Mining Gaze for Contrastive Learning toward Computer-Assisted Diagnosis

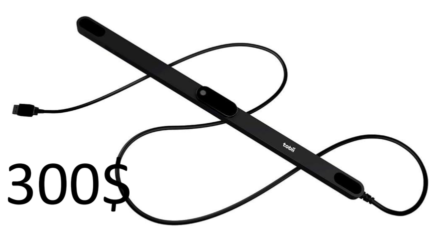
Zihao Zhao [1]*, Sheng Wang [1, 2, 3]*, Qian Wang [1, 4], and Dinggang Shen [1, 3, 4], ✉

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
[2] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[3] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China
[4] Shanghai Clinical Research and Trial Center, Shanghai, China
dgshen@shanghaitech.edu.cn

AAAI-24 / IAAI-24 / EAAI-24

EYE-TRACKING

Eye-tracker inside



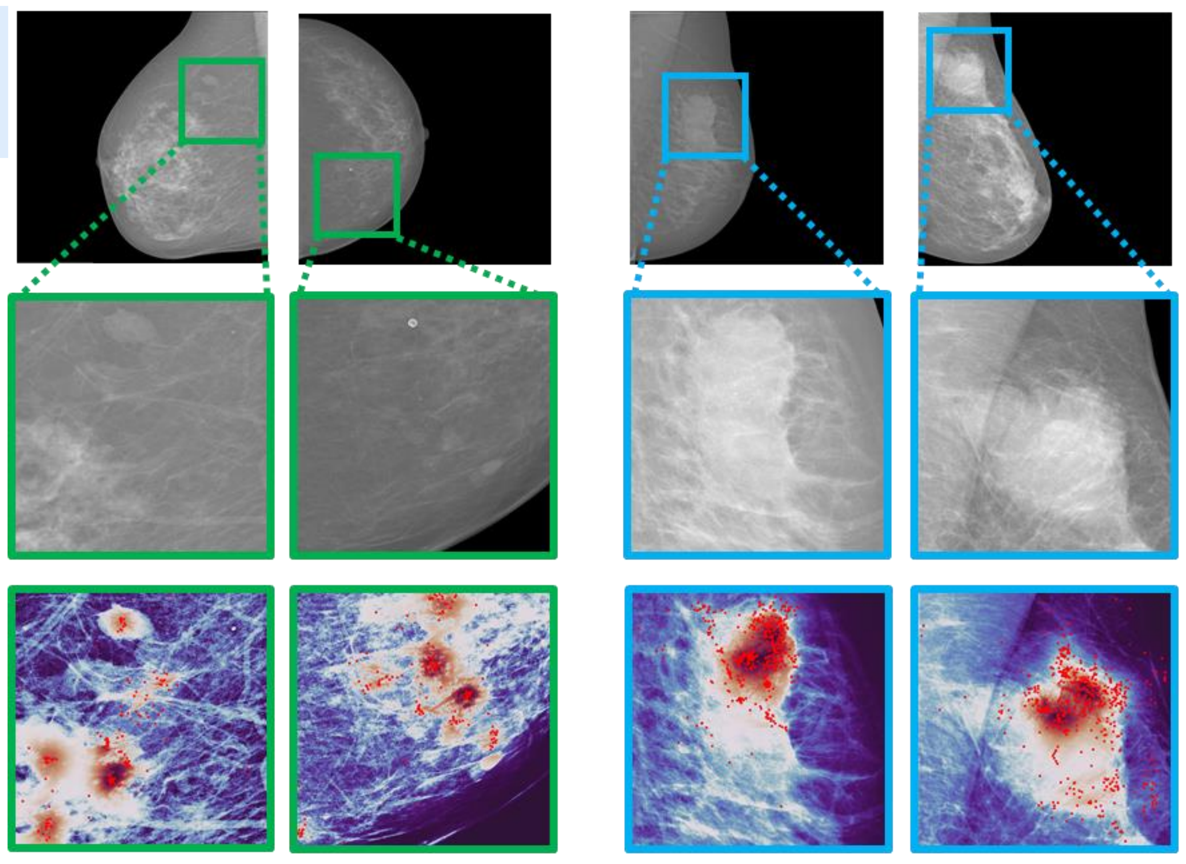
300$

Regular screen-based

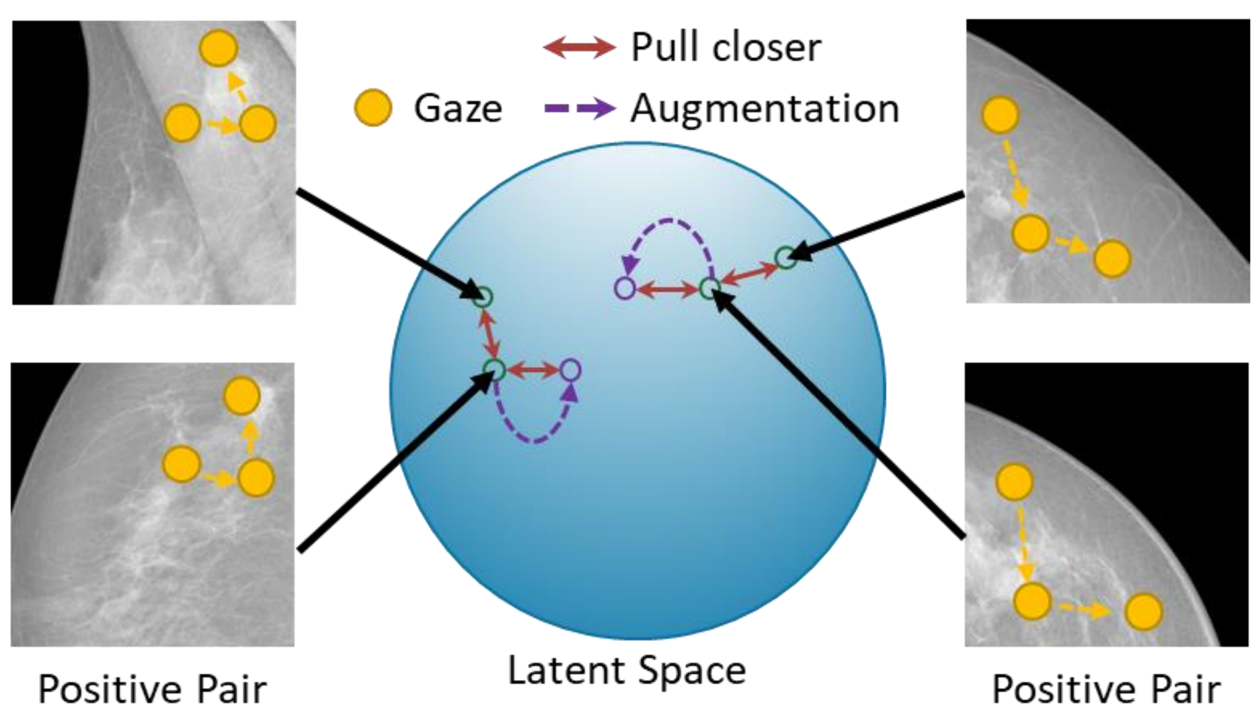
- **Eye-tracking (gaze)**, has involved to be a **low-cost** and **popular** tool to understand and interpret reasoning and clinical decision in 2024.
- Apple Vision Pro, Meta's latest Quest headset and Sony PSVR2, all have eye-trackers inside.

INTUITIVE FINDINGS

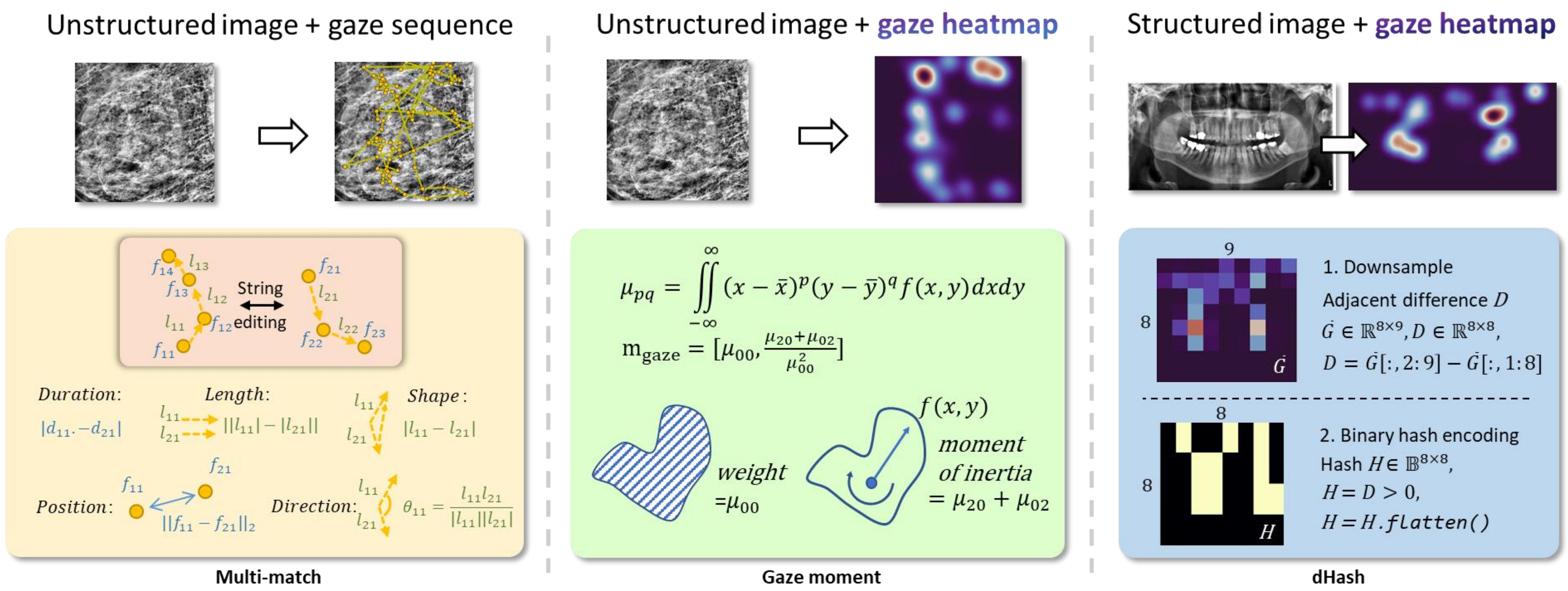
- When read by a radiologist, **semantically similar regions** usually demonstrate **similar distributions of gaze points**.
- Benign lesion often attracts a scattered gaze distribution, whereas malignant parts have a more centralized distribution



METHOD



Philosophy: Images with similar radiologist gaze patterns are considered as positive pairs and be pulled closer in the latent space.

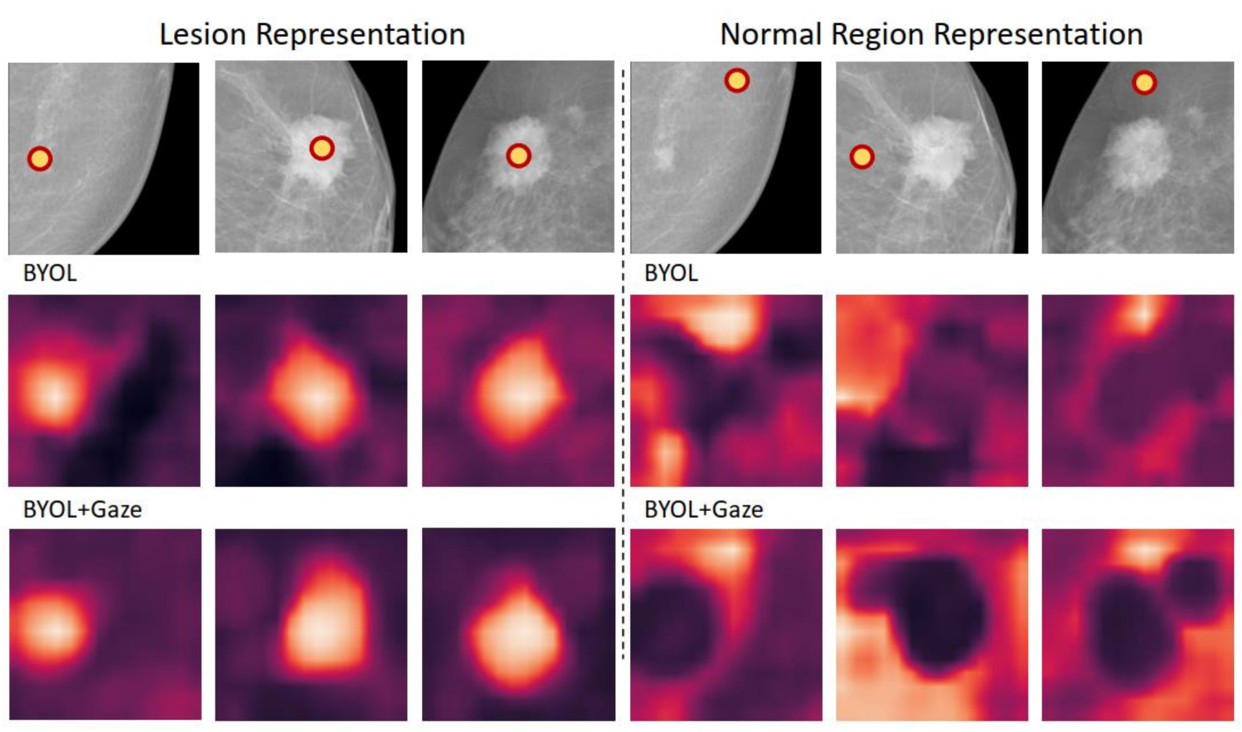


Gaze Similarity Evaluation: To align with radiologists' varied reading patterns for different medical images, we design three schemes for gaze similarity evaluation, distinguishing unstructured and structured images.

RESULTS

Method	RN18			RN50			RN101		
	M-AUC	AUC	ACC	M-AUC	AUC	ACC	M-AUC	AUC	ACC
From-scratch	71.39±3.26	73.98±4.55	76.76±4.79	68.01±2.41	71.44±4.38	75.41±4.22	69.89±2.56	67.27±3.71	73.78±2.02
ImageNet	83.43±1.98	82.38±2.84	80.38±2.34	89.73±0.89	86.17±1.23	82.97±1.08	88.90±2.98	87.50±0.89	85.63±2.26
MoCo	82.19±3.05	84.69±2.53	82.43±1.71	89.52±2.15	89.44±0.90	81.62±1.38	92.28±2.86	91.03±1.88	86.22±1.58
MoCo+McGIP	**85.07±2.43**	**88.37±1.73**	**83.51±1.01**	**92.74±1.87**	**91.44±2.08**	**85.68±1.38**	**93.06±1.73**	**92.58±2.92**	**87.03±0.66**
BYOL	90.42±2.31	90.59±1.48	83.78±0.85	93.84±1.72	87.96±1.71	85.95±1.83	93.82±3.44	**90.39±2.08**	86.49±0.89
BYOL+McGIP	**95.83±0.63**	**94.96±1.13**	**85.14±0.85**	**97.07±0.75**	**93.80±0.79**	**87.57±1.01**	**95.46±2.67**	90.09±3.08	**86.76±0.54**
SimSiam	91.10±3.26	91.81±1.63	83.51±1.01	93.11±2.26	86.56±2.99	86.27±1.79	92.26±1.11	**90.26±1.14**	85.68±1.08
SimSiam+McGIP	**95.30±1.16**	**94.62±1.34**	**85.95±1.38**	**95.30±0.78**	**89.22±0.95**	**88.65±1.38**	**96.85±0.63**	90.08±1.51	**87.84±1.01**

Consistent Improvements: McGIP acts as a plug-and-play technique for different frameworks and backbones.

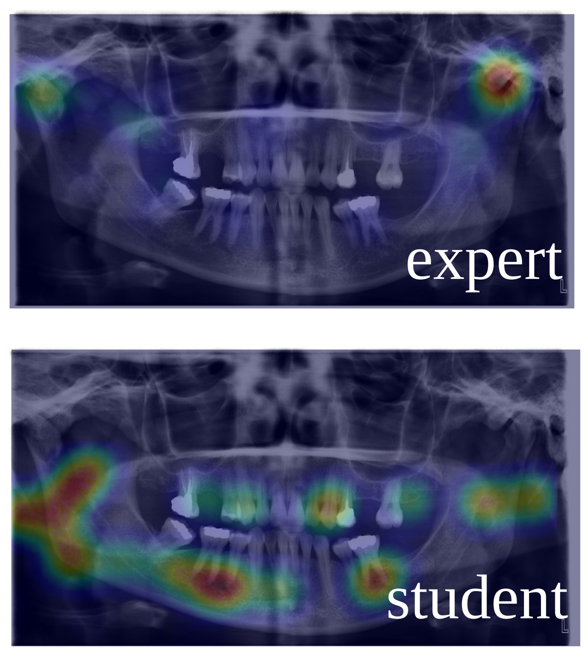


Better Spatial Sensing: A point is selected and compared with all other points. Brighter color denotes more similar points.

Method		INbreast			Tufts	
		M-AUC	AUC	ACC	AUC	ACC
RN18	GT	93.01±1.78	90.86±1.32	84.32±0.66	60.53	59.00
	Gaze	**95.83±0.63**	**94.96±1.13**	**85.14±0.85**	**62.91**	**65.00**
RN50	GT	96.02±0.88	88.96±1.61	85.14±0.85	59.29	58.00
	Gaze	**97.07±0.75**	**93.80±0.79**	**87.57±1.01**	**61.35**	**67.50**
RN101	GT	94.81±2.48	**90.60±2.97**	85.95±1.08	59.17	63.50
	Gaze	**95.46±2.67**	90.09±3.08	**86.76±0.54**	**61.14**	**64.50**

Better than Supervised Contrastive Learning: Comparison between gaze and ground-truth (GT) one-hot labels, revealing the superiority of gaze data (default contrastive learning method: BYOL)

LIMITATIONS



Expertise demanding

CONCLUSIONS

- Gaze data can enhance the effectiveness of contrastive learning methods in a model-agnostic manner.
- Compared to ground truth labels, gaze data consistently outperforms across various datasets, which illustrates its effectiveness in uncovering visual semantics.