Technische Universität München – Faculty of Informatics
Chair for Computer Aided Medical Procedures (Prof. Nassir Navab)
Practical Course: Machine Learning in Medical Imaging SS24

Causal Video Generation in Medical Imaging

1 General Info

Project Title: Causal Video Generation in Medical Imaging
Supervisors: Azade Farshad, Yousef Yeganeh
Contact Email: azade.farshad@tum.de, y.yeganeh@tum.de

2 Background and Motivation

Video generation involves synthesizing realistic and diverse video frames from a given input. In medical imaging, video generation can have various applications, such as data augmentation, anomaly detection, and surgical flow simulation. However, most existing methods for video generation rely on pixel-level information and ignore the high-level semantic structure of the scene as well as the temporal causal relationships [2]. Our project seeks to address this by leveraging higher-level inputs, such as action triplets, to guide the video generation. Specifically, scene graphs offer a visual representation of a scene's objects, their attributes, and the relationships between them, encapsulating both semantic and spatial details of images. Dynamic scene graphs corresponding to a video provide information about causal relations and the order of events and actions that should occur. In this project, we aim to produce video frames based on a dynamic scene graph that follows the causal relations provided in the graph. This graph will detail the evolving changes in the scene as time progresses.

3 Project Abstract

The project consists of the following steps: 1) Causal modeling of the actions performed in the surgery [4, 5], 2) Video Generation: We will build on top of a conditional stable video diffusion model [1] to generate a video that follows the provided actions and the cause / effect accurately. The methods would be trained and evaluated on the CholecT50 [3] dataset that contains videos of surgeries and the actions performed during the surgery.

4 Technical Prerequisites

- Good background in machine learning and deep learning
- Experienced in PyTorch
- Experienced in Python
- Experience with Generative Models

5 Benefits

- Weekly supervision and discussions
- Possible novelty of the research
- The results of this work are intended to be published in a conference or journal

6 Work packages and Time-plan

	Description	# Students
WP1	Familiarizing with the literature.	4
WP2	Implementing the baselines	4
WP3	Improving the baselines and validation on relevant datasets	4
	Midterm Presentation	4
WP4	Implementing the model	4
WP5	Finalizing the results and evaluation	4
	Final Presentation	4

Table 1: Project Timeline

References

[1] Andreas Blattmann, Robin Rombach, Huan Ling, Tim Dockhorn, Seung Wook Kim, Sanja Fidler, and Karsten Kreis. Align your latents: High-resolution video synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22563–22575, 2023.

[2] Aneesh Komanduri, Xintao Wu, Yongkai Wu, and Feng Chen. From identifiable causal representations to controllable counterfactual generation: A survey on causal generative modeling. *arXiv preprint arXiv:2310.11011*, 2023.

[3] Chinedu Innocent Nwoye, Tong Yu, Cristians Gonzalez, Barbara Seeliger, Pietro Mascagni, Didier Mutter, Jacques Marescaux, and Nicolas Padoy. Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos. *Medical Image Analysis*, 78:102433, 2022.

[4] Fabio De Sousa Ribeiro, Tian Xia, Miguel Monteiro, Nick Pawlowski, and Ben Glocker. High fidelity image counterfactuals with probabilistic causal models. In *International Conference on Machine Learning*, pages 7390–7425. PMLR, 2023.

[5] Nitish Srivastava and Russ R Salakhutdinov. Multimodal learning with deep boltzmann machines. *Advances in neural information processing systems*, 25, 2012.