LEAF: 通过高效编码器蒸馏实现医学图像分割中对齐特征的潜在扩散

Qilin Huang[1], Tianyu Lin[2], Zhiguang Chen[1], and Fudan Zheng[1] ✉

[1] School of Computer Science and Engineering, Sun Yat-sen University, China
zhengfd5@mail.sysu.edu.cn
[2] Department of Biomedical Engineering, Johns Hopkins University, United States

Abstract. 利用扩散模型的强大功能，在医学图像分割任务中取得了相当有效的结果。然而，现有方法通常直接转移原始训练过程，而没有针对分割任务进行具体调整。此外，常用的预训练扩散模型在特征提取方面仍存在不足。基于这些考虑，我们提出了一种名为 LEAF 的医学图像分割模型，该模型基于潜空间扩散模型。在微调过程中，我们用直接预测分割图的方式替换了原始的噪声预测模式，从而减少了分割结果的方差。我们还采用特征蒸馏方法，将卷积层的隐藏状态与基于变压器的视觉编码器的特征对齐。实验结果表明，我们的方法在不同疾病类型的多个分割数据集上增强了原始扩散模型的性能。值得注意的是，我们的方法并不改变模型架构，也不增加推理阶段的参数或计算数量，使其具有高度的效率。项目页面：https://leafseg.github.io/leaf/

Keywords: Latent Diffusion · Feature Alignment · Efficient.

1 介绍

扩散模型 [20] 在多项图像生成任务中取得了成功的成果，展示了其作为生成高维视觉数据的可扩展方法的能力。由于其强大的能力，最近的研究开始探索其在其他视觉任务中的应用潜力。例如，DMP [11] 适配文本到图像的扩散模型，以在若干密集预测任务中获得准确的估计。Marigold [10] 直接微调了稳定扩散 [17] ，用于图像条件下的深度生成，在多个深度估计数据集上实现了最先进的（SOTA）性能，同时能够对未见过的数据进行零样本迁移。

鉴于扩散模型的强大能力，众多研究探索了其在医学图像分割中的应用 [24,25,26] ，展示了其显著的潜力并激发了社区中日益增长的研究兴趣。然而，这些方法通常直接采用原始的扩散模型训练过程，并常常引入过于复杂的模块来增强特征表示。虽然提高了性能，这些设计带来了计算效率低下的问题，并且掩盖了分割和生成任务之间的基本差异。相比之下，SDSeg [13] 使用潜在扩散模型，并通过单步逆过程提高了推理速度和准确性。尽管如此，这种方法仍然未能解决分割目标（如逐像素分类）和生成建模原则（如噪声预测）之间的内在差异。已经有一些先前的工作研究了替代的参数化方法 [19,1] 来生成详细且逼真的自然图像。然而，这些方法要么依赖于多步渐进生成，要么提供了相当的评估指标，因此对这一任务的见解有限。

同时，广泛使用的预训练扩散模型是基于卷积 U-Net 架构。许多研究指出，Transformer 架构 [22] 可以有效增强特征提取，虽然这也增加了计算成本和参数数量。此外，一些研究 [12] 表明，从单张图像中估计像素级几何属性需要对场

景有全面的理解，仅预测输入空间的结果不足以学习到稳健的表示。因此，在速度和性能之间实现良好平衡仍然是当前工作中的一个重要挑战。蒸馏方法提供了一个有前景的解决方案。值得注意的是，近期的 REPA 方法 [27] 通过对齐两个 Transformers [22] 的特征来加速模型收敛，从而改善了生成性能。

出于这些考虑，我们提出了 LEAF（一种用于特征对齐的高效编码器蒸馏潜变量扩散方法）。我们分析了扩散公式，发现使用噪声预测在分割任务中可能并不是最佳方案，因为它倾向于放大预测误差。因此，我们用一种直接预测样本的方法替代了这种方法。此外，我们实施了一种新颖而简单的蒸馏方法，以增强卷积网络的特征表示，使模型能够将其特征与从强大的基于 Transformer 的模型中获得的特征对齐。这种对齐在推理过程中不增加任何额外的计算开销或参数增加的情况下，提升了分割性能。

总之，我们的主要贡献如下：

- 我们将最初用于生成任务的扩散模型中的高方差 ϵ 预测替换为更适合分割任务的低方差 x_0 预测，并提供相应的数学解释。
- 我们设计了一种高效的特征对齐方法，通过蒸馏一个强大的视觉编码器来丰富 U-Net 的表示，从而提高多种疾病的多个医学影像数据集的分割性能。
- 我们的方法允许在推理期间移除对齐模块，不会带来额外的计算或内存开销。此外，这种即插即用的方法不会改变模型的内部结构，并且可以轻松移植到其他基于扩散的模型中。

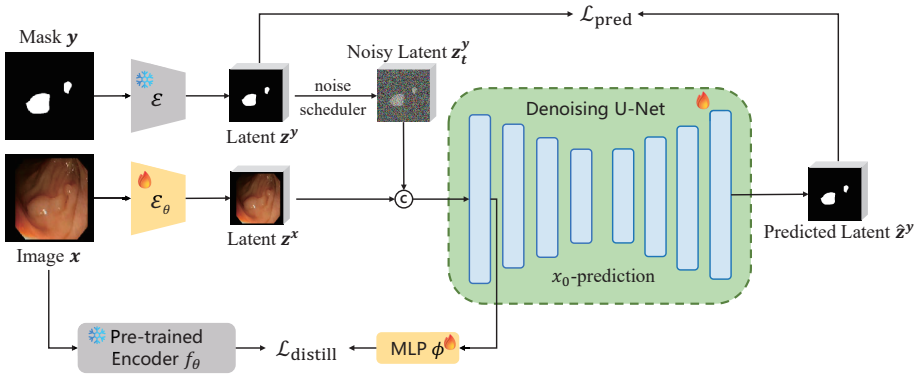

Fig. 1. LEAF 训练流程的示意图，c 表示连接，解码器 $\mathcal{D}$ 被省略，噪声调度器是方差保持型的，例如 $\alpha_t^2 + \sigma_t^2 = 1$。

2 方法

我们方法的框架如图 1 所示。对于调节方法，我们遵循 SDSeg [13] 的过程。给定一个地面真值分割图 y，我们使用冻结的编码器 $\mathcal{E}$ 将其映射到潜变量 z^y，并通过方程 (1) 添加噪声以获得噪声变量 z_t^y，其中 ϵ 是随机高斯噪声，α_t 和 σ_t

是由噪声调度器预定义的一组超参数；通常，α_t 减少而 σ_t 增加。对于图像 x ，我们使用可学习的编码器 $\mathcal{E}_\theta$ ，以 $\mathcal{E}$ 的权重初始化，将其映射到 z^x 。然后，我们将 $\text{concat}(z^x; z_t^y)$ 作为输入连接到 U-Net 并获得输出 $\hat{z}^y$ 。

$$z_t = \alpha_t z + \sigma_t \epsilon \tag{1}$$

2.1 参数化类型

当前扩散模型主要用于生成任务，预测目标普遍使用以下两种方法：(1) ϵ 预测 [9] ，其中模型学习预测噪声 ϵ ；(2) $\boldsymbol{v}$ 预测 [19] ，其中模型学习预测由方程 (2) 定义的速度。

$$\boldsymbol{v} := \alpha_t \epsilon - \sigma_t z \tag{2}$$

这些参数化类型可用于通过公式 (3) 估计原始图像：

$$\hat{z} = \begin{cases} (z_t - \sigma_t \hat{\epsilon})/\alpha_t & \text{, If } \epsilon\text{-prediction} \\ \alpha_t z_t - \sigma_t \hat{\boldsymbol{v}} & \text{, If } \boldsymbol{v}\text{-prediction} \\ \hat{z} & \text{, If } x_0\text{-prediction} \end{cases} \tag{3}$$

由于我们在训练期间冻结了解码器 $\mathcal{D}$ ，因此 $\hat{x}$ 中的错误主要来自 $\hat{z}$ 。请注意，当重建 z_t 时，两种参数化方法都涉及基于方差的系数；随着 $t \to T$ ，σ_t 增加而 α_t 减少。在 $t = T$ 使用单步逆过程的扩散模型中，这进一步放大了估计中的误差。

因此，我们提出 x_0 -prediction [9] 方法更适合用于图像分割。使用 x_0 -prediction，扩散模型直接输出 $\hat{z}$ ，而不引入额外的缩放系数，从而避免了不必要的错误。与其他两种预测方法相比，这种方法产生更稳定和准确的结果。总之，对于使用 ϵ -prediction 或 $\boldsymbol{v}$ -prediction 预训练的扩散模型，我们直接使用 x_0 -prediction 进行微调，以便它直接预测分割图。相应的损失如方程 (4) 中所示，使用了 SDSeg 中实现的 $L1$ 损失。

$$\mathcal{L}_{\text{pred}} = \mathcal{L}_{L1}(\hat{z}^y, z^y) \tag{4}$$

2.2 特征对齐

在最近的研究中，增强扩散模型提取特征的能力通常通过修改模型架构来实现。如在 TransUNet [4] 和 Diff-Trans [5] 中所示，Transformer 架构能够有效地提高编码器的特征提取能力，但它们也显著增加了模型参数数量和计算成本。为了促使基于 U-Net 的潜变量扩散模型学习丰富的表征，我们引入一种正则化策略来增强卷积的能力，使其能够捕获 Transformer 架构学习到的表征。

受 REPA [27] 启发，我们利用预训练的自监督强大视觉编码器 f_θ ，例如 DINOv2 [14] 或 CLIP [16] ，作为提供鲁棒视觉表征的基础模型。它将一个干净的图像 x 作为输入并产生隐藏状态 $h = f_\theta(x) \in \mathbb{R}^{L \times D}$ ，其中 L 是块的数量，D 是嵌入维数。在去噪 U-Net 的编码器块中，我们获得一个特征图 $m \in \mathbb{R}^{C \times H \times W}$

并将其重塑为 $\mathbb{R}^{HW\times C}$ ，条件是 $HW = L$ 。然后，我们使用多层感知器（MLP）ϕ 来投射 m ，得到 $\phi(m) \in \mathbb{R}^{L\times D}$ ，并基于余弦相似性计算一个蒸馏损失：

$$\mathcal{L}_{\text{distill}} = \mathcal{L}(h, \phi(m)) = -\frac{1}{N}\sum_{i=1}^{N}(\frac{h_i \cdot m_i}{\|h_i\| \cdot \|m_i\|}) \tag{5}$$

我们将这个损失添加到上面的预测损失中，总的目标损失如公式 (5) 所示，其中 λ 是控制蒸馏对齐强度的一个正常数。

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pred}} + \lambda\mathcal{L}_{\text{distill}} \tag{6}$$

2.3 推理

在推理过程中，我们用标准高斯噪声 z_T^y 初始化分割图，并使用 $\mathcal{E}_\theta$ 将输入图像编码为 z^x 。然后将拼接后的特征 $(z_T^y; z^x)$ 输入到 U-Net 中。值得注意的是，在此阶段我们移除了预训练的视觉编码器和 MLP，确保与原始模型相比没有引入额外的参数。根据 SDSeg [13] ，我们应用单步逆过程来获得 $\hat{z}^y$ ，然后通过 $\mathcal{D}$ 解码到像素空间，生成最终的分割图。

3 实验结果

3.1 实验设置

为了全面评估所提出的方法，我们在四个公共医学图像分割任务上进行实验：(1) 从视网膜眼底图像中分割视杯 (REFUGE2 (REF) [15])，(2) 从结肠镜图像中分割息肉 (CVC-ClinicDB (CVC) [2])，(3) COVID-19 病灶分割 (QaTa-Covid19 (Qata) [7])，以及 (4) 从皮肤镜图像中分割皮肤病灶 (ISIC 2018 [6,21])。我们使用平均 Dice 系数和平均 IoU 作为主要评估指标。对于 REFUGE2，我们使用 SDSeg 中定义的数据分区。对于 ISIC 2018，我们采用了训练-测试比例为 $7:3$ ，而对于 CVC，我们使用了 80:10:10 的数据分区。另一方面，QaTa 使用了默认的训练和测试集。

我们在 PyTorch 平台上实现了 LEAF，并在单个 NVIDIA A800 GPU 上进行训练和评估。所有图像均调整为 256×256 的分辨率。预训练的无条件潜在扩散模型基于 LDM-KL-8 [17] 。为了优化模型，我们使用了标准的 AdamW 优化器，批量大小为 4 。学习率设置为 4×10^{-5} ，并使用暖启动常数学习率调度器。为了处理拼接输入，我们将 U-Net 输入层从 4 个通道复制到 8 个通道，并通过将原始权重减半来初始化其权重，如 Marigold [10] 中所述。预训练的视觉编码器是 DINOv2 [14] 。

3.2 性能比较

我们在各种评估数据集中进行了广泛的实验，以评估 LEAF 的有效性，如表 1 所示。LEAF 是针对潜在扩散模型的一种通用方法，不包含针对特定医学成像模态的领域特定模块。因此，我们的比较主要集中在具有较强泛化能力的模型上。为了与 SDSeg 进行公平的比较，我们在相同配置下，在我们的框架内重新

Table 1. 我们提出的模型与现有的 SOTA 医学分割模型的性能比较。

	Model	REF		CVC		QaTa		ISIC2018	
		Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
CNN/Transformer-based	U-Net[18]	80.1	-	85.6	80.5	79.0	69.5	87.6	77.9
	TransUNet[4]	85.6	-	92.0	87.8	78.6	69.1	88.7	79.7
	Swin-UNet[3]	84.3	-	91.4	87.4	78.1	68.3	-	-
Diffusion-based	MedSegDiff[24]	86.3	78.2	92.4	88.9	76.5	67.2	85.5	74.7
	SDSeg[13]	88.7	80.9	93.6	89.3	77.6	68.0	88.1	79.7
Proposed	LEAF	89.5	81.5	95.2	90.9	80.2	71.0	90.5	84.1

训练了它。MedSegDiff 在 CVC 上进行了重新评估。QaTa 和 ISIC2018 的结果直接引用自 MMDSN [23] 和 BGDiffSeg [8] ，而 REF 的结果来自 SDSeg [13] ，CVC 的结果来自 Diff-Trans [5] 。

如表 1 所示，LEAF 在涉及各种类型医学图像的数据集上优于所有基线模型，验证了其有效性和普适性。虽然与 SDSeg 共享相同的 U-Net 架构，但我们的方法替换了其参数化类型，并将卷积层与从基于 Transformer 的编码器提取的特征对齐，实现了显著的性能提升。

3.3 消融研究

微调流程的消融 我们将使用原始 ϵ 预测且不进行特征对齐的模型作为基线（表格 2 中的第一行）。从表中可以看出，仅仅将预测方法从 ϵ 预测更换为 $\boldsymbol{v}$ 预测就能得到显著的性能提升。此外，切换到不带缩放因子的 x_0 预测可以进一步提高模型性能。最后，特征对齐相比其他配置实现了最高的性能。尽管这些特征来自不同的模型架构，且 DINOv2 并未在医学图像上进行微调，但它仍能改善分割性能。我们强调，这些改进在所有评估的数据集上都是一致的，并且具有非平凡的幅度。

Table 2. 参数化和特征对齐的消融研究。灰色行突出显示在训练期间从视觉编码器中提取的模型特征，以便更清晰地进行比较。

Parameterization Types	Feature Alignment	REF		CVC		QaTa		ISIC2018	
		Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
ϵ -prediction	✗	88.47	79.59	90.15	83.68	74.27	63.80	87.67	80.13
ϵ -prediction	✓	87.61	78.43	91.63	87.10	74.56	63.97	87.34	79.48
$\boldsymbol{v}$ -prediction	✗	89.21	80.92	93.75	89.32	79.10	69.74	90.35	83.91
$\boldsymbol{v}$ -prediction	✓	89.30	80.88	94.89	90.44	79.32	69.98	90.52	84.11
x_0 -prediction	✗	89.21	79.08	94.49	89.94	79.08	69.85	90.39	83.94
x_0 -prediction	✓	89.53	81.45	95.17	90.94	80.15	71.04	90.54	84.18

我们研究了控制对齐强度的超参数 λ 和视觉编码器的模型大小，结果如表 3 所示。首先，λ 的最佳值通常在不同的数据集上不同，这可能与数据的分布和固有特性有关。其次，不同 λ 值对模型性能的影响不显著，例如，ISIC2018 上的 dice 评分最大绝对差为 0.3 ，而在 QaTa 上则超过 1.0 。我们认为这是因为

QaTa 中的图像包含更多的结构信息，使它们对蒸馏强度更敏感。总体而言，模型在 $\lambda > 0$ 的情况下表现优于在 $\lambda = 0$ 的情况下，这进一步验证了特征对齐的有效性。

Table 3. 超参数 λ 对特征对齐的影响。

λ	REF		CVC		QaTa		ISIC2018	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
0	89.21	79.08	94.49	89.94	79.08	69.85	90.39	83.94
0.15	89.39	81.19	95.07	90.77	79.62	70.37	90.24	83.83
0.25	89.41	81.24	94.97	90.57	79.74	70.65	90.54	84.06
0.50	89.33	81.12	94.21	89.35	80.05	70.87	90.43	84.06
0.75	89.53	81.45	95.01	90.67	79.98	70.93	90.51	84.05
1.0	89.44	81.27	95.17	90.94	79.87	70.77	90.34	83.90
1.25	89.43	81.27	95.07	90.76	80.15	71.04	90.30	83.93

3.4 质量结果

稳定性 扩散模型是非确定性模型；因此，许多先前的模型尝试通过多次运行模型并以某种方式整合结果作为最终结果来减少分割的不确定性，这增加了模型的推理速度。LEAF 的分割结果是通过单次运行获得的，因此有必要展示其稳定性，结果如表 4 所示。

Table 4. 稳定性实验。我们选择了 10 个不同的随机种子进行 10 次推断，并计算了 Dice 系数（%）的标准偏差。

Parameterization Types	REF	CVC	QaTa	ISIC2018
ϵ -prediction	0.09	0.32	0.14	0.29
x_0 -prediction	0.05	0.11	0.08	0.06

实验结果表明，与 x_0 预测方法相比，ϵ 预测方法的方差确实更大。此外，使用 x_0 预测方法训练的模型的标准差大多是 10^{-2} 的量级，表明差异非常小。这个差异显著小于我们所提出的方法所带来的改进，进一步证明了我们方法的有效性和稳定性。

此外，我们在不同的医学分割任务中可视化分割结果，如图 2 所示。

4 结论

在本文中，我们提出了 LEAF，一种高效且广泛的框架，用于微调潜在扩散模型以进行医学图像分割。我们研究了参数化类型的影响，并建议使用 x_0 预测参数化进行分割任务。我们还通过对视觉编码器进行蒸馏，介绍了一种简单的特征对齐方法，为基于 CNN 的 U-Net 提供了更好的表示。LEAF 在推理时不增加成本，并且可以轻松适应其他基于 LDM 的分割模型。

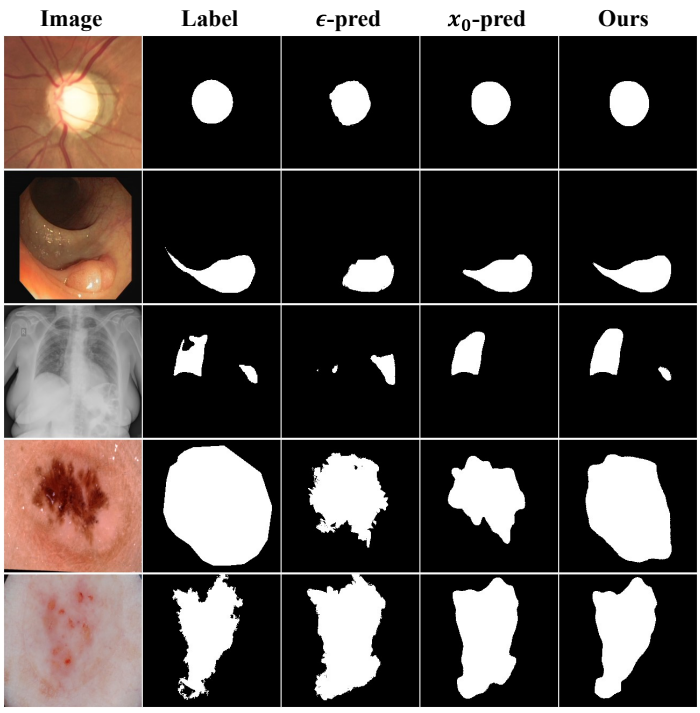

Fig. 2. 分割结果的可视化。

Acknowledgments. This study was funded by the Guangdong S & T Program (Grant No. 2024B0101040005), the National Natural Science Foundation of China (Grant No. 62461146204), the National Key Research and Development Program of China (Grant No. 2021YFB0300103), and the Pazhou Lab program (Grant No. PZL2023KF0001.)

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Benny, Y., Wolf, L.: Dynamic dual-output diffusion models. In: CVPR. pp. 11472–11481. IEEE (2022)
2. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., de Miguel, C.R., Vilariño, F.: WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Comput. Medical Imaging Graph. 43 , 99–111 (2015)
3. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: ECCV Workshops (3). Lecture Notes in Computer Science, vol. 13803, pp. 205–218. Springer (2022)
4. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image

segmentation through the lens of transformers. Medical Image Analysis p. 103280 (2024)
5. Chowdary, G.J., Yin, Z.: Diffusion transformer u-net for medical image segmentation. In: MICCAI (4). Lecture Notes in Computer Science, vol. 14223, pp. 622–631. Springer (2023)
6. Codella, N.C.F., Gutman, D.A., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N.K., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC). In: ISBI. pp. 168–172. IEEE (2018)
7. Degerli, A., Kiranyaz, S., Chowdhury, M.E.H., Gabbouj, M.: Osegnet: Operational segmentation network for covid-19 detection using chest x-ray images. In: ICIP. pp. 2306–2310. IEEE (2022)
8. Guo, Y., Cai, Q.: BGDiffSeg: a Fast Diffusion Model for Skin Lesion Segmentation via Boundary Enhancement and Global Recognition Guidance . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15009. Springer Nature Switzerland (October 2024)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851. Curran Associates, Inc. (2020)
10. Ke, B., Obukhov, A., Huang, S., Metzger, N., Daudt, R.C., Schindler, K.: Repurposing diffusion-based image generators for monocular depth estimation. In: CVPR. pp. 9492–9502. IEEE (2024)
11. Lee, H., Tseng, H., Lee, H., Yang, M.: Exploiting diffusion prior for generalizable dense prediction. In: CVPR. pp. 7861–7871. IEEE (2024)
12. Li, A.C., Prabhudesai, M., Duggal, S., Brown, E., Pathak, D.: Your diffusion model is secretly a zero-shot classifier. In: ICCV. pp. 2206–2217. IEEE (2023)
13. Lin, T., Chen, Z., Yan, Z., Yu, W., Zheng, F.: Stable diffusion segmentation for biomedical images with single-step reverse process. In: MICCAI (8). Lecture Notes in Computer Science, vol. 15008, pp. 656–666. Springer (2024)
14. Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P., Li, S., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jégou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: Dinov2: Learning robust visual features without supervision. Trans. Mach. Learn. Res. 2024 (2024)
15. Orlando, J.I., Fu, H., Breda, J.B., van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P., Kim, J., Lee, J., Lee, J., Li, X., Liu, P., Lu, S., Murugesan, B., Naranjo, V., Phaye, S.S.R., Shankaranarayana, S.M., Bogunovic, H.: REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical Image Anal. 59 (2020)
16. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. In: ICML. Proceedings of Machine Learning Research, vol. 139, pp. 8748–8763. PMLR (2021)
17. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR. pp. 10674–10685. IEEE (2022)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (3). Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer (2015)

19. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. In: ICLR. OpenReview.net (2022)
20. Sohl-Dickstein, J., Weiss, E.A., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: ICML. JMLR Workshop and Conference Proceedings, vol. 37, pp. 2256–2265. JMLR.org (2015)
21. Tschandl, P., Rosendahl, C., Kittler, H.: Descriptor : The ham 10000 dataset , a large collection of multi-source dermatoscopic images of common pigmented skin lesions (2018), https://api.semanticscholar.org/CorpusID:263789934
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: NIPS. pp. 5998–6008 (2017)
23. Wang, H., Zhang, Z., Zhang, Y., Zhang, J., Ou, Y., Sun, X.: Multi-modal diffusion network with controllable variability for medical image segmentation. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 3817–3822 (2024). https://doi.org/10.1109/BIBM62325.2024.10822810
24. Wu, J., Fang, H., Zhang, Y., Yang, Y., Xu, Y.: Medsegdiff: Medical image segmentation with diffusion probabilistic model. In: International Conference on Medical Imaging with Deep Learning (2022)
25. Wu, J., Ji, W., Fu, H., Xu, M., Jin, Y., Xu, Y.: Medsegdiff-v2: Diffusion-based medical image segmentation with transformer. In: AAAI. pp. 6030–6038 (2024)
26. Xing, Z., Wan, L., Fu, H., Yang, G., Zhu, L.: Diff-unet: A diffusion embedded network for volumetric segmentation. CoRR 10326 (2023)
27. Yu, S., Kwak, S., Jang, H., Jeong, J., Huang, J., Shin, J., Xie, S.: Representation alignment for generation: Training diffusion transformers is easier than you think. In: International Conference on Learning Representations (2025)

A 特征对齐的稳定性

我们通过在 QaTa 数据集上测试不同的随机种子来评估超参数 λ 的稳定性。结果表明，性能的轻微波动主要是由于随机性，确认了 λ 的稳健性。

Table 5. 不同 λ 值下的 Dice 分数

λ	0.15	0.25	0.5	0.75	1.0	1.25
Dice	79.88 ± 0.11	79.82 ± 0.14	79.81 ± 0.06	79.89 ± 0.01	80.06 ± 0.01	80.17 ± 0.01

B 噪音调度器的影响

我们还研究了不同噪声调度器对分割性能的影响。结果表明，调度器的选择对模型结果有显著影响，因此需要进一步研究以确定最佳配置。

Table 6. 贝塔调度和相应 Dice 分数的比较

Beta Schedule	linear	linear	linear	scaled linear	scaled linear
	0.0001–0.02	0.00085–0.012	0.0015–0.0155	0.0001–0.02	0.0015–0.0155
Dice	79.58	76.29	79.59	77.64	77.70