

Porównanie metod dekonwolucji barwników w obrazach histopatologicznych barwionych hematoksyliną i eozyną

Oliwia Woźniak

1 Cel projektu

Celem pracy było porównanie efektywności różnych metod dekonwolucji barwników w obrazach histopatologicznych barwionych hematoksyliną i eozyną (H&E). W analizie będzie brane pod uwagę:

- dokładności rekonstrukcji obrazu po dekonwolucji
- zdolności separacji składników barwnikowych
- zgodności z teoretycznym rozkładem barwników
- czasu przetwarzania różnych metod
- wpływu technik normalizacji na efektywność dekonwolucji

Porównam trzy metody dekonwolucji: metodę referencyjną, metodę Ruiforka oraz metodę falkową z analizą niezależnych składowych (Wavelet + ICA), w połączeniu z różnymi technikami normalizacji obrazów.

2 Wstęp teoretyczny

2.1 Podstawy dekonwolucji barwników

Wszystkie analizowane metody opierają się na fundamentalnym modelu optycznej gęstości (OD) wynikającym z prawa Lamberta-Beera, które opisuje absorpcję światła w materiale:

$$OD = -\log_{10} \left(\frac{I}{I_0} \right) \quad (1)$$

gdzie I to intensywność obrazu dla danego piksela, a $I_0 = 255$ to intensywność tła (niezabarwiony obszar).

W przestrzeni OD zakłada się liniową separowalność barwników, co pozwala na przedstawienie problemu dekonwolucji jako równania macierzowego:

$$\mathbf{OD}_{m \times 3} = \mathbf{C}_{m \times n} \cdot \mathbf{M}_{n \times 3}^T \quad (2)$$

gdzie \mathbf{M} to macierz barwników (wektory barw w przestrzeni OD), \mathbf{C} to macierz stężeń, m to liczba pikseli, a n to liczba barwników. Różnice między metodami polegają na sposobie estymacji macierzy \mathbf{M} i rozwiązania tego równania.

2.2 Metody dekonwolucji

2.2.1 Metoda referencyjna

Metoda referencyjna stanowi nadzorowane podejście do dekonwolucji, wymagające *a priori* znanej macierzy barwników \mathbf{M}_{ref} .

Podstawy teoretyczne:

- **Podejście nadzorowane:** Wymaga znanej macierzy barwników określonej dla danego protokołu barwienia

- **Brak etapu uczenia lub estymacji** - dekonwolucja przeprowadzana jest poprzez bezpośrednie rozwiązywanie równania liniowego
- **Reprodukowalność:** Gwarantuje identyczne wyniki dla różnych obrazów przy użyciu tej samej macierzy referencyjnej

Algorytm:

1. Transformację do przestrzeni OD zgodnie z prawem Lambert-Berra
2. Rozwiązywanie równania liniowego wykorzystującego pseudoodwrotność Moore'a-Penrose'a:

$$\mathbf{C} = \mathbf{M}_{\text{ref}}^t \cdot OD \quad (3)$$

3. Rekonstrukcję obrazu:

$$I_{\text{recon}} = I_0 \cdot 10^{-\mathbf{M}_{\text{ref}} \cdot \mathbf{C}} \quad (4)$$

Zastosowana macierz referencyjna dla barwienia H&E:

$$\mathbf{M}_{\text{ref}} = \begin{bmatrix} 0.65 & 0.07 \\ 0.70 & 0.99 \\ 0.29 & 0.11 \end{bmatrix} \quad (5)$$

gdzie pierwsza kolumna odpowiada hematoksylinie, a druga eozynie.

2.2.2 Metoda Ruiforka

Metoda Ruiforka jest nienadzorowanym podejściem wykorzystującym analizę głównych składowych (PCA) do automatycznej estymacji macierzy barwników bez wymagania wstępnej wiedzy o barwnikach.

Podstawy teoretyczne:

- **Liniowa separowalność:** W przestrzeni OD, piksele barwione pojedynczym barwnikiem leżą wzdłuż linii prostych wychodzących z początku układu współrzędnych
- **Dominujące kierunki barwników:** Wektory własne obliczone za pomocą PCA na foregroundzie odpowiadają poszukiwanym kierunkom barwników w przestrzeni OD
- **Automatyczna estymacja:** Macierz barwników jest wyznaczana na podstawie statystycznych właściwości obrazu

Algorytm:

1. Konwersję do OD maskowanie tła poprzez usunięcie pikseli o niskiej OD:

$$\text{mask} = \|OD_i\|_2 > \tau \quad (6)$$

2. Analizę PCA foregroundu:

$$\text{PCA}(OD_{\text{foreground}}) \rightarrow \text{wektory własne } V \quad (7)$$

3. Estymację macierzy barwników poprzez transpozycję wektorów własne:

$$M = V^T \quad (8)$$

4. Dekonwolucję:

$$C = M^t \cdot OD \quad (9)$$

2.2.3 Metoda falkowa z analizą niezależnych składowych (Wavelet + ICA)

Metoda zaproponowana przez Najah Alsubaie, Shan E Ahmed Raza oraz Nasir Rajpoot wykorzystuje analizę niezależnych składowych (ICA) w dziedzinie falkowej do dekonwolucji barwników w obrazach histopatologicznych.

Podstawy teoretyczne:

- **Relaksacja założenia niezależności:** Barwniki mogą być statystycznie zależne, ale ich wąskopasmowe reprezentacje falkowe mogą zapewnić niezależność
- **Reprezentacja wielorozdzielcza:** Dekompozycja falkowa wydobywa informację teksturalną związaną z poszczególnymi składnikami tkankowymi
- **Selekcja podpasm:** Wybór podpasm o najmniejszym rozkładzie Gaussa zwiększa niezależność statystyczną składowych

Algorytm:

1. Dekompozycję falkową każdego kanału barwnego:

$$d_c = [d_r, d_g, d_b]^T \rightarrow \{d_{c,a}^l, d_{c,h}^l, d_{c,v}^l, d_{c,o}^l\}_{l=1}^L \quad (10)$$

gdzie L to liczba poziomów dekompozycji, a, h, v, o to odpowiednio pasma: aproksymacji, poziome, pionowe i diagonalne

2. Konstrukcję pasm trzykanałowych i selekcję 20 podpasm o najwyższej wartości kurtozy:

$$K = \frac{E(u - \mu)^4}{\sigma^4} - 3 \quad (11)$$

3. Analizę ICA wybranych pasm:

$$\mathbf{X} = \mathbf{MS} \rightarrow \mathbf{S} = \mathbf{WX} \quad (12)$$

gdzie \mathbf{W} to macierz demieszająca

4. Estymację macierzy barwników poprzez regresję liniową

5. Dekonwolucję:

$$\mathbf{C} = \mathbf{M}^t \mathbf{OD} \quad (13)$$

2.3 Metody normalizacji

2.3.1 Normalizacja specyficzna dla barwników (Stain-Specific)

- **Zasada działania:** Dekonwolucja obrazu na składowe barwników, klasteryzacja każdego kanału na foreground/background oraz dopasowanie statystyk do obrazu referencyjnego
- **Przestrzeń barwna:** Przestrzeń gęstości optycznej (OD) z późniejszą rekonalwolucją do RGB
- **Równanie normalizacji:**

$$C_{\text{norm}} = (C_{\text{orig}} - \mu_{\text{src}}) \cdot \frac{\sigma_{\text{trg}}}{\sigma_{\text{src}}} + \mu_{\text{trg}} \quad (14)$$

gdzie C to stężenie barwnika, μ to średnia, σ to odchylenie standardowe

2.3.2 Normalizacja Reinharda

- **Zasada działania:** Transformacja do przestrzeni LAB i statystyczne dopasowanie średniej oraz odchylenia standardowego
- **Przestrzeń barwna:** Przestrzeń LAB (Luminance, A, B)
- **Równanie normalizacji:**

$$LAB_{\text{norm}} = (LAB_{\text{src}} - \mu_{\text{src}}) \cdot \frac{\sigma_{\text{trg}}}{\sigma_{\text{src}}} + \mu_{\text{trg}} \quad (15)$$

2.3.3 Normalizacja kwantylowa

- **Normalizacja kwantylowa wszystkich pikseli:** Niezależne mapowanie kwantylów histogramów dla każdego kanału RGB:

$$I_{\text{norm}}^c = F_{\text{trg}}^{-1}(F_{\text{src}}(I_{\text{src}}^c)) \quad (16)$$

gdzie F to dystrybuanta empiryczna, c to kanał RGB

- **Normalizacja kwantylowa mapy kolorów:** Tworzenie mapy unikalnych kolorów i ich statystyczna normalizacja z interpolacją kolorów

2.4 Metody segmentacji

Metody te służą do identyfikacji pikseli reprezentujących rzeczywiste zabarwienie tkanki, co umożliwia statystyczne dopasowanie wyłącznie istotnych obszarów obrazu poprzez separację obszarów znacząco zabarwionych (foreground) od tła (background). W implementacji wykorzystano trzy metody klastracji:

- **K-średnich** - partycjonowanie danych na 2 klastry minimalizujące sumy kwadratów odległości punktów od centroidów ich klastrów.
- **Gaussian Mixture Models (EM)** - probabilistyczne modelowanie mieszaniny dwóch rozkładów Gaussa
- **Variational Bayesian GMM (VB)** - bayesowskie podejście z automatyczną regularyzacją

3 Implementacja systemu

3.1 Środowisko i biblioteki

System zaimplementowałem w języku Python 3.12.11 z pomocą Jupyter Notebook z wykorzystaniem następujących kluczowych bibliotek:

- **OpenCV (cv2):** Przetwarzanie obrazów, konwersje przestrzeni barwnych, operacje morfologiczne
- **NumPy:** Obliczenia numeryczne, operacje na macierzach, algebra liniowa
- **Scikit-learn:** Algorytmy uczenia maszynowego (K-means, GMM, Bayesian GMM, ICA, PCA)
- **PyWavelets:** Dekompozycja falkowa i analiza wielorozdzielcza
- **Matplotlib:** Wizualizacja wyników i tworzenie wykresów
- **SciPy:** Funkcje statystyczne, optymalizacja i narzędzia naukowe

3.1.1 Proces analizy

1. **Przygotowanie danych:** Normalizacja i preprocessing obrazów
2. **Dekonwolucja:** Zastosowanie różnych metod ekstrakcji barwników
3. **Segmentacja:** Testowanie jakości segmentacji na znormalizowanych obrazach
4. **Analiza statystyczna:** Porównanie metryk jakościowych (MSE, PSNR, SSIM)
5. **Walidacja krzyżowa:** Testowanie między-partiami (cross-batch validation)

3.1.2 Metryki analizy

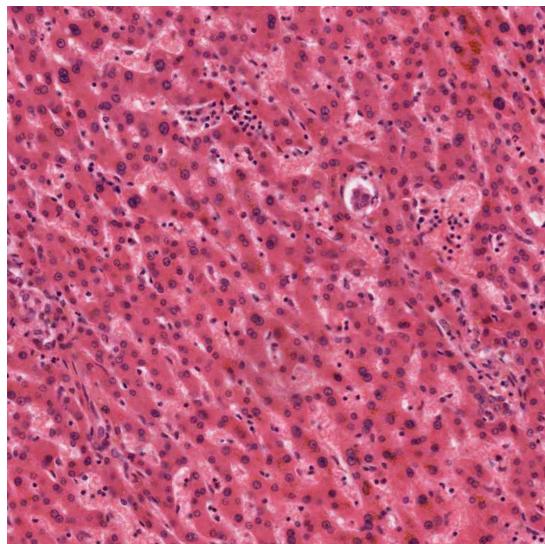
- **MSE (Mean Squared Error)**: Błąd średniokwadratowy rekonstrukcji
- **PSNR (Peak Signal-to-Noise Ratio)**: Stosunek sygnału do szumu
- **SSIM (Structural Similarity)**: Podobieństwo struktur
- **Between-class variance**: Wariancja międzyklasowa barwników
- **Stain correlation**: Korelacja między składowymi barwników
- **Reconstruction accuracy**: Dokładność rekonstrukcji obrazu

Pełny kod jest dostępny na platformie GitHub:

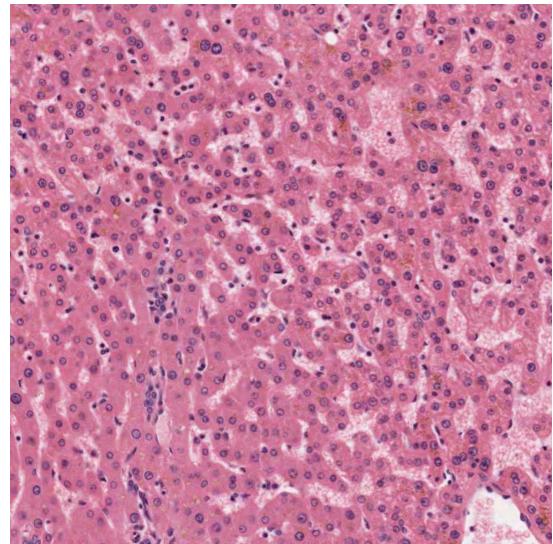
https://github.com/olwozniak/Colour_deconvolution

4 Porównanie metod

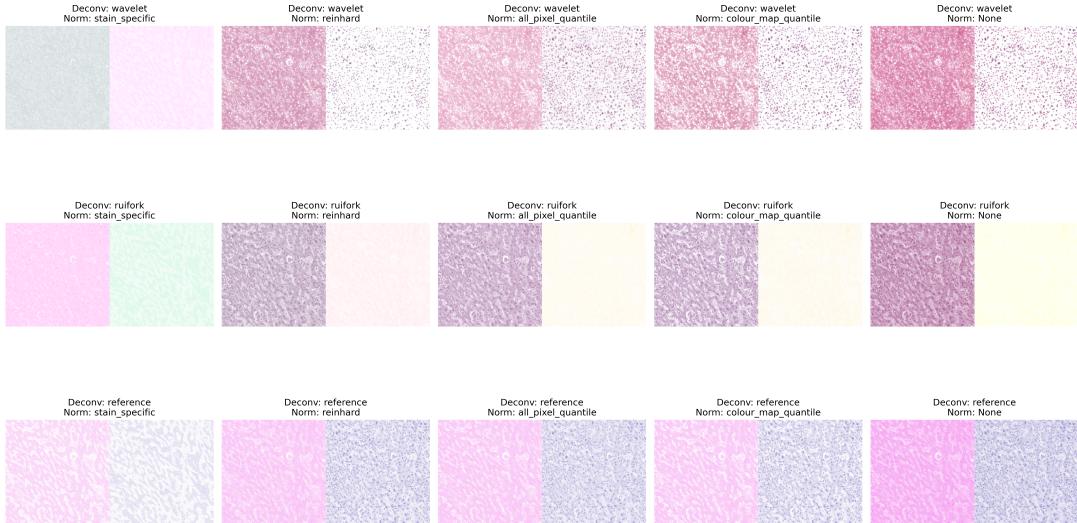
W celu porównania powyższych metod zaimplementowałem je za pomocą języka Python w środowisku Jupyter Notebook oraz przygotowałem skrypty porównujące i wizualizujące efektywność metod. Do analizy wykorzystałem poniższy obraz oryginalny oraz jego wersję wzorcową do procesu normalizacji.



Rysunek 1: Oryginalny obraz

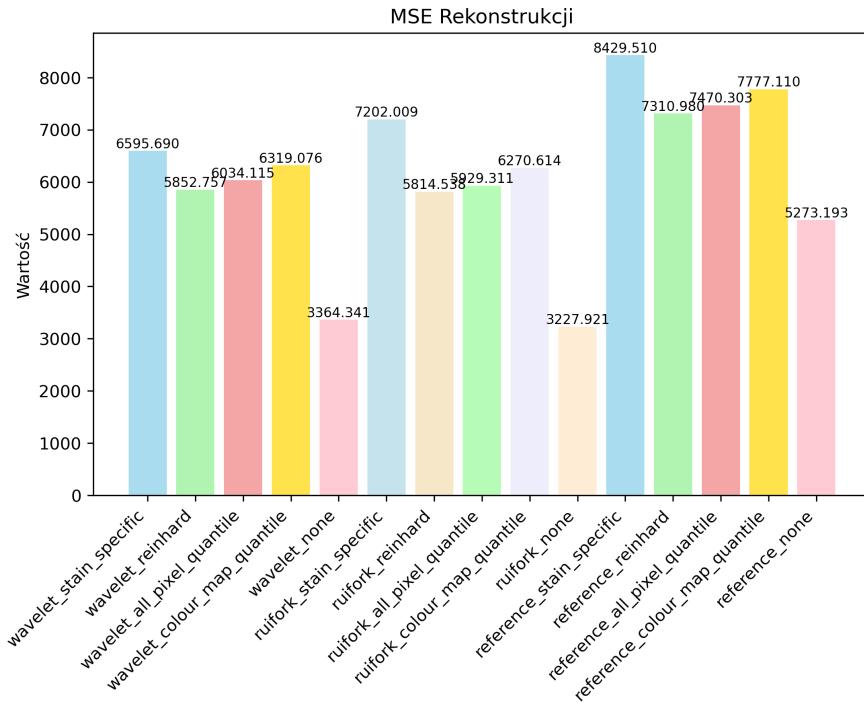


Rysunek 2: Obraz wzorcowy



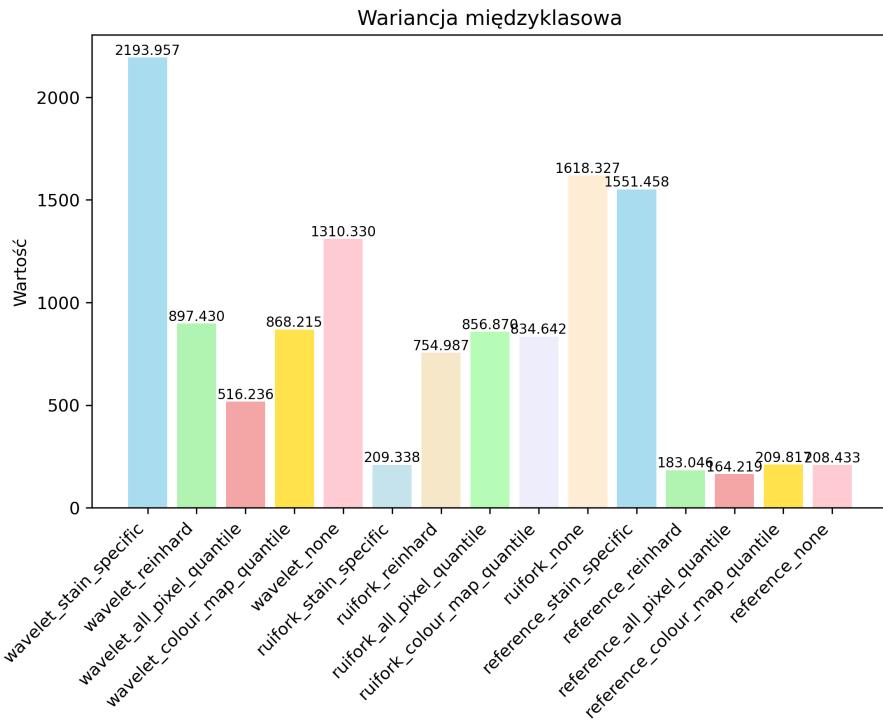
Rysunek 3: Wizualne porównanie metod segmentacji barwników

Klasyczne metody (Ruirok, Reference) dają obrazy najbardziej zbliżone do standardowego barwienia H&E, natomiast metoda falkowa (Wavelet + ICA) pozwala lepiej rozróżnić tekstury i struktury tkankowe, co może być korzystne przy automatycznej analizie obrazu, choć wizualnie odbiega od tradycyjnego wyglądu preparatu. Normalizacja (szczególnie Reinhard) zwiększa porównywalność próbek, zgodnie z teorią wyrównywania rozkładu kolorów.



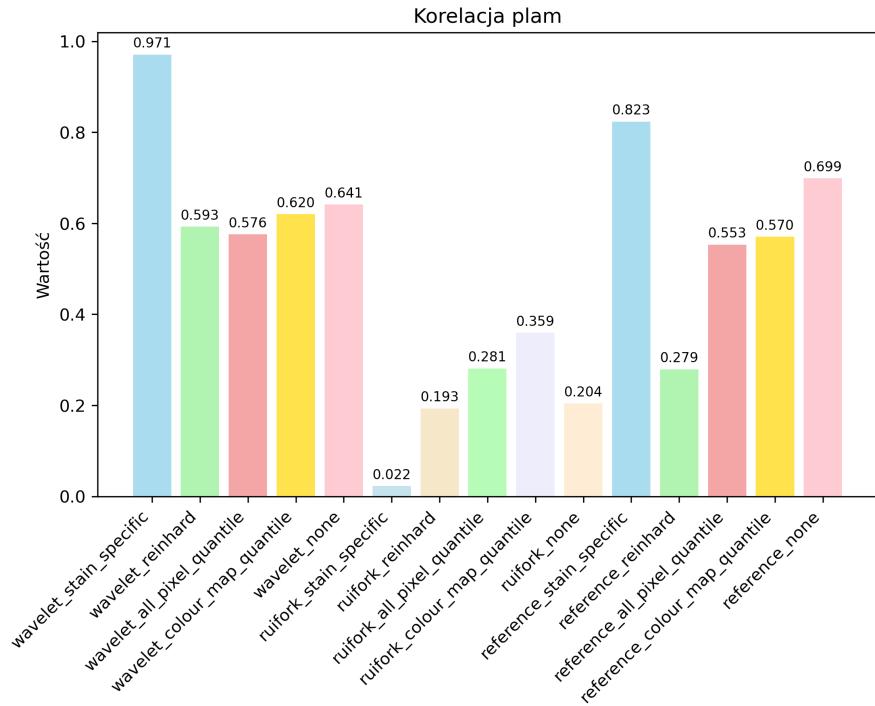
Rysunek 4: MSE Rekonstrukcji

Jeżeli spojrzymy na MSE rekonstrukcji, czyli średni błąd kwadratowy między oryginalnym obrazem a zrekonstruowanym, łatwo zauważyc, że najniższe wartości otrzymujemy dla kombinacji rufork + none i wavelet + none.



Rysunek 5: Wariancja międzyklasowa

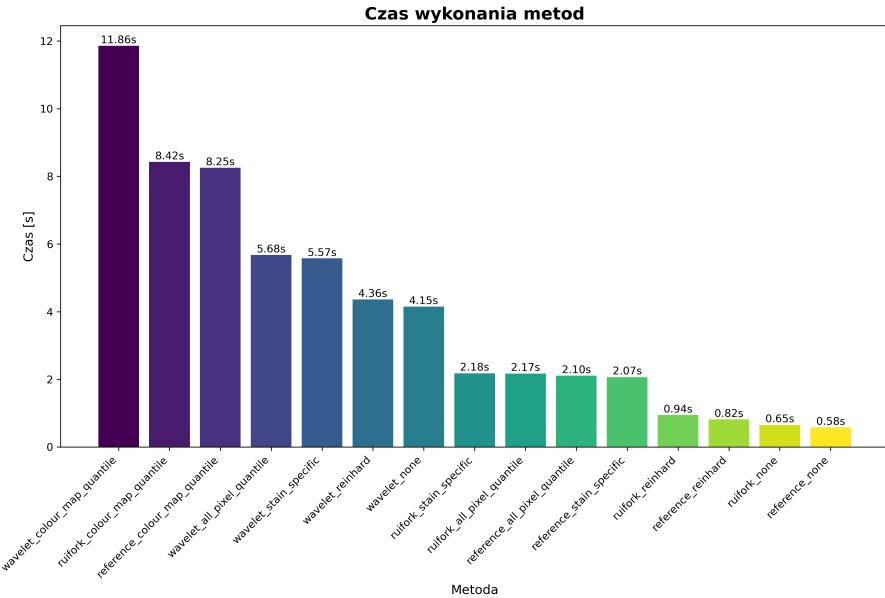
Z drugiej strony, najlepszy wynik dla wariancji międzyklasowej otrzymujemy dla kombinacji wavelet + stain_specific oraz lekko niższe reference + stain_specific lub none. Czyli dla tych metod udaje się najlepiej rozdzielić poszczególne komponenty barwników.



Rysunek 6: Korelacja plam

Jak spojrzymy na korelacje plam, czyli wskaźnik zgodności, między teoretycznym a uzyskanym rozkładem barwnika, można zauważyc, że ponownie zdecydowanie najlepsze wyniki otrzymujemy dla wavelet

+ stain_specific oraz reference + stain_specific, jednak tutaj reference + none ma jeden z najniższych wyników.

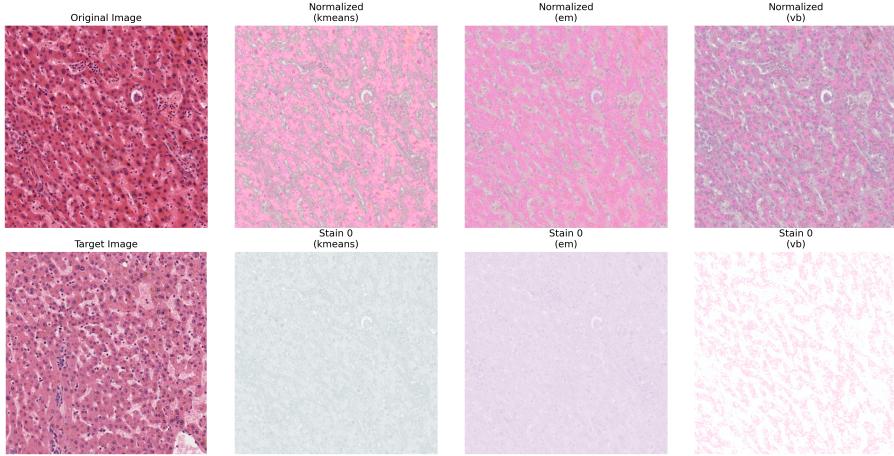


Rysunek 7: Porównanie czasowe

Metoda	Średni czas [s]	Min [s]	Maks [s]	l. pomiarów
Czasy dekonwolucji				
wavelet	6.323	4.150	11.858	5
ruifork	2.874	0.652	8.424	5
reference	2.764	0.582	8.254	5
Czasy normalizacji				
stain_specific	3.272	2.066	5.573	3
reinhard	2.040	0.817	4.360	3
all_pixel_quantile	3.317	2.104	5.675	3
colour_map_quantile	9.512	8.254	11.858	3
none	1.795	0.582	4.150	3

Tabela 1: Czasy przetwarzania metod dekonwolucji i normalizacji

Na podstawie dokładnych pomiarów poszczególnych metod dekonwolucji oraz normalizacji, można zauważać zdecydowanie dłuższy czas przetwarzanie dla metody wavelet oraz colour_map_quantile, co udowadnia nam również wykres.



Rysunek 8: Wyniki różnych sposobów klastracji

Metryka	Najlepsza metoda (wartość)
Separacja plam	vb (356.0462)
Korelacja plam	em (0.0369)
Błąd rekonstrukcji	vb (9612.4493)
Kontrast plamy 0	vb (8.4842)
Kontrast plamy 1	vb (14.1595)
Czas	k-średnich (29.82s)

Tabela 2: Najlepsze metody klastracji

Na podstawie przeprowadzonych testów widać zdecydowaną przewagę klastracji metodą GMM (vb), jednak ma ona również swoje wady, gdyż w testach czasowych wypada zdecydowanie najsłabiej, za to metoda k-średnich wypada zdecydowanie najlepiej czasowo, gdzie GMM potrzebowało aż 203.23s

5 Wnioski

- **Metoda falkowa z ICA** okazała się najbardziej efektywna w separacji składników barwnikowych, osiągając najwyższe wartości wariancji międzyklasowej i korelacji plam, szczególnie w połączeniu z normalizacją stain-specific. Jest to preferowana metoda w zastosowaniach wymagających precyzyjnej analizy składników tkankowych.
- **Normalizacja obrazów** znaczco wpływa na efektywność dekonwolucji. Normalizacja stain-specific dała najlepsze wyniki w poprawie separacji barwników, podczas gdy brak normalizacji (none) zapewnił najlepszą dokładność rekonstrukcji.
- **Kompromis między dokładnością a czasem przetwarzania** jest wyraźnie widoczny. Metoda Wavelet + ICA, choć najwolniejsza, zapewnia najlepszą separację barwników, podczas gdy metoda Ruiforka oferuje najlepszy stosunek jakości do czasu przetwarzania.
- **Metoda referencyjna** stanowi dobry kompromis między wszystkimi analizowanymi parametrami, oferując zrównoważone wyniki we wszystkich kategoriach. Jednak dla innego zestawu danych, zmiana wartości macierzy barwień może okazać się uciążliwa lub powodować niedokładne wyniki.

Literatura

- [1] Ruifrok, A. C., & Johnston, D. A. (2001). Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology*, 23(4), 291-299.
- [2] Alsubaie et al. - 2016 - Stain deconvolution of histology images via independent component analysis in the wavelet domain
- [3] Magee et al. - Colour Normalisation in Digital Histopathology Images
- [4] Hoffman et al. - 2014 - Comparison of normalization algorithms for cross-batch analysis