Evaluating Dataset Watermarking for Fine-tuning Traceability of Customized Diffusion Models: A Comprehensive Benchmark and Removal Approach

Supplementary Material

The supplementary material provides additional details that complement the main paper. It includes dataset descriptions, implementation details, extended experimental results, qualitative visualizations, and full prompt lists that could not be included in the main manuscript due to space constraints. Unless otherwise specified, all supplementary experiments follow the same evaluation protocol, metric settings, and training pipeline as those used in the main paper, in order to ensure consistency and comparability across all reported results.

1. Dataset Details

In this section, we provide additional information about the three datasets used in Sec. 4 of the main paper. These datasets are selected to cover three representative protection scenarios, namely facial identity protection, virtual object protection, and artistic style protection.

- ***CelebA*** [5]: This dataset contains face images of celebrities. For our experiments, each image is paired with a descriptive caption generated by LLaVA [4]. Since the original dataset is substantially larger than required for diffusion model fine-tuning in our setting, we randomly sample 1,000 image-caption pairs for experiments. This dataset is mainly used to evaluate the preservation and traceability of identity-related content.
- ***Pokémon*** [6]: This dataset consists of 833 high-quality Pokemon images, each accompanied by a text caption produced by the BLIP [3] captioning model. Compared with real-image benchmarks, this dataset emphasizes stylized and character-centric visual concepts, making it suitable for evaluating watermark behavior in virtual object customization settings.
- ***WikiArt*** [9]: This dataset contains 81,444 artwork images collected from WikiArt.org, with annotations such as artist, genre, and style. Owing to its broad stylistic diversity and rich artistic textures, WikiArt serves as a challenging benchmark for studying artistic style protection and the transferability of watermark-related effects under fine-tuning.

Together, these three datasets provide complementary evaluation scenarios and enable a more comprehensive analysis of watermark robustness, traceability, and removal effectiveness under different data characteristics.

2. Experimental Settings

All experiments are implemented in Python 3.10 with PyTorch 2.7.1, and are conducted on Ubuntu 20.04 using four NVIDIA A800 GPUs. Unless explicitly stated otherwise, the same computational environment and evaluation pipeline are used throughout all experiments.

2.1. Models and Datasets

We consider four representative dataset watermarking methods, namely DIAGNOSIS [8], DiffusionShield [1], SIREN [2], and WatermarkDM [10], as the primary protection baselines. These methods embed imperceptible watermarks into the original training images, with the goal of enabling the tracing of unauthorized data usage during downstream diffusion model customization. Following the setup in the main paper, we evaluate them on three high-resolution datasets, CelebA-HQ [5], Pokémon [6], and WikiArt [9], corresponding to face protection, virtual object protection, and artistic style protection, respectively.

2.2. Implementation Details

During watermark embedding, we follow the default settings of each watermarking method and resize all images to 512×512. During diffusion model fine-tuning, we uniformly adopt Stable Diffusion v1.4 (SD1.4) [7] as the default backbone and apply the same prompt template across all four fine-tuning strategies, so that the comparison is not affected by differences in prompting format. Each fine-tuning run is based on 10 training images sampled from the corresponding dataset. During evaluation, we use 50 prompts for each dataset, and generate five images per prompt. The final FID and CLIP similarity scores are reported as the average over the generated samples. This unified setup ensures that the observed performance differences mainly reflect the effects of watermarking and data removal, rather than confounding factors introduced by inconsistent fine-tuning or generation conditions.

3. Common Distortion Processing

To evaluate watermark robustness under realistic perturbations, we consider three commonly used natural distortions: Gaussian blur, JPEG compression, and Gaussian noise. These distortions are designed to simulate typical image degradation encountered in practical scenarios such as image storage, transmission, and post-processing.

Algorithm 1 Apply Image Distortions: Gaussian Blur, JPEG Compression, and Gaussian Noise

Require: Original image I; JPEG quality $q = 15$; Gaussian noise std $\sigma = 0.3$
Ensure: Distorted images: I_{blur}, I_{jpeg}, I_{noise}
Gaussian Blur:
$I_{\text{blur}} \leftarrow \text{GaussianBlur}(I, \text{kernel\_size} = 31, \sigma = 0)$
JPEG Compression:
$I_{\text{jpeg}} \leftarrow \text{JPEGEncode}(I, \text{quality} = q)$
Gaussian Noise:
$I_{\text{norm}} \leftarrow I/255$
Sample $\varepsilon \sim \mathcal{N}(0, \sigma^2)$
$I_{\text{noise}} \leftarrow \text{clip}(I_{\text{norm}} + \varepsilon, 0, 1)$
$I_{\text{noise}} \leftarrow I_{\text{noise}} \times 255$
Return: $I_{\text{blur}}, I_{\text{jpeg}}, I_{\text{noise}}$

3.1. Image Blur

For Gaussian blur, we convolve the image with a two-dimensional Gaussian kernel. In our implementation, a 31×31 kernel is used, and the standard deviation is automatically determined by the underlying image processing function. This distortion mainly smooths local image details and attenuates high-frequency information.

$$I_{\text{blur}}(x, y) = (I * G)(x, y) = \sum_{u=-k}^{k} \sum_{v=-k}^{k} I(x-u, y-v) \cdot G(u, v), \tag{1}$$

$$G(u, v) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right), \tag{2}$$

where I is the original image, G is a two-dimensional Gaussian kernel, σ is standard deviation. We used a 31×31 kernel, which automatically calculates the corresponding standard deviation.

3.2. JPEG Compression

For JPEG compression, the image is first partitioned into 8×8 blocks, transformed into the frequency domain by the discrete cosine transform (DCT), and then quantize the frequency domain coefficients $C(u, v)$, as follows:

$$C_q(u, v) = \text{round}\left(\frac{C(u, v)}{Q(u, v)}\right), \tag{3}$$

where $Q(u, v)$ is the standard quantization matrix. The distortion mainly comes from this quantization operation. The JPEG quality level used in the code is 15.

3.3. Gaussian Noise

For Gaussian noise, zero-mean Gaussian noise is independently added to each color channel, followed by clipping the perturbed pixel values to the valid range, the process is as follows:

$$I_{\text{noisy}}(x, y, c) = \text{clip}\left(I(x, y, c) + \mathcal{N}(0, \sigma^2), 0, 1\right), \tag{4}$$

where $\mathcal{N}(0, \sigma^2)$ represents Gaussian noise with mean 0 and variance x; the $clip$ function clamps the pixel values to the interval $[0, 1]$. The distortion is added independently in the dimension of the color channel c.

The complete distortion pipeline is summarized in Algorithm 1, which provides a unified description of the three perturbation processes used in our experiments.

4. Experimental Analysis

In the main paper, we adopt Stable Diffusion v1.4 (SD1.4) as the default diffusion backbone, following the evaluation setting commonly used in prior watermarking studies. This choice is also consistent with the configuration adopted by most mainstream baselines, which facilitates fair comparison under a unified experimental protocol. To examine whether the conclusions of this work extend beyond a single diffusion architecture, we further conduct supplementary experiments on SDXL. The results reported in Table 1 show that the overall benchmark observations remain consistent when moving from SD1.4 to a stronger backbone, indicating that the conclusions drawn in the main paper are not specific to SD1.4. In particular, the relative performance trends across different fine-tuning strategies are largely preserved, suggesting that the benchmark captures stable properties of dataset watermarking rather than artifacts of a particular model choice.

We further report the performance of DeAttack on SDXL in Table 2. The results show that DeAttack remains effective under a different diffusion architecture, demonstrating that its removal capability is not tightly coupled to the original SD1.4 setting. This observation suggests that the method has reasonable architectural generalizability and can transfer to stronger generation backbones without substantial degradation in effectiveness.

To further assess the robustness of DeAttack, we additionally evaluate it on multiple datasets and under different fine-tuning strategies, with the corresponding results presented in Table 3. The overall trends remain consistent with those observed in the main paper, indicating that the effectiveness of DeAttack is not limited to a single dataset or a specific fine-tuning configuration. Instead, the method exhibits stable behavior across different data distributions and customization settings, which further supports its practical applicability.

In addition, the main paper reports experimental statistics under different data protection ratios, where the clean setting corresponds to a protection ratio of 0%. Since this condition already appears in the main paper as the unpro-

Table 1. Experimental results of SDXL.

FT Method	CelebA-HQ			Pokémon			WikiArt		
	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑
T2I	0.1771	284.35	96.35	0.1915	264.73	98.44	0.1137	374.39	70.32
LoRA	0.2287	263.33	90.32	0.2106	216.57	92.83	0.1324	378.41	70.21
DB	0.2249	285.64	88.73	0.2559	279.40	92.28	0.2232	350.25	83.36
TI	0.1917	291.32	92.59	0.2437	243.59	90.35	0.2139	360.78	81.59

Table 2. Results of DeAttack on SDXL.

FT Method	CelebA-HQ			Pokémon			WikiArt		
	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑
T2I	0.1934	284.23	52.59	0.1972	272.49	54.12	0.1359	343.74	58.89
LoRA	0.2389	254.89	51.52	0.2109	296.37	49.45	0.1994	373.15	55.57
DB	0.2322	258.71	58.73	0.2775	234.31	54.10	0.2021	354.69	56.74
TI	0.2301	302.29	55.37	0.2697	265.91	49.86	0.2103	356.98	57.21

tected reference case, we further provide in Table 4 the results of natural distortion under the clean setting for completeness. These results offer a clearer reference point for understanding model behavior in the absence of watermark protection, and they facilitate comparison with the corresponding results under non-zero protection ratios.

5. Visualization

5.1. Frequency domain distribution

This section analyzes how different watermarking methods, as well as natural distortions, affect the frequency-domain characteristics of the original image. As shown in Figure 1, we visualize the perturbation patterns introduced by DIAGNOSIS [8], SIREN [2], and DiffusionShield [1], Gaussian noise, and Gaussian blur using heatmaps in the Fourier domain. Specifically, the Fourier transform is applied to project pixel-level perturbations into the frequency space, where brighter yellow regions indicate larger-magnitude deviations and greener regions indicate relatively smaller changes.

From a spatial perspective, the center of the spectrum corresponds to low-frequency components, while regions farther from the center correspond to higher-frequency components. Based on this visualization, different watermarking methods exhibit distinct spectral characteristics. DIAGNOSIS tends to produce stronger perturbations concentrated in the low-frequency region, whereas DiffusionShield shows a comparatively more distributed response across both low- and high-frequency bands. This suggests that different watermarking methods encode traceability signals in different spectral patterns, which may partly explain differences in robustness and removability observed in the main experiments.

The two natural distortions also exhibit recognizable frequency-domain behaviors. Gaussian noise introduces dispersed perturbations, while Gaussian blur leads to more pronounced low-frequency changes and stronger degradation in the overall spectrum. Compared with Gaussian noise, Gaussian blur produces more visually concentrated and larger-magnitude changes, indicating that it may interfere more strongly with the spectral structure of the original image. Overall, these visualizations provide an intuitive explanation of how watermark-induced perturbations differ from natural distortions, and help clarify why some watermarking methods are more resilient or more vulnerable under certain post-processing operations.

5.2. Fine-tuning generation

This section presents qualitative examples of images generated after fine-tuning on watermarked data, with comparisons across four watermarking methods and four fine-tuning strategies. The purpose of these visualizations is to provide an intuitive view of how watermark embedding influences downstream generation behavior under different customization settings.

As shown in Figure 2, the examples on the Pokemon dataset highlight the impact of watermarking in a stylized, character-centric domain. Since Pokemon images typically contain clear contours, saturated colors, and relatively simple semantic compositions, visual differences across methods can be more directly observed in object appearance, color consistency, and structural fidelity. In contrast, Figure 3 presents results on the WikiArt dataset, which contains

Table 3. Results of DeAttack on three datasets and fine-tuning method.

FT Method	CelebA-HQ			Pokémon			WikiArt		
	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑
T2I	0.2018	286.39	54.84	0.1934	265.37	53.10	0.1090	362.25	56.76
DB	0.2589	253.51	59.86	0.2931	233.59	55.56	0.2280	356.20	54.74
TI	0.2503	292.30	56.25	0.2901	268.81	50.86	0.2581	348.24	52.78

Table 4. Results of clean data under different natural distortion.

Distortion Type	Text-to-Image			LoRA			DreamBooth			Textual Inversion		
	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑	CLIP-T↑	FID↓	Acc.↑
Blur(w/o)	0.2274	230.81	N/A	0.2621	241.74	N/A	0.2558	223.45	N/A	0.2617	243.79	N/A
Blur(w)	0.2150	217.95	N/A	0.2412	273.47	N/A	0.2533	240.02	N/A	0.2489	239.61	N/A
JPEG(w/o)	0.2175	235.92	N/A	0.2580	243.39	N/A	0.2512	232.12	N/A	0.2594	242.72	N/A
JPEG(w)	0.2163	220.83	N/A	0.2386	278.88	N/A	0.2513	244.59	N/A	0.2502	241.45	N/A
Noise(w/o)	0.1975	240.36	N/A	0.2039	275.44	N/A	0.2201	236.86	N/A	0.2332	252.81	N/A
Noise(w)	0.1992	229.62	N/A	0.2014	279.23	N/A	0.2201	269.35	N/A	0.2293	243.89	N/A

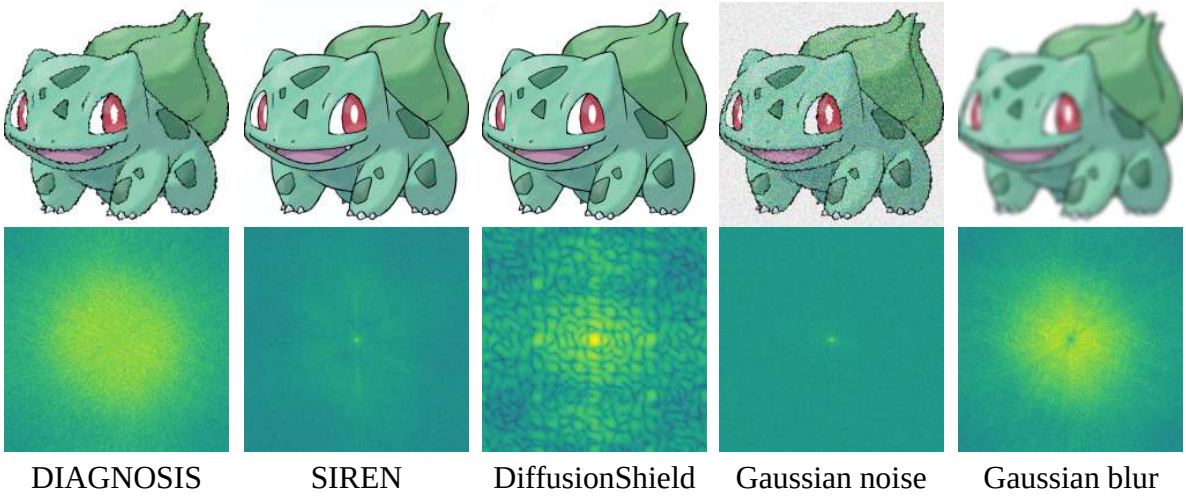

Figure 1. Heatmap visualization of frequency-domain changes caused by watermarking methods and natural distortion.

diverse artistic styles, painterly textures, and more complex visual abstractions. In this setting, the influence of watermarking is reflected not only in object-level structure, but also in style transfer behavior, brushstroke patterns, and texture coherence.

Taken together, these qualitative comparisons show that the visual effects of watermark embedding and removal are not limited to a single visual domain. Instead, they can be observed across both structured cartoon-like images and highly diverse artistic images. This further supports the generality of the benchmark and provides complementary evidence to the quantitative results reported in the main paper and the supplementary tables.

6. Prompts for generation

This section provides the complete set of textual prompts used for image generation after fine-tuning the diffusion model. For each dataset, we use 50 prompts designed to cover a broad range of semantic content and stylistic variations, so that the evaluation is not biased toward a narrow subset of concepts. More specifically, the prompts for CelebA emphasize facial attributes, accessories, and appearance-related descriptions; the prompts for Pokemon focus on character identity, shape, color, and cartoon-style object descriptions; and the prompts for WikiArt cover diverse artistic genres, painting styles, historical aesthetics, and compositional patterns.

Figure 2. Pokémon dataset fine-tuning results.

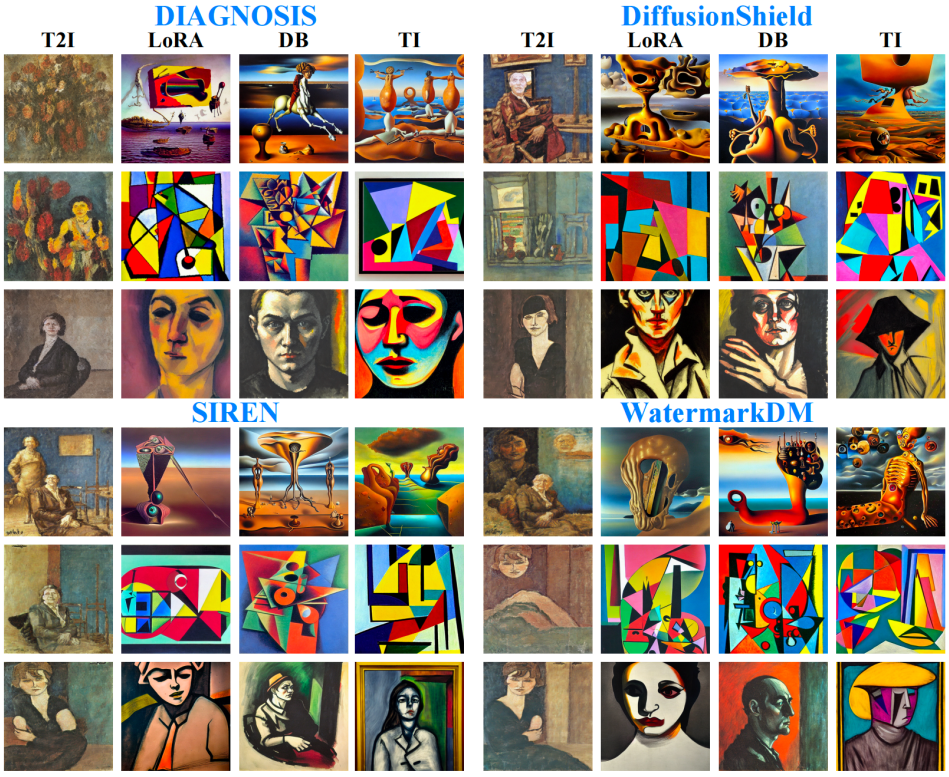

Figure 3. WikiArt dataset fine-tuning results.

Table 5. List of CelebA dataset prompts.

ID	Prompt Description
1	A young woman with red hair, heart-shaped face, small nose, large brown eyes, glasses, and a necklace. Likely a young adult.
2	A young bald man with a beard and round face, wearing a football helmet. He has thick lips and a large nose.
3	A blonde young woman with a heart-shaped face, small nose, thin lips, wearing a black dress and a necklace. Smiling.
4	A young woman with dark hair, heart-shaped face, full lips, straight nose, and a necklace. Likely in her late teens.
5	A bald man with glasses, wide face, thick lips, wearing a suit and standing at a microphone. He is an adult.
6	An elderly white man with glasses, wide face, large nose, thin mouth, and beard. Wearing a tie and brown jacket.
7	A young blonde woman with glasses, small nose, wide mouth, brown eyes, a bracelet, and a nose piercing. No facial hair.
8	A smiling young man with a round face, glasses, beard, brown eyes, and wearing a suit and tie. A flag is behind him.
9	A man with a wide face, thick mustache, black hair, and glasses. Smiling, with no other accessories visible.
10	A young woman with dark hair, brown cat-like eyes, heart-shaped face, wide mouth, small nose, and black dress.
11	A man with blonde hair, wide face, small nose, thick lips, large eyes, and glasses. Wearing a white shirt and smiling.
12	A young woman with a narrow heart-shaped face, dark hair, large eyes, small nose, thin lips, wearing pink dress and necklace.
13	A smiling young blonde woman with heart-shaped face, brown eyes, small nose, glasses, thick eyebrows, and a necklace.
14	A young man with glasses, dark hair, large nose, black eyes, strong jawline, and straight mouth. Likely a young adult.
15	A woman with blonde bobbed hair, red dress, heart-shaped face, small nose, smiling. Possibly young, not clearly elderly.
16	A young blonde woman with large expressive eyes, small nose, thin mouth, glasses, white shirt, and heart-shaped face.
17	A smiling young blonde woman with heart-shaped face, blue eyes, glasses, small wide nose, thick lips, and jewelry.
18	A young woman with brown eyes, glasses, thick lips, pearl necklace, heart-shaped face, small nose, and a smile.
19	A young woman with auburn hair, glasses, brown eyes, wide mouth, heart-shaped face, and a necklace.
20	A young woman with a ponytail, dark hair, glasses, small nose, full mouth, black clothes, and heart-shaped face.
21	An elderly bald man with a beard, wide thick face, large nose, wearing a jacket and hat. Seated in front of a camera.
22	A young man with shaved sides and ponytail, large brown eyes, wide upturned nose, glasses, beard, and rectangular face.
23	A smiling young blonde woman with heart-shaped face, pointed nose, full lips, earrings, and brown eyes. No glasses.
24	A young blonde woman with wide face, small wide nose, thick lips, large earrings, necklace, and a smile.
25	A young bald man with glasses, beard, large nose, wide face, thick eyebrows, and a black jacket. Likely young adult.
26	A goofy young man with round face, large eyes, small nose, thin mouth, glasses, and a playful smile.
27	A young man with shaved head, large nose, wide open eyes, black hoodie, wide mouth. Possibly a teen or young adult.
28	A young woman in pink dress with round face, glasses, pink bow, large brown eyes, holding a cherry in her mouth.
29	An elderly bald man with glasses, large nose, thick mustache, red-striped shirt, tie, and a big smile.
30	A man with long hair, beard, glasses, large nose, wide mouth, wearing a black shirt and jacket. Likely adult.
31	A young blonde woman with heart-shaped face, large brown eyes, small nose, necklace, and a thin mouth. Wearing a dress.
32	A smiling woman with almond-shaped eyes, wide nose, large mouth, pink dress, blonde hair. Likely a young adult.
33	A young woman with red shirt, heart-shaped face, large dark eyes, full lips, small nose, necklace, and ponytail.
34	A smiling young woman with heart-shaped face, small nose, large brown eyes, pink bathing suit, and pink headband.
35	A young man in a suit and tie with round face, small nose, brown eyes, glasses, beard, and a thin mouth.
36	A young woman with long black hair, glasses, small nose, heart-shaped face, necklace, and a thin mouth. Likely teen.
37	A smiling young woman with heart-shaped face, wide mouth, large eyes, necklace, and ponytail. Possibly young adult.
38	An elderly person with wide face, large black eyes, round nose, bushy eyebrows, suit, tie, and hat.
39	A smiling young woman with curly dark hair, large brown eyes, full lips, small nose, necklace, and heart-shaped face.
40	A man with beard, sunglasses, black suit and tie, large nose, wide mouth, and prominent chin. Handsome appearance.
41	An elderly bald man with white beard, glasses, very wide face, small nose, thick mouth, wearing a suit and tie.
42	A woman with round face, glasses, blonde hair, thick lips, blue shirt, small wide nose, and a friendly smile.
43	A man with shaved head and beard, blue shirt, wide mouth, large nose, small blue eyes, and a youthful appearance.
44	A young woman with dark straight hair, glasses, brown eyes, small nose, full lips, necklace, bracelet, and oval face.
45	A thin-faced bald man with glasses, large nose, close-set eyes, suit and tie, looking directly at the camera.
46	A young woman with heart-shaped face, dark hair, large expressive eyes, full mouth, small nose, and earrings.
47	A young woman with blue hair, red dress, large round eyes, thick lips, wide face, necklace, and blue eyes.
48	A person with long black hair, white shirt, large black eyes, wide nose, glasses, and nose piercing. Teen or adult.
49	A smiling young blonde woman with round face, small nose, blue eyes, glasses, necklace, and red background.
50	A young man with round face, glasses, full beard, small nose, wearing a suit and tie. Appears formally dressed.

Table 6. List of Pokémon dataset prompts.

ID	Prompt Description
1	a drawing of a green pokemon with red eyes
2	a cartoon monkey flying with a bone in its mouth
3	a drawing of a purple dragon with spikes on it's head
4	a drawing of a cat sitting on top of a flower
5	a pink bird with orange eyes and a pink tail
6	a cartoon bee with a big smile on it's face
7	a blue cartoon character with a target in his hand
8	a cartoon bird with a green leaf on its head
9	a drawing of a blue dinosaur with wings
10	a drawing of a green pokemon sitting on top of a leaf
11	a very cute looking pokemon type
12	a drawing of a shark with its mouth open
13	a cartoon character with a mushroom on his head
14	a drawing of a cat wearing a helmet
15	a drawing of a cat with a pink tail
16	a cartoon elephant with a red nose and orange ears
17	a drawing of a black and white animal with horns
18	a drawing of a purple and white animal
19	a drawing of a red and yellow insect
20	a drawing of a green and yellow lizard
21	a drawing of a blue and orange pokemon
22	a drawing of a gray and white pokemon
23	a cartoon picture of a green vegetable with eyes
24	a drawing of a green cartoon character with a sad look
25	a cartoon giraffe with a ball in its mouth
26	a cartoon bird with a hat on its head
27	a cartoon dog is standing in a pose
28	a drawing of a star with a red eye
29	a cartoon turtle with a tree on its back
30	a drawing of a pink cartoon character
31	a drawing of a fox with wings on it's back
32	a blue and yellow cartoon character with its mouth open
33	a cartoon mouse with a pink shirt and tie
34	a cartoon character with a yellow shirt and blue pants
35	a drawing of a fish with a horn on it's head
36	a drawing of a white and red pokemon
37	a drawing of a blue fish with yellow eyes
38	a cartoon bunny flying through the air
39	a drawing of a small animal with a pink nose
40	a blue and white cartoon character flying through the air
41	a green and yellow toy with a red nose
42	a drawing of a woman in a pink dress with a dragon head
43	a cartoon character with a magnifying glass
44	a drawing of a blue sea turtle holding a rock
45	a cartoon bear with a ring around its neck
46	a cartoon cat is holding onto a leash
47	a cartoon rat with its mouth open and it's mouth wide open
48	a green bird with a red tail and a black nose
49	a cartoon sheep is kicking a soccer ball
50	a close up of a cartoon character with big eyes

Table 7. List of WikiArt dataset prompts.

ID	Prompt Description
1	surreal oil painting, Salvador Dalí style, hyper-detailed, high quality
2	romantic landscape, 19th-century French painting, soft brushwork, ultra high-res
3	cubist still life, abstract geometric shapes, Picasso-inspired, vibrant colors
4	impressionist river scene, vivid brush strokes, Claude Monet style, realistic lighting
5	art nouveau floral pattern, elegant flowing lines, Alphonse Mucha inspired, intricate details
6	German expressionist portrait, emotional color palette, dramatic, cinematic lighting
7	Russian avant-garde constructivist poster, vintage style, bold typography, clean vector
8	hyper-realistic Baroque portrait, dramatic chiaroscuro, Rembrandt style, 8K
9	abstract color field painting, Rothko inspired, vivid colors, minimalist
10	Italian Renaissance fresco, mythological figures, high detail, realistic faces
11	minimalist geometric abstraction, Malevich style, pure shapes, modern design
12	medieval illuminated manuscript, gold leaf, intricate patterns, ancient calligraphy
13	gothic cathedral interior, stained glass, atmospheric light, photorealistic
14	Japanese woodblock print, Hokusai style, traditional ukiyo-e, fine linework
15	fauvist landscape, intense color contrasts, Matisse inspired, expressive painting
16	surreal dreamscape, Magritte style, hyper-realistic, conceptual art
17	art deco poster, glamorous 1920s woman, vintage illustration, high detail
18	Russian symbolist painting, mystical, ethereal lighting, rich textures
19	rococo palace interior, pastel colors, ornate details, photorealistic
20	Dutch golden age still life, flowers and fruits, realistic lighting, master painting
21	pre-Raphaelite portrait, medieval-inspired, flowing hair, detailed textile
22	Chinese ink landscape, shan shui style, misty mountains, traditional painting
23	Bauhaus modernist architectural drawing, clean lines, geometric composition
24	Italian futurist cityscape, motion blur, dynamic angles, vibrant
25	Byzantine mosaic, religious icon, gold tesserae, intricate details
26	Spanish romantic painting, dramatic history scene, vivid brushwork, realistic
27	symbolist fantasy scene, allegorical figures, mystical atmosphere, high detail
28	social realism mural, workers, propaganda style, bold colors, large format
29	abstract expressionist painting, chaotic brushstrokes, Pollock style, large canvas
30	Venetian rococo carnival scene, masked figures, ornate costumes, detailed
31	French rococo pastoral painting, elegant people, romantic light, high realism
32	Russian lubok folk art, storytelling style, bright colors, naive art
33	Egyptian revival decorative motif, hieroglyphs, ancient style, symmetrical pattern
34	neoclassical sculpture study, idealized human figure, marble texture, photorealistic
35	academic classical painting, mythological subject, realistic anatomy, dramatic light
36	Neue Sachlichkeit portrait, German realism, neutral colors, intense gaze
37	surrealist collage, Max Ernst style, dreamlike, high-res details
38	pre-Columbian inspired pattern, tribal geometric symbols, earthy colors
39	gothic illuminated manuscript page, ornate borders, medieval style, hyper-detailed
40	classical Greek vase painting, heroic myth scene, terracotta style, authentic
41	romantic seascape, stormy sky, 19th-century painting style, high detail
42	Renaissance-inspired religious altarpiece, golden halos, realistic faces, dramatic
43	art brut, outsider art style, raw brushstrokes, expressive emotion
44	Victorian fairy painting, delicate wings, flower garden, high detail
45	expressionist cityscape, angular architecture, dramatic colors, thick brush strokes
46	post-impressionist village scene, vivid colors, Van Gogh style, swirling strokes
47	orientalist painting, Middle Eastern architecture, rich textures, historical
48	pop art reinterpretation, classical sculpture, bright bold colors, high contrast
49	suprematist non-objective composition, simple shapes, modernist, clean vector
50	primitivist figure painting, tribal inspiration, earthy colors, simplified forms

By using a relatively diverse prompt set for each domain, we aim to better evaluate how watermarking and watsermark removal affect the generation quality, semantic alignment, and style preservation of fine-tuned diffusion models. The full prompt lists are reported in Table 5, Table 6, and Table 7, respectively, to facilitate reproducibility and future comparison.

References

[1] Yingqian Cui, Jie Ren, Han Xu, Pengfei He, Hui Liu, Lichao Sun, Yue Xing, and Jiliang Tang. Diffusionshield: A watermark for data copyright protection against generative diffusion models. *ACM SIGKDD Explorations Newsletter*, 26(2): 60–75, 2025. 1, 3

[2] Boheng Li, Yanhao Wei, Yankai Fu, Zhenting Wang, Yiming Li, Jie Zhang, Run Wang, and Tianwei Zhang. Towards reliable verification of unauthorized data usage in personalized text-to-image diffusion models. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 2564–2582. IEEE, 2025. 1, 3

[3] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*, pages 12888–12900. PMLR, 2022. 1

[4] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023. 1

[5] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015. 1

[6] Justin N. M. Pinkney. Pokemon blip captions. https://huggingface.co/datasets/lambdalabs/pokemon-blip-captions/, 2022. 1

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 1

[8] Zhenting Wang, Chen Chen, Lingjuan Lyu, Dimitris N Metaxas, and Shiqing Ma. Diagnosis: Detecting unauthorized data usages in text-to-image diffusion models. In *12th International Conference on Learning Representations, ICLR 2024*, 2024. 1, 3

[9] Wikiart. Wikiart: Visual art encyclopedia. https://www.wikiart.org/, 2016. 1

[10] Yunqing Zhao, Tianyu Pang, Chao Du, Xiao Yang, Ngai-Man Cheung, and Min Lin. A recipe for watermarking diffusion models. *arXiv preprint arXiv:2303.10137*, 2023. 1