Evaluation of Domain Specific Encoding Strategy as Compared to Transfer Learning Approaches in Digital Pathology

Zihang Fang[1], Alex Baras[1]

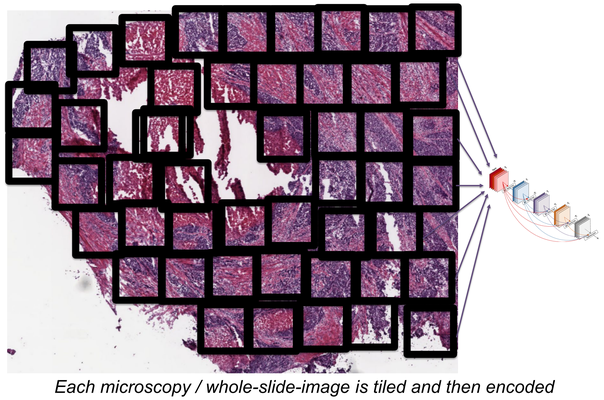
[1]Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD

Background

End to end featurization of high-resolution whole slide images (WSIs) in pathology can be challenging. Common workarounds involve tiling WSIs into smaller images of reasonable size, extracting features from each tile, and then aggregating over the whole WSI. Due to complexity of end-to-end WSI level machine learning tasks and the small amounts of tile-level annotations from WSIs (relative to other image machine learning datasets), the more common approach in digital pathology has been the use of models pre-trained on out-of-domain data (transfer learning). Domain specific feature extractors can still be trained through unsupervised or weakly-supervised settings, although their effectiveness remains unanswered.

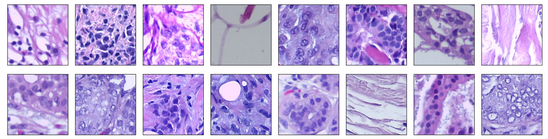
Methods

We use the ICIAR 2018 BACH dataset for training and evaluation. This dataset contains 400 microscopy images of three classes: normal (200), in-situ carcinoma (100), and invasive carcinoma (100). We develop and train a densely connected convolutional variational autoencoder (VAE) with a 64-dimensional bottleneck and later use its encoder as the domain-specific feature extractor. We also use a DenseNet201 pre-trained from ImageNet for a baseline comparison, which by design has a bottleneck feature size of 1920. We consider two algorithms that return WSI-level features from a set of tile-level features of that WSI: 1) global percentile pooling (GPP), which performs pooling across the entire set of tiles at distinct percentiles (0th, 25th, 50th, 75th, 100th) and concatenates these into a feature vector for the WSI; 2) attention based multiple instance learning (aMIL).

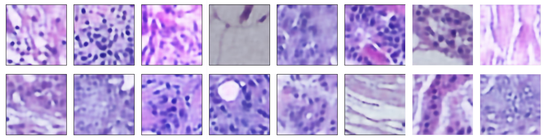


Each microscopy / whole-slide-image is tiled and then encoded

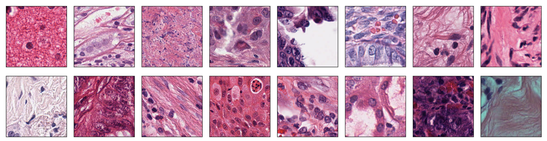
Results



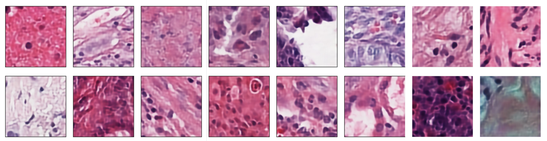
Tiles randomly sampled from ICIAR dataset



Reconstructed tiles from VAE trained on ICIAR dataset

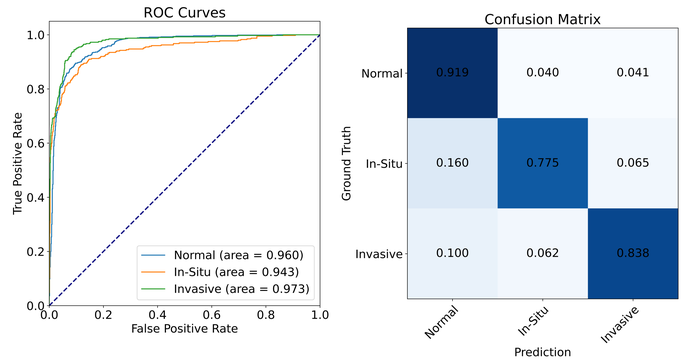


Tiles randomly sampled from TCGA dataset

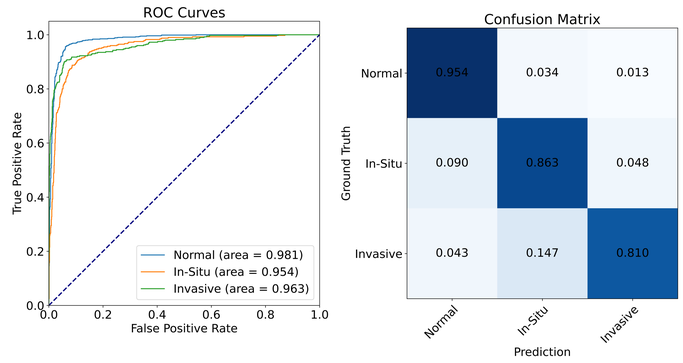


Reconstructed tiles from VAE trained on TCGA dataset

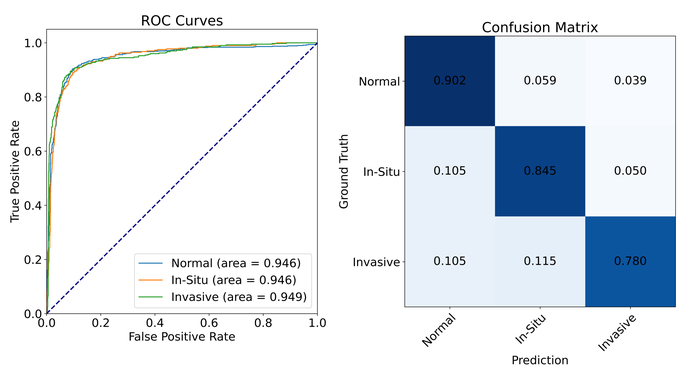
The densely connected convolutional VAE (used as the domain specific feature extractor) trained with a β value of 1e-3 was able to converge at an MSE loss of 3e-3. The tile reconstructions were reviewed by a pathologist and considered nearly indistinguishable from input. This encoder of the VAE had slightly over 400 thousand parameters, and the DenseNet201 architecture had roughly 18 million. DenseNet201 pretrained on ImageNet as the tile feature extractor achieved an overall classification accuracy of 86.3% using GPP and an accuracy of 89.5% using aMIL. These results outperformed the domain specific pre-trained VAE encoder (BACH data) both with (85.7% using GPP and 82.0% using aMIL) and without (70.4% using GPP and 66.8% using aMIL) allowing for fine-tuning during supervised learning.



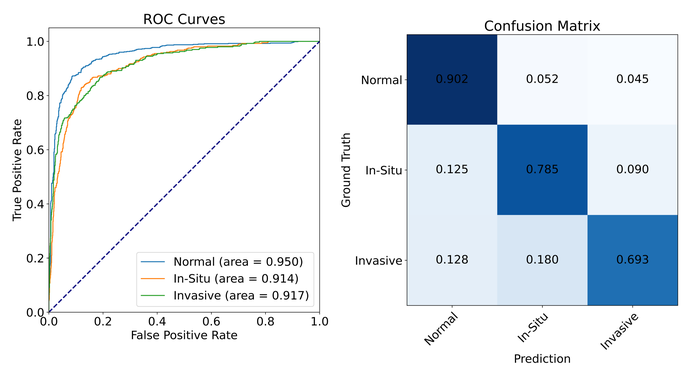
Tile features were extracted using pre-trained DenseNet201, WSI-level features were acquired using GPP



Tile features were extracted using pre-trained DenseNet201, WSI-level features were acquired using attention based MIL



Tile features were extracted using fine-tuned VAE, WSI-level features were acquired using GPP



Tile features were extracted using fine-tuned VAE, WSI-level features were acquired using attention based MIL

Conclusion

A domain specific VAE encoder trained by BACH dataset is outperformed by pre-trained DenseNet201 as a feature extractor. Fine tuning the domain specific encoder helps closing the gap, but only if global percentile pooling is used. This contrasts with the use of pretrained DenseNet201 wherein aMIL increased performance. We may expect this inconsistency to be overcome with the introduction of a larger and more comprehensive dataset, such as the TCGA which we are currently exploring and will add to this work. It should be noted that the domain specific encoder of the VAE had orders of magnitude smaller number of parameters and size of latent representation while performing nearly as well as a pre-trained DenseNet201. Future work will include enlarging the parameter size and latent representation of the domain specific VAE encoder.

Contact

zfang15@jh.edu
https://www.linkedin.com/in/zihang-fang-0729/
baras@jhmi.edu
https://www.linkedin.com/in/alexander-baras-md-phd-03986586/