Accounting for Dependencies in Deep Learning Based Multiple Instance Learning for Whole Slide Imaging

Medical Deep-Learning 2nd Paper review

Kyonggi Univ. Smart I.O.T Lab 이상민 (2021년 11월 24일)

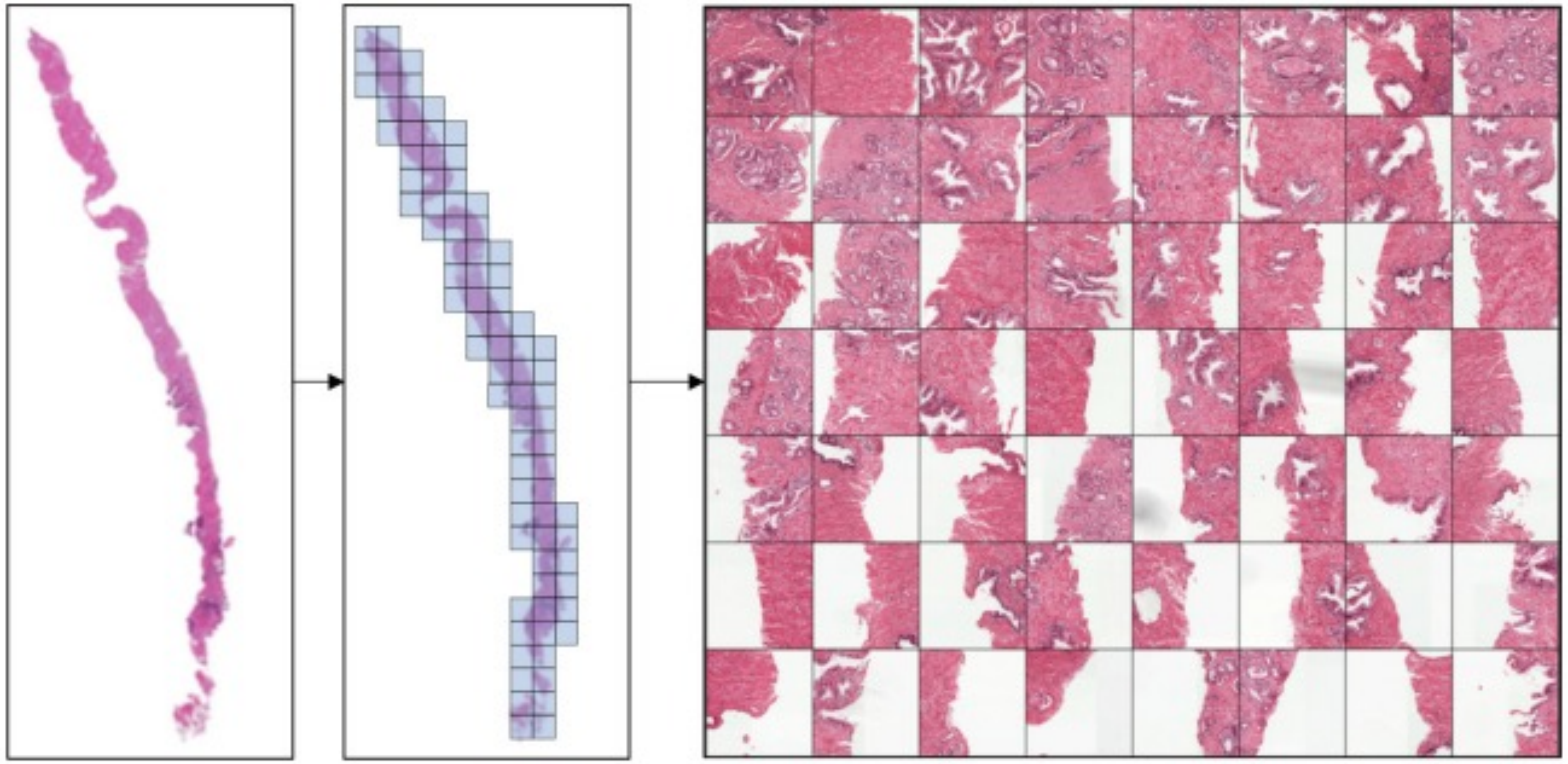
목차

01. Introduction

Multiple instance learning

01. Introduction

Patch extraction from Whole slide images



Bag (instance set)

Patches(Instance)

KYONGGI UNIVERSITY

02. Method

H(Bag) = hk(instance), k는 instance의 개수

$$\mathbf{H} = \{\mathbf{h}_1, \ldots, \mathbf{h}_K\}$$

Y는 제공되는 Bag의 라벨

$$Y = \begin{cases} 0, & \text{iff all } y_k = 0, \\ 1, & \text{iff any } y_k = 1. \end{cases} = \max_k\{y_k\}$$

Attention weights of patch embeddings

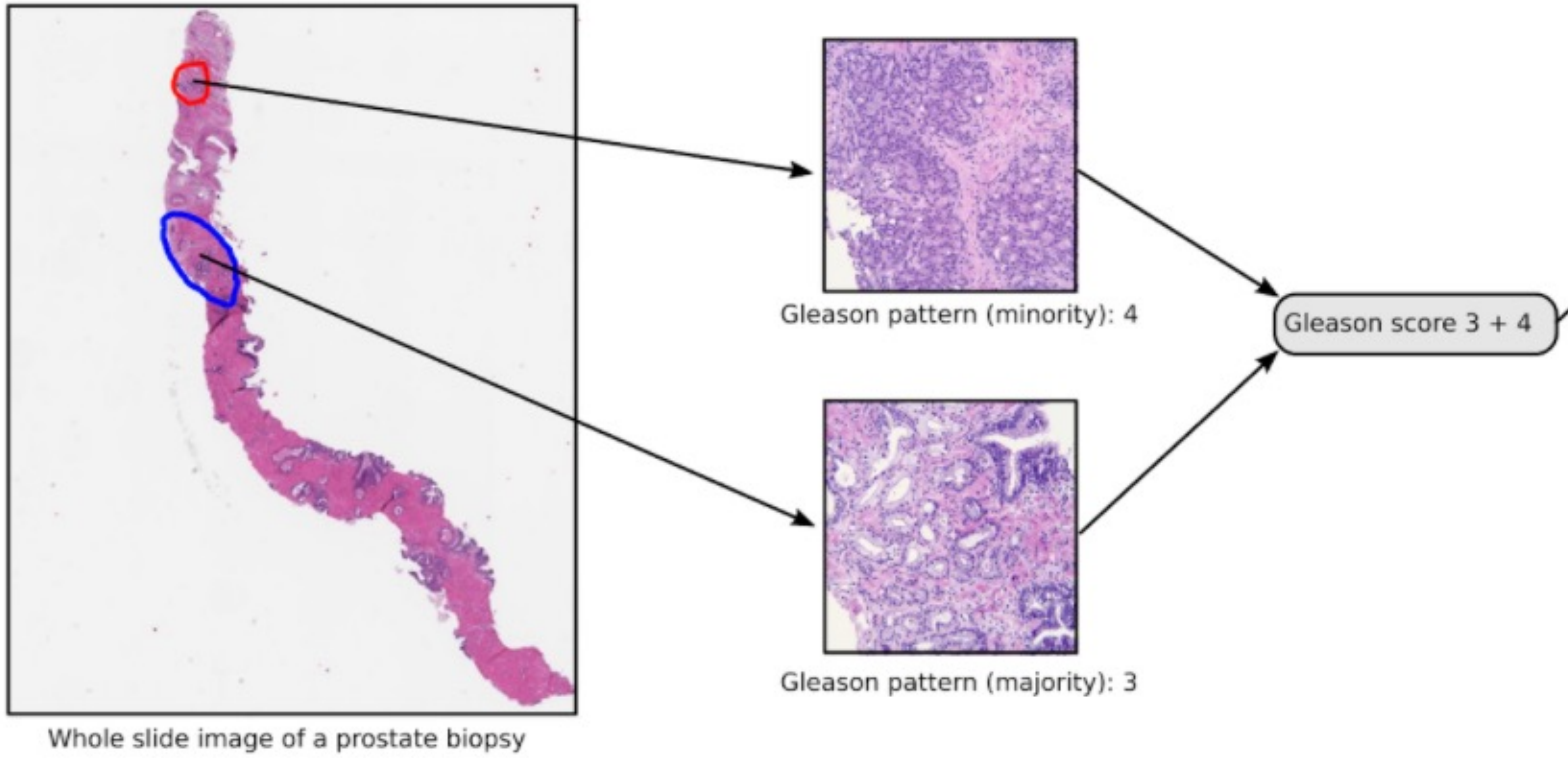
$$\mathbf{a} = softmax(\tanh(\mathbf{HV})\mathbf{w})$$

Linear combination of Patch embedding

$$\mathbf{z} = \sum_{k=1}^{K} a_k \mathbf{h}_k = \mathbf{Ha}$$

02. Method

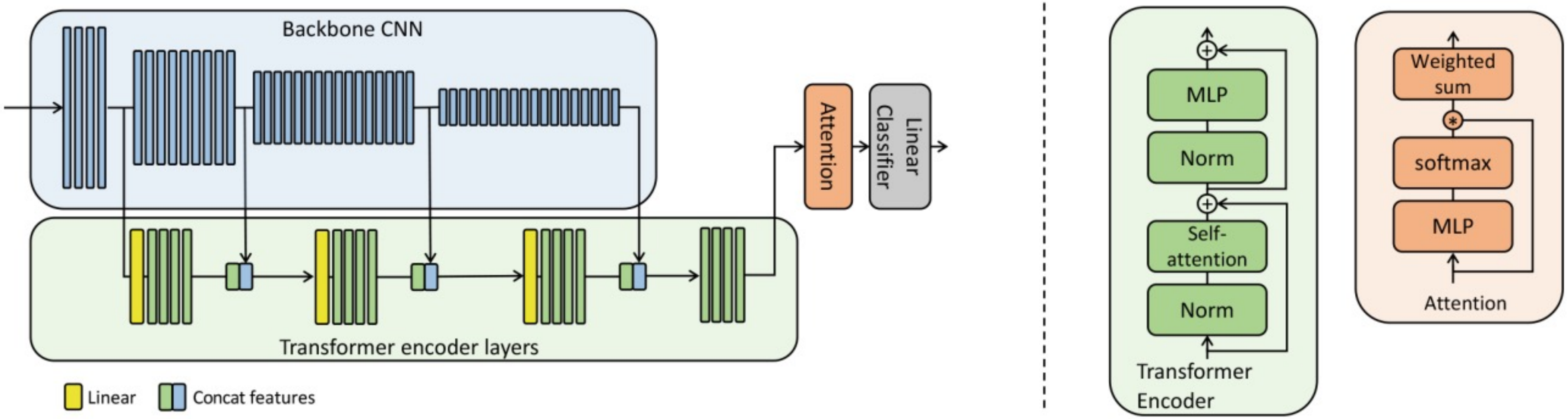
Dependency Between Instances



Gleason score	ISUP grade
3+3	1
3+4	2
4+3	3
4+4	4
3+5	4
5+3	4
4+5	5
5+4	5
5+5	5

02. Method

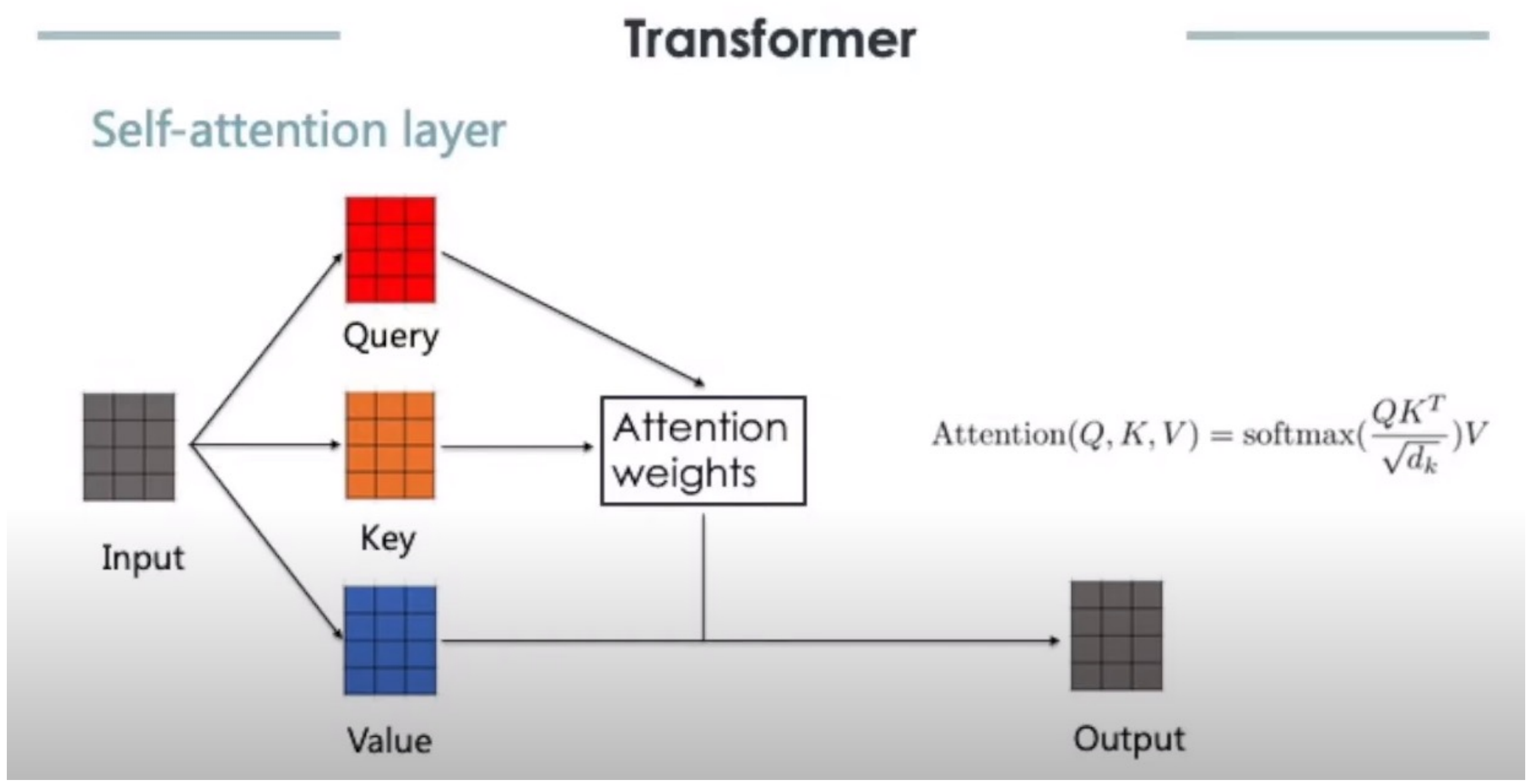
Model architecture



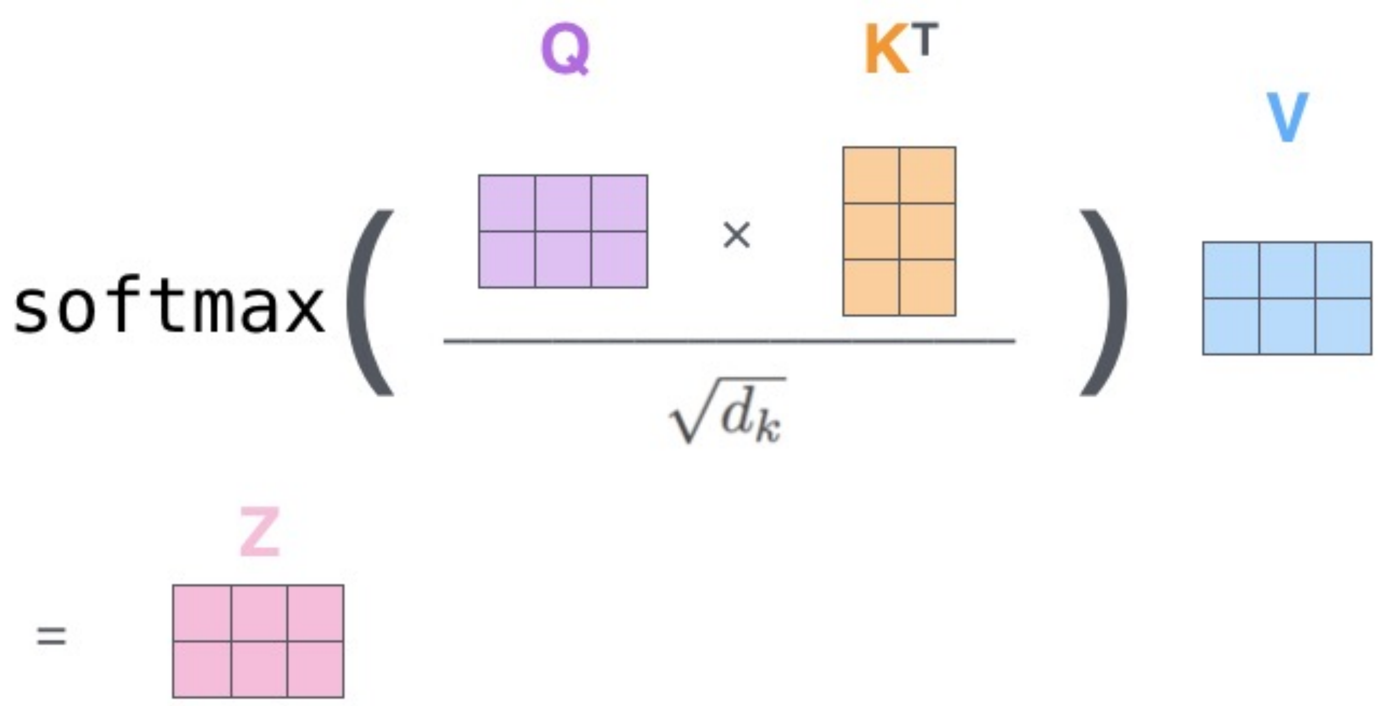
Network input = B(batch size) x N(인스턴스의 개수) x 3 x W x H

02. Method

Self-attention

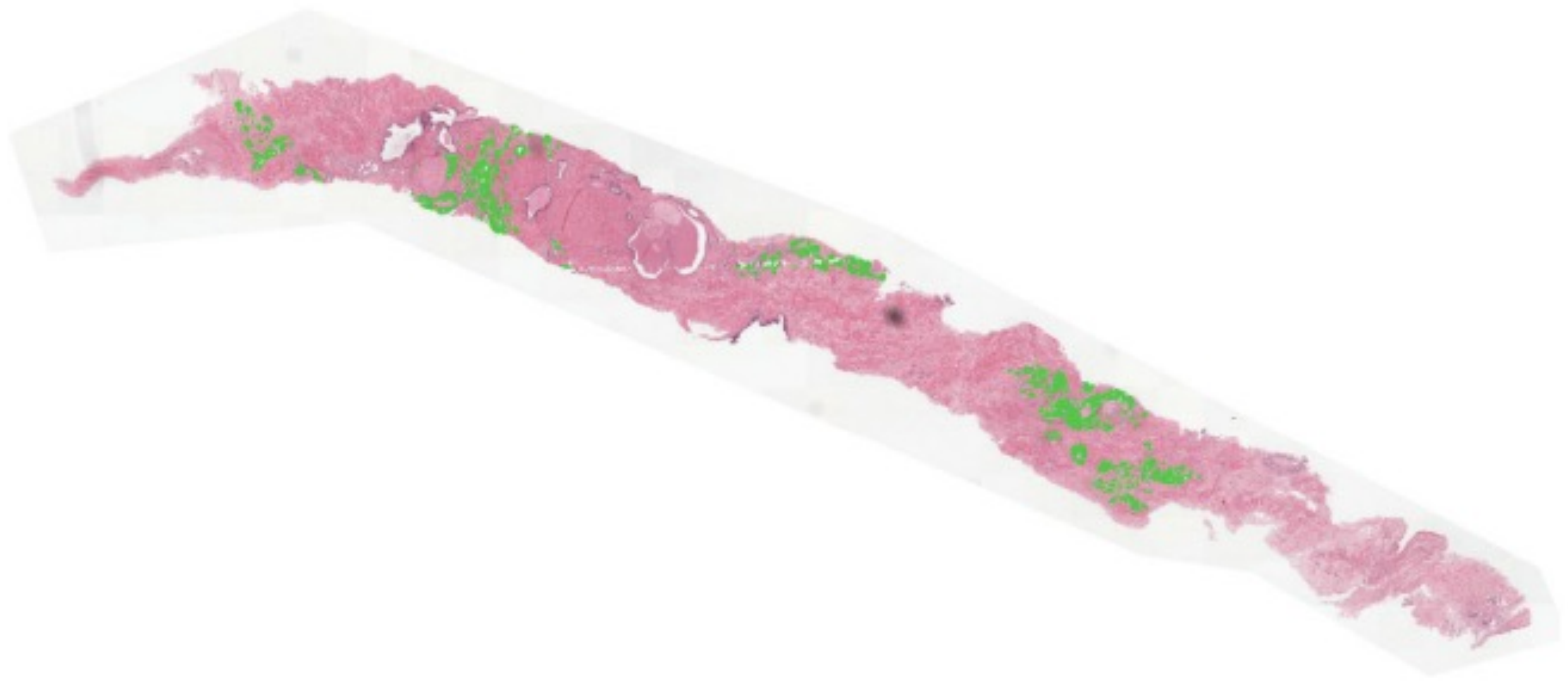


Scaled dot product attention

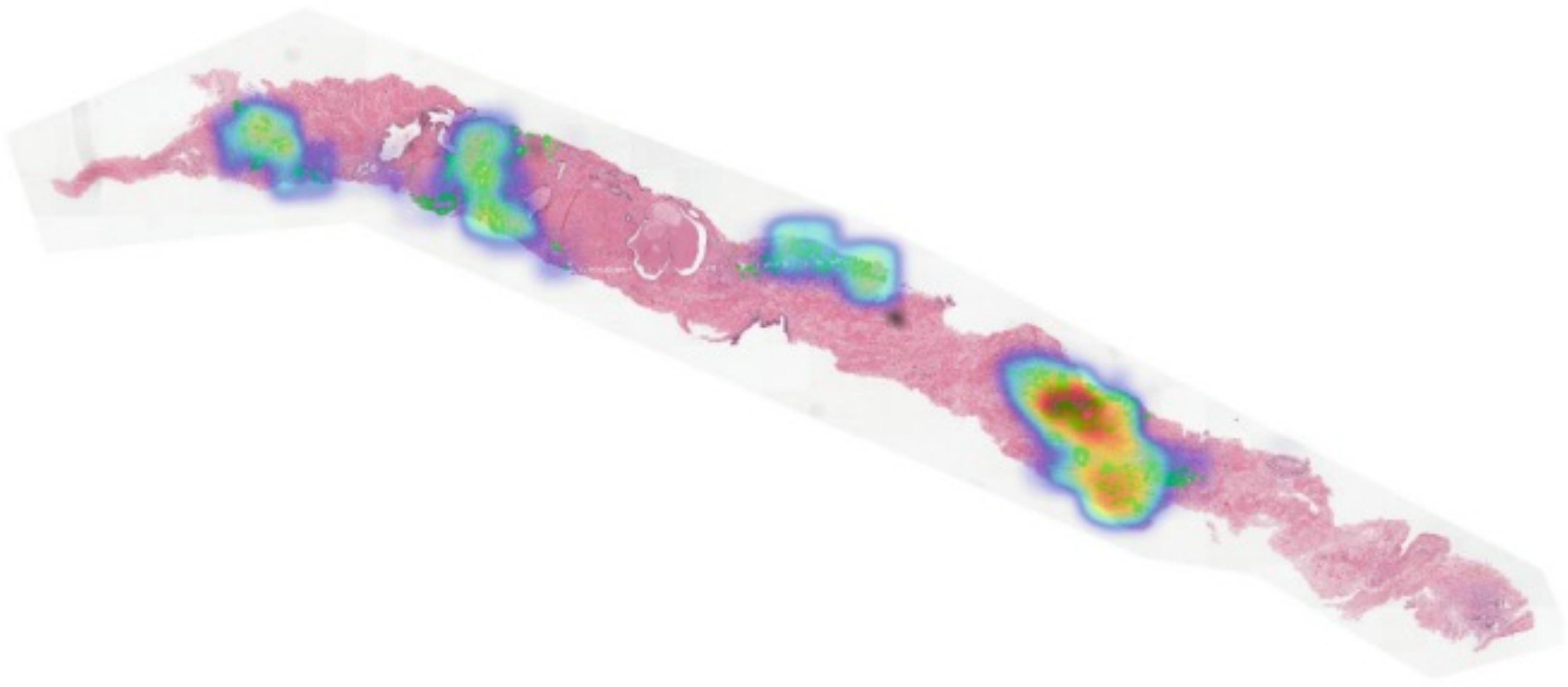


02. Method

Instance Level Semi-supervision and Pseudo-labeling



(a)



(b)

(a): 실제 암 영역을 녹색 부분으로 마스킹한 overlay

(b) : 예측된 pseudo-labels을 통한 시각화 overlay

02. Method

Sliding-window approach

02. Method

어떠한 문제인가?

총 손실 = bag의 손실 + 패치 별 손실

$$L = L_{bag} + \lambda \sum_{k} L_{patch}$$

Bag-level의 예측 출력

$$\mathbf{c} = sigm(\mathbf{Wz}) = sigm(\mathbf{WHa})$$

출력 클래스 수에 대한 Linear projection

$$\mathbf{c}_k = sigm(\mathbf{Wh}_k)$$

Linear combination of Patch embedding

$$\mathbf{z} = \sum_{k=1}^{K} a_k \mathbf{h}_k = \mathbf{Ha}$$

02. Method

Pseudo-labels assignmnet

Pseudocode 1: Pseudo-labels assignment

```
Train N MIL models (N=5)
for all patches in the bag do
    Run inference on patches for each image                       각 이미지에 대한 패치 수준의 추론 실행
    Ensemble predictions of attention weights a and instance classes c_k   Attention weight a와 instance classes Ck의 앙상블 예측
    if bag label is not zero then
        for patches with top 10% of highest attention a weights, assign the
        ensembled labels as pseudo-label      Attention weight a가 상위 10%에 속할 경우 앙상블 레이블을 pseudo-label로 할당
        for patches with top 10% of lowest attention a weights, assign the zero
        labels as the pseudo-label      Attention weight a가 하위 10%에 속할 경우 0으로 pseudo-label로 할당
        otherwise flag the patch as unknown pseudo-label
    else      나머지 80%의 Patch에 대해서는 알 수 없음으로 flag를 지정
        assign zero pseudo-labels for all patches, since here we know that all
        patches must have zero labels
    end      모든 패치 레이블이 0이여야 하기 때문에 모든 패치에 대한 라벨을 0으로 할당
end
```

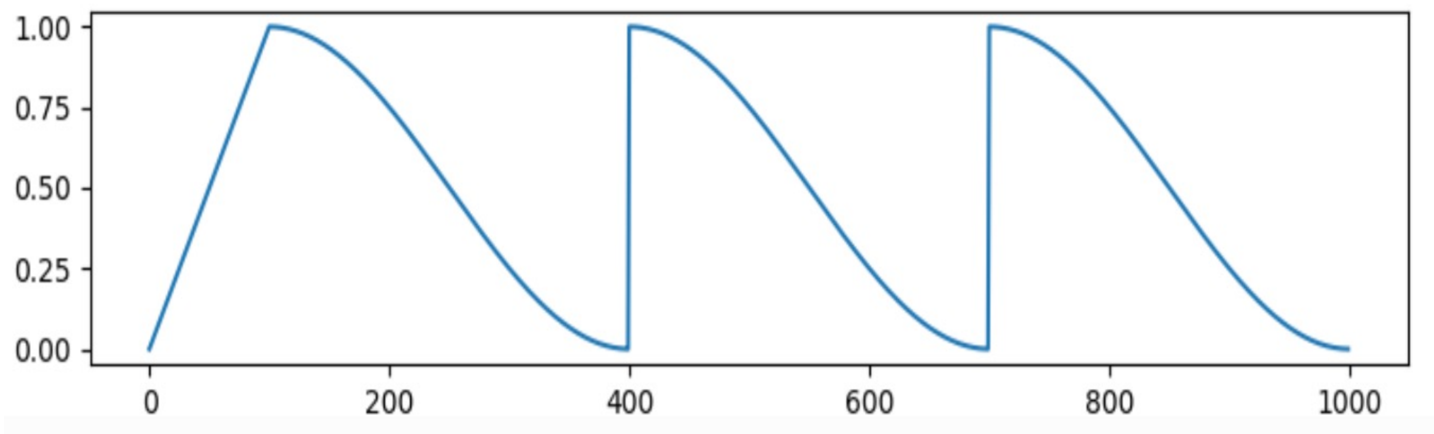
KYONGGI UNIVERSITY

03. Experiments

Experiment in progress

Library	Pytorch
GPU	NVIDIA Tesla V100 16GB
Backbone CNN	ResNet50
Pre-trained	ImageNet
Optimizer	Adam, initial learning rate a0 = 3e-4(0.0003), CNN, transformer parameter = 3e-5(0.00003)
Epochs	50 (Cosine learning rate scheduler)
Weight decay	0.1(transformer layers), No dropout
Tune the parameters	5-fold cross validations

Cosine learning rate scheduler

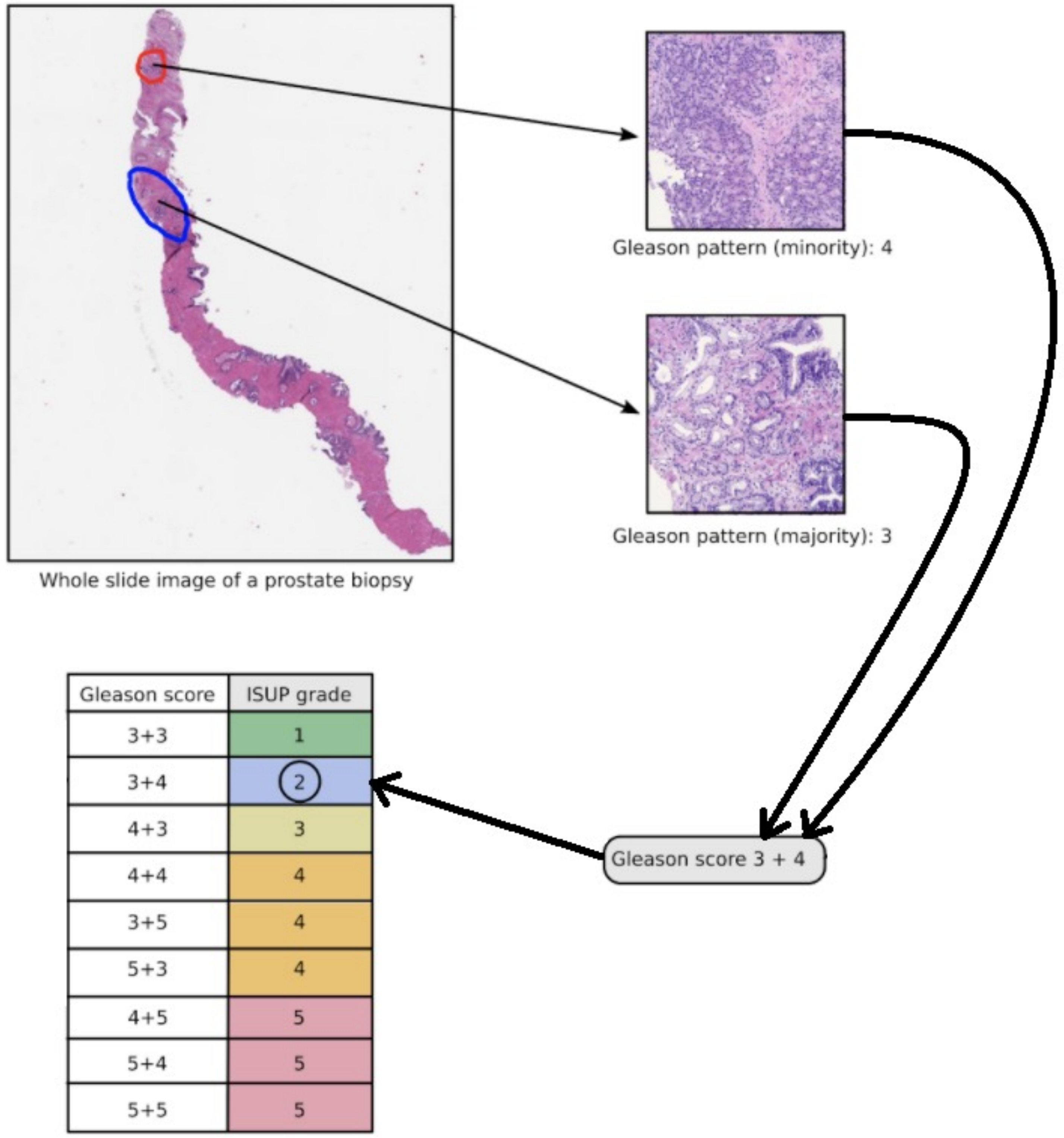


KYONGGI UNIVERSITY

03. Experiments

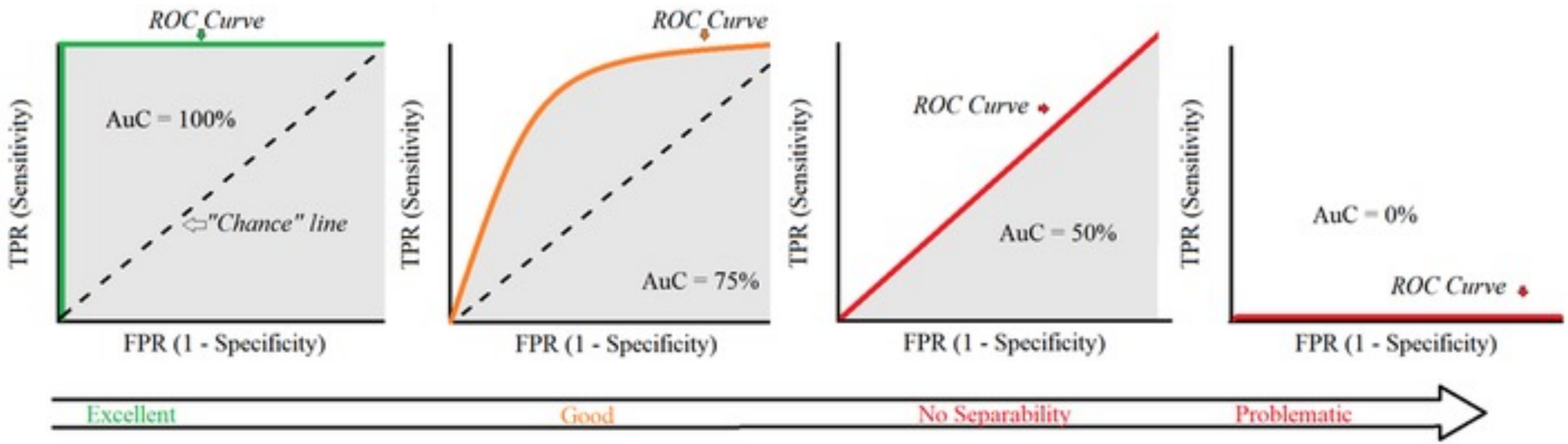
PANDA Dataset, Patch Selection

PANDA Dataset	
Name	Cancer grade assessment (PANDA) Challenge dataset
amount	11K whole-slide images
Size	25,000px x 25,000px RGB
Used Image size	4x smaller than the highest
Patch Selection	
Size	224px x 224px gird random offsets to ensure randomness.
	only keep the foreground patch.
Data input size	16 x K x 3 x 224 x 224 (batch size = 16) K = 56, random subset, GPU limit



04. Conclusion

Area under the Curve (AOC)



04. Conclusion

Quadratic Weighted Kappa (QWK)

$$K = \frac{P_A - P_C}{1 - P_C}$$

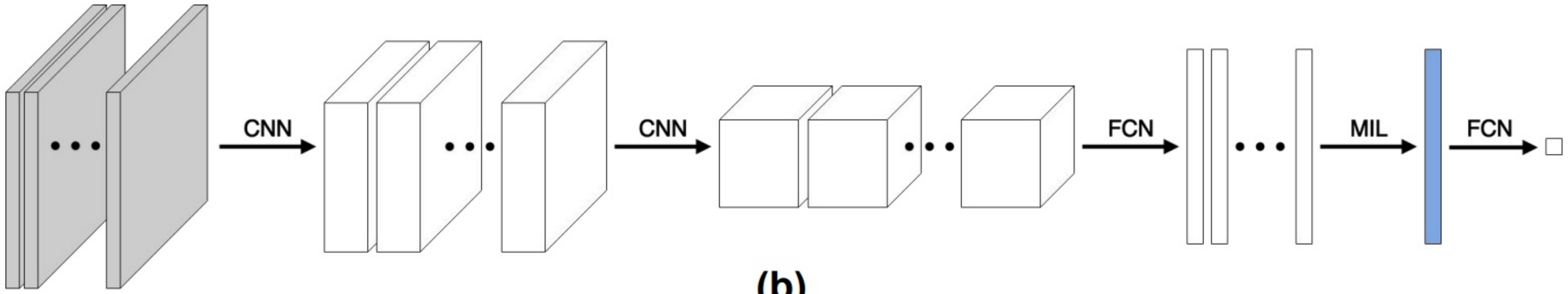
Pa : 2명의 평가자간 일치 확률
Pc : 우연히 두 평가자에 의하여 일치된 평가를 받을 비율
K : 0과 1 사이의 값을 가짐

점수(K)에 따른 일치도	
> 0.000	거의 없는 일치도
0.000 ~ 0.200	약간의 일치도
0.201 ~ 0.4	어느정도 일치도
0.401 ~ 0.600	적당한 일치도
0.601 ~ 0.800	상당한 일치도
0.801 ~ 1.000	완벽한 일치도

04. Conclusion

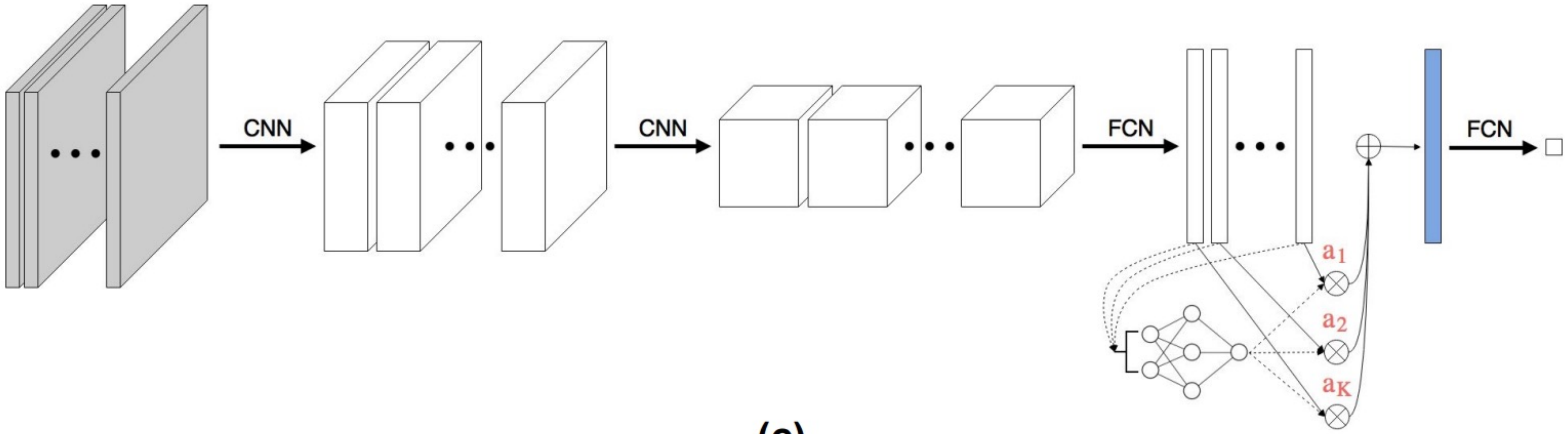
Table 1. Evaluations results on PANDA dataset

Attention MIL :



(b)

Gated attention MIL :



(c)

04. Conclusion

Table 1. Evaluations results on PANDA dataset (Patch-Wise Pseudo-labels)

Max MIL : 최대값을 활용한 다중 인스턴스 학습

	Accuracy	AUC	QWK	Leaderboard
Attention MIL [14]	0.793 ± 0.035	0.983 ± 0.021	0.948 ± 0.036	0.915 ± 0.086
Gated attention MIL [14]	0.795 ± 0.037	0.981 ± 0.011	0.936 ± 0.042	0.914 ± 0.069
Max MIL [3]	0.770 ± 0.055	0.973 ± 0.048	0.910 ± 0.053	0.868 ± 0.091
Transformer MIL	0.801 ± 0.014	0.988 ± 0.015	0.960 ± 0.034	0.930 ± 0.012
Pyramid Transformer MIL	$\mathbf{0.805} \pm 0.011$	$\mathbf{0.989} \pm 0.018$	$\mathbf{0.961} \pm 0.032$	$\mathbf{0.932} \pm 0.015$

04. Conclusion

Table 2. MIL adding pseudo-labels

	QWK (val)	QWK (Leaderboard)
Attention MIL [14] + Pseudo-labels	0.9502 ± 0.0319	0.9304 ± 0.0542
Transformer MIL + Pseudo-labels	0.9614 ± 0.0367	0.9347 ± 0.0353
Pyramid Transformer MIL + Pseudo-labels	$\mathbf{0.9652} \pm 0.0168$	$\mathbf{0.9365} \pm 0.0513$
First place - Panda kaggle challenge [2]	–	0.94085
Second place - Panda kaggle challenge [2]	–	0.93768
Third place - Panda kaggle challenge [2]	–	0.93480
Pyramid Transformer MIL (ours, ensemble of 10)	–	**0.94136**

[a] https://www.kaggle.com/c/prostate-cancer-grade-assessment/leaderboard

#	△pub	Team Name	Notebook	Team Members	Score	Entries	Last
1	▲ 21	PND			0.94085	105	1y
2	▲ 3	Save The Prostate	</> (ENS_XIE_2FOL...		0.93768	263	1y
3	▲ 188	Mikhail Druzhinin			0.93480	14	1y

04. Conclusion

- Instances의 의존성을 파악하기 위한 Transformer module
- Pseudo-labels을 사용한 instance-level의 supervision loss 제안

→ 새로운 Deep-Learning 기반 MIL 학습 접근 법을 제안

기존의 SOTA를 달성한 방법은 수백 개의 이미지만으로 성능을 파악하였지만,
본 연구에서는 11000개 이상의 이미지를 포함하는 PANDA Challenge의 전립선 전체 슬라이드 이미지 데이터 세트에 대한 평가를 진행하였습니다.

모델의 성능의 평가를 위해서 Kaggle의 test dataset을 활용하였으며, 앙상블을 진행하지 않은 단일 모델의 성능이 순위권에 속한 모델의 성능과 동등하다는 결과를 냄을 알 수 있었습니다.

Visual transformers는 Classification convolutional neural network을 완전히 대체할 수 있음을 보임으로서 Transformer blocks만을 기반으로 MIL 학습 모델을 만들 수 있는 가능성을 보여주었습니다.

Question

감사합니다!