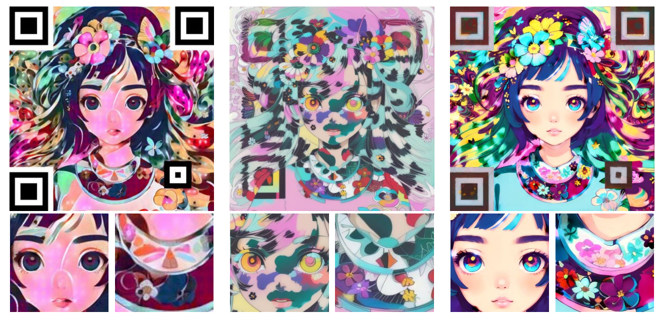
GladCoder: Stylized QR Code Generation with Grayscale-Aware Denoising Process

Yuqiu Xie[1], **Bolin Jiang**[1], **Jiawei Li**[2], **Naiqi Li**[*1], **Bin Chen**[*3], **Tao Dai**[4], **Yuang Peng**[1] and **Shu-Tao Xia**[1]
[1]Tsinghua University
[2]Huawei Manufacturing[†]
[3]Harbin Institute of Technology, Shenzhen
[4]Shenzhen University
{jieyq22, jbl22}@mails.tsinghua.edu.cn, li-jw15@tsinghua.org.cn, linaiqi@sz.tsinghua.edu.cn, chenbin2021@hit.edu.cn, daitao@szu.edu.cn, yuangpeng.cn@gmail.com, xiast@sz.tsinghua.edu.cn

Abstract

Traditional QR codes consist of a grid of black-and-white square modules, which lack aesthetic appeal and meaning for human perception. This has motivated recent research to beautify the visual appearance of QR codes. However, there exists a trade-off between the visual quality and scanning-robustness of the image, causing outputs of previous works are simple and of low quality to ensure scanning-robustness. In this paper, we introduce a novel approach GladCoder to generate stylized QR codes that are personalized, natural, and text-driven. Its pipeline includes a Depth-guided Aesthetic QR code Generator (DAG) to improve quality of image foreground, and a GrayscaLe-Aware Denoising (GLAD) process to enhance scanning-robustness. The overall pipeline is based on diffusion models, which allow users to create stylized QR images from a textual prompt to describe the image and a textual input to be encoded. Experiments demonstrate that our method can generate stylized QR code with appealing perception details, while maintaining robust scanning reliability under real world applications.



(a) ArtCoder (b) QRBTF (c) Ours

Figure 1: Examples of ArtCoder [Su *et al.*, 2021], QRBTF [Ni *et al.*, 2023], and ours outputs.

1 Introduction

Nowadays, the Quick Response (QR) code [ISO, Geneva Switzerland ISOIEC 18004 2000] is applied in a broad range of applications such as business cards, mobile payments, and advertising. Traditional QR codes are of matrix form comprising black-and-white square modules, which are aesthetically displeasing and meaningless for human perception.

Stylized QR codes can better capture people's attention to increase link visits, and many works have investigated various ways to improve visual quality of QR codes [Chu *et al.*, 2013; Aliva *et al.*, 2018; Su *et al.*, 2021; Cox., 2012; Xu *et al.*, 2021; Lin *et al.*, 2015].

Among these works, most of them are based on predefined generation rules and styles, which lack flexibility and personalization. State-of-the-art stylized QR code generation method, ArtCoder [Su *et al.*, 2021], uses a Neural Style Transfer network to generate outputs. Therefore, it is hard for ArtCoder to generate natural and high quality images. As shown in Fig. 1, ArtCoder outputs have obvious black/white pixels that negatively influence the visual perception.

Recent advancements of image generation models (e.g., Stable Diffusion [Rombach *et al.*, 2022], DALLE [Ramesh *et al.*, 2021]), show outstanding ability to generate flexible and realistic images, paving the way for high-quality stylized QR code generation. [Ni *et al.*, 2023] proposed QRBTF to embed QR code images into Diffusion model generated images, using ControlNet [Zhang and Agrawala, 2023]. As shown in Fig. 1, in the generated image produced by QRBTF, there are noticeable black and white blocks that adversely affect the image quality. This issue stems from two main reasons. Firstly, the random distribution of black and white patterns in the QR code leads to irregular areas of brightness and darkness in the generated image, which becomes particularly noticeable when the image content features close-ups of characters or objects. Secondly, the utilization of Brightness ControlNet in the QR code image generation process prioritizes consistency in grayscale between the image and the QR code,

[*]Corresponding author.
[†]The work was done during internship at Huawei Manufacturing.

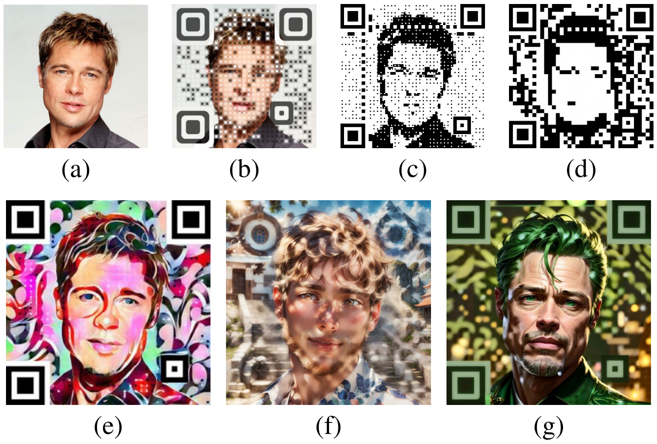


Figure 2: (a) Blended image. (b) Visualead [Aliva *et al.*, 2018]. (c) Halftone QR code [Chu *et al.*, 2013]. (d) Qart code [Cox., 2012]. (e) ArtCoder [Su *et al.*, 2021]. (f) QRBTF [Ni *et al.*, 2023]. (g) Ours.

neglecting considerations for visual quality.

In this paper, we present a novel method GladCoder for creating stylized QR codes, which ensures the scanning-robustness of aesthetic QR codes while generating realistic, high-quality image content. The pipeline takes **a textual prompt** and **a text content** to be coded into QR as input. The generated images not only conform to the provided textual prompt, but also can be recognized by QR code scanners to extract the encoded text content.

We first use a diffusion model to generate an initial image from the textual prompt as a reference image. To mitigate the impact of random QR code patterns on the visual quality of the image, we propose a depth-guided aesthetic QR code generation. The foreground region of the image is extracted by the depth estimation and assigned with low QR priority. Then, the QR patterns are generated under the guidance of the assigned priority, with a focus on high-priority regions and fewer occurrences in low-priority (image foreground) areas. With the aesthetic QR code in hand, we propose a GrayscaLe-Aware Denoising (GLAD) for QR blending to generate the final QR image. Our GLAD relies on pre-trained diffusion networks to generate artistic QR images. Different from using Brightness ControlNet [Ni *et al.*, 2023], our GLAD involves a grayscale-matching loss to ensure the generated images can be recognized by QR scanners. The grayscale-matching loss allows us to more accurately control specific pixels in the image to improve scanning-robustness without compromising the visual quality of the image.

To summarize, our work has the following key contributions: 1) Aesthetic QR code generation process is improved with depth guidance for solving mismatch between QR codes and natural images; 2) GrayscaLe-Aware Denoising (GLAD) process is proposed for QR code Blending; 3) We introduce GladCoder, a diffusion-model-based method, to create high quality stylized QR code images from text inputs.

2 Related Work

In this section, we examine prior research on blend-type stylized QR codes [Xu *et al.*, 2019], which involve incorporating specific images into QR codes. These approaches typically fall into four categories: module deformation, module reshuffling, Neural Style Transfer (NST) , and Diffusion Model.

Module-Deformation: Methods based on module deformation involve reshaping and reducing the areas occupied by square modules in QR codes. Subsequently, these modified regions are used to incorporate images. Two noteworthy examples of such methods are Visualead [Aliva *et al.*, 2018] and Halftone QR codes [Chu *et al.*, 2013]. Visualead [Aliva *et al.*, 2018] enhances the visual appeal of QR codes by distorting modules while maintaining a clear contrast between the modules and the inserted images Fig.2(b). In contrast, Halftone QR codes [Chu *et al.*, 2013] break down each module into 3×3 sub-modules, preserving the color of the central submodules, and then align the remaining sub-modules with a halftone map corresponding to the integrated image Fig. 2(c).

Module-Reshuffle: Recent approaches that focus on module reshuffling draw inspiration from the pioneering work Qart code [Cox., 2012]. Qart code suggests using the Gauss-Jordan Elimination Procedure to rearrange the positions of modules in order to align with characteristics of blended images Fig.2(d). Subsequently, to enhance the visual quality of QR codes, later research has developed various strategies for module reshuffling based on different image features, such as regions of interest [Xu *et al.*, 2021], central saliency [Lin *et al.*, 2015], and global grayscale values [Xu *et al.*, 2019].

NST-Based Method: Xu et al. [Xu *et al.*, 2019] are the first to introduce the NST technique for generating stylized QR codes. They presented SEE (Stylized aEsthEtic) QR codes, which are both personalized and machine-readable. Their approach aimed to address the challenge of style transfer potentially affecting scanning reliability. Another method, Artcoder [Su *et al.*, 2021], proposed by Su et al., improved the integration of blended images into the QR codes while maintaining scanning robustness Fig.2(e). However, the effectiveness of these methods largely relies on how well the original, style, and code images align. Users may find it challenging to achieve a specific desired style. Furthermore, the generated images may still be easily recognized as QR codes due to the conspicuous presence of the black/white pixels.

Diffusion-Model-Based Method: Our method may falls in a new category that uses diffusion models to generate output images [Zhang *et al.*, 2023; Jiang *et al.*, 2024]. As various controllable image generation methods for diffusion model [Dong *et al.*, 2022; Shi *et al.*, 2023; Xie *et al.*, 2023; Kim *et al.*, 2023; Zhang and Agrawala, 2023; Ye *et al.*, 2023] are proposed, blending QR codes into diffusion-generated images becomes feasible. Ni et al. [Ni *et al.*, 2023] first proposed a novel way to train a ControlNet model [Zhang and Agrawala, 2023] that can change grayscale pictures to color pictures, and the model can also be used to embed QR codes into generated images. However, images generated with only ControlNet generally face a trade-off between image quality and sacannability, leading to results with many rectangular shapes and elements arranged in a disorderly manner.

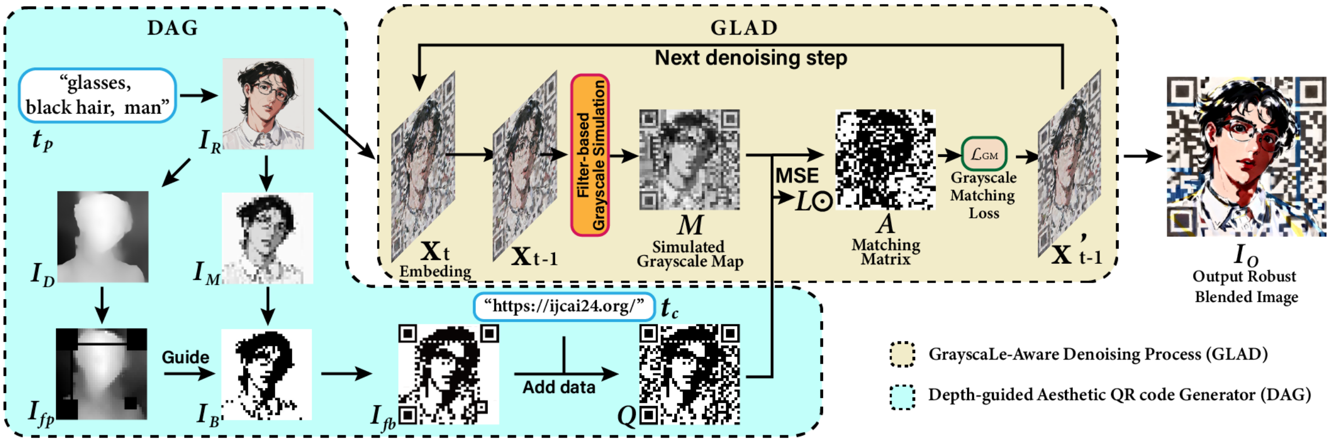


Figure 3: Framework of GladCoder. The pipeline first utilizes depth information as guidance for aesthetic QR code construction, then applies GLAD process to the reference image and the constructed QR code, resulting in a robust blended image.

3 Method

3.1 Framework

The workflow of our method is outlined in Fig. 3. It takes in two text string inputs: t_p is for guidance of image content, and t_c is scanning result. The final output is a scanning-robust image that can be interpreted as t_c by scanners, while under human vision, it is still a harmonious image related to t_p.

The principle goal of our method is to solve the mismatch between the QR code and the reference image, whether it is generated or given. On the one hand, GladCoder utilizes an depth-guided aesthetic QR code generation process, constructing an aesthetic QR code image that is more compatible with reference image than standard QR codes. On the other hand, GladCoder enhances the scanning robustness based on the aesthetic QR code, while keeping the image roughly the same, through a grayscale-aware denoising process.

3.2 Depth-Guided Aesthetic QR Code Generator

With Depth-guided Aesthetic QR code Generator (DAG), GladCoder takes in a reference image I_R and code content t_c, aiming to get an aesthetic QR code Q that is encoded by t_c, while looks similar to I_B, the binary and modulized version of I_R.

The idea is that, for a specific text content, there can actually be many choices of QR codes to embed, and some QR code modules are not evenly arranged in spatial regions. Leveraging this property, we place the smoother QR modules in the foreground of the image, thereby achieving better visual quality.

Following the method proposed in Qart [Cox., 2012], once the version and content of QR code is set, a fixed number of modules of the final output Q can be arbitrarily set to black or white. An algorithm is needed to determine which part of the modules is aligned to I_B. Our method propose a new way to select these flexible modules.

GladCoder aims to solve the mismatch between randomly distributed QR codes and certain types of natural images. These images normally contains one or multiple characters or objects, consisting the foreground and occupying the main part of image. When prioritizing the place they occupies, it would be easier to reproduce these subjects in latter generation. Furthermore, for those small elements placed in background, even with randomly distributed codes, we can still rely on Diffusion Models to express them correctly. Therefore, our method chooses to use depth information as the indicator for the fill preference I_{fp}, putting the most emphasis on foreground elements.

For a certain QR code version v, the number of modules in one row or column of the QR code is $M_{num} = 17 + 4 \times v$. Given a reference image I_R, the pipeline first apply a series of image processes, getting a $M_{num} \times M_{num}$ matrix I_M, which is resized from the grayscale version of I_R. Then I_M is turned into its binary version I_B with:

$$I_B(i,j) = \begin{cases} 0, & I_M(i,j) \leq \tau \\ 1, & I_M(i,j) > \tau \end{cases} \tag{1}$$

Where τ is the mean value of I_M.

On the other hand, I_R also goes through a depth estimator, resized to $M_{num} \times M_{num}$, getting the corresponding depth image I_D. After setting all the fixed modules, e.g. modules of finder/timing/alignment patterns and version/format information, into zero, GladCoder get the fill preference I_{fp} for guiding latter code generation.

During code generation, GladCoder will first set modules have higher value in I_{fp} to bits the same as that in I_B. After running out of the flexible modules, the remaining modules are calculated and set based on the code content t_c, according to the method raised in [Cox., 2012].

3.3 Grayscale-Aware Denoising Process

After constructing the aesthetic QR code, a GrayscaLe-Aware Denoising(GLAD) process is applied to reference image I_R, achieving a scanning-robust image I_O. The idea of this process is that, when generating final outputs, there is no need to control all the pixels. Since code readers only uses central pixels of a module to define its value, the rest pixels should be generated freely, instead of still being affected by white/black constraints. The GLAD process is a modified

(a) Early steps (b) Latter steps

Figure 4: Examples of simulated grayscale map M and matching matrix A.

version of image-to-image diffusion process. After each step of denoising to latent X_t, a grayscale-matching loss L_{RM} is calculated and backpropagated to X_{t-1}. The process gives I_R a minimal revision, making the output I_O function the same as the QR code Q.

To be specific about the GLAD process, after getting the latent X_{t-1}, it is decoded to an image O, which GladCoder regards as an image consists of $M_{num} \times M_{num}$ modules of $a \times a$ pixels. Then, a convolution layer to extract image grayscale map, is applied to the image O.

The convolution Layer, which is a filter-based grayscale simulation, has kernel size $a \times a$ and stride a. Its kernel weight K is defined as:

$$K_{(i,j)} = \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}} \tag{2}$$

Therefore, the simulated grayscale map M of the decoded image is :

$$M_{M_k} = \sum_{(i,j) \in M_k} K_{(i,j)} \cdot O_{M_k}(i,j) \tag{3}$$

Where M_k is the kth module, $O_{M_k}(i,j)$ is the grayscale value of the (i,j) pixel in M_k. The reason why kernel weight K is set by a Gaussian distribution is that QR code scanners detect value of modules based only on central pixels. Thus, feature map calculation should also put greater emphasis on those central pixels.

Next, GladCoder performs a module-by-module comparison between the simulated grayscale map M and the code target Q, getting an matching matrix A:

$$A_{(i,j)} = \begin{cases} 0, & |M_{(i,j)} - Q_{(i,j)}| \leq 128 \\ & \& \, |M_{(i,j)} - 128| > \delta \\ 1, & otherwise \end{cases} \tag{4}$$

Where δ is a predefined threshold. After that, a grayscale-matching loss L_{GM} can be calculated through:

$$L_{GM} = MSE(M \odot A, Q \odot A) \tag{5}$$

Finally, a backpropagation can help our pipeline find X'_{t-1}, a more scanning robust version of latent X_{t-1}. The output of our pipeline $I_O = Decode(X'_0)$.

As shown in Fig. 4, when the latent of input image X undergoes GLAD process, its simulated grayscale map M increasingly resembles the aesthetic QR code Q, indicating successful blending of the code into the image.

4 Experiments

In the following sections, experiments are performed to explore three aspects of our method, image generation quality, scanning robustness, and ablation study.

4.1 Implementation

Experimental Setting: All experiments are conducted on a computer with a NVIDIA RTX 3090Ti GPU. When doing scanning robustness related experiments, all output images are displayed on a 23.8-inch, 144Hz, 1920×1080 IPS-panel monitor screen. The base models we used for generating images are RevAnimated, Aniflatmix, and MajicMIX realistic. All three models are fine-tuned version of "stable-diffusion-v1.5", having expertise in different image categories.

Parameter Setting: In scanning robustness enhancement process, by default, $\sigma = 3$ in equation 2, $\delta = 50$ in equation 4, and learning rate is 0.005. All images are generated in size 592×592 (37×37 modules, each module of size 16×16), for fair comparison with ArtCoder.

4.2 Generation Quality

Comparison With Other Methods

Fig. 5 shows the comparison between other state-of-the-art stylized QR code generation methods (ArtCoder [Su *et al.*, 2021] and QRBTF [Ni *et al.*, 2023]) and GladCoder. For fair comparison with the two methods, reference images and aesthetic QR codes are used as ArtCoder inputs (reference images are used as both content and style images), and text promts, reference images, as well as aesthetic QR codes are used as QRBTF inputs. Overall, our method gives outputs with the best visual quality, because it resembles the reference images better, with finer details and more natural appearances.

ArtCoder results can barely retain the appearance of reference images, with little detail and tedious colors. As for QRBTF, as long as the mismatch between codes and contents is solved by our aesthetic QR codes, it can produce visually pleasing and scannable images with ControlNet models. Ours Grayscale-Aware Denosing process improves the visual quality one step further. By only restricting the central pixels, GladCoder outputs have better image quality, with not only less rectangular shapes, but also more vivid colors. Therefore, GladCoder outperforms ArtCoder and QRBTF in image quality, with much more natural appearance.

Generation Variety

Fig. 6 shows different GladCoder outputs from one text prompt. Since Diffusion models have variety when generating images, our pipeline can also generate various output images based on the same prompt, offering users more choices.

4.3 Scanning Robustness

In this section, a series of experiments are conducted to evaluate the scanning robustness of our stylized QR codes.

Analysis of Preserving Scanning Robustness

Fig. 7 displays why GladCoder outputs can achieve high image quality while maintaining scanning robustness.

First, in space dimension, there is no need to fill the whole module with ideal code module color, and code readers also only checks center pixels. GladCoder makes sure that pixels near the center of each modules are matched with corresponding code modules, which gives space for diffusion models to create. Then, in color dimension, there is also no need to fill

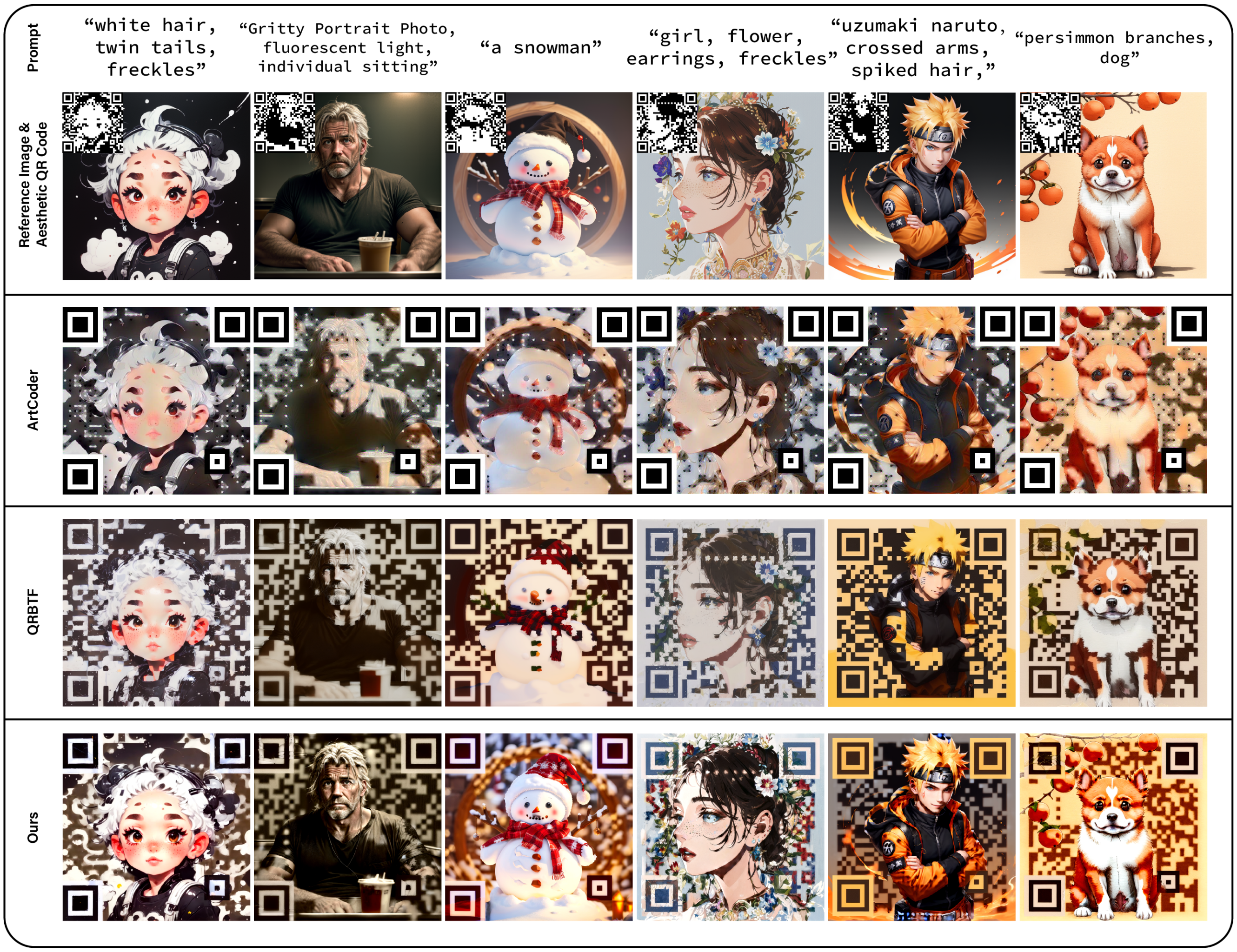


Figure 5: Comparison with other stylized QR code generation methods: ArtCoder [Su *et al.*, 2021] and QRBTF [Ni *et al.*, 2023].

in pure black or white into central pixels. Any color that its grayscale value is away from the average value of the whole image, for a certain distance, can be recognized correctly by code readers. Therefore, it gives a wider range of colors for GladCoder to fill in pixels. With this two dimensions of freedom, GladCoder is able to achieve better image quality with consistent scanning robustness.

Influences of Parameter δ

Parameter δ in equation 4 controls the grayscale threshold that GladCoder decides whether a module is valid or not. As shown in Fig. 8, as δ is set higher, each module must be blacker/whiter to be classified as a valid module. Otherwise, its corresponding value in activate matrix will be set to one and updated by the backpropagation.

The quality-robustness trade-off still exists in our method. As the module colors become more extreme, the outputs become more visually unpleasant. In our experiments, outputs with $\delta = 40$ is enough for direct image scan, and those with $\delta = 50$ is robust for daily app scanning, which is revealed in the next section.

Influences of Code Reader and Scanning Distance

The scanning robustness under different mobile phone apps and different distances is evaluated as follows. First, 50 output images of each method (ArtCoder, QRBTF, GladCoder) are randomly selected. All the images are generated in resolution 592×592. Then, each image is displayed on screen in three frequently used size: $4cm^2$, $7cm^2$, and $10cm^2$. We scan each image 5 times, at distance of 30cm, with each app and each size, recording the average times of successful scanning and decoding time (a successful scanning meaning the image is decoded in 5 seconds).

As the experimental results shown in Table 1, ArtCoder performs better with larger codes while QRBTF performs better with smaller codes. This is because ArtCoder only focus on central pixels of modules, which would be less obvious when code is small; Although QRBTF roughly keeps the colors and rectangular shapes, the creativity of diffusion model and ControlNet may still change central pixels accidentally. GladCoder, taking the advantages of the two methods, make sure central pixels are valid with grayscale-matching loss, and spread the colors with generation ability of diffusion models,

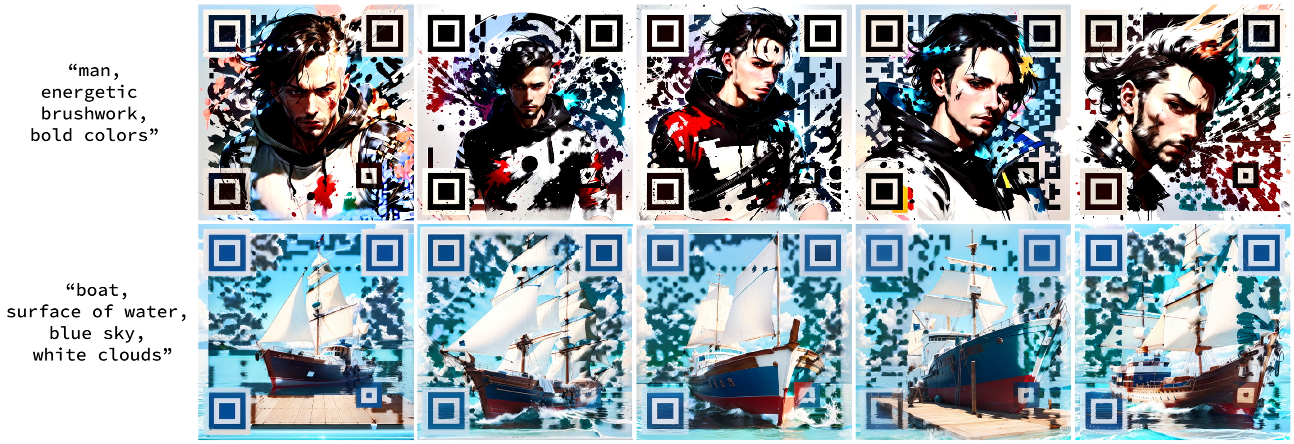


Figure 6: Different outputs from same prompt, exhibiting variety and consistency of our method.

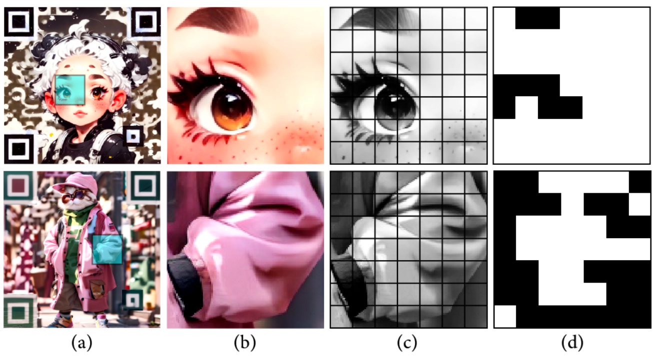


Figure 7: Analysis of preserving scanning robustness. (a) Ours outputs. (b) Enlarged views. (c) Grayscale results. (d) Ideal codes.

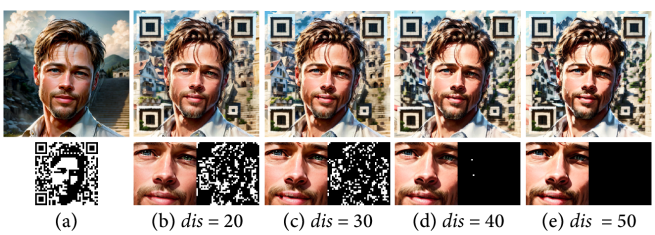


Figure 8: Influence of parameter δ. (a) Reference image and QR code. (b-e) Outputs above, below left is enlarged view, below right is error matrix with threshold 40.

end up performing well in both small and large sizes.

Among all four apps, Wechat and Tiktok have stronger scanning ability that all three methods perform well. QRBTF results can hardly be scanned by Alipay in large size, and the reason might be Alipay would auto zoom-in when a code can't be scanned, which makes it even harder to scan a QRBTF code. It seems like Facebook QR code scanner has a higher standard for code, which makes it unable to scan ArtCoder and QRBTF results in most situations, proving that GladCoder results has better scanning robustness.

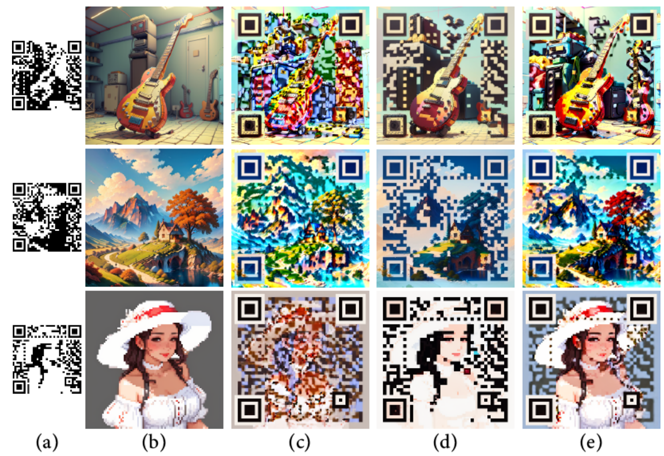


Figure 9: Ablation experiment for our framework. (a) Aesthetic QR codes. (b) Reference images. (c) w.o. Aesthetic QR codes (using standard QR codes). (d) w.o. GLAD process (using QRBTF for robustness enhancement). (e) Ours.

4.4 Ablation Study

Ablation for Our Framework

A ablation study about the two parts of our framework is conducted in Fig. 9. As shown in Fig. 9 (c), without aesthetic QR codes, outputs will face the same problem that occurs when QRBTF is enforced character poses (Fig. 11 (c)). Therefore, matching the QR code and image content is the essential part to improve image quality.

GLAD process aligns central pixels of each module to its corresponding code module color, so scanning robustness enhancement is essential to ensure the functionality of QR code. Even with other method to enhance scanning robustness, as shown in Fig. 5 (d), Brightness and Light Composition ControlNet models from QRBTF [Ni *et al.*, 2023] is applied to reference images and aesthetic QR codes. The models will treat all pixels in one module equally, and push them to extreme colors. With enhancement of QRBTF, large proportion of modules still retains rectangular shape, which makes the results look worse than GladCoder outputs.

Figure 10: Stylized results of ArtCoder [Su *et al.*, 2021] and ours. (a) Reference image. (b) Style images. (c) ArtCoder outputs. (d) Text prompts. (e) Our outputs.

Method	App	$4cm^2$	$7cm^2$	$10cm^2$
Artcoder	Wechat	92%/1.93s	96%/1.57s	98%/1.29s
	Alipay	86%/2.33s	98%/1.62s	**98%/1.67s**
	Facebook	0%/-	5%/1.19s	86%/1.15s
	Tiktok	98%/0.78s	**100%/0.75s**	**100%/0.74s**
QRBTF	Wechat	**100%/0.95s**	**100%/0.96s**	**99%/0.95s**
	Alipay	100%/0.97s	61%/1.14s	20%/1.37s
	Facebook	0%/-	0%/-	0%/-
	Tiktok	96%/0.77s	83%/0.82s	72%/0.84s
Ours	Wechat	100%/0.96s	100%/1.02s	98%/1.01s
	Alipay	**100%/0.98s**	**98%/1.35s**	94%/1.66s
	Facebook	**98%/0.96s**	**98%/0.93s**	**98%/0.89s**
	Tiktok	**100%/0.75s**	**100%/0.75s**	99%/0.78s

Table 1: Average scanning success rate and decoding time, under different methods and apps. ArtCoder performs better with larger codes while QRBTF performs better with smaller codes. Ours performs well in both small and large sizes.

Therefore, each of our improvement is critical to achieve a visually pleasing and scanning robust result.

4.5 More Discussion and Application

Stylization

Since ArtCoder is based on Neural Style Transfer networks, we conduct an experiment to compare the stylization of ArtCoder and our method. As shown in Fig. 10, ArtCoder results are generated with (a) as content image and (b) as style images, while our results are generated with only (d) as input. Our method offers a easier control over outputs. Users only need to give out text prompts about the contents and styles they want about the output images, rather than finding references by themselves.

Plugin Test of Our Aesthetic QR Codes

In this section, we test the effectiveness of our aesthetic QR code generation process, over QRBTF. The pipeline utilizes a text-to-image stable diffusion framework with ControlNet models. Reference images are used as depth ControlNet inputs and QR codes are used as Brightness and Light Composition ControlNet model inputs. As shown in Fig. 11, (c) shows QRBTF results with standard QR codes, (d) shows QRBTF results with our aesthetic QR codes. When using standard QR code, outputs retain the outline of reference images, but they are still visually unpleasant because of the mismatch between QR codes and reference images. When using aesthetic QR codes, outputs can better imitate the reference images, proving that our aesthetic QR code generation process is also effective to QRBTF.

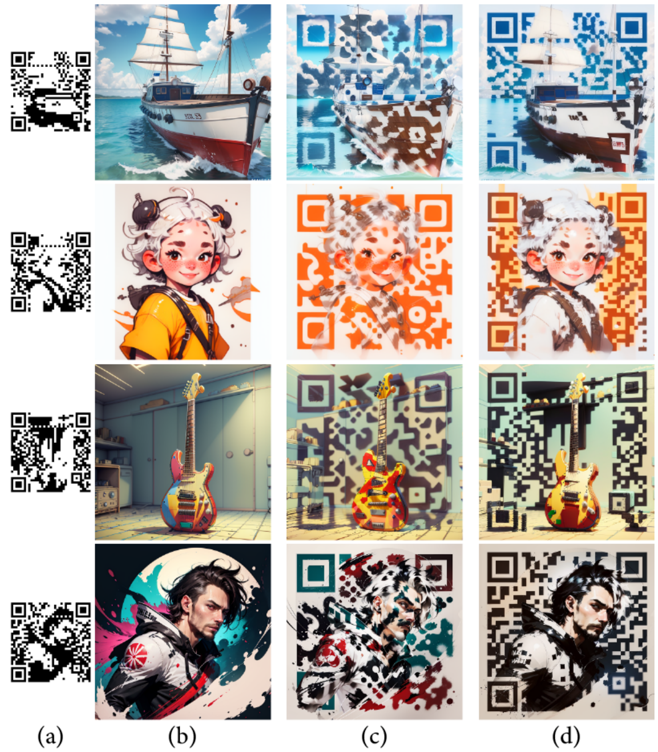


Figure 11: Ours depth-guided aesthetic QR codes can be seamlessly integrated to improve other method's performance. (a) Aesthetic QR codes. (b) Reference images. (c) QRBTF with standard QR codes. (d) QRBTF with our aesthetic QR codes.

5 Conclusion

In this paper, we introduce GladCoder, a method for creating high-quality images that seamlessly combine aesthetic QR codes and reference images generated by text prompts in a harmonious manner. To tackle the challenge of maintaining scanning reliability while preserving high-quality content, our method utilizes two parts of process: aesthetic QR code generation prioritized by depth, and further robust enhancement with grayscale-aware denoising process. These stages help strike a balance in quality-robustness trade-off. Extensive experiments confirm that our stylized QR code images excel in blending content and code, resulting in output images with finer details.

References

[Aliva *et al.*, 2018] N. Aliva, U. Peled, and F. Itamar. Visualead. http://www.visualead.com, 2018.

[Chu *et al.*, 2013] Hung-Kuo Chu, Chia-Sheng Chang, Ruen-Rone Lee, and Niloy J Mitra. Halftone QR codes. *ACM transactions on graphics (TOG)*, 32(6):1–8, 2013.

[Cox., 2012] R. Cox. Qartcodes. http://research.swtch.com/qart, 2012.

[Dong *et al.*, 2022] Ziyi Dong, Pengxu Wei, and Liang Lin. Dreamartist: Towards controllable one-shot text-to-image generation via contrastive prompt-tuning. *arXiv preprint arXiv:2211.11337*, 2022.

[ISO, Geneva Switzerland ISOIEC 18004 2000] ISO. Information technology automatic identification and data capture techniques code symbology QR code. *Int. Org. Standard.*, Geneva, Switzerland, ISO/IEC 18004: 2000.

[Jiang *et al.*, 2024] Bolin Jiang, Yuqiu Xie, Jiawei Li, Naiqi Li, Yong Jiang, and Shu-Tao Xia. Cagen: Controllable anomaly generator using diffusion model. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3110–3114, 2024.

[Kim *et al.*, 2023] Yunji Kim, Jiyoung Lee, Jin-Hwa Kim, Jung-Woo Ha, and Jun-Yan Zhu. Dense text-to-image generation with attention modulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7701–7711, 2023.

[Lin *et al.*, 2015] Shih-Syun Lin, Min-Chun Hu, Chien-Han Lee, and Tong-Yee Lee. Efficient QR code beautification with high quality visual content. *IEEE transactions on multimedia*, 17(9):1515–1524, 2015.

[Ni *et al.*, 2023] Hao Ni, Boyu Chen, Zhaohan Wang, and Zhiyong Chen. QRBTF. http://qrbtf.com, 2023.

[Ramesh *et al.*, 2021] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pages 8821–8831. PMLR, 2021.

[Rombach *et al.*, 2022] 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.

[Shi *et al.*, 2023] Yujun Shi, Chuhui Xue, Jiachun Pan, Wenqing Zhang, Vincent YF Tan, and Song Bai. Dragdiffusion: Harnessing diffusion models for interactive point-based image editing. *arXiv preprint arXiv:2306.14435*, 2023.

[Su *et al.*, 2021] Hao Su, Jianwei Niu, Xuefeng Liu, Qingfeng Li, Ji Wan, Mingliang Xu, and Tao Ren. Artcoder: an end-to-end method for generating scanning-robust stylized QR codes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2277–2286, 2021.

[Xie *et al.*, 2023] Jinheng Xie, Yuexiang Li, Yawen Huang, Haozhe Liu, Wentian Zhang, Yefeng Zheng, and Mike Zheng Shou. Boxdiff: Text-to-image synthesis with training-free box-constrained diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7452–7461, 2023.

[Xu *et al.*, 2019] Mingliang Xu, Hao Su, Yafei Li, Xi Li, Jing Liao, Jianwei Niu, Pei Lv, and Bing Zhou. Stylized aesthetic QR code. *IEEE transactions on multimedia*, 21(8):1960–1970, 2019.

[Xu *et al.*, 2021] Mingliang Xu, Qingfeng Li, Jianwei Niu, Hao Su, Xiting Liu, Weiwei Xu, Pei Lv, Bing Zhou, and Yi Yang. ART-UP: A novel method for generating scanning-robust aesthetic QR codes. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, 17(1):1–23, 2021.

[Ye *et al.*, 2023] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023.

[Zhang and Agrawala, 2023] Lvmin Zhang and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. *arXiv:2302.05543*, 2023.

[Zhang *et al.*, 2023] Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, and Shu-Tao Xia. Unsupervised surface anomaly detection with diffusion probabilistic model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6782–6791, 2023.