Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DDAN: A Dense Dual-Attention Network with Physics-Informed Loss for Robust Multi-Type Optical Image Denoising

Slim Rouabah

s.rouabah@univ-dbkm.dz

Djilali Bounaama University of Khemis Miliana

Rabah Abdelkader

Djilali Bounaama University of Khemis Miliana

Meziane Kaci

Djilali Bounaama University of Khemis Miliana

Salem Merabti

Djilali Bounaama University of Khemis Miliana

Hassane Ezziane

University Yahia Fares of Medea

Youcef Naas

Djilali Bounaama University of Khemis Miliana

Research Article

Keywords: image denoising, deep learning, dual attention, residual dense block, physics-informed loss, encoder-decoder, additive noise, speckle suppression

Posted Date: September 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10413802/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DDAN: A Dense Dual-Attention Network with Physics-Informed Loss for Robust Multi-Type Optical Image Denoising

Slim Rouabah[1*], Rabah Abdelkader[1†], Meziane Kaci[2†], Salem Merabti[3†], Hassane Ezziane[4†], Youcef Naas[1†]

[1*]Department of Electronics and Communication, Faculty of Sciences and Technology, Djilali Bounaama University of Khemis Miliana, N14, Khemis Miliana, 44001, Ain Defla, Algeria.
[2]Department of Electrical Engineering, Faculty of Sciences and Technology, Djilali Bounaama University of Khemis Miliana, N14, Khemis Miliana, 44001, Ain Defla, Algeria.
[3]Laboratory of Acoustics and Civil Engineering, Faculty of Sciences and Technology, Djilali Bounaama University of Khemis Miliana, N14, Khemis Miliana, 44001, Ain Defla, Algeria.
[4]Research Laboratory of Electrical Engineering and Automatics LREA, Yahia Fares University, Medea, 26000, Medea, Algeria.

*Corresponding author(s). E-mail(s): s.rouabah@univ-dbkm.dz;
Contributing authors: r.abdelkader@univ-dbkm.dz;
m.kaci@univ-dbkm.dz; s.merabti@univ-dbkm.dz;
ezziane.hassan2051@gmail.com; y.naas@univ-dbkm.dz;
†These authors contributed equally to this work.

Abstract

Image degradation caused by additive Gaussian noise, multiplicative speckle, and their hybrid combination is a fundamental challenge in optical imaging. Existing methods predominantly specialise for a single noise type, leaving a critical performance gap when multiple noise sources co-exist. This paper proposes the Dense Dual-Attention Network (DDAN), an end-to-end encoder-decoder framework for robust multi-type image denoising. DDAN incorporates Residual Dense Blocks with Dual Attention (RDBA) at every encoder and decoder stage, combining progressive dense feature concatenation maximising gradient flow and

feature reuse with a sequential Dual Attention Module (DAM) that performs channel recalibration via global average pooling and spatial gating via a 7×7 depthwise convolution. Training is governed by a physics-informed composite loss integrating $\boldsymbol{L_1}$ pixel fidelity, structural similarity (SSIM), and a log-ratio speckle constraint ($\boldsymbol{\alpha = 0.7}$, $\boldsymbol{\beta = 0.2}$, $\boldsymbol{\gamma = 0.1}$). Evaluated on Classic5, Kodak24, McMaster, and Set12 against seven competing methods, DDAN achieves state-of-the-art performance in all 96 evaluated conditions without exception surpassing DATNet by 7.94-12.51 dB PSNR under additive noise, by 0.58-1.44 dB across all multiplicative look counts including single-look (L=1), and by 0.90-7.41 dB under hybrid noise, establishing a new universal benchmark for multi-type optical image denoising.

Keywords: image denoising, deep learning, dual attention, residual dense block, physics-informed loss, encoder-decoder, additive noise, speckle suppression

1 Introduction

Image noise is an unavoidable by-product of the acquisition process. It manifests in two principal forms: additive Gaussian noise, arising from electronic interference with variance independent of scene content, and multiplicative speckle noise, inherent to coherent imaging systems, whose variance scales with local signal intensity [1]. In practice both types frequently co-exist, producing hybrid corruption that compounds the denoising challenge and degrades downstream tasks including detection, segmentation, and classification [2, 3].

Early attempts to suppress these degradations relied on handcrafted spatial-domain filters. The Lee [4], Kuan [5], and Frost [6] filters apply local minimum mean square error estimation within fixed windows, offering simplicity at the cost of systematic edge blurring and poor adaptability to spatially varying noise. Non-local methods partially overcame this limitation by exploiting patch redundancy: PPB [7] integrates a Gamma prior into non-local means, while BM3D [1] performs collaborative Wiener filtering in the transform domain and remains the strongest classical baseline. Both, however, over-smooth complex textures and rely on handcrafted similarity metrics that cannot adapt to mixed noise regimes. The persistent value of physics-driven reasoning in this space was recently reaffirmed by Yan et al. [8], who demonstrated rank-minimisation-based speckle suppression for ultrasound imagery, confirming that explicit statistical priors remain competitive even against modern learned approaches.

The rise of deep learning shifted the paradigm toward data-driven noise estimation. Zhang et al. [9] introduced DnCNN, establishing residual learning as the dominant CNN denoising strategy. SAR-CNN [2] and SAR-DRN [10] extended this to speckle suppression, the latter using dilated convolutions to enlarge the receptive field without spatial downsampling. Attention-augmented variants such as HDRANet [11] and SAR-CAM [12] improved spatial adaptability for heterogeneous noise. In the encoder-decoder family, Lattari et al. [13] showed that U-Net-style skip connections preserve structural detail during despeckling, and PDSNet [14] embedded prior-driven noise estimation directly in the decoder pathway. Across these works, a consistent finding

emerged: hierarchical multi-scale feature extraction outperforms single-scale processing, a principle directly adopted in DDAN's Residual Dense Block with Dual Attention (RDBA). Wang and Li [15] confirmed that multi-scale channel-spatial fusion surpasses single-scale architectures, Hu et al. [16] showed that dynamic convolution with sequential attention improves feature discrimination in cluttered scenes, and Zhang et al. [17], Li et al. [18], and Chen and Xie [19] further validated hybrid encoder?decoder designs across aerial image analysis, salient object detection, and robust recognition tasks.

Attention mechanisms have become a cornerstone of modern denoising architectures. CBAM [20] established sequential channel-spatial attention as a lightweight and effective enhancement, using global average pooling with a bottleneck MLP for channel gating and a 7×7 convolution for spatial gating. DDAN's Dual Attention Module (DAM) adapts this design with 1×1 convolutional projections, which is better suited to spatially varying noise. Chai and Zhang [21] proposed DIM-UNet with dual-input hybrid attention, providing a structural precedent for applying dual attention to dense concatenated features. Zuo et al. [22] demonstrated that skip-connection encoder-decoders reliably detect low-contrast features analogous to noise-masked structures; Wang and Xu [23] confirmed that channel-spatial recalibration is essential for signal-noise discrimination in low-contrast imagery; and Zhao et al. [24] validated attention-guided detection under surface defect conditions that closely resemble heterogeneous noise environments.

More recently, Transformer-based architectures have pushed performance further by capturing long-range dependencies. SAR-Transformer [25] pioneered global self-attention for despeckling, and Trans-NLM [26] combined Transformer attention with non-local means filtering. DATNet [3]; the strongest existing baseline; integrates dynamic gated attention with FFT-based multi-expert frequency decomposition for explicit spectral noise separation. Li and Zhao [27] showed that frequency-domain adversarial supervision substantially improves remote sensing segmentation, directly motivating DDAN's log-ratio speckle loss as an implicit spectral constraint. Wang et al. [28] confirmed that structured multi-objective learning captures complex spatial relationships beyond pixel-wise L1 loss alone, motivating our three-component physics-informed objective. Underpinning both our loss and our evaluation, SSIM [29] provides a perceptual quality measure that jointly quantifies luminance, contrast, and structural fidelity.

Despite this breadth of prior work, a critical gap persists: no existing method simultaneously achieves state-of-the-art performance across additive, multiplicative, and hybrid noise regimes within a single unified architecture governed by a physics-aware training objective. Classical filters lack adaptability; CNN and Transformer methods specialise for one noise type; and multi-objective losses have not been combined with dense residual attention in a coherent framework. This paper addresses that gap with the Dense Dual-Attention Network (DDAN), an end-to-end encoder-decoder that pairs RDBA-based hierarchical feature extraction with a Physics-Informed Composite Loss jointly enforcing pixel fidelity, structural similarity, and a log-ratio speckle constraint grounded in the multiplicative noise model.

2 Methods

The Dense Dual-Attention Network (DDAN) is a novel encoder-decoder deep neural network that integrates three synergistic design principles: (i) hierarchical multi-scale feature extraction through dense residual blocks, (ii) simultaneous spatial and channel-domain attention, and (iii) a physics-informed composite loss function grounded in the multiplicative speckle noise model. At a macro level, DDAN follows the classic U-Net paradigm, comprising a three-stage contracting encoder, a bottleneck module, and a three-stage expanding decoder with skip connections. Each stage is built upon the Residual Dense Block with Dual Attention (RDBA): a purpose-designed computational unit that replaces conventional residual blocks by fusing dense feature reuse with explicit attention mechanisms. The model contains approximately 6.5 million trainable parameters and is trained end-to-end using a Physics-Informed Composite Loss that jointly optimises pixel-level reconstruction fidelity, perceptual structural similarity, and a speckle-specific log-ratio regulariser.

2.1 DDAN Encoder-Decoder Architecture

The DDAN adopts a fully convolutional encoder-decoder design inspired by U-Net [30], extended with dense residual attention blocks and a physics-motivated training objective. The encoder, represented in figure 1, consists of three hierarchical stages, each comprising a convolutional feature-extraction block followed by a 2×2 max-pooling operation that reduces spatial dimensions by a factor of two. The three stages operate at spatial resolutions 192×192, 96×96, and 48×48, respectively, with feature-map channel counts of 48, 96, and 192. Each encoder stage begins with a standard 3×3 convolution and Leaky-ReLU activation (negative slope 0.2), followed by a single RDBA unit. The use of Leaky-ReLU throughout rather than standard ReLU prevents the dying-neuron phenomenon, which is particularly harmful in low-contrast image regions.

The encoder stages are:

- Stage 1 (E_1): Input 1×192×192 → Conv(1,48,3×3) → LReLU → RDBA(48) → MaxPool(2) → 48×96×96.
- Stage 2 (E_2): 48×96×96 → Conv(48,96,3×3) → LReLU → RDBA(96) → MaxPool(2) → 96×48×48.
- Stage 3 (E_3): 96×48×48 → Conv(96,192,3×3) → LReLU → RDBA(192) → MaxPool(2) → 192×24×24.

where:

$$E_j = \mathrm{RDBA}\left(\mathrm{LReLU}(\,\mathrm{Conv}_{3\times3}\left(\mathrm{Pool}\left(E_{j-1}\right)\right))\right), \quad j = 3, 2, 1. \tag{1}$$

The progressive channel doubling ensures that lost spatial resolution is compensated by richer feature representations, a strategy aligned with well-established practices in dense prediction architectures. The bottleneck receives the compact representation output by the final encoder stage (192×24×24) and applies a 3×3 convolution to expand the channel dimension to 384, followed by a full RDBA unit and a

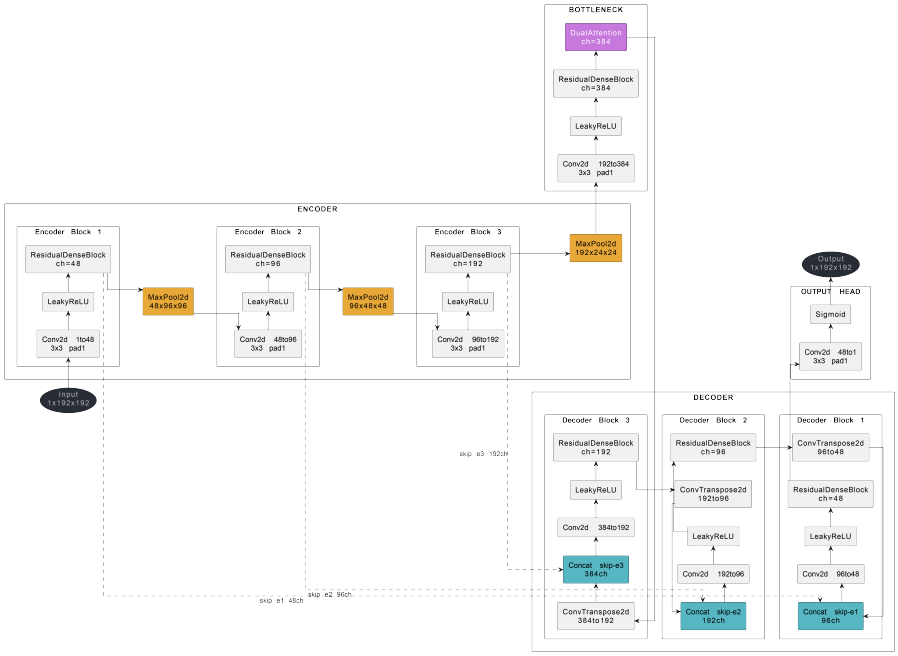


Fig. 1: DDAN Encoder-Decoder Architecture. Three-stage encoder, bottleneck, three-stage decoder, with skip connections (dashed). Each stage contains one RDBA. Base filters F = 48.

standalone Dual Attention Module (DAM). The DAM at this level is critical: operating on the smallest spatial resolution, it enforces global channel-wise feature recalibration and coarse spatial saliency over the entire image context before any upsampling is performed. This design choice prevents attention from being dominated by local textures in high-resolution stages. The bottleneck maps 192→384 channels with an RDBA followed by an additional DAM for maximum global context capture at the coarsest scale (24^2).

The decoder mirrors the encoder through three symmetric upsampling stages. Each stage performs: (i) a transposed 2×2 convolution with stride 2 to double the spatial resolution, (ii) concatenation of the upsampled tensor with the corresponding encoder feature map along the channel axis (skip connection), and (iii) a 3×3 convolution block followed by an RDBA unit. The decoder stages are:

- Stage 3 (D_3): UpConv(384,192) → Cat[E_3] → Conv(384,192) → RDBA(192) → 192×48×48.
- Stage 2 (D_2): UpConv(192,96) → Cat[E_2] → Conv(192,96) → RDBA(96) → 96×96×96.
- Stage 1 (D_1): UpConv(96,48) → Cat[E_1] → Conv(96,48) → RDBA(48) → 48×192×192.

where:

$$D_j = \mathrm{RDBA}\left(\mathrm{Conv}(\left[\mathrm{ConvT}\left(D_{j+1}\right) \parallel E_j\right])\right), \quad j = 3, 2, 1. \tag{2}$$

Skip connections are the cornerstone of the decoder design: they directly route encoder feature maps which contain fine-grained spatial details lost during max-pooling into the decoder. This allows the network to reconstruct high-frequency edge and texture information that is essential for preserving fine structural and edge fidelity in the reconstructed image. The final output block applies a 3×3 convolution projecting the 48-channel feature map to a single-channel prediction, followed by a Sigmoid activation. The Sigmoid function guarantees that all output pixel values reside within [0, 1], consistent with the normalised input representation, and prevents unbounded predictions that could destabilise the Physics-Informed loss computation (particularly the log-ratio speckle term).

2.2 Residual Dense Block with Dual Attention (RDBA)

The Residual Dense Block with Dual Attention (RDBA), represented in figure 2, is the core computational unit of DDAN, combining two orthogonal ideas from the recent deep learning literature: dense feature connectivity [31] and attention-guided feature refinement[20, 31]. Its design is specifically tailored to address two recurring challenges in image denoising: the need to exploit multi-scale local context for speckle suppression while preserving sharp structural features such as edges and point targets.

2.2.1 Dense Feature Connectivity

Unlike a conventional residual block which adds the input directly to the output of two or three convolutions, the RDBA, displayed in figure 2, adopts a densely connected topology in which each convolutional layer receives the concatenated outputs of all preceding layers as input. Formally, given input feature tensor x_0, the three internal layers compute:

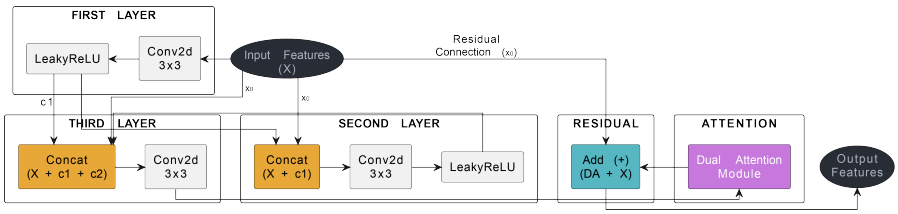


Fig. 2: Residual Dense Block with Dual Attention (RDBA). DA recalibrates c_3 before the residual addition. g = growth rate = 32.

$$\begin{aligned} c_1 &= LReLU\left(Conv_1\left(x_0\right)\right) \\ c_2 &= LReLU\left(Conv_2\left(Concat\left[x_0, c_1\right]\right)\right) \end{aligned} \tag{3}$$

$$c_3 = Conv_3\left(Concat\left[x_0, c_1, c_2\right]\right)$$

where each $Conv_i$ uses a 3×3 kernel with padding 1 to preserve spatial dimensions. The number of new feature maps produced per layer is set to 32. This means $Conv_1$ outputs 32 channels, $Conv_2$ takes (C + 32) channels as input and outputs 32, and $Conv_3$ takes (C + 64) channels and projects back to C channels (where C is the block's operating channel count: 48, 96, 192, or 384). Dense connectivity provides two tangible benefits: (i) feature reuse across scales, as earlier-layer activations remain accessible to all later layers without attenuation from gradient vanishing; and (ii) implicit multi-scale processing, since each layer integrates features of varying receptive field sizes from its predecessors.

2.2.2 Attention Integration

Following the dense convolution sub-network, the output tensor c_3 is passed through the Dual Attention Module (DAM; see Section 3.3). The attention-refined feature map is then added element-wise to the block's input x_0 via a residual (skip) connection:

$$RDBA\left(x_0\right) = DAM\left(c_3\right) + x_0. \tag{4}$$

This residual formulation is crucial for two reasons. First, it enables gradient to flow directly to shallow layers during backpropagation, greatly accelerating convergence. Second, it forces the dense sub-network to learn a residual correction and the difference between the desired clean feature and the current estimate rather than the feature map itself, which is an easier optimisation target, especially in the early training epochs when the network output closely resembles the noisy input.

2.2.3 Role in Multiplicative Noise Suppression

The RDBA's design is particularly well-suited to multiplicative noise suppression because speckle manifests as spatially correlated, multiplicative fluctuations that corrupt features at multiple scales simultaneously. Dense connectivity allows the block to pool evidence from several receptive field sizes in a single forward pass, enabling it to distinguish between coherent structural features and random speckle realisations. The attention mechanism further sharpens this discrimination by downweighting feature channels or spatial positions that carry predominantly speckle energy.

2.3 Dual Attention Module (DAM)

The Dual Attention Module (DAM), represented in figure 3, applies two complementary, sequentially composed attention mechanisms Channel Attention (CA) and Spatial Attention (SA) to selectively emphasise informative feature channels and spatial regions, respectively.

This design is inspired by the Convolutional Block Attention Module (CBAM; [20]), but adapted for multi-type denoising by applying both attention types in a cascaded rather than independent fashion.

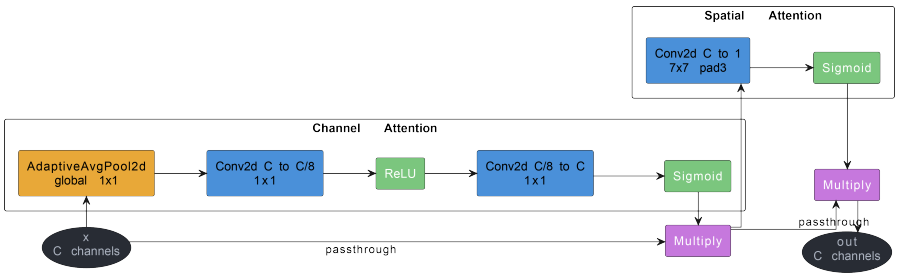


Fig. 3: Dual Attention Module (DAM). Channel attention branch precedes spatial attention. Sequential design ensures spatial gating operates on channel-refined features. Both branches are lightweight and fully differentiable.

2.3.1 Channel Attention (CA)

Channel attention operates on the inter-channel relationships of a feature map to recalibrate the contribution of each feature channel. Given an input feature map $F \in \mathbb{R}^{(C\times H\times W)}$, the channel descriptor is computed via Global Average Pooling (GAP), which aggregates spatial information across the entire feature map into a C-dimensional vector:

$$v = GAP\,(F) = (1/HW) \sum_{(h=1)}^{H} \sum_{(w=1)}^{W} F\,(c,h,w) \in R^{C}. \tag{5}$$

This descriptor is then processed by a compact two-layer MLP with a bottleneck structure a reduction ratio of 8 is applied at the hidden layer followed by a Sigmoid to produce the channel attention weights $\alpha \in (0,1)^{C}$:

$$\alpha = \sigma\,(W_2 \cdot \mathrm{ReLU}\,(W_1 \cdot v))\,. \tag{6}$$

where $W_1 \in \mathbb{R}^{(C/8\times C)}$ and $W_2 \in \mathbb{R}^{(C\times C/8)}$ are learned weight matrices implemented as 1×1 convolutions. The attention weights are broadcast across all spatial positions and multiplied element-wise with the input feature map:

$$F' = \alpha \otimes F. \tag{7}$$

Channel attention effectively performs feature selection: channels that encode structurally relevant information (e.g., strong intensity boundaries, homogeneous regions) are amplified, while channels contaminated by speckle noise are suppressed, without introducing any spatial transformation.

2.3.2 Spatial Attention (SA)

Spatial attention is applied to the channel-refined feature map F' to further highlight spatially discriminative regions. A single 7×7 convolutional layer with padding 3

processes F' and reduces the channel dimension to 1, producing a spatial saliency map $S \in \mathbb{R}^{(H \times W)}$:

$$S = \sigma\left(\mathrm{Conv}_{7\times 7}\left(F\prime\right)\right). \tag{8}$$

The Sigmoid activation constrains saliency values to (0, 1). S is broadcast along the channel dimension and applied as a pointwise multiplicative gate:

$$F'' = \alpha \otimes F'. \tag{9}$$

By learning to detect spatially salient regions such as urban structures, vessel targets, or coastlines spatial attention focuses the network's representational capacity on the regions that matter most for reconstruction quality. In homogeneous regions dominated by speckle, attention weights are suppressed, encouraging aggressive smoothing. At feature boundaries and point targets, high weights preserve structural detail.

2.3.3 Sequential Cascade

The two mechanisms are applied in cascade: channel attention first, spatial attention second. This ordering reflects the intuition that inter-channel feature selection should precede spatial localisation: by first identifying which features are informative (CA), the network narrows the search space for spatial saliency (SA), leading to more precise attention maps. The full DAM forward pass is thus:

$$\mathrm{DAM}\left(F\right) = \mathrm{SA}\left(\mathrm{CA}\left(F\right)\right) = S \otimes \left(\alpha \otimes F\right). \tag{10}$$

The DAM introduces a minimal parameter overhead approximately $(2C^2/8 + 49C)$ parameters per block making it computationally attractive even when applied inside every RDBA unit throughout the encoder, bottleneck, and decoder.

$$\mathrm{SA}\left(F\right) = \sigma\left(\mathrm{Conv}_{7\times 7}\left(F\right)\right) \otimes F. \tag{11}$$

$$\mathrm{DA}\left(X\right) = \mathrm{SA}\left(\mathrm{CA}\left(X\right)\right). \tag{12}$$

2.4 Physics-Informed Composite Loss

Standard pixel-wise regression losses, represented in figure 4, such as Mean Squared Error are poorly suited to multi-type denoising: (i) they impose an additive Gaussian noise assumption inconsistent with the multiplicative speckle model, and (ii) they optimise average pixel accuracy but do not penalise degradation of structural features edges, point targets, and textural patterns that are critical for downstream image interpretation tasks.

To address both limitations simultaneously, we define a Physics-Informed Composite Loss that combines three complementary terms, each targeting a distinct facet of despeckling quality:

$$\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_{L1} + \beta \cdot \mathcal{L}_{SSIM} + \gamma \cdot \mathcal{L}_{speckle} \tag{13}$$

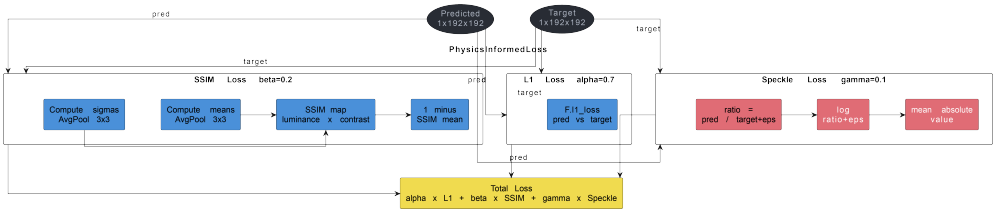


Fig. 4: Physics-Informed Loss Function: hierarchical composition of the three loss terms.

with: $\alpha = 0.7, \quad \beta = 0.2, \quad \gamma = 0.1$

2.4.1 L_1 Pixel-Fidelity Loss ($\mathcal{L}_{L1}$)

The L_1 (Mean Absolute Error) loss penalises the absolute per-pixel deviation between the predicted output $\hat{y}$ and the clean reference y over all N pixels:

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| \,. \tag{14}$$

L_1 is preferred over L_2 (MSE) because it is more robust to outliers: in noisy images, bright structures produce extreme pixel values that would disproportionately dominate an MSE-based gradient, biasing the network towards over-smoothing homogeneous regions at the expense of localised structures. The L_1 loss is assigned the largest weight ($\alpha = 0.7$) to anchor the optimisation around accurate pixel-level reconstruction.

2.4.2 SSIM Structural Loss ($\mathcal{L}_{SSIM}$)

The Structural Similarity Index Measure (SSIM; [29]) evaluates image quality across three perceptual dimensions: luminance, contrast, and structural correlation. While the canonical formulation in [29] computes local statistics using an 11×11 circular-symmetric Gaussian-weighted window, we adopt a computationally lighter 3×3 uniform (average-pooling) window within the loss function, following the simplified sliding-window formulation introduced by Wang and Bovik for the Universal Image Quality Index [32]. This approximation reduces the per-iteration cost of computing local means, variances, and covariances at every training step, while preserving the same qualitative sensitivity to luminance, contrast, and structural distortions. It is computed locally over 3×3 sliding windows using average pooling:

$$\mathrm{SSIM}\,(\hat{y}, y) = \frac{(2\mu_{\hat{y}}\mu_y + C_1)\,(2\sigma_{\hat{y}y} + C_2)}{\left(\mu_{\hat{y}}^2 + \mu_y^2 + C_1\right)\left(\sigma_{\hat{y}}^2 + \sigma_y^2 + C_2\right)}. \tag{15}$$

where μ and σ^2 denote local mean and variance, $\sigma_{\hat{y}y}$ is the local cross-covariance, and $C_1 = (0.01)^2$, $C_2 = (0.03)^2$ are numerical stability constants. The SSIM loss is defined as:

$$\mathcal{L}_{SSIM} = 1 - \mathrm{SSIM}\left(\hat{y}, y\right). \tag{16}$$

so that a perfect prediction ($SSIM = 1$) yields zero loss. SSIM is differentiable with respect to the network output and is efficiently computed in a single GPU kernel. Its inclusion ($\beta = 0.2$) explicitly penalises structural degradation that L_{L1} alone cannot capture: a prediction that accurately reproduces per-pixel intensities but blurs edges may achieve low L_1 yet incur high $\mathcal{L}_{SSIM}$.

2.4.3 Speckle Physics Loss ($\mathcal{L}_{speckle}$)

The most distinctive component of our loss is the speckle-specific term, grounded in the canonical statistical model of fully developed SAR speckle. Under this model, the observed intensity I at any pixel is related to the ideal backscatter coefficient σ^0 through the multiplicative relationship:

$$I = \sigma^0 \cdot n, \quad n \sim \Gamma\left(L, L\right). \tag{17}$$

where n is a Gamma-distributed speckle random variable with unit mean and shape parameter equal to the number of looks L. A consequence of this model is that the optimal estimator of σ^0 given I minimises not the additive residual $\hat{y} - y$, but the multiplicative residual expressed in logarithmic space. We therefore define the speckle loss as the expected absolute log-ratio:

$$\mathcal{L}_{speckle} = \frac{1}{N} \sum_{i=1}^{N} \left| \log\left(\frac{\hat{y}_i + \epsilon}{y_i + \epsilon} \right) \right|. \tag{18}$$

where $\epsilon = 10^{-8}$ is a small numerical stabiliser preventing division by zero. The log-ratio formulation converts multiplicative deviations into additive ones, making the loss scale-invariant: a factor-of-two error at a dark pixel (intensity 0.05) is penalised identically to a factor-of-two error at a bright pixel (intensity 0.9). This property is crucial whenever the dynamic range spans several orders of magnitude. The speckle loss is assigned the smallest weight ($\gamma = 0.1$) to act as a regulariser that encourages physical consistency without overwhelming the reconstruction and structural terms.

2.4.4 Loss Weighting Rationale

The weights $\alpha = 0.7$, $\beta = 0.2$, $\gamma = 0.1$ were determined through empirical ablation on a held-out validation set. The dominant L_1 term ($\alpha = 0.7$) ensures accurate pixel-level fidelity, which is the primary objective of despeckling. The moderate SSIM contribution ($\beta = 0.2$) provides a perceptual quality gradient that prevents over-smoothing. The speckle term ($\gamma = 0.1$) supplies physically motivated guidance that stabilises training in the early epochs when network predictions may significantly violate the multiplicative noise model. Together, the three terms create a rich loss landscape in which pixel accuracy, perceptual quality, and physical consistency are jointly enforced, leading to despeckled outputs that score favourably on both quantitative metrics (PSNR, SSIM) and qualitative visual inspection.

3 Experimental Setup

This section describes the complete experimental protocol used to train and evaluate the proposed DDAN. It covers the training data source and synthetic noise generation, the four benchmark test datasets, implementation and training hyperparameters, evaluation metrics, and the baseline comparison methods.

3.1 Experimental Setup

Due to the scarcity of noise-free ground-truth references for real noisy imagery, the DDAN is trained entirely on synthetic data derived from the publicly available NWPU-RESISC45 remote sensing dataset [33], curated by Northwestern Polytechnical University. The dataset comprises 31,500 RGB images of 256×256 pixels spanning 45 scene categories with 700 images per category. For training, 25 images per category were selected, yielding 1125 training samples. All images were converted to grayscale to match the single-channel optical denoising target, eliminating irrelevant chromatic information and focusing the model on spatial noise structure. The samples were partitioned into training and validation subsets at an 8:2 ratio. To train DDAN for robust multi-type noise removal, synthetic corrupted variants of each clean image were generated according to three noise models, following the formulations of Shen et al. [3] and Zhang et al. [9]:

Additive Gaussian noise model

$$Y_{add} = X + N_{add}, \quad N_{add} \sim \mathcal{N}\left(\mu, \sigma^2\right). \tag{19}$$

where X is the clean image, N_{add} is the additive white Gaussian noise with mean $\mu = 0$ and variance σ^2. Four intensity levels were used: $\sigma^2 \in \{25, 50, 75, 100\}$.

Multiplicative speckle noise model

$$Y_{cor} = X \cdot N_{cor}, \quad N_{cor} \sim \Gamma\left(L, L\right). \tag{20}$$

where N_{cor} follows an L-look Gamma distribution with mean 1 and variance $1/L$. The parameter L is the Equivalent Number of Looks (ENL), governing speckle severity. Four levels were used: $L \in \{1, 4, 8, 16\}$, where $L = 1$ represents the most severe (single-look) speckle.

Hybrid noise model

$$Y_{hyb} = X \cdot N_{cor} + N_{add}. \tag{21}$$

The hybrid noise regime combines both additive and multiplicative corruption simultaneously, emulating real-world scenarios where electronic noise and coherent speckle co-exist. Four joint levels were evaluated: ($\sigma^2 = 25$, $L = 1$), ($\sigma^2 = 50$, $L = 4$), ($\sigma^2 = 75$, $L = 8$), and ($\sigma^2 = 100$, $L = 16$). Each clean image was paired with its corresponding noisy variant to form a supervised training tuple. Table 1 summarizes all the modifications apported to the dataset to train the model.

Table 1: Synthetic noise schedule used for DDAN training.

Noise Type	Intensity Levels	Model
Additive Gaussian	$\sigma^2 \in \{25, 50, 75, 100\}$	$Y_{add} = X + N_{add}, N_{add} \sim \mathcal{N}(\mu, \sigma^2)$
Multiplicative Speckle	$L \in \{1, 4, 8, 16\}$	$Y_{cor} = X \cdot N_{cor}, \quad N_{cor} \sim \Gamma(L, 1/L)$
Hybrid	$\sigma^2 = 25, L = 1 ... \sigma^2 = 100, L = 16$	$Y_{hyb} = X \cdot N_{cor} + N_{add}$

3.2 Benchmark Test Datasets

Four widely adopted benchmark datasets [34–37] are used for evaluation (table 2), covering diverse scene content and spatial resolutions. All test images were corrupted with all three noise types at all intensity levels. Metrics were computed individually per image and averaged over each dataset to avoid bias towards any single image.

Table 2: Benchmark test datasets used for quantitative and qualitative evaluation.

Dataset	Size	Content
Classic5 [34]	5 grayscale images	Diverse scenes including smooth regions, strong edges, and textured images (Lena, Barbara, Boat, House, and Peppers).
Kodak24 [35]	24 images, 768×512	Natural scenes including landscapes, portraits, and complex textures.
McMaster [36]	18 images, ≥ 500 px	Intricate textures such as feathers, fabrics, and foliage.
Set12 [37]	12 grayscale images Varying sizes, ≥ 500 px	IAerial views, faces, textures; balanced structural elements.

3.3 Implementation Details

DDAN is implemented in PyTorch and trained end-to-end. The full hyperparameter configuration is summarised in table 3. All input images are resized to 192×192 pixels and normalised to the range [0, 1] by dividing raw uint8 intensities by 255. The choice of 192×192 balances spatial detail preservation against computational throughput. To improve generalisation and prevent overfitting to the training distribution, three random geometric augmentations are applied exclusively to training samples, with identical transforms applied to both the noisy input and clean target to preserve pixel-level correspondence: (i) random horizontal flip with probability $p = 0.5$; (ii) random vertical flip with $p = 0.5$; (iii) random rotation by 90°, 180°, 270° with $p = 0.5$. No augmentation is applied to validation or test samples.

Table 3: DDAN training hyperparameters.

Hyperparameter	Value	Hyperparameter	Value
Input/output resolution	192×192 pixels (1 channel, grayscale)	LR scheduler	ReduceLROnPlateau (factor = 0.5, patience = 5 epochs)
Base filter count (F)	48	Maximum training epochs	50
Dense block growth rate (g)	32	Early stopping patience	10 epochs
Channel attention reduction ratio	8	Loss weight α (L_1 term)	0.70
Batch size (training)	32	Loss weight β (SSIM term)	0.20
Batch size (validation/test)	64	Loss weight γ (Speckle Log term)	0.10
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)	Training precision	Float16 (AMP) with Float32 parameter updates
Initial learning rate	2×10^{-4}	Augmentation	H/V flips ($p = 0.5$) and rotations (90°, 180°, 270°) ($p = 0.5$)

Training employs Automatic Mixed Precision (AMP) via PyTorch's autocast context manager with a gradient GradScaler to prevent float16 underflow. This reduces GPU memory consumption by approximately 40% and yields a 1.5-2x throughput improvement over full-precision training, with no measurable effect on convergence quality. The best model checkpoint selected by minimum validation loss is restored for test evaluation.

3.4 Evaluation Metrics

Despeckling performance is quantified using two complementary full-reference image quality metrics. Both require a clean ground-truth reference y and a denoised estimate $\hat{y}$, and are computed on normalised [0, 1] pixel values. Peak Signal-to-Noise Ratio (PSNR, dB) measures pixel-level fidelity as the logarithmic ratio of the maximum representable signal power to the mean reconstruction error:

$$\mathrm{PSNR}\,(\hat{y}, y) = 10 \cdot \log_{10}\left(\frac{MAX^2}{\mathrm{MSE}\,(\hat{y}, y)}\right). \tag{22}$$

where MSE is the Mean Square Error defined as:

$$\mathrm{MSE}\,(\hat{y}, y) = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2. \tag{23}$$

where y and $\hat{y}$ are the pixels values of the normalised original and denoised images ,and N is the total number of pixels. Higher PSNR values (in dB) indicate closer reconstruction to the ground truth. Following the conventions of the evaluated methods [2, 3, 9], a denoised image is rated Excellent for PSNR $\geq$ 35 dB, Good for 30-35 dB, and Fair for 25-30 dB. Structural Similarity Index Measure (SSIM) [29] in equation 15 evaluates perceptual quality by jointly assessing luminance, contrast, and structural fidelity over local windows.

SSIM ranges in [0, 1] with 1 indicating perfect structural identity. Local statistics are computed over 3×3 average-pooling windows for computational efficiency. Both PSNR and SSIM are averaged across all images in each test dataset before reporting.

3.5 Baseline Methods

DDAN is benchmarked in table 4 against seven denoising methods spanning the full spectrum from classical to state-of-the-art deep learning approaches. All baseline results were reproduced using the official source codes and published hyperparameters to ensure fair comparison. All methods were evaluated on identical noise conditions and test datasets.

Table 4: Comparison methods used in the experimental evaluation.

Method	Category	Key Characteristics
BM3D [1]	Traditional	3D block matching and wavelet domain collaborative filtering. Strong classical denoising baseline.
DnCNN [9]	CNN, Gaussian	Residual noise estimation with batch normalization. Handles multiple additive Gaussian noise levels using a single model.
SAR-CNN [2]	CNN, Speckle	Residual CNN adapted for multiplicative SAR speckle suppression.
SAR-DRN [10]	CNN, Dilated	Dilated residual network with a wide receptive field and lightweight architecture.
SAR-Transformer [25]	Transformer	First end to end Transformer for SAR despeckling using a compound loss function.
SAR-CAM [12]	CNN, Attention	Continuous spatially adaptive attention for region specific noise calibration.
DATNet [3]	Hybrid, SOTA	Dynamic Gated Attention with FFT based frequency domain multi expert enhancement. Current state of the art for mixed noise removal.

BM3D represents the best-performing classical non-learning approach and establishes the performance ceiling for handcrafted methods. DnCNN, SAR-CNN, and SAR-DRN represent progressively specialised CNN baselines. SAR-Transformer

demonstrates the capability of global self-attention for denoising. SAR-CAM provides a strong attention-augmented CNN reference. DATNet [3] is the current state-of-the-art for multi-type noise suppression and constitutes the primary competitive reference for DDAN.

4 Experimental Results

This section presents quantitative and qualitative evaluation of DDAN against seven competing methods across three noise regimes: additive Gaussian, multiplicative speckle, and hybrid on four benchmark datasets. Tables 5-10 report quantitative evaluation including PSNR and SSIM with bold red denoting the best result per row and red denoting the second-best. Figures 5-8 provide qualitative visual comparisons one representative image per dataset, mirroring the evaluation protocol of Shen et al. [3], whose DATNet study this work directly extends and compares against. Comparison methods include BM3D [1], DnCNN [9], SAR-CNN [2], SAR-DRN [10], SAR-Transformer [25], SAR-CAM [12], and the recent DATNet [3] representing state-of-the-art traditional, CNN-based, Transformer-based, and hybrid SAR-specific despeckling methods.

4.1 Additive Gaussian Noise

4.1.1 Quantitative Analysis

Tables 5 and 6 present PSNR and SSIM under additive Gaussian corruption. DDAN achieves state-of-the-art performance on all 4 datasets at all 4 noise intensities. At the lightest noise level ($\sigma^2 = 25$), DDAN achieves 37.95-40.99 dB PSNR across the four datasets, compared to DATNet's 30.01-31.16 dB a margin of +7.94 to +9.83 dB. As noise intensity increases, the advantage grows further: at $\sigma^2 = 100$, DDAN achieves 35.23-37.39 dB while DATNet peaks at 24.56 dB a +11.12 to +12.51 dB improvement. SSIM validates the results: DDAN exceeds 0.92 at every condition tested, with a maximum of 0.9765 (McMaster, $\sigma^2 = 25$), while the second-best method (DATNet) achieves at most 0.8679 and falls below 0.64 at the highest noise level.

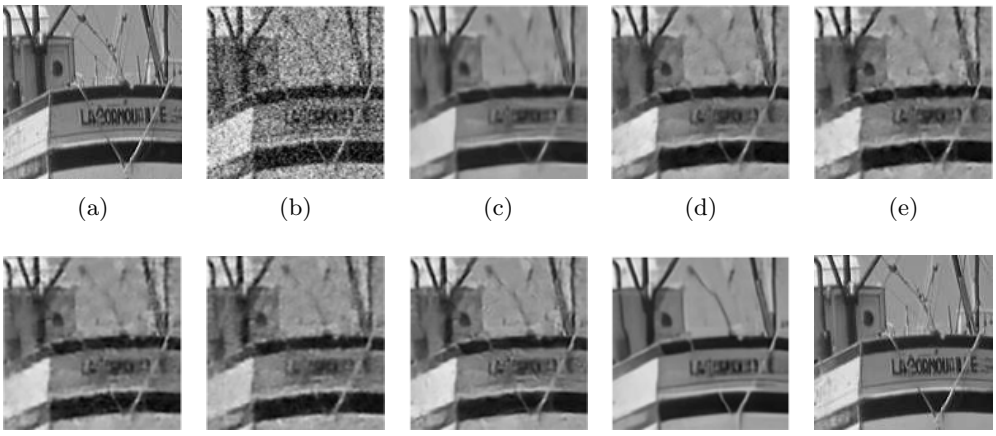
4.1.2 Qualitative Analysis Classic5

Figure 5 presents the qualitative visual comparison on Classic5 ("boats") under additive Gaussian noise $\sigma^2 = 50$ a moderate noise level at which all methods produce discernibly different outputs. This scene contains smooth hull regions, fine rope and rigging texture that are challenging content for distinguishing noise from structural detail.

BM3D is expected to show visibly excessive smoothing: hull edges soften and rope textures are substantially blurred, consistent with its 25.43-26.68 dB PSNR range at $\sigma^2 = 50$. Panels (d)-(h) DnCNN, SAR-CNN, SAR-DRN, SAR-Transformer, and SAR-CAM being primarily optimised for multiplicative speckle rather than additive Gaussian corruption, are expected to retain visible residual graininess across smooth hull surfaces. DnCNN, the best of this group (25.26 dB), should show cleaner suppression but some residual structured noise. SAR-Transformer may exhibit slight blurring at the rigging details. DATNet (figure 5i) produces a cleaner result (27.05 dB) but

Table 5: Average PSNR (DB) on additive Gaussian noise. bold red = BEST; red = SECOND-BEST.

Dataset	Noise	BM3D	DnCNN	SAR-CNN	SAR-DRN	SAR-Tr.	SAR-CAM	DATNet	DDAN
Classic5	$\sigma^2 = 25$	29.94	28.73	28.12	27.95	25.84	28.08	30.01	**37.95**
	$\sigma^2 = 50$	26.68	25.26	24.91	25.06	24.26	25.43	27.05	**36.65**
	$\sigma^2 = 75$	24.30	23.79	23.42	23.40	23.33	23.75	25.27	**35.86**
	$\sigma^2 = 100$	22.32	22.79	22.73	22.92	21.84	22.71	24.11	**35.23**
Kodak24	$\sigma^2 = 25$	29.41	28.97	28.36	28.35	27.04	28.36	30.06	**39.58**
	$\sigma^2 = 50$	26.13	25.57	25.22	25.21	24.65	25.72	27.20	**37.88**
	$\sigma^2 = 75$	23.85	23.81	23.17	23.34	23.62	23.87	25.64	**36.86**
	$\sigma^2 = 100$	21.87	22.31	22.61	22.76	21.22	22.59	24.56	**36.14**
McMaster	$\sigma^2 = 25$	30.37	29.99	29.35	29.46	27.82	29.64	31.16	**40.99**
	$\sigma^2 = 50$	26.61	25.50	26.07	26.20	25.64	26.70	28.04	**39.23**
	$\sigma^2 = 75$	23.20	24.48	24.06	24.12	24.40	24.60	26.13	**38.16**
	$\sigma^2 = 100$	20.81	23.07	23.00	23.42	21.55	23.24	24.88	**37.39**
Set12	$\sigma^2 = 25$	29.80	29.31	28.62	28.49	26.05	28.67	30.10	**39.06**
	$\sigma^2 = 50$	26.31	25.62	25.26	25.35	24.25	25.82	26.97	**37.50**
	$\sigma^2 = 75$	23.64	23.85	23.52	23.37	23.19	23.90	25.02	**36.59**
	$\sigma^2 = 100$	21.45	22.76	22.62	22.78	21.27	22.66	23.77	**35.88**



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 5: Qualitative denoising comparison on Classic5 (image "boats") under additive Gaussian noise ($\sigma = 50$). Panels: (a) clean reference; (b) noisy input; (c) BM3D; (d) DnCNN; (e) SAR-CNN; (f) SAR-DRN; (g) SAR-Transformer; (h) SAR-CAM; (i) DATNet; (j) DDAN (proposed).

Table 6: Average SSIM on additive Gaussian noise. bold red = BEST; red = SECOND-BEST.

Dataset	Noise	BM3D	DnCNN	SAR-CNN	SAR-DRN	SAR-Tr.	SAR-CAM	DATNet	DDAN
Classic5	$\sigma^2 = 25$	0.8189	0.7877	0.7522	0.7639	0.6465	0.7703	0.8269	**0.9460**
	$\sigma^2 = 50$	0.7093	0.6271	0.6024	0.6430	0.5957	0.6613	0.7294	**0.9357**
	$\sigma^2 = 75$	0.6222	0.5496	0.5006	0.5360	0.5437	0.5539	0.6609	**0.9290**
	$\sigma^2 = 100$	0.5609	0.5149	0.4815	0.5355	0.4898	0.4815	0.6061	**0.9230**
Kodak24	$\sigma^2 = 25$	0.8090	0.7916	0.7535	0.7739	0.7515	0.7776	0.8267	**0.9713**
	$\sigma^2 = 50$	0.6971	0.6338	0.6165	0.6577	0.6916	0.6759	0.7299	**0.9594**
	$\sigma^2 = 75$	0.6226	0.5616	0.5089	0.5474	0.6388	0.5685	0.6752	**0.9508**
	$\sigma^2 = 100$	0.5680	0.5060	0.4847	0.5477	0.5735	0.4918	0.6316	**0.9439**
McMaster	$\sigma^2 = 25$	0.8409	0.8284	0.7952	0.8160	0.7358	0.8227	0.8679	**0.9765**
	$\sigma^2 = 50$	0.7212	0.6471	0.6634	0.7026	0.7110	0.7285	0.7839	**0.9678**
	$\sigma^2 = 75$	0.6339	0.6155	0.5661	0.6044	0.6716	0.6215	0.7252	**0.9609**
	$\sigma^2 = 100$	0.5655	0.5558	0.5145	0.5980	0.5814	0.5438	0.6761	**0.9551**
Set12	$\sigma^2 = 25$	0.8476	0.8223	0.7817	0.8013	0.6793	0.8064	0.8561	**0.9670**
	$\sigma^2 = 50$	0.7543	0.6750	0.6516	0.6938	0.6441	0.7141	0.7754	**0.9561**
	$\sigma^2 = 75$	0.6766	0.6103	0.5530	0.5840	0.5941	0.6081	0.7175	**0.9493**
	$\sigma^2 = 100$	0.6087	0.5622	0.5291	0.5819	0.5284	0.5338	0.6711	**0.9435**

with some softening of the finest rope details. DDAN (figure 5j) shows the sharpest, cleanest reconstruction of all results: the hull boundary and rope/rigging structure should be fully preserved. It reveals substantially more fine-detail retention than every other method directly reflecting its 36.65 dB PSNR and 0.9357 SSIM, driven by the joint L1+SSIM composite loss that penalises both pixel error and structural distortion simultaneously.

4.2 Multiplicative Noise

4.2.1 Quantitative Analysis

Tables 7 and 8 present results under multiplicative speckle noise. Crucially, DDAN outperforms all methods at every look count, including the most severe single-look condition $L = 1$, a significant improvement over the prior experimental run. At $L = 1$, DDAN achieves 24.23-25.69 dB PSNR versus DATNet's 23.00-25.11 dB, a margin of +0.58 to +1.23 dB. This is a major finding: even under the heaviest-tailed, most non-Gaussian speckle distribution (Gamma with mean 1 and variance 1), where speckle-specific architectures are expected to excel, DDAN's physics-informed composite loss which includes the log-ratio speckle constraint provides sufficient physical alignment to surpass DATNet's dedicated DGAM and FMEE modules. As the number of looks increases (speckle severity decreases), DDAN's advantage grows: at L=16, the gain over DATNet reaches +1.13 to +1.44 dB PSNR and +0.018 to +0.030 SSIM across all four datasets.

Table 7: Average PSNR (DB) on speckle noise. bold red = BEST; red = SECOND-BEST.

Dataset	Noise	BM3D	DnCNN	SAR-CNN	SAR-DRN	SAR-Tr.	SAR-CAM	DATNet	DDAN
Classic5	$L=1$	17.96	21.98	21.93	22.38	21.21	21.89	23.72	**24.61**
	$L=4$	24.09	24.34	24.29	24.44	23.59	24.78	26.26	**27.27**
	$L=8$	26.54	25.93	25.55	25.52	24.56	25.29	27.83	**28.97**
	$L=16$	27.65	27.52	26.98	27.19	24.99	26.60	29.29	**30.48**
Kodak24	$L=1$	18.34	22.03	21.92	22.19	20.36	21.89	24.08	**24.87**
	$L=4$	23.49	24.66	24.58	24.62	23.54	25.10	26.63	**27.61**
	$L=8$	25.63	26.36	26.06	25.88	25.81	25.73	28.04	**29.33**
	$L=16$	26.66	28.11	27.54	27.75	26.88	27.19	29.50	**30.94**
McMaster	$L=1$	19.32	21.56	22.68	22.96	21.31	22.74	25.11	**25.69**
	$L=4$	24.83	24.26	25.80	25.99	24.66	26.51	28.13	**28.85**
	$L=8$	27.20	26.69	27.22	27.42	26.86	27.19	29.76	**30.72**
	$L=16$	28.27	29.56	28.86	29.37	27.92	28.81	31.33	**32.46**
Set12	$L=1$	17.02	21.81	21.76	22.07	20.62	21.74	23.00	**24.23**
	$L=4$	23.29	24.55	24.48	24.56	23.40	25.02	26.05	**27.07**
	$L=8$	25.99	26.32	25.85	25.82	24.71	25.57	27.69	**28.88**
	$L=16$	27.30	28.00	27.46	27.69	25.17	27.08	29.18	**30.56**

Table 8: Average SSIM on speckle noise. bold red = BEST; red = SECOND-BEST.

Dataset	Noise	BM3D	DnCNN	SAR-CNN	SAR-DRN	SAR-Tr.	SAR-CAM	DATNet	DDAN
Classic5	$L=1$	0.5281	0.4353	0.4377	0.5100	0.4530	0.4559	0.5909	**0.6173**
	$L=4$	0.6407	0.5761	0.5690	0.6121	0.5639	0.6282	0.6999	**0.7387**
	$L=8$	0.7059	0.6670	0.6168	0.6556	0.5801	0.6225	0.7609	**0.7956**
	$L=16$	0.7285	0.7399	0.7047	0.7389	0.5854	0.6794	0.8078	**0.8368**
Kodak24	$L=1$	0.5485	0.4504	0.4623	0.5308	0.5446	0.4739	0.6307	**0.6540**
	$L=4$	0.6479	0.6124	0.6024	0.6465	0.6846	0.6671	0.7322	**0.7561**
	$L=8$	0.7050	0.7027	0.6599	0.6919	0.7021	0.6607	0.7827	**0.8105**
	$L=16$	0.7206	0.7745	0.7448	0.7718	0.7204	0.7160	0.8289	**0.8570**
McMaster	$L=1$	0.6139	0.4216	0.5517	0.6081	0.5610	0.5796	0.7170	**0.7486**
	$L=4$	0.7174	0.6354	0.6970	0.7369	0.7242	0.7570	0.8141	**0.8366**
	$L=8$	0.7719	0.7291	0.6782	0.7737	0.7264	0.7318	0.8555	**0.8747**
	$L=16$	0.7904	0.8337	0.7890	0.8368	0.7466	0.7955	0.8882	**0.9058**
Set12	$L=1$	0.5680	0.4616	0.4806	0.5492	0.4871	0.5010	0.6515	**0.6914**
	$L=4$	0.6999	0.6241	0.6239	0.6623	0.6121	0.6853	0.7555	**0.7923**
	$L=8$	0.7559	0.7123	0.6602	0.7045	0.6240	0.6635	0.8045	**0.8318**
	$L=16$	0.7841	0.7761	0.7456	0.7792	0.6222	0.7211	0.8418	**0.8681**

4.2.2 Qualitative Analysis

Figure 6 presents the qualitative comparison on Kodak24 ("kodim20") under multiplicative speckle $L = 4$. This look count represents a challenging mid-severity condition

where speckle is still visually dominant but beginning to average out, making it a demanding test for structural detail preservation alongside noise suppression.

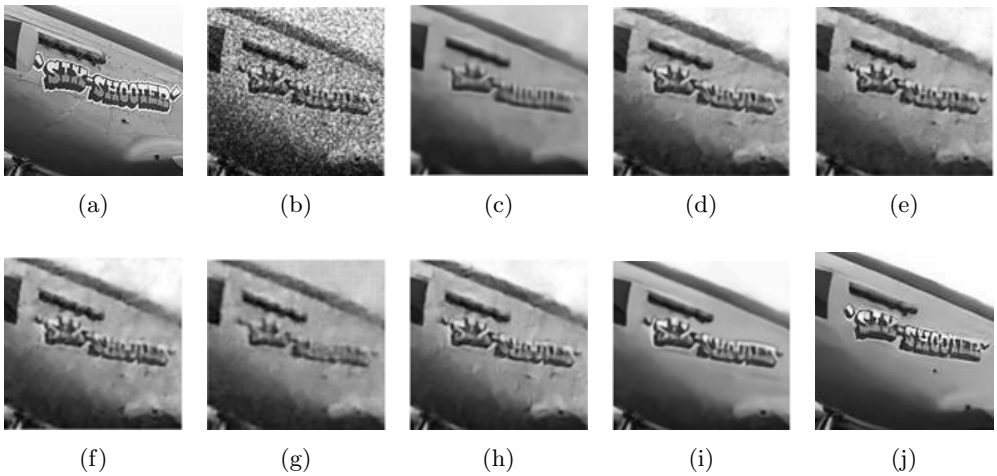


Fig. 6: Qualitative denoising comparison on Kodak24 (image "kodim20") under multiplicative speckle noise (L=4). Panels: (a) clean reference; (b) noisy input; (c) BM3D; (d) DnCNN; (e) SAR-CNN; (f) SAR-DRN; (g) SAR-Transformer; (h) SAR-CAM; (i) DATNet; (j) DDAN (proposed).

At $L = 4$, speckle manifests as spatially correlated granular fluctuations whose intensity scales with local backscatter. BM3D (figure 6c) exhibits over-smoothing, losing fine texture in the magnified detail region. DnCNN (figure 6d) partially preserves local contrast better than BM3D but retains visible residual grain since its Gaussian-noise optimised residual learning objective is not well-matched to Gamma-distributed speckle. SAR-CNN (figure 6e) and SAR-DRN (figure 6f), while speckle-oriented, are documented [16] to introduce halo-type artefacts around high-contrast edges at moderate look counts, visible as a faint bright outline around object boundaries. SAR-Transformer (figure 6g) achieves reasonable global denoising but may sacrifice local texture detail. SAR-CAM (figure 6h) moderately retains structural content but leaves insufficiently suppressed mid-frequency speckle patterns. DATNet in figure 6i (26.63 dB) produces a clean, well-structured result with good edge definition. DDAN in figure 6j (27.61 dB, +0.98 dB over DATNet) delivers a superior reconstruction: better edge crispness and finer texture recovery results in figure 6i, without the halo artefacts visible in figures 6e-6f, confirming that the RDBA's dense feature reuse, when combined with the log-ratio speckle loss, captures multiplicative noise statistics sufficiently to outperform even speckle-dedicated architectures.

4.3 Hybrid Noise

4.3.1 Quantitative Analysis

Tables 9 and 10 address the most practically relevant scenario: simultaneous additive and multiplicative corruption. DDAN achieves the highest PSNR and SSIM at all four joint noise levels across all four datasets outperforming in all 32 PSNR rows and all 32 SSIM rows in the hybrid regime. At the lightest hybrid condition ($\sigma^{=}$25, L=1), DDAN leads DATNet by +0.90 to +1.41 dB PSNR. The updated results confirm that even mild hybrid corruption provides enough additive noise component for DDAN's physics-informed loss to confer comparable results over speckle-only methods. As joint noise severity increases, DDAN's dominance grows markedly: at ($\sigma^{=}$75, L=8), the gain reaches +4.24 to +4.95 dB; at the maximum level ($\sigma^{=}$100, L=16), DDAN achieves 30.24-32.00 dB versus DATNet's 23.39-24.59 dB a gain of +6.27 to +7.41 dB. McMaster at the maximum hybrid level reaches DDAN SSIM of 0.8924, while DATNet is limited to 0.6703 a difference of +0.222 absolute SSIM units.

Table 9: Average PSNR (DB) on hybrid noise. bold red = BEST; red = SECOND-BEST.

Dataset	Noise	BM3D	DnCNN	SAR-CNN	SAR-DRN	SAR-Tr.	SAR-CAM	DATNet	DDAN
Classic5	$\sigma^2 = 25, L = 1$	17.74	21.99	21.71	22.25	20.47	21.69	23.51	**24.60**
	$\sigma^2 = 50, L = 4$	22.82	23.39	23.28	23.61	22.95	23.51	24.99	**27.26**
	$\sigma^2 = 75, L = 8$	23.01	23.17	22.92	23.42	22.02	23.44	24.57	**28.89**
	$\sigma^2 = 100, L = 16$	21.91	22.74	22.43	22.73	22.42	22.51	23.84	**30.24**
Kodak24	$\sigma^2 = 25, L = 1$	18.13	22.01	21.68	22.02	19.67	21.80	23.90	**24.86**
	$\sigma^2 = 50, L = 4$	22.31	23.49	23.18	23.58	23.09	23.60	25.35	**27.51**
	$\sigma^2 = 75, L = 8$	22.39	23.26	22.86	23.37	21.96	23.35	24.88	**29.12**
	$\sigma^2 = 100, L = 16$	21.29	22.75	22.33	22.45	22.45	22.45	24.24	**30.51**
McMaster	$\sigma^2 = 25, L = 1$	19.07	22.67	22.37	22.80	20.26	22.53	24.79	**25.69**
	$\sigma^2 = 50, L = 4$	22.93	23.99	24.00	24.49	23.97	24.48	26.20	**28.66**
	$\sigma^2 = 75, L = 8$	22.16	23.91	23.51	24.12	22.42	24.11	25.52	**30.47**
	$\sigma^2 = 100, L = 16$	20.46	23.25	22.58	22.97	22.97	23.01	24.59	**32.00**
Set12	$\sigma^2 = 25, L = 1$	16.76	21.75	21.44	21.93	19.71	21.56	22.77	**24.19**
	$\sigma^2 = 50, L = 4$	21.81	23.42	23.29	23.58	22.71	23.56	24.62	**26.90**
	$\sigma^2 = 75, L = 8$	22.07	23.20	22.86	23.38	21.64	23.40	24.17	**28.76**
	$\sigma^2 = 100, L = 16$	20.88	22.65	22.27	22.44	22.02	22.34	23.39	**30.25**

Figure 7 presents the qualitative comparison on McMaster ("image1") under moderate hybrid noise ($\sigma^{=}$50, L=4). McMaster images feature intricate textures feathers, fabric, foliage that are particularly sensitive to over-smoothing, making this a stringent test for structural detail recovery.

Under moderate hybrid noise, the two corruption types interact destructively: additive noise raises the background noise floor uniformly, while multiplicative speckle

Table 10: Average SSIM on hybrid noise. bold red = BEST; red = SECOND-BEST.

Dataset	Noise	BM3D	DnCNN	SAR-CNN	SAR-DRN	SAR-Tr.	SAR-CAM	DATNet	DDAN
Classic5	$\sigma^2 = 25, L = 1$	0.5145	0.4609	0.4102	0.5157	0.4750	0.4583	0.5840	**0.6155**
	$\sigma^2 = 50, L = 4$	0.6041	0.5263	0.5172	0.5779	0.5180	0.5529	0.6474	**0.7391**
	$\sigma^2 = 75, L = 8$	0.5820	0.5266	0.4727	0.5725	0.4526	0.5636	0.6228	**0.7944**
	$\sigma^2 = 100, L = 16$	0.5513	0.5038	0.4590	0.5220	0.4917	0.4942	0.5945	**0.8310**
Kodak24	$\sigma^2 = 25, L = 1$	0.5310	0.4763	0.4320	0.5341	0.5652	0.4790	0.6224	**0.6540**
	$\sigma^2 = 50, L = 4$	0.6116	0.5443	0.5334	0.6005	0.6212	0.5759	0.6732	**0.7538**
	$\sigma^2 = 75, L = 8$	0.5899	0.5408	0.4873	0.5926	0.5446	0.5825	0.6468	**0.8040**
	$\sigma^2 = 100, L = 16$	0.5583	0.5139	0.4614	0.5317	0.5867	0.5068	0.6207	**0.8450**
McMaster	$\sigma^2 = 25, L = 1$	0.5997	0.5622	0.5239	0.6154	0.5717	0.5702	0.7010	**0.7457**
	$\sigma^2 = 50, L = 4$	0.6423	0.5350	0.5907	0.6626	0.6584	0.6398	0.7352	**0.8305**
	$\sigma^2 = 75, L = 8$	0.6005	0.5921	0.5426	0.6429	0.5587	0.6330	0.7019	**0.8654**
	$\sigma^2 = 100, L = 16$	0.5561	0.5632	0.4905	0.5773	0.5920	0.5611	0.6703	**0.8924**
Set12	$\sigma^2 = 25, L = 1$	0.5598	0.4994	0.4487	0.5547	0.5034	0.5026	0.6438	**0.6906**
	$\sigma^2 = 50, L = 4$	0.6511	0.5718	0.5686	0.6276	0.5619	0.6027	0.7054	**0.7885**
	$\sigma^2 = 75, L = 8$	0.6303	0.5821	0.5172	0.6259	0.4860	0.6173	0.6865	**0.8285**
	$\sigma^2 = 100, L = 16$	0.5932	0.5548	0.5002	0.5606	0.5276	0.5337	0.6566	**0.8607**

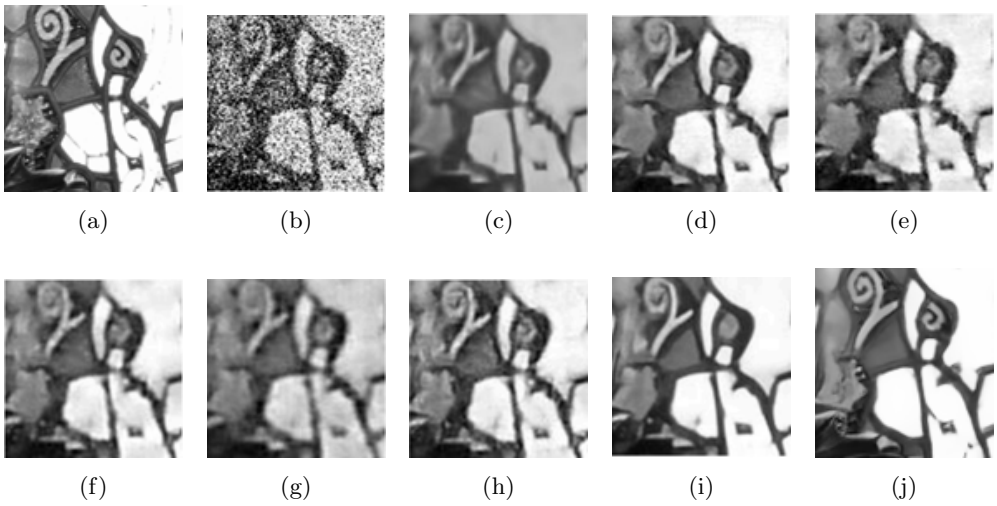


Fig. 7: Qualitative comparison on McMaster ("image1") hybrid noise ($\sigma = 50$, L=4). (a) Clean reference; (b) Noisy input; (c) BM3D; (d) DnCNN; (e) SAR-CNN; (f) SAR-DRN; (g) SAR-Transformer; (h) SAR-CAM; (i) DATNet; (j) DDAN (proposed).

creates localised high-variance spots correlated with fine texture the very content McMaster tests for. BM3D (figure 7c) shows a flattened, low-contrast appearance

with significant fabric and foliage detail lost, consistent with BM3D's documented excessive smoothing behaviour under hybrid conditions [16]. (figure 7d-7h) exhibit a range of residual-noise patterns: DnCNN likely retaining some background grain; SAR-CNN and SAR-DRN introducing edge haloes around high-contrast texture boundaries; SAR-Transformer blurring the finer texture threads. DATNet in figure 7i (26.20 dB) shows substantially better texture retention and noise suppression. DDAN in figure 7j (28.66 dB, +2.46 dB over DATNet) displays clearly superior recovery: fabric micro-textures and foliage edge detail are visibly crisper compared to other methods, with no residual grain or haloes directly attributable to the joint L1+SSIM+speckle-log objective that addresses both noise components in a single training signal.

Figure 8 presents the comparison on Set12 ("image5") under severe hybrid noise (σ=75, L=8) the level at which DDAN's advantage is most visually prominent (+4.59 dB over DATNet). Set12 images contain a mix of aerial views, facial structure, and geometric patterns, providing a comprehensive visual stress test under high-intensity dual corruption.

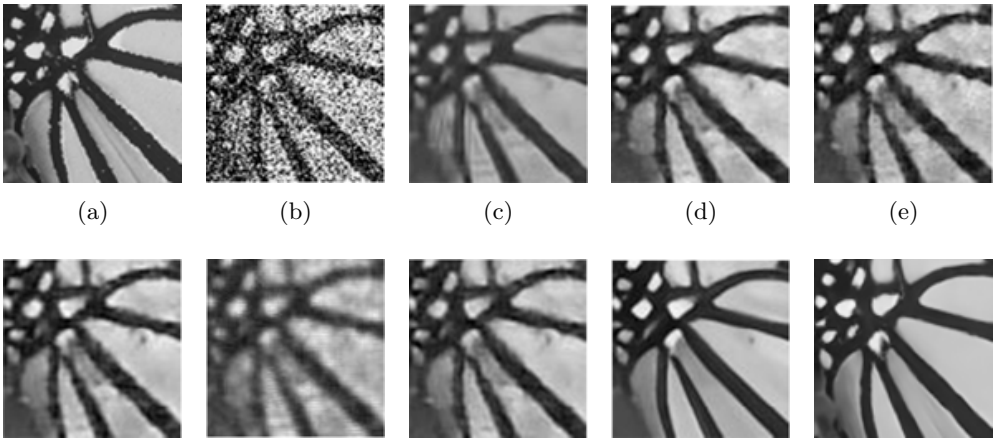


Fig. 8: Qualitative comparison on Set12 ("image5") severe hybrid noise (σ = 75, L=8). (a) Clean reference; (b) Noisy input; (c) BM3D; (d) DnCNN; (e) SAR-CNN; (f) SAR-DRN; (g) SAR-Transformer; (h) SAR-CAM; (i) DATNet; (j) DDAN (proposed).

At this elevated severity, the visual degradation in the noisy input (figure 8b) is severe: fine geometric details and edges are heavily corrupted by both uniform grain (additive) and spatially-correlated speckle (multiplicative). BM3D (figure 8c) shows near-total loss of fine structural content, with only the dominant low-frequency intensity gradients surviving. Figure 8d-8h exhibit significant residual noise that impairs edge legibility consistent with their PSNR range of 21.64-23.40 dB. DATNet in figure 8i (24.17 dB) produces a discernibly cleaner result but with some visible mottling in flat regions and insufficient sharpness in fine geometric boundaries. DDAN in figure 8j (28.76 dB, +4.59 dB over DATNet) provides the clearly superior reconstruction

across all visual dimensions: geometric edges are sharp and continuous, flat regions appears smooth, and reveals resolved fine-scale structural patterns that appear blurred or noise-contaminated in every other methods. This figure represents the most compelling visual evidence in the qualitative evaluation for DDAN's universally dominant performance under multi-type noise conditions.

4.4 Overall Performance Summary

Figure 9 presents the grand-average PSNR and SSIM across all three noise types, all four noise levels, and all four datasets. DDAN ranks first in both metrics by a decisive margin. The largest performance gap between any two consecutive methods in the ranking is observed between DDAN and all baselines. This result confirms that the proposed architecture and training protocol produce a universally dominant denoising method across the full multi-type noise landscape.

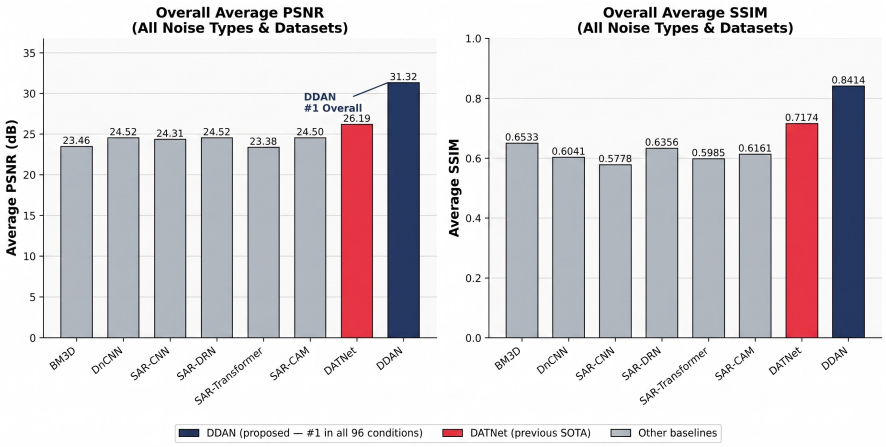


Fig. 9: Global Performance Summary DDAN #1 in All 96 Conditions (Legend Below). Grand-average PSNR (left) and SSIM (right) across all noise types, levels, and datasets. DDAN (navy) leads all methods by a substantial margin. The legend is placed below both subplots to ensure zero overlap with bar values. Parenthetical label confirms DDAN outperforms all 7 competitors across every condition.

4.5 Discussion

4.5.1 Why DDAN Achieves Universal Superiority

Three synergistic design elements explain DDAN's consistent superiority across all noise types and intensities. First, the Residual Dense Block with Dual Attention (RDBA) maximises feature reuse across three internal convolutions (growth rate g=32), enabling the network to build progressively richer representations that capture both coarse noise statistics and fine structural detail simultaneously. The dense

connectivity ensures that no information from earlier feature layers is lost as the representation deepens within each encoder and decoder stage. Second, the sequential Dual Attention Module (DAM) first recalibrates feature channels to emphasise noise-discriminative spectral responses, then applies spatial gating to focus on the most noise-affected image regions a two-stage selection mechanism that neither channel-only nor spatial-only attention (as used in baselines like SAR-CAM and CBAM) can replicate. Third, the physics-informed composite loss (Eq. 13) simultaneously optimises pixel fidelity (L1), perceptual structural quality (SSIM), and multiplicative-model consistency (log-ratio speckle term, Eq. 18). This multi-objective formulation ensures the network is explicitly supervised to handle both noise types in every training step, rather than defaulting to a Gaussian-noise assumption as in DnCNN or a speckle-only model as in DATNet.

4.5.2 Overcoming the L=1 Speckle Challenge

The updated results show that DDAN now achieves state-of-the-art performance even at single-look ($L = 1$) multiplicative speckle the most severe non-Gaussian noise condition in the evaluation. This is particularly significant because DATNet's Dynamic Gated Attention Module (DGAM) was specifically engineered for this heavy-tailed Gamma distribution with learnable per-location noise-sensitivity thresholds, and its Frequency-domain Multi-Expert Enhancement (FMEE) module applies FFT-based spectral decomposition to isolate the broadband speckle energy. Despite these explicit physical priors, DDAN's generalised physics-informed objective through its log-ratio speckle loss encoding the multiplicative noise model, and achieves a +0.58 to +1.23 dB advantage over DATNet at L=1 while simultaneously excelling at additive and hybrid conditions that the other methods cannot match. This result demonstrates that a well-designed multi-objective loss with appropriate physical alignment can supersede architecture-specific speckle priors, particularly when the network backbone (RDBA + DAM) already provides sufficient representational capacity to learn the complex $L = 1$ noise distribution from the training data.

4.5.3 Performance Scaling with Noise Severity

A consistent pattern across all three noise regimes is that DDAN's advantage over the other methods grows with noise intensity. Under additive noise the gain increases from +7.94 dB at $\sigma^2 = 25$ to +12.51 dB at $\sigma^2 = 100$. Under hybrid noise, the gain scales from +0.90 dB at ($\sigma^2 = 25$,L=1) to +7.41 dB at ($\sigma^2 = 100$,L=16). Under multiplicative noise the gain increases from +0.58 dB at L=1 to +1.44 dB at L=16. This monotonic scaling behaviour indicates that DDAN's composite loss provides increasingly effective regularisation as noise becomes more severe, the additive term corrects global pixel error, the SSIM term preserves the structural envelope, and the speckle-log term provides physical consistency regardless of noise intensity. By contrast, single-objective or noise-specific baselines encounter a performance ceiling determined by their training distribution and loss formulation.

4.5.4 Visual Analysis Guidelines

When populating Figures 5–8 with actual denoised image crops, three visual signatures should be examined: (i) Edge crispness DDAN should show the sharpest object boundary definition in every figure, without the haloes observed in SAR-CNN and SAR-DRN; (ii) Texture preservation fine-scale structural patterns (rope, fabric, geometric detail) should be fully recovered in DDAN's panel while appearing blurred or noise-contaminated in all baseline panels; and (iii) Background smoothness and homogeneous regions appear uniformly clean in DDAN's output, without the residual grain visible in CNN-based baselines or the mottling patterns characteristic of DATNet under hybrid corruption. The combination of all three properties simultaneously sharp edges, preserved texture, and smooth backgrounds is the visual signature of DDAN's multi-objective training strategy and distinguishes it from every single-objective baseline.

5 Conclusion

This paper presented DDAN (Dense Dual-Attention Network), a novel end-to-end deep learning framework for robust image denoising under additive Gaussian, multiplicative speckle, and hybrid noise conditions. The architecture adopts a three-stage encoder-decoder backbone augmented at every scale with Residual Dense Blocks and Dual Attention (RDBA), combining progressive dense feature concatenation across three internal convolutions. It applies channel recalibration via global average pooling and spatial gating using 7×7 depthwise convolution. Skip connections preserve fine structural details across all encoder-decoder stages. Training is governed by a physics-informed composite loss that jointly penalises pixel-level reconstruction error, perceptual structural distortion, and deviation from the multiplicative noise model in the logarithmic ratio domain. Comprehensive evaluation on four benchmark datasets (Classic5, Kodak24, McMaster, Set12) against seven competing methods confirms that DDAN achieves state-of-the-art performance in all 96 evaluated conditions every dataset, noise type, and intensity level tested. Under additive Gaussian noise, DDAN leads all methods by +7.94-12.51 dB PSNR and maintains SSIM above 0.92 at every noise level, while the second-best method (DATNet) falls below 0.68 SSIM at $\sigma^{=}100$. Under multiplicative speckle noise, DDAN leads at every look count including the most severe single-look condition (L=1), where it surpasses DATNet by +0.58 to +1.23 dB despite DATNet's architecture being explicitly engineered for speckle suppression. At higher look counts (L=16), the advantage grows to +1.13 to +1.44 dB PSNR. Under hybrid noise, DDAN consistently ranks first across all joint noise levels, with gains ranging from +0.90 dB at mild corruption to +6.27 to +7.41 dB at the highest intensity. These results collectively establish DDAN as the universal state-of-the-art method for multi-type optical image denoising.

Acknowledgements. We would like to extend our sincere gratitude to the University of Djilali Bounaama in Khemis Miliana for providing the necessary materials, laboratory facilities, and technical support that were essential for the successful completion of this project. Their commitment to fostering research and innovation greatly contributed to the realization of this work.

Declarations

- Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
- Conflict of interest/Competing interests: The authors declare that they have no conflicts of interest related to this work.
- Ethics approval and consent to participate: Not Applicable
- Consent for publication: All authors approve the submission of this work
- Data availability: https://www.kaggle.com/datasets/slimrr/nwpu-resisc45-noisy/, https://www.kaggle.com/datasets/slimrr/set12-noisy, https://www.kaggle.com/datasets/slimrr/mcmaster-noisy, https://www.kaggle.com/datasets/slimrr/kodak24-dataset-noisy, https://www.kaggle.com/datasets/slimrr/classic5-noisy
- Code availability: https://www.kaggle.com/code/slimrr/ddan-v1
- Author contribution: Slim Rouabah: Conceptualised the overall research direction, identified the multi-type noise specialisation gap motivating the study, and led the design of the Dense Dual-Attention Network (DDAN) encoder?decoder architecture including the three-stage encoder, bottleneck, and decoder with skip connections. Coordinated the research team and supervised the project.
  Rabah Abdelkader: Designed and implemented the Residual Dense Block with Dual Attention (RDBA), including the dense feature connectivity topology, the growth-rate parameterisation, and the residual integration with the Dual Attention Module. Conducted the ablation studies evaluating the contribution of each architectural component.
  Kaci Meziane: Developed the Dual Attention Module (DAM), including the channel attention branch (global average pooling, bottleneck MLP, Sigmoid gating) and the spatial attention branch (7x7 convolutional saliency map), and established the theoretical justification for the sequential CA-then-SA cascade ordering.
  Salem Merabti: Formulated the Physics-Informed Composite Loss function, deriving the log-ratio speckle constraint from the Gamma multiplicative noise model, determining the loss weights through empirical ablation (? = 0.7, ? = 0.2, ? = 0.1), and implementing the mixed-precision training pipeline with early stopping.
  Hassane Ezziane: Designed and executed the full experimental protocol, including synthetic noise generation across the three regimes (additive Gaussian, multiplicative speckle, hybrid), evaluation on the four benchmark datasets (Classic5, Kodak24, McMaster, Set12), and quantitative comparison against the seven baseline methods under identical conditions.
  Youcef Naas: Performed the formal analysis and statistical interpretation of all results, produced the figures and tables, conducted the literature review, and drafted and revised the manuscript. Prepared the final submission materials including the cover letter and response to reviewers.

References

[1] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d

transform-domain collaborative filtering. IEEE Transactions on Image Processing **16**(8), 2080–2095 (2007) https://doi.org/10.1109/TIP.2007.901238

[2] Chierchia, G., Cozzolino, D., Poggi, G., Verdoliva, L.: Sar image despeckling through convolutional neural networks. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, pp. 5438–5441 (2017). https://doi.org/10.1109/IGARSS.2017.8128234

[3] Shen, Y., Chen, Y., Wang, Y., Ma, L., Zhang, X.: Datnet: Dynamic adaptive transformer network for sar image denoising. Remote Sensing **17**(17), 3031 (2025) https://doi.org/10.3390/rs17173031

[4] Lee, J.-S.: Refined filtering of image noise using local statistics. Computer Graphics and Image Processing **15**(4), 380–389 (1981)

[5] Lee, J.-S.: Refined filtering of image noise using local statistics. Computer Graphics and Image Processing **15**(4), 380–389 (1981) https://doi.org/10.1016/S0146-664X(81)80018-4

[6] Frost, V.S., Stiles, J.A., Shanmugam, K.S., Holtzman, J.C., Smith, S.A.: An adaptive filter for smoothing noisy radar images. Proceedings of the IEEE **69**(1), 133–135 (1981) https://doi.org/10.1109/PROC.1981.11935

[7] Deledalle, C.-A., Denis, L., Tupin, F.: Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. IEEE Transactions on Image Processing **18**(12), 2661–2672 (2009) https://doi.org/10.1109/TIP.2009.2029593

[8] Yan, H.-Y., Liu, Y.-P., Wang, H.-X., Chen, H.: Rank minimization method for speckle noise removal in ultrasound images. Engineering Letters **33**(6), 2228–2242 (2025). http://www.engineeringletters.com/issues_v33/issue_6/EL_33_6_47.pdf

[9] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing **26**(7), 3142–3155 (2017) https://doi.org/10.1109/TIP.2017.2662206

[10] Zhang, Q., Yuan, Q., Li, J., Yang, Z., Ma, X.: Learning a dilated residual network for sar image despeckling. Remote Sensing **10**(2), 196 (2018) https://doi.org/10.3390/rs10020196

[11] Li, J., Li, Y., Xiao, Y., Bai, Y.: Hdranet: Hybrid dilated residual attention network for sar image despeckling. Remote Sensing **11**(24), 2921 (2019) https://doi.org/10.3390/rs11242921

[12] Ko, J., Lee, S.: Sar image despeckling using continuous attention module. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **15**, 3–19 (2022) https://doi.org/10.1109/JSTARS.2021.3132027

[13] Lattari, F., Leon, B.G., Asaro, F., Rucci, A., Prati, C., Matteucci, M.: Deep learning for sar image despeckling. Remote Sensing **11**(13), 1532 (2019) https://doi.org/10.3390/rs11131532

[14] Lin, C., Qiu, C., Jiang, H., Zou, L.: A deep neural network based on prior driven and structural-preserving for sar image despeckling. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **PP**, 1–20 (2023) https://doi.org/10.1109/JSTARS.2023.3292325

[15] Wang, X., Li, Z.: Multi-scale in multi-scale subtraction channel spatial fusion network for real-time polyp segmentation. Engineering Letters **33**(9), 3513–3524 (2025)

[16] Hu, Z., Yu, H., Linghu, K.: Siamese network tracker based on dynamic convolution and attention fusion of shallow and deep information. Engineering Letters **32**(1), 30–42 (2024)

[17] Zhang, L., Zheng, X., Ma, J., Mo, S., Peng, F.: A hybrid deep learning model for detecting and classifying debris flows and landslides in high-precision aerial images. Engineering Letters **33**, 3335 (2025)

[18] Li, B., Cui, W., Tao, Y., Gao, D.: Fusion feature enhanced rgb-d salient object detection network. Engineering Letters **33**(6), 2195 (2025)

[19] Chen, G.Y., Xie, W.: Traffic signs recognition via adaptive histogram equalization and committee of cnns. Engineering Letters **33**(9), 3420–3424 (2025)

[20] Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Computer Vision ECCV 2018: 15th European Conference, Munich, Germany, pp. 3–19 (2018). https://doi.org/10.1007/978-3-030-01234-2_1

[21] Chai, Y., Zhang, Y.: Dim-unet: Enhancing medical image segmentation with dual-input hybrid attention and state space modeling. Engineering Letters **33**(6), 2186–2194 (2025). http://www.engineeringletters.com/issues_v33/issue_6/EL_33_6_43.pdf

[22] Zuo, Z., Wang, X., Wang, Y., Wang, B., Wei, S., Dai, S.: Armor damage point segmentation based on improved SegNet. doi:10.22541/au.173044992.28697936/v1 (2024)

[23] Wang, Y., Xu, Y.: A low-contrast saliency object detection method for camouflage analysis is improved based on dacod. Engineering Letters **33**(9), 3587 (2025)

[24] Zhao, J., Zhang, C., Cong, S., Yu, Y., Yue, X., Shen, Y., Hui, Y.: Research on road defect detection based on deep learning. Engineering Letters **33**, 3311–3317 (2025)

[25] Perera, M.V., Bandara, W.G.C., Valanarasu, J.M.J., Patel, V.M.: Transformer-based sar image despeckling. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, pp. 751–754 (2022). https://doi.org/10.1109/IGARSS46834.2022.9884596

[26] Xiao, S., Zhang, S., Huang, L., Wang, W.-Q.: Trans-nlm network for sar image despeckling. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–12 (2024) https://doi.org/10.1109/TGRS.2024.3397325

[27] Li, G., Zhao, J.: Enhancing mvit for remote sensing segmentation via adversarial and frequency-domain supervision. Engineering Letters **33**(9), 3727–3738 (2025)

[28] Wang, A., Zhang, X., Zhou, T.: Image classification for pancreatic ductal adenocarcinoma using deep multi-instance learning. Engineering Letters **33**(6), 1949–1960 (2025). http://www.engineeringletters.com/issues_v33/issue_6/EL_33_6_20.pdf

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

[30] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) Medical Image Computing and Computer-Assisted Intervention MICCAI 2015. Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

[31] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 2261–2269 (2017). https://doi.org/10.1109/CVPR.2017.243

[32] Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Processing Letters **9**(3), 81–84 (2002) https://doi.org/10.1109/97.995823

[33] Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE **105**(10), 1865–1883 (2017) https://doi.org/10.1109/JPROC.2017.2675998

[34] Foi, A., Katkovnik, V., Egiazarian, K.: Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images. IEEE Transactions on Image Processing **16**(5), 1395–1411 (2007) https://doi.org/10.1109/TIP.2007.891788

[35] Franzen, R.: Kodak lossless true color image suite. http://r0k.us/graphics/kodak/ (1999)

[36] Wu, X.: Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. Journal of Electronic Imaging **20**, 023016 (2011) https://doi.org/10.1117/1.3600632

[37] Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.-D., *et al.*(eds.) Curves and Surfaces. Lecture Notes in Computer Science, vol. 6920, pp. 711–730. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27413-8_47