

IMA_iGen: A Cross-Modal Image-to-Music Generation with a Non-Deep Learning Core

Artificial Intelligence Generated Content (AIGC) has grown rapidly, particularly with tools like ChatGPT. However, AI music generation lags behind AI art and writing due to its complex structure and required musical expertise. Current methods rely on deep learning, using large datasets for music generation, but face challenges including data collection and preparation, copyright risks, and high computing costs.

In developing a musically robust song generation method, knowledge of music theory, literature, and linguistics is required to emulate the intuitive thinking process of musical artists while not relying on existing music to train and develop such algorithms. Therefore, we propose a novel cross-modal Image-to-Music Artificial intelligence Generation (IMA_iGen) framework driven by an innovative non-deep learning music core. This music core takes in lyrical rhythmic information derived from images and purely utilizes novel algorithms while leveraging correlations between lyrics and music, as inspired by [1][2]. We developed a web tool, GenAIM, that allows users, regardless of their expertise, to generate music with customizable features such as key signatures, instruments, and sheet music display as shown in Figure 1. As a reliable co-pilot for composers and entertainment, this unique copyright-free approach to AI non-lyrical music generation from images combines latent rhythms, melodies, music theory, and composer intuition to create natural, human-sounding music that bridges visual and auditory art forms across modalities.

In the framework, different images are inputted into Large Language Models (LLMs) to automatically generate Lyrical Rhythmic Information (LRI) of multiple phrases. Preference prompts are added to customize the generated lengths before the LRI is employed to construct the rhythmic score. LRI is extracted to determine the time signature [3]. Phrases are identified through punctuation with accents and keywords influencing the total number of measures. Keywords are then inserted into stressed beats within each measure following the order specified by the lyrics. During the pitch construction, the insertion processes work in a feedback loop, with randomly generated pitches adapted to music theory before being refined for smoother transitions and varied phrasing. Upon generating the music, the architecture presents the results in a MIDI or MusicXML sheet music file format for better readability and compatibility. Our web tool, GenAIM, is built on the Amazon Web Services (AWS) platform, and displays and plays the sheet music.

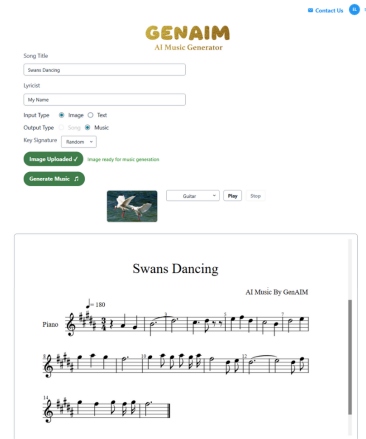


Figure 1: Image-to-music generation.

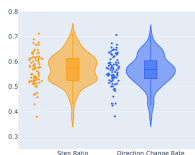


Figure 2: Melodic smoothness metrics and their scores.

To evaluate the generated music, we used 20 distinct images and the corresponding 75 AI-generated pieces as well as key confidence, the measure of the central tonal tendency that is derived from a metric of music21 [4]. The average key confidence is at 85.3%, indicating a higher alignment of the key signatures with the generated melodies. Figure 2 displays the results of two melodic smoothness metrics. A higher step ratio and a moderate direction change rate that corresponds to a smoother melody further shows that overall, our approach produces diverse, human-like compositions.

References:

- [1] C. C. Liao, D. Liao, and J. Guessford, "Multimodal lyrics-rhythm matching," in 2022 IEEE Int. Conf. on Big Data (BigData), 2022, pp. 3622–3630.
- [2] C. C. Liao, D. Liao, and E. L. Zhang, "Relationships between Keywords and Strong Beats in Lyrical Music," in 2024 IEEE BigData, 2024, pp. 3191–3199.
- [3] C. C. Liao, D. Liao, and J. Guessford, "Automatic time signature determination for new scores using lyrics for latent rhythmic structure," in 2023 IEEE Intl. Conf. on Big Data (BigData), 2023, pp. 4485–4494.
- [4] S. Cuthbert and C. Ariza, "Music21: A toolkit for computer-aided musicology and symbolic music data," in International Society for Music Information Retrieval, 2010, pp. 637–642.