[Supplementary Material]
DiffQRCoder: Diffusion-based Aesthetic QR Code Generation with Scanning Robustness Guided Iterative Refinement

Jia-Wei Liao[1,2] Winston Wang[1*] Tzu-Sian Wang[1*] Li-Xuan Peng[1*] Ju-Hsuan Weng[1,2]
Cheng-Fu Chou[2] Jun-Cheng Chen[1]

[1] Research Center for Information Technology Innovation, Academia Sinica,
[2] National Taiwan University

A. Grayscale Conversion

We denote $\mathcal{G}(\cdot)$ as the grayscale operator, which is defined as:

$$\mathcal{G}(\mathbf{x}) = c_r\mathbf{x}^r + c_g\mathbf{x}^g + c_b\mathbf{x}^b,$$

where $\mathbf{x}^r$, $\mathbf{x}^g$ and $\mathbf{x}^b$ are R, G and B channels of the image $\mathbf{x}$, respectively. The coefficients $c_r = 0.299$, $c_g = 0.587$, and $c_b = 0.114$ are chosen according to the YCbCr color space standards for grayscale conversion.

B. Scanning Robust Perceptual Guidance (SRPG)

B.1. Learned Perceptual Image Patch Similarity (LPIPS)

Traditional image-level similarity metrics, which typically compare pixels directly, often fail to align with human perception. To address this issue, Zhang et al. [11] employed the pre-trained NN-based feature extractors, such as VGG and AlexNet, to transform images into a feature space for comparison. Given our focus on assessing "aesthetics," a high-level and abstract semantic concept, we employ LPIPS for a more appropriate evaluation. LPIPS loss $\mathcal{L}_{\text{LPIPS}}(\mathbf{x}, \hat{\mathbf{x}})$ is defined as:

$$\mathcal{L}_{\text{LPIPS}}(\mathbf{x}, \hat{\mathbf{x}}) = \sum_{l,i,j} \frac{1}{h_l w_l} \|\omega^l \odot (\psi^l(\mathbf{x})_{i,j} - \psi^l(\hat{\mathbf{x}})_{i,j})\|_2^2,$$

where $\psi^l(\cdot)$ denotes features extracted from the l-th layer, $(h_l,\ w_l)$ are the height and width of $\psi^l(\mathbf{x})$, and ω^l is a channel-wise scaling vector.

B.2. Derivation of Conditional Probability Term in Generalized Classifier Guidance

Song et al. [8, 9] established a connection between the score function $\nabla_{\mathbf{z}_t} \log p(\mathbf{z}_t)$ and the noise estimation function $\epsilon_\theta(\mathbf{z}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}})$ via Tweedie's Formula [4]

$$\epsilon_\theta(\mathbf{z}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}}) = -\sqrt{1-\bar{\alpha}_t}\nabla_{\mathbf{z}_t} \log p(\mathbf{z}_t). \tag{1}$$

Inspired by [3], to perform conditional sampling, we substitute the score function with a conditional probability term $p(\mathbf{z}_t|\mathbf{y})$. Then we rewrite the conditional probability term using Bayes' Theorem. Specifically, we define the updated score estimate $\hat{\epsilon}_t$ with condition $\mathbf{y}$ at timestep t as:

$$\begin{aligned}
\hat{\epsilon}_t &:= -\sqrt{1-\bar{\alpha}_t}\nabla_{\mathbf{z}_t} \log p(\mathbf{z}_t|\mathbf{y}) \\
&= -\sqrt{1-\bar{\alpha}_t}\nabla_{\mathbf{z}_t} \log\left(\frac{p(\mathbf{z}_t)p(\mathbf{y}|\mathbf{z}_t)}{p(\mathbf{y})}\right) \\
&= -\sqrt{1-\bar{\alpha}_t}\left(\nabla_{\mathbf{z}_t} \log p(\mathbf{z}_t) + \nabla_{\mathbf{z}_t} \log p(\mathbf{y}|\mathbf{z}_t)\right) \\
&= \epsilon_\theta(\mathbf{z}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}}) - \sqrt{1-\bar{\alpha}_t}\nabla_{\mathbf{z}_t} \log p(\mathbf{y}|\mathbf{z}_t).
\end{aligned}$$

Following [1], we define F as the guidance function, thus the final updated estimated score becomes:

$$\hat{\epsilon}_t = \epsilon_\theta(\mathbf{z}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}}) + \sqrt{1-\bar{\alpha}_t}\nabla_{\mathbf{z}_t} F(\mathbf{z}_t, \mathbf{y}). \tag{2}$$

B.3. Derivation of SRPG Gradient

In this section, we derive the gradient of the guidance function. Given the expression

$$\tilde{\mathbf{x}}_{0|t} = \mathcal{D}_\theta\left(\frac{1}{\sqrt{\bar{\alpha}_t}}\left(\tilde{\mathbf{z}}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\tilde{\mathbf{z}}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}})\right)\right), \tag{3}$$

which involves the VAE decoder calculation, we must apply the Chain Rule to derive the gradient.

Consequently, the gradient of our proposed generalized classifier guidance function F_{SRP} can be derived as follows:

$$\begin{aligned}
\nabla_{\tilde{\mathbf{z}}_t} F_{\text{SRP}}(\tilde{\mathbf{z}}_t, \tilde{\mathbf{y}}, \hat{\mathbf{x}}) &= \lambda_1 \nabla_{\tilde{\mathbf{z}}_t}\mathcal{L}_{\text{SR}}(\tilde{\mathbf{x}}_{0|t}, \tilde{\mathbf{y}}) + \lambda_2 \nabla_{\tilde{\mathbf{z}}_t}\mathcal{L}_{\text{LPIPS}}(\tilde{\mathbf{x}}_{0|t}, \hat{\mathbf{x}}) \\
&= \left(\lambda_1 \frac{\partial \mathcal{L}_{\text{SR}}(\tilde{\mathbf{x}}_{0|t}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{x}}_{0|t}} + \lambda_2 \frac{\partial \mathcal{L}_{\text{LPIPS}}(\tilde{\mathbf{x}}_{0|t}, \hat{\mathbf{x}})}{\partial \tilde{\mathbf{x}}_{0|t}}\right) \cdot \frac{\partial \mathcal{D}_\theta(\tilde{\mathbf{z}}_{0|t})}{\partial \tilde{\mathbf{z}}_{0|t}} \cdot \\
&\qquad \frac{1}{\sqrt{\bar{\alpha}_t}}\left(1 - \sqrt{1-\bar{\alpha}_t}\frac{\partial \epsilon_\theta(\tilde{\mathbf{z}}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}})}{\partial \tilde{\mathbf{z}}_t}\right).
\end{aligned}$$

[*] Equal contributions.

where $\hat{\mathbf{x}}$ indicates the reference image generated from Stage-1.

Finally, substitute the conditional score term with F_{SRP}, the estimated score at timestep t becomes:

$$\hat{\epsilon}_t = \epsilon_\theta(\tilde{\mathbf{z}}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}}) + \sqrt{1-\bar{\alpha}_t}\nabla_{\tilde{\mathbf{z}}_t} F_{\text{SRP}}(\tilde{\mathbf{z}}_t, \tilde{\mathbf{y}}, \hat{\mathbf{x}}). \quad (4)$$

C. Details of Our Proposed Two-stage QR Code Generation Pipeline

C.1. Qart

Qart [2] transforms traditional QR codes with user-specified target patterns by exploiting the padding modules. We leverage its capability to create similar patterns of the reference image $\hat{\mathbf{x}}$ from Stage-1 and the target QR code $\mathbf{y}$, forming a better target QR code $\tilde{\mathbf{y}}$ for the Stage-2 ControlNet conditioning.

C.2. Two-stage QR Code Generation Algorithm

Algorithm 1 Two-stage QR Code Generation Pipeline with Iterative Refinement

```
Input: QR code image $\mathbf{y}$, prompt embedding $\mathbf{e}_p$, QR code image embedding $\mathbf{e}_{\text{code}}$, UNet $\epsilon_\theta(\cdot,\cdot,\cdot,\cdot)$, VAE encoder $\mathcal{E}_\theta(\cdot)$ VAE decoder $\mathcal{D}_\theta(\cdot)$, sequence $\{\bar{\alpha}_t\}_{t=1}^T$, guided weights $\lambda_1, \lambda_2 > 0$, error rate $\mathcal{E}(\cdot,\cdot)$, and QR code error correction capacity $\tau$.
$\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.                                  ▷ Stage-1
for $t = T$ to 1 do
    $\hat{\epsilon} \leftarrow \epsilon_\theta(\mathbf{z}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}})$.
    $\mathbf{z}_{t-1} \leftarrow \sqrt{\frac{\bar{\alpha}_{t-1}}{\bar{\alpha}_t}}\left(\mathbf{z}_t - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\hat{\epsilon}$.
end for
$\hat{\mathbf{x}} \leftarrow \mathcal{D}_\theta(\mathbf{z}_0)$.
$\tilde{\mathbf{y}} \leftarrow \text{Qart}(\hat{\mathbf{x}}, \mathbf{y})$.
$\tilde{\mathbf{z}}_T \leftarrow \sqrt{\bar{\alpha}_T}\mathcal{E}(\hat{\mathbf{x}}) + \sqrt{1-\bar{\alpha}_T}\epsilon_T, \epsilon_T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.   ▷ Stage-2
for $t = T$ to 1 do
    $\tilde{\mathbf{z}}_{0|t} \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\tilde{\mathbf{z}}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\tilde{\mathbf{z}}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}})\right)$.
    $\tilde{\mathbf{x}}_{0|t} \leftarrow \mathcal{D}_\theta(\tilde{\mathbf{z}}_{0|t})$.
    if $\mathcal{E}(\tilde{\mathbf{x}}_{0|t}, \tilde{\mathbf{y}}) \geq \tau$ then
        $F_{\text{SRP}}(\tilde{\mathbf{z}}_t, \tilde{\mathbf{y}}, \hat{\mathbf{x}}) \leftarrow \lambda_1\mathcal{L}_{\text{SR}}(\tilde{\mathbf{x}}_{0|t}, \tilde{\mathbf{y}}) + \lambda_2\mathcal{L}_{\text{LPIPS}}(\tilde{\mathbf{x}}_{0|t}, \hat{\mathbf{x}})$.
        $\hat{\epsilon}_t \leftarrow \epsilon_\theta(\tilde{\mathbf{z}}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}}) + \sqrt{1-\bar{\alpha}_t}\nabla_{\tilde{\mathbf{z}}_t} F_{\text{SRP}}(\tilde{\mathbf{z}}_t, \mathbf{y}, \hat{\mathbf{x}})$.
    else
        $\hat{\epsilon}_t \leftarrow \epsilon_\theta(\tilde{\mathbf{z}}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}})$.
    end if
    $\tilde{\mathbf{z}}_{t-1} \leftarrow \sqrt{\frac{\bar{\alpha}_{t-1}}{\bar{\alpha}_t}}\left(\tilde{\mathbf{z}}_t - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\hat{\epsilon}$.
end for
$\mathbf{x}_0 \leftarrow \mathcal{D}_\theta(\tilde{\mathbf{z}}_0)$.
return $\mathbf{x}_0$.
```

D. More Details of Experiments

Our implementation primarily utilizes the `diffusers` library [10] from Hugging Face. Tab. 1 outlines the parameters for the various methods used in our experiments; parameters not specified here are set to their default values.

Method	Parameters
QRBTF (Test in June 2024)	Size: 1152px Padding ratio: 0.2 Anchor style: square Correct level: 15%
QR Code AI Art	ControlNet conditioning scale: 1.1 Strength: 0.9 Guidance scale: 7.5 Sampler: DPM++ Karras SDE Seed: 6745177115
QR Diffusion	QR code weight: 1.65
QR Code Monster	ControlNet conditioning scale: 1.35

Table 1. Parameter settings in our experiments.

D.1. Implementation Details of Simulating Scanning Angles

The QR codes are randomly chosen from our generated results. These codes are then rotated by 0, 15, 30, and 45 degrees using CSS (Fig. 1). A code is considered scannable if it can be scanned within 3 seconds.

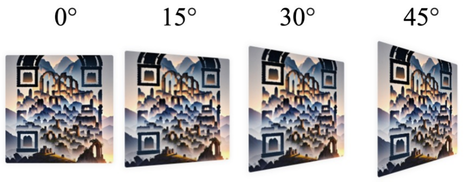


Figure 1. Visualization in different scanning angles of QR code using CSS. Zoom in for better scannability.

D.2. Implementation Details of Scanning with Different Scanners

We chose three widely used QR code scanners for scannability assessment: the built-in scanners on the iPhone 13 and Pixel 7, and the QR Verify software scanner powered by the WeChat decoding algorithm. Our experiment involves scanning 30 aesthetic QR codes ten times for each aesthetic QR code, then calculating the Scanning Success Rate (SSR).

D.3. Visualization of QR Code Module Error

We analyze the robustness of the generated results through error analysis. According to SRL, the scanning robustness can be maintained as long as the modules after sampling and binarization yield identical results as the target QR code, regardless of pixel color changes within these modules. Fig. 2 indicates that our aesthetic QR codes display irregular colors and shapes in their modules. Despite

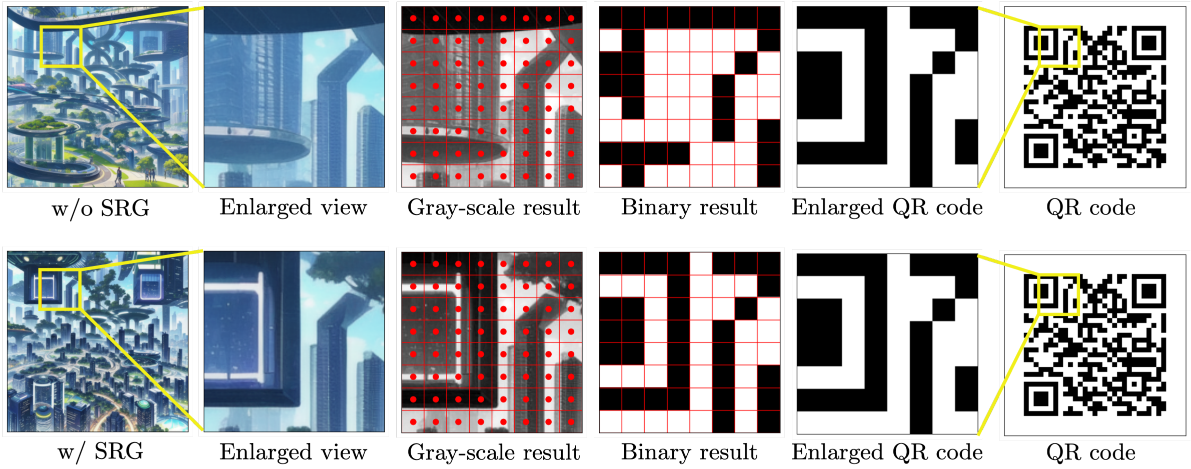


Figure 2. Visual illustration of error analysis.

undergoing sampling and binarization, the modules remain consistent with the original QR code. This suggests that our aesthetic QR codes are robust and readable by a standard QR code scanner.

D.4. Error Analysis

D.4.1 Analysis of QR Code Error Rate

The QR error rate can be computed using the following formula:

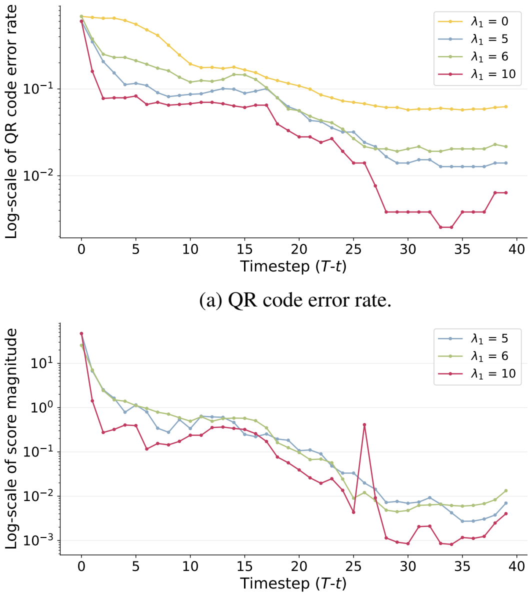
$$\mathcal{E}(\mathbf{x},\mathbf{y}) = \frac{1}{N}\sum_{k=1}^{N}\phi(\mathbf{x}_{M_k},\mathbf{y}_{M_k}), \tag{5}$$

where $\mathbf{y}$ is the target QR code, $\mathbf{x}$ is our decoded code, N is the number of modules. The function $\phi(\mathbf{x}_{M_k},\mathbf{y}_{M_k})$ measures whether the module M_k can be correctly decoded , as defined in Eq. 5 in main paper.

As illustrated in Fig. 3a, we set the perceptual guidance scale, λ_2, to 0 and examine the error rates of a sample across various scanning robust guidance scales, λ_1, during iterative refinement steps. We observe a marked reduction in error within the first five iterations under our proposed guidance. In contrast, without our guidance, i.e., when $\lambda_1 = 0$, the decrease in error occurs more gradually. Additionally, we visualize QR code errors at different timesteps to better understand the progression of error reduction, the error modules are marked in red, see Fig. 6 in main paper. Appendix D.4 further demonstrates how we visualize the error modules.

D.4.2 Analysis of the Score Magnitude

Furthermore, we analyze the change in score magnitude $\|\nabla_{\tilde{\mathbf{z}}_t} F_{\text{SRL}}(\tilde{\mathbf{z}}_t,\mathbf{y})\|_F$ across different values of λ_1. We observe that the score magnitudes decrease over the iterations, suggesting that the effects of guidance diminish over time. This trend is illustrated in Fig. 3b.



(a) QR code error rate.

(b) Score magnitude $\|\nabla_{\tilde{\mathbf{z}}_t} F_{\text{SRP}}(\tilde{\mathbf{z}}_t,\mathbf{y})\|_F$.

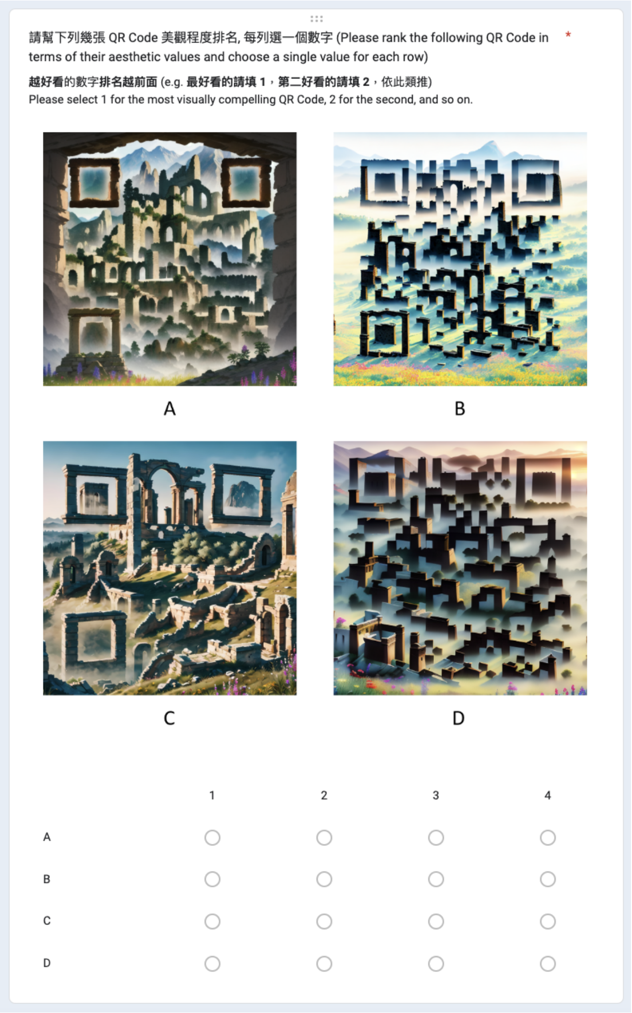
Figure 3. Error Analysis.

E. User Study

We conduct a user study with 387 participants. Our subjective test is authorized by the Academia Sinica IRB committee under the approval number AS-IRB-HS 24031.

E.1. Privacy Issues

We obtain consent from all participants before they participate in the survey. Additionally, we disclose our data processing policy, which includes the immediate destruc-



請幫下列幾張 QR Code 美觀程度排名, 每列選一個數字 (Please rank the following QR Code in terms of their aesthetic values and choose a single value for each row) *

越好看的數字**排名越前面** (e.g. **最好看的請填 1，第二好看的請填 2**，依此類推)
Please select 1 for the most visually compelling QR Code, 2 for the second, and so on.

	1	2	3	4
A	○	○	○	○
B	○	○	○	○
C	○	○	○	○
D	○	○	○	○

Figure 4. Sample question.

tion of data after compiling the statistical report, and clarify that no sensitive personal data is collected. Furthermore, participants are informed that they can withdraw from the survey at any time.

E.2. Question Details

Fig. 4 presents a sample of the questions included in our questionnaire, where participants will view four aesthetic QR codes. These codes are generated by QR Diffusion [6], QR Code AI Art [5], QRBTF [7], and our DiffQRCoder. Participants are then asked to rank the options A, B, C, and D based on their perceived aesthetic appeal.

E.3. Average Ranking Calculation

To evaluate the results of the user study, we calculate the weighted average rank for each QR code by summing the products of all ranks and their corresponding frequencies, then dividing by the total number of participants. For example, if 20 participants rank a QR code as 1, 10 participants rank it as 2, and 100 participants rank it as 3, the average rank of that QR code can be calculated as follows:

$$\frac{1 \times 20 + 2 \times 10 + 3 \times 100}{130} = 2.615.$$

F. Limitation and Future Work

Our approach showcases the significant capability of creating aesthetic QR codes, outperforming existing methods. However, it sometimes does not guarantee 100% scannability and requires hyperparameter adjustments to optimize results. To address this, we apply post-processing to refine our outputs. Our future work aims at improving the approach into a hyperparameter-insensitive and end-to-end pipeline without post-processing. Additionally, we plan to enhance controllability using image-to-image methodologies to enable more personalized aesthetic QR code generation.

G. Societal Impacts

Our proposed approach has potential vulnerabilities, including the risk of being used for phishing, spamming, or disseminating false or inappropriate content. To mitigate these risks, we can implement preventive measures such as URL filtering and prompt blacklisting.

References

[1] Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *ICLR*, 2024. 1

[2] R. Cox. Qartcodes, 2012. 2

[3] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *NeurIPS*, 2021. 1

[4] Bradley Efron. Tweedie's formula and selection bias. *J. Am. Stat. Assoc.*, 2011. 1

[5] huggingface projects. Qr-code ai art, 2023. 4

[6] QR Diffusion Inc. Qr diffusion, 2024. 4

[7] IoC Lab. Qrbtf, 2023. 4

[8] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *NeurIPS*, 2019. 1

[9] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2020. 1

[10] Patrick von Platen, Suraj Patil, Anton Lozhkov, Pedro Cuenca, Nathan Lambert, Kashif Rasul, Mishig Davaadorj, Dhruv Nair, Sayak Paul, William Berman, Yiyi Xu, Steven Liu, and Thomas Wolf. Diffusers: State-of-the-art diffusion models. `https://github.com/huggingface/diffusers`, 2022. 2

[11] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 1

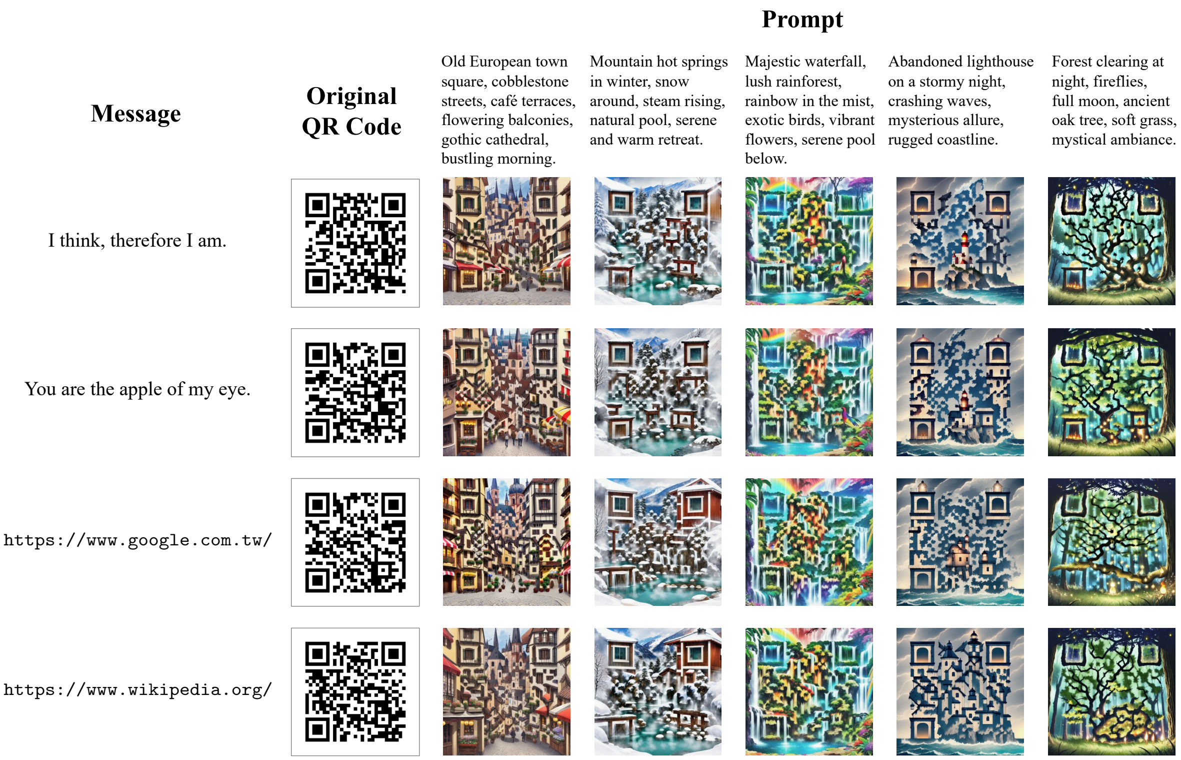

Prompt
Message
Original QR Code
Old European town square, cobblestone streets, café terraces, flowering balconies, gothic cathedral, bustling morning.
Mountain hot springs in winter, snow around, steam rising, natural pool, serene and warm retreat.
Majestic waterfall, lush rainforest, rainbow in the mist, exotic birds, vibrant flowers, serene pool below.
Abandoned lighthouse on a stormy night, crashing waves, mysterious allure, rugged coastline.
Forest clearing at night, fireflies, full moon, ancient oak tree, soft grass, mystical ambiance.
I think, therefore I am.
You are the apple of my eye.
https://www.google.com.tw/
https://www.wikipedia.org/

Figure 5. Qualitative results for different QR code messages.

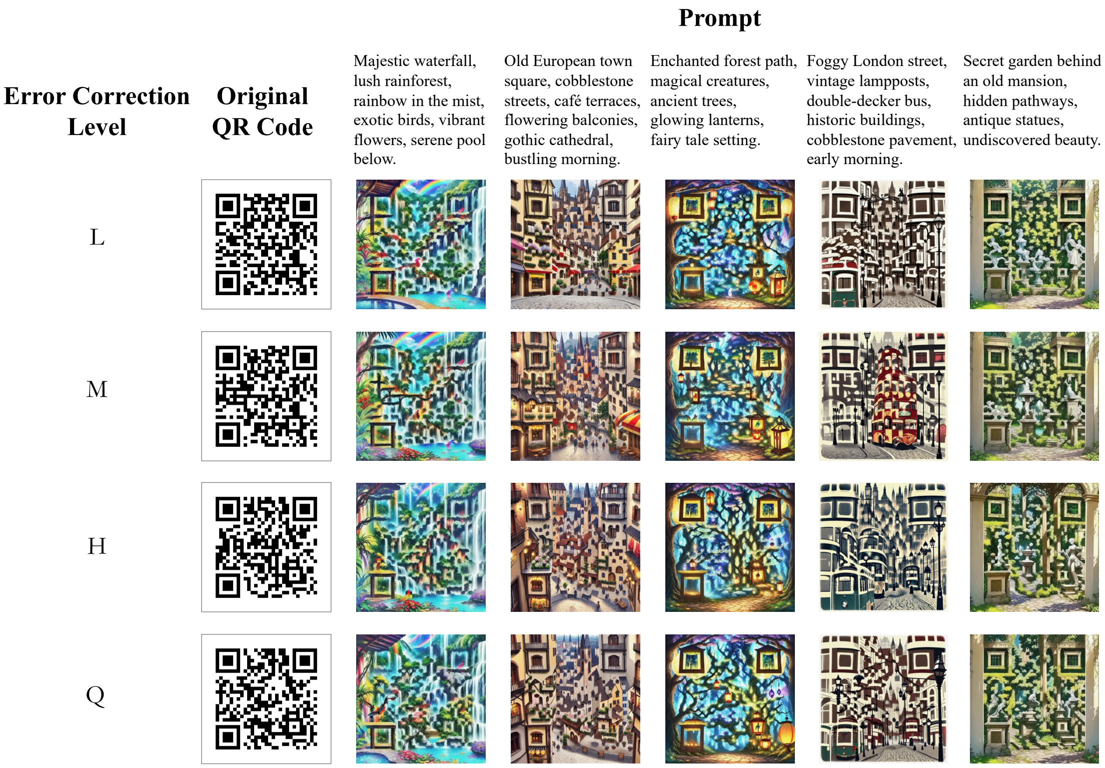
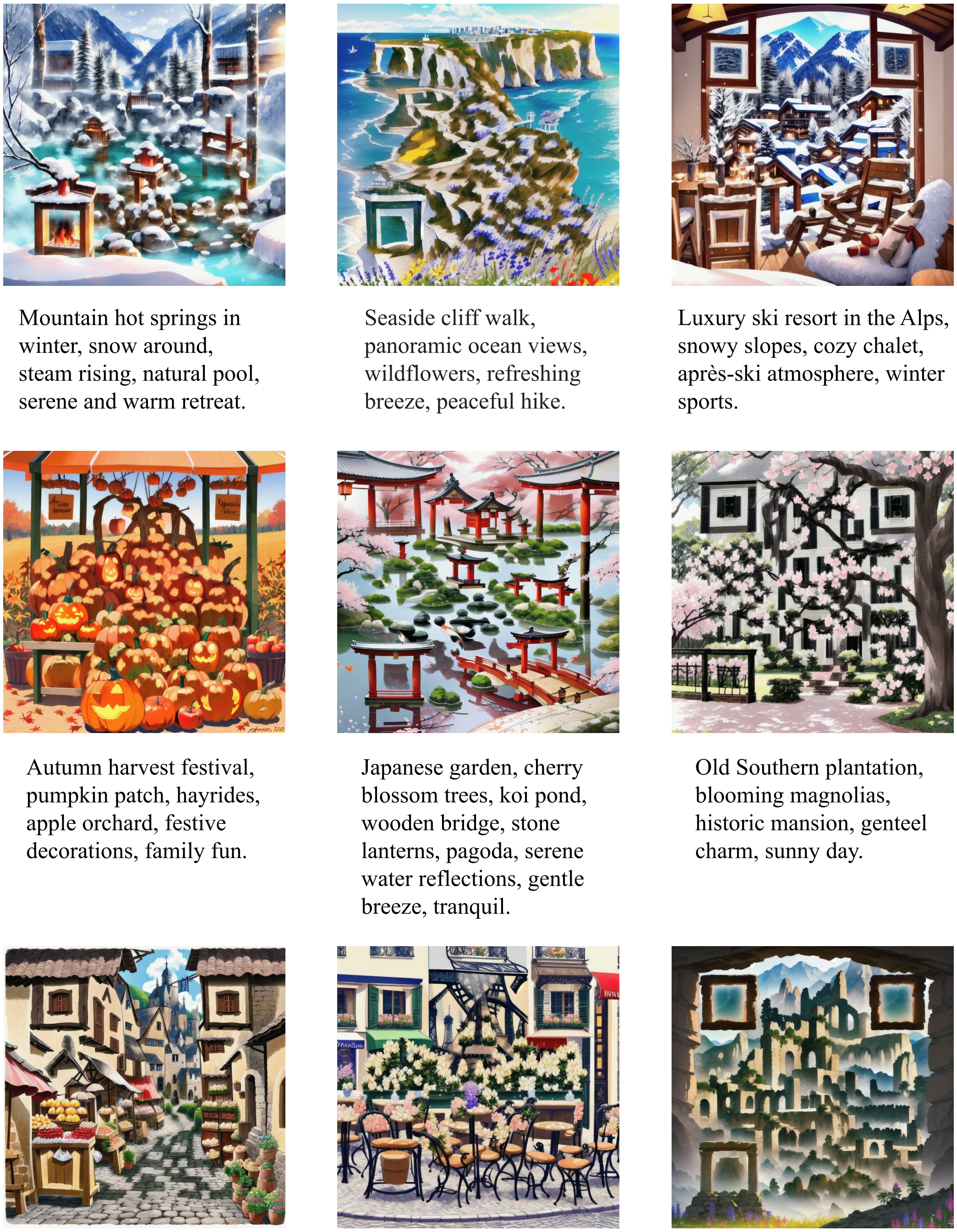

Prompt
Error Correction Level
Original QR Code
Majestic waterfall, lush rainforest, rainbow in the mist, exotic birds, vibrant flowers, serene pool below.
Old European town square, cobblestone streets, café terraces, flowering balconies, gothic cathedral, bustling morning.
Enchanted forest path, magical creatures, ancient trees, glowing lanterns, fairy tale setting.
Foggy London street, vintage lampposts, double-decker bus, historic buildings, cobblestone pavement, early morning.
Secret garden behind an old mansion, hidden pathways, antique statues, undiscovered beauty.
L
M
H
Q

Figure 6. Qualitative results for different QR code error correction levels.



Mountain hot springs in winter, snow around, steam rising, natural pool, serene and warm retreat.

Seaside cliff walk, panoramic ocean views, wildflowers, refreshing breeze, peaceful hike.

Luxury ski resort in the Alps, snowy slopes, cozy chalet, après-ski atmosphere, winter sports.

Autumn harvest festival, pumpkin patch, hayrides, apple orchard, festive decorations, family fun.

Japanese garden, cherry blossom trees, koi pond, wooden bridge, stone lanterns, pagoda, serene water reflections, gentle breeze, tranquil.

Old Southern plantation, blooming magnolias, historic mansion, genteel charm, sunny day.

Medieval village, stone cottages, cobblestone streets, market stalls, villagers in traditional attire, castle in the distance, noon, lively atmosphere.

Parisian café street in spring, outdoor tables, blooming flowers, Eiffel Tower in the distance, lively atmosphere.

Ancient city ruins at dawn, misty mountains in the background, wildflowers, crumbling stone structures, first light of day.

Figure 7. More qualitative results and corresponding prompts.