Image Generation using Variational Autoencoders

Purnima Sai Koumudi Panguluri, Kishore Kumar Kamarajugadda

Abstract**: *The proposed system generates new images from the existing images using variational autoencoders. The autoencoder aims to map the input image to a multivariate normal distribution in the latent space. Variational autoencoder transforms input image into a remarkable output by reducing the reconstruction and KL divergence losses. The primary advantage of implementing variational autoencoder over the other autoencoders is that it follows a specific probability distribution called Gaussian distribution and results in generating high quality images.***

Keywords: Variational Autoencoders, Principal Component Analysis, Orthogonal Transformation, KL Divergence, Encoder and Decoder.

I. INTRODUCTION

Over the recent years, deep learning research has evolved remarkably in industrial and trading domains. The state-of-the-art techniques escalated the possibility of interpreting many more complicated AI tasks. The end to end problem solving techniques like image classification, speech recognition, music generation acquired much popularity due to the supremacy in terms of accuracy. It’s potential to extract high-level features from the input in a progressive approach, abolishes the demand for domain expertise and dedicated feature extraction. But from the past few years, a rapid accomplishment of autoencoders has been perceived. Gradients calculated by backpropagation and mini-batch are used to train these exceptional instances of feed-forward non-recurrent neural networks. Autoencoders vigorously applied one-class classification in various applications [1]. The intrinsic purpose is that the weights that are procured from stacked autoencoders can be employed for deep neural network models instead of using randomly initialized weights. Additionally, variational autoencoders are used to generate images as well as consequential binary semantic hash codes for text documents. Ever since 2013, the progress of autoencoders favored predominantly. Autoencoders are unsupervised learning algorithms and play a significant role in deep learning architectures. They are the elementary parts of neural networks which notably contain two parts, one being the encoder and the other, the decoder. The primary objective of the encoder is to take an input image, pass the image through a series of convolutional layers and compress the input into lower dimensional vectors. The lower dimensional vector is then decompressed by the decoder to produce an image of the original dimension. The salient feature to note is that the decoder uses the convolutional transpose layers that doubles the size of the input tensor. Autoencoders are more powerful and malleable than linear factor analysis methods and principal component analysis because of non-linear dimensionality reduction i.e., they reduce both linear and non-linear data. The weights that are produced during the training process, minimizes the error between the reconstructed and the native image. Inherently, encoder comprehends to avoid the random noise resulting in lesser noise generation. One of the interesting problems in autoencoders is image generation. Autoencoders directly map the input image to a single point in the latent vector space resulting in non-uniform distribution of points. Encoder attempts to map *X* to *u*, i.e., f*(x)* → *u* and decoder maps the output of the encoder to the original image, i.e., f*(u)* → *X*. Some of the innate applications of autoencoders are image generation, denoising the data, dimensionality reduction for visualization, image compression, feature extraction, recommendation systems and many more.

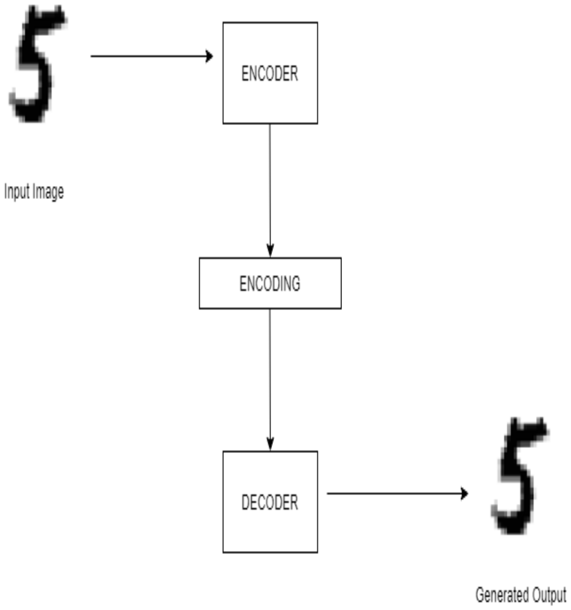

Fig. 1. Autoencoder Diagram

Revised Manuscript Received on March 30, 2020.
***** Correspondence Author

*** Purnima Sai Koumudi Panguluri**, Department of CSE, Faculty of Science and Technology, ICFAI Foundation for Higher Education, Hyderabad, India, pangulurikoumudi@gmail.com

Kishore Kumar Kamarajugadda, Department of ECE, Faculty of Science and Technology, ICFAI Foundation for Higher Education, Hyderabad, India, kkishore@ifheindia.org

However, an essential hindrance caused by the autoencoders in image generation is that there is no conventional distribution of the data i.e., distribution of the data is uncertain resulting in inadequate formation of images. The Variational Autoencoder endures the complication by attaining a probability distribution called Gaussian distribution. Unlike autoencoder, variational autoencoder attempts to map each of an input image to a multivariate gaussian distribution about a point in a latent vector space. Hence, we recommend using variational autoencoders for image generation which can map the given input to a joint normal distribution. In our problem, we use human faces.

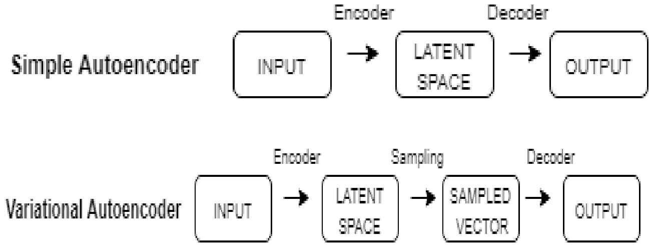

Fig. 2. Difference between Simple Autoencoder and Variational Autoencoder

II. LITERATURE SURVEY

A. Orthogonal Variation

Dimensionality reduction is the obligatory technique while dealing with image datasets. There are a number of dimensionality reduction algorithms like multidimensional scaling, PCA, t-Distributed Stochastic Neighbor Embedding, linear discriminant analysis and many more. Principal component analysis is the predictable and traditional dimensionality reduction technique implemented in machine learning and deep learning applications.

The prevalent orthogonal transformation used in PCA convert higher dimensional set of features into lower dimensional vectors i.e., linear composition of corresponding variables into uncorrelated variables [2]. The orthogonal projections operate with eigenvalues, eigenvectors and these when applied to set of faces produce eigenfaces which are used for facial recognition and reconstruction systems. The limitation of PCA is the linear non-correspondence of features as the transformations are orthogonal and hence inadequate to reinforce with non-linear representation of data [3].

B. Fuzzy Image Formation

Auto encoders are comprehensible neural network structures that avails backpropagation and are trained through the fundamental miscellaneous algorithms from Restricted Boltzmann Machines to more standard traditional techniques like stochastic gradient descent. Besides the activation function, autoencoder and PCA are analogous to each other [4]. Typically, autoencoders exploit non-linear activation functions like Rectified linear unit and sigmoid functions. Activation function potentially used by the encoders are given by,

$$z = \sigma\left(Wx + b\right) \tag{1}$$

Correspondingly, the decoder takes the output of the encoder and decodes the image using the following equation,

$$x^{'} = \sigma^{'}\left(W^{'}z + b^{'}\right) \tag{2}$$

The loss function required to train the autoencoders through standard backpropagation mechanism is given by,

$$\Delta(x, x^{'}) = \| \, x = x^{'} \, \|^{2} = \| \, x - \sigma^{'}\left(W^{'}(z) + b^{'}\right) \|^{2} \tag{3}$$

However, discontinuity and discrete latent space resulted in obscure image formations.

C. Perpetual Latent Space

Unlike autoencoder, the primitive isolated property of variational autoencoder is the continuous latent space that permits easier interpolation i.e., the decoder confirms that the points residing in the same community produce similar images [5]. The probability distribution allows encoder to output a vector of two variables i.e., mean and standard deviation vectors which help in data modelling [6]. Thus, it generates lucid images by sampling the distribution

III. DATASET

The dataset taken here is the Flickr-Faces-HQ (FFHQ) dataset [7] which is comprised of 70,000 high-quality images at a resolution of 1024X1024. The images having a satisfactory coverage of accessories such as eyeglasses, hats and many more were gathered under non-restrictive licenses. The following figure proclaims some of the examples of faces taken from FFHQ dataset.

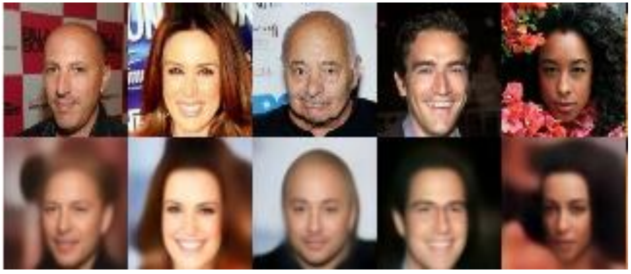

Fig. 3. Examples of faces from FFHQ Face Database

IV. PROPOSED METHOD

The unsupervised learning used for the given problem is Variational Autoencoder commonly known as VAE. Initially variational autoencoder encodes the given input and maps it to a mean and a variance vector. The point is then sampled from this probability distribution and is decoded to evaluate the reconstruction error. Due to the gaussian distribution, VAE generates flattened depiction of input. The correlation does not exist between any two dimensions and hence the covariance matrix configured is diagonal.

The logarithm of the variance ranges from (-∞,∞) and the value of epsilon is sampled from normal distribution. The point z in the latent space is produced from the following equation,

$$z = \mu + exp(log(\sigma^{2}) * \varepsilon) \tag{4}$$

A. The Encoder

The implementation of encoder uses 4 piled convolutional 2D layers sequentially followed by Leaky RELU activation function. The Leaky RELU used here is to implant dying RELU problem encountered [8]. These convolutional 2D layers incorporated with odd filters are applied to the designated input to originate a feature map. They capture progressive exemplary features of input. After every single convolutional layer [9], batch normalization technique is implemented to regulate, fine-tune and scale parameters in order to accelerate the learning process. Eventually, the last layer is flattened and coupled with the dense layers of mean and log variance to generate two-dimensional latent space.

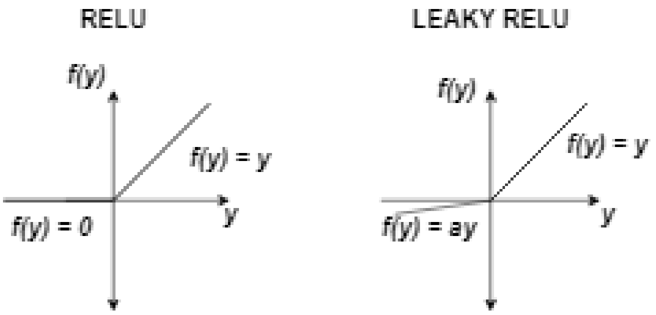

Fig. 4. RELU Vs Leaky RELU

B. The Decoder

The decoder attains the reshaped tensor from the encoder and comprises 4 stacked convolutional transpose layers. These layers perform up-sampling in an optimal pattern and contain reasonable weights. Rear layers adhere with fully connected i.e., the dense layers and decodes it into the primary image domain.

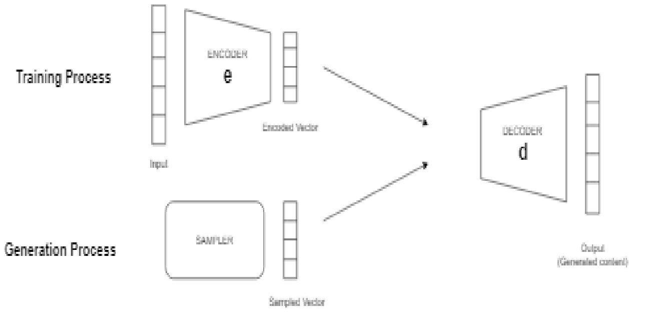

Fig. 5. Encoder and Decoder in Variational Autoencoder

C. The Loss

Diminishing the loss obtained is one of the pivotal factors in deep and machine learning models. Some of the commonly used loss functions are Mean Squared Error, Mean Absolute Error and cross Entropy Loss etc. Variational autoencoders follow a particular category of loss function called Kullback-Leibler divergence, popularly known as KL divergence [10]. KL divergence estimates the divergence between any two probability distributions. KL divergence contracts the information loss while dealing with imprecise distributions [11]. Hence KL loss is directly proportional to mean and log variance i.e., loss becomes 0 when mean and log variance reaches to 0. The loss function used is the combination of KL divergence loss and reconstruction loss which is measured using r_loss_factor. The key factor to note is that by increasing the reconstruction loss [12], the regulatory consequence of KL loss decreases and by reducing the reconstruction loss, we acquire poorly formed images.

$$D_{KL}(a||b) = \sum_{i=1}^{N} a(x_i) \cdot (log\, a(x_i) - log\, b(x_i))$$

(5)

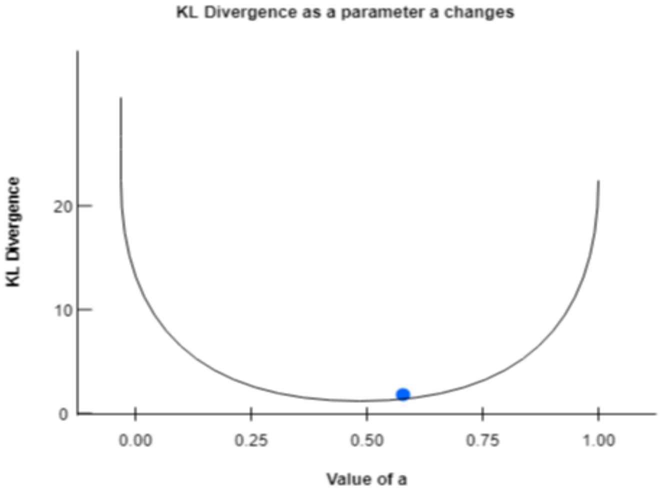

Fig. 6. KL Divergence Loss Graph

V. TRAINING AND EVALUATION

Distribution of Points- Contrary to standard autoencoder, variational autoencoder follows a probability distribution called normal distribution. The impact of KL divergence decreases if the dimensions significantly differ from standard normal distribution. Hence to make them identical, the reconstruction loss has to be minimized.

Training Details- The reconstruction loss is implemented using mean squared error and KL loss is computed at the end. The aggregate loss obtained is determined by taking the average of reconstruction and KL loss values. The optimizer used during the compilation process is Adam with a learning rate of 0.0005. Ultimately, the model was trained for 30 epochs with a batch size of 150.

Latent Space Arithmetic- Arithmetic operations on vectors can be computed after encoding the given input. Latent space arithmetic aids in converting one type of attribute into another i.e., distinct versions of faces can be formed with a relevant variation in feature. Vector arithmetic in latent space can be calculated using the expression given below,

$$w = w_{new} + (\alpha * \eta)$$

(6)

VI. RESULTS

Figure 7 shows the output that is produced after passing through the encoder and the decoder. The vector arithmetic helped in attaining contemporary outcomes such as face morphing and feature exploitation. Figure 8 shows the KL divergence loss graph. The KL divergence decreases gradually and produces finer images.

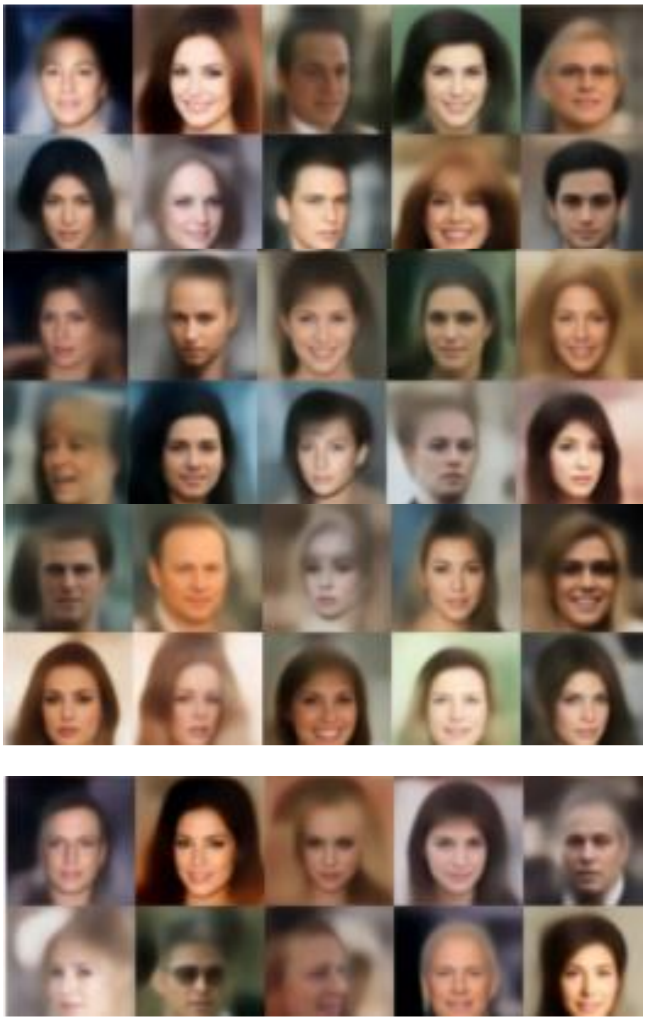

Fig 7. Sample output images generated

The reconstructed image formed after applying variational autoencoder is so satisfying and firmly resembles the original input. This signifies the capability of variational autoencoders for capturing and fine-tuning high level features in images.

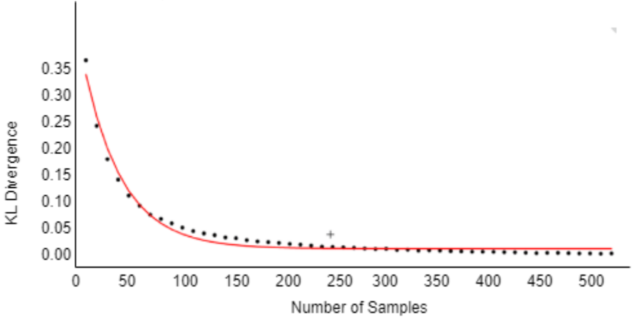

Fig 8. Example of KL Divergence Graph Vs Samples

VI. CONCLUSION

With the increase in the complexity of images, the output generated by standard autoencoders are unfocused due to the non-conventional distribution. Moreover, the latent space of standard autoencoder is uncertain and results in over fitting. In the proposed paper, variational autoencoders generated new faces from the existing human faces by simply selecting the points from normal distribution. The entire process took place by training 2 models (the encoder and the decoder). The encoder uses convolutional layers whereas the decoder decodes using convolutional transpose layers. The latent space is uniform and the training of variational autoencoders is regularized to avoid overfitting. The losses i.e., KL divergence loss and reconstruction loss are minimized during the training process, thus resulting in the generation of good quality images. Additionally, the vector arithmetic of latent space helped in converting one type of attribute into another.

REFERENCES

1. Baldi, Pierre. "Autoencoders, unsupervised learning, and deep architectures." Proceedings of ICML workshop on unsupervised and transfer learning. 2012.
2. Geladi, Paul, et al. "Principal component analysis of multivariate images." Chemometrics and Intelligent Laboratory Systems 5.3 (1989): 209-220.
3. K.Kishore Kumar and P.Trinatha Rao," Age-Invariant Face Recognition using Multiple Descriptors along with Modified Dimensionality Reduction Approach", Multimedia Tools and Applications, Springer,Volume 78, Issue 19,PP 27639 -27661,June 2019.
4. Wang, Yasi, Hongxun Yao, and Sicheng Zhao. "Auto-encoder based dimensionality reduction." Neurocomputing 184 (2016): 232-242.
5. Doersch, Carl. "Tutorial on variational autoencoders." arXiv preprint arXiv:1606.05908 (2016).
6. K.Kishore Kumar and P.Trinatha Rao," Extract features from the Periocular region to identify the Age using Machine Learning Algorithms," Journal of Medical Systems, Springer, Volume 43, Issue 196, PP 1-15,May 2019.
7. Zhu, Jiapeng, Deli Zhao, and Bo Zhang. "LIA: Latently Invertible Autoencoder with Adversarial Learning." arXiv preprint arXiv:1906.08090 (2019).
8. Xu, Bing, et al. "Empirical evaluation of rectified activations in convolutional network." arXiv preprint arXiv:1505.00853 (2015).
9. O'Shea, Keiron, and Ryan Nash. "An introduction to convolutional neural networks." arXiv preprint arXiv:1511.08458 (2015).
10. Chen, Xi, et al. "Variational lossy autoencoder." arXiv preprint arXiv:1611.02731 (2016).
11. Vidyasagar, Mathukumalli. "Kullback-Leibler divergence rate between probability distributions on sets of different cardinalities." 49th IEEE Conference on Decision and Control (CDC). IEEE, 2010.
12. K.AbhiroopTejomay and K.Kishore Kumar, "Sketch to Photo Conversion using Cycle-Consistent Adversarial Networks", International Journal of Innovative Technology and Exploring Engineering, Volume 9, Issue 4, PP 2467-2471, February 2020.

AUTHORS PROFILE

Purnima Sai Koumudi Panguluri, graduated from ICFAI Foundation of Higher Education with a Bachelor of Technology degree in the field of Computer Science and Engineering. She is an IEEE student member

Kishore Kumar is Chairperson – Academic Instruction, Faculty of Science and Technology, ICFAI Foundation of Higher Education, Hyderabad, India.