IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 4, 2021, accepted August 16, 2021, date of publication August 23, 2021, date of current version August 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3106707

Augmented Noise Learning Framework for Enhancing Medical Image Denoising

SWATI RAI, JIGNESH S. BHATT, AND S. K. PATRA, (Senior Member, IEEE)
Indian Institute of Information Technology (IIIT) Vadodara, Gandhinagar 382027, India

Corresponding author: Jignesh S. Bhatt (jignesh.bhatt@iiitvadodara.ac.in)

This work was supported by the Indian Institute of Information Technology Vadodara (IIITV).

ABSTRACT Deep learning attempts medical image denoising either by directly learning the noise present or via first learning the image content. We observe that residual learning (RL) often suffers from signal leakage while dictionary learning (DL) is prone to Gibbs (ringing) artifacts. In this paper, we propose an unsupervised noise learning framework that enhances denoising by augmenting the limitation of RL with the strength of DL and vice versa. To this end, we propose a ten-layer deep residue network (DRN) augmented with patch-based dictionaries. The input images are presented to patch-based DL to indirectly learn the noise via sparse representation while given to the DRN to directly learn the noise. An optimum noise characterization is captured by iterating DL/DRN network against proposed loss. The denoised images are obtained by subtracting the learned noise from available data. We show that augmented DRN effectively handles high-frequency regions to avoid Gibbs artifacts due to DL while augmented DL helps to reduce the overfitting due to RL. Comparative experiments with many state-of-the-arts on MRI and CT datasets (2D/3D) including low-dose CT (LDCT) are conducted on a GPU-based supercomputer. The proposed network is trained by adding different levels of Rician noise for MRI and Poisson noise for CT images considering different nature and statistical distribution of datasets. The ablation studies are carried out that demonstrate enhanced denoising performance with minimal signal leakage and least artifacts by proposed augmented approach.

INDEX TERMS Augmented noise learning, deep residue network, denoising, dictionary learning, inverse ill-posed problem, unsupervised learning.

I. INTRODUCTION

Noise is the unwanted energy which is generally intervened during the acquisition, transmission, and/or reconstruction of an image. Though the noise cannot be altogether eliminated, however, it can be reduced at acquisition time. Post-processing of acquired imagery using signal processing algorithms is used to reduce its effects. In such applications, denoising is a major challenge for the researchers [1]–[3]. Denoising is an inverse ill-posed problem [4] which is classically addressed by specifying a forward model and then invert it for the unknowns [5]. Recent developments are exploring the use of deep learning techniques for the denoising [6]–[10].

Denoising is the fundamental step in the medical diagnosis [11], [12], since doctors and medical practitioners most often rely on these processed images. In particular, magnetic resonance imaging (MRI) and computed tomography (CT) including low-dose CT (LDCT) scans are used to observe the internal structure as well as any defects like tumors or injuries present inside the body. Generally, MRI, CT and LDCT images are affected by the noise due to fluctuations in temperature of the scanning room, disturbance in the scanning machines and/or patient's movement during the image acquisition. Due to the noise, magnitudes of the pixel/voxel in the images/image stack are perturbed which lead to artifacts and loss of details in the images. It makes the diagnosis and disease prediction complicated.

Therefore, the main considerations involved in medical image denoising algorithms include: a) edges in the denoised images should be preserved, i.e., filtering performed for denoising should not blur out the finer details of imagery and while at the same time, b) the visual quality of the denoised images should be preserved and improved, and (c) minimal signal leakage should be observed in the denoised images. To this end, we propose a novel unsupervised deep learning method augmenting patch-based dictionary learning (DL)

The associate editor coordinating the review of this manuscript and approving it for publication was Cheng Chin.

and residual learning (RL) in order to learn the noise in augmented manner and construct a novel dictionary-based deep residue network (DRN) for denoising of 2*D* and 3*D* MRI and CT data including LDCT images. We perform ablation studies on the proposed DL and DRN parts in order to show signal (image) leakage [13] in the estimated denoised images and for better dissemination of results.

Rest of the paper is arranged as follows: we begin with literature review in Section II. Section III provides motivation and defines the problem. Section IV explains the methodology used for denoising 2*D*/3*D* MRI and CT data as well as LDCT images. We also present an analysis of the proposed approach. The results obtained after implementing our proposed model to the noisy MRI/CT/LDCT datasets along with the ablation studies, quantitative, and quantitative comparisons with state-of-the-art are shown in Section V. Finally, the paper is concluded in Section VI with possible future direction.

II. RELATED WORK

Over the years, various medical image denoising methods have been proposed [14]–[20]. By and large, four broad philosophies are adopted: (a) filtering, (b) transformation, (c) statistical, and (d) learning-based methods. With the recent advances in computer technology and available resources, learning-based methods have gained a lot of attention. Hence, we review the learning-based approaches for denoising the MRI, CT and LDCT images.

The learning-based denoising approaches can be divided into three categories: supervised learning, semi-supervised learning, and unsupervised learning. In supervised learning, the model is trained with available data sets from which it can learn features called pre-learning or it can learn these features simultaneously during image reconstruction. It is found that the images are denoised using the supervised learning approach by incorporating wavelet transform (WT), curvelet transform (CuT), and optimization techniques in machine learning frameworks [21]. The compressed sensing (CS) technique is used in denoising MRI images and called as CS-MRI. The CS is included with a dictionary learning approach to learn an overcomplete dictionary using k-singular value decomposition (K-SVD) method to give a sparse representation of an image [22]. CS-MRI is used to reconstruct MRI images consuming less acquisition time in a supervised way [23]. Again dictionary learning is used along with CS to reconstruct MRI images by training the model with denoised images [24]. Subsequently, Bayesian approach is implemented with dictionary learning to denoise the MRI images [25]. Recently, the deep learning approach is explored with the classical methods to denoise the MRI images [26]. Supervised learning is practiced to enhance the quality of MRI, CT, and LDCT images by removing noise and reducing the artifacts from them [27]. Very recently, directionality component is added to enhance the dictionary learning for MRI image reconstruction [28]. However, when only DL is used to denoise the image, there is formation of Gibbs artifact, also known as ringing artifact in the high-frequency regions of the images [29]. Along with this, there is a min-imax problem if proper sparse thresholding is not chosen to estimate the sparse coefficients from the dictionary and given images [30], [31].

A semi-supervised deep learning approach is used to reduce the noise from LDCT images without using original projection data by training the model with less number of denoised images [32], [33]. The LDCT images are mapped to their respective normal-dose part in a patch-by-patch manner using a deep convolutional neural network (CNN). Again, for LDCT images, a residual encoder-decoder CNN (RED-CNN) is formed by autoencoders and deconvolutional network which help in noise removal along with structural preservation and lesion detection [34]. This uses normal-dose and LDCT images to train the network. Deep feed-forward CNN is then used to reduce noise from the images taking lesser number of clean images [35]. This uses residual learning (RL) while batch normalization is used for regularization. Recently, a work on generative adversarial network (GAN) is carried out to denoise the medical images [36]. GAN is modified to Wasserstein GAN (WGAN) in order to denoise the MRI images in a semi-supervised manner [37]. A conveying path-based convolutional encoder-decoder (CPCE) is used to denoise the LDCT images using transfer learning concept within GAN framework [38]. Also modularized adaptive processing neural network (MAP-NN) is employed to denoise the LDCT images in progressive manner [39].

It is a well-known fact in medical imaging that the availability of training dataset and ground truth are scare to train the model with supervised or semi-supervised settings. Therefore, a better approach is to investigate unsupervised learning models that can learn on their own only from the available images and could generate high-quality denoised medical images. A lot of attention is being given to LDCT as it reduces the risk on patients. To give promising results for CT images and to keep the crucial information intact, GAN is combined with perceptual similarity and Wasserstein distance using unsupervised learning [40]. In order to denoise LDCT images, a tunable CycleGAN is proposed using only one generator that makes use of adaptive instance normalization (AdaIN) layers [41]. A deep neural network is recently trained in an unsupervised way using Poisson unbiased risk estimator (PURE) to denoise the LDCT image [42]. The RL is a promising method to learn the noise from the images, still, overfitting occurs as it starts learning the image content along with the noise [43], [44]. Due to this, a part of image content is leaked in the estimated noise content. This leads to correlation among the learned noise and obtained denoised image. The use of skip connections in the RL method add to blur effect in the denoised image. It is to match the dimensions of skip layer and previous layer one has to use either convolution or zero-padding [45]. Therefore, the use of RL in unsupervised manner also limits denoising performance.

III. MOTIVATION AND PROBLEM FORMULATION

In this section, we first provide motivation and then define the problem for enhancing denoising of MRI, CT, and LDCT images. In general, one may either (directly) learn the noise from given images or learn an image representation which in turn provides the noise content (indirect learning). It is known that the RL is a widely used contemporary method to learn noise (residue) from the images while dictionary learning (DL) is a well-accepted image representation method, using which, one may indirectly learn the noise.

We observe that, a residue network, in general, invades parts of image content into the learned noise, i.e., very often may allow part of the image content into the learned noise, and exhibits signal (image) leakage in the corresponding denoised image as shown in Fig. 1 (a) and 1 (c). This shows the local correlation values between estimated noise and denoised image using [13]. Here, blue color indicates less signal leakage. One can observe that a part of image information is learnt along with the noise and contributed to signal leakage by showing higher local correlation values. It mainly happens due to overfitting by a residue network and hence degrade the denoising performance (Fig. 1 (a) and 1 (c)). One has to do extra convolutions or zero-padding to match the different dimensions at skip connections, which may introduce blurring in the denoised image. The RL process very often applies a heuristic threshold to decide on normalizing the learned noise. Hence, we note the degradation in denoising performance when using a residue network.

On the other hand, one may first employ learning the image representation, and then indirectly extract the residue (noise). To this end, the use of dictionary learning (DL) is popular for its representational ability from available data. However, it is found that such indirectly noise learned denoised images shows Gibbs (ringing) artifacts (Fig. 1(b) and 1 (d)), as well as minimax problems due to sparsity threshold while constructing the dictionary. Consequently, it also leads to poor denoising performance for medical imagery.

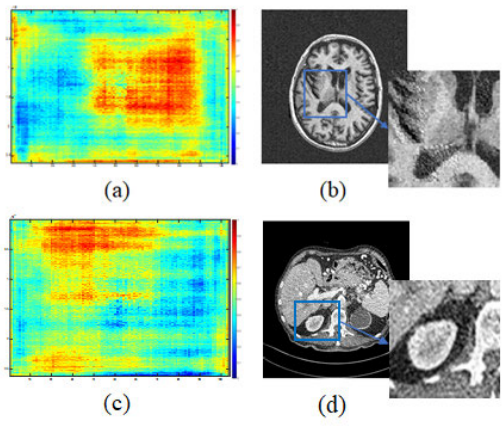

FIGURE 1. Illustrations on medical images: MRI brain scans [46]: (a) signal leakage, and (b) Gibbs artifact; LDCT images [47]: (c) signal leakage, and (d) Gibbs artifact.

Hence, we are motivated to enhance the medical image denoising performance and address the limitations of both the RL and DL approaches. To this end, we propose a dictionary-based DRN approach. We develop an augmented noise learning framework that augments the RL with the DL and vice versa.

Problem Statement: Given a medical MRI/CT/LDCT 2*D*/3*D* image/stack, our objective is to enhance the denoising process of 2*D*/3*D* image/stack such that critical contents especially at edges in the estimated denoised images are preserved while visual information (quality) of the resultant (denoised) images is improved. At the same time, the signal leakage phenomena and Gibbs artifacts are reduced in the estimated denoised images. In this work, we resort to the well-accepted data model for denoising medical images [2], [14], [35], [48],

$$\mathbf{Y} = \mathbf{X} + \mathbf{R}, \tag{1}$$

where $\mathbf{Y}$ is available (given) image, $\mathbf{X}$ is the corresponding denoised image (unknown), and $\mathbf{R}$ is the noise. We conveniently consider that MRI images are corrupted by the Rician noise [15], [49], [50] while CT/LDCT images are corrupted by the Poisson noise [18], [51], [52]. Hence, given the $\mathbf{Y}$ image, our objective is to estimate the noise $\widehat{\mathbf{R}}$ and in turn evaluate the denoised image $\widehat{\mathbf{X}}$ (referring to equation 1), such that estimated denoised image $\widehat{\mathbf{X}}$ is close to the $\mathbf{X}$, both qualitatively and analytically.

IV. METHODOLOGY

Block diagram of the proposed augmented noise learning approach is shown in Fig. 2. It mainly comprises of two parts: dictionary learning (DL) and deep residue network (DRN). The MRI and CT data are available in the form of a 3*D* image cube. We develop a model for both 3*D* and 2*D* processing of the MRI/CT including LDCT data considering different generations of the scanning machines. Note that while the proposed framework is generalized for 3*D* and 2*D* processing, nevertheless, user can choose to perform either 3*D* block or 2*D* image processing.

For 3D Processing: Here, we consider a block of image cube comprising of voxels. These volume patches (3*D* blocks) $\{\mathbf{Y}_i\}_{i=1}^{l}$ where i is the index of an image and l is the total images in the 3*D* cube. For voxel processing, each 3*D* block is of dimension $N \times N \times Q$ voxels. These images are first given to the decomposition stage. Here such 3*D* block is divided into overlapping block (volume) patches $\{\mathbf{P}_j\}_{j=1}^{r}$ each of size $n \times n \times q$ voxels, where $n \ll N$, and j is the index of the block patch chosen from a total of r block patches. These patches $\{\mathbf{P}_j\}_{j=1}^{r}$ are fed to the DL and DRN parts for processing, again one block patch $\mathbf{P}_j$ at a time.

For 2D Processing: In this case, we consider one image at a time from the image cube, each image has a dimension of $N \times N$ pixels and is decomposed into image patches of size $n \times n$ pixels. Now, these obtained patches $\{\mathbf{P}_j\}_{j=1}^{r}$ are given to the DL and DRN parts for further processing, again one image patch $\mathbf{P}_j$ at a time.

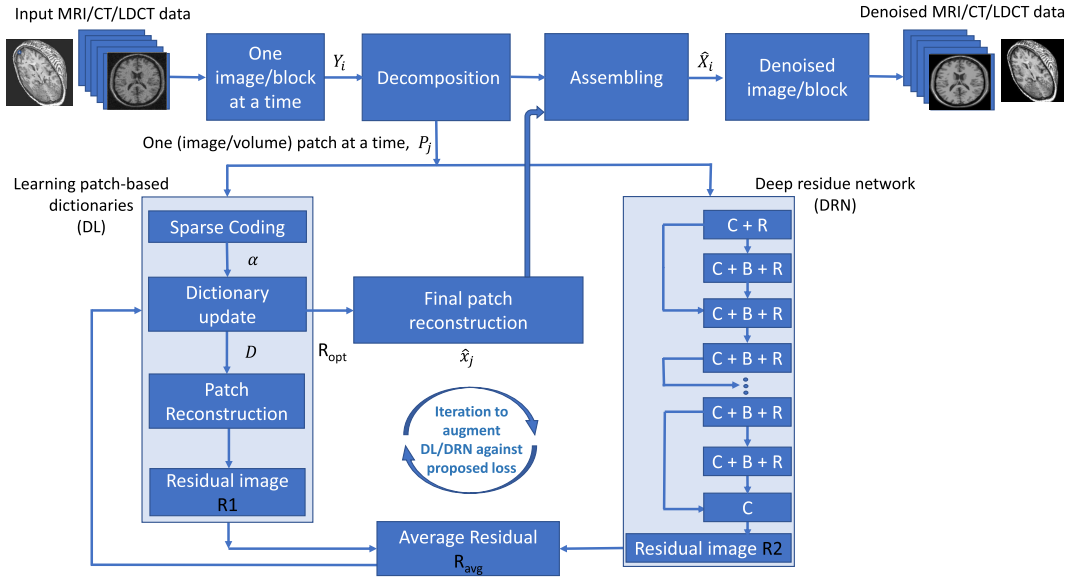

FIGURE 2. Block diagram of the proposed augmented noise learning DL/DRN approach for MRI/CT/LDCT denoising. C = Convolution, R = ReLU, and B = Batch normalization. All the functions are in 3*D* for voxel (block) processing and 2*D* for pixel (image) processing.

Referring to Fig. 2, there are three steps in the DL part: (a) sparse coding, (b) dictionary update, and (c) patch reconstruction. The role of DL part is to provide efficient representation of input MRI/CT/LDCT so that, in turn, we have estimate of noise content via the sparse representation of information. This is an indirect way of learning noise characteristics of the data. To start the DL process, we use an initial dictionary $\mathbf{D}_{init}$ of size $m \times k \times q$ for block processing and of size $m \times k$ for image processing, obtained using the discrete cosine transform. We consider an overcomplete dictionary since it has basis vectors greater than the dimension of the input patch vector, which allows to better capture underlying characteristics of the data. One may notice that for medical images, capturing the underlying information is vital for better processing and the final diagnosis. With the initial dictionary $\mathbf{D}_{init}$ and available patch $\mathbf{P}_j$, we first obtain the sparse coefficient $\boldsymbol{\alpha}_j$ of dimension $k \times 1 \times q$ for block processing and of $k \times 1$ for image processing, i.e., the sparse representation of a patch $\mathbf{P}_j$ is considered as:

$$\mathbf{P}_j \approx \mathbf{D}_{init}\ \boldsymbol{\alpha}_j, \tag{2}$$

where the sparse coefficients of a image/block patch is computed using orthogonal matching pursuit [53] as,

$$\widehat{\boldsymbol{\alpha}}_j = \min_{\boldsymbol{\alpha}_j} \left(\frac{1}{2} ||\mathbf{P}_j - \mathbf{D}_{init}\boldsymbol{\alpha}_j||_2^2 + \mu ||\boldsymbol{\alpha}_j||_0 \right). \tag{3}$$

Here μ is the regularization parameter. We now estimate sparse dictionary $\mathbf{D}$ using the estimated sparse coefficients $\widehat{\boldsymbol{\alpha}}_j$,

$$\mathbf{D} = \arg\min_{\mathbf{D}} \sum_{j=1}^{n} ||\mathbf{P}_j - \mathbf{D}\widehat{\boldsymbol{\alpha}}_j||_2^2 \quad \text{such that} \quad ||\boldsymbol{\alpha}_j||_0 \leq s, \tag{4}$$

where s is the sparsity. To this end, in order to update the dictionary, we employ K-SVD algorithm [22]. Now this updated dictionary $\mathbf{D}$ and the estimated sparse coefficients $\widehat{\boldsymbol{\alpha}}_j$ are used to reconstruct denoised image/block patch $\widehat{\mathbf{X}}_j$ as:

$$\widehat{\mathbf{X}}_j = \mathbf{D}\ \widehat{\boldsymbol{\alpha}}_j. \tag{5}$$

The residual patch $\mathbf{R1}_j$ can now be extracted by taking absolute difference of estimated denoised image/block patch $\widehat{\mathbf{X}}_j$ and available input image/block patch $\mathbf{P}_j$ as,

$$\mathbf{R1}_j := |\mathbf{P}_j - \widehat{\mathbf{X}}_j|, \quad \forall j. \tag{6}$$

See that in equation (6) we have used absolute subtraction between the given patch and estimated denoised patch referring to our data model equation (1). Note that the residue patch $\mathbf{R1}_j$ consists of part of the noise contents due to representational limitations at the time of image acquisition. Thus, the proposed DL part indirectly learned the noise characteristics from the MRI/CT/LDCT data. However, DL suffers from minimax problem if proper thresholding value is not set. Along with this ringing artifact also occurs at the high frequency regions. To overcome these limitations we iteratively use the DRN part as discussed below.

Now the residual learning (RL) using proposed DRN (Fig. 2) is designed to directly learn the noise characteristics present in the patches. As shown in Fig. 2, we pass the image/block patch $\mathbf{P}_j$ through the proposed DRN having depth t comprises of the following layers: (a) First layer (C + R): C stands for the convolution process that is performed between a patch and a filter. Note that it will be 2*D* convolution for image processing while 3*D* convolution for block (volume) processing. Convolution helps to extract the features of the image/block and generate feature maps.

In particular, 84 filters of size 3×3 are employed that give rise to 84 feature maps. Then rectified linear units (R) are used to introduce the non-linearity by using the $max(0, \cdot)$ function. (b) Second layer to $(t-1)$ layer (C + B + R): Here, batch normalization (B) is introduced in between C and R. The B acts as a regularizer term and helps the network to use higher learning rates which in turn uplifts the denoising performance. Note that there are skip connections added in between alternate layers in DRN (Fig. 2). The layers having same dimension receive identity connection from the previous layer. While convolution layer is added in between the identity connection if dimension of recent input and previous input data is different. (c) Last layer (C): Finally convolution is performed to give the residual $\mathbf{R2}_j$ (noise) part learned from the input image/block patch. In this way, the proposed augmented noise learning framework directly learns the noise from image/block MRI and CT including LDCT data.

Hence we now have learned the noise in the form of $\mathbf{R1}_j$ residue using the DL, while $\mathbf{R2}_j$ residue using the DRN. We construct residue $\mathbf{R}_{avg_j}$ by averaging $\mathbf{R1}_j$ and $\mathbf{R2}_j$, pixel-by-pixel, to preserve the noise characteristics learned by representation content (DL) and residue (DRN). One may note that it essentially improves the denoising performance unlike other residual learning based approaches who heuristically decide a threshold to normalize the residue [35]. The averaged residue $\mathbf{R}_{avg_j}$ is then fed back to the DL stage in order to again update the estimated dictionary $\mathbf{D}$, as shown in Fig. 2:

$$\mathbf{D} = \arg\min_{\mathbf{D}} \lambda \sum_{j=1}^{r} ||\mathbf{P}_j - \mathbf{D}\boldsymbol{\alpha}_j||_2^2 + \mu||\boldsymbol{\alpha}_j||_0 + \frac{1}{r}||\mathbf{R}_{avg_j} - \mathbf{R1}_j||^2, \quad (7)$$

where λ is regularization parameter. Now the updated dictionary $\mathbf{D}$ is used to generate the optimum residue $\mathbf{R}_{opt}$ and give the final estimated denoised patch $\widehat{\mathbf{X}}_{opt_j}$ as:

$$\widehat{\mathbf{X}}_{opt_j} := |\mathbf{P}_j - \mathbf{R}_{opt_j}|, \quad \forall j. \quad (8)$$

The entire process from equations (2) to (8) is repeated for all the r image/block patches of an input image/block $\mathbf{Y}_i$. Finally estimated denoised patches are assembled to form an entire estimated denoised image/block $\widehat{\mathbf{X}}_i$ (Fig. 2). Note that the patches are overlapping, therefore, the voxels and pixels in the overlapping regions in the $\widehat{\mathbf{X}}_i$ are considered by local patch-level averaging. Finally, as shown in Fig. 2, the process is repeated for each image/block in the stack of MRI/CT/LDCT images and estimate corresponding denoised stack of MRI/CT/LDCT images.

A. ANALYSIS OF PROPOSED AUGMENTED NOISE LEARNING METHOD

In this subsection, we analyse the proposed approach for its effectiveness and conduct analytical comparison with respect to state-of-the-art approaches. It is observed that noise is most often mixed in MRI, CT and LDCT images during the image acquisition process. Therefore we begin with imaging processes of MRI, CT, and LDCT. In MRI, a patient's body is exposed to a very strong magnetic field, radio waves, and magnetic field gradients [54]. During the acquisition, both the frequency and phase of the MRI signals, called raw MRI data, are accumulated in a temporary image space and then inverse Fourier transform is computed to form a grayscale MRI image. It is found that in MRI, the probability density function (PDF) of noise follows the Rician distribution [15]. Hence, referring to the data model in equation (1), one may write the conditional PDF of MRI data as,

$$p_{\mathbf{Y}}(\mathbf{Y}|\mathbf{X}) = \frac{\mathbf{Y}}{\sigma^2} e^{\frac{-(\mathbf{X}^2+\mathbf{Y}^2)}{2\sigma^2}} I_0\left(\frac{\mathbf{X}\,\mathbf{Y}}{\sigma^2}\right), \quad (9)$$

where $\mathbf{Y}$ is the acquired image, σ is the noise variance, $\mathbf{X}$ is the noiseless image intensity level (unknown), and $I_0(\cdot)$ is the zeroth order modified Bessel function used to induce smoothness in the curve. In CT scans, a thin beam of X-rays is passed through a patient's body from the source that is captured by the X-ray detectors, located opposite to the X-ray source [55]. These signals are processed by the computer and cross-sectional images of the patient's body are generated. In CT images, the most common noise is the Poisson noise [18]. This is mainly due to the usage of X-rays and scanning methods in the generation of the CT scans. The probability mass function (PMF) of CT data can thus be written in reference to the data model equation (1) as,

$$p(\mathbf{Y}|\mathbf{X}) = \frac{e^{-\mathbf{X}t}(\mathbf{X}t)^{\mathbf{Y}}}{\mathbf{Y}!}, \quad (10)$$

where $\mathbf{Y}$ is the amount of photons (image intensities) measured over time interval t by the sensor element, and $\mathbf{X}$ is the expected amount of photon (corresponding denoised image content) per unit time. Similarly, LDCT images are created using very low radiation dose compared to a conventional CT scan. The LDCT is preferred for patients having high risk of getting infected by cancer or other disease. However, use of low-dose of radiation affects the quality of CT image and hence provide more challenges for denoising [38]. This image acquisition process also many times is corrupted with Poisson noise. It can be seen from equations (9) and (10) that both Rician and Poisson noises mainly affect the magnitudes of the MRI and CT including LDCT images, respectively. Practically, the MRI and CT as well as LDCT images are perturbed primarily due to following causes [56], [57]: (a) Ambient temperature is not maintained inside the scanner room. Typical range of temperature required to maintain is 23° to 24° Celsius. Any variation beyond the said temperature range can generate artifacts in the acquired images. (b) Number of detectors used to capture the images. More number of detectors can reduce the scanning time, however, probability of noise is also increased with more detectors. (c) Any movement of patient during scanning leads to artifacts and loss of finer details in the images.

Now let us first consider the signal leakage phenomena when we apply denoising algorithm in post-processing

of acquired imagery. It occurs when a part of information contents of image is leaked as the noise part in the estimated denoised imagery. Therefore, the recorded pixel values (magnitude) get affected. Due to this, correlation exists between the learned noise and the corresponding estimated denoised image. We observe in the ablation studies that when only DRN part is used, some of the image information content is also learned as the noise due to overfitting. Also one has to use skip connections to match the dimension between the skip connection layer and previous layer to perform extra convolution or zero-padding. This leads to formation of blur patches in the resultant denoised image. In order to address these problems we augment the DRN with an efficient sparse representation method to learn the noise indirectly from the images, i.e., proposed DL part in our framework. As we are using patch-based dictionaries, during the dictionary update stage values are adjusted so that they do not overfit the data. On the other hand, when only DL part is used then problem like artifact formation called Gibbs (ringing) artifact in the high frequency regions is observed and minimax problem can occur if proper sparse thresholding is not applied. These problems are augmented by DRN part as residual learning (DRN) handles the high frequency regions efficiently through the convolution, batch normalisation and ReLU layers. In this way, proposed DL and DRN parts work in augmented manner, hence, possible loss of information in one part is augmented by another part. Thus our proposed approach enhances the medical image denoising process.

Next, type of scanning machines impacts the image acquisition process. Third generation or below machines typically generate the $2D$ scans that can later be converted into $3D$ data while fourth and fifth generation machines directly provide $3D$ data as the output. Hence, in proposed approach we have considered options of denoising process on both $2D$ and $3D$ medical data.

We observe that many researchers have explored the volumetric data procedure for denoising of the medical images. In block-matching and 3D filtering (BM3D) [58], the similar image patches are stacked to form $3D$ blocks and filtering is done on all the blocks. The inverse transform is then performed to get them back into $2D$ form. On the other hand, non-local means (NLM) [59] processes the voxels by $2D$ filtering with a search and a neighborhood window. This is used to find out the similarity between the pixels and a parameter to control the degree of smoothness in an image. Underlying assumption is noisy patches will find the similarity with other patches containing the noise. Therefore, as a side effect, the information present in the edges is lost and the edges become blur. In anisotropic diffusion filter (ADF) [60], voxels of the images are considered by combining domain and range Gaussian filters in order to find the geometric and photometric distances. The final estimated intensity value of a pixel is calculated by taking the average of the geometric and photometric distances among the pixels inside the selected spatial window. Hence, it enlarges the edge widths and makes them more blurry. Notably several recent approaches use $2D$ images for medical image denoising. Recently, RED-CNN [34] takes the $2D$ images as input and combines autoencoders and deconvolutional networks to preserve the image structures. More recently CNN-RL [35] accepts the images in $2D$ form and instead of learning the mapping function of an image, it predicts the latent clean image. In total variation (TV) [61], regularization is controlled in a way that more denoising process is applied in smooth regions and lesser at edge (discontinuity) regions of each image. Volumetric ($3D$) data is used in residual encoder-decoder Wasserstein generative adversarial network (RED-WGAN) [37] to denoise the images. They have used four convolutional and deconvolutional layers in generator and discriminator parts each along with visual geometry group (VGG) network to learn the features from images. However, boundaries in image are pixelated as entire $3D$ image cube is considered at once and processed. In our proposed approach, we consider one block of $3D$ data at a time and construct dictionaries of overlapping block patches as well as learn noise from residual learning using the $3D$ blocks for a better denoising process. Considering different generations of scanning machines, $2D$ image processing is also included. We have generalized the proposed approach to consider each slice (image) of the MRI, CT, and LDCT data independently for denoising process (Fig. 2). These images/blocks are given to the augmented framework in the form of patches where we directly and indirectly learn the noise characteristics and obtain the respective denoised images/blocks.

It is found that patch-based methods effectively smoothen the homogeneous regions as well as preserves the finer details in an image. Our proposed model also learns patch-based dictionaries for each image/block from the set of input images. In TV [61], the patches of an image are used in the edge detection scheme. When the TV norm of an image is too low it leads to over-smoothing and only edges are preserved if the norm is high. Therefore the approach [61] is sensitive to TV norm for denoising patches. The patch-based approach is also adopted in BM3D [58]. In K-SVD [22] a similar approach to the K-means method is adopted, however, a single dictionary is learned for entire image. The CPCE [38] network also uses patch-based method with transfer learning concept inside GAN framework to perform the denoising process on LDCT data. Firstly a $2D$ CPCE model is trained in GAN framework and then it is directly extended to $3D$ model so as to employ $3D$ spatial details using transfer learning concept and enhance the denoising process. However, noise is still present in the homogeneous regions. MAP-NN [39] is also used to denoise LDCT data in progressive manner using patch-based method. Here each module helps to denoise the image gradually though the quality of edges is compromised. In the proposed approach, we are learning patch-based dictionaries for overlapping image/block patches of a dataset. These dictionaries have knowledge of images as well as the residues learnt from the DRN part. Hence the critical information remains intact and edges are also preserved.

In our approach, we learned the noise characteristics from the image using the proposed dictionary-based DRN. In CNN-RL [35], the residue is learned from the deep CNN layers and it is multiplied by a constant factor to normalize the elements. This normalized residue is then subtracted from the given image to form a denoised image. It may be noted this way the denoised image may observe the loss of information at the edges due to a single scaling factor to normalize the residue. Unlike it, our proposed unsupervised method considers the average of residue learned from dictionary learning and DRN. It essentially avoids the heuristic of hard coding the normalization factor. In RED-CNN [34] the autoencoders and deconvolutional networks are used to preserve the structures in order to reduce noise from the images. This approach may over smoothed the edges as the data is compressed by the encoders and decoders. The PURE [42] method uses skip connections between the layers to form the residue and denoise the LDCT images. It uses modified U-Net network and full-dose CT data as ground truth images although noise is present in the images. AdaIN-based tunable CycleGAN [41] selects random patches to denoise LDCT images using a single switchable generator. The generator follows U-Net architecture and discriminator uses PatchGAN to estimate the noise pattern and obtain the denoised images. Note that the signal leakage is present in such network. Our proposed augmented approach uses the constraints to update the patch-based dictionaries and learn the residues that better handles the ill-posed nature of the problem.

V. EXPERIMENTS AND EVALUATION

In this section, we evaluate our proposed unsupervised learning approach by conducting experiments on different MRI, CT, and LDCT images. We begin by providing details of the datasets, machine specifications, parameter settings, and training. We then compare and analyze the results obtained by our approach with state-of-the-art approaches including signal leakage phenomena. We also discuss the results obtained by carrying out ablation studies in our proposed framework.

A. DATASETS

We use the synthetic MRI dataset available at https://syntheticmr.com by adding different levels of Rician noise in the ground truth MRI image. For synthetic CT dataset, we have used the Shepp-logan image available in MATLAB R2020a and then added different levels of Poisson noise to it. The real MRI and CT images are considered as available from the cancer imaging archive (TCIA) [46]. It is an open-access database of medical images available for the research. We have also conducted additional experiment on benchmark LDCT dataset from Mayo clinic [47]. It is an open access library consisting of full-dose and corresponding low-dose CT dataset. The library consists total of 10,112,591 scans that includes low-dose non-contrast CT scans of head, chest, and abdomen. The collection comprises of both full and corresponding simulated lower dose levels including 99 head scans, 100 chest scans, and 100 abdomen scans. In our experiment, data is in digital imaging and communications in medicine (DICOM) format.

B. MACHINE SPECIFICATION

All the algorithms are implemented in PARAM Shavak GPU-based supercomputer powered with two multicore CPUs, each with fourteen cores. It has NVIDIA GP100 accelerator card and 96 GB RAM. We have also used Intel Core i7-9750H CPU @ 2.60GHz with 20 GB RAM to add noise in the synthetic datasets for MRI and CT images. The programming is done using Python 3.7 and major libraries include matplotlib, skimage, numpy, scipy, pytorch, and pydicom.

C. SETUP AND PARAMETERS

In our experiment, **(a) for** $3D$ **denoising**: we have used 400 slices of $3D$ MRI and 350 slices of CT images having 256×256 voxels, each voxel is of resolution $1\ mm \times 1\ mm \times 1\ mm$. We have added different levels of noise in MRI and CT images. Then we test our model using 50 and 45 slices of $3D$ MRI and CT real datasets, respectively. As shown in Fig. 2 we select one block of $3D$ data at a time and decompose it into overlapping block patches. We have a total of 64 block patches each of dimension $32 \times 32 \times 8$ voxels within a block of $256 \times 256 \times 192$ voxels. Then each block patch is provided to the DL and DRN parts in order to generate the sparse vector and the $3D$ dictionary, respectively, to reconstruct the denoised patches. The DRN in RL consists of 10 layers where the first layer is a combination of $3D$ convolution and $3D$ ReLU. The $3D$ convolution consists of 84 filters of dimension $3 \times 3 \times 8$ followed by $max3D$ operation to introduce the non-linearity. Then $3D$ batch normalization is added in the next 10 layers to uplift the denoising performance by using higher learning rates. The final layer is the $3D$ convolution layer that gives the learned residue. This is used with the residue obtained from the DL part to make an average residue $\mathbf{R}_{avg}$. The averaged residue $\mathbf{R}_{avg}$ is again fed back to update the dictionary in DL part. Note that now the learned sparse dictionary can efficiently reduce Rician noise and Poisson noise from the MRI and CT images, respectively. **(b) for** $2D$ **denoising**: we have used 1500 slices of MRI and 1000 slices of CT images. Each image has a dimension of 512×512 pixels. Then we test our model using 442 and 250 image slices for real $2D$ datasets of MRI and CT images, respectively. We have a total of 84 patches each of dimension 64×64 pixels within an image of 512×512 pixels. Note that now the learned sparse dictionary can efficiently reduce Rician noise and Poisson noise from the MRI and CT images, respectively. **(c) for LDCT denoising**: we have used 1720 slices of LDCT images each having dimension of 512×512 pixels. We used 500 slices to test the model using the dictionary obtained. We divide each image into patches with dimension 64×64 pixels. These patches are given to the proposed framework and we obtain the respective denoised images.

The regularization parameters λ and μ are chosen after performing many trials on different noise levels on

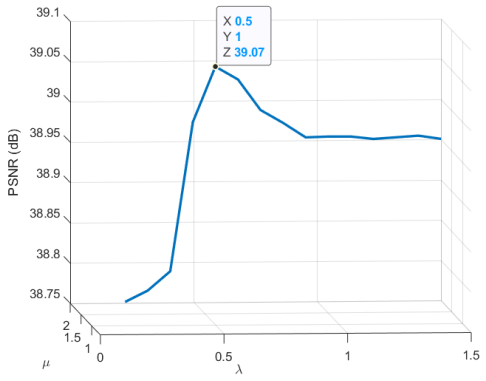

FIGURE 3. Sensitivity analysis of regularization parameters λ and μ (equation (7)).

3D and 2D images for MRI and CT including LDCT images. Fig. 3 shows the values of peak signal-to-noise ratio (PSNR) obtained at different range of λ and μ. As shown in the figure, they are fixed to $\lambda = 0.5$ and $\mu = 1$ as they show highest PSNR values and help avoiding minimax and overfitting problems. For implementing other comparative approaches, we use the optimal values of parameters as available in respective papers [22], [34], [35], [37]–[39], [41], [42], [58]–[61].

D. TRAINING

We have trained our augmented DL-DRN framework with synthetic 3D/2D MRI and CT images separately. First, we train our framework with volumetric (3D) slices of synthetic MRI data with different Rician noise levels to obtain the residue that in turn helps to get the learned dictionary. This dictionary is used on the test MRI dataset to estimate the denoised image. Similarly, we then use 2D MRI dataset to obtain the corresponding denoised image. We have then similarly trained and tested the noise learning framework for 3D/2D CT images with different levels of Poisson noise. We have also separately trained our framework with benchmark LDCT Mayo dataset. This dataset consists of chest scans at 10% of regular dose along with head and abdomen scans at 25% of normal dose. The slice thickness is 1.5 mm with image size 512 $\times$ 512 pixels.

E. RESULT ANALYSIS

We first display/discuss the visual results and then present the quantitative analysis with different performance metrics. Ablation studies have been carried out on the proposed framework to analyze the contribution of parts. Fig. 4 (a) is the given MRI image and Fig. 4 (b) shows the denoised image obtained by applying only the DRN part. One can observe in the zoomed portion that the edges are preserved but still noise is present in the image. When we apply only the DL part of the framework on the given MRI image, edges are not well preserved as shown in Fig. 4 (c). Also, it is observed that DL is able to reduce Gaussian and Poisson noise effectively as compared to Rician noise. Now by augmenting both DL and DRN parts (proposed framework), we obtain clear edges and a well-denoised image that can be witnessed in Fig. 4 (d). The major constraints of medical imaging are matched by the proposed augmented noise learning framework. Similar results are obtained during ablation studies on CT images as shown in Fig. 5. Fig. 5 (d) shows the denoised image obtained by the proposed augmented noise learning framework that has clear boundaries and reduced noise unlike Fig. 5 (b) and 5 (c).

We now compare the proposed approach with state-of-the-art methods. Fig. 6 shows the denoising results of synthetic 2D MRI data. We add 5% of Rician noise in the ground truth image (Fig. 6 (a)) and apply different algorithms to denoise it. Fig. 6 (b) shows that CNN-RL [35] maintains the outer details but inner details are pixelated. In Fig. 6 (c) RED-CNN algorithm [34] used to denoise the image also maintains the outer part of the edges whereas the inner details are still pixelated however it has improved results than CNN-RL image. In Fig. 6 (d) we see that K-SVD [22] successfully preserves the edges to some extent but inner details are not clear. In Fig. 6 (e) from the zoom portion, one can observe that BM3D [58] can maintain the outer edges but inner details of the image are appearing blur. Even the proposed approach result, Fig. 4 (f) obtained for 2D image is also preserving the edges. The quantitative evaluation is presented in Table 1 where one can observe calculated values of PSNR [62], structural similarity index measure (SSIM) [62], and root mean square error (RMSE) [63] of an estimated denoised image by implementing different algorithms at various noise levels in the imaging. It can be seen that the PSNR and SSIM of the denoised image estimated by 2D proposed approach is higher than other methods while the RMSE value is at a low when compared to other methods.

Further, the results obtained on synthetic 3D MRI data by adding 5% of Rician noise in the ground truth image (Fig. 7 (a)) are shown in Fig. 7. Fig. 7 (b) shows TV approach [61] is not able to preserve the edges of the images. From Fig. 7 (c) one can observe that NLM [59] is not able to maintain the sharpness in edges as they get distorted. Fig. 7 (d) shows that ADF [60] blurs the image and edge width is also increased. Fig. 7 (e) shows the denoised image obtained by RED-WGAN [37], zoom portion reveal the pixelated high frequency regions. In Fig. 7 (f) one can see that the image generated by the proposed unsupervised approach for 3D is nearly close to the ground truth image as well as it preserves both the inner and outer details of the image. Table 2 verifies the visual results by showing the quantitative results obtained by applying these methods on given images and shows that the proposed approach performs better than other methods.

Fig. 8 shows the visual results of the denoised image reconstructed from different denoising methods applied on synthetic 2D CT images obtained by adding 5% Poisson noise to the Shepp-logan dataset. Shepp-logan dataset is the synthetic dataset that is widely used for research purposes. In Fig. 8 (b) CNN-RL [35] maintains the structure of the image, however, due to the formation of grainy structures in the image edges are not distinct. Fig. 8 (c) shows that RED-CNN [34] reduces

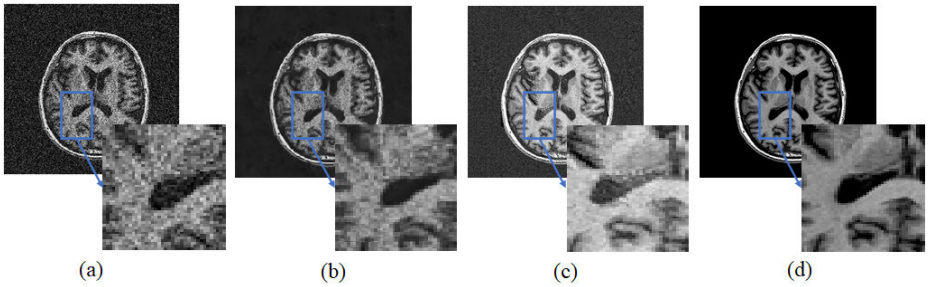

FIGURE 4. Ablation studies performed on 2*D* MRI brain scans [46]: (a) given image, (b) only DRN part in proposed approach, (c) only DL part in proposed approach, and (d) proposed augmented noise learning framework. It is evident that proposed augmented learning enhances the denoising performance.

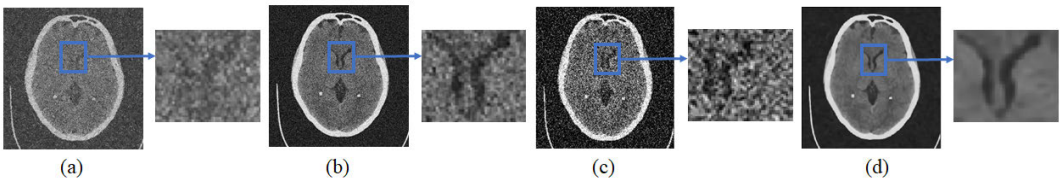

FIGURE 5. Ablation studies performed on 2*D* CT brain scans [46]: (a) given image, (b) only DRN part in proposed approach, (c) only DL part in proposed approach, and (d) proposed augmented noise learning framework. It is evident that proposed augmented learning enhances the denoising performance.

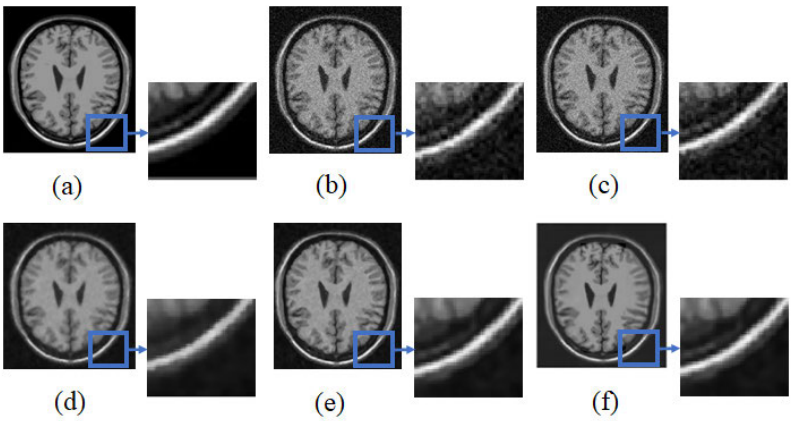

FIGURE 6. Denoising by different algorithms by adding 5% Rician noise in synthetic 2*D* MRI image: (a) Ground truth, (b) CNN-RL [35], (c) RED-CNN [34], (d) K-SVD [22], (e) BM3D [58], and (f) Proposed 2*D*.

TABLE 1. Average error scores by different approaches at different levels of Rician noises on synthetic 2*D* MRI images.

Measures	Ideal value	Rician noise levels	Methods				
			CNN-RL [35]	RED-CNN [34]	K-SVD [22]	BM3D [58]	Proposed 2D
PSNR (dB) [62]	High	5 %	33.685	35.663	33.252	36.213	**39.421**
		10 %	33.462	34.673	32.336	35.121	**37.071**
		15 %	32.747	31.982	31.032	28.164	**35.185**
SSIM [62]	1	5 %	0.801	0.824	0.809	0.832	**0.881**
		10 %	0.791	0.813	0.803	0.822	**0.869**
		15 %	0.724	0.762	0.736	0.792	**0.826**
RMSE [63]	0	5 %	19.033	17.261	18.904	18.072	**17.055**
		10 %	19.678	18.862	19.053	18.147	**17.119**
		15 %	20.024	19.381	21.165	20.012	**18.673**

the noise to some level, however, the structure is shifted a bit from its original position. Fig. 8 (d) shows that K-SVD [22] is not able to smooth the homogeneous regions that are making the image pixelated. Fig. 8 (e) indicates that overall the noise

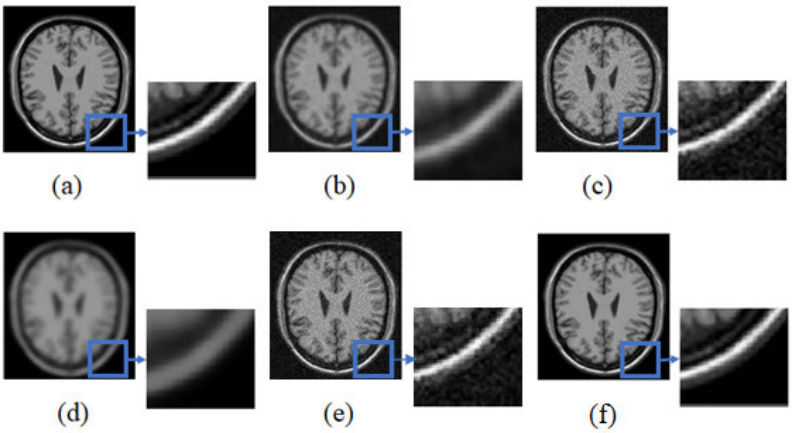

FIGURE 7. Denoising by different algorithms by adding 5% Rician noise in synthetic 3*D* MRI image: (a) Ground truth, (b) TV [61], (c) NLM [59], (d) ADF [60], (e) RED-WGAN [37], and (f) Proposed 3*D*.

TABLE 2. Average error scores by different approaches at different levels of Rician noises on synthetic 3*D* MRI images.

Measures	Ideal value	Rician noise levels	Methods				
			TV [61]	NLM [59]	ADF [60]	RED-WGAN [37]	Proposed 3D
PSNR (dB) [62]	High	5 %	28.928	29.166	28.461	38.772	**42.992**
		10 %	28.733	28.872	28.124	37.236	**41.870**
		15 %	28.181	28.729	28.022	35.524	**40.297**
SSIM [62]	1	5 %	0.697	0.721	0.708	0.781	**0.899**
		10 %	0.676	0.709	0.672	0.744	**0.893**
		15 %	0.670	0.683	0.618	0.712	**0.874**
RMSE [63]	0	5 %	24.117	22.061	23.073	19.329	**16.771**
		10 %	24.892	22.755	24.161	19.894	**16.892**
		15 %	26.883	24.819	24.442	20.112	**17.329**

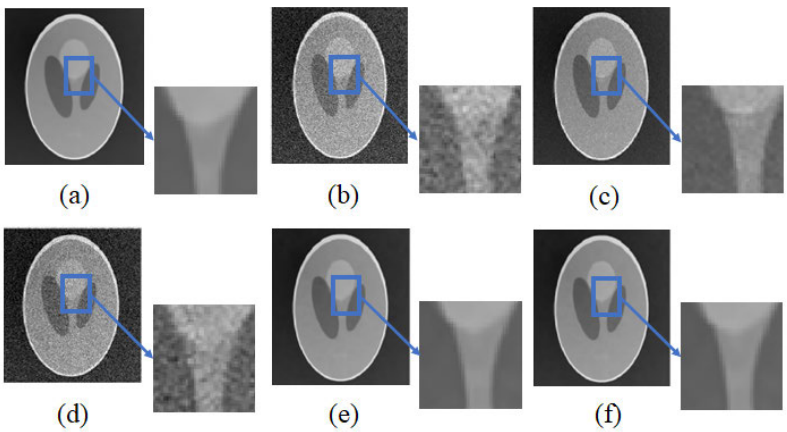

FIGURE 8. Denoising by different algorithms by adding 5% Poisson noise to the Shepp-logan 2*D* CT image: (a) Ground truth, (b) CNN-RL [35], (c) RED-CNN [34], (d) K-SVD [22], (e) BM3D [58], and (f) Proposed 2*D*.

is reduced using BM3D method [58] however finer details can be improved. The result obtained in Fig. 8 (f) from proposed approach for denoising image in 2*D* method also decreases the noise level to some extent. Table 3 shows the respective quantitative results of all the methods for different noise levels and it is visible that PSNR [62] and SSIM [62] of the proposed approach are higher than other approaches and RMSE [63] value of the proposed approach is at a low.

Fig. 9 shows the denoised images obtained on synthetic 3*D* CT images. In Fig. 9 (a) we observe that TV output [61] is blur and one cannot differentiate the boundaries. Fig. 9 (b) is the NLM [59] denoised output in which noise is visible.

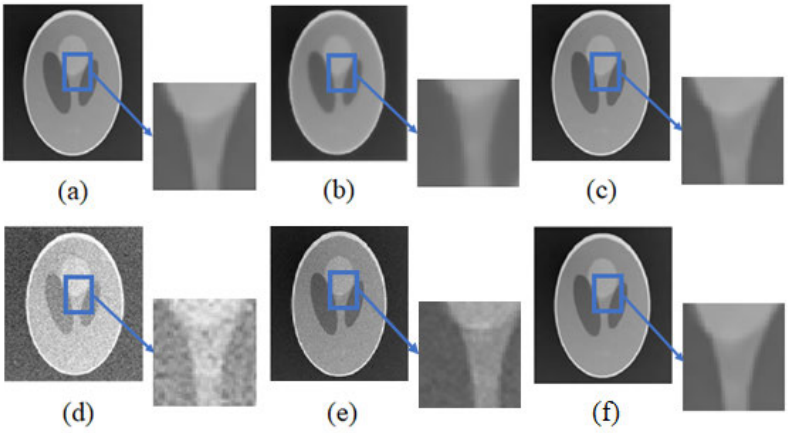

FIGURE 9. Denoising by different algorithms by adding 5% Poisson noise to the Shepp-logan 3*D* CT image: (a) Ground truth, (b) TV [61], (c) NLM [59], (d) ADF [60], (e) RED-WGAN [37], and (f) Proposed 3*D*.

TABLE 3. Average error scores by different approaches at different levels of Rician noises on synthetic 3*D* MRI images.

Measures	Ideal value	Rician noise levels	Methods				
			TV [61]	NLM [59]	ADF [60]	RED-WGAN [37]	Proposed 3D
PSNR (dB) [62]	High	5 %	28.928	29.166	28.461	38.772	**42.992**
		10 %	28.733	28.872	28.124	37.236	**41.870**
		15 %	28.181	28.729	28.022	35.524	**40.297**
SSIM [62]	1	5 %	0.697	0.721	0.708	0.781	**0.899**
		10 %	0.676	0.709	0.672	0.744	**0.893**
		15 %	0.670	0.683	0.618	0.712	**0.874**
RMSE [63]	0	5 %	24.117	22.061	23.073	19.329	**16.771**
		10 %	24.892	22.755	24.161	19.894	**16.892**
		15 %	26.883	24.819	24.442	20.112	**17.329**

TABLE 4. Average error scores by different approaches at different levels of Poisson noise on synthetic 3*D* CT images.

Measures	Ideal value	Poisson noise levels	Methods				
			TV [61]	NLM [59]	ADF [60]	RED-WGAN [37]	Proposed 3D
PSNR (dB) [62]	High	5 %	29.988	30.601	28.459	35.056	**43.129**
		10 %	29.843	30.212	28.336	34.962	**42.263**
		15 %	28.434	29.221	28.363	34.032	**40.663**
SSIM [62]	1	5 %	0.643	0.692	0.623	0.698	**0.802**
		10 %	0.691	0.689	0.622	0.665	**0.800**
		15 %	0.669	0.661	0.613	0.634	**0.797**
RMSE [63]	0	5 %	26.115	24.421	26.148	22.187	**17.948**
		10 %	26.862	25,124	26.924	23.371	**18.041**
		15 %	28.834	26.224	28.363	23.672	**18.685**

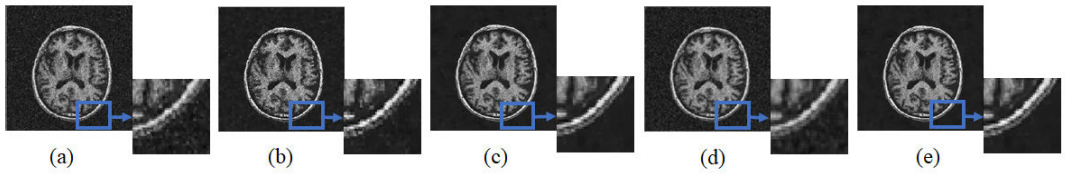

FIGURE 10. Denoising by different algorithms on real 2*D* MRI brain scans [46]: (a) CNN-RL [35], (b) RED-CNN [34], (c) K-SVD [22], (d) BM3D [58], and (e) Proposed 2*D*.

In Fig. 9 (c) we see that ADF [60] blurs the image that degrades the visual quality of the image. Fig. 9 (d) is the RED-WGAN [37] denoised image in which the edges are a bit blur. Fig. 9 (e) shows the result obtained from the proposed unsupervised learning approach in 3*D* way and one can observe that the reconstructed image is close to the

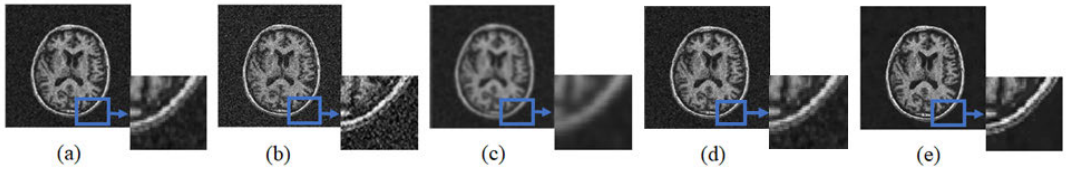

FIGURE 11. Denoising by different algorithms on real 3*D* MRI brain scans [46]: (a) TV [61], (b) NLM [59], (c) ADF [60], (d) RED-WGAN [37], and (e) Proposed 3*D*.

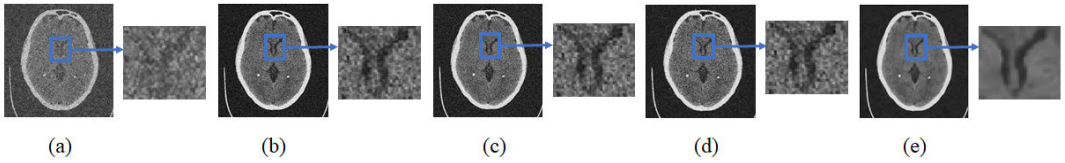

FIGURE 12. Denoising by different algorithms on real 2*D* CT brain scans [46]: (a) CNN-RL [35], (b) RED-CNN [34], (c) K-SVD [22], (d) BM3D [58], and (e) Proposed 2*D*.

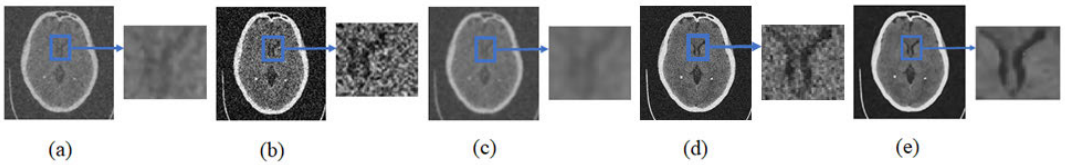

FIGURE 13. Denoising by different algorithms on real 3*D* CT brain scans [46]: (a) TV [61], (b) NLM [59], (c) ADF [60], (d) RED-WGAN [37], and (e) Proposed 3*D*.

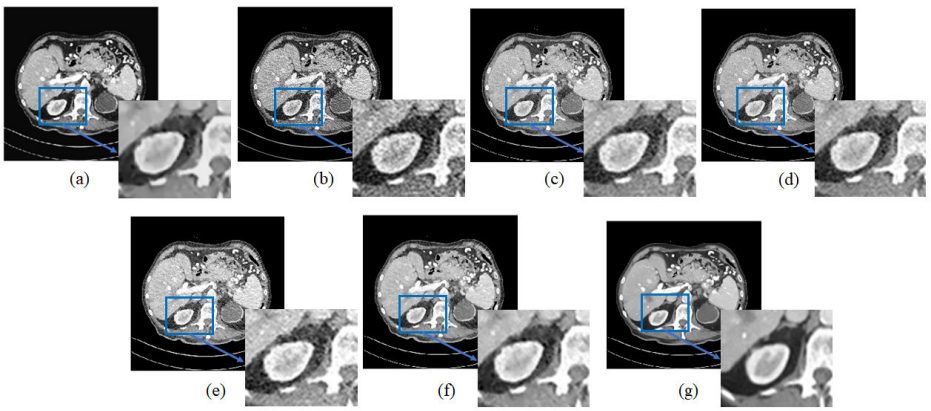

FIGURE 14. Denoising by different algorithms on the benchmark low-dose CT (Mayo clinic) scans of abdomen on 32-year male with id L033 [47]: (a) Normal dose image, (b) Corresponding LDCT image, (c) PURE [42], (d) MAP-NN [39], (e) CPCE [38], (f) AdaIN CycleGAN [41], and (g) Proposed.

ground truth image (Fig. 9 (a)) and also the Poisson noise is reduced to a great extent. Respective average error scores are shown in Table 4.

Fig. 10 shows the qualitative results on the real 2*D* MRI dataset [46]. In Fig. 10 (a) one can observe that the result of CNN-RL [35] is pixelated. Fig. 10 (b) is the result of RED-CNN method [34] and here too the edges are not clear and pixelated. Fig. 10 (c) shows the K-SVD output [22] and again the finer details are missing. Fig. 10 (d) is the result obtained by BM3D [58] method and one can notice that outer edges are sharp, however, inner details are blur. In Fig. 10 (e) one can see that the proposed approach has reduced noise to a greater extent.

Qualitative results can be seen in Fig. 11 on the real 3*D* MRI dataset [46]. In Fig. 11 (a), TV [61] is used to denoise the image, and one may see that the edges are not as clear that degrades the visual quality of image. Fig. 11 (b) shows the result obtained by NLM method [59] and one can see that edges are not clear and noise is still present in the image that decreases the visual quality of image. In Fig. 11 (c) one can

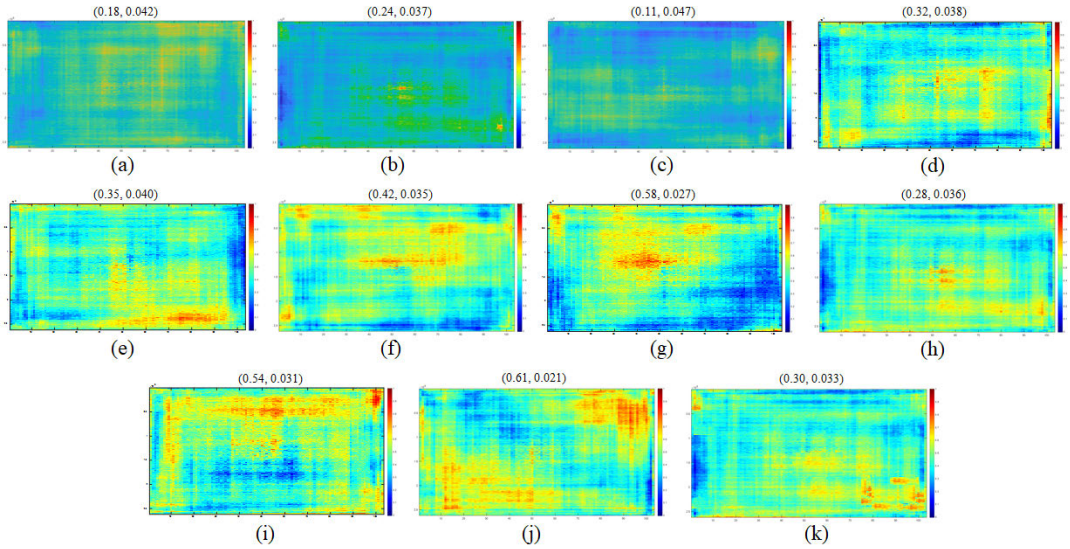

FIGURE 15. Signal leakage phenomena after denoising MRI and CT datasets: local cross-correlation indices computed between estimated noise and corresponding denoised data cube over randomly selected 1000 images using different approaches including ablation result on proposed approach. The brackets on the top indicate (mean, standard deviation) values of each map. (a) only DRN part in proposed approach, (b) only DL part in proposed approach, (c) Proposed augmented noise learning framework, (d) CNN-RL [35], (e) RED-CNN [34], (f) K-SVD [22], (g) TV [61], (h) BM3D [58], (i) NLM [59], (j) ADF [60], and (k) RED-WGAN [37]. It is evident that proposed augmented noise learning process (c) shows minimal signal leakage by the least correlation values and shortest range of local cross-correlation indices, i.e., enhances the medical image denoising process in MRI and CT datasets.

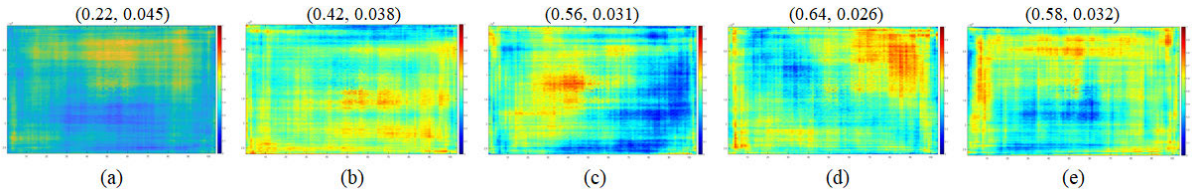

FIGURE 16. Signal leakage phenomena after denoising the benchmark Mayo low-dose CT datasets: local cross-correlation indices computed between estimated noise and corresponding denoised data cube over randomly selected 1000 images after implementing different algorithms. The brackets on the top indicate (mean, standard deviation) values of each map. (a) Proposed augmented noise learning framework, (b) PURE [42], (c) MAP-NN [39], (d) CPCE [38], and (e) AdaIN CycleGAN [41]. It is evident that proposed augmented noise learning process (a) shows minimal signal leakage by the least correlation values and shortest range of local cross-correlation indices, i.e., enhances the medical image denoising process on low-dose CT datasets.

observe that the entire image generated by ADF method [60] is blur and visual quality is also poor. In Fig. 11 (d) one can see that the denoised image has unclear boundary regions using RED-WGAN [37] approach. Fig. 11 (e) shows the result obtained from the proposed unsupervised 3*D* learning approach and it is visible that our model can reduce noise to a much lower level as compared to other approaches.

Fig. 12 shows the results obtained by applying the denoising methods on real 2*D* CT datasets [46]. In Fig. 12 (a) we observe that CNN-RL [35] is not able to preserve the edges due to which one cannot differentiate between the boundaries. Fig. 12 (b) shows that RED-CNN [34] maintains the edges however noise content is still present in the image. The output in Fig. 12 (c) shows that the structure is maintained by K-SVD [22] however details in the image are lost. In Fig. 12 (d) it is clear that BM3D [58] is able to preserve the edges however noise is still present in homogeneous regions of the image. Fig. 12 (e) shows the denoised image obtained from proposed approach and one can observe from the zoom portion that boundaries are preserved along with the reduction in noise.

Qualitative results obtained by applying different denoising algorithms on real 3*D* CT data are shown in Fig. 13. Fig. 13 (a) shows that the image generated by applying TV [61] method on the input image is not able to preserve the structure in the image. In Fig. 13 (b) one can see that the image obtained by NLM [59] still is very noisy. Fig. 13 (c) shows that ADF [60] makes the image blur thus any part of the image is not clearly visible. The denoised image obtained by using RED-WGAN [37] approach is shown in Fig. 13 (d) and one can see the pixelated regions in the zoom portion along side the edges. In Fig. 13 (e) one can observe that proposed unsupervised learning approach performs better than other existing seven state-of-the-art approaches by preserving the edges and maintaining the visual quality.

We have also obtained the qualitative results by applying different denoising algorithms on the benchmark Mayo LDCT datasets as shown in Fig. 14. Fig. 14 (a) shows the normal dose CT image and Fig. 14 (b) shows the corresponding low-dose CT image of abdomen. In Fig. 14 (c) one can observe that the image obtained by PURE [42] still has a bit of noise present. Fig. 14 (d) shows the image generated by MAP-NN [39] also has some grainy patches present. In Fig. 14 (e) we can see that noise is not reduced properly using CPCE [38] network. In the zoom portion of Fig. 14 (f) one can see that a tinch of noise is still present in the image generated by AdaIN CycleGAN [41]. Fig. 14 (g) shows that the denoised image obtained by our proposed approach has minimal noise and maintains the visual quality by preserving the edges.

Finally, we provide the signal (image) leakage analysis results in Fig. 15 and 16. Many times the model learns some of the image information content as noise (overfitting) and consequently correlation exists between the estimated noise and corresponding denoised image. We calculate such correlations for different denoising algorithms and compare with proposed approach including ablation study on our framework. In ideal scenarios, the local cross-correlation between noise and denoised image should be zero (0). Referring to Fig. 15, the blue portions indicate low correlation values and the red portions indicate high correlation values (more signal leakage). Note that these maps are obtained after running the codes on 1000 images. Fig. 15 (a) shows the correlation map when only DRN part is applied to the given images. The correlation values are in between 0 and 0.61 with mean and standard deviation as 0.18 and 0.042, respectively. When only the DL part is applied, the range of correlation values is higher, i.e., 0 to 0.92 with $(0.24, 0.037)$ as shown in Fig. 15 (b). The result obtained by applying the proposed augmented DL/DRN approach is shown in Fig. 15 (c). See that the range of correlation values are now drastically reduced to 0 to 0.53 $(0.11, 0.047)$ with most of the values close to zero. We also tested the performance of other approaches for signal leakage as shown in Fig. 15 (d) to 15 (k). Note that the approaches that do not directly estimate noise, we calculate it by subtracting the estimated denoised image from the available data. From all the correlation values shown in Fig. 15, one can observe that very little correlation exists when applying the proposed augmented unsupervised noise learning framework (Fig. 15 (c)) on the images. We also provide the signal leakage results for LDCT dataset in Fig. 16. One can observe from Fig. 16 (a) that correlation map using the proposed approach exhibits minimal correlation with mean 0.22 and standard deviation 0.045. We have also shown the signal leakage map of other approaches in Fig. 16 (b) to 16 (e) and they show high signal leakage that is evident from higher correlation values in the maps when compared to the proposed approach. Hence from the correlation maps one can see that proposed approach effectively enhances the denoising process with a minimal signal leakage.

VI. CONCLUSION AND FUTURE WORK

We have presented a novel augmented unsupervised noise learning approach for enhancing medical image denoising considering input as $2D$ and $3D$ for image/voxel processing. The proposed dictionary based DRN handles both the Rician noise and Poisson noise present in the MRI and CT/LDCT images, respectively. Our model learns the patch-based dictionaries in order to learn noise indirectly and augment with the residue (noise) contents learn directly from the available MRI/CT/LDCT images using proposed DRN. Note that the proposed approach does not require the clean (denoised) images for training the model, unlike many deep learning-based recent approaches. We have better handled the ill-posed nature of the problem by choosing the optimum regularization parameters that have been estimated from the data. Dictionary-based DRN reduces the noise from the images by preserving the edges of the images and maintaining their visual quality (without losing details) which is evident from the results. Ablation study conducted on DL and DRN parts further evaluate the efficacy of the proposed augmented noise learning process. We further showed that the proposed method has minimal signal leakage with least Gibbs (ringing) artifacts in the estimated denoised image, and enhances the medical image denoising. In future, one would like to design unsupervised framework for restoration of medical images that would address any degradation in the multimodal images along with the noise.

ACKNOWLEDGMENT

The authors would like to thank the editor, the associate editor, and anonymous reviewers for their insightful and helpful comments. They are thankful to Gujarat Council on Science and Technology (GUJCOST) for providing PARAM Shavak GPU-based supercomputer to their institute and allow them to conduct the exhaustive experiments in this research work. They also thank to a team of qualified technicians, including Vikas Patel, Gujarat, India, for providing practical information on medical imaging techniques and to Dr. Shrishail Gajbhar for helpful discussions.

REFERENCES

[1] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.
[2] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 755–768, Feb. 2014.
[3] K. Ravani, S. Saboo, and J. S. Bhatt, "A practical approach for SAR image despeckling using deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 2957–2960.
[4] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia, PA, USA: SIAM, 1998.
[5] T. Tirer and R. Giryes, "Back-projection based fidelity term for ill-posed linear inverse problems," *IEEE Trans. Image Process.*, vol. 29, pp. 6164–6179, Dec. 2020.
[6] E. López-Rubio and M. N. Florentín-Núñez, "Kernel regression based feature extraction for 3D MR image denoising," *Med. Image Anal.*, vol. 15, no. 4, pp. 498–513, Aug. 2011.
[7] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[8] P. Shende, M. Pawar, and S. Kakde, "A brief review on: MRI images reconstruction using GAN," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2019, pp. 139–142.
[9] D. Karimi, H. Dou, S. K. Warfield, and A. Gholipour, "Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101759.
[10] K. Choi, J. S. Lim, and S. K. Kim, "StatNet: Statistical image restoration for low-dose CT using deep learning," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1137–1150, Oct. 2020.
[11] M. Sharma, J. S. Bhatt, and M. V. Joshi, "Early detection of lung cancer from CT images: Nodule segmentation and classification using deep learning," in *Proc. 10th Int. Conf. Mach. Vis. (ICMV)*, vol. 10696, Apr. 2018, Art. no. 106960.
[12] V. S. Deshpande and J. S. Bhatt, "Bayesian deep learning for deformable medical image registration," in *Proc. Int. Conf. Pattern Recognit. Mach. Intell.* Cham, Switzerland: Springer, 2019, pp. 41–49.
[13] R. Anvari, A. R. Kahoo, M. Mohammadi, N. A. Khan, and Y. Chen, "Seismic random noise attenuation using sparse low-rank estimation of the signal in the time–frequency domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 5, pp. 1612–1618, May 2019.
[14] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles, "New methods for MRI denoising based on sparseness and self-similarity," *Med. Image Anal.*, vol. 16, no. 1, pp. 18–27, Jan. 2012.
[15] J. Mohan, V. Krishnaveni, and Y. Guo, "A survey on the magnetic resonance image denoising methods," *Biomed. Signal Process. Control*, vol. 9, pp. 56–69, Jan. 2014.
[16] P. Kaur, G. Singh, and P. Kaur, "A review of denoising medical images using machine learning approaches," *Current Med. Imag. Rev.*, vol. 14, no. 5, pp. 675–685, Sep. 2018.
[17] M. Diwakar and M. Kumar, "A review on CT image noise and its denoising," *Biomed. Signal Process. Control*, vol. 42, pp. 73–88, Apr. 2018.
[18] D. Thanh, P. Surya, and L. M. Hieu, "A review on CT and X-ray images denoising methods," *Informatica*, vol. 43, no. 2, pp. 505–515, Jun. 2019.
[19] P. Liu, M. D. El Basha, Y. Li, Y. Xiao, P. C. Sanelli, and R. Fang, "Deep evolutionary networks with expedited genetic algorithms for medical image denoising," *Med. Image Anal.*, vol. 54, pp. 306–315, May 2019.
[20] B. Wen, S. Ravishankar, L. Pfister, and Y. Bresler, "Transform learning for magnetic resonance image reconstruction: From model-based learning to building neural networks," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 41–53, Jan. 2020.
[21] L. Zhou, J. D. Schaefferkoetter, I. W. K. Tham, G. Huang, and J. Yan, "Supervised learning with cyclegan for low-dose FDG PET image denoising," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101770.
[22] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[23] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.
[24] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, May 2011.
[25] Y. Huang, J. Paisley, Q. Lin, X. Ding, X. Fu, and X.-P. Zhang, "Bayesian nonparametric dictionary learning for compressed sensing MRI," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5007–5019, Dec. 2014.
[26] J. V. Manjón and P. Coupe, "MRI denoising using deep learning," in *Patch-Based Techniques in Medical Imaging*. Cham, Switzerland: Springer, 2018, pp. 12–19.
[27] T. Higaki, Y. Nakamura, F. Tatsugami, T. Nakaura, and K. Awai, "Improvement of image quality at CT and MRI using deep learning," *Jpn. J. Radiol.*, vol. 37, no. 1, pp. 73–80, Jan. 2019.
[28] A. Arun, T. J. Thomas, J. S. Rani, and R. S. S. Gorthi, "Efficient directionality-driven dictionary learning for compressive sensing magnetic resonance imaging reconstruction," *J. Med. Imag.*, vol. 7, no. 1, 2020, Art. no. 014002.
[29] D. Du, Z. Pan, P. Zhang, Y. Li, and W. Ku, "Compressive sensing image recovery using dictionary learning and shape-adaptive DCT thresholding," *Magn. Reson. Imag.*, vol. 55, pp. 60–71, Jan. 2019.
[30] A. Jung, Y. C. Eldar, and N. Görtz, "On the minimax risk of dictionary learning," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1501–1515, Mar. 2015.
[31] S. Singh, B. Póczos, and J. Ma, "Minimax reconstruction risk of convolutional sparse dictionary learning," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2018, pp. 1327–1336.
[32] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network," *Biomed. Opt. Exp.*, vol. 8, no. 2, pp. 679–694, 2017.
[33] V. Cheplygina, M. de Bruijne, and J. P. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Med. Image Anal.*, vol. 54, pp. 280–296, May 2019.
[34] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Image Process.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
[35] W. Jifara, F. Jiang, S. Rho, M. Cheng, and S. Liu, "Medical image denoising using convolutional neural network: A residual learning approach," *J. Supercomput.*, vol. 75, no. 2, pp. 704–718, Feb. 2019.
[36] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101552.
[37] M. Ran, J. Hu, Y. Chen, H. Chen, H. Sun, J. Zhou, and Y. Zhang, "Denoising of 3D magnetic resonance images using a residual encoder–decoder Wasserstein generative adversarial network," *Med. Image Anal.*, vol. 55, pp. 165–180, Jul. 2019.
[38] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.
[39] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Mach. Intell.*, vol. 1, no. 6, pp. 269–276, 2019.
[40] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.
[41] J. Gu and J. C. Ye, "AdaIN-based tunable CycleGAN for efficient unsupervised low-dose CT denoising," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 73–85, 2021.
[42] K. Kim, S. Soltanayev, and S. Y. Chun, "Unsupervised training of denoisers for low-dose CT reconstruction without full-dose ground truth," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1112–1125, Oct. 2020.
[43] D. Lee, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing MRI," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 15–18.
[44] M. S. Ebrahimi and H. K. Abadi, "Study of residual networks for image recognition," 2018, *arXiv:1805.00325*. [Online]. Available: http://arxiv.org/abs/1805.00325
[45] F. Liu, X. Ren, Z. Zhang, X. Sun, and Y. Zou, "Rethinking skip connection with layer normalization," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 3586–3598.
[46] ACRIN. (2020). *The Site is Funded by the National Cancer Institute's (NCI) Cancer Imaging Program, and the Contract is Operated by the University of Arkansas for Medical Sciences*. [Online]. Available: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=50135264#1619a826e02b4ce8b09f6214aca48665
[47] C. McCollough, B. Chen, D. Holmes, X. Duan, Z. Yu, L. Yu, S. Leng, and J. Fletcher, "Data from low dose CT image and projection data [data set]. The cancer imaging archive," Cancer Imag. Arch., Nat. Med. Imag. Arch., Tech. Rep., 2020, doi: 10.7937/9npb-2637.
[48] P. Gravel, G. Beaudoin, and J. A. De Guise, "A method for modeling noise in medical images," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1221–1232, Oct. 2004.
[49] L. He and I. R. Greenshields, "A nonlocal maximum likelihood estimation method for Rician noise reduction in MR images," *IEEE Trans. Med. Imag.*, vol. 28, no. 2, pp. 165–172, Feb. 2009.
[50] S. Aja-Fernández, C. Alberola-López, and C.-F. Westin, "Noise and signal estimation in magnitude MRI and Rician distributed images: A LMMSE approach," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1383–1398, Aug. 2008.
[51] N. D. Mascarenhas, S. S. Furuie, and A. L. S. Portal, "Global projection estimation methods for the tomographic reconstruction of images with Poisson noise," *IEEE Trans. Nucl. Sci.*, vol. 40, no. 6, pp. 2008–2013, Dec. 1993.
[52] L. Ma, L. Moisan, J. Yu, and T. Zeng, "A dictionary learning approach for Poisson image deblurring," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1277–1289, Jul. 2013.

[53] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[54] G. A. Wright, "Magnetic resonance imaging," *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 56–66, Jan. 1997.
[55] D. J. Goodenough and K. E. Weaver, "Overview of computed tomography," *IEEE Trans. Nucl. Sci.*, vol. NS-26, no. 1, pp. 1661–1667, Feb. 1979.
[56] G. J. Hoiting, "Measuring MRI noise," Ph.D. dissertation, Univ. Groningen, Groningen, The Netherlands, 2005.
[57] F. E. Boas and D. Fleischmann, "CT artifacts: Causes and reduction techniques," *Imag. Med.*, vol. 4, no. 2, pp. 229–240, 2012.
[58] 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, Aug. 2007.
[59] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles, "MRI denoising using non-local means," *Med. Image Anal.*, vol. 12, no. 4, pp. 514–523, 2008.
[60] K. Krissian and S. Aja-Fernandez, "Noise-driven anisotropic diffusion filtering of MRI," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2265–2274, Oct. 2009.
[61] A. Ben Said, R. Hadjidj, and S. Foufou, "Total variation for image denoising based on a novel smart edge detector: An application to medical images," *J. Math. Imag. Vis.*, vol. 61, no. 1, pp. 106–121, Jan. 2019.
[62] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[63] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study," *J. Comput. Commun.*, vol. 7, no. vol. 3, pp. 8–18. 2019.

SWATI RAI received the B.E. degree in electronics and telecommunications engineering from BVCOEW, Pune University, India, in 2015, and the M.Tech. degree in electronics and telecommunications from SIT, Symbiosis International University, Pune, India, in 2017. She is currently a full-time Ph.D. Research Scholar with the Indian Institute of Information Technology Vadodara (IIITV), Gandhinagar, India. Her research interests include signal processing, deep learning, and wireless communication.

JIGNESH S. BHATT received the B.E. degree in electronics and communications engineering from Hemchandracharya North Gujarat University, India, the M.Tech. degree in communication systems from SVNIT, Surat, India, and the Ph.D. degree in information and communication technology from DA-IICT, Gandhinagar, India.

He has been working as an Assistant Professor with the Indian Institute of Information Technology Vadodara (IIITV), India, since 2014. He has successfully developed image registration software for meteorological payloads at the Department of Space, ISRO, India. He has coauthored a book *Regularization in Hyperspectral Unmixing* (SPIE Press, USA). His research interests include remote sensing, signal processing, computer vision, deep learning, and inverse ill-posed problems, especially in hyperspectral imagery and medical data.

S. K. PATRA (Senior Member, IEEE) received the B.Sc. (Eng.) degree in electronics and telecommunication engineering from UCE Burla, in 1986, the M.Sc. (Eng.) degree in electronics system and communication specialization from NIT Rourkela (formerly known as REC Rourkela), and the Ph.D. degree from The University of Edinburgh, U.K., in 1998.

He has served the Defence Research and Development Organization (DRDO), for a short period before moving to the National Institute of Technology, Rourkela, in 1989. He is currently serving as the Director for the Indian Institute of Information Technology Vadodara, Gujarat, India. He has supervised 16 Ph.D. students and published over 60 articles in refereed journals besides more than 150 conference papers. His research interests include adaptive signal processing, fuzzy systems, wireless communication, and machine learning for wireless communication networks and related fields. He was a recipient of the prestigious Common-Wealth Scholarship.