DiffQRCoder: Diffusion-based Aesthetic QR Code Generation with Scanning Robustness Guided Iterative Refinement

WACV 2025
TUCSON, ARIZONA • FEB 28 - MAR 4

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

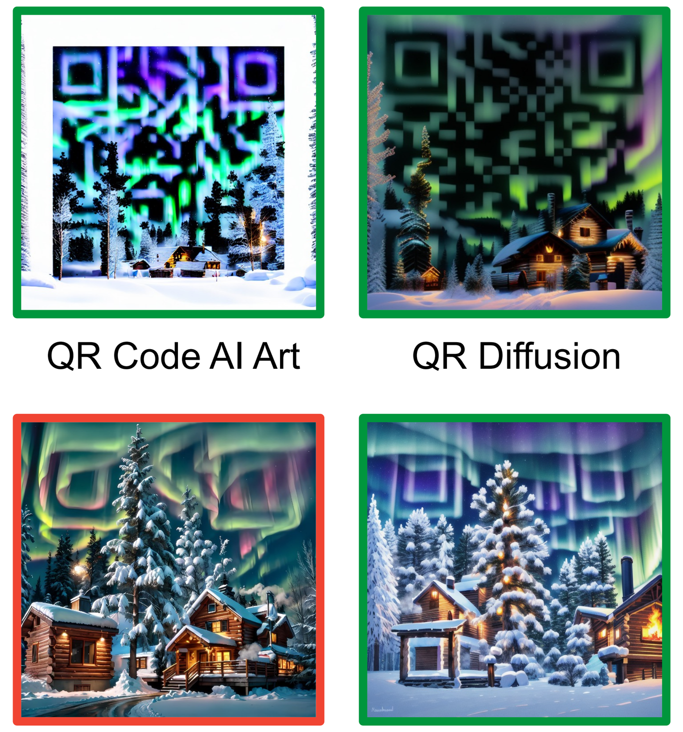
Project Page

Aesthetic QR codes generated from DiffQRCoder

Original QR Code | Winter wonderland, fresh snowfall, evergreen trees, cozy log cabin, smoke rising from chimney, aurora borealis in night sky. | Cherry blossom festival, pink petals floating in the air, traditional lanterns, peaceful river, people in kimonos, sunny day. | Majestic waterfall, lush rainforest, rainbow in the mist, exotic birds, vibrant flowers, serene pool below.

Motivation

Most Diffusion-based aesthetic QR code generation struggle to balance scannability and aesthetics. Although QRBTF generate visually appealing QR codes, they lack scanning robustness. Conversely, QR Code AI Art and QR Diffusion produce better scanning robust QR codes but are visually less appealing. Our approach can generate both attractive and scannable QR codes.



Green: scannable, **Red:** unscannable

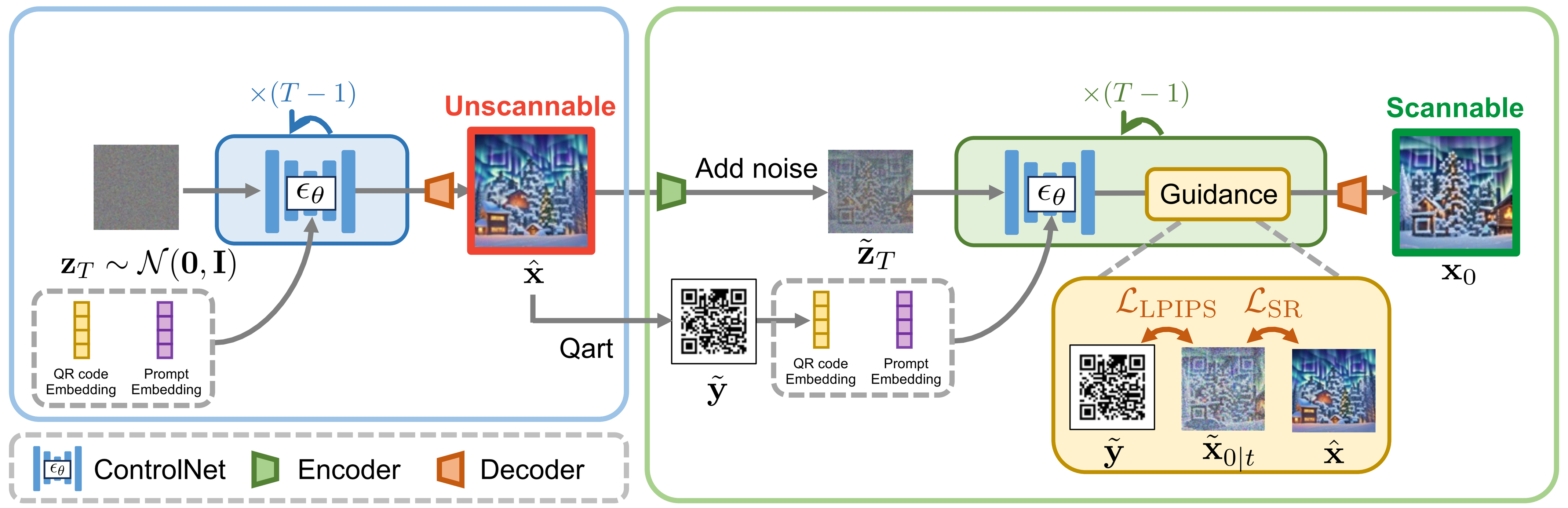
Contribution

- We propose a two-stage iterative refinement framework with **Scanning Robust Perceptual Guidance (SRPG)** to create scanning-robust, visually appealing QR codes without training.
- We develop **Scanning Robust Manifold Projected Gradient Descent (SR-MPGD)**, enhancing the Scanning Success Rate through latent space optimization.
- Our pipeline improves SSR from 60% to nearly 100% compared to ControlNet-only methods, maintaining aesthetics as validated by user evaluations.

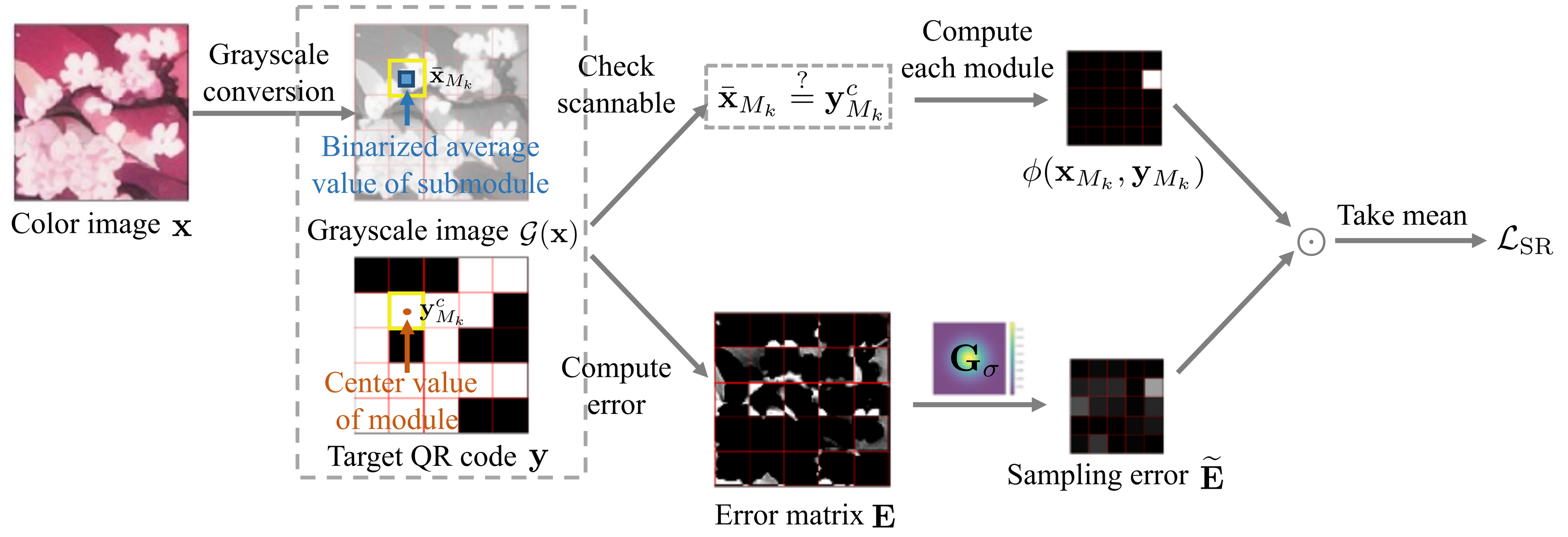
Methodology

Two-stage Iterative Refinement Pipeline

- **Stage-1:** Utilize the pre-trained ControlNet to generate an attractive yet unscannable QR code $\hat{\mathbf{x}}$.
- **Stage-2:** Convert the QR code from Stage-1 into a latent $\tilde{\mathbf{z}}_T$ by adding Gaussian noise and transforming the target QR code $\mathbf{y}$ to $\tilde{\mathbf{y}}$, which has a more similar pattern as $\hat{\mathbf{x}}$, using Qart. Finally, feed the latent and the transformed code into ControlNet, guided by SRPG, to create an aesthetic QR code with scannability.



SRPG

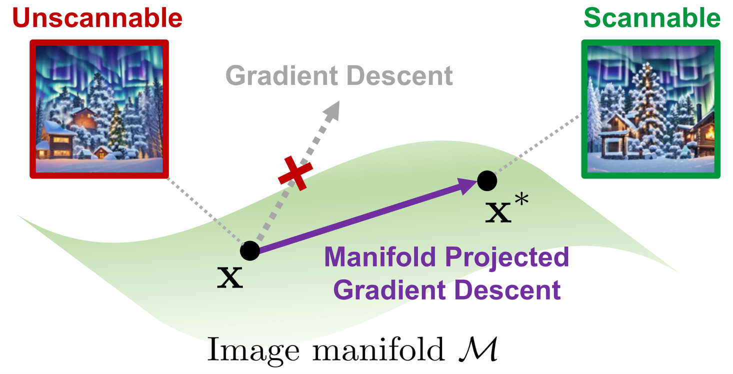


$$\hat{\epsilon}_t = \epsilon_\theta(\tilde{\mathbf{z}}_t, t, \mathbf{e}_p, \mathbf{e}_{\text{code}}) + \sqrt{1-\bar{\alpha}_t}\nabla_{\tilde{\mathbf{z}}_t}[\lambda_1 \mathcal{L}_{\text{SR}}(\tilde{\mathbf{x}}_{0|t}, \tilde{\mathbf{y}}) + \lambda_2 \mathcal{L}_{\text{LPIPS}}(\tilde{\mathbf{x}}_{0|t}, \hat{\mathbf{x}})]$$

SR-MPGD Post-Processing

Optimization Formulation: $\min_{\mathbf{x} \in \mathcal{M}} \mathcal{L}_{\text{SR}}(\mathbf{x}, \mathbf{y})$

$$\mathbf{z} \leftarrow \mathbf{z} - \gamma \nabla_{\mathbf{z}}[\mathcal{L}_{\text{SR}}(\mathcal{D}(\mathbf{z}), \mathbf{y}) + \lambda \mathcal{L}_{\text{LPIPS}}(\mathcal{D}(\mathbf{z}), \mathbf{x}_0)]$$



Experiments

(a) Encoded message: I think, therefore I am!

(b) Encoded message: https://www.google.com.tw/

Quantitative Result

Method	SSR ↑	CLIP-aes. ↑	CLIP-score ↑	Avg-rank ↓
QR Code AI Art [13]	90%	5.7003	0.2341	2.71
QR Diffusion [15]	96%	5.5150	0.2780	3.18
QRBTF [18]	56%	**7.0156**	**0.3033**	**1.86**
DiffQRCoder (Ours)	**99%**	6.8233	0.2992	2.25

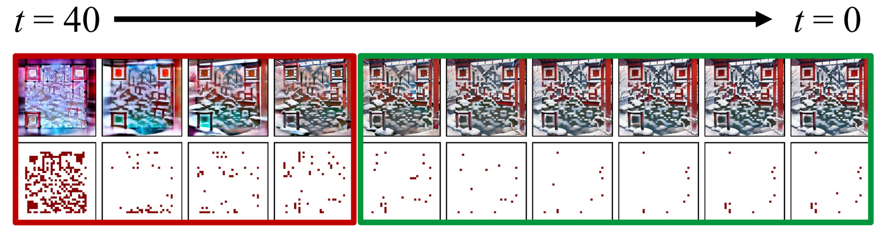
Table 1: Quantitative comparison with other methods.

- **SSR:** Utilize qr-verify to assess the scanning success rate
- **CLIP-aes:** Utilize CLIP aesthetic predictor to quantify the aesthetic
- **CLIP-score:** Utilize CLIP to quantify the text-image alignment
- **Avg-rank:** Perform user subjective aesthetic preference study

Ablation Study

Stage	λ_1	λ_2	SR-MPGD	CLIP-aes. ↑	SSR ↑
Stage-1-only	-	-		7.0661	60%
Two-stage	400	0		6.7860	86%
Two-stage	500	0		6.7259	88%
Two-stage	600	0		6.7183	94%
Two-stage	1000	0		6.5667	93%
Two-stage	400	0	✓	6.7567	98%
Two-stage	500	0	✓	6.7097	100%
Two-stage	600	0	✓	6.7002	99%
Two-stage	1000	0	✓	6.5629	99%
Two-stage	500	2		6.8600	90%
Two-stage	500	3		6.8744	89%
Two-stage	500	5		6.8357	89%
Two-stage	500	10		6.8409	88%
Two-stage	500	2	✓	6.8204	98%
Two-stage	500	3	✓	6.8233	99%
Two-stage	500	5	✓	6.7779	100%
Two-stage	500	10	✓	6.8040	97%

Table 2: Ablations for our proposed pipeline.



Timestep 40 ~ 25 (Unscannable) | Timestep 25 ~ 1 (Scannable)

Degree	0°	15°	30°	45°
SSR ↑	100%	100%	100%	97%

Table 3: Scannability of different rotated angles.

Level	L (7%)	M (15%)	Q (25%)	H (30%)
SSR ↑	96%	100%	100%	100%

Table 4: Scannability of different QR code error correction levels.

Acknowledgement: This research is supported by National Science and Technology Council, Taiwan (R.O.C), under the grant number of NSTC-112-2634-F-002-006, NSTC-112-2222-E-001-001-MY2, NSTC-113-2634-F-001-002-MBK, NSTC-113-2221-E-002-201 and Academia Sinica under the grant number of AS-CDA-110-M09.