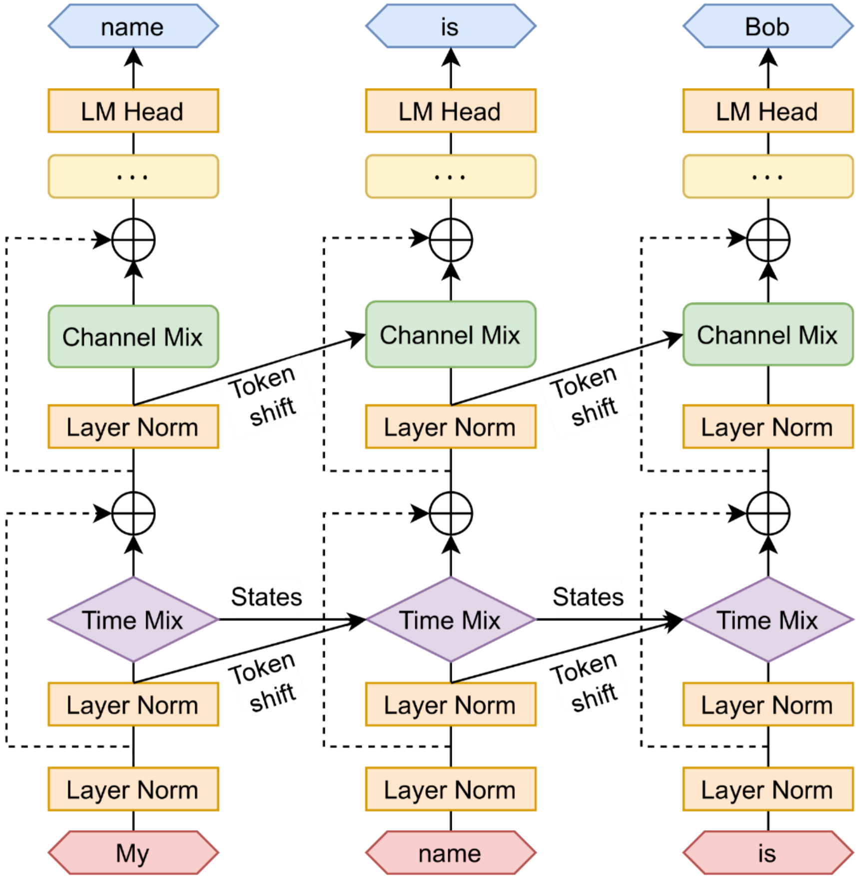
DRWKV: Focusing on Object Edges for Low-Light Image Enhancement

Xuecheng Bai[1*], Yuxiang Wang[2*], Boyu Hu[3], Qinyuan Jie[1], Chuanzhi Xu[2], Kechen Li[4], Hongru Xiao[5], Vera Chung[2]

[1] Shenyang Ligong University, Shenyang, China
[2] The University of Sydney, NSW, Australia
[3] University of International Business and Economics, Beijing, China
[4] Nanjing University of Aeronautics and Astronautics, Nanjing, China
[5] Tongji University, Shanghai, China
*Equal contribution
bai_xuecheng@163.com, ywan0192@uni.sydney.edu.au

RWKV Model Introduction



Schematic Diagram of the RWKV Model

➢ **The limitations of RNNs and Transformers**

⭐ RNN has the **vanishing gradient** problem.

⭐ RNN cannot perform **parallel training**, limiting its scalability.

⭐ Transformer has **quadratic time and space complexity**, resulting in high computational costs and memory usage.

➢ **What is the RWKV Model?**

RWKV is a RNN-based large language model that dispenses with the self-attention mechanism. Through its innovative architectural design, it achieves **a synergy of parallelized training and recurrent inference**:

⭐ **Training phase:** Drawing on the parallel computing paradigm of Transformers

⭐ **Inference phase:** Adopting RNN-style recurrent state propagation, it only needs to maintain the current hidden state without relying on KV caching, resulting in constant memory usage that is independent of context length.

⭐ **Core innovation:** Replacing traditional Softmax self-attention with **WKV computation**, it reduces the **computational complexity from O(n2) (as in Transformers) to O(n) linear scale**, leading to a 10–100× improvement in inference speed while supporting "unbounded" linear scaling of context length.

RWKV Model Introduction

➢ **RWKV Name Explanation (Comparison with Transformer QKV)**

⭑ R (Receptance): **A gating mechanism**, dynamically controls the degree of acceptance and forgetting of historical information.

⭑ W (Weight): **A trainable temporal attenuation parameter**.

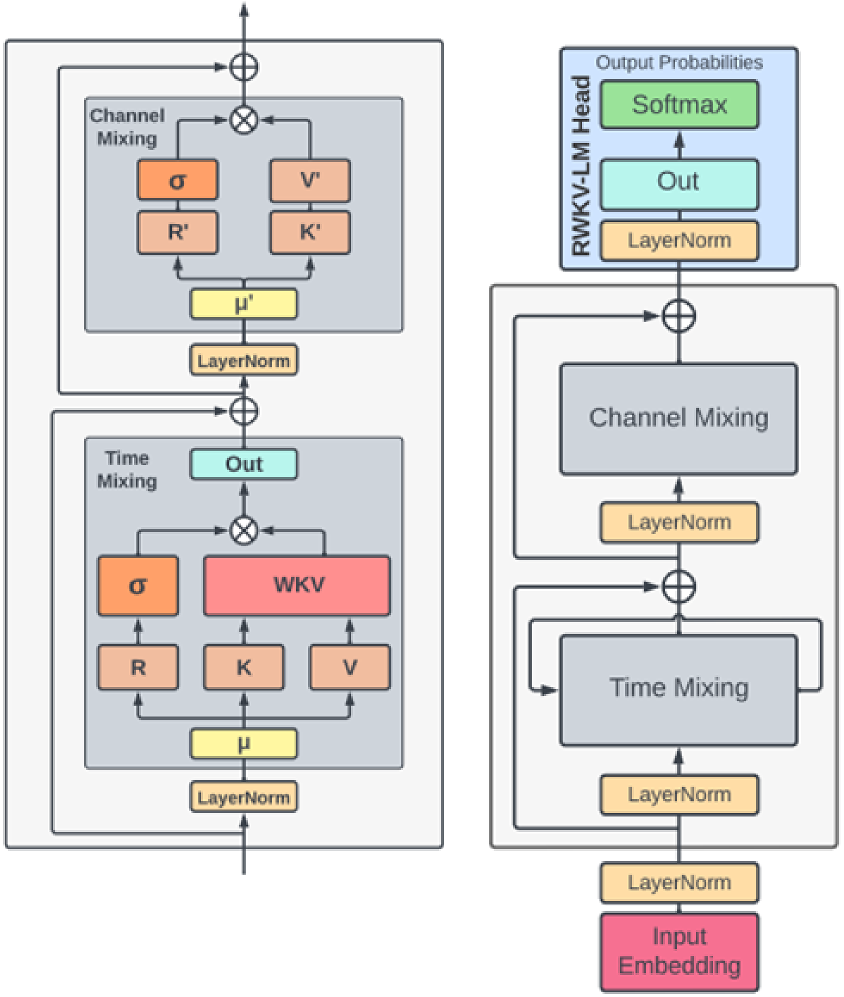
⭑ K (Key) & V (Value): The key and value vectors, which **serve a similar role to K/V in Transformer** by storing feature information of historical tokens.

➢ **Architecture of RWKV**

The model is composed of multiple stacked residual blocks, each containing two core submodules: **Time Mixing and Channel Mixing**

⭑ Time Mixing：Modeling temporal dependencies via **token shift**, which **fuses the current token's features with those from the previous 1–2 positions**. Combined with **W's positional decay and R's gating**, it dynamically retains or forgets history

⭑ Channel Mixing：For cross-channel interaction, a **linear projection** followed by a **GeGLU** variant dynamically controls feature flow, boosting efficiency and expressiveness

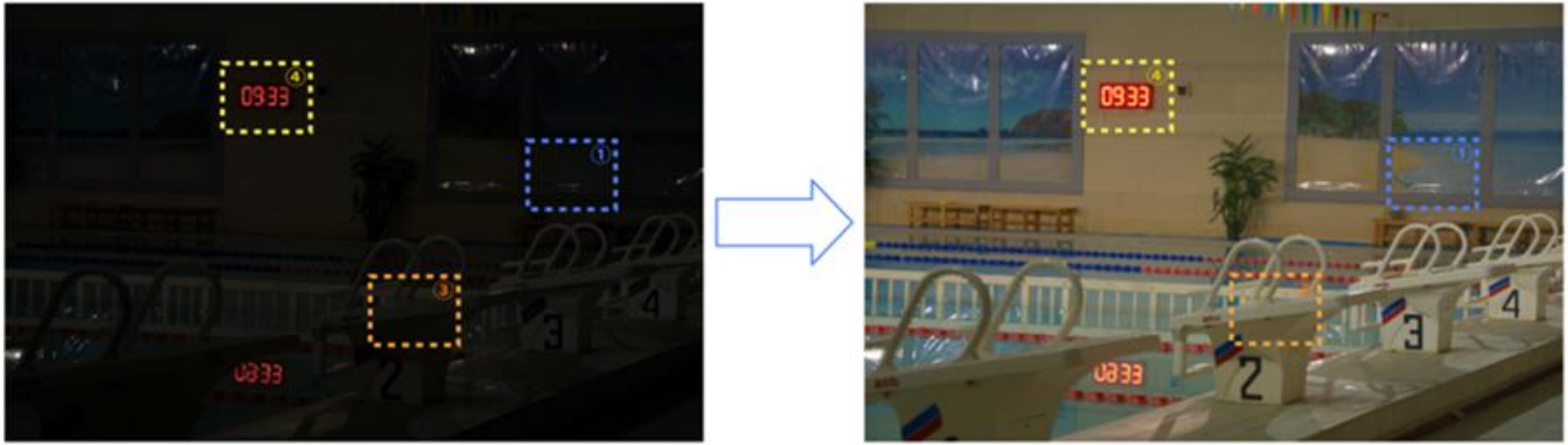


Schematic Diagram of the RWKV Model

Core pain points in low-light imaging

The main pain points in low-light imaging are concentrated in the following **five areas**:

1. **Insufficient brightness**, with an overall dark and blurry appearance;
2. **Severe image noise**, interfering with the recognition of effective information;
3. **Loss of critical details**, making it difficult to discern core content;
4. **Color representation distortion**, deviating from the true scene tones;
5. Poor contrast, with **indistinct differentiation between light and dark levels**.



Visualization of Core Pain Points

Introduction to Innovations

- **Innovation 1: Global Edge Retinex Theory**

Based on the Retinex model, it **decouples global illumination and edge structures** for refined global/local illumination restoration, addressing noise amplification, edge distortion, and insufficient decoupling in traditional low-light methods.

- **Innovation 2: Evolutionary Scanning Mechanism**

Adapted to the **curved Riemannian manifold distribution of low-light edges**, it solves inaccurate Euclidean space modeling and poor edge continuity capture of traditional methods.

- **Innovation 3: Bilateral Spectrum Alignment Block**

It optimizes **luminance and chrominance**, resolving overexposure, color distortion, and insufficient matching between edge details and spectral features caused by luminance-chrominance imbalance.

- **Innovation 4: Loss Function**

Constraining the model from five dimensions (**decomposition consistency, edge sparsity, illumination smoothness, artifact suppression, parameter regularization**), it solves false edges, uneven illumination, residual artifacts, and poor generalization in low-light image enhancement.

Global Edge Retinex Theory

Dilemmas of Traditional Retinex

Decomposing the formula $I = R \cdot L + N$ reveals the following issues：

Indistinction between artifacts and noise：

Halation, streaking and other artifacts induced by enhancement under low illumination are often conflated with sensor noise, leading to a lack of targeted design for denoising and artifact suppression.

Coupling of illumination and edges：

The absence of independent edge constraints tends to impair the edge details and structural integrity during the illumination adjustment process

Core Idea of GER

Five-Dimensional Decoupling:

Edge features (E) and artifacts (S) are explicitly decoupled, enabling independent control over illumination, structure and noise.

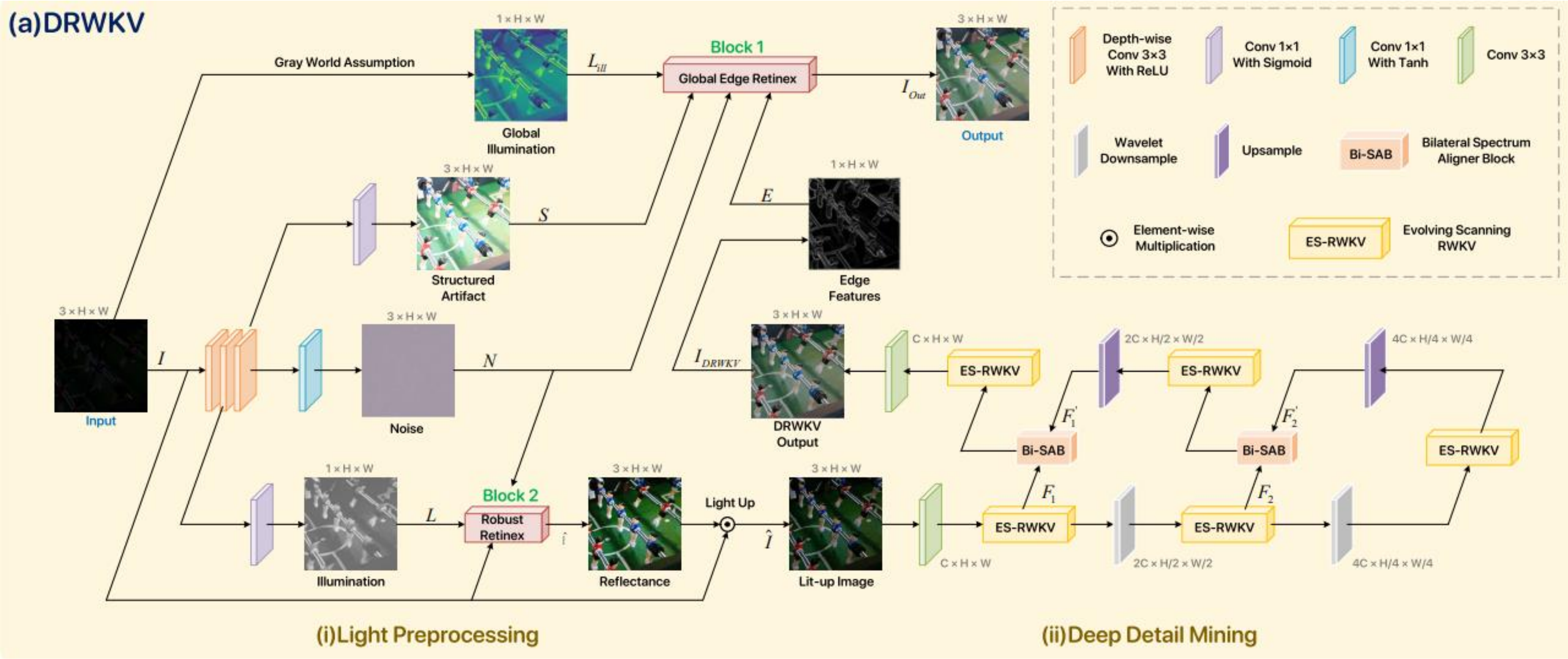
GER Mathematical Formulation

$$I = (R + \alpha \cdot E) \odot L_{ill} + \beta \cdot N + \gamma \cdot S$$

$$R(x, y) = \frac{I(x, y) - N(x, y)}{L_{ill}(x, y) + \epsilon}$$

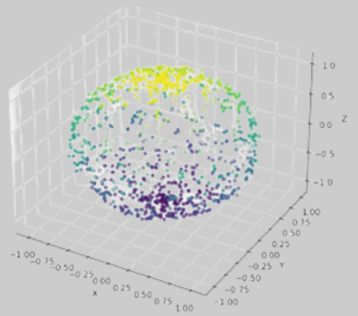
$\alpha \cdot E$	$\beta \cdot N$	$\gamma \cdot S$
Explicit Edge Prior (Extracted via Scharr)	**Noise Decoupling Term** (Independently Modeled)	**Artifact Suppression Term** (Artifacts)

Global Edge Retinex Theory



Evolutionary Scanning Mechanism

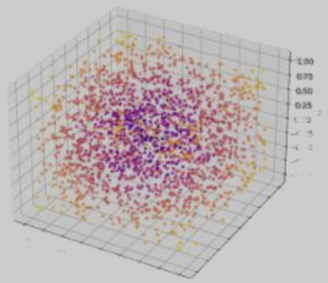
The Problem



Euclidean

Traditional CNNs adopt fixed grids and sliding windows
Result: Disrupts continuous edges

The Solution



Riemannian

Low-light edges exhibit a curved manifold structure
Requirement: A scanning mechanism that "flows" along the geometric shape is needed

Inspiration

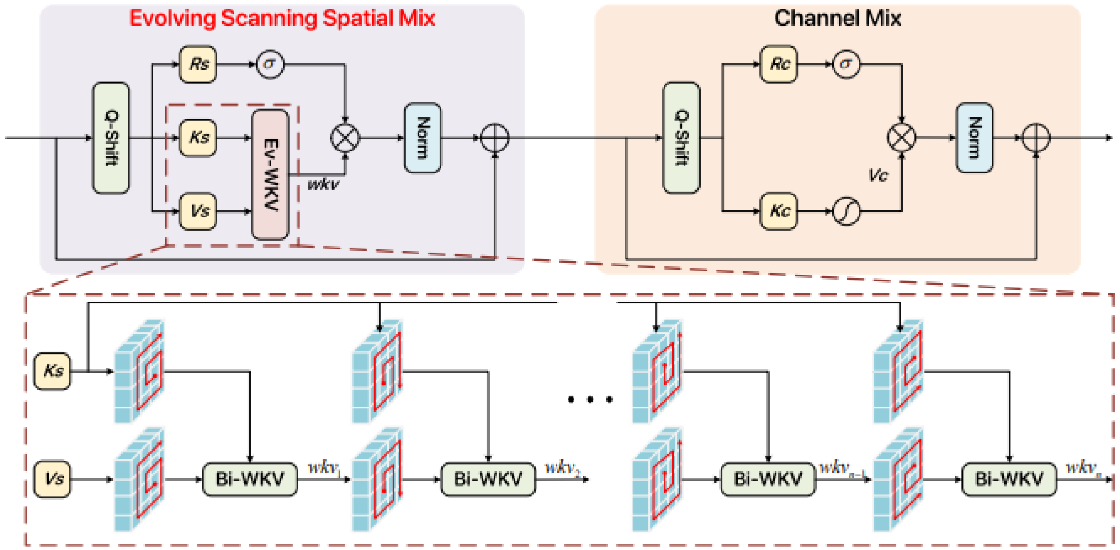
Drawing on the **Archimedean Spiral**, the discontinuity in the 2D space is transformed into the continuity of a 1D time series.

$$r(\theta) = a + b\theta$$

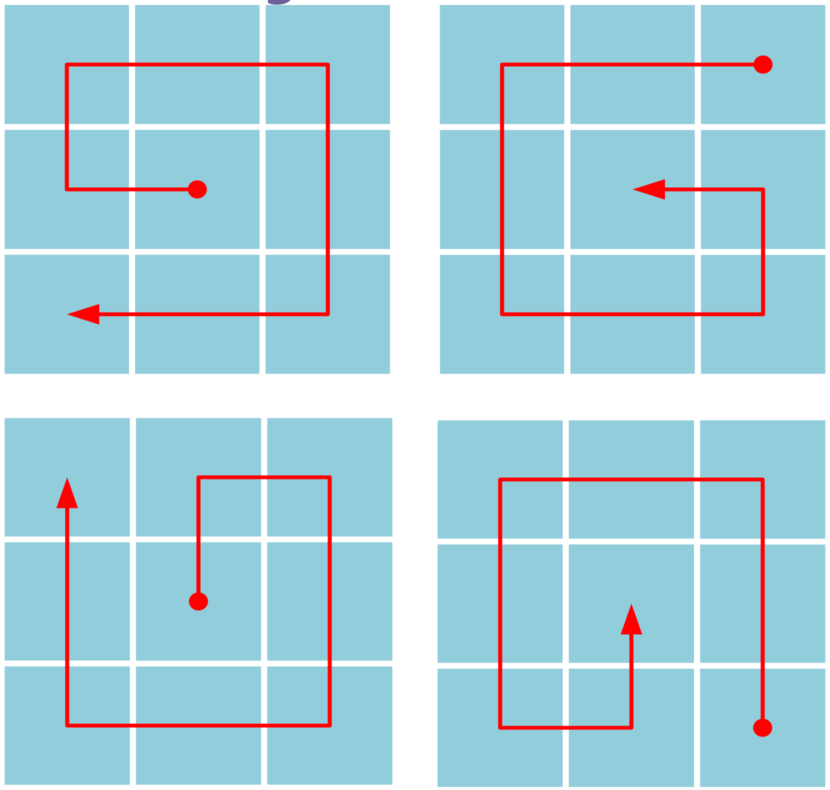
Evolutionary Scanning Mechanism

How it works

- **Four-Directional Cyclic Shifting:** Cyclic shifting in the sequence of Up → Right → Down → Left to capture omni-directional dependencies.
- **Centripetal to Centrifugal Scanning:** Scanning radiates outward from the center of the feature map to the surrounding areas.



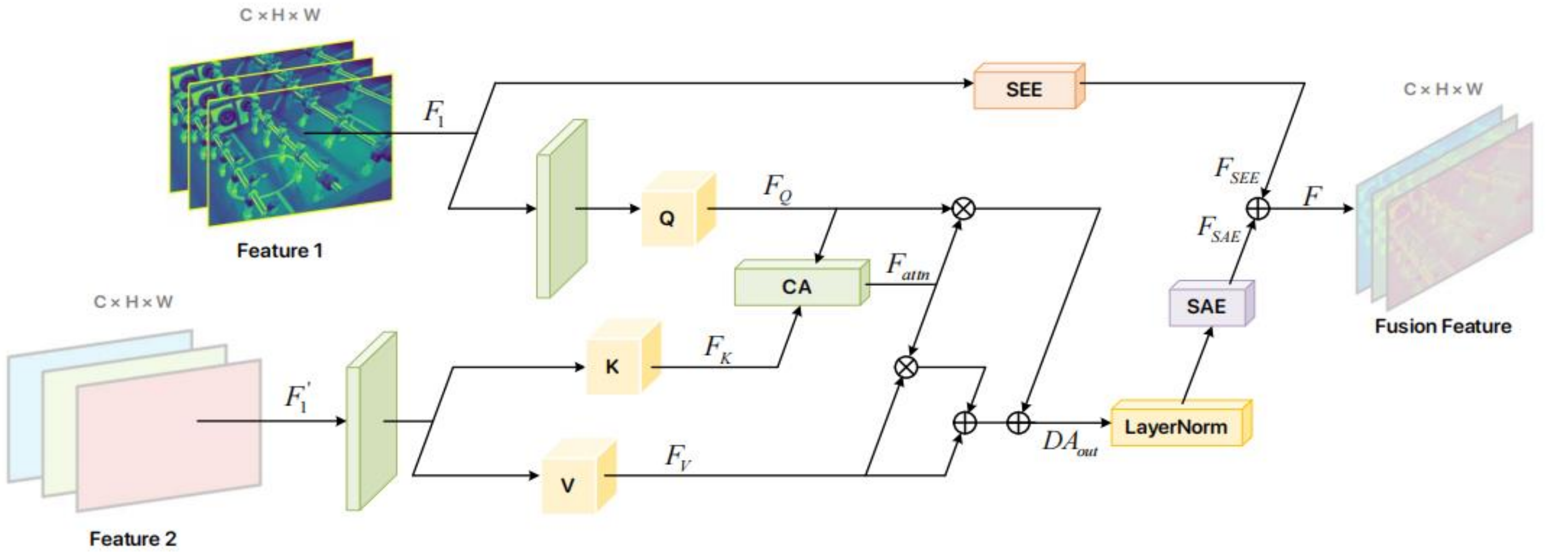
Scanning Architecture



Why is this approach valid?

This topological structure preserves the **spatial continuity of object edges** during the process of gradual variation to the maximum extent and avoids the fragmentation caused by sliding windows.

Bilateral Spectrum Alignment Block



The core idea is to enable the encoder's luminance features and ES-RWKV's detail features to **"perceive" and fuse with each other**:

1. **Input Preparation & QKV Projection:** Receive low-level luminance features (encoder) and high-level detail features (ES-RWKV).

2. **Cross-Attention Feature Alignment:** L2-normalize Q/K, calculate cross-covariance attention matrix, weight-fuse value features, and align luminance-chrominance spatial features.

3. **Spectrum-Edge Dual-Branch Enhancement:** Optimize low-illumination spectral details via deep convolution and residual connections; extract bidirectional XY gradients with Scharr operator to strengthen edge textures.

4. **Feature Fusion & Output:** Element-wise add optimized spectral and edge features, output enhanced features with balanced luminance, accurate colors and clear edges.

Algorithm 1 Bi-SAB: Bilateral Spectrum Aligner Block

Input: Low-level Feature F_1: (B, C_i, H_i, W_i)
High-level Feature F_1': (B, C_i, H_i, W_i)
Output: Bi-SAB Fusion Feature F: (B, C_i, H_i, W_i)

/*NOTE: Technical details are provided in the supplementary material.*/

/*Compute the query (Q), key (K), and value (V) representations.*/
$F_Q : (B, C_i, H_i, W_i) \leftarrow \text{Conv}_{3\times3}(F_1)$
$F_K : (B, C_i, H_i, W_i) \leftarrow \text{Conv}_{3\times3}(F_1')$
$F_V : (B, C_1, H_i, W_i) \leftarrow \text{Conv}_{3\times3}(F_1')$
/*Compute Cross Attention (CA) feature representations.*/
$F_{attn} : (B, C_1, H_i, W_i) \leftarrow CA(F_Q, F_K)$
/*Compute Feature Difference Adjustment (FDA) feature representations.*/
$Att_{out} : (B, C_1, H_i, W_i) \leftarrow F_{attn} \times F_K$
$F_{out} : (B, C_1, H_i, W_i) \leftarrow \rho \cdot Att_{out} + F_V, \rho = 0.2$
$FDA : (B, C_1, H_i, W_i) \leftarrow F_{attn} \times F_Q + F_{out}$
/*Compute Scharr Edge Enhancement (SEE) feature representations.*/
$F_{SEE} : (B, C_i, H_i, W_i) \leftarrow SEE(F_1)$
/*Compute Spectral Alignment Enhancer (SAE) feature representations.*/
$F_{SAE} : (B, C_1, H_i, W_i) \leftarrow SAE(LN(FDA))$
/*Compute Bi-SAB feature representations.*/
$F : (B, C_1, H_i, W_i) \leftarrow F_{SAE} + F_{SEE}$
return F

Loss Function

Total Objective

$$L_{total} = \lambda_1 L_{recon} + \lambda_2 L_{sparse} + \lambda_3 L_{smooth} + \lambda_4 L_{artifact} + \lambda_5 L_{reg}$$

Reconstruction Loss

$$||I - \hat{I}||_1$$

Basic pixel constraint
Ensures overall structural consistency of the image

Edge Sparsity Loss

$$||E||_1$$

Suppresses artifacts from edge enhancement
Preserves true edge structure

Illumination Smoothness Loss

$$|\nabla L| \cdot \exp(-\lambda|\nabla I|)||$$

Ensures natural illumination transition
Avoids local abrupt changes

Halo Suppression Loss

$$||S||_1 + \delta \cdot TV(S)$$

Minimizes halo energy / spatial regularization
Removes Halo and Block artifacts

Parameter Regularization

$$\alpha^2 + \beta^2 + \gamma^2$$

Constraints GER parameter range
Prevents overfitting, improves generalization

Loss Function

(a)w/o L_{recon} (b)w/o L_{sparse} (c)w/o L_{smooth}

(d)w/o $L_{artifact}$ (e)w/o L_{reg} DRWKV

Comparative Experiment of GER

Table 5. The impact of different α values on the performance of the LOLv2-Real dataset.

α	0.05	0.10	0.15	**0.20**	0.25	0.30	0.35
PSNR↑	23.56	23.89	24.02	**24.12**	23.84	23.97	23.94
SSIM↑	0.798	0.815	0.814	**0.832**	0.792	0.821	0.814
NIQE↓	4.012	3.967	3.941	**3.926**	3.951	3.989	4.173

Table 6. The impact of different β values on the performance of the LOLv2-Real dataset.

β	-0.100	-0.075	**-0.050**	-0.025	0
PSNR↑	23.72	23.98	**24.12**	23.86	23.61
SSIM↑	0.803	0.821	**0.832**	0.817	0.804
NIQE↓	3.995	3.948	**3.926**	3.973	4.003

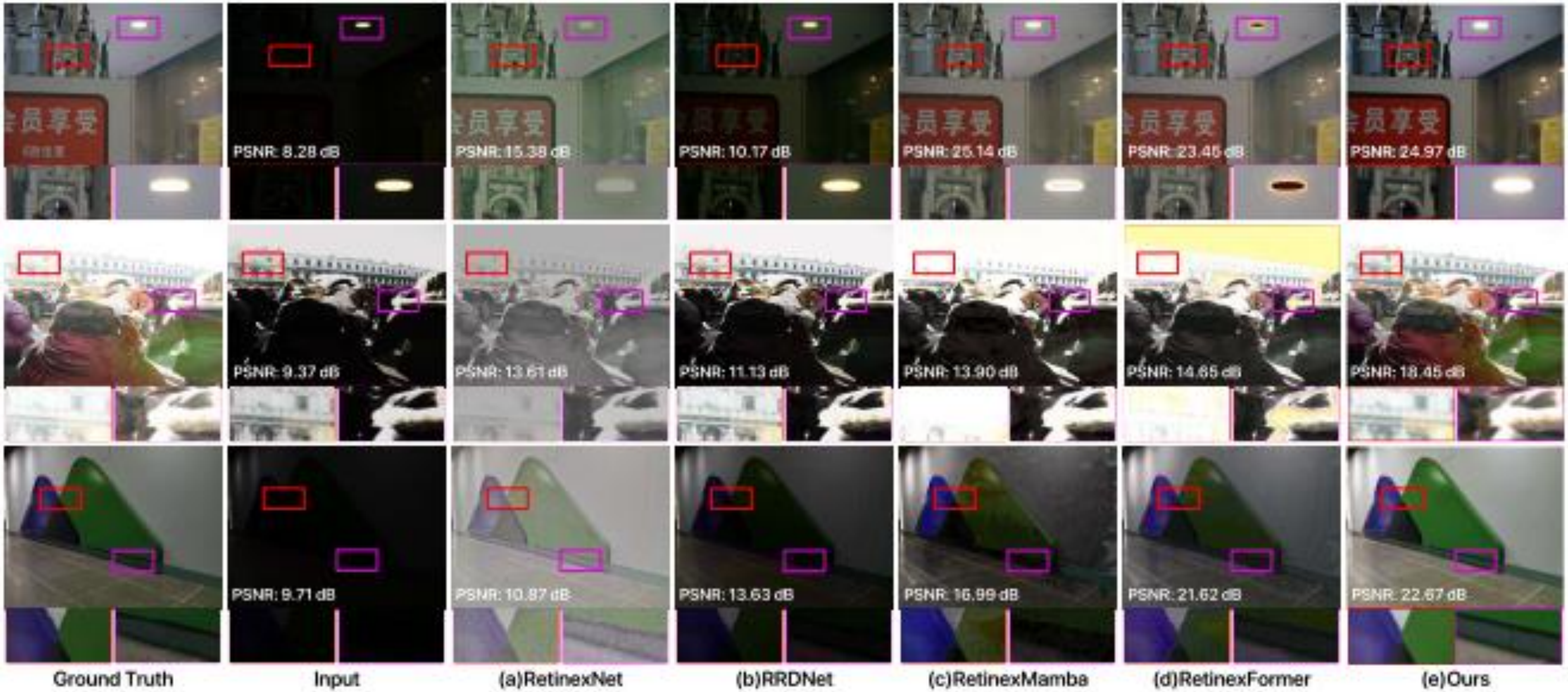
Table 7. The impact of different γ values on the performance of the LOLv2-Real dataset.

γ	0.05	0.10	0.15	**0.20**	0.25	0.30	0.35
PSNR↑	22.74	23.65	23.91	**24.12**	23.99	23.78	23.84
SSIM↑	0.804	0.801	0.823	**0.832**	0.826	0.812	0.814
NIQE↓	4.112	4.023	3.957	**3.926**	3.962	3.998	4.472

Comparison of different Block combinations on the LOLv2 Real benchmark dataset.

Config	**Block1**	**Block2**	**PSNR↑**	**SSIM↑**	**NIQE↓**
(I)	Robust Retinex	Robust Retinex	22.57	0.741	4.124
Ours	Robust Retinex	GER	**24.12**	**0.832**	**3.926**

Comparative Experiment



Comparative Experiment

Method	Params	GFLOPs	LSRW-Huawei			LOLv2_Real			LOLv2_Syn			SDSD_indoor			SDSD_outdoor		
			PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓
RetinexNet [46]	0.8	587.47	15.61	0.414	7.235	16.10	0.407	9.425	17.14	0.756	5.405	20.84	0.617	7.924	20.96	0.629	9.947
KinD [53]	8.0	34.99	15.77	0.548	5.517	18.67	0.772	9.221	15.22	0.542	6.124	20.98	0.597	7.221	21.65	0.621	8.714
RRDNet [55]	0.1	2.10	14.66	0.541	7.725	15.21	0.514	9.912	16.67	0.667	4.894	21.07	0.601	6.921	20.04	0.647	8.997
Zero-DCE [12]	0.7	5.21	14.86	0.559	4.215	14.12	0.512	8.652	14.93	0.531	5.507	21.23	0.752	6.941	20.39	0.691	8.014
SCI [27]	0.09	0.06	14.78	0.526	3.667	17.30	0.541	8.077	14.96	0.721	4.899	20.43	0.714	6.574	21.33	0.701	7.778
LLFormer [42]	13.2	22.52	16.23	0.642	3.947	20.56	0.801	9.162	24.42	0.914	4.614	25.66	0.832	6.627	28.45	0.821	8.011
UHDFormer [39]	0.3	48.37	21.07	0.604	4.621	21.59	0.804	7.230	22.60	0.903	5.476	28.13	0.875	7.021	22.75	0.732	6.657
Retinexformer [4]	1.6	15.51	22.24	0.701	2.976	21.65	0.835	4.735	25.10	0.925	3.971	28.96	0.879	5.441	28.96	0.896	7.201
IAT [8]	0.09	1.44	20.12	0.694	4.217	20.30	0.752	5.232	22.96	0.856	5.512	19.97	0.713	4.177	19.97	0.711	6.417
WalMaFa [37]	39.7	14.41	21.04	0.698	3.112	22.49	0.851	4.365	25.10	0.945	4.275	29.67	0.915	3.541	28.94	0.891	3.955
MambaIR [13]	20.4	60.66	21.14	0.704	3.004	20.11	0.802	4.928	24.75	0.922	4.522	25.11	0.873	3.661	26.35	0.510	4.477
RetinexMamba [2]	4.6	34.75	20.88	0.629	3.104	22.34	0.826	4.771	24.71	0.932	4.503	28.21	0.893	3.976	26.22	0.866	4.011
MambaLLIE [47]	4.4	20.85	20.64	0.627	3.047	22.14	0.821	4.021	24.81	0.940	4.071	29.74	0.902	4.417	28.92	0.869	4.417
DRWKV (Ours)	9.7	1.67	22.34	0.703	2.944	24.12	0.832	3.926	25.02	0.947	3.941	30.26	0.922	3.441	28.96	0.891	3.954

Low-Light Object Tracking

Without low-light enhancement

With DRWKV low-light enhancement

Thank You
for Your Attention

Q&A

DRWKV: WACV 2026 Accepted Paper