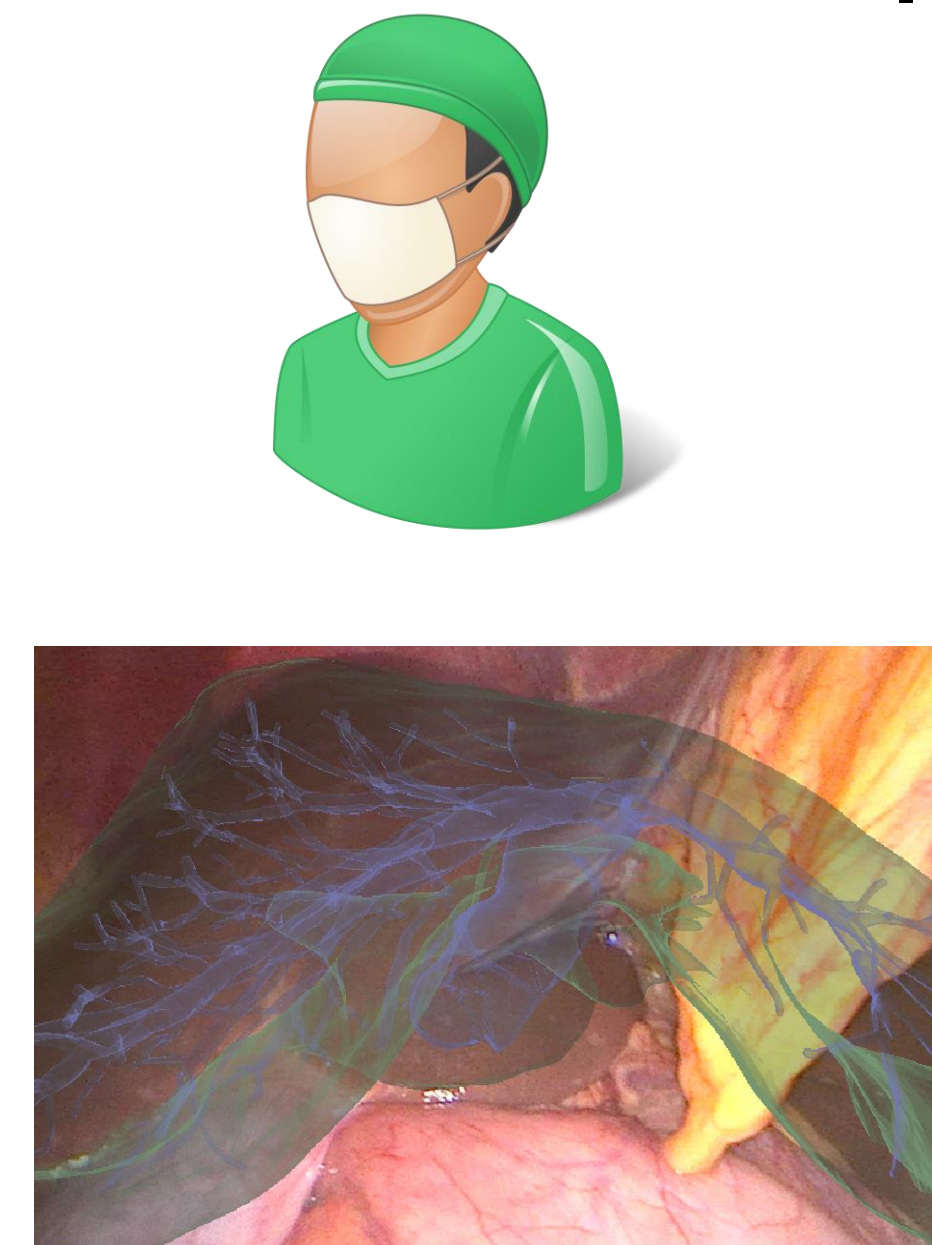# Data Augmentation for Surgical Scene Segmentation using Anatomy-Aware Diffusion Models
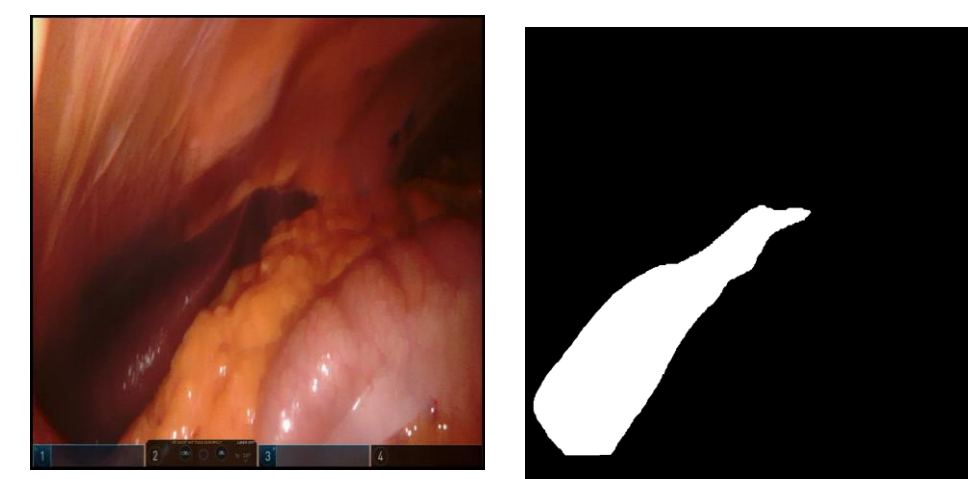
Danush Kumar Venkatesh[1,2], Dominik Rivoir[2], Micha Pfeiffer[2], Fiona Kolbinger[1,3], Stefanie Speidel[1,2]

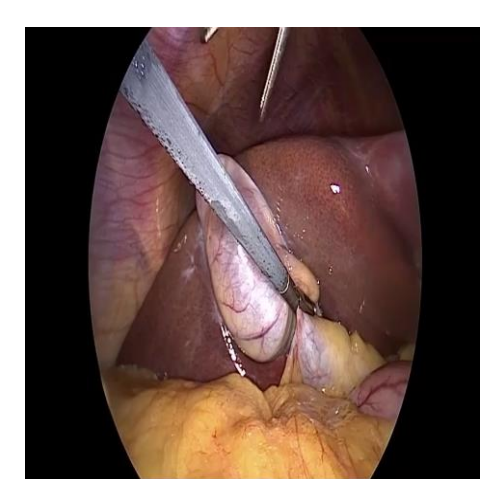WACV 2025 — TUCSON, ARIZONA • FEB 28 - MAR 4

Introduction

- Laparoscopic procedures
- Computer assistance during surgery
- **Requirement:** Understanding the surgical scene
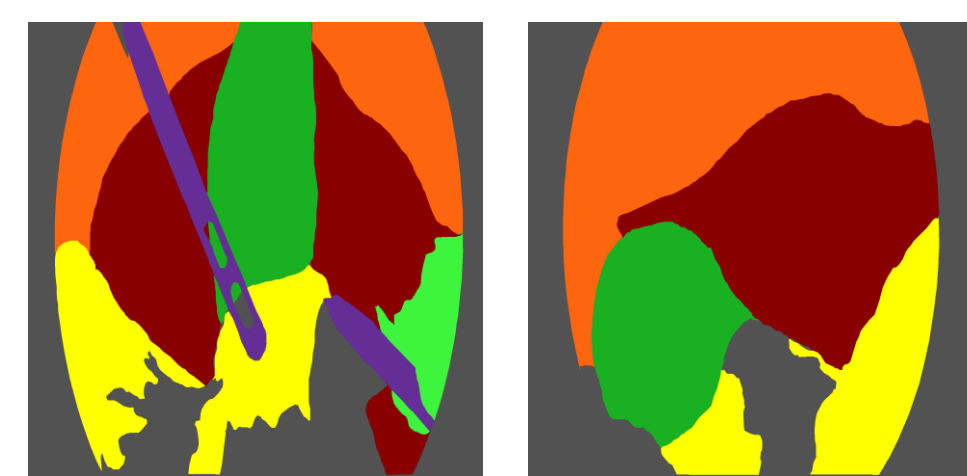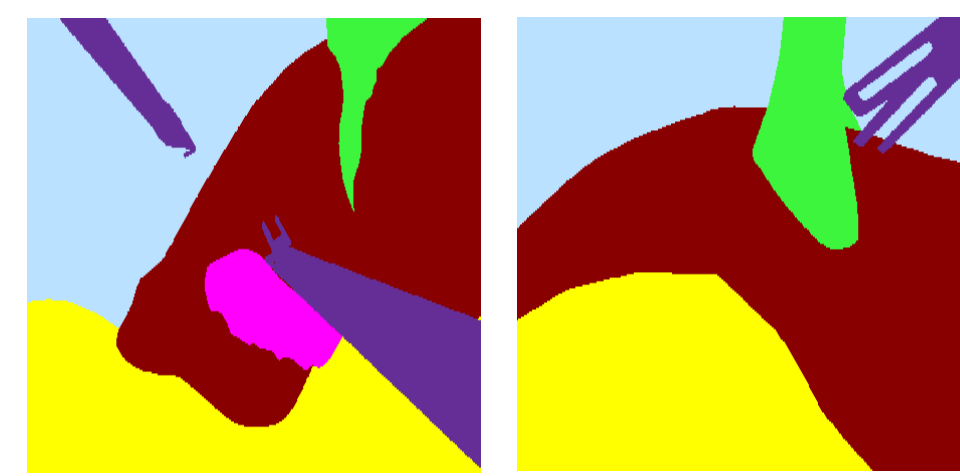
Problem Statement

Binary annotation

Multi-class annotation

➤ Full scene understanding requires multi-class annotation
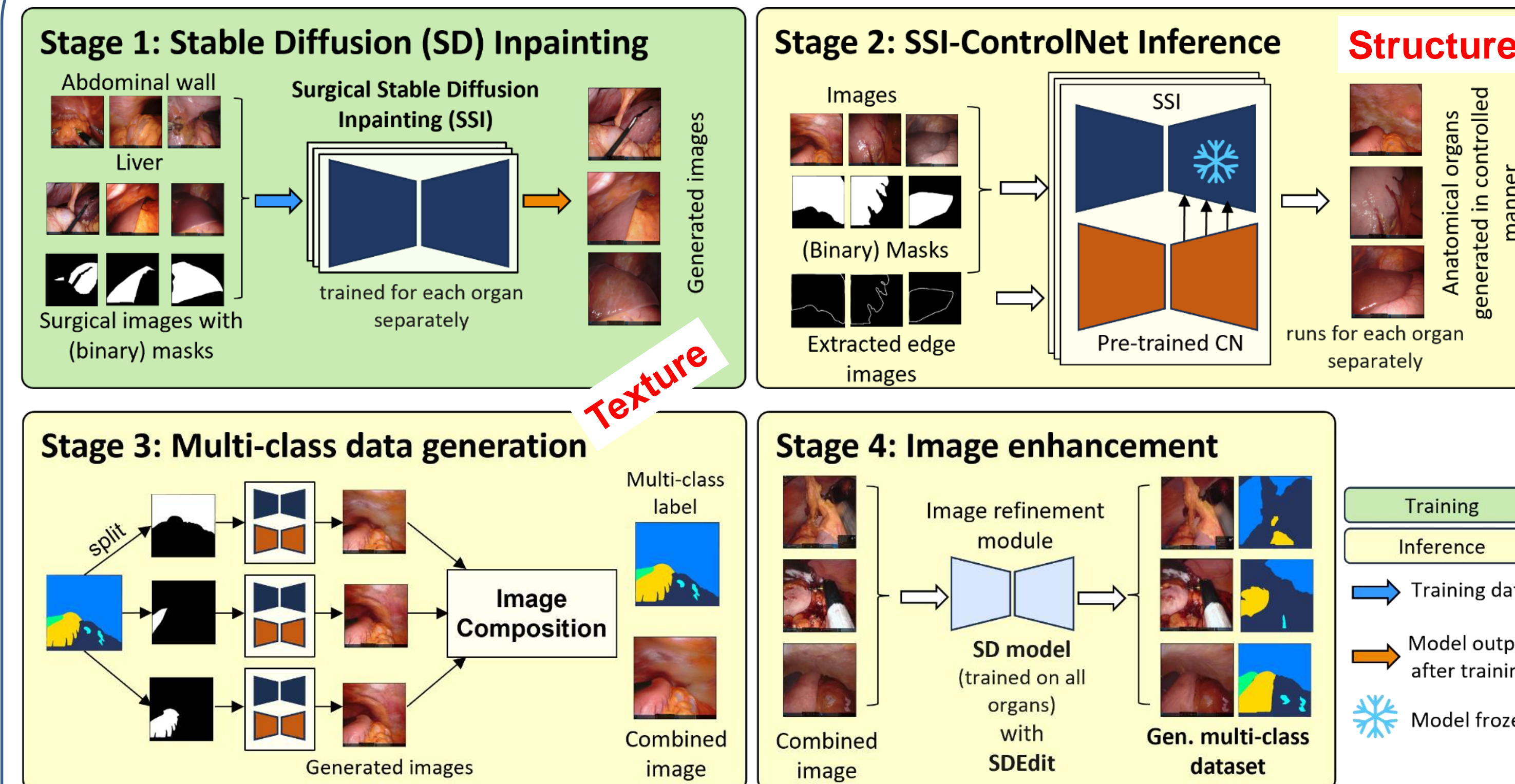
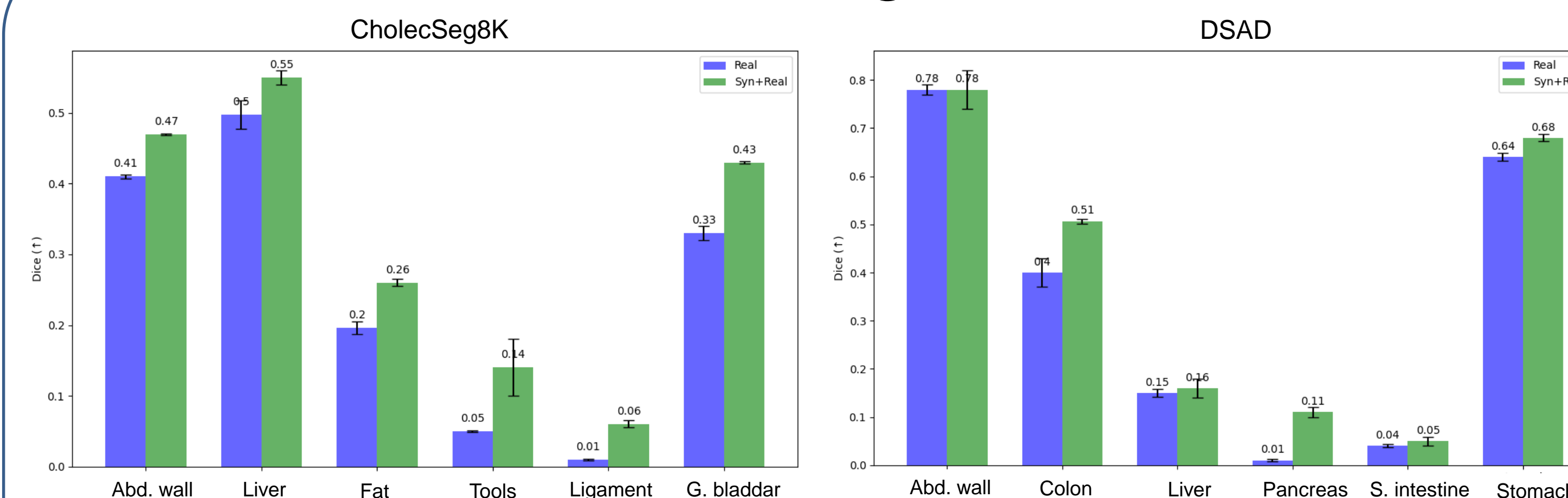**Available segmentation masks :**

Real surgical masks

Surgical simulation masks

**Can we generate multi-class surgical images ?**

Method

Stage 1: Stable Diffusion (SD) Inpainting

Abdominal wall — Liver — Surgical images with (binary) masks → Surgical Stable Diffusion Inpainting (SSI) — trained for each organ separately → Generated images

Texture

Stage 2: SSI-ControlNet Inference — Structure

Images — (Binary) Masks — Extracted edge images → SSI / Pre-trained CN → Anatomical organs generated in controlled manner — runs for each organ separately

Stage 3: Multi-class data generation

split → Image Composition → Generated images — Multi-class label — Combined image

Stage 4: Image enhancement

Combined image → Image refinement module — SD model (trained on all organs) with SDEdit → Gen. multi-class dataset

Training / Inference / Training data / Model output after training / Model frozen

Qualitative Results

Real image / Segmentation Mask. / Generated images

Semantic mask — SPADE — SPADE-vae — Pix2PixHD — ControlNet — ControlNet-HED — T2i-Adapter — T2i-Adapter-CY — Ours

Semantic Segmentation

CholecSeg8K — Abd. wall (0.41 / 0.47), Liver (0.49 / 0.55), Fat (0.2 / 0.26), Tools (0.05 / 0.14), Ligament (0.01 / 0.06), G. bladdar (0.33 / 0.43)

DSAD — Abd. wall (0.78 / 0.78), Colon (0.41 / 0.51), Liver (0.15 / 0.16), Pancreas (0.01 / 0.11), S. intestine (0.04 / 0.05), Stomach (0.64 / 0.68)

Real / Syn+Real

Training scheme	Unet++ [69]			DV3+ [8]			UperNet-Tiny [64]		
	Dice(↑)	IOU (↑)	HD (↓)	Dice (↑)	IOU (↑)	HD (↓)	Dice (↑)	IOU (↑)	HD (↓)
Real with no-aug	0.50±0.03	0.36±0.01	101.03±0.12	0.50±0.01	0.36±0.01	115.36±3.82	0.56±0.01	0.47±0.02	118.37±6.62
Real with color-aug	0.52±0.01	0.38±0.02	**98.95±2.05**	0.53±0.01	0.39±0.01	101.54±0.19	0.59±0.01	0.45±0.01	110.93±1.42
Real with color+spatial-aug	0.61±0.05	0.49±0.04	109.09±0.52	0.58±0.01	0.45±0.01	108.14±1.07	0.61±0.04	0.50±0.05	108.63±1.51
Ours only *Syn*	0.53±0.03	0.40±0.01	110.65±1.31	0.53±0.01	0.41±0.02	108.66±1.18	0.56±0.01	0.44±0.01	109.41±2.09
Ours-*SS-Syn* + Real	**0.67±0.01**	**0.54±0.01**	107.10±0.49	0.64±0.01	0.51±0.05	**95.86±8.25**	0.65±0.03	0.53±0.02	**95.76±2.49**
Ours-*Syn* + Real	0.64±0.03	0.51±0.01	101.96±1.43	**0.68±0.01**	**0.56±0.01**	95.93±6.89	**0.67±0.01**	**0.54±0.01**	99.97±2.24
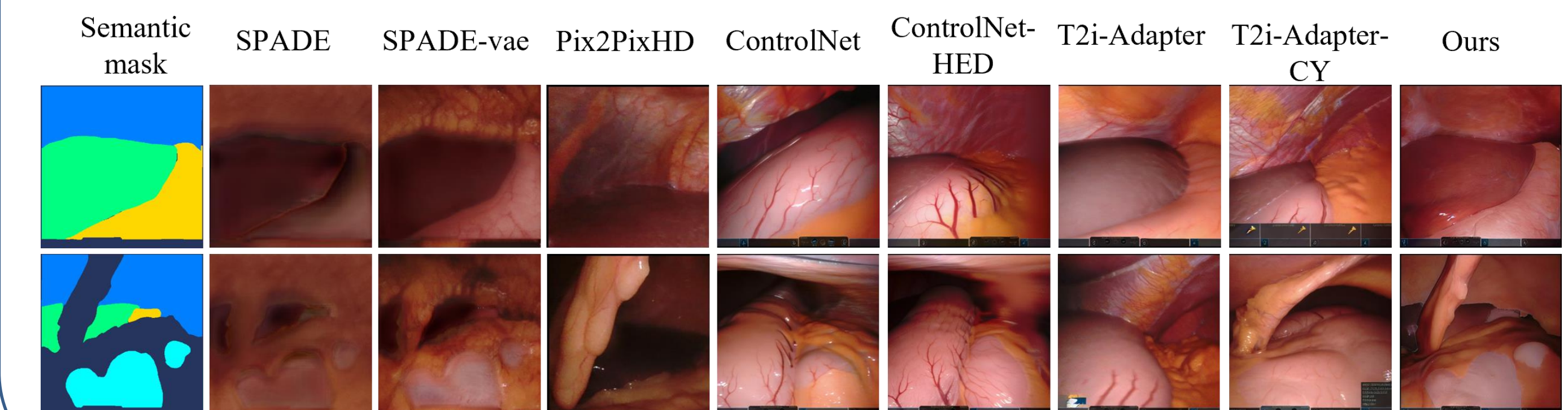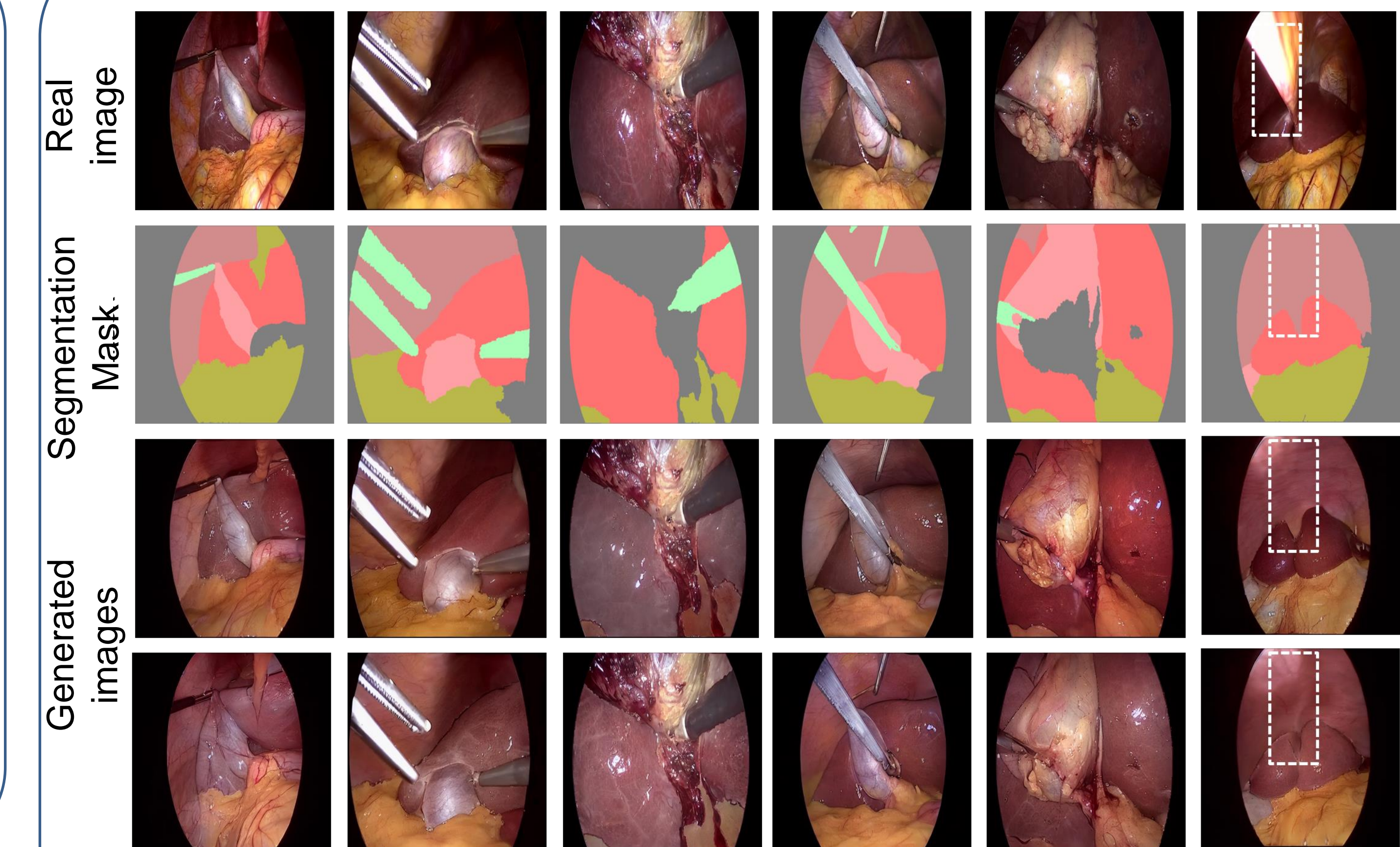
Conclusion

- Multi-stage diffusion approach
- Specific control of each anatomy
- Extension to other anatomies

Code & Paper

References:
Micha et al. Data-driven Navigation for Laparoscopic Liver Surgery, 2021
Rombach et al. High resolution image sysnthesis with latent diffusion models,2021
Zhang et al.Adding conditional control to t2l diffusion models, 2021
Meng et alSDEdit:Guided image synthesis & editing with SDE, 2021