Query-Level Alignment for End-to-End Lesion Detection with Human Gaze

Yan Kong[3,1†], Zhixiang Peng[1†], Yuan Yin[3], Yonghao Li[3], Jiangdong Cai[3], Sheng Wang[3,4], Qian Wang[3,5(✉)], Yuqi Fang[1,2(✉)], and Caifeng Shan[1,2(✉)]

[1] School of Intelligence Science and Technology, Nanjing University, Nanjing, China
[2] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
{cfshan,yqfang}@nju.edu.cn
[3] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
[4] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[5] Shanghai Clinical Research and Trial Center, Shanghai, China
qianwang@shanghaitech.edu.cn

Abstract. Lesion detection for medical image is crucial in computer-aided diagnostic systems, enabling early disease identification and enhancing clinical decision-making. Existing lesion detection models primarily rely on bounding boxes for supervision, which overemphasize lesion boundaries while neglecting critical internal features, potentially resulting in misdetections. In contrast, clinicians' gaze, which reflects the visual focus during diagnosis, captures internal semantic patterns of lesions, providing a more informative supervisory signal than conventional annotations. Inspired by this insight, we propose a gaze-driven detection framework for enhancing lesion identification accuracy. Specifically, our framework introduces three key gaze-prioritized innovations: 1) an *adaptive gaze kernel* that prioritizes diagnostically significant high-magnification regions, 2) a *gaze-guided assignment module* that establishes query-level gaze-region correspondence, and 3) a *query-level consistent loss* that aligns detection model attention with clinicians' gaze patterns. By incorporating clinicians' expertise through gaze data, our method improves lesion detection accuracy and clinical interpretability. In addition, our method can be designed as a plug-and-play module, which maintains compatibility with mainstream object detectors. To validate the effectiveness of our method, we employ two public and one private datasets, and extensive experiments demonstrate its superiority over existing approaches. Furthermore, we contribute a pioneering gaze-tracking dataset with 1,669 precise gaze annotations, establishing a new benchmark for gaze-driven research in object detection. The dataset and code is available at https://github.com/YanKong0408/GAA-DETR.

Keywords: Medical lesion detection, Eye-tracking, Candida detection, Breast tumor detection

† Yan Kong and Zhixiang Peng contributed equally to this study.

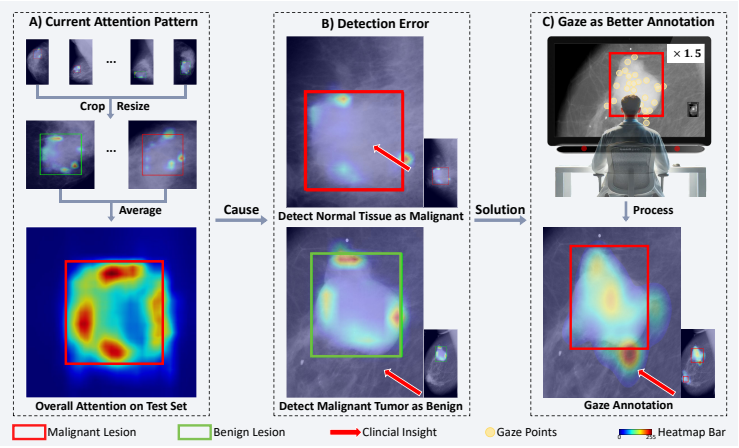

Fig. 1. Illustration of the motivation. (A) DETR's predictions and overall attention on the test set demonstrate a tendency to focus on lesion boundaries. (B) Examples of errors caused by boundary-focused attention for neglecting clinicians' expertise (morphology and peritumoral vasculature). (C) Clinicians' gaze patterns that capture diagnostic-relevant features can be regarded as better supervisory annotations.

1 Introduction

Lesion detection plays a critical role in clinical practice by enhancing diagnostic accuracy and efficiency through precise localization of anatomical structures and pathological findings [1,6]. It also serves as an integral component of automated diagnosis systems to streamline workflows and enable reliable AI-driven healthcare solutions at scale [29,22].

Despite its clinical importance, current lesion detection models rely solely on bounding box supervision, demonstrating a critical flaw [7,18]: heavily depending on intensity contrast at lesion boundaries while neglecting internal characteristics within the lesion (Fig. 1 (A)). This boundary-focused attention pattern often leads to detection failure, as exampled in Fig. 1 (B): in one case incorrectly detecting normal tissue as tumor by ignoring internal textures (top), and in another misclassifying a distinctly malignant tumor by disregarding peritumoral vasculature (bottom). The correlation between attention focusing on boundaries and poor detection performance is also evidenced by the results in Section 4. These failure cases reveal the limitation of conventional detection paradigms and highlight the urgent need for more comprehensive diagnostic annotations.

Clinical gaze patterns offer a promising solution to this limitation. Emerging research suggests that intergrating decision-related gaze patterns could significantly enhance computer-aided diagnosis systems [19], because eye-tracking data reflect the visual focus during diagnosis and capture internal semantic patterns of lesions, providing a more informative supervisory signal absent in current data-driven models [25,21,23]. For example, the focus of gaze on peritumoral

vessels and the morphological context in mammography (Fig. 1 (C)) highlights how gaze patterns can capture subtle diagnostic signatures that conventional boxes-driven detection models often overlook.

However, current gaze-integrated detection systems face dual challenges in both *architectural design* and *data availability*. From the perspective of architectural design, existing frameworks suffer from inadequate spatial specificity of gaze heatmaps [17,12], mandatory reliance on gaze inputs during inference [15], and inefficient utilization of gaze data at lesion-level [16,2]. From the aspect of available data, research is limited by scarce public gaze datasets [19], while existing works [11,3,9] lack raw temporal-spatial gaze sequences and lesion-specific regions-of-interest (ROIs), restricting the spatiotemporal analysis of gaze.

To address these limitations, we propose **G**aze-**A**ligned-**A**ttention **De**tection **Tr**ansformer (GAA-DETR) – a gaze-prioritized detection framework and introduce the first medical gaze-tracking dataset (2,450 images, 1,669 open-sourced). Our framework integrates three key innovations: *an adaptive gaze kernel* enabling magnification-aware refinement of diagnostic regions; *a gaze-guided Hungarian matching module* establishing query-attention semantic alignment; and *a query-level consistency loss* enforcing spatial correspondence between model attention and clinicians' visual focus. The dataset combines mammograms and cervical pathology images with comprehensive gaze annotations, representing the first public gaze benchmark for lesion detection research. Experiments demonstrate our method enhances detection accuracy and anatomical plausibility while maintaining compatibility with mainstream detectors through modular design.

We summarize our major contributions as follows:

- **Attention phenomenon discovery**: We observe that traditional models focus on lesion boundaries but overlook internal features, causing errors, whereas clinicians' gaze captures these critical decision-relevant features.
- **Architectural innovations**: We propose a novel detection framework that integrates clinical gaze priors via query-level attention alignment.
- **Open-source dataset**: We contribute the first large-scale gaze-tracking dataset for lesion detection with comprehensive annotations.
- **Empirical validation**: Extensive experiments demonstrate superior performance, clinical interpretability and compatibility of our method.

2 Method

In this section, we present our end-to-end GAA-DETR model through three components: a data transformation pipeline incorporating novel adaptive gaze kernel, a concise overview of the model architecture, and the loss calculation module. Notably, our framework does not require gaze during inference.

2.1 Adaptive Gaze Data Transformation

This module is to transform raw gaze data into query-level ground-truth representations that can be effectively utilized by our model. This section starts

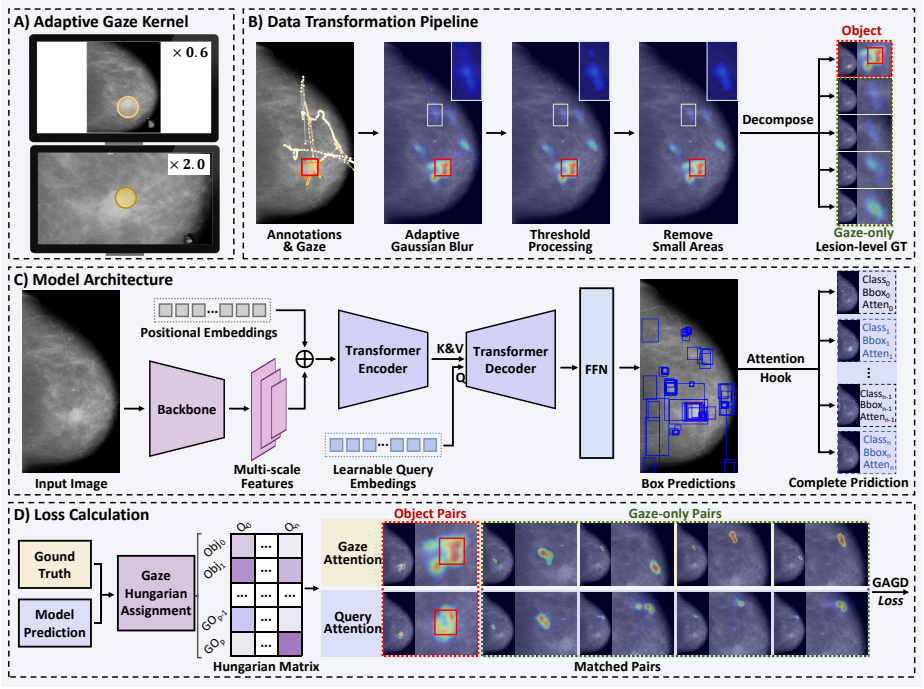

Fig. 2. Overview of the proposed method. (A) Adaptive Gaze Kernel: the focal region of gaze point can be modeled as a Gaussian kernel. (B) Data Transformation Pipeline transforms gaze data into query-level ground-truth. (C) Model Architecture generates predictions. (D) Loss Calculation aligns model attention and clinician gaze patterns.

with the innovative Adaptive Gaze Kernel algorithm for sequence-to-heatmap conversion and followed by the overall data transformation pipeline.

Adaptive Gaze Kernel: Radiologists often zoom in on high-resolution images to examine critical local regions in detail. Clinical observations reveal a correlation between magnification levels and the diagnostic significance of focal regions. Based on this insight, we propose a novel adaptive gaze kernel algorithm to leverage magnification information for improved heatmap generation. As demonstrated in Fig. 2 (A), the focal region of individual gaze point during diagnosis can be modeled as a Gaussian kernel [24]. Our implementation adapts kernel properties to magnification factors: higher factors correspond to smaller, more intense kernels ($\sigma = f/\beta$). The adaptive Gaussian blur is formalized as:

$$\text{Heatmap}(x, y) = \text{Normalize}\left(\sum_{(x_i, y_i) \in P} \frac{\lambda f}{2\pi} \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma^2}\right)\right), \quad (1)$$

where P denotes gaze point coordinates, λ controls intensity scaling, f represents the magnification factor, and β is the proportionality constant.

Data Transformation Pipeline: As illustrated in Fig. 2 (B), our pipeline first generates raw heatmap through adaptive Gaussian blurring, followed by noise reduction via thresholding and small-area removal [12]. The processed heatmap is subsequently decomposed into two clinically meaningful components: *object attention* (contiguous regions within bounding boxes exceeding 50% of total area, representing lesion-specific clinician focus) and *gaze-only attention* (remaining regions indicating significant potential clinical interest, inspired by [12]).

2.2 Model Architecture

Our model, based on the DETR [5] architecture, processes input images through a backbone to extract multi-scale features, which are added with positional embeddings and fed into a transformer encoder. The encoder outputs generate keys and values, while learnable queries undergo refinement through cross-attention mechanisms in the transformer decoder. Subsequently, a feed forward network (FFN) performs regression tasks, yielding both category and location predictions. The learnable queries can be regarded as candidate boxes' features, and their respective attention weights can be hooked during model's forward pass.

2.3 Loss Calculation

Our proposed loss calculation module starts with Gaze Hungarian Assignment to establish pairings between ground-truth annotations and predicted boxes. This approach enhances the semantic understanding of queries via gaze supervision, enabling efficient alignment of feature representations with clinician focus. Attention alignment module contributes to robust convergence through improved matching stability. To formalize this assignment, we define the Gaze Hungarian Matrix with lesion-containing queries aligned before gaze-only queries:

$$L_{\text{match}}(y_i, y_{\sigma(i)}) = \begin{cases} -\hat{P}_{\sigma(i)}(c_i) + L_{\text{B}}(b_i, b_{\sigma(i)}) + L_{\text{A}}(a_i, a_{\sigma(i)}), & y_i \in \text{Object} \\ \max(L_C^{Obj}) + \max(L_{\text{B}}^{Obj}) + L_{\text{A}}(a_i, a_{\sigma(i)}), & y_i \in \text{Gaze-only} \end{cases}, \quad (2)$$

where y_i denotes the ground truth, $y_{\sigma(i)}$ represents predicted results, b is the bbox, and a repesents the heatmap, L_C denotes focal classification loss, L_{B} is L1 loss for box regression, and L_{A} corresponds to MSE loss for attention alignment.

Then, a novel **G**aze **A**ttention **G**uidance **D**etection (**GAGD**) loss effectively integrates gaze semantic information into the model. We define it as:

$$L_{GAGD} = \lambda_1 L_C + \lambda_2 L_{\text{B}} + \lambda_3 L_A^{Obj} + \lambda_4 L_A^{Gaze-only}, \quad (3)$$

where weighting coefficients λ_1, λ_2, λ_3 and λ_4 maintain component balance. The gaze-only attention term guides query regression toward clinical regions of interest (e.g., lesion-like confounding areas). This mechanism not only enhances semantic understanding through additional contextual cues to distinguish between normal and abnormal regions, but also enables efficient feature utilization by increasing query dispersion. Overall, our GAGD Loss achieves holistic optimization of detection accuracy while encoding clinicians' attention patterns.

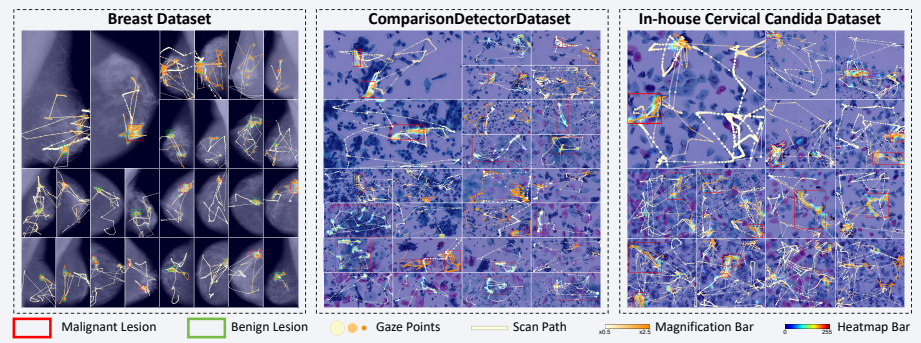

Fig. 3. Dataset Overview. The datasets include annotations for malignant lesions (red) and benign lesions (green), eye-tracking data such as gaze points (orange), scan path (yellow) and instant magnification of the image, alongside the processed gaze heatmaps.

The framework eliminates post-processing dependencies (e.g., NMS) through its one-to-one matching paradigm, establishing an end-to-end detection pipeline, while efficiently enhancing the detection performance through synergistic alignment of computational focus with diagnostic priorities from gaze.

3 Dataset

Our dataset comprising 2 public and 1 private subsets, totaling 2450 images with eye-tracking data. The comprehensive gaze data for both public subsets will be open-sourced, establishing the first gaze dataset for lesion detection.

Gaze-tracking data were collected using the Tobii 4C eye-tracker at 100 Hz with the software [24]. The tracking module was non-disruptively embedded within radiologists' diagnostic annotation workflows. Periodic recalibration (every 80 images) ensured measurement accuracy. Annotations were physician-verified or cross-validated, with quality control removing substandard data.

The dataset comprises 2,450 medical images with comprehensive gaze annotations shown in Fig. 3. Gaze potentially highlights tumor morphology and peritumoral vasculature to aid breast tumor detection [10] and fungal hyphae and spore structures to improve cervical candida detection [4]. It includes:

Breast Dataset: 857 border-trimmed mammograms sourced from [20,13], with 367 malignant and 389 benign tumor from a 3-year-experienced researcher.

ComparisonDetector (CD) Dataset: 781 cervical TCT images with 702 annotations, sourced from [14] and labeled by a 2-year-experienced researcher.

In-House Cervical Candida Dataset: 812 cervical TCT images with 1234 annotations, labeled by two researchers (2/3-year-experienced).

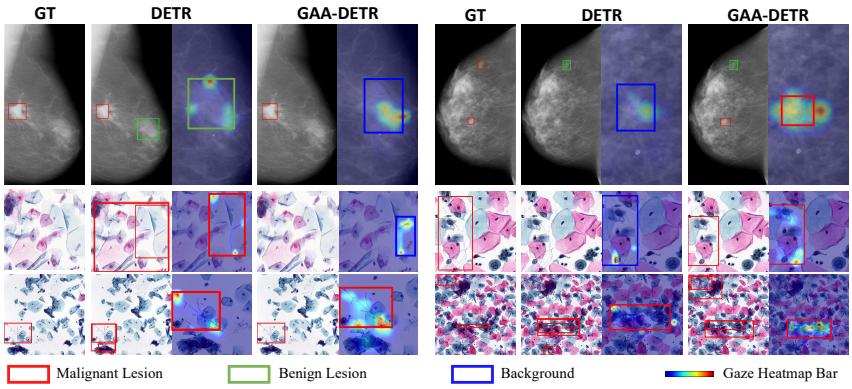

Fig. 4. Qualitative results. Our GAA-DETR focuses on internal clinical features and improves detection performance in both radiological and pathological tasks.

4 Experiments

Experimental Settings To ensure experimental consistency, all models were trained on a NVIDIA RTX A6000 GPU for an equivalent epoch. The dataset was randomly split at a ratio of 6:1:1 for training/validation/testing. We report test set performance of the best validation checkpoint using standard detection metrics [5] ($AP_{0.5:0.95}$, $AP_{0.5}$, $AP_{0.75}$, AR_{10}, AR_{100}) and our novel Attention Distribution (AD) metric. Given that x_c, y_c, w, h denote coordinates of the predicted box, $AD = \frac{1}{N}\sum_{x,y} \mathrm{A}(x,y)P(x,y)\frac{(x-x_c)(y-y_c)}{wh}$. Higher AD scores reflect a stronger concentration of model attention around box boundaries.

Table 1. Comparative performance metrics of ours and other gaze-based models.

Dataset	Method	$AP_{0.5:0.95}$	$AP_{0.5}$	$AP_{0.75}$	AR_{10}	AR_{100}	AD
Breast Dataset	MDF-Net	0.142	0.298	0.161	0.399	0.402	0.193
	DETR+CG-CAM	0.182	0.337	0.150	0.411	0.436	**0.113**
	Gaze-DETR	0.188	0.344	0.188	**0.453**	**0.507**	0.167
	GAA-DETR(Ours)	**0.199**	**0.351**	**0.205**	0.434	0.483	0.124
CD Dataset	MDF-Net	0.131	0.287	0.109	0.299	0.320	0.142
	DETR+CG-CAM	0.101	0.175	0.088	0.376	**0.593**	0.145
	Gaze-DETR	0.133	0.325	0.097	**0.400**	0.408	0.162
	GAA-DETR (Ours)	**0.158**	**0.340**	**0.115**	0.396	0.567	**0.131**
In-house Dataset	MDF-Net	0.119	0.347	0.054	0.278	0.302	0.396
	DETR+CG-CAM	0.165	0.427	0.103	0.339	0.483	0.367
	Gaze-DETR	0.170	0.442	0.093	0.346	0.502	0.330
	GAA-DETR(Ours)	**0.188**	**0.504**	**0.123**	**0.360**	**0.545**	**0.317**

Comparison with Other Gaze-based Detection Methods Benchmarked against the feature fusion-based MDF-Net [8], image-level alignment method DETR+CG-CAM [28], and gaze denoising method Gaze-DETR [12], as shown in Table 1, our GAA-DETR demonstrates leading performance across all evaluation metrics, indicating optimal gaze information utilization.

We also present qualitative results in the Fig. 4. Our method effectively focuses on peritumoral vessels and morphological context in mammographic images, as well as hyphal and spore structures in pathological candida, thus enhancing the detection performance. Both qualitative results and quantitative AD metrics reveal a correlation between higher performance and internal focus, supporting our hypothesis that boundary-focused attention is less effective.

Table 2. Performance comparison of different models with and without our GAA.

Method	Breast Dataset			CD Dataset			In-house Dataset		
	$AP_{0.5:0.95}$	AR_{100}	AD	$AP_{0.5:0.95}$	AR_{100}	AD	$AP_{0.5:0.95}$	AR_{100}	AD
DETR	0.170	0.468	0.193	0.115	0.555	0.142	0.143	0.487	0.396
GAA-DETR	0.199	0.483	<u>0.124</u>	0.158	0.567	<u>0.131</u>	0.188	0.544	0.317
Dino	<u>0.283</u>	0.589	0.188	0.230	<u>0.668</u>	0.133	<u>0.193</u>	<u>0.594</u>	0.327
GAA-Dino	**0.294**	<u>0.602</u>	0.164	<u>0.242</u>	**0.698**	**0.112**	**0.204**	**0.600**	**0.301**
RT-DETR	0.220	0.593	0.185	0.238	0.648	0.171	0.149	0.584	0.387
GAA-RT-DETR	0.244	**0.609**	**0.110**	**0.258**	0.664	0.159	0.164	0.575	<u>0.309</u>

Compatibility of our method To demonstrate the compatibility of our method and further validate its effectiveness, we extend our method to current mainstream detection models: the foundational DETR [5], the state-of-the-art Dino [26], and the real-time optimized RT-DETR [27]. As depicted in the Table 2, our approach consistently delivered performance improvements across all frameworks and datasets, affirming the effectiveness and compatibility of our method.

Table 3. Ablation Study evaluating the impact of the adaptive gaze kernel (AGK), gaze Hungarian assignment (GHA), and Gaze-Attention Guided Detection loss (GAGD).

Method			Breast Dataset		CD		In-house Dataset	
AGK	GHA	GAGDL	$AP_{0.5:0.95}$	AR_{100}	$AP_{0.5:0.95}$	AR_{100}	$AP_{0.5:0.95}$	AR_{100}
			0.170	0.468	0.115	0.544	0.143	0.487
	✓		0.175	0.477	0.121	0.537	0.153	0.480
		✓	0.179	0.481	0.138	**0.658**	0.165	0.480
	✓	✓	<u>0.180</u>	**0.501**	<u>0.145</u>	0.492	<u>0.180</u>	<u>0.487</u>
✓	✓	✓	**0.199**	<u>0.483</u>	**0.158**	<u>0.567</u>	**0.188**	**0.544**

Ablation Study We performed an ablation study on GAA-DETR to evaluate the impact of each component. The results demonstrate the utility of each module, with particularly significant improvements observed from the Gaze Adaptive

Kernel and GAGD Loss components. These findings highlight the importance of lesion-level alignment and the processing of gaze data.

5 Conclusion

Our study introduces GAA-DETR, a novel detection framework that integrates gaze data to enhance lesion detection through query-level alignment. By focusing on clinically relevant features that often overlooked by traditional boundary-based methods, our approach achieves superior performance in both radiological and pathological tasks, underscoring the value of gaze-pattern integration. We further contribute a novel gaze dataset for lesion detection with comprehensive gaze annotations, offering a platform for further research into gaze-informed clinical tools. We hope this contribution will facilitate further research into reliable AI solutions for healthcare that integrate human diagnostic expertise.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgement. This work was supported by AI & AI for Science Project of Nanjing University (0205-14380344, 4810-14380006).

References

1. Albuquerque, C., Henriques, R., Castelli, M.: Deep learning-based object detection algorithms in medical imaging: Systematic review. Heliyon **11**(1) (2025)
2. Bhattacharya, M., Jain, S., Prasanna, P.: Gazeradar: A gaze and radiomics-guided disease localization framework. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 686–696. Springer (2022)
3. Bigolin Lanfredi, R., Zhang, M., Auffermann, W.F., Chan, J., Duong, P.A.T., Srikumar, V., Drew, T., Schroeder, J.D., Tasdizen, T.: Reflacx, a dataset of reports and eye-tracking data for localization of abnormalities in chest x-rays. Scientific Data **9**(1), 350 (2022)
4. Cai, J., Xiong, H., Cao, M., Liu, L., Zhang, L., Wang, Q.: Progressive attention guidance for whole slide vulvovaginal candidiasis screening. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 233–242. Springer (2023)
5. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on Computer Vision. pp. 213–229. Springer (2020)
6. Fang, Y., Chen, C., Yuan, Y., Tong, K.y.: Selective feature aggregation network with area-boundary constraints for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. pp. 302–310. Springer (2019)
7. He, L., Todorovic, S.: Destr: Object detection with split transformer. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 9377–9386 (2022)

8. Hsieh, C.: Human-centred multimodal deep learning models for chest x-ray diagnosis. In: IJCAI. pp. 7085–7086 (2023)
9. Hsieh, C., Ouyang, C., Nascimento, J.C., Pereira, J., Jorge, J., Moreira, C.: Mimic-eye: Integrating mimic datasets with reflacx and eye gaze for multimodal deep learning applications. PhysioNet (version 1.0. 0) (2023)
10. Ji, C., Du, C., Zhang, Q., Wang, S., Ma, C., Xie, J., Zhou, Y., He, H., Shen, D.: Mammo-net: Integrating gaze supervision and interactive information in multi-view mammogram classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 68–78. Springer (2023)
11. Karargyris, A., Kashyap, S., Lourentzou, I., Wu, J.T., Sharma, A., Tong, M., Abedin, S., Beymer, D., Mukherjee, V., Krupinski, E.A., et al.: Creation and validation of a chest x-ray dataset with eye-tracking and report dictation for ai development. Scientific Data **8**(1), 92 (2021)
12. Kong, Y., Wang, S., Cai, J., Zhao, Z., Shen, Z., Li, Y., Fei, M., Wang, Q.: Gaze-detr: Using expert gaze to reduce false positives in vulvovaginal candidiasis screening. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 133–143. Springer (2024)
13. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific Data **4**(1), 1–9 (2017)
14. Liang, Y., Tang, Z., Yan, M., Chen, J., Liu, Q., Xiang, Y.: Comparison detector for cervical cell/clumps detection in the limited data scenario. Neurocomputing **437**, 195–205 (2021)
15. Luís, A., Hsieh, C., Nobre, I.B., Sousa, S.C., Maciel, A., Jorge, J., Moreira, C.: Integrating eye-gaze data into cxr dl approaches: A preliminary study. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). pp. 196–199. IEEE (2023)
16. Ma, C., Jiang, H., Chen, W., Li, Y., Wu, Z., Yu, X., Liu, Z., Guo, L., Zhu, D., Zhang, T., et al.: Eye-gaze guided multi-modal alignment for medical representation learning. Advances in Neural Information Processing Systems **37**, 6126–6153 (2025)
17. Mariam, K., Afzal, O.M., Hussain, W., Javed, M.U., Kiyani, A., Rajpoot, N., Khurram, S.A., Khan, H.A.: On smart gaze based annotation of histopathology images for training of deep convolutional neural networks. IEEE Journal of Biomedical and Health Informatics **26**(7), 3025–3036 (2022)
18. Meng, D., Chen, X., Fan, Z., Zeng, G., Li, H., Yuan, Y., Sun, L., Wang, J.: Conditional detr for fast training convergence. In: Proceedings of the IEEE/CVF International conference on Computer Vision. pp. 3651–3660 (2021)
19. Moradizeyveh, S., Tabassum, M., Liu, S., Newport, R.A., Beheshti, A., Di Ieva, A.: When eye-tracking meets machine learning: A systematic review on applications in medical image analysis. arXiv preprint arXiv:2403.07834 (2024)
20. Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: Inbreast: toward a full-field digital mammographic database. Academic Radiology **19**(2), 236–248 (2012)
21. Noton, D., Stark, L.: Scanpaths in eye movements during pattern perception. Science **171**(3968), 308–311 (1971)
22. Ren, X., Chu, S., Ji, G., Zhao, Z., Zhao, J., Qiang, Y., Wei, Y., Wang, Y.: Omsf2: optimizing multi-scale feature fusion learning for pneumoconiosis staging diagnosis through data specificity augmentation. Complex & Intelligent Systems **11**(1), 109 (2025)

23. Song, Y., Wang, X., Yao, J., Liu, W., Zhang, J., Xu, X.: Vitgaze: gaze following with interaction features in vision transformers. Visual Intelligence **2**(1), 1–15 (2024)
24. Wang, S., Ouyang, X., Liu, T., Wang, Q., Shen, D.: Follow my eye: Using gaze to supervise computer-aided diagnosis. IEEE Transactions on Medical Imaging **41**(7), 1688–1698 (2022)
25. Wang, S., Zhao, Z., Zhuang, Z., Ouyang, X., Zhang, L., Li, Z., Ma, C., Liu, T., Shen, D., Wang, Q.: Learning better contrastive view from radiologist's gaze. Pattern Recognition **162**, 111350 (2025)
26. Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L.M., Shum, H.: DINO: DETR with improved denoising anchor boxes for end-to-end object detection. In: The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. pp. 1–19. OpenReview.net (2023)
27. Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J.: Detrs beat yolos on real-time object detection. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 16965–16974 (2024)
28. Zhu, H., Salcudean, S., Rohling, R.: Gaze-guided class activation mapping: Leverage human visual attention for network attention in chest x-rays classification. In: Proceedings of the 15th International Symposium on Visual Information Communication and Interaction. pp. 1–8 (2022)
29. Zhu, X., Li, X., Ong, K., Zhang, W., Li, W., Li, L., Young, D., Su, Y., Shang, B., Peng, L., et al.: Hybrid ai-assistive diagnostic model permits rapid tbs classification of cervical liquid-based thin-layer cell smears. Nature Communications **12**(1), 3541 (2021)