# TopoDiT-3D: Topology-Aware Diffusion Transformer with Bottleneck Structure for 3D Point Cloud Generation

BMVC 2025


Zechao Guan[1], Feng Yan[2], Shuai Du[1], Lin Ma[2], Qingshan Liu[1†]

[1] Southeast University, Nanjing, China [2] Meituan Company, Beijing, China

† Corresponding author; https://github.com/Zechao-Guan/TopoDiT-3D

Email: zechaoguan@gmail.com

Homepage: https://zechao-guan.github.io/


## Abstract


Recent advancements in Diffusion Transformer (DiT) models have significantly improved 3D point cloud generation. However, existing methods primarily focus on local feature extraction while overlooking global topological information, such as voids, which are crucial for maintaining shape consistency and capturing complex geometries. To address this limitation, we propose TopoDiT-3D, a Topology-Aware Diffusion Transformer with a bottleneck structure for 3D point cloud generation. Specifically, we design the bottleneck structure utilizing Perceiver Resampler, which not only offers a mode to integrate topological information extracted through persistent homology into feature learning, but also adaptively filters out redundant local features to improve training efficiency. Experimental results demonstrate that TopoDiT-3D outperforms state-of-the-art models in visual quality, diversity, and training efficiency. Furthermore, TopoDiT-3D demonstrates the importance of rich topological information for 3D point cloud generation and its synergy with conventional local feature learning.


## Research Gap

### Gap 1: How to extract and utilize the rich topological information inherent in point clouds ?

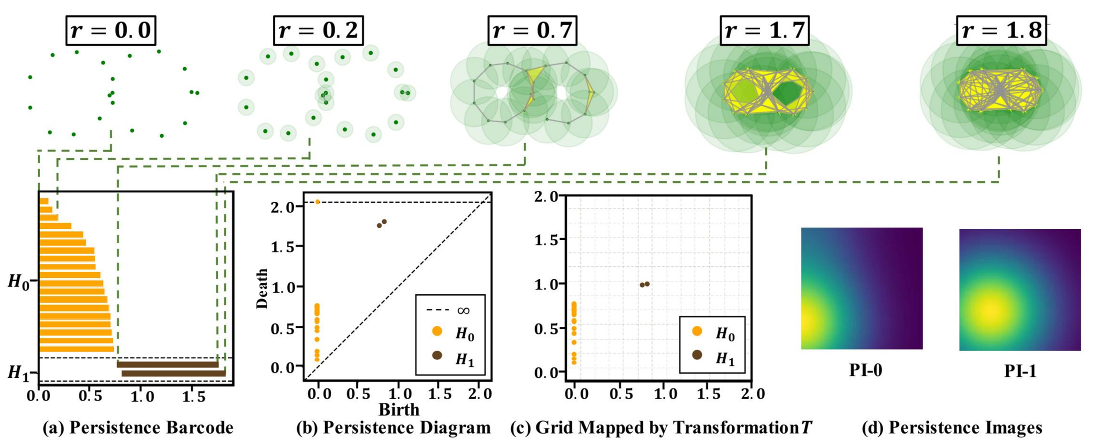


Figure 1: Vietoris-Rips Filtration and Persistent Homology

As the filtration radius $r$ increases, simplices are incrementally added to the complex, giving rise to the birth and death of topological features:

- Connected components (0D) and loops (1D) appear and vanish at specific values of $r$.
- Their lifespans are encoded as persistence barcodes (a) or, equivalently, as points in a persistence diagram (b)
- Applying a transformation $T$, the diagram is converted into persistence images (c, d), yielding grid-based, vectorizable representations of 0D and 1D persistent features.

### Gap 2: How to effectively integrate these topological features to enhance point cloud generation task while improving efficiency ?

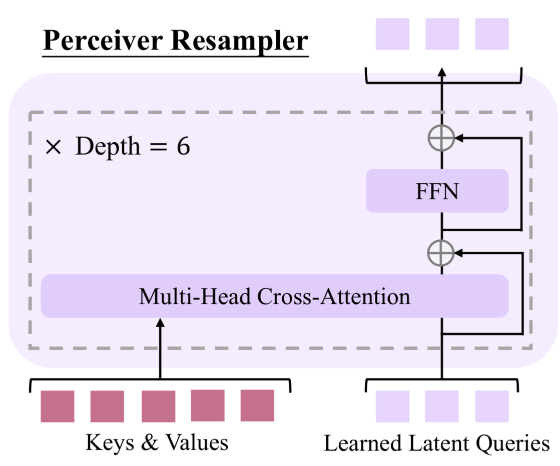


Figure 2: The Perceiver Resampler module.

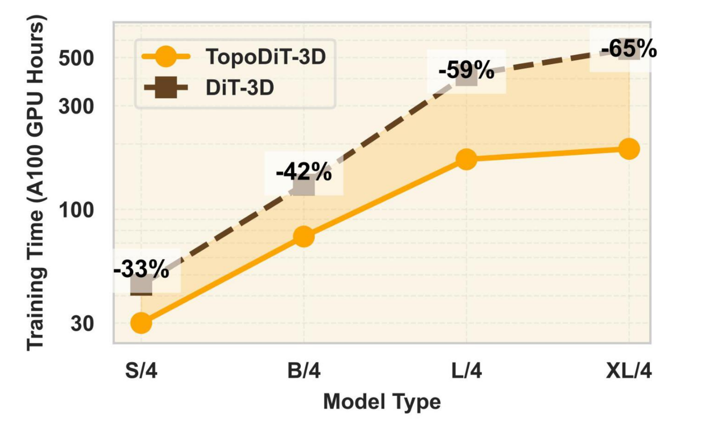


Figure 3: Comparison of training efficiency between TopoDiT-3D and DiT-3D.

Perceiver Resampler (Upsampling)

Bottleneck Structure

(DownSampling) Perceiver Resampler

Bottleneck Design with Perceiver Resampler

- Fuses topological features from persistent homology into representation learning.
- Adaptively prunes redundant patch tokens via Perceiver Resampler.
- Boosts training efficiency without sacrificing multiscale structure.

### Our Contributions

- TopoDiT-3D: A diffusion transfor-mer with a Perceiver-based bottleneck that injects global topological priors and prunes redundancy for efficient training.
- Persistent homology enables global shape understanding, pushing perfor-mance beyond DiT-3D and strong DDPM / non-DDPM baselines.
- Both topology and bottleneck design matter—ablation confirms each is criti-cal for capturing global structure and local detail.

### To address those limitations, we propose TopoDiT-3D, a Topology-Aware Diffusion Transformer with a bottleneck structure for 3D point cloud generation.

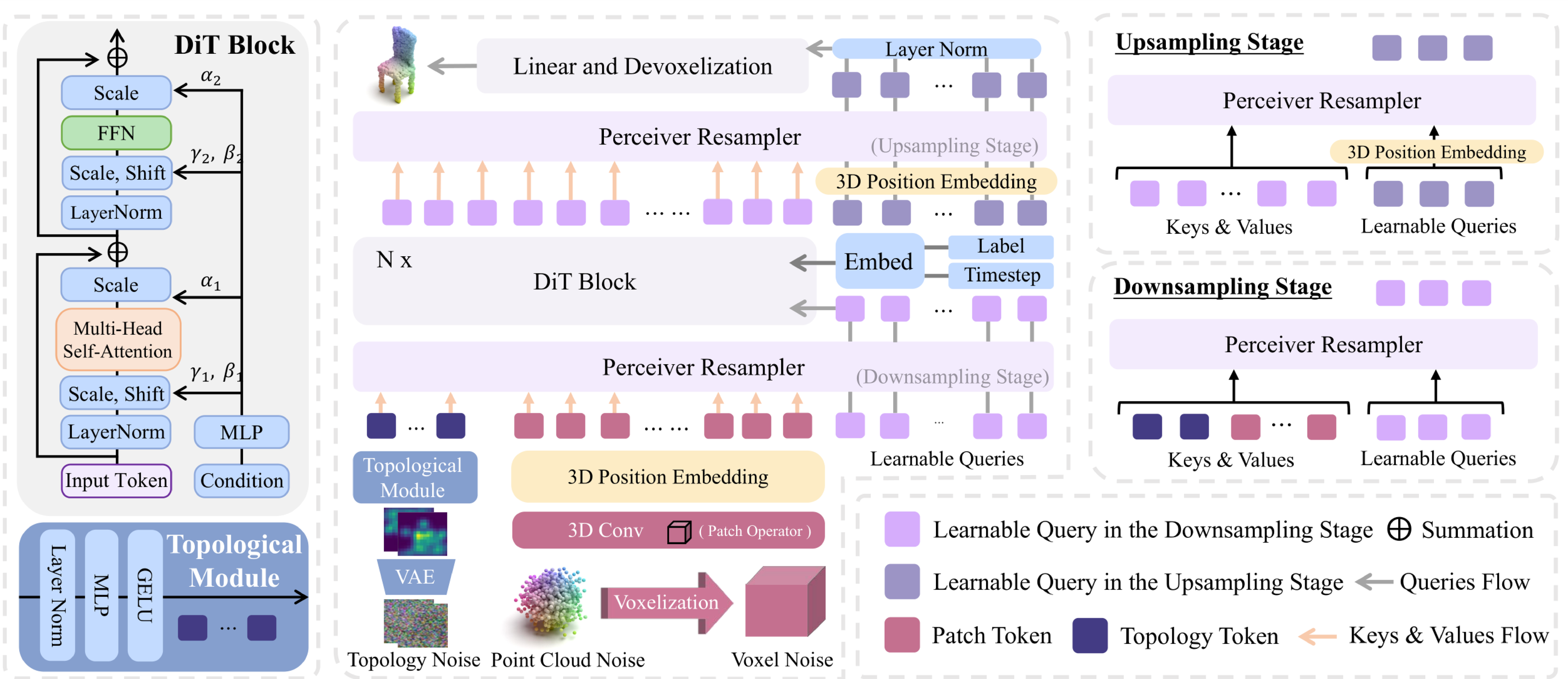


Figure 4: The illustration depicts the architecture of TopoDiT-3D.

## Topological Guidance in Generation

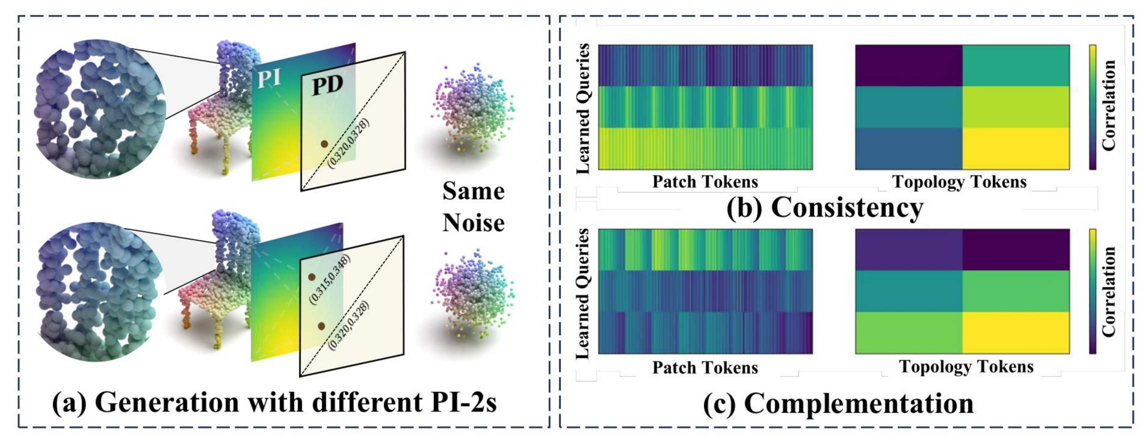


Figure 5: Vietoris-Rips Filtration and Persistent Homology

Topological Information Enhances Detail

- Using the same noise, richer persistence images produce point clouds with finer structural details (e.g., chair backrest voids; Fig.4(a)).

Two integration modes revealed by latent query analysis

- Consistency: Topology and patch tokens jointly inform global structure (Fig.4(b)).
- Complementarity: Topology tokens directly encode global shape, bypassing local redundancy (Fig.4(c)).

## ShapeNet Dataset

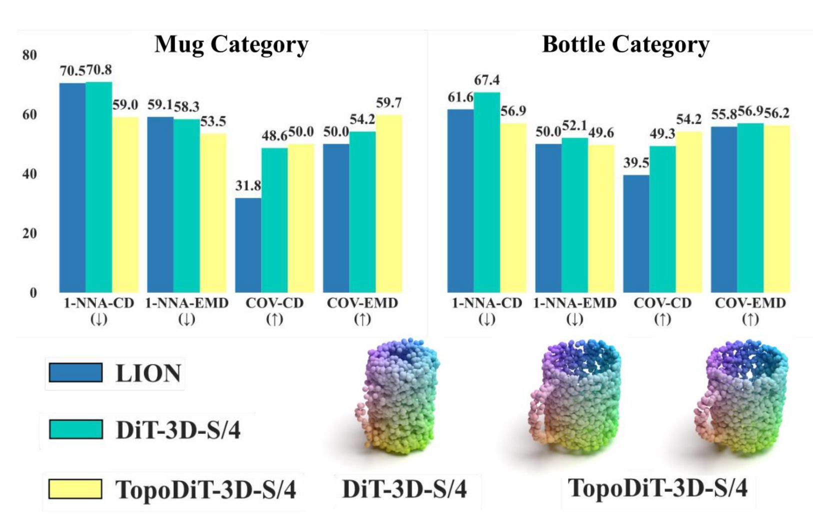


Figure 6: The result of mug and bottle categories on ShapeNet.

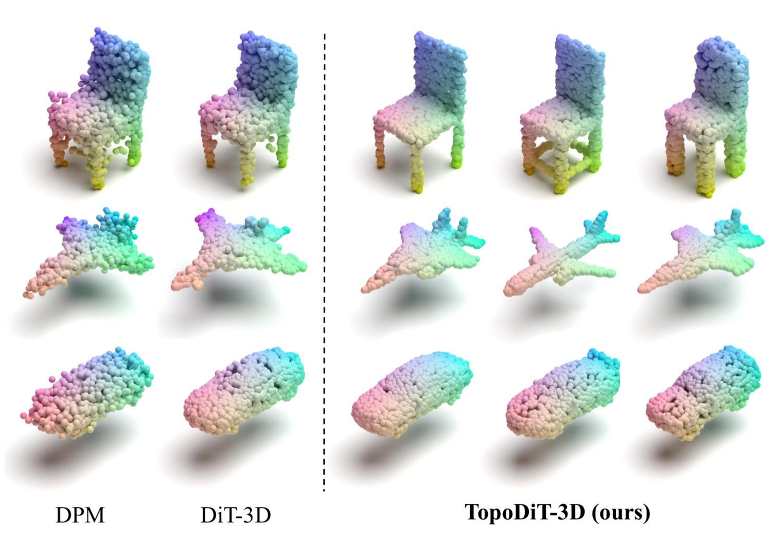


Figure 7: The visualizations of 3D point cloud generation.

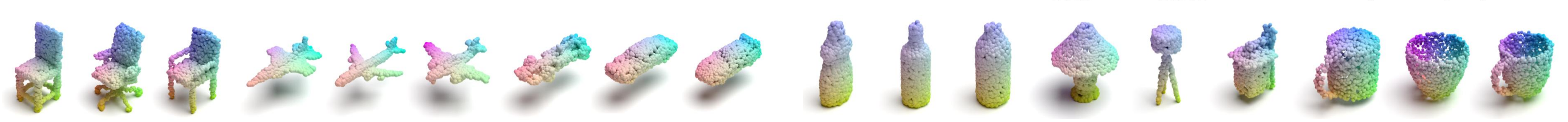

Table 1: Experimental results compare TopoDiT-3D with SOTA methods on 3D point cloud generation on ShapeNet.

| Method | Chair 1-NNA(↓) CD | Chair 1-NNA(↓) EMD | Chair COV(↑) CD | Chair COV(↑) EMD | Airplane 1-NNA(↓) CD | Airplane 1-NNA(↓) EMD | Airplane COV(↑) CD | Airplane COV(↑) EMD | Car 1-NNA(↓) CD | Car 1-NNA(↓) EMD | Car COV(↑) CD | Car COV(↑) EMD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| r-GAN [redacted] | 83.69 | 99.70 | 24.27 | 15.13 | 98.40 | 96.79 | 30.12 | 14.32 | 94.46 | 99.01 | 19.03 | 6.539 |
| l-GAN (CD) | 68.58 | 83.84 | 41.99 | 29.31 | 87.30 | 93.95 | 38.52 | 21.23 | 66.49 | 88.78 | 38.92 | 23.58 |
| l-GAN (EMD) | 71.90 | 64.65 | 38.07 | 44.86 | 89.49 | 76.91 | 38.27 | 38.52 | 71.16 | 66.19 | 37.78 | 45.17 |
| PointFlow [redacted] | 62.84 | 60.57 | 42.90 | 50.00 | 75.68 | 70.74 | 47.90 | 46.41 | 58.10 | 56.25 | 46.88 | 50.00 |
| SoftFlow [redacted] | 59.21 | 60.05 | 41.39 | 47.43 | 76.05 | 65.80 | 46.91 | 47.90 | 64.77 | 60.09 | 42.90 | 44.60 |
| SetVAE [redacted] | 58.84 | 60.57 | 46.83 | 44.26 | 76.54 | 67.65 | 43.70 | 48.40 | 59.94 | 59.94 | 49.15 | 46.59 |
| DPF-Net [redacted] | 62.00 | 58.53 | 44.71 | 48.79 | 75.18 | 65.55 | 46.17 | 48.89 | 62.35 | 54.48 | 45.74 | 49.43 |
| DPM [redacted] | 60.05 | 74.77 | 44.86 | 35.50 | 76.42 | 86.91 | 48.64 | 33.83 | 68.89 | 79.97 | 44.03 | 34.94 |
| PVD [redacted] | 57.09 | 60.87 | 36.68 | 49.24 | 73.82 | 64.81 | 48.88 | 52.09 | 54.55 | 53.83 | 41.19 | 50.56 |
| LION [redacted] | 53.70 | 52.34 | 48.94 | 52.11 | 67.41 | 61.23 | 47.16 | 49.63 | 53.41 | 51.14 | 50.00 | 56.53 |
| GET3D [redacted] | 75.26 | 72.49 | 43.36 | 42.77 | - | - | - | - | 75.26 | 72.49 | 15.04 | 18.38 |
| MeshDiffusion [redacted] | 53.69 | 57.63 | 46.00 | 46.71 | 66.44 | 76.26 | 47.34 | 42.15 | 81.43 | 87.84 | 34.07 | 25.85 |
| DiT-3D [redacted] | 49.11 | 50.73 | 52.45 | 54.32 | 62.35 | 58.67 | 53.16 | 54.39 | 48.24 | 49.35 | 50.00 | 56.38 |
| TopoDiT-3D (ours) | **46.91**(-2.2) | **46.43**(-4.3) | **54.51**(+2.06) | **56.62**(+2.3) | **53.18** (-9.17) | **50.61**(-8.06) | **64.70**(+11.54) | **61.27**(+6.88) | **44.46**(-3.78) | **43.46**(-5.89) | **55.56**(+5.56) | **62.50**(+6.12) |

## Ablation Studies

Table 2: Ablation of the comp-onents of TopoDiT-3D.

| Model | Bottleneck Structure | Position Emdding | Topological Information | 1-NNA(↓) CD | 1-NNA(↓) EMD | COV(↑) CD | COV(↑) EMD |
|---|---|---|---|---|---|---|---|
| DiT-3D-S/4 | ✗ | ✗ | ✗ | 56.31 | 55.82 | 47.21 | 50.75 |
| TopoDiT-3D-S/4 | ✓ | ✗ | ✗ | 57.76 (+1.45) | 59.39 (+3.57) | 49.80 (+2.59) | 48.75 (-2.00) |
| | ✗ | ✓ | ✗ | 57.85 (+1.54) | 56.65 (+0.83) | 51.85 (+4.64) | 51.65 (+0.90) |
| | ✓ | ✓ | ✗ | 52.86 (-3.45) | 54.84 (-0.98) | 52.56 (+5.35) | 51.05 (+0.30) |
| | ✓ | ✓ | ✓ | **49.84** (-6.47) | **46.46** (-9.36) | **58.28** (+11.07) | **60.69** (+9.94) |

Both components are essential—each contributes uniquely to the performance of TopoDiT-3D.

Table 3: Ablation on TopoDiT-3D model sizes.

| Model | 1-NNA(↓) CD | 1-NNA(↓) EMD | COV (↑) CD | COV (↑) EMD |
|---|---|---|---|---|
| DiT-3D-XL/4 | 48.24 | 49.35 | 50.00 | 56.38 |
| TopoDiT-3D-S/4 | 57.95 | 59.09 | 55.68 | 51.98 |
| TopoDiT-3D-B/4 | 45.31 | 43.89 | 55.39 | 58.23 |
| TopoDiT-3D-L/4 | 44.48 | 44.88 | **58.52** | **63.06** |
| TopoDiT-3D-XL/4 | **44.46** | **43.46** | 55.68 | 62.50 |

The bottleneck structure does not violate the model scaling. For more interesting conclusions, you can refer to the paper