

ISBN 979-12-5965-477-9



9 791259 654779

Quaderni del Dipartimento di Linguistica  
n. 31/2023-24

Quaderni del Dipartimento di Linguistica  
Università della Calabria

n. 31/2023-24



CACUCCI  EDITORE  
BARI

---

**Quaderni del Dipartimento di Linguistica  
Università della Calabria**

---

n. 31/2023-24

**Direttore**

Roberto Guarasci, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

**Comitato scientifico**

Francesco Altimari, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Mario Caligiuri, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Carlo Spartaco Capogreco, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Mario Caterini, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Angela Costabile, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Belinda Crawford, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Ines Crispini, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Maria Mirabelli, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Luciano Romito, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Anna Rovella, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

Giuseppe Spadafora, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

**Responsabile di redazione**

Assunta Caruso, U<sub>NIVERSITÀ DELLA CALABRIA</sub>

# Quaderni del Dipartimento di Linguistica

## Università della Calabria

La rivista “Quaderni del Dipartimento di Linguistica” è una rivista scientifica inherente gli ambiti di ricerca del Dipartimento di Culture, Educazione e Società ed è censita dall’ANVUR per l’Area 10 – Scienze dell’antichità, filologico-letterarie e storico-artistiche e l’Area 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche.

Lo scopo della rivista è di creare un’occasione in cui studiosi, docenti e ricercatori possano condividere idee sulla ricerca e didattica svolta nel campo di studio proposto, attraverso uno sguardo collaborativo e innovativo. La rivista include diverse sotto-aree con l’intento di incoraggiare la condizione di una più ampia gamma di esperienze, risultati, scambi di idee. I contributi potranno essere presentati in diverse lingue allo scopo di rafforzare l’aspetto multilingue della rivista ed evidenziare la pluralità di culture e stili comunicativi con i quali la comunità universitaria entra in contatto durante la quotidianità accademica.



*Quaderni del Dipartimento di Linguistica*  
UNIVERSITÀ DELLA CALABRIA  
**31/2023-24**

**Echi di cittadinanza. Ricerca, esperienze e  
contaminazioni tra Calabria e Argentina.**

*a cura di*  
Silvia Morelli  
Anna Rovella  
Mario Francisco Benvenuto  
Rossella Michienzi



---

**PROPRIETÀ LETTERARIA RISERVATA**

---

© 2024 Cacucci Editore – Bari  
Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220  
<http://www.cacuccieditore.it> e-mail: [info@cacucci.it](mailto:info@cacucci.it)

Ai sensi della legge sui diritti d'Autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilms, registrazioni o altro, senza il consenso dell'autore e dell'editore.

# Sommario

Introduzione	9
<i>Silvia Morelli</i> , Premisa	11
<i>Maria Giuseppina Bartolo</i> , Interventi <i>Evidence-Based</i> per prevenire e contrastare il bullismo e il cyberbullismo	13
<i>Francesco Craig</i> , Autolesionismo non suicidario come manifestazioni del malessere psicologico adolescenziale	33
<i>Anna Lisa Palermi</i> , La fenomenologia del bullismo e del cyberbullying a scuola	47
<i>Natalia Forlini, Fernando Fariás</i> , El impacto subjetivo del trabajo grupal en el ámbito educativo	65
<i>Verónica Torres, Guillermo Dezorzi</i> , La adolescencia desde una concepción del sujeto	75
<i>Mario F. Benvenuto, Rossella Michienzi</i> , Escenarios inéditos de enseñanza-aprendizaje en la didáctica de Lenguas Extranjeras	95
<i>Donata Chiricò</i> , L'educazione democratica alla prova della sordità	119
<i>Paula Marini, Franco Carbonari</i> , Temporalidades, vínculos y lo escolar en el presente	133
<i>Lucia Montesanti, Francesca Veltri</i> , Politiche di protezione e processi di socializzazione dei minori non accompagnati in Italia, tra impegno civile e prospettive giuridiche	157
<i>Rocco Servidio, Martina Basilico</i> , Il gioco online problematico: uno studio esplorativo	183
<i>María Isabel Pozzo</i> , Prolegómenos para el diálogo intercultural y la comunicación científica en educación	227
<i>Giovanna Vingelli</i> , L'educazione alla sessualità	253
<i>Paula Caldo</i> , ¿Dónde están las mujeres en los archivos?	277
<i>Martin Critelli</i> , L'applicazione di reti neurali convoluzionali per l'estrazione automatica della conoscenza dei beni culturali da archivi multimediali	297
<i>Adriana Hereñú</i> , Socialización de experiencias de indagación en fuentes documentales y testimonios como enfoque metodológico en investigación educativa	309
<i>Claudia Lanza</i> , Dal riconoscimento della scrittura manuale alla categorizzazione automatica dei dati	331
<i>Anna Rovella</i> , Resilienti o persistenti?	343



# L'applicazione di reti neurali convoluzionali per l'estrazione automatica della conoscenza dei beni culturali da archivi multimediali

Martin Critelli\*

**Abstract:** The increasing application of Artificial Intelligence (AI) technologies is radically transforming the ways of analyzing multimedia files, affecting the possibilities of retrieving hidden information. These opportunities are much more evident in the field of cultural heritage, wherein the assets are characterized by a heterogeneity of elements that can be well-documented and described in archival documents. The present article aims to highlight the opportunities for applying convolutional neural networks (CNNs) on archival documents in the form of images or audiovisuals related to cultural heritage. The work aims to theoretically identify the main techniques based on CNNs to explore the content of these types of documents to automatically extract key information related to cultural heritage useful to promote knowledge discovery and dissemination services.

**Keywords:** Cultural Heritage, Knowledge Extraction, Archival Documents, Convolutional Neural Networks, Digitalization.

## 1. Introduzione

All'interno del vasto dominio culturale, gli archivi, certamente, rappresentano una fonte di primaria importanza dal momento che conservano e rendono fruibile la memoria di intere comunità e popoli. Le forme documentarie attraverso cui il sapere è custodito sono molteplici e ognuna di esse presenta proprie caratteristiche. La digitalizzazione offre sicuramente dei vantaggi facilitando l'accesso ai documenti. Negli ultimi anni abbiamo assistito ad un incremento esponenziale di progetti di digitalizzazione con la finalità di rendere disponibili via web gli oggetti del patrimonio culturale quali, ad esempio, dipinti, manoscritti, materiale librario, archivistico e altro. Il digitale è divenuto l'ambiente attraverso cui accedere, indagare ed

---

<sup>1</sup> \* Sorbonne University Abu Dhabi, SAFIR, Centre for Humanities, Language, and Education, Emirati Arabi Uniti. martin.critelli@sorbonne.ae

esplorare il patrimonio culturale. In virtù del processo di digitalizzazione, le immagini acquistano maggiore valore trasmettendo contenuti informativi che caratterizzano il patrimonio culturale, anche nella forma intangibile. Ovviamente, la rapida evoluzione delle tecnologie apre nuovi orizzonti e, ad esempio, la recente diffusione delle tecniche di AI può contribuire allo sviluppo di strumenti di estrazione e organizzazione delle informazioni basati, perlopiù, su modelli statistici e probabilistici per estrarre elementi informativi come classi di metadati, entità testuali, oggetti visivi, argomenti etc. (Xu et al. 2024). Incoraggianti sono i risultati ottenuti dall'applicazione di reti neurali convoluzionali (*Convolutional Neural Networks – CNNs*) nell'analisi delle immagini. Si tratta di una tecnologia di Deep Learning (DL) in grado di indagare il contenuto degli oggetti digitali multimediali mediante la segmentazione e il riconoscimento delle forme presenti in immagini e video.

Il presente lavoro, partendo dall'analisi delle architetture di reti neurali convoluzionali prova a evidenziare i vantaggi della loro applicazione al patrimonio culturale, anche archivistico, nei processi di estrazione della conoscenza nascosta e nella rappresentazione delle relazioni tra entità in attività di valorizzazione e disseminazione. L'articolo delinea una panoramica sulle tecniche e, soprattutto, le modalità, attraverso cui le reti neurali convoluzionali possono essere applicate a documenti d'archivio di tipo immagini o immagini in movimento.

## 2. Il funzionamento di una rete neurale convoluzionale

Una delle principali ragioni che rende le reti neurali convoluzionali così efficaci nel trattamento di immagini statiche e dinamiche è la capacità di elaborare, velocemente e contemporaneamente, i diversi input visivi di cui questi documenti si compongono. Per comprendere il funzionamento delle architetture di reti neurali e la loro possibile applicazione a file di tipo immagini o video, si è inteso fornire, di seguito, alcuni elementi descrittivi di base di tale tecnologia evidenziando le peculiarità che la contraddistinguono<sup>1</sup>.

Esaminando la struttura di base delle reti neurali convoluzionali possiamo distinguere:

- *Input layer*, il livello dei dati in ingresso da processare;
- *Hidden layer(s)*, costituiscono il livello in cui avviene l'elaborazione dei dati nelle architetture delle reti neurali;

---

<sup>1</sup> Per maggiori dettagli e approfondimenti tecnici si suggerisce la lettura di specifici manuali o articoli.

- *Output layer*, ultimo livello dell'architettura, all'interno del quale i dati provenienti dagli *Hidden layers* vengono processati per restituire il risultato finale pianificato.

Ciò che distingue le CNNs dalle altre tipologie di reti neurali come le *Recurrent Neural Network* (RNNs) sono le operazioni che hanno luogo negli *Hidden layers* dove gli input visivi vengono processati per elaborare il risultato desiderato. All'interno di tali strati è possibile individuare ulteriori *layers* caratterizzati da specifiche operazioni che consentono di processare i pixel che compongono il file multimediale per eseguire le attività di classificazione, riconoscimento etc.:

- *Convolutional layer*: l'operazione di convoluzione ha la finalità di definire una “*features map*”, ovvero una mappa delle caratteristiche, che individui tra i pixel dell'immagine in input le caratteristiche essenziali per classificarla. La componente principale della convoluzione è il kernel, la matrice di pesi che viene sovrapposta ai pixel dell'immagine per l'estrazione delle features. In ogni sovrapposizione i valori dei pixel vengono moltiplicati e sommati con quelli del kernel per ottenere in output la mappa delle caratteristiche che evidenzia i punti più rilevanti dell'immagine. Per le funzioni appena descritte, nelle architetture CNNs la convoluzione assume un ruolo core nell'analisi e nella successiva classificazione.
- *Pooling layer*: agisce sull'output del livello precedente per ridurre il numero di parametri che la rete dovrà continuare ad analizzare. Come per la convoluzione, anche in questo caso viene applicato un filtro a cui non è associato alcun peso. A seconda del tipo di pooling utilizzato, il filtro agisce in modo da individuare il pixel con il valore massimo o con il valore medio per ogni parte analizzata, che comporrà l'output di questa operazione. Uno dei vantaggi del pooling consiste nel diminuire il numero di caratteristiche da processare con la conseguente riduzione di potenza computazionale richiesta per l'elaborazione dell'immagine. La riduzione della potenza di calcolo nel processamento delle immagini costituisce una delle più importanti sfide che l'AI è chiamata a risolvere. All'interno delle CNNs non vi è limite di costruzione dei blocchi, pertanto, possono essere definiti più strati di convoluzione o di pooling, prima che entri in azione lo strato di classificazione.
- *Fully Connected Layer* (FCL): rappresenta la parte finale di ogni rete, che deve combinare tutte le caratteristiche estratte negli strati

precedenti per produrre il risultato finale. Tali strati possono essere adottati in numero variabile, in base alla tipologia di rete.

- *Loss function*: o *error function*, è una funzione che definisce l'accuratezza del modello, delineandone la capacità predittiva con la misurazione della discrepanza tra i valori predetti e i valori reali di riferimento (*groundtruth*).

Le CNNs sono in grado indagare i contenuti degli oggetti multimediali per diverse finalità. È importante sottolineare che la caratteristica di questo tipo di rete neurale è la capacità di segmentare il contenuto multimediale permettendo poi l'applicazione di specifici algoritmi per l'elaborazione delle diverse attività. I progressi registrati nello sviluppo delle reti neurali convoluzionali hanno permesso l'implementazione di modelli sempre più performanti in grado di analizzare con estrema efficacia gli oggetti multimediali attraverso tecniche quali:

- *image o video classification*: per la classificazione dell'oggetto digitale secondo categorie fornite a priori (Dhruv e Naskar 2020);
- *face detection and recognition*: per il riconoscimento di persone e in particolare del viso (Zahid et al. 2023);
- *deepfake detection*: impiegato per l'individuazione di contenuti falsi creati artificiosamente (Tan et al. 2023);
- *object detection*: per l'individuazione e la localizzazione di oggetti, animali o persone all'interno di immagini e video (Ji et al. 2021).

Ciascuna di queste tecniche è finalizzata all'esecuzione di attività complesse e produce, in termini di performance, risultati molto differenziati (Stanford University Human-Centered Artificial Intelligence 2024). Comprendere in che modo e attraverso quali tecniche le reti neurali convoluzionali processino le immagini, statiche e dinamiche, consente maggiore consapevolezza in merito all'applicabilità di tale tecnologia in un determinato contesto.

### 3. Le modalità di applicazione delle reti neurali convoluzionali su documenti d'archivio

Negli archivi analogici, i documenti digitalizzati sono trasformati in file di immagini e una delle sfide maggiori è costituita dalla possibilità di rendere le rappresentazioni di immagini statiche in formati leggibili dalle macchine riducendo il margine di errore di lettura dell'originale a tassi non rilevanti. Progettare e sviluppare sistemi in grado di riconoscere automaticamente il testo richiede, necessariamente, l'analisi di vari fattori quali, ad esempio, lo stile grafico o il materiale scrittoriale che incidono sulla qualità

dell'output. In tale contesto si parla di OCR (*Optical Character Recognition*) per indicare il riconoscimento automatico di caratteri a stampa, di HTR (*Handwritten Text Recognition*) e di ICR (*Intelligent Charachter Recognition*) riferendosi ad avanzati sistemi in grado di riconoscere la scrittura a mano. I sistemi implementati per tali finalità sono caratterizzati da un'operazione di preprocessing dell'immagine con l'obiettivo di prepararla alle fasi successive. Dal risultato di questa preliminare operazione, viene elaborata la segmentazione geometrica dell'immagine finalizzata a individuare le zone di testo da cui partire per l'estrazione delle *features* del testo e, infine, il riconoscimento dei caratteri (Sarika, Sirisala, e Velpuru 2021). All'interno di tale processo, le reti convoluzionali possono svolgere una funzione essenziale incrementando, sia sotto il profilo qualitativo che quantitativo in termini di tempo, le performance di tali sistemi.

Attualmente, i sistemi di OCR sono impiegati in innumerevoli soluzioni, sia open source che proprietarie, risultando estremamente efficaci sui diversi layout di testo (es. a colonne, direzione della scrittura etc.). Software quali Tesseract (Shafait e Smith 2010), consentono, ad esempio, una lettura automatica affidabile di immagini di testo a stampa. Diversamente il riconoscimento della scrittura a mano richiede ancora ottimizzazione per il raggiungimento di risultati soddisfacenti. Si tratta, in questo caso, di un problema estremamente complesso che deve fare i conti con diversi elementi legati alla scrittura come, ad esempio lo stile del singolo scrivente. Per questo genere di attività vengono impiegati i sistemi di HTR e ICR (AlKendi et al. 2024). Quest'ultimi, tradizionalmente basati sull'uso di specifici lessici di riferimento – *Lexicon based* – come thesauri o vocabolari per classificare le parole riconosciute all'interno del testo, prevedono ad oggi l'applicazione, anche congiunta, di *Recurrent Neural Networks* e *Convolutional Neural Networks* (CNNs). Sono attualmente presenti sul mercato diverse soluzioni software che sfruttano le tecnologie appena richiamate. Tra questi, sono particolarmente noti gli applicativi open source come *Transkribus* (Kahle et al. 2017) ed *eScriptorium* (Kiessling et al. 2019) che, attraverso una forte partecipazione della comunità, mettono a disposizione un gran numero di modelli addestrati per il riconoscimento di diversi stili di scritture e lingue. Uno degli aspetti più significativi nel riconoscimento del testo dai documenti, semplificato attraverso l'apporto di CNNs, è l'individuazione di testo presente in tabelle e grafici, largamente presenti, specie le prime, nei documenti moderni e contemporanei. Si tratta in questo caso di oggetti informativi che possono essere trasformati in dati *machine readable* e quindi processabili per estrarre conoscenza nascosta (Rovella et al. 2024).

Se in ambito archivistico l'elaborazione del testo è, ad oggi, un task che trova ampia applicabilità, meno diffuso è invece l'uso delle CNNs per l'e-

secuzione di tecniche quali *object detection* e *video activity recognition*. Tali tecniche, utilizzate anche in altri settori dei beni culturali relativamente alla classificazione di oggetti come, ad esempio, manufatti o elementi architettonici (Cosovic e Jankovic 2020; Zhou et al. 2018), all'individuazione di punti di deterioramento in siti culturali (Mishra e Lourenço 2024) o nel riconoscimento dello stile artistico nei dipinti (Smirnov e Eguizabal 2018), possono essere potenzialmente impiegate in archivio per il trattamento di immagini e audiovideo per l'estrazione di oggetti informativi. Negli archivi, la combinazione di tecniche per l'analisi dei contenuti può risultare particolarmente efficace nella lettura di una tipologia di documenti, come quelli multimediali.

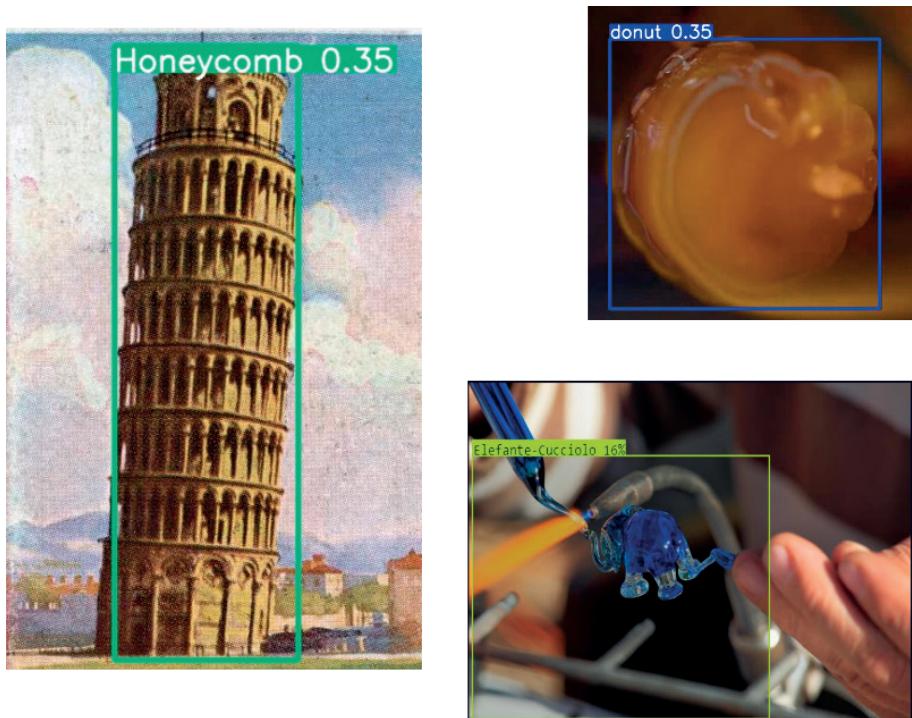
Nel caso specifico delle tecniche di *object detection*, applicabili sia a immagini che video, le reti neurali convoluzionali permettono di individuare persone, oggetti o animali immersi negli scenari più diversi e posti in condizione di scarsa visibilità. I risultati ottenuti in questo genere di attività (Stanford University Human-Centered Artificial Intelligence 2024) consentono di delineare le modalità attraverso cui applicare le CNNs allo scopo di individuare gli elementi che connotano il patrimonio culturale. A tal scopo, dunque, le tecniche di *object detection* possono essere finalizzate all'individuazione di oggetti particolari, o comuni, come strumenti musicali, artigianato, prodotti enogastronomici e manifatturieri tipici. Con riferimento al patrimonio culturale italiano, è possibile menzionare gli strumenti di lavorazione del vetro, i liutai o anche i diversi oggetti impiegati nell'alpismo solo per citare alcuni esempi di beni riconosciuti come patrimonio immateriale mondiale dall'UNESCO e documentati attraverso il materiale archivistico. Su tale tipologia documentale, l'*object detection* può essere applicato anche con la finalità di riconoscere e classificare il materiale degli oggetti individuati (Schwartz e Nishino 2020). Si tratta di un'operazione potenzialmente strategica per il patrimonio intangibile, con particolare attenzione alla componenete folkloristica, per la presenza di oggetti manifatturieri o abiti legati a tradizioni e ceremonie di cui diviene possibile recuperarne informazioni in merito la loro creazione o finalità. È altresì importante sottolineare che il riconoscimento del materiale pone in essere una sfida diversa rispetto all'individuazione degli oggetti; infatti, laddove quest'ultima attività si basa su attributi evidenti (es. colore, forma etc.), il riconoscimento dell'immateriale implica l'analisi di altro genere di proprietà non altrettanto facilmente analizzabili, richiedendo una stretta e quanto mai necessaria supervisione umana. A tal fine, le immagini conservate negli archivi potrebbero costituire la base di conoscenza per addestrare i sistemi basati su CNNs a riconoscere queste singole sfaccettature.

Un altro aspetto centrale nel contesto qui delineato è rappresentato dai movimenti, e più in generale dalle azioni e dai gesti che all'interno di una

comunità danno forma a pratiche, usanze e tradizioni. Anche quest'ultime, infatti, espresse attraverso danze, processioni e attività di vario genere, costituiscono parte del retaggio e *know-how* culturale di una comunità. Negli archivi audiovisivi, una parte del materiale conservato è caratterizzata da documentari aventi per oggetto ad esempio le attività e le usanze proprie della cultura popolare. Su tali tipologie documentali, le reti neurali convoluzionali possono essere applicate mediante le tecniche di *video activity recognition* con la finalità di classificare i movimenti riscontrabili nei segmenti di video. Le tecniche, in particolare, si basano sulla iniziale segmentazione del video in frame, utilizzando parametri impostati dall'utente, successivamente processati da specifici modelli di reti neurali convoluzionali con l'obiettivo di classificare gli elementi conoscitivi estraibili. Tale operazione avviene comparando ciascun frame con quello immediatamente precedente e successivo. Il confronto con i frame sequenziali ha un duplice obiettivo: da un lato permette di classificare il singolo elemento, dall'altro di verificare la continuità o la conclusione dell'azione di cui, ad esempio l'elemento fa parte. Si pensi ad una postura come elemento e al movimento come attività di cui il primo fa parte. Applicate su video del patrimonio culturale, le tecniche di riconoscimento automatico delle attività possono essere finalizzate a rintracciare e classificare i diversi elementi che danno forma a pratiche, balli e danze proprie delle singole comunità. Un esempio di ciò è quanto sperimentato per il riconoscimento di movimenti e gesti nelle forme del Tai Chi, patrimonio immateriale della cultura cinese, dove tali tecniche sono state applicate per riconoscere le diverse posture e gesti espressi in quell'arte (Li, Zou, e Li 2024). Similmente, in Italia, dove esistono danze tipiche, tradizioni e prassi (es. alpinismo) che definiscono il retaggio culturale di singole comunità, l'applicazione di tali tecniche può risultare vantaggiosa nella definizione di attività volte alla tutela e valorizzazione del patrimonio. Quest'ultime possono essere dunque impiegate per rintracciare le caratteristiche proprie di ciascuna attività del patrimonio culturale, permettendo di evidenziare le peculiarità e gli elementi, talvolta ignoti, che determinano e differenziano questo aspetto immateriale della cultura.

Applicate singolarmente, le reti neurali convoluzionali permettono l'estrazione di elementi diversi tra loro in base alla natura del documento e agli obiettivi imposti. Un aspetto importante da evidenziare riguarda, infine, la possibilità di usare congiuntamente tali tecniche per capitalizzare i diversi oggetti informativi. Infatti, considerata l'elevata eterogeneità delle fonti archivistiche conservati in Italia, le reti neurali convoluzionali potrebbero costituire uno strumento dalle eccezionali potenzialità per esplorare testi, immagini e audiovisivi al fine di rintracciare anche elementi di rimando valorizzando la rete di relazioni che ad esempio caratterizza un archivio.

Tuttavia, una riflessione va fatta sulla disponibilità eventuale di dataset annotati di dominio da utilizzare per le reti neurali. Nonostante i risultati positivi raggiunti in casi specifici, attualmente, sotto una prospettiva allargata, si è lontani da una reale capitalizzazione della varietà di elementi che caratterizzano il patrimonio culturale. Difatti, pur disponendo di innumerevoli modelli affidabili, l'applicabilità di tali sistemi è condizionata dai dati di addestramento. Un aspetto che diviene ancora più evidente nel settore culturale, dove la mancanza di specifici dataset impedisce di capitalizzare la peculiare eterogeneità di elementi, aumentando l'occorrenza di errori legati al bias informativo di dataset non coerenti con il dominio (Fig. 1).



**Figura 1.** Risultato della sperimentazione di *Object Detection* condotta impiegando un modello di YOLO v8x addestrato sul dataset Google Open Image v.7 in cui la Torre pendente di Pisa è stata riconosciuta come un alveare, un anello lavorato a vetro come ciambella e, infine, un oggetto raffigurante un elefante come un cucciolo di elefante.

Definire una *knowledge base* è un'attività che richiede, necessariamente, varie competenze e collaborazione tra figure professionali diverse in grado di stabilire gli aspetti metodologici e tecnologici. A tale scopo, l'esperto di

dominio è chiamato a individuare le peculiarità e a collaborare con i profili tecnici e informatici nella definizione del dataset per l'addestramento. Per la realizzazione di tale attività, possono essere impiegati strumenti in grado di guidare l'annotazione, sia sotto l'aspetto tecnico e tecnologico (Lin 2018) che metodologico. Gli archivi, in questo senso, hanno il potenziale informativo per dare forma a tali dataset, e gli archivisti le competenze metodologiche quantomai indispensabili in questo processo. Pertanto, un approccio non solo tecnologico ma soprattutto metodologico, volto a definire dataset di qualità potrebbe rappresentare una interessante prospettiva soprattutto in considerazione dei risultati positivi ottenuti dall'applicazione dell'*object detection* e *video activity recognition*.

#### 4. Conclusioni e prospettive

Il presente lavoro ha avuto come obiettivo quello di delineare, in una prospettiva teorica, le modalità attraverso cui le reti neurali convoluzionali possono essere funzionali all'estrazione automatica della conoscenza da documenti archivistici come immagini e video. Partendo dal materiale archivistico digitale, sia nativo che scansionato, la finalità imposta in questo articolo è quella di evidenziare la possibilità di orientare le tecnologie attualmente disponibili in favore del patrimonio culturale per collegare tra loro oggetti ed entità del dominio. Allo stato attuale, lo sviluppo di tali tecnologie, unitamente all'implementazione di algoritmi sempre più sofisticati apertamente disponibili, consente la realizzazione di attività dal forte impatto per il settore culturale. Tuttavia, per garantire una corretta elaborazione delle tecniche citate, soprattutto in considerazione della peculiarità del patrimonio culturale, è necessario prestare attenzione alla scarsa disponibilità di dataset relativi a tale dominio. Occorre, dunque, definire un approccio specifico per trattare un dominio così vasto. A tal fine, le tecnologie possono costituire un valido strumento, ma solo a condizione che esse siano modellate in base al dominio e non viceversa. Comprendere l'importanza dei dataset, e in che modo questi influiscano sui risultati finali, è fondamentale per avere contezza delle criticità che contrassegnano l'attuale scenario applicativo delle tecnologie di reti neurali, specialmente nel settore culturale nel quale si registra una scarsa o inesistente disponibilità di dati aperti per l'addestramento e lo sviluppo di modelli. Quest'ultima condizione rappresenta forse l'ostacolo più significativo per l'implementazione di apposite applicazioni nel dominio del patrimonio culturale. Tale constatazione assume un maggior peso alla luce dei risultati ottenuti dalle

medesime tecnologie e tecniche in altri settori quali la medicina, la guida autonoma, gli assistenti vocali etc., dove l'ampia disponibilità di dataset contribuisce al perfezionamento di tali tecnologie.

## Riferimenti bibliografici

- AlKendi, Wissam, Franck Gechter, Laurent Heyberger, and Christophe Guyeux. 2024. “Advancements and Challenges in Handwritten Text Recognition: A Comprehensive Survey.” *Journal of Imaging* 10 (1): 18. <https://doi.org/10.3390/jimaging10010018>.
- Cosovic, Marijana, and Radmila Jankovic. 2020. “CNN Classification of the Cultural Heritage Images.” In *2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH)*, 1–6. East Sarajevo, Bosnia and Herzegovina: IEEE. <https://doi.org/10.1109/INFOTEH48170.2020.9066300>.
- Dhruv, Patel, and Subham Naskar. 2020. “Image Classification Using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN): A Review.” In *Machine Learning and Information Processing*, a cura di Debabala Swain, Prasant Kumar Pattnaik, e Pradeep K. Gupta, 1101:367–81. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore. [https://doi.org/10.1007/978-981-15-1884-3\\_34](https://doi.org/10.1007/978-981-15-1884-3_34).
- Ji, Yuzhu, Haijun Zhang, Zhao Zhang, and Ming Liu. 2021. “CNN-Based Encoder-Decoder Networks for Salient Object Detection: A Comprehensive Review and Recent Advances.” *Information Sciences* 546 (febbraio):835–57. <https://doi.org/10.1016/j.ins.2020.09.003>.
- Kahle, Philip, Sebastian Colutto, Gunter Hackl, and Gunter Muhlberger. 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. Kyoto: IEEE. <https://doi.org/10.1109/ICDAR.2017.307>.
- Kiessling, Benjamin, Robin Tissot, Peter Stokes, and Daniel Stokl Ben Ezra. 2019. “eScriptorium: An Open Source Platform for Historical Document Analysis.” In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 19–19. Sydney, Australia: IEEE. <https://doi.org/10.1109/ICDARW.2019.90032>.

- Li, Xiongfeng, Limin Zou, and Haojie Li. 2024. “Tai Chi Movement Recognition and Precise Intervention for the Elderly Based on Inertial Measurement Units and Temporal Convolutional Neural Networks.” *Sensors* 24 (13): 4208. <https://doi.org/10.3390/s24134208>.
- Lin, Tzu Ta. 2018. «LabelImg». Python. Label Studio. <https://github.com/HumanSignal/labelImg/releases/tag/v1.8.1>.
- Mishra, Mayank, and Paulo B. Lourenço. 2024. “Artificial intelligence-assisted visual inspection for cultural heritage: State-of-the-art review.” *Journal of Cultural Heritage* 66:536–50. <https://doi.org/10.1016/j.culher.2024.01.005>.
- Ptucha, Raymond, Felipe Petroski Such, Suhas Pillai, Frank Brockler, Vatsala Singh, and Paul Hutkowski. 2019. “Intelligent Character Recognition Using Fully Convolutional Neural Networks.” *Pattern Recognition* 88 (aprile):604–13. <https://doi.org/10.1016/j.patcog.2018.12.017>.
- Rovella, Anna, Assunta Caruso, Martin Critelli, and Francesca M.C. Messiniti. 2024. “Knowledge Extraction, Research Projects and Archives Management.” *AIDAinformazioni* 41 (2): 145–56. <https://doi.org/10.57574/596532918>.
- Sarika, Naragudem, Nageswararao Sirisala, and Muni Sekhar Velpuru. 2021. “CNN based Optical Character Recognition and Applications.” In *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 666–72. Coimbatore, India: IEEE. <https://doi.org/10.1109/ICICT50816.2021.9358735>.
- Schwartz, Gabriel, and Ko Nishino. 2020. “Recognizing Material Properties from Images.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (8): 1981–95. <https://doi.org/10.1109/TPAMI.2019.2907850>.
- Shafait, Faisal, and Ray Smith. 2010. “Table Detection in Heterogeneous Documents.” In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, 65–72. Boston Massachusetts USA: ACM. <https://doi.org/10.1145/1815330.1815339>.
- Smirnov, Stanislav, and Alma Eguizabal. 2018. “Deep learning for object detection in fine-art paintings.” In *2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo)*, 45–49. Cassino FR, Italy: IEEE. <https://doi.org/10.1109/MetroArchaeo43810.2018.9089828>.

- Stanford University Human-Centered Artificial Intelligence. 2024. “Artificial Intelligence Index Report 2024.” [https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI\\_2024\\_AI-Index-Report.pdf](https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf).
- Tan, Chuangchuang, Huan Liu, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, and Yunchao Wei. 2023. “Rethinking the Up-Sampling Operations in CNN-based Generative Network for Generalizable Deepfake Detection.” arXiv. <http://arxiv.org/abs/2312.10461>.
- Wigington, Curtis, Chris Tensmeyer, Brian Davis, William Barrett, Brian Price, and Scott Cohen. 2018. “Start, Follow, Read: End-to-End Full-Page Handwriting Recognition.” In *Computer Vision – ECCV 2018*, a cura di Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, e Yair Weiss, 11210:372–88. Lecture Notes in Computer Science. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-01231-1\\_23](https://doi.org/10.1007/978-3-030-01231-1_23).
- Xu, Derong, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, Yang Wang, and Enhong Chen. 2024. “Large Language Models for Generative Information Extraction: A Survey.” arXiv. <http://arxiv.org/abs/2312.17617>.
- Zahid, Shaik Mohammed, T. Nashiya Najesh, Salman. K, Shaik Ruhul Ameen, and Anooja Ali. 2023. “A Multi Stage Approach for Object and Face Detection using CNN.” In *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, 798–803. Coimbatore, India: IEEE. <https://doi.org/10.1109/ICCES57224.2023.10192823>.
- Zhou, Jun, Yuhang Lu, Kang Zheng, Karen Smith, Colin Wilder, and Song Wang. 2018. “Design Identification of Curve Patterns on Cultural Heritage Objects: Combining Template Matching and CNN-based Re-Ranking.” arXiv. <http://arxiv.org/abs/1805.06862>.