MamTiff-CAD: Multi-Scale Latent Diffusion with Mamba+ for Complex Parametric Sequence

Liyuan Deng[1], Yunpeng Bai[2], Yongkang Dai[1], Xiaoshui Huang[3], Hongping Gan[1], Dongshuo Huang[1], Hao Jiacheng[4], Yilei Shi[1*]

[1]Northwestern Polytechnical University & [2]National University of Singapore & [3]Shanghai Jiao Tong University & [4]Nanchang University

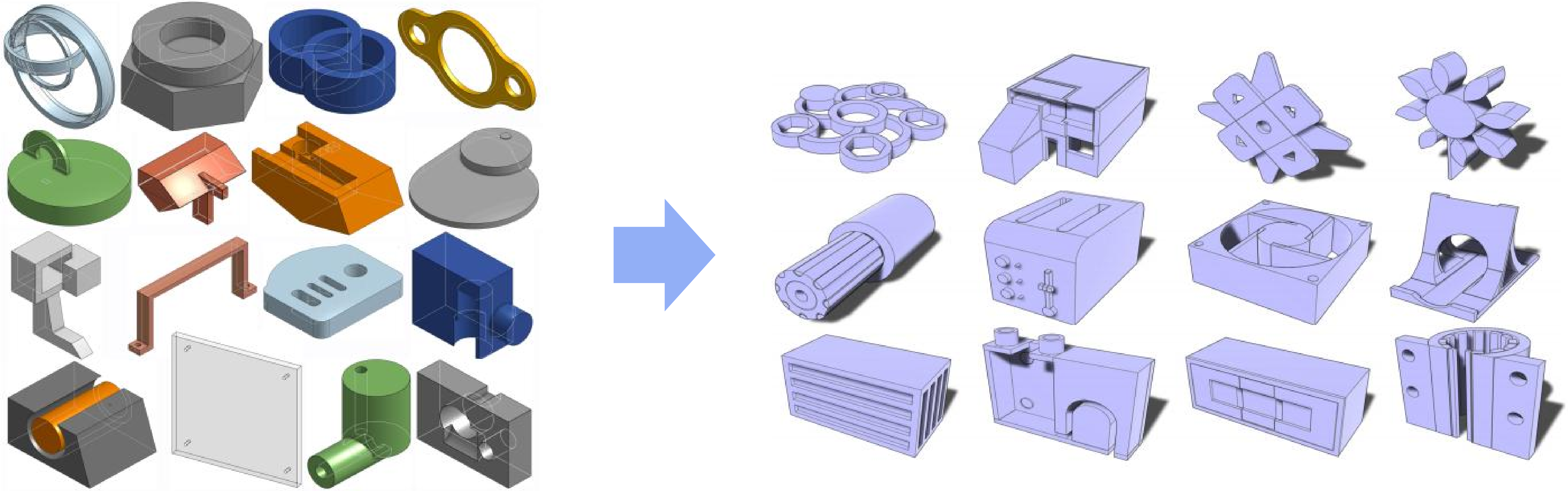
Introduction

Background:

- Industrial-grade CAD relies on *parametric command sequences* to record both geometry and design intent, thereby enabling editable and iterative design.
  In industrial production, design workflows often involve **hundreds of commands**, where traditional approaches face significant bottlenecks in **long-sequence modeling and computational efficiency**.

Research goal:

- Our objective is to learn the **generative distribution of long-sequence CAD commands in latent space**, enabling the synthesis of **more complex and executable CAD models**.

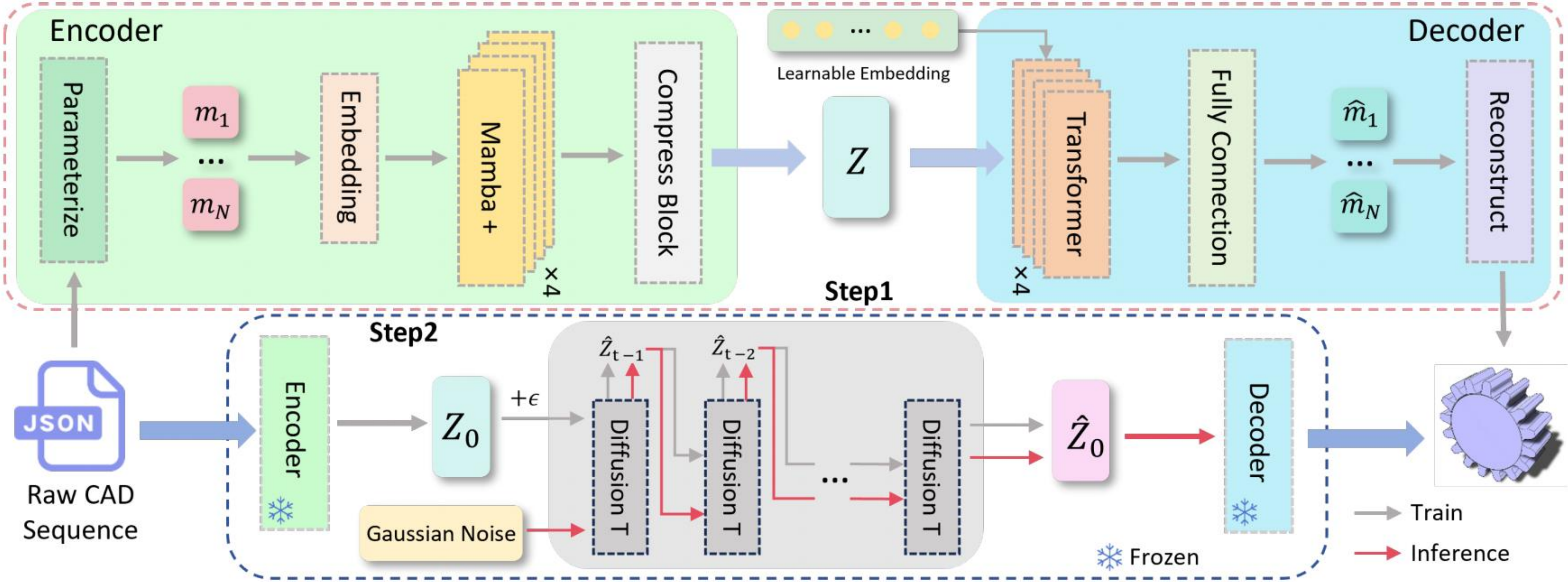


Method

- **Framework**

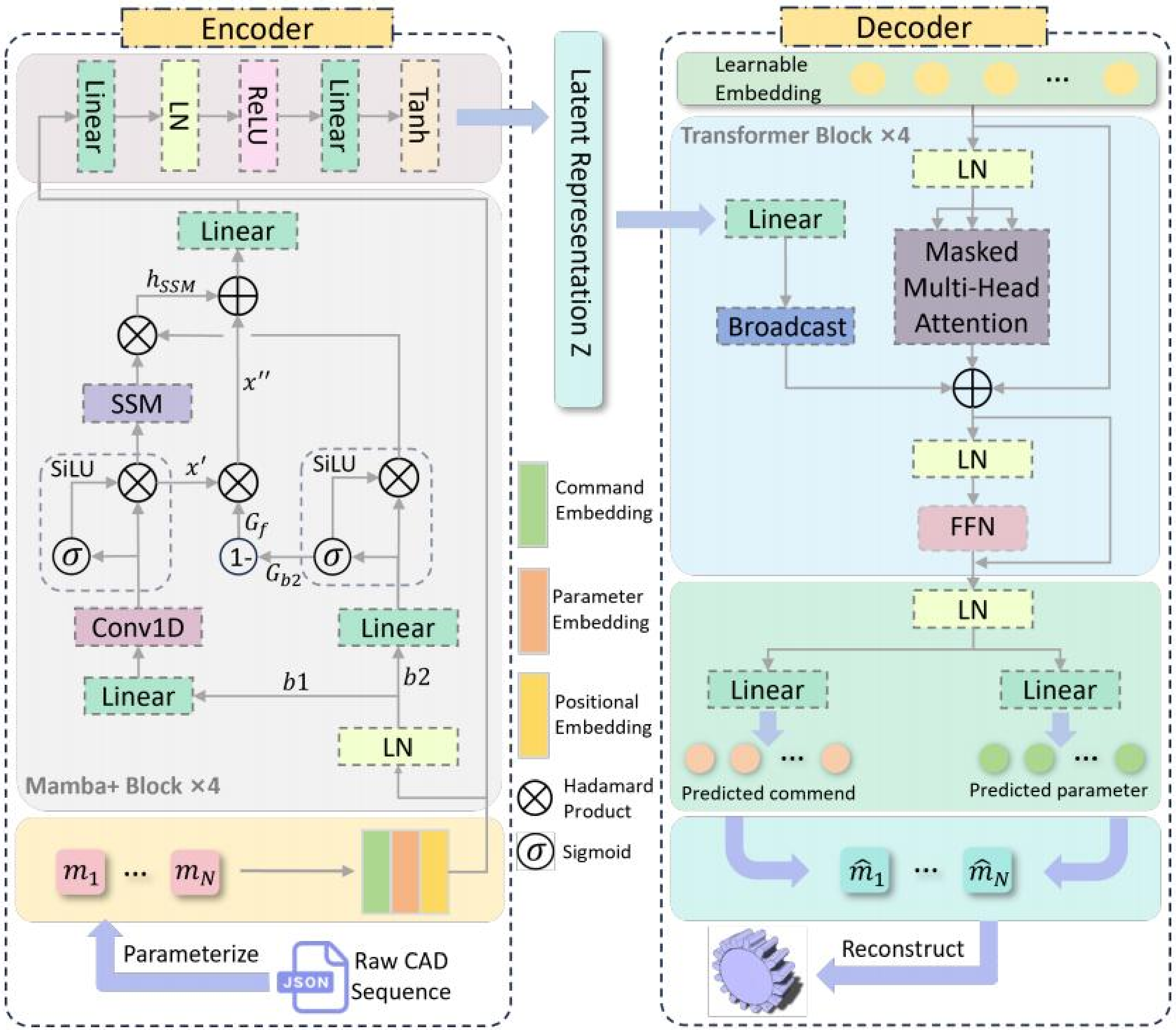
We propose a two-stage training framework, **MamTiff-CAD**:

Step 1: Autoencoder **Step 2:** Latent-space diffusion generation



Method

- **Step 1: Autoencoder（Mamba+ & Transformer)**



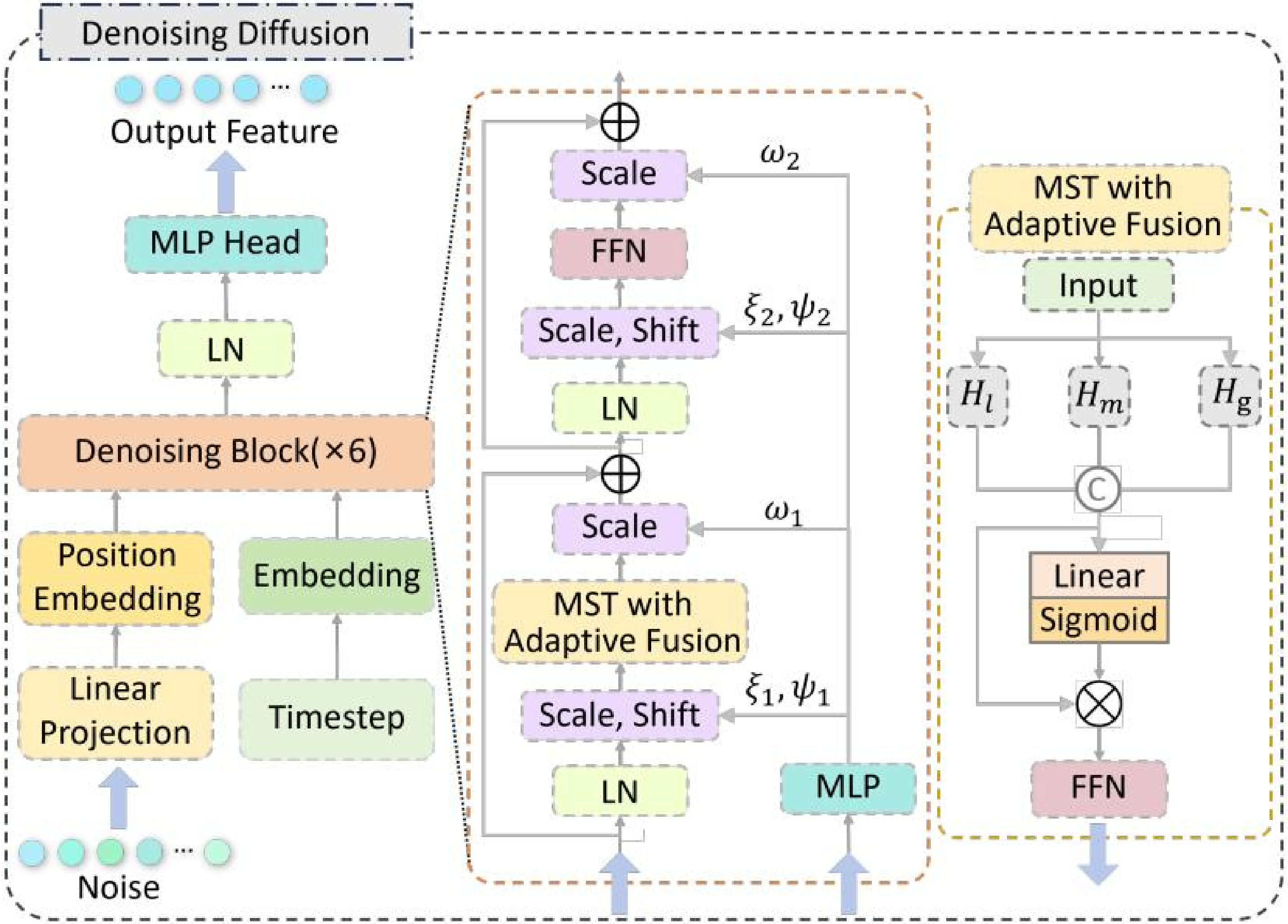
Representation:
Each command consists of a type C_i and **16 quantized parameters**.The sequence length is capped at N_c = 256.

Mamba+ Encoder:
An enhanced SSM encoder with an additional **forget gate** G_f to preserve historical information and mitigate long-range dependency degradation.

Transformer Decoder:
A **non-autoregressive** decoder that takes the latent vector Z and learnable positional embeddings as input, predicting the **entire sequence of command types and parameters in parallel**.

Method

- **Step 2： Latent-Space Diffusion Generation**



Multi-Scale Attention:
Attention windows of **64 / 128 / 256**, capturing **local geometry**, **mid-range topology**, and **global consistency** respectively.

Adaptive Fusion:
Gate-based fusion mechanism for integrating outputs from multiple attention scales.

Sequence-Aware Positional Encoding:
Sine-based positional encoding with learnable weights, enabling better **sequence-awareness** during generation.

Dataset

- **Dataset ABC-256**

Total of **13,705 samples** (10,964 training / 1,370 validation / 1,371 testing) with an **average sequence length of 99**.

Dataset	Total	Average Length	1-10	11-60	61-128	129-256
DeepCAD	178,238	15	44.58	55.42	-	-
ABC-256	13,705	99	-	-	82.89	17.11

Experiment

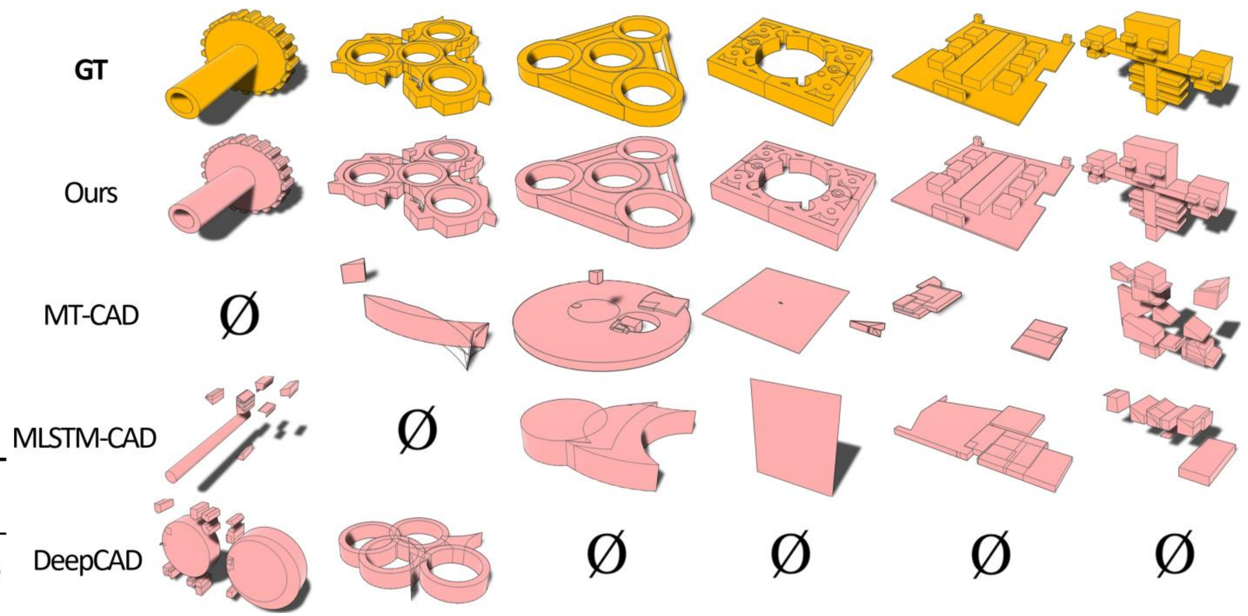
- **Reconstruction Comparison**

Method	$ACC_c \uparrow$	$ACC_p \uparrow$	MCD $\downarrow$	IR $\downarrow$	SR $\uparrow$
DeepCAD	92.24	75.93	41.02	33.11%	70.46%
MT-CAD	89.72	66.87	121.35	39.89%	63.97%
MLSTM-CAD	86.09	65.55	112.89	42.53%	59.85%
OURS	**99.99**	**99.93**	**0.75**	**8.50%**	**93.93%**

- **Fusion360 Generalization**

Method	$ACC_c \uparrow$	$ACC_p \uparrow$	MCD $\downarrow$	IR $\downarrow$	SR $\uparrow$
DeepCAD	93.35	77.99	104.76	19.41%	82.57%
MT-CAD	91.85	60.18	300.21	11.81%	90.17%
MLSTM-CAD	88.42	62.13	261.80	23.39%	80.97%
OURS	**99.99**	**97.99**	**1.44**	**5.70%**	**95.16%**

- **Reconstruction Visualization Comparison**



Experiment

- **Unconditional Generation Analysis**

Method	MMD ↓	JSD ↓	COV ↑	Unique ↑	Novel ↑	SR ↑
DeepCAD	2.66	6.49	56.66%	75.8	88.0	23.96%
SkexGen	2.31	4.53	57.76%	80.5	96.9	75.26%
HNC-CAD	1.63	4.25	62.03%	89.2	91.8	80.86%
OURS	**1.32**	**3.19**	**65.31%**	**99.6**	**99.4**	**85.38%**

- **Effectiveness of Mamba+ Module**

Method	ACC_c ↑	ACC_p ↑	MCD ↓	IR ↓	SR ↑
Transformer	77.29	63.62	64.30	67.46%	34.28%
Mamba	99.98	99.90	0.76	10.35%	92.01%
Mamba+	**99.99**	**99.93**	**0.75**	**8.50%**	**93.93%**

- **Impact of Multi-Scale Transformer (MST)**

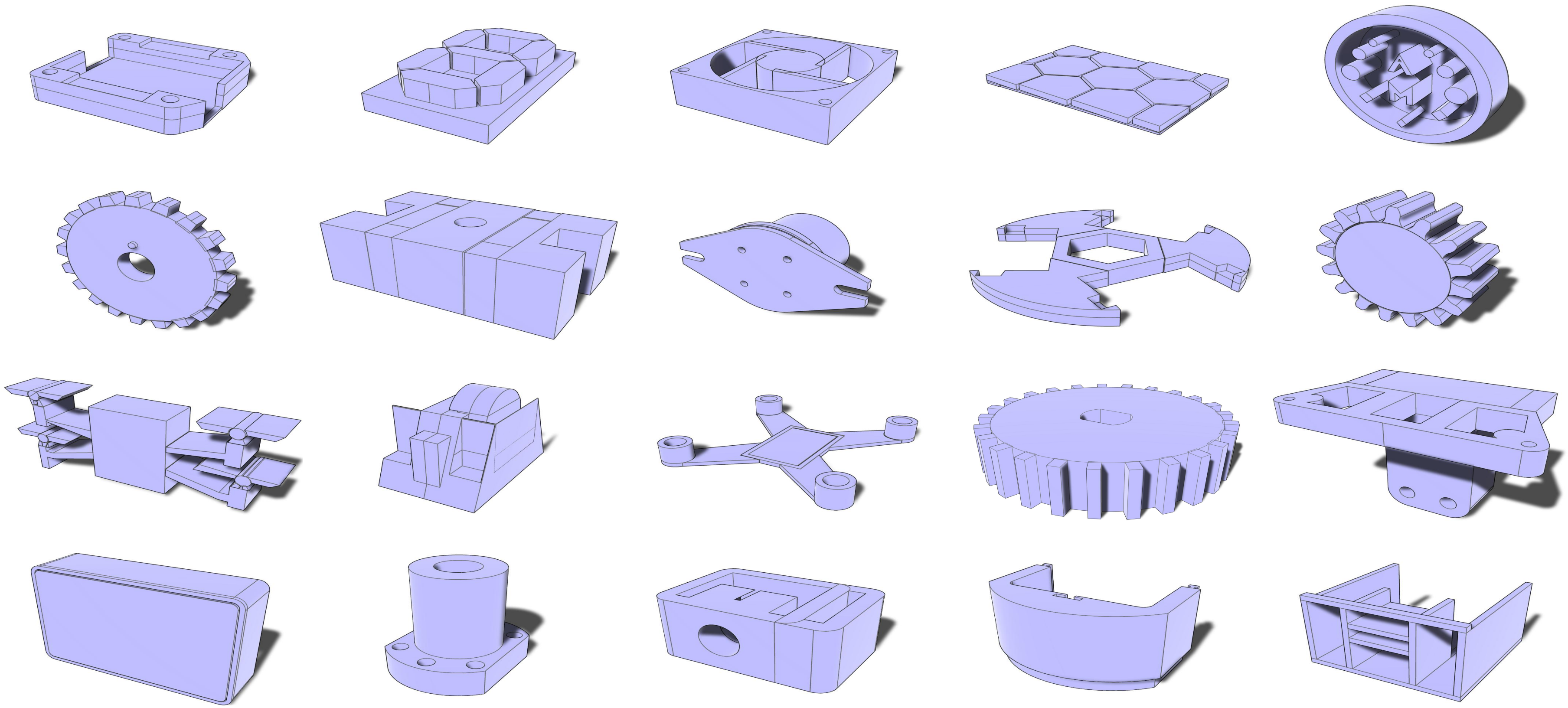
Method	MMD ↓	JSD ↓	COV ↑	Unique ↑	Novel ↑	SR ↑
w/o MST	1.47	4.92	61.69%	**99.5**	**99.6**	77.05%
w/ MST	**1.32**	**3.19**	**65.31%**	99.6	99.4	**85.38%**

- **Conclusion & Outlook**

1. **MamTiff-CAD** effectively supports the generation of long-sequence and complex CAD models.
2. Introduced the **ABC-256 dataset** as a new benchmark.
3. **Future work:** extend to more CAD command types and integrate B-rep representations.

Result Show

This is a visual display of our partially generated results



Thank you