Digital Approaches to Handwritten Text Recognition in Complex Art-Historical Archives: Towards a Robust Text Model for the Giovanni Battista Cavalcaselle Manuscripts

Jenny Palermo[1,2,*,†]

[1]Sapienza Università di Roma, piazzale Aldo Moro 5, 00185, Roma, Italy.
[2]Università degli Studi di Udine, via Palladio 8, 33100, Udine, Italy.

Abstract

This extended abstract is based on an ongoing doctoral project focused on the development of digitization and automation strategies applied to the archival holdings of the connoisseur and art historian Giovanni Battista Cavalcaselle (1819–1897), held at the Biblioteca Nazionale Marciana in Venice. The ultimate goal is to develop a system to make the contents of Cavalcaselle's legacy more easily accessible and searchable online, while also aiming to outline a replicable model useful for the valorization, promotion, and preservation of the historical memory of other complex and heterogeneous collections. The focus of this paper is, in particular, on the development of a handwritten text recognition HTR-based OCR model for the recognition and subsequent automatic transcription of the manuscript texts. The main challenge lies in the extreme heterogeneity and visual-textual complexity of the Cavalcaselle's legacy: pages often feature rapid cursive writing, marginal notes, sketches, corrections, abbreviations, and faded or damaged areas, thus challenging conventional OCR approaches. We then detail the development of a customized HTR model using the Transkribus platform, which currently achieves a character error rate (CER) of 5.71%. In addition, a brief exploratory test with a generative multimodal vision-language model (Qwen-VL) was carried out as a contextual probe rather than a systematic benchmark. We also describe our ongoing efforts to build a robust and generalizable text recognition model, capable of interpreting the most complex folios. This work aims to contribute to the broader field of digital humanities by proposing a methodologically sound, AI-assisted framework for unlocking complex, multimodal historical archives.

Keywords

handwritten text recognition, optical character recognition, Transkribus, Giovanni Battista Cavalcaselle, archival digitalization, manuscripts

1. Introduction

Giovanni Battista Cavalcaselle was one of the most influential art historians and connoisseurs of the 19th century. For all biographical information on Cavalcaselle, refer to Levi's monograph [1]. His legacy was donated to the Biblioteca Nazionale Marciana in Venice in 1904 by Angela Rovea, his widow. It represents a unique laboratory of art-historical practice, containing thousands of pages of notes, sketches, transcriptions, press clippings, and annotated drawings [2]. Since its full digitization and open-access release in 2019 [3], the collection has become a vital resource for research. Among the recent works, created thanks to the online archive consultation, we can mention the publications of Mazzaferro [4], Galassi [5] and Santi [6]. However, its lack of analytical indexing remains a critical barrier: researchers cannot efficiently retrieve references to specific artworks, locations, or individuals embedded in the manuscript pages. Cavalcaselle's materials are extremely diverse and particularly complex encompassing heterogeneous physical supports and highly variable textual and visual content [7]. During the investigation, the engagement with the connoisseur's handwritten notes was inevitable, as they were often written in a hasty, irregular cursive, filled with personal abbreviations, and sometimes inscribed on faded or damaged paper. Transcription thus emerged as a fundamental prerequisite not

IRCDL 2026: 22nd Conference on Information and Research Science Connecting to Digital and Library Science, February 19-20, 2026, Modena, Italy
jenny.palermo@uniroma1.it;jenny.palermo@uniud.it (J. Palermo)
0009-0007-1779-538X (J. Palermo)

only for comprehension but also for achieving the core objective of this research: the automated analytical indexing and multimodal recognition of Cavalcaselle's documentation. Given the objective difficulty of interpreting these notes, it was necessary to use advanced HTR technologies, focusing in particular on the Transkribus platform [8], [9]. In addition, a brief exploratory probe was conducted using a recent generative multimodal vision-language model (Qwen-VL), to qualitatively assess whether general-purpose AI models could complement specialized HTR pipelines in the analysis of complex manuscript materials.

2. Methodology: Building a Custom HTR Pipeline in Transkribus

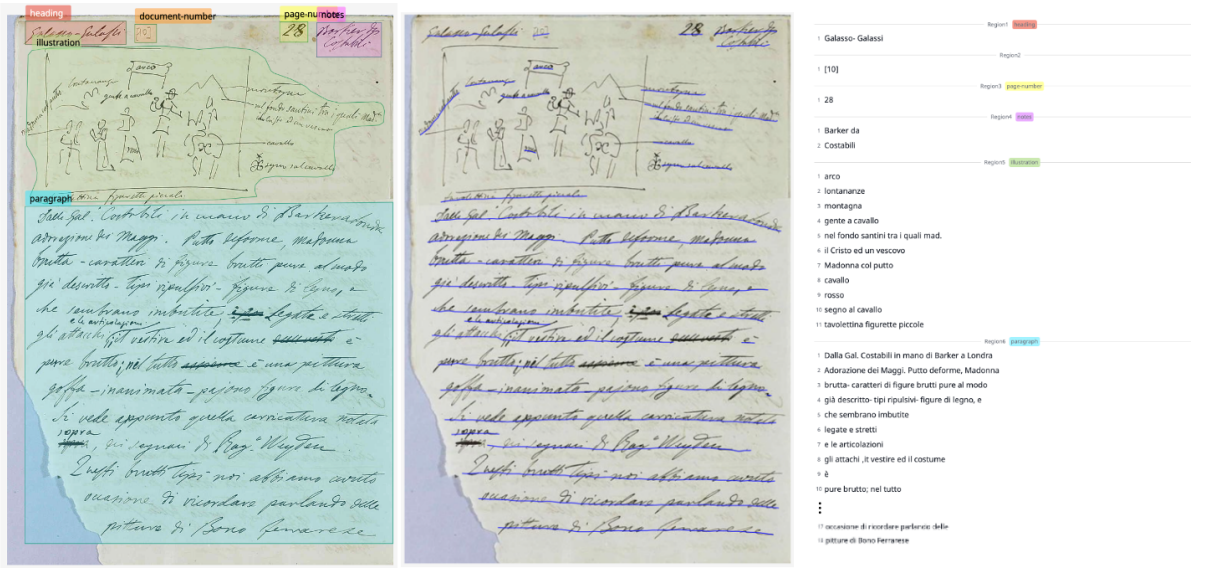
Several open-source HTR and annotation frameworks have been proposed in recent years, including Kraken integrated within the eScriptorium virtual research environment [10], [11] OCRopus-derived engines such as Calamari [12], and annotations platform such as CVAT [13]. These platforms and tools offer high configurability: they typically require local installation, GPU resources, and advanced technical expertise. Transkribus [14] was selected for this study, pertinent to the National PhD in Heritage Science, because of its integrated end-to-end workflow, combining layout analysis, baseline detection, text recognition, and model training within a single environment. Moreover, Transkribus provides a quite intuitive interface that allows scholars without extensive machine learning expertise to iteratively train and evaluate models on diverse historical material, making it particularly suitable for exploratory research on complex art-historical archives.

Transkribus is a software platform developed as part of the EU-funded READ and tranSkriptorium projects [15], [16] designed for the transcription and analysis of historical documents based on artificial intelligence. It relies on machine learning to train models that recognize the handwriting of specific authors, periods, or document types. The combination of an intuitive interface and advanced features make it a valuable tool for the digital humanities, suitable for both individual and collaborative research, capable of delivering satisfactory results even for users without advanced computer skills [17]. It is no coincidence that in a 2022 article Nockels, Gooding, Ames and Terras state that «(HTR) technology is now a mature machine learning tool, becoming integrated in the digitization processes of libraries and archives, speeding up the transcription of primary sources and facilitating full text searching and analysis of historical texts at scale. [...] 381 papers from 2015 to 2020 were gathered from Google Scholar, Scopus, and Web of Science, then grouped and coded into categories using quantitative and qualitative approaches. Published research that mentions Transkribus is international and rapidly growing. Transkribus features primarily in archival and library science publications, while a long tail of broad and eclectic disciplines, including history, computer science, citizen science, law and education, demonstrate the wider applicability of the tool. The most common paper categories were humanities applications (67%), technological (25%), users (5%) and tutorials (3%) » [18].

For the processing of Cavalcaselle's manuscripts, we implemented a three-step workflow. In the following, the terminology used by the Transkribus platform (Field Model, Baseline Model, and Text Model) is retained for clarity, but each component corresponds to well-established stages in HTR pipelines, namely layout segmentation, baseline detection, and handwritten text recognition.

2.1. Layout Segmentation (Field Model)

The first step of the workflow consists of layout segmentation, implemented in Transkribus through a so-called Field Model. This model is trained to identify and classify distinct layout elements, called regions, within each manuscript page, such as main text blocks, marginal annotations, page numbers and non-textual elements (e.g. sketches and drawings) (Figure 1). Moreover, the model is trained not only to detect layout regions, but also to automatically assign a specific tag (structural tag) to each region (e.g. illustration, page-number, paragraph) in order to enable the structured encoding of layout elements in a XML file, including their spatial coordinates. Furthermore, the structural tags, allow text recognition to be selectively applied to specific regions rather than to the entire page (e.g. exclusion of illustration regions).



(a) Layout Segmentation (Field Model)

(b) Baseline Detection (Baseline Model)

(c) Handwritten Text Recognition (Text Model)

Figure 1: Schematic representation of the pipeline in Transkribus, illustrating the sequential stages of layout segmentation, baseline detection, and handwritten text recognition on a Cavalcaselle manuscript folio. Source: https://fondocavalcaselle.bibliotecanazionalemarciana.it:8443/imagesF2/900/2024%2001-28/2024%2012%20jpg/2024%20%2012%20009.jpg

This step is crucial for the Cavalcaselle corpus, where textual and visual components frequently coexist. This layout segmentation model establishes a structural representation of the page that enables subsequent processing stages.

2.2. Baseline Detection (Baseline Model)

The second step of the workflow consists of baseline detection, i.e. the identification of individual lines of text within previously segmented textual regions. In Transkribus, baselines are represented as polygonal lines tracing the lower contour of each text line (Figure 1). This step is particularly critical for Cavalcaselle's manuscripts, where the spatial organization of writing frequently departs from a regular horizontal layout. In many folios, annotations are dispersed across the page, written in multiple directions, and interwoven with sketches and other visual elements, resulting in curved, oblique, or fragmented textual trajectories. Reliable baseline detection is therefore a prerequisite for the subsequent development of a robust text recognition model capable of handling non-linear manuscripts structures.

2.3. Handwritten Text Recognition (Text Model)

The final stage of the pipeline consists of handwritten text recognition (HTR), aimed at the automatic transcription of the textual content of the manuscripts. This component relies on a supervised machine-learning model trained on manually transcribed lines, aligned with detected baselines (Figure 1-c). Due to the high variability of Cavalcaselle's handwriting, an initial training set larger than the standard recommendation was required. While Transkribus typically recommends an initial corpus of 50-70 sheets to train a model on a single author, the present study relied on 178 transcribed sheets, selected to be as homogeneous as possible in terms of writing style and legibility. More complex and chaotic folios were intentionally excluded at this stage. Fine-tuning an HTR on Cavalcaselle's writing, turned out to be more complicated than expected: his handwriting varies significantly depending on the sheet, ranging from moments of greater order and clarity to others where it is very chaotic and garbled.

The HTR model was developed through an iterative training strategy. Few intermediate models were trained and subsequently applied to additional manuscript pages in order to accelerate dataset expansion: less accurate model outputs were manually corrected rather than transcribed from scratch, enabling a progressive enlargement of the training corpus. This bootstrapping approach allowed the gradual inclusion of increasingly heterogeneous material while maintaining annotation efficiency. After some training and fine-tuning cycles, the current HTR model achieves a Character Error Rate (CER) of 5.71% calculated on the validation set during the training phase. The reported text recognition model was trained on 161 handwritten pages, representing the training set, and 17 handwritten pages used for validation, for a total of 178 pages, corresponding to 19,795 words and 3,443 lines of text. The validation set therefore represents approximately 10% of the total pages (Figure 1).

3. Toward a Robust Text Model for Complex Manuscripts

Despite these advances, the current model still struggles with the most intricate sheets: those where sketches intersect with text, corrections overwrite original lines, or marginal annotations are spatially dispersed across the page. These folios represent the main challenge for this stage of the doctoral research.

To assess the model's capability, it was applied to a deliberately more challenging test set, consisting of sixteen folios characterized by dense corrections, rapid cursive writing, overlapping sketches, and material deterioration. On this complex test set, the model achieved an average character error rate (CER) of 11.07% compared to 5.71% on the validation set, highlighting both its potential and its current limitations when confronted with highly heterogeneous manuscript material.

The objective is therefore to develop a robust text recognition model capable of handling the full heterogeneity of Cavalcaselle's manuscripts, rather than only the neatest and most regular pages. This is currently being pursued by expanding the training set to include more complex pages and by fine-tuning the existing HTR model, using the current model with a 5.71% CER (calculated on the validation set) as a reference model.

In essence, this research explores the limits of current HTR technologies when applied to Cavalcaselle's dense and multimodal notational practices. Achieving reliable transcription for such material would unlock previously inaccessible content and provide a replicable methodological framework for other historical, art-historical or humanistic archives in general, characterized by similar complexity.

4. Brief exploratory test with generative multimodal AI: (Qwen-VL)

To contextualize recent developments in multimodal artificial intelligence, a brief exploratory test was conducted using Qwen-VL, a large vision-language model [19]. The model was used in a zero-shot setting on a small selection of manuscripts pages to assess its capabilities in transcription and layout understanding. This short test was intended as an exploratory probe to assess qualitative results rather than systematic benchmark or evaluation.

On relatively neat manuscript pages, Qwen produced partially readable transcriptions but frequently confused visually similar characters in Cavalcaselle's cursive handwriting (e.g., "r" vs. "n"), failed to correctly interpret deletions and overwriting, and substantially failed on complex folios where text is intertwined with sketches and marginalia, resulting in incomplete or hallucinated outputs.

However, Qwen demonstrated notable layout-awareness, producing accurate descriptions of spatial text distribution (e.g., identifying marginal and corner annotations). This suggests potential for document layout interpretation but not for reliable scholarly transcription. In contrast, the specialized and trainable architecture of Transkribus, with its explicit separation between layout segmentation and text recognition, provides the precision and controllability required for rigorous digital humanities research.

5. Conclusion

The digitization of Cavalcaselle's legacy has opened up new academic horizons, but its full potential remains locked in thousands of complex manuscript pages. The present work demonstrates that customized HTR models thanks to platforms such as Transkribus, when developed through careful, iterative and domain-informed methodology, can achieve high accuracy even on very challenging historical manuscripts.

Looking ahead, several limitations of the current approach open concrete avenues for future work. The present Text Model, although increasingly robust, still falls short when confronted with the most intricate folios, where Cavalcaselle's handwriting overlaps with sketches, rethinks, erasures and marginalia that disrupt conventional OCR workflows. A next step will therefore involve the development of a more comprehensive and "organic" text model capable of handling these highly multimodal pages. This will require expanding the training corpus to include the most challenging specimens, systematically testing the limits of Transkribus' architecture, and seeking to reduce the CER as far as possible. Should the platform prove insufficient for certain classes of documents, the project will also explore alternative HTR or multimodal AI solutions to ensure the full recoverability of Cavalcaselle's documentary legacy. Beyond the specific case, this work aims to contribute a replicable framework to unlock similar complex art-historical archives, enabling richer modes of access, analysis and scholarly interpretation.

The pursuit of an "organic" textual model represents more than a technical optimization: it is a methodological commitment to preserve the material and intellectual integrity of historical sources. By expanding the boundaries of what OCR can recognize, the goal is to transform the Cavalcaselle archive from a static digital collection into a dynamic, searchable and interconnected knowledge space.

Declaration on Generative AI

During the preparation of this work, the author used ChatGPT-5.2 in order to: Grammar and spelling check. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] Donata Levi, Cavalcaselle: il pioniere della conservazione dell'arte italiana, Barataria: Saggi, G. Einaudi, 1988.

[2] Jenny Palermo, Archivi in transizione. strategie digitali per la valorizzazione del fondo cavalcaselle, in: Le prossime sfide per i beni culturali. ricerca, competenze e professioni a fronte di cambiamenti climatici, sostenibilità e transizione digitale, volume 2025, Edizioni Arcadia Ricerche, 2025, pp. 923–927.

[3] Biblioteca Nazionale Marciana, Fondo cavalcaselle, https://fondocavalcaselle.bibliotecanazionalemarciana.it:8443/FondoCavalcaselleWeb/frame.htm, 2026. Accessed: 2026-02-06.

[4] Giovanni Mazzaferro, Il giovane Cavalcaselle, 2023.

[5] Cristina Galassi, L'occhio del conoscitore. Le ricognizioni di Cavalcaselle e le opere della Galleria Nazionale dell'Umbria nel taccuino XI della Biblioteca Marciana, volume 9, Aguaplano, 2023.

[6] Elena Santi, L'amor dell'arte è una febbre che mi aggitta continuamente. il viaggio di giovanni battista cavalcaselle in spagna (1852), Horti Hesperidum 2 (2022) 327–378.

[7] Silvia Marcon, Il fondo di giovan battista cavalcaselle a venezia, in: Giovanni Battista Cavalcaselle 1819–2019, ZeL Edizioni, Treviso, 2019, pp. 59–74.

[8] Transkribus platform, https://www.transkribus.org, 2026. Accessed: 6 February 2026.

[9] J. Palermo, Strategie digitali e sostenibilità per la conservazione consapevole del patrimonio documentario. i casi studio del fondo cavalcaselle a venezia e del fondo crowe a londra, in: Lo stato dell'arte 23, Edifir Edizioni Firenze, 2025, pp. 113–117.

[10] Kraken Development Team, Kraken htr engine, https://github.com/mittagessen/kraken, 2026. Accessed: 2026-02-06.
[11] eScriptorium Development Team, escriptorium virtual research environment, https://escriptorium.rich.ru.nl, 2026. Accessed: 2026-02-06.
[12] Calamari Development Team, Calamari ocr/htr engine, https://github.com/Calamari-OCR/calamari, 2026. Accessed: 2026-02-06.
[13] CVAT.ai, Computer vision annotation tool (cvat), https://www.cvat.ai/, 2026. Accessed: 2026-02-06.
[14] READ Consortium, Read: Recognition and enrichment of archival documents, https://eadh.org/projects/read, 2026. Accessed: 2026-02-06.
[15] tranSkriptorium Consortium, transkriptorium european project, https://www.transkriptorium.com/, 2026. Accessed: 2026-02-06.
[16] Philip Kahle, Sebastian Colutto, Günter Hackl, Günter Mühlberger, Transkribus-a service platform for transcription, recognition and retrieval of historical documents, in: 2017 14th iapr international conference on document analysis and recognition (icdar), volume 4, IEEE, 2017, pp. 19–24.
[17] Guenter Muehlberger, Louise Seaward, Melissa Terras, Sofia Ares Oliveira, Vicente Bosch, Maximilian Bryan, Sebastian Colutto, H. Déjean, M. Diem, S. Fiel, et al., Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study, Journal of documentation 75 (2019) 954–976.
[18] Joe Nockels, Paul Gooding, Sarah Ames, Melissa Terras, Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of transkribus in published research, Archival science 22 (2022) 367–392.
[19] Alibaba Cloud, Qwen chat, https://chat.qwen.ai, 2026. Accessed: 2026-02-06.