Identifying Explainable Latent Features in Digital Whole Slide Pathology Images

Katherine Xu[1], Mirage Modi[1], Qin Ma[1,2], Jordan Krull[1,2]

[1] Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA. [2]The Pelotonia Institute for Immuno-oncology, The Ohio State University Comprehensive Cancer Center, Columbus, OH 43210, USA.

Motivation

- Hematoxylin and Eosin stained tissue slides are an abundant source of information, which is standard practice in pathology and readily accessible at most institutions as digital slide images.
- The human body is composed of numerous distinct organ systems, made up of combinations of various fundamental tissues, which are organized multicellular components.
- A large comparative study of these “bulding blocks” is currently infeasible for humans, due to the sheer size of data.
- Deep learning is a capable solution to extracting information from images.

Data, WSI QC, and processing

Data

- 26,000 WSIs at 40x objective from GTEx
- 40 tissues represented from 29 organs
- Select 128 x 128px (0.64µm x 0.64µm) random patches from each image.
- ~800,000 patches equally representing all 40 tissues (<1% of any single slide)

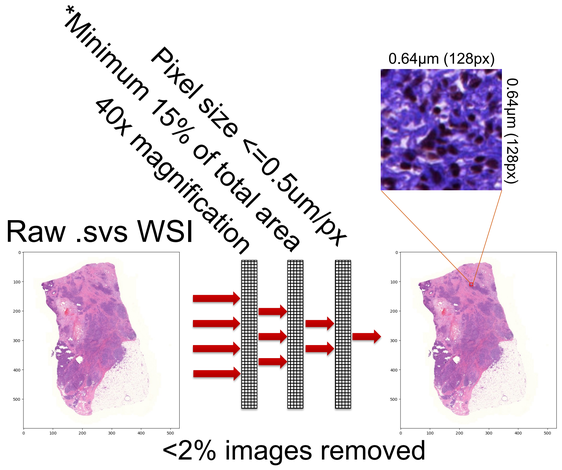

Figure 1 – WSI filtration and training image generation.
Raw WSIs are filtered for tissue content, magnification, and appropriate microns per pixel. Randomly selected tissue patches were then extracted from the resulting list of images (>98% original list) at 0.64um square and resized to 128px square.

Small, high-resolution images contain little information on tissue of origin

Poor performance on tissue classification from training images

- Training ~800k GTEx image patches to predict tissue of origin (40 classes)
- Trained using 80/10/10 split on DenseNet-121 and VGG-16 for 20 initial epochs unless not converged
- DenseNet converged quickly and overgeneralized on some classes
- VGG-16 overgeneralized to the training set

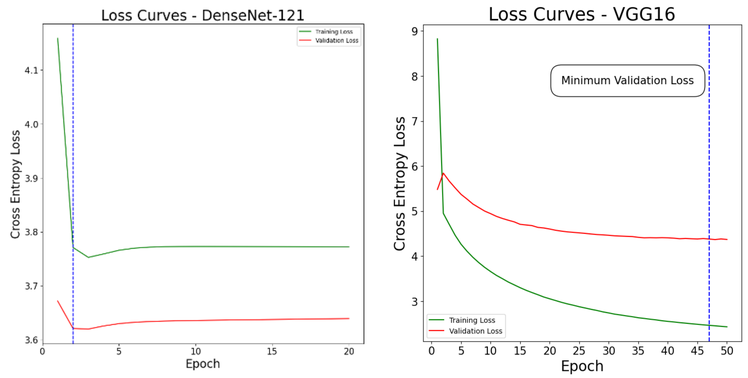

Figure 2 – WSI patch tissue classification training loss curves.
(Left) Loss curve for the best performing DenseNet-121 over 20 epochs, where each epoch evaluates from the entire 80% training set. (Right) Loss curve for the best performing VGG-16 over 50 epochs. Green = Training Loss & Red = Valdiation Loss.

DenseNet displayed poor prediction accuracy and preference for specific tissues

- Trained DenseNet was applied to the test set for class prediction
- Average validation set % correct across all tissues was ~8%
- Most tissues resulted in predictions among 6 classes of tissue

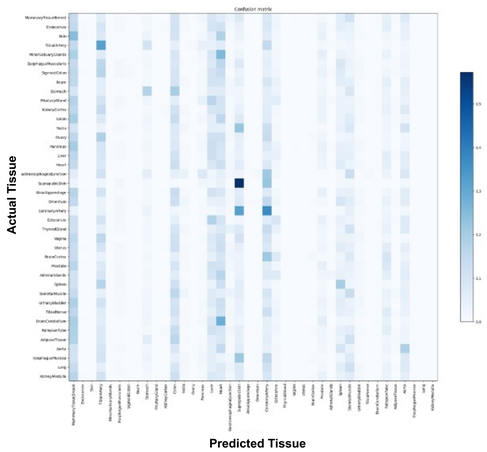

Figure 3 – Actual and predicted tissue confusion matrix from DenseNet-121.
Confusion matrix representing percentage of each training class predicted for each other training class. The cell values represent percent of row (actual class).

Extracting features using an autoencoder

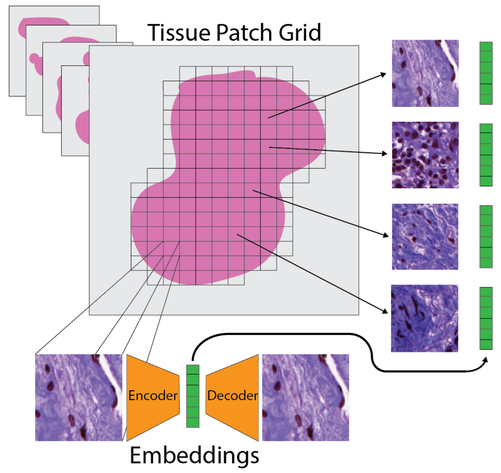

Figure 4 – Tissue patch and embedding feature extraction scheme from WSIs.
Cartoon representation of tissue patch extraction and generation of patch embeddings from a simple autoencoder.

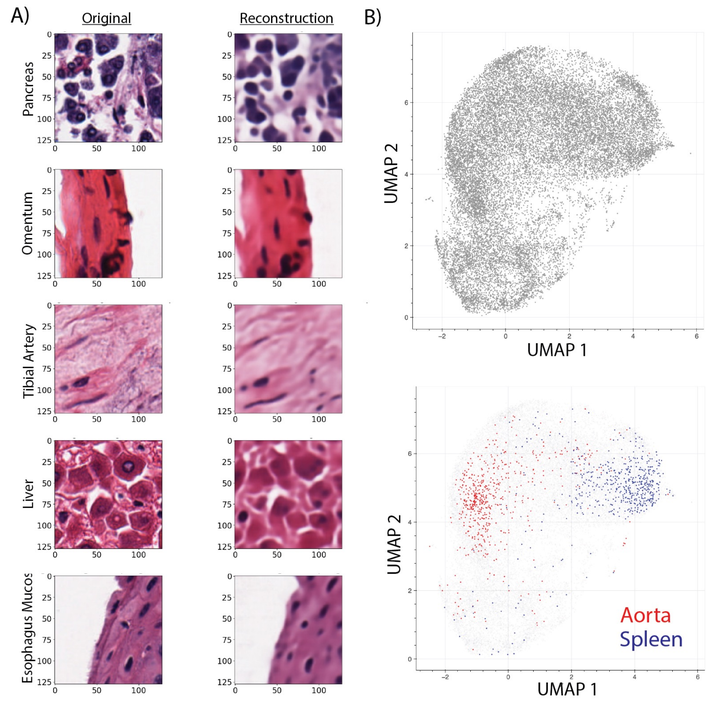

Figure 5 – Embedding feature representations from a convolutional autoencoder.
(A) Image reconstruction performance from the trained autoencoder. 5 representative tissues are chosen. (B) UMAP dimension reduction representations of training patch embeddings. Representative tissues demonstrate the embeddings are capable of discerning tissues of very different origins.