Medical Image Generation using Segmentation Guided Diffusion

Florian Hunecke, Chin Ju Chen

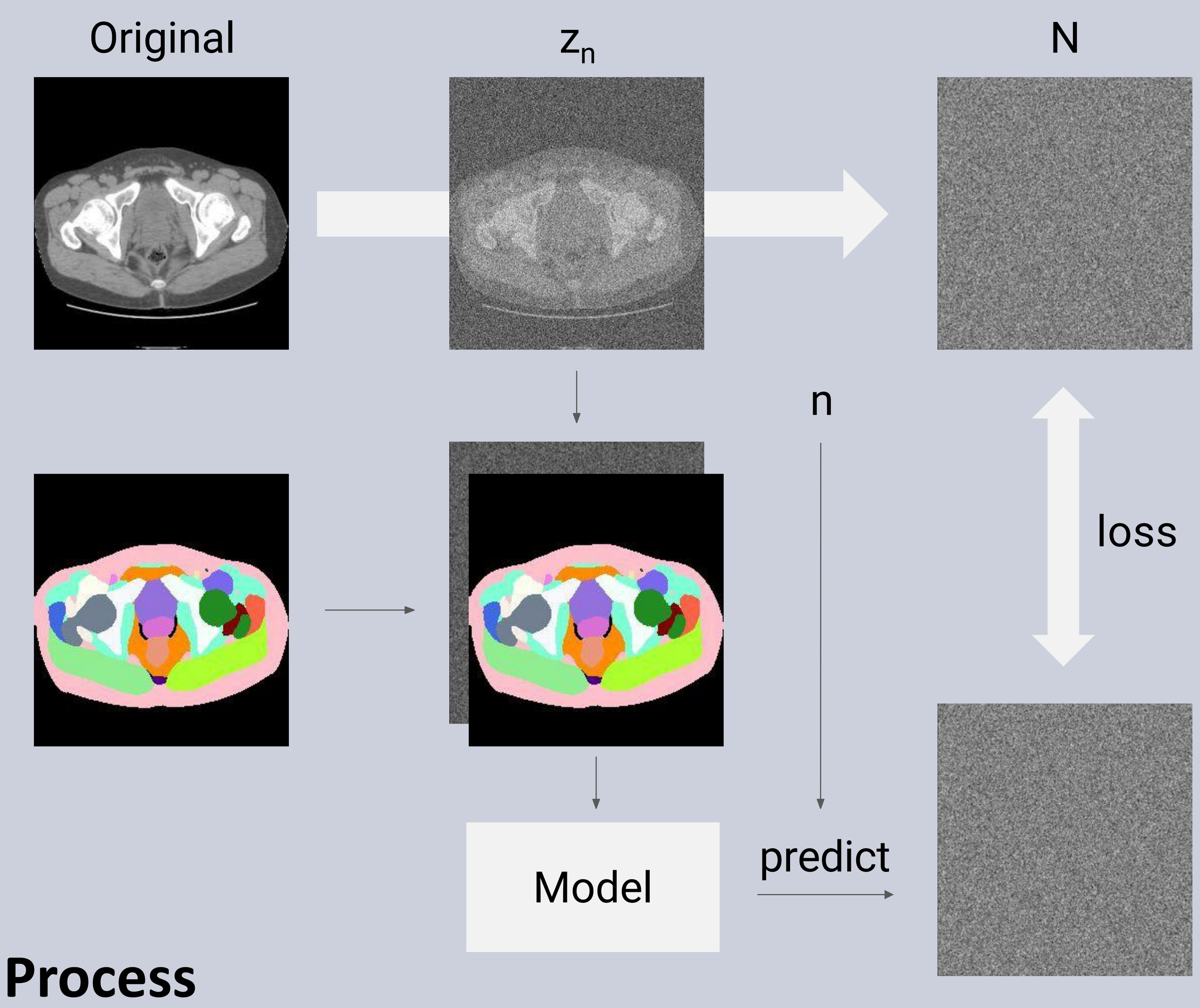
Background and Motivation

- Deep learning has proven highly effective for medical image segmentation, transforming CT or MRI scans into precise tissue and organ delineations
- In this study, we employ a state-of-the-art generative diffusion model to invert this process, synthesizing accurate and high-fidelity medical images from segmentation masks
- The reverse diffusion process is conditioned on the masks to preserve the distribution of anatomical landmarks and regions

Methods

Segmentation Guided Diffusion (SGD) [1]

- **Forward Pass**: Iteratively adding noise to the input image
- **Backward Pass**: In addition to the noised image, the denoising model receives information about the desired segmentation
- Common methods of enriching the noise: concatenation, addition



Training Process

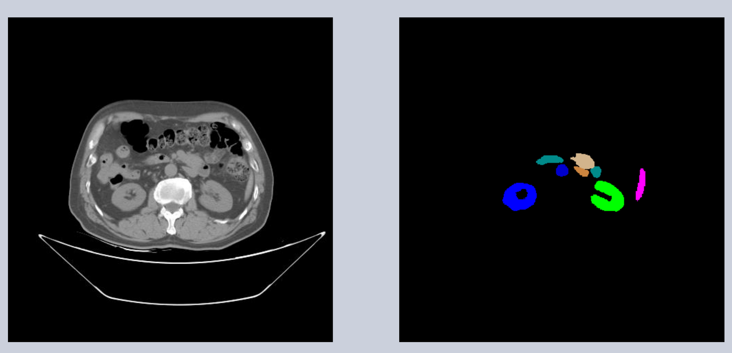
- The models were separately trained on a total of 260k CT and 40k MRI 2D images slices
- For performance comparison, training was also done on the different subsets of slice directions (axial, coronal, sagittal)

Hyperparameter	Image Shape	Batch Size	Learning Rate	Model Type
SGD	256	16	1e-4	UNet
2D Segmentor	256	64	1e-2	UNet

Datasets

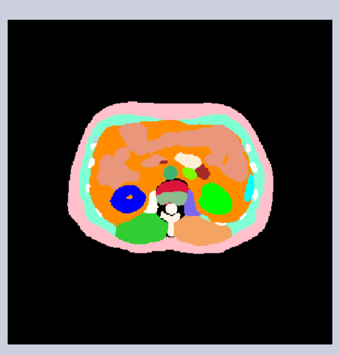
AMOS [2]

- Longgang District Hospital, China
- 500 CT and 100 MRI scans covering 15 organ categories
- Assessment: 2018 – 2021



TotalVibeSegmentor [3]

- Segmentation model trained on NAKO and UK Biobank
- 71 structures for full torso segmentation

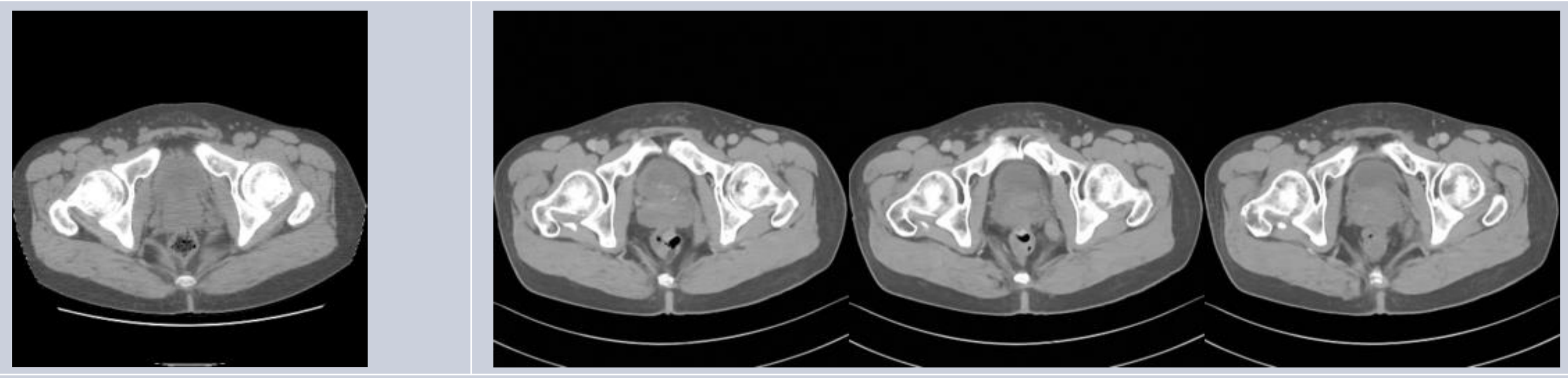


Generations

Original | **Segmentation** | **Generation**

- Reconstruction (CT to CT)

- Multi-Inference (CT to CT)



- Multi-modal (CT to MRI)

Evaluation

- We evaluate the generations using typical metrics to compare the similarity of two image(sets) like PSNR, SSIM and FID
- Since our generations are not supposed to match the original scan, an auxiliary model was trained to segment the generated images
- Dice Score is then used to assess the label overlap, with $Dice_{orig}$ representing the score achieved by the auxiliary model when segmenting the original images, serving as an upper bound

	CT axial	CT	MRI axial	MRI
PSNR ↑	19.06	20.26	13.07	12.44
SSIM ↑	0.637	0.618	0.210	0.198
FID ↓	47.66	61.50	75.04	100.4
$Dice_{orig}$ ↑	0.925	0.927	0.793	0.848
$Dice_{gen}$ ↑	0.923	0.927	0.778	0.807

Conclusion

- The results clearly show that a segmentation-conditioned generation of high-quality medical images is possible
- Although this method cannot reconstruct details of the original image, it may easily be used to create diversity in future datasets
- The evaluation reveals differences between the imaging techniques, potentially originating from their different value distributions
- Further experiments could be conducted on adapting image preprocessing and diffusion noise scheduling

References

[1] Konz, Nicholas, et al. "Anatomically-controllable medical image generation with segmentation-guided diffusion models." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer Nature Switzerland, 2024.

[2] Ji, Yuanfeng, et al. "Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation." *Advances in neural information processing systems* 35 (2022): 36722-36732.

[3] Graf, Robert, et al. "TotalVibeSegmentator: Full Torso Segmentation for the NAKO and UK Biobank in Volumetric Interpolated Breath-hold Examination Body Images." *arXiv preprint arXiv:2406.00125* (2024).

Klinikum rechts der Isar
Technische Universität München