Schedule On the Fly: Diffusion Time Prediction for Faster and Better Image Generation

Supplementary Material

A. Dataset Details

In Section 4, we provide a brief overview of our training dataset. Below, we present a more detailed description of the dataset.

A.1. Dataset Source

We curated training prompts from three high-quality datasets, COCO [24], Laion-art [54] and COYO-11M [6].

For COCO, the original captions in the training split is used. For Laion-art, we caption the images with Florence-2 [62] to obtain the text prompts. For COYO-11M, we retained its original captioning by Llava-next [26]. Fig. 7 illustrates the distribution of prompt length after tokenization.

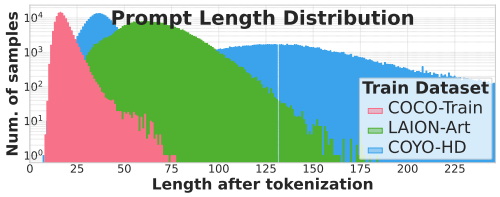


Figure 7. Prompt length distribution.

To ensure diversity during training, we design a filtering pipeline to select a diverse subset of these prompts.

A.2. Filtering Pipeline

Our filtering pipeline, depicted in Figure 8, aims to select diverse and high-quality prompts by analyzing key linguistic features, such as nouns, prepositions, and adjectives. To assess diversity, we utilize *WordNet* [38] to count the number of valid nouns, adjectives, and prepositions in each prompt.

Prompts are ranked based on these counts. Those exhibiting the highest scores are included in the training set. The selection process prioritizes **noun-diversity**, ensuring a balanced representation across categories such as *Person*, *Animal*, *Plant*, *Artifacts (Large/Small Objects)*, and *Natural Views*. Next, prompts with high **preposition-diversity** are selected, emphasizing those that contain spatial and relational terms (e.g., `near`, `on`). Finally, prompts are evaluated for **adjective-diversity**, with particular focus on adjectives describing *color* and *shape*, to enhance descriptive richness.

The process is iterative, selecting the top prompts in order of their noun, adjective, and preposition diversity scores

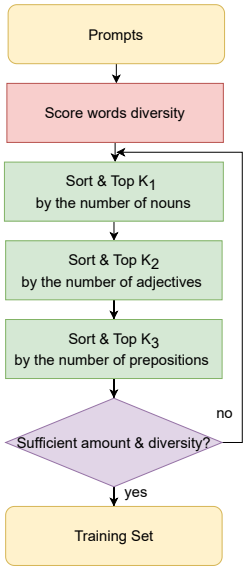


Figure 8. The filtering pipeline used to construct the training set.

until the desired data quantity (51,200 prompts, as specified in Section 4.1) and diversity are achieved. As shown in Figure 9, the diversity of the training set has significantly enhanced after filtering.

B. Applicaion on Denoising Diffusion Probabilistic Models (DDPMs)

Before the widespread use of flow-matching models, diffusion models dominated image generation, such as DDPM [13] and LDM [45]. They demonstrate several differences from flow models during inference. First, they directly predict the noise added to the data at each step, rather than predicting the velocity vector field. Second, both the diffusion time t and the noise level are typically discrete, unlike continuous time in flow-matching models.

In this section, we integrate TPM into Stable Diffusion v1.5 [45], demonstrating that TPM also works with DDPMs. At diffusion time t_{n-1}, to determine the next diffusion time t_n, TPM predicts two real-valued parameters, a_n and b_n. These parameters define α_n and β_n, which shape the Beta distribution of the diffusion time decay rate r_n,

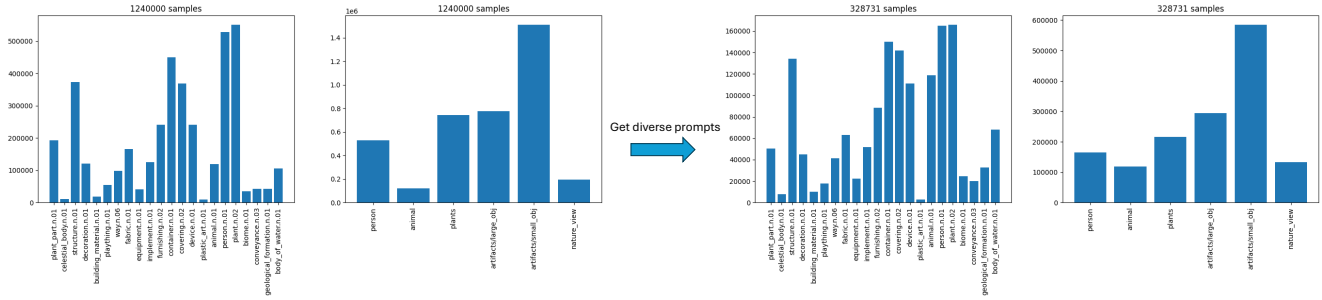


Figure 9. The statistics of prompts from the re-captioned high-resolution Laion-Art dataset, before and after filtering, highlight an improvement in diversity. Our dataset demonstrates greater variety compared to the original.

Method	Inference Steps	FID	CLIP-T	Aesthetic v2	Pick Score	HPSv2.1
DPM-Solver++ [30]	10	21.08	**0.314**	**5.219**	21.30	24.00
DPM-Solver++ & AYS [46]	10	18.81	0.313	5.156	21.31	**24.40**
DPM-Solver++ & GITS [7]	10	18.32	0.313	5.108	21.19	23.72
DPM-Solver++ & TPM	9.89	**18.31**	0.313	5.118	**21.32**	24.29

Table 4. Experiments on Stable Diffusion v1.5 [45], and 9.89 is the average number of steps TPM used. The best results are highlighted in bold.

similar to the flow-matching model version of TPDM introduced in 3.2. Then t_n is obtained using Eq. 6 and quantized to the nearest discrete diffusion time to obtain the corresponding noise level.

We compare the performance of TPDM with other scheduler optimization methods applied to diffusion models. The evaluation metrics are calculated using 5,000 prompts from the COCO 2017 validation set, in line with the evaluation protocol as in Table 1. Notably, our findings reveal that TPM can be combined with higher-order solvers, such as DPM-Solver++ [30], resulting in a significant improvement of -2.79 in FID, surpassing the performance of other methods as reported in Tab. 4.

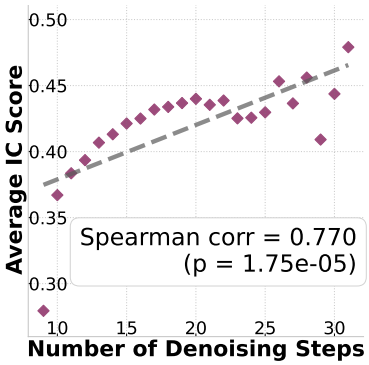


Figure 10. Correlation between the number of steps and image complexity (measured by ICNet [12]).

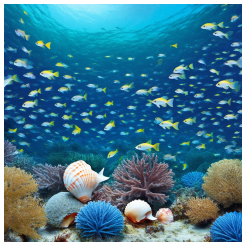
C. Analysis of the Predicted Schedule

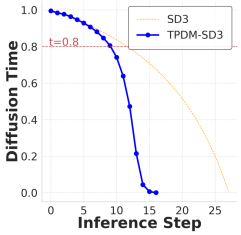
Fig. 11 provides more examples to show the noise schedule predicted by TPDM.

We observe that TPDM tends to allocate more inference steps to higher noise levels to generate complex details and layouts. Take Fig. 11(a) for an example where multiple objects of various sizes are generated with a complex visual layout – TPDM allocates 10 out of 17 steps to more noisy diffusion time with $t > 0.8$. This way, TPDM more efficiently spends its inference steps on denoising noisier samples at the earlier stage so that it can add more complex visual layouts and details as early as possible to eventually generate high quality results. It avoids the problem in the benchmark diffusion model that may waste many evenly allocated steps to denoise cleaner images at the later stage. The steep curve in Fig. 11(c) indicates that, for simple generation, TPDM reduces the diffusion time much faster to almost zero within only 13 steps, instead of evenly allocating 28 steps in the benchmark model. Fig. 11(e) and 11(f) also visualize the results for extremely long prompts. They exhibit a similar trend.

Moreover, Fig. 4 shows that simple prompts lead to faster decrease in diffusion time with fewer steps, while complex ones lead to more steps. We use ICNet [12] to score the image complexity, and calculate the average score of images generated with different number of steps in evaluation set. As shown in Fig. 10, it has a high correlation efficient (0.770), which supports our assumption that complex images are dynamically generated with more steps by TPDM.

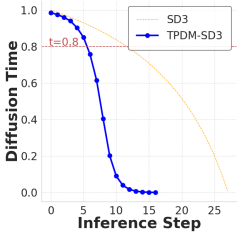
Prompt: Seabed wonderland, schools of fish, seashells, multi-sized corals.





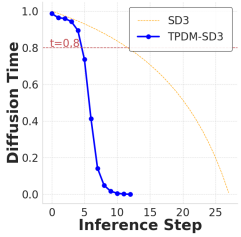
(a)

Prompt: Arctic ice, polar bears, drifting floes, aurora.



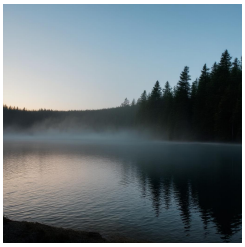
(b)

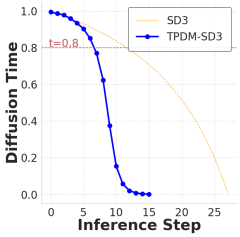
Prompt: Snowy peak, soaring eagle, icy winds, blue sky.



(c)

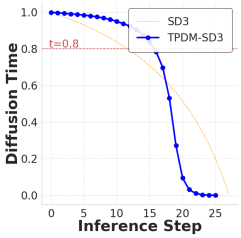
Prompt: Serenity over a lake, mist rising, dawn's first light, silent forest.





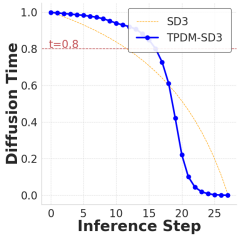
(d)

Prompt: An ornate, golden invitation letter with intricate calligraphy. The text reads 'Your Presence is Requested at the Royal Feast' in elegant, swirling script. The letter is illuminated by soft candlelight and rests on a royal velvet cushion. The background features a grand palace with towering spires and lush gardens, with a small scroll tucked inside the envelope.



(e)

Prompt: An ancient, leather-bound book with embossed mystical symbols, open on a wooden desk in a dimly lit library. The pages are yellowed, filled with arcane illustrations and handwritten notes in an unknown script. A faint, ethereal glow emanates from the book, casting soft light on a quill and ink pot nearby. A crystal ball sits on the desk, reflecting the glow. Through an arched window, a full moon shines in a starry night sky, and a raven perches on the windowsill, its eyes fixed on the book. In the background, shelves are lined with other mysterious tomes, some with spines adorned with gems or strange markings.



(f)

Figure 11. Prompts, images and predicted schedules.

D. Prompts to Generate Images in Fig. 1

1. 8k uhd A man looks up at the starry sky, lonely and ethereal, Minimalism, Chaotic composition Op Art.
2. A deep-sea exploration vessel descending into the pitch-black ocean, its powerful lights illuminating the glowing, alien creatures that inhabit the abyss. A massive, ancient sea creature with bioluminescent patterns drifts into view, its eyes glowing as it watches the explorers from the shadows of an underwater cave.
3. Half human, half robot, repaired human.
4. A baby painter trying to draw very simple picture, white background.
5. an astronaut sitting in a diner, eating fries, cinematic, analog film.

6. Van Gogh painting of a teacup on the desk.
7. A galaxy scene with stars, planets, and nebula clouds.
8. A hidden, forgotten city deep in a jungle, with crumbling stone temples overgrown by thick vines. In the heart of the city, a mysterious glowing artifact lies on an ancient pedestal, surrounded by an eerie mist. Strange symbols shimmer faintly on the stone walls, waiting to be uncovered.
9. A lone astronaut stranded on a desolate planet, gazing up at the sky. The planet's surface is cracked and barren, with glowing, unearthly ruins scattered across the horizon. In the distance, a massive, alien ship slowly descends, casting an eerie shadow over the landscape.