Low-Complexity Convolutional Neural Network for Salt and Pepper Noise Removal in Digital Images

Prudhvi Raj Budumuru, K N V N Gupta, M Sri Hari, J Anand Sai Santosh, Satya Sri Naga Venkat N, K Yogi Suhas Reddy, B Elisha Raju and G Prasanna Kumar

Vishnu Institute of Technology, Bhimavaram, Andhra Pradesh, INDIA

[prajece2005@gmail.com](mailto:prajece2005@gmail.com)

**Abstract:** Digital images have been utilized widely in medical, satellite, and security applications and they may deteriorate with unwanted information acknowledged as image noise by the reasons of inappropriate capturing, transmission, and storage. Salt and Pepper noise is one of the significant issues in digital images, it creates black and white spots on the image and results loss of particular information. Hence, Image denoising is one of the key concepts in image restoration to recover the ground truth image from input noisy image. In this work, proposed a low-complexity CNN model in terms of layers and computational time for salt and pepper noise removal.

**Keywords:** CNN; Denoising; Digital Images; Image Noise; Image Restoration; Salt and Pepper noise

Introduction:

Salt and Pepper noise was one of the significant categories of image noise. It has been replacing the original pixel intensity values with either minimum gray level ‘0’ or maximum gray level ‘255’ (i.e., 8-bit image) in a uniform manner, and unaffected pixels remain the same. Hence, these noisy images appear with black and white spots and are also termed as fixed valued impulse noise defined in (1).

(1)

Image denoising is one of the prominent tasks to perform subsequent image processing steps involved after the pre-processing stage in the applications such as; diagnosis of tumors from clinical images, object identification in satellite images, and face recognition in security camera images.

Spatial domain algorithms are familiarly named as non-linear filtering schemes; conventional Standard Median Filter (SMF), filtering scheme along with intensity threshold switching mechanism [10], filtering scheme with threshold and rank ordered switching mechanisms, adaptive window size methods and statistical parameters-based methods were designed and operated directly on pixel gray values to take off the salt and pepper image noise. Most of these had performed satisfactorily at low and mid-range noise density conditions only and had edge-preserving problems at higher noise density situations.

Convolutional Neural Network (CNN) is the solution to overcome the aforementioned drawbacks of non-linear methods [11-12]. Convolutional neural networks are providing good solutions in the fields of medical and satellite for various applications such as image restoration, segmentation, and classification.

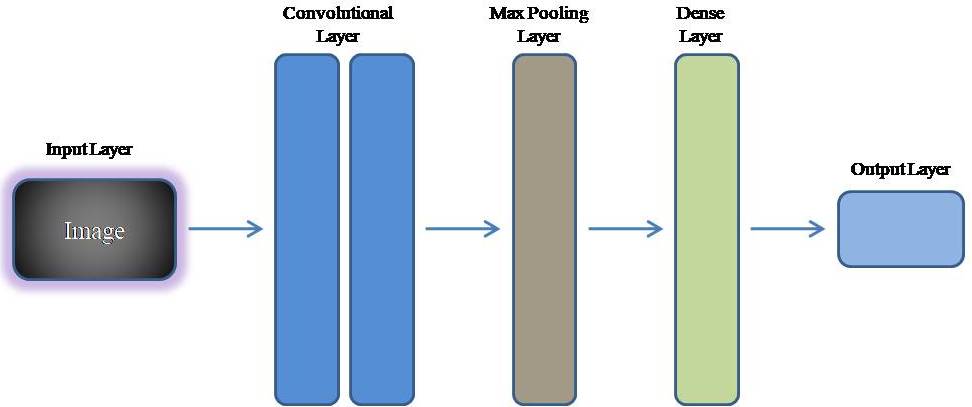


Figure 1: General architectural flow of CNN

Figure 1 shows the generalized architectural flow of CNN. It consists input layer, convolutional layers, max-pooling layers, dense layers or fully connected layers, and finally output layer. The CNN methods or approaches are capable of reconstructing fruitful images from various noisy input images of individual and mixed types. Various researchers have proposed CNN and RCNN models for the removal of Salt and Pepper noise over the past two decades.

Literature Survey:

Frequency Separation Network-based Restoration:

The authors [1] introduced a novel approach called the Frequency Separation Network (FSN) for Single-Image Super-Resolution (SISR). Unlike existing methods that assume simple degradation models, FSN addresses real-world image complexities. It utilizes Gaussian filters in a Frequency Separation (FS) module to gradually separate low and high frequencies. These are then processed through different feature extraction modules and merged using an Adaptive Feature Fusion (AFF) module to reconstruct high-resolution images. FSN focuses on high-frequency information, ensuring accurate restoration of important details in real-world images. Experimental results demonstrate superior visual quality and generalization ability, particularly in scenarios with realistic and natural textures.

Deblurring Learning Framework for Remote Sensing Images:

A blind deblurring learning framework was proposed by the authors [2] as a solution to the problem of deteriorated remote sensing photographs. Whereas deep learning approaches are opaque, traditional approaches depend on preconceived notions or lack transparency. Based on network design theory, the suggested framework updates blurring kernels and pictures via alternating shrinkage threshold rounds. Using multiscale prior feature extraction, an attention mechanism, and a generalized shrinkage threshold, it presents a learnable blur kernel proximal mapping module (KPMM) and a deep proximal mapping module in the picture domain. Optimizing image restoration using deep geometric prior feature learning is the main goal of the Multiscale Generalized Shrinkage Threshold Network (MGSTNet) that is finally produced.

To overcome the shortcomings of previous research that only addressed enlarging the receptive field, this work [3] presents Uniwin, a revolutionary picture super-resolution model. Combining sliding window and shifted window attention is essential for balancing the interactions of global and local. Shifted widow operation is employed to enhance the global interactions by increasing receptive field while getting the exact information of local patterns and the Uniwin merges the information of global and local patterns through this operation. To sum up, Uniwin offers a practical solution by combining local and global attention methods, improving the performance of picture super-resolution.

Transformer-based methods to blind super-resolution:

The research [4] introduces a unique degradation-aware self-attention-based Transformer model to overcome the difficulty of using Transformer-based techniques for blind super-resolution (SR). Contrastive learning is used in this model to handle unknown noise in input photos. The method uses a degradation-aware Transformer to extract global semantic features and combines CNN and Transformer components in the SR network to extract local features modified by degradation information. The suggested model outperforms current techniques and reaches state-of-the-art performance on multiple benchmark datasets. Notably, it surpasses DASR by 0.94 dB and KDSR by 0.26 dB with PSNRs of 32.43 dB on a ×2 scale on the Urban100 dataset and 26.62 dB on a ×4 scale. These results establish a new standard.

Impact of adverse weather conditions:

The impact of bad weather on Advanced Driver Assistance Systems (ADAS) and the difficulties of implementing high-performance Transformers for picture restoration on platforms with limited resources was discussed by the authors [5]. The authors provide a workable approach by introducing a Linear Sparse Transformer (LSFormer) intended for neuromorphic computing systems. LSFormer provides superior picture restoration over conventional Transformers by resolving issues with Softmax attention difficulty. Additionally, the article suggests an implementation strategy that reduces deployment complexity at the edge and is based on neuromorphic computing platforms. Tests using different image restoration tasks show that LSFormer outperforms the state-of-the-art networks. Furthermore, an evaluation against current hardware implementations confirms the viability and excellence of the suggested neuromorphic computing-friendly scheme.

Proposed Method:

Deep learning-based image denoising is a widely employed technique in contemporary image processing [11]. In the realm of classification challenges, CNN-based image processing stands out as the predominant approach, extending its application to solving regression-based problems as well. The current focus in image denoising revolves around regression techniques, where the input image undergoes continuous mapping to generate the output image. The previous method [12] of DnCNN reduced the Gaussian Noise up to a level but was not effective for the full denoising of Noise so our work with RCNN was introduced to impact the level of noise. The proposed method of RCNN is accomplished for the removal of Gaussian noise with the residual networks with are effectively back-propagated weights for the updated network. The input image is considered as Al and output as Zl, the weights or the cost function is updated based on the calculation of Error which reflects the training accuracy of the network and the error function will be calculated using the following in (2)

*R*(*ϕ*)=1/*M*​∑*l*=1*M*​∥*W*(*Al*​,*ϕ*)−*Zl*​∥ (2)

* R(ϕ): The loss function with respect to the parameters ϕ in the CNN.
* M: The total number of samples in the dataset.
* l: An index indicating the individual samples in the dataset, ranging from 1 to M.
* W(Al​,ϕ): represents the output of CNN with parameters
* Zl​: The ground truth or actual output corresponding to the l-th input Al​.
* ∥⋅∥∥⋅∥: The norm, which measures the magnitude of the difference between the predicted output W(Al​,ϕ) and the ground truth Zl

The loss function R(ϕ) is calculated as the average of the norms overall M samples in the dataset, providing a measure of the difference amongst actual datasets and the expected output. Minimizing this loss function is a common objective in training CNNs, aiming to improve model performance in various tasks such as image classification, object detection, or image denoising.

A few noisy images were sent to the network as input during the testing phase, and the network produced restored images as output.



Figure 2: CNN architecture for Image deconvolution

The general architecture of the CNN for the image deconvolution is shown in Figure 2 which consists of three stages of hidden layers between the input and output layers. The first stage consists of the combination of the convolution and the Relu Network layers which are fully connected types. Along with these two layers, we are introducing Batch Normalization in the second stage of the hidden layers for normalization of the extracted neurons that are so-called features of training. Finally, before the output layer, the traditional convolution layer is modified with the transpose convolution layer. But this architecture is slightly modified to well suit our proposal which is as shown in Figure 2. In order to decrease the gradients in the CNN we are adding some layers which will also decrease the Cost function. Gradients should be decreased in RCNN with the help of skip connections which are the key concepts involved. When the gradient function gets close to zero, training the network becomes significantly more challenging. To address this issue, residual networks with skip connections might be employed. The derivatives are calculated in the backpropagation to ensure the effective training and also some skip connections are introduced in the architecture to stop the derivatives from becoming null.



Figure 3: Proposed RCNN architecture for Image deconvolution

The proposed network is shown in the figure 3, it is the combination of U-net and Res-Net. This network is the best fit for our application of denoising the impulse noise. The skip connections avoid the network for overfitting problems. The second stage of the hidden layer is used to decrease the size of the features matrix whereas the third stage in the hidden layer is responsible for improving the feature matrix size suitable for the valuable predictions in the stage of the training phase. The number of skip connection stages is three in the proposed design. This network achieves higher accuracy for homogenous areas gives pixel-wise detailed analysis and is robust.

Simulation Results

Training the RCNN network involves a dataset of 12 images, shown in Figure 4. The training process spans 30 epochs with 𝜎 set to 50, employing Adam's optimizer. The training carried out on a CPU, takes around 21 hours, and the outcomes are shown in Figure 4. Our proposed method was simulated with the help of the Deep Learning Toolbox in MATLAB.

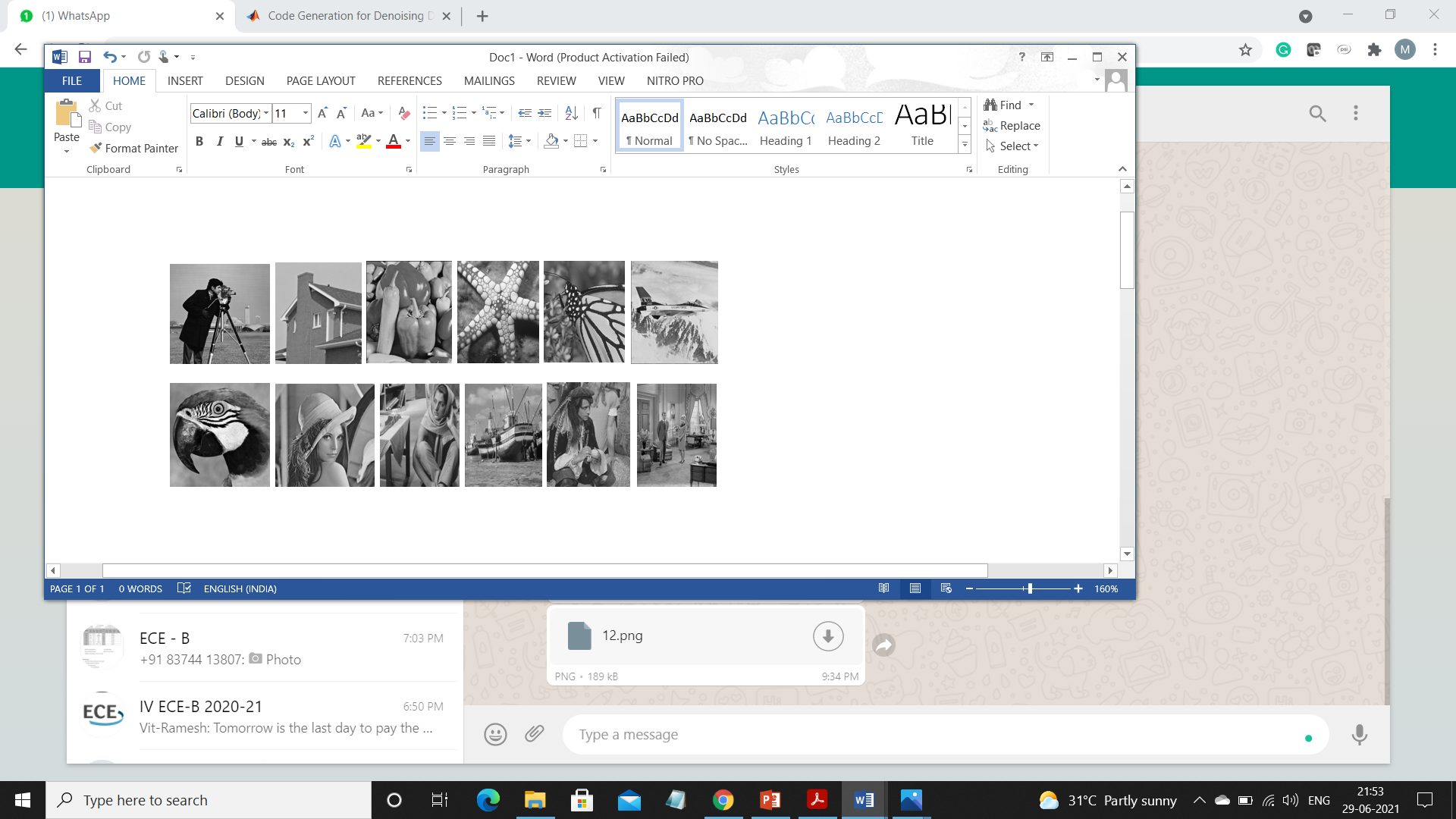


Figure 4: Sample Images of Data

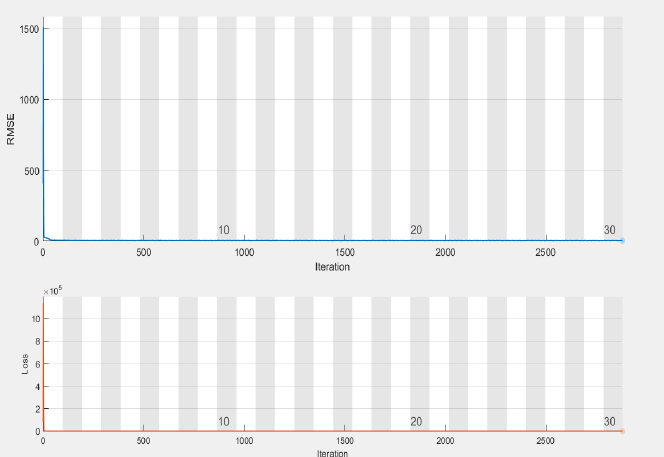


Figure 5: Training loss curves

|  |  |
| --- | --- |
| **(a)** | **(b)** |
|  |  |
| Figure 8: [7] 26.1 dB | Figure 9: [8] 25.9 dB |
|  |  |
| Figure 10: [9] 26.4 dB | Figure 11: Proposed 27.2 dB |
|  |  |

Figure 6. Image Denosing of Proposed with 50%.

The image-denoising outcomes for the parrot image are showcased in Figure 6, alongside a comparative analysis with existing deconvolution methods, emphasizing the PSNR values of the proposed method in contrast to established approaches. Notably, the proposed technique delivers a sharp image with an impressive PSNR value of 27.2 dB, surpassing the performance of prevalent methods. Table 1 provides more information on the denoising performance by contrasting the suggested RCNN with the current dnCNN using Adam's optimizer. The dataset including 12 photos is represented by the first column, which is labeled images 1 through 12. These findings clearly show that the suggested technique performs better for each image of SSIM and in terms of PSNR.

Table 1: COMPARISION OF PROPOSED METHOD WITH EXISTING

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Image | PSNR  DnCNN | PSNR  Residual | SSIM  DnCNN | SSIM  Residual |
| 1 | 20.43 | 23.49 | 0.89 | 0.69 |
| 2 | 21.63 | 24.49 | 0.87 | 0.76 |
| 3 | 21.00 | 24.45 | 0.84 | 0.74 |
| 4 | 20.79 | 22.64 | 0.91 | 0.64 |
| 5 | 21.28 | 24.00 | 0.8 | 0.77 |
| 6 | 21.01 | 22.71 | 0.86 | 0.71 |
| 7 | 21.09 | 27.20 | 0.87 | 0.72 |
| 8 | 21.78 | 26.51 | 0.85 | 0.75 |
| 9 | 20.46 | 22.68 | 0.92 | 0.61 |
| 10 | 21.00 | 24.80 | 0.84 | 0.62 |
| 11 | 21.46 | 25.12 | 0.85 | 0.64 |
| 12 | 21.03 | 24.46 | 0.86 | 0.60 |

Conclusion:

This paper introduces RCNN, a fresh approach to image denoising that leverages highly deep networks. However, training such deep networks is challenging due to slow convergence rates, often caused by issues like gradient vanishing and explosion. We use skip connection operations and residual learning to overcome these obstacles. Using Adam's optimizer, the RCNN network is trained for image deconvolution on a dataset of twelve images. When using Adam's optimizer to improve our suggested method, the PSNR and SSIM values are better improved when compared with the existing algorithms of Image Denoising. To improve our outcomes even more, we plan to investigate the possibility of training on different image datasets in subsequent studies.

References:

1. W. Guan, H. Li, D. Xu, J. Liu, S. Gong and J. Liu, "Frequency Generation for Real-World Image Super-Resolution," in IEEE Transactions on Circuits and Systems for Video Technology.
2. Y. Feng, Y. Yang, X. Fan, Z. Zhang and J. Zhang, "A Multiscale Generalized Shrinkage Threshold Network for Image Blind Deblurring in Remote Sensing," in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-16, 2024.
3. G. Cho and Y. S. Choi, "Image Super-Resolution with Unified-Window Attention," in IEEE Access, doi: 10.1109/ACCESS.2024.3368436.
4. Q. Liu, P. Gao, K. Han, N. Liu and W. Xiang, "Degradation-Aware Self-Attention Based Transformer for Blind Image Super-Resolution," in IEEE Transactions on Multimedia.
5. D. Xie, H. Xiao, X. Hu, Y. Zhou, G. Zhou and S. Duan, "Neuromorphic Vision Restoration Network for Advanced Driver Assistance System," in IEEE Transactions on Consumer Electronics.
6. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Trans. image Process., vol. 16, no. 8, pp. 2080–2095, 2007.
7. S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 2862–2869.
8. D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” in 2011 international conference on computer vision, 2011, pp. 479–486.
9. Levin, B. Nadler, F. Durand, and W. T. Freeman, “Patch complexity, finite pixel correlations and optimal denoising,” in European Conference on Computer Vision, 2012, pp. 73–86.
10. R. Sravani, K. Srinivas and P. Raj.B, "Removal Of Low-DENSITY Impulse Noise From Grayscale Images Using New Modified Hybrid MedianFilTerWith Various Sliding Window Sizes," 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2018, pp. 1750-1754.
11. Sankar, M.R. et al. (2024). Optimization of Single Image Dehazing Based on Stationary Wavelet Transform. In: Pareek, P., Gupta, N., Reis, M.J.C.S. (eds) Cognitive Computing and Cyber Physical Systems. IC4S 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 536. Springer, Cham.
12. Soni Sharmila, K., Manikanta, S.P., Santosh Kumar Patra, P., Satyanarayana, K., Ramesh Chandra, K. (2024). An Efficient Denoising of Medical Images Through Convolutional Neural Network. In: Pareek, P., Gupta, N., Reis, M.J.C.S. (eds) Cognitive Computing and Cyber Physical Systems. IC4S 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 536. Springer, Cham.