




# Normalizzazione di slide istologiche

 Materials	<a href="#">A method for normalizing histology slides for quantitative analysis.pdf</a>
 Reviewed	

## COSA E' IL METODO MACENKO

L'idea principale dietro il metodo Macenko è quella di trasformare un'immagine in uno spazio in Densità Ottica (OD), stimare i vettori delle macchie primarie usando la Decomposizione di Valori Singoli (SVD), separare l'immagine in base alle componenti delle macchie, normalizzare queste componenti in una distribuzione di riferimento, e poi ricostruire l'immagine normalizzata. Questo processo standardizza la distribuzione dei colori in slide delle macchie, facendo analisi successive più coerenti e affidabili. Nella ricerca e diagnosi medica, i colori delle immagini dei tessuti delle macchie devono essere coerenti. Diversi laboratori e procedure possono produrre immagini con colori variabili, facendo diventare così le analisi accurate più difficili da realizzare. Il metodo Macenko standardizza questi colori, assicurando coerenza in tutte le immagini. Questo processo, chiamato normalizzazione delle macchie, aiuta sia i patologi umani che i tools automatizzati nel fare valutazioni più accurate, e crea risultati più affidabili nelle ricerche e diagnosi mediche.

## QUALI SONO LE SFIDE NEL PROCESSO DI COLORAZIONE DELLE SLIDE DI NORMALIZZAZIONE?

Diverse sfide saltano fuori dalla grande varietà nel processo di colorazione:

1)

### VARIABILITA' DELLA COLORAZIONE

- Differenze di protocollo → Diversi laboratori possono usare protocolli diversi, portando a differenze in termine di intensità di colore e tonalità
- Lotti di reagenti → La variabilità tra lotti di reagenti di macchie può influire sul colore e l'intensità delle macchie
- Preparazione delle slide → Le differenze nelle tecniche di preparazione delle slide possono introdurre incoerenza nella colorazione

2)

### IMPATTO NELL'ANALISI QUANTITATIVA

- Misurazioni incoerenti → La variabilità nella colorazione può portare a misurazioni incoerenti, influenzando l'affidabilità delle analisi quantitative
- Analisi automatizzate → Gli algoritmi di analisi automatizzata di immagini fanno affidamento su data input coerenti. Le variazioni nella colorazione possono risultare in output non accurati o non affidabili da parte di questi algoritmi
- Studi comparativi → Confrontare campioni di tessuti da diverse sorgenti diventa difficile se la colorazione non è coerente, riducendo la validità degli studi comparativi

3)

### RIPRODUCIBILITA' E AFFIDABILITA'

- Riproducibilità → Assicurare riproducibilità di risultati attraverso tutte le slide e gli studi è cruciale per la ricerca scientifica
- Accuratezza diagnostica → Nelle impostazioni cliniche, la coerenza della colorazione è essenziale per diagnosi accurate. La variabilità può portare a interpretazioni errate delle caratteristiche istologiche

## IL METODO MACENKO COME SUPERA QUESTE SFIDE?

Il metodo Macenko affronta queste sfide fornendo un approccio sistematico per normalizzare le variazioni di colore nelle slide istologiche. Ecco come colma le necessità di coerenza e affidabilità:

1)

### STANDARDIZZAZIONE DELLA COLORAZIONE

- Normalizzando la distribuzione di colore delle slide delle macchie in uno standard di riferimento, il metodo Macenko assicura coerenza nella colorazione in tutte le slide, i laboratori e gli studi

2)

### ANALISI QUANTITATIVA AVANZATA

- Le slide normalizzate forniscono un input coerente per le analisi quantitative, migliorando l'accuratezza e l'affidabilità delle misurazioni
- Gli algoritmi di analisi automatizzata delle immagini lavorano meglio con dati standardizzati, portando a output più affidabili


3)

### COMPARABILITA' MIGLIORATA

- La normalizzazione permette confronti significativi tra campioni di tessuti da diverse sorgenti, aumentando la validità degli studi comparativi

4)

## RIPRODUCIBILITA' AUMENTATA

- Standardizzare il processo di colorazione migliora la riproducibilità dei risultati, una pietra angolare della 
- La colorazione coerente supporta diagnosi accurate e affidabili nelle impostazioni cliniche

## COME FUNZIONA IL METODO MACENKO?

Il metodo Macenko è una tecnica ampiamente usata per la normalizzazione di colori nelle slide istologiche, in particolare per le slide H&E. Esso implica diversi step per trasformare un'immagine in uno spazio di colore normalizzato, riducendo variabilità e migliorando la coerenza per l'analisi quantitativa.

### Spiegazione step-by-step:

[1° passo]

#### CONVERSIONE DELL'IMMAGINE IN UNO SPAZIO OD

◦ Il primo step per convertire un'immagine RGB in uno spazio OD. Questa trasformazione è basata sulla legge di Beer-Lambert, la quale afferma che l'assorbimento della luce che passa attraverso un materiale è proporzionale alla sua concentrazione

◦ La formula di conversione è

$OD = -\log(I/I_0)$ , con  $[I]$  l'intensità di pixel nell'immagine e  $[I_0]$  l'intensità della luce di riferimento (normalmente 255 per un'immagine a 8 bit)

[2° passo]

#### MATRICE DI STIMA DELLE MACCHIE

◦ Usa la Decomposizione dei Valori Singoli (SVD) per trovare le componenti del colore principale nello spazio OD

$$OD_{matrix} = U \Sigma V^T$$

◦ Estrae i primi due vettori singolari da  $V$  per formare la matrice delle macchie  $S$

$$S = [V_1 \ V_2]$$

[3° passo] **MACCHIE SEPARATE**

◦ Usa i vettori di colorazione stimate per separare l'immagine nelle sue componenti. Questo comporta la proiezione dei valori OD nei vettori di colorazione

◦ I valori OD sono proiettati nei vettori, separando l'immagine in diversi canali di colorazione

[4°passo]

### NORMALIZZAZIONE DELLE MACCHIE

- Normalizzare le componenti delle macchie scalandole per farle combaciare con la distribuzione di riferimento. Questo step porta l'intensità delle macchie a uno standard coerente, riducendo la variabilità tra slide
- Questo comporta la scalatura delle componenti come per esempio le loro proprietà statistiche per farle combaciare con una slide di riferimento
- Aggiusta le componenti per farle combaciare con una distribuzione bersaglio, tipicamente derivata da un'immagine di riferimento.

$$C_{\text{normalized}} = \frac{C - \mu_C}{\sigma_C} \times \sigma_T + \mu_T$$

Con  $[\mu_c]$  e  $[\sigma_c]$  la deviazione media e standard delle concentrazioni correnti, mentre  $[\mu_t]$  e  $[\sigma_t]$  la deviazione media e standard delle concentrazioni target.

[5°passo]

### RICOSTRUZIONE DELL'IMMAGINE

- La ricostruzione dell'immagine normalizzata dalle componenti normalizzate. Converti i valori OD nello spazio RGB per ottenere l'immagine normalizzata
- I valori OD normalizzati sono riconvertiti nello spazio RGB per produrre l'immagine normalizzata finale
- Le componenti normalizzate vengono combinate nell'immagine RGB per raggiungere la colorazione standardizzata

$$OD_{\text{normalized}} = C_{\text{normalized}} \times S$$

$$I_{\text{normalized}} = I_o \times e^{-OD_{\text{normalized}}}$$

Con  $I_{\text{normalized}}$  l'immagine RGB normalizzata.

## IMPLEMENTAZIONE DEL METODO MACENKO IN PYTHON

Ecco un'implementazione in python usando le librerie "numpy" e "opencv".

[1°passo]

### IMPORTARE LE LIBRERIE

Bisogna importare le librerie necessarie come `cv2` (OpenCV), `numpy` (per le operazioni numeriche) e PCA da `sklearn.decomposition`.



-OpenCV è un software-libreria open-source per il machine learning, l'immagine processing e la computer vision (appunto OpensourceComputerVisionlibrary-OpenCV). Usando questa libreria si possono processare immagini e video per identificare oggetti, facce, o addirittura la scrittura umana.

-NumPy è un pacchetto per il processamento di array in generale. Fornisce un array object ad alte performance e multidimensionale, oltre a tools per lavorare con questi array. E' un pacchetto fondamentale per il computing scientifico con python. Inoltre può essere usato anche come contenitore multidimensionale di dati generici.

-PCA (principal component analysis) è una risoluzione al problema che si pone nel virtualizzare dati con molti attributi in input, migliora la virtualizzazione e aumenta l'accuratezza.

```
import cv2
import numpy as np
from sklearn.decomposition import PCA
```

## [2° passo] CARICARE E MOSTRARE L'IMMAGINE

-Caricare l'immagine dal percorso specifico usando OpenCV( `cv2.imread()` )

-Mostrare l'immagine originale usando

`cv2.imshow()` e aspettare la key press per continuare ( `cv2.waitKey(0)` )

-Chiudere tutte le finestre usando

`cv2.destroyAllWindows()`

```
image_path='/Users/edoardoarca/Downloads/HEImage.jpg'
img=cv2.imread(image_path)
cv2.imshow('Original Image', img)
cv2.waitKey(0)
cv2.destroyAllWindows()
```

## [3° passo] CONVERTIRE L'IMMAGINE RGB IN UNO SPAZIO OD

-Definire una funzione

`rgb_to_od()` per convertire l'immagine RGB in uno spazio OD

-Evitare le divisione per zero impostando i valori dei pixel da 0 a 1

-Applicare la trasformazione con

```
-np.log(img/255)
```

```
def rgb_to_od(img):  
    img=img.astype(np.float32)  
    img[img==0]=1 #per evitare divisioni per zero  
    return -np.log(img/255)  
img_od=rgb_to_od(img)
```

[4°passo]STIMARE LA MATRICE DELLE MACCHIE CON IL PCA

-Rimodellare l'immagine OD in un array 2D

-Usare il PCA per stimare la matrice con 2 componenti

-Trasporre le componenti per avere la matrice

```
reshaped_od=img_od.reshape((-1,3))  
pca=PCA(n_components=2)  
pca.fit(reshaped_od)  
stain_matrix=pca.components_.T
```

[5°passo]SEPARARE LE MACCHIE DALL'IMMAGINE OD

-Calcolare il prodotto scalare dell'immagine OD rimodellata e la pseudo-inversa della matrice

```
stains=np.dot(reshaped_od,np.linalg.pinv(stain_matrix).T)
```

[6°passo]RIMODELLARE LE MACCHIE PER COMBACIARE CON LA FORMA DELL'IMMAGINE ORIGINALE

-Rimodellare l'array delle macchie per farlo combaciare con la forma dell'immagine originale

```
stains=stains.reshape(img.shape[0],img.shape[1],-1)
```

[7°passo]NORMALIZZARE LE MACCHIE

-Definire la media e standard deviazione del target

-Normalizzare le macchie usando le deviazioni

```
target_means=np.array([0.5],[0.5])  
target_stds=np.array([0.2],[0.2])
```

```
stains_normalized=(stains - np.mean(stains,axis=(0,1))) / np.std(stains,axis=(0,1))
stains_normalized=stains_normalized * target_std + target_means
```

#### [8°passo] RICOSTRUIRE LE MACCHIE NORMALIZZATE IN UN'IMMAGINE RGB

- Calcolare il prodotto scalare delle macchie normalizzate rimodellate e la trasposta della matrice
- Rimodellare il risultato per combaciare con la forma dell'immagine originale
- Usare la funzione esponenziale (`np.exp()`) per invertire la trasformazione logaritmica

```
od_normalized=np.dot(stains_normalized.reshape((-1,2)), stain_matrix.T)
od_normalized=od_normalized.reshape(img.shape)
img_reconstructed=np.exp(-od_normalized)
```

#### [9°passo] RITAGLIARE I VALORI, CONVERTIRE IN 8-BIT E MOSTRARE L'IMMAGINE RICOSTRUITA

- Ritagliare i valori dei pixel nell'intervallo [0,1] usando `np.clip()`
- Convertire l'immagine in un formato a 8 bit usando `(img_reconstructed * 255).astype(np.uint8)`
- Mostrare l'immagine ricostruita usando `cv2.imshow()`

```
img_reconstructed=np.clip(img_reconstructed,0,1)
img_reconstructed=(img_reconstructed * 255).astype(np.uint8)
cv2.imshow('Reconstructed Image', img_reconstructed)
cv2.waitKey(0)
cv2.destroyAllWindows()
```

## CONCLUSIONE

Il processo di normalizzazione delle macchie usando il metodo Macenko aiuta a far risultare le immagini istopatologiche più coerenti e affidabili. Questo lo fa convertendo le immagini in uno spazio speciale chiamato OD, poi tramite tecniche matematiche corregge i colori a un livello standard. La standardizzazione è cruciale perché assicura che i dottori e i ricercatori possono accuratamente analizzare e comparare queste immagini, portando a migliori diagnosi e trattamenti per i pazienti.



*Quali sono i benefici di usare tecniche di normalizzazione come il metodo Macenko?*

I benefici includono riproducibilità aumentata e comparabilità di risultati attraverso diversi laboratori, impatto di variazioni di colore ridotto nell'analisi delle immagini e nell'interpretazione, e accuratezza elevata nella diagnostica e ricerca di malattie basate su immagini istopatologiche

*Può la normalizzazione alterare le informazioni biologiche presenti nelle immagini?*

Le tecniche di normalizzazione sono progettate per preservare le informazioni biologiche sottostanti mentre corregge l'intensità del colore per la standardizzazione. Il bersaglio è minimizzare gli artefatti tecnici senza compromettere le caratteristiche essenziali del tessuto analizzato

*Ci sono limitazioni o considerazioni mentre si implementa la normalizzazione?*

Una considerazione potrebbe stare nel scegliere il metodo di normalizzazione appropriato in base all'applicazione specifica, capendo il potenziale impatto di certi modelli di colorazione o caratteristiche, e assicurando la validità e l'ottimizzazione dei parametri per risultati affidabili e accurati

*La normalizzazione è pertinente solo a immagini istopatologiche o può essere applicata per altri tipi di imaging medico?*

Mentre la normalizzazione delle macchie è comunemente usata nell'istopatologia, tecniche simili possono essere applicate ad altri tipi di imaging medico come l'imaging immunoistochimico e fluorescente, dove la coerenza del colore è cruciale per avere analisi e interpretazioni accurate

*Quali misure possono essere adottate per validare l'efficacia delle tecniche di normalizzazione?*

Gli studi di validità sono essenziali per valutare la performance e l'affidabilità dei metodi di normalizzazione. Questo include comparare le immagini normalizzate con dati di verità di base, valutando l'impatto in task analitiche specifiche, e conducendo confronti tra laboratori per assicurare coerenza e riproducibilità