Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Response-decomposed residual filtering for joint low-light enhancement and deblurring

Minhui Kang

zebukang@hotmail.com

Jiangxi Electric Vocational and Technical College

Peizhe Long

Shandong Agricultural University

Depei Cheng

Shandong Agricultural University

Yiyao Zhang

Jiangxi Electric Vocational and Technical College

Jiayun Zheng

Jiangxi Electric Vocational and Technical College

Yuxin Zhu

Shanghai Normal University

Research Article

Keywords: Low-light restoration, Joint enhancement and deblurring, Dynamic filtering, Response decomposition, Paired evaluation

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11620707/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Response-decomposed residual filtering for joint low-light enhancement and deblurring

Minhui Kang[1*†], Peizhe Long[2†], Depei Cheng[2], Yiyao Zhang[1], Jiayun Zheng[1], Yuxin Zhu[3]

[1*]School of Power Engineering, Jiangxi Electric Vocational and Technical College, Nanchang, China.

[2]School of Information Science and Engineering, Shandong Agricultural University, Tai'an, China.

[3]College of Life Sciences, Shanghai Normal University, Shanghai, China.

*Corresponding author(s). E-mail(s): zebukang@hotmail.com;

†Minhui Kang and Peizhe Long contributed equally to this work.

Abstract

Joint low-light enhancement and deblurring require local corrections that restore visibility and detail together. We propose response-decomposed residual filtering at LEDNet's native dynamic-kernel interfaces. The original generator is retained, while an added route combines a signed kernel-sum increment with four learned zero-sum shape bases. Encoding features predict the sum increment; the same features, augmented by the degraded red–green–blue (RGB) image, predict the shape coefficients. The route operates at three scales and starts from zero output before joint training. Paired fixed-budget comparisons examine the complete model, conditioning source, kernel-sum branch and carrier choice. A prespecified model trained for 60,000 updates is evaluated on 1,800 official-test pairs from 30 videos. Relative to paired LEDNet, the structural similarity index (SSIM) increases by 0.004656 and learned perceptual image patch similarity (LPIPS) decreases by 0.011445, with improvements in 26/30 and 30/30 videos. The peak signal-to-noise ratio (PSNR) difference is **−0.045441** dB; PSNR improvement is not established. The route adds 226,880 parameters and approximately 17.7% measured forward latency. These results establish a within-LEDNet SSIM/LPIPS benefit with additional cost; public DarkIR-m remains a stronger system reference under a different training protocol.

Keywords: Low-light restoration, Joint enhancement and deblurring, Dynamic filtering, Response decomposition, Paired evaluation

1 Introduction

Low-light blurred images require visibility and spatial detail to be recovered together. Increasing intensity reveals scene content but does not reconstruct displaced edges; detail recovery must also accommodate the information available in a dark input. LEDNet addresses this joint task with an enhancement encoder, a deblurring decoder and spatially adaptive feature filtering [1]. Because it already predicts local kernels, the question here is how to organize a supplementary correction while preserving the original generator's learned aggregation capability.

Low-light restoration systems differ in how they combine enhancement and detail recovery. DMPHNet organizes enhancement and noise reduction through decomposition, multiscale restoration and illumination adjustment [2], whereas joint-task systems use efficient convolutional blocks (DarkIR) [3], multi-state modeling (URWKV) [4], or Fourier-domain decoupling (FDN) [5]. NAFNet provides a general convolutional restoration backbone [6]. LEDNet's native dynamic-filtering interfaces offer a specific setting for studying how a supplementary correction should be represented and conditioned. We therefore distinguish paired design comparisons within LEDNet from comparisons between complete restoration systems.

Dynamic Filter Networks and Kernel Prediction Networks learn input-dependent filtering [7, 8]. BPN predicts burst-dependent spatiotemporal bases to filter input frames [9]; KBNet mixes shared learned bases for local feature aggregation [10]; AHFNet weights prescribed high-pass bases for video deblurring [11]. Our model, denoted LEDNet-P3 (P3), instead adds a signed sum coordinate and a learned centered shape expansion to an existing, unrestricted kernel. The distinction is its native residual placement, sum/shape representation and conditional routing; the backbone and added route are jointly trained.

An unconstrained tap increment leaves its responses to neighborhood means and local variation implicit. We make the added constant-neighborhood response explicit through one signed kernel-sum coordinate and restrict the remaining correction to four learned zero-sum bases. The restriction applies to the increment, while the original generator remains trainable. Encoding features predict the sum; a degraded red–green–blue (RGB) projection supplements those features only for shape coefficients. This routing supplies direct image information for spatial modulation while retaining a feature-only sum path.

We introduce a structured residual correction at LEDNet's native dynamic-filtering interfaces. The design combines an explicit kernel-sum coordinate, a restricted zero-sum shape expansion and feature/RGB conditioning while retaining the original generator. Function-preserving initialization makes the added route compatible with the unmodified backbone at the start of training. Paired comparisons examine the complete design and its conditioning, branch and carrier choices. A prespecified official-test evaluation quantifies the resulting SSIM/LPIPS gains and computational cost; the scope of these findings is discussed alongside the results.

2 Response-decomposed residual restoration model

P3 combines a single-scale kernel correction with multiscale integration and joint training.

2.1 Overall architecture and native filtering interfaces

Figure 1 distinguishes feature propagation, kernel prediction and RGB conditioning. The enhancement encoder, pyramid pooling modules (PPM), curve nonlinear units (CurveNLU), bottleneck, decoder and side outputs are retained. The encoder widths are $[32, 64, 128, 128]$. Corrections operate at the three native filtering resolutions: 1/8, 1/4 and 1/2, with 128, 128 and 64 channels, respectively. The decoder connects these resolutions and produces the restored RGB image; the complete configuration is given in Online Resource 1, Section S8.1.

At scale s, the enhanced encoder feature f_s predicts the original kernel $k_s^0 = G_s^0(f_s)$. The decoder provides the feature map u_s to which that kernel is applied. For channel c and location p, let $u_{s,c,p} \in \mathbb{R}^{25}$ denote the corresponding 5×5 neighborhood. The proposed route changes the kernel argument of the existing operator:

$$k_{s,c,p} = k^0_{s,c,p} + \Delta k_{s,c,p}. \tag{1}$$

The original generator is neither frozen nor replaced. At each scale, G_s^0 uses three 3×3 Conv–ReLU pairs and a final 1×1 convolution producing $25C_s$ kernel coefficients per location. The native aggregation and downstream decoder remain in place. The native feature-filtering operator (denoted FAC in the figures) retains replicate padding and LeakyReLU with slope 0.2. Thus the modification is a feature-kernel increment, not an image-domain post-processing CNN or an additional output skip connection.

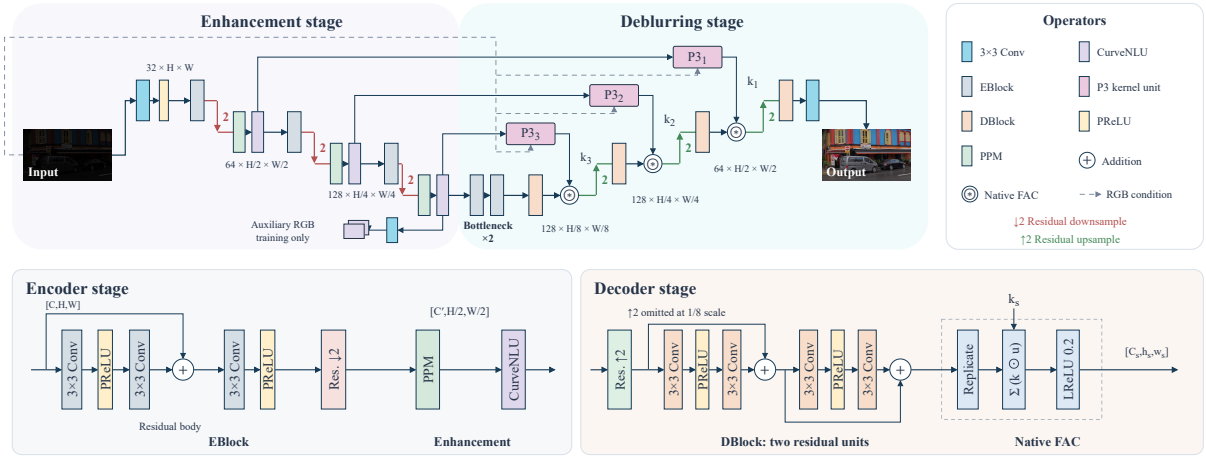


Fig. 1 Multiscale LEDNet-P3 architecture and native encoder/decoder operations. PPM and CurveNLU are retained. At each scale, the P3 kernel unit preserves the original generator and adds the correction detailed in Fig. 2; its output enters native FAC, rather than an encoder–decoder feature addition. Dashed paths carry RGB conditioning. Sampling arrows denote the residual sampling within the corresponding blocks. The bottom panels expand the native residual units and FAC; the auxiliary RGB branch is used for training only.

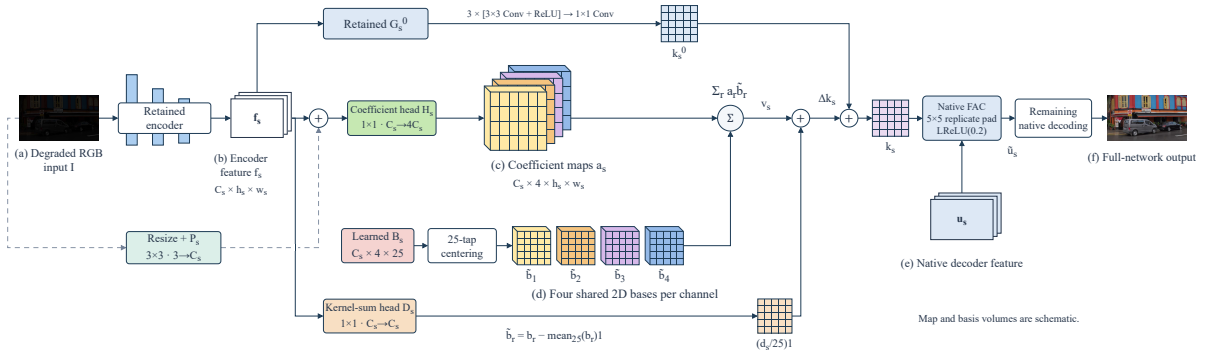


Fig. 2 Response-decomposed correction at one native filtering scale. (a,b) The input and retained encoder supply the RGB condition and feature. (c,d) Four position-varying coefficient maps weight four shared, tap-centered 2D bases per channel. The shape term and feature-only sum contribution form the increment, which is added to the original kernel. (e) Native FAC filters the decoder feature with one inherited activation; subsequent decoding produces (f). Internal map/basis volumes are schematic. Input/output images are saved final-evaluation predictions for case 0164/0039.

2.2 Response-decomposed kernel correction

Figure 2 expands the complete added route at one scale. Let C_s be the channel count and $h_s \times w_s$ the feature resolution. The operators D_s, P_s and H_s denote the sum head, RGB projection and coefficient head, respectively. For brevity, scalar/vector equations suppress scale, channel and position indices where unambiguous.

Feature-conditioned kernel-sum response. A 1×1 convolution maps C_s input channels to C_s outputs, providing one signed increment per channel and location:

$$d_s = D_s(f_s), \qquad \Delta k^{\mathrm{sum}} = \frac{d}{25}\mathbf{1}. \tag{2}$$

No positivity activation or unit-sum normalization is imposed. With $\bar{u} = \mathbf{1}^\top u/25$, this component adds $d\bar{u}$ to the local pre-activation filtering response. Its role is a constant-neighborhood response coordinate, not a directly measured illumination variable.

RGB-assisted shape coefficients. The degraded RGB image is resized bilinearly to the feature resolution, with `align_corners=false`, to form z_s. A 3×3 projection maps its three channels to C_s. Element-wise addition, not concatenation,

combines this projection with the full encoding feature:

$$a_s = H_s(f_s + P_s(z_s)). \tag{3}$$

The 1×1 coefficient head maps C_s channels to $4C_s$, interpreted as four coefficients for each channel and position. RGB is supplied directly to this coefficient route only; the sum head in Eq. 2 reads f_s.

Centered bases and spatial mixing. For each scale and channel, four learned 5×5 tap vectors are shared over spatial positions. They are centered across the 25 taps before being mixed:

$$\begin{aligned} \widetilde{b}_{s,c,r} &= b_{s,c,r} - \frac{\mathbf{1}^\top b_{s,c,r}}{25}\mathbf{1}, \\ v_{s,c,p} &= \sum_{r=1}^{4} a_{s,c,r,p}\widetilde{b}_{s,c,r}. \end{aligned} \tag{4}$$

Each position-wise coefficient weights a complete basis vector before the four vectors are summed. Consequently $\mathbf{1}^\top v = 0$. Four centered bases span at most four dimensions in the 24-dimensional zero-sum tap space; linear independence is not imposed. These are logical dimensions, not a required kernel materialization.

Kernel synthesis and native application. The two components are added to the original kernel, rather than applied sequentially:

$$\begin{aligned} \Delta k = \frac{d}{25}\mathbf{1} + v, \qquad \mathbf{1}^\top \Delta k = d, \\ \Delta k^\top u = d\bar{u} + v^\top(u - \bar{u}\mathbf{1}). \end{aligned} \tag{5}$$

For a constant neighborhood $u = \alpha\mathbf{1}$, the centered component vanishes and the added response is $d\alpha$; otherwise, the second term responds to local variation. These identities describe the increment before the single inherited activation, leaving the original kernel unrestricted. Online Resource 1 (Section S1, Fig. S1) illustrates the basis expansion and equivalent local application.

2.3 Multiscale integration and joint training

Each filtering scale has its own sum head, RGB projection, coefficient head and bases. The three interfaces therefore use

$$(s, C_s) \in \{(1/8, 128), (1/4, 128), (1/2, 64)\}.$$

At each interface, the projection output and f_s have the same channel count and spatial dimensions. The basis layout is logically $C_s \times 4 \times 25$; the coefficient field has four values per channel at each spatial position. The same construction is repeated at all three scales without sharing its learned parameters across scales. Together these additions contain 226,880 parameters.

Function-preserving initialization. The weights and biases of D_s and H_s are initialized to zero, while basis entries have nonzero random initialization with standard deviation 0.02. Thus $d_s = a_s = 0$ and $\Delta k_s = 0$ initially, preserving the unmodified backbone's function at the same initialization. The backbone, heads, projection and bases are then trained jointly from scratch.

Inherited training objective. The correction introduces no new target or loss. Using $\widehat{I}$ for the final restored image and $I^\star$ for its paired target, the existing objective is

$$\begin{aligned} \mathcal{L}_{\text{train}} = \mathcal{L}_1(\widehat{I}, I^\star) + 0.01\mathcal{L}_{\text{VGG19}}(\widehat{I}, I^\star) \\ + 0.8\mathcal{L}_{\text{side}}. \end{aligned} \tag{6}$$

The inherited 3×3 RGB side head produces $\widehat{I}_a$ from the enhanced 1/8 encoding feature. Its target T_a is $I^\star$ bicubically resized to the side-output dimensions, with `align_corners=false`. The side term is $\mathcal{L}_{\text{side}} = \mathcal{L}_1(\widehat{I}_a, T_a) + 0.01\mathcal{L}_{\text{VGG19}}(\widehat{I}_a, T_a)$; the factor 0.8 therefore weights both auxiliary losses. This prediction is not an input to either P3 head. All variants retain the same objective and optimization settings (Section 3.1); the perceptual-loss settings are detailed in Online Resource 1, Section S8.1.

3 Experimental results and analysis

3.1 Dataset and training configuration

Data roles. LOL-Blur provides paired low-light blurred inputs and sharp normally exposed targets [1]. Its 170-video training portion (10,200 pairs) is split by source into 153 fitting videos (9,180 pairs) and 17 development videos (1,020 pairs). The development set has been used for exploration; the same split is retained across training runs. The official test comprises 1,800 pairs from 30 source-disjoint videos, with 60 frames per

video. The inherited loss in Eq. 6 is evaluated on training pairs only.

Paired fixed-budget training. Structural variants train from scratch for 60,000 updates using 256×256 patches, batch one and Adam with initial learning rate 10^{-4}, betas $(0.9, 0.99)$ and cosine decay to 10^{-7}. Shared initial parameters and actual sample, crop and augmentation streams are paired within each comparison; added parameters are initialized identically in the feature-conditioned control and LEDNet-P3. All endpoints are fixed before training, without development-score selection. The primary run supplies the final paired models. An independently initialized repeat run and the branch and carrier studies each retain their own paired controls. Further training details appear in Online Resource 1, Section S8.1.

Final model selection and public-system references. The final paired LEDNet and LEDNet-P3 models were fixed at 60,000 updates before official-test access; neither the repeat run nor later design comparisons replaced them. These two models and the public DarkIR-m checkpoint formed the prespecified PSNR/SSIM/LPIPS evaluation, with one complete evaluation per model. The released paper LEDNet was evaluated later as an additional public-system reference and did not influence model selection. Public DarkIR-m and LEDNet use the standard 170-video training portion, with training histories distinct from the paired models. DarkIR-m also differs in crop, batch and optimization recipe; checkpoint-specific pretraining and full training-compute records remain unavailable (Online Resource 1, Sections S8.1 and S8.3).

Quality measurements. Original 640×1120 inputs use whole-image FP32 inference without tiling, test-time augmentation or exposure correction. Each model retains its native normalization, padding and crop-back. Outputs are clipped, mapped and rounded to RGB8 before common RGB PSNR, RGB SSIM and LPIPS-VGG evaluation [12]. LPIPS uses package 0.1.4, model version 0.1, CPU execution and inputs mapped to $[-1, 1]$. Video means are averaged equally, with 60 frames per video. The primary official contrast is LEDNet-P3–LEDNet; public-system contrasts are interpreted separately. Online Resource 1, Sections S2 and S8, specifies the summaries and evaluation definitions. The numerical records are retained by the authors.

Post-hoc natural image quality evaluator (NIQE) [13] and lightness-order error (LOE) [14] descriptors use separate fixed-checkpoint replays on the same pairs. $\mathrm{LOE}_{\mathrm{low}}$ references the degraded input and $\mathrm{LOE}_{\mathrm{high}}$ the sharp target. Their definitions and scope are given in Online Resource 1, Section S4.

3.2 Analysis of results

Official-test quality is followed by separate development comparisons, fixed visual examples and measured cost.

3.2.1 Official-test comparison on LOL-Blur

Under the common evaluator (Table 1), LEDNet-P3 increases SSIM by +0.004656 and reduces LPIPS by 0.011445 relative to paired LEDNet. The changes are favorable in 26/30 videos for SSIM and 30/30 for LPIPS. These results support an improvement in structural similarity and learned perceptual distance for the prespecified model under the paired training protocol. Public LEDNet and DarkIR-m are separate system references with different training histories. The released paper LEDNet uses native RGB range $[0, 1]$; its post-hoc result does not replace the paired baseline. Starred rows report published context rather than a common-protocol ranking.

The mean PSNR difference is −0.045441 dB, so a PSNR improvement is not established. PSNR changes are favorable in 16 videos and unfavorable in 14; these counts describe direction rather than magnitude. The saved video differences range from −2.150967 to +2.420956 dB (Online Resource 1, Fig. S2). Thus the near-zero overall difference does not imply negligible changes in every video or equivalence between the models.

DarkIR-m achieves better means on all three primary metrics: LEDNet-P3–DarkIR-m is −2.923944 dB PSNR, −0.095290 SSIM and +0.106041 LPIPS. Different training histories limit architecture-only attribution. DarkIR-m's published LPIPS (0.162) differs from our 0.202576. The protocol comparison in Online Resource 1, Section S8.3, identifies range, quantization and device-path differences but does not quantify their effects or fully explain the discrepancy; the common-evaluator result is retained.

Table 1 Quantitative comparison on LOL-Blur. Published reports and our fixed-model evaluation.

Method	PSNR↑	SSIM↑	LPIPS↓	NIQE↓	LOE_{low} ↓	LOE_{high} ↓
KinD++* [15]	21.2600	0.7530	0.3590	/	/	/
DRBN* [16]	21.7800	0.7680	0.3250	/	/	/
DeblurGAN-v2* [17]	22.3000	0.7450	0.3560	/	/	/
MIMO-UNet* [18]	22.4100	0.8350	0.2620	/	/	/
NAFNet* [6]	25.3600	0.8820	0.1580	/	/	/
RetinexFormer* [19]	26.0200	0.8870	0.1810	/	/	/
LEDNet (paired)	**24.1204**	0.7823	0.3201	**7.3130**	280.62	**209.02**
LEDNet (public) [1]	26.0592	0.8476	0.2172	5.5721	238.23	112.60
DarkIR-m [3]	26.9989	0.8822	0.2026	5.5371	239.01	86.10
LEDNet-P3 (ours)	24.0750	**0.7869**	**0.3086**	7.3687	**279.43**	210.40

* Published scores: DarkIR (2025), Table 1, including earlier LOL-Blur retraining reports; trailing zeros align display precision. A slash (/) denotes an unreported metric; no NIQE/LOE comparison is made for these literature rows. The exact NAFNet variant/checkpoint remains unconfirmed. Unstarred rows: our 1,800-pair evaluations. Paired-model/DarkIR NIQE/LOE use post-hoc replays; public LEDNet uses one saved-output set for all six metrics. PSNR is in dB. Bold compares paired LEDNet/LEDNet-P3 only. Protocols differ, particularly LPIPS; no cross-source superiority is claimed.

The post-hoc descriptors are mixed for LEDNet-P3 relative to LEDNet: LOE_{low} decreases, whereas NIQE and LOE_{high} increase. These results do not establish a broader restoration advantage.

3.2.2 Controlled model and design comparisons

Table 2 places configurations and metrics in three separate training blocks. The condition-source control replaces RGB by means of three nearly equal feature-channel partitions, preserving the full feature path and parameter count. The branch control and carrier control each use a newly trained full-response model, distinct from the primary training run; these panels are not a factorial experiment.

Complete model and conditioning source. Table 2(a) compares LEDNet-P3 with LEDNet and the feature-summary control; Table 3 reports the two fixed-endpoint training runs separately. LEDNet-P3 improves SSIM and LPIPS over both references in each run, whereas PSNR differences are unfavorable in the primary run and favorable in the repeat run. The feature-conditioned control also improves SSIM/LPIPS over LEDNet in both runs. Thus the results support RGB assistance relative to this feature-summary alternative for SSIM and LPIPS, without attributing the complete benefit to RGB alone.

Neither run meets the original PSNR screening references of +0.15 dB against LEDNet and +0.10 dB against the feature-conditioned control; later favorable metrics do not replace those criteria. Complete means and all within-run contrasts remain in Online Resource 1, Table S3.

Kernel-sum branch. Table 2(b) compares a newly trained full-response model with Shape-only, which sets $d = 0$ throughout training and evaluation. The full model is better by +0.142813 dB PSNR, +0.002436 SSIM and −0.001636 LPIPS. This supports retaining the sum branch in the tested design. Active added parameters differ (226,880 versus 189,696), so the intervention includes the branch's capacity and is not a capacity-matched mechanism test.

Carrier choice. Table 2(c) replaces $d\mathbf{1}/25$ by de_0, with e_0 the center-tap unit vector, using a separate paired uniform control. All three metrics worsen. The outcome supports retaining the uniform carrier under this protocol. Both carriers sum to d, but the change adds $d(e_0 - \mathbf{1}/25)$, a fixed zero-sum direction: it changes the spatial response as well as its parameterization. It does not establish a universal carrier rule. Complete design comparisons and earlier 20k component observations are reported in Online Resource 1, Tables S4 and S5.

3.2.3 Visual comparison of restored images

Figure 3 shows the three official-test cases selected before inference by a fixed identity-based rule. The input, three restored outputs and target share the same display geometry and fixed crop. Case

Table 2 Configuration and saved development quality in three separate paired studies, trained for 60,000 updates, 1,020 pairs per model. Controls are trained within panels; full-response rows are different checkpoints, not a pooled factorial ablation.

Variant	Kernel sum	Zero-sum shape	Auxiliary condition	Carrier	PSNR (dB)↑	SSIM↑	LPIPS↓
(a) Primary run: original model and conditioning source							
LEDNet	×	×	—	—	22.49	0.7837	0.3036
Feature-cond.	✓	✓	Feature summary	Uniform	22.44	0.7845	0.2999
LEDNet-P3	✓	✓	RGB	Uniform	22.42	0.7878	0.2935
(b) Branch control: full response vs Shape-only							
Full (branch control)	✓	✓	RGB	Uniform	22.45	0.7879	0.2950
Shape-only ($d = 0$)	×	✓	RGB	n/a	22.31	0.7854	0.2967
(c) Carrier control: Uniform vs Center							
Uniform (carrier control)	✓	✓	RGB	Uniform	22.51	0.7885	0.2940
Center	✓	✓	RGB	Center	22.31	0.7869	0.2964

✓/× mark *added* branches; all rows retain the native generator. The auxiliary condition feeds the shape-coefficient route. Shape-only has 189,696 active added parameters versus 226,880 for full variants; its carrier is inactive at $d = 0$. Function capacity is not matched, and gains are interpreted within panels only.

identities and display conventions are detailed in Online Resource 1, Sections S2–S3.

The shop sign, awning and vehicle/window boundaries in Fig. 3(a) show similar overall visibility in LEDNet and LEDNet-P3, with remaining edge softness. In Fig. 3(b), sign and doorway structures remain blurred in both models, while DarkIR-m separates more boundaries. The bright lights, signage and dark surfaces in Fig. 3(c) likewise show small within-LEDNet differences and residual blur relative to the target. These fixed examples contextualize the complete-set results without establishing human preference or branch-specific mechanisms. Online Resource 1 (Section S3, Fig. S3) examines the same crops using signed local RGB error differences.

Table 3 Paired development differences at fixed-60k: LEDNet-P3 minus the indicated reference, on the same 1,020 pairs per round. Training runs are reported separately. Complete three-model means and the other saved contrasts appear in Online Resource 1.

Run	Reference	ΔPSNR (dB)	ΔSSIM	ΔLPIPS
Primary	LEDNet	−.073826	+.004190	−.010120
	Feature-cond.	−.017445	+.003305	−.006420
Repeat	LEDNet	+.053448	+.005049	−.006485
	Feature-cond.	+.043488	+.001605	−.001311

3.2.4 Computational cost and scope of the improvement

LEDNet-P3/LEDNet have 7.636/7.409 M parameters and synchronized forward means of 352.98/299.88 ms/image, respectively (+17.7% for P3); DarkIR-m has 3.322 M parameters and 167.00 ms/image (Table 4). The observed SSIM/LPIPS benefit therefore comes with additional parameters and latency. The design provides a specific native-interface correction: it exposes the increment's kernel sum, restricts its shape component and preserves the original generator's aggregation capability.

The eight-channel LEDNet operator uses synchronous CPU intermediate storage, whose internal transfers are included in synchronized forward time. Input transfer, output conversion and CPU metrics are excluded. DarkIR-m uses its native path. Hardware policies, latency distributions and training-cost measurements are detailed in Online Resource 1, Section S7. These measurements characterize the evaluated implementations. The 60,000-update, batch-one, 153-video contract defines a controlled paired study, rather than the performance ceiling of sufficiently trained LEDNet or a general deployment advantage.

Attribution and applicability. The primary comparison quantifies observed differences between two prespecified paired models under the

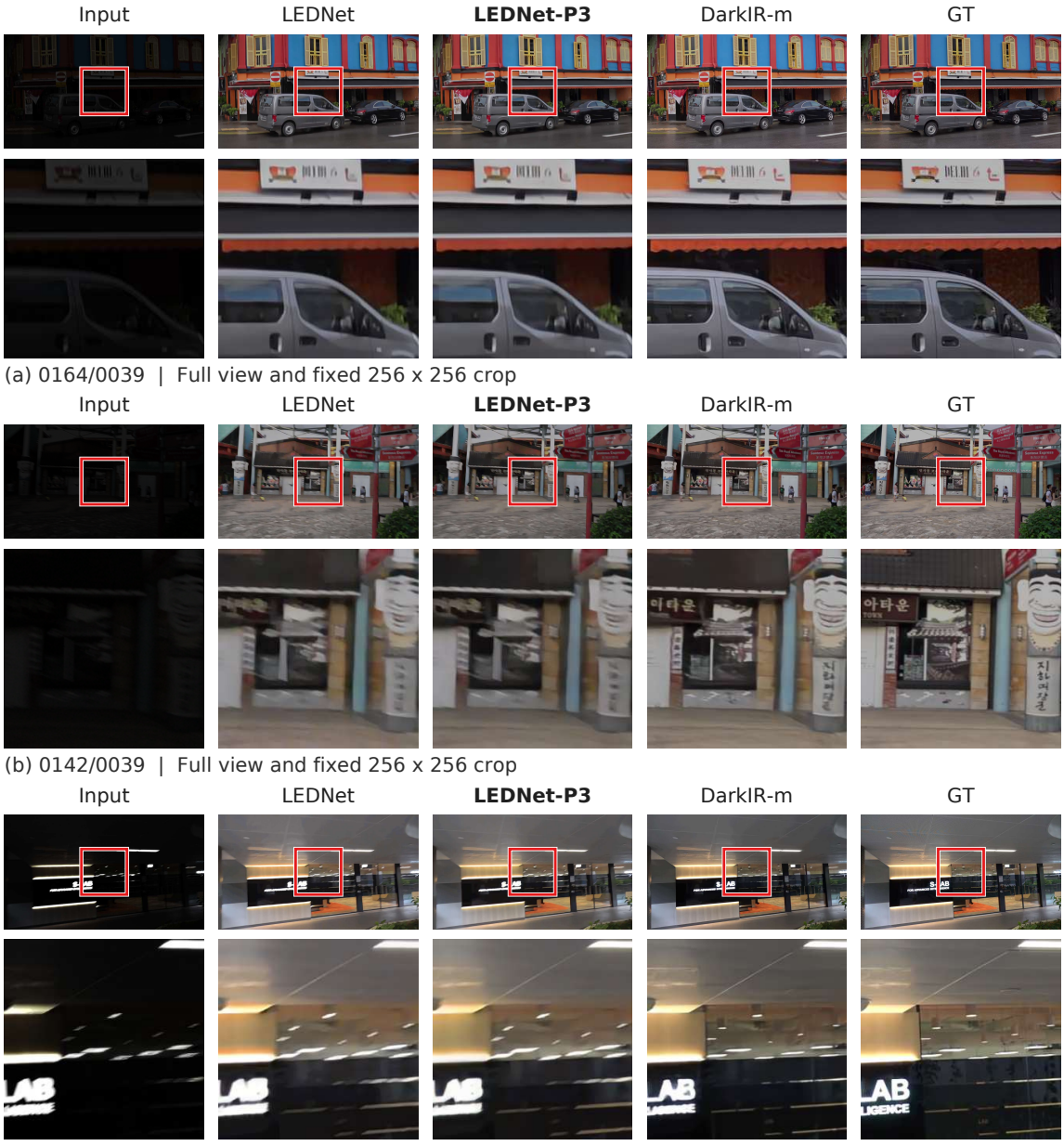


Fig. 3 Prespecified official-test predictions for (a) 0164/0039, (b) 0142/0039 and (c) 0215/0039. Each case shows the full image and the common crop at $[432, 192, 688, 448]$ in the original 640×1120 image. Red boxes are display-only overlays marking the crop; the underlying saved images are unchanged. All columns use the same display geometry.

stated protocol. It supports the complete LEDNet intervention but does not isolate response decomposition from a parameter-comparable ordinary dynamic-kernel residual at the same interface. Branch removal also changes active capacity; conditioning and carrier comparisons address their respective choices. The zero-sum constraint does not alone define a strict high-pass response, and the response algebra does not identify physical illumination and blur components. Fixed-model video differences do not establish stability across independent training runs or a necessary perception–distortion tradeoff.

Table 4 Measured inference cost on RTX 5060, 640 × 1120, batch 1, FP32. Synchronized forward means exclude the first five calls. Memory is PyTorch allocated/reserved peak MiB, not total GPU occupancy. M denotes 10^6 parameters. Hardware, timing scope and latency distributions are detailed in Online Resource 1, Section S7.

Model	Params (M)	Forward mean (ms/img)	GPU peak (MiB) allocated / reserved
LEDNet	7.409	299.88	6512.65 / 6518
LEDNet-P3	7.636	352.98	6513.52 / 6520
DarkIR-m	3.322	167.00	1085.44 / 1182

In the separate DarkIR adaptation, response-decomposed filtering shows no established advantage over ordinary-residual filtering; its SSIM and LPIPS results are unfavorable relative to native DarkIR (Online Resource 1, Tables S6–S7). This feature-residual interface differs from LEDNet's native kernel interface and neither supplies the missing direct control nor establishes universal failure. CNN post-training also shows no established additional multiscale-fusion benefit (Online Resource 1, Table S8). Both explorations retain separate identities, leave the final paired models unchanged and do not use the exposed official test for design selection.

4 Conclusions

We presented response-decomposed residual filtering at LEDNet's native dynamic-kernel interfaces. An explicit signed kernel-sum coordinate, a restricted zero-sum shape expansion and feature/RGB routing organize the added correction while retaining the original generator and decoder. Zero-output initialization preserves the backbone function at the start of joint training. Under the fixed training protocol, the prespecified model improves SSIM and LPIPS over paired LEDNet, and separate paired comparisons support the retained design choices. This benefit does not extend to an established PSNR improvement and comes with added parameters and latency; public DarkIR-m remains stronger. The findings support a bounded within-LEDNet contribution.

Statements and Declarations

Funding. This work was supported by the Science and Technology Research Project of the Jiangxi Provincial Department of Education, China (Grant No. GJJ2406408).

Competing interests. The authors declare no competing interests.

Author contributions.
Minhui Kang: Conceptualization, Methodology, Investigation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.
Peizhe Long: Conceptualization, Methodology, Software, Data curation, Investigation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing.
Depei Cheng: Methodology, Software, Investigation, Formal analysis, Validation, Writing – review & editing.
Yiyao Zhang: Data curation, Formal analysis, Writing – original draft, Writing – review & editing.
Jiayun Zheng: Validation, Writing – original draft, Writing – review & editing.
Yuxin Zhu: Writing – review & editing.

Data and code availability. LOL-Blur was obtained from its original authors [1]; the authors confirm permission for research use and publication of the example images. Numerical records are retained by the authors. The P3 code, configurations and final checkpoint are available on Zenodo at https://doi.org/10.5281/zenodo.23918535. Online Resource 1 describes the experiments and extended results.

Ethics and consent. Not applicable: this study did not involve human participants, identifiable personal data or animal experiments.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process. During the preparation of this work, the authors used ChatGPT and OpenAI Codex (OpenAI) for language editing and LaTeX typesetting. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

[1] Zhou, S., Li, C., Loy, C.C.: LEDNet: Joint low-light enhancement and deblurring in the dark. In: European Conference on Computer Vision (2022). https://shangchenzhou.com/projects/LEDNet/

[2] He, M., Wang, R., Wang, Y., Zhou, F., Guo, N.: DMPH-Net: a deep multi-scale pyramid hybrid network for low-light image enhancement with attention mechanism and noise reduction. Signal, Image and Video Processing 17, 4533–4542 (2023). https://doi.org/10.1007/s11760-023-02687-9

[3] Feijoo, D., Benito, J.C., Garcia, A., Conde, M.V.: DarkIR: Robust low-light image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10879–10889 (2025).

[4] Xu, R., Niu, Y., Li, Y., Xu, H., Liu, W., Chen, Y.: URWKV: Unified RWKV model with multi-state perspective for low-light image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 21267–21276 (2025).

[5] Tu, L., Wu, J., Wang, C., Meng, D., Jin, Z.: Fourier-based decoupling network for joint low-light image enhancement and deblurring. IEEE Transactions on Image Processing 34, 5184–5199 (2025). https://doi.org/10.1109/TIP.2025.3592559

[6] Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: European Conference on Computer Vision, pp. 17–33 (2022). https://doi.org/10.1007/978-3-031-20071-7_2

[7] Jia, X., De Brabandere, B., Tuytelaars, T., Van Gool, L.: Dynamic filter networks. In: Advances in Neural Information Processing Systems 29, pp. 667–675 (2016).

[8] Mildenhall, B., Barron, J.T., Chen, J., Sharlet, D., Ng, R., Carroll, R.: Burst denoising with kernel prediction networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2502–2510 (2018).

[9] Xia, Z., Perazzi, F., Gharbi, M., Sunkavalli, K., Chakrabarti, A.: Basis prediction networks for effective burst denoising with large kernels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11844–11853 (2020).

[10] Zhang, Y., Li, D., Shi, X., He, D., Song, K., Wang, X., Qin, H., Li, H.: KBNet: Kernel basis network for image restoration. arXiv preprint arXiv:2303.02881 (2023). https://doi.org/10.48550/arXiv.2303.02881

[11] Ji, B., Yao, A.: High-pass kernel prediction for efficient video deblurring. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2442–2452 (2025).

[12] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595 (2018).

[13] Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "Completely Blind" image quality analyzer. IEEE Signal Processing Letters 20(3), 209–212 (2013). https://doi.org/10.1109/LSP.2012.2227726

[14] Wang, S., Zheng, J., Hu, H.-M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing 22(9), 3538–3548 (2013). https://doi.org/10.1109/TIP.2013.2261309

[15] Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. International Journal of Computer Vision 129(4), 1013–1037 (2021). https://doi.org/10.1007/s11263-020-01407-x

[16] Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3063–3072 (2020).

[17] Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8878–8887 (2019).

[18] Cho, S.-J., Ji, S.-W., Hong, J.-P., Jung, S.-W., Ko, S.-J.: Rethinking coarse-to-fine approach in single image deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4641–4650 (2021).

[19] Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage Retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 12504–12513 (2023).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf