CRAFT: Aligning Diffusion Models with Fine-Tuning Is Easier Than You Think

Supplementary Material

A. Detailed Proof of Theorem 3.1

We provide a fully detailed derivation of the proposed lower bound for clarity.

Step 0: Restate the empirical objective.

Recall the empirical Monte Carlo objective:

$$\hat{J}(\theta) = \mathbb{E}_{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\{\boldsymbol{x}_0^{(i)}\}_{i=1}^{G}\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c})}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{p_\theta(\boldsymbol{x}_0^{(i)}|\boldsymbol{c})}{p_{\theta_{\text{old}}}(\boldsymbol{x}_0^{(i)}|\boldsymbol{c})}\cdot\hat{A}_i\right], \tag{5}$$

where

$$\hat{A}_i = \frac{r(\boldsymbol{c},\boldsymbol{x}_0^{(i)}) - \text{mean}\{r(\boldsymbol{c},\boldsymbol{x}_0^{(1)}),\ldots,r(\boldsymbol{c},\boldsymbol{x}_0^{(G)})\}}{\text{std}\{r(\boldsymbol{c},\boldsymbol{x}_0^{(1)}),\ldots,r(\boldsymbol{c},\boldsymbol{x}_0^{(G)})\}+\epsilon}, \quad \frac{1}{G}\sum_{i=1}^{G}\hat{A}_i = 0. \tag{6}$$

Step 1: Express in terms of log-likelihood difference.

Define the log-likelihood difference:

$$\Delta_i(\theta) := \log p_\theta(\boldsymbol{x}_0^{(i)}|\boldsymbol{c}) - \log p_{\theta_{\text{old}}}(\boldsymbol{x}_0^{(i)}|\boldsymbol{c}). \tag{7}$$

Then (5) becomes

$$\hat{J}(\theta) = \mathbb{E}_{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\{\boldsymbol{x}_0^{(i)}\}_{i=1}^{G}\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c})}\left[\frac{1}{G}\sum_{i=1}^{G}\exp\left(\Delta_i(\theta)\right)\cdot\hat{A}_i\right]. \tag{8}$$

Step 2: ELBO form for diffusion models.

Now according to the evidence lower bound (ELBO) of the log-likelihood in diffusion models [6], we have

$$\begin{aligned}
\log p_\theta(\boldsymbol{x}_0^{(i)}|\boldsymbol{c}) &= \log\int p_\theta(\boldsymbol{x}_{0:T}^{(i)}|\boldsymbol{c})\,\mathrm{d}\boldsymbol{x}_{1:T}^{(i)} \\
&= \log\int q(\boldsymbol{x}_{1:T}^{(i)}|\boldsymbol{x}_0^{(i)})\frac{p_\theta(\boldsymbol{x}_{0:T}^{(i)}|\boldsymbol{c})}{q(\boldsymbol{x}_{1:T}^{(i)}|\boldsymbol{x}_0^{(i)})}\mathrm{d}\boldsymbol{x}_{1:T}^{(i)} \\
&\geq \mathbb{E}_{q(\boldsymbol{x}_{1:T}^{(i)}|\boldsymbol{x}_0^{(i)})}\left[\log\frac{p_\theta(\boldsymbol{x}_{0:T}^{(i)}|\boldsymbol{c})}{q(\boldsymbol{x}_{1:T}^{(i)}|\boldsymbol{x}_0^{(i)})}\right] \\
&= -\mathop{\mathbb{E}}_{\substack{t\sim\text{Uniform}(\{1,\ldots,T\})\\ \boldsymbol{\epsilon}_t^{(i)}\sim\mathcal{N}(0,I)}}\left[\underbrace{\frac{1}{2\sigma_t^2}\cdot\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}_{w(t)}\cdot\left\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2\right] + \text{const.}
\end{aligned} \tag{9}$$

So we can simply denote

$$\log p_\theta(\boldsymbol{x}_0^{(i)}|\boldsymbol{c}) = -\mathop{\mathbb{E}}_{\substack{t\sim\text{Uniform}(\{1,\ldots,T\})\\ \boldsymbol{\epsilon}_t^{(i)}\sim\mathcal{N}(0,I)}}\left[w(t)\cdot\left\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2\right] + C_i, \tag{10}$$

where C_i does not depend on θ. Similarly, for the old parameters:

$$\log p_{\theta_{\text{old}}}(\boldsymbol{x}_0^{(i)}|\boldsymbol{c}) = -\mathop{\mathbb{E}}_{\substack{t\sim\text{Uniform}(\{1,\ldots,T\})\\ \boldsymbol{\epsilon}_t^{(i)}\sim\mathcal{N}(0,I)}}\left[w(t)\cdot\left\|\boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2\right] + C_i', \tag{11}$$

Subtracting gives the exact identity:

$$\Delta_i(\theta) = -\mathop{\mathbb{E}}_{\substack{t\sim\text{Uniform}(\{1,\dots,T\})\\ \boldsymbol{\epsilon}_t^{(i)}\sim\mathcal{N}(0,I)}}\left[w(t)\cdot\left(\left\|\boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2-\left\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2\right)\right]+C_i-C_i'. \tag{12}$$

Define

$$M_i^\theta := \mathop{\mathbb{E}}_{\substack{t\sim\text{Uniform}(\{1,\dots,T\})\\ \boldsymbol{\epsilon}_t^{(i)}\sim\mathcal{N}(0,I)}}\left[w(t)\cdot\left\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2\right],\quad M_i^{\theta_{\text{old}}} := \mathop{\mathbb{E}}_{\substack{t\sim\text{Uniform}(\{1,\dots,T\})\\ \boldsymbol{\epsilon}_t^{(i)}\sim\mathcal{N}(0,I)}}\left[w(t)\cdot\left\|\boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2\right]. \tag{13}$$

Then $\Delta_i(\theta) = -M_i^\theta + M_i^{\theta_{\text{old}}} + C_i - C_i'$.

Step 3: Factor out constants.

The term $C_i - C_i'$ does not depend on θ, so

$$\exp\left(\Delta_i(\theta)\right) = \exp\left(C_i - C_i'\right)\cdot\exp\left(-M_i^\theta + M_i^{\theta_{\text{old}}}\right), \tag{14}$$

where $\exp\left(C_i - C_i'\right)$ can be absorbed into a constant $\tilde{C}$. Hence

$$\hat{J}(\theta) = \tilde{C} + \mathbb{E}_{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\{\boldsymbol{x}_0^{(i)}\}_{i=1}^G\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c})}\left[\frac{1}{G}\sum_{i=1}^G\exp\left(-M_i^\theta + M_i^{\theta_{\text{old}}}\right)\cdot\hat{A}_i\right]. \tag{15}$$

Step 4: Taylor expansion via small learning rate.

Assume $\theta = \theta_{\text{old}} + \eta g$, with $\eta \to 0$. The noise predictor $\boldsymbol{\epsilon}_\theta$ is a smooth function of θ, so we can do a first-order Taylor expansion, i.e., $f(\theta_{\text{old}} + \eta g) = f(\theta_{\text{old}}) + \eta\nabla_\theta f(\theta_{\text{old}})\cdot g + O(\eta^2)$, we have

$$\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c}) = \boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c}) + \eta\nabla_\theta\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})|_{\theta_{\text{old}}}\cdot g + O(\eta^2). \tag{16}$$

Thus the difference

$$\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c}) - \boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c}) = O(\eta). \tag{17}$$

Consequently, the squared error inside M_i^θ can be expanded:

$$\begin{aligned}
&\left\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2\\
=&\left\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})+\boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2\\
=&\left\|\boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right\|^2+2\left(\boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_t^{(i)}\right)^\top\left(\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})-\boldsymbol{\epsilon}_{\theta_{\text{old}}}(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c})\right)+O(\eta^2).
\end{aligned} \tag{18}$$

Taking expectation over t and $\boldsymbol{\epsilon}_t^{(i)}$ gives

$$M_i^\theta = M_i^{\theta_{\text{old}}} + O(\eta) \Longrightarrow M_i^\theta - M_i^{\theta_{\text{old}}} = O(\eta), \tag{19}$$

justifying the linearization:

$$\exp\left(-M_i^\theta + M_i^{\theta_{\text{old}}}\right) = \exp\left(-\left(M_i^\theta - M_i^{\theta_{\text{old}}}\right)\right) = 1 - \left(M_i^\theta - M_i^{\theta_{\text{old}}}\right) + O(\eta^2). \tag{20}$$

Step 6: Substitute back $\exp\left(-M_i^\theta + M_i^{\theta_{\text{old}}}\right)$ in terms of diffusion MSE.

Finally,

$$\begin{aligned}
\hat{J}(\theta) &= \tilde{C} + \mathbb{E}_{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\{\boldsymbol{x}_0^{(i)}\}_{i=1}^G\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c})}\left[\frac{1}{G}\sum_{i=1}^G \exp\left(-M_i^\theta + M_i^{\theta_{\text{old}}}\right)\cdot\hat{A}_i\right] \\
&= \tilde{C} + \mathbb{E}_{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\{\boldsymbol{x}_0^{(i)}\}_{i=1}^G\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c})}\left[\frac{1}{G}\sum_{i=1}^G \left(1 - \left(M_i^\theta - M_i^{\theta_{\text{old}}}\right) + O(\eta^2)\right)\cdot\hat{A}_i\right] \\
&= \tilde{C} + \underbrace{\mathbb{E}_{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\{\boldsymbol{x}_0^{(i)}\}_{i=1}^G\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c})}\left[\frac{1}{G}\sum_{i=1}^G \hat{A}_i\right]}_{=0} + \underbrace{\mathbb{E}_{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\{\boldsymbol{x}_0^{(i)}\}_{i=1}^G\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c})}\left[\frac{1}{G}\sum_{i=1}^G M_i^{\theta_{\text{old}}}\cdot\hat{A}_i\right]}_{\text{constant}} \\
&\quad - \mathbb{E}_{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\{\boldsymbol{x}_0^{(i)}\}_{i=1}^G\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c})}\left[\frac{1}{G}\sum_{i=1}^G M_i^\theta\cdot\hat{A}_i\right] + O(\eta^2),
\end{aligned} \tag{21}$$

substituting the definition of M_i^θ, we have

$$\hat{J}(\theta) \geq C - \mathop{\mathbb{E}}_{\substack{\boldsymbol{c}\sim\mathcal{D}_{\boldsymbol{c}},\\ \{\boldsymbol{x}_0^{(i)}\}_{i=1}^G\sim p_{\theta_{\text{old}}}(\cdot|\boldsymbol{c}),\\ t\sim\text{Uniform}(\{1,\ldots,T\})\\ \boldsymbol{\epsilon}_t^{(i)}\sim\mathcal{N}(0,I)}}\left[\frac{1}{G}\sum_{i=1}^G \hat{A}_i w(t)\cdot\left\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t^{(i)},t,\boldsymbol{c}) - \boldsymbol{\epsilon}_t^{(i)}\right\|^2\right]. \tag{22}$$

This completes the proof.

B. Supplementary Experimental Results

Evaluation Bias on PicScore. In this work, we exclude PickScore from our primary evaluation metrics to ensure a fair and unbiased comparison. Most state-of-the-art baselines (e.g., Diff-DPO, SPO, SmPO) are explicitly optimized using the Pick-a-Pic dataset. Since the PickScore evaluator is trained on the exact same preference distribution, using it to evaluate these methods introduces significant **in-domain bias**, where high scores may reflect dataset overfitting rather than generalized generation quality. To avoid this data leakage and demonstrate the robustness of our method, we adopt a comprehensive suite of four independent metrics: **HPSv2.1**, **ImageReward**, **Aesthetic Score (AES)**, and **MPS**. These evaluators are derived from diverse data sources distinct from the baselines' training sets, providing a neutral ground for comparison. By validating our approach across this multi-dimensional framework, we ensure that the reported performance reflects genuine improvements in human alignment and aesthetic quality, fully substantiating the effectiveness of our method without relying on potentially biased indicators.

Geneval Results on SDXL. We extend our evaluation to the larger SDXL architecture to verify the scalability and effectiveness of our method. As presented in Table 12, **CRAFT** achieves the highest 'Overall' score of **57.97**, surpassing the base SDXL model (55.05) as well as recent alignment baselines such as Diff-DPO (57.23) and SmPO (57.86). Notably, CRAFT demonstrates superior capability in ensuring object presence and spatial fidelity. It secures state-of-the-art results in the 'Single Object' (**99.06**) and 'Two Object' (**86.36**) categories. The significant margin in the 'Two Object' metric, outperforming the second-best method by nearly 6 points, highlights CRAFT's robustness in complex multi-subject generation scenarios. Furthermore, CRAFT also leads in the 'Position' category with a score of **14.50**, while maintaining highly competitive performance in attribute binding and color consistency. These results confirm that CRAFT effectively enhances the controllability and compositional alignment of large-scale text-to-image models.

More Visulization. Figure 8 provide additional visually appealing samples generated by CRAFT-SDXL, showcasing both its general aesthetic superiority and its specific structural integrity. Consistent with the main text, the prompts for these images are sourced from the HPDv2, Parti-Prompt, and Pick-a-Pic test sets, affirming performance across diverse domains. Figure 6 includes results from ControlNet, demonstrating CRAFT's superior control fidelity when guided by conditioning inputs

Figure 6. **ControlNet Visualization: Superior Control Fidelity and Data Efficiency of CRAFT.** Qualitative comparison using diverse ControlNet conditions (**Canny** and **Depth**) against existing fine-tuning methods (SDXL, Diff-DPO, MaPO, SmPO). CRAFT consistently demonstrates strong geometric control fidelity and generates visually superior images across all modalities. This high robustness and aesthetic quality are achieved despite utilizing significantly fewer training samples, highlighting CRAFT's superior data efficiency and generalization capability.

Table 7. **End-to-end cost in H100 hours.** For CRAFT, we report both preprocessing and fine-tuning cost. Even with candidate generation and filtering included, CRAFT remains substantially more efficient than prior methods.

SDXL			SD1.5		
Method	**Hours**	**Ratio**	**Method**	**Hours**	**Ratio**
Diff-DPO	∼638.1	26.5×	Diff-DPO	∼80.6	9.2×
MaPO	∼545.6	22.6×	Diff-KTO	∼415.6	47.2×
CRAFT	∼(4.0 + 20.1)	**1.0**×	**CRAFT**	∼(1.9 + 6.9)	**1.0**×

(such as Canny and Depth maps). These samples confirm the model's robustness: the high aesthetic quality is consistently maintained even when the geometric structure is strictly constrained by external control signals, which is a key indicator of model generalization beyond simple style transfer.

End-to-End Efficiency. To provide a fairer efficiency comparison, we report the end-to-end cost of our pipeline, including candidate generation, reward-based filtering, and the final fine-tuning stage. As shown in Table 7, CRAFT remains substantially more efficient than existing baselines even after including preprocessing overhead. This is important because our method does not assume access to large-scale preference pairs; instead, it constructs a compact training set from prompts with a comparatively light self-curation stage.

Additional Preference-Optimization Baselines. We further compare CRAFT against recent preference-optimization baselines. Table 8 reports score-based comparisons against InPO using its public checkpoint. Table 9 summarizes the comparison against DSPO [48] using the reward win-rate statistics reported in its original paper, since official checkpoints are unavailable. Across all reported datasets and metrics, CRAFT remains consistently stronger.

Table 8. **Comparison with InPO on SDXL.**

Model	HPDv2			Parti-Prompt			Pick-a-Pic		
	HPS	AES	IR	HPS	AES	IR	HPS	AES	IR
InPO	30.79	5.837	1.058	29.27	5.756	1.024	30.30	5.870	0.978
CRAFT	**32.67**	**6.031**	**1.312**	**31.10**	**5.976**	**1.252**	**32.18**	**6.080**	**1.308**

Table 9. **Comparison with DSPO on SDXL.**

Model	HPDv2			Parti-Prompt			Pick-a-Pic		
	HPS	AES	IR	HPS	AES	IR	HPS	AES	IR
DSPO	83.47	51.41	70.09	81.80	57.84	73.47	80.00	54.20	68.60
CRAFT	**97.84**	**73.00**	**84.38**	**94.73**	**81.50**	**80.70**	**96.40**	**74.00**	**86.00**

Table 10. **Vanilla SFT vs. CRAFT.**

Method	HPS	AES	IR
Vanilla SFT	30.91	6.040	1.110
CRAFT	**32.18**	**6.080**	**1.308**

Table 11. **T2I-CompBench.**

Model	Attr. Bind.			Obj. Rela.		Complex
	C	S	T	Sp.	N-Sp.	
SD1.5	0.384	0.376	0.407	0.102	0.309	0.391
CRAFT	**0.495**	**0.446**	**0.456**	**0.139**	**0.311**	**0.405**

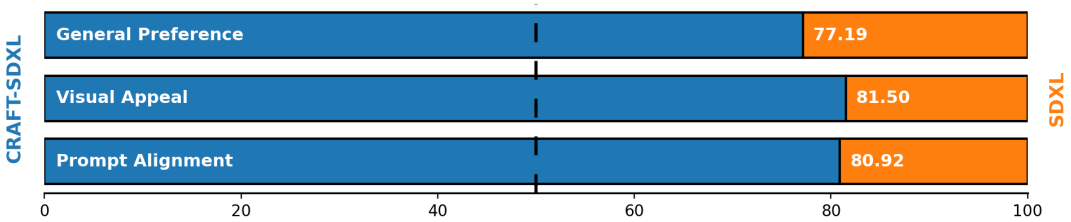

Figure 7. **User study results.** Human evaluation corroborates the automatic reward-based metrics and favors CRAFT over the compared baselines.

Isolating the Effect of Advantage Weighting. To separate the contribution of data selection from the contribution of the training objective, we train a vanilla SFT baseline on the exact same Top-100 filtered dataset used by CRAFT. Table 10 shows that advantage weighting yields clear gains over vanilla SFT, especially on HPS and ImageReward, demonstrating that the performance gain does not come solely from data filtering.

Human Evaluation and Semantic Faithfulness. Since pretrained reward models would be biased, we additionally report a user study in Figure 7. The human preference trend is consistent with our automatic evaluations, providing complementary evidence that the observed gains are not merely artifacts of the reward models. We also report T2I-CompBench in Table 11, which supports that prompt refinement does not degrade compositional or semantic faithfulness.

Prompt Refinement, Reward Balancing, and Dataset Diversity. We use Qwen-Plus to rewrite original prompt into several more descriptive variants, which are used only for candidate generation. Importantly, all reward computation and filtering are performed with respect to the original prompt, so candidates with semantic drift are naturally disfavored during selection. For composite reward filtering, we first map the three reward signals to comparable scales before applying the final weights $(0.4, 0.4, 0.2)$ for HPS, PickScore, and AES. These coefficients should therefore be interpreted as relative importance weights rather than raw-scale coefficients. The ablation results in Table 6 indicate that the full combination is the most reliable in practice, while Table 5 shows that a compact Top-100 subset is sufficient for strong performance. In the final selected set, we did not observe exact duplicate prompts, and the strong performance across HPDv2, Parti-Prompt, Pick-a-Pic, GenEval, and T2I-CompBench suggests that the few-shot setting does not bias the model toward a narrow prompt pattern.

C. Discussion

Limitations. Despite the promising results achieved by CRAFT, we acknowledge two primary limitations in our current framework. First, regarding the algorithmic paradigm, CRAFT fundamentally operates as an offline Supervised Fine-Tuning approach. While efficient, it relies on a static dataset constructed prior to training. This formulation inherently limits the model's data utilization efficiency and exploration capability compared to online Reinforcement Learning (RL) strategies (e.g., PPO or the recent GRPO), which can dynamically explore the generation space and potentially reach a higher performance ceiling. Second, regarding the data construction strategy, our current framework still selects training samples at the image level after generation, rather than explicitly modeling the quality of the initial noise itself. Recent studies suggest that

Table 12. **GenEval results** of our **CRAFT** and baseline methods for SDXL. The highest value is shown in bold, the second highest is underlined, and models marked with * are reproduced strictly following the official code and datasets.

Method	Single Object	Two Object	Counting	Colors	Position	Attribute Binding	Overall
SDXL	97.50	70.96	42.81	**88.30**	11.00	21.00	55.05
Diff-DPO	<u>98.75</u>	<u>80.56</u>	**45.62**	86.70	11.00	20.75	57.23
SmPO*	<u>98.75</u>	79.55	<u>44.69</u>	86.70	10.50	**27.00**	<u>57.86</u>
SPO*	97.81	80.05	38.44	84.04	<u>11.75</u>	20.50	55.43
CRAFT (Ours)	**99.06**	**86.36**	36.88	<u>87.23</u>	**14.50**	<u>23.75</u>	**57.97**

high-quality or prompt-aware noise can be selected, optimized, or learned to consistently improve sample quality [22, 47]. Incorporating such noise-aware criteria into our reward filtering pipeline is a promising yet currently unexplored direction. Third, regarding the application scope, our current experimental validation is exclusively confined to Text-to-Image (T2I) generation. Although the principles of preference-free alignment are theoretically universal, we have not yet addressed the unique challenges present in other modalities, such as the temporal consistency required for Text-to-Video (T2V) or the geometric constraints essential for Text-to-3D generation.

Future Directions. Building upon these observations, our future work will focus on three strategic expansions. To transcend the offline limitations, we aim to evolve CRAFT into an *online* learning framework. By implementing an iterative "generate-train" loop, we can continuously update the training data with the model's own evolving distributions. This on-policy approach will ensure the model receives sustained and increasingly precise gradient signals throughout the training process, bridging the gap between SFT and RL. We also plan to enrich our filtering stage with adaptive noise selection criteria, where the filtering form and quality standards can vary across prompts, models, or downstream objectives, potentially benefiting from learned noise priors and reflection-based sampling signals [2, 2, 22, 47]. Finally, to broaden the application scope, we plan to adapt the CRAFT algorithm to a wider array of generative tasks. Specifically, we will investigate how our composite reward filtering and fine-tuning can be integrated into T2V and T23D pipelines, exploring whether the robust alignment capabilities demonstrated in 2D images can effectively generalize to complex temporal and spatial dimensions. In particular, recent video inference and distillation techniques, such as weak-to-strong video distillation and mixed image-video samplers [31, 32], suggest that alignment-time fine-tuning and video-specific inference improvements may be fruitfully combined.

Broader Opportunities. Our work is also connected to several broader directions in the diffusion literature. A recent survey summarizes the fundamentals, challenges, and future directions of diffusion alignment [16]. From the data perspective, diffusion dataset condensation and dataset pruning suggest that compact yet informative training subsets can substantially reduce training cost [8, 44], which is consistent with our motivation of aligning models from extremely limited filtered data. From the evaluation perspective, recent work shows that T2I comparisons can be distorted by guidance-scale bias [40], reinforcing our use of multiple complementary metrics rather than relying on a single score. Meanwhile, CRAFT focuses on improving model alignment during fine-tuning, but it is naturally compatible with a range of orthogonal inference-time advances. Recent methods improve generation quality or efficiency through collect-reflect-refine pipelines, sparse attention approximations, or carefully designed inference heuristics for alternative generative backbones [13, 30, 33]. Beyond alignment and synthesis, diffusion models have also shown promise for zero-shot retrieval and classification [12, 23]. Understanding how these data-centric, inference-centric, and downstream semantic capabilities interact with preference-aligned diffusion models is another valuable direction for future research.

Figure 8. **Qualitative results of CRAFT-SDXL.** After CRAFT fine-tuning, the model demonstrates the capability to generate visually superior images with exceptional aesthetic quality.

Figure 9. **Qualitative results of CRAFT-SDXL.** After CRAFT fine-tuning, the model demonstrates the capability to generate visually superior images with exceptional aesthetic quality.