Certainty Pooling for Multiple Instance Learning
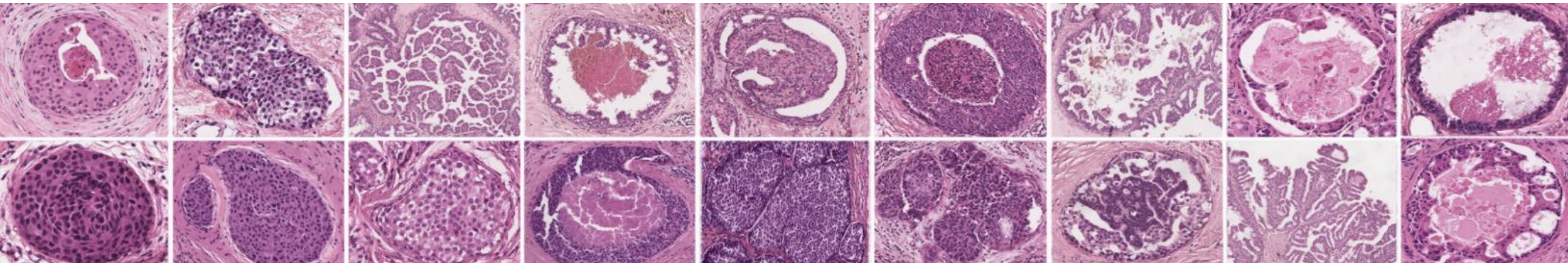
Jacob Gildenblat[1], Ido Ben-Shaul[1], Zvi Lapp[1], Eldad Klaiman[2]

[1] SagivTech Ltd. / DeePathology.ai

[2] Early Biomarker Discovery Oncology, Pharma Research and Early Development, Roche Innovation Center Munich

The Team

- Jacob Gildenblat, SagivTech

- Ido Ben-Shaul, SagivTech

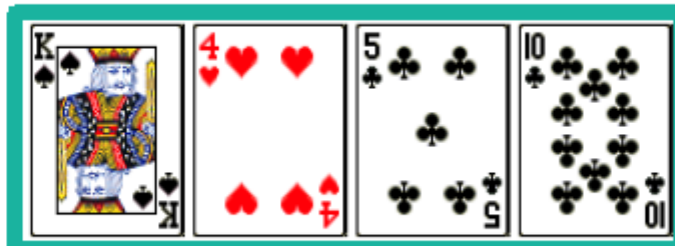- Eldad Klaiman, Early Biomarker Discovery Oncology, pRED, RICM

- Amal Lahiani, Jacob Gildenblat, Irina Klaman, Nassir Navab, Eldad Klaiman, "Generalizing multistain immunohistochemistry tissue segmentation using one-shot color deconvolution deep neural networks." IET Image Processing (2019). https://arxiv.org/abs/1805.06958
- Lahiani, Amal, Jacob Gildenblat, Irina Klaman, Shadi Albarqouni, Nassir Navab, and Eldad Klaiman. "Virtualization of tissue staining in digital pathology using an unsupervised deep learning approach." ECDP (2019). https://arxiv.org/abs/1810.06415
- Gildenblat, Jacob, and Eldad Klaiman. "Self-Supervised Similarity Learning for Digital Pathology." MICCAI-*COMAPY 2019 arXiv:1905.08139* (2019).
- Eldad Klaiman, Jacob Gildenblat, Ido Ben-Shaul, Astrid Heller, Konstanty Korski, Astrid Kiermaier, Fabien Gaire, Prediction of biomarker status, diagnosis and outcome from histology slides using deep-learning based hypothesis free 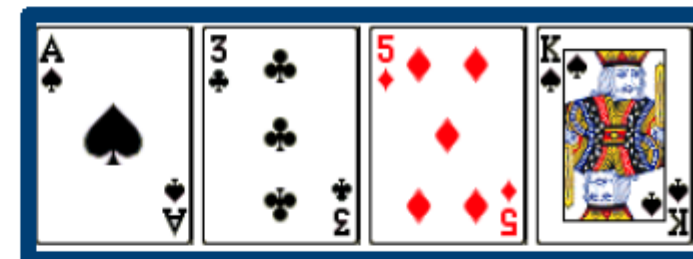feature extraction –ASCO 2019, June 1st, 08:00, Hall A https://www.linkedin.com/feed/update/urn:li:activity:6544867419603685376

Multiple Instance Learning

Has the card

Has the card

Does this have the card?

Does not have the card

Does not have the card

What is the card?
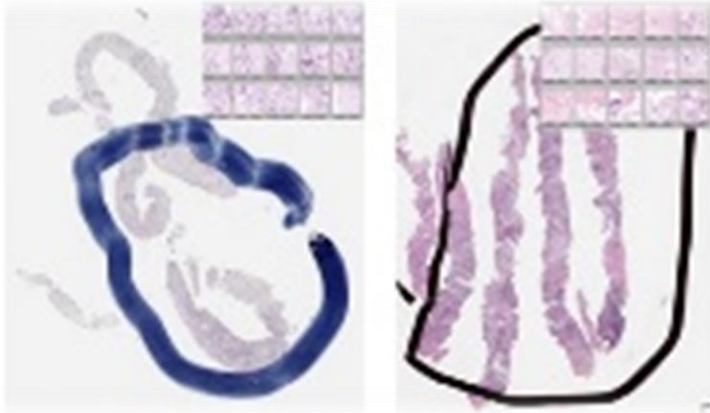
Instance

Bag

Klaiman, Eldad, et al. "Prediction of biomarker status, diagnosis and outcome from histology slides using deep learning-based hypothesis free feature extraction." (2019): 3140-3140.
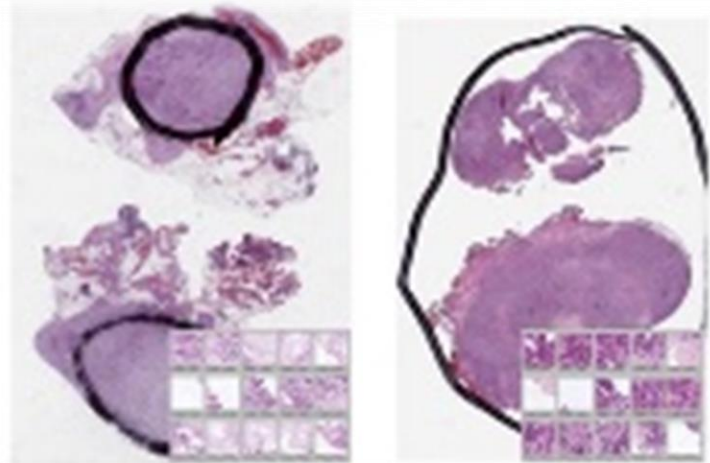
Multiple Instance Learning in Pathology

Patient responds to treatment

No response to treatment

Can this patient benefit from treatment A ?

Klaiman, Eldad, et al. "Prediction of biomarker status, diagnosis and outcome from histology slides using deep learning-based hypothesis free feature extraction." (2019): 3140-3140.
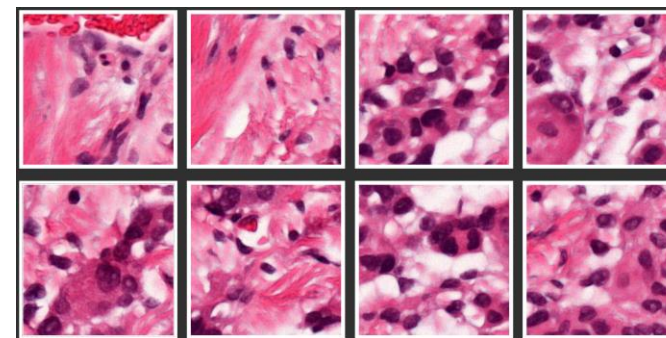
Motivation

- Multiple Instance Learning is more difficult when the predictive ratio is very low such as in Pathology
- We wish to interpret the top tiles that contributed to the prediction
- Need to deal with uncertainty - what happens when the model sees tiles it never saw before, like new artifacts?
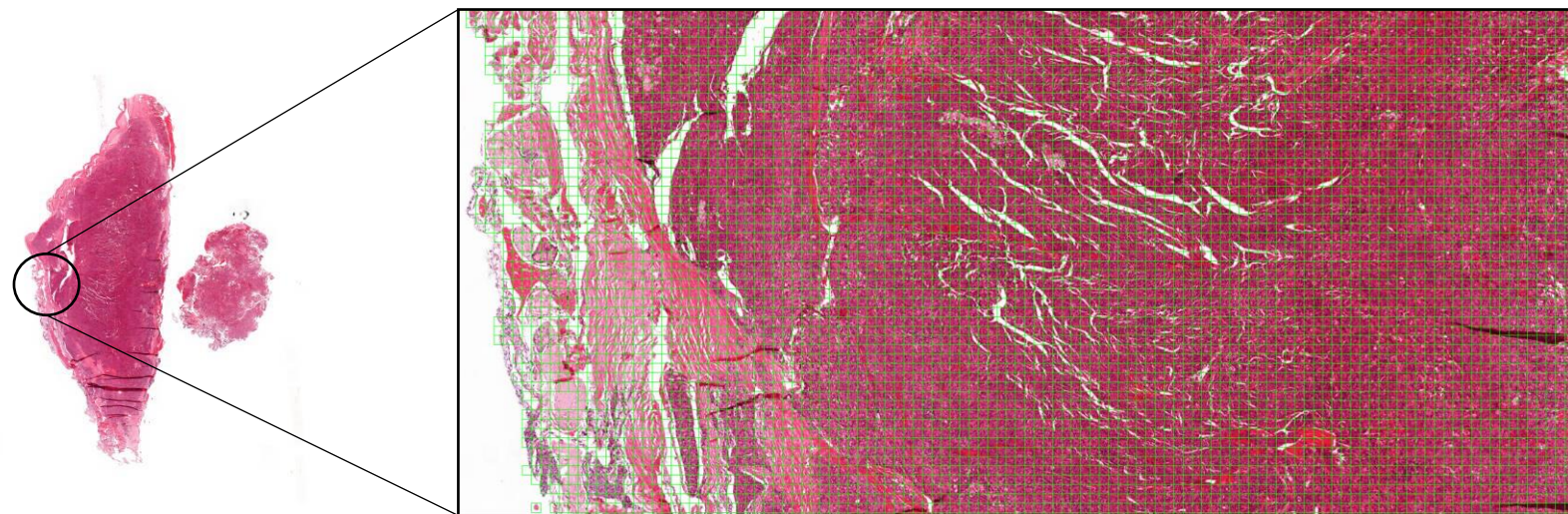
Most Important Tiles

Formalization

Multiple Instance Learning(MIL) – model for aggregating **Instance-Level** information to get a **Bag-Level** prediction.

A bag is composed of

$$\{x_1, \dots, x_{K_m}\}, \qquad K_m\text{- size of Bag}$$

A binary label $Y \in \{0, 1\}$ is associated with every bag.

MIL Assumption: For bag m,

$$Y_m = \begin{cases} 0, & y_k = 0, \forall k \in \{1, \dots, K_m\} \\ 1, & otherwise \end{cases}$$

Pooling Operators

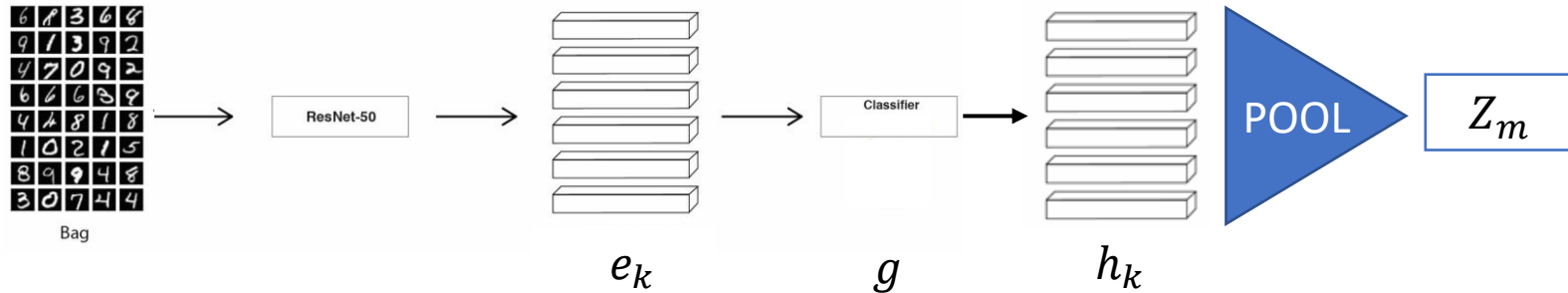In MIL, a pooling operator is applied to aggregate the **instance** embeddings or predictions to create a **bag** output.

Z_m - The Bag-level prediction is aggregated by:

Max-Pooling

$$Z_m := \max_{k \in \{1,\dots,K_m\}} (h_{km})$$

Mean-Pooling

$$Z_m := \frac{1}{K_m} \sum_{k=1}^{K_m} h_{km}$$

A more general formulation has been proposed in [1] by assigning every instance a learned **attention weight**:

$$Z_m := g\left(\sum_{k=1}^{K_m} a_{km} e_{km} \right)$$

Where g is the Bag-Level classifier, e_{km} is the instance embedding, a_{km} is the instance level attention, $k \in \{1, \dots, K_m\}$, in bag m.

[1] Ilse, M., Tomczak, J. M., and Welling, M., "Attention-based Deep Multiple Instance Learning", 35th International Conference on Machine Learning, ICML 2018.

Uncertainty in Deep Learning

Monte-Carlo Dropout[1] is a method for approximating uncertainty in Deep Learning.

- Apply Dropout during inference
- Run forward pass multiple times on the same input data
- Asses certainty by variation of model performance on inputs

In [2] uncertainty was used for multi-task learning by weighting individual task losses to create a global loss function.

[1] Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data.In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. pp. 1183{1192. JMLR. org (2017)
[2] Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7482{7491 (2018)

Method – Certainty Pooling

Intuition – Use certainty to model attention.

We use the Mean-STD MC Dropout method for certainty calculation.

Let X_k be the vector of MC Dropout predictions on instance k.

The instance certainty C_k as:

$$C_k = C(X_k) = \frac{1}{\sigma(X_k) + \epsilon}$$

Where σ is the standard deviation operator and ϵ is a small number that prevents division by zero.

$$Z_m = h_{k^*,m} \text{ where } k^* = \arg\max_k (C_k h_{k,m})$$
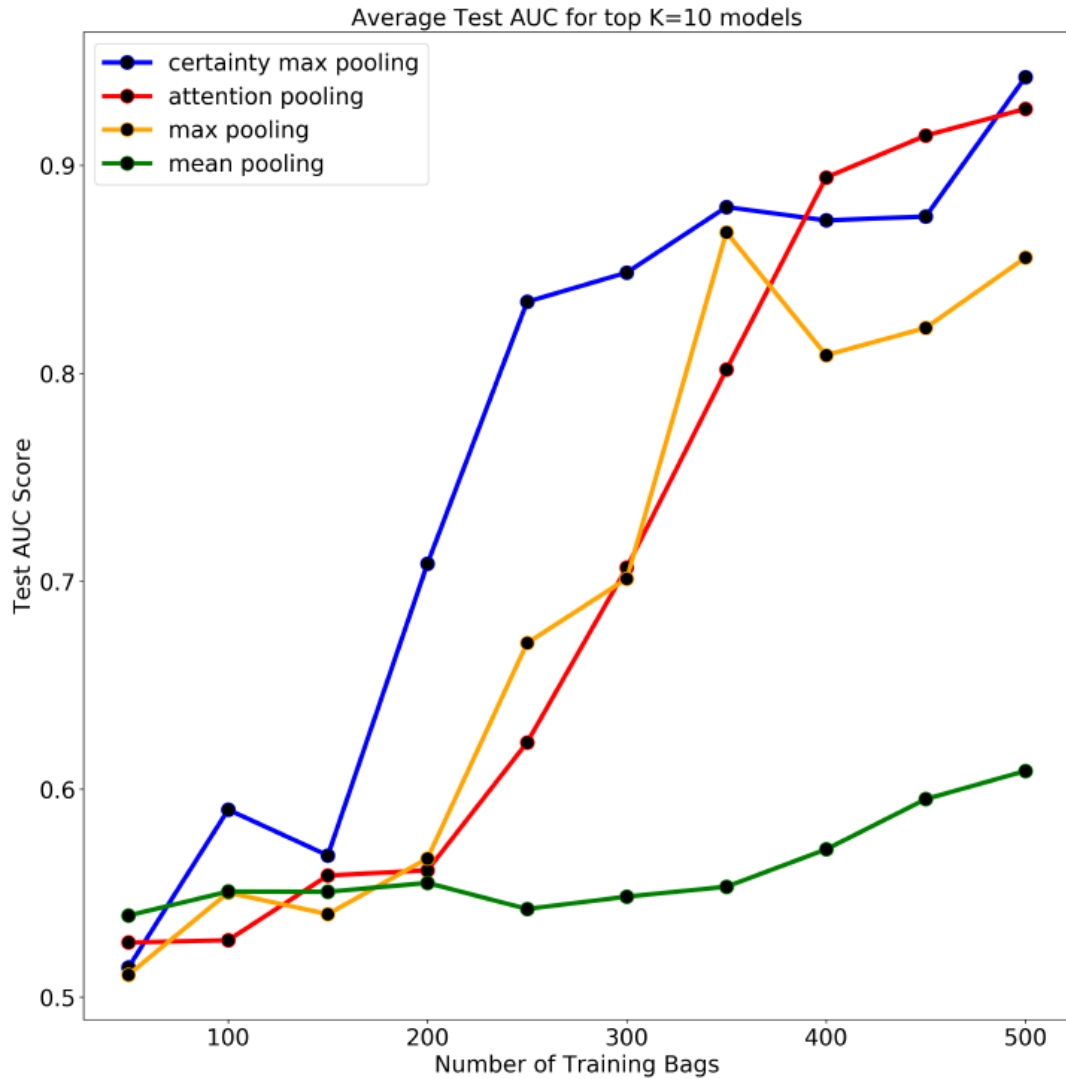
Method – Certainty Pooling
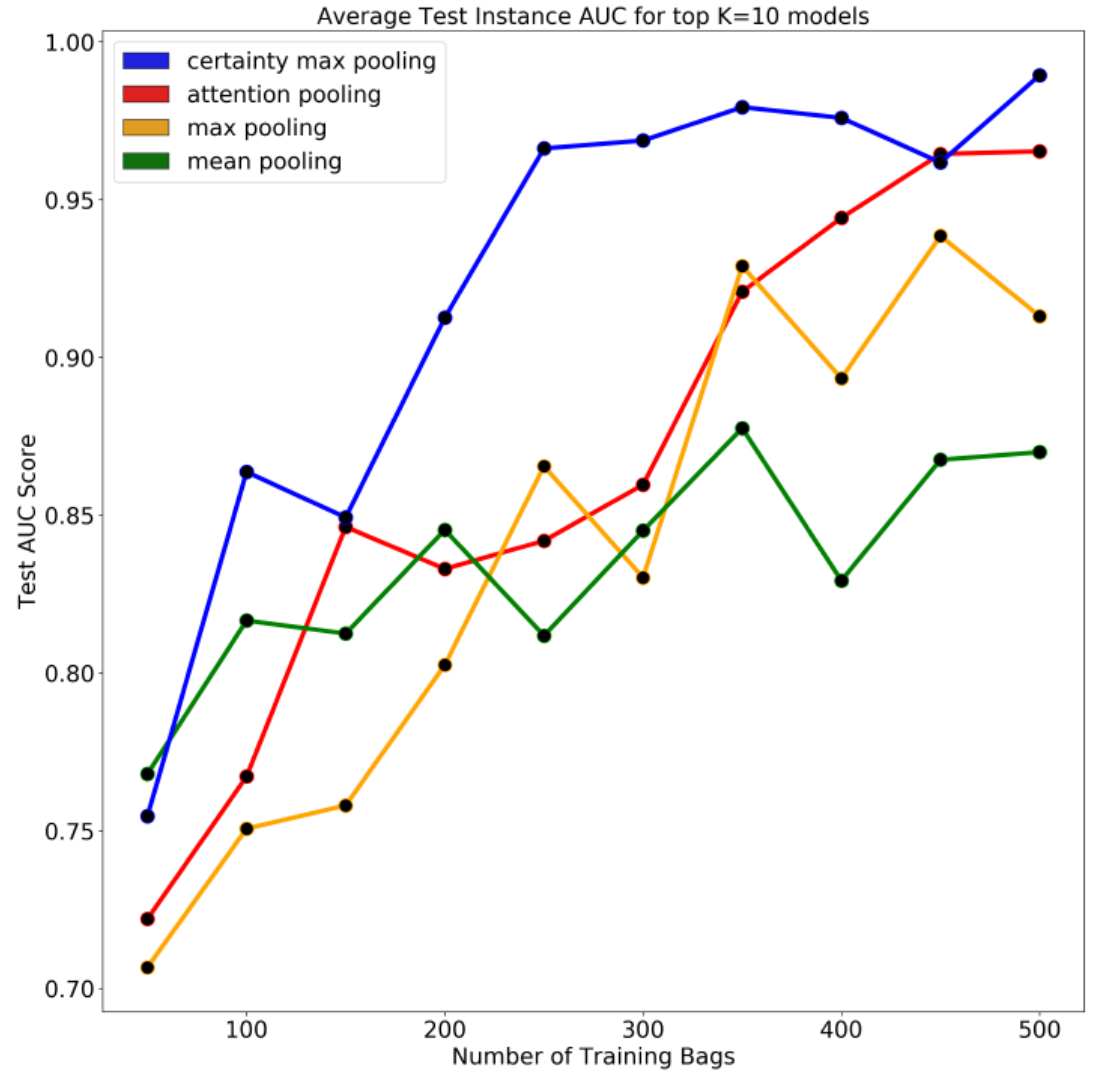
Experiments - Low Evidence Ratio MNIST-Bags

MNIST – Bag Dataset(proposed in [1])

- Instances are MNIST digits, where the bags can have varying size.
- Bags containing the digit 9 are labeled as positive bags and bags without the digit 9 are labeled negative
- As opposed to [1], we look at evidence ratio of 1%(1 – 9 image for every 100 images in a positive bag).

[1] Ilse, M., Tomczak, J. M., and Welling, M., "Attention-based Deep Multiple Instance Learning", 35th International Conference on Machine Learning, ICML 2018.
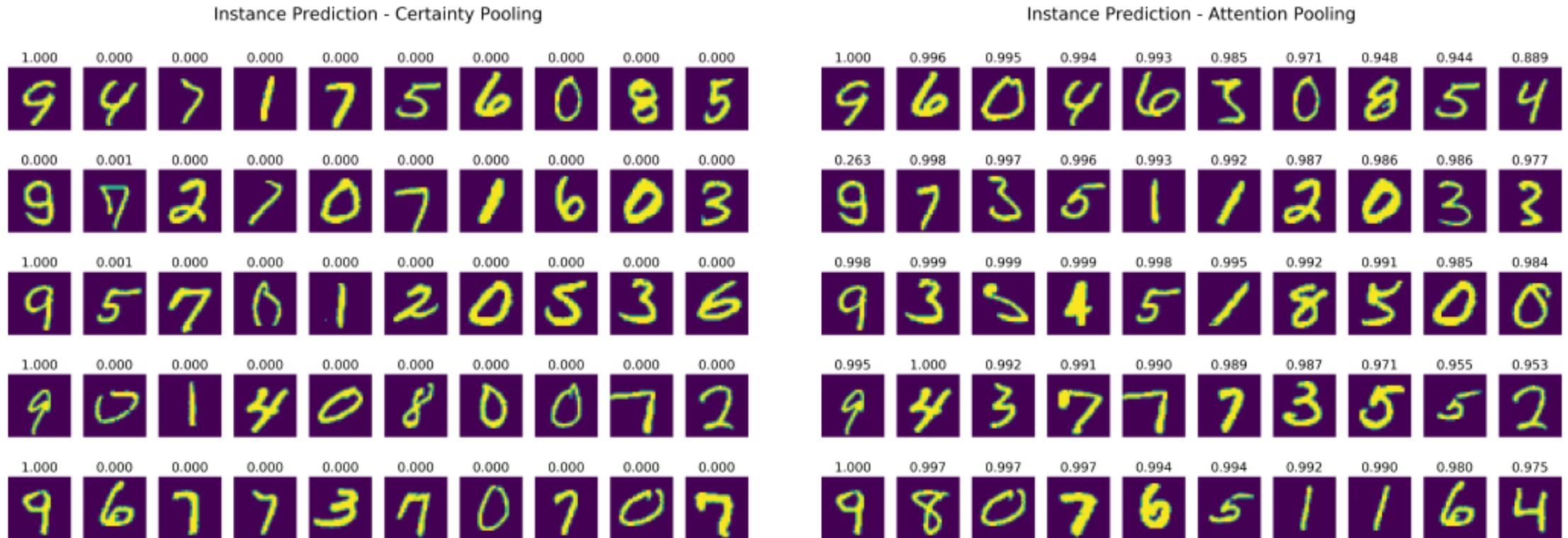
Experiments - Low Evidence Ratio MNIST-Bags

Testing dataset bag prediction AUCs for models trained with different sizes of low evidence ratio MNIST bag datasets. In all experiments we consider the average of the top K=10 runs.

Average instance prediction AUC values for the top K=10 models.

Experiments - Low Evidence Ratio MNIST-Bags

Example instance level prediction values for Certainty Pooling MIL (left) and Attention Pooling MIL (right). The positive instance followed by the highest value instances in the bag ordered in descending order are presented and the value above each image represents the prediction value for that instance.

Camelyon16 Lymph Node Metastasis Detection Challenge

- 400 H&E stained WSIs taken from sentinel lymph nodes, which are either healthy or exhibit metastases of some form.

- The dataset contains both slide level annotations, i.e. healthy or contains-metastases and pixel level segmentation masks per slide denoting the metastases.

- Tile Size 256x256, without overlap at 20X resolution.

- Tissue Detection was used to discard Background Tiles. We apply a simple LAB color space stain normalization.

- Bag consists of all tiles in a Slide.

- We test affects of varying dataset size by randomly sampling percentage of 270 Train WSIs.
- Test Set – 130 WSIs. We holdout 25% of the train set for validation.

# Camelyon16 Lymph Node Metastasis Detection Challenge

- Following [1] we avoid learning directly on the instance
images because of the computational cost and instead we first extract 2048 length
features for every instance using Resnet50 pre-trained on ImageNet.

- Similar to [2], we randomly sample 128 instances during train from every bag at each Epoch.

[1][Campanella, G., Silva, V.W.K., Fuchs, T.J.: Terabyte-scale deep multiple in-stance learning for classifcation and localization in pathology. arXiv preprint arXiv:1805.06983 (2018)

[2]Couture, H.D., Marron, J.S., Perou, C.M., Troester, M.A., Niethammer, M.: Multiple instance learning for heterogeneous images: Training a cnn for histopathology. In: International conference on Medical Image Computing and Computer-Assisted Intervention. pp. 254{262. Springer (2018)
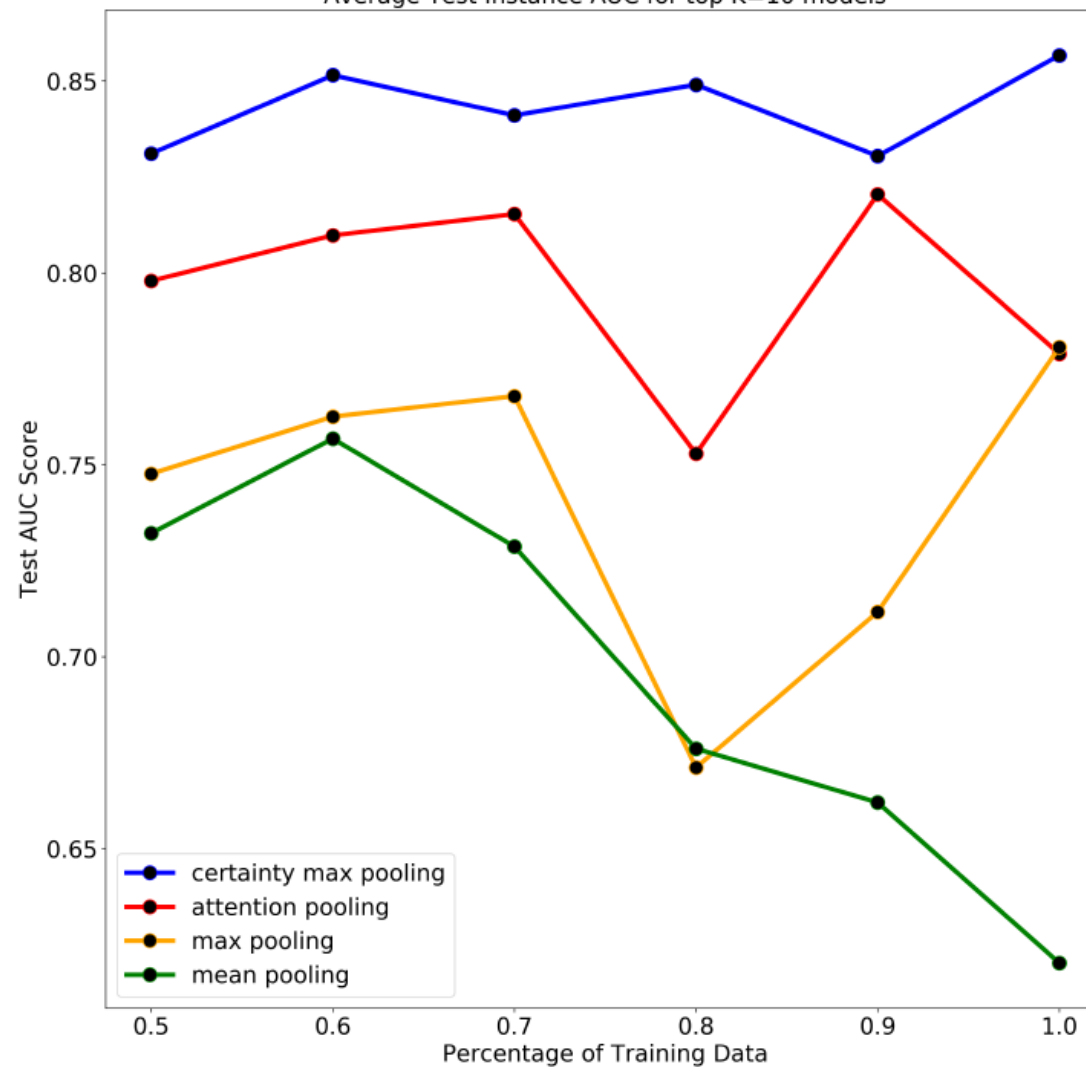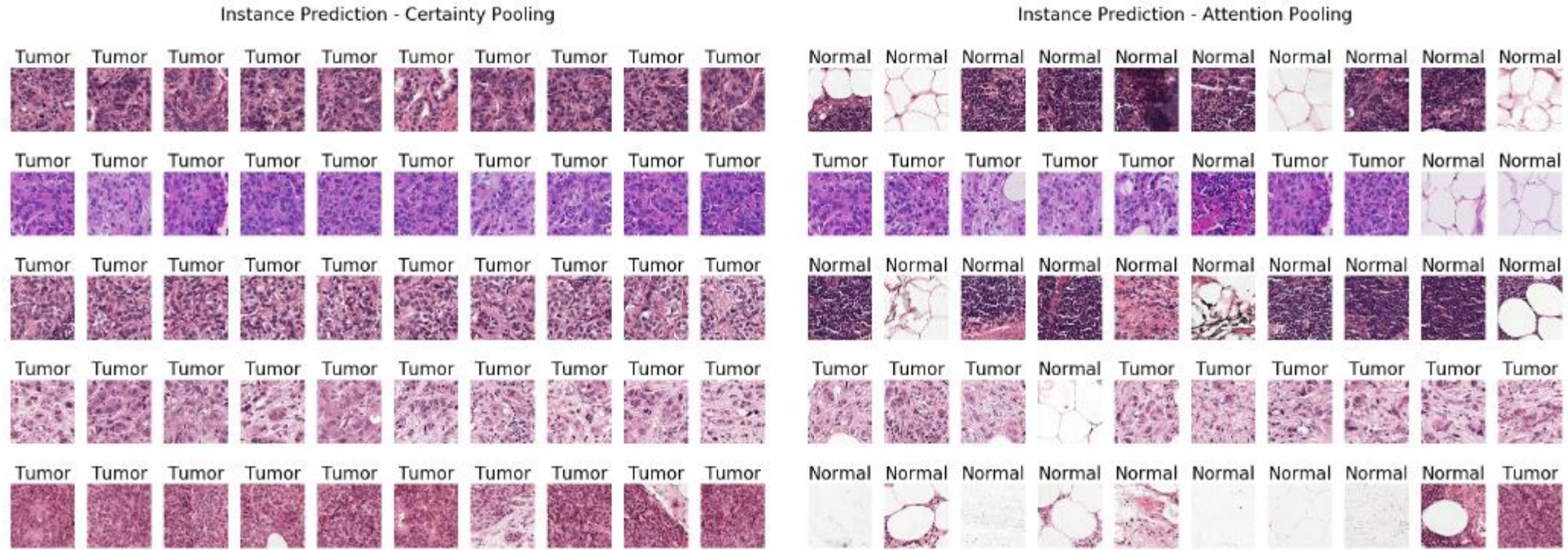
Camelyon16 Lymph Node Metastasis Detection Challenge
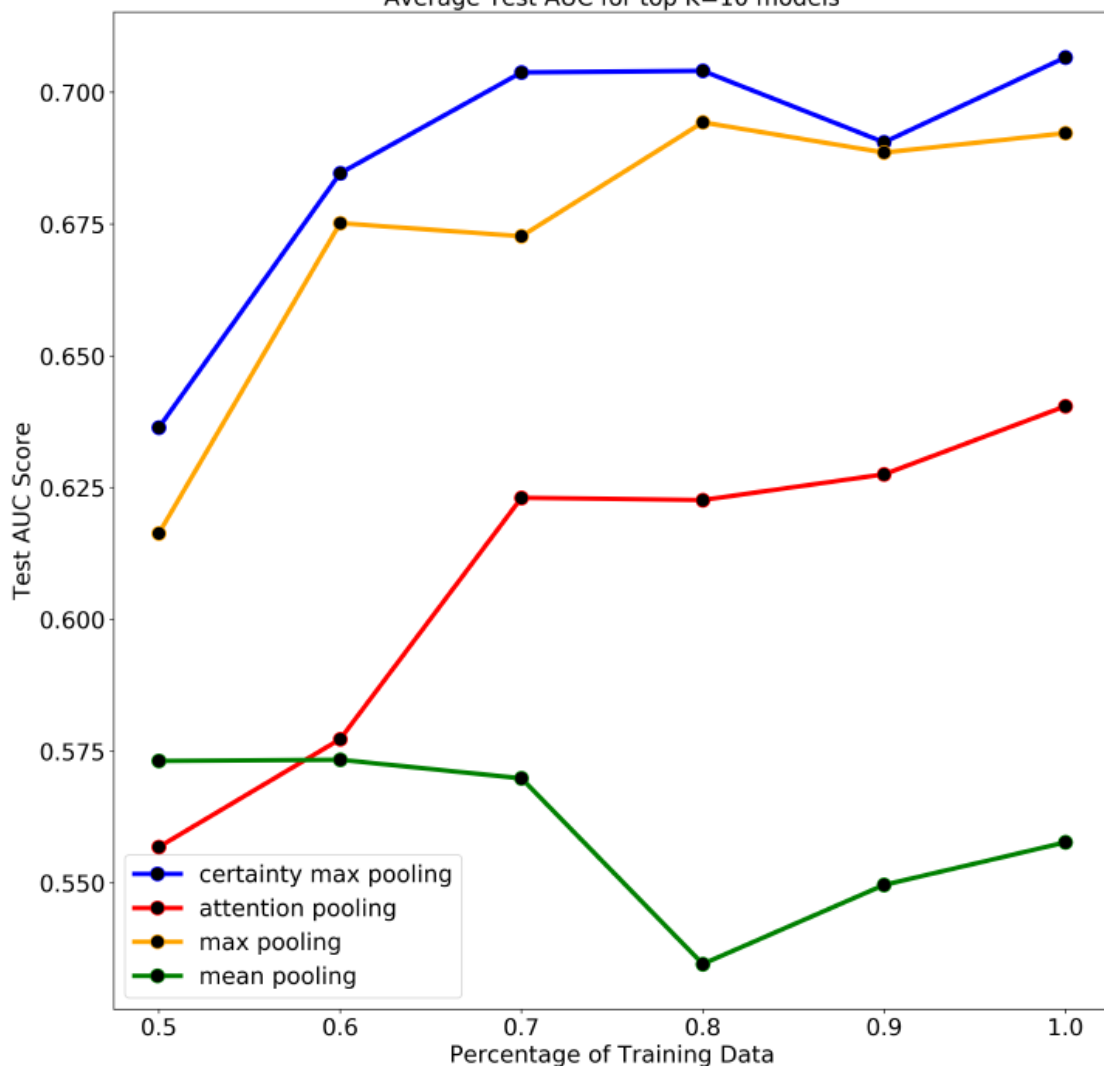
Camelyeon16 testing set average top K=10 global bag prediction AUCs for different training dataset sizes.

Camelyeon16 testing set average top K=10 global instance prediction AUCs for different training dataset sizes.

Camelyon16 Lymph Node Metastasis Detection Challenge

Example instance level prediction tiles for Certainty Pooling MIL (left) and Attention Pooling MIL (right). The top instances sorted by predicted instance (tile) value for 5 "Tumor" labeled slides. Above each tile the instance level label from the Camelyon16 mask labels.

Summary

- We present a novel MIL approach that excels in low evidence ratio domains

- We demonstrate our results in a hard synthetic task and in a challenging real world problem

- Future work includes Learning to output the certainty for every tile from the network instead of using MC dropout

For further questions - ido@deepathology.ai