CVPR
JUNE 3-7, 2026
DENVER
COLORADO

# Efficient Real-Time Raw-to-Raw Denoising for Extreme Low-Light Ultra HD Video on Mobile Devices

**Charantej Reddy Pochimireddy, Subhasmita Sahoo, Apoorva Verma, Palavalli Shyam, Swapnil Malviya, Sarvesh, Raj Narayana Gadde**

*Samsung R&D Institute India - Bangalore*

# Challenges

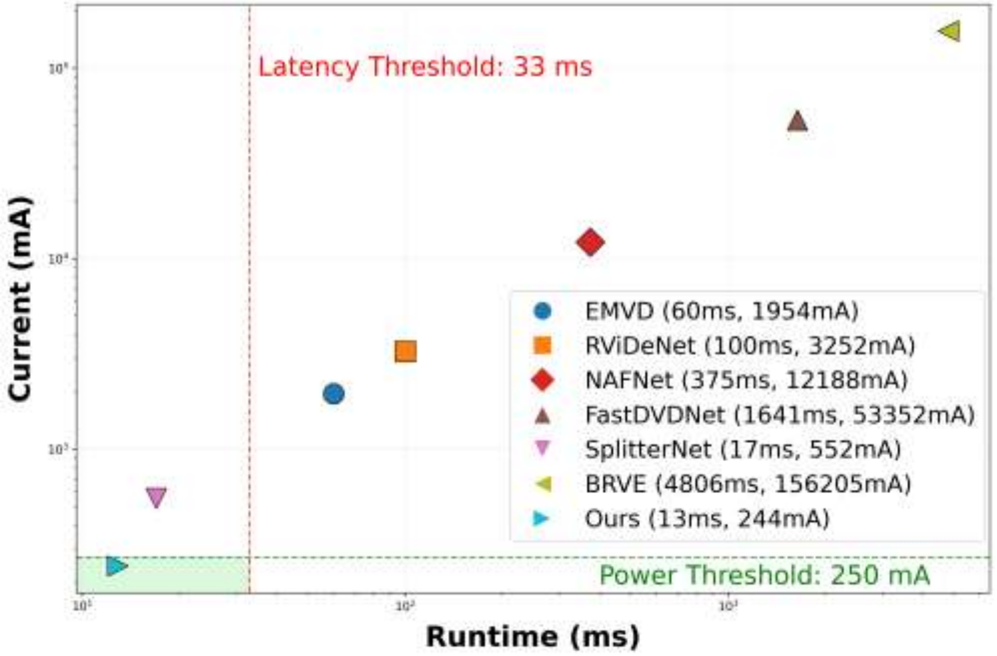
- **Extreme Low Light:** Low signal strength and short exposure times at high frame rates lead to amplification of sensor noise. Fine details are also suppressed, thus degrading visual quality.

- **Small Sensor Size:** Mobile imaging sensors have tiny pixels which capture very few photons in extreme low-light, leading to massive shot and read noise.

- **Limited Compute:** Mobile devices have limited compute and stringent thermal constraints, limiting the use of complex state-of-the-art models.

- **Strict Latency/Power:** For 4K 30fps videos, per-frame processing must be under 33ms and power consumption below 250mA to prevent battery drain.

Target Lux Range: 0lx to 2lx



# Our Contributions

- Synthetic RAW data generation
- Real-world captures, GT verification

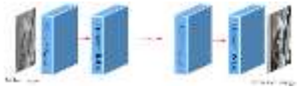
**Data Preparation**

**Model Deployment**

- Model packaging
- Integration into ISP pipeline
- Model evaluation

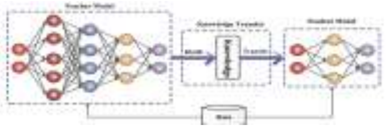
***Raw Low-Light Video Denoiser***

**Model Training**



- Computationally efficient model architecture
- Multi-step training strategy

**Model Optimization**



- Distillation, weight reorientation
- Re-parameterization (layer fusion)

## Dataset Strategy

*Hybrid data creation pipeline merging synthetic data with real tripod-captured mobile bursts.*

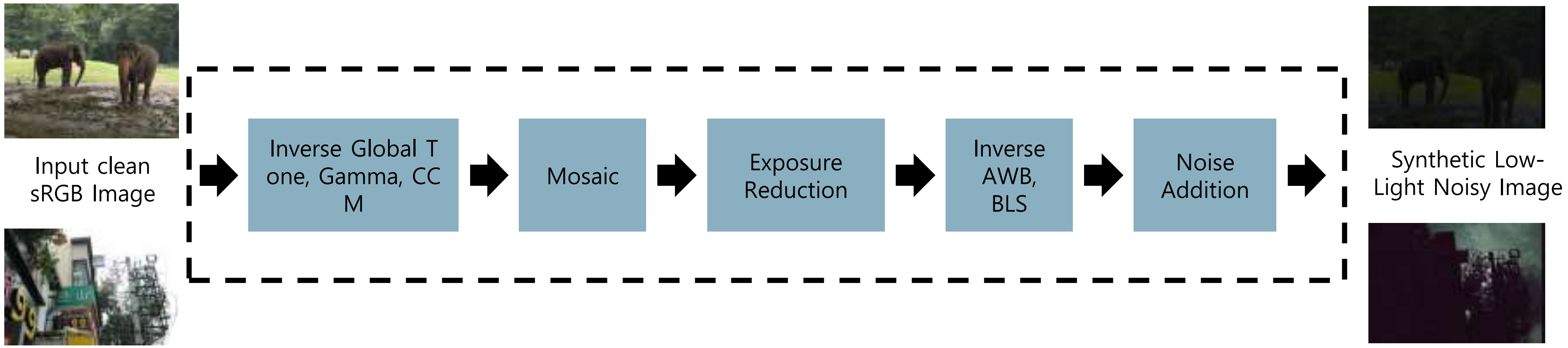
## Model Architecture

*Low-complexity, modular architecture designed specifically for raw-to-raw processing on NPU.*

## Runtime Optimizations

*Holistic deployment path including repar ameterization, spatial restructuring, and INT16 quantization for efficient NPU execution.*

# Dataset Strategy

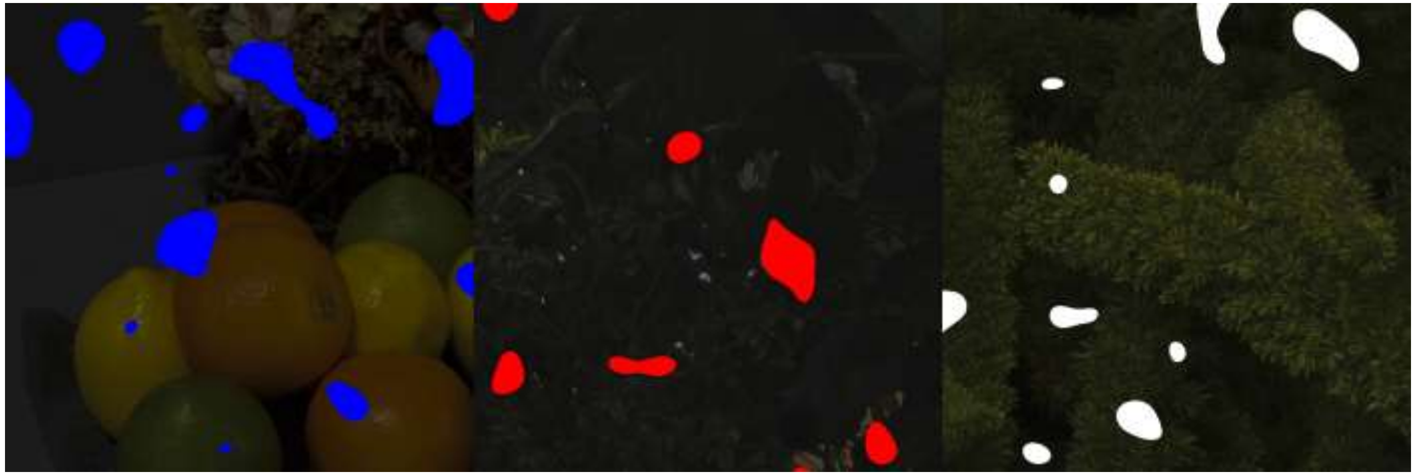



**Synthetic Pipeline:** Unprocess high-quality sRGB images back to pseudo-raw and apply calibrated Poisson-Gaussian noise.

# Dataset Strategy

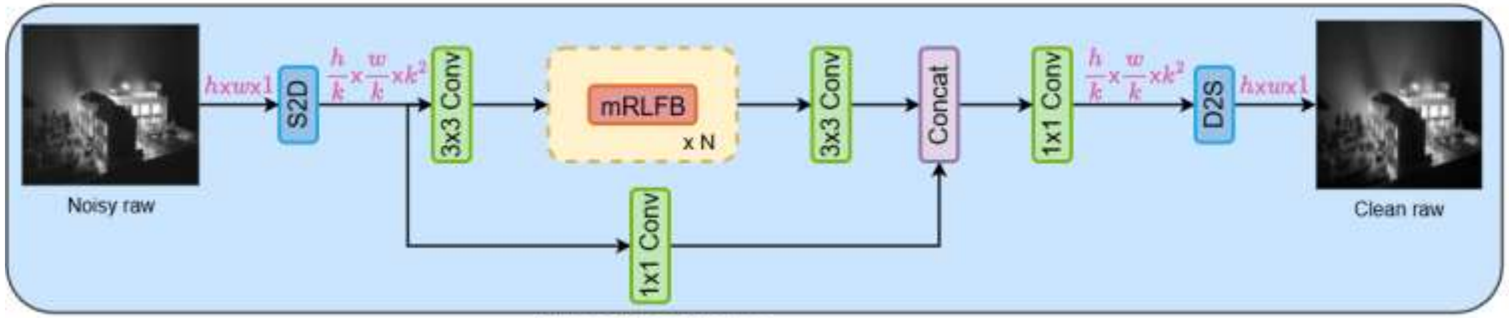

**Real Static Captures:** Lighting controlled tripod-captured videos with ground truth generated via burst averaging + residual denoising.

# Dataset Strategy



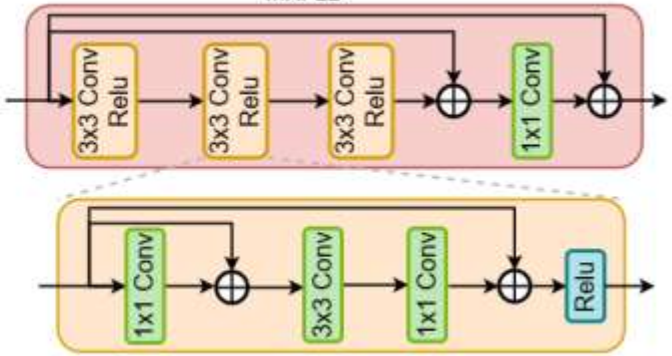
**Controlled Blob Addition:** Simulate the occurrence of point light sources and prevent edge bleeding around high-to-low intensity transition areas.

# Model Architecture and Training Methodology



Full model architecture

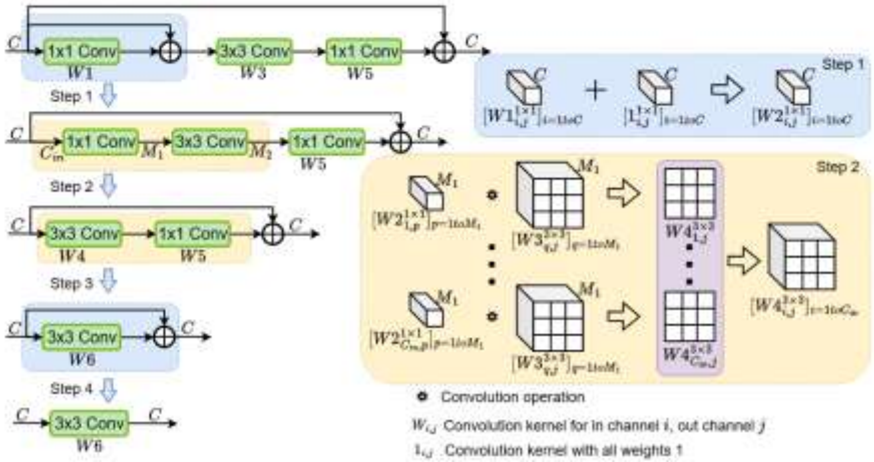
mRFLB

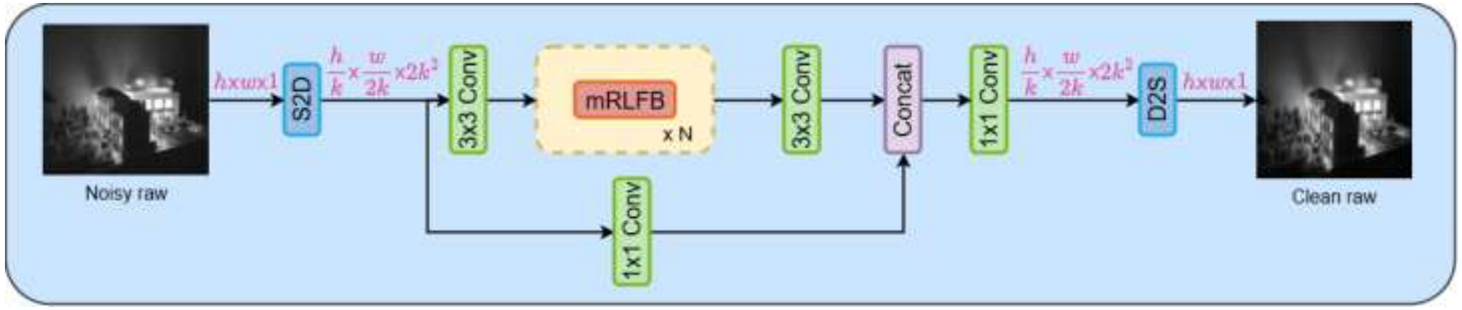


Reparameterizable block for 3x3 conv in mRFLB

- **Efficient Raw Video Denoising Model**: RLFN-inspired, optimized for *mobile devices* with *re-parameterizable structures, heavy attention module removed,* and *strategically placed skip connections for efficient information flow*.
- **UHD Video Processing Optimization**: Utilize $k \times k$ *Space-to-Depth (S2D)* to downsample resolution while expanding channels, preserving colour channel relationships, balancing computational efficiency and output quality.
- **Mobile-Optimized Feature Processing**: A cascade of $N = 4$ *mobile-optimized Residual Local Feature Blocks (mRLFBs)* for efficient feature refinement without *attention modules*.
- **Feature Fusion and Output**: Aggregates *deep features* and *shallow features* through $3 \times 3$ and $1 \times 1$ convolutions respectively, followed by *feature fusion* and *Depth-to-Space (D2S)* to reconstruct denoised output.
- **Composite Loss for Training:** Use *raw reconstruction loss* and *chromatic aberration loss* for raw fidelity and colour consistency.

# Runtime Optimizations






- **Structural Re-parameterization**: Replaces multi-branch block with a single $3 \times 3$ convolution via kernel fusion to reduce memory overhead and improve runtime efficiency.
- **Knowledge Distillation**: Transfers knowledge from a high-capacity teacher model ($N = 4,\ k = 4, d = 32$) to a compact student model ($N = 4, k = 4,$ $d = 16$) using output-space and intermediate-feature guidance loss.
- **Spatial Resolution Reduction**: Halves spatial resolution during inference, doubling channel depth through weight reconfiguration to exploit NPU parallelism.
- **Quantization**: Post-training int16 quantization with per-channel symmetric weights and per-tensor activations, with fine-tuning for quality retention.

# Results

Noisy

Ours

| Method | | | Real-world | | Synthetic→Real | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Runtime (ms) ↓ | Current (ms) ↓ | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ | tOF ↓ x10 | tLP ↓ x100 | Avg. flicker ↓ $x10^3$ | VMAF ↑ |
| **NAFNet** | 375 | 12187 | 59.09 | 0.9838 | 61.49 | 0.9857 | 0.81 | 1.7 | 1.2 | 67.11 |
| **BRVE** | 4806 | 156204 | 59.55 | 0.9993 | 60.42 | 0.994 | 0.61 | 1.26 | 2.46 | 66.87 |
| **SplitterNet** | 16.97 | 551 | 53.83 | 0.9735 | 55.74 | 0.9860 | 1.56 | 3.99 | 2.9 | 40.74 |
| **Ours** | 12.91 | 244 | 60.18 | 0.9883 | 60.54 | 0.9886 | 0.55 | 0.7 | 0.78 | 70.43 |

# Conclusion

- **End-to-end Framework:** We propose a comprehensive, end-to-end framework for real-time raw-to-raw video denoising specifically tailored for extreme low-light environments on mobile devices.

- **Hardware-friendly model design:** We develop a highly efficient mRLFB architecture to drastically reduce computational load while preserving spatial textures and temporal stability.

- **Efficient Real-Time Performance:** Our method achieves 4K video denoising in under 13ms per frame with a power draw of less than 250mA, while maintaining similar perceptual quality and temporal consistency compared to other state-of-the-art methods.

# References


1. M.Chu, Y. Xie, J. Mayer, L. Leal-Taix´e, and N. Thuerey, *“Learning temporal coherence via self-supervision for gan-based video generation,” ACM Transactions on Graphics (TOG), vol. 39, no. 4, pp. 75–1, 2020.*
2. P. Ebelin, G. Denes, T. Akenine-Mo¨ller, K. A˚ stro¨m, M. Oskarsson, and W. H. McIlhagga*, “Estimates of temporal edge detection filters in huma n vision,” ACM Transactions on Applied Perception, vol. 21, no. 2, pp. 1–25, 2024.*
3. Z. Li, *“On vmaf’s property in the presence of image enhancement operations,” 2021.*
4. G. Zhang, Y. Zhang, X. Yuan, and Y. Fu, *“Binarized lowlight raw video enhancement,” in CVPR, 2024.*
5. L. Chen, X. Chu, X. Zhang, and J. Sun, *“Simple baselines for image restoration,” in ECCV. Springer, 2022.*
6. R. Flepp, A. Ignatov, R. Timofte, and L. Van Gool, *“Realworld mobile image denoising dataset with efficient baselines,” in CVPR, 2024*.
7. C. Chen, Q. Chen, J. Xu, and V. Koltun. *Learning to see in the dark. In Proceedings of the IEEE conference on computer vision and pattern recogni tion, pages 3291–3300, 2018.*
8. A. Abdelhamed, S. Lin, and M. S. Brown. *A high-quality denoising dataset for smartphone cameras. In Proceedings of the IEEE conference on co mputer vision and pattern recognition, pages 1692–1700, 2018.*
9. T. Brooks, B. Mildenhall, T. Xue, J. Chen, D. Sharlet, and J. T Barron. *Unprocessing images for learned raw denoising. In Proceedings of the IEEE/ CVF conference on computer vision and pattern recognition, pages 11036–11045, 2019.*
10. Alessandro Foi. *Clipped noisy images: Heteroskedastic modeling and practical denoising. Signal Processing, 89 (12):2609–2629, 2009.*