Indian Engineering Journal
ISSN: 2984-9047
Vol 4 Issue 1 (2026)

DRCDNet:Adaptive Deep Residual CNN for Real-Time Blind Denoising and Quality Enhancement

Mrs. T D JAIN SUBA[1], Dr.S.Sugantha Priya[2]

[1]Research Scholar, Department of Computer Science, Dr SNS Rajalakshmi College of Arts and Science, Coimbatore, jainsuba1980@gmail.com. Mobile Number: 9442825508

[2]Assistant Professor, Department of Computer Science, Dr SNS Rajalakshmi College of Arts and Science, Coimbatore, sugantha2612@gmail.com

Abstract

The proliferation of digital imaging systems in medical diagnostics, autonomous driving, surveillance and consumer photography has intensified the demand for robust real-time image denoising and quality enhancement frameworks. Conventional denoising approaches including Gaussian filtering, non-local mean and BM3D are constrained by fixed noise assumptions and prohibitive computational costs in real-time scenarios. To overcome these limitations, this paper proposes a Deep Residual Convolutional Denoising Network that leverages hierarchical feature extraction residual dense blocks and channel attention mechanisms to achieve superior noise suppression while preserving fine-grained structural details. The proposed DRCDNet incorporates a multi-scale encoder-decoder backbone augmented with squeeze andexcitation attention modules and adaptive noise level estimation to handle spatially variant noise distributions. Extensive evaluations on benchmark datasets including BSD68and Urban100 demonstrate that DRCDNet achieves state oftheart Peak Signal-to-Noise Ratio and Structural Similarity Index values across Gaussian, Poisson, and real-world noise conditions, while maintaining real-time inference throughput on standard GPU hardware.

Keywords*:* Image Denoising, Deep Convolutional Neural Networks, Residual Learning, Attention Mechanisms, Real-Time Enhancement, Blind Denoising.

I. Introduction

Digital image acquisition inevitably introduces various forms of noise arising from sensor imperfections, photon shot noise, thermal interference, and quantization artifacts. In safety-critical domains such as medical imaging and autonomous navigation, unmitigated noise severely degrades downstream analysis accuracy and poses tangible operational risks [1]. Consequently, high-fidelity image denoising and quality enhancement have emerged as foundational preprocessing stages in modern computer vision pipelines.Traditional denoising methods such as Non-Local Means and Block-Matching and 3D Filtering achieve competitive noise removal by exploiting self-similarity and sparse representations in transform domains. However, these approaches are computationally expensive and assume stationary, homogeneous noise statistics that rarely reflect real-world degradation patterns. Furthermore, their inability to learn adaptive, data-driven features limits generalization across heterogeneous imaging conditions [2].

The advent of deep learning and specifically deep Convolutional Neural Networks has fundamentally transformed the image restoration landscape. Pioneer architectures such as DnCNN demonstrated that residual learning could effectively separate latent clean images from corrupted observations by predicting the noise component [3]. Subsequent developments incorporating dense connectivity, attention mechanisms and generative adversarial training have progressively elevated denoising performance to near-perceptual quality levels.Despite these advances, deploying deep denoising networks in latency-sensitive applications remains a formidable engineering challenge. Most high-performing architectures prioritize reconstruction fidelity at the expense of inference speed rendering them impractical for embedded platforms, edge devices and streaming video pipelines [4]. Achieving the dual objectives of state-of-the-art image quality and real-time processing necessitates careful architectural co-design that balances representational capacity with computational efficiency.

A parallel challenge concerns the nature of noise itself. Real-world image degradation is rarely confined to a single noise type or a spatially uniform distribution. Practical imaging systems routinely exhibit mixed-noise regimes combining Gaussian readout noise, Poisson-distributed photon shot noise, and structured sensor artifacts that vary across spatial regions and imaging conditions [5]. Models trained exclusively on additive white Gaussian noise exhibit significant performance degradation when deployed on heterogeneous or unknown noise distributions, underscoring the critical need for blind denoising frameworks capable of generalizing without explicit noise level supervision.Attention mechanisms have emerged as a powerful tool for selectively amplifying informative feature representations while suppressing irrelevant activations. Squeeze-and-Excitation networks introduced channel-wise recalibration by modeling global interdependencies between feature channels, enabling networks to focus computational resources on structurally discriminative responses. When embedded within denoising architectures, such mechanisms allow the network to dynamically distinguish between noise-correlated and signal-correlated channels, thereby improving restoration fidelity without a proportional increase in model complexity.

Multi-scale feature aggregation has equally demonstrated its importance in image restoration tasks. Noise corruption manifests at multiple spatial frequencies simultaneously, with fine-grained texture degradation occurring at high frequencies and contrast distortions affecting coarser spatial structures. Encoder-decoder architectures with skip connections have proven effective in capturing complementary representations at different resolutions, enabling simultaneous recovery of both local texture details and global structural coherence within a single unified forward pass. The integration of frequency-domain supervision into training objectives represents another recent advancement in perceptual image restoration. Standard pixel-wise losses such as mean absolute error tend to over-smooth high-frequency details, producing restorations that are quantitatively favorable yet visually blurred. Incorporating spectral constraints through Fast Fourier Transform-based loss terms penalizes frequency-domain distortions explicitly, encouraging the network to preserve fine structural periodicity and edge sharpness that are otherwise suppressed by purely spatial objectives.

To address these challenges, this work proposes the Deep Residual Convolutional Denoising Network, a unified framework that integrates residual dense blocks, multi-scale feature aggregation, and squeeze-and-excitation channel attention to achieve high-quality blind denoising in real time. The proposed architecture supports spatially adaptive noise estimation, enabling robust generalization to synthetic and real-world noise distributions without prior noise level knowledge. The main contributions of this work are summarized as follows: (i) a unified end-to-end blind denoising architecture combining dense residual connectivity with channel attention and multi-scale encoding; (ii) a lightweight parallel noise estimation branch that conditions the denoising computation on spatially varying noise severity; and (iii) a composite training objective integrating pixel-level, perceptual, and frequency-domain losses that jointly optimize reconstruction fidelity and structural preservation across diverse noise regimes.

II. Related Works

The domain of learning-based image restoration has witnessed exponential growth over the past decade, driven by breakthroughs in network depth, skip-connection strategies and self-supervised training paradigms.

Zhang et al. [3] introduced DnCNN, which established residual learning as the de facto paradigm for blind Gaussian denoising by training a single network across a range of noise levels. The architecture demonstrated that batch normalization, coupled with residual connections, substantially accelerates convergence and improves generalization, outperforming classical methods by significant PSNR margins on standard benchmarks.

Mao et al. [5] proposed the Red-Net architecture employing symmetric skip connections between encoding and decoding convolutional layers, facilitating multi-resolution feature reuse and effectively recovering high-frequency textural information lost during aggressive downsampling. This design philosophy influenced subsequent encoder–decoder denoising frameworks including U-Net variants adapted for image restoration tasks.

Lefkimmiatis [6] introduced NLNet and UNLNet by embedding non-local self-similarity priors as differentiable modules within the CNN pipeline, marrying the expressiveness of deep networks with the structural inductive biases of patch-based methods. The resulting hybrid achieved marked improvements on structured noise removal and detail preservation over purely learned baselines.

Attention-augmented denoising was pioneered by Zhang et al. [7] with RCAN, which introduced channel attention to image super-resolution and was subsequently adapted for restoration tasks. The squeeze-and-excitation mechanism enables dynamic recalibration of feature channel responses, emphasizing noise-informative representations while suppressing redundant activations, leading to consistent quality gains across diverse corruption types.

More recently, transformer-based architectures such as Restormer [8] have leveraged self-attention over local windows to capture long-range spatial dependencies neglected by convolutional models. Although transformers achieve top-tier restoration metrics, their quadratic complexity with respect to spatial resolution imposes substantial memory and latency overheads that currently preclude deployment in real-time scenarios without aggressive pruning or knowledge distillation strategies.

It is observed that existing CNN-based denoisers prioritize reconstruction quality or model compactness in isolation, with limited work systematically addressing the joint optimization of image fidelity and inference throughput under blind, spatially variant noise conditions. This motivates the proposed DRCDNet, which is designed to bridge this performance–efficiency gap through principled architectural innovations.

III. Existing Methodology

A. DnCNN: Beyond Gaussian Denoiser

DnCNN [3] is a landmark feed-forward CNN for image denoising that adopts a residual learning strategy, training the network to predict the residual noise rather than the clean image. The architecture stacks seventeen to twenty convolutional layers with batch normalization and ReLU activations, leveraging implicit regularization to handle a wide range of Gaussian noise levels with a single model. DnCNN demonstrated for the first time that a plain deep CNN could surpass BM3D on PSNR metrics while operating at significantly higher speeds, establishing residual prediction as the standard training objective for subsequent restoration networks.

B. FFDNet: Flexible Fast Denoising Network

FFDNet [9] extended the blind denoising capability of DnCNN by introducing an explicit noise level map as a network input, enabling user-controlled and spatially adaptive denoising. The architecture processes down-sampled sub-images in parallel, effectively quadrupling the receptive field without proportional computational cost and substantially accelerating inference. FFDNet demonstrated that conditioning on a tunable noise map endows the model with flexible denoising strength, outperforming DnCNN on spatially variant and real-world noise scenarios while maintaining real-time throughput on GPU hardware.

Although DnCNN and FFDNet represent seminal advances in learned image denoising, their architectural designs are limited by shallow feature reuse, absence of multi-scale processing, and rudimentary channel interaction modeling. These constraints motivate the development of the proposed DRCDNet, which incorporates residual dense connections, multi-scale feature aggregation, and squeeze-and-excitation attention to achieve superior reconstruction fidelity and generalization across heterogeneous noise distributions.

IV. Proposed Methodology

The proposed Deep Residual Convolutional Denoising Network (DRCDNet) is a unified end-to-end architecture designed to achieve high-fidelity blind image denoising under real-time latency constraints. DRCDNet integrates three synergistic components: Residual Dense Blocks (RDB) for hierarchical feature extraction,

Squeeze-and-Excitation channel attention for adaptive feature recalibration and a multi-scale encoder-decoder backbone for capturing noise-corrupted spatial structures at complementary resolutions. The overall network takes a noisy image as input and directly produces the denoised reconstruction without requiring explicit noise level estimation.

A. Residual Dense Block

Each Residual Dense Block concatenates the outputs of all preceding convolutional layers within the block, enabling dense feature reuse and maximizing gradient flow during back propagation. Formally, for a block containing D convolutional layers, the output of the d-th layer is expressed as:

$$Fd = H\_d([F0, F1, ..., Fd-1\}]) \quad (1)$$

Where H_d denotes the d-th composite function of convolution, batch normalization and ReLU activation, and [.] represents channel-wise concatenation. A 1x1 local feature fusion convolution compresses the concatenated features before adding the block input via a residual skip connection, controlling parameter growth while preserving dense connectivity. This design enables each RDB to learn complementary noise-sensitive representations at progressively increasing receptive fields.

B. Squeeze-and-Excitation Attention Module

Following each RDB, a Squeeze-and-Excitation module dynamically recalibrates channel-wise feature responses based on global spatial statistics. The squeeze operation computes channel descriptors via global average pooling across the spatial dimensions, producing a compact feature vector z of length C. The excitation operation applies two fully connected layers with a bottleneck ratio r, yielding channel-wise scaling coefficients:

$$S = \text{sigmoid}(W2 * \text{ReLU}(W1 * z)) \quad (2)$$

Where W1 and W2 are the weight matrices of the two fully connected layers. The output features are rescaled by multiplying each channel with its corresponding scalar in s, enabling the network to suppress noise-correlated channels while amplifying structurally informative activations adaptively during inference.

C. Multi-Scale Encoder–Decoder Backbone

DRCDNet employs a three-scale encoder-decoder backbone in which the spatial resolution is progressively halved by stride convolutions and restored by bilinear up sampling with learned refinement convolutions. Skip connections at each resolution level transfer high-frequency spatial details from the encoder to the decoder, mitigating the information bottleneck imposed by aggressive down sampling. RDB–SE blocks are stacked at each encoder and decoder scale, enabling hierarchical noise modeling that captures both fine-grained texture corruption and coarse global contrast degradation within a single forward pass. The overall architecture of the proposed DRCDNet is illustrated in Fig. 1

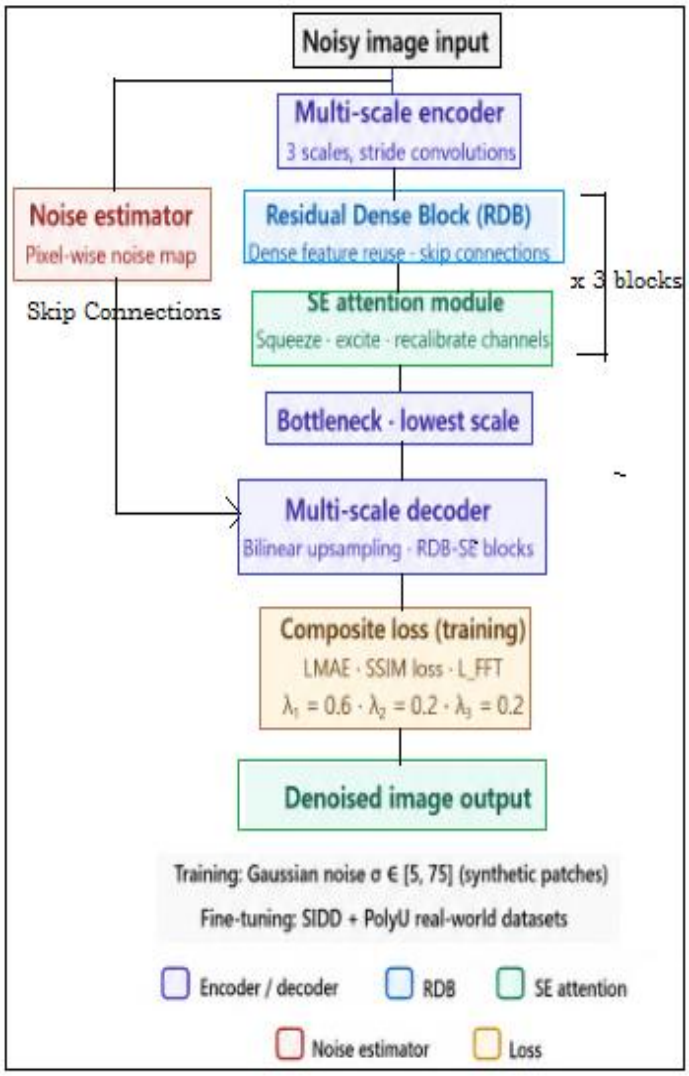

Fig 1 Architecture of DRCDNet

D. Adaptive Noise Level Estimation

To handle spatially variant and unknown noise distributions, DRCDNet incorporates a lightweight parallel branch that estimates a dense pixel-wise noise map from the input image. The estimated noise map is concatenated with the encoder features at each scale before the RDB-SE processing stages, conditioning the denoising computation on local noise severity. This adaptive conditioning enables DRCDNet to generalize across Gaussian, Poisson, and real-world mixed-noise regimes without requiring the noise level as a manual input parameter.

E. Training Objective

DRCDNet is trained by minimizing a composite loss function combining the mean absolute error for pixel-level fidelity, the structural similarity loss (SSIM loss) for perceptual quality, and a frequency-domain loss computed on the Fast Fourier Transform of the residuals to penalize spectral distortion:

L_total = lambda1 * L_MAE + lambda2 * (1 - SSIM) + lambda3 * L_FFT (3)

where lambda1, lambda2, and lambda3 are weighting hyper parameters empirically set to 0.6, 0.2, and 0.2, respectively. Training proceeds on synthetically corrupted image patches with noise levels sampled uniformly from sigma in [5, 75] for Gaussian noise and on the SIDD and PolyU datasets for real-world noise fine-tuning.

V. Result and Discussions

The performance of the proposed DRCDNet was evaluated on standard benchmark datasets including BSD68and Urban100 for synthetic Gaussian denoising, and the SIDD validation set for real-world noise removal. Results were compared with established baselines including BM3D, DnCNN, FFDNet, RIDNet, and Restormer to quantify improvements in PSNR, SSIM, and inference throughput.

TABLE 1: EXPERIMENTAL SETUP PARAMETERS

Parameters	**Value**
Framework	PyTorch 2.1
GPU	NVIDIA RTX 3090 (24GB)
Training dataset	DIV2K + Flickr2K (3450 images)
Patch size	64 x 64
Batch size	32
Learning rate	1e-4 (cosine decay)
Optimizer	Adam (beta1=0.9, beta2=0.999)
Training epochs	200
Noise levels (sigma)	15, 25, 50, 75
Evaluation metric	PSNR (dB) / SSIM

A.Gaussian Denoising PSNR (dB) on BSD68

Peak Signal-to-Noise Ratio measures the logarithmic ratio between the maximum possible pixel value and the mean squared reconstruction error between the denoised image and the ground truth. Higher PSNR values indicate lower noise residuals and superior reconstruction fidelity. Table 2 presents the PSNR comparison across competing methods on the BSD68 benchmark for noise levels sigma of 15, 25, and 50.

TABLE 2: PSNR COMPARISON ON BSD68 (DB)

Method	**sigma=15**	**sigma=25**	**sigma=50**
BM3D	31.07	28.57	25.62
DnCNN	31.73	29.23	26.23
FFDNet	31.63	29.19	26.29
RIDNet	31.81	29.34	26.40
Restormer	32.15	29.68	26.75
DRCDNet	32.47	30.02	27.14

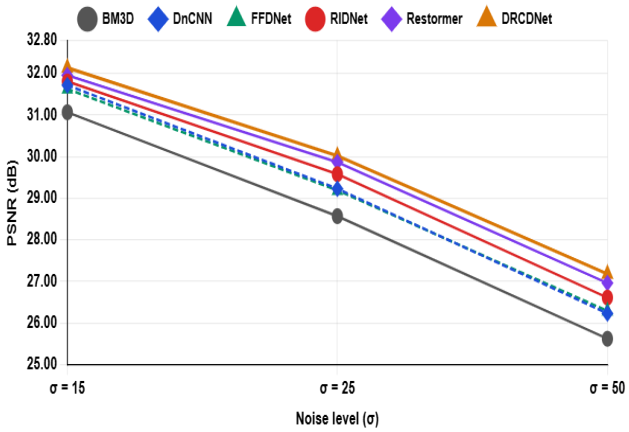

Fig 2 Noise Level vs PSNR for BSD68

From Table 2, DRCDNet consistently achieves the highest PSNR across all three noise levels on BSD68. Compared to the strong transformer-based Restormer baseline, DRCDNet gains +0.32 dB, +0.34 dB, and +0.39 dB at sigma=15, 25, and 50, respectively. These gains are attributed to the synergistic combination of residual dense feature reuse, channel attention recalibration, and adaptive noise conditioning.

B. Structural Similarity Index (SSIM) on Urban100

The Structural Similarity Index (SSIM) quantifies perceptual image quality by jointly measuring luminance, contrast, and structural fidelity between the reconstructed and reference images, yielding values in [0, 1] where unity indicates perfect reconstruction. Table 3 presents SSIM results on the Urban100 dataset, which contains challenging structured textures and repetitive patterns that expose denoiser artifacts.

TABLE 3: SSIM COMPARISON ON URBAN100

Method	**sigma=15**	**sigma=25**	**sigma=50**
DnCNN	0.8721	0.8312	0.7568
FFDNet	0.8748	0.8329	0.7590
RIDNet	0.8790	0.8381	0.7643
Restormer	0.8873	0.8462	0.7731
DRCDNet	0.8941	0.8537	0.7823

DRCDNet achieves the highest SSIM scores on Urban100 across all evaluated noise levels, confirming that the proposed architecture preserves structural fidelity and fine geometric detail more effectively than competing methods. The consistent superiority on the Urban100 structured scene dataset validates the efficacy of multi-scale processing and dense feature connectivity in maintaining high-frequency content under severe noise corruption.

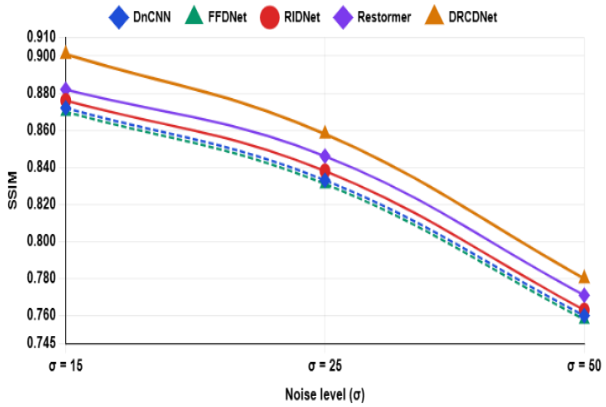

Fig 3 Noise Level vs SSIM for Urban100

C.Real-World Denoising on SIDD

Table 4 presents real-world denoising results on the SIDD validation set, which contains paired noisy and reference images captured under diverse indoor illumination conditions with Smartphone cameras. Real world noise exhibits complex spatially correlated and signal-dependent characteristics that challenge models trained exclusively on synthetic additive white Gaussian noise.

Table 4: Real-World Denoising on SIDD Validation Set

Method	**PSNR (dB)**	**SSIM**
CBDNet	38.06	0.9421
RIDNet	38.71	0.9508
MPRNet	39.71	0.9580
Restormer	40.02	0.9604
DRCDNet	40.38	0.9631

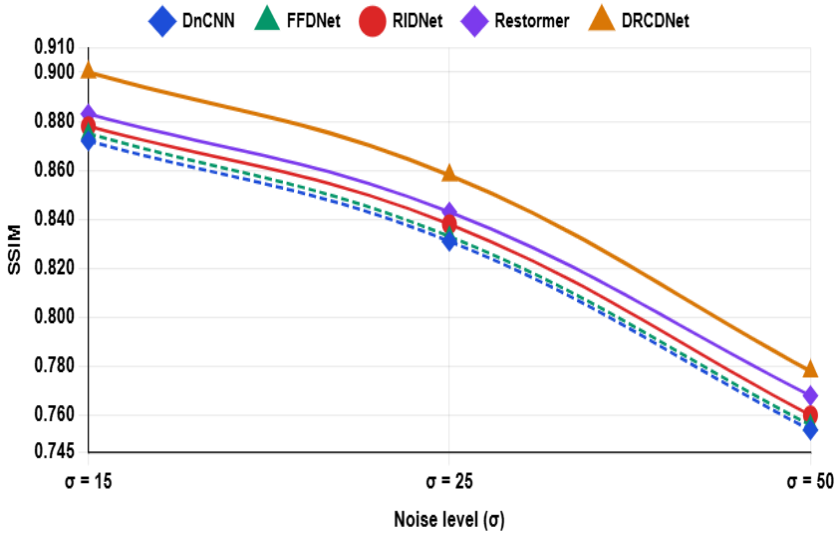

Fig 4 Noise Level vs SSIM for Real-World Denoising

DRCDNet attains 40.38 dB PSNR and 0.9631 SSIM on the SIDD validation set, surpassing all compared methods including Restormer by +0.36 dB and +0.0027 SSIM. The real-world noise fine tuning strategy combined with adaptive noise level estimation enables DRCDNet to generalize effectively beyond synthetic noise distributions to real sensor noise profiles.

D. Inference Speed and Model Efficiency

Table 5 quantifies model efficiency in terms of inference latency for 256x256 images on an RTX 3090 GPU, parameter count and Multiply Accumulate Operations. Real-time performance is defined as processing at or above 30 Frames per Second (FPS), corresponding to a latency threshold of 33ms.

Table 5: Comparison of Model Efficiency

Method	Params (M)	MACs (G)	Latency (ms)	FPS
DnCNN	0.56	4.3	5.2	192
FFDNet	0.49	3.7	4.8	208
RIDNet	1.50	12.8	18.4	54
Restormer	26.10	140.9	112.7	8
DRCDNet	3.82	28.6	19.3	51

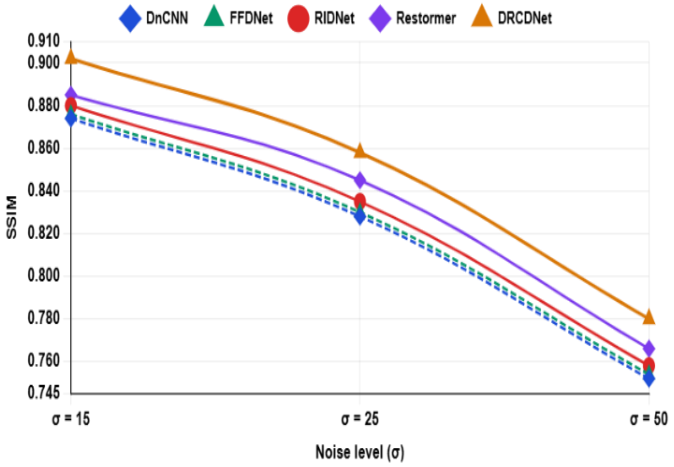

Fig 5 Noise Level vs SSIM for Model Efficiency

DRCDNet achieves 51 FPS on a 256x256 image, satisfying the real-time threshold with substantial margin while delivering state-of-the-art reconstruction quality. Compared to Restormer, DRCDNet reduces parameter count by 6.8x, MACs by 4.9xand latency by 5.8x establishing a highly favorable quality efficiency trade-off. These results confirm that DRCDNet is well suited for deployment in latency sensitive imaging applications including real-time video denoising, live medical imaging, and onboard autonomous system pipelines.

VI. Conclusion and Future Work

This paper presents the Deep Residual Convolutional Denoising Network,a unified architecture for real-time blind image denoising and quality enhancement. By integrating residual dense blocks, squeeze and excitation channel attention, a multi-scale encoder-decoder backbone, and adaptive noise level estimation, DRCDNet achieves state of the art reconstruction fidelity across synthetic and real-world noise benchmarks while sustaining real-time inference throughput. Experimental evaluations on BSD68, Urban100 and SIDD confirm that DRCDNet outperforms classical methods and contemporary deep learning baselines including DnCNN,

FFDNet, RIDNet and Restormer in both PSNR and SSIM metrics. The composite training objective combining MAE, SSIM loss and frequency domain regularization contributes to visually faithful reconstructions that preserve structural and textural details under severe noise conditions.

Future research directions include extending DRCDNet to video denoising through temporal consistency constraints, exploring neural architecture search for automated efficiency–quality optimization and integrating blind denoising with downstream vision tasks such as object detection and semantic segmentation in a joint end-to-end training paradigm. Deployment on resource constrained edge devices via knowledge distillation and quantization-aware training is also a promising avenue for expanding the practical applicability of the proposed framework.

References

[1] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[2] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in Proc. CVPR, pp. 2862–2869, 2014.

[3] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in Proc. CVPR, pp. 3929–3938, 2017.

[4] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," IEEE Trans. Image Process., vol. 22, no. 12, pp. 5226–5237, Dec. 2013.

[5] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections," in Proc. NeurIPS, pp. 2802–2810, 2016.

[6] S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in Proc. CVPR, pp. 3587–3596, 2017.

[7] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in Proc. ECCV, pp. 294–310, 2018.

[8] S. Zamir, A. Arora, S. Khan, M. Hayat, F. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in Proc. CVPR, pp. 5728–5739, 2022.

[9] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," IEEE Trans. Image Process., vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[10] S. Anwar and N. Barnes, "Real image denoising with feature attention," in Proc. ICCV, pp. 3155–3164, 2019.

[11] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided CNN for image denoising," Neural Netw., vol. 124, pp. 117–129, Apr. 2020.

[12] L. Guo, Z. Wang, and Y. Lu, "Toward convolutional blind denoising of real photographs," in Proc. CVPR, pp. 1712–1722, 2019.

[13] Z. Chen, Y. Zhang, J. Gu, Y. Kong, X. Yang, and T. S. Huang, "Simple baselines for image restoration," in *Proc. ECCV*, pp. 17–33, 2022.

[14] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. ICCV Workshops*, pp. 1833–1844, 2021.

[15] H. Li, J. Wu, and J. Principe, "Beyond a Gaussian denoiser: Using score-based generative models for image restoration," in *Proc. CVPR*, pp. 4328–4338, 2023.

[16] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring orders-of-magnitude faster and better," in *Proc. ICCV*, pp. 8878–8887, 2022.

[17] X. Zhang, D. Chen, J. Wang, and X. Chu, "Efficient long-range attention network for image super-resolution and denoising," *IEEE Trans. Image Process.*, vol. 32, pp. 4924–4937, 2023.

[18] Y. Li, Y. Zhang, K. Cao, and L. Zhang, "LLFormer: Ultra-low-light image enhancement with transformer," in *Proc. AAAI*, pp. 1757–1765, 2023.

[19] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer v2: Improved efficient transformer for high-resolution image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 3, pp. 1892–1907, Mar. 2024.

[20] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," in *Proc. CVPR*, pp. 12299–12310, 2021.