受量子理论启发的青光眼筛查方法

郭璠[1]，刘文韬[2]，程欣[2]，孙蔓苓[2]
（1.中南大学人工智能学院 长沙 410083；2.中南大学自动化学院 长沙 410083）

摘要：现有的基于卷积神经网络（CNN）的青光眼筛查方法，因局部感受野的限制，存在全局建模能力弱、对位置信息不敏感的问题。而视觉 Transformer 则存在计算复杂度高、在中小规模数据集上易过拟合的不足。为此，本文针对青光眼筛查问题，提出一种受量子理论启发的处理方法。该方法的核心是设计量子启发式动态锚点双向特征交互机制，通过动态生成复数锚点、相位－振幅联合调制以及规范连接驱动的特征聚合，在降低计算复杂度的同时，实现高效的局部－全局特征融合。此外，所提方法还引入全局上下文信息增强模块与自交互模块，有效提升关键特征的定位能力和应对尺度变化的泛化能力。在 AIROGS 等公开数据集上的实验表明，所提方法在参数量与计算量更低的条件下，取得了相对较优的青光眼筛查性能。

关键词：青光眼筛查；量子理论；动态锚点；特征交互；特征融合
中图分类号：TP391.41
文献标志码：A

Glaucoma Screening Method Inspired by quantum theory

Guo Fan[1], Liu Wentao[2], Cheng Xin[2], Sun Manling[2]
1. School of Artificial Intelligence, Central South University, Changsha 410083
2. School of Automation, Central South University, Changsha 410083

Abstract: Existing glaucoma screening methods based on Convolutional Neural Networks (CNNs) suffer from weak global modeling capability and insensitivity to positional information, owing to the limitation of local receptive fields. Meanwhile, Vision Transformers suffer from high computational complexity and are prone to overfitting on small- and medium-scale datasets. To address these issues, this paper proposes a quantum theory-inspired method for glaucoma screening. The core of the proposed method is a quantum-inspired dynamic anchor bidirectional feature interaction mechanism, which dynamically generates complex-valued anchors, performs joint phase-amplitude modulation, and conducts canonical connectivity-driven feature aggregation. This mechanism achieves efficient local-global feature fusion while reducing computational complexity. In addition, a global context enhancement module and a self-interaction module are introduced to effectively improve the localization of key features and the generalization ability for scale variations. Experiments on public datasets such as AIROGS demonstrate that the proposed method achieves relatively good performance in glaucoma screening with fewer parameters and lower computational cost.

Key words: Glaucoma screening, Quantum theory, Dynamic anchors, Feature interaction, Feature fusion

0 引言

青光眼疾病已经发展成为世界第二大不可逆的致盲眼科慢性疾病[1]。目前针对青光眼筛查问题，基于深度学习的筛查算法已成为解决该问题的主流方法。此类方法采用数据驱动方式自动完成特征工程与分类建模，在数据充足情况下分类性能往往优

收稿日期：XXXX–XX–XX；**修回日期：**XXXX–XX–XX
通信作者：郭璠，guofancsu@163.com
基金项目：湖南省自然科学基金资助项目（No.2023JJ30697,No.2025JJ50396）；长沙市自然科学基金资助项目（No.kq2208286）
Foundation Items：The Natural Science Foundation of Hunan Province (No.2023JJ30697, No.2025JJ50396), The Natural Science Foundation of Changsha (No.kq2208286)

于传统手工特征建模。基于深度学习的视觉处理方法主要包括CNN系列网络方法和Transformer系列网络方法[2]。这两类网络的典型代表性工作有：德国亚琛工业大学的RWTH-CuP团队[3]采用Vision Transformer (ViT) 或其变种作为主干网络进行青光眼筛查，北京协和医院的PUMCH-eye团队[4]采用DenseNet-121模型捕捉血管走向、黄斑位置等全局上下文信息。在网络架构方面，He等人[5]提出用于图像识别的深度残差学习网络，Liu等人[6]基于Vision Transformer提出视觉Transformer架构Swin Transformer，Yu等人[7]提出PoolFormer模型家族中的平衡型版本PoolFormer-S24，以及Meta AI团队开发了全卷积掩码自动编码器框架ConvNeXt V2[8]。此外，2024年提出的RMT (Retentive Networks Meet Vision Transformers) 是一种受自然语言处理领域RetNet启发的视觉骨干网络，旨在解决Vision Transformer (ViT)缺乏显式空间先验和计算复杂度高的问题，RMT-S是该系列中的标准版本配置[9]。但上述属于CNN和Transformer网络的这些方法仍受限于网络架构固有的根本性约束。例如，CNN网络方法主要依靠卷积层来提取局部空间特征，其全局建模能力弱且对位置信息不敏感。Transformer网络方法基于自注意力机制，适合需要全局语义推理的任务，然而自注意力机制的时间复杂度远高于卷积操作，且不具备局部归纳偏置，因此在中小规模数据集上易过拟合。

为了克服上述问题，本文提出一种受量子启发式动态锚点双向特征交互机制。该机制在控制计算量、参数量的同时，能达到相对较优的青光眼筛查效果。此外，所提方法还引入了两个即插即用模块，即全局上下文信息增强模块和自交互模块。前者可以利用全局一致性和语义一致性来提升模型对于关键特征的检测定位能力，后者通过对多尺度特征信息的提取，极大地增强了模型自适应处理不同尺度变化样本的泛化能力。具体来说，本文提出的量子启发式锚点特征交互机制主要整合了现有特征提取架构与量子物理学的相关概念理论，其主要贡献可归纳如下：

1）复数锚点动态生成。不同于传统实数锚点或固定复数参数，本文方法通过轻量级网络生成包含实部（位置）与虚部（相位）的动态复数锚点，首次将量子态参数化与锚点自适应结合。并通过锚点与像素特征的双向交互来完成线性复杂度下的局部-全局特征融合提取；

2）相位-振幅联合调制。所提方法通过相位旋转和振幅衰减模块来模拟量子态的旋转与测量过程，实现特征在复数域的联合变换，弥补现有方法分离处理的不足；

3）规范连接驱动的特征聚合。所提方法引入规范场论中的相位差计算，通过规范连接增强像素与锚点的交互一致性，解决传统方法对几何变换鲁棒性不足这一问题。

总体而言，本文在量子启发的特征处理框架下，实现了动态锚点生成、相位旋转-振幅衰减调制、规范对称性保持的深度融合，为轻量级高效图像特征提取提供了新的解决方案。

1 相关知识

1.1 锚点注意力机制

在锚点注意力机制方面，Lee等人[10]提出的ISAB方法使用固定的诱导点，作为可学习的参数独立于输入数据，而没有依赖于输入特征去动态生成锚点坐标，尚未实现数据驱动的锚点生成。同时，也未涉及复数表示，仅限于提取全局特征。为此，2020年，Hudson等人[11]提出了GA-Set Transformer。该改进版Transformer在GA-Set Transformer基础上加入生成对抗的采样机制，使诱导点能动态选择重要输入子集，但是更侧重样本选择而非空间特征变换。2021年，Jaegle等人[12]提出了PerceiverIO网络，其核心思想是使用一组固定大小的潜节点与输入做交叉注意力，将任意高维输入投射到潜空间，再经多次自注意力，最后交叉到所需输出。同年，Zeng等人[13]提出了一种近似自我注意力算法，用少量锚点近似全局自注意力的键-值矩阵，实现了线性复杂度。但由于二者都没有空间坐标意识，因此从一维自然语言处理领域引入到图像领域后，对二维空间的灵活感知性能仍需提升。

1.2 与量子物理启发的神经网络

在量子物理启发的神经网络方面，主要围绕复数特征与相位变换、振幅与相位场建模、规范对称性与特征聚合三个方面展开研究。目前已有的相关工作可主要归纳为：

1）复数特征与相位变换。早期Trabelsi等人[14]提出引入复数卷积、复数批归一化与复数激

活的思想，实现端到端复数网络，以提升语音与信号处理性能。但其复数乘法成本约为4倍的实数乘法会带来过大的计算开销。Ng等人[15]将复数卷积核与激活进一步加上量子计算理论中的相位衰减来模拟量子耗散，同样在原CNN网络的基础上计算量翻倍，对硬件支持要求高。Zhang等人[16]提出了一种基于Vision Transformer的混合量子视觉转换器。该架构的复数注意力头中用相位矩阵乘以实/虚部特征，直接在Transformer中植入量子叠加/干涉机理，以提升注意力表达。然而其训练不稳定，需要精心设计初始化与正则。

2）振幅与相位场建模。从频域视角出发，通过在特征融合中显式重组振幅与相位分量，Chen等人[17]所提方法提高了医学重建/分割任务的质量，适合需要特征全局稳定性的任务。但该方法对空间局部细节与空间定位建模较弱，针对分类任务中需要精细定位的细节无法有效识别。2022年提出的SiSPRNet方法[18]通过端到端网络同时估计振幅与相位，实现了高精度、无需迭代的相位恢复，但此方法对自然图像领域的泛化性有待验证。Tang等人[19]认为纯多层感知器（Multilayer Perceptron, MLP）架构可以实现与CNN和transformer竞争的性能，为此他们提出了一种将图像块视为波，显式分离振幅/相位并在MLP中重组的方法。其架构能动态聚合token，具备网络轻量且保留局部细节与全局频域信息的优势，在多项基准上超越了Transformer。但不足之处在于该方法不仅需额外的快速傅里叶变换运算开销，还对低频偏置较为依赖。

3）规范对称性与特征聚合。2018至2025年期间，Cohen等人[20]一直致力于研究从流形上gauge场的纤维丛出发，设计等变卷积算子，现已形成了较为完善的理论框架，并且可以推广到任意Lie群。但其从量子态理论出发的网路结构，实现起来复杂度过高。Bulusu等人[21]将标准化流与规范对称性结合，通过规范等变变换（gauge-equivariant变换）保证采样分布不破坏规范不变量。该方法能高效生成满足规范不变量的样本，理论上保证严格满规范对称性。但主要该方法针对格点量子色动力学，不易直接移植到大规模计算机视觉任务。同样的，文献[22-24]也聚焦于晶格量子色动力学的研究，不直接面向图像特征提取，要应用到图像研究领域需要大量的桥接工作。目前为止，在经典针对自然图像的计算机视觉任务中，直接以“规范连接”算子形式出现、并结合锚点机制的相关工作仍非常稀少。

2　本文方法

2.1　方法框架

本文提出的动态双向锚点交互机制，受到Kerenidis等人[25]探讨量子线路的启发，通过动态生成“复数锚点”并在像素-锚点特征双向交互中引入相位和振幅操作，实质上模拟了量子态的概率振幅与相位演化过程。核心流程包括：复数锚点生成、振幅衰减（量子测量）、相位旋转（量子门作用）、规范连接（量子规范场）、局部/全局特征聚合与残差融合。每一步均可在量子物理框架下找到对应对象。例如，振幅衰减对应测量塌缩，相位旋转对应单量子比特相位门，规范连接对应规范场连接子。其好处在于：通过相位编码引入更丰富的全局/局部分布信息，且天然具备正交性约束，减少锚点间冗余。利用相干叠加与干涉机制提升特征表达能力，引入局部–全局协同的“量子纠缠”式信息流动，并借助规范不变性获得更强的几何/拓扑鲁棒性。此外，所提方法还采用全局上下文信息增强（Global Context Information Enhancement Module，GCIEM）模块[26]与自交互（Self-Interaction Module，SIM）模块[27]来提升模型的性能。这两个模型都具有极强的泛化能力，可以做到即插即用，适用于多种任务。量子启发式锚点特征交互机制整体架构如图1所示。

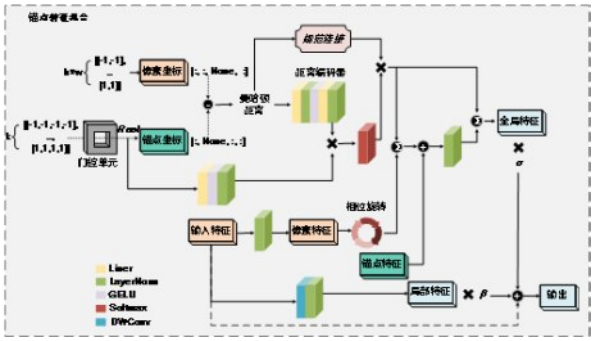

图1　所提量子启发式锚点特征交互机制整体结构图

2.2　量子启发式锚点特征交互机制

（1）　动态锚点生成器

锚点生成器（Dynamic Anchor Generator，DAG）作为架构的基础模块之一，其设计灵感源

自量子物理中对微观粒子状态的动态描述。这一概念启发我们动态生成锚点坐标，而非采用传统的固定可学习参数方式。由于传统检测网络中的锚点难以随图像内容自适应，受量子态可变振幅与相位的启发，本模块动态生成复数坐标锚点，将每个锚点视为一个二维量子“位置”振幅与“相位”振幅的组合。该模块首先通过池化层对输入特征图进行压缩，将全局信息凝聚为一个低维向量。这一过程类似于量子物理中的信息坍缩，将高维特征空间中的信息压缩到一个更易于处理的维度。设输入特征图

$$\hat{A}\left(|\psi\rangle\right)=tanh\left(\mathrm{W2LN}\left(\mathrm{GELU}\left(\mathrm{W1}|\psi\rangle\right)\right)\right)\in[\,-1,1]^{4K}$$

其中，W1 与 W2 表示两层映射层，LN 表示层归一化操作。具体而言，首先通过第一层映射层将全局特征映射到相同维度空间，接着利用高斯误差线性单元引入非线性，再经过层归一化操作进行归一化，最后将特征映射为 $4K$ 维向量。其中每 4 个值对应一个锚点的复数坐标，即横纵坐标的实部和虚部。重塑后得每个锚点的复数坐标 $\left\{\left(x_k^{\mathrm{R}},x_k^{\Im},y_k^{\mathrm{R}},y_k^{\Im}\right)\right\}_{k=1}^{K}$，$K$ 表示锚点的数量。上述生成过程可抽象为：

$$z=\mathcal{F}(g) \tag{3}$$

在式（3）中，$\mathcal{F}$ 表示映射函数。将坐标值限制在 [-1, 1]范围内，与图像像素坐标的归一化范围保持一致，这类似于量子物理中对物理量取值范围的限制。值得注意的是，在动态更新锚点位置以外，实验也尝试过设计自适应机制，如基于熵的锚点数量预测，来动态设置锚点数量，避免固定 K 的限制。然而，实验发现动态改变锚点数量极易导致每次前向传播时，锚点相关的张量维度不匹配的问题。解决方案是将每个样本的锚点数量填充到统一的最大值 K_{max}，并用掩码标记无效锚点。但多次实验结果证明，该方案对模型性能增益不明显。

2） *相位旋转*

相位旋转（Phase Rotation，PR）模块的设计借鉴了量子力学中量子态的相位特性。在量子系统中，量子态的相位对其叠加和干涉现象起着关键作用，这种相位变化能够改变量子态之间的相互作用。在量子力学中，相位门 $R(\theta)=e^{i\theta}$ 引入全局或局部相位，影响干涉结果而不改变概率振幅模长。

本模块首先基于输入特征的全局均值生成相位参数。对于输入特征张量 $X\in\mathbb{R}^{B\times N\times C}$（N=H×W），通过计算其在空间维度（$n$ 维度）上的均值得到全局特征 $g_x\in\mathbb{R}^{B\times C}$。再如图 2 所示，通过相位参数生成模块对输入的实数特征变换为复数域特征，相位参数生成模块同样包含多层线性变换和激活函数，其输出的 θ 被拆分为 $cos\theta$ 和 $sin\theta$ 两部分，分别对应相位旋转中的余弦值和正弦值。

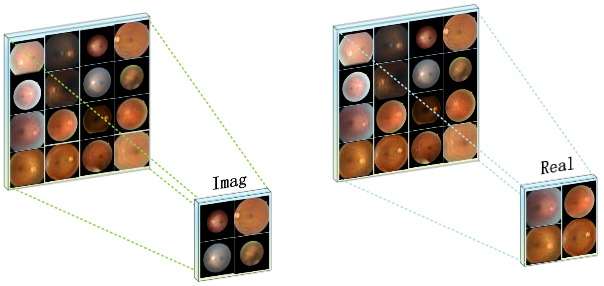

图2　将输入相位旋转模块的实数特征变换为复数域特征

根据量子力学中复数形式的量子态变换公式，基于 θ 对每个特征向量施加相位旋转，形式如图 3 所示，得到旋转后的实部 x^{R} 和虚部 $x^{\Im}$：

$$\left(x^{\mathrm{R}}+ix^{\Im}\right)\mapsto\left(x^{\mathrm{R}}+ix^{\Im}\right)e^{i\theta}\#\tag{4}$$

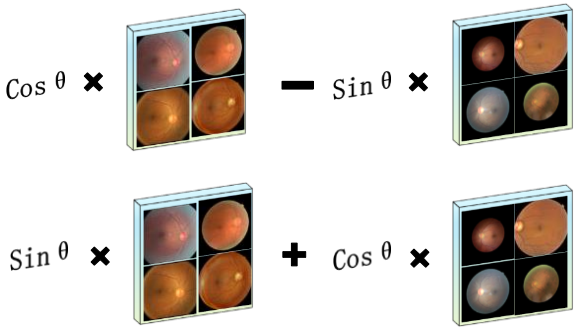

图3　对复数域特征进行复数乘法来实现相位旋转

具体来说：

$$\begin{cases} x^{\mathrm{R}} = x^{\mathrm{R}}cos\theta - x^{\Im}sin\theta, \\ x^{\Im} = x^{\mathrm{R}}sin\theta + x^{\Im}cos\theta. \end{cases} \tag{5}$$

最后，将旋转后的实部和虚部合并，得到经过相位旋转的输出特征，这一过程模拟了量子态在相位变化下的演化，增强了特征的表达能力。

(3) 振幅衰减

振幅衰减（Amplitude Attenuation，AA）模块的设计理念来源于量子物理中的测量过程。在量子系统中，对量子态进行测量会导致量子态发生坍缩，其概率遵循一定的衰减规律。本模块通过对锚点坐标的处理，模拟了这一量子测量中的振幅衰减现象，类似量子振幅阻尼信道，将复数振幅按坐标信息生成门控因子$\gamma \in (0,1)$，实现对锚点坐标振幅的自适应调制：

$$\left(x_k^{\mathrm{R}}, x_k^{\Im}, y_k^{\mathrm{R}}, y_k^{\Im}\right) \mapsto \gamma \odot \left(x_k^{\mathrm{R}}, \cdots, y_k^{\Im}\right) \tag{6}$$

模块中的门控是由映射层和Sigmoid函数组成的。在输入复数锚点坐标后，门控会根据坐标信息来生成衰减因子。映射层再对坐标进行线性变换， Sigmoid函数则起到将输出值限制在[0, 1]范围内的作用，使其符合概率衰减的取值要求。相比于直接进行简单的均值振幅衰减，可学习的门控网络层可以在不明显增加复杂度的情况下比原先设计的量子线路在准确率上有0.44%的提升。

(4) 局部相位场

局部相位场（Local Phase Field，LPF）模块用于捕捉图像局部空间的相位信息，设计思想与量子力学中对空间局部量子态的描述相关。在量子场论中，空间不同位置的量子场具有不同的相位，这些相位差异反映了物理量在空间中的分布特性。该模块通过轻量卷积层对输入特征图进行卷积操作，将通道维度压缩为1，再通过Tanh激活函数得到局部相位偏移ϕ。对于输入特征图$X \in \mathbb{R}^{B \times C \times H \times W}$经过卷积操作后得到$\phi \in \mathbb{R}^{B \times H \times W}$，其计算过程可表示为：

$$\phi_{b,h,\omega} = \mathcal{T}\left(\sum_{c=1}^{C} \omega_c x_{b,c,h,\omega} + b\right) \tag{7}$$

其中，$\mathcal{T}$表示Tanh函数，ω_c和b分别为卷积核的权重和偏置。局部相位偏移ϕ反映了图像局部区域的相位变化情况。通过将其与像素坐标进行融合，能够为后续的特征处理引入空间局部的相位信息，增强模型对空间结构的感知能力，类似于量子场在空间中传递信息的过程。

(5) 规范连接

规范连接（Gauge Connection，GC）模块的设计基于量子规范理论，该理论描述了物理系统在规范变换下的不变性。在网络中，通过计算像素与锚点之间的相对坐标，生成规范连接，以模拟量子系统中规范场的作用。量子场论中，规范连接$U = \exp(i\int A_\mu dx^\mu)$描述不同点间相位并行移位。本模块将锚点与像素的相对坐标映射到相位增量$\Delta\theta$，构造复数连接权重。

具体来说，对于输入的相对坐标$r \in \mathbb{R}^{B \times N \times K \times 2}$（N=H×W为像素数量，K为锚点数量），首先将其展平为$r_f \in \mathbb{R}^{B \times N \times K \times 2}$，然后通过映射函数生成相位增量$\Delta\theta$。设其映射函数为T，则：

$$\Delta\theta = \mathrm{T}\left(r_f\right) \tag{8}$$

最后，根据复数指数形式生成规范连接U，其表达式如（9）所示。规范连接的引入使得网络在特征交互过程中具有类似于量子规范理论中的不变性，增强了模型对不同空间位置特征关系的建模能力，确保了特征传递过程中的稳定性和一致性。

$$U_{n,k} = e^{i\Delta\theta\left(r_{n,k}\right)}, r_{n,k} = \left(x_n - x_k, y_n - y_k\right) \tag{9}$$

(6) 锚点特征交互机制

锚点特征交互（Anchor Feature Interaction Module，AFIM）模块是整个架构的核心，它是一个多模块协同的量子启发式特征交互模块。将上述各个模块有机结合，实现了基于锚点的特征混合与交互。其设计全面融入了量子物理的思想，构建了一个高效的特征处理系统。

模块首先为输入特征图的每个像素生成归一化的二维坐标，像素与锚点坐标的统一归一化使得后续操作，如相对位置计算，可以独立于输入特征图的具体尺寸（h, w），从而支持可变分辨率的输入。与此同时，利用可学习轻量网络动态生成复数锚点坐标，替代固定预设或手工设计的锚点，每个锚点k的坐标由实部$(x_k^{\mathrm{R}}, y_k^{\mathrm{R}})$与虚部$(x_k^{\Im}, y_k^{\Im})$组成，整体视为量子态振幅与相位的结合：

$$z_k = x_k^{\mathrm{R}} + ix_k^{\Im}, \ \omega_k = y_k^{\mathrm{R}} + iy_k^{\Im} \tag{10}$$

虚部携带“相位”信息，用于后续相位调制。实部用于空间定位。生成过程是输入特征$X \in \mathbb{R}^{B \times H \times W \times C}$经全局平均池化得到状态向量，通过两层全连接与GELU、LayerNorm，再经Tanh限

制于[-1,1]，生成$4K$维输出，重塑后得复数锚点坐标张量。然后模拟量子测量导致的概率坍缩，使锚点位置的生成依赖于输入特征的局部统计特性，避免固定锚点的先验偏差，将坐标振幅按可学习门控$\gamma_k \in (0,1)^4$自适应衰减：

$$\tilde{Z}_b[k] = \gamma_k \odot Z_b[k] \quad (11)$$

$$\gamma_k = \sigma\left(W_a Z_b[k]\right) \quad (12)$$

上式中，σ为 Sigmoid 函数，$\odot$为逐元素乘。此步骤可以提升坐标鲁棒性并抑制噪声影响。取实部作为最终的动态锚点位置，同时分离出虚部用于后续相位投影。然后利用相位场模块提取空间依赖的相位偏移$\phi \in \mathbb{R}^{B\times H\times W}$。通过 1×1 卷积和 Tanh 函数激活，将通道维度压缩为1，生成逐像素的相位偏移，类比量子场论中空间各点的相位分布。在此基础上，受如图4所示的RMT模型[9]启发，通过计算锚点与像素之间的曼哈顿距离，为模型提供更细粒度的几何关系。此外，我们在实践中发现：保留坐标差的正负性，并不会因为加强了特征相对位置的精准表示，而达到提升模型性能的效果。相反，实验证明计算相对坐标后进一步取绝对值是更好的选择。

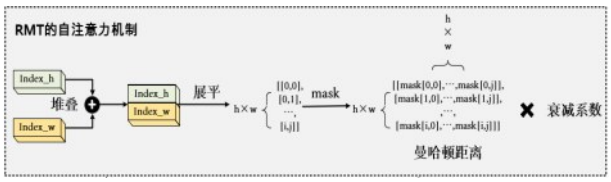

图4　RMT网络[9]提出的创新自注意力机制

所提方法还引入了规范连接模拟量子规范场理论中的局部对称性，以确保特征交互在空间变换下的不变性，提升模型对平移、旋转等几何变换的鲁棒性。在聚合像素特征到锚点特征的更新过程中，首先通过距离编码器得到距离嵌入。这里距离嵌入的计算，类似于量子物理中粒子间相互作用的强度与距离相关。然后插入相位信息，将锚点的虚部相位通过相位投影层网络投影到特征维度，得到锚点相位信息，并将其与距离嵌入相乘，使得距离嵌入中融入了相位信息，体现了量子物理中波函数相位对粒子状态的影响。对距离嵌入取绝对值并进行裁剪，防止数值溢出。将像素级特征进行层归一化后再对其应用相位旋转，模拟量子态的相位变化。

在通道维度上对距离嵌入应用 softmax 函数，得到在锚点维度K上归一化的权重张量，这类似于量子物理中的概率分布，为后续的加权聚合提供权重。接着将其与规范连接的绝对值相乘进行不变聚合。然后沿通道维度加权聚合像素特征，得到每个像素对锚点的贡献，并对其进行平均处理。将所有像素对锚点的贡献求和并扩展维度后，与可学习的锚点特征相加，再经过层归一化得到锚点表示。

在此基础上进行锚点特征到像素特征的广播。通过爱因斯坦求和操作，将锚点表示与加权后的距离嵌入相乘，得到全局特征。由此完成了像素特征聚合到锚点，再由锚点特征广播到像素特征的双向交互。其过程可以抽象为像素特征ψ_n按照$\omega_{n,k}$加权平均后得到锚点特征：

$$\delta_{a_{b,k}} = \frac{1}{N}\sum_n \omega_{n,k}\psi_{b,n} \quad (13)$$

$$\tilde{a}_{b,k} = a_k + \delta_{a_{b,k}} \quad (14)$$

利用权重$\omega_{n,k}$再将锚点特征$\tilde{a}_{b,k}$广播回像素：

$$\hat{\psi}_{b,n} = \omega_{n,k}\tilde{a}_{b,k} \quad (15)$$

同时，借鉴RMT[9]网络的思想对输入特征进行局部轻量级深度可分离卷积和层归一化提取局部细节$L \in \mathbb{R}^{B\times H\times W\times C}$。受到Swin Transformer[6]模型的启发，设计两个可学习的层缩放参数λ_g、λ_l分别对全局特征和局部特征进行缩放。再将输入特征、缩放后的全局特征和局部特征进行残差相加，最后附加上残差后归一化得到输出特征。上述过程可表示如下：

$$Y = X + \lambda_g\hat{\psi} + \lambda_l L \quad (16)$$

大量实验结果表明：层缩放技术通过在训练初期将这些缩放参数初始化为零，让网络优先学习局部特征。之后随着训练的进行，再逐渐学习全局特征。最终利用残差后归一化层使特征分布更加稳定，从而避免了在反向传播过程中极其容易出现的梯度爆炸情况。

上述整个流程中多处受到量子理论的启发。例如，相对坐标的计算类似于量子物理中粒子间的相对位置关系。规范连接的引入确保了特征交互在空间变换下的不变性，与量子规范场理论中的规范变换相对应。相位信息的插入和相位旋转操作模拟了量子态的相位特性。softmax 函数得到权重类似于量子物理中的概率分布。

3　实验结果与分析

3.1　数据集及评价指标

所提方法在青光眼筛查领域的公开数据集AIROGS[28]上进行主要训练与评估，并与专门针对青光眼筛查的人工智能挑战赛上，参赛优胜队伍所达到的各项指标成绩进行横向比较。AIROGS数据集包含来自约60000名患者和500个不同筛查中心的约101442张图像。所有图像均由人类专家分配，标签为referable glaucoma与no referable glaucoma。实验中，我们按照9:1的比例来分配训练集与测试集，对所有图像进行统一的裁剪、填充、缩放处理，尺寸大小默认统一为224×224，处理后的效果如图5所示。

在网络参数配置方面，实验中我们设置epoch为300，batchsize设置为64，optimizer默认为adamw，学习率默认为0.0005，锚点数量设置为8。以上超参数可根据实验硬件配置微调。训练所用设备为NVIDIA RTX 4090D(24GB) GPU，采用了单机多卡的分布式训练方式。

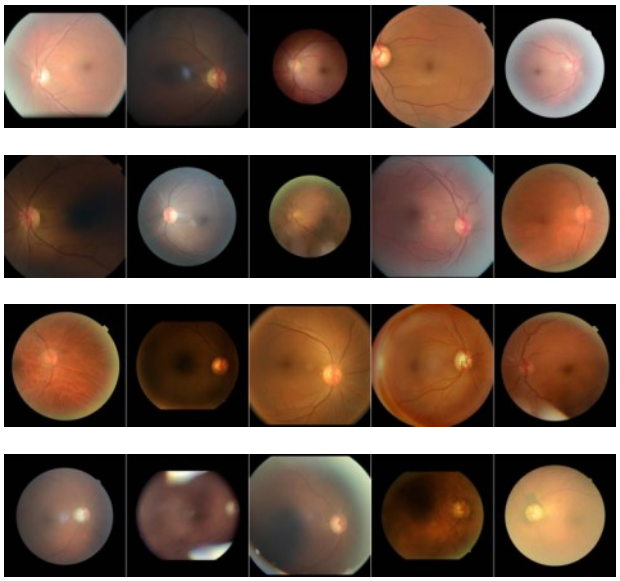

图5　AIROGS数据集图像数据处理后效果，前三行图像为正常样本，最后一行为患者样本

AIROGS挑战赛官方评测指标为pAUC@90-100%特异性、95%特异性下灵敏度（SE@95SP），其官方测试集不对外公开释放，仅可线上Docker提交评测。本文基于公开可获取的AIROGS开发数据集，采用9:1分层随机划分得到本地训练子集与本地测试子集，采用指标：青光眼检测灵敏度、特异性检测、受试者工作特征曲线下面积、准确率。需要说明，本地划分数据集得到的结果，与挑战赛封闭测试集线上提交结果不能直接等价对比。其中，灵敏度表示实际为青光眼患者中，被正确检测为青光眼的比例，公式为：

$$sensitivity = \frac{TP}{TP + \mathrm{FN}} \tag{17}$$

上式中，TP（真阳性）为实际是青光眼且被正确诊断为青光眼的例数FN（假阴性）为实际是青光眼但被误诊为非青光眼的例数。

特异性表示实际非青光眼患者中，被正确检测为非青光眼的比例，公式为：

$$specificity = \frac{TN}{TN + \mathrm{FP}} \tag{18}$$

其中，TN（真阴性）为实际非青光眼且被正确诊断为非青光眼的例数，FP（假阳性）为实际非青光眼但被误诊为青光眼的例数。

曲线下面积（AUC）通常基于受试者工作特征曲线（ROC曲线，以灵敏度为纵轴、（1-特异性）为横轴）计算。常用梯形法近似积分，公式为：

$$AUC = \sum_{i=1}^{n-1} \frac{\left(FP_{i+1} - \mathrm{FP}_i\right)\cdot\left(TP_{i+1} + \mathrm{TP}_i\right)}{2} \tag{19}$$

在式（19）中，FP_i和TP_i为ROC曲线上第i个点的假阳性率和真阳性率。AUC值越接近1，诊断准确性越高。此外，还采用准确率指标来全面评价各模型的综合性能。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

3.2　实验结果

（1）　消融实验

为了验证所提方法各模块的有效性，我们在AIROGS数据集上做了一组消融实验来比较各模块的作用大小。表1列出了消融实验过程中各模块组合训练后的各项指标。实验中以不附加动态锚点生成器（DAG）、相位旋转（PR）、振幅衰减（AA）、局部相位场（LPF）、规范连接（GC）模块的锚点特征交互机制作为Base模块，表1即给出了针对各关键模块的消融实验结果。

实验设置遵循控制变量原则，所有组合均保持训练轮次、学习率、batch size等超参数一致，仅通过增减目标模块验证性能贡献，排除了训练配置差异对实验结果的干扰，保证了消融结果的可靠

性。从该表可以看出，动态锚点生成器（DAG）可根据眼底图像特征自适应调整锚点分布，解决了固定锚点对微小视盘病变适配性差的问题。规范连接（GC）模块能有效聚合多尺度上下文特征，增强模型对不同程度青光眼病灶的感知能力。自交互（SIM）模块可挖掘特征内部依赖关系，强化关键病理特征的响应。这三个模块对模型的性能提升最为关键。同时，相位旋转（PR）、振幅衰减（AA）、局部相位场（LPF）、全局上下文信息增强（GCIEM）等其它模块的加入也不同程度地提升了模型的性能。经统计，所提方法在青光眼检测灵敏度（Sen）、特异性检测（Spe）、受试者工作特征曲线下面积（AUC）这三个指标下的统计结果分别为0.8594、95.86%与0.9944，整体准确率（Acc）为95.45%。

（2） 模型鲁棒性测试

表3中展示了临床常见图像退化场景下的鲁棒性测试结果。具体做法是在AIROGS测试集上对输入图像分别施加高斯模糊、亮度偏移、高斯噪声和椒盐噪声等扰动，训练好的模型不再进行微调，直接测试各项指标变化。结果显示，在常见退化条件下模型性能仅出现小幅下降，说明动态锚点生成、振幅衰减门控、规范连接以及GCIEM/SIM带来的全局上下文与多尺度建模能力，能够在一定程度上增强模型对图像质量波动的适应性。

（3） 在其它测试集上的性能分析

考虑到所提方法在单一AIROGS数据集上的表现可能不具备泛化性，因此实验中又引入GAMMA[29]和REFUGE[30]两个青光眼数据集。其中GAMMA数据集中的300个样本对应于276名中国患者（42%为女性）。青光眼占样本的50%，其中早期占52%，中期占28.67%，晚期占19.33%。眼底图像是使用分辨率为2000×2992像素的KOWA相机和分辨率为1934×1956像素的Topcon TRC-NW400相机采集的。REFUGE挑战赛发布的数据集包含1200张眼底图像的数据集，其中包含视杯视盘的真实分割和临床青光眼标签。

图6中展示了ResNet-50[5]、ResNet-101[5]、SwinTransformerV2-T[6]、PoolFormer-S24[7]、ConvNeXtsV2-Y[8]、RMT-S[9]、本文方法这7种不同的青光眼筛查算法在AIROGS训练集上训练后，在AIROGS、GAMMA、REFUGE这三个测试集上的AUC值分布。在REFUGE、GAMMA、AIROGS三个测试集上，本文模型AUC指标在对比方法中数值相对较优；受限于数据集样本量，未做严格统计显著性检验，该结果仅作为数值参考。由该图可知，各方法AUC值的分布在AIROGS测试集上分布最密集，在另外两个数据集上分布比较分散，这也符合各算法均在AIROGS数据集上训练的事实。

（4） 与已有模型的性能对比

1） 与其他代表性模型的分类精度对比

表4统计了本文方法、AIROGS比赛的顶尖参赛方案，以及近年在大规模青光眼公开数据集上性能表现优异的多个网络。实验中我们选取了参数量大小相近的模型版本，对比了它们在AIROGS数据集上的性能表现。

表4中，PUMCH-eye团队在AIROGS挑战赛线上封闭测试集灵敏度、特异性分别为0.8533、0.9500，0.9898为该团队图像不可分级判别任务的AUCU指标，并非青光眼筛查ROC-AUC指标。该线上结果基于挑战赛私有封闭测试集，与本文本地划分数据集的实验结果仅做参考性对比，二者数

表1　　**所提方法关键模块的消融实验**

Module	AIROGS Datasets			
	Sen		Spe	AUC	Acc
Base	0.8223	0.9374	0.9753	93.37%
Base+DAG	0.8319	0.9456	0.9835	93.81%
Base+DAG+PR	0.8437	0.9502	0.9858	94.12%
Base+DAG+PR+AA	0.8464	0.9519	0.9872	94.26%
Base+DAG+PR+AA+LPF+GC	0.8515	0.9523	0.9893	94.60%
Base+DAG+PR+AA+LPF+GC+GCIEM	0.8521	0.9529	0.9917	94.79%
Base+DAG+PR+AA+LPF+GC+GCIEM+SIM	**0.8594**	**0.9586**	**0.9944**	**95.45%**

表2　不同锚点数量K的消融实验结果

K	AIROGS Datasets			
	Sen	Spe	AUC	Acc
2	0.8458	0.9501	0.9881	94.21%
4	0.8522	0.9540	0.9912	94.78%
6	0.8567	0.9569	0.9931	95.18%
8	0.8594	0.9586	0.9944	95.45%
10	0.8576	0.9571	0.9936	95.29%
12	0.8569	0.9564	0.9932	95.21%

表3　模型鲁棒性测试结果

退化类型	AIROGS Datasets			
	Sen	Spe	AUC	Acc
无退化	0.8594	0.9586	0.9944	95.45%
高斯模糊（σ=1）	0.8526	0.9541	0.9915	94.88%
高斯模糊（σ=2）	0.8448	0.9489	0.9872	94.13%
亮度（+20%）	0.8558	0.9562	0.9928	95.12%
亮度（−20%）	0.8519	0.9537	0.9911	94.76%
高斯噪声（σ=0.03）	0.8507	0.9532	0.9904	94.69%
椒盐噪声（p=0.02）	0.8485	0.9510	0.9891	94.43%

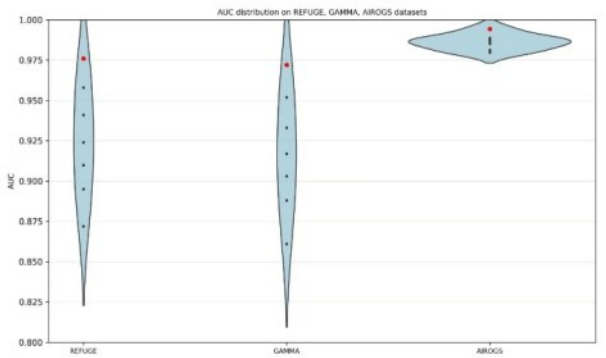

图6　在REFUGE、GAMMA、AIROGS数据集上的比较

据集分布存在差异，不可直接横向等同。本文基础模型Ours，即不包含GCIEM与SIM模块的量子启发式锚点特征交互机制，在特异性上达到0.9523，略高于 PUMCH-eye team，同时其灵敏度和AUC分别达到0.8515和0.9893，已接近竞赛第一名水平，说明所提QAFIM机制本身具备较强的青光眼筛查能力。在此基础上，加入全局上下文信息增强模块（GCIEM）和自交互模块（SIM）后的完整方法进一步将灵敏度、特异性和AUC提升至0.8594、0.9586 和 0.9944，分别比 PUMCH-eye team 高出0.61、0.86 和 0.46 个百分点，其中灵敏度达到85.94%，与人类评分员86%的水平十分接近。

此外，表4还显示，本文方法相较于ResNet、Swin Transformer、PoolFormer、ConvNeXt V2 和RMT 等代表性通用视觉骨干具有更优或更稳定的综合表现。值得注意的是，当将所提方法分别与CNN和Transformer 领域的代表性高性能网络ConvNeXt V2 与 RMT 结合时，ConvNeXt V2-T+Ours的灵敏度和 AUC 分别提升至0.8521 和 0.9901，RMT-S+Ours 的灵敏度、特异性和AUC分别提升至0.8529、0.9533和0.9909，均优于对应原始骨干网络。这说明所提量子启发式锚点特征交互机制不仅可以作为独立特征提取结构使用，也能够与CNN和Transformer类架构形成良好的互补关系。此外，除SegImgNet的灵敏度值最佳外，Ours模型在特异性和AUC这两个指标均取得了相对较好的统计结果。再结合GCIEM与SIM后，所提Ours+GCIEM+SIM方法在全局上下文建模、多尺度特征融合和关键区域定位方面得到进一步增强，从而在三个指标上均取得了表中相对较优的综合分类性能。

2） 与其它代表性模型的参数量对比

表5给出了所提方法与其他代表性方法参数量的比对，可以看出所提方法在参数量方面显著低于ResNet-101 和 ConvNeXtsV2-T 的 4000 万参数量。与RMT-S、ResNet-50和SwinTransformer-T的2500万左右参数量相比，具有较大优势。但与轻量级模型 PoolFormer-S24 的 2100 万左右参数量相比略高。

在计算效率方面，从图7可以看出：本文方法的总计算量为3.6G FLOPs，计算效率显著优于绝大多数对比方法，仅略高于 PoolFormer-S24 的3.4G FLOPs，二者差距仅为0.2G FLOPs，差异相对较小。该结果验证了本文方法在架构设计中注重计算效率优化，能够通过较少的计算量实现优异的检测性能。因此，在临床青光眼筛查的实际落地场景中，具备良好的部署价值与应用潜力。本实验未开展统计显著性检验，不能从统计学意义上证明该方法显著优于其他对比算法，仅展示了多次运行观测得到的性能数值对比。

表4 不同方法在 AIROGS 数据集上青光眼筛查性能对比

Model or Team	AIROGS Datasets		
	Sen	Spe	AUC
RWTH-CuP team[3]	0.8396	0.95	0.9852
PUMCH-eye team[4]	0.8533	0.95	0.9898
ResNet-50[5]	0.8414	0.9502	0.9867
ResNet-101[5]	0.8359	0.9465	0.9813
SwinTransformerV2-T[6]	0.8322	0.9423	0.9798
PoolFormer-S24[7]	0.8486	0.9527	0.9887
ConvNeXtsV2-T[8]	0.8469	0.9506	0.9885
RMT-S[9]	0.8468	0.9520	0.9878
MedViT[31]	0.9254	0.9073	0.9703
Multi-GlaucNet[32]	0.9493	0.8441	0.9665
VisionDeep-AI[33]	0.9305	0.8997	0.9704
SegImgNet[34]	**0.9493**	0.9341	0.9852
Ours	**0.9393**	0.9545	0.9852
ConvNeXtsV2-T+Ours	0.8515	**0.9523**	**0.9893**
RMT-S+Ours	0.8521	0.9511	0.9901
Ours+GCIEM+SIM	**0.8594**	**0.9586**	**0.9944**

表中 ResNet-50~RMT-S 为本文复现结果。Multi-GlaucNet、VisionDeep-AI、SegImgNet、RWTH-CuP、PUMCH-eye 为原始文献报道结果，数据集划分、训练设置不完全一致，仅作参考对比。

表5　所提机制与其它代表性方法参数量比对

Model	Parameters(M)
ResNet-50[5]	25.56
ResNet-101[5]	44.55
SwinTransformerV2-T[6]	26.44
PoolFormer-S24[7]	21.50
ConvNeXtsV2-T[8]	40.24
RMT-S[9]	25.94
Ours	**22.23**

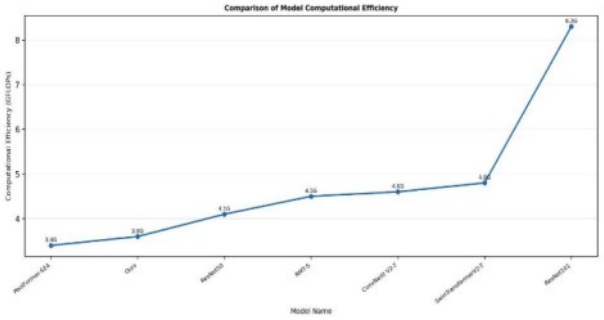

图7　本文所提机制与其他方法的计算效率对比折线图

4　结语

本文针对现有基于卷积神经网络CNN与视觉Transformer的青光眼筛查方法在全局建模能力、位置敏感性，以及计算复杂度等方面存在的局限性，提出了一种受量子理论启发的青光眼筛查方法。该方法构建了量子启发式动态锚点双向特征交互机制，通过引入复数域动态锚点生成、相位-振幅联合调制、规范连接驱动的特征聚合策略，有效模拟了量子态的演化过程，可在显著降低模型计算复杂度的同时，实现局部细节与全局语义的高效融合。此外，结合全局上下文信息增强模块与自交互模块，进一步提升了模型对细微病变区域的定位精度，以及对不同尺度样本的泛化适应能力。在AI-ROGS等公开数据集上的实验结果表明：所提方法在参数量与计算开销较低的前提下，可取得优于或相当于目前主流方法的筛查性能，从而验证了将量子物理思想引入医学影像分析任务的有效性与可行性。尽管本文方法在青光眼筛查任务中表现优异，但在极端难例下的鲁棒性方面仍有提升空间。未来的研究工作将致力于探索更复杂的量子场论模型框架，并将该模型框架扩展至其他眼科疾病乃至更广泛的医学影像智能诊断领域。

参考文献：

[1] Tham Y C, Li X, Wong T Y, 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

[2] Khan S, Naseer M, Hayat M, Zamir S W, Khan F S, Shah M. Transformers in Vision: A Survey[J]. ACM Computing Surveys. 2021, 54(10): 1-41

[3] Fu H, Cheng J, Xu Y, et al. Disc-Aware Ensemble Network for Glaucoma Screening from Fundus Image[J]. IEEE Trans Med Imaging, 2018, 37(11): 2493-2501

[4] Wang H, Sun H, Fang Y, et al. A workflow for computer-aided diagnosis of glaucoma[C]. Proceedings of the 2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC), 2022

[5] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016

[6] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the Proceedings of the IEEE/CVF International conference on computer vision, 2021

[7] Yu W, Luo M, Zhou P, et al. Metaformer is actually what you need for vision[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022

[8] Woo S, Debnath S, Hu R, et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders[C]. Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023

[9] Fan Q, Huang H, Chen M, et al. RMT: Retentive networks meet vision transformers[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024

[10] Lee J, Lee Y, Kim J, et al. Set transformer: A framework for attention-based permutation-invariant neural networks[C]. Proceedings of the International conference on machine learning, 2019

[11] Hudson D A, Zitnick L. Generative adversarial transformers[C]. Proceedings of the International conference on machine learning, 2021.

[12] Jaegle A, Borgeaud S, Alayrac JB, et al. Perceiver IO: A general architecture for structured inputs & outputs[C]. Proceedings of the International Conference on Learning Representations, 2021

[13] Xiong Y, Zeng Z, Chakraborty R, et al. Nyströmformer: A nyström-based algorithm for approximating self-attention[C]. Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2021

[14] Trabelsi C, Bilaniuk O, Zhang Y, et al. Deep complex networks. arXiv preprint arXiv:170509792, 2017

[15] Ng K, Song T. Hybrid Quantum-Classical Convolutional Neural Networks for Image Classification in Multiple Color Spaces. arXiv preprint arXiv:240602229, 2024

[16] Zhang H, Zhao Q, Zhou M, et al. HQViT: Hybrid quantum vision transformer for image classification. arXiv preprint arXiv:250402730, 2025

[17] Chen G, Peng P, Ma L, et al. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain [C]. Proceedings of the Proceedings of the IEEE/CVF international conference on computer vision, 2021.

[18] Ye Q, Wang L-W, Lun D P. SiSPRNet: end-to-end learning for single-

shot phase retrieval[J]. Opt Express, 2022, 30(18): 31937-31958.

[19] Tang Y, Han K, Guo J, et al. An image patch is a wave: Phase-aware vision mlp[C]. Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022

[20] Cohen T, Weiler M, Kicanaoglu B, et al. Gauge equivariant convolutional networks and the icosahedral CNN[C]. Proceedings of International conference on Machine learning, 2019.

[21] Bulusu S, Favoni M, Ipp A, et al. Generalization capabilities of translationally equivariant neural networks[J]. Physical Review D, 2021, 104 (7): 1-28

[22] Favoni M, Ipp A, Muller D I, et al. Lattice gauge equivariant convolutional neural networks[J]. Phys Rev Lett, 2022, 128(3): 1-6.

[23] Lehner C, Wettig T. Gauge-equivariant neural networks as preconditioners in lattice QCD[J]. Physical Review D, 2023, 108(3): 1-13.

[24] Knuttel D, Lehner C, Wettig T. Gauge-equivariant multigrid neural networks[C]. Proceedings of the 40th International Symposium on Lattice Field Theory, 2024

[25] Kerenidis I, Mathur N, Landman J, et al. Quantum vision transformers [J]. Quantum, 2024, 8: 1-20

[26] Guo F, Liu W T, Tang Jin. A saliency detection-inspired method for optic disc and cup segmentation[J]. Medical Image Analysis, 2026, 107: 1-16

[27] Pang Y W, Zhao X Q, Zhang L H, Lu H C. Multi-scale interactive network for salient object detection[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, pp. 9410-9419

[28] De Vente C, Vermeer K A, Jaccard N, et al. AIROGS: Artificial intelligence for robust glaucoma screening challenge[J]. IEEE Trans Med Imaging, 2023, 43(1): 542-557

[29] Wu J, Fang H, Li F, et al. Gamma challenge: glaucoma grading from multi-modality images[J]. Medical Image Analysis, 2023, 90: 1-12.

[30] Orlando J I, Fu H, Breda J B, et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs[J]. Medical Image Analysis, 2020, 59: 1-21

[31] Manzari O N, Ahmadabadi H, Kashiani H, et al. MedViT: a robust vision transformer for generalized medical image classification[J]. Computers in biology and medicine, 2023, 157: 106791.

[32] Xiong H, Long F, Alam M S, et al. Multi-GlaucNet: A multi-task model for optic disc segmentation, blood vessel segmentation and glaucoma detection[J]. Biomedical Signal Processing and Control, 2025, 99: 106850.

[33] Joshi R C, Sharma A K, Dutta M K. VisionDeep-AI: Deep learning-based retinal blood vessels segmentation and multi-class classification framework for eye diagnosis[J]. Biomedical Signal Processing and Control, 2024, 94: 106273.

[34] O Luo X, Zhao S, Zong Y, et al. SegImgNet: Segmentation-Guided Dual-Branch Network for Retinal Disease Diagnoses[C]//Proceedings of the AAAI Symposium Series. 2025, 5(1): 19-24.

郭璠（1982− ），女，湖南临澧人，博士，中南大学副教授，主要研究方向为图像处理、模式识别、人工智能等。

刘文韬（2000− ），男，中南大学硕士生，主要研究方向为计算机视觉、医学影像处理、多模态大模型等。

程欣（2003− ），女，中南大学在读硕士生，主要研究方向为图像处理、人工智能等。

孙蔓岑（1988− ），女，中南大学在读博士生，主要研究方向为图像增强、机器视觉、智能感知技术等。