ChordGAN: Symbolic Music Style Transfer with Chroma Feature Extraction

Conan Lu, Washington

ABSTRACT

This project introduces ChordGAN, a generative adversarial network that transfers the style elements of tonal music genres. Early attempts at symbolic music style transfer have faced the challenge of preserving specific content features like chords while simultaneously changing style, which includes aspects like rhythm and ornamentation. ChordGAN seeks to learn the rendering of harmonic structures into notes using a novel process that embeds chroma feature extraction within the training process. In notated music, the chroma representation approximates chord notation as it only takes into account the pitch class of musical notes, representing multiple notes collectively as a density of pitches over a short time period. Chroma is used in this work to distinguish critical style features from content features and improve the consistency of transfer. ChordGAN uses conditional GAN architecture and appropriate loss functions, paralleling image-to-image translation algorithms. Pop, jazz, and classical datasets were used for training purposes. To evaluate the success of the transfer, I used two metrics: Tonnetz distance, to measure harmonic similarity, and a separate genre classifier, to measure transfer realism. The success of the transfer is evidenced by the high independent genre classifier accuracy rate and near-zero Tonnetz distance. Using chroma feature extraction offers significant advantages, as the format improves transfer consistency while requiring less data to train. ChordGAN can be utilized as a tool for musicians to study compositional techniques. In addition, experiments using lead sheets in conjunction with ChordGAN present the possibility of automatic music generation.

MOTIVATION

The concept of **style transfer** has intrigued both musicians and artificial intelligence researchers. In essence, the objective is to use a model to reimagine a piece of music in one style as a target style by modifying the note configurations. Think: pop song as a sonata. The base logic is that a piece should retain is **content features** while changing its learned **style features**, however, defined.

Previous implementations (MIDI-VAE, CycleGAN) do not explicitly encode the difference between content and style features.

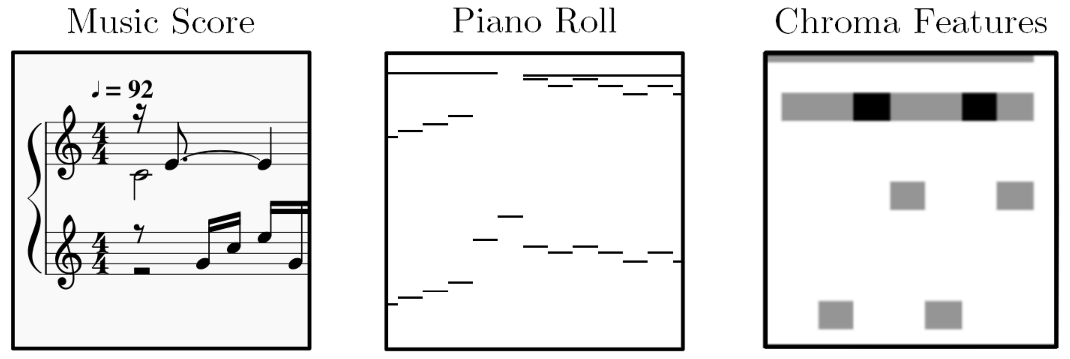
PROPOSAL

ChordGAN accomplishes style transfer in the symbolic domain of music using a novel technique, combining conditional generative adversarial network (GAN) architecture and the chromagram representation of music.

I focus my lens of compositional style as the rendering of harmonic structure into notes.

The **chromagram** is a pitch class vector that outlines the density of each chromatic note over a series of measures. As the harmonic structure of a piece of music, this format can be used as a base for learning style.

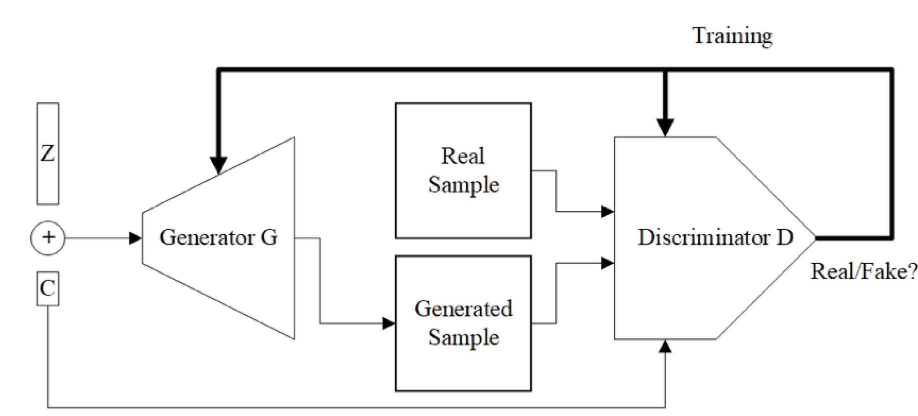
Music Score Piano Roll Chroma Features



MODEL ARCHITECTURE

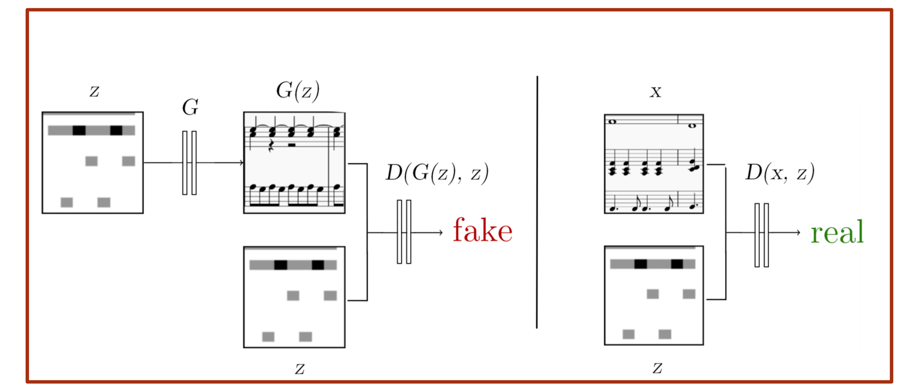
Base GAN

The Conditional GAN architecture features two networks: the **generator** and **discriminator**. The generator's objective is to create the most realistic music sample, while the discriminator tries to distinguish between real and generated samples.



Unique Additions

- Conditional GAN uses additional parameter (chroma) for discriminator. I used chroma, allowing model to map specific harmonic structures to notes.
- Generator input (z) changed from random noise to chroma.



IMPLEMENTATION

Dataset

Datasets used genres of pop, general classical, general jazz, Bach, Mozart, and Haydn. Pop, jazz, and Bach evaluated in-depth.
100 sixteen-measure MIDI files each.

Pre-Processing

MIDI converted into piano roll and chroma format with Python (Pretty MIDI library and custom scripts). All pieces with non-standard time signatures were filtered out.

Training

EVALUATIONS

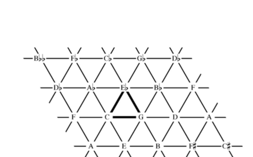
Content Evaluation: Tonnetz

I use **Tonnetz** to measure the content (defined as harmonic structure) similarity of pieces before and after transfer. Tonnetz is a graph representation of the harmonies within a piece of music, and the distance between two Tonnetz matrices is the Tonnetz distance. A lower Tonnetz distance between two pieces of music signifies more similar harmonic structures.

Tonnetz distances for fifty experimental (post-transfer / pre-transfer) pairs and fifty control (post-transfer / random) pairs in each target genre were calculated. Results shown in Table 1.

	Experimental	Control
Pop	0.000	0.044
Jazz	0.000	0.048
Classical	0.000	0.049

Table 1



Style Evaluation: Genre Classifier

Style evaluation is measured via an independent genre classifier. The classifier is a convolutional neural network (CNN) that uses the spectral data and piano roll to categorize music into different genres.

With independent data, genre prediction accuracy was 80%. Once classifier was finished training, fifty post-transfer samples were given as input. Results shown in Table 2.

Baseline	80%
Pop	68%
Jazz	74%
Classical	64%

Table 2

DEMO EXAMPLE

Original: Pop Sample *(Katy Perry, I Kissed A Girl)*

Output: Classical Transfer *(Bach Preludes)*

CONCLUSION

Based on style and content evaluations, ChordGAN presents a successful novel approach to style transfer in the symbolic domain. It specifically targets the style involved in rendering harmonic structures to notes. Limitations include narrow scope of style—no long-term structure—and uneven training for genres like jazz.

This model presents significant implications for accessible computational creativity, as it requires relatively small amounts of data to draw inferences while maintaining consistency.

One major application of this work is in automatic music generation, especially in collaboration with humans. By substituting chroma for lead sheets and initial chords created by a user, ChordGAN has the capacity to be a useful compositional tool. Beyond music, the novel split between style and content features can improve models for areas like speech synthesis and image manipulation.

REFERENCES

Argamon, S., Burns, K., & Dubnov, S. (2010). The structure of style: Algorithmic approaches to understanding manner and meaning. Springer Publishing Company, Incorporated. doi: 10.5555/1869899

Brunner, G., Konrad, A., Wang, Y., & Wattenhofer, R. (2018). Midi-vae: Modeling dynamics and instrumentation of music with applications to style transfer. Brunner, G., Wang, Y., Wattenhofer, R., & Zhao, S. (2018).

Symbolic music genre transfer with cyclegan. Cambouropoulos, E. (2016).

The harmonic musical surface and two novel chord representation schemes. In D. Meredith (Ed.), Computational music analysis (pp. 31–56). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-25931-4 2 doi: 10.1007/978-3-319-25931-4 2

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). Generative adversarial networks. Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2018). Image-to-image translation with conditional adversarial networks.

Raffel, C., & Ellis, D. P. (2014). Intuitive analysis, creation and manipulation of midi data with pretty midi. In Proceedings of the 15th international conference on music information retrieval late breaking and demo papers (Vol. 15).

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2019). A hierarchical latent vector model for learning long-term structure in music.

Wang, C.-i., & Dubnov, S. (2014, 10). Guided music synthesis with variable markov oracle.. doi: 10.13140/2.1.2171.2329