ConTiCoM-3D: A Continuous-Time Consistency Model for 3D Point Cloud Generation

Sebastian Eilermann René Heesch Oliver Niggemann
Institute for Artificial Intelligence
Helmut Schmidt University
Hamburg, Germany
{sebastian.eilermann,rene.heesch,oliver.niggemann}@hsu.hamburg

Abstract

*Fast and accurate 3D shape generation from point clouds is essential for real-world applications such as robotics, AR/VR, and digital content creation. We present **ConTiCoM-3D**, a continuous-time consistency model that generates 3D shapes directly in point space, without relying on discretized diffusion steps, pre-trained teacher models, or latent-space encodings. Our approach combines a TrigFlow-inspired continuous noise schedule with a Chamfer Distance-based geometric loss, providing stable training in high-dimensional point sets while avoiding costly Jacobian-vector products. This enables efficient one- to two-step inference with high geometric fidelity. Unlike previous methods that require iterative denoising or latent decoders, ConTiCoM-3D operates entirely in continuous time with a time-conditioned neural network, achieving fast generation. Extensive experiments on the ShapeNet benchmark demonstrate that our method matches or surpasses leading diffusion and latent consistency models in both quality and efficiency, establishing ConTiCoM-3D as a practical solution for scalable 3D shape generation.*

1. Introduction

Fast and accurate 3D shape generation from point clouds is essential for downstream tasks in robotics [47], autonomous driving [49], medicine [11, 31], and Augmented Reality (AR)/Virtual Reality (VR) [19], where real-time inference is often required under hardware constraints. In such settings, generative models must balance geometric fidelity with stringent runtime and memory limitations. This rules out approaches that depend on hundreds of diffusion steps, expensive teacher–student distillation, or latent space compression, all of which are poorly suited for geometry-sensitive point cloud domains.

To address these challenges, a wide variety of 3D generative paradigms have been proposed, including variational autoencoders (VAEs) [14, 19, 33], generative adversarial networks (GANs) [1, 40], flow-based models [3, 48], and increasingly popular diffusion models [11, 13, 29, 30, 44, 53]. VAEs and GANs offer fast inference, but struggle with mode collapse and limited diversity. Flow-based models achieve diversity and tractability, but require iterative sampling, which limits scalability [10, 18, 21, 47, 48]. Diffusion models produce high-quality samples but are slowed by multi-step denoising [29, 30, 53], while latent diffusion accelerates sampling at the cost of structural artifacts from encoder-decoder bottlenecks [8, 49]. Point space approaches such as Point Straight Flow (PSF) [47] reduce inference time but rely on hand-crafted heuristics that hinder generalization [10].
Finally, consistency models [45] promise one- to two-step generation, but existing training strategies do not scale to 3D: distillation-based CMs inherit the weaknesses of their diffusion teachers and fail without reliable perceptual metrics [28, 45], Jacobian–vector product (JVP)-based supervision is costly and unstable for high-dimensional point clouds [2, 42], and latent space CMs [7, 8] compress geometry through autoencoding, introducing artifacts.

In this work, we introduce **ConTiCoM-3D**, a **Con**tinuous-**Ti**me **Co**nsistency **M**odel for **3D** point cloud generation that operates directly in raw point space. Unlike prior diffusion- or latent-based CMs, ConTiCoM-3D avoids reliance on autoencoding, hand-crafted flows, teacher–student supervision, and unstable JVP-based losses. Instead, it leverages a Chamfer Distance–based geometric loss with a deterministic time-dependent weighting schedule, combined with a TrigFlow-inspired continuous noise schedule, to achieve stable training and efficient one- to two-step inference. The source code for the algorithm and for reproducing the evaluation results is available at https://github.com/SEilermann/ConTiCoM_3D.

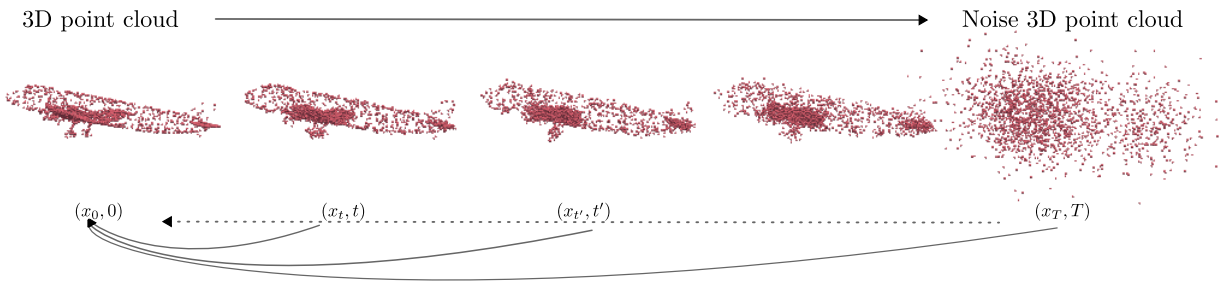

Figure 1. ConTiCoM-3D models a TrigFlow-inspired forward flow ODE (dashed line) that perturbs a clean point cloud x_0 into noisy samples x_t, $x_{t'}$, and x_T. A time-conditioned model $f_\theta\,(x_t, t)$ learns to reverse this flow (dotted arrows), enabling single-step reconstruction and enforcing consistency along the continuous forward process without iterative denoising.

Our main contributions are:

- **ConTiCoM-3D**, the first continuous-time consistency model for 3D point cloud generation that operates directly in raw point space, without relying on latent encodings, hand-crafted flows or teacher models.
- A **geometry-aware loss** that combines Chamfer Distance with a deterministic time-dependent weighting schedule, providing geometric consistency without costly or unstable JVPs.
- Empirical evidence on ShapeNet, showing that ConTiCoM-3D achieves competitive one- to two-step sampling quality, surpassing diffusion and latent Consistency Models while enabling real-time inference.

2. Background

Our continuous-time 3D point cloud generation approach is based on diffusion models, Flow-Matching (FM) techniques, and Consistency Models (CMs).

2.1. Diffusion Models and Their Variants

Recent advances in generative modeling have unified denoising diffusion probabilistic models (DDPMs) [13] and score-based methods [43, 44] within a continuous-time framework, where data is generated by solving an ODE that transforms noise into structure. Although DDPMs remain widely used, alternatives such as FM [24] and TrigFlow [27] simplify training and sampling by directly learning velocity fields. These approaches improve stability, better align signal-to-noise ratios (SNRs), and are well-suited for consistency-based training.

Diffusion models [44] reverse a standard forward noising process that perturbs a clean sample x_0 into $x_t = \alpha_t x_0 + \sigma_t z$, where $z \sim \mathcal{N}(0, I)$ is Gaussian noise. DDPMs [13] train a noise predictor ϵ_θ and generate samples by solving the probability flow ODE (PF-ODE):

$$\dot{x}_t = \frac{d\log\alpha_t}{dt} x_t + \left(\frac{d\sigma_t}{dt} - \frac{d\log\alpha_t}{dt}\sigma_t\right)\epsilon_\theta(x_t, t). \quad (1)$$

FM [24] uses a linear schedule $\alpha_t = 1 - t$, $\sigma_t = t$, and trains a velocity field v_θ to match the reverse time velocity $v_t = \sigma_t' z - \alpha_t' x_0$, where $\alpha_t' = \frac{d\alpha_t}{dt}$ and $\sigma_t' = \frac{d\sigma_t}{dt}$. The objective is minimizing the following:

$$\mathbb{E}_{x_0,z,t}\left[w(t)\left\|v_\theta(x_t, t) - (\sigma_t' z - \alpha_t' x_0)\right\|^2\right], \quad (2)$$

where $w(t)$ is a scalar time-dependent weighting function. Sampling then follows the simplified ODE $\frac{dx_t}{dt} = v_\theta(x_t, t)$.

TrigFlow [27] adopts a spherical interpolation schedule with $\alpha_t = \cos(t)$ and $\sigma_t = \sin(t)$ for $t \in [0, \frac{\pi}{2}]$, yielding $x_t = \cos(t)x_0 + \sin(t)z$. It trains a time-conditioned velocity predictor F_θ, scaled by a fixed noise level $\sigma_d > 0$, using:

$$\mathbb{E}_{x_0,z,t}\left[w(t)\left\|\sigma_d F_\theta\left(\frac{x_t}{\sigma_d}, t\right) - (\cos(t)z - \sin(t)x_0)\right\|^2\right], \quad (3)$$

Samples are generated via:

$$\frac{dx_t}{dt} = \sigma_d F_\theta\left(\frac{x_t}{\sigma_d}, t\right). \quad (4)$$

This formulation ensures consistent SNRs, simplifies time conditioning, and improves numerical stability, making it well suited for training consistency models. We adopt TrigFlow as the forward process for ConTiCoM-3D (see Sec. 4).

2.2. Continuous-Time Consistency Models

CMs [45] directly learn a mapping $f_\theta(x_t, t)$ that predicts the clean sample x_0 from noisy input x_t at time t. A common parameterization is as follows:

$$f_\theta(x_t, t) = c_{\text{skip}}(t)x_t + c_{\text{out}}(t)F_\theta(x_t, t), \quad (5)$$

where $c_{\text{skip}}(t)$ and $c_{\text{out}}(t)$ are scalar functions that ensure $f_\theta(x_0, 0) = x_0$, and F_θ is a neural network.

Discrete-time vs. continuous-time. CMs can be trained in discrete or continuous time. Discrete-time models minimize discrepancies between adjacent steps with a stop-gradient teacher:

$$\mathbb{E}_{x_t,t}\left[d\big(f_\theta(x_t,t), f_\theta^-(x_{t-\Delta t}, t-\Delta t)\big)\right], \tag{6}$$

where f_θ^- is a stop-gradient copy and $d(\cdot,\cdot)$ is a distance metric such as squared error. Although effective, this approach typically requires a pre-trained diffusion teacher, increasing cost and constraining performance.

Continuous-time CMs [27] avoid discretization by taking $\Delta t \to 0$, leading to:

$$\mathcal{L}_{\text{cont.CM}} = \mathbb{E}_{x_t,t}\left[w(t)\left\langle f_\theta(x_t,t), \tfrac{d}{dt} f_\theta(x_t,t)\right\rangle\right]. \tag{7}$$

Although elegant in theory, this requires JVPs to compute $\frac{d}{dt} f_\theta$ [2, 42], which scale poorly in high-dimensional 3D point clouds.

Practical workarounds. Alternative strategies such as distillation [28, 45] introduce the dependence on pre-trained diffusion teachers, which both increase training cost and limit performance at the teacher's quality level. Latent consistency models [7, 8] sidestep these issues through compression, but suffer from encoder-decoder bottleneck artifacts. As a result, directly applying existing CM objectives to point clouds leads to instability and degraded geometry.

TrigFlow parameterization. Under the TrigFlow path, the continuous-time CM has a simple closed form:

$$f_\theta(x_t,t) = \cos(t)x_t - \sin(t)\,\sigma_d F_\theta\left(\tfrac{x_t}{\sigma_d}, t\right), \tag{8}$$

where $\sigma_d > 0$ is the data scale. This form avoids ODE solvers and supports efficient training. However, existing objectives still rely on JVPs or distillation, leaving open the question of how to train CMs for 3D efficiently and robustly.

Motivation for ConTiCoM-3D. These limitations motivate ConTiCoM-3D: a teacher-free, JVP-free, latent-free continuous-time CM tailored to 3D point clouds. Unlike DDPMs and FM models that require many inference steps, ConTiCoM-3D enables fast single-step sampling. In contrast to JVP-based continuous-time consistency models [2], we avoid unstable Jacobian–vector products, which are prohibitively expensive for unordered 3D point clouds. Our approach also avoids teacher-student distillation [45], which relies on strong perceptual metrics that are not available in point space, and sidesteps latent compression [8], which introduces shape artifacts. Together, these design choices allow our model to operate robustly and efficiently in the raw point cloud domain without teachers, JVPs, or latent bottlenecks, while preserving geometric fidelity through a lightweight Chamfer reconstruction term (see Sec. 4).

3. Related Work

Adversarial and Autoencoding Approaches. Early 3D generative models adopted adversarial frameworks such as r-GAN [1], TreeGAN [40], SP-GAN [22], and PDGN [16]. While pioneering, these methods suffer from training instability, mode collapse, and under-constrained geometry. PDGN alleviates some of these issues by progressively generating multi-resolution point clouds with a shape-preserving adversarial loss, but it remains constrained by adversarial optimization. Variational autoencoders (VAEs) [14, 19, 33] and graph-based extensions such as GCA [50] improved stability but compress geometry through latent bottlenecks, limiting the fidelity for fine-grained shape reconstruction.

Flow-Based Models. Flow-based methods such as PointFlow [48] and ShapeGF [3] model continuous distributions over point sets. They offer exact likelihoods and diverse sampling, but require expensive ODE integration at inference. Later refinements like SoftFlow [18], DPF-Net [21], and PSF [47] reduce inference time, but often rely on handcrafted priors or dynamics that limit generalization.

Diffusion Models. Diffusion models [13, 44] and their 3D extensions [11, 29, 30, 53] achieve state-of-the-art fidelity by progressively denoising from Gaussian noise. However, they require hundreds of steps, preventing real-time deployment. Latent diffusion variants, such as LION [49], MLPCM [8], and MeshDiffusion [26], accelerate inference but introduce discretization artifacts from encoder-decoder compression.

Consistency Models. CMs [45] promise one- to two-step generation by enforcing forward process alignment in continuous flows. Distillation-based CMs [28, 45] inherit the limitations of their diffusion teachers and are weak in 3D domains without reliable perceptual metrics. Continuous-time CMs [2, 27, 42] replace teachers with JVP supervision, which is costly and unstable for high-dimensional point sets. Latent space CMs [7, 8] reduce training cost but compress geometry, introducing bottleneck artifacts. In addition, in [8] it is introduced a latent space CM based approach with teacher distillation MLPCM(TM) and without teacher distillation MLPCM(LCM). Recent extensions include multistep CMs [12], inverse-flow consistency [52], and one-step acceleration methods such as SANA [5] or alignment-based strategies [38]. While effective in images or robotics [28], these designs either trade efficiency for fidelity or depend on surrogate objectives. In contrast, ConTiCoM-3D is the first CM to operate *directly in raw point space* without teachers, JVPs, or latent bottlenecks,

Algorithm 1 ConTiCoM-3D: JVP-free training with flow-matching and Chamfer reconstruction

Require: Dataset $\mathcal{X}$; velocity net F_θ; symmetric squared Chamfer CD; data scale σ_d; $T_{\max} = \pi/2$; $\lambda_{\min}, \lambda_{\max}$; $T_{\max} = \pi/2$

1: **for** each epoch **do**
2: **for** each batch $x_0 \sim \mathcal{X}$ **do**
3: Sample $t \sim \mathcal{U}(0, T_{\max})$, $z \sim \mathcal{N}(0, \sigma_d^2 I)$
4: Construct noisy sample $x_t = \cos(t)\, x_0 + \sin(t)\, z$ (Eq. 9)
5: Compute velocity $v_\theta = F_\theta(x_t/\sigma_d, t)$
6: Predictor $f_\theta(x_t, t) = \cos(t)\, x_t - \sin(t)\, \sigma_d\, v_\theta$ (Eq. 10)
7: **FM loss:** $\mathcal{L}_{\text{FM}} = \| \sigma_d\, v_\theta - (\cos(t)\, z - \sin(t)\, x_0) \|_2^2$ (Eq. 12)
8: **Chamfer loss:** $\mathcal{L}_{\text{CD}} = \text{CD}(f_\theta(x_t, t), x_0)$ (Eq. 13)
9: **Weight:** $\lambda_{\text{CD}}(t) = \lambda_{\min} + (\lambda_{\max} - \lambda_{\min}) \cos^2(t)$
10: **Total:** $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{FM}} + \lambda_{\text{CD}}(t)\, \mathcal{L}_{\text{CD}}$ (Eq. 14)
11: Update θ by SGD/Adam
12: **end for**
13: **end for**

using only analytic flow supervision and a Chamfer-based reconstruction loss.

Summary. Overall, the landscape of 3D generative modeling reveals a persistent trade-off. Adversarial and autoencoding methods [1, 14, 19, 22, 33, 40] are efficient but limited in fidelity and diversity, while flow- and diffusion-based models [3, 10, 11, 13, 18, 21, 29, 30, 44, 48, 53] provide strong geometric quality but require many iterative steps. Latent and multistep consistency models [5, 7, 8, 12, 49] mitigate cost, but either introduce artifacts or add algorithmic complexity. Moreover, existing CM variants do not transfer cleanly to unordered 3D point clouds. Distillation-based CMs [28, 45] rely on perceptual similarity metrics that are effective in image space but unavailable in point space, leading to weak supervision signals. Continuous-time CMs with JVP supervision [2, 42] are computationally prohibitive and often unstable when applied to high-dimensional shapes. Latent space CMs [7, 8] mitigate these costs but compress geometry through encoder-decoder bottlenecks, introducing structural distortions. These limitations motivate the design of **ConTiCoM-3D**, a continuous-time CM that is teacher-free, JVP-free, and latent-free, combining analytic FM with Chamfer-based supervision to enable robust one- to two-step point cloud generation directly in raw point space.

4. Method

We propose **ConTiCoM-3D**, the first continuous-time consistency model that operates directly in raw point space and enables single-step point cloud generation. Unlike previous approaches, ConTiCoM-3D is *teacher-free, JVP-free, and latent-free*: it avoids diffusion teachers, costly Jacobian supervision, and lossy latent compression. Our method relies on two complementary objectives: an analytic FM regression target and a lightweight Chamfer reconstruction term. Together, these provide geometry-aware, permutation-invariant supervision that is both stable and efficient, enabling real-time inference with one to four sampling steps.

The complete training loop is summarized in Algorithm 1, where each step corresponds to its formal definition in the text.

4.1. Problem Definition

Let $\mathcal{X} = \{X_i\}_{i=1}^N$ be a dataset of different 3D point clouds, with each $X_i \in \mathbb{R}^{M\times 3}$ containing unordered points. Following TrigFlow [27] (see Sec. 2.1), the forward process is as follows:

$$x_t = \cos(t)\, x_0 + \sin(t)\, z, \qquad z \sim \mathcal{N}(0, \sigma_d^2 I), \tag{9}$$

with $t \in [0, T_{\max}]$ and $T_{\max} = \frac{\pi}{2}$ (see Alg. 1, l. 3-4).

The predictor admits a closed-form expression derived by substituting the analytic TrigFlow velocity into the CT-CM parameterization of Eq. 8:

$$f_\theta(x_t, t) = \cos(t)\, x_t - \sin(t)\, \sigma_d\, F_\theta(x_t/\sigma_d, t). \tag{10}$$

This form enforces $f_\theta(x, 0) = x$ and realizes the continuous-time consistency formulation described in Sec. 2.2 (see Alg. 1, l. 6). Here, CD denotes the symmetric squared Chamfer distance, normalized by M points.

4.2. Model Training

Motivation. Temporal derivative supervision in continuous-time CMs usually requires JVPs [2], which are unstable and computationally heavy in high-dimensional geometry. Distillation-based supervision [17, 45] depends on diffusion teachers and fails in 3D due to the absence of

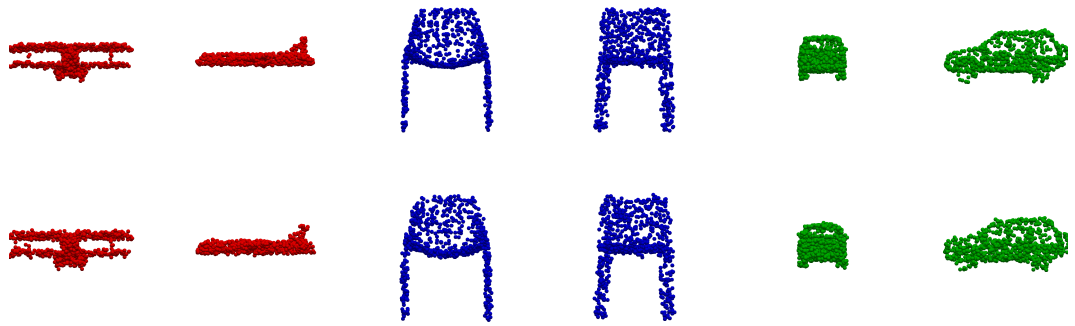

Figure 2. Single-class generation results (*airplane, chair, car*). Top row: samples generated by ConTiCoM-3D with $S = 1$ step. Bottom row: samples from MLPCM [8]. ConTiCoM-3D produces sharper structures and more consistent global geometry while maintaining diversity, whereas MLPCM exhibits compression artifacts from latent bottlenecks.

robust perceptual metrics. We therefore adopt a *single-time, JVP-free* supervision that combines an analytic flow target with a lightweight Chamfer term. This design is efficient, unbiased, and stable for unordered point clouds.

Analytic flow matching. The TrigFlow forward process admits an analytic velocity

$$\frac{dx_t}{dt} = -\sin(t)\, x_0 + \cos(t)\, z, \tag{11}$$

yielding the regression objective

$$\mathcal{L}_{\text{FM}} = \left\| \sigma_d\, F_\theta(x_t/\sigma_d, t) - (\cos(t)\, z - \sin(t)\, x_0) \right\|_2^2, \tag{12}$$

corresponding to line 7 in Alg. 1 and related to Flow Matching [24]. Unlike secant-based consistency losses [6], this is unbiased, variance-reduced, and avoids higher-order derivatives.

Proposition 1 (Closed-form recovery)**.** *If F_θ matches the analytic velocity, then $f_\theta(x_t, t) = x_0$ for all $t \in [0, T_{\max}]$.*

This implies that perfect velocity regression suffices for exact reconstruction, without the need for iterative refinement (see proof in Appendix 8).

Chamfer reconstruction. To encourage geometric fidelity, we add a permutation-invariant Chamfer loss [9]:

$$\mathcal{L}_{\text{CD}} = \text{CD}(f_\theta(x_t, t), x_0). \tag{13}$$

This corresponds to line 8 in Alg. 1.

Total objective. The final training loss is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{FM}} + \lambda_{\text{CD}}(t)\, \mathcal{L}_{\text{CD}}. \tag{14}$$

Time-dependent Chamfer weighting. We use a deterministic time-dependent weight that emphasizes geometric refinement near the data manifold:

$$\lambda_{\text{CD}}(t) = \lambda_{\min} + (\lambda_{\max} - \lambda_{\min}) \cos^2(t) \tag{15}$$

with $\lambda_{\min} = 0.1$ and $\lambda_{\max} = 0.3$. This assigns higher Chamfer weight near $t = 0$ and lower weight near $t = T_{\max}$.

Eq. 14 corresponds to line 9 in Alg. 1.

Unlike previous CMs [8, 45], we train without an EMA teacher.

4.3. Sampling and Inference

Single-step generation. Sampling requires only one evaluation at $T = T_{\max}$:

$$\hat{x}_0 = f_\theta(x_T, T), \qquad x_T \sim \mathcal{N}(0, \sigma_d^2 I). \tag{16}$$

This enables practical single-step point cloud generation, unlike diffusion models that require hundreds of denoising steps.

Few-step refinement. For optional higher fidelity, we define a total number of inference steps $S \in \mathbb{N}$ and partition $[0, T_{\max}]$ into S uniform time intervals of size $\Delta = T_{\max}/S$. We integrate the TrigFlow PF-ODE in reverse time using explicit Euler steps:

$$x_{t-\Delta} = x_t - \Delta\, \sigma_d\, F_\theta(x_t/\sigma_d, t), \tag{17}$$

with $t = T_{\max}, T_{\max} - \Delta, \ldots, \Delta$. A Heun proposal [27] may optionally be used for local error control, while accepted states follow the Euler update. Empirically, small values of S already suffice for high-fidelity generation with minimal cost [5, 27, 45].

5. Experiments

We evaluate ConTiCoM-3D on standard 3D point cloud generation benchmarks in both single-class and multi-class

Table 1. Single-class generation on ShapeNet (*airplane, chair, car*) following PointFlow [48]. We report **1-NNA**↓ computed using Chamfer Distance (CD) and Earth Mover's Distance (EMD); lower is better. Training and test data are normalized globally to $[-1, 1]$.

Method	Airplane		Chair		Car	
	CD	EMD	CD	EMD	CD	EMD
r-GAN [1]	98.40	96.79	83.69	99.70	94.46	99.01
1-GAN(CD) [1]	87.30	93.95	68.58	83.84	66.49	88.78
1-GAN(EMD) [1]	89.49	76.91	71.90	64.65	71.16	66.19
PointFlow [48]	75.68	70.74	62.84	60.57	58.10	56.25
DPF-Net [21]	75.18	65.55	62.00	58.53	62.35	54.48
SoftFlow [18]	76.05	65.80	59.21	60.05	64.77	60.09
SetVAE [19]	76.54	67.65	58.84	60.57	59.94	59.94
DPM [29]	76.42	86.91	60.05	74.77	68.89	79.97
PVD [53]	73.82	64.81	56.26	53.32	54.55	53.83
LION [49]	67.41	61.23	53.70	52.34	53.32	54.55
DiT-3D [30]	69.42	65.08	55.59	54.91	53.87	53.02
MeshDiffusion [26]	66.44	76.26	53.69	57.63	81.43	87.84
MLPCM(TM) [8]	65.12	58.70	51.48	50.06	51.17	48.92
MLPCM(LCM) [8]	67.22	60.32	53.63	51.74	53.82	52.75
NSOT [15]	68.64	61.85	55.51	57.63	59.66	53.55
3DQD [23]	**56.29**	**54.78**	55.61	52.94	55.75	52.80
PVD-DDIM [41]	76.21	69.84	61.54	57.73	60.95	59.35
PSF [47]	71.11	61.09	58.92	54.45	57.19	56.07
ShapeGF [3]	80.00	76.17	68.96	65.48	63.20	56.53
ConTiCoM-3D (S=1)	64.89	61.77	54.30	50.52	53.30	51.40
ConTiCoM-3D (S=2)	70.20	66.56	**50.90**	**49.24**	**50.83**	**48.77**

settings. The experimental design focuses on three aspects: The fidelity and diversity of generated point clouds, the efficiency of inference in terms of sampling speed, and the robustness of the method, as demonstrated through ablation studies.

5.1. Experimental Setup

Following prior work [8, 48, 49], we evaluate the generative performance on ShapeNet [4] for single-class generation (*airplane, chair, car*) and on ShapeNet-vol [32] for multi-class generation across 13 categories. All shapes are uniformly resampled to 2048 points. The evaluation is based on **1-Nearest Neighbor Accuracy (1-NNA)** computed under Chamfer Distance (CD) and Earth Mover's Distance (EMD) [48]. Lower scores indicate a better balance between fidelity and diversity. We report results under both global normalization (as in Table 1) and per-shape normalization (as in Table 4), following common practice in prior work.

Our model builds on a Point-Voxel CNN (PVCNN) [25] U-Net backbone, augmented with PointNet++ [35] set abstraction and feature propagation modules. Time conditioning is introduced at the bottleneck via sinusoidal embeddings. Training is performed for 4 days on two NVIDIA L40S GPUs using the Adam optimizer [20] with a fixed learning rate of 1×10^{-4}. More architectural and training details are provided in the Appendix 7.

Table 2. Inference efficiency comparison across methods. We report inference steps and runtime (seconds).

Model / Method	Steps	Time (s)
LION (DDPM) [49]	1000	27.09
LION (DDIM) [49]	1000	27.09
LION (DDIM) [49]	100	3.07
LION (DDIM) [49]	10	0.47
DPM [29]	1000+	22.80
PVD [53]	1000	29.90
PVD-DDIM [41]	100	3.15
1-GAN [1]	1	**0.03**
SetVAE [19]	1	**0.03**
PSF [47]	1	0.04
MLPCM [8]	1	<0.5
ConTiCoM-3D (ours)	1	0.22
	2	0.37
	4	0.42

5.2. Single-Class 3D Point Cloud Generation

We first consider the setting where the models are trained separately for each category (*airplane, chair, car*), following the protocol of PointFlow [48]. A visualization of results is given in Figure 2.

Quantitative Results. Table 1 and Table 4 show that ConTiCoM-3D consistently outperforms GAN- and VAE-based baselines and is competitive with recent diffusion and consistency models. With only two inference steps, our method achieves the lowest EMD scores and competitive CD performance across all three categories, while only requiring two evaluations at inference. Compared to latent approaches such as LION [49] and MLPCM [8], ConTiCoM-3D operates directly in the raw point space, avoiding encoder–decoder artifacts and enabling temporally stable and spatially precise generation. Moreover, its runtime is significantly faster than multistep diffusion models, offering a strong balance between speed and fidelity.

Sampling time: We report wall-clock sampling time per shape (seconds) with batch size 1. Note that runtimes can vary across implementations and hardware. We follow the evaluation setting used for our experiments.

Results. As shown in Table 2, ConTiCoM-3D achieves near real-time sampling (0.22-0.42 s for 1-4 steps), providing a large speedup over diffusion models that require hundreds to thousands of denoising steps (e.g., PVD and DPM). While distilled one-step methods such as PSF and MLPCM can be faster, they often trade off geometric fidelity, as reflected by their CD/EMD results in Sec. 5 [9, 48]. Overall,

Figure 3. Noise interpolation results with ConTiCoM-3D.

ConTiCoM-3D offers a practical balance between sampling speed and shape quality in the 1–4 step regime.

Interpolation. We probe representation continuity by interpolating between noise vectors $z_1, z_2 \sim \mathcal{N}(0, I)$. Intermediate shapes are generated with $z_\alpha = (1-\alpha)z_1 + \alpha z_2$ and $\hat{x}_0 = f_\theta(x_T, T_{\max})$. Figure 3 shows smooth and structurally consistent transitions, confirming that ConTiCoM-3D learns a continuous geometry-aware flow.

5.3. Ablation Studies

We perform extensive ablations to evaluate the contribution of each design choice in ConTiCoM-3D. The main findings are summarized here, while the complete numerical results are provided in Appendix 10, where all tables are reported.

Loss components. We disentangle the effect of the analytic FM loss and the Chamfer reconstruction loss. As shown in Appendix 10, FM-only training produces diverse but geometrically imprecise shapes, while Chamfer-only training improves fidelity at the cost of mode collapse. The combination of FM and Chamfer with adaptive weighting achieves the best balance across 1-NNA, MMD, COV, and JSD metrics.

Noise schedule. We compare TrigFlow with linear and cosine schedules. The results (Appendix 10) confirm that TrigFlow provides superior stability and fidelity for few-step generation, validating our choice of forward process.

Sampling steps. We examine the inference quality for $S = 1, 2, 4$ with both Euler and Heun solvers. As detailed in Appendix 10, performance improves significantly from one to two steps, with diminishing returns beyond. Heun provides slight gains for single-step sampling at a modest computational cost.

5.4. Multi-Class 3D Shape Generation

Next, we evaluate unconditional multi-class generation on ShapeNet-vol, where a single model is trained jointly across 13 categories. This setting is more challenging due to the multimodal nature of the shape distribution.

6. Discussion and Conclusion

We presented **ConTiCoM-3D**, the first continuous-time consistency model for 3D point cloud generation that operates directly in raw point space. Unlike prior approaches that rely on diffusion teachers, Jacobian-vector products, or latent compression, ConTiCoM-3D combines an analytic flow-matching objective with a lightweight Chamfer reconstruction loss, yielding stable training and efficient one- to two-step inference. This design preserves geometric fidelity while avoiding the bottlenecks and instabilities of previous paradigms.

Experiments on ShapeNet benchmarks demonstrate that ConTiCoM-3D achieves state-of-the-art performance among teacher-free methods. In single-class settings, our model outperforms GAN- and VAE-based baselines and surpasses recent diffusion and latent consistency models in both Chamfer and Earth Mover's metrics, while requiring only two evaluations at inference. In the more challenging multi-class generation task, ConTiCoM-3D generalizes effectively across 13 categories, achieving the lowest 1-NNA scores under both CD and EMD. Qualitative results further show smooth interpolations and coherent shape transitions, confirming that our geometry-aware training encourages continuous and robust generative flows.

The ablation studies highlight three key findings. First, Chamfer supervision is essential for improving spatial fidelity without sacrificing sample diversity. Second, the TrigFlow schedule is critical for stable few-step generation, outperforming linear and cosine baselines. Third, most of the quality improvements saturate at two sampling steps, confirming the efficiency of our design. Together, these results validate that our contributions are both necessary and sufficient for efficient high-quality 3D generation.

Despite strong efficiency and competitive fidelity, our study has several limitations. We primarily evaluate on ShapeNet-vol with fixed 2048-point clouds, and do not test real scans or settings with partial/noisy and uneven-density point clouds. Our evaluation focuses on geometric distances (CD/EMD and related set metrics) and does not include semantic or task-driven evaluation. Some categories with thin structures exhibit a larger quality gap in very few-step sampling. In addition, we do not report full multi-seed confidence intervals for every table due to compute constraints. Future work will extend ConTiCoM-3D to scan-like data distributions, larger and more diverse datasets, higher point resolutions, and more comprehensive evaluation protocols.

Finally, ConTiCoM-3D offers a scalable and practical solution for fast 3D point cloud synthesis, making it well suited for interactive robotics, AR/VR, and medical applications where both speed and accuracy are critical. Future work may extend our framework toward conditional

Table 3. Generation metrics (1-NNA↓) on 13 classes of ShapeNet-vol.

Method	CD	EMD
Tree-GAN [40]	96.80	96.60
PointFlow [48]	63.25	66.05
ShapeGF [3]	55.65	59.00
SetVAE [19]	79.25	95.25
PDGN [16]	71.05	86.00
DPF-Net [21]	67.10	64.75
DPM [29]	62.30	86.50
PVD [53]	58.65	57.85
LION [49]	51.85	48.95
MLPCM(TM) [8]	50.17	47.84
MLPCM(LCM) [8]	53.85	52.45
ConTiCoM-3D (S=1)	49.30	46.50
ConTiCoM-3D (S=2)	**48.90**	**45.21**

Table 4. Single-class generation results on ShapeNet dataset from PointFlow [48]. We use 1-NNA↓ as generation metric for evaluation. Training and test data normalized individually into $[-1, 1]$.

Method	Airplane		Chair		Car	
	CD	EMD	CD	EMD	CD	EMD
LION [49]	76.30	67.04	56.50	53.85	59.52	49.29
Tree-GAN [40]	97.53	99.88	88.37	96.37	89.77	94.89
SP-GAN [22]	94.69	93.95	72.58	83.69	87.36	85.94
PDGN [16]	94.94	91.73	71.83	79.00	89.35	87.22
GCA [50]	88.15	85.93	64.27	64.50	70.45	64.20
ShapeGF [3]	81.23	80.86	58.01	61.25	61.79	57.24
MLPCM(TM) [8]	**73.28**	**63.08**	56.20	53.16	58.31	**47.74**
MLPCM(LCM) [8]	75.56	66.85	58.58	55.32	61.28	49.91
ConTiCoM-3D (S=1)	74.98	70.01	57.68	56.78	59.77	53.05
ConTiCoM-3D (S=2)	74.22	68.54	**56.11**	**53.09**	**55.43**	54.22

and guided generation, scene-level modeling, and handling high-resolution or partial scans. Another promising direction is to explore hybrid training objectives that integrate consistency with diffusion-style guidance, potentially improving controllability and robustness in more complex 3D settings.

Acknowledgments

This research as part of the projects LaiLa and EKI, which are funded by dtec.bw -Digitalization and Technology Research Center of the Bundeswehr which we gratefully acknowledge. dtec.bw is funded by the European Union - NextGenerationEU.

References

[1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, pages 40–49. PMLR, 2018. 1, 3, 4, 6

[2] Nicholas M Boffi, Michael S Albergo, and Eric Vanden-Eijnden. How to build a consistency model: Learning flow maps via self-distillation. *arXiv preprint arXiv:2505.18825*, 2025. 1, 3, 4

[3] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely, and Bharath Hariharan. Learning gradient fields for shape generation. In *Computer Vision–ECCV 2020*, pages 364–381. Springer, 2020. 1, 3, 4, 6, 8

[4] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 6, 3, 4

[5] Junsong Chen, Shuchen Xue, Yuyang Zhao, Jincheng Yu, Sayak Paul, Junyu Chen, Han Cai, Song Han, and Enze Xie. Sana-sprint: One-step diffusion with continuous-time consistency distillation. *arXiv preprint arXiv:2503.09641*, 2025. 3, 4, 5

[6] Yiding Chen, Yiyi Zhang, Owen Oertell, and Wen Sun. Convergence of consistency model with multistep sampling under general data assumptions. *arXiv preprint arXiv:2505.03194*, 2025. 5

[7] Quan Dao, Khanh Doan, Di Liu, Trung Le, and Dimitris Metaxas. Improved training technique for latent consistency models. *arXiv preprint arXiv:2502.01441*, 2025. 1, 3, 4

[8] Bi'an Du, Wei Hu, and Renjie Liao. Multi-scale latent point consistency models for 3d shape generation. In *arXiv preprint arXiv:2412.19413*, 2024. 1, 3, 4, 5, 6, 8

[9] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017. 5, 6, 1, 3

[10] Leonhard Faubel, Thomas Woudsma, Leila Methnani, Amir Ghorbani Ghezeljhemeidan, Fabian Buelow, Klaus Schmid, Willem D van Driel, Benjamin Kloepper, Andreas Theodorou, Mohsen Nosratinia, et al. Towards an mlops architecture for xai in industrial applications. *arXiv preprint arXiv:2309.12756*, 2023. 1, 4

[11] Paul Friedrich, Julia Wolleb, Florentin Bieder, Florian M Thieringer, and Philippe C Cattin. Point cloud diffusion models for automatic implant generation. In *International conference on medical image computing and computer-assisted intervention*, pages 112–122. Springer, 2023. 1, 3, 4

[12] Jonathan Heek, Emiel Hoogeboom, and Tim Salimans. Multistep consistency models. *arXiv preprint arXiv:2403.06807*, 2024. 3, 4

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1, 2, 3, 4

[14] Michael Hohmann, Sebastian Eilermann, Willi Großmann, and Oliver Niggemann. Design automation: A condi-

tional vae approach to 3d object generation under conditions. In *2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–8. IEEE, 2024. 1, 3, 4

[15] Ka-Hei Hui, Chao Liu, Xiaohui Zeng, Chi-Wing Fu, and Arash Vahdat. Not-so-optimal transport flows for 3d point cloud generation. *arXiv preprint arXiv:2502.12456*, 2025. 6

[16] Le Hui, Rui Xu, Jin Xie, Jianjun Qian, and Jian Yang. Progressive point cloud deconvolution generation network. In *Computer Vision–ECCV 2020*, pages 397–413. Springer, 2020. 3, 8

[17] Chenru Jiang, Chengrui Zhang, Xi Yang, Jie Sun, Yifei Zhang, Bin Dong, and Kaizhu Huang. Consistency diffusion models for single-image 3d reconstruction with priors. *arXiv preprint arXiv:2501.16737*, 2025. 4, 3

[18] Hyeongju Kim, Hyeonseung Lee, Woo Hyun Kang, Joun Yeop Lee, and Nam Soo Kim. Softflow: Probabilistic framework for normalizing flow on manifolds. In *Advances in Neural Information Processing Systems*, pages 16388–16397, 2020. 1, 3, 4, 6

[19] Jinwoo Kim, Jaehoon Yoo, Juho Lee, and Seunghoon Hong. Setvae: Learning hierarchical composition for generative modeling of set-structured data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15059–15068, 2021. 1, 3, 4, 6, 8

[20] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6, 1

[21] Roman Klokov, Edmond Boyer, and Jakob Verbeek. Discrete point flow networks for efficient point cloud generation. In *European Conference on Computer Vision*, pages 694–710. Springer, 2020. 1, 3, 4, 6, 8

[22] Ruihui Li, Xianzhi Li, Ka-Hei Hui, and Chi-Wing Fu. Sp-gan: Sphere-guided 3d shape generation and manipulation. *ACM Transactions on Graphics (TOG)*, 40(4):1–12, 2021. 3, 4, 8

[23] Yuhan Li, Yishun Dou, Xuanhong Chen, Bingbing Ni, Yilin Sun, Yutian Liu, and Fuzhen Wang. Generalized deep 3d shape prior via part-discretized diffusion process. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16784–16794, 2023. 6

[24] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022. 2, 5, 1

[25] Zhijian Liu, Haotian Tang, Yujun Lin, and Song Han. Point-voxel cnn for efficient 3d deep learning. In *Advances in Neural Information Processing Systems*, 2019. 6, 1

[26] Zhen Liu, Yao Feng, Michael J Black, Derek Nowrouzezahrai, Liam Paull, and Weiyang Liu. Meshdiffusion: Score-based generative 3d mesh modeling. *arXiv preprint arXiv:2303.08133*, 2023. 3, 6

[27] Cheng Lu and Yang Song. Simplifying, stabilizing and scaling continuous-time consistency models. *arXiv preprint arXiv:2410.11081*, 2024. 2, 3, 4, 5, 1

[28] Guanxing Lu, Zifeng Gao, Tianxing Chen, Wenxun Dai, Ziwei Wang, Wenbo Ding, and Yansong Tang. Manicm: Real-time 3d diffusion policy via consistency model for robotic manipulation. *arXiv preprint arXiv:2406.01586*, 2024. 1, 3, 4

[29] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2837–2845, 2021. 1, 3, 4, 6, 8

[30] Shentong Mo, Enze Xie, Ruihang Chu, Lanqing Hong, Matthias Niessner, and Zhenguo Li. Dit-3d: Exploring plain diffusion transformers for 3d shape generation. *Advances in neural information processing systems*, 36:67960–67971, 2023. 1, 3, 4, 6

[31] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 523–540. Springer, 2020. 1

[32] Songyou Peng, Chiyu Jiang, Yiyi Liao, Michael Niemeyer, Marc Pollefeys, and Andreas Geiger. Shape as points: A differentiable poisson solver. *Advances in Neural Information Processing Systems*, 34:13032–13044, 2021. 6

[33] Christoph Petroll, Sebastian Eilermann, Philipp Hoefer, and Oliver Niggemann. A generative neural network approach for 3d multi-criteria design generation and optimization of an engine mount for an unmanned air vehicle. *arXiv preprint arXiv:2311.03414*, 2023. 1, 3, 4

[34] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 3

[35] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 6, 1, 3

[36] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 3

[37] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000. 3

[38] Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your flow: Scaling continuous-time flow map distillation. *arXiv preprint arXiv:2506.14603*, 2025. 3

[39] Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022. 4

[40] Dong Wook Shu, Sung Woo Park, and Junseok Kwon. 3d point cloud generative adversarial network based on tree structured graph convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3859–3868, 2019. 1, 3, 4, 8

[41] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 6

[42] Yang Song and Prafulla Dhariwal. Improved techniques for training consistency models. *arXiv preprint arXiv:2310.14189*, 2023. 1, 3, 4

[43] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019. 2

[44] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 1, 2, 3, 4

[45] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023. 1, 2, 3, 4, 5

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 1

[47] Lemeng Wu, Dilin Wang, Chengyue Gong, Xingchao Liu, Yunyang Xiong, Rakesh Ranjan, Raghuraman Krishnamoorthi, Vikas Chandra, and Qiang Liu. Fast point cloud generation with straight flows. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9445–9454, 2023. 1, 3, 6

[48] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4541–4550, 2019. 1, 3, 4, 6, 8

[49] Xiaohui Zeng, Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler, and Karsten Kreis. Lion: Latent point diffusion models for 3d shape generation. In *Advances in Neural Information Processing Systems*, pages 10021–10039, 2022. 1, 3, 4, 6, 8

[50] Dongsu Zhang, Changwoon Choi, Jeonghwan Kim, and Young Min Kim. Learning to generate 3d shapes with generative cellular automata. *arXiv preprint arXiv:2103.04130*, 2021. 3, 8

[51] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 3

[52] Yuchen Zhang and Jian Zhou. Inverse flow and consistency models. In *Forty-second International Conference on Machine Learning*, 2025. 3

[53] Linqi Zhou, Yilun Du, and Jiajun Wu. 3d shape generation and completion through point-voxel diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5826–5835, 2021. 1, 3, 4, 6, 8