

# Learning a low dimensional manifold of real cancer tissue with PathologyGAN

Adalberto Claudio Quiros

Roderick Murray-Smith

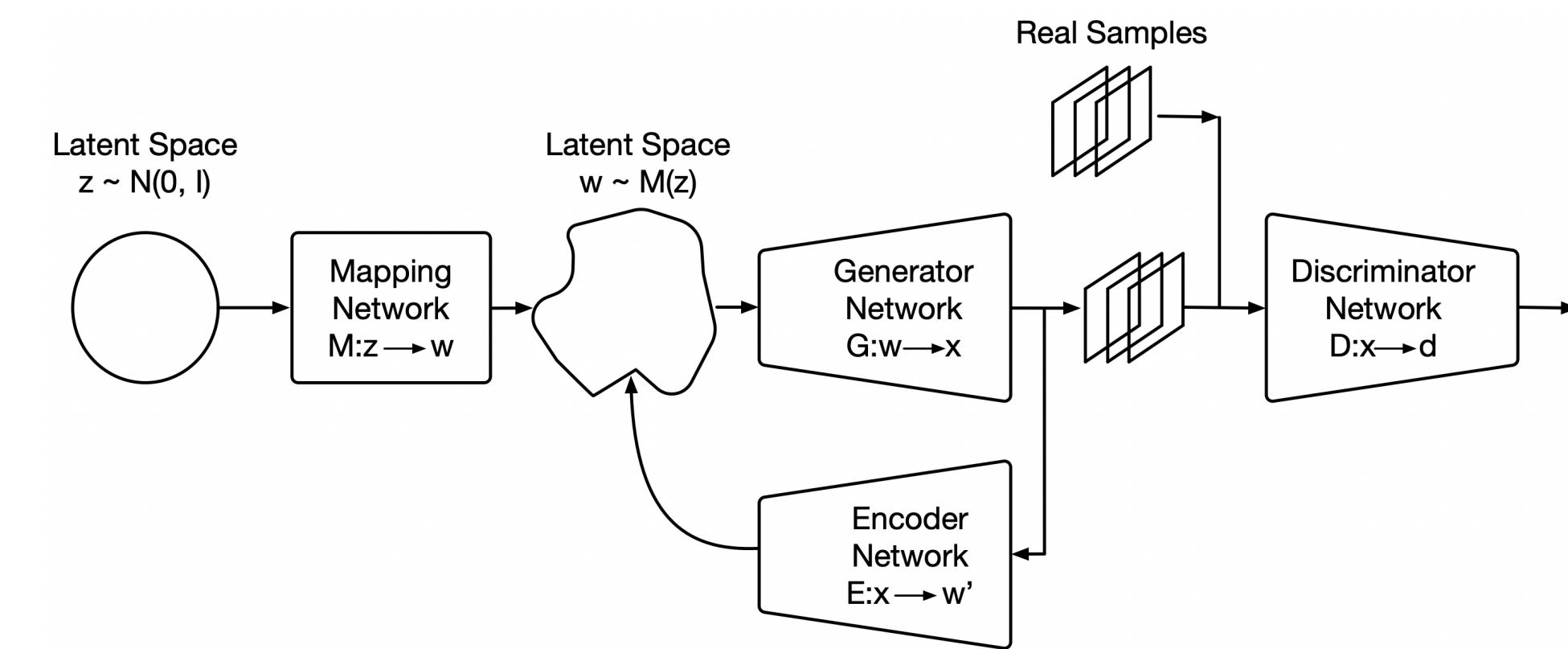
Ke Yuan

School of Computing Science, University of Glasgow

## Abstract

- Deep generative models with representation learning properties provide an alternative path to further understand cancer tissue phenotypes, capturing tissue morphologies.
- We present a deep generative model that learns to simulate high-fidelity cancer tissue images while mapping the real images onto an interpretable low dimensional latent space.
- Here we provide examples of how the latent space holds morphological characteristics of cancer tissue (e.g. tissue type or cancer, lymphocytes, and stromal cells).

## PathologyGAN Model



## GAN Loss: Relativistic Average Discriminator

$$\begin{aligned} L_{Dis} &= -\mathbb{E}_{x_r \sim \mathbb{P}} [\log (\tilde{D}(x_r))] - \mathbb{E}_{x_f \sim \mathbb{Q}} [\log (1 - \tilde{D}(x_f))], \\ L_{Gen} &= -\mathbb{E}_{x_f \sim \mathbb{Q}} [\log (\tilde{D}(x_f))] - \mathbb{E}_{x_r \sim \mathbb{P}} [\log (1 - \tilde{D}(x_r))], \\ \tilde{D}(x_r) &= \text{sigmoid}(C(x_r) - \mathbb{E}_{x_r \sim \mathbb{Q}} C(x_r)), \\ \tilde{D}(x_f) &= \text{sigmoid}(C(x_f) - \mathbb{E}_{x_r \sim \mathbb{P}} C(x_r)). \end{aligned}$$

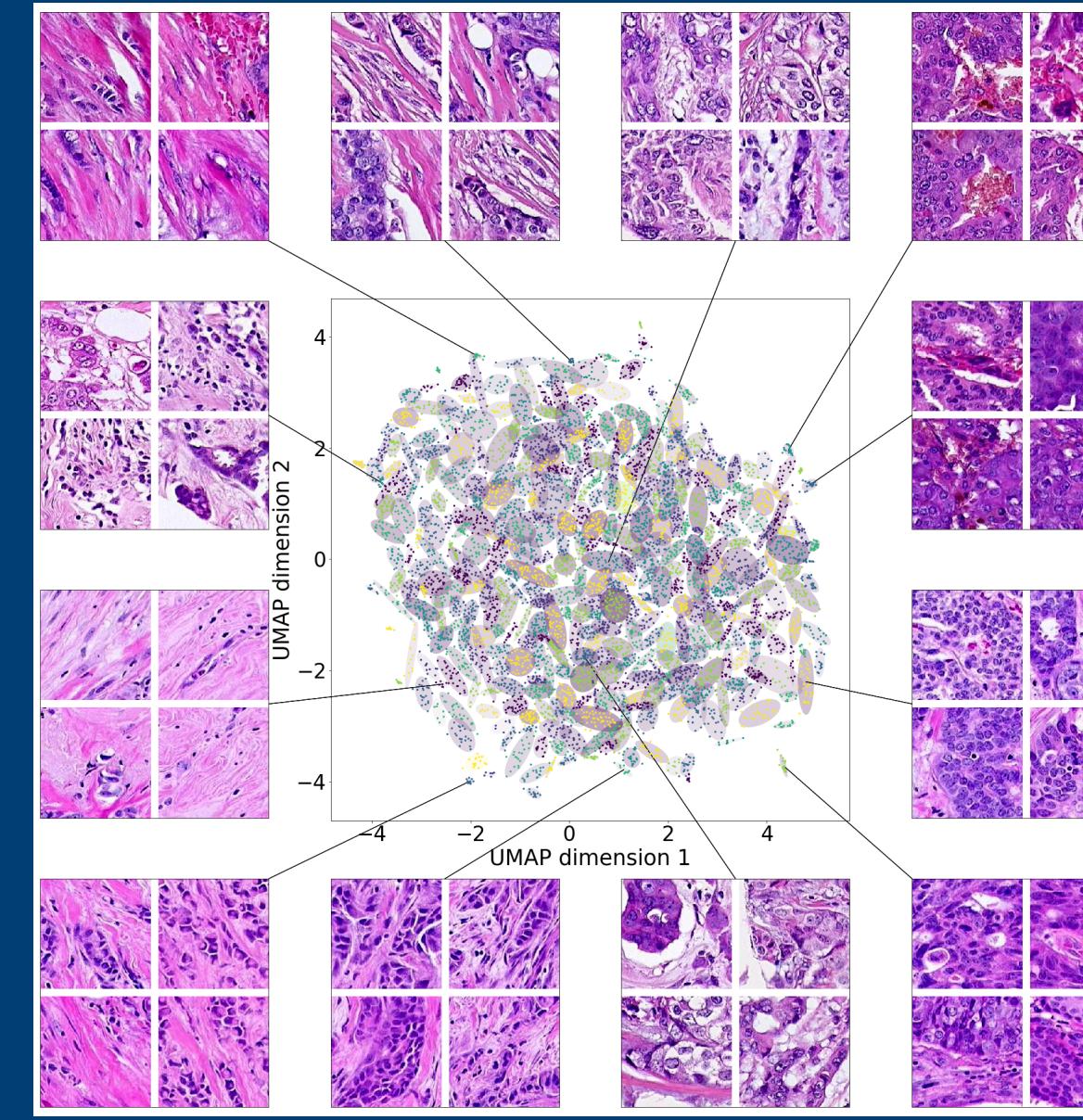
## Encoder Loss: MSE Reconstruction of Latent Vectors

$$L_{Enc} = \mathbb{E}_{z \sim P_z} \left[ \frac{1}{n} \sum_{i=1}^n (w_i - w'_i)^2 \right]$$

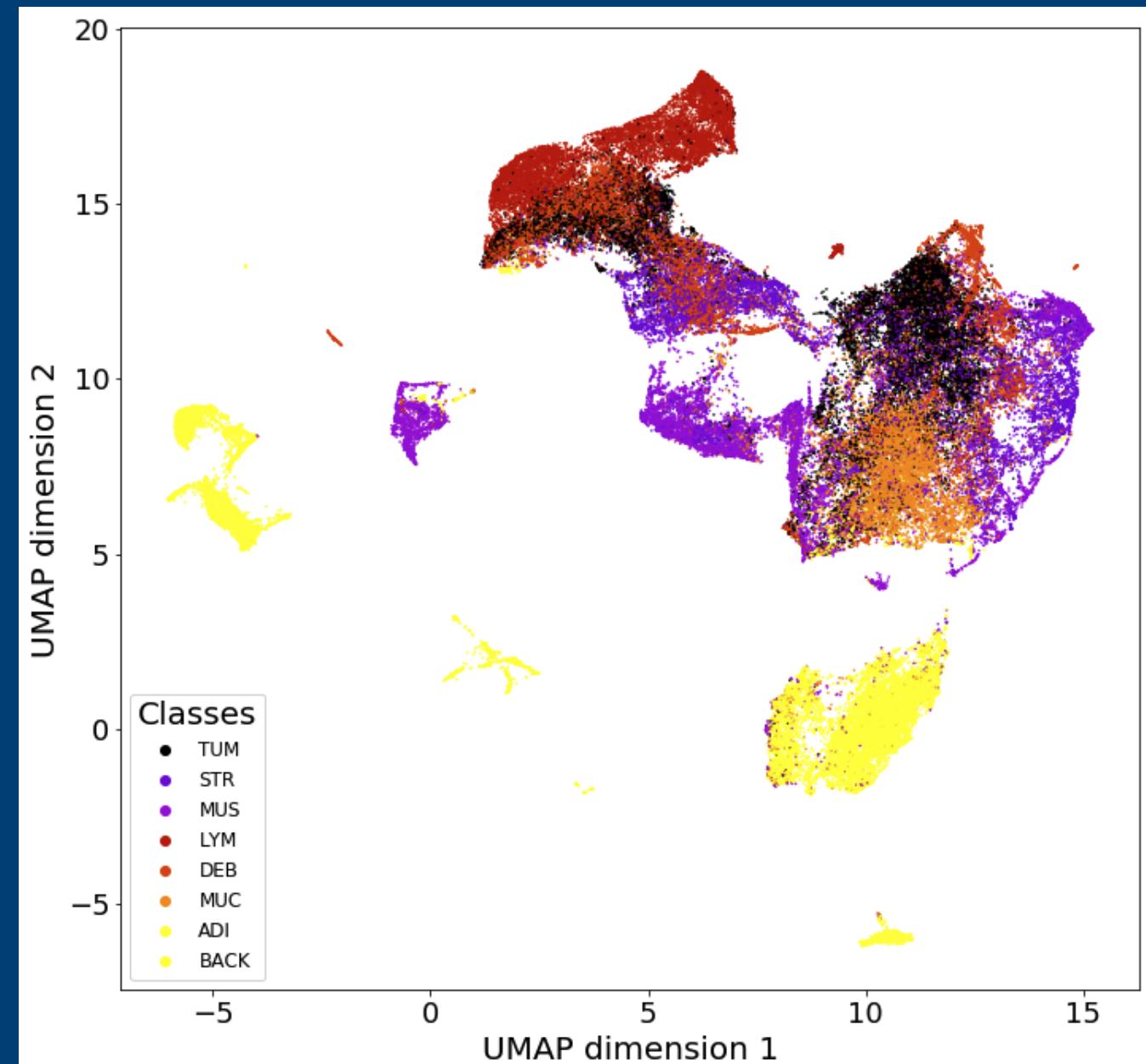
where  $n = \dim(w)$ ,  $w' = E(G(w))$ ,  $w \sim M(z)$ .

# Visualization of the Latent Space

- H&E Breast cancer dataset from Netherlands Cancer Institute (NKI) and the Vancouver General Hospital (VGH).
- Projection of real tissue images onto the GAN's latent space.
- Distinct regions of the latent space contain tissue samples with common properties, such as tissue type, stroma, lymphocyte, or cancer cells.



## Tissue Type Classification

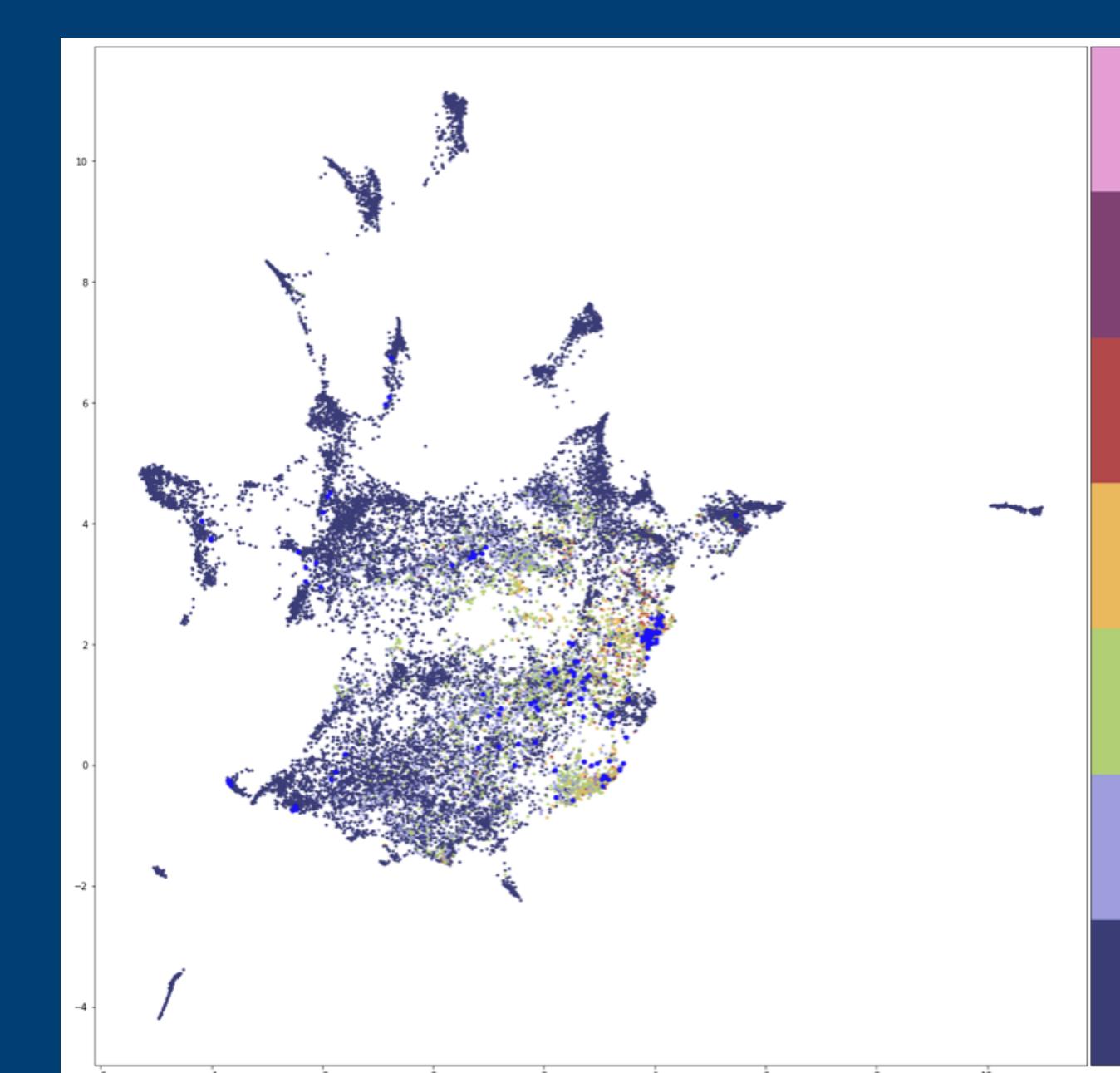
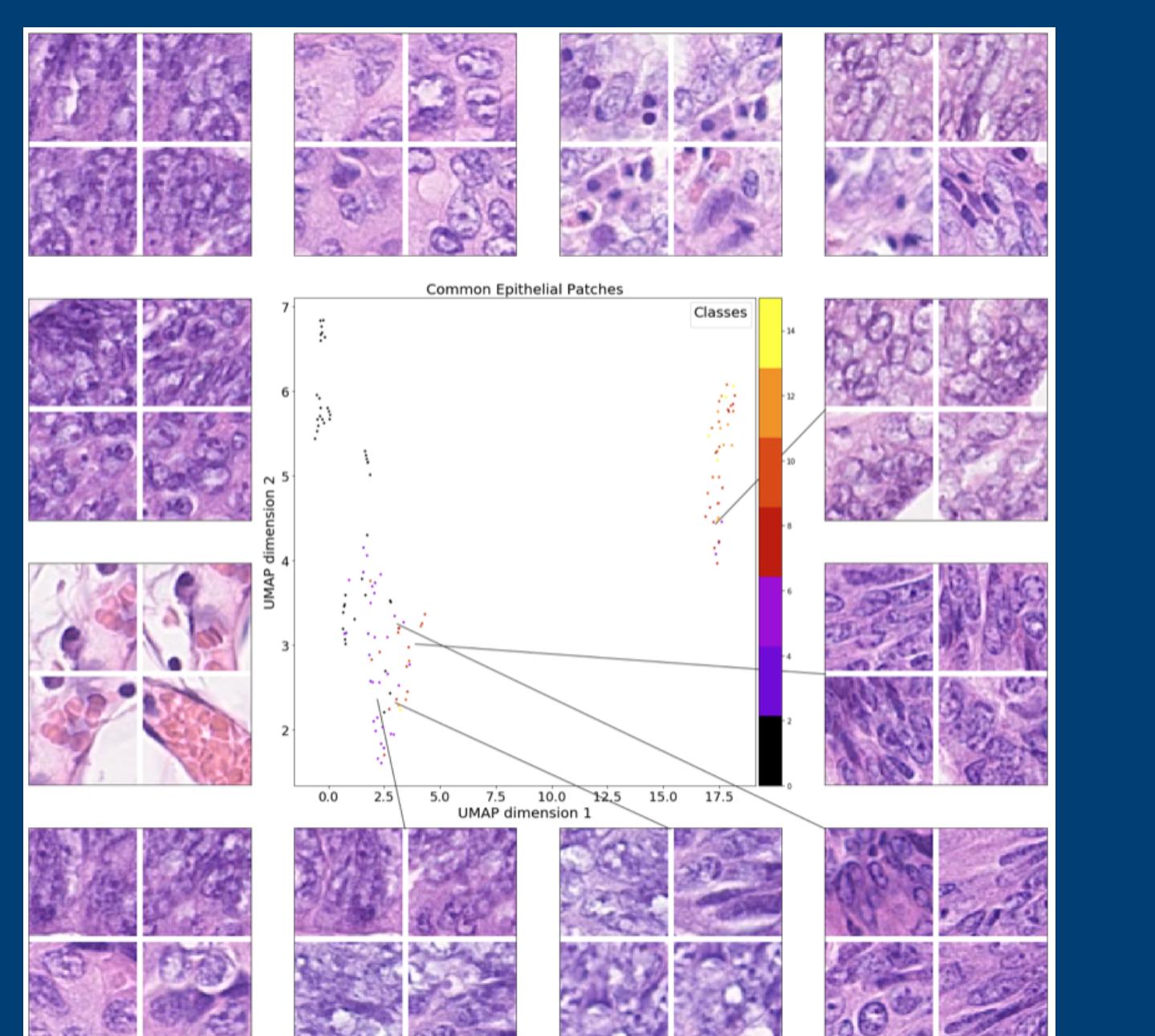


- H&E Colorectal cancer dataset from National Center for Tumor diseases (NCT, Germany) [2].
- Each tissue image has an associated tissue label: adipose, background, debris, lymphocytes, mucus, smooth muscle, normal colon mucosa, cancer-associated stroma, and colorectal adenocarcinoma epithelium (tumor).
- Logistic regression trained over latent vectors:

Test accuracy: 87%							
Tumor	Stroma	Muscle	Lymph.	Debris	Mucosa	Adipose	Background
91.8%	38.2%	78.4%	92%	82.6%	90.4%	90.1%	99.4%

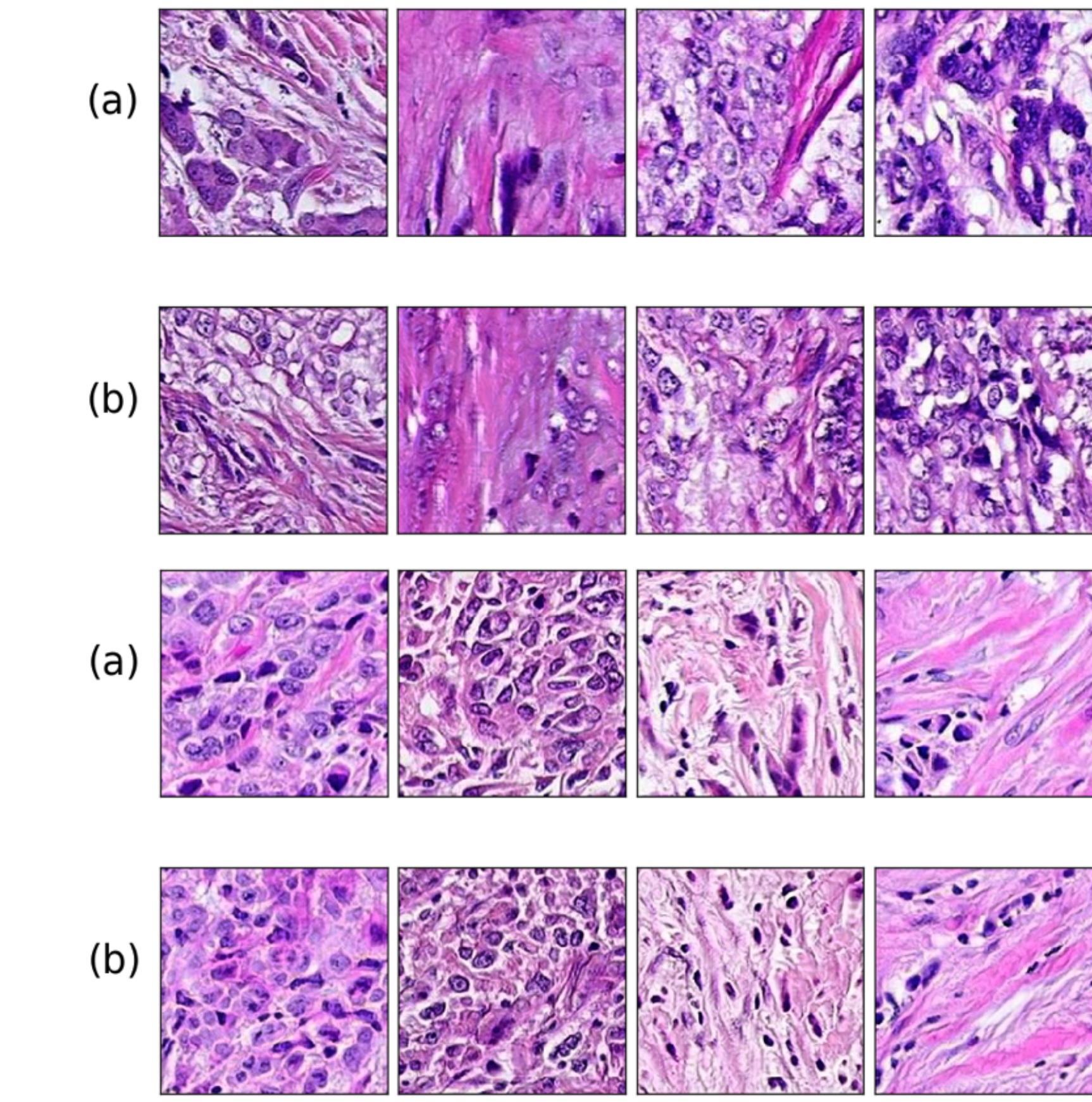
## Multiple Instance Learning over Latent Representations

- We use the attention-based deep MIL [1] over our latent representations. This framework highlights which samples of the bag of instances are relevant for the label or outcome, assigning specific attention weights to each of the samples.
- H&E Colorectal dataset [4] containing 100 tissue images with annotations of epithelial cells. We assign the label of one or zero based on the presence of epithelial cells and divide each image into  $56 \times 56$  patches.
- We run a 10-fold cross validation achieving an accuracy of 89%.
- Top 10% of most weighted tissue patches are concentrated in areas with high density of epithelial cells. UMAP reduction of all latent presentations, assigning color based on the number of epithelial cells in the patch, top weighted patches are highlighted in blue.



## Real Tissue Reconstructions

Real tissue images (a) and their reconstructions (b), images are paired by columns. Reconstructions follow the real tissue attributes.



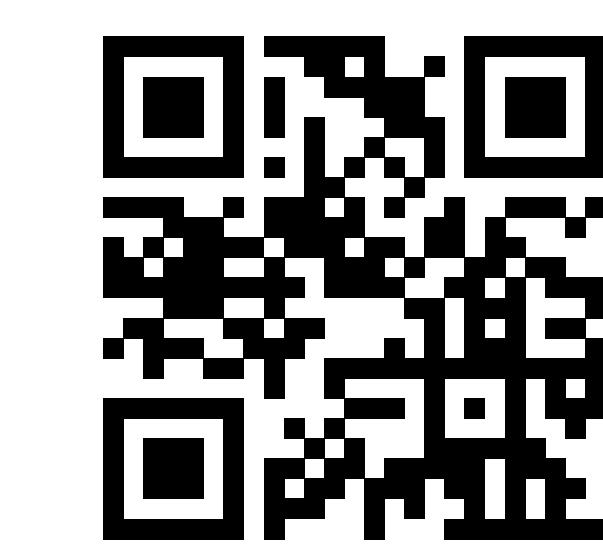
## Conclusions

PathologyGAN[3] captures distinct phenotype characteristics of cancer tissue. We tested the general applicability of our representations in three different settings:

- Latent space visualization.
- Tissue type classifier over latent representations.
- Multiple Instance Learning.

## References

- [1] Maximilian Ilse, et al. Attention-based deep multiple instance learning, ICML 2018.
- [2] Jakob Nikolas Kather, et al. 100,000 histological images of human colorectal cancer and healthy tissue, April 2018.
- [3] Adalberto Claudio Quiros, et al. Pathologygan: Learning deep representations of cancer tissue. MIDL 2020.
- [4] K. Sirinukunwattana, et al. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Transactions on Medical Imaging, 2016.



Take a picture  
to  
download the  
full paper



University  
of Glasgow