Super Resolution Image Denoiser: Toward Real-world Noise Removal

P. Mohit Harsh
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
21951a04a3@iare.ac.in

V. Teja Vardhan
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
21951a04q8@iare.ac.in

V. Vishnu Vardhan
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
21951a04q9@iare.ac.in

J. Manoj Naidu
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
21951a04r4@iare.ac.in

Dr. V. Padmanabha Reddy
Electronics and Communication Engineering
Institute of Aeronautical Engineering
Hyderabad, India
v.padmanabhareddy@iare.ac.in

***Abstract*—In recent years, deep learning has significantly advanced image denoising techniques, with methods such as transformers and GANs pushing the boundaries of performance. However, these state-of-the-art models are typically resource-intensive, requiring vast computational power, large datasets, and extensive training time, which also hampers real-time inference speed. To address these challenges, we propose the Super-Resolution Image Denoiser Network (SRIDNet), a lightweight yet highly effective model designed to achieve near state-of-the-art results with a fraction of the data and computational resources. SRIDNet significantly reduces the model size while maintaining competitive performance in image denoising and super-resolution tasks. We evaluated our model on multiple datasets, where it demonstrated high PSNR scores and the fastest inference times compared to existing models. Our results confirm that SRIDNet offers an efficient solution for real-time image denoising, making it a practical alternative to more cumbersome models without compromising on quality.**

***Index Terms*—Convolution Neural Network (CNN), Adversarial Neural Network (GAN), SRIDNet, PSNR, Transformers.**

I. Introduction

Image denoising is a critical preprocessing step in many computer vision and image processing applications, aimed at restoring corrupted images by reducing noise while preserving important structural details. Traditional denoising techniques, such as Gaussian filtering and wavelet transforms, have long been employed for this task but are often limited by their inability to adapt to complex and diverse noise patterns encountered in real-world images. In recent years, the rise of deep learning has revolutionized the field of image denoising, offering powerful data-driven solutions capable of learning intricate noise distributions and producing visually appealing results.

Convolutional Neural Networks (CNNs), in particular, have become the dominant paradigm for image denoising, leveraging their ability to capture local spatial hierarchies through multiple layers of convolutional filters. These models have significantly outperformed classical approaches, achieving state-of-the-art results by learning from large datasets of noisy and clean image pairs. Recent advancements have also introduced more sophisticated architectures such as Generative Adversarial Networks (GANs) and Transformer models, further improving denoising performance by capturing global dependencies and generating more realistic outputs.

Despite their impressive results in controlled environments, these state-of-the-art models often come with high computational costs, extensive memory requirements, and the need for large training datasets. These factors limit their practicality in real-world applications, particularly in resource-constrained environments where hardware and data availability are restricted.

To address these challenges, we propose the Super Resolution Image Denoiser Network (SRIDNet), a novel approach that strikes a balance between denoising performance and computational efficiency. SRIDNet is designed to deliver competitive results with significantly lower hardware demands and reduced model complexity, making it suitable for deployment in practical settings. By integrating super-resolution techniques with an efficient denoising mechanism, SRIDNet provides a robust solution for noise reduction without sacrificing speed or resource efficiency. This paper presents the architecture,

training methodology, and performance evaluation of SRID-Net, showcasing its advantages in terms of both accuracy and efficiency across various datasets and noise levels.

II. Literature Survey

Over the past decade, deep convolutional neural networks (CNNs) have significantly advanced high-level vision tasks like visual recognition, motion analysis, and object segmentation. More recently, CNNs have been applied to low-level vision tasks, such as super-resolution (SR), image denoising, and compression artifact reduction, where they are trained to map low-quality images to high-quality outputs, typically aiming to remove noise or minimize artifacts. While "deeper is better" is a widely accepted principle in high-level vision tasks—evidenced by networks like VGG, GoogleNet, and ResNet achieving substantial breakthroughs—this principle has not shown as much impact in low-level vision tasks. Despite the use of networks with 20 to 30 layers, such as DnCNN [1] and RED-Net, the performance gains in low-level tasks have been modest compared to earlier methods. This is because low-level vision tasks rely more on pixel-level features, where depth is less crucial. Instead, statistical priors, like non-local similarity or pixel distribution patterns (e.g., Gaussian noise), play a key role in enhancing the accuracy of these tasks, offering a more effective solution to image degradation issues.

A. Existing Work

In CBDNet [2], an asymmetric loss function was employed to enhance the model's ability to generalize to real-world noise scenarios, while also facilitating convenient interactive denoising. For training, they utilized a dataset comprising 400 images from BSD500 [4], 1600 images from Waterloo [11], and 1600 images from the MIT-Adobe FiveK dataset [10]. A batch size of 32 was selected, with each patch being 128 × 128 in size. The model was trained over 40 epochs, starting with a learning rate of 10^{-3} for the first 20 epochs, followed by a learning rate of 5×10^{-4} for fine-tuning during the remaining epochs. The model demonstrated PSNR scores of 30.78 on the SIDD dataset [5] and 38 on the DND dataset. Processing a 512x512 image takes approximately 0.4 seconds.

In their work on RIDNet [3], the authors introduce a CNN-based denoising model specifically designed for both synthetic noise and real-world noisy images. This model is a single-blind denoising network for real noisy images, differing from prior methods. To enhance the network's ability to learn and improve feature extraction, they incorporated a restoration module along with feature attention, which adjusts the channel-wise features by considering the interdependencies between channels. Additionally, they implemented LSC, SSC, and SC mechanisms, allowing low-frequency information to bypass the network, enabling it to focus on learning residuals. The model's architecture consists of four Enhanced Attention Mechanism (EAM) blocks, where most convolutional layers have a kernel size of 3x3, except for the final layer in the enhanced residual block and feature attention units, which utilize a 1x1 kernel. To ensure feature map dimensions remain consistent, the model applies zero-padding to the 3x3 convolutions. Each layer operates with 64 channels, and a downscaling factor of 16 is applied in the feature attention process, reducing the number of feature maps. The model processes a 512x512 image in approximately 0.2 seconds during evaluation and achieved PSNR scores of 31.38, 39.23, and 38.71 on the BSD68 [4], DnD [6], and SIDD [5] datasets, respectively.

The NBNet [7] architecture is built upon a modified UNet framework. NBNet incorporates four encoder and decoder stages, where feature maps are downsampled using strided convolutions in the encoder and upsampled using deconvolutions in the decoder. Skip connections between encoder and decoder stages transfer low-level features, and the primary innovation lies in the introduction of Subspace Attention (SSA) modules within these skip connections. Unlike conventional UNet architectures, where low- and high-level feature maps are directly fused, NBNet leverages SSA modules to project low-level features into a signal subspace guided by upsampled high-level features before fusion. This projection allows the model to better capture global structure information while preserving local details, improving denoising performance. It achieved PSNR scores of 29.16, 39.62 and 39.75 on BSD68 [4], DnD [6] and SIDD [5] datasets respectively.

DANet [2] is the latest architecture introduced for real-world image denoising tasks. Its core concept revolves around unfolding in-camera processing pipelines or learning the noise distribution through a generative adversarial network (GAN). In this dual adversarial framework, three components require optimization: the denoiser (R), the generator (G), and the discriminator (D). To stabilize training, they incorporated the gradient penalty technique from WGAN-GP [3], ensuring the discriminator adheres to a 1-Lipschitz constraint by adding an additional gradient penalty term, d. DANet includes two hyperparameters—one primarily affecting the performance of the denoiser R, while the other directly impacts the generator G. This architecture achieved PSNR scores of 39.25 and 39.79 on the SIDD and DND benchmark datasets, respectively.

B. Evaluation Metrics

In image processing, Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM) are commonly used metrics to assess the quality of images, particularly for evaluating how similar a processed image is to its reference image.

1) 1. Mean Squared Error (MSE): MSE is a pixel-based metric that calculates the average of the squared differences between corresponding pixels of two images—typically, an original (reference) image and a processed (distorted or denoised) image.

$$MSE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I(i,j) - K(i,j))^2$$

Where:

- $I(i,j)$ is the pixel value at position (i,j) in the original image.
- $K(i,j)$ is the pixel value at position (i,j) in the processed image.
- m and n are the image dimensions.

Interpretation:

- Lower MSE indicates less error, meaning the processed image is closer to the original.
- Higher MSE implies a larger difference between the two images.

However, MSE doesn't take human visual perception into account and is sensitive to large differences, even if small changes are hard to perceive.

2) 2. Peak Signal-to-Noise Ratio (PSNR): PSNR is a more perceptually aligned metric that expresses the ratio between the maximum possible pixel value of the image and the error (noise) introduced by the processing, measured via MSE.

$$PSNR = 10 \times \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

Where:

- MAX_I is the maximum possible pixel value (255 for an 8-bit image).
- MSE is the mean squared error between the two images.

Interpretation:

- Higher PSNR indicates that the processed image is of better quality and closer to the original.
- A PSNR above 30 dB is generally considered acceptable for most image processing tasks, though in applications like medical imaging, higher values are preferred.

PSNR is easy to compute and gives a high-level indication of image quality, but it still does not correlate well with human visual perception.

3) 3. Structural Similarity Index Measure (SSIM): SSIM is designed to better reflect the human perception of image quality by considering changes in structural information. It compares luminance, contrast, and structural properties between two images.

$$SSIM(I,K) = \frac{(2\mu_I\mu_K + C_1)(2\sigma_{IK} + C_2)}{(\mu_I^2 + \mu_K^2 + C_1)(\sigma_I^2 + \sigma_K^2 + C_2)}$$

Where:

- μ_I and μ_K are the means of the original and processed images, respectively.
- σ_I^2 and σ_K^2 are the variances of the original and processed images, respectively.
- σ_{IK} is the covariance between the two images.
- C_1 and C_2 are constants to stabilize the division in case of weak denominator values.

Interpretation:

- SSIM values range from -1 to 1, with 1 indicating perfect similarity and 0 or negative values representing no correlation.
- SSIM is more aligned with human visual perception as it emphasizes structure and texture rather than pixel-by-pixel differences.

4) Summary of the Metrics::

- *MSE* focuses purely on pixel-wise differences.
- *PSNR* relates the error to the maximum possible signal value, providing a better general indication of quality.
- *SSIM* captures perceptual quality by evaluating structural similarity, making it more sensitive to changes that are noticeable to the human eye.

SSIM is often considered a more reliable metric for human-perceived image quality, while MSE and PSNR are useful for mathematical comparison and quick estimation of errors.

C. Datasets

1) SIDD [5]: In recent years, there has been a significant shift from DSLR and point-and-shoot cameras to smartphone imaging, where the smaller apertures and sensor sizes of smartphones result in higher noise levels. Despite extensive research on smartphone image denoising, the field lacked a comprehensive dataset of real noisy images with corresponding high-quality ground truth. To address this gap, this dataset presents a systematic approach for estimating ground truth for noisy smartphone images, creating a new dataset—the Smartphone Image Denoising Dataset (SIDD)—comprising approximately 30,000 images from 10 scenes under varying lighting conditions, captured using five different smartphone cameras. This dataset enables benchmarking of denoising algorithms, and it has been demonstrated that CNN-based methods trained on this high-quality dataset outperform models trained with alternative strategies, such as using low-ISO images as proxy ground truth.

2) BSD [4]: The Berkeley Segmentation Dataset (BSD) [4] is a widely used collection of images for evaluating segmentation and image processing algorithms, including image denoising. It contains several subsets with varying numbers of images, each providing a rich source of diverse natural images.

1. BSD68 [4]: This subset consists of 68 images selected from the larger dataset. It is often used as a benchmark for evaluating image segmentation and denoising algorithms due to its manageable size and representative diversity.

2. BSD100 [4]: This version includes 100 images, expanding on the BSD68 [4] set. The increased number of images allows for more comprehensive testing and validation of algorithms, offering a broader range of textures, colors, and structures.

3. BSD300 [4]: This dataset features 300 images, providing an even more extensive set of natural images. The larger size enhances the dataset's ability to assess the performance of denoising methods across a wider variety of scenarios, making it useful for researchers looking to develop and test robust algorithms.

Overall, these datasets serve as standard benchmarks for evaluating the effectiveness of image denoising techniques and help researchers compare their methods against established results in the field.

3) Urban100 [9]: The dataset consists of 100 high-resolution images that capture various urban scenes, including buildings, streets, and other architectural features. These images contain a range of textures, colors, and structures, which make them suitable for testing the robustness of denoising algorithms. The dataset consists of 100 high-resolution images that capture various urban scenes, including buildings, streets, and other architectural features. These images contain a range of textures, colors, and structures, which make them suitable for testing the robustness of denoising algorithms. The dataset includes a wide variety of scenes, ensuring that models trained or tested on Urban100 [9] can generalize well across different types of urban imagery. While the original images are clean, researchers often add synthetic noise (e.g., Gaussian noise) at various levels to create noisy versions for denoising tasks. This setup allows for a controlled evaluation of how well different algorithms can recover the original clean images from their noisy counterparts.

D. Drawbacks of existing methods

1) High Computational Complexity and Resource Demand: State-of-the-art denoising models typically consist of millions of trainable parameters, leading to significant computational demands. While these models perform exceptionally well in controlled laboratory environments, their need for vast datasets and computational resources makes them impractical for many real-world applications.

2) Inefficiency in Image Processing Speed: Current state-of-the-art models require 0.2 to 0.4 seconds to denoise a single 512x512 image. This processing time limits their usability in scenarios where real-time or high-speed image processing is essential.

III. Proposed Architecture

To mitigate the aforementioned limitations, we propose the Super-Resolution Image Denoiser Network (SRIDNet), a lightweight architecture designed for high speed denoising of images while maintaining the state of the art results. This Convolutional Neural Network (CNN) architecture is split into two primary blocks: the Super-Resolution Block and the Denoiser Block, with each serving a distinct purpose while contributing to the overall objective of producing high-quality, denoised images.

A. Input Layer

Input Shape: $(112, 112, 3)$

The model accepts an input image of size 112×112 with three color channels (RGB). This compact image resolution is chosen to reduce computational complexity and memory requirements while ensuring efficiency in model training and inference.

B. Super-Resolution Block

Output Shape: $(224, 224, 3)$

Trainable Parameters: 110,515

The Super-Resolution Block is responsible for upscaling the input image from (112×112) to (224×224) by leveraging a series of *convolutional layers* and residual connections. This block effectively doubles the spatial resolution of the image and forms the encoder part of the architecture.

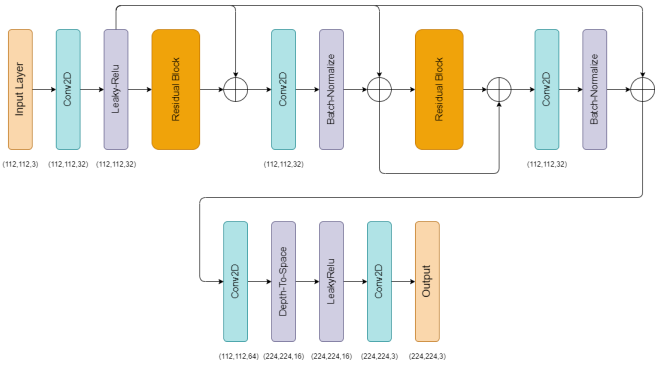

Fig. 1. Super-Resolution Block

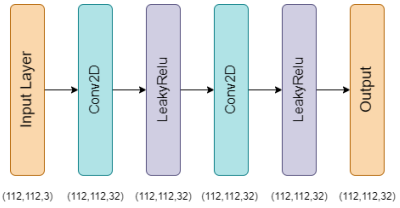

Fig. 2. Residual Block

- *Initial Convolution Layer:* A *Conv2D* layer with 32 filters and a kernel size of 3×3 is applied to the input, followed by a *LeakyReLU activation*. This step extracts low-level features from the image.
- Residual Blocks: Two *Residual Blocks* are employed. Each block consists of two convolutional layers with a kernel size of 3×3 and 32 filters, followed by *Batch Normalization* and *LeakyReLU* activations. These blocks capture intricate features while addressing the vanishing gradient problem. Skip connections are added to improve feature flow and gradient propagation, enhancing model training.
- *Final Convolution and Upsampling:* The final convolutional layer in the Super-Resolution Block has 64 filters, followed by a *Depth-to-Space operation* (also known as Pixel Shuffling), which rearranges the tensor to increase its spatial resolution to 224×224. This technique provides an efficient way to upscale the image.
- *Output:* A *Conv2D* layer with 3 filters reconstructs the RGB image, which now has double the spatial dimensions (224×224) compared to the input.

C. Denoiser Block

Output Shape: $(112, 112, 3)$

Trainable Parameters: 344,259

The Denoiser Block performs the task of noise removal by processing the upscaled image from the Super-Resolution Block. This component can be considered a *decoder block*, reducing the resolution back to (112×112) while preserving

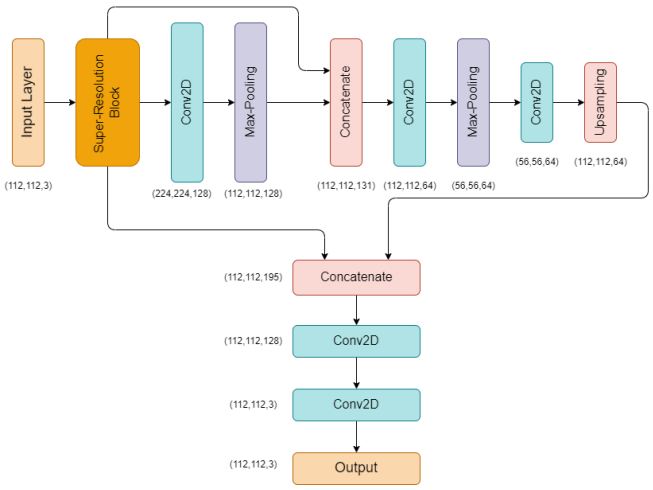

Fig. 3. Model Architecture

important visual features through skip connections and convolution operations.

- *Convolution Layers:* The first convolution layer in the Denoiser Block has 128 filters, followed by a *MaxPooling2D* layer that downscales the image to (112×112) for processing.
- *Concatenation with Input:* A skip connection is established where the original input image is concatenated with the downsampled version of the upscaled image, producing an intermediate feature map of size $(112 \times 112 \times 131)$. This fusion ensures the model retains essential features from the initial input.
- *Downsampling and Upsampling:*
  - The image is further downsampled using another convolutional layer with 64 filters, followed by another *MaxPooling2D* operation that reduces the spatial size to (56×56).
  - An *UpSampling2D* layer increases the resolution back to (112×112).
- *Concatenation with Intermediate Features:* Skip connections are introduced again by concatenating the upsampled feature map with previous intermediate layers, producing a feature map of size $(112 \times 112 \times 195)$.
- *Final Convolutions:* The concatenated feature map is passed through a convolutional layer with 128 filters and finally reduced to 3 channels (RGB) using a *Conv2D* layer with 3 filters. The output is the denoised image, reconstructed to its original resolution of (112×112).

D. Working Principle

- The architecture leverages the *Super-Resolution Block* to enhance the resolution of the input image from (112×112) to (224×224). This block employs a combination of *convolutional layers*, *residual connections*, and *depth-to-space transformation* to ensure that the upscaled image retains fine details without introducing significant artifacts.
- Following this, the *Denoiser Block* takes over, utilizing downsampling and upsampling techniques, combined with skip connections, to effectively remove noise from the upscaled image. The final result is a clean, denoised image at the original resolution of (112×112).
- *Skip connections* play a crucial role in both blocks by preserving critical features from the earlier stages and preventing information loss during downsampling and upsampling.

E. Summary of Parameters

- *Total Parameters:* 454,902
- *Trainable Parameters:* 454,774
- *Non-trainable Parameters:* 128

This architecture is highly optimized for the task of image denoising, leveraging the efficiency of the *Super-Resolution Block* to work with smaller input patches, significantly reducing the amount of data and time required for training and inference. Furthermore, the *Denoiser Block* ensures that the model can effectively remove noise while retaining critical image features, providing high-quality denoised images as output.

IV. Implementation

A. Data Preprocessing

1) SIDD [5]: In our study, we utilized images from the SIDD [5] dataset, resizing them to dimensions of (2576, 1456, 3). From each image, we extracted patches with a resolution of (112, 112, 3). For training, we randomly selected 7,000 patches, while 3,000 patches were allocated for validation. The final 10 images from the dataset were reserved for testing, which, after patching, yielded 2,990 test patches.

2) Urban100 [9]: In our study, we utilized the Urban100 [9] dataset to extract a total of 80 images for the training set. These images were cropped to a size of $560 \times 560 \times 3$ pixels. Subsequently, we generated patches of size $112 \times 112 \times 3$ from each image, resulting in a total of 2,000 patches designated for training with a validation split of 0.1. For the testing phase, we selected 20 remaining images from the dataset, which were similarly cropped to $560 \times 560 \times 3$ pixels. Following the patch generation process, we obtained 500 patches of size $112 \times 112 \times 3$ for testing purposes.

B. Model Training

The Super-Resolution Image Denoiser Network (SRIDNet) was trained using the TensorFlow and Keras frameworks on image patches of size (112, 112, 3). The training process was carried out in three stages, each with progressively reduced learning rates to enhance model performance and stability. Initially, the model was trained for 20 epochs with a learning rate of 0.001. This was followed by an additional 20 epochs at a reduced learning rate of 0.0001, and a final 10 epochs at a further reduced learning rate of 0.00001.

The Adam optimizer was employed to minimize the mean squared error (MSE) loss function, with Peak Signal-to-Noise Ratio (PSNR) used as the evaluation metric. The batch size

was set to 16 for all training stages. On the Urban100 dataset, the model required approximately 45 seconds per epoch, while on the SIDD dataset, it required 130 seconds per epoch. These timings reflect the computational complexity and dataset size differences between the two datasets.

V. Results and Comparisons

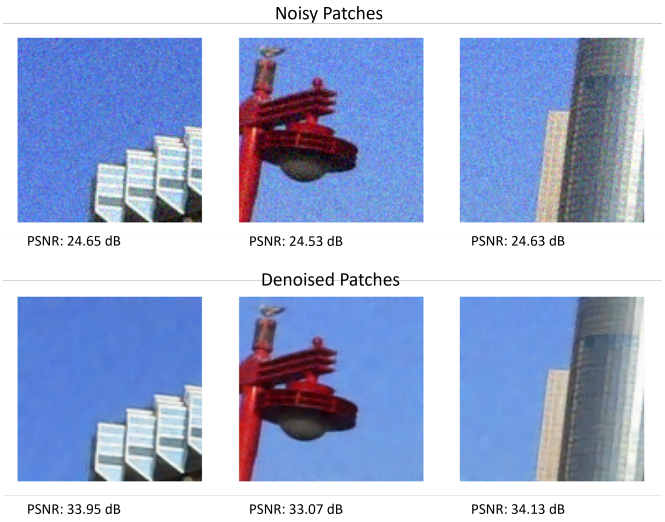

Fig. 4. PSNR results on $112 \times 112 \times 3$ image patches ($\sigma = 15$) from Urban100 dataset

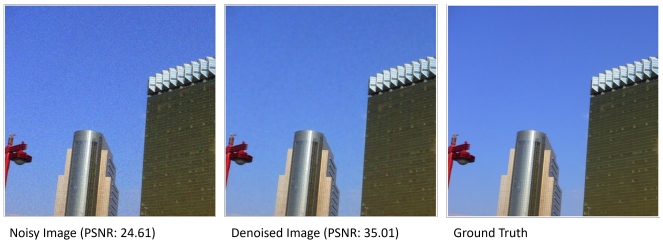

Fig. 5. PSNR results on an entire image ($\sigma = 15$) of size $560 \times 560 \times 3$ from Urban100 dataset

```
1/1 [==============================] - 0s 24ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 23ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 20ms/step
1/1 [==============================] - 0s 22ms/step
1/1 [==============================] - 0s 19ms/step
Avg PSNR:  tf.Tensor(34.230599650363615, shape=(), dtype=float64) Max PSNR:  tf.Tensor(35.81651203385877, shape=(),
dtype=float64)
```

Fig. 6. Avg. PSNR and evaluation time per image in tensorflow and keras for images at noise level ($\sigma = 15$) from Urban100 dataset

The proposed Super-Resolution Image Denoiser Network (SRIDNet) was evaluated on the Urban100 [9] dataset, with the last 10 images used as the test set. The evaluation process was conducted with TensorFlow, and inference was completed in an average time of 20-24 ms per image of size $560 \times 560 \times 3$ at noise level ($\sigma = 15$). The model's performance was assessed using Peak Signal-to-Noise Ratio (PSNR) as the evaluation metric. The results indicate that SRIDNet achieved an average PSNR of 34.23 dB across the test images, with a maximum PSNR of 35.82 dB.

Model	σ	PSNR(dB)
DnCNN	15	32.98
	25	30.81
	50	27.59
FFDNet	15	33.83
	25	31.40
	50	28.05
DRUNet	15	34.81
	25	32.60
	50	29.61
SwinIR	15	35.13
	25	32.90
	50	29.82
ours	15	**34.23**
	25	**32.01**
	50	**28.67**

TABLE I
URBAN100 IMAGE DENOISING RESULTS AND COMPARISON WITH EXISTING MODELS

The proposed Super-Resolution Image Denoiser Network (SRIDNet) was evaluated on the Urban100 dataset and compared against state-of-the-art denoising models, including DnCNN [1], FFDNet [12], DRUNet [13], and SwinIR [14], across different noise levels ($\sigma = 15, 25, 50$). The evaluation metric is Peak Signal-to-Noise Ratio (PSNR) as presented in table 1.

```
10/10 [==============================] - 1s 104ms/step
10/10 [==============================] - 1s 91ms/step
10/10 [==============================] - 1s 91ms/step
10/10 [==============================] - 1s 91ms/step
10/10 [==============================] - 1s 91ms/step
10/10 [==============================] - 1s 91ms/step
10/10 [==============================] - 1s 91ms/step
10/10 [==============================] - 1s 91ms/step
10/10 [==============================] - 1s 91ms/step
10/10 [==============================] - 1s 91ms/step
tf.Tensor(39.50148010907847, shape=(), dtype=float64)
```

Fig. 7. Avg. PSNR and evaluation time of SIDD images in tensorflow and keras

The proposed model was evaluated on the SIDD dataset and benchmarked against several state-of-the-art denoising models, including DANet [16], VDN [15], RIDNet [3], and CBDNet [2]. The model achieved an average PSNR of 39.5 on the SIDD dataset, with an average inference time of 95 ms for images sized 2576×1456, which were divided into 299 patches of size 112×112

Figures 8 and 9 illustrate the denoising results of low-light and bright-light images from the SIDD [5] dataset, respectively, using $(112 \times 112 \times 3)$ image patches.

Table II presents a comparison of denoising performance across different models on the SIDD [5] dataset, showing that our model outperformed others by achieving a higher average PSNR. While DnCNN [1] and FFDNet [12]—models

Fig. 8. Denoising results on SIDD dataset images

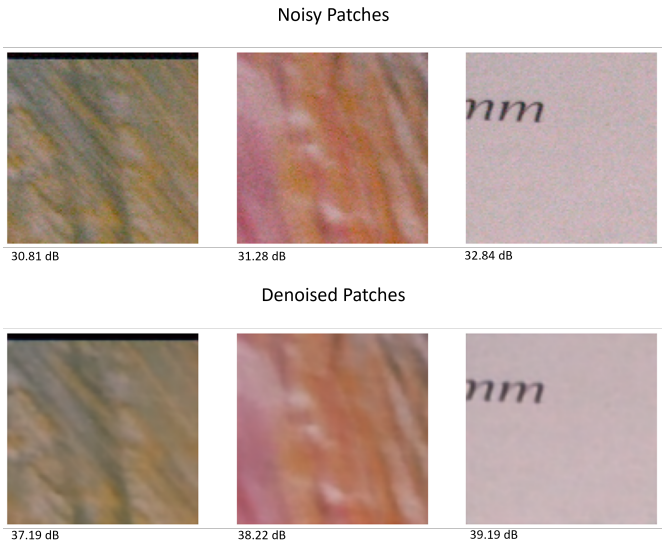

Fig. 9. Denoising results on SIDD dataset images

of similar size—performed well on images with additive Gaussian white noise (AGWN) from Urban100 [9] dataset, they underperformed on real-world photographic images from the SIDD [5] dataset. In contrast, our model demonstrated robust performance across both datasets with greater efficiency.

Table III compares the inference times of the models evaluated in this study. While our model is faster than all models except FFDNet [12], it offers more consistent results on both the Urban100 and SIDD [5] datasets, where FFDNet [12] struggles with real-world images.

VI. Conclusion

In conclusion, this research presents SRIDNet, a novel deep learning architecture for image denoising that effectively addresses the limitations of current state-of-the-art models, such as high computational demands, extensive training data requirements, and slow inference times. By designing a lightweight model that integrates super-resolution and denoising tasks, SRIDNet achieves competitive results with significantly reduced resource consumption. Extensive testing on both synthetic (AGWN) and real-world noisy images from the Urban100 and SIDD datasets demonstrates the model's robustness and efficiency. With PSNR scores of 34.23 on Urban100 and 39.50 on SIDD, SRIDNet delivers near state-of-the-art performance while maintaining the smallest model size and fastest inference time. This balance of efficiency and effectiveness positions SRIDNet as a promising solution for practical image denoising applications, particularly in environments with limited computational resources.

Model	PSNR (dB)
CBDNet	30.78
RIDNet	38.71
VDN	39.28
DANet	39.47
DnCNN	26.21
FFDNet	29.20
Ours	**39.50**

TABLE II
Comparison of results with state of the art models on SIDD [5] dataset

Model	Inference Time (s)
DnCNN	0.0314
FFDNet	0.0071
DRUNet	0.0733
SwinIR	-
CBDNet	0.4
RIDNet	0.2
VDN	-
DANet	-
Ours	**0.0208**

TABLE III
Comparison of inference time with different models.

VII. Acknowledgment

We would like to extend our heartfelt gratitude to those who made this project possible through their constant guidance and support. Our deepest appreciation goes to our guide, Dr. V. Padmanabha Reddy, Assistant Professor, for his invaluable guidance and continuous cooperation throughout the project. We also want to thank the teaching and non-teaching staff for their support, and express our deepest gratitude to our parents, friends, and well-wishers for their assistance in completing this project report successfully.

References

[1] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing, 26:3142–3155, 2017.

[2] S. Guo, Z. Yan, K. Zhang, W. Zuo, L. Zhang: Toward Convolutional Blind Denoising of Real Photographs. In: CVPR (April 2019).

[3] S. Anwar, N. Barnes.: Real Image Denoising with Feature Attention. In: ICCV (March 2020).

[4] Stefan Roth and Michael J Black. Fields of experts. IJCV, 2009. 1, 2, 5, 6, 7

[5] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In CVPR, 2018. 5, 8

[6] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. arXiv preprint arXiv:1707.01313, 2017. 5, 7

[7] S. Cheng1, Y. Wang1, H. Huang, D. Liu, H. Fan and S. Liu.: NBNet: Noise Basis Learning for Image Denoising with Subspace Projection. In: CVPR (May 2021).

[8] Sara, U. , Akter, M. and Uddin, M. (2019) Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. Journal of Computer and Communications, 7, 8-18. doi: 10.4236/jcc.2019.73002.

[9] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), June 2015.

[10] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in Proc. 24th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2011.

[11] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo Exploration Database: New challenges for image quality assessment models," IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 1004-1016, Feb. 2017.

[12] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. TIP, 2018. 1, 2, 5, 6, 7, 8

[13] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021. 1, 2, 7, 8

[14] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image Restoration Using Swin Transformer," arXiv preprint arXiv:2108.10257, 2021.

[15] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, "Variational denoising network: Toward blind noise modeling and removal," in Advances in Neural Information Processing Systems 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 1690-1701. Available: http://papers.nips.cc/paper/8446-variational-denoising-network-toward-blind-noise-modeling-and-removal.pdf.

[16] Z. Yue, Q. Zhao, L. Zhang, and D. Meng, "Dual adversarial network: Toward real-world noise removal and noise generation," in Proc. Eur. Conf. Comput. Vis. (ECCV), Aug. 2020.