Study on Denoising Variational Auto-encoders

Prof.Jajati Keshari Sahoo[1] and Rusafid Mirza Pottengad[2]

[1]Department of Mathematics, BITS Pilani K.K Birla Goa Campus
[2]Department of Mathematics, BITS Pilani K.K Birla Goa Campus

***Abstract*—Removing noise from the original image is still a challenging problem for researchers. There have been several published algorithms and each approach has its assumptions, advantages, and limitations.In this paper we investigate the ability of a particular class of auto-encoders known as variational autoencoders to denoise manually corrupted images.We will be using the MNIST handwritten digit dataset for our work.**

***Index Terms*—Deep Learning, Convolution neural networks, GPU, Image Denoising**

I. Introduction

An autoencoder is a type of artificial neural network used to learn efficient data encodings in an unsupervised manner The aim of an autoencoder is to learn a representation (encoding) for a set of data, typically for dimensionality reduction, by training the network to ignore signal noise. Along with the reduction side, a reconstructing side is learnt, where the autoencoder tries to generate from the reduced encoding a representation as close as possible to its original input, hence its name. Recently, the autoencoder concept has become more widely used for learning generative models of data.

Variational autoencoder models inherit the autoencoder architecture, but make strong assumptions concerning the distribution of latent variables. They use a variational approach for latent representation learning, which results in an additional loss component and a specific estimator for the training algorithm called the Stochastic Gradient Variational Bayes (SGVB) estimator. Varoiational autoencoders are powerful generative models.We study the reconstructive ability of a variational autoencoder comprised of fully connected linear layers and how it improves the classification accuracy of manually corrupted MNIST dataset.

II. BACKGROUND

A. Non-Deep Learning Methods

The image denoising technology is one of the most important branches of image processing technologies and is used as an example to show the development of the image processing technologies in last 20 years [1]. Buades et al. [2] proposed a non-local algorith mmethod to deal with image denoising. Lan et al. [3] fused the belief propagation inference method and Markov Random Fields (MRFs) to address image denoising. Dabov et al. [4] proposed to transform grouping similar two-dimensional im-age fragments into three-dimensional data arrays to improve sparisty for image denoising.These conventional methods produced amazing results but suffered from two challenges.One was these methods were non convex hence we have to manually set the parameters, Second they refer to a convex optimization problem in the test stage requiring high computational power.

B. Deep Learning Methods

In recent years researchers have found that deep learning methods with deeper architectures can automatically learn and find more suitable image features than the traditional methods [5].Big data and GPU are also essential to improve their learning abilities[6].Convolutional neural network (CNN) is one of the most typical and successful deep learning network for image processing [7]. CNN was originated LeNet from 1998 and it was successfully used in hand-written digit recognition, achieving excellent performance [8]. DiracNets [9], IndRNN [10] andvariational U-Net [11] also provide us with many competitive technologies for image processing. These deep networks are also widely applied in image denoising, which is the branch of image processing technologies. For example, the combination of kernel-prediction net and CNN is used to obtain denoised image[12]. BMCNN utilizes NSS and CNN to deal with image denoising [13]. GAN is used to remove noise from noisy image [14].

III. Proposed Method

A. Datasets

We have performed our study on the MNIST handwritten dataset which consists of 60000 28x28 grey scale images of various handwritten digits labelled from 0-9. We have manually corrupted each image in the dataset by adding noise from gaussian distribution of mean 0 and a range of values for standard deviation as observed in Fig.1.

B. Models

1) Classification Model: We have defined a classification model to classify the digits in to their respective labels using the architecture shown in the Fig.2. The classifier was trained with the negative log likelihood function given by the equation (1), where y(i) is the predicted output and n is total number of instances in the batch

$$L = \frac{1}{n}\sum_{i=1}^{n}\left[ln(y_i)\right] \tag{1}$$

.

2) Generative Model : We have defined a variational autoencoder model for image to image translation using the following architecture given in the Fig.3. The VAE(variational autoencoder) was trained with KL divergence loss coupled with binary cross entropy loss.The KL divergence between two probability distributions simply measures how much they diverge from each other. Minimizing the KL divergence here means optimizing the probability distribution parameters to closely resemble that of the target distribution

$$L = \frac{1}{m}\sum_{k=1}^{m}\left[t_k ln(y_k) + (1-t_k) ln(1-y_k)\right] \tag{2}$$

$$D_{KL} = \frac{1}{2}\sum_{k}\left(exp(\Sigma(X)) + \mu^2(X) - 1 - \Sigma(X)\right) \tag{3}$$

Equation (2) is the binary cross entropy loss where y(k) is the predicted output and t(k) is the true output and m is number of instances.In equation (3) X is the data we want to model and k is summed up over all instances .

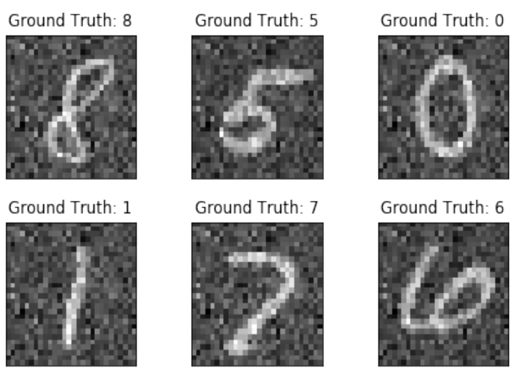

Fig. 1. Images after adding gaussian noise.

IV. Experiment and Results

The dataset was normalized by using min max normalization of minimum value of 0 and maximum value of 1.It was trained on the classification model giving an accuracy of 96 percent. The dataset was then manually corrupted using gaussian noise of mean 0 and standard deviation of values in the range between 0.01 and 1.The Fig.1 shows us the images after corruption by gaussian noise of 0.2 standard deviation.The classification model was then run on this dataset and its results were plotted against standard deviation of gaussian noise.

The VAE(variational auto-encoder) model was trained on the original uncorrupted images to reconstruct it back from the VAEs(variational auto-encoder) encoded dimension.The corrupted images were then fed into the VAE(variational auto-encoder) which attempted to denoise the images and the denoised images then were fed to the classification model and its results were plotted against standard deviation of gaussian noise.

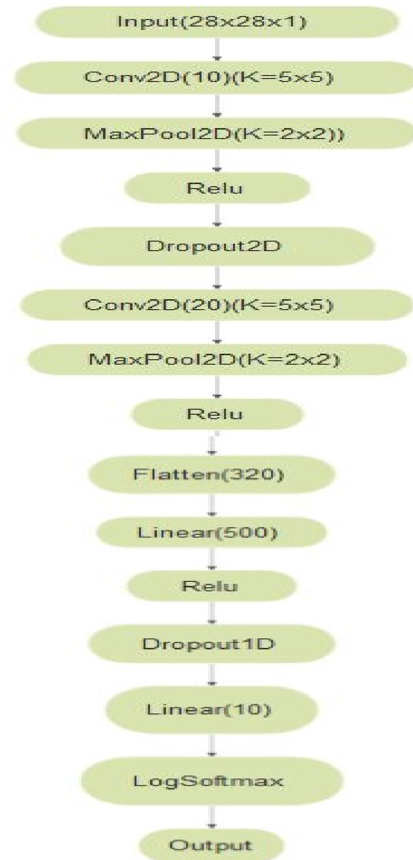

Fig. 2. Architecture of classification model.

V. Conclusion

The reconstructed images and noisy images were giving almost the same classification accuracy when the noise factor(standard deviation of gaussian noise) was below 0.1. From 0.15 to 0.25 the reconstructed images gave better classification accuracy as compared to the noisy images.beyond that the reconstructed images performs worse in classification as compared to noisy images.The results are plotted in Fig.4.

Acknowledgment

This work was supported by the supervisor of the course MATH F376 Prof.Jajati Keshari Sahoo.

References

[1] Jun Xu, Lei Zhang, Wangmeng Zuo, David Zhang, and Xiangchu Feng. Patch group based nonlocal self-similarity prior learning for image denoising. In Proceedings of the IEEE international conference on computer vision, pages 244252, 2015.

[2] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 6065. IEEE, 2005.

[3] Xiangyang Lan, Stefan Roth, Daniel Huttenlocher, and Michael J Black. Efficient belief propagation with learned higher-order markov random fields. In European conference on computer vision, pages 269282. Springer, 2006.

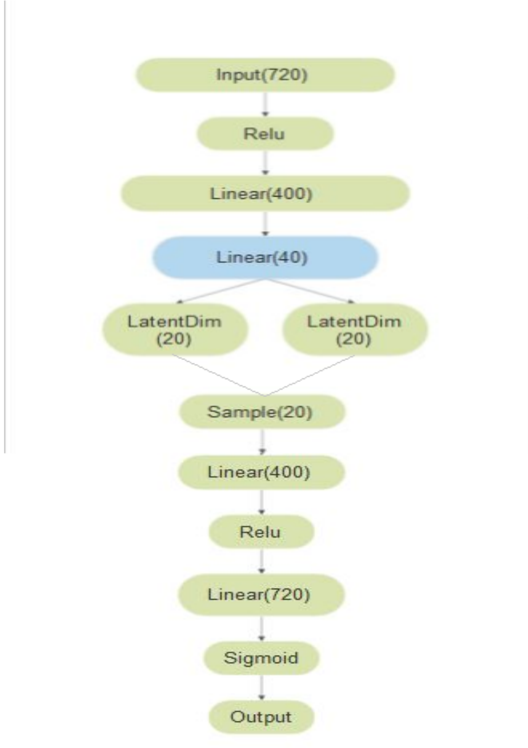

Fig. 3. Architecture of variational auto-encoder model.

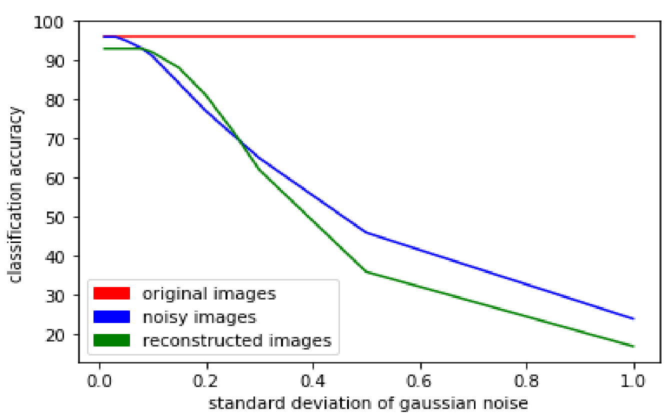

Fig. 4. Classification accuracy vs standard deviation of guassian noise.

[4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Im-age denoising by sparse 3-d transform-domain collaborative filtering.IEEE Trans-actions on image processing, 16(8):20802095, 2007.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information pro-cessing systems, pages 10971105, 2012.

[6] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip Torr. Deeply supervised salient object detection with short connections. In 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pages53005309. IEEE, 2017.

[7] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: A convolutional neural-network approach.IEEE transactions on neural networks, 8(1):98113, 1997.

[8] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard EHoward, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to hand written zip code recognition.Neural computation, 1(4):541551, 1989.

[9] Sergey Zagoruyko and Nikos Komodakis. Diracnets: training very deep neural networks without skip-connections.arXiv preprint arXiv:1706.00388, 2017.

[10] Shuai Li, Wanqing Li, Chris Cook, Ce Zhu, and Yanbo Gao. Independently recur-rent neural network (indrnn): Building a longer and deeper rnn. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 54575466, 2018.

[11] Patrick Esser, Ekaterina Sutter, and Bj orn Ommer. A variational u-net for conditional appearance and shape generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 88578866, 2018.

[12] Steve Bako, Thijs Vogels, Brian McWilliams, Mark Meyer, Jan Nov ak, Alex Harvill, Pradeep Sen, Tony Derose, and Fabrice Rousselle. Kernel-predicting convolutional networks for denoising monte carlo renderings.ACM Trans. Graph,36(4):97, 2017.

[13] Byeongyong Ahn and Nam Ik Cho. Block-matching convolutional neural network for image denoising.arXiv preprint arXiv:1704.00524, 2017.

[14] Subarna Tripathi, Zachary C Lipton, and Truong Q Nguyen.Correction by projection: Denoising images with generative adversarial networks.arXiv preprint arXiv:1803.04477, 2018.