

A FEW-SHOT NEURAL APPROACH FOR LAYOUT ANALYSIS OF MUSIC SCORE IMAGES

Francisco J. Castellanos¹ - Antonio Javier Gallego¹ - Ichiro Fujinaga²

¹ University Institute for Computing Research, University of Alicante, Spain

² Schulich School of Music, McGill University, Montreal, Canada

{fcastellanos, jgallego}@dlsi.ua.es, ichiro.fujinaga@mcgill.ca

Presented at the 24th International Society for Music Retrieval Conference, 5-9, 2023.

1. INTRODUCTION

- Layout analysis is a common step within the traditional Optical Music Recognition (OMR) workflow.
- State-of-the-art methods require a great amount of annotated data obtained by hand, being a high-cost and error-prone task.
- Proposal: to integrate a few-shot learning strategy.

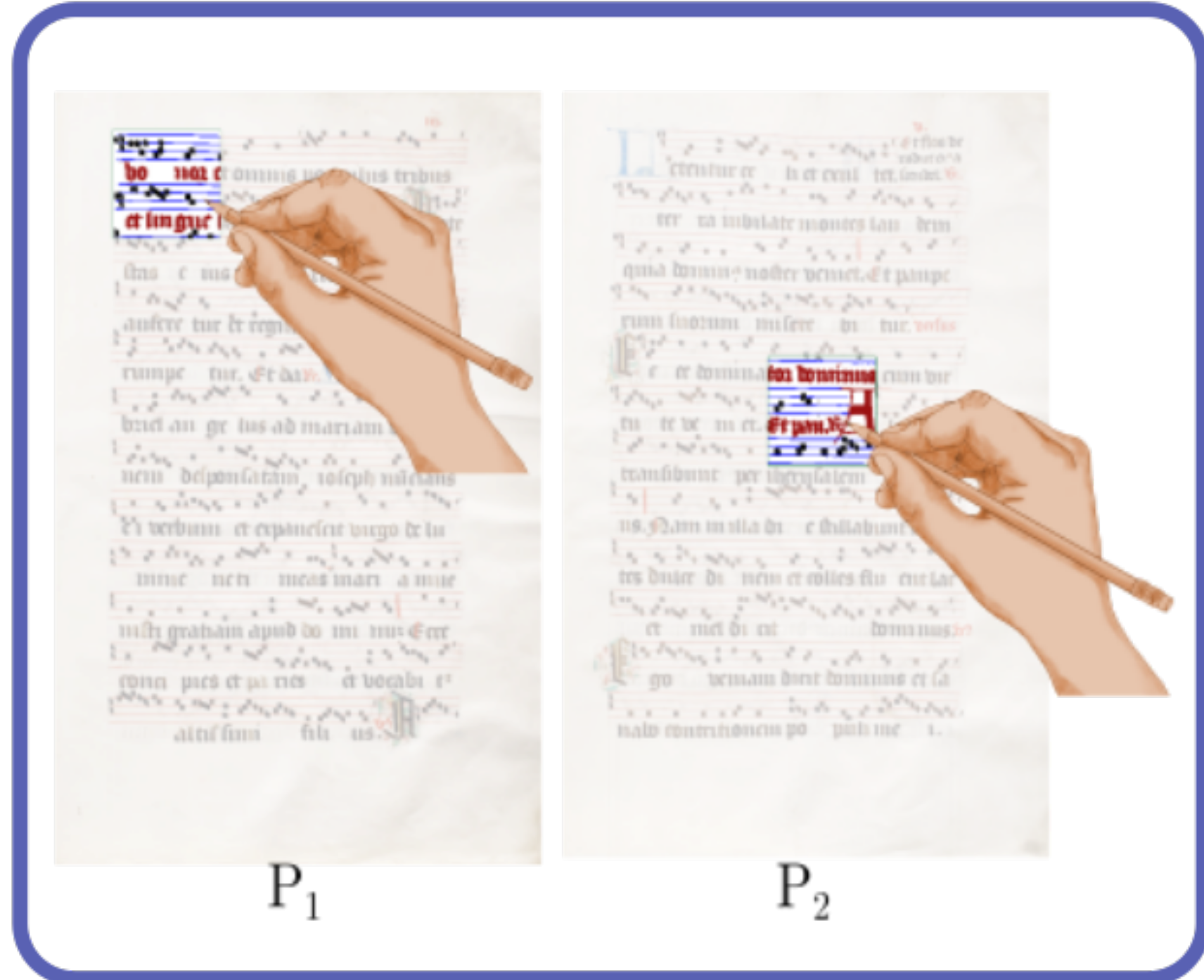
2. CURRENT STATE

- SAE-based framework:** it uses a series of supervised U-net networks, the so-called *Selectional Auto-Encoders* (SAE).
 - It requires labeled data for each new manuscript.
- Few-shot learning:** strategy in which scarce annotated data is employed to learn the task at issue.

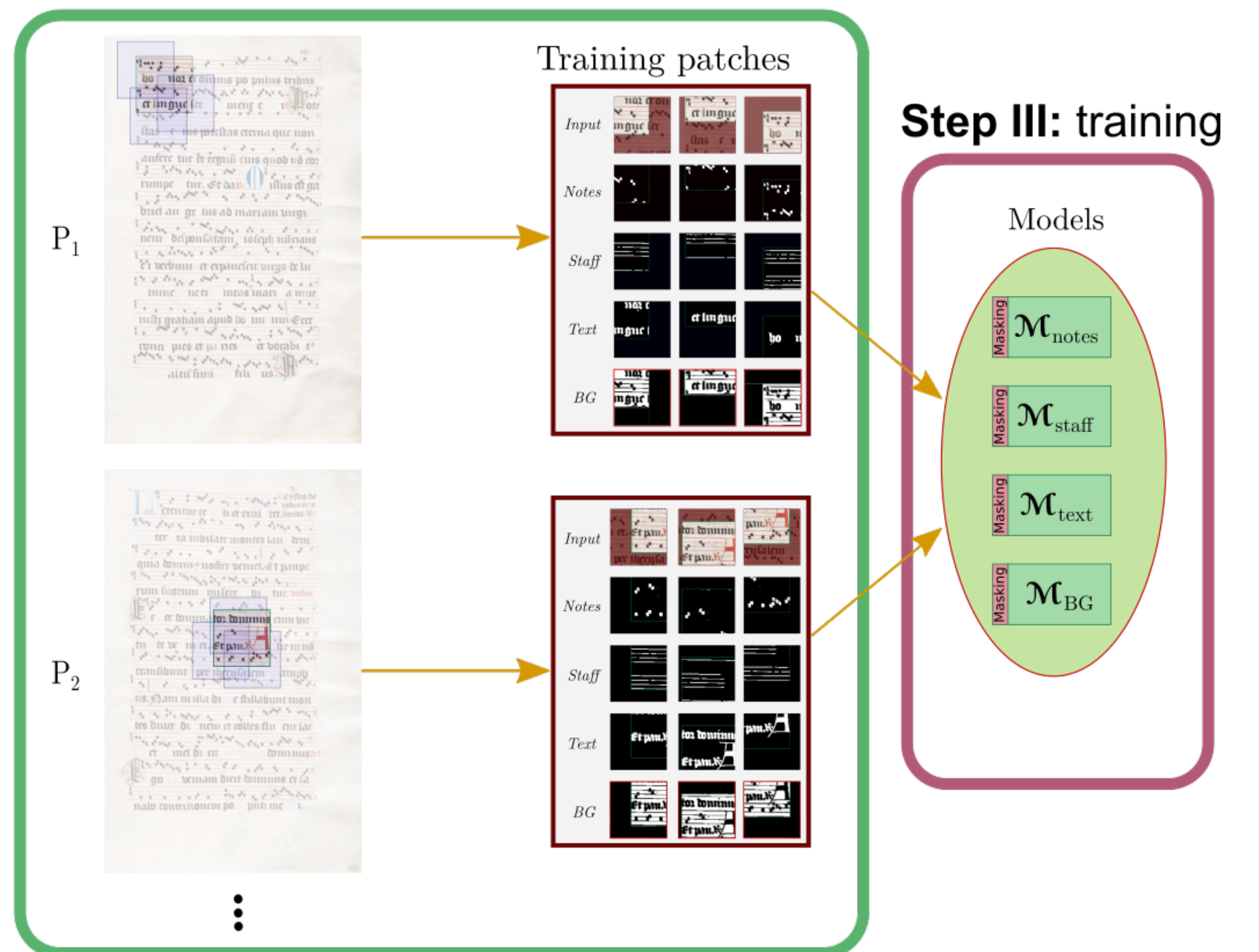
3. FEW-SHOT LEARNING FOR LAYOUT ANALYSIS

- Our approach uses partial annotations and **extracts random patch samples** around the available annotations.
- Our model **includes a masking layer** to ignore those pixels not annotated within the random patch samples employed for training.

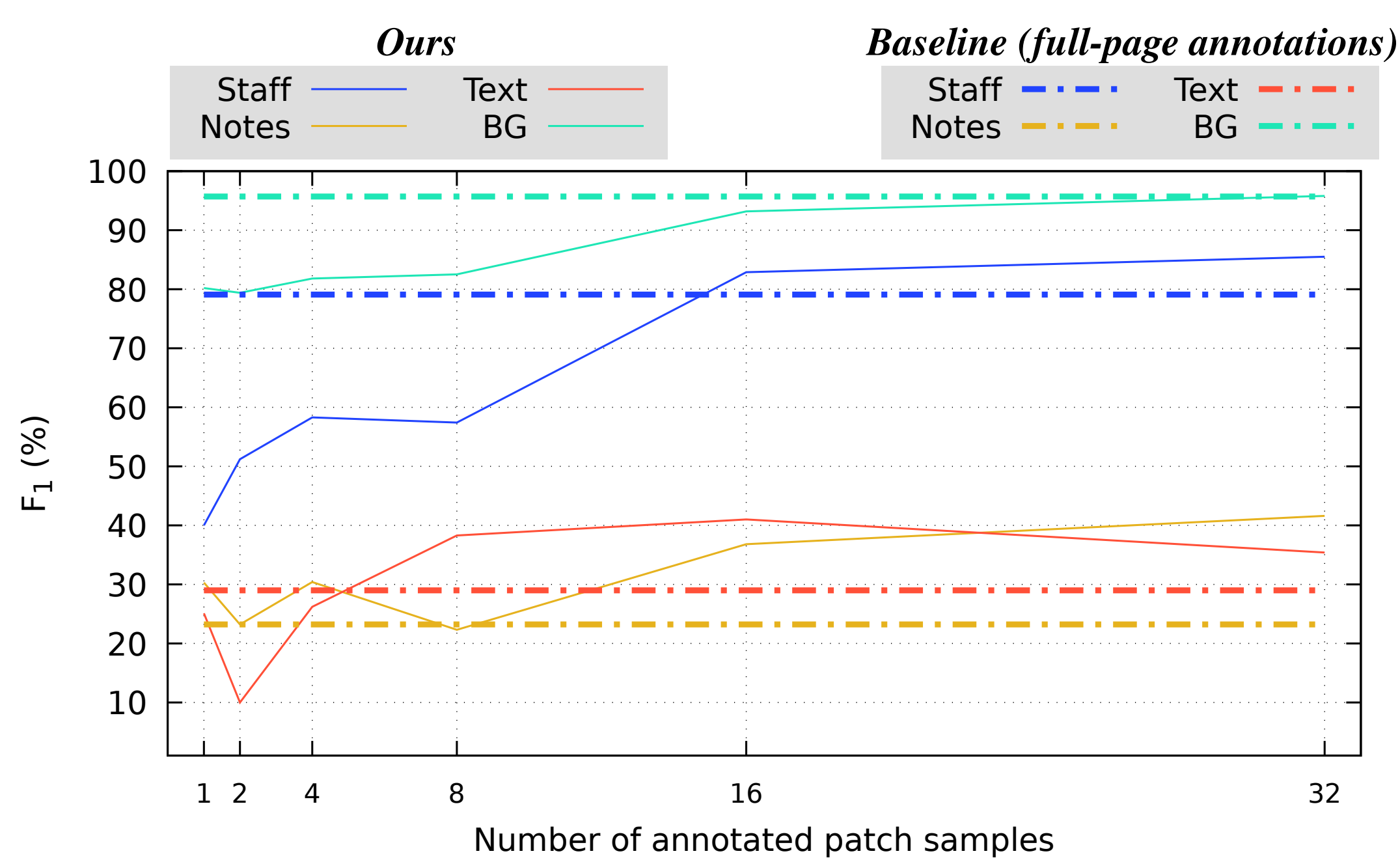
Step I: manual partial annotations



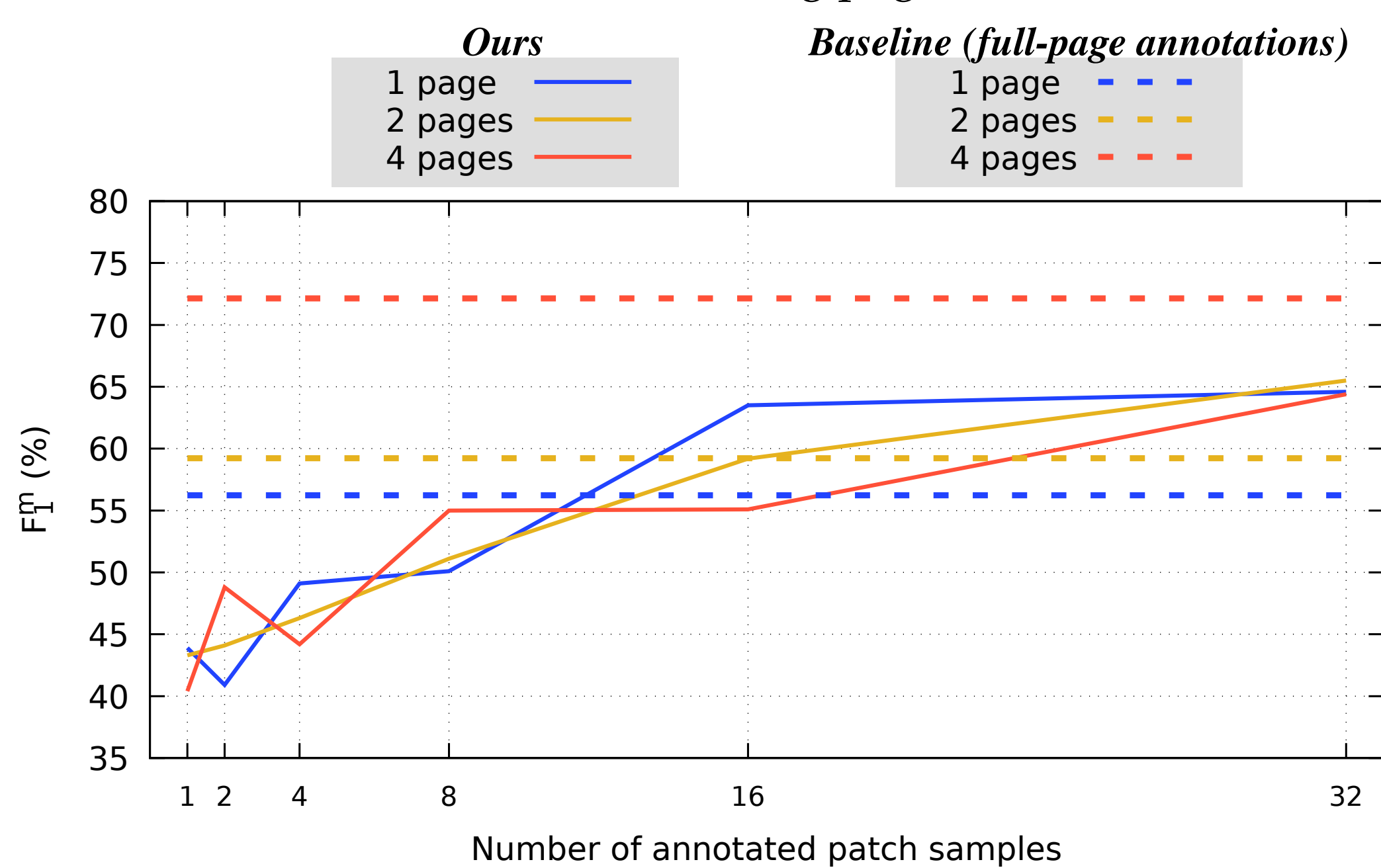
Step II: random patch-sample extraction



4. RESULTS

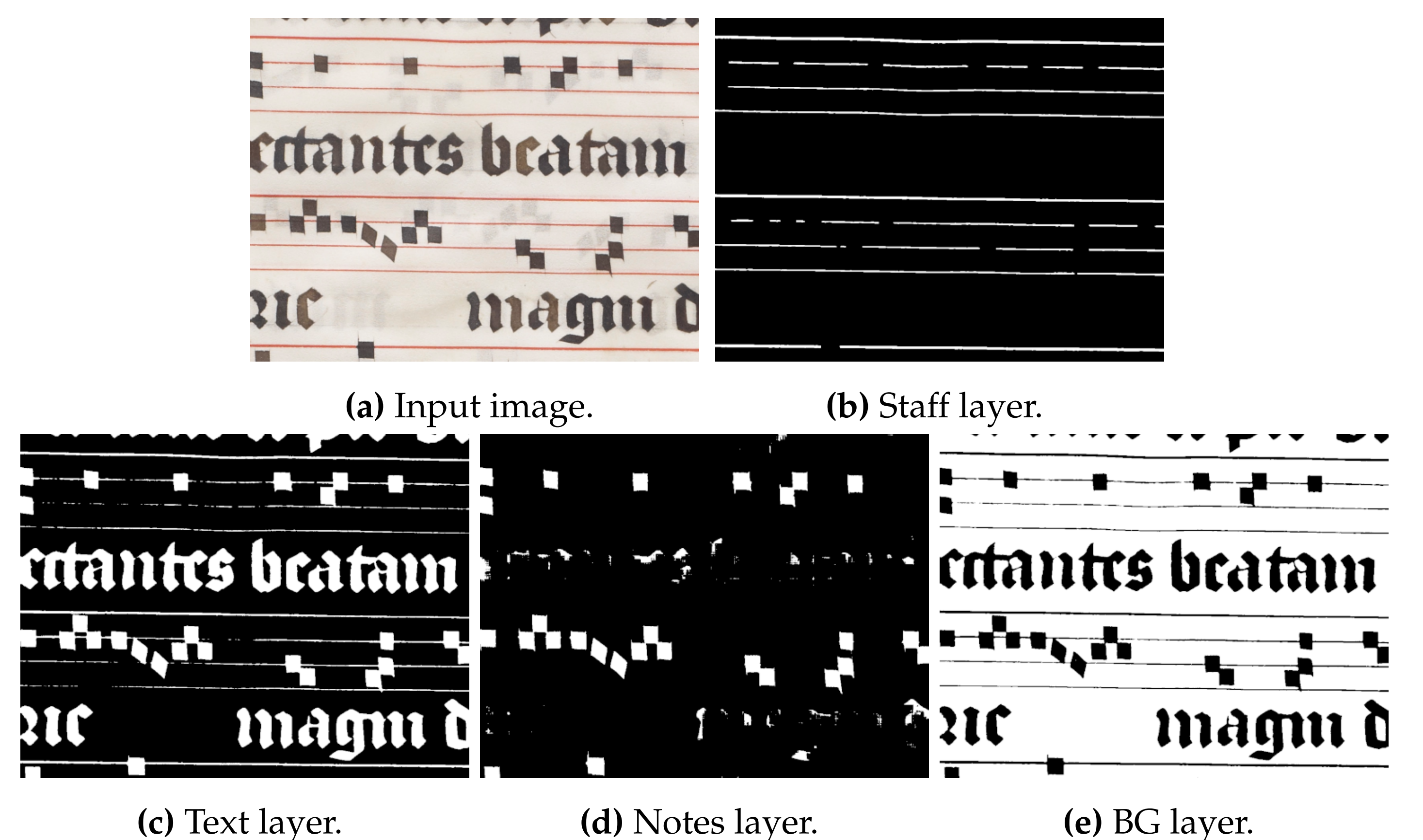


(a) Results for one training page in F-score.



(b) Results for multiple training pages in macro F-score.

5. QUALITATIVE EVALUATION



6. CONCLUSIONS

- Our approach is a potential solution for few-shot scenarios.
- It enables reducing ground-truth requirements.
- Annotating 32 patch samples in one page** yields competitive results ($F_1^M = 65.5\%$) with respect to annotating 4 full pages ($F_1^M = 72\%$).
- Transfer, incremental, and active learning may be explored.