Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

# Survey on Deep Learning in Music using GAN

Rajat Kulkarni, Rutik Gaikwad, Rudraksh Sugandhi,
Pranjali Kulkarni, Shivraj Kone
RMD Sinhgad Technical Institute Campus

***Abstract*:- Generating realistic music is one of the exciting tasks in the field of deep learning. We studied various models that used Generative Adversarial Network (GANs), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN) and based on our survey we analyzed that GAN provides an effective training to generate music through use of collection of midi files.**

***Keywords:- Generating realistic music(GAN), SVM, RNN, LSTM***

## I. INTRODUCTION

Music composition is a challenging craft that has been a way for artists to express themselves ever since the dawn of civilization. We studied various papers of CNN, GAN, LSTM and analyzed that GAN is better option for generating music. In CNN a bar graph or wave chart is used to map the musical notes of the midi files. Then the images are sent to the GAN, the GAN then generates a new set of musical notes. In RNN we feed the midi file format directly to RNN, RNN has the capability of remembering the previous states of the musical notes and then produce the next notes. In LSTM model they used two LSTM
networks that worked like GAN. One LSTM acted as generator while other was the Generator Generative adversarial networks (GANs) are a class of neural network architectures designed with the aim of generating realistic data. Midi file data is given to GAN as from of input on which the GAN learn and give output.

GAN approach involves training two neural models with conflicting objectives, one generator (G), and one discriminator (D), forcing each other to improve output image file is again converted back to the music file and represents the file output of our model.After the generation of new image files of music by the two LSTM networks of the GAN model, we use a SVM(Support Vector Machine)Classifier to classify the image files and get the best image file as the final output from the classifier. Then the output image file is again converted back to the music file and represents the file output of our model.

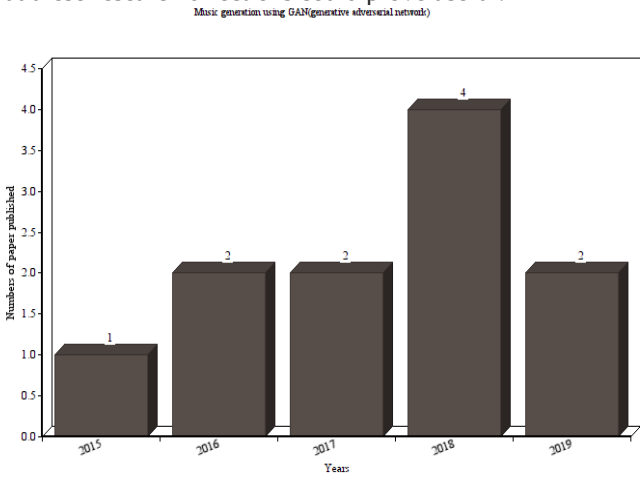
## II . LITERATURE SURVEY

[1]Allen Huang states previous work in music generation has mainly been focused on creating a single melody.[1] More recent work on polyphonic music modelling, centered around time series probability density estimation, has met some partial success. One of the earliest papers on deep learning-generated music, written by Chen et al , generates one music with only one melody and no harmony. The authors also omitted dotted notes, rests, and all chords.[1]Midi files are structured as a series of concurrent tracks, each containing a list of meta messages.
[1]They used a 2-layered Long Short Term Memory (LSTM) recurrent neural network (RNN) architecture to produce a character level model to predict the next note in a sequence. [1]In their midi data experiments, they treated a midi message as a single token, whereas in piano roll experiment, they treated each unique combination of notes across all time steps as a separate token.
[1]Their architecture allowed the user to set various hyper parameters such as number of layers, hidden unit size, sequence length, batch size, and learning rate. [1]They also anneal their learning rate when they see that the rate of training error is decreasing slowly.
[1]The conclusion by Allen Haung is to show that a multi-layer LSTM, character-level language model applied to two separate data representations is capable of generating music that is at least comparable to sophisticated time series probability density techniques prevalent in the literature.
[2]Li-Chia Yang proposed CNN-GAN based system named MidiNet which converts a noise into midi files using convolutional neural networks (CNNs).[2] In this model using CNN for generating melody (a series of MIDI notes). [2]In addition to the generator, it uses a discriminator to learn the distributions of melodies, making it a generative adversarial network (GAN).
[2]It uses random noises as input to generator CNN. [2]The goal of the generator is to transform random noises into real midi file. [2]Meanwhile, a discriminator CNN that takes input from generator and predicts whether it is from a real or a generated midi, thereby informs the generator how to appear to be real. [2]This amounts to a generative adversarial network (GAN), which learns the generator and discriminator iteratively. It shows that it can be powerful alternative to RNNs.
[3]A generative adversarial network (GAN) is a machine learning (ML) model in which two neural networks compete with each other to become more accurate in their predictions. [3]GANs typically run unsupervised and use a cooperative zero-sum game framework to learn. [3]They propose a new framework for estimating generative models via an adversarial process, in which they simultaneously train two models: a generative model G
that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. [3]The training procedure for G is to maximize the probability of D making a mistake.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019


Fig1: GAN Implementation

Fig.1.[3]The two neural networks that make up a GAN are referred to as the generator and the discriminator.[3] The generator is a convolutional neural network and the discriminator is a deconvolutional neural network. [3]The goal of the generator is to artificially manufacture outputs that could easily be mistaken for real data. [3]The goal of the discriminator is to identify which outputs it receives have been artificially created. [3]This paper has demonstrated the viability of the adversarial modeling framework, suggesting that these research directions could prove useful.




The above diagram shows number of GAN related papers published each year from 2015 to 2019.It was observed that it was highest in the year 2018. GAN has a lot of potential and variety of applications in various fields.

## III. SYSTEM APPROACH

The system we propose uses the GAN(Generative Adversarial Network) model. The Formula Below gives the cost function of the GAN model

[3]Cost Function=min G max D V (D, G) = Ex~pdata(x) [log D(x)] + Ez~pz(z) [log(1 − D(G(z)))].

GAN uses two LSTM(Long Short Term Memory) networks. LSTM's are the advanced form of RNN(Recurrent Neural Networks). LSTM's takes the output from the previous time step and uses information from it to generate a new output for the current time step. It is generally used in NLP(Natural Language Processing). In Music Generation information about the previous tone must also be recorded in order to generate the next tone in the music. LSTM's are therefore the best option to carry out music generation.

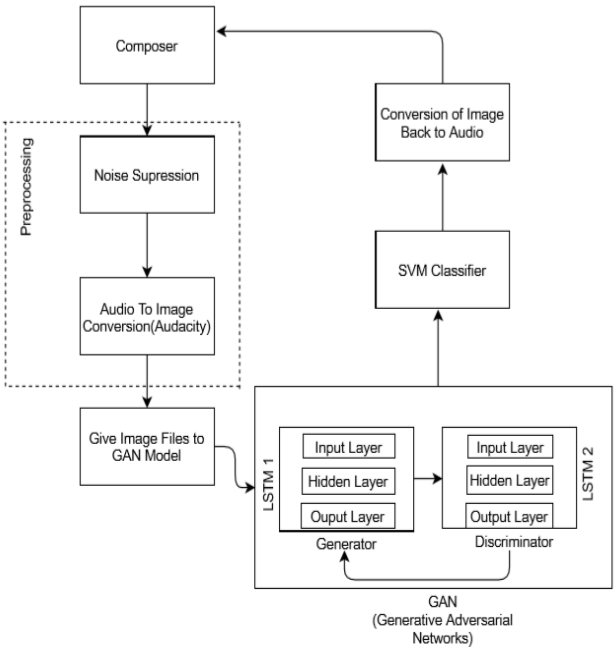



Fig.2: System Architecture

$$\min_{\theta} C\sum_{i=1}^{m}\left[y^{(i)}cost_1(\theta^T x^{(i)}) + (1-y^{(i)})cost_0(\theta^T x^{(i)})\right] + \frac{1}{2}\sum_{i=1}^{n}\theta_j^2$$

[4]The Above Formula represents the cost function of the SVM Classifier

In our proposed model we will be using two LSTM networks. One network will act as a generator while other network will be acting as a discriminator. The Generator LSTM network will produce some random noise as output. The Discriminator LSTM network will then differentiate between the sample music file and the generated noise by the generator. Based on the error observed between the original music sample and the generated sample the discriminator will then update the generator about it and the generator network will then makes changes in its weights accordingly. This process will continue until the discriminator is no longer able to distinguish between the original music sample and the generated sample. The newly generated network will then will be able to produce new music from a given samples of music files.Then finally the GAN model produces a new set of images files from the original ones's.These new set of image files are then given to the SVM(Support Vector Machine) Classifier ,it takes the newly generated image files from the GAN model and classifies them to get the best image file.The best image file is then converted back to the image file. This music files is the final output of our project.

## IV. CONCLUSION

The use of deep learning techniques for the creation of music is nowadays getting increased attention. In this paper, we studied and analyzed the various deep learning neural networks to generate musical content. We have analyzed and compared various systems and technologies proposed by various researchers. GAN uses two neural networks which are Generator and Discriminator which works concurrently. Due to this it makes efficient use of RNN and CNN. The generated

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 09, September-2019

music cannot yet compare to the music in the training data, by human judgement. The reasons for this remain to be explored.

## V. ACKNOWLEDGEMENT

Authors are thankful to Mr. Shivraj Kone and Faculty of RMD Sinhgad school of Engineering and Technology, Pune for providing the facility to carry out the research work.

## VI. REFERENCES

[1] J. Pons, "Deep learning for music information research," pp. 1–8, 2015.
[2] L.-C. Yang, S.-Y. Chou, and Y.-H. Yang, "MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation," no. March 2017, 2017.
[3] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," pp. 1–9, 2014.
[4] P. W. Wang and C. J. Lin, "Support vector machines," in *Data Classification: Algorithms and Applications*, 2014.
[5] T. Simonite, "Computer, write me a song," *Technology Review*. 2016.
[6] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, 2002.
[7] A. P. Viswanathan, "Music Genre Classification," *Int. J. Eng. Comput. Sci.*, 2016.
[8] Y. H. Yang and H. H. Chen, "Machine recognition of music emotion: A review," *ACM Transactions on Intelligent Systems and Technology*. 2012.
[9] G. Kamhi, A. Novakovsky, A. Tiemeyer, and A. Wolffberg, "MAGENTA," 2009.
[10] Y. E. Kim *et al.*, "Music emotion recognition: A state of the art review," in *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010*, 2010.
[11] H. H. Mao, "DeepJ: Style-Specific Music Generation," in *Proceedings - 12th IEEE International Conference on Semantic Computing, ICSC 2018*, 2018.
[12] J. P. Briot and F. Pachet, "Deep learning for music generation: challenges and directions," *Neural Computing and Applications*, 2018.
[13] H. W. Dong, W. Y. Hsiao, L. C. Yang, and Y. H. Yang, "Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment," in *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 2018.
[14] A. Flexer, D. Schnitzer, M. Gasser, and G. Widmer, "Playlist generation using start and end songs," in *ISMIR 2008 - 9th International Conference on Music Information Retrieval*, 2008.