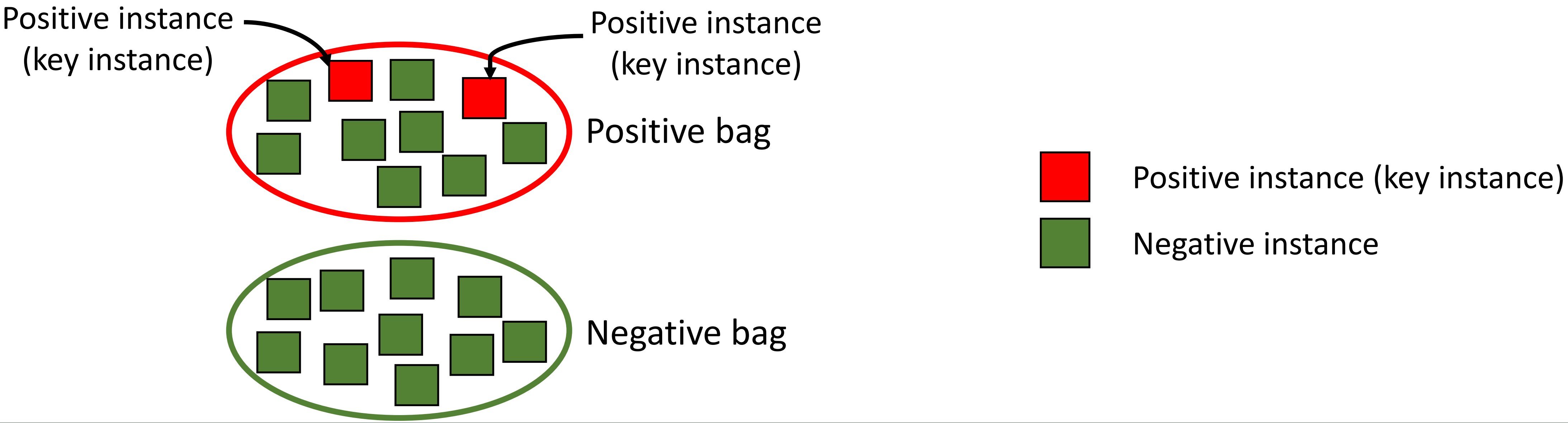
Sparse Network Inversion for Key Instance Detection in Multiple Instance Learning

ICPR20

Beomjo Shin[1], Junsu Cho[1], Hwanjo Yu[1]*, Seunjin Choi[2]
Pohang University of Science and Technology, South Korea[1]
BARO AI, South Korea[2]

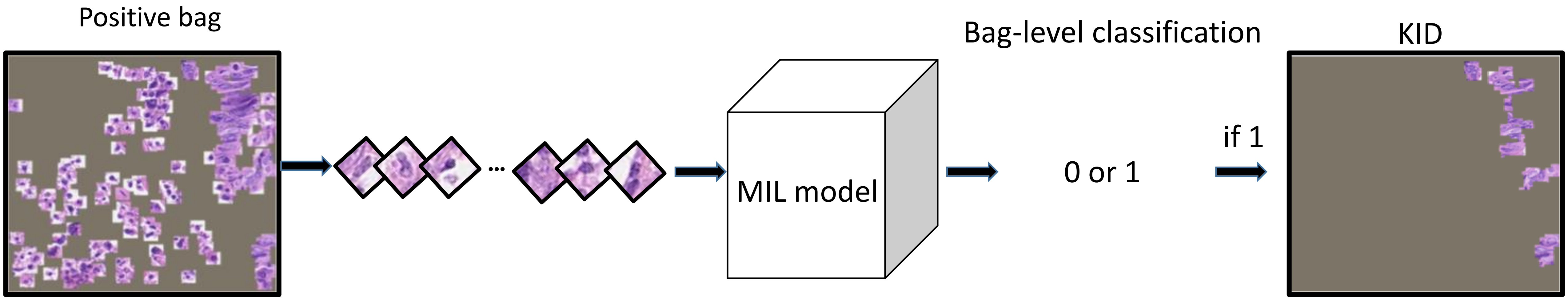
Multiple Instance Learning (MIL)

- Weakly Supervised Learning
- Property
  - Bag: **<u>group of multiple instances</u>**
  - Positive bag: negative instances + **<u>at least</u>** one positive instance (key instance)
  - Negative bag: only negative instances



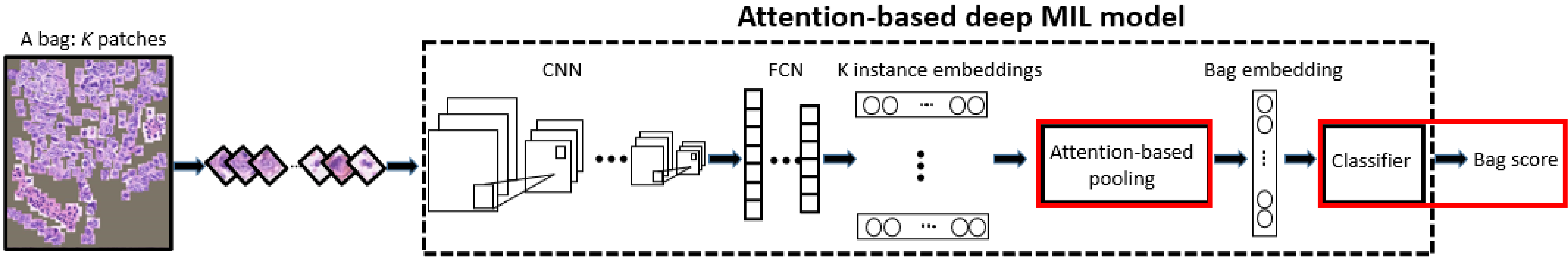
Multiple Instance Learning (MIL)

- Tasks: Bag-level classification and Key Instance Detection (KID)
  - Input: bag
  - Output: bag-level label and instance-level labels in a positive bag
  - Application: image classification with ROIs



Conventional Method

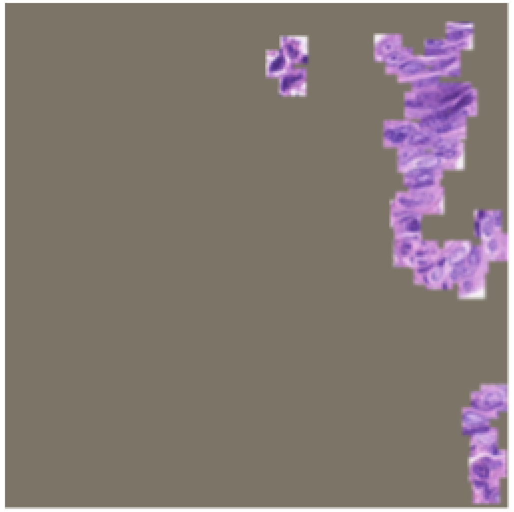
- Attention-based deep Multiple Instance Learning (MIL) model
  - Bag-level classification: bag score
  - Key instance detection :attention scores if bag-level prediction is positive



1) Ilse et al., Attention-based deep multiple instance learning, ICML 2018

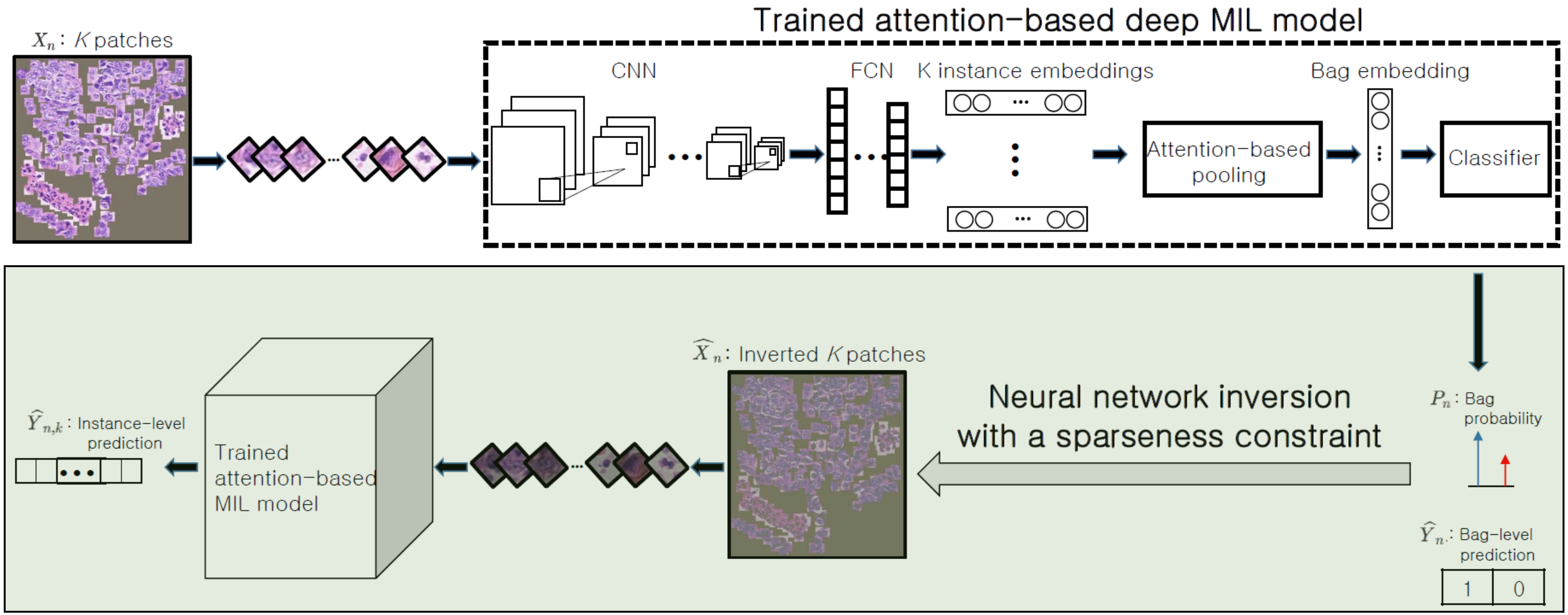
Limitation of Conventional Method

- Limitation: **Low Key Instance Detection (KID) Performance**
  - The model cannot learn every key instance in detail



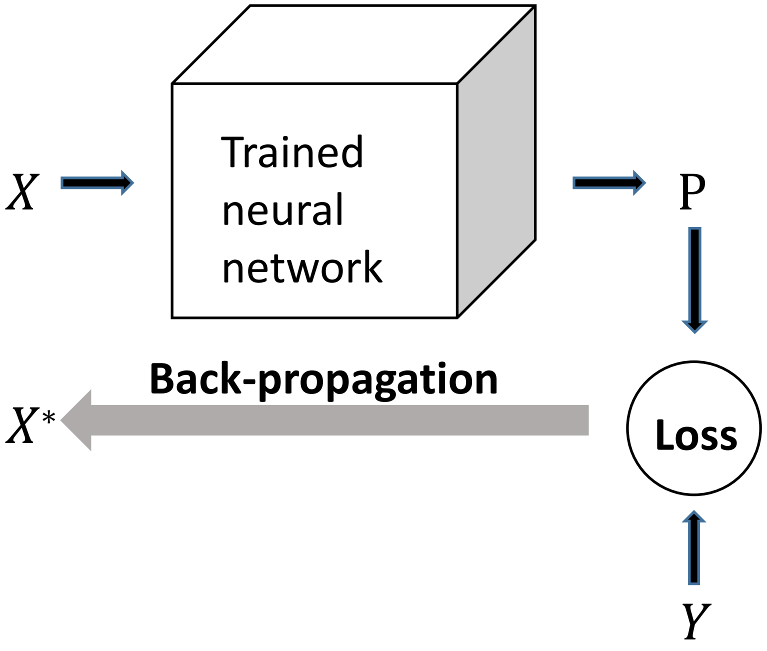
Ground truth **Attention model**

Sparse Network Inversion



Neural Network Inversion

- Task: **Finding Optimal Input Pattern**
  - Input: random initialized input, label
  - Output: optimal input pattern



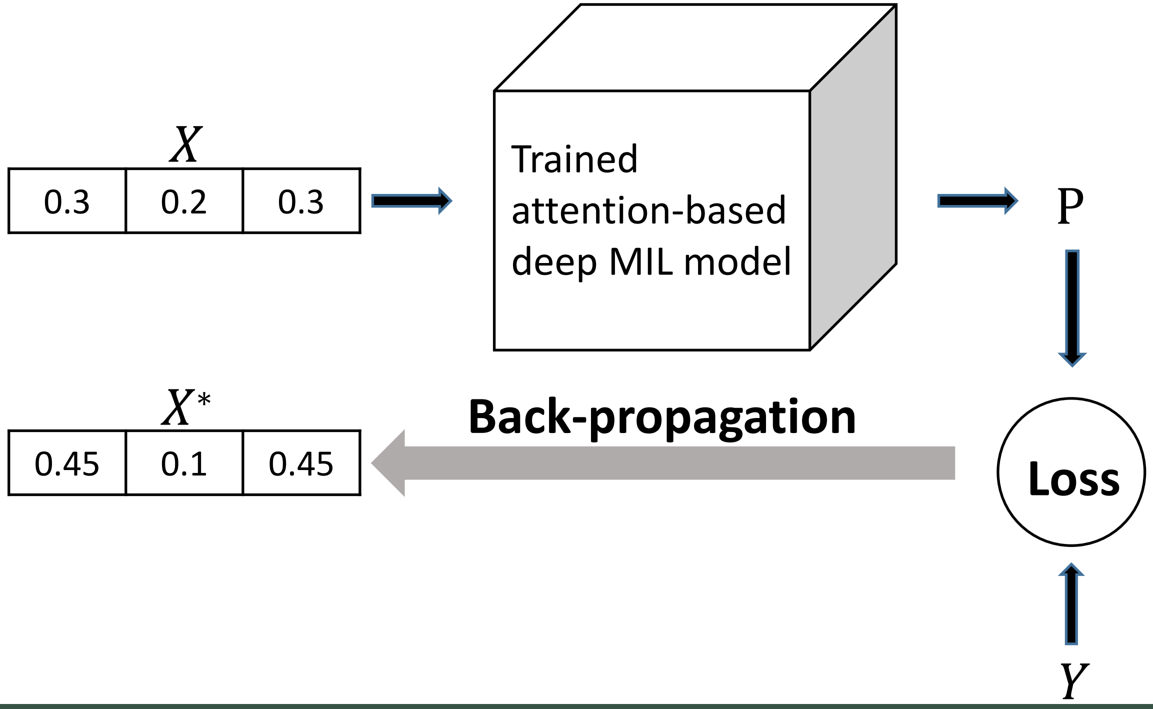
$$\boldsymbol{X^{t+1} = X^{t} - \eta \frac{\partial Loss}{\partial X^{t}}}$$

X: random initialized input
Y: target label
P: output
X^*: optimal input pattern
Loss: loss of model

1) Kindermann et al., Inversion of neural networks by gradient descent, Parallel Computing 1990
2) Hoskins et al., Iterative inversion of Neural networks and its application to adaptive control, Trans. Neur.Netw., 1992

Sparse Network Inversion for KID in MIL

- For KID in MIL
  - **Random initialized input → Input bag**
  - **Label → Predicted label**
  - **No constraints → Sparseness constraint**



$$l(X) = -\hat{Y}\log P - \left(1-\hat{Y}\right)\log(1-P)$$

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

X: input bag
P: bag score
$\hat{Y}$: predicted label
Loss: loss of model
X^*: optimal input pattern

Experimental Setting

- Datasets
  - MNIST-based image MIL dataset
  - Two histopathology datasets (COLON CANCER, BREAST CANCER)
- Baseline Methods
  - Instance-space paradigm methods
    - Instance-space paradigm method that use max pooling
    - Instance-space paradigm method that use mean pooling
  - Attention-based deep MIL model

Dataset	Positive	Negative
MNIST-based image MIL	50%	50%
COLON CANCER	51%	49%
BREAST CANCER	44.8%	55.2%

Positive bags vs Negative bags

Dataset	Positive	Negative
MNIST-based image MIL	10.2%	89.8%
COLON CANCER	53.5%	46.5%
BREAST CANCER	8.5%	91.5%

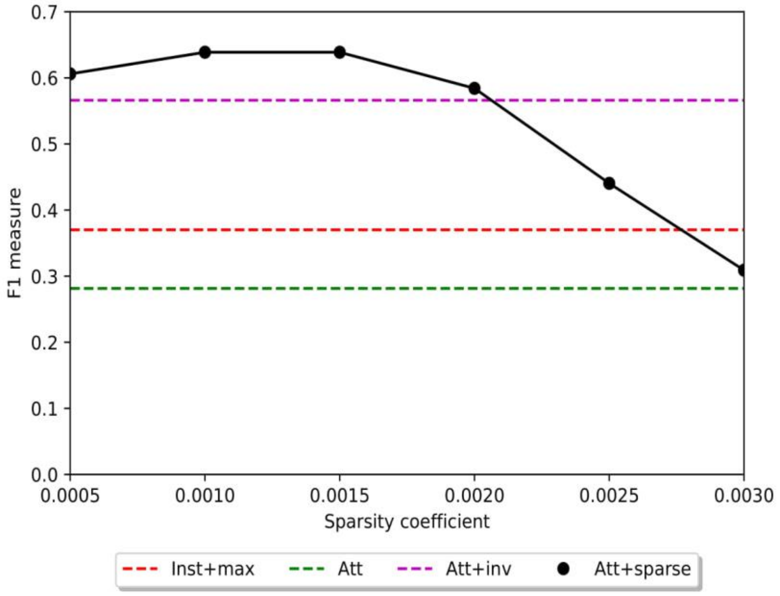
Positive instances vs Negative instances

Result and Analysis

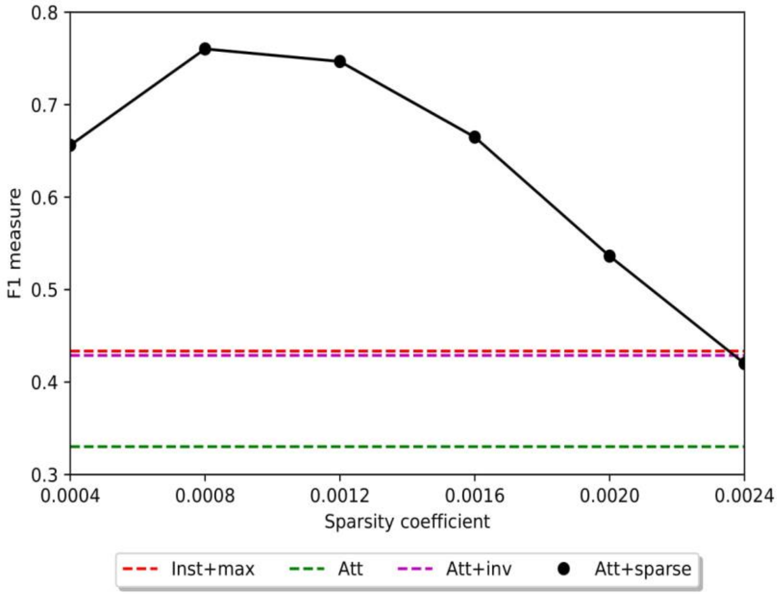
Method	MNIST-based image MIL	COLON CANCER	BREAST CANCER
Inst+max	0.804 ± 0.232	0.868 ± 0.025	0.536 ± 0.062
Inst+mean	0.708 ± 0.095	0.798 ± 0.023	0.612 ± 0.038
Attention	0.996 ± 0.008	0.909 ± 0.02	0.718 ± 0.054

Bag-level Classification Performance (Accuracy)

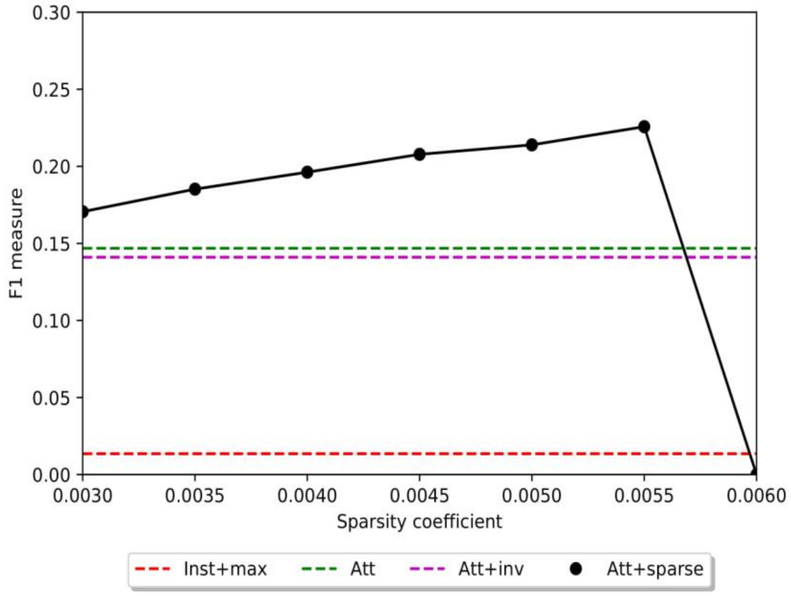
Result and Analysis



Mnist-based image MIL dataset



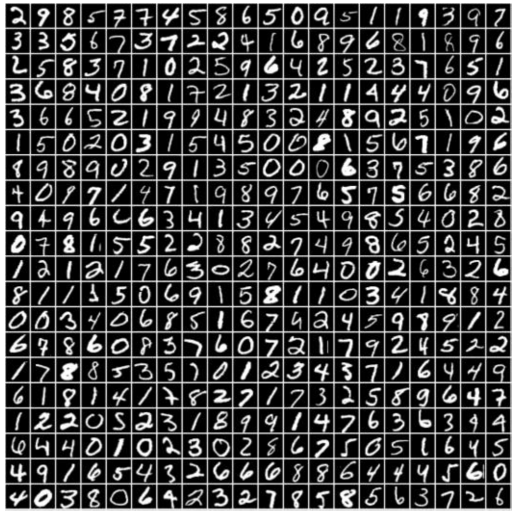
Colon cancer dataset



Breast cancer dataset

KID Performance (F1 measure)

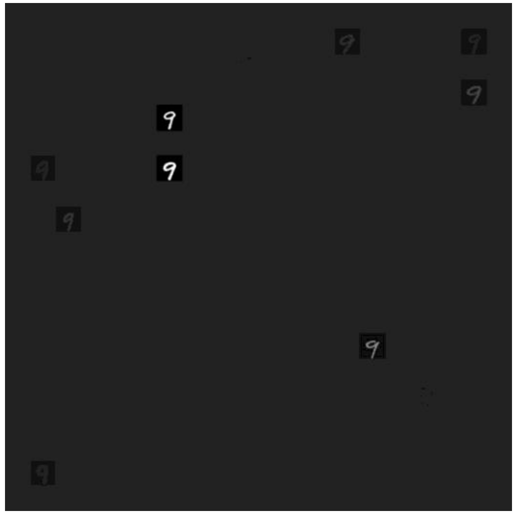
Result and Analysis



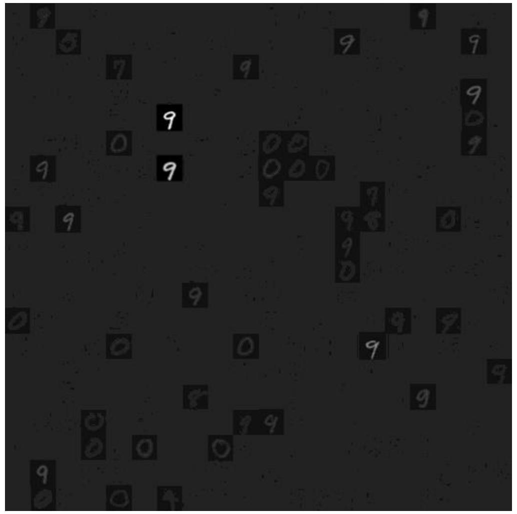
a. All patches

b. All refined patches

Result and Analysis



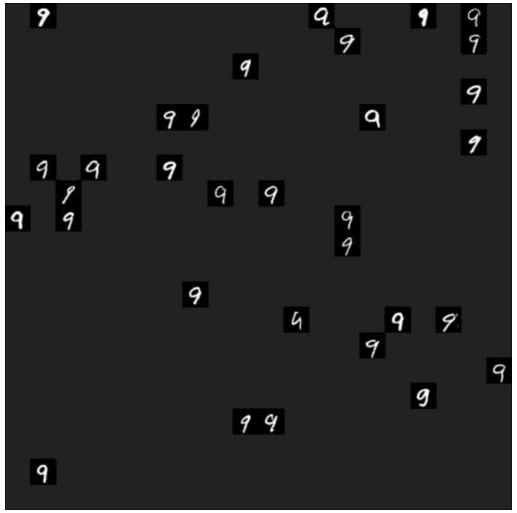
c. Heatmap of att



d. Heatmap of sparse

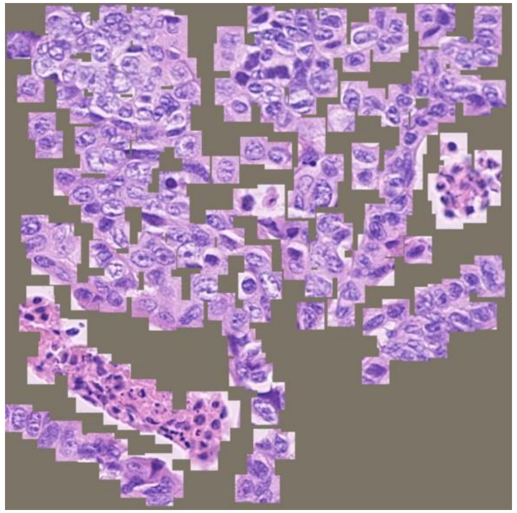
e. Visualization result of att

f. Visualization result of sparse

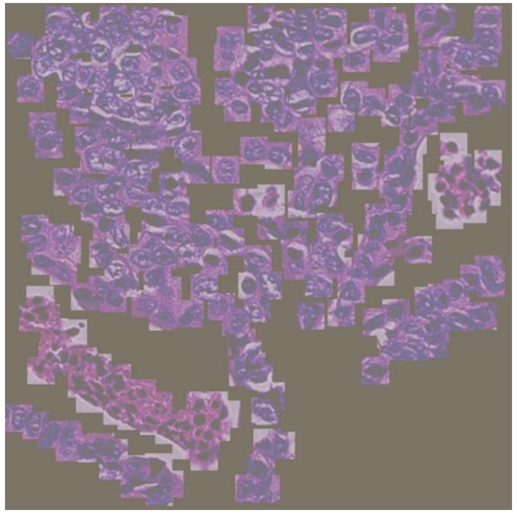


g. Ground truth

Result and Analysis

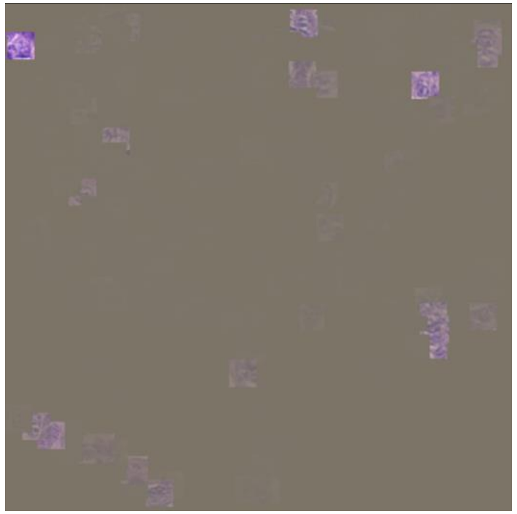


a. All patches

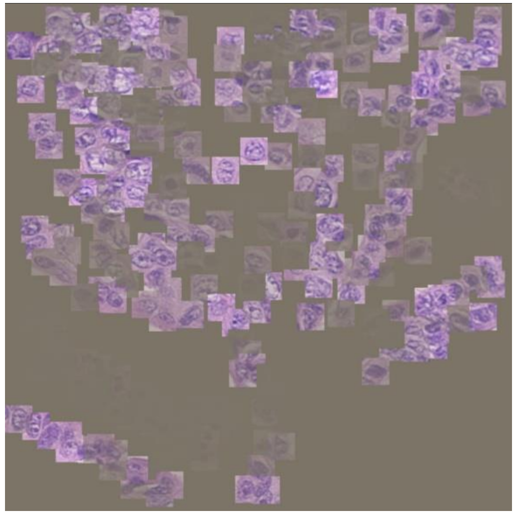


b. All refined patches

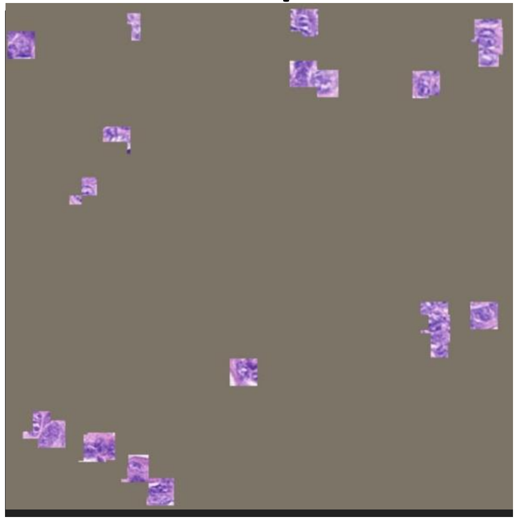
Result and Analysis



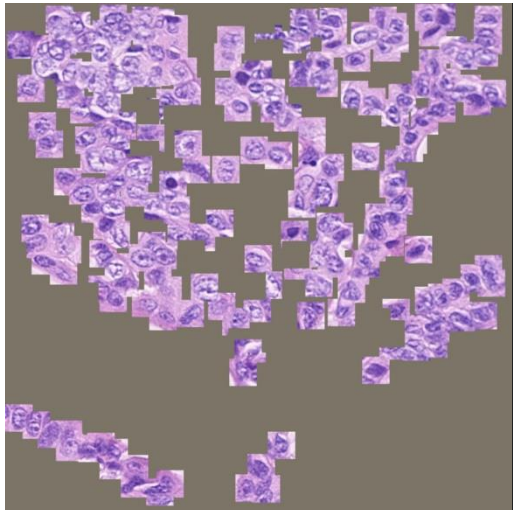
c. Heatmap of att



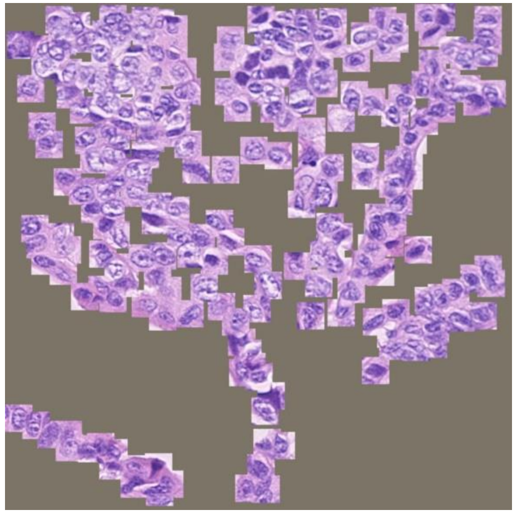
d. Heatmap of sparse



e. Visualization result of att



f. Visualization result of sparse



g. Ground truth