Noise Diffusion for Enhancing Semantic Faithfulness in Text-to-Image Synthesis

Supplementary Material

A. Theoretical Analysis

In this section, we present the proof of the inequality $\|\nabla_\xi^2 s(\xi)\|_2 \overset{a.e.}{\leq} c$. First, we prove that this inequality holds within the domain of the function $s(\cdot)$. Then, we prove that the region where the derivatives of the function are undefined forms a set of zero measure, which guarantees the effectiveness of our method.

Proof of the inequality. Considering a neural network $f(\cdot)$ with the following structure:

$$f(x) = g_k\left(\omega_k \cdots g_1\left(\omega_1 x + b_1\right) + b_2 \cdots + b_k\right), \tag{1}$$

where x is the input, $g_i(\cdot)$ is the activation function for the i-th layer, ω_i is the weight matrix, and b_i is the bias vector for the i-th layer. Using the chain rule, the first derivative of $f(x)$ can be expressed by:

$$f'(x) = \frac{\partial g_k}{\partial z_k} \cdot \frac{\partial z_k}{\partial x}, \tag{2}$$

where $z_k = \omega_k g_{k-1}\left(z_{k-1}\right) + b_k$. The second derivative is given by:

$$f''(x) = \frac{\partial^2 g_k}{\partial z_k^2} \cdot \left(\frac{\partial z_k}{\partial x}\right)^2 + \frac{\partial g_k}{\partial z_k} \cdot \frac{\partial^2 z_k}{\partial x^2}. \tag{3}$$

The complexity of the derivative mainly comes from the activation functions, as the first derivative of a linear function is constant and its second derivative is zero. Table. 1 presents the bounds of the first and second derivatives for basic activation functions widely used in neural networks. For smooth functions, both their first and second derivatives are bounded over the entire domain. For non-smooth activation functions, their first and second derivatives are bounded within the domain where they are defined. Let $\mathbf{R}$ represent the entire space and define the domain of the second derivative of f as follows:

$$\mathrm{dom}(g'') = \mathrm{dom}(g_1'') \cap \mathrm{dom}(g_2'') \cap \cdots \cap \mathrm{dom}(g_k''). \tag{4}$$

Therefore, for any given $z \in \mathrm{dom}(g'')$, we have the following bounds:

$$\|g_i'(z)\|_2 \leq M_i, \quad \|g_i''(z)\|_2 \leq C_i, \quad \|\omega_i\|_2 \leq W_i, \tag{5}$$

where M_i, C_i, W_i $i = 1, \ldots, k$, are positive constants. The first derivative of z_k with respect to x in terms of z_{k-1} can be expressed as:

$$\frac{\partial z_k}{\partial x} = \omega_k \cdot \mathrm{diag}\left(g_{k-1}'\left(z_{k-1}\right)\right) \cdot \frac{\partial z_{k-1}}{\partial x}. \tag{6}$$

And the second derivative of z_k with respect to x in terms of z_{k-1} can be expressed as:

$$\begin{aligned} \frac{\partial^2 z_k}{\partial x^2} &= \omega_k \cdot \left[\mathrm{diag}\left(g_{k-1}''\left(z_{k-1}\right)\right) \cdot \left(\frac{\partial z_{k-1}}{\partial x}\right)^2\right] \\ &\quad + \omega_k \cdot \left[\mathrm{diag}\left(g_{k-1}'\left(z_{k-1}\right)\right) \cdot \frac{\partial^2 z_{k-1}}{\partial x^2}\right]. \end{aligned} \tag{7}$$

Now we use the recursive method to prove that f'' is bounded. For the derivative at the first layer:

$$\left\|\frac{\partial z_1}{\partial x}\right\|_2 = \|\omega_1\|_2 \leq W_1, \quad \left\|\frac{\partial^2 z_1}{\partial x^2}\right\|_2 = 0. \tag{8}$$

Assume that for z_{i-1}, its first and second derivatives are bounded:

$$\left\|\frac{\partial z_{i-1}}{\partial x}\right\|_2 \leq U_{i-1}, \quad \left\|\frac{\partial^2 z_{i-1}}{\partial x^2}\right\|_2 \leq V_{i-1}. \tag{9}$$

Then the first derivative of the i-th layer satisfies:

$$\begin{aligned} \left\|\frac{\partial z_i}{\partial x}\right\|_2 &\leq \|\omega_i\|_2 \cdot \left\|g_{i-1}'\left(z_{i-1}\right)\right\|_2 \cdot \left\|\frac{\partial z_{i-1}}{\partial x}\right\|_2 \\ &= W_i \cdot M_{i-1} \cdot U_{i-1}. \end{aligned} \tag{10}$$

The second derivative also satisfies:

$$\begin{aligned} \left\|\frac{\partial^2 z_i}{\partial x^2}\right\|_2 &\leq \|\omega_i\|_2 \cdot \left\|g_{i-1}''\left(z_{i-1}\right)\right\|_2 \cdot \left\|\frac{\partial z_{i-1}}{\partial x}\right\|_2^2 \\ &\quad + \|\omega_i\|_2 \cdot \left\|g_{i-1}'\left(z_{i-1}\right)\right\|_2 \cdot \left\|\frac{\partial^2 z_{i-1}}{\partial x^2}\right\|_2 \\ &= W_i \cdot \left(C_{i-1} \cdot U_{i-1}^2 + M_{i-1} \cdot V_{i-1}\right). \end{aligned} \tag{11}$$

Let

$$\begin{aligned} U_i &= W_i \cdot M_{i-1} \cdot U_{i-1}, \\ V_i &= W_i \cdot \left(C_{i-1} \cdot U_{i-1}^2 + M_{i-1} \cdot V_{i-1}\right), \end{aligned} \tag{12}$$

then we have:

$$\left\|\frac{\partial z_i}{\partial x}\right\|_2 \leq U_i, \quad \left\|\frac{\partial^2 z_i}{\partial x^2}\right\|_2 \leq V_i. \tag{13}$$

In this way we have proven that $f''(x)$ is bounded within its domain $\mathrm{dom}(g'')$.

Proof that the inequality holds almost everywhere. Let $\mu(\cdot)$ denote the measure of sets, then we need to prove $\mu(\mathrm{dom}(g'')) = \mu(\mathbf{R})$. Reviewing those non-smooth activation functions, their derivatives are only undefined at specific points, and the set of such points is very sparse in

Activation Function	Mathematical Form	First Order Derivative Properties	Second Order Derivative Properties
ReLU	$g(x) = \max(0, x)$	$g'(x) = 1 \cdot I(x > 0) + 0 \cdot I(x < 0)$	$g''(x) = 0$ (except at $x = 0$)
Sigmoid	$g(x) = \frac{1}{1+e^{-x}}$	$\left\|g'(x)\right\| \leq \frac{1}{4}$	$\left\|g''(x)\right\| \leq \frac{1}{4}$
Tanh	$g(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	$\left\|g'(x)\right\| \leq 1$	$\left\|g''(x)\right\| \leq 2$
Leaky ReLU	$g(x) = x \cdot I(x > 0) + \alpha x \cdot I(x \leq 0)$	$g'(x) = 1 \cdot I(x > 0) + \alpha \cdot I(x < 0)$	$g''(x) = 0$ (except at $x = 0$)
Swish	$g(x) = x \cdot \frac{1}{1+e^{-x}}$	$\left\|g'(x)\right\| < 2$	$\left\|g''(x)\right\| < 2$
ELU	$g(x) = x \cdot I(x > 0) + \alpha(e^{x} - 1) \cdot I(x \leq 0)$	$\left\|g'(x)\right\| \leq \max(0, \alpha)$	$g''(x) \leq \|\alpha\|$ (except at $x = 0$)
GELU	$g(x) = x \cdot \Phi(x)$	$\left\|g'(x)\right\| \leq 1$	$\left\|g''(x)\right\| \leq 0.5$
Mish	$g(x) = x \cdot \tanh\left(\ln\left(1 + e^{x}\right)\right)$	$\left\|g'(x)\right\| \leq 1.5$	$\left\|g''(x)\right\| \leq 2$
Softplus	$g(x) = \ln\left(1 + e^{x}\right)$	$\left\|g'(x)\right\| \leq 1$	$\left\|g''(x)\right\| \leq \frac{1}{4}$
Softmax	$g_i(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}$	$\left\|g'_i(x)\right\| \leq 1$	$\left\|g''_i(x)\right\| \leq 1$

Table 1. Properties of commonly used activation functions and their first and second derivatives.

high-dimensional space. Take ReLU as an example. The points where its derivative is not defined occur only when the input is zero or when the input is mapped to zero after a linear transformation. Given that the weights in neural networks are typically close to full rank, the set of inputs where ReLU's derivative is undefined forms a lower-dimensional manifold, which is sparse in high-dimensional space. Consequently, the measure of this set is zero. Therefore $\mu(\mathcal{R} \setminus \mathrm{dom}(g''_i)) = 0$, $i = 1, \dots, k$. Since

$$\mathbf{R} \setminus \mathrm{dom}(g'') = (\mathbf{R} \setminus \mathrm{dom}(g''_1)) \cup \cdots \cup (\mathbf{R} \setminus \mathrm{dom}(g''_k)), \quad (14)$$

then we can obtain that

$$\begin{aligned} \mu(\mathbf{R} \setminus \mathrm{dom}(g'')) \leq \mu(\mathbf{R} \setminus \mathrm{dom}(g''_1)) + \cdots \\ + \mu(\mathbf{R} \setminus \mathrm{dom}(g_k)) = 0. \end{aligned} \quad (15)$$

Therefore, we can conclude that the region where f'' is undefined forms a set of measure zero, and the inequality for f'' is satisfied almost everywhere. Therefore, the function $s(\cdot)$ used to compute the score of latent variables in this paper also satisfies this condition, meaning that there exists a constant c so that for any ξ in the Gaussian space, we have $\|\nabla^2_{\xi} s(\xi)\|_2 \leq c$ almost everywhere.

B. Hyper-parameters

B.1. Number oF Denoising Timesteps

To investigate the impact of the number of denoising timesteps T. We select the cases of $T = 10, 20, 50, 80, 100$ and use the prompt "an apple and a pear" as an example, conducting the experiment with 25 random seeds. The quantitative results are shown in Fig. 1. Regardless of the number of timesteps, our method consistently leads to significant improvements. Generally speaking, increasing the number of timesteps can enhance the VQA score of the initial images; however, the VQA score after increasing the timesteps still leaves room for improvement. Fig. 2 shows an extreme example: increasing T from 10 to 100 does not result in a substantial change in the image content. In contrast, after optimizing the noisy latent variable, the generated images have a significant improvement in alignment

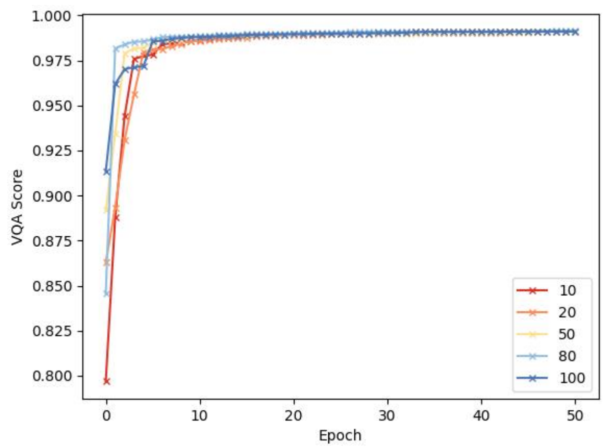

Figure 1. The average VQA score of the generated images across different epochs under various settings of the number of denoising timesteps.

with the prompt across different settings. This demonstrates that our method enhances the faithfulness across different numbers of timesteps. When it comes to the choice of T, selecting smaller values can significantly speed up inference time while still yielding satisfactory images. If speed is the primary concern, one could opt for $T = 20$ or even $T = 10$. However, to balance stability with performance, $T = 50$ is a recommended choice based on the current performance of diffusion models.

B.2. Number of Candidate Noises

The number of candidate noises N should be large enough to ensure that the optimization progresses in the direction of increasing the VQA score. Since the latent variable exists in a high-dimensional space, it is unlikely to find a noise that leads to the step difference perfectly aligning with the gradient through random sampling. Instead, the criterion for setting N is to ensure that the lower bound is sufficiently large. We set N to different values: 10, 20, 50, 80 and 100, and observe the statistical characteristics of $\nabla_{z_T} s(z_T) v / \|v\|_2^2$ under these settings. As shown in Fig. 3, when N increases from 10 to 20, the lower quartile (Q1) im-

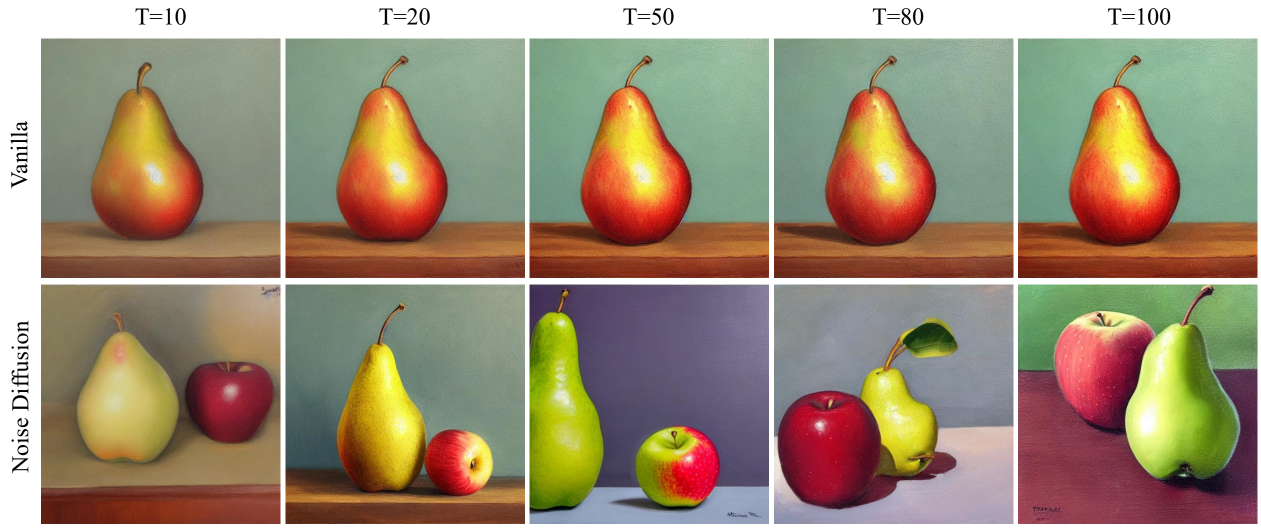

Figure 2. An example of an extreme case. Increasing the number of denoising timesteps does not lead to an improvement in the alignment of the generated image with the prompt. However, under different settings of the number of timesteps, our method consistently enhances the faithfulness.

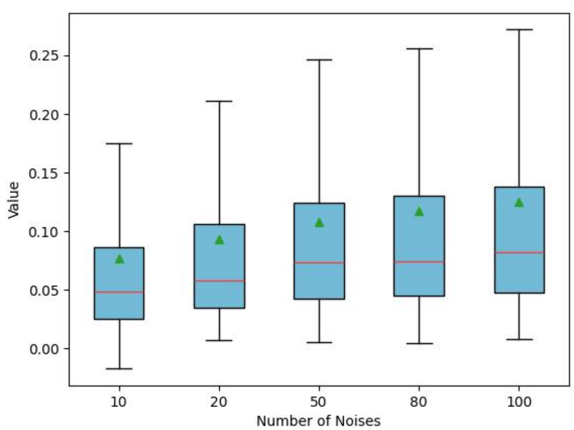

Figure 3. The statistical characteristics of $\nabla_{z_T} s(z_T)v/\|v\|_2^2$ with different values of N during the optimization.

proves from 0.026 to 0.035. When N reaches 50, it further increases to 0.043, demonstrating the effect of increasing the number of candidate noises. As N increases to 100, Q1 reaches 0.048. Although the rate of improvement begins to slow down, increasing N further will still lead to continued growth. Therefore, although we set N as 50 in our experiments, increasing N undoubtedly has a positive effect on the optimization.

C. Datasets

We present the prompt datasets used in our main experiment. Table. 2 shows the dataset for simple cases, which includes a total of 276 prompts, categorized into three groups: animals, animals-objects, and objects. The complex cases are shown in Table. 3, where we have a total of 100 prompts, which similarly include animals, animals-objects and objects, but with additional spatial relationships such as "on," "under," "above," and "below." These added relationships increase the complexity of the prompts, forcing text-to-image models to better understand and reflect the logical structure within the prompts.

D. More Results

We present additional results for vanilla settings and our method on SD V-1.4, V-1.5, V-2.0 and V-2.1 in Fig. 4, Fig. 5, Fig. 6 and Fig. 7, respectively.

an elephant and a rabbit	a dog and a frog	a bird and a mouse	a monkey and a frog	a horse and a monkey	a bird and a turtle
a bird and a lion	a lion and a monkey	a horse and a turtle	a bird and a monkey	a bear and a frog	a bear and a turtle
a dog and an elephant	a dog and a horse	a turtle and a mouse	a cat and a turtle	a dog and a mouse	a cat and an elephant
a cat and a bird	a dog and a monkey	a lion and a mouse	a bear and a lion	a bird and an elephant	a lion and a turtle
a dog and a bird	a bird and a rabbit	an elephant and a turtle	a lion and an elephant	a cat and a rabbit	a dog and a bear
a dog and a rabbit	a cat and a bear	a bird and a horse	a rabbit and a mouse	a bird and a bear	a bear and a monkey
a horse and a frog	a cat and a horse	a frog and a rabbit	a bear and a mouse	a monkey and a rabbit	a cat and a dog
a lion and a frog	a frog and a mouse	a dog and a lion	a lion and a rabbit	an elephant and a frog	a frog and a turtle
a cat and a lion	a horse and a rabbit	a cat and a monkey	a bear and a rabbit	a turtle and a rabbit	an elephant and a monkey
a bird and a frog	a lion and a horse	a bear and a horse	a bear and an elephant	a horse and a mouse	a dog and a turtle
a monkey and a mouse	a cat and a frog	a monkey and a turtle	a horse and an elephant	a cat and a mouse	an elephant and a mouse
a horse with a glasses	a bear with a glasses	a monkey and a red car	an elephant with a bow	a frog and a purple balloon	a mouse with a bow
a bird with a crown	a turtle and a yellow bowl	a rabbit and a gray chair	a dog and a black apple	a rabbit and a white bench	a lion and a yellow clock
a turtle and a gray backpack	an elephant and a green balloon	a monkey and a orange apple	a lion and a red car	a lion with a crown	a bird and a purple bench
a rabbit and a orange backpack	a rabbit and a orange apple	a monkey and a green bowl	a frog and a red suitcase	a monkey and a green balloon	a cat with a glasses
a bear and a blue clock	a cat and a gray bench	a bear with a crown	a lion with a bow	a bear and a red balloon	a bird and a black backpack
a horse and a pink balloon	a turtle and a yellow car	a lion with a glasses	a cat and a yellow balloon	a horse and a yellow clock	a dog with a glasses
a horse and a blue backpack	a frog with a bow	an elephant with a glasses	a mouse and a red bench	a bird and a brown balloon	a monkey and a yellow backpack
a turtle and a pink balloon	a cat and a red apple	a monkey and a brown bench	a rabbit with a glasses	a bear and a gray bench	a turtle and a blue clock
a monkey and a blue chair	a turtle and a blue chair	a turtle with a bow	an elephant and a black chair	a mouse and a purple chair	a bear and a white car
a lion and a black backpack	a dog with a crown	a horse and a green apple	a dog and a gray clock	a dog and a purple car	a dog and a gray bowl
a monkey with a bow	a mouse and a blue clock	a bird and a black bowl	a horse and a white car	a mouse and a pink apple	a bear and a orange backpack
an elephant and a yellow clock	a bird and a green chair	a mouse and a black balloon	a turtle and a white bench	a bird with a bow	a turtle with a crown
a bird and a yellow car	a frog and a orange car	a dog and a pink bench	a frog with a crown	a frog and a green bowl	a frog and a pink bench
a horse with a bow	a bird and a yellow apple	a monkey with a crown	a cat and a blue backpack	a turtle and a pink apple	a dog and a orange chair
a horse and a green suitcase	an elephant with a crown	a monkey and a orange suitcase	a turtle and a orange suitcase	a lion and a gray apple	a mouse with a crown
a mouse with a glasses	a horse and a brown bowl	a monkey and a yellow clock	a turtle with a bow	a dog and a brown backpack	a cat and a purple bowl
a lion and a white bench	a rabbit and a blue bowl	a lion and a brown balloon	a horse and a pink chair	an elephant and a green bench	a rabbit and a white balloon
an elephant and a pink backpack	a lion and a orange suitcase	an elephant and a orange apple	an elephant and a green suitcase	a horse with a crown	a bear with a bow
a rabbit and a yellow suitcase	a horse and a blue bench	a dog and a green suitcase	a mouse and a red car	a cat and a black chair	a bear and a red suitcase
a rabbit and a gray clock	a bear and a pink apple	a lion and a white chair	a rabbit with a crown	a mouse and a purple bowl	a frog and a black apple
a rabbit with a bow	a mouse and a pink suitcase	a lion and a pink bowl	a frog and a black chair	a frog and a green clock	a bear and a white chair
an elephant and a brown car	a turtle with a glasses	a cat and a black suitcase	a cat and a yellow car	a frog and a yellow backpack	a bird and a black suitcase
a cat with a crown	a rabbit and a yellow car	a cat with a bow	a bird and a white clock	a cat and a green clock	a bear and a purple bowl
a monkey with a glasses	a frog with a glasses	an elephant and a green bowl	a bird with a glasses	a dog and a blue balloon	a mouse and a brown backpack
a pink crown and a purple bow	a blue clock and a blue apple	a blue balloon and a orange bench	a pink crown and a red chair	a orange chair and a blue clock	a purple bowl and a black bench
a green glasses and a black crown	a purple chair and a red bow	a yellow glasses and a black car	a orange backpack and a purple car	a white balloon and a white apple	a brown suitcase and a black clock
a yellow backpack and a purple chair	a gray backpack and a green clock	a blue crown and a red balloon	a gray suitcase and a black bowl	a brown balloon and a pink car	a black backpack and a green bow
a blue balloon and a blue bow	a orange bowl and a purple apple	a orange bowl and a purple apple	a brown chair and a white bench	a purple crown and a blue suitcase	a yellow bow and a orange bench
a yellow glasses and a brown bow	a red glasses and a red suitcase	a pink bow and a gray apple	a gray crown and a white clock	a black car and a white clock	a brown bowl and a green clock
a green backpack and a yellow crown	a orange glasses and a pink clock	a purple chair and a orange bowl	a orange suitcase and a brown bench	a white glasses and a orange balloon	a yellow backpack and a gray apple
a green bench and a red apple	a gray backpack and a yellow glasses	a green glasses and a yellow chair	a white glasses and a gray apple	a gray suitcase and a brown bow	a white car and a black bowl
a purple car and a pink apple	a gray crown and a purple apple	a orange car and a red bench	a red suitcase and a blue apple	a red backpack and a yellow bowl	a red bench and a yellow clock
a black backpack and a pink balloon	a blue suitcase and a gray balloon	a yellow glasses and a gray bowl	a white suitcase and a white chair	a purple crown and a blue bench	a yellow bow and a pink bowl
a green backpack and a brown suitcase	a green glasses and a black bench	a white bow and a black clock	a red crown and a black bowl	a green chair and a purple car	a white chair and a gray balloon
a pink chair and a gray apple	a yellow suitcase and a yellow car	a green backpack and a purple bench	a black crown and a red car	a green balloon and a pink bowl	a purple balloon and a white clock

Table 2. Prompt dataset for simple cases.

an elephant under a rabbit	a dog under a frog	a monkey under a frog	a bird on a turtle	a bird above a lion
a bird on a monkey	a dog on an elephant	a cat on an elephant	a dog under a monkey	a lion under a mouse
a bird above an elephant	a lion under a turtle	a dog below a bird	a bird above a rabbit	an elephant under a turtle
a cat under a rabbit	a dog on a bear	a cat above a bear	a bird on a horse	a bear under a monkey
a cat on a horse	a frog under a mouse	a dog on a lion	an elephant under a frog	a cat on a lion
a horse under a rabbit	a cat below a monkey	an elephant below a monkey	a bird above a frog	a cat above a frog
a monkey under a turtle	a cat above a mouse	an elephant under a mouse	a monkey on a red car	a frog below a purple balloon
a turtle under a yellow bowl	a rabbit on a gray chair	a rabbit under a white bench	a lion below a yellow clock	an elephant below a green balloon
a monkey above a green bowl	a bear below a blue clock	a bear under a crown	a mouse under a red bench	a bear under a gray bench
a turtle below a blue clock	a monkey on a blue chair	a bear below an orange backpack	a bird on a green chair	a mouse below a black balloon
a turtle on a white bench	a bird above a yellow car	a monkey under a crown	a cat above a blue backpack	a dog on an orange chair
a monkey on an orange suitcase	a mouse below a crown	a horse under a brown bowl	a monkey below a yellow clock	a lion on a white bench
a rabbit under a blue bowl	a dog above a green suitcase	a cat under a black chair	a lion under a pink bowl	a frog on a black chair
a frog on a green clock	a pink crown under a purple bow	a blue clock under a blue apple	a blue balloon on an orange bench	a pink crown under a red chair
a purple bowl on a black bench	an orange backpack on a purple car	a brown suitcase below a black clock	a gray backpack under a green clock	a brown balloon above a pink car
a purple crown on a blue suitcase	a yellow bow above an orange bench	a pink bow on a gray apple	a gray crown below a white clock	a black car under a white clock
a brown bowl below a green clock	a green backpack under a yellow crown	a purple chair under an orange bowl	an orange suitcase under a brown bench	a yellow backpack under a gray apple
a green bench under a red apple	a gray suitcase under a brown bow	a white car under a black bowl	a purple car under a pink apple	a gray crown under a purple apple
a red suitcase under a blue apple	a red backpack under a yellow bowl	a black backpack below a pink balloon	a blue suitcase below a gray balloon	a white suitcase on a white chair
a green backpack under a brown suitcase	a white bow on a black clock	a red crown under a black bowl	a yellow suitcase on a yellow car	a green balloon above a pink bowl

Table 3. Prompt dataset for complex cases.

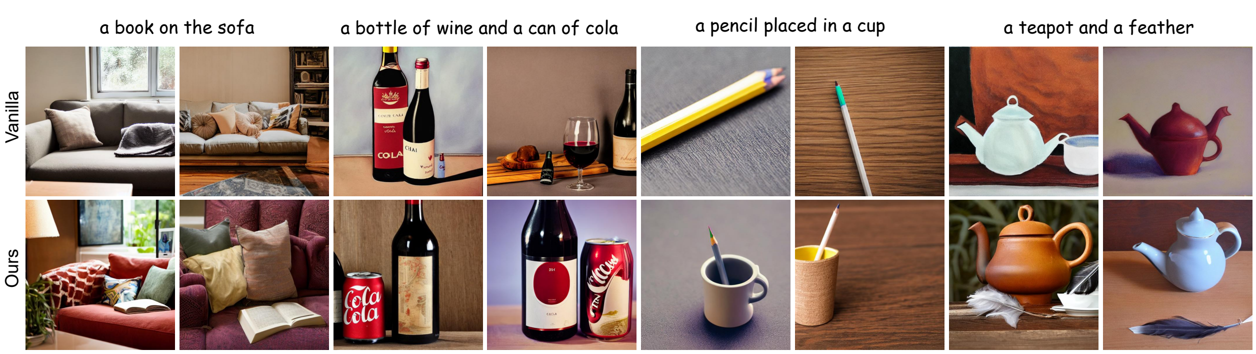

Figure 4. Outputs of Vanilla settings and our method for SD V-1.4.

Figure 5. Outputs of Vanilla settings and our method for SD V-1.5.

Figure 6. Outputs of Vanilla settings and our method for SD V-2.0.

Figure 7. Outputs of Vanilla settings and our method for SD V-2.1.