南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

本科生毕业设计（论文）

题　　目：基于眼动引导的深度多示例学习的糖尿病视网膜病变检测

姓　　名：侯伊林

学　　号：11912636

系　　别：计算机科学与工程系

专　　业：计算机科学与技术

指导教师：刘 江

2023 年　5 月　8　日

诚信承诺书

1.本人郑重承诺所呈交的毕业设计（论文），是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。

2.除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。

3.本人承诺在毕业论文（设计）选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4.在毕业论文（设计）中对侵犯任何方面知识产权的行为，由本人承担相应的法律责任。

作者签名：侯伊林

2023 年 5 月 8 日

基于眼动引导的深度多示例学习的糖尿病视网膜病变检测

侯伊林

（计算机科学与工程系　指导教师：刘江）

[摘要]：在糖尿病引起的各种眼部疾病中，糖尿病视网膜病变是最常见且致盲风险最高的。通常病变在早期无明显症状，一旦症状显现，患者往往已经错过最佳治疗时期，因此实现对糖网的早期准确诊断具有重要意义。利用计算机技术辅助眼底图像进行大规模筛查糖网可以帮助医生提高早期糖网诊断的精确度，而计算机辅助诊断方法在检测早期糖网的表现并不令人满意。本文主要研究将眼科医生的眼动追踪信息应用到深度学习模型中，从而提高模型对糖网在眼底图像上的分类准确率。

首先，本文采集了 1020 张医生诊断时的眼动注意力图，对比了二值融合和加权融合两种注意力融合方式，并验证了注意力融合能够提高模型分类准确率的有效性。其次，本文将眼动注意力注入深度多示例学习模型中，通过多示例框架对各个注意力区域进行线性学习。将上述数据集以 512*512 的大小输入网络，模型在测试集上的分类准确率达到 63%。在此基础上，本文提出在眼底图像上基于注意力分布建立图结构，使用视觉图卷积（ViG）提取图像特征，强化了节点与节点之间的信息交流，更加拟合医生的诊断策略。将相同的数据集上应用到为加入眼动信息的 ViG 和加入眼动信息的 ViG 上，在测试集上的准确率从 62.5%提升到 74%，有效提升了深度学习模型对早期糖网的诊断效果。

[关键词]：计算机辅助诊断；糖尿病视网膜病变；眼动追踪；深度学习；多示例学习

[ABSTRACT]: Among the various eye diseases caused by diabetes, diabetic retinopathy is the most common with the highest risk of blindness. Usually, the lesions have no obvious symptoms in the early stage, and once the symptoms appear, the patients often have missed the optimal treatment period, so it is of great significance to realize the accurate early diagnosis of diabetic retinopathy. Using computer-aided fundus images for large-scale screening of diabetic retinopathy can help doctors improve the accuracy of early diagnosis, but the performance of computer-aided diagnosis on early diabetic retinopathy is not satisfactory. This paper mainly studies the application of eye movement tracking information of ophthalmologists to the deep learning model, to improve the model's classification accuracy of diabetic retinopathy in fundus images.

Firstly, this paper collected 1020 eye movement graphs of doctors during diagnosis, compared two attention fusion methods, the binary fusion and the weighted fusion, and verified the effectiveness of attention fusion to improve the accuracy of model classification. Secondly, this paper infuses eye-movement attention into the deep multi-instance learning model and uses the multi-instance framework for linear learning of each attention area. The classification accuracy of the model on the test set reaches 63% by inputting the data set 512*512 into the network. On this basis, this paper proposes to establish a map structure based on attention distribution in fundus images and use visual image convolution (ViG) to extract image features, which can strengthen the information exchange between nodes and better fit the diagnosis strategy of doctors. When the same data set was applied to the ViG with eye movement information and the ViG with eye movement information, the accuracy of the test set

was increased from 62.5% to 74%, which effectively improved the diagnosis effect of the deep learning model on the early sugar web.

[Keywords]: Computer-aided diagnosis; Diabetic retinopathy; Eye tracking; Deep learning; Multiple instance learning

目录

1. 绪论

1.1 研究背景及意义

糖尿病（Diabetes Mellitus）是一种以高血糖为特征的代谢性疾病，位列 2019 年世界卫生组织发布的全球十大死亡原因之一。在高基数的人口基础上，加之人口老龄化加剧，我国的糖尿病患病率逐年攀升。根据 2021 年国际糖尿病联盟（International Diabetes Federation，IDF）统计，我国的糖尿病患者人数已达 1.4 亿。在糖尿病引起的各种眼部疾病中，糖尿病视网膜病变（以下简称“糖网”）是最常见也是致盲风险最高的。其成因是慢性进行性糖尿病导致的视网膜微血管渗漏和阻塞从而引起一系列的眼底病变，如微动脉瘤、硬性渗出、棉絮斑、新生血管、玻璃体增值、黄斑水肿甚至视网膜脱离。我国糖尿病患者中的糖网患病率高达 24.7%-37.5%，每年有 300-400 万糖尿病患者因糖网失明。

根据国际糖尿病视网膜病变临床分期标准，糖网依照是否出现新生血管分为非增殖性病变（NPDR）和增殖性病变（PDR）两大类，其中非增殖性病变又分为轻度、中度、和重度三种级别[1]。通常病变在早期无明显症状，一旦症状显现，患者往往已经错过最佳治疗时期。医学专家建议，糖尿病患者应每年定期进行眼底检查，从而避免疾病的进一步恶化。因此，实现对糖网的早期准确诊断具有重要意义。

然而，糖网的早期诊断在临床上依旧面临一些挑战：1.早期糖网在眼底彩照中仅表现为微动脉瘤，临床上漏诊率极高。2.早期糖网的临床诊断对医生专业水平要求较高，医生需要经过专业培训和一定的经验累积才能做出较为准确的诊断。3.我国糖尿病患者群体庞大，而眼科医生数量较少，这导致在大规模筛查糖网时医疗资源紧张的问题。因此，利用计算机技术辅助眼底图像进行大规模筛查糖网可以帮助医生提高早期糖网诊断的精确度，同时可以缓解不平衡的医患供需现状。

1.2 国内外研究现状

1.2.1 糖尿病视网膜病变研究现状

随着计算机技术和医学影像技术的不断发展，国内外许多学者对糖网的计算机辅助诊断方法进行了相关研究。在使用范围较广的深度学习分类架构（如 ResNet、

Inception_V3、DenseNet 和 SE-Net）的基础上，研究者提出了几种用于糖网诊断二分类和多分类算法[2,3]。为了减轻不同糖网类中数据不平衡的现像，一些研究采用了不同的数据增强技术和类加权策略来训练深度学习模型[4,5]。此外，最近有研究者提出了基于注意力的糖网分类模型，以提高糖网检测结果的可解释性。Li 等人[6]提出了一种新的基于疾病间注意力的深度学习框架，通过提取疾病间的关联，对糖网和糖尿病黄斑水肿进行联合分类。早期糖网的临床病变微动脉瘤是眼底图像最小的病变之一，其大小常常小于 125um[7]。一些研究侧重于通过微动脉瘤检测或分割对糖网进行早期检测或诊断，这是糖网检测中一个更困难的挑战[8]。Kandemir 等人[9]提出了一种基于图像分块的多示例学习算法用于诊断糖网，首先将图像分割成不重叠的相等部分，然后将每个区域提取的特征作为示例，建立数据包与示例之间的关系，从而对未知图像进行分类。Quellec 等人[10]同样是基于图像分块的多示例学习思想，建立参考数据集和对应的相关及不相关的图像块，通过不断优化示例和包的权重来计算每个图像块的相关度，最后利用相关度分析对未知图像进行预测。

1.2.2 基于眼动追踪技术的医学影像诊断研究现状

眼动追踪技术已成为心理学[11]、神经科学[12]、教育[13]、工业[14]、生物医学工程[15]等多个领域视觉行为和人类行为的技术手段之一。研究证实，被视为视觉注意的眼动注意力图可以帮助解释临床专家在诊断中的决策。眼动数据主要用于医学图像分析、图像标注和疾病辅助诊断方面。由于现实中医学图像标注的成本较高，研究人员提出了一些基于眼动追踪的标注方法代替人工标注。Stember 等人[16]试图将眼动追踪标注作为病变（脑膜瘤和乳房肿块）和器官（肝脏、肾脏和心脏）分割任务的训练标签。从眼球追踪标注中学习的分割结果与从人工标注中学习的分割结果的平均 Dice 相似系数超过 0.85。此外，他们的另一项工作[17]提出了一个新颖的图像标注框架，将眼动跟踪和语音识别结合在一起，以定位每个 MRI 扫描的脑肿瘤。这种注释方法在标记病变位置时可以获得 92%的准确率，通过监督学习，最终的检测网络可以达到 85%的准确率。同样，Saab 等人[18]为医学图像诊断任务提出了一个观察性监督框架。他们的实验验证了使用临床专家的凝视信息作为唯一的监督标签取得了接近人工标签的性能。凝视数据作为辅助监督任务，进一步使他们的 CAD 模型的性能得到了较大的提升。

1.3 主要研究内容

目前，大部分检测糖网的方法只能在较为严重的病变眼底彩照和正常眼底彩照之间取得较理想的分类性能，由于早期糖网病变的眼底彩照与正常情况极为相似、病灶微小且位置零散不固定，对于早期的糖网检测一直未能达到理想效果。因此，本文通过眼动追踪技术获取医生诊断时的注意力，将医生注意力作为临床先验知识注入深度学习模型，实现基于眼底图像的早期糖网的精确诊断。本文主要研究内容与贡献如下：

（1） 构建基于眼动的深度多示例学习模型。为了提高深度学习模型对早期糖网诊断的精确度，本文提出了一种基于眼动的深度多示例学习模型，通过提取医生阅片时的眼动注意力，将每张图像的所有注意力区域作为多示例的一个包，在多示例学习的框架下通过深度学习对图像进行二分类。

（2） 构建基于眼动的图卷积神经网络模型。在深度多示例学习架构中，眼动部分作为重点区域在网络中学习，但眼动区域之间的关联性被忽略。为了强化眼动区域之间的关联性，我们提出对一张图像中的多个眼动区域建立图结构，使用视觉图神经网络（Vision GNN, ViG）提取图像特征并进行分类任务。ViG 模型在测试集上的表现相比与深度多示例学习得到了显著提升。

2. 论文相关背景知识

2.1 眼动实验配置

眼睛是心灵的窗户，也是大脑与外部环境进行信息交流的窗口。当人们进行对图像的视觉观察时，仅仅用言语和文字很难准确地表达出自己的兴趣所在，而眼睛注视点的运动状态能够客观地记录大脑的认知过程和对图像感兴趣的区域[19]。当我们注视吸引我们注意的某个特征时，我们并不是按照该物体或场景来感知的，我们会使用通过注视点获取的信息在我们的视觉皮层进行图像的构建，大脑会将连续的注视点获得的视觉图像结合到物体或场景上。因此，使用眼动追踪技术捕获观察者的眼动模式能够间接表达出进行视觉观察过程中的思维认知流程和注意力分布。

本文的眼动实验部分使用的设备是 Tobii Pro Spectrum 屏幕式眼动仪，该设备由尺寸为 55cm×18cm×6cm 的眼动仪硬件和 24 英寸的高清显示器组成，确保被试者在不被干扰的条件下在任何位置都能获得全面的观看体验。眼动仪在使用时需要通过网线连接外部计算机，并在计算机中安装制造商提供的应用程序（即 Tobii Pro Lab），便于我们通过简单的参数设置快速构建实验环境。本文的眼动实验采集由两名专业的眼科医生协助完成。在进行糖网诊断之前，眼科医生需要坐在屏幕前固定的椅子上，保持视线垂直于屏幕，调整眼睛到屏幕的距离在 55 ~ 60cm 的范围内，以便眼动仪准确地捕捉眼球的运动。在此之后，需要对每个眼科医生进行标准的眼动追踪校准测试，以适应个人的视觉特征。最后，眼科医生观察显示器上出现的眼底图像进行糖网诊断，我们采集 60Hz 频率下眼科医生的眼动追踪数据。实验完成后，我们将眼动追踪数据保存在计算机上，用于生成眼动凝视图，进行进一步的分析和使用。

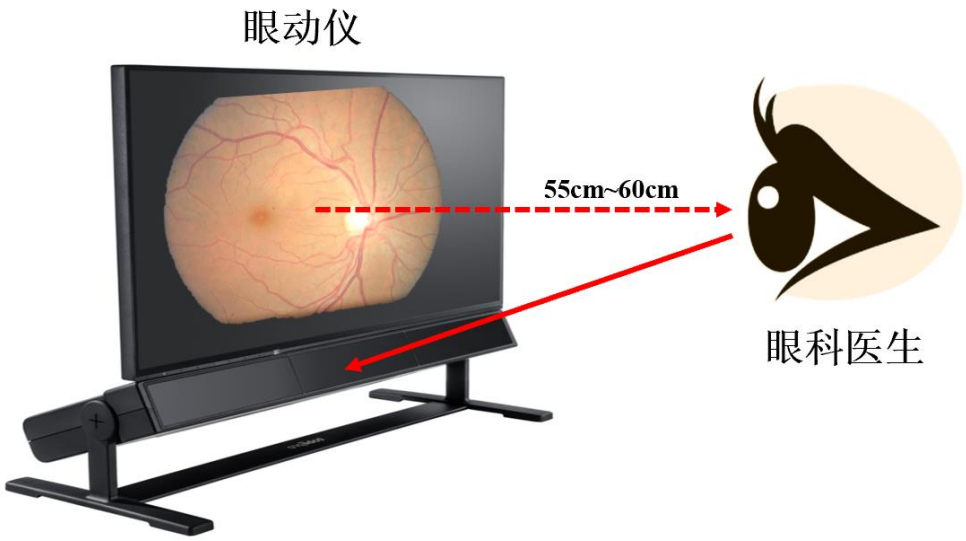


图 1 眼动仪示意图

2.2 深度学习基础

2.2.1 卷积神经网络

本文主要研究基于深度学习的糖网自动分类模型，用于辅助医生的诊断。深度学习模型是近年来进行图像处理最热门的技术，其强大的性能在图像分类任务中表现卓越。以卷积神经网络（CNN）为代表的各种深度学习模型虽然网络结构不同，其基本单元和组成是相近的。下面将简要介绍卷积神经网络的基本结构。

CNN 主要由这五类层构成：输入层（Input layer）、卷积层（Convolutional layer）、线性整流层（Rectified Linear Units layer，即激活函数）、池化层（Pooling layer）和全连接层（Fully-Connected layer）。

卷积层是 CNN 最重要的层次，通过对输入数据进行卷积操作，提取输入的不同特征。图像卷积一般通过滑动窗口实现，卷积运算本质上就是在滑动窗口和输入数据的局部区域间做点积。通过不断堆叠这一操作，即可实现对原始图像不同深度的特征提取。将原始图像记为二维函数$f(x,y)$，使用卷积核$g(x,y)$对该图像进行卷积运算，得到的输出记为$\mathrm{h}(x,y)$，该过程可表示为

$$h(x,y) = f(x,y) * \mathrm{g}(\mathrm{x},\mathrm{y}),$$

由于卷积计算是线性的，而实际中大部分问题是非线性的，为了拟合更复杂的特征，需要引入激活函数增强网络的特征描述能力。三种常用的激活函数分别是 Sigmoid、Tanh 和 ReLU。Sigmoid 函数由于其单增以及反函数单增等性质，常被用作神经网络的阈值函数，将变量映射到（0, 1），可以用来做二分类。双曲正切，将变量映射到（-1, 1）。Tanh 是 Sigmoid 的变形，与 Sigmoid 不同的是，Tanh 是 0 均值的。因此，实际应用中，Tanh 会比 Sigmoid 更好。修正线性单元，它的作用是如果计算出的值小于 0，就让它等于 0，否则保持原来的值不变，克服了前两个激活函数会出现的梯度消失问题。

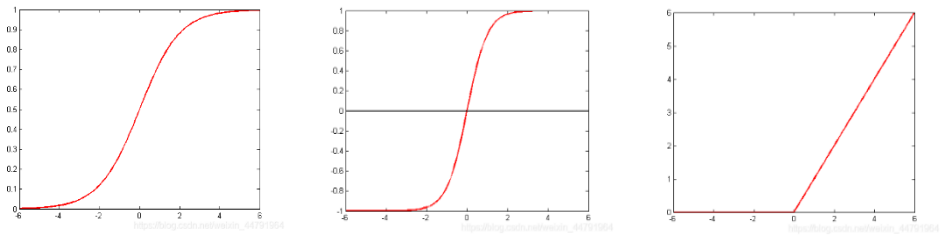


图 2 三种常见激活函数图像

池化层的主要目的是通过下采样对提取的特征信息进行进一步的抽象，能够降低网络压力，减少计算量。池化操作能够保留图像中最重要的特征，降低维度，去除冗余的信息。池化主要包括最大化池化和平均池化两种方式。

全连接层作为网络的最后一层，其特点是将上一层的每个结点与当前层的所有节点相连，从而把网络提取导的特征综合起来。由于全连接层连接的结点数量最多，其参数也是最多的。在分类网络模型中，全连接层一般出现在最后一层利用所有的特征

推断模型的预测结果。

2.2.2 图神经网络

传统的深度学习方法在提取欧式空间数据特征的效果令人惊叹，但在实际应用场景中许多数据来自非欧式空间，此时传统的深度学习方法在处理数据的表现常常不尽如人意。在许多场景中，图结构可以用来表达数据样本中各个节点间的相互依赖关系。图神经网络[20]（GNN）就是将图结构和神经网络结合，利用图结构中的信息传递提取样本特征。

图神经网络的计算过程主要是聚合相邻的节点。每个节点除了包含自身的信息，还包含相邻节点的信息，而随着网络深度的增加，一个节点不仅包含相邻的信息，还有相邻的相邻，以此扩展下去。如图，以具有多图卷积层的卷积神经网络为例。图卷积层通过聚合相邻节点的特征信息来封装每个节点的隐藏表示。特征聚合后，对结果输出进行非线性变换，通过堆叠多层，每个节点的最终隐藏表示接收来自更远邻居的消息。

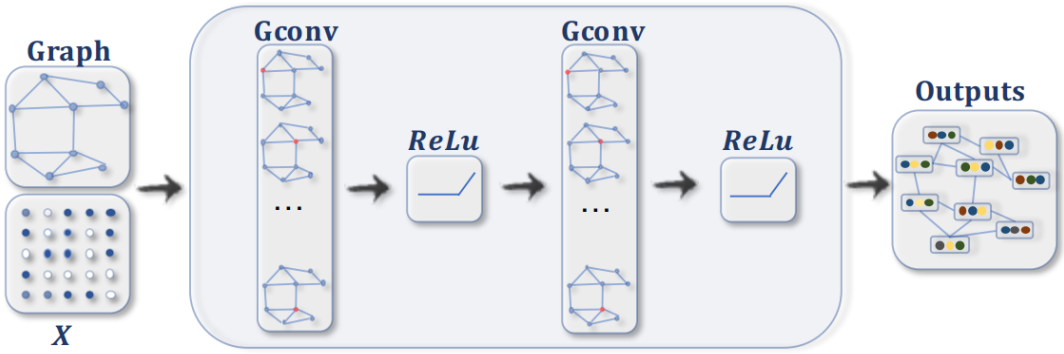


图 3 图卷积网络架构

3. 眼动数据采集与融合实验

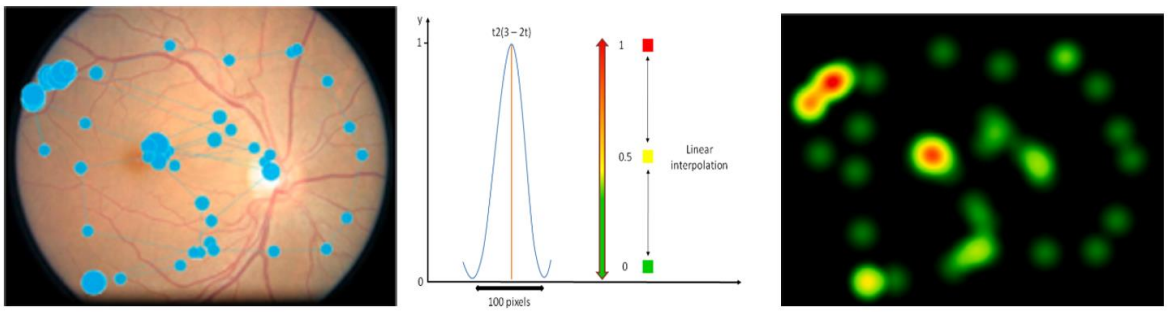
为了获取质量较高的眼动追踪数据，使得深度学习模型能有效学习医生阅片时的注意力，本文规范化了眼动数据采集的实验流程，并进行预实验证明医生的眼动数据能够提升深度学习对糖网诊断的准确率。

3.1 眼动数据采集实验流程

本文从 IDRID 等公共数据集选取正常和早期糖网眼底图像，首先对眼底图像进

行了清洗和归一化处理。为了降低图像中无意义的内容比例，使眼底内容的比例最大化，我们对每张图像进行裁剪并缩放到统一的尺寸，即 800×800。然后将标准化后的眼底图像导入 Tobii Pro Lab。在医生阅片过程中，每一张眼底图像以全屏模式显示在眼动仪显示屏中央，同时眼动仪不断记录被试者的眼球运动信息。在诊断完成后，眼动追踪数据(包括扫视点、注视点、注视时间和眼睛扫描路径)存储在计算机中。其中，注视点可以用来生成眼动凝视图。此外，眼动凝视图上的每个注视点使用高斯曲线的近似值来实现注视点周围值的分布，默认的多项式函数是 t2(3-2t)，再赋予不同的颜色，使得每个注视点的颜色更为平滑，得到最终的眼动热力图（如图 4）。为了减轻视觉误差的影响，我们调整函数的半径为 55 像素。

经过眼动数据采集，最终获得 1020 张眼动数据样本，将样本按原始数据是否为糖网分成两组。其中 499 张正常数据的眼动热力图，521 张早期糖网的眼动热力图，每张眼动热力图与眼底彩照原图一一对应，尺寸比例相同。



(a) 眼动凝视图↵　(b) 高斯曲线拟合赋值颜色↵　(c) 眼动热力图↵

图 4 眼动热力图生成原理

3.2 眼动数据融合实验

3.2.1 眼动数据融合方式构建

为了将眼动数据包含的医生注意力先验知识融入神经网络模型，我们构建了两种注意力融合方式。首先，我们将导出的眼动热力图与眼底彩照原图叠加，去噪，得到融合后的二值化眼动图，从而获得医生诊断过程中的主要关注区域。同时，我们在二

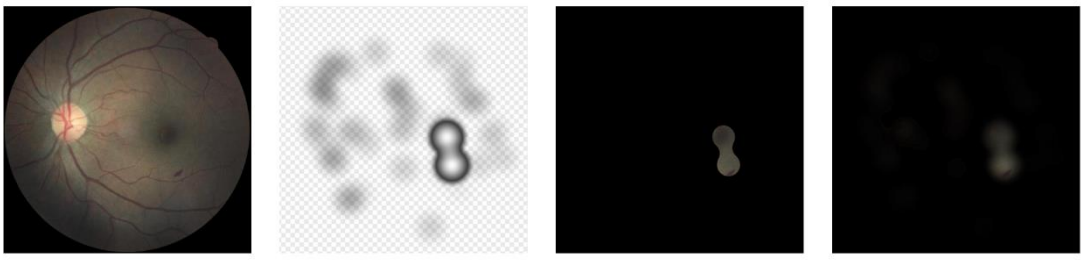


图 4 眼动融合示意图

值化融合眼动图的基础上加入权重，得到加权融合眼动图。二值化融合眼动图和加权融合眼动图如图所示。加权融合眼动图被用于神经网络学习医生的诊断过程中对病灶的关注度，从而提升模型对糖网的检测能力。

3.2.2 眼动数据融合有效性证明

构建好眼动数据的融合方式后，我们设计了一个实验来验证学习医生的注意力是否有助于提高模型对糖网诊断的精确度。在本实验中，我们使用了四个经典分类网络（ResNet18、ResNet34、ResNet50 和 Inception_V3）分别对原图、眼动热力图、二值化融合图、和加权融合图进行糖网分类准确率的预测。为了保证两类数据比例近似 1：1，实验中使用的糖网数据总共 392 张，其中训练数据 297 张，包括正常眼底 150 张、早期糖网 147 张，测试数据 95 张，包括正常眼底 50 张、早期糖网 45 张。所有使用的图片大小均为 512×512。

表 1　眼动数据融合有效性实验结果

模型	图像尺寸	准确率			
		原图	眼动热力图	二值融合	加权融合
Resnet18	512	69.47%	76.84%	61.05%	77.89%
Resnet34	512	53.68%	69.47%	61.05%	73.68%
Resnet50	512	55.78%	75.78%	60.00%	65.26%
Inception_v3	512	56.84%	72.63%	60.00%	69.47%

眼动数据融合有效性实验的实验结果如表 1 所示。加权融合眼动图和眼动热力图对糖网的检测准确率高于二值融合图和眼底图像原图对糖网的准确率，其中，当加权融合图使用 ResNet18 模型分类时的准确率最高，为 77.89%。实验证明了将医生的注意力数据注入深度学习模型中学习可以有效提高模型对糖网的检测能力。

4. 基于眼动的深度多示例学习糖尿病视网膜病变检测方法

4.1 基于眼动的深度多示例学习网络结构

基于眼动数据能够提升模型准确率的前提，本文提出了一种基于眼动的深度多示例学习糖网自动分类模型，其网络架构如图 6 所示。该模型由多示例学习框架和卷积神经网络内核组成。

多示例学习最早由 Dietterich 等人[21]在药物活性预测中引入。在多示例学习中，

训练样本是由多个示例组成的包，包是有概念标记的，但示例本身没有概念标记。如果一个包中至少包含一个正例，那么该包是一个正包，否则为反包。由此可见，多示例学习的原则与医生诊断糖网的过程是极其相似的。 传统的多示例学习模型通常将整张图片切割成 16（或 64）小块，每一个小块对应一个示例，那么一张图片的所有小块可以构成一个包。本项目将多示例思想与眼动技术融合，将加权融合眼动图中的所有眼动区域（即医生的关注区域）截取出来，每个区域为一个示例，一张图片的所有眼动区域为一个包。对于一个示例来说，若该部分包含病灶区域（即微血管瘤），则该示例为一个反例；反之若不包含病灶区域，则为正例。若整张图片包含一个有病灶的小块，那么这张图片将被诊断是糖网；若整张图片不包含任何一个有病灶的小块，那么我们视这张眼底图正常。

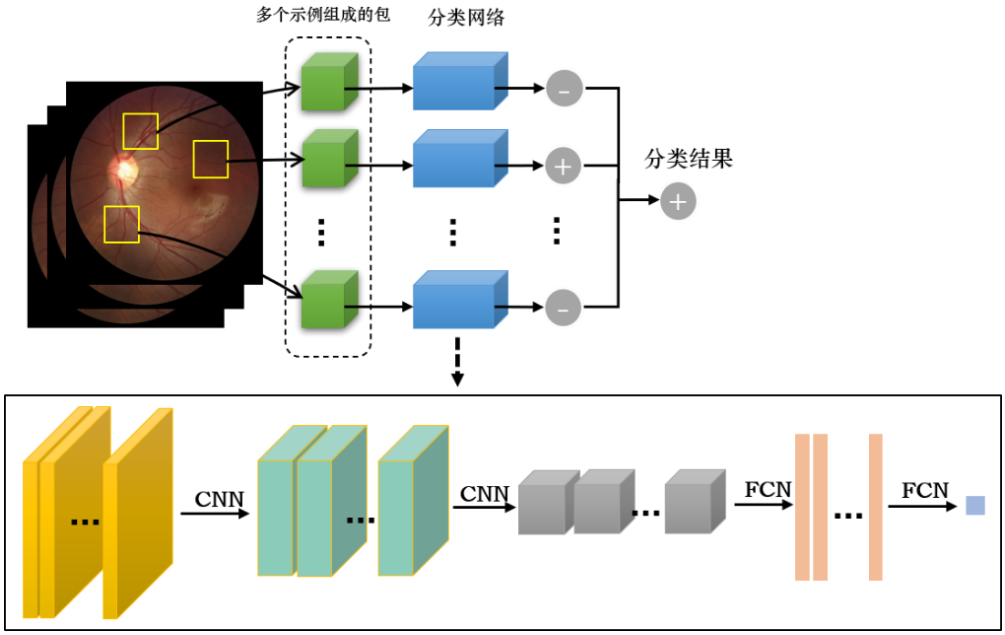


图 5 基于眼动的深度多示例学习模型网络架构

深度多示例学习模型中的分类网络主要有四层结构。第一层卷积神经网络（CNN）为数据输入层，将图像作为包输入，同时对图像进行归一化等预处理。第二层为卷积计算层，分割图像作为包中的样例，同时运用卷积计算提取图像的各部分特征。第三层由全卷积神经网络（FCN）组成，通 过将特征空间映射到样例的图像空间，生成样例的特征向量。第四部分由全连接层组成，得到样例的分类结果 0/1（正常或异常）。

4.2 实验与分析

我们使用上文提到的眼底彩照数据集（1020 张）和眼动数据集（1020 张）完成深度多示例学习实验。其中训练集 620 张，验证集 200 张，测试集 200 张，数据集

具体构成如表 2 所示。

表 2 数据集组成

	眼底彩照	眼动热力图	训练集	验证集	测试集
糖网 0	499	499	299	100	100
糖网 1	521	521	321	100	100

具体实验环境如表所示。

表 3 实验环境配置

操作系统	Linux
CPU	Intel(R) Xeon(R) Platinum 8375C
GPU	NVIDIA GA102GL [RTX A6000]
内存	1.0T
编译器	Vim
仿真语言与框架	Python+Pytorch

训练过程中，使用 ReLU 激活函数，使用 Adam 优化器进行梯度优化，学习率为 5e-4，权值衰减 10e-5，批次大小为 16，训练 100 次迭代。

为了客观评价方法的性能，本文主要以该模型分类的准确率作为模型的评价指标。真阳性 TP（True Positive）是预测为正且实际为正的病灶数，假阳性 FP（False Positive）是预测为正且实际为负的病灶数，真阴性 TN（True Negative）是预测为负且实际为负的病灶数，假阴性 FN（False Negative）是预测为负且实 际为正的病灶数。准确率（Accuracy）可表示为:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}。$$

4. 3 实验结果与分析

该模型在测试集上的准确率最高为 63%，结合模型在训练数据集的表现较好，可判断模型为过拟合。

在原定实验计划中，本文将深度多示例学习作为主要模型进行研究。由于模型实验结果较差，本文在多示例学习的基础上拓展了新方法进行对糖网的自动分类。本文目前以新方法为主要实验完成对糖网的检测，本节实验结果将作为对照组出现在后续实验中。

5. 基于视觉图神经网络的糖尿病视网膜病变检测方法

5.1 基于眼动的视觉图神经网络结构

在以上的深度多示例学习架构中，图像中的眼动区域作为重点区域在网络中学习，但每个眼动区域的关联信息和顺序信息被忽略了。图是一种广义的数据结构，将图像视为图形对于视觉感知来说往往更加灵活和有效，因此，我们提对一张图像中的多个眼动区域建立图结构，使用视觉图卷积（Vision GNN, ViG）[22]体系结构提取图像特征，并在此基础上加入医生注意力信息，优化网络性能，网络结构如图 7 所示。ViG 模型

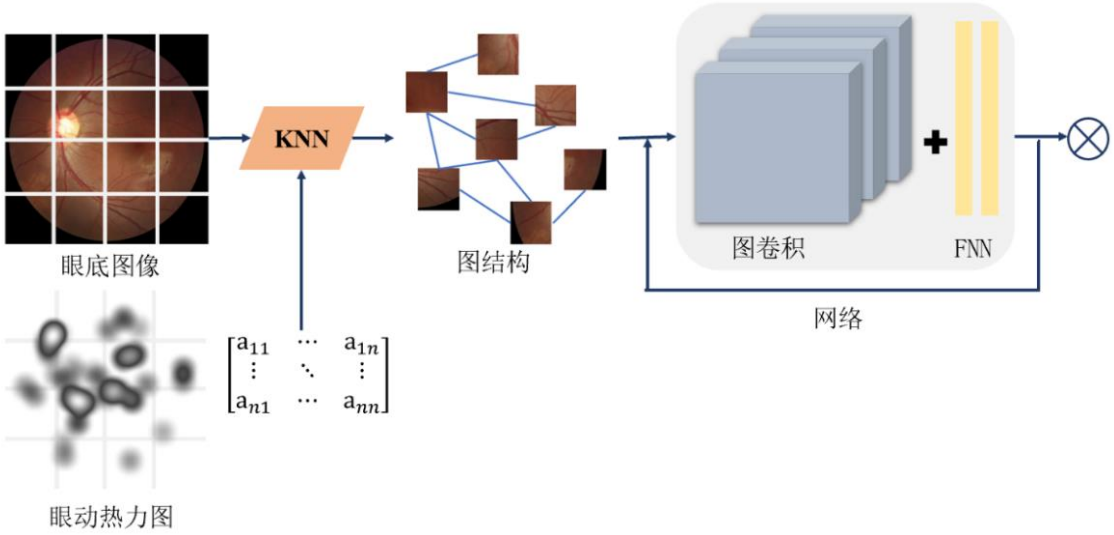


图 6 基于眼动的视觉图卷积网络架构

可以实现所有节点之间的信息变换和交换，而这一行为模式与眼科医生的真实诊断过程更加匹配。

对于一张眼底图像，首先将图片划分为 n*n 个块，然后进行特征变换得到每一个块对应的特征x_i，因此有：$X = [x_1, \dots, x_n]$。这些特征可以看作是一组无序的节点，表示为$V = [v_1, \dots, v_n]$。对于每个节点找到最近的 K 个邻居$\mathcal{N}(v_i)$，然后加入一条有向边e_{ij}从v_j到v_i，所有的边表示为$\mathcal{E}$。这样我们得到了一个图$\mathcal{G} = (\mathcal{V}, \mathcal{E})$。通过将图像视为图数据，可以利用图卷积神经网络提取特征。

与此同时，我们对眼底图像对应的眼动热力图做相同的分块操作，并将分块后的眼动热力图抽象成一个 n*n 大小的二维注意力矩阵，记为$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$，$a_{ij}$为眼动热力图对应小块的平均值计算。将注意力矩阵与原眼底图像相乘意味着注入了注意力权重，因此有：$X = A * X$。进行 KNN 计算时模型会对注意力区域，也就是疑似

病灶区域，做更多关注。

图卷积层通过聚合相邻节点的特征来实现节点之间的信息交换。本文在图卷积前后应用线性层，将节点特征投影到同一个域，增加特征多样性，在图卷积后插入一个非线性激活函数以避免层坍塌。图卷积模块可表示为：

$$Y = \sigma(GraphConv(XW_{in}))W_{out} + X$$

为了进一步提高特征转换能力并缓解过平滑现象，在每个节点上使用前馈网络（FFN）。FFN 模块是一个简单的多层感知机，包含两个完全连接的层。FFN 模块可表示为：

$$Z = \sigma(YW_1)W_2 + Y$$

图卷积模块和 FFN 模块叠加而成的 ViG 模块是网络的基本构建单元，ViG 随着层的叠加能够保持特征多样性，学习出判别性的表征。

5.2 实验与分析

训练过程中，使用 ReLU 激活函数，使 SGD 优化器进行梯度优化，初始学习率为 1e-3，权值衰减 10e-4，动量因子为 0.9，批次大小为 32，训练 100 次迭代。在网络训练前，初始化学习率为 1e-3，若验证机损失在十个迭代内不下降，则将学习率乘以 0.1，学习率最低不得低于 1e-5。

5.3 实验结果与分析

本研究首先搭建了未加入医生注意力的 ViG 框架，该模型在测试上的准确率最高达到 62.5%。将眼动注意力加入模型后，使用相同参数和相同数据集训练模型，模型在测试集上的准确率最高达到 74%。对比传统深度多示例学习和未加入注意力的 ViG 模型，模型的分类性能都得到了大大的提升。

实验证明，眼动注意力能够帮助 AI 模型在较复杂的任务学习中提升性能，同时证明基于图的深度学习模型在复杂非线性的特征提取任务表现更出色。

6. 总结与展望

本文通过使用眼动追踪技术制作了眼科医生诊断糖网时的眼动注意力数据集，数据量为 1020 张。在实验中，眼动仪能够采集医生的阅片诊断过程、诊断依据、以及诊断结论。基于此数据集，探究了两种眼动注意力与原始图像的融合方式，并对于不同融合方式在四种基础分类模型上进行了实验，选择出分类准确率最高的加权融合方式。实验证明基于眼动多模态数据能够提高 AI 模型诊断疾病的准确性。

在医疗资源紧张、标注成本较高的情况下，将医学先验知识与 AI 模型相结合具

有很大意义。本文设计了两个基于眼动的深度学习框架完成早期糖网的二分类任务。第一个框架是传统的多示例学习模型，将眼动区域对应的原始图像部分作为多个示例直接输入卷积神经网络学习。第二个框架是基于图结构的视觉图卷积神经网络模型，眼动注意力作为一个权值矩阵叠加到已经输入网络的原始输入中。后者的图结构在特征提取上更加灵活，图卷积强化了各个节点间的信息交换，也因此在测试集上表现更好，相比于多示例学习模型 63%的准确率，达到了 74%的准确率。

参考文献

[1] 姚毅,赵军平,马志中等.糖尿病眼底病防治指南[J].中国实用眼科杂志,2001(02):83-95.

[2] Tao L， Yg A， Kai W , et al. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening - ScienceDirect[J]. Information Sciences, 2019, 501:511-522.

[3] Zhang W， Zhong J， Yang S , et al. Automated Identification and Grading System of Diabetic Retinopathy Using Deep Neural Networks[J]. Knowledge-Based Systems, 2019, 175(JUL.1):12-25.

[4] Pratt H， Coenen F， Broadbent D M , et al. Convolutional Neural Networks for Diabetic Retinopathy[J]. Procedia Computer Science, 2016.

[5] Zhou Y， Wang B， He X , et al. DR-GAN: Conditional Generative Adversarial Network for Fine-Grained Lesion Synthesis on Diabetic Retinopathy Images.[J]. IEEE Journal of Biomedical and Health Informatics, 2020.

[6] Li X， Hu X， Yu L , et al. CANet: Cross-disease Attention Network for Joint Diabetic Retinopathy and Diabetic Macular Edema Grading[J]. 2019(5).

[7] Kumar S， Adarsh A， Kumar B , et al. An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation[J]. Optics & Laser Technology, 2020, 121:105815-.

[8] Vm A， Ssk A， Uk B . Automated Microaneurysms Detection for Early Diagnosis of Diabetic Retinopathy: A Comprehensive Review[J]. Computer Methods and Programs in Biomedicine Update, 2021.

[9] Kandemir M，Hamprecht F A．Computer-aided diagnosis from weak supervision: a benchmarking study［J］．Computerized Medical Imaging and Graphics，2015，42: 44-50．

[10] Quellec G，Lamard M，Abrmoff M D，et al． A multiple-instance learning framework for diabetic retinopathy screening［J］． Medical Image Analysis，2012，16(6) : 1228-1240．

[11] Sterna R， Cybulski A， Igras-Cybulska M , et al. Psychophysiology, eye-tracking and VR: exemplary study design[C]// 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE, 2021.

[12] Stéphanie Philippe， Lourdeaux D， Leroy L , et al. Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review[J]. International Journal of HumanComputer Interaction, 2022, 38(9):801-824.

[13] Jarodzka H， Skuballa I， Gruber H . Eye-Tracking in Educational Practice: Investigating Visual Perception Underlying Teaching and Learning in the Classroom[J]. Educational Psychology Review, 2021, 33(4):1-10.

[14] Martinez-Marquez D， Pingali S， Panuwatwanich K , et al. Application of Eye Tracking Technology in Aviation, Maritime, and Construction Industries: A Systematic Review[J]. Sensors (Basel, Switzerland), 21(13):4289.

[15] Oliveira J S， Franco F O， Revers M C , et al. Computer-aided autism diagnosis based on visual attention models using eye tracking[J]. Scientific Reports.

[16] Stember J N， Celik H， Krupinski E , et al. Eye Tracking for Deep Learning Segmentation Using Convolutional Neural Networks[J]. Journal of Digital Imaging, 2019(3).

[17] Stember J N， Celik H， Gutman D , et al. Integrating Eye-Tracking and Speech Recognition Accurately Annotates MRI Brain Images for Deep Learning: Proof of Principle[J]. 2020.

[18] Khaled Saab, Sarah M. Hooper, Nimit S. Sohoni, Jupinder Parmar, Brian Pogatchnik, Sen Wu, Jared A. Dunnmon, Hongyang R. Zhang, Daniel Rubin, Christopher Ré, Observational supervision for medical image classification using gaze data, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 603–614

[19]李双. 乳腺钼靶诊断中的视觉感知研究[D]. 杭州电子科技大学, 2011.

[20] Wu Z， Pan S， Chen F , et al. A Comprehensive Survey on Graph Neural Networks[J]. IEEE transactions on neural networks and learning systems, 2021(1):32.

[21] Dietterich T G，Lathrop R H，Lozano-Pérez T． Solving the multiple instance problem with axis-parallel rectangles［J］． Artificial Intelligence，1997，89(1-2) : 31-71．

[22] Han K, Wang Y, Guo J, et al. Vision gnn: An image is worth graph of nodes[J]. arXiv preprint arXiv:2206.00272, 2022.

致谢

行文至此，忽觉光阴似箭。
感谢我的导师刘江教授和姜泓羊博士后对我的毕设和科研学习的指导。
感谢我的父母生我养我成人。
感谢我的朋友们与我相伴成长。
感谢美食、音乐、羽毛球。
感谢自己，祝我们永远勇敢且自由。