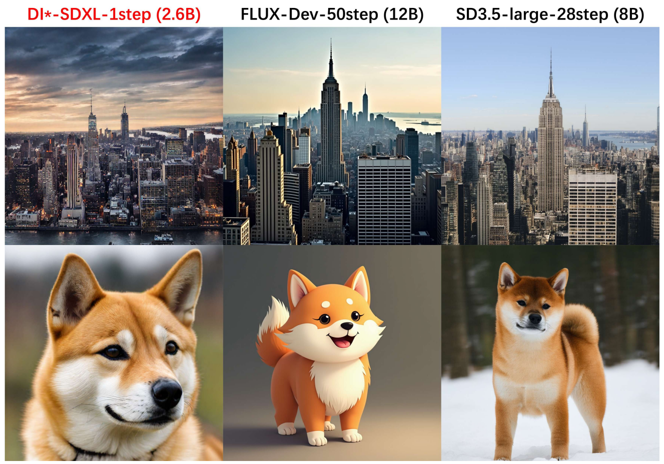
David and Goliath: Small One-step Model Beats Large Diffusion with Score Post-training

Weijian Luo*[1] Colin Zhang[1] Debing Zhang[1] Zhengyang Geng[2]



Algorithm 1: Diff-Instruct* Pseudo Code.

Input: prompt dataset $\mathcal{C}$, generator $g_\theta(\boldsymbol{x}_0|\boldsymbol{z},\boldsymbol{c})$, prior distribution p_z, reward model $r(\boldsymbol{x},\boldsymbol{c})$, reward model scale α_{rew}, CFG reward scale α_{cfg}, reference diffusion model $\boldsymbol{s}_{ref}(\boldsymbol{x}_t|c,\boldsymbol{c})$, assistant diffusion $\boldsymbol{s}_\psi(\boldsymbol{x}_t|t,\boldsymbol{c})$, forward diffusion $p(\boldsymbol{x}_t|\boldsymbol{x}_0)$ (2.1), assistant diffusion updates rounds K_{TA}, time distribution $\pi(t)$, diffusion model weighting $\lambda(t)$, generator IKL loss weighting $w(t)$.

while *not converge* **do**

freeze θ, update ψ for K_{TA} rounds by

1. sample prompt $\boldsymbol{c}\sim\mathcal{C}$; sample time $t\sim\pi(t)$; sample $\boldsymbol{z}\sim p_z(\boldsymbol{z})$;
2. generate fake data: $\boldsymbol{x}_0=\mathrm{sg}[g_\theta(\boldsymbol{z},\boldsymbol{c})]$; sample noisy data: $\boldsymbol{x}_t\sim p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)$;
3. update ψ by minimizing loss: $\mathcal{L}(\psi)=\lambda(t)\|\boldsymbol{s}_\psi(\boldsymbol{x}_t|t,\boldsymbol{c})-\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)\|_2^2$;

freeze ψ, update θ using SGD:

1. sample prompt $\boldsymbol{c}\sim\mathcal{C}$; sample time $t\sim\pi(t)$; sample $\boldsymbol{z}\sim p_z(\boldsymbol{z})$;
2. generate fake data: $\boldsymbol{x}_0=g_\theta(\boldsymbol{z},\boldsymbol{c})$; sample noisy data: $\boldsymbol{x}_t\sim p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)$;
3. explicit reward: $\mathcal{L}_{rew}(\theta)=-\alpha_{rew}r(\boldsymbol{x}_0,\boldsymbol{c})$;
4. CFG reward: $\mathcal{L}_{cfg}(\theta)=\alpha_{cfg}\cdot w(t)\{\boldsymbol{s}_{ref}(\mathrm{sg}[\boldsymbol{x}_t]|t,\boldsymbol{c})-\boldsymbol{s}_{ref}(\mathrm{sg}[\boldsymbol{x}_t]|t,\varnothing)\}^T\boldsymbol{x}_t$;
5. score-regularization: $\mathcal{L}_{reg}(\theta)=-w(t)\{\mathbf{d}'(\boldsymbol{s}_\psi(\boldsymbol{x}_t|t,\boldsymbol{c})-\boldsymbol{s}_{ref}(\boldsymbol{x}_t|t,\boldsymbol{c}))\}^T\{\boldsymbol{s}_\psi(\boldsymbol{x}_t|t,\boldsymbol{c})-\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)\}$;
6. update θ by minimizing DI* loss: $\mathcal{L}_{DI*}(\theta)=\mathcal{L}_{rew}(\theta)+\mathcal{L}_{cfg}(\theta)+\mathcal{L}_{reg}(\theta)$;

end

return θ,ψ.

One-step Diffusion Model?

Since the pioneering work of **Diff-Instruct**, **one-step generative models** map latent noises directly to data samples in a single forward pass; making them preferred with **very strong performance and efficiency;**

The life of a one-step diffusion model.

Problem formulation:

$$\theta^* = \operatorname*{argmin}_\theta \mathbb{E}_{\boldsymbol{x}_0\sim p_\theta(\boldsymbol{x}_0|\boldsymbol{c})}\Big\{\big[-\alpha r(\boldsymbol{x}_0,\boldsymbol{c})\big]+D(p_\theta,p_{ref})\Big\} \quad (3.1)$$

$$\mathcal{L}_{Orig}(\theta)=\mathbb{E}_{\substack{\boldsymbol{z}\sim p_z,\\ \boldsymbol{x}_0=g_\theta(\boldsymbol{z},\boldsymbol{c})}}\big[-\alpha r(\boldsymbol{x}_0,\boldsymbol{c})\big]+\mathbf{D}^{[0,T]}(p_\theta,p_{ref}) \quad (3.3)$$

If we take the Score-based divergence as the proximal regularization:

$$\mathbf{D}^{[0,T]}(p_\theta,p_{ref}) := \int_0^T w(t)\mathbb{E}_{\pi_t}\mathbf{d}(\boldsymbol{s}_{p_{\theta,t}}-\boldsymbol{s}_{q_t})\mathrm{d}t. \quad (3.2)$$

We can practically minimize the learning objective with:

$$\mathcal{L}_{DI*}(\theta)=\mathbb{E}_{\substack{\boldsymbol{z}\sim p_z,\\ \boldsymbol{x}_0=g_\theta(\boldsymbol{z})}}\Big[-\alpha r(\boldsymbol{x}_0,\boldsymbol{c}) + \int_{t=0}^T w(t)\mathbb{E}_{\substack{\boldsymbol{x}_t|\boldsymbol{x}_0\\ \sim p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)}}\Big\{-\mathbf{d}'(\boldsymbol{y}_t)\Big\}^T \Big\{\boldsymbol{s}_{p_{\mathrm{sg}[\theta]},t}(\boldsymbol{x}_t)-\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t|\boldsymbol{x}_0)\Big\}\mathrm{d}t\Big]. \quad (3.4)$$

with $\boldsymbol{y}_t := \boldsymbol{s}_{p_{\mathrm{sg}[\theta]},t}(\boldsymbol{x}_t)-\boldsymbol{s}_{q_t}(\boldsymbol{x}_t)$.

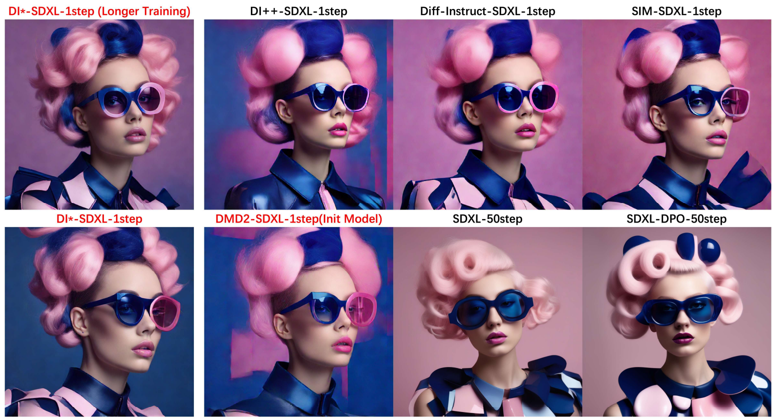
The **key theorem** shows the equivalence of DI* loss and general score-divergence (**refer to Score Implicit Matching**):

Theorem 3.1. Under mild assumptions, if we take the sampling distribution in (3.2) as $\pi_t = p_{\mathrm{sg}[\theta],t}$, then the gradient of (3.3) w.r.t θ is the same as (3.4):

$$\frac{\partial}{\partial\theta}\mathcal{L}_{Orig}(\theta)=\frac{\partial}{\partial\theta}\mathcal{L}_{DI*}(\theta).$$

Table 3: Quantitative evaluations of models on **HPSv2.1** scores. We compare open-sourced models regardless of their base model and architecture. † indicates our implementation.

MODEL	ANIMATION↑	CONCEPT-ART↑	PAINTING↑	PHOTO↑	AVERAGE↑
50STEP-SDXL-BASE(PODELL ET AL., 2023)	30.85	29.30	28.98	27.05	29.05
50STEP-SDXL-DPO(WALLACE ET AL., 2024)	32.01	30.75	30.70	28.24	30.42
28STEP-SD3.5-LARGE	31.89	30.19	30.39	28.01	30.12
50STEP-FLUX-DEV	32.09	30.44	31.17	29.09	30.70
1STEP-DMD2-SDXL(YIN ET AL., 2024)	29.72	27.96	27.64	26.55	27.97
1STEP-DIFF-INSTRUCT-SDXL(LUO ET AL., 2024B)	31.15	29.71	29.72	28.20	29.70
1STEP-SIM-SDXL(LUO ET AL., 2024C)	31.97	30.46	30.13	28.08	30.16
1STEP-DI++-SDXL(LUO, 2024)	31.19	29.88	29.61	28.21	29.72
1STEP-DI*-SDXL(OURS)	32.26	30.57	30.10	27.95	30.22
1STEP-DI*-SDXL(OURS, LONGER TRAINING)	33.22	31.67	31.25	28.62	31.19
1STEP-DI*-SDXL-DDPO(OURS, LONGER TRAINING)	**33.92**	**32.80**	**32.71**	**29.62**	**32.26**



Prompt: art collection style and fashion shoot, in the style of made of glass, dark blue and light pink, paul rand, solarpunk, camille vivier, …