基于脉冲神经网络的近眼红外瞳孔检测

朱厚帅[1] 汪 耀[1]
ZHU Houshuai WANG Yao

摘 要 眼动交互是人机交互领域的一个重要研究方向，而瞳孔检测是眼动交互的核心环节，其性能直接决定后续意图识别的上限。当前主流瞳孔检测算法普遍面临参数量大、运行功耗高的问题。针对这一问题，文章提出了轻量化瞳孔检测模型 Pupil-YOLO，以 YOLOv8 为基础优化模型骨干，然后在模型中引入脉冲神经元实现低功耗；为提升特征提取效率，构建了 P-Block1 轻量特征提取模块；为降低脉冲神经元的量化误差，设计了固定整数 A-LIF 神经元。实验结果表明，所提模型相较最新瞳孔检测方法 YOLOv8n-eye，在精度相差 0.8% 的情况下，参数量仅为其 1/18，为眼动交互技术提供了新的技术路径。

关键词 瞳孔检测；脉冲神经网络；人机交互；计算机视觉

doi：10.3969/j.issn.1672-9528.2026.01.041

0 引言

瞳孔检测（pupil detection, PD）通过对瞳孔位置与形态的精准定位，为眼动分析、凝视估计与视觉感知等任务提供基础支撑，是眼动交互系统的核心环节 [1-2]。其在医疗辅助 [3]、心理学实验 [4] 与可穿戴交互设备等场景中具有重要应用价值。

瞳孔检测方法主要可分为基于模型的方法和基于深度学习的方法。其中，基于模型的方法通常利用瞳孔与虹膜边缘等几何特征来定位瞳孔位置 [5]。这类方法由于引入了特定的先验假设，因而在只需少量训练数据的情况下即可实现有效检测。然而，这种简化假设也限制了其对复杂场景的适应能力，使其在面对外观变化较大、光照不均或反射干扰等情况下的检测效果不够理想。

另一方面，基于深度学习的瞳孔检测方法不再依赖人工设计的特征，而是通过端到端的方式直接从眼部图像中学习特征并预测瞳孔位置 [6]。此类方法能够在外观变化较大、光照复杂等情况下保持较好的检测精度。基于深度学习的方法尽管提升了模型对复杂场景的适应性，但其往往伴随着庞大的参数规模与较高的计算开销。

脉冲神经网络（spiking neural networks, SNNs），通过离散的脉冲事件进行信息编码与传播，具备天然的稀疏性与事件驱动特性，因而能够在推理过程中显著降低功耗 [7]。凭借其独特潜力，正逐步成为计算机视觉领域的前沿研究方向。一些研究已经尝试将 SNN 应用于目标检测任务，如 EMS-YOLO[8] 模型。虽然采用脉冲神经网络的模型可以大幅降低功耗，但是由于脉冲神经网络采用二值编码方式，会存在较大的量化误差。

为应对上述局限性，本文提出了基于脉冲神经网络的近眼红外瞳孔检测模型 Pupil-YOLO。该模型在保留 YOLO 架构的基础上进行轻量化设计，完成了自底向上的结构重构。在此基础上又设计了一种高效脉冲特征提取模块 P-Block1，以实现在较低参数的情况下对瞳孔区域有效信息的充分提取。此外，为减少脉冲神经元二值量化带来的误差，受 SFA 方法 [9] 和 Luo 等 [14] 启发，将神经元输出的脉冲发放模式调整为 0~4 之间的整数值，从而减小二值量化带来的量化误差。

本文所提出的方法，在 openEDS 数据集上得到了验证，并展现出了较高的检测精度和具有绝对优势的模型参数。此项工作的主要贡献有：首次将脉冲神经网络引入到近眼红外瞳孔检测领域；提出了高效的脉冲特征提取模块 P-Block1；优化传统的 YOLO 架构提出适用于脉冲神经网络的近眼红外瞳孔检测架构；在 openEDS 数据集上仅仅使用 0.15×10^6 参数，便实现了 99.5% 的 mAP-@50 和 95.6% 的 mAP@50:95 检测精度，相比于同架构的 ANN 网络 [10] 参数大幅降低。

1 方法

本节围绕所提出的 Pupil-YOLO 模型展开，其面向静态图像数据集。首先阐述网络的宏观架构设计，随后依次说明轻量级脉冲特征提取模块 P-Block1，最后，介绍固定整数脉冲神经元 A-LIF。

1.1 网络架构

脉冲神经网络的输入可表示为 $\boldsymbol{X}\in\mathbf{R}^{T\times C\times H\times W}$，其中 T 是时间步长；C 是通道；$H\times W$ 是空间分辨率。对于静态图像。为了利用脉冲神经网络的时空能力，通常的做法是将静态图

1. 武汉轻工大学电气与电子工程学院 湖北武汉 430023

像重复，并将其作为每个时间步长 T 的输入。这被称为直接输入编码，直接输入编码在小时间步长下能够实现更高的准确性[11]，尤其在深层架构和大规模数据集上优势更明显。图 1 给出了 Pupil-YOLO 的整体架构。在宏观架构上，本文借鉴了 YOLOv8 中“Backbone–Neck–Head”的经典检测结构布局，在保持整体拓扑结构的基础上，进行了全局重构与轻量化适配。如图 1 所示，整个 Pupil-YOLO 网络由 3 个主要部分组成，Backbone 负责提取图像的时空特征。为实现较高效率的特征提取，设计了轻量脉冲特征提取模块 P-Block1，并在每一层后接入下采样操作以逐步压缩空间维度增加感受野。Neck 部分以轻量化脉冲卷积替代原 YOLOv8 中的 PAF-PN 结构，实现跨层特征整合与上采样操作。

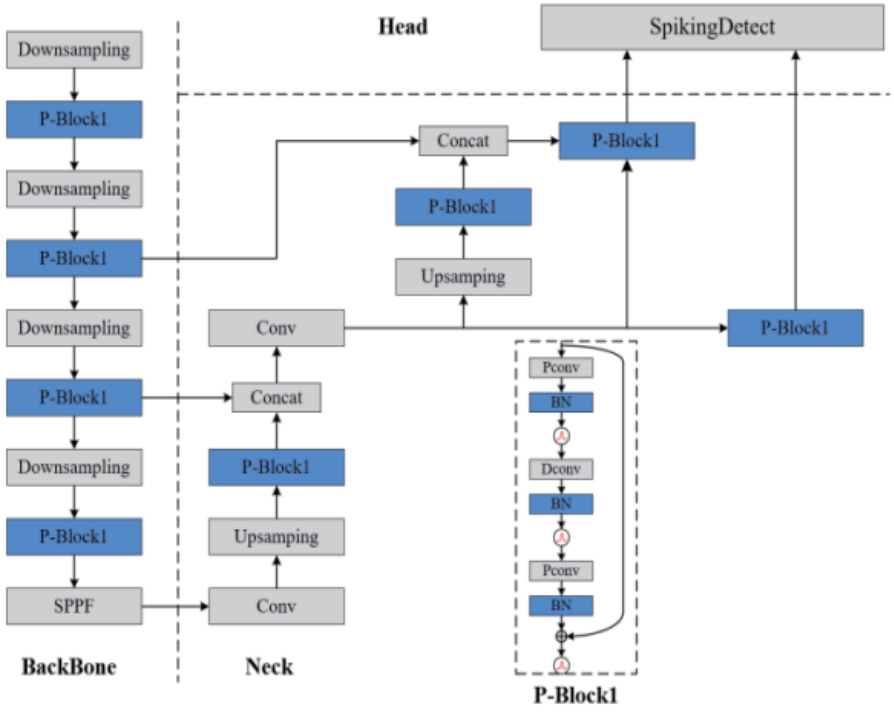

图 1 Pupil-YOLO 架构

1.2 P-Block1 特征提取模块

为实现在低参数下利用脉冲神经元充分提取瞳孔特征，本文借鉴 MobileNetV2 中的深度可分离卷积[12]设计，提出一种轻量化脉冲特征提取模块 P-Block1，其结构如图 2 所示。

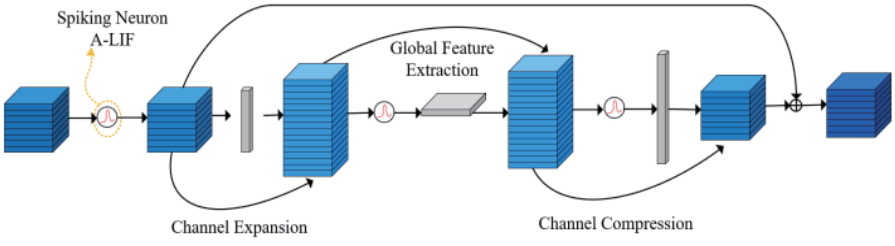

图 2 轻量化脉冲特征提取模块 P-Block1

深度可分离卷积采用“1×1 卷积升维，3×3 深度卷积，1×1 卷积降维”的架构，在多个移动端视觉任务中被证明能够显著减少参数量，同时保持良好的特征表达能力，因而广泛应用于轻量化网络设计。受此启发，本文将深度可分离卷积引入脉冲神经网络（SNN），构建 P-Block1 模块。在该模块中，传统激活函数被替换为脉冲神经元，从而构建出基于脉冲神经元的深度可分离卷积结构。具体而言，该结构首先通过 1×1 卷积扩充通道维度，接着采用 7×7 卷积提取全局特征，再通过 1×1 卷积提取空间特征并压缩通道，每一步之后均插入脉冲神经元作为激活单元。

由于脉冲神经元的稀疏性，在时序上的信息传递具有不连续性，单独依赖局部脉冲特征提取容易忽略低频脉冲中潜在的重要信息。为解决该问题，P-Block1 在完成深度可分离卷积后，引入了一个连接，将神经元激活后的输入特征与经过改进后的深度可分离卷积提取后的特征进行融合，其模块可标识为：

$$\boldsymbol{X}=\left\{x^{t}\right\}_{t=1}^{T}\in\mathbf{R}^{T\times C\times H\times W} \tag{1}$$

$$\boldsymbol{X}'=\boldsymbol{X}+\mathrm{DPConv}(\boldsymbol{X}) \tag{2}$$

式中：$\boldsymbol{X}$ 为脉冲激活后的输入特征；DPConv($\boldsymbol{X}$) 为改进后的深度可分离卷积。DPConv(·) 可表示为：

$$\boldsymbol{X}_1=PConv_1(S(\boldsymbol{X})) \tag{3}$$

$$\boldsymbol{X}_2=DConv(S(\boldsymbol{X}_1)) \tag{4}$$

$$\boldsymbol{X}_3=PConv_2(S(\boldsymbol{X}_2)) \tag{5}$$

式中：$\boldsymbol{X}$ 表示输入特征；$PConv_1$ 表示 1×1 卷积用于扩展通道数量提升通道表达能力；$PConv$ 表示 7×7 卷积，对扩展后的通道进行空间建模以提取全局特征；$PConv_2$ 表示 1×1 卷积用来压缩通道维度，恢复输出维度；S 表示脉冲神经元激活。

1.3 脉冲神经元 A-LIF

脉冲神经元是通过其膜电位的积分与发放机制实现的信息提取，将连续输入编码转换为稀疏的脉冲序列。具有软重置的泄漏积分与发放脉冲神经元 LIF[13]，因其在生物合理性和计算复杂性之间达到了平衡。成为构建脉冲神经网络中最常用的神经元，其动态特性为：

$$\boldsymbol{U}[t]=\boldsymbol{H}[t-1]+\boldsymbol{X}[t] \tag{6}$$

$$\boldsymbol{S}[t]=\Theta(\boldsymbol{U}[t]-V_{th}) \tag{7}$$

$$\boldsymbol{H}[t]=\beta(\boldsymbol{U}[t]-\boldsymbol{S}[t]) \tag{8}$$

式中：t 表示离散时间步；$\boldsymbol{U}[t]$ 表示膜电位，用于整合上一时刻的时间信息 $\boldsymbol{H}[t-1]$ 与当前时刻的空间输入 $\boldsymbol{X}[t]$；$\Theta(\cdot)$ 表示 Heaviside 阶跃函数：当 $x \geq 0$，取 1，否则取 0。若 $\boldsymbol{U}[t]$ 超过发放阈值 V_{th}，神经元将发放一个脉冲 $\boldsymbol{S}[t]$，并在随后的更新中 $\boldsymbol{U}[t]$ 进行减法复位；若未超过阈值，则不发生发放。随后，$\boldsymbol{U}[t]$ 以衰减因子 β 衰减并写入为 $\boldsymbol{H}[t]$，其中 β 示膜电位的衰减常数。为便于表述，下文将上述的发放过程抽象为一层脉冲神经元 SN(·)，其输入为膜电位张量 $\boldsymbol{U}$，输出为脉冲张量为 $\boldsymbol{S}$。

然而，传统 LIF 神经元采用二值输出表示是否发放脉冲，虽然强化了稀疏性，但表达能力受限，难以体现复杂输入的差异性。为提升表达精度，Luo 等 [14] 提出了 I-LIF 神经元，其将输出扩展为多个离散整数值，从而在不显著增加计算复杂度的前提下提升网络性能。其动态特性为：

$$\boldsymbol{U}[t] = \boldsymbol{H}[t-1] + \boldsymbol{X}[t] \tag{9}$$

$$\boldsymbol{S}[t] = \text{Clip}\left(\text{round}\left(\boldsymbol{U}[t]\right), 0, D\right) \tag{10}$$

$$\boldsymbol{H}[t] = \beta\left(\boldsymbol{U}[t] - \boldsymbol{S}[t]\right) \tag{11}$$

式中：round(·) 表示取整数运算；Clip($\boldsymbol{x}$, min, max) 表示 $\boldsymbol{x}$ 限幅区间在 [min, max]；D 表示超参数，即 I-LIF 神经元可以发放的整数上限。

在本研究中，考虑到近红外眼部图像中主要区域瞳孔、虹膜、眼白、背景在红外图像上存在显著区分，基于 I-LIF 构建了 A-LIF 神经元，使用固定的 0~4 之间离散整数进行脉冲量化，通过在相同的时间步给出的不同脉冲强度来区分不同的信息。其动态特性为：

$$\boldsymbol{U}[t] = \boldsymbol{H}[t-1] + \boldsymbol{X}[t] \tag{12}$$

$$\boldsymbol{S}[t] = \text{Clip}\left(\text{round}\left(\boldsymbol{U}[t]\right), 0, 4\right) \tag{13}$$

$$\boldsymbol{H}[t] = \beta\left(\boldsymbol{U}[t] - \boldsymbol{S}[t]\right) \tag{14}$$

式中：Clip($\boldsymbol{x}$, 0, 4) 表示 $\boldsymbol{x}$ 限幅区间在 [0, 4]。图 3 展示了三种脉冲神经元的量化过程。

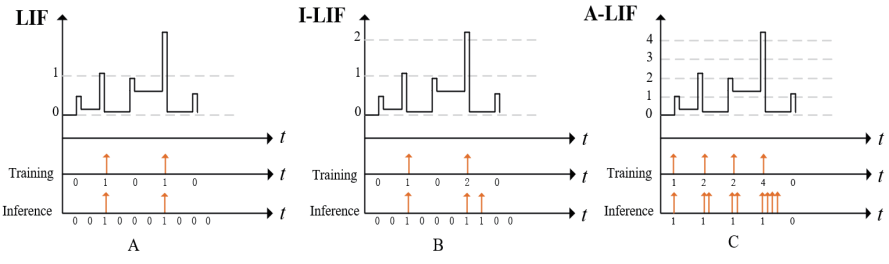

图 3 三种脉冲神经元

2 实验结果及其分析

2.1 实验设计

目前公开的近红外眼球数据集有很多，其中 CASIA 数据集被广泛用于瞳孔检测 [15]，但 CASIA 数据集数据量较小，无法表现瞳孔在各种角度下的情况。为了验证所提方法在近眼红外图像上的检测性能，选取数据量更大的公开可用数据集 openEDS 进行实验。图 4 为两个数据集中的近眼红外图像。

（a）CASIA 数据集　　（b）openEDS 数据集

图 4 两个数据集中的近眼红外图像

OpenEDS（open eye dataset）是由 Facebook Reality Labs 构建的一个高质量、公开可用的眼球数据集，专为虚拟现实（VR）应用中的眼动追踪研究和瞳孔检测而设计。该数据集使用定制的头戴式显示设备（HMD），配备两个同步红外摄像头，在 200 Hz 的高帧率受控光照条件下采集，涵盖了 152 名受试者的眼部图像数据。数据集包括四个子集 127 597 张带有精确像素级分割标注的图像，91 200 帧视频序列，以及 143 对由角膜地形仪获取的左眼和右眼点云数据。在本次实验中，使用了 27 431 张红外眼图进行训练，2 743 张红外眼图用于验证，2 743 张红外眼图用于测试。

在本研究中，使用 PyTorch 深度学习框架对图像数据集进行了训练和测试。硬件配置主要包括运行频率为 2.50 GHz 的英特尔 (R) 至强 (R) 铂金 8255C CPU 和配备 4 个 24 GB 内存的 NVIDIA GeForce RTX 4090 GPU。

为了评估本文算法的性能，选取了参数量、精度均值在 IoU=0.5 时的值（mAP@50）、精度均值在 0.5~0.95 之间的值（mAP@50:95），以及功耗作为指标。具体而言，人工神经网络和脉冲神经网络的功耗可以用公式表示为：

$$E_{\text{ANN}} = O^2 \times C_{\text{in}} \times C_{\text{out}} \times k^2 \times E_{\text{MAC}} \tag{15}$$

$$E_{\text{SNN}} = \left(T \times D\right) \times \text{fr} \times O^2 \times C_{\text{in}} \times C_{\text{out}} \times k^2 \times E_{\text{AC}} \tag{16}$$

式中：o 为特征输出尺寸；c_{in} 与 c_{out} 分别为输入通道数和输出通道数；k 为卷积核大小；fr 为平均脉冲发放频率；T 为时间步长；D 为训练阶段整数激活的上限。

采用 SNN 常用的评估方法 [16]。所有运算均假设在 45 nm 工艺上的 32 位浮点实现，其中乘加运算功耗 E_{MAC}=4.6 PJ，纯加法运算功耗 E_{AC}=0.9 PJ，由上述公式可知，SNN 的低功耗主要得益于其稀疏的加法型计算；脉冲越少，计算越稀疏。

和平均精度的计算用公式表示为：

$$\text{AP} = \int_0^1 P\left(R\right) \text{d}R \tag{17}$$

$$\mathrm{mAP}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{AP}_i \tag{18}$$

其中，mAP@0.5 指在 IoU 阈值为 0.5 时的平均精度；mAP@50:95 指在多个 IoU 阈值从 0.5 到 0.95，步长 0.05 下计算得到的平均精度的均值。后者能够更全面评估。

2.2 实验结果

本文在 OpenED 数据集上对所提出的方法进行了全面评估，在所有实验中，将衰减因子 β 设置为 0.25，学习率设置为 0.01，并采用 SGD 优化器。

模型在 4 个 NVID-IA-4090GPU 上以 50 的批大小训练 200 个周期。在本文方法中，推理时间步长以 $T\times D$ 的形式报告，例如，1×4 表示的含义是 T=1，D=4，其中 D 表示在直接编码模式下的静态图像重复次数。主要结果如表 1 所示。

表 1 实验结果

Model	Param /10^6	Power /mJ	$T\times D$	mAP@50 /%	mAP@50:95 /%
YOLOv5	21.2	112.5	1	99.5	97.1
DETR[17]	41.0	197.5	1	96.3	88.2
PVT[18]	32.9	520.3	1	88.2	76.9
YOLOv8n-eye[10]	2.79	—	1	99.6	96.4
EMS-YOLO[8]	26.9	29.0	4	90.1	60.3
SpikingYOLO[14]	13.2	23.1	1×4	99.5	96.2
Pupil-YOLO(Our)	0.153	8.9	1×4	99.5	95.6
Pupil-YOLO(Our)	0.108	6.3	1×4	99.5	93.8

实验结果表明，本文所提出的 Pupil-YOLO 模型在瞳孔检测中显著降低了模型参数量和功耗。仅用 0.15×10^6 的参数便获得了 99.5% 的 mAP@50 和 95.3% 的 mAP@50:95。与 EMS-YOLO 相比，参数仅为其 1/179，检测精度 mAP@50:95 提升了 35.3%。与目前最先进的 SNN 通用目标检测算法 SpikingYOLO 相比，虽然在 mAP@50:95 上检测精度下降了 0.6% 但是参数是其 1/88。此外以目前最先进的近眼红外瞳孔检测模型 YOLOv8n-eye[10] 相比参数在精度相差 0.8% 的情况下，参数量仅为其 1/18。

图 5 显示了模型在 OpenEDS 数据集上训练与验证过程中的性能曲线。上行表示训练阶段的各项损失与检测指标变化情况，下行对应验证阶段结果。

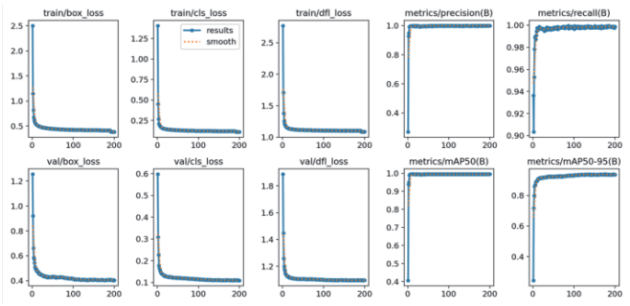

图 5 型训练与验证指标监控图

图 6 显示了模型在各种角度下的瞳孔检测情况，即使在瞳孔显示不完整的情况下也可以精准识别。

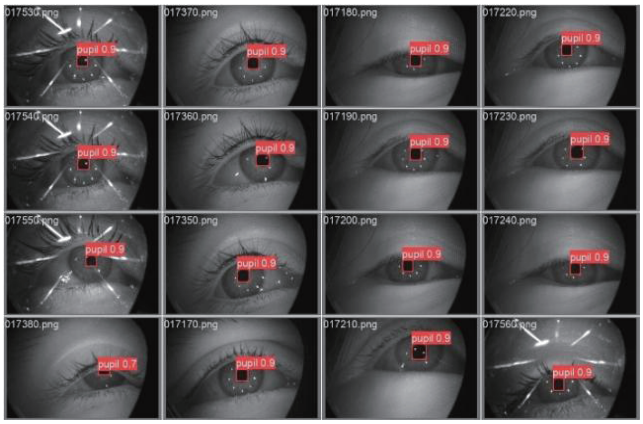

图 6 不同情况下瞳孔识别的情况

2.3 消融实验

为阐明所提方案中各不同模块的单独贡献，针对 openEDS 数据集执行了一系列消融研究。实验数据如表 2 所示，其内容为：EXP1 作为基准实验，采用图 1 的架构进行训练。EXP2 将 EXP1 中的 P-Block1 模块换成 SNNConv[14]，SNNConv 模块是一种将脉冲神经元与 CNN 简单结合的卷积模块。EXP3 将 EXP1 中的神经元的量化值由 0~4 改为 0~1 量化。本文以模型 mAP@50:95、模型参数和脉冲神经元时间步作为评价标准。从 EXP2 和 EXP1 的对比中可以看出，P-Block1 尽可能保留最大特征的同时减少参数，在相同精度下参数为其 1/5。最后通过 EXP1 和 EXP3 可以看出，量化上线越大，量化误差越小。但如果过度提高量化的上限，将导致增加脉冲发放的个数，使其功耗上升。

表 2 消融实验

	Param/10^6	mAP@50:95/%	$T\times D$	Epoch	量化范围
EXP1	0.153	95.3	1×4	200	0~4
EXP2	0.82	95.3	1×4	200	0~4
EXP3	0.153	91.9	1×4	200	0~1

3 总结

本文面向端侧资源受限场景，提出轻量化瞳孔检测框架 Pupil-YOLO，在保持 YOLO 宏观架构的基础上进行轻量化设计并同时引入 SNN 机制实现低功耗；通过构建轻量特征块 P-Block1 实现在低参数下的特征提取；在神经元建模上提出固定整数泄漏的 A-LIF，并将激活量化至 0~4，降低量化误差精度，使推理走整数路径、契合事件稀疏性与嵌入式算力。基于 openEDS 数据集的实验表明，在仅仅 0.15×10^6 参数规模下，模型实现 99.5%mAP@50 与 95.3%mAP @50:95 的

检测精度，在复杂角度与遮挡场景下仍具鲁棒性。总体而言，Pupil-YOLO 以轻量结构、脉冲计算、整数量化的协同设计在精度、效率与可部署性之间取得均衡。当前局限在于验证数据以 openEDS 为主尚未系统量化多平台能耗与跨域泛化；后续将扩展红外近眼数据与跨设备测试，联动虹膜分割与凝视估计构建端侧一体化流程，并探索事件相机融合与脉冲时序注意力。

参考文献：

[1]JETTER H C, SCHRÖDER J H, GUGENHEIMER J,et al.Transitional interfaces in mixed and cross-reality: a new frontier?[C]//ISS Companion'21: Companion Proceedings of the 2021 Conference on Interactive Surfaces and Spaces. NewYork: ACM, 2021: 46-49.

[2]CAMPOS-CASTILLO E. HIERRO-GUTIERREZ E G D.Eye control and accessibility in computers[C]//XLVII Mexican Conference on Biomedical Engineering.Berlin: Springer, 2025: 219-227.

[3]CAMPBELL I, BECKERS E, SHARIFPOUR R, et al. Impact of light on task-evoked pupil responses during cognitivetasks[J].Journal of sleep research, 2024, 33(4): 14101.

[4] CHENG Y H, YUAN X Y, JIANG Y.Eye pupil signals life motion perception[J].Attention perception and psychophys, 2023, 86: 579-586.

[5]WILDENMANN U, SCHAEFFEL F.Variations of pupil centration and their effects on video eye tracking[J].Ophthalmic physiologic optic, 2013, 33(6): 634-641.

[6]EIVAZI S, SANTINI T, KESHAVARZI A, et al.Improving real-time CNN-based pupil detection through domain-specific data augmentation[C]//ETRA '19: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications,NewYork:ACM, 2019:1-6.

[7]VERMA G, BINDAL N, NISAR A,et al.Advances in neuromorphic spin-based spiking neural networks: a review[J]. IEEE nanotechnology mag, 2021,15(5):33-44.

[8]SU Q Y, CHOU Y H, HU Y F,et al.Deep directly-trained spiking neural networks for object detection[EB/OL].(2023-07-27)[2025-12-21].https://doi.org/10.48550/arXiv.2307.11411.

[9]YAO M, QIU X R, HU T X,et al.Scaling spike-driven transformer with efficient spike firing approximation training[EB/OL].(2024-11-25)[2025-12-31].https://doi.org/10.48550/arXiv.2411.16061.

[10]XUE K J, WANG J S, WANG H. Research on pupil center localization detection algorithm with improved YOLOv8[J]. Applied sciences, 2024,14(5):6661.

[11]KIM Y, PARK H, MOITRA A,et al.Rate coding or direct coding: which one is better for accurate, robust, and energy-efficient spiking neural networks?[C]//ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway:IEEE,2022:71-75.

[12] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway:IEEE,2018[2025-12-30]. https://ieeexplore.ieee.org/abstract/document/8578572.DOI: 10.1109/CVPR.2018.00474.

[13]MAASS W.Networks of spiking neurons: the third generation of neural network models[J].Neural networks,1997,10(9):1659–1671.

[14]LUO X H, YAO M, CHOU Y H, et al.Integer-valued training and spike-driven inference spiking neural network for high-performance and energy-efficient object detection[C]//Computer Vision – ECCV 2024.Berlin: Springer, 2024: 253-272.

[15]XIONG Q, ZHANG X M, WANG X Z,et al.Robust Iris-localization algorithm in non-cooperative environments based on the improved YOLOv4 model[J].Sensors, 2022, 22(24): 9913.

[16]YAO M, ZHAO G S, ZHANG H Y, et al.Attention spiking neural networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2023, 45(8): 9393-9410.

[17]CARION N, MASSA F, SYNNAEVE G, et al.End-to-end object detection with transformers[EB/OL].(2020-05-28)[2025-12-31].https://doi.org/10.48550/arXiv.2005.12872.

[18]WANG W H, XIE E Z, LI X, et al.Pyramid vision transformer: a versatile backbone for dense prediction without convolutions[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 548-558.

【作者简介】

朱厚帅（1998—），男，湖北武汉人，硕士研究生，研究方向：信息感知与机器学习。

汪耀（1991—），男，湖北武汉人，硕士研究生，研究方向：信号处理和机器学习。

（收稿日期：2025-11-24 修回日期：2026-01-09）