

基于深度特征匹配的白内障术中眼球旋转角度检测方法

赵文涛¹⁾, 续欣莹¹⁾, 谢琨²⁾, 程兰¹⁾, 张喆^{1)*}

¹⁾(太原理工大学电气与动力工程学院 太原 030024)

²⁾(太原理工大学信息与计算机学院 晋中 030600)

(zhangzhe@tyut.edu.cn)

摘要: 为了实现对白内障手术中眼球旋转角度的测量, 提出基于深度特征匹配的白内障术中眼球旋转角度检测方法, 即对术前基准图像与术中图像中角膜缘周围区域的特征点进行提取与匹配, 以计算术中眼球旋转角度。针对角膜缘周围纹理特征丰富但相似性高、特征易受手术进程与器械干扰而发生显著变化等问题, 提出一种结合注意力卷积模块和自适应跳层连接的自监督图像局部特征提取与描述模型。在注意力卷积模块中, 利用坐标注意力机制加强模型对于方向和空间位置信息的准确感知, 并通过条件参数化深度卷积提升模型容量, 增强对于特征信息的表示能力; 在自适应跳层连接中, 通过融合深层语义信息和浅层结构信息获得对特征点更具区分性的描述。在 CATARACT 数据集上的实验结果表明, 所提模型在各误差限下的特征点平均匹配精度均优于其他同类模型; 所提方法的眼球旋转角度测量误差为 0.740° , 实时检测速度为 36.675 帧/s, 满足白内障手术中眼球旋转角度检测精度和实时性的要求。

关键词: 眼球旋转角度检测; 白内障手术; 深度学习; 特征提取与描述; 注意力机制

中图法分类号: TP391.41

DOI: 10.3724/SP.J.1089.2024.19997

Detection Method of Eye Rotation Angle in Cataract Surgery Based on Depth Feature Matching

Zhao Wentao¹⁾, Xu Xinying¹⁾, Xie Jun²⁾, Cheng Lan¹⁾, and Zhang Zhe^{1)*}

¹⁾(College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024)

²⁾(College of Information and Computer, Taiyuan University of Technology, Jinzhong 030600)

Abstract: In order to measure the eye rotation angle during cataract surgery, a detection method of eye rotation angle in cataract surgery based on depth feature matching is proposed, which extracts and matches the feature points of the preoperative reference image and the intraoperative image around the corneal limbus to calculate the intraoperative eye rotation angle. The texture features around the limbus are rich but their similarity is high, and the features are susceptible to obvious changes due to the interference of the surgical process and instruments. To solve these problems, a self-supervised image local feature extraction and description model is proposed, which combines attention convolution block (AttConvBlock) and adaptive skip connection. First, AttConvBlock enhances the model's accurate perception of orientation and spatial location information by coordinate attention. Besides, AttConvBlock improves the capacity of the model through conditionally parameterized depthwise convolutions, which can enhance the representation ability of the model for feature information. Furthermore, the adaptive skip connection fuses deep semantic information

收稿日期: 2022-10-08; 修回日期: 2023-03-01. 基金项目: 国家自然科学基金(62073232); 山西省自然科学基金(202103021224056); 山西省科技合作交流专项(202104041101030). 赵文涛(1996—), 男, 硕士研究生, 主要研究方向为计算机视觉、医学图像处理; 续欣莹(1979—), 男, 博士, 教授, 硕士生导师, CCF 会员, 主要研究方向为计算机视觉、智能控制; 谢琨(1979—), 女, 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为机器学习与计算智能、情感分析与生成、智能推荐系统、医学图像处理; 程兰(1984—), 女, 博士, 副教授, 硕士生导师, 主要研究方向为计算机视觉、机器人视觉自主定位与建图; 张喆(1991—), 女, 博士, 讲师, CCF 会员, 论文通信作者, 主要研究方向为计算机视觉、模式识别。

and shallow structural information, which contributes to a more discriminative description of feature points. The experimental results on the CATARACT dataset show that the proposed model has higher mean matching accuracy under each error limit than other compared models. Additionally, the mean rotation error of the proposed method is 0.740° , and the real-time detection speed is 36.675 frames per second, meeting the requirements of accuracy and real-time detection of eye rotation angle in cataract surgery.

Key words: eye rotation angle detection; cataract surgery; deep learning; feature extraction and description; attention mechanism

随着白内障手术技术^[1]和屈光性人工晶状体的不断发展和日益成熟,白内障手术正由单一的复明手术向可以提升患者视觉质量的屈光性手术迈进.在屈光性白内障手术中,角膜轴位与人工晶状体散光轴的准确对位影响着患者术后视觉恢复的效果,相关研究表明,1°的轴位偏差可导致矫正的散光量减少 3.3%^[2].

传统的白内障手术方案中,医生会在术前对患者的角膜子午线两端进行有痕标记^[3],保证手术时的准确对位;然而这种标注方法存在固有的缺陷,如接触式标记可能会在术中褪色,并带来一定的感染风险.虽然一些商用系统,如 VERION 数字导航系统^[4]和 CALLISTO

eye 导航系统^[5]可以在屈光性白内障手术中实现角膜散光轴位的无接触式标记,但其技术细节尚未披露. Zhai 等^[6]公开了一套计算机辅助下的白内障术中散光轴位标记方案,通过与术前基准图像对比,实时地检测术中每帧图像中眼球的旋转角度,实现对散光轴位的标记;该方案对角膜缘周围提取出的多个感兴趣区域(region of interest, ROI)进行跟踪,通过计算全体 ROI 的平均位移实现对眼球旋转角度的测量,但 ROI 的选择过程较为烦琐,且区域易受眼球挤压形变的影响,导致 ROI 的跟踪失败.为此,本文提出一种基于特征点检测与匹配的眼球旋转角度测量方法,通过对基准图像与术中图像中角膜缘周围区域的密集特征点进行提取并匹配为多组特征点对,计算每组特征点对的旋转角度,用于估计眼球的旋转角度.基于特征点匹配的方法可以更精准地描述角膜缘周围特征的移动情况,使得对眼球旋转角度的计算更准确,能在一定程度上提高测量的容错能力和鲁棒性.

眼球旋转角度的跟踪测量还依赖于眼球中心位置和眼球大小的精确测定,可以通过成熟的语义分割模型(如 U-Net^[7], DeepLabV3+^[8], BiSeNet^[9]

等)对眼球区域进行分割,使用椭圆拟合算法对分割后的眼球区域边界进行拟合以确定眼球的几何参数,本文将其作为基本的前置任务,不在本文中讨论.一般的特征匹配需要经过 4 步: (1) 提取特征点; (2) 计算局部特征描述符; (3) 通过匹配模型进行匹配; (4) 剔除异常匹配. 其中,前 2 步是得到更多成功匹配的关键,本文针对角膜缘周围特征的提取与描述展开工作.

传统的特征点匹配模型依赖于手工设计特征提取与描述策略.尺度不变特征变换(scale invariant feature transform, SIFT)^[10]是计算机视觉中经典的传统局部特征提取与描述符计算方法,它依靠图像尺度空间模型获取具有尺度不变性的特征点.其他基于手工设计的特征描述方法还有 SURF^[11], ORB^[12], BRISK^[13]等.传统的特征点提取模型易受图像质量与场景变化的影响,导致误匹配率较高,在实践场景中表现较差.当前,特征匹配的深度学习方法通常使用卷积神经网络,从数据中学习更鲁棒的特征点检测器和局部描述符. DeTone 等^[14]提出的 SuperPoint 模型能够同时预测稀疏的特征点和描述符,通过人造几何数据训练一个 MagicPoint 网络作为特征点提取器,并利用该网络提取数据集中的特征点并送入 SuperPoint 模型训练特征点和描述符,所提取的特征点仅由几何数据的连接点所定义,提取到的特征点数量较为稀疏. Dumanu 等^[15]提出的 D2-Net 仅由单个卷积神经网络同时预测一个密集的特征检测器和特征描述符,实现了检测器与描述符的高度耦合,并直接使用描述符通道上的局部最大值作为特征点.在 D2-Net 的基础上, Revaud 等^[16]提出的 R2D2 增加对特征点可重复性和描述符可靠性的估计,增强了对特征点的检测和描述能力;然而, R2D2 模型的特征提取网络较为简单,其特征描述符的提取只依赖于网络提取的深度特征信息,忽略了浅层特征中的颜色、纹理、边缘等信息对特征描述的重要性.

角膜缘周围的毛细血管区域蕴含着大量相似的纹理细节等结构信息,且手术过程中易受到手术操作与器械挤压,导致特征区域发生明显的变化,增加了对特征描述的难度.为了准确地对角膜缘周围特征点提取与描述,本文提出结合注意力卷积模块(attention convolution block, AttConvBlock)和自适应跳层连接的特征提取与描述模型.本文将R2D2模型作为基准模型,引入坐标注意力机制(coordinate attention, CA)^[17],使得模型在保留空间精确位置信息的同时,建立像素特征点对于不同方向特征的长程依赖关系,增强了描述符对空间位置的描述能力,减少跨区域错误匹配的发生;引入条件参数化深度卷积(conditionally parameterized depthwise convolutions, CondDwConv),在提升模型容量的同时,依据不同的输入自适应地调整卷积核参数,有助于提升网络对于不同特征的代表能力;最后通过一条简单而高效的跳层连接将浅层丰富的细节信息与深层语义信息进行自适应融合,增强网络对于细节特征的提取能力.本文在公开的白内障手术视频数据集 CATARACT^[18]上的实验结果表明,当目标区域特征受手术操作与器械干扰导致区域特征发生变化时,本文方法与模型仍能获得较为精确的匹配.

1 本文方法

提取与描述基准图像与术中图像角膜缘周围区域的特征点,并根据特征点匹配结果,实现白内障手术过程中眼球旋转角度的检测.本文方法的流程如图1所示,步骤如下:

Step1. 图像眼球区域裁剪.为了提升模型的推理性能,通过眼球椭圆几何参数 (x, y, r_a, r_b) 对图像进行自适应裁剪,只保留眼球及角膜缘周围区域.其中, (x, y) , r_a 和 r_b 分别表示眼球中心坐标、眼球半长轴和眼球半短轴,它们均由前置任务获得,本文实验中直接采用眼球几何参数真值.裁剪区域中心为眼球中心 (x, y) ,区域宽度与高度均为 $2 \times (r_a + W_G)$,其中, W_G 表示环宽增益,实验中设置 $W_G = 30$,约为眼球平均直径的1/10.

Step2. 提取图像特征点及其描述符.分别提取基准图像与术中图像中的所有特征点及其描述符.

Step3. 目标匹配区域确定及特征点筛选.将位于角膜缘周围的环状区域作为特征点的目标匹配区域,环状区域中心为眼球中心 (x, y) ,外圆半径为 $r_a + W_G$,内圆半径为 r_b ;筛选出落在目标匹配区域内的特征点.

Step4. 特征点匹配.对于基准图像与术中图像中

落在目标匹配区域内的特征点,通过最近邻匹配算法进行匹配,即计算基准图像与术中图像中提取的特征描述符间的 L_2 距离,将对应距离最小且小于匹配阈值($N = 0.85$)的特征描述符作为最佳匹配,得到 M 对匹配特征点对.

Step5. 旋转角度计算.根据每个特征点对与眼球中心的位置关系计算旋转角度,得到旋转角度序列 $S_{\text{Rot}} = [r_1, r_2, \dots, r_M]$.剔除旋转角度序列中的明显异常值,若所估计的旋转角度满足

$$|r_i - M_e(S_{\text{Rot}})| > T, i \in [1, M],$$

则将此旋转角度与对应的特征点对判定为异常值.其中, $M_e(S_{\text{Rot}})$ 表示旋转角度序列的中位数; T 表示剔除阈值,实验中设置 $T = 20$.剔除异常角度后,旋转角度序列为 $S'_{\text{Rot}} = [r'_1, r'_2, \dots, r'_m]$, $m \leq M$,并将其中位数作为最终所估计的眼球旋转角度 $R_p = M_e(S'_{\text{Rot}})$.

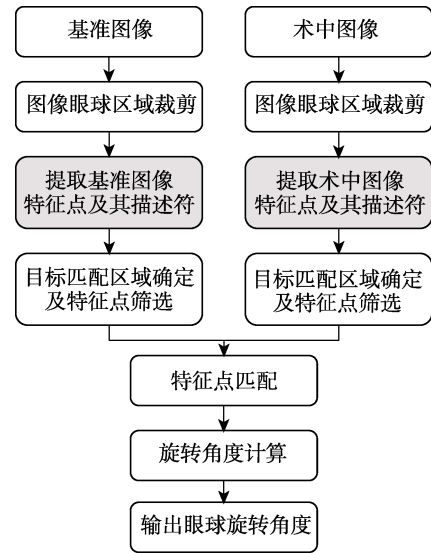


图1 本文方法流程

特征点与描述符检测的准确度直接影响特征点匹配的精度,进而影响眼球旋转角度测量的精度.针对白内障手术场景中角膜缘周围区域特征不易区分且易受手术操作发生明显变化的情况,本文提出更准确的特征点提取与描述模型.

2 本文模型

R2D2模型中将 $L_2\text{-Net}$ ^[19]作为特征提取网络,并对其改进以提高网络的计算效率,即将 $L_2\text{-Net}$ 中最后1个 8×8 卷积层替换为3个 2×2 卷积层.对R2D2的特征提取网络进行深入地探索与改进,提升其在白内障手术场景中对图像特征的提取和描述能力,获得更精细而准确的特征融合信息和空

间位置特征信息. 本文模型的结构如图 2 所示, 将 RGB 三通道图像 $I \in \mathbb{R}^{H \times W \times 3}$ 作为输入, 经过特征提取网络得到对应的 3 个输出: 第 1 个输出为描述符 $X \in \mathbb{R}^{H \times W \times 128}$, 每个像素位置处的 128 维特征向量为该特征点提供特有的描述信息; 第 2 个输出对应描述符的可靠性 $R \in [0, 1]^{H \times W \times 1}$, 表征每个像素位置处所估计的描述符的可靠程度, 即判别性; 第 3 个输出对应特征点位置的可重复性 $S \in [0, 1]^{H \times W \times 1}$, 其目标是提供稀疏且可重复的关键特征点的位置. 本文模型在特征提取网络中设计并添加了 AttConvBlock, 使得网络更加关注感兴趣特征信息和特征的空间位置, 同时为特征建立长程依赖关系; 其次, 通过简单而高效的自适应跳层连接, 实现跨层次跨尺度的特征融合和信息交互.

2.1 AttConvBlock

融合 CondDwConv 和 CA, 设计如图 2 所示称为 AttConvBlock 的特征提取模块. 在 AttConvBlock 中, 输入特征首先通过 CondDwConv 获得对输入敏感的特征输出; 然后通过 CA 获得带有方向和空间注意力信息的特征图; 通过 1×1 卷积实现不同特征通道间的信息交互, 并在 AttConvBlock 中使用残差连

接防止特征退化, 增强对特征的重用能力.

2.1.1 CondDwConv

深度卷积神经网络性能的提升在很大程度上来自于模型表示力的提升, 动态卷积可在引入较少参数的情况下, 有效地提升模型的容量, 增强对特征的表示力. 其中, 条件参数化卷积(conditionally parameterized convolutions, CondConv)^[20]为每个输入示例学习特有的卷积核参数, 即将卷积核参数化为多个专家知识的线性组合. 对于输入特征图 $x \in \mathbb{R}^{H \times W \times C}$, CondConv 的输出特征可表示为

$$f(x) = \sigma((\alpha_1 W_1 + \alpha_2 W_2 + \dots + \alpha_n W_n) * x).$$

其中, $W_i \in \mathbb{R}^{k \times k \times C_d \times C_o}$ 表示第 i 个专家核参数; k 表示卷积核的大小; C_d 表示卷积核的深度; C_o 表示卷积核的数量(即对应输出特征图的通道数); n 表示专家的数量; σ 表示激活函数; 权重系数 α_i 通过路由函数对输入特征 x 的学习得到, $i \in [1, n]$. 路由函数由全局平均池化、全连接层与 Sigmoid 激活函数构成. CondConv 可通过提升专家数量达到提升网络容量的目的, 并且可以有效地针对不同输入自适应地计算卷积核参数, 使得卷积更好地学习与表达输入特征.

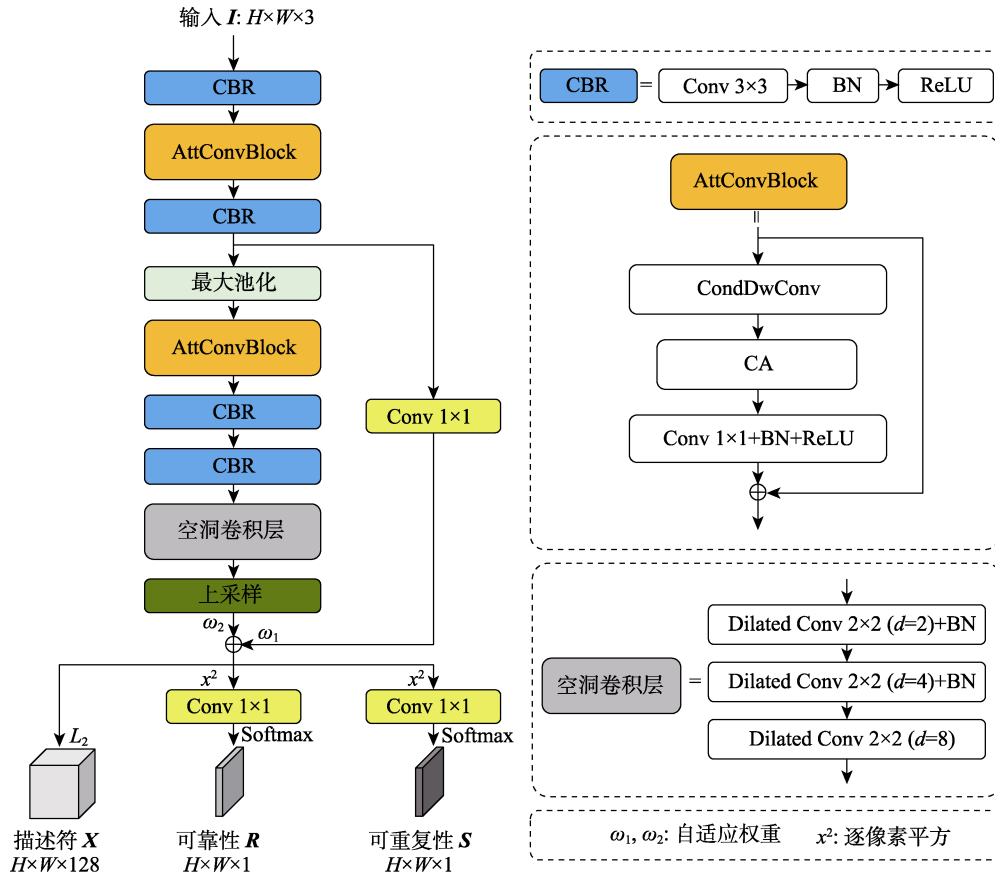


图 2 本文模型结构

将 CondConv 中的普通卷积操作替换为深度卷积^[21]操作, 称为 CondDwConv, 如图 3 所示. 深度卷积的每个深度为 1 的卷积核只与输入特征图的一个通道进行卷积, 即核参数 $W \in \mathbb{R}^{k \times k \times 1 \times C}$, 进一步降低卷积过程中的参数量和计算成本.

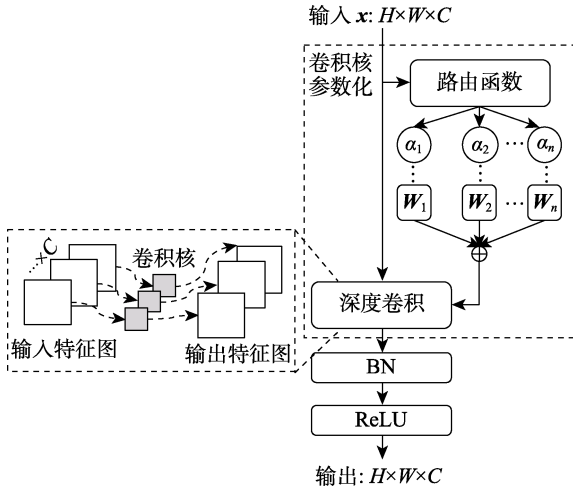


图 3 CondDwConv 的结构

2.1.2 CA

视觉注意力机制^[22]使得网络重点关注图像中重要的特征信息而抑制不相关的信息, 可以看作是对图像特征的动态权重调整过程; 同时, 注意力机制可以在一定程度上建立长程像素特征间的依赖关系. 作为一种即插即用的特征增强模块, 注意力机制已被很多计算机视觉任务采用^[23-24]. 本文引入 CA 提升模型对于关键特征信息的空间位置描述能力.

CA 将特征的位置信息嵌入到通道注意中, 其实现过程如图 4 所示. 对输入特征分别在 X 和 Y 方向进行全局池化, 获得在 2 个不同方向上的特征编码; 然后进行通道拼接操作, 实现 2 个空间方向特征的聚合, 并经过 1 个 1×1 卷积、批标准化(batch normalization, BN)和非线性激活函数 h-swish 生成注意力特征图, 其中, 比例因子 r 用于控制通道数的规模以减少计算量; 再次分别沿着 X 和 Y 方向对注意力特征图进行通道分离, 并分别通过 1×1 卷积和 Sigmoid 激活函数获得特征在 2 个不同方向上的注意力权重, 将其与原始输入特征矩阵作点乘. CA 对 2 个不同方向进行特征变换操作, 使得网络可以沿着一个空间方向捕捉特征间的长程依赖关系, 同时沿着另一个空间方向保留特征的精确位置信息. CA 的引入使得特征提取网络具有感知方向和空间位置信息的能力, 同时建立特征间远距离的依赖关系, 增强了网络生成的描述符对特征的代表能力.

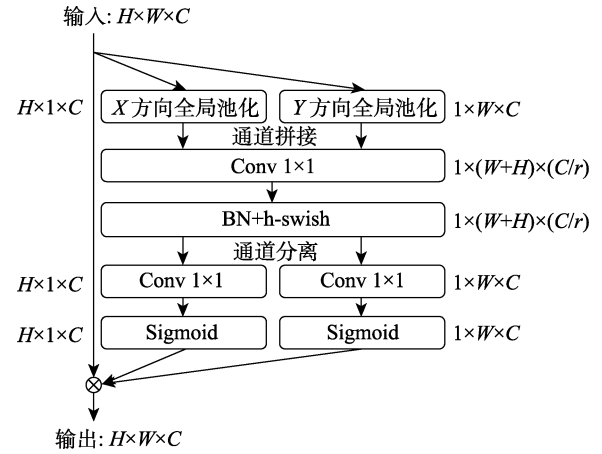


图 4 CA 的结构

2.2 自适应跳层连接

下采样可以有效地降低数据维度, 减少网络参数量, 同时增大感受野; 然而, 过多的下采样会导致大量有用特征信息的丢失, 对于图像特征提取与描述任务十分不利. 因此, 本文模型仅在特征提取网络中使用一次最大池化操作, 将特征图的高与宽下采样为原始尺寸的 $1/2$, 提高网络的计算效率, 并使得特征图仍保持较高的分辨率, 避免大量有用特征信息的丢失.

随着网络深度的加深, 模型更倾向于学习一种抽象的语义特征信息. 然而在特征描述网络中, 浅层的颜色、纹理等丰富的结构信息对于局部特征的表达也十分重要, 浅层特征图具有更精细化的高分辨率特征信息. 为此, 通过在模型中添加一条跳层连接(如图 2 所示), 对不同层次、不同尺度的特征进行自适应融合. 将浅层特征图通过 1×1 卷积, 在保持分辨率的同时提升通道数得到特征图 $F_1 \in \mathbb{R}^{H \times W \times C}$; 将深层特征图通过双线性插值上采样为与原图尺度一致的特征图 $F_2 \in \mathbb{R}^{H \times W \times C}$, 通过自适应生成权重 w_1 与 w_2 实现特征图的自适应融合, 得到特征图 $F = w_1 F_1 + w_2 F_2$.

2.3 损失函数

与 R2D2 采用的方法相似, 本文通过自监督的方式进行特征点提取与描述符生成的训练. 其损失函数可分为特征点可重复性损失函数与描述符可靠性损失函数.

2.3.1 特征点可重复性损失函数

设 I 和 I' 分别为同一场景下的 2 幅图像, 令 $U \in \mathbb{R}^{H \times W \times 2}$ 表示其真实的映射关系. 如果图像 I 和 I' 为自然图像, 则 U 可以使用光流或立体匹配算法进行估计; 如果 I' 为 I 通过变换而生成的图像, 那么 U 是已知的变换矩阵. 令 S 和 S' 分别表

示图像 I 和 I' 的可重复性得分图, S'_U 为得分图 S' 经过 U 变换后得到; 则将图像的可重复性表征为 S 中的所有局部最大值与 S'_U 中的所有局部最大值相对应. 为了避免图像中遮挡、伪影和边界效应的影响, 可将 S 和 S'_U 划分成大小为 $N \times N$ 的可重叠补丁集合 $P = \{p\}$, 并计算每个局部补丁区域余弦相似度的均值. 因此, 损失函数可定义为

$$L_{\cosim}(I, I', U) = 1 - \frac{1}{|P|} \sum_{p \in P} \cosim(S[p], S'_U[p]).$$

其中, $S[p] \in \mathbb{R}^{N^2}$ 表示将从 S 中提取的大小为 $N \times N$ 的补丁 p 扁平化为一维向量; $S'_U[p]$ 同理. 为了避免网络将 S 和 S'_U 错误地预测为恒定的常量图而使得余弦相似度最小, 引入损失函数 L_{peaky} 来最大化可重复性得分图的局部峰值, 公式为

$$L_{\text{peaky}}(I) = 1 - \frac{1}{|P|} \sum_{p \in P} \left(\max_{(i,j) \in p} S_{ij} - \text{mean}_{(i,j) \in p} S_{ij} \right).$$

最终, 特征点可重复性损失函数被定义为 $L_{\cosim}$ 与 L_{peaky} 的加权函数, 公式为

$$L_{\text{rep}}(I, I', U) = L_{\cosim}(I, I', U) + \frac{1}{2} (L_{\text{peaky}}(I) + L_{\text{peaky}}(I')).$$

2.3.2 描述符可靠性损失函数

不同特征点的描述符应尽可能地具有区分性, 为图像 I 中每个像素 (i, j) 定义一个尺寸为 O 的补丁, 将其与图像 I' 中所有的其他补丁计算平均精度 (average precision, AP)^[25], 优化补丁描述符的匹配; 并为每个描述符 $X_{ij} \in \mathbb{R}^{128}$ 预测一个置信度 $R_{ij} \in [0, 1]$, 判别此特征描述符的可靠性. 描述符可靠性损失函数定义为

$$L_{\text{APK}}(i, j) = 1 - [\text{AP}(i, j) R_{ij} + \kappa (1 - R_{ij})].$$

其中, 超参数 $\kappa \in [0, 1]$ 表示对于每个补丁的最小期望 AP, 本文实验中设置 $\kappa = 0.5$.

训练模型的总损失函数公式为

$$L_{\text{total}} = L_{\text{rep}} + L_{\text{APK}}.$$

3 实验结果与分析

3.1 实验数据及实验细节

3.1.1 数据集

本文将 Aachen Day-Night 数据集^[26]和 Oxford and Paris 数据集^[27]作为预训练数据集, 将 CATARACT 数

据集^[18]作为主训练数据集与测试集. CATARACT 数据集中共包含 20 位白内障患者的手术视频序列, 并划分为 3 组: 10 个视频作为训练集, 3 个视频作为验证集, 7 个视频作为测试集. 图像原始分辨率均为 1920×1080 像素, 为了保证网络的推理速度, 将图像分辨率调整为 960×540 像素.

3.1.2 训练与测试细节

实验使用 Pytorch-1.10 深度学习框架实现, 服务器平台为 Intel Core i7-10700 2.9GHz CPU, 24GB RAM, NVIDIA Geforce RTX 2080Ti GPU.

首先在预训练数据集上对模型进行训练, 使用 Adam 优化器, 设置优化器平滑参数 $\beta_1 = 0.9$, $\beta_2 = 0.999$, 初始学习率为 10^{-4} , 批大小为 6, 权重衰减为 5×10^{-4} , 共训练 20 个轮次; 然后在 CATARACT 的训练集上对模型进行微调训练, 使用 Adam 优化器, 设置学习率为 10^{-6} , 批大小为 6, 权重衰减为 5×10^{-4} , 共训练 10 个轮次.

在 CATARACT 测试集上对训练好的模型进行评估, 测试集包含 7 个视频, 共 53 个图像序列, 每个序列包含 50 帧图像. 将每个视频的第 1 帧作为基准图像, 其余的图像均与基准图像进行特征点匹配以计算眼球的旋转角度.

3.2 评价指标

本文使用评估指标——旋转角度误差 (rotation error, RE) 评价模型的精度^[6], 即 $RE = |R_T - R_P|$. 其中, R_T 表示旋转角度的真值; R_P 表示旋转角度的估计值. RE 越小, 表示估计的旋转角度精度越高.

为了进一步评估特征点匹配的准确度, 本文定义特征点的平均匹配精度 (mean matching accuracy, MMA), 表示在某一误差限下正确匹配的特征点对数占有效匹配对数的百分比, 公式为

$$MMA(K) = \frac{1}{N} \sum_{i=1}^N \frac{\|(|S'_{\text{Rot}} - R_T| < K)\|_i}{\|S'_{\text{Rot}}\|_i}.$$

其中, N 表示测试集的样本数量; $\|S'_{\text{Rot}}\|$ 表示剔除异常角度后的旋转角度序列的长度, 即有效匹配特征点对数; $\|(|S'_{\text{Rot}} - R_T| < K)\|$ 表示旋转角度序列 S'_{Rot} 中满足预测旋转角度与真实旋转角度的绝对误差小于误差限 K 的个数. 本文将误差限 K 分别设置为 0.5° , 1° , 2° 和 5° .

本文还定义了平均匹配对数 (mean matching pairs, MMP) 和平均有效匹配对数 (mean valid matching pairs, MVMP), 公式分别为

$$\text{MMP} = \frac{1}{N} \sum_{i=1}^N \|S_{\text{Rot}}\|_i;$$

$$\text{MVMP} = \frac{1}{N} \sum_{i=1}^N \|S'_{\text{Rot}}\|_i.$$

其中, $\|S_{\text{Rot}}\|$ 表示最近邻匹配能匹配到的特征点总对数.

3.3 对比实验

在 CATARACT 测试集上, 将本文模型与 5 种主流的特征点提取与描述模型进行对比实验. 对比模型有 SIFT^[10], SURF^[11], SuperPoint^[14], D2-Net^[15] 和 R2D2^[16]; 所有模型均采用相同的匹配策略和异

常点剔除方法.

各模型在测试集上的 RE 评估结果如表 1 所示. 可以看出, 在手术场景发生较大变化时, 传统的 SIFT 和 SURF 模型无法对特征进行准确的区分性描述, 导致错误匹配数量上升, 可能会影响对异常角度的正确剔除, 使得 RE 明显增大; 基于深度学习的特征点提取与描述模型可以更好地适应白内障手术中特征区域的变化, 对特征的描述能力更强; 本文模型在各视频序列上的整体性能最优, 具有较低的平均 RE, 其值为 $0.740^\circ \pm 0.208^\circ$, 表明该模型可以更为准确地测量眼球旋转角度.

表 1 在 CATARACT 测试集上 6 种模型的 RE 对比

(°)

模型	视频序列							平均
	1	2	3	4	5	6	7	
SIFT ^[10]	2.353±3.061	0.936±0.334	1.594±1.233	0.672±0.420	1.211±0.764	12.801±10.899	1.044±0.963	2.944±4.056
SURF ^[11]	3.038±3.118	1.457±0.746	1.534±1.856	0.922±0.634	0.995±0.442	9.035±9.585	1.166±1.459	2.592±2.712
SuperPoint ^[14]	1.896±2.526	0.943±0.230	0.754±0.338	0.505±0.212	1.125±0.511	1.936±1.727	0.610±0.215	1.110±0.544
D2-Net ^[15]	3.477±6.004	1.632±1.068	0.935±0.441	0.645±0.398	0.801±0.332	1.505±0.799	1.206±0.934	1.457±0.890
R2D2 ^[16]	1.650±2.319	0.724±0.227	0.859±0.811	0.890±0.566	0.566±0.293	0.789±0.352	0.499±0.220	0.854±0.352
本文	1.216±1.154	0.654±0.220	0.797±0.479	0.702±0.401	0.564±0.275	0.678±0.256	0.568±0.256	0.740±0.208

注. 粗体表示最优值.

6 种模型的特征点 MMA 与匹配对数对比如表 2 所示.

表 2 6 种模型的特征点 MMA 与匹配对数对比

模型	MMA				MVMP	MMP	异常匹配
	0.5°	1°	2°	5°			
SIFT ^[10]	0.166	0.311	0.502	0.662	93	476	383
SURF ^[11]	0.180	0.339	0.549	0.714	47	169	122
SuperPoint ^[14]	0.208	0.397	0.667	0.892	34	35	1
D2-Net ^[15]	0.165	0.320	0.566	0.867	82	82	0
R2D2 ^[16]	0.233	0.436	0.696	0.875	101	107	6
本文	0.236	0.443	0.718	0.902	85	90	5

注. 粗体表示最优值.

6 种模型的特征点匹配结果如表 3 所示, 其中, 蓝色连线表示总的匹配点对(不进行异常点对剔除), 绿色连线表示误差限为 2° 下被认为是正确匹配的特征点对. 可以看出, 在手术场景发生改变时, SIFT 和 SURF 模型的 MMA 较低, 异常匹配数量上升, 无法对眼球旋转角度进行可靠的计算, 导致 RE 较高; SuperPoint 模型由于对人造几何数据集交叉点的学习, 特征点的定位精度与 MMA 较高, 但 MMP 较低, 表明提取到的特征点较为稀疏,

导致旋转角度计算时可依赖的特征点对较少; D2-Net 模型的异常匹配最少, 但是网络缺乏对特征点稳定性的考量, 特征点检测与描述精度不高, MMA 较低, 且当眼球区域明显形变时无法有效地提取特征点; R2D2 模型增加了对特征点可重复性和描述符可靠性的学习, 使得 MMA 的提升较为明显; 本文模型增强对于细节特征和空间位置信息的提取能力, 减少了较为明显的跨区域匹配错误, 使得在各误差限下的特征点的 MMA 进一步提升, 且能在较少的异常匹配下执行较为密集的特征点匹配. 当目标区域特征受到手术操作与器械干扰导致区域特征发生明显形变时, 该模型仍具有较优的特征提取与描述能力, 匹配更为准确.

6 种模型的性能对比如表 4 所示, 其中, 检测速度包括眼球区域裁剪时间、模型特征点提取的推理时间、特征点筛选与匹配时间和旋转角度的计算时间. 可以看出, 本文模型的检测速度为 36.675 帧/s, 基本可以满足白内障术中实时检测的任务需求; 与 R2D2 模型相比, 本文模型参数量有所减少, 这是得益于在模型设计中采用了深度卷积操作. 但是, 由于引入的 CA 计算较为复杂, 同时自适应跳层连接降低了模型的并行度, 增加了模型运行的

时间成本,在一定程度上使得检测速度下降;而本文模型通过较小的速度代价获得了更为精细化地

提取与描述特征点的能力,对白内障术中眼球旋转角度的测量更为精准.

表 3 6 种模型的特征点匹配结果示例

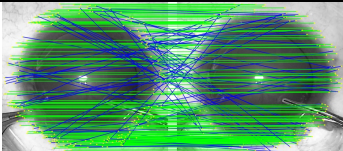
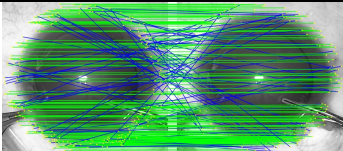
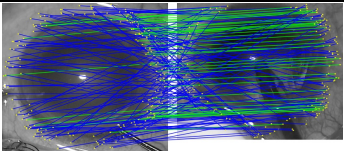
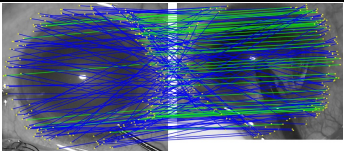
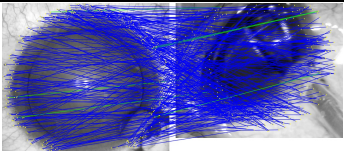
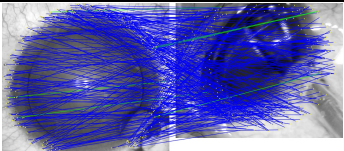
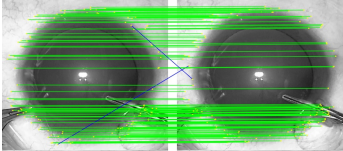
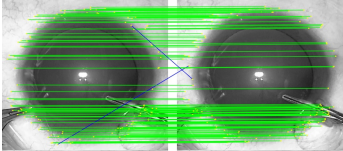
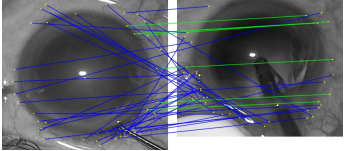
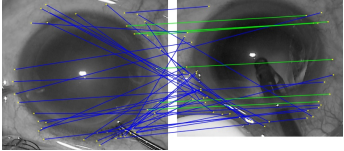
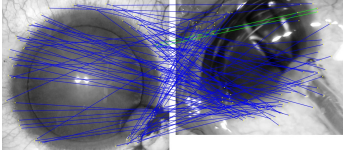
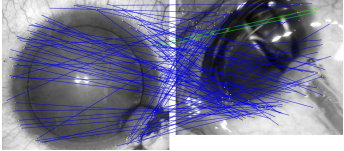
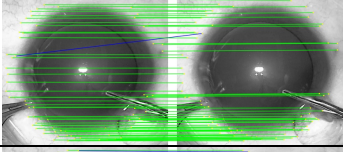
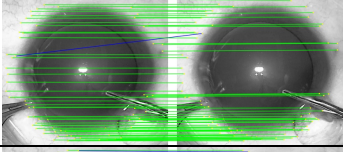
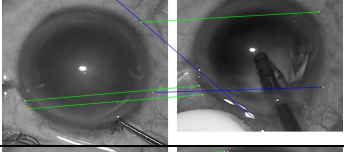
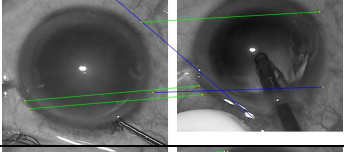
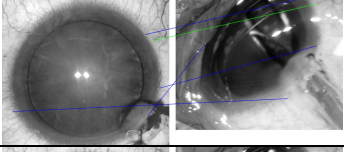
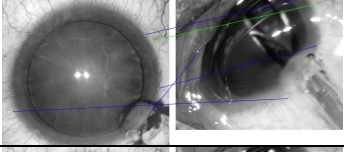
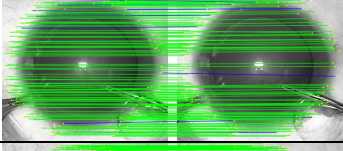
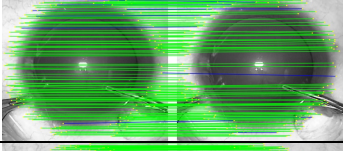
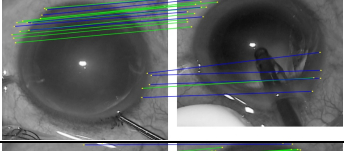
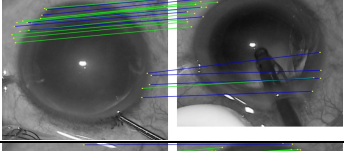
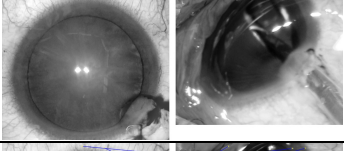
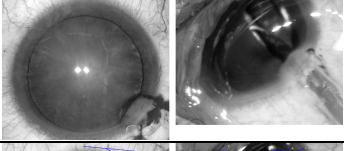
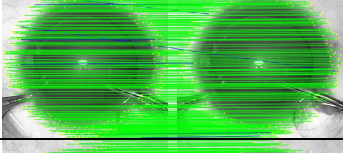
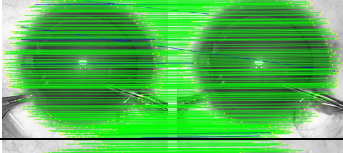
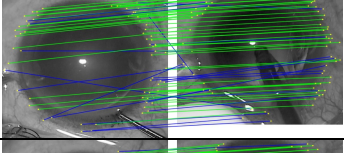
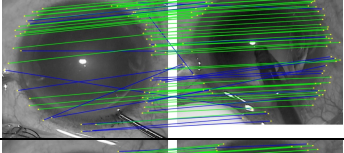
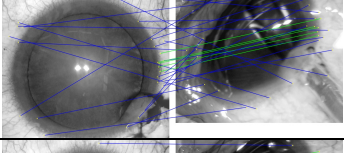
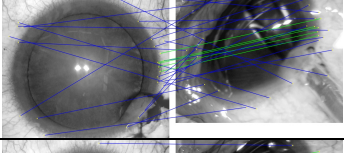
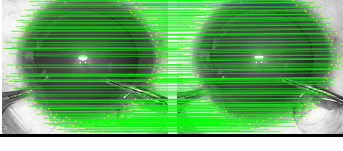
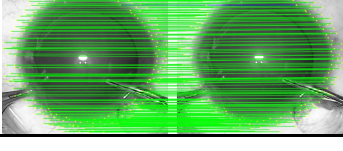
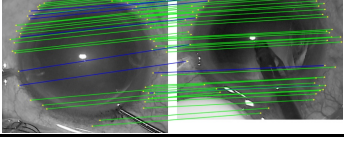
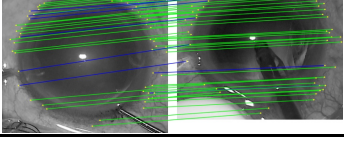
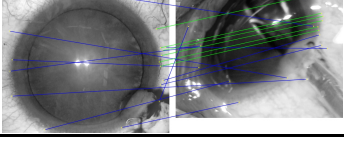
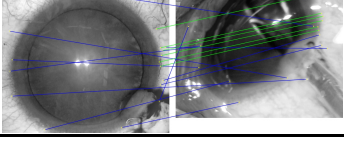
模型	简单场景(未受较大程度干扰)		普通场景(受器械干扰并轻微遮挡)		困难场景(受严重挤压并图像模糊)	
	基准图像	术中图像	基准图像	术中图像	基准图像	术中图像
SIFT ^[10]						
SURF ^[11]						
SuperPoint ^[14]						
D2-Net ^[15]						
R2D2 ^[16]						
本文						

表 4 6 种模型的性能对比

模型	检测速度/(帧·s ⁻¹)	参数量	描述符维度
SIFT ^[10]	12.179		128
SURF ^[11]	21.225		64
SupPoint ^[14]	71.786	1.304M	256
D2-Net ^[15]	19.225	7.635M	512
R2D2 ^[16]	42.344	0.484M	128
本文	36.675	0.456M	128

3.4 消融实验

将 R2D2 作为基准网络,对其分别引入 AttConvBlock 和自适应跳层连接,进一步提升网络对于特征的提取与描述能力.通过消融实验,验证本文方法中添加各模块的效果,结果如表 5 所示.可以看出,增加 AttConvBlock 和自适应跳层连接均会带来一定程度的提升,其中,自适应跳层

连接带来的性能增益较为明显.消融实验的特征点匹配结果如表 6 所示.可以看出,因为基准网络缺乏对细节和空间位置感知能力,容易导致较为明显的跨区域匹配错误;示例 1 中,基准网络错误地将毛细血管纹理与手术器械进行匹配;增加 AttConvBlock 后,网络对特征的空间位置较为敏感,减少了特征点跨区域匹配错误;增加自适应跳层连接后,网络对纹理细节更加敏感,可以较容易地依据不同的浅层特征信息区分毛细血管与手术器械;另外,当眼球区域受器械操作影响较小时(示例 2),角膜缘周围纹理受光照影响导致细节不明显时(示例 3)和眼球受到明显挤压形变导致特征模糊时(示例 4)中,基准网络均可能出现明显的跨区域匹配错误,而 AttConvBlock 和自适应跳层连接的引入,增强了模型对于不同位置、不同细节特征的描述与判别能力,在一定程度上减少了明显

表 5 各模块的消融实验结果

模型	RE/(°)	MMA				MVMP	MMP
		0.5°	1°	2°	5°		
基准网络	0.854±0.352	0.233	0.436	0.696	0.875	101	107
基准网络+AttConvBlock	0.814±0.316	0.232	0.438	0.711	0.893	74	77
基准网络+自适应跳层连接	0.798±0.212	0.241	0.449	0.715	0.880	91	95
本文	0.740±0.208	0.236	0.443	0.718	0.902	85	90

表 6 消融实验的特征点匹配结果示例

模型	示例 1		示例 2		示例 3		示例 4	
	基准图像	术中图像	基准图像	术中图像	基准图像	术中图像	基准图像	术中图像
基准网络								
基准网络+AttConvBlock								
基准网络+自适应跳层连接								
本文								

错误匹配的发生. 实验结果表明, 与基准网络模型相比, 本文模型对角膜缘周围特征点的提取与描述更准确, 对眼球旋转角度的检测更精准, RE 更低, 更适合于白内障手术中眼球旋转角度测量这一应用场景.

3.5 不同参数对眼球 RE 的影响实验

上述实验中, 将特征点最近邻匹配的 N 、目标匹配区域的 W_G 与异常角度的 T 分别设置为 0.85,

30, 20. 为了进一步探究参数设置对眼球旋转角度测量的影响, 通过控制变量法对各个参数进行实验, 结果如图 5 所示. 可以看出, 不同参数设置对眼球的 RE 均有一定程度的影响, D2-Net 模型易受 N 的影响, 而 SIFT 和 SURF 模型受 W_G 与 T 的影响较大; 在不同参数条件下, 与其他模型相比, 本文模型仍保持较优的测量精度, RE 更低, 且对参数变化的鲁棒性更强.

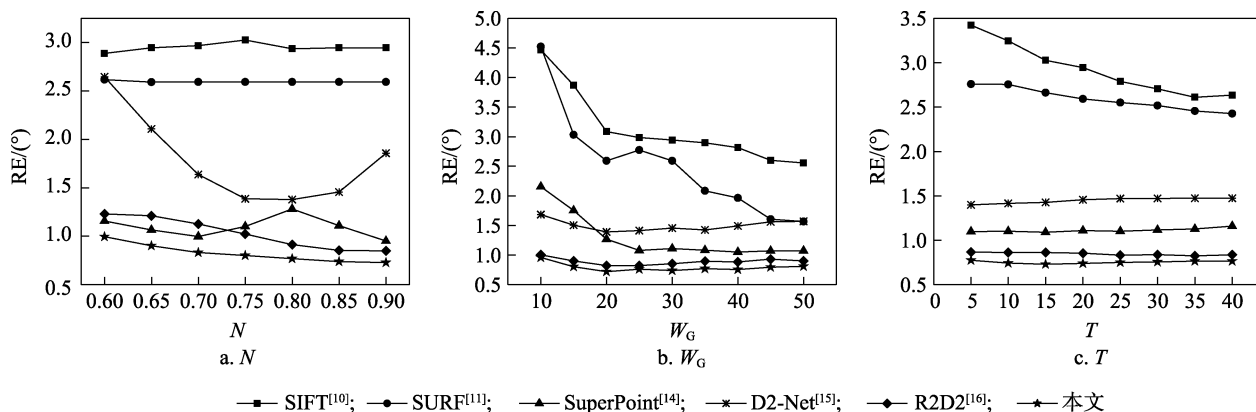


图 5 不同参数对眼球 RE 的影响结果对比

4 结 语

本文提出一种基于深度神经网络的特征点提取与描述模型,并用于白内障手术中眼球旋转角度检测.通过结合 AttConvBlock 与自适应跳层连接,增强了网络对纹理细节和空间位置的描述能力,促使网络生成更具判别性与鲁棒性的特征描述符.在 CATARACT 数据集上的实验结果表明,本文模型可以在一定程度上应对术中眼球区域形变与手术器械干扰对特征描述的影响,更好地对相似特征进行区分性描述,使得特征点匹配精度更高,旋转角度检测更精准.

但是,本文工作也存在一些不足:

(1) 本文模型虽然能更好地应对器械干扰与眼球受挤压情况下术中图像与基准图像的特征点匹配问题,但当眼球受到严重挤压时(如表 3 中困难场景),其匹配精度仍较低.未来将继续提升模型在严重形变、运动模糊、器械遮挡等困难场景下的特征提取与描述能力.

(2) 本文方法需要根据语义分割等前置任务获得的眼球几何参数对目标区域内的特征点进行筛选,从而计算旋转角度,无法端到端地由基准图像与术中图像直接测量眼球旋转角度.未来将对该方法进行改进,将语义分割模型与特征点提取与匹配模型融合到同一网络模型中,以实现端到端的眼球旋转角度测量.

参考文献(References):

- [1] Jiang Shumin, Han Qi. The choice and progress of cataract operative method and intraocular lens[J]. Medical Recapitulate, 2010, 16(9): 1388-1390(in Chinese)
(姜淑敏, 韩琪. 白内障手术方式及人工晶状体的选择与进展[J]. 医学综述, 2010, 16(9): 1388-1390)
- [2] Felipe A, Artigas J M, Díez-Ajenjo A, *et al.* Residual astigmatism produced by toric intraocular lens rotation[J]. Journal of Cataract & Refractive Surgery, 2011, 37(10): 1895-1901
- [3] Wan Jiayu, Zhao Chunmei, Chen Yanhui, *et al.* Status of axial position labeling methods for astigmatic correction type toric intraocular lens[J]. Recent Advances in Ophthalmology, 2019, 39(11): 1096-1100(in Chinese)
(万佳昱, 赵春梅, 陈艳辉, 等. 散光矫正型人工晶状体轴位标记方法研究现状[J]. 眼科新进展, 2019, 39(11): 1096-1100)
- [4] Levis I, Lahti K, Nalibotski B. Intraocular lens alignment: US, 20100208199[P]. 2010-08-19
- [5] Moeller G, Seiwert A, Perez M, *et al.* Surgical microscopy system and method for performing eye surgery: US, 20110157553A1[P]. 2011-06-30
- [6] Zhai Y X, Zhang G H, Zheng L S, *et al.* Computer-aided intraoperative toric intraocular lens positioning and alignment during cataract surgery[J]. IEEE Journal of Biomedical and Health Informatics, 2021, 25(10): 3921-3932
- [7] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C] //Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241
- [8] Chen L C, Zhu Y K, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 833-851
- [9] Yu C Q, Wang J B, Peng C, *et al.* BiSeNet: bilateral segmentation network for real-time semantic segmentation[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 334-349
- [10] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [11] Bay H, Tuytelaars T, van Gool L. SURF: speeded up robust features[C] //Proceedings of the 9th European Conference on Computer Vision. Heidelberg: Springer, 2006: 404-417
- [12] Rublee E, Rabaud V, Konolige K, *et al.* ORB: an efficient alternative to SIFT or SURF[C] //Proceedings of the International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2011: 2564-2571
- [13] Leutenegger S, Chli M, Siegwart R Y. BRISK: binary robust invariant scalable keypoints[C] //Proceedings of the International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2011: 2548-2555
- [14] DeTone D, Malisiewicz T, Rabinovich A. SuperPoint: self-supervised interest point detection and description[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2018: 337-33712
- [15] Dusmanu M, Rocco I, Pajdla T, *et al.* D2-Net: a trainable CNN for joint description and detection of local features[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 8084-8093
- [16] Revaud J, Weinzaepfel P, De Souza C, *et al.* R2D2: repeatable and reliable detector and descriptor[C] //Proceedings of the 33rd International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2019: Article No.1113
- [17] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 13708-13717
- [18] Zhai Y X, Zhang G H, Zheng L S, *et al.* Cataract surgery dataset for eye positioning and alignment[DB/OL]. [2022-10-08]. <https://dx.doi.org/10.21227/wykf-fc47>
- [19] Tian Y R, Fan B, Wu F C. L_2 -Net: deep learning of discriminative patch descriptor in euclidean space[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 6128-6136
- [20] Yang B, Bender G, Le Q V, *et al.* CondConv: conditionally pa-

- parameterized convolutions for efficient inference[C] //Proceedings of the 33rd International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2019: Article No. 117
- [21] Howard A G, Zhu M L, Chen B, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications[OL]. [2022-10-08]. <https://doi.org/10.48550/arXiv.1704.04861>
- [22] Hassanin M, Anwar S, Radwan I, *et al.* Visual attention methods in deep learning: an in-depth survey[OL]. [2022-10-08]. <https://doi.org/10.48550/arXiv.2204.07756>
- [23] Sun Meijun, Lyu Chaozhang, Han Yahong, *et al.* Weakly supervised surface defect detection based on attention mechanism[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(6): 920-928(in Chinese)
(孙美君, 吕超章, 韩亚洪, 等. 弱监督学习下的融合注意力机制的表面缺陷检测[J]. 计算机辅助设计与图形学学报, 2021, 33(6): 920-928)
- [24] Li Xuanye, Hao Xingwei, Jia Jingong, *et al.* Human action recognition method based on multi-attention mechanism and spatiotemporal graph convolution networks[J]. Journal of Computer-Aided Design & Computer Graphics, 2021, 33(7): 1055-1063(in Chinese)
(李炫烨, 郝兴伟, 贾金公, 等. 结合多注意力机制与时空图卷积网络的人体动作识别方法[J]. 计算机辅助设计与图形学学报, 2021, 33(7): 1055-1063)
- [25] He K, Lu Y, Sclaroff S. Local descriptors optimized for average precision[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 596-605
- [26] Sattler T, Maddern W, Toft C, *et al.* Benchmarking 6DOF outdoor visual localization in changing conditions[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 8601-8610
- [27] Radenovic F, Iscen A, Tolias G, *et al.* Revisiting oxford and Paris: large-scale image retrieval benchmarking[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 5706-5715