arXiv:2302.03003v4 [eess.IV] 9 Apr 2023

OTRE：最优传输引导的无配对图像到图像转换与增强正则化的结合

Wenhui Zhu[1*]、Peijie Qiu[2*]、Oana M. Dumitrascu[3]、、Jacob M. Sobczak[3]、、Mohammad Farazi[1]、、Zhangsihao Yang[1]、、Keshav Nandakumar[1]、、YalinWang[1]

[1] 美国亚利桑那州立大学计算与增强智能学院[2] 美国密苏里州圣路易斯华盛顿大学麦凯利工程学院[3] 美国亚利桑那州斯科茨代尔梅奥诊所神经内科

摘要。非散瞳眼底彩色照相（CFP）因无需瞳孔扩张而广泛应用，但易受操作人员、系统缺陷或患者相关因素影响，导致图像质量不佳。准确的医学诊断和自动化分析都要求视网膜图像具有最佳质量。本文利用 *Optimal Transport (OT)* 理论，提出一种无配对图像到图像转换方案，将低质量视网膜彩色眼底照片映射为高质量对应图像。此外，为提高图像增强流程在临床实践中的灵活性、鲁棒性和适用性，我们对一种先进的基于模型的图像重建方法——去噪正则化进行了扩展，将最优传输引导的图像到图像转换网络学习到的先验融入其中。我们将其命名为 *regularization by enhancing (RE)*。我们在三个公开可用的视网膜图像数据集上验证了集成框架OTRE，评估增强后的图像质量及其在多项下游任务中的表现，包括糖尿病视网膜病变分级、血管分割和糖尿病病灶分割。实验结果表明，我们提出的框架优于一些先进的无监督方法和一种先进的有监督方法。

关键词： 眼底彩色照相 · 图像增强 · 最优传输 · 通过增强进行正则化 · 无监督学习。

1 引言

眼底彩色照相（彩色眼底照片）被广泛用于常规诊断各种眼部疾病。基于免散瞳彩色眼底照片，人们正在开发用于即时医疗疾病筛查的自动化分析。 [1], 此外，研究

* 两位作者对本文的贡献相同。我们的代码已公开发布于 https://github.com/Retinal-Research/REDOT。

旨在发掘彩色眼底照片筛查神经退行性疾病（如阿尔茨海默病）的潜力 [2]。无论是人工分析还是计算机辅助分析方法，都更适用于高质量视网膜彩色眼底照片。免散瞳视网膜彩色眼底照片对患者和医疗服务提供者都更为友好，但容易受到噪声影响，例如，由于光线传输干扰、失焦、瞳孔异常或人工操作不当 [3], ，可能出现阴影伪影和模糊，导致彩色眼底照片质量低下。彩色眼底照片退化，例如血管被遮挡，以及病变缺失或出现伪影，都会导致诊断解读不准确。将低质量视网膜彩色眼底照片增强为高质量对应图像，对许多下游任务至关重要，例如糖尿病视网膜病变（DR）、血管分割、糖尿病视网膜病变病灶分割等。Shen 等人 [3] 提出了面向临床的眼底增强网络（cofe-Net），其输入为由眼底退化模型合成的退化图像与无损高质量图像对。然而，现实中采集成对的含噪—无噪视网膜训练数据既困难又昂贵。Lehtinen 等人 [4] 将成对含噪—无噪图像放宽为在相同条件下获得的含噪图像对，并指出从期望上看，含噪数据趋近于无噪数据。Krull 等人 [5] 通过预测盲点像素周围的像素，将 [4] 扩展为一种自监督训练方案。这类自监督方法强烈假设，导致图像退化的“噪声”与像素/图像无关。然而，与自然图像不同，视网膜眼底图像中的噪声构成更为复杂，因此也更难建模。

无监督方法近来受到广泛关注。此类过程通常被建模为端到端的图像到图像转换任务。许多针对该任务的早期探索 [6,7,8,9] 都基于生成对抗网络（GANs），将源域 Y 映射到目标域 X。由于存在大量映射歧义，当输入与目标图像未配对时，这种映射会变得更具挑战性。为缩小庞大的搜索空间， [9,6] 通过任务特定正则化对生成对抗网络进行正则化，而 CycleGAN [8] 则提出了一种通用正则化方法，称为循环一致性。然而，计算开销大、病变遭到破坏以及引入不存在的血管结构等问题，限制了 CycleGAN 在视网膜眼底图像中的应用。为降低 CycleGAN 的计算复杂度，Wang 等人 [10] 提出了一种用于无监督图像去噪的最优传输引导生成对抗网络（OTTGAN），该方法仅使用一个生成器和一个判别器。尽管它在自然图像去噪方面取得了显著成果，但采用均方误差代价作为度量，且缺乏高质量的图像一致性，导致在我们的任务中血管和病变结构遭到破坏或过度篡改。

在此，我们提出了一种基于最优传输（OT）和去噪正则化 [11] 方法的集成式无监督端到端图像增强框架。我们提出的新颖OT公式最大限度地保留了增强图像与低质量图像之间的结构一致性（例如病变、血管结构和视盘），以防止对重要结构的过度篡改。为进一步提升该方法对不同分布图像的灵活性和鲁棒性，并增强其在缺乏充足训练数据的真实临床实践中的适用性 [12,11,13], ，我们采用所提出的通过增强进行正则化

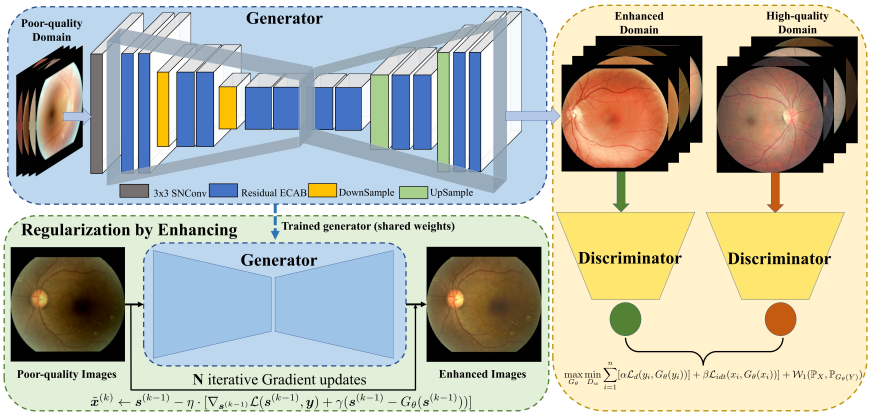


图1：我们提出的方法框架包括基于最优传输引导生成对抗网络的眼底增强网络和通过增强进行正则化的框架，其中 SNConv 表示谱归一化卷积层，Residual ECAB 表示带有高效通道注意力的残差块 [14]。判别器的架构采用自 [15]。

（RE）对增强图像进行细化。RE是去噪正则化（RED）方法的一种变体 [11]，其先验由最优传输引导网络学习。本文的贡献主要有三点：（1）我们提出了一种新颖的、基于最优传输引导生成对抗网络的无监督端到端视网膜图像增强训练方案，并采用最大信息保持一致性机制，防止对病变和结构的过度篡改。（2）我们引入了一个RE模块，用于细化最优传输模块的输出，从而提升系统的灵活性、鲁棒性和适用性。据我们所知，本研究首次弥合了最优传输引导生成模型与基于模型的增强框架之间的鸿沟。这是一种通用方法，可应用于其他结构保持型医学图像增强研究。（3）我们在三个大型视网膜成像队列上开展了大量实验，结果表明，我们提出的方法优于无监督方法和最先进的监督方法。

2 方法

从退化图像 $y \sim Y$ 中恢复无损图像 $x \sim X$ 可以在贝叶斯框架下表述为一种变分正则化

$$\hat{x} = \operatorname*{argmin}_x f(x) + R(x), \tag{1}$$

其中， f 是数据保真项，用于衡量复原结果与受损数据之间的一致性； R 是正则化/先验项。现代基于深度学习的图像复原旨在通过最小化经验风险$\mathbb{E}_{x,\ y}[\mathcal{L}(f_\theta(y)\ x)]$, 来训练端到端回归器，其中 f_θ 是由 θ参数化的神经网络， $\mathcal{L}$ 是损失函数。近期的研究进展 [12,11,13] 表明

将学习得到的图像回归器嵌入基于模型的复原框架能够提升其性能。本研究将基于模型的正则化与学习得到的图像复原回归器相结合，构建了一个灵活、稳健的框架，用于增强低质量视网膜眼底图像。我们的框架包含两个主要模块，如图1所示：1）一种由OT引导的无监督生成对抗网络学习方案，作为回归器将低质量图像增强，以实现公式1中的 f_θ ； 2）显式正则化项RE，即 $R(x)$，用于细化在第一个模块中得到的训练后生成器网络。这两个模块级联在一起。整个框架迭代运行，直至两个模块均收敛。

2.1 最优传输引导的无配对图像到图像转换

设 $\mu \sim \mathbb{P}_X$ 和 $\nu \sim \mathbb{P}_Y$ 分别是定义在目标域和源域概率流形上的两个概率测度。将质量从域 Y 传输到 X （$Y \to X$）的*Monge* 最优传输问题可定义为

$$\inf \int_Y C(y, T(y)) dv(y) \tag{2}$$

其中， C（$\cdot$， $\cdot$）是将 y 传输到 T（y）的代价。在所有可能的 ν-可测映射中， T 所对应的最小代价给出了最优传输 $u = T^*$（ν）。直观而言，式 2 中定义的传输与图像到图像转换的目标相一致：寻找从源域到目标域的最优映射，我们将其定义为域传输。我们将所提出的最优传输引导的图像到图像转换方案转化为一个优化问题。

定义 1. 由最优质量传输所建议的从源域到目标域$Y \to X$ 的图像到图像转换可表示为

$$\inf \int_Y C(y, T(y)) dv(y), \quad \textbf{\textit{subject to}} \quad u = T^*(\nu) \tag{3}$$

进一步将最优传输映射 T 参数化为神经网络 T_θ后，公式 3 可离散化为

$$\min_\theta \mathbb{E}_{y \sim \mathbb{P}_Y}[C(y, T_\theta(y))], \quad \textbf{subject to} \quad \mathbb{P}_{T_\theta(Y)} = \mathbb{P}_X \tag{4}$$

通过应用拉格朗日乘子，公式 4 可放宽为以下约束优化问题

$$\min_\theta \mathbb{E}_{y \sim \mathbb{P}_Y}[C(y, T_\theta(y))] + \lambda d(\mathbb{P}_{T_\theta(Y)}, \mathbb{P}_X), \tag{5}$$

同样，将目标域中的某个测量值进行传输，也会在目标域中产生另一个测量值 X，然而，我们并不希望目标域中的测量值之间存在差异。为防止网络过度学习或生成意外的测量值，我们引入了恒等代价约束。同时，该约束还用于保持目标域中的一致性。将这一项添加到公式 5 中，可表示为：

$$\min_\theta \mathbb{E}_{y \sim \mathbb{P}_Y}[C(y, T_\theta(y))] + \mathbb{E}_{x \sim \mathbb{P}_X}[C(x, T_\theta(x))] + \lambda d(\mathbb{P}_{T_\theta(Y)}, \mathbb{P}_X), \tag{6}$$

算法 1 最优传输引导的无配对图像到图像转换。

要求： 学习率 η、批量大小 m、梯度惩罚权重 λ，以及
一致性损失权重 $\alpha \leq 1$、恒等损失权重 β。
要求： 判别器初始参数 w_0、生成器初始参数 θ_0。
当未收敛时 **执行**
使用 $\{g_i\}_{i=1}^m$采样一批低质量图像 $\boldsymbol{y} = \{y_i\}_{i=1}^m \sim \mathbb{P}_Y$。
使用 $\{g_i\}_{i=1}^m$采样一批高质量图像 $\boldsymbol{x} = \{x_i\}_{i=1}^m \sim \mathbb{P}_X$。
对于 $i = 1, ..., m$，**执行**
随机采样一个 $\epsilon \sim U[0, 1]$。
$\tilde{x}_i \leftarrow G_\theta(y_i)$
$\hat{x}_i \leftarrow \epsilon x_i + (1 - \epsilon)\tilde{x}_i$
$\mathcal{L}_{D_w}(i) \leftarrow D_w(\tilde{x}_i) - D_w(x_i) + \lambda(||\nabla_{\hat{x}_i} D_w(\hat{x}_i)||_2 - 1)_+^2$
结束 for 循环
$w \leftarrow w + \eta \cdot$ 均方根传播算法$(w, \nabla_w \frac{1}{m} \sum_{i=1}^m \mathcal{L}_{D_w}(i))$
$\mathcal{L}_{G_\theta} \leftarrow \frac{1}{m} \sum_{i=1}^m -D_w(G_\theta(\boldsymbol{y})) + \alpha \mathcal{L}_d(\boldsymbol{y}, G_\theta(\boldsymbol{y})) + \beta \mathcal{L}_{idt}(\boldsymbol{x}, G_\theta(\boldsymbol{x}))$
$\theta \leftarrow \theta - \eta \cdot \text{RMSProp}(\theta, \nabla_\theta \mathcal{L}_{G_\theta})$
结束 while 循环

该项定义为恒等约束。其中， $d(\cdot, \cdot)$衡量$\mathbb{P}_X$ 与$\mathbb{P}_{T_\theta(Y)}$之间的散度， λ 为权重参数。值得注意的是，我们使用相同的代价表示方式，但恒等项作为目标域中的约束 X，与源域和目标域之间的最优传输无关。

命题1。 假设使用 *Wasserstein-1*距离 $\mathcal{W}_1(\cdot, \cdot)$ 来衡量 $\mathbb{P}_X$ 与 $\mathbb{P}_{T_\theta(Y)}$之间的散度，式*6*给出了从 $Y \to X$进行无配对图像到图像转换的对抗训练方案，如下所示：

$$\begin{aligned} &\max_{G_\theta} \min_{D_w} \mathbb{E}_Y[\mathcal{L}_d(y, G_\theta(y))] + \mathbb{E}_X[\mathcal{L}_{idt}(x, G_\theta(x))] + \lambda \mathcal{W}_1(\mathbb{P}_X, \mathbb{P}_{G_\theta(Y)}) \\ &\mathcal{W}_1(\mathbb{P}_X, \mathbb{P}_{G_\theta(Y)}) = \sup_{||D_w||_L \leq 1} \mathbb{E}_X[D_w(x)] - \mathbb{E}_Y[D_w(G_\theta(y))] \end{aligned} \tag{7}$$

其中， G_θ 是由 θ参数化的生成器； D_w是由 w参数化的判别器，是一个*1-Lipschitz*函数； $\mathcal{L}_d$ 和 $\mathcal{L}_{idt}$ 分别表示域传输成本和恒等约束成本。为更好地保留重要结构（例如病变），我们确保无配对输入具有匹配的疾病标签 g_i（如果存在），其中 g 表示疾病类型。

在我们的实验中，通过梯度惩罚 [16] 来实现*1-Lipschitz* 约束。域 传输和恒等约束采用相同的成本函数，详情如下。

信息保持一致性机制在我们的任务中，所提出的最优传输引导的无配对图像到图像转换主要关注两点：1）确保转换前后底层信息（例如视盘、病变和血管）保持一致；2）尽可能缩小原问题（公式 4）与对偶问题（公式 5）之间的对偶间隙。我们

将介绍我们围绕这两个主要关注点设计的信息保持一致性机制。

CycleGAN 通过引入 L_1 范数作为损失函数来解决第一个问题，从而强制保持低频一致性，使我们趋近于最优中位数。此外，还引入了块判别器，通过在块级别强制保持局部结构一致性来捕获高频成分。块判别器需要指定具有预定义感受野的架构，通常会形成一个“浅层”判别器。然而，我们使用 CycleGAN 进行的早期实验表明，它破坏了病变结构，并引入了原本不存在的血管。受块判别器启发，我们采用多尺度结构相似性指数（SSIM） [17]作为一致性损失函数 $\mathcal{L}_d$，其定义为 $\mathcal{L}_d(y,\ G_\theta(y)) = 1 - \mathbf{SSIM}_{MS}(y,\ G_\theta(y))$。仿照 CycleGAN，我们还引入了恒等损失 $\mathcal{L}_{idt}$ ，以确保高质量输入能够产生高质量增强结果，其定义为 $\mathcal{L}_{idt}(x,\ G_\theta(x)) = 1 - \mathbf{SSIM}_{MS}(x,\ G_\theta(x))$。我们还使用 UNet [18] 作为生成器，以帮助低层语义从低质量域流向高质量域。以下定理 [17] 为我们的损失函数定义提供了理论保证。

定理 1 ([17])。 我们证明，结构相似性指数在局部是拟凸的，这使原问题与对偶问题之间的对偶间隙最小化，并且弱对偶成立。

为更好地平衡恒等损失、域损失以及 $\mathbb{P}_X$ 与 $\mathbb{P}_{G_\theta(y)}$之间的散度，最终目标函数被改写为

$$\max_{G_\theta} \min_{D_w} \sum_{i=1}^{n} [\alpha \mathcal{L}_d(y_i, G_\theta(y_i)) + \beta \mathcal{L}_{idt}(x_i, G_\theta(x_i))] + \mathcal{W}_1(\mathbb{P}_X, \mathbb{P}_{G_\theta(Y)}), \tag{8}$$

其中， α, β 分别是域损失和恒等损失的权重参数。我们提出的最优传输引导的非配对图像增强训练方案的算法见算法 1。

2.2 通过增强进行正则化

通过去噪进行正则化（RED） [11] 是一种可直接使用的基于模型的框架，无需修改模型架构即可利用各种现有的卷积神经网络先验来引导图像复原。我们将以去噪器为核心的 RED 思路推广为一种更通用的方法，利用从所提出的最优传输引导增强网络中学习到的图像先验。在端到端训练样本不足时，我们将增强过程构建为一种图像先验，用于引导任意测试图像的复原。我们提出的通过增强进行正则化（RE）的目标函数为

$$\hat{x} = \underset{x}{\operatorname{argmin}}\, \mathbb{E}_x[\mathcal{L}(x, y)] + \gamma R(x) \text{ with } R(x) = \frac{1}{2} x^T (x - G_\theta(x)), \tag{9}$$

算法 2 通过增强进行正则化。

要求： 步长 η，正则化强度 γ，容差 **tol**，生成器 G_θ
要求： 初始 $\tilde{\boldsymbol{x}}^{(0)}$，$\boldsymbol{s}^{(0)} = \tilde{\boldsymbol{x}}^{(0)}$，$t^{(0)} = 1$
  当未收敛时**执行**
    $t^{(k)} = \frac{1}{2}(1 + \sqrt{1 + 4(t^{(k-1)})^2})$
    $\mathbf{Der}(\boldsymbol{s}^{(k-1)}) = \nabla_{\boldsymbol{s}^{(k-1)}} \mathcal{L}(\boldsymbol{s}^{(k-1)}, \boldsymbol{y}) + \gamma(\boldsymbol{s}^{(k-1)} - G_\theta(\boldsymbol{s}^{(k-1)}))$
    $\tilde{\boldsymbol{x}}^{(k)} \leftarrow \boldsymbol{s}^{(k-1)} - \eta \cdot \mathbf{Der}(\boldsymbol{s}^{(k-1)})$
    $\boldsymbol{s}^{(k)} \leftarrow \tilde{\boldsymbol{x}}^{(k)} + \frac{t^{(k-1)}-1}{t^{(k)}}(\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^{(k-1)})$
    如果 $||\tilde{\boldsymbol{x}}^{(k)} - \tilde{\boldsymbol{x}}^{(k-1)}|| \leq \mathbf{tol} \cdot ||\tilde{\boldsymbol{x}}^{(k-1)}||$ **则**
      $break$
    结束 if
  结束 while

其中，γ 控制正则化强度，$\mathcal{L}$ 表示多尺度结构相似性损失。RE 先验的梯度形式简单：

$$\nabla_x R(x) = x - G_\theta(x), \tag{10}$$

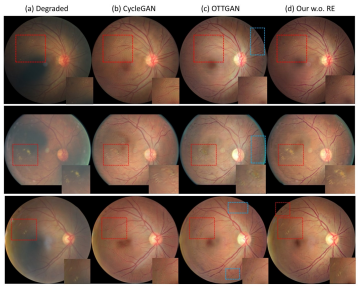
在 G_θ 局部齐次且具有对称雅可比矩阵的条件下，G_θ 的1-利普希茨约束可进一步保证G_θ 的被动性，从而得到一个凸目标函数。我们通过谱归一化 [19]对生成器 G_θ 中每个卷积层权重的谱半径进行正则化，以近似满足1-利普希茨约束。在优化阶段，我们选择加速梯度下降，迭代逼近最优解。增强正则化 的迭代优化过程见算法 2。

3 实验结果

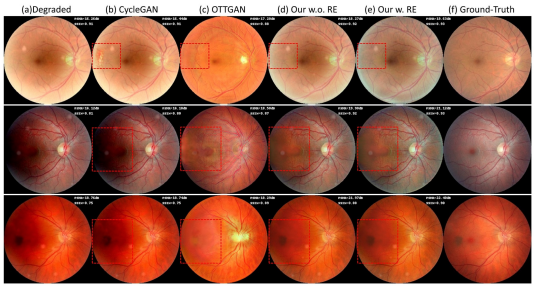
我们在真实干净图像可用（全参考评估）和不可用（无参考评估）的场景下进行了大量实验。我们研究了糖尿病视网膜病变分级、血管分割和病灶分割这三项下游任务，以进一步评估所提方法的性能。由人类眼科医生进行目视检查，以评估无参考评估的性能。我们使用标准 ResNet-50 [20] 和 UNet [18] 训练并测试下游任务。

3.1 数据集

我们提出的方法在三个公开可用的视网膜彩色眼底照片数据集上进行了广泛评估：EyeQ数据集 [21]，、DRIVE数据集 [22]，和IDRID数据集 [23]。EyeQ数据集经人工标注，分为三个质量等级：良好、可用和拒绝。我们使用7886张训练图像和8161张测试图像（良好和拒绝）进行训练和评估。我们在包含40名受试者的DRIVE数据集上评估了所提方法在血管分割任务中的表现。IDRID数据集包含81名受试者的像素级标注，标注对象包括微动脉瘤（MA）、软性渗出物（SE）、出血（HE）和硬性渗出物



(A) Human Expert's visual inspection of enhanced images from our low2high model.



(B) Visual comparison of enhanced images of our low2high model to other unsupervised competitors.

图2：（A）高亮方框表示病变和血管结构的对比。可以明显看出，其他所有方法都改变了病变或血管的结构。（B）红色方框表示与我们的方法相比生成了额外结构，以及噪声抑制效果不明显的区域。

（EX），用于评估我们的方法在糖尿病视网膜病变病灶分割任务中的表现。所有图像均经过中心裁剪，并调整至 256×256 的尺寸。

3.2 实验设计

在无参考评估中，我们提出的最优传输引导图像到图像转换生成对抗网络使用EyeQ数据集中的 7886 张训练图像进行训练，图像由未配对的低质量图像和高质量图像组成。该模型定义为低质量到高质量（$low2high$）模型。在全参考评估中，模型使用 EyeQ 高质量训练数据集的一个子集进行训练，其中包括通过 [3] 获得的退化图像和未配对的高质量图像。该模型定义为退化到高质量（$deg2high$）模型。算法 1 中的疾病标签 g 为 EyeQ 中的糖尿病视网膜病变分级标签。训练期间采用了数据增强，包括随机水平/垂直翻转、随机裁剪和随机旋转，以防止过拟合。所有模型均使用 RMSprop优化器训练 200 个训练轮次，判别器和生成器的初始学习率分别为 $1\times\ 10^{-4}$ 和 5×10^{-5}，每 100 个训练轮次衰减 10 倍。low2high 和 deg2high 模型的最优超参数均为 $\alpha = 60$、$\beta = 20$ 。在测试阶段，通过网格搜索确定最优超参数 γ，搜索范围为 1×10^{-3} 至 1×10^{-4}，所有实验的迭代次数均为 400。所有方法均使用 PyTorch 实现，论文录用后将在 GitHub 上公开代码。

3.3 无参考质量评估

在缺少真实干净图像作为参照的情况下，评估增强质量颇具挑战。我们考虑结合糖尿病视网膜病变分级任务与人类专家的目视检查来评估增强效果。糖尿病视网膜病变分级任务可作为判断增强后是否保留病灶信息的标准。我们按照 [24] 中的实验设置，在高质量图像上训练 ResNet-50 模型，并使用

	糖尿病视网膜病变分级			专家评估		
方法	准确率	卡帕系数	ROC	LCR	BCR	GESR
CycleGAN	0.7148	0.5378	0.9083	0.449	**0.0**	0.347
OTTGAN	0.6996	0.5105	0.8995	0.429	0.102	0.490
OTRE	**0.7767**	**0.6814**	**0.9403**	**0.020**	0.040	**0.326**

表1：采用不同方法增强后，糖尿病视网膜病变分级任务的评估指标，包括病灶结构变化率（LCR）、背景颜色变化率（BCR）和生成额外结构比率（GESR）。

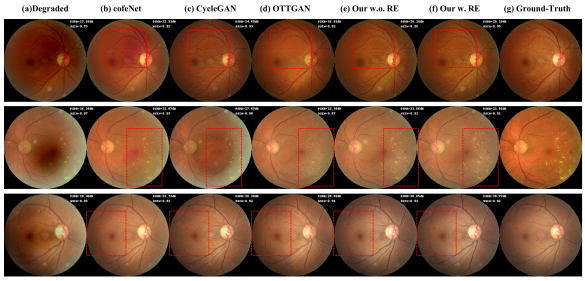
不同方法处理的低质量图像及其增强结果对模型进行评估。增强效果将通过分类准确率、受试者工作特征曲线下面积（AU-ROC）和科恩κ系数（卡帕系数）来衡量。

为供人类专家进行目视检查，我们随机选取了50张低质量图像，并采用不同的增强方法进行处理。目视检查用于测量：1）病灶结构变化率（LCR）；2）主要背景颜色变化率（BCR）；3）生成不存在结构的比率（GESR）。为确保实验公平，评估首先由三名接受过设计方案培训的志愿者进行，随后由眼科医生最终确认。

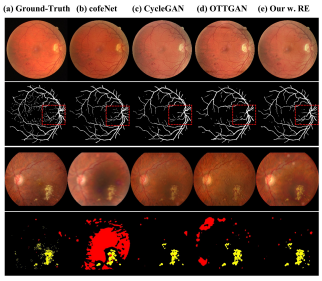
图2（A）展示了采用不同无监督增强方法得到的一些结果。所有方法都能提升图像质量。然而，我们的方法在降低噪声的同时，能更好地保留病变和血管结构。我们设计了两项实验，以验证我们的方法能够最大限度地保留信息（表1）。首先，我们对增强图像（low2high）应用糖尿病视网膜病变分级算法，并评估其分级准确率。如表1所示，OTRE在所有三项指标上均优于其他方法，尤其是卡帕系数指标高出20%以上。人类专家的检查也证实，我们的方法能够最大限度地保留信息。我们的方法在LCR和GESR方面表现最佳，其中LCR显著提升，甚至比另外两种方法高出40%以上。

3.4 全参考质量评估

对于全参考评估，我们遵循 Shen 等人提出的退化模型 [3] 对高质量图像进行退化，为每个数据集合成低质量图像。训练数据集由从 EyeQ 训练数据集中选取的 6500 张高质量图像，以及从 EyeQ 训练数据集中另外 6500 张互不重叠的高质量图像退化生成的 6500 张低质量图像组成。测试数据集由来自 EyeQ 测试数据集的 500 张图像、完整的 DRIVE 数据集和完整的 IDRID 数据集组成。我们采用常用的峰值信噪比（PSNR）和结构相似性指数（SSIM）来评估增强低质量图像的质量。为进一步验证所提出方法在下游任务上的有效性，我们评估了所提出方法在血管分割和糖尿病视网膜病变病灶分割任务上的性能。



(A) deg2high model evaluation based on the full-reference experiment.



(B) The visualization of downstream segmentation tasks.

图 3：（A）红色高亮区域展示了结构保留与额外结构生成的对比。从视觉上看，图像增强效果良好，但峰值信噪比不高，因此有必要进一步进行下游任务评估。（B）红色高亮区域表示细血管分叉分割结果的对比。

首先，我们评估了增强图像与其对应高质量图像之间的一致性。我们开展了两项不同的实验，分别采用无参考训练模型（第3.3节，low2high）和全参考训练模型（退化到高质量）。我们还测试了引入 RE模块是否能够提升我们的

		EyeQ		DRIVE		IDRID	
	方法	峰值信噪比	结构相似性指数	峰值信噪比	结构相似性指数	峰值信噪比	结构相似性指数
监督式	cofe-Net	23.11	**0.910**	21.87	0.767	20.25	0.825
无监督	CycleGAN **(low2high)**	18.57	0.836	18.78	0.705	19.13	0.799
	CycleGAN **(退化到高质量)**	22.75	0.895	21.92	0.766	21.56	0.855
	OTTGAN **(low2high)**	18.93	0.859	19.20	0.723	19.70	0.828
	OTTGAN **(退化到高质量)**	23.69	0.894	21.61	0.750	21.93	0.839
	不含增强正则化的 OTRE **（low2high）**	20.39	0.878	20.03	0.733	20.50	0.837
	带增强正则化的 OTRE **（low2high）**	21.08	0.880	20.61	0.740	20.55	0.836
	不带增强正则化的 OTRE **(退化到高质量)**	24.29	0.906	22.40	0.772	21.51	**0.860**
	带增强正则化的 OTRE **(退化到高质量)**	**24.63**	0.905	**22.81**	**0.794**	**22.05**	0.852

表2：在当前退化测试数据集上，使用无参考训练数据（第3.3节，low2high）和全参考训练数据（退化到高质量）训练的无监督方法的结果比较。在两个数据集上均评估了带有/不带有 RE 模块的 OTRE 框架。有监督方法（coef-Net）仅使用退化数据集进行训练和评估。

	血管分割					EX		HE		
方法	ROC	PR	F1	SE	SP	受试者工作特征曲线 PR曲线	F1值 受试者工作特征曲线	PR曲线		F1
cofe-Net	0.923	0.787	0.714	0.644	**0.977**	0.926	0.442	0.469	0.807	0.103 0.090
CycleGAN	0.910	0.762	0.696	0.622	0.975	0.900	0.474	0.347	0.845	0.155 0.141
OTTGAN	0.900	0.739	0.667	0.581	0.976	0.912	0.507	**0.512**	0.855	0.107 0.145
OTRE	**0.927**	**0.796**	**0.726**	**0.672**	0.975	**0.934**	**0.529**	0.441	**0.894**	**0.233 0.273**

表3：DRIVE队列中血管分割以及IDRID数据集 [23]中糖尿病病变（硬性渗出物和出血）的结果比较 [22] 。与其他监督和无监督方法相比，OTRE表现更优。（ROC：受试者工作特征曲线下面积，PR：精确率-召回率曲线下面积，F1：F1分数，SE：灵敏度，SP：特异度）。

结果。图2（B）和图3（A）展示了一些图像示例，表2报告了数值结果。如表2所示，除EyeQ的结构相似性指数指标外，我们的OTRE在三个不同数据集上的表现均优于所有其他监督和无监督方法，峰值信噪比分别达到最高值24.63、22.81和22.05。值得注意的是，尽管最先进的监督方法（cofe-Net）使用配对图像进行训练，而OTRE并未使用配对图像训练，OTRE仍优于cofe-Net。有趣的是，我们的无参考训练模型（$low2high$）在未见过的退化噪声上也取得了有竞争力的结果。我们还发现，加入增强正则化模块后，性能有所提升。图2（B）和图3（A）也进一步有力地证明，我们的方法能够保留结构并实现更好的噪声抑制。

为了进一步验证我们方法的优越性，我们使用来自DRIVE和IDRID数据集的真实标注数据，研究了两个下游分割任务。由于分割任务的训练和测试完全基于增强图像，未添加任何预处理或额外技巧，因此在病变分割任务中，我们只考虑了易于训练的大块病变，例如硬性渗出物和出血。如表3所示，OTRE方法在三个分割任务中均取得了出色的结果。在所有分割结果中，其受试者工作特征曲线和PR曲线指标均为最高，并在F1值指标上取得了3项结果中的2项最佳成绩。从图3（B）所示的一些图像示例中可以清楚看到，其他方法存在增强效果不明显的问题，导致血管和病变结构发生改变。

4 结论与未来工作

本研究将 OT 引导的基于 GAN 的增强网络与 RE 模块相结合，并在三个数据集上取得了良好结果，性能超过或媲美最先进的无监督和监督方法。当前系统的一个局限是，它假设所有输入数据都可用；但在实际临床应用中，有些图像会完全损坏。若能通过筛查流程判断输入图像是否可用，将使我们的工作更具实用性。我们将在未来的工作中对此进行研究。

致谢。本研究部分得到了美国国立卫生研究院的资助（R21AG065942、R01EY032125 和 R01DE030286）。

参考文献

1. Wolf, R.M., Channa, R., Abramoff, M.D., Lehmann, H.P.：自主即时糖尿病视网膜病变筛查对患有糖尿病的儿科患者的成本效益。JAMA Ophthalmol **138**(10), 1063–1069（2020年10月） 2. Cheung, C.Y., 等：基于视网膜照片检测阿尔茨海默病的深度学习模型：一项回顾性、多中心病例对照研究。Lancet Digit Health **4**(11), e806–e815（2022） 3. Shen, Z., Fu, H., Shen, J., Shao, L.：低质量视网膜眼底图像的建模与增强。IEEE医学影像学汇刊 **40**(3), 996–1006（2021） 4. Lehtinen, J., 等：Noise2Noise：无需干净数据的图像复原学习。国际机器学习会议 **80**, 2965–2974（2018） 5. Krull, A., 等：Noise2void——从单张含噪图像中学习去噪。IEEE计算机学会计算机视觉与模式识别会议论文集，第2129–2137页（2019） 6. Bousmalis, K., Silberman, N., 等：利用生成对抗网络进行无监督像素级域适应。计算机视觉与模式识别会议。（2016） 7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.：利用条件对抗网络进行图像到图像转换。计算机视觉与模式识别会议。（2017） 8. Zhu, J., Park, T., Isola, P., Efros, A.A.：利用循环一致对抗网络进行无配对图像到图像转换。计算机视觉与模式识别会议，第2242–2251页（2017） 9. Liu, M.Y., Tuzel, O.：耦合生成对抗网络。神经信息处理系统进展（2016） 10. Wang, W., Wen, F., Yan, Z., Liu, P.：用于无监督去噪学习的最优传输。IEEE PAMI，第1–1页（2022） 11. Romano, Y., Elad, M., Milanfar, P.：小引擎也能发挥大作用：去噪正则化（RED）。SIAM影像科学杂志 **10**(4), 1804–1844（2017） 12. Ryu, E., Liu, J., Wang, S., Chen, X., Wang, Z., Yin, W.：经过适当训练的去噪器可确保即插即用方法收敛。PMLR **97**, 5546–5557（2019） 13. Lucas, A., Iliadis, M., Molina, R., Katsaggelos, A.K.：利用深度神经网络解决成像逆问题：超越解析方法。IEEE信号处理杂志 **35**(1), 20–36（2018） 14. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.：ECA-Net：面向深度卷积神经网络的高效通道注意力。IEEE计算机视觉与模式识别会议（2020） 15. Ledig, C., Theis, L., 等：利用生成对抗网络实现照片级单图像超分辨率。计算机视觉与模式识别会议。（2016） 16. Gulrajani, I., Ahmed, F., Arjovsky, M., 等：改进Wasserstein生成对抗网络的训练。神经信息处理系统进展 **30** （2017） 17. Brunet, D., Vrscay, E.R., Wang, Z.：结构相似性指数的数学性质。IEEE图像处理汇刊 **21**(4), 1488–1499（2012） 18. Ronneberger, O., Fischer, P., Brox, T.：U-Net：用于生物医学图像分割的卷积网络。MICCAI，第234–241页（2015） 19. Miyato, T., 等：生成对抗网络的谱归一化。国际学习表征会议（2018） 20. He, K., Zhang, X., Ren, S., Sun, J.：用于图像识别的深度残差学习。IEEE计算机学会计算机视觉与模式识别会议论文集（2016）

21. Fu, H., Wang, B., Shen, J., 等：不同色彩空间下视网膜图像质量评估网络的评估。MICCAI，第48–56页（2019） 22. Staal, J., 等：视网膜彩色图像中的基于脊线的血管分割。IEEE医学影像学汇刊 **23**(4), 501–509（2004） 23. Porwal, P., 等：IDRID：糖尿病视网膜病变筛查研究数据库。Data**3**(3)（2018） 24. Zhu, W., Qiu, P., Lepore, N., Dumitrascu, O., Wang, Y.：自监督等变正则化协调多实例学习：联合进行可转诊糖尿病视网膜病变分类和病灶分割。第18届国际医学信息处理与分析研讨会（SIPAIM）（2022）