Journal of Mathematical Finance and Risk Management
ISSN: 3080-9983 (Print), ISSN: 3080-9991 (Online) | Vol. 2, No. 1, 2026

Sensor-Calibrated Poisson–Gaussian Modeling and Adaptive Algorithms for RAW Image Denoising

Yijia Chen *

Hangzhou Foreign Languages School, Hangzhou, Zhejiang, China

* **Corresponding author**: (Email: jia.chy@outlook.com)

Abstract: RAW images preserve information close to the camera sensor, but they also contain noise whose strength changes with brightness. This study develops a denoising workflow for monochrome 10-bit RAW video from an industrial camera. The camera noise was first measured with a Photon Transfer Curve (PTC). Two independently recorded grayscale videos produced the fitted relation variance ≈ 0.106185 × mean + 2.001, supporting a Poisson–Gaussian noise model. The calibrated model was then used in a variance-stabilizing transform before BM3D, creating a strong classical baseline. A multi-frame NAFNet system was trained with synthetic noise generated from the PTC and reference images obtained from real camera bursts. Because a single network did not balance noise removal and edge preservation well, the final system combined a clean-image branch with an edge-preserving branch. A light wavelet refinement with automatically selected strength was applied only to flat regions. On 11 fixed development-evaluation clips covering 0.5–250 frames per second, blind BM3D achieved a mean Denoise–Edge Score (DES) of 0.8324, calibrated VST+BM3D achieved 0.8882, and the final P150 system achieved 0.9589 while passing an edge-safety test. The results show that camera measurement and carefully limited processing can improve both classical and learning-based denoising. The study also discusses the limits of its camera-specific dataset and internal evaluation metric.

Keywords: Adaptive algorithm, BM3D, image denoising, NAFNet, photon transfer curve, Poisson–Gaussian model, RAW imaging.

1. Introduction

Image noise is created by both light and camera electronics. The number of photons reaching a sensor changes randomly, especially when the exposure is short. Electronic readout adds another source of variation. In a RAW image, these effects have not yet been hidden by a full camera processing pipeline. As a result, smooth surfaces may appear grainy, while weak lines and small text may become difficult to recognize. Denoising should reduce this random variation without removing real structure.

A simple smoothing filter cannot solve both parts of the problem. Strong smoothing removes more grain but can blur edges; weak smoothing protects detail but leaves visible noise. Another difficulty is that RAW noise is signal dependent: bright and dark pixels do not have the same noise variance. A method that assumes one fixed noise level may therefore use the wrong strength in different parts of the image. A measured noise model can connect the physical behavior of the sensor to the parameters used by a denoising algorithm [1].

This study asks whether a sensor-calibrated hybrid pipeline can outperform a strong classical baseline while preserving edges. The work has four main stages. First, a Photon Transfer Curve is used to estimate the relation between brightness and noise variance. Second, the measured model is combined with a variance-stabilizing transform (VST) and BM3D [1–3]. Third, a multi-frame NAFNet model learns from both synthetic and real-camera data [4, 5]. Finally, a small adaptive refinement is limited to flat regions so that strong edges are protected. The focus is the complete experimental path from measurement to evaluation, rather than the invention of each individual algorithm.

The study is positioned between physical noise modeling and computational image restoration. Photon-transfer measurement has long been used to connect sensor output with shot and read noise [8], while variance-stabilizing transforms provide a simple mathematical bridge from signal-dependent noise to algorithms designed for nearly constant variance [9]. Multi-frame computational photography provides a second source of information because independent frames contain related scene structure but partly independent noise [10]. These ideas motivate a pipeline in which measurement, classical filtering, and learning are used for different purposes rather than treated as competing solutions.

2. Background

A useful approximation for RAW sensor noise is the Poisson–Gaussian model. Shot noise is related to the random arrival of photons and grows with the expected signal. Read noise is produced by the camera electronics and is treated as approximately signal independent. For a measured pixel value y, the relation can be summarized by a linear variance model:

$$\mathrm{Var}(y) = aE[y] + b \tag{1}$$

Equation (1) is the PTC model used in this study.

In this equation, a is connected to the signal-dependent part of the noise, while b represents the approximately fixed part. A PTC estimates these values by recording a stable grayscale target over many frames and plotting temporal variance against mean brightness. Foi et al. described practical methods for fitting Poisson–Gaussian models to RAW data [1]. The model is an approximation, but it provides a clear test: if variance grows nearly linearly with mean brightness, the measured camera behavior supports its use.

BM3D is a strong classical denoising method. It searches for similar image patches, groups them, suppresses weak transform coefficients, and combines the reconstructed patches [2]. BM3D works best when the added noise has approximately constant variance. Since RAW noise does not

naturally meet this assumption, a generalized variance-stabilizing transform can be applied before BM3D. The transform makes noise strength more even across brightness levels, and an inverse transform returns the result to the original digital-number range [1, 3].

$$z = (2/a)\sqrt{\max(ay + 3a^2/8 + b, 0)} \quad (2)$$

Equation (2) shows the generalized variance-stabilizing transform in the black-level-corrected DN domain.

Neural networks can learn image features that are difficult to describe with a fixed filter. NAFNet was selected because it provides an efficient restoration architecture [4]. However, real noisy frames do not have perfectly clean matching targets. This project therefore combined synthetic input–target pairs with approximate references made from real bursts. The general possibility of learning without perfectly clean labels has also been demonstrated by Noise2Noise [5]. In this study, temporal averaging and BM3D were used as structured references instead of relying only on noisy pairs.

Several related methods help explain the design choices. Bilateral filtering changes its smoothing strength according to spatial distance and intensity difference [11]. Non-local means uses repeated structures within an image [12], an idea also central to the patch grouping of BM3D. Translation-invariant wavelet processing reduces artifacts caused by a fixed transform grid [13]. Standard full-reference measures such as SSIM are valuable when clean targets are available [14], while real-noise benchmarks such as DND and SIDD show the importance of testing on camera data rather than synthetic Gaussian noise alone [15, 16]. The present work uses a smaller camera-specific dataset, so its scores are reported only as within-project comparisons.

3. Experimental Method

The camera produced monochrome 1920 × 1200 RAW frames with 10-bit values. After decoding, the useful range was approximately 0–1023 digital numbers (DN), and the black-level pedestal was near 60 DN. Two grayscale-chart videos were recorded independently for calibration. Within each brightness group, the mean and temporal variance were calculated after accounting for black level. The first fit had slope 0.106221, intercept 2.0334, and $R^2 \approx 0.9995$. The second had slope 0.106149, intercept 1.9695, and a similar R^2. Their average, $a = 0.106185$ and $b = 2.001456$, was stored in the processing code [1].

The data were divided by purpose. Approximately 200 relatively clean PixelShift images were used for synthetic pretraining. Random patches were sampled, and Poisson–Gaussian noise generated from the fitted PTC was added to produce known input–target pairs. Thirty-one real-camera bursts were then used to reduce the difference between simulated and real noise. After frame alignment, temporal averages and offline BM3D outputs provided approximate references. BM3D was more useful in flat areas, while the temporal reference helped prevent excessive edge smoothing [1, 2, 5].

The classical baseline accounted for black level, applied a generalized VST using the fitted parameters, ran BM3D in the stabilized domain, and returned the result to DN values. A simpler blind comparison did not use the PTC or VST. Instead, it estimated one noise value from robust differences between neighboring frames. This comparison isolated the value of calibration: both methods used BM3D, but only one used the measured signal-dependent model [1–3].

The learning system used a short sequence of frames as input. Synthetic pretraining first taught the model a basic mapping from calibrated noisy patches to clean patches. Real-burst training then adapted the model to camera-specific patterns. An early single NAFNet reached a mean DES of about 0.851, below the calibrated BM3D result. Visual inspection showed that stronger denoising cleaned flat regions but sometimes weakened text and narrow contours. Edge-focused training had the opposite behavior. The final system therefore used two NAFNet branches: one favored noise removal, and the other favored edge preservation. A soft edge map blended their outputs according to local image structure [4].

$$\hat{y} = (1 - w)c + we, \qquad 0 \le w \le 1 \quad (3)$$

Here c is the clean-branch output, e is the edge-branch output, and w increases near detected edges.

After the two branches were combined, the remaining noise level was estimated from flat regions. A light wavelet-based refinement was then applied only where the structure mask indicated that strong edges were absent. Its strength was selected automatically from a small set of candidates. The implementation used established ideas from variance stabilization, wavelet shrinkage, and data-based risk estimation [3, 6, 7]. These advanced components were treated as supporting tools rather than new algorithms. Their practical purpose was simple: remove weak residual grain from smooth areas without processing important boundaries again.

$$k^* = \arg\min \mathrm{SURE}(k), \qquad k \in \{0.8, 1.0, 1.2\} \quad (4)$$

Equation (4) states only the selection rule; the established SURE derivation is not claimed as a contribution of this study.

Eleven RAW clips with frame rates from 0.5 to 250 fps formed a fixed development-evaluation set. Because no perfectly clean single-frame ground truth was available, the main metric was a project-defined Denoise–Edge Score (DES). An aligned temporal trimmed mean provided an approximate reference, and unreliable pixels were masked. DES combined 60% flat-region noise reduction and 40% edge fidelity. A separate edge-safety test required minimum edge fidelity on three difficult clips. A method was rejected if it raised average DES but failed this safety test. DES is not a public standard and should not be compared directly with PSNR or SSIM values from other datasets.

$$\mathrm{DES} = 0.6\mathrm{NG} + 0.4\mathrm{EF} \quad (5)$$

NG denotes normalized flat-region noise gain and EF denotes normalized edge fidelity.

All methods were evaluated on the same decoded frames, temporal references, masks, and regions of interest. Mean scores were checked together with per-clip behavior and visual crops. Major design choices were compared separately: blind versus calibrated BM3D, one network versus two branches, and fixed versus noise-adaptive refinement. Failed changes were recorded. In particular, aggressive sharpening could raise measured edge energy while creating unnatural contours, so the edge-safety test was necessary to prevent a misleading improvement.

The main symbols and parameters were kept small enough to interpret directly. The PTC uses only slope a and intercept b. The blend uses one local weight w between zero and one. DES uses two normalized terms: noise reduction and edge fidelity. The final wavelet stage tests only three candidate strength multipliers. This limited parameter set was chosen so

that each adaptive decision could be inspected instead of being hidden inside a large search.

The two grayscale fits also provide a basic repeatability check. Their slope difference was 0.00007186, or about 0.068% of the deployed mean slope. Their intercept difference was 0.06392 DN^2, or about 3.19% of the deployed mean intercept. These ranges are not formal confidence intervals because only two videos were used, but they show that the signal-dependent slope was highly repeatable while the smaller read-noise term was less stable. The average was therefore used as a practical calibration rather than claimed as an exact physical constant.

Table 1. Main parameters and their meanings

Parameter	Value	Interpretation
a	0.106185	PTC slope; signal-dependent noise
b	2.001456 DN^2	PTC intercept; fixed-noise term
Black level	≈60 DN	Sensor pedestal
Input frames	4	Frames supplied to NAFNet
SURE candidates	0.8, 1.0, 1.2	Wavelet strength multipliers
DES weights	0.6 / 0.4	Noise gain / edge fidelity

4. Results

Blind BM3D achieved a mean DES of 0.8324. Calibrated VST+BM3D increased the mean to 0.8882, a gain of 0.0558 without using a neural network. This result supports the main modeling decision: adapting BM3D to the measured sensor noise was better than estimating one raw-domain noise level. The early single NAFNet scored about 0.851 and did not beat the calibrated classical baseline. Therefore, the final improvement cannot be explained simply by replacing BM3D with artificial intelligence [1–4].

The two-branch design and noise-adaptive processing produced a stable improvement over BM3D. The final P150 pipeline reached 0.9589 mean DES on the same 11 clips. This was 0.0707 above calibrated VST+BM3D and 0.1265 above blind BM3D. P150 also passed the edge-safety test. Later refinement produced smaller gains than calibration and branch specialization, suggesting that the main progress came from matching the method to the sensor and separating flat-region cleaning from edge protection.

Not every experiment improved the result. Heavy sharpening was rejected because it damaged edge quality. A noisy-target fine-tuning attempt achieved about 0.887, close to but below calibrated BM3D. Several alternative transforms and threshold rules showed little or no improvement. These negative results reduced unnecessary complexity and helped identify the components that made a consistent difference.

The ablation sequence separates three sources of improvement. Calibration raised BM3D from 0.8324 to 0.8882. A single learning branch did not surpass that baseline. Branch specialization and local fusion produced the first stable improvement, residual-noise adaptation produced a larger later gain, and the final flat-region wavelet stage made a smaller improvement near saturation. This ordering is important because it shows that the final result was not created by the last module alone.

Table 2. Main ablation stages on 11 clips

Stage	Mean DES	Main change
Blind BM3D	0.8324	No calibration
VST+BM3D	0.8882	PTC calibration
Single NAFNet	0.8510	One learning branch
Dual fusion	0.906	Clean + edge branches
Exposure schedule	0.929	Frame-rate-aware processing
Noise gate	0.944	Residual-noise adaptation
P150	0.9589	Flat-region wavelet refinement

5. Discussion

The clearest result is that sensor calibration improved a well-established classical method. The PTC provided two measured parameters rather than a manually chosen filter strength. After variance stabilization, BM3D operated under an assumption closer to its intended noise model. This shows why applied mathematics is important in image processing: a simple model can connect physical measurement to an algorithm and produce a measurable improvement [1–3].

The second result concerns the conflict between denoising and detail preservation. One network was not equally strong in every region. The two-branch system made this conflict easier to control because flat areas and edges could favor different outputs. This does not prove that two networks are necessary for every camera, but it was effective for the available training data. The design also remained understandable: each branch had a clear role, and the blend depended on a visible property of the image [4].

The final refinement was intentionally limited. By that stage, the output was already relatively clean. Applying another strong filter over the whole image would have risked losing small structures for a minor reduction in grain. Restricting the refinement to flat regions was therefore more important than the exact wavelet variant. The detailed shrinkage and threshold-selection methods came from existing research [3, 6, 7]; the contribution of this project was their controlled placement within a calibrated camera pipeline.

Several limitations prevent the reported score from being treated as a general benchmark. Only one monochrome camera family was studied. The 11 clips were repeatedly checked during development, so they are not a completely untouched final test set. DES is an internal metric based on an imperfect temporal reference, and the scenes were mainly indoor and nearly static. Multi-frame processing may produce alignment artifacts in fast motion. The P150 pipeline is also more complex and slower than one BM3D or NAFNet pass.

The reported precision should not be confused with statistical certainty. DES values are numerical outputs of a fixed pipeline, but the dataset contains only 11 clips and does not support a strong population-level claim. In addition, the clips were reused during development. The gain of 0.0707 over calibrated BM3D is therefore evidence for this controlled camera experiment, not a confidence interval for unseen cameras. A stronger study would report per-clip dispersion, confidence intervals over newly collected scenes, and results on a public benchmark [15, 16].

The mathematical contribution is intentionally limited. This work does not introduce a new Poisson distribution, BM3D transform, NAFNet architecture, SURE estimator, or BiShrink rule. Its contribution is the measured connection among these existing ideas: the PTC supplies camera

parameters, variance stabilization aligns the noise with a classical baseline, local blending separates two restoration goals, and a restricted risk-selected refinement targets only the residual error. Stating this boundary makes the work easier to reproduce and avoids claiming ownership of established methods.

Future work should freeze all design choices before collecting a new test set. Additional sensors would show whether repeating the PTC calibration is sufficient or whether the network must also be retrained. Public noisy-clean data would allow PSNR, SSIM, and perceptual metrics to be reported alongside DES. Motion-aware frame selection could improve dynamic scenes, and model compression could make the two-branch system easier to deploy. These steps would test whether the current camera-specific result can become a more general method.

6. Conclusion

This study developed a RAW image denoising workflow beginning with camera measurement. Two grayscale recordings produced a stable linear relation between signal mean and noise variance. Using this model with a variance-stabilizing transform improved BM3D from 0.8324 to 0.8882 mean DES. The result demonstrates that calibration alone can improve denoising before a neural model is added.

A single NAFNet did not outperform calibrated BM3D, so the final system separated noise removal and edge preservation into two branches. A light, flat-region refinement removed remaining grain without processing strong boundaries again. The final P150 system reached 0.9589 mean DES and passed the edge-safety test on 11 development-evaluation clips.

The main lesson is that denoising quality depends on the full experimental process: measure the sensor, match the algorithm to the noise model, compare against a strong baseline, and reject changes that damage visible structure. The result remains specific to one camera and a limited dataset, but it provides a clear and reproducible foundation for future testing.

References

[1] Foi, A., Trimeche, M., Katkovnik, V., & Egiazarian, K. (2008). Practical Poissonian–Gaussian noise modeling and fitting for single-image raw-data. IEEE Transactions on Image Processing, 17(10), 1737–1754. https://doi.org/10.1109/TIP.2008.2001399

[2] Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing, 16(8), 2080–2095. https://doi.org/10.1109/TIP.2007.901238

[3] Mäkitalo, M., & Foi, A. (2013). Optimal inversion of the generalized Anscombe transformation for Poisson–Gaussian noise. IEEE Transactions on Image Processing, 22(1), 91–103. https://doi.org/10.1109/TIP.2012.2202675

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

[5] Lehtinen, J., Munkberg, J., Hasselgren, J., et al. (2018). Noise2Noise: Learning image restoration without clean data. In Proceedings of the 35th International Conference on Machine Learning (Vol. 80, pp. 2965–2974).

[6] Şendur, L., & Selesnick, I. W. (2002). Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. IEEE Transactions on Signal Processing, 50(11), 2744–2756. https://doi.org/10.1109/TSP.2002.804091

[7] Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. The Annals of Statistics, 9(6), 1135–1151. https://doi.org/10.1214/aos/1176345632

[8] Janesick, J. R. (2001). Scientific charge-coupled devices. SPIE Press.

[9] Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. Biometrika, 35(3–4), 246–254. https://doi.org/10.2307/2332343

[10] Hasinoff, S. W., Sharlet, D., Geiss, R., et al. (2016). Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics, 35(6), Article 192. https://doi.org/10.1145/2980179.2980254

[11] Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and color images. In Proceedings of the 6th International Conference on Computer Vision (pp. 839–846). https://doi.org/10.1109/ICCV.1998.710815

[12] Buades, A., Coll, B., & Morel, J.-M. (2005). A non-local algorithm for image denoising. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 60–65). https://doi.org/10.1109/CVPR.2005.38

[13] Coifman, R. R., & Donoho, D. L. (1995). Translation-invariant de-noising. In A. Antoniadis & G. Oppenheim (Eds.), Wavelets and statistics (pp. 125–150). Springer. https://doi.org/10.1007/978-1-4612-2544-7_9

[14] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

[15] Plötz, T., & Roth, S. (2017). Benchmarking denoising algorithms with real photographs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1586–1595). https://doi.org/10.1109/CVPR.2017.168

[16] Abdelhamed, A., Lin, S., & Brown, M. S. (2018). A high-quality denoising dataset for smartphone cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1692–1700). https://doi.org/10.1109/CVPR.2018.00182