Self-supervised graph representations of WSIs

Oscar Pina OSCAR.PINA@UPC.EDU
Verónica Vilaplana VERONICA.VILAPLANA@UPC.EDU
Technical University of Catalonia

Abstract

In this manuscript we propose a framework for the analysis of whole slide images (WSI) on the cell entity space with self-supervised deep learning on graphs and explore its representation quality at different levels of application. It consists of a two step process in which the cell level analysis is performed locally, by clusters of nearby cells that can be seen as small regions of the image, in order to learn representations that capture the cell environment and distribution. In a second stage, a WSI graph is generated with these regions as nodes and the representations learned as initial node embeddings. The graph is leveraged for a downstream task, region of interest (ROI) detection addressed as a graph clustering. The representations outperform the evaluation baselines at both levels of application, which has been carried out predicting whether a cell, or region, is tumor or not based on its learned representations with a logistic regressor.

Keywords: Computational histopathology, graph neural networks, self-supervised learning

1. Introduction

The digitization of pathology slides into multi-resolution whole-slide images (WSI) has revolutionized the clinical practitioners' work routine and has become a standard method for disease diagnosis and characterization by allowing experts to work at different scales and resolutions. In addition, its adoption has triggered the field of computational histopathology, which explores the application of machine learning algorithms, among others, for computer-aided diagnosis from pathology slides.

Convolutional neural networks (CNN) have been successfully employed at all ranges of applications, from entity detection and classification (i.e. cell detection) to WSI classification (Srinidhi et al., 2021). WSI processing is usually carried out by splitting the slide into smaller image patches that are independently processed by a CNN and finally aggregated if the application requires it. Recently, graph representations of histology images have been proposed in order to model the relationships between biological entities of the slide, such as cell graphs (CG) (Zhou et al., 2019), (Jaume et al., 2021b) and tissue graphs (TG) (Pati et al., 2022), (Anklin et al., 2021), as well as to leverage patch positional and contextual information with patch graphs (PG) (Levy et al., 2021). The nodes of a CG are the cells of an image, and its connectivity usually is defined by connecting those cells that are close to each other (i.e. kNN graph, radius graph), assuming that nearby cells are likely to interact. However, a WSI can contain thousands or millions of cells, which presents computational challenges, so this approach is commonly addressed with smaller image tiles. Alternatively, PGs are intended to represent larger images, such as WSIs, by creating a graph whose nodes are patch representations obtained with a CNN in a previous step, and connecting

© 2022 O. Pina & V. Vilaplana.

those patches that are adjacent or overlap, so that the patch representations can be extended to include neighborhood information. The analysis is addressed with Graph Neural Networks (GNN) (Hamilton, 2021), a framework for the application of neural networks to graph structured data. Additionally, there is an increasing literature on the interpretability of this family of models, which encourages their usage in medical applications (Jaume et al., 2021b).

We propose a GNN framework for the processing of WSIs at cell-level and apply it to the automatic detection of regions of interest (ROI), which is crucial towards developing reliable computer-aided diagnoses methods that can contribute to pathologists' practice. (Anklin et al., 2021), (Ozen et al., 2021). In general, ROIs can be of any size and any shape, which adds an extra level of difficulty to the task. We define a graph level GNN encoder to embed local CGs created with subsets of cells to a vector space. Afterwards, a graph is created from a WSI employing the previous representations as initial node attributes, in order to perform a graph clustering into different ROIs. The models are trained in a self-supervised manner, firstly by maximizing the mutual information (MI) between local and global representations and subsequently by optimizing the spectral modularity of the cluster assignments. The representations are firstly evaluated at cell level to detect tumor cells with a dataset of lung WSIs and patches with cell annotations and at region level to predict tumor regions in breast cancer WSIs, which is also addressed as a clustering task.

2. Methodology

In this section we describe the steps of the proposed method. In Section 2.1 we introduce the preprocessing steps needed, namely background removal and cell detection. Then, we explain how these cells are spatially clustered to create subgraphs of cells that represent image regions in Section 2.2 and how these subgraphs are embedded to vector representations with a GNN encoder in Section 2.3. In Section 2.4 we describe how the vector representations of these subgraphs (or regions) are employed to define a graph that represents a WSI and finally, in Section 2.5, this graph is leveraged to carry out a cluster analysis for ROI detection.

2.1. Preprocessing

Given an input WSI, the preprocessing step consists of the nuclei detection and characterization. The positions of the N cells nuclei $P = \{p_i\}_{i=1}^N$ are extracted with a HoVerNet (Graham et al., 2019) trained on the PanNuke (Gamper et al., 2020) dataset provided by Histocartography package (Jaume et al., 2021a). In this initial project, the input cell features $X = \{x_i\}_{i=1}^N$ are their morphological features such as of the nuclei and the convex hull, eccentricity, diameter, extent, major and minor axis, perimeter and solidity.

Prior to feeding the HoVerNet for nuclei detection, we must split the image into smaller patches and remove those that correspond to background and contain no information. In order to optimize this process, we first run a tissue detection algorithm on a downsampled version of the WSI to shortlist the set of candidate patches. Afterwards, the amount of tissue of the remaining patches is estimated and those with less than 20% of tissue are discarded as well.

2.2. Local subgraph representation

The WSI is represented as a set of overlapping subgraphs of cells. To define the subgraphs, we cluster the cells following the implementation proposed in (Lu et al., 2020), but performing a soft cluster assignment that yields to overlapped clusters. The motivation roots in the fact that it allows different samples, or subgraphs, to share information, which is more convenient in next steps. Formally, given the set of centroid positions of all cells in the WSI, $P = \{p_i\}_{i=1}^{N}$, we randomly subsample them by a factor of r_s, which is a hyperparameter, so that it results in a subset of $P' = \{p_j\}_{j=1}^{r_s N}$. Afterwards, agglomerative clustering is applied to cluster the cells of P' into K clusters based on their spatial proximity, $C' = \{c_j\}_{j=1}^{r_s N}$, $c_j \in \{1, ..., K\}$. The number of clusters K is fixed as a small ratio r_c of the amount of subsampled cells $K = r_c|P'|$, bein r_c another hyperparameter. Finally, for every cell i, $p_i \in P$, we get the set of labels of its $k_{overlap}$ nearest neighbors, which leads to an overlapping clustering when the nearest neighbors in P' of a cell in P are allocated to distinct clusters. The connectivity of every subset of cells is generated with the Delaunay triangulation algorithm. The final result is a set of K overlapped subgraphs $\mathbf{G} = \{G_k = (X_k, P_k, A_k)\}_{k=1}^{K}$.

2.3. Local subgraph encoder

The objective is to learn a function $f_\Theta : (X, P, A) \in \mathbb{R}^{|V_G|\times d} \times \mathbb{R}^{|V_G|\times 2} \times V_G \times V_G \to \mathbb{R}^D$ with a neural network parametrized by Θ that embeds a subgraph $G = (X, P, A)$ to one single vector representation $\mathbf{z}_G \in \mathbb{R}^D$.

2.3.1. Encoder architecture

The mapping f_Θ is a composition of a permutation equivariant message passing encoder h_ϕ and a permutation invariant global pooling g_ψ, $f_\Theta = g_\psi \circ h_\phi$. The message passing h_ϕ has been implemented with a L-layer GIN (Xu et al., 2019), whose node-update equation is $x_i^{(l+1)} = h_\phi^{(l)}\left((1+\epsilon)x_i^{(l)} + \sum_{j\in N(i)} x_j^{(l)}\right)$. The equivariant encoder h_ϕ is completed with a LSTM-based jumping knowledge (Xu et al., 2018) layer that obtains the final representation of node i, z_i, taking the output after every message passing layer $x_i^{(1)}, ..., x_i^{(L)}$ and performing a weighted sum $z_i = \sum_l s_i^{(l)} x_i^{(l)}$. The coefficients $s_i^{(l)}$ are calculated with a bidirectional LSTM so that the network is able to attend at different hops of the node's neighborhood.

Finally, the graph-level representation $\mathbf{z}_G$ is obtained with the permutation invariant global pooling g_ψ, which has been implemented as a concatenation of *mean*, *variance*, *max* and *min* poolings followed by a fully connected neural network (FCNN) that projects the aggregations to a lower dimensional space:

$$\mathbf{z}_G = g_\psi\left(\frac{1}{|V_G|}\sum_{i\in V_G} z_i \;||\; \frac{1}{|V_G|}\sum_{i\in V_G}(z_i-\mu)^2 \;||\; \max_{i\in V_G} z_i \;||\; \min_{i\in V_G} z_i\right) \tag{1}$$

2.3.2. Encoder self-supervised training

The lack of fine-grained annotations motivates the usage of self-supervised methods to train f_Θ. We have employed InfoGraph (Sun et al., 2019) scheme, which is proposed as an extension of Deep Graph Infomax (Veličković et al., 2019) to learn graph-level embeddings

by maximizing the MI between global and local representations. Concretely, given a dataset of graphs $\mathbf{G} = \{G_k \in \mathbf{G}\}_{k=1}^{K}$, the parameters θ are optimized by maximizing:

$$\widehat{\theta}, \widehat{\vartheta} = arg\ max \sum_{G \in \mathbf{G}} \frac{1}{|V_G|} \sum_{i \in V_G} I_{\theta,\vartheta}(\mathbf{z}_i; \mathbf{z}_G) \tag{2}$$

where $I_{\theta,\vartheta}$ is the neural MI estimation function, which is calculated by a neural network discriminator T_ϑ parametrized by ϑ.

2.4. WSI graph representation

The embeddings of the local subgraphs of a WSI, $\mathbf{Z} = \{\mathbf{z}_k\}_{k=1}^{K}$ and their positions $\mathbf{P} = \{\mathbf{p}_k\}_{k=1}^{K}$, where every $\mathbf{p}_k$ is computed as the centroid of the convex hull created by the cells belonging to subgraph k, are employed to construct a graph that represents the WSI $\mathbf{G} = (\mathbf{Z}, \mathbf{P}, \mathbf{A})$. The nodes of the graph are the subgraphs of cells previously defined, their initial node attributes are the embeddings obtained after the encoder f_Θ of Section 2.3 and its connectivity is defined as the Delaunay triangulation of the convex hull centroids $\mathbf{P}$, following the method of (Lu et al., 2020). Moreover, we apply a threshold to every connection to remove those links connecting distant nodes.

2.5. WSI graph clustering

The representation of the WSI is leveraged to carry out an automatic ROI detection process addressed as a self-supervised task with a clustering algorithm. On the one hand, classic clustering methods such as k-Means on the embeddings $\mathbf{Z}$ would ignore the global positional information $\mathbf{P}$ of these convex hulls. On the other hand, graph clustering methods use the graph topology $\mathbf{A}$ but may fail at efficiently including the representations $\mathbf{Z}$, which in general can be high dimensional.

In order to consider both the node embeddings and their local context, we propose to employ a GNN-based clustering method. Concretely, Deep Modularity Network (DMoN) (Anton Tsitsulin and Müller, 2020), which given the graph $\mathbf{G}$ groups the nodes into M clusters by obtaining a soft assignment matrix $\mathbf{S} \in \mathbb{R}^{|V_\mathbf{G}| \times M}$ with a message passing encoder f_φ followed by a FCNN f_τ and a softmax activation function,

$$\mathbf{S} = softmax(f_\tau(f_\varphi(\mathbf{Z}, \mathbf{A}))) \tag{3}$$

We have followed DMoN authors' instructions (Anton Tsitsulin and Müller, 2020) and the encoder f_φ has been implemented with a T-layer Graph Convolutional Network (GCN). The cost function of the model is

$$\mathcal{L}_{DMoN} = -\frac{1}{2|E_\mathbf{G}|} tr(\mathbf{S}^T \mathbf{B} \mathbf{S}) + \frac{\sqrt{M}}{|V_\mathbf{G}|} \left\| \sum_i \mathbf{S}_i^T \right\| - 1 \tag{4}$$

where B is the modularity matrix $B = A - \frac{dd^T}{2|E_\mathbf{G}|}$. The first term of the equation is derived from a spectral relaxation of the modularity maximization problem and the second one is intended to avoid the cluster assignments collapsing to one single cluster (Anton Tsitsulin and Müller, 2020). Finally, cluster-level representations and adjacency can be obtained with $\mathbf{H}_M = \mathbf{S}^T \mathbf{H}$ and $\mathbf{A}_M = \mathbf{S}^T \mathbf{A} \mathbf{S}$, where $\mathbf{H}$ is the matrix of node embeddings after f_φ.

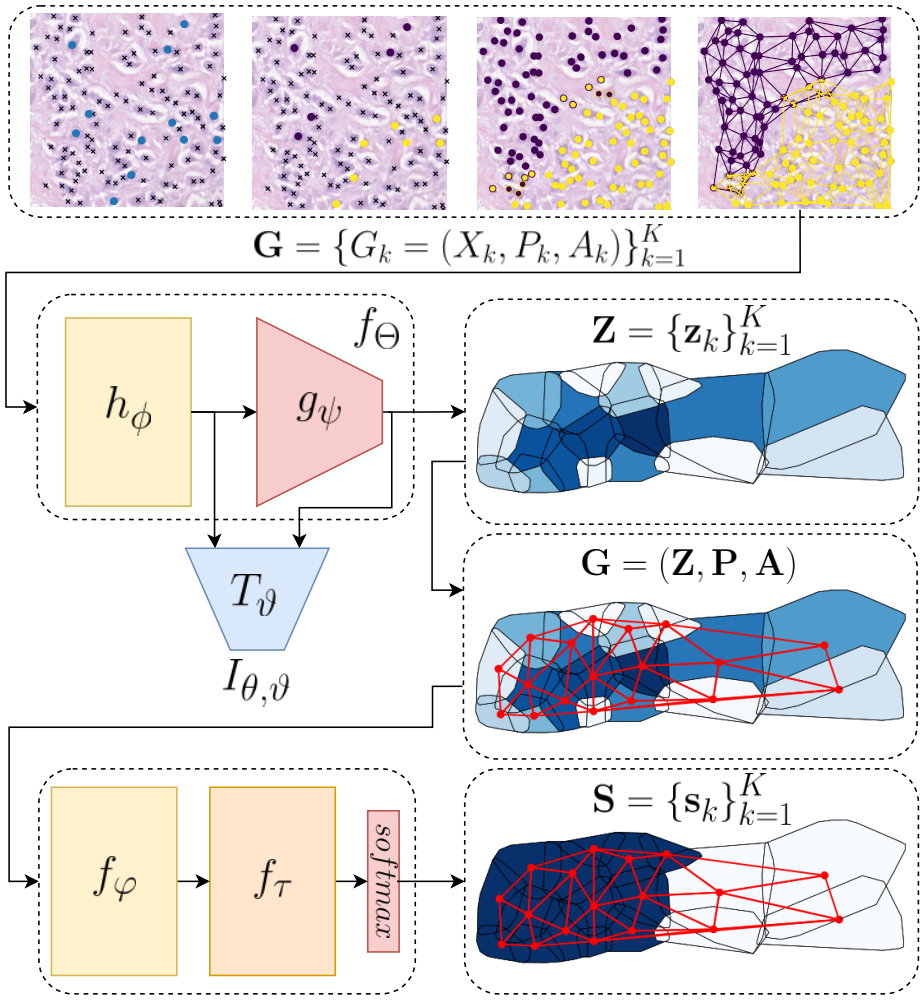


Figure 1: Overview of the proposed method.
(i) The cells are clustered with overlapped cluster assignments and a graph is created for every cluster with the Delaunay triangulation algorithm $\mathbf{G} = \{G_k\}$. (ii) These are fed into an encoder which outputs a vector representation for every graph $\mathbf{Z} = \{\mathbf{z}_k\}$ that can be seen as regions of the image (iii). (iv) Delaunay triangulation between regions centroids creates a graph representing the WSI. (v) The WSI graph is input to a GNN that performs soft cluster assignments, (vi) $\mathbf{S} = \{\mathbf{s}_k\}$.

3. Experiments and Evaluation

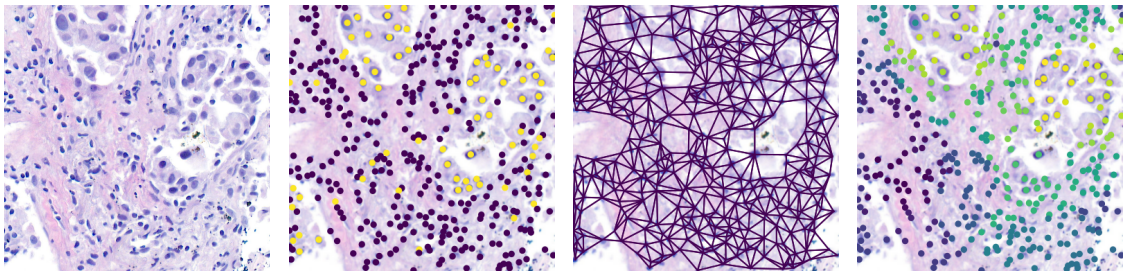
3.1. Data

The data consists of a set of 20 H&E breast cancer and lung cancer WSIs of average size 200.000×100.000 pixels provided by *Institut Català de la Salut (ICS)*. Breast slides include partially annotated regions as tumor or normal. Concretely, we have extracted a total of 8728 image cell clusters from breast slides, from which 5948 have no annotations and are employed to train the encoder and 2780 are labelled as tumor or normal. On the other hand, lung slides have no annotatios, instead, we have a set of patches extracted from these images with cell level annotations, also classified as tumor or normal.

3.2. Self-supervised local embedding

In this section we evaluate the quality of the self-supervised embeddings. When defining the architecture of $f_\Theta = g_\psi \circ h_\phi$, we have set $L = 5$ message passing layers and 1 layer LSTM for h_ϕ, and the global pooling projection g_ψ is obtained with a 2-layer FCNN. We have trained two encoders separetely, one with the breast WSIs with no annotations and another one employing the lung slides. The set of cells P of every WSI is subsampled by a ratio of $r_s = 1 \times 10^{-2}$ to settle P' and carry out the hierarchical clustering to a set of $K = r_c|P'|$ clusters, $r_c = 2 \times 10^{-1}$. For training, have employed Adam optimizer, a learning rate of 1×10^{-3} and a batch size of 128. We perform independent evaluations at cell level and region level by taking the representations, fitting a logistic regressor with 20% of the annotated data and predicting on the remaining 80%, as it is usually done in self-supervised learning literature.

Lung slides, cell level evaluation. The evaluation is performed at cell level employing the annotated patches of lung slides and the encoder trained on lung WSIs. A cell graph is created for every patch with the Delaunay triangulation and they are fed into the trained encoder. The process is shown in Figure 2. The node level embeddings z_i output by h_ϕ are taken and used to predict whether the corresponding cell (node) is tumoral or not with a linear classifier. The results are contrasted with using the raw cell morphological features x_i in Table 1.



(a) Image patch (b) Cell annotations (c) Cell graph (d) Cell embeddings

Figure 2: Lung cell graph.
(a) Lung 1024×1024 pixels patch. (b) Cell level annotations as tumor (yellow) or normal (blue). (c) Cell graph connectivity. (d) Cells colored according to their embedding z_i.

Breast slides, region level evaluation. Given the available region annotations of 8 WSIs, we label every subgraph $G_k = (X_k, P_k, A_k)$ as tumor or normal. We evaluate the subgraph embeddings $\mathbf{z}_k$ by predicting these labels with a linear classifier given $\mathbf{z}_k$. Additionally, in Table 2 we compare the performance with two other input spaces: (i) the number of cells $|V_k|$ and (ii) the concatenation of *mean*, *variance*, *max* and *min* of the cell raw input features X_k, $\overline{\mathbf{x_k}}$.

Table 1: Linear model prediction performance for tumor cells.

Input	Accuracy	Precision	Recall	F1
Random	50.0 ± 1.0	34.3 ± 1.4	49.5 ± 1.6	40.5 ± 1.4
x_i	78.6 ± 0.6	76.2 ± 3.1	54.7 ± 1.9	63.7 ± 1.2
z_i	$\mathbf{87.5 \pm 0.7}$	$\mathbf{83.3 \pm 1.2}$	$\mathbf{79.5 \pm 1.5}$	$\mathbf{81.3 \pm 0.9}$

Table 2: Linear model prediction performance for tumor regions.

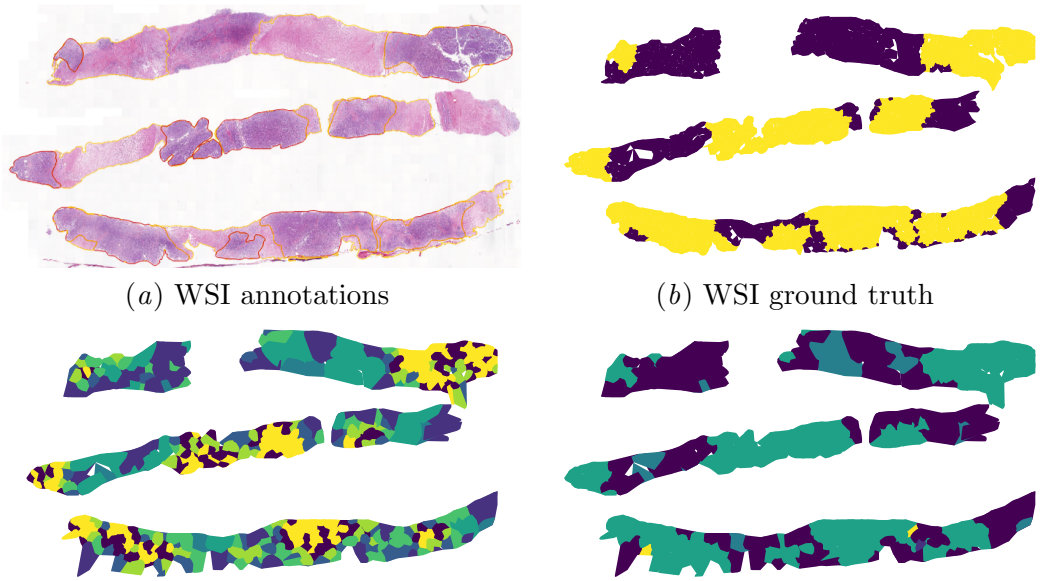
Method	Accuracy	Precision	Recall	F1
Random	50.4 ± 0.6	79.8 ± 0.5	50.5 ± 0.9	61.9 ± 0.7
$\mathbf{\lvert V_k \rvert}$	72.0 ± 0.6	72.2 ± 0.6	$\mathbf{100.0 \pm 0.0}$	83.7 ± 0.4
$\overline{\mathbf{x_k}}$	78.6 ± 0.6	79.8 ± 1.3	94.6 ± 2.0	86.4 ± 0.2
$\mathbf{z_k}$	$\mathbf{86.5 \pm 0.6}$	$\mathbf{88.7 \pm 0.9}$	92.7 ± 1.2	$\mathbf{90.6 \pm 0.4}$

3.3. Clustering for ROI detection

We have trained an independent model for every WSI forwarding the entire graph at once, which is computationally tractable, in order to avoid graph-based batching strategies for node-level tasks that could influence clustering performance. The encoder f_φ consists of 1 message passing layer. The model has been trained with Adam optimizer, a learning rate of 1×10^{-4} and a weight decay of 2×10^{-1}. The clustering performance is evaluated qualitatively by visual inspection of the assignments, shown in Figure 3 and quantitatively in terms of normalized mutual information (NMI), adjusted rand score (ARI) and F score with the binary ground truth labels, exposed in Table 3.

Table 3: Clustering performance.

Clustering	ARI	NMI	F
Random	0.0 ± 0.0	0.0 ± 0.0	11.6 ± 0.4
k-Means	18.3 ± 6.3	19.5 ± 6.5	30.3 ± 3.6
DMoN	$\mathbf{19.8 \pm 20.2}$	$\mathbf{21.2 \pm 15.7}$	$\mathbf{60.9 \pm 10.3}$



(*a*) WSI annotations (*b*) WSI ground truth

(*c*) k-Means clusters (*d*) DMoN clusters

Figure 3: WSI clustering

4. Discussion

Cell clustering strategy Aiming to reduce the size of the WSIs that can contain millions of cells, these are split into small spatial clusters. An alternative method could be to split the WSI into small image patches and define a cell graph for every one of them. Indeed, patches are more frequent in the literature to encourage the usage of CNNs. However, the grid structure of patches does not align with the cell distribution, so the process can lead to graphs corresponding to stroma regions with almost no cells. Spatial clustering of the cells, instead, ensures the presence of cells at every sample.

Hyperparameters At the time of subsampling the cells to cluster them and create the subgraphs, the hyperparameters r_s and r_c must be chosen in a balance between computation and space. Being n the number of data points, the time and space complexity of the hierarchical clustering are $O(n^3)$ and $O(n^2)$, respectively. Taking into account the large number of cells contained by a WSI, the ratio r_s must be low enough to make the clustering feasible. Once the subsampling ratio is fixed, the number of clusters ratio r_c plays the role of determining how many cell clusters will the process output, and therefore, how big are they going to be. The value $r_c = 2e - 1$ is the minimum we could set to obtain meaningful clusters, otherwise the process could lead to many one-cell clusters.

Cell features In our experiments, we have employed simple cell morphological features as initial cell attributes rather than more sophisticated representations, such as visual features over a small patch centered on the cell-nuclei (Jaume et al., 2021a). The reason is that these morphological features actually align with pathologists' methods and we are interested in demonstrating the usefulness of the contextualization effect of GNNs in this domain.

Self-supervision The GNN encoder has been trained in a self-supervised manner due to the lack of annotations, which results in a task agnostic encoder that is able to characterize distinct, any-shape regions of a WSI based on their cell distribution. The evaluation of the method at both cell level and region level demonstrates that that the model captures cell environmental information and successfully aggregates it to obtain coarse region embeddings, yielding to multi level, hierarchical representations. We have implemented the most simple version of InfoGraph (Sun et al., 2019), but this could be extended with, for instance, prior matching and feature disentanglement. Alternatively, non-contrastive approaches could also be explored.

Linear evaluation As it is common in the self-supervised learning literature, we have followed the linear evaluation protocol to evaluate our method, which basically consists of freezing the encoder weights, getting the representations and fitting a linear classifier with a small part of the available annotations. The complexity of the tasks addressed, how the input data is processed (i.e. cell clusters that lead to any-shape regions rather than squared patches), the cell features employed, the self-supervision training scheme and the scarce annotations available for the evaluation difficult the settlement of a baseline. For this reason, we evaluate the model with respect to the raw input features.

Graph clustering Clustering is addressed with GNNs to leverage both node feature and topological information. Intuitively, the cluster assignments shown in Figure 3 suggest that DMoN clustering returns well defined regions, which makes the algorithm more suitable for ROI detection, whereas the output of k-Means is a noisy cluster assignment that is not coherent with the graph topology. Quantitatively, DMoN also outperform k-Means in terms of NMI, ARI and F scores. The high standard deviation in the metrics of DMoN is caused by a low performance in a subset of the slides that are almost completely annotated as tumor whereas the network is trained to avoid cluster collapse.

Architectures The cell graph encoder h_ϕ has been implemented with a $L = 5$ layer GIN model. The GIN message passing scheme has been chosen since it is well known to be powerful for graph-level tasks. Additionally, we have not noticed any improvement when increasing the number of layers whereas it limits other parameters such as the hidden dimensionality or the batch size due to memory constraints. Nonetheless, we have seen that the number of layers is important for the clustering model f_φ, which, the deeper it is, the more prone to collapse.

5. Conclusions

We present a self-supervised learning based framework for the analysis of WSIs that learns representations for overlapped small regions of the slide based on the cells that belong to them and, in a second stage, creates a graph of regions to represent the WSI. The model is trained with no annotations and the scarce annotations available for unseen images during training are employed to evaluate the representations at both cell level and region level and it has been shown that they outperform the raw input features. In future work, we will extend the framework by exploring other self-supervised training strategies, including more sophisticated cell level attributes and developing an end-to-end version of the method.

Acknowledgments

This work has been supported by the Spanish Research Agency (AEI) under project PID2020-116907RB-I00 of the call MCIN/ AEI /10.13039/501100011033 and the FI-AGAUR grant funded by Direcció General de Recerca (DGR) of Departament de Recerca i Universitats (REU) of the Generalitat de Catalunya.

References

Valentin Anklin, Pushpak Pati, Guillaume Jaume, Behzad Bozorgtabar, Antonio Foncubierta-Rodríguez, Jean-Philippe Thiran, Mathilde Sibony, Maria Gabrani, and Orcun Goksel. Learning whole-slide segmentation from inexact and incomplete labels using tissue graphs, 2021. URL https://arxiv.org/abs/2103.03129.

Bryan Perozzi Anton Tsitsulin, John Palowitch and Emmanuel Müller. Graph clustering with graph neural networks. In *Proceedings of the 16th International Workshop on Mining and Learning with Graphs (MLG)*, 2020.

Jevgenij Gamper, Navid Alemi Koohbanani, Simon Graham, Mostafa Jahanifar, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, and Nasir Rajpoot. Pannuke dataset extension, insights and baselines. *arXiv preprint arXiv:2003.10778*, 2020.

Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, page 101563, 2019.

William L. Hamilton. Graph representation learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 14(3):1–159, 2021.

Guillaume Jaume, Pushpak Pati, Valentin Anklin, Antonio Foncubierta, and Maria Gabrani. Histocartography: A toolkit for graph analytics in digital pathology, 2021a. URL https://arxiv.org/abs/2107.10073.

Guillaume Jaume, Pushpak Pati, Behzad Bozorgtabar, Antonio Foncubierta, Anna Maria Anniciello, Florinda Feroce, Tilman Rau, Jean-Philippe Thiran, Maria Gabrani, and Orcun Goksel. Quantifying explainers of graph neural networks in computational pathology. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8102–8112, 2021b. doi: 10.1109/CVPR46437.2021.00801.

Joshua Levy, Christian Haudenschild, Clark Barwick, Brock Christensen, and Louis Vaickus. Topological feature extraction and visualization of whole slide images using graph neural networks. *Pac Symp Biocomput*, 26:285–296, 2021.

Wenqi Lu, Simon Graham, Mohsin Bilal, Nasir Rajpoot, and Fayyaz ul Amir Afsar Minhas. Capturing cellular topology in multi-gigapixel pathology images. *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 05 2020.

Yigit Ozen, Selim Aksoy, Kemal Kösemehmetoğlu, Sevgen Önder, and Ayşegül Üner. Self-supervised learning with graph neural networks for region of interest retrieval in histopathology. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6329–6334, 2021. doi: 10.1109/ICPR48806.2021.9412903.

Pushpak Pati, Guillaume Jaume, Antonio Foncubierta-Rodríguez, Florinda Feroce, Anna Maria Anniciello, Giosue Scognamiglio, Nadia Brancati, Maryse Fiche, Estelle Dubruc, Daniel Riccio, Maurizio Di Bonito, Giuseppe De Pietro, Gerardo Botti, Jean-Philippe Thiran, Maria Frucci, Orcun Goksel, and Maria Gabrani. Hierarchical graph representations in digital pathology. *Medical Image Analysis*, 75:102264, 2022. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2021.102264. URL https://www.sciencedirect.com/science/article/pii/S1361841521003091.

Chetan L. Srinidhi, Ozan Ciga, and Anne L. Martel. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67:101813, 2021. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2020.101813. URL https://www.sciencedirect.com/science/article/pii/S1361841520301778.

Fan-Yun Sun, Jordan Hoffman, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. In *International Conference on Learning Representations*, 2019.

Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep Graph Infomax. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=rklz9iAcKQ.

Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. Representation learning on graphs with jumping knowledge networks. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5453–5462. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/xu18c.html.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

Yanning Zhou, Simon Graham, Navid Alemi Koohbanani, Muhammad Shaban, Pheng-Ann Heng, and Nasir M. Rajpoot. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In *2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27-28, 2019*, page 388–398. IEEE, IEEE, 2019. doi: 10.1109/ICCVW.2019.00050. URL https://doi.org/10.1109/ICCVW.2019.00050.