DVAE: Deep Variational Auto-Encoders for Denoising Retinal Fundus Image

Biswajit Biswas, Swarup Kr Ghosh and Anupam Ghosh

Abstract Deep learning performs as a computational tool with various potential utilities in ophthalmology. Retinal infections of the eye need to analyze small retinal vessels, microaneurysms, and exudates in the diagnosis of retinal diseases. Due to the appearance of various noises in the fundus images, the retinal vasculature is too complicated to be analyzed for retinal conditions. In this work, we have focused on the field of advanced deep learning in which plethora of architecture is available with the increase in dimension and flexibility of the retinal fundus images. Removal of noise is an essential part to better visibility of noisy fundus and thus a deep learning method for degraded retinal fundus image restoration scheme has been suggested in this investigation. A deep convolutional denoising auto-encoder method based on total variational multi-norm loss function minimization with batch normalization approach has been introduced for restoration of the fundus. The proposed scheme is utilized to restore the perceptible structural details of fundus as well as to decrease the noise level. Moreover, the speed of the network for target noisy images is faster compared to that of other models after fine-tuning of the network with the dropout mechanism. The retinal image databases such as DRIVE, STARE, and DIARETDB1 have been adopted to assess image denoising effects. Our approach to increase the visibility of fundus images by noise reduction through a deep training method has significantly delivered better performance without losing image details along with having fast convergence rate.

B. Biswas
Department of Computer Science & Engineering,
University of Calcutta, Kolkata 700098, India
e-mail: biswajit.cu.08@gmail.com

S. K. Ghosh (✉)
Department of Computer Science & Engineering,
Maulana Abul Kalam Azad University of Technology, Kolkata 700064, India
e-mail: swarupg1@gmail.com

A. Ghosh
Department of Computer Science & Engineering,
Netaji Subhash Engineering College, Kolkata 700152, India
e-mail: anupam.ghosh@rediffmail.com

S. Bhattacharyya et al. (eds.), *Hybrid Machine Intelligence for Medical Image Analysis*, Studies in Computational Intelligence 841,
https://doi.org/10.1007/978-981-13-8930-6_10

Keywords Deep learning · Image denoising · Convolutional auto-encoder · Variational method · Retinal image

1 Introduction

Nowadays, deep learning has become a remarkably powerful tool for image processing and application [1]. It has considerable application in classification and prediction problem since the year 2012. Medical imaging is a challenging task due to illumination property of medical image such as brain MRI, chest X-ray, retinal image, lung MRI, and liver MRI. Medical imaging has two parts in general: first one is disease classification or anomaly detection from different medical images, and secondly, image pre-processing step. Image pre-processing is a crucial part in medical imaging since initial prediction is more important to the respective doctor for an initial assessment before going for automated diagnosis [1, 2]. There are several image pre-processing techniques such as image restoration, image enhancement, and denoising of image which increase the clarity of the image as well as give the details of the part of image (edge, nodes, vessels, etc.) [3, 4]. Before post-processing of images to detect some anomaly, pre-processing is equally important for further processing. Lots of image restoration, image enhancement, and image denoising methods have been developed by several researchers [5–7]. The concept of deep learning in image processing increases the accuracy as well as the robustness and hence we proposed a deep-learning-based image denoising method which is applied on retinal fundus image [8].

In this chapter, we have suggested a denoising model based on deep variational auto-encoders for denoising the retinal image. Lots of automatic anomaly detection algorithm X-ray images or MR images have been proposed since the last few decades. Diagnosis of retinal image is equally important to detect different diseases such as diabetic retinopathy, cataract, and glaucoma earlier to prevent blindness [9]. According to several ophthalmologist and eye experts, diabetes is predictable and considered to be one of the primary reasons of blindness, which is known as diabetic retinopathy (DR) [10]. Early diagnosis is very important for proper treatment because it prevents the visual impairment caused by diabetes. For diagnosis, DR causes damage to small blood vessels occurring in the retina [11]. Images of a patient with DR exhibit red and yellow spots indicative of hemorrhages and exudates. There are three types of DR: background diabetic retinopathy (BDR), proliferate diabetic retinopathy (PDR) and severe diabetic retinopathy (SDR). In BDR stage, arteries in the retina become dilated and show leakage, forming small, dot-like hemorrhages. In PDR phase, circulation problems cause areas of the retina to become oxygen-deprived or ischemic and new fragile vessels to develop to sustain the oxygen levels within the retina. It can cause blood leakage in the retina and vitreous, spots or floaters, along with decreased vision. In the SDR stage, more blood vessels are blocked, blocking blood supply to several areas of the retina along with continued abnormal vessel growth and scar tissue. Hard exudates is one of the primary characteristics of DR which

belongs to non-proliferate diabetic retinopathy (NPDR) and the size can vary from tiny specks to large patches with sharp edges [9, 10].

Recently, deep learning has became a very popular tool for feature extraction and classification on large datasets since deep learning in contrast urges to solve the problem end-to-end and creation of abstract features is a nice property of it. Deep learning takes a crucial part in medical imaging such as image segmentation, image registration, image restoration, or different disease diagnosis for the large volume of data. The architecture of ANNs consists of neurons with activation functions and adjusted weight, and it is trained as supervised approach [12]. To adjust the weight in ANNs, back propagation neural network based on gradient decent algorithm takes a significant role in computational intelligence since last few decades. The main drawback of gradient decent is that it has an over fitting problem with the increase in the size of data since the algorithm is often trapped by its local minima. The deep neural network is popular in recent research since it overcomes the overfitting problem for large data as well as it can assign non-random initial value in pre-procedure training phase. In 2009, ImageNet's competition has drawn attention to many computer vision researchers all over the world and hence journey of deep learning has been started. There are four important deep learning architecture such as restricted Boltzmann machines (RBMs), deep belief networks (DBNs), convolutional neural networks (CNNs), and auto-encoder (AE) [13].

There is certain loss of information within the image during the capture of the image because of some limitation of X-ray hardware and therefore the image contains noise. Many vital structures do not seem to be perceived properly since medical images are poorly lit up. Lots of uncertainties are present within the image in imprecise form that leads to inaccurate diagnosis. Thus, it is essential to remove noise from the image or restore the image. The objective of denoising of an image is similar to restoring an image to another form that is appropriate for further processing [14]. Image denoising plays a crucial role in the image pre-processing stage in digital imaging applications as it restores the images from corrupted images [15]. The objective of image denoising is to restore the maximum feature details of the original image through the process of elimination of undesired noise. Various denoising methods have been suggested by several researchers, including neighborhood filtering, median filtering, total variation minimization (TV), and non-local mean (NLM) denoising algorithms [16, 17]. Conventional image denoising algorithms, such as Gaussian filter and median filters are satisfactory for pulling out image noise, still some blurred edge and texture regions are present in the image [15, 17]. Recently, deep-learning-based image denoising scheme has presented outstanding performance compared to other traditional image denoising methods. Zhang et al. (2014) proposed an adaptive bilateral-filter-based image denoising framework in which entropy-based feature extraction is used to restore medical image [5]. It is the most fashionable image domain-based denoising strategy which are used bilateral filter recently. A novel artifact suppressed dictionary learning scheme was suggested by Chen et al. (2015) for low-dose CT image processing [14]. A medical image denoising method using convolutional denoising auto-encoders presented by Gondara et al.[18]. All the aforementioned scheme have suffered from bias and variance issues as well as have

long training time. In this investigation, we represent an image denoising strategy in the light of a deep variational denoising auto-encoder for retinal images called as DVAE and assess in applications in ophthalmology by looking at existing image denoising algorithms [19–21].

An auto-encoder (AE) is an alternative category of feed-forward neural network where the number of input layer and output layer are indistinguishable, called an autoassociator, that is designed for unsupervised learning. This can be one form of multilayer perceptron (MLP) in auto-association fashion that can be used for dimension reduction and data compression in an information system. Basically, AE is prepared to encode the input data and turn it into an abstract illustration by encoder function so that the input vectors are reorganized in that representation. Fundamentally, it tries to approximate the identity mapping in this procedure. Vincent et al. (2011) described the denoising auto-encoder (DAE) which deliberately adds noise into the training data and trains the AE with these adulterated information [22]. DAE can recover the original data and make it noise-free through the training process leading to a superior robustness. The main drawback of DAE is that we can be injected noise at input level and training criteria is intractable. In auto-encoder, the latent space where the encoded vectors lie may be discrete or may allow easy interpolation [20]. In this investigation, the design deep variational auto-encoder (DVAE) which is tractable with lower bound has been applied on the retinal image for removal of noise [19–21].

In this work, we have considered a deep denoising auto-encoder (DA) that supported variational technique to restore noise-free retinal image. First, the auto-encoder model maps the features vector into latent plane through its variational distribution function that reduces the dimension of features vector. For more noisy data, we have measured multiple-loss function to realize quick optimization in convolutional model throughout train periods. We get an inclination to verify overall performance of the proposed methodology and compare with different three state-of-the-art methods using a public standard benchmark database in an enormous satisfaction in training session. The experiential results determine that the planned methodology not only restores retinal images but additionally preserves image details a lot accurately and minimize number of epoch than most privileged technique. The primary contributions of this article are as follows:

- We analyze the deep denoising auto-encoder algorithms and propose a deep variational auto-encoder method (DVAE) for the removal of noise from retinal image and restoration to the exploitation of the benefits of pre-processing steps.
- Extracted features from convolutional model map into latent space which reduce the dimension of the data and restore the noise-free image with less number of features with robustness.
- Deep variational auto-encoder (DVAE) exploits the multi-level loss function using Kullback–Leibler divergence (KLD) in proposed scheme to discourse restoration for corrupted retinal images.
- We have designed DA based on total variation model to develop clarity of the retinal image for efficient elimination of noise with a very less number of epoch

and the accomplishment of the proposed scheme surpasses from state-of-the-art algorithms while achieving the best restoration results.

The rest of this chapter is structured as follows: In Sect. 2, we have discussed the preliminaries and mathematical model. The proposed methodology and designed retinal image restoration scheme in detail have been discussed in Sect. 3. The experimental results and discussion have been explored in Sect. 4, and finally conclusions of this work have been given in Sect. 5.

2 Preliminaries

This section concisely discusses the theories of deep variational auto-encoder (D-VAE) [19, 21] and the basic framework of retinal image denoisification method [20]. The proposed algorithm deep variational auto-encoder method (DVAE) is established on the basis of following perception.

2.1 Denoise Model for Retinal Images

According to mathematical definition [20, 22], we consider that a given $\mathbf{X} \in \mathbb{R}^{m \times n}$ noisy retinal image and $\mathbf{y} \in \mathbb{R}^{m \times n}$ is its corresponding noise-free image. Then, the interrelation between $\mathbf{X}$ and $\mathbf{y}$ can be written as follows [18]:

$$\mathbf{X} = \eta\,(\mathbf{y})$$

where $\eta : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ denotes the extortion of the additive noise that distorts the normal retina. To recover the corrupted image, the noise reduction algorithm (viz. a function) is mainly converted into the original once while finding a function ϕ that holds the following criterion [14]:

$$\phi = \arg\min \|\phi\,(\mathbf{x}) - \mathbf{y}\| \tag{1}$$

where ϕ is deployed as the best approximation of η^{-1} and $\|\cdot\|$ is l_1 norm.

2.2 Variational Auto-Encoder (VAE) for Noisy Retina

For retinal image restoration, we have chosen unsupervised-learning-based deep variational auto-encoder (DVAE) which regenerates a noise-free image from its noisy version with self learning [21, 23]. Let $\mathbf{X}$ be a noisy retinal image and we have a vector of latent variables z taken from high dimensional space $\mathscr{Z}$ that can be easily sampled

to some probability density function (pdf) $P(z)$ is defined over $\mathscr{Z}$. We can define a family of deterministic function $f(z;\theta)$ parameterized by θ as $f : \mathscr{Z} \times \Theta \rightarrow \mathscr{X}$. Our objective is to optimize θ such that we can sample z from $P(z)$ with high probability $f(z;\theta)$ and it is defined by [21]:

$$P(X) = \int P(X|z;\theta)P(z)dz \tag{2}$$

By the definition of total probability, $f(z;\theta)$ is substituted by the distribution $P(X|z;\theta)$. In variational auto-encoder (VAE), the choice of output distribution is often Gaussian distribution since it generates similar sample and hence we may write $P(X|z;\theta = \mathscr{N}(z|\mu, \sigma * \mathbf{I})$, where mean and standard deviation of sample are μ and σ respectively [20, 21] and $\mathbf{I}$ denote identity matrix. The network $P(X|z;\theta)$ is known as "generative network," and it is intractable due to complex nonlineraity of neural network.

The parameterized distribution from Eq. (2) has been introduced into the proposed neural network architecture for input z over the parameterized distribution X, where θ is the weight vector of the neural network. The result of VAE is also the approximate distribution $\bar{P}$ on the observation where X is defined as $\bar{P}(X|z;\phi)$ (called inference network) for variational parameter ϕ, also the weight vector of network and z and X are the output and input, respectively. The primary objective of VAE is to maximize the variational lower bound which is defined as [20, 21]:

$$\begin{aligned} \log\ P(X;\theta) &\geq \mathbb{E}_{\bar{P}(z|X;\phi)}\left[\log \frac{P(X|z;\theta)}{\bar{P}(z|X;\phi)}\right] \\ &= \mathbb{E}_{\bar{P}(z|X;\phi)}[P(X|z;\theta) - \mathbb{KLD}(\bar{P}(z|X;\phi)||P(z))] \end{aligned} \tag{3}$$

where $\mathbb{KLD}$ is the Kullback-Leibler Divergence is derived as: [6, 23]:

$$\mathbb{KLD}(\bar{P}(z|X;\phi)||P(z)) = \mathbb{E}_{z\sim\bar{P}}[\log\ (\bar{P}(z) - \log\ P(z|X)]$$

The image restoration accuracy of an auto-encoder is calculated by the Eq (3) whose noise is injected in the hidden layer, i.e., output inference layer.

3 Retinal Image Restoration Scheme by Using Deep Variational Auto-Encoder

The proposed image reconstruction scheme is based on a deep learning approach that determines the sequences of noisy image as concurrent input and provides the sequences of noise-free output image (normal image) on a huge numbers of train-test data pairs, which consists of a train input sequence and also the desired target sequence. In this investigation, we have used deep variational auto-encoder for retinal

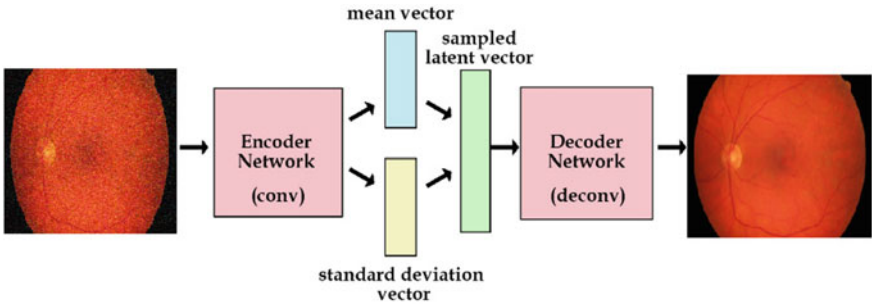

Fig. 1 Architecture of proposed auto-encoder

image restoration since mean (μ) vector and standard deviation (σ) vector of Gaussian distribution control the area of encoding of an input and the variation of area from mean. In encoding phase, image patches are selected at random and the decoder learns from all point in latent space, as well as nearby points of the reference latent space and hence the decoder is exposed to a range of variations of the encoding of the same input. Now, the model is designed with a certain degree of local variation by varying the encoding of one sample in latent space. Extracted features are mapped into the latent space which reduce the dimension of the features vectors using latent variables with probability distribution parameters in distribution function.

There are several values of vectors μ and σ for different samples, and hence the encoder can learn to produce different μ for different classes. So, a number of cluster generation depend on μ, and then minimize σ and μ accordingly. The decoder can efficiently reconstruct the training data based on the values of μ and σ. In order to optimize the model, we have suggested the Kullback–Leibler divergence ($\mathbb{KLD}$) into the loss function defined on Eq. (3). The KL divergence basically determines the divergence between two probability distributions. We have utilized KL divergence into loss function to closely resemble the target distribution, and optimize the probability distribution parameters μ and σ. The convergence rate of the proposed model is lesser than that of the other deep learning model due to dimension reduction property which reduces the computational cost.

We address the most deficiency by updating the DVAE architecture to incorporate "generative network" and "inference network" stages by batch normalization theme. Later, we continue this procedure by deploying the preparation of training data and also the perfect loss function with optimizer throughout training. Figure 1 shows the schematic block diagram of the proposed deep variational auto-encoder model (DVAE).

3.1 Training

In this section, we have illustrated data augmentation in order to construct the training data. Beginning with 300 slices of retinal image for each train sets during training

period, for each image, we initiate a noisy image at one frame per data sets, adding the Gaussian additive noise in training phase [5]. We make 512×512 images during explanation, but the training is tested with resizing as 128×128 patch which are randomly chosen for each training sequence. A sequence of 50 consecutive frames has been used to give abundant temporal substances in the training phase.

The variational EM algorithm has been applied to train the model defined on Eq. (3) which is discussed in the following steps:

- **E-step**: We update ϕ such that the approximate distributation $\bar{P}(z|X;\phi)$ will be close to the posterior distribution $P(z|X;\theta)$ by keeping θ fixed.
- **M-step**: We update θ to increase the marginal log-likelihood by keeping ϕ fixed.

In order to train the whole network of DVAE, parameter ϕ is applied in back-propagation considering z as a function of Gaussian noise of the output of encoder. The parameter θ of generative network also rapidly updates in backpropagation step, and hence we can update both parameters θ and ϕ together by stochastic gradient descent algorithm.

In data normalization, we have randomly chosen small rotation of the training sequence in 2–3 degrees in train increasing capture more temporal details of changing directions. We additionally choose a random gamma correction within the range [0, 1] one by one to every color channel, and put on the network to the complete sequence [1].

There is a simple way of training the proposed DVAE with denoising principle in three steps as: First, a corrupted input is sampled as $\bar{X}^{(i)} \sim P(\bar{X}|X)$, where $\bar{X}$ is the corrupted retinal image of the original version X. Second, sample $z^{(j)} \sim \bar{P}(z|\bar{X}^{(i)})$ and third, the generative network $P(X|z^{(j)};\theta)$ reconstructs the sample and parameters are optimized by the Eq. (5).

3.2 Loss Function

The computed error between training targets and outputs from network during the training phase is described as a loss function. The most ordinary MSE (mean squared error) from the train image S and the target image T is defined by the $\mathscr{L}_2$ loss functions. However, another loss function, $\mathscr{L}_1$ loss instead of $\mathscr{L}_2$ can be reduced the blemish artifacts from reconstructed images [7]. As mentioned above, the first loss term is a spatial $\mathscr{L}_1$ loss in the basis of $\mathbb{KLD}$, denoted as $\mathscr{L}_z$ applicable for a distinct image in the training phase [14]. The second loss term is a spatial $\mathscr{L}_1$ loss, stand for $\mathscr{L}_s$ for a particular image in the training stage [18]:

$$\mathscr{L}_s = \frac{1}{N} \sum_{j=1}^{N} \|\mathscr{S}_j - \mathscr{T}_j\| \tag{4}$$

where $\mathscr{S}_j$ and $\mathscr{T}_j$ are jth pixel of the resultant and goal image correspondingly.

The loss function of DVAE is decided by lower bound of the variational model which is approximated by Monte Carlo sampling as:

$$\begin{aligned}\mathscr{L}_v &= \mathbb{E}_{P(\bar{X}|X)}\mathbb{E}_{\bar{P}(z|\bar{X})}\left[\log \frac{P(X, z; \theta)}{\bar{P}(z|\bar{X}; \phi)}\right] \\ &\simeq \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\log \frac{P(X, z^{(j|i)}; \theta)}{\bar{P}(z^{(j|i)}|\bar{X}^{(i)}; \phi)}\end{aligned} \tag{5}$$

where $\bar{X}^{(i)} \sim P(\bar{X}|X)$ and $z^{(j|i)} \sim \bar{P}(z|\bar{X}^{(i)})$. Equation (5) represents the estimator of DVAE for all reconstruction parameters.

The $\mathscr{L}_2$ spatial loss gives a well image metric which is used to measure the overall outliers. To highlight the variances in adequate information (edge, corner, etc.), we use a Laplacian operator with square root of absolute $\mathscr{L}_1$ in the domain of gradient loss [7]

$$\mathscr{L}_l = \frac{1}{N}\sum_{j=1}^{N}\sqrt{\|\nabla\mathscr{S}_j - \nabla\mathscr{T}_j\|} \tag{6}$$

where individual gradient $\nabla(\cdot)$ is calculated based on high frequency error norm (HFEN), which is a metric used for medical image comparison [6]. The Laplacian of Gaussian kernel (3×3) has been used for edge detection, but it is sensitive to noise and hence to make image pre-smoothed, first Gaussian filter is used for edge detection [14]. To authenticate that the combined loss ends up in an enhancement above the spatial loss $\mathscr{L}_s$, it has been calibrated by the structural similarity metric (SSIM) [24] on tuning sequence of fundus after 50 epochs with 128 batch size of training. SSIM shown an enhancement for $\mathscr{L}_s$ from 0.9308 to 0.9736 for the fused loss.

4 Experiments

The effectiveness of our algorithm, along with comparisons is demonstrated on three standard publicly available databases, viz., DRIVE [25, 26] and STARE [27, 28], and DIARETDB1 [29] in this section.

4.1 Description of Datasets

In this work, we have used two standard publicly available databases DRIVE [25, 26] and STARE [27, 28]. The DRIVE database [25] consists of 40 color fundus images obtained from a diabetic retinopathy screening program in Netherlands. The images are compressed in JPEG format and are captured in digital form from a Canon CR5

non-mydriatic 3CCD camera at 45 degree field of view (FOV). The size of each image is 565 × 584 pixels, 8 bit per color channel and has a FOV of approximately 540 pixels in diameter. The 40 images have been divided into training set and test set, each containing 20 images.

The STARE [27, 28] database contains a total of 400 raw images of which small version of images can be viewed 100 at a time. The images are captured in digitized slides by TopCon TRV-50 fundus camera at 35 degree FOV. The size of each image is 700 × 605 pixels, 8 bit per color channel and has a FOV of approximately 550 pixels in diameter.

The DIARETDB1 [29] database consists of 89 fundus images. The images were captured with the same 50 degree FOV digital camera in Finland. The size of each image is 1500 × 1152 pixels, 8 bit per color channel. This database is mainly used to detect the hard exudates, hemorrhages, and microaneurysms.

4.2 Experimental Setup of DVAE

A total of 529 ordinary retinal images of size 565 × 584 obtained from aforementioned databases archive as normal-dose dataset in training phase. As stated above, the corresponding dataset of noise fundus images have been achieved by introducing Gaussian noise into the routine retinal images for the simulation of the proposed method DVAE. The salt-and-pepper noise plus with Gaussian noise is utilized to test the performances of the suggested model DVAE. In the initial configuration, $\mu_1 = 0, \sigma_1 = 30$ control the noise level of images. Each retinal image is corrupted by uniform impulse noise $p = 25\%$ and Gaussian noise $\sigma_1 = 30$, which are shown in Figs. 2b, and 3b for example.

In the training phase, we have randomly selected 400 normal (cleaned) and corresponding additive noise fundus pairs. Patch size p is set to 128 × 128 pixels. We have evaluated many parameter accumulation for the reconstruction of retinal image, and therefore Gaussian distribution is approximated to $\mathbb{N}(0, \sigma * I)$ in cost function Eq. (5), where σ is the hyper-parameter and I denote the identity matrix. With the increase in hidden layers and neurons, duration of training increases without noticeable performance gain, therefore the number of layers L are set to 16 for all cases. All the experiments are performed on Python 3.6 using Keras and Opencv-python on an Ubuntu 18.04 LTS (Intel *i*58400 CPU and 16 GB RAM).

In the test phase, we have imposed noise in patches of test images are managed by the trained proposed DVAE, so we have used the resultant patches to regenerate the reconstructed image. Three metrics have been adapted for statistical quantification. The primary one is the root mean square error (**MSE**) [30], and later are picture quality score(**PQS**) [24] universal quality index (**UQI**) [31] and structural similarity index metric (**SSIM**) [24, 30], respectively. The highest value among them indicates minimum distortion and better reconstruction except MSE.

Three different baseline methods are compared for the performance evaluation, such as “An adaptive bilateral filter based framework for image denoising” abbre-

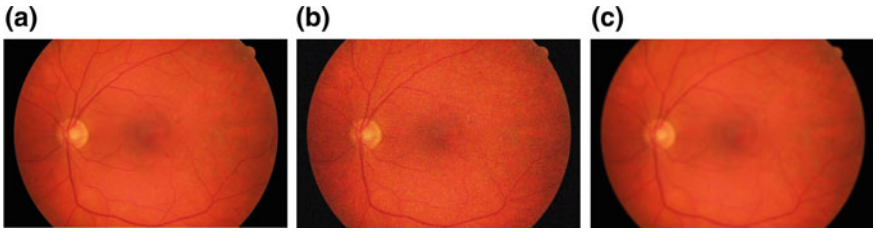

Fig. 2 Simulation with the proposed scheme. **a** The original image. **b** Noised image. **c** Reconstructed image (epoch = 50, batch-size = 64)

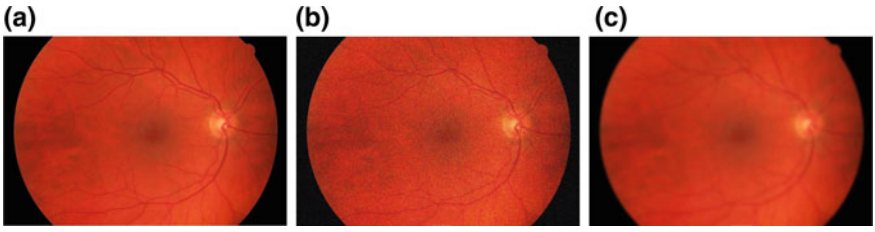

Fig. 3 Simulation with the proposed scheme. **a** The original image. **b** Noised image **c** Reconstructed image (epoch = 50, batch-size = 64)

viated for convention as (ABFD) [5], "Artifact suppressed dictionary learning for low-dose CT image processing" (SDLA) [14], and "Medical image denoising using convolutional denoising auto-encoders" (CDA) [18]. ABFD is the most fashionable image-domain-based denoising strategy which use bilateral filter recently. CDA is projected low-dose CT reconstruction ways supported by neural networks. All the connected parameters within the baseline techniques are fixed as per the recommendation of the primary investigation.

4.3 Analysis of Different Parameters

The efficiency of the proposed DVAE method depends on different parameters of the algorithm. Here, we have examined the two key factors: the amount of training instance and noise level introduced in the databases that have been discussed in this section.

The training data: We have emphasized on the effect of the size of training data from three databases, viz., "DRIVE," "STARE," and "DIARETDB1." In the first phase, 400 image pairs have been randomly chosen from the databases as training data. In the next phase, data augmentation has been taken to magnify the training data supporting the initial 400 image pairs using Keras API. The transformations enclosed rotation (by 2–3), horizontal and vertical flipping, and scaling (varying scale factors

Table 1 Quantitative evaluation of various methods

Image	Method	MSE	PQS	UQI	SSIM
Figure 4	ABFD [5]	0.0276	26.87dB	0.846	0.824
	SDLA [14]	0.0227	27.52dB	0.833	0.841
	CDA [18]	0.0195	28.62dB	0.875	0.887
	DVAE	**0.0187**	**29.12dB**	**0.884**	**0.892**
Figure 5	ABFD [5]	0.0269	25.95dB	0.861	0.830
	SDLA [14]	0.0224	26.62dB	0.873	0.835
	CDA [18]	0.0202	28.14dB	0.884	0.878
	DVAE	**0.0182**	**29.52dB**	**0.902**	**0.921**
Figure 6	ABFD [5]	0.0308	27.53dB	0.894	0.872
	SDLA [14]	0.0238	28.45dB	0.803	0.847
	CDA [18]	0.0214	28.86dB	0.876	0.852
	DVAE	**0.0206**	**29.68dB**	**0.895**	**0.927**
Figure 7	ABFD [5]	0.0263	26.17dB	0.784	0.802
	SDLA [14]	0.0207	26.52dB	0.804	0.785
	CDA [18]	0.0187	27.35dB	0.878	0.825
	DVAE	**0.0176**	**28.72dB**	**0.882**	**0.853**

0.2–0.5). Lastly, similar operations for data augmentation have been executed on the new elite where a hundred image pairs images are achieved.

Noise level: In practice, consistent to noise levels from the training and test sets should not be similar; consequently, it is essential to justify the validity of the suggested method completely for various datasets with different noise levels. However, one testing set with a Gaussian noise level, $\mu_1 = 0$, $\sigma_1 = 30$ and salt-paper noise with $p = 25\%$ area unit is employed in the analysis. The training set preserved a similar noise level CDA [18]. The objective evaluations are shown in Table 1. It is determined that for a given noise level with large epochs the CDA and DVAE have higher MSE, PQS, and UQI values than different baseline methods. Figure 2 and Fig. 3 demonstrate the reconstruction of the retinal image by the proposed DVAE where (b) is a noisy version of (a).

4.4 Qualitative Analysis

In test phase, one sequence frame from the test set has been selected to assess the performance of the proposed DAVE. Figures 4, 5, 6, and 7 demonstrated the results achieved from a retinal image by proposed DVAE along with all baseline methods. Figures 4, 5, 6, and 7b show the noise retinal images and all vessels (especially thin vessels) disappeared in the images. Each one of the techniques can productively stifle the noise; however, in Figs. 4, 5, 6, and 7c, ABFD can be affected by conspicuous

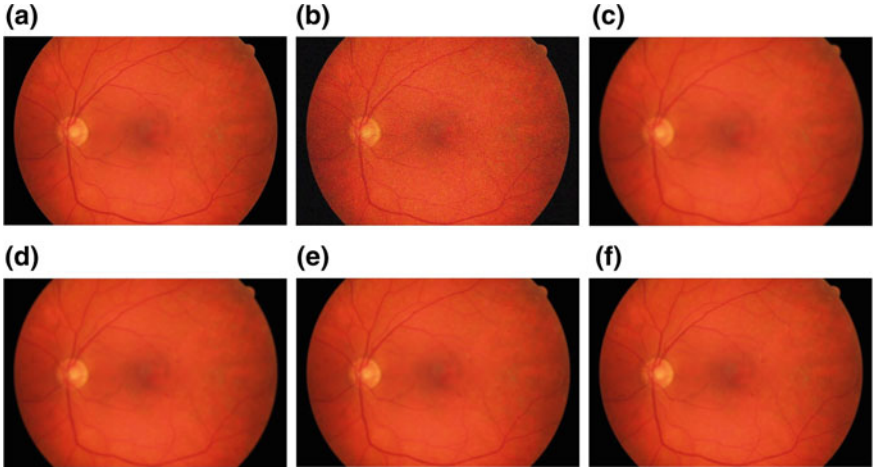

Fig. 4 Results of a Retina (in Fig. 2b) image for comparison with baseline methods. **a** Cleaned image; **b** noised image; **c** ABFD; **d** SDLA; **e** CDA; and **f** DVAE

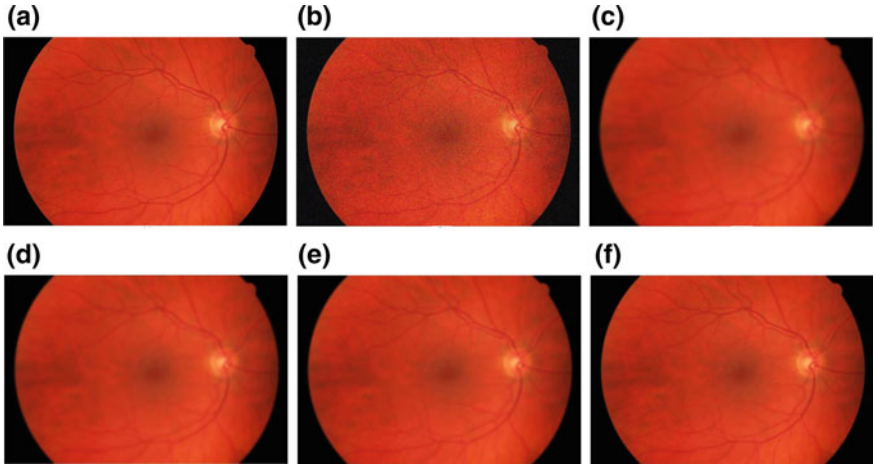

Fig. 5 Results of a Retina (in Fig. 3b) image for comparison with baseline methods. **a** Cleaned image; **b** noised image; **c** ABFD; **d** SDLA; **e** CDA; and **f** DVAE

square effect and oversmoothed a few tiny vessels in the retinal border region. CDA retains major details than SDLA, but the tiny spot in the center region of retinal images still has the noisy appearance with the low contrast levels. Deep-learning-based techniques CDA can protect more subtle elements but the noise and artifacts. Nonetheless, CDA oversmoothed few low contrast domains that are indicated in Figs. 4, 5, 6, and 7e.

Three baseline strategies (ABFD, SDLA, and CDA) have been put in place. It is detected in Figs. 4, 5, 6, and 7c–e that the three strategies preserve the tiny vessels

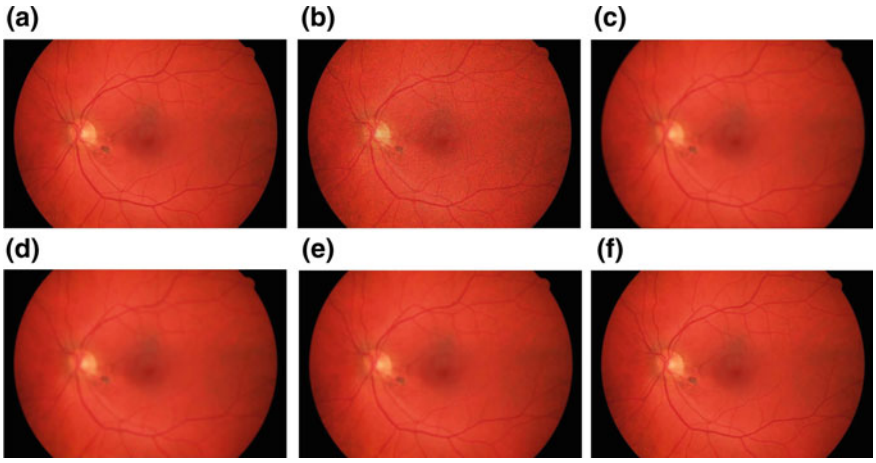

Fig. 6 Results of a Retinal image for comparison with baseline methods. **a** Cleaned image; **b** noised image; **c** ABFD; **d** SDLA; **e** CDA; and **f** DVAE

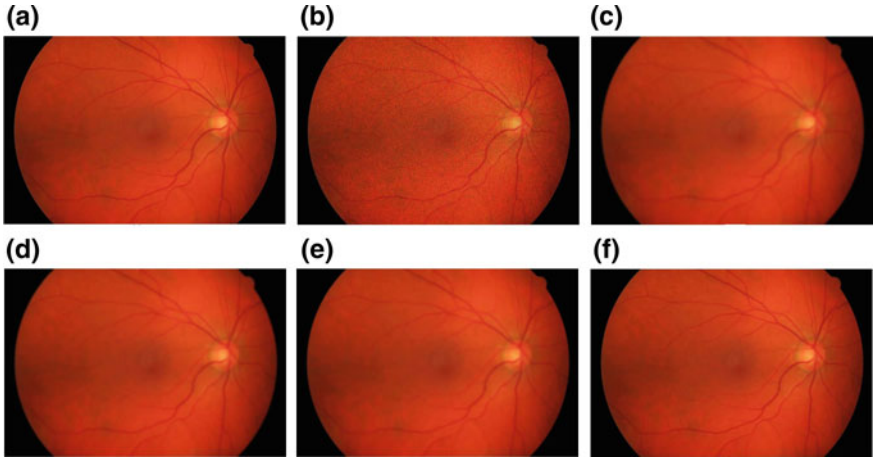

Fig. 7 Results of a Retinal image for comparison with baseline methods. **a** Cleaned image; **b** noised image; **c** ABFD; **d** SDLA; **e** CDA; and **f** DVAE

much distinct than different methods. DVAE additionally would have obtained the most effective performance within the distinctive low region if we had circumstantially determined all resultant data. Additional methods, mainly ABFD and SDLA failed to eliminate impulse noise completely. The proposed methodology DVAE can be preserved additionally by keeping noise-free pixels unchanged on multi-loss utilization and hence we can determine that the details of the restored image by the proposed methodology DVAE has produced the best visual effect compared to other methodologies.

4.5 Quantitative Analysis

The objective performance of the proposed method DVAE is compared with three existing methods as mentioned above shown in Figs. 4, 5, 6, and 7. The statistical results for the four metrics (MSE, PQS, UQI, and SSIM) are shown in Table 1 with bold font with the highest score of DVAE. For all the images, DVAE achieved better scores of PSNR and PQS, UQI and SSIM and interestingly, CDA also achieved better performance with respect to MSE, PQS, UQI, and SSIM, which is the same as DVAE, and it indicates that both are able to sustain more details in image restoration task. Table 1 shows the statistical validation of Figs. 4, 5, 6, and 7c–f in the test phase. Table 1 proves that the suggested DVAE method obtained ambitious achievement with CDA and excellent reconstruction than other baseline methods.

The performances of proposed techniques DVAE are compared with three baseline methods such as ABFD [5], SDLA [14], and CDA [18] in Table 1. The UQI obtained by baseline methods such as ABFD and SDLA are in the range of 0.78 to 0.88, whereas deep-learning-based techniques CDA in UQI are close to the proposed algorithm. SSIM for proposed technique DVAE is more important than ABFD, SDLA, and CDA and close is to 0.9 in all cases. So, we realize that edges of vessels are not well sustained using ABFD and SDLA, whereas the CDA and suggested method DVAE are able to do so. So, we can state that Table 1 reveals all image quality metrics with improved CDA and proposed technique DVAE than ABFD and SDLA methods. Metric indices of performance metric in Table 1 show superior noise reduction capabilities of the proposed DVAE scheme. In Fig. 8a–b shown that the test and validation loss of the DVAE with a certain number of epochs, respectively. The proposed model DVAE model has taken very less number of epoch (approximate 50) in training phase.

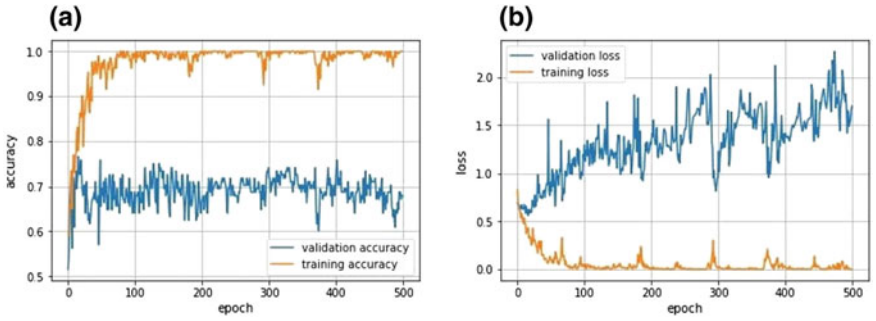

Fig. 8 Performance evaluation of the proposed method. **a** Curve for accuracy, and b curve for validation

5 Conclusion

In this work, deep learning that is based on variational auto-encoder has been recommended for retinal image restoration. The proposed method can increase the picture quality proficiently and rapidly as well as preserve image details more accurately since variational lower bound yield a sensible corruption function which is tighter than the standard lower bound on noisy image. The extracted features map into latent space using latent variable with Gaussian distribution which reduce the dimension of the data and converge for few iterations compared to other deep learning model. Both the subjective and quantitative results have exhibited and investigated its capacity to decrease noise and maintain the structure details. Proposed algorithm has reduced the computational cost because of its fast convergence for the diagnosis of several malfunctions from retinal image such as diabetic retinopathy, glaucoma, and exudates at the first screening of an ophthalmology. It is more robust and flexible than DAE since it uses approximate posterior distribution on input data. Some times variational model fails to restore images due to the presence of high amount of complex noises in the images. In future work, other issues the retinal images for diabetic retinopathy detection, optic disk detection, or exudates detection via deep learning will be exploited.

References

1. Sonka, M., Hlavac, V., Boyle, R.: Image processing analysis and machine vision. Thomson Learn (2001)
2. McBee, M.P., et al.: Deep Learning in Radiology. Elsevier, Academic Radiology (2018)
3. Moraru, L., Obreja, C.D., Dey, N., Ashour, A.S.: Dempster-shafer fusion for effective retinal vessels diameter measurement. Soft Comput. Based Med. Image Anal. 149-160 (2018)
4. Chakraborty, S., Mukherjee, A., Chatterjee, D., Maji, P., Acharjee, S., Dey, N.: A semi-automated system for optic nerve head segmentation in digital retinal images. In: 2014 International Conference on Information Technology (ICIT), IEEE (2014)
5. Zhang, Y., Tian, X., Ren, P.: An adaptive bilateral filter based framework for image denoising. Neurocomputing (2014)
6. Zhang, Y., Zhang, W.H., Chen, H., Yang, N.L., Li, T.Y., Zhou, J.L.: Few-view image reconstruction combining total variation and a high-order norm. Int. J. Imaging Syst. Technol. **23**(3), 249–255 (2013)
7. Hang, Z., Gallo, O., Frosio, I., Kautz, J.: Loss Functions for Image Restoration with Neural Networks. IEEE Transactions on Computational Imaging, ACM, arxiv.org/abs/1511.08861 (2016)
8. Abramoff, M.D., Garvin, M.K., Sonka, M.: Retinal imaging and image analysis. IEEE Rev. Biomed. Eng. **3**, 169–208 (2010)
9. Ghosh, S.K., Ghosh, A., Chakrabarti, A.: VEA: Vessels extraction algorithm and a novel wavelet analyser for diabetic retinopathy detection. Int. Journal of Image and Graphics-Worldscientific **18**(2) (2018)
10. Singer, D.E., Nathan, D.M., Fogel, H.A., Schachat, A.P.: Screening for diabetic retinopathy. Ann. Intern. Med. **116**, 660–671 (1992)
11. Kanski JJ, Bowling B (2012) Synopsis of Clinical Ophthalmology. Elsevier Health Sci

12. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
13. Goodfellow, I., Bengio, Y., et al.: Deep Learning. MIT Press (2016)
14. Chen, Y., Shi, L., Feng, O., Yang, J., Shu, H., et al.: Artifact suppressed dictionary learning for low-dose CT image processing. IEEE Trans. Med. Imaging **33**(12), 2271–2292 (2014)
15. Zhang, Y., Wang, Y., et al.: Statistical iterative reconstruction using adaptive fractional order regularization. Biomed. Opt. Exp. **7**(3), 1015–1029 (2016)
16. Li, Z., et al.: Adaptive nonlocal means filtering based on local noise level for CT denoising. Med. Phys. **41**(1) (2014)
17. Huang, Z.K., Li, Z.H.: Comparison of different image denoising algorithms for Chinese calligraphy images. Neurocomputing **188**, 102–112 (2016)
18. Gondara, L.: Medical image denoising using convolutional denoising autoencoders. Pro. of the IEEE Int. Con. Data Mining Workshops (ICDMW 2016) pp. 241–246 (2016)
19. Doersch, C.: Tutorial on variational autoencoders. Mach. Learn. arxiv.org/abs/1606.05908 (2016)
20. Im, D.J., Ahn, S., Memisevic, R., Bengio, Y.: Denoising Criterion for Variational auto-encoding framework. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17) arXiv:1511.06406 (2017)
21. Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A.: Variational autoencoder for deep learning of images, Labels and Captions. In: 30th Conference on Neural Information Processing Systems (NIPS 2016), arXiv:1609.08976 (2016)
22. Vincent, P.: A connection between score matching and denoising autoencoders. Neural Comput. **23**(7), 1661–1674 (2011)
23. Mao, X.J., Shen, C., Yang, Y.B.: Image restoration using convolutional autoencoders with symmetric skip connections. In: Proceeding of NIPS (2016)
24. Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity. IEEE Tran. of Image Processing **13**(4) (2004)
25. Niemeijer, M., Staal, J.J., Ginneken, B., Loog, M., Abramoff, M.D.: Comparative study of retinal vessel segmentation methods on a new publicly available database In: Fitzpatrick, J.M., Sonka M. (eds.), SPIE Medical Imaging, SPIE 5370, pp. 648-656 (2004)
26. Available from: https://www.isi.uu.nl/Research/Databases/DRIVE/
27. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imaging **19**(3), 203–210 (2000)
28. Available from: http://cecas.clemson.edu/~ahoover/stare/
29. Kalesnykiene, V., Kamarainen, J. K., Voutilainen, R., Pietil, J., Klviinen, H., Uusitalo, H.: DIARETDB1 diabetic retinopathy database and evaluation protocol (2007)
30. Gu, K., Zhai, G., Yang, X., Zhang, W.: Using free energy principle for blind image quality assessment. IEEE Trans. Multimedia **17**(1), 50–63 (2015)
31. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett. **9**(3) (2002)