第 43 卷 第 20 期/2023 年 10 月/光学学报 研究论文

光学学报

基于深度学习的视网膜OCT图像无监督去噪方法

吴广义，袁卓群，梁艳梅*

南开大学现代光学研究所，天津市微尺度光学信息技术科学重点实验室，天津 300350

摘要 以散斑噪声为主的噪声干扰严重影响视网膜光学相干层析(OCT)图像质量。深度学习是一种有效的去噪方法。但对活体成像而言，其很难获取多帧配准的真值图像，这影响了监督学习方法的效果。提出一种无监督深度残差稀疏注意力网络用于视网膜OCT图像去噪，并分别从视觉评价和数值评价两方面与传统的三维块匹配滤波去噪算法和经典的深度学习去噪网络进行对比。研究了监督学习与无监督学习策略下3种卷积神经网络的去噪性能，并利用公开的视网膜OCT图像数据集进行泛化能力测试。实验结果表明：所提算法的视觉评价和数值评价均具有良好的降噪效果，可以实现视网膜OCT图像高质量降噪，具有较强的泛化性，而且与监督学习相比，无监督学习在数据集不充分时仍能获得较好的降噪性能，可以有效地辅助医生进行准确高效的临床诊断。

关键词 光学相干层析技术；视网膜；图像去噪；深度学习；无监督学习

中图分类号 O436 **文献标志码** A **DOI**：10.3788/AOS230720

1 引 言

光学相干层析(OCT)技术由于其高分辨率和无创成像等优点，已成为多种眼部疾病安全有效的诊断工具[1-2]，如年龄相关性黄斑变性(AMD)[3]和糖尿病性黄斑水肿(DME)[4]，其并被视作眼科疾病诊断领域的“金标准”。然而，OCT技术在成像过程中不可避免地存在噪声[5]，如电路的热噪声、探测器的散粒噪声和相干引入的散斑噪声，这些噪声会降低OCT图像的对比度和分辨率，为像素级别的视网层分割[6]和厚度测量[7]带来困难。因此，在对视网膜OCT图像进行降噪的同时最大程度地保留图像的分层和边缘等结构细节信息具有十分重要的意义，其能辅助医生对眼部疾病进行诊断以及后续治疗。

OCT图像降噪方法分为硬件和软件两类。基于硬件的方法需要对OCT系统的架构进行改进和优化[8]，但这样会使系统结构变得更为复杂，造价也更为昂贵。基于软件的方法是利用数字图像处理算法来执行所采集OCT图像的后处理。传统的数字图像去噪方法大致可分为以下几类：基于滤波的方法，如均值滤波器[9]和高斯滤波器[10]等，利用人工设计的滤波器进行降噪，通常会模糊图像中的特征和边缘信息；基于块匹配的方法，如非局部均值(NLM)[11]和三维块匹配滤波(BM3D)[12]等，需要针对不同的噪声水平进行复杂的参数调整；基于变换的方法[13-14]在变换域(频率域或小波域)中处理退化的OCT图像，但在变换域中引入的任何意外伪影都会扩散到整个图像；基于图像稀疏性的方法，如字典学习[15]，虽然可以有效地抑制噪声，但去噪后的图像过于平滑，可能会导致图像细节丢失；多帧*B*-scan图像平均是常用的去除OCT散斑噪声的方法，但其会增加采样时间，且人眼不自主地运动会降低图像平均后的效果，容易出现伪影[16-17]。

近年来，一些基于深度学习的方法，特别是卷积神经网络(CNN)，已经被广泛地应用于医学图像处理任务，如图像分类[18]、图像分割[19]等，同时这些方法也为图像去噪提供了新思路。针对视网膜OCT图像的去噪，Ma等[20]提出了一种条件生成对抗网络(cGAN)用于OCT图像降噪，选择U形网络(U-Net)[21]作为cGAN中的生成器以生成低噪图像，视觉几何组(VGG)[22]作为判别器用以甄别真值图像和降噪图像，并向目标函数中引入与边缘信息显式相关的损失，增强对边缘信息的保持能力，该方法在性能和泛化能力方面优于传统方法。Qiu等[23]利用修剪的去噪卷积神经网络(DnCNN)[24]对视网膜OCT图像进行降噪，并对比研究了包括感知敏感、平均绝对值误差和平均平方误差等8种损失函数对OCT图像去噪表现的影响，证明该方法在保留图像细节方面优于NLM和BM3D。Dai等[25]利用多层卷积和反卷积构建自编码

收稿日期：2023-03-27；**修回日期**：2023-04-30；**录用日期**：2023-05-06；**网络首发日期**：2023-05-16

基金项目：国家自然科学基金(61875092)、京津冀基础研究合作专项(19JCZDJC65300)、天津市科技支撑重点项目(17YFZC-SY00740)

通信作者：*ymliang@nankai.edu.cn

器，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结果，能够灵活地满足不同降噪任务的使用需求，该方法取得了优于传统滤波器和BM3D的降噪效果。

与传统算法相比，基于深度学习的降噪方法在图像质量，特别是在保留图像边缘细节方面取得较好的改善，但模型的训练需要大量多帧平均的真值图像以构成Clean-Noisy图像对并供神经网络学习映射函数。然而，多帧平均的真值图像的获取较为复杂且耗时较长，不利于降噪方法的推广。近年来，Lehtinen等[26]提出了一种无监督深度学习去噪策略，即Noise2Noise（N2N）策略。与监督学习方式相比，N2N策略仅使用成对的噪声图像数据集来训练深度学习模型，实现类似监督学习的去噪性能。Gisbert等[27]率先将N2N策略应用到视网膜OCT图像去噪中，选用U-Net作为去噪网络，取得了较好的降噪效果。Huang等[28]利用改进的SRResNet，并结合N2N策略实现了对不同样品的OCT图像实时降噪。

本文提出了一种基于N2N训练策略的无监督深度残差稀疏注意力网络（DRSA-Net）用于视网膜OCT图像去噪。只需在同一样本位置获得两个*B*-scan OCT图像构成Noisy-Noisy图像对，将其中一个噪声图像作为输入，另一个噪声图像作为标记进行训练，无需多帧平均的无噪声图像作为标签。分别从定性的视觉评价和定量的数值评价两方面对所提的DRSA-Net与经典的深度学习去噪网络（U-Net、DnCNN）进行分析，研究这3种网络模型对视网膜OCT图像降噪的有效性，并与传统的BM3D方法进行比较。然后，对比监督学习与无监督学习策略下3种CNN的去噪效果。最后，在公开的视网膜OCT图像数据集中进行泛化能力测试。

2　方 法

2.1　N2N无监督训练原理

基于深度学习的图像降噪任务通常被认为是回归模型，具体可表示为

$$\arg\min_{\theta} E_{(x,y)}\left\{L\left[f_\theta(x),y\right]\right\}, \tag{1}$$

式中：$\arg\min(\ )$为使目标函数取最小值时的变量值；f为神经网络的映射函数；x为输入噪声图像；y为真值图像；E为观测值的期望值；L为损失函数；θ为神经网络参数。

若忽略输入的噪声图像x和真值图像y之间的关联，根据贝叶斯定理这个过程可表示为

$$\arg\min_{\theta} E_x\left\{E_{y|x}\left\{L\left[f_\theta(x),y\right]\right\}\right\}。 \tag{2}$$

由式(2)可知，理论上神经网络可以将噪声图像x和真值图像y的损失函数的优化过程独立，这视为分别对其计算数学期望。因此，用期望值等于真值图像y的含噪数据来替换真值，可以保证神经网络的拟合效果不变。N2N训练策略的详细数学推导见文献[26]。

OCT图像的噪声在大样本下均值为0，能够用噪声图像x对真值图像y进行替换。

此时，若输入和标签图像都采用噪声图像，则网络的目标函数可表示为

$$\arg\min_{\theta}\sum_i L\left(f_\theta(\hat{x}_i),\hat{y}_i\right)。 \tag{3}$$

为了保证N2N训练策略的有效性，必须满足两个前提条件：1）x_i和y_i的噪声是独立的；2）x_i和y_i具有相同的无噪声目标。OCT图像噪声随机且相互独立，并对样品同一个位置多帧OCT图像进行采集可以保证无噪声目标具有相同的结构，所以OCT图像满足N2N去噪策略的前提条件。

2.2　DRSA-Net结构

基于U-Net对多尺度信息的充分利用及DnCNN的残差学习和批量标准化的优点，并结合空洞卷积[29]和注意力机制[30]，提出一种DRSA-Net，该网络由4个模块组成：局部稀疏注意力模块（LSAB）、深度提取模块（DEB）、全局注意力模块（GAB）、残差模块（RB），网络结构如图1所示。

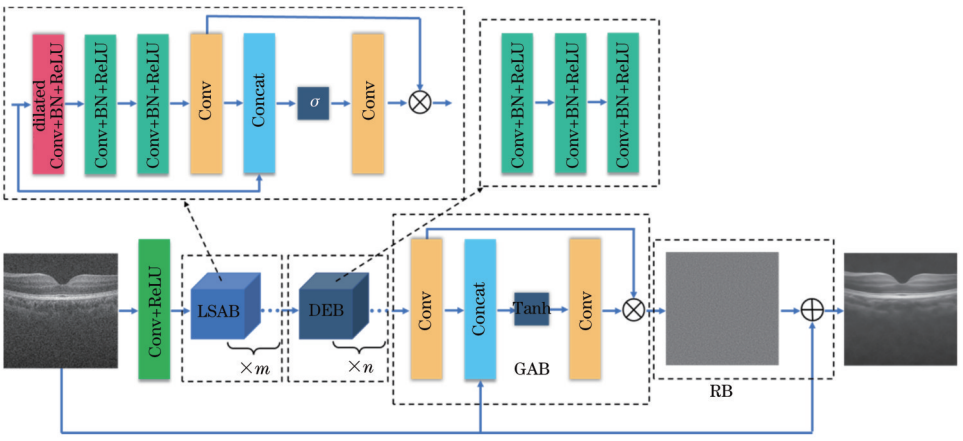

图1　DRSA-Net网络结构图

Fig. 1　Network structure of DRSA-Net

图像输入网络中首先由一个Conv和ReLU激活函数提取出原始特征图，并将其作为下一模块的输入。

LSAB采用普通卷积和空洞卷积搭配使用的策略，组成稀疏块，在扩大感受野的同时减少了参数的使用，避免了空洞卷积的“网格效应”[31]。包括3种类型的卷积层：空洞卷积（dilated Conv）+批归一化（BN）+ReLU、Conv+BN+ReLU、Conv。空洞卷积核的空洞率设置为2，卷积核的大小设置为3×3。Concat层是特征融合层，利用残差连接将经过稀疏卷积前后的特征进行融合，充分利用不同尺度的信息，提高去噪效果。选用σ激活函数来增加模块的非线性表达能力，并应用注意力机制来指导CNN训练去噪模型，使用前一阶段的学习结果来指导现阶段的噪声学习，让网络自动将更多的计算资源倾注于任务中的重要特征，以此来提高效率。

随着网络深度的增加，较深的网络可能会受到浅层对深层的削弱影响[32]，且使用空洞卷积会损失降噪精度。为了解决此问题，提出由3个Conv+BN+ReLU组成的用于图像特征深度提取的DEB，通过长路径对LSAB提取出的特征进行深度提取，并与LSAB进行搭配达到最佳的降噪效果。

类似于LSAB中的局部注意力，GAB是对全局特征的注意力再分配。其利用最后一个卷积核大小为1×1的Conv将获得的特征压缩为向量，并将其作为调整前一阶段的权重。GAB模块将局部注意力与全局注意力相结合，充分利用局部和全局信息来挖掘鲁棒性更强的特征，并将注意力机制应用在多尺度层面，在提取局部信息的同时有效关注全局语义信息，实现对数据的深入挖掘。

RB借鉴DnCNN中的残差学习策略，在隐藏层中将真实的图像从噪声图像中去除，从而利用噪声信息对网络进行训练，在不改变去噪性能的同时加快收敛速度。

此外，LSAB数量m和DEB数量n可调，其可根据数据集的大小、任务的复杂度而改变，避免盲目堆叠层数使网络层数过深，从而导致网络退化[33]。

2.3 数据集

为了更好地验证去噪网络和训练策略的有效性，选用在视网膜OCT图像降噪任务中被广泛使用的Duke大学公开数据集[34]来训练和评估网络的去噪效果。该数据集采用Bioptigen公司的840 nm光谱域光学相干层析成像（SD-OCT）系统获取，其轴向分辨率为4.5 μm，对患有AMD或健康的13名受试者的眼睛以视网膜中央凹为中心的正方形区域进行扫描。数据集共含有39组图像，每组由5张相邻的图像构成，像素大小为450 pixel×450 pixel。

N2N无监督训练策略要求训练的图像对具有除噪声外一致的结构，即在同一个位置多次采样的Noisy-Noisy图像对。将5张连续位置的噪声图像分别标号为1、2、3、4、5，通用的数据集构建方法都会采用奇-偶分组原则[35-36]，但这种方式没有充分利用数据集，并不适用于图像较少的情况。为了最大程度地匹配N2N的去噪原理以达到最好的去噪效果，将1、2、3、4构成噪声数据集Noisy 1，2、3、4、5构成噪声数据集Noisy 2，组成1-2、2-3、3-4、4-5的Noisy-Noisy图像对，相比于奇-偶分组，这种分组方式能够在最大程度地保持两张噪声图像组织结构信息相似的情况下获得更丰富的数据集。

真值图像由5张连续位置的噪声图像采用多帧平均法合成。真值图像集与噪声数据集Noisy 1和Noisy 2构成噪声-真值图像对，用于后续的监督训练，并与无监督训练策略进行对比。在训练时为了便于网络快速收敛，将输入的450 pixel×450 pixel噪声图像随机裁剪为128 pixel×128 pixel，并把每张图像重复裁剪50次，构成一个分别含有7800张图像的噪声数据集Noisy 1、噪声数据集Noisy 2和真值图像数据集Clean的数据集，并按照11∶2的比例划分训练集和测试集。这些图像分别构成无监督学习Noisy 1-Noisy 2噪声图像对和监督学习Clean-Noisy 1图像对进行对比训练。

2.4 实验环境和参数设置

使用3种类型的CNN（U-Net、DnCNN和DRSA-Net），并利用BM3D作为传统降噪算法的代表作对比试验，实验流程如图2所示。实验在配备16 GB操作内存、RTX2060显卡和英特尔I7 CPU的计算机上进行；去噪网络基于PyTorch构建、训练和测试。批大小设置为16，学习率为0.001，Adam为优化器。最大迭代次数设置为100，在100次迭代中具有最小损失值的模型被保存以供进一步评估。

选取L_2作为损失函数，L_2可表示为

$$L_2=\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}\left[I_{\mathrm{d}}(x,y)-I_{\mathrm{l}}(x,y)\right]^2, \tag{4}$$

式中：W、H为图像的宽、高，W、H的值均为128；$I_{\mathrm{d}}(x,y)$、$I_{\mathrm{l}}(x,y)$分别为去噪图像和标签图像的灰度值。

2.5 评价指标

主观视觉评价是通过志愿者的主观感知来评价图像的质量，对志愿者的评价结果取平均来量化不同模型的降噪效果。具体来说，邀请10名志愿者对降噪图像进行打分，通过比较不同网络输出的降噪图像，打分区间为1～5分，取平均值作为最终的视觉评价结果。

客观数值评价选取4种常用的定量指标对图像的去噪效果进行评价，包括峰值信噪比（PSNR）、结构相似性（SSIM）、等效视数（ENL）、边缘保留指数（EPI）。

PSNR为去噪图像和真值图像对应像素之间的差异，是最经典的降噪评价指标，具体可表示为

$$f_{\mathrm{PSNR}}=10\lg\left[\frac{\mathrm{Max}(\boldsymbol{I})^2}{f_{\mathrm{MSE}}}\right], \tag{5}$$

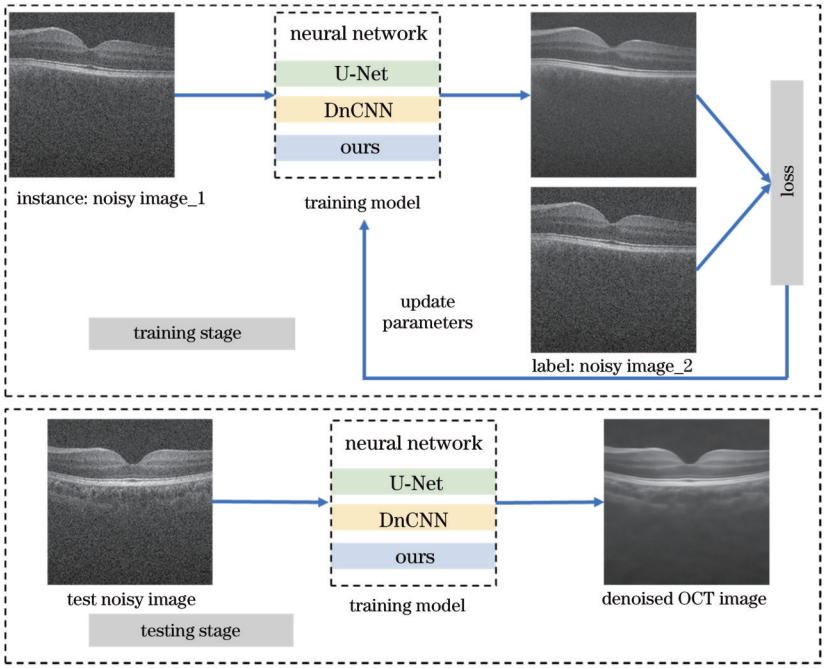

图 2　视网膜 OCT 图像无监督训练实验流程图

Fig. 2　Flow chart of retinal OCT image unsupervised training experiment

式中：f_{PSNR} 为 PSNR 的值；Max(I)为图像 I 中像素的最大灰度值；f_{MSE} 为去噪图像 I_{d} 和真值图像 I_{c} 之间的均方误差，f_{MSE} 可表示为

$$f_{\mathrm{MSE}}=\frac{1}{W\times H}\sum_{x=0}^{W-1}\sum_{y=0}^{H-1}\left\|I_{\mathrm{d}}(x,y)-I_{\mathrm{c}}(x,y)\right\|^{2}。\tag{6}$$

SSIM 从亮度 l、对比度 c、结构 s 等方面的相似度来比较去噪图像和真值图像之间的纹理差异，具体可表示为

$$f_{\mathrm{SSIM}}=\left[l(I_{\mathrm{d}},I_{\mathrm{c}})^{\alpha}\cdot c(I_{\mathrm{d}},I_{\mathrm{c}})^{\beta}\cdot s(I_{\mathrm{d}},I_{\mathrm{c}})^{\gamma}\right],\tag{7}$$

式中：$l(I_{\mathrm{d}},I_{\mathrm{c}})$、$c(I_{\mathrm{d}},I_{\mathrm{c}})$、$s(I_{\mathrm{d}},I_{\mathrm{c}})$ 分别可表示为

$$l(I_{\mathrm{d}},I_{\mathrm{c}})=\frac{2\mu_{\mathrm{d}}\mu_{\mathrm{c}}+c_1}{\mu_{\mathrm{d}}^2+\mu_{\mathrm{c}}^2+c_1},\tag{8}$$

$$c(I_{\mathrm{d}},I_{\mathrm{c}})=\frac{2\sigma_{\mathrm{d}}\sigma_{\mathrm{c}}+c_2}{\sigma_{\mathrm{d}}^2+\sigma_{\mathrm{c}}^2+c_2},\tag{9}$$

$$s(I_{\mathrm{d}},I_{\mathrm{c}})=\frac{\sigma_{\mathrm{dc}}+c_3}{\sigma_{\mathrm{d}}\sigma_{\mathrm{c}}+c_3},\tag{10}$$

式中：μ_{d}、μ_{c} 分别为 I_{d}、I_{c} 的平均值；σ_{d}、σ_{c} 分别为 I_{d}、I_{c} 的标准差；σ_{dc} 为 I_{d} 和 I_{c} 的协方差；c_1、c_2、c_3 为常数。α、β、γ 通常设置为 1。

ENL 通常用于测量降噪图像中均匀区域的平滑度。图像中第 i 个感兴趣区域(ROI)上的 ENL 可表示为

$$f_{\mathrm{ENL},i}=\frac{\mu_i^2}{\sigma_i^2},\tag{11}$$

式中：$f_{\mathrm{ENL},i}$ 为 ENL 的值；μ_i、σ_i 分别为图像中第 i 个 ROI 的平均值、标准偏差。实验中选取 3 个 ROI 来计算平均 ENL。

EPI 是反映降噪图像边缘细节保持程度的性能度量。图像中第 i 个 ROI 上的 EPI 可表示为

$$f_{\mathrm{EPI}}=\frac{\sum_{W}\sum_{H}\left|I_{\mathrm{n}}(x+1,y)-I_{\mathrm{n}}(x,y)\right|}{\sum_{W}\sum_{H}\left|I_{\mathrm{d}}(x+1,y)-I_{\mathrm{d}}(x,y)\right|},\tag{12}$$

式中：f_{EPI} 为 EPI 的值；I_{n}、I_{d} 为原始噪声图像和去噪图像。在整个图像上，计算该指数可能不是描述边缘细节保持程度的准确指标，因为降噪后图像中均匀区域的梯度将变小。因此，只计算图像边界附近的 ROI。

3　结果与分析

3.1　无监督训练结果

经实验测试，对本视网膜 OCT 图像去噪任务，DRSA-Net 选取 $m=4$、$n=2$ 时具有最佳降噪性能。图 3 为原始噪声图像 Noisy、5 帧平均图像 ground truth、BM3D 去噪图像和 N2N 无监督训练策略下 3 种 CNN 模型的视网膜 OCT 图像的去噪结果，包括 DnCNN-N2N、U-Net-N2N 和 Ours-N2N。

为了更好地分辨视网膜的多层结构，在有信号区域选择 3 个 ROI，分别为图中的Ⅰ、Ⅱ和Ⅲ矩形框。结果表明，BM3D 对含有高噪声的原始 OCT 图像具有一定的去噪效果，但其会引入一些块状的模糊结构，使图像过于平滑而导致纹理结构被破坏。N2N 无监督训练策略下的 3 种深度学习模型都在保留视网膜层结构信息的情况下实现较好的去噪性能，噪声去除程度较高，视网膜多层结构清晰。通过分析 10 名志愿者的主观视觉评价，3 种 CNN 的总体去噪性能排序为：Ours-N2N＞U-Net-N2N＞DnCNN-N2N。

对背景区域而言，Ours-N2N 中的噪声明显减少，

可以得到较为干净的背景，但在 U-Net-N2N 和 DnCNN-N2N 中仍可以观察到一些噪声，如图 3 中背景区域处箭头所示。对信号区域而言，视网膜 OCT *B*-scan 图像的边缘信息和层的精细结构在 Ours-N2N 中保持较好。U-Net-N2N 倾向于破坏层内细节和层边界，同时会引入一些层间的模糊结构，如图(e)中Ⅱ处箭头所示；而 DnCNN-N2N 导致图(c)的Ⅰ处箭头所指的层边界退化以及Ⅲ处箭头所指的外界膜模糊。

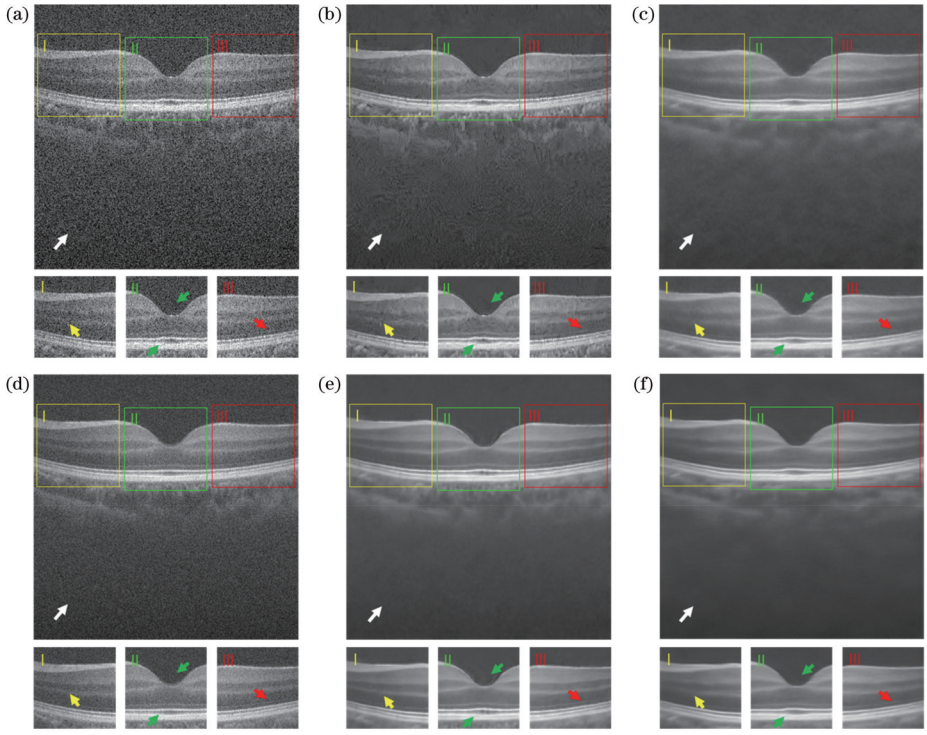

图 3　视网膜 OCT 图像降噪结果。(a)原始噪声图像；(b) BM3D 去噪图像；(c) DnCNN-N2N 去噪图像；(d) 5 帧平均 ground truth 图像；(e) U-Net-N2N 去噪图像；(f) Ours-N2N 去噪图像

Fig. 3　Noise reduction in retinal OCT images. (a) Original noisy image; (b) denoised images of BM3D; (c) denoised image of DnCNN-N2N; (d) 5-frame average ground truth images; (e) denoised image of U-Net-N2N; (f) denoised image of Ours-N2N

3.2　监督训练与无监督训练结果对比

图 4 为监督学习 Noise2Clean(N2C)与无监督学习 N2N 视网膜 OCT 图像降噪结果对比。通过分析 10 名志愿者的主观视觉评价，3 种监督学习模型的总体去噪性能排序为：Ours-N2C＞U-Net-N2C＞DnCNN-N2C。可以发现监督学习的去噪结果十分接近 ground truth 图像，但仍具有较多噪声，而无监督学习的 3 种模型噪声去除程度都较高，能提供更清晰的结构和边缘信息。一般来说，同一个网络模型采取监督学习的回归效果会优于无监督学习。造成图 4 中同一个网络结构下无监督学习的降噪效果优于监督学习的主要原因是 ground truth 图像的噪声去除程度较低。当高质量的多帧平均视网膜 ground truth OCT 图像获取困难时，不依赖于真值图像的无监督训练策略体现出优势，该策略的噪声去除程度更高，视网膜多层结构更清晰。

表 1 为 BM3D 算法以及 3 种网络监督学习和无监督学习降噪图像的客观数值评价指标。由表 1 可知，与原始噪声图像相比，监督学习和无监督学习模型在图像的各项评价指标方面都实现较大的提升，且相比于传统的块匹配算法 BM3D，基于深度学习的降噪算法将去噪时间缩短两个数量级，可以在 1 s 内完成对 OCT 图像的高水平降噪。对 N2N 无监督学习而言，所提模型在 PSNR、SSIM、EPI 参数上表现最好，U-Net 取得最高的 ENL 值，这与 U-Net 将 OCT 图像处理得过于平滑有关，可以解释图 4(g) Ⅱ中箭头处的层间过于平滑的模糊结构；对监督学习而言，所提模型在 PSNR、EPI、ENL 参数上表现最好，DnCNN 在 SSIM 和去噪时间两方面取得更好的结果。

从数值上看，监督学习的大部分指标优于无监督学习，但这与视觉评价相反。PSNR 和 SSIM 是基于 ground truth 图像计算得出的，而 5 帧平均的 ground truth 图像噪声去除程度较低，因为监督学习网络输出的最理想的降噪图像就是 ground truth，其在噪声去除程度较低的情况下能够获得高于无监督学习网络的分数。此外，无监督学习的 EPI 不仅低于监督学习，还低于传统的去噪算法 BM3D，因为 EPI 是基于原始噪声图像计算的，BM3D 算法和 5 帧平均的 ground truth 下的监督训练网络保留较多的去噪图像的噪声，这与原始高噪图像较为接近，所以带来更高的 EPI。

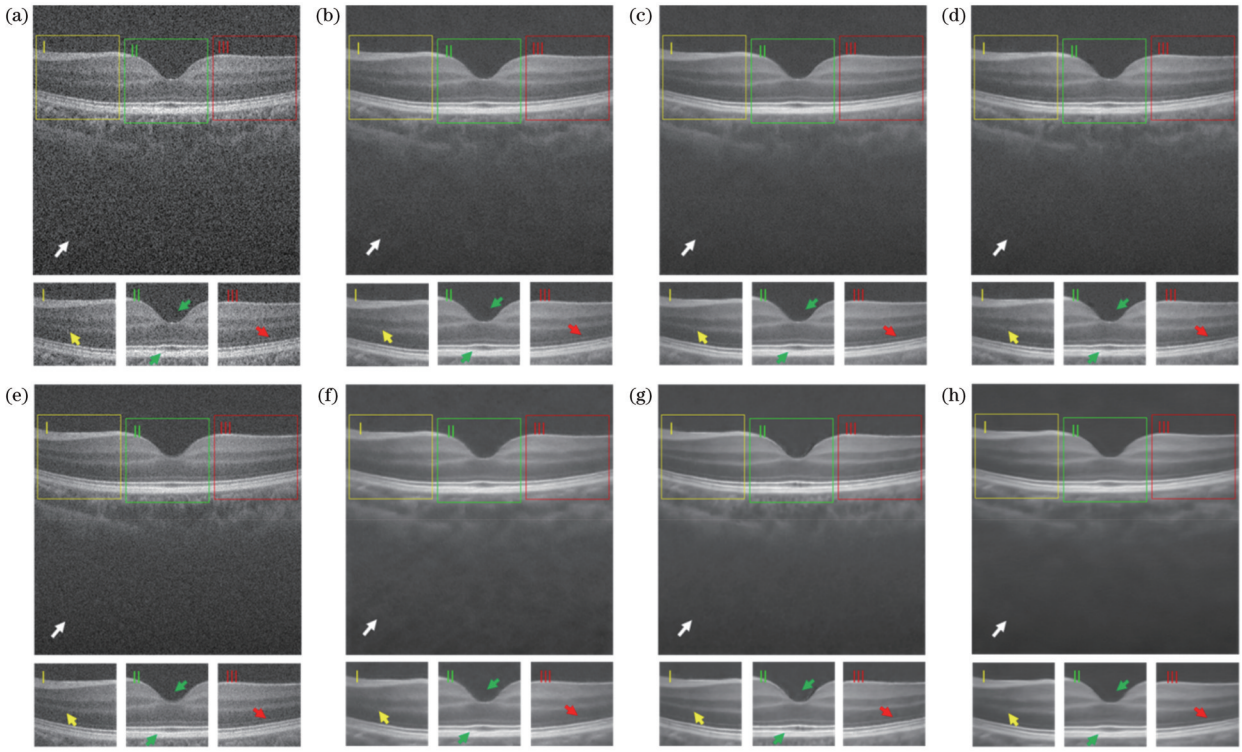

图 4 监督学习与无监督学习视网膜 OCT 图像降噪结果。(a) 原始噪声图像;(b) DnCNN-N2C 去噪图像;(c) U-Net-N2C 去噪图像;(d) Ours-N2C 去噪图像;(e) 5 帧平均 ground truth 图像;(f) DnCNN-N2N 去噪图像;(g) U-Net-N2N 去噪图像;(h) Ours-N2N 去噪图像

Fig. 4 Noise reduction results of supervised learning and unsupervised learning retinal OCT images. (a) Original noisy image; (b) denoised image of DnCNN-N2C; (c) denoised image of U-Net-N2C; (d) denoised image of Ours-N2C; (e) 5-frame average ground truth image; (f) denoised image of DnCNN-N2N; (g) denoised image of U-Net-N2N; (h) denoised image of Ours-N2N

表 1 监督学习与无监督学习去噪数值评价结果

Table 1 Results of supervised learning and unsupervised learning denoising numerical evaluation

Method		PSNR	SSIM	EPI	ENL	Time /s
Baseline	Noisy	19.073±0.065	0.391±0.002		14.938±1.693	
	BM3D	23.937±0.181	0.253±0.101	0.281±0.051	434.819±111.647	127
Supervised learning model	DnCNN	25.418±0.213	0.506±0.007	0.295±0.023	220.663±14.472	0.48
	U-Net	24.852±0.322	0.498±0.006	0.304±0.024	255.427±20.642	0.70
	Ours	25.447±0.191	0.494±0.007	0.312±0.025	279.760±27.946	0.53
Unsupervised N2N model	DnCNN	24.234±0.183	0.284±0.007	0.242±0.027	768.128±137.606	0.48
	U-Net	24.543±0.212	0.287±0.008	0.243±0.028	1601.956±573.328	0.70
	Ours	24.582±0.225	0.289±0.008	0.262±0.026	1304.384±466.983	0.53

在衡量图像均匀区域平滑程度的 ENL 这项指标中,无监督学习网络明显高于监督学习网络。由图 4 可知,无监督学习噪声水平去除程度更高,图像更为平滑,这与视觉评价结果相同。

4 讨 论

降噪任务是 OCT 图像处理中最重要的部分之一,为医生对视网膜疾病的初步诊断和后续图像的处理和分析奠定基础。尽管两种经典的深度学习去噪网络 U-Net 和 DnCNN 能实现较好的降噪性能,但采用监督学习策略,需大量高质量的 Clean-Noisy 图像对来构成的数据集,而传统的块匹配去噪算法 BM3D 在降噪水平和时间代价上的表现一般。基于 N2N 训练策略,提出一种无监督 DRSA-Net 用于视网膜 OCT 图像降噪,在视觉评价和数值评价方面都优于经典的降噪算法。当数据集不充分时,无监督学习仍能获得较好的降噪性能,与监督学习相比是更优的选择。

从 Duke 大学另一组公开的视网膜数据集[15]中选取 12 对像素大小为 900 pixel×450 pixel,且配准良好的 Clean-Noisy 图像对并进行去噪网络泛化性测试。一方面评估去噪网络在不同数据集之间的泛化能力,另一方面验证对上述 ground truth 图像噪声去除程度

低引起的PSNR和SSIM两个指标视觉评价与数值评价不符的猜想。该数据集提供由40帧图像平均获得的ground truth。

图5为原始噪声图像Noisy、40帧平均图像ground truth、BM3D去噪图像和N2N无监督训练策略下3个CNN模型(DnCNN-N2N、U-Net-N2N、Ours-N2N)的视网膜OCT图像的去噪结果。

由图5可知,40帧平均ground truth图像的噪声去除程度较5帧平均得到大幅提升。选择3个ROI以更好地分辨视网膜的多层结构,分别为图5中Ⅰ、Ⅱ和Ⅲ矩形框。无监督训练策略下3种深度学习模型都在保留视网膜多层结构信息的情况下实现较好的去噪性能,具有较强的泛化性。原始Noisy图像中图5Ⅰ处箭头所指白色斑点结构为视网膜血管,在高噪声背景下并不明显,3种无监督模型在去噪后都能准确清晰地分辨出具有高对比度的血管。Ours-N2N获得更好的视觉性能,在背景噪声去除方面,能够优于ground truth获得更干净的背景,如背景区域处箭头所示;在结构信息保持方面,其具有更清晰的层结构和更均匀的层,如图5Ⅱ、Ⅲ处箭头所示。

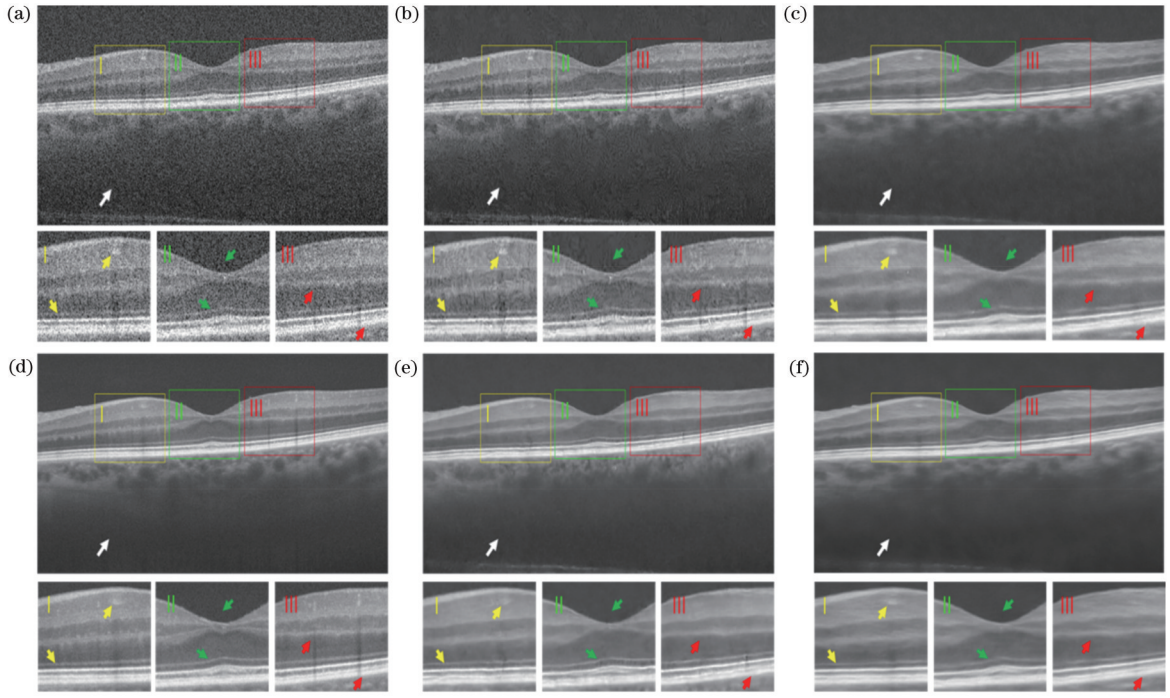

图5 无监督学习泛化能力测试。(a) 原始噪声图像;(b) BM3D去噪图像;(c) DnCNN-N2N去噪图像;(d) 40帧平均ground truth图像;(e) U-Net-N2N去噪图像;(f) Ours-N2N去噪图像

Fig. 5 Unsupervised learning generalization ability test. (a) Original noisy image; (b) denoised images of BM3D; (c) denoised image of DnCNN-N2N; (d) 40-frame average ground truth images; (e) denoised image of U-Net-N2N; (f) denoised image of Ours-N2N

图6为监督学习与无监督学习3种网络的去噪性能对比。由图6可知,无监督训练的噪声去除和结构信息保留能力优于监督训练,具有更好的对比度和视觉效果。

表2为BM3D算法以及3种网络监督学习和N2N无监督学习去噪结果的数值评价指标。与原始噪声图像相比,监督模型和无监督模型在各项图像评价指标方面都得到巨大的提升,其均可对OCT图像进行高水平降噪。所提算法在绝大多数的数值评价指标上都能获得领先,这与视觉评价结果相符。无监督学习在数值评价的两个关键指标PSNR和SSIM方面均低于监督学习,这与视觉评价结果相反,这是ground truth图像质量较低造成的。在使用40帧平均的ground truth图像后,无监督学习的PSNR和SSIM大幅度领先于监督学习,验证了上述猜想,也说明基于N2N的无监督学习在不具有充足的、高质量的Clean-Noisy图像对的情况下仍能获得较好的降噪效果,其与监督学习相比是更优的选择。

设计消融实验来验证网络结构中各模块的作用。表3为DEB+GAB+RB、LSAB+GAB+RB、LSAB+DEB+RB、LSAB+DEB+GAB、LSAB+DEB+GAB+RB等5种网络的去噪实验。由表5可知,LSAB+DEB+GAB+RB组合在各项评价指标方面均更优,这充分说明网络结构中各个模块对高质量降噪都具有贡献作用。DEB+GAB+RB在缺少LSAB后,衡量图像均匀区域平滑程度的ENL显著下降,这说明LSAB在提升图片平滑度中具有不可或缺的作用;LSAB+GAB+RB在衡量去噪图像质量的两个关键指标PSNR和SSIM中表现最差,这说明特征DEB对网络性能提升具有关键性;LSAB+DEB+RB

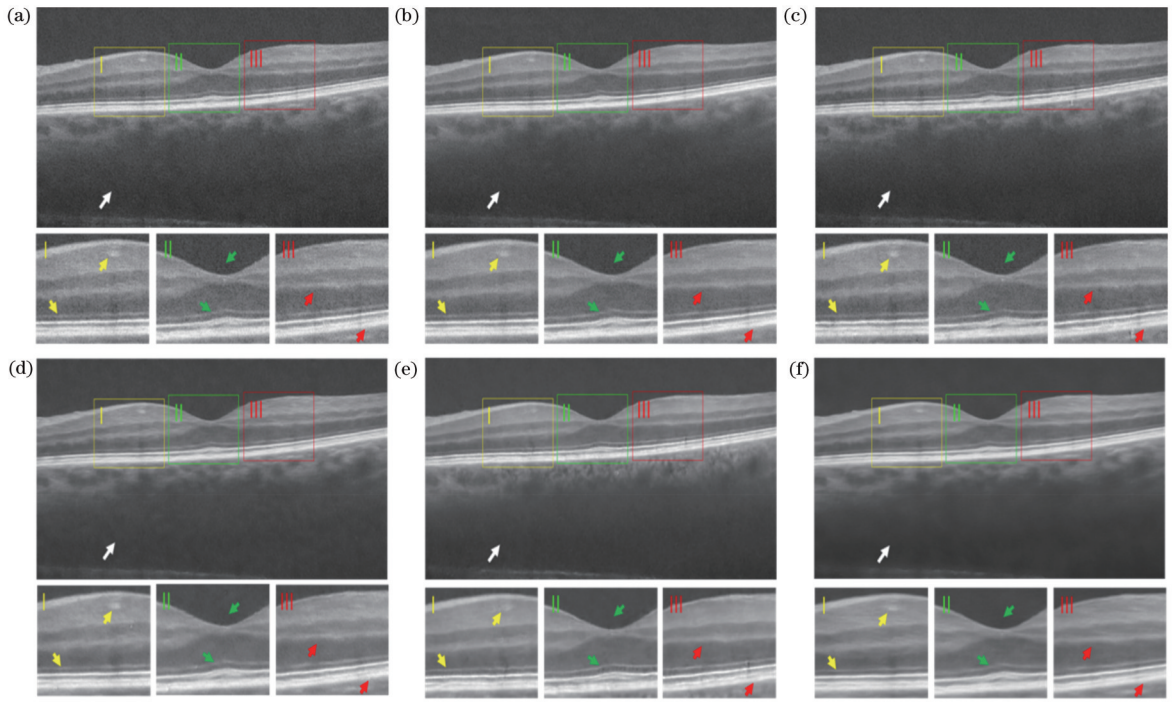

图 6　监督学习与无监督学习泛化能力对比。(a)DnCNN-N2C 去噪图像;(b)U-Net-N2C 去噪图像;(c)Ours-N2C 去噪图像;(d)DnCNN-N2N 去噪图像;(e)U-Net-N2N 去噪图像;(f)Ours-N2N 去噪图像

Fig. 6　Comparison of generalization ability between supervised and unsupervised learning. (a) Denoised image of DnCNN-N2C; (b) denoised image of U-Net-N2C; (c) denoised image of Ours-N2C; (d) denoised image of DnCNN-N2N; (e) denoised image of U-Net-N2N; (f) denoised image of Ours-N2N

表 2　监督学习与无监督学习去噪数值评价结果

Table 2　Results of supervised learning and unsupervised learning denoising numerical evaluation

Method		PSNR	SSIM	EPI	ENL	Time /s
Baseline	Noisy	18.193±0.273	0.134±0.024		14.074±3.681	
	BM3D	28.035±1.457	0.550±0.044	0.268±0.035	418.037±106.729	256
Supervised learning model	DnCNN	28.699±1.301	0.554±0.036	0.277±0.027	216.478±37.898	0.79
	U-Net	27.956±1.186	0.569±0.035	0.287±0.028	241.361±31.499	1.04
	Ours	29.317±1.419	0.593±0.034	0.293±0.030	262.771±34.762	0.92
Unsupervised N2N model	DnCNN	29.673±1.195	0.665±0.012	0.166±0.037	628.151±182.529	0.79
	U-Net	30.950±1.565	0.702±0.012	0.176±0.036	1266.897±338.543	1.04
	Ours	31.172±1.706	0.706±0.013	0.194±0.035	1029.639±220.714	0.92

在各项评价指标中均大幅落后于含有 GAB 的 LSAB+DEB+GAB+RB,这证明 GAB 能更好地指导网络进行高效去噪;缺少 RB 的 LSAB+DEB+GAB 的各项评价指标与 LSAB+DEB+GAB+RB 相当,但 RB 可以利用噪声信息进行训练,在不改变去噪性能的同时加快模型收敛速度。上述网络单帧图像的降噪时间差异较小,均在 1 s 以内,其能够实现对 OCT 图像的快速降噪。

表 3　网络不同模块消融实验去噪结果

Table 3　Ablation experimental results of different modules of network

Module	PSNR	SSIM	EPI	ENL	Time/s
DEB+GAB+RB	28.962±1.632	0.703±0.015	0.131±0.031	159.334±13.502	0.73
LSAB+GAB+RB	28.737±1.264	0.613±0.026	0.133±0.072	879.870±207.982	0.69
LSAB+DEB+RB	30.301±1.446	0.647±0.012	0.150±0.035	903.572±343.217	0.73
LSAB+DEB+GAB	30.802±1.427	0.701±0.012	0.127±0.033	937.421±225.073	0.75
LSAB+DEB+GAB+RB	31.172±1.706	0.706±0.013	0.194±0.035	1029.639±220.714	0.92

5 结 论

针对视网膜OCT图像中的噪声干扰以及活体成像中高质量多帧平均图像获取困难的问题，提出了一种不依赖于真值图像的无监督深度残差稀疏注意力降噪算法。通过注意力机制搭配稀疏卷积核完成数据间信息高效充分的挖掘，并基于N2N训练策略利用噪声图像对完成高质量的训练，在保留视网膜OCT图像多层结构信息的同时实现高水平的降噪。分别从视觉评价和PSNR、SSIM、EPI、ENL等4种数值评价对传统的去噪算法和经典的深度学习网络进行对比分析，监督学习与无监督学习的去噪效果及对公开视网膜OCT图像数据集的泛化能力测试的实验结果表明，所提的降噪算法在各项评价指标上均取得较好的结果，具有较强的泛化性。与监督学习相比，无监督学习在数据集不充分时仍能够获得较好的降噪性能。

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] Drexler W, Morgner U, Ghanta R K, et al. Ultrahigh-resolution ophthalmic optical coherence tomography[J]. Nature Medicine, 2001, 7(4): 502-507.

[3] Cukras C, Wang Y D, Meyerle C B, et al. Optical coherence tomography-based decision making in exudative age-related macular degeneration: comparison of time- vs spectral-domain devices[J]. Eye, 2010, 24(5): 775-783.

[4] Virgili G, Menchini F, Casazza G, et al. Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy[J]. The Cochrane Database of Systematic Reviews, 2015, 1: CD008081.

[5] 朱晓农, 毛幼馨, 梁艳梅, 等. 光学相干层析系统噪音分析(Ⅱ): 时域OCT和频域OCT[J]. 光子学报, 2007, 36(3): 457-461.
Zhu X N, Mao Y X, Liang Y M, et al. Noise analyses of optical coherence tomography systems (Ⅱ): Fourier domain and time domain OCT systems[J]. Acta Photonica Sinica, 2007, 36(3): 457-461.

[6] 贺琪欲, 李中梁, 王向朝, 等. 基于光学相干层析成像的视网膜图像自动分层方法[J]. 光学学报, 2016, 36(10): 1011003.
He Q Y, Li Z L, Wang X Z, et al. Automated retinal layer segmentation based on optical coherence tomographic images[J]. Acta Optica Sinica, 2016, 36(10): 1011003.

[7] Balasubramanian M, Bowd C, Vizzeri G, et al. Effect of image quality on tissue thickness measurements obtained with spectral domain-optical coherence tomography[J]. Optics Express, 2009, 17(5): 4019-4036.

[8] 袁治灵, 陈俊波, 黄伟源, 等. 基于稳健性主成分分析算法的光学相干层析成像去除散斑噪声的研究[J]. 光学学报, 2018, 38(5): 0511002.
Yuan Z L, Chen J B, Huang W Y, et al. Speckle noise reduction of optical coherence tomography based on robust principle component analysis algorithm[J]. Acta Optica Sinica, 2018, 38(5): 0511002.

[9] 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.

[10] Deng G, Cahill L W. An adaptive Gaussian filter for noise reduction and edge detection[C]//1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, October 31-November 6, 1993, San Francisco, CA, USA. New York: IEEE Press, 2002: 1615-1619.

[11] 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.

[12] 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.

[13] Mayer M A, Borsdorf A, Wagner M, et al. Wavelet denoising of multiframe optical coherence tomography data[J]. Biomedical Optics Express, 2012, 3(3): 572-589.

[14] Zhang A Q, Xi J F, Sun J T, et al. Pixel-based speckle adjustment for noise reduction in Fourier-domain OCT images [J]. Biomedical Optics Express, 2017, 8(3): 1721-1730.

[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] Baghaie A, Yu Z Y, D'Souza R M. Involuntary eye motion correction in retinal optical coherence tomography: hardware or software solution? [J]. Medical Image Analysis, 2017, 37: 129-145.

[17] 屈慧, 汪毅, 娄世良, 等. 纯随机相位板散斑去相关光学相干层析成像[J]. 光学学报, 2023, 43(1): 0111002.
Qu H, Wang Y, Lou S L, et al. Speckle decorrelation optical coherence tomography with pure random phase plate[J]. Acta Optica Sinica, 2023, 43(1): 0111002.

[18] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. JAMA, 2016, 316 (22): 2402-2410.

[19] 孙正, 王树雁. 深度学习在血管内光学相干层析成像中的应用现状[J]. 激光与光电子学进展, 2022, 59(22): 2200002.
Sun Z, Wang S Y. Application of deep learning in intravascular optical coherence tomography[J]. Laser & Optoelectronics Progress, 2022, 59(22): 2200002.

[20] 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.

[21] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]// Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention - MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.

[22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2023-02-04]. https://arxiv.org/abs/1409.1556.

[23] 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.

[24] Zhang K, Zuo W M, Chen Y J, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.

[25] 代豪, 杨亚良, 岳献, 等. 基于模块化降噪自编码器的视网膜OCT图像降噪方法[J]. 光学学报, 2023, 43(1): 0110001.
Dai H, Yang Y L, Yue X, et al. Denoising method of retinal OCT images based on modularized denoising autoencoder[J]. Acta Optica Sinica, 2023, 43(1): 0110001.

[26] Lehtinen J, Munkberg J, Hasselgren J, et al. Noise2Noise: learning image restoration without clean data[EB/OL]. (2018-03-12)[2023-02-04]. https://arxiv.org/abs/1803.04189.

[27] Gisbert G, Dey N, Ishikawa H, et al. Improved denoising of optical coherence tomography via repeated acquisitions and unsupervised deep learning[J]. Investigative Ophthalmology & Visual Science, 2020, 61(9): PB0035.

[28] Huang Y, Zhang N, Hao Q. Real-time noise reduction based on ground truth free deep learning for optical coherence tomography [J]. Biomedical Optics Express, 2021, 12(4): 2027-2040.
[29] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23)[2023-02-04]. https://arxiv.org/abs/1511.07122.
[30] Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: a survey[J]. Computational Visual Media, 2022, 8(3): 331-368.
[31] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE Press, 2018: 1451-1460.
[32] Tai Y, Yang J, Liu X M, et al. MemNet: a persistent memory network for image restoration[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4549-4557.
[33] Duvenaud D, Rippel O, Adams R P, et al. Avoiding pathologies in very deep networks[EB/OL]. (2014-02-24)[2023-02-01]. https://arxiv.org/abs/1402.5836.
[34] 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.
[35] Wu D F, Kim K, Li Q Z. Low-dose CT reconstruction with Noise2Noise network and testing-time fine-tuning[J]. Medical Physics, 2021, 48(12): 7657-7672.
[36] Qiu B, You Y F, Huang Z Y, et al. N2NSR-OCT: simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning[J]. Journal of Biophotonics, 2021, 14(1): e202000282.

Unsupervised Denoising of Retinal OCT Images Based on Deep Learning

Wu Guangyi, Yuan Zhuoqun, Liang Yanmei*

Institute of Modern Optics, Nankai University, Tianjin Key Laboratory of Micro-Scale Optical Information Science and Technology, Tianjin 300350, *China*

Abstract

Objective Optical coherence tomography (OCT) is employed as a safe and effective diagnostic tool for a variety of ophthalmic diseases due to its high resolution and non-invasive imaging, which is regarded as the "gold standard" in ophthalmic disease diagnosis. However, various kinds of noise, especially speckle noise, seriously affect the quality of retinal OCT images to reduce the contrast and resolution, which makes it difficult to segment and measure retinal sublayer thickness at the pixel level. Therefore, it is of significance to reduce the noise of retinal OCT images and retain structural details such as layering and edges of the images to the greatest extent. The deep learning-based noise reduction method shows advantages in image quality, especially in preserving edge details. However, for *in vivo* imaging, it is difficult to obtain a large number of multi-frame registration ground truth images, which affects the performance of the supervised learning method. Therefore, the realization of unsupervised denoising independent of ground truth images is vital in the clinical diagnosis of eye diseases.

Methods We propose an unsupervised deep residual sparse attention network (DRSA-Net) based on the Noise2Noise training strategy for retinal OCT image denoising. DRSA-Net consists of local sparse attention block (LSAB), depth extraction block (DEB), global attention block (GAB), and residual block (RB). The TMI_2013OCT dataset publicly provided by Duke University is selected and preprocessed, and a total of 7800 Clean-Noisy and Noisy-Noisy image pairs are obtained. The proposed DRSA-Net is compared with the classical deep learning denoising networks U-Net and DnCNN from two aspects of qualitative visual evaluation and quantitative numerical evaluation and is also compared with the traditional BM3D algorithm. Then the denoising effects of three convolutional neural networks under supervised learning and unsupervised learning strategies are compared. Finally, generalization ability tests and network module ablation experiments are performed based on another public retinal OCT image dataset.

Results and Discussions The results of unsupervised training denoising (Fig. 3) show that the built model has better denoise and intra-layer fine structure preservation ability for retinal OCT images. U-Net-N2N tends to destroy the details and boundary of layers and introduces some fuzzy structures among layers. DnCNN-N2N brings degradation of layer boundary and blurring of the outer limiting membrane. The comparison between the results of supervised training and unsupervised training (Fig. 4) indicates that when ideal ground truth images cannot be provided, the denoised images of the supervised learning model have more noise, while the unsupervised learning model has a higher denoise degree and can provide clearer structures and edge information. The denoising numerical evaluation results of supervised learning and unsupervised learning (Table 1) show that compared with the original noise images, the supervised learning and

unsupervised learning models realize great improvement in various evaluation indexes of the images. Additionally, compared with the traditional block matching algorithm BM3D, the denoising algorithm based on deep learning reduces the denoising time by two orders of magnitude. High-quality noise reduction of OCT images can be achieved within 1 s, and the proposed algorithm can get ahead of most evaluation indexes regardless of what kind of training strategy is adopted. The test results of generalization ability of unsupervised learning (Fig. 5) show that our proposed model has better generalization ability among different datasets, and can obtain a cleaner background than ground truth in terms of background denoise. In terms of structural information retention, it has clearer interlayer structures and more uniform layers. The results of ablation experiments on different modules of the denoising network proposed (Table 3) indicate that the combination of LSAB+DEB+GAB+RB is better in various evaluation indexes, which fully demonstrates the contribution of each module in the network structure to high-quality noise reduction.

Conclusions We put forward an unsupervised depth residual sparse attention denoising algorithm independent of ground truth images to solve the noise interference in retinal OCT images and the difficulty of acquiring high-quality multi-frame average images in *in vivo* imaging. The attention mechanism is combined with sparse convolution kernel to complete the information mining between data efficiently and fully, and the Noise2Noise training strategy is adopted to complete the high-quality training with noise images, which achieves a high level of noise reduction and preserves the multi-layer structure information of retinal OCT images. The traditional denoising algorithm and the classical deep learning network are compared and analyzed from the visual evaluation and numerical evaluation including PSNR, SSIM, EPI, and ENL respectively. The denoising effect of supervised learning and unsupervised learning and the experimental results of the generalization ability test on the public retinal OCT image dataset show that the proposed noise reduction algorithm yields good results in various evaluation indexes and has strong generalization. Compared with supervised learning, unsupervised learning can still obtain better noise reduction performance under insufficient data sets.

Key words optical coherence tomography; retina; image denoising; deep learning; unsupervised learning