INTRODUCTION TO REVOLUTIONIZING IMAGE DENOISING USING CNN WITH DOWN-SAMPLED SUB-IMAGES

1Sushma U K, 2Dileep Kumar M J, 3Sundresh Madalu

1Lecturer, 2Asst. Professor, 3Asst. Professor

1VTU,
2Nitte(Deemed to be University,
3VTU

Abstract: Image restoration is the operation of taking a corrupt/noisy image and estimating the clean, original image. Corruption may come in many forms such as motion blur, noise and camera mis-focus. There have a lot of approaches to remove noise from an affected image and to reconstruct a clean image with high visual quality. Some of the non-CNN related approaches are much efficient to remove noise from affected images. But these methods also show some issues like complex optimization, Computational expenses, etc. while performing denoising. To beat the disadvantages of these denoising approaches, CNN related methodologies are created to enhance the denoising performance. These denoising strategies are in acceptable considerations for their efficient performance in image rebuilding. They utilize some techniques like residual learning to enhance the denoising performance. So this paper provides a review on some CNN related image denoising approaches.

Keywords: Denoising, CNN, Residual Learning
1. Introduction

Today, there is hardly a technical endeavour that digital image processing is without having some sort of influence on [1]. Digital image processing has so many different uses that it would be beneficial to organize it in some way to try and encompass its breadth. Organizing pictures into categories based on their source (such as visual, X-ray, and other) is one of the simplest methods to gain a fundamental grasp of the range of applications for image processing. The electromagnetic energy spectrum serves as the primary energy source for pictures used today. Computers create the fake visuals that are utilized in modeling and visualization. Image processing is any type of signal processing used in computer science and electronics engineering where the input is an image, such as a picture or a frame of video, and the output can either be another image or a set of parameters or characteristics connected to the image. Over the past 50 years, image processing techniques have advanced significantly and are now employed for a variety of reasons:

- Improving the visual appearance of the image for further analysis
- Preparing images for measurement of the features and structures present
- Most image processing techniques involve treating the image as a two-dimensional signal and applying standard signal processing techniques to it.

Some of the important image processing operations are:

- Image restoration: It relates to the elimination or mitigation of visual degradations. By utilizing a priori knowledge of the degradation phenomena, restoration strategies ultimately aim to recreate or restore a damaged image.
- Image enhancement: It may be thought of as the transformation of one picture into another, enhancing the appearance and feel of the image for computer analysis or human visual perception.
- Image segmentation: It is the division of a digital picture into several sections. The split or separation of a picture into sections with comparable attributes or qualities is known as segmentation.
- Image compression: It entails reducing the size of the image without sacrificing the image's quality to an unacceptable degree. More photos may be saved in a given amount of disc or memory space because of the smaller file sizes. Additionally, it speeds up the process of downloading or sending a picture over the internet.

The suggested research looked at cutting-edge restoration and enhancement filters and focused on picture restoration and enhancement processes. The advantages and disadvantages of the current sophisticated filters have been examined as part of this inquiry. Three fuzzy logic-based soft computing strategies have been presented to address the shortcomings of these sophisticated filters. We have strong tools to describe and analyze human knowledge as fuzzy if-then rules thanks to soft computing approaches like fuzzy logic. The
strength of fuzzy logic lies in its ability to effectively handle ambiguity and vagueness, making it suited for managing image processing problems.

1.1. Concept of image and denoising

In the present era of digital image application requirement, there is a need to have an efficient image denoising and restoration method as these images are often taken in poor conditions [4]. The need for image improvement cannot be compromised albeit state of the art modern equipment are deployed for capturing images.

Images are ideally encoded with gray level matrix or values of colour. The dependency of image accuracy is basically on its blur and noise contents, wherein blur is an intrinsic factor based of image acquisition system have finite number of samples.

For example, the image can be defined as “x” and de-noising operator as “Dh” where the filtering parameter considered to be “h”. This can be defined by the noise method and image difference as $x - Dhx$.

1.2. Sources of noise

Noises are inevitable in images during the time of acquisition or transmission; various factors and processes are attributed to its presence in an image. The quantification of noise is decided based on the number of pixel corrupted in an image [5]. The following factors are considered to be the prime sources for the presence of noise in an image.

i. Environmental condition prevailing during image acquisition which affect the imaging sensor.

ii. Noise may be introduced in an image due to insufficient levels of light and sensor temperature.

iii. An image may also be corrupted with noise due to transmission interference.

iv. Presence of dust particles in the screen scanner may also be a cause for introducing noise in an image.

1.3. Types of noises

Types of noise present in an image can be broadly classified as additive or multiplicative form. These noises are originated during image capturing process and are unavoidable. It is generated from grain short noise of an ideal photon of an image [6]. Corrupted noisy signals are produced, while the original signals of an image are added with additive noise signal and termed as additive noise model. Noise signals getting multiplied during the process of image capturing with the original signal is said to be multiplicative noise model. Based on these considerations, various noise types are briefly explained below:
Gaussian Noise

The additive noise model independently present in each pixel of an image is termed as amplifier noise or Gaussian noise. This presence is mainly because, added use of amplification, while capturing blue color channel of an image as compared to green or red channel. Presence of this noise in the dark area of image at constant level is considered as major step towards de-noising of an image [7].

Salt and Pepper noise

Images having dark pixels in bright region and bright pixels in dark regions are considered as having salt and pepper (Impulse) noise. The main causes of this noise are by deceased pixel, analog to digital converters error, and bit error during transmission. Elimination of this noise is possible by applying dark frame subtraction and interpolation of dark and bright pixels [8].

Speckle noise (Multiplicative noise)

Speckle noise is the granular noise, which is inherently present in an image and cast for degrading the quality of active radar and Synthetic Aperture Radar (SAR) images. It creates random fluctuations in conventional radar results while returning signal from an object, which is not larger than in single processing element. This noise SAR generally causes more serious difficulties in an image interpretation [9]. This noise is caused by coherent processing backscattered signal received from multiple distributed targets.

Uniform noise

The uniform noise is also called as quantization noise. It is caused by quantization of the image pixel distributed uniformly across the image to a number of distinct levels. In this noise, the gray values of the noise level are distributed across the image in a specified range [10]. It can also be used for generating different types of noise distribution and to degrade images for evaluating using restoration algorithms. This also gives most neutral unbiased noise.

Poisson Noise (Photon Noise)

Specific types of electronic noises that occur when finite number of electron energy particles carried in a circuit and photon in optical device that is small enough to detect statistical fluctuation in a measurement called Poisson noise [11]. This is also cast wherein the number of photons being sensed by a sensor is not adequate to provide detectable statistical information. It generally has root mean square value proportional to square root intensity of the image.
1.4. Denoising process

The process of denoising to be applied is based on an algorithm that can remove the parameters of noise in a noisy image without affecting the inherent features of the same, using various denoising methods. The major complication in this process is that it makes use of various complicated algorithms and cumbersome models that have their own advantages and disadvantages [12].

Application of wavelet provides an excellent parameter in the field of image denoising based on its characteristics like sparsely and multi-resolution structure [13]. To cater to the need, wavelet transform has gained increased popularity in the last few decades; various algorithms have been developed in the wavelet domain as observed in published papers and articles relating to image denoising.

Deep learning, a subset of machine learning which in turn is a subset of artificial intelligence (AI) has networks capable of learning things from the data that is unstructured or unlabeled. The approach utilized in this project is Convolutional Neural Networks (CNN). It uses the Haar-cascade classifiers which help us in the detection of objects.

1.5 Convolutional Neural Networks (CNN).

The convolutional neural network, or CNN for brief, could also be a specialized kind of neural network model designed for working with two-dimensional image data, although they're going to be used with one-dimensional and three-dimensional data. Central the convolutional neural network is the convolutional layer that gives the network its name. This layer performs an operation known as “convolution”.

In the context of a convolutional neural network, a convolution may be a linear operation that involves the multiplication of a group of weights with the input, very similar to a standard neural network. As long as the technique was designed for two-dimensional input, the multiplication is performed between an array of input file and a two-dimensional array of weights, called a filter or a kernel.

The filter is smaller than the input file and therefore the before the sort of multiplication applied between a filter-sized patch of the input and the filter may be a scalar product. A scalar product is that the element-wise multiplication between the filter-sized patch of the input and filter, which is then summed, always leading to one value. Because it leads to 1 value, the operation is conventionally represented and mentioned because the “scalar product”.

Using a filter smaller than the input is intentional because it allows an equivalent filter (set of weights) to be multiplied by the input array multiple times at distinct points on the input. Specifically, the filter is applied systematically to every overlapping part or filter-sized patch of the input file, left to right, top to bottom.
This systematic application of an equivalent filter across a picture may be a powerful idea. If the filter is meant to detect a selected sort of feature within the input, then the appliance of that filter systematically across the whole input image allows the filter a chance to get that feature anywhere within the image.

This capability is usually represented and mentioned as translation invariance, e.g. the total altogether concern in whether the feature is present instead of where it should had been present.

2. LITERATURE SURVEY

In this study, Handa, P. and Krishan, B.. et al.. [15] offer a combination strategy based on the contrast limited adaptive histogram equalization and Retinex algorithm. The Retinex algorithm is wavelet-based and includes a Gaussian filter and adaptive histogram equalization. Before the Retinex technique, which employed low frequency components to improve the picture, the image is first enhanced using CLAHE, then decomposed using Daubechies wavelet. In order to smooth the picture, a Gaussian filter is employed last. For the investigation of quality enhancement and denoising, the dataset of maize leaf disease is employed. The findings demonstrate that the suggested strategy enhances quality by lowering the noise of the maize leaf photos. High accuracy disease detection and classification systems for maize leaves may be created using these improved pictures.

In this study, Hu, B. et al..[16] suggested a quality-improving method based on texture search, histogram redistribution, denoising, and contrast enhancement. First, a texture-based search technique is suggested to increase the sparsity of 4D block matching clustering, which would enhance the accuracy of denoising. Then, spatial contrast is improved while spectral data is kept intact by using histogram redistribution and Poisson fusion. The suggested approach is quantitatively assessed using synthesized noise data from publicly available hyper spectral datasets, and the experimental results are analyzed using a variety of criteria. Simultaneously, classification tasks were employed to check the increased data's accuracy. The
outcomes demonstrate that the suggested technique is effective in enhancing the quality of hyper spectral data.

By employing two illumination beams to activate the material, Van Nhu, L. introduces a greater spatial resolution for confocal scanning fluorescence microscopy in this research. The comparison of the imaging results between the proposed method and the confocal scanning fluorescence microscopy with the Gaussian illumination beam has shown the proposed method's capacity to increase resolution.

A thorough review of deep learning-based techniques for post-processing MR pictures to improve image quality and remove artifacts is given by Chen, Z., et al. in this publication [18]. We want to offer a literature review on deep learning methods for MR image enhancement to researchers working in MRI or other research areas, such as computer vision and image processing. We explore the existing drawbacks of AI's use in MRI and highlight potential avenues for advancements in the future. We emphasize the significance of a rigorous evaluation of the explanations offered in the deep learning age as well as the generalizability of deep learning algorithms in medical imaging.

3. Problem Statement

A difficult job in computer vision and image processing is image restoration, which aims to restore the real or original picture from a damaged or degraded version. The objective is to fix different distortions and flaws that were caused during picture capture, transmission, or storage and return the image to its original, high-quality state.

Image restoration's primary goal is to improve the image's visual quality, sharpness, and details while minimizing artifacts, noise, blur, and other degradations. Various methods and approaches are applied throughout the restoration process to restore lost or damaged information and enhance the perceived quality of the image. The CNN architecture may be modified for noise reduction applications to produce patterns that bypass the vanishing gradient bottleneck. When creating CNN techniques, technical ideas and expertise are combined with an understanding of the many types of noise and noise models. However, noise features are of a continuous nature and require a model constructed from scratch because pre-trained CNN models were employed in the majority of research. The creation of such a model leaves potential for modification and improvement. But it takes a lot of time and computing resources to create a model from scratch. Deep learning techniques may replace traditional ones as a result of the usage of spatial patterns in CNN architecture. Contrary to popular belief, CNN is not a "black box," and features visualization techniques offer a reliable foundation for noise reduction; yet, the computational time and space still pose the biggest hurdle.
4. Proposed System

In this study, we utilize a noise level map $M$ that can be adjusted to provide input to the denoising model, allowing it to adapt to varying noise conditions. In order to make the denoiser work better, a reversible down sampling operator is introduced to divide the input picture into four smaller sub-images. For a grayscale image, $C = 1$, but for a color image, $C = 3$. To ensure that the trade-off between noise reduction and detail preservation is consistently handled by the noise level map without introducing any visual distortions, we employ the orthogonal initialization strategy for convolution filters.

**Network Architecture**

Figure 4.1 shows the DSS-NL-Mapping-Multi-noise Net framework. The first stage involves reformatting the noisy input picture $y$ into four lower-resolution copies using a reversible down sampling operator. We then use a noise map $M$, whose magnitude may be changed, to combine the down sampled portions of the image, to produce a tensor $\tilde{y}$ of size $W/4*H/4*C$.

The next CNN employs a series of three $3x3$ convolution layers and the input tensor $y$. Each layer is composed of three different types of processes: convolution (Conv), rectified linear units (ReLU), and batch normalization (BN). To be more precise, "Conv+ReLU" for the first convolution layer, "Conv+BN+ReLU" for the middle layers, and "Conv" for the final convolution layer are all utilized. The sizes of the feature map are preserved between convolutions by using zero-padding. By applying an up scaling operation as the inverse of the down sampling procedure used in the input stage, the estimated clean image $x$ with dimensions $WHC$ is produced. Because it operates on down sampled sub-images, the DSS-NL-Mapping-Multi-noiseNet does not require dilated convolution to further enlarge the receptive field.

Considering the trade-off between complexity and performance, we empirically found that 15 convolution layers were best for grayscale pictures while 12 layers were best for colour images. We used a value of 64 for grayscale and 96 for colour for feature maps. We alter our settings differently for colour and...
black-and-white images for a variety of reasons. First off, because to the significant dependencies between the R, G, and B channels, the model is more likely to benefit from inter-channel dependency when fewer convolution layers are employed. The second is that while colour pictures have more channels as input, feature maps need more channels. Our findings demonstrate that adding more feature maps enhances the colour image denoising performance. Colour images can have an average PSNR increase of 0.15dB using DSS-NL-Mapping-Multinoise Net across a variety of settings and noise circumstances. 12-layer DSS-NL-Mapping-Multinoise Net for colour pictures performs somewhat worse than 15-layer FFDNet for grayscale images, as will be addressed in Section IV-F. We discovered that 12 convolution layers and 96 feature maps provide the optimum performance and efficiency for colour image denoising.

**Noise Level Map**

We first take a look back at model-based image denoising approaches to understand what makes them capable of managing noises of varied intensities in order to boost the flexibility of the CNN-based denoiser. The bulk of model-based denoising algorithms concentrate on this problem:

\[
\hat{x} = \arg \min_{x} \frac{1}{2\sigma^2} \|y - x\|^2 + \lambda \Phi(x)
\]  

(1)

Where \(\frac{1}{2\sigma^2}\|y - x\|^2\) is the data fidelity term with noise level \(\sigma\), (\(x\)) is the associated regularization term for the image prior, and adjusts the relative importance of this term and the one that controls data fidelity. It's crucial to keep in mind that determines the practical trade-off between noise suppression and detail retention. If the amount of noise is too low, a lot of noise will remain; if it is too high, the details will be smoothed out along with the noise. Equation (1)'s solution may be used to construct an implicit function with certain optimization techniques, as demonstrated by

\[
\hat{x} = f(y, \sigma, \lambda; \Theta)
\]  

(2)

Since \(\lambda\) can be engrossed into \(\sigma\), Eqn. (2) can be redrafted as

\[
\hat{x} = f(y, \sigma; \Theta)
\]  

(3)

One can alter the noise level or to balance the retention of detail with noise reduction. (3) Model-based techniques may be used to pictures with varied degrees of noise simply by adjusting the value of in Eqn.

It appears suitable to use CNN to learn an explicit charting of Eqn. (3), which accepts the noise image and noise level as input. However, because to their varied dimensionalities, it is not simple to feed \(y\) and into CNN as inputs. By expanding the noise level into a noise level map \(M\), which is based on patch-based
denoising approaches that really set for each patch, we address the dimensionality mismatching issue. No component of \( M \) is anything other than. Given this, we can rewrite Equation (3) as

\[
\hat{x} = f(y, M, \Theta)
\]

The fact that \( M \) can be generalized to multi-channel degradation maps for usage with more expansive \( N(0,) \) with a zero mean and covariance matrix in the RGB colour space, should not be overlooked. A single CNN model is expected to automatically handle a broad range of noise models with diverse features since \( M \) may not be uniform.

It has been reported that batch normalization and residual learning for a basic CNN may be combined to streamline training and generally improve performance, hence facilitating the eradication of AWGN. Because the residual output (noise) has a Gaussian distribution, batch normalization's Gaussian normalization stage is rather simple. When only one noise level is being considered, the denoising network benefits the most from this task-specific advantage.

Instead, we use DSS-NL-Mapping-Multi-noise Net to explore a wide range of noise using a noise level map as input. It is therefore an exciting idea to reevaluate the residual learning and batch normalization combo for conventional CNN. According to experimental findings, batch normalization may always hasten training regardless of whether a denoising network uses residual or non-residual learning techniques. For instance, with batch normalization, residual learning converges more quickly than non-residual learning, although their final outcomes are nearly equal after fine-tuning. It is really feasible to train a basic network without using the residual learning approach when the network depth is moderate (less than 20 nodes) by using advanced CNN training and design techniques like ReLU, batch normalization, and Adam [28-29]. To make things simple, we avoid residual learning in network design.

We utilize the ADAM technique to minimise the following loss function in order to maximise DSS-NL-Mapping-Multinoise Net.

\[
\ell(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \| F(Y_i, M_i; \Theta) - x_i \|^2
\]

The learning rate falls from 103 to 104 as the training error reaches a plateau. When the training error does not change after five iterations, we mix the parameters of each batch normalisation with the nearby convolution filters. A further 50 iterations of fine-tuning the DSS-NL-Mapping-Multi-noise Net model are performed after that at a slower learning rate of 106. The remaining hyper-parameters for ADAM are left at their factory default values. A mini-batch size of 128 is employed during training, and data augmentation via rotation and flipping is also incorporated.
Proposed Contrast Adjustment

A median filter and an image intensity adjustment are used by the following Contrast Adjustment model to further enhance the quality of the original image from the Multi-noise Net output. The first step in adjusting image intensity is to change the intensity values in image I to other values in image J. By increasing the lowest and highest one percent of pixel values to their maximum, this modification increases the contrast of the resulting image. (ii) A median operation is performed on the two-dimensional image I in a median filter. Both of these procedures—adjusting the image's intensity and using a median filter—improve the final image's quality.

5. Dataset Description

In entirely image denoising means work on noisy images under three dissimilar noise variances \( \sigma \in [30, 50, 75] \). For a thorough analysis of the test photos, we employ the following datasets: BSD68, CBSD68, Set14, and Set24. 174 photos from the independent test batch of 74 photographs make up the whole dataset. Below Figure displays a few photos from the sample dataset. Those are broadly used testing images. All the dataset images are evaluated in two different sizes 256×256 and 512×512.

![Figure 5.1: Some of the widely used testing images.](image)

Metrics of denoising performance

PSNR and SSIM are used as performance metrics to assess image denoising techniques. A ground truth picture \( x \) can be used as a starting point to determine an image's PSNR.

\[
PSNR(x, \hat{x}) = 10 \log_{10} \left( \frac{255}{\| x - \hat{x} \|_2} \right)^2
\]

In addition, the SSIM index is intended by
The two variables' respective means, variances, and covariances are given here \((x)\), together with two constants \((C_1 \text{ and } C_2)\) to prevent instability. Quantitative metrics cannot adequately capture the visual quality of a set of images, but they must be compared. The preservation of edges and textures is essential for assessing denoising algorithms.

\[
SSIM(x, \hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{\mu_x^2 + \mu_{\hat{x}}^2 + c_1(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}
\]  

(7)
Figure- Comparative analysis of all denoising results by plotting the histogram with respect to reference image.

The proposed model achieved nearly 31% to 35% of PSNR and achieved 0.89 to 0.95 of SSIM on all four different data set, when the Salt and pepper noise level is 20% to 40%.

When the noise level is 10%, the PSNR values are 40.19%, 40.05%, 41.17% and 39.75% on BSD68, CBSD68, Set14 and Set24.

While comparing with Gaussian noise, the proposed model achieved better performance on Salt and pepper noise, even at the highest level of noise percentage.

For instance, the SSIM is 0.771, 0.808, 0.738 and 0.812, where the PSNR is 27.18%, 27.92%, 26.58% and 27.92% on all different datasets for the noise level at 80%.

While comparing with all datasets, the proposed model achieved performance on PSNR and SSIM on Set14 dataset for the noise level varies from 60% to 80%.
6. CONCLUSION

In this paper, we present the creation of a novel CNN model called DSS-NL-Net, whose function is to conduct discriminative denoising quickly, effectively, and flexibly. Several methods, including the use of noise level maps as input and denoising in the down sampled sub-image space as the denoising method, were employed in the construction and training of the network to accomplish this aim. The results using synthetic images with AWGN showed that the proposed model can not only produce cutting-edge outcomes when input noise level matches ground truth noise level, but also has the ability to robustly control the trade-off between noise reduction and detail preservation when using synthetic images. The suggested model's effectiveness in managing inhomogeneous noise has been validated on the basis of simulations on pictures with spatially variable ACGN, as shown by the findings. The final comparison showed the suggested network's improved speed performance when comparing the running times of other approaches, including BM3D. The suggested network offers a workable solution to CNN denoising applications that overcomes the aforementioned concerns due to its adaptability, efficacy, and efficiency. This next contribution will involve the addition of a layer of intelligence at the acquisition level that will be able to automatically assess an image's quality and texture based on many elements including noise level, blur, illumination, and shooting circumstances. In order for the suggested method to always perform at its best, this is to be utilised to automatically modify the settings.

7. Applications

- Applications in Medical Field
- Application in Gaming and Entertainment
- Forensic analysis application
- Facial Recognition and Identification
- Accident Investigation
- Traffic Management Centers
- Weather Conditions Monitoring
- Parking Space Detection
- Autonomous Vehicles
8. References


