Realistic Endoscopic Synthetic Dataset Generation Through Surgical Simulation and Diffusion Models

Sabina Martyniak[1], Joanna Kaleta[1,2], Diego Dall'Alba[1,3], Szymon Płotka[1,4], Przemysław Korzeniowski[1]

[1]Sano Centre for Computational Medicine, [2]Warsaw University of Technology, [3]University of Verona, [4]University of Warsaw
Keywords: Diffusion Models, surgical simulation, Synthetic Data Generation

1. Introduction

Computer-assisted intervention (CAI) is a research field focused on enhancing the safety, efficiency, and cost-effectiveness of medical procedures by minimizing errors and complications [1]. Within CAI, Laparoscopic Cholecystectomy (LC) has gained significant attention [2]. However, LC presents technical challenges due to limited visibility and the use of laparoscopic instruments, leading to potential complications like bile duct injury (BDI) [3]. To address these issues, generating synthetic data is a promising solution, but bridging the domain gap between synthetic and real data remains a challenge. We propose SimuScope, a pipeline that requires minimal real data, preserves annotation integrity, and has the potential to support the development of robust DL models for surgical assistance.

2. Description of the problem

To address these complexities, CAI systems leveraging Deep Learning (DL) methods have been proposed. These systems rely on deep learning models trained on complex and difficult-to-annotate data. Generating synthetic data is a promising solution, but bridging the domain gap between synthetic and real data remains a challenge.

3. Related work

Recently, several approaches have been proposed for generating synthetic data with realistic characteristics. While GAN-based approaches show potential, they have limitations, such as early convergence of discriminators and instability of adversarial training, leading to mode collapse and reduced diversity in generated data [4]. Diffusion models (DMs) [5] emerge as a promising alternative, surpassing GANs in computer vision tasks. Although multiple generative works exist in the medical and surgical fields, a gap remains in surgical data generation, particularly for fully labeled simulator-based data with accurate and detailed instrument-tissue interactions [6,8]. Based on simulation data, [6] generates large fully labeled realistic endoscopic images, addressing the minimal data requirement. However, this work also relies on a very simple simulator lacking tool-tissue interactions.

4. Solution to the problem

We use a custom simulator built in Unity3D integrated with an XPBD (Extended Position-Based Dynamics) solver implemented in C/C++. This allows real-time soft-tissue simulation with grasping, cutting, clipping, tearing, and thermo-coagulation. Anatomical models include liver, gallbladder, cystic duct and artery, simulated as tetrahedral meshes. Our method involves adding a new style to the SD model and using it to generate realistic images from synthetic ones. We begin by fine-tuning SD based on LoRA [7]. Then, the fine-tuned Laparoscopic Cholecystectomy Stable Diffusion model is employed to generate realistic images. For realistic tissue generation, we use text-guided image-to-image inference with additional control. The input to the model is raw simulator image. The proposed method is depicted in (Fig. 1)

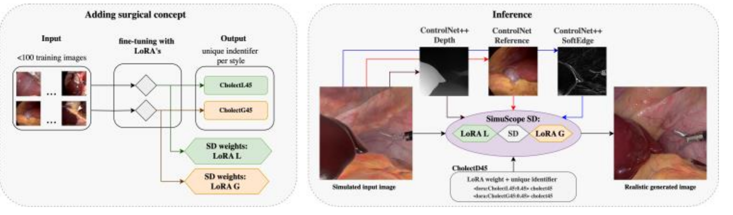

Fig.1. SimuScope uses a Stable Diffusion model, fine-tuned with LoRA and conditioned by ControlNet, to generate realistic surgical images from simulator input.

5. Conclusions and future work

SimuScope successfully generates realistic, diverse, and semantically consistent synthetic surgical images from a lightweight training process using only 100 frames of real data. It achieves a high mean Intersection over Union (mIoU) of 70.65%, confirming excellent preservation of semantic labels. While the Fréchet Inception Distance (FID) is slightly higher than the baseline, the Kernel Inception Distance (KID) of 0.0690 is comparable and acceptable, showing a strong balance between realism and label fidelity. The use of ControlNet is crucial for enhancing texture and color. The main limitation is the lack of temporal coherency between frames; future work will aim to address this by focusing on video-to-video generation.

Acknowledgements. This work is supported by the EU's Horizon 2020 program under grant agreement no. 857533 (Sano) and the International Research Agendas program of the Foundation for Polish Science, co-financed by the European Regional Development Fund. We thank NVIDIA Corporation for donating DGX Clara hardware.

References

1. L. Maier-Hein, M. Eisenmann, D. Sarikaya, K. M¨arz, T. Collins, A. Malpani, J. Fallert, H. Feussner, S. Giannarou, P. Mascagni, et al. *Surgical data science–from concepts toward clinical translation. Medical image analysis*, 76:102306, 2022.
2. C. I. Nwoye, T. Yu, C. Gonzalez, B. Seeliger, P. Mascagni, D. Mutter, J. Marescaux, and N. Padoy. Rendezvous: *Attention mechanisms for the recognition of surgical action triplets in endoscopic videos*. Medical Image Analysis, 78:102433, 2022.
3. T. Tokuyasu, Y. Iwashita, Y. Matsunobu, T. Kamiyama, M. Ishikake, S. Sakaguchi, K. Ebe, K. Tada, Y. Endo, T. Etoh, et al. *Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy. Surgical endoscopy*, 35:1651–1658, 2021.
4. N. Huang, A. Gokaslan, V. Kuleshov, and J. Tompkin. The GAN is dead; long live the GAN! a modern baseline GAN. In *ICML 2024 Workshop on Structured Probabilistic Inference & Generative Modeling*, 2024.
5. J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.
6. J. Kaleta, D. Dall'Alba, S. Płotka, and P. Korzeniowski. Minimal data requirement for realistic endoscopic image generation with stable diffusion. *International journal of computer assisted radiology and surgery*, 19(3):531–539, 2024.
7. E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. LoRA: Low-rank adaptation of large language models. In International Conference on Learning Representations, 2022.
8. B. Kim, I. Han, and J. C. Ye. Diffusemorph: Unsupervised deformable image registration using diffusion model. In *European Conference on Computer Vision*, pages 347–364. Springer, 2022.

This work was originally presented at the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).