International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct 2026, PP 01-18

CAMSD-Net for SIDD-Small Real-World Image Denoising Using Lightweight Multi-Scale Deep Learning Framework

Arpit Urmaliya[1], Jitendra Kumar Maitra[2]
[1]Research scholar, Department of Mathematics and Computer Science
Rani Durgavati University, Jabalpur, Madhya Pradesh India
[2]Professor Department of Mathematics and Computer Science
Rani Durgavati University, Jabalpur,Madhay Pradesh India

Abstract. Effective image denoising is a challenging task for real-world smartphone images due to the complex noise resulting from the limited response of the sensors and the varying illumination conditions. The paper proposes and tests a light weight deep learning framework called Convolutional Adaptive Multi-Scale Denoising Network CAMSD-Net to restore high quality image from noisy smartphone photos in the meanwhile retaining the structure and visual quality. The proposed methodology employs the SIDD-Small sRGB dataset with 160 pairs of noisy and clean images which are systematically validated, with 70% used for training, 15% for validation, and 15% for testing, and with random 256×256 image patches extracted from each image, leading to a fixed 384 paired image patches used for testing. The network is implemented based on multi-scale convolutional network, local spatial learning, dilated contextual feature extraction, feature fusion and skip connections, and is trained for 30 epochs with the Adam optimizer and learning rate of 0.001. Experimental results show the superior performance of CAMSD-Net V2 over classical denoising schemes and the first version of the network. The proposed model achieves the highest Peak Signal-to-Noise ratio (PSNR) of 36.5564 dB, and the Structural Similarity Index (SSIM) of 0.925337, and minimizes the Mean Squared Error (MSE) of 0.00032221, Root Mean Squared Error (RMSE) of 0.016248, and Mean Absolute Error (MAE) of 0.012848. In addition, the proposed framework is effective in real world smartphone image restoration as evidenced by its performance on 356/384 test patches (92.71%) with respect to the improvement of PSNR.

Keywords: Smartphone Image Denoising, Deep Learning, Convolutional Adaptive Multi-Scale Denoising CAMSD-Net, SIDD-Small Dataset, Multi-Scale Convolutional Network.

I. Introduction

Everyday photography, social media, medical documentation, and intelligent vision—these are the most important functions of today's imaging and camera uses, which are mainly handled by smartphones. The call for good quality images in various lighting environments, such as low-light scenes, indoor environments, and difficult lighting situations has grown as they are widely used [1]. However, images captured by the smartphone camera are subject to many kinds of noises, which causes degradation in visual quality and makes subsequent computer vision applications like object detection, image segmentation, object recognition and scene understanding less accurate [2], [3], [4], [5]. In the real world, there will always be noise present, which makes the pre-processing of image restoration an important process for modern imaging systems.In smartphone photography, image noise is much more complex and distinct from the synthetic Gaussian noise that is usually employed in traditional image denoising research. Multiple physical processes in the camera pipeline are responsible for real smartphone noise: photon shot noise, electronic read noise, sensor imperfection, color channel imbalance and image signal processing operations.

These noise components are signal dependent, spatially variable and can often interact with each other within color channels. Therefore, denoising techniques designed under simplified synthetic noise conditions may not have a good generalisation to real-world context of a smartphone image.The early image denoising methods used mathematical filtering technique like Gaussian filtering, median filtering, Wiener filtering and Non-Local Means (NLM). All these techniques try to subtract from unwanted noise, using local neighbour information or patch similarity across an image. While they can help to reduce noise to some degree, they often cause undesirable artifacts like edge blurring, smoothing of textures and loss of fine structural detail. Not only is noise reduction important for imaging applications, but maintaining image detail and keeping the natural textures of the image is also important as the applications need to deliver increasingly greater visual fidelity.

With the recent leaps in the state of the art of deep learning, the community has begun to explore new methods for image restoration that can learn complex mappings directly from noisy images to their clean counterparts. Convolutional Neural Networks (CNNs) have shown to be very successful in learning hierarchical spatial information while maintaining relevant image structures. Deep learning models, such as DnCNN, CBDNet, RIDNet and other CNN-based models have achieved significant performance improvement in denoising compared to traditional filtering methods. Recently, models incorporating attention mechanisms, multi-scale feature extraction and transformer networks have been able to model long-range dependencies and retain very fine image features, which has further improved the performance of deep networks. However, there are many approaches currently in existence which are relatively complex. Architectures that demand high computational needs, optimization techniques that restrict possible implementation in imaging systems with limited resources [6], [7], [8].

One of the most important steps made in the field of real world denoising is the release of a dataset called Smartphone Image Denoising Dataset (SIDD) where noisy and clean smartphone images are paired and acquired under realistic imaging conditions. As it is the real noise collected from various lighting conditions and multiple smartphone devices, SIDD is one of the most widely accepted benchmarks to test real-image denoising algorithms [9][10]. The SIDD-Small subset contains carefully aligned pairs of noisy and ground-truth images suitable for supervised learning, with the same characteristics as real smartphone sensor noise. This has spurred further research into algorithms that can solve real-world denoising problems and not just use synthetic noise models.While significant progress has been made in the denoising performance of deep learning methods, it is still an important research challenge to obtain good noise reduction performance while preserving details.

Too much smoothing can cause the loss of important texture information, while too little denoising can cause some residual noise that may affect the subsequent viewing tasks. Moreover, different types of images and channel-dependent noise create a demand for architectures that can extract both local and global representations of image content and texture simultaneously. This study focuses on a direct deep-learning denoising framework specifically for real-world smartphone images [11], [12]. in response to these challenges. This proposed CAMSD-Net uses multi-scale convolutional networks to learn a direct mapping from noisy RGB images to clean RGB images, with features designed to extract local spatial information, to use the dilated receptive field to represent the contexts of the image, and to fuse the features and to refine the outputs. Unlike conventional prediction methods, the network restores images from the image to image, trying to maintain structural information while the complex real-world noises are reduced effectively.

The SIDD-Small sRGB dataset as well as a carefully designed experimental pipeline, with an emphasis on reproducibility and data integrity, are the focus of the present work. To avoid any scene leakage between the training, validation and test sets, the dataset is divided at the image level prior to patch extraction. Furthermore, the methodology incorporates data verification, patch quality control, classical baseline preparation and standardized evaluation metrics to create a consistent analysis framework of real world smartphone image denoising.

II. Literature Review

Lee et al., 2026 [13] To tackle the challenge of balancing computational efficiency and modeling long-range dependency, explores a practical CNN–Transformer hybrid network for real-world image denoising. The goal is to get denoising results competitive with the state-of-the-art and decrease the computational cost when deployed on resource limited devices. The methodology consists of using combinations of strategic CNN–Transformer blocks, eliminating nonlinear operations and searching the CNN architecture under explicit resource constraints, with the aid of NAFNet and Restormer. Experimental evaluation on the SIDD and DND datasets demonstrates 39.98–40.05 dB PSNR and 0.958–0.961 SSIM on SIDD, while DND achieves 39.73–39.91 dB PSNR and 0.959–0.961 SSIM. The models only need 7.18–16.02 M parameters and 20.44–44.49 G MACs, and the cross-validations show strong generalization and practical efficiency.

Jang et al., 2025 [14] presents Deep-Feature Auxiliary Guidance (DFAG) method for enhancing the accuracy of image restoration while maintaining the same model structures and inference time. The goal is to improve the quality and stability of the deep feature learning in multi-scale image restoration networks. However, in the methodology, the auxiliary losses are added to the feature maps at each multi-scale level during training and the extra DFAG modules are stripped during inference, thus eliminating the need for additional computational burden. Multi-scale encoder-decoder restoration models are tested across a range of experiments in real image denoising, gaussian denoising, motion deblurring, JPEG artifact reduction and super-resolution with consistent improvement in each of these tasks. The results validate DFAG as an effective training method for enhancing the restoration performance while retaining the original architecture and inference efficiency.

Tang et al., 2024 [15] In order to alleviate the expensive procurement of real-world noisy images and the creation of labeled datasets, presents a self-supervised image denoising approach called Multi-Scale Blind-Spot Network with Adaptive Feature Fusion (MA-BSN). The goal is to achieve better denoising performance and overcome the limitations of existing self-supervised methods, such as spatial noise correlation and small receptive fields. The methodology is based on multi-scale blind-spot architecture to generate sub-images, a depth-wise convolutional Transformer network to obtain the global feature, and an Adaptive Feature Fusion module to refine the feature in the region through attention mechanism. Experiments on real world noise benchmark datasets, SIDD and DND, show good denoising result, with DND reaching 38.41 dB PSNR and 0.940 SSIM, outperforming the state-of-the-art self supervised methods available in the literature.

Zhang & Zhou, 2023 [16] propose Denoise Transformer to overcome the drawbacks of CNN-based approaches such as limited receptive field and color drift, which are caused by the difficulty in extracting texture and color information from the images. The goal is to enhance the ability to remove noise and to maintain both global and local image information without a priori knowledge of noise intensity or noise type. The methodology consists of Context-aware Denoise Transformer (CADT) with dual global and local branches, hierarchical network for residual noise distribution learning and Secondary Noise Extractor (SNE) for efficient global noise extraction. The final denoised image is then reconstructed from the blind spots. The experimental results for the SIDD benchmark are 50.62 dB PSNR and 0.990 SSIM, which have excellent visual performance for the sRGB, Raw-RGB and grayscale data sets.

Y. Liu et al., 2022 [17] presents FDN, an invertible denoising network which focuses on image denoising as a distribution learning and disentangling problem instead of a standard feature-based denoising problem. The goal is to understand the distribution of noisy images and recover clean images by operating on the image latent representations, eliminating distortions and artifacts, without making any assumptions on the distribution of clean images or the distribution of noise. The methodology consists of the assumption of noisy image as a joint distribution of clean image and noise and the distribution disentanglement via the proposed FDN framework. An experimental evaluation shows the efficiency in the removal of synthetic additive white Gaussian noise from category specific and remote sensing images. Results also indicate that FDN is superior to the previously published methods in real-image denoising in terms of the number of parameters used and the speed of processing.

TABLE I. LITERATURE SUMMARY

Author/Year	Research Gap	Method	Finding
Cho et al., 2026 [18]	MSE-trained denoisers tend to favor bright regions, resulting in poorer detail recovery in dark regions under signal-dependent noise.	Brightness Bias-Robust Denoising (BBRD); brightness-band error normalization with empirical noise variance and Group-DRO.	Improved all brightness bands, with up to +0.45 dB PSNR in dark bands, +0.32 dB in bright bands, and +0.65 dB aggregate PSNR on SIDD.
R. Liu et al., 2025 [19]	Existing CNN and Transformer restoration methods have limitations in simultaneously capturing local spatial and global channel information.	Multi-Stage progressive image restoration Network (MSTNet) using window-based and channel-level Transformers with feature fusion.	Demonstrated effective deblurring and denoising performance and outperformed existing state-of-the-art methods on general and colon image datasets.
Tiantian et al., 2024 [20]	Deep denoising networks often require excessive depth, increasing computational complexity while affecting efficient feature extraction.	Lightweight progressive residual and attention fusion network using dense blocks, progressive residual fusion, and CAFFM.	Improved PSNR, SSIM, and FSIMc under Gaussian and natural noise levels of 15–50, outperforming more than 20 methods across six datasets.
Ko et al., 2026 [21]	Real-world noise synthesis is limited by the scarcity of aligned clean-noisy pairs and camera metadata.	YeTI using a Reconstruction Autoencoder and one-step Conditional Diffusion Transformer trained from two noisy observations.	Generated realistic signal-dependent noise and enabled denoisers trained with synthesized images to achieve strong real-world performance on DND.
Ryou et al., 2025 [22]	Real-world restoration performance is constrained by the quality of available ground-truth images.	Ground-truth enhancement using super-resolution, adaptive frequency masks, frequency-domain mixup, and a lightweight output refinement network.	Consistently improved restored-image quality, with experiments and user studies validating supervision enhancement and output refinement.

III. Research Methodology

The proposed methodology is systematically deep-learning workflow for denoising image by using a smartphone camera image, which is reproducible, consistent, and reliable for experimental implementation, including dataset validation, preprocessing, patch generation, and CAMSD-Net training and standardized evaluation process.

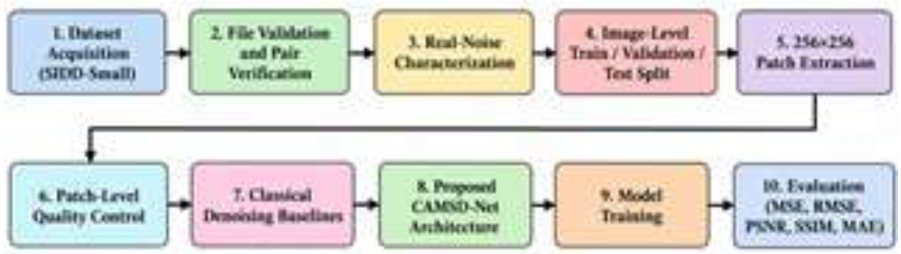

Fig. 1 Proposed Flowchart

A. SIDD-Small Dataset Acquisition and File Validation

The proposed methodology starts with the acquisition of the SIDD-Small sRGB dataset that includes pairs of noisy and clean smartphone images under a wide range of indoor lighting conditions with different mobile camera sensor types. In order to obtain accurate correspondence between noisy images and their ground truth, it is necessary to set up the noisy–ground truth pairs before pre-processing, as the model is supervised image denoising. A full validation procedure ensures your file is available, your image dimensions and RGB channel consistency, your file formats and that you have a complete pair within all of your scene folders. Automated integrity checks also ensure the spatial alignment of every noisy image with its clean reference image, and detect missing, duplicated and corrupted images. By passing this validation step you have a valid data structure, you don't let invalid samples into the training pipeline and you have a consistent ground to do reproducible image denoising experiments.

B. Noisy–Ground Truth Pair Verification and Data Integrity

A special pair-verification phase verifies that each noisy image is properly paired to the reference image. The supervised image denoising requires to be paired at the pixel level, so that even if there is a slight pairing error, the performance of the model will be affected. The two images in each pair are checked for uniformity of spatial resolution, uniformity of RGB channel, exact file names and scene level organization. The automated validation scripts detect missing samples, duplicate samples and incorrectly paired samples and provide summary statistics to ensure that the dataset is complete. To avoid scene leakage and train and evaluate with only valid image pairs, pair verification is performed prior to the patch extraction. This helps ensure data consistency and reliability, which is crucial for future data feature extraction and optimization using deep learning.

C. Real-Noise Characterization

The proposed methodology investigates the statistical properties of the noise of the real smartphones prior to training the model in the SIDD-Small dataset. Normalization is performed for each pair of noisy–clean images, and the resulting difference images are considered to be the camera noise. The noise behavior is quantified by the statistical measures, such as residual mean, standard deviation, variance, RMSE and MAE. Each analysis is done for the Red, Green and Blue channel, and brightness based analysis uses the correlation between illumination intensity and noise variation. This stage does not assume a given noise distribution, and provides the possibility to use a multi-scale convolutional network, which can learn complex noise distributions across both channels and space.

D. Image-Level Dataset Splitting

Before extracting patches, the SIDD-Small dataset is split into three equal parts: 70% to the training set, 15% to the validation set, and 15% to the test set to ensure unbiased training and reliable performance evaluation. The split is done at the original image level, thus avoiding the possibility of scene leakage during supervised learning between the various subsets. To ensure the repeatability of experiments, a fixed random seed has been used. The validation set assists the process of model optimization and the testing set is separate for the last evaluation. Each created patch is guaranteed to contain the same partition of its parent image, which keeps the integrity of the datasets through the training pipeline. This is an image-level splitting process that greatly reduces overfitting and serves as a solid approach for assessing the generalization performance of the proposed deep-learning model.

E. Random 256×256 Patch Extraction

High-resolution smartphone images need significant processing power and memory to train directly on these images. Thus, the proposed approach is based on a patch based training method that extracts fixed size image patches from all the identified image pairs while maintaining exact spatial correspondence between the noisy and clean images.

The network is trained to learn local image structures, textures and noise patterns with multiple patches of each original image, which are typically 256 x 256 pixels. The fixed random seed ensures that the same patch location can be extracted from different experiments, helping to ensure experimental consistency.

When cropping noisy and clean images, the same spatial coordinates are used for supervised correspondence. This ensures that every patch with noise has an exactly aligned patch with target during training.

The benefit of using smaller patches is that it will require less GPU memory, more number of samples to be trained, and expose the network to a variety of different characteristics in each original photograph. Rather than just depending on a handful of high-resolution images, the model is trained using thousands of localized examples that feature different textures, edges and lighting angles.

The patch-based strategy thus reduces the computation time while keeping the spatial information essential for learning the accurate image denoising representations.

F. Patch-Level Quality Control

Following the patch extraction, a complete quality-control process is carried out to ensure that all the patches created meet the needed preprocessing requirements before being inputted into the model-training pipeline. This stage takes care that all training samples have the same formatting, size and match with the corresponding ground-truth image.

First, all the removed patches are checked to ensure that each patch occupies a valid spatial size and that the RGB channels are consistent for the patch, and that the noisy and clean images are properly paired. Inconsistent patches, that may cause inconsistencies in the supervised learning process, are detected by automated verification routines by identifying incorrectly cropped, duplicated or mismatched patches.

Further quality checks include aspects of the image content, including near-black areas, low-information regions, brightness variation and consistency of pixel values. These observations are not deleted from the dataset, but are recorded instead in order to maintain transparency of the data and to facilitate data reproducibility when challenging samples are encountered.

Simple descriptive statistics are also computed to ensure that the retrieved patch collection is representative of the diversity in the original smartphone images. This quality-control step helps to prevent propagation of the preprocessing errors into the training process and keeps the resulting patch dataset from being affected by unrealistic imaging conditions.

The proposed methodology will enable a reliable and reproducible dataset preparation pipeline to be validated prior to the beginning of optimization, which can then be used in deep-learning based image denoising experiments.

G. Classical Denoising Baseline Preparation

To compare the proposed deep-learning framework with the classical image-denoising techniques, a set of classical image-denoising techniques are introduced into the methodology. These baseline approaches offer a unified baseline framework to compare the proposed architecture with the same preprocessing and testing conditions.

Four denoising methods are chosen: Gaussian filtering, Median filtering, Wiener filtering and Non-Local Means (NLM). These methods are some of the most common spatial-domain filtering methods and variations in noise suppression and detail preservation.

The parameters for each algorithm are pre-defined and the same set of testing patches are used for each algorithm. In order to provide the same input conditions to all methods, the noisy images of the input remain unchanged prior to baseline processing. Consistency in evaluations is achieved by using the same evaluation data set and avoiding different methods of preprocessing.

The inclusion of the traditional filtering techniques is not to change the proposed model, but to create a clear experimental standard. All these baselines are typical methods that have been adopted in image restoration so they serve as good references for future deep-learning experiments.

The procedures and inputs used in all techniques are kept the same to ensure the overall experimental approach is reproducible and scientific.

H. Proposed CAMSD-Net Architecture

Proposed CAMSD-Net aims to be a lightweight multi-scale convolutional neural network for direct image-to-image denoising of real-world smartphone images.

The architecture is different from the residual-learning methods which estimate the noise component in isolation. The initial convolutional layer is the first layer in the network that is designed to extract low-level visual features from the input image.

The extracted representations are then passed to two complementary parallel branches. A spatial branch is used to sample the local texture and fine image structures while a dilated context branch is used to expand the receptive field to include more context, but without loss of spatial resolution.

Both branches output features are concatenated and then fused with the channel information effectively by a 1×1 convolutional fusion layer. The resulting image is progressively enhanced from each of the refinement layers to create the final restored image. Important structural information is preserved across the network with the help of the detail-preserving skip connection

The last convolutional layer produces three channel output of rgb which directly corresponds to the predicted clean image. The multi-scale design allows the network to capture complex spatial relationships while maintaining efficient computation for real-world image restoration applications.

I. Model Training Configuration

During the optimization process, the proposed CAMSD-Net is trained by supervised learning with all the noisy image patches and their corresponding clean reference ones. A proper training configuration is used to guarantee stability and convergence of the trained system, computational efficiency and repeatability of repeated experiments.Optimization of the network is performed with the Adam optimizer, which optimizes the model parameters through backpropagation adaptively. An initial learning rate is used for the sake of optimizing for convergence speed without compromising optimization stability, and a mini-batch strategy is used to enhance the efficiency of GPU training in memory usage.Training is done across several epochs to enable the network to see many different real-noise patterns generated from thousands of image patches.

Gradient thresholding is introduced to increase stability in the optimization process and smoothness in the updates of the parameters within the training process.During optimization a separate validation dataset is used so that the learning behaviour can be monitored without impacting the independence of the testing partition. The goal of training is to reduce the difference between the predicted clean images and the corresponding ground-truth targets by minimizing a pixel-level regression loss throughout training.By fixing all optimization parameters during the experimental process, future comparisons between different architectural variations can be done consistently, reproducibly and in a scientific way.

J. Evaluation Framework

The last step of the suggested methodology is to create a standardized assessment structure to evaluate the proposed CAMSD-Net once trained with a model. For an adequate comparison of the performances of different denoising methods and architectural variations, a consistent evaluation protocol is necessary.The framework establishes a common test procedure, where each method is assessed based on the same independent test data obtained during the image level partitioning process. Identical testing conditions ensure that the results of the different experiments are not due to different input samples or to preprocessing available in each experiment.Five quantitative image quality measures are chosen as the main evaluation measures: mean squared error (MSE), root mean squared error (RMSE), peak signal to noise ratio (PSNR), structural similarity index (SSIM), and mean absolute error (MAE). These are all based on reconstruction accuracy at the level of individual pixels, and structural similarity between the reconstructed and reference images.

1. Mean Squared Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(I_i - \hat{I}_i\right)^2 \qquad (1)$$

2. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{MSE} \qquad (2)$$

3. Peak Signal-to-Noise Ratio (PSNR)

$$PSNR = 10log_{10}\left(\frac{MAX^2}{MSE}\right) \qquad (3)$$

4. Structural Similarity Index (SSIM)

$$SSIM = \frac{(2\mu I\mu\hat{I} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \qquad (4)$$

5. Mean Absolute Error (MAE)

The framework also includes numerical evaluation as well as patch-level comparisons and procedures for conducting statistical analysis that can be applicable across model versions. Implementing the evaluation protocol before the experimentation increases the reproducibility of the experiment, allows for objective comparison of approaches, and will provide a scientifically rigorous assessment procedure for future experiments without evaluation bias.

IV. Result & Discussion

This section evaluates the proposed CAMSD-Net in the real-world image denoising of smartphone images on the SIDD-Small sRGB dataset. The quantitative evaluation of the proposed model is done using MSE, RMSE, PSNR, SSIM, and MAE. The results are compared to the classical denoising methods and two versions of the proposed architecture, CAMSD-Net V1 and CAMSD-Net V2. For all experiments, the test protocol used consisted of 384 pairs of image patches, making the evaluations consistent and reproducible.

Experimental Configuration

The proposed model was trained with a dataset of SIDD-Small sRGB which included 160 pairs of noisy and clean smartphone images. Prior to patch extraction the data set was split into 70% training, 15% validation, and 15% testing. Each image produced sixteen pairs of patches of 256 x 256 x size pixels. The 3 is converted to a testing set of

384 patches. The training process was carried out for 30 epochs with the learning rate of 0.001 and mini-batch size of 8.

Table II. Experimental configuration for CAMSD-Net evaluation.

Parameter	Value
Dataset	SIDD-Small sRGB
Image Pairs	160
Train/Validation/Test	112 / 24 / 24
Test Patches	384
Patch Size	256 × 256 × 3
Training Epochs	30
Optimizer	Adam
Learning Rate	0.001
Batch Size	8

To make a fair comparison between all the denoising methods, the experiment was conducted according to a standardized training condition and fixed testing protocol.

Classical Denoising Results

TFour traditional denoising methods: Gaussian filtering, Median filtering, Wiener filtering, and Non-Local Means (NLM) were first compared with the proposed model. These methods were used as benchmarks prior to testing the deep-learning framework.

Table III. Performance Comparison of Classical Denoising Methods.

Method	RMSE	PSNR (dB)	SSIM	MAE
Noisy Input	0.048014	28.5390	0.648801	0.036632
Gaussian	0.027633	33.2455	0.810441	0.020880
Median	0.033632	31.5337	0.762686	0.024934
Wiener	0.025234	33.9790	0.842260	0.018386
NLM	0.031774	34.4833	0.824815	0.023430

Compared to conventional methods, the Wiener filtering had the highest SSIM score (0.842260), and NLM had the highest PSNR score (34.4833 dB). These results provide significant traditional baselines to be used to compare with the proposed deep-learning architecture.

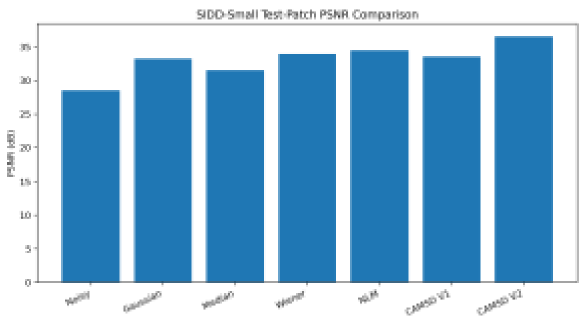

Fig. 2 Comparison on the fixed 384-patch test set.

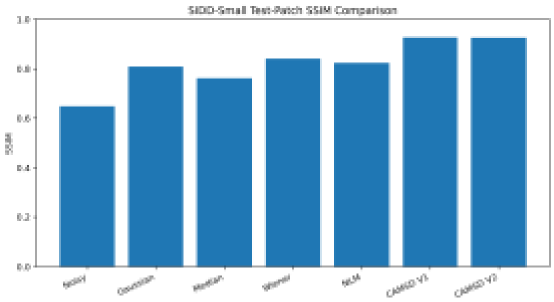

Fig. 3 SSIM comparison on the fixed 384-patch test set. The noisy-input SSIM shown here uses the classical-baseline evaluation value.

CAMSD-Net V1 Results

The initial proposed network was compared to the original noisy images to assess its performance with respect to direct clean image regression. The model has shown significant improvement in all the metrics.

Table. IV Camsd-Net V1 Performance Compared With The Noisy Input.

Metric	Noisy	CAMSD-Net V1	Improvement
MSE	0.00376649	0.00059051	−84.32%
RMSE	0.048014	0.022693	−52.72%
PSNR (dB)	28.5390	33.5463	+5.0074
SSIM	0.648801	0.926875	+0.278074
MAE	0.036632	0.018585	−49.26%

The results showed that CAMSD-Net V1 could achieve a high quality of images and significantly minimize the reconstruction error, indicating that it was possible to remove the real noise from the smartphone images directly from the clean image prediction.

CAMSD-Net V2 Results

The second version of the proposed architecture added multi-scale feature refinement and detail preserving learning, and kept the direct denoising framework. The highest overall performance was obtained by CAMSD-Net V2 in the experiments that were completed.

Table V. Camsd-Net V2 Performance Compared with The Noisy Input.

Metric	Noisy	CAMSD-Net V2	Improvement
MSE	0.00376649	0.00032221	−91.44%
RMSE	0.048014	0.016248	−66.16%
PSNR (dB)	28.5390	36.5564	+8.0174
SSIM	0.648801	0.925337	+0.276536
MAE	0.036632	0.012848	−64.93%

The PSNR and error values of CAMSD-Net V2 were the highest and the lowest, respectively, showing that more effectively real-world smartphone noise was suppressed and high structural similarity was maintained.

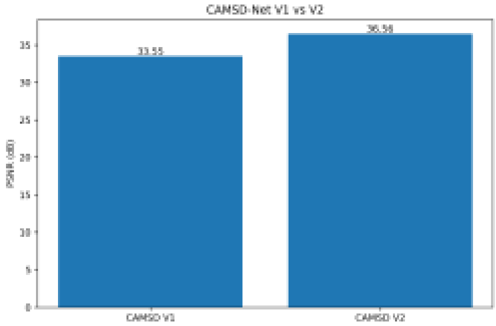

Fig. 4 CAMSD-Net V1 versus V2 PSNR.

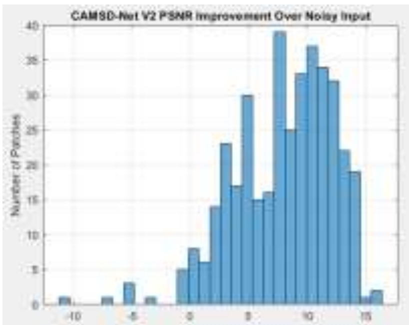

Fig. 5 Existing MATLAB plot of CAMSD-Net V2 PSNR improvement over the noisy input.

CAMSD-Net V1 versus V2 Comparison

The two proposed models were directly compared with the same fixed testing protocol. The comparison shows the improvements gained by implementing the improved multi-scale architecture of CAMSD-Net V2.

Table VI. Performance Comparison Between Camsd-Net V1 and Camsd-Net V2.

Metric	V1	V2	V2 − V1
MSE	0.00059051	0.00032221	−0.00026831
RMSE	0.022693	0.016248	−0.006446
PSNR (dB)	33.5463	36.5564	+3.0101
SSIM	0.926875	0.925337	−0.001539
MAE	0.018585	0.012848	−0.005737

Additional patch-level analysis showed that:

1. PSNR improved on 356 out of 384 patches (92.71%).
2. PSNR decreased on only 28 patches (7.29%).

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct 2026, PP 01-18

3. MSE decreased by 45.44% compared with V1.
4. RMSE decreased by 28.41%.
5. MAE decreased by 30.87%.

These enhancements were also confirmed statistically by both paired t-test and Wilcoxon signed-rank test with statistically significant improvements in PSNR values and a large practical effect was observed from Cohen's d = 1.2354.

V2 achieved a fairly small degradation in the average SSIM compared to V1, but it achieved better pixel level restoration by reducing reconstruction errors and enhancing PSNR in most of the test patches.

Qualitative Analysis

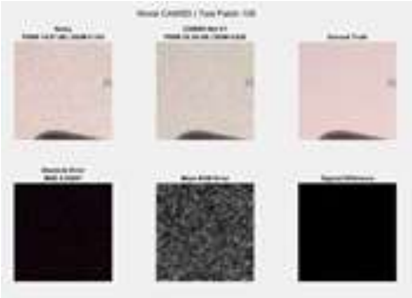

Fig. 6 CAMSD-Net V1 worst test patch example: strong noise removal but remaining error around fine structure.

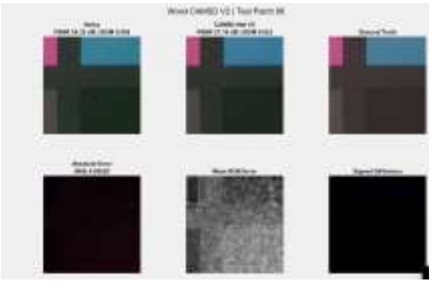

Fig. 7 CAMSD-Net V2 worst test patch example (Patch 96): PSNR 24.33→27.18 dB and SSIM 0.558→0.932.

The results of CAMSD-Net are shown qualitatively for the most difficult test patches in V1 and V2 in Figures 5 and 6, respectively. For smartphone noise reduction, CAMSD-Net V1 performs very well in suppressing the visible noise, as shown in Figure 5, with minor reconstruction errors in fine texture and curved edge areas. As shown in Figure 6, the restored images obtained by CAMSD-Net V2 show a clearer and more coherent image, with PSNR increasing from 24.33 dB to 27.18 dB, and SSIM from 0.558 to 0.932. The absolute error maps reveal that reconstruction errors are low throughout in the smooth areas with the primary errors being manifested as minor residual errors in the edges of the images, and largely reflecting the increased error in the mean RGB rather than a failure to reconstruct significant errors.

G. Comparison with Published SIDD Methods

For comparison, a few published denoising methods, which were reported to have been assessed in the context of SIDD, are compared to the proposed model. These are therefore only contextually valid comparisons, as the current work has a custom split of the images into 112/24/24 with a fixed 384 patch testing protocol.

Table VII. Comparison with Published Sidd Methods.

Published method	SIDD PSNR	SIDD SSIM	Relation to CAMSD V2
CVF-SID (S^2)[23]	34.71	0.913	Lower
AP-BSN + R^3[24]	35.79	0.925	PSNR lower; SSIM nearly equal
Noise2Variance[25]	34.91	0.915	Lower
CAMSD-Net V2 (this work)	36.5564*	0.925337*	-

The highest PSNR (36.5564 dB) and a comparable SSIM (0.925337) in the adopted experimental protocol give good evidence that CAMSD-Net V2 is superior at denoising. It achieves smaller reconstruction errors than published SIDD methods, while maintaining the structural details, suggesting that its lightweight multi-scale convolutional architecture has the ability to restore images from smartphone phones in the real world more effectively.

H. Discussion

The experimental results show that the proposed CAMSD-Net can restore the real world smartphone images corrupted by complex noises. The proposed network outperformed state-of-the-art denoising techniques in terms of PSNR and significantly reduced reconstruction errors, demonstrating its superior performance in restoring the image at the pixel level. The multi-scale feature-learning strategy used in CAMSD-Net V2 helped to achieve better noise suppression with high structural similarity. These are the experimental findings which fulfilled the research objectives. First, they developed a real-world image denoising system for the smartphone using the deep-learning framework, which was called CAMSD-Net. The second goal (model evaluation on SIDD-Small) was done by applying a standard experimental protocol with fixed training and testing partitions. Last, the comparison with the conventional denoising method and the comparison between CAMSD-Net V1 and V2 showed that the improved architecture achieved a better denoising effect on the smartphone image dataset, while the denoising effect was consistent during most of the test samples.

V. Conclusion

By using SIDD-Small sRGB dataset, this study successfully designed and tested CAMSD-Net, which is a lightweight multi-scale convolutional neural network for real-world image denoising on a smartphone camera. The proposed framework was systematically designed comprising of dataset validation, splitting of images at image level, extraction of patches from images, quality control, supervised learning and standardized evaluation to guarantee reliable and repeatable experimentation. In contrast to traditional filter-based methods, CAMSD-Net directly maps noisy and clean images, and can successfully remove the camera noise in complex real-world images while maintaining important image structures and fine textures.

The experimental results showed that the best performance of the proposed CAMSD-Net V2 (0.925337 PSNR, 0.016248 RMSE, 0.012848 MAE, 0.00032221 MSE) was achieved. The model also gave superior PSNR result on 356 out of 384 test patches, showing consistent performance over a variety of smartphone images. The comparative analysis also demonstrated that the proposed architecture outperforms the conventional denoising techniques when using the experiment conditions mentioned in the paper, and also achieved significant improvements compared with CAMSD-Net V1 by improving multi-scale features. In summary, the research goals were met, which show that the CAMSD-Net is a reliable, efficient and effective approach to restoring a real-world smartphone image at a high visual quality and with high structural fidelity.

References

1. W. Wang et al., “Computerized Medical Imaging and Graphics Diffusion model for medical image denoising , reconstruction and translation,” Comput. Med. Imaging Graph., vol. 124, no. June, p. 102593, 2025, doi: 10.1016/j.compmedimag.2025.102593.
2. K. Gong, K. A. Johnson, G. El Fakhri, Q. Li, and T. Pan, “PET image denoising based on denoising diffusion probabilistic model,” vol. 51, no. 2, pp. 358–368, 2025, doi: 10.1007/s00259-023-06417-8.PET.
3. R. T. Sadia, J. Chen, and J. Zhang, “CT image denoising methods for image quality improvement and radiation dose reduction,” J. Appl. Clin. Med. Phys., vol. 25, no. 2, pp. 1–17, 2024, doi: 10.1002/acm2.14270.
4. F. Hashimoto, Y. Onishi, K. Ote, H. Tashima, A. J. Reader, and T. Yamaya, “Deep learning - based PET image denoising and reconstruction : a review,” Radiol. Phys. Technol., pp. 24–46, 2024, doi: 10.1007/s12194-024-00780-3.
5. C. Tian, M. Zheng, W. Zuo, B. Zhang, Y. Zhang, and D. Zhang, “Multi-stage image denoising with the wavelet transform,” Pattern Recognit., vol. 134, pp. 1–33, 2023, doi: 10.1016/j.patcog.2022.109050.
6. W. Wu, S. Liu, Y. Zhou, Y. Zhang, and Y. Xiang, “Dual Residual Attention Network for Image Denoising,” 2023.
7. J. Yuan, F. Zhou, Z. Guo, X. Li, and H. Yu, “HCformer : Hybrid CNN - Transformer for LDCT Image Denoising,” J. Digit. Imaging, vol. 36, no. 5, pp. 2290–2305, 2023, doi: 10.1007/s10278-023-00842-9.
8. N. You, L. Han, D. Zhu, and W. Song, “Research on Image Denoising in Edge Detection Based on Wavelet Transform,” Appl. Sci., vol. 13, no. 3, 2023, doi: 10.3390/app13031837.

9. V. Kulikov, S. Yadin, M. Kleiner, and T. Michaeli, "SinDDM : A Single Image Denoising Diffusion Model," pp. 1–11, 2023.
10. J. J. Huang and P. L. Dragotti, "WINNet: Wavelet-Inspired Invertible Network for Image Denoising," IEEE Trans. Image Process., vol. 31, pp. 4377–4392, 2022, doi: 10.1109/TIP.2022.3184845.
11. K. Zhang et al., "Practical Blind Image Denoising via Swin-Conv-UNet and Data Synthesis," vol. 20, no. December, pp. 822–836, 2023, doi: 10.1007/s11633-023-1466-0.
12. S. Izadi, D. Sutton, and G. Hamarneh, "Image Denoising in the Deep Learning Era," 2022.
13. A. Lee, E. Hwang, and D. Kim, "A Practical CNN–Transformer Hybrid Network for Real-World Image Denoising," Mathematics, vol. 14, no. 1, 2026, doi: 10.3390/math14010203.
14. C. Jang, D. Ji, and N. I. Cho, "Boosting the Performance of Image Restoration Models Through Training With Deep-Feature Auxiliary Guidance," IEEE Access, vol. 13, no. October, pp. 173398–173414, 2025, doi: 10.1109/ACCESS.2025.3617488.
15. H. Tang, W. Zhang, H. Zhu, and K. Zhao, "Self-Supervised Real-World Image Denoising Based on Multi-Scale Feature Enhancement and Attention Fusion," IEEE Access, vol. 12, pp. 49720–49734, 2024, doi: 10.1109/ACCESS.2024.3384577.
16. D. Zhang and F. Zhou, "Self-Supervised Image Denoising for Real-World Images With Context-Aware Transformer," IEEE Access, vol. 11, no. February, pp. 14340–14349, 2023, doi: 10.1109/ACCESS.2023.3243829.
17. Y. Liu, S. Anwar, Z. Qin, P. Ji, S. Caldwell, and T. Gedeon, "Disentangling Noise from Images: A Flow-Based Image Denoising Neural Network," Sensors, vol. 22, no. 24, pp. 1–16, 2022, doi: 10.3390/s22249844.
18. S. Cho, Z. Chen, and X. Li, "The Devil is in the Dark Pixels: Toward Brightness Bias-Robust Denoising," pp. 1–17, 2026.
19. R. Liu et al., "MSTNet: a multi-stage progressive network with local–global transformer fusion for image restoration," Complex Intell. Syst., vol. 11, no. 6, pp. 1–16, 2025, doi: 10.1007/s40747-025-01892-y.
20. W. Tiantian, Z. Hu, and Y. Guan, "An efficient lightweight network for image denoising using progressive residual and convolutional attention feature fusion," Sci. Rep., vol. 14, no. 1, pp. 1–20, 2024, doi: 10.1038/s41598-024-60139-x.
21. J. Ko, B. W. Lim, S. Lee, D. Kim, and T. H. Kim, "YeTI: You Only Need Two Noisy Images for Real-World sRGB Noise Generation," 2026.
22. D. Ryou, I. Ha, S. Chu, and B. Han, "Beyond the Ground Truth: Enhanced Supervision for Image Restoration," 2025.
23. S. Son and K. M. Lee, "Self-Supervised Image Denoising by Disentangling Noise from Image," 2022.
24. W. Lee, S. Son, and K. M. Lee, "AP-BSN : Self-Supervised Denoising for Real-World Images via Asymmetric PD and Blind-Spot Network," pp. 17725–17734, 2022.

25. H. Tan, “Noise2Variance : Dual networks with variance constraint for self-supervised real-world image denoising,” no. May, pp. 3251–3261, 2024, doi: 10.1049/ipr2.13170.