



山东大学学报(医学版)

Journal of Shandong University(Health Sciences)

ISSN 1671-7554,CN 37-1390/R

《山东大学学报(医学版)》网络首发论文

题目: 青光眼影像人工智能深度学习研究现状与展望
作者: Carol Y.Cheung, 冉安然
收稿日期: 2020-09-04
网络首发日期: 2020-10-20
引用格式: Carol Y.Cheung, 冉安然. 青光眼影像人工智能深度学习研究现状与展望
[J/OL]. 山东大学学报(医学版).
<https://kns.cnki.net/kcms/detail/37.1390.R.20201019.1149.002.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

文章编号:1671-7554(2020)11-0000-00

DOI:10.6040/j.issn.1671-7554.0.2020.1249



Carol Y. Cheung, 香港中文大学眼科和视觉科学系助理教授, 博士研究生导师。主要研究领域是“眼部成像”, 其理论基础为眼睛是人类循环和神经系统的“窗口”。在香港和新加坡有十多年从事眼部成像研究的经验, 致力于开发和应用图像分析以及人工智能技术对糖尿病性视网膜病变、青光眼和阿尔茨海默氏病的研究。

目前从事多项研究, 包括开发和应用新型成像技术, 以及探索如何利用这些技术改善临床工作流程和公共卫生。为人们理解和认识主要眼病以及脑部疾病带来新的角度, 促进用先进的眼部成像技术进行更有针对性、有效的疾病筛查, 实现防盲和脑部疾病的早期检测。在国际索引的同行评审期刊发表 200 多篇有影响力的论文, 并撰写 10 本书的相关章节, 专注于与视网膜、脑部疾病有关的视网膜成像技术。

国际科学期刊和国际资助基金会的定期审稿人。亚太眼科影像学会秘书长, 亚太远视眼科学会理事会理事兼财务主管。海峡两岸医学与健康交流协会视网膜血管疾病委员会的名誉会员, 以及中国医学教育协会智能医学专业委员会智能眼科学组的理事会成员。

青光眼影像人工智能深度学习研究 现状与展望

Carol Y. Cheung, 冉安然

(香港中文大学眼科与视觉科学系, 香港 999077)

摘要:青光眼是一组异质性神经退行性疾病, 其特征是视网膜神经节细胞及其轴突逐渐消失, 现已成为全球不可逆性失明的主要原因。人工智能(AI)是由机器展示的智能, 而深度学习(DL)是其中一个基于深度神经网络的分支, 在医学成像领域取得了重大突破。在青光眼影像方面, 已有越来越多的研究将 DL 应用于眼底图像以及光学相干断层扫描(OCT), 以检测青光眼性视神经病变。有很好的结果显示, 将 DL 技术整合到影像中进行青光眼评估是高效、准确的, 这可能会解决当前实践和临床工作流程中的一些难题。但是, 未来进一步的研究对于解决现存挑战至关重要, 例如为不同研究之间的图像标记建立标准, 将“黑匣子”的学习过程进行可视化, 提高模型在未知数据集上的泛化能力, 开发基于 DL 的实际应用程序, 以及建立合理的临床工作流程, 进行前瞻性验证和成本效益分析等。这篇综述总结了 AI 应用于青光眼影像的最新研究现状, 并讨论了对临床的潜在影响和未来的研究方向。

关键词: 青光眼影像; 人工智能; 光学相干断层扫描; 眼底照相; 深度学习

中图分类号: R574

文献标志码: A

Artificial intelligence deep learning in glaucoma imaging: current progress and future prospect

CHEUNG Y. Carol, RAN Anran

(Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong SAR 999077, China)

收稿日期: 2020-09-04; 网络出版时间:

网络出版地址:

通讯作者: Carol Y. Cheung. E-mail: carolcheung@cuhk.edu.hk

Abstract: Glaucoma is a group of heterogeneous neurodegenerative diseases, characterized by the gradual loss of retinal ganglion cells and their axons, and has now become the major reason of irreversible blindness worldwide. Artificial intelligence (AI) is intelligence demonstrated by machines. Deep learning (DL) is a subset of AI based on deep neural networks, and it has made great breakthroughs in medical imaging. In glaucoma imaging, research interests have been increasing on applying DL in fundus photographs and optical coherence tomography (OCT) for glaucomatous optic neuropathy (GON) detection. Promising results show that the incorporation of DL technology in imaging for glaucoma assessment is efficient and accurate, which could potentially address some gaps in the current practice and clinical workflow. However, further research is crucial in tackling some existing challenges, such as setting a standard for ground truth labelling among different studies, visualizing the learning process in the “black box”, improving the model generalizability on unseen datasets, developing the DL-powered infrastructure for real-world implementation, establishing a practical clinical workflow, conducting prospective validation and cost-effectiveness analysis. This review summarizes recent studies on the application of AI on glaucoma imaging, discusses the potential clinical impact and future research directions.

Key words: Glaucoma imaging; Artificial intelligence; Optical coherence tomography; Fundus photography; Deep learning

青光眼是一组异质性神经退行性疾病,其特征是视网膜神经节细胞及其轴突逐渐消失^[1],现已成为全球不可逆性失明的主要原因。预计到 2040 年,青光眼患者人数将达到 1.118 亿^[2]。青光眼在视乳头和视网膜层的结构改变称为青光眼性视神经病变 (glaucomatous optic neuropathy, GON),包括视网膜神经纤维层 (retinal nerve fibre layer, RNFL),神经节细胞以及内部丛状层 (ganglion cell with inner plexiform layer, GCIPL) 厚度变薄,杯盘比增大,盘沿变窄等。同时,青光眼还会导致功能性损伤,如视野缺损等。该疾病早期通常无明显症状,周边视力也无明显下降,很多患者在疾病的晚期察觉症状并就医,而此时的视功能损伤已不可逆。因此,及早发现、及时治疗 and 定期随访对于预防青光眼致盲至关重要。然而,现阶段并无切实有效的青光眼早期筛查方案^[3-5]。

在 GON 的评估中,眼部影像起到重要作用。眼底照相是常用的成像方式之一,它记录了视乳头和视网膜的俯视图,临床医生随后对图像进行主观分析,例如是否存在杯盘比增大,RNFL 缺失等。其他技术,包括共聚焦扫描激光检眼镜 (confocal scanning laser ophthalmoscopy, CSLO),扫描激光偏振仪 (scanning laser polarimetry, SLP) 和光学相干断层扫描 (optical coherence tomography, OCT) 等,可以对视神经和视网膜结构进行客观、立体和定量的测量。现阶段,眼底照相和 OCT 是在青光眼评估中应用最为广泛的影像技术^[6]。

然而,通过眼底彩照评估 GON 受到临床医生

主观因素的影响,研究表明,不同专家评估结果一致性并不理想^[7]。而通过传统 OCT 报告对 GON 的判断需要将测得的 RNFL 厚度与 OCT 中内置的正常人数据库进行比较,但临床上没有用于诊断或排除青光眼的 RNFL 厚度特定阈值。此外,RNFL 厚度测量受多种因素影响,例如图像质量差或近视等^[8-10]。因此,青光眼影像,包括眼底图像和 OCT 的解读都需要大量有经验的青光眼专家或训练有素的技术员,这既费时又费力,很难在人群筛查中实现。

“人工智能” (artificial intelligence, AI) 最早是由 John McCarthy 于 1956 年提出的,其定义是通过机器模拟人类智能^[11]。机器学习 (machine learning, ML) 是 AI 的分支,它主要利用计算机系统编程来执行任务或预测结果^[12-13]。深度学习 (deep learning, DL) 是 ML 的一个子领域 (图 1)。而卷积神经网络 (convolutional neural network, CNN) 的引入为 DL 用于图像分类和模式识别的发展带来了重大突破^[14]。近年来,随着计算机技术,大数据采集和成像方式的进步,DL 在由影像学驱动的专业领域发挥着越来越重要的作用,例如放射科,皮肤科和眼科等^[15-20]。研究表明,基于 DL 的医学图像分析系统在疾病检测方面的准确性与临床医生或训练有素的人员相当甚至更好^[21-22]。一些研究还证明了将 DL 算法用于人群疾病筛查和检测的潜力和可行性^[23-24]。本文总结了 DL 应用于青光眼影像的最新研究进展以及未来的发展方向。

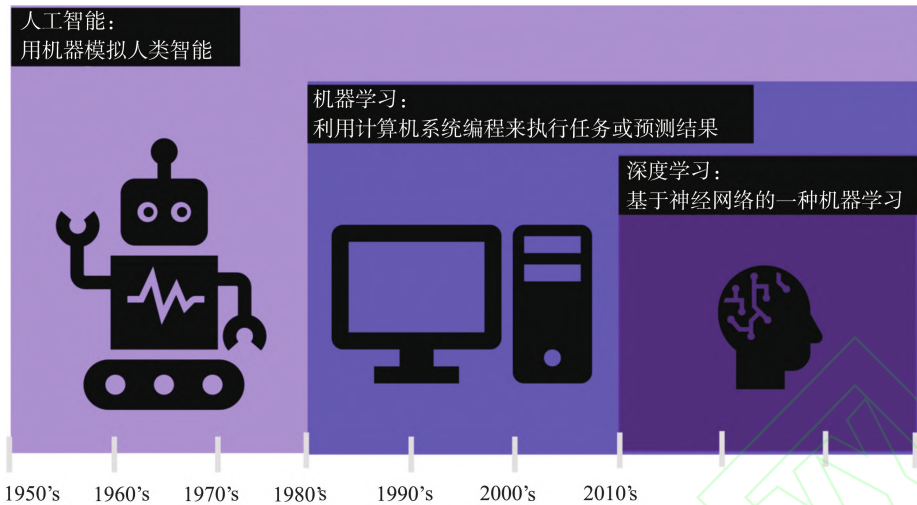


图 1 人工智能、机器学习以及深度学习
Fig.1 Artificial intelligence, machine learning and deep learning

1 深度学习方法概述

1.1 深度学习与传统机器学习相比的优势 传统 ML 图像识别主要依赖于人造特征的定向提取以获得良好的性能,其困难在于必须选择每张给定图像中的重要特征。由于医学图像中的病理变化在形状和大小上呈现出很大的个体差异,因此,准确创建病理特征并定向提取成为一项艰巨的任务。除此之外,传统 ML 模型存在过拟合可能性,泛化能力有限,即在某个数据集上训练好的 ML 算法应用于另一个未知数据集时可能性能变差。

与之对比,DL 主要通过构建具有很多隐层的神经网络结构,以大量数据进行训练,实现对图像中特征的自主提取和学习,而不需要创建任何域特定的输入功能。DL 是一个端到端的学习过程,以带标注的数据集作为输入,最终分类或者预测作为输出。学习过程中则利用多层非线性信息处理,将每一层网络结构的输出作为下一层网络结构的输入,通过每一层产生适当的特征层层递进,最终提升分类或预测准确性。因此,理论上,只要使用尽可能多样化的庞大数据集进行训练,DL 可以具有更高的泛化能力^[25-27]。

1.2 卷积神经网络在医学图像处理中的优势 CNN 是一种非全连接的神经网络结构,其将输入信息自动转换为多个抽象级别并学习不同影像中的特征,可实现更好的泛化^[27]。使用不同网络构建的 CNN 有些许结构差异,但总的来说,CNN 的标准模型包括输入层、卷积层、池化层(或次抽样层)和非线性层。卷积和池化层是 CNN 特征性结构,通常

组合为模块,模块彼此堆叠以形成深层网络。卷积层由多个特征平面构成,用于提取特征,而每个特征平面由神经元构成。它是权重的网格结构集与从输入中不同空间位置得出的相似网格结构之间的点积运算,此操作对于图像分析尤其有用。CNN 的引入,加速了 DL 在医学影像分析中的应用^[28-29]。现阶段,迁移学习、二维 CNN、三维 CNN、多任务 CNN 等最新方法已在医学图像的自动疾病检测中显示出可喜的成果^[30-31]。

1.3 建立深度学习模型的基本过程 表 1 及图 2 总结描绘了 DL 模型开发(即训练集和调试集)以及网络性能的评估(即内部和外部测试集)所需的基本数据集。通常,训练集是一个较大数据集,用于让网络学习和自主提取图像特征。而调试集是一个较小数据集,用于实时监控网络性能并根据结果调整架构。如果网络在训练集表现良好,但在调试集上的拟合效果较差,则存在过度拟合问题。此时应调整网络参数或改变某些架构,并观察学习曲线以找到最佳的训练停止点,避免过度拟合的问题^[32]。训练和调试完成后,使用另一个独立的内部测试数据集来检验最终性能。这三个数据集通常是根据特定比例从同一个大数据集分离出来。而为了进一步验证模型在未知数据集上的性能以及其泛化能力,需要其他独立的数据集作为外部验证。如果 DL 模型在包括内部和外部所有验证数据集中都具有良好性能,则意味着该模型具有好的泛化能力。基于 DL 的疾病检测模型需要用准确标注的大型数据集进行训练,建立合适的网络架构,并具有良好的泛化能力,以最终应用于不同的临床中心^[33]。

表 1 深度学习模型开发与性能评估的基本数据设置

Tab.1 The basic data sets for the development and performance evaluation of deep learning models

术语	说明
训练集	与调试集和内部验证集来自同一个大的数据集,是基于一定比例划分且互不重合,用于建立深度学习模型。研究者可构建具有不同参数甚至不同网络架构的深度学习模型来学习训练集中青光眼性视神经改变的相关特征。
调试集	与训练集和内部验证集来自同一个大的数据集,是基于一定比例划分且互不重合,用于实时评估不同模型在训练期间的性能。研究者可根据实时观察的学习曲线进行参数调节,确定训练停止节点,并最终选择最佳深度学习模型。
内部测试集	与训练集和调试集来自同一个大的数据集,是基于一定比例划分且互不重合,用于评估所选的最佳深度学习模型在该未知数据集上的性能。临床研究中,研究者需汇报模型在内部测试集上的性能,通常包括受试者工作特征曲线下面积,灵敏度,特异性,以及准确性等。
外部测试集	一个或多个独立的数据集,与训练集、测试集和内部验证集来自不同数据集,用于评估上述深度学习模型在其他未知数据集上的性能,验证其临床通用性。研究者同样需汇报模受试者工作特征曲线下面积,灵敏度,特异性,以及准确性等。当深度学习模型在内部以及外部测试集都具有较好性能时,即证明该模型有良好的通用性。
K-倍交叉验证	主要用于数据量有限时。将整个大的数据集分为 k 个相等的小数据集,其中 (k-1) 个合并为训练集用于模型开发,剩下的一个作为测试集评估模型性能。该训练过程重复 k 次,每次选择其中一个小数据集用作测试集。研究者应汇报所有测试集的平均性能,通常为 $\bar{x}\pm s$ 。
留一法交叉验证	主要用于数据量非常有限时,是一种特殊类型的 K-倍交叉验证,此时 K 等于数据总数(即标注图像的总数)。仅留下一张图像来测试模型性能,其余图像全部用于训练。该训练过程同样重复 k 次,研究者应汇报所有测试集的平均性能,通常为 $\bar{x}\pm s$ 。

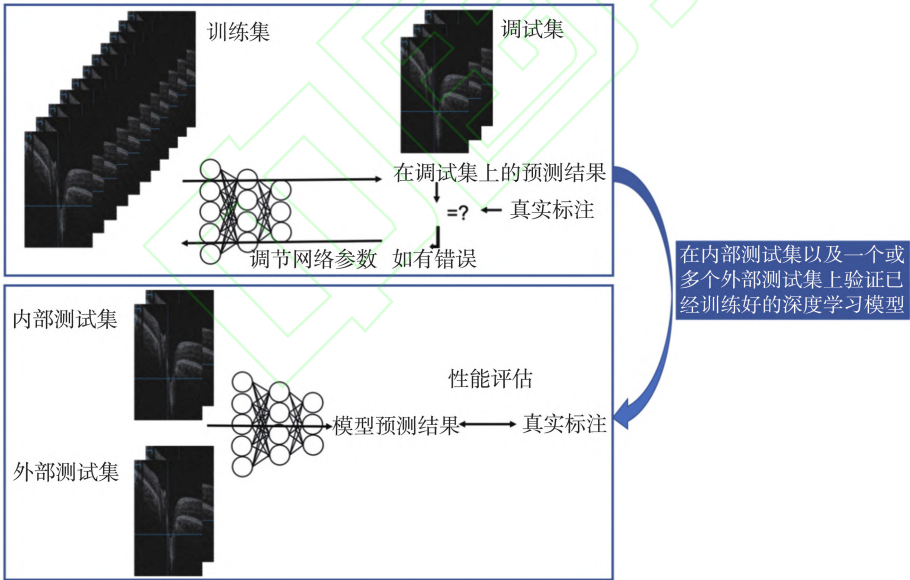


图 2 深度学习模型的开发与性能评估基础步骤

Fig.2 The basic steps for the development and performance evaluation of deep learning models

2 深度学习在青光眼眼底图像的应用

前文提到,杯盘比增大和 RNFL 厚度变薄是 GON 的主要特征,因此,DL 应用于青光眼眼底图像主要有两种方式:①从眼底图像中分割视杯和视盘,得出较为准确的杯盘比,或直接对整体图像分类是否存在青光眼性视神经损害;②“机器到机器”的方法,即从眼底照片预测 OCT 测量值,例如 RNFL 厚

度,Bruch’s 膜开口距离与最小盘沿宽度 (bruch’s membrane opening-minimum rim width, BMO-MRW) 等。

2.1 基于眼底图像的分割或分类深度学习模型

研究已证实,应用 DL 在眼底图像上精准分割视杯和视盘的可行性^[34-37],但仅有杯盘比并不能全面反映青光眼性视神经改变,应评估整张眼底图像以判断是否存在 GON。几项研究证明了 DL 有潜力应用于眼底图像以检测 GON。例如,Li 等^[38]采用了

48116 张眼底图像训练并验证了 DL 模型,用于检测需转诊的 GON。该项研究中有 21 位眼科医生对所有数据进行标注,确定同一张图片的最终标注需要至少 3 位医师达成一致,这在很大程度保证了数据标注的准确性。而该 DL 模型在内部验证集的受试者工作特征曲线下面积(area under the receiver operating characteristic curve, AUROC)为 0.986,灵敏度和特异性分别为 95.6%和 92.0%。该研究还发现生理性大视杯或高度/病理性近视会导致 DL 模型错判。Liu 等^[39]的另一项研究使用了 269601 张眼底照片进行 DL 模型的训练和验证,而在内部验证中,该 DL 模型的 AUROC 为 0.996,灵敏度和特异性分别为 96.2%和 97.7%。值得一提的是,该 DL 模型在拥有不同图像质量的多中心外部验证集上都表现了良好的性能,且热力图显示,模型认为有助于判断 GON 的区域与 RNFL 走形一致。

2.2 基于眼底图像的“机器到机器”深度学习模型

尽管以上研究证明,DL 模型在眼底图像上检测 GON 的准确性很高,其仍存在一定局限性。首先,数据标注的参考标准需要基于专家对眼底照片的主观判读;其次,这些 DL 算法仅能从视神经和视网膜俯视图提取信息,无法考虑到内部或整个垂直结构中青光眼病变相关的信息。因此有一些深度学习模型使用“机器到机器”方法从眼底图像预测相应的 OCT 测量值,以获得更客观和定量的估计。该方法即是将一张眼底图像对应的 OCT 测量的参数作为参考标准来训练 DL 模型,而验证时,则从未知眼底图像中预测 OCT 参数,从而与真实数值对比。根

据 Felipe 等^[40]研究结果,从 DL 获得的预测 RNFL 厚度与实际测量值显示出很强的相关性,并且预测的平均绝对误差(MAE)约为 7 μm 。使用这些预测值区分青光眼和正常眼的 AUROC(0.944)与使用实际 RNFL 厚度值的 AUROC(0.940)几乎相同。热力图显示,对于 DL 算法而言,靠近视盘和相邻的 RNFL 区域是用于预测的重要部位。

Thompson 等^[41]结果表明,BMO-MRW 预测值和观测值之间存在很强的相关性,MAE 为 27.8 μm 。而使用预测值和真实值区分青光眼和正常眼的 AUROC 分别为 0.945 和 0.933,热力图显示 DL 模型的关注区域也是视盘。“机器到机器”的 DL 模型可应用于没有 OCT 仪器的边远区域或非眼科诊所,能通过预测 OCT 测量值的方式,提供定量和客观参考,提高眼底图像检测 GON 的准确性。

3 深度学习在青光眼 OCT 图像的应用

虽然眼底照相更便捷经济,可能适用于资源匮乏的区域或初级诊疗中心,OCT 可进行三维立体成像,并测得客观数据,对于青光眼的检测更准确,近十多年临床上应用得也越来越多。目前,已经有研究实现了用 OCT 相关数据或者图像来训练 DL 模型,以检测 GON。这些 DL 模型主要分三类,具有不同类型的输入:①传统 OCT 报告中的测量参数,厚度图,偏差图和俯视图;②未分割的二维扫描图;③未分割的三维立体扫描图。见图 3。

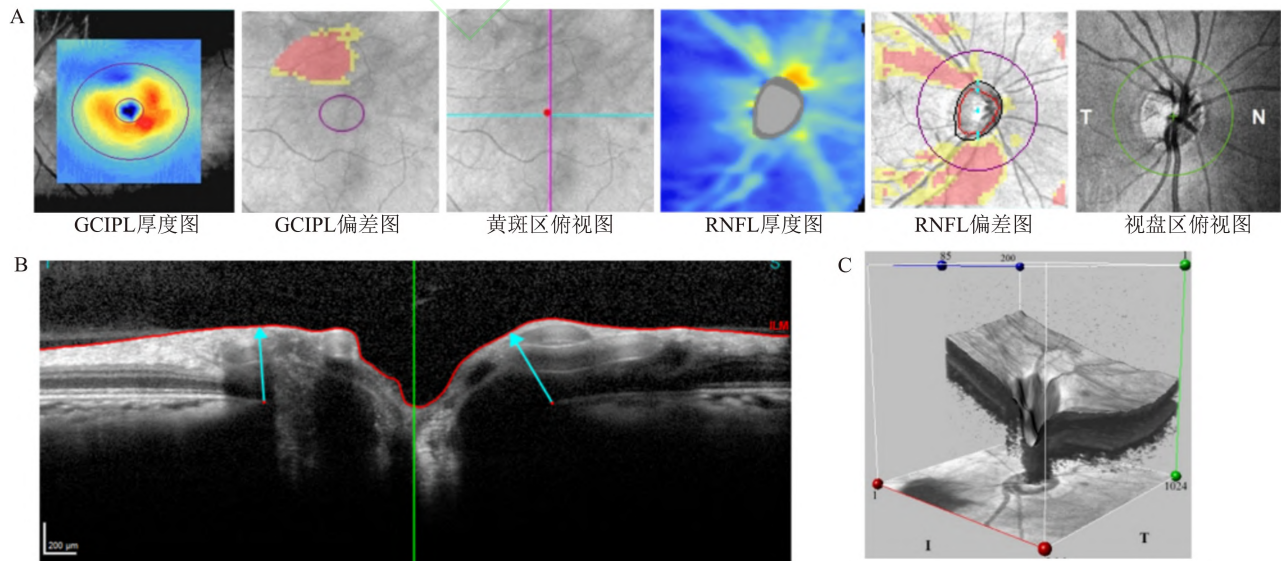


图 3 基于光学相干断层扫描(optical coherence tomography, OCT)不同类型输入的深度学习模型
A:传统 OCT 报告;B:视盘区未分割二维扫描图;C:视盘区立体扫描图。
Fig.3 The deep learning models based on different input of OCT
A: Traditional OCT reports; B: Optic disc segmentation-free 2D scans; C: Optic disc volumetric scans

3.1 基于 OCT 报告中测量参数或图像的深度学习模型 从传统的 OCT 报告中,可以提取测量的定量参数,比如 RNFL 和 GCIPL 厚度等,以及厚度图、偏差图和俯视图等图像,来作为 DL 模型的输入。

Asaoka 等^[42]使用 8×8 网格黄斑区 RNFL 和 GCIPL 厚度作为输入,构建了 DL 模型以诊断早期青光眼。为了防止因数据量小出现过拟合,他们还进行了迁移学习。这是一种适用于小数据量的方法,它将已经在海量自然图像上完成预训练的 DL 模型作为另一个模型训练的起点,使得另一个模型所需训练数据大大减少^[43]。他们发现使用了迁移学习的 DL 模型大大提升了性能,而且优于其他两种传统的机器学习方法。

Lee 等^[44]进行的另一项研究使用了 OCT 视盘和黄斑区图像作为输入,分别是:①GCIPL 厚度图;②GCIPL 偏差图;③RNFL 厚度图;④RNFL 偏差图。通过上述图像训练了 4 个独立的 DL 模型,随后将 4 种图像的集成特征用来建立一个“整体训练模型”。该模型取得了最佳性能,证明 4 种图像中提取的信息集合在一起更有助于区分青光眼和正常眼。

3.2 基于 OCT 未分割二维扫描图训练的深度学习模型 前述研究基于 OCT 测得的 RNFL 或 GCIPL 厚度,该厚度可能会受到分割误差的影响。Thompson 等^[45]开发了一种基于 OCT 未分割二维扫描图的 DL 模型,避免了分割误差。他们发现该方法在检测青光眼结构变化方面比使用传统 RNFL 厚度性能更好。

Wang 等^[46]也使用了 OCT 二维扫描图进行 DL 模型训练,并进一步加入了配对的视野相关参数作为附加信息。他们应用了半监督学习和多任务技术等先进的技术,发现这些方法有助于提高 DL 模型检测青光眼的性能。

3.3 基于 OCT 未分割三维立体扫描图的深度学习模型 Maetschke 等^[47]训练了一个 DL 模型,直接将未分割的 OCT 立体扫描图像分类为健康或青光眼。这项研究还生成了热力图来展示 DL 模型认为对青光眼检测很重要的区域。他们发现视盘、视杯、盘沿以及筛板区域与 DL 模型识别是否有青光眼显著相关。

Ran 等^[48]在更大的数据集上开发并验证了三维 DL 模型,该数据集包括 6921 张视盘区三维立体扫描图。在这项研究中,使用立体扫描图训练的三维 DL 模型明显优于使用二维俯视图训练的二维 DL 模型。该三维 DL 模型的性能也可与两名拥有十多年青光眼临床诊疗经验的专家相媲美。生成的热

力图显示,三维 DL 模型检测 GON 具有最大判别力的区域与眼科医生在临床中通常观察的区域相似。这项研究的优势之一是拥有覆盖多个地区且包含不同种族的三个独立验证数据集,结果显示,该模型在所有验证集中具有普遍良好的性能,证明其在完全未知的新数据集上也具有泛化能力。

总而言之,利用 OCT 相关数据或图像训练的 DL 模型在青光眼检测方面都表现出良好的性能。而不同的输入形式存在不同的利弊:①OCT 测量的参数训练 DL 模型需要的计算能力较少,但图像伪影和分割误差可能导致参数本身不准确;②厚度图和偏差图可以在某种程度上表现 RNFL 或 GCIPL 厚度的分布,但仍不能直接显示结构变化;③视盘的二维扫描图可以直接获得视盘周围区域的结构信息,但无法全面展示其他部位的青光眼特征性变化;④而三维立体扫描图可以充分利用所有与青光眼病变相关的信息,但是训练过程需要更多的图形处理器(graphic processing unit, GPU)内存,耗费更长的时间。

4 讨 论

尽管 DL 模型具有合理的效率和准确性,能潜在地应用于青光眼影像,为青光眼的检测或筛查节省时间,减轻人力负担,但目前仍存在一定局限:①由于 DL 模型自动提取和学习特征,因此很难展示和解释其确切的学习过程和机制(即“黑匣子”问题)。尽管已有研究生成了热力图来证明 AI 所关注的区域,DL 模型得出预测值的原理仍是一个未知过程。②尚不能完全确保现有 DL 模型在未知数据集中的泛化能力。③其他结构或病理变化(例如与高度/病理性近视有关的变化)可能会影响 DL 模型对青光眼评估的性能。④所有的 DL 模型都只基于回顾性研究的数据集训练及验证,尚未知其临床实用性。因此,未来仍需更进一步的研究以提高 DL 模型在青光眼影像中的应用。从临床角度来看,建立一个具有准确标注的大型数据库至关重要,该数据库必须尽可能包含来自全球多个中心的数据,以保证数据的多样性,减少由数据相关因素(例如种族/种族、疾病严重程度、成像技术、成像仪器的差异)引起的偏差。而为了收集大型数据集以进行 DL 系统开发,需要针对数据收集、共享、标注、参考标准和 DL 模型结果报告建立一套标准化的准则。

从技术角度来看,DL 方法本身的主要限制之一是需要大量高质量的准确标注数据进行训练和验

证,以获得临床能接受的诊断性能。然而,训练庞大数据集需要耗费大量 GPU 内存,因此,模型压缩或输入预处理等技术需要提升,以减少 GPU 内存。而且,如上所述,在现实环境中获得大规模高质量的医学图像仍然是一个很大的挑战,技术上应利用更先进的数据增强方法,例如编程模拟病变以融入正常图像^[49],或将真实病变融入正常或异常图像^[50]的其他位置,以及生成对抗网络(generative adversarial networks, GAN)^[51]。此外,现有的 DL 模型多采用监督学习,因此,缺乏高质量、可靠的标注也会降低 DL 训练的有效性。为了从技术上解决此问题,未来可以采用半监督或无监督学习的方法,以进一步完善现有的 DL 模型并解决标注数据有限的问题。这些方法已经在脑肿瘤分割^[52],健康物联网服务中的临床决策支持系统^[53]以及 OCT 图像分析中有所研究^[54]。此外,尽管一些 DL 模型在内部验证中表现良好,但由于设备和成像技术的多样性,眼部生理解剖结构的差异性,数据分布的不平衡性,导致在来源不同临床中心的数据集的外部验证性能变差,因此,未来还应探索如何使 DL 模型获得更高的泛化能力以应用于不同的临床中心。此外,现有大部分深度学习模型主要应用于青光眼的眼后节影像,而眼前节影像(如前节 OCT)^[55],或青光眼影像与视野检查的结合^[56],以及图像分割与病灶检测技术结合^[57]等研究方向也具有相当的应用前景。未来应致力于研发深度学习系统对青光眼进行综合性检测。

以上临床和技术问题解决后,训练好的 DL 模型可以进一步工程化,建立一套 AI 青光眼影像分析系统,应用在不同级别诊疗中心,以辅助青光眼检测甚至人群筛查的临床工作流程。例如,患者在临

床中心进行青光眼影像检查,AI 系统将自动进行图像质量控制,图像质量合格后,系统将生成是否有 GON 以及是否转诊至青光眼专科进行下一步诊疗的报告。见图 4。在未来实际应用于临床之前,有以下研究方向:①首要保障患者的数据隐私和安全(包括道德和法律方面)。例如使用用户名以及密码身份验证防止未授权访问,以及保存患者资料于本地服务器等。此外,临床医生应熟悉 AI 青光眼影像分析系统的工作原理,以保护患者的安全。②应该建立和开发全自动系统(例如,从影像机器提取图像并将其发送到 GPU 服务器)来促进工作流程的有效性,以进一步实现临床应用。③应收集临床实时数据,进行前瞻性研究,以分析在当前临床工作流程中应用 AI 青光眼影像分析系统的成本效益,效率和准确性。此外,还应根据前瞻性研究结果随时改进 AI 模型,以长期监控 AI 系统的泛化能力。④该 AI 分析系统输出结果的可解释性对于临床上确保患者安全也至关重要。未来应有进一步的研究以更好地对“黑匣子”进行可视化,从而提高 AI 系统的可解释性。⑤还应建立合理的临床工作流程以及提高临床工作者对 AI 系统的认识度和接受度。尽管实现 AI 全自动青光眼影像分析是最理想的选择,但因为现实世界中的许多情况不能在 AI 模型的训练和验证中完全覆盖,基于安全考虑,临床医师仍应发挥重要的监督作用。因此,远程青光眼诊疗可能是未来一种可行的方案^[58],比如在筛查环境中应用 AI 系统以在青光眼影像中进行疾病自动检测,而临床医师可以远程监督,确保转诊决策的准确性。

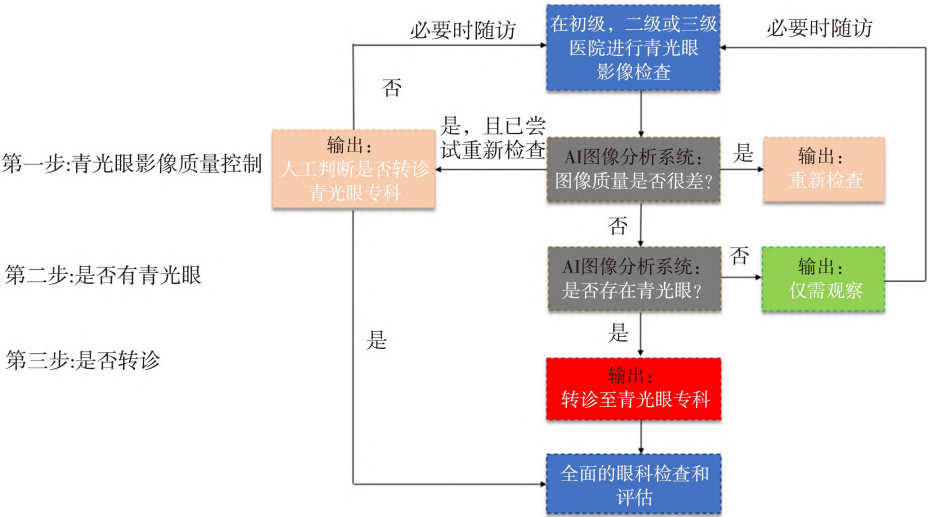


图 4 人工智能青光眼影像分析系统在临床中辅助青光眼检测或筛查的工作流程
Fig.4 The workflow of artificial intelligence-based glaucoma imaging analysis system for glaucoma detection or screening in clinics

综上所述,研究已证实人工智能,尤其是深度学习应用于在青光眼影像分析,从而辅助青光眼检测是有效、准确和有前景的。未来需进一步研究并解决现有的挑战,例如为不同研究之间的图像标记建立标准,将 AI“黑匣子”可视化,提高模型在未知数据集中的泛化能力,开发基于 DL 的实际应用软件,建立合理的临床工作流程,进行前瞻性研究和成本效益分析等。

参考文献:

- [1] Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review [J]. *JAMA*, 2014, 311(18): 1901-1911.
- [2] Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis [J]. *Ophthalmology*, 2014, 121(11): 2081-2090.
- [3] Weinreb RN, Leung CKS, Crowston JG, et al. Primary open-angle glaucoma [J]. *Nat Rev Dis Primers*, 2016, 2: 16067. doi: 10.1016/S0140-6736(04)16257-0.
- [4] Jonas JB, Aung T, Bourne RR, et al. Glaucoma [J]. *Lancet*, 2017, 390(10108): 2183-2193.
- [5] Momont AC, Mills RP. Glaucoma screening: current perspectives and future directions [J]. *Semin Ophthalmol*, 2013, 28(3): 185-190.
- [6] Schuman JS. Detection and diagnosis of glaucoma: ocular imaging [J]. *Invest Ophthalmol Vis Sci*, 2012, 53(5): 2488-2490.
- [7] Rossetto JD, Melo LAS, Campos M, et al. Agreement on the evaluation of glaucomatous optic nerve head findings by ophthalmology residents and a glaucoma specialist [J]. *Clin Ophthalmol*, 2017, 11: 1281-1284. doi: 10.2147/OPHTH.S140225.
- [8] Cheung CY, Leung CK, Lin D, et al. Relationship between retinal nerve fiber layer measurement and signal strength in optical coherence tomography [J]. *Ophthalmology*, 2008, 115(8): 1347-1351.
- [9] Cheung CY, Chan N, Leung CK. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: impact of signal strength on analysis of the RNFL map [J]. *Asia Pac J Ophthalmol (Phila)*, 2012, 1(1): 19-23.
- [10] Biswas S, Lin C, Leung CK. Evaluation of a myopic normative database for analysis of retinal nerve fiber layer thickness [J]. *JAMA Ophthalmol*, 2016, 134(9): 1032-1039.
- [11] Andresen SL. John McCarthy: father of AI [J]. *IEEE Intell Syst*, 2002, 17(5): 84-85.
- [12] Simon A DMS, Venkatesan S, Babu DRR. An overview of machine learning and its applications [J]. *International Journal of Electrical Sciences & Engineering*, 2015, 1(1): 22-24.
- [13] Shinde PP, Shah S. A review of machine learning and deep learning applications [J]. 2018 Fourth International Conference on Computing Communication Control and Automation (Iccubea), Pune, India, 2018: 1-8. doi: 10.1109/ICCUBEA.2018.8697857.
- [14] Aggarwal CC. Training deep neural networks[M]//*Neural Networks and Deep Learning*. Cham: Springer International Publishing, 2018: 105-167. doi: 10.1007/978-3-319-94463-0_3.
- [15] Becker AS, Mueller M, Stoffel E, et al. Classification of breast cancer from ultrasound imaging using a generic deep learning analysis software: a pilot study [J]. *Br J Radiol*, 2017: 20170576. doi: 10.1259/bjr.20170576.
- [16] Walsh SLF, Calandriello L, Silva M, et al. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study [J]. *Lancet Respir Med*, 2018, 6(11): 837-845.
- [17] Brinker TJ, Hekler A, Enk A, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task [J]. *Eur J Cancer*, 2019, 111: 148-154. doi: 10.1016/j.ejca.2019.02.005.
- [18] de Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease [J]. *Nat Med*, 2018, 24(9): 1342-1350.
- [19] Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes [J]. *JAMA*, 2017, 318(22): 2211-2223.
- [20] Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning [J]. *Cell*, 2018, 172(5): 1122-1131.
- [21] Hekler A, Utikal JS, Enk AH, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images [J]. *Eur J Cancer*, 2019, 118: 91-96. doi: 10.1016/j.ejca.2019.06.012.
- [22] Maron RC, Weichenthal M, Utikal JS, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks [J]. *Eur J Cancer*, 2019, 119: 57-65. doi: 10.1016/j.ejca.2019.06.013.
- [23] Dascalu A, David EO. Skin cancer detection by deep learning and sound analysis algorithms: A prospective clinical study of an elementary dermoscope [J]. *EBio-Medicine*, 2019, 43: 107-113. doi: 10.1016/j.ebiom.

- 2019.04.055.
- [24] Al-Antari MA, Al-Masni MA, Kim TS. Deep learning computer-aided diagnosis for breast Lesion in digital mammogram [J]. *Adv Exp Med Biol*, 2020, 1213: 59-72. doi: 10.1007/978-3-030-33128-3_4.
 - [25] O'Mahony N, Campbell S, Carvalho A, et al. Deep learning vs. traditional computer vision [J]. *Adv Intell Syst*, 2020, 943: 128-144. <https://arxiv.org/pdf/1910.13796.pdf>.
 - [26] Wang JJ, Ma YL, Zhang LB, et al. Deep learning for smart manufacturing: Methods and applications [J]. *J Manuf Syst*, 2018, 48: 144-156.
 - [27] Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review [J]. *Neural Comput*, 2017, 29(9): 2352-449.
 - [28] Aggarwal CC. Convolutional Neural Networks. In: *Neural Networks and Deep Learning* [J]. Springer Cham, 2018. doi: 10.1007/978-3-319-94463-0_8.
 - [29] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks [J]. *Commun ACM*, 2017, 60(6): 84-90.
 - [30] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
 - [31] Benuwa B, Zhan YZ, Ghansah B, et al. A review of deep machine learning [J]. *Int J Eng Res Afr*, 2016, 24: 124-36. doi: 10.4028/www.scientific.net/JERA.24.124.
 - [32] Ying X. An overview of overfitting and its solutions [J]. 2018 International Conference on Computer Information Science and Application Technology, 2019, 1168. doi: 10.1088/1742-6596/1168/2/020222.
 - [33] Aggarwal CC. Teaching Deep Learners to Generalize. In: *Neural Networks and Deep Learning*. Springer, Cham; 2018. doi: 10.1007/978-3-319-94463-0_4.
 - [34] Zilly J, Buhmann JM, Mahapatra D. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation [J]. *Comput Med Imag Grap*, 2017, 55: 28-41. doi: 10.1016/j.compmedimag.2016.07.012.
 - [35] Yu S, Xiao D, Frost S, et al. Robust optic disc and cup segmentation with deep learning for Glaucoma detection [J]. *Comput Med Imaging Graph*, 2019, 74: 61-71. doi:10.1016/j.compmedimag.
 - [36] Sevastopolsky A. Optic disc and cup segmentation methods for Glaucoma detection with modification of U-Net convolutional neural network [J]. *Pattern Recognit Image Anal*, 2017, 27(3): 618-624.
 - [37] Fu H, Cheng J, Xu Y, et al. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation [J]. *IEEE Trans Med Imaging*, 2018, 37(7): 1597-1605.
 - [38] Li Z, He Y, Keel S, et al. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs [J]. *Ophthalmology*, 2018, 125(8): 1199-1206.
 - [39] Liu H, Li L, Michael Wormstone I, et al. Development and validation of a deep learning system to detect glaucomatous optic neuropathy using fundus photographs [J]. *JAMA Ophthalmol*, 2019, 137(12): 1353-1360.
 - [40] Medeiros FA, Jammal AA, Thompson AC. From machine to machine: an oct-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs [J]. *Ophthalmology*, 2019, 126(4): 513-521.
 - [41] Thompson AC, Jammal AA, Medeiros FA. A deep learning algorithm to quantify neuroretinal rim loss from optic disc photographs [J]. *Am J Ophthalmol*, 2019, 201: 9-18. doi: 10.1016/j.ajo.
 - [42] Asaoka R, Murata H, Hirasawa K, et al. Using Deep Learning and transfer learning to accurately diagnose early-onset glaucoma from macular optical coherence tomography images [J]. *Am J Ophthalmol*, 2019, 198: 136-145. doi: 10.1016/j.ajo.
 - [43] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? [J]. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2014(2): 3320-3328. doi: 10.5555/2969033.2969197.
 - [44] Lee J, Kim YK, Park KH, et al. Diagnosing glaucoma with spectral-domain optical coherence tomography using deep learning classifier [J]. *Journal of Glaucoma*, 2020, 29(4): 287-294.
 - [45] Thompson AC, Jammal AA, Berchuck SI, et al. Assessment of a segmentation-free deep learning algorithm for diagnosing glaucoma from optical coherence tomography scans [J]. *JAMA Ophthalmol*, 2020, 138(4): 333-339.
 - [46] Wang X, Chen H, Ran AR, et al. Towards multi-center Glaucoma OCT image screening with semi-supervised joint structure and function multi-task learning [J]. *Med Image Anal*, 2020, 63: 101695. doi: 10.1016/j.media.
 - [47] Maetschke S, Antony B, Ishikawa H, et al. A feature agnostic approach for glaucoma detection in OCT volumes [J]. *PLoS One*, 2019, 14(7): e0219126.
 - [48] Ran AR, Cheung CY, Wang X, et al. Detection of glaucomatous optic neuropathy with spectral-domain optical coherence tomography: a retrospective training and validation deep-learning analysis [J]. *Lancet Digit Heal*, 2019, 1(4): 172-182.
 - [49] Badano A, Graff CG, Badal A, et al. Evaluation of dig-

- ital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial [J]. *Jama Netw Open*, 2018, 1(7): e185474.
- [50] Cha KH, Petrick N, Pezeshk A, et al. Reducing overfitting of a deep learning breast mass detection algorithm in mammography using synthetic images [J]. *Proc SPIE* 10950, Medical Imaging 2019: Computer-Aided Diagnosis, 1095004. doi: 10.1117/12.2512604.
- [51] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets [J]. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2014 (2): 2672-2680. doi: 10.5555/2969033.2969125.
- [52] Sun Y, Zhou CF, Fu YW, et al. Parasitic Gan for Semi-Supervised Brain Tumor Segmentation [J]. *Ieee Image Proc*, 2019: 1535-1539. doi: 10.1109/ICIP.2019.8803073
- [53] Yang Y, Nan F, Yang P, et al. GAN-Based Semi-Supervised Learning Approach for Clinical Decision Support in Health-IoT Platform [J]. *IEEE Access*, 2019, 7: 8048-8057. doi: 10.1109/ACCESS.
- [54] Wang X, Tang FY, Chen H, et al. UD-MIL: Uncertainty-driven deep multiple instance learning for OCT image classification [J]. *IEEE J Biomed Health Inform*, 2020; 1. doi: 10.1109/jbhi.
- [55] Fu HZ, Li F, Sun X, et al. AGE challenge- angle closure glaucoma evaluation in anterior segment optical coherence tomography [J]. *Medical Image Analysis*, 2020 (66): 101798. doi: 10.1016/j.media.2020.101798.
- [56] Mariottoni E, Datta S, Dov D, et al. Artificial intelligence mapping of structure to function in glaucoma [J]. 2020, 9(2): 19.
- [57] Yu S, Xiao D, Frost Shaun, et al. Robust optic disc and cup segmentation with deep learning for glaucoma detection [J]. *Computerized Medical Imaging Analysis Graphics*, 2019 (74): 61-71. doi: 10.1016/j.comp-medimag.
- [58] Tan NYQ, Friedman DS, Stalmans I, et al. Glaucoma screening: where are we and where do we need to go? [J]. *Curr Opin Ophthalmol*, 2020, 31(2): 91-100.

(编辑:徐苗蓁)