

基于图像生成的多模态视网膜图像配准方法

於佳乐, 黄坤, 张潇, 陈强

(南京理工大学计算机科学与工程学院, 江苏 南京 210094)

[摘要] 针对多模态视网膜的全局粗配准, 提出了一种基于图像生成的多模态视网膜配准方法. 不同于当前主流的提取多模态视网膜图像血管再进行配准的方法, 使用 GAN 模型对不同模态的视网膜图像进行像素级映射, 再通过特征点匹配的方式计算图像仿射矩阵, 实现图像粗配准. 基于彩色眼底图像与荧光素血管造影图像的实验结果表明, 该方法与当前主流的视网膜粗配准方法相比, 具有快速且鲁棒的优势.

[关键词] 图像配准, 多模态, 视网膜图像, 图像生成

[中图分类号] TP391 [文献标志码] A [文章编号] 1672-1292(2023)01-0010-08

Multi-Modal Retinal Image Registration Method Based on Image Generation

Yu Jiale, Huang Kun, Zhang Xiao, Chen Qiang

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: A multi-modal retinal registration method based on image generation is proposed for global coarse registration of multi-modal retinal images. Unlike the current mainstream methods that extract retinal vascular structures for registration, this method uses GAN model to perform pixel-level mapping of different modal retinal images. Then, the affine matrix is calculated through feature point matching to achieve image rough registration. Experimental results based on color fundus images and fluorescein angiography images demonstrate that this method has the advantages of faster speed and robust performance compared with current mainstream retinal registration methods.

Key words: image registration, multi-modal, retinal images, image generation

视网膜是人体中独一无二的可通过非侵入式仪器观察血管的部位, 这为血流系统和神经系统疾病的研究提供了一个成熟的环境. 通过对眼底特征的详细临床观察, 可确定各种长期疾病的早期指标, 如糖尿病、中风、高血压和心理疾病等. 视网膜图像可通过不同的成像方式获得, 如彩色眼底 (CF)、红外扫描 (IR)、荧光素血管造影 (FA)、光学相干断层扫描 (OCT) 等, 以便得到眼睛的多种表征^[1]. 在临床诊断过程中往往需要对多种模态的、不同时刻的视网膜图像进行横向与纵向的观察. 多模态视网膜配准可完成视网膜图像的对齐, 以便眼科医生更便利地确认不同时刻不同模态视网膜特征的表现.

由于不同模态图像之间不仅存在分辨率、对比度、亮度的差异, 且在外表上也存在明显不同, 多模态视网膜配准仍是一个挑战^[2]. 近年来提出的视网膜图像配准的主流方法可归纳为两类: 基于区域的方法和基于特征点匹配的方法. 基于区域的方法通过找到转换参数来最小化两幅图像之间的相关性或互信息^[3], 从而获得图像的全局对齐. 该方法计算成本高, 对图像失真、噪声等引起的外观变化、光照变化敏感, 多用在小的旋转和局部变形的情况下^[4]. 基于特征点匹配的方法通过显著特征检测和描述、特征匹配和几何模型估计实现. 该方法根据提取的特征描述符之间的对应关系寻找最佳的变换参数, 可以得到更好的几何变形处理能力, 常被用于图像配准和拼接任务中^[5-6].

特征点检测算法曾被广泛研究和应用于多模态视网膜配准. Harris-PIIFD 设计了针对多模态和低质

收稿日期: 2022-09-15.

基金项目: 国家自然科学基金项目 (62172223 和 61671242)、中央高校基本科研业务费专项资金项目 (30921013105).

通讯作者: 陈强, 教授, 博士生导师, 研究方向: 图像处理和分析. E-mail: chen2qiang@njnu.edu.cn

量的视网膜图像的配准框架,提出了一种高度分区的局部特征描述子 PIIFD^[7-8]. SURF-PIIFD-RPM 改进了 Harris-PIIFD 的框架,利用 SURF 关键点检测解决特征的重复性,并提出了一种鲁棒的特征点匹配方式^[9-10]. Zhang 等^[11]使用 URSIFT 检测关键点,利用 PIIFD 描述子和 AGMM 进行错配去除,得到了更好的配准结果.除了传统方法,深度方法也逐渐被应用于视网膜配准中. Wang 等^[12]提出了一种基于弱监督的深度神经网络进行全局粗配准,通过 GAN 网络风格迁移的方式训练血管分割网络,再通过 SuperPoint 网络^[10]提取特征点位置和描述,最后用离群点抑制网络完成错配去除,实现了深度模型在多模态视网膜配准中的应用.

利用血管分割的图像预处理方法常被用于同模态、跨模态的视网膜配准任务中^[13-14].但由于视网膜图像的血管分割标注困难,公开数据集往往只包含对 CF 图像的血管分割,血管分割任务有较大的成本与难度.当患者眼底因疾病出现出血等情况时,血管被遮挡会进一步提高血管分割的难度.

本文提出一种新的多模态视网膜配准框架,使用生成对抗网络(GAN)对不同模态的视网膜图像进行像素级映射,只需为每对图像人工标注 10 个左右的配准点对,计算粗配准仿射矩阵,即可进行训练.所提框架的基本流程是:先基于修改后的 Pix2Pix 网络^[12]进行跨模态转换,再通过 SuperPoint 提取特征点位置和描述,最后通过 SuperGlue^[15]进行特征点匹配和错配去除,使用得到的特征点对计算图像的仿射变换矩阵,从而得到两幅图像的粗配准结果.

1 相关工作

1.1 模态转换

大量研究证明,基于 CNN 的生成器可以学习 CF 图像和 FA 图像之间的像素级映射^[16-17].为了获得更真实的生成性能,大多数方法都采用图像级对抗损失来训练生成器,即使用一个额外的鉴别器来限制输出图像的样式和纹理,使之与真实目标图像相似.本文基于 Pix2Pix 网络^[18]和 Pix2PixHD 网络^[19]训练了一个视网膜图像模态转换网络,使得不同模态的视网膜图像可在传统方法及预训练的深度网络中提取到相似的特征描述.

1.2 特征点提取

计算机视觉中常用的特征检测和描述方法也可用于视网膜配准任务.传统方法包括尺度不变特征变换(SIFT)^[20]、加速鲁棒特征(SURF)^[21]及 Harris 角点检测^[8]等.针对多模态和低质量视网膜图像,文献[10]提出了一种适用于多模态的部分强度不变特征描述子 PIIFD. PIIFD 被广泛应用于多模态视网膜配准框架中,以提高受疾病影响的图像对的鲁棒性^[7,10-11].

基于深度模型的特征检测方法近年逐渐取代了原有的传统算法. Choy 等提出了一种基于全卷积网络的特征描述方法(UCN),用于检测逐像素的特征描述^[22]. Yi 等提出了利用深度模型的特征检测流程(LIFT),包括特征检测器、方向估计器和特征描述符^[23]. SuperPoint 网络^[10]克服了一个编码器和两个解码器的限制,在一次前向传递中联合获得关键点和描述符,性能优于 LIFT 和 UCN. SuperPoint 被广泛应用于视网膜图像的配准框架中,作为特征点检测和特征描述符^[12,24-25].

1.3 特征点配对与过滤

局部特征匹配通常由检测关键点、计算视觉描述符后进行关键点匹配、错配滤除、估计几何变换来完成.其中关键点匹配与错配滤除主要采用最近邻搜索和随机样本一致性(RANSAC^[26])检测的方式. SuperGlue^[27]受 Transformer^[16]的启发,使用自我(图像内)和交叉(图像间)注意的方法利用关键点的空间关系和视觉外观进行匹配,可在短时间内完成关键点匹配与错配滤除.

2 多模态视网膜图像粗配准方法

如图 1 所示,本文提出的多模态视网膜图像粗配准方法由 3 个网络组成,分别用于模态转换、特征检测和描述、特征点匹配和异常点去除.当待配准图像中同时存在 CF 图像和 FA 图像时,训练的模态转换网络 Pix2PixHD 会将 CF 图像转换成 FA 图像,再使用特征提取网络 SuperPoint 分别对固定图像和浮动图像提取特征点位置及其描述.根据所提取特征,使用 SuperGlue 网络实现特征点匹配及计算每对特征点配对的置信度.最后选取特定置信度以上的配对点计算浮动图像的仿射变换矩阵.

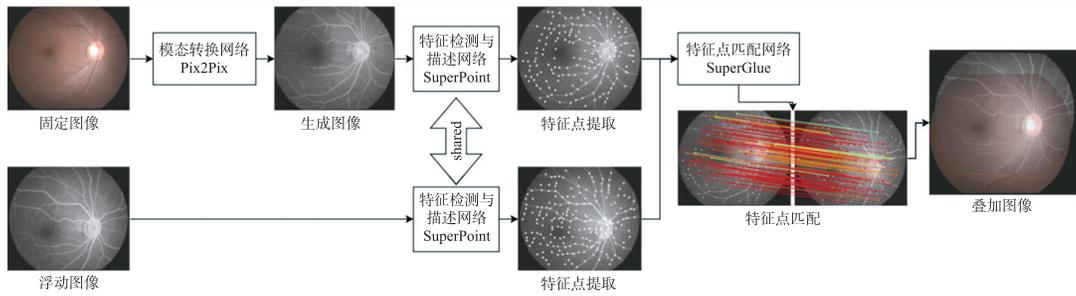


图 1 多模态视网膜图像粗配准框架

Fig. 1 Global coarse registration framework for multi-modal retinal images

2.1 模态转换网络

对抗生成网络 GAN 主要包括一个生成器 G 和一个判别器 D . 本文参考 Pix2Pix 网络和 Pix2PixHD 网络设计了从 CF 图像到 FA 图像的基于条件对抗生成的模态转换网络. 如图 2 所示, 令 CF 图像为 x , FA 图像为 y , 生成器 G 通过训练将 x 映射为 FA 图像 $G(x)$, 鉴别器 D 通过训练判断输入的 (x, y) 和 $(x, G(x))$ 是真实图像还是生成的图像. 条件对抗生成网络通过生成器与鉴别器的对抗博弈实现网络模型的训练:

$$\min_G \max_D \mathcal{L}_{GAN}(G, D). \quad (1)$$

本文保留了 Pix2Pix 网络中的设计, 将 Unet^[28] 作为生成器 G . 相较于 Pix2PixHD 网络中使用的由粗到细的生成器, Unet 可以更小的网络模型和更快的速度得到图像配准任务中所需的图像精度. 判别器 D 参考 Pix2PixHD 网络的多尺度判别器, 使用了 Pix2Pix 中的 PatchGAN, 并将原始尺寸和 2 倍降采样图像分别输入两个判别器 D_1, D_2 , 其对抗目标可表示为:

$$\min_G \max_{D_1, D_2} \sum_{k=1,2} \mathcal{L}_{GAN}(G, D_k). \quad (2)$$

通过在不同尺度下对图像进行判别, 使得生成图像在全局和细节上都更接近真实图像. 训练生成器 G 时除了使用对抗损失, 还在鉴别器的每层特征提取器上对真实和合成图像提取的张量计算 $L1$ 距离, 在层内平均、层间求和后得到 \mathcal{L}_{FM} 损失, 定义为:

$$\mathcal{L}_{FM}(G, D_k) = \mathbb{E}_{(x,y)} \sum_{i=1}^T \frac{1}{N_i} [\| D_k^{(i)}(x, y) - D_k^{(i)}(x, G(x)) \|_1], \quad (3)$$

式中, T 为鉴别器的总层数; N_i 为每层的元素个数. 类似的, 使用预训练的 VGG-net19^[29] 提取真实和合成图像的部分层张量计算 $L1$ 距离, 在层内平均、层间求和后得到 \mathcal{L}_{VGG} 损失, 定义为:

$$\mathcal{L}_{VGG}(G, VGG) = \mathbb{E}_{(x,y)} \sum_i \frac{1}{N_i} [\| VGG^{(i)}(x) - VGG^{(i)}(G(x)) \|_1], \quad (4)$$

式中, $S = \{2, 7, 12, 21, 30\}$. 从而得到最终的 GAN 损失函数:

$$\min_G \left(\max_{D_1, D_2} \sum_{k=1,2} \mathcal{L}_{GAN}(G, D_k) \right) + \sum_{k=1,2} \mathcal{L}_{FM}(G, D_k) + \mathcal{L}_{VGG}(G, VGG). \quad (5)$$

2.2 特征检测与描述网络

本文使用 SuperPoint 网络^[27] 对视网膜图像进行特征检测和描述. 该网络由一个编码器和两个解码器构成, 其结构如图 3 所示. 编码器部分由两个 3×3 的卷积核和 ReLU 激活函数及一个 2×2 的最大值池化堆叠构成. VGG 风格的编码器将输入图像 $I \in \mathbf{R}^{H \times W}$ 映射到一个空间维度更小、信道深度更大的中间张量 $\beta \in \mathbf{R}^{\frac{H}{8} \times \frac{W}{8} \times 256}$ 中, 以便完成后续的解码操作. SuperPoint 的解码器分为两个部分, 一个为输出特征点热力图的特

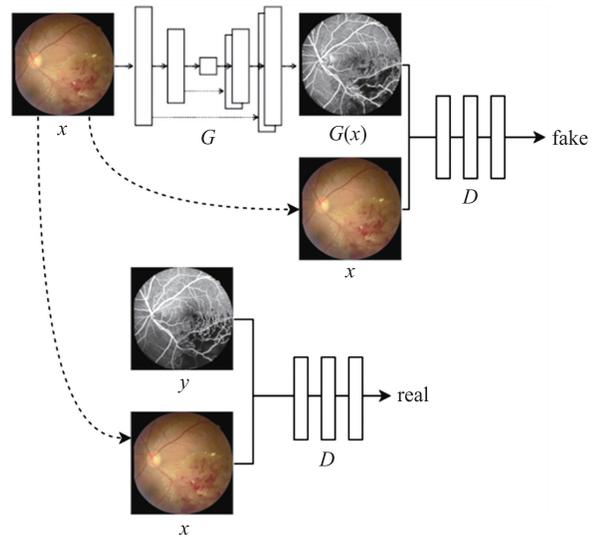


图 2 条件 GAN 网络实现 CF 向 FA 的映射

Fig. 2 A conditional GAN structure mapping CF to FA

征点解码器,另一个为输出特征点描述的描述符解码器. 特征点解码器将 $X \in \mathbf{R}^{\frac{H}{8} \times \frac{W}{8} \times 65}$ 中前 64 个通道数据对应图像不重叠的 8×8 个区域,第 65 个通道作为无兴趣点的垃圾箱维度,对 X 在通道上计算 Softmax 后即可得每个像素的被认定为特征点的置信度,最后将前 64 个通道 $\mathbf{R}^{\frac{H}{8} \times \frac{W}{8} \times 64}$ 重塑到对应位置得到 $\mathbf{R}^{H \times W \times 1}$ 的热力图. 该方法不需进行上采样,可节省大量时间,且避免了上采样可能带来的不必要的棋盘伪影. 描述符解码器对 $D \in \mathbf{R}^{\frac{H}{8} \times \frac{W}{8} \times 256}$ 进行双三次插值,再进行 $L2$ 归一化得到逐像素的描述符.

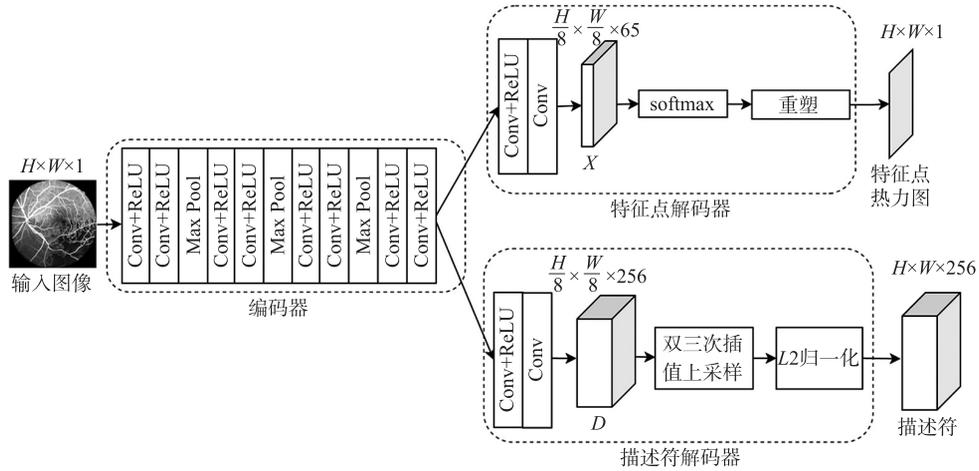
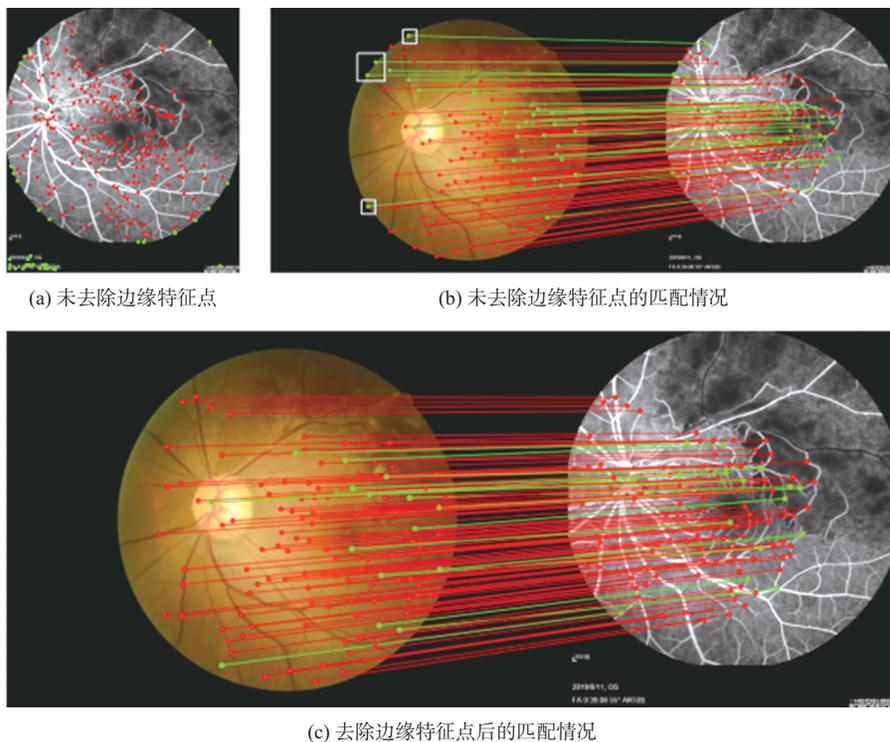


图 3 SuperPoint 网络结构

Fig. 3 SuperPoint network architecture

本文直接使用预训练的 SuperPoint 网络模型对视网膜图像进行特征点提取与描述. 该模型先在大规模的合成数据集以及真实图像数据集上训练得到. 其中,合成数据集包含了各种带有明显边缘特征和角特征的模拟图像,因而 SuperPoint 可很好地提取血管拐角处的特征点^[25]. 本文在关键点概率图上进行 4 像素的非极大值抑制. 视网膜图像往往不会充满整幅图像,故将像素值小于 3 的区域作为掩码区域过滤其中的特征点,同时用 3×3 的矩阵对掩码区域做膨胀操作,有效去除边缘及部分图像四周字符上的关键点. 图 4(a)展示了 SuperPoint 提取的特征点,其中包含了图像边缘及左下角字符上的特征点. 使用掩码方



(c) 去除边缘特征点后的匹配情况

图 4 掩码前后对比

Fig. 4 Comparison whether to use mask

式去除(a)中绿色的关键点,从而减少其对图像配准的影响.图 4(b)、(c)展示了滤除前后的特征点匹配情况,其中绿色线条表示错误或误差较大的配对点,白色框选标记了边缘特征点导致的错误匹配.

2.3 特征点匹配网络

SuperGlue^[15]是一种特征匹配网络,同时具备特征匹配与外点滤除的能力. SuperPoint 和 SuperGlue 的框架常被应用于各类特征点匹配问题. 本文使用预训练的 SuperGlue 模型代替传统的 RANSAC 算法进行特征匹配.

SuperGlue 主要由注意力神经网络和最优匹配层组成. 注意力神经网络对关键点位置和描述符编码,并使用交替的自我注意力和交叉注意力层增强其匹配能力,再通过 sinkhorn 算法迭代匹配度得分矩阵算出最优的特征点匹配. 本文保留了 SuperGlue 的原始参数,根据经验滤除置信度小于最高置信度 40% 的匹配点对,从而得到较为全面且准确的匹配点集.

3 实验

3.1 数据集

本文在两个数据集上对所提方法进行了验证:数据集 1 为江苏省人民医院所提供的 133 只视网膜分支静脉栓塞(BRVO)患者眼睛的眼底照相,每只眼睛包含一幅分辨率为 2 976×1 984 的 CF 图像和一幅分辨率为 768×770 的 FA 图像. 数据集 2 为来自文献[30]的 59 只眼睛的眼底照相,其中 29 只来自正常人,30 只来自糖尿病视网膜病变(DR)患者,每只眼睛包含一幅分辨率为 720×576 的 CF 图像和一幅分辨率为 720×576 的 FA 图像.

本试验为每对 CF-FA 图像人工标注了 10 个以上的匹配点对,通过匹配点对计算仿射矩阵作为数据集的真实标签 M_{gt} .

3.2 训练模态转换网络

通过仿射矩阵 M_{gt} 将图像进行扭曲得到配准后的图像,保留图像重叠部分,并通过缩放和填充的方式将图像处理成 512×512 像素. 为了增加数据量以提高模型的泛化能力,对输入图像随机旋转、翻转和裁剪进行数据增广,将 CF-FA 图像对裁剪为 256×256 像素对模态转换网络进行训练,超参数设置如表 1 所示.

表 1 模态转换网络超参数
Table 1 Hyperparameters of modal conversion networks

超参数名称	参数值
Batch_size	1
Iterations	100 000
Lr of Generator	0.000 1
Lr of Discriminator	0.000 4

为了得到完整测试数据集,对 BVRO 数据集进行三折交叉实验,即将 133 对图像随机分为 3 份,进行 3 次网络训练,每次分别取一份作为测试集,另两份作为训练集,最后合并 3 次训练得到的测试集输出作为下一步特征点提取的输入. 同理,对糖尿病视网膜病变数据集进行二折交叉实验,分别将编号为奇数或偶数的图像作为训练集,另一半作为测试集.

图 5 展示了模态转换网络的结果. 图 5(a)和(b)分别为输入的 CF 图像和 FA 图像,(c)是模型生成的 FA 图像. 可以看出,生成的图像很好地保留了 CF 图像中不明显的、细小的血管,并将其映射到了对比度更强的 FA 图像上,为下一步特征点提取与匹配提供了良好的助力.

3.3 评价指标

以浮动图像上人工标注的配准点为锚点,计算其经过标签仿射矩阵 M_{gt} 和估计的仿射矩阵 M_{pred} 后得到的位置,通过对比计算最大像素距离来验证配准性能. 最大绝对值误差 MAE 定义为:

$$MAE = \max_{p \in \text{Points}} \| T(p, M_{gt}) - T(p, M_{pred}) \|_1, \quad (6)$$

式中,Points 表示浮动图像上的点集; p 表示点集中的一个点; T 运算表示使用仿射矩阵 M 计算点映射的像素位置. 将图像缩放和填充到 768×768 后计算 MAE,小于 10 即配准成功,该方法被普遍作为全局粗配准是否配准成功的依据^[12,31].

3.4 结果分析

测试平台配置:CPU 为 12th Gen Intel(R) Core(TM) i5-12600K 3.69 GHz;GPU 为 GeForce GTX 1060 3GB;系统为 Windows 11;运行环境为 Matlab 2021b、Python 3.9;深度学习框架为 PyTorch.

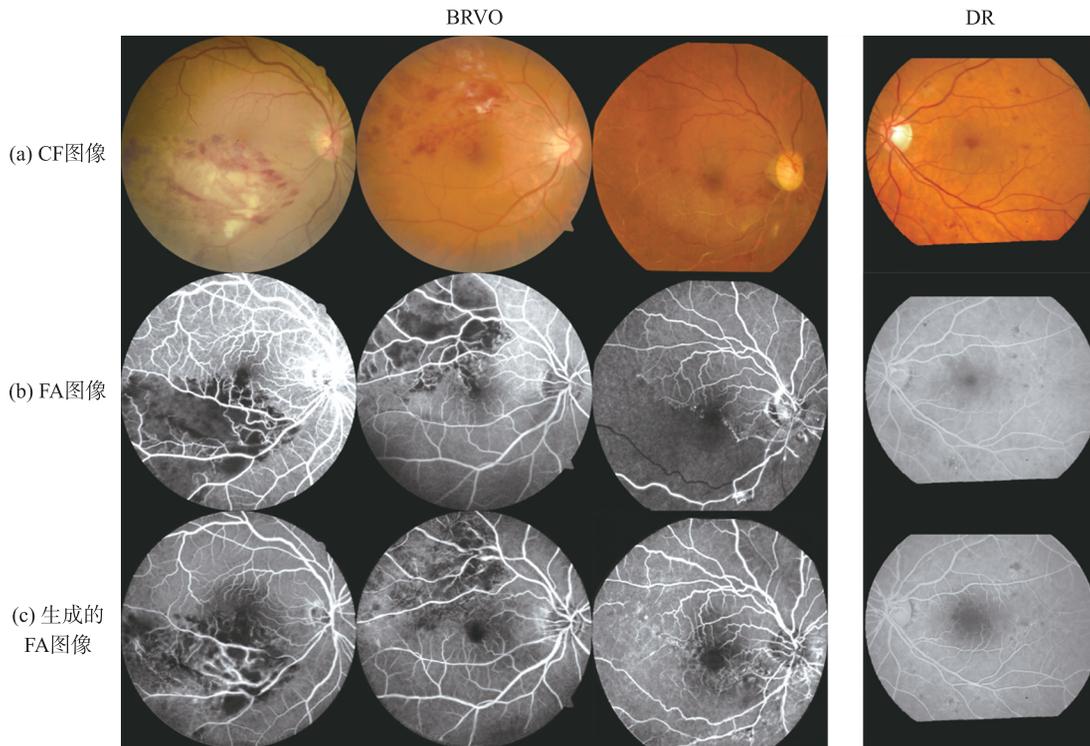


图5 模态转换结果

Fig. 5 Results of modal conversion

表2所示为模态转换网络的用时. 表3定量地展示了本文方法与其他方法^[10-11]的对比, 其中, SURF-PIIFD-RPM 和 URSIFT-PIIFD-AGMM 是专为跨模态视网膜配准设计的传统方法, SuperPoint-SuperGlue 为本文使用的配准方法. 表3对比了直接使用原始图像和将本文模态转换作为图像预处理后执行这3种方法的配准结果. 可以看出, 模态转换有效提高了各类方法的配准正确率, 本文框架在配准成功率和配准用时上均超过了传统的配准方法.

表2 模态转换网络用时

Table 2 Time spent in modal conversion

数据集	CPU 用时/(s/幅)	GPU 用时/(s/幅)
BRVO	1.25	0.31
DR	1.39	0.38

表3 不同方法的结果对比

Table 3 Results of different methods

数据集	方法	MAE<5	MAE<10	CPU(GPU) 运行时间/对
BRVO(133对)	SURF-PIIFD-RPM ^[10]	48.1%(64)	70.7%(94)	3.58 s(-)
	SURF-PIIFD-RPM(with modal conversion)	85.0%(113)	91.7%(122)	5.58 s(4.65 s)
	URSIFT-PIIFD-AGMM ^[11]	3.0%(4)	11.3%(15)	24.15 s(-)
	URSIFT-PIIFD-AGMM(with modal conversion)	23.3%(31)	51.9%(69)	19.31 s(18.86 s)
	SuperPoint-SuperGlue ^[15,27]	0%(0)	0.8%(1)	0.43 s(0.12 s)
	SuperPoint-SuperGlue(with modal conversion)(ours)	85.0%(113)	94.0%(125)	1.71 s(0.41 s)
DR(59对)	SURF-PIIFD-RPM	88.1%(52)	93.2%(55)	4.05 s(-)
	SURF-PIIFD-RPM(with modal conversion)	95.0%(56)	96.6%(57)	6.71 s(5.70 s)
	URSIFT-PIIFD-AGMM	33.6%(21)	54.2%(32)	26.14 s(-)
	URSIFT-PIIFD-AGMM(with modal conversion)	39.0%(23)	64.4%(38)	27.34 s(25.22 s)
	SuperPoint-SuperGlue	40.7%(24)	55.9%(33)	0.41 s(0.11 s)
	SuperPoint-SuperGlue(with modal conversion)(ours)	88.1%(52)	98.3%(58)	1.80 s(0.45 s)

如图6所示, 多模态转换网络有效提高了各类多模态配准方法的结果. 图6(a)表示使用原始图像配准, (b)表示对CF图像进行模态转换后进行配准. (x)所示为最终得到的匹配点对, 红色线段为像素绝对值误差小于10的点对, 即正确点对, 绿色反之. (y)所示为配准图像的棋盘展示. 对比(a)和(b)可知, 模态转换预处理方法有效提高了3种配准方法的匹配点对数量及配准结果. 对比3种方法的(b)图可知, URSIFT-PIIFD-AGMM的配准结果最好, 计算得到的匹配点都在10像素距离误差以内; 本文的 SuperPoint-

SuperGlue 方法部分匹配点误差较大,但对粗配准的结果影响很小.

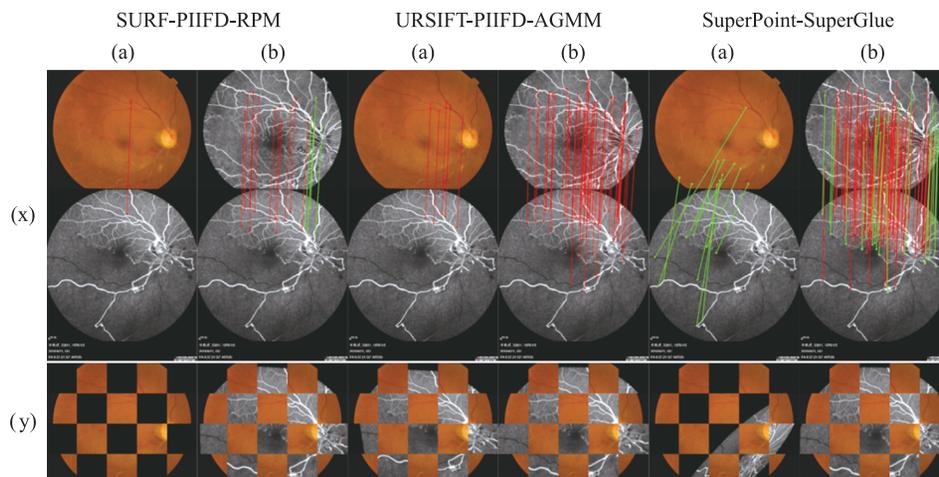


图 6 配准结果

Fig. 6 The registration results

4 结论

本文提出了一种基于视网膜模态转换的多模态视网膜配准框架,训练了一个视网膜模态转换网络作为图像预处理,进而使用基于特征点匹配的方法进行图像配准.该方法可通过容易得到的粗配准标签进行训练,并能有效提高多种配准方法的正确率.本文使用 SuperPoint-SuperGlue 匹配特征点,在速度与正确率上都得到了更优的结果.

[参考文献] (References)

- [1] MACGILLIVRAY T J, TRUCCO E, CAMERON J R, et al. Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions[J]. *The British Journal of Radiology*, 2014, 87(1040): 20130832.
- [2] ZHANG L, FAN H, ZHANG J, et al. Multi-modal and multi-vendor retina image registration[J]. *Biomedical Optics Express*, 2018, 9(2): 410-422.
- [3] WANG Y Q, ZHANG J K, AN C H, et al. A segmentation based robust deep learning framework for multimodal retinal image registration[C]//*Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE, 2020.
- [4] MA J Y, JIANG X Y, FAN A X, et al. Image matching from handcrafted to deep features: a survey[J]. *International Journal of Computer Vision*, 2021, 129(1): 23-79.
- [5] 汪前进, 朱斌, 李存华. 基于特征点的图像拼接方法的研究与应用[J]. *南京师范大学学报(工程技术版)*, 2016, 16(3): 48-53.
- [6] 李鹏程, 曾毓敏, 张梦. 基于改进 Harris 的图像拼接算法[J]. *南京师范大学学报(工程技术版)*, 2014, 14(1): 70-75.
- [7] CHEN J, TIAN J, LEE N, et al. A partial intensity invariant feature descriptor for multimodal retinal image registration[J]. *IEEE Transactions on Biomedical Engineering*, 2010, 57(7): 1707-1718.
- [8] HARRIS C, STEPHENS M. A combined corner and edge detector[C]//TAYLOR C J. *Proceedings of the Alvey Vision Conference*. Manchester, UK: Alvey Vision Club, 1988.
- [9] BAY H, TUYTELAARS T, GOOL L V. Surf: speeded up robust features[C]//*Proceedings of the 9th European Conference on Computer Vision*. Graz, Austria: Springer, 2006.
- [10] WANG G, WANG Z C, CHEN Y F, et al. Robust point matching method for multimodal retinal image registration[J]. *Biomedical Signal Processing and Control*, 2015, 19: 68-76.
- [11] ZHANG H T, LIU X H, WANG G, et al. An automated point set registration framework for multimodal retinal image[C]//*Proceedings of the 24th International Conference on Pattern Recognition (ICPR)*. Beijing, China: IEEE, 2018.

- [12] WANG Y Q, ZHANG J K, CAVICHINI M, et al. Robust content-adaptive global registration for multimodal retinal images using weakly supervised deep-learning framework[J]. *IEEE Transactions on Image Processing*, 2021, 30:3167–3178.
- [13] SHE Y Y, ZHOU M, LI Q L, et al. Retinal image registration based on features of vessel-segmented image[C]//*Proceedings of the 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. Shanghai, China: IEEE, 2021.
- [14] AN C H, WANG Y Q, ZHANG J K, et al. Self-supervised rigid registration for multimodal retinal images[J]. *IEEE Transactions on Image Processing*, 2022, 31:5733–5747.
- [15] SARLIN P E, DETONE D, MALISIEWICZ T, et al. Superglue: Learning feature matching with graph neural networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE, 2020.
- [16] SCHIFFERS F, YU Z K, ARGUIN S, et al. Synthetic fundus fluorescein angiography using deep neural networks[C]//*Proceedings of the Bildverarbeitung für die Medizin 2018*. Erlangen, Germany: Springer, 2018.
- [17] LI K, YU L Q, WANG S J, et al. Unsupervised retina image synthesis via disentangled representation learning[C]//*Proceedings of the 4th International Workshop on Simulation and Synthesis in Medical Imaging*. Shenzhen, China: Springer, 2019.
- [18] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, USA: IEEE, 2017.
- [19] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE, 2018.
- [20] LOWE D G. Object recognition from local scale-invariant features[C]//*Proceedings of the Seventh IEEE International Conference on Computer Vision*. Kerkyra, Greece: IEEE, 1999.
- [21] BAY H, TUYTELAARS T, GOOL L V. Surf: speeded up robust features[C]//*Proceedings of the 9th European Conference on Computer Vision*. Graz, Austria: Springer, 2006.
- [22] CHOY C B, GWAK J Y, SAVARESE S, et al. Universal correspondence network[C]//*Proceedings of the 30th International Conference on Neural Information Processing Systems*. Barcelona, Spain: NIPS, 2016.
- [23] YI K M, TRULLS E, LEPETIT V, et al. Lift: Learned invariant feature transform[J]. *arXiv Preprint arXiv:1603.09114*, 2016.
- [24] 吴玲玉, 蓝洋, 夏海英. 基于卷积神经网络的眼底图像配准研究[J]. *广西师范大学学报(自然科学版)*, 2021, 39(5): 122–133.
- [25] ZHANG J K, WANG Y Q, DAI J, et al. Two-step registration on multi-modal retinal images via deep neural networks[J]. *IEEE Transactions on Image Processing*, 2021, 31:823–838.
- [26] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 24(6): 381–395.
- [27] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: self-supervised interest point detection and description[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Salt Lake City, USA: IEEE, 2018.
- [28] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[J]. *arXiv Preprint arXiv:1505.04597*, 2015.
- [29] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv Preprint arXiv:1409.1556*, 2014.
- [30] ALIPOUR S H M, RABBANI H, AKHLAGHI M R. Diabetic retinopathy grading by digital curvelet transform[J]. *Computational and Mathematical Methods in Medicine*, 2012(4): 761901.
- [31] DAN T T, HU Y, HAN C, et al. Fusion of multi-source retinal fundus images via automatic registration for clinical diagnosis[J]. *Neurocomputing*, 2021, 459:370–382.

[责任编辑:严海琳]