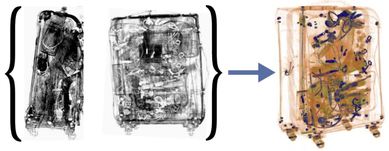
Unaligned 2D to 3D Translation with Conditional Vector-Quantized Code Diffusion using Transformers

Abril Corona-Figueroa, Sam Bond-Taylor, Neelanjan Bhowmik,
Yona Falinie A. Gaus, Toby P. Breckon, Hubert P. H. Shum, Chris G. Willcocks
Durham University, UK

ICCV23 PARIS

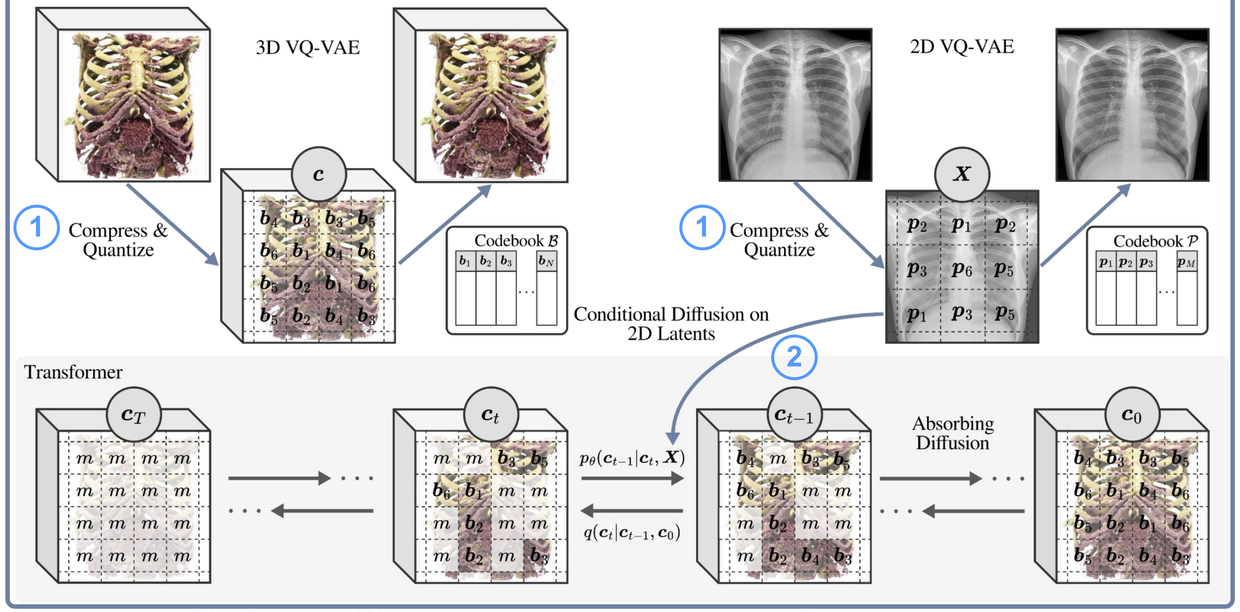
Motivation



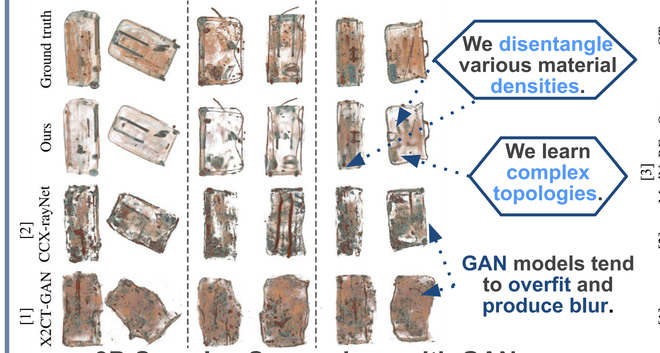
- Can we model **complex 3D** objects **from two 2D** images?
- **Unaligned inputs** for real-world applications.
- Improve **feature learning** and **speed** up 3D sampling.

Conditional Diffusion using Transformers

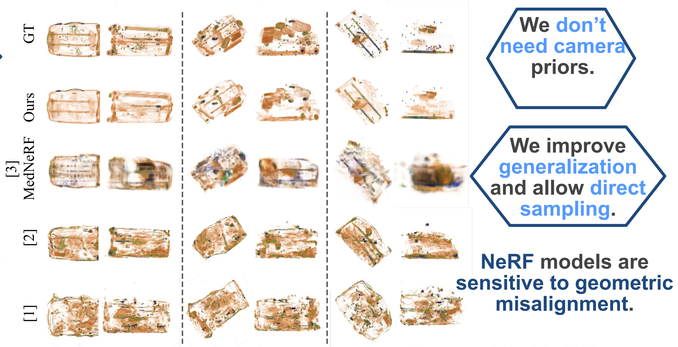
1. **Unpaired compression:** Learn rich Vector-Quantized discrete 2D and 3D spaces independently.
2. Model $p(c|\boldsymbol{X})$ with a **diffusion model** parameterized with an **unconstrained transformer**.
   - $\boldsymbol{c}$: VQ codes of 3D data.
   - $\boldsymbol{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$: set of VQ codes of all 2D views.



Modeling Objects in Baggage Screening



3D Samples Comparison with GAN



3D-aware Projections Comparison with NeRF

Modeling Chest Structures

Quantitative Evaluation

Baggage Security Screening dataset

Method	↓ NLL	↑ Density	↑ Coverage	↑ SSIM	↑ PSNR
[1] X2CT-GAN	N/A	0.95	0.80	0.655	34.68
[2] CCX-rayNet	N/A	1.28	0.89	0.886	35.45
Ours	**0.007**	**2.01**	**0.97**	**0.899**	**39.45**

LIDC-IDRI (chest) dataset

[1] X2CT-GAN	N/A	0.87	0.88	0.321	19.68
[2] CCX-rayNet	N/A	1.41	**0.98**	0.386	22.66
Ours	**0.10**	**1.42**	0.97	**0.436**	**25.05**

Contributions

- ★ Full-coverage attention allows unaligned inputs.
- ★ Domain-invariance (e.g. imaging modalities).
- ★ Fast high-quality 3D samples.
- ★ Data likelihood representation.
- ★ Generative control (e.g. feature level).
- ★ Mode coverage (i.e. 2D & 3D distributions).
- ★ Global context of conditional 2D inputs.

Not requisite of:

- ❑ Camera priors like in NeRF.
- ❑ Local alignment of inputs like in CNNs.
- ❑ Many input views (i.e. 2 views suffice).
- ❑ Continuous latent representations.
- ❑ Hierarchical architectures.
- ❑ Deep architectures with skip-connections.

Project Page

[1] Xingde Ying, *et al.* X2CT-GAN: Reconstructing CT from bi-planar X-rays with generative adversarial networks. CVPR, 2019.
[2] Md Aminur Rab Ratul, *et al.* CCX-rayNet: A Class Conditioned Convolutional Neural Network For Biplanar X-Rays to CT Volume. ISBI, 2021.
[3] Abril Corona-Figueroa, *et al.* MedNeRF: Medical Neural Radiance Fields for Reconstructing 3D-aware CT-Projections from a Single X-ray, EMBC, 2022.