

IlluminAI: un sistema di navigazione interattivo per i manoscritti miniati rinascimentali

Valeria Minisini¹, Giorgio Gosti², Bruno Fanini³

¹ Sapienza Università di Roma, Italia – valeria.minisini@uniroma1.it

² CNR-ISPC, Italia – giorgio.gosti@cnr.it

³ CNR-ISPC, Italia – bruno.fanini@cnr.it

ABSTRACT (ITALIANO)

Nell'ultimo decennio, la digitalizzazione massiccia del patrimonio culturale ha reso consultabile una quantità inedita di volumi manoscritti, difficilmente accessibili per ragioni conservative, riversando in rete migliaia di pagine e dando nuovo impulso agli studi comparativi sull'evoluzione dei loro apparati decorativi. La conseguente necessità di sviluppare strumenti automatici e intuitivi per agevolare la ricerca di specifiche componenti ha portato a individuare nell'intelligenza artificiale un imprescindibile campo di sperimentazione non solo per il riconoscimento dei caratteri ma anche per l'analisi del layout e l'identificazione dei soggetti rappresentati nelle miniature figurative. Il contributo presenta "IlluminAI", un'interfaccia per il confronto tra manoscritti miniati basata su una rete neurale profonda addestrata con un dataset originale al fine di creare un sistema di visualizzazione immersiva per la valorizzazione dei codici rinascimentali attraverso le raffigurazioni contenute e la loro iconografia specifica. Dopo una breve introduzione per contestualizzare il progetto, verrà descritto il modello con particolare attenzione ai dati e alla tecnica usata per l'addestramento prima di esporre i risultati finora ottenuti. Concluderemo quindi presentando il sistema di visualizzazione in fase di elaborazione e i futuri sviluppi previsti.

Parole chiave: transfer learning; reti neurali profonde; manoscritti miniati; iconografia; realtà virtuale

ABSTRACT (ENGLISH)

IlluminAI: an interactive navigation system for Renaissance illuminated manuscripts

In the last decade, the massive cultural heritage digitization has made available an unprecedented quantity of manuscript volumes that would be difficult to access for conservation reasons, pouring thousands of pages online and giving new impetus to comparative studies on the evolution of their decorative apparatuses. The resulting need to develop intuitive automatic tools to facilitate the search for specific components has led to identify artificial intelligence as an essential field of experimentation not only for character recognition but also for layout analysis and the identification of subjects represented in figurative miniatures. The paper presents "IlluminAI", a comparison interface for illuminated manuscripts based on a deep neural network trained on an original dataset with the aim of creating an immersive visualization system for the enhancement of the Renaissance codices through the illustrations contained and their specific iconography. After a brief introduction to contextualize the project, the model will be described, with particular attention to the data and the technique used for training before presenting the results obtained so far. We will conclude by exposing the visualization system under development and the project future steps.

Keywords: transfer learning; deep neural network; illuminated manuscripts; iconography; virtual reality

1. INTRODUZIONE

L'autonomia della miniatura come genere artistico è una conquista relativamente recente, consolidata dagli studi dello storico dell'arte Otto Pächt (1902-1988) e dal relativo ciclo di lezioni tenute presso l'Università di Vienna nel semestre invernale del 1967-68, in seguito pubblicate in un celeberrimo volume (Pächt, 1984). Sebbene nell'immaginario collettivo la produzione dei codici miniati sia tradizionalmente associata al periodo medievale, alcuni dei volumi più riccamente decorati giunti fino a noi risalgono al Rinascimento che segna l'ultima grande stagione della miniatura in Europa, resa fiorente da artisti specializzati e pittori come Lorenzo Monaco, Taddeo Crivelli e Beato Angelico (Alexander, 1994).

Questa particolare tipologia di bene culturale, difficilmente accessibile fisicamente a causa della fragilità dei suoi materiali, ha trovato nuova vita grazie a diversi progetti di digitalizzazione (Cappellotto, 2020) che stanno mettendo a disposizione degli studiosi e del grande pubblico un'inedita quantità di opere raramente in precedenza uscite dai depositi. L'inevitabile perdita parziale delle informazioni strettamente legate alla fisicità dell'oggetto, non replicabili nel facsimile virtuale (Campagnolo, 2024) anche quando acquisito con tecniche avanzate (Pietroni et al., 2023), è compensata dalla visibilità ottenuta su portali facilmente

raggiungibili e dotati di visualizzatori come Mirador¹ capaci di comparare esemplari ripresi da fonti differenti grazie all'ormai ampia adozione dell'*International Image Interoperability Framework* (IIIF)². Nonostante ciò, l'interesse collettivo verso i codici miniati è ancora contenuto rispetto all'attrazione esercitata da altre tipologie di opere continuando a provenire, per la maggior parte, da un numero relativamente esiguo di esperti a causa delle barriere linguistiche esistenti e della complessità insita nel decifrare la scrittura manoscritta.

Sono osservabili oggi due diverse tendenze nella digitalizzazione dei volumi, legate alla tipologia di istituzione proprietaria del bene e al suo pubblico di riferimento: se le biblioteche riproducono infatti l'intero codice, i musei si limitano a creare copie digitali solamente delle pagine più interessanti dal punto di vista artistico, ovvero quando presenti scene con figure, decorazioni e capilettera ornamentali. Mentre il primo approccio preserva l'artefatto nella sua completezza fisica e contenutistica rendendolo però un prodotto impegnativo da esplorare anche se opportunamente indicizzato, il secondo è più idoneo a catturare l'attenzione dell'utente comune che naviga quotidianamente sul web ma, per farlo, riduce drasticamente il volume alla sua sola qualità estetica.

Oltre alla componente testuale che costituisce, al tempo stesso, il principale veicolo e il maggiore ostacolo alla comunicazione dei contenuti, proprio le miniature che più affascinano il pubblico non accademico possono essere utilizzate per costruire percorsi esperienziali accessibili all'interno dei codici attraverso un linguaggio visivo padroneggiato indipendentemente dal paese di provenienza e dal grado di istruzione posseduto.

Il contributo presenta la rete neurale profonda alla base di "IlluminAI", un'interfaccia di visualizzazione immersiva in fase di sviluppo progettata per valorizzare i volumi manoscritti digitalizzati agevolando lo spostamento tra le pagine grazie al riconoscimento automatico degli elementi in esse contenuti, comparare esemplari diversi e mettere in relazione le eventuali illustrazioni presenti con riproduzioni di opere d'arte dal medesimo soggetto sfruttando la comune iconografia.

Dopo alcuni brevi cenni allo stato dell'arte, si procederà a presentare la composizione del dataset appositamente formato e il sistema di classificazione delle immagini adottato prima di esporre la tecnica elaborata per addestrare il modello attraverso un approccio di apprendimento automatico interattivo (Porter et al., 2013) impiegato per annotare rapidamente tre gruppi di immagini dalle dimensioni crescenti. Infine, nella sezione conclusiva dopo i risultati finora ottenuti verranno prospettati i futuri sviluppi.

2. DEEP LEARNING E MANOSCRITTI

L'impiego del machine learning e delle reti neurali profonde per facilitare lo studio dei documenti storici è un campo di ricerca ormai consolidato (Lombardi & Marinai, 2020) che si differenzia da altri settori della computer vision per problemi specifici, aggiungendo agli ostacoli propri del riconoscimento automatico dei caratteri quelli dell'individuazione di particolari elementi all'interno delle immagini (Esposito et al., 2008). La maggioranza dei dataset riguardanti i testi manoscritti e a stampa antichi sono formati a partire da una lingua condivisa, dal periodo di creazione e l'istituzione proprietaria delle risorse (Nikolaidou et al., 2022) qualificandosi spesso come poco rappresentativi e non performanti su materiale diverso da quello d'addestramento. A differenza di quanto accade poi con le immagini naturali, i giornali (Lee et al., 2020) e le collezioni museali (Ypsilantis et al., 2022), la dimensione è sempre relativamente contenuta con pochi esemplari completi al loro interno come nel dataset HBA (Mehri et al., 2017) formato a partire da undici volumi disponibili su Gallica, il portale digitale della Biblioteca Nazionale di Francia (BnF). Nonostante l'aumento delle risorse disponibili, i dataset possono necessitare anche di venire integrati con nuovi elementi creati sinteticamente (Rahal et al., 2024), strategia particolarmente utile nella ricerca focalizzata sulle decorazioni e sull'analisi del layout.

Oltre al riconoscimento dei caratteri (Islam & Iacob, 2023), l'applicazione più studiata con decine di sperimentazioni pubblicate, l'attenzione su elementi grafici come i capilettera ornamentali per l'estrazione delle lettere contenute (Coustaty et al., 2011) o sulle figure per la classificazione dei gesti attraverso rilevatori basati sul template (Schlecht et al., 2011) ha contribuito a spostare gradualmente l'interesse dalla parola alle decorazioni che l'accompagnano. Per quanto riguarda nello specifico le miniature, la loro diversa cromia è stata sfruttata per individuarle (Borghesani et al., 2014) e tentativi di riconoscimento

¹ Per approfondire la piattaforma open-source Mirador: <https://projectmirador.org/> (cons. 09/04/2025).

² Maggiori informazioni sul Framework IIIF sono disponibili sul sito ufficiale: <https://iiif.io/> (cons. 09/04/2025).

iconografico sono stati effettuati impiegando l'algoritmo di object detection YOLO (Manoni, 2023) con buoni risultati, soprattutto se usato contestualmente il sopracitato standard IIIF (Aouinti et al., 2022). Una piattaforma online particolarmente interessante con il medesimo scopo, inoltre, è Mandragore (Jacquot, 2018) che attinge materiale dalle collezioni della BnF e dotata, dal 2022, di una nuova versione con una buona ma tradizionale interfaccia grafica³.

Nel campo storico-artistico, tra le diverse tipologie di reti neurali (Zhao et al., 2021), l'architettura ResNet (He et al., 2016) è stata spesso utilizzata (Milani & Fraternali, 2021) per le buone performance normalmente ottenute nell'attività di etichettatura e rilevazione oltre a essere facilmente addestrabile, anche su esigue quantità di dati, in quanto la sua implementazione di funzioni residue consente una migliore propagazione degli errori.

3. IL PROGETTO

3.1 La composizione del dataset

Per rendere interattivo il sistema di visualizzazione immersiva "IlluminAI", un'apposita rete neurale profonda come Minisini et al. (2024) è stata addestrata su un dataset originale al fine di individuare le pagine manoscritte in gruppi eterogenei di riproduzioni, identificare i diversi elementi di layout al loro interno oltre a distinguere le eventuali illustrazioni presenti anche da motivi decorativi e iniziali ornamentali. A tale scopo, il training del modello è stato effettuato non soltanto su riproduzioni di codici ma considerando un più articolato repertorio contenente testi a stampa, opere d'arte per la maggior parte bidimensionali quanto oggetti frequentemente presenti nei cataloghi museali e spesso decorati anch'essi con soggetti figurativi.

In totale, il dataset si compone ad oggi di 46.014 immagini provenienti esclusivamente da database istituzionali come la Library of Congress di Washington, il Metropolitan Museum of Art di New York e la J. Paul Getty Museum Collection di Los Angeles, ma anche da realtà europee come la Staatsbibliothek di Berlino, il Catalogo Generale dei Beni Culturali italiano e la Koninklijke Bibliotheek dei Paesi Bassi la cui collezione è ampiamente visibile sul portale Europeana. Il ricorso a fonti diverse ha consentito di avere un buon grado di eterogeneità nelle condizioni di scansione e scatto nonché nella qualità delle riproduzioni, sempre comunque materiale in pubblico dominio o protetto da licenze Creative Common CC-BY e CC-BY-NC scelte in previsione del rilascio finale.

Essendo il progetto incentrato sui manoscritti, il numero di immagini a loro dedicate costituisce il 68% del totale a cui si aggiungono le riproduzioni di incunaboli e volumi a stampa, attestate sul 15%, mentre il restante 17% è formato da opere d'arte e oggetti musealizzati. Le pagine, in particolare, provengono da più di duecentotrenta volumi differenti, centotrentacinque dei quali sono stati inclusi nella loro interezza, con una media di trecento immagini ciascuno, per coprire la più ampia varietà di layout e stili decorativi possibile. Compresi in un ampio arco cronologico, esteso dal IX al XX secolo, hanno principalmente tema sacro e presentano illustrazioni sia a pagina intera che dalle dimensioni più contenute e inserite nel corpo del testo.

Ogni riproduzione è stata classificata utilizzando un sistema alfanumerico che permette di risalire rapidamente all'ente proprietario e ottenere informazioni in merito al soggetto raffigurato, al periodo storico di realizzazione quanto all'autore o, se sconosciuto, almeno alla sua area geografica d'origine. Tramite poi l'aggiunta di un prefisso relativo alla tipologia di bene digitalizzato, il materiale è stato inoltre suddiviso in due macrocategorie, volumi (P) e arte (A), quanto in dieci sottogruppi: fogli manoscritti (P.01), pagine stampate (P.02), dipinti (A.01), incisioni (A.02), disegni (A.03), sculture (A.04), vetrate (A.05), arazzi (A.06), stampe d'arte (A.07) e oggetti di natura differente (A.08) come monete, ceramiche antiche e porcellane.

Un esempio di classificazione così prodotta è "P.01_S(Mark)_14BMGetty" dove "P.01" indica la natura di foglio manoscritto, "S(Mark)" identifica il soggetto della miniatura presente come San Marco Evangelista mentre "14" è il secolo di creazione, "BM" le iniziali dell'autore e "Getty" è la collezione di provenienza. Per le pagine, la sequenza che segue la specifica della raffigurazione può essere comune a più esemplari in quanto associa in maniera univoca tutti i fogli originariamente parte dello stesso volume con il codice identificativo a esso assegnato, a sua volta collegato a un record contenente le principali informazioni bibliografiche, il permalink alla risorsa primaria e l'eventuale IIIF Manifest.

³ Il portale è raggiungibile al seguente indirizzo: <https://mandragore.bnf.fr/> (cons. 09/04/2025).

Il sistema sinora esposto ha consentito di contenere significativamente il rischio di inserire copie multiple di un'opera in casi come le incisioni e, allo stesso tempo, associare con facilità le immagini dotate di una medesima raffigurazione per aumentare i soggetti scarsamente rappresentati.

Le pagine prive di miniature sono state invece classificate sulla base della presenza di decorazioni o, in assenza di queste, di testo e spartiti musicali considerando la caratteristica dominante per bilanciare prima la composizione del dataset e, successivamente, i tre gruppi usati per addestrare il modello dalle dimensioni progressivamente crescenti di 400, 4.000 e 41.614 elementi.

3.2 L'addestramento del modello

Il primo passo nella creazione del modello è stato implementare un ciclo d'addestramento incrementale iterativo articolato in quattro fasi per limitare l'apporto umano e accelerare il processo di etichettatura che, in analoghi progetti, costituisce uno dei momenti più dispendiosi temporalmente.

Il procedimento adottato è il seguente: innanzitutto, un esperto di dominio annota o corregge le etichette di un dataset più piccolo che un esperto d'intelligenza artificiale impiega in seguito per addestrare la rete neurale utilizzando la tecnica del transfer learning e l'ottimizzazione degli iperparametri. Il modello ottenuto, quindi, viene usato per prevedere le etichette di un dataset più grande che l'esperto di dominio successivamente controlla per identificare specifici bias del sistema e, quando necessario, provvede a correggere manualmente ricorrendo alla piattaforma open-source Label Studio (Tkachenko et al., 2025) prima che il processo venga ripetuto per l'insieme successivo.

Nello specifico, inizialmente sono state annotate manualmente solo le immagini facenti parti del più contenuto gruppo da 400 esemplari per segnalare la presenza o co-presenza nelle pagine di illustrazioni (Fig), decorazioni (Deco), testo (Text), spartiti musicali (Mus) e capilettera ornamentali (Let) con un minimo di una fino a un massimo di quattro etichette per un totale di 818. L'annotazione "Fig" è stata usata per contraddistinguere anche tutte le immagini della categoria "Arte" oltre a venire assegnata alle lettere abitate e istoriate.

Quanto ottenuto è stato ripartito destinando il 70% per il training e il restante 30% al test del modello, basato sull'architettura ResNet, al quale sono stati dati come compiti nella prima sperimentazione la distinzione tra volumi e arte, fogli manoscritti e pagine a stampa oltre al rilevamento della presenza dei cinque elementi di layout sopra elencati.

La tecnica del transfer learning, ricorrendo sia al Fine Tuning su tutti i pesi (Fine-Tuning) che mantenendo le caratteristiche degli strati più profondi fisse e riaddestrando solamente l'ultimo strato completamente connesso (Fixed Features), è stata impiegata con una ResNet-18 preaddestrata su ImageNet (Russakovsky et al., 2015) scelta tenendo in considerazione le dimensioni contenute del dataset a disposizione e la tendenza dei modelli con pochi parametri a essere meno soggetti a incorrere in un regime di sovrapparametrizzazione. Sono stati anche applicati due metodi distinti di addestramento con discesa del gradiente: discesa del gradiente stocastica (SGD) e Adam. Per SGD abbiamo utilizzato un exponential learning rate scheduler con velocità di apprendimento iniziale di 0,001, momentum 0,9, dimensione dello step learning rate scheduler 7 e gamma 0,1. Nel caso di Adam, invece, è stato fissato un learning rate pari a 0,001.

In quasi tutti i casi, il Fine Tuning su tutti i pesi e SGD hanno dato i risultati migliori anche se la differenza è contenuta. Il tasso di accuratezza registrato con questa combinazione è particolarmente elevato nella distinzione delle pagine dall'arte (97%) e del materiale manoscritto da quello a stampa (95%) quanto nell'individuazione di testo (96%) e musica (97%). Il valore più basso, invece, è stato registrato per le iniziali miniate (82%) a causa principalmente dall'ambiguità intrinseca tra semplici lettere ornamentali e iniziali istoriate.

Considerati i risultati promettenti, i modelli con il più elevato tasso di accuratezza per ogni target sono stati utilizzati per prevedere automaticamente le etichette del secondo gruppo, ottenendo buoni risultati nel rilevamento di elementi visivi diversi dal testo come decorazioni e illustrazioni, soprattutto nel caso di fogli con bordi ornamentali importanti oppure occupati da scene a pagina intera. Una certa confusione, però, persiste tra le iniziali e le miniature inserite nel corpo del testo, spesso confuse dal sistema e classificate come "Let".

Su 4.000 immagini solo 1.529 (38%) hanno richiesto un intervento manuale, risparmiando diverse ore di lavoro per raggiungere un risultato di 7.204 annotazioni complessive. Confrontando le previsioni con le etichette corrette, una particolare criticità è stata rilevata per gli spartiti musicali a stampa, non riconosciuti in quanto tali, con un impatto significativo sul numero di etichette assegnate dal sistema per "Mus", solo 251 sulle 358 esatte, quanto ripercussioni su "Text", passato da 3.170 a 2.855. Anche

l'etichetta "Let" ha subito una flessione, da 966 a 933, mentre le classi "Fig" e "Deco" sono aumentate, rispettivamente da 1.488 a 1.762 e da 1.214 a 1.296.

La precisione delle previsioni è stata complessivamente considerata soddisfacente con un tasso del 97% per il riconoscimento delle miniature figurative, dello 88% per il testo, del 83% per le decorazioni ma solamente del 66% per le iniziali miniate.

Nonostante lo sbilanciamento di alcune categorie, abbiamo comunque ottenuto dei modelli con un buon livello di discriminazione e si è quindi proceduto ad addestrare nuovamente la rete neurale adottando la medesima suddivisione di 70% per il training e 30% per il test.

Mentre con il primo gruppo abbiamo potuto però mantenere il learning rate costante, nel secondo sia per SGD che per Adam è stato necessario variare il valore tra 0,00001 e 0,1 per raggiungere i risultati visibili in Figura 1. Nel caso del Fine Tuning e SGD, rispetto l'insieme precedente è stato osservato un miglioramento significativo nell'accuratezza registrata nel riconoscimento delle decorazioni (92%) e delle lettere ornamentali che, seppur rimanendo il valore più basso, raggiungono anch'esse il 91%.

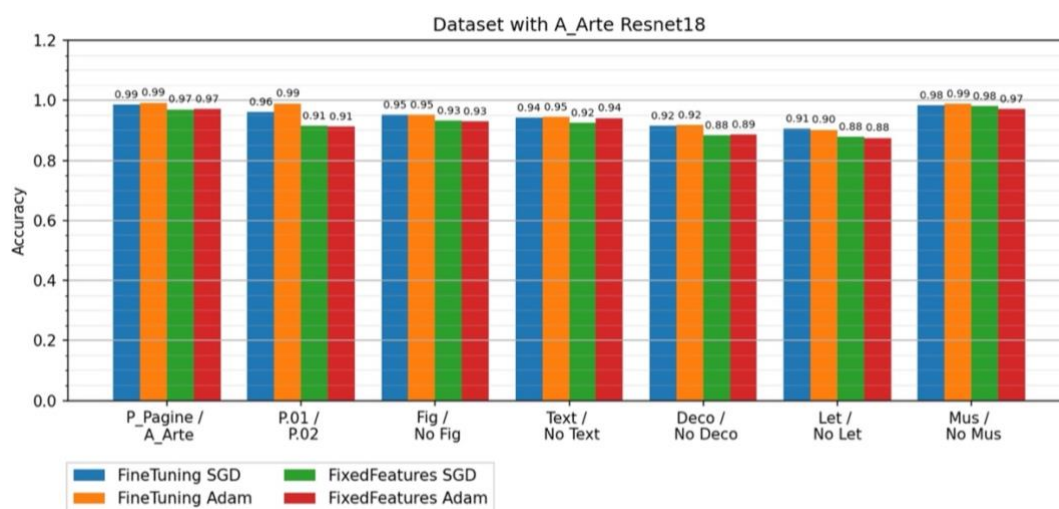


Figura 1. Grafico dei risultati ottenuti con il secondo gruppo da 4.000 immagini.

L'ultimo passaggio nel processo di addestramento ha visto ripetere quanto finora esposto per le rimanenti immagini del terzo e più consistente gruppo in cui, a differenza dei precedenti incentrati sulla figura umana, sono confluiti in numero consistente paesaggi e vedute cittadine quanto tipologie di opere quali vasi, reperti archeologici, arredi e rilegature di volumi, tutti parte della sottoclasse generica A.08.

Questa volta, solamente il 35% delle previsioni ottenute sulle 41.614 immagini analizzate è incorsa in modifica, portando a un totale di 63.677 etichette di cui 48.719 assegnate alle sole pagine manoscritte con una media di due a esemplare. Una flessione negativa tra tag previsti e corretti è stata osservata per "Let", passato con la revisione da 6.949 a 6.737, mentre per le altre etichette si è registrato un aumento con "Fig" che sale da 12.978 a 13.665, "Deco" da 9.032 a 9.217, "Text" da 29.394 a 29.428 e "Mus" da 4.604 a 4.630. Attraverso un esame approfondito dei falsi positivi è emersa la tendenza del modello a interpretare erroneamente gli elementi vegetali e le architetture quanto il persistere delle difficoltà nel riconoscere la presenza di note musicali nelle pagine a stampa oltre al testo.

Il tasso di precisione del modello, infine, si è contraddistinto per un generale miglioramento raggiungendo il 99% per il testo, le illustrazioni e gli spartiti musicali, il 96% per le decorazioni mentre per le iniziali miniate è stato registrato un più netto incremento arrivando a 88%.

4. CONCLUSIONI E FUTURI SVILUPPI

Ad oggi, la prima fase del progetto si è conclusa e i test di riconoscimento sugli elementi di layout svolti finora ci hanno dato risultati molto incoraggianti poiché livelli di accuratezza superiori all'80% non si ottengono spesso con dataset così piccoli e, ancora più raramente, vengono mantenuti incrementando l'entità e la varietà all'interno dei gruppi. La presenza di alcune categorie sottorappresentate non sembra avere impattato significativamente sui valori e, pertanto, non si è ritenuto per il momento necessario ricorrere alla data augmentation o a metodi generativi per la creazione di ulteriori esempi sintetici. L'obiettivo, una volta riaddestrato il modello per l'ultima volta con tutte le immagini etichettate, è passare gradualmente dalla classificazione alla segmentazione semantica delle miniature utilizzando le zone di

attivazione ottenute come guida per posizionare automaticamente le caselle sulle aree delle pagine da elaborare. Le maschere così ricavate saranno utilizzate per arricchire un sistema web3D immersivo e interattivo, basato sul framework open-source ATON (Fanini et al., 2021), sviluppato dal CNR-ISPC e tra i servizi del progetto H2IOSC - *Humanities and cultural Heritage Italian Open Science Cloud*, che metterà l'utente in condizione di ispezionare e interrogare spazialmente grandi quantità di pagine in un ambiente 3D virtuale, accostare immagini simili e confrontarle con altre opere di analogo soggetto attraverso l'uso di comuni web browser.

Inoltre, le riproduzioni saranno corredate da informazioni e spiegazioni per contestualizzare i manoscritti non solo a livello storico ma culturale. In questo modo intendiamo dimostrare come, attraverso l'uso delle nuove tecnologie, sia possibile elaborare strategie di valorizzazione innovative capaci di riutilizzare materiale già acquisito e rendere artefatti, originariamente prerogativa di pochi, accessibili a tutti liberando un delicato bene culturale dai suoi limiti fisici.

RINGRAZIAMENTI

Questa ricerca è finanziata dal Progetto H2IOSC - *Humanities and cultural Heritage Italian Open Science Cloud* (<https://www.h2iosc.cnr.it/>, ultimo accesso 09/04/2025), finanziato dall'Unione europea NextGenerationEU - Piano Nazionale di Ripresa e Resilienza (PNRR), Mission 4 "Education and Research", Componente 2 "From research to business", Investment 3.1 "Fund for the realization of an integrated system of research and innovation infrastructures", Action 3.1.1 "Creation of new research infrastructures strengthening of existing ones and their networking for Scientific Excellence under Horizon Europe", codice progetto IR0000029-CUP-B63C22000730005. Ente di implementazione CNR.

BIBLIOGRAFIA

- Alexander, J. G. J. (1994). *The Painted Page: Italian Renaissance Book Illumination, 1450-1550*, Prestel Pub.
- Aouinti, F., Eyharabide, V., Fresquet, X., & Billiet, F. (2022). Illumination Detection in IIIF Medieval Manuscripts Using Deep Learning. *Digital Medievalist*, 15(1), 1–18. <https://doi.org/10.16995/dm.8073>
- Borghesani, D., Grana, C., & Cucchiara, R. (2014). Miniature illustrations retrieval and innovative interaction for digital illuminated manuscripts. *Multimedia Systems*, 20, 65–79. <https://doi.org/10.1007/s00530-013-0315-3>
- Campagnolo, A. (2024). Codificare il codice. La materialità del libro e la sua digitalizzazione verso una codicologia digitale. *Umanistica Digitale*, 8(17), 1–23. <https://doi.org/10.6092/issn.2532-8816/16694>
- Cappellotto, A. (2020). From codex to apps: the medieval manuscript in the age of its digital reproduction. *Umanistica Digitale*, 4(9), 1–18. <https://doi.org/10.6092/issn.2532-8816/11459>
- Coustaty, M., Pareti, R., Vincent, N., & Ogier, J.-M. (2011). Towards historical document indexing: extraction of drop cap letters. *International Journal on Document Analysis and Recognition (IJDAR)*, 14(3), 243–254. <https://doi.org/10.1007/s10032-011-0152-x>
- Esposito, F., Ferilli, S., Basile, T. M. A., & Di Mauro, N. (2008). Machine Learning for Digital Document Processing: from Layout Analysis to Metadata Extraction. In S. Marinai & H. Fujisawa (Eds.), *Studies in Computational Intelligence: Vol. 90. Machine Learning in Document Analysis and Recognition* (pp. 105-138). https://doi.org/10.1007/978-3-540-76280-5_5
- Fanini, B., Ferdani, D., Demetrescu, E., Berto, S., & D'Annibale, E. (2021). ATON: An Open-Source Framework for Creating Immersive, Collaborative and Liquid Web-Apps for Cultural Heritage. *Applied Sciences*, 11(22), Article 11062. <https://doi.org/10.3390/app112211062>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Islam, M. A., & Jacob, I. E. (2023). Manuscripts Character Recognition Using Machine Learning and Deep Learning. *Modelling*, 4(2), 168–188. <https://doi.org/10.3390/modelling4020010>
- Jacquot, O. (2018, October 22). Mandragore, base des manuscrits enluminés de la BnF. *Carnet de recherche de la Bibliothèque nationale de France*. <https://doi.org/10.58079/m3f3>
- Lee, B. C. G., Mears, J., Jakeway, E., Ferriter, M., Adams, C., Yarasavage, N., Thomas, D., Zwaard, K., &

- Weld, D. S. (2020). The Newspaper Navigator Dataset: Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America. *CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, New York, NY, USA, 3055–3062. <https://doi.org/10.1145/3340531.3412767>
- Lombardi, F., & Marinai, S. (2020). Deep Learning for Historical Document Analysis and Recognition—A Survey. *Journal of Imaging*, 6(10), Article 110. <https://doi.org/10.3390/jimaging6100110>
- Manoni, P. (2023). AI4MSS: un esperimento di intelligenza artificiale alla Biblioteca Apostolica Vaticana. In G. Bergamin & T. Possemato (Eds.), *Guardando oltre i confini: partire dalla tradizione per costruire il futuro delle biblioteche: studi e testimonianze per i 70 anni di Mauro Guerrini* (pp. 231–244). AIB - Associazione Italiana Biblioteche. <https://doi.org/10.1400/294065>
- Mehri, M., Héroux, P., Mullot, R., Moreux, J.-P., Coüasnon, B., & Barrett, B. (2017). HBA 1.0: A Pixel-based Annotated Dataset for Historical Book Analysis. *HIP '17: Proceedings of the 4th International Workshop on Historical Document Imaging and Processing*, New York, NY, USA, 107–112. <https://doi.org/10.1145/3151509.3151528>
- Milani, F., & Fraternali, P. (2021). A Dataset and a Convolutional Model for Iconography Classification in Paintings. *Journal on Computing and Cultural Heritage*, 14(4), Article 46. <https://doi.org/10.1145/3458885>
- Minisini, V., Gosti, G., & Fanini, B. (2024). Transfer learning for Renaissance illuminated manuscripts: starting a journey from classification to interpretation. In R. Damiano, S. Ferilli, M. Striani, G. Silvello & B. Sassoli de' Bianchi (Eds.), *CEUR Workshop Proceedings: Vol. 3865. Artificial Intelligence for Cultural Heritage 2024* (pp. 30–35). https://ceur-ws.org/Vol-3865/04_paper.pdf
- Nikolaidou, K., Seuret, M., Mokayed, H., & Liwicki, M. (2022). A survey of historical document image datasets. *International Journal on Document Analysis and Recognition (IJDAR)*, 25, 305–338. <https://doi.org/10.1007/s10032-022-00405-8>
- Pächt, O. (1984). *Buchmalerei des Mittelalters: eine Einführung* (D. Thoss & U. Jenni). Prestel-Verlag.
- Pietroni, E., Chirivì, A., Fanini, B., & Bucciero, A. (2023). An Innovative Approach to Shape Information Architecture Related to Ancient Manuscripts, Through Multi-layered Virtual Ecosystems. From Codex4D to DataSpace. In L. T. De Paolis, P. Arpaia & M. Sacco (Eds.), *Lecture Notes in Computer Science: Vol. 14219. Extended Reality: International Conference, XR Salento 2023* (pp. 247–267). Springer. https://doi.org/10.1007/978-3-031-43404-4_16
- Porter, R., Theiler, J., & Hush, D. (2013). Interactive Machine Learning in Data Exploitation. *Computing in Science and Engineering*, 15(5), 12–20. <https://doi.org/10.1109/MCSE.2013.74>
- Rahal, N., Vögtlin, L., & Ingold, R. (2024). Approximate ground truth generation for semantic labeling of historical documents with minimal human effort. *International Journal on Document Analysis and Recognition (IJDAR)*, 27, 335–347. <https://doi.org/10.1007/s10032-024-00475-w>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115, 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Schlecht, J., Carqué, B., & Ommer, B. (2011). Detecting gestures in medieval images. *2011 18th IEEE International Conference on Image Processing*, Brussels, Belgium, 1285–1288. <https://doi.org/10.1109/ICIP.2011.6115669>
- Tkachenko, M., Malyuk, M., Holmanyuk, A., & Liubimov, N. (2020–2025). Label Studio: data labeling software. Retrieved from <https://github.com/HumanSignal/label-studio>
- Ypsilantis, N.-A., Garcia, N., Han, G., Ibrahimi, S., Van Noord, N., & Tolas, G. (2022). The Met Dataset: Instance-level Recognition for Artworks. arXiv preprint arXiv:2202.01747
- Zhao, W., Zhou, D., Qiu, X., & Jiang, W. (2021). Compare the performance of the models in art classification. *PLoS ONE*, 16(3). <https://doi.org/10.1371/journal.pone.0248414>