Article

An Adaptive Edge-Guided Dual-Network Framework for Decoding-Oriented QR Code Motion Deblurring

Jianping Li , Dongyang Guo , Wenjie Li , Wei Zhao *

Posted Date: 3 June 2026

doi: 10.20944/preprints202606.0184.v1

Keywords: motion deblurring; QR code deblurring; edge-guided attention; image prior; adaptive network

 doi:10.20944/preprints202606.0184.v1

Article

An Adaptive Edge-Guided Dual-Network Framework for Decoding-Oriented QR Code Motion Deblurring

Jianping Li [1,2], Dongyang Guo [3], Wenjie Li [4] and Wei Zhao [4,*]

[1] Guangdong Provincial Key Lab of Robotics and Intelligent Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2] University of Chinese Academy of Sciences, Beijing 101408, China
[3] School of Nano Science and Technology, University of Science and Technology of China, Suzhou 215127, China
[4] Shenzhen University of Advanced Technology, Shenzhen 518107, China
* Correspondence: zhao.wei@siat.ac.cn

Abstract

Unlike natural image deblurring, which primarily prioritizes perceptual quality, Quick Response (QR) code deblurring aims to ensure successful decoding. QR codes are characterized by highly structured patterns with sharp edges, which provide strong structural priors for restoration. However, existing deep learning methods rarely exploit these priors explicitly. To address this limitation, we propose the Edge-Guided Attention Block (EGAB), which incorporates explicit edge priors into a Transformer architecture. Built upon EGAB, we develop the Edge-Guided Restormer (EG-Restormer) to improve the decoding rate for severely blurred QR codes. For mildly blurred inputs, we introduce a Lightweight and Efficient Network (LENet) that achieves rapid deblurring with minimal computational overhead. To leverage the complementary strengths of both networks, we further integrate EG-Restormer and LENet into an Adaptive Dual-network (ADNet), which dynamically selects the appropriate restoration branch according to the input blur severity of the input, making it particularly suitable for real-time QR code deblurring. Extensive experiments demonstrate the effectiveness of our approaches. Specifically, EG-Restormer achieves state-of-the-art decoding performance, improving the decoding success rate on QRData by 8.67 percentage points under GoPro-only training and 1.33 percentage points after GoPro pre-training followed by QRData fine-tuning. Moreover, ADNet reduces inference latency by 19% while maintaining high decoding accuracy, making it well-suited for real-time QR code deblurring.

Keywords: motion deblurring; QR code deblurring; edge-guided attention; image prior; adaptive network

1. Introduction

Quick Response (QR) codes have become ubiquitous in daily life, with applications in mobile payment, logistics, product traceability, industrial automation, and human–machine interaction. A QR code is a two-dimensional matrix barcode composed of black and white modules arranged on a regular square grid, together with functional patterns such as finder markers, timing patterns, and alignment patterns, as shown in Figure 1 [1]. However, practical QR scanning is frequently performed in dynamic environments, where cameras may be handheld, mounted on mobile robots, or embedded in resource-constrained devices. These acquisition scenarios render captured QR code images highly susceptible to motion blur, typically induced by camera shake, object motion, or relative movement between the imaging device and the target scene [2,3]. This degradation poses a significant challenge, as even mild blur may corrupt module boundaries, distort finder patterns, or weaken timing patterns, leading to decoding failure. Although substantial progress has been made in natural image deblurring, restoring motion-blurred QR codes remains a distinct and practically important challenge.

 doi:10.20944/preprints202606.0184.v1

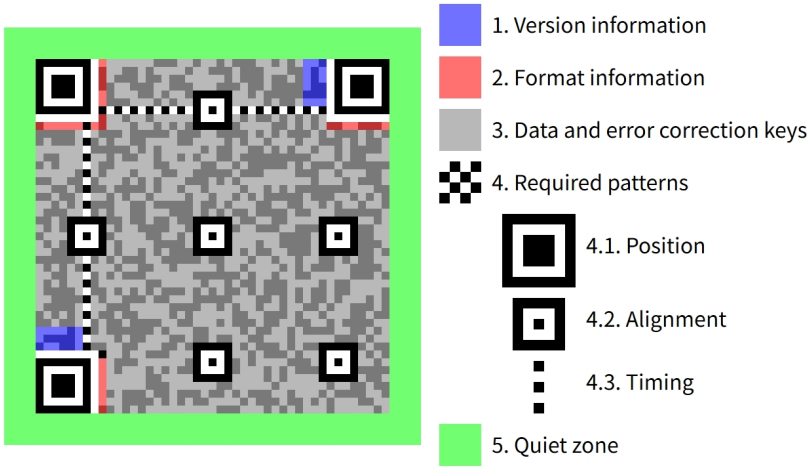


Figure 1. Structure of a QR code, highlighting functional elements. Source: Wikipedia [1].

Different from natural image deblurring, which often emphasizes perceptual quality or pixel-level fidelity, QR code deblurring is inherently a decoding-oriented restoration task [5]. For a restored natural image, small local artifacts may still be visually acceptable. In contrast, a QR code must preserve strict geometric layouts and binary structures to be successfully decoded. This fundamental difference changes the objective of restoration: instead of hallucinating visually plausible textures, a QR code deblurring model should recover sharp module boundaries, regular grid layouts, and structurally consistent black-white transitions. Therefore, the key challenge is not only to improve conventional image quality metrics such as PSNR or SSIM, but also to maximize the decoding success rate under challenging blur conditions [6].

To restore blurred QR codes, most of previous optimization-based methods have exploited QR-code-specific priors to constrain the solution space. For example, Van Gennip et al. [7] proposed a regularization approach that uses corner features of QR codes to estimate the point spread function (PSF), while Rioux et al. [8] incorporated barcode symbology constraints into kernel estimation through Kullback–Leibler divergence. Sörös et al. [4] proposed a fast blur removal algorithm for QR code by estimating the blur from the salient edges in an iterative optimization scheme involving sharpening, blur estimation, and decoding. These methods demonstrate that QR codes contain strong structural priors that can facilitate restoration. However, these optimization-based approaches typically rely on simplified blur models and hand-crafted assumptions, making them less effective for complex, spatially varying, or severe motion blur in practical scenarios.

With the rapid development of deep learning, learning-based methods have become the dominant paradigm for image restoration and deblurring. Convolutional neural networks (CNN), multi-stage architectures, generative adversarial networks (GAN), attention mechanisms, and Transformer-based models have achieved remarkable performance on natural image deblurring benchmarks [9–13,15]. Representative methods include progressive multi-stage restoration networks [12], U-shaped Transformer architectures [9], efficient high-resolution restoration Transformers [11], nonlinear activation-free networks [15], frequency-domain Transformers [13], and selective-frequency models [14]. These methods are effective for natural images because they can capture rich textures, long-range dependencies, and complex degradation patterns. Nevertheless, directly applying them to QR code deblurring is suboptimal, since they are designed primarily for perceptual or fidelity-oriented reconstruction rather than decoding-oriented structural recovery.

Several learning-based approaches have been developed specifically for QR code or barcode restoration. Pu et al. [16] proposed a dual CNN framework for restoring degraded 2D barcodes. Li et al. [17] combined a deep convolutional network with an adaptive thresholding technique to address motion blur and uneven illumination. Dong et al. [18] and Wang et al. [19] further explored GAN-based methods for restoring severely motion-blurred QR codes. Although these methods [16–19] adapt general deblurring models to the QR code domain, most of them still depend mainly on implicit

 doi:10.20944/preprints202606.0184.v1

feature learning. They rarely encode the domain-specific structural priors of QR codes explicitly, which may result in outputs that appear visually sharp but still contain distorted module boundaries, damaged finder patterns, or inconsistent grid structures, ultimately leading to decoding failure.

This limitation motivates us to resort to the intrinsic structure of QR codes. As illustrated in Figure 1, a standard QR code consists of binary modules arranged on a regular two-dimensional grid, together with functional components such as finder patterns, timing patterns, alignment patterns, and data modules. Such highly structured patterns produce gradient statistics that are substantially different from those of natural images. As shown in Figure 2(b), the gradient histogram of QR code images exhibits stronger and more concentrated edge responses than that of natural images. Moreover, these strong gradients are not randomly distributed; instead, they follow highly regular horizontal and vertical spatial arrangements, as illustrated in Figure 2(c). These observations indicate that edge information is not merely a low-level cue but a structural prior closely related to QR code readability. These properties suggest that explicit edge modeling can serve as an effective structural prior for QR code deblurring, especially when motion blur severely weakens module boundaries.

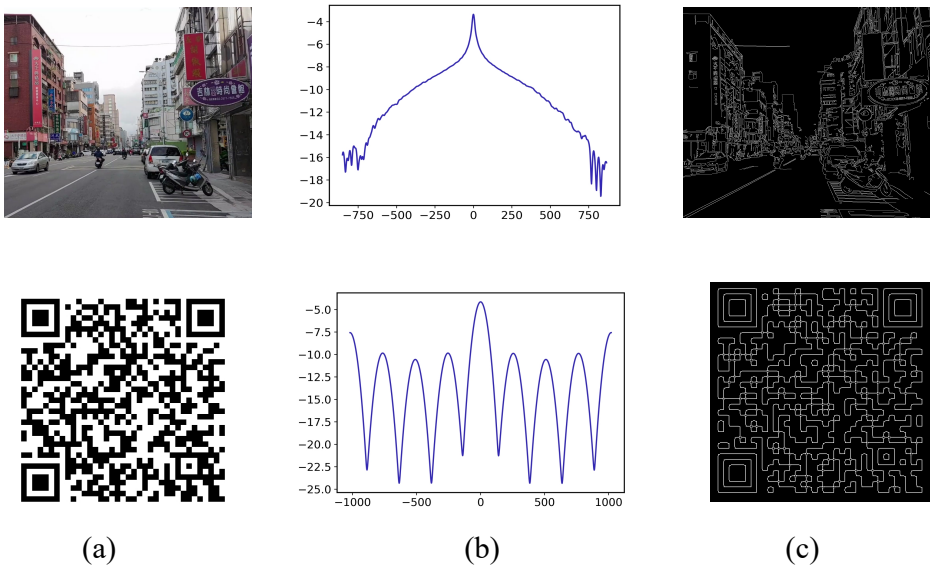


Figure 2. Comparison of gradient statistics: natural image vs QR code. (**a**) original sharp image; (**b**) log-scale gradient histograms of (a); (**c**) gradient magnitudes of (a).

To effectively leverage this edge prior, two critical challenges must be addressed. First, existing learning-based QR code deblurring methods typically learn restoration mappings in a fully data-driven manner, without explicitly modeling multi-directional edge priors that are essential for preserving QR code structures. As a result, they may produce visually acceptable but structurally unreliable outputs, leading to decoding failure. Second, practical QR code scanning is commonly performed on mobile, embedded, or edge devices, where computational resources and latency are strictly constrained. A large restoration network may be necessary for severely blurred inputs but inefficient for mildly blurred ones, while a lightweight model may be efficient but insufficient for challenging cases. Therefore, an effective QR code deblurring system should not only restore severe blur accurately, but also allocate computation adaptively according to input difficulty.

To address these challenges, we propose an edge-guided and adaptive framework for QR code motion deblurring. The framework consists of three key components. First, we introduce the Edge-Guided Attention Block (EGAB), which explicitly extracts multi-directional edge features and integrates them into a Transformer attention mechanism. By incorporating edge priors into feature aggregation, EGAB directs the network to focus on structurally critical regions, improving the accuracy of module boundary restoration. Based on EGAB, we construct the Edge-Guided Restormer (EG-Restormer), a Transformer-based restoration network designed for severely blurred QR codes. Second, to efficiently

 doi:10.20944/preprints202606.0184.v1

restore mildly blurred inputs, we further propose a Lightweight and Efficient Network (LENet), which achieves fast restoration with significantly lower computational cost. Finally, we integrate EG-Restormer and LENet into an Adaptive Dual-network (ADNet), which dynamically selects the appropriate restoration branch according to the estimated blur severity. This adaptive mechanism balances decoding accuracy and inference efficiency, making the proposed framework suitable for real-time QR code applications.

The main contributions of this work are summarized as follows:

- We propose the Edge-Guided Attention Block (EGAB), which explicitly incorporates multi-directional edge priors into a Transformer architecture. Based on EGAB, we develop EG-Restormer, an edge-guided Transformer-based network that significantly improves the decoding success rate of severely blurred QR codes.
- We introduce a Lightweight and Efficient Network (LENet) for fast restoration of mildly blurred QR codes, and further integrate it with EG-Restormer into an Adaptive Dual-network (ADNet). The proposed ADNet dynamically selects the appropriate restoration branch according to input blur severity, achieving a favorable balance between decoding accuracy and inference efficiency.
- We propose an acceleration-based non-uniform motion blur synthesis algorithm to generate a QR code dataset with real-world-like blur characteristics. On this dataset, extensive experiments demonstrate that the proposed EG-Restormer achieves state-of-the-art performance in decoding rate, while ADNet substantially reduces inference latency without sacrificing decoding accuracy.

2. Related Work

2.1. Prior Knowledge for Two-Tone Image Deblurring

Image deblurring is a highly ill-posed inverse problem, as multiple latent sharp images and blur kernels may produce the same blurred observation. To regularize the solution space, traditional deblurring methods often introduce image priors into the restoration process. For natural image deblurring, traditional methods typically exploit statistical assumptions such as sparse image gradients, L_0 gradient, total variation regularization, dark channel priors, or kernel constraints to estimate the blur kernel and recover the latent image [3,22–25]. These priors are effective for natural images, where edges and textures provide useful cues for kernel estimation and latent image recovery. However, two-tone images, such as document text, 1D barcodes, and QR codes, exhibit visual statistics that are substantially different from natural images [26]. They are usually composed of piecewise constant foreground/background regions, sparse but strong gradients, and sharp structural boundaries. These properties have inspired a series of deblurring methods that explicitly incorporate domain-specific priors.

Document and text image deblurring. Document and text images share several properties with barcode-like patterns, such as sharp foreground-background transitions and limited intensity levels. To exploit these properties for deblurring, Chen et al. [27] proposed a document image deblurring algorithm by exploiting the distinct characteristics of document images and introducing a content-aware prior based on foreground segmentation. Pan et al. [28,29] further proposed an L_0-regularized intensity and gradient prior for text image deblurring. Later, Jiang et al. formulated a two-tone prior to guide the latent image toward a piecewise constant image with two dominant gray levels [30]. These methods demonstrate that task-specific priors can provide stronger constraints than generic natural image priors for recovering sharp and readable structures. Nevertheless, text and document images do not contain the deterministic symbology and grid layouts of QR codes, making QR code restoration a more structured and decoding-sensitive problem.

1D barcode deblurring. For 1D barcode restoration, traditional methods often modeled barcode recovery as a signal recovery or statistical inference problem. For example, Turin and Boie [31] introduced a deterministic expectation–maximization algorithm to recover degraded barcode signals, improving barcode readability under blur and noise. Kresić-Jurić et al. [32] further applied hidden Markov models to barcode decoding, where barcode edge locations were inferred from noisy scan

 doi:10.20944/preprints202606.0184.v1

signals. These methods show that barcode-specific structural information can effectively constrain the deblurring process. However, their focus on one-dimensional or relatively simple barcode structures makes them less suitable for QR codes, which involve two-dimensional grid layouts, finder patterns, timing patterns, and module-level spatial dependencies.

Traditional QR code deblurring. Due to their explicit and deterministic structural priors, QR codes are particularly suitable for prior-driven restoration. Choksi et al. [33] proposed an anisotropic total variation regularized L_1 approximation model for denoising and deblurring 2D barcodes. Their formulation favors axis-aligned rectangular structures, which are consistent with barcode geometry. Van Gennip et al. [7] proposed a regularization-based blind deblurring and denoising framework for QR codes by incorporating the finder patterns of the QR symbology into known regularization methods. Rioux et al. [8] further introduced a Kullback–Leibler divergence based blind deblurring method that incorporates 2D barcode symbology constraints into both kernel estimation and latent image recovery. Zheng et al. [34] proposed a blind QR code deblurring method using intensity and gradient priors of positioning patterns, where the known structural properties of finder patterns are used to guide sub-regional blur estimation. These studies highlight the importance of QR-code-specific priors in constraining the deblurring process and improving structural recovery.

Different from both optimization-based prior methods and purely data-driven deblurring networks, our method explicitly integrates QR-code-specific edge priors into a Transformer-based architecture. In particular, the proposed Edge-Guided Attention Block (EGAB) extracts multi-directional edge features and uses them to guide attention learning. This design combines the flexibility of deep feature learning with the reliability of structural QR code priors, enabling more robust decoding-oriented deblurring under severe motion blur.

2.2. Self-Attention and Transformers for Image Deblurring

Deep neural networks have become the dominant paradigm for image restoration and deblurring. Convolutional neural networks have achieved remarkable progress in image deblurring due to their strong local feature extraction capability and efficient hierarchical representation. However, convolution-based methods are inherently limited by local receptive fields and static convolution kernels, making it difficult to adaptively model long-range dependencies that are important for severely degraded images. Self-attention provides a powerful alternative by allowing each spatial location to aggregate information from other locations according to content-adaptive similarities. As a result, vision Transformers and their variants have been widely studied in image restoration.

Several Transformer-based architectures have shown strong potential for image restoration. Vision Transformers were first popularized for image recognition by Dosovitskiy et al. [35]. IPT [36] applies large-scale Transformer pretraining to multiple restoration tasks. SwinIR [37] employs shifted-window self-attention to reduce computational cost while preserving local contextual modeling. Uformer [9] proposes a U-shaped Transformer architecture with locally-enhanced window attention for image restoration. These methods show that Transformer-based architectures can effectively capture long-range dependencies and achieve competitive or superior restoration performance. However, standard self-attention has quadratic complexity with respect to spatial resolution, while window-based attention restricts the interaction range. These limitations can be particularly problematic for high-resolution deblurring, where both efficiency and global context are important.

Restormer [11] addresses this challenge by proposing an efficient Transformer architecture specifically designed for high-resolution image restoration. Its core component is the Multi-Dconv Head Transposed Attention (MDTA) module as shown in Figure 3. Unlike standard multi-head self-attention, which computes attention over the spatial dimension, MDTA computes self-attention across the channel dimension. This transposed attention formulation substantially reduces computational complexity and makes the model suitable for high-resolution restoration tasks. Moreover, MDTA introduces local context before attention computation by using 1×1 convolution for channel mixing and depth-wise convolution for spatially local feature aggregation. Therefore, MDTA combines the global modeling ability of self-attention with the locality bias of convolution.

 doi:10.20944/preprints202606.0184.v1

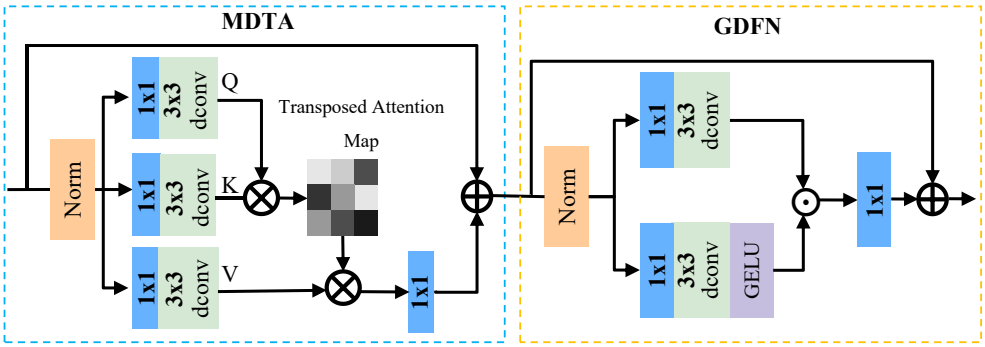


Figure 3. The Transformer block of the Restormer, some details are omitted for clarity. The block consists of a multi-Dconv head transposed attention (MDTA) and a Gated-Dconv feed-forward network (GDFN).

Another important component of Restormer is the Gated-Dconv Feed-Forward Network (GDFN) as shown in Figure 3. In a standard Transformer, the feed-forward network independently transforms features through linear layers and nonlinear activation. GDFN modifies this design by incorporating a gating mechanism and depth-wise convolutions. Specifically, it expands the feature channels, applies depth-wise convolution to encode local spatial context, and uses an element-wise gating operation to control information flow. This design suppresses less informative features while allowing useful structural and textural cues to propagate through the network hierarchy.

Consequently, the combination of MDTA and GDFN enables Restormer to achieve strong performance on various image restoration tasks, including motion deblurring, defocus deblurring, deraining, and denoising. Although Restormer provides an effective backbone for image restoration, it is originally designed for generic natural images and does not explicitly incorporate QR-code-specific structural priors. For QR code deblurring, long-range dependency modeling alone is insufficient. The network also needs to preserve sharp binary transitions, regular grid structures, and reliable module boundaries to ensure successful decoding. Motivated by this observation, our work builds upon the efficient Restormer framework but embeds explicit edge priors into the attention mechanism. By guiding attention with QR code edge information, the proposed EGAB encourages the network to focus on the structural boundaries that are most critical for module-level recovery and decoding success.

3. Proposed Methods

As shown in Figure 4, this paper proposes an Adaptive Dual-network (ADNet) model for QR code motion deblurring, which integrates two pre-trained U-shape QR code deblurring networks: EG-Restormer and LENet. The dual-network design enables ADNet to adaptively select the appropriate model according to the blur severity of the input QR code, thereby optimizing both restoration performance and inference efficiency. EG-Restormer is designed to restore severely blurred QR codes by leveraging the proposed EGAB to explicitly incorporate edge priors, which are essential for successful decoding. In contrast, LENet is a lightweight and efficient model designed for mildly blurred QR codes, utilizing the Simple Gate Depthwise Convolution Block (SGDB) for fast deblurring. The ADNet framework assigns blurred QR codes to the appropriate model through the Blur Severity-based Routing (BSR) module, ensuring efficient allocation of computational resources while maximizing the decoding success rate. We first describe the overall pipeline of ADNet architecture (as shown in Figure 4). Then, we provide detailed explanations of the core components of proposed dual-network framework: **(a)** EG-Restormer, **(b)** LENet components, and **(c)** BSR module.

 doi:10.20944/preprints202606.0184.v1

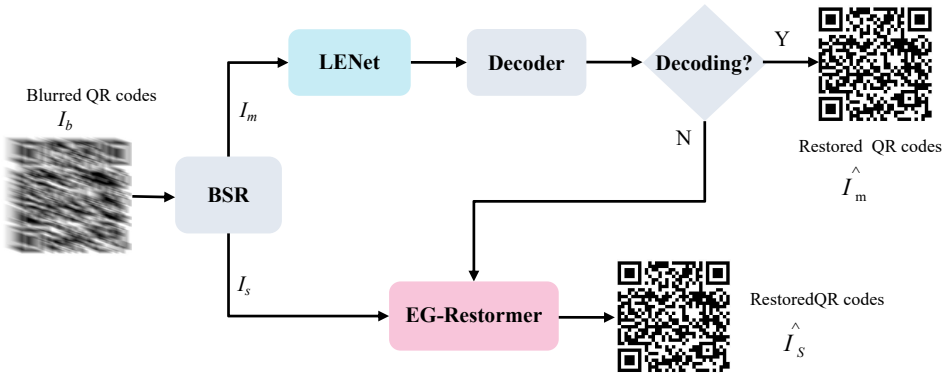


Figure 4. The overall architecture of the proposed ADNet framework, which consists of a dual-network with a powerful EG-Restormer and an efficient LENet, a decoder and the BSR module.

Overall Pipeline. Given a blurred input QR code $I_b \in \mathbb{R}^{H\times W\times 3}$, where H and W denote the height and width, respectively, the Blur Severity-based Routing (BSR) module first classifies I_b into two categories according to blur severity: severely blurred inputs I_s and mildly blurred inputs I_m. Severely blurred inputs I_s are routed to EG-Restormer model to obtain a sharpened output $\hat{I}_s$. While mildly blurred inputs I_m are first processed by LENet for fast deblurring. The output of LENet is then evaluated for successful decoding by a decoder to determine whether successful decoding can be achieved. If decoding fails, the output is rerouted to EG-Restormer for further refinement. Otherwise, the output of LENet $\hat{I}_m$ is regarded as the final restored output of ADNet. Therefore, the proposed ADNet is designed not only to ensure successful decoding of QR codes, but also to improve the efficiency of the deblurring model.

3.1. Edge-Guided Restormer (EG-Restormer)

To build an effective Transformer model for restoring severely blurred QR codes, this study explicitly leverages the unique structural properties of QR codes, particularly their sharp edges, which are critical for successful decoding. We propose EG-Restormer, a Transformer-based architecture that explicitly incorporates edge priors into multi-head self-attention through the Edge-Guided Attention Block (EGAB). We first introduce the overall architecture of EG-Restormer, followed by a detailed description of the EGAB and its core component, the Edge Feature Extractor (EFE) module.

As shown in Figure 5, EG-Restormer adopts a U-shaped hierarchical Transformer architecture that progressively extracts and restores multiple-scale features. Given a severely blurred input QR code I_s,EG-Restormer first extracts low-level features F_0 through a convolutional layer. The extracted features are then fed into a four-layer encoder, where the spatial resolution is progressively downsampled while the channel dimension is expanded for deep feature extraction. The decoder then iteratively upscales the bottleneck features F_l to high-resolution representations F_r. Finally, a convolutional layer is applied to generate a residual QR code R, which is added to the input for restoration. Both the encoder and decoder contain multiple Edge-guided Transformer Blocks (EGTBs), which guide edge-aware feature restoration and help preserve critical structural details for successful decoding.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 June 2026 doi:10.20944/preprints202606.0184.v1

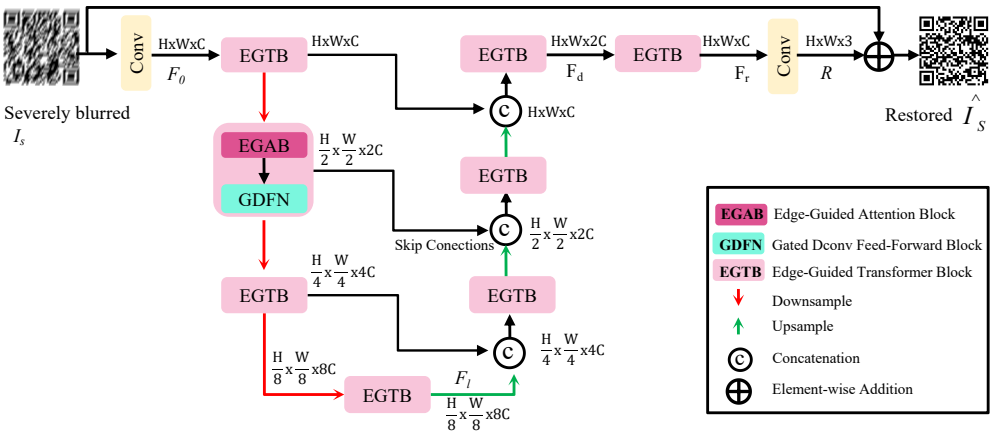
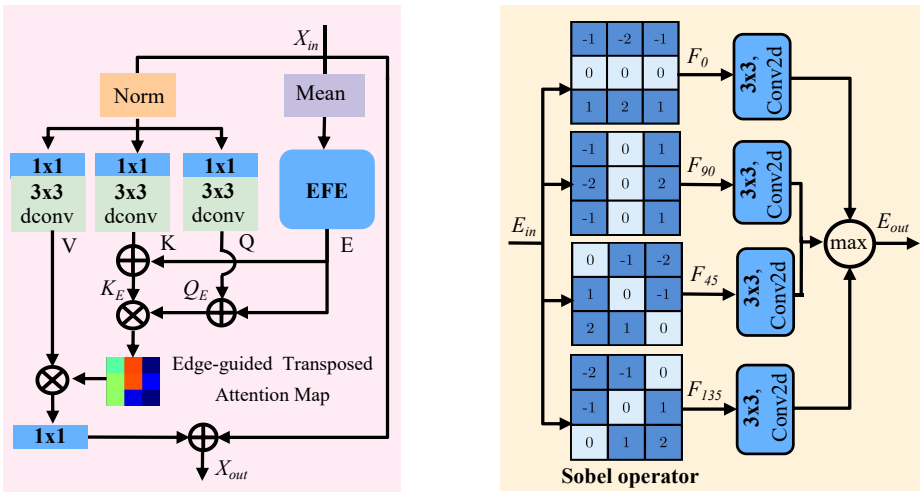


Figure 5. Architecture of the EG-Restormer for severely blurred QR code deblurring. Our EG-Restormer employs a U-shaped architecture with multiple edge-guided Transformer blocks (EGTB). The core module of EGTB is edge-guide attention block (EGAB) that incorporates an explicit edge prior into the attention block.

3.1.1. Edge-Guided Attention Block (EGAB)

Edge sharpening is crucial for successful QR code decoding, since module boundaries, finder patterns, and timing patterns are mainly characterized by sharp black-white transitions. Conventional learning-based methods usually learn such structural information implicitly, which may lead to visually plausible but structurally unreliable restoration results. However, directly modeling such priors in conventional attention mechanisms may introduce high computational complexity. To address this issue, we propose the Edge-Guided Attention Block (EGAB), as illustrated in Figure 6(a), which integrates QR code edge priors by modulating the query and key in multi-Dconv head transposed attention (MDTA) module [11].



(a) EGAB (b) EFE

Figure 6. The core module of Edge-guided Transformer Block and its components. (a) The Edge-Guided Attention Block (EGAB)–it uses the Edge Feature Extractor (EFE) module to inject an explicit edge feature E into the attention mechanism. (b) The structure of the EFE module.

EGAB first generates an explicit edge prior map E from the input feature X_{in} using the proposed Edge Feature Extractor (EFE), a lightweight module designed to capture multi-directional structural edges. Meanwhile, query (Q), key (K), and value (V) matrices are generated from X_{in} for implicit feature representation. The extracted edge prior is then used to modulate Q and K, producing edge-aware query and key features Q_E and K_E. Finally, edge-guided transposed attention is computed

 doi:10.20944/preprints202606.0184.v1

based on the modulated query-key correlation. This design allows spatial edge structures, especially QR module boundaries and finder-pattern contours, to contribute more strongly to the attention computation. The detailed formulation of EGAB is presented as follows.

Given an input feature map $X_{in} \in \mathbb{R}^{C\times H\times W}$, EGAB first generates query, key, and value features through a 1×1 convolution followed by a 3×3 depth-wise convolution:

$$Q, K, V = \mathrm{DWConv}_{3\times 3}(\mathrm{Conv}_{1\times 1}(X_{in})), \tag{1}$$

where $Q, K, V \in \mathbb{R}^{C\times H\times W}$. Meanwhile, to obtain a compact structural representation for edge extraction, channel-wise mean aggregation is applied to X_{in}:

$$X_m = \mathrm{Mean}_c(X_{in}), \quad X_m \in \mathbb{R}^{1\times H\times W}. \tag{2}$$

The proposed Edge Feature Extractor (EFE) is then applied to X_m to produce an explicit edge prior map:

$$E = \mathrm{EFE}(X_m), \quad E \in \mathbb{R}^{1\times H\times W}. \tag{3}$$

The edge map is used to modulate the query and key features before computing transposed attention:

$$\begin{aligned} Q_E &= Q \odot (1 + \alpha E), \\ K_E &= K \odot (1 + \alpha E), \end{aligned} \tag{4}$$

where $\odot$ denotes element-wise multiplication and α is a fixed edge modulation coefficient. In our implementation, α is set to 0.1. Since E has the shape $\mathbb{R}^{1\times H\times W}$, it is broadcast along the channel dimension to match the shape of Q and K. In this way, spatial positions with stronger edge responses contribute more to the query-key correlation computation.

After edge modulation, Q_E, K_E, and V are reshaped into multi-head representations:

$$\hat{Q}_E, \hat{K}_E, \hat{V} \in \mathbb{R}^{N_h\times C_h\times HW}, \tag{5}$$

where N_h denotes the number of attention heads and $C_h = C/N_h$ is the number of channels per head. Following MDTA, $\hat{Q}_E$ and $\hat{K}_E$ are normalized along the spatial dimension. The edge-guided transposed attention is then computed as:

$$A = \mathrm{Softmax}\left(\bar{Q}_E \bar{K}_E^T \cdot T\right), \tag{6}$$

where $A \in \mathbb{R}^{N_h\times C_h\times C_h}$ is the channel-wise attention map and $T \in \mathbb{R}^{N_h\times 1\times 1}$ is a learnable temperature parameter for each attention head. Finally, the output feature is obtained by:

$$X_{out} = X_{in} + \mathrm{Conv}_{1\times 1}(A\hat{V}). \tag{7}$$

Although the attention matrix in MDTA is computed across channels rather than spatial positions, the proposed edge modulation still injects spatial structural information into the attention process. Specifically, the edge map reweights the spatial responses of Q and K before the channel-wise correlation is computed. Therefore, edge regions such as QR module boundaries and finder-pattern contours contribute more strongly to the transposed attention, encouraging the network to preserve decoding-critical structures during deblurring.

3.1.2. Edge Feature Extractor (EFE)

The Edge Feature Extractor (EFE), as illustrated in Figure 6(b), is designed to extract a compact and explicit structural prior from the input feature map. Specifically, EFE is designed to capture multi-directional edge features, which serves as an explicit structural prior for the subsequent attention mechanism in EGAB. Given a single-channel input feature map $\mathbf{E}_{\mathrm{in}} \in \mathbb{R}^{1\times H\times W}$, EFE applies four

 doi:10.20944/preprints202606.0184.v1

parallel Sobel operators to capture edge features along four principal orientations: 0°, 45°, 90°, and 135°. The corresponding Sobel kernels are defined as:

$$\mathbf{K}_{0^\circ} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \mathbf{K}_{45^\circ} = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \mathbf{K}_{90^\circ} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \mathbf{K}_{135^\circ} = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix}$$

Each directional feature is obtained by convolving the input with the corresponding fixed Sobel kernel, producing four orientation-specific edge feature maps. Formally, the directional edge features are computed as:

$$\mathbf{F}_\theta = \mathbf{E}_{\text{in}} * \mathbf{K}_\theta, \quad \theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\} \tag{8}$$

where $*$ denotes the convolution operation. To further refine these features, each $\mathbf{F}_\theta$ is passed through a learnable *conv*2*d* layer, allowing the network to adaptively enhance salient edge structures while suppressing irrelevant patterns.

The final output edge map $\mathbf{E}_{\text{out}}$ is obtained by selecting the maximum response across all orientations:

$$\mathbf{E}_{\text{out}} = \max_\theta \big(conv2d(\mathbf{F}_\theta)\big), \quad \theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}\big) \tag{9}$$

where $E_{out} \in \mathbb{R}^{1\times H\times W}$ is the final edge prior map. This max-pooling operation across orientations ensures that the most prominent edges dominate the final representation, thereby providing a robust structural prior that guides the attention mechanism toward critical regions. By explicitly modeling multi-directional edges, EFE enhances the sensitivity of EGAB to essential QR code structures, which is particularly beneficial for challenging restoration tasks such as deblurring and decoding under severe degradation.

3.2. *Lightweight and Efficient Network (LENet)*

Similarly, LENet is designed for fast deblurring of mildly blurred QR codes. as shown in Figure 7, it adopts a four-level U-shaped encoder-decoder architecture. Both the encoder and decoder contain Simple Gate Depthwise Convolution Block (SGDB) to efficiently extract and refine features. The decoder output is then passed to an Edge Sharpening Block (ESB), which employs a 3x3 depthwise convolution for spatial context refinement and edge enhancement.

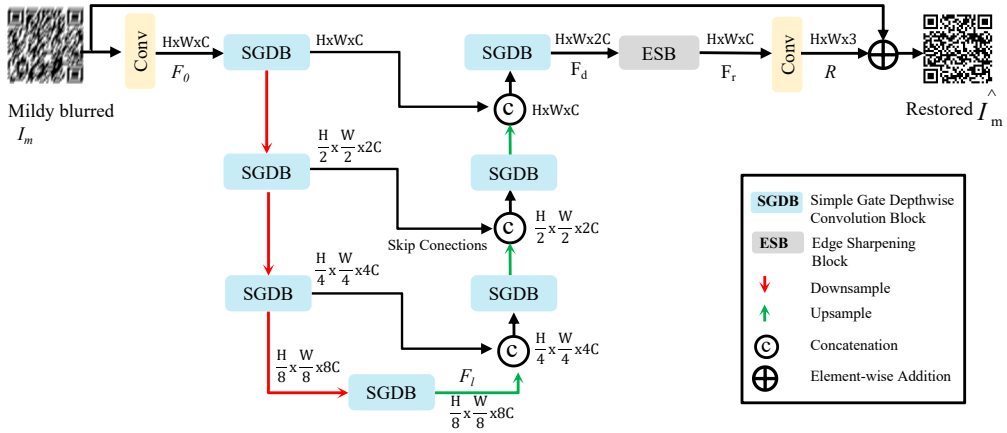


Figure 7. The overall architecture of LENet. we employs a U-shaped encoder-decoder architecture with multiple Simple Gate Depthwise Convolution Blocks (SGDB).

 doi:10.20944/preprints202606.0184.v1

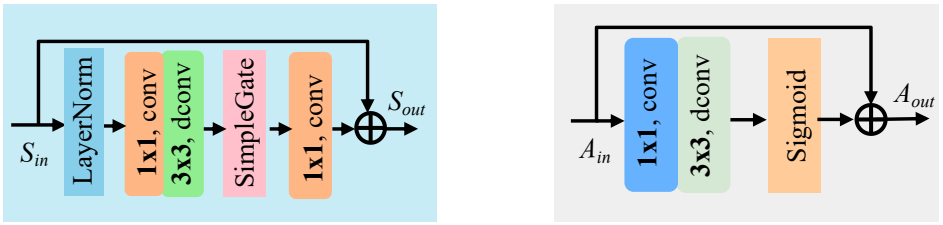
3.2.1. Simple Gate Depthwise Convolution Block (SGDB)

To efficiently extract and refine features for mildly blurred QR code deblurring for real-time QR code deblurring, we design a lightweight Simple Gate Depthwise Convolution Block (SGDB) in the U-shaped encoder–decoder network. Inspired by efficient architectures such as MobileNetV2 [20] and NAFNet [15], SGDB follows a streamlined processing paradigm that decouples channel mixing and spatial aggregation while maintaining strong representational capacity.

As shown in Figure 8(a), the input feature S_{in} is processed through: layer normalization (LN), 1x1 convolution (Conv) for channel expansion, 3x3 depthwise convolution (Conv2d) for spatial context aggregation, SimpleGate (SG) [15], and another 1x1 convolution for channel contraction, with a weighted residual connection producing refined output S_{out}. The overall operation can be formulated as:

$$S_{out} = Conv(SG(Conv2d(Conv(LN(S_{in}))))) + S_{in} \tag{10}$$

This design follows a simple yet effective restoration paradigm, in which normalization, channel expansion, depthwise spatial modeling, and gated feature selection are unified in a residual learning framework. As a result, SGDB achieves a strong balance between efficiency and representational power, making it well-suited for lightweight QR code deblurring on mobile devices.



(a) SGDB (b) ESB

Figure 8. The core blocks of the LENet: (a)Simple Gate Depthwise Convolution Block (SGDB) that performs efficient feature extraction and refinement, and (b) Edge Sharpening Block (ESB) that performs spatial context refinement and edge enhancement.

3.2.2. Edge Sharpening Block (ESB)

To further enhance high-frequency details and refine structural edges in the reconstructed QR codes, we introduce an Edge Sharpening Block (ESB), which is appended after the output of the U-shaped encoder–decoder backbone, as shown in Figure 8(b). ESB is designed to selectively emphasize edge-related features while preserving global content consistency through a lightweight residual formulation.

Given the intermediate feature map $A_{in} \in \mathbb{R}^{C\times H\times W}$ produced by the decoder, ESB generates a learned edge-aware attention map through a sequence of channel mixing, spatial context aggregation, and nonlinear gating operations. First, a 1×1 convolution is applied to A_{in} for channel mixing and feature fusion without altering spatial resolution. Next, the transformed features are processed by a 3×3 depthwise convolution for spatial context aggregation, which effectively enlarges the receptive field while maintaining computational efficiency. The resulting representation is passed through a sigmoid function to generate a soft spatial attention map where values are normalized into [0,1] to indicate pixel-wise importance of edge-related regions. Finally, a residual connection is employed to preserve the fidelity of the reconstructed QR codes while selectively enhancing structural details. The overall operation of ESB is computed as:

$$A_{out} = Sigmoid(Conv2d(Conv(A_{in}))) + A_{in} \tag{11}$$

The proposed ESB is designed as a lightweight edge-aware refinement module. By generating an attention map that emphasizes edge-related features and applying it in a residual manner, ESB

 doi:10.20944/preprints202606.0184.v1

enhances critical structural details without introducing significant computational overhead or causing over-sharpening. Therefore, it is well-suited for efficient deblurring of mildly blurred QR codes in LENet.

3.3. Blur Severity-based Routing (BSR) Module

To achieve both a high decoding success rate and computational efficiency, we introduce the Blur Severity-based Routing (BSR) module, as illustrated in Figure 4. This module categorizes input QR codes I_b as mildly blurred I_m or severely blurred I_s QR codes using a pre-calibrated threshold τ on Laplacian variance (LV), a sharpness metric in which a higher LV indicates less blur. This threshold is derived from LENet's performance on a dataset sampled from training data. Specifically, τ is calculated as the midpoint between the minimum LV of the blurriest decodable restored QR codes and the maximum LV of the sharpest non-decodable ones:

$$\tau = \frac{\min(\mathcal{L}_{\text{decodable}}) + \max(\mathcal{L}_{\text{non-decodable}})}{2} \tag{12}$$

where $\mathcal{L}_{\text{decodable}}$ and $\mathcal{L}_{\text{non-decodable}}$ denote the set of LV scores for QR codes that are decodable and non-decodable after LENet restoration, respectively. This adaptive routing strategy ensures that mildly blurred QR codes are processed efficiently by LENet, while severely blurred ones are assigned to EG-Restormer for stronger restoration capability. In this way, the proposed framework optimizes both performance and efficiency in real-world applications. The algorithmic implementation of the BSR module is detailed in Algorithm 1.

Algorithm 1 Blur Severity-based Routing Strategy

Input: I_b: a blurred QR code,
  M_E: the large network (EG-Restormer).
  M_L: the lightweight network (LENet).
  τ: a pre-calibrated blur severity threshold.
  $\mathcal{M}(\cdot)$: a blur metric function (Laplacian variance).
  $\mathcal{D}(\cdot)$: a QR code decoder.
Output: Restored QR code $\hat{I}$.
function BLURROUTER($I_b, M_E, M_L, \tau, \mathcal{M}, \mathcal{D}$)
    $lv \leftarrow \mathcal{M}(I_b)$ ▷ Calculate the blur severity score
    if $lv > \tau$ **then** ▷ Case 1: mild blur detected
        $\hat{I}_L \leftarrow M_L(I_b)$ ▷ Route to LENet
        if $\mathcal{D}(\hat{I}_L)$ succeeds **then**
            return $\hat{I}_L$
        else
            $\hat{I}_E \leftarrow M_E(I_b)$ ▷ Fallback to EG-Restormer
            return $\hat{I}_E$
        end if
    else ▷ Case 2: severe blur detected
        $\hat{I}_E \leftarrow M_E(I_b)$ ▷ Route to EG-Restormer
        return $\hat{I}_E$
    end if
end function

4. Experiments

4.1. Datasets

We constructed a synthetic QR code Dataset (QRData), containing 1,822 images degraded by non-uniform motion kernels for training and evaluation. To evaluate the effectiveness of the proposed QR code deblurring network in real-world scenarios, it is crucial to train and test the model on images that mimic physical camera motion. Conventional linear motion blur, which assumes constant velocity, often fails to capture the complex dynamics of handheld camera shake. Therefore, we propose a

 doi:10.20944/preprints202606.0184.v1

non-uniform motion blur synthesis algorithm that incorporates an acceleration parameter to simulate realistic trajectory dynamics.

(1) Mathematical Model

The degradation of a sharp QR code image I can be modeled as the convolution with a motion blur kernel (Point Spread Function, PSF) K, followed by the addition of noise N:

$$B = I * K + N \tag{13}$$

where B is the resulting blurred image. Unlike standard linear kernels, our kernel K is generated using a non-linear temporal mapping. Let L denote the total motion length and θ denote the motion angle. We define the normalized linear time as $t_{lin} \in [0, 1]$. To introduce non-uniformity, we apply a power-law transformation using an acceleration factor a:

$$t_{acc} = (t_{lin})^a \tag{14}$$

where $a > 1$ or $a < 1$ correspond to different acceleration or deceleration profiles, respectively.

(2) Kernel Generation Process

The spatial coordinates (x, y) of the motion trajectory within a kernel of size $S \times S$ are determined by:

$$\begin{cases} x(t_{acc}) = t_{acc} \cdot L\cos(\theta) - \frac{L\cos(\theta)}{2} + \frac{S}{2} \\ y(t_{acc}) = t_{acc} \cdot L\sin(\theta) - \frac{L\sin(\theta)}{2} + \frac{S}{2} \end{cases} \tag{15}$$

To ensure that the generated kernel is physically plausible, weights are assigned to each sampled point along the trajectory. The residence time at each coordinate determines its contribution to the PSF intensity. Bilinear interpolation is used to map these continuous coordinates onto the discrete pixel grid of K, ensuring smooth and realistic blur transitions.

The complete non-uniform motion blur kernel synthesis procedure is summarized in Algorithm 2.

Algorithm 2 Non-uniform Motion Blur Kernel Synthesis

Input: Length L, Angle θ, Acceleration factor a, Kernel size S
Output: Non-uniform motion blur kernel K
Initialize K as an $S \times S$ zero matrix
Sample N_p points uniformly to create linear time vector $t_{lin} = \{0, \ldots, 1\}$
Compute accelerated time vector: $t_{acc} = (t_{lin})^a$
for each $t \in t_{acc}$ **do**
    $x \leftarrow t \cdot L\cos(\theta) - \frac{L\cos(\theta)}{2} + \frac{S}{2}$
    $y \leftarrow t \cdot L\sin(\theta) - \frac{L\sin(\theta)}{2} + \frac{S}{2}$
    Calculate weight w based on local velocity $\frac{d}{dt}(t_{acc})$
    Perform *Bilinear Interpolation* to distribute weight w to 4-neighbor pixels in K
end for
Normalize K such that $\sum K_{i,j} = 1$
return K

Representative samples from QRData are shown in Figure 9, illustrating the diversity of blur severity and motion directions produced by the proposed non-uniform motion blur synthesis algorithm. In the data generation pipeline, the kernel size S is set adaptively according to L. The motion length L is sampled from $[20, 61]$ pixels, and the angle θ is sampled from $[0, \pi]$. The acceleration factor a is stochastically selected from $[0.3, 0.7]$ to generate diverse non-linear motion patterns. This strategy effectively bridges the domain gap between synthetic datasets and real-world QR code distortions captured by mobile devices.

 doi:10.20944/preprints202606.0184.v1

(a) Sharp QR code (b) L = 20, θ=0° (c) L = 40, θ=45° (d) L = 50, θ=90° (e) L = 60, θ=135°

Figure 9. Example QR codes from the proposed QRData dataset. (a) Sharp ground-truth QR code. (b)–(e) Synthesized non-uniform motion blurred QR codes with different motion parameters: (b) mild horizontal blur (L = 20, $\theta = 0^\circ$), (c) medium diagonal blur (L = 40, $\theta = 45^\circ$), (d) severe vertical blur (L = 50, $\theta = 90^\circ$), (e) severe opposite diagonal blur (L = 60, $\theta = 135^\circ$). The acceleration factor a<1 produces decelerating motion trajectories.

In addition to the constructed QRData dataset, we also employ the GOPRO dataset for training. The GOPRO dataset contains 2,103 training images and 1,111 test images of various scenes with motion blur, providing a rich source of realistic blurred images for pre-training our models before fine-tuning on the QRData dataset. This two-stage training strategy enables the models to learn general deblurring features from the GOPRO while adapting to the specific structural characteristics of QR code deblurring in QRData.

4.2. Metrics

Our models are trained using a two-stage transfer learning strategy: pre-training on the GoPro [21] dataset (2,103 training images, 1,111 test images), followed by fine-tuning on QRData (1,521 training images, 301 testing images). For performance evaluation, we construct a test set of 150 motion blurred QR codes, including 100 synthetic and 50 real-world captures under realistic conditions. The evaluation metrics include Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM), and Decoding Rate (DR), which is defined as the percentage of successfully decoded QR codes. Model complexity is evaluated using parameter count and FLOPs (for a 512 $\times$ 512 image), while inference efficiency is measured by the Average inference time (Avg_time) per image in seconds.

4.3. Implementation Details

EG-Restormer is configured with [4, 6, 6, 8] Transformer blocks and [1, 2, 4, 8] attention heads across four levels, using 48 base channels [11]. All models are trained using the AdamW optimizer (β_1=0.9, β_2=0.999, weight decay = 0.0001). The initial learning rate is set to $3e^{-4}$ and gradually reduced to $1e^{-6}$ via cosine annealing. EG-Restormer employs progressive training for 400K iterations with the following (patch size, batch size) pairs: [(128^2, 64), (160^2, 40), (192^2, 32), (256^2, 16), (320^2, 8), (384^2, 8)]. LENet is trained for 1,000K iterations with a fixed patch size of 256×256 and a batch size of 8. For data augmentation, we use flipping, rotation, and shuffling. For the decoder of ADNet, the Zbar library is used for all decoding attempts.

All experiments are conducted on a system running Ubuntu 20.04.6 LTS, with Python 3.8.2 and PyTorch 2.3.1. The hardware platform consists of an Intel(R) Xeon(R) Gold 6226R CPU @ 2.90 GHz and four NVIDIA RTX A6000 GPUs. To ensure a fair comparison, all models are implemented in PyTorch and trained under the same hardware configuration.

4.4. Results

We compare EG-Restormer with state-of-the-art image deblurring methods, including NAFNet-32 [15] and Restormer [11]. Quantitative results for two training strategies are reported in Tables 1 and 2.

Pre-training on GoPro only. We first evaluate the quantitative performance of models pre-trained exclusively on the GoPro dataset. EG-Restormer is compared with two state-of-the-art image deblurring methods, NAFNet-32 and Restormer, as shown in Table 1. Among all methods, EG-Restormer achieves the best performance across all three metrics: with a PSNR of 10.99 dB, which is

 doi:10.20944/preprints202606.0184.v1

0.39 dB higher than Restormer and 0.79 dB higher than NAFNet-32; an SSIM of 0.582, which is 0.086 higher than Restormer and 0.133 higher than NAFNet-32; and a DR of 58.67%, which is 8.67% higher than Restormer and 18.00% higher than NAFNet-32. These results demonstrate that, even without QR-specific training data, the explicit edge guidance mechanism of EG-Restormer effectively restores structural integrity under severe motion blured conditions, leading to improved decodability.

Table 1. Comparison of deblurring methods trained on GoPro (evaluated on QRData test set.)

Methods	**GoPro**		
	DR(%)	**PSNR**	**SSIM**
LENet (Ours)	32.67	9.78	0.421
NAFNet-32 [15]	40.67	10.20	0.449
Restormer [11]	50.00	10.60	0.496
EG-Restormer (Ours)	**58.67**	**10.99**	**0.582**

Fine-tuning on QRData. After fine-tuning on the synthetic QRData dataset, all methods exhibit substantial performance improvements, as shown in Table 2. Restormer achieves the best reconstruction fidelity, obtaining the highest PSNR of 18.15 dB and SSIM of 0.766. In contrast, EG-Restormer attains the highest decodability, with a DR of 90.00%, surpassing Restormer by 1.33% and NAFNet-32 by 8.67%. Notably, although EG-Restormer yields a lower PSNR of 14.88 dB than Restormer, it achieves superior decoding performance. This observation suggests that, for QR code deblurring, conventional pixel-wise reconstruction metrics are not fully aligned with downstream decoding accuracy. Instead, preserving structural characteristics critical to machine readability, such as edge sharpness, module boundary integrity, and pattern continuity, plays a more significant role in successful decoding. Since EG-Restormer explicitly incorporates edge-aware priors, it better maintains these structural cues under severe blur degradation.

Table 2. Comparison of deblurring methods trained on GoPro + QRData (evaluated on QRData test set).

Methods	**GoPro + QRData**		
	DR(%)	**PSNR**	**SSIM**
LENet (Ours)	49.33	11.37	0.572
NAFNet-32 [15]	81.33	15.56	0.695
Restormer [11]	88.67	**18.15**	**0.766**
EG-Restormer (Ours)	**90.00**	14.88	0.666

These results highlight the limitations of relying solely on fidelity-oriented metrics for evaluating QR code restoration and demonstrate the importance of integrating structural priors into deblurring frameworks. More broadly, they indicate that practical QR restoration systems should be optimized toward decoding-oriented objectives rather than only pixel-level similarity.

Qualitative analysis. To validate the effectiveness of our EG-Restormer, Figure 10 presents a qualitative comparison among EG-Restormer, Restormer, and NAFNet on severely motion-blurred QR codes. The input blur severely smears edges, disrupts the finder patterns, and makes individual modules difficult to recognize. After deblurring, NAFNet fails to recover fine structures and produces discontinuous stripes that cannot be recognized by the decoder. Restormer removes part of the blur but still leaves visible trailing artifacts and blurred boundaries; as a result, the restored QR code remains undecodable. In contrast, EG-Restormer reconstructs sharp and continuous edges, restores the integrity of all three finder patterns, and preserves clear module separations, producing a clean and successfully decodable QR code. This qualitative result confirms that explicit edge priors are critical for achieving practical decodability under severe non-uniform motion blur, whereas pixel-wise accuracy alone is insufficient.

 doi:10.20944/preprints202606.0184.v1

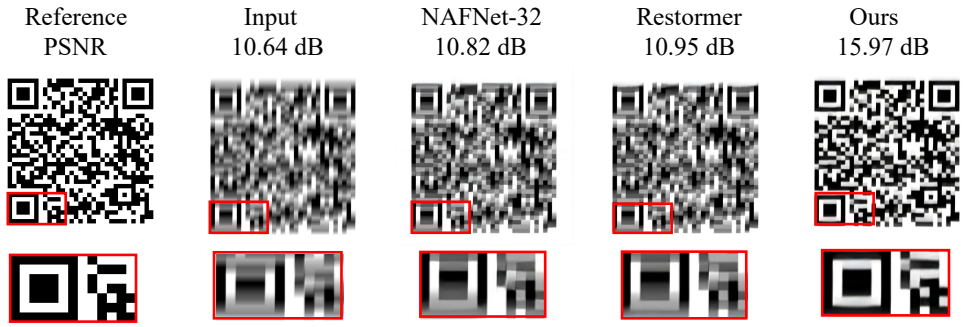


Figure 10. Qualitative comparison of QR code deblurring on the test set.

Complexity analysis. Table 3 reports the model complexity and runtime of the compared methods. LENet, our lightweight baseline, requires only 0.28 M parameters and 3.87 G FLOPs, with an average inference time of 0.23 s per image, which is substantially lower than those of the other methods. This makes LENet suitable for real-time QR code deblurring scenarios, although at the cost of lower decodability, with a DR of 49.33% after joint training, as shown in Table 2. Among the high-performance models, NAFNet-32 contains 17.11 M parameters and 64.44 G FLOPs, achieving an average inference time of 0.794 s. Restormer and our EG-Restormer have similar parameter counts (≈26.1 M) and FLOPs (≈565 G), with EG-Restormer introducing only a slight overhead (+0.15 G FLOPs and +0.072 s) due to its explicit edge guidance modules. Despite this modest increase in complexity, EG-Restormer achieves the highest decodability rate (90.00%), confirming that the additional computational cost is justified for QR code deblurring tasks, where decoding success is the primary objective.

Table 3. Comparison of model complexity and runtime for different deblurring methods.

Methods	Params.(M)	Flops(G)	Avg_time(s)
LENet (Ours)	**0.28**	**3.87**	**0.23**
NAFNet-32 [15]	17.11	64.44	0.794
Restormer [11]	26.13	564.96	0.838
EG-Restormer (Ours)	26.13	565.14	0.910

Ablation Study. As shown in Table 4, we conduct an ablation study on the proposed blur severity-based routing strategy by comparing it with three baseline settings: using EG-Restormer alone, using LENet alone, and randomly selecting a restoration branch (ADNet-Random). The results clearly demonstrate the effectiveness of the proposed routing strategy. ADNet achieves the same DR of 90% as EG-Restormer while reducing Avg_time by 19%, from 0.910s to 0.737s. Compared with random routing, ADNet improves DR by 18%. This results validates that the blur severity-based routing strategy effectively balances decoding performance and inference efficiency by intelligently assigning each input to the appropriate network.

Table 4. Ablation studies for blur severity-based routing unit.

Methods	Avg_time(s)	DR(%)
Restormer (baseline)	0.838	**88.67**
EG-Restormer	0.910	**90**
LENet	**0.28**	49.33
ADNet (random)	0.457	72
ADNet (ours)	0.737	**90**

 doi:10.20944/preprints202606.0184.v1

5. Discussion

The results show that incorporating domain-specific priors is particularly effective for QR code motion deblurring. Unlike general image restoration, the main goal of QR code deblurring is not only to improve visual quality, but also to recover the structural information required for successful decoding. Since QR codes contain regular spatial patterns and sharp edges, explicitly modeling edge priors is more suitable than relying solely on implicit feature learning. In this work, the proposed EGAB enhances the attention mechanism by injecting edge information, which helps the network focus on the high-frequency structures that are most critical to QR code readability. This explains why EG-Restormer achieves stronger decoding performance than general-purpose deblurring models.

Another important finding is that image quality metrics and decoding performance are not always fully consistent. Although PSNR and SSIM are standard indicators for restoration quality, they cannot completely reflect whether a restored QR code can be successfully recognized by a decoder. In practical applications, the decoding rate is a more task-oriented and meaningful metric. Our experiments show that EG-Restormer achieves the best decoding rate while maintaining competitive reconstruction quality, suggesting that explicit edge-aware restoration is better aligned with the final objective of QR code recognition.

The proposed ADNet further demonstrates the importance of balancing restoration accuracy and computational efficiency. In real-world mobile deployment, processing all inputs with a large model is inefficient, especially because mildly blurred QR codes may already be recoverable using a lightweight network. By using Laplacian variance as a blur indicator, ADNet routes easy samples to LENet and reserves EG-Restormer for more challenging cases. The ablation results confirm that this adaptive strategy preserves the decoding performance of the stronger model while reducing average inference time, showing its practical potential for real-time QR code deblurring scenarios.

Despite these advantages, several limitations remain. First, the routing threshold in the LVR module is manually calibrated, which may reduce robustness across different devices or imaging conditions. Second, the current dataset size is still limited. Third, the framework mainly focuses on motion blur, while real scanning scenarios may also include defocus blur, illumination variation, compression artifacts, perspective distortion, and occlusion. Future work could explore learnable routing strategies, more comprehensive multi-degradation restoration, and decoding-oriented training objectives. More broadly, the idea of combining explicit structural priors with adaptive computation may also be extended to other structured visual targets, such as text images and industrial markers.

6. Conclusions

In this paper, we proposed a domain-specific deblurring by directly integrating edge priors into a Transformer-based network. The proposed Edge-Guided Attention Block (EGAB) enables the network to focus on restoring the sharp edges that are crucial for QR code recognition. Furthermore, the adaptive dual-network framework, ADNet, intelligently combines a lightweight network with a more powerful restoration network, significantly improving inference efficiency without sacrificing performance. Comprehensive experiments demonstrate that the proposed method achieves state-of-the-art performance in QR code restoration, with superior decoding rates, competitive image quality metrics, and improved inference efficiency. This work provides a promising direction for adapting general vision models to specialized tasks with strong structural priors, potentially benefiting related tasks such as text image deblurring and face image deblurring. The proposed method has practical implications for improving the reliability of QR code scanning in real-time applications. Future research could explore more sophisticated routing strategies, larger and more diverse datasets, and extensions to other types of structured visual data.

Author Contributions: Conceptualization, J.L.,W.L and W.Z.; methodology, J.L.; software, J.L.; validation, J.L. and W.Z.; formal analysis, J.L.; investigation, J.L. and D.G; resources, W.Z.; data curation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, W.Z.; visualization, J.L.; supervision, W.Z.; project administration, W.Z.; funding acquisition, W.Z. All authors have read and agreed to the published version of the manuscript.

 doi:10.20944/preprints202606.0184.v1

Funding: This research was funded by Guandong Key Laboratory for New Generation IoT Research and Development,Project No.:2024KSYS009.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The original contributions presented in this study are included in the article. The source code and trained models for ADNet will be made publicly available upon acceptance at https://github.com/leejianping/ADNet.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wikipedia, QR Code. Available online: https://en.wikipedia.org/wiki/QR_code (accessed on 8 May 2026).
2. Zhang, K.; Ren, W.; Luo, W.; Lai, W.-S.; Stenger, B.; Yang, M.-H.; Li, H. Deep image deblurring: A survey. *arXiv* **2022**, arXiv:2201.10700.
3. Cho, S.; Lee, S. Fast motion deblurring. In Proceedings of the ACM SIGGRAPH Asia 2009 Papers, Yokohama, Japan, 16–19 December 2009; pp. 1–8.
4. Sörös, G.; Semmler, S.; Humair, L.; Hilliges, O. Fast blur removal for wearable QR code scanners. In Proceedings of the ACM International Symposium on Wearable Computers, Osaka, Japan, 7–11 September 2015; pp. 117–124.
5. Munoz-Mejias, D.; Gonzalez-Diaz, I.; Diaz-de-Maria, F. A low-complexity pre-processing system for restoring low-quality QR code images. *IEEE Trans. Consum. Electron.* **2011**, *57*, 1320–1328.
6. Alam, N.; Sagar, A.S.M.S.; Zhang, W.; Jin, T.; Dosset, A.; Dang, L.M.; Moon, H. A comprehensive study on enhanced QR extraction techniques with deep learning-based verification. *Appl. Intell.* **2025**, *55*, 676.
7. Van Gennip, Y.; Athavale, P.; Gilles, J.; Choksi, R. A regularization approach to blind deblurring and denoising of QR barcodes. *IEEE Trans. Image Process.* **2015**, *24*, 2864–2873.
8. Rioux, G.; Scarvelis, C.; Choksi, R.; Hoheisel, T.; Marechal, P. Blind deblurring of barcodes via Kullback–Leibler divergence. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 77–88.
9. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. A general U-shaped Transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17662–17672.
10. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8183–8192.
11. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H. Restormer: Efficient Transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 5718–5729.
12. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.-H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14821–14831.
13. Kong, L.; Dong, J.; Ge, J.; Li, M.; Pan, J. Efficient frequency domain-based Transformers for high-quality image deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023.
14. Cui, Y.; Tao, Y.; Bing, Z.; Ren, W.; Gao, X.; Cao, X.; Huang, K.; Knoll, A. Selective frequency network for image restoration. In Proceedings of the Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
15. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple baselines for image restoration. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 17–33.
16. Pu, H.; Fan, M.; Yang, J.; Lian, J. Quick response barcode deblurring via doubly convolutional neural network. *Multimed. Tools Appl.* **2019**, *78*, 897–912.
17. Li, J.; Zhang, D.; Zhou, M.; Cao, Z. A motion blur QR code identification algorithm based on feature extracting and improved adaptive thresholding. *Neurocomputing* **2022**, *493*, 351–361.
18. Dong, H.; Liu, H.; Li, M.; Ren, F.; Xie, F. An algorithm for the recognition of motion-blurred QR codes based on generative adversarial networks and attention mechanisms. *Int. J. Comput. Intell. Syst.* **2024**, *17*, 83.

 doi:10.20944/preprints202606.0184.v1

19. Wang, B.; Xu, J.; Zhang, J.; Li, G.; Wang, X. Motion deblur of QR code based on generative adversative network. In Proceedings of the 2nd International Conference on Algorithms, Computing and Artificial Intelligence, Sanya, China, 20–22 December 2019; pp. 166–170.
20. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
21. Nah, S.; Kim, T.H.; Lee, K.M. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3883–3891.
22. Fergus, R.; Singh, B.; Hertzmann, A.; Roweis, S.T.; Freeman, W.T. Removing camera shake from a single photograph. In Proceedings of the SIGGRAPH Conference, Boston, MA, USA, 30 July–3 August 2006; pp. 787–794.
23. Xu, L.; Jia, J. Two-phase kernel estimation for robust motion deblurring. In Proceedings of the European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010; pp. 157–170.
24. Krishnan, D.; Tay, T.; Fergus, R. Blind deconvolution using a normalized sparsity measure. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 233–240.
25. Pan, J.; Sun, D.; Pfister, H.; Yang, M.-H. Deblurring images via dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2315–2328.
26. Lv, X.-G.; Liu, J.; Li, F.; Yao, X.-L. Blind motion deconvolution for binary images. *J. Comput. Appl. Math.* **2021**, *393*, 113500.
27. Chen, Y.; Su, D.; Zheng, J. An effective document image deblurring algorithm. In Proceedings of the International Conference on Document Analysis and Recognition, Beijing, China, 18–21 September 2011; pp. 885–889.
28. Pan, J.; Hu, Z.; Su, Z.; Yang, M.-H. Deblurring text images via L_0-regularized intensity and gradient prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2901–2908.
29. Pan, J.; Hu, Z.; Su, Z.; Yang, M.-H. L_0-regularized intensity and gradient prior for deblurring text images and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 342–355.
30. Jiang, X.; Yao, H.; Zhao, S. Text image deblurring via two-tone prior. *Neurocomputing* **2017**, *242*, 1–14.
31. Turin, W.; Boie, R.A. Bar code recovery via the EM algorithm. *IEEE Trans. Signal Process.* **1998**, *46*, 354–363.
32. Kresić-Jurić, S.; Madej, P.; Santosa, F. Applications of hidden Markov models in bar code decoding. *Pattern Recognit. Lett.* **2006**, *27*, 1665–1672.
33. Choksi, R.; Van Gennip, Y.; Oberman, A.M. Anisotropic total variation regularized L_1-approximation and denoising/deblurring of 2D bar codes. *Inverse Probl. Imaging* **2010**, *4*, 11–33.
34. Zheng, X.; Li, H.; Zhou, Y. Blind deblurring of QR code using intensity and gradient prior of positioning patterns. *Multimed. Tools Appl.* **2024**, *83*, 14873–14895.
35. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations, Virtual, 3–7 May 2021.
36. Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Gao, W. Pre-trained image processing Transformer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12299–12310.
37. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image restoration using Swin Transformer. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Montreal, QC, Canada, 11–17 October 2021; pp. 1833–1844.