[1,*]Yifan Li
[2]Xinyu Hu

The Style Transformation of Gu-Zheng Music Based on the Improved Model of Cyclegan

Abstract: - In recent years, generative adversarial networks have excelled in the field of image style migration, however their performance in the field of music has been mediocre. Existing music style migration does not work well for style migration of gu-zheng music. In order to solve these problems, we first extract the features of gu-zheng music and the Mel-spectrum features, then use CycleGAN to do style transformation on the combined features and Mel-spectrum features, and then use WaveNet vocoder to decode the migrated spectrograms, and finally achieve the style migration with gu-zheng music. The proposed model was evaluated on the publicly available dataset FMA, and the average style migration rate of the compliant music reached 94.07%. Compared to other algorithms, the music produced by this method outperformed other algorithms in terms of style migration rate and audio quality.

Keywords: CycleGAN; gu-zheng music;WaveNet vocoder.

I. INTRODUCTION

Old-fashioned music composition [1] is an emerging form of music that combines modernity with Chinese tradition. It has a large number of traditional Chinese elements, classical lyrics, beautiful melodies and a strong focus on mood [2]. Most of the music are accompanied by folk instruments to retain the traditional musical flavour.

Not only is old-fashioned music used to enhance the atmosphere of online games, but since Jay Chou, a Taiwanese Chinese pop singer, and Fang Wenshan, a lyricist, have collaborated to create songs such as "The Maiden", "East Wind Breaking" and "Chrysanthemum Terrace", the entertainment industry has received a strong response and a "Chinese wind" has been blowing [3-5]. The music of ancient styles was born out of the boom of the Internet, and some fans of ancient styles have reworked the music of Chinese game on the Internet.

The gu-zheng [6] is a strung instrument with a long history in China. The origin of the gu-zheng has been discussed by experts, and some schools of thought have suggested that it is most likely to have evolved from the four, and that the appearance and string set-up of the gu-zheng during the Han dynasty can be regarded as almost identical [7].

The number of strings on gu-zheng has varied from five, six, seven, nine, ten, eleven, twelve, thirteen, fourteen, fifteen and sixteen strings from ancient times to the present [8]. With the increasing variety of gu-zheng repertoire and the new demands on the range of sound, the number of gu-zheng strings has increased in recent times to 21, 23 and 25 strings. In ancient time, the strings were made of silk and metal, but nowadays they are made of nylon, which is wrapped around the metal strings to protect them and prolong their life, and can be coloured for easy identification [9]. The modern zither body is a resonating box made of wooden panels, long and flat but curved in appearance, with two or three round holes in the bottom. The frame is made of a variety of materials, and the quality of sound varies according to the wood used. The best woods are rosewood, but mahogany and rosewood can also be used. The head and tail of a gu-zheng have a front and a back beam. The top has strings and the sides have tuned shafts. The top of the gu-zheng has a pillar for each string, which is also known as a goose pillar because it looks like a flock of geese in flight. The gu-zheng is strung in the pentatonic scale, but it can also be tuned to the seventh scale, and there are other ways of setting the strings [10].

[1]Modern Service College,Jiangxi Metallurgical Vocational and Technical College,XinYu, JiangXi, 338015,China.

[2]Department of musicology,Xinghai Conservatory of Music,GuangZhou, GuangDong,510006, China.

*Corresponding Email:cxcy2022@sina.com

The gu-zheng is played using an armour, which is more commonly used today in the form of a tortoiseshell gourd. Traditionally, the gu-zheng was played "with the right hand and the left with the rhythm". However, due to the increasing complexity of modern zheng music, it is sometimes necessary to shift the left hand to the right [11-13].

Figure 1: Diagram of the gu-zheng.

In recent years, due to the rise of social networking sites, the art of cover music has flourished, taking many forms, such as the re-writing of a song by an arranger and its presentation as a song, or the playing of a cover instrument, or even the re-editing of a music file using a computer. However, nowadays, people are more and more interested in the quality of entertainment and therefore produce more musical work than in the past. Although there is an irreplaceable value of human creativity in music, it cannot be created in large quantities for a short period of time [14]. Moreover, the human act of music-making is often limited to those who have received musical training. This paper therefore proposes a novel way of accelerating the rate of production of music transcriptions, particularly to Chinese gu-zheng music, in order to capitalise on the Chinese music trend. The music conversion method is based on an artificial intelligence model that can process the music data in bulk and is also suitable for users who do not have a musical background but wish to cover the gu-zheng music.

II. RELATED WORK

The current academic work that has been done on music style migration is mainly on pure music timbre style migration of common instruments. The music undergoing processing can be divided into raw audio and non-raw audio. For example, [15] generates the final converted midi audio by converting the midi format audio into a piano rolling matrix, which is then fed into CycleGAN for training. A significant advantage of this method is the low computational overhead and the more variable style of the converted purely instrumental music [16].

However, the disadvantages are obvious: firstly, it is not possible to process the original audio, and secondly, only the playing dimension can be used for pattern shifting by this method. The Timbretron proposed by [17] extracts the CQT features of the audio frequencies, then converts them into timbre by CycleGAN, and then converts the converted CQT features into the original audio by training the vocoder in advance. However, because it is a stylistic transformation on a single timbrel domain, when an abrupt voice (e.g. a human voice) is present in a song, the intensity of that sound is significantly reduced.

The network can achieve conversion from one timbre domain to multiple timbrel domains, but as the algorithm requires training multiple decoders for different styles, the machine needs to suffer from a huge computational overhead and the model lacks generality as a new decoder needs to be trained when processing a new kind of music [18]. The above algorithms have shown good results in their respective research directions.

A. CycleGAN

The generative adversarial network is a deep learning model, a class of implicit generative models proposed by [19]. The model generates high quality output by playing the two modules (generative and discriminative) of the framework against each other. Assume that G is the generator, D is the discriminator, $P_{\text{data}}(x)$ is the distribution

of real samples and x is sampled from that distribution, and $P_z(z)$ is the distribution of potential codes z of x. Then the objective equation is:

$$G, D = \min_G \max_D \mathrm{E}_{x\sim P_{\text{data}}(x)}[\log D(x)] + \mathrm{E}_{z\sim P_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

CycleGAN, each of which needs to learn a mapping from the domain to the corresponding domain. The loss function of CycleGAN contains two adversarial losses in addition to a cyclic consistency loss to preserve its input structure. As shown in equation (2):

$$\mathrm{L}_{cyc}(G, F) = \mathrm{E}_{x\sim P_{ral}(x)}\left[\| F(G(x)) - x\|_1\right] + \mathrm{E}_{y\sim P_{ral}(y)}\left[\| F(G(y)) - y\|_1\right] \quad (2)$$

Where G denotes the forward conversion of CycleGAN and F denotes the backward conversion of CycleGAN.

B.Our Programme

This chapter provides details on the collection and pre-processing of data and the setting of training parameters.

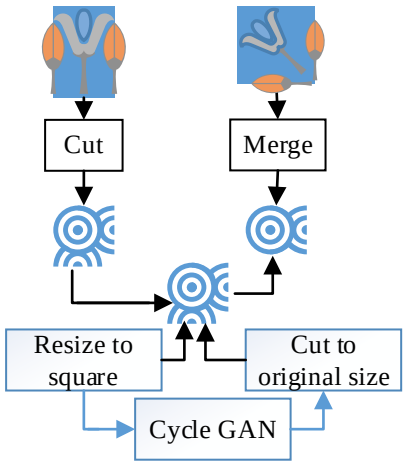

Figure 2: Flowchart of gu-zheng Music Style Transformation with CycleGAN.

As CycleGAN is an image to image conversion, we had to convert the collected music files into image files properly. As showed in Figure 2, in the pre-processing section, we first collected the music files by recording and cut the collected wave files on a per-beat basis. After cutting, the files were then converted to music using Photosounder [20], first audio editor or synthesiser to use a completely graphical approach to music creation and editing. Because of its powerful and versatile synthesis algorithms, it is capable of producing any sound possible. Finally, to facilitate the training of CycleGAN, if the image is not a multiple of 4 pixels on a side, the image is complemented by a black screen, with the original file being placed on a black square image (see Figures 3 and 4).

In the post-processing section, the images produced by CycleGAN were first trimmed back to their original size, then the images were simulated as echo files using Photosounder, and finally the wave files were synthesised into a complete tune in 1-second increments using Audio joiner [21].

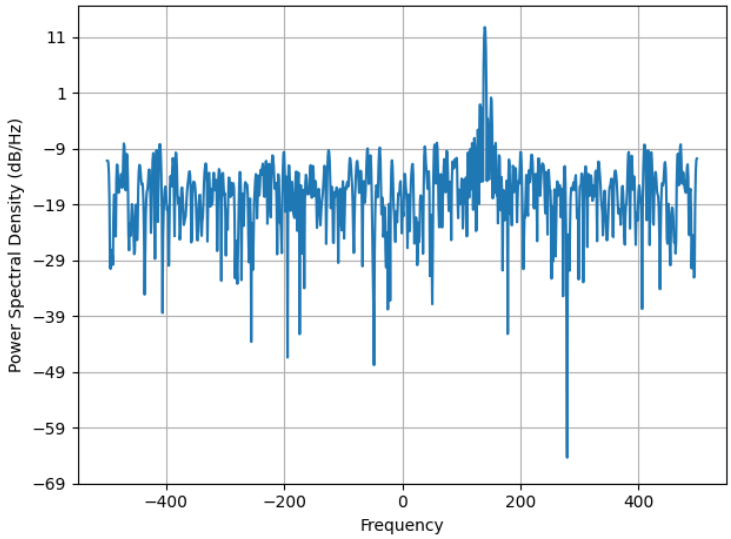

Figure 3: Bmp files converted from Photosounder to wav files in 1 second increments （Dimensions: 100*571 ）

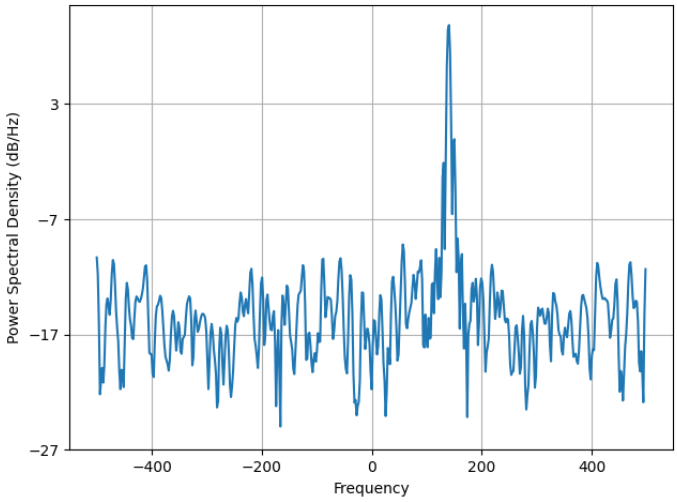

Figure 4: Bmp file with black background patch（Dimensions: 1024*1024

In this paper, we apply CycleGAN [22] to the generation of gu-zheng music. To ensure the quality of the generation, the model structure consists of a generator and an authenticator. The two models have tried to challenge each other until the generator creates a dataset that is almost indistinguishable from the real data. $G_{SynthPiano_{G}uzheng}$ the generator, which helps to transfer images from the simulated piano domain to the gu-zheng domain. Another generator, called $F_{Guzheng_{S}ynthPiano}$, helps to transfer images back into the synthetic piano domain. The generator is a model architecture of encoders and decoders. The system consists of two stride-2 volumes, which are used to train the image's nine residual blocks for conversion from synthetic piano to Chinese gu-zheng at resolutions higher than 1024 x 1024 and two 1⁄2 step volume steps [23]. The remaining blocks are used to ensure that subsequent layers retain the attributes of the previous layers. For differentiator D_{Guzheng}, a 70×70 PatchGAN is used to classify the true and false image patches. Cyclic consistency loss is used to reduce image discrepancies by updating the image input to $G_{SynthPiano_{G}uzheng}$ with the image output from $F_{Guzheng_{S}ynthPiano}$. The full objective function is given by (Eq. 3), with a recommended λ of 10 and a learning rate of 0.0002 for all networks. Since the source and target domains do not need to be correlated in any case, such a construction is suitable for music style ripping, where it is not easy to find the corresponding data in the source and target domains [25-27].

$$\begin{aligned}
&\mathrm{LosS}_{adv}\left(G, D_y, X, Y\right) = E_{y\sim Pdata}(y)[\log Y(y)] + E_{x\sim Pdata}(x)\left[1 - D_y(G(x))\right] \\
&\mathrm{Loss}_{cyc}(\mathrm{G},\mathrm{F}) = E_{x\sim Pdata(x)}\left[\| F\left(G(x) - x\right\|_1\right] + E_{y\sim Pdata(y)}\left[\| F\left(G(y) - y\right\|_1\right] \quad (3) \\
&\mathrm{Loss}_{full} = \mathrm{Loss}_{adv} + \lambda\, \mathrm{Loss}_{cyc}
\end{aligned}$$

C.Experimental Results and Analysis

In this chapter, the experimental results of the proposed methodology are presented, and related discussions, analyses and comparisons are made. It can be divided into three main sections, 1 Introduction to the dataset, 2 Scoring methods, 3 Experimental results [28].

In this experiment, a total of five Shaanxi-style gu-zheng pieces were collected, together with their corresponding piano versions, and in order to allow the model to learn a better Shaanxi accent, a special collection of Shaanxi practice pieces in the pressed tone was also included. In order to increase the amount of information available, in addition to the original tunes of the pieces, versions with 1 key up or down have also been collected as showed in Table 1.

Table 1 Training repertoire

Song	Qu Chang
Qin sangqu	5’24’’
Baihuayin	5’28’’
Cloud suit	12’54’’
Qin tuqing	8’20’’
Press to play the etude	3’19’’

D.Grading Method

We used human perception to investigate the ability of the method in this paper to preserve musical compositions. A total of 40 subjects were asked to listen to the gu-zheng music produced by our method, 20 of whom had studied the instrument and 20 of whom had never studied it. The subjects were then given a score on a scale of 1 to 5, with higher scores indicating higher similarity, based on their impressions of Chinese gu-zheng music [29].

E.Experimental Results

We averaged the similarity score of the 20 subjects who had learned the gu-zheng and obtained a score of 4.1, while the 20 subjects who had not learned the gu-zheng averaged their score and obtained a score of 4.5.

Table 2 Assessment results

Score /Subject	20 peoplehave experience forplaying gu-zheng	20 peopleno experience forplaying gu-zheng	Total experimentsubject
Average score	4.1	4.5	4.3

At the beginning of the experiment, λ_1 was set to a fixed value to demonstrate the superiority of Algorithm λ_2 over the line fading scheme proposed by [8] using the fading scheme presented in this paper. The two schemes were experimented on separately according to the previously set judging criteria, and the results were used for the next series of selected values for the hypermastigote. From Tables 3 and 4, it can be seen that the style conversion rate (TR) of Cy-cleGAN improved significantly with a small difference in audio quality, with the forward style conversion rate (TR) [30] produces an average music quality (AQR) [31]of 85.44%, the algorithm also shows good results in processing music with vocals, with some robustness.

Table 3 λ_2 Audio quality pass rates and style migration rates resulting from the use of linear attenuation

Object domain	Forward		Backward	
	AQP	TR	AQP	TR
Classical	86.65	93.6	80.64	92.4
Blues	87.65	97.4	75.31	94.5
Country	82.23	93.9	79.07	94.3
Jazz	88.5	89.5	78.2	88.3
Classical	79.95	91.1	75.39	89.4
Folk	84.68	94.2	80.12	91.1

Table 4 λ_2 Audio quality pass rates and style migration rates resulting from the use of non-linear attenuation

Object domain	Forward		Backward	
	AQP	TR	AQP	TR
Classical	85.78	96.2	82.34	95.2
Blues	89.32	97.8	79.44	93.9
Country	86.43	95.5	79.87	94.2
Jazz	82.22	92.5	80.2	89.2
Classical	81.35	95.3	78.25	91.4
Folk	86.62	94.3	81.86	91.8

After every 200 iterations, the value of the loss function, which includes the loss of the discriminator and the loss of the generator, is recorded and the waveforms of the loss values of the discriminator and the loss values of the generator are plotted on a graph, as shown in Figure 5. It can be seen that when the algorithm iterates to about 125000 steps, the losses of both the generator and the discriminator converge to a local minimum. And the generated music style and the target music style are basically close to each other, and the algorithm tends to be stable [31]. Due to computational constraints, 125 000 is chosen as the number of iterations of the algorithm in this paper, and the selection of other hypermastigote is explored on this basis.

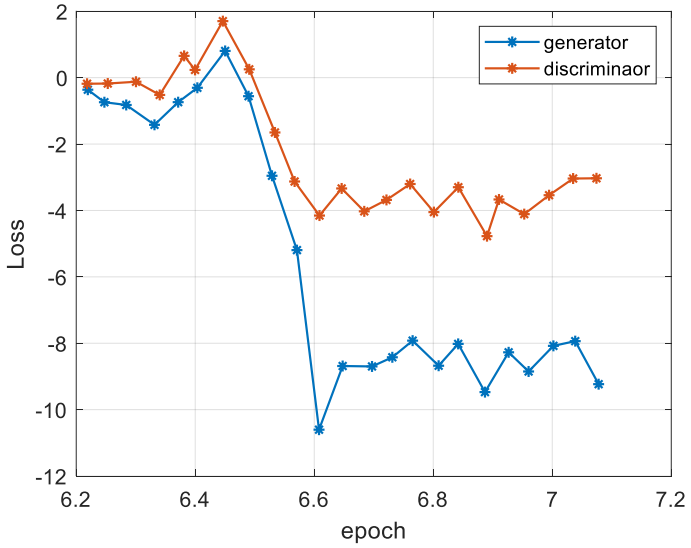

Figure 5: Forward model loss curve.

III.CONCLUSIONS

In this paper, we present a novel method for transforming the musical style of the gu-zheng. This method has the ability to produce good results, as evidenced by empirical evaluations. Furthermore, the method can be used to produce Chinese gu-zheng music even for those who do not play the instrument. In terms of future prospects, as

this model can only convert from analogue piano to gu-zheng music, it is hoped that a more flexible framework can be explored in the future to accommodate a wider range of instruments. It is also hoped that in the future, artificial intelligence can be used to solve the sound and image conversion part of this thesis. The music collected for this thesis was not recorded in a professional studio, and it is hoped that in the future more professional equipment and space can be used to avoid external interference and sound reflections.

ACKNOWLEDGEMENTS

There is no specific funding to support this research.

REFERENCES

[1] Li H, Zeng D, Chen L, et al. Immune Multipath Reliable Transmission with Fault Tolerance in Wireless Sensor Networks[C]//International Conference on Bio-Inspired Computing: Theories and Applications. Springer, Singapore, 2016: 513-517.

[2] He, S., Liao, W., Yang, M. Y., Yang, Y., Song, Y. Z., Rosenhahn, B., & Xiang, T. (2021). Context-Aware Layout to Image Generation with Enhanced Object Appearance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15049-15058).

[3] Srisombut, R. (2014). Chinese Cultural Music" Gu Zheng" in Thailand. Fine Arts Journal: Srinakharinwirot University, 18(2), 218-224.

[4] Kulkarni, R., Gaikwad, R., Sugandhi, R., Kulkarni, P., & Kone, S. (2019). Survey on deep learning in music using GAN. International Journal of Engineering Research & Technology, 8(9), 646-648.

[5] C. H. Cao, Y. N. Tang, D. Y. Huang, G. WeiMin, and Z. Chunjiong, "IIBE: an improved identity-based encryption algorithm for wsn security," Security and Communication Networks, vol. 2021, Article ID 8527068, 8 pages, 2021.

[6] Gu, Z. (2001). Economies of scale in the gaming industry: An analysis of casino operations on the Las Vegas Strip and in Atlantic City. The Journal of Hospitality Financial Management, 9(1), 1-15.

[7] LIU, M. D., & LIU, M. X. (2011). An elementary introduction to the skills in playing the composition Rush to Nadam Fair with Joy with Wang's unique technique by Gu Zheng. Journal of Zhenjiang College.

[8] Chen, J., Fan, C., Zhang, Z., Li, G., Zhao, Z., Deng, Z., & Ding, Y. (2021). A Music-driven Deep Generative Adversarial Model for Guzheng Playing Animation. IEEE Transactions on Visualization and Computer Graphics.

[9] Hayashi, T., Tamamori, A., Kobayashi, K., Takeda, K., & Toda, T. (2017, December). An investigation of multi-speaker training for WaveNet vocoder. In 2017 IEEE automatic speech recognition and understanding workshop (ASRU) (pp. 712-718). IEEE.

[10] Xiaoqian, H., Karin, K., & Chuangprakhon, S. (2021). The Guzheng Music in Henan Province, China. Review of International Geographical Education Online, 11(5), 2755-2765.

[11] Shi, X. J., Cai, Y. Y., & Chan, C. W. (2007). Electronic music for bio-molecules using short music phrases. Leonardo, 40(2), 137-141.

[12] Chen, P. H., Liang, K. W., & Chang, P. C. (2020, September). Music Conversion from Synthetic Piano to Chinese Guzheng Using Image-based Deep Learning Technique. In 2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan) (pp. 1-2). IEEE.

[13] Peng, S. (2016, January). The Research on Integration of Music and Technology in Guzheng Teaching. In 2016 2nd International Conference on Education Technology, Management and Humanities Science (pp. 133-136). Atlantis Press.

[14] Adiga, N., Tsiaras, V., & Stylianou, Y. (2018, April). On the use of WaveNet as a statistical vocoder. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5674-5678). IEEE.

[15] Li, S., Xu, K., & Zhang, H. (2018, May). Research on virtual Guzheng based on Kinect. In AIP Conference Proceedings (Vol. 1967, No. 1, p. 040010). AIP Publishing LLC.

[16] Mazurek, P., & Oszutowska-Mazurek, D. (2020, December). String Plucking and Touching Sensing using Transmissive Optical Sensors for Guzheng. In 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV) (pp. 1143-1149). IEEE.

[17] Shen, J., Wang, R., & Shen, H. W. (2020). Visual exploration of latent space for traditional Chinese music. Visual Informatics, 4(2), 99-108.

[18] Schedl, M. (2019). Deep learning in music recommendation systems. Frontiers in Applied Mathematics and Statistics, 5, 44.

[19] Tan, X., & Li, X. (2021, October). A Tutorial on AI Music Composition. In Proceedings of the 29th ACM International Conference on Multimedia (pp. 5678-5680).

[20] Martin-Gutierrez, D., Peñaloza, G. H., Belmonte-Hernandez, A., & García, F. Á. (2020). A multimodal end-to-end deep learning architecture for music popularity prediction. IEEE Access, 8, 39361-39374.

[21] Chen, G. F. (2021, June). Music sheet score recognition of Chinese Gong-che notation based on Deep Learning. In 2021 International Conference on Big Data Analysis and Computer Science (BDACS) (pp. 183-190). IEEE.

[22] Fessahaye, F., Perez, L., Zhan, T., Zhang, R., Fossier, C., Markarian, R., ... & Oh, P. (2019, January). T-recsys: A novel music recommendation system using deep learning. In 2019 IEEE international conference on consumer electronics (ICCE) (pp. 1-6). IEEE.

[23] Brand, M. (2001). Chinese and American music majors: Cross-cultural comparisons in motivation and strategies for learning and studying. Psychology of Music, 29(2), 170-178.

[24] Chen, P. H., Liang, K. W., & Chang, P. C. (2020, September). Music Conversion from Synthetic Piano to Chinese Guzheng Using Image-based Deep Learning Technique. In 2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan) (pp. 1-2). IEEE.

[25] Briot, J. P., & Pachet, F. (2020). Deep learning for music generation: challenges and directions. Neural Computing and Applications, 32(4), 981-993.

[26] Kulkarni, R., Gaikwad, R., Sugandhi, R., Kulkarni, P., & Kone, S. (2019). Survey on deep learning in music using GAN. International Journal of Engineering Research & Technology, 8(9), 646-648.

[27] Cai, Z. (2020). Usage of deep learning and blockchain in compilation and copyright protection of digital music. IEEE Access, 8, 164144-164154.

[28] Carnovalini, F., Rodà, A., Harley, N., Homer, S. T., & Wiggins, G. A. (2021). A New Corpus for Computational Music Research andA Novel Method for Musical Structure Analysis. In Audio Mostly 2021 (pp. 264-267).

[29] Guan, F., Yu, C., & Yang, S. (2019, July). A gan model with self-attention mechanism to generate multi-instruments symbolic music. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-6). IEEE.

[30] Tsivian, M., Qi, P., Kimura, M., Chen, V. H., Chen, S. H., Gan, T. J., & Polascik, T. J. (2012). The effect of noise-cancelling headphones or music on pain perception and anxiety in men undergoing transrectal prostate biopsy. Urology, 79(1), 32-36.

[31] Li, S., Jang, S., & Sung, Y. (2019). Automatic melody composition using enhanced GAN. Mathematics, 7(10), 883.