IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)
e-ISSN: 2278-1684, p-ISSN: 2320-334X, Volume 22, Issue 6 Ser. 1 (Nov - Dec 2025), PP 17-21
www.iosrjournals.org

Portable Ear Canal Visualization Device And AI Recognition System: A YOLO-Based Lesion Detection Scheme

Siyuan Wang, Wenting Wang, Xiaoru Fang, Yang Yang
School Of Mechanical And Automotive Engineering, Shanghai University Of Engineer Science, Shanghai, China

Abstract:
To address the issues in external auditory canal (EAC) lesion diagnosis, such as reliance on professional experience, poor portability of traditional equipment, and low primary screening efficiency, a dedicated portable visualization device for the EAC was designed. Additionally, an intelligent EAC lesion recognition scheme based on the YOLO algorithm was proposed, constructing an integrated "acquisition-preprocessing-recognition-feedback" system. Focusing on the anatomical characteristics of the EAC, the device adopts a short-focus macro lens (with a focal length of 3-8 mm) and an annular fill-light module, and optimizes the handheld lightweight design. It can accurately capture images of the EAC wall mucosa, cerumen areas, and lesion regions, and supports wireless real-time transmission. This solves the pain points of traditional equipment, including difficulty in adapting to the narrow space of the EAC, blurred imaging, and insufficient safety for ordinary household operations.
For the EAC images collected by the device— which are susceptible to hair occlusion and uneven illumination, and whose lesions are mostly small targets such as local redness and swelling, ulcers, and tiny foreign bodies— a targeted preprocessing workflow was designed. Data quality was improved through operations such as adaptive noise reduction, hair occlusion weakening, and image cropping, and a dedicated dataset covering 4 scenarios (normal EAC, EAC inflammation, EAC foreign bodies, and EAC ulcers) was constructed. Based on the optimized model structure of YOLOv8-nano, combined with the DINO self-supervised visual representation learning framework, the anchor box sizes adapted to small EAC targets were re-clustered to improve the detection accuracy of local lesions.
The system can quickly output lesion types, confidence levels, and preliminary care recommendations, enabling primary screening of EAC lesions without professional operation. This scheme balances portability, low cost, and intelligence. It allows primary screening of EAC lesions without professional doctors' operation, providing an efficient solution for primary medical care and ordinary household health monitoring, and thus has significant practical value and promotion prospects.
Keywords: *YOLO Algorithm; Ear Canal Lesions; Image Recognition; Object Detection; Artificial Intelligence; Auxiliary Diagnosis*

Date of Submission: 04-11-2025 Date of Acceptance: 14-11-2025

I. Introduction

In recent years, with the rapid development of science and technology and the progress of social economy, the level of medical care has been significantly improved. Health issues related to body parts such as the eyes, respiratory tract, and oral cavity have received great attention from the public and society. However, against this backdrop, ear canal health has been widely overlooked.

Even when only discussing foreign bodies in the EAC, from August 2009 to July 2010, the Department of Otorhinolaryngology of Beilun District People's Hospital in Ningbo, Zhejiang Province, admitted 76 patients with EAC foreign bodies. Among them, there were 50 males and 26 females; 41 patients were aged 2-8 years, 13 were aged 9-18 years, and 22 were aged 19-72 years, covering populations of different age groups. When such a situation occurs, seeking medical attention in a timely manner is undoubtedly necessary. Ordinary people may not be able to confirm the presence of foreign bodies in the ear canal, which can lead to more serious complications. Foreign bodies in the ear canal may still be recognizable, but diseases such as EAC inflammation and EAC ulcers are even more difficult for ordinary people to distinguish. Misjudgment of symptoms and neglect of ear canal health may cause patients to take incorrect self-treatment methods, leading to the deterioration of the condition.

Therefore, during self-testing of ear canal health, if primary screening of EAC lesions can be completed without professional operation, it will be of great help for us to judge our health status. This study focuses on the actual needs of EAC lesion diagnosis, designs a dedicated portable visualization device for the EAC, proposes a hybrid detection model integrating YOLO and DINO technologies, and constructs an integrated intelligent "acquisition-preprocessing-recognition-feedback" system. The main research contents of this study are as follows:
(1) Design a portable visualization device adapted to the EAC to realize high-definition image acquisition and high-speed transmission in narrow spaces;
(2) Design a targeted image preprocessing workflow;
(3) Propose a YOLO-DINO hybrid detection model and optimize feature extraction and fusion strategies;

II. Design Of The Dedicated Portable Visualization Device For The External Auditory Canal

The device is composed of four parts: an imaging module, a wearing and positioning module, a holding and control module, and a transmission module. It has a length of 20 cm and a width of 6 cm, meeting the requirements of portable use. The innovations of the device mainly focus on the wearing and positioning module and the holding and control module.

Wearing and Positioning Module: In-Ear Adaptive Structure

Drawing on the wearing principle of in-ear headphones, an in-ear positioning structure that fits the physiological structure of the EAC was designed to ensure acquisition depth and usage safety:

- Size Adaptation Design: The in-ear part is made of soft silica gel, with 3 replaceable sizes (S/M/L) designed to adapt to the EAC diameters (0.5-1 cm) of different populations. The silica gel material has a certain elasticity, which can fit the ear canal wall closely without causing a sense of oppression, avoiding the displacement of the device during the acquisition process;
- Depth Limitation Design: The length of the in-ear part is fixed at 1.5 cm. When the device is fully worn, the camera is exactly 1 cm inside the entrance of the EAC. This not only ensures that images of the deep part of the EAC can be captured but also prevents damage to the eardrum due to excessive insertion, eliminating operational risks from the structural design;
- Safety Protection Design: The surface of the silica gel is smooth without edges and corners, and the edges are rounded. No sharp components are exposed. At the same time, it has good biocompatibility. It meets medical standards through skin irritation tests and can be reused (with disposable disinfection sleeves), reducing the risk of cross-infection.

Figure1: *Silicone Protective Cover for the In-Ear Section*

Holding and Control Module: Ergonomic Grip

To solve the problem of shaking during handheld acquisition, an ergonomic grip structure was designed to improve operational stability: The grip adopts an arc-shaped curved surface design, which fits the natural holding posture of the palm. The surface of the grip is provided with anti-slip patterns to increase the holding friction, ensuring stability even if the hands sweat slightly. The length of the grip is suitable for single-hand holding by adults, and fingers can naturally rest on the function buttons for easy operation.

Figure2: *Rendering of the Device's Grip Section*

Summary

In addition to the innovative in-ear adaptive structure and ergonomic grip, the device also includes an imaging module and a transmission module. Through structural optimization and function integration, it has successfully achieved high-definition image acquisition in the narrow space of the EAC. At the same time, it has the characteristics of portability, safety, ease of operation, and high-speed transmission, providing high-quality and stable data source support for the subsequent YOLO-DINO hybrid model and adapting to the actual use scenarios of primary medical care and home self-examination.

III. Module Framework Of The AI Recognition System

Acquisition Layer: Portable EAC Visualization Device

As the core data source customized for EAC detection, the device takes "adapting to the EAC, portability, and high imaging quality" as its core goals, providing high-quality original data for the YOLO-DINO hybrid model.

Preprocessing Layer: Targeted Image Optimization Module

In view of the characteristics of EAC images—"hair occlusion, lens distortion, and small lesion regions"—a targeted preprocessing workflow was designed to improve data quality and lay a foundation for the collaborative work of YOLO and DINO:

1. Image Enhancement and Interference Suppression:
   - Illumination Correction: The adaptive CLAHE algorithm is used to adjust the image contrast to the range of 0.9-1.3, highlighting the grayscale difference between lesion regions (such as congested and red-swollen mucosa) and normal tissues;
   - Hair Occlusion Weakening: Combining morphological filtering and inpainting algorithms, hair pixel regions with a length of ≥5 mm in the image are identified and repaired, reducing the interference of hair on lesion feature extraction without damaging lesion details.
2. Image Correction and Standardization:
   - Distortion Correction: Based on the internal parameter matrix calibrated by the device, perspective transformation is used to correct radial distortion, ensuring that the EAC wall presents a regular cylindrical contour;
   - Region of Interest (ROI) Cropping and Scaling: Redundant regions are automatically cropped to retain the core EAC region, which is then scaled to 640×640 pixels (compatible with the input requirements of YOLO and DINO) to improve the model operation efficiency.
3. Data Annotation and Format Conversion: The LabelImg tool is used to annotate 4 types of targets ("normal EAC", "EAC inflammation", "EAC foreign bodies", and "EAC ulcers"), with an annotation accuracy of ≥96%. The annotated XML files are converted into TXT format (compatible with YOLO) and COCO format (compatible with DINO) respectively, and the dataset is divided into training set, validation set, and test set in a ratio of 7:2:1.

Algorithm Layer: YOLO-DINO Hybrid Detection Model

To address the limited ability of the single YOLO algorithm to detect tiny and weak-feature lesions in the EAC, a hybrid model was constructed by integrating the advanced feature extraction capability of DINO and the real-time detection advantage of YOLO, achieving a balance between "accuracy and speed":

1. Dual-Backbone Feature Extraction:
   - YOLOv8-nano Backbone Network: Responsible for extracting shallow detailed features of EAC images (such as lesion edges and local textures), and ensuring real-time performance relying on its lightweight advantage;
   - DINO Feature Encoder: The Transformer-based feature extraction module of DINO is introduced. Through the self-attention mechanism, the model's ability to capture deep semantic features (such as the global context of lesions) is enhanced, effectively suppressing interference from non-lesion regions (such as normal mucosa and residual cerumen). There have been previous applications of the DINO model in the medical field, and its effectiveness has been proven in gastrointestinal endoscopy.
2. Cross-Scale Feature Fusion: A bidirectional feature fusion module was designed to fuse the shallow detailed features output by YOLO and the deep semantic features output by DINO. For tiny lesions (such as small foreign bodies), the weight of shallow features is increased; for diffuse lesions (such as extensive EAC inflammation), the weight of deep features is increased, thereby improving the detection accuracy of lesions of different sizes.
3. Guided Detection Head Design:
   - Adaptive Anchor Boxes: Based on the statistical characteristics of the EAC lesion dataset, 3 groups of anchor boxes ([12,18], [25,35], [45,55] pixels) were re-clustered to match the sizes of common EAC lesions;

- DINO-Guided Attention: The attention map output by DINO is embedded into the YOLO detection head to guide the model to focus on potential lesion regions, reducing false detections caused by similar textures (such as cerumen and ulcers).

4. Training and Post-Processing Strategies:

- Training: Transfer learning is adopted. The YOLO backbone network loads the pre-trained weights of YOLOv8-nano, and the DINO encoder loads the pre-trained weights of ImageNet;
- Post-Processing: Soft-NMS is combined with DINO-based false detection suppression—detection boxes with a confidence level <0.55 are filtered out, and the global features of DINO are used to judge and eliminate false detection boxes in non-lesion regions, further reducing the false detection rate.

Application Layer: Result Visualization and Auxiliary Feedback

With "ease of use and practicality" as the core, the application layer realizes the visualization and practical output of detection results:

1. Real-Time Visualization: The terminal APP displays the original EAC image in real time, with red detection boxes (marking lesion positions) and lesion type labels (such as "EAC inflammation") overlaid;
2. Data Management: Automatically stores image files and detection reports (including lesion type, confidence level, and acquisition time), and supports querying historical records by date/lesion type, facilitating subsequent comparative diagnosis;
3. Auxiliary Feedback: Outputs personalized preliminary care recommendations based on detection results (such as "Suspected EAC inflammation: Keep the EAC dry and avoid water entry; seek medical attention promptly if symptoms persist for more than 3 days"), guiding users to handle the condition scientifically without professional medical knowledge.

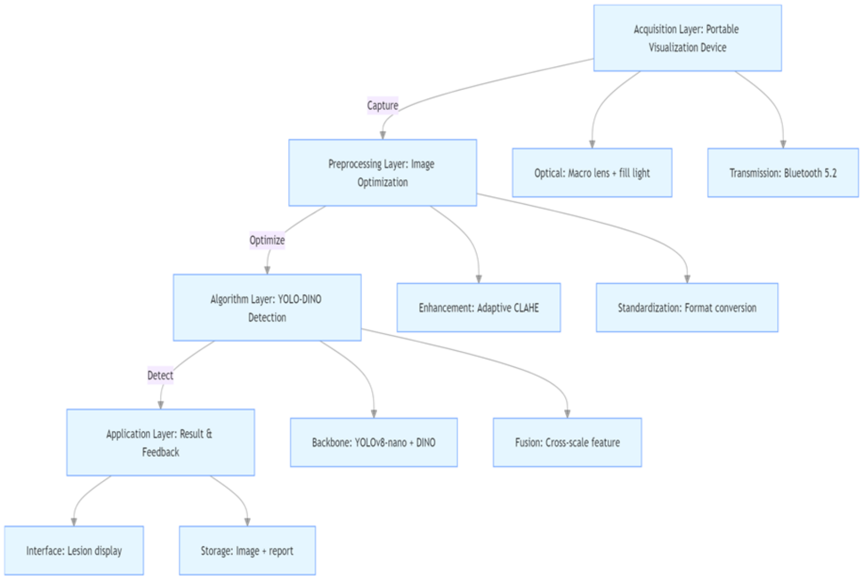

Figure3: *System Overall Design Framework Diagram*

IV. Application Prospects

The innovation of this study lies in the realization of collaborative innovation of "hardware adaptation + algorithm optimization". The device is custom-designed according to the anatomical characteristics of the EAC, and the model is optimized and improved according to the characteristics of ear canal images. The in-depth integration of the two improves the practicality and reliability of the system. In the future, it can be applied in primary medical care scenarios, home health monitoring scenarios, and medical science popularization and teaching scenarios.

The core design idea of this system can be extended to other narrow-space and small-target detection scenarios, such as the primary screening of lesions in parts like the nasal cavity and oral mucosa. By optimizing the optical structure of the device and model training data, multi-part adaptation can be achieved. In the future, the applicability and promotion value of the system will be further improved by expanding multi-center datasets, introducing multi-modal fusion technology to optimize the model, and upgrading the device function to support video acquisition.

V. Conclusion

Aiming at the core pain points in EAC lesion diagnosis—"poor equipment adaptability, insufficient detection accuracy, and high operation threshold"—the design and development of a dedicated portable EAC visualization device and a YOLO-DINO hybrid detection system were completed. A portable visualization device adapted to EAC detection was designed, and an integrated "acquisition-preprocessing-recognition-feedback" system was constructed. The terminal APP supports the visualization of detection results, data storage, and output of care recommendations, and can be operated without professional medical knowledge, adapting to the scenarios of primary medical care and home self-examination.

The innovation of this study lies in the realization of collaborative innovation of "hardware adaptation + algorithm optimization". The device is custom-designed according to the anatomical characteristics of the EAC, and the model is optimized and improved according to the characteristics of ear canal images. The in-depth integration of the two improves the practicality and reliability of the system. This study was partially funded by the 2025 Shanghai University Student Innovation and Entrepreneurship Training Program of China. Project Number: cx2501039.

References

[1]. Ji, K., Xia, G. H., Liu, S. S. (2011). 76 Cases Of Foreign Bodies In The External Auditory Canal[J]. Chinese Journal Of Ophthalmology And Otorhinolaryngology, 11(1): 55.

[2]. Tian, Y., Ye, Q., Doermann, D. (2025). Yolov12: Attention-Centric Real-Time Object Detectors[EB/OL]. (2025)[2025-11-09]. Https://Arxiv.Org/Abs/2502.12524.

[3]. Ansari, M. M. (2025). Academic Sharing: Evaluating The Performance Of Grounding DINO, YOLO, And DINO In Vascular Stenosis Detection Based On The ARCADE Dataset[EB/OL]. (2025-04-11)[2025-11-09]. Https://Blog.Csdn.Net/Weixin_51607793/Article/Details/147144838