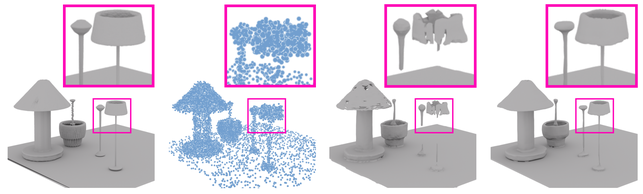
ALTO: Alternating Latent Topologies for Implicit 3D Reconstruction

Zhen Wang[1*], Shijie Zhou[1*], Jeong Joon Park[2], Despoina Paschalidou[2],
Suya You[3], Gordon Wetzstein[2], Leonidas Guibas[2], Achuta Kadambi[1]
University of California, Los Angeles[1], Stanford University[2] , DEVCOM[3]

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Contributions

- We introduce an iterative technique to blend the strengths of different latent topologies for high-fidelity conditional neural fields generation.
- We propose an attention-based decoder that replaces naive linear interpolation of feature-grids or computationally expensive point-wise attention while keeping compute burden in check.
- We demonstrate performance and runtime improvements over the highest-quality previous method, as well as performance improvements over all other baselines.



References

[1] Chang, Angel X., et al. "Shapenet: An information-rich 3d model repository." arXiv preprint arXiv:1512.03012 (2015).

[2] Peng, Songyou, et al. "Convolutional occupancy networks." In ECCV, pp. 523-540, 2020.

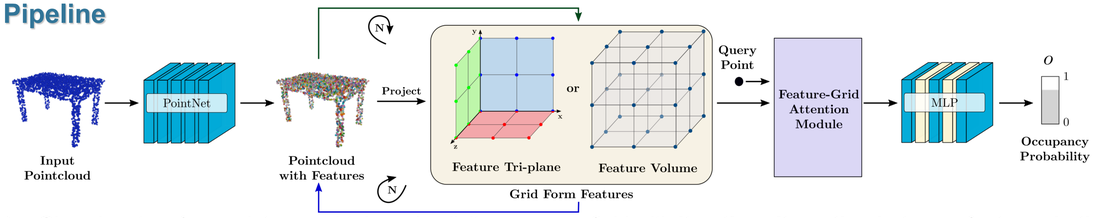
[3] Dai, Angela, et al. "Scannet: Richly-annotated 3d reconstructions of indoor scenes." In CVPR, pp. 5828-5839. 2017.

[4] Boulch, Alexandre, and Renaud Marlet. "Poco: Point convolution for surface reconstruction." In CVPR, pp. 6302-6314. 2022.

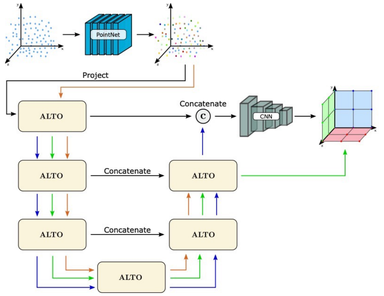
[5] Kazhdan, Michael, and Hugues Hoppe. "Screened poisson surface reconstruction." ACM Transactions on Graphics (ToG) 32.3 (2013): 1-13.

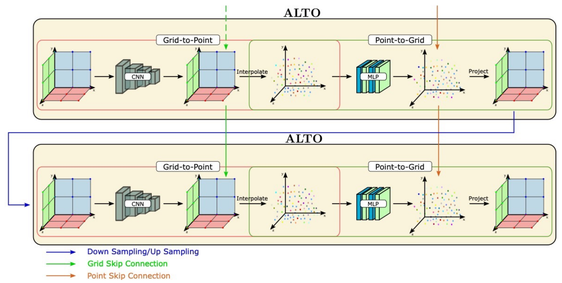
Method

Pipeline



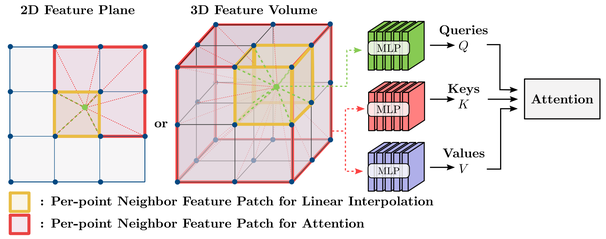
- Given input surface points, we obtain an implicit occupancy field with **iterative alternation between features in the forms of points and 2D or 3D grids**. Then we decode the occupancy values for query points with a learned attention-based interpolation from neighboring grids.





- As an example, we show the ALTO block instantiated by alternating between two latent topologies: point and triplanes via an “in-network” fashion, i.e. within each level of an hourglass framework U-Net.

Feature-Grid Attention Module

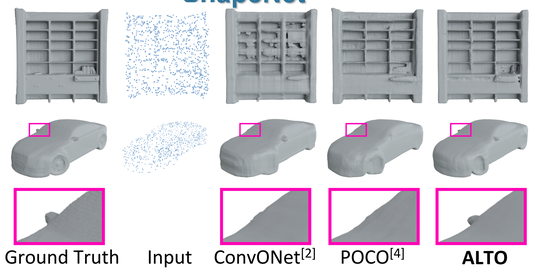


- Attention-based decoder on neighboring grids (2D or 3D)

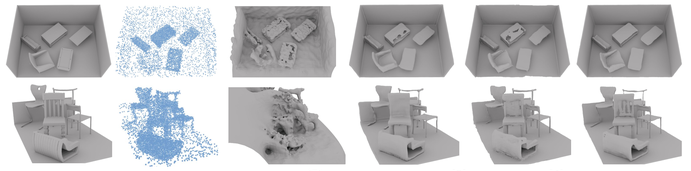
https://visual.ee.ucla.edu/alto.htm/

Experimental Results

ShapeNet[1]



Synthetic Room[2]



Generalization to ScanNet[3]