Weining WANG, Jiahui LI, Yifan LI, Xiaofen XING, 2024. Style-conditioned music generation with Transformer-GANs. *Frontiers of Information Technology & Electronic Engineering*, 25(1):106-120.
https://doi.org/10.1631/FITEE.2300359

Style-conditioned music generation with Transformer-GANs

Key words: Music generation; Style-conditioned; Transformer; Music emotion

First author: Weining WANG
E-mail: wnwang@scut.edu.cn
ORCID: https://orcid.org/0009-0006-0589-8157

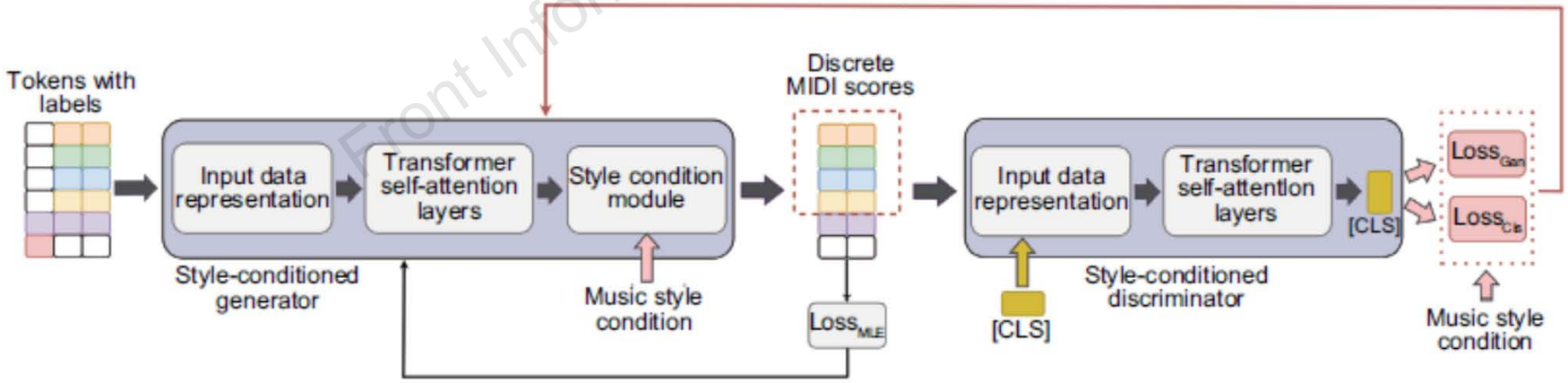
Music generation

- Recently, various algorithms have been developed for generating appealing music. Researchers are committed to studying two fundamental attributes: structural awareness and interpretive ability.

Algorithm	Main ideas
Music Transformer	Concerned with learning long musical sequences, music Transformer aims to recognize the varying degrees of repetition and variation between notes.
Museformer	Designing fine-grained and coarse-grained attention for the music generator, Museformer aims to better model the note sequence.
MG-VAE	Focusing on the stylistic differences of Chinese folk songs, MG-VAE implements the control of generated folk songs through special stylistic notation.
CPS	Concerned with exploring how the model can be controlled, CPS designs a new data representation method with style information for style-conditioned generation.

Method

- We propose a model for music generation under style control, namely style-conditioned Transformer-GANs (SCTG). Style information is effectively used to enhance the interpretive ability of the model and a style-conditioned patch discriminator is designed to consider structural awareness.
- Two specific losses, the music style category loss (for interpretive ability) and the music style information adversarial loss (for structural awareness), are designed in the discriminator.



Major results

- We use some existing objective (PR, NPC, SC) and subjective (H, R, O) metrics to evaluate the effectiveness of our model.
- We use a pre-trained music style classification model to evaluate the style consistency (CA). We also propose a new metric SD for it.
- We choose two widely used datasets for our experiments.

Dataset	Model	|PR|	|NPC|	|SC|	SD	CA	H	R	O
EMOPIA	Emopia	2.07	1.54	0.0012	16.62	52.5%	3.13	3.39	3.25
	Sulun et al. (2022)'s	4.52	1.63	0.0044	24.89	63.0%	3.46	3.17	3.52
	SCTG (ours)	**0.95**	**1.34**	**0.0012**	**12.45**	**69.5%**	**3.65**	**3.60**	**3.75**
Pianst8	Emopia	1.39	**0.88**	0.043	33.47	29.5%	3.22	3.34	3.15
	Sulun et al. (2022)'s	1.44	1.41	0.038	39.05	54.0%	3.39	3.45	3.44
	SCTG (ours)	**1.30**	0.90	**0.036**	**29.37**	**67.0%**	**3.92**	**3.81**	**3.96**

Major results (Cont'd)

- In the analysis of musical emotions, the arousal of music can be easily observed based on note density and note length. Thus, we also compare these metrics for the generated music with the original training data.

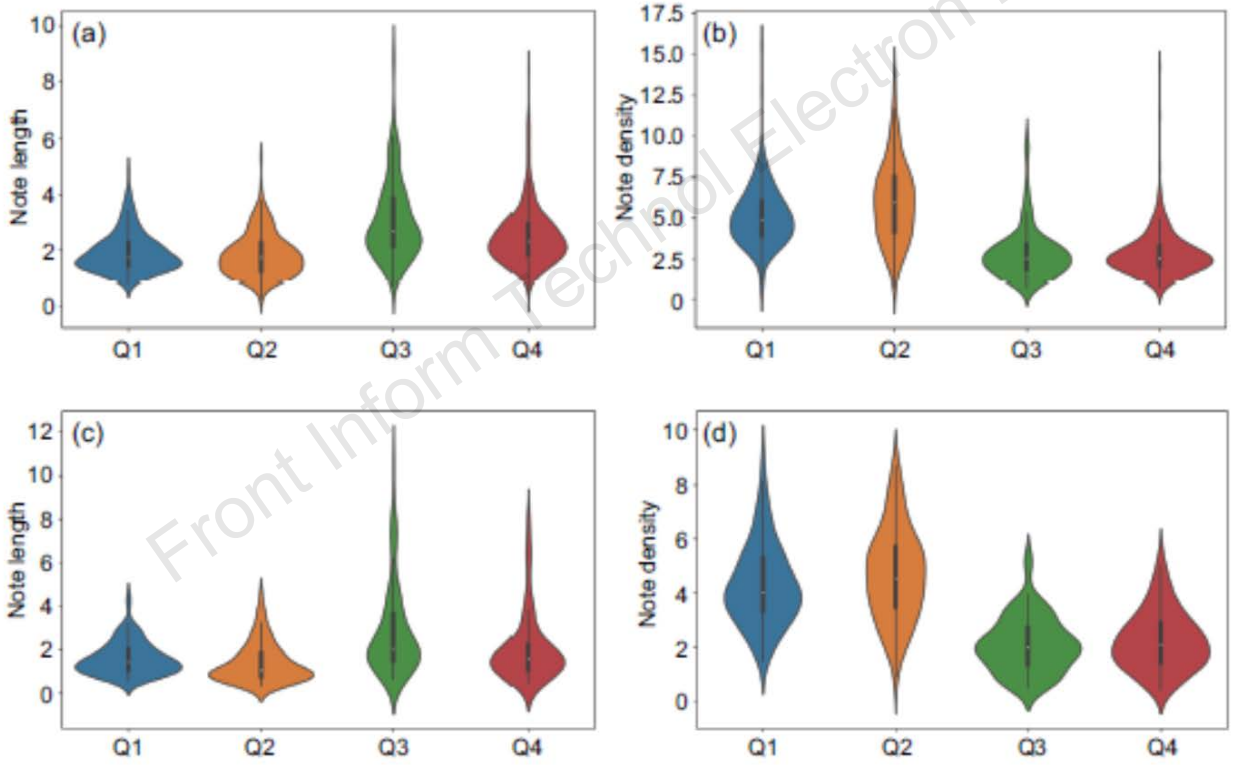


Fig. 5 Note length and note density distributions of the original dataset EMOPIA and the generated data: (a) note length of the original data; (b) note density of the original data; (c) note length of the generated data; (d) note density of the generated data

Conclusions

- We have proposed a novel music generation model, SCTG, which can generate a complete musical composition from scratch based on a specified target style.
- We innovatively embedded style information in our proposed style-conditioned linear Transformer. We also designed a style-conditioned patch discriminator with two innovative losses to enhance the interpretive ability and structural awareness of the model.
- In addition, to evaluate the style consistency, we defined a metric for the first time. Extensive experiments on two public datasets showed the effectiveness of our approach.
- In the future, we will build a richer dataset and continue to explore style-conditioned music generation. We will also further investigate the influence of style information on the inference stage of the model.

References

[1] Huang CZA, Vaswani A, Uszkoreit J, et al., 2019. Music Transformer: generating music with long-term structure. Proc 7th Int Conf on Learning Representations.

[2] Huang YS, Yang YH, 2020. Pop music Transformer: beat-based modeling and generation of expressive pop piano compositions. Proc 28th ACM Int Conf on Multimedia, p.1180-1188. https://doi.org/10.1145/3394171.3413671

[3] Yu BT, Lu PL, Wang R, et al., 2022. Museformer: Transformer with fine- and coarse-grained attention for music generation. Proc 36th Conf on Neural Information Processing Systems, p.1376-1388.

[4] Luo J, Yang XY, Ji SL, et al., 2020. MG-VAE: deep Chinese folk songs generation with specific regional styles. Proc 7th Conf on Sound and Music Technology, p.93-106. https://doi.org/10.1007/978-981-15-2756-2_8

[5] Wang WP, Li XB, Jin C, et al., 2022. CPS: full-song and style-conditioned music generation with linear transformer. Proc IEEE Int Conf on Multimedia and Expo Workshops, p.1-6. https://doi.org/10.1109/ICMEW56448.2022.9859286

[6] Hung HT, Ching J, Doh S, et al., 2021. EMOPIA: a multimodal pop piano dataset for emotion recognition and emotion-based music generation. Proc 22nd Int Society for Music Information Retrieval Conf, p.318-325.

[7] Sulun S, Davies MEP, Viana P, 2022. Symbolic music generation conditioned on continuous-valued emotions. *IEEE Access*, 10:44617-44626. https://doi.org/10.1109/ACCESS.2022.3169744

[8] Dong HW, Yang YH, 2018. Convolutional generative adversarial networks with binary neurons for polyphonic music generation. Proc 19th Int Society for Music Information Retrieval Conf, p.190-196.

[9] Yang LC, Lerch A, 2020. On the evaluation of generative models in music. *Neur Comput Appl*, 32(9):4773-4784. https://doi.org/10.1007/s00521-018-3849-7

[10] Zhang N, 2023. Learning adversarial transformer for symbolic music generation. *IEEE Trans Neur Netw Learn Syst*, 34(4):1754-1763. https://doi.org/10.1109/TNNLS.2020.2990746

Weining WANG, associate professor and master's supervisor at the School of Electronic and Information Engineering, South China University of Technology. Her research interests include computer vision, machine learning, multimedia information processing, and medical image processing. She has published over 40 papers, obtained 14 national invention patents, and held 7 software copyrights. She has served as a reviewer for journals such as *Neurocomputing*, *Image and Vision Computing*, *Journal of Automation*, *Journal of Computer-Aided Design & Computer Graphics*, and *Journal of Computer Science*. She is a member of the Pattern Recognition Committee of the Chinese Association for Artificial Intelligence, as well as a member of IEEE, CCF, and the Guangdong Society of Image and Graphics. She has led projects funded by the National Natural Science Foundation, Guangdong Natural Science Foundation, and the Ministry of Education Engineering Research Center Open Project, among others.