EÖTVÖS LORÁND
UNIVERSITY | BUDAPEST

FLORA:Flow-based Latent-informed Optimization for 3D proton-CT Reconstruction with Spatial Attention

HUN-REN WIGNER GPU DAY 2026. 06. 28-29.

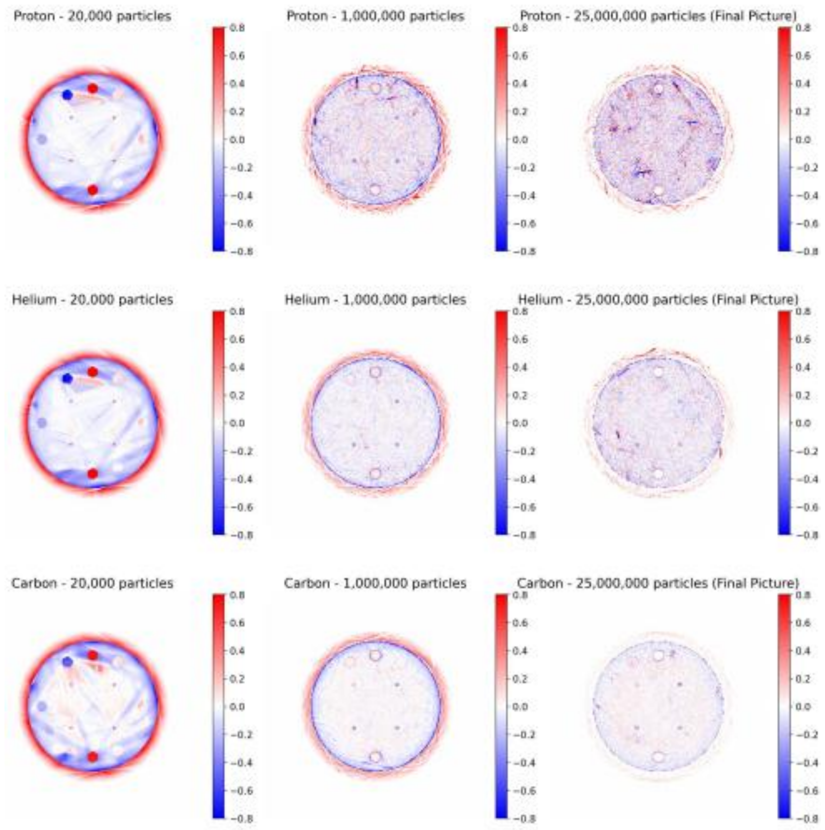
Bence Dudás, PhD Student
Eötvös Loránd University, Theoretical Physics Department

Motivation

Proton Computed Tomography

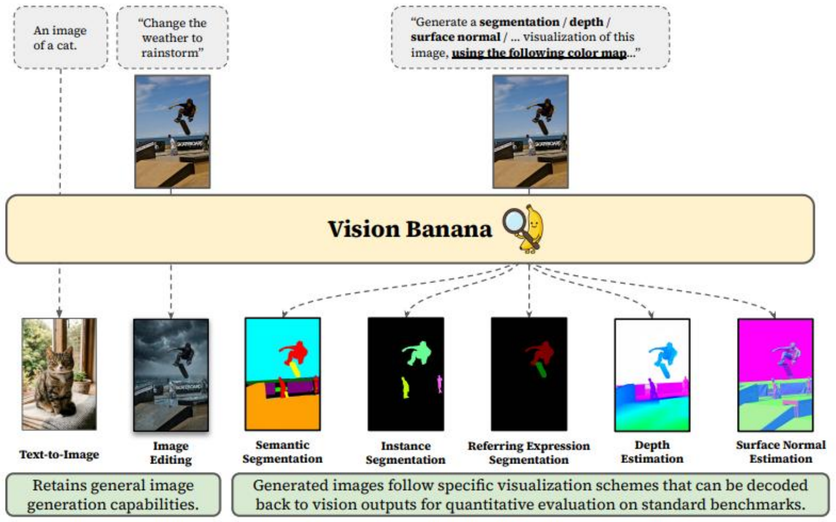
Traditional reconstruction

- Iterative reconstruction algorithm.
  - Needs a large amount of particle to converge.
- Currently focusing on 2D reconstruction.
  - Can we directly jump into 3D?
- Adding external medical knowledge is hard.
  - What is the gender or age of the patient?
  - What are the medical conditions?
  - Can these reduce the computational need?
- Advancements in GENAI shows potential.

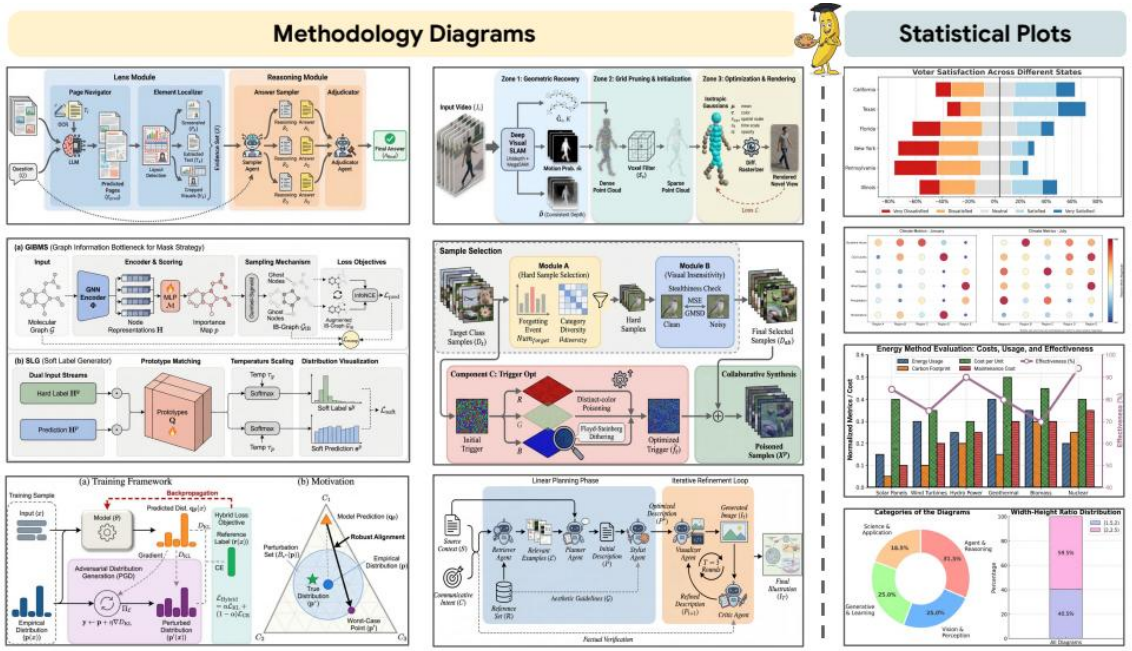


GENERATIVE AI

- Generative AI is an extremely promising field of machine learning.
- Many of these models can overperform specified predictive models.
  - Image segmentation.
  - Depth Estimation
  - Project workflow visualization
  - Coding
  - Statistical analysis.
- Agentic AI Scientist show potential.
- Can it be applied to medical AI?



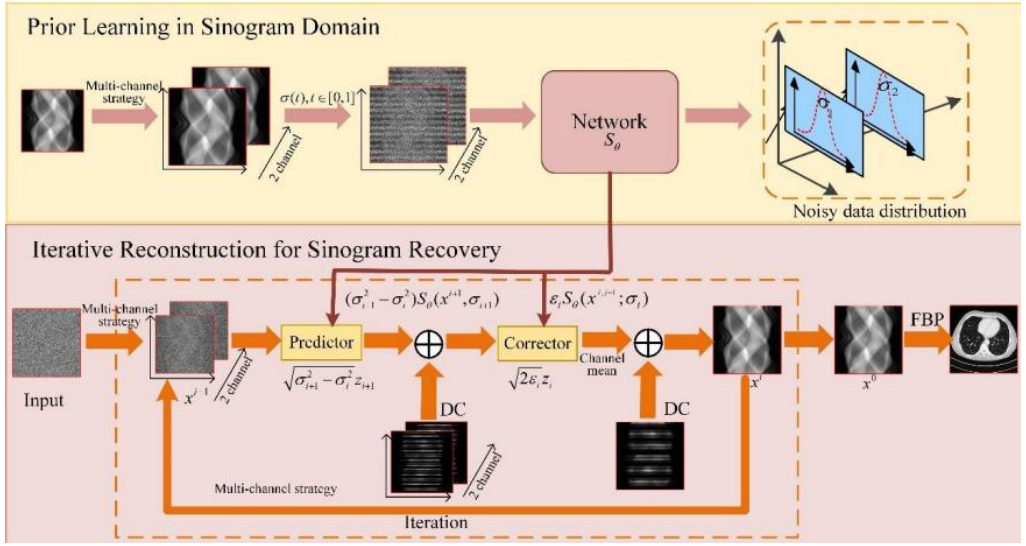
Valentin G. et al (2026), Image Generators are Generalist Visin Learners; DeepMind



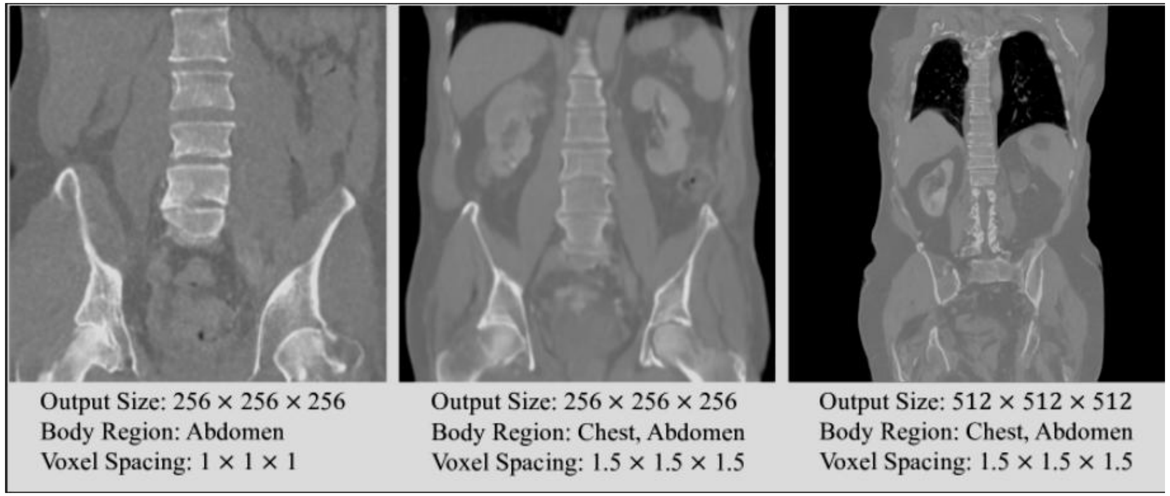
Dawei Z. et al (2026), PaperBanan: Automating Academic Illustration for AI Scientist; Google

GENAI in Medical Images

- Research shows that traditional CT can be reconstructed from sparse-view by generative AI.
- MAISI already showed it can generate biologically consistant images.
  - Segmentators trained on these perform better.
  - This is unconditional image generation.
- How to achieve Relative Stopping Power (RSP) reconstruction from detector data with GENAI?
- Can we add external medical knowledge?
  - Diagnosis, age, etc.
  - Speed up reconstruction.
  - Increase accuracy.



Bing G. et al. (2022), Generative Modeling in Sinogram Domain for Sparse-view CT Reconstruction

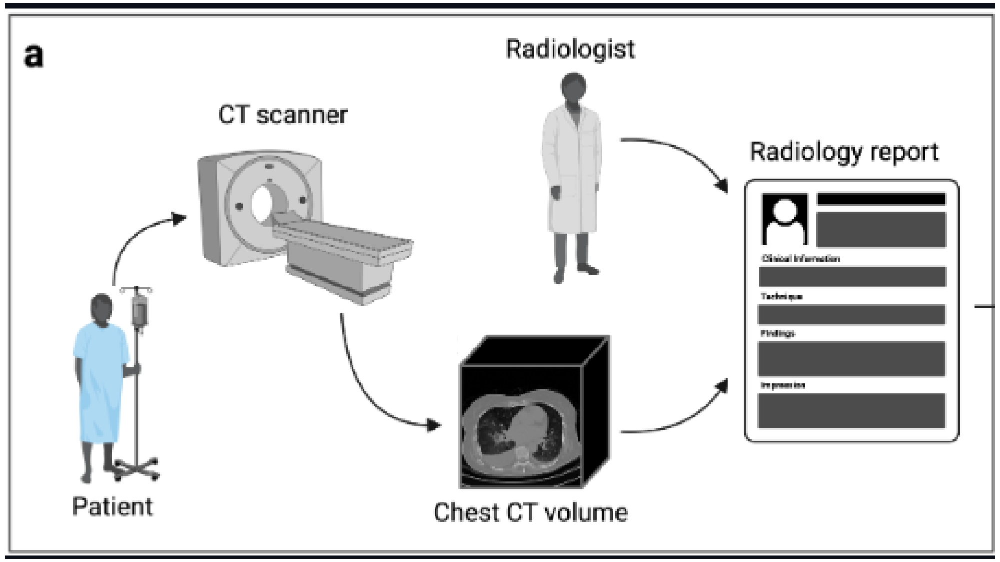


Pengfei, G. et al. (2024), MAISI: Medical AI for Synthetic Imaging; NVIDIA MONAI

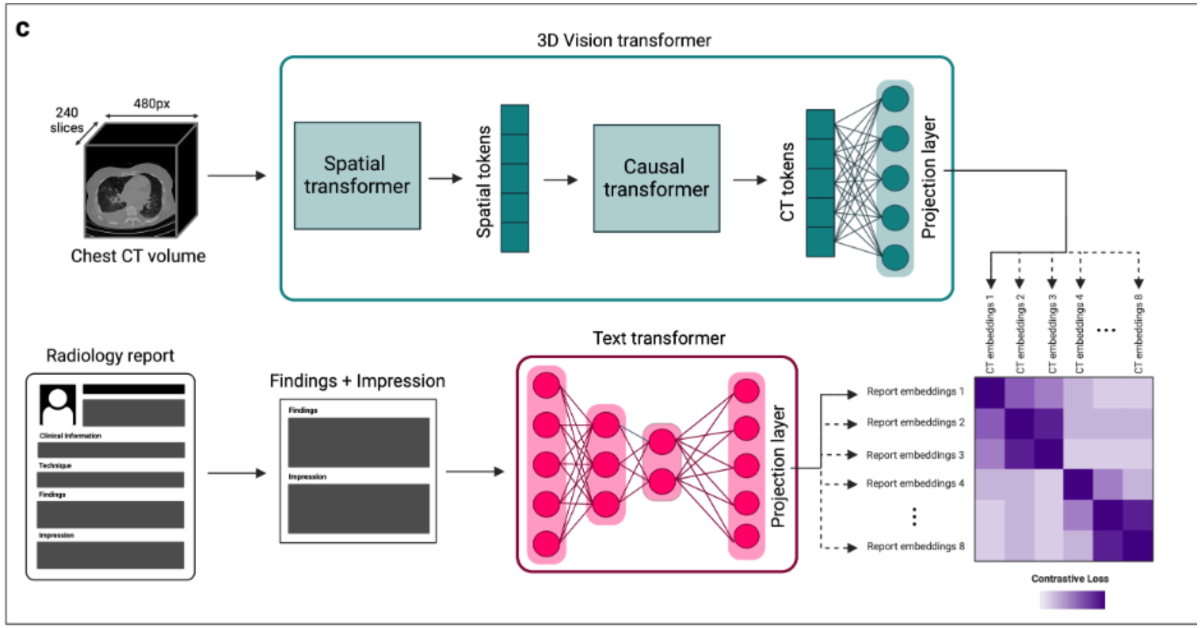
Data acquisition

CT-RATE

- 21TB of multimodal chest CT dataset.
  - – CT-RATE from HuggingFace.
  - – 21,304 unique patients.
- For our reconstruction subset is sufficient.
  - – Using only 2000 unique patients.
  - – This is sufficient for conditioned image generation.
- Expert metadata is available.
  - – Clinical diagnosis.
  - – One patient can have multiple conditions.
  - – pCT needed for treatment planning not for diagnosis.
- Tumor masks included.
  - – Only used for evaluation/test.



Ibrahim E. H. et al (2024), Generalist Foundation Models from a Multimodal Dataset for 3D Computed Tomography



Simulations

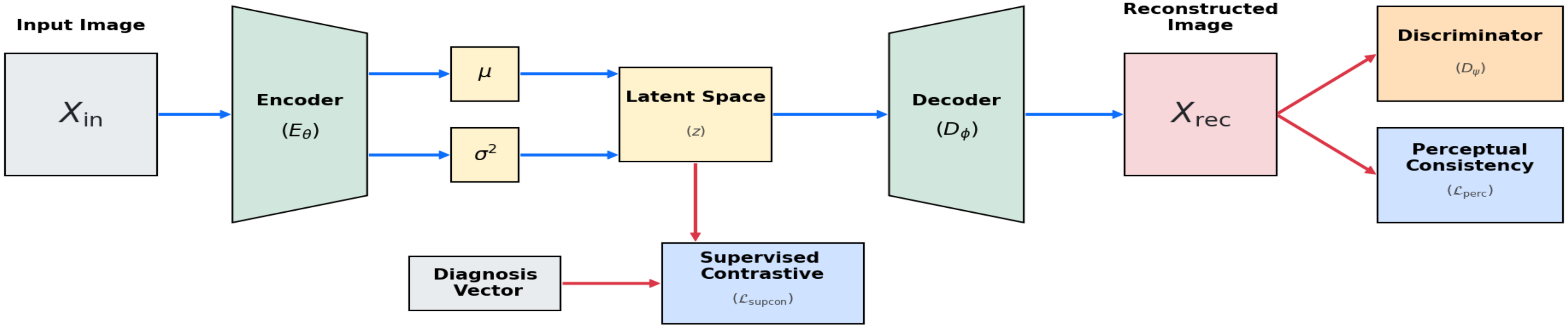
- Phantom iradiations are simulated with OpenGate.
- CT converted into ground truth RSP.
  - CT gives Hounsfeld Unit (HU).
  - For every building material(bones,tissues) calculated corrsesponding RSP.
  - HU converted to material, material to measured RSP.
- Particle scattering simulations run on GT RSP volumes.
- Lung simulations show huge uncertainity.
  - Expected: Lung is filled with air bubbles.

Material	HU	RSP	Std
Air	-1050	0.000	0.000
Lung	-950	0.499	0.330
Adipose	-120	0.955	0.003
Soft Tissue	19	1.038	0.003
Connective Tissue	80	1.058	0.003
Marrow Bone (02)	200	1.127	0.003
Marrow Bone (03)	300	1.153	0.003
Marrow Bone (05)	500	1.217	0.003
Marrow Bone (07)	700	1.280	0.004
Marrow Bone (09)	900	1.344	0.003
Marrow Bone (11)	1100	1.409	0.004
Marrow Bone (14)	1400	1.511	0.005
Marrow Bone (15)	1500	1.542	0.004
Amalgam Tooth	1640	1.602	0.013
Metal Implants	2300	2.125	0.015
Metal Implants	3000	3.162	0.022

Image reconstruction

MEDICALLY CONDITIONED AUTOENCODER

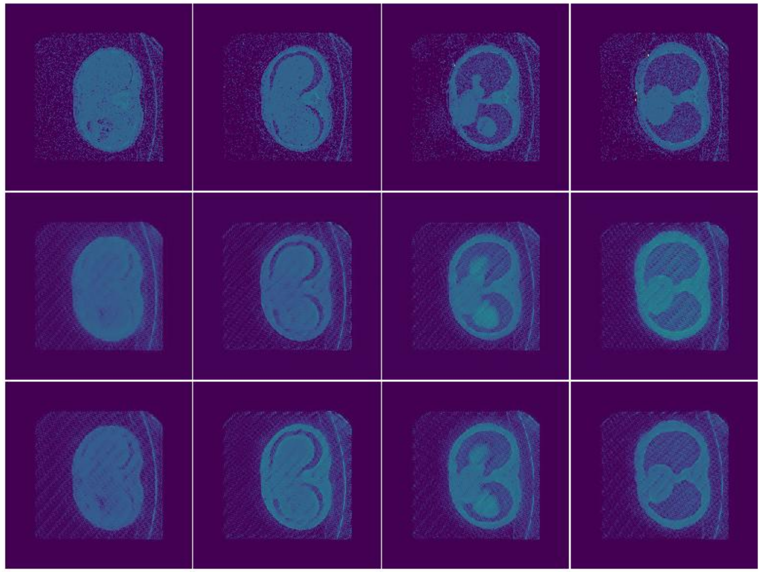
- Variational Autoencoder to generate latent space.
  - 3D Volumes of size (256,256,64) yield high computational cost.
- Only MSE for reconstruction not enough
  - Often yields biologically wrong images.
- Supervised Contrastive Loss on latent space
  - Enable diagnosis based sampling.
- Adversarial discriminator on reconstruction.
  - Force anatomic cosistency.
- Perceptual loss to measure biological features.
  - Similarity between feature maps from a pretrained model.



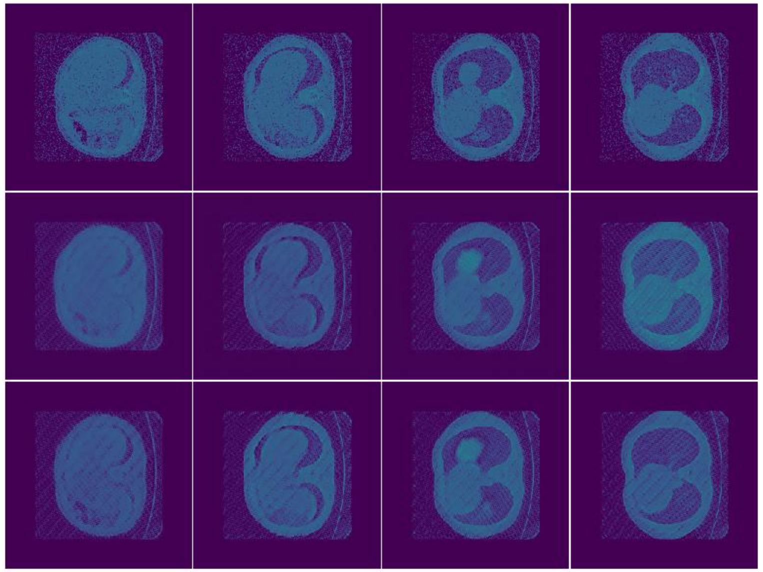
Unconditioned Reconstruction

- The Variational autoencoder generates biologically consistent images.
- Experimenting to use Exponential Moving Average to generate even better outputs.
- Visible structures with sharp borders.
- Average volume Mean Squared Error is low (~0.005).
- Air bubble in the lung often smoothed into the tissue.
  - This is not necessarily a problem for radiology.
  - Due to the amount of air bubble RSP of the lung is extremely noisy.

Sample 3 | Cond: 6-9 | Top: Real | Mid: Vanilla | Bot: EMA

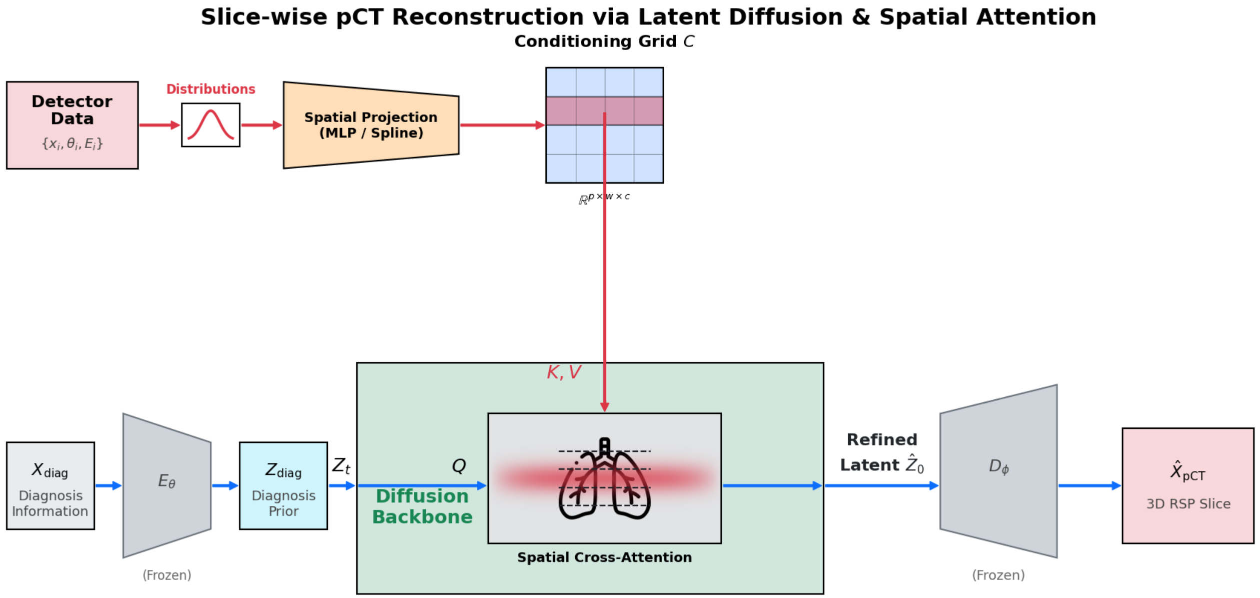


Sample 5 | Cond: 2-4 | Top: Real | Mid: Vanilla | Bot: EMA



LATENT FLOW WITH DETECTOR SIGNAL CONDITIONING

- Detector signals converted into distributions.
  - Particle scattering information, used for traditional reconstruction (scattering angles and residual energies)
  - For every slice and every angle.
- Frozen Encoder-Decoder to focus transition.
  - Alternative: unfrozen decoder with small LR.
- Generate latent volume from diagnosis.
  - Clinical condition is known.
  - Generate latent from corresponding distribution.
- Flow Matching to generate target from prior.
  - Less stochastic then DDPM.
- **Spatial attention to take physics into account.**
  - **Proton scattering can reach adjecent slices.**
  - **Traditional algorithm based on MLP.**



Summary

Summary

FLORA

- Converted a large amount of 3D volumes for training a deep learning model.
  - Calculated the RSP for every RSP found in chest CT volumes.
- Introduced FLORA a Generative AI based RSP reconstruction pipeline.
- Trained a VAE-GAN with 0.005 MSE as the first stage of our workflow.
  - Latent space regularized by the medical conditions.
  - Can start reconstruction from diagnosis.
- Spatial Cross-Attention for a Flow-Matching backbone to fit physics.
  - Still need extensive studies.

Future works

- Determine the the optimal number of angles.
  - We can compare to traditional method, but the approaches are too different.
- The optimal projection of the detector data.
  - Distribution conversion is usefull due to compute need.
  - Direct MLP projections could be applied.
- Using non-frozen decoder with smaller learning rate might yield better outputs.
  - Totally unfreezing probabbly destroy biological structure.
- Experimenting with the diffusion backbone.

Support

SUPPORTED BY THE DKOP-23 DOCTORAL EXCELLENCE PROGRAM OF THE MINISTRY FOR CULTURE AND INNOVATIONFROM THE SOURCE OF THE NATIONAL RESEARCH, DEVELOPMENT AND INNOVATION FUND.

We acknowledge the Digital Government Development and Project Management Ltd. for awarding us access to the Komondor HPC facility based in Hungary.

THANK YOU!