Unsupervised Clustering for Nuclei Segmentation at Multiple Levels in H&E Histopathology Images

University of Hertfordshire UH

Mahendra Nath Reddy E, Dr. Paul Moggridge;
Department of Computer Science

LinkedIn

Code Repository

Abstract

We propose an unsupervised method for nuclei clustering in H&E-stained histopathology images using three independent approaches: superpixel-level, pixel-level, and stain-level clustering. Each level leverages different features—spatial, textural, and hematoxylin intensity—to generate interpretable nuclei clusters. Our results are visually validated and enable phenotyping and downstream analysis without the need for labelled data.

Project Aim & Objectives

To develop and evaluate an unsupervised, multi-phase clustering-based pipeline for accurate segmentation of individual nuclei in H&E-stained histopathology images.

- Investigate clustering methods: K-Means, Fuzzy C-Means (FCM), Gaussian Mixture Models.
- Apply clustering at three levels: superpixel, pixel, and Hematoxylin channel.
- Isolate and segment individual nuclei through morphological post-processing.
- Use consensus clustering to combine outputs and enhance robustness.
- Evaluate results using instance-level metrics (TP, FP, FN) visually.
- Analyse clustering effectiveness for both tissue-level and nuclei-level segmentation.

Introduction

Nuclei morphology is a key indicator of tissue pathology. Nuclei segmentation in H&E-stained histopathology images is vital for disease grading and tumour analysis.

Supervised deep learning methods require extensive labelled data, limiting scalability. Variability in shape, size, and staining makes automated analysis challenging.

This work proposes an unsupervised, clustering-based framework for nuclei segmentation and phenotype discovery, reducing annotation dependency while preserving biological interpretability.

Methodology

Our work is structured into three phases, progressively refining nuclei segmentation from tissue regions to detailed nuclei-level clusters in H&E-stained histopathology images using unsupervised clustering.

The framework combines superpixel and pixel-level clustering with stain-specific feature analysis for enhanced segmentation accuracy.

Each phase is evaluated visually; TP/FP/FN metrics against ground truth.

Superpixel Clustering:

- SLIC superpixel generation from H&E images
- Cluster mean colours using KMeans, FCM, GMM
- Combine masks via consensus clustering
- Output: Overlay of superpixels + consensus mask

Pixel-Level Clustering:

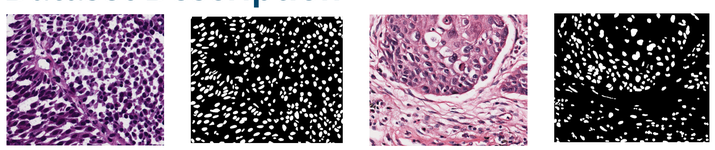
- Flatten the RGB image and normalise pixel data
- Cluster using KMeans, GMM, FCM
- Dice score comparison with ground truth nuclei
- Output: Pixel-wise nuclei cluster masks overlays

Stain Feature Clustering:

- Macenko normalisation + H-channel extraction
- Feature vector: (x, y, intensity)
- Cluster nuclei based on extracted features and evaluate purity
- Output: Nuclei instance, Cluster overlays

1 — 2 — 3

Dataset Description



Bladder (Urothelial Carcinoma) | Liver (Lung squamous Cell carcinoma)

We use the MoNuSeg 2018 (Multi-Organ Nuclei Segmentation) dataset, which provides high-resolution H&E-stained histopathology images with expert-annotated nuclei masks.

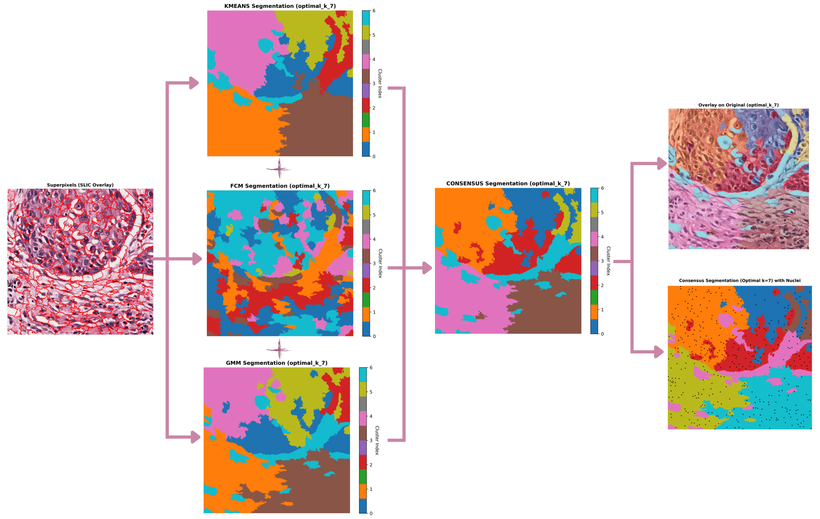
- Contains H&E-stained histopathology images with expert-annotated nuclei masks.
- Hematoxylin and Eosin (H&E) staining enhances contrast for tissue and nuclei structures.
- Tissue samples are from multiple patients with tumours in different organs, from various hospitals.

Phase 1: Superpixel-Level Clustering

SLIC superpixels are generated and clustered using K-means, FCM, and Gaussian Mixture Model (GMM). A consensus mask is created to segment coarse tissue regions.

- Clusters SLIC superpixels to capture tissue regions
- Consensus combines K-Means, FCM, and GMM
- Limitation: Fails to isolate individual nuclei

SLIC Superpixel Generation → Clustering on Superpixels (KMeans, GMM, FCM) → Consensus Mask → Tissue Regions

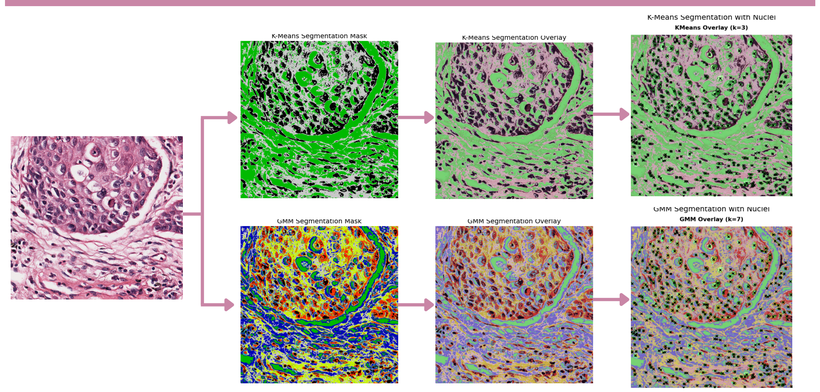


Phase 2: Pixel-Level Clustering

Clustering is applied at the pixel level to refine nuclei localisation, but boundaries remain ambiguous due to overlapping nuclei.

- Pixel-wise clustering improves nuclear region focus
- Only KMeans & GMM used due to memory limits
- Limitation: Boundaries remain blurry; reduced precision

Pixel Flattening → Pixel-wise Clustering (KMeans, GMM) → Reshape Clusters to Image → Nuclei Region Masks

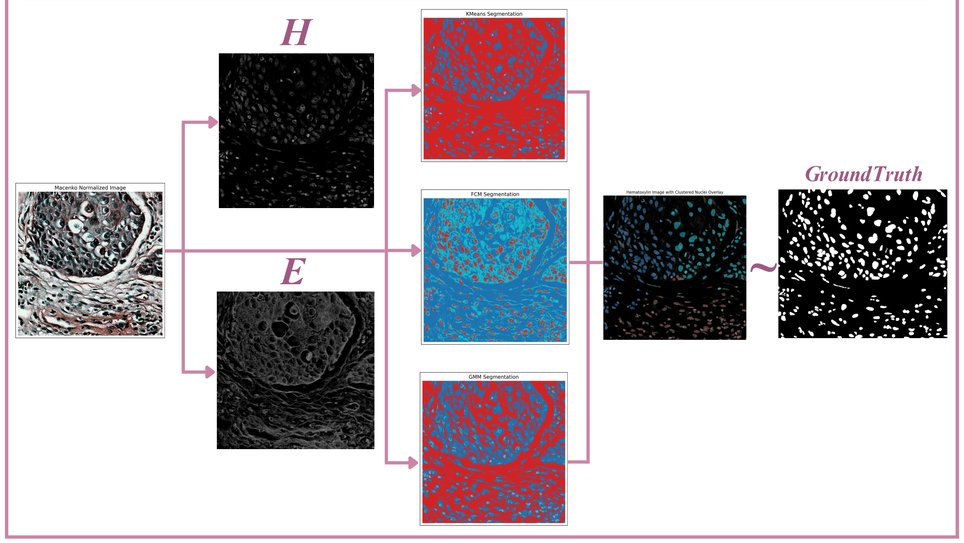


Phase 3: Nuclei Instance Stain Extraction

Color deconvolution isolates the Hematoxylin channel. Clustering on this stain-rich channel captures nuclei shape better, enabling more precise instance-level segmentation.

- Clustering on Hematoxylin channel (nuclei-specific stain)
- Best visual and metric match to ground truth
- Significant TP increase, improved nuclei separation

Input Image → Stain Normalization & Channel Separation → Feature Extraction (Intensity, Edges) → Nucleus-Level Clustering → Nuclei Instance Segmentation



Evaluations and Discussion

Superpixel-level clustering (SLIC):

- Provides spatial and regional grouping
- Lacks fine detail for instance-level separation

Pixel-level clustering:

- Improves focus on nuclear regions
- Boundaries remain fuzzy and less distinct

Hematoxylin (H) channel clustering:

- Leverages stain-specific contrast
- Yields the most accurate nuclei separation

Across all phases, segmentation improves:- True Positives (TP) increase | False Positives (FP) and False Negatives (FN) decrease | Indicates the critical role of feature selection in unsupervised segmentation

Conclusion

This project investigates unsupervised clustering-based nuclei segmentation in H&E-stained histopathology images across three levels: superpixel, pixel, and stain-specific. We demonstrate that:

- Unsupervised clustering can approximate nuclei segmentation without labelled data.
- Phase 3 (H-channel / stain clustering) provides the most reliable instance-level detection.
- The approach offers a scalable alternative for annotation-scarce biomedical image analysis.

Further Work

- Optimised consensus techniques to scale clustering at finer image resolutions, as it was excluded at the pixel level due to high memory and computational costs.
- Apply instance separation, morphological refinement, and shape analysis to enhance segmentation accuracy and interpretability.
- Investigate unsupervised or semi-supervised convolutional neural networks to segment individual nuclei.
- Explore GANs/VAEs for synthetic pathology images to enhance data diversity and clustering robustness, and generalisation.

References

- Kumar, N. et al. (2019). Multi-Organ Nuclei Segmentation Challenge (MoNuSeg). IEEE Transactions on Medical Imaging.
- Macenko, M. et al. (2009). A method for normalising histology slides for quantitative analysis. ISBI.
- Bezdek, J. C. et al. (1984). Fuzzy C-means clustering. Computers & Geosciences, 10(2–3), 191–203.
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer. (Chapter on Gaussian Mixture Models)
- Strehl, A., & Ghosh, J. (2002). Cluster ensembles — A knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research, 3, 583–617.