PPS-Ctrl：可控的模拟到现实转换用于结肠镜深度估计

Xinqi Xiong[1], Andrea Dunn Beltran[1], Jun Myeong Choi[1], Marc Niethammer[2], and Roni Sengupta[1]

[1] University of North Carolina at Chapel Hill, Chapel Hill, USA
[2] University of California, San Diego, USA

Abstract. 准确的深度估计提升了内窥镜导航和诊断，但在临床环境中获取真实深度是具有挑战性的。合成数据集常被用于训练，但域间差距限制了对真实数据的泛化。我们提出了一种新颖的图像到图像翻译框架，该框架在生成临床数据的真实纹理的同时保留结构。我们创新性地结合了 Stable Diffusion 和 ControlNet，以从每像素阴影（PPS）地图中提取的潜在表示为条件。PPS 捕捉到表面光照效果，比深度图提供更强的结构约束。实验表明，我们的方法比基于 GAN 的 MI-CycleGAN 产生更真实的翻译，并改进了深度估计。我们的代码可在 https://github.com/anaxqx/PPS-Ctrl 上公开获取。

1 引言

结直肠癌是全球最致命的恶性肿瘤之一 [1]，结肠镜检查仍然是检测肠道病变（如息肉和结肠癌）的金标准 [17]。在结肠镜检查中准确的深度估计可以增强空间理解和程序指导，对于开发辅助导航和诊断工具起着至关重要的作用。提升估计精度有助于引导探索未调查的区域 [23]，检测盲点 [12] 和息肉 [21]，以及机器人导航系统 [23,3]。

由于大多数内窥镜不支持高质量的 3D 感测，临床结肠镜检查数据集通常缺乏真实深度标注。因此，大多数基于深度学习的深度估计模型依赖于合成数据集，因为在那里可以轻松生成深度标注 [5,10,15]。然而，合成数据集通常缺乏在真实临床数据中存在的细致纹理、真实光照条件和复杂解剖结构，如覆盖粘液的表面和息肉变化，这使得在合成数据上训练的模型难以泛化到真实临床数据。

模拟到现实的图像翻译已被探索以减轻这一差距，主要通过基于 GAN 的方法 [22,10]。然而，GAN 在模式崩溃、结构不一致和训练不稳定方面存在困难。扩散模型提供了一种更稳定的替代方案 [19]，但它们需要更强的约束来在复杂环境如内窥镜中保持结构。

我们提出了一种新颖的模拟到真实图像翻译框架，利用扩散模型代替 GAN，并引入每像素阴影（PPS）图 [15] 作为结构约束，而不是深度图，用于通过 ControlNet 引导稳定扩散模型。我们进一步引入了一种编码器-解码器架构，从 PPS 图提取更有意义的结构信息作为 ControlNet 的输入。每像素阴影图计算由深度图定义的表面上每个点的光照效果，并考虑到靠近内窥镜且面向内窥镜的点应比远离的点接收到更大的光强。使用 PPS 图而不是深度进行显性条件处理，可以在提高现实感的同时，实现翻译图像与源图像结构之间更好的对齐。

总而言之，我们工作的贡献如下：

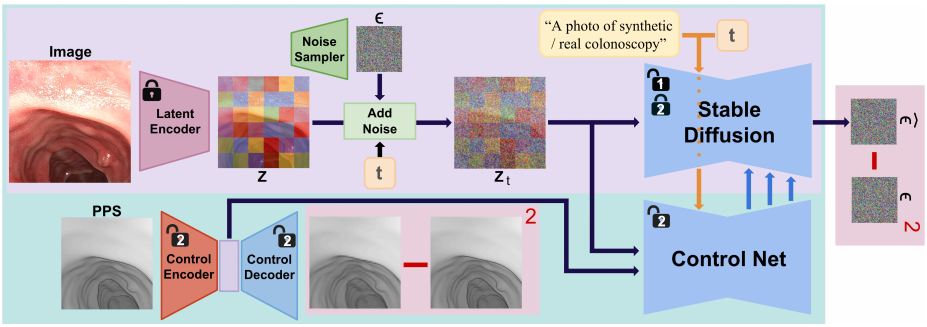

Fig. 1. Our method takes a depth map from any synthetic colon datasets as input and generates textures similar to real endoscopy videos while preserving the depth in the generated image. Our proposed pipeline consists of a Stable Diffusion model that can capture image statistics of both real and synthetic domains using text prompts (trained in Stage 1), and a ControlNet that guides the diffusion model to perform depth-preserving image generation through a latent encoding of a Per-Pixel Shading (PPS) map (trained in Stage 2).

- 我们表明，每像素着色（PPS）图比深度图更适合作为模拟到真实转换的结构约束。
- 我们提出了一种新颖的结构保持仿真到现实图像转换框架，该框架使用 Stable Diffusion 和 ControlNet，并结合了一种编码器-解码器架构，该架构经过训练以从 PPS 地图中提取有意义的结构信息以指导 ControlNet。
- 我们展示了相比于最新的基于 GAN 的方法 MI-CycleGAN [22]，在逼真度和深度保留方面的改进，这通过在 C3VD（幻影）→ Colon10K（临床）翻译中 FID 指标 40 % 的提升以及在 C3VD 数据集上的下游深度预测准确性的 20 % 提升所证明，对于 SimCol3D（模拟）→ C3VD（幻影）翻译。

2 相关工作

结肠镜数据集来源于临床操作 [2,14]、模型数据 [4] 和完全合成的数据 [4,28]。尽管合成和模型数据集提供了深度和姿态标注，但它们缺乏诸如动态光照、镜面反射和组织变形等现实世界的复杂性。使用基于 GAN 的方法 [9] 进行的不成对图像翻译结合深度图 [10,29] 或互信息 [22] 已被探索以保持结构完整性。然而，GAN 可能不稳定，容易模型崩溃，并可能表现出有限的多样性。最近，由于其改进的模式覆盖和稳定性 [19]，稳定扩散被应用于临床结肠镜合成 [8,24] 和外科数据的模拟到现实翻译中 [11,20]，但保持结构完整性仍然是一个挑战 [6,26,7]，我们在本文中通过设计一个新颖的 PPS 控制的扩散模型来解决这一问题。

3 方法

我们的目标是执行内窥镜图像的合成到真实图像翻译，同时利用一种新颖的 Stable Diffusion [19] 和 ControlNet [27] 改进方法保留原始合成图像的深度，如

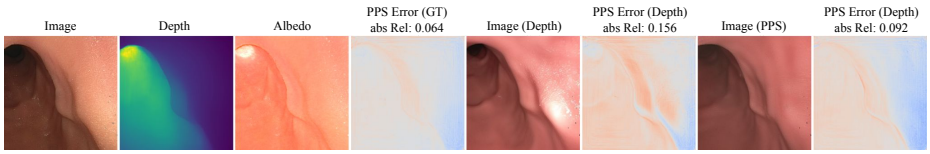

Fig. 2. 我们计算了真实的每像素阴影图与原始图像（第 4 列）、通过深度条件生成的图像（第 6 列）和 PPS 条件生成的图像（第 8 列）之间的误差，以显示 PPS 条件显著减少了结构不一致性，并与原始图像非常匹配。

图 1 所示。我们首先使用文本条件来区分它们，在合成源和真实目标领域进行微调 Stable Diffusion，如 Sec. 3.1 所述。在 Sec. 3.2 中，我们概述了基于潜在逐像素阴影（PPS）特征的 ControlNet [27] 的新颖改进方法，这些特征是通过自动编码器获得的，显著改善了现有技术的深度保留图像翻译质量。

3.1 使用文本在稳定扩散中进行域分离

当在多个域上的图像上天真地微调 Stable Diffusion (SD) 时，我们观察到它最终会在两个不同领域之间混合特征，从而降低适应的有效性。因此，我们在训练和推理过程中通过基于文本的领域分离来精炼源域和目标域的 SD。为实现这一目标，领域分离通过文本条件控制：p_{source} = " 合成结肠镜检查的照片" 和 p_{target} = " 真实结肠镜检查的照片"，其中 p_{source} 和 p_{target} 分别是对应于合成和真实结肠镜检查图像的文本提示。通过使用文本提示强制领域分离，我们确保 SD 生成模型为每个领域保持不同的图像特征。最后，我们使用标准的去噪扩散模型损失函数，在时间步 t 使用参数 θ^S 、文本提示 p 和图像潜代码 z_t 训练 SD 模型：

$$\mathcal{L} = \mathbb{E}_{z_t,\epsilon\sim\mathcal{N}(0,1),t,p}\left[\|\epsilon - \epsilon_{\theta^S}(z_t,t,p)\|_2^2\right] . \tag{1}$$

3.2 用于深度保持的 PPS-Ctrl

现有的用于内容保留图像生成的技术 [6,26] ，通常使用基于深度图的 ControlNet [27] 来指导 Stable Diffusion 生成保留深度的高质量图像。然而，在内窥镜检查中，这些技术会对离摄像机和光源很近（近距离）或很远（远距离）的区域生成不一致的表面颜色，导致图像的深度感知降低。我们注意到，如图 2 所示，计算的每像素阴影场，即每个像素表面法线与入射光的内积，在近距离和远距离区域显著不同于生成的图像强度。为了解决这个问题，我们提出了基于每像素阴影（PPS）的调节，它能够有效捕捉细尺度的光照变化，并提供更加稳定、几何一致的调节条件。

每像素着色（PPS）已被应用于近场照明光度计立体 3D 重建 [13] ，其中光源靠近表面放置，并且在内窥镜 3D 估计中也被证明效果良好 [15,3,18] 。基于 PPS 在基于光照的 3D 重建中的成功应用，我们提出使用 PPS 表示作为 ControlNet 的条件输入，以指导稳定扩散模型在生成逼真纹理的同时保持源图像的深度。正式来说，给定一个合成深度图 $D(u,v)$ 和相机内参 K ，我们估计法向量为 $N(\boldsymbol{x}) = \frac{\partial \boldsymbol{x}/\partial u \times \partial \boldsymbol{x}/\partial v}{|\partial \boldsymbol{x}/\partial u \times \partial \boldsymbol{x}/\partial v|}$ ，其中 $\boldsymbol{x} = K^{-1}[u,v,D(u,v),1]^\top$ 是对应于图像平面中像素 (u,v) 的 3D 表面点。对于一个位于 p_c 的相机和光源，每个点 $\boldsymbol{x}$

接收方向为 $L^d(\boldsymbol{x}) = \frac{\boldsymbol{x}-\boldsymbol{p}_c}{\|\boldsymbol{x}-\boldsymbol{p}_c\|_2}$ 的光，其强度通过平方反比减弱，假设没有角度衰减，类似于 [13,15,3] 。最后，我们可以将每个表面点的每像素着色表示为：

$$\mathcal{PPS}(\boldsymbol{x}) = L^a(\boldsymbol{x}) \times (L^d(\boldsymbol{x})^T \boldsymbol{N}(\boldsymbol{x})) . \tag{2}$$

为了将 $\mathcal{PPS}$ 整合到 ControlNet 中，我们使用一个可学习的控制编码器 $E^C(\cdot)$ ，将 $\mathcal{PPS}$ 编码成特征图 （$f = E^C(\mathcal{PPS})$ ）。为了进一步确保基于 PPS 的条件的保留，我们引入了一个额外的控制解码器 $D^C(\cdot)$ ，从特征图 f 中重建条件 $\mathcal{PPS}$ 。

$$\mathcal{L}_{\mathcal{D}} = \left[\left\| D^C(E^C(\mathcal{PPS})) - \mathcal{PPS} \right\|_2^2 \right] . \tag{3}$$

控制编码解码器架构能够防止特征退化并提取最有意义的几何特征，在后续的 ControlNet 降噪阶段中保留控制信号。我们在表格 1 中训练翻译数据，通过训练验证了控制编码解码器对下游深度图预测任务的重要性。最后，我们联合训练 ControlNet （θ^C ），输入包括 PPS 特征图 f 、时间步长为 t 的图像潜在代码 z_t 以及文本提示 p ，同时使用如下损失函数进行控制编码解码器的训练：

$$\mathcal{L} = \mathcal{L}_D + \mathbb{E}_{z_t, f, \epsilon \sim \mathcal{N}(0,1), t, p} \left[\|\epsilon - \epsilon_{\theta^C}(z_t, f, t, p)\|_2^2 \right] . \tag{4}$$

总结一下，我们使用以下步骤进行模拟到真实的转换：
• 训练, 阶段 1：在模拟和真实领域上使用特定于领域的文本提示 p_{source} 和 p_{target} 微调 SD，优化公式 1。
• 训练, 阶段 2：通过优化公式 4 在来自模拟领域的图像和 PPS（或深度）上训练 ControlNet 和编码器-解码器，使用对应于模拟领域的文本提示的 SD。
• 测试：从模拟领域采样一个深度图，使用采样的深度计算 PPS，然后运行 ControlNet + SD，使用对应于真实领域的文本提示 p_{target} 生成一张既保留输入模拟领域图像深度又类似于真实领域的图像。

3.3 实现细节

我们的控制解码器 D^C 由 4 个残差块组成，并采用控制编码器 E^C 的转置架构。我们以 2e-6 的学习率微调 Stable Diffusion v2 模型，并以 1e-4 的学习率训练 ControlNet 10000 步。我们对两个模型都使用批大小 16 和 AdamW 优化器。在推理时，我们应用 DDIM 调度器并采样 20 步。所有输入被调整为 512 × 512。所有实验都使用 4xA6000 GPU 进行训练。

4 结果

4.1 实验装置

我们考虑两组不同的实验。(i) SimCol3D → C3VD: 用于在 C3VD 上数值评价深度图预测准确性和图像翻译质量；(ii) C3VD → Colon10K: 用于在真实数据上定性评价图像翻译和深度预测质量。对于 ControlNet 训练，我们重用源数据集的训练拆分。
数据集我们使用一个模拟的结肠数据集 SimCol3D [16] ，一个幻影结肠数据集 C3VD [4] ，以及一个临床数据集 Colon10K [14] ，来执行从源到目标图像的翻译，然后在目标域上进行深度图预测。

Table 1. 我们通过在 SimCol3D → C3VD 翻译数据上使用各种技术训练来评估 C3VD 数据集上的深度估计准确性。我们考虑没有翻译时的下限表现，即在 SimCol3D 上训练，以及在 C3VD 自身上训练时的上限表现。我们还进行了额外的消融研究，以证明对深度图进行 Per-Pixel Shading(PPS) 条件处理的有效性，以及使用控制编码器-解码器架构以提取更有意义的潜在编码的有效性。

Training Data	Condition	DepthAnything		
		RMSE ↓	Abs_{rel} ↓	$\delta < 1.1$ ↑
SimCol3D	Lower Bound	4.753	0.117	0.592
SimCol3D → C3VD	MI-CycleGAN [22]	4.662	0.100	0.619
SimCol3D → C3VD	Ours-Depth	4.444	0.095	0.664
SimCol3D → C3VD	Ours-Depth w. D^C	4.099	0.093	0.672
SimCol3D → C3VD	Ours-PPS	4.135	0.095	0.663
SimCol3D → C3VD	Ours-PPS w. D^C	3.740	0.088	0.692
C3VD	Upper Bound	2.277	0.055	0.839

指标我们使用源到目标翻译数据与原目标数据之间的 Frecet Inception Distance（FID）来评估图像翻译质量。我们使用均方根误差（RMSE）、绝对相对误差（AbsRel）以及在 10 % 以内的实际深度值的像素百分比 $\delta < 1.1$ 来评估深度估计性能。
基线我们与一种先进的用于内窥镜的深度保持图像到图像翻译算法 MI-CycleGAN [22] 进行比较。

4.2 深度估计

为了评估我们图像翻译的结构保留能力，我们比较了多个经过微调的 DepthAnything [25] 模型的性能。更具体地说，我们使用与原始源数据中的相应深度图配对的翻译图像训练每个模型，然后在目标域的图像上进行测试。在表 1 中，我们将使用我们翻译数据进行训练的深度估计准确性与使用现有方法 MI-CycleGAN [22] 为 SimCol3D → C3VD 数据进行的翻译进行比较。我们还评估了仅使用未翻译的 SimCol3D 数据训练并在 C3VD 数据上评估的模型（作为下限）以及在 C3VD 数据上训练和测试的模型（作为上限）。我们观察到，我们的方法在 RMSE 方面显著优于 MI-CycleGAN [22] 提高了 20 % ，并且在 Abs_{rel} 和 $\delta < 1.1$ 上有类似的改进。

在缺乏临床数据的真实深度标注的情况下，我们定性评估将 C3VD → Colon10K 翻译对深度预测的影响（图 **??** ）。当在我们方法翻译的数据上训练时，相比于未翻译的数据，生成的深度图在靠近摄像头的区域更有效地捕捉到了结肠的轮廓（在倾斜和正面视图中均有突出）。此外，当将使用我们翻译的图像训练的模型与使用 MI-CycleGAN [22] 翻译的模型进行比较时，我们观察到在准确表示结肠较远区域方面有类似的改善（特别是在管道视图中尤为明显）。

4.3 图像转换

我们在表 2 中展示了领域转换的有效性，其使用 FID 衡量的图像翻译质量。针对我们的方法，MI-CycleGAN [22] ，以及无翻译，我们考虑了两个不同的翻译

场景（SimCol3d → C3VD 和 C3VD → Colon10K）。我们的翻译增强了纹理真实感（图 **??** ），减少了过度反光和褪色等伪影，并比 MI-CycleGAN [22] 基线更好地保留了源图像结构。

Table 2. 我们使用 FID（越低越好）来评估 SimCol3D → C3VD 和 C3VD → Colon10K 在三种方法中的图像翻译质量：无翻译、MI-CycleGAN [22] 翻译和我们的翻译。请注意，这只衡量纹理翻译质量，忽略结构保留。

Translation	SimCol3D → C3VD			C3VD → Colon10K		
Algorithm	No Trans.	MI-cGAN	Ours	No Trans.	MI-cGAN	Ours
FID ↓	0.545	0.591	0.437	0.527	0.498	0.297

4.4 消融研究

我们展示了我们关键贡献的有效性：(i) 使用 PPS 图代替深度图进行 ControlNet 条件设置，以及（ii）使用控制编码-解码架构在 ControlNet 中提取更有意义的结构保留潜在编码，而不仅仅使用潜在编码器。我们在 SimCol3D → C3VD 上使用下游深度估计任务，并在表 1 中展示了结果。结果表明，用 PPS 替换深度图一致地改善了 RMSE、绝对相对误差以及 $\delta < 1.1$ 指标，确认 PPS 提供了更丰富的几何提示以更好地保留结构。将解码器 D^C 集成到 ControlNet 进一步增强了在所有条件下的深度估计性能。结合这两项进步，我们在 D^C 设置下使用 Ous-PPS 获得了最佳性能。

尽管我们的方法在仿真到真实转移过程中学会了添加由于粘液层引起的镜面反射，但我们不能控制镜面反射的数量，因为我们没有使用基于物理的材料建模来显式地对其进行建模。我们的模型也难以保持或添加细微特征，例如小血管（见图 **??** ，第 4 行，第 1 列），这可以在未来通过显式条件解决。

5 结论

我们提出了 PPS-Ctrl，这是一种使用 Stable Diffusion 与 ControlNet 的模拟到真实图像翻译框架，基于像素级阴影（PPS）以增强结肠镜图像的结构保真度和真实感。我们的实验表明，PPS-Ctrl 在实现照片般真实的结果方面优于现有的纹理转移方法。此外，通过在我们翻译的图像上训练深度估计模型，我们观察到在域外临床数据上性能有所提升，这凸显了我们的方法在弥合合成与真实内窥镜数据集之间域间差距的有效性。

6 致谢

这项工作由国家卫生研究院（NIH）项目 # 1R21EB035832 “下一代内镜视频的 3D 建模”支持。

References

1. Araghi, M., Soerjomataram, I., Jenkins, M., Brierley, J., Morris, E., Bray, F., Arnold, M.: Global trends in colorectal cancer mortality: projections to the year 2035. International journal of cancer **144**(12), 2992–3000 (2019)
2. Azagra, P., Sostres, C., Ferrández, Á., Riazuelo, L., Tomasini, C., Barbed, O.L., Morlana, J., Recasens, D., Batlle, V.M., Gómez-Rodríguez, J.J., et al.: Endomapper dataset of complete calibrated endoscopy procedures. Scientific Data **10**(1), 671 (2023)
3. Beltran, A.D., Rho, D., Niethammer, M., Sengupta, R.: Nfl-ba: Improving endoscopic slam with near-field light bundle adjustment. arXiv preprint arXiv:2412.13176 (2024)
4. Bobrow, T.L., Golhar, M., Vijayan, R., Akshintala, V.S., Garcia, J.R., Durr, N.J.: Colonoscopy 3d video dataset with paired depth from 2d-3d registration. Medical image analysis **90**, 102956 (2023)
5. Cheng, K., Ma, Y., Sun, B., Li, Y., Chen, X.: Depth estimation for colonoscopy images with self-supervised learning from videos. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 119–128. Springer (2021)
6. Cheng, T.Y., Sharma, P., Markham, A., Trigoni, N., Jampani, V.: Zest: Zero-shot material transfer from a single image. In: European Conference on Computer Vision. pp. 370–386. Springer (2024)
7. Choi, J.M., Wang, A., Peers, P., Bhattad, A., Sengupta, R.: Scribblelight: Single image indoor relighting with scribbles. arXiv preprint arXiv:2411.17696 (2024)
8. Du, Y., Jiang, Y., Tan, S., Wu, X., Dou, Q., Li, Z., Li, G., Wan, X.: Arsdm: colonoscopy images synthesis with adaptive refinement semantic diffusion models. In: International conference on medical image computing and computer-assisted intervention. pp. 339–349. Springer (2023)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
10. Jeong, B.H., Kim, H.K., Son, Y.D.: Depth estimation of endoscopy using sim-to-real transfer. arXiv preprint arXiv:2112.13595 (2021)
11. Kaleta, J., Dall' Alba, D., Płotka, S., Korzeniowski, P.: Minimal data requirement for realistic endoscopic image generation with stable diffusion. International journal of computer assisted radiology and surgery **19**(3), 531–539 (2024)
12. Kim, B.S., Cho, M., Chung, G.E., Lee, J., Kang, H.Y., Yoon, D., Cho, W.S., Lee, J.C., Bae, J.H., Kong, H.J., et al.: Density clustering-based automatic anatomical section recognition in colonoscopy video using deep learning. Scientific Reports **14**(1), 872 (2024)
13. Lichy, D., Sengupta, S., Jacobs, D.W.: Fast light-weight near-field photometric stereo. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12612–12621 (2022)
14. Ma, R., McGill, S.K., Wang, R., Rosenman, J., Frahm, J.M., Zhang, Y., Pizer, S.: Colon10k: a benchmark for place recognition in colonoscopy. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 1279–1283. IEEE (2021)
15. Paruchuri, A., Ehrenstein, S., Wang, S., Fried, I., Pizer, S.M., Niethammer, M., Sengupta, R.: Leveraging near-field lighting for monocular depth estimation from

endoscopy videos. In: European Conference on Computer Vision. pp. 473–491. Springer (2024)
16. Rau, A., Bano, S., Jin, Y., Azagra, P., Morlana, J., Kader, R., Sanderson, E., Matuszewski, B.J., Lee, J.Y., Lee, D.J., et al.: Simcol3d—3d reconstruction during colonoscopy challenge. Medical Image Analysis **96**, 103195 (2024)
17. Rex, D.K., Schoenfeld, P.S., Cohen, J., Pike, I.M., Adler, D.G., Fennerty, B.M., Lieb, J.G., Park, W.G., Rizk, M.K., Sawhney, M.S., et al.: Quality indicators for colonoscopy. Official journal of the American College of Gastroenterology| ACG **110**(1), 72–90 (2015)
18. Rodríguez-Puigvert, J., Batlle, V.M., Montiel, J., Martinez-Cantin, R., Fua, P., Tardós, J.D., Civera, J.: Lightdepth: Single-view depth self-supervision from illumination decline. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21273–21283 (2023)
19. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695 (2022)
20. Venkatesh, D.K., Rivoir, D., Pfeiffer, M., Speidel, S.: Surgical-cd: Generating surgical images via unpaired image translation with latent consistency diffusion models. arXiv preprint arXiv:2408.09822 (2024)
21. Wan, J., Chen, B., Yu, Y.: Polyp detection from colorectum images by using attentive yolov5. Diagnostics **11**(12), 2264 (2021)
22. Wang, S., Paruchuri, A., Zhang, Z., McGill, S., Sengupta, R.: Structure-preserving image translation for depth estimation in colonoscopy. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 667–677. Springer (2024)
23. Wu, Z., Jin, Y., Qiu, L., Han, X., Wan, X., Li, G.: Toder: Towards colonoscopy depth estimation and reconstruction with geometry constraint adaptation. arXiv preprint arXiv:2407.16508 (2024)
24. Xie, Y., Wang, J., Feng, T., Ma, F., Li, Y.: Ccis-diff: A generative model with stable diffusion prior for controlled colonoscopy image synthesis. arXiv preprint arXiv:2411.12198 (2024)
25. Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H.: Depth anything: Unleashing the power of large-scale unlabeled data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10371–10381 (2024)
26. Zhang, F., You, S., Li, Y., Fu, Y.: Atlantis: Enabling underwater depth estimation with stable diffusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11852–11861 (2024)
27. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3836–3847 (2023)
28. Zhang, S., Zhao, L., Huang, S., Ye, M., Hao, Q.: A template-based 3d reconstruction of colon structures and textures from stereo colonoscopic images. IEEE Transactions on Medical Robotics and Bionics **3**(1), 85–95 (2020)
29. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)