Multiple Instance Learning with Spatial Attention for ROP Case Classification, Instance Selection and Abnormality Localization

Xirong Li[*], Wencui Wan[†], Yang Zhou[‡], Jianchun Zhao[‡], Qijie Wei[‡], Junbo Rong[†], Pengyi Zhou[†]
Limin Xu[†], Lijuan Lang[†], Yuying Liu[†], Chengzhi Niu[†], Dayong Ding[‡], and Xuemin Jin[†]

[*]Renmin University of China

[†]The First Affiliated Hospital of Zhengzhou University, China

[‡]Visionary Intelligence Ltd, Beijing, China

ICPR 2020

Jan. 12, 2021

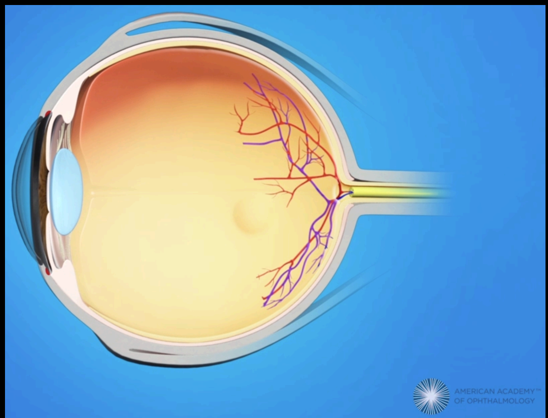
Retinopathy of Prematurity (ROP)

Preterm infants may get high oxygen exposure

- Disorganized growth of retinal vessels

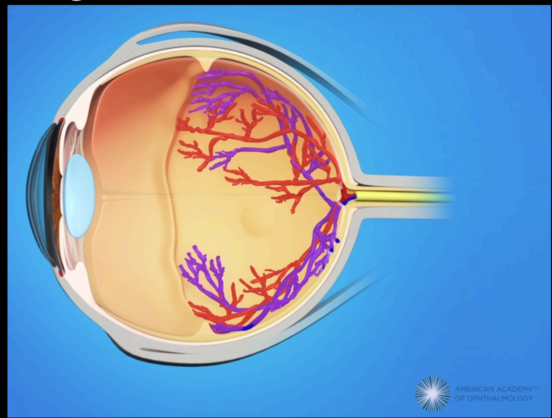
Progression process of ROP

Vessels grow in peripheral retina



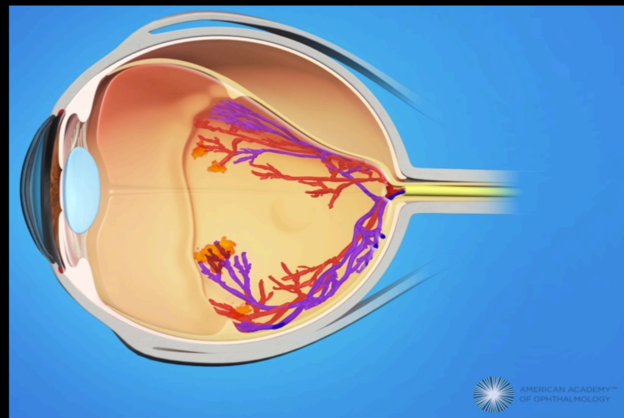
…

Ridge is formed



…

Retina is detached



ROP is a leading cause of visual loss in childhood

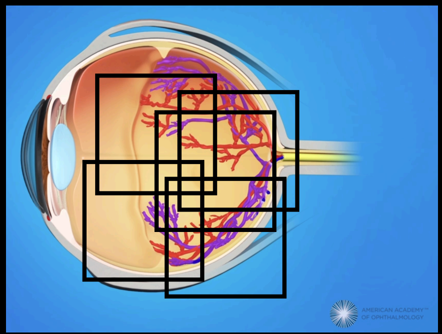
Source: American Academy of Ophthalmology. https://www.youtube.com/watch?v=OyaUpwSYe0w

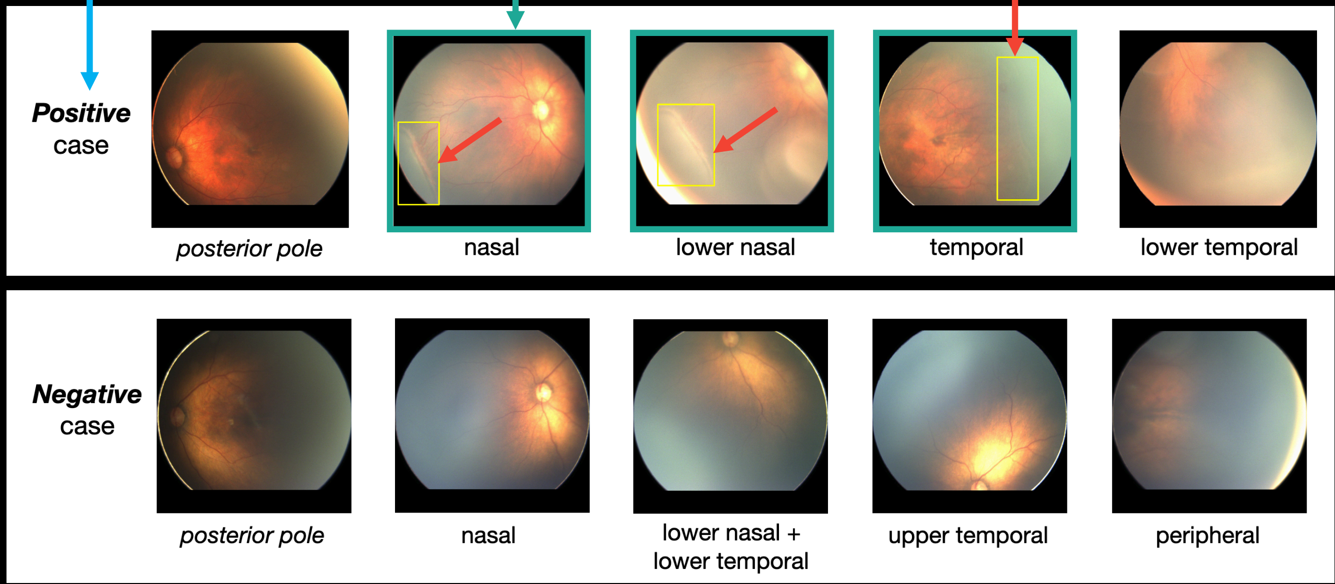
ROP Diagnosis and Three Tasks

Requires multiple color fundus images per eye to cover distinct zones of the retina

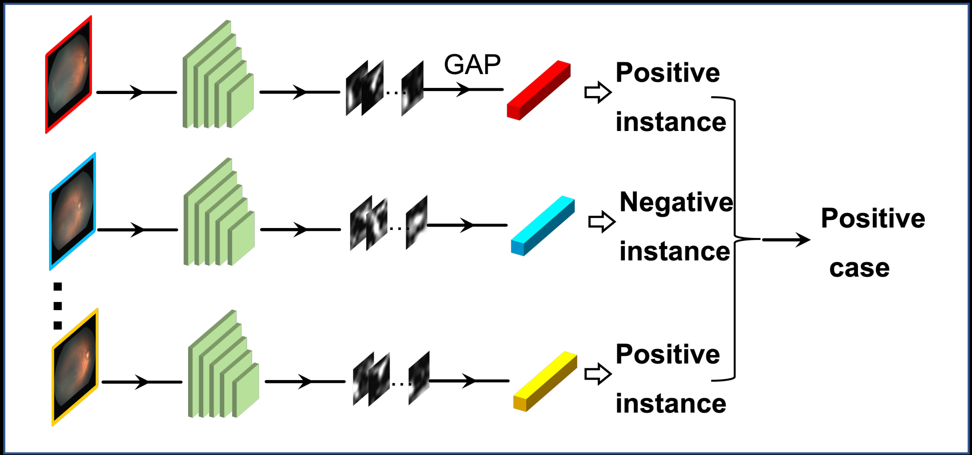
- Different from retinal disease recognition for adults where a single image is used
- Three questions to answer: whether ROP-positive / which instances / which part?

1. Case classification 2. Instance selection 3. Abnormality localization





State-of-the-art



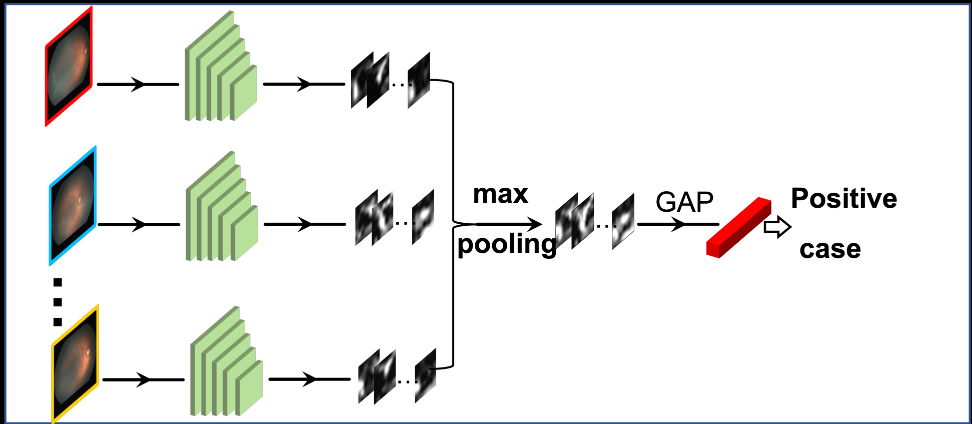
Instance-level classification[Zhang *et al*. OMIA 2019]

Pro:

- Instance selection
- Case classification

Con:

- Instance-level annotations are costly
- Abnormality localization not covered



Multiple instance learning [Hu *et al*. TMI 2018]

Pro:

- Need only case-level annotations

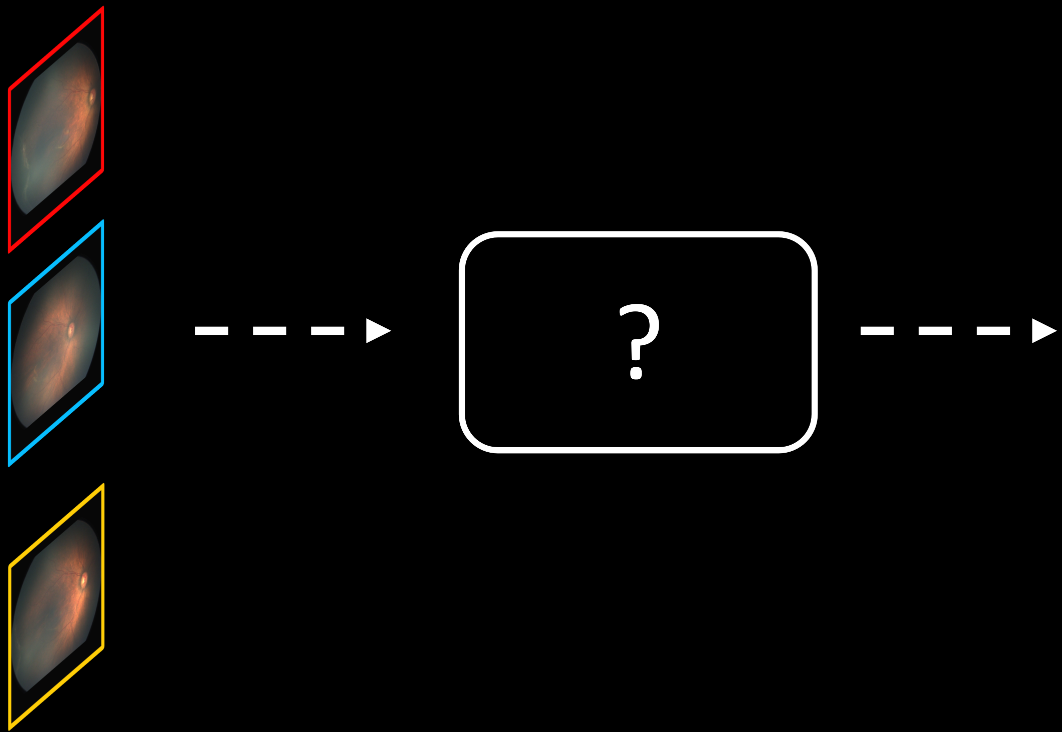
Con:

- Instance selection not covered
- Abnormality localization not covered

Challenge We Want to Tackle

How to solve the three tasks related to ROP diagnosis in a unified framework?

- Given only case-level labels



1. Case classification
2. Instance selection
3. Abnormality localization

Deep Multiple Instance Learning with Spatial Attention

MIL-SA: One network for all

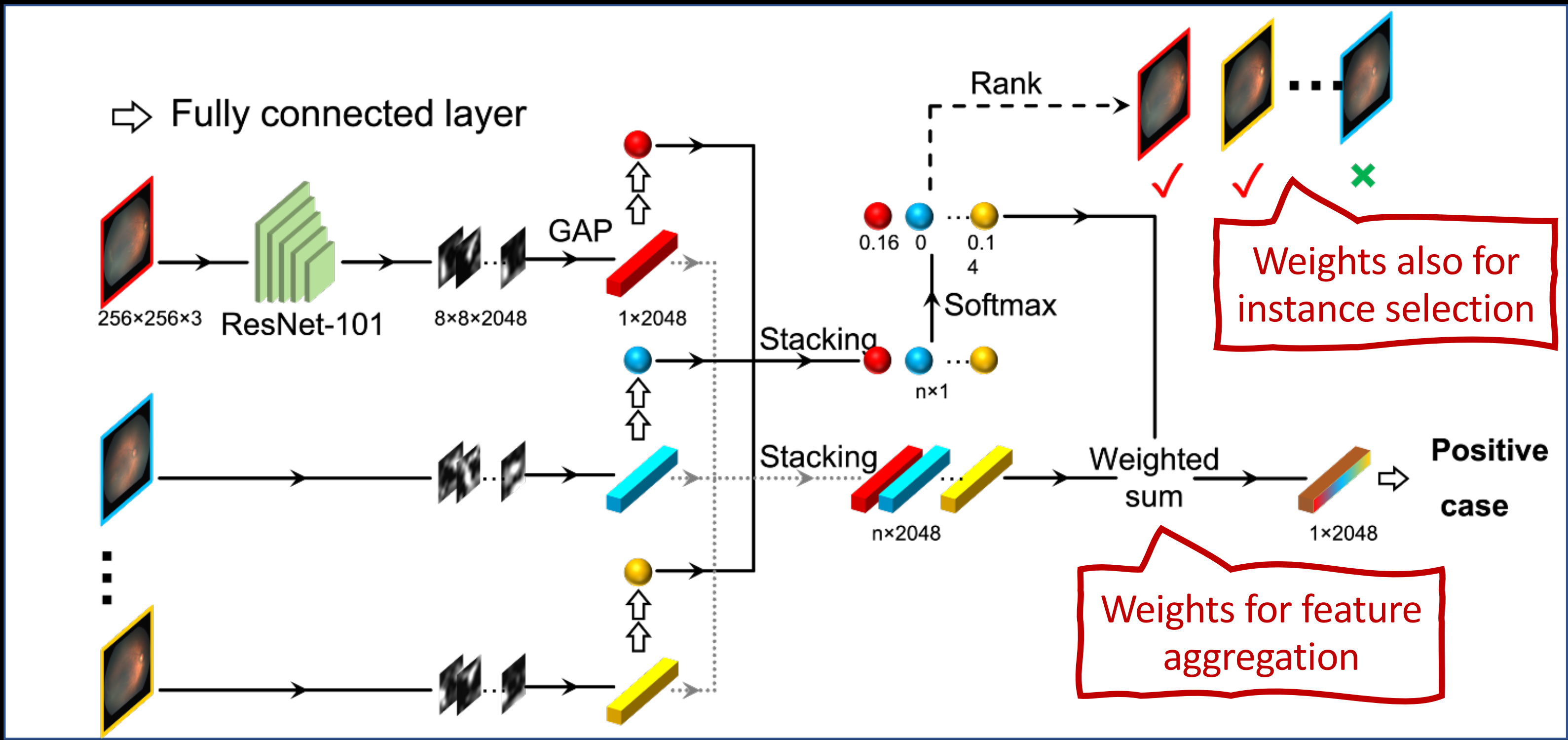
MIL-SA

- Deep MIL with instance attention
  - 1. Case classification
  - 2. Instance selection
- Pixel-level spatial attention → 3. Abnormality localization

MIL-SA: Part I

Deep MIL with instance attention

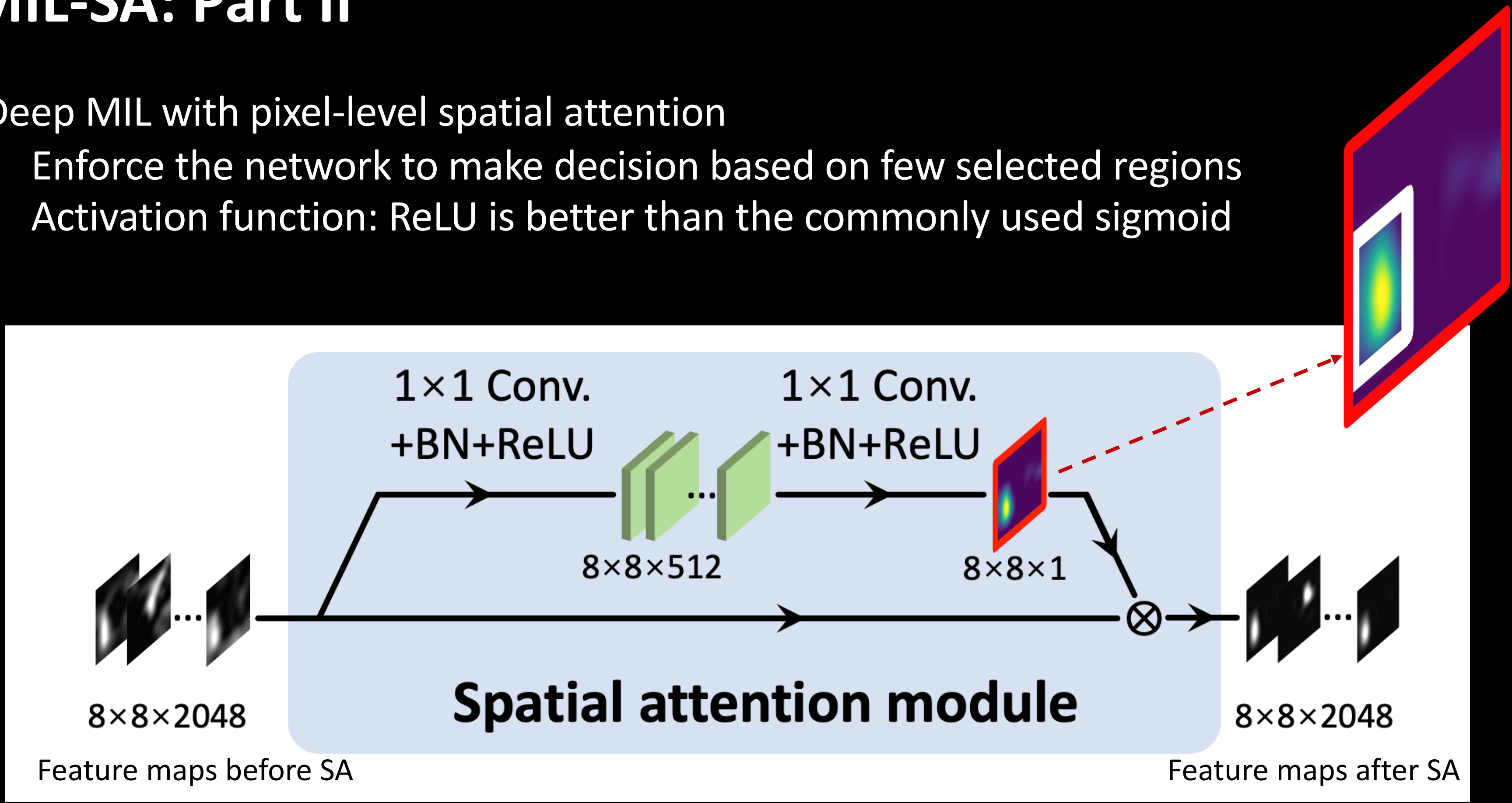
- A self-attention module to produce instance-level weights[Ilse *et al.* ICML 2018]



MIL-SA: Part II

Deep MIL with pixel-level spatial attention

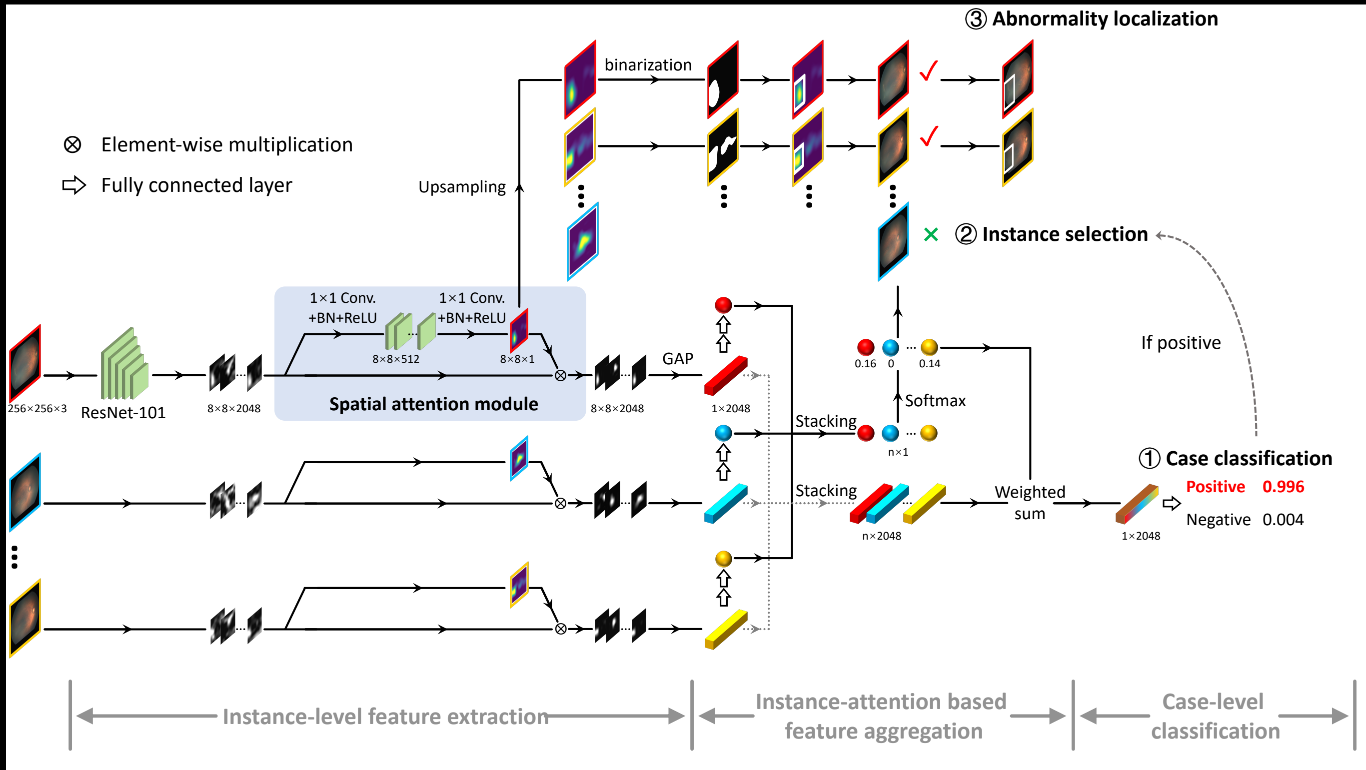
- Enforce the network to make decision based on few selected regions
- Activation function: ReLU is better than the commonly used sigmoid



MIL-SA: Part I + Part II

One network for all

- End-to-end
- Learn only from case-level annotations



Evaluation

Data split	No. of cases		No. of instances		No. of instances with specific lesions			
	Positives	*Negatives*	*Positives*	*Negatives*	*DL*	*ridge*	*EFP*	*VDT*
training	859	1,014	2,599	6,995	748	980	458	1,086
validation	51	112	251	488	31	75	124	135
test	41	109	191	529	25	73	73	90

DL: Demarcation Line
EFP: Extraretinal Fibrovascular Proliferation
VDT: Vascular Dilatation and Tortuosity

Data

- Expert-labeled 2,186 cases and 11,053 image instances
- Instance-level and region-level annotations are merely used for evaluation

Baselines

- MIL-max: MIL with max pooling
- MIL-att: MIL with instance attention

	Case classification	Instance selection	Abnormality localization
MIL-max	✓	✕ (fixed with Grad-CAM)	✕ (fixed with Grad-CAM)
MIL-att	✓	✓	✕ (fixed with Grad-CAM)

Experiment 1. ROP Case Classification

Model	Sensitivity	Specificity	F1	Accuracy	AUC
MIL-max [6]	**0.9512**	0.9083	0.9292	0.9200	0.9758
MIL-att [20]	**0.9512**	0.9358	0.9434	0.9400	0.9655
proposed *MIL-SA*	0.9268	**0.9725**	**0.9491**	**0.9600**	**0.9895**

The proposed MIL-SA outperforms the baseline in terms of F1, Accuracy and AUC.

Experiment 2. Instance Selection

Metric: Rank-based Average Precision (AP)

	Overall	Range of the positive rate per case			
		$(0, 0.3)$	$[0.3, 0.5)$	$[0.5, 0.7)$	$[0.7, 1]$
Case number (percentage)	41 (100%)	1 (2.4%)	4 (9.8%)	8 (19.5%)	28 (68.3%)
Models:					
MIL-max + Grad-CAM	0.8991	0.5833	0.8833	0.8534	0.9257
MIL-att	0.9694	**1.0**	0.9375	0.9172	0.9877
proposed *MIL-SA*	**0.9811**	**1.0**	**1.0**	**0.9302**	**0.9922**

MIL-SA shows advantages especially for cases with lower positive rates.

Experiment 3. Abnormality Localization

Metric: Intersection over Union (IoU)

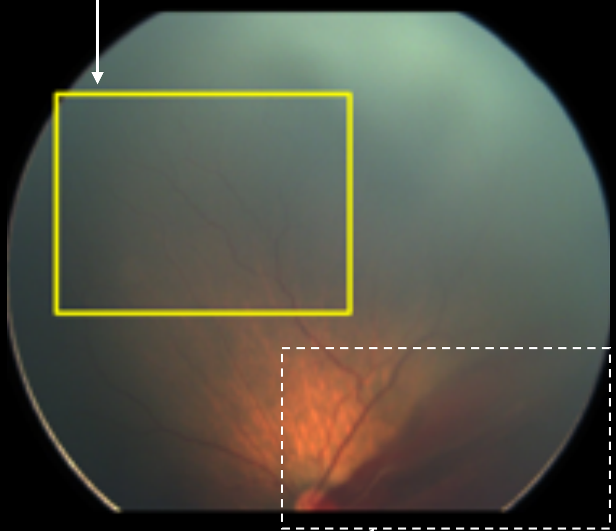
Model	Overall	DL	ridge	EFP	VDT
MIL-max + Grad-CAM	0.2594	0.2489	0.2942	0.3562	0.2258
MIL-att + Grad-CAM	0.2956	0.2853	0.3573	0.3879	0.2046
proposed *MIL-SA*	**0.3615**	**0.3814**	**0.4023**	**0.4621**	**0.2374**

DL: *demarcation line*
EFP: *extraretinal fibrovascular proliferation*
VDT: *vascular dilatation and tortuosity*

MIL-SA localizes abnormal regions more accurately.

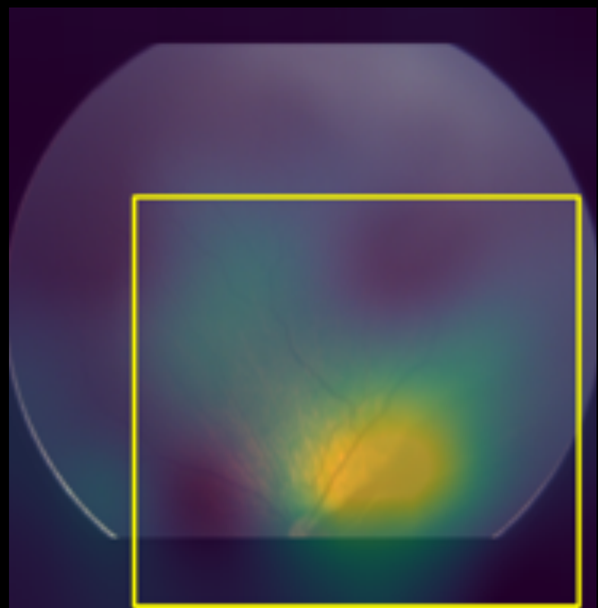
Abnormality Localization: A Close-up View

VDT (Vascular Dilatation and Tortuosity)

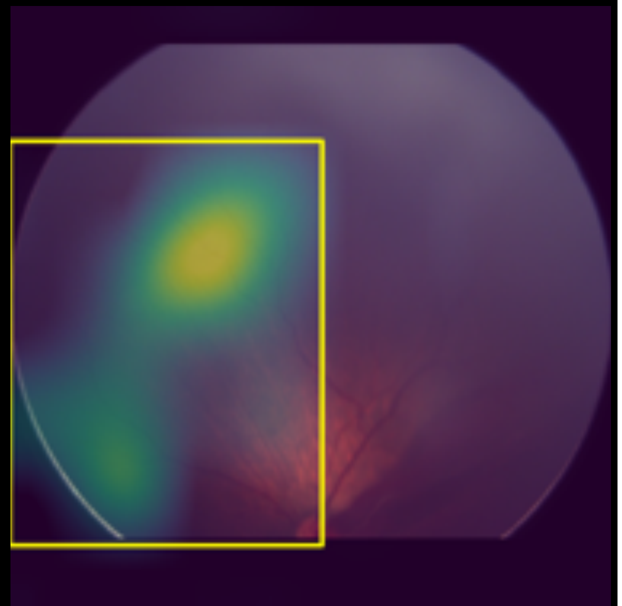


Blot hemorrhage
(irrelevant w.r.t. ROP)

MIL-Max + Grad-CAM



MIL-SA



Take-Home Messages

MIL-SA as the first deep network to achieve three tasks for ROP diagnosis, given only case-level annotations

Promising performance for automated ROP diagnosis

- Case classification -> AUC of 0.9895
- Instance selection -> AP of 0.9811
- Abnormality localization -> IoU of 0.3615 (much room for future improvement)

contact: xirong@ruc.edu.cn