Cell Detection with Transformers – A Paradigm Shift from Segmentation to Detection in Digital Pathology

Oscar Pina[1], Eduard Dorca[2], Verónica Vilaplana[1]

[1] Image Processing Group (GPI), Universitat Politècnica de Catalunya - BarcelonaTech (UPC), Spain
[2] Anatomia Patològica, Hospital Universitari de Bellvitge (HUB), Spain

Introduction
Cell nuclei detection and classification are fundamental tasks in digital pathology, enabling biomarker quantification and tumor microenvironment analysis. Although the clinically relevant objective is to detect and classify individual cells, these tasks are typically addressed using segmentation-based methods. While effective, segmentation introduces significant computational overhead. To overcome these limitations, we propose CellNuc-DETR, a transformer-based detection model that directly detects and classifies cells, offering a more efficient and scalable alternative to segmentation-based approaches.

Material and methods
CellNuc-DETR employs a Deformable-DETR backbone with a Swin Transformer encoder, leveraging multi-scale deformable attention to detect and classify individual nuclei without requiring pixel-wise segmentation. The model is trained on the PanNuke dataset and evaluated across multiple datasets, including CoNSeP and MoNuSeg, to assess generalization. Efficiency analyses are conducted by benchmarking inference speed on WSIs against segmentation-based approaches.

Results and discussion
CellNuc-DETR achieves an F1-score of 0.84 on PanNuke and 0.78 on CoNSeP, achieving state-of-the-art performance. The model reduces post-processing time by a factor of 40 compared to HoVer-Net and achieves a 2.5× speed-up in inference time compared to CellViT while maintaining comparable classification accuracy. The in-device sliding window approach further enhances efficiency, enabling rapid WSI processing with minimal GPU overhead.

Conclusion
By eliminating complex post-processing steps and reducing computational demands, CellNuc-DETR enables scalable, high-throughput cell analysis in digital pathology. These results advocate for a paradigm shift from segmentation-based methods to transformer-based detection approaches, offering a practical and efficient solution for WSI analysis in both clinical and research settings.

Key words: Cell Detection, Transformers, Deep Learning, Whole Slide Image