Cutting 10 Hours to 30 Seconds: Deep Learning Precision Cropping of Strabismus Eye Regions for Workflow Optimization and Standardized AI Model Training Data Preprocessing

Dawen Wu, Yanfei Li, Zeyi Yang, Teng Yin, Xiaohang Chen, Jingyu Liu, Wenyi Shang, Bin Xie, Guoyuan Yang, Haixian Zhang, Long Qian Liu

Submitted to: Journal of Medical Internet Research
on: March 24, 2025

Table of Contents

Cutting 10 Hours to 30 Seconds: Deep Learning Precision Cropping of Strabismus Eye Regions for Workflow Optimization and Standardized AI Model Training Data Preprocessing

Dawen Wu[1*] Dr med; Yanfei Li[2*] MSc; Zeyi Yang[1]; Teng Yin[2]; Xiaohang Chen[1]; Jingyu Liu[2]; Wenyi Shang[2]; Bin Xie[3]; Guoyuan Yang[1]; Haixian Zhang[2]; Long Qian Liu[1] DProf, PhD

[1] Department of Optometry and Vision Sciences, West China Hospital, Sichuan University, Chengdu, China Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China Chengdu CN
[2] Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China Chengdu CN
[3] Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China Chengdu CN
[*]these authors contributed equally

Corresponding Author:
Long Qian Liu DProf, PhD

Department of Optometry and Vision Sciences, West China Hospital, Sichuan University, Chengdu, China
Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China
Chengdu 610041, Sichuan Province, China
Chengdu
CN

Abstract

Background: Privacy protection for strabismus patients necessitates de-identification of facial information in ocular gaze photographs used for AI model development, outpatient medical records, telemedicine, and academic exchange. This requires cropping full-face images to include only the eye region. However, conventional preprocessing methods are time-consuming, labor-intensive, and plagued by non-standardized practices, highlighting the need for an efficient and standardized solution.

Objective: This study aimed to develop an AI-driven management platform for strabismus patients, leveraging preprocessing algorithms based on ocular gaze photographs. The platform uses advanced algorithms to optimize eye region cropping, improving accuracy, efficiency, and standardization in both clinical workflows and AI model training data preprocessing.

Methods: This retrospective and prospective cross-sectional study used 5832 images from inpatients and outpatients, capturing three gaze positions under varying conditions. The model was evaluated using precision, recall, and mean average precision (mAP) at various Intersection over Union (IoU) thresholds through 5-fold cross-validation on an inpatient dataset and tested on an independent outpatient dataset. Expert validation confirmed alignment with clinical standards, and a control experiment with five optometry specialists compared manual and automated cropping efficiency for real-time applications.

Results: The model achieved a mean precision of 1, recall of 1, mAP50 of 0.995, and mAP95 of 0.893 across the 5-fold cross-validation set. On the independent test set, the precision was 1, recall was 1, mAP50 was 0.995, and mAP95 was 0.801. Expert validation of 48 images confirmed adherence to clinical and research standards. A control experiment with five optometry specialists demonstrated a reduction in image preparation time from 10 hours to 30 seconds for 900 photos.

Conclusions: The AI-driven management platform for strabismus patients integrates an intelligent cropping algorithm, electronic archives for comprehensive patient histories, and a patient-physician interaction module. This platform not only optimizes the eye region cropping process, improving both accuracy and efficiency, but also offers a complete solution for strabismus care. The system accelerates clinical workflows with seamless uploads to electronic health records, while ensuring patient privacy during remote consultations. This innovative approach has the potential to significantly enhance data preparation, patient care, and clinical practices in ophthalmology. Clinical Trial: This retrospective and prospective cross-sectional study received approval from the Ethics Committee of West China Hospital, Sichuan University, China (2023 Review No. 1477).

(JMIR Preprints 24/03/2025:74402)
DOI: https://doi.org/10.2196/preprints.74402

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Cutting 10 Hours to 30 Seconds: Deep Learning Precision Cropping of Strabismus Eye Regions for Workflow Optimization and Standardized AI Model Training Data Preprocessing

Dawen Wu, MD[1,3*], Yanfei Li[2*], Zeyi Yang[1,3], Teng Yin[2], Xiaohang Chen[1,3], Jingyu Liu[2], Wenyi Shang[2], Bin Xie[1], Guoyuan Yang, MD[1,3*], Haixian Zhang, PhD[2*], Longqian Liu, MD[1,3*]

Affiliations:

[1]Department of Ophthalmology, West China Hospital, Sichuan University, Chengdu, China

[2]Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China

[3]Department of Optometry and Vision Sciences, West China Hospital, Sichuan University, Chengdu, China

* Dawen Wu, and Yanfei Li contributed equally to this work.

[*]**Corresponding Author:**

Longqian Liu, MD, PhD

Department of Ophthalmology, West China Hospital, Sichuan University

Chengdu, 610041, China

Tel: +86 18980601759

Email: b.q15651@hotmail.com

Haixian Zhang, PhD

Machine Intelligence Laboratory, College of Computer Science, Sichuan University

Tel: +86 18980917399

Email: zhanghaixian@scu.edu.cn

Guoyuan Yang, MD

Department of Ophthalmology, West China Hospital, Sichuan University

Chengdu, 610041, China

Tel: +86 18980603049

Email: 31981077@qq.com

Abstract

Background

This study aimed to develop an AI-driven management platform for strabismus patients, leveraging preprocessing algorithms based on ocular gaze photographs. The platform uses advanced algorithms to optimize eye region cropping, improving accuracy, efficiency, and standardization in both clinical workflows and AI model training data preprocessing.

Methods

This retrospective and prospective cross-sectional study used 5832 images from inpatients and outpatients, capturing three gaze positions under varying conditions. The model was evaluated using precision, recall, and mean average precision (mAP) at various Intersection over Union (IoU) thresholds through 5-fold cross-validation on an inpatient dataset and tested on an independent outpatient dataset. Expert validation confirmed alignment with clinical standards, and a control experiment with five optometry specialists compared manual and automated cropping efficiency for real-time applications.

Results

The model achieved a mean precision of 1, recall of 1, mAP50 of 0.995, and mAP95 of 0.893 across the 5-fold cross-validation set. On the independent test set, the precision was 1, recall was 1, mAP50 was 0.995, and mAP95 was 0.801. Expert validation of 48 images confirmed adherence to clinical and research standards. A control experiment with five optometry specialists demonstrated a reduction in image preparation time from 10 hours to 30 seconds for 900 photos.

Conclusions

The AI-driven management platform for strabismus patients integrates an intelligent cropping algorithm, electronic archives for comprehensive patient histories, and a patient-physician interaction module. This platform not only optimizes the eye region cropping process, improving both accuracy and efficiency, but also offers a complete solution for strabismus care. The system accelerates clinical workflows with seamless uploads to electronic health records, while ensuring patient privacy during remote consultations. This innovative approach has the potential to significantly enhance data preparation, patient care, and clinical practices in ophthalmology.

Keywords: Artificial intelligence; strabismus; Deep learning; Intelligent cropping; gaze photos; AI management platform

Introduction

Strabismus, marked by binocular misalignment, affects 0.8% to 6.8%(1) of children but can manifest at any age, impacting visual function, appearance, and social interactions including romantic relationships, learning, and employment opportunities(2). Early detection and treatment are crucial, as they can considerably improve outcomes, underscoring the importance of timely intervention in managing this disorder.

Recent initiatives have been directed towards crafting a dependable artificial intelligence (AI) system for diagnosing strabismus, utilizing ocular alignment photographs(3, 4). This process requires image pre-processing to isolate and focus on the eye region, ensuring the exclusion of identifiable facial features for privacy protection and optimal dataset preparation. Historically, the development of strabismus screening AI has been encumbered by two manual and laborious tasks: cropping photographs and correcting for head tilts to achieve a horizontal alignment of the eyes(5). This specificity in image preparation is not only critical for AI model training but also permeates clinical and academic settings, where there's a need to share or upload ocular alignment images that are devoid of any extraneous facial features. Presently, the process of converting full-face photographs to ones that exclusively display the eye region is predominantly a manual, time-consuming endeavor. Hence, the introduction of an automated, efficient algorithm for eye region cropping promises to significantly alleviate the workload of medical professionals and reduce healthcare costs.

Previous researches have explored automating eye region cropping using algorithms like Dlib-toolkit(6, 7) and Faster R-CNN(8). However, these approaches have encountered obstacles, such as the necessity for comprehensive facial landmarks and the challenge of accommodating variations in head tilt, which have impacted the preservation of photo cropping accuracy and the integrity of image features. Consequently, the developed algorithms demand considerable requirements from the original full-face photographs and the automatically cropped photos often necessitate manual readjustment, limiting the practical application of the developed algorithms.

Building upon the YOLOv8(9) backbone, we developed an advanced rotating bounding box detection algorithm to identify the eye region while accounting for head tilt angles. This method involves initial cropping of full-face images, followed by alignment correction (as illustrated in Figure 1), ensuring accurate eye region detection. Additionally, existing solutions are limited to initial screening and often lack functionalities for patient referral and long-term management, which can result in gaps in care for strabismus patients. To bridge this gap, we have developed an AI-driven management platform for strabismus, available as a mobile applet compatible with devices such as smartphones and tablets. This platform integrates digital archives for storing patient test results and prescriptions, ensuring comprehensive records of patient history. Furthermore, it features a patient-physician interaction module, facilitating direct communication for better follow-up care. Together, these features provide a seamless, continuous care experience from initial screening to long-term monitoring.

Methods

1. Data Collection

This retrospective and prospective cross-sectional study received approval from the Ethics Committee of West China Hospital, Sichuan University, China (2023 Review No. 1477). Figure 1 delineates the study's comprehensive workflow. Data were amassed from two distinct datasets spanning various age groups, one derived from inpatients and the other from outpatients. Both datasets included photographs of strabismus patients across Primary Gaze, Secondary Gaze, and Tertiary Gaze positions, with each patient contributing one, four, and four photographs for these positions respectively, amounting to a total of nine distinct eye positions, showcasing variability in distances. A critical aspect of this methodology is its adaptability to variations in facial positioning, diverse head tilt angles, facial features across different age groups, and obstructions like fingers or cotton swabs used to retract eyelids.

The inpatient dataset, consisting of 5400 ocular gaze photographs from 600 patients, was utilized to construct a 5-fold cross-validation set for the training process. The dataset was stratified and randomly split into training and validation sets, following an 80% to 20% ratio. These images were captured from patients both pre- and post-surgery, admitted to the Department of Ophthalmology at West China Hospital, China, from January 2018 to October 2023, using a Canon EOS M50 Mark II single-lens reflex autofocus digital camera, with a 24-million-pixel resolution and equipped with a lens-mounted flash. The camera was positioned at a distance ranging from 33 centimeters to 1 meter from the subjects. To evaluate the model's performance under diverse imaging conditions, our study incorporated an outpatient-based dataset, consisting of 432 photographs from 48 patients, collected from August 1, 2023, to October 1, 2023. This dataset, featuring a variety of imaging devices, served as an independent test set. Photographs of 12 patients were captured using the EOS M50 Mark II (Canon Inc., Tokyo, Japan), while the remaining 36 patients' photographs were taken with the Huawei (Huawei Inc., China), OPPO (OPPO Inc., China) series, and other leading mobile phone brands. The capturing devices were positioned at distances ranging from 33 centimeters to 1 meter from the patients.

2. **Data annotation**

In this study, we employed roLabelImg 3.0 software for precise eye region annotation in strabismus patient photographs, a necessary step for developing our intelligent cropping algorithm. This tool allows for the detailed marking of rotated rectangle regions, crucial for accurately capturing the varied orientations of patients' heads and eyes. A Senior Ophthalmologist and an AI Medicine Expert from the Computer Science department collaboratively annotated each image, with every annotated image subsequently reviewed by a Veteran Ophthalmologist boasting over 35 years of clinical and research experience. This ensured the identification of the eye region with high accuracy. The capability of roLabelImg to create adjusted bounding boxes for rotations was especially beneficial for our dataset, which includes photographs where faces and eyes often deviate from the horizontal. This process enabled precise annotations, accommodating head tilts and eyelid positions, which are typical in strabismus cases. These annotations are critical for training and testing the model, providing it with a comprehensive understanding of the eye region's geometry, and significantly improving its performance in accurately cropping the eye region from full-face photographs.

3. Data processing

To maintain the original aspect ratio, all images were resized to 640 × 640 pixels via letterboxing. Moreover, beyond employing standard data augmentation techniques like random stretching, Gaussian blurring, and contrast adjustment, a crucial enhancement strategy was the implementation of Mosaic data augmentation. This technique amalgamates four separate images into a single composite, substantially broadening the training dataset's variety. Consequently, it fortifies the model's capability to navigate and interpret complex visual environments, thereby improving its overall robustness and performance in diverse conditions, as illustrated in Supplementary Material Figure 1.

4. Model Development and Performance Evaluation

4.1 Model Architecture and Training Process

Considering the model's speed and efficiency, we adopted YOLOv8 as the backbone. The backbone component incorporates the C2f module to improve gradient flow throughout the network, enabling certain upsampling operations to be executed with a single CNN layer and thus simplifying the network architecture. Network pruning was achieved via sparse training by selectively eliminating parameters based on their impact on model performance. For the head section, a decoupled head along with an anchor-free mechanism was implemented. This design segregates bounding box regression and classification into distinct pathways. Specifically, for this task, which requires only bounding box regression without the need for classification, such architecture minimizes the influence of classification on detection tasks, thereby boosting detection efficacy. The model was trained over 300 epochs, employing a linear learning rate schedule with a warmup phase. SGD served as the optimizer, with a batch size of 128 and an initial learning rate of 0.01. Furthermore, a foundational weight decay of 0.005 was applied during training. This comprehensive approach not only optimizes model performance but also enhances computational efficiency, supporting streamlined and effective object detection.

4.2 Optimization Strategy

Considering the characteristics of this task, we utilized only regression loss as the model's total loss. The regression loss comprises Distribution Focal Loss (DFL) and ProbIoU Loss, with DFL_Reg_max set to 16 by default. ProbIoU harnesses the properties of rotated boxes, enhancing the conventional IOU calculation by incorporating angle information, thereby facilitating the computation of rotated box IOU (refer to Section 4.3 for details). Unlike modeling a single Dirac distribution for bounding box coordinates, DFL loss regresses an arbitrary distribution to model the boundary box. Given the complexity of precisely defining the boundaries of eye sockets, slightly larger boxes are considered acceptable. DFL loss allows the network to focus promptly on values near the labels, maximizing probability density at the labels and thus constraining the model's predictions within an acceptable range. This method employs the cross-entropy function to optimize probabilities around the label y, concentrating the network distribution near the label values. The final overall loss is derived by combining these two components with appropriate weighting.

$$B_1=\frac{1}{4}\frac{(a_1+a_2)(y_1-y_2)^2+(b_1-b_2)(x_1-x_2)^2+2(c_1+c_2)(x_2-x_1)(y_1-y_2)}{(a_1+a_2)(b_1+b_2)-(c_1+c_2)^2}$$

$$B_2=\frac{1}{2}\ln\frac{(a_1+a_2)(b_1+b_2)+(c_1+c_2)^2}{4\sqrt{(a_1 b_1-c_1^2)\text{¿¿¿}}}$$

$$L_{Prob}(p,q)=1-ProbIoU(p,q)\in[0,1]$$

$$L_{DFL}(S_i,S_{i+1})=-((y_{i+1}-y)\log(S_i)+(y-y_i)\log(S_{i+1}))$$

$$L_{total}=0.8*L_{Prob}+0.2*L_{DFL}(S_i,S_{i+1})$$

4.3 IoU Calculation for Skewed Rectangles

We implement a method for calculating the skew IoU (Intersection over Union), incorporating triangulation. Figure 2 illustrates the geometric principles. Our objective is to calculate the IoU for each pair of skew rectangles. Consider rectangle ABCD as the prediction and rectangle EFGH as the artificially designed, real rectangle. First, we add the intersection points of the two rectangles to set P (e.g., points I, J, K, L in Figure 2B), including the vertices of one rectangle located within the other into set P (e.g., points A, C in Figure 2B). Then, we sort the points in set P to form a convex polygon (e.g., polygon AIJCKL in Figure 2B). Triangulation yields a set of triangles (shown in Figure 2B as △AIJ, △AJC, △ACK, △AKL). The polygon's area is the sum of these triangles' areas. Finally, this setup allows us to calculate the IoU using the intersection to union ratio formula:

$$IoU=\frac{Area(I)}{Area(rectangle_{ABCD})+Area(rectangle_{EFGH})-Area(I)}$$

4.4 Evaluation metrics

The trained model's performance is evaluated using precision rate, recall rate, and Mean Average Precision (mAP) at various Intersection over Union (IoU) thresholds. Precision rate measures the proportion of correctly predicted eye region boxes, while recall rate evaluates how well the model identifies all relevant eye regions. The mAP50 assesses precision at a 50% IoU threshold, and mAP95 averages precision across IoU thresholds from 50% to 95%, offering a comprehensive measure of model accuracy(10). The evaluation process was conducted on two datasets: a five-fold cross-validation set composed of 5400 eye position photos and an independent test set consisting of 432 eye position photos. The cross-validation was used to ensure the model's robustness, with precision, recall, and mAP metrics calculated for each fold. The final model was selected based on the best-performing fold, and this model was then applied to the independent test set to further validate its performance across the aforementioned metrics. Additionally, an expert in pediatric and strabismus ophthalmology with 35 years of clinical and research experience was consulted to evaluate whether the model's eye region recognition and cropping performance meet clinical and research standards (see Results section for actual images of eye region recognition and cropping by the model).

To further validate the model's efficiency, a control experiment was conducted in which five optometry specialists each manually cropped and uploaded 81 eye position photos from 9 patients, with each patient contributing 1 photo in Primary Gaze, 4 in Secondary Gaze, and 4 in Tertiary Gaze positions. The average time per patient required by each physician to crop and upload these photos was recorded and compared to the average time taken by the model for the same task. This comparison aims to explore the model's potential in reducing image processing time and assess its suitability for real-time clinical applications.

Results

1. Baseline Characteristics of the Study Participants

In this study, we harnessed a dataset from 648 patients with 5832 images across three gaze positions. The demographic and image attributes of participants within our study's datasets are detailed in Supplementary Material Table 1, with the distributions of Ocular Alignment Angle (OAA) (Training set: median: 0, range: [-12.603, 5.737]; Testing set: median: 0, range: [-5.713, 5.737]) and Eye Region Bounding Box Area (ERBBA) (Training set: median: 0.134, range: [0.022, 0.305]; Testing set: median: 0.127, range: [0.024, 0.249]) depicted in Supplementary Material Figure 2.A, 2.B respectively. These two variables are pivotal in the clinical acquisition of ocular position photographs. The OAA, measuring the angle (radians) between the line connecting the eyes' inner canthi and the horizontal plane, is particularly significant due to the frequent occurrence of head tilt in patient images. This metric enables the model to autonomously correct for head tilt, ensuring the eyes are aligned horizontally. The ERBBA, expressed as a percentage, reflects the eye region's proportion of the entire image frame, highlighting the importance of accounting for non-uniform shooting distances and different device magnifications, which lead to bounding boxes of the eye region that vary significantly in size. This aspect of variability highlights our model's adaptability and robustness, efficiently navigating the complexities introduced by diverse photographic conditions to achieve precise detection outcomes.

2. Model Performance Evaluation

2.1 Quantitative Model Performance and Responsiveness

The model achieved a mean precision and recall of 1.000 (95% confidence interval

[CI]: 1.000-1.000), an mAP50 of 0.995 (95% CI: 0.995-0.995), and an mAP95 of 0.893 (95% CI: 0.870-0.918) across the 5-fold cross-validation set. The model's steady convergence during training is visualized in Figure 3. On the independent test set, precision and recall were 1.000, with an mAP50 of 0.995 and mAP95 of 0.801, as detailed in Table 1. Manually cropping and uploading 9 photos per patient took specialists an average of 6.52 minutes, with a near-normal time distribution across physicians (Supplementary Material Figure 3). By contrast, the model completed the task in just 27 milliseconds per photo, totaling 0.243 seconds for all 9 photos, underscoring its efficiency for real-time clinical use.

2.2 Qualitative Expert Validation

The model's eye region detection and cropping were qualitatively assessed by a pediatric and strabismus ophthalmology expert with over 35 years of experience, focusing on 48 randomly selected images spanning various distances, age groups, and skin tones, including obstructions such as fingers and cotton swabs, as illustrated in Supplementary Material Figure 4. This review garnered strong approval for the precision and outcome quality. Figure 4 showcases examples of head tilt correction and ocular region cropping for two patients. Such detailed evaluation confirms the model's adherence to clinical and research standards, showcasing its capability in accurate eye region recognition and cropping.

3. AI-Driven Management Platform

The AI-driven management platform developed in this study not only enhances the precision of strabismus eye region cropping but also integrates functionalities that improve overall patient care (as shown in Figure 5).

3.1 Patient Nine-Position Ocular Gaze Upload and Information Collection

The platform collects nine-position ocular gaze photographs, which are critical for strabismus screening, diagnosis, and AI model training. These photographs are captured by an ophthalmologist in a clinical setting, ensuring the accuracy and consistency of the data. In addition, essential patient data, including name, gender, date of birth, UCVA, refractive data, axial length, angle of deviation, and binocular vision function test results—previously recorded on paper—are digitally stored. This process is guided and assisted by the attending physician, ensuring that all examination results are accurately uploaded, thus creating a comprehensive patient profile.

3.2 Digital Medical Records

A secure digital record system allows for easy access to patients' diagnostic results and prescriptions. Both patients and healthcare providers can retrieve and review these records, ensuring that clinical decisions are based on up-to-date information. This feature also supports the integration of historical data, promoting continuity of care and aiding in both treatment progress tracking and informed clinical decision-making.

3.3 Patient-Physician Interaction Module

The platform includes an interactive module that allows seamless communication between patients and healthcare providers. This module facilitates consultation, sharing of symptoms, and response to inquiries, reducing the likelihood of missed follow-ups. Communication through text, images, and videos further improves the consultation process, providing personalized and efficient care.

Discussion

Strabismus screening, traditionally manual and reliant on physician expertise, demands significant patient-doctor cooperation, often resulting in subjective outcomes. Recent efforts focus on developing AI systems using photographs for quicker, more objective diagnoses(5). However, the advancement of strabismus screening AI has been significantly impeded by the laborious and prone-to-error task of manually cropping photographs to determine ocular alignment, particularly when correcting for head tilts in full-face images to achieve horizontal eye alignment. In clinical environments, such as outpatient clinics, there exists a parallel demand for an algorithm that enables clinicians to rapidly crop full-face photographs to isolate the eye region for inclusion in electronic health records, thereby preserving patient privacy and enhancing the quality of diagnostic imaging. This requirement is echoed in academic and telemedicine contexts, where the accuracy and precise alignment of eye region images are paramount. It is crucial to address these longstanding challenges by leveraging advanced technological solutions.

Previously proposed automatic eye region cropping algorithm was primarily based on the Dlib-toolkit facial feature point recognition function(6, 7, 11). This function utilizes a facial feature point detector to extract 68 facial landmarks (as shown in Figure 6.A) and locates the eye region based on the detected facial positional information for cropping. However, this method relies on additional facial landmarks to identify eye region markers, which becomes problematic in clinical settings where patients often wear masks, only the upper half of the face is photographed for clarity, or fingers and cotton swabs obstruct the extraction of facial landmarks during eyelid retraction. For such photos, the eye region cropping algorithm developed with the Dlib-toolkit facial feature point recognition function fails to operate. Furthermore, as the Dlib-toolkit function is encapsulated within the Pytorch platform, it cannot be modified.

Employing object detection algorithms presents an alternative approach for cropping the eye region in full-face photographs. Presently, object detection algorithms are divided into two principal methodologies: two-stage and one-stage detection. Two-stage detectors, such as Faster R-CNN(8), R-FCN(12), and Mask R-CNN(13), initiate by generating a multitude of region proposals within images, which are then classified and refined. This sequential process underpins the nomenclature of two-stage detection. There has been attempts to automate eye region cropping using the Faster R-CNN object detection algorithm(14), with the results shown in Figure 6.B. However, this algorithm faces challenges such as misalignment of the eyes on the same horizontal plane affecting diagnostic accuracy, and variation in the aspect ratio of cropped images due to patient head tilt, resulting in loss of original features when resized for model input. Consequently, images often require manual readjustment following automated cropping by the algorithm to rectify these discrepancies.

One-stage detectors, such as YOLO(15), SSD(16) and M2Det(17), streamline object detection by directly predicting categories and locations in a single step, bypassing proposal box generation. This efficiency renders them faster and more suitable for real-time tasks with limited computational resources than their two-stage counterparts. Within the array of one-stage detectors, the YOLO architecture distinguishes itself by striking an optimal balance between speed and precision, facilitating swift and dependable object recognition within images. Specifically, in medical applications, YOLO's application spans cancer detection(18), skin lesion segmentation(19), and pill identification(20), contributing to heightened diagnostic accuracy and streamlined treatment protocols(10).

Our study builds upon the YOLOv8(9) backbone, leveraging its advanced capabilities to develop a novel ocular gaze photograph preprocessing algorithm specifically tailored for intelligent cropping of the eye region in full-face photographs of strabismus patients. This algorithm adeptly accounts for head tilt angles during eye region detection, initiating with the cropping of the patient's full-face photo to isolate the eye region, followed by realigning the cropped image to ensure the eyes are aligned horizontally. It accurately crops images with eye tilt angles up to 5.74 radians, demonstrating robustness and enhancing its utility in clinical and telemedicine settings where precise identification is crucial. Furthermore, our model has achieved outstanding performance, with a precision rate of 1, recall of 1, mAP50 of 0.995, and mAP95 of 0.893, validated by ophthalmic clinicians for its efficacy in eye region cropping. The integration of both quantitative results and qualitative expert feedback in this dual evaluation method illustrates the comprehensive validation strategy of our study, reinforcing the model's suitability for both clinical practice and advanced research. Moreover, in a clinical setting, a physician would require approximately 10 hours to manually crop and adjust 900 eye position photos from 100 patients—a colossal endeavor prone to human error and inconsistency in the aspect ratio of eye position photos. By automating this process, our developed model excels in precision, completing the image rectification in less than 30 seconds. This significantly enhances the accuracy and efficiency of image preparation for strabismus screening, diagnosis, and surgery. Additionally, the model uniquely identifies the eye region as the primary target and offers a user-adjustable cropping boundary, allowing for the inclusion of surrounding areas based on user preference. This feature enhances the model's flexibility, catering to varying clinical needs.

Currently, many clinical records for strabismus patients are still maintained in paper format, leading to issues with the non-digital and unstructured collection and annotation of data. These challenges result in incomplete or inaccurate multimodal medical data, which limits the broader application and development of AI research in this field. Digital products in strabismus care remain scarce, and previous AI-driven screening and diagnostic efforts have typically only focused on delivering diagnostic results, lacking support for patient referrals and long-term follow-up. This gap often leads to missed follow-up visits, hindering comprehensive treatment and obstructing the collection of diverse, multimodal, and longitudinal data from patients(21). Our platform integrates electronic archives to create a secure and efficient system for storing patient test results and prescriptions. This integration allows for comprehensive and accurate tracking of patient histories, enabling informed clinical decision-making. Additionally, the platform enhances patient management by preventing missed appointments and ensuring timely treatment. By leveraging advanced data management tools, it facilitates large-scale population data collection, which in turn supports improved intervention strategies and prognosis predictions. This innovation represents a major step forward in strabismus management, with the potential to significantly improve patient outcomes and optimize the delivery of eye care services.

In summary, this study presents a novel approach in ophthalmology, developing an advanced ocular gaze photograph preprocessing algorithm for precise eye region cropping in strabismus patients. By automating a traditionally manual and error-prone process, our method significantly enhances the accuracy and efficiency of image preparation for strabismus screening and AI model training, thereby improving clinical workflows and diagnostic quality. Additionally, we developed a comprehensive AI-driven management platform, which integrates digital archives and patient-physician interaction modules, further optimizing patient care from screening through to long-term follow-up.

Declarations:

Ethics approval and consent to participate:

This retrospective and prospective cross-sectional study received approval from the Ethics Committee of West China Hospital, Sichuan University, China (2023 Review No. 1477).

Availability of data and materials:

DATA AVAILABILITY:

The data that support the findings of this study are available from the corresponding author, [LLQ], upon reasonable request.

CODE AVAILABILITY:

The algorithms of this study are available for download at: https://github.com/victor250214384/yolov8obb-for-strbismus

Competing interests:

The authors declare that they have no competing interests.

Funding:

This study was funded by the Peak Discipline Development Fund under the '1·3·5' Project of West China Hospital, Sichuan University (Grant No. ZYAI24033), Fang Qianxun-Tang Zeyuan Ophthalmic Clinical Medicine Charity Project (0040206107039) and the National Natural Science Foundation of China (Grant No. 82070996).

Author Contributions:

Dawen Wu was responsible for data collection, neural network construction, and was a major contributor in writing the manuscript. Yanfei Li and Haixian Zhang were responsible for code development and manuscript writing. Zeyi Yang and Xiaohang Chen was responsible for manuscript writing and manuscript editing. Guoyuan Yang and Teng Yin were involved in dataset selection. Jingyu Liu and Wenyi Shang contributed to the development of the model. Longqian Liu was responsible for research design, manuscript writing, manuscript editing and funding support. All authors read and approved the final manuscript.

Acknowledgements:

Not applicable

Reference

1. Chia A, Roy L, Seenyen L. Comitant horizontal strabismus: an Asian perspective. Br J Ophthalmol. 2007;91(10):1337-40.
2. Coats DK, Paysse EA, Towler AJ, Dipboye RL. Impact of large angle horizontal strabismus on ability to obtain employment. Ophthalmology. 2000;107(2):402-5.
3. Liu LQ, Wu DW. [Challenges and prospects in the application of artificial intelligence for ocular disease screening and diagnosis]. Chinese Journal of Ophthalmology. 2024;60(6):484-9.
4. Ming S, Yao X, Guo X, Guo Q, Xie K, Chen D, et al. Performance of ChatGPT in Ophthalmic Registration and Clinical Diagnosis: Cross-Sectional Study. J Med Internet Res. 2024;26:e60226.
5. Wu D, Li Y, Zhang H, Yang X, Mao Y, Chen B, et al. An artificial intelligence platform for the screening and managing of strabismus. Eye (Lond). 2024.
6. Kazemi V, Sullivan J, Ieee, editors. One Millisecond Face Alignment with an Ensemble of Regression Trees. 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2014 Jun 23-28; Columbus, OH. NEW YORK: Ieee; 2014.
7. King DE. Dlib-ml: A Machine Learning Toolkit. J Mach Learn Res. 2009;10:1755-8.
8. Girshick R, Ieee, editors. Fast R-CNN. IEEE International Conference on Computer Vision; 2015 Dec 11-18; Santiago, CHILE. NEW YORK: Ieee; 2015.
9. G. Jocher AC, and J. Qiu. “YOLO by Ultralytics” [Internet]. https://github.com/ultralytics/ultralytics2023 February. [cited 2023].
10. Terven J, Córdova-Esparza D-M, Romero-González J-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction. 2023;5(4):1680-716.
11. Huang X, Lee SJ, Kim CZ, Choi SH. An automatic screening method for strabismus detection based on image processing. PLoS One. 2021;16(8):e0255643.
12. Dai JF, Li Y, He KM, Sun J, editors. R-FCN: Object Detection via Region-based Fully Convolutional Networks. 30th Conference on Neural Information Processing Systems (NIPS); 2016; Barcelona, SPAIN. LA JOLLA: Neural Information Processing Systems (Nips); 2016.
13. He KM, Gkioxari G, Dollár P, Girshick R, Ieee, editors. Mask R-CNN. 16th IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22-29; Venice, ITALY. NEW YORK: Ieee; 2017.
14. Zheng C, Yao Q, Lu J, Xie X, Lin S, Wang Z, et al. Detection of Referable Horizontal Strabismus in Children's Primary Gaze Photographs Using Deep Learning. Transl Vis Sci Technol. 2021;10(1):33.
15. Redmon J, Divvala S, Girshick R, Farhadi A, Ieee, editors. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Seattle, WA. NEW YORK: Ieee; 2016.
16. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al., editors. SSD: Single Shot MultiBox Detector. 14th European Conference on Computer Vision (ECCV); 2016 Oct 08-16; Amsterdam, NETHERLANDS. CHAM: Springer International Publishing Ag; 2016.
17. Qijie Z, Tao S, Yongtao W, Zhi T, Ying C, Ling C, et al. M2Det: A Single-Shot Object Detector based on Multi-Level Feature Pyramid Network arXiv. arXiv (USA). 2018:8 pp.- pp.
18. Nie YL, Sommella P, O'Nils M, Liguori C, Lundgren J, Ieee, editors. Automatic Detection of Melanoma with Yolo Deep Convolutional Neural Networks. 7th E-Health and Bioengineering Conference (EHB); 2019 Nov 21-23; Grigore T Popa Univ Med & Pharmacy, Iasi, ROMANIA. NEW YORK: Ieee; 2019.
19. Ünver HM, Ayan E. Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm. Diagnostics. 2019;9(3):21.
20. Tan L, Huangfu TR, Wu LY, Chen WY. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. BMC Med Inform Decis Mak. 2021;21(1):11.
21. Wu D, Huang X, Chen L, Hou P, Liu L, Yang G. Integrating artificial intelligence in strabismus management: current research landscape and future directions. Exp Biol Med (Maywood). 2024;249:10320.

Table 1 Performance of the Model on Five-fold Cross-Validation and Independent Test Sets

Data set	Precision	Recall	mAP50	mAP95
Five-fold cross-validation set Mean, (95% CI)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)	0.995 (0.995, 0.995)	0.893 (0.870, 0.918)
Independent test set	1.000	1.000	0.995	0.801

mAP50, mean average precision at 50% intersection over union threshold.
mAP95, mean average precision averaged over intersection over union thresholds from 0.5 to 0.95.
95% CI, 95% confidence interval.

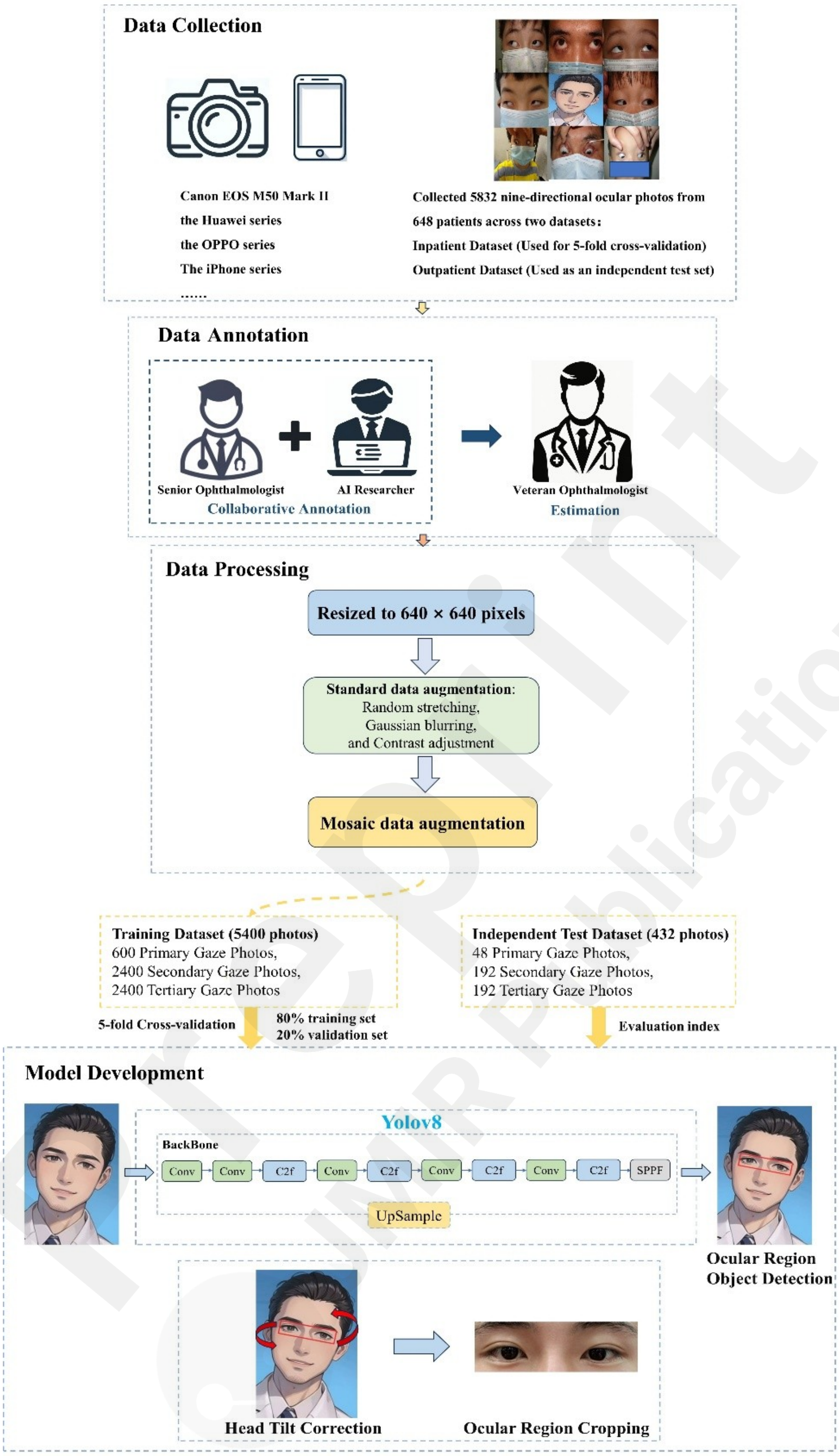


Figure 1 Workflow Overview for Intelligent Eye Region Cropping in Strabismus Patient Photographs

This Workflow Overview illustrates the comprehensive process undertaken in our study for intelligent eye region cropping in strabismus patient photographs. Beginning with data collection from 648 patients, we meticulously annotated eye regions, processed images for uniformity, and developed the model through systematic training and optimization strategies. Our approach encompasses both technical rigor in model architecture and practical considerations in data handling, ensuring precise eye region detection across diverse ocular positions.

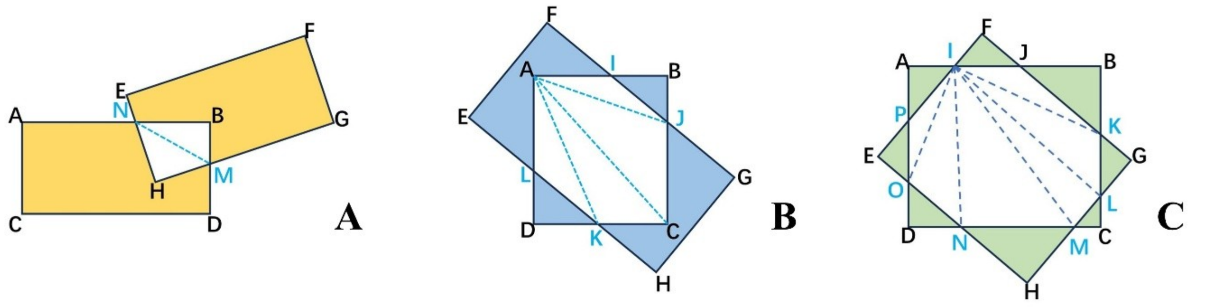


Figure 2 Geometric Principles of Skew IoU Calculation Using Triangulation
Figure 2 depicts the process of skew IoU calculation through triangulation, illustrating the geometric principles involved. Rectangles ABCD (prediction) and EFGH (real) intersect to form set P with points I, J, K, L, and vertices A, C. These points are sorted to create convex polygon AIJCKL, which is then divided into triangles △AIJ, △AJC, △ACK, △AKL for area calculation. This method enables precise IoU measurement by incorporating the angle information of skew rectangles.

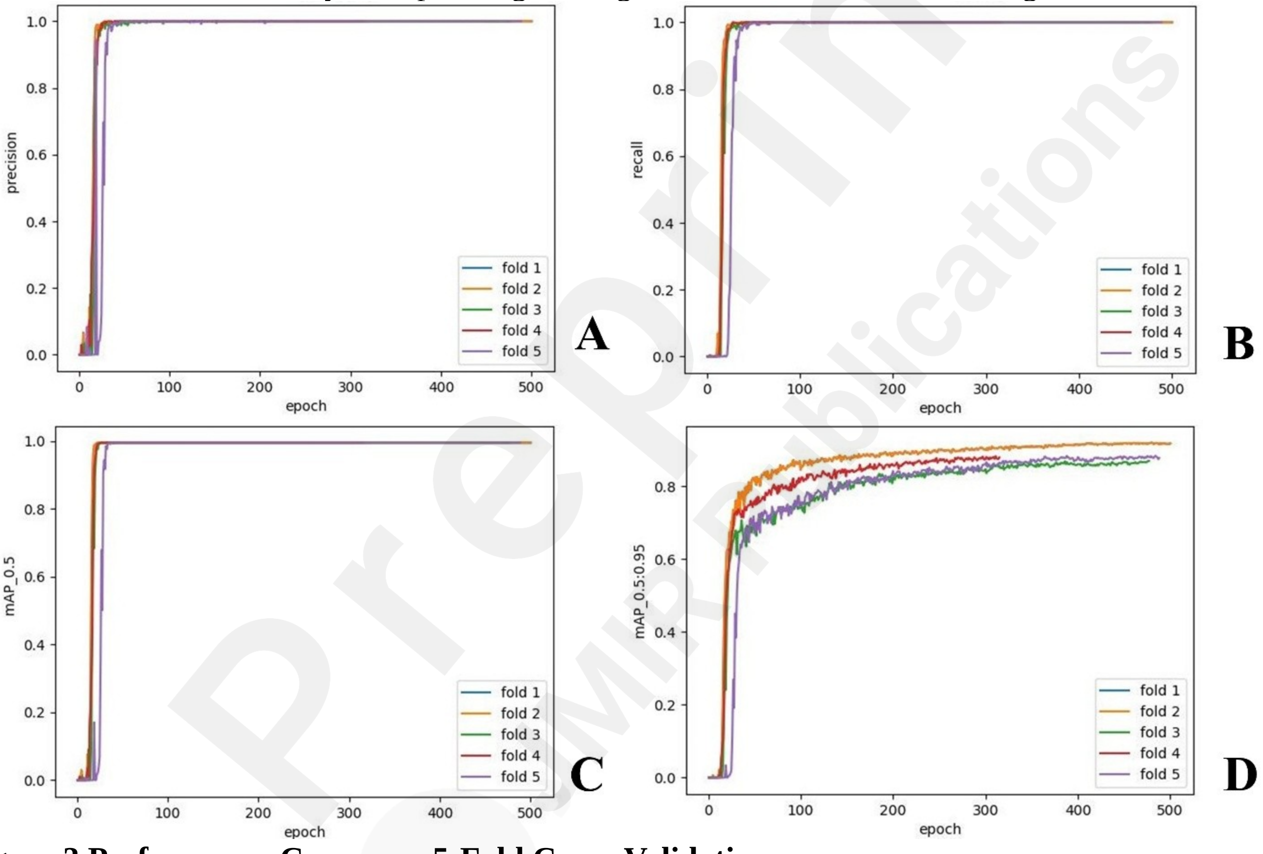


Figure 3 Performance Curves on 5-Fold Cross-Validation
Figure 3 demonstrates the model's performance across five validation folds during training. Subfigure A shows precision curves, all approaching 1.0, indicating high accuracy in eye region detection. Subfigure B highlights recall curves, also nearing 1.0, reflecting robust identification of true positives. Subfigure C presents mAP50, confirming the model's precision across all folds, while Subfigure D illustrates mAP95, depicting consistent improvement at stricter IoU thresholds.

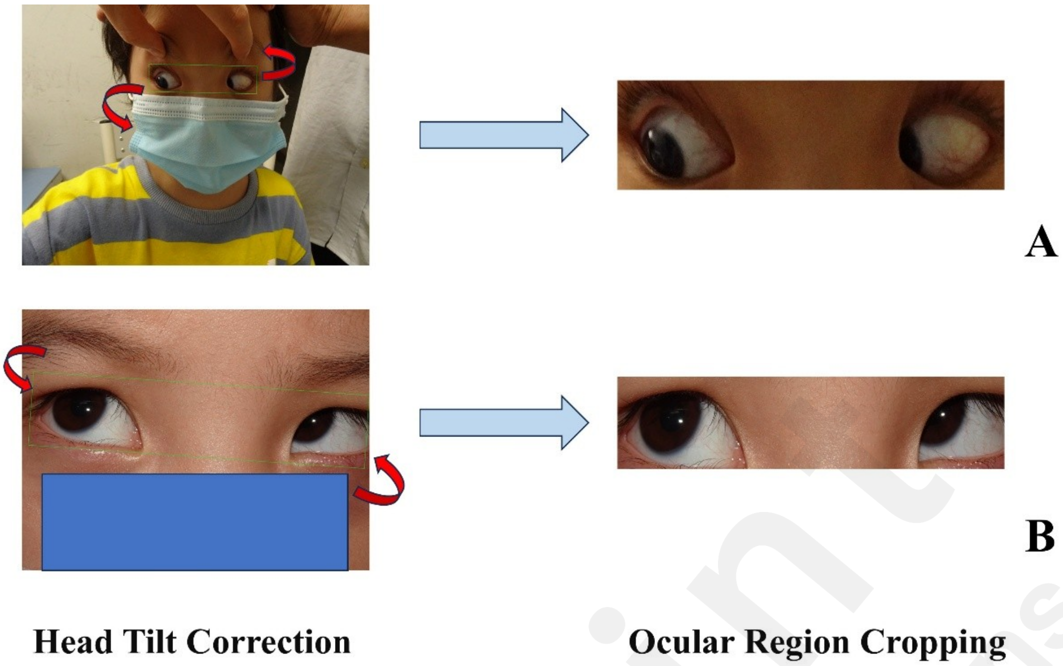


Figure 4 Head Tilt Correction and Ocular Region Cropping in Strabismus Patient Images
Figure 4 illustrates the model's detection, cropping, and head tilt correction across two patient photographs. In Figure 4.A, the Ocular Alignment Angle (OAA) is 0.457° and the Eye Region Bounding Box Area (ERBBA) is 0.054, while in Figure 4.B, the OAA is 2.382° and the ERBBA is 0.171. The model isolates the eye region and adjusts the image orientation to ensure horizontal eye alignment, offering user-adjustable cropping boundaries to accommodate varying clinical needs and enhance flexibility.

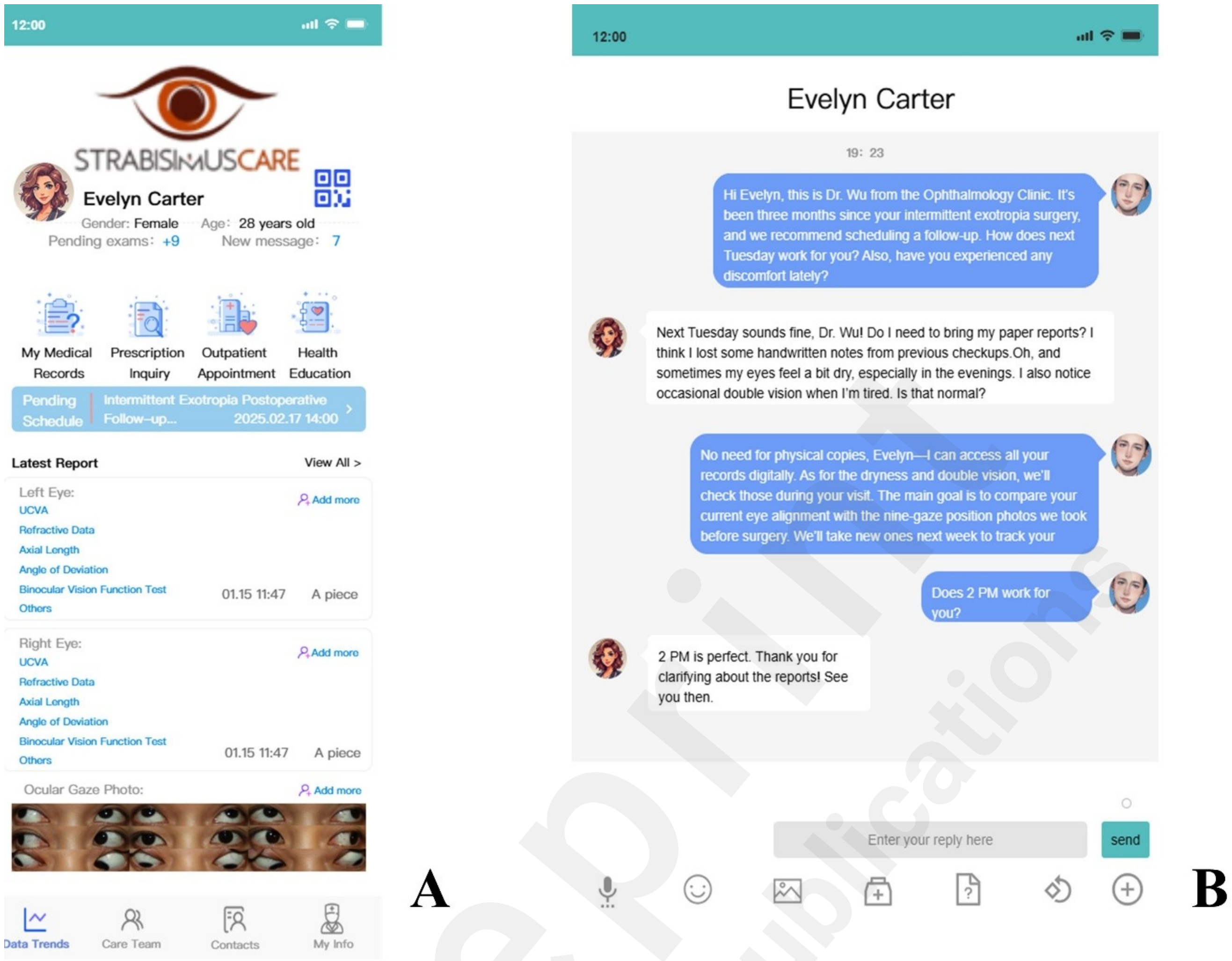


Figure 5 The Interface of the AI-Driven Management Platform for Strabismus Patient Care
Figure 5.A presents the digital medical records interface, where patients and healthcare providers can access and manage comprehensive patient histories, including diagnoses, prescriptions, and eye examination results. Figure 5.B depicts the doctor-patient interaction module, allowing seamless communication through text, images, and video, facilitating effective consultation and follow-up care.

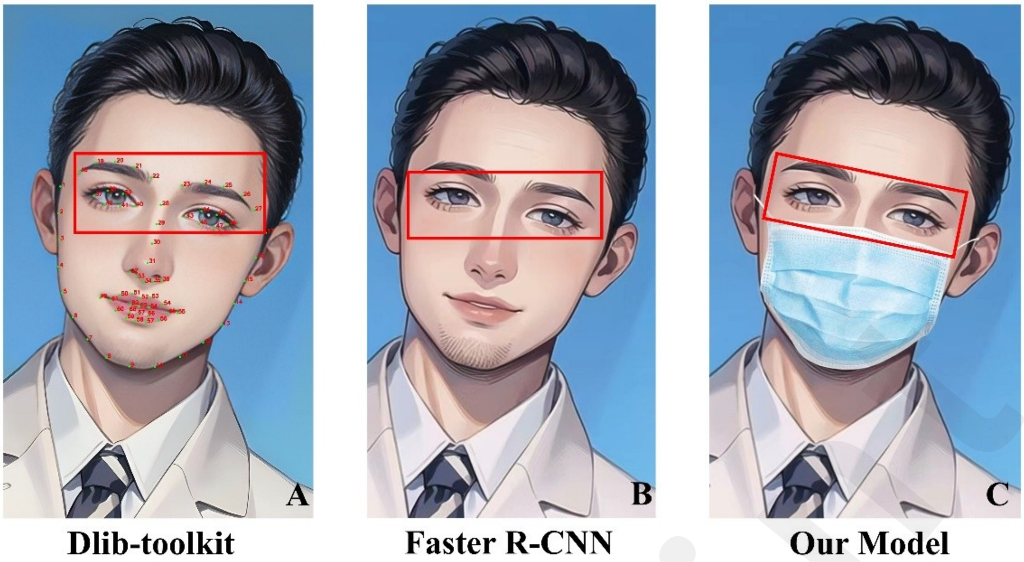


Figure 6 Comparative Evaluation of Eye Region Cropping Algorithms from Full-Face Patient Photographs

Figure 6.A illustrates the eye region cropping using the Dlib-toolkit facial feature point recognition function, which extracts 68 facial landmarks to identify and crop the eye region. The effectiveness of this method is contingent on the visibility of facial landmarks, which can be hindered by obstructions such as masks, fingers, or cotton swabs during eyelid retraction.

Figure 6.B showcases the application of the Faster R-CNN object detection algorithm for eye region cropping. This method generates region proposals followed by classification and refinement, but it can be challenged by the misalignment of eyes and the variability in the aspect ratio of cropped images, often necessitating manual adjustments post-cropping.

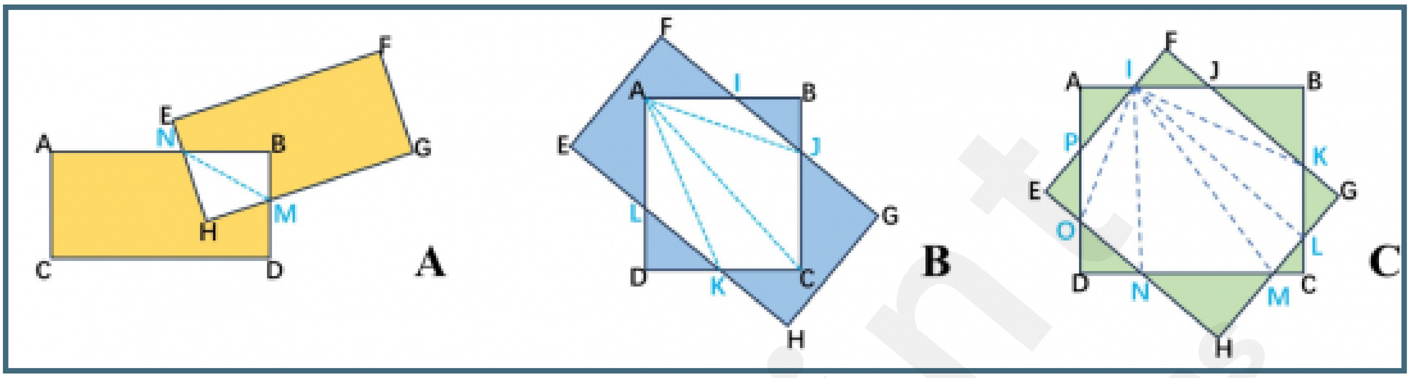
Figure 6.C displays the results of our proprietary algorithm for eye region cropping, specifically designed to address head tilts and to ensure horizontal alignment of the eyes. This innovative approach automates the cropping process while maintaining the original features of the eye region, reflecting significant improvements over prior methods.

Supplementary Files

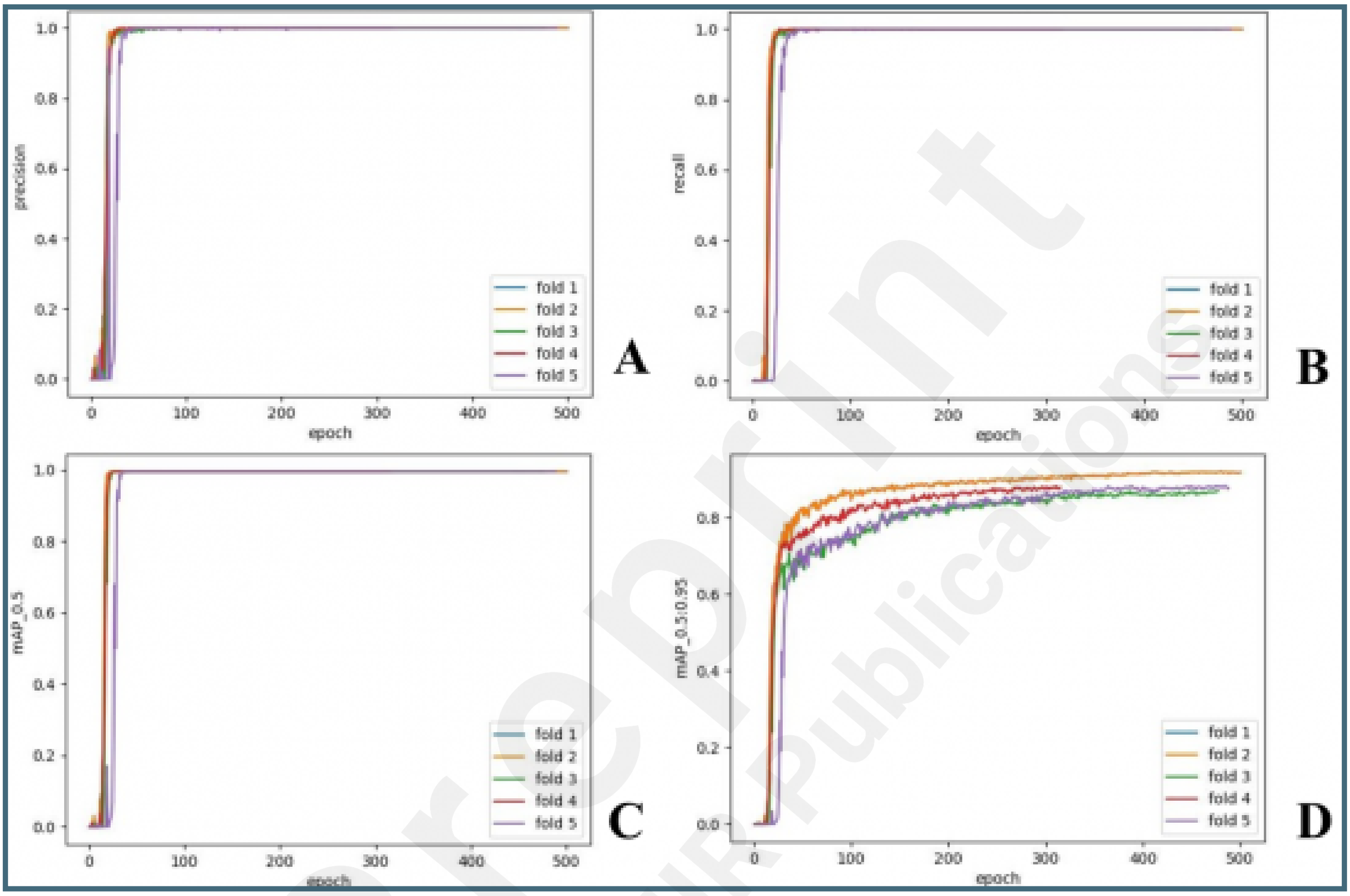
Figures

Workflow Overview for Intelligent Eye Region Cropping in Strabismus Patient Photographs This Workflow Overview illustrates the comprehensive process undertaken in our study for intelligent eye region cropping in strabismus patient photographs. Beginning with data collection from 648 patients, we meticulously annotated eye regions, processed images for uniformity, and developed the model through systematic training and optimization strategies. Our approach encompasses both technical rigor in model architecture and practical considerations in data handling, ensuring precise eye region detection across diverse ocular positions.

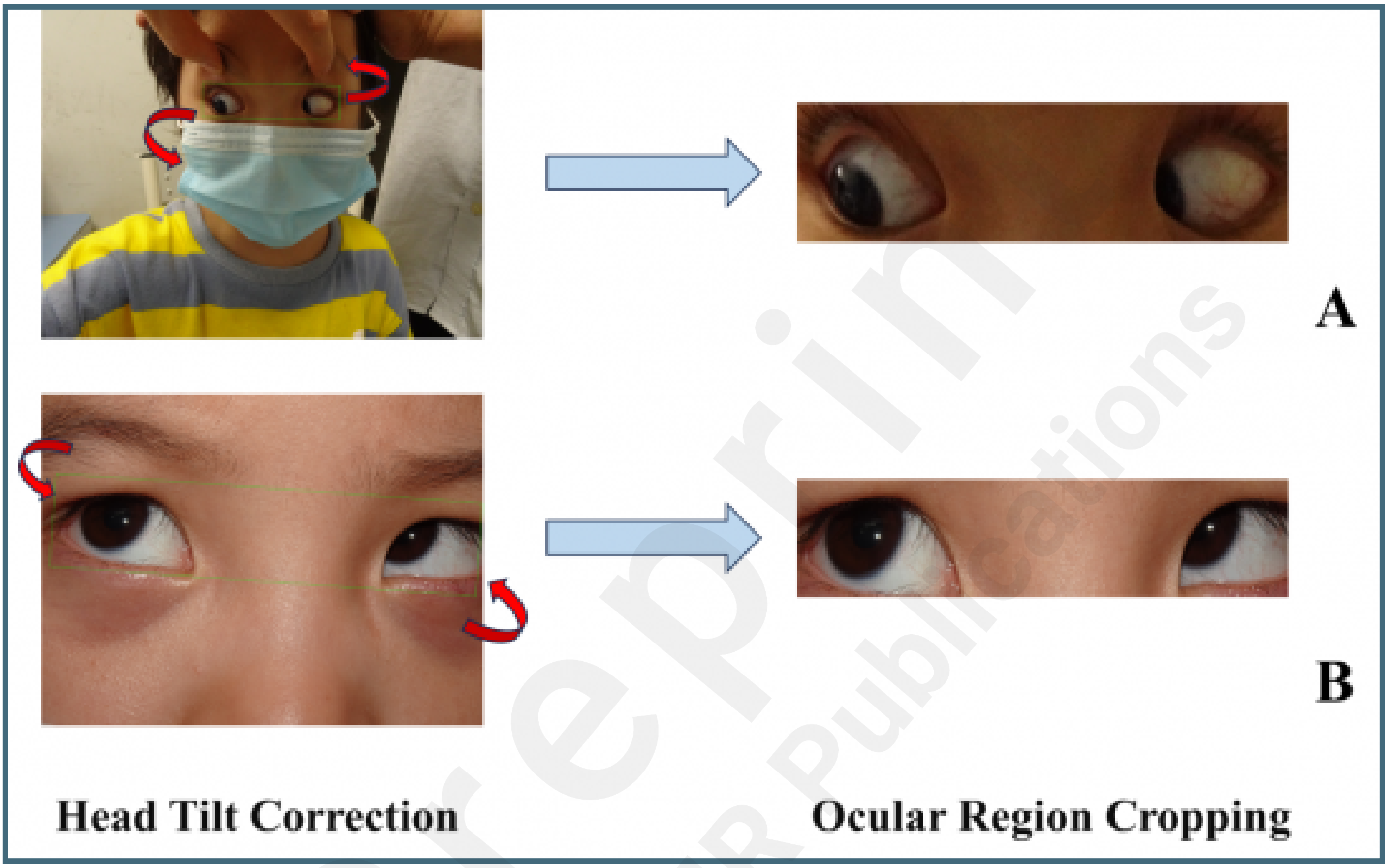
Geometric Principles of Skew IoU Calculation Using Triangulation Figure 2 depicts the process of skew IoU calculation through triangulation, illustrating the geometric principles involved. Rectangles ABCD (prediction) and EFGH (real) intersect to form set P with points I, J, K, L, and vertices A, C. These points are sorted to create convex polygon AIJCKL, which is then divided into triangles ?AIJ, ?AJC, ?ACK, ?AKL for area calculation. This method enables precise IoU measurement by incorporating the angle information of skew rectangles.



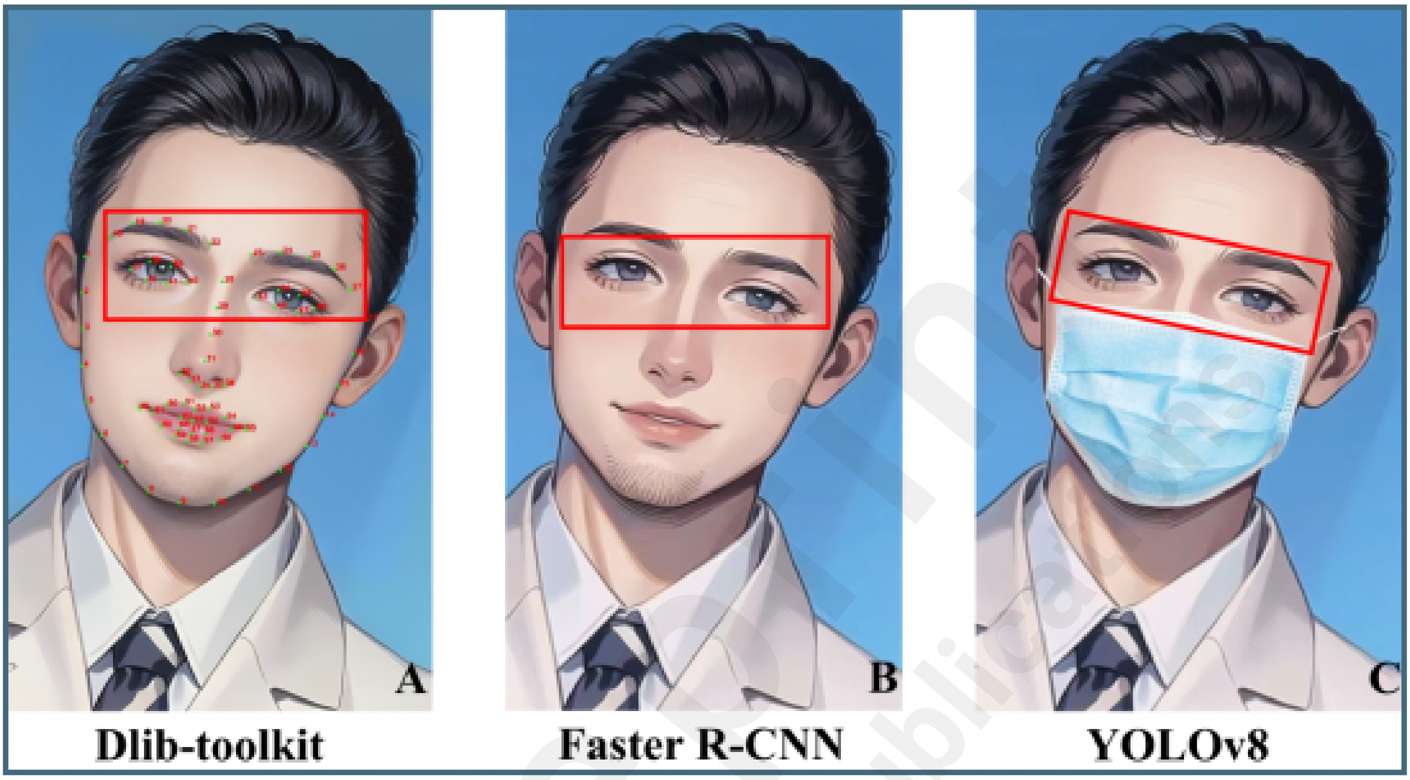
Performance Curves on 5-Fold Cross-Validation Figure 3 demonstrates the model's performance across five validation folds during training. Subfigure A shows precision curves, all approaching 1.0, indicating high accuracy in eye region detection. Subfigure B highlights recall curves, also nearing 1.0, reflecting robust identification of true positives. Subfigure C presents mAP50, confirming the model's precision across all folds, while Subfigure D illustrates mAP95, depicting consistent improvement at stricter IoU thresholds.



Head Tilt Correction and Ocular Region Cropping in Strabismus Patient Images Figure 4 illustrates the model's detection, cropping, and head tilt correction across two patient photographs. In Figure 4.A, the Ocular Alignment Angle (OAA) is 0.457° and the Eye Region Bounding Box Area (ERBBA) is 0.054, while in Figure 4.B, the OAA is 2.382° and the ERBBA is 0.171. The model isolates the eye region and adjusts the image orientation to ensure horizontal eye alignment, offering user-adjustable cropping boundaries to accommodate varying clinical needs and enhance flexibility.



The Interface of the AI-Driven Management Platform for Strabismus Patient Care Figure 5.A presents the digital medical records interface, where patients and healthcare providers can access and manage comprehensive patient histories, including diagnoses, prescriptions, and eye examination results. Figure 5.B depicts the doctor-patient interaction module, allowing seamless communication through text, images, and video, facilitating effective consultation and follow-up care.



Multimedia Appendixes

The following are Supplementary Materials.
URL: http://asset.jmir.pub/assets/bfe2857a8973313b2513df5d6571c916.docx