A MULTIMODAL APPROACH FOR AN APPLICATION OF HAND WRITTEN CHARACTER RECOGNITION

Dr.K.Selvani Deepthi

Associate Professor,Deparment of Computer Science and Engineering,

Anil Neerukonda Institute of Technology and Sciences (Autonomous),

Visakhapatnam, India.

Email:selvanideepthi.cse@anits.edu.in

B.Rajesh

Assistant Professor, Department of Computer Science and Engineering,

GIT, GITAM (Deemed University), Visakhapatnam, India.

Email:rajeshit1236@gmail.com

S V S S Lakshmi

Assistant Professor, Department of Computer Science and Engineering,

Anil Neerukonda Institute of Technology and Sciences (Autonomous),

Visakhapatnam, India.

Email:svsslakshmi.cse@anits.edu.in

Abstract: Over a decade of research topics, Deep learning is a prominent domain that has complex categories such as image synthesis and classification. Generative Adversarial Networks is one of the methods in deep learning that are essentially two networks which compete against each other to win a game. In 2014 Generative Adversarial Network (GAN) has been introduced to support various applications such as computer vision and natural language processing to achieve efficient performance. Over the past few years GAN has brought a potential growth for research. The author has studied various applications of GAN, out of which image synthesis is well learned. The author has recognized a drawback that images which are synthesized using GAN have low resolution. Hence, a Convolution neural network was proposed to obtain high resolution images. The main objective of this paper is to create new images from the existing images which will be helpful when the data set is of small size and it reduces the time in creating the anime pictures for the designers .By experimental results, the author has identified that Convolution Neural Network produced better clarity images with less number of epochs compared to artificial neural networks.

Keywords: Convolutional neural network (CNN), Deep learning, Generative adversarial network(GAN)

I. INTRODUCTION

Generative Adversarial Networks abbreviated as GAN uses Deep Learning techniques. Deep Learning is a sub field of AI. Deep Learning methods help computer to recognize things as a human does[1][8][19]. GAN's primarily consists of two neural networks namely Generator and Discriminator. Generator creates new data from the existing ones and the discriminator identifies whether pictures are fake or real .If yes the generator tries to create new images until discriminator fails to recognize. This is the actual conflict that is present between Generator and Discriminator. There are many number of applications using GAN's such as Text to image , noise removal in images , Facebook uses GAN methods to replace closed eyes of a person with his own eyes in another picture, medical field etc[4]. GAN plays a major role in generating new images from the existing ones thus reducing human time and effort. In GAN, both the networks need to be optimized and also function simultaneously. Training GAN is the most difficult part and many researches are going on to make this process easy. So that the networks will find a point where both will have minimum loss functions and do their job perfectly. In the adversarial net framework, the generative model is made to observe the dataset thoroughly, a discriminative model learns to distinguish whether the sample is from the original data or not. Here the generative model can be thought of as similar to a team of counterfeiters, trying to generate fake currency and without detection. The discriminative model is similar to the police who detect the counterfeit currency [9]. The networks of generator and discriminator are typically implemented by multi-layer networks which are either convolution or fully-connected layers .These models try to find all the statistical distributions of training data, thereby helping in creating or synthesizing new samples from the learned distributions. Adversarial training is complicated as both generator and discriminator needs to get optimized simultaneously and hence GAN's are very unstable. There will be many cases where generator fails in understanding the distribution of the original data which user wants.

There are two main threads of research on GAN.

1) Generator
2) Discriminator

Generator:
The task of Generator is to generate fake data i.e., n from real data set and train the discriminator to correctly predict them as real. The value of n can be any natural number between 1 and infinity.
Discriminator: The task of Discriminator Network is to take input either from the real data or from the generator and try to predict whether the input is real or generated[3][14][15][19].

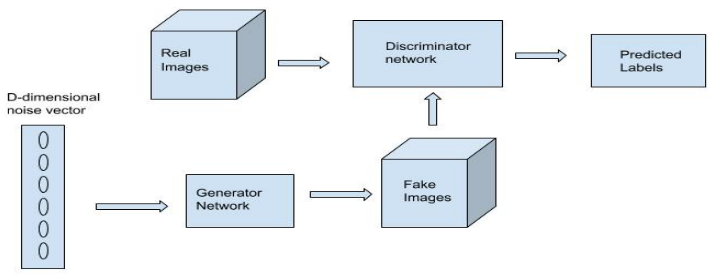

Figure 1: Generative Adversarial Network Architecture

The rest of paper is organized as follows: Section II describes the literature survey related to GAN .Section III describes the proposed system and section IV presents the experiment results followed by conclusion in section V.

II.LITERATURESURVEY

Phillip Isola et al [12] proposed "Image to image translation with conditional networks". These networks not only learn the mapping from input image to output image and also learn a loss function to train this mapping. The author has demonstrated an approach at synthesizing the photos from label maps, reconstructing the objects from edge maps and colorizing the images among other tasks. In this paper the author has taken significant steps in the direction of convolution neural networks becoming the common workhorse behind a wide variety of image prediction problems that traditionally would require very different loss formulations. CNN's learn to minimize a loss function and the process is an automatic that thereby a lot of manual effort still goes into designing effective losses. The author has implemented a new type of GAN's such as Patch GAN and U-Net based architecture. After this in order to decrease the loss function they have used stochastic gradient descent (SGD). Han Zhang et al [8] proposed "text to photo-realistic image synthesis with stacked generative adversarial networks". In this paper the author has demonstrated the Synthesizing photo-realistic images from text descriptions and it is a challenging problem in computer vision and has many practical applications. Qiang Huang et al [13] has proposed “synthesis of images by two-stage generative adversarial networks”. In this paper the author has proposed a technique divide-and-conquer using generative adversarial networks (GANs) to explore how a machine can represent colorful pictures (bird) using a huge amount of training data and they simulate the procedure of an artist drawing a picture, where one begins with drawing objects contours and edges and then paints them different colors. They adopted two GAN models to process basic visual features including shape, texture and color. The first GAN model to generate object shape and the Second GAN model to paint the black and white image based on the knowledge learned by running their experiments on 600 color images. The experimental results show that the use of their approach can generate good quality synthetic images, comparable to real ones. Xian Wu et al[17] has present “a survey of image synthesis and editing with generative adversarial networks. It produces a reasonable and realistic image by using generative adversarial network. Leon A. Gatys et al [10] has demonstrated “a neural algorithm of Artistic style”. In fine art, especially painting, humans have mastered the skill to create unique visual experiences through composing a complex interplay between the content and style of an image. However, in other key areas of visual perception such as object and face recognition near-human performance was recently demonstrated by a class of biologically inspired vision models called Deep Neural Networks. The author has introduced an artificial system based on a Deep Neural Network that creates artistic images of high perceptual quality. The system uses neural representations to separate and recombine content and style of arbitrary images, providing a neural algorithm for the creation of artistic images. Moreover, in light of the striking similarities between performance-optimized artificial neural networks and biological vision, their work offers a path forward to an algorithmic understanding of how humans create and perceive artistic image. Tero karras et al [15] has proposed a paper “Progressive growing of GAN’s for improved quality”. In this paper , a new training methodology for generative adversarial networks. The main idea of this paper is to generate high resolution by adding new layers for the model with the use of generator and discriminator. This both speeds the training up and greatly stabilizes it, allowing us to produce images of unprecedented quality of images. Finally, they suggest a new metric for evaluating GAN results, both in terms of image quality and variation.

III. PROPOSED SYSTEM

This section will demonstrate the proposed system for image generation i.e GAN which can be developed using basic neural network and later adding convolutional neural network to it.

3.1 System Architecture-

The system architecture is represented in the form of flow chart as shown in figure 2. This represents the basic architecture of implementation of GAN for fake image synthesis, which consists of a generator and discriminator using convolution neural network method.

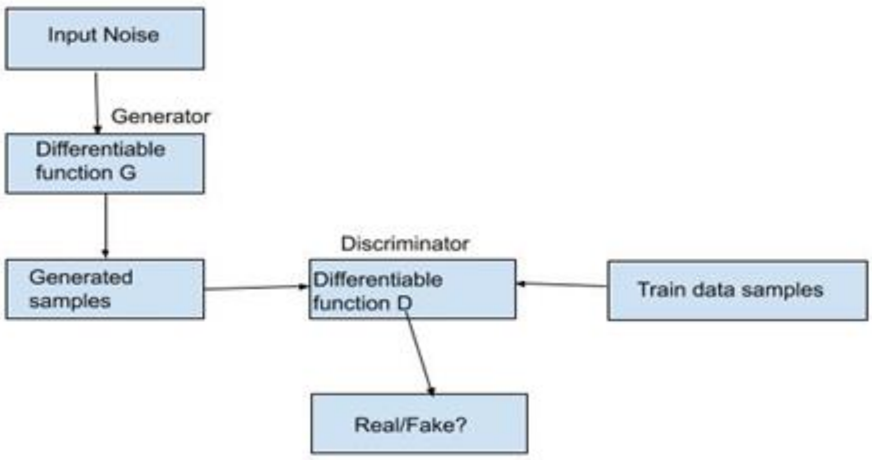

Figure 2: System Architecture

The Figure 2 shows the two important components of GAN i.e Generator and Discriminator where Generator tries to create new fake images in such a way that discriminator don’t notice that it is fake . In this way it fools discriminator thereby creating new images. Generative Adversarial Networks are composed of two models:

- The first model is called a Generator and it aims to generate new data similar to the expected one. The Generator could be assimilated to a human art forger, which creates fake works of art.

- The second model is named the Discriminator. This model's goal is to recognize if an input data is 'real' belongs to the original dataset or if it is 'fake' generated by a forger. In this scenario, a Discriminator is analogous to the police (or an art expert), which tries to detect artworks as truthful or fraud.

3.2 Algorithm steps-

Step 1: Getting the Data- In this step gathering of data is done. The dataset which we have taken is MNIST-Modified National Institute of Science and Technology. This dataset contains 10,000 images of handwritten digits from 0 to 9.Each number is going to have 1000 images.

Step 2: Initialize parameters-We first define methods to initialize both the filters for the convolution layers and the weights for the dense layers. The number of filters to be chosen is completely based on user.

Step 3: Define the back propagation operations-To compute the gradients that will force the network to update its weights and optimize its objective, we need to define methods that back propagate gradients through the convolutional and max pooling layers. And pooling can be either max pooling or average pooling based on user need.

Step 4: Building the network-We now define a method that combines the forward and backward operations of a convolution neural network. It takes the network's parameters and hyper parameters as inputs and spits out the gradients:

Step 5: Training the network-To efficiently force the network's parameters to learn meaningful representations, we use the Adam optimization algorithm.

IV. EXPERIMENT AND RESULT

In the Experimental results, the data set the author has taken standard MNIST dataset. It contains handwritten digits.It contains 60,000 training images and 10,000 testing images for validation. This dataset contains handwritten digits from 0-9 as shown in figure 3.

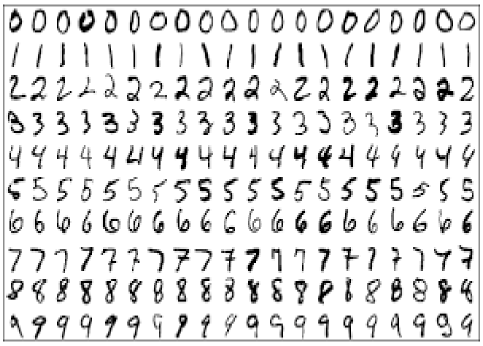

Figure 3 Sample Data set: Handwritten digits

The author has implemented GAN with ANN and CNN. After performing all the operations of GAN the author has observed a lot of changes in the output of ANN and CNN using GAN. The images generated by CNN using GAN are more clear and they are obtained with less training that is with less number of epochs. For CNN based GAN the clarity images have started coming at 90^{th} epoch where as ANN from after 140^{th} epochs.

Figure 4 ANN for Epoch 1

Figure 5 ANN for Epoch 9

Figure 6 ANN for Epoch 150

Figure 7 ANN for Epoch 200

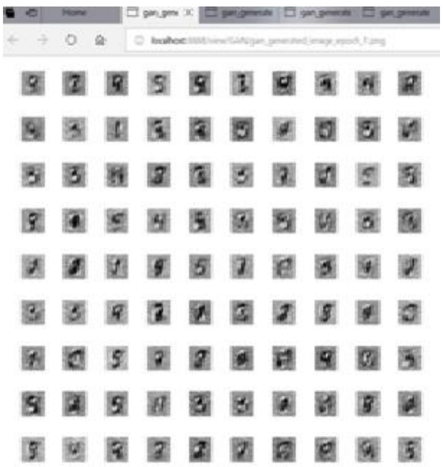

Figure 8 CNN using GAN for Epoch 1

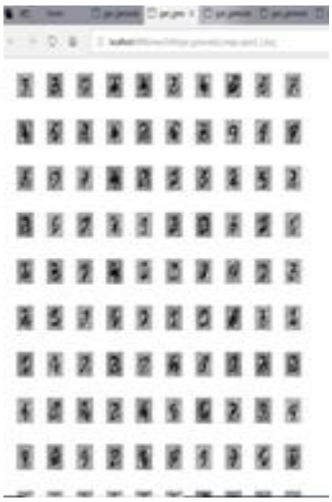

Figure 9 CNN using GAN for Epoch 9

Figure 10 CNN using GAN for Epoch 50

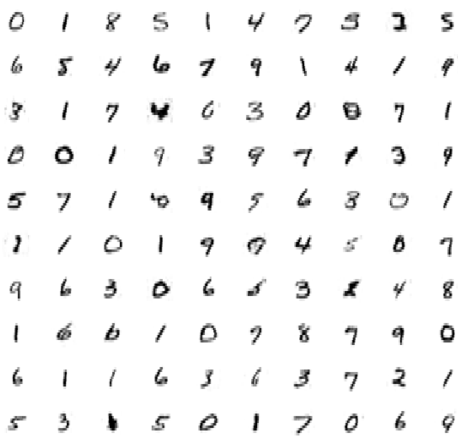

Figure 11 CNN using GAN for Epoch 150

V. CONCLUSION

In this paper, the author has studied how to use GAN and various applications of GAN, out of which image synthesis is well learned. The author has implemented ANN and CNN using GAN and found that CNN using GAN shows the clarity images while comparing ANN with less number of epochs. The author has create new images from the existing images which will be helpful when the data set is of small size and it reduces the time in creating the anime pictures. The data set used in this work is hand written recognition from MNIST dataset.

REFERENCES:

[1] Alec Radford & Luke Metz, “Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks” AI Research New York, NY arXiv:1511.06434v2 [cs.LG] (Jan 2016).

[2] Andrew Brock, Heriot, “Large Scale Gan Training For High Fidelity Natural Image Synthesis”Watt University arXiv:1809.11096v2 [cs.LG] (Feb 2019).

[3] Chaoyue Wang , Chang Xu , Xin Yao , Dacheng Tao, “Evolutionary Generative Adversarial Networks”, Centre for Artificial Intelligence, School of Software, University of Technology Sydney, Australia arXiv:1803.00657v1 [cs.LG] (Mar 2018).

[4] Dr.R.Parthasarathi , G.Vijayanand , B.Arun , G.R. Gopinath , K.Ranjith kumar, “ ANN Based Modeling for Performance and Exhaust Emission of DI Diesel Engine using Emulsified Diesel Fuel”Swansea printing technology limited,Tanga Journal,volume 14,pp:2175-2182, (2018).

[5] Fuzhou Gong, Zigeng Xia “Generate the corresponding Image from Text Description using Modified GAN-CLS Algorithm” , computerscience, ArXiv:1806-11302v1(2018)

[6] Goodfellow, I. J., Mirza, M., Courville, A., and Bengio, Y. (2013b). “Multi-prediction deep Boltzmann machines”. In NIPS (2013).

[7] Goodfellow, I. J., Warde-Farley, D., Lamblin, P., Dumoulin, V., Mirza, M., Pascanu, R., Bergstra, J., Bastien, F., and Bengio, Y. (2013c). “Pylearn2: a machine learning research library”. arXiv preprint arXiv:1308.4214.

[8] Han Zhang, Tao Xu.Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., and Bengio, Y, “Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks Maxout networks”. arXiv:1612.03242v2 [cs.CV] (Aug 2017)

[9] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Nets” (2014) a statistic machine learning . arXiv preprint arXiv:1406.2661v1.

[10] Leon A.Gatys, Alexander S.Eckeer, Mattias Bethge, “A Neural Algorithm of Artistic style” computer science, arXiv:1508.06576v2 (2015)

[11] Nicolas Vecoven , Damien Ernst, Antoine Wehenkel , Guillaume Drion ,”Introducing Neuromodulation in Deep Neural Networks to Learn Adaptive Behaviours”(2018) Machine Learning (cs.LG); Neural and Evolutionary Computing (cs.NE); Machine Learning (stat.ML) arXiv:1812.09113 .

[12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros , “Image-to-Image Translation with Conditional Adversarial Networks” computer science, arXiv:1611.07004v3 (2018)

[13] Qiang Huang,Philip J.B Jackson,Mark D.Plumbley,Wenwu Wang, “Synthesis of Images by two stage Generative Adversarial Networks”, International Conference on Acoustics, Speech, and Signal Processing, ICASSP (April 2018)

[14]. Samaneh Azadi1, Matthew Fisher, Vladimir Kim, Zhaowen Wang, Eli Shechtman, “ Multi-Content GAN for Few-Shot Font Style Transfer”(2018).

[15]Tero Karras , Samuli Laine ,Timo Aila, “progressive growing of GANs for improve quality, stability and variation” in ICLR(2018)

[16] Ugur Demir, Gozde Unal, “Patch-Based Image Inpainting with Generative Adversarial Networks”, Computer Vision and Pattern Recognition, arXiv: 1803.07422(2018)

[17] Wei Yang, Wanli Ouyang, Xiaolong Wang, Jimmy Ren, Hongsheng Li, Xiaogang Wang, “Learning 3D Human Pose Estimation in the Wild by Adversarial” Computer Vision and Pattern Recognition arXiv:1803.09722 (2018)

[18] Xian Wu ; Kun Xu ; Peter Hall, “A survey of image synthesis and editing with generative adversarial networks” Tsinghua Science and Technology (Volume: 22 , Issue: 6 , **DOI:** 10.23919/TST.2017.8195348 (December 2017).

[19]Xiangyu Liu, Yunhong Wang, Qingjie Liu, “PSGAN: A Generative Adversarial Network For Remote Sensing Image Pan – Sharpening” IEEE Explorer **,** 25th IEEE International Conference on Image Processing (ICIP) DOI: 10.1109/ICIP.2018.8451049 (2018).