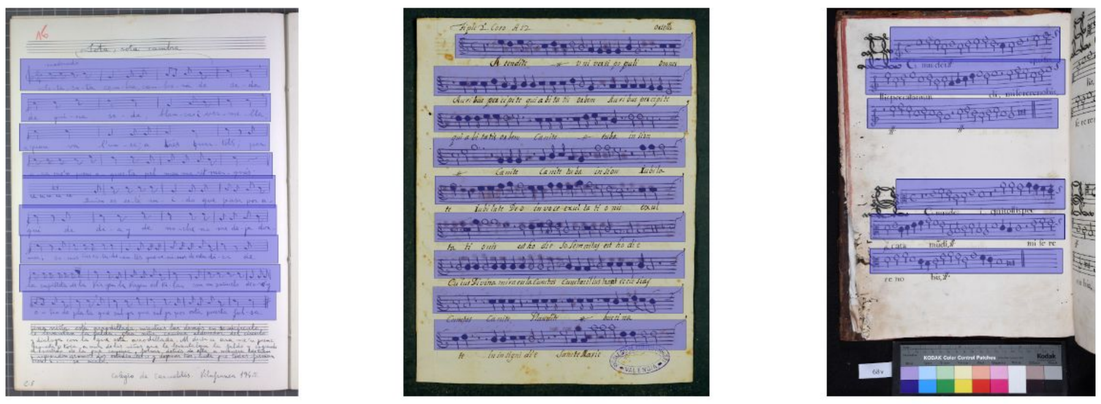
Fig. 1: Scheme of the proposed few-shot method.

Experiments

Corpora



(a) Capitan (b) Seils (c) FMT-M



(d) FMT-C (e) Patriarca (f) Guatemala

Fig. 2: Examples of the corpora considered with their ground truth.

Staff retrieval experiments

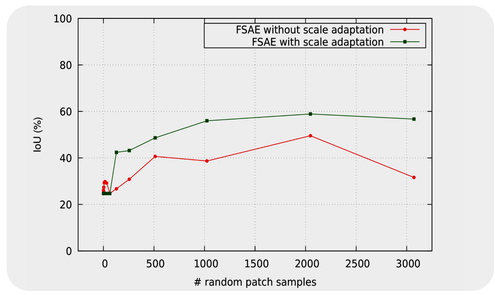


Fig. 2: Study of the number of random samples training with only 1 labeled staff.

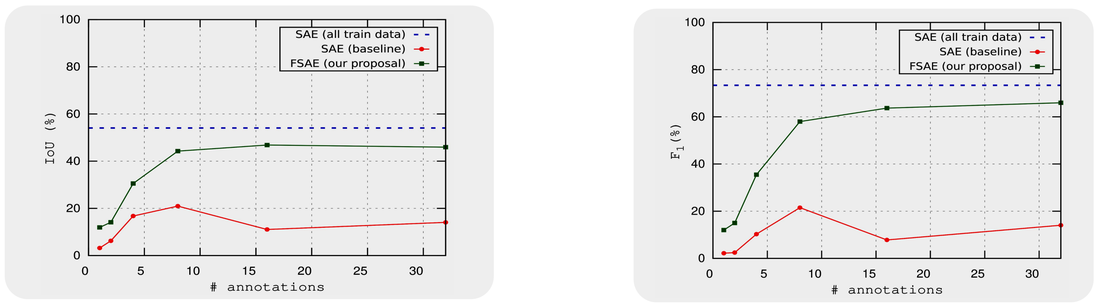
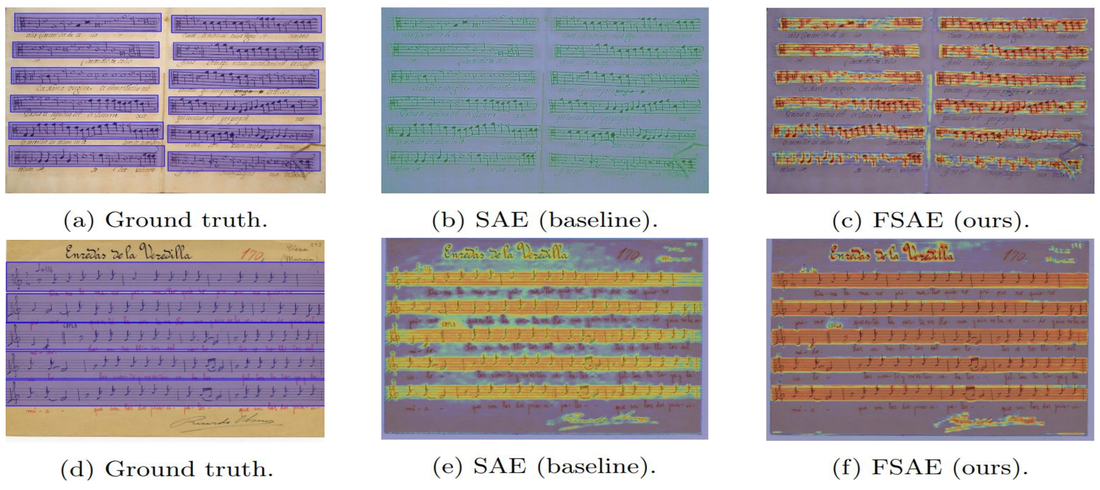


Fig. 3: Average results with respect to the number of labeled staves (λ).

	SOTA (with all training data)			Few-shot scenario (with λ training annotated staves)					
				SAE			FSAE (ours)		
Metric	RetinaNet	YOLO	SAE	λ = 8	λ = 16	λ = 32	λ = 8	λ = 16	λ = 32
F_1 (%)	75.1	88.6	73.4	21.5	8.3	14.1	57.0	63.7	66.0

Tab. 1: Average F_1 (%)$^{IoU=0.5}$ results comparing SOTA with the few-shot cases.



(a) Ground truth. (b) SAE (baseline). (c) FSAE (ours).

(d) Ground truth. (e) SAE (baseline). (f) FSAE (ours).

Fig. 4: Qualitative results with SAE (SOTA) and FSAE.

Transcription experiments

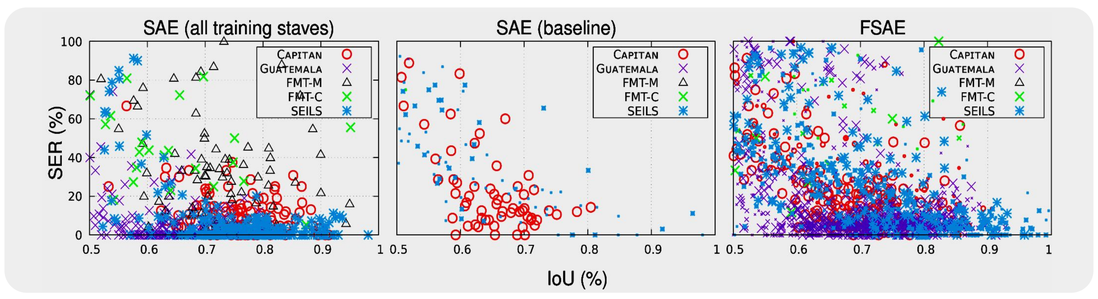
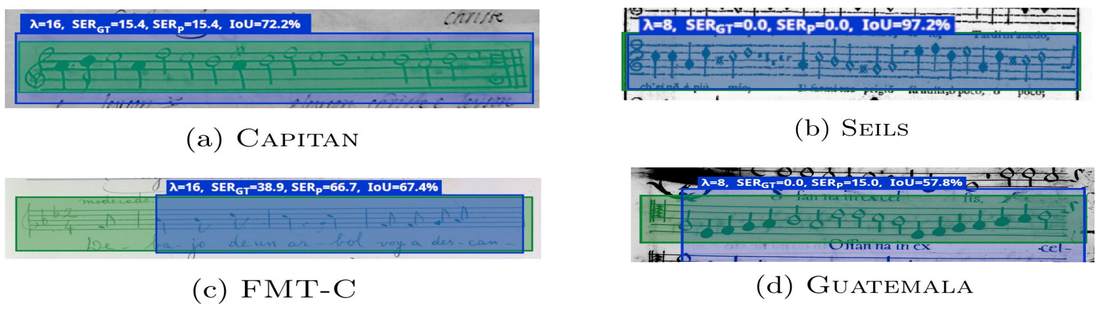


Fig. 5: Transcription results in terms of Symbol Error Rate (%).



(a) Capitan (b) Seils

(c) FMT-C (d) Guatemala

Fig. 6: Examples of retrieved staves by using λ training staves.

Conclusions

- We introduced a novel layout analysis framework for OMR working under few-shot conditions.
- Our approach, FSAE, enables partial manual annotations to train a robust staff-retrieval model.
- Annotating between 8 and 32 staves is sufficient to obtain competitive performances.

Universitat d'Alacant
Universidad de Alicante

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES AGENCIA ESTATAL DE INVESTIGACIÓN

Funded by the European Union
NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia