A Robust System for Noisy Image Classification Combining Denoising Autoencoder and Convolutional Neural Network

Sudipta Singha Roy
Institute of Info. and Comm. Technology
Khulna University of Engineering & Technology
Khulna, Bangladesh

Sk. Imran Hossain, M. A. H. Akhand
Dept. of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh

Kazuyuki Murase
Graduate School of Engineering
University of Fukui
Fukui 910-8507,
Japan

Abstract—**Image classification, a complex perceptual task with many real life important applications, faces a major challenge in presence of noise. Noise degrades the performance of the classifiers and makes them less suitable in real life scenarios. To solve this issue, several researches have been conducted utilizing denoising autoencoder (DAE) to restore original images from noisy images and then Convolutional Neural Network (CNN) is used for classification. The existing models perform well only when the noise level present in the training set and test set are same or differs only a little. To fit a model in real life applications, it should be independent to level of noise. The aim of this study is to develop a robust image classification system which performs well at regular to massive noise levels. The proposed method first trains a DAE with low-level noise-injected images and a CNN with noiseless native images independently. Then it arranges these two trained models in three different combinational structures: CNN, DAE-CNN, and DAE-DAE-CNN to classify images corrupted with zero, regular and massive noises, accordingly. Final system outcome is chosen by applying the winner-takes-all combination on individual outcomes of the three structures. Although proposed system consists of three DAEs and three CNNs in different structure layers, the DAEs and CNNs are the copy of same DAE and CNN trained initially which makes it computationally efficient as well. In DAE-DAE-CNN, two identical DAEs are arranged in a cascaded structure to make the structure well suited for classifying massive noisy data while the DAE is trained with low noisy image data. The proposed method is tested with MNIST handwritten numeral dataset with different noise levels. Experimental results revealed the effectiveness of the proposed method showing better results than individual structures as well as the other related methods.**

Keywords—Image denoising; denoising autoencoder; cascaded denoising autoencoder; convolutional neural network

I. Introduction

Categorization of objects from images is a complex perceptual task and is termed as image classification. Classification of images utilizes multispectral data. The underlying multispectral pattern of the data of each individual pixel is utilized as the quantitative basis for classification [1]. In the past decade, image classification has shown major advances in terms of classification accuracy. In recent times, image classification models are rapidly being used in various application fields, such as handwritten numeral recognition [2], recognition of traffic signs from roadside boards [3]-[5], segmentation of Magnetic Resonance Image (MRI) [6], identification of chest pathology [7], face detection from images [8] and so on. Existing models are categorized into unsupervised and supervised modes.

Unsupervised classification based models try to find out the underlying representation in the input images without considering whether the images are labeled or not. One conventional model of this genre is stacked autoencoders (SAE) [9]-[11]. With an intention to learn features, SAE stacks shallow autoencoders which at first encodes the original input image to a vector of lower dimension and then decodes this vector to the original representation of the image. Shin et al. [12] showed the application of stacked sparse autoencoders (SSAEs) to classify medical images which made a noteworthy promotion in terms of classification accuracy. Norouzi et al. [13] inaugurated stacked convolutional restricted Boltzmann machine (SCRBM) where they applied a modified training process rather than the conventional one for individual restricted Boltzmann machine (RBM) and finally combined them in a stacked manner to implement the deep architecture. Later, Lee et al. [14] instigated another variant of deep belief network (DBN) called convolutional DBN (CDBN) by placing convolutional RBMs (CRBM) instead of traditional RBMs at each layer and then joined the layers in a convolutional structure to ensure the construction of a hierarchical model and it produced better feature representation [15]. With the practically identical considerations, Zeiler et al. [16], [17] modified traditional sparse coding technique [18] to build deconvolutional model that decomposes the input data in a convolutional way, at the same time, maintains a sparsity constraint. In contrast to conventional sparse coding technique, this approach produces mid-level delineations of data with more affluent learned features.

Unlike unsupervised classifiers, supervised classification based models require labeled data to complete their training process. In this category, deep neural network (DNN) does the task efficiently implementing the idea of human visual system.

Layer-wise pre-training and fine-tuning makes DNN successful in image processing tasks such as classification, feature extraction etc. Convolutional neural network (CNN) [19]-[22] is the most successful hierarchical deep neural network structure. Shared weights, three-dimensionally arranged and locally connected neurons make the architecture of CNN distinctive to ordinary neural networks and contribute to its superior performance to most of the image classification algorithms [23]. The unique characteristics of CNN such as weight sharing and preservation of the corresponding locality, which make the deep architecture the most suitable for 2D images to conserve a better epitome, are the outcome of using convolution and following subsampling layer. Right now, CNN based models are being used vastly in 2D material identification and various cases [3]-[8].

One major challenge in image classification tasks is the presence of noise that corrupts the original shape of the objects in the image and makes it difficult for the classifiers to be used in real life scenarios. Unlike human visual system, which is capable of classifying objects ignoring a certain amount of perturbation present in the image, these classifiers suffer in quality if the test image contains noise. Although DNN based methods outperformed others in image classification, their performances are deteriorated during classification of noisy images. However, it is quite impossible to work with noiseless images in practical cases. During acquisition and transition phases, corruption of digital images due to noise is common. As DNN based models are trained to work with noiseless images, their accuracy noticeably drops when they are applied in real life applications. The main reason that works behind the occurrence of this incident is the affection of the DNN based models to the training data. Because of this sensitive behavior towards the training data, often these models perform misclassification if the test data is subject to a significant amount of noise and distortion [24].

It is an open challenge to develop image classification systems for the real-life noisy environment. Lu and Weng [25] investigated different image classification models and finally came to a conclusion that denoising images prior classification is the best possible way to make the DNN based models more compatible with practical cases. Their survey gives the evidence of the fact that training classifiers with noisy images may enhance the precision a tad; however, it is not satisfiable. So, applying image denoising techniques before feeding the image to the classifier has become a compulsory to fit the DNN based classifiers in real life scenario.

A number of researches have been conducted to recover the true image from the noisy form by applying both spatial and transform domain [26]. Several pioneer image denoising researches used wavelet transformation techniques [27], partial differential equation based approaches [28]-[30], and conveyed scant coding approaches [18], [31], [32]. Singh et al. [33] introduced a multi-class classifier for images which are adulterated with Gaussian noise. To accomplish image denoising they utilized NeighShrink thresholding over the wavelet coefficients to wipe out wavelet coefficients which are responsible for the noise present in the image and find out just the useful ones. However, these denoising approaches face problems in case of heavily noised image and are computationally complex. In the process of image denoising using spatial filtering techniques images gets blurred, where transfer domain filtering models are time-consuming as well as computationally complex.

Recently, artificial neural network (ANN) based models are being adopted in image denoising tasks. A variant of autoencoder (AE) named denoising autoencoder (DAE) [34], [35] has been introduced to serve the purpose of image denoising and shows a better performance compared to the traditional ones. In DAE the initial input gets corrupted by arbitrary noise then it is trained to restore the original image from its' corrupted version. In [36] Vincent et al. stacked a number of denoising autoencoders and established a deep network named stacked denoising autoencoder (SDAE) which is widely implemented for unsupervised learning. Agostinelli et al. [37] developed an adaptive multi-column DNN combining multiple stacked sparse DAEs (SSDAE), where the multi-column architecture empowers the model to deal with images corrupted by not one type but three different types of noises. Utilizing non-linear optimization technique, they figured out the most favorable column weights at first and then individual models were trained to make them anticipate those optimal weights. Incorporating the idea of AEs and convolutional operation Masci et al. [38] introduced convolutional autoencoder (CAE) which can preserve better spatial locality. CAE is based on CNN and it learns to reconstruct the images at the output end from the input image set applying convolutional approach so that the kernels convolve over the 2D images and at each layer generates more abstract feature maps. In order to use this convolutional structure of AEs for image denoising task, Gondara [39] deployed DAEs along with CAEs. She utilized the DAE, at first, to denoise medical images and then CAEs to generate a better representation of the images. Xu et al. [40] implemented a deep CNN architecture that can find out the features of blur degradation present in an image.

Du et al. [41] introduced stacked convolutional denoising autoencoder (SCDAE). To maintain a hierarchical structure, they arranged a stack of DAEs in a convolutional manner. Additionally, they embedded a whitening layer in front of each and every convolutional layer to enclose the input feature maps. Most recently, Roy et al. [42] applied convolutional denoising autoencoder (CDAE) followed by a DAE and arranged them in a cascaded manner, rather than in a stacked way to deal with data subject to massive noise. They showed that if two AE based models are individually trained to denoise images subject to regular level of noise, the cascaded architecture of them can show a great performance in case of denoising massive noisy images. Still, these models suffer from one limitation: their performances require the presence of a quite same proportion of noise in both training and testing dataset. To fit the DNN based image classifiers in real life scenario, models should be able to work with variable level of noise i.e., regular to massive level of noise.

The aim of this study is to develop a robust image classification system which performs well in any noise level with a minimized computational cost, at the same time, omits the requirement of arranging multiple training of the system with images containing variable proportion of noise separately to deal with images subject to variable level of noise. The proposed method first trains a DAE with low-level noise injected images and a CNN with noiseless native images independently. Then it arranges the two trained models in three different combinational ways: CNN, DAE-CNN, and DAE-DAE-CNN. Finally, it combines the outcomes of these three combinations for system outcome. The motivation of such arrangement is the adaptation of noise in DAE and image classification ability of CNN. Since CNN is trained with native images without noise it is well for noiseless image classification. On the other hand, DAE-CNN and DAE-DAE-CNN structures perform well for low-level and high-level noisy cases, respectively [42]. In DAE-CNN, DAE first removes noises from noisy input images and then CNN is fed with these restored images for classification purpose. In DAE-DAE-CNN, two pre-trained DAEs are cascaded together and followed by a CNN. First DAE denoises the input images which are further filtered by the next DAE and therefore CNN gets the restored images, which is better suited to classify the test images in case they are adulterated with massive noise even though both DAEs are same and trained with the low noise level. The winner-take-all combination gives system output emphasizing confidence of individual structure and thus the proposed model performs well to classify images for noiseless to high-level noise cases. In this study, the proposed method is tested with MNIST handwritten numeral dataset and its performances are compared with other related methods.

The rest of the paper is designed as follows. Section II describes the proposed robust system for noisy image classification along with some preliminaries for better understanding. Section III shows the result of the proposed method as well as performance comparison with some other existing related research works. Finally, a brief conclusion of this work is presented in Section IV.

II. Robust System for Noisy Image Classification

In practical life, image classification models suffer from noise, injected in the image while acquiring and transmitting, as well as other imperfections existing in the image. Exiting systems are found to be effective for a fixed level of noise on which they are trained. On the other hand, this study investigated a robust system which performs well in classification of images in spite of the varying level of noise present in the image. The proposed method first trains a DAE with low noise level and a CNN independently. The main novelty of the proposed system is the innovative combinational arrangements of the trained DAE and CNN for three different structures. This section first describes the training of individual DAE and CNN briefly; and then explains the proposed system.

A. Review of DAE and CNN

The main computational components of the proposed system are DAE and CNN. Well studied standard DAE and CNN architectures are considered in this study. For a better understanding as well as to make the paper self-contained, DAE and CNN are presented briefly.

1) Denoising Autoencoder (DAE): Autoencoder (AE) is a three-layered neural network, which is unsupervised and deterministic in nature. It maps the input into a hidden representation through encoder and then decoder maps it back to a reconstruction, which is of the same shape of the input. DAE, unlike basic AE, forces the hidden layer to capture the information about how the inputs are statistically dependent on each other instead of learning trivial features [14]. This is done by the corruption of input dataset x into $\tilde{x}$ stochastically ($q_D(\tilde{x}|x)$) which is mapped into a hidden representation y.

$$y=\mu(W\,(q_D(\tilde{x}|x)+b))\ , \qquad (1)$$

where W is the weight of the input-hidden layer and q_D represents the type of distribution with a certain probability D. q_D depends on two parameters: one, the distribution of the original input x and two, the type of noise corrupting the images. Clinched alongside practical scenarios, binomial noise is utilized while working with black and white pictures whereas, to color pictures uncorrelated Gaussian noise is superior suiting. At that point, $\tilde{x}$ is mapped to a low dimensional hidden depiction y using nonlinear deterministic function μ. Finally, this hidden representation gets mapped into a reconstruction z which has as close as possible resemblance to input x. This process also passes through another nonlinear deterministic function $\emptyset$.

$$z=\emptyset(W\acute{}y+b\acute{}) \qquad (2)$$

where W' is the weight of the hidden-output layer. Thus, DAE is capable to generate representations of features, which is suitable for the classification task. The architecture of DAE is shown in Fig. 1. The training of DAE requires it to be fed with noisy images putting the corresponding native images at the output layer. During backpropagation, this model learns to filter out the underlying noise from the input image and reconstruct a noiseless one. The detailed description, as well as training of DAE, is available in the previous studies [42].

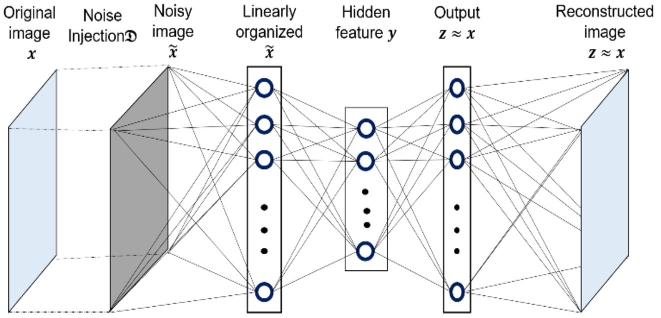

Original image x
Noise Injection $\mathfrak{D}$
Noisy image $\tilde{x}$
Linearly organized $\tilde{x}$
Hidden feature y
Output $z \approx x$
Reconstructed image $z \approx x$

Fig. 1. Denoising Autoencoder (DAE) architecture.

2) Convolutional Neural Network (CNN): CNN [19] is a variant of neural network popular for object detection and segmentation task. Regular neural network or multi-layer perceptron has some limitations: it suffers from overfitting; it ignores the fact that there is a strong correlation among neighborhood pixels and it is sensitive to any kind of transformation of the image. CNN overcomes these problems by ensuring spatial local connection, weight sharing, and subsampling. The operation of a CNN is done on the premises of two basic operations: convolution and subsequent subsampling. Convolution operation forces a kernel, which is an organization of weights and bias, to convolve over input feature map (IFM) which in the end results in a convolved feature map (CFM). Throughout the convolution operation, the same kernel is applied to each and every small segment of each IFM, which is called the local receptive field (LRF), to acquire every specific point of the CFM. Throughout this process, both weights sharing among each and every position as well as the preservation of special locality are done simultaneously. From an IFM the CFM can be calculated by.

$$CFM_{(x,y)} = \tau(\sum_{i=1}^{K_{\hbar}} \sum_{j=1}^{K_{w}} K_{(i,j)} * IFM_{(x+i,y+j)} + \beta) \qquad (3)$$

where τ and * symbolize the activation function and the 2-D convolution operation accordingly. The bias of the applied kernel $K_{\hbar} \times K_{w}$ is symbolized by β . To conduct the experiment here, relu is used as the activation function, whereas for every latent map single bias is used.

The feature map, obtained from convolution operation, is processed by applying the following subsampling layer in order to gain a simplified form. This procedure of simplification is accomplished by choosing important features from a locale and discarding whatever is left of the ones [41]. Having different sub-sampling methods available, throughout the experiments here, max-pooling [21] has been utilized. Max-pooling operation picks the most important feature over non-covering sub-regions and this process can be defined as:

$$FM(x,y) = \mathscr{s}\left(\sum_{i=0}^{R-1} \sum_{j=0}^{C-1} CFM_{(xR-1+i,yC-1+j)}\right) \qquad (4)$$

where $\mathscr{s}$ symbolizes the max-pooling operation over the pooling locale and the size of the pooling area is represented as $R \times \mathrm{C}$ matrix.

Fig. 2 shows the most studied CNN architecture which is considered in this study. The CNN has two convolutional layers of filter size of 5×5 and the subsampling layer with a pool size of 2×2. A subsampling layer follows each convolutional layer. The convolutional and pooling layers together extract the features of the image. There is a fully connected layer and the input of which is the output of the second subsampling layer. This layer uses the extracted feature to classify the image depending upon the training dataset. The parameters of the network, as well as the kernel, get updated during the training process until the desired accuracy is achieved. The detail description of CNN training is also available in the previous studies [2], [42].

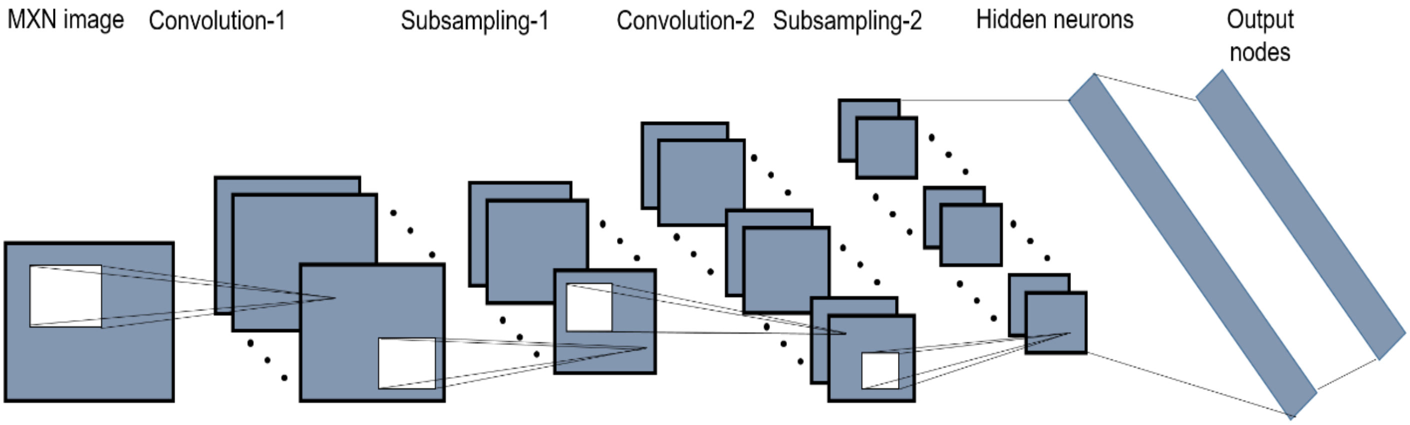


Fig. 2. CNN architecture for classification.

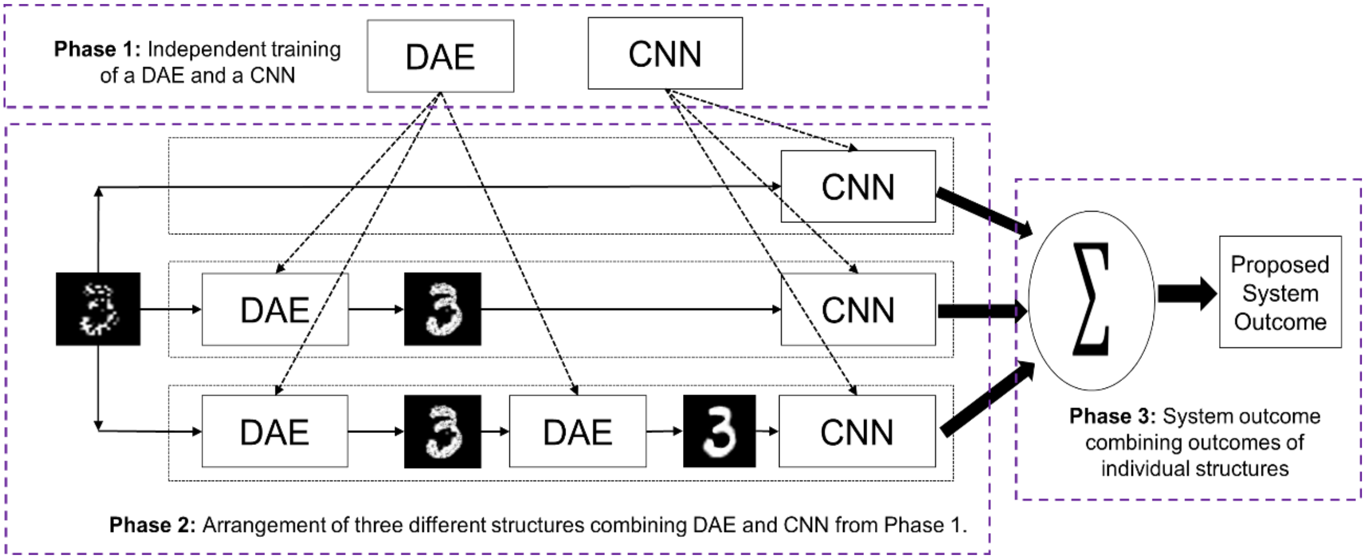

Phase 1: Independent training of a DAE and a CNN
DAE
CNN
CNN
DAE
CNN
DAE
DAE
CNN
∑
Proposed System Outcome
Phase 2: Arrangement of three different structures combining DAE and CNN from Phase 1.
Phase 3: System outcome combining outcomes of individual structures

Fig. 3. Proposed robust system for noisy image classification combining three different structures with DAE and CNN.

B. Proposed Robust System Combining Different Structures with DAE and CNN

Fig. 3 is the topological structure of the proposed Robust System based on DAE and CNN (RSDAECNN). Proposed RSDAECNN has three different functional phases. It trains a single DAE and a single CNN in Phase 1. In Phase 2, three different structures are arranged by copying the same DAE and CNN from Phase 1. In Phase 3, system outcome is prepared by combining outcomes of individual structures. Although proposed system consists of three DAEs and three CNNs in different structure layers, the DAEs and CNNs are the copy of same DAE and CNN trained in Phase 1. Thus, trainings of the DAE and the CNN in Phase 1 are the main computational elements in the proposed system.

Training of a single DAE and a single CNN only in Phase 1 makes the proposed system computationally efficient. The CNN is trained with native images and is used for classification purpose. The DAE is trained to restore the native image from the noisy image. In this study, the DAE is trained for regular level noise. Classification with CNN after restoring images through DAE might be helpful for noisy image classification. Since DAE is trained with regular level noise, different structures with different organizations of this DAE are managed to make the system adaptable to real-life environment where the noise level is not defined.

The main novelty of the proposed model is the innovative arrangements of pre-trained DAE and CNN to produce three different structures where each of the structures is responsible for dealing with images corrupted by a specific noise level. Each individual structure shown in Phase 2 has significant motivation to use in the proposed system. The CNN, common in all three structures, is used for classification purpose as CNN outperforms all other models in case of image classification [23]. To classify noiseless images CNN alone is good enough as CNN is trained with native images. This is the motivation of first structure with a CNN only in the system as seen in Phase 2 of Fig. 3. However, images corrupted by noise require a prior denoising step to improve the classification accuracy of the image classifier. For this purpose in the proposed system, DAE is used as the image denoiser. This DAE is trained only once with images corrupted by regular level noise (such as 20%). So, only a single DAE is sufficient enough to reconstruct the corresponding noiseless native form by filtering the images subject to regular level of noise. From this point of view, a DAE-CNN is placed as the second structure to emphasize the classification of images which are corrupted by regular level of noise. As the DAE and CNN are already trained separately, the DAE-CNN needs no further training. In DAE-CNN, DAE filters images subject to noise and then CNN classify the restored images.

A different structure DAE-DAE-CNN is developed to emphasize classification of images with massive noise because DAE-CNN structure is not sufficient enough to classify images in case of massive noise present in them. The DAE is trained with regular noise only and can't reconstruct native images which are corrupted with massive level of noises, such as if the percentage of noise in the images are around 50%. Roy et al. [42] showed that cascaded architectures of DAEs, where each of the DAEs is trained with 20% noisy images, can reconstruct images of good quality even if the noise level present in the images is 50%. Following this idea, a cascaded DAE-DAE is arranged as the image denoiser in the third structure DAE-DAE-CNN. In DAE-DAE-CNN, both the DAEs are the same in terms of architecture as well as all the corresponding parameters as they are copies of trained DAE in Phase 1. CNN is also the duplicate of the trained CNN in Phase 1 as like other two structures. Therefore, no additional training is required for this structure. In DAE-DAE-CNN operation, at first the image is filtered by the first DAE, then the intermediate image is further filtered by the following DAE. So, the pre-trained CNN is sufficient enough to classify the reconstructed image from massive noisy images after they are filtered by DAE-DAE. However, this model doesn't suit in case the level of noise in the image is not that much because restoration through DAE-DAE might overshoot to different images.

The proposed robust system combines the outcomes of the three structures, which are specialized to different noise levels while classifying an image subject to unknown level of noise. Among the three structures, structure with CNN alone is best suited for noiseless image classification. With DAE, DAE-CNN and DAE-DAE-CNN structures are suitable for images with comparatively less and heavier noise levels, respectively. An image with unknown level of noise is fed to all three structures at the same time and generates different outcomes. In Phase 3, winner-take-all combination is employed to generate outcome of proposed RSDAECNN system. Combination of outcomes from several individual systems is generally used in ensemble of classifiers and winner-take-all combination emphasizes the individual best confident system [43]-[44]. Therefore, the outcome of the proposed system will be correct classification selecting the outcome of the most confident structure. As an example, if the input image is noiseless, system outcome might come from the structure with CNN alone. On the other hand, system outcome might come from DAE-CNN and DAE-DAE-CNN for input image with less and heavier noise levels, respectively. As an example, if an image of '3' with massive noise is placed to the system, DAE-DAE will restore the original image as shown in Fig. 3 and CNN will classify it correctly.

C. Significance of the Proposed Model

There are several notable differences between the existing models and the proposed one on the premises of noisy image classification. Existing noisy image classification methods are found suitable for defined noise level. To work with less noisy data these models need to be trained with less noisy data whereas to classify massive noisy data the training data set should be corrupted by similar proportionate of noise. However, the proposed model can work with zero to massive level of noise due to the innovative arrangement of trained DAE and CNN for three different structures.

This model also omits the necessity of the system to be trained for images with different noise levels. It requires a single DAE to be trained with images containing regular level of noise and a CNN with noiseless images. Instead of using multiple training it places different arrangements of this trained DAE and CNN to deal with images carrying different proportion of noise. Thus, it reduces the pre-processing time for preparing the training dataset.

One more significant contribution of this work is the computational efficiency. To develop the proposed model with three DAEs and three CNNs, only one DAE and one CNN are trained independently. The cascaded DAEs in DAE-DAE-CNN also contains same trained DAE. Innovative arrangements of a trained DAE and a trained CNN makes the system computationally efficient.

III. Performance Evaluation

This section investigates the performances of proposed RSDAECNN on the benchmark image dataset MNIST numeral images [19]. This section first gives the description of the dataset and the experimental setup used to work over this dataset and afterward the results of the experiments conducted on images of different noise levels and lastly looks at the capability of the proposed model against existing ones. This model is implemented in Matlab R2017a. The performance analysis has been conducted on MacBook Pro Laptop (CPU: Intel Core i5 @ 2.70 GHz and RAM: 8.00 GB) in OS-X Yosemite environment.

A. Dataset Description and Experimental Setup

MNIST database [19] consists of 70000 sample gray-scaled images of handwritten digits collected from individuals having different writing styles. There are two sets of data: training set, which consists of 60000 images, and testing set of 10000 images. For each of 10 digits there are 6000 training samples and 1000 testing samples. Images in this dataset are of size 28×28.

In order to conduct a fair analysis of the proposed model's performance against the existing ones, a uniform experimental environment is required. The DAE used here has 784 input nodes as the images in the MNIST dataset are of 28×28 size and DAE can be fed with linearized data only. DAE includes 784 input neurons, 500 hidden neurons and 784 output neurons. The input of DAE is a linearly oriented noisy image of size of 28×28 whereas the output is the linearly oriented raw image of size 28×28.

CNN, the only classifier used here, is trained with clean images of 28×28 size. The CNN used here is two layered where each layer contains a convolution layer and a following subsampling layer. The kernels and other parameters are initialized randomly. The filter used for the convolution task in both layer is a 5×5 matrix. This filter slides over the original image and for every position the dot product is calculated which results in the feature map. The size of the feature map is 24×24 and the depth is 6 as the number of filters used is 6. Afterwards, max pooling is applied separately on each feature map with a spatial neighborhood of 2×2 window and the size of the feature map becomes 12×12. It is followed by another convolution and pooling operation with the same sized kernels and pooling region as before, which further reduces the size of the feature map to 192 as the depth of convolution layer used here is 12. The output of the second pooling layer acts as the input of the fully connected layer, which calculates the output probabilities for each class. So, there would be 192 nodes in the hidden layer. The data in this benchmark dataset is distributed among 10 classes. That's why the CNN used here contains 10 nodes in the output layer.

In order to deal with noises in the images, all the images in the training set are corrupted with 20% random noise. For, the testing purpose, the test dataset is used once as it is, then corrupted with 10% noise, afterward, they are adulterated with 20% noise and finally to check the performance of our model with massive noisy images, we increased the level of noise included in the images to 50%. To add noise in the training and testing image samples zero masking noise has been used where a random matrix is initialized with the same size of training data with some of the pixels within the data being randomly OFF having probability of 20%. For testing purpose, another three random matrices of same size are initialized where 10%, 20% and 50% data are randomly turned OFF. These matrices are multiplied with the raw images to generate the noisy images.

B. Experimental Result and Analysis

This section evaluates the classification performance of the proposed system against MNIST dataset on the premises of various proportionate of noise present in the image to validate its performance in case of dealing with variable level of noise. To simulate the performance of the proposed system for real world scenario where images can be noisy but prior knowledge about the level of noises is not possible, different level of noises has been added to the dataset because MNIST does not carry noises.

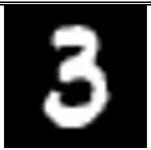
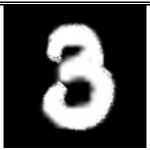
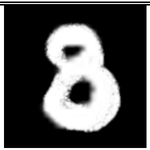
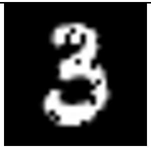
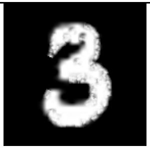
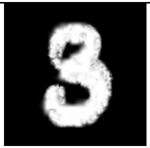
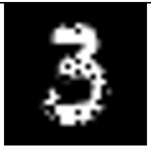
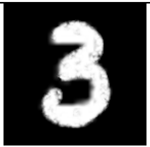
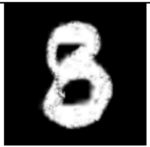
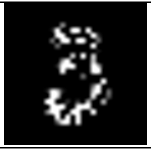
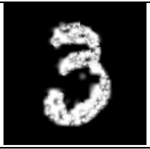
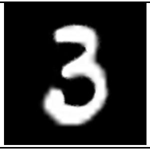
In this study, we implemented masking noise where fraction of the pixels of input image is forced to be zero having probability of 0%, 10%, 20% and 50%. At first a detailed presentation has been given for a sample image containing different noise levels as well as the reconstructed ones from DAE and DAE-DAE and finally the classification results. Experimental results for the dataset are collected for individual structures as well as proposed RSDAECNN system and are compared with other prominent methods. The performance of the system is analyzed on the basis of image reconstruction as well as classification accuracies represented by both confusion matrices and accuracy graphs.

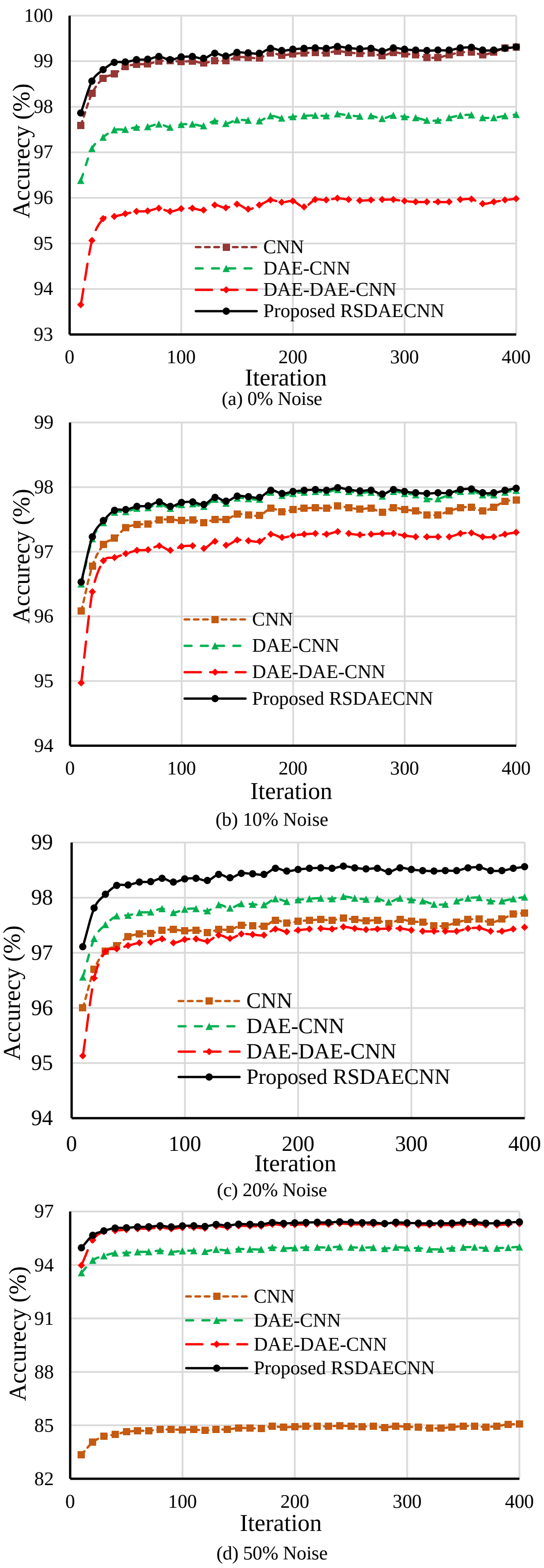
Table I delicates the outputs of image denoising step applying DAE and DAE-DAE architectures as well as the obtained classification results from three different structures (i.e., CNN, DAE-CNN, DAE-DAE-CNN) as well as the classification result of the proposed PSDAECNN. Images of '3' with different noise levels are considered as inputs of the proposed system those are classified with individual structures and generate system outcome. In case of noiseless image, it is seen that first structure (i.e., only CNN) correctly classified the image as "3". However, the reconstructed image obtained from a DAE seems more like numeral "8" and whenever it is filtered by DAE-DAE the image turned into the image of numeral "8". There remain two reasons behind the occurrence. Firstly, the DAEs being used here is the pre-trained DAE which is learned to reconstruct noiseless image from a noisy version of it. In case it is fed with a noiseless image it is not possible for it to know whether the image contains no noise and tries to reconstruct an image taking the input image as a noisy image. Secondly, the structures of numeral "3" and "8" are quite same.

So, DAE takes the input image and reconstruct an image like the numeral “8”. The DAE-DAE architecture is a two-layered cascaded form of the same DAE. The same scenario happens with it also. So, both the DAE-CNN and DAE-DAE-CNN misclassify the image as numeral “8”. Still, the proposed model classifies this image accurately as “3” because single CNN classified it correctly with more confident level. In case of both 10% and 20% noisy form of the very same image, only DAE-DAE-CNN misclassifies it as numeral “8”. In case of DAE-DAE-CNN structure, the image is first filtered by the frontier DAE. As the proportion of noise present in the image is less, this frontier DAE is sufficient enough to output the noiseless and good quality image. This reconstructed image is then again fed to the following DAE which also takes it as noisy image and tries to denoise it which in the end outputs a disordered image which looks like numeral “8”. The scenario is different in case of 50% noisy images. This time without any additional denoising technique CNN classifies it as numeral “5” whereas, both the DAE-CNN and the DAE-DAE-CNN classify it correctly. Though DAE-CNN classifies it correctly, from the figure it is clearly observable that the quality of the image reconstructed by DAE-DAE is far better and more like the original one as it is in case of the reconstructed one from the single DAE. Finally, the proposed RSDAECNN classified correctly all four cases although individual structures generate different outcomes.

Fig. 4 shows test set image classification accuracy of the proposed RSDAECNN system along with individual structures (i.e., CNN, DAE-CNN and DAE-DAE-CNN) for 0%, 10%, 20% and 50% noisy images up to 400 epochs. For 0% noise (Fig. 4(a)), structure with CNN alone achieved the highest classification accuracy among three individual structures and showed classification accuracy of 99.31%. On the other hand, classification accuracies of DAE-CNN and DAE-DAE-CNN are 97.83% and 95.99% accordingly. The reason behind these two models poor performance compared to single CNN is that CNN is trained with noiseless native images. Whenever any noiseless image is fed to a DAE trained with 20% noisy images, the DAE tries to convert the shape of the image to some other form assuming that the image is corrupted by 20% noise and results in producing a deformed image. The scenario is worse in case cascaded DAE is used. So, logically DAE-CNN and DAE-DAE-CNN perform worse compared to single CNN. However, because of using winner-takes-all model for final class label selection, the proposed model shows a better classification accuracy than these two models and same as the single CNN. For 10% noise (Fig. 4(b)) DAE-CNN is shown best suited individual structure because DAE is trained with 20% noise level. For this case performance of the proposed method is same as DAE-CNN. DAE-CNN is showed as best individual structure for 20% noise (Fig. 4(c)), but interestingly proposed model performed better than DAE-CNN for this case. On the other hand, for 50% noise case, DAE-DAE-CNN outperformed CNN and DAE-CNN. The reason behind is already explained that cascaded DAEs perform well than single DAE in case of image with massive noise as they are both trained at 20% noise level; after the first DAE works on a noisy image the second one gets an image with relatively less noise which gets further denoised. In such heavy noise case, proposed method showed the similar performance of DAE-DAE-CNN. Finally, considering all the scenarios the proposed model performs the best for noiseless to heavy noise cases.

TABLE I. SAMPLE OF ORIGINAL IMAGES WITH AND WITHOUT NOISE AND THEIR RECONSTRUCTION USING DAE, DAE-DAE AS WELL AS THE CLASSIFICATION RESULT OF CNN, DAE-CNN, DAE-DAE-CNN AND THE PROPOSED MODEL

Noise Level	**Input Image**	**Reconstructed Image**		**Classification through Individual Structure**			**Classification of Proposed RSDAECNN Combining Individual Structures**
		DAE	**DAE-DAE**	**CNN**	**DAE-CNN**	**DAE-DAE-CNN**	
0%				**3**	**8**	**8**	**3**
10%				3	3	8	3
20%				3	3	8	3
50%				5	3	3	3



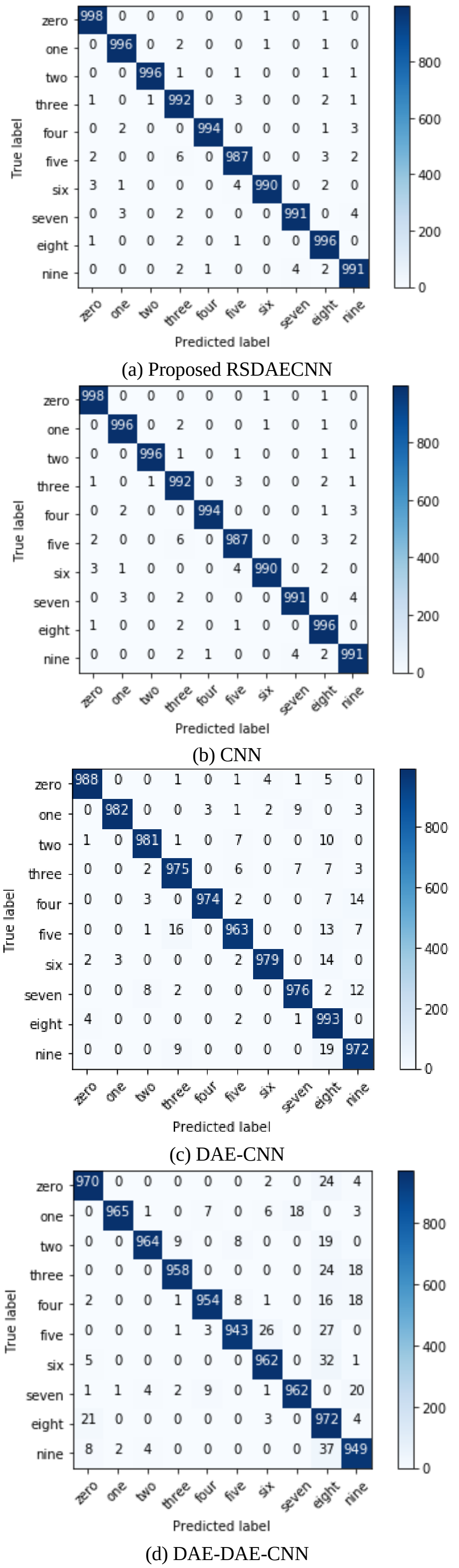
(a) 0% Noise

(b) 10% Noise

(c) 20% Noise

(d) 50% Noise

Fig. 4. Test set recognition accuracy with different noise levels.



	zero	one	two	three	four	five	six	seven	eight	nine
zero	998	0	0	0	0	0	1	0	1	0
one	0	996	0	2	0	0	1	0	1	0
two	0	0	996	1	0	1	0	0	1	1
three	1	0	1	992	0	3	0	0	2	1
four	0	2	0	0	994	0	0	0	1	3
five	2	0	0	6	0	987	0	0	3	2
six	3	1	0	0	0	4	990	0	2	0
seven	0	3	0	2	0	0	0	991	0	4
eight	1	0	0	2	0	1	0	0	996	0
nine	0	0	0	2	1	0	0	4	2	991

(a) Proposed RSDAECNN

	zero	one	two	three	four	five	six	seven	eight	nine
zero	998	0	0	0	0	0	1	0	1	0
one	0	996	0	2	0	0	1	0	1	0
two	0	0	996	1	0	1	0	0	1	1
three	1	0	1	992	0	3	0	0	2	1
four	0	2	0	0	994	0	0	0	1	3
five	2	0	0	6	0	987	0	0	3	2
six	3	1	0	0	0	4	990	0	2	0
seven	0	3	0	2	0	0	0	991	0	4
eight	1	0	0	2	0	1	0	0	996	0
nine	0	0	0	2	1	0	0	4	2	991

(b) CNN

	zero	one	two	three	four	five	six	seven	eight	nine
zero	988	0	0	1	0	1	4	1	5	0
one	0	982	0	0	3	1	2	9	0	3
two	1	0	981	1	0	7	0	0	10	0
three	0	0	2	975	0	6	0	7	7	3
four	0	0	3	0	974	2	0	0	7	14
five	0	0	1	16	0	963	0	0	13	7
six	2	3	0	0	0	2	979	0	14	0
seven	0	0	8	2	0	0	0	976	2	12
eight	4	0	0	0	0	2	0	1	993	0
nine	0	0	0	9	0	0	0	0	19	972

(c) DAE-CNN

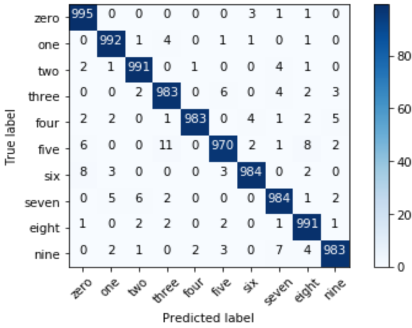
	zero	one	two	three	four	five	six	seven	eight	nine
zero	970	0	0	0	0	0	2	0	24	4
one	0	965	1	0	7	0	6	18	0	3
two	0	0	964	9	0	8	0	0	19	0
three	0	0	0	958	0	0	0	0	24	18
four	2	0	0	1	954	8	1	0	16	18
five	0	0	0	1	3	943	26	0	27	0
six	5	0	0	0	0	0	962	0	32	1
seven	1	1	4	2	9	0	1	962	0	20
eight	21	0	0	0	0	0	3	0	972	4
nine	8	2	4	0	0	0	0	0	37	949

(d) DAE-DAE-CNN

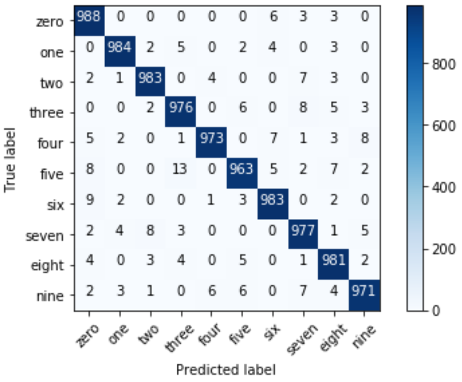
Fig. 5. Confusion matrices of test set with 0% noise for proposed RSDAECNN and individual structures (CNN, DAE-CNN and DAE-DAE-CNN).

Fig. 5 shows the confusion matrixes of test set with 0% noise for the proposed RSDAECNN system along with individual structures after 400 epochs. It clearly observed from the figure for the noiseless case that single CNN (Fig. 5(b)) and proposed model (Fig. 5(a)) performs batter than DAE-CNN and DAE-DAE-CNN achieving a fair accuracy. It is also visible from the figure that all the individual structures (i.e., CNN, DAE-CNN and DAE-DAE-CNN) and the proposed system performed worst for the numeral “5”. Among them DAE-DAE-CNN performs worst misclassifying this numeral 57 times out of 1000 samples. Single CNN, DAE-CNN and the proposed model classifies it correctly for 987, 963 and 987 times accordingly. In case of DAE-DAE-CNN it is noticeable that most of the digits are misclassified as numeral “8”; 179 samples out of 1000 samples are misclassified as numeral “8”. This incident is more frequent for numeral “9”, “6”, “5”, “3” and “2”. The reason behind this incident is that the two layered cascaded DAE is fed with the noiseless images this time, where both the DAEs are trained with 20% noisy data. So, this denoiser deforms the shape of the images even if the images contain no noise which in the end misled the CNN classifier. The scenario is almost same for the DAE-CNN also. Still, proposed model performs well because of using winner-takes-all in the end for final selection process. As this method chooses the one, based on maximum node value, misclassification by two models doesn’t affect the overall performance of the proposed model. The best classification accuracy is found for numeral “0”. The proposed model, CNN, DAE-CNN, DAE-DAE-CNN classify it 998, 998, 988, 970 times accordingly.

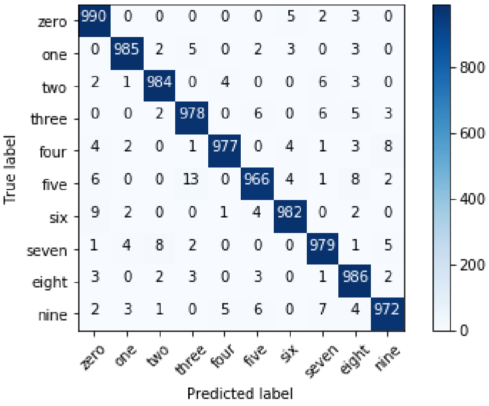
Fig. 6 shows the confusion matrixes of test set with 20% noise for the proposed RSDAECNN system along with individual structures (i.e., CNN, DAE-CNN, DAE-DAE-CNN) after 400 epochs. In such noisy case, all four models performed best for digit ‘0’ and worst for digit ‘5’. The proposed model, CNN, DAE-CNN, and DAE-DAE-CNN classify numeral ‘0’ correctly in 995, 988, 990 and 986 cases out of 1000 cases, respectively. On the other hand, for ‘5’ the true classifications by the methods are 970, 963, 966 and 959 for proposed model, CNN, DAE-CNN, and DAE-DAE-CNN, respectively. On the basis of overall performance, DAE-CNN is the best and DAE-DAE-CNN is the worst among individual structures. In this case CNN performs better than the DAE-DAE-CNN but performance degraded with respect 0% noise case (Fig. 5) architecture which is logical as explained earlier. On the other hand, proposed RSDAECNN is better than best individual structure (i.e., DAE-CNN).



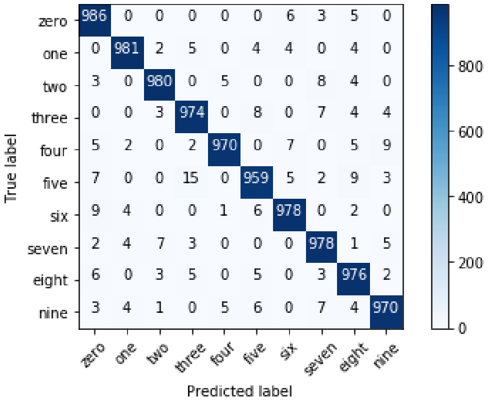
(a) Proposed RSDAECNN



(b) CNN



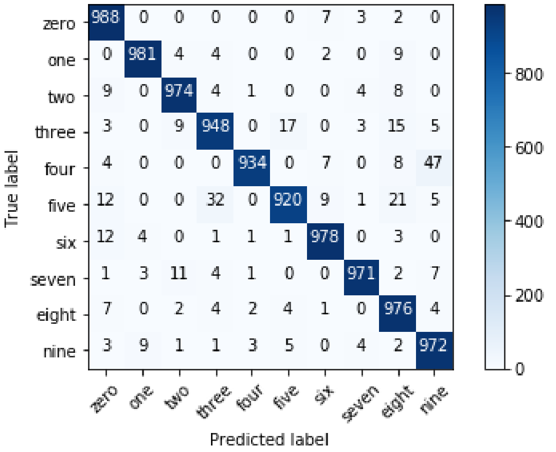
(c) DAE-CNN



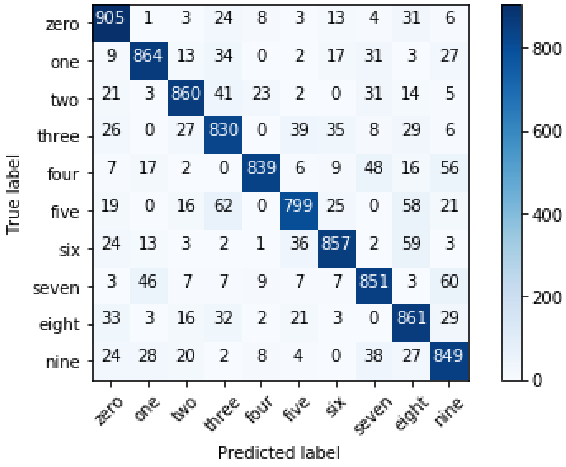
(d) DAE-DAE-CNN

Fig. 6. Confusion matrices of test set with 20% noise for proposed RSDAECNN and individual structures (CNN, DAE-CNN and DAE-DAE-CNN).

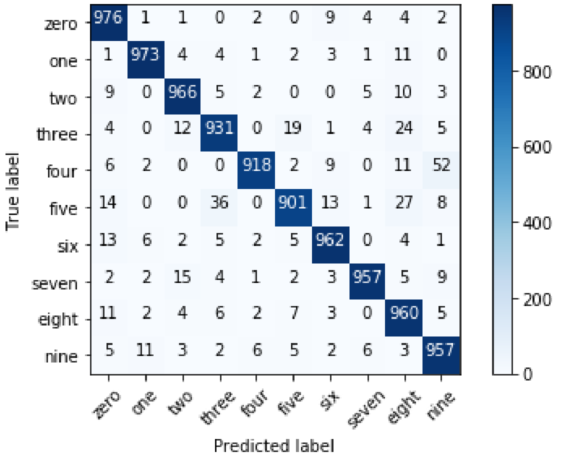
Fig. 7 shows the confusion matrixes of test set with 50% noise for the proposed RSDAECNN system along with individual structures (i.e., CNN, DAE-CNN, and DAE-DAE-CNN) after 400 epochs. This confusion matrix gives the evidence of the fact that proposed model is best suited even if the image is distorted by massive proportion of noise. This time CNN performs the worst; it classifies numerals “1” to “9” correctly only on 864, 860, 830, 839, 799, 857, 851, 861, 849 cases out of 1000 samples for each of them. Apart from classifying numeral “0”, in each and every time its classification accuracy is below 90% and for the numeral “5” its accuracy is even below 80%. In such huge noise, DAE-DAE-CNN is the best among individual structures. On the other hand proposed model classifies the numerals “0” to “9” on 988, 981, 974, 948, 934, 920, 978, 971, 976 and 972 cases accordingly. The performance of the proposed model is comparably better than the DAE-DAE-CNN structure in case of classifying such massive noisy image data. The confusion matrixes presented in Fig. 5 to 7 clearly revealed the effectiveness of the proposed system to work well to classify images with noise free to heavy noise scenario.



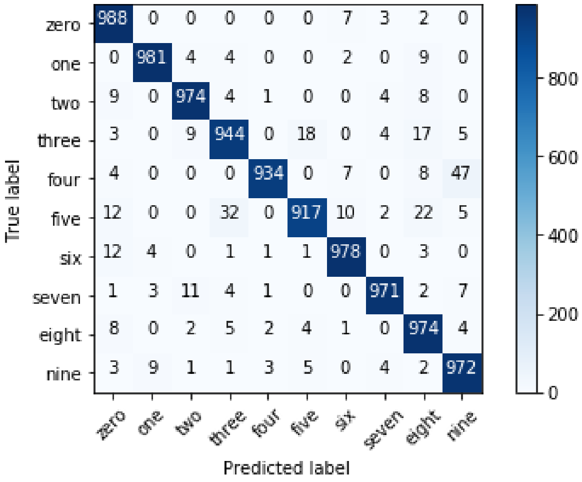
(a) Proposed RSDAECNN



(b) CNN



(c) DAE-CNN



(d) DAE-DAE-CNN

Fig. 7. Confusion matrices of test set with 50% noise for proposed RSDAECNN and individual structures (CNN, DAE-CNN and DAE-DAE-CNN).

TABLE II. A COMPARATIVE DESCRIPTION OF PROPOSED RSDAECNN NOISY IMAGE CLASSIFIER WITH SOME CONTEMPORARY METHODS

The Work Reference		**Classification**	**Noise Level**	**Recognition Accuracy**
Bengio et al. [19]		DBN	0%	98.50%
		Deep net	0%	98.40%
		Shallow net	0%	95.00%
Glorot [45]		Sparse rectifier neural network	25%	98.43%
Vincent et al. [35]		DAE	10%	97.20%
Vincent et al. [36]		SVM	25%	98.37%
		SDAE-3	25%	98.50%
Self-Implemented	CNN	CNN	0%	99.31%
			10%	97.88%
			20%	97.76%
			50%	85.15%
	DAE-CNN [42]	CNN	0%	97.83%
			10%	97.95%
			20%	98.01%
			50%	95.01%
	DAE-DAE-CNN [42]	CNN	0%	95.99%
			10%	97.31%
			20%	97.47%
			50%	96.32%
	Proposed RSDAECNN	CNN	0%	**99.31%**
			10%	**97.98%**
			20%	**98.56%**
			50%	**96.41%**

Table II shows the comparative result analysis of the proposed RSDAECNN model with some other prominent noisy image classifiers. In extent, it describes the particular feature(s) of particular models while classifying noisy images. It is a highly mentionable issue that most of the existing models employ additional feature extraction techniques, whereas, proposed model overcomes the necessity of applying additional feature extraction techniques. The results presented in the table for CNN, DAE-CNN, DAE-DAE-CNN and proposed RSDAECNN are the tabular forms which have already been explained in the previous section. Results of other existing methods are collected from corresponding papers. It is notable that existing methods are tested for different individual noise levels. However, the proposed RSDAECNN has outperformed other models for any noise level. For noise-free case, as an example, Bengio et al. showed accuracy 98.50% and proposed method showed 99.31% accuracy. For 10% noise case, Vincent et al. [35] showed 97.20% accuracy and proposed method showed 97.98%. On the other hand, no existing method presented accuracy for heavy noise (i.e., 50%) and their outcome might be dramatically worse. However, for 50% noise, the performance of proposed method degraded little but outperformed other individual structures CNN, DAE-CNN, and DAE-DAE-CNN. The achieved accuracy for such heavy noisy case is 96.41%. Finally, the results presented in the table clearly revealed the effectiveness of the proposed system for classifying noisy images adulterated with variable level of noise.

IV. Conclusions

Considering real life scenario, it is usual for an image data to be noisy. Pre-processed noiseless images can be classified at ease with the help of existing classification methods. However, for a supervised classifier, it is difficult to deal with the noisy data directly fed to it and failure to classify is quite certain. In this paper, autoencoders are implemented to restore the image from its noisy version and then the reconstructed image is forwarded to a classifier. Another important consideration is that having prior knowledge about the proportion of noise carried by image data is not possible. Keeping all these facts in mind, an innovative model is investigated which includes CNN, DAE-CNN, and DAE-DAE-CNN. This model excludes the necessity to train it for different levels of noise. Being noise independent, the proposed model showed better performance on MNIST dataset compared to other models in terms of classifying images with noises ranging from zero to massive which also ensures its capability of learning hierarchical representations.

Several future research directions are opened from this study. The three-layered architecture investigated in this study is found efficient. Future researches can be conducted by stacking layers with some optimization algorithms to get better performance. Various AEs rather than DAE can also be employed to check whether the image reconstruction process improves or not. Furthermore, proposed model is noise level independent but not noise type independent. The method is tested with images corrupted by only random noise and might perform well for only one type of noise by which it is trained with. To make the system more robust and more applicable in real life scenarios it should be further upgraded so that it would be both noise level independent as well as noise type independent.

References

[1] T. M. Lillesand and R. W. Kiefer, “Remote Sensing and Image Interpretation,” Geological Magazine, vol. 132, issue 2, pp. 248-249, 1995.

[2] M. A. H. Akhand, M. Ahmed, M. H. Rahman and M. M. Islam, “Convolutional Neural Network Training incorporating Rotation based Generated Patterns and Handwritten Numeral Recognition of Major Indian Scripts,” IETE Journal of Research (TIJR), Taylor & Francis, vol. 63, pp. 1-19, 2017.

[3] F. J. Huang, and Y. LeCun, “Large-scale learning with svm and convolutional nets for generic object recognition” Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1-8, 2006.

[4] D. C. Cireşan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, “High-performance neural networks for visual object classification”, arXiv Preprint arXiv:1102.0183, 2011.

[5] D. C. Cireşan, U. Meier, J. Masci, and J. Schmidhuber , “A committee of neural networks for traffic sign classification,” Proceedings of the 2011 International Joint Conference on Neural Networks (IJCNN), pp. 1918-1921, 2011.

[6] J. C. Bezdek, L. O. Hall, and L. Clarke, “Review of MR image segmentation techniques using pattern recognition,” Medical Physics, vol. 20, issue. 4, pp. 1033-1048, 1993.

[7] Y. Bar, I. Diamant, L. Wolf and H. Greenspan, “Deep learning with non-medical training used for chest pathology identification,” Proceedings of Society for Optics and Photonics, pp. 94140V-94140V. doi: 10.1117/12.2083124, 2015.

[8] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, “Subject independent facial expression recognition with robust face detection using a convolutional neural network,” Neural Networks, vol. 16, issue 5, pp. 555-559, 2003.

[9] H. Bourlard, and Y. Kamp, “Auto-association by multilayer perceptrons and singular value decomposition,” Biological Cybernetics, vol. 59, issue 4, pp. 291-294, 1988.

[10] Y. Bengio, “Learning deep architectures for AI. Foundations and trends® in Machine Learning,” vol. 2, issue 1, pp. 1-127, 2009.

[11] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” Parallel distributed processing: explorations in the microstructure of cognition, vol. 1, pp. 318-362, 1986.

[12] H. C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, “Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, issue 8, pp. 1930-1943, 2013.

[13] M. Norouzi, M. Ranjbar, and G. Mori, “Stacks of convolutional restricted boltzmann machines for shift-invariant feature learning,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVRP), pp. 2735-2742, 2009.

[14] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations,” Proceedings of the 26th Annual International Conference on Machine Learning, pp. 609-616, 2009.

[15] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” Neural Computation, vol. 18, issue 7, pp. 1527-1554, 2006.

[16] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, “Deconvolutional networks,” Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2528-2535, 2010.

[17] M. D. Zeiler, G. W. Taylor, and R. Fergus, “Adaptive deconvolutional networks for mid and high-level feature learning,” Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2018-2025, 2011.

[18] B. A. Olshausen and D. J. Field, “Sparse coding with an overcomplete basis set: A strategy employed by V1?,” Vision Research, vol. 37 issue 23, pp. 3311-3325, 1997.

[19] Y. Bengio, P. Lamblin, D. Popovici and H. Larochelle, “Greedy layer-wise training of deep networks” , In Proc of Advances in 19 th neural information processing systems, pp. 153-160, Dec 2007.

[20] S. Behnke, “Hierarchical Neural Networks for Image Interpretation”, volume 2766 of Lecture Notes in Computer Science. Sprnger, 2003.

[21] D. Scherer, A. Müller, S. Behnke, “Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition,” 20th International Conference on Artificial Neural Networks (ICANN), Thessaloniki, Greece, Springer. pp. 92–101, 2010.

[22] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks”, in Proc. Neural Information Processing Systems, pp. 1097–1105, 2012.

[23] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN features off-the-shelf: an astounding baseline for recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806-813, 2014.

[24] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, “Intriguing properties of neural networks,” arXiv Preprint arXiv:1312.6199, 2013.

[25] D. Lu, and Q. Weng, “A survey of image classification methods and techniques for improving classification performance.” International Journal of Remote Sensing, vol. 28, issue 5, 823-870, 2007.

[26] M.C. Motwani, M.C. Gadiya, R.C. Motwani, F.C. Harris, “Survey of Image Denoising Techniques”, Proc. of GSP 2004, Santa Clara, CA, pp. 27-30, 2004.

[27] R. R. Coifman, and D. L. Donoho, “Translation-invariant de-noising,” Wavelets and Statistics, vol. 103, pp. 125-150, 1995.

[28] P. Perona, and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, issue7, pp. 629-639, 1990.

[29] L. I. Rudin, and S. Osher, “Total variation based image restoration with free local constraints,” Proceedings of the IEEE International Conference on Image Processing, vol. 1, pp. 31-35, 1994.

[30] O. Subakan, B. Jian, B. C., Vemuri and C. E. Vallejos, “Feature preserving image smoothing using a continuous mixture of tensors,” Proceedings of the 11th International Conference on Computer Vision (ICCV), pp. 1-6, 2007.

[31] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” IEEE Transactions on Image Processing, vol. 15, issue 12, 3736-3745, 2006.

[32] J. Mairal, F., Bach, J. Ponce, and G. Sapiro, “Online dictionary learning for sparse coding”, Proceedings of the 26th Annual International Conference on Machine Learning, pp. 689-696, 2009.

[33] A. K. Singh, V. P. Shukla, S. R Biradar., and S. Tiwari, “Multiclass Noisy Image Classification Based on Optimal Threshold and Neighboring Window Denoising.” International Journal of Computer Engineering Science (IJCES), vol. 4, issue 3, pp. 1-11, 2014.

[34] Cheema, T.A., I. Qureshi and M. Naveed A., “Blur and Image Restoration of Nonlinearly Degraded Images Using Neural Networks Based on Nonlinear ARMA Model”, Proc. INMIC, pp. 102-107.

[35] P. Vincent, H. Larochelle, Y. Bengio, and P. A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” Proceedings of the 25th International Conference on Machine Learning, 1096-1103, 2008.

[36] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.A. Manzagol, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,” The Journal of Machine Learning Research, vol. 11, issue 3371–3408, 2010.

[37] F. Agostinelli,M. R. Anderson, and H. Lee, “Adaptive multi-column deep neural networks with application to robust image denoising,” Proceedings of the Advances in Neural Information Processing Systems, pp. 1493-1501, 2013.

[38] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, , ”Stacked convolutional auto-encoders for hierarchical feature extraction.” , In Proc. International Conference on Artificial Neural Networks. Springer Berlin Heidelberg, pp. 52-59, 2011.

[39] L. Gondara, “Medical image denoising using convolutional denoising autoencoders,” Proceedings of the 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pp. 241-246, 2016.

[40] L. Xu, J. S. Ren, C. Liu, and J. Jia, “Deep convolutional neural network for image deconvolution,” Proceedings of the Advances in Neural Information Processing Systems, pp. 1790-1798, 2014.

[41] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, “Stacked convolutional denoising auto-encoders for feature representation,” IEEE Transactions on Cybernetics, vol. 47, issue 4, pp. 1017-1027, 2017.

[42] S. S. Roy, M. Ahmed and M. A. H. Akhand, “Classsification of massive noisy image using auto-encoders and convolutional neural network,” 2017 8th International Conference on Information Technology (ICIT), pp. 971-979, 2017.

[43] M. A. H. Akhand, and K. Murase, “Ensembles of Neural Networks based on the Alteration of Input Feature Values,” International Journal of Neural Systems, vol. 22, issue 1, pp. 77-87, 2012.

[44] M. A. H. Akhand, Md. Monirul Islam, and K. Murase, “A Comparative Study of Data Sampling Techniques for Constructing Neural Network Ensembles,” International Journal of Neural Systems, vol. 19, issue 2, pp. 67-89, 2009.

[45] X. Glorot, A. Bordes, and Y. Bengio. ”Deep Sparse Rectifier Neural Networks.” , In Proc of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11). Vol. 15. pp. 315-323, 2011.