

Appendix / Supplemental Material

A Derivation of Forward Process in 3DDM

We start the derivation of the 3DDM forward process from timestep $t - 1$ to t with our proposed diagonal matrix $\beta_t^E = \text{diag}(\beta_{t,x}^E, \beta_{t,y}^E, \beta_{t,z}^E)$. The Equation 1 indicates the success of the diagonal matrix β_t^E being incorporated into the proposed anisotropic energy-aware forward process by adding anisotropic noise.

$$X^t = \text{diag} \left(\sqrt{(1 - \beta_{t,x}^E)}, \sqrt{(1 - \beta_{t,y}^E)}, \sqrt{(1 - \beta_{t,z}^E)} \right) \cdot X^{t-1} + \text{diag} \left(\sqrt{\beta_{t,x}^E}, \sqrt{\beta_{t,y}^E}, \sqrt{\beta_{t,z}^E} \right) \cdot \epsilon_1 \quad (1)$$

where X^t and X^{t-1} are the noisy 3D point clouds in timesteps t and $t - 1$. $\beta_t^E = \text{diag}(\beta_{t,x}^E, \beta_{t,y}^E, \beta_{t,z}^E)$, which is achieved from the energy-aware β_t rescaling model proposed in our 3DDM. ϵ_1 is a noisy 3D point cloud sampled the standard 3D Gaussian noise.

Therefore, the mean μ_q and the covariance Σ_q of the probability of $q(X^t | X^{t-1})$ from timestep $t - 1$ to t in forward process are achieved from the reparameterization trick.

$$\begin{aligned} q(X^t | X^{t-1}) &= \mathcal{N}(X^t | \mu_q(X^t, t), \Sigma_q(X^t, t)) \\ \mu_q &= \text{diag} \left(\sqrt{(1 - \beta_{t,x}^E)}, \sqrt{(1 - \beta_{t,y}^E)}, \sqrt{(1 - \beta_{t,z}^E)} \right) \cdot X^{t-1} \\ \Sigma_q &= \text{diag} \left(\beta_{t,x}^E, \beta_{t,y}^E, \beta_{t,z}^E \right) \end{aligned} \quad (2)$$

We then derive the noisy 3D point cloud X^t in any timestep t can be achieved from the noise X^0 in timestep 0 in the proposed anisotropic energy-aware forward process.

Following the analogies in Equation 1, the diffusion process from timestep $t - 2$ to $t - 1$ is defined by Equation 3.

$$X^{t-1} = \text{diag} \left(\sqrt{(\alpha_{t-1,x}^E)}, \sqrt{(\alpha_{t-1,y}^E)}, \sqrt{(\alpha_{t-1,z}^E)} \right) \cdot X^{t-2} + \text{diag} \left(\sqrt{\beta_{t-1,x}^E}, \sqrt{\beta_{t-1,y}^E}, \sqrt{\beta_{t-1,z}^E} \right) \cdot \epsilon_2 \quad (3)$$

where $\alpha_{t-1,x}^E = 1 - \beta_{t-1,x}^E$. ϵ_2 is also a noisy 3D point cloud sampled the standard 3D Gaussian noise.

Substituting Equation 3 into Equation 1, the 3D point cloud in timestep t can be directly achieved from the timestep $t - 2$.

$$\begin{aligned} X^t &= \text{diag} \left(\sqrt{(1 - \beta_{t,x}^E)}, \sqrt{(1 - \beta_{t,y}^E)}, \sqrt{(1 - \beta_{t,z}^E)} \right) \\ &\quad \cdot \text{diag} \left(\sqrt{(\alpha_{t-1,x}^E)}, \sqrt{(\alpha_{t-1,y}^E)}, \sqrt{(\alpha_{t-1,z}^E)} \right) \cdot X^{t-2} \\ &\quad + \text{diag} \left(\sqrt{\beta_{t-1,x}^E}, \sqrt{\beta_{t-1,y}^E}, \sqrt{\beta_{t-1,z}^E} \right) \cdot \epsilon_2 \\ &\quad + \text{diag} \left(\sqrt{\beta_{t,x}^E}, \sqrt{\beta_{t,y}^E}, \sqrt{\beta_{t,z}^E} \right) \cdot \epsilon_1 \\ &= \text{diag} \left(\sqrt{\alpha_{t,x}^E \alpha_{t-1,x}^E}, \sqrt{\alpha_{t,y}^E \alpha_{t-1,y}^E}, \sqrt{\alpha_{t,z}^E \alpha_{t-1,z}^E} \right) \cdot X^{t-2} \\ &\quad + \text{diag} \left(\sqrt{\alpha_{t,x}^E (\beta_{t-1,x}^E)}, \sqrt{\alpha_{t,y}^E (\beta_{t-1,y}^E)}, \sqrt{\alpha_{t,z}^E (\beta_{t-1,z}^E)} \right) \\ &\quad \cdot \epsilon_2 + \text{diag} \left(\sqrt{1 - \alpha_{t,x}^E}, \sqrt{1 - \alpha_{t,y}^E}, \sqrt{1 - \alpha_{t,z}^E} \right) \cdot \epsilon_1 \\ &= \text{diag} \left(\sqrt{\alpha_{t,x}^E \alpha_{t-1,x}^E}, \sqrt{\alpha_{t,y}^E \alpha_{t-1,y}^E}, \sqrt{\alpha_{t,z}^E \alpha_{t-1,z}^E} \right) \cdot X^{t-2} \\ &\quad + \text{diag} \left(\sqrt{(\bar{\rho}_{(t-1:t),x}^E)}, \sqrt{(\bar{\rho}_{(t-1:t),y}^E)}, \sqrt{(\bar{\rho}_{(t-1:t),z}^E)} \right) \cdot \epsilon \end{aligned} \quad (4)$$

where $\bar{\rho}_{(t-1:t),x}^E = 1 - \bar{\alpha}_{(t-1:t),x}^E = 1 - \alpha_{t,x}^E \cdot \alpha_{t-1,x}^E$. Since the linear combinations of 3D Gaussian noise still follow a 3D Gaussian distribution, ϵ_1 and ϵ_2 can be merged into a single term ϵ , which combines along the x, y , and z axes.

Following the analogies in Equation 4, the noisy 3D point cloud X^t in any timestep t can be achieved from the noise X^0 in timestep 0, where it is the clean 3D point cloud.

$$X^t = \text{diag} \left(\sqrt{\bar{\alpha}_{t,x}^E}, \sqrt{\bar{\alpha}_{t,y}^E}, \sqrt{\bar{\alpha}_{t,z}^E} \right) \cdot X^0 + \text{diag} \left(\sqrt{(1 - \bar{\alpha}_{t,x}^E)}, \sqrt{(1 - \bar{\alpha}_{t,y}^E)}, \sqrt{(1 - \bar{\alpha}_{t,z}^E)} \right) \cdot \hat{\epsilon} \quad (5)$$

where $\bar{\alpha}_{t,x}^E = \alpha_{t,x}^E \cdot \alpha_{t-1,x}^E \cdot \alpha_{t-2,x}^E \cdot \dots \cdot \alpha_{1,x}^E$, and $1 - \bar{\alpha}_{t,x}^E = 1 - \alpha_{t,x}^E \cdot \alpha_{t-1,x}^E \cdot \alpha_{t-2,x}^E \cdot \dots \cdot \alpha_{1,x}^E$. $\hat{\epsilon}$ is the standard 3D Gaussian noise used to simulate random perturbations during the diffusion process.

Finally, we can achieve the probability of $q(X^{(0:T)})$ in the proposed anisotropic energy-aware forward process.

$$q(X^{(0:T)}) = q(X^0) \prod_{t=1}^T q(X^t | X^{t-1}) \quad (6)$$

$$q(X^t | X^{t-1}) = \mathcal{N}(X^t | \mu_q(X^t, t), \Sigma_q(X^t, t))$$

where the mean μ_q and the covariance Σ_q in the forward process are defined in Equation 2.

B Derivation of Reverse Process in 3DDM

The reverse process aims to achieve the 3D point cloud from timestep t to $t - 1$, which is opposite to the forward process. This section derives the μ_θ and the Σ_θ from timestep t to $t - 1$ in reverse process.

First, Bayes' theorem is used to achieve the reverse probability $p_\theta(X^{t-1} | X^t, c)$.

$$p_\theta(X^{t-1} | X^t, c) = \underbrace{q(X^t | X^{t-1}, X^0)}_{\textcircled{1}} \frac{\overbrace{q(X^{t-1} | X^0)}^{\textcircled{3}}}{\underbrace{q(X^t | X^0)}_{\textcircled{2}}} \quad (7)$$

where $\textcircled{1}$ and $\textcircled{2}$ are achieved from the forward process in Equations 1 and 5. where c denotes the latent features of partial 3D point clouds. Since c does not participate in the forward process, we remove c in $\textcircled{1}$, $\textcircled{2}$ and $\textcircled{3}$.

Therefore, the means and the covariances of $\textcircled{1}$ and $\textcircled{2}$ can be achieved from the reparameterization trick, as defined by Equations 8 and 9, respectively.

$$\begin{aligned} \textcircled{1} &\sim N \left(\text{diag} \left(\sqrt{\alpha_{t,x}^E}, \sqrt{\alpha_{t,y}^E}, \sqrt{\alpha_{t,z}^E} \right) \right) \cdot X^{t-1}, \\ &\quad \text{diag} \left(\sqrt{(1 - \alpha_{t,x}^E)}, \sqrt{(1 - \alpha_{t,y}^E)}, \sqrt{(1 - \alpha_{t,z}^E)} \right) \end{aligned} \quad (8)$$

$$\begin{aligned} \textcircled{2} &\sim N \left(\text{diag} \left(\sqrt{\bar{\alpha}_{t,x}^E}, \sqrt{\bar{\alpha}_{t,y}^E}, \sqrt{\bar{\alpha}_{t,z}^E} \right) \right) \cdot X^0, \\ &\quad \text{diag} \left(\sqrt{(1 - \bar{\alpha}_{t,x}^E)}, \sqrt{(1 - \bar{\alpha}_{t,y}^E)}, \sqrt{(1 - \bar{\alpha}_{t,z}^E)} \right) \end{aligned} \quad (9)$$

Substituting $\textcircled{3}$ into Equation 9, the mean and the covariance of $\textcircled{3}$ can be achieved.

$$\begin{aligned} \textcircled{3} &\sim N \left(\text{diag} \left(\sqrt{\bar{\alpha}_{t-1,x}^E}, \sqrt{\bar{\alpha}_{t-1,y}^E}, \sqrt{\bar{\alpha}_{t-1,z}^E} \right) \right) \cdot X^0, \\ &\quad \text{diag} \left(\sqrt{\bar{\rho}_{(1:t-1),x}^E}, \sqrt{\bar{\rho}_{(1:t-1),y}^E}, \sqrt{\bar{\rho}_{(1:t-1),z}^E} \right) \cdot \hat{\epsilon}_{t-1} \end{aligned} \quad (10)$$

where $\bar{\rho}_{(1:t-1),x}^E = 1 - \bar{\alpha}_{t-1,x}^E$.

Second, the means and covariances of ①, ② and ③ in Equations 8, 9 and 10 are substituted in the probability density function of the 3D Gaussian distribution in Equation 11.

$$f(x) = \frac{1}{(2\pi)^{\frac{3}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \cdot \exp \left(-\frac{1}{2} \underbrace{\underbrace{(x - \mu)^\top \cdot \Sigma^{-1} \cdot (x - \mu)}_{④}}_{④} \right) \quad (11)$$

where $M = \frac{1}{(2\pi)^{\frac{3}{2}} \cdot |\Sigma|^{\frac{1}{2}}}$. Since M is a constant, it is omitted in the subsequent calculations.

Substituting Equation 8 into ④:

$$\begin{aligned} &= \left(X^t - \text{diag} \left(\sqrt{\alpha_{t,x}^E}, \sqrt{\alpha_{t,y}^E}, \sqrt{\alpha_{t,z}^E} \right) \cdot X^{t-1} \right)^\top \\ &\quad \cdot \text{diag} \left(\sqrt{(1 - \alpha_{t,x}^E)}, \sqrt{(1 - \alpha_{t,y}^E)}, \sqrt{(1 - \alpha_{t,z}^E)} \right)^{-1} \\ &\quad \cdot \left(X^t - \text{diag} \left(\sqrt{\alpha_{t,x}^E}, \sqrt{\alpha_{t,y}^E}, \sqrt{\alpha_{t,z}^E} \right) \cdot X^{t-1} \right) \\ &= \left(X^t \right)^\top \cdot \text{diag} \left(\frac{1}{\beta_{t,x}^E}, \frac{1}{\beta_{t,y}^E}, \frac{1}{\beta_{t,z}^E} \right) \cdot X^t \\ &\quad - \left(X^t \right)^\top \cdot \text{diag} \left(\frac{\sqrt{\alpha_{t,x}^E}}{1 - \alpha_{t,x}^E}, \frac{\sqrt{\alpha_{t,y}^E}}{1 - \alpha_{t,y}^E}, \frac{\sqrt{\alpha_{t,z}^E}}{1 - \alpha_{t,z}^E} \right) \cdot X^{t-1} \quad (12) \\ &\quad - \left(X^{t-1} \right)^\top \cdot \text{diag} \left(\frac{\sqrt{\alpha_{t,x}^E}}{1 - \alpha_{t,x}^E}, \frac{\sqrt{\alpha_{t,y}^E}}{1 - \alpha_{t,y}^E}, \frac{\sqrt{\alpha_{t,z}^E}}{1 - \alpha_{t,z}^E} \right) \cdot X^t \\ &\quad \underbrace{- \left(X^{t-1} \right)^\top \cdot \text{diag} \left(\frac{\alpha_{t,x}^E}{1 - \alpha_{t,x}^E}, \frac{\alpha_{t,y}^E}{1 - \alpha_{t,y}^E}, \frac{\alpha_{t,z}^E}{1 - \alpha_{t,z}^E} \right) \cdot X^{t-1}}_{⑤} \\ &\quad + \underbrace{\left(X^{t-1} \right)^\top \cdot \text{diag} \left(\frac{\alpha_{t,x}^E}{1 - \alpha_{t,x}^E}, \frac{\alpha_{t,y}^E}{1 - \alpha_{t,y}^E}, \frac{\alpha_{t,z}^E}{1 - \alpha_{t,z}^E} \right) \cdot X^{t-1}}_{⑥} \end{aligned}$$

Substituting Equation 9 into ④:

$$\begin{aligned} &= \left(X^t - \text{diag} \left(\sqrt{\bar{\alpha}_{t,x}^E}, \sqrt{\bar{\alpha}_{t,y}^E}, \sqrt{\bar{\alpha}_{t,z}^E} \right) \cdot X^0 \right)^\top \\ &\quad \cdot \text{diag} \left(\sqrt{(1 - \bar{\alpha}_{t,x}^E)}, \sqrt{(1 - \bar{\alpha}_{t,y}^E)}, \sqrt{(1 - \bar{\alpha}_{t,z}^E)} \right)^{-1} \\ &\quad \cdot \left(X^t - \text{diag} \left(\sqrt{\bar{\alpha}_{t,x}^E}, \sqrt{\bar{\alpha}_{t,y}^E}, \sqrt{\bar{\alpha}_{t,z}^E} \right) \cdot X^0 \right) \\ &= \left(X^t \right)^\top \text{diag} \left(\frac{1}{1 - \bar{\alpha}_{t,x}^E}, \frac{1}{1 - \bar{\alpha}_{t,y}^E}, \frac{1}{1 - \bar{\alpha}_{t,z}^E} \right) \cdot X^t \\ &\quad - \left(X^t \right)^\top \cdot \text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t,x}^E}}{1 - \bar{\alpha}_{t,x}^E}, \frac{\sqrt{\bar{\alpha}_{t,y}^E}}{1 - \bar{\alpha}_{t,y}^E}, \frac{\sqrt{\bar{\alpha}_{t,z}^E}}{1 - \bar{\alpha}_{t,z}^E} \right) \cdot X^0 \\ &\quad - \left(X^0 \right)^\top \cdot \text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t,x}^E}}{1 - \bar{\alpha}_{t,x}^E}, \frac{\sqrt{\bar{\alpha}_{t,y}^E}}{1 - \bar{\alpha}_{t,y}^E}, \frac{\sqrt{\bar{\alpha}_{t,z}^E}}{1 - \bar{\alpha}_{t,z}^E} \right) \cdot X^t \\ &\quad + \left(X^0 \right)^\top \cdot \text{diag} \left(\frac{\bar{\alpha}_{t,x}^E}{1 - \bar{\alpha}_{t,x}^E}, \frac{\bar{\alpha}_{t,y}^E}{1 - \bar{\alpha}_{t,y}^E}, \frac{\bar{\alpha}_{t,z}^E}{1 - \bar{\alpha}_{t,z}^E} \right) \cdot X^0 \quad (13) \end{aligned}$$

Substituting Equation 10 into ④:

$$\begin{aligned} &= \left(X^{t-1} - \text{diag} \left(\sqrt{\bar{\alpha}_{t-1,x}^E}, \sqrt{\bar{\alpha}_{t-1,y}^E}, \sqrt{\bar{\alpha}_{t-1,z}^E} \right) \cdot X^0 \right)^\top \\ &\quad \cdot \text{diag} \left(\sqrt{(\bar{\beta}_{t-1,x}^E)}, \sqrt{(\bar{\beta}_{t-1,y}^E)}, \sqrt{(\bar{\beta}_{t-1,z}^E)} \right)^{-1} \\ &\quad \cdot \left(X^{t-1} - \text{diag} \left(\sqrt{\bar{\alpha}_{t-1,x}^E}, \sqrt{\bar{\alpha}_{t-1,y}^E}, \sqrt{\bar{\alpha}_{t-1,z}^E} \right) \cdot X^0 \right) \\ &= \underbrace{\left(X^{t-1} \right)^\top \cdot \text{diag} \left(\frac{1}{\bar{\rho}_{(1:t-1),x}^E}, \frac{1}{\bar{\rho}_{(1:t-1),y}^E}, \frac{1}{\bar{\rho}_{(1:t-1),z}^E} \right) \cdot X^{t-1}}_{⑦} \\ &\quad - \underbrace{\left(X^{t-1} \right)^\top \cdot \text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t-1,x}^E}}{1 - \bar{\alpha}_{t-1,x}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,y}^E}}{1 - \bar{\alpha}_{t-1,y}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,z}^E}}{1 - \bar{\alpha}_{t-1,z}^E} \right) \cdot X^0}_{⑧} \\ &\quad - \left(X^0 \right)^\top \cdot \text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t-1,x}^E}}{1 - \bar{\alpha}_{t-1,x}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,y}^E}}{1 - \bar{\alpha}_{t-1,y}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,z}^E}}{1 - \bar{\alpha}_{t-1,z}^E} \right) \cdot X^{t-1} \\ &\quad + \left(X^0 \right)^\top \cdot \text{diag} \left(\frac{\bar{\alpha}_{t-1,x}^E}{1 - \bar{\alpha}_{t-1,x}^E}, \frac{\bar{\alpha}_{t-1,y}^E}{1 - \bar{\alpha}_{t-1,y}^E}, \frac{\bar{\alpha}_{t-1,z}^E}{1 - \bar{\alpha}_{t-1,z}^E} \right) \cdot X^0 \quad (14) \end{aligned}$$

The aim of the reverse process is to predict $\boldsymbol{\mu}_\theta$ and $\boldsymbol{\Sigma}_\theta$ from timestep t to $t-1$, therefor, ④ can be written as:

$$\begin{aligned} &\quad \left(X^{t-1} - \boldsymbol{\mu}_\theta \right)^\top \boldsymbol{\Sigma}_\theta^{-1} \left(X^{t-1} - \boldsymbol{\mu}_\theta \right) \\ &= \underbrace{\left(X^{t-1} \right)^\top \cdot \boldsymbol{\Sigma}_\theta^{-1} \cdot X^{t-1}}_{⑨} - \underbrace{\left(X^{t-1} \right)^\top \cdot \boldsymbol{\Sigma}_\theta^{-1} \cdot \boldsymbol{\mu}_\theta}_{⑩} \quad (15) \\ &\quad - \boldsymbol{\mu}_\theta^\top \cdot \boldsymbol{\Sigma}_\theta^{-1} \cdot X^{t-1} + \boldsymbol{\mu}_\theta^\top \cdot \boldsymbol{\Sigma}_\theta^{-1} \cdot \boldsymbol{\mu}_\theta \end{aligned}$$

To derive $\boldsymbol{\Sigma}_\theta^{-1}$ in ⑨, two terms ⑤ and ⑦ in Equations 12 and 14 are added, which correspond to the additive exponents in the probability density function of the 3D Gaussian distribution.

$$\begin{aligned} &\textcircled{⑤} + \textcircled{⑦} = \left(X^{t-1} \right)^\top \\ &\quad \cdot \text{diag} \left(\frac{1 - \bar{\alpha}_{t,x}^E}{\beta_{t,x}^E \bar{\rho}_{(1:t-1),x}^E}, \frac{1 - \bar{\alpha}_{t,y}^E}{\beta_{t,y}^E \bar{\rho}_{(1:t-1),y}^E}, \frac{1 - \bar{\alpha}_{t,z}^E}{\beta_{t,z}^E \bar{\rho}_{(1:t-1),z}^E} \right) \cdot X^{t-1} \quad (16) \end{aligned}$$

where $\bar{\rho}_{(1:t-1),x}^E = 1 - \bar{\alpha}_{t-1,x}^E$.

Therefore, the $\boldsymbol{\Sigma}_\theta^{-1}$ is:

$$\text{diag} \left(\frac{1 - \bar{\alpha}_{t,x}^E}{(1 - \bar{\alpha}_{t-1,x}^E) \beta_{t,x}^E}, \frac{1 - \bar{\alpha}_{t,y}^E}{(1 - \bar{\alpha}_{t-1,y}^E) \beta_{t,y}^E}, \frac{1 - \bar{\alpha}_{t,z}^E}{(1 - \bar{\alpha}_{t-1,z}^E) \beta_{t,z}^E} \right) \quad (17)$$

The $\boldsymbol{\Sigma}_\theta$ is the reciprocal of $\boldsymbol{\Sigma}_\theta^{-1}$:

$$\begin{aligned} \boldsymbol{\Sigma}_\theta &= \\ &\quad \text{diag} \left(\frac{(1 - \bar{\alpha}_{t-1,x}^E) \beta_{t,x}^E}{1 - \bar{\alpha}_{t,x}^E}, \frac{(1 - \bar{\alpha}_{t-1,y}^E) \beta_{t,y}^E}{1 - \bar{\alpha}_{t,y}^E}, \frac{(1 - \bar{\alpha}_{t-1,z}^E) \beta_{t,z}^E}{1 - \bar{\alpha}_{t,z}^E} \right) \quad (18) \end{aligned}$$

To derive $\boldsymbol{\mu}_\theta$ in ⑩, two terms ⑤ and ⑧ in Equations 12 and 14 are added, which correspond to the additive exponents in the probability density function of the 3D Gaussian distribution.

$$\begin{aligned} &\textcircled{⑤} + \textcircled{⑧} = - \left(X^{t-1} \right)^\top \cdot \boldsymbol{\Sigma}_\theta^{-1} \cdot \boldsymbol{\mu}_\theta = \\ &\quad - \left(X^{t-1} \right)^\top \cdot \text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t-1,x}^E}}{\bar{\rho}_{(1:t-1),x}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,y}^E}}{\bar{\rho}_{(1:t-1),y}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,z}^E}}{\bar{\rho}_{(1:t-1),z}^E} \right) \\ &\quad \cdot X^0 + \text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t-1,x}^E}}{1 - \bar{\alpha}_{t-1,x}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,y}^E}}{1 - \bar{\alpha}_{t-1,y}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,z}^E}}{1 - \bar{\alpha}_{t-1,z}^E} \right) \cdot X^t \quad (19) \end{aligned}$$

Multiplying Equation 19 by Σ_θ , the expression for μ_θ is achieved.

$$\begin{aligned} \mu_\theta &= \Sigma_\theta \cdot \Sigma_\theta^{-1} \cdot \mu_\theta \\ &= \text{diag} \left(\frac{\bar{\rho}_{(1:t-1),x}^E \cdot \beta_{t,x}^E}{1 - \bar{\alpha}_{t,x}^E}, \frac{\bar{\rho}_{(1:t-1),y}^E \cdot \beta_{t,y}^E}{1 - \bar{\alpha}_{t,y}^E}, \frac{\bar{\rho}_{(1:t-1),z}^E \cdot \beta_{t,z}^E}{1 - \bar{\alpha}_{t,z}^E} \right) \\ &\quad \cdot \left(\text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t-1,x}^E}}{1 - \bar{\alpha}_{t-1,x}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,y}^E}}{1 - \bar{\alpha}_{t-1,y}^E}, \frac{\sqrt{\bar{\alpha}_{t-1,z}^E}}{1 - \bar{\alpha}_{t-1,z}^E} \right) \cdot X^0 + \right. \\ &\quad \left. \text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t,x}^E}}{1 - \bar{\alpha}_{t,x}^E}, \frac{\sqrt{\bar{\alpha}_{t,y}^E}}{1 - \bar{\alpha}_{t,y}^E}, \frac{\sqrt{\bar{\alpha}_{t,z}^E}}{1 - \bar{\alpha}_{t,z}^E} \right) \right) \quad (20) \\ &= \text{diag} \left(\frac{\beta_{t,x}^E \cdot \sqrt{\bar{\alpha}_{t-1,x}^E}}{\bar{\rho}_{(1:t-1),x}^E}, \frac{\beta_{t,y}^E \cdot \sqrt{\bar{\alpha}_{t-1,y}^E}}{\bar{\rho}_{(1:t-1),y}^E}, \frac{\beta_{t,z}^E \cdot \sqrt{\bar{\alpha}_{t-1,z}^E}}{\bar{\rho}_{(1:t-1),z}^E} \right) \\ &\quad \cdot X^0 + \text{diag} \left(\frac{\sqrt{\bar{\alpha}_{t,x}^E} \cdot \bar{\rho}_{(1:t-1),x}^E}{1 - \bar{\alpha}_{t,x}^E}, \frac{\sqrt{\bar{\alpha}_{t,y}^E} \cdot \bar{\rho}_{(1:t-1),y}^E}{1 - \bar{\alpha}_{t,y}^E}, \right. \\ &\quad \left. \frac{\sqrt{\bar{\alpha}_{t,z}^E} \cdot \bar{\rho}_{(1:t-1),z}^E}{1 - \bar{\alpha}_{t,z}^E} \right) \cdot X^t \end{aligned}$$

Replacing X^0 with X^t using Equation 5, the final μ_θ is:

$$\begin{aligned} \mu_\theta &= \text{diag} \left(\frac{1}{\sqrt{\alpha_{t,x}^E}}, \frac{1}{\sqrt{\alpha_{t,y}^E}}, \frac{1}{\sqrt{\alpha_{t,z}^E}} \right) \cdot \left(X^t - \right. \\ &\quad \left. \text{diag} \left(\frac{\beta_{t,x}^E}{\sqrt{1 - \bar{\alpha}_{t,x}^E}}, \frac{\beta_{t,y}^E}{\sqrt{1 - \bar{\alpha}_{t,y}^E}}, \frac{\beta_{t,z}^E}{\sqrt{1 - \bar{\alpha}_{t,z}^E}} \right) \cdot \epsilon_\theta(X^t, t, c) \right) \quad (21) \end{aligned}$$

where $\epsilon_\theta(X^t, t, c)$ is the predicted 3D Gaussian noise deriving from replacing X^0 with X^t using Equation 5.

C Derivation of Anisotropic Quadratic Loss in 3DDM

To obtain the most plausible reconstruction of x^0 and to find the model parameters θ that maximize the likelihood of x^0 , the maximum likelihood estimation is adopted. The term $p_\theta(x^0|c)$ denotes the marginal probability of the generated point cloud conditioned on c . The variational evidence lower bound (ELBO) is constructed for the logarithm of the marginal probability $p_\theta(x^0|c)$ in 3DDM.

$$\begin{aligned} \log p_\theta(X^0|c) &= \log E_{q(X^{(1:T)}|X^0)} \left[\frac{p_\theta(X^{(0:T)}|c)}{q(X^{(1:T)}|X^0)} \right] \\ &\geq E_{q(X^{(1:T)}|X^0)} \left[\log \frac{p_\theta(X^{(0:T)}|c)}{q(X^{(1:T)}|X^0)} \right] \quad (22) \end{aligned}$$

where c denotes the latent features of partial 3D point clouds.

To construct the loss function, we take the negative of both sides of the equation:

$$\begin{aligned} -\log p_\theta(X^0|c) &\leq E_{q(X^{(1:T)}|X^0)} \left[\log \frac{q(X^{(1:T)}|X^0)}{p_\theta(X^{(0:T)}|c)} \right] \\ &= L_{VLB} \end{aligned} \quad (23)$$

where L_{VLB} is the variational lower bound.

Then, the L_{VLB} is written as:

$$\begin{aligned} L_{VLB} &= E_{q(X^{(1:T)}|X^0)} \left[\log \frac{q(X^{(1:T)}|X^0)}{p_\theta(X^{0:T}|c)} \right] \\ &= E_{q(X^{1:T}|X^0)} \left[\log \frac{q(X^1|X^0) \dots q(X^T|X^0)}{p_\theta(X^T) \dots p_\theta(X^1|X^0, c)} \right] \\ &= E_{q(X^{(1:T)}|X^0)} \left[\log \frac{\Pi_{t=1}^T q(X^{(t)}|X^{t-1})}{p_\theta(X^T) \cdot \Pi_{t=1}^T p_\theta(X^{t-1}|X^t, c)} \right] \\ &= E_q(X^{(1:T)}|X^0) \left[-\log p_\theta(X^T) \right. \\ &\quad \left. + \Sigma_{t=1}^T \log \frac{q(X^t|x^{t-1})}{p_\theta(X^{t-1}|X^t, c)} \right] \\ &= E_q(X^{(1:T)}|X^0) \left[-\log p_\theta(X^T) \right. \\ &\quad \left. + \underbrace{\Sigma_{t=2}^T \log \frac{q(X^t|X^{t-1}, X^0)}{p_\theta(X^{t-1}|X^t, c)}}_{\textcircled{7}} + \log \frac{q(X^1|X^0)}{p_\theta(X^0|X^1, c)} \right] \end{aligned} \quad (24)$$

Applying Bayes' theorem to decompose $\textcircled{7}$:

$$\begin{aligned} q(X^t|X^{t-1}, X^0) &= \frac{q(X^t, X^{t-1}, X^0)}{q(X^{t-1}, X^0)} \\ &= \frac{q(X^{t-1}|X^t, X^0) q(X^t|X^0) q(X^0)}{q(X^{t-1}|X^0) q(X^0)} \\ &= \frac{q(X^{t-1}|X^t, X^0) q(X^t|X^0)}{q(X^{t-1}|X^0)} \quad (25) \end{aligned}$$

Substituting Equation 25 into Equation 24:

$$\begin{aligned} L_{VLB} &= E_q(X^{(1:T)}|X^0) \left[-\log p_\theta(X^T) + \Sigma_{t=2}^T \right. \\ &\quad \left. \log \frac{q(X^{t-1}|X^t, X^0) q(X^t|X^0)}{p_\theta(X^{t-1}|X^t, c) q(X^{t-1}|X^0)} + \log \frac{q(X^1|X^0)}{p_\theta(X^0|X^1, c)} \right] \\ &= E_q(X^{(1:T)}|X^0) \left[-\log p_\theta(X^T) \right. \\ &\quad \left. + \Sigma_{t=2}^T \log \frac{q(X^{t-1}|X^t, X^0)}{p_\theta(X^{t-1}|X^t, c)} + \log \frac{q(X^1|X^0)}{p_\theta(X^0|X^1, c)} \right. \\ &\quad \left. + \underbrace{\Sigma_{t=2}^T \log \frac{q(X^t|X^0)}{q(X^{t-1}|X^0)}}_{\textcircled{8}} \right] \quad (26) \end{aligned}$$

$\textcircled{8}$ is then simplified:

$$\begin{aligned} \textcircled{8} &= \log \frac{q(X^2|X^0)}{q(X^1|X^0)} + \dots + \log \frac{q(X^T|X^0)}{q(X^{T-1}|X^0)} \\ &= \log q(X^2|X^0) - \log q(X^1|X^0) + \log q(X^3|X^0) \\ &\quad - \log q(X^2|X^0) + \dots + \log q(X^T|X^0) \\ &\quad - \log q(X^{T-1}|X^0) \\ &= -\log q(X^1|X^0) + \log q(X^T|X^0) = \log \frac{q(X^T|X^0)}{q(X^1|X^0)} \quad (27) \end{aligned}$$

Substituting $\textcircled{8}$ into Equation 26, L_{VLB} is further reorganized by using the property that the expectation of a sum

is equal to the sum of expectations.

$$\begin{aligned}
L_{VLB} &= E_{q(X^{(1:T)}|X^0)} \left[-\log p_\theta(X^T) + \Sigma_{t=2}^T \right. \\
&\quad \log \frac{q(X^{t-1}|X^t, X^0)}{p_\theta(X^{t-1}|X^t, c)} + \log \frac{q(X^T|X^0)}{q(X^1|X^0)} \\
&\quad \left. + \log \frac{q(X^t|X^0)}{p_\theta(X^0|X^1, c)} \right] \\
&= E_{q(X^{(1:T)}|X^0)} \left[\log \frac{q(X^T|X^0)}{p_\theta(X^T)} + \Sigma_{t=2}^T \right. \\
&\quad \log \frac{q(X^{t-1}|X^t, X^0)}{p_\theta(X^{t-1}|X^t, c)} - \log p_\theta(X^0|X^1, c) \left. \right] \quad (28) \\
&= E_{q(X^{(1:T)}|X^0)} \log \frac{q(X^T|X^0)}{p_\theta(X^T)} + \Sigma_{t=2}^T E_{q(X^{1:T}|X^0)} \\
&\quad \left[\log \frac{q(X^{t-1}|X^t, X^0)}{p_\theta(X^{t-1}|X^t, c)} \right] - E_{q(X^1|X^0)} \left[\log p_\theta(X^0|X^1, c) \right] \\
&= E_{q(X^T|X^0)} \left[\log \frac{q(X^T|X^0)}{p_\theta(X^T)} \right] + \Sigma_{t=2}^T E_{q(X^t, X^{t-1}|X^0)} \\
&\quad \left[\log \frac{q(X^{t-1}|X^t, X^0)}{p_\theta(X^{t-1}|X^t, c)} \right] - E_{q(X^1|X^0)} \left[\log p_\theta(X^0|X^1, c) \right]
\end{aligned}$$

$q(X^t, X^{t-1} | X^0)$ is decomposed using the joint probability factorization:

$$q(X^t, X^{t-1} | X^0) = q(X^t | X^0) q(X^{t-1} | X^t, X^0) \quad (29)$$

Substituting Equation 29 into Equation 28, the logarithm of the marginal probability $p_\theta(x^0|c)$ in 3DDM is written as:

$$\begin{aligned}
\log p_\theta(X^0|c) &\geq L_{VLB} = \\
&\underbrace{D_{KL}(q(X^T|X^0) \| p_\theta(X^T))}_{\text{prior matching term}} - \underbrace{E_{q(X^1|X^0)} [\log p_\theta(X^0|X^1, c)]}_{\text{reconstruction term}} \quad (30) \\
&+ \underbrace{\sum_{t=2}^T E_{q(X^t|X^0)} [D_{KL}(q(X^{t-1}|X^t, X^0) \| p_\theta(X^{t-1}|X^t, c))]}_{\text{denoising matching term}}
\end{aligned}$$

Finally, the following derivation of the loss function has been presented in our main paper.

D Implementation Details

Our 3DDM was implemented using PyTorch and trained on an Nvidia GeForce 2080Ti GPU with 12 GB memory. We train the 3DDM for 2000 epochs with a batch size of 32, an Adam optimizer and a learning rate of 0.001.

Encoder. The encoder for extracting latent features of partial 3D point clouds is PointNet (Qi et al. 2017). Specifically, the 3D point clouds are fed into a 3-128-256-512 MLP with the ReLU nonlinearity followed by a max-pooling to obtain a global 512-dimensional feature. Then, the feature is fed into a 512-256-128-256 MLP with the ReLU nonlinearity. Thus, the latent features of 256 dimensions are obtained. We also proved that different state-of-the-art encoders can be adopted in our 3DDM for feature extraction in ablation studies of the main paper.

Diffusion in 3DDM. The number of steps T in the diffusion process is 20. The detailed derivations of the forward

process, the reverse process and the loss function are derived in Appendix A, B and C, respectively. The reverse diffusion net is implemented using nine layers of a variant of MLPs, as defined by Equation 31.

$$\begin{aligned}
h^{l+1} &= (w_1 h^l + b_1) \odot \text{Sigmod}(w_2 C + b_2) + w_3 C \\
C &= \text{concat}[c, \beta_t^E, \sin(\beta_t^E), \cos(\beta_t^E)] \\
\beta_t^E &= \text{diag}(\beta_{t,x}^E, \beta_{t,y}^E, \beta_{t,z}^E)
\end{aligned} \quad (31)$$

where h^l and h^{l+1} are input and output. The first layer of the input is the 3D point cloud. c denotes the latent features of partial 3D point clouds. w_1, w_2, w_3, b_1 and b_2 are all trainable parameters. The dimensions of these MLPs are 3-128-256-512-1024-1024-256-128-3, and we use the LeakyReLU nonlinearity between the layers. The BatchNorm is used at the first eight layers. The final output 3D point cloud contains 8,192 points.

E Dataset Details

Traing and Testing Datasets

The ShapeNet34 dataset (Yu et al. 2021), constructed from the larger ShapeNet repository (Chang et al. 2015), consists of 34 object categories and exhibits greater diversity and incompleteness. It serves as a standard benchmark to evaluate the model’s capability to complete 3D shapes under varying levels of missing data. In this dataset, each partial input comprises 2,048 points, while the complete shapes (ground truth) contain 8,192 points. The training set includes 46,765 samples, and the test set has 5,750 samples.

To assess model robustness under different levels of incompleteness, the testing dataset is divided into three categories with three different missing rates: easy (25%), median (50%), and hard (75%), which enables structured evaluation across increasing difficulty levels.

Unseen Evaluation Datasets

To assess out-of-distribution generalization, we evaluate all models on unseen datasets that were never trained on, which contain ShapeNet21 (Yu et al. 2021) for new object categories, and ScanNet (Dai et al. 2017) and KITTI (Geiger, Lenz, and Urtasun 2012) for indoor and outdoor real-world scenes. These datasets introduce diverse challenges, including sensor noise, sparsity, and occlusion, reflecting the complexity of real-world 3D perception tasks.

Unseen ShapeNet21 (Yu et al. 2021) comprises the remaining 21 untrained categories from ShapeNet (Chang et al. 2015), totaling 2,305 samples. Similar to ShapeNet34, this dataset contains partial observations with varying missing rates (25%, 50%, and 75%).

Unseen real-world Indoor ScanNet scenes (Dai et al. 2017) provide annotated 3D semantic reconstructions of indoor environments. This dataset evaluates the model’s adaptability to domain shifts in sensor characteristics and point density. Particularly, ScanNet poses a high generalization challenge due to its large-scale, dense 3D semantic scenes.

Unseen real-world outdoor KITTI scenes (Geiger, Lenz, and Urtasun 2012) are collected using LiDAR sensors

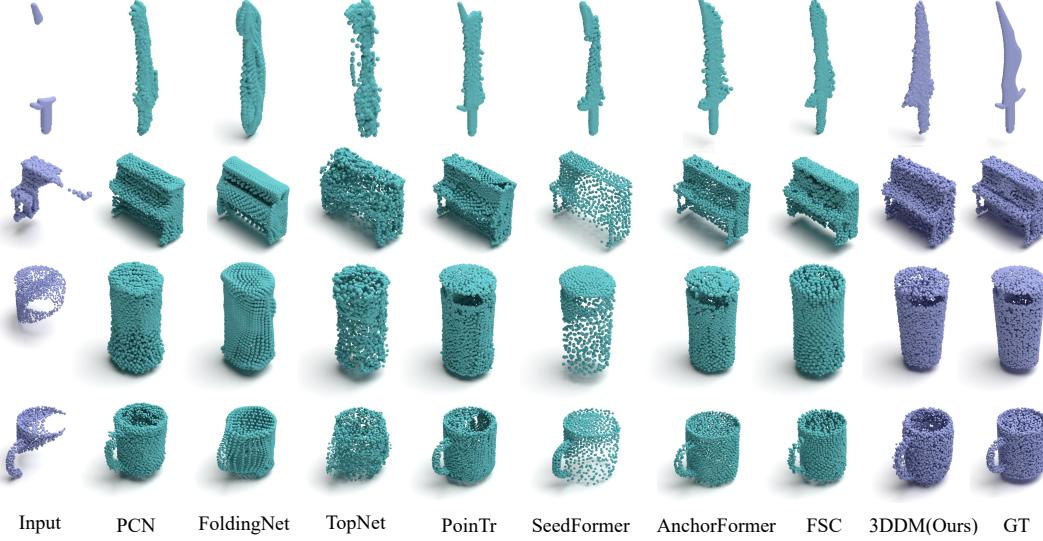


Figure 1: Qualitative completion results on ShapeNet34 datasets

Methods	airplane	bus	Bottle	Guitar	telephone	laptop	train
	F1-S / F1-M / F1-H ↑						
PCN (Yuan et al. 2018)	0.34/0.33/0.32	0.18/0.17/0.16	0.24/0.22/0.16	0.65/0.64/0.56	0.27/0.26/0.22	0.20/0.19/0.17	0.22/0.21/0.20
FoldingNet (Yang et al. 2018)	0.27/0.26/0.24	0.18/0.16/0.14	0.18/0.17/0.11	0.49/0.47/0.37	0.24/0.23/0.16	0.20/0.19/0.16	0.18/0.17/0.15
TopNet (Tchapmi et al. 2019)	0.18/0.18/0.17	0.10/0.09/0.07	0.09/0.08/0.05	0.40/0.36/0.29	0.14/0.15/0.10	0.11/0.10/0.09	0.12/0.11/0.08
PoinTr (Yu et al. 2021)	0.54/0.57/0.50	0.37/0.41/0.40	0.41/0.42/ 0.37	0.68/0.72/0.66	0.41/0.46/0.44	0.39/0.42/ 0.40	0.40/0.43/ 0.41
AnchorFormer (Chen et al. 2023)	0.45/0.46/0.42	0.28/0.28/0.27	0.33/0.32/0.27	0.75/0.75/0.67	0.32/0.32/0.30	0.30/0.31/0.29	0.32/0.32/0.30
SeedFormer (Zhou et al. 2022)	0.15/0.14/0.13	0.06/0.06/0.05	0.09/0.09/0.07	0.83/0.36/0.30	0.07/0.07/0.06	0.06/0.05/0.05	0.09/0.08/0.07
FSC (Wu et al. 2024)	0.36/0.39/0.31	0.17/0.21/0.23	0.13/0.19/0.18	0.61/0.68/0.69	0.18/0.19/0.19	0.25/0.26/0.26	0.18/0.24/0.25
3DDM (Ours)	0.60/0.59/0.53	0.45/0.45/0.41	0.44/0.43/0.36	0.82/0.80/0.73	0.49/0.48/0.45	0.44/0.43/0.39	0.44/0.44/0.40

Table 1: Evaluations on the ShapeNet34 dataset. Seven categories of results are randomly selected and displayed. The average F Score@1%(F1) has been calculated to evaluate each method. F1-S, F1-M and F1-H denote the F1 score on the simple, median and hard modes with the missing rates of 25%, 50% and 75%. (Higher F1 is better.)

Methods	Printer	Mailbox	Tower	Washer	Remote	Skateboard	Microwave
	F1-S/F1-M/F1-H ↑						
PCN (Yuan et al. 2018)	0.07/0.06/0.06	0.16/0.02/0.14	0.18/0.17/0.15	0.09/0.08/0.07	0.27/0.26/0.21	0.21/0.19/0.16	0.08/0.08/0.07
FoldingNet (Yang et al. 2018)	0.06/0.06/0.05	0.10/0.09/0.07	0.13/0.12/0.09	0.09/0.08/0.07	0.24/0.23/0.15	0.16/0.14/0.12	0.08/0.07/0.06
TopNet (Tchapmi et al. 2019)	0.04/0.04/0.03	0.08/0.07/0.06	0.08/0.08/0.07	0.05/0.04/0.03	0.14/0.13/0.08	0.15/0.13/0.07	0.05/0.04/0.03
PoinTr (Yu et al. 2021)	0.32/0.33/0.31	0.42/0.42/0.36	0.41/0.41/0.36	0.31/0.32/0.30	0.42/0.47/ 0.45	0.48/0.51/ 0.42	0.31/0.32/0.31
AnchorFormer (Chen et al. 2023)	0.19/0.19/0.18	0.45/0.41/0.33	0.34/0.33/0.27	0.18/0.17/0.17	0.37/0.37/0.35	0.44/0.43/0.33	0.17/0.17/0.16
SeedFormer (Zhou et al. 2022)	0.04/0.04/0.03	0.14/0.13/0.11	0.13/0.12/0.10	0.03/0.03/0.03	0.10/0.09/0.08	0.13/0.11/0.08	0.03/0.03/0.03
FSC (Wu et al. 2024)	0.07/0.07/0.06	0.19/0.18/0.15	0.19/0.18/0.15	0.07/0.07/0.07	0.19/0.20/0.19	0.29/0.28/0.22	0.07/0.06/0.06
3DDM(Ours)	0.32/0.33/0.31	0.44/0.43/0.36	0.42/0.41/0.36	0.32/0.32/0.30	0.49/0.47/0.41	0.52/0.51/0.38	0.32/0.32/0.31

Table 2: Evaluations on the unseen ShapeNet21 that contains 21 unseen categories. Seven categories of results are randomly selected and displayed. (Higher F1 is better.)

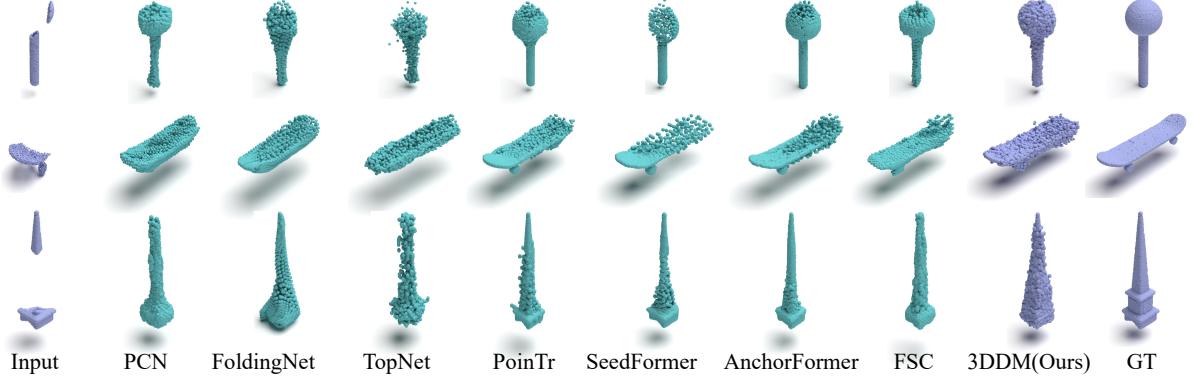


Figure 2: Qualitative completion results on the unseen categories in the ShapeNet21 dataset that were not trained.

and capture real-world driving environments. The data exhibit sparsity, sensor noise, and significant occlusion. The evaluation focuses on 1,000 car instances extracted via 3D bounding boxes. In contrast to synthetic datasets of ShapeNet, KITTI features naturally incomplete and sparse point clouds, providing a rigorous benchmark for real-scene generalization.

F Additional Experiments on ShapeNet34 and unseen ShapeNet21

We randomly select the results of seven categories in ShapeNet34, as reported Table 1. Figure 1 presents qualitative comparisons. PCN and FoldingNet suffer from distortion and structural loss, while PoinTr and AnchorFormer often produce sparse or incomplete shapes. In contrast, 3DDM yields high-quality and more coherent outputs, even under severe incompleteness.

We also randomly select the results of seven unseen categories in Shapenet21 that were not trained, as reported Table 2. 3DDM ranks among the top methods across most categories and occlusion levels, significantly outperforming PoinTr and AnchorFormer in challenging cases. Figure 2 further shows that 3DDM demonstrates strong and stable generalization to unseen categories, making it a promising foundation for future cross-domain applications.

G Additional Qualitative Comparisons on ScanNet Dataset

Figure 3 shows qualitative completion results in a ScanNet scene. Most methods struggle to generalize, producing sparse or blurry reconstructions. In contrast, our 3DDM recovers complete and coherent scene geometry with higher point density. Notably, it reconstructs ground surfaces and fine structures like chair legs, demonstrating strong structural understanding and robustness in real-world scenes.

Figure 4 presents object-level results on a wall, a floor, chairs and tables. PCN, FoldingNet and TopNet often produce oversmoothed shapes and miss local details. PoinTr, AnchorFormer, and SeedFormer better capture global structure but still struggle with fine details, especially in occluded

or incomplete regions. In contrast, our 3DDM faithfully reconstructs overall object structures, including key support components, even under severe sparsity or noise. This highlights the strong generalization across varying sensor modalities and acquisition conditions.

H Additional Qualitative Comparisons on KITTI Dataset

Figures 5 and 6 show qualitative comparisons of state-of-the-art methods at both the scene and single-object levels, respectively.

Our 3DDM generates semantically coherent and structurally complete 3D shapes, even under challenging conditions such as heavy occlusion (e.g. car numbers 2 and 5), limited visible regions (e.g. car numbers 1, 7 and 8), or significant variations in car scales. In contrast, PoinTr and AnchorFormer often suffer from fragmented structures. Notably, PoinTr has a severe problem on the scales of completion results for unseen real-world scenes, making it difficult to faithfully reconstruct real-world object geometries, as shown in Figure 6.

Importantly, our 3DDM achieves completion results without any task-specific adaptation to KITTI data. It is directly transferred from training on the ShapeNet dataset, yet still exhibits robust performance and strong adaptability under sparse and previously unseen data. This further demonstrates the excellent cross-domain generalization capability of our approach.

I Ablation Studies

Experiments	F1-S ↑ / F1-M ↑ / F1-H ↑	CD-S ↓ / CD-M ↓ / CD-H ↓
MSE loss	0.68 / 0.66 / 0.61	1.05 / 1.07 / 1.27
Isotropic noise (w/o phy)	0.67 / 0.65 / 0.60	1.08 / 1.11 / 1.29
Simpler aniso-noise (w/o phy)	0.79 / 0.76 / 0.68	0.74 / 0.77 / 1.02
3DDM(Ours)(w/ phy)	0.83 / 0.82 / 0.73	0.64 / 0.66 / 0.86

Table 3: Ablation analyses of MSE loss, isotropic noise (without physics) and simpler anisotropic noise (without physics).

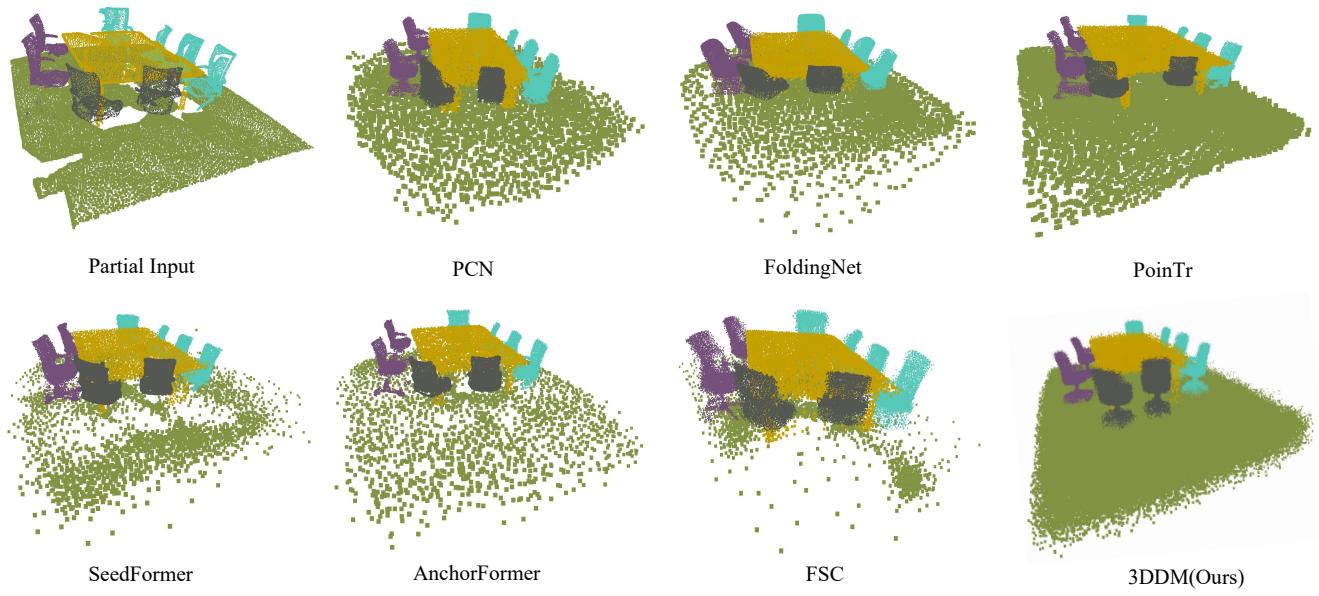


Figure 3: Qualitative completion results on the unseen real-world ScanNet that was not trained.

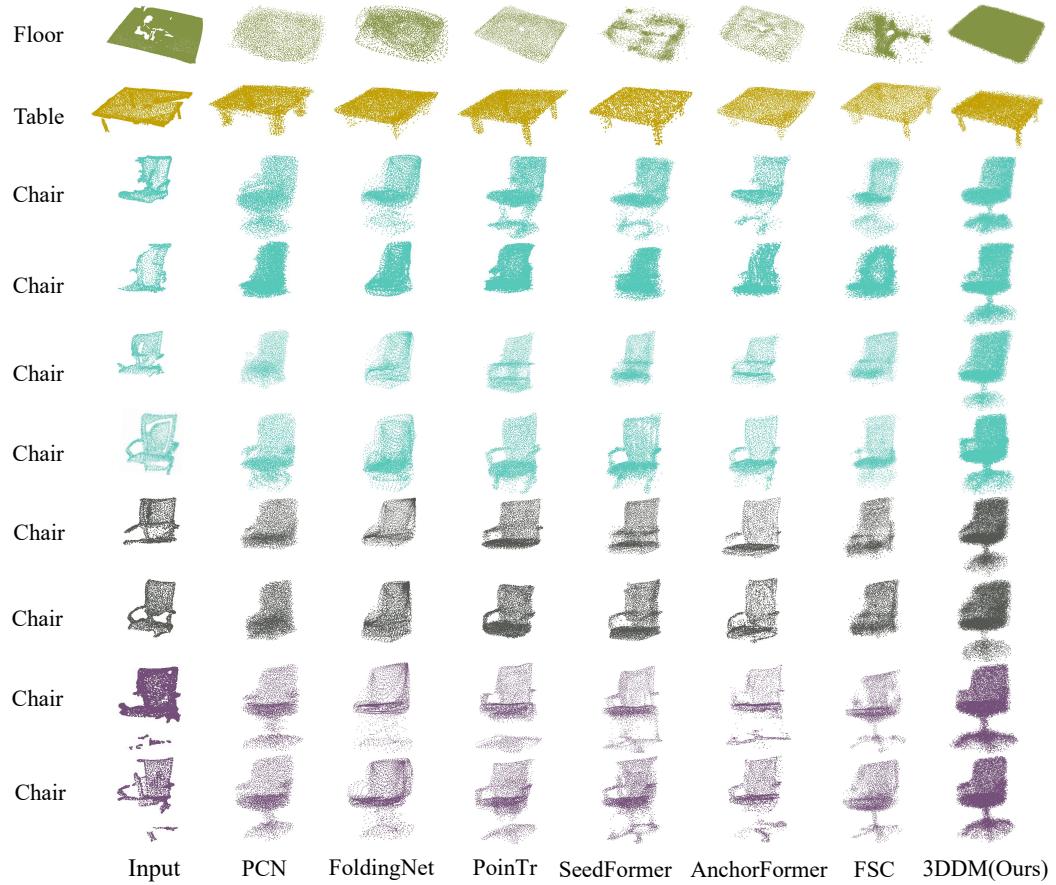


Figure 4: Details of completion results in Figure 3. Most state-of-the-art methods do not generate the missing legs of chairs on untrained data. Our 3DDM successfully generate these missing legs.

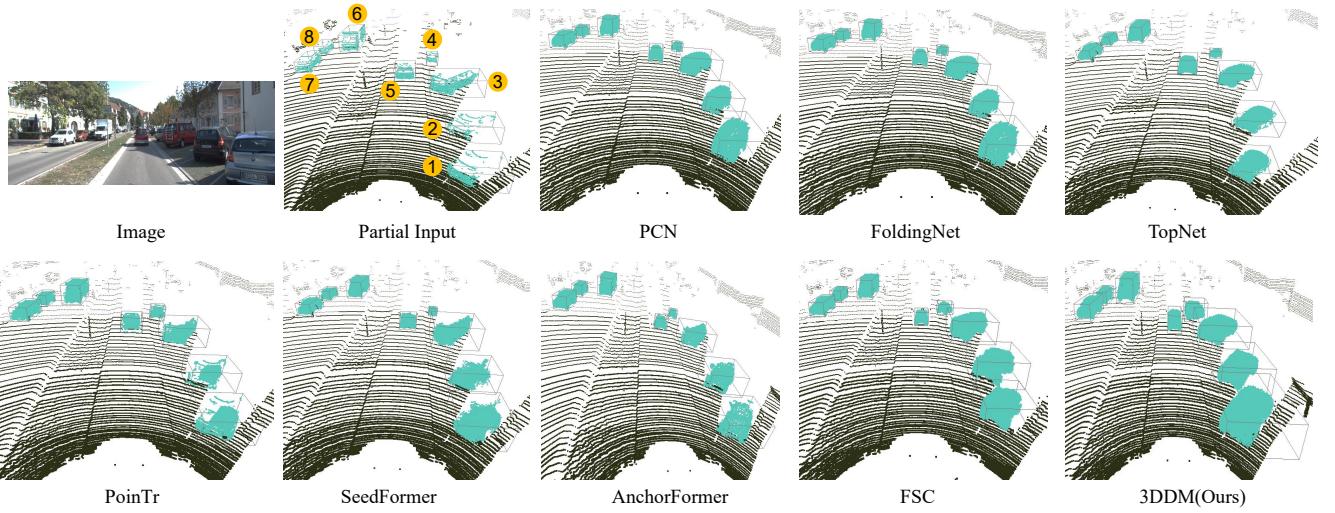


Figure 5: Qualitative completion results on the unseen real-world KITTI cars that were not trained.

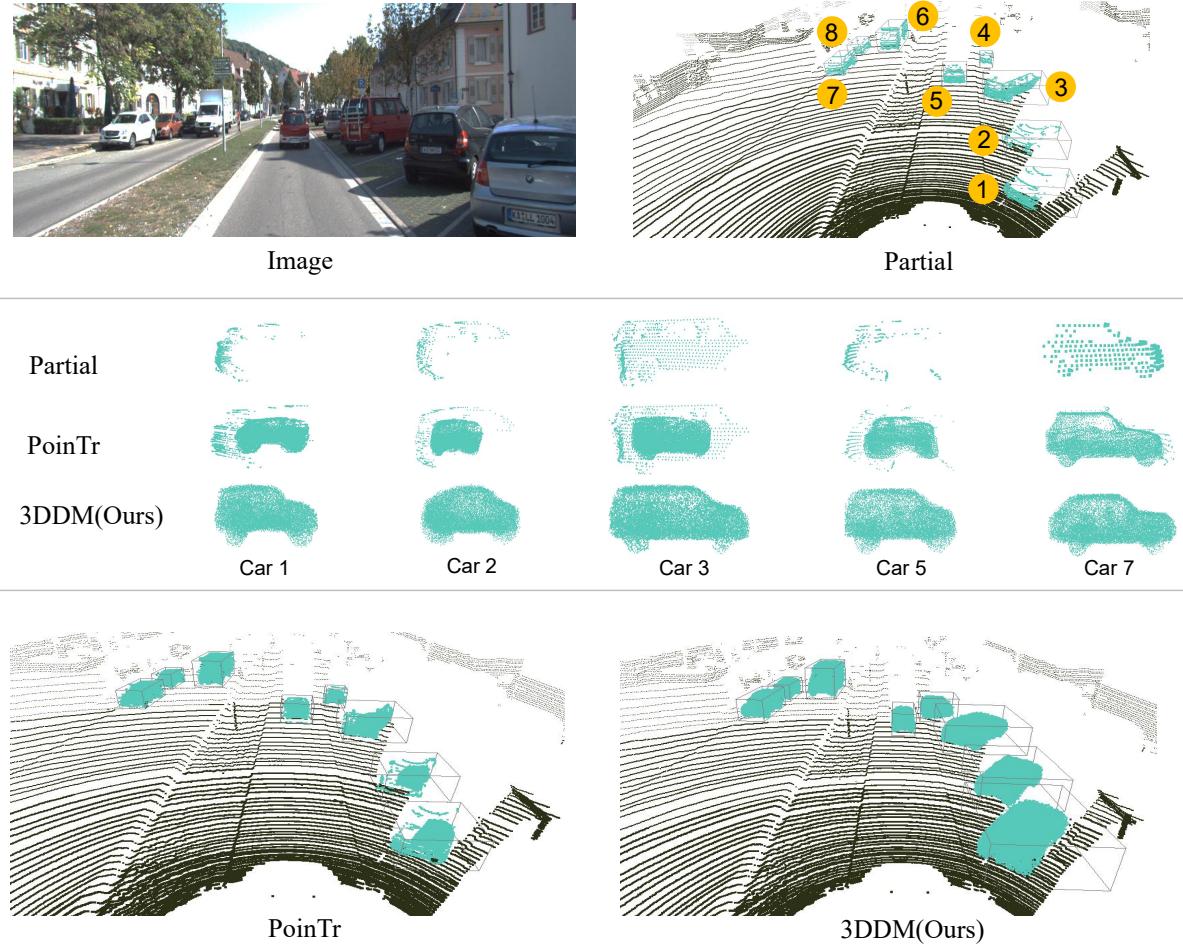


Figure 6: Details of completion results in Figure 5. PoinTr has a severe problem on the scales of completion results for unseen real-world scenes, while our 3DDM generates more faithful and accurate results in both scales and geometry.

We further conduct ablation studies to analyze the MSE loss, isotropic noise (without physics) and simpler anisotropic noise (without physics) with ours 3DDM. Our loss improves CD by 0.4 over MSE by learning per-axis denoising that MSE ignores. Compared to isotropic noise, our anisotropic noise improves CD by 0.43. Isotropies enforce equal x/y/z noise, missing anisotropic 3D dynamics. Our physical anisotropy improves CD by 0.16 over simpler anisotropic noise, as our physical 3D dynamics add excellent interpretability, unlike simpler schemes. The results are reported in Table 3. Our 3DDM learns physical anisotropic noise through thermodynamic evolution, improving accuracy and interpretability.

References

- Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; Xiao, J.; Yi, L.; and Yu, F. 2015. ShapeNet: An Information-Rich 3D Model Repository. Technical report, Stanford University — Princeton University — Toyota Technological Institute at Chicago.
- Chen, Z.; Long, F.; Qiu, Z.; Yao, T.; Zhou, W.; Luo, J.; and Mei, T. 2023. Anchorformer: Point cloud completion from discriminative nodes. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13581–13590.
- Dai, A.; Chang, A. X.; Savva, M.; Halber, M.; Funkhouser, T.; and Nießner, M. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5828–5839.
- Geiger, A.; Lenz, P.; and Urtasun, R. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3354–3361.
- Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 652–660.
- Tchapmi, L. P.; Kosaraju, V.; Rezatofighi, H.; Reid, I.; and Savarese, S. 2019. Topnet: Structural point cloud decoder. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 383–392.
- Wu, X.; Wu, X.; Luan, T.; Bai, Y.; Lai, Z.; and Yuan, J. 2024. FSC: Few-point Shape Completion. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 26077–26087.
- Yang, Y.; Feng, C.; Shen, Y.; and Tian, D. 2018. FoldingNet: Point Cloud Auto-encoder via Deep Grid Deformation. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Yu, X.; Rao, Y.; Wang, Z.; Liu, Z.; Lu, J.; and Zhou, J. 2021. Pointr: Diverse point cloud completion with geometry-aware transformers. In *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12498–12507.
- Yuan, W.; Khot, T.; Held, D.; Mertz, C.; and Hebert, M. 2018. Pcn: Point completion network. In *In International Conference on 3D Vision*, 728–737. IEEE.
- Zhou, H.; Cao, Y.; Chu, W.; Zhu, J.; Lu, T.; Tai, Y.; and Wang, C. 2022. Seedformer: Patch seeds based point cloud completion with upsample transformer. In *In European Conference on Computer Vision*, 416–432. Springer.