Face2QR: A Unified Framework for Aesthetic, Face-Preserving, and Scannable QR Code Generation

Xuehao Cui*, **Guangyang Wu***, **Zhenghao Gan**, **Guangtao Zhai**, **Xiaohong Liu**†
Shanghai Jiao Tong University
{cavosamir, wu.guang.young, ganzhenghao,
zhaiguangtao, xiaohongliu}@sjtu.edu.cn

Abstract

Existing methods to generate aesthetic QR codes, such as image and style transfer techniques, tend to compromise either the visual appeal or the scannability of QR codes when they incorporate human face identity. Addressing these imperfections, we present Face2QR—a novel pipeline specifically designed for generating personalized QR codes that harmoniously blend aesthetics, face identity, and scannability. Our pipeline introduces three innovative components. First, the ID-refined QR integration (IDQR) seamlessly intertwines the background styling with face ID, utilizing a unified Stable Diffusion (SD)-based framework with control networks. Second, the ID-aware QR ReShuffle (IDRS) effectively rectifies the conflicts between face IDs and QR patterns, rearranging QR modules to maintain the integrity of facial features without compromising scannability. Lastly, the ID-preserved Scannability Enhancement (IDSE) markedly boosts scanning robustness through latent code optimization, striking a delicate balance between face ID, aesthetic quality and QR functionality. In comprehensive experiments, Face2QR demonstrates remarkable performance, outperforming existing approaches, particularly in preserving facial recognition features within custom QR code designs. Codes are available at https://github.com/cavosamir/Face2QR.

1 Introduction

Quick Response (QR) codes, due to their capability to store a substantial amount of data and their ease of accessibility through basic camera devices, have become an exceedingly widespread medium for the representation of information in the digital era [13, 47, 3, 4, 36]. With the wide application of QR codes in social context, there has been increasing needs for customizing QR codes to include **personal identity** and **aesthetic allure**. However, such needs cannot be fulfilled by the dull appearance of common QR codes, which contain only black and white modules.

With the widespread application of QR codes across diverse fields, related technologies are also developing at a rapid pace. While techniques employing image transformation [5, 6, 13, 46] and style transferring [36, 47] can partially retain facial features, their perceptual quality and aesthetic adaptability are limited. On the other hand, generative model-based approaches [11, 43] can produce QR code images of superior quality and diversity, yet they pose challenges in controlling the generated content, particularly in preserving human facial characteristics. To address these limitations and ensure faithful preservation of face identity within a customized and scannable QR code image, we introduce a novel pipeline, named Face2QR. This approach achieves a balanced compromise between face ID preservation, aesthetic appeal, and scannability for QR code images.

* Equal contribution.
† Corresponding author.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Figure 1: Face images (first row) and QR code images (second row) generated by Face2QR. Our QR codes not only faithfully maintain face ID, but also showcase remarkable scanning resilience and aesthetic quality.

The primary challenges lie in effectively integrating three key aspects: face ID, aesthetic quality, and scanable QR pattern, which can be summarized as follows: **(1) Combination of face ID and background.** Achieving a harmonious balance between strict facial ID preservation and diverse customized background styles within a unified pipeline presents a notable challenge. Methods reliant on style transfer [36, 47] often yield facial textures that appear unnatural, while those based on image transfer [5, 6, 13, 46] may introduce visible artifacts in the facial region; **(2) Conflict between face ID and QR code pattern.** While prior generative model-based techniques [43] have demonstrated the ability to control the QR code pattern using QR blueprints, they have struggled to exclude these patterns from the facial region, resulting in unnatural shadows and undesirable artifacts. However, directly removing these patterns from the facial region can make the image unscannable. Thus, achieving a balance between maintaining visual quality in the facial region and ensuring the correctness of the QR pattern presents a formidable obstacle; **(3) Balance between aesthetics and scannability.** As revealed in [43], generated images often exhibit a tendency towards being unscannable, necessitating enhancements to their scannability through post-processing. However, globally adjusting brightness can compromise the natural appeal of the facial region. Thus, novel region-based enhancement methods are worth considering to address this challenge.

To address these challenges, the proposed Face2QR pipeline offers a solution for generating personalized QR codes that strike a balance between aesthetics, facial ID preservation, and scannability. We propose ID-refined QR integration (IDQR) to seamlessly combine background and face ID, and ID-aware QR ReShuffle (IDRS) to solve the conflict between face ID and QR code pattern. Specifically, IDQR applies a unified SD-based framework to ensure that the generated images have a uniform style. Stable Diffusion (SD) models are guided by two sets of control networks, corresponding to face refinement and QR pattern respectively, to achieve separate control in face region and background. IDRS utilizes the flexibility of QR code encoding and reshuffles the modules to make the QR pattern compatible with face ID. Finally, we use ID-preserved Scannability Enhancement (IDSE) to enhance scan robustness through latent code optimization, achieving a new trade-off between face ID, aesthetics and scannability. Figure 1 shows the QR images generated by Face2QR. It is worth noting that the generated QR images are not only the reprints of the provided references, but also have improved aesthetics to align with the generated background, guided by text prompts (*e.g.*, the style and color of clothes have been adjusted accordingly).

The contributions of this work can be summarized as:

- We propose a novel pipeline that holistically integrates aesthetic appealing, facial ID, and scannability to deliver a customized personal representation in QR codes.

- We introduce the ID-refined QR integration (IDQR) for seamlessly integrating face ID with background, the ID-aware QR ReShuffle (IDRS) for solving conflicts between face ID and QR pattern, and the ID-preserved Scannability Enhancement (IDSE) for optimizing scan robustness while maintaining face ID and aesthetic quality.

- Our Face2QR achieves the State-Of-The-Art (SOTA) performance in generating the ID-preserved aesthetic QR codes, compared with previous methods.

2 Related Works

Quick Response (QR) Code. As QR codes emerging as a key connector between real and virtual worlds, there is increasing interest in enhancing the visual appeal of normally monochromatic QR codes. Halftone QR codes [5] offers a design where QR code patterns align with a given image in a thematically cohesive manner. QRImage and Artup [13, 46] explore ways to encode colorful imagery within a QR code. Other advances [35, 36] have been made in artistic style transfer to increase aesthetic appearance of QR codes. To further customize QR code and obscure overt QR code markers, Chen et al. [2, 4, 23] crafted encoding schemes that consider human visual perception, thus making these patterns less intrusive. TPVM [12] has gone further to conceal QR codes within video content, exploiting the discrepancies in frame capture rates between human vision and digital screens. Similarly, advancements have sought to keep data imperceptible yet accessible through various stealth mechanisms [10, 9, 37, 16, 42, 17].

Diffusion Based Models. Image manipulation and generation techniques powered by deep learning have made strides in recent years [41, 45, 21, 44, 31, 33, 32], with generative models being at the forefront of this development [51, 24, 28, 26]. Novel diffusion-based models such as GLIDE [26], DALLE-2 [28], and Latent Diffusion models [30] have come into prominence. Notably, the Stable Diffusion model [30] moves the denoising steps into the latent dimension of a variational autoencoder, which significantly optimizes the generation process in terms of data volume and training time. In parallel, new research has introduced various techniques for modulating the diffusion process. Structural condition interventions have been successfully implemented by ControlNet [52] and T2I-Adapter [25]. On a different note, BLIP-Diffusion [20] and SeeCoder [48] have made progress on steering generative outcomes based on stylistic aspects.

Identity Preserved Generative Models. In the field of ID-preserving image generation, research focuses on maintaining semantic face attributes while generating images that have wide real-world applications. Studies have generally split between techniques requiring test-time fine-tuning, such as Low-Rank Adaptation [15], and newer optimization-free methods such as Face0 [38], PhotoMaker [22], and FaceStudio [49], which integrate facial embeddings into the generation process in different ways. While techniques like IP-Adapter [50] strive for identity consistency by using embeddings from recognition models, they face challenges in compatibility with pre-trained models and ensuring facial fidelity. Most recent work like InstantID [40] use a pluggable module that does not demand fine-tuning and can work seamlessly with available pre-trained diffusion models to achieve high-quality face preservation in generated images.

3 Method

The overall structure of Face2QR is shown in Figure 2. The pipeline unfolds through three stages, represented by blue, red and green arrows. Given a user-customized face image f, QR Code m, text prompts c and random noise z_0, the first stage uses the ID-refined QR integration (IDQR) module to generate an initial QR image I^g. The IDQR module includes a pre-trained SDXL model (denoted as $\mathcal{SD}$), an InstantID [40] network (denoted as C_{id}) and a QR Controller [43] (denoted as C_{qr}). Stage 1 can be expressed as:

$$I^g = \mathcal{SD}(c, z_0 | \mathcal{C}_{qr}(m, c, z_0), \mathcal{C}_{id}(f, c, z_0)). \tag{1}$$

The InstantID network preserves the facial identity information in the generated images, while the QR Controller guides the luminance distribution of the images.

However, as shown in Figure 2, the initial output image from the first stage contains a significant error rate (over 43%). This issue arises from the inherit conflict between two control signals: the foreground face information and the background QR patterns, which are incompatible in the center regions. These conflicts lead to unavoidable QR code errors, presenting a core challenge in our pipeline. To address this, we design the ID-aware QR ReShuffle (IDRS) module to harmonize these conflicts and regenerate the image using a fine-grained QR blueprint I_b. As illustrated in Figure 2, this reduces the error rate by more than half. Finally, we use the ID-preserved Scannability Enhancement (IDSE) module to refine the result I^s in latent space, further improving its scanning robustness without compromising the overall visual quality. In the following, we introduce the second and third stages in details.

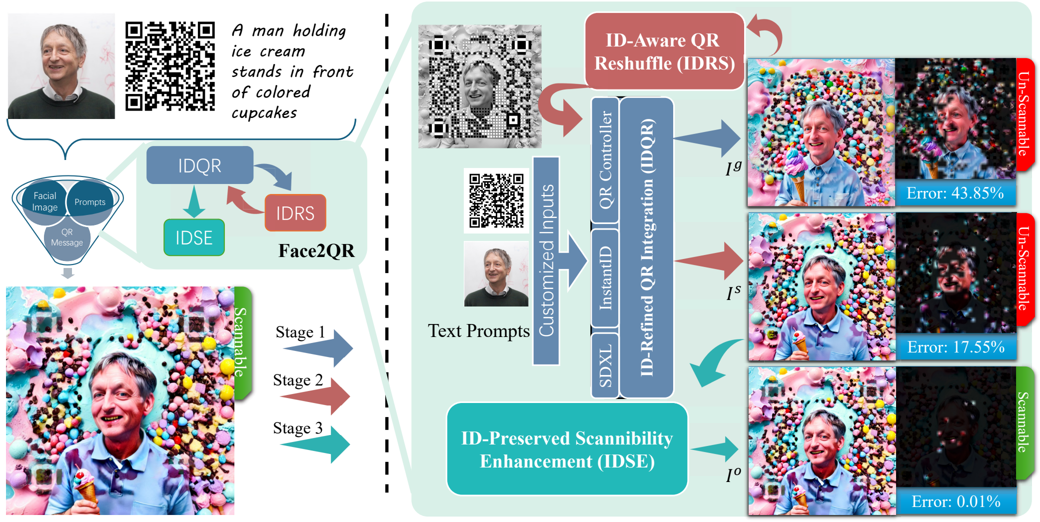


Figure 2: The pipeline of Face2QR is a training-free process for generating ID-consistent and scannable QR code images. Our pipeline has three stages, represented by blue, red, and green arrows. The IDRS module resolves conflicts between human identity and QR patterns during the control process, while the IDSE module reduces coding errors to ensure the output is scannable.

3.1 ID-Aware QR ReShuffle

As revealed in [43], a fine-grained QR blueprint can effectively control the generator. To resolve control conflicts in the facial region, we design a novel blueprint that makes facial information and QR patterns compatible. By leveraging the dynamic characteristics of QR code encoding, we can adaptively rearrange the QR modules. Specifically, we maintain the brightness distribution of the facial region and reshuffle the remaining black and white modules accordingly.

First of all, we binarize $I^g \in \mathbb{R}^{H \times W \times 3}$ into module-wise binary information $\mathbf{E} \in \mathbb{R}^{n^2}$. By dividing I^g into $n \times n$ modules each of size $a \times a$, and let θ_j be the set of pixel coordinates of the j-th module in I^g, the extracted information code $\mathbf{E}$ is given by:

$$\mathbf{E}_j = \begin{cases} 0, & \text{if } \operatorname{avg}(I^g(\theta_j)) < \tau, \\ 1, & \text{if } \operatorname{avg}(I^g(\theta_j)) \geq \tau, \end{cases} \tag{2}$$

where $\operatorname{avg}(\cdot)$ denotes the mean pixel value of the squared patch of size $a \times a$. The binarization uses a threshold τ, typically set to 128 for a total of 256 grayscale levels.

As shown in Figure 3 (left), the binarized QR code is un-scannable due to a significant error rate. To address this, we fix the facial and marker region within $\mathbf{E}$, then rearrange the remaining codes to align with the encoded information. To locate the facial region, we use a pre-trained face recognition model to obtain the binary facial mask $M_f \in \mathbb{R}^{H \times W}$. Let the set $\Delta_f = \{j \mid \operatorname{avg}(M_f(\theta_j)) = 1\}$ represent the indices of information codes in $\mathbf{E}$ that correspond to the facial region, and let the Δ_m represent the indices of marker codes. Our goal is to obtain a new information code $\tilde{\mathbf{E}}$ which is partially modified from $\mathbf{E}$ to make the QR decoder $\mathbb{D}$ extract lossless information:

$$\min |\mathbb{D}(\tilde{\mathbf{E}}) - \mathbb{D}(m)|, \tag{3}$$

$$\text{s.t.} \quad \tilde{\mathbf{E}}_j = \mathbf{E}_j, \text{for } j \in \Delta_f \cup \Delta_m, \tag{4}$$

To ensure the resultant $\tilde{\mathbf{E}}$ can be decoded to the target message, aligning with original QR code m, we re-generate the error correction code [29] in $\tilde{\mathbf{E}}$.

Afterwards, we expand the binary information of $\tilde{\mathbf{E}}$ to image space. We use adaptive-halftone to combine the texture information of I^g with binary code information of $\tilde{\mathbf{E}}$ in an adaptive manner, resulting in the blueprint $I_b \in \mathbb{R}^{H \times W}$. Note that we leave the facial region unmodified to maintain rich facial features without compromising scanning robustness. The resultant blueprint I_b is then fed

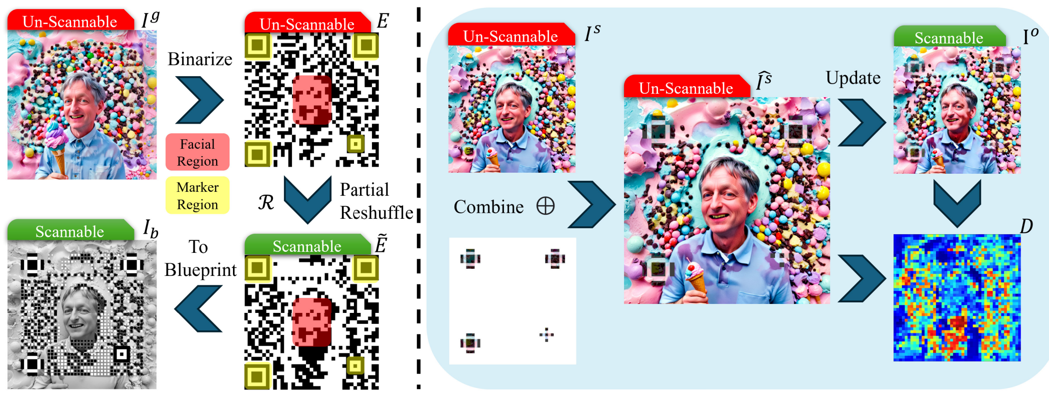


Figure 3: Illustration of IDRS (left) and IDSE (right). In IDRS, we maintain the information codes within the face and marker regions (red and yellow masks) and remap the remaining modules accordingly. In IDSE, we strengthen the finder and alignment pattern, and update in latent space using adaptive loss to enhance scannability. Visualization D shows the difference between I^o and $\widehat{I^s}$. Compared with uniform loss, adaptive loss modifies face region more gently.

into $\mathcal{SD}$ for the second generation:

$$I^s = \mathcal{SD}(c, z_0 | \mathcal{C}_{qr}(I_b, c, z_0), \mathcal{C}_{id}(f, c, z_0)). \tag{5}$$

Compared with the first generation in Equation 1, both controllers in stage 2 include facial information to mitigate conflicts. As shown in Figure 2, the result of stage 2 reduces errors by more than half compared to stage 1, while consistently preserving face identity information.

3.2 Scannability Enhancement

The resultant QR image I^s from stage 2 contains a certain QR pattern and consistently reveals face identity, but it is still unscannable by common QR decoders. In this part, we design the ID-Preserved Scannability Enhancement (IDSE) module to achieve the following two goals: 1) minimize modifications to the QR image (especially for facial region) to ensure its scannability; 2) enhance the marker region to better harmonize it without compromising scanning robustness. As illustrated in Figure 3 (right), we first strengthen the finder and alignment pattern of I^s, and then refine it using dynamic code loss to reach a harmonious balance between face ID, visual appeal and scannability.

3.2.1 Marker Harmonziation

The functional regions of a QR code, especially the finder and alignment patterns, are crucial for the decoder to locate the QR code. Therefore, these patterns on I^s are strengthened to generate $\widehat{I^s}$. Specifically, for pixel $\mathbf{p} \in \theta_k$ where $k \in \Delta_m$, we have:

$$\widehat{I^s}(\mathbf{p}) = \begin{cases} I^s(\mathbf{p}) - \min(I^s(\mathbf{p}) - \tau(1+\lambda), 0), & \text{if } \mathbf{E}_k = 1, \\ I^s(\mathbf{p}) - \max(I^s(\mathbf{p}) - \tau(1-\lambda), 0), & \text{if } \mathbf{E}_k = 0, \end{cases} \tag{6}$$

where $\lambda \in (0, 1)$ is a hyper-parameter, typically set to 0.8 by default. This threshold-based enhancement helps ensure that the functional regions of the output QR image are easily located.

3.2.2 Spatially Dynamic Loss Function

Inspired by [43], we use gradient descent to update the latent code of $\widehat{I^s}$ to optimize certain loss function. However, instead of using a fixed loss function with constant coefficients, we propose to leverage a spatially dynamic loss function.

Given a pretrained VQ-VAE [39] with the encoder $\mathcal{E}$ and the decoder $\mathcal{D}$, the optimization is given by:

$$\hat{z} = \underset{z}{\operatorname{argmin}}\, \mathcal{L}(\mathcal{D}(z), I_b, \widehat{I^s}), \tag{7}$$

where $z \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 4}$ is the latent code. The loss function $\mathcal{L}$ consists of an aesthetic content loss $\mathcal{L}_a$ and a spatially dynamic code loss $\mathcal{L}_c$:

$$\hat{z} = \underset{z}{\operatorname{argmin}}\{\mathcal{L}_c(\mathcal{D}(z), I_b) + \mathcal{L}_a(\mathcal{D}(z), I^s)\}. \tag{8}$$

We initialize z to $\mathcal{E}(\widehat{I^s})$, and use Adam [19] as the optimizer with a learning rate of 0.002 to iteratively update z until convergence. Finally, the output $I^o = \mathcal{D}(\hat{z})$ achieves robust scannability and high visual quality.

Adaptive Code Loss. A simulated decoder [36] using a 2D Gaussian kernel can extract module-wise information consistent with common QR decoders. The variance σ of the Gaussian kernel is a key factor in balancing visual quality and scanning robustness. However, in our scenario, we want the facial region to be smooth and the background region to be lossless. Therefore, we propose a spatially dynamic code loss. Let $Z = \mathcal{D}(z)$, the loss of j-th module is calculated by:

$$s_j = w(j) \times \text{avg}\{[Z(\theta_j) - I_b(\theta_j)] \odot G(j)\}, \tag{9}$$

where $\odot$ denotes the Hadamard product. $G(j) \in \mathbb{R}^{a \times a}$ is a weighting kernel, and $w(j)$ is a weighting factor defined by:

$$G(j) = \begin{cases} G_{\sigma_f}, & \text{if } j \in \Delta_f, \\ G_{\sigma_b}, & \text{otherwise} \end{cases}; \; w(j) = \begin{cases} w_f, & \text{if } j \in \Delta_f, \\ w_b, & \text{otherwise}, \end{cases} \tag{10}$$

where G_σ is a 2D Gaussian kernel with variance σ. The specific settings for the hyper-parameters w_f, w_b, σ_f, and σ_b can be found in the experiments section. Finally, the adaptive code loss is computed by:

$$\mathcal{L}_c(Z, I_b) = \sum_{j=1}^{n^2} w(j) \times \text{avg}\{[Z(\theta_j) - I_b(\theta_j)] \odot G(j)\}. \tag{11}$$

Gaussian distribution with bigger σ is flatter, which helps equalize the color within the module when updating the latent code. Although this makes modules easier to decode after iterations, bigger σ might create unnatural shadow in the face region. On the other hand, Gaussian distribution with smaller σ effectively regulates only the central region of a module, making the modules remain unscannable even after updates.

This problem is addressed by utilizing adaptive loss for different regions, *i.e.*, applying smaller weight w_f and σ_f in the face region to prevent distortion on face, and relatively larger w_f and σ_f in remaining region to maintain balance between scannability and aesthetic quality.

Aesthetic Content Loss. To ensure the retention of aesthetic qualities while preserving face ID and enhancing scannability, we use the aesthetic content loss to retain essential visual characteristics. It is quantified by calculating L^2-Wasserstein distance [1] (denoted as D_{W2}) of feature representations between Z and $\widehat{I^s}$ as follows:

$$\mathcal{L}_a(Z, \widehat{I^s}) = \sum_i D_{W2}(g_i(Z), g_i(\widehat{I^s})), \tag{12}$$

where g_i is feature representations from a pre-trained VGG-19 [34] network at layer i. The aesthetic content loss reflects the global aesthetic quality of the image. By optimizing both code loss and content loss, IDSE module adeptly balances the aesthetic quality, face-preserving, and scannability and creates optimal customized QR code images.

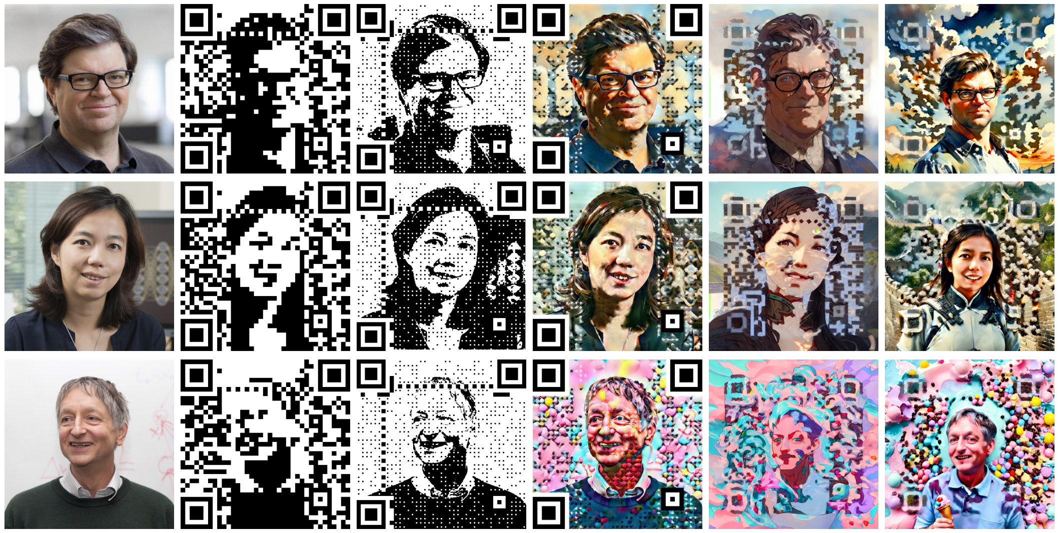
4 Experiments

4.1 Experimental Setup and Configuration

We implemented our pipeline in Python using the PyTorch framework and conducted experiments on an NVIDIA GeForce 4090 GPU. The scannability of QR images is tested using a 27-inch IPS display monitor with a refresh rate of 144Hz. In our experiments, we set control strengths for the InstantID network [40] and QR Controller at 0.8 and 1.4, respectively. The parameter λ in the marker harmonization process defaults to 0.8. The VAE configuration is consistent with the SD model. The

Table 1: Visual comparison of different methods.

Input	QArt [6]	Halftone [5]	ArtCoder [36]	Text2QR [43]	Face2QR



face recognition model AntelopeV2 from InsightFace [14] assists the generation of face mask M_f in IDRE. The VGG-19 architecture, pre-trained on the MS-COCO dataset, facilitates the feature map extraction in IDSE. The Adam optimizer powers the optimization within IDSE, performing 300 iterations at a learning rate of 0.002. Default settings for σ_f, σ_b, w_f, and w_b are 1.5, 3.0, 1.0, and 15.0 respectively. We produce QR code in version 5, each with 37×37 modules. For clarity, we define e as the number of error modules in QR image I^o (excluding finder and alignment pattern areas), and e_f as the number of error modules within the face region. Our dataset for comparative analysis contains 200 uniquely stylized QR images, each 1024×1024 pixels in size, with diverse visual content and artistic styles. To more accurately assess the preservation of face identity, we define the face feature distance d as the cosine similarity between the facial features (extracted using ArcFace [7]) of the generated QR image I^o and the original face image f.

4.2 Qualitative Comparison

Aesthetic Quality. In our comparative study, we evaluate our approach against several state-of-the-art aesthetic QR code generation techniques, including QArt [6], Halftone QR code [5], ArtCoder [36] and Text2QR [43], as detailed in Table 1. QArt, Halftone QR and Text2QR take the original face image f as the primary input, except that Text2QR takes in additional prompt input c. As ArtCoder is based on neural-style transfer technique, we employ f and I^g to serve as the content reference and the style reference respectively. The results show that Artcoder tends to render the texture of style image to face region, causing unwanted distortion on the face. Text2QR, on the other hand, cannot preserve face ID due to lack of specific control mechanisms for the face region. In contrast, our QR codes are adept at harmoniously integrating face ID, background and QR pattern, thereby achieving superior visual quality as well as scannability.

Identity Preservation. The comparison between original face image f and the generated image I^o is shown in Table 2. The face ID is well preserved in the final generated QR image I^o, with minimal change in haircut or facial expression, which can be further customized by users by adding prompt. The facial region is consistent with the background in style, and the QR pattern is blended seamlessly into the picture. We also compare the generated image I^o with output of the baseline pipeline InstantID [40] in Table 3, which shows that our pipeline achieves a similar level of identity preservation as the baseline. The outcomes displayed in Table 4 demonstrate that Face2QR consistently generates high-quality images across various poses.

Table 2: Visual comparison of face ID preservation in face image f and generated QR image I^o. I_1^o and I_2^o are generated from f_1, and I_3^o and I_4^o are generated from f_2. Face feature distance d is measured between pairs of I^o and f.

f_1	I_1^o	I_2^o	f_2	I_3^o	I_4^o

4.3 Quantitative Comparsion

Scanning Robustness. In this study, we assess the scanning robustness of our QR images using different scanning applications. We first generate a batch of 20 aesthetically pleasing QR codes, each with a dimension of 1,024 $\times$ 1,024 pixels. These QR images are then displayed on a high-definition monitor in three standard sizes: 3cm $\times$ 3cm, 5cm $\times$ 5cm, and 7cm $\times$ 7cm. During our controlled test, smartphones are held at a fixed distance of 25cm from the display, and each code is scanned for 3 seconds from different angles. Over a total of 50 trials, the percentage of successful scans is recorded in Table 5. The results reveal an average success rate exceeding 94%, showcasing high reliability of the generated QR images in diverse practical settings. It is also noted that QR images that fail the test in 3s can eventually be scanned if given more time. The scanning success rate is similar to that of Text2QR [43], as presented in our comparative analysis.

Table 3: Visual comparison of identity preservation with InstantID [40].

Input	InstantID [40]	Face2QR

Table 4: Generated QR images (second row) using face images (first row) with different poses.

$\sim 90°$	$\sim 45°$	$0°$

Subjective Study. Figure 4 presents a user study consisting of 30 participants to evaluate 150 QR images (50 for each methods) generated by different methods (the approval from Institutional Review Board is obtained). Participants are asked to choose the better one from a pair of pictures in the aspect of face ID preservation and aesthetic quality. Each pair of QR images are generated by different methods using the same face image as input. The percentages represent how many times users prefer the results of a method over the other. Our results are preferred by most users.

Objective Study. Table 6 shows the statistical performance measured by taking the average of 100 samples. We use the feature distance d, varying from -1 to 1, as a quantifiable measure for the

Table 5: Scannability success rates of QR codes across various decoders at different sizes and angles.

Decoders	Success Rate (%)					
	$(3\text{cm})^2$		$(5\text{cm})^2$		$(7\text{cm})^2$	
	45°	90°	45°	90°	45°	90°
	Face2QR (ours)					
Scanner	100	<u>94</u>	100	100	100	100
TikTok	100	100	100	100	100	100
WeChat	96	100	100	100	<u>94</u>	100
	Text2QR [43]					
Scanner	96	96	100	100	100	100
TikTok	100	100	100	100	100	100
WeChat	100	<u>94</u>	100	100	<u>94</u>	100

Table 6: Comparison of average face feature distance d and average Aesbench scores B_a. [Key: **Best**]

	ArtCoder [36]	Text2QR [43]	Face2QR
d	0.50	0.43	**0.51**
B_a	62.0	87.5	**90.1**

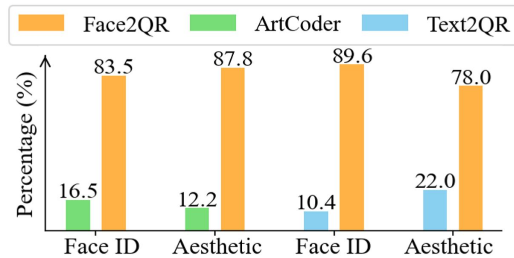


Figure 4: User study of different methods.

preservation of face ID. The higher the distance d, the generated face is more consistent with the original face image. We also use AesBench tool [8], which assigns aesthetic scores ranging from 0 to 100 (with higher scores denoting better aesthetics), to objectively evaluate the aesthetic quality of generated pictures. The results indicate that our approach exceeds competing methods in all evaluated metrics, confirming its capability to produce QR iamges with faithfully preserved face ID and high aesthetic quality.

4.4 Ablation Study

Table 7: IDRE Ablation Study: Compared with result obtained by completing the entire stage 2 (rightmost column), skipping stage 2 (leftmost column) or conducting stage 2 without IDRE (middle column) will result in more error modules e_f in the face region. The distribution of error modules, highlighted as bright areas, is depicted in the second row.

I^g	w/o. IDRE	with IDRE

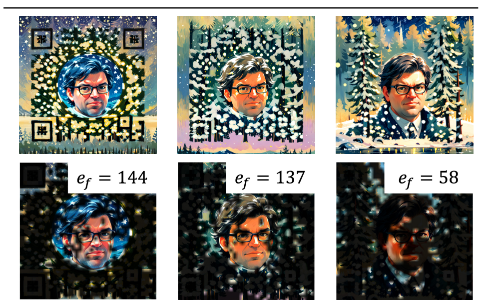


Table 8: IDSE Ablation Study: We examine the influence of $w_f, w_b, \sigma_f, \sigma_b$ to the generated QR image I^o. Model 1 ($w_f = w_b, \sigma_f = \sigma_b$) tends to create undesirable shadow in the face region. Model 2 ($w_f = w_b, \sigma_f < \sigma_b$) leads to an increased number of error modules e. Model 3 ($w_f < w_b, \sigma_f < \sigma_b$) reaches a balance between face ID and scannability. Second row zooms in on the face region.

I^s	Model 1	Model 2	Model 3

IDRE Module. Table 7 illustrates how skipping stage 2 or the IDRE module affects the scannability of resultant QR image I^s. In stage 2, IDRE first rearranges the QR modules to construct a blueprint I_b in which the QR pattern is compatible with face ID, and then SD model generates I^s guided by I_b. If stage 2 is bypassed, the output I^g from stage 1 is used as the input for stage 3 directly. If IDRE is omitted, a normal QR code is used to guide the QR pattern generation in stage 2. The result shows that skipping stage 2 or IDRE results in a substantial increase in the number of error modules e_f within the facial area, which significantly reduces the scannability of the QR code.

IDSE Module. In stage 3, the IDSE module leverages adaptive code loss, which is key to maintaining face identity and simultaneously decreasing number of error modules. This loss function is determined by two parameter pairs (σ_f and σ_b for Gaussian kernel; w_f and w_b for loss strength). Table 8 presents images produced by the IDSE with varying parameter configurations. The comparison shows that a uniform σ leads to distortions in the facial area, and a universal loss weight w will increase the number of error modules and compromise the generated image's scannability. A clearer comparison is given by the normalized image difference visualization D shown in Table 9. The visualization D demonstrates the module-wise difference between the QR images before and after IDSE. The results demonstrate that the adaptive loss makes the modification in the face region gentler than uniform loss, reaching a better balance between face identity and scannability.

Table 9: Image difference visualization D of uniform loss (first row) and adaptive loss (second row).

I^o	D	face region

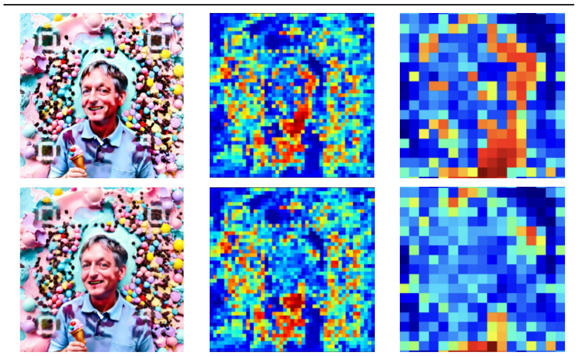


Table 10: Bad cases caused by failure of generative model.

Input	Output

5 Conclusion

In summary, we introduce Face2QR, an innovative pipeline that seamlessly integrates face ID, aesthetic design and scannability in the generation of QR codes. By introducing three key modules, *i.e.* IDQR for integrating face ID with aesthetic background, IDRS for resolving conflict between face ID and QR pattern, and IDSE for enhancing scannability while preserving face ID and aesthetic quality, our pipeline is able to balance between three inherently conflicting control signals and generate QR codes that preserve face ID, aesthetic quality and scannability at the same time. Extensive experiments demonstrate that Face2QR significantly outperforms previous methods, establishing a new benchmark for generating ID-preserved aesthetic QR codes.

Limitations. Our method is still constrained by some limitations of generative models. Although generative models are powerful, they can produce inconsistent results and often require substantial computing power to generate detailed, high-resolution images. Some typical bad cases caused by failure of generative models are shown in Table 10. As these computational models become more advanced, we can anticipate further improvements in the accuracy, speed, and overall aesthetic quality of the generated QR codes.

Broader Impact. By enhancing the visual appeal and personal connection of QR codes, our work has the potential to revolutionize their use in entertainment, social media, marketing, and personal memorabilia, transforming them from mere tools for information transfer into objects of personal expression and aesthetic value. Looking forward, we anticipate that future work will not only refine these methods but also explore their integration into various technological ecosystems, consistently enriching the social and functional aspects of QR codes.

Acknowledgement. The work was supported in part by the National Natural Science Foundation of China under Grant 62301310 and 62225112, and in part by Sichuan Science and Technology Program under Grant 2024NSFSC1426.

References

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[2] Changsheng Chen, Wenjian Huang, Lin Zhang, and Wai Ho Mow. Robust and Unobtrusive Display-to-Camera Communications via Blue Channel Embedding. *IEEE Transactions on Image Processing*, 28(1):156–169, 2018.

[3] Changsheng Chen, Wenjian Huang, Baojian Zhou, Chenchen Liu, and Wai Ho Mow. PiCode: A New Picture-Embedding 2D Barcode. *IEEE Transactions on Image Processing*, 25(8):3444–3458, 2016.

[4] Changsheng Chen, Baojian Zhou, and Wai Ho Mow. RA Code: A Robust and Aesthetic Code for Resolution-Constrained Applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(11):3300–3312, 2018.

[5] 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.

[6] Russ Cox. Qartcodes. https://research.swtch.com/qart, 2012.

[7] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[8] Huang et al. Aesbench: An expert benchmark for multimodal large language models on image aesthetics perception. *arXiv preprint arXiv: 2401.08276*, 2024.

[9] Han Fang, Dongdong Chen, Feng Wang, Zehua Ma, Honggu Liu, Wenbo Zhou, Weiming Zhang, and Neng-Hai Yu. TERA: Screen-to-Camera Image Code with Transparency, Efficiency, Robustness and Adaptability. *IEEE Transactions on Multimedia*, 24:955–967, 2022.

[10] Han Fang, Weiming Zhang, Hang Zhou, Hao Cui, and Nenghai Yu. Screen-Shooting Resilient Watermarking. *IEEE Transactions on Information Forensics and Security*, 14(6):1403–1418, 2018.

[11] Anthony Fu. Stylistic qr code with stable diffusion. https://antfu.me/posts/ai-qrcode, 2023.

[12] Zhongpai Gao, Guangtao Zhai, and Chunjia Hu. The Invisible QR Code. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pages 1047–1050, 2015.

[13] Gonzalo J Garateguy, Gonzalo R Arce, Daniel L Lau, and Ofelia P Villarreal. QR Images: Optimized Image Embedding in QR Codes. *IEEE Transactions on Image Processing*, 23(7):2842–2853, 2014.

[14] Jia Guo, Xiang An, Jinke Yu, Jing Yang, Alexandros Lattas, Baris Gecer, and Jiankang Deng. Insightface: A 2d and 3d face analysis project, 2023.

[15] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[16] Jun Jia, Zhongpai Gao, Kang Chen, Menghan Hu, Xiongkuo Min, Guangtao Zhai, and Xiaokang Yang. RIHOOP: Robust Invisible Hyperlinks in Offline and Online Photographs. *IEEE Transactions on Cybernetics*, pages 1–13, 2020.

[17] Jun Jia, Zhongpai Gao, Dandan Zhu, Xiongkuo Min, Guangtao Zhai, and Xiaokang Yang. Learning invisible markers for hidden codes in offline-to-online photography. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[18] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of StyleGAN. In *Proc. CVPR*, 2020.

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[20] Dongxu Li, Junnan Li, and Steven CH Hoi. Blip-diffusion: Pre-trained subject representation for controllable text-to-image generation and editing. *arXiv preprint arXiv:2305.14720*, 2023.

[21] Wenhao Li, Guangyang Wu, Wenyi Wang, Peiran Ren, and Xiaohong Liu. Fastllve: Real-time low-light video enhancement with intensity-aware look-up table. In *ACM Int. Conf. Multimedia*, 2023.

[22] Zhen Li, Mingdeng Cao, Xintao Wang, Zhongang Qi, Ming-Ming Cheng, and Ying Shan. Photomaker: Customizing realistic human photos via stacked id embedding. *arXiv preprint arXiv:2312.04461*, 2023.

[23] Zehua Ma, Xi Yang, Han Fang, Weiming Zhang, and Nenghai Yu. Oacode: Overall aesthetic 2d barcode on screen. *IEEE Transactions on Multimedia*, 2023.

[24] Maxim Maximov, Ismail Elezi, and Laura Leal-Taixé. Ciagan: Conditional identity anonymization generative adversarial networks. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020.

[25] Chong Mou, Xintao Wang, Liangbin Xie, Jian Zhang, Zhongang Qi, Ying Shan, and Xiaohu Qie. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models. *arXiv preprint arXiv:2302.08453*, 2023.

[26] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.

[27] Pexels. Pexels: Free stock photos, royalty free stock images & videos. https://www.pexels.com/, 2024.

[28] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.

[29] Irving S Reed and Gustave Solomon. Polynomial codes over certain finite fields. *Journal of the society for industrial and applied mathematics*, 8(2):300–304, 1960.

[30] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[31] Zhihao Shi, Xiaohong Liu, Chengqi Li, Linhui Dai, Jun Chen, Timothy N. Davidson, and Jiying Zhao. Learning for unconstrained space-time video super-resolution. *IEEE Trans. Broadcast.*, 68(2):345–358, 2022.

[32] Zhihao Shi, Xiaohong Liu, Kangdi Shi, Linhui Dai, and Jun Chen. Video frame interpolation via generalized deformable convolution. *IEEE Trans. Multim.*, 24:426–439, 2022.

[33] Zhihao Shi, Xiangyu Xu, Xiaohong Liu, Jun Chen, and Ming-Hsuan Yang. Video frame interpolation transformer. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022.

[34] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Yoshua Bengio and Yann LeCun, editors, *ICCV*, 2015.

[35] Hao Su, Jianwei Niu, Xuefeng Liu, Qingfeng Li, Ji Wan, and Mingliang Xu. Q-Art Code: Generating Scanning-robust Art-style QR Codes by Deformable Convolution. In *ACM Int. Conf. Multimedia*, 2021.

[36] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021.

[37] Matthew Tancik, Ben Mildenhall, and Ren Ng. Stegastamp: Invisible Hyperlinks in Physical Photographs. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020.

[38] Dani Valevski, Danny Lumen, Yossi Matias, and Yaniv Leviathan. Face0: Instantaneously conditioning a text-to-image model on a face. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–10, 2023.

[39] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

[40] Qixun Wang, Xu Bai, Haofan Wang, Zekui Qin, and Anthony Chen. Instantid: Zero-shot identity-preserving generation in seconds. *arXiv preprint arXiv:2401.07519*, 2024.

[41] Wenyi Wang, Guangyang Wu, Weitong Cai, Liaoyuan Zeng, and Jianwen Chen. Robust prior-based single image super resolution under multiple gaussian degradations. *IEEE Access*, 8:74195–74204, 2020.

[42] Eric Wengrowski and Kristin Dana. Light Field Messaging with Deep Photographic Steganography. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019.

[43] Guangyang Wu, Xiaohong Liu, Jun Jia, Xuehao Cui, and Guangtao Zhai. Text2qr: Harmonizing aesthetic customization and scanning robustness for text-guided qr code generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8456–8465, 2024.

[44] Guangyang Wu, Xiaohong Liu, Kunming Luo, Xi Liu, Qingqing Zheng, Shuaicheng Liu, Xinyang Jiang, Guangtao Zhai, and Wenyi Wang. Accflow: Backward accumulation for long-range optical flow. In *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023.

[45] Guangyang Wu, Lili Zhao, Wenyi Wang, Liaoyuan Zeng, and Jianwen Chen. Pred: A parallel network for handling multiple degradations via single model in single image super-resolution. In *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2019.

[46] 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 Trans. Multim. Comput. Commun. Appl.*, 17(1):25:1–25:23, 2021.

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

[48] Xingqian Xu, Jiayi Guo, Zhangyang Wang, Gao Huang, Irfan Essa, and Humphrey Shi. Prompt-Free Diffusion: Taking" Text" out of Text-to-Image Diffusion Models. *arXiv preprint arXiv:2305.16223*, 2023.

[49] Yuxuan Yan, Chi Zhang, Rui Wang, Yichao Zhou, Gege Zhang, Pei Cheng, Gang Yu, and Bin Fu. Facestudio: Put your face everywhere in seconds. *arXiv preprint arXiv:2312.02663*, 2023.

[50] 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.

[51] Liming Zhai, Qing Guo, Xiaofei Xie, Lei Ma, Yi Estelle Wang, and Yang Liu. A3gan: Attribute-aware anonymization networks for face de-identification. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 5303–5313, 2022.

[52] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023.

A Appendix

A.1 Additional Experiments

A.1.1 Additional Results

Our Face2QR pipeline is generalizable to real faces, generated realistic faces, and cartoon faces. The experimental results in Table 11 demonstrate that facial identities are well preserved and seamlessly blended into the background in all generated QR images, showcasing the effectiveness of Face2QR across these three face types.

A.1.2 Visualization of Intermediate Results

In Table 14, we show the intermediate results of Face2QR. Here, I^g represents the output of stage 1, I_b signifies blueprint image generated by IDRE, and I^s denotes the results of regeneration results from stage 2. The blueprint I_b guides both the generation of I^s in IDQR module within stage 2, and the IDSE process in stage 3 to generate QR image I^o with a harmonious balance between face ID, aesthetic quality and scannability. Table 12 presents results from different iterations in the IDSE process. Additionally, Table 15 presents the prompts and models used in the generation of the aforementioned QR image samples.

A.1.3 Loss Curve & Running Time

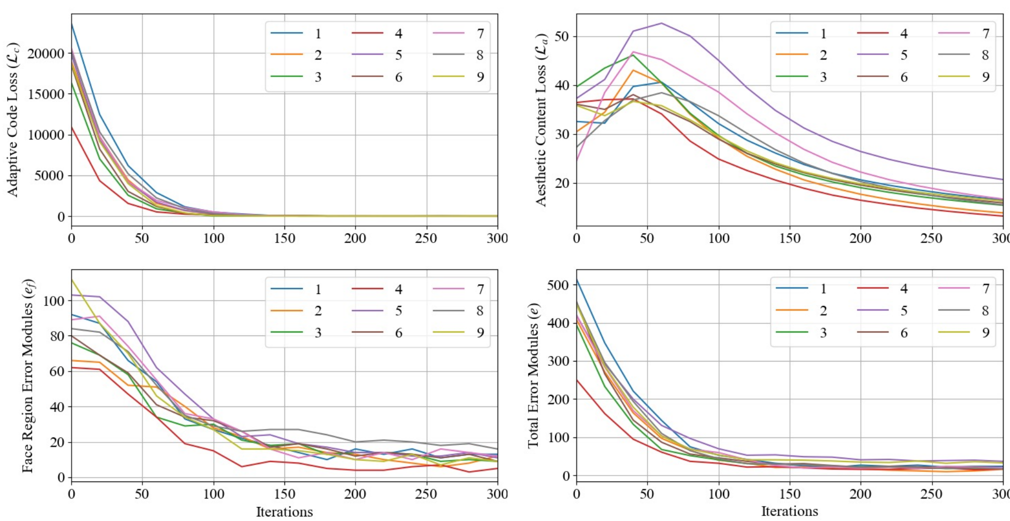


Figure 5: Curve of different metrics during IDSE. We show metric curves for diverse samples, each represented by distinct colors. These curves illustrate metric variations over 300 iterations.

In stage 3, we use IDSE to enhance the scannability of I^s by updating the its latent code. The dynamic loss function consists of aesthetic content loss $\mathcal{L}_a$ and adaptive code loss $\mathcal{L}_c$. Both losses apply at the same time and helps the update process to converge sooner. The total number of error module e acts as a indicator of scannability, and the error module number in the face region e_f helps visualize the modification process in the face region. Figure 5 illustrates the above four metrics. The IDSE process converges in about 120 seconds when executed on an NVIDIA 4090 GPU to enhance images of size 1024×1024 pixels.

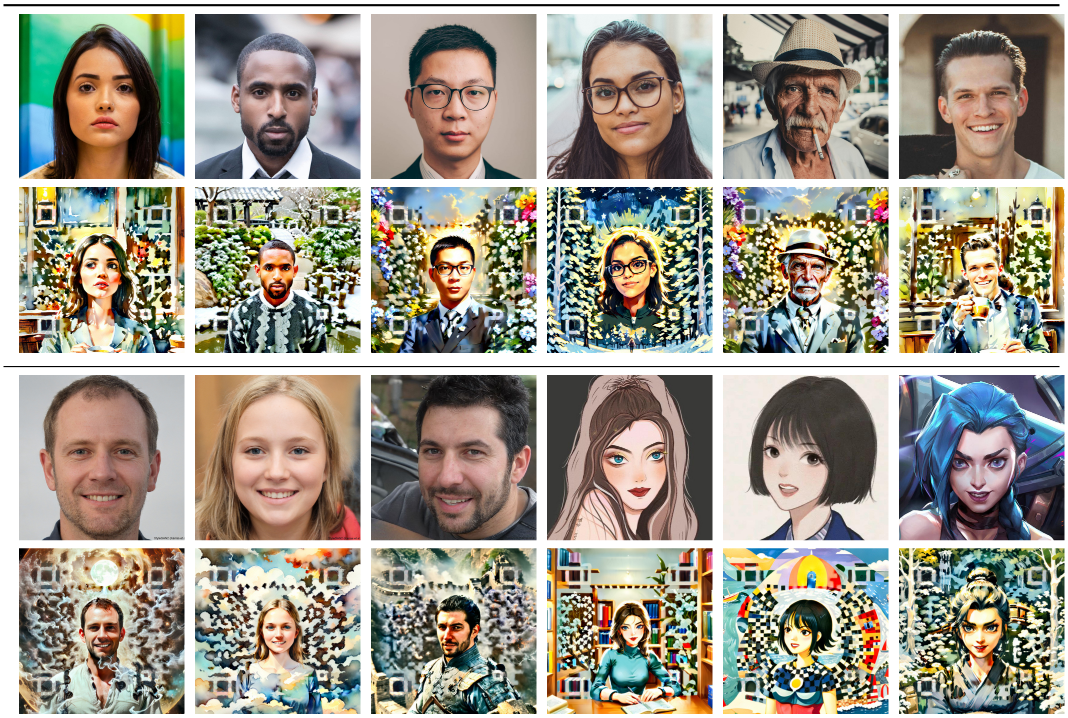
A.2 Bad Cases

In addition to bad cases shown in Table 10, we present suboptimal cases when the face image and the prompt are conflict with each other in Table 13. For example, the first row in Table 13 shows the case when a face image of a woman and the prompt "A male man" are given at the same time.

A.3 The Interface for User Study

The scoring interface of user study is shown in Figure 6. We adopt the pair-wise comparisons for subjective study rather than absolute ratings since the former is relatively more accurate in general.

Table 11: Real face images of ordinary people (Row 1) collected from [27], generated realistic face (left three on Row 3) using StyleGAN2 [18] and cartoon faces (right three on Row 3) with corresponding QR images I^o (Rows 2 and 4).



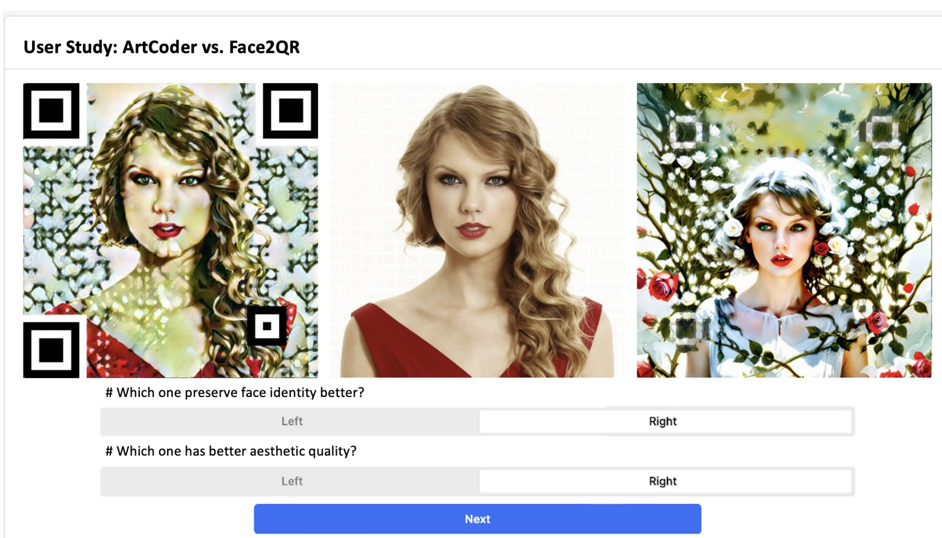


Figure 6: The user interface.

Table 12: Visualization of IDSE process at different iteration steps.

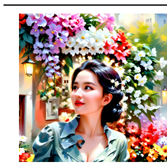
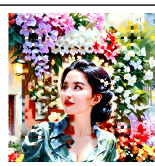
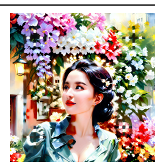
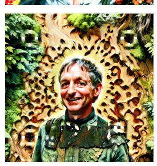
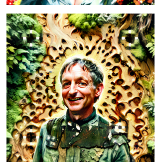
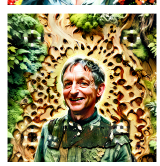
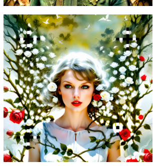
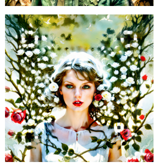
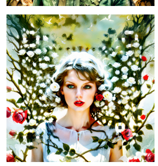
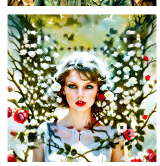
Input	Iteration 100	Iteration 200	Iteration 300
			
			
			

Table 13: Bad I^s results caused by conflict between face image and prompt.

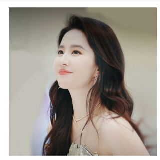
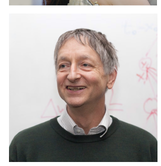
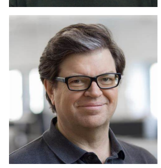
Input	I^s
	
	
	

Table 14: Visualization of intermediate results during our aesthetic QR code generation pipeline.

Input	I^g	I_b	I^s	$\hat{I}^s$	I^o
un-scannable	un-scannable	scannable	un-scannable	un-scannable	scannable

Table 15: Prompts for generated QR codes in the paper. All images are generated with size of 1,024×1,024. The generative model is uniformly SDXL Unstable Diffusers YamerMIX.

Sample	Prompt
Figure 1 Col 1	"A female woman, face in the middle. white bird sitting on a branch of roses, digital watercolor illustration of a meadow with white roses in the morning light, detailed fantastic background of Salvador Dali, waterhouse, Canaletto, watercolor art, intricate, complex contrast, HDR, sharp, soft cinematic volumetric lighting, the background is lost in haze. The foreground is brightly lit. 4k"
Figure 1 Col 2	"A male man, face in the middle. J. R. R. Tolkien-inspired landscape photo, a magical landscape inspired by J. R. R. Tolkien The Lord of the Rings, hilly path, bathed in a breathtaking play of sunlight splashing on surfaces, presents bark textures with color gradients in wood-earth tones, Jungle, mossy rock formations, complicated plants. HDR, Creating a photorealistic, asymmetrical composition, complicated details, very detailed, by Greg Rutkowski"
Figure 1 Col 3	"A female woman, face in the middle. olpntng style, ink wash in green and gold tones, Landscape of lotus flowers in the foreground over a lake, muted colours, wet on wet technique, sketch ink watercolor style with a hint of orange and white by Wu Guanzhong, Truong Lo, Mary Jane Ansell, Agnes Cecile, muted splatter art, gold ink splatter, faded dripping paints. green monochrome, soft impressionistic brushstrokes, oil painting, heavy strokes, dripping paint, oil painting, heavy strokes, paint dripping"
Figure 1 Col 4	"A male man, face in the middle. flat stylized pine trees, winter landscape with starry night sky and lake, painterly, acrylic painting, trending on pixiv fanbox, palette knife and brush strokes, style of makoto shinkai jamie wyeth james gilleard edward hopper greg rutkowski studio ghibli genshin impact"
Figure 1 Col 5	"A female woman, face in the middle. (best quality:1.5), (intricate emotional details:1.5), (sharpen details), (ultra detailed), (cinematic lighting), sorcerer's ancient library, ,floating candles, mystical artifacts, magical books, oxfort Key Elements:"
Figure 1 Col 6	"A male man, face in the middle. colorful birds sitting on top of a pink flower, fantasy, parrot by Adam MarczyÅ ski, fantasy art, art of alessandro pautasso, glowing oil,detailed beautiful animals, artwork in the style of guweiz"
Table 1 Row 1	"A male man, face in the middle. painted clouds and landscape background, Watercolor, trending on artstation, sharp focus, studio photo, intricate details, highly detailed, by greg rutkowski"
Table 1 Row 2	"A female woman, face in the middle. UHD, (masterpiece) Landscape of the Great Wall of China, smoke effects, trending on artstation, sharp focus, intricate details, highly detailed,"
Table 1 Row 3	"A male man, face in the middle. A ocean of pastel pink blue and lilac ice cream, with a boat made of candy, waves"
Table 2 Col 2	"A male man, face in the middle. Hatsune Mecha Tech Sense HD Wallpaper, ultra hd, realistic, vivid colors, highly detailed, UHD drawing, pen and ink, perfect composition, beautiful detailed intricate insanely detailed octane render trending on artstation, 8k artistic photography, photorealistic concept art, soft natural volumetric cinematic perfect light"
Table 2 Col 3	"A male man, face in the middle. in the style of james gilleard, SamDoesArts, art by Sam Yang, absolute beauty birth'd from fragile chaos, mandelbulb dress, insanely detailed, full of life, animated"
Table 2 Col 5	"A male man, face in the middle. impressionist landscape of a Japanese garden in winter with a bridge over a pond"
Table 14 Row 7&8	"A female woman, face in the middle. Highly detailed beautiful landscape, vintage style, bright colors, atmospheric lighting flowers, cinematic composition, digital painting, elegant, beautiful, high detail, by Willem Haenraets, trending on artstation, sharp focus, studio photo, intricate details, highly detailed, by greg rutkowski"

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: In abstract, the main contributions of this paper are emphasized. Furthermore, in the last paragraph of the introduction, these contributions are clearly listed again.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Please refer to Section 5.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: This paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide all needed information to reproduce the main experimental results of this paper in Section 4. Our code will be released upon publication.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We consider publishing the code at https://github.com/cavosamir/Face2QR once we complete our patent application process.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All experiments details are illustrated in Section 4.1.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: This paper mainly conducts qualitative comparisons and subjective experiments. Therefore, the corresponding error bars are not applicable.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Computational resources have been described in Section 4.1.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: This work is conducted in accordance with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Please refer to the Section 5.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The creators or original owners of assets (e.g., code, data, models), used in the paper, are properly credited, and the license and terms of use are explicitly mentioned and are properly respected.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The new assets introduced in the paper are well documented alongside the assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: This paper includes the full text of instructions given to participants and screenshots, and the human subjects are paid at least the minimum wage in the country of the data collector, following the NeurIPS Code of Ethics.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes]

    Justification: There is no such potential risks aware for research with human subjects in this paper. We have obtained the IRB approval and also adhere to the NeurIPS Code of Ethics.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.