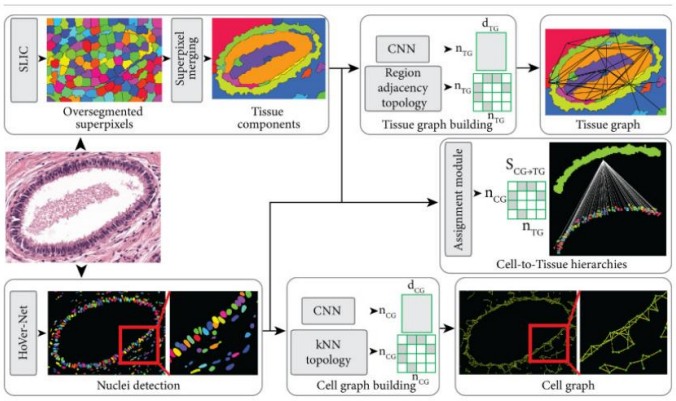
IEEE PAPER SUMMARY

Integration of Deep Learning and Graph Theory for Analysing Histopathology Whole-slide Images

PUBLISHED BY: HYUNG JUNG *et al*



SUMMARISED BY :

DHATHRI GURRAM **PES1UG20CS141**

GAURAV DNYANESH MAHAJAN **PES1UG20CS150**

SECTION C

ABSTRACT:

Characterization of collagen deposition in immunostained pictures is relevant to a variety of clinical diseases, particularly HIV infection. To accomplish quantitative diagnosis, accurate segmentation of these collagens and extraction of representative features of underlying illnesses are critical steps. While a first order statistic obtained from segmented collagens can be useful in describing progression of the disease over time, it does not reflect morphological alterations or spatial arrangements.

AIM:

We show how to extract essential histopathology signals reflecting underlying disease development from histopathology **whole-slide images (WSIs)** using a comprehensive workflow. Good results have been achieved in agreement with the expected pathology.

Model used is:

- CNN for histopathological WSI segmentation.
- Parallel processing is applied to convert 100K ~ 150K segmented collagen fibrils into a single collective attributed relational graph.
- graph theory is applied to extract topological and relational information from the collagenous framework.

BACKGROUND:

Histopathology studies signs of disease using the microscopic examination of a biopsy or surgical specimen. The sections are dyed with different stains to visualize different components of the tissue microscopically. The underlying tissue architecture is preserved, allowing for a thorough perspective of illness and its effects on tissues. As a result, histopathology image diagnosis remains the "gold standard" in diagnosing a wide range of illnesses, including nearly all kinds of cancer.

Previously, some articles in the field of computational image analysis used "hand-crafted" methods to do quantitative analysis of histopathology pictures. To capture a wide spectrum of variability, the "hand-crafted" approach may necessitate considerable efforts.

WHY DEEP LEARNING?

Deep learning has had numerous successes in image classification challenges, as demonstrated by AlexNet, VGGNet, and ResNet as well as biomedical image segmentation tasks. Deep learning's capacity to derive high-level features that generalise effectively in a data-driven manner is the primary driver of its unheard-of performance.

Histopathology images have distinctive characteristics which set them apart from other images. These traits are high resolutions, complex appearances, diverse magnifications, various stains, and the corresponding differences in semantic interpretations.

HOW GRAPH THEORY?

One of the most promising methods for histopathology image analysis is the use of graph-based techniques as they are flexible fictive tools receiving major interest for their expressive capability in modelling topological and relational information between image components.

Graph-based methods are also frequently applied to regions of interest (ROIs) sampled from histopathology WSIs due to the limited computational resources and lack of advanced parallel processing techniques, limiting comprehensive views of the tissue structure, organization, and a deeper understanding of the underlying disease conditions, which are the main advantages of analysing the WSIs.

METHODOLOGY:

• Ultrahigh resolution images at 20x magnification are divided into small non overlapping patches for analysis.
• The initial ground truth labels are utilized for training a SVM to generate additional ground truth labels.
• The implementation utilizes 3 x 3 convolutions with reflective padding to keep input and output image sizes same.
• Emphasis is made on demonstrating an integrated approach using deep learning and graph theory for analysing network-forming collagens.
• Performance comparison between different CNN models for analysing a collagenous network has also been addressed.
• Dice coefficient loss has been utilized as a loss function and minimized during the training.
• During the training, a model with the highest validation dice score has been saved and it achieved 0.9444 dice score for the testing set.
• Segmentation patches created by the neural network are collected and stitched together to generate a segmentation mask of the input WSI.

Graph Theory:
• Each segmented collagen is converted into an undirected attributed relational graph.
• Before conversion connected components are labelled using 8-connectivity.
• Nodes and branches have been assigned depending on the number of adjacent neighbours and converted into the ARGs
• The nodes, edges, and edge lengths are saved in comma separated value (CSV) formatted files.
• Then the recorded values for each WSI are combined into slide-level undirected graph

objects.
• Global features are computed for each connected component graph object, using parallelization in the Python multiprocessing module, and then aggregated for the whole slide image.

RESULT AND CONCLUSION:

RESULT:

- Histopathology WSIs of rhesus macaques at different time points (healthy, post infection of SIV, and after treatment using combination antiretroviral therapy (cART) with antifibrotic agent) have been analysed using the pipeline proposed in this work.

- A collagen density heatmap is generated based on the ratio between the collagen positive pixels and the total number of pixels within a 50 × 50 pixel area (2,500 pixels). The follicles located in the medullary cords area in the SIV infected lymph node, which were manually identified and reviewed by pathologists, show lower collagen density compared to the surrounding compartments
- After treatment with the cART and antifibrotic agent, the collagen density decreased significantly in the medullary cords where the collagens are densely deposited, which corresponds to the expected disease progression.

- Global features that are extracted using the collective ARGs reveal underlying changes of the collagen deposition. The diameter and radius (the maximum and minimum shortest path lengths from one node to any other node in the graph, respectively) increase after the SIV infection, indicating the elongation of the collagens.
- The giant connected component ratio (GCCR, a ratio between the number of nodes in the largest connected component in the graph and total the number of nodes) shows a different trend compared to the previous features: a significant increase after the treatment. This observation indicates that the collagen with the largest number of nodes relatively remain the same, while the collagens with smaller number of nodes are treated more effectively resulting in a decrease in the overall total number of nodes.

CONCLUSION:

- The authors have proposed an integrated approach that combines collagen segmentation generated by a CNN with graph theoretic post-processing. This enables extraction of quantifiable measures from immunohistochemistry stained WSIs to evaluate pathogenicity induced by collagen deposition at the scale of whole tissue sections.
- The segmented collagens within the WSI are converted into a single collective ARG using parallel processing and then global features are extracted.

- The global features show the clear advantage of capturing morphological evolutions of each collagen fibre and the spatial arrangements associated with the underlying pathological changes.

FUTURE:

Future research will focus on extracting and contextualizing key quantifiable histopathology features representing underlying disease progression. This will greatly facilitate the analysis of various meshwork-structures in histopathology whole-slide images in clinical applications.

Extension of the proposed method into 3D volume and applying different CNN models should further increase the accuracy of estimating biological changes occurring in the tissue samples.

Additionally, other deep learning models like Long Short-Term Memory [LSTM] and Gated Recurring Unit [GRU] networks, which have been known to provide a great accuracy on image data could be deployed to obtain better accuracy.