Sparse Network Inversion for Key Instance Detection in Multiple Instance Learning

Beomjo Shin, Junsu Cho, Hwanjo Yu* Seungjin Choi
POSTECH, South Korea BARO AI, South Korea

Multiple Instance Learning (MIL)

- MIL learns relationship between a set of instances called bag and binary bag label.
- MIL includes two tasks: bag-level classification, which predicts a binary label for a bag; Key Instance Detection (KID), which detects positive instances (key instances) when a bag is categorized as a positive bag.
- Our goal is to **improve KID performance of MIL model that uses attention mechanism** [1] while maintaining the bag-level classification performance.

Attention-based deep MIL model

- Attention-based deep MIL model [1] works in following two steps. The MIL model uses bag score for bag-level classification and uses attention scores for KID when bag-level prediction is positive.

1) Make a bag embedding by aggregating K instance embeddings through attention-based pooling.

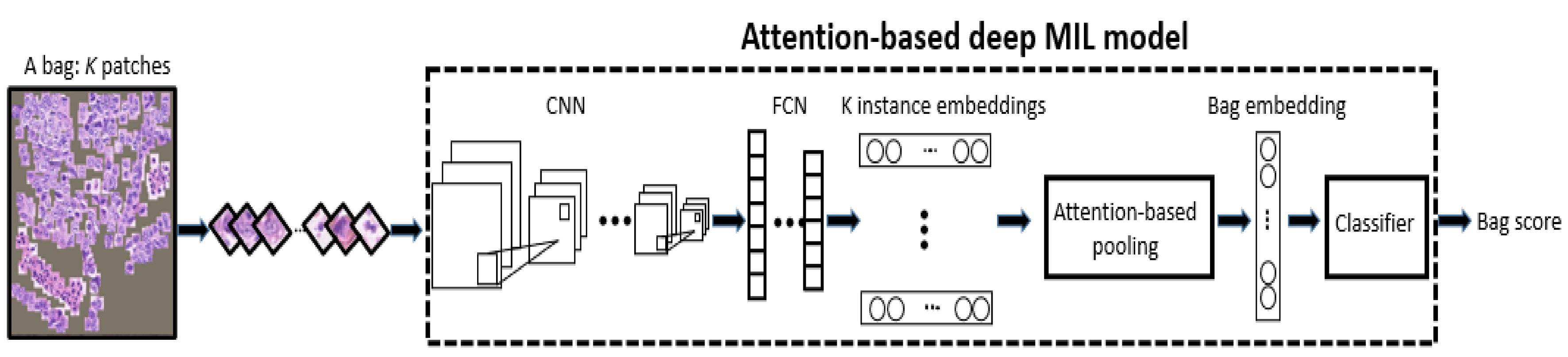
Bag embedding *Attention score of k^{th} instance of n^{th} bag*

$$z_n = \sum_{j=1}^{K} a_{n,j} g_{n,j} \qquad a_{n,k} = \frac{\exp\left\{w^\top \tanh(V g_{n,k}^\top)\right\}}{\sum_{j=1}^{K} \exp\left\{w^\top \tanh(V g_{n,j}^\top)\right\}}$$

2) Calculate a bag score using classifier.

Bag embedding

$$P_n = \text{sigmoid}\left(H\left(z_n\right)\right)$$



Limitation of Attention-based deep MIL model

- Attention-based deep MIL model has limits in KID performance due to the following reasons.

  - Attention-based deep MIL model is trained by optimizing negative log likelihood for the bag-level label, and the bag in MIL is labeled as positive bag if at least one of the instances in the bag is positive instance. Thus, to minimize the negative log likelihood, It is advantageous for the MIL model to focus only on few distinguishable key instances, not all key instances.

Ground truth **Attention model**

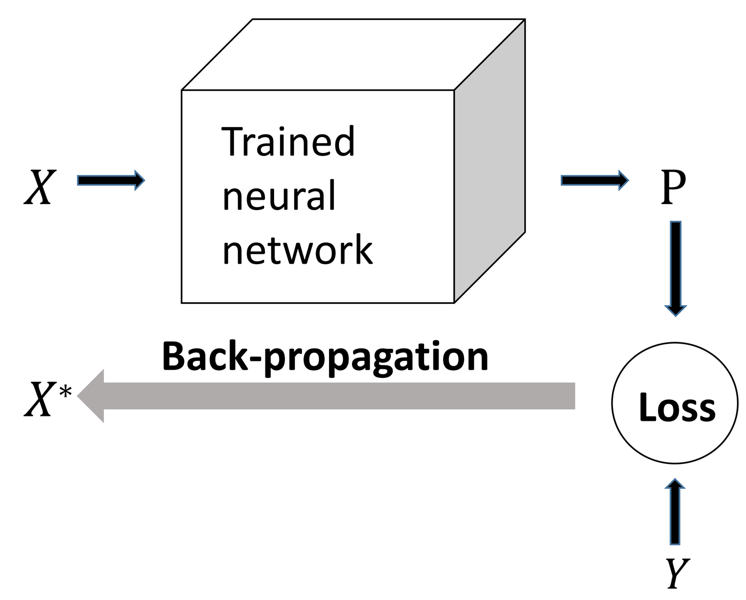
Our Key Idea

- We solve the problem that attention-based deep MIL model focuses only on few distinguishable key instances by removing the constraint that input data cannot be changed.

Neural Network Inversion with a sparseness constraint (Sparse Network Inversion)

Neural Network Inversion

- Find optimal input pattern by using trained neural network, random initialized input, and target label [2, 3].



$$X^{t+1} = X^t - \eta \frac{\partial Loss}{\partial X^t}$$

Sparse Network Inversion for KID in MIL

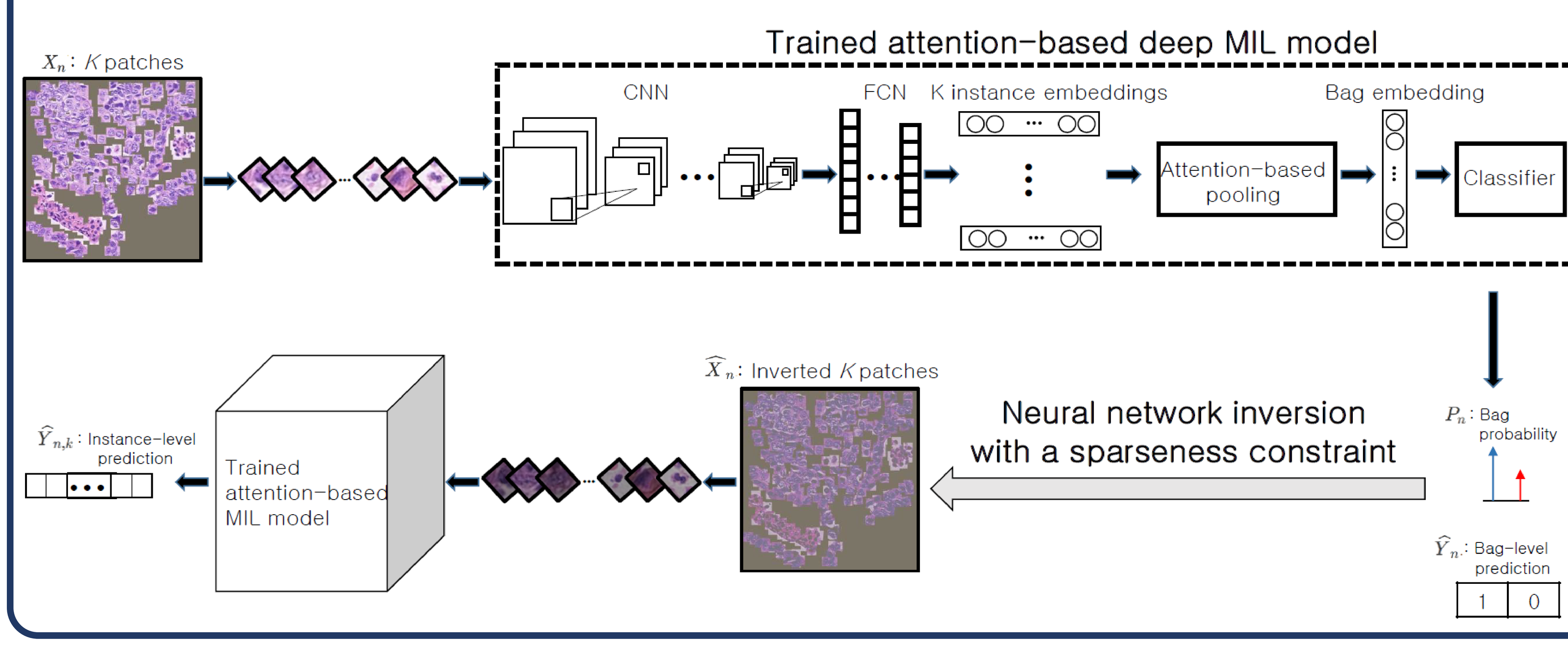
- To apply Sparse Network Inversion in inference time, we change random initialized input to input bag, target label to predicted label. In addition, we incorporate a sparseness constraint.

Loss of Sparse Network Inversion

$$Loss(X) = -\hat{Y}\log P - \left(1-\hat{Y}\right)\log(1-P) + \lambda\|X\|_1$$

Update equation of Sparse Network Inversion

$$X_n^{t+1} = s_\lambda(X_n^t - \eta\nabla l(X_n^t))$$



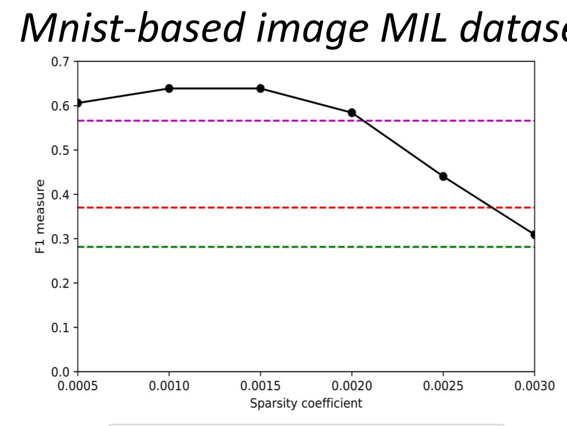
Experimental Results

- Attention-based deep MIL model achieves the **best bag-level classification performance.**

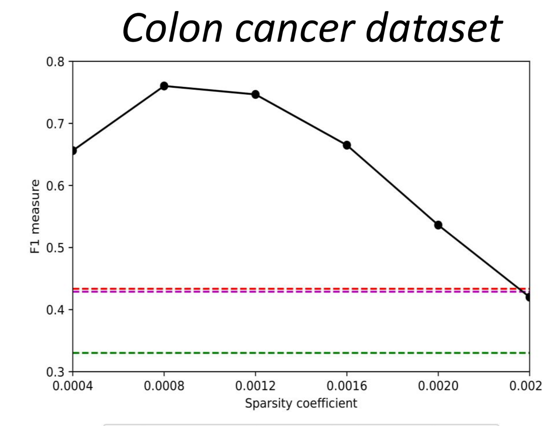
Method	MNIST-based image MIL	COLON CANCER	BREAST CANCER
Inst+max	0.804 ± 0.232	0.868 ± 0.025	0.536 ± 0.062
Inst+mean	0.708 ± 0.095	0.798 ± 0.023	0.612 ± 0.038
Attention	0.996 ± 0.008	0.909 ± 0.02	0.718 ± 0.054

- Sparse Network Inversion **significantly improves the KID performance** while maintaining bag-level classification performance.

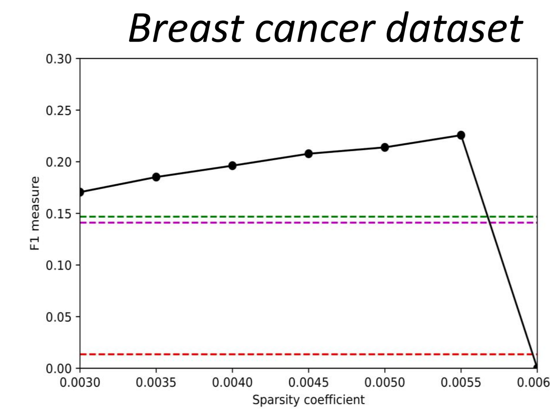
Mnist-based image MIL dataset



Colon cancer dataset



Breast cancer dataset



[1] Ilse et al., Attention-based deep multiple instance learning, ICML 2018
[2] Kindermann et al., Inversion of neural networks by gradient descent, Parallel Computing 1990
[3] Hoskins et al., Iterative inversion of Neural networks and its application to adaptive control, Trans. Neur.Netw., 1992