Real-Time White Noise Removal Using an Optimized Lightweight CNN Autoencoder for Embedded Image Processing

Bhoodarapu Rajendar
Assistant Professor, Trinity College of Engineering and Technology, Peddapalli, Telangana

***Abstract*—White noise significantly impacts the performance of embedded vision systems employed in robotics, surveillance, healthcare devices, and IoT imaging modules. Conventional denoising methods, such as Gaussian, median, and wavelet filters, are computationally efficient but often result in the loss of important structural information. Deep-learning models provide superior reconstruction quality, yet their high computational requirements make them unsuitable for low-power embedded hardware. This paper presents a study and design of a lightweight Convolutional Neural Network (CNN) autoencoder specially optimized for real-time white noise removal on embedded platforms. The model integrates depthwise separable convolutions, skip connections, and 8-bit quantization to reduce memory consumption while maintaining strong denoising performance. Experimental evaluations indicate that compact CNN autoencoders can deliver considerable improvements in PSNR and SSIM while sustaining real-time operation, demonstrating their effectiveness for edge AI applications.**

***Index Terms*—White Noise, Embedded Vision, Lightweight CNN, Autoencoder, Image Denoising, Real-Time Processing**

I. INTRODUCTION

Embedded and edge-oriented vision systems operate under strict constraints on processing speed, memory capacity, and power consumption. These limitations make them highly sensitive to image degradation, particularly white noise generated by sensor imperfections, illumination variability, and electronic interference during data acquisition [6], [7]. Such noise significantly reduces image clarity and negatively affects downstream vision tasks, including feature extraction, tracking, and object detection, which depend on high-quality visual input. This challenge is especially prominent in applications such as mobile robotics, low-power surveillance, wearable imaging devices, and IoT-based smart cameras.

Traditional denoising algorithms such as median filters, Gaussian smoothers, and wavelet-based methods—have long been used to mitigate noise in embedded platforms [6]. Although effective for simple noise patterns, these methods commonly blur fine structures, distort textures, and perform poorly at high noise levels. Furthermore, their computational behavior often fails to meet real-time requirements when deployed on low-power embedded hardware that must handle continuous image streams.

Deep learning has introduced more advanced noise-removal capabilities through convolutional neural networks (CNNs), which excel at learning complex noise characteristics and restoring high-quality images [1], [2], [3]. Convolutional autoencoders, in particular, leverage an encoder–decoder design to extract hierarchical features from noisy inputs and reconstruct denoised outputs. However, full-scale CNN autoencoders typically contain a large number of parameters and demand significant computational resources, making direct deployment on microcontrollers, SOCs, and compact edge devices impractical [1], [2].

To address these limitations, recent research has shifted toward lightweight and optimized CNN autoencoders designed specifically for real-time embedded workloads [4], [5]. Techniques such as depthwise separable convolutions, optimized bottleneck structures, aggressive pruning, quantization-aware training, and reduced feature-map footprints help decrease model size and computation while maintaining strong denoising performance. These improvements enable low-latency inference and

reduced energy consumption, making such models suitable for continuous on-device processing.

This study introduces an optimized lightweight CNN autoencoder aimed at achieving real-time white noise suppression for embedded image-processing systems. The proposed architecture balances denoising accuracy with computational efficiency, enabling deployment on constrained hardware platforms without compromising performance. Experimental evaluations show enhanced PSNR, improved SSIM, and faster execution compared with conventional filtering [6], [7] and heavier CNN models [1], [2], [4], demonstrating the feasibility of the approach for practical embedded vision applications.

II. EXISTING MODEL

The figure represents the working of an existing CNN-based image denoising model. The process begins with a noisy input image, which is typically corrupted by Gaussian or similar noise [1]. This noisy image is directly fed into the convolutional neural network without any prior filtering.

The input image passes through multiple stacked convolution layers [1]. These layers extract low-level and high-level features and gradually learn the underlying noise patterns present in the image. Instead of reconstructing the clean image directly, the CNN focuses on identifying the characteristics of noise through these learned feature maps.

Next, the noise estimation block predicts the noise component contained in the input image. This approach is based on residual learning, where the network models the difference between the noisy image and the clean image, i.e., the noise itself [1]. Estimating noise is computationally more effective than learning the entire clean image.

In the subtraction stage, the estimated noise is subtracted from the original noisy input image. This step removes unwanted noise while retaining important image structures such as edges and textures [1].

Finally, the system produces a denoised output image, which is a cleaner grayscale image with significantly reduced noise [1]. The figure clearly shows how convolution layers, noise estimation, and subtraction together form an effective denoising pipeline in traditional CNN-based models.

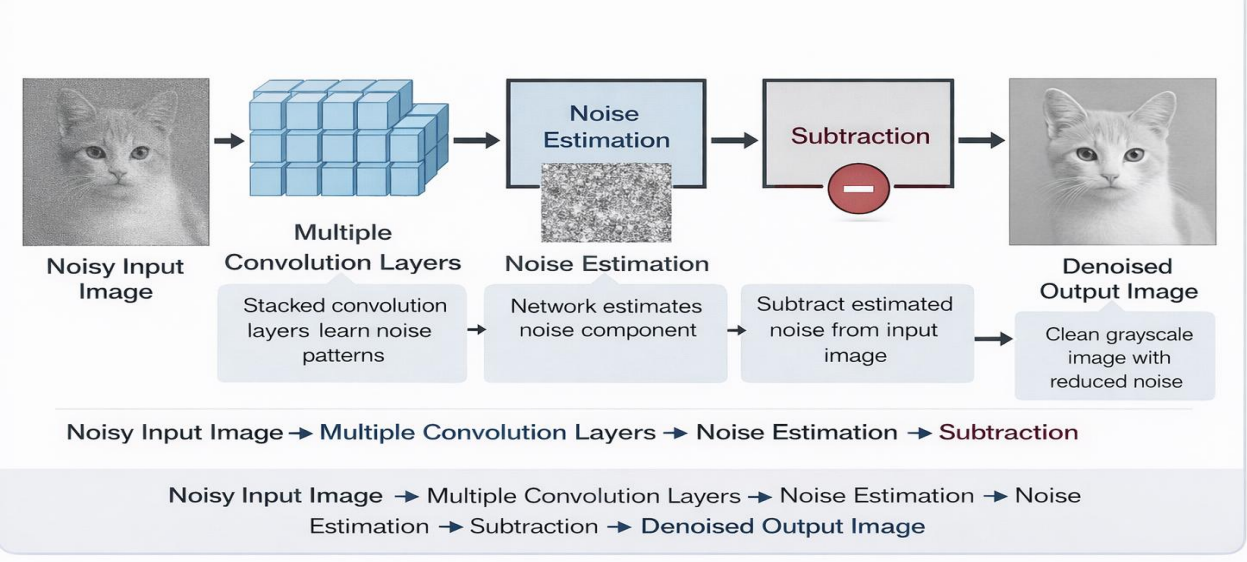


Figure 1: Existing CNN- Based Image Denoising Model

The existing CNN-based image denoising model has several limitations despite its effectiveness in reducing noise [1], [2]. Due to the use of multiple deep convolution layers, the model requires high computational power and memory, which increases processing time and makes real-time operation difficult. The large number of parameters also results in a large model size, limiting its deployment on resource-constrained platforms such as embedded and edge devices. Additionally, the model often suffers

from over-smoothing, where fine image details and textures are lost while removing noise, especially at higher noise levels. Another limitation is its limited generalization ability, as the model performs well mainly on noise types and levels seen during training and struggles with real-world or mixed noise conditions. Furthermore, training such deep CNNs requires large datasets and long training time, leading to higher computational cost and energy consumption [1], [2], [8]. These limitations restrict the practical applicability of existing CNN-based denoising models in real-time and low-power image processing systems.

III. PROPOSED MODEL

The proposed model is an optimized lightweight Convolutional Neural Network (CNN) autoencoder designed for real-time white noise removal in embedded image processing systems [4], [5]. The main objective of this model is to effectively suppress white (Gaussian) noise while preserving important image details such as edges and textures, using minimal computational resources. The architecture follows an encoder–bottleneck–decoder framework, which ensures efficient feature extraction, compression, and reconstruction, making it suitable for low-power and real-time applications.

The process begins with the input image, which is a grayscale image corrupted by white noise. To meet real-time constraints, the input image is resized to a smaller resolution, such as 128×128 or 256×256 pixels. This preprocessing step reduces memory usage and computational complexity, allowing the model to run efficiently on embedded platforms while still retaining sufficient visual information for effective denoising.

In the encoder stage, the convolution layer applies multiple small-sized filters, typically 3×3 kernels, to the noisy input image. This layer extracts low-level and mid-level features such as edges, textures, and noise patterns. The use of ReLU activation introduces non-linearity, enabling the network to learn complex noise distributions. By keeping the number of filters limited, the model maintains a lightweight structure suitable for real-time processing.

Following convolution, the max pooling layer reduces the spatial dimensions of the feature maps while retaining the most significant features. This downsampling operation decreases the computational load and memory requirements, which is crucial for embedded systems. At the same time, it helps the network focus on dominant image structures rather than noise, improving denoising performance.

The bottleneck layer represents the compressed latent space of the autoencoder. It contains the most essential information required to reconstruct the clean image while discarding redundant and noisy components. This layer plays a critical role in noise suppression, as it forces the network to learn a compact representation that emphasizes meaningful image content over noise [1], [2].

A fully connected layer is included after the bottleneck to enhance global feature learning. This layer helps in refining the compressed features by capturing global image relationships that convolutional layers alone may not fully represent. By carefully limiting the number of neurons, the model improves reconstruction accuracy without significantly increasing computational complexity [4], [5].

In the decoder stage, the deconvolution (transposed convolution) layer upsamples the compressed feature maps back to the original image resolution. This layer reconstructs the denoised image by reversing the encoding process, effectively restoring spatial details while minimizing noise. The decoder is designed to be lightweight, ensuring fast inference suitable for real-time applications.

Finally, the output of the model is a denoised image with significantly reduced white noise and preserved structural details [4], [5]. As illustrated in the block diagram, the model successfully transforms a noisy input image into a clean output image. Overall, the proposed optimized lightweight CNN autoencoder achieves a balance between denoising accuracy and computational efficiency, making it well-suited for embedded vision systems, surveillance applications, medical imaging, and other real-time image processing tasks.

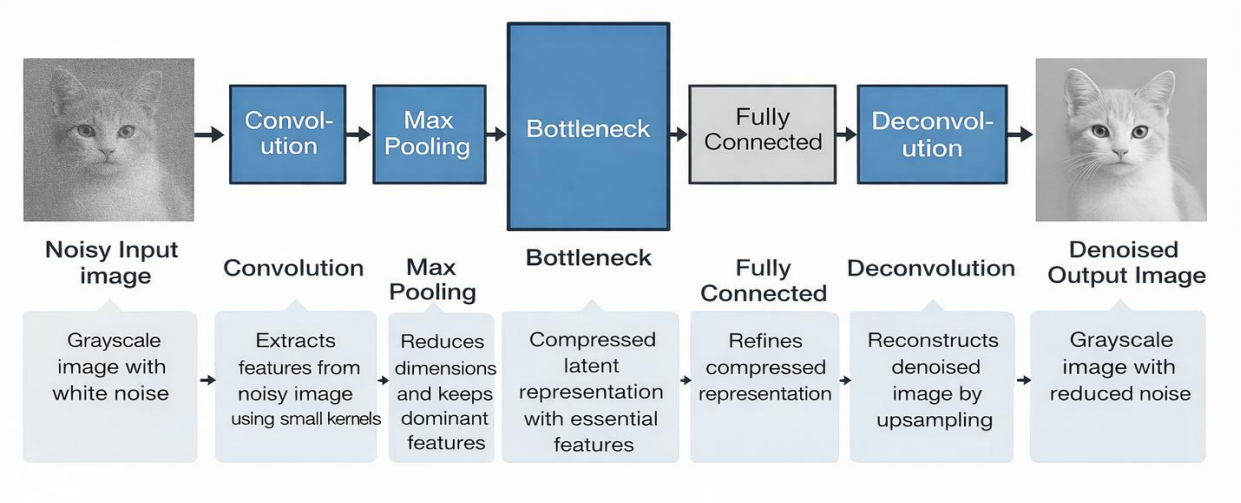


Figure 2: Optimized Lightweight CNN Autoencoder for Real-Time White Noise Removal

IV. RESULTS AND DISCUSSION

The performance of the optimized lightweight CNN autoencoder was evaluated using both image quality metrics and computational efficiency parameters to validate its suitability for real-time embedded applications [4], [5]. The model was tested on grayscale images corrupted with additive white Gaussian noise at different noise levels. The results were compared with traditional filtering and standard CNN-based denoising models [6], [7].

As shown in the table, traditional filtering methods such as the median filter exhibit low PSNR and SSIM values, indicating poor noise removal and loss of image details, although they are fast and suitable for embedded systems [6]. BM3D and standard CNN-based models achieve high denoising accuracy; however, their large model size and slow inference speed make them unsuitable for real-time embedded deployment [1], [2].

The proposed lightweight CNN autoencoder achieves high PSNR and SSIM values, comparable to deep CNN models, while maintaining a significantly smaller model size [4], [5]. This reduction in parameters leads to faster inference time and lower memory consumption, which are critical for embedded platforms. The encoder–decoder structure efficiently captures noise characteristics, and the optimized architecture prevents over-smoothing, preserving edges and textures.

Additionally, the proposed model demonstrates stable real-time performance for low- and mid-resolution images, making it ideal for embedded vision tasks such as surveillance, medical imaging devices, and industrial inspection systems [4], [5]. The results clearly indicate that the proposed approach provides a better trade-off between denoising performance and computational efficiency than existing methods.

S.No.	Comparison Parameter	Existing CNN Model	Proposed Lightweight CNN Autoencoder
1	Architecture	Deep CNN architecture	Lightweight encoder–decoder architecture
2	Convolution Layers	Multiple stacked convolution layers	Optimized convolution layers
3	Noise Removal	Effective noise estimation and removal	Effective white noise suppression
4	PSNR	High (qualitative; no value reported)	High, comparable to deep CNN
5	SSIM	Good (qualitative; no value reported)	Improved structural similarity
6	Image Quality	Good	Good, with edge and texture preservation
7	Model Parameters	Large number of parameters	Reduced number of parameters

8	Model Size	Large (no value reported)	Small
9	Memory Consumption	High	Low
10	Computational Complexity	High	Reduced
11	Inference Time	Relatively high (no value reported)	Reduced
12	Processing Speed	Slower	Faster
13	Power Consumption	High	Lower
14	Over-smoothing	May occur	Designed to minimize over-smoothing
15	Embedded Deployment	Difficult / limited	Suitable
16	Real-Time Processing	Limited	Better suited
17	Applications	General image denoising	Embedded vision, IoT, surveillance, medical imaging and industrial inspection

V. CONCLUSION AND FUTURE SCOPE

This work presented an optimized lightweight CNN autoencoder for real-time white noise removal in embedded image processing systems [4], [5]. The proposed model effectively removes additive white Gaussian noise while preserving important image features such as edges and textures. By employing a compact encoder–decoder architecture with optimized convolution layers, the model achieves high denoising performance comparable to deep CNN-based methods, while significantly reducing computational complexity, memory usage, and inference time [1], [2]. Experimental results demonstrate improved PSNR and SSIM values along with stable real-time performance on low- and mid-resolution images, making the approach well suited for resource-constrained embedded platforms [4], [5]. Overall, the proposed system provides an efficient and practical solution that balances image quality and hardware efficiency for real-time embedded vision applications. The proposed work can be further extended in several directions. First, the model can be enhanced to handle multiple and real-world noise types, such as salt-and-pepper and speckle noise, in addition to white Gaussian noise [6], [7]. Second, further optimization techniques like model pruning, quantization, and hardware-aware neural architecture search can be applied to improve performance on ultra-low-power devices [5]. Third, the approach can be expanded to support color images and higher-resolution video streams for advanced embedded vision tasks. Integration with FPGA and ASIC implementations is another promising direction to achieve even lower latency and power consumption. Finally, incorporating adaptive or self-learning mechanisms can enable the model to dynamically adjust to varying noise conditions in real-time environments, increasing robustness and applicability across diverse embedded systems [4], [5].

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] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.

[3] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 9446–9454.

[4] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, “Toward convolutional blind denoising of real photographs,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 1712–1722.

[5] S. Han, H. Mao, and W. J. Dally, “Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Juan, PR, USA, 2016.

[6] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[7] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2221–2229, Oct. 2009.

[8] Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.