

Research Paper



Matched, mismatched and blind: How well does a patch-based neural denoiser trained on six photographs generalise across noise levels and image domains?

Bilal Öztürk*

*Department of Software Engineering, Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey.

Article Info

Article History:

Received: 29 June 2026

Revised: 28 August 2026

Accepted: 18 September 2026

Published: 22 October 2026

Keywords:

Image Denoising

Multilayer Perceptron

Residual Learning

BM3D

Domain Shift

Scientific Imaging



ABSTRACT

Learned denoisers are frequently reported to outperform classical filters, but they are usually trained and tested at the same noise level and on similar images. This study examines what happens when either assumption is broken. A plain multilayer perceptron (MLP) that removes noise from 9×9 image patches by residual learning was trained on only six natural photographs, either at a single noise level (MLP-25, $\sigma = 25$) or over a range of levels (MLP-blind, $\sigma = 5-55$). It was compared with Gaussian, median, bilateral, total-variation (TV), wavelet, non-local means (NLM) and BM3D denoising on eleven held-out real images from three domains: natural photographs, textures and scientific images from microscopy, ophthalmology and astronomy. The classical methods were calibrated on the training images and applied with an automatically estimated noise level. At the matched noise level, MLP-25 was second only to BM3D (29.96 dB versus 31.19 dB mean PSNR) and exceeded NLM by 0.49 dB in 18 of 22 image-seed cases (Holm-adjusted $p = 0.005$). Its advantage transferred to scientific images, where its gap to BM3D (0.99 dB) was smaller than on natural photographs (1.70 dB). Under noise-level mismatch, however, MLP-25 failed: at $\sigma = 50$ it reached only 19.40 dB, 5.15 dB below a simple Gaussian filter, and at $\sigma = 10$ it was worse than the unprocessed noisy input in 7 of 22 cases. Training over a range of noise levels removed this failure at a cost of 0.62 dB at $\sigma = 25$, and MLP-blind was the second-best method at $\sigma = 50$ (25.47 dB). BM3D remained the most accurate method on average in every combination of domain and noise level. Domain shift from photographs to scientific images was therefore much less harmful than noise-level shift, and learned denoisers intended for practical use should be trained blind or paired with reliable noise estimation.

Corresponding Author:

Bilal Öztürk

Department of Software Engineering, Faculty of Engineering, Istanbul Aydin University, Istanbul, Turkey.

Email: bilalo@aydin.edu.tr

Copyright © 2026 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Noise is unavoidable in every imaging system, from smartphone cameras to fluorescence microscopes and telescopes, and denoising is often the first step of an image-analysis pipeline. For two decades the field was led by carefully engineered methods that exploit the redundancy of natural images, among them total-variation regularisation [1], [2], wavelet shrinkage [3], [4], non-local means (NLM) [5] and block-matching 3-D filtering (BM3D) [6]. Chatterjee and Milanfar [7] and Levin and Nadler [8] even asked whether denoising was approaching fundamental limits. [9] Then showed that a plain multilayer perceptron (MLP), trained on a very large number of image patches, could match BM3D, and residual convolutional networks such as DnCNN [10] and FFDNet [11] subsequently surpassed it on standard benchmarks [12].

Two assumptions underlie most of these comparisons. The first is that the noise level at test time equals, or is known to, the level used in training. The second is that test images resemble training images, typically natural photographs. Neither holds reliably in practice. Noise levels vary with exposure, sensor gain and acquisition settings [13], and many users of denoising work with images, such as microscopy, medical or astronomical data, that look nothing like the photographs on which learned denoisers are trained [14], [15]. Benchmarks on real photographs have also shown that rankings obtained on synthetic noise do not always carry over [16], [17].

This paper asks a deliberately practical question: if a small learned denoiser is trained on a handful of photographs, how much does it lose when the noise level or the image domain changes, and how does that compare with the classical methods it is meant to replace? The contributions are as follows:

- A controlled benchmark of seven classical denoisers and two variants of a residual patch-based MLP on eleven held-out real images from three domains (natural photographs, textures and scientific images) at three noise levels, with paired statistical testing.
- Evidence that an MLP trained on only six photographs is the second-best method, after BM3D, when the test noise level matches the training level, and that this advantage transfers to microscopy, retinal and astronomical images.
- Evidence that noise-level mismatch is far more damaging than domain shift: a denoiser trained at one noise level can perform worse than a Gaussian filter, or even worse than no denoising, at other levels, while training over a range of levels largely removes this failure.
- Visual comparisons on real images that illustrate the characteristic artefacts of each method.

2. RELATED WORK

2.1 Classical Denoising

Linear smoothing and the median filter remove noise at the cost of blurring edges and fine detail. The bilateral filter [18] weights neighbours by both spatial and intensity distance and so preserves edges better. Total-variation (TV) denoising [1], efficiently solved by Chambolle's algorithm [2], favours piecewise-constant images. Wavelet shrinkage [3], [19] thresholds coefficients in a sparsifying transform, and BayesShrink [4] adapts the threshold to each subband. Non-local means [5] averages pixels whose surrounding patches are similar, exploiting self-similarity across the whole image, and BM3D [6] combines block matching with collaborative filtering in a 3-D transform domain; it has remained a reference method

for Gaussian noise for more than fifteen years, with later refinements for correlated noise [20]. Learned-dictionary methods [21] and patch priors [22] represent intermediate approaches that learn image statistics from data.

2.2 Learned Denoising

[9] Trained large MLPs on millions of noisy-clean patch pairs and showed that they could compete with BM3D, but also observed that networks trained at one noise level performed poorly at others. Introduced residual learning for denoising [10], in which the network predicts the noise rather than the clean image, following the residual principle [23]; FFDNet [11] added the noise level as an explicit input to handle a range of levels with one network. Self-supervised methods such as Noise2Noise [24] and Noise2Void [15] remove the need for clean training images, and content-aware restoration [14] has brought learned denoising to fluorescence microscopy. [12] Review the rapidly growing literature. Most of these methods are evaluated on standard photographic datasets, and the generalisation of small learned denoisers to other image domains and to mismatched noise levels has received comparatively little systematic attention.

3. METHODOLOGY

3.1 Images

All experiments used real images distributed with the open-source scikit-image [25] and scikit-learn [26] libraries, which document their provenance and licences. Six natural photographs were used for training: an astronaut portrait, a cup of coffee, a cat, a rocket on its launch pad, a Chinese pagoda and a flower. Eleven different images formed the test set, divided into three domains (Figure 1): natural photographs (a photographer with a camera, a tray of coins and a clock), textures (brick, grass and gravel), and scientific images (a microscopy image of a cell, an immunohistochemistry section, a retinal fundus photograph, the Hubble Deep Field and a lunar surface image). Colour images were converted to grey scale, all images were scaled to intensities in $[0, 1]$, images larger than 256 pixels on their shorter side were downsampled with anti-aliasing, and a central 256×256 crop was taken. Because of this cropping, the clock and cell images are dominated by smooth regions, which explains their comparatively high reconstruction quality.

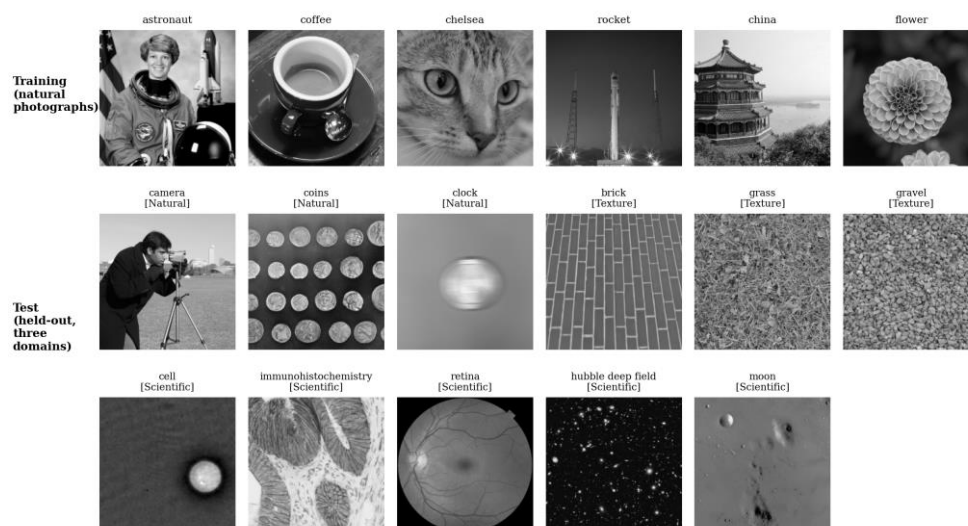


Figure 1. The Real Images Used in the Study After Grey-Scale Conversion and Cropping to 256×256 Pixels. Top Row: the Six Natural Photographs Used to Train the Neural Denoisers. Lower Rows: the Eleven Held-Out Test Images from Three Domains. All Images are Distributed with Scikit-Image and Scikit-Learn

Additive white Gaussian noise with standard deviation $\sigma \in \{10, 25, 50\}$ on the 0–255 intensity scale was added to each test image, with two independent noise realisations per image and level, and the result was clipped to the valid range. This gave 22 image–seed cases per noise level. Reconstruction quality was measured by the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [27] with respect to the clean image.

3.2 Classical Denoisers

Seven classical methods were evaluated: a Gaussian filter, a median filter, the bilateral filter [18], TV denoising with Chambolle’s algorithm [2], wavelet denoising with BayesShrink soft thresholding [4], NLM [5] (5×5 patches, 13×13 search window) and BM3D [6], [20]. To keep the comparison realistic, no method was told the true noise level of a test image. Instead, the noise standard deviation $\hat{\sigma}$ was estimated from each noisy image with the wavelet-based median absolute deviation estimator [19] implemented in scikit-image. The single free parameter of each of the first five methods was calibrated on the six training images by a grid search at each noise level, maximising mean PSNR (Table 1); at test time, the parameter of the noise level closest to $\hat{\sigma}$ was used. Wavelet shrinkage, NLM (whose filtering strength was set as a calibrated multiple of $\hat{\sigma}$) and BM3D received $\hat{\sigma}$ directly. The estimator was accurate on average (mean $\hat{\sigma} = 10.8, 24.4$ and 45.6 for $\sigma = 10, 25$ and 50), although it overestimated the noise on textures and underestimated it at $\sigma = 50$.

Table 1. Parameters of the Classical Denoisers Calibrated on the Six Training Images (Grid Search Maximising Mean PSNR at Each Noise Level)

Method	Parameter	Search Grid	$\sigma = 10$	$\sigma = 25$	$\sigma = 50$
Gaussian filter	Kernel standard deviation (pixels)	0.5–3.0	0.5	1.0	1.5
Median filter	Window size (pixels)	3, 5, 7	3	5	7
Bilateral filter	Intensity σ (spatial $\sigma = 2$ pixels)	0.05–0.30	0.05	0.15	0.30
TV (Chambolle)	Regularisation weight	0.02–0.25	0.02	0.08	0.18
NLM	Filtering strength h as a multiple of $\hat{\sigma}$	0.6–1.2	0.8	0.6	0.6
Wavelet (BayesShrink)	Uses $\hat{\sigma}$ directly	–	–	–	–
BM3D	Uses $\hat{\sigma}$ directly	–	–	–	–

3.3 Patch-Based Neural Denoiser

The learned denoiser follows the plain-MLP approach [9] combined with the residual learning [10] (Figure 2). The network receives a 9×9 noisy patch (81 values), passes it through two fully connected hidden layers of 512 rectified linear units and outputs an 81-value estimate of the noise in the patch, which is subtracted from the input. At inference, the network is applied to every overlapping 9×9 patch of the image and the overlapping estimates of each pixel are averaged. Training data consisted of 120,000 patch pairs drawn at random from the six training photographs and their horizontally flipped and rotated versions. Two models were trained with the Adam optimiser [28] (batch size 256, learning rate 10^{-3} , early stopping on a 10% validation split): MLP-25, trained with $\sigma = 25$ only (40 epochs, 255 s on one CPU core), and MLP-blind, trained with σ drawn uniformly from 5 to 55 for each patch (36 epochs, 242 s). Neither model was given the noise level at test time. The networks were implemented with scikit-learn [26].

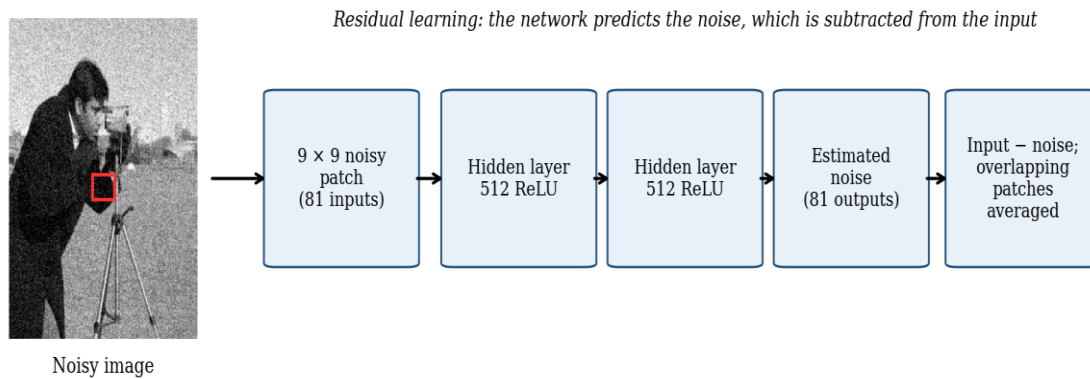


Figure 2. Architecture of the Patch-Based Residual MLP Denoiser. Each Overlapping 9×9 Patch of the Noisy Image is Mapped to an Estimate of Its Noise, which is Subtracted; the Overlapping Clean-Patch Estimates are then Averaged

3.4 Statistical Analysis

At each noise level, the nine denoisers were compared with the Friedman test [29] using the 22 image-seed cases as blocks, and six pre-specified pairs were compared with the Wilcoxon signed-rank test [30] on PSNR, with Holm correction over all 18 tests [31]. The two noise realisations of the same image are not independent with respect to image content, so the tests are interpreted together with effect sizes and the number of cases in which each method was better [32].

4. RESULTS AND DISCUSSION

4.1 Overall Performance

Table 2 reports mean PSNR and SSIM over all eleven test images. BM3D was the most accurate method at every noise level, reaching 35.82, 31.19 and 27.23 dB at $\sigma = 10$, 25 and 50, and it had the best mean rank at each level (1.09–1.18 of 9). Among the other methods, the ranking depended strongly on the noise level. At $\sigma = 25$, the level at which MLP-25 was trained, the network was second (29.96 dB, SSIM 0.808), ahead of NLM (29.47 dB) and TV (29.21 dB). At $\sigma = 10$ and $\sigma = 50$, however, MLP-25 fell to sixth and last place, respectively. MLP-blind was more consistent: fourth at $\sigma = 10$, fifth at $\sigma = 25$ and second at $\sigma = 50$ (25.47 dB). The Friedman test rejected equal performance at every level ($\chi^2 = 102.6$, 147.9 and 127.2; all $p < 10^{-17}$).

Table 2. Mean PSNR (dB) and SSIM over the Eleven Test Images and Two Noise Realisations (22 Cases per Noise Level). The Best Value in Each Column is in Bold

Method	PSNR $\sigma = 10$	PSNR $\sigma = 25$	PSNR $\sigma = 50$	SSIM $\sigma = 10$	SSIM $\sigma = 25$	SSIM $\sigma = 50$
Noisy input	28.25	20.48	14.84	0.659	0.369	0.190
Gaussian	31.13	27.54	24.56	0.780	0.687	0.554
Median	30.60	26.49	24.08	0.810	0.624	0.473
Bilateral	31.51	25.45	21.18	0.831	0.569	0.328
TV	33.49	29.21	24.41	0.885	0.781	0.546
Wavelet	32.66	27.98	24.77	0.865	0.728	0.594
NLM	34.43	29.47	25.12	0.902	0.775	0.566
BM3D	35.82	31.19	27.23	0.928	0.844	0.719
MLP-25	32.14	29.96	19.40	0.847	0.808	0.331
MLP-blind	33.36	29.34	25.47	0.885	0.780	0.620

4.2 Visual Comparison on Real Images

Figure 3 and Figure 4 show the reconstructions of four test images at $\sigma = 25$, with an enlarged region in each panel. On the camera image, BM3D (30.86 dB) and MLP-25 (30.10 dB) scored highest, and TV produced the flattened, piecewise-constant regions typical of total-variation regularisation. On the gravel texture, all methods removed some genuine texture along with the noise; BM3D and the two MLPs retained more of the stone boundaries (24.6–24.9 dB) than TV and NLM (23.8 dB). On the immunohistochemistry section, BM3D and MLP-25 again preserved the fine cellular structure best (27.82 and 27.47 dB). The Hubble Deep Field was the exception among these examples: both MLPs (27.7 dB) were slightly less accurate than TV and NLM (28.0–28.3 dB), and faint galaxies were partly suppressed by every method, illustrating how difficult it is to separate small, low-contrast objects from noise. In this image, the large black background also means that clipping to the valid intensity range alters the noise distribution, which is not represented in the training data.

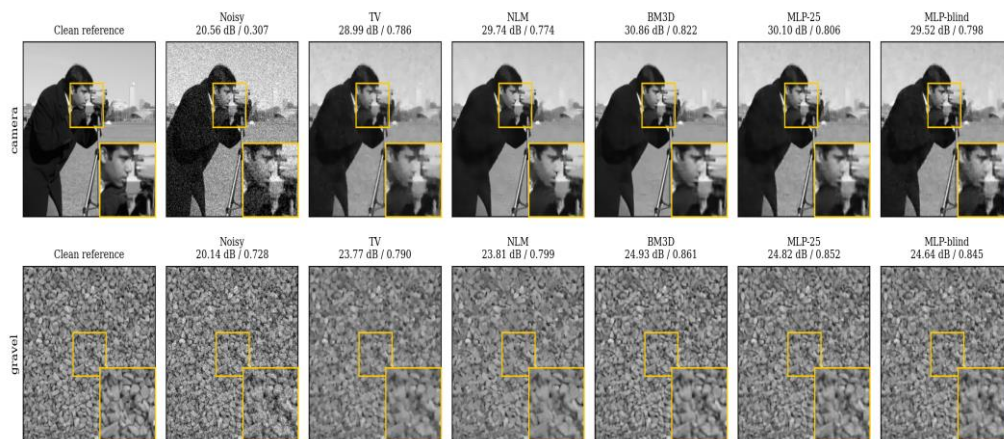


Figure 3. Denoising Results at $\sigma = 25$ on a Natural Photograph (Camera) and a Texture (Gravel). Numbers Give PSNR and SSIM for Each Panel; the Inset Enlarges the Region Marked in Yellow

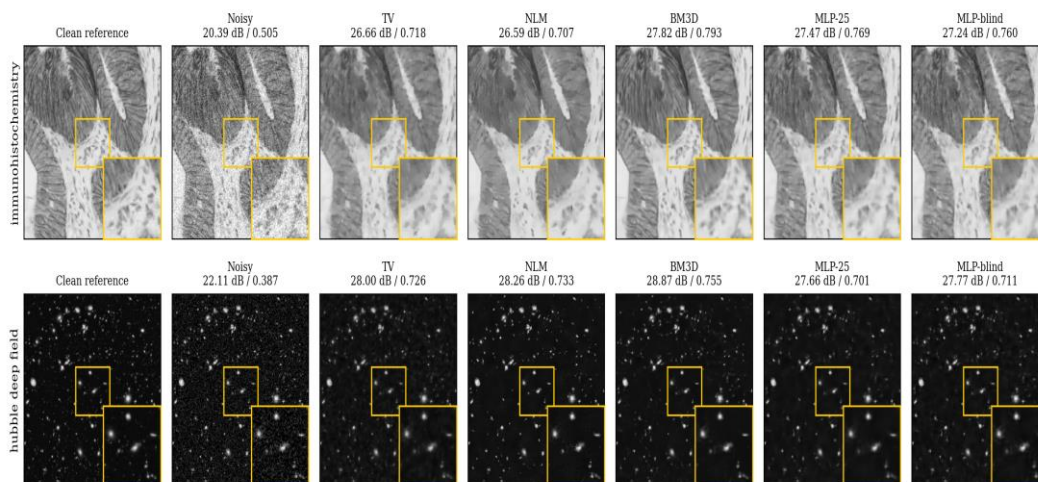


Figure 4. Denoising Results at $\sigma = 25$ on Two Scientific Images: an Immunohistochemistry Section and the Hubble Deep Field. Numbers Give PSNR and SSIM for Each Panel; the Inset Enlarges the Region Marked in Yellow

4.3 Generalisation across Image Domains

Table 3 breaks down PSNR by domain. Textures were the hardest domain for every method, with PSNR 4–6 dB below the other domains at the same noise level, because fine stochastic detail is statistically

similar to noise. The median filter was particularly harmful on textures, where at $\sigma = 10$ it produced images worse than the noisy input (26.52 versus 28.13 dB). Importantly, MLP-25, trained exclusively on natural photographs, did not lose its advantage on scientific images at the matched noise level. At $\sigma = 25$ its gap to BM3D was 1.70 dB on natural photographs, 0.99 dB on scientific images and 1.15 dB on textures, and it outperformed NLM by 0.14, 0.40 and 0.98 dB in the three domains. There was thus no evidence that the shift from photographs to microscopy, ophthalmic and astronomical images degraded the learned denoiser relative to the classical methods.

Table 3. Mean PSNR (dB) by Image Domain and Noise Level. Natural and Texture Domains Contain Three Images and the Scientific Domain Five, Each with Two Noise Realisations

Domain	σ	Gaussian	Median	Bilateral	TV	Wavelet	NLM	BM3D	MLP-25	MLP-Blind
Natural	10	31.31	31.73	32.58	34.55	34.08	36.58	37.87	34.77	34.70
	25	28.15	27.61	26.78	30.85	29.49	31.52	33.36	31.66	31.04
	50	25.27	24.87	21.67	25.34	26.19	27.04	29.33	19.79	26.91
Texture	10	30.01	26.52	28.93	30.61	29.39	30.53	32.40	27.50	29.82
	25	24.57	21.89	22.02	24.81	24.12	25.29	27.42	26.27	25.67
	50	21.89	19.99	19.03	21.70	21.22	21.24	23.62	17.93	22.55
Scientific	10	31.71	32.37	32.42	34.59	33.77	35.48	36.63	33.36	34.67
	25	28.96	28.58	26.70	30.86	29.39	30.75	32.14	31.15	30.53
	50	25.72	26.05	22.17	25.47	26.05	26.30	28.13	20.05	26.36

4.4 The Cost of Noise-Level Mismatch

In contrast to domain shift, a mismatch between training and test noise levels was severe (Figure 5). At $\sigma = 50$, MLP-25 improved the noisy input by only 4.6 dB and was 5.15 dB worse than a simple Gaussian filter, losing to it in 20 of 22 cases; its reconstructions retained a large amount of structured residual noise, because the network had learned to remove noise of a smaller amplitude. At $\sigma = 10$ the failure took the opposite form: the network removed noise of the amplitude it expected, erasing genuine fine detail, and on grass, gravel and immunohistochemistry (and one Hubble realisation) its output was worse than the unprocessed noisy image, 7 of 22 cases in total. Training over a range of noise levels removed both failures. MLP-blind outperformed MLP-25 by 6.07 dB at $\sigma = 50$ (22 of 22 cases) and by 1.22 dB at $\sigma = 10$ (18 of 22), at a cost of 0.62 dB at $\sigma = 25$ (Table 4). It was statistically indistinguishable from NLM at $\sigma = 25$ and $\sigma = 50$ (adjusted $p = 0.78$ and 0.26) and significantly worse than NLM at $\sigma = 10$ (by 1.07 dB), but it never collapsed.

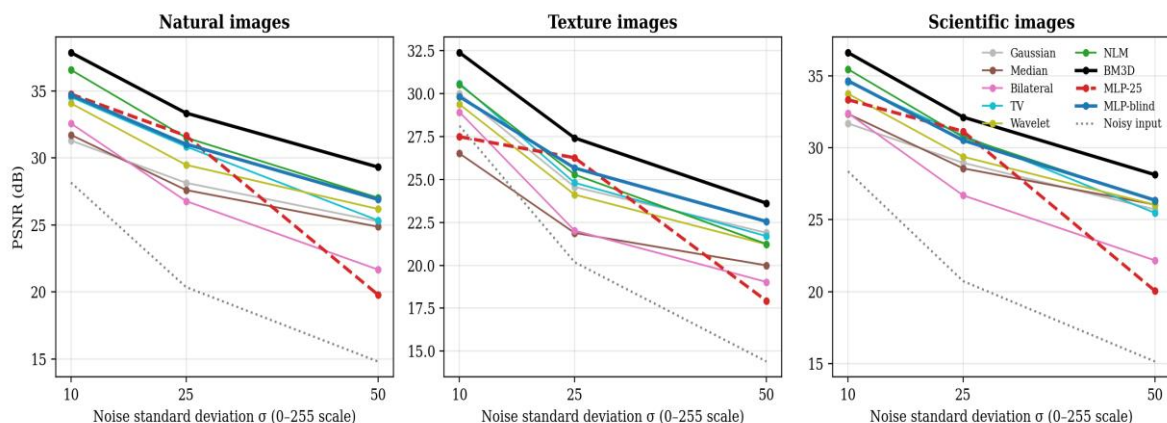


Figure 5. Mean PSNR as a Function of the Noise Level for Each Image Domain. The Dashed Red Curve (MLP-25) Peaks at its Training Noise Level ($\sigma = 25$) and Falls Below Classical Methods Elsewhere, Whereas the Blind Network (Blue) Degrades Gracefully

Table 4. Paired Comparisons of PSNR at Each Noise Level (Wilcoxon Signed-Rank Test over 22 Image-Seed Cases; Holm Correction over 18 Tests)

σ	Comparison (A vs B)	Mean Difference (dB)	Cases A Better	Adjusted p
10	BM3D vs MLP-25	+3.67	22 / 22	< 0.001
	MLP-25 vs NLM	-2.28	2 / 22	< 0.001
	MLP-blind vs MLP-25	+1.22	18 / 22	0.046
	MLP-blind vs NLM	-1.07	8 / 22	0.011
	MLP-25 vs Gaussian	+1.01	12 / 22	0.776
	BM3D vs MLP-blind	+2.46	22 / 22	< 0.001
25	BM3D vs MLP-25	+1.23	20 / 22	< 0.001
	MLP-25 vs NLM	+0.49	18 / 22	0.005
	MLP-blind vs MLP-25	-0.62	2 / 22	< 0.001
	MLP-blind vs NLM	-0.13	8 / 22	0.776
	MLP-25 vs Gaussian	+2.42	22 / 22	< 0.001
	BM3D vs MLP-blind	+1.84	22 / 22	< 0.001
50	BM3D vs MLP-25	+7.83	22 / 22	< 0.001
	MLP-25 vs NLM	-5.72	0 / 22	< 0.001
	MLP-blind vs MLP-25	+6.07	22 / 22	< 0.001
	MLP-blind vs NLM	+0.35	12 / 22	0.256
	MLP-25 vs Gaussian	-5.15	2 / 22	< 0.001
	BM3D vs MLP-blind	+1.75	20 / 22	< 0.001

4.5 Discussion

4.5.1 Principal Findings

A small residual MLP, trained in about four minutes on a single CPU using only six photographs, was the best of the non-BM3D methods at its training noise level and generalised well to scientific images it had never seen. The same network, however, failed badly when the noise level changed, performing worse than a Gaussian filter at $\sigma = 50$ and worse than no denoising on several images at $\sigma = 10$. Training over a range of noise levels made the network robust at a modest cost in matched-level accuracy. BM3D remained the most accurate method on average in every combination of domain and noise level. These results reproduce, at small scale and with real images, the observation of [9] that plain networks can approach BM3D, and they quantify the noise-level sensitivity that motivated noise-level-aware architectures such as FFDNet [11].

4.5.2 Why Noise-Level Shift Hurts More Than Domain Shift

A residual denoiser learns the statistics of the noise more directly than those of the image: it must decide, for each patch, which part of the signal is noise. The amplitude of Gaussian noise is a single global parameter, and a network that has only ever seen one amplitude learns an implicit threshold tied to it; when the true amplitude is larger, it leaves noise behind, and when it is smaller, it removes image detail as if it were noise. Local image structure, by contrast, consists of edges, gradients, blobs and textures that are

shared across photographs, microscopy and astronomy at the 9×9 patch scale. This explains why the network transferred well across domains but not across noise levels, and why textures, whose local statistics resemble noise, were the hardest domain for every method.

4.5.3 Implications for Practice

For users who wish to apply learned denoisers to their own images, four recommendations follow. First, a learned denoiser should either be trained over the full range of noise levels expected in use or be paired with a reliable noise estimate and a model for each level; a single-level model is unsafe when acquisition conditions vary. Second, the concern that denoisers trained on photographs cannot be used on scientific images appears to be less serious than commonly assumed, at least for additive Gaussian noise and small patch-based models, although specialised training remains preferable when data are available [14], [15]. Third, BM3D remains a strong and training-free baseline that should be included in any comparison. Fourth, denoising of textured or low-contrast scientific content, such as faint astronomical sources, should be validated visually, because all methods, learned or classical, suppressed genuine structure in these images.

4.5.4 Limitations

The study used synthetic additive white Gaussian noise; real sensor noise is signal-dependent and spatially correlated [13], [16], [17], and both classical and learned methods may behave differently on it. The test set comprised eleven images, and the domains were represented by three to five images each, so domain-level conclusions should be confirmed on larger collections. The neural denoisers were small fully connected networks trained on six images; deeper convolutional architectures trained on large datasets would be more accurate, and their sensitivity to mismatch may differ. Several calibrated parameters of the classical methods lay at the edge of their search grids (Table 1), so slightly better settings may exist. Finally, PSNR and SSIM [27] do not fully capture perceptual quality or the preservation of scientifically relevant structures.

5. CONCLUSION

This study examined how a patch-based residual MLP denoiser trained on only six photographs generalises across noise levels and image domains, using eleven real test images from photography, texture, microscopy, ophthalmology and astronomy. When the noise level matched training, the network was second only to BM3D (29.96 versus 31.19 dB), outperformed NLM, TV and wavelet shrinkage, and retained this position on scientific images. When the noise level did not match, the same network performed worse than a Gaussian filter at high noise and removed genuine detail at low noise, in some cases producing images worse than the noisy input. Training over a range of noise levels eliminated these failures at a cost of 0.62 dB at the matched level. Noise-level shift was therefore a far greater threat to learned denoising than domain shift. Future work will extend the analysis to real, signal-dependent noise, to convolutional and self-supervised denoisers, and to larger collections of scientific images.

Acknowledgments

The authors thanks the scikit-image and scikit-learn projects for making the test images freely available, the developers of the open-source BM3D implementation, and the researchers whose work is cited in this article.

Funding Information

This research has received no external funding.

Author Contribution Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Bilal Öztürk	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

Conflict of Interest Statement

No conflict of interest.

Informed Consent

Not Applicable.

Ethical Approval

Not Applicable.

Data Availability

This study draws on publicly available images distributed with the scikit-image and scikit-learn libraries. No new datasets were created. The analysis code is available from the corresponding author on reasonable request.

REFERENCES

- [1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992. [doi.org/10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F)
- [2] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004. doi.org/10.1023/B:JMIV.0000011325.36760.1e
- [3] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994. doi.org/10.1093/biomet/81.3.425
- [4] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000. doi.org/10.1109/83.862633
- [5] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005, pp. 60–65. doi.org/10.1109/CVPR.2005.38
- [6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007. doi.org/10.1109/TIP.2007.901238
- [7] P. Chatterjee and P. Milanfar, "Is denoising dead?," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 895–911, 2010. doi.org/10.1109/TIP.2009.2037087
- [8] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011, pp. 2833–2840. doi.org/10.1109/CVPR.2011.5995309
- [9] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2012, pp. 2392–2399. doi.org/10.1109/CVPR.2012.6247952
- [10] 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. doi.org/10.1109/TIP.2017.2662206

- [11] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018. doi.org/10.1109/TIP.2018.2839891
- [12] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," *Neural Networks*, vol. 131, pp. 251–275, 2020. doi.org/10.1016/j.neunet.2020.07.025
- [13] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, 2008. doi.org/10.1109/TIP.2008.2001399
- [14] M. Weigert et al., "Content-aware image restoration: Pushing the limits of fluorescence microscopy," *Nature Methods*, vol. 15, pp. 1090–1097, 2018. doi.org/10.1038/s41592-018-0216-7
- [15] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void – Learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2019, pp. 2129–2137. doi.org/10.1109/CVPR.2019.00223
- [16] T. Plötz and S. Roth, "Benchmarking denoising algorithms with real photographs," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2017, pp. 1586–1595. doi.org/10.1109/CVPR.2017.294
- [17] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2018, pp. 1692–1700. doi.org/10.1109/CVPR.2018.00182
- [18] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Computer Vision*, 1998, pp. 839–846. doi.org/10.1109/ICCV.1998.710815
- [19] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995. doi.org/10.1109/18.382009
- [20] Y. Mäkinen, L. Azzari, and A. Foi, "Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching," *IEEE Transactions on Image Processing*, vol. 29, pp. 8339–8354, 2020. doi.org/10.1109/TIP.2020.3014721
- [21] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006. doi.org/10.1109/TIP.2006.881969
- [22] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. IEEE Int. Conf. Computer Vision*, 2011, pp. 479–486. doi.org/10.1109/ICCV.2011.6126278
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 770–778. doi.org/10.1109/CVPR.2016.90
- [24] J. Lehtinen et al., "Noise2Noise: Learning image restoration without clean data," in *Proc. 35th Int. Conf. Machine Learning, PMLR*, vol. 80, 2018, pp. 2965–2974. proceedings.mlr.press/v80/lehtinen18a.html
- [25] S. van der Walt et al., "scikit-image: Image processing in Python," *PeerJ*, vol. 2, p. e453, 2014. doi.org/10.7717/peerj.453
- [26] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. jmlr.org/papers/v12/pedregosa11a.html
- [27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. doi.org/10.1109/TIP.2003.819861
- [28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learning Representations*, 2015. doi.org/10.48550/arXiv.1412.6980
- [29] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937. doi.org/10.1080/01621459.1937.10503522
- [30] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945. doi.org/10.2307/3001968
- [31] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979. www.jstor.org/stable/4615733

- [32] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," Journal of Machine Learning Research, vol. 7, pp. 1–30, 2006. jmlr.org/papers/v7/demsar06a.html

How to Cite: : Bilal Öztürk (2026). Matched, mismatched and blind: How well does a patch-based neural denoiser trained on six photographs generalise across noise levels and image domains?. Journal of Artificial Intelligence, Machine Learning and Neural Network, 6(2), 134-145. <https://doi.org/10.55529/jaimlnn.61.134.145>

BIOGRAPHIE OF AUTHOR



Bilal Öztürk, is a member of the Department of Software Engineering, Faculty of Engineering, Istanbul Aydın University, Istanbul, Turkey. His research covers image processing, image transmission and compression, and the software engineering of image-based systems. He is interested in the behaviour of image restoration and classification methods when they meet data unlike their training material, and in the practical evaluation of such methods. Email: bilalo@aydin.edu.tr