Master's thesis

Deep learning methods for histological image processing

Learning, and showing, which regions of a gigapixel slide belong together

Author
Ing. Patrik Kozlík

Supervisor
prof. Ing. Vanda Benešová, PhD.

Faculty
FIIT STU Bratislava, 2026

Motivation

A gigapixel object, cut into tiles and averaged

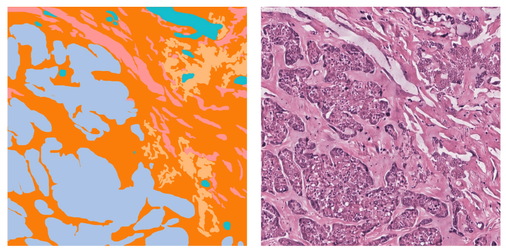
Reading histological slides is what identifies the biomarkers that guide treatment, but the images are enormous and slow to analyse. AI in digital pathology can significantly support doctors and help save time. Our research focuses on two main problems that limit its use. Gigapixel slides are processed patch by patch and pooled into an unordered bag, so the long-range relations a pathologist reads holistically are lost. On the other hand, those predictions are unexplainable — a model that cannot show which regions drove its output cannot be deployed. This work builds those local–global relations into the architecture, exposes them as an inspectable graph, and validates them on synthetic slides with known structure.

- **Relations carry the diagnosis.** One slide analysed is 161 352 × 78 618 px. Tile pipelines embed thousands of tiles independently, then pool them as an unordered bag, discarding exactly relations between them.
- **No explanation, no clinic.** A model that cannot show which regions drove its output cannot be deployed.
- **Attentions weights** inside of mechanisms **are not trustworthy** as an single source of explainability.

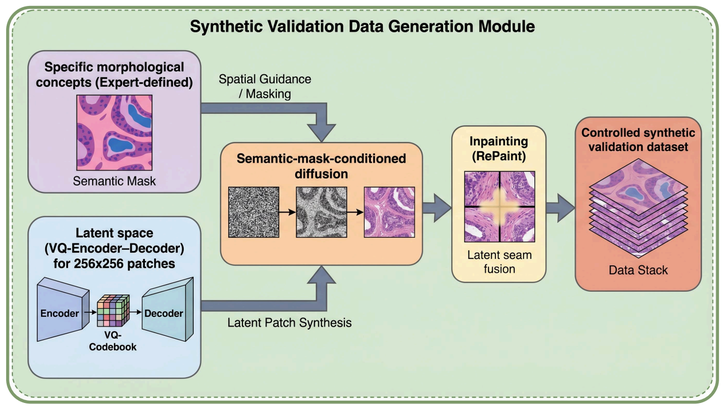
Validation data

Synthetic slides with ground truth by construction

- **Inpainting as a stitching operator.** Mask-conditioned latent diffusion over a VQ-VAE latent space generates the tissue.
- **Qualitative and quantitative evaluation**. Seams stay invisible to Segment Anything and to superpixels, and 86.7 % of 15 lay readers found the target structures.



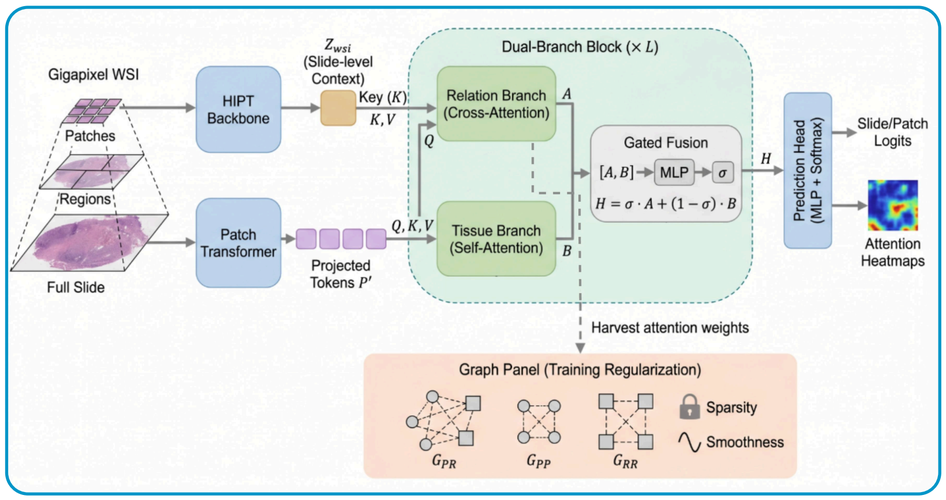
Example of semantic mask and synthetically generated region of 4096x4096px



Three-group BRACS task (benign / atypical / malignant), 500 slides, 467 patients, patient-grouped folds

Model

Graph RA-HIPT: Architecture revealing relations



- **Two branches, fused by a learned gate** - The *relation branch* lets a region's patch tokens attend over the whole-slide context. *Tissue branch* lets them attend over themselves. Every head yields its own relations, so explainability is built in rather than added afterwards.
- **No diagnostic labels in training** - The objective is purely structural (attention sparsity, smoothness and neighbourhood diversity). Labels are used only afterwards, to score the representation with a linear probe.
- **Graph RA-HIPT** - Turns attention into patch to region, patch to patch and learned region to region graphs regularised for sparsity, diversity and anti-collapse (VICReg-style), keeping long-range relations alive. Embeddings are disentangled into three subspace views.

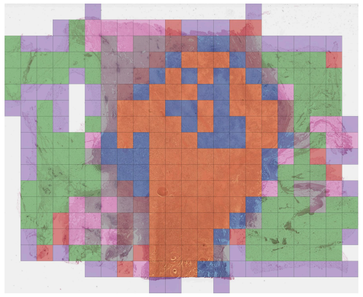
Results

The representation carries the signal, the graph shows where

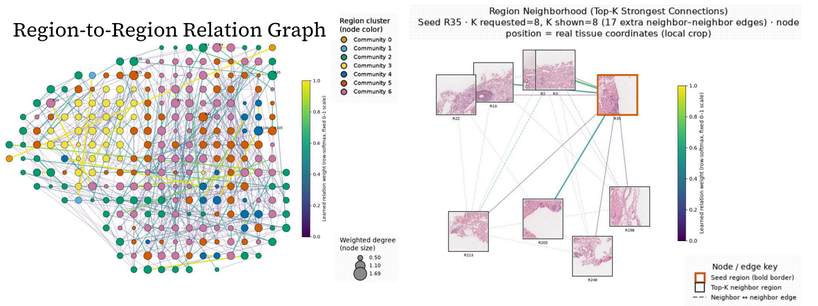
Stable neighbourhoods - Top-5 neighbourhoods agree at Jaccard 0.87–0.94 across region subsamples, and links join similar tissue across the slide rather than merely adjacent tiles.

Aggregator, same folds	frozen HIPT bAcc / AUC	Graph RA-HIPT bAcc / AUC
Mean pooling	0.347 / 0.564	**0.487 / 0.699**
Max pooling	0.357 / 0.600	**0.484 / 0.706**
ABMIL	0.344 / 0.527	**0.479 / 0.699**
CLAM-SB	0.332 / 0.504	**0.479 / 0.684**
CLAM-MB	0.324 / 0.537	**0.481 / 0.692**
TransMIL	0.486 / 0.653	**0.493 / 0.683**

Three-group BRACS task (benign / atypical / malignant), 500 slides, 467 patients, patient-grouped 5-k folds



BRACS_1232 - Invasive carcinoma. **Learned relation communities** separate the invasive mass (orange), its border (blue) and the peripheral fatty and stromal tissue (green) without ever seeing a label. Edges are drawn at true slide coordinates, so a pathologist can inspect them in HistoView or QuPath.



Visualization of region-to-region graph with the top-k strongest connections on the different WSI

- **Read by a pathologist. A pathologist confirmed the structures** the first relation graphs found, and their relations. A reader study with the **Biomedical Research Center SAS** follows.

Accepted paper: P. Kozlík, V. Benešová, Context-Aware Synthesis of Histological Whole-Slide Images Using Latent Diffusion Inpainting, ICMV2026, Budapest