Umanistica Digitale – ISSN: 2532–8816 – n. 25, 2026: Special Issue
D. Scutelnic, A. Trovati, M. De Michiel, C. Daffara – *Reading the Unreadable: Artificial Intelligence for Text Interpretation in Degraded Manuscripts*
DOI: https://doi.org/10.60923/issn.2532-8816/25619

# Reading the Unreadable: Artificial Intelligence for Text Interpretation in Degraded Manuscripts

Dumitru Scutelnic

Department of Computer Science, University of Verona
dumitru.scutelnic@univr.it

Anna Trovati

Department of Computer Science, University of Verona
anna.trovati@gmail.com

Margherita De Michiel

Department of Legal, Language, Interpreting and Translation Sciences, University of Trieste
mdemichiel@units.it

Claudia Daffara

Department of Computer Science, University of Verona
claudia.daffara@univr.it

## Abstract

This study addresses the methodological and epistemological implications of employing artificial intelligence in the study of severely damaged manuscripts, proposing a paradigm shift from traditional Handwritten Text Recognition (HTR) systems that fail under regimes of severe material degradation. Three reference computer vision architectures are compared on the multispectral MADAM dataset: object detection (Faster R-CNN), semantic segmentation (U-Net), and instance segmentation (Mask R-CNN), evaluated on incremental training subsets ranging from 6 to 84 patches. Semantic segmentation outperforms both object detection and



Copyright © 2026 The Author(s)
The text in this work is licensed under the Creative Commons BY License.
https://creativecommons.org/licenses/by/4.0/

instance segmentation across most metrics, achieving stable performance from 24 annotated patches; conversely, Faster R-CNN and Mask R-CNN architectures require larger datasets to avoid divergence during training (Carcagnì et al. 2026). However, U-Net encounters a structural limitation in case of thin or nearly transparent strokes where residual ink is insufficient to support segmentation; under such conditions, textual completion requires the expert's judgment. This work, therefore, justifies a human-in-the-loop (HITL) approach in which algorithmic analytical capabilities and philological, linguistic, and semiotic expertise operate in a complementary manner throughout the entire recognition pipeline. A new hermeneutic paradigm, where the acronym AI alludes to a necessary collaboration between Artificial Intelligence and Artisan Intelligence – for the safeguarding of Artistic Intelligence of humankind.

**Keywords:** Multispectral imaging; damaged manuscripts; semantic segmentation; deep learning; human-in-the-loop; interepistemic paradigm; AIUCD2025

*L'articolo affronta le implicazioni metodologiche ed epistemologiche dell'impiego dell'intelligenza artificiale nello studio di manoscritti storici gravemente danneggiati, proponendo un cambio di paradigma rispetto ai sistemi tradizionali di Handwritten Text Recognition (HTR), fallimentari in simili condizioni di analisi. Tre architetture di riferimento della computer vision sono messe a confronto sul dataset multispettrale MADAM: object detection (Faster R-CNN), segmentazione semantica (U-Net) e segmentazione di istanza (Mask R-CNN), valutate su sottoinsiemi incrementali di addestramento da 6 a 84 patch. La segmentazione semantica supera l'object detection e la segmentazione di istanza secondo la maggior parte delle metriche utilizzate, raggiungendo prestazioni stabili a partire da 24 patch annotate; le architetture Faster R-CNN e Mask R-CNN, per contro, richiedono dataset più ampi per evitare che l'addestramento risulti instabile ((Carcagnì et al., 2026)). U-Net presenta tuttavia un limite nei tratti molto sottili o resi quasi trasparenti dal degrado, dove l'inchiostro residuo non è sufficiente a sostenere la classificazione del tratto; in tali condizioni, il completamento richiede il giudizio dell'esperto. La ricerca qui presentata propone un metodo human-in-the-loop in cui la capacità del modello e competenze filologiche, linguistiche e semiotiche operano in sinergia lungo l'intera pipeline di riconoscimento. Un nuovo paradigma di indagine, dove la sigla IA allude a una necessaria collaborazione tra Intelligenza Artificiale e Intelligenza Artigianale – per la salvaguardia dell'Intelligenza Artistica dell'umanità.*

**Parole chiave:** Imaging multispettrale; manoscritti degradati; segmentazione semantica; deep learning; human-in-the-loop; paradigma interepistemico; AIUCD2025

## 1. Introduction

"Manuscripts don't burn" – so goes the famous quote uttered by the Devil-Woland in M.A. Bulgakov's masterpiece *The Master and Margarita* (Bulgakov 1967, 316). Alluding to the eternal value of the Word as an expression of the self – and, fundamentally, to the eternal value of the literary word as an expression of the highest demands of civilization – this formula also appears to fit the material reality of manuscripts as concrete historical documents. "Rukopisi ne goryat": the formula itself seems to imply a "superior" challenge to the volatile nature of the substrate primarily intended for writing – paper. At the same time, it suggests the idea of a Text as the result of successive stratifications that, for a correct semantic decoding, may require the intervention of complex semiotic and hermeneutic synergies. In this sense, the analysis of manuscripts emerges as a field of inquiry endowed with an unexpected heuristic power, defining itself "at the boundary" (Bakhtin 1999, 61) of multiple humanistic disciplines (philology, linguistics, semiotics, hermeneutics) in a renewed relationship with hard sciences (computer science, optical physics, material diagnostics).

Manuscripts represent one of the most vulnerable components of cultural heritage. Traumatic events or natural degradation processes – such as flooding, combustion, surface abrasions, or the wear of time – radically alter the physical surface of the artifact. In degraded manuscripts, the text is not necessarily destroyed in its physical form; rather, the progressive reduction of visible contrast between the graphic stroke and the material background renders the text difficult to decipher with the naked eye. The recovery of such witnesses therefore requires methodologies capable of operating on the threshold of textual legibility, extracting the latent graphic signal retained by the physical medium to bridge the gap between material preservation and textual interpretation.

Over the last decade, non-invasive imaging methods, spanning from traditional multispectral imaging to recent thermal quasi-reflectography (TQR), have been used to capture data across the UV, VIS, and IR bands to uncover hidden information (Tonazzini et al. 2019; Perino et al. 2024; Cimino et al. 2023). Machine learning models extract complex textual representations from large annotated datasets, shifting character recognition from deterministic template matching to generalized statistical learning (Wang, Xiao, and Li 2021; Singh, Rohilla, and Sharma 2025). Broader surveys of deep learning applications to historical document analysis (Lombardi and Marinai 2020) and of the datasets underpinning this research (Nikolaidou et al. 2022) confirm that binarization and dataset scarcity remain persistent bottlenecks across scripts and degradation profiles. The workflow has progressed from early convolutional neural networks (CNNs) optimized for isolated character segmentation (Liu et al. 2016) to advanced transformer-based architectures, such as TrOCR (Li et al. 2021) and HTR-VT (Li et al. 2024). However, while these modern frameworks accurately process entire lines of handwritten text, their high performance remains strictly conditional on the legibility of the initial graphic signal.

Against this backdrop, a severely degraded manuscript introduces a radical domain shift from the clean datasets on which established software ecosystems and transcription platforms – such as Transkribus (Kahle et al. 2017), eScriptorium (Kiessling et al. 2019), and OCR4All (Reul et al. 2019) – are trained. Optimized for highly legible scripts, uniform inks, and homogeneous backgrounds, these systems achieve state-of-the-art accuracy only within their native domain. When the substrate is heavily affected by visual noise, including tide lines and degradation residues, and the ink contrast drops drastically, they undergo a sudden, non-linear catastrophic failure. Despite recent advancements, AI models remain strictly dependent on domain-specific training data; when encountering unique, non-reproducible degradation profiles, neural networks lack the semantic and contextual inference that allows human experts to navigate illegibility. This interpretive gap carries precise methodological implications, which are extensively discussed later.

Consequently, this domain shift necessitates the methodological repositioning proposed in this study, which occupies the intrinsically interdisciplinary crossroads of digital or computer-aided palaeography (Stokes 2009; Hassner et al. 2013). Within this framework, philological expertise lato sensu provides the necessary linguistic and historical context, diagnostic imaging uncovers visually obscured data, and machine learning processes the resulting high-dimensional datasets.

This interdisciplinary synergy underpins the research conducted within the AIPAD (Artificial Intelligence and Physics for Art Diagnostics) project (Daffara et al. 2025) and, specifically, the development of the MADAM (Manuscript Annotated Dataset based on Multispectral Imaging) dataset (Carcagnì et al. 2026). This dataset leverages high-resolution multispectral stacks in the UV-VIS-IR, advanced imaging techniques in the Mid-IR (Cimino et al. 2023), and micro-profilometry integration (Mazzocato, Cimino, and Daffara 2024) to benchmark diverse machine learning models on degraded materials. The most significant contribution represents the construction of a ground truth annotation at the individual character level, to support evaluation

of three computer vision paradigms based on the same material. Based on this implementation, this study argues that shifting the target from standard textual transcription to character-level instance segmentation reconfigures the annotation task itself. This approach evaluates various architectures that have previously seen limited application in the recognition of degraded handwritten text, while simultaneously and actively reintegrating philological expertise into the training and validation cycles.

The remainder of this paper is structured as follows. Section 2 addresses the phenomenology of degradation from the perspective of text deciphering and the limitations of semantic inference. Section 3 translates these challenges into computer vision frameworks, discussing the limitations of traditional HTR and introducing object detection, semantic segmentation, and instance segmentation alternatives. Section 4 outlines the character-level annotation methodology and the constraints of the labeling taxonomy. Section 5 presents a proof-of-concept case study comparing three computer vision paradigms: object detection, semantic segmentation, and instance segmentation under conditions of severe data scarcity, and evaluates efficiency in terms of data scale across the different architectures. Section 6 provides a comprehensive discussion of the experimental results and architectural implications, while Section 7 concludes the paper by delineating future research directions.

## 2. The challenge of degraded manuscript interpretation

Computer-aided analysis of damaged manuscripts requires decoupling three layers: the material layer (the physical substrate and ink altered by degradation), the graphic layer (the visual system of strokes, ductus, and ligatures), and the textual layer (the underlying text sequence to be reconstructed). Physical decay propagates from the material to the graphic layer, resulting in fragmented strokes and structural noise caused by background degradation. To reconstruct the text, the human reader inherently combines bottom-up perceptual data (grapheme shapes) with top-down contextual knowledge (lexical and semantic expectations), bypassing the need for a clear visual signal (Rayner 1998; McClelland and Rumelhart 1981).

Under severe degradation, top-down contextual constraints become the decisive factor in textual reconstruction; rather than linear reading, the human reader formulates probabilistic linguistic hypotheses based on residual stroke fragments and domain knowledge. However, this reliance on semantic inference presents structural limitations. When the visible graphic data is excessively truncated or ambiguous, the number of potential lexical candidates expands combinatorially, rendering the textual resolution undecidable. Furthermore, contextual inference exposes the reader to interpretive confirmation bias, where a dominant textual hypothesis is systematically prioritized over contrary physical evidence – a methodological vulnerability long recognized in critical text editing (van Lit 2019; Stokes 2009).

In addition to contextual constraints, limitations in interpreting arise from the degradation of minimal distinctive geometric features – such as stems, loops, or diacritics – due to abrasions or severe contrast loss (Figure 1). This induces three primary classes of high-risk ambiguity: the loss of secondary cross-strokes, truncated vertical ascenders or descenders (where partial erosion misaligns otherwise distinct graphemes), and the disappearance of diacritics. Conversely, well-preserved regions provide a localized dictionary of the author's calligraphic variants, guiding the reader in text decoding.

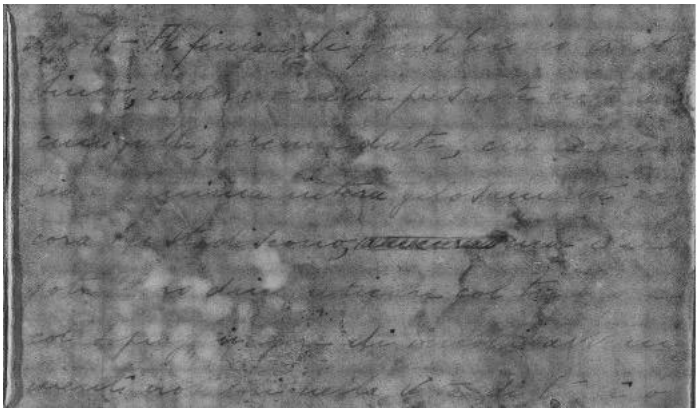


Figure 1: Critical markers distinguishing structurally similar graphemes (e.g., closed loops for a/u, full vertical strokes for l/t, and stem and tittle for i/1 in 19th- and 20th-century cursives). Under optimal legibility, these high-frequency features provide sufficient visual redundancy; under severe degradation, however, they drop below the threshold of perceptual robustness.

## 3. Bridging human perception and computer vision

### *3.1. HTR state of the art and structural limitations*

Recent advancements in Handwritten Text Recognition (HTR) focus on line-level pipelines that integrate convolutional backbones for visual feature extraction with recurrent networks or transformers for sequential decoding, trained end-to-end via Connectionist Temporal Classification (CTC) loss or encoder-decoder schemes (Wang, Xiao, and Li 2021). Transcription platforms such as Transkribus (Kahle et al. 2017), eScriptorium (Kiessling et al. 2019), and OCR4All (Reul et al. 2019) leverage these architectures, while transformer-based models like TrOCR (Li et al. 2023) and HTR-VT (Li et al. 2024) deliver state-of-the-art accuracy, provided the input signal remains clear.

However, when applied to severely degraded manuscripts, these sequential frameworks undergo a non-linear performance collapse due to the simultaneous failure of two architectural prerequisites:

[1] **Stroke-Substrate Separability**. Visual feature extraction presupposes a homogeneous background. This condition is systematically violated on surfaces where degradation artifacts share optical and spectral properties with the ink stroke.

[2] **Feature Corruption**. CTC loss (Graves et al. 2006) relies on a stable and informative feature distribution at each timestep to perform forward-backward alignment across the text line. Physical stroke fragmentation – abrasions, fading, structural lacunae – corrupts or removes this signal locally, degrading the alignment rather than the sequence structure itself.

Consequently, interventions targeting only the sequential decoder remain ineffective, necessitating upstream solutions to restore the input signal.

### *3.2. Bounding-box constraints in object detection*

An alternative approach treats characters as discrete entities using object detection architectures like Faster R-CNN (Ren et al. 2017), YOLO (Redmon et al. 2016), SSD (Liu et al. 2016), or RetinaNet (Lin et al. 2017) to predict bounding boxes and class labels (Srivastava et al. 2021). Although Baena et al. (2024) recently repositioned object detection via Detection-based Text Line Recognition (DTLR) using synthetic pre-training, detection-based pipelines have historically been marginalized due to the challenge of isolating characters with ligatures without sequential context, alongside the high burden of character-level annotation.

Furthermore, applying object detection to heavily damaged scripts presents a geometric limitation. An axis-aligned rectangular bounding box inherently encapsulates both the target grapheme stroke and its immediate local background. While this background is uniform in legible documents, in severely degraded manuscripts it contains complex visual noise, including tide lines, halos, and mold residues. The bounding box thus incorporates competing non-textual signals, leading to data ambiguity during training. This induces overfitting to specific background patterns, limits generalization, and propagates errors during inference, resulting in false positives on stains and false negatives on faint strokes. Mitigating this through dataset expansion is unfeasible due to the scarcity of comparable degradation profiles.

### *3.3. Pixel-Level reformulation: semantic and instance segmentation*

To overcome bounding-box limitations, the unit of analysis can be shifted from the character-object to the pixel level. Semantic segmentation (Long, Shelhamer, and Darrell 2015; Ronneberger, Fischer, and Brox 2015; Chen et al. 2018) assigns a class label to every pixel, classifying it as either ink stroke or substrate. Instance segmentation (He et al. 2017; Kirillov et al. 2020; Chen et al. 2019) adds an investigative layer by treating each character as an independent instance defined by its own precise mask, bounding box, and class. CNN-based pixel classification is well established in document binarization, where classic thresholding methods (Otsu 1979; Sauvola and Pietikäinen 2000) have been progressively superseded by fully convolutional and encoder-decoder architectures trained end-to-end on pixel-wise ground truth (Tensmeyer and Martinez 2017; He and Schomaker 2019; Dhali, de Wit, and Schomaker 2026), including variants specifically adapted to multispectral inputs (Hollaus, Brenner, and Sablatnig 2019). While CNN-based segmentation is established in document binarization and text line extraction (Kiessling et al. 2019; Pratikakis et al. 2018; Ntirogiannis, Gatos, and Pratikakis 2014), its use for simultaneous ink isolation and character recognition on highly degraded surfaces remains unaddressed.

This pixel-wise approach decomposes the recognition task into two distinct sub-problems:

[1] **Ink-Substrate Separation**. A low-level binary task governed by physical and optical features, such as local contrast, absorbance, and multi-spectral profiles across the UV-VIS-NIR bands.

[2] **Character Classification**. A fine-grained categorical assignment task.

Semantic segmentation isolates the ink stroke to output a clean binary mask, reverting the workflow to standard HTR pipelines and enabling transfer learning from legible corpora. Conversely, instance segmentation addresses both sub-problems jointly; it isolates the ink from the background noise and classifies each grapheme as an autonomous unit, operating independently of line-level stroke continuity.

## 4. An innovative paradigm for text annotation: from traditional OCR and HTR to instance semantic segmentation

Traditional annotation for training Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) systems relies on plain text transcriptions, digitizing text line by line and occasionally incorporating layout information. However, no further data is given about single ink strokes, whether it be their position on the page or their exact shape. This approach performs greatly when faced with well-preserved manuscripts but proves inadequate as a ground truth for the MADAM dataset (Carcagnì et al. 2026) and the objectives of the AIPAD project (Daffara et al. 2025). To maximize information density, this study adopts character-level instance segmentation utilizing pixel-wise masks and categorical classes.

A linguist with expertise in text semiotics manually annotated a high-resolution multispectral stack of 18 pages from the dataset (for a total of 108 contrast enhanced patches) using a digital drawing pad and the open-source platform Label Studio. The result was a set of pixel-accurate masks replicating the exact shape of each ink stroke, sorted into specific labelling classes. This instance segmentation ground truth inherently encompasses the other analytical approaches: it provides binary masks (separating ink from substrate noise) for semantic segmentation, and categorization via bounding boxes for object detection. This unified annotation framework enables the unbiased experimental comparison presented in Section 5. Although instance segmentation is highly resource-intensive, it represents the optimal methodology for complex, severely degraded manuscripts where standard line-level HTR models fail due to compromised substrate and ink conditions.

### *4.1. Labelling class system*

To adopt the most effective labelling system to annotate the dataset, classes to be used were defined by physicists and computer vision specialists in synergy with experts in textual semiotic and linguistic analysis. Each letter (both capital and small), each figure, each punctuation mark, each ordinal indicator was annotated as a single stand-alone class (Maheswari et al. 2024), resulting in almost 80 possible labels tailored on the Italian manuscript used as case study. Across the full corpus, 16,078 character instances were annotated (11,953 for training, 2,331 for validation, and 1,794 for testing), distributed over these 80 classes with a markedly long-tailed distribution: the most frequent classes (e.g., '.', 'a', 'I', 'o') present hundreds of occurrences, while several rare classes (e.g., 'w', 'x', 'H') are represented by a single annotated occurrence. It should be noted that to increase the model's ability to recognize ink strokes, colon, semicolon and the letter "i" were broken down into their components – dot and comma, stem and tittle respectively – as cases in which only one of the elements is readable aren't rare.

Additionally, a specific class was introduced to isolate all other ink strokes: graphic elements such as dashes after a full stop or lines separating items in lists, as well as unreadable, not uniquely identified characters. Creating a dedicated class for these additional or ambiguous cases is vital to prevent mismatches in the stroke-to-character correspondence during model training.

Particular attention was paid to ligatures and cursive script variations, where a single grapheme assumes different shapes based on its relative position and adjacent characters. Annotating ligatures as intrinsic components of the characters, preserving their differentiation, ensures a larger sample of variations within the dataset and increases the model's capability to generalize during the process of deciphering inputs, thus mitigating the challenges posed by handwriting irregularities.

More specifically, two distinct annotation levels were created, one "actual", reproducing solely the physically visible ink, the other one "reconstructed"; in the latter, minor gaps in ink stroke continuity were restored (e.g., closing letters) and characters rendered nearly transparent by degradation were annotated. Nonetheless, in both annotation levels wider lacunae representing missing letters or words were left unfilled, ensuring maximum adherence to the physical signal extracted via multispectral imaging. All interpretative decisions regarding the translation dimension of a transcription (Stevens and Burg 1997) – whether they concern text normalization, layout, or editorial intervention (e.g., modernizing spelling, expanding abbreviations, reporting inaccuracies) – were excluded at this stage. These reconstructions are deferred to a subsequent workflow phase of the project, where lexical and contextual coherence is restored through a collaborative human-machine framework coupling linguistic semiotic expertise with a Large Language Model (LLM) trained to maximize its multimodal semantic potential, following recent precedents in the linguistic and contextual restoration of lacunose ancient texts (Assael et al. 2022; Assael et al. 2025).

## 5. Proof-of-concept on a 20th-century damaged manuscript

The case study focuses on samples from an early 20th-century handwritten notebook severely compromised by flooding events. The industrial paper substrate exhibits the signs typical of this degradation, consisting mainly of structural damage in the form of planar distortion at different scales (from warping and buckling to localized cockling), differential ink fading, and a dense layering of tide lines and mold stains that alter surface reflectance. This makes the handwritten text illegible to the naked eye, causing standard OCR and HTR engines operating in the visible range to fail.

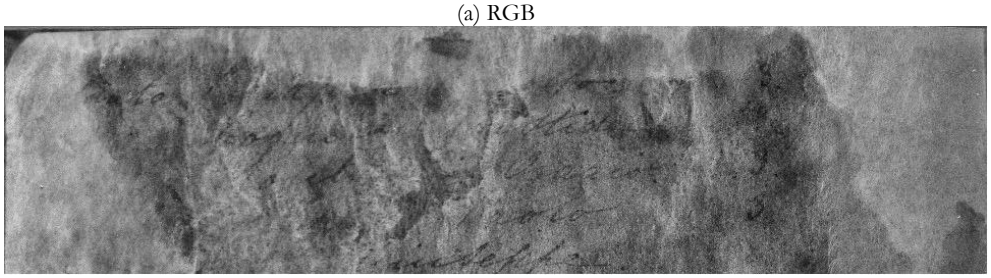
To overcome these limitations, the proposed methodology couples AI processing with multispectral data acquisition, thus enabling an analysis in the physical domain based on the distinct spectral signatures of the materials. While UV imaging is sensitive to the surface features and IR imaging probes subsurface layers, the combination of these techniques, integrated with thermal quasi-reflectography (TQR) imaging (Cimino et al. 2023), aims to maximize the information extracted from the layered manuscript.

Specifically, in the presented case study, the pipeline exploits the dataset MADAM (Carcagnì et al. 2026) of the AIPAD Project (Daffara et al. 2025), which includes multispectral patches of degraded manuscripts acquired with the Phase One imaging platform in the UV-VIS-NIR range (19 bands, up to 1 micron). Manual annotation was performed on the 660 nm band (Figure 2), where the contrast between ink and paper substrate is maximized. Because the multispectral stack is spatially registered, the geometric metadata defined at 660 nm can be re-projected across all the bands.

Every single visible character was segmented and classified; partially damaged or indecipherable characters were labelled into a special class. The resulting instance map isolates the ink signal from background noise, organizing it into discrete textual units.



(a) RGB

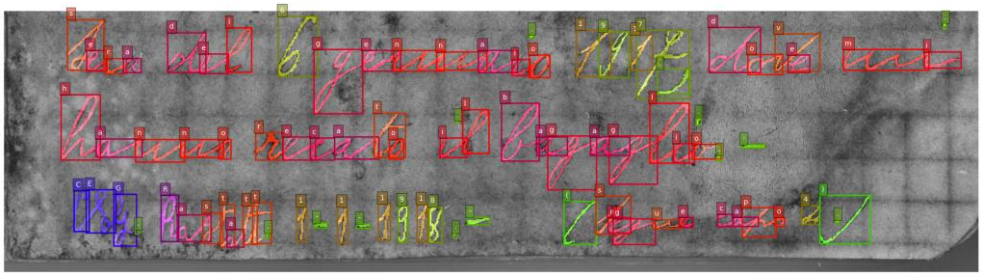


(b) 660 nm

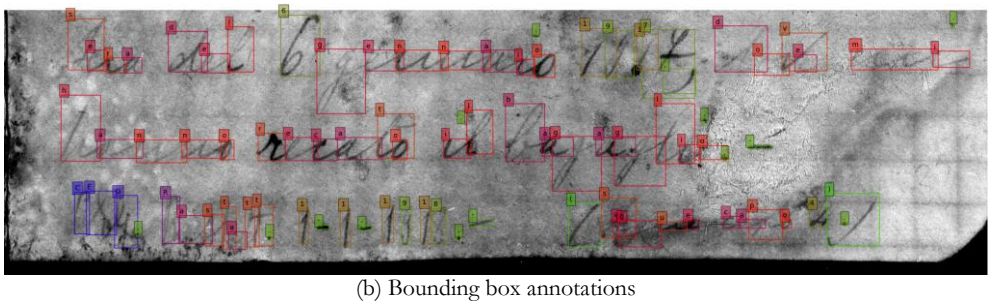
Figure 2: Exemplary patch from the severely damaged 20th-century notebook: (a) RGB VIS image; (b) 660 nm band image. The handwriting, illegible to the naked eye in the RGB VIS image, is rendered partially detectable through the multispectral imaging technique.

### *5.1. Annotation scheme*

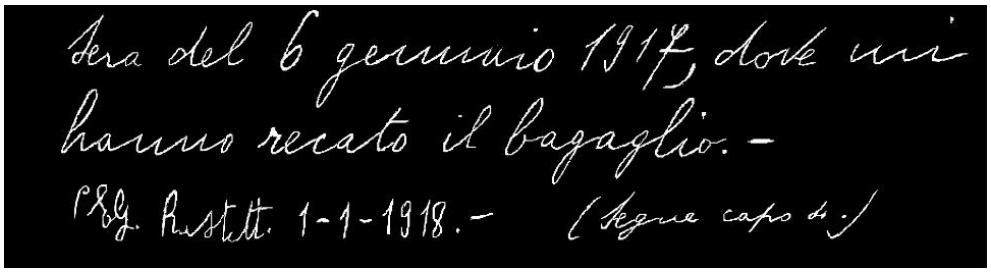
The annotation scheme was designed to enable automatic derivation of ground truth in three distinct formats, one for each of the architectural families being compared. Figure 3 (a) depicts the output of multi-class instance segmentation with pixel-wise masks: the mask colours correspond to character classes, showing the isolation of the stroke from the substrate by tracing its exact shape. Figure 3 (b) shows the same patch annotated with rectangular bounding boxes, required by object detection: the inclusion of large regions of degraded substrate within each bounding box constitutes an uncontrollable source of noise during the training phase. Figure 3 (c) shows the binarized mask required by semantic segmentation, where each pixel is labelled as either ink stroke or substrate.



(a) Multi-class pixel-wise instance annotations



(b) Bounding box annotations




(c) Binary mask

Figure 3: The three ground truth formats automatically derived from the same instance annotation: (a) multi-class pixel-wise instance masks, in which different colours correspond to different character categories and follow the morphology of the stroke; (b) axis-aligned bounding boxes for object detection, which enclose each character together with a surrounding portion of the degraded substrate; (c) binary mask for semantic segmentation, in which each pixel is labelled as either ink stroke or background.

### *5.2. Data-scale efficiency analysis*

The performance evaluation of the segmentation and detection models employs standard metrics: mean Average Precision (mAP), Intersection over Union (IoU), Recall, Precision, and the Dice coefficient for pixel-wise verification. The philological accuracy of the recovered text remains subject to expert qualitative review, distinct from segmentation quality.

The evaluated architectures are Faster R-CNN (Ren et al. 2017) for object detection, U-Net (Ronneberger, Fischer, and Brox 2015) for semantic segmentation, and Mask R-CNN (He et al. 2017) for instance segmentation. All models were trained on a single NVIDIA RTX 3090 Ti GPU (24 GB VRAM). Input images were used at their native resolution of 3584×960 px, without downscaling, to preserve the fine strokes of individual characters. Rather than fixing a uniform batch size across architectures, the largest batch size that fit in GPU memory was determined independently for each model via an empirical probe, including a worst-case validation-time pass, yielding batch sizes of 27 for Faster R-CNN, 2 for Mask R-CNN, and 1 for U-Net. Gradient accumulation was applied on top of these, over 2 steps for Faster R-CNN and Mask R-CNN, and over 4 steps for U-Net, giving effective batch sizes of 54, 4, and 4, respectively. Training patches were augmented with random contrast and brightness jitter. The learning rate for Adam optimization was scaled linearly with each model's actual effective batch size, yielding $6.75\times10^{-4}$, $5.0\times10^{-5}$, and $5.0\times10^{-5}$ for Faster R-CNN, Mask R-CNN, and U-Net, respectively. All models used a plateau-based learning-rate scheduler tracking validation performance (mAP for the two

detection models, Dice for U-Net). Training stopped early after 30 (Faster R-CNN), 25 (Mask R-CNN), or 20 (U-Net) epochs without improvement, subject to a minimum-epoch floor (50/40/20 respectively) to avoid premature termination during the flat initial plateau in validation mAP/Dice characteristic of this fine-grained, small-object task. The seed was set to 42 for reproducibility.

Faster R-CNN and Mask R-CNN both rely on a ResNet-50 backbone within a Feature Pyramid Network, pretrained on ImageNet and COCO, respectively; U-Net, by contrast, is a compact four-level encoder-decoder trained from scratch, with a single-channel output mapping each pixel to ink or substrate.

The dataset consists of 18 multispectral pages, from which the 660 nm band was selected for subsequent analysis; six patches were cropped from each page, yielding a total of 108 patches, divided into 84 for training (14 pages), 12 for validation (2 pages), and 12 for testing (2 pages), using a page-level split to avoid data leakage. Each architecture was trained on incremental subsets of $N \in \{6, 12, 24, 36, 48, 60, 72, 84\}$ patches, to characterize behaviour under varying data constraints.

As shown in Figure 4, U-Net reaches its performance plateau with few patches, whereas Faster R-CNN and Mask R-CNN require more examples before converging. This difference is consistent with the fact that, for a given annotated image, semantic segmentation provides several training examples on the order of the number of pixels in the patch. In contrast, object detection provides one example per annotated object: the ratio between the two regimes spans several orders of magnitude, leading to different saturation levels on the learning curve.

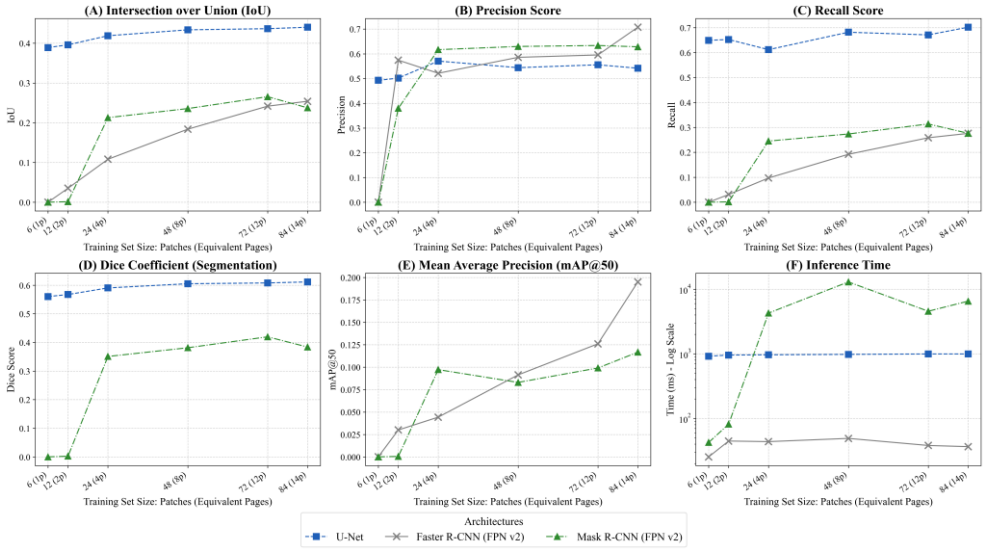



Figure 4: Learning curves and data efficiency analysis as a function of training set size N.

Table 1 reports the performance of the three models obtained at the maximum training set size considered, i.e., at N = 84 patches. The evaluation of these absolute values is subordinate to the qualitative comparison among the three approaches, which is the primary focus: specifically, the behaviour of each model in discriminating between the ink stroke and the substrate.

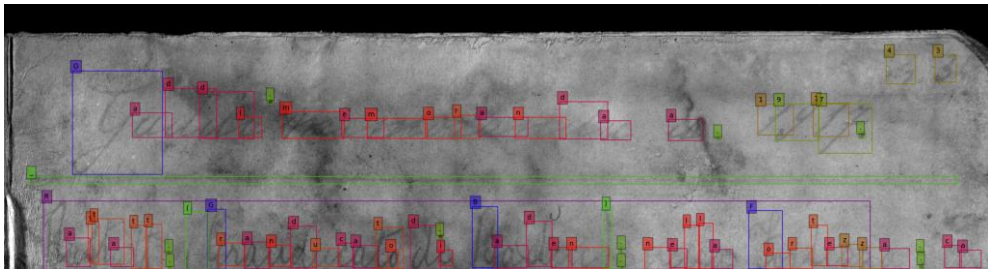
| Model | IoU | Precision | Recall | Dice | mAP@50 |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| ***Faster R-CNN*** (object detection) | 0.253 | 0.707 | 0.275 | - | 0.195 |
| ***U-Net*** (semantic segmentation) | 0.440 | 0.541 | 0.701 | 0.611 | - |
| ***Mask R-CNN*** (instance segmentation) | 0.237 | 0.628 | 0.276 | 0.383 | 0.116 |

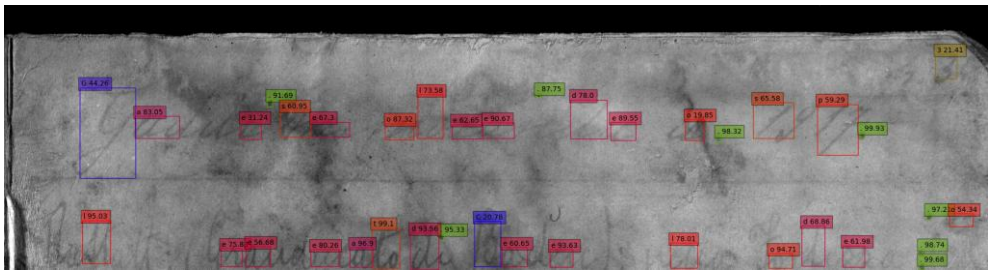
Table 1: Performance comparison of the three selected architectures at maximum training volume (84 patches of the MADAM dataset, evaluated on an independent test set of 12 patches). Average values represent precision and recall computed at the character/region level. Faster R-CNN, U-Net, and Mask R-CNN represent the object detection, semantic segmentation, and instance segmentation regimes, respectively. The evaluation is purely qualitative, focusing on a structural comparison of architectural frameworks rather than absolute performance.

With the same dataset, the semantic segmentation model (U-Net) achieves an IoU exceeding 44%, whereas object detection yields incomparably low mAP values, remaining far below operational utility. The pixel-wise approach effectively separates the stroke from the background, whereas the bounding-box approach fails to generalize (Figure 5).

False negatives occur in areas compromised by stains and sediments, where the algorithm fails to detect characters with broken strokes or highly similar strokes. False positives occur when the bounding box incorporates the degraded background, leading the network to mistake stains or abrasions for characters during training. The tests confirm that a pixel-level loss function is sufficient to separate the ink from the background in a stable, reliable manner.



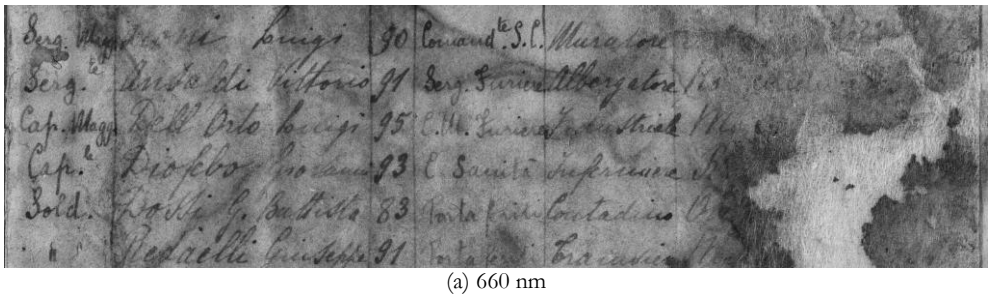
(a) Ground truth



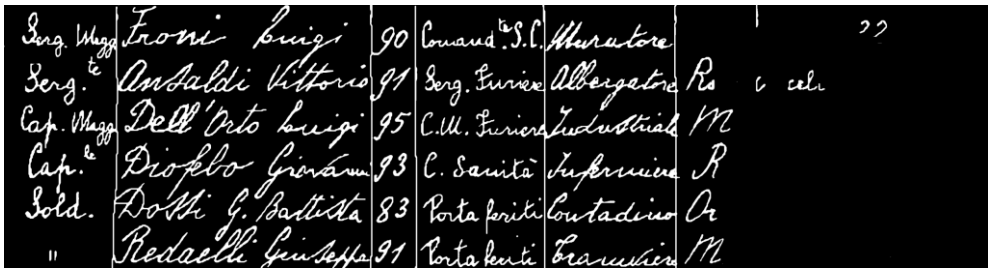
(b) Predictions

Figure 5: Result of the Faster R-CNN architecture on degraded text. Manual annotation (ground truth) (a), the predictions (b). The image shows how textural noise and the ambiguity caused by graphically similar characters mislead the model, leading to obvious false negatives (ignored characters) and false positives (spots mistaken lines).

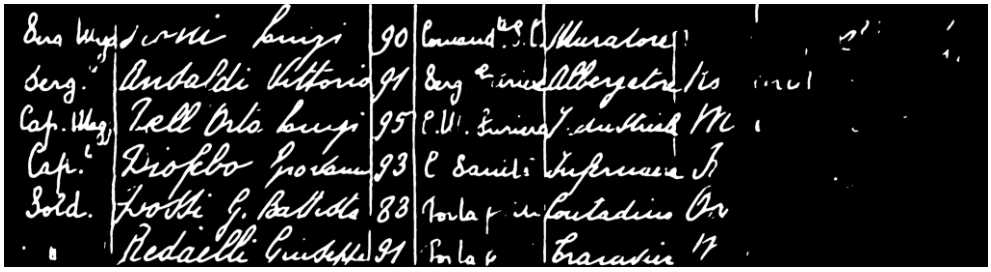
Figure 6 shows a highly degraded manuscript region, the corresponding reference manual annotation, and the U-Net predictive output. We observe that where the ink stroke thins or its spectral signature blends with the flooded support, the network disrupts the binary classification, producing a mask with discontinuities within a single character and between adjacent characters. This behaviour highlights the limitations of the purely neural approach. Full pipeline effectiveness requires a human-in-the-loop (HITL) regime (Holzinger 2016, Wu et al. 2022), in which expert intervention occurs during the refinement of masks, by reconstructing the continuity of missing strokes and during final validation, facilitating the downstream use of conventional HTR systems through transfer learning from legible handwritten text. This integrated pipeline will be the subject of subsequent publications.



(a) 660 nm



(b) Ground truth

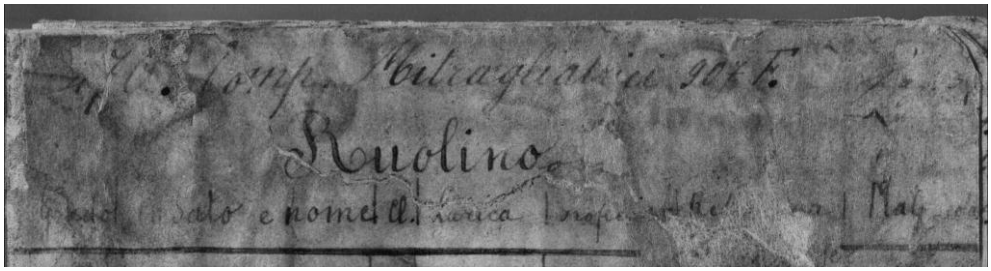


(c) Predictions

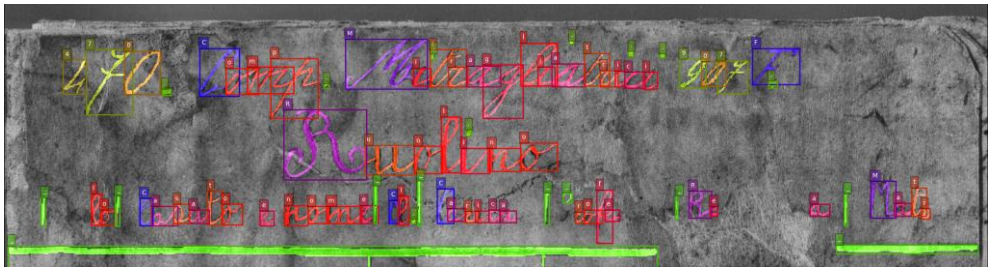
Figure 6: Qualitative analysis of pixel-wise segmentation (U-Net) on highly degraded text segments. (a) 660 nm image, (b) manually annotated ground-truth mask, (c) model-predicted mask. Note the severe fragmentation in the prediction at the faded features, compared to the continuity operated by the expert linguist based on their deductions.

Figure 7 illustrates the Mask R-CNN output on a representative patch of the manuscript, showing the 660 nm image, the instance-level ground truth, and the model prediction, with masks, bounding boxes, and confidence scores. In legible regions (e.g., the word "Ruolino" at

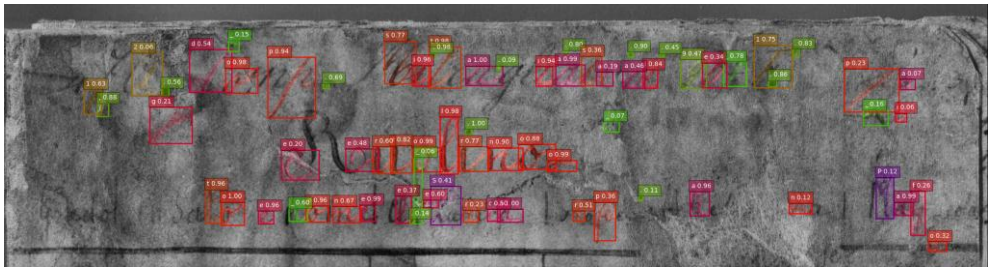
the center), predictions achieve high scores (0.90–1.00) and the masks overlap closely with the ground truth. In partially compromised regions, scores vary widely (0.30–0.90) and masks are less faithful to the contours. In regions of severe degradation, both false positives on substrate residues and false negatives on weak strokes are observed. This behaviour demonstrates the inadequacy of standard instance segmentation architectures when operating on small datasets in highly degraded contexts.



(a) 660 nm



(b) Ground truth



(c) Predictions

Figure 7: Qualitative analysis of instance segmentation (Mask R-CNN) on a damaged manuscript: (a) 660 nm image, (b) annotated ground truth, (c) predicted instances.

## 6. Discussion

The results demonstrate that semantic segmentation outperforms both object detection and instance segmentation under data-scarcity conditions. The optimal learning rate varies according to the training set size, with the three architectures exhibiting distinct responses. Notably, U-Net reaches stable performance already with few patches, as is observed consistently across the metrics.

To avoid training divergence, Faster R-CNN requires a reduced learning rate. This requirement stems from the bounding box, which, by construction, includes wide portions of degraded substrate; thus, the model struggles to discriminate text from high-variability noise on a reduced dataset. Convergence is slower and reaches lower Recall and IoU values (Table 1).

Mask R-CNN occupies an intermediate position. While its detection module inherits sensitivity to degraded substrates, its segmentation component introduces a pixel-wise loss that partially mitigates these effects. The model achieves high Precision, but lowers IoU, Recall and Dice values than U-Net, indicating that while individual character localization is accurate, the generated masks are less faithful to the actual contour of the stroke.

Qualitatively, U-Net shows a superior robustness, reaching acceptable performance at reduced training set sizes. It effectively captures the fine morphological details of the stroke where the physical signal is sufficiently consistent, facilitating the subsequent stage of textual interpretation.

A possible extension for future research is the use of generative frameworks, such as Generative Adversarial Networks (GANs) and diffusion models for reconstructing missing strokes, in line with recent work in digital heritage restoration (Cao et al. 2020; Zhao et al. 2025). However, this direction requires significantly larger datasets than those currently available in MADAM and introduces a distinct epistemological boundary: generative networks propose statistically plausible completions but cannot verify the historical accuracy of the text. Consequently, the final validation stage of the recognition pipeline strictly requires the intervention of a human expert – a philologist or palaeographer – to critically confirm the absolute correctness of the decoding process. Within this human-in-the-loop workflow, the model generates statistical proposals, while the expert retains final interpretive authority, suggesting a new reading of the "AI" paradigm through unprecedented interrelations between the applicative power of Artificial Intelligence and the philological practice of an "Artisan Intelligence", in the aim of safeguarding the Artistic Intelligence of humankind.

## 7. Conclusion and future directions

This study addresses handwritten text recognition in severely damaged manuscripts, configuring the task as a signal extraction problem under data scarcity. Utilizing the MADAM dataset, built according to an instance-based annotation scheme, three machine learning architectures (object detection, semantic segmentation, and instance segmentation) have been compared. The experimental results demonstrate that semantic segmentation, and U-Net in particular, reaches acceptable performance at reduced training scales, whereas object detection and instance segmentation architectures require larger datasets.

Future research will progress in two main areas. First, the MADAM dataset will be expanded with new case studies and will include scripts in multiple alphabets, such as Cyrillic. Second, on the algorithmic side, the multi-stage pipeline will be refined to include stroke extraction followed by a transcription stage applied to the binarized signal. Special emphasis will be placed on implementing transfer learning from pre-trained HTR models.

In conclusion, the methodological comparison of these computational frameworks carried out in a context characterised by data scarcity and high variability shows the structural limitations of each approach. The integration of philological expertise into the human-in-the-loop pipeline is

not a provisional solution awaiting more powerful models; rather, it constitutes the definitive methodological structure required in the analysis of degraded manuscript heritage.

## Acknowledgements

This research is funded by the European Union – NextGenerationEU, M4C2 component, investment 1.1, PRIN PNRR 2022 Project: «Artificial Intelligence and Physics for Art Diagnostics (AIPAD): the killer application of the historical manuscripts and palimpsests» P2022SFEPN – CUP B53D23029240001 – Grant Assignment Decree No. 1372 (1/9/2023) by the Italian MUR.

## References

Assael, Yannis, Thea Sommerschield, Brendan Shillingford, Mahyar Bordbar, John Pavlopoulos, Marita Chatzipanagiotou, Ion Androutsopoulos et al. 2022. "Restoring and Attributing Ancient Texts Using Deep Neural Networks." *Nature* 603 (7900): 280–283. https://doi.org/10.1038/s41586-022-04448-z.

Assael, Yannis, Thea Sommerschield, Alison Cooley, Brendan Shillingford, John Pavlopoulos, Priyanka Suresh, Bailey Herms et al. 2025. "Contextualizing Ancient Texts with Generative Neural Networks." *Nature* 645 (8079): 141–147. https://doi.org/10.1038/s41586-025-09292-5.

Baena, Raphael, Syrine Kalleli, and Mathieu Aubry. 2024. "General Detection-based Text Line Recognition." In *Advances in Neural Information Processing Systems* 37 *(NeurIPS 2024)*, edited by Amir Globerson, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub Tomczak, and Cheng Zhang, 42388–42404. San Diego: Neural Information Processing Systems Foundation, Inc. (NeurIPS). https://doi.org/10.52202/079017-1342.

Bakhtin, Mikhail. 1999. *Problems of Dostoevsky's Poetics*. Translated by Caryl Emerson. 8th ed. Minneapolis: University of Minnesota Press.

Bulgakov, Mikhail. 1967. *The Master and Margarita*. Translated by Michael Glenny. London: The Harvill Press.

Cao, Jianfang, Zibang Zhang, Aidi Zhao, Hongyan Cui, and Qi Zhang. 2020. "Ancient Mural Restoration Based on a Modified Generative Adversarial Network." *Heritage Science* (8): 7. https://doi.org/10.1186/s40494-020-0355-x.

Carcagnì, Pierluigi, Marco Del Coco, Marco Leo, Laura Gazzani, Marco Malagodi, Melania Paturzo, and Claudia Daffara. 2026. "MADAM: Manuscript Annotated Dataset Based on Multispectral Imaging for Handwritten Text Enhancement and Restoration." In *The Future of Heritage Science and Technologies II. Florence Heri-Tech 2024*: *Lecture Notes in Mechanical Engineering*, edited by Rocco Furferi, Lapo Governi, Yary Volpe, Francesco Gherardini, and Kate Seymour, 572–584. Cham: Springer. https://doi.org/10.1007/978-3-031-98379-5_43.

Chen, Liang-Chieh, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation." In *Computer Vision – ECCV 2018. Lecture Notes in Computer Science (vol 11211)*, edited by

Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, 833–851. Cham: Springer. https://doi.org/10.1007/978-3-030-01234-2_49.

Chen, Kai, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng et al. 2019. "Hybrid Task Cascade for Instance Segmentation." In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*: 4969–4978. Long Beach: IEEE. https://doi.org/10.1109/CVPR.2019.00511.

Cimino, Dafne, Giacomo Marchioro, Pietro De Paolis, and Claudia Daffara. 2023. "Evaluating the Integration of Thermal Quasi-reflectography in Manuscript Imaging Diagnostic Protocols to Improve Non-invasive Materials Investigation." *Journal of Cultural Heritage* (62): 72–77. https://doi.org/10.1016/j.culher.2023.04.009.

Daffara, Claudia, Nicole de Manincor, Laura Gazzani, Sara Mazzocato, Dumitru Scutelnic, Anna Trovati, Paolo Pellegrini et al. 2025. "Artificial Intelligence and Physics for Art Diagnostics: First Results from 'AIPAD' Project." In *Proc. SPIE 13569, Optics for Arts, Architecture, and Archaeology (O3A) X*, edited by Haida Liang, and Roger Groves, 135691A. Bellingham: SPIE. https://doi.org/10.1117/12.3066653.

Dhali, Maruf A., Jan Willem de Wit, and Lambert Schomaker. 2026. "BiNet: A Deep Encoder-Decoder Network for Binarizing Degraded Ancient Manuscripts." In *Document Analysis and Recognition – ICDAR 2025 Workshops. Lecture Notes in Computer Science (vol 16226)*, edited by Lianwen Jin, Richard Zanibbi, and Veronique Eglin, 166–193. Cham: Springer. https://doi.org/10.1007/978-3-032-09371-4_11.

Graves, Alex, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks." In *ICML '06: Proceedings of the 23rd international conference on Machine learning*, edited by William W. Cohen, and Andrew Moore, 369–376. New York: Association for Computing Machinery. https://doi.org/10.1145/1143844.1143891.

Hassner, Tal, Malte Rehbein, Peter A. Stokes, and Lior Wolf. 2013. "Computation and Palaeography: Potentials and Limits." *Dagstuhl Manifestos* 2 (1): 14–35. https://doi.org/10.4230/DagMan.2.1.14.

He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. "Mask R-CNN." In *2017 IEEE International Conference on Computer Vision (ICCV)*: 2980–2988. Long Beach: IEEE. https://doi.org/10.1109/ICCV.2017.322.

He, Sheng, and Lambert Schomaker. 2019. "DeepOtsu: Document Enhancement and Binarization Using Iterative Deep Learning." *Pattern Recognition* (91): 379–390. https://doi.org/10.1016/j.patcog.2019.01.025.

Hollaus, Fabian, Simon Brenner, and Robert Sablatnig. 2019. "CNN Based Binarization of MultiSpectral Document Images." In *2019 International Conference on Document Analysis and Recognition (ICDAR)*: 533–538. Long Beach: IEEE. https://doi.org/10.1109/ICDAR.2019.00091.

Holzinger, Andreas. 2016. "Interactive Machine Learning for Health Informatics: When Do We Need the Human-in-the-loop?." *Brain Informatics* 3 (2): 119–131. https://doi.org/10.1007/s40708-016-0042-6.

Kahle, Philip, Sebastian Colutto, Günter Hackl, and Günter Mühlberger. 2017. "Transkribus – A Service Platform for Transcription, Recognition and Retrieval of Historical Documents."

In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*: 19–24. Long Beach: IEEE. https://doi.org/10.1109/ICDAR.2017.307.

Kiessling, Benjamin, Robin Tissot, Peter Stokes, and Daniel Stökl Ben Ezra. 2019. "eScriptorium: An Open Source Platform for Historical Document Analysis." In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*: 19–24. Long Beach: IEEE. https://doi.org/10.1109/ICDARW.2019.10032.

Kirillov, Alexander, Yuxin Wu, Kaiming He, and Ross Girshick. 2020. "PointRend: Image Segmentation as Rendering." In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*: 9796–9805. Long Beach: IEEE. https://doi.org/10.1109/CVPR42600.2020.00982.

Li, Minghao, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang et. al. 2023. "TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models." In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, edited by Brian Williams, Yiling Chen, and Jennifer Neville, 13094–13102. Washington: AAAI Press. https://doi.org/10.1609/aaai.v37i11.26538.

Li, Yuting, Dexiong Chen, Tinglong Tang, and Xi Shen. 2024. "HTR-VT: Handwritten Text Recognition with Vision Transformer." *Pattern Recognition* (158): 110967. https://doi.org/10.1016/j.patcog.2024.110967.

Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. "Focal Loss for Dense Object Detection." In *2017 IEEE International Conference on Computer Vision (ICCV)*: 2999–3007. Long Beach: IEEE. https://doi.org/10.1109/ICCV.2017.324.

Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. 2016. "SSD: Single Shot MultiBox Detector." In *Computer Vision – ECCV 2016. Lecture Notes in Computer Science (vol 9905)*, edited by Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, 21–37. Cham: Springer. https://doi.org/10.1007/978-3-319-46448-0_2.

Lombardi, Francesco, and Simone Marinai. 2020. "Deep Learning for Historical Document Analysis and Recognition – A Survey." *Journal of Imaging* 6 (10): 110. https://doi.org/10.3390/jimaging6100110.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 3431–3440. Long Beach: IEEE. https://doi.org/10.1109/CVPR.2015.7298965.

Maheswari, S. Uma, P. Uma Maheswari, and G. R. S. Aakaash. 2024. "An Intelligent Character Segmentation System Coupled with Deep Learning Based Recognition for the Digitization of Ancient Tamil Palm Leaf Manuscripts." *Heritage Science* (12): 342. https://doi.org/10.1186/s40494-024-01438-4.

Mazzocato, Sara, Dafne Cimino, and Claudia Daffara. 2024. "Integrated Microprofilometry and Multispectral Imaging for Full-field Analysis of Ancient Manuscripts." *Journal of Cultural Heritage* (66): 110–116. https://doi.org/10.1016/j.culher.2023.11.014.

McClelland, James L., and David E. Rumelhart. 1981. "An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings." *Psychological Review* 88 (5): 375–407. https://doi.org/10.1037/0033-295X.88.5.375.

Ntirogiannis, Konstantinos, Basilis Gatos, and Ioannis Pratikakis. 2014. "ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014)." In *2014 14th International Conference on Frontiers in Handwriting Recognition*: 809–813. Long Beach: IEEE. https://doi.org/10.1109/ICFHR.2014.141.

Nikolaidou, Konstantina, Mathias Seuret, Hamam Mokayed, and Marcus Liwicki. 2022. "A Survey of Historical Document Image Datasets." *International Journal on Document Analysis and Recognition (IJDAR)* 25 (4): 305–338. https://doi.org/10.1007/s10032-022-00405-8.

Otsu, Nobuyuki. 1979. "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1): 62–66. https://doi.org/10.1109/TSMC.1979.4310076.

Perino, Michela, Lucilla Pronti, Candida Moffa, Michela Rosellini, and Anna Candida Felici. 2024. "New Frontiers in the Digital Restoration of Hidden Texts in Manuscripts: A Review of the Technical Approaches." *Heritage* 7 (2): 683–696. https://doi.org/10.3390/heritage7020034.

Pratikakis, Ioannis, Konstantinos Zagori, Panagiotis Kaddas, and Basilis Gatos. 2018. "ICFHR 2018 Competition on Handwritten Document Image Binarization (H-DIBCO 2018)." In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*: 489–493. Long Beach: IEEE. https://doi.org/10.1109/ICFHR-2018.2018.00091.

Rayner, Keith. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124 (3): 372–422. https://doi.org/10.1037/0033-2909.124.3.372.

Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. "You Only Look Once: Unified, Real-Time Object Detection." In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 779–788. Long Beach: IEEE. https://doi.org/10.1109/CVPR.2016.91.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6): 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Reul, Christian, Dennis Christ, Alexander Hartelt, Nico Balbach, Maximilian Wehner, Uwe Springmann, Christoph Wick et al. 2019. "OCR4all–An Open-Source Tool Providing a (Semi-)Automatic OCR Workflow for Historical Printings." *Applied Sciences* 9 (22): 4853. https://doi.org/10.3390/app9224853.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science (vol 9351)*, edited by Nassir Navab, Joachim Hornegger, William Wells, and Alejandro Frangi, 234–241. Cham: Springer. https://doi.org/10.1007/978-3-319-24574-4_28.

Sauvola, Jaakko, and Matti Pietikäinen. 2000. "Adaptive Document Image Binarization." *Pattern Recognition* 33 (2): 225–236. https://doi.org/10.1016/S0031-3203(99)00055-2.

Singh, Sukhdeep, Sudhir Rohilla, and Anuj Sharma. 2025. "An Inclusive Review on Deep Learning Techniques and Their Scope in Handwriting Recognition." *Electronic Letters on Computer Vision and Image Analysis* 24 (1): 51–80. https://doi.org/10.5565/rev/elcvia.1986.

Srivastava, Shrey, Amit Vishvas Divekar, Chandu Anilkumar, Ishika Naik, Ved Kulkarni, and V. Pattabiraman. 2021. "Comparative Analysis of Deep Learning Image Detection Algorithms." *Journal of Big Data* (8): 66. https://doi.org/10.1186/s40537-021-00434-w.

Stevens, Michael E., and Steven B. Burg. 1997. *Editing Historical Documents: A Handbook of Practice*. Lanham: Rowman Altamira.

Stokes, Peter A. 2009. "Computer-Aided Palaeography, Present and Future." In *Kodikologie und Paläographie im Digitalen Zeitalter / Codicology and Palaeography in the Digital Age*, edited by Malte Rehbein, Patrick Sahle, and Torsten Schassan, 309–338. Norderstedt: BoD.

Tensmeyer, Chris, and Tony Martinez. 2017. "Document Image Binarization with Fully Convolutional Neural Networks." In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*: 99–104. Long Beach: IEEE. https://doi.org/10.1109/ICDAR.2017.25.

Tonazzini, Anna, Emanuele Salerno, Zienab A. Abdel-Salam, Mohamed Abdel Harith, Luciano Marras, Asia Botto, Beatrice Campanella et al. 2019. "Analytical and Mathematical Methods for Revealing Hidden Details in Ancient Manuscripts and Paintings: A Review." *Journal of Advanced Research* (17): 31–42. https://doi.org/10.1016/j.jare.2019.01.003.

van Lit, L.W. Cornelis. 2019. *Among Digitized Manuscripts: Philology, Codicology, Paleography in a Digital World*. Leiden: Brill. https://doi.org/10.1163/9789004400351.

Wang, Yintong, Wenjie Xiao, and Shuo Li. 2021. "Offline Handwritten Text Recognition Using Deep Learning: A Review." *Journal of Physics: Conference Series* (1848): 012015. https://doi.org/10.1088/1742-6596/1848/1/012015.

Wu, Xingjiao, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. 2022. "A Survey of Human-in-the-loop for Machine Learning." *Future Generation Computer Systems* (135): 364–381. https://doi.org/10.1016/j.future.2022.05.014.

Zhao, Fahuna, Hui Ren, Zhibin Su, Xian Zhu, and Chengya Zhang. 2025. "Diffusion-based Heterogeneous Network for Ancient Mural Restoration." *Heritage Science* (13): 206. https://doi.org/10.1038/s40494-025-01719-6.