Johannes Thyroff

Diffusion-based 3D Shape completion

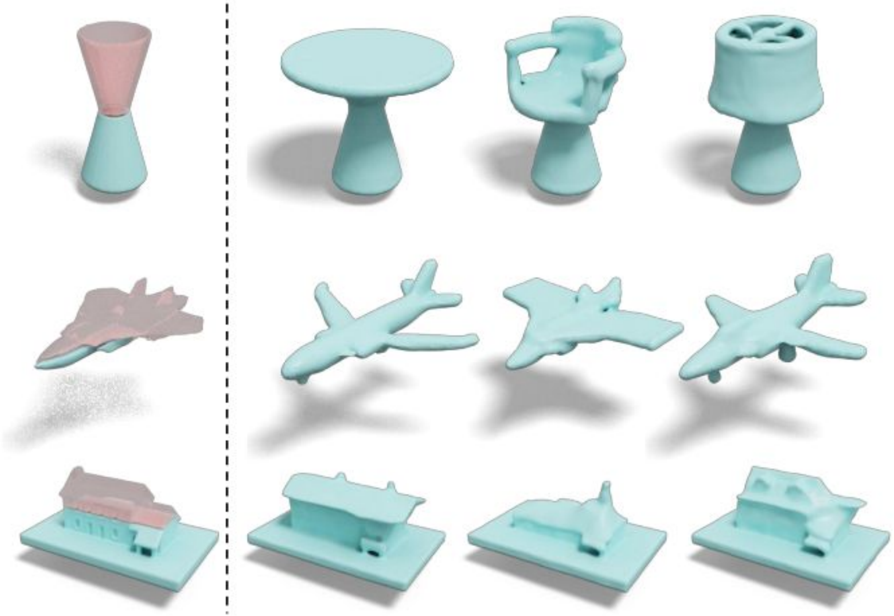
Technische Universität München

+

Johns Hopkins
Whiting School
of Engineering

Outline

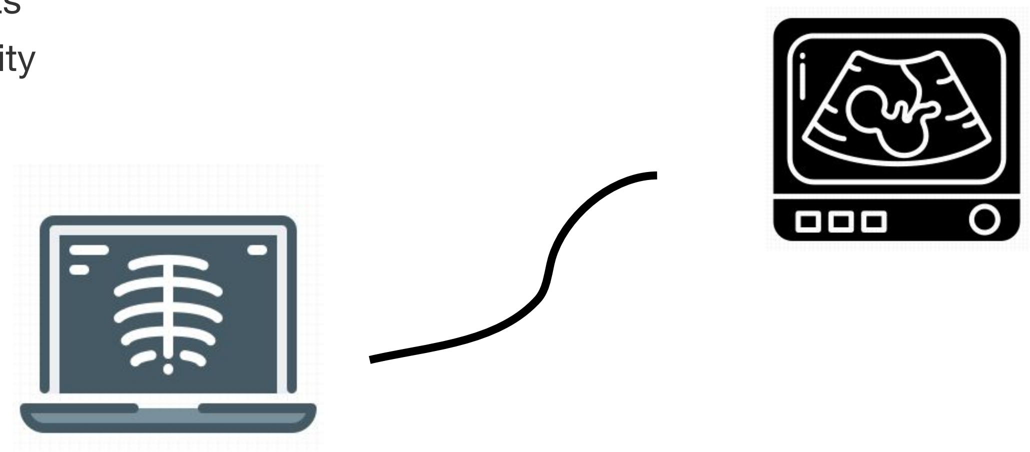
- Motivation and Terminology
- Methodology
- Experiment Results
- Discussion
- Review



[7]

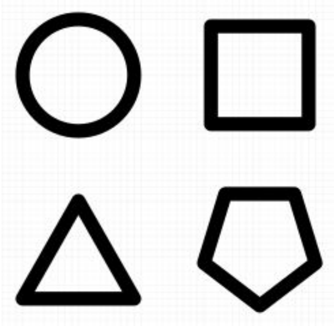
Motivation

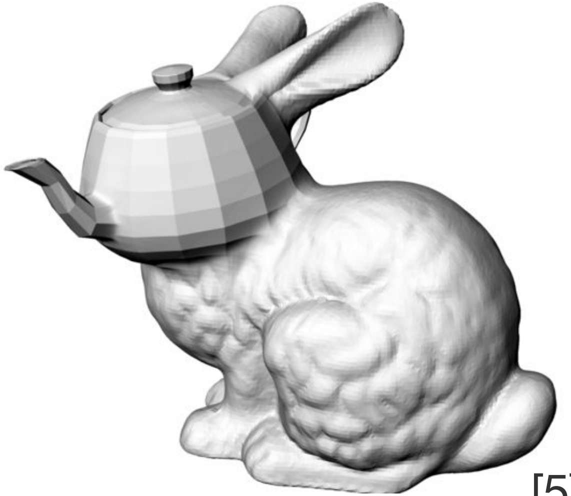
- 3D shape completion in a variety of fields
- Medical use:
  - data augmentation
  - synthetic datasets
  - change of modality



Goals [11]

1. Realistic
2. Diverse
3. Coherent reconstruction - GT resemblance

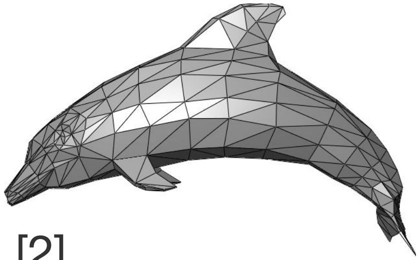




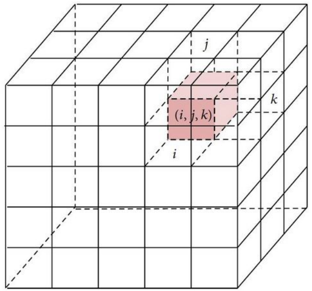
[5]

3D Shape Representations

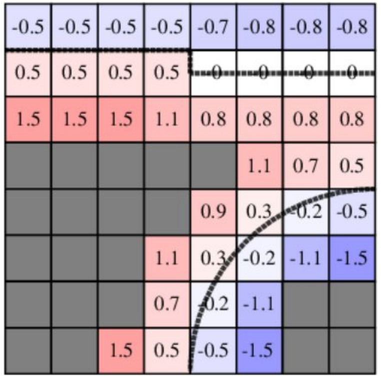
- Voxel Grids
- Polygon Meshes
- Signed Distance Function
- Neural fields



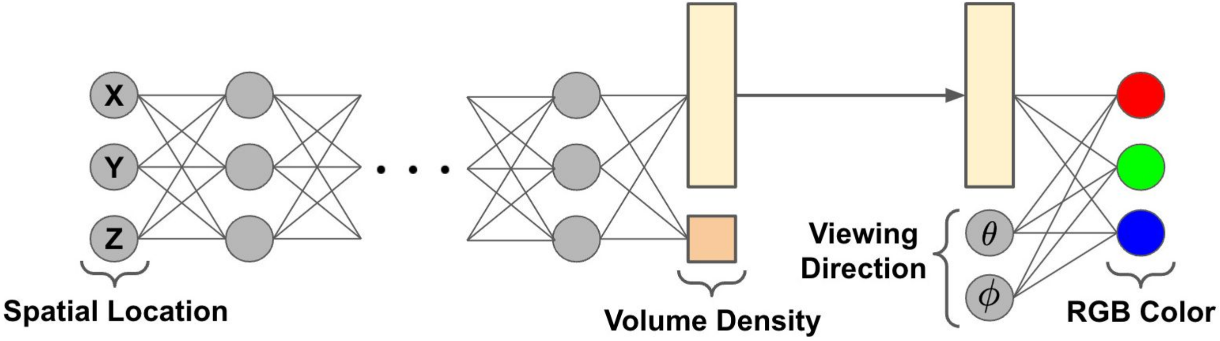
[2]



[1]



[3]

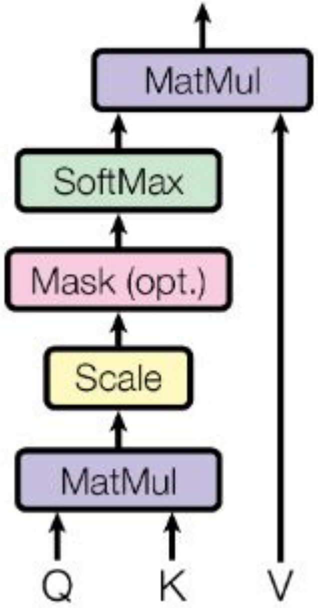


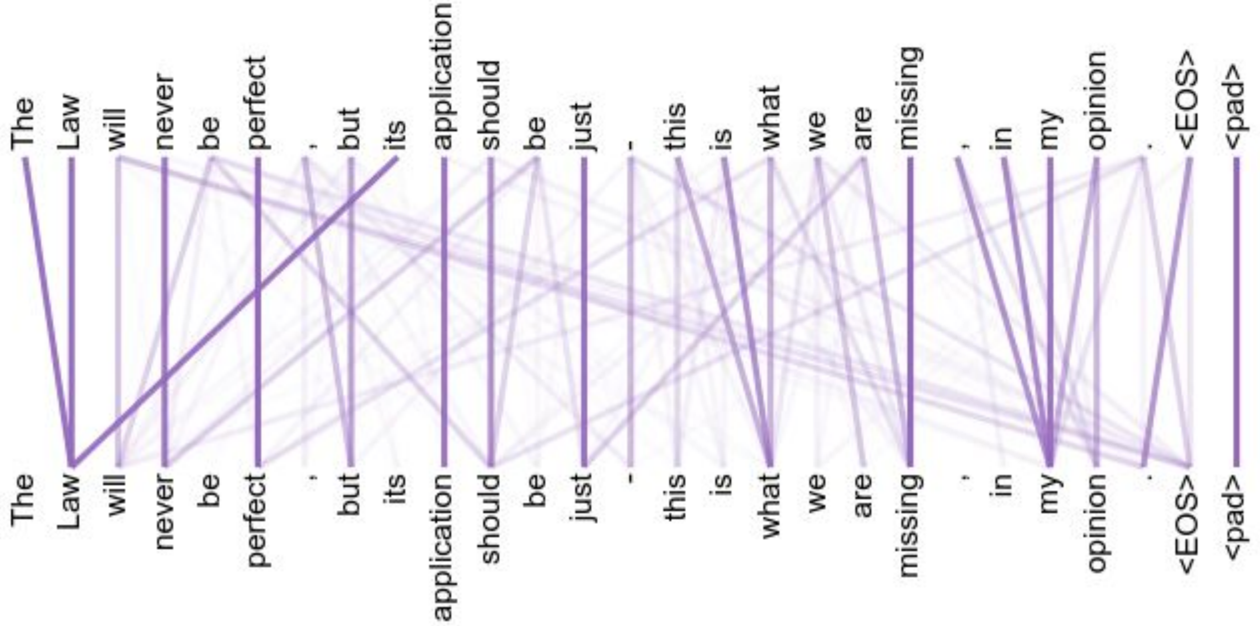
[4]

Goals and Metrics

Realistic	(Visual Inspection)
Diverse	Total Mutual Difference (TMD ↑)
Coherent reconstruction - GT resemblance	Unidirectional Hausdorff Distance (UHD ↓) Chamfer Distance (CD ↓) F-Score (↑)

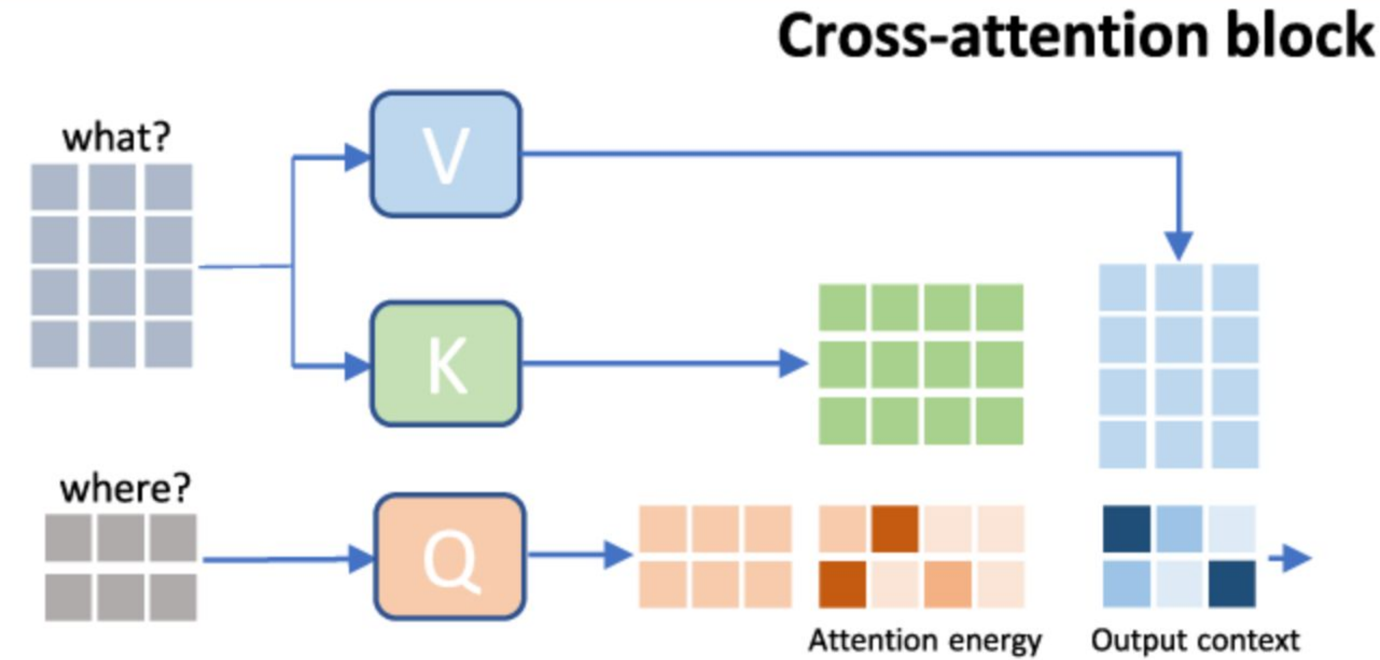
Attention


MatMul
SoftMax
Mask (opt.)
Scale
MatMul
Q
K
V


The Law will never be perfect , but its application should be just - this is what we are missing , in my opinion . <EOS> <pad>
The Law will never be perfect , but its application should be just - this is what we are missing , in my opinion . <EOS> <pad>

[6]

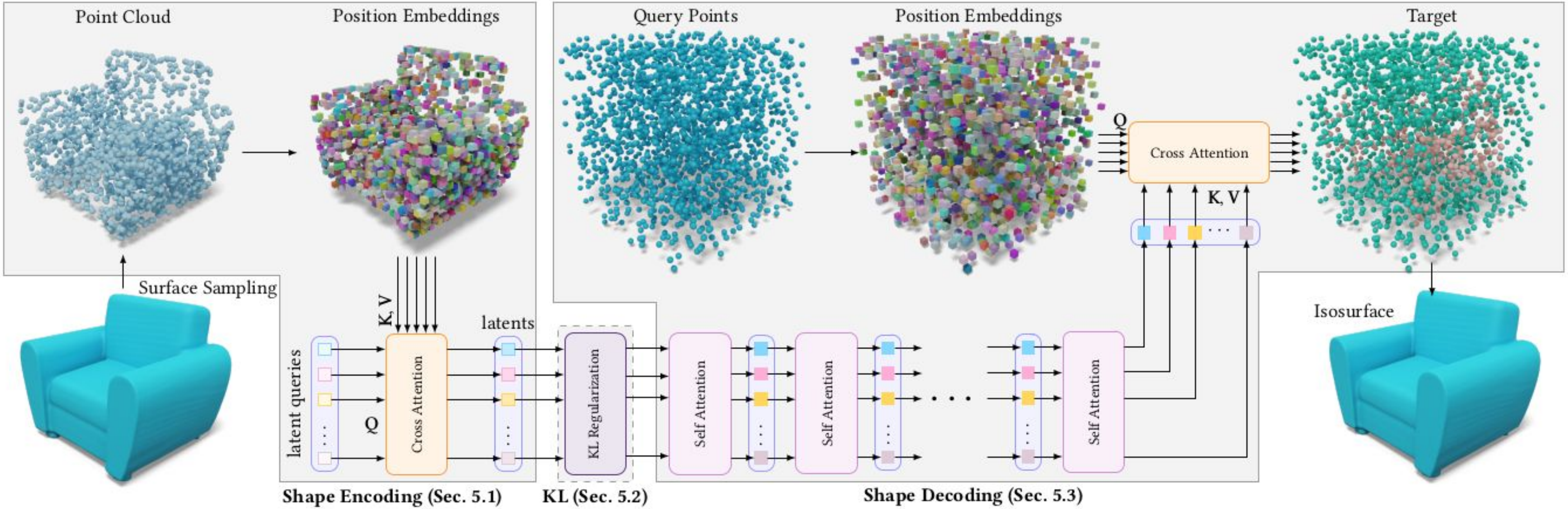
Cross Attention



$$Cross\_attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{C/h}}) \cdot V$$

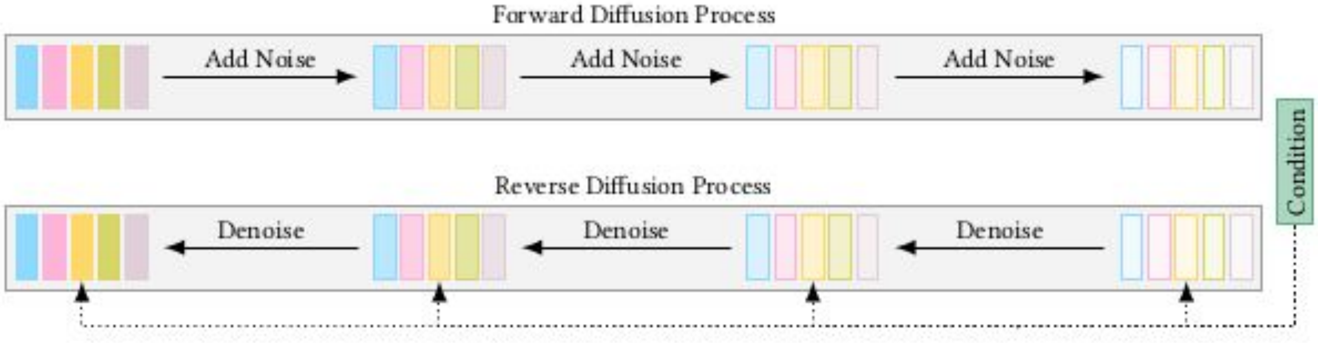
[11]

Methodology



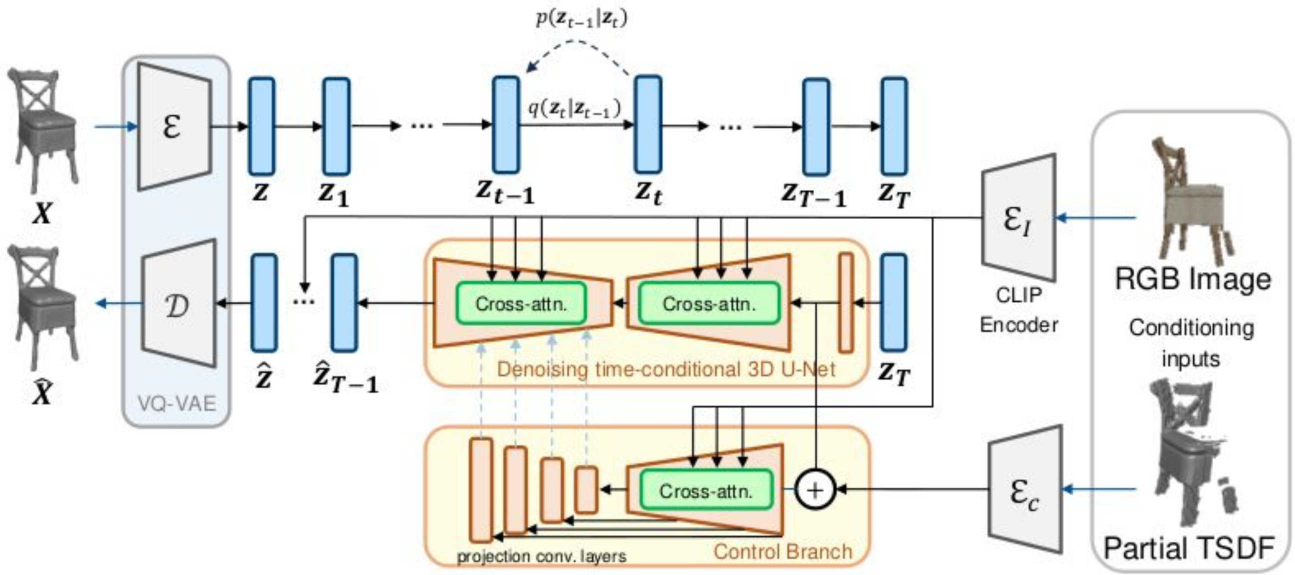
3DShape2VecSet Framework [8]

Methodology



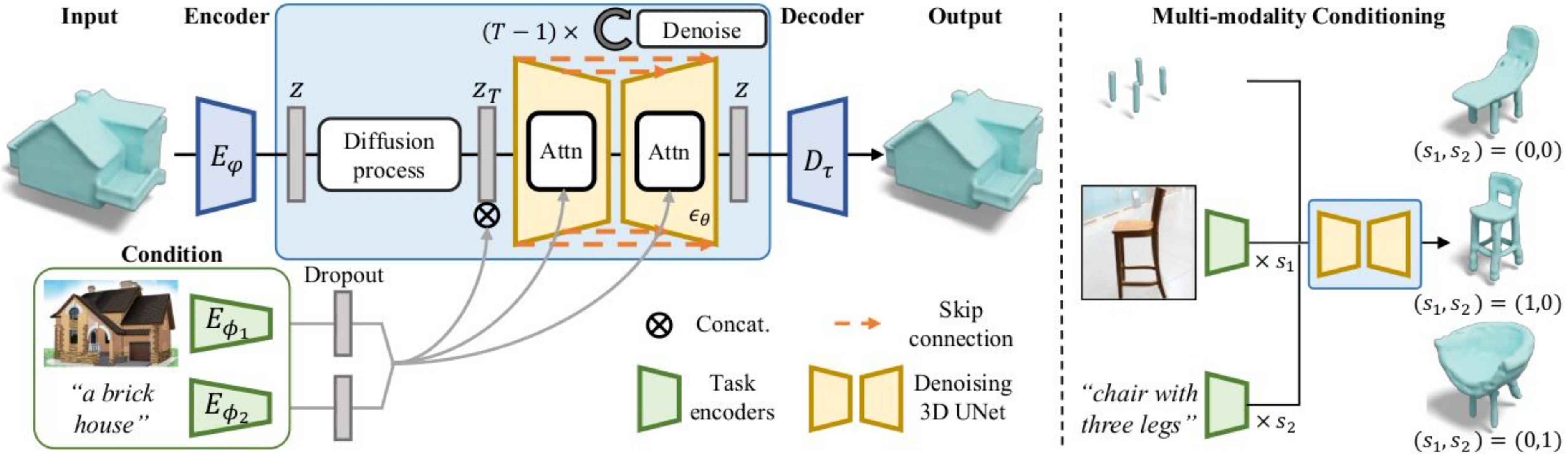
3DShape2VecSet Diffusion Process [8]

Methodology



SC-Diff Framework [9]

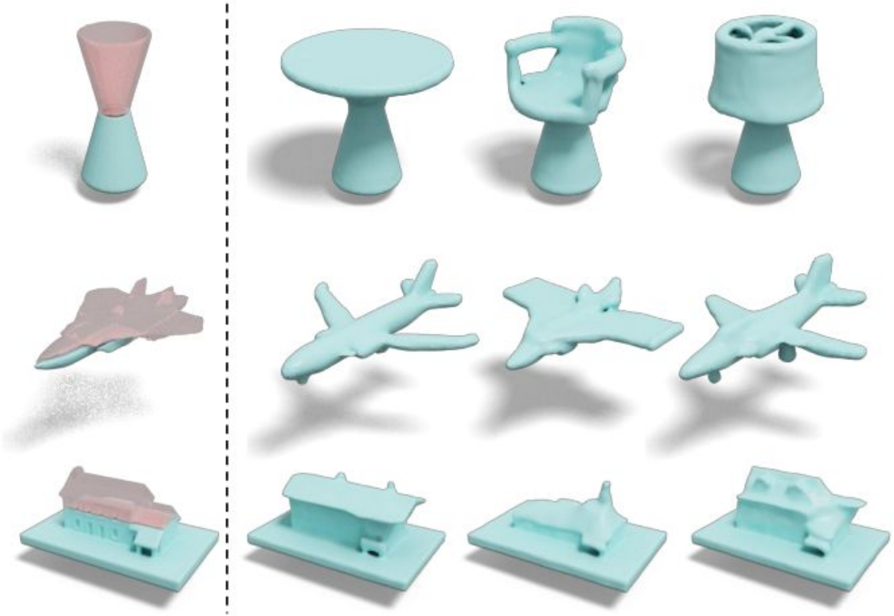
Methodology



SDFusion Framework [10]

Datasets

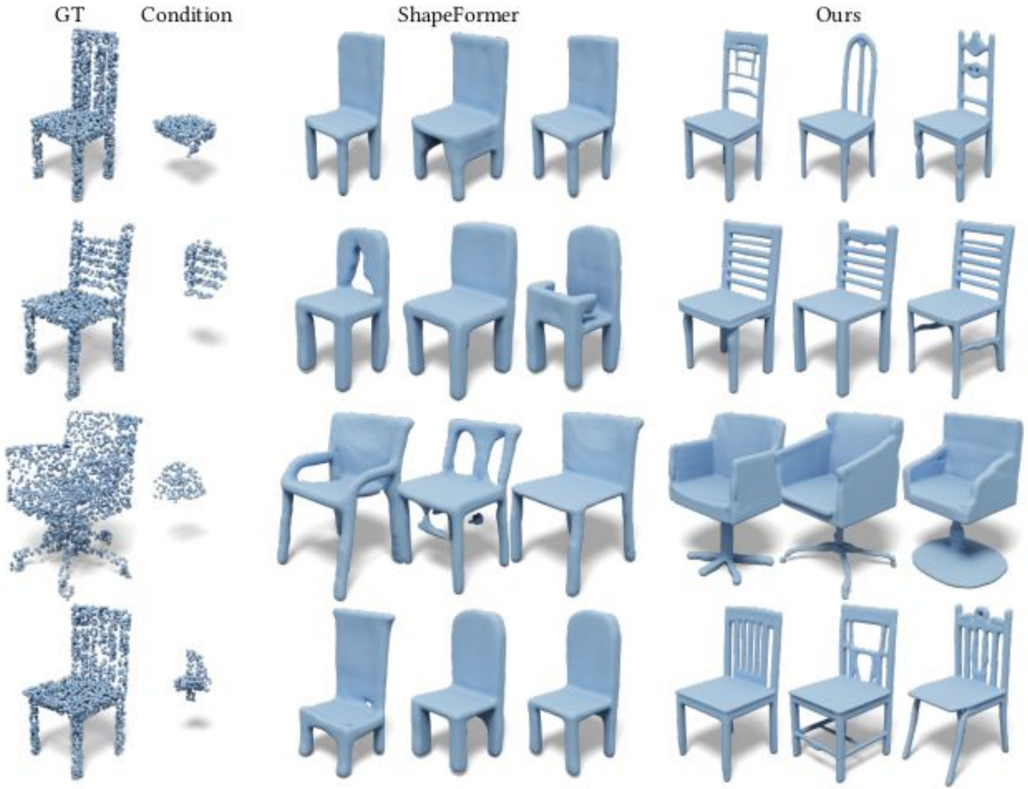
- ShapeNet
- BuildingNet



[7]

Results

3DShape2VecSet [8]



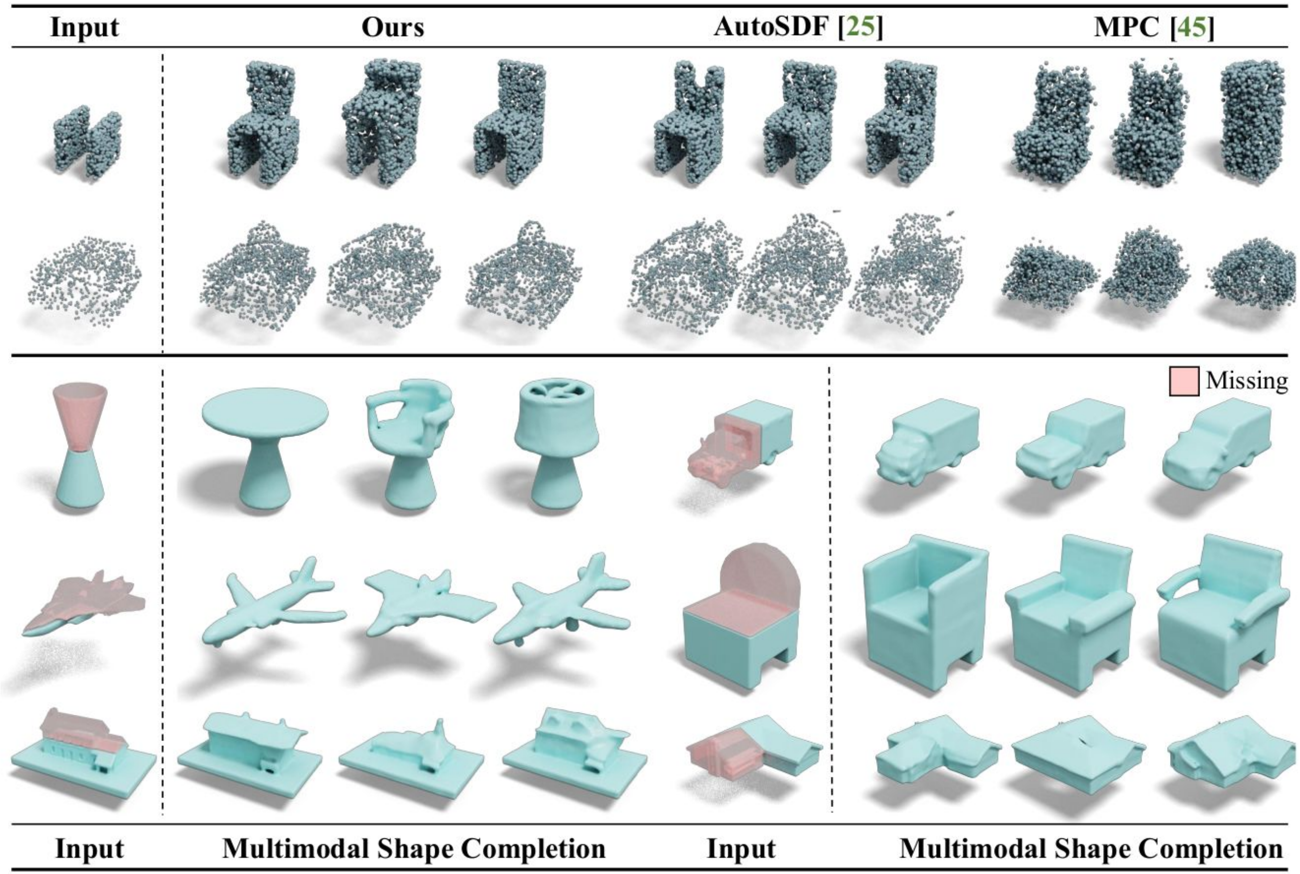
Results

SDFusion [10]

	ShapeNet		BuildingNet	
Method	UHD ↓	TMD ↑	UHD ↓	TMD ↑
MPC [45]	0.0627	0.0303	0.1350	0.0467
AutoSDF [25]	0.0567	0.0341	0.1208	0.0649
Ours	**0.0557**	**0.0885**	**0.1116**	**0.0745**

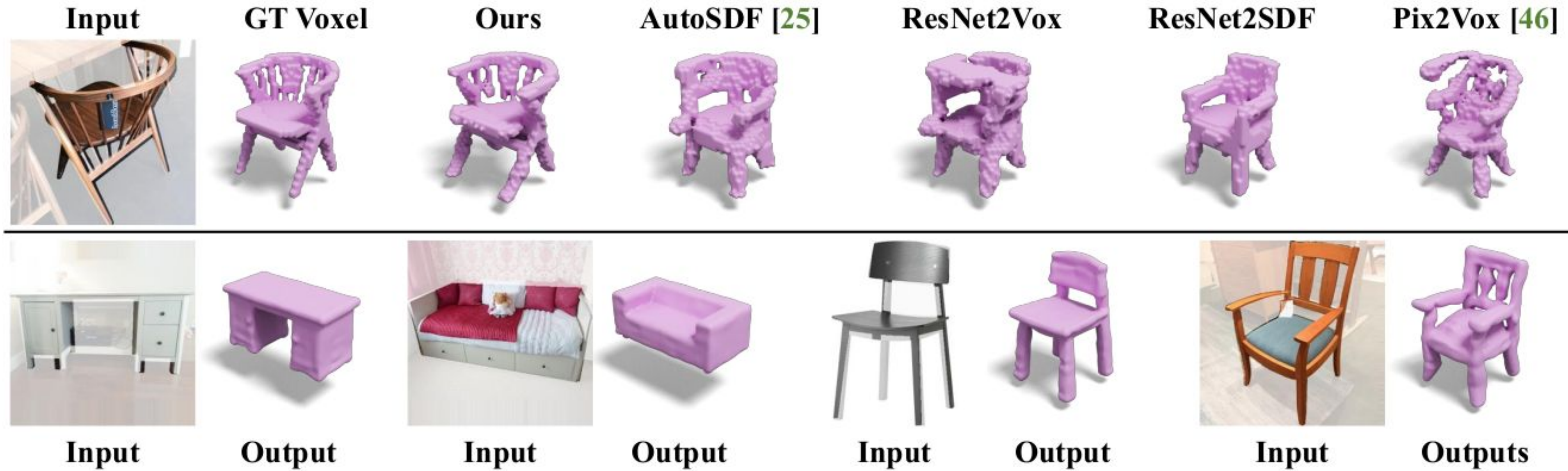
Results

SDFusion [10]



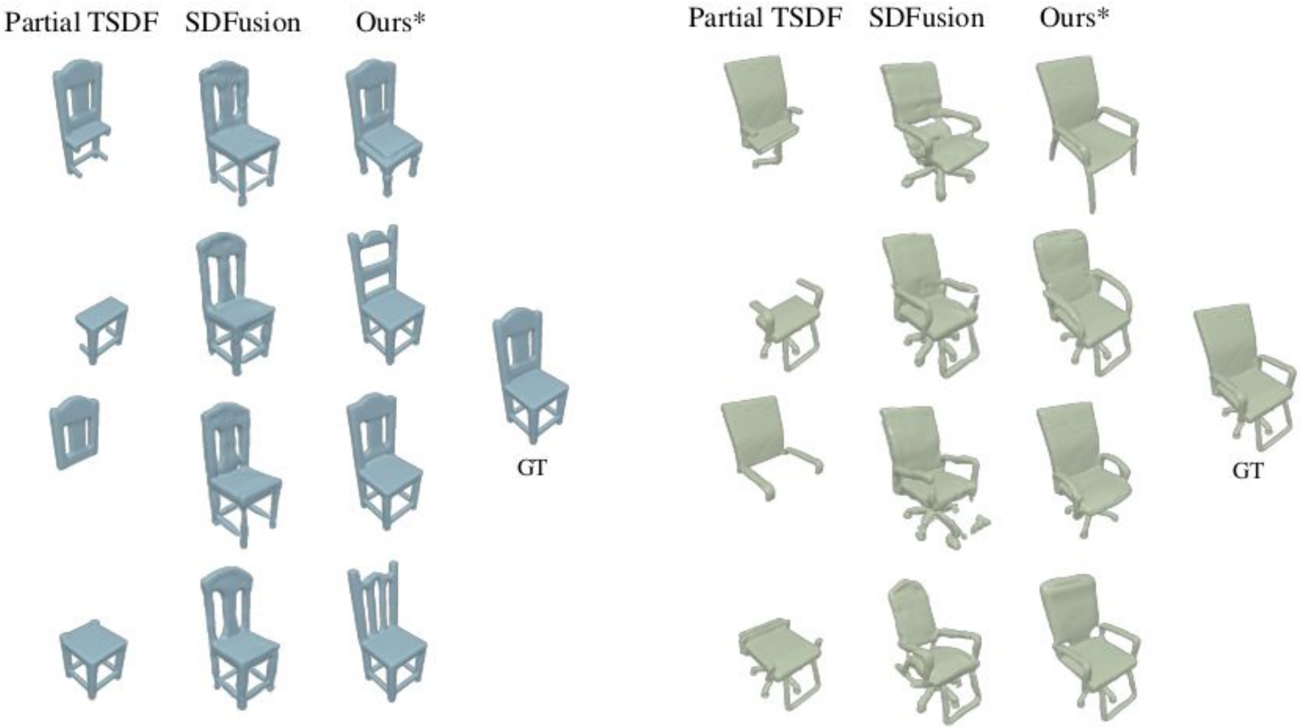
Results

SDFusion [10]



Results

SCDiff [9] vs
SDFusion [10]

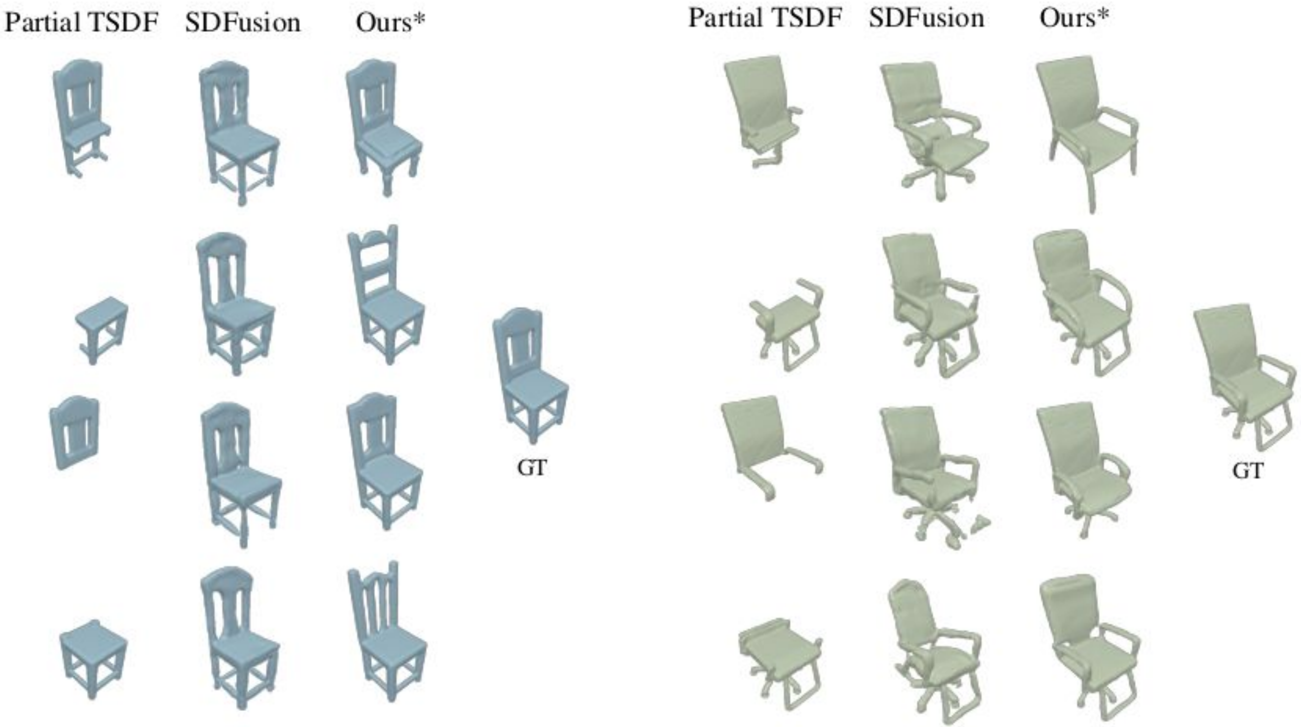


Discussion and Review

- Long time to draw samples in the diffusion process
- Dimensionality & Processing Speed
- **Mitigation**: Diffusion process in the latent space

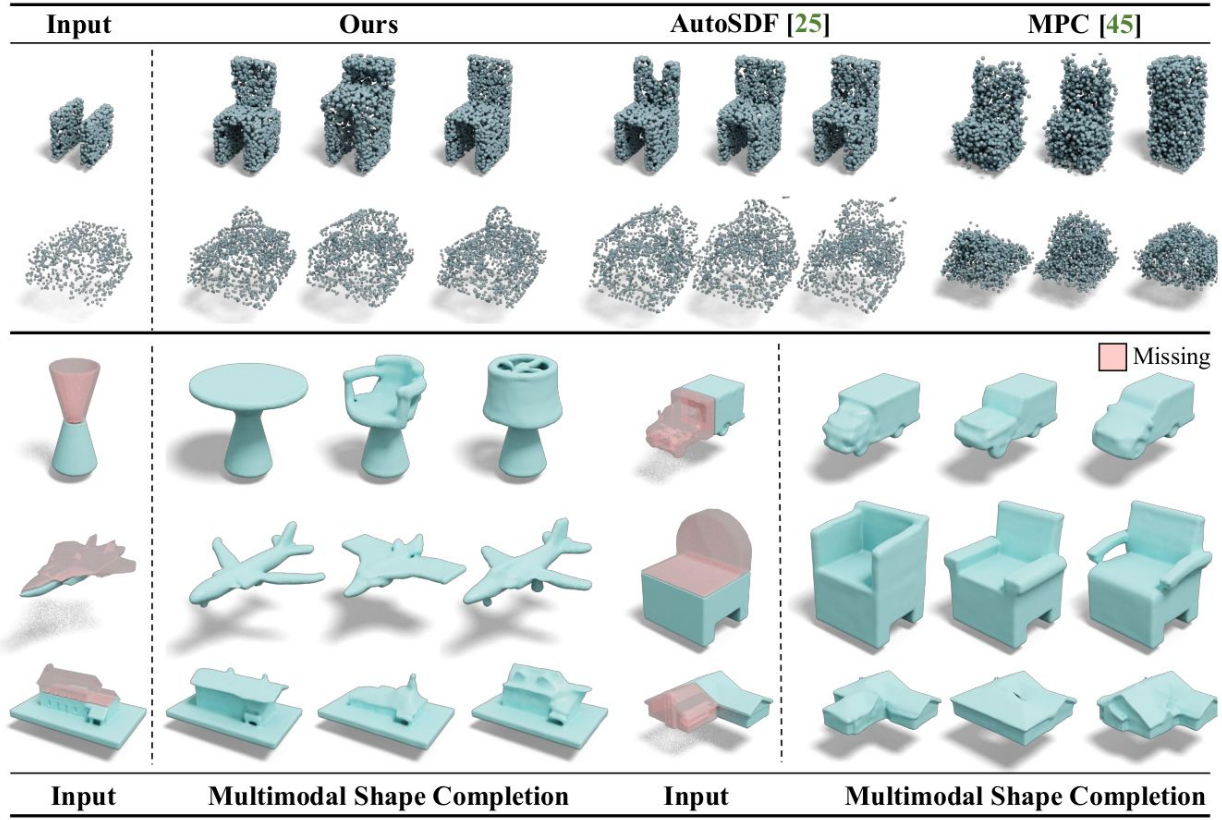
Discussion and Review - SDFusion Diversity

SCDiff [9] vs
SDFusion [10]



Discussion and Review - SDFusion Diversity

SDFusion [10]



Summary

- SOTA diffusion based 3D shape completion approaches yield promising results - also for **multimodal** conditioned input -> cross-attention

- **Key challenges**: Data acquisition, dimensionality, computational limits, drawing samples in the denoising process takes some time

- Common **mitigation**: Diffusion processes in latent space, subsampling
- Future work:
    - Better benchmarks
    - More compute
    - Development of new models

Backup

Chamfer Distance

$$d_{CD}(S_1, S_2) = \frac{1}{|S_1|} \sum_{x \in S_1} \min_{y \in S_2} ||x - y||_2 + \frac{1}{|S_2|} \sum_{y \in S_2} \min_{x \in S_1} ||y - x||_2$$

F-Score

$$F(d) = \frac{2P(d)R(d)}{P(d) + R(d)}$$

Precision ~ accuracy
Recall ~ completeness

Bildquellen

Pictograms: https://icons8.de/icons/set/voxel

Voxel Grid: [1] https://discourse.vtk.org/t/vtkmarchingcubes-vtkdiscretemarchingcubes-does-not-produce-a-closed-mesh-surface/4312

Polygon Mesh: [2]

https://en.wikipedia.org/wiki/Polygon_mesh#/media/File:Dolphin_triangle_mesh.png

TSDF: [3]

Daun, Kevin, Stefan Kohlbrecher, Jürgen Sturm and Oskar von Stryk. “Large Scale 2D Laser SLAM using Truncated Signed Distance Functions.” *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)* (2019): 222-228.

Neural Field: [4]

https://cameronrwolfe.substack.com/p/understanding-nerfs

Teabunny: [5]

https://cohost.org/snippid/post/6371906-teabunny

Attention: [6]

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need, http://arxiv.org/abs/1706.03762, (2023). https://doi.org/10.48550/arXiv.1706.03762.

3DShapes: [7]

Cheng, Y.-C., Lee, H.-Y., Tulyakov, S., Schwing, A., Gui, L.: SDFusion: Multimodal 3D Shape Completion, Reconstruction, and Generation, http://arxiv.org/abs/2212.04493, (2023). https://doi.org/10.48550/arXiv.2212.04493.

Cross Attention: [11]

Nguyen, Ngoc-Quang & Park, Sejeong & Gim, Mogan & Kang, Jaewoo. (2024). MulinforCPI: enhancing precision of compound–protein interaction prediction through novel perspectives on multi-level information integration. Briefings in Bioinformatics. 25. 10.1093/bib/bbad484.

Quellen

3DShape2VecSet [8]:

Zhang, B., Tang, J., Niessner, M., Wonka, P.: 3DShape2VecSet: A 3D Shape Representation for Neural Fields and Generative Diffusion Models, http://arxiv.org/abs/2301.11445, (2023). https://doi.org/10.48550/arXiv.2301.11445.

SC-Diff [9]:

Galvis, J.D., Zuo, X., Schaefer, S., Leutengger, S.: SC-Diff: 3D Shape Completion with Latent Diffusion Models, http://arxiv.org/abs/2403.12470, (2024). https://doi.org/10.48550/arXiv.2403.12470.

SDFusion[10]:

Cheng, Y.-C., Lee, H.-Y., Tulyakov, S., Schwing, A., Gui, L.: SDFusion: Multimodal 3D Shape Completion, Reconstruction, and Generation, http://arxiv.org/abs/2212.04493, (2023). https://doi.org/10.48550/arXiv.2212.04493.

DiffComplete[11]:

Chu, R., Xie, E., Mo, S., Li, Z., Nießner, M., Fu, C.-W., Jia, J.: DiffComplete: Diffusion-based Generative 3D Shape Completion, http://arxiv.org/abs/2306.16329, (2023). https://doi.org/10.48550/arXiv.2306.16329.

Diffusion-based 3D Shape completion

Technische Universität München

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING