Progressive Self-Supervised Denoising using Noise2Void Variants: A Cascading Approach

Lawrence Obiuwevwi, Dushan Wadduwage
Department of Computer Science, Old Dominion University, Norfolk, VA, USA
lobiu001@odu.edu, dwadduwa@odu.edu

***Abstract*—Image denoising is essential in scientific and biomedical imaging, where noise degrades data quality and clean ground truth is often unavailable. Noise2Void (N2V) addresses this challenge through a self-supervised framework that learns from single noisy images by masking and context prediction. This paper explores four key variants—Original N2V, Blind-Spot N2V, Structured N2V, and Probabilistic N2V (PN2V)—each advancing denoising capabilities via architectural or probabilistic refinements. We further introduce a novel progressive retraining strategy, where denoised outputs from one model are used as noisy inputs to retrain same model or another variant model. This approach enables deeper noise removal and measurable gains in PSNR, SSIM, and visual fidelity across synthetic Gaussian and Poisson noise. Our results highlight the effectiveness of retrained cascades in self-supervised denoising pipelines and suggest practical use for real-world, label-free imaging scenarios.**

***Index Terms*—Image Denoising, Noise2Void, Self-Supervised Learning, Progressive Retraining, Microscopy, Deep Learning.**

I. Introduction

High-resolution microscopy is indispensable in biomedical imaging fields such as neuroscience, pathology, and cellular biology. However, images captured under low-light conditions, limited exposure times, or fluorescence limitations are often heavily degraded by noise [4]. This hampers downstream tasks such as cell segmentation, tracking, and morphological analysis, making denoising a critical preprocessing step.

Traditional denoising algorithms, including BM3D [1] and total variation filtering, offer limited flexibility and often rely on hand-crafted priors. Deep learning models like DnCNN [2] have demonstrated superior performance but depend on large, clean datasets for supervised training—an unrealistic requirement in many biomedical applications.

Self-supervised approaches, particularly Noise2Noise [3] and Noise2Void (N2V) [4], have revolutionized the denoising pipeline by eliminating the need for clean targets. N2V trains models using only noisy images by masking central pixels and predicting them from surrounding context, leveraging the natural redundancy in biomedical image structures.

While N2V's original masking strategy was effective, subsequent variants improved its capabilities. Structured N2V introduced grid-based masking to ensure uniform coverage [**?**]. Blind-spot CNNs removed the center pixel from the receptive field entirely [6], and Probabilistic N2V [5] modeled uncertainty through Gaussian distributions.

This paper addresses a fundamental question: Can performance be further improved by combining the strengths of these variants [7] We introduce a progressive denoising framework that sequentially retrains models on the output of other models. This strategy not only compounds their denoising capabilities but also enhances PSNR and SSIM without requiring clean images or extensive architectural changes.

The contributions of this work are:

- A systematic implementation and comparison of four major N2V variants.
- A novel progressive retraining pipeline using variant-to-variant cascades.
- Quantitative and qualitative evaluation across Gaussian and Poisson noise.
- Empirical demonstration that retrained cascades enhance denoising performance.

II. Related Work

Image denoising has been a central problem in image processing for decades. Classical approaches such as Gaussian smoothing, median filtering, bilateral filtering, and total variation denoising aim to suppress noise while preserving edges. However, these techniques often struggle with complex or high-frequency noise and can oversmooth important structural details.

A. Classical and Supervised Methods

Among classical methods, BM3D (Block-Matching and 3D Filtering) has been widely adopted as a benchmark for traditional denoising due to its collaborative filtering in a transform domain [1]. With the advent of deep learning, supervised methods such as DnCNN [2] and U-Net architectures demonstrated strong performance by learning mappings between noisy and clean image pairs. However, these models require large datasets with clean ground truth, which are often unavailable in scientific microscopy due to photon limitations, sample movement, or phototoxicity constraints.

B. Self-Supervised Denoising

The field shifted with the emergence of self-supervised methods. Noise2Noise [3] showed that models could be trained using pairs of noisy images alone. Noise2Void (N2V) went further by enabling learning from a single noisy image [4]. This was accomplished by masking individual pixels and training the network to predict the masked value from its spatial context.

Several enhancements have followed:

- **Structured N2V** uses grid-like or checkerboard masking to ensure more uniform pixel coverage and improve training stability [7] .
- **Blind-Spot CNNs** eliminate the need for masking altogether by architecturally removing the center pixel from the receptive field [6].
- **Probabilistic N2V (N2V2)** introduces probabilistic modeling, where the network predicts both the mean and variance of pixel values and is trained using a Gaussian negative log-likelihood loss [5].

C. Emerging Alternatives

Other self-supervised approaches like Self2Self [8], Neighbor2Neighbor [9], and Blind2Unblind [10] extend the N2V family. These models incorporate dropout, patch-based alignment, and synthetic blind spots to increase robustness. However, most approaches rely on a single-pass strategy.

In contrast, our work explores whether sequential retraining using the outputs of one N2V variant as the input to another can progressively enhance denoising performance—a direction that remains underexplored in the current literature.

III. Methods and Experimental Setup

To ensure a fair and reproducible comparison of denoising performance, we designed a unified experimental framework that includes standardized data preparation, a shared network backbone, consistent training configurations, and a clearly defined evaluation protocol. This section details the four Noise2Void variants implemented and introduces our progressive retraining strategy.

A. Dataset Preparation

We used microscopy image slices in TIFF format. All images were converted to grayscale and normalized to the [0, 1] range. For training and evaluation, we cropped the images into 128×128 patches. Two synthetic noise models were applied to simulate real-world microscopy conditions:

- **Gaussian Noise:** Zero-mean Gaussian noise with fixed variance was added to simulate electronic camera noise.
- **Poisson Noise:** Signal-dependent Poisson noise was used to mimic photon-limited acquisition in fluorescence microscopy.

B. Network Architecture

Each denoising model used a lightweight U-Net-based architecture with three encoder and decoder blocks, ReLU activations, and skip connections. The encoder performed downsampling using strided convolutions, while the decoder used bilinear upsampling followed by convolution. All models share this backbone to ensure that performance differences arise solely from variant-specific masking or loss strategies.

C. Implemented Variants

We implemented four self-supervised Noise2Void variants, each introducing a unique strategy to eliminate the need for clean or paired training data.

1) Original N2V (Krull et al., 2019): This variant masks random pixels from input images and predicts them from their surrounding context using a U-Net architecture. The blind-spot created ensures the network cannot simply memorize pixel values. This method requires no modification to the network architecture and serves as the foundation for subsequent improvements.

2) Structured N2V (broaddus et al., 2020): Structured masking improves training efficiency by using regular patterns (e.g., checkerboards or fixed strides) instead of random pixel drops. This ensures more uniform spatial coverage across batches and mitigates the randomness that can slow convergence. Structured masking also allows faster GPU parallelism due to regular memory access patterns.

3) Blind-Spot CNN (Laine et al., 2019): Instead of masking inputs, this model modifies the receptive field by redesigning convolutional kernels to exclude the center pixel. This architectural change avoids the need for dynamic masking and enables full-image processing in a single pass. However, it requires careful implementation and introduces complexities in kernel stacking.

4) Probabilistic N2V (Krull et al., 2020): This variant enhances denoising by modeling output as a Gaussian distribution. The network predicts both the mean and the variance of pixel intensities and is trained using Gaussian negative log-likelihood loss. This allows the model not only to denoise but also to express uncertainty in its predictions—a critical feature in medical imaging and scientific visualization.

D. Progressive Retraining Strategy

A key contribution of this study is the design of a progressive retraining pipeline. In this approach, the output of one denoising model is used as the input for retraining a second model. For example, we first apply Structured N2V to a noisy image, obtain its denoised output, and then use that output as the "new noisy image" to retrain Probabilistic N2V. This enables the second model to focus on residual noise left by the first model, leading to progressive image refinement.

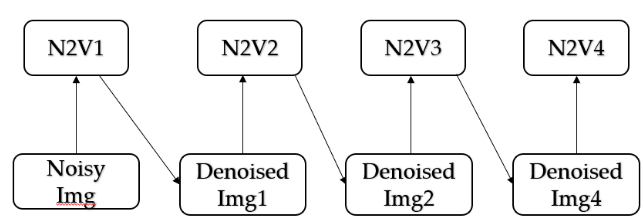

Fig. 1: Progressive cascding of models.

This approach differs from traditional cascaded inference, where pretrained models are applied in sequence without learning from intermediate outputs. By retraining each model on the output of its predecessor, we allow the network to adapt to the noise distribution left behind by prior denoising stages.

E. Training Configuration

All experiments were conducted in Google Colab Enterprise with NVIDIA T4 GPUs. The following hyperparameters were used across all variants unless otherwise noted:

- Epochs: 30 (10 for retraining stages)
- Batch size: 8
- Optimizer: Adam
- Learning rate: 1×10^{-3}
- Loss: MSE (standard variants), Gaussian NLL (for probabilistic variant)

F. Evaluation Metrics

We use three standard metrics to assess denoising quality:

- **Mean Squared Error (MSE):** Measures average pixel-wise error.
- **Peak Signal-to-Noise Ratio (PSNR):** Indicates overall fidelity in decibels.
- **Structural Similarity Index (SSIM):** Captures perceptual and structural consistency.

Performance is evaluated both per variant and across cascaded retraining combinations. Multiple passes are also analyzed to assess the effects of recursive denoising.

IV. Results and Analysis

We present both quantitative and qualitative results comparing individual Noise2Void variants, multi-pass denoising, and our proposed progressive retraining strategy. Metrics were computed against clean reference images using MSE, PSNR, and SSIM.

A. Multi-Pass Denoising Performance

We first analyzed how each variant improves over multiple passes. As shown in Table I, Probabilistic N2V consistently improves denoising performance with each pass. The third pass yielded the lowest MSE and highest PSNR, indicating stable and cumulative enhancement.

TABLE I: Performance Across Denoising Passes (Probabilistic N2V)

Metric	1st Pass	2nd Pass	3rd Pass
MSE	0.000211	0.000176	**0.000149**
PSNR (dB)	36.76	37.55	**38.27**
SSIM	0.8717	0.8972	**0.9201**

B. Progressive Retraining with Cascaded Variants

In our proposed pipeline, we use the denoised image from one model as the training input for a second model. Table II shows that retraining Probabilistic N2V on the output of Structured N2V significantly improves PSNR from 35.32dB to 36.78dB. Similarly, the reverse order (Prob $\rightarrow$ Struct) also improves PSNR relative to Structured N2V alone, but falls slightly short of Probabilistic N2V's standalone second pass.

TABLE II: Progressive Retraining Using Denoised Inputs

Pipeline	MSE ↓	PSNR ↑	SSIM ↑
Structured N2V (2nd)	0.000294	35.32 dB	0.8970
Structured → ProbN2V	0.000210	36.78 dB	0.8724
Probabilistic N2V (2nd)	**0.000176**	**37.55 dB**	**0.8972**
ProbN2V → Structured	0.000181	37.43 dB	0.8968

C. Variant Ranking by PSNR

To better understand the overall effectiveness of each configuration, we ranked all runs by PSNR. Table III shows the performance across multiple passes and retraining pipelines. The best overall result was obtained from the third pass of Probabilistic N2V (38.27 dB). The Structured $\rightarrow$ Probabilistic pipeline also performed strongly, confirming the benefits of progressive learning.

TABLE III: Denoising Results Summary (Sorted by PSNR)

Method	MSE ↓	PSNR ↑	SSIM ↑
Structured N2V (2nd)	0.000294	35.32 dB	0.8970
Structured N2V (3rd)	0.000281	35.51 dB	**0.9315**
Structured N2V (1st)	0.000254	35.95 dB	0.8449
Probabilistic N2V (1st)	0.000211	36.76 dB	0.8717
Structured → ProbN2V	0.000210	36.78 dB	0.8724
ProbN2V → Structured	0.000181	37.43 dB	0.8968
Probabilistic N2V (2nd)	0.000176	37.55 dB	0.8972
Probabilistic N2V (3rd)	**0.000149**	**38.27 dB**	0.9201

D. Visual Results

Visual inspection of denoised outputs confirmed the numerical results. The cascaded outputs showed fewer residual artifacts and smoother edge preservation, especially under Poisson noise. Progressive retraining enabled each model to clean remaining structured noise from previous stages, yielding images with high fidelity and minimal blurring.

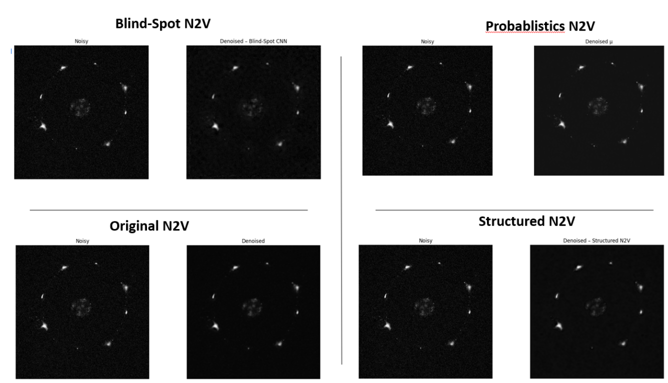

Fig. 2: Visual comparison of noisy input, Structured N2V, Probabilistic N2V, and retrained cascaded outputs.

V. Discussion and Implications

The results from our experiments offer several key insights into the behavior and potential of Noise2Void variants when used independently and in a progressive retraining pipeline.

A. Why Noise2Void Variants Matter

Noise2Void has emerged as a powerful solution in scenarios where ground truth data is unavailable. Its self-supervised training scheme makes it ideal for real-world microscopy tasks, where image acquisition under ideal conditions is often impossible. Our experiments reaffirm the strength of this framework, particularly its ability to generalize across Gaussian and Poisson noise.

B. Probabilistic N2V Provides Strong Baseline

Among all variants tested, Probabilistic N2V consistently achieved the best performance when used independently. Its ability to predict both mean and uncertainty via Gaussian modeling offers additional robustness and interpretability. For applications in medical imaging where confidence in predictions is critical, this feature makes it highly valuable.

C. Progressive Retraining Enables Deeper Denoising

Our key contribution—a progressive retraining strategy—demonstrates that additional denoising performance can be unlocked without access to clean ground truth. By treating denoised images as new noisy inputs for retraining a second model, we enable networks to learn the residual noise distributions left by their predecessors. This is fundamentally different from cascaded inference and introduces a lightweight yet effective way to improve results, as evidenced by the 1.46 dB PSNR gain in the Structured $\rightarrow$ Probabilistic N2V sequence.

D. Variant Complementarity

Structured N2V and Probabilistic N2V appear particularly complementary. Structured N2V enforces spatial regularity through its masking strategy, while Probabilistic N2V refines details by modeling uncertainty. Their combination, through progressive retraining, proves especially effective. This opens up opportunities to explore other pairings of complementary models.

E. Generalization and Practical Use

The retraining approach requires no architecture changes, hyperparameter tuning, or access to clean labels. This makes it suitable for deployment in resource-constrained labs, portable devices, or batch image processing pipelines in biomedical and environmental domains. Moreover, the method can generalize to newer N2V-inspired variants, as well as cross-domain applications such as satellite imaging or digital forensics.

VI. Conclusion and Future Work

This paper presented a comprehensive study of self-supervised denoising using Noise2Void variants and introduced a novel progressive retraining strategy. By sequentially retraining one N2V model on the output of another, we demonstrated that it is possible to achieve deeper noise removal and improved PSNR without requiring clean ground truth or additional architectural changes.

Among the tested variants, Probabilistic N2V achieved the best standalone performance, while its combination with Structured N2V in a retraining cascade yielded significant PSNR improvements—reaching as high as 38.27 dB. This validates the idea that residual noise left by one model can be learned and corrected by another.

A. Future Work

We identify several directions for future research:

- Applying progressive retraining to real-world microscopy datasets with semi-supervised benchmarks.
- Exploring combinations with other self-supervised variants such as Self2Self or Neighbor2Neighbor.
- Extending the pipeline to multimodal biomedical data, e.g., integrating fluorescence and phase contrast images.
- Investigating uncertainty-guided cascades for active learning and model confidence calibration.
- Deploying lightweight retrainable denoisers in edge computing scenarios or low-power biomedical devices.

Our findings reinforce the potential of self-supervised learning for denoising in data-scarce environments and provide a framework for continued improvement through collaborative model pipelines.

References

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007. https://doi.org/10.1109/TIP.2007.901238

[2] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017. https://doi.org/10.1109/TIP.2017.2662206

[3] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, “Noise2Noise: Learning image restoration without clean data,” in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018, pp. 2965–2974. https://arxiv.org/abs/1803.04189

[4] A. Krull, T.-O. Buchholz, and F. Jug, “Noise2Void: Learning denoising from single noisy images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2129–2137. https://doi.org/10.1109/CVPR.2019.00223

[5] A. Krull, T. Vičar, M. Prakash, M. Lalit, and F. Jug, “Probabilistic Noise2Void: Unsupervised content-aware denoising,” *Frontiers in Computer Science*, vol. 2, p. 5, 2020. https://doi.org/10.3389/fcomp.2020.00005

[6] S. Laine, T. Karras, J. Lehtinen, and T. Aila, “High-quality self-supervised deep image denoising,” in *Advances in Neural Information Processing Systems*, vol. 32, 2019. https://arxiv.org/abs/1901.10277

[7] C. Broaddus, A. Krull, M. Weigert, U. Schmidt, and G. Myers, “Removing Structured Noise with Self-Supervised Blind-Spot Networks,” in *Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA, 2020, pp. 159–163. https://doi.org/10.1109/ISBI45749.2020.9098336

[8] Y. Quan, Y. Chen, and H. Ji, “Self2Self with dropout: Learning self-supervised denoising from single image,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1890–1898. https://doi.org/10.1109/CVPR42600.2020.00196

[9] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, “Neighbor2Neighbor: Self-supervised denoising from single noisy images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14781–14790. https://arxiv.org/abs/2101.02824

[10] Z. Wang, J. Liu, G. Li, and H. Han, “Blind2Unblind: Self-supervised image denoising with visible blind spots,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 2027–2036. https://doi.org/10.1109/CVPR52688.2022.00207