Appendix

A. Training and Inference Details

We train the generative model PVD [89] with a batch size of 128 for 10k iterations, and adopt Adam optimizer with learning rate 2×10^{-4}. For *airplane*, we set $\beta_0 = 10^{-5}$ and $\beta_T = 0.008$. For other categories, we set $\beta_0 = 10^{-4}$ and $\beta_T = 0.01$.

In terms of reconstruction model PC^2/ CCD-3DR, we train it with a batch size of 16 for 100k iterations, and adopt Adam optimizer with a learning rate increasing from 10^{-5} to 10^{-3} in the first 2k iterations then decaying to 0 in the remaining iterations. For all the categories, we set $\beta_0 = 10^{-5}$ and $\beta_T = 0.008$.

For the training of BDM-merging, we adopt a similar strategy as the reconstruction model, except that we reduce the total iterations to 20k and scale the learning rate schedule accordingly.

During inference, we set the number of denoising steps as 1000. We divide the denoising process into three distinct stages, *i.e*. early (timesteps 1000–872), middle (timesteps 872–128), and late (timesteps 128–0). We conduct our Bayesian denoising steps in the early and late stages. To be more specific, every 32 timesteps, a Bayesian denoising step is executed for a duration of 16 timesteps. Subsequently, we forward a standard reconstruction process for 16 timesteps, followed by another Bayesian denoising step.

B. Extended Object Categories

Leveraging PVD as our prior model, we follow its settings and adopt official pre-trained weights, which are only on three categories. Each category is trained with different hyperparameters. To further illustrate the effectiveness of BDM, we add two new categories of ShapeNet-R2N2 in Tab. A.1.

Method	Sofa						Table					
	10%		50%		100%		10%		50%		100%	
	CD↓	F1↑	CD↓	F1↑	CD↓	F1↑	CD↓	F1↑	CD↓	F1↑	CD↓	F1↑
CCD-3DR	63.20	0.444	47.83	0.482	43.16	0.501	79.41	0.523	77.51	0.520	67.13	0.538
BDM-merging	62.29	0.460	45.18	0.500	41.43	0.517	78.25	0.535	75.94	**0.538**	65.24	**0.560**
BDM-blending	**61.54**	**0.471**	**44.31**	**0.516**	**41.94**	**0.520**	**74.18**	**0.547**	**73.46**	0.526	**64.56**	0.557

Table A.1. Results on two additional categories of ShapeNet-R2N2, *i.e*. *sofa* and *table*.

C. Alternative Prior Model

To validate the robustness of our proposed BDM, we explore its performance with alternative generative priors. Specifically, we replace the PVD-based generative diffusion model [89] with DiT-3D [54], an extension of DiT [58], which uniquely applies the denoising process to voxelized point clouds. Unlike the PVCNN architecture [47], DiT-3D leverages a Transformer-based framework, and therefore we only experiment on BDM-blending. For proof-of-concept purposes, we take 10% of the chairs on ShapeNet as the training data for the reconstruction model (CCD-3DR). As demonstrated in Tab. A.2, BDM brings consistent improvement regardless of the prior model utilized. Also, we show some qualitative visualizations in Fig. A.1. The results illustrate the effectiveness of our BDM across different generative diffusion priors.

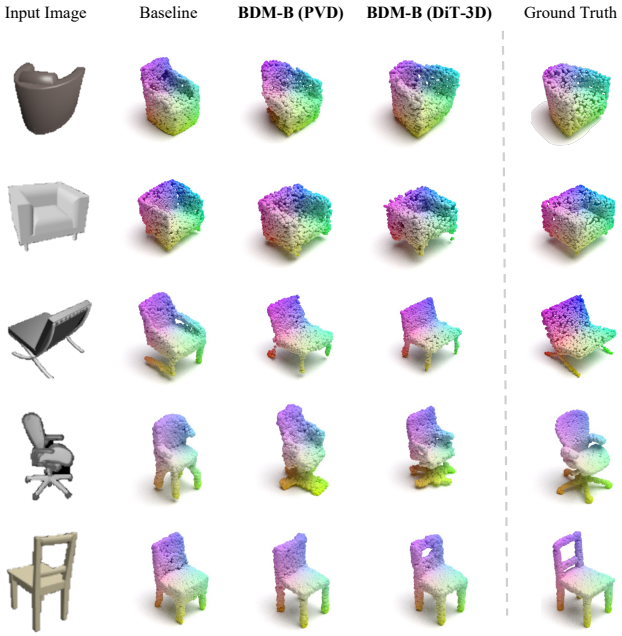

Figure A.1. Visualization of taking DiT-3D as prior compared with PVD on chairs.

	CD ↓	F1 ↑
CCD-3DR (baseline)	89.79	0.418
+ PVD [89]	79.26	0.441
+ DiT-3D [54]	80.77	0.431

Table A.2. Reconstruction results of taking DiT-3D as prior compared with PVD, evaluated with Chamfer Distance and F-Score@0.01.

D. Different Gaussian Noises

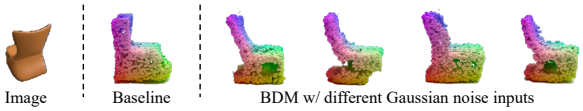

Figure A.2. Predictions over different initial Gaussian noise.

One key advantage of Bayesian methods is that it allows to obtain distributions over predicted outputs, allowing to measure prediction uncertainties. The uncertainties in diffusion model originate from initial Gaussian noise. To investigate the correlation between the outputs and different

initial Gaussian noise $\mathbf{X}_T$, we evaluate BDM-blending with 10 different initial noise inputs, and report mean and variance in Tab. A.3. In addition, we visualize a chair sample in Fig. A.2, featuring consistent valid reconstructions across different noise inputs despite minor shape differences. BDM transforms the vertical chair-back to a tilted and curved one, better aligned with the 2D image.

	CD ↓	F1 ↑
CCD-3DR	89.79	0.418
$\mathbf{X}_T$ ablate	**79.61** (0.048)	**0.441** (2.2e-5)

Table A.3. Mean and variance w.r.t. different initial Gaussian noise.

E. Details of Human Evaluation

For each generated 3D point cloud, we render multiple images of it, as shown in Fig. A.3.

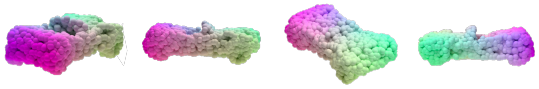

Figure A.3. Multi-view rendering.

From the evaluation results, we can see our BDM-M and BDM-B both outperform CCD-3DR, while the quality of BDM-M is more favored than BDM-B. However, this superiority of BDM-M is not evident from the CD and F1.

As discussed in [67, 79], CD is susceptible to mismatched local density and F1 does not fully address such issue. These metrics may not align with human preference, as shown in Fig. A.4. Therefore, according to the human evaluation, BDM-M yields more visually appealing outcomes whereas blending has stronger quantitative results, which confirms the effectiveness of our BDM-M approach.

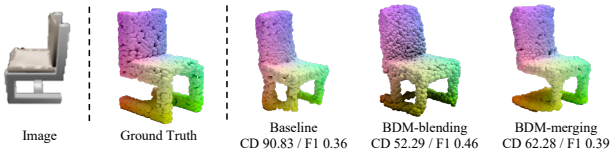

Figure A.4. Quantitative results vs actual visualization.