A. Details of the Derivations

From Eq. (1) to Eq. (2). As in [90], we introduce a latent oracle O defined on the whole denoising chain $\boldsymbol{x}_{0:T}$, such that:

$$o(\boldsymbol{x}_0) = \mathbb{E}_{p_\theta(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[O(\boldsymbol{x}_{0:T})\right]. \tag{8}$$

Then, starting from Eq. (1), we have:

$$\begin{aligned}
&\max_\theta \mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_0\sim p_\theta(\boldsymbol{x}_0|I)}\left[o(\boldsymbol{x}_0)\right] - \beta\mathbb{D}_{\mathrm{KL}}\left[p_\theta(\boldsymbol{x}_0|I)\|p_{\mathrm{ref}}(\boldsymbol{x}_0|I)\right] \\
\geq &\max_\theta \mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_0\sim p_\theta(\boldsymbol{x}_0|I)}\left[o(\boldsymbol{x}_0)\right] - \beta\mathbb{D}_{\mathrm{KL}}\left[p_\theta(\boldsymbol{x}_{0:T}|I)\|p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)\right] \\
= &\max_\theta \mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_{0:T}\sim p_\theta(\boldsymbol{x}_{0:T}|I)}\left[O(\boldsymbol{x}_{0:T})\right] - \beta\mathbb{D}_{\mathrm{KL}}\left[p_\theta(\boldsymbol{x}_{0:T}|I)\|p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)\right] \\
= &\beta\max_\theta \mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_{0:T}\sim p_\theta(\boldsymbol{x}_{0:T}|I)}\left[\log Z(I) - \log\frac{p_\theta(\boldsymbol{x}_{0:T}|I)}{p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)\exp(O(\boldsymbol{x}_{0:T})/\beta)/Z(I)}\right],
\end{aligned} \tag{9}$$

where $Z(I) = \sum_{\boldsymbol{x}_{0:T}} p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)\exp(O(\boldsymbol{x}_{0:T})/\beta)$ is a normalizing factor independent of θ. Since

$$\mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_{0:T}\sim p_\theta(\boldsymbol{x}_{0:T}|I)}\left[\log\frac{p_\theta(\boldsymbol{x}_{0:T}|I)}{p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)\exp(O(\boldsymbol{x}_{0:T})/\beta)/Z(I)}\right] = \mathbb{D}_{\mathrm{KL}}\left[p_\theta(\boldsymbol{x}_{0:T}|I)\|p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)\exp(O(\boldsymbol{x}_{0:T})/\beta)/Z(I)\right] \geq 0 \tag{10}$$

with equality if and only if the two distributions are identical, the optimal $p^\star_\theta(\boldsymbol{x}_{0:T}|I)$ of the right-hand side of Eq. (9) has a unique closed-form solution:

$$p^\star_\theta(\boldsymbol{x}_{0:T}|I) = p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)\exp(O(\boldsymbol{x}_{0:T})/\beta)/Z(I). \tag{11}$$

Therefore,

$$O(\boldsymbol{x}_{0:T}) = \beta\log Z(I) + \beta\log\frac{p^\star_\theta(\boldsymbol{x}_{0:T}|I)}{p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)} \tag{12}$$

for any $I \in \mathrm{supp}(\mathcal{I})$.

We can then obtain Eq. (2) by plugging Eq. (12) into Eq. (8).

From Eq. (4) to Eq. (5). Since sampling from $p_\theta(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0, I)$ is intractable, we follow [90] and replace it with $q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)$:

$$\begin{aligned}
\mathcal{L}_{\mathrm{DRO}} &:= \min \mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_0\sim\mathcal{X}_I,\boldsymbol{x}_{1:T}\sim q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[(1-2o(\boldsymbol{x}_0))\log\frac{p_\theta(\boldsymbol{x}_{0:T}|I)}{p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}|I)}\right] \\
&= \min \mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_0\sim\mathcal{X}_I,\boldsymbol{x}_{1:T}\sim q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[(1-2o(\boldsymbol{x}_0))\sum_{t=1}^{T}\log\frac{p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, I)}{p_{\mathrm{ref}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, I)}\right] \\
&= \min T\mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_0\sim\mathcal{X}_I,t\sim\mathcal{U}(0,T),\boldsymbol{x}_t\sim q(\boldsymbol{x}_t|\boldsymbol{x}_0),\boldsymbol{x}_{t-1}\sim q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0,\boldsymbol{x}_t)}\left[(1-2o(\boldsymbol{x}_0))\log\frac{p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, I)}{p_{\mathrm{ref}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, I)}\right] \\
&= \min T\mathbb{E}_{I\sim\mathcal{I},\boldsymbol{x}_0\sim\mathcal{X}_I,t\sim\mathcal{U}(0,T),\boldsymbol{x}_t\sim q(\boldsymbol{x}_t|\boldsymbol{x}_0)}\Bigg[(1-2o(\boldsymbol{x}_0))\Big(\\
&\quad \mathbb{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)\|p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, I)\right] - \mathbb{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)\|p_{\mathrm{ref}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, I)\right]\Big)\Bigg].
\end{aligned} \tag{13}$$

Recall that for diffusion models p_θ and p_{ref}, the distributions $q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)$, $p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, I)$ and $p_{\mathrm{ref}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, I)$ are all Gaussian. Therefore, the KL divergence on the right-hand side of Eq. (13) can be re-parameterized analytically using $\boldsymbol{\epsilon}_\theta$. After some algebra, and removing all terms independent of θ, this yields Eq. (5).

B. Additional Training Details

All hyperparameters are listed in Tab. 4. We did *not* extensively tune these parameters: the LoRA parameters and the β used in $\mathcal{L}_{\mathrm{DPO}}$ follow [43], and the rectified flow noise level t sampling uses the distribution from TRELLIS [101].

Loss formulation	$\mathcal{L}_{\text{DRO}}$	$\mathcal{L}_{\text{DPO}}$
Optimization		
Optimizer	AdamW	AdamW
Learning rate	5×10^{-6}	5×10^{-6}
Learning rate warmup	Linear 2,000 iterations	Linear 2,000 iterations
Weight decay	0.01	0.01
Effective batch size	48	48
Training iterations	4,000	8,000
Precision	`bf16`	`bf16`
LoRA		
Rank	64	64
α	128	128
Dropout	0	0
Miscellaneous		
Rectified flow t sampling	$\text{LogitNorm}(1, 1)$	$\text{LogitNorm}(1, 1)$
β in $\mathcal{L}_{\text{DPO}}$	—	500

Table 4. DSO training details and hyperparameter settings.

Method	Alarm clock		Motorcycle	
	% Stable↑	Rot.↓	% Stable↑	Rot.↓
TRELLIS [101]	67.5	14.14°	44.4	46.53°
TRELLIS + DSO	**85.0**	**5.58°**	**58.1**	**36.75°**

Table 5. DSO enhances the model's ability to generate assets that remain stable under gravity from in-the-wild images of stable objects.

C. Additional Evaluation Details

For evaluation, the 3D models are generated by TRELLIS [101] and DSO fine-tuned TRELLIS using the default setting: 12 sampling steps in stage 1 with classifier-free guidance 7.5 and 12 sampling steps in stage 2 with classifier-free guidance 3. Under this setting, generating *one* model takes 10 seconds on average on an NVIDIA A100 GPU. By contrast, Atlas3D [7] takes 2 hours to generate a model using SDS and PhysComp [21] takes on average 15 minutes to optimize *one* model output by TRELLIS on our hardware.

We use MuJoCo [88] for rigid body simulation for evaluation. The 3D models are assumed to be rigid and uniform in density. We run the simulation for 10 seconds, at which almost all objects have reached the steady state.

D. Additional Results

D.1. Additional Evaluation Results

To demonstrate that the enhanced physical soundness achieved through DSO is not limited to a specific simulation environment, we report the evaluation results in Isaac Gym [53] and under perturbations in Tab. 6. For the evaluation under perturbations, we choose 4 maximum perturbation angles $\theta_{\max}$ and perform 100 simulation runs with each $\theta_{\max}$ where the generated 3D models are initially rotated by a random angle $\theta \in (-\theta_{\max}, \theta_{\max})$, following Atlas3D [7]. We then report the average stability rate of the 100 runs. In Tab. 6, TRELLIS post-trained with only MuJoCo feedback via DSO outperforms all baselines under all simulation settings, showing that the improved physical soundness generalizes well to different simulation environments.

Method	MuJoCo					Isaac Gym
	w/o perturbation	$\theta_{\max} = 0.01$	$\theta_{\max} = 0.02$	$\theta_{\max} = 0.04$	$\theta_{\max} = 0.08$	w/o perturbation
Full evaluation set (65 objects)						
TRELLIS [101]	85.1	84.8	84.2	82.5	77.2	97.3
Atlas3D [7]	69.4	70.3	70.2	66.3	61.8	88.7
TRELLIS + DSO (w/ $\mathcal{L}_{\text{DPO}}$)	<u>95.1</u>	<u>94.8</u>	<u>94.1</u>	<u>92.6</u>	<u>88.0</u>	<u>99.3</u>
TRELLIS + DSO (w/ $\mathcal{L}_{\text{DRO}}$)	**99.0**	**98.8**	**98.6**	**97.2**	**93.7**	**99.6**
Partial evaluation set (11 unstable objects)						
TRELLIS [101]	54.5	54.0	53.8	48.5	41.5	93.9
TRELLIS + PhysComp [21]	80.3	76.9	76.1	72.6	<u>67.7</u>	83.9
TRELLIS + DSO (w/ $\mathcal{L}_{\text{DPO}}$)	<u>82.6</u>	<u>82.0</u>	<u>80.7</u>	<u>77.5</u>	67.5	<u>98.5</u>
TRELLIS + DSO (w/ $\mathcal{L}_{\text{DRO}}$)	**95.5**	**95.4**	**95.0**	**93.9**	**85.4**	**100.0**

Table 6. **Results** evaluated under different simulation settings.

Figure 7. DSO fine-tuned TRELLIS (**ours**) is more likely to generate physically sound 3D objects when conditioned on *real-world* images of challenging categories.

D.2. Additional Comparison with Post-Processing Baselines

In Tab. 7, we compare DSO with a naive post-processing baseline that cuts the mesh flat just above the lowest vertex, following Atlas3D [7]. This method is less effective at stabilizing meshes and significantly degrades geometric quality, as reflected in the higher Chamfer distance (Tab. 7).

Method	Enforcing flat at height z				DSO (Ours)
	$z = 0.05$	$z = 0.1$	$z = 0.15$	$z = 0.2$	
% Stable	94.2	90.5	93.2	<u>95.8</u>	**99.0**
Chamfer Distance	<u>0.0502</u>	0.0537	0.0591	0.0662	**0.0440**

Table 7. **Comparison** with post-processing baselines.

D.3. Additional Results on In-the-Wild Images

To assess the generalization of DSO fine-tuned models in generating physically sound 3D objects from real-world images, we curate a set of 30 CC-licensed images for each category: stable alarm clocks and motorcycles supported by kickstands. We select these two categories because the base model, TRELLIS, struggles to generate physically stable versions of these objects. The results are reported in Tab. 5, with *randomly sampled* examples visualized in Fig. 7. As is evident, DSO enhances the model's ability to generate assets that remain stable under gravity from in-the-wild images of stable objects.

E. Additional Discussions

A deeper analysis of DRO *vs*. DPO. We further analyze the similarities and differences between $\mathcal{L}_{\text{DRO}}$ and $\mathcal{L}_{\text{DPO}}$. Both losses are monotonic functions of $o = \|\boldsymbol{\epsilon}^w - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^w, t)\|_2^2 - \|\boldsymbol{\epsilon}^w - \boldsymbol{\epsilon}_{\text{ref}}(\boldsymbol{x}_t^w, t)\|_2^2 - \left(\|\boldsymbol{\epsilon}^l - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^l, t)\|_2^2 - \|\boldsymbol{\epsilon}^l - \boldsymbol{\epsilon}_{\text{ref}}(\boldsymbol{x}_t^l, t)\|_2^2\right)$. In Fig. 8, we plot each loss (**left**) and its derivative with respect to o (**right**, log-scale). A key difference is that $\frac{d\mathcal{L}_{\text{DRO}}}{do}$ is constant, while $\frac{d\mathcal{L}_{\text{DPO}}}{do}$ decays exponentially as o decreases. As a result, o tends to plateau during optimization of $\mathcal{L}_{\text{DPO}}$. This leads to faster convergence with $\mathcal{L}_{\text{DRO}}$, although extended training may harm performance.

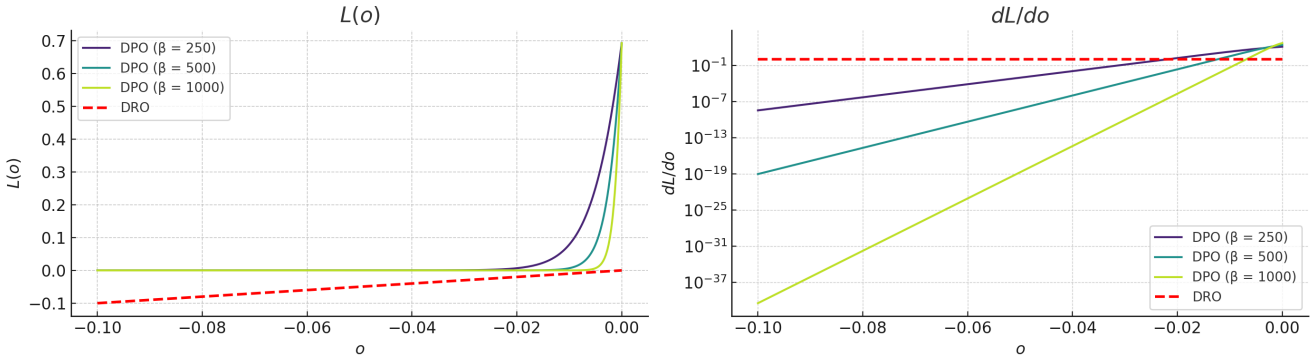

Figure 8. **Plots** of $\mathcal{L}_{\text{DRO}}$ and $\mathcal{L}_{\text{DPO}}$ and their derivatives.

Scaling behaviors when optimizing $\mathcal{L}_{\text{DRO}}$. In Sec. 4.5, we analyzed how DSO scales when optimizing $\mathcal{L}_{\text{DPO}}$. Here, we present the corresponding scaling behavior for $\mathcal{L}_{\text{DRO}}$. As shown in Tab. 8, performance peaks at $4,000$ training steps, after which the geometry quality noticeably degrades—consistent with our earlier analysis. Scaling with training data follows a similar trend to that observed for $\mathcal{L}_{\text{DPO}}$ in Fig. 6b.

Training steps	2000	3000	4000	5000
% Stable	91.5	96.9	**99.0**	98.7
Chamfer D.	0.0473	0.0464	**0.0440**	0.0853

Table 8. **Scaling behavior with training compute** of $\mathcal{L}_{\text{DRO}}$.

F. Limitations and Future Work

DSO's self-improving scheme relies on the base model generating at least some positive samples, and hence may be less effective for base models where such samples are rare. DSO opens up new possibilities for integrating physical constraints into generative models, enhancing their applicability in real-world scenarios where adherence to such constraints is crucial.