

MUNSELL SOIL FINDER E PYWALL: STRUMENTI PER LA DOCUMENTAZIONE CROMATICA E IL RILIEVO DEI PARAMENTI MURARI IN ARCHEOLOGIA

1. INTRODUZIONE

Questo lavoro presenta due applicazioni sviluppate per supportare il ricercatore nella documentazione cromatica dei materiali archeologici e nel rilievo formale dei paramenti murari. Munsell Soil Finder consente di caricare una foto, selezionare un punto o un'area e ottenere la notazione Munsell corrispondente; PyWALL esamina immagini ortorettificate di paramenti murari, distingue i mattoni dalla malta ed estrae i contorni in formato vettoriale¹. Munsell Soil Finder e PyWALL sono applicazioni desktop dotate di interfaccia grafica, sviluppate in Python secondo una tecnica qui definita “Vibe Coding per l'Archeologia”: il ricercatore, mediante istruzioni in linguaggio naturale (prompt), descrive a un assistente IA le funzionalità desiderate del software e il modello genera il codice corrispondente. Questo approccio rende lo sviluppo di strumenti computazionali accessibile anche a ricercatori privi di competenze informatiche avanzate.

Il processo si articola in cinque fasi: 1) ideazione (definizione del problema via prompt); 2) generazione del codice; 3) validazione (esecuzione e annotazione degli errori); 4) iterazione guidata (correzioni e test fino all'ottenimento di una versione stabile); 5) consolidamento (ottimizzazione e stesura della documentazione)².

2. PROBLEMI OPERATIVI CHE HANNO MOTIVATO LO SVILUPPO DEI TOOL

Standardizzare la documentazione cromatica è una sfida metodologica dell'archeologia contemporanea. Il sistema Munsell (MUNSELL 1915) è lo standard di riferimento (BLOCH *et al.* 2021, 134-135). I valori si determinano per confronto visivo con le Munsell Soil-Color Charts raccolte nel Munsell Soil-Color Book. Ciascuna tavola (chart) è dedicata a una tinta (hue) e presenta una griglia di chip cromatici semi-lucidi (color chips) a valore (value) e croma (chroma) variabili, ognuno affiancato da un foro. L'operatore, portando l'oggetto sotto il foro, individua il chip più simile. Questa procedura presenta limiti noti: soggettività percettiva degli operatori, sensibilità alle condizioni di

¹ I due software, prototipi di ricerca sviluppati e utilizzati dall'autore, saranno resi disponibili alla comunità scientifica tramite il rilascio del codice sorgente, dei dataset di validazione e della relativa documentazione tecnica.

² Il codice sorgente di entrambe le applicazioni è stato generato da ChatGPT o3-mini-high (OpenAI). Le versioni più recenti e stabili sono state consolidate da Claude Opus 4.5 (Anthropic).

illuminazione (spettro, intensità e direzione della luce, che possono alterare la resa cromatica percepita), variabilità dovuta all'usura delle tavole (GERHARZ *et al.* 1988; GOODWIN 2000; SÁNCHEZ-MARAÑÓN *et al.* 2005). A ciò si aggiunge il costo non trascurabile del book, che può rappresentare un limite alla sua diffusione, soprattutto tra studenti e liberi professionisti con risorse ridotte. Tutte queste criticità rendono necessari strumenti alternativi, capaci di favorire una documentazione cromatica più oggettiva, standardizzata e accessibile. Da questa esigenza nasce l'applicazione Munsell Soil Finder.

Se nella documentazione cromatica il limite principale è la soggettività, nel rilievo murario pesa il tempo richiesto per scontornare in CAD i materiali visibili nelle ortofoto. Da questa esigenza operativa nasce PyWALL, concepito per produrre automaticamente disegni vettoriali (DXF).

3. STATO DELL'ARTE

3.1 *Munsell Soil Finder*

Munsell Soil Finder si colloca nel panorama delle soluzioni per la conversione RGB-Munsell. I principali attori industriali, come X-Rite, offrono dispositivi portatili (ad es. Capsure) che misurano sull'oggetto i valori CIELAB e restituiscono il colore standard più vicino nelle librerie selezionate, incluso il Munsell Soil-Color Book³. Sebbene assicurino accuratezza grazie a illuminazione e calibrazione integrate, tali strumenti presentano svantaggi legati alla natura chiusa (algoritmi proprietari non documentati) e ai costi elevati.

Parallelamente, si è diffuso un ecosistema di applicazioni che usano la fotocamera dei comuni smartphone (GÓMEZ-ROBLEDO *et al.* 2013; HAN *et al.* 2016). Tra queste, ChromaMagic: pensata per il contesto artistico, è in grado di riconoscere in qualsiasi foto i corrispondenti valori Munsell. Sebbene l'app offra portabilità e facilità d'uso, non è chiara la logica di conversione cromatica; inoltre, non essendo prevista una procedura di calibrazione dell'immagine, i risultati possono essere poco affidabili (<https://chromamagic.com/>). In ambito accademico si segnalano soluzioni sviluppate specificamente per l'uso archeologico, come ad esempio ARCA: una desktop application che integra acquisizione dell'immagine, campionamento manuale, stima automatica della notazione Munsell e generazione del report (MILOTTA *et al.* 2017).

Esistono infine soluzioni open-source usate in ambito tecnico-scientifico che implementano algoritmi RGB-Munsell basati su diversi approcci (CENTORE 2012). Tra queste rientrano librerie software e pacchetti come Colour Science, munsellinterpol, munsell.js e MunsellSpace: strumenti che offrono

³ CIELAB (o $L^*a^*b^*$) è uno spazio colore standardizzato dalla Commission Internationale de l'Éclairage (CIE), definito da tre coordinate: L^* (luminosità), a^* (verde-rosso) e b^* (blu-giallo). Consente confronti oggettivi tra colori mediante il calcolo di distanze euclidee (ΔE).

motori di color matching robusti, ma che richiedono competenze di programmazione e risultano quindi poco accessibili a utenti generici⁴.

3.2 PyWALL

La segmentazione è un processo automatico che suddivide l'immagine in regioni omogenee assegnando ogni pixel a una categoria. In particolare, nel rilievo di muri in laterizio, i pixel si classificano come “mattoni”, “malta” o “altro”, producendo una maschera (binaria o multiclasse) da cui estrarre i contorni. I primi tentativi di segmentazione automatica di elementi da muro si basavano su algoritmi con regole geometriche e cromatiche predefinite, la cui accuratezza dipendeva dalla taratura manuale dei parametri (soglie di colore e peso dei filtri) da parte dell'operatore. Inoltre, tali metodi risultavano più efficaci in presenza di paramenti ben leggibili, caratterizzati da un netto contrasto tra elementi costruttivi e giunti: condizioni spesso assenti nei contesti reali (RIVEIRO *et al.* 2015; VALERO *et al.* 2018).

Negli ultimi anni, l'applicazione del Deep Learning (DL) ha fatto progredire sensibilmente questo campo, migliorando sia la precisione del riconoscimento, sia la capacità di adattarsi a scenari variabili (IBRAHIM *et al.* 2020; LOVERDOS, SARHOSIS 2023; VANDENABEELE *et al.* 2024). Tuttavia, l'impiego del DL comporta almeno due problematiche: la necessità di dataset ampi e vari per addestrare i modelli; il costo computazionale (il processo per segmentare immagini ad alta risoluzione richiede GPU performanti e tempi di calcolo rilevanti). Nonostante ciò, i benefici in accuratezza e automazione stanno rendendo questo metodo lo standard nella ricerca.

4. FUNZIONAMENTO DEI SOFTWARE

4.1 Munsell Soil Finder

Munsell Soil Finder si basa su *real.dat*: un dataset di 2734 campioni messo a disposizione dal Munsell Color Science Laboratory⁵. All'interno del file, per ogni campione si trovano la notazione Munsell (H V/C) e le coordinate colorimetriche CIE xyY (calcolate usando l'Illuminante C)⁶. Munsell

⁴ Il filone di ricerca sull'automazione della codifica Munsell ha prodotto contributi rilevanti fino al 2025 in tre aree tematiche: sviluppo del sistema ARCA (MILOTTA *et al.* 2018, 2020); approcci neurali per la stima RGB→Munsell (SOLÍS *et al.* 2022; NODI *et al.* 2025); smartphone-based colorimetry (SINCLAIR, KABIR 2022; NODI *et al.* 2023).

⁵ Il file è disponibile nella pagina “Munsell Color Science Lab Educational Resources” del Rochester Institute of Technology (RIT): <https://www.rit.edu/science/munsell-color-science-lab-educational-resources>.

⁶ CIE xyY è una rappresentazione standard del colore definita dalla CIE: x e y sono le coordinate di cromaticità (tinta/saturazione), mentre Y esprime la luminanza relativa (chiarezza). L'Illuminante C è uno standard CIE che simula una luce diurna media (temperatura di colore ~6774K).

Soil Finder opera su una tavolozza di riferimento che integra questi 2734 campioni con alcuni valori interpolati⁷. All'avvio, l'applicazione converte le coordinate xyY di tutti i campioni in valori CIELAB ($xyY \rightarrow XYZ \rightarrow XYZ_{D65} \rightarrow L^*a^*b^*$), costruendo così un database interrogabile mediante ricerca nearest-neighbor (albero KD).

Il programma offre due modalità di analisi: singolo pixel o “Region of Interest” (ROI) di 11×11 pixel. Al click dell'utente sull'immagine, il programma acquisisce il colore selezionato, lo converte da RGB in CIELAB e lo confronta con le triplette dell'albero KD per identificare rapidamente i 3 valori Munsell più vicini (basandosi sulla distanza euclidea)⁸. In una seconda fase di raffinamento, calcola per questi candidati il ΔE_{2000} (ASTM D 2244-05) – metrica standardizzata e percettivamente più accurata che quantifica le differenze cromatiche come percepite dall'occhio umano – riordinandoli per differenza percettiva anziché distanza euclidea.

4.2 PyWALL

PyWALL offre due modalità di segmentazione: clustering non supervisionato (K-Means) e DL (DynUNet pre-addestrato). Il software non introduce quindi un nuovo algoritmo di segmentazione. La sua originalità risiede piuttosto nell'orchestrazione di soluzioni già consolidate nella computer vision e nel machine learning, entro un'unica pipeline operativa orientata al rilievo archeologico.

4.2.1 Segmentazione mediante clustering K-Means

Il primo approccio usa l'algoritmo K-Means, un metodo di clustering che raggruppa i pixel in base alla loro somiglianza su più caratteristiche. Per ogni pixel, il software costruisce un profilo composto da 5 descrittori: i 3 valori colore ($L^*a^*b^*$), il valore di prossimità a un bordo rilevato dall'algoritmo Canny e quello che esprime il livello di ruvidità intorno al pixel (indice di texture locale). K-Means riceve questi dati e trova la migliore suddivisione matematica in 3 cluster, separando i gruppi di pixel cromaticamente e morfologicamente simili. Ottenuti i cluster, il software li interpreta in base all'indice di texture (il quinto descrittore). Per ciascun cluster, PyWALL calcola il valore medio di tale indice; confronta questi valori medi e identifica

⁷ Confrontando real.dat con il Munsell Soil-Color Book si rilevano alcune lacune: Cromi dispari (C=1, C=3) necessari per distinguere colori a bassa saturazione; Valore intermedio V=2.5 per descrivere suoli scuri ricchi di materia organica; scala completa dei grigi neutri e dei bianchi. Il problema è stato risolto grazie alla libreria munsellinterpol, che ha calcolato le coordinate xyY dei campioni mancanti.

⁸ In modalità ROI i pixel dell'area vengono convertiti in CIELAB e la mediana dei valori $L^*a^*b^*$ è usata per la stima del colore, riducendo l'influenza di pixel anomali dovuti a rumore o micro-variazioni dell'immagine.

il cluster che ha il valore più basso in assoluto, cioè quello caratterizzato da minore variabilità locale della luminosità. A questo punto, il codice prende una decisione: “questo cluster, essendo il più liscio, è la malta”. Tutti i pixel che non appartengono al cluster della malta vengono uniti e considerati “mattoni”. Da questa classificazione deriva una maschera binaria: un’immagine in cui ogni pixel assume uno dei due valori possibili. I pixel classificati come malta vengono impostati a 0 (nero), mentre quelli classificati come mattone vengono impostati a 255 (bianco).

4.2.2 Segmentazione mediante DL

Il secondo approccio integra un modello DynUNet pre-addestrato, cioè una rete neurale convoluzionale della famiglia U-Net già disponibile nel repository TopoMortar e addestrata su un dataset supervisionato specifico per materiali da costruzione⁹.

4.2.3 Post-processing ed esportazione vettoriale

Indipendentemente dal metodo, la maschera può essere rifinita con strumenti interattivi di correzione manuale. Una volta consolidata la maschera, i contorni dei mattoni vengono visualizzati ed esportati in DXF come polilinee chiuse.

5. TEST

5.1 Munsell Soil Finder

Per validare l’affidabilità sono stati condotti due test specifici sul riconoscimento cromatico.

5.1.1 Test in condizioni operative realistiche: obiettivi e procedura

Il funzionamento del software si basa sull’analisi di immagini digitali, non sulla misurazione diretta dell’oggetto. Questa mediazione tecnologica comporta una vulnerabilità intrinseca: ogni passaggio della catena di acquisizione – illuminazione ambientale, risposta spettrale del sensore, profilo colore, compressione JPEG – può perturbare i valori cromatici dell’oggetto, compromettendo l’accuratezza della determinazione Munsell. Per valutare l’influenza di tali fattori, è stata condotta una validazione sperimentale basata sulla riproduzione digitale delle Munsell Soil-Color Charts. Le tavole sono state fotografate includendo nel fotogramma una ColorChecker Passport (X-Rite), uno strumento di calibrazione che fornisce tre target di riferimento

⁹ <https://github.com/jmlipman/TopoMortar>. L’addestramento supervisionato usa coppie di dati: ogni immagine è accompagnata dalla sua “soluzione” corretta (una maschera disegnata a mano da un operatore umano che delinea perfettamente mattoni e malta).

(White Balance, Classic e Creative Enhancement). La presenza del target Classic in ogni scatto è fondamentale: le sue 24 patch certificate costituiscono un riferimento oggettivo per verificare l'accuratezza cromatica. Se correttamente riprodotti, è ragionevole presumere che anche le Munsell Charts mantengano la fedeltà cromatica.

Le acquisizioni sono state effettuate in formato RAW (NEF) con fotocamera Nikon D7100, all'aperto, in una giornata di cielo sereno e nelle ore centrali, ma sempre in condizioni di ombra. Per identificare le condizioni di illuminazione più favorevoli al funzionamento del software, le chart sono state fotografate in tre situazioni diverse: all'ombra di un gazebo bianco, all'ombra di un porticato e all'ombra di un edificio collocato in un'area verde. Per ciascuno scenario è stato generato uno specifico profilo DCP (Digital Camera Profile) di correzione cromatica.

La procedura di acquisizione ha seguito questo protocollo standardizzato: prima è stato eseguito il bilanciamento del bianco "in camera", inquadrando il target White Balance della ColorChecker Passport, quindi sono state fotografate le tavole Munsell avendo cura di includere sempre il target Classic nello stesso fotogramma. La pipeline di elaborazione ha previsto i seguenti step: conversione RAW→DNG, creazione profilo DCP tramite software ColorChecker Passport, sviluppo in RawTherapee (applicazione profilo, correzione esposizione, affinamento della neutralità cromatica sulla patch Neutral 8, conversione in spazio colore sRGB), esportazione in JPEG 8 bit (qualità massima).

Prima di sottoporre le immagini all'analisi con Munsell Soil Finder, è stato necessario verificare che l'intera pipeline di elaborazione avesse preservato correttamente i valori cromatici. Il controllo è stato condotto mediante un'applicazione appositamente sviluppata che campiona i valori RGB delle 24 patch del target Classic presenti in ogni fotogramma e calcola la differenza cromatica rispetto ai valori certificati dal produttore. Sono state ammesse alla fase successiva solo le immagini il cui profilo cromatico, valutato sulle 24 patch del target Classic mediante differenze ΔE_{2000} rispetto ai valori di riferimento X-Rite, è risultato di qualità adeguata alla documentazione¹⁰. A questo punto, sulle immagini validate è stato possibile testare le prestazioni di Munsell Soil Finder sui chip delle tavole Munsell (Fig. 1).

Per ciascun chip sono stati eseguiti 5 campionamenti in modalità ROI 11×11 (uno centrale e quattro periferici), che sono stati confrontati con le notazioni Munsell attese mediante uno script sviluppato appositamente. I risultati sono sintetizzati attraverso due indici: accuratezza (la percentuale di

¹⁰ Il tool di validazione classifica i profili su quattro livelli (eccellente, buono, accettabile, scadente) basati sul ΔE_{2000} medio delle 24 patch. Sono state ammesse le immagini con profilo classificato come "eccellente" o "buono" e con nessuna patch in classe "scadente".

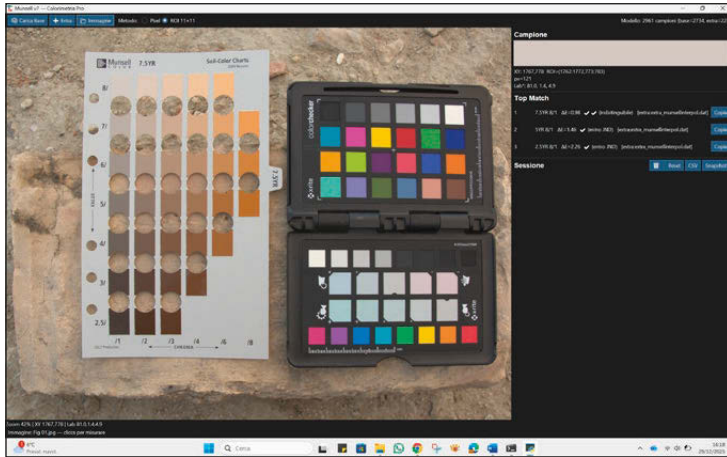


Fig. 1 – Interfaccia di Munsell Soil Finder durante l'analisi di una Munsell Soil-Color Chart. Il pannello destro visualizza i risultati di una misurazione eseguita sul chip 7.5YR 8/1 in modalità ROI 11×11 pixel.

chip correttamente identificati)¹¹ e robustezza (che misura la stabilità interna del riconoscimento). Quest'ultima è calcolata sui soli chip PASS e misura il grado di accordo tra le cinque ROI: valori prossimi a 1 indicano che le diverse aree campionate sullo stesso chip restituiscono sistematicamente la notazione Munsell corretta, anche in presenza di micro-variazioni superficiali del chip¹².

5.1.2 Test in condizioni operative realistiche: risultati e discussione

I risultati evidenziano performance diverse a seconda delle condizioni di acquisizione (Tab. 1). La configurazione sotto il gazebo bianco, caratterizzata da illuminazione naturale diffusa e uniforme, ha prodotto i migliori risultati con accuratezza del 97% e robustezza 0,96. Sotto il porticato si osserva maggiore variabilità: l'accuratezza oscilla tra 71% e 100%, con robustezza compresa tra 0,87 e 1,00. Questa oscillazione non sembra dipendere dalle caratteristiche intrinseche delle chart (numero di chip, distribuzione delle tinte, range di chroma), bensì dalle specifiche condizioni locali di ripresa (presenza di riflessi, distribuzione non uniforme della luce, micro-ombre). Nel parco, accanto all'edificio, l'accuratezza scende invece al 63% con robustezza 0,88.

¹¹ Un chip è considerato correttamente identificato (PASS) quando almeno 3 campionamenti su 5 convergono sulla stessa notazione Munsell e questa coincide con il valore atteso. L'indice di accuratezza è calcolato come rapporto tra chip PASS e chip totali analizzati.

¹² Per ciascun chip PASS viene calcolato un valore di accordo pari al rapporto tra il numero di ROI che restituiscono la notazione corretta e le cinque ROI analizzate; l'indice complessivo di robustezza corrisponde alla media di questi valori.

Posizione	Foto	Tinta	Numero di chip	Indice di accuratezza	Indice di robustezza del consenso
Porticato	DSC_7905	5GY	8	100%	0,98
Porticato	DSC_7905	10Y	8	100%	0,88
Porticato	DSC_7907	N	7	100%	1
Porticato	DSC_7907	10GY	7	100%	0,94
Porticato	DSC_7907	5GY	7	100%	0,94
Gazebo	DSC_7469	7.5YR	35	97%	0,96
Porticato	DSC_7901	7.5YR	35	86%	0,87
Porticato	DSC_7902	2.5Y	31	84%	0,91
Porticato	DSC_7907	10Y	7	71%	0,96
Parco	DSC_7928	5GY	8	63%	0,88
Parco	DSC_7928	10Y	8	63%	0,88

Tab. 1 – Prestazioni di Munsell Soil Finder. La tabella riporta l'indice di accuratezza e l'indice di robustezza del consenso rilevati nei tre scenari di acquisizione (porticato, gazebo, parco).

Le prestazioni eccellenti sotto il gazebo confermano l'alta affidabilità del software con illuminazione diffusa e uniforme. Tale condizione minimizza i riflessi speculari sui chip semi-lucidi, consentendo al sensore di catturare i pigmenti reali anziché artefatti luminosi. Questo risultato è metodologicamente importante perché dimostra che le prestazioni intrinseche dell'algoritmo di color matching sono solide quando opera su dati fotografici di qualità adeguata.

Al contrario, sotto il porticato l'accuratezza risulta più variabile (71-100%) e nel parco decisamente inferiore (63%). Ciò dipende dal fatto che in assenza di illuminazione uniformemente diffusa, i riflessi speculari sulla superficie semi-lucida dei chip alterano localmente i valori RGB registrati nell'immagine, in funzione dell'angolo di incidenza della luce e della posizione del sensore. Nel parco, questo effetto può essere amplificato dalle superfici circostanti (pareti dell'edificio e vegetazione) che riflettono la luce solare creando sorgenti luminose secondarie con angoli di incidenza variabili, producendo pattern di riflessi complessi e difficilmente controllabili.

Un aspetto significativo riguarda la robustezza del consenso, che resta elevata anche negli scenari più sfavorevoli (0,87-0,88). Poiché l'indice è calcolato solo sui chip correttamente identificati, questi valori indicano che, quando il riconoscimento riesce, è sostenuto dal consenso della gran parte delle cinque ROI – tipicamente 4-5 su 5 – e non su un accordo marginale (valori come 0,87-0,88 sono ben sopra la soglia minima teorica del PASS, che è 3 su 5 = 0,60). L'algoritmo è dunque internamente stabile: le perdite di accuratezza non derivano da sue fluttuazioni, ma da alterazioni esterne dei valori cromatici (riflessi).



Fig. 2 – I reperti ceramici utilizzati per il test comparativo “uomo *vs* macchina”.

Nel complesso, l'esperimento conferma che Munsell Soil Finder è affidabile quando opera su immagini acquisite con illuminazione diffusa, calibrate via ColorChecker e validate sul piano cromatico. La riduzione di accuratezza osservata nelle configurazioni meno favorevoli non dipende dal motore di color matching, ma dalla perturbazione dei valori RGB introdotta dai riflessi sui chip semi-lucidi.

Questa limitazione ha però rilevanza ridotta nel contesto applicativo per cui il software è stato sviluppato: i reperti archeologici presentano tipicamente superfici opache o semi-opache, significativamente meno soggette a riflessi speculari.

5.1.3 Confronto uomo *vs* macchina: obiettivi e procedura

Il test aveva due obiettivi: 1) confrontare le notazioni restituite dal software con quelle assegnate da osservatori umani in condizioni controllate; 2) misurare la coerenza inter-osservatore, così da valutare il peso della variabilità percettiva nel metodo tradizionale basato sul confronto visivo. Sei frammenti ceramici di età tardomedievale sono stati fotografati in studio



Fig. 3 – Un operatore attribuisce la notazione Munsell a uno dei reperti del test.

con illuminazione artificiale e ColorChecker Passport (Fig. 2)¹³. Le immagini, elaborate secondo la pipeline già descritta (§ 5.1.1), sono state analizzate con Munsell Soil Finder. Gli stessi campioni sono stati poi valutati visivamente da 10 osservatori, ciascuno dei quali ha assegnato la notazione Munsell ritenuta più appropriata mediante confronto diretto con il Munsell Soil-Color Book (Fig. 3). Gli osservatori presentavano una diversa esperienza nell'impiego del Munsell: su una scala da 0 (nessuna esperienza) a 2 (molta esperienza), circa metà dei soggetti si autovalutava come moderatamente esperta (0,5-1,0), due si collocavano al massimo (2,0) e due al minimo (0).

I dati sono stati processati mediante script automatico per produrre due confronti: 1) software *vs* osservatori (60 confronti totali: 10 osservatori $\times$ 6 campioni); 2) inter-osservatori (con 10 persone si formano 45 coppie possibili; per 6 campioni fa 270 confronti). I confronti sono stati classificati

¹³ L'illuminazione è stata fornita da due softbox con lampade D65 (luce diurna standard), così da ridurre al minimo le variazioni cromatiche introdotte dalla sorgente. Le stesse condizioni sono state mantenute durante le valutazioni degli osservatori, in modo da rendere le annotazioni cromatiche più attendibili e confrontabili con quelle prodotte dal software.

Categoria di accordo	Numero di casi	Percentuale	Descrizione
Perfect	23	38,3%	L'osservatore ha identificato la stessa identica notazione Munsell dell'app
Livello 1 (lv1)	7	11,7%	L'osservatore ha mantenuto la tinta corretta, con una differenza minima (1 gradino) su Value o Chroma
Livello 2 (lv2)	4	6,7%	L'osservatore ha mantenuto la tinta corretta, con una differenza contenuta (2 gradini) su Value o Chroma
Cambio di tinta	26	43,3%	L'osservatore ha classificato il campione con una tinta (Hue) diversa da quella dell'app

Tab. 2 – Categorie di accordo tra software e osservatori (60 confronti: 10 osservatori × 6 campioni).

Campione	Accordo (n)	Accordo (%)	Cambio di tinta (n)	Cambio di tinta (%)	Totale
1	5	50,0%	5	50,0%	10
2	9	90,0%	1	10,0%	10
3	2	20,0%	8	80,0%	10
4	9	90,0%	1	10,0%	10
5	2	20,0%	8	80,0%	10
6	7	70,0%	3	30,0%	10
Totale	34	56,7%	26	43,3%	60

Tab. 3 – Accordo tra software e osservatori per ciascun campione ceramico. La colonna Accordo include sia le coincidenze perfette sia gli scostamenti lievi entro la stessa tinta (lv1, lv2); i cambi di tinta, considerati discordanti, sono riportati separatamente.

in categorie ordinate per grado crescente di scostamento: coincidenza perfetta, quando Hue, Value e Chroma risultano identici; scostamenti lievi entro la stessa tinta, distinti in lv1 e lv2; cambio di tinta (hue_change), considerato critico perché implica l'attribuzione del campione a una diversa famiglia cromatica.

5.1.4 Confronto uomo *vs* macchina: risultati e discussione

Assumendo le notazioni fornite dal software come riferimento operativo ottenuto mediante una pipeline fotografica calibrata e replicabile, emerge che – su 60 confronti complessivi – solo il 38,3% delle valutazioni umane coincide. Includendo gli scostamenti lievi (lv1, lv2), la concordanza sale al 56,7%. Resta quindi un 43,3% di casi in cui gli osservatori attribuiscono una tinta completamente diversa rispetto al software: il dato non va interpretato come scarto rispetto a un “gold standard” assoluto, ma come misura della variabilità introdotta dal confronto visivo rispetto a una procedura digitale deterministica, ripetibile a parità di acquisizione e calibrazione (Tab. 2). La concordanza varia fortemente tra i campioni (Tab. 3): i nn. 2 e 4 raggiungono il 90% di accordo con il software, mentre i nn. 3 e 5 scendono al 20%,

con ben l'80% degli osservatori che sceglie una tinta alternativa. Questo evidenzia che la difficoltà della valutazione visiva dipende dal colore specifico del campione, soprattutto per tonalità collocate in aree di transizione tra tinte adiacenti. I campioni nn. 1 e 6 mostrano valori intermedi (50% e 70% di accordo rispettivamente).

Anche la concordanza inter-osservatore è bassa: nei 270 confronti incrociati, solo nel 28,9% dei casi due osservatori danno esattamente la stessa notazione Munsell a un dato campione. Includendo gli scostamenti lievi, l'accordo arriva al 61,9%, ma resta un sostanziale 38,1% in cui due osservatori assegnano tinte completamente diverse allo stesso campione ceramico.

5.2 PyWALL

Per testare uno strumento di segmentazione destinato al rilievo murario non basta valutarne la velocità: occorre dimostrare che il disegno prodotto rispecchi le forme reali. In PyWALL, il prodotto fondamentale del processo è la maschera binaria mattone/malta, da cui il software estrae i contorni (mediante algoritmi di contour detection) e li esporta in DXF. Se la classificazione dei pixel è imprecisa, l'errore si propaga inevitabilmente ai contorni e quindi al rilievo finale. Per questo motivo la validazione si concentra sulla maschera, confrontandola con maschere di riferimento create mediante annotazione manuale da parte di un operatore esperto.

5.2.1 Test sull'affidabilità e la correttezza della segmentazione: obiettivi e procedura

Il confronto quantitativo tra maschere automatiche e maschere di riferimento è stato svolto da un'applicazione sviluppata ad hoc, che calcola quattro indicatori standard: IoU (Intersection over Union, area dei pixel marcati come "mattone" sui quali software e riferimento concordano, rapportata all'unione delle due aree mattone), Precisione (quota di pixel marcati come "mattone" dal software che sono davvero mattone nel riferimento), Recall (quota di pixel "mattone" del riferimento effettivamente intercettati dal software) e F1-score (sintetizza Precisione e Recall in un unico indice bilanciato). Il risultato è un set di valori per ciascuna immagine e una media finale, che restituisce una misura complessiva della qualità della segmentazione.

La validazione è stata condotta su quattro casi. Il muro 1 è un'immagine sintetica, priva di rumore, utilizzata come riferimento in condizioni ideali; i muri 2 e 3 rappresentano situazioni più vicine al reale ma ancora favorevoli, con bassa variabilità cromatica e materica e, nel terzo caso, con malta particolarmente leggibile e contrasto più netto tra materiali. Il muro 4, infine, costituisce lo scenario più critico, caratterizzato da malta cromaticamente eterogenea, bordi sfumati ed erosione superficiale, condizioni che rendono meno univoca la definizione dei contorni e richiedono approcci più robusti.

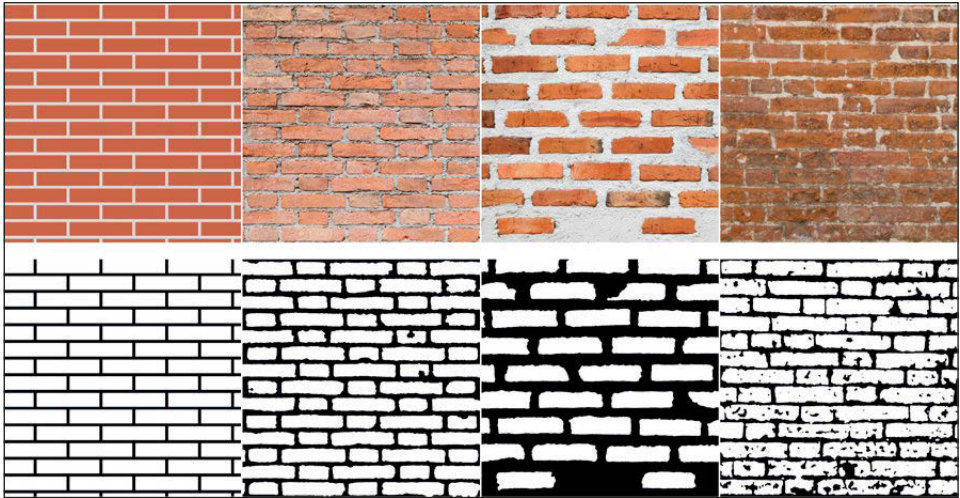


Fig. 4 – Immagini originali (sopra) e maschere di segmentazione prodotte da PyWALL (sotto).

Caso	Scenario	Metodo	IoU	Precisione	Recall	F1-score
Muro 1	Immagine sintetica ideale	K-Means	1,000	1,000	1,000	1,000
Muro 3	Malta ben leggibile (contrasto netto)	K-Means	0,974	0,980	0,994	0,987
Muro 2	Bassa variabilità cromatica e materica	K-Means	0,920	0,996	0,923	0,958
Muro 4	Malta eterogenea, bordi sfumati ed erosione	Deep Learning	0,883	0,987	0,893	0,938
Media	–	–	0,944	0,991	0,953	0,971

Tab. 4 – Metriche di segmentazione automatica su quattro scenari di complessità crescente, calcolate confrontando le maschere automatiche con le annotazioni manuali di riferimento.

5.2.2 Test sull’affidabilità e la correttezza della segmentazione: risultati e discussione

I risultati della validazione mostrano come le prestazioni della segmentazione varino sistematicamente in funzione della complessità dello scenario analizzato (Fig. 4, Tab. 4)¹⁴. Nel muro 1 (immagine sintetica ideale), la segmentazione coincide perfettamente con il riferimento manuale (tutti gli indicatori sono pari a 1,000). Passando ai casi reali, nel muro 2, caratterizzato da bassa variabilità cromatica e materica, l’IoU resta elevato (0,920) e

¹⁴ Per interpretare correttamente i valori riportati, è utile ricordare che le metriche di segmentazione sono espresse su una scala da 0 a 1, dove valori superiori a 0,9 sono generalmente considerati eccellenti nella letteratura tecnica; quelli compresi tra 0,8 e 0,9 sono buoni; quelli inferiori a 0,8 indicano segmentazioni che richiedono significative correzioni manuali.

la Precisione è quasi unitaria (0,996). Il Recall, leggermente inferiore (0,923), indica invece che circa l'8% dei pixel di mattone presenti nella maschera di riferimento non sono stati intercettati dall'algoritmo e sono stati erroneamente classificati come malta.

Nel muro 3, dove il contrasto tra materiali è netto, le prestazioni sono eccellenti (IoU=0,974; F1=0,987). Il Recall particolarmente elevato (0,994) conferma che quando la distinguibilità cromatica è buona, l'algoritmo K-Means è capace di individuare quasi tutti i pixel appartenenti ai mattoni con pochissime omissioni. Il muro 4 presenta condizioni nettamente più avverse: malta eterogenea, bordi sfumati ed erosione superficiale. In questo caso, la segmentazione K-Means si è rivelata inadeguata; tuttavia, grazie alla modalità DL, i risultati rimangono competitivi (IoU=0,883; F1=0,938), sebbene inferiori alla media dei casi precedenti.

In sintesi, la segmentazione K-Means (muri 1-3) mostra ottima efficacia quando le condizioni del paramento permettono una buona separazione cromatica tra materiali. Al contrario, quando il degrado rende ambigui i confini (muro 4), l'approccio DL diventa necessario. È importante sottolineare che i valori riportati sono stati ottenuti da una pipeline completamente automatica, senza alcun intervento correttivo manuale. In contesti applicativi reali, una revisione selettiva della maschera mediante lo strumento "gomma" di PyWALL consente di correggere errori localizzati della segmentazione – ad esempio piccole aree erroneamente classificate come malta all'interno delle facce dei laterizi – e di migliorare la qualità del contorno vettoriale esportato.

6. INTELLIGENZA ARTIFICIALE E ARCHEOLOGIA DELL'ARCHITETTURA: IL CASO DELLE MURA MEDIEVALI DI RIMINI

Gli applicativi sviluppati sono stati sperimentati su un tratto della cinta urbana di Rimini, lungo via Bastioni Meridionali (Fig. 5). Questo complesso architettonico rappresenta un importante esempio di fortificazione tardomedievale, benché non sia stato ancora oggetto di uno studio stratigrafico di dettaglio (LEONI 2012).

6.1 *Strumenti e metodi*

PyWALL è stato utilizzato per ricavare dalle ortofoto di 3 campioni murari il rilievo formale dei laterizi (Fig. 5). I campioni 1 e 2 – murature in buone condizioni con malta ben conservata e contrasto netto tra materiali – hanno prodotto segmentazioni automatiche di qualità sufficiente per l'estrazione vettoriale diretta. Il campione 3, invece, presenta criticità analoghe al "muro 4" della validazione (§ 5.2.2), talvolta accentuate (malta degradata, bordi sfumati ed erosione superficiale). Di conseguenza, anche la segmentazione DL ha prodotto contorni con discontinuità ed errori (Fig. 6), rendendo necessario



Fig. 5 – Cinta urbana di Rimini, settore meridionale (via Bastioni Meridionali). Il riquadro indica il campione 3. Il n. 1 si trova nella parete orientale della torre adiacente, il n. 2 sulla cinta, 100 m più a E del campione 3.

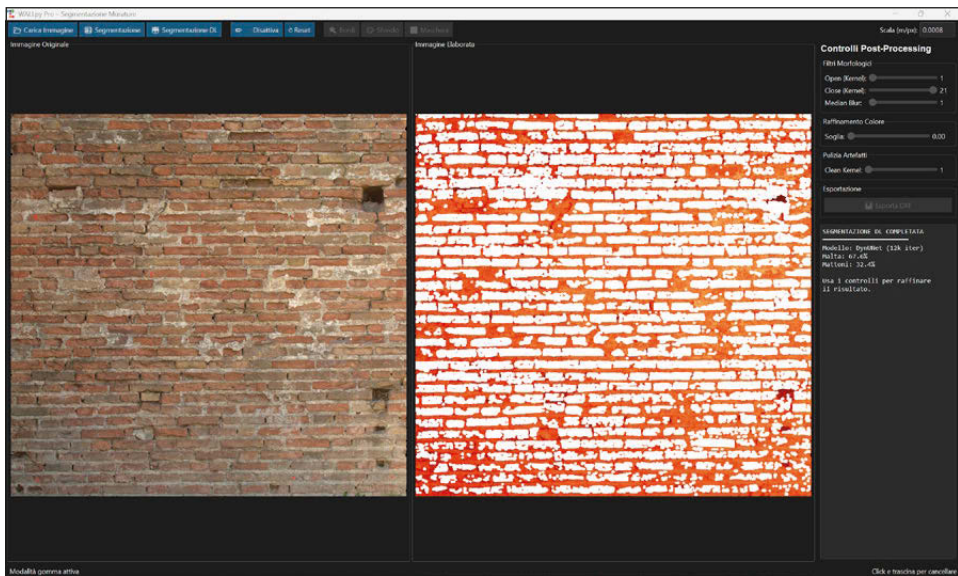


Fig. 6 – Segmentazione Deep Learning (DynUNet) del campione 3 all'interno di PyWALL.

un intervento correttivo per ottenere un risultato utilizzabile (rimozione dei vettori errati e digitalizzazione delle porzioni problematiche sull'ortofoto) (Fig. 7). Le dimensioni dei mattoni sono state misurate sulla muratura tramite una riga metallica. Per ciascuna misura (lunghezza, larghezza e spessore) si è registrato anche lo stato di conservazione, distinguendo cioè i casi in cui una dimensione è integra, rotta o non determinabile (assenza di visibilità dei bordi).

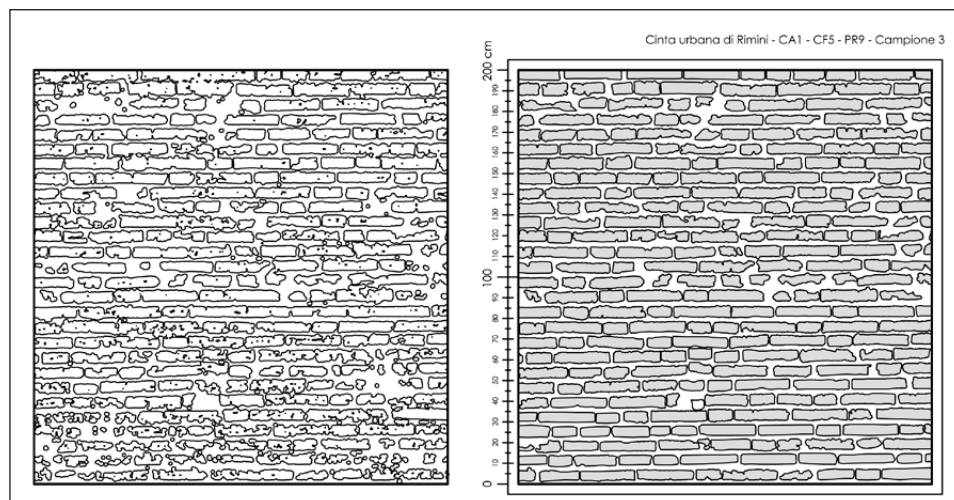


Fig. 7 – Limiti della segmentazione su paramento degradato (campione 3). A sinistra, vettori dalla maschera: i contorni presentano discontinuità ed errori. A destra, risultato dopo correzione manuale (rimozione vettori errati e digitalizzazione selettiva in CAD).

Sul campione 3 è stata condotta anche l'analisi cromatica con Munsell Soil Finder. Le foto sono state scattate interponendo un pannello diffusore per ottenere una luce omogenea, replicando così le condizioni operative dello scenario “gazebo”. Poiché il colore dei laterizi non è perfettamente omogeneo, sulle immagini acquisite sono stati eseguiti, per ciascun mattone, cinque campionamenti cromatici (uno centrale e quattro periferici)¹⁵. Per esplorare congiuntamente caratteristiche cromatiche e dimensionali, i mattoni interi sono stati analizzati mediante *fuzzy c-means*¹⁶. A differenza dei metodi tradizionali a categoria rigida, l'approccio fuzzy calcola per ogni mattone il grado di appartenenza (membership) a ogni cluster: valori prossimi a 1 indicano un'appartenenza netta, mentre quelli intorno a 0,5 intercettano i mattoni con caratteristiche ibride, frequenti in materiali coevi soggetti a variabilità produttiva.

¹⁵ Per ciascun laterizio, i dati sono stati sintetizzati calcolando la mediana delle coordinate CIELAB (L^* , a^* , b^*), l'intervallo interquartile (IQR) di L^* come indicatore di variabilità interna, e identificando la notazione Munsell dominante e la relativa quota percentuale sui cinque campionamenti.

¹⁶ L'analisi separa i laterizi disposti “di fianco” e “di testa” per evitare falsi raggruppamenti dovuti alla miscelazione di lunghezze e larghezze. Per ciascun gruppo, il clustering considera solo i mattoni con dimensioni intere, utilizzando un vettore a cinque dimensioni: mediana di L^* , a^* e b^* , dimensione primaria (lunghezza per “fianco” e larghezza per “testa”) e spessore. Prima del clustering, i dati sono normalizzati via Z-score per equiparare il peso delle componenti metriche e colorimetriche.

Gruppo	Cluster	N	Appart. media	Dimensione principale media (cm)	Spess. medio (cm)	L* medio	a* medio	b* medio	Notazione Munsell dominante	Quota notaz. dominante
fianco	1	15	0,72	26,68	4,92	34,53	7,51	14,96	7.5YR 3/3	0,59
fianco	2	12	0,71	26,85	4,93	37,93	11,21	19,12	5YR 3/4	0,65
testa	1	13	0,8	11,75	5,27	38,76	11,48	19,23	5YR 4/4	0,52
testa	2	12	0,72	11,94	5,08	35,78	6,67	14,85	10YR 4/3	0,48

Tab. 5 – Risultati del clustering fuzzy sui laterizi integri del campione 3. La tabella riporta per ogni cluster la notazione Munsell dominante, le dimensioni medie e l'appartenenza media: valori elevati indicano cluster ben definiti, mentre valori inferiori segnalano la presenza di elementi ibridi.

Campione	Posizione	Rotti	Integri	Stato nd	Lung. (cm)	Larg. (cm)	Spes. (cm)
1	Torre (CF4, PR8)	26,92%	69,23%	3,85%	26,84	11,87	5,08
2	Cinta (CF9, PR17)	39,47%	60,53%	0,00%	27,05	11,80	5,15
3	Cinta (CF5, PR9)	41,14%	32,91%	25,95%	26,76	11,85	5,01

Tab. 6 – Confronto tra campioni della cinta. Il n. 1 (torre) mostra frammentazione inferiore rispetto ai nn. 2 e 3 (cinta), compatibile con selezione del materiale. Le dimensioni medie dei laterizi sono sostanzialmente omogenee.

6.2 Risultati e interpretazione archeologica

Data la maggiore criticità operativa e la necessità di documentare i limiti del metodo in contesti reali, la discussione si concentra sul campione 3. Il dataset comprende 158 laterizi: 91 disposti di fianco, 26 di testa e 41 non determinabili (25,95%) per assenza di visibilità dei bordi. Dei 117 laterizi misurabili, 52 risultano integri (32,91%) – 27 di fianco e 25 di testa – mentre 65 sono rotti (41,14%), una percentuale significativa compatibile con pratiche di adattamento in opera e/o reimpiego. Il clustering fuzzy applicato ai 52 mattoni integri ha identificato due cluster cromatico-dimensionali per ciascuna disposizione (Tab. 5).

Negli elementi di fianco, le dimensioni medie dei due cluster risultano molto simili (lunghezza ~26,68-26,85 cm; spessore ~4,92-4,93 cm), mentre i valori cromatici differiscono: il cluster 1 presenta valori CIELAB corrispondenti a un colore più scuro e meno saturo ($L^*=34,5$; $a^*=7,5$; $b^*=15,0$) rispetto al cluster 2 ($L^*=37,9$, $a^*=11,2$, $b^*=19,1$). Le rispettive notazioni Munsell dominanti sono 7.5YR 3/3 nel cluster 1 (15 mattoni) e 5YR 3/4 nel cluster 2 (12 mattoni). La separazione avviene quindi principalmente per aspetto cromatico, non per formato, suggerendo due scenari: laterizi dello stesso formato provenienti da forniture con impasti o cotture diverse, oppure differenze superficiali (patine o alterazioni) che influenzano l'aspetto senza implicare necessariamente produzioni distinte.

Tra gli elementi di testa emergono analogamente due cluster: 13 mattoni con notazione prevalente 5YR 4/4 e 12 con 10YR 4/3. La larghezza media è simile (~11,75 *vs* ~11,94 cm), mentre lo spessore differisce lievemente (~5,27 *vs* ~5,08 cm). Le misure cromatiche sulle teste mostrano però una minore concordanza tra i cinque campionamenti rispetto ai fianchi: l'IQR mediano di L* (variabilità di L* tra i cinque campionamenti del singolo mattone) è 2,1 contro 1,7, e la concordanza della notazione Munsell tra i cinque campionamenti (Tab. 5, colonna "Quota notaz. dominante") scende da una media di 0,62 sui fianchi a 0,50 sulle teste. I due indicatori convergono nel suggerire maggiore variabilità superficiale, probabilmente per residui di malta, croste o esposizione agli agenti atmosferici. Nel complesso, il campione 3 mostra un modulo dimensionale coerente (~27×12×5 cm) e due famiglie cromatiche riconducibili a fattori produttivi (differenze di impasto e/o cottura) e/o post-deposizionali (alterazioni superficiali, patine, croste, residui di malta).

Il confronto tra i tre campioni evidenzia significativa variabilità nella frammentazione: il campione 1 (torre) presenta il 26,92% di laterizi rotti contro ~40% nei campioni 2 e 3 (cinta) (Tab. 6). Questo scarto è compatibile con una possibile selezione intenzionale: nei tratti strutturalmente più rilevanti sarebbero stati impiegati laterizi meglio conservati per garantire posa più regolare e maggiore stabilità.

Gli scarti dimensionali tra i campioni restano contenuti (<3 mm), confermando l'uso di un modulo uniforme che, confrontato con gli studi mensiocronologici pregressi, si colloca tra la fine del XV e i primi decenni del XVI secolo, sebbene la larghezza risulti lievemente inferiore ai valori di riferimento (LEONI 2016).

7. CONCLUSIONI

Questo lavoro aveva un duplice obiettivo: da un lato, dimostrare l'efficacia del Vibe Coding come paradigma per lo sviluppo di strumenti computazionali in archeologia; dall'altro, validare due applicazioni concrete che rispondono a esigenze operative specifiche della documentazione archeologica. I risultati ottenuti permettono di trarre conclusioni significative su entrambi i livelli.

Lo sviluppo di Munsell Soil Finder e PyWALL tramite modelli di intelligenza artificiale ha confermato che ricercatori senza formazione informatica specialistica possono creare strumenti software funzionanti e validati scientificamente. Tuttavia, il processo non è stato privo di sfide. Le fasi di debugging e ottimizzazione hanno richiesto la comprensione dei concetti algoritmici sottostanti. Il Vibe Coding, pertanto, facilita ma non elimina completamente la necessità di alfabetizzazione computazionale di base.

La validazione di Munsell Soil Finder conferma che, in condizioni di acquisizione controllate – illuminazione diffusa, calibrazione tramite Color-Checker e verifica colorimetrica delle immagini – il software può raggiungere un'accuratezza superiore al 95 % nell'identificazione delle notazioni Munsell. Il dato ne indica l'affidabilità come strumento per la documentazione cromatica standardizzata, purché la fase fotografica sia condotta secondo una procedura rigorosa. Il confronto con le valutazioni umane mostra inoltre il principale vantaggio della procedura digitale: assumendo le notazioni prodotte dal software come riferimento operativo ottenuto da una pipeline calibrata e ripetibile, solo il 38,3% delle attribuzioni visive coincide perfettamente con tale riferimento, mentre la concordanza inter-osservatore scende al 28,9%. Questi risultati evidenziano il peso della variabilità percettiva nel metodo tradizionale basato sul confronto visivo e, al tempo stesso, il valore della ripetibilità offerta da una procedura computazionale. Tale affidabilità, tuttavia, non è indipendente dalla qualità dell'immagine: illuminazione non uniforme e riflessi speculari possono alterare i valori cromatici registrati e ridurre l'accuratezza del riconoscimento.

La validazione di PyWALL mostra che la segmentazione automatica può raggiungere risultati molto elevati quando il paramento presenta una buona distinguibilità cromatica tra mattoni e malta. In queste condizioni, l'approccio K-Means produce maschere sufficientemente affidabili per l'estrazione dei contorni. Nei casi più critici, caratterizzati da malta eterogenea, erosione superficiale e bordi sfumati, lo stesso approccio può invece risultare inadeguato; in tali situazioni, la modalità Deep Learning offre una soluzione più robusta, pur richiedendo talvolta una revisione manuale selettiva. Nella pratica archeologica, ciò implica che la scelta tra K-Means e DL debba essere valutata caso per caso, in base alle caratteristiche conservative e cromatiche del paramento. Il contributo originale di PyWALL non consiste dunque nell'introduzione di un algoritmo inedito, ma nell'integrazione di metodi consolidati entro una pipeline operativa specificamente progettata per il rilievo murario, validata sperimentalmente e verificata su un caso di studio reale.

L'applicazione congiunta di PyWALL e Munsell Soil Finder – integrata da valutazioni autoptiche – ha permesso di acquisire nuovi dati sulla costruzione della cinta urbana di Rimini. La presenza di laterizi dimensionalmente omogenei suggerisce l'uso di matrici di formatura riconducibili a uno stesso modulo, mentre la variabilità cromatica può dipendere da fattori produttivi – differenze nell'impasto o nelle condizioni di cottura – e/o da alterazioni successive alla messa in opera, come patine e fenomeni superficiali. Il modulo dimensionale trova confronti con produzioni tardomedievali-rinascimentali; la minore frammentazione osservata nella torre rispetto ai tratti di cinta è compatibile con una possibile selezione intenzionale del materiale, basata su criteri qualitativi.

Munsell Soil Finder e PyWALL sono due strumenti sviluppati secondo il paradigma del Vibe Coding e dimostrano come modelli di intelligenza artificiale possano rendere accessibile lo sviluppo di software a partire da esigenze metodologiche della ricerca archeologica. Il rilascio open-source del codice sorgente e dei dataset di validazione, previsto per entrambe le applicazioni, è un passo necessario per la riproducibilità dei risultati e per la verifica indipendente dell'efficacia del paradigma stesso.

ANDREA FIORINI

Dipartimento di Storia Culture Civiltà (DiSci), Sezione di Archeologia
Alma Mater Studiorum, Università di Bologna
andrea.fiorini6@unibo.it

BIBLIOGRAFIA

- ASTM International 2005, *Standard Practice for Calculation of Color Tolerances and Color Differences from Instrumentally Measured Color Coordinates*, ASTM D 2244-05.
- BLOCH L.C., HOSEN J.D., KRACHT E.C., LEFEBVRE M.J., LOPEZ C.J., WOODCOCK R., KEEGAN W.F. 2021, *Is it better to be objectively wrong or subjectively right? Testing the accuracy and consistency of the Munsell capture spectrophotometer for archaeological applications*, «Advances in Archaeological Practice», 9, 2, 132-144 (<https://doi.org/10.1017/aap.2020.53>).
- CENTORE P. 2012, *An open-source inversion algorithm for the Munsell renotation*, «Color Research & Application», 37, 6, 455-464 (<https://doi.org/10.1002/col.20715>).
- GERHARZ R.R., LANTERMANN R., SPENNEMANN D.R. 1988, *Munsell color charts: A necessity for archaeologists?*, «The Australian Journal of Historical Archaeology», 6, 88-95.
- GÓMEZ-ROBLEDO L., LÓPEZ-RUIZ N., MELGOSA M., PALMA A.J., CAPITÁN-VALLVEY L.F., SÁNCHEZ-MARAÑÓN M. 2013, *Using the mobile phone as Munsell soil-colour sensor: An experiment under controlled illumination conditions*, «Computers and Electronics in Agriculture», 99, 200-208 (<https://doi.org/10.1016/j.compag.2013.10.002>).
- GOODWIN C. 2000, *Practices of color classification*, «Mind, Culture, and Activity», 7, 1-2, 19-36 (<https://doi.org/10.1080/10749039.2000.9677646>).
- HAN P., DONG D., ZHAO X., JIAO L., LANG Y. 2016, *A smartphone-based soil color sensor: For soil type classification*, «Computers and Electronics in Agriculture», 123, 232-241 (<https://doi.org/10.1016/j.compag.2016.02.024>).
- IBRAHIM Y., NAGY B., BENEDEK C. 2020, *Deep Learning-based masonry wall image analysis*, «Remote Sensing», 12, 23, 3918 (<https://doi.org/10.3390/rs12233918>).
- LEONI N. 2012, *Le mura bassomedievali di Rimini: problemi interpretativi e osservazioni preliminari*, «Sibrium», 26, 202-231 (<https://doi.org/10.17613/18eda-nvp71>).
- LEONI N. 2016, *Mensiocronologia dei laterizi nel centro storico di Rimini: prime analisi*, «Archeologia dell'Architettura», 21, 137-151 (<https://doi.org/10.1400/248558>).
- LOVERDOS D., SARHOSIS V. 2023, *Geometrical digital twins of masonry structures for documentation and structural assessment using machine learning*, «Engineering Structures», 275, art. 115256 (<https://doi.org/10.1016/j.engstruct.2022.115256>).
- MILOTTA F.L.M., FURNARI G., QUATTROCCHI C., PASQUALE S., ALLEGRA D., GUELI A.M., STANCO F., TANASI D. 2020, *Challenges in automatic Munsell color profiling for cultural heritage*, «Pattern Recognition Letters», 131, 135-141 (<https://doi.org/10.1016/j.patrec.2019.12.008>).
- MILOTTA F.L.M., STANCO F., TANASI D. 2017, *ARCA (Automatic Recognition of Color for*

- Archaeology*): A desktop application for Munsell estimation, in S. BATTIATO, G. GALLO, R. SCHETTINI, F. STANCO (eds.), *Image Analysis and Processing. ICIAP 2017. Lecture Notes in Computer Science*, vol. 10485, Cham, Springer (https://doi.org/10.1007/978-3-319-68548-9_60).
- MILOTTA F.L.M., STANCO F., TANASI D., GUELI A.M. 2018, *Munsell color specification using ARCA (Automatic Recognition of Color for Archaeology)*, «Journal on Computing and Cultural Heritage», 11, 4, art. 17, 1-15 (<https://doi.org/10.1145/3216463>).
- MUNSELL A.H. 1915, *Atlas of the Munsell Color System*, Malden, Wadsworth, Howland & Company, Inc., Printers.
- NODI S.S., PAUL M., ROBINSON N., WANG L., REHMAN S.U. 2023, *Determination of Munsell soil colour using smartphones*, «Sensors», 23, 6, art. 3181 (<https://doi.org/10.3390/s23063181>).
- NODI S.S., PAUL M., ROBINSON N., WANG L., REHMAN S.U., KABIR M.A. 2025, *Munsell soil colour prediction from the soil and soil colour book using patching method and deep learning techniques*, «Sensors», 25, 1, art. 287 (<https://doi.org/10.3390/s25010287>).
- RIVEIRO B., CONDE-CARNERO B., GONZÁLEZ-JORGE H., ARIAS P., CAAMAÑO J.C. 2015, *Automatic creation of structural models from point cloud data: The case of masonry structures*, «ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences», II-3/W5, 3-9 (<https://doi.org/10.5194/isprsannals-II-3-W5-3-2015>).
- SÁNCHEZ-MARAÑÓN M., HUERTAS R., MELGOSA M. 2005, *Colour variation in standard soil-colour charts*, «Soil Research», 43, 7, 827-837 (<https://doi.org/10.1071/SR04169>).
- SINCLAIR R., KABIR M.A. 2022, *Understanding the effect of smartphone cameras on estimating Munsell soil colors from imagery*, in 2022 *International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (Sydney 2022), IEEE, 1-8 (<https://doi.org/10.1109/DICTA56598.2022.10034624>).
- SOLÍS M., MUÑOZ-ALVARADO E., PEGALAJAR M.C. 2022, *The transformation of RGB images to Munsell soil-color charts*, «Uniciencia», 36, 1, 559-568 (<https://doi.org/10.15359/ru.36-1.36>).
- VALERO E., BOSCHÉ F., FORSTER A., HYSLOP E. 2018, *Historic digital survey: Reality capture and automatic data processing for the interpretation and analysis of historic architectural rubble masonry*, in R. AGUILAR, D. TORREALVA, S. MOREIRA, M.A. PANDO, L.F. RAMOS (eds.), *Structural analysis of historical constructions: An interdisciplinary approach*, RILEM Bookseries, vol. 18, Cham, Springer, 388-396 (https://doi.org/10.1007/978-3-319-99441-3_41).
- VANDENABEELE L., LOVERDOS D., PFISTER M., SARHOSIS V. 2024, *Deep Learning for the segmentation of large-scale surveys of historic masonry: A new tool for building archaeology applied at the Basilica of St Anthony in Padua*, «International Journal of Architectural Heritage», 18, 11, 1749-1761 (<https://doi.org/10.1080/15583058.2023.2260771>).

ABSTRACT

This paper presents two software applications for quantitative archaeological documentation: Munsell Soil Finder and PyWALL. Both tools were developed using AI-assisted natural language programming (Vibe Coding), demonstrating its viability for domain-specific archaeological software development. The two tools address recurrent operational bottlenecks: the standardization of color recording and time-intensive masonry documentation. Munsell Soil Finder maps pixel samples from calibrated photographs to Munsell (H V/C) notation via KD-tree nearest-neighbour search in CIELAB and ΔE_{2000} refinement, using 2734 reference samples plus interpolated values. PyWALL segments bricks and mortar in orthorectified images via K-Means clustering (Lab + edge/texture features) or a pre-trained DynUNet, extracts contours, and exports to DXF. Validation comprised: 1) chart tests with ColorChecker profiling (accuracy 63-100%, robustness ≥ 0.87); 2) human-versus-software and inter-observer

comparisons (38.3% exact agreement with the software output; 28.9% among observers), highlighting the variability of visual assessment; 3) segmentation benchmarking (IoU 0.883-1.00; F1 0.938-1.00). A case study on Rimini's medieval walls employs fuzzy c-means on combined chromatic and metric features to identify a consistent brick module (~27×12×5 cm) and two clusters per brick orientation. The clusters are separated primarily by CIELAB values rather than format, suggesting variability in clay composition or firing conditions, while fragmentation rates range from 26.92% to 41.14%.