Automatic Stage Lighting Control: Is it a Rule-Driven Process or Generative Task?

Zijian Zhao[1,5] (zzhaock@connect.ust.hk), Dian Jin[2,5], Zijing Zhou[3,5], Xiaoyu Zhang[4]*

[1]The Hong Kong University of Science and Technology [2]The Hong Kong Polytechnic University
[3]The University of Hong Kong [4]City University of Hong Kong [5]Tokamak Disruption Band

Problem Setup

Primary Light Generation: Given a music sequence $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$, find a mapping function $\mathrm{f}(\cdot)$ that transfers it to the corresponding stage light sequence $\mathcal{Y} = \{y_1, y_2, \ldots, y_n\}$, i.e., $\mathcal{X} \xrightarrow{\mathrm{f}} \mathcal{Y}$.

- n: Number of frames.
- $x_i \in \mathbb{R}^m$: Mel spectrum with m bands.
- $y_i \in \mathbb{R}^3_*$: Primary light represented in Hue, Saturation, Value (HSV) color space.

Previous Rule-Based Solutions

Classification models are used to divide music pieces into several categories (e.g., by style, chord, and emotion) and manually define stage light patterns for each category. These methods mainly suffer from the following shortcomings:

- **Low Interpretability with No Creativity:** The manually pre-defined light patterns are not only limited but also lack interpretability [7].
- **Limited to the Classification Model:**
  - **Coarse Grain:** For instance, some chord classification methods only support 24 major and minor chords, ignoring other colorful chords. Similarly, most style classifications support only broad categories like pop, jazz, and rock, overlooking finer categories. For instance, pop black metal and suicidal black metal have completely different expressions and styles but may both be classified as metal or even rock, sharing similar light patterns, which is unacceptable.
  - **Low Classification Performance:** Classification tasks in the Music Information Retrieval (MIR) field remain challenging due to limited open-source datasets [8]. Insufficient classification accuracy can significantly impact the overall system efficacy.

Our Contributions

- In this paper, we propose a novel perspective for Automatic Stage Lighting Control (ASLC) by framing it as a generative task rather than a rule-based classification process. To support this claim, we develop an end-to-end deep learning framework validated through both quantitative analysis and human evaluation. Results demonstrate that our method closely matches human lighting engineering performance, outperforming conventional rule-based solutions.
- We introduce Skip-BART, a novel end-to-end deep learning framework that addresses the aforementioned generative task for ASLC. Building on BART, our approach incorporates adapted embedding and head layers to support both music and light modalities. Additionally, we introduce a novel skip-connection module to enhance the relationship between music and light in a fine-grained context. We also implement pre-training and transfer learning mechanisms to improve model performance under limited training data. During inference, a Restricted Stochastic Temperature-Controlled (RSTC) sampling method is employed to ensure both diversity and stability in the generated results.
- We present the first self-collected stage lighting dataset, named Rock, Punk, Metal, and Core - Livehouse Lighting (RPMC-L^2). To address ethical and copyright concerns, we release only the processed features and ground truth labels.

Proposed End-to-End Solution

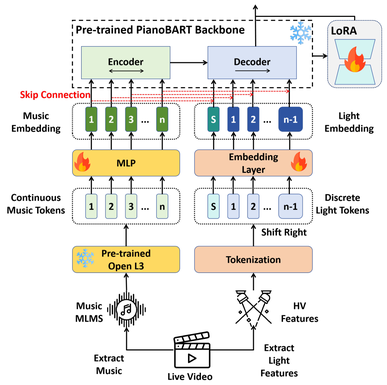

Fig. 1: Network Architecture

A. Network Architecture:

- **Feature Extractor**: OpenL3 [1] for music input & Embedding Layer for tokenized light input
- **Backbone**: Frozen pre-trained PianoBART [6] with trainable LoRA [5]
- **Skip-Connection Module**: Concatenate music and light within each frame before inputting to decoder, enhancing the rhythm alignment

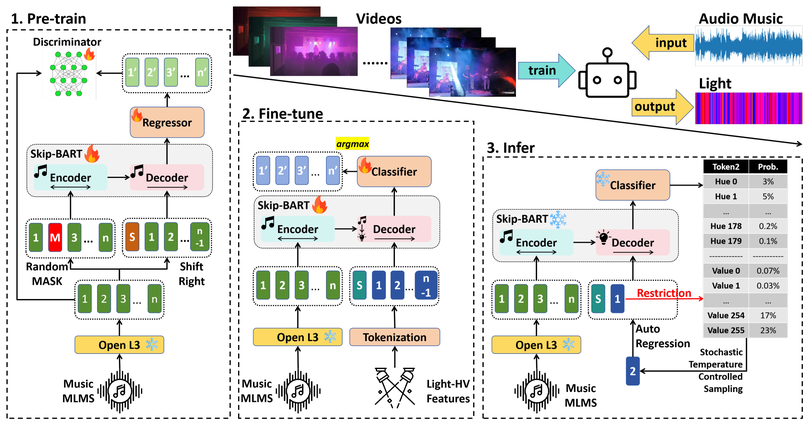

Fig. 2: Workflow

B. Workflow:

- **Pre-training:** Music MLM pre-training enhanced by adversarial learning [9], which is utilized to distinguish between the recovered sequence and the original sequence. (Note: Light data is not utilized in this stage to avoid information leakage.)
- **Fine-tuning:** End-to-end training with the maximum likelihood estimation (MLE) objective (θ: network parameters):

$$\theta^* = \arg\max_{\theta} \mathbf{E}_{\mathcal{X},\mathcal{Y},t}[\log \mathrm{P}(y_t|\mathcal{X}, y_{1:t-1}; \theta)] \tag{1}$$

- **Inference:** Add range restrictions to Stochastic Temperature-Controlled sampling [4] to avoid overshooting and model degradation.

Fig. 3: Light Example

Experiment

A. Quantitative Analysis: We first measure the similarity between different methods and the ground truth using RMSE and MAE. The results demonstrate that our proposed Skip-BART achieves superior performance compared to previous rule-based methods and a series of ablation studies, suggesting its efficiency.

Method	Hue RMSE	Value RMSE	Hue MAE	Value MAE
Rule-based [3]	48.67	93.39	43.43	86.55
Skip-BART	**36.13**	**60.74**	**28.72**	**51.27**
w/o skip connection	36.89	68.33	29.44	58.34
w/o light embedding	51.04	<u>67.25</u>	41.50	54.87
train from scratch	<u>36.63</u>	67.49	<u>28.83</u>	57.22
pre-train w/o [MASK]	49.97	64.45	42.07	52.63
pre-train w/o discriminator	50.40	68.09	41.52	56.54

Tab. 1: Quantitative results. Bold = best, underline = second best.

B. Human Study: We then conduct a human study, asking 38 participants to evaluate the lighting produced by different methods across six dimensions [2], scoring from 1 to 7 (the higher, the better). The results show that our Skip-BART achieves performance similar to that of real human engineers, significantly outperforming previous rule-based methods.

Method	Emotion	Impact	Rhythm	Smoothness	Atmosphere	Surprise
Ground Truth	4.50	**4.48**	**4.61**	**4.62**	**4.49**	**4.34**
Rule-based [3]	3.12	2.65	2.54	2.56	2.77	2.35
Skip-BART	**4.69**	<u>4.39</u>	4.50	4.32	<u>4.32</u>	<u>3.83</u>
w/o skip connection	4.31	3.78	<u>4.54</u>	<u>4.43</u>	4.11	3.50

Tab. 2: Human evaluation result. Bold = best, underline = second best.

References

[1] Aurora Linh Cramer et al. "Look, listen, and learn more: Design choices for deep audio embeddings". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 3852–3856.

[2] Matthias Erdmann, Markus von Berg, and Jochen Steffens. "Development and evaluation of a mixed reality music visualization for a live performance based on music information retrieval". In: *Frontiers in Virtual Reality* 6 (2025), p. 1552321.

[3] Shih-Wen Hsiao, Shih-Kai Chen, and Chu-Hsuan Lee. "Methodology for stage lighting control based on music emotions". In: *Information sciences* 412 (2017), pp. 14–35.

[4] Wen-Yi Hsiao et al. "Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 1. 2021, pp. 178–186.

[5] Edward J Hu et al. "Lora: Low-rank adaptation of large language models." In: *ICLR* 1.2 (2022), p. 3.

[6] Xiao Liang et al. "Pianobart: Symbolic piano music generation and understanding with large-scale pre-training". In: *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 2024, pp. 1–6.

[7] James McDonald et al. "Illuminating music: Impact of color hue for background lighting on emotional arousal in piano performance videos". In: *Frontiers in Psychology* 13 (2022), p. 828699.

[8] Khalid Zaman et al. "A survey of audio classification using deep learning". In: *IEEE access* 11 (2023), pp. 106620–106649.

[9] Zijian Zhao et al. "CSI-BERT2: A BERT-inspired Framework for Efficient CSI Prediction and Classification in Wireless Communication and Sensing". In: *IEEE Transactions on Mobile Computing* (2025).