

DIFFUSION ADVERSARIAL REPRESENTATION LEARNING FOR SELF-SUPERVISED VESSEL SEGMENTATION

Boah Kim, Yujin Oh and Jong Chul Ye

ICLR 2023

摘要

医学图像中的血管分割是诊断血管疾病和制定治疗计划的重要任务之一。尽管学习型分割方法得到了广泛研究，但在监督方法中需要大量的地面真实标签，而混乱的背景结构使得神经网络难以在无监督的情况下分割血管。为了解决这个问题，本文引入了一种新颖的扩散对抗表示学习（DARL）模型，该模型利用了一种具有对抗学习的去噪扩散概率模型，并将其应用于血管分割。特别是，针对自监督的血管分割，DARL 通过扩散模块学习背景信号，使生成模块能够有效提供血管表示。此外，通过基于提出的可切换的空间自适应去标准化的对抗学习，我们的模型估计合成的假血管图像以及血管分割掩模，进一步使模型捕捉与血管相关的语义信息。一旦训练完成，该模型可以在单一步骤中生成分割掩模，并可应用于冠状动脉造影和视网膜图像的一般血管结构分割。对各种数据集的实验结果表明，我们的方法在很大程度上优于现有的无监督和自我监督血管分割方法。

关键词：图像分割；血管结构；扩散模型；自监督学习

1 引言

该文献的选题背景集中在医学图像处理领域的血管分割任务上。在临床诊断血管疾病和制定治疗计划的过程中，血管分割是不可或缺的步骤。特别是在冠状动脉疾病的诊断中，X射线血管造影图像通过向血管注射对比剂以提高血管可视性。然而，由于低对比度、运动伪影、复杂的血管结构等因素，传统的图像分割方法在这一领域面临着挑战。虽然基于深度学习的分割方法得到了广泛研究，但在监督方法中需要大量的地面实况标签，而无监督方法由于复杂的背景结构很难使神经网络有效地进行血管分割。因此，该文献提出了一种新颖的方法，即扩散对抗表示学习（DARL），该方法是一种自监督学习的方法，其中模型通过学习背景信号和对抗学习的方式，在无需地面实况标签的情况下进行血管分割，通过结合去噪扩散概率模型和对抗学习，实现了在无监督和自我监督条件下的血管分割。这种方法通过自监督学习，使用扩散模块学习背景信号，使生成模块能够有效提供血管表示。通过对抗学习，该模型

不仅估计合成的假血管图像，还生成血管分割掩模，从而进一步捕获血管的语义信息。这种方法有望在医学图像处理中取得更好的血管分割性能，尤其是在缺乏标签数据的情况下，为医学影像分析提供更有前景的解决方案。

2 相关工作

该文章提到了与其研究内容相关的工作，主要分为两个方面：一是与扩散模型相关的工作，二是与自监督血管分割相关的工作。

2.1 扩散模型

扩散模型 (Diffusion Model) 是生成模型之一，通过学习真实图像的分布来采样逼真的数据。特别是，去噪扩散概率模型 (Denoising Diffusion Probabilistic Model, DDPM) [1] (Ho 等, 2020) 在图像生成方面表现出色，采用了得分匹配。具体而言，DDPM 学习马尔可夫链将高斯噪声分布 $x_T \sim \mathcal{N}(0, I)$ 转换为目标分布 x_0 。在正向扩散过程中，通过以下逐步添加噪声到数据：

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I) \quad (1)$$

其中 $\beta_t \in [0, 1]$ 是固定的方差。因此，从数据 x_0 得到的噪声目标 x_t 分布表示为：

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_0, (1 - \alpha_t)I) \quad (2)$$

其中， $\alpha_t = \prod_{s=1}^t (1 - \beta_s)$ 。然后，DDPM 被训练来近似反向扩散过程：

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t); \sigma_t^2 I) \quad (3)$$

其中， σ_t^2 是一个固定的方差，而 θ 是具有噪声预测器 θ 的参数化均值：

$$\mu_\theta(x_t; t) = \frac{1}{\sqrt{(1 - \beta_t)}} \left(x_t - \frac{\beta_t}{\sqrt{(1 - \alpha_t)}} \epsilon_\theta(x_t; t) \right) \quad (4)$$

因此，在生成过程中，通过迭代去噪步骤，可以从高斯噪声中获得样本： $x_{t-1} = \mu_\theta(x_t; t) + \sigma_t z$ ，其中 $z \sim \mathcal{N}(0, I)$ 。

通过这种随机过程，DDPM 提供了多样化的真实样本，并在许多领域中得到了广泛的使用，包括超分辨率 [2] [3]，修复 [4] 和着色 [5] 然而，在语义分割的应用研究中仍然有限。尽管一些研究 [6] 最近提出来解决高层次视觉问题，但它们需要有标注的数据来训练模型。

文章引入了去噪扩散概率模型 (DDPM) 作为主要研究方向，该模型用于建模数据分布和采样多样性图像。DDPM 是一种生成模型，通过学习从高斯噪声到目标分布的逆扩散过程，成功应用于多个低级计算机视觉任务，如超分辨率、修复和着色。文章还指出，尽管 DDPM 在这些低级任务中表现出色，但在高级视觉任务，如语义分割，其应用尚未得到深入研究。因此，文章提出了一种扩散对抗表示学习 (DARL) 模型，将扩散模型与对抗学习相结合，以用于自监督血管分割任务。

2.2 自监督血管分割

对于血管分割任务来说，由于血管具有复杂的结构和大量微小分支，难以获得用于监督学习的精细标签。传统的监督学习方法需要大量标记的数据进行训练，而无监督方法在处理医学图像时由于复杂的背景结构和低对比度通常效果较差。尽管这种标签稀缺问题可以通过半监督或无监督学习来缓解，但在不使用任何监督的情况下，以合理性能分割微小血管的方法相对较少。事实上，最近使用自然图像训练的无监督学习方法在未见数据集上具有很强的泛化能力 [7] [8] [9]，因此它们可以轻松地适应医学图像分割任务。然而，由于血管造影具有混乱的背景因素和复杂的血管结构等独特特征，任何为自然图像分割设计的无监督方法在应用于噪声血管造影图像的血管分割时可能性能下降。作为无监督学习的一种形式，自监督学习也被引入到各种医学图像分割任务中，利用从数据本身生成的自监督标签高效地学习目标表示，并展现出其潜力 [10] [11] (Mahmood 等, 2019; Ma 等, 2021; Oh & Ye, 2021)。具体而言，Ma 等人 [10] (2021) 引入了一个端到端的对抗学习框架，采用 CycleGAN 的结构 [12] 进行血管分割，学习实际的血管造影生成，将分形引导的伪标签添加到背景图像中。然而，用于合成血管生成的简单算术操作通常无法生成逼真的伪血管图像，因此使用不真实的合成图像训练对抗网络难以产生最佳的分割性能。

3 本文方法

通过提出 DARL 模型，文章致力于结合扩散模型的生成能力和对抗学习的优势，以实现自监督血管分割任务，并在实验中展示了该方法在不同数据集上显著优于现有方法的性能。

3.1 本文方法概述

本文提出了一种新颖的扩散对抗表示学习 (DARL) 模型, 如图 4 所示, 专门设计用于医学图像的自监督血管分割。该模型关注的是血管造影图像, 这些图像是在向血管注射对比剂之前和之后拍摄的, 并且还包括背景图像。由于扫描时间的不同, 这两种图像具有不同的对比度, 并且由于患者移动而无法对齐。因此, 如图 1 所示, 该文献的 DARL 模型是在未配对的血管造影图像和背景图像上进行训练的。它结合了去噪扩散概率模型和对抗学习, 并将其应用于血管分割。该模型通过自监督学习, 使用扩散模块学习背景信号, 以便生成模块可以有效地提供血管表示。通过对抗学习, 该模型估计合成的假血管图像以及血管分割掩模, 进一步捕获有关血管的语义信息。

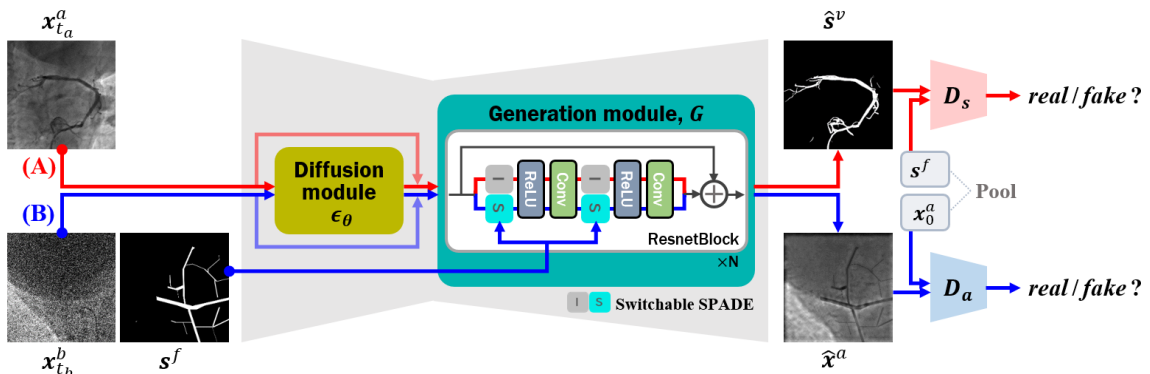


图 1. 自监督血管分割的扩散对抗表示模型示意图。

在路径 (A) 中, 给定真实的噪声血管造影图像 $x_t^A a$, 模型会估计血管分割掩码 x^v 。在路径 (B) 中, 给定一个真实的噪声背景图像 $x_t^B b$ 和一个类似血管的分形掩膜 s^f , DARL 模型生成一个合成血管成像图像 x^a 。

3.2 扩散模块

文中所用到的扩散模块 (Diffusion Module) 是用于估计潜在特征的关键组成部分。具体来说, 扩散模块共包含两个过程: (1) 正向扩散过程: 对输入数据进行噪声添加, 并通过逆扩散过程进行去噪。在给定的时间步骤内, 通过逐渐添加噪声, 模块能够学习从高斯噪声分布到目标分布的逆过程。这有助于模块从输入数据中提取有用的特征。(2) 逆向扩散过程: 扩散模块采用逆扩散过程来学习图像的分布并估计输入数据的有意义的潜在特征。逆扩散是指从目标分布生成高斯噪声分布的过程。这个逆过程的学习有助于模块捕获输入数据的关键信息。

扩散模块的训练目标是最小化逆扩散模型的损失函数。该损失函数基于模块的预测和实际逆扩散过程中添加的噪声之间的差异。通过这个目标, 扩散模块能够学习有效的潜在特征表示, 这对于后续生成模块和分割任务非常重要。该文所用的扩散模型的损失函数是遵循 DDPM [1] (Denoising Diffusion Probabilistic Model) 训练的标准损失

$$L_{\text{diff}}(\theta) = \mathbb{E}_{t, x_0, \epsilon} \left[\left\| \epsilon - \theta \left(\sqrt{\alpha_t x_0 + \sqrt{1 - \alpha_t} \epsilon}, t \right) \right\|_2^2 \right] \quad (5)$$

具体而言, DARL 模型由扩散模块 θ 用于估计潜在特征, 生成模块 G 用于估计血管分割掩膜 \hat{s}^v 和合成血管造影图 \hat{x}^a , 以及两个鉴别器 (D_s 、 D_a) 用于区分血管掩膜和血管造影图的真假图像。在生成血管造影图时, 我们向生成模块提供了 Ma 等 [13] (2021 年) 提出的血管样式分形掩膜 s_f , 以执行基于语义布局的图像合成。此外, 为了有效估计分割地图和血管造影图, 我们设计了生成模块, 其中包含新颖的可切换 SPADE 层, 其中 SPADE [14] 有助于符合当前任务的语义图像合成。

3.3 生成模块

生成模块 (Generation Module) 利用从扩散模块获得的潜在特征, 生成模块估计血管分割掩膜 s^v 和合成的血管造影图像 x^a 。具体来说, 生成模块采用可切换的 SPADE 层的方法, 该方法基于空间自适应标准化 [14] (Spatially-Adaptive Denormalization, SPADE), 用于根据语义掩膜进行图像合成。生成模块输入噪声背景图像 (Background images) 和血管样式的分形掩膜 (Vessel-like fractal mask), 然后生成合成的血管图像。这种方法通过在生成图像时考虑语义布局, 有助于更有效地生成包含血管结构的合成图像。文章中还引入了可变换的 SPADE 层的概念, 这是 SPADE 层的变体。SPAED 层在图像合成中广泛用于根据语义掩膜规范化特征图。可变换的 SPAED 层的特点是能够根据输入是否有语义掩膜来选择性地进行规范化, 具有更强的灵活性。

生成模块的训练目标涉及两个路径: (A) 给定真实的血管图像, 生成模块通过扩散模块的潜在特征估计血管分割掩膜; (B) 给定真实的背景图像和血管样式的分形掩膜, 生成模块生成合成的血管图像。训练目标包括扩散损失 (diffusion loss)、对抗损失 (adversarial loss)

以及循环一致性损失（cycle consistency loss）等。损失函数如式 (2)：

$$L_G(\theta, G, D_s, D_a) = L_{\text{diff}}(\theta) + \alpha L_{G_{\text{adv}}}(\theta, G, D_s, D_a) + \beta L_{\text{cyc}}(\theta, G) \quad (6)$$

$L_{\text{diff}}(\theta)$ 是扩散损失，用于衡量生成的图像与真实图像之间的差异。 $L_{G_{\text{adv}}}(\theta, G, D_s, D_a)$ 是对抗生成损失，衡量生成模块欺骗判别器的能力。 $L_{\text{cyc}}(\theta, G)$ 是循环一致性损失，用于保持生成的图像在经过两个路径（A 和 B）的循环操作后保持一致性。这个总体损失函数的目标是通过平衡扩散损失、对抗生成损失和循环一致性损失来训练生成模块和扩散模块。参数 α 和 β 是用于平衡不同损失项的超参数。

4 复现细节

4.1 与已有开源代码对比

本文虽然开源了代码，但却未提供生成血管样式分形掩模（Vessel-like fractal mask）的具体方法。在对这一部分代码进行复现时，我参考了 [13]，并进行了相应的实现。然而，在复现的过程中，我发现生成的血管样式掩膜过于单一，而且生成的样式并不完全符合血管本身的固有拓扑结构。鉴于这一问题，我进一步进行了创新性的工作，提出了一种改进的生成模型，旨在生成更多样、更真实的血管样式掩膜。

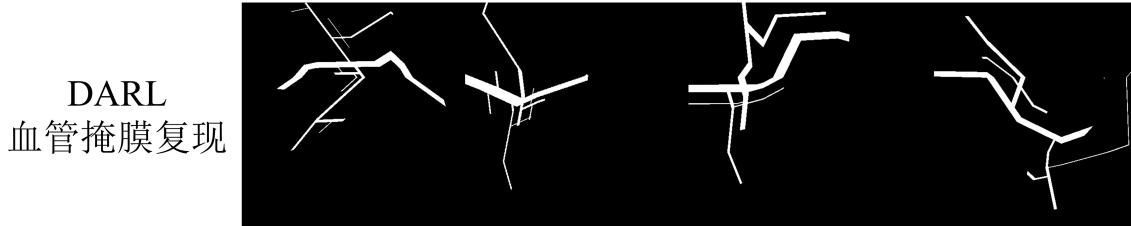


图 2. DARL 复现掩膜生成图

4.2 创新点

在改进的过程中，为了解决上述问题，我采用了一种全新的方法来生成更丰富和更真实的血管样式掩膜。在这个过程中，我融合了 CotrolNet [15] 中的技术，并对生成算法进行了调整，通过微调并控制 Stable Diffusion [16]，以期望更好地捕捉血管的复杂结构和拓扑特征。

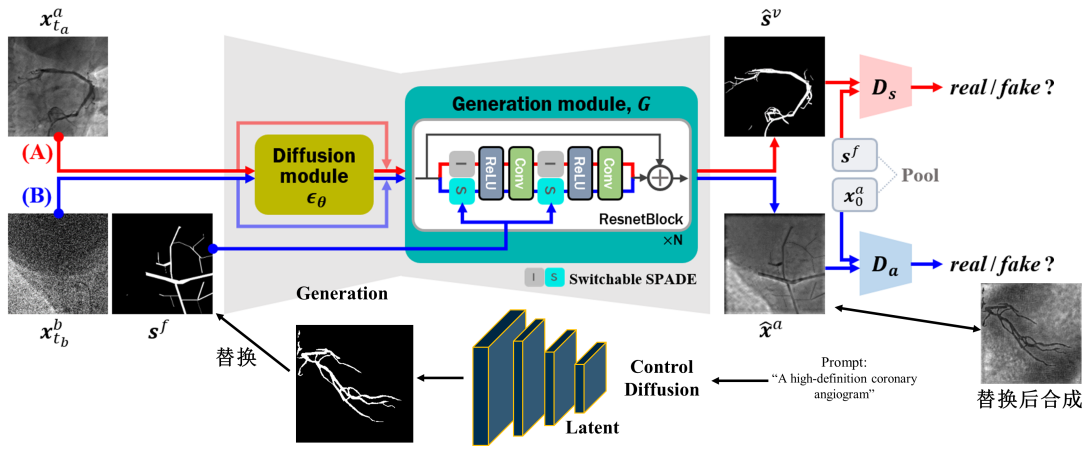


图 3. 模型改进示意图

我的创新性工作在于提出了一种改进的生成模型，该模型能够更准确地生成具有更真实血管形态的掩膜。通过引入 CotrolNet 中的技术，我成功改善了血管样式掩膜的生成过程，使其更富有变化和真实感。此外，通过对 Stable Diffusion 的微调和控制，生成的掩膜能够更好地反映血管的复杂性，包括微小分支和拓扑关系。这种调整不仅提高了生成模型的性能，还使其更适用于医学图像分割等任务。这一改进为血管分割领域带来了更可靠和高效的生成模型，为未来的医学影像分析提供了更精准的工具。



图 4. 改进后的掩膜生成图

通过后续实验的验证，这一创新不仅显著提高了生成血管样式的准确性，同时使得生成的掩膜更适用于医学图像分割任务。这表明我的改进不仅在生成模型的性能方面取得了显著进展，而且对于解决医学影像处理中的实际问题具有实质性的影响。通过对生成算法的巧妙改进，我期望为医学影像处理领域提供更可靠和实用的工具。这一创新有望促进相关研究的进展，为更广泛的医学图像分析提供更精准、高效的解决方案，有望在未来为医学影像处理领域带来新的突破。

5 实验结果分析

图 3 展示了在冠状动脉造影数据集上 DARL 方法和我所改进的方法的血管分割掩膜。方法的定量评估结果呈现在表 1 中。对结果的分析如下。与 DARL 的比较：当我们将所提出的方法与基准方法进行比较时，我们的模型更准确地分割了包括微小分支在内的血管结构。同时，如表 1 所示，与原论文的方法相比，改进后的模型始终以较大的优势实现了 SOTA 性能。具体来说，我们的改进在精度分数上呈现出显著提高，这进一步验证了改进后的分割掩膜更贴近血管的拓扑结构。这种改进使得模型更有效地区分前景血管结构，并消除噪声背景中的假阳性信号，突显了在血管分割任务中的优势。

表 1. XCAD 血管造影图像血管分割的定量评价结果

XCAD 测试数据集 *			
	Iou	Dice	Precision
DARL	0.4708+-0.0776	0.6363+-0.0727	0.7013+-0.1166
DARL 复现	0.4478+-0.0800	0.6143+- 0.0784	0.7137+-0.1254
改进方法	0.4628+-0.0772	0.6246+-0.0769	0.7007+-0.1054

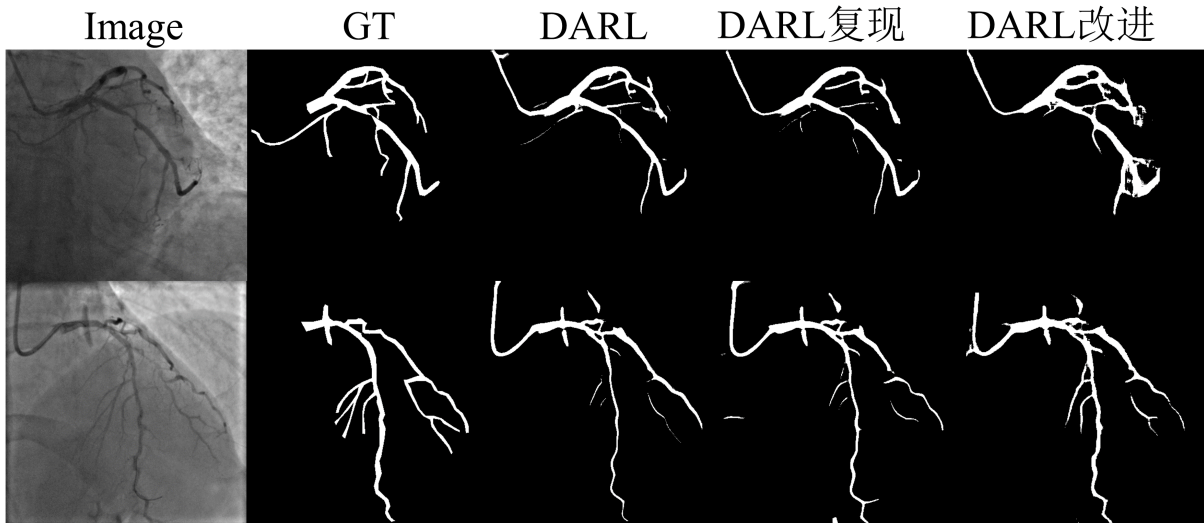


图 5. 血管造影图像血管分割的视觉对比结果

6 总结与展望

这篇文章介绍了一种新的方法，即扩散对抗表示学习 (DARL)，用于自监督血管分割。文章指出，在医学图像中进行血管分割是诊断血管疾病和治疗规划中的重要任务。传统的监督学习方法需要大量的标注数据，而无监督学习方法面临背景结构复杂，使神经网络难以有效分割血管的问题。为了解决这一问题，文章引入了 DARL 模型，该模型利用具有对抗学习的去噪扩散概率模型，并将其应用于血管分割任务。特别是，对于自监督血管分割，DARL 通过扩散模块学习背景信号，让生成模块有效地提供血管表示。通过基于提出的可切换空间自适应去规范化的对抗学习，该模型估计合成的伪血管图像以及血管分割掩膜，进一步使模型

捕捉与血管相关的语义信息。在整个文献的复现过程中，我分析了其生成的血管样式掩膜的样式存在着过于单一，而且生成的样式并不完全符合血管本身的固有拓扑结构。鉴于这一问题，我进一步进行了创新性的工作，提出了一种改进的生成模型，旨在生成更多样、更真实的血管样式掩膜。通过后续的实验说明，这一创新不仅提高了生成血管样式的准确性，还使得生成的掩膜更适用于后续的医学图像分割任务。

未来，我计划深入研究血管结构自身的拓扑性质，并将这一关键信息融入整个分割网络。通过充分挖掘血管系统的拓扑结构，我期望提高分割模型对冠状动脉血管的精确性，进一步优化医学图像分割的性能和准确性。这一方法有望进一步优化分割算法，使其更好地适应不同解剖结构和病理变化，为医学图像分析领域的进一步发展贡献一份力量。我的研究将聚焦于整合拓扑性质，推动自监督分割方法在医学图像处理中的应用，为未来的临床实践提供更准确和可靠的辅助工具，为医学影像领域的创新和进步注入新的动力。

参考文献

- [1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS’20, Red Hook, NY, USA, 2020. Curran Associates Inc.
- [2] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12413–12422, June 2022.
- [3] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2023.
- [4] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models, 2022.
- [5] Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models, 2022.
- [6] Tomer Amit, Tal Shaharbandy, Eliya Nachmani, and Lior Wolf. Segdiff: Image segmentation with diffusion probabilistic models, 2022.
- [7] Luke Melas-Kyriazi, Christian Rupprecht, Iro Laina, and Andrea Vedaldi. Deep spectral methods: A surprisingly strong baseline for unsupervised semantic segmentation and localization. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8354–8365, 2022.
- [8] Euijoon Ahn, Dagan Feng, and Jinman Kim. A spatial guided self-supervised clustering network for medical image segmentation. page 379–388, Berlin, Heidelberg, 2021. Springer-Verlag.
- [9] Mickaël Chen, Thierry Artières, and Ludovic Denoyer. *Unsupervised Object Segmentation by Redrawing*. Curran Associates Inc., Red Hook, NY, USA, 2019.
- [10] Yuxin Ma, Yang Hua, Hanming Deng, Tao Song, Hao Wang, Zhengui Xue, Heng Cao, Ruhui Ma, and Haibing Guan. Self-supervised vessel segmentation via adversarial learning. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7516–7525, 2021.
- [11] Yujin Oh and Jong Chul Ye. Cxr segmentation by adain-based domain adaptation and knowledge distillation. page 627–643, Berlin, Heidelberg, 2022. Springer-Verlag.

- [12] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.
- [13] Yuxin Ma, Yang Hua, Hanming Deng, Tao Song, Hao Wang, Zhengui Xue, Heng Cao, Ruhui Ma, and Haibing Guan. Self-supervised vessel segmentation via adversarial learning. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7516–7525, 2021.
- [14] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2332–2341, 2019.
- [15] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3836–3847, October 2023.
- [16] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.