

基于模块化降噪自编码器的视网膜 OCT 图像降噪方法

代豪^{1,2,3}, 杨亚良^{1,2*}, 岳献^{1,2,3}, 陈燊^{1,2,3}

¹中国科学院光电技术研究所, 四川 成都 610209;

²中国科学院自适应光学重点实验室, 四川 成都 610209;

³中国科学院大学, 北京 100049

摘要 针对光学相干层析成像(OCT)过程中,光线散射、目标微动和硬件抖动等原因引起的噪声干扰,尤其是视网膜 OCT 图像中存在的严重噪声干扰问题,提出了一种基于模块化降噪自编码器的渐进式 OCT 图像降噪方法。使用多层卷积和反卷积构建自编码器,以模块化深度神经网络的架构为基础搭建了具有多个自编码器模块的神经网络,每个自编码器模块可依次输出降噪程度逐渐升高的过程结果,以满足不同的使用需求。以均方误差、峰值信噪比和结构相似度作为降噪结果的评价指标,对编码器模块数量 T 的研究结果表明,所设计的编码器在 $T=4$ 时具有最佳性能。利用所提方法和各种主流方法对正常眼和病眼的视网膜 OCT 图像进行降噪处理,结果表明所提方法在各项指标上均取得最优结果,可以有效地对视网膜 OCT 图像进行降噪处理和大幅提升图像的质量。

关键词 图像处理; 视网膜光学相干层析图像; 图像降噪; 降噪自编码器; 深度学习

中图分类号 TP391.4 **文献标志码** A

DOI: 10.3788/AOS220815

1 引言

光学相干层析成像(OCT)是一种非侵入和无损伤的内部“光学切片”观察技术,具有高分辨率、高灵敏度和快速等优点,已被广泛应用于生物、医学、工业亚表面检测和文物鉴定等领域中^[1-2]。OCT 技术在眼科中应用最为成功和广泛,其已成为眼科临床上必不可少的日常检查工具和疾病诊断金标准。OCT 具有高纵向分辨能力,该能力使其成为眼底视网膜在体成像的强大技术手段,这是因为视网膜的纵向结构异常复杂,在不到 0.5 mm 的厚度范围内包含着多层组织结构 and 分界膜,对它们的分层观察和定量测量有助于找到病变的原因。

受到组织的高散射性和成像过程中目标的微动或硬件的抖动等因素的影响,OCT 图像中总是携带着以散斑噪声为主的噪声信号^[3],它会淹没有用的目标信号、使细节信息丢失和降低图像的对比度^[4],从而使得后续处理环节精度下降。对于视网膜 OCT 成像而言,由于视网膜是半透明组织,由其返回至探测器的光信号非常微弱,再加上眼组织像差和眼球微颤引起的目标抖动等因素的干扰,故噪声干扰非常严重,这给后续的视网膜亚层分割和厚度测量等工作带来了困难^[5-6]。

因此,自视网膜 OCT 成像技术出现以来,如何对图像进行降噪处理以提高图像的质量一直受到研究人员和产业界的高度关注。

已有多多种 OCT 图像降噪方法被提出,其中提出最早和应用最为广泛的是基于多帧 B-Scan 图像(纵截面内的二维图像)平均的方法^[7-8],该方法通过对同一目标区域进行多次扫描(横向一维往返扫描)来获取多幅 OCT 图像,再将它们配准并平均成一幅图像来降低噪声,但该方法存在着信号采集时间较长、图像错位(由信号采集过程中发生的微位移引起)和图像易模糊等缺陷^[9]。一些经典的图像处理算法也被用于 OCT 图像降噪,包括基于滤波^[10-12]、基于块匹配^[13-14]和基于稀疏表示的方法^[15-16]等。基于滤波的方法主要包括均值滤波^[10]、中值滤波^[11]和高斯滤波^[12]等,基于块匹配的方法包括非局部均值滤波^[13]和三维块匹配(BM3D)算法^[14]等。这些方法虽然取得了一定的降噪效果,但是会过度平滑图像,从而导致边缘对比度降低和细节信息丢失^[17]。此外,基于块匹配的方法通常包含多个步骤,当图像数据量较大时,存在着耗时长和任务复杂度高问题。Fang 等^[15]提出了一种针对谱域 OCT(SD-OCT)图像的字典学习降噪算法,这是一种基于稀疏表示的方法。Zhang 等^[16]在分析 SD-OCT 图像噪声的

收稿日期: 2022-03-17; 修回日期: 2022-04-27; 录用日期: 2022-06-20; 网络首发日期: 2022-06-30

基金项目: 国家自然科学基金(61575205)、中国科学院光电技术研究所前沿部署项目

通信作者: *ylyang@ioe.ac.cn

基础上,提出了基于噪声估计和全局字典的稀疏表示的降噪算法。这类方法需要根据先验信息来估计噪声水平和手动调节表征稀疏程度的超参数,若噪声水平估计不准确,则降噪效果会变得不理想。

近年来的大量研究成果和临床实践经验均表明,深度学习在眼科图像处理领域的应用日益广泛且效果显著。在 OCT 领域中,深度学习主要被用于图像的分割^[18-19]、分类^[20-21]和降噪^[22-24]等,而其在图像降噪方面的应用还只处于起始阶段。用于 OCT 图像降噪的深度学习网络目前主要有降噪自编码器(DAE)和生成对抗网络(GAN)。Vincent 等^[22]提出了将多个 DAE 堆叠而成的深度神经网络,取得了高质量的降噪结果。Ma 等^[23]提出了一种边缘敏感条件 GAN,用于商用 OCT 成像仪的图像降噪。Halupka 等^[24]提出了一种具有感知相似性的 GAN,以增强商用仪器 OCT 图像的质量。这些基于深度学习的方法通过最小化定义的损失函数来实现降噪前后的端到端映射,这意味着降噪的程度是不可控的。然而,在实际环境中,噪声的大小有可能发生变化,其程度不可控将会导致模型的泛化能力降低。对于医生而言,患者的病情不同,所要求的降噪程度也会不同,基于端到端的深度学习网络限制了对降噪程度的自主把控。若按同一标准降噪,难免会出现部分图像降噪不彻底而另一部分图像过度降噪的问题,前者会导致图像中仍含有残余噪声,而后者会导致图像模糊不清,故均应避免。因此,Shan 等^[25]提出了一种通过克隆多个自适应神经网络模块来实现端到过程的图像降噪方法,并且在计算成像领域中取得了显著的降噪效果。

为了降低视网膜 OCT 图像的噪声,并针对基于端到端的深度神经网络的不足,本文将模块化的方法应用到视网膜 OCT 图像领域中,提出了一种基于端到过程的深度学习网络来实现视网膜 OCT 图像的降噪。采用模块化自编码器结构对视网膜 OCT 图像进行循环降噪,将每个自编码器网络输出的降噪图像作为下

一个自编码器网络的输入,直到输出的降噪图像达到预期要求,从而实现端到过程的降噪,使得医生可根据任务需求来对降噪过程进行自主控制。本文使用杜克大学公开的视网膜 OCT 数据集来训练搭建的深度神经网络,并基于测试集对所提模型和其他主要方法进行比较,以验证所提方法的可行性和优越性。

2 方 法

2.1 降噪自编码器原理

DAE 是一种将损坏或有噪声的数据点作为输入,经过训练后能够提取出输入数据的高维深层特征,从而输出未损坏或干净的原始数据点的神经网络^[19]。DAE 的原理如图 1 所示,包含编码器和解码器两个部分,首先将 \tilde{x} 映射到隐藏层中编码成 h ,这一过程可表示为

$$h = f(\tilde{x}) = \sigma(w\tilde{x} + b), \quad (1)$$

式中: w 是编码权重; b 是偏置量; σ 是激活函数,如 Sigmoid、ReLU 和 ELU 等; h 是 \tilde{x} 在高维潜在空间中的表达; $f(\cdot)$ 为编码函数。编码之后又将 h 解码成 y ,这一过程可表示为

$$y = g(h) = \sigma(w'h + b'), \quad (2)$$

式中: w' 是解码权重; b' 是偏置量; y 是 \tilde{x} 经过自编码器重构后的样本; $g(\cdot)$ 为解码函数。因此,可以得到重构误差为

$$L(y, \tilde{x}) = \|y - \tilde{x}\|_2^2, \quad (3)$$

对训练集的每个样本的误差求和,得到关于 w 和 b 的误差函数

$$J(w, b) = \sum_{i=1}^n L(y_i, \tilde{x}_i). \quad (4)$$

因此,对 DAE 的训练目的就是找到最优的参数 w 和 b ,以使生成的降噪图像和目标图像之间的重构误差最小,从而迫使 DAE 能够提取输入图像的高维深层特征。然后,将噪声点根据数据特征映射到估计的高概率点上,从而达到图像降噪的目的^[26]。

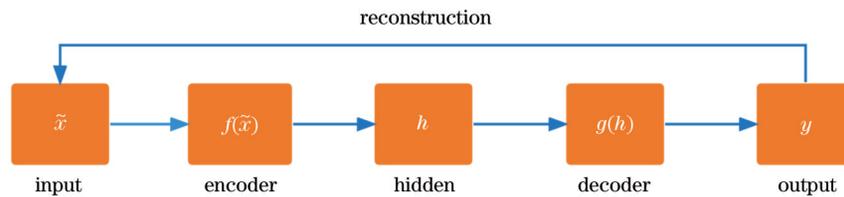


图 1 DAE 原理示意图

Fig. 1 Schematic diagram of DAE principle

2.2 模块化降噪自编码器结构

所提模块化降噪自编码器(MDAE)由多个相同的自编码器组成,模块化自编码器的结构原理可描述为

$$I_{\text{out}} = m^{(T)}(I_{\text{in}}) = (m \circ m \circ \dots \circ m)(I_{\text{in}}) \approx K, \quad (5)$$

式中: I_{in} 为输入的带有噪声的 OCT 图像; I_{out} 为网络输出的 OCT 图像; K 为理想的无噪声的 OCT 图像; m 代表一个自编码器模块; \circ 为操作符,用于将多个函数进行组合处理; T 为自编码器模块的数量,其取值由具体任务决定。

不同于直接输出单幅降噪图像的模型,MDAE是端到过程的模型,可以输出多幅降噪图像,实现渐进式交互降噪操作,其网络结构如图2所示。每一个自编码器模块均包含编码器和解码器两个部分:编码器由4个卷积层构成,用于提取图像特征,每个卷积层包含

32个大小为 3×3 的卷积核,使用ReLU作为激活函数;解码器由反卷积层构成,用于解码潜在表达,反卷积层包含32个大小为 3×3 的反卷积核。为了减少参数量和um提高训练速度,所有模块共享参数,即所有模块的参数相同。

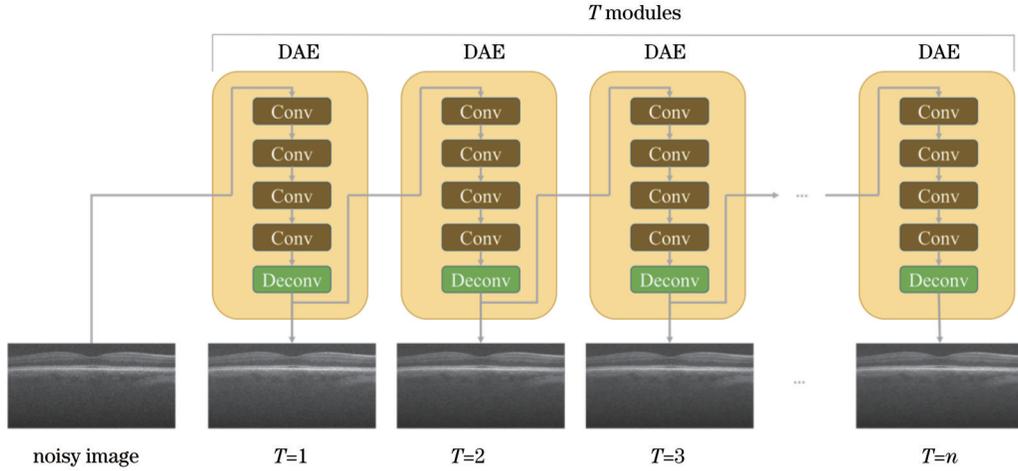


图2 MDAE的网络结构

Fig. 2 Network structure of MDAE

参数 T 的选择可由以下三个因素中最早满足者决定:1)医生通过肉眼直观感受当前降噪图像是否达到临床使用要求,如达到则停止降噪,并将当前输出的降噪图像作为最终结果;2)设定评价指标阈值,若当前模块输出图像的评价指标相较于上一模块的变化小于该阈值,说明继续降噪已无明显收益,则停止降噪;3)若有非常高的降噪要求,则将 T 设置为模型本身或计算能力所允许的最大值,达到该值后停止工作。因此,该工作模式可根据任务需求来灵活调节参数 T ,实现过程可控,其中因素1)有利于人工参与,而因素2)和因素3)可保证无人值守时MDAE的自动运行。

2.3 数据集

自编码器是一种神经网络,对它的训练也需要大量数据的支持。本文使用杜克大学公开的TMI_2013OCT数据集^[27],这些视网膜OCT图像由BiopTigen公司的840 nm波段SD-OCT成像仪(轴向分辨率约为 $4.5 \mu\text{m}$)采集。图像来自正常人群和患有年龄相关性黄斑变性(AMD)患者,包含了72组训练样本对和18组测试样本对,每组样本对均由含噪声的初始图像和多帧平均降噪后的图像(视为无噪声的理想图像)组成,训练集和测试集的比例为4:1。利用训练集对搭建的MDAE网络进行训练,为了考察网络的泛化能力,本文结果部分用到的正常眼和AMD病患眼的图像均来自于测试集。

为避免训练数据量不足引起的模型欠拟合问题的出现,需采用数据增强手段来扩充训练集规模。散斑噪声是一种乘性噪声,服从正态分布,可表示为

$$p(I) = \frac{1}{\sigma_v} \exp\left(-\frac{I}{\sigma_v}\right), \quad (6)$$

式中: σ_v 是图像 I 有噪声时的方差。

为尽可能获取真实OCT图像的散斑噪声,需进行三步操作:1)将图像读取为数值矩阵 X ,使用标准三维正态分布 $N(r, c, h)$ 生成随机值 α ,其中 r, c 和 h 分别为图像的行数、列数和通道数;2)设置一散斑噪声强度系数 β ,其取值范围为 $0 \sim 1$,平均间隔取10份;3)通过计算 $(1 + \alpha\beta)X$ 来得到新的数值矩阵,再将该矩阵图像化,从而得到带噪声的图像。对训练集中的每一幅图像都随机生成10幅加入噪声的图像,从而使训练集的规模增加10倍,再将像素数为 900×450 的图像裁剪成100幅像素数为 90×45 的无重叠子图,最后可得到79200组训练样本对。

2.4 实验环境和参数设置

MDAE模型的搭建基于开源深度学习框架Keras。实验设备的内存大小为16 GB,同时使用Nvidia K80 GPU加速模型训练。为避免均方误差(MSE)损失函数带来的图像过度平滑问题^[29]的出现,本文采用均方绝对误差作为损失函数。学习率会影响网络收敛速度和精度,综合考虑数据规模和训练时间两个因素后,本文将初始学习率设置为0.001。同时,为避免采用固定学习率导致的训练过程中梯度下降陷入局部最优解的问题的出现,采用Adam优化器根据梯度的一阶矩估计和二阶矩估计来动态地调整模型的学习率。经多次实验,本工作将每次训练大小(batch size)设置为1000,即每次训练1000幅图片,80次完成一个训练轮次(epoch)。训练轮次设置为2000,以确保

损失函数不再减小和收敛到最小值,此时模型训练可达到预期效果。

2.5 评价指标

为了客观评价降噪后的图像质量,并将所提方法和其他方法进行定量比较,本文采用MSE、峰值信噪比(PSNR)和结构相似度(SSIM)作为评价指标,其表达式为

$$E_{\text{MSE}} = \frac{1}{wh} \sum_{p=0}^{w-1} \sum_{q=0}^{h-1} [I(p, q) - K(p, q)]^2, \quad (7)$$

$$R_{\text{PSNR}} = 10 \cdot \lg \left(\frac{M_I^2}{E_{\text{MSE}}} \right), \quad (8)$$

$$M_{\text{SSIM}}(I, K) = \frac{(2\mu_I\mu_K + c_1)(2\sigma_{IK} + c_2)}{(\mu_I^2 + \mu_K^2 + c_1)(\sigma_I^2 + \sigma_K^2 + c_2)}, \quad (9)$$

式中: w 和 h 分别为图像的宽度和高度; M_I 是图像的最大灰度值; μ_I 和 σ_I 是图像 I 的均值和标准差; μ_K 和 σ_K 是图像 K 的均值和标准差; σ_{IK} 是图像 I 和图像 K 的协方差; c_1 和 c_2 是常量。MSE通过计算 I 和 K 的误差来衡量降噪处理后图像与理想无噪声图像的差别,其值越小表明结果越接近真实图像。PSNR通过计算图像最大灰度值和噪声的比值来体现图像中信号的相对强

度,值越大表明对噪声的抑制越强、对信号的保留越充分。SSIM逐点比较图像 I 和图像 K 中各像素点的相似度,以描述降噪处理后图像与理想无噪声图像的相似程度,值越小说明图像细节和结构保留得越多。

3 结果

3.1 渐进式降噪

所提端到过程的神经网络由多个自编码器模块组成,每个模块都会输出一幅降噪程度不同的图像。模块数量 T 会影响降噪的程度,但并非 T 越大,图像降噪越好,这是因为模型的学习能力受限于深度网络的层数。当网络过深时,会出现梯度消失现象^[30],误差无法完成反向传播,导致模型无法学习到图像的特征,从而输出无意义的结果。大量实验表明,本工作在 T 超过4时,模型无法完成对图像特征的学习。因此,本文将 T 定为4,即所提模型由4个DAE网络构成。图3展示了一幅原始视网膜OCT图像依次经过每个模块后的降噪输出结果。经过第一次降噪后图像中的散斑噪声有了明显下降,随后降噪效果逐步小幅提升,经过4次降噪后图像质量有了明显提升。

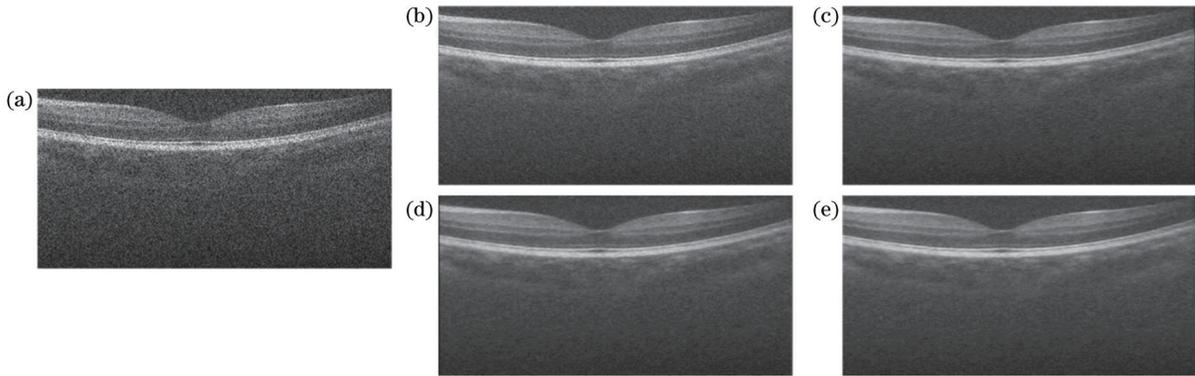


图3 不同模块数量时的降噪图像。(a)原图;(b) $T=1$;(c) $T=2$;(d) $T=3$;(e) $T=4$

Fig. 3 Denoised images by using different modules. (a) Original image; (b) $T=1$; (c) $T=2$; (d) $T=3$; (e) $T=4$

为了定量评价采用不同 T 时的降噪能力,计算了 T 为1~4时的MSE、PSNR和SSIM指标,结果如图4所示。在经过第一个模块降噪后,MSE有了显著下

降,随后每个模块的MSE仍在降低,但幅度越来越小。PSNR和SSIM在经过第一个模块降噪后均取得了大幅提高,之后经过每个模块后提升幅度逐渐变小。这

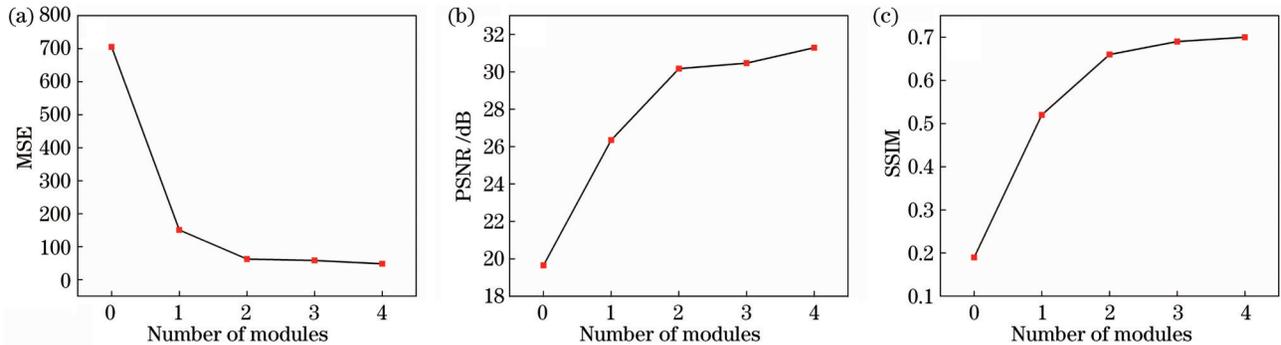


图4 评价指标随模块数量的变化。(a) MSE;(b) PSNR;(c) SSIM

Fig. 4 Evaluation metrics varying with number of modules. (a) MSE; (b) PSNR; (c) SSIM

些指标表明所提端到过程的模型可以渐进式地对图像进行降噪, $T=4$ 是本工作综合考虑降噪性能和运行时间等因素后的最佳选择。

3.2 降噪结果对比

为了评价所提模型的性能, 从测试集中随机选取正常眼的视网膜 OCT 图像进行测试, 并和其他 4 种主要方法进行比较, 具体包括: 经典的基于滤波的方法——高斯滤波^[12]和均值滤波^[10]方法; 鉴于 BM3D 方法^[14]在传统图像降噪领域中具有较高的地位, 也将其列为对比目标; 所提模型基于 DAE, 故将 DAE 的另一种改进方法——堆叠降噪自编码器 (SDAE) 方法^[22]也

作为对比目标。利用上述 5 种方法对如图 5(a) 所示的带噪声的原图进行处理, 所得结果如图 5(b)~(f) 所示。图 5(b)、(c) 为高斯滤波和均值滤波方法的结果, 二者性能较为接近, 图像中均存在残余噪声和过度平滑图像的问题, 并且均值滤波方法会导致边缘对比度降低。图 5(d) 是 BM3D 方法的结果, 虽然降噪效果不错, 但是结构出现了过度锐化。图 5(e) 是 SDAE 方法的结果, 虽然降噪效果极为显著, 但是图像模糊较为严重。图 5(f) 是所提 MDAE 方法的结果, 与 SDAE 方法相比, 在保持显著降噪效果的前提下, 保留了更多的图像细节信息。

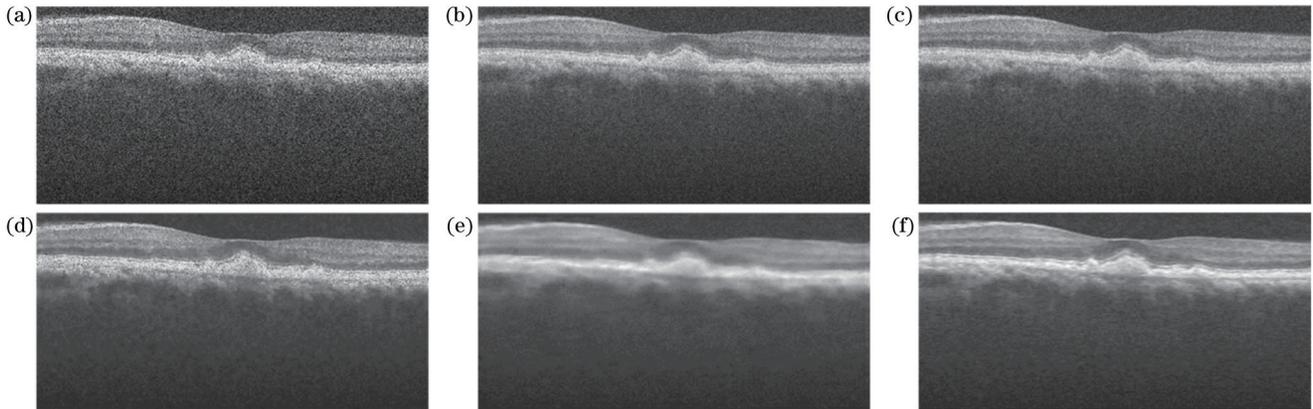


图 5 不同方法对正常眼视网膜 OCT 图像的降噪结果。(a) 原图; (b) 高斯滤波; (c) 均值滤波; (d) BM3D; (e) SDAE; (f) MDAE
Fig. 5 Denoised results of retinal OCT image of normal eye obtained by different methods. (a) Original image; (b) Gaussian filter; (c) mean filter; (d) BM3D; (e) SDAE; (f) MDAE

计算各方法所得降噪图像的评价指标以作定量比较, 结果列于表 1 中。原图的指标也列于表 1 中, 以便考察各方法的性能提升效果。由于高斯滤波和均值滤波方法是对像素点进行加权平均, 故 MSE 和 SSIM 明显劣于其他方法, 但 PSNR 与其他方法的差距并不大。BM3D 和 SDAE 方法均取得了较好的降噪结果, 但后者输出的图像略有模糊, 故 SSIM 指标不及前者。所提 MDAE 方法在三项指标上均取得了最优结果, 其中 MSE 和 PSNR 显著优于其他方法, SSIM 略优于其他方法, 表明所提方法对于正常眼视网膜 OCT 图像具有最优性能。

表 1 不同方法对正常眼视网膜 OCT 图像降噪结果的定量比较

Table 1 Quantitative comparison of denoised results obtained by different methods for retinal OCT image of normal eye

Method	MSE	PSNR /dB	SSIM
Original image	861.62	18.78	0.19
Gaussian filter	94.18	28.39	0.59
Mean filter	93.28	28.43	0.58
BM3D	82.54	28.96	0.67
SDAE	82.65	28.95	0.66
MDAE	63.40	30.10	0.69

为进一步考察所提模型对病眼视网膜 OCT 图像的降噪性能, 从测试集中随机选取了 AMD 患者的视网膜 OCT 图像进行测试。图 6 展示了各种方法的降噪结果, 原图中的箭头标示了病变区域。可以看出: 经过 MDAE 降噪后, 信息几乎没有损失, 图像复原度最高; 其他各方法的结果与正常眼的结果类似。各方法指标的定量比较如表 2 所示, MDAE 的各项指标均是所有方法中最优的, 其中 MSE 显著优于其他各方法, PSNR 相比于次优的 SDAE 仍有 0.63 dB 的提升, SSIM 也略优于其他方法, 表明所提方法对于病眼视网膜 OCT 图像也具有最优性能。

为衡量各方法的时间消耗情况, 利用它们对 170 幅像素大小为 900×450 的 OCT 图像进行降噪处理, 然后求取处理一幅图像的平均时间, 结果也列于表 2 中。高斯滤波和均值滤波方法凭借极少的计算量而有着绝对的时间优势, 但降噪效果非常不理想。BM3D 方法最为耗时, 处理一幅图像的平均时间接近 7.0 s, 这在临床实际应用中是难以忍受的。所提 MDAE 方法处理每幅图像的平均时间约为 0.26 s, 虽然与其他方法相比并无优势, 也离实时降噪的要求有着不小的差距, 但是降噪通常属于图像后处理阶段, 该时间消耗仍属于可接受的水平。

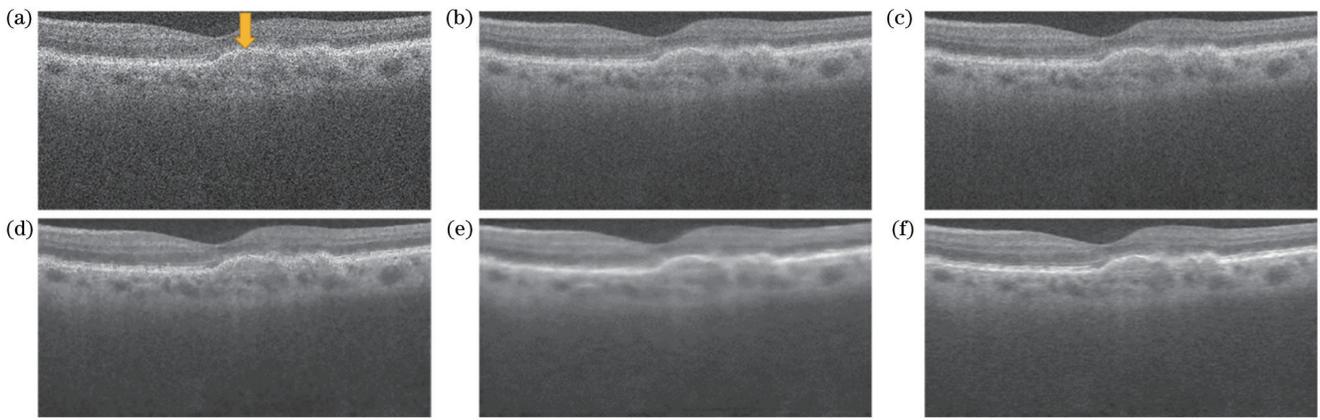


图 6 不同方法对病眼视网膜 OCT 图像的降噪结果。(a)原图;(b)高斯滤波;(c)均值滤波;(d) BM3D;(e) SDAE;(f) MDAE
Fig. 6 Denoised results of retinal OCT image of diseased eye obtained by different methods. (a) Original image; (b) Gaussian filter; (c) mean filter; (d) BM3D; (e) SDAE; (f) MDAE

表 2 不同方法对病眼视网膜 OCT 图像降噪结果的定量比较
Table 2 Quantitative comparison of denoised results obtained by different methods for retinal OCT image of diseased eye

Method	MSE	PSNR / dB	SSIM	Time /ms
Original image	963.94	18.29	0.16	
Gaussian filter	96.09	28.30	0.55	0.37
Mean filter	91.70	28.51	0.54	1.33
BM3D	72.71	29.54	0.66	6901.00
SDAE	68.97	29.74	0.67	45.71
MDAE	59.72	30.37	0.68	257.12

4 讨 论

相比于未作处理的视网膜 OCT 原图,利用所提方法得到的正常眼图像的降噪结果的峰值信噪比提高了 11.32 dB,病眼图像提高了 12.08 dB,噪声水平有了大幅下降,这为后续的高精度图像分析处理提供了可能。所提方法可通过控制模块数量 T 来控制降噪程度,以便更好地应对复杂的临床应用场景。同时,所提方法对于降噪程度的控制,只依赖参数 T ,而无需调节其他参数,省去了使用者的学习成本,这对于专注于临床任务的医生来说非常友好。另外,可考虑将每个模块的输入加入原图,来解决当 $T > 4$ 时出现的梯度消失现象,这将是本工作的后续研究方向。

虽然所提方法的直接应用目标是视网膜 OCT 图像,但是其并无特殊要求或使用范围等限制,故实现该方法的通用思路原则上也适用于其他领域的 OCT 图像。所提方法能在噪声干扰比较严重的视网膜 OCT 图像上取得显著的降噪效果,故相信在其他领域也能发挥不错的效果。考虑到深度学习的方法效果高度依赖于训练集的质量和数量,若将其应用于其他领域的 OCT 图像中,则建议根据该领域的数据集对模块的网络结构和数量进行调整,并对网络进行重新训练,使网络学习到该领域图像的特征,从而使网络的降噪性

能在具体任务中达到最佳。因此,将所提方法成功转移至其他领域的一个关键因素就是该领域具有满足要求的训练数据集。

5 结 论

针对 OCT 图像存在的噪声干扰问题和现有降噪方法只能输出最终的降噪图像的缺点,提出了一种基于 MDAE 的视网膜 OCT 图像降噪方法,实现了端到过程地输出降噪程度不同的多幅图像,其中就有性能最佳或最符合使用需求的结果。通过在公开数据集上的测试发现,模块数量为 4 时,所提方法可取得最优的降噪效果。以 MSE、PSNR 和 SSIM 作为评价指标,在正常眼和患病眼的视网膜 OCT 图像上,将所提方法与其他主流方法进行了定量比较。结果表明,所提方法在各项评价指标上均取得了较好结果,可以有效地对视网膜 OCT 图像进行降噪处理和大幅提升图像的质量。

参 考 文 献

- [1] Huang D, Swanson E A, Lin C P, et al. Optical coherence tomography[J]. Science, 1991, 254(5035): 1178-1181.
- [2] 赵蒙蒙, 鲁贞贞, 朱书缘, 等. 基于变分自编码器的眼科光学相干断层成像图像生成[J]. 光学学报, 2021, 41(14): 1417001. Zhao M M, Lu Z Z, Zhu S Y, et al. Generation of optical coherence tomography images in ophthalmology based on variational auto-encoder[J]. Acta Optica Sinica, 2021, 41(14): 1417001.
- [3] 刘新文, 王惠南, 钱志余. 小波变换对 OCT 图像的降噪处理[J]. 光子学报, 2006, 35(6): 935-939. Liu X W, Wang H N, Qian Z Y. Denoising process of OCT image based on wavelet transform[J]. Acta Photonica Sinica, 2006, 35(6): 935-939.
- [4] Schmitt J M, Xiang S H, Yung K M. Speckle in optical coherence tomography[J]. Journal of Biomedical Optics, 1999, 4(1): 95-105.
- [5] Ciulla T A, Amador A G, Zinman B. Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies[J]. Diabetes Care, 2003, 26(9): 2653-2664.
- [6] Jager R D, Mieler W F, Miller J W. Age-related macular

- degeneration[J]. The New England Journal of Medicine, 2008, 358(24): 2606-2617.
- [7] Jørgensen T M, Thomadsen J, Christensen U, et al. Enhancing the signal-to-noise ratio in ophthalmic optical coherence tomography by image registration: method and clinical examples[J]. Journal of Biomedical Optics, 2007, 12(4): 041208.
- [8] Alonso-Caneiro D, Read S A, Collins M J. Speckle reduction in optical coherence tomography imaging by affine-motion image registration[J]. Journal of Biomedical Optics, 2011, 16(11): 116027.
- [9] Qiu B, Huang Z Y, Liu X, et al. Noise reduction in optical coherence tomography images using a deep neural network with perceptually-sensitive loss function[J]. Biomedical Optics Express, 2020, 11(2): 817-830.
- [10] Deshpande S D, Er M H, Venkateswarlu R, et al. Max-mean and max-median filters for detection of small targets[J]. Proceedings of SPIE, 1999, 3809: 74-83.
- [11] Brownrigg D R K. The weighted median filter[J]. Communications of the ACM, 1984, 27(8): 807-818.
- [12] Deng G, Cahill L W. An adaptive Gaussian filter for noise reduction and edge detection[C]//IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, October 31-November 6, 1993, San Francisco, CA, USA. New York: IEEE Press, 1993: 1615-1619.
- [13] Aum J, Kim J H, Jeong J. Effective speckle noise suppression in optical coherence tomography images using nonlocal means denoising filter with double Gaussian anisotropic kernels[J]. Applied Optics, 2015, 54(13): D43-D50.
- [14] Chong B, Zhu Y K. Speckle reduction in optical coherence tomography images of human finger skin by wavelet modified BM3D filter[J]. Optics Communications, 2013, 291: 461-469.
- [15] Fang L Y, Li S T, Nie Q, et al. Sparsity based denoising of spectral domain optical coherence tomography images[J]. Biomedical Optics Express, 2012, 3(5): 927-942.
- [16] Zhang X, Li Z L, Nan N, et al. Denoising algorithm of OCT images via sparse representation based on noise estimation and global dictionary[J]. Optics Express, 2022, 30(4): 5788-5802.
- [17] Aydogan O, Alberto B, Adrien E, et al. Speckle reduction in optical coherence tomography images using digital filtering[J]. Journal of the Optical Society of America A, 2007, 24(7): 1901-1910.
- [18] Schlegl T, Waldstein S M, Bogunovic H, et al. Fully automated detection and quantification of macular fluid in OCT using deep learning[J]. Ophthalmology, 2018, 125(4): 549-558.
- [19] 唐艳红, 陈允照, 刘明迪, 等. 基于 CNN 和改进的图搜索分割 OCT 图像中的视网膜层[J]. 激光与光电子学进展, 2020, 57(24): 241702.
- Tang Y H, Chen Y Z, Liu M D, et al. Segmentation of retinal layers in OCT images based on CNN and improved graph search[J]. Laser & Optoelectronics Progress, 2020, 57(24): 241702.
- [20] Karri S P K, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration[J]. Biomedical Optics Express, 2017, 8(2): 579-592.
- [21] 连超铭, 钟舜聪, 张添福, 等. 光学相干断层扫描视网膜图像的迁移学习分类[J]. 激光与光电子学进展, 2021, 58(1): 0117002.
- Lian C M, Zhong S C, Zhang T F, et al. Transfer learning-based classification of optical coherence tomography retinal images[J]. Laser & Optoelectronics Progress, 2021, 58(1): 0117002.
- [22] Vincent P, Laroche H, Lajoie I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. Journal of Machine Learning Research, 2010, 11: 3371-3408.
- [23] Ma Y H, Chen X J, Zhu W F, et al. Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN[J]. Biomedical Optics Express, 2018, 9(11): 5129-5146.
- [24] Halupka K J, Antony B J, Lee M H, et al. Retinal optical coherence tomography image enhancement via deep learning[J]. Biomedical Optics Express, 2018, 9(12): 6205-6221.
- [25] Shan H M, Padole A, Homayounieh F, et al. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction[J]. Nature Machine Intelligence, 2019, 1(6): 269-276.
- [26] Shimobaba T, Endo Y, Hirayama R, et al. Autoencoder-based holographic image restoration[J]. Applied Optics, 2017, 56(13): F27-F30.
- [27] Fang L Y, Li S T, McNabb R P, et al. Fast acquisition and reconstruction of optical coherence tomography images via sparse representation[J]. IEEE Transactions on Medical Imaging, 2013, 32(11): 2034-2049.
- [28] Bashkansky M, Reintjes J. Statistics and reduction of speckle in optical coherence tomography[J]. Optics Letters, 2000, 25(8): 545-547.
- [29] Eckert M P, Bradley A P. Perceptual quality metrics applied to still image compression[J]. Signal Processing, 1998, 70(3): 177-200.
- [30] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6(2): 107-116.

Denoising Method of Retinal OCT Images Based on Modularized Denoising Autoencoder

Dai Hao^{1,2,3}, Yang Yaliang^{1,2*}, Yue Xian^{1,2,3}, Chen Shen^{1,2,3}

¹Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, Sichuan, China;

²Key Laboratory of Adaptive Optics, Chinese Academy of Sciences, Chengdu 610209, Sichuan, China;

³University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

Objective Optical coherence tomography (OCT) has the characteristics of high resolution, high sensitivity and high speed. However, affected by factors such as the high scattering of tissue, the micro-movement of target, and the jitter of

hardware during imaging process, OCT images always carry noise dominated by speckle noise, which reduces the accuracy of the subsequent processing. How to denoise the image to improve the image quality has been highly concerned. Current denoising methods based on deep learning are almost end-to-end, which means that the denoising degree is uncontrollable. However, the noise intensity may be different in different cases, and its uncontrollable denoising degree will lead to the reduction of the generalization ability of the model. For doctors, the denoising degree required is different depending on the patient's condition. The end-to-end deep learning network limits the autonomous control of the denoising degree. Therefore, achieving end-to-process denoising is of great significance in clinical applications.

Methods The TMI_2013OCT dataset publicly available from Duke University is used in this work, which is obtained from the normal population and patients with age-related macular degeneration (AMD). In order to avoid the under-fitting problem of the model caused by insufficient training data, data augmentation is used to expand the size of training set to 79200 pairs. Using multi-layer convolution and deconvolution to build an autoencoder, a modularized denoising autoencoder (MDAE) is built based on the architecture of a modular deep neural network. Each autoencoder module can sequentially output a image with gradually increased denoising degree. Process results meet different usage requirements. In order to reduce the amount of parameters and improve the training speed, all modules share parameters (i. e. , all modules have the same parameters). Mean square error (MSE), peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as evaluation metrics.

Results and Discussions To quantitatively evaluate the denoising capability with different number of modules (T), the MSE, PSNR and SSIM are calculated when T varies from 1 to 4 (Fig. 4). The MSE has dropped significantly after the denoising of the first module, and then is still decreased but the magnitude is getting smaller and smaller after the denoising of the subsequent each module. Both PSNR and SSIM have been greatly improved after denoising in the first module, and the magnitude gradually becomes smaller after subsequent each module. These metrics show that the proposed end-to-process model can denoise the image progressively, and $T=4$ is the best choice in this work after considering factors such as denoising performance and running time. To evaluate the performance of the proposed model, retinal OCT images of normal eyes are randomly selected from the test set for testing and compared with Gaussian filtering, mean filtering, block-matching and 3D filtering (BM3D), and stacked denoising autoencoder (SDAE) methods (Fig. 5). On the premise of maintaining a significant denoising effect, the proposed method preserves more image details, and has the best results in all the metrics compared to others (Table 1). To further examine the denoising performance of the proposed method on retinal OCT images of diseased eyes, images from AMD patients are randomly selected from the test set for testing (Fig. 6). After MDAE denoising, there is almost no information loss, and the image restoration is the highest. All the metrics of MDAE except time are the best among all methods (Table 2), indicating that the proposed method also has the best performance for retinal OCT images of diseased eyes. Gaussian filtering and mean filtering methods have absolute time advantages with very little calculation, but the denoising effect is very unsatisfactory. The BM3D method is the most time-consuming, and the average time to process an image is close to 7.0 s, which is unbearable in practical clinical applications. The average time for MDAE method to process each image is about 0.26 s. Although it has no advantage compared to other methods and is far from the requirement of real-time denoising, denoising usually belongs to the post-processing stage of a image, and this time consumption is still at an acceptable level.

Conclusions Compared to the original retinal OCT image without processing, the PSNR of the denoised result obtained by the proposed method is increased by 11.32 dB and 12.08 dB for the images from normal eyes and diseased eyes, respectively, and the noise level is greatly reduced. This provides the possibility for subsequent high-precision image processing and analysis. The proposed method can control the denoising degree by controlling the number of modules, so as to be more suitable for complex clinical applications. At the same time, the proposed method only relies on the parameter T to control the denoising degree without adjusting other parameters, which saves the user's learning cost and is very friendly to doctors who focus on clinical tasks.

Key words image processing; retinal optical coherence tomography image; image denoising; denoising autoencoder; deep learning