International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V69I2P110

Volume 69 Issue 2, 64-67, February 2021
© 2021 Seventh Sense Research Group®

Original Article

A System for Identifying Synthetic Images using LSTM: A Deep Learning Approach

Hemanth Somasekar[1], Kavya Naveen [2]

[1]*Computer science Engineering, RNS Institute of Technology*

[2]*Artifciial Intelligence & Machine learning Engineering, RNS Institute of Technology.*
Bangalore, Karnataka, India.

Received Date: 31 December 2020
Revised Date: 15 February 2021
Accepted Date: 17 February 2021

Abstract *- In the current scenario Generative Adversarial Network (GAN) is generating more exhilaration in various fields with an amazing growth of it can be seen over a few years. It is very much successful in generating synthetic images over natural images. These are unsupervised neural networks that are capable of creating new image samples based on the training process they have adapted from the information that has been fed to them. On the other hand, Long Short Term Memory (LSTM) is one type of Recurrent Neural Network (RNN) mainly used in the domain facing sequence prediction issues. In this paper, the GAN is considered a Generator, and the LSTM is considered a Discriminator. The work of the generator is to produce synthetic images out of random samples. Based on the fine-tune training, it can produce a perfect fake image that is difficult to identify as a real one. The same is fed to the LSTM network along with the real images, and the fine-tune training is performed to get more perfect synthetic images. Both facial datasets, as well as abstract art dataset available open-source, is taken for training and testing. From this research, it is proven that Generative Adversarial Network (GAN) and Long Short-Term Memory (LSTM) are the networks utilized, and the accuracies were found to be 58.53% and 72.68%, respectively, which explicitly proves that synthetic images are more clearly identified by the LSTM over GANs.*

Keywords *- Generative Adversarial Network, Long Short-Term Memory.*

I. INTRODUCTION

Machine Learning (ML) plays a pivotal role in contemporary trade and investigation. The recital of machine learning mainly evolves based on the algorithms and the neural network models for supporting the computer systems. Based on the submitted samples called training data, it generates arithmetical models mechanically and provides decisions without any human intervention.

Deep Learning, on the other hand, provides solitary intelligent decisions by creating an artificial neural network capable of learning by itself. It needs a lot of data for learning, and it provides better accuracy from the experience it gained from the learning predictions.

Neural Network is a division of ML with a sequence of algorithms. It follows a certain process that imitates human brain operation that ventures in identifying the hidden relationships among the set of data. There is also a mathematical possibility to derive the reason behind the predictions.

A Generative Adversarial Network (GAN) is a type of neural network architecture generating new data instances based on the training with the probable samples. It is categorized as unsupervised learning with the capability of generating fresh data points with a little bit of deviation learned from the exact data distribution. At the same time, it is extremely a challenging task to study the accurate distribution of data. Hence the data distribution needs to be modeled by influencing the neural networks in such a way that it suits the exact distribution.

GAN's are trained in an intelligent way that handles the problem as a supervised learning problem with the help of two models, namely generator model, and discriminator model, by maintaining a perfect symmetry between both the models. The former model deals with the new samples, whereas the latter presumes the samples either as genuine or counterfeit.

Long Short Term Memory (LSTM) is an artificial recurrent neural network (RNN) architecture with feedback acquaintances that has taken a pivotal role in the field of deep learning. It has the capability of processing both single as well as sequence data. It has the capability of studying order reliance in succession forecasting problems and also works very well with complex problems like speech recognition, machine translation, etc.

II. RELATED WORK

A new generative model has been developed for acquiring adversarial training along with a novel performance measure which helps both the models to contend against one another. This resulted in producing visual samples with higher quality compared to the single-step model. Also, the proposed model comes with a new metric that capitulates the performance comparatively better than the current state-of-the-art generative models [1].

Supervised Adversarial Network (SAN) is a novel method that resembles the functionality of basic GAN with a small improvisation for detecting the most relevant images. Similar to GAN, it also deals with two novel aspects called as G-Network and D-Network that work on the conversion of true images into fake images, and based on the fake images, the latter is trained respectively. The D-network is introduced with a new technique called a convolution-comparison layer to produce superior images with the help of Simple Linear Iterative Clustering pixels [2].

Numerous realistic applications face defying while generating images from text metaphors. Frequently a general meaning of the metaphors is generated; on the other hand, the essential important minutiae of the same are missed out. This has been overcome with the new proposed methodology, Stacked Generative Adversarial Networks (Stack GAN), which helps in spawning 256×256 photo-realistic icons with higher resolution are stipulated on text metaphors [3].

A binary generative adversarial network (BGAN) is an unsupervised hashing technique used for entrenching the image to binary codes resulting in the generation of reasonable images like true images. The experiment is carried out in the CIFAR-10, NUSWIDE, and Flickr datasets that are considered to be the traditional datasets with an outcome of 107% in performance compared to the existing hash technique [4].

A Label Denoising Adversarial Network (LDAN) is introduced to artificial data to coach a deep Convolution Neural Network for lighting regression on genuine face metaphors. This method deals with precise facts resulting in 100,000 times faster execution time compared to the existing optimization-based lighting evaluation methods [5].

A novel method is introduced for improvising GAN in an efficient way aiming at human dealings. This method enhances the rate of optimistic communication by humans [6].

The GAN is embedded with a variation auto-encoder called Variation Generative Adversarial Networks (VGAN), dealing with two novel aspects. Firstly, it accepts loss for the discriminative and classifier network along with a mean inconsistency intention for the generative network resulting in constant training. Secondly, it uses using pairwise feature matching to configure the generated images by accepting an encoder network to gain knowledge of the connection existing between the dormant and the genuine image space [7].

In Remote Sensing Field Very High Spatial Resolution (VHSR), large-scale SAR image databases are taken into account for which the image generation test is carried out for finding the accuracy of artificial data. [8]

A new method is named Stacked Generative Adversarial Networks (Stack GANs) that works in two levels. The first level converts the given text details into an image with low resolution. The next level generates a high-resolution image by taking the output from the first level along with the text account as inputs. The experimental results surpass other modern methods by providing the most significant photo images [9].

It has been proved with the experiments that for the qualified unlabeled data, the mechanism named Dual GAN generates appreciable progress in the image outcome. It has also been proved that Dual GAN is comparatively better than conditional GAN as the latter deals with the qualified labeled data [10].

A novel method called Conditional Cycle GAN is introduced for generating face images using attribute and identity-guided face image generation technique. This technique helps to overcome the adversarial loss by involving a conditional feature vector while feeding the input in both networks. In the identity identity-preserving technique, the taken input with low resolution is compared with the identity image to produce a high-resolution face image. Whereas in the face transfer technique, with the help of attributes provided, the face image can be transferred from one gender to the other [12].

A new architecture for GAN is introduced that deals with unsupervised learning without any human intervention for the division of complex attributes as well as for dissimilar images that have been spawned. This new architecture enhances the modern methods in two main fields like allocation for interpolation and entanglement for differences resulting in an extremely diverse and premium human faces dataset [13].

A survey has been carried out on the low-level statistics of images generated by state-of-the-art deep generative models. For this, the following models, namely Variational auto-encoder (VAE), Wasserstein generative adversarial network (WGAN), and deep convolutional generative adversarial network (DCGAN), were skilled on various dataset like ImageNet dataset, a large set of cartoon frames from animations. The survey resulted in revealing the capability of capturing the spirit of natural sceneries. It also provided a novel dimension that can appraise the models. It has also suggested possible directions to improve image generation models [14].

The survey deals with the generation of premium metaphors, an assortment of metaphor generation, and constant preparation. It has been identified that modern GANs like BigGAN and PROGAN are able to fabricate premium metaphors and assortment metaphors in the field of computer hallucination. Meantime it also has been identified that GANs are very best at images rather than videos [15].

A novel method called Conditional CycleGAN is introduced for generating face images using attribute and identity-guided face image generation techniques. This technique helps to overcome the adversarial loss by involving a conditional feature vector while feeding the input in both networks. In the identity identity-preserving technique, the taken input with low resolution is compared with the identity image to produce a high-resolution face image. Whereas in the face transfer technique, with the help of attributes provided, the face image can be transferred from one gender to the other [12].

A new architecture for GAN is introduced that deals with unsupervised learning without any human intervention for the division of complex attributes as well as for dissimilar images that have been spawned. This new architecture enhances the modern methods in two main fields like allocation for interpolation and entanglement for differences resulting in an extremely diverse and premium human faces dataset [12].

A survey has been carried out on the low-level statistics of images generated by state-of-the-art deep generative models. For this, the following models, namely Variational auto-encoder (VAE), Wasserstein generative adversarial network (WGAN), and deep convolutional generative adversarial network (DCGAN), were skilled on various dataset like ImageNet dataset, a large set of cartoon frames from animations. The survey resulted in revealing the capability of capturing the spirit of natural sceneries. It also provided a novel dimension that can appraise the models. It has also suggested possible directions to improve image generation models [13].

The survey deals with the generation of premium metaphors, an assortment of metaphor generation, and constant preparation. It has been identified that modern GANs like Big GAN and PROGAN are able to fabricate premium metaphors and assortment metaphors in the field of computer hallucination. Meantime it also has been identified that GANs are very best at images rather than videos [14].

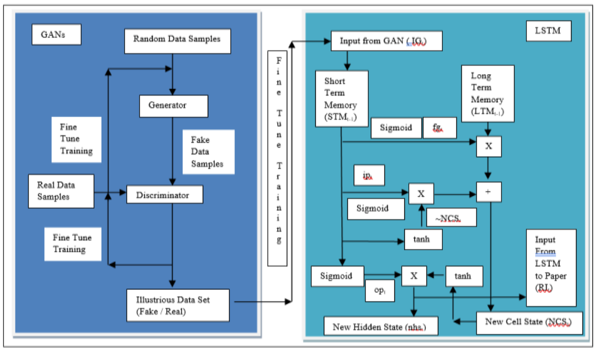


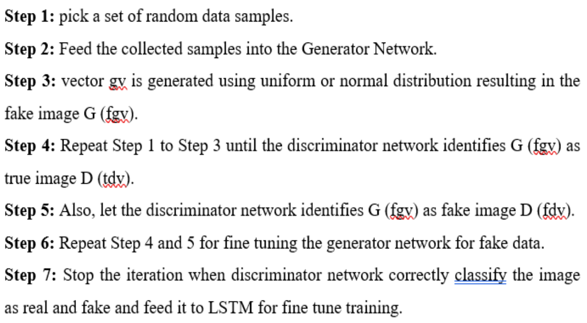
Fig. 1 Block Diagram of Proposed Methodology

GAN block consists of a Generator network and a Discriminator network to which the random sample images are fed into the former, for which the synthetic data are obtained as an output from the former is fed into the latter along with some real images. The fine-tuning training is performed on both the networks in order to obtain perfect synthetic data from the generator network, and also, the discriminator is fine-tuned in such a way that it can identify the fake images.

III. PROPOSED METHODOLOGY

In this paper, Long Short-Term Memory (LSTM), an artificial recurrent neural network (RNN) architecture, is used for training the images obtained from GAN, and accuracies are compared between GAN and LSTM. The entire work is caged into a major block comprising two sub-blocks, namely GAN and LSTM, in which the former acts as a Generator and the latter acts as a Discriminator, respectively. The work of the Generator is to get the random sample values and generate fake data samples, and the same has to be fed into the Discriminator along with the real data samples. The work of the LSTM acting as a Discriminator has the capability of recognizing the synthetic data generated based on the fine-tune training.

IV. PSEUDO CODE FOR GAN



```
Step 1: pick a set of random data samples.
Step 2: Feed the collected samples into the Generator Network.
Step 3: vector gv is generated using uniform or normal distribution resulting in the fake image G (fgv).
Step 4: Repeat Step 1 to Step 3 until the discriminator network identifies G (fgv) as true image D (tdv).
Step 5: Also, let the discriminator network identifies G (fgv) as fake image D (fdv).
Step 6: Repeat Step 4 and 5 for fine tuning the generator network for fake data.
Step 7: Stop the iteration when discriminator network correctly classify the image as real and fake and feed it to LSTM for fine tune training.
```

Fig. 2 Pseudo Code for GAN

fgv	-	Fake generated Vector / image
tdv	-	True Discriminated Vector / image

LSTM is comprised of memory blocks, where each memory block is comprised of three propagative units known as input gate, output gate, and forget gate. Other than this, it has memory cells connected in a recurrent way consisting of a repetition of simple little units like this, which take as an input the past, a new input, and produce a new prediction and connect to the future. Now, what's in the middle of that is typically a simple set of layers with some weights and linearities.

The gating values for each gate get controlled by a tiny logistic regression on the input parameters. Each of them has its own set of shared parameters. And there's an additional hyperbolic tension sprinkled to keep the outputs between -1 and 1. Also, it's differentiable all the way, which means it can optimize the parameters very easily. All these little gates help the model keep its memory for longer when it needs to and ignore things when it should.

The sigmoid activation function used in neural networks has an output boundary of (0, 1), and α is the offset parameter to set the value at which the sigmoid evaluates to 0. The sigmoid function often works fine for gradient descent as long as the input data x is kept within a limit. For large values of x, y is constant. Hence, the derivatives dy/dx (the gradient) equates to 0, which is often termed as the vanishing gradient problem. This is a problem because when the gradient is 0, multiplying it with the loss (actual value - predicted value) also gives us 0, and ultimately networks stop learning.

The following are the equations used for gates in LSTM:

fg_t=sigmoid(mmul(wt_{fg},IG_t)+mmul(Ut_{fg},STM_{t1})
+ofs_{fg}) Eqn – (1)

ip_t=sigmoid(mmul(wt_{ip},IG_t)+mmul(Ut_{ip},STM_{t1})
+ofs_{ip}) Eqn – (2)

op_t=sigmoid(mmul(wt_{op},IG_t)+mmul(Ut_{op},STM_{t1})
+ofs_{op}) Eqn – (3)

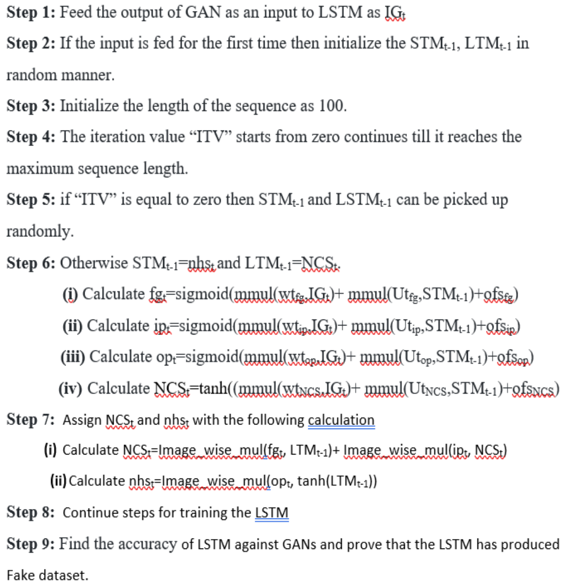
NCS_t=tanh((mmul(wt_{NCS},IG_t)+mmul(Ut_{NCS},STM_{t1})
+ofs_{NCS}) Eqn – (4)

NCS_t=Image_wise_mul(fg_t,LTM_{t-1})+
Image_wise_mul(ip_t, NCS_t) Eqn–(5)

nhs_t=Image_wise_mul(op_t, tanh(LTM_{t-1}))Eqn – (6)

fg_t	-	Forget gate
ip_t	-	Input gate
op_t	-	Output gate
Wt	-	Weight of respective gate neurons
Ut	-	Update of respective gate neurons
Ofs	-	Biases of respective gate neurons
STM_{t-}	-	Input from GANs (at timestamp t-1)
LTM_{t-1}	-	Real Images (at timestamp t-1)
NCS	-	New Cell State (at timestamp t)
nhs	-	New Hidden State (at timestamp t)
mmul	-	Matrix Multiplication

V. PSEUDO CODE FOR LSTM



```
Step 1: Feed the output of GAN as an input to LSTM as $IG_t$
Step 2: If the input is fed for the first time then initialize the $STM_{t-1}$, $LTM_{t-1}$ in random manner.
Step 3: Initialize the length of the sequence as 100.
Step 4: The iteration value "ITV" starts from zero continues till it reaches the maximum sequence length.
Step 5: if "ITV" is equal to zero then $STM_{t-1}$ and $LSTM_{t-1}$ can be picked up randomly.
Step 6: Otherwise $STM_{t-1}$=$nhs_t$ and $LTM_{t-1}$=$NCS_t$.
        (i) Calculate $fg_t$=sigmoid(mmul($wt_{fg}$,$IG_t$)+ mmul($Ut_{fg}$,$STM_{t-1}$)+$ofs_{fg}$)
        (ii) Calculate $ip_t$=sigmoid(mmul($wt_{ip}$,$IG_t$)+ mmul($Ut_{ip}$,$STM_{t-1}$)+$ofs_{ip}$)
        (iii) Calculate $op_t$=sigmoid(mmul($wt_{op}$,$IG_t$)+ mmul($Ut_{op}$,$STM_{t-1}$)+$ofs_{op}$)
        (iv) Calculate $NCS_t$=tanh((mmul($wt_{NCS}$,$IG_t$)+ mmul($Ut_{NCS}$,$STM_{t-1}$)+$ofs_{NCS}$)
Step 7: Assign $NCS_t$ and $nhs_t$ with the following calculation
        (i) Calculate $NCS_t$=Image_wise_mul($fg_t$, $LTM_{t-1}$)+ Image_wise_mul($ip_t$, $NCS_t$)
        (ii) Calculate $nhs_t$=Image_wise_mul($op_t$, tanh($LTM_{t-1}$))
Step 8: Continue steps for training the LSTM
Step 9: Find the accuracy of LSTM against GANs and prove that the LSTM has produced Fake dataset.
```

Fig. 3 Pseudo Code for LSTM

VI. RESULTS AND COMPARISON

The following graphs in figure 4 and figure 5 depict the comparison of losses and accuracies of the two networks, i.e., GAN and LSTM, respectively. Both the networks are trained till less training loss is obtained. The accuracy of GAN was found to be 58.53%, and for LSTM, it was 72.68%, where the networks are represented in the x-axis and accuracies are measured on the y-axis.

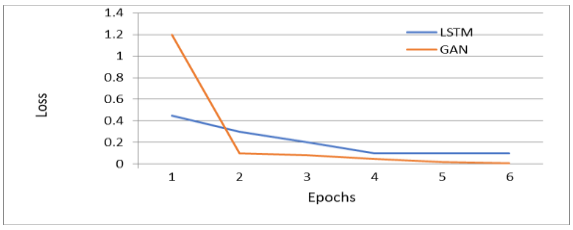


Fig. 4 Comparison of Losses for GAN and LSTM

VII. CONCLUSION

In this paper, both the networks, namely GAN and LSTM, are performed with fine-tune training in such a way that the synthetic images are identified perfectly without any discrimination. The block diagram clearly depicts the functionality of GAN as well as the LSTM, which are fed with proper datasets available open-source. Both the networks generate a dataset that can be categorized as trained as well as a test dataset. In this research, both facial images, as well as scenery images, are trained and tested, which resulted in the accuracy of 58.53% and 72.68% for the networks GAN and LSTM, respectively.

REFERENCES

[1] Daniel Jiwoong Im, Chris Dongjoo Kim, Hui Jiang, and Roland Memisevic. Generating images with recurrent adversarial networks. arXiv:1602.05110v5 [cs.LG] 13 (2016).

[2] Hengyue Pan and Hui Jiang. Supervised Adversarial Networks for Image Saliency Detection. arXiv:1704.07242v2 [cs.CV] 26 (2017).

[3] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang and Dimitris Metaxas. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. arXiv:1612.03242v2 [cs.CV] 5 (2017).

[4] Jingkuan Song. Binary Generative Adversarial Networks for Image Retrieval. arXiv:1708.04150v1 [cs.CV] 8 (2017).

[5] Hao Zhou, Jin Sun, Yaser Yacoob, and David W. Jacobs. Label Denoising Adversarial Network (LDAN) for Inverse Lighting of Face Images. arXiv:1709.01993v1 [cs.CV] 6 (2017).

[6] Andrew Kyle Lampinen, David So, Douglas Eck and Fred Bertsch. Improving generative image models with human Interactions. arXiv:1709.10459v1 [cs.CV] 29 (2017).

[7] Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, Gang Hua. CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training. arXiv:1703.10155v2 [cs.CV] 12 (2017).

[8] Dimitrios Marmanis, Wei Yao, Fathalrahman Adam, Mihai Datcu, Peter Reinartz, Konrad Schindler, Jan Dirk Wegner and Uwe Stilla. Artificial generation of big data for improving image classification: a generative adversarial network approach on sar data. arXiv:1711.02010v1 [cs.CV] 6 (2017).

[9] Han Zhang, Tao Xu, Hong sheng Li, Shaoting Zhang, Xiaogang WangXiaolei, HuangDimitris, and N. Metaxas. StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks. arXiv:1710.10916v3 [cs.CV] 28 (2018).

[10] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. arXiv:1704.02510v4 [cs.CV] 9 (2018).

[11] Vaibhav Kumar, Recurrent Neural Network based Language Modeling for Punjabi ASR SSRG International Journal of Computer Science and Engineering 7(9) (2020) 7-13.

[12] Yongyi Lu, Yu-Wing Tai, and Chi-Keung Tang. Attribute-Guided Face Generation Using Conditional CycleGAN. arXiv:1705.09966v2 [cs.CV] 14 (2018).

[13] Tero Karras, Samuli Laine and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv:1812.04948v3 [cs.NE] 29 (2019)

[14] Yu Zeng, Huchuan Lu, and Ali Borji. Statistics of Deep Generated Images. arXiv:1708.02688v5 [cs.CV] 24 (2019).

[15] Zhengwei Wang, Qi She, Tom´as E. Ward. Generative Adversarial Networks in Computer.