Journal of Medical Systems (2019) 43: 234
https://doi.org/10.1007/s10916-019-1371-9

IMAGE & SIGNAL PROCESSING

An Intelligent Recurrent Neural Network with Long Short-Term Memory (LSTM) BASED Batch Normalization for Medical Image Denoising

R. Rajeev[1] · J. Abdul Samath[2] · N. K. Karthikeyan[3]

Received: 18 March 2019 / Accepted: 5 June 2019 / Published online: 15 June 2019

Abstract
The process of denoising of medical images that are corrupted by noise is considered as a long established setback in the signal or image processing domain. An effective system for denoising in order to remove white, salt and also pepper noises by means of merging the Long Short-Term Memory, otherwise known as LSTM, based Batch Normalization and Recurrent Neural Network or RNN techniques have been proposed in this research paper. The images of the lung CT are considered as an input in this particular work. Following this, an effectual batch size is calculated by employing the method of Particle Swarm Optimization (PSO). To denoise the image, Recurrent Neural Network or RNN algorithm were proposed, here to reduce the internal covariate shift present in the neural networks, the Long Short-Term Memory or LSTM based Batch Normalization is brought-in. With respect to SNR or Peak Signal to Noise Ratio and Mean Square Error (MSE), operations were assessed. This algorithm is considered as competitive to other denoising schemes which have been confirmed by the experimental outcomes.

Keywords Image denoising · Recurrent neural network · Long short-term memory · Batch normalization

Introduction

The development observed in the field of imaging, computing as well as communication technologies denotein recent years the occurrence of a rapid growth in the domain of image and video applications. Simultaneously, a rising demand has been observed for high image qualities. Nevertheless, during the process of acquisition and transmission, the signal that is captured by a camera is prone to noise. Hence, the process of denoising persists as an important issue in numerous tasks concerning image processing and this has drawna great dealof research interest in the precedent decades [1–12].

In the past decades, a wide range of methodologies for image denoising have been exploited, which consist of filtering-based methods [13], diffusion based methods [14], total variation-sourced techniques [15, 16], wavelet/curvelet-sourced techniques [17–19], sparse representation-sourced techniques [20–23], non-local self-similarity or NSS- sourced techniques [23–30] and so on. NSS or non-local self-similarity models are well-liked, in the state-of-the art methods, this, in turn, implements the fine denoising performance, mainly for images which possess textures that are regular and repetitive. This is because, the methods, which are non-local means-based were assumed to be superior over images with repetitive and regular textures, dissimilar to the discriminative training sourced approaches which usually give rise to better results on images which have either smooth regions and irregular textures [31]. Moreover, NSS or non-local self-similarity models unavoidably experience two major shortcomings [23, 29, 30]. Primarily, the models regularly

This article is part of the Topical Collection on *Image & Signal Processing*

✉ R. Rajeev
rajeevathira@gmail.com

J. Abdul Samath
abdul_samath@yahoo.com

N. K. Karthikeyan
karthiaish1966@gmail.com

[1] Department of Information Technology, Sri Ramakrishna College of Arts and Science (Formerly SNR Sons College), Coimbatore, Tamil Nadu, India

[2] Department of Computer Science, Chikkanna Government Arts College, Tiruppur 641 602, India

[3] Department of Information Technology, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

states a specific feature for the process of denoising, which ultimately results in the algorithm it doesn't implement the good denoising jobs by assuming entire images. Secondly, the models generally are non-convex and engage enormous parameters which are manually chosen, so it offers a little scope to improve the denoising performance.

Enormous discriminative learning approaches were brought-in recently for learning image prior models like Trainable Non-Linear Reaction Diffusion (TNRD) model [32]. TNRD replica is enforced by unfolding number of steps for gradient descent inference process. Moreover, TNRD is naturally limited to stated priors forms. Specifically, priors that are executed in TNRD on basis of model analysis that is restricted in seizing entire image structures' features.

TNRD model is assumed as a part of deep neural network further more this states as feed-forward deep network. Furthermore, we have few deep neural networks for the process of denoising of image [33–35]. Jain and Seung [33] enforced the initial deep neural networks for the process of image denoising. When distinguished to Markov random field model, the deep neural networks on image denoising possess alike or even better representation power. For image denoising [34], Burger et al. anticipated multi-layer perceptron. Moreover, it links deep neural networks and sparse coding which are pre-trained with denoising auto-encoder [35]. Though such deep neural networks achieves superior performance concerning image denoising, which fail at the process of successfully exploring the native features of image, as MLP networks are cascaded by them.

CNN or Convolution Neural Network has fascinated enormous researchers, since the existence of finest capabilities of self-learning through enormous data and it doesn't demand the strict selection of the characteristics, but only the guide to learning for achieving the preferred purpose, which is greatly used in the image preprocessing fields, such as image super-resolution [36–38]. Few researchers tries to utilize Deep Convolution Neural Network (CNN) for image denoising [39, 40]. According to a non-local image model was suggested by Lefkimmiatis [41], a new deep network architecture for image denoising for grayscales well as colour images. According to the author, this method is definitely a Non-local self-similarity or NSS method, which enforces the Deep CNN for learning the various parameters regarding the NSS method. In order to implement the task of denoising, where in an image which is contaminated is entered into the Deep CNN along with the concern latent clean image as the output, the Deep CNN is utilized. Furthermore, the stochastic gradient descent method makes use of the strategy for training purposes, which maximizes time consumption pertaining to training. Though, the vital issue with regard to convolutional architecture was considered as the proper depth computation, since performing image denoising along with additive white Gaussian noise is enforced only with CNN models. It doesn't assume the real complex noises, because, we are in need of improvement in the performance of this classifier.

For denoising the images to resolve this crisis, in this particular work [58], LSTM based Batch Normalization in addition to Recurrent Neural Network (RNN) were suggested. Initially, processing the input image is done. Next to this, we understand that residual learning as well as batch normalization can be of a great advantage for the RNN learning because they not only maximize the speed of training however enhances denoising performance. Consequently, latent clean image by the process of separating is achieved, since the noise image that is predicted from the image is contaminated. Moreover, the authority of network's depth on denoising recital is discussed. Furthermore, a specific number of network parameters are discussed. Lastly, series of comparison were performed to authenticate de-noising technique for proposed image.

The following research paper is trailed as mentioned below: In Section "Related work", current denoising schemes were indicated. Section "Proposed methodology" states the RNN based denoising implementation in detail. In Section "Results and discussion", we reveal experimental outcomes and discussions. Conclusions are explained in Section "Conclusion".

Related work

Here, Image denoising based existing proposals along with advantages and disadvantages were discussed. A novel approach for image denoising on deep convolution neural network is projected by Zhang et al., [42]. DCNN is designed to accomplish noise image, dissimilar to other learning-based methods. Hence, we can accomplish the latent clear image by partitioning noise image from contaminated image. Gradient clipping eliminates gradient explosions plus facilitates quick convergence of the network, at the time of training stage. A superior recital with anticipated denoising technique is accomplished by experiments, which is distinguished to the state-of-the-art denoising approaches. Result specifies that denoising method possesses suppressing ability different noises with varied noise levels through one single denoising replica.

An adaptive learning process is to learn patch-based image priors for image denoising process was suggested by Luo et al., [27]. The novel algorithm, known as the expectation-maximization or EM adaptation, takes basic prior learned from anexternal database that is generic and alters it to the image which is noisy to produce a particular prior. Dissimilar to the current methodologies which joined external and internal statistics through means of ad hoc, anticipated algorithm is developed from perspective which is Bayesian hyper-prior. There exist two contributions. Firstly, complete

derivation of EM adaptation algorithm is offered to enhance computational complexity is demonstrated. Secondly, when the latent clean image is not present, we demonstrate how the expectation-maximization or EM adaptation could be altered on the basis of pre-filtering. The experimental outcomes illustrate that adaptation algorithm suggested produces improve denoising outcomes constantly when compared to the one which doesn't have an adaptation and is considered to be superior to many state-of-the-art algorithms.

Xiong et al., [43] projected a novel image denoising algorithm which is based on adaptive signal modeling as well as regularization. This enhances the image's quality by the process of regularizing every image patch by employing bandwise distribution modeling in the transform domain. Rather than employing a global model by considering all the patches present in an image, it employs adaptive models which are content-dependent in order to address the non-stationarity of image signals as well as the diversity prevalent amongst various transform bands. The distribution model is adaptively calculated for every patch separately. This differs with patch locations and also changes for varied bands. Specifically, we take into consideration the anticipated distribution in order to possess a non-zero expectation. Patches are excluded if they are considered to be inappropriate so that such a model that is adaptively learned is considered as more accurate than the one which is global. Restoration of the image is ultimately done by means of bandwise adaptive soft-thresholding, which is based on a Laplacian approximation of the similar-patch group transform coefficients distribution. Experimental outcomes reveal that scheme that was anticipated surpasses enormous state-of-the-art methods of denoising considering both qualities that are objective and perceptual.

A Convolutional Neural Network (CNN) for the purpose of image denoising which accomplishes performance of diverse approaches was anticipated by Liu et al., [44]. Secondly, we suggest a deep neural network solution which surges two modules pertaining to image denoising and many other high-level tasks, respectively, while it employs the joint loss only for the purpose of updating the denoising network by means of back-propagation. We show that on one hand, the denoiser that was proposed possesses the generality for it to overcome the performance degradation witnessed in various tasks that are of high-level vision. Alternatively, through the guidance of the high-level vision information, more visually appealing outcomes can be generated by the denoising network. This is the first work which is investigating the advantage of image semantics exploitation, concurrently for image denoising as well as vision tasks that are of high-level by means of deep learning, to our knowledge.

Godard et al., [45] employed strategy of burst-capture and executed the intelligent integration through recurrent CNN. construct new, multiframe architecture so that it appears as addition to every single frame denoising replica, and design to manage a random quantity of noisy input frames. It accomplishes state of the art denoising outcomes on burst dataset enhancing multi-frame techniques which are best published, like the VBM4Dand FlexISP. Lastly, various other applications pertaining to image enhancement were discovered by the process of incorporating content obtained from multiple frames furthermore display that DNN generalizes image super-resolution effectually.

Remez et al., [46] confirmed how the quality of reconstruction enhances when a denoiser is attentive of the content typepresent in the image. At one end, we initially propose a novelcompletely convolutional deep neural network architecture that is observed to be simple but powerful as it accomplishes the state-of-the-art performance without even being present as class-aware. We additionally display that a considerable increase in performance of up to 0.4 dB PSNR can be attained by building our network class-aware, specifically, by the process of fine-tuning it for images that belong to a particular semantic class. Depending on the highly flourishing existing image classifiers, this particular research supports employing a class-aware approach in all the tasks pertaining to image enhancement.

Zhang et al., [47] took a move ahead by the process of investigating feed-forward denoising convolutional neural networks to accept development observed in extremely learning algorithm, deep architecture, and regularization technique into image denoising. Particularly, batch normalization and residual learning are employed to quicken training process and boost denoising recital. Unlike current discriminative denoising replicas often guide specific model for additive white Gaussian noise purpose at certain noise level, DnCNNs possesses the ability to manage Gaussian denoising with noise level that is unknown (which is, blind Gaussian denoising). DnCNN totally eliminates the latent clean image in hidden layers with the residual learning strategy. This characteristic feature inspires us to guide one DnCNN model to deal with several tasks pertaining to general image denoising like single image super-resolution, Gaussian denoising and JPEG image deblocking. DnCNN model does not just exhibit high efficiency in several tasks pertaining to image denoising, however effectively executes by gaining profit from GPU computing which has been demonstrated through our extensive experiments.

A novel image denoising technique enforcing multiple-minimum cuts, were proposed by Sato et al., [48], which are sourced on maximum-flow neural network algorithm on minimum cut which is according to nonlinear resistive circuit analysis. MF-NN consists of two characteristics which are not distributed by conventional minimum cut algorithm: One characteristic is multiple-minimum cuts were concurrent, while other characteristic is to be suitable for execution of hardware. With the help of features of MF-NN, we can recognize

new solutions for the two issues pertaining to conventional graph-cuts.

Al-Sbou [49] displayed a comprehensive performance evaluation of employing the neural networks as a tool for reduction of noise. The suggested approach comprises of utilizing both mean as well as median statistical functions for the purpose of estimating the training pattern's output pixels concerning the neural network. This employs a fraction of the image pixel that is degraded in order to produce the training pattern for the system. A range of test images, noise levels and neighborhoods sizes are utilized. Depending uponutilizing samples of pixel neighborhoods which are degraded as inputs, the output of the recommended approach offered a good image denoising performance that displayed promising results which are qualitative and quantitative pertaining to the degraded noisy images with respect to PSNR, MSE as well as visual tests.

Proposed methodology

The proposed LSTM based batch normalization in addition to RNN based image denoising procedure has been elucidated in the following section.

System overview

Primarily, the CT images of the lung are captured from the SIMBA database and processed, and following this the RNN is suggested to perform image denoising. Considering the process of RNN, the LSTM with batch normalization is put forth in order to bring about improvements in the PSNR of the proposed scheme, whereas in batch normalization, the size of the batch is optimally chosen by utilizing PSO. Through experimental results it is observed that the RNN proposed achieved better performances in comparison to other denoising schemes.

Considering the experiments that were conducted, to devise the neural network's training set, we simulate the artificially degraded images by the application of the degradation model. The image considered is convoluted first with a selected noise operation like Gaussian white noise as well as salt and pepper noise after which noise is added to it at unusual rate occurrence. An image which is corrupted is given as follows:

$$y(i,j) = x(i,j) + n(i,j) \tag{1}$$

Where y(i,j) is observed value, x(i,j) is true (original) value and n(i,j) is noise perturbation at pixel (i,j). The MLP's learning phase makes an effortin order tocapture innate space relations concerning degraded pixels and relate them to the pixels that are non-degraded. The image data that is degraded is supplied as input to the MLP and the image which is non-degraded as the respective output in the learning process which is supervised.

Batch normalization

In optimization [50], Whitening is considered as a regular procedure that has shown to lessen the convergence rates. Through widening the deep neural networks, we can assume a random layer as accepting samples from distribution formed by available layer. In distribution, there occurs a modification, at the time of training, creating any layer but initially, we are in-charge not only for learning finest representation as well for getting accustomed to modifying input distribution, nothing but Internal Covariate Shift, decreasing theorized to assist in training process [51]. To minimize this shift regarding internal covariate, will be able to whiten every network's layer. However, this usually appears to be extremely computational demand. Batch normalization [51] computes process of whitening by intermediate indications being standardized makes use of statistics of current mini-batch. Provided, mini-batch x, sample variance and sample mean of every feature k with mini-batch axis can be expressed as:

$$\overline{X}_k = \frac{1}{m}\sum_{i=1}^{m} x_{i,k} \tag{2}$$

$$\sigma_k^2 = \frac{1}{m}\left(x_{i,k} - \overline{x}_k\right)^2 \tag{3}$$

Where m denotes the mini-batch's size. The mini-batch is finely selected by utilizing PSO to enhance PSNR performance in RNN.

Particle swarm optimization

PSO algorithm is infuriated by agroup of migrating birds' social behaviour of attempting to arrive atanvague destination. In the flock, every solution is named as a 'bird' and is identified as 'particle'. A particle corresponds to chromosome in Genetic Algorithms (GA). Unlike GAs, developmental procedure in PSO doesn't proceed by creating original birds from parent. As an alternative, birds present in population develop social behavior and this leads to movement in the direction of a destination. The current work refers to the objective function of this PSO which is to optimally select the batch size in the RNN pertaining to training.

Communication amidst a group of birds occurs when they fly together. From every particular direction, each bird appears, and this leads to collective communication and they recognize bird that is present in best location. As a result, every bird flies in the best bird's direction in the course of velocity that is sourced on current location. All birds are

examined for search space from its novel local location, and same procedure is reiterated till arrival of flock at destination which is favoured. Social interaction and intelligence can be observed in the procedure, with the intention that the birds find out through local search, i.e., from its own experience and from others which is termed as global search.

This procedure is kicked off with N, collection of random particles. i^{th} particle is specified by its position as in S-dimensional space, in which S indicates number of variables. During the process, every particle i perceives three values which are as followed, current position (X_i), best position in previous cycles (P_i), flying velocity (V_i). These three values are indicated as:

$$\begin{aligned} &\text{Current position } X_i = (x_{i1}, x_{i2}, \ldots, x_{iS}) \\ &\text{Best previous position } P_i = \left(p_{i1}, p_{i2}, \ldots, p_{iS}\right. \\ &\text{Flying velocity } V_i = (v_{i1}, v_{i2}, \ldots v_{is}) \end{aligned} \tag{4}$$

Considering every time interval (cycle), position (P_g) of best particle (g) is computed as best fitness amongst all particles.

Hence, every particle revises its velocity $V_i$to be closer to best particle g, as trails.

$$\begin{aligned} New\ V_i = \omega \times current\ V_i + c_1 \times rand() \times (P_i - X_i) + c_2 \\ \times Rand() \times (P_i - X_i) \end{aligned} \tag{5}$$

With new velocity V_i, updated particle position becomes:

$$\begin{aligned} New\ position\ X_i = current\ position\ X_i + New\ V_i \\ V_{max} \geq V_i \geq -V_{max} \end{aligned} \tag{6}$$

where c_1 and $c_2$indicate two positive constants termed learning factors (typically $c_1 = c_2 = 2$); rand () and Rand () represent two random functions in range [0, 1], V_{max}denotes upper limit on maximum particle velocity change, ωrepresents an inertia weight used as an enhancement to direct prior velocities' influence on current velocity. Theglobal search and local search is balanced by ω; and it is launched to reduce linearly with time from 1.4–0.5 value. Global search initiates huge weight and diminishes with time to favour local search over global search.

Second term as noticed in eq. (6) specifies cognition or private judgment of particle in comparison to prevailing position to its own position. Third term in eq. (6) specifies social collaboration prevalent amongst particles and also makes comparison among particle's current positions to best particle. In addition, to manage the changes which happens in the particles' velocities, the velocity change's lower and upper bounds restricted to value that is user-specified of V_{max}. While executing particle's new position eq. (5), particle, then, flees towards it. Consequently, chief factors enforced in PSO are number of generation cycles; population size (number of birds); maximum change of particle velocity V_{max} and ω.

Initialize an arbitrary number of nodes and indicated by either N or k. Relative edge weight G by initializing weight factor$\omega$is computed, and then fitness function of all nodes are estimated. For every sensor node distance, we determine best position as $pBest$. If fitness (i) is superior than$pBest$, $pBest(i) = fitness\ (i)$ function is fulfilled and execution completed, store the data in the nodes. In case the condition is not met, allocate energy cost to fitness value then evaluate temporary energy cost for every node. Compute temporary energy cost and it should be fulfilled with respect to each other and then terminate algorithm and data are successfully stored in nodes for purpose of future retrieval.

Pseudo-code of PSO algorithm:

1. Randomly create a population of agents

$$X_i = (P_1, P_2, P_3, \ldots\ldots\ldots, P_N)$$

2. Position of each particle with respect to objective function is estimated. In such case, total operational cost provided by C is considered by every particle and their fitness is evaluated (objective function minimization)
3. Cycle =1
4. Repeat
5. Velocity update of particles according to formula

$$\begin{aligned} V_i(t) = V_i(t-1) + C_i r_i(pbest(t) - x_i(t-1)) \\ + C_2 r_2(gbest(t) - x_i(t-1)) \end{aligned}$$

c = acceleration factor. r = random values among 1 and 0.

6. Calculate velocity to determine if range of

$$V_{max} \leq V_i \leq V_{min}$$

7. Move particles to new position

$$X_i(t) = X_i(t-1) + V_i(t)$$

8. Calculate to make certain that there's no exceeding of limits.

9. Compare fitness evaluation of particle with preceding pbest. Supposing the existing value is superior than prior pbest, then pbest value equal to current value is set along with pbest location equal to present location in dimensional search space, N.
10. Compare best current fitness with population gbest. If current value is superior than population gbest, then gbest to current best position is set with fitness value to current fitness value.
11. Check if criterion for stopping is attained. If not update cycle and return to step (5).
12. Complete when stopping criterion, herein is number of iterations, has been reached.

By employing these statistics, standardization of each feature is done as follows

$$\hat{x}_k = \frac{x_k - \bar{x}_k}{\sqrt{\sigma_k^2 + \epsilon}} \tag{7}$$

Where ϵ indicates small positive constant, to enhance numerical stability.

Nonetheless, reduction in the representational power of the layer happens as standardizing the intermediate activations. Then, additional learnable parameters γ and β are brought-in by batch normalization, which shift and scale data correspondingly, leading to form layer.

$$BN(X_k) = \gamma_k \hat{X}_k + \beta_k \tag{8}$$

By setting γ_k to σ_k and $\beta_k$to $\bar{x}_k$, recovery of the original layer representation can be performed by the network. So, a standard feed forward layer present in neural network

$$y = \phi(Wx + b) \tag{9}$$

Where W represents weights matrix, b points to bias vector, x indicates input of layer and φ indicates arbitrary activation function, batch normalization is utilized as trails:

$$y = \phi(BN(Wx)) \tag{10}$$

We observe that the bias vector has been avoided, because the standardization cancels its effect. Provided, that the normalization is now assumed to be network part, process of back propagation demands to be utilized for broadcasting gradients via mean and variance calculations. Statistics of mini-batch cannot be enforced, at time of test time. As an alternative, we can compute them by process of either forwarding various training mini-batches via network or by process of averaging statistics, or by preserving running average estimated over every mini-batch which is noticed while training.

Recurrent neural networks (RNNs)

RNN widens Neural Networks to sequential data. Provided, an input vector sequence $(x_1, \ldots, x_T)$, they result in a sequence of hidden states $(h_1, \ldots, h_T)$, which are evaluated at time step t and is given by

$$h_t = \phi(W_h h_{t-1} + W_x x_t) \tag{11}$$

Where$W_h$points to the recurrent weight matrix, $W_x$indicates the input-to-hidden weight matrix, and φ represents arbitrary activation function.

In case, a contact to whole input sequence, but from future ones can also be utilized, permitting bidirectional RNNs [52].

$$\overrightarrow{h}_t = \phi\left(\overrightarrow{W}_h \overrightarrow{h}_{t-1} + \overrightarrow{W}_x x_t\right) \tag{12}$$

$$h_t = \phi\left(W_h h_{t-1} + W_x x_t\right) \tag{13}$$

$$h_t = \left[\overrightarrow{h}_t : h_t\right]$$

where [x: y] indicates x and y concatenation. At last, RNNs stack by enforcing h as input to another different RNN, thus constructing architectures which are deeper [53].

$$h_t^l = \phi\left(W_h h_{t-1}^l + W_x h_t^{l-1}\right) \tag{14}$$

In vanilla RNNs, φ, activation function is generally considered to be a sigmoid function, like hyperbolic tangent. Training of networks is recognized as specifically arduous, as gradients are vanishing and exploding [54].

Long short-term memory

Long Short-Term Memory (LSTM), frequently enforced recurrent structure. It generally manages the issue of the vanishing gradient, which is noticed in vanilla RNNs by the process of incorporating the gating functions into its state dynamics. LSTM upholds hidden vector h as well as cell vector c accountable for controlling state updates and outputs process, at the time of time step. Particularly, we delineate the calculation during time step t as mentioned below [55]:

$$\begin{aligned} i_t &= sigmoid(W_{hi} h_{t-1} + W_x x_t) \\ f_t &= sigmoide\left(W_{hf} h_{t-1} + W_{hf} x_t\right) \\ c_t &= f_t \odot c_{t-1} + i_t \tanh \odot (W_{hc} h_{t-1} + W_{xc} x_t) \\ o_t &= sigmoide(W_{ho} h_{t-1} + W_{hx} x_t + W_{co} c_t) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{15}$$

Wheresigmoid(·) specifies logistic sigmoid function, tanh specifies hyperbolic tangent function, $W_{h\cdot}$ specifies recurrent weight matrices and $W_{x\cdot}$ specifies input-to-hiddent weight

matrices. i_t, f_t and ot are correspondingly input, forget and output gates, and $c_t$is cell.

Batch normalization for RNNs

From eq. (11), to process batch normalization to an RNN and is defined as given below:

$$h_t = \phi(BN(W_h h_{t-1} + W_x x_t)) \tag{16}$$

However, in trials, then enforce the batch normalization in this specific fashion, the training procedure wasn't benefitted, or we recommend making use of batch normalization to input-to-hidden transition only ($W_x x_t$), i.e. as given below:

$$h_t = \phi(W_h h_{t-1} + BN(W_x x_t)) \tag{17}$$

This concept is analogous to dropout [56] and enforced to RNNs [57]: batch normalization is functioned on connections that are vertical (that is from one layer to another) and not on the connections which are horizontal (which is in recurrent layer). Similar principle for LSTMs is enforced: after performing multiplication, batch normalization process is only used, with input-to-hidden weight matricesW_x.The images are normalized according to the procedures mentioned above.

Results and discussion

The proposed procedure requires no prior knowledge of the noise this represents a good advantage of the proposed approach compared to numerous current techniques that assume noise model to be known like the Gaussian noise. In reality, it will not hold true owing to nature of noise. The experimental results presented here are based on adding white noise, salt and pepper noise with zero mean and different variance values to demonstrate performance of proposed algorithm. The proposed technique was implemented using MATLAB. The experimental work is performed on SIMBA dataset. Database comprises of image set of 50 low-dose documented CT scans for identification. CT scans are acquired in single breath hold with 1.25 mm slice thickness. Nodules location detected by radiologist is offered.

Figures 1, 2, and 3 images are taken from the SIMBA dataset. Peak to Signal Noise Ratio (PSNR) and MSE are standard criteria reported in the literature for quantitative evaluation of the effectiveness of proposed image denoising methods. MSE and PSNR are metrics by which they evaluate absolute difference among two signals that are entirely quantifiable. PSNR is ratio amongst reference signal and distorted signal in image with decibels. If higher PSNR, then closer the distorted image to original image. Usually, higher PSNR value correlate to higher quality image, however it is not common in all case. MSE is average squared difference amongst distorted

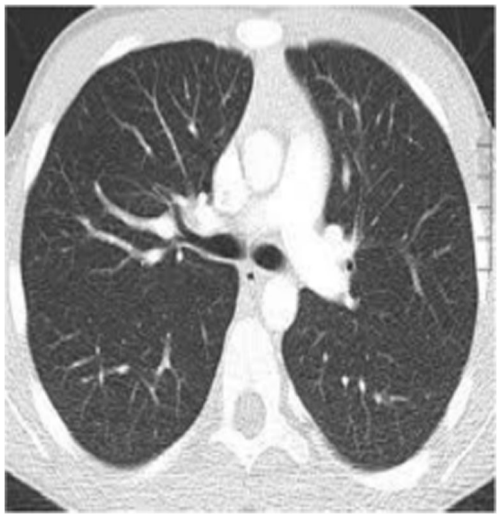

Fig. 2 Input image 2

Fig. 1 Input image 1

Fig. 3 Input image 3

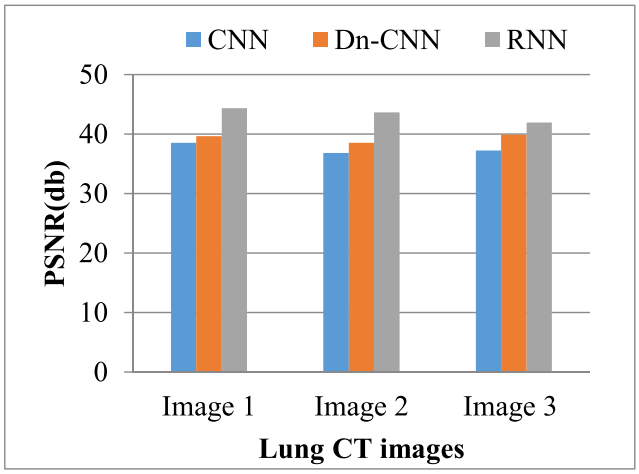

Fig. 4 Experimental results of PSNR comparison for lung CT images

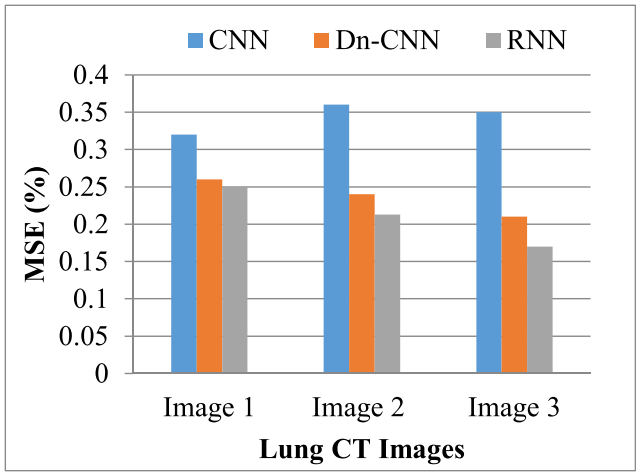

Fig. 6 Experimental results of MSE comparison for lung CT images

image and reference image. It is evaluated pixel-by-pixel by adding squared differences of pixels and partitioning total pixel count. Both MSE and PSNR are very important in image and video quality monitoring and are computed as below,

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2 \quad (18)$$

$$PSNR = 10.\log_{10}\left(\frac{MAX^2}{MSE}\right) \quad (19)$$

Where I and K are the original and the distorted images,respectively.$m$and n are number of pixels in both images (dimensions of images) and MAX equal to maximum probable pixel value.

Figure 4 illustrates PSNR of the input images based on the mean function. From this figure, it can be noticed that as the input image increases, the PSNR increases. Performance of anticipated RNN scheme is contrast with prevailing image denoising schemes like DnCNN, CNN. The PSNR results of the proposed RNN schema is higher after the noise applying when compare to existing denoising methods is shown in Fig. 4.

Figure 5 illustrates the PSNR of the lung CT images based on the white, salt and pepper noises. From this figure, it can be noticed that as the proposed RNN attained high PSNR performance compared than existing image denoising schemes like DnCNN, CNN. Due to the optimal batch selection and batch normalization the proposed RNN scheme attained better results.

Figure 6 illustrates the MSE of the lung CT images based on the mean function. From this figure, it can be noticed that as the input image increases, the MSE increases. Performance of anticipated RNN scheme is contrast to prevailing image denoising schemes like DnCNN, CNN. The MSE results of the proposed RNN schema is lesser after the noise applying when compare to existing denoising methods is shown in Fig. 6.

Figure 7 illustrates the MSE of the lung CT images based on the white, salt and pepper noises. From this figure, it can be noticed that as the proposed RNN attained less MSE error rate

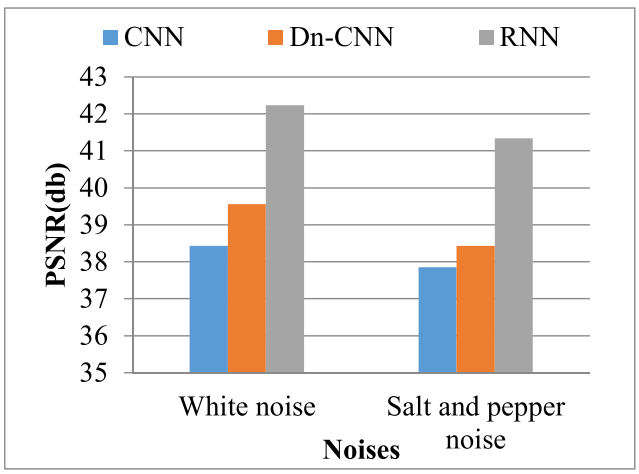

Fig. 5 Experimental results of PSNR comparison against noises

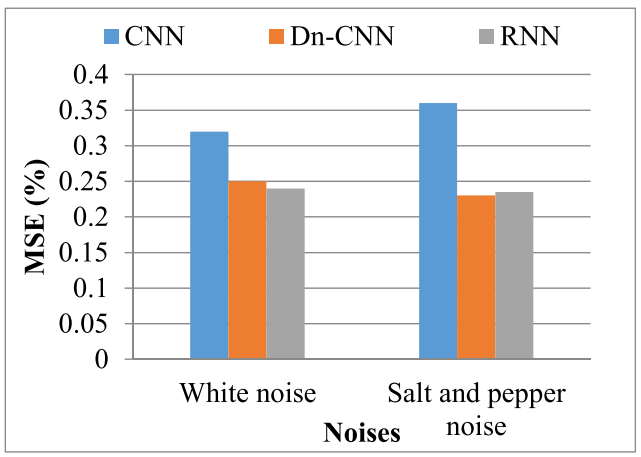

Fig. 7 Experimental results of MSE comparison against noises

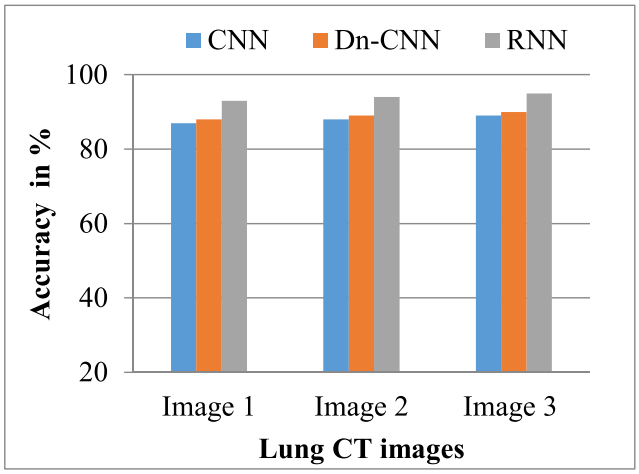

Fig. 8 Experimental results of accuracy comparison for lung CT images

compared than existing image denoising schemes like DnCNN, CNN. Due to the LSTM based batch normalization the proposed RNN scheme attained better results.

From this Fig. 8, it can be noticed that as the proposed RNN attained high accuracy compared than existing image denoising schemes like DnCNN, CNN.

Conclusion

Here, an effective denoising scheme to remove white, salt and pepper noises by combining LSTM based Batch Normalization and RNN techniques. In this work, first, the input images are processed. Then, an effective batch size is computed by using PSO. Finally, the RNN algorithm is proposed for denoising the image. In RNN, LSTM based Batch Normalization introduced to reduce internal covariate shift in neural networks. The PSNR and MSE results show that the proposed RNN attained better results compared than other denoising schemes like DnCNN and CNN. Comparison experimental outcome validates superior denoising capability of anticipated technique and specify that it offers an effective solution to image denoising. In future, some other denoising algorithms will focus to improve the PSNR performance result.

Compliance with Ethical Standards

Conflict of Interest The authors have no conflict of interests and the paper has not been submitted elsewhere.

Human and Animal Rights This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent The work does not involve any human or animal participants. The datasets used in the work are taken from free online sources.

References

1. Chatterjee, P., and Milanfar, P., Patch-based near-optimal image denoising. IEEE Trans. Image Process. 21(4):1635–1649, 2012.
2. Zhang, J., Xiong, R., Zhao, C., Ma, S., and Zhao, D., Exploiting image local and nonlocal consistency for mixed Gaussian-impulse noise removal. Proc. IEEE Int. Conf. Multimedia Expo, Jul.:592–597, 2012.
3. Chen, Y., and Liu, K. J. R., Image denoising games. IEEE Trans. Circuits Syst. Video Technol. 23(10):1704–1716, 2013.
4. Rajwade, A., Rangarajan, A., and Banerjee, A., Image denoising using the higher order singular value decomposition. IEEE Trans. Pattern Anal. Mach. Intell. 35(24):849–862, 2013.
5. Mäkitalo, M., and Foi, A., Noise parameter mismatch in variance stabilization, with an application to Poisson–Gaussian noise estimation. IEEE Trans. Image Process. 23(12):5348–5359, 2014.
6. Zuo, W., Zhang, L., Song, C., Zhang, D., and Gao, H., Gradient histogram estimation and preservation for texture enhanced image denoising. IEEETrans. Image Process. 23(6):2459–2472, 2014.
7. Peng, H., Rao, R., and Dianat, S. A., Multispectral image denoising with optimized vector bilateral filter. IEEE Trans. Image Process. 23(1):264–273, 2014.
8. Talebi, H., and Milanfar, P., Global image denoising. IEEE Trans. ImageProcess. 23(2):755–768, 2014.
9. Zhang, J., Zhao, D., Xiong, R., Ma, S., and Gao, W., Image restoration using joint statistical modeling in a space-transform domain. IEEE Trans. Circuits Syst. Video Technol. 24(6):915–928, 2014.
10. Huang, C.-T., Bayesian inference for neighborhood filters with application in denoising. Proc. IEEE Conf. Comput. Vis. Pattern Recognit.:1657–1665, 2015.
11. Chen, F., Zeng, X., and Wang, M., Image denoising via local and nonlocal circulant similarity. J. Vis. Commun. Image Represent. 30:117–124, 2015.
12. Chouzenoux, E., Jezierska, A., Pesquet, J.-C., and Talbot, H., A convex approach for image restoration with exact Poisson–Gaussian likelihood. SIAM J. Imag. Sci. 8(4):2662–2682, 2015.
13. Tomasi, C., and Manduchi, R., Bilateral filtering for gray and color images. '. Int.Conf. Computer Vision:839–846, 1998.
14. Perona, P., and Malik, J., Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. 1(7):629–639, 1990.
15. Rudin, L. I., Osher, S., and Fatemi, E., Nonlinear total variation based noise removal algorithms. Physica D 1(1):259–268, 1992.
16. Osher, S., Burger, M., Goldfarb, D. et al., An iterative regularization method for total variation-based image restoration. Multiscale Model. Simul. 5(2):460–489, 2005.
17. Donoho, L., De-noising by soft-thresholding. IEEE Trans. Inf. Theor. 1(3):613–627, 1995.
18. Chang, S. G., Yu, B., and Vetterli, M., Adaptive wavelet thresholding for image denoising and compression. IEEE Trans. Image Process. 1(9):1532–1546, 2000.
19. Starck, J. L., Candés, E. J., and Donoho, D. L., The curvelet transform for image denoising. IEEE Trans. Image Process. 1(6):670–684, 2002.
20. Elad, M., and Aharon, M., Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. 3(12):3736–3745, 2006.
21. Mairal, J., Bach, F., Ponce, J. et al., Non-local sparse models for image restoration. Int. Conf. Computer Vision 6:2272–2279, 2009.
22. Dong, W., Zhang, L., Shi, G. et al., Non locally centralized sparse representation for image restoration. IEEE Trans. Image Process. 6(4):1620–1630, 2013.
23. Xu, J., Zhang, L., Zuo, W. et al., Patch group based nonlocal self-similarity prior learning for image denoising. Int. Conf. Computer Vision 6:244–252, 2015.

24. Dabov, K., Foi, A., Katkovnik, V. et al., Image denoising by sparse 3-Dtransform-domain collaborative filtering. IEEE Trans. Image Process. 6(8):2080–2095, 2007.
25. Ji, H., Liu, C., Shen, Z. et al., Robust video denoising using low rank matrix completion. Computer Vision and Pattern Recognition 2:1791–1798, 2010.
26. Gu, S., Zhang, L., Zuo, W. et al., Weighted nuclear norm minimization withapplication to image denoising. Computer Vision and Pattern Recognition 6:2862–2869, 2014.
27. Luo, E., Chan, S. H., and Nguyen, T. Q., Adaptive image denoising by mixture adaptation. IEEE Trans. Image Process. 25(10):4489–4503, 2016.
28. Fei, C., Lei, Z., and Huimin, Y., External patch prior guided internal clustering for image denoising. Int. Conf. Computer Vision 6:7–13, 2015.
29. Gu, S., Zhang, L., Zuo, W. et al., Weighted nuclear norm minimization withapplication to image denoising. Computer Vision and Pattern Recognition 3:2862–2869, 2014.
30. Yuan, X., Shuhang, G., Yan, L. et al., Weighted schatten p-normminimization for image denoising and background subtraction. IEEE Trans. Image Process. 8(10):4842–4857, 2016.
31. Burger, H. C., Schuler, C., and Harmeling, S., Learning how to combine internal and external denoising methods. Pattern Recogn.:121–130, 2013.
32. Chen, Y., and Pock, T., Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration to appear. IEEE Trans. Pattern Analysis and Machine Intelligence 8(10):1–1, 2016.
33. Jain, V., and Seung, S., Natural image denoising with convolutional networks. Adv. Neural Inf. Proces. Syst.:769–776, 2009.
34. Burger, H. C., Schuler, C. J., and Harmeling, S., Image denoising: can plain neuralnetworks compete with BM3D? Computer Vision and Pattern Recognition:2392–2399, 2012.
35. Xie, J., Xu, L., and Chen, E., Image denoising and in painting with deep neural networks. Adv. Neural Inf. Proces. Syst. 1:341–349, 2012.
36. Dong, C., Loy, C. C., He, K. et al., Learning a deep convolutional network for image super-resolution. European Conf. Computer Vision:184–199, 2014.
37. Wenzhe, S., Jose, C., Ferenc, H. et al., Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Computer Vision and Pattern Recognition 12:27–30, 2016.
38. Dong, C., Chen, C. L., and Tang, X., Accelerating the super-resolution convolutional neural network. European Conf. Computer Vision:391–407, 2016.
39. Wang, X., Tao, Q., Wang, L. et al., Deep convolutional architecture fornatural image denoising. Int. Conf. Wireless Communications and Signal Processing 5(53):1–4, 2015.
40. Koziarski, M., and Cyganek, B., Deep neural image denoising. Int. Conf. Computer Vision and Graphics:163–173, 2016.
41. Lefkimmiatis, S., Non-local color image denoising with convolutional neural networks', arXiv, 2016
42. Zhang, F., Cai, N., Wu, J., Cen, G., Wang, H., and Chen, X., Image denoising method based on a deep convolution neural network. IET Image Process., 2017.
43. Xiong, R., Liu, H., Zhang, X., Zhang, J., Ma, S., Wu, F., and Gao, W., Image denoising via bandwise adaptive modeling and regularization exploiting nonlocal similarity. IEEE Trans. Image Process. 25(12):5793–5805, 2016.
44. Liu, D., Wen, B., Liu, X., & Huang, T. S., When Image Denoising Meets High-Level Vision Tasks: A Deep Learning Approach. arXiv preprint arXiv:1706.04284, 2017.
45. Godard, C., Matzen, K., and Uyttendaele, M., Deep Burst Denoising. arXiv preprint arXiv:1712.05790, 2017.
46. Remez, T., Litany, O., Giryes, R., & Bronstein, A. M., Deep class aware denoising. arXiv preprint arXiv:1701.01698, 2017.
47. Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L., Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans. Image Process. 26(7):3142–3155, 2017.
48. Sato, M., Otake, T., Aomori, H., & Tanaka, M., New image denoising method using multiple-minimum cuts based on maximum-flow neural network. In Circuit Theory and Design (ECCTD), 2015 European Conference on (pp. 1-4). IEEE, 2015.
49. Al-Sbou, Y. A., Artificial neural networks evaluation as an image denoising tool. World Appl. Sci. J. 17(2):218–227, 2012.
50. LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R., Efficient backprop, In Neural networks: Tricks of the trade, pp. 9–48. Springer, 2012.
51. Ioffe, S., and Szegedy, C., Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167, 2015.
52. Schuster, M., and Paliwal, K. K., Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45(11):2673–2681, 1997.
53. Pascanu, R., Mikolov, T., and Bengio, Y., On the difficulty of training recurrent neural networks. arXiv preprint arXiv:1211.5063, 2012.
54. Hochreiter, S., and Schmidhuber, J., Long short-term memory. Neural Comput. 9(8):1735–1780, 1997.
55. Gers, F. A., Schraudolph, N. N., and Schmidhuber, J., Learning precise timing with LSTM recurrent networks. The Journal of Machine Learning Research 3:115–143, 2003.
56. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15(1):1929–1958, 2014.
57. Zaremba, W., Sutskever, I, and Vinyals, O., "Recurrent neural network regularization,"arXiv preprint arXiv:1409.2329, 2014.
58. Laurent, C., Pereyra, G., Brakel, P., Zhang, Y., Bengio, Y, Batch normalized recurrent neural networks. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016.