Learning to Represent Whole Slide Images by Selecting Cell Graphs of Patches

Yinan ZHANG*, Beril Besbinar*, Pascal Frossard

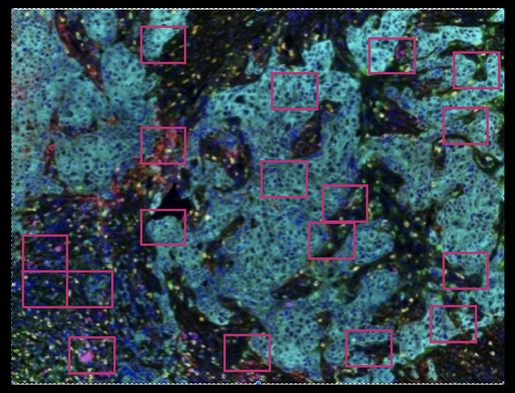
Signal Processing Laboratory (LTS4), École Polytechnique Fédérale de Lausanne (EPFL)

MIDL Lübeck 2021

MOTIVATION

Advances in multiplex biomarker imaging systems have enabled the study of complex spatial biology within the tumor microenvironment. Yet, access to big datasets of such slides with accompanying clinical data is often limited. Moreover, in practice, only some regions of interest (RoIs) are available at high resolution.

Here, we focus on datasets with few images and without labels and aim to learn representations for slides, which are described by cell-graphs[1] of RoIs. We choose a task of patient identification that leads our new model to select RoIs with discriminative properties and infer patient-specific features that can be used later for any task via transfer learning.



PROPOSED METHOD

We construct cell graphs for each ROI, where each cell corresponds to a node and edges are determined by thresholding the spatial distance between cell locations. The initial cell feature is one-hot encoded phenotype. We embed each cell graph by applying successive layers of GraphSAGE[2] and DiffPool[3], which allows us to consider multi-hop relationship of each cell with its neighbors.

For the pre-training phase, we use a permutation invariant aggregation function to obtain a slide (patient) embedding for self identification task with cross-entropy loss. After pre-training, using the learned graph embedding module, we learn to sample graphs jointly with the classification task, using Gumbel-Softmax trick[4] to have an end-to-end differentiable pipeline. Different learning rates are used for updating the graph embedding and the graph sampling modules.

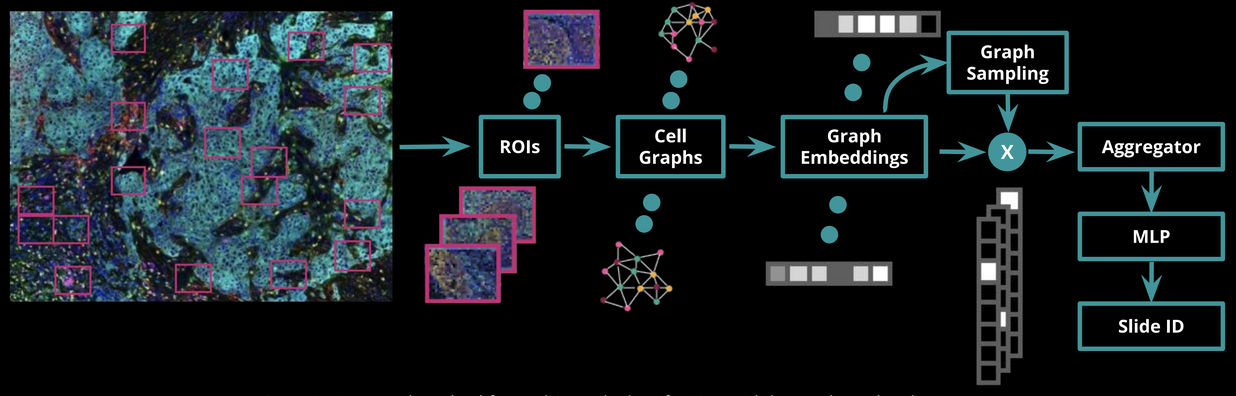


Figure 1: Proposed method for multi-graph classification with learned graph selection.

DATASET

We create a synthetic dataset of 10 patients, each of which is represented by 15 cell graphs. The total number of nodes for each graph is fixed to 1000. We assume six different types of cells, two of which represent *tumour* and *stroma* cells, while the rest resemble lymphocytes.

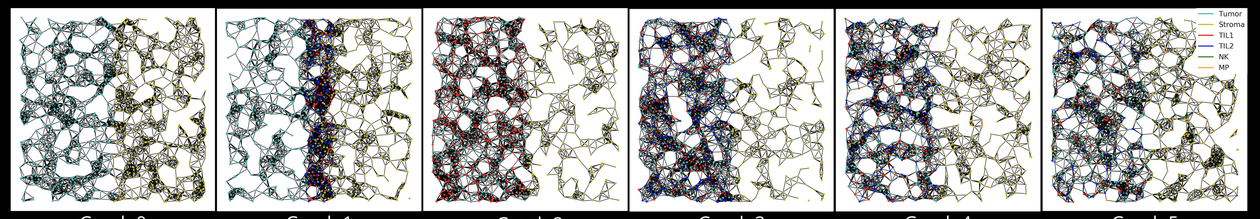
To characterize 10 patients, we create five discriminative cell graph distributions with different combinations of lymphocyte types and counts with different spatial organizations, as described by Table 1.
A sample from each distribution can be seen on Figure 2.

Table 1: Characteristics of Synthetic Dataset

	Ratio of Different Types of Cells						Discriminative Patches									
	Tumour	Stroma	T-Cells	B-Cells	NK-Cells	Macrophages	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
Patch 1	0.40	0.40	0.10	0.10	0.00	0.00	X	X	X	X						
Patch 2	0.30	0.30	0.40	0.00	0.00	0.00	X				X	X	X			
Patch 3	0.30	0.30	0.20	0.20	0.00	0.00		X			X			X	X	
Patch 4	0.30	0.30	0.15	0.15	0.00	0.00			X			X		X		X
Patch 5	0.30	0.30	0.10	0.10	0.10	0.10				X			X		X	X

* Contributed Equally, The first author performed the work while at LTS4, EPFL

EXPERIMENTAL SETUP



Graph 0 Graph 1 Graph 2 Graph 3 Graph 4 Graph 5

Figure 2: Samples from the different graph distributions

We split the set of patches for each patient into three equal parts for training/validation/testing by keeping the number of different discriminative patches similar for each subset.

During the pre-training, we experimented with different aggregation functions, and different architectures. For the graph sampling module, we tried using the approximation of categorical sampling variable (gumbel-softmax, soft) as well as the real one-hot version (gumbel-softmax, hard).

We used the validation set to determine these hyperparameters and design choices.

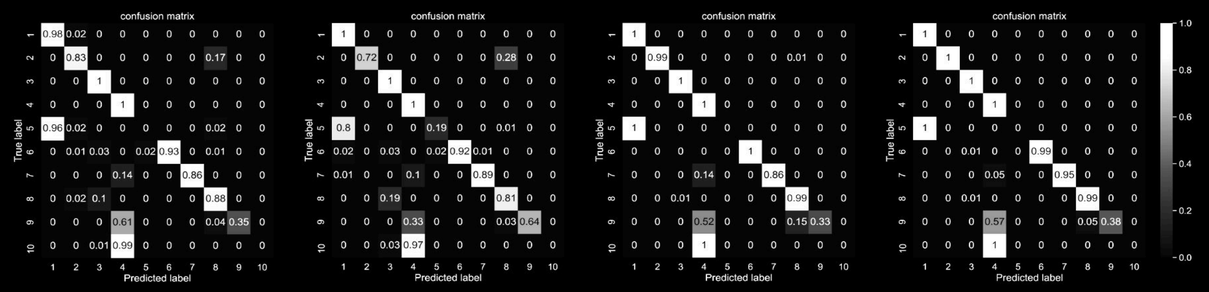
RESULTS

For pre-training, the identification (classification) test accuracy reaches to 100%.

Table 2 reports both the classification accuracy and patch selection performance when we sample different number of patches for each patient. Figure 3 further depicts the classification performance when we sample different number of graphs at test time. The reported accuracy values are the average of 100 Monte Carlo simulations.

Table 2: The quantitative results for different experimental setups

agg	gumbel softmax	# of patches selected	acc (both)	acc (one)	acc (none)	classi acc
sum	soft	2	0.538	0.460	0.002	0.800
sum	hard	2	0.380	0.606	0.014	0.700
mean	soft	1	0	0.984	0.016	0.700
mean	soft	2	0.479	0.519	0.002	0.800
mean	soft	3	0.584	0.416	0	0.700
mean	soft	4	0.670	0.330	0	0.700
mean	hard	2	0.447	0.554	0.009	0.800
max	soft	2	0.282	0.678	0.040	0.700
max	hard	2	0.258	0.736	0.006	0.700



K=1 K=2 K=3 K=4

Figure 3: The average confusion matrices of Monte Carlo simulations with different number of selected patches (K) for each patient

REFERENCES

[1] Bulent Yener. Cell-graphs: image-driven modeling of structure-function relationship.Communications of the ACM, 60(1):74–84, 2016

[2] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In Advances in neural information processing systems, pages 1024–1034, 2017.

[3] Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. In Advances in neural information processing systems, pages 4800–4810, 2018.

[4] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144, 2016.