



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2021
07
VOL.26

ISSN1006-8961
CN11-3758/TB



中国多媒体大会(ChinaMM2020)专刊



第26卷第7期 (总第303期)

2021年7月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿, 均视为同意在本刊网站及CNKI等全文数据库出版, 所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用, 所有内容, 未经许可, 不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院空天信息创新研究院

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 吴一戎

编辑出版 《中国图象图形学报》编辑出版委员会

通信地址 北京市海淀区北四环西路19号

邮 编 100190

电子信箱 jig@aircas.ac.cn

电 话 010-58887035

网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

(邮政信箱: 北京399信箱 邮编: 100048)

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Aerospace Information Research Institute, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong

Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics

Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China

Zip code 100190

E-mail jig@aircas.ac.cn

Telephone 010-58887035

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation

(P.O.Box 399, Beijing 100048, P.R.China)

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

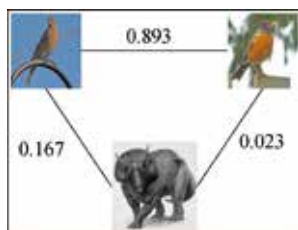
ISSN 1006-8961

CODEN ZTTXFZ

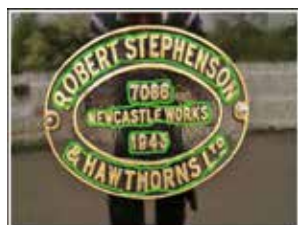
国外发行代号 M1406

国内邮发代号 82-831

国内定价 60.00元



类别语义相似性监督的小样本图像识别(第1594页)



像素聚合和特征增强的任意形状场景文本检测(第1614页)



时序特征融合的视频实例分割(第1692页)

综述

面向上行流媒体的压缩感知视频流技术前沿

刘浩, 黄荣, 袁浩东 1545

图像处理和编码

耦合保持投影哈希跨模态检索

闵康凌, 张国宾, 王磊, 李丹萍 1558

面向视觉搜索的空间局部敏感哈希方法

黄小燕, 孙彬, 杨展源, 朱映映, 田奇 1568

图像分析和识别

融合视觉关系检测的电力场景自动危险预警

高明, 左红群, 柏帆, 田清阳, 葛志峰, 董兴宁, 甘甜 1583

类别语义相似性监督的小样本图像识别

徐鹏帮, 桑基韬, 路冬媛 1594

局部双目视差回归的目标距离估计

张羽丰, 李昱希, 赵明璧, 喻晓源, 占云龙, 林巍峭 1604

像素聚合和特征增强的任意形状场景文本检测

师广琛, 巫义锐 1614

相位一致性指导的全参考全景图像质量评价

夏雨蒙, 王永芳, 王闯 1625

图像理解和计算机视觉

特征金字塔结构的时序行为识别网络

何嘉宇, 雷军, 李国辉 1637

多模态多层次事件网络的谣言检测

李莎, 张怀文, 钱胜胜, 方全, 徐常胜 1648

多模态零样本人体动作识别

吕露露, 黄毅, 高君宇, 杨小汕, 徐常胜 1658

足球视频球员感知跟踪算法

冯思佳, 宋子恺, 于俊清, 何云峰, 管涛 1668

融合时空图卷积的多人交互行为识别

成科扬, 吴金霞, 王文杉, 荣兰, 詹永照 1681

时序特征融合的视频实例分割

黄泽涛, 刘洋, 于成龙, 张加佳, 王轩, 漆舒汉 1692

医学图像处理

多尺度深度特征提取的肝脏肿瘤CT图像分类

毛静怡, 宋余庆, 刘哲 1704

3D多尺度深度卷积神经网络肺结节检测

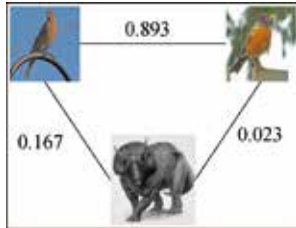
孙华聪, 彭延军, 郭燕飞, 张晓庆 1716

结合多通道注意力的糖尿病性视网膜病变分级

顾婷菲, 郝鹏翼, 白琮, 柳宁 1726

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Few shot image recognition based on class semantic similarity supervision(P1594)



Arbitrary shape scene-text detection based on pixel aggregation and feature enhancement (P1614)



Video instance segmentation based on temporal feature fusion (P1692)

Review

Survey on compressive sensing video stream for uplink streaming media

Liu Hao, Huang Rong, Yuan Haodong 1545

Image Processing and Coding

Structure-preserving hashing with coupled projections for cross-modal retrieval

Min Kangling, Zhang Guobin, Wang Lei, Li Danping 1558

Locality-sensitive hashing approach based on semantic space for visual retrieval

Huang Xiaoyan, Sun Bin, Yang Zhanyuan, Zhu Yingying, Tian Qi 1568

Image Analysis and Recognition

Visual relationship detection-based emergency early-warning description generation in electric power industry

Gao Ming, Zuo Hongqun, Bai Fan, Tian Qingyang, Ge Zhifeng, Dong Xingning, Gan Tian 1583

Few shot image recognition based on class semantic similarity supervision

Xu Pengbang, Sang Jitao, Lu Dongyuan 1594

Object distance estimation based on stereo regional disparity regression

Zhang Yufeng, Li Yuxi, Zhao Mingbi, Yu Xiaoyuan, Zhan Yunlong, Lin Weiyao 1604

Arbitrary shape scene-text detection based on pixel aggregation and feature enhancement

Shi Guangchen, Wu Yirui 1614

Phase consistency guided full-reference panoramic image quality assessment algorithm

Xia Yumeng, Wang Yongfang, Wang Chuang 1625

Image Understanding and Computer Vision

Temporal action detection based on feature pyramid hierarchies

He Jiayu, Lei Jun, Li Guohui 1637

Multi-modal multi-level event network for rumor detection

Li Sha, Zhang Huaiwen, Qian Shengsheng, Fang Quan, Xu Changsheng 1648

Multimodal-based zero-shot human action recognition

Lyu Lulu, Huang Yi, Gao Junyu, Yang Xiaoshan, Xu Changsheng 1658

Players-aware tracking algorithm in soccer video

Feng Sijia, Song Zikai, Yu Junqing, He Yunfeng, Guan Tao 1668

Multi-person interaction action recognition based on spatio-temporal graph convolution

Cheng Keyang, Wu Jinxia, Wang Wenshan, Rong Lan, Zhan Yongzhao 1681

Video instance segmentation based on temporal feature fusion

Huang Zetao, Liu Yang, Yu Chenglong, Zhang Jiajia, Wang Xuan, Qi Shuhan 1692

Medical Image Processing

CT image classification of liver tumors based on multi-scale and deep feature extraction

Mao Jingyi, Song Yuqing, Liu Zhe 1704

3D multi-scale deep convolutional neural networks in pulmonary nodule detection

Sun Huacong, Peng Yanjun, Guo Yanfei, Zhang Xiaoqing 1716

Diabetic retinopathy grading based on multi-channel attention

Gu Tingfei, Hao Pengyi, Bai Cong, Liu Ning 1726

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2021)07-1726-11

论文引用格式: Gu T F, Hao P Y, Bai C and Liu N. 2021. Diabetic retinopathy grading based on multi-channel attention. Journal of Image and Graphics, 26(07): 1726-1736 (顾婷菲, 郝鹏翼, 白琮, 柳宁. 2021. 结合多通道注意力的糖尿病性视网膜病变分级. 中国图象图形学报, 26(07): 1726-1736)
[DOI:10.11834/jig.200518]

结合多通道注意力的糖尿病性视网膜病变分级

顾婷菲¹, 郝鹏翼^{1,2*}, 白琮^{1,2}, 柳宁³

1. 浙江工业大学计算机科学与技术学院, 杭州 310023; 2. 浙江省可视媒体智能处理技术研究重点实验室, 杭州 310023;
3. 上海交通大学电子信息与电气工程学院, 上海 200240

摘要: **目的** 糖尿病性视网膜病变(diabetic retinopathy, DR)是一种常见的致盲性视网膜疾病,需要患者在早期就能够被诊断并接受治疗,否则将会造成永久性的视力丧失。能否检测到视网膜图像中的微小病变如微血管瘤,是糖尿病性视网膜病变分级的关键。然而这些病变过于细小导致使用一般方法难以正确地辨别。为了解决这一问题,本文提出了一种基于多通道注意力选择机制的细粒度分级方法(fine-grained grading method based on multi-channel attention selection, FGMA)用于糖尿病性视网膜病变的分级。**方法** 该方法结合了细粒度分类方法和多通道注意力选择机制,通过获取局部特征提升分级的准确度。此外考虑到每一层通道特征信息量与分类置信度的关系,本文引入了排序损失以优化每一层通道的信息量,用于获取更加具有信息量的局部区域。**结果** 使用两个公开的视网膜数据集(Kaggle 和 Messidor)来评估提出的细粒度分级方法和多通道注意力选择机制的有效性。实验结果表明:FGMA在 Kaggle 数据集上进行的五级分类任务中相较于现有方法,在平均准确度(average of classification accuracy, ACA)上取得了 3.4%~10.4% 的提升。尤其是对于病变点最小的 1 级病变,准确率提升了 11%~18.9%。此外,本文使用 FGMA 在 Messidor 数据集上进行二分类任务。在推荐转诊/不推荐转诊分类上 FGMA 得到的准确度(accuracy, Acc)为 0.912,比现有方法提升了 0.1%~1.9%,同时 AUC(area under the curve)为 0.962,比现有方法提升了 0.5%~9.9%;在正常/不正常分类上 FGMA 得到的准确度为 0.909,比现有方法提升了 2.9%~8.8%,AUC 为 0.950,比现有方法提升了 0.4%~8.9%。实验结果表明,本文方法在五分类和二分类上均优于现有方法。**结论** 本文所提细粒度分级模型,综合了细粒度提取局部区域的思路以及多通道注意力选择机制,可以获得较为准确的分级结果。

关键词: 糖尿病性视网膜病变(DR);病变分级;细粒度分级;深度学习;多通道注意力选择机制;局部特征提取

Diabetic retinopathy grading based on multi-channel attention

Gu Tingfei¹, Hao Pengyi^{1,2*}, Bai Cong^{1,2}, Liu Ning³

1. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China;
2. Key Laboratory of Visual Media Intelligent Processing Technology of Zhejiang Province, Hangzhou 310023, China;
3. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: **Objective** Diabetic retinopathy (DR) is a common blinding retinal disease that cannot be cured in the later stage, and requires patients to be diagnosed and treated at an early stage; otherwise, it causes permanent vision loss. The

收稿日期:2020-08-26;修回日期:2021-02-08;预印本日期:2021-02-15

* 通信作者: 郝鹏翼 haopy@zjut.edu.cn

基金项目: 国家自然科学基金项目(U1908210, 61801428, U20A20196); 浙江省自然科学基金项目(LR21F020002)

Supported by: National Natural Science Foundation of China(U1908210, 61801428, U20A20196); Natural Science Foundation of Zhejiang Province, China(LR21F020002)

prevalence of diabetic retinopathy is extremely high in China, and is in the stage of rapid growth. At present, China has become the country that has the largest number of patients with diabetic retinopathy. Diagnosis of DR is usually performed by analyzing fundus medical images. Detection of microscopic lesions such as microaneurysms in retinal images is necessary in grading diabetic retinopathy with neural networks. This condition requires the attention mechanism to simulate the focus of the human eyes and focus on the local area with information. However, most of the present methods only consider the attention in the spatial domain and ignore the information in the channel attention, which cause difficulty in distinguishing the small lesions. To solve this problem, a fine-grained grading method based on multi-channel attention selection (FGMAS) mechanism is proposed for the grading of diabetic retinopathy in this paper. **Method** This method combines fine-grained classification with a multi-channel attention selection mechanism. First, the structure of fine-grained classification is adopted to improve the recognition accuracy of small differences between categories by obtaining local regional features. Then, the characteristics of different feature layers in the channel domain with different information content are used to select high-information channels. The model establishes the relationship between information content and classification confidence, and obtains the lesion area that is conducive to classification results. Finally, the local and global features are combined to improve the accuracy of classification. In addition, considering the relationship between the channel characteristic information of each layer and the classification confidence, this study also introduces Rank_loss to optimize the channel information of each layer. The loss function enables the regions with higher classification confidence to have higher information content and obtain better classification results. **Result** Two open retinal datasets (Kaggle and Messidor) are used to evaluate the effectiveness of the proposed fine-grained grading method and multi-channel attention selection mechanism. The experimental results show that FGMAS performs a five-level classification on the Kaggle dataset with better results than the existing method, with an average accuracy of 0.577, which is 3.4%~10.4% higher than the accuracy of other methods. The first category shows small lesion points, which are difficult to distinguish in other methods. However, the accuracy rate of 0.301 can be obtained through FGMAS proposed in this paper, which is better than other methods with the improvement of 11%~18.9%, such as 0.190 of VGGNet with Extra Kernel/LGI (VNKK/LGI). Meanwhile, FGMAS is used in the Messidor dataset to perform a dichotomous task, including recommended reference/non-reference and normal/abnormal classification. In the reference/non-reference classification task, the experimental results are 0.912 of accuracy and 0.962 of AUC (area under the curve), which is superior to the existing methods by 0.1%~1.9% and 0.5%~9.9%, respectively. In the normal/abnormal classification task, the experimental results are 0.909 of accuracy and 0.950 of AUC, which are improved by 2.9%~8.8% and 0.4%~8.9% respectively, compared with existing methods. In addition, parameter experiments are set up in this study, and the function of each parameter and optimal parameter selection result are analyzed in detail. **Conclusion** This study proposes a fine-grained grading model that combines the fine-grained classification and multi-channel attention models. In addition, Rank_loss combines the ranking result and information of every layer. It is used to obtain the local feature area, which is beneficial to the classification result. According to the experimental results, the model can obtain good results in five-classification and two-classification tasks.

Key words: diabetic retinopathy (DR); lesion grading; fine-grained grading; deep learning; multi-channel attention selection; local feature extraction

0 引言

糖尿病性视网膜病变 (diabetic retinopathy, DR) 是糖尿病最常见的并发症之一,也是人类致盲的主要原因之一。在中国糖尿病患者中,糖尿病性视网膜病变的患病率高达 23%,且该视网膜病变在晚期无法治愈,因此早期的诊断就显得尤为重要 (中华医学会眼科学会眼底病学组,2014)。然而糖

尿病性视网膜病变没有早期的预警信号,人们对其的诊断只能由电子摄像机等设备拍摄眼底图,交由专业的医生进行分析之后才能给出,这极度依赖于医生的专业知识和相关设备的配置,通常会花几天时间来进行诊断。因此,当前大量的潜在患者和少量的经验人员给糖尿病性视网膜病变的检测带来了极大的困难。

为解决因人工诊断带来的问题,研究人员提出了多种人工设计特征的传统机器学习方法对视网膜

病变进行分级。例如数学形态学、视网膜病变跟踪、阈值化和可变形模型等方法。Pires 等人(2017)和 Rocha 等人(2012)就使用了词袋视觉文字表达和 BossaNova、Fisher 等载体检测病变,使用从病变探测器中获得的概率分数来表示视网膜图像属于哪一级别。Salami 等人(2014)使用改良的局部熵阈值算法从属于健康和糖尿病视网膜图像的眼底图中提取血管。再通过统计矩阵和主成分分析的方法,提取不同的特征并组合到一个特征向量中,输入顺序最小优化(sequential minimal optimal, SMO)分类器,以得到分类结果。但是,专业知识是手工设计特征的先决条件,选择合适的功能需要深入研究各种相关知识和烦琐的参数设置。因此,基于手工设计特征的技术不能得到很好的推广。

庞大数据集的可用性和图形处理单元提供的强大计算能力极大地推动了对深度学习算法的研究。研究人员已经针对糖尿病性视网膜病变提出了多种基于深度学习的算法,以分析视网膜眼底图像,并取得了较好的结果。例如 Chen 等人(2018)提出了轻量级网络 SI2DRNet-v1 以及 6 种增强方法来进一步提高 DR 病变检测的性能。Bravo 和 Arbeláez(2017)研究了不同预处理方法的影响,并使用基于 VGG16(Visual Geometry Group 16)的架构将它们结合起来,在糖尿病视网膜病变分级中取得了良好的效果。 M^2 CNN (multi-cell multi-task convolutional neural networks)(Zhou 等,2018)是一种逐步增加深度神经网络深度和输入图像分辨率的多细胞结构,在提高分类精度的同时减少了训练时间。Vo 和 Verma(2016)提出了一种混合颜色空间 LGI,与神经网络相结合用于 DR 识别。虽然这些算法都致力于通过各种方法提取病变特征,但仍然存在对微小病变识别性能不足的问题。

糖尿病性视网膜病变分级最大的挑战就是眼底病变图的类别精度相比于其他图像类别更加精细,过于微小的病变点使得类间差异非常细微,难以分辨。这与细粒度分类问题非常契合。细粒度分类问题指的是对某一类中的子类进行分类,通常通过微小的局部差异来区分类别。因此引入细粒度分级来实现视网膜病变的分级。而对于细粒度分级而言,如何找到并获取有效的局部区域决定了细粒度分级模型的性能。

本文采用了注意力机制来获取局部区域。现有

的注意力机制,如 Zoom-in-net (Wang 等,2017)、BiRA-Net(Zhao 等,2019)等通常是针对空间的,在空间上对注意力图中的像素点通过加权的方式进行空间概率建模,从而使需要关注的像素点拥有较高的权重和更多的关注。但是却忽视了通道方向的注意力,只是简单地利用注意力机制作为提取特征的方法,对于每一层通道都一视同仁地进行相同处理。但是事实上,每一层通道所包含的信息是不同的。并且通道方向的注意力可以看做是根据上下文的需求选择语义属性的过程(Chen 等,2017)。当想要预测微血管瘤病变时,通道注意力机制将给能够预测到微血管瘤病变的通道层分配更多的权重,通过获取信息量高的通道注意力图,模型不只是关注像素点的位置,而是可以获得特定的某些对分类具有较高影响力的病变点。因此,为了更好地识别眼底病变图中微小的病变点,本文提出了一种多通道注意力选择机制,将重点着眼于通道域上的注意力,区分不同通道上的信息,根据每个通道所包含的信息量,对通道注意力图进行信息量排序和选择,并在排序和选择的过程中,引入排序损失,通过对通道置信度和信息量的约束,使其保持同样的递增递减关系,从而获得更具影响力的局部区域。

图 1 给出了本文提出的基于多通道注意力选择机制的细粒度分级方法(fine-grained grading method based on multi-channel attention selection, FGMAS)的网络框架图。框架主要由多通道注意力选择、全局特征提取、局部特征提取、特征融合等 4 部分组成。本文所提出的 FGMAS 在 Kaggle (EyePACS, 2015)和 Messidor(Decencière 等,2014)两个公开数据集上分别进行了五分类和二分类实验。实验结果表明,通过引入多通道注意力选择机制获取局部特征的方法可以有效获取局部病变特征,从而提高糖尿病性视网膜病变的细粒度分级性能,且本文方法优于现有的方法。

1 多通道注意力选择

1.1 问题描述

对于一个给定的特征矩阵 $F \in \mathbf{R}^{C \times H \times W}$,其中, C 是特征矩阵的通道数, H 、 W 分别是特征矩阵的长和宽。多通道注意力选择的目标是提取 F 每一层通道的信息量 $S = \{S_1, S_2, S_3, \dots, S_M\}$, S_i ($i \in [1,$

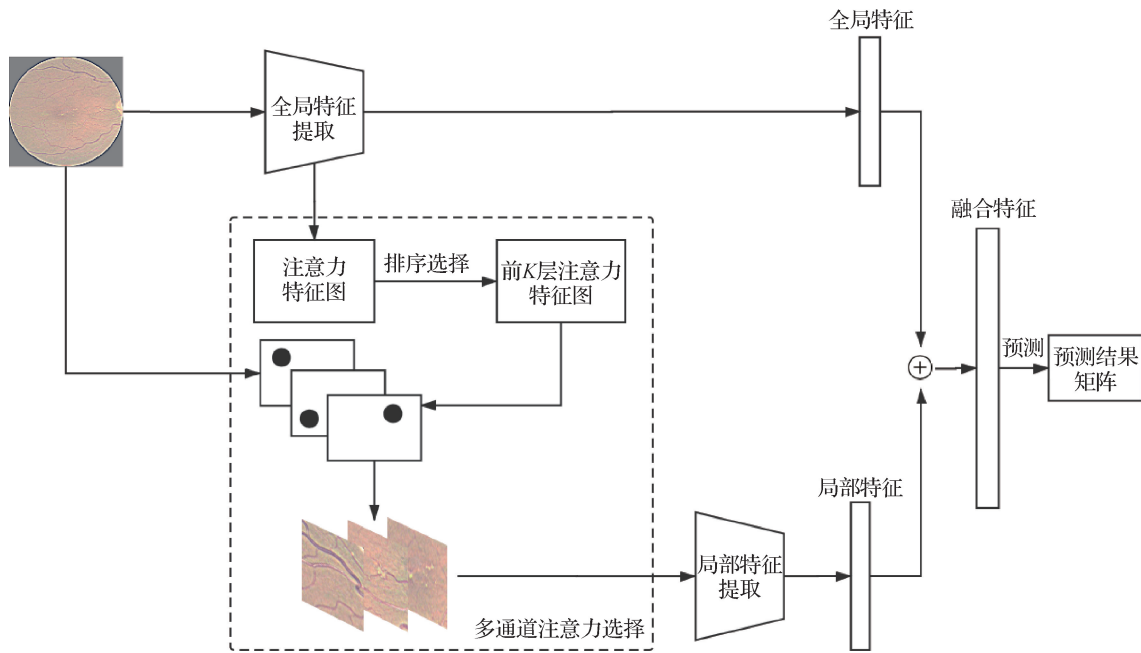


图1 FGMA网络框架图

Fig.1 FGMA network

$M\} \in \mathbf{R}$ 为第 i 层通道的信息量, 并通过损失函数约束获得具有高信息量的局部区域 $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \dots, \mathbf{R}_K\}$, $\mathbf{R}_i (i \in [1, K])$ 为信息量排名为 i 的局部区域。这些局部区域应为糖尿病性视网膜病变的细粒度分级带来有利信息。

1.2 信息量提取

通道方向的注意力可以理解为每个特征在某一通道上的信号分量, 也可以称之为通道的权重, 该权重有大有小。可以利用某一通道的信息量来表示该通道对于分级结果的重要性和相关度, 相关度越高的通道层, 也就是在分级过程中模型需要去注意的通道层, 因此本文通过提取通道的信息量来获取对分类结果拥有更大影响力的通道层。

首先将特征矩阵 $\mathbf{F}$ 作为输入, 通过 3 个卷积层得到通道数为 M 的注意力特征图 $\mathbf{A} \in \mathbf{R}^{M \times H \times W}$ 。可以将注意力特征图视为是将图像特征矩阵压缩通道数之后得到的, 具体压缩的通道数 M 是一个可自由设定的超参数。该特征图中每一层通道的注意力图都可以看做是一个特征矩阵, 都自适应聚焦于图像的某个特定部位。对于通道注意力, 可以将注意力特征图视为 $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_M\}$, 其中 $\mathbf{A}_i (i \in [1, M]) \in \mathbf{R}^{H \times W}$ 为第 i 层通道上的注意力特征。

接下来, 使用卷积的注意力机制模块 (convolu-

tional block attention module, CBAM) (Woo 等, 2018) 提取每一层特征的信息量。模型先对注意力特征图进行归一化, 将像素大小量化到 $[0, 1]$ 区间, 即

$$\mathbf{A}_i = \frac{\mathbf{A}_i - \min(\mathbf{A}_i)}{\max(\mathbf{A}_i) - \min(\mathbf{A}_i)} \quad (1)$$

然后计算基于长宽的全局最大池化 Maxpool 和全局平均池化 Avgpool , 获得大小为 $1 \times M$ 的通道描述 $\mathbf{F}_{\max}, \mathbf{F}_{\text{avg}}$ 。再将它们分别经过多层感知机操作 (multilayer perceptron, MLP) 后进行逐像素的加操作。最后使用 Sigmoid 激活函数进行激活, 生成最终的信息量分数图 $\mathbf{S} = \{S_1, S_2, S_3, \dots, S_M\}$, 其中 $S_i (i \in [1, M]) \in \mathbf{R}$ 为第 i 层通道上的信息量, 与每一层注意力特征图 $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_M\}$ 一一对应。总体信息量分数计算为

$$T(\mathbf{A}) = f(\text{MLP}(\text{Maxpool}(\mathbf{A})) + \text{MLP}(\text{Avgpool}(\mathbf{A}))) = f(\mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{\max})) + \mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{\text{avg}}))) \quad (2)$$

式中, 权重因子 $\mathbf{W}_0 \in \mathbf{R}^{c/r \times c}$, $\mathbf{W}_1 \in \mathbf{R}^{c \times c/r}$, C 是通道数, r 是减少率, 通过卷积操作实现, f 是 Sigmoid 操作, $\mathbf{W}_0$ 后面使用了 ReLU 作为激活函数。

1.3 局部区域选择

这部分对每个特征提取器获得的注意力特征图进行排序选择, 即对不同通道的特征层进行排序选择, 获取信息量大的特征层, 从而得到有利于提升细

粒度分级精度的局部区域 $R = \{R_1, R_2, R_3, \dots, R_K\}$ 。

通过已经压缩通道的注意力特征图获取了所需要的特征和信息量。接下来首先要对信息量 S 进行排序, 获得前 K 名的信息量 $S' = \{S'_1, S'_2, \dots, S'_K\}$ 以及对应的特征层编号索引, 其中 $S'_1 \geq S'_2 \geq \dots \geq S'_K$, 根据特征层的索引选取相对应的前 K 个特征层 $\{A'_1, A'_2, \dots, A'_K\}$ 。

然后根据提取到的前 K 层注意力特征图生成一个遮罩, 将图中大于阈值 $\theta \in [0, 1)$ 的元素置为 1, 其他元素置为 0, 即

$$A'_v(i, j) = \begin{cases} 1 & A'_v(i, j) \geq \theta \\ 0 & A'_v(i, j) < \theta \end{cases} \quad (3)$$

式中, $A'_v(i, j) = 1$ 所表示的区域就是分级中需要的局部区域。模型计算局部区域所占矩形区域, 将其映射到原图中, 在原图中分割出相应区域并上采样至模型输入的图片大小, 将其作为一个局部区域, 每一幅图经过上述操作, 总共获得 K 个局部区域 $R = \{R_1, R_2, R_3, \dots, R_K\}$, 如图 2 所示。总体来说, 遮罩操作可以视为一个截图的策略, 截取图像中的某一区域来作为局部区域进行后续操作。上文所述的阈值 θ 作为一个超参数可以根据具体问题进行调节, 文中设置为 0.5。

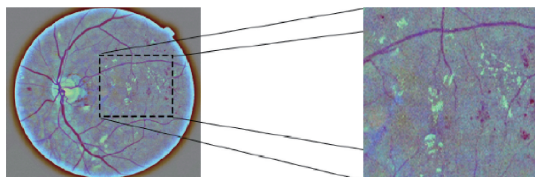


图 2 局部区域提取
Fig. 2 Extracting local area

在实验过程中发现, 虽然上述方法选取了信息量前 K 特征层, 但是不同的特征层可能会聚焦在图像相同的部位, 也就是识别图像中的同一物体, 为了让特征图可以注意到其他部位, 在提取注意力特征图之后, 对其进行通道 Dropout 操作, 与全连接层的神经元 Dropout 相似, 以固定概率 p 随机丢弃一部分特征层, 当信息量最高的特征层被丢弃, 意味着模型暂时忽略图中最具代表性的部分, 被迫提高识别性较差部分的信息量, 这将鼓励网络关注到另外的特征图, 提高了视网膜病变分级模型的鲁棒性和目标定位的精度。

在通道 Dropout 的过程中, 随机生成一个只有 0、1 值的向量 $m \in \mathbf{R}^{1 \times K}$, 将其在通道维度与特征图相乘

$$A'' = A' \times m \quad (4)$$

式中, $m \sim \text{Bernoulli}(p)$, 表示服从伯努利分布。

2 基于多通道注意力选择的 DR 分级

2.1 特征提取及融合

在训练阶段, 一方面, 模型将图像 X 输入全局特征提取器, 获得全局特征 $F \in \mathbf{R}^{C \times H \times W}$ 。将全局特征输入多通道注意力选择机制获取 K 个局部区域, 将局部区域分别输入局部特征提取器对模型进行训练, 强制模型注意这些局部区域并获取局部特征 $F' \in \mathbf{R}^{K \times C \times H \times W}$, 可视为 $F' = \{F'_1, F'_2, \dots, F'_K\}$, F'_i 为第 i 个局部区域获得的局部特征。再将全局特征和局部特征分别输入全连接层, 得到每幅图的分类置信度 $C = \{C_0, C_1, C_2, \dots, C_K\}$, 其中 C_0 是全局特征的分类置信度, C_i ($i \in [1, K]$) 是局部特征的分类置信度, 如图 3 所示。通过排序损失函数对模型优化, 使得分类置信度与信息量具有相同的顺序, 驱使模型获取置信度更高的区域。

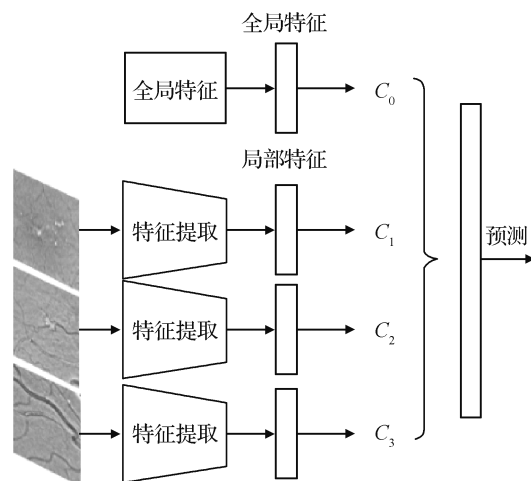


图 3 特征提取及融合

Fig. 3 Feature extraction and fusion

另一方面, 模型将全局特征 F 以及局部特征 F' 进行特征融合, 在通道维度连接成一个特征矩阵, 输入全连接层映射为概率矩阵, 表示图像属于各个疾病级别的概率。

本文使用经过 ImageNet (Russakovsky 等, 2015) 预训练的 DenseNet (Huang 等, 2017) 作为特征提取

器,且全局特征提取器和局部特征提取器共享参数。

2.2 损失函数及优化

在获取通道特征图的时候,需要对特征区域进行约束,所以结合多个损失函数以使得不同通道的注意力聚焦区域具有辨别性和多样性。在一般计算机视觉中,常用的分类损失函数是交叉熵损失函数

$$\begin{aligned} \text{loss}(x, \text{class}) = & -\log\left(\frac{\exp(x[\text{class}])}{\sum_j \exp(x[j])}\right) = \\ & -x[\text{class}] + \log\left(\sum_j \exp(x[j])\right) \end{aligned} \quad (5)$$

式中, x 是得到的分类置信度, class 是类别数, class 并不参与计算,只是作为一个置信度的索引。

2.2.1 联合损失

联合损失基于交叉熵损失,通过模型可以获得局部特征和全局特征。可以将获取局部特征图的行为视为数据增强,将局部区域图像当成增强之后的数据进行分析,所以分别计算局部特征和全局特征各自的分类损失

$$L_C = -\sum_i^K \text{loss}(f_c(\mathbf{F}'_i)) - \text{loss}(f_c(\mathbf{F})) \quad (6)$$

式中, f_c 是计算每一块特征区域属于正确分类概率的置信度函数,损失函数的第1部分是对于所有局部区域的交叉熵损失求和,第2部分是计算全局区域的交叉熵损失。

此外,本文还定义了局部特征和全局特征融合之后的分类损失为

$$L_S = -\text{loss}(f_c(\mathbf{F}, \mathbf{F}'_1, \mathbf{F}'_2, \dots, \mathbf{F}'_K)) \quad (7)$$

最终,将联合损失记为

$$L = \alpha \cdot L_C + \beta \cdot L_S \quad (8)$$

2.2.2 排序损失

在本文中采用多通道注意力选择机制,涉及排序问题,因此文中采取成对排序损失来获得最具有信息量的特征图。通过成对排序损失,可以优化多通道注意力选择模型,在生成注意力特征图时更加关注信息量大的通道以及对应区域,使得当特征图信息量大的时候,该特征图对于正确分类的置信度也大。

在成对排序方法中(Burges 等,2005),学习到排序问题被归结为一个分类问题。即学习一个在一对数值中选择正确数值的二分类器。假设函数 $F(X_i, X_j)$ 只取 $\{1, 0\}$ 中的一个值,其中 $F(X_i, X_j) =$

0 表示 X_i 排在 X_j 之前, $F(X_i, X_j) = 1$ 表示 X_i 排在 X_j 之后。在模型中获得了局部区域分类置信度 $\mathbf{C} = \{C_1, C_2, \dots, C_k\}$,信息量排名前 K 层特征图的信息量 $\mathbf{S}' = \{S'_1, S'_2, \dots, S'_K\}$,希望当 C_i 越大时,其所对应的 S'_i 也能越大,所以将所有对的损失定义为

$$L_R(\mathbf{S}', \mathbf{C}) = \sum_{(i,j): C_i < C_j} f(S'_j - S'_i) \quad (9)$$

式中, f 为 Hinge loss。最终使得当 $C_i < C_j$ 时, $S'_i < S'_j, S'_i = T(A_i)$ 。

损失函数的反向传播可以表示为

$$\begin{aligned} \frac{\partial L_R(\mathbf{S}', \mathbf{C})}{\partial \mathbf{W}_R} = & \sum_{(i,j): C_i < C_j} f'(S'_j - S'_i) \cdot \\ & \left(\frac{\partial L(x)}{\partial \mathbf{W}_R} \Big|_{x=A_j} - \frac{\partial L(x)}{\partial \mathbf{W}_R} \Big|_{x=A_i} \right) \end{aligned} \quad (10)$$

式中, f' 为 Hinge loss 的求导, $\frac{\partial L_R}{\partial \mathbf{W}_R}$ 为 L_R 对 $\mathbf{R}$ 的求导。

2.2.3 总损失函数

综上,该模型的总损失函数定义为

$$L_{\text{total}} = \alpha \cdot L_C + \beta \cdot L_S + L_R \quad (11)$$

式中, α 和 β 是超参数,在本文中设定 $\alpha = \beta = 1$,并采用 Adam(adaptive moment estimation)优化算法对损失函数进行优化。

2.3 整体流程算法总结

本文算法主要是通过对特征矩阵 $\mathbf{F}$ 进行通道信息量的提取和排序,获得信息量较大的特征层。再对特征层进行遮罩操作,得到具有高信息量的局部区域和局部特征。最后将局部特征与全局特征相融合,获得分类结果。FGMAS 算法流程为:

输入:图像 $\mathbf{X}$, batch size 值 m , 超参数 M 、 K 。

输出:对图像 $\mathbf{X}$ 的分类结果。

1) 初始化神经网络参数,预处理图像 $\mathbf{X}$

2) Repeat

对 $\mathbf{X}$ 提取全局特征 $\mathbf{F}$

提取注意力特征图 $\mathbf{A} = \{A_1, A_2, A_3, \dots, A_M\}$

计算每一层信息量 $\mathbf{S} = \{S_1, S_2, S_3, \dots, S_M\}$

排序并获取前 K 层特征图 $\{A'_1, A'_2, \dots, A'_K\}$ 和

信息量 $\{S'_1, S'_2, \dots, S'_K\}$

获得高信息量局部区域 $\{R_1, R_2, R_3, \dots, R_K\}$

计算细粒度分类结果

$P = f_p\{\mathbf{X}, R_1, R_2, R_3, \dots, R_K\}$

计算局部特征和全局特征的分类置信度

$\{C_0, C_1, C_2, \dots, C_k\} = f_c(\{\mathbf{F}, \mathbf{F}'_1, \mathbf{F}'_2, \dots, \mathbf{F}'_K\})$

计算损失函数 L_{total} 并用梯度下降进行反向传播
3) 直到总损失函数的收敛阈值或迭代次数达到上限

4) 将待预测样本输出模型, 获得分类结果。

上述算法中, f_p 为计算分类结果函数, f_c 为计算分类置信度函数。

3 实验与分析

3.1 数据集及评价指标

Kaggle 数据集由 EyePACS 提供, 于 2015 年在 Kaggle 上公开, 共有 35 126 幅训练图像和 53 576 幅测试图像。这些图像由不同型号和品牌的设备拍摄并以不同的分辨率进行储存。医学专家根据疾病的严重程度, 将图像的病变程度诊断为 5 个阶段, 标记为 $\{0, 1, 2, 3, 4\}$, 分别对应健康、轻度、中度、重度和增生性 DR。其中健康图像 25 810 幅, 轻度图像 2 443 幅, 中度图像 5 292 幅, 重度图像 873 幅, 增生性 DR 图像 708 幅。可见, 健康图像远多于具有 DR 病变图像。

Messidor 数据集是 Messidor 项目合作伙伴提供的公开数据集, 包括 1 200 幅视网膜图像, 每幅图像都被诊断为从 0 到 3 共 4 个 DR 阶段, 分别被标记为 $\{0, 1, 2, 3\}$ 。1 200 幅图像中, 无 DR 图像 546 幅, 第 1 阶段图像 254 幅, 第 2 阶段图像 247 幅, 第 3 阶段图像 153 幅。相较于 Kaggle 数据集, Messidor 数据集虽然数据量较小但是没有类间不平衡问题。

1) 数据增强。因为 Kaggle 数据集类间数量极度不平衡, 数据集中具有 DR 病变的图像数量远小于健康的图像数量, 所以在训练过程中采用数据增强, 在解决数据不平衡的同时减小过拟合, 增强方法包括: 将图像进行随机上下水平翻转、 $0^\circ \sim 360^\circ$ 的随机角度中心旋转, 以及 $1 \sim 1.3$ 倍的随机缩放。

2) 图像预处理。Kaggle 数据集的一个挑战就是其图像在分辨率和质量上的巨大差异, 所以需要对其进行预处理, 本文采用 EP (eyes-image preprocessing) 方法 (Ratchasima, 2015) 对图像进行处理, 原始图像有黑色背景, 对其先进行灰度化, 然后设定阈值生成掩膜, 裁剪去除黑色区域并将大小调整为 448×448 像素, 再将原始图像与经过高斯滤波之后的图像进行图像混合加权, 最终得到相同大小并且具有更加明显病变特征的图, 如图 3 所示。

3) 评价指标。在 Kaggle 数据集上, 由于 Kaggle 数据集类间极度不平衡, 即使将所有图像分到第 0 级也能获得较高的正确率, 如果直接使用正确率来进行评估不能客观地表示分类结果, 所以本文使用分类准确率均值 (average of classification accuracy, ACA) 作为实验的评价指标。通过混淆矩阵来计算每一类中正确分类的数量以及错误分类的类别和数量, 然后对混淆矩阵对角线中的正确率进行均值计算。此外, 使用 F1 的微观平均版本和宏观平均版本, 分别记为 Micro F1 和 Marco F1 来评价多分类的结果。F1 作为精度和召回率之间的调和平均值。Micro F1 是对不同类别的个体真阳性、假阳性和假阴性进行汇总之后的 F1 得分; Marco F1 是所有类的 F1 得分的平均值。

在 Messidor 数据集上, 因为 Messidor 数据集没有非常严重的类间不平衡问题, 所以本文使用正确率 (accuracy, Acc) 和受试者工作特征曲线 (receiver operating characteristic curve, ROC) 下的面积大小 AUC (area under the curve) 作为实验的二分类评价指标。

3.2 实验结果和分析

使用 Pytorch 来构建模型并在 GTX2080 Ti GPU 上进行训练和测试。在训练中, 设定 batch size 为 4, 总 epoch 为 30, 学习率为 0.000 1, 使用 Adam 优化器进行优化。损失函数超参数 $\alpha = \beta = 1$ 。

3.2.1 参数设置

针对 M 、 K 这两个超参数, 本文在 Kaggle 数据集上做了分析实验。参数 K 控制着获取局部区域的数量, 当 $K = 0$ 时, 只用全局特征进行分类, 不使用局部特征。 K 越大, 本文所提取的局部特征越多, 但是特征维度也会大大增加, 使得训练和测试所使用的时间也会越长, 本文设置 K 在 $\{4, 3, 2, 1, 0\}$ 的取值范围内进行实验。结果如表 1 所示, 当采用更大的 K 时, 精度变化很小, 但是所用的时间大大增加, 因此最终根据运行时间以及正确率等多方面考虑, 选取 $K = 3$ 进行实验。并且 $K = 3$ 与 $K = 0$ 的对比, 可视为消融实验, 证明了所提方法的有效性。

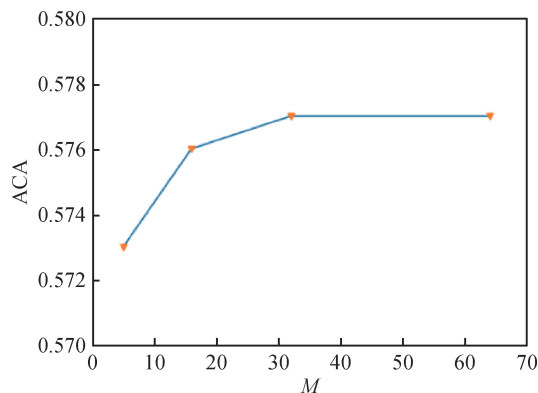
然后, 本文将 M 的取值设置在 $\{64, 32, 16, 5\}$ 范围内, 并进行了实验。 M 表示对特征图降维之后得到的注意力特征图的通道数, 让排序选取的时候有更多的选择, 设置 $M = 5$ 的目的是想探究能否使每一个特征层对应每一类的特征, 但是效果并没有优

表1 参数 K 分析结果Table 1 The analysis results of K

K	ACA	时间/(min/批次)
0	0.557	17
1	0.570	25
2	0.572	35
3	0.577	43
4	0.578	57

注:加粗字体表示本文选用 K 值所对应的结果。

于其他数值。最终根据实验情况如图4分析,本文采取 $M=32$ 来进行实验。

图4 参数 M 分析结果Fig.4 The analysis results of M

3.2.2 在 Kaggle 数据集上的结果和分析

本文在 Kaggle 数据集上进行了实验,和最近的方法 VNXX/LGI (VGGNet with extra kernel/LGI), CKML/LGI (combined kernels with multiple losses network/LGI) (Vo 和 Verma, 2016), BiRA-Net, Bravo 进行了比较,比较结果如表2所示。从表2可以看出,在整体分类准确度均值上,本文方法比其他方法具有更好的结果,达到了 0.577,比其中效果最好的 BiRA-Net 高 0.034。Marco F1 和 Micro F1 的结果也优于其他方法。本文算法的 Micro F1 高达 0.853,表明对于所有类的图像来说,本文方法具有更多的真正例(true positive, TP)。与 BiRA-Net 相比,本文算法能够更好地分辨健康的图像与含有病变的图像,对 0 级图像具有较高的分类精度,而同时大量的 0 级图像导致较高的 Micro F1 结果。

为了证明所提方法对于含有微小病变点的分类是有效的,本文又绘制了混淆矩阵(confusion matrix),

表2 在 Kaggle 上的分类结果对比

Table 2 Comparison of classification results on Kaggle

	ACA	Marco F1	Micro F1
VNXX/LGI	0.512	—	—
CKML/LGI	0.473	—	—
Bravo 等人(2017)	0.505	0.508	0.505
BiRA-Net	0.543	0.573	0.544
FGMAS(本文)	0.577	0.610	0.853

注:—表示暂无该数据,加粗字体为该评价指标下最优结果。

横坐标为预测标签(predicted label),纵坐标为真实标签(true label),如图5所示。由图5中的混淆矩阵中不难发现,本文方法在总体精度上具有较好的结果,尤其是对于最难分类的1级病变,FGMAS具有0.301的正确率,远高于 BiRA-Net 的0.187、CKML/LGI 的0.112 和 VNXX/LGI 的0.190。1级病变的主要症状为有微血管瘤,其难以分类最主要的原因就在于微血管瘤都过于微小且难以察觉,所以大多数方法都会将它分到正常类也就是0级。而本文方法通过获取局部特征,能够更好地提取到局部特征,对于这些微小的病变的察觉和分级有更好的效果。

3.2.3 在 Messidor 数据集上的结果和分析

为了进一步评估性能,将所提出的 FGMAS 网络应用于另一公开数据集 Messidor 进行糖尿病性视网膜病变分级。并与 VNXX/LGI, CKML/LGI, Zoom-in-Net, comprehensive computer-Aided design (Sánchez 等, 2011), Fisher Vector (Pires 等, 2017), Dynamic Shape Features-RFcara (DSF-RFcara) (Seoud 等, 2016), ensemble-based system for microaneurysm detection (ESMD) (Antal 和 Hajdu, 2012) 等方法进行比较。由于 Messidor 和 Kaggle 采用不同的标注尺度 (Messidor: 0—3, Kaggle: 0—4), Wang 等人 (2017) 提出两个二分类任务来实现在该数据集上的算法验证:推荐转诊与不推荐转诊之间的分类,正常和不正常之间的分类。根据是否有明显的黄斑水肿风险,专家将 Messidor 0 级和 1 级定义为不推荐转诊,而 2 级和 3 级定义为推荐转诊。在 Messidor 上分级为 0 也就是无病变的图像定义为正常,否则为不正常。

由于 Messidor 只有 1 200 幅图像,对于训练 CNNs 来说,图像的数量是很少的。所以本文首先

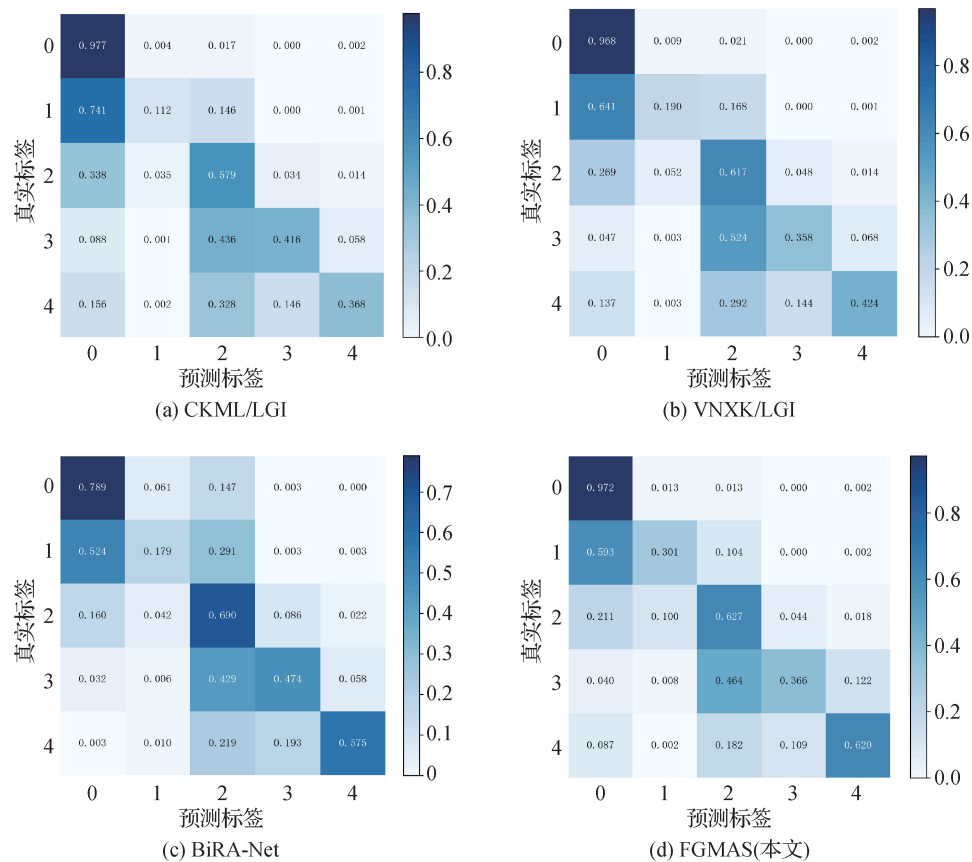


图5 混淆矩阵

Fig. 5 Confusion matrix((a) CKML/LGI;(b) VNXX/LGI;(c) BiRA-Net;(d) FGMA(ours))

在 Kaggle 上预训练 FGMA 模型,之后训练一个全新的二分类全连接层。在训练过程中采用的训练验证数据比为 5 : 1,并训练两个批次,所得分类器用于 Messidor 数据集的二分类任务。实验结果如表 3 和表 4 所示。

表 3 推荐转诊/不推荐转诊的结果			
Table 3 The results of reference/non-reference			
	对比方法	AUC	Acc
推荐转诊/ 不推荐转诊	Fisher Vector	0.863	–
	Zoom-in-Net	0.957	0.911
	VNXX/LGI	0.887	0.893
	CKML/LGI	0.891	0.897
	Comprehensive CAD	0.910	–
	FGMA(本文)	0.962	0.912

注:–表示暂无该数据,加粗字体为该评价指标下最优结果。

从表 3 和表 4 中可以看出,对于正常/不正常分类,本文方法正确率达到了0.909,高于VNXX/LGI

表 4 正常/不正常的结果			
Table 4 The results of normal/abnormal			
	对比方法	AUC	Acc
正常/ 不正常	DSF-RFcara	0.916	–
	Zoom-in-Net	0.921	0.905
	VNXX/LGI	0.870	0.871
	CKML/LGI	0.862	0.858
	ESMD	0.900	0.820
	FGMA(本文)	0.950	0.909

注:–表示暂无该数据,加粗字体为该评价指标下最优结果。

和 CKML/LGI 方法,在 AUC 上也达到了 0.950 的好成绩,高于其他方法。对于推荐转诊/不推荐转诊分类,本文方法达到了 0.912 的正确率和 0.962 的 AUC,相比于其他方法都取得了较好的结果。

4 结 论

本文提出一种有效的基于多通道注意力选择机

制的细粒度分级方法 (FGMAS) 对糖尿病性视网膜病变进行分级。考虑到糖尿病性视网膜病变有些过于细小导致使用一般方法难以正确地辨别, 算法借鉴细粒度分类网络结构来帮助提取到微小病变的局部特征。此外, 由于细粒度分类中局部特征区域的提取非常关键, 本文提出了多通道注意力选择机制优化该网络结构, 以获取更加准确的局部特征区域。本文在两个公开的糖尿病性视网膜病变的眼底数据集上与多个先进的分类算法进行了实验比较, 同时对方法中的关键参数进行了实验分析。实验结果表明, 本文方法在糖尿病性视网膜病变五分级和二分级任务中都取得了较好的性能。

接下来, 将在现有基础上进一步开展深入研究, 对糖尿病性视网膜病变图像中的病变点进行可视化, 进一步研究类别间的差异。此外, 将对病变区域的检测和分割展开研究, 进一步提升糖尿病性视网膜病变的分级精度。

参考文献 (References)

- Antal B and Hajdu A. 2012. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Transactions on Biomedical Engineering*, 59(6): 1720-1726 [DOI: 10.1109/TBME.2012.2193126]
- Bravo M A and Arbeláez Pablo A. 2017. Automatic diabetic retinopathy classification//13th International Symposium on Medical Information Processing and Analysis. San Andres Island, Colombia; SPIE: 105721E. 1-10 [DOI: 10.1117/12.2285939]
- Burges C, Shaked T, Renshaw E, Lazier A, Deeds M, Hamilton N and Hullender G. 2005. Learning to rank using gradient descent//Proceedings of the 22nd International Conference on Machine Learning. New York, USA: ACM: 89-96 [DOI: 10.1145/1102351.1102363]
- Chen L, Zhang H W, Xiao J, Nie L Q, Shao J, Liu W and Chua T S. 2017. SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6298-6306 [DOI: 10.1109/CVPR.2017.667]
- Chen Y W, Wu T Y, Wong W H and Lee C Y. 2018. Diabetic retinopathy detection based on deep convolutional neural networks//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, Canada: IEEE: 1030-1034 [DOI: 10.1109/ICASSP.2018.8461427]
- Decencière E, Zhang X W, Cazuguel G, Lay B, Cochener B, Trone C, Gain P, Ordonez R, Massin P, Erginay A, Charton B and Klein J C. 2014. Feedback on a publicly distributed image database: the messidor database. *Image Analysis and Stereology*, 33(3): 231-234 [DOI: 10.5566/ias.1155]
- EyePACS. 2015. Diabetic-Retinopathy-Detection [EB/OL]. [2021-04-15]. <https://www.kaggle.com/c/diabetic-retinopathy-detection>
- Fundus ophthalmology Group of Chinese Medical Association. 2014. Chinese guidelines for clinical diagnosis and treatment of diabetic retinopathy. *Chinese Journal of Ophthalmology*, 50(11): 851-865 (中华医学会眼科学会眼底病学组. 2014. 我国糖尿病视网膜病变临床诊疗指南. *中华眼科杂志*, 50(11): 851-865) [DOI: 10.3760/cma.j.issn.0412-4081.2014.11.014]
- Huang G, Liu Z, van der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Pires R, Avila S, Jelinek H F, Wainer J, Valle E and Rocha A. 2017. Beyond lesion-based diabetic retinopathy: a direct approach for referral. *IEEE Journal of Biomedical and Health Informatics*, 21(1): 193-200 [DOI: 10.1109/JBHI.2015.2498104]
- Rocha A, Carvalho T, Jelinek H F, Goldenstein S and Wainer J. 2012. Points of interest and visual dictionaries for automatic retinal lesion detection. *IEEE Transactions on Biomedical Engineering*, 59(8): 2244-2253 [DOI: 10.1109/TBME.2012.2201717]
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z H, Karpathy A, Khosla A, Bernstein M, Berg A C and Li F F. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3): 211-252 [DOI: 10.1007/s11263-015-0816-y]
- Rathasima N. 2019. APTOS: Eye Preprocessing in Diabetic Retinopathy [CP/OL]. [2021-04-15]. <https://www.kaggle.com/rathasima/aptos-eye-preprocessing-in-diabetic-retinopathy>
- Salami M J E, Khorshidtalab A, Baali A and Aibinu A M. 2014. Classification of retinal images based on statistical moments and principal component analysis//Proceedings of 2014 International Conference on Computer and Communication Engineering. Kuala Lumpur, Malaysia: IEEE: 92-95 [DOI: 10.1109/ICCCE.2014.37]
- Sánchez C, Niemeijer M, Dumitrescu A V, Suttorp-Schulten M S A, Abramoff M D and van Ginneken B. 2011. Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data. *Investigative Ophthalmology and Visual Science*, 52(7): 4866-4871 [DOI: 10.1167/iovs.10-6633]
- Seoud L, Hurtut T, Chelbi J, Cheriet F and Langlois J M P. 2016. Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE Transactions on Medical Imaging*, 35(4): 1116-1126 [DOI: 10.1109/TMI.2015.2509785]
- Vo H H and Verma A. 2016. New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space//2016 IEEE International Symposium on Multimedia. San Jose, USA: IEEE: 209-215 [DOI: 10.1109/ism.2016.0049]

- Wang Z, Yin Y X, Shi J P, Fang W, Li H S and Wang X G. 2017. Zoom-in-Net: deep mining lesions for diabetic retinopathy detection//Proceedings of Medical Image Computing and Computer Assisted Intervention. Quebec City, Canada: Springer: 267-275 [DOI: 10.1007/978-3-319-66179-7_31]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Zhao Z Y, Zhang K R, Hao X J, Tian J, Chua M C H, Chen L and Xu X. 2019. BiRA-Net: bilinear attention net for diabetic retinopathy grading//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 1385-1389 [DOI: 10.1109/ICIP.2019.8803074]
- Zhou K, Gu Z W, Liu W, Luo W X, Cheng J, Gao S H and Liu J. 2018. Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading//Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Honolulu, USA: IEEE: 2724-2727 [DOI: 10.1109/EMBC.2018.8512828]

作者简介



顾婷菲, 1995 年生, 女, 硕士研究生, 主要研究方向为基于深度学习的医学图像处理。

E-mail: gtf_atoll@163.com



郝鹏翼, 通信作者, 女, 副教授, 主要研究方向为智能图像处理。

E-mail: haopy@zjut.edu.cn

白琮, 男, 副教授, 主要研究方向为多媒体信息处理。

E-mail: congbai@zjut.edu.cn

柳宁, 男, 副研究员, 主要研究方向为多媒体通信与信息处理。E-mail: ningliu@sjtu.edu.cn