E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2024)08-2426-13

论文引用格式:Wang Z, Huo G L, Lan H, Hu J M and Wei X. 2024. Fundus image enhancement algorithm based on convolutional dictionary diffusion model. Journal of Image and Graphics, 29(08):2426-2438(王珍, 霍光磊, 兰海, 胡建民, 魏宪. 2024. 基于卷积字典扩散模型的眼底图像增强算法. 中国图象图形学报, 29(08):2426-2438)[DOI:10.11834/jig.230595]

基于卷积字典扩散模型的眼底图像增强算法

王珍[1,2],霍光磊[3*],兰海[2],胡建民[4],魏宪[5]

1. 福建农林大学机电工程学院,福州　350108; 2. 中国科学院福建物质结构研究所,福州　350002; 3. 泉州通维科技有限责任公司,泉州　362008; 4. 福建医科大学医学技术与工程学院,福州　350122; 5. 华东师范大学软件工程学院,上海　200062

摘　要:**目的**　视网膜眼底图像广泛用于临床筛查和诊断眼科疾病,但由于散焦、光线条件不佳等引起的眼底图像模糊,导致医生无法正确诊断,且现有图像增强方法恢复的图像仍存在模糊、高频信息缺失以及噪点增多问题。本文提出了一个卷积字典扩散模型,将卷积字典学习的去噪能力与条件扩散模型的灵活性相结合,从而解决了上述问题。**方法**　算法主要包括两个过程:扩散过程和去噪过程。首先向输入图像中逐步添加随机噪声,得到趋于纯粹噪声的图像;然后训练一个神经网络逐渐将噪声从图像中移除,直到获得一幅清晰图像。本文利用卷积网络来实现卷积字典学习并获取图像稀疏表示,该算法充分利用图像的先验信息,有效避免重建图像高频信息缺失和噪点增多的问题。**结果**　将本文模型在EyePACS数据集上进行训练,并分别在合成数据集DRIVE(dgital retinal images for vessel extraction)、CHASEDB1(child heart and health study in England)、ROC(retinopathy online challenge)和真实数据集RF(real fundus)、HRF(high-resolution fundus)上进行测试,验证了所提方法在图像增强任务上的性能及跨数据集的泛化能力,其评价指标峰值信噪比(peak signal-to-noise ratio,PSNR)和学习感知图像块相似度(learned perceptual image patch similarity,LPIPS)与原始扩散模型(learning enhancement from degradation,Led)相比平均分别提升了1.992 9 dB和0.028 9。此外,将本文方法用于真实眼科图像下游任务的前处理能够有效提升下游任务的表现,在含有分割标签的DRIVE数据集上进行的视网膜血管分割实验结果显示,相较于原始扩散模型,其分割指标对比其受试者工作特征曲线下面积(area under the curve,AUC),准确率(accuracy,Acc)和敏感性(sensitivity,Sen)平均分别提升0.031 4,0.003 0和0.073 8。**结论**　提出的方法能够在保留真实眼底特征的同时去除模糊、恢复更丰富的细节,从而有利于临床图像的分析和应用。

关键词:眼底图像增强;卷积字典学习;稀疏表示;扩散模型;条件扩散模型

Fundus image enhancement algorithm based on convolutional dictionary diffusion model

Wang Zhen[1,2], Huo Guanglei[3*], Lan Hai[2], Hu Jianmin[4], Wei Xian[5]

1. College of Mechanical and Electrical Engineering, Fujian Agriculture and Forestry University, Fuzhou　350108, China; 2. Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou　350002, China; 3. Quanzhou Tongwei Technology Co., Ltd., Quanzhou　362008, China; 4. School of Medical Technology and Engineering, Fujian Medical University, Fuzhou　350122, China; 5. Software Engineering Institute, East China Normal University, Shanghai　200062, China

Abstract: **Objective** Retinal fundus images have important clinical applications in ophthalmology. These images can be used to screen and diagnose various ophthalmic diseases, such as diabetic retinopathy, macular degeneration, and glau-

收稿日期:2023-08-29;**修回日期**:2023-11-08;**预印本日期**:2023-11-15

*通信作者:霍光磊　hgl2124@hitqz.com

基金项目:泉州市科技项目(2023C009R,2022C004L)

Supported by:Quanzhou Science and Technology Plan Project (2023C009R, 2022C004L)

coma. However, the acquisition of these images is often affected by various factors in real scenarios, including lens defocus, poor ambient light conditions, patient eye movements, and camera performance. These issues often lead to quality problems such as blurriness, unclear details, and inevitable noise in fundus images. Such poor-quality images pose a challenge to ophthalmologists in their diagnostic work. For example, blurred images will lead to the absence of detailed information about the morphological structure of the retina, which causes difficulty for the physicians to accurately localize and identify abnormalities, lesions, exudations, and other conditions. Existing enhancement methods for fundus images have progressed in improving image quality. However, some problems still exist, such as image blurring, artifacts, missing high-frequency information, and increased noise. Therefore, in this study, we propose a convolutional dictionary diffusion model, which combines convolutional dictionary learning with conditional diffusion model. This algorithm aims to cope with the abovementioned problems of low-quality images to provide an effective tool for fundus image enhancement. Our approach can improve the quality of fundus images and enable physicians to increase diagnostic confidence, improve assessment accuracy, monitor treatment progress, and ensure better care for patients. This method will contribute to ophthalmic research and provide more opportunities for prospective healthcare management and medical intervention, which positively impacts patients' ocular health and overall quality of life. **Method** The algorithm consists of two parts: simulation of diffusion process and inverse denoising process. First, random noise is gradually added to the input image to obtain a purely noisy image. Then, a neural network is trained to gradually remove the noise from the image until a clear image is finally obtained. This study takes the blurred fundus image as the conditional information to better preserve the fine-grained structure of the image. Collecting blurred-clear fundus image pairs is difficult. Thus, synthetic fundus dataset is widely used for training. Therefore, a Gaussian filtering algorithm is designed to simulate the defocus blur images. In the training process, the conditional information and the noisy image are first spliced and fed into the network, and the abstract features of the image are extracted by continuously reducing the image size through downsampling. This procedure can significantly reduce the time and space complexity of the sparse representation calculation. Then, the convolutional network is used to implement convolutional dictionary learning and obtain the sparse representation of the image. Given that the self-attention mechanism can capture non-local similarity and long-range dependency, this study adds self-attention to the convolutional dictionary learning module to improve the reconstruction quality. Finally, hierarchical feature extraction is achieved by feature concatenation to realize information fusion between different levels and better use local features in the image. The downsampled feature is recovered to the original image size by an inverse convolutional layer. The model minimizes the negative log-likelihood loss, which represents the difference in probability distribution between the generated image and the original image. After the model is trained, a clear fundus image is generated by gradually removing the noise from a noisy picture with a blurred image as conditional input. **Result** The proposed method was evaluated on EyePACS dataset, and multiple experiments were performed on synthetic datasets DRIVE (digital retinal images for vessel extraction), CHASEDB1 (child heart and health study in England), ROC (retinopathy online challenge), realistic datasets RF (real fundus) and HRF (high-resolution fundus) to demonstrate the generalizability of our model. Experimental results show that the evaluation metrics peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity (LPIPS) are improved on average by 1.992 9 and 0.028 9, respectively, compared with the original diffusion model (learning enhancement from degradation (Led)). Moreover, the proposed approach was used as a preprocessing module for downstream tasks. The experiment on retinal vessel segmentation is adopted to prove that our approach can benefit the downstream tasks in clinical application. The results of segmentation experiments on the DRIVE dataset show that all the segmentation metrics improve compared with the original diffusion model. Specifically, the area under the curve (AUC), accuracy (Acc), and sensitivity (Sen) are improved by 0.031 4, 0.003 0, and 0.073 8 on average, respectively. **Conclusion** The proposed method provides a practical tool for fundus image deblurring and a new perspective to improve the quality and accuracy of diagnostic. This approach has a positive impact on patients and ophthalmologists and is expected to promote further development in the interdisciplinary research of ophthalmology and computer science.

Key words: fundus image enhancement; convolutional dictionary learning; sparse representation; diffusion model; conditional diffusion model

0 引 言

眼底图像是一种重要的诊断工具，广泛应用于眼科疾病的诊断和监测(Schmidt-Erfurth 等，2018)。然而，眼底图像的质量往往受到多种因素的影响，如成像设备限制、镜头散焦、光线条件和患者合作度等。这些因素可能导致眼底图像出现模糊、低对比度、不均匀照明和噪声等问题(如图 1 所示)，从而导致医生无法做出准确的诊断和制定有效的治疗方案。一项基于 UK BioBank(MacGillivray 等，2015)的研究表明，进行视网膜成像的 68 544 名参与者中，大约有 30% 的眼底图像由于质量不足而无法被眼科医生识别。在不可处理的图像中，大约 80% 的图像完全或大部分模糊。因此将低质量眼底图像进行增强，去除模糊得到清晰的眼底图像是非常有意义的。

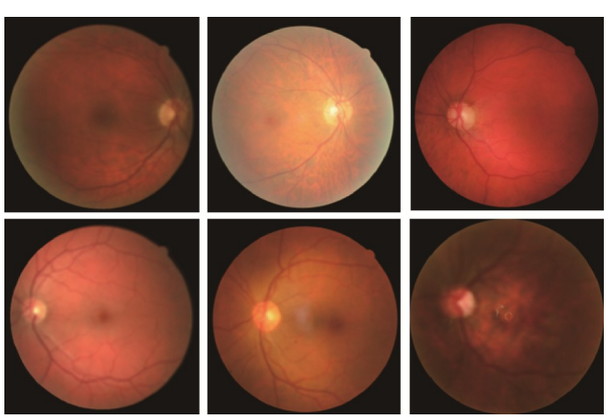


图 1 模糊眼底图像

Fig. 1 Blurred fundus images

由于深度学习技术强大的图像表示能力，其在计算机视觉和医学图像分析领域得到了广泛应用。在卷积神经网络的方法中，Gaudio 等人(2020)在去雾理论基础上，提出了一种像素颜色放大理论来促进视网膜图像的血管分割任务。Zamir 等人(2020)提出多尺度残差块的特征融合策略，确保感受野可以在不牺牲原始特征细节的情况下动态调整。Zhu 等人(2023)为了减轻低质量图像与增强图像之间的信息不一致，提出了一种信息一致性机制，以最大限度地保持源图像与增强图像之间的结构(视杯视盘、血管、病变)一致性。然而，眼底图像往往具有较高的分辨率要求，基于卷积神经网络(convolutional neural network，CNN)的模型在高分辨率图像重建方面存在一定限制，无法恢复图像中的细微结构，并且对噪声和伪影的鲁棒性较弱，导致重建结果模糊或伪结构。扩散模型(Sohl-Dickstein 等，2015)是一种使用随机过程结合神经网络的无监督学习方法，直到 2020 年，经典的去噪扩散概率模型(denoising diffusion probabilistic models，DDPM)(Ho 等，2020)才得以提出。迄今为止，扩散模型已广泛应用于各种生成建模任务，如图像生成(Rombach 等，2022)、图像超分辨率(Saharia 等，2022)和图像转换(Wang 等，2022)等，具有更好的鲁棒性，能够处理更复杂的图像，以及更可控的生成过程和更稳定的训练动态。而扩散模型通常结合 U-Net 网络进行训练，该类型网络首先将图像输入网络进行下采样，下采样过程会逐渐降低图像的空间分辨率，细小的特征可能在下采样过程中丢失或被模糊化，而眼底图像相比于自然图像具有更丰富的细节，所以该模型不能直接适用于眼底图像。

在图像重构问题中，卷积字典学习可以自动从数据中学习到适应的字典，自适应字典可以更好地捕捉数据中的特征和结构(王丽芳 等，2019)；在去噪任务中，稀疏表示可以有效地过滤噪声，使得重建的图像更接近于原始清晰图像。卷积字典学习在图像处理领域得到了广泛的应用，如图像超分辨率(Swapna 等，2015)、图像分类(Liang 和 Li，2016)和图像去噪、修复(焦丽娟 等，2017)等任务中，并取得了较好的性能。

针对上述问题，本文提出了基于卷积字典扩散模型的眼底图像增强算法。将卷积字典学习模块作为 U-Net 网络下采样后的中间层，以更有效地过滤噪声、捕捉图像结构信息并提取重要特征。该模型首先将条件信息和噪声图像拼接并输入网络进行下采样，再利用卷积神经网络实现卷积字典学习并获取图像的稀疏编码，最后，上采样至原图大小。模型通过最小化负对数似然损失，度量生成图像与原始图像之间的概率分布差异，逐渐调整生成图像的分布，使其逼近原始图像的真实分布，从而提高生成图像的质量和逼真度。模型训练完成后，能够在给定噪声图像和模糊图像输入的情况下，通过逐步将噪声从图像中移除来生成清晰眼底图像。

1 理论基础

1.1 卷积字典学习

卷积字典学习是一种基于卷积神经网络的字典学习方法，其目标是学习适合输入数据的卷积核和稀疏编码，以更好地进行信号重构。卷积形式的稀疏表示是用一组线性滤波器$\{\boldsymbol{d}^j\}_{j=1}^m$代替非结构化字典$\boldsymbol{D}$，通常将卷积字典学习问题表示为优化目标函数，具体为

$$\arg\min_{\{d_m\},\{z_{m,n}\}} \frac{1}{2}\sum_{i=1}^{n}\left\|\boldsymbol{x}_i-\sum_{j=1}^{m}\boldsymbol{d}^j*\boldsymbol{z}_i^j\right\|_2^2+\lambda\sum_{j=1}^{m}\left\|\boldsymbol{z}_i^j\right\|_1 \quad (1)$$
$$\text{s.t.}\left\|\boldsymbol{d}^j\right\|_2\leqslant 1, j=1,\cdots,m$$

式中，为了避免滤波器与系数之间的尺度模糊，需要对滤波器的范数$\boldsymbol{d}^j$进行约束。$*$代表卷积操作，$\boldsymbol{z}_i^j$为稀疏编码，$\|\cdot\|_1$为L1正则项，$\lambda\in\mathbf{R}^+$为稀疏加权系数，$\boldsymbol{x}_i\in\mathbf{R}^n$为训练图像，$m$和$n$分别表示滤波器和训练图像的数量。

式(1)中的稀疏编码$\boldsymbol{z}_i^j$和信号恢复问题可由迭代阈值收缩算法(iterative shrinkage-thresholding algorithm，ISTA)(Gregor和LeCun，2010)求解，具体为

$$\boldsymbol{z}_i^{j+1}=ST\left(\boldsymbol{z}_i^j+\alpha\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{D}\boldsymbol{z}_i^j),\alpha\lambda\right) \quad (2)$$

式中，步长参数$\alpha=1/\sigma_{\max}$，$\sigma_{\max}$表示矩阵的最大特征值$\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}$，$\boldsymbol{D}^{\mathrm{T}}$为字典$\boldsymbol{D}$的转置。将每次更新的结果进行阈值处理。常用的软阈值函数ST定义为

$$ST(\boldsymbol{x},\theta)=\text{sign}(\boldsymbol{x})\max(|\boldsymbol{x}|-\theta,0) \quad (3)$$

式中，$\text{sign}(\boldsymbol{x})$是$\boldsymbol{x}$的符号函数，$\theta=\alpha\lambda$，软阈值函数$ST$通过减去阈值并将负值截断为零来促进稀疏性。

1.2 去噪扩散概率模型(DDPM)

扩散模型是一种新兴的生成模型，与以往的生成模型在建模方式和采样方法等方面存在差异。它基于概率模型进行建模，使用条件概率分布来描述噪声图像和清晰图像之间的关系，通过最大化条件概率来生成图像。

1.2.1 正向扩散过程

在正向扩散过程中，如图2所示，从左向右，给定目标数据分布$\boldsymbol{x}_0\sim q(\boldsymbol{x})$，定义一个马尔可夫过程$q$，逐渐向目标数据中添加噪声得到$(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_T)$，经过$T$次添加，最终数据分布$\boldsymbol{x}_T$变成一个各项独立的高斯分布，具体为

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})=N\left(\boldsymbol{x}_t;\sqrt{1-\beta_t}\,\boldsymbol{x}_{t-1},\beta_t I\right) \quad (4)$$

式中，给定方差$\{\beta_t\in(0,1)\}_{t=1}^t$，均值是由固定值$\beta_t$和数据$\boldsymbol{x}_{t-1}$决定的，$I$代表标准正态分布的标准差，$I=1$。

从标准分布$N(0,I)$采样出z，由参数重整化$\boldsymbol{x}=\mu+\sigma*z$，得到逐步加噪公式，具体为

$$\boldsymbol{x}_t=\sqrt{\alpha_t}\,\boldsymbol{x}_{t-1}+\sqrt{1-\alpha_t}\,z_{t-1} \quad (5)$$

式中，$\alpha_t=1-\beta_t$，$\boldsymbol{x}_t$可表示为$\boldsymbol{x}_0$和噪声变量$\boldsymbol{z}_t$的线性组合。具体为

$$\boldsymbol{x}_t=\sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0+\sqrt{1-\bar{\alpha}_t}\,\tilde{z}_t \quad (6)$$

式中，$\bar{\alpha}_t=\prod_{t=1}^{T}\alpha_t$，$\tilde{z}_t\sim N(0,I)$。即任意时刻的$q(\boldsymbol{x}_t)$可以通过给定的$\boldsymbol{x}_0$和$\beta_t$计算得出。

假设给定$(\boldsymbol{x}_t,\boldsymbol{x}_0)$，得到的$\boldsymbol{x}_{t-1}$后验分布为

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)=N\left(\boldsymbol{x}_{t-1};\tilde{\mu}_t(\boldsymbol{x}_t,\boldsymbol{x}_0),\tilde{\beta}_t I\right) \quad (7)$$

式中，均值$\tilde{\mu}_t(\boldsymbol{x}_t,\boldsymbol{x}_0)$和方差$\tilde{\beta}_t$可由贝叶斯公式及代数完全平方得到，分别为

$$\tilde{\mu}_t=\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}z_t\right)$$
$$\tilde{\beta}_t=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\times\beta_t \quad (8)$$

即在给定$\boldsymbol{x}_0$的情况下，后验高斯分布的均值只与t时刻的随机正态分布$\boldsymbol{z}_t$有关，这为之后的反向去噪过程提供了基础。

1.2.2 反向去噪过程

反向去噪过程是从高斯噪声中恢复高质量数据，如图2所示，从右向左，给定噪声图$\boldsymbol{x}_t$，通过逐步去噪，直至还原出清晰图像$\boldsymbol{x}_0$。首先定义一个数据分布为高斯分布$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$的反向马尔可夫过程，具体为

$$p_\theta(\boldsymbol{x}_{t-1},\boldsymbol{x}_t)=N\left(\boldsymbol{x}_{t-1};\mu_\theta(\boldsymbol{x}_t,t),\sigma_t^2 I\right) \quad (9)$$

式中，$\sigma_t^2=\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$，参考式(8)，$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$的均值

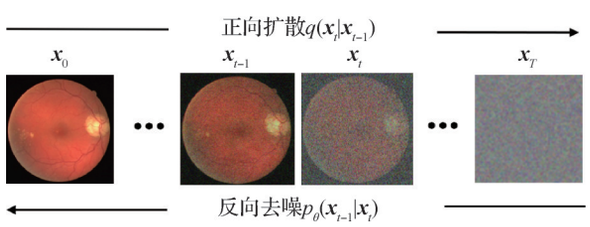


图2 扩散框架

Fig. 2 Diffusion structure

可以表示为

$$\mu_\theta = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} z_\theta(\boldsymbol{x}_t, t)\right) \tag{10}$$

式中，z_θ为模型预测的高斯噪声，θ是模型的参数。

最后，通过参数重整化，得到模型迭代的表达式，具体为

$$\boldsymbol{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} z_\theta(\boldsymbol{x}_t, t)\right) + \sigma_t z \tag{11}$$

2 模型框架

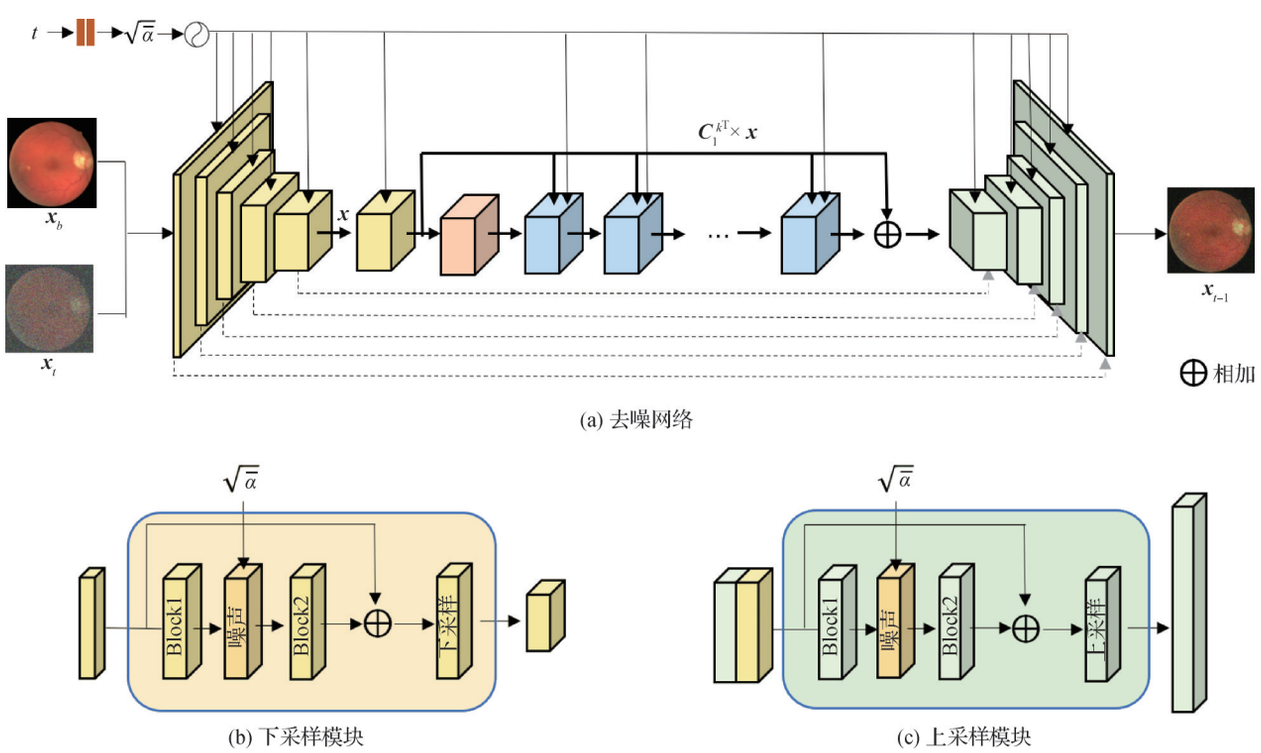
本文首次将卷积字典学习与条件扩散模型进行结合，应用到眼底图像增强领域。将卷积字典学习模块作为U-Net网络下采样后的中间层，以更有效地过滤噪声、捕捉图像结构信息并提取重要特征。同时将模糊眼底图像$\boldsymbol{x}_b$作为先验条件信息和噪声图$\boldsymbol{x}_t$拼接，再输入网络进行下采样，从而缓解去噪过程中的不确定性，引导模型能够更精确地恢复图像。此外，为避免训练过程中网络层数增加而导致训练困难和参数量剧增的问题，本文采用Dhariwal和Nichol(2021)方法中基于Wide-Resnet(Zagoruyko和Komodakis，2017)的一种减小残差网络深度、增大残差网络宽度的结构作为模型的上下采样模块。最后通过上采样模块得到重建图像$\boldsymbol{x}_{t-1}$。条件去噪扩散概率模型(conditional denoising diffusion probabilistic models，CDDPM)如图3(a)所示，其中包括上采样模块、卷积字典学习模块和下采样模块。模型的网络结构将在2.1节具体介绍。

此外，受限于在真实的眼底图像数据集中，高质量图像通常占主导地位，而低质量图像的数量相对有限，且目前广泛使用人工合成的眼底数据集。所以，本文利用高斯滤波模拟散焦模糊生成模糊图像$\boldsymbol{x}_b$。眼底图像模糊退化模型在2.2节具体介绍。

2.1 条件去噪扩散概率模型CDDPM

1)通过下采样不断缩小图像尺寸来提取图像的抽象特征，并且减少计算稀疏表示消耗的时间。下采样块结构如图3(b)所示。

2)使用卷积字典学习获取图像的稀疏编码，本文将其过程展开为一个递归的卷积神经网络，利用卷积算子代替矩阵乘法，卷积滤波器当做字典，提取图像的特征信息。可表示为



(a) 去噪网络

(b) 下采样模块

(c) 上采样模块

图3 CDDPM框架

Fig. 3 CDDPM framework ((a) denoising network; (b) downsampling module; (c) upsampling module)

$$\begin{cases} z^0 = 0, k = 1, 2, \cdots, K-1 \\ z^{k+1} = ST\left(z^k + \boldsymbol{C}_1^{k^{\mathrm{T}}} * (\boldsymbol{x} - \boldsymbol{C}_2^k * z^k), v^k\right) \\ z^{\mathrm{ISTA}} = z^K \\ \hat{\boldsymbol{x}} = \boldsymbol{D} * z^{\mathrm{ISTA}} \end{cases} \tag{12}$$

式中，z^k 为稀疏编码，$*$ 代表卷积操作，$\boldsymbol{O} = \{[\boldsymbol{C}_1^{k^{\mathrm{T}}}, \boldsymbol{C}_2^k, v^k]_{k=0}^{K-1}, D\}$ 为可学习参数集合，$\boldsymbol{C}_2^k$ 和 D 为卷积合成算子，$\boldsymbol{C}_1^{k^{\mathrm{T}}}$ 为卷积分析算子（灰度或彩色图像），非负阈值 $v^k \in \mathbf{R}^M$ 对应式（1）中权重 λ，Shrinkage 激活函数表示为软阈值函数 ST。$\hat{\boldsymbol{x}}$ 为重构图像。迭代过程如图 4 所示。

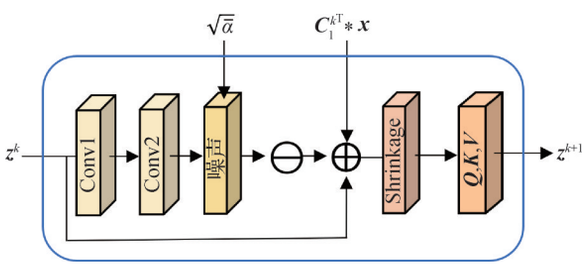


图 4　稀疏编码模块

Fig. 4　Sparse coding module

图 4 中，Conv1 为 $\boldsymbol{C}_1^{k^{\mathrm{T}}}$，Conv2 为 $\boldsymbol{C}_2^k$，卷积字典学习的目标函数由式（1）可表示为

$$L = \frac{1}{2}\left\| \boldsymbol{x} - \hat{\boldsymbol{x}} \right\|_2^2 \tag{13}$$

参数集合 $\boldsymbol{O}$ 通过最小化式（13）不断更新。

由于自注意力（self-attention）可以捕捉非局部自相似性和长程依赖关系，所以本文将自注意力添加到卷积字典学习模块，以提高重建质量。具体为

$$f_{\mathrm{Attention}}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = f_{\mathrm{softmax}}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{14}$$

式中，$\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbf{R}^{(H \times W) \times C}$ 由输入特征图 $\boldsymbol{X}$ 与可训练参数矩阵：$\boldsymbol{W}^Q$、$\boldsymbol{W}^K$ 和 $\boldsymbol{W}^V$ 分别相乘得到，H, W, C 分别代表输入特征图的高、宽和通道数，d_k 为张量的维度大小。

3）使用特征拼接实现特征融合，促进不同层次间的信息传递，更有效地利用图像中的局部特征。同时，通过反卷积层将下采样的特征映射还原到原始图像尺寸，得到输出图像 $\boldsymbol{x}_{t-1}$。上采样模块结构如图 3（c）所示。

训练过程中，扩散时间 T 通过 Transformer 在每个残差块中进行位置嵌入。本文设置时间步长 T = 3 000，通过位置编码生成带有位置信息的编码张量，具体为

$$PE_n^{(k)} = \begin{cases} \sin(w_i n) & k = 2i \\ \cos(w_i n) & k = 2i+1 \end{cases} \tag{15}$$

式中，$w_i = \dfrac{1}{10\,000^{2i/d}}$，$i = 0, 1, \cdots, \left(\dfrac{d}{2} - 1\right)$，$\boldsymbol{PE}_n \in \mathbf{R}^d$ 表示位置向量，$PE_n^{(k)}$ 表示位置向量里的第 k 个元素。将其嵌入模块中，有助于模型更好地理解输入数据中的位置关系和顺序信息。

此外，参考 Nichol 和 Dhariwal（2021）所提出的基于余弦的方差调度，具体为

$$\begin{aligned} \bar{\alpha}_t &= \frac{f(t)}{f(0)} \\ f(t) &= \cos\left(\frac{\dfrac{t}{T} + s}{1+s} \times \frac{\pi}{2}\right)^2 \end{aligned} \tag{16}$$

式中，$T = (t_1, t_2, \cdots, t_T)$ 为迭代次数，s = 0.008，以帮助扩散模型获得较低的非线性规划。

由于原始模型反向去噪过程中需要 1 k～2 k 次迭代，导致生成速度较为缓慢。所以，本文不再将 t 直接输入到模型，而是使用一种分层抽样方法来模仿离散抽样策略（Chen 等，2020）。定义一个 T = 100 的噪声调度，在分割片段 $(\sqrt{\bar{\alpha}_{t-1}}, \sqrt{\bar{\alpha}_t})$ 中均匀采样得到噪声水平 $\sqrt{\bar{\alpha}}$ 作为输入，仅通过 100 次迭代便完成了去噪。

2.2　基于高斯滤波的散焦模糊退化模型

在眼底成像过程中，设备的自动对焦功能可能由于不稳定或误差而未能准确对焦，从而导致眼底图像出现部分或完全模糊的情况。为了模拟这一现象，本文参考 Shen 等人（2021）的退化方法，将散焦模糊退化模型定义为

$$\boldsymbol{x}_{\mathrm{blur}} = \boldsymbol{x} * G_B\left(r_B + \sigma_B\right) + n \tag{17}$$

式中，$\boldsymbol{x}$ 表示原始眼底图像，$*$ 表示卷积操作，G_B 表示高斯滤波器，调整模糊半径 r_B 可以改变模糊的范围，σ_B 为空间常数，n 为噪声。本文将 σ_B 定义为 15 + 10 × $random$，r_B 定义为 $[\sigma_B/3, \sigma_B/2]$，$random$ 为一个随机数，取值为 0～1。模型结构如图 5 所示。

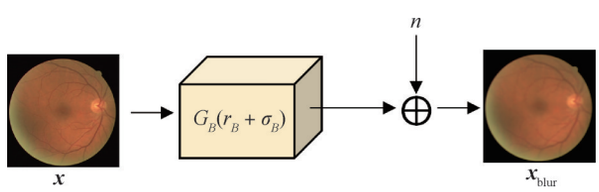


图 5　退化模型

Fig. 5　Degradation module

2.3 损失函数

由于DDPM是无条件概率扩散模型，其生成过程通过逐步从纯高斯噪声中去噪来生成清晰图像。因此，在去噪过程中，图像中的结构细节可能会发生变化，从而导致图像失真。所以本文将模糊眼底图像作为先验信息输入到模型中，由式(11)得

$$\boldsymbol{x}_{t-1}=\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}z_\theta\left(\boldsymbol{x}_b,\boldsymbol{x}_t,\sqrt{\bar{\alpha}}\right)\right)+\sigma_t z \tag{18}$$

通过神经网络学习来近似反向去噪过程中的条件概率分布$p_\theta\left(\boldsymbol{x}_{t-1}|\boldsymbol{x}_b,\boldsymbol{x}_t\right)$。由于方差是给定的，所以损失函数$L_\theta$旨在最小化$\mu_\theta$和$\tilde{\mu}_t$的差异，具体为

$$L_\theta=E_{\boldsymbol{x}_b,\boldsymbol{x}_0,z,\sqrt{\bar{\alpha}}}\left[\frac{1}{2\sigma_t^2}\left\|\tilde{\mu}_t\left(\boldsymbol{x}_t,\boldsymbol{x}_0\right)-\mu_\theta\left(\boldsymbol{x}_b,\boldsymbol{x}_t,\sqrt{\bar{\alpha}}\right)\right\|^2\right]=E_{\boldsymbol{x}_b,\boldsymbol{x}_0,z,\sqrt{\bar{\alpha}}}\left[\frac{{\beta_t}^2}{2\sigma_t^2\alpha_t\left(1-\bar{\alpha}_t\right)}\left\|z_t-z_\theta\left(\sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0+\sqrt{1-\bar{\alpha}_t}\,z_t,\boldsymbol{x}_b,\sqrt{\bar{\alpha}}\right)\right\|^2\right] \tag{19}$$

Ho等人(2020)提出去除系数会使得训练更稳定，所以本文将模型的目标损失函数定义为

$$L_\theta=E_{\boldsymbol{x}_b,\boldsymbol{x}_0,z,\sqrt{\bar{\alpha}}}\left[\left\|z_t-z_\theta\left(\sqrt{\bar{\alpha}_t}\,\boldsymbol{x}_0+\sqrt{1-\bar{\alpha}_t}\,z_t,\boldsymbol{x}_b,\sqrt{\bar{\alpha}}\right)\right\|^2\right] \tag{20}$$

模型训练完成后，通过噪声水平$\sqrt{\bar{\alpha}}$、噪声图$\boldsymbol{x}_t$和模糊眼底图像$\boldsymbol{x}_b$预测出正向过程中所加噪声，代入式(18)得到上一时刻噪声图$\boldsymbol{x}_{t-1}$，然后通过迭代去噪以生成清晰图像$\boldsymbol{x}_0$。

3 实验与分析

3.1 实验环境及数据集

为保证实验结果，本文所有实验均在统一实验环境中进行，环境配置参数如表1所示。

在模型训练前，首先进行数据预处理，将图像的尺寸统一处理为512×512像素。在训练期间，该模型使用Adam优化器进行优化，学习率初始化为3×10^{-6}，并进行了2 M次迭代训练。由于训练时图像的显存较大，批处理大小较小，因此采用组归一化代替批归一化，以获得更好的性能。

本文从EyePACS(Cuadros和Bresnick，2009)糖尿病视网膜病变检测数据集中选择8 400幅眼底图像，作为训练集中的高质量图像部分，并将其使用眼底图像退化模型合成了8 400幅模糊眼底图像，组成训练集的低质量图像部分；另外，还选取了50幅眼底图像，同样合成了50幅模糊眼底图像作为测试集。此外，为验证模型的泛化性能，分别在不同数据集上进行泛化性实验，包括使用相同退化模型合成的DRIVE(dgital retinal images for vessel extraction)(Staal等，2004)、CHASEDB1(child heart and health study in England)(Fraz等，2012)、ROC(retinopathy online challenge)(Niemeijer等，2010)、RF(real fundus)(Deng等，2022)和HRF(high-resolution fundus)(Köhler等，2013)数据集。其中，DRIVE数据集含有40幅768×584像素分辨率的眼底图像，训练集和测试集均为20幅。CHASEDB1数据集含有28幅999×960像素的彩色高清眼底图像。DRIVE数据集和CHASEDB1数据集中的每幅眼底图像均有对应的视网膜血管分割图，可用于后续的视网膜血管分割实验。ROC数据集包含了100幅具有参考价值的视网膜图像，由国际公认的视网膜专家提供。其中训练集和测试集均为50幅。同时，为了探索模型在临床应用中的潜力，在真实临床模糊眼底图像RF和HRF数据集上实验。其中，RF数据集由120组2 560×2 560像素分辨率的高质量和低质量的临床眼底图像对组成。HRF数据集使用佳能CR-1眼底相机，从18名受试者的同一只眼睛上采集了18对图像。每一对图像包括一幅由于相机失焦造成清晰度下降的模糊图像和一幅高清图像。这两幅图像共享大致相同的视野，唯一的差异是在两次获取之间眼球微小的运动引起的变化。

表1 实验环境

Table 1 Experiment environment

实验环境	环境配置
操作系统	Centos7
处理器	Intel (R) Core (TM) i7-6700 CPU @ 3.40 GHz
编程语言	Python3.6.9
开发工具	Pycharm11.0.11
深度学习框架	PyTorch1.9.0
CUDA	11.1
显卡型号	NVIDA GeForce RTX 3090

3.2 评价指标

图像质量评价指标是评价重建图像质量的重要标准，本文选取传统像素级指标峰值信噪比（peak signal-to-noise ratio，PSNR）来评估图像重建质量。PSNR是一种衡量重构图像与原始图像之间相似程度的指标，通过比较像素值之间的均方差来评估图像的质量。同时，为了更好地模拟人眼对图像的感知，衡量图像之间的感知差异，选取LPIPS（learned perceptual image patch similarity）（Zhang等，2018）作为视觉感知指标。LPIPS使用深度学习模型来学习图像感知的特征表示，并计算图像之间的相似性，更加注重图像的语义内容，能够更好地衡量图像之间的视觉差异。LPIPS的计算可以表示为

$$f_{\mathrm{LPIPS}}(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=1}^{l} w_i \times d_i(x,y) \tag{21}$$

式中，原始图像$\boldsymbol{X}$与重建图像$\boldsymbol{Y}$经过卷积神经网络的特征提取，对每一层输出的特征进行激活后归一化处理。l表示特征层数，$d_i(x,y)$表示在第i层特征空间中的特征距离，w_i是对应层的权重；最后将各层级的距离加权求和得到LPIPS值。LPIPS的取值为0～1，值越小，表示两幅图像之间的感知差异越小。

3.3 对比方法

1）CLAHE（Setiawan等，2013）。为了降低彩色视网膜图像在采集过程中产生的噪声效应，在红色、绿色和蓝色通道中采用直方图均衡化增强的方法来改善彩色视网膜图像质量。

2）Cofe-Net（Shen等，2021）。使用卷积神经网络构造保留眼底结构模块和去除低质量因素模块，增加伪影感知的误差度量，精确校正图像。

3）Arc-Net（Li等，2022）。考虑到合成数据与真实数据的差距，引入了域自适应算法，进一步提升了基于合成数据的眼底图像增强模型的性能。

4）Cycle-GAN（Wan等，2022）。在未配对的数据中学习从低质量域到高质量域的合适映射，从而实现低质量图像增强。

5）T-Cycle-GAN（Alimanov 和 Islam，2022）。基于转换器和生成对抗网络的眼底图像增强方法，将转换器代替传统CNN编码器，收敛速度更快。

6）Led（Cheng等，2023）。基于扩散模型的框架，从数据驱动的退化模型中学习高低质量图像的映射，用于眼底图像增强。

3.4 实验结果与分析

3.4.1 图像去模糊结果

本文以PSNR和LPIPS作为评价指标，分别在6个数据集上与6种眼底眼底增强方法进行了定量比较，包括基于传统方法的CLAHE（contrast limited adaptive histogram equalization），2种基于CNN的方法Cofe-Net（clinically oriented fundus enhancement network）和Arc-Net（annotation-free restoration network），2种基于GAN（generative adversarial network）的方法Cycle-GAN和T-Cycle-GAN，以及1种基于DDPM的方法Led（learning enhancement from degradation）。实验结果如表2所示，本文算法在6个数据集上的平均PSNR和LPIPS指标均达到了最优值，分别为29.671 0 dB和0.101 5。其中，CLAHE方法在6个数据集的PSNR和LPIPS指标均为最差，平均PSNR和LPIPS分别为20.920 0 dB和0.295 1。这是因为直方图均衡化无法很好地保持图像的局部对比度，导致图像质量不佳。而Cofe-Net和Arc-Net方法使用卷积神经网络可以更精确地校正图像，其LPIPS指标明显有所提升。Cycle-GAN方法使用更复杂的循环生成对抗网络模型，较Cofe-Net和Arc-Net方法，各项指标均有所提升，平均PSNR和LPIPS分别为28.488 0 dB和0.132 6。T-Cycle-GAN方法添加转换器作为编码器，平均PSNR和LPIPS指标均达到了次优值，分别为29.116 8 dB和0.106 3。LED方法的各项指标优于Cofe-Net和Arc-Net方法，但略逊于Cycle-GAN和T-Cycle-GAN方法。平均PSNR和LPIPS分别为27.678 1 dB和0.130 4。本文算法仅在RF数据集上的表现略逊于CycleGAN方法，但在其余5个数据集上的PSNR和LPIPS均优于CycleGAN方法。实验结果表明，本文算法相较于其他方法在图像增强任务上具有更优异的性能及跨数据集泛化能力。

接着，为了评估不同眼底图像增强方法的鲁棒性，本文对其在不同数据集上的PSNR和LPIPS评价指标结果进行方差计算。如表3所示，其中，$\sigma^2(P)$和$\sigma^2(L)$分别代表PSNR和LPIPS的方差，方差越小表示方法的鲁棒性越强。由表3可知，本文算法仅在CHASEDB1数据集上的$\sigma^2(P)$略高于Led方法，但在整体平均方差上取得了最优值；同时，在所有数据集上的$\sigma^2(L)$均为最小。实验数据表明，本文算法在眼底图像增强任务中表现出比其他方法更强的鲁棒性。

表2　眼底图像增强算法对比结果

Table 2　Comparison results of fundus image enhancement algorithms

方法	EyePACS		DRIVE		CHASEDB1		RF		HRF		ROC	
	PSNR/dB	LPIPS	PSNR/dB	LPIPS	PSNR/dB	LPIPS	PSNR/dB	LPIPS	PSNR/dB	LPIPS	PSNR/dB	LPIPS
CLAHE	21.961 9	0.303 6	20.986 4	0.255 1	20.607 5	0.284 3	21.096 6	0.385 2	21.046 6	0.267 0	19.820 9	0.275 1
Cofe-Net	25.476 0	0.088 8	21.334 3	0.145 6	21.274 9	0.197 2	24.952 8	0.222 1	18.745 7	0.145 9	20.581 4	0.178 4
Arc-Net	20.809 5	0.115 1	21.540 4	0.121 6	21.018 5	0.129 3	22.510 0	0.183 5	22.854 3	0.134 8	20.571 7	0.135 5
Cycle-GAN	30.387 2	0.105 4	27.928 1	0.125 0	32.343 8	0.128 0	**26.785 3**	0.180 0	24.134 6	0.121 8	29.349 1	0.135 6
T-Cycle-GAN	30.400 5	0.096 8	29.690 9	0.102 4	32.366 8	0.081 1	26.137 4	0.181 4	26.049 0	0.085 2	30.056 1	0.090 8
Led	26.464 5	0.121 7	28.810 7	0.110 8	30.003 8	0.106 6	26.378 4	0.182 6	26.236 4	0.142 5	28.174 5	0.118 4
本文	**31.802 3**	**0.082 7**	**30.153 3**	**0.075 7**	**32.374 4**	**0.077 0**	26.379 1	**0.175 5**	**27.208 4**	**0.119 7**	**30.108 3**	**0.078 5**

注：加粗字体表示各列最优结果。

表3　眼底图像增强算法在不同数据集上定量指标的方差

Table 3　Variance of quantitative metrics of fundus image enhancement algorithms on different datasets

方法	EyePACS		DRIVE		CHASEDB1		RF		HRF		ROC	
	$\sigma^2(P)$/dB^2	$\sigma^2(L)$	$\sigma^2(P)$/dB^2	$\sigma^2(L)$	$\sigma^2(P)$/dB^2	$\sigma^2(L)$	$\sigma^2(P)$/dB^2	$\sigma^2(L)$	$\sigma^2(P)$/dB^2	$\sigma^2(L)$	$\sigma^2(P)$/dB^2	$\sigma^2(L)$
CLAHE	9.962 2	0.001 6	10.986 4	0.007 1	6.687 4	0.006 6	8.005 4	0.002 9	8.390 1	0.007 0	9.820 9	0.002 7
Cofe-Net	11.253 5	0.000 8	18.764 7	0.001 8	4.681 5	0.001 8	5.229 4	0.003 2	10.235 4	0.001 5	7.392 7	0.001 8
Arc-Net	5.007 0	0.000 8	2.272 0	0.000 7	5.274 1	0.000 8	6.367 0	0.001 5	7.543 9	0.000 3	3.661 5	0.000 3
Cycle-GAN	6.699 5	0.000 7	2.387 2	0.001 2	4.112 2	0.002 3	7.608 6	0.001 7	10.263 0	0.001 5	4.384 1	0.001 9
T-Cycle-GAN	17.608 6	0.001 1	10.224 8	0.001 0	6.696 9	0.001 0	10.716 7	0.002 2	13.073 6	0.003 7	3.760 0	0.000 7
Led	3.175 0	0.001 3	2.442 6	0.002 3	**1.967 8**	0.002 7	5.269 2	0.001 5	12.107 9	0.001 1	3.101 6	0.001 5
本文	**3.080 3**	**0.000 2**	**1.560 8**	**0.000 5**	2.426 3	**0.000 7**	**4.998 6**	**0.001 4**	**6.783 0**	**0.000 9**	**3.020 4**	**0.000 1**

注：加粗字体表示各列最优结果。

此外，为描述各方法在合成图像和临床图像的定性比较，分别选取了2幅合成眼底图像和临床眼底图像，如图6和图7所示，每组图中上图为不同方法增强后的眼底图像，下图为局部细节放大图。由图6和图7可知，CLAHE、Cofe-Net和Arc-Net方法的结果过于平滑，无法恢复模糊眼底图像。Cycle-GAN方法虽然可以重建出更多的高频边缘细节，但是没有保持真实的眼底结构，存在明显的伪影，将会影响医生的诊断准确性。T-Cycle-GAN方法添加转换器后，伪影效果得到明显改善，但仍有较强的模糊感并伴随着噪点和结构不清晰。Led方法在去除模糊的同时，眼部特征结构也更加明显，但是并未恢复出细小的视网膜血管。相比之下，本文算法能够在不引入伪影的情况下恢复出更细粒度的内容和结构细节，不仅提高了临床诊断的准确性，同时为临床医学图像的分析与应用（如视网膜血管分割等任务）奠定了坚实的基础。

3.4.2　视网膜血管分割结果

增强眼底图像的最终目的是为了更好地服务于临床工作，提高医生诊断的准确性。为了验证不同增强方法的临床应用潜力，本文使用Liu等人（2022）的视网膜血管分割方法，在含有分割标签的DRIVE数据集上进行视网膜血管分割实验，并将受试者工作特征曲线下面积（area under the curve，AUC）、准确率（accuracy，Acc）和敏感性（sensitivity，Sen）作为分割精度指标，指标值越高表示分割结果越接近于分割标签。受试者工作特征（receiver operating characteristic，ROC）曲线图如图8所示，不同方法的分割结果如表4所示。本文算法的分割精度指标均取得了最优结果。

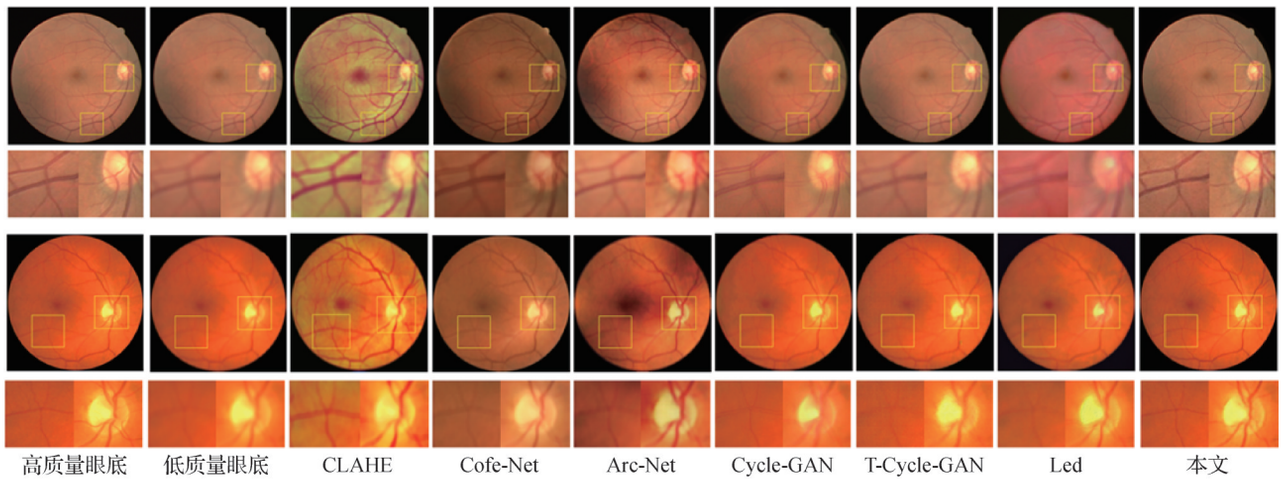


图 6　合成视网膜图像增强对比结果

Fig. 6　Comparison results of synthetic retinal image enhancement

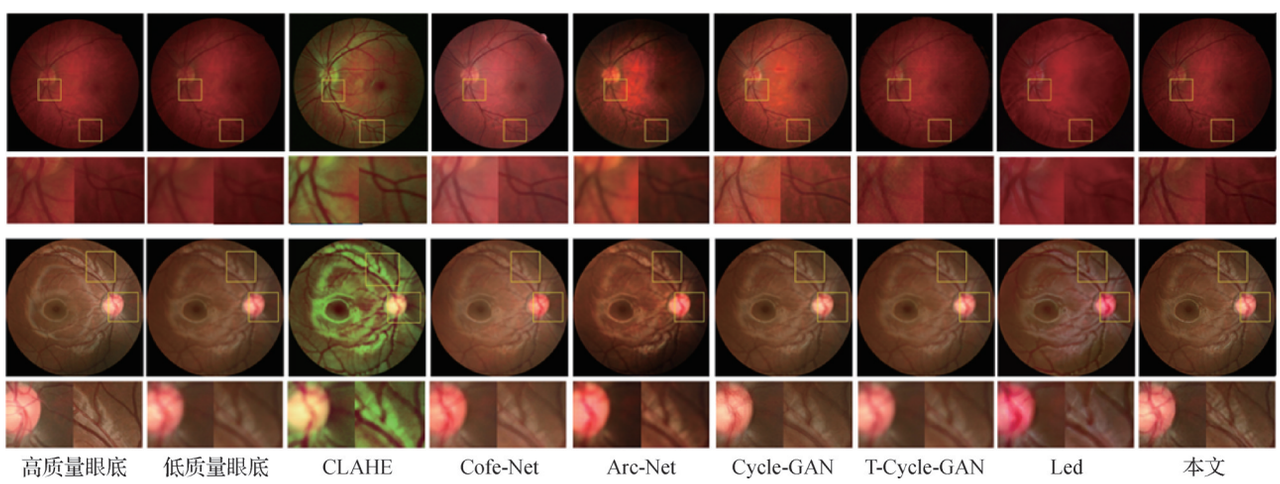


图 7　临床真实视网膜图像增强对比结果

Fig. 7　Comparison results of clinical real retinal image enhancement

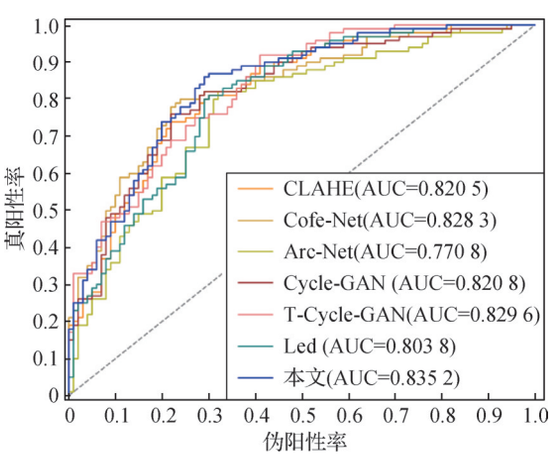


图 8　各方法受试者工作特征曲线图

Fig. 8　Receiver operating characteristic curves for each method

表 4　视网膜血管分割对比结果

Table 4　Comparison results of retinal vessel segmentation

方法	AUC	Acc	Sen
CLAHE	0.820 5	0.901 4	0.362 7
Cofe-Net	0.828 3	0.897 2	0.296 2
Arc-Net	0.770 8	0.896 1	0.159 8
Cycle-GAN	0.820 8	0.901 1	0.284 8
T-Cycle-GAN	0.829 6	0.906 1	0.307 2
Led	0.803 8	0.905 0	0.306 3
本文	**0.835 2**	**0.908 0**	**0.380 1**

注：加粗字体表示各列最优结果。

可视化对比结果如图 9 所示，从上至下依次为不同方法增强后的眼底图像、视网膜血管分割图和对应方法的局部细节放大图。由图 9 可知，CLAHE 方法分割的血管并不连续，这是因为直方图均衡化重新调整像素值的分布，可能导致一些细微的图像

特征丢失或模糊。Cofe-Net方法使用视网膜结构激活模块，分割出了较完整的血管结构。Arc-Net方法使用无监督域自适应来增强图像，但源域和目标域的图像分布可能存在模式不匹配的情况，从而无法准确地转换为目标域，导致其视神经头部区域血管结构分割并不完整。Cycle-GAN方法专注于图像不同域之间的转换，生成图像可能出现伪结构，导致分割出一些不存在的血管。T-Cycle-GAN方法在Cycle-GAN方法的基础上添加了注意力机制，能够更好地捕捉重要特征和信息。Led方法实现了视网膜血管的增强，但由于选用U-Net网络框架，导致对视神经头部区域血管恢复的仍不完整。通过对比，本文算法分割的血管与分割标签最为接近，不仅有效地去除了图像模糊，而且更完整地分割出了视神经头部区域复杂的血管特征。

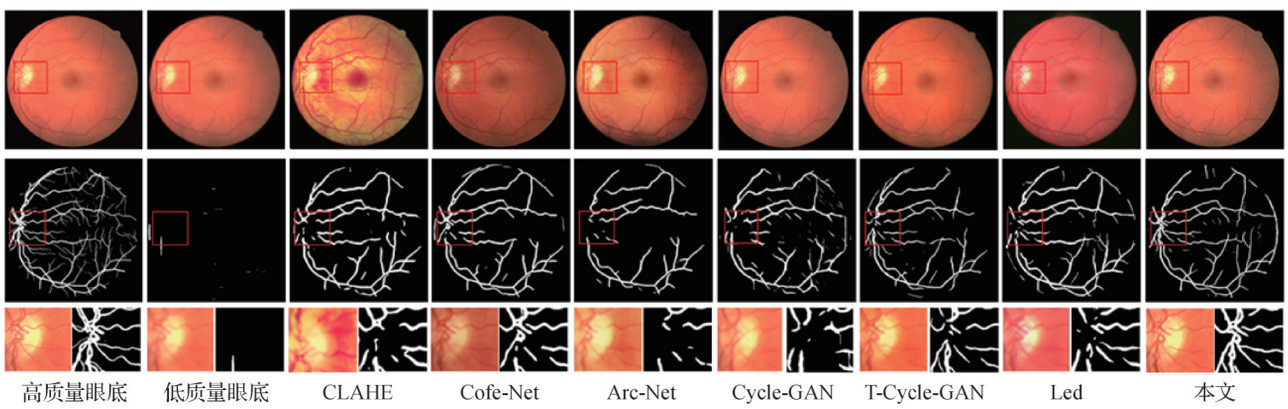


图9　不同方法的视网膜血管分割结果

Fig. 9　Results of retinal vessel segmentation with different methods

3.5　消融实验

本文模型通过上/下采样阶段更有效地捕获图像的多层次特征，提升模型对图像内容的理解能力。但随着采样模块数量的增加，模型的深度和参数数量也会增加，可能会导致内存不足或训练时间过长。为验证不同采样模块数量对模型性能的影响，分别将其设置为3、5、6和7，并记录每轮训练所需的时间以及前30轮的增强效果，如图10所示。综合考虑到模型性能、计算资源和运行时间，发现当模块数量设置为5时，模型的效率最佳。

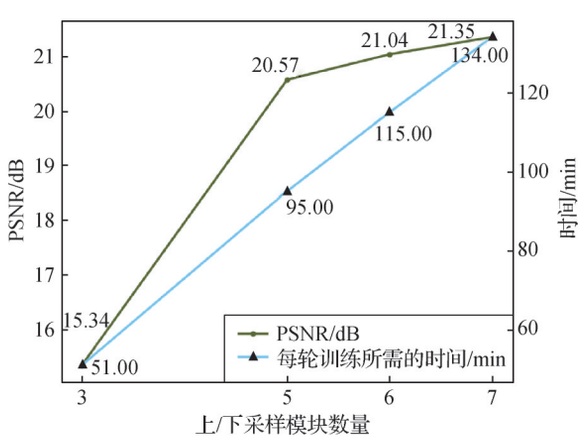


图10　不同模块数量对模型性能的影响

Fig. 10　Impact of different number of modules on model performance

此外，为了探索卷积字典学习模块、自注意力机制和条件扩散模型的引入对模型性能的影响。本文在EyePACS测试集上进行消融实验，实验结果如表5所示。

表5　消融实验

Table 5　Ablation experiments

扩散模型	卷积字典学习	自注意力	PSNR/dB	LPIPS
×	√	√	23.376 5	0.205 9
√	×	×	26.464 5	0.121 7
√	×	√	28.161 7	0.110 1
√	√	×	30.897 2	0.096 2
√	√	√	**31.802 3**	**0.082 7**

注：加粗字体表示各列最优结果，其中，“×”为不使用，“√”为使用。

4　结　论

针对现有眼底图像增强方法存在模糊和高频信

息缺失等问题，本文提出了一种基于卷积字典扩散模型的眼底图像增强算法。将卷积字典学习能够有效过滤噪声、捕捉图像结构信息和提取重要特征的优势与条件扩散模型强大的生成能力相结合，提高模型的数据表征能力和灵活适应性。本文与6种经典图像增强方法进行对比，在多个数据集上进行了测试和泛化性评估，并通过对实验结果分析和可视化分析，证明了所提算法具有优异的性能、跨数据集泛化能力和鲁棒性。此外，在视网膜血管分割实验中，分割指标AUC、Acc和Sen均取得了最优值，证明了所提算法具有较高的临床应用潜力。

然而，本文算法的模型复杂度相对较高，在计算速度上没有明显的优势，限制了该算法在实际应用中的可行性。因此，未来的研究方向将着重于开发高效的网络模型，如设计和采用轻量级深度学习架构，应用模型剪枝和轻量化技术，以降低模型的复杂度和参数数量，实现更高的图像处理速度。

参考文献（References）

Alimanov A and Islam M B. 2022. Retinal image restoration using Transformer and cycle-consistent generative adversarial network//2022 International Symposium on Intelligent Signal Processing and Communication Systems. Penang, Malaysia: IEEE: 1-4 [DOI: 10.1109/ispacs57703.2022.10082822]

Chen N X, Zhang Y, Zen H G, Weiss R J, Norouzi M and Chen W. 2020. WaveGrad: estimating gradients for waveform generation [EB/OL]. [2023-08-17]. https://arxiv.org/pdf/2009.00713.pdf

Cheng P J, Lin L, Huang Y J, He H Q, Luo W H and Tang X Y. 2023. Learning enhancement from degradation: a diffusion model for fundus image enhancement [EB/OL]. [2023-08-17]. https://arxiv.org/pdf/2303.04603.pdf

Cuadros J and Bresnick G. 2009. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. Journal of Diabetes Science and Technology, 3(3): 509-516 [DOI: 10.1177/193229680900300315]

Deng Z, Cai Y F, Chen L, Gong Z, Bao Q Q, Yao X, Fang D, Yang W M, Zhang S C and Ma L. 2022. RFormer: Transformer-based generative adversarial network for real fundus image restoration on a new clinical benchmark. IEEE Journal of Biomedical and Health Informatics, 26(9): 4645-4655 [DOI: 10.1109/jbhi.2022.3187103]

Dhariwal P and Nichol A. 2021. Diffusion models beat GANs on image synthesis [EB/OL]. [2023-08-17]. https://arxiv.org/pdf/2105.05233.pdf

Fraz M M, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka A R, Owen C G and Barman S A. 2012. An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Transactions on Biomedical Engineering, 59(9): 2538-2548 [DOI: 10.1109/tbme.2012.2205687]

Gaudio A, Smailagic A and Campilho A. 2020. Enhancement of retinal fundus images via pixel color amplification//Proceedings of the 17th International Conference on Image Analysis and Recognition. Póvoa de Varzim, Portugal: Springer: 299-312 [DOI: 10.1007/978-3-030-50516-5_26]

Gregor K and LeCun Y. 2010. Learning fast approximations of sparse coding//Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: Omnipress: 399-406

Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 6840-6851

Jiao L J, Wang W J, Zhao Q S and Cao J F. 2017. Image denoising based on sparse representation of neighbor local OMP. Jouranl of Image and Graphics, 22(11): 1486-1492 (焦丽娟, 王文剑, 赵青杉, 曹建芳. 2017. 近邻局部OMP稀疏表示图像去噪. 中国图象图形学报, 22(11): 1486-1492) [DOI: 10.11834/jig.170105]

Köhler T, Budai A, Kraus M F, Odstrcilik J, Michelson G and Hornegger J. 2013. Automatic no-reference quality assessment for retinal fundus images using vessel segmentation//The 26th IEEE International Symposium on Computer-Based Medical Systems. Porto, Portugal: IEEE: 95-100 [DOI: 10.1109/cbms.2013.6627771]

Li H, Liu H F, Hu Y, Fu H Z, Zhao Y T, Miao H P and Liu J. 2022. An annotation-free restoration network for cataractous fundus images. IEEE Transactions on Medical Imaging, 41(7): 1699-1710 [DOI: 10.1109/tmi.2022.3147854]

Liang H and Li Q. 2016. Hyperspectral imagery classification using sparse representations of convolutional neural network features. Remote Sensing, 8(2): #99 [DOI:10.3390/rs8020099]

Liu W T, Yang H H, Tian T, Cao Z W, Pan X P, Xu W J, Jin Y and Gao F. 2022. Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation. IEEE Journal of Biomedical and Health Informatics, 26(9): 4623-4634 [DOI: 10.1109/JBHI.2022.3188710]

MacGillivray T J, Cameron J R, Zhang Q L, El-Medany A, Mulholland C, Sheng Z Y, Dhillon B, Doubal F N, Foster P J, Trucco E, Sudlow C and UK Biobank Eye and Vision Consortium. 2015. Suitability of UK biobank retinal images for automatic analysis of morphometric properties of the vasculature. PLoS ONE, 10(5): #e0127914 [DOI: 10.1371/journal.pone.0127914]

Nichol A and Dhariwal P. 2021. Improved denoising diffusion probabilistic models [EB/OL]. [2023-08-17]. https://arxiv.org/pdf/2102.09672.pdf

Niemeijer M, Van Ginneken B, Cree M J, Mizutani A, Quellec G, Sanchez C I, Zhang B, Hornero R, Lamard M, Muramatsu C, Wu X Q, Cazuguel G, You J, Mayo A, Li Q, Hatanaka Y, Cochener B,

Roux C, Karray F, Garcia M, Fujita H and Abramoff M D. 2010. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. IEEE Transactions on Medical Imaging, 29(1): 185-195 [DOI: 10.1109/tmi. 2009. 2033909]

Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 10674-10685 [DOI: 10.1109/cvpr52688.2022.01042]

Saharia C, Ho J, Chan W, William, Salimans T, Fleet D J and Norouzi M. 2022. Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 4713-4726 [DOI:10.1109/TPAMI.2022.3204461]

Schmidt-Erfurth U, Sadeghipour A, Gerendas B S, Waldstein S M and Bogunović H. 2018. Artificial intelligence in retina. Progress in Retinal and Eye Research, 67: 1-29 [DOI: 10.1016/j.preteyeres. 2018.07.004]

Setiawan A W, Mengko T R, Santoso O S and Suksmono A B. 2013. Color retinal image enhancement using CLAHE//Proceedings of 2013 IEEE International Conference on ICT for Smart Society. Jakarta, Indonesia: IEEE: 1-3 [DOI: 10.1109/ictss. 2013. 6588092]

Shen Z Y, Fu H Z, Shen J B and Shao L. 2021. Modeling and enhancing low-quality retinal fundus images. IEEE Transactions on Medical Imaging, 40(3): 996-1006 [DOI: 10.1109/tmi.2020.3043495]

Sohl-Dickstein J, Weiss E A, Maheswaranathan N and Ganguli S. 2015. Deep unsupervised learning using nonequilibrium thermodynamic [EB/OL]. [2023-08-17]. https://arxiv.org/pdf/1503.03585.pdf

Staal J, Abrùmoff M D, Niemeijer M, Viergever M A and Van G B. 2004. Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging, 23(4): 501-509 [DOI: 10.1109/tmi.2004.825627]

Swapna T R, Indu D and Chakraborty C. 2015. Macular region enhancement of fundus fluorescein angiogram images using super resolution via sparse representation and quality analysis. Procedia Computer Science, 58: 586-592 [DOI: 10.1016/j.procs.2015.08.077]

Wan C, Zhou X T, You Q J, Sun J, Shen J X, Zhu S J, Jiang Q and Yang W H. 2022. Retinal image enhancement using cycle-constraint adversarial network. Frontiers in Medicine, 8: #793726 [DOI: 10.3389/fmed.2021.793726]

Wang L F, Dou J L, Qin P L, Lin S Z, Gao Y and Zhang C C. 2019. Medical image fusion using double dictionary learning and adaptive PCNN. Journal of Image and Graphics, 24(9): 1588-1603 (王丽芳, 窦杰亮, 秦品乐, 蔺素珍, 高媛, 张程程. 2019. 双重字典学习与自适应PCNN相结合的医学图像融合. 中国图象图形学报, 24(9): 1588-1603) [DOI: 10.11834/jig.180667]

Wang W L, Bao J M, Zhou W G, Chen D D, Chen D, Yuan L and Li H Q. 2022. SinDiffusion: learning a diffusion model from a single natural image [EB/OL]. [2023-08-17]. https://arxiv.org/pdf/2211.12445.pdf

Zagoruyko S and Komodakis N. 2017. Wide residual networks [EB/OL]. [2023-08-17]. https://arxiv.org/pdf/1605.07146.pdf

Zamir S W, Arora A, Khan S, Hayat M, Khan F S, Yang M H and Shao L. 2020. Learning enriched features for real image restoration and enhancement//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 492-511 [DOI: 10. 1007/978-3-030-58595-2_30]

Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 586-595 [DOI: 10.1109/cvpr.2018.00068]

Zhu W H, Qiu P J, Farazi M, Nandakumar K, Dumitrascu O M and Wang Y L. 2023. Optimal transport guided unsupervised learning for enhancing low-quality retinal images [EB/OL]. [2023-08-17]. https://arxiv.org/pdf/2302.02991.pdf

作者简介

王珍，女，硕士研究生，主要研究方向为机器学习与模式识别。E-mail:980439577@qq.com

霍光磊，通信作者，男，技术副总监，主要研究方向为自动驾驶与机器学习。E-mail:hgl2124@hitqz.com

兰海，男，博士研究生，主要研究方向为机器学习与模式识别。E-mail:lanhai09@fjirsm.ac.cn

胡建民，男，教授，博士生导师，主要研究方向为眼底病、眼视光及低视力。E-mail:doctorhjm@163.com

魏宪，男，研究员，博士生导师，主要研究方向为机器学习与模式识别。E-mail:xian.wei@tum.de