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

Supplementary Material: Results and Experimental Details

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.
*Correspondence: zebukang@hotmail.com.
Minhui Kang and Peizhe Long contributed equally to this work.
Journal: *Signal, Image and Video Processing*.

Reading guide. Section S1 expands the computation; Sections S2–S4 support the test and visual results. Sections S5–S6 retain separate controls and explorations. Costs, protocols and access conditions are given in Sections S7–S9.

S1 Kernel-basis expansion and local application

Figure S1 details the tap centering, position-wise expansion and local application in main-paper Section 2.2. Shared bases and location-dependent coefficients have distinct roles.

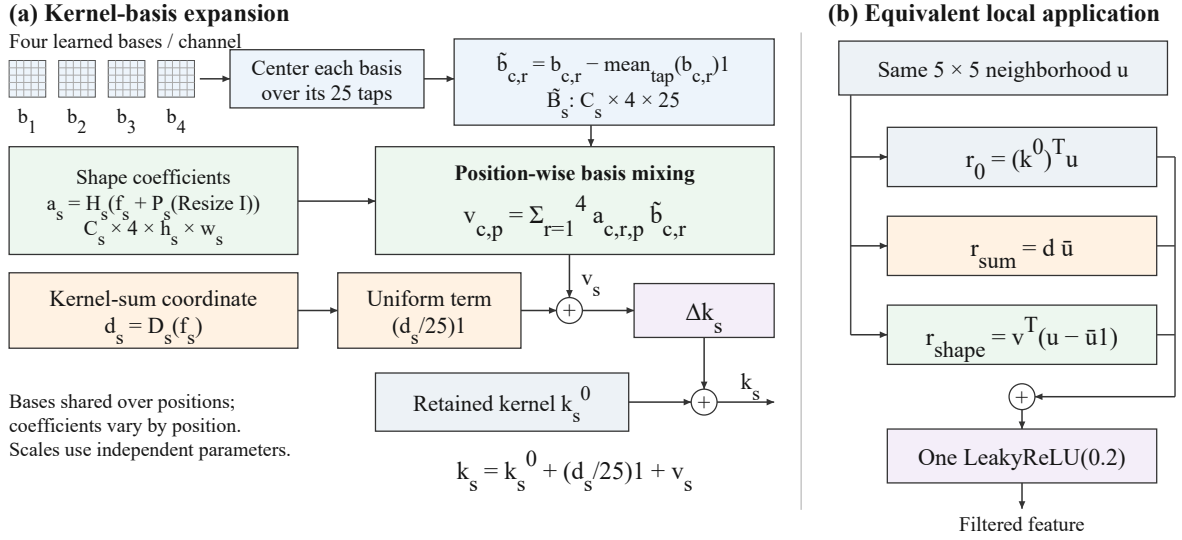


Figure S1: Computational schematic, not measured kernels or responses. (a) Four 5×5 bases per channel are centered over 25 taps and mixed by position-wise coefficients; the sum term and retained kernel are then added. (b) Three linear responses on the same neighborhood are summed before one inherited LeakyReLU(0.2). Dimensions omit batch and denote logical representations, not required tensor materialization.

Expansion and conditioning. For a fixed scale s and channel c, the four centered bases satisfy

$$\widetilde{b}_{s,c,r,t} = b_{s,c,r,t} - \frac{1}{25}\sum_{t'=1}^{25} b_{s,c,r,t'}, \qquad r = 1, \ldots, 4.$$

Centering uses the tap axis only, without mixing channels or averaging the four bases. The logical dimensions are $\widetilde{B}_s \in \mathbb{R}^{C_s \times 4 \times 25}$ and $a_s \in \mathbb{R}^{C_s \times 4 \times h_s \times w_s}$. At each position, four scalar coefficients weight the four complete 25-tap bases before summation. Bases are shared over positions, coefficients vary with position, and scales have independent parameters. Degraded RGB enters only the coefficient route directly; the sum head reads f_s.

One local application. Let u be the replicate-padded 5×5 neighborhood for one channel and location, and $\bar{u} = \mathbf{1}^T u/25$. Before activation, the synthesized kernel gives

$$(k^0 + \Delta k)^T u = (k^0)^T u + d\bar{u} + v^T(u - \bar{u}\mathbf{1}).$$

The retained operator sums these linear contributions before one LeakyReLU(0.2), not through sequential or separately activated filters. Here k^0 comes from the retained generator inside the current P3 network, rather than the complete prediction of the independently trained LEDNet baseline.

Interpretation of the response parameterization. The coordinate d exposes the increment's constant-neighborhood response while centered bases restrict its shape; the original generator stays trainable. Fixing $d = 0$ reduces active capacity. Changing to a center carrier adds $d(e_0 - \mathbf{1}/25)$, rather than preserving the spatial response. Table S4 evaluates these choices separately.

Relation to kernel-basis filtering. KBNet mixes learned bases through position-dependent coefficients [6]; AHFNet mixes prescribed high-pass bases [7]. P3 instead structures an increment to a retained kernel. Its advantage over an ordinary residual at this interface remains unresolved.

S2 Fixed-model official-test differences

The official LOL-Blur test contains 1,800 image pairs from 30 videos, with 60 frames per video. The LEDNet and LEDNet-P3 models are the prespecified primary-run endpoints after 60,000 updates. Table S1 gives the local absolute means and saved contrasts; Figure S2 shows their distribution across videos. Public DarkIR-m and the later released-LEDNet evaluation remain separate system references with different training histories.

Table S1: Local official-test results on the same 1,800 pairs. Paired LEDNet/LEDNet-P3 share a training protocol; public-system references have different training histories. These rows are separate from the published reports in main-paper Table 1.

(a) Absolute means on 1,800 pairs

Model	PSNR (dB)↑	SSIM↑	LPIPS↓
LEDNet (paired primary run)	24.120416	0.782292	0.320062
LEDNet-P3 (paired primary run)	24.074974	0.786948	0.308617
Public systems: different training histories			
DarkIR-m (prespecified reference)	26.998919	0.882238	0.202576
LEDNet (public, post-hoc)	26.059236	0.847589	0.217238

(b) Paired differences and video directions

Contrast	Metric	Difference	Better/worse
P3−LEDNet	PSNR (dB)	−0.045441	16/14
	SSIM	+0.004656	26/4
	LPIPS	−0.011445	30/0
P3−DarkIR-m	PSNR (dB)	−2.923944	5/25
	SSIM	−0.095290	0/30
	LPIPS	+0.106041	0/30

P3 denotes LEDNet-P3. Counts are over 30 videos; there are no ties. Positive PSNR/SSIM and negative LPIPS differences favor P3. Differences retain reported precision and are not reconstructed from rounded means. P3–DarkIR-m describes two systems, not a matched architecture comparison. The public-LEDNet row is copied from its saved post-hoc summary and does not replace the paired baseline or define a new primary contrast.

For the paired models, SSIM improves in 26/30 videos and LPIPS in 30/30. PSNR differences range from -2.150967 to +2.420956 dB; the near-zero aggregate change does not imply negligible changes in every video. Figure S2 preserves the same 30-video order across metrics. The 16/30 positive PSNR signs describe frequency, not magnitude. These metric-specific results are not a human-preference experiment.

Let $m^{M}_{v,i}$ denote a saved primary-metric score for frame i of video v and fixed model M. The equal-video aggregation and paired video difference are

$$\overline{m}^{M}_{v} = \frac{1}{60}\sum_{i=1}^{60} m^{M}_{v,i}, \qquad \overline{m}^{M} = \frac{1}{30}\sum_{v=1}^{30} \overline{m}^{M}_{v},$$
$$\Delta_v = \overline{m}^{\mathrm{P3}}_{v} - \overline{m}^{\mathrm{LEDNet}}_{v}.$$

Differences pair the same frames and videos. Means average 60 frame scores within each video, then videos equally. All displayed values and directions are copied from saved records and describe the fixed models, rather than stability across training runs.

The video-level results are retained in the authors' research records; the figure displays the saved differences without recomputing the underlying metrics.

Relation to the fixed visual examples. Main-paper Figure 3 uses cases 0164/0039, 0142/0039 and 0215/0039, selected before the original evaluation by a fixed identity-based rule. The same central 256×256 crop, $[432, 192, 688, 448]$, is used for each original 640×1120 image. Red rectangles only indicate this crop; they are display overlays rather than

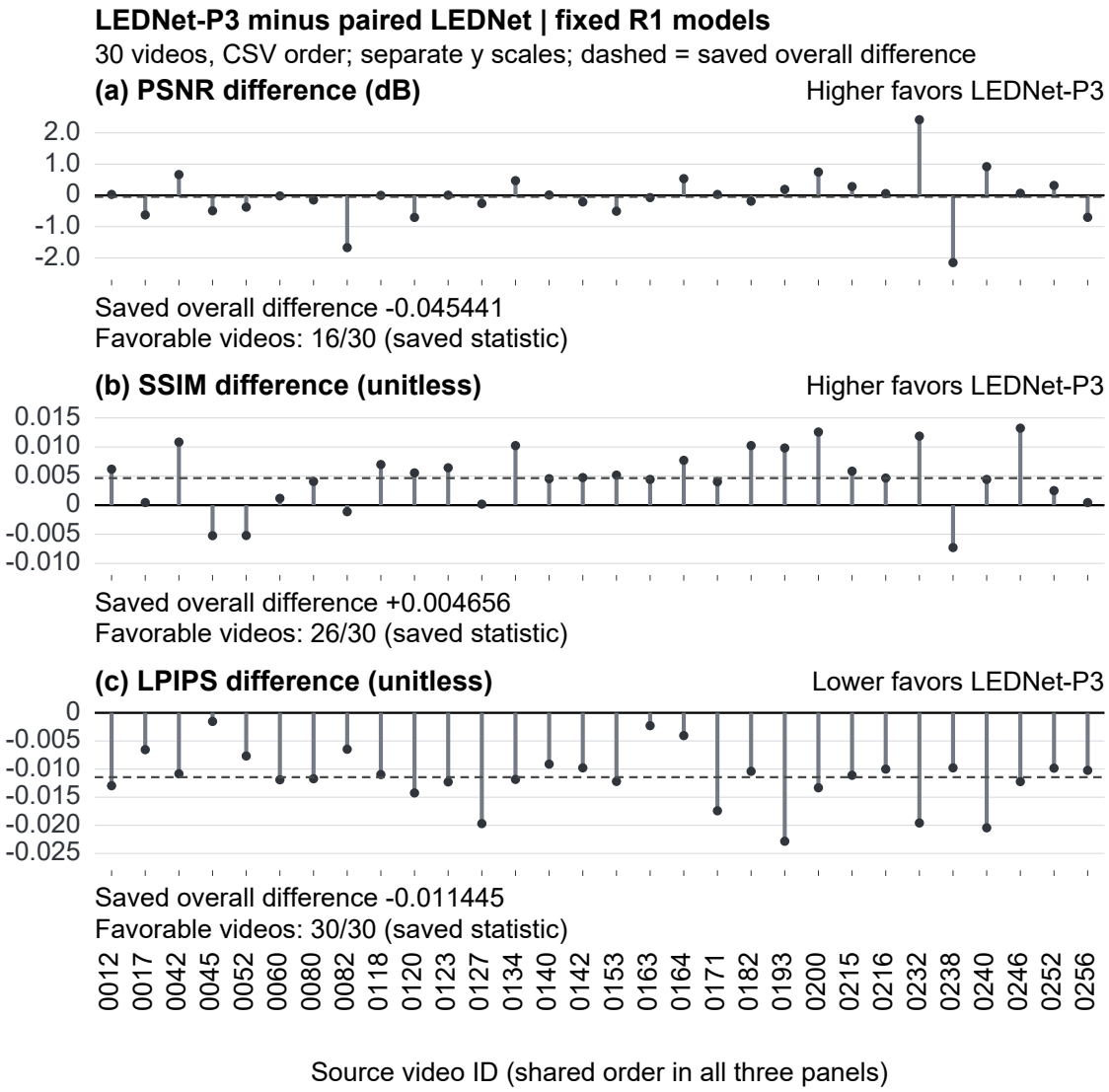


Figure S2: Per-video quality differences for the fixed models. All 30 videos retain the same order across panels. Solid lines mark zero; dashed lines and annotations reproduce the saved overall differences and statistics.

changes to the saved inputs, predictions or targets. The video summaries do not assign scores to these individual frames. Section S3 examines the same crops; their provenance accompanies the files.

S3 Local restoration differences

The fixed cases in main-paper Figure 3 also allow a direct pixel-level comparison of the paired models. Figure S3 uses the original formal-evaluation input, LEDNet prediction, LEDNet-P3 prediction and target, rather than the later descriptor replay. The common crop is $[432, 192, 688, 448]$ in (x_0, y_0, x_1, y_1) coordinates, with exclusive right/bottom limits: rows 192–447 and columns 432–687 of each 640×1120 image.

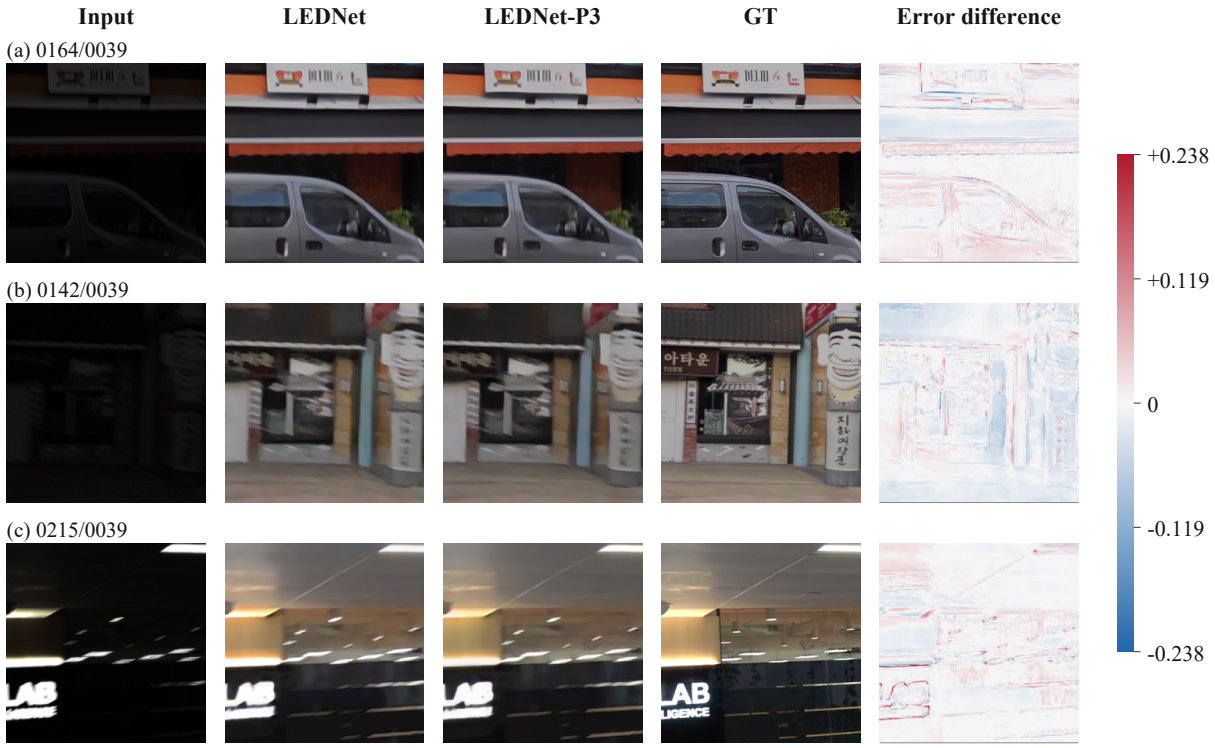


Figure S3: Saved predictions and post-hoc local RGB error differences for (a) 0164/0039, (b) 0142/0039 and (c) 0215/0039. Each RGB8 crop is converted to floating point and divided by 255. With $E_M(p) = \text{mean}_c\,|I_M(p,c) - \text{GT}(p,c)|$, the final column displays $D = E_{\text{LEDNet}} - E_{\text{P3}}$. Red ($D > 0$) favors P3 locally; blue ($D < 0$) favors LEDNet. All cases share a zero-centered symmetric scale, $[-0.2379085, +0.2379085]$, set by the maximum absolute D over the three fixed crops. This is a visualization of saved RGB pixel errors, not an LPIPS map, a new complete-test evaluation or evidence for a perceptual mechanism.

Local restoration differences. The paired predictions remain visually close. In case (a), differences follow the sign edges, awning and vehicle/window outline; red and blue neighborhoods coexist around these boundaries, while lettering and vehicle detail remain softer than the target. Case (b) shows mixed differences near the storefront sign, doorway and vertical trim. Both predictions still blur the text and facade detail that is clearer in the target. In case (c), changes concentrate around ceiling-light edges, signage and dark panel boundaries, with residual blur in both restorations. Many smoother areas are near white on the shared difference scale. The map therefore localizes small, spatially mixed changes rather than showing uniformly better edges or recovered texture.

The shared scale retains unfavorable differences without per-image stretching. These three crops do not explain the official SSIM/LPIPS differences, establish human preference or replace any saved metric result.

S4 Supplemental image-quality descriptors

NIQE and lightness-order error (LOE) were evaluated after the original analysis for the three prespecified fixed models on the same 1,800 pairs. Their separate inference pass is not assumed to reproduce the original RGB8 pixels bitwise. The public-LEDNet row comes from its later six-metric evaluation, in which all six metrics use one saved output set. These post-hoc descriptors supplement the primary PSNR/SSIM/LPIPS results.

Result. The additional descriptors are mixed for LEDNet-P3: input-reference LOE is lower than for LEDNet, whereas NIQE and target-reference LOE are higher. They therefore do not establish an additional overall quality advantage.

Table S2: Post-hoc equal-video means on 1,800 pairs. All descriptors are lower-is-better; the public-LEDNet row has the separate system identity described in Section S8.3. $\mathrm{LOE}_{\mathrm{low}}$ references the degraded input; $\mathrm{LOE}_{\mathrm{high}}$ references the sharp target.

Model	NIQE↓	$\mathrm{LOE}_{\mathrm{low}}$ ↓	$\mathrm{LOE}_{\mathrm{high}}$ ↓
LEDNet	7.312998	280.620471	209.015647
LEDNet-P3	7.368696	279.434401	210.402462
DarkIR-m (public model)	5.537144	239.005692	86.100837
LEDNet (public, post-hoc)	5.572117	238.232147	112.601764

DarkIR-m has lower values on all three descriptors, but remains a system reference with a different training recipe.

NIQE definition. NIQE [4] uses the BasicSR implementation and its supplied natural-scene statistics, without fitting a new model to LOL-Blur. RGB8 outputs are converted to grayscale using OpenCV on float32 RGB/255, mapped to $[0, 255]$ and rounded. The metric uses nonoverlapping 96×96 blocks and two scales. Its internal complete-block crop is retained, with no external border crop. The pinned implementation and natural-statistics identity are documented in the authors' research records.

Lightness-order definition. Lightness is $L(p) = \max\{R(p), G(p), B(p)\}$. Float64 lightness fields are resized from 640×1120 to 50×88 using separable antialiased Keys bicubic interpolation, pixel-center mapping and symmetric extension, without requantization. For $N = 4400$ sampled locations,

$$\mathrm{LOE}(L_r, L_o) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left[\mathbf{1}\{L_r(i) \geq L_r(j)\} \ \mathrm{XOR} \ \mathbf{1}\{L_o(i) \geq L_o(j)\} \right]. \tag{S1}$$

Here L_o is output lightness and L_r is either the degraded input or paired target [5]. Ties use $\geq$; normalization is by N, not N^2. Each descriptor is averaged over frames within each video and then equally across videos. No exposure adjustment is applied. Input-reference LOE describes preservation of lightness order in the blurred input; it is not a deblurring-fidelity or human-preference measure. The descriptors use different reference images, so lower $\mathrm{LOE}_{\mathrm{low}}$ can coexist with higher $\mathrm{LOE}_{\mathrm{high}}$. Their ordering is descriptive; neither is an exposure-correction score.

S5 Development repetitions and design choices

S5.1 Two fixed-budget training repetitions

Table S3 reports the primary and independently initialized repeat studies. Both use the same 153-video fitting set and 17-video development set, with a fixed endpoint of 60,000 updates. Within each comparison, shared initialization and the actual sample, crop and augmentation streams are paired. The feature-conditioned control preserves the full feature path and parameter count, replacing the RGB auxiliary input with means of three nearly equal feature-channel partitions.

Result. LEDNet-P3 improves SSIM and LPIPS over both controls in both repetitions. PSNR differences are negative in the primary run and positive in the repeat run, so the results do not establish a stable PSNR improvement. The feature-conditioned control also improves SSIM and LPIPS over LEDNet, showing that the complete benefit cannot be assigned to RGB access alone. Neither repetition reaches the original PSNR screening references of +0.15 dB against LEDNet and +0.10 dB against the feature-conditioned control. The two rounds are not pooled, and the favorable repeat-run PSNR direction does not replace the primary-run official-test result.

S5.2 Kernel-sum branch and carrier

The branch study compares the full response with Shape-only, which fixes the kernel-sum increment to $d = 0$. The carrier study compares the uniform carrier $d\mathbf{1}/25$ with the center carrier de_0. Each study has its own newly trained control; neither control is the official-test model.

Table S3: Development results on 1,020 pairs. The two training runs remain separate; the primary run supplies the prespecified official-test model.

Training run	Model	PSNR (dB)↑	SSIM↑	LPIPS↓
Primary	LEDNet	22.492249	0.783652	0.303570
	Feature-conditioned	22.435868	0.784536	0.299870
	LEDNet-P3	22.418423	0.787842	0.293450
Repeat	LEDNet	22.371705	0.781247	0.305541
	Feature-conditioned	22.381666	0.784691	0.300366
	LEDNet-P3	22.425153	0.786296	0.299055
Training run	**Contrast**	**ΔPSNR (dB)**	**ΔSSIM**	**ΔLPIPS**
Primary	P3−LEDNet	−0.073826	+0.004190	−0.010120
	P3−Feature-conditioned	−0.017445	+0.003305	−0.006420
	Feature-conditioned−LEDNet	−0.056381	+0.000885	−0.003700
Repeat	P3−LEDNet	+0.053448	+0.005049	−0.006485
	P3−Feature-conditioned	+0.043488	+0.001605	−0.001311
	Feature-conditioned−LEDNet	+0.009960	+0.003443	−0.005174

Differences retain the reported values and are not reconstructed from rounded absolute entries.

Table S4: Separate 60k design comparisons on 1,020 development pairs. Comparisons are within panels, not across independently trained controls.

Variant or contrast	PSNR (dB)↑	SSIM↑	LPIPS↓
(a) Kernel-sum branch			
Full response (branch control)	22.448668	0.787880	0.295035
Shape-only	22.305854	0.785444	0.296671
Full−Shape-only	+0.142813	+0.002436	−0.001636
(b) Carrier choice			
Uniform (carrier control)	22.514801	0.788482	0.293950
Center	22.313944	0.786925	0.296384
Center−Uniform	−0.200857	−0.001557	+0.002434

Result. Retaining the kernel-sum branch improves all three metrics in its paired study. Full response and Shape-only have 226,880 and 189,696 active added parameters, respectively; this is a branch intervention rather than a capacity-matched mechanism test. The uniform carrier is better than the center carrier on all three metrics in the separate carrier study. Both carriers sum to d, but their difference is $d(e_0 - \mathbf{1}/25)$, a fixed zero-sum spatial direction. Replacing the carrier therefore changes the response to nonuniform neighborhoods as well as its parameterization; the comparison does not establish a universal carrier rule.

S5.3 Earlier component and conditioning observations

Table S5 retains the earlier 20k comparison of branch restrictions and conditioning location. These results use a different schedule and a reused historical full-model control, rather than the later same-round 60k controls. They address whether either response component alone, uncentered bases, or direct RGB access to the sum head was preferable in that earlier setting. They are not additional replications of the main 60k experiment.

Table S5: Earlier 20k exploratory results on the development set. These models are distinct from both final paired training runs.

Variant	PSNR (dB)↑	SSIM↑	LPIPS↓
Full response	22.447309	0.762169	0.329312
Shape-only	22.426905	0.760474	0.330509
Kernel-sum-only	22.285632	0.755724	0.336209
Uncentered bases	22.408910	0.758882	0.332856
RGB also supplied to sum head	22.349666	0.758121	0.331894

The full response has the best point estimates in this exploratory set. Supplying RGB directly to the sum head does not

improve the tested model. These observations provide context for the retained design; differences in active capacity and the historical control prevent attribution of each change to an isolated mechanism.

Separate author-confirmed historical round. A separate author-confirmed multi-seed round reports LEDNet-P3 minus LEDNet and LEDNet-P3 minus the feature-conditioned control PSNR differences of +0.151002/+0.102211 dB. Its complete per-seed statistics and training/evaluation contract are not specified in this manuscript. These accepted historical results are kept separate, without recalculation or pooling, and do not enter the primary/repeat comparisons or support their claims.

S6 Exploratory transfer and post-training

S6.1 Adaptation at a DarkIR feature-residual interface

A separate adaptation adds a zero-initialized 5×5 residual after the first DarkIR-m decoder stage, at one-quarter resolution with 128 channels. The ordinary predictor uses two linear layers, $128 \to 28 \to 3200$, so that mapping has rank at most 28; the response-decomposed route retains the uniform carrier, four centered bases and separate feature/RGB conditions. Initialization and training streams are paired, but close parameter counts do not imply matched functional capacity. All three models train from scratch for 60k updates on 136 videos and are evaluated on a 17-video holdout (1,020 pairs). The original development set and official test are not used. The holdout previously belonged to the LEDNet fitting set, so it is not historically unused external data. Training uses 256^2 patches, batch one, FP32 and AdamW: learning rate 5×10^{-4}, betas (0.9, 0.9), weight decay 10^{-3} and cosine decay to 10^{-6}. All three models share the objective

$$\begin{aligned}\mathcal{L} = \mathcal{L}_1(y, g) + 0.01P(y, g) + 50E(y, g)\\ + \mathcal{L}_1(N_8(s), g) + 0.01P(N_8(s), g).\end{aligned} \tag{S2}$$

Here y is the restored image, g its target, and s the side output. P is the implemented frozen VGG19 feature-L_1 loss with five weights $(1/32, 1/16, 1/8, 1/4, 1)$ and no additional ImageNet normalization. E is the Gaussian/Laplacian residual MSE with replicate boundaries, and N_8 is nearest-neighbor upsampling of the side output. This is the shared implemented objective for this study, not a claim of reproducing every detail of DarkIR's published training. Native input normalization, whole-image FP32 inference, RGB8 conversion and the common metric definitions are retained. These are constrained training runs, not the public DarkIR-m model.

Table S6: DarkIR adaptation on 1,020 holdout pairs. These independently trained models are not the public system in the main comparison.

Model	Parameters	PSNR (dB)↑	SSIM↑	LPIPS↓
Native DarkIR-m	3,321,638	24.440582	0.785853	0.341269
Ordinary residual	3,421,634	24.496358	0.784288	0.341910
Response-decomposed	3,420,582	24.455982	0.782232	0.343255

Table S7: Saved contrasts in the DarkIR adaptation study. Positive PSNR/SSIM and negative LPIPS favor the first model. Counts are over 17 holdout videos.

Contrast	Metric	Difference	Better/worse
Response−ordinary	PSNR (dB)	−0.040376	5/12
	SSIM	−0.002056	4/13
	LPIPS	+0.001346	4/13
Response−native	PSNR (dB)	+0.015400	7/10
	SSIM	−0.003621	2/15
	LPIPS	+0.001986	3/14
Ordinary−native	PSNR (dB)	+0.055776	9/8
	SSIM	−0.001565	6/11
	LPIPS	+0.000641	5/12

Result. All point estimates for response-decomposed minus ordinary residual are unfavorable. Relative to native DarkIR, the response-decomposed model also has worse SSIM and LPIPS means, with unfavorable directions in 15/17 and 14/17 videos, respectively. The primary contrast is response-decomposed minus ordinary residual. The saved differences and direction counts do not establish an additional benefit of response decomposition in this adaptation, nor do they demonstrate equivalence. This distinct feature-residual interface does not replace an ordinary-residual control at the native LEDNet interface, and the result does not imply universal failure.

S6.2 CNN post-training with a frozen LEDNet-P3

Two heads—an ordinary CNN and a multiscale feature-fusion CNN—are trained for 10,000 updates on the same data and objective, with matched sample and augmentation streams. The final primary-run LEDNet-P3 remains frozen. Evaluation uses the original 1,020-pair development set, which is exploratory rather than an independent confirmation set.

Table S8: Post-training results. Neither CNN head is part of the final method.

Model	PSNR (dB)↑	SSIM↑	LPIPS↓
Frozen LEDNet-P3	22.418423	0.787842	0.293450
P3 + ordinary CNN	22.436965	0.787751	0.293389
P3 + multiscale CNN	22.438926	0.787766	0.293402
Multiscale−ordinary	+0.001960	+0.000015	+0.000012

Result. Both heads improve PSNR by approximately 0.02 dB over the frozen model, without a three-metric improvement. The multiscale head adds only 0.001960 dB over the ordinary head and has a slightly higher LPIPS. The tested post-training setting therefore provides no substantive additional evidence for multiscale fusion. Saved contrasts retain their reported precision rather than being reconstructed from rounded means.

S7 Computational cost

Measurements use an RTX 5060, 640×1120 inputs, batch one, FP32, PyTorch 2.7.1 and CUDA 12.8. Synchronized model-call latency excludes the first five calls. LEDNet/LEDNet-P3 include their internal synchronous CPU storage transfers; DarkIR-m uses its native path. These are deployment measurements. Table S9 extends main-paper Table 4 with latency distributions; Table S10 gives training costs.

Table S9: Measured inference cost at the original image size. Parameter counts are exact; main-paper Table 4 displays them in millions and converts forward time to ms/image. Peak GPU memory denotes PyTorch allocated/reserved memory, not total device occupancy.

Model	Parameters	Mean (s/image)	Median / p90 (s/image)	GPU peak (MiB) allocated / reserved
LEDNet	7,408,893	0.299882	0.300545 / 0.384937	6512.65 / 6518
LEDNet-P3	7,635,773	0.352979	0.347483 / 0.381148	6513.52 / 6520
DarkIR-m	3,321,638	0.167003	0.160176 / 0.184838	1085.44 / 1182

Table S10: Original training cost with 256^2 patches and batch one. Memory is peak allocated GPU memory.

Model	Mean full iteration (s)	Training memory (MiB)
LEDNet	0.126484	5274.31
Feature-conditioned	0.132653	5316.17
LEDNet-P3	0.135920	5316.17

Result. LEDNet-P3 adds 226,880 parameters and approximately 17.7% mean forward latency. The original complete training iteration is 7.460% slower, with 0.794% greater allocated training memory. These measurements do not establish an efficiency advantage. DarkIR-m's public training cost is unavailable, so its inference latency is not a substitute for training compute. Forward time excludes input transfer, output conversion and CPU metrics. cuDNN benchmarking is enabled for LEDNet/P3 and disabled for DarkIR-m; matrix-multiplication TF32 is disabled, cuDNN TF32 is enabled and determinism is disabled. These policies limit cross-system attribution. Loading, I/O, saving and CPU-metric accounts are retained in the authors' research records.

S8 Training and evaluation details

S8.1 Model configuration and training objective

Data and model roles. LOL-Blur [1] is divided by source into 153 fitting videos (9,180 pairs), 17 development videos (1,020 pairs) and 30 official-test videos (1,800 pairs). The primary and repeat training runs in main-paper Table 3 use

random seeds 10 and 11, respectively. Seeds do not change the split. Official-test LEDNet/P3 use seed10 60k endpoints, not best-score selection. Public DarkIR-m [2] uses the standard 170-video training portion and a different reported crop, batch and optimization recipe; checkpoint-specific complete pretraining and training-compute records are unavailable. It is not a locally trained, matched structural control. In contrast, the three DarkIR models in Section S6 are trained from scratch on their own 136-video fitting split; their scores must not be mixed with those of the public model.

Optimization and paired streams. The fixed-60k LEDNet comparisons train from scratch for 60,000 updates using 256 × 256 patches, batch one and Adam (initial learning rate 10^{-4}; betas 0.9/0.99), with cosine decay to 10^{-7}. There is no added weight decay, warmup, EMA or mixed precision. Initialization and actual sample/crop/augmentation streams are paired within each comparison. The network and response parameterization are defined in main-paper Section 2. The encoder widths are [32, 64, 128, 128], with the retained connection=false configuration. PPM pooling bins are (1, 2, 3, 6), and CurveNLU uses three curve updates. The filtering scales (1/8, 1/4, 1/2) have (128, 128, 64) channels, respectively.

Filtering interface and initialization. At each scale, the original generator consists of three 3×3 Conv–ReLU pairs and a 1×1 projection to $25C_s$ coefficients. The added 1×1 sum and coefficient heads output C_s and $4C_s$ channels; the 3×3 RGB projection maps three channels to C_s. Four 5×5 bases per channel are centered over 25 taps, shared over spatial positions but not across scales. The two output heads have zero weights and biases at initialization, while the bases have random initialization with standard deviation 0.02. Thus the added kernel initially vanishes. No positivity constraint, unit-sum normalization, clipping bound or additional activation is imposed on the increment. The inherited FAC uses replicate boundaries and LeakyReLU with slope 0.2.

Inherited loss. All LEDNet variants retain the same pixel and perceptual objective: main-image L_1 plus 0.01 times a VGG19 feature-L_1 loss, with an auxiliary side-output loss weighted by 0.8. The side term also contains L_1 and the 0.01-weighted perceptual term. Its target is bicubically resized with align_corners=false. VGG19 uses conv1_2, conv2_2, conv3_4 and conv4_4 with equal layer weights; images are mapped from [-1, 1] to [0, 1] before ImageNet normalization. The inherited side output is a training auxiliary, not an input to the P3 heads.

S8.2 Image conversion and primary metrics

Input and output conversion. LEDNet/P3 inputs use $I/127.5 - 1$, reflect padding to a multiple of eight and crop-back. Outputs are clipped to [-1, 1], mapped to [0, 255], rounded and converted to RGB8. DarkIR-m uses $I/255$, native zero padding and crop-back; outputs are clipped to [0, 1] and converted to RGB8 in the corresponding way. Whole-image inference uses original 640×1120 images without tiling, test-time augmentation or exposure correction.

Primary metrics. PSNR is computed from RGB image-level mean-squared error. SSIM is computed per RGB channel with an 11×11 Gaussian window ($\sigma = 1.5$) and valid-window treatment, with no extra evaluation-border crop. LPIPS [3] uses package 0.1.4, model version 0.1 and its learned VGG16 branch on CPU. RGB8 images are mapped to [-1, 1] before evaluation; no second range normalization is applied. This VGG16 evaluation metric is distinct from the VGG19 training loss.

Aggregation and comparison unit. Means first average frames within each video and then videos equally. Paired contrasts use the same frames and source videos. Their displayed means and direction counts are descriptive of the fixed evaluated models, not training-seed or human-preference evidence.

S8.3 Comparison roles and the published LPIPS difference

Public DarkIR-m obtains PSNR 26.998919, SSIM 0.882238 and LPIPS 0.202576 here; the published values are 27.00, 0.883 and 0.162, respectively. Differences in input-range handling, quantization and CPU/GPU paths were identified, but their effects were not measured and the discrepancy remains unresolved. Publication-specific outputs and environment are unavailable. We retain the common-evaluator result without declaring the published value erroneous or merging the two protocols.

Released paper LEDNet, post-hoc system reference. The additional public-LEDNet row in main-paper Table 1 evaluates the officially released `lednet.pth`, distinct from the paired primary-run baseline and from the separately released retrained model. The published data contract uses the full 170-video training portion, rather than the 153 fitting videos used by the paired models; checkpoint-specific training-compute details remain unverified. Native input/output use [0, 1], reflect padding to a multiple of eight and crop-back. All six metrics use the same 1,800 saved RGB8 outputs and the definitions above. This post-hoc evaluation does not alter the final paired models, select P3 or match training budgets. The previously published 25.74/0.850/0.224 PSNR/SSIM/LPIPS report remains literature context, not a source for the new row. Checkpoint identity and saved numerical records are retained in the authors' research records.

S9 Code and data 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 and are not included in the Supplementary Material. The P3 code, configurations and final checkpoint are available on Zenodo at `https://doi.org/10.5281/zenodo.23918535`. Dataset access remains through its original authors. The public-LEDNet results retain their separate identity and protocol. The Supplementary Material does not redistribute the dataset, third-party weights or a complete environment.

References

[1] S. Zhou, C. Li and C. C. Loy. LEDNet: Joint low-light enhancement and deblurring in the dark. *ECCV*, 2022.
[2] D. Feijoo, J. C. Benito, A. Garcia and M. V. Conde. DarkIR: Robust low-light image restoration. *CVPR*, pp. 10879–10889, 2025.
[3] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. *CVPR*, pp. 586–595, 2018.
[4] A. Mittal, R. Soundararajan and A. C. Bovik. Making a completely blind image quality analyzer. *IEEE Signal Process. Lett.*, 20(3):209–212, 2013.
[5] S. Wang, J. Zheng, H.-M. Hu and B. Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.*, 22(9):3538–3548, 2013.
[6] Y. Zhang et al. KBNet: Kernel basis network for image restoration. *arXiv:2303.02881*, 2023.
[7] B. Ji and A. Yao. High-pass kernel prediction for efficient video deblurring. *WACV*, pp. 2442–2452, 2025.