·专家论坛·

多模态深度学习及其在眼科人工智能的应用展望

李锡荣[1,2,3]

中国人民大学 [1] 数据工程与知识工程教育部重点实验室 [2] 信息学院人工智能与媒体计算实验室，北京 100872
[3] 北京致远慧图科技有限公司人工智能实验室，北京 100872

电话：010-82504345，E-mail：xirong@ruc.edu.cn

【摘要】深度学习的强学习能力和高易用性使其成为当前主流机器学习算法和医学人工智能的核心技术。鉴于医学影像在健康筛查、疾病诊断、精准治疗、预后评估等诸多任务中的关键作用，用于医学影像结构分析与语义理解的深度学习正成为重要的交叉学科研究方向。在临床场景中，医生为了实现更精准的诊断，往往需要同时参考不同类型、不同模态的影像样本进行综合分析和判断。本文介绍面向此类场景的多模态深度学习的基本概念和工作原理，结合具体案例分析多模态深度学习在眼科领域的研究进展、应用情况及技术挑战，并对该技术的应用前景作出展望。

【关键词】多模态深度学习；眼科；人工智能；辅助诊断
【中图分类号】R77；TP18　　【文献标志码】A　　【文章编号】1674-9081(2021)05-0602-06
DOI：10.12290/xhyxzz.2021-0500

Multi-modal Deep Learning and Its Applications in Ophthalmic Artificial Intelligence

LI Xirong[1,2,3]

[1]Key Laboratory of DEKE, [2]AIMC Lab, School of Information, Renmin University of China, Beijing 100872, China
[3]Vistel AI Lab, Visionary Intelligence Ltd., Beijing 100872, China

Tel: 86-10-82504345, E-mail: xirong@ruc.edu.cn

【**Abstract**】Deep learning, for its powerful learning capability and high usability, has been a prevalent algorithm of machine learning and a core technique for artificial intelligence (AI) in medicine and healthcare. Due to the importance of medical imaging in many tasks such as health screening, disease diagnosis, precise treatment, and prognosis prediction, deep learning of structural analysis and semantic understanding for medical images is becoming an important interdisciplinary research direction. In clinical scenarios, in order to achieve a more accurate diagnosis, doctors need to simultaneously refer to multiple modalities of medical imaging for a comprehensive analysis and judgment. This article introduced the basic concepts and working principles of multimodal deep learning in such scenarios, reviewed recent research progress on applying multi-modal deep learning in both generic medical fields and ophthalmology, and discussed technical challenges and also envision potential applications of multi-modal deep learning in AI-assisted ophthalmology.

【**Key words**】multi-modal deep learning; ophthalmology; artificial intelligence; assisted diagnosis

基金项目：北京市自然科学基金面上项目（4202033）；北京市自然科学基金-海淀原始创新联合基金（19L2062）；北京市科委医药协同创新专项课题（Z191100007719002）
引用本文：李锡荣. 多模态深度学习及其在眼科人工智能的应用展望［J］. 协和医学杂志，2021，12（5）：602-607. doi：10.12290/xhyxzz.2021-0500.

Funding: Beijing Natural Science Foundation (4202033); Beijing Natural Science Foundation Haidian Original Innovation-Joint Fund (19L2062); the Pharmaceutical Collaborative Innovation Research Project of Beijing Science and Technology Commission (Z191100007719002)

Med J PUMCH, 2021,12(5):602-607

以深度学习为代表的新一代人工智能 (artificial intelligence, AI) 技术对各行各业的影响是前所未有的。例如美国科学家利用 AI 辅助新型冠状病毒疫苗研发[1], 训练深度卷积网络根据咳嗽声音筛查新型冠状病毒肺炎患者[2]; 日本农民利用深度学习模型根据黄瓜品相对其进行自动分类[3], 等等。这种影响的形成, 与深度学习自身的技术特点密不可分。

深度学习是一种以深层人工神经网络为架构, 以原始数据为输入, 以目标任务为输出, 具备端到端 (end-to-end) 学习能力的机器学习算法[4-5]。相比传统机器学习算法, 深度学习具有强学习能力和高易用性的特殊优势。以图像分类任务为例, 传统方法分为特征提取 (feature extraction) 和分类器训练 (classifier training) 两个阶段。前者负责从原始图像样本中提取与当前指定任务相关的向量化的视觉特征, 而后者基于视觉特征和样本所对应的类别标签, 寻找最优分类决策边界。这两个阶段之间并不存在反馈机制。分类器训练只能在既定特征空间进行, 即使不同类别的样本在该特征空间缺乏区分性。与之相反, 深度学习将特征提取和分类器训练纳入一个神经网络框架中, 输入数据经过多层神经网络, 逐次提取表达能力更强的视觉特征, 最后经任务层给出分类结果。任务层获得的错分信息经后向传播 (back propagation) 反馈给特征层, 使其不断调整、优化特征提取过程, 从而实现特征提取和分类器训练的联合优化。值得进一步指出的是, 由于传统方法天然缺乏联合优化能力, 因此特征提取 [一些文献称之为特征工程 (feature engineering)[6]] 非常关键, 往往需要密集的领域知识和大量的经验式设计。相比之下, 深度学习的特征提取过程更为精简, 相同或相似的神经网络架构可用于解决传统意义上完全不同的两个任务 (如图像分类和文本分类)。

鉴于医学影像在健康筛查、疾病诊断、精准治疗、预后评估等诸多任务中的关键作用, 用于医学影像结构分析与语义理解的深度学习正成为重要的交叉学科研究方向。由于眼睛是全身唯一活体能够直接观察到血管和神经的部位, 关于该部位的多种类型医学影像如眼底彩照 (color fundus photography, CFP)、超广角眼底图像 (ultra-wide-field fundus images, UWF)、光学相干断层成像 (optical coherence tomography, OCT)、裂隙灯照片等 (图 1) 具有无创、非侵入、经济等优点, 因此发展眼科 AI 对于在不同年龄段开展大规模眼健康筛查具有重要意义。

以 CFP 为例, 眼科 AI 涉及结构分析 (左右眼识别、黄斑定位、视杯视盘分割、血管提取等) 和语义理解 (图像质量评估、眼底病灶分割、眼底疾病识别等) 两大类任务。近年来关于特定子任务的代表性研究案例逐年增多 (表 1)。例如, 谷歌 2016 年发表于 *JAMA* 的研究[7], 首次证实了利用深度卷积网络从单张后极部 CFP 中识别糖尿病视网膜病变 (diabetic retinopathy, DR) 的可行性。谷歌下属的 DeepMind 公司于 2018 年在 *Cell* 发文表明, 以 OCT 图像序列作为输入的 AI 模型在多个病种的转诊判断上, 有望达到临床专家的水平[9]。北京协和医院的新近研究证实, 基于单张 CFP, AI 模型在 10 余种常见眼底疾病的识别精度上已可媲美住院医师[19]。

上述眼科 AI 方向的工作均以单一类型影像 (如 CFP、OCT、UWF 等) 作为 AI 模型的输入。而在临床实践中, 医生为了实现更精准的诊断, 往往需同时参考不同类型、不同模态的影像样本进行综合分析、交叉验证和判断。以 CFP 和 OCT 为例, 考量二者成像部位的物理位置关系可以发现, CFP 反映的是视网

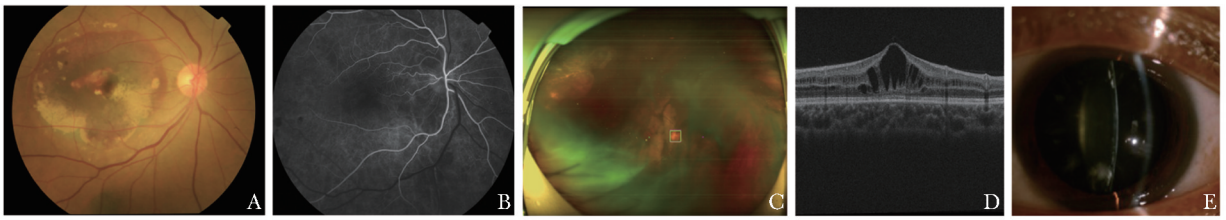

图 1 不同类型眼科影像示例

A. 眼底彩照; B. 荧光素眼底血管造影; C. 超广角眼底图像; D. 光学相干断层成像; E. 裂隙灯照片 (斜照法)

表 1 单模态深度学习在眼科领域的应用举例

年份（年）	研究者	任务	单模态输入
2016	Gulshan 等[7]	DR 转诊/非转诊分类	单张眼底彩照
2017	Burlina 等[8]	AMD 分级	单张眼底彩照
2018	Kermany 等[9]	多病种识别	OCT 图像序列
2018	Wei 等[10]	激光斑检测	单张眼底彩照
2019	Lai 等[11]	左右眼识别	单张眼底彩照
2019	Xu 等[12]	核性白内障分级	单张裂隙灯照片
2019	Yang 等[13]	视盘-黄斑联合定位	单张超广角眼底图像
2020	Wu 等[14]	异常检测	单张 OCT B-scan 图像
2020	Ding 等[15]	视盘/视杯分割	单张眼底彩照
2020	Ding 等[16]	RNFLD 检测	单张眼底彩照
2020	Wei 等[17]	眼底病灶分割，DR 分级	单张眼底彩照
2020	Li 等[18]	ROP 检测	多张眼底彩照
2021	Li 等[19]	多病种识别	单张眼底彩照
2021	Zhang 等[20]	多病种识别	单张超广角眼底图像

DR：糖尿病视网膜病变；AMD：年龄相关性黄斑变性；OCT：光学相干断层成像；RNFLD：视神经纤维层缺损；ROP：早产儿视网膜病变

膜平面，而 OCT 图像反映的是视网膜切面，两种不同模态的影像包含的信息存在互补性。为充分利用不同模态影像之间的互补性，需要从单模态深度学习转向多模态深度学习。

1 多模态深度学习的原理

关于模态（modality）一词，既往文献为了覆盖尽可能多的研究领域，其定义要么语焉不详，要么过于抽象[21-22]。考虑到 AI 辅助诊断的背景，本文给出如下定义：模态是对由一种特定类型装置采集的具有相同表达形式的数据的总称。根据该定义，CFP 是一种模态，而 OCT 是另外一种模态，因此图 1 亦可视为不同模态的眼科影像。上述定义也区分了数据本身的多样性（diversity）和模态在概念上的根本差异。因个体因素（如具体设备型号、拍摄者、被拍摄者、拍摄条件等）导致的影像上的差异，不能形成一个单独的模态。同一模态的样本因数据采集过程中的系统性偏差形成的风格各异的数据集合，称为域（domain）[23]。

相比单模态深度学习，多模态深度学习架构的一个重要特性是其数据层要具备同时接受不同模态输入的能力。在其学习过程中，不但要充分提取和利用各个模态内部的有用信息，同时要挖掘各模态之间的互补性并进行有效的多模态信息融合，以实现较单模态网络更优的性能。根据融合发生的位置，多模态深度学习包括数据层、特征层和任务层融合 3 种范式（图 2）。

数据层融合将不同模态的样本混在一起作为“单模态”输入，强制神经网络在训练过程中提取与模态无关的特征[24]（图 2A）。这种范式的优点是可以直接使用现有的单模态架构，缺点是对模态之间的空间关联性要求较高，不适用于类似 CFP 和 OCT 这两种空间上正交的模态。

特征层融合尝试在各个模态的特征提取过程中融合不同模态的信息（图 2B）。浅层特征仍保留相当多的原始数据信息，而深层次的特征包含更多与任务相关的语义特征，因此一般选择在深层特征上进行融合。常见的融合算法有简单的特征向量拼接[25]和旨在获取高阶关联信息的双线性池化（bilinear pooling）、张量融合（tensor fusion）等[26]。

任务层融合是将基于各个模态分别给出的预测结果进行融合[27]（图 2C），因此，在概念上可以看成是多个单模态网络的集成。各个网络既可以独立并行训练，也可以联合训练。对比 3 种范式，数据层融合实现最简单，但适用范围较窄；特征层融合的适用范围广、模型学习能力强，但对融合模块的设计和训练数据量也提出了更高要求；任务层融合则介于二者之间。在实践中选取何种范式，需具体问题具体分析。目前，第 2 种范式是研究者采用的主流方案。

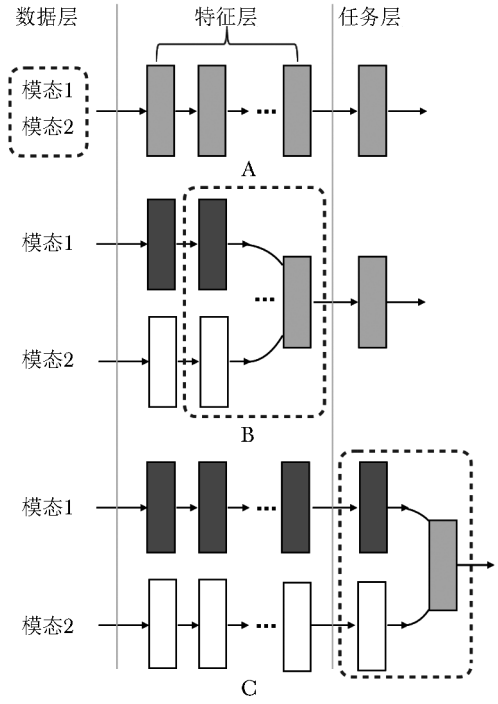

图 2 多模态深度学习的 3 种范式（虚线方框）
A. 数据层融合；B. 特征层融合；C. 任务层融合

2 多模态深度学习在医学领域的应用

多模态深度学习在医学领域的最新应用主要集中于各类肿瘤/癌症的辅助诊断和预后预测方面（表2）。例如，针对乳腺癌分类任务，Wang 等[28]提出了一种多模态分类网络。该网络以普通超声、彩色多普勒超声、剪切波弹性成像、应变弹性成像 4 种不同模态的图像同时作为输入，并在特征层以特征拼接的形式实现多模态信息融合。

针对脑肿瘤患者的总生存期预测任务，Zhou 等[29]将总生存期分为短期（<10 个月）、中期（10~15 个月）、长期（>15 个月）3 类，从而将一个连续值的回归问题简化为三分类问题。研究者提出了一种多模态、多通道分类网络，接受 4 种模态的 MR 影像作为输入；为降低计算复杂度，引入了预处理模块，将三维 MR 立体图像投影得到不同方向的二维平面图。与 Wang 等[28]的研究类似，该研究从不同模态图像提取的特征以及肿瘤大小、患者年龄等辅助信息，也是通过特征拼接的形式实现了多模态信息的融合。

鉴于特征拼接的局限性，研究者们尝试探索更复杂、表达能力更强的多模态融合策略。Chen 等[26]以组织病理学图像和基因组特征为多模态输入，构建了针对癌症诊断与预后预测任务的模型。该模型采用张量融合（tensor fusion）提取组织病理学图像特征和基因组特征之间的关联关系。Jiang 等[30]尝试利用静脉期 CT 和动脉期 CT 图像的互补性以实现更准确的胰腺分割。神经网络先分别从静脉期 CT 和动脉期 CT 图像中提取不同层次的深度特征，之后进行多层次、选择性特征融合。

上述特征融合策略，无论是简单的特征拼接还是相对复杂的具备学习能力的融合，均是由研究者根据其经验人工设计。为了克服人工设计的局限性，Peng 等[31]针对癌细胞远端转移预测任务，尝试通过网络结构搜索（network architecture search，NAS）在训练过程中动态确定对于融合 PET 和 CT 特征最有效的特征融合网络。尽管该研究表明了 NAS 在性能上的优势，由于其本身需要额外的训练数据，这种动态生成的网络结构存在过拟合的风险。此外，NAS 以性能为导向，由此获得的网络结构较之前人工设计的网络的可解释性较差。

3 眼科 AI 的多模态深度学习

3.1 探索

相较于其他医学领域，眼科 AI 的多模态深度学习应用目前仍处于起步阶段（表3）。笔者在主流期刊检索到该方向的首篇应用成果发表于 2019 年[32]，内容为多模态年龄相关性黄斑变性（age-related macular degeneration，AMD）分类问题。研究者采用了一种双流（two-stream）网络架构，分别从 CFP 和 OCT B-scan图像中提取相关特征，之后将不同模态特征进行拼接，再输入到后续的分类任务层，实现正常眼底/干性 AMD/湿性 AMD 的三分类。Xu 等[33]采用了类似的网络架构，并将任务进一步细分为四分类问题（正常眼底/干性 AMD/湿性 AMD/息肉状脉络膜血管病变）。上述研究结果均表明，相比仅采用 CFP 或 OCT 图像的单模态网络，多模态网络在分类准确率方面明显提升，初步显示了多模态深度学习在眼科 AI 上的应用潜力。

不同于以 CFP 和 OCT 图像作为多模态输入，Li 等[24]尝试将 CFP 和经生成对抗网络[34]合成的荧光素眼底血管造影（fluorescein fundus angiography，FFA）混在一起，通过数据层融合，引导神经网络在训练过程中学习模态无关而与任务相关的视觉特征。因此，该技术方案在本质上可以视为一种比基于传统底层图像处理技术更为复杂的数据增强。

北京协和医院在国际视觉与眼科研究协会 2021 年会上报告的一项工作[27]表明，以 CFP 和 OCT 图像序列为输入的多模态深度学习模型也可用于同时检测

表 2 多模态深度学习在医学领域的应用举例

年份（年）	研究者	任务	多模态输入	融合层级	融合策略
2020	Wang 等[28]	乳腺癌分类	普通超声，彩色多普勒超声，剪切波弹性成像，应变弹性成像	特征层	特征拼接
2020	Zhou 等[29]	脑肿瘤患者总生存期预测	4 种模态（T1、T1ce、T2、FLAIR）的 MR 影像	特征层	特征拼接
2020	Chen 等[26]	癌症诊断与预后预测	组织病理学图像，基因组特征	特征层	张量融合
2020	Jiang 等[30]	胰腺分割	静脉期 CT，动脉期 CT	特征层	多层次选择性特征融合
2020	Peng 等[31]	癌细胞远端转移预测	PET，CT	特征层	网络结构搜索

表 3 多模态深度学习在眼科领域的应用举例

年份（年）	研究者	任务	多模态输入	融合层级	融合策略
2019	Wang 等[32]	AMD 分类	眼底彩照，OCT 图像	特征层	特征拼接
2020	Xu 等[33]	AMD/PCV 分类	眼底彩照，OCT 图像	特征层	特征拼接
2020	Li 等[24]	特定眼底疾病识别	眼底彩照，算法合成 FFA	数据层	样本混合
2021	Yang 等[27]	多种眼底疾病识别	眼底彩照，OCT 图像序列	任务层	平均得分

AMD、OCT：同表 1；PCV：息肉状脉络膜血管病变；FFA：荧光素眼底血管造影

多种常见致盲性眼底疾病，如 DR、AMD、视网膜前膜、病理性近视等。相比之前的工作，除检测病种数量增加外，在 OCT 分支网络中引入了一种深度多示例学习模块[18]，可直接接受整个 OCT 图像序列，无须人工选择 OCT B-scan 图像作为多模态网络的输入。

3.2 挑战

虽然上述探索得出了令人鼓舞的研究结果，但眼科 AI 的多模态深度学习仍存在相当多的技术挑战需要攻克，主要集中于数据和算法两个层面。

数据层面，相比单模态场景，多模态数据存在配对要求，其前期原始数据采集和后期人工标注的难度及成本显著增加。因此，需加强各相关单位合作机制创新，以获得更多的多模态研发数据；此外，在数据高效深度学习（data-efficient deep learning）方面需进行技术创新，以在训练数据规模受限的条件下实现有效的多模态学习。

算法层面，尽管现有的研究结果表明，多模态模型总体性能优于单模态模型，但在特定病种中，多模态模型并不总能超过在该病种上表现最优的单模态模型。单一模态影像并不能覆盖所有疾病特征。比如 DR 作为血管病，特征表现面积较大，CFP 相比 OCT 可反映更多的疾病信息；而黄斑水肿的特征反映在视网膜层次厚度和结构的变化上，OCT 的优势则更明显。如何设计更加智能的、具有自主选择能力的多模态信息融合机制是值得深入探索的研究课题[35]。

3.3 前景

需要指出的是，由于现有关于多模态眼科 AI 的研究相对较少，多模态深度学习在病种亚型分类、分期和相应的处置建议推荐等方面，较单模态的优势尚未充分体现。以干性 AMD 为例，玻璃膜疣是干性 AMD 的特征性临床表现，在早期阶段，玻璃膜疣较小，OCT 相比 CFP 更容易观察到这一表现。理论上可以利用不同模态影像在病种不同阶段的不同适应性，实现更细粒度的分类，从而推荐更恰当的处置建议。

在数据形态上，现有研究主要考虑融合不同模态的影像，而在临床实践中，患者信息除影像数据外，还有非影像数据，比如定性的病史、定量的视光检查结果等。当前，这些非影像数据存在记录不准确或不完整等问题。随着电子病历系统的普及和建设水平的提高，能够有效融合影像和非影像数据的多模态 AI 有望在青少年近视综合防控、成人慢病管理、个性化医疗保健等多个应用场景发挥关键作用。

4 小结

深度学习是当前医学人工智能的核心技术。现有研究结果表明，在眼底疾病辅助诊断方面，多模态深度学习较基于单一模态的技术方案在识别性能上存在明显优势。发展面向眼科的多模态深度学习技术具有广阔的应用前景。由于多模态影像对于眼底疾病诊断的高效性和必要性，眼底成像设备已呈现“一体化”和“低成本化”的趋势，多模态 AI 辅助诊断具有巨大的普及空间。此外，眼底作为非侵入式观察全身健康状况的“窗口”，对于慢性病进展的检测和管理起着重要提示作用。我们有理由相信，多模态眼底分析在眼科以外的医疗健康领域也有着巨大的需求和应用潜力。

利益冲突：无

志谢：感谢北京致远慧图科技有限公司丁大勇博士对本文的建议，中国人民大学博士生林海斓在本文修订方面提供的帮助。

参 考 文 献

[1] Etzioni O, Decario N. AI can help scientists find a COVID-19 vaccine [EB/OL]. [2021-06-16]. https://www.wired.com/story/opinion-ai-can-help-find-scientists-find-a-covid-19-vaccine.

[2] Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings [J]. IEEE Open J Eng Med Biol, 2020, 1: 275-281.

[3] Zeeberg A. D. I. Y. Artificial intelligence comes to a Japanese family farm [EB/OL]. [2021-06-16].https://www.newyorker.com/tech/annals-of-technology/diy-artificial-intelligence-

comes-to-a-japanese-family-farm.
[4] Bengio Y. Learning deep architectures for AI [G]. Foundations and Trends® in Machine Learning, 2009, 2: 1-127.
[5] Schmidhuber J. Deep learning in neural networks: An overview [J]. Neural Netw, 2015, 61: 85-117.
[6] Zheng A, Casari A. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists [M]. New York: O'Reilly Media Inc., 2018.
[7] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs [J]. JAMA, 2016, 316: 2402-2410.
[8] Burlina PM, Joshi N, Pekala M, et al. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks [J]. JAMA Ophthalmol, 2017, 135: 1170-1176.
[9] Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning [J]. Cell, 2018, 172: 1122-1131. e9.
[10] Wei Q, Li X, Wang H, et al. Laser scar detection in fundus images using convolutional neural network [C]. ACCV, 2018: 191-206.
[11] Lai X, Li X, Qian R, et al. Four models for automatic recognition of left and right eye in fundus images [C]. MMM, 2019: 507-517.
[12] Xu C, Zhu X, He W, et al. Fully deep learning for slit-lamp photo based nuclear cataract grading [C]. MICCAI, 2019: 513-521.
[13] Yang Z, Li X, He X, et al. Joint localization of optic disc and fovea in ultra-widefield fundus images [C]. MLMI, 2019: 453-460.
[14] Wu J, Zhang Y, Wang J, et al. AttenNet: Deep attention based retinal disease classification in OCT images [C]. MMM, 2020: 565-576.
[15] Ding F, Yang G, Wu J, et al. High-order attention networks for medical image segmentation [C]. MICCAI, 2020: 253-262.
[16] Ding F, Yang G, Ding D, et al. Retinal nerve fiber layer defect detection with position guidance [C]. MICCAI, 2020: 745-754.
[17] Wei Q, Li X, Yu W, et al. Learn to segment retinal lesions and beyond [C]. ICPR, 2020: 7403-7410.
[18] Li X, Wan W, Y. Zhou, et al. Deep multiple instance learning with spatial attention for ROP case classification, instance selection and abnormality localization [C]. ICPR, 2020: 7293-7298.
[19] Li B, Chen H, Zhang B, et al. Development and evaluation of a deep learning model for the detection of multiple fundus diseases based on colour fundus photography [J]. Br J Ophthalmol, 2021. doi: 10. 1136/bjophthalmol-2020-316290.
[20] Zhang C, He F, Li B, et al. Development of a deep-learning system for detection of lattice degeneration, retinal breaks, and retinal detachment in tessellated eyes using ultra-wide-field fundus images: A pilot study [J]. Graefes Arch Clin Exp Ophthalmol, 2021, 259: 2225-2234.
[21] Zhang C, Yang Z, He X, et al. Multimodal intelligence: Representation learning, information fusion, and applications [J]. IEEE J Sel Top Signal Process, 2020, 14: 478-493.
[22] Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: A survey and taxonomy [J]. IEEE Trans Pattern Anal Mach Intell, 2018, 41: 423-443.
[23] Wang J, Tian K, Ding D, et al. Unsupervised domain expansion for visual categorization [J]. ACM Trans Multimedia Comput Commun Appl, 2021. https: //arxiv. org/abs/2104. 00233.
[24] Li X, Jia M, Islam M T, et al. Self-supervised feature learning via exploiting multi-modal data for retinal disease diag-nosis [J]. IEEE Trans Med Imaging, 2020, 39: 4023-4033.
[25] Wang W, Xu Z, Yu W, et al. Two-stream CNN with loose pair training for multi-modal AMD categorization [C]. MICCAI, 2019: 156-164.
[26] Chen RJ, Lu MY, Wang J, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis [J]. IEEE Trans Med Imaging, 2020. doi: 10. 1109/TMI. 2020. 3021387.
[27] Yang J, Yang Z, Mao Z, et al. Bi-modal deep learning for recognizing multiple retinal diseases based on color fundus photos and OCT images [C]. ARVO Annual Meeting, 2021.
[28] Wang J, Miao J, Yang X, et al. Auto-weighting for breast cancer classification in multi- modal ultrasound [C]. MICCAI, 2020: 190-199.
[29] Zhou T, Fu H, Zhang Y, et al. M2Net: Multi-modal multi-channel network for overall survial time prediction of brain tumor patients [C]. MICCAI, 2020: 221-231.
[30] Jiang X, Luo Q, Wang Z, et al. Multiphase and multi-level selective feature fusion for automated pancreas segment from CT images [C]. MICCAI, 2020: 460-469.
[31] Peng Y, Bi L, Fulham M, et al. Multi-modality information fusion for radiomics-based neural architecture search [C]. MICCAI, 2020: 763-771.
[32] Wang W, Xu Z, Yu W, et al. Two-stream CNN with loose pair training for multi-modal AMD categorization [C]. MICCAI, 2019: 156-164.
[33] Xu Z, Wang W, Yang J, et al. Automated diagnoses of age-related macular degeneration and polypoidal choroidal vasculopathy using bi-modal deep convolutional neural networks [J]. Br J Ophthalmol, 2021, 105: 561-566.
[34] Zhu J, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]. ICCV, 2017: 2223-2232.
[35] Li X, Zhou Y, Wang J, et al. Multi-modal multi-instance learning for retinal disease recognition [C]. ACMMM, 2021. doi: 10. 1145/3474085. 3475418.

（收稿：2021-06-28 录用：2021-07-29 在线：2021-08-19）
（本文编辑：李娜）