

Coloring Music: Bridging Music and Color Palettes for Graphic Design

Takayuki Nakatsuka, Masahiro Hamasaki, Masataka Goto
National Institute of Advanced Industrial Science and Technology (AIST)

✉ takayuki.nakatsuka@aist.go.jp

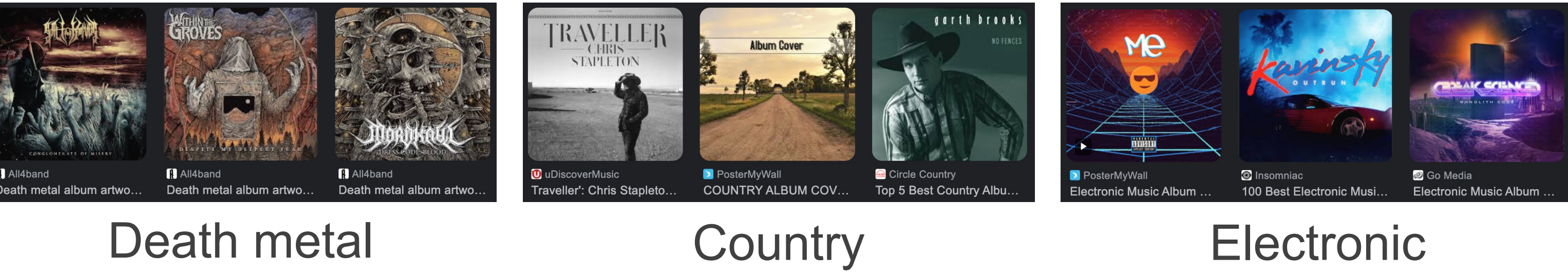


Key Contributions

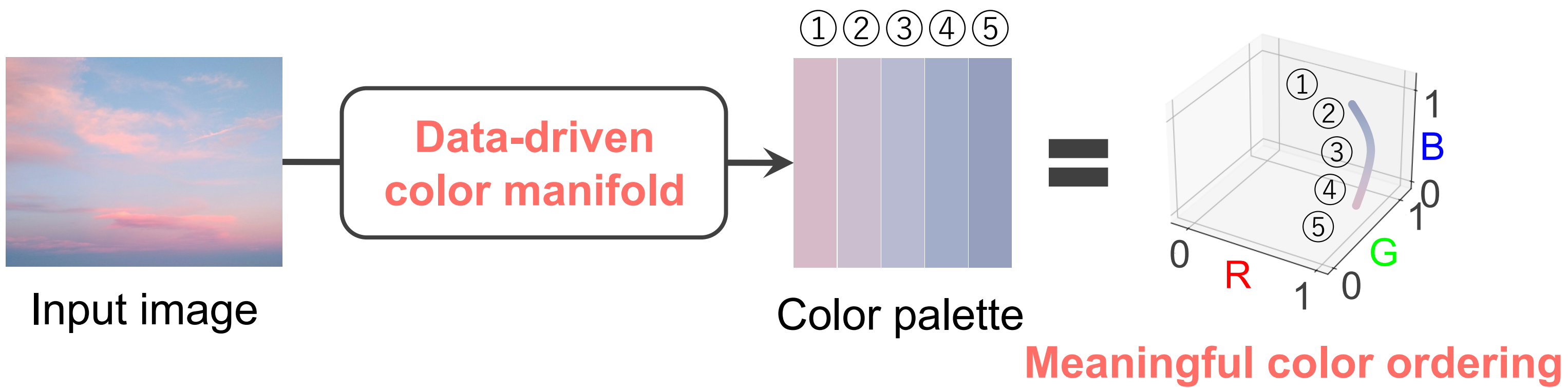
- A) Our study finds **the relationship between auditory and visual expression**, specifically through the colors chosen for music cover images.
- B) Our proposed framework employs deep learning techniques to **estimate appropriate color palettes from musical pieces** for graphic design.
- C) Our framework can **streamline the design process** by automating or helping the process of color selection.

A) Relationship Between Music and Colors Used for Designing Music Cover Images

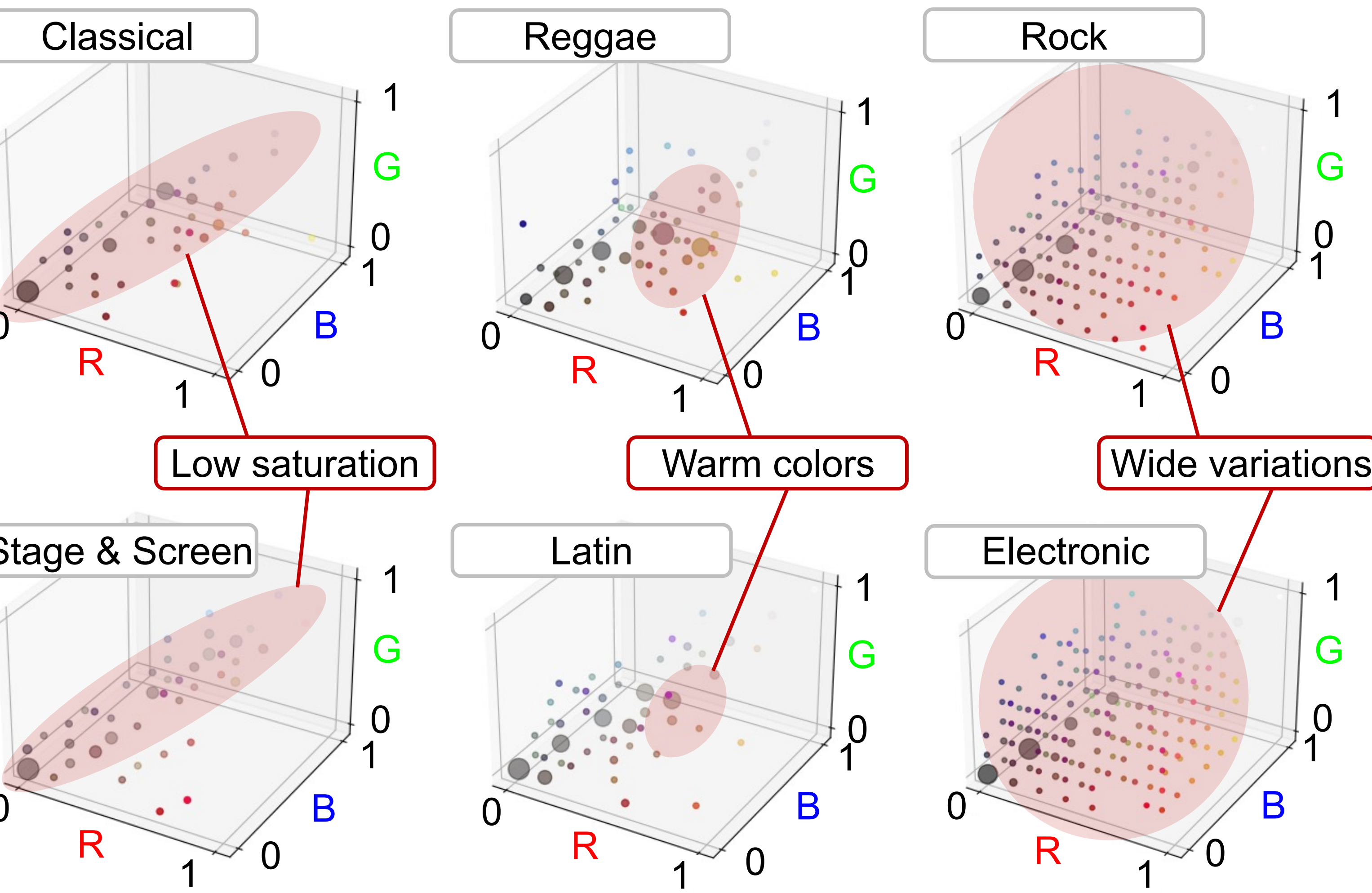
- Different music genres display **distinctive characteristics in the colors used in the cover images**, reflecting the characteristics of the corresponding music style.



- We investigate **the relationship between musical pieces and the color palettes** used in their corresponding cover image.
- To extract representative colors from cover images, we leverage **data-driven color manifold** [Nguyen+ '15].
- We here utilized **one representative color** extracted from each image for this preliminary study A).
- We used **1~5 colors** for the experiments in B).



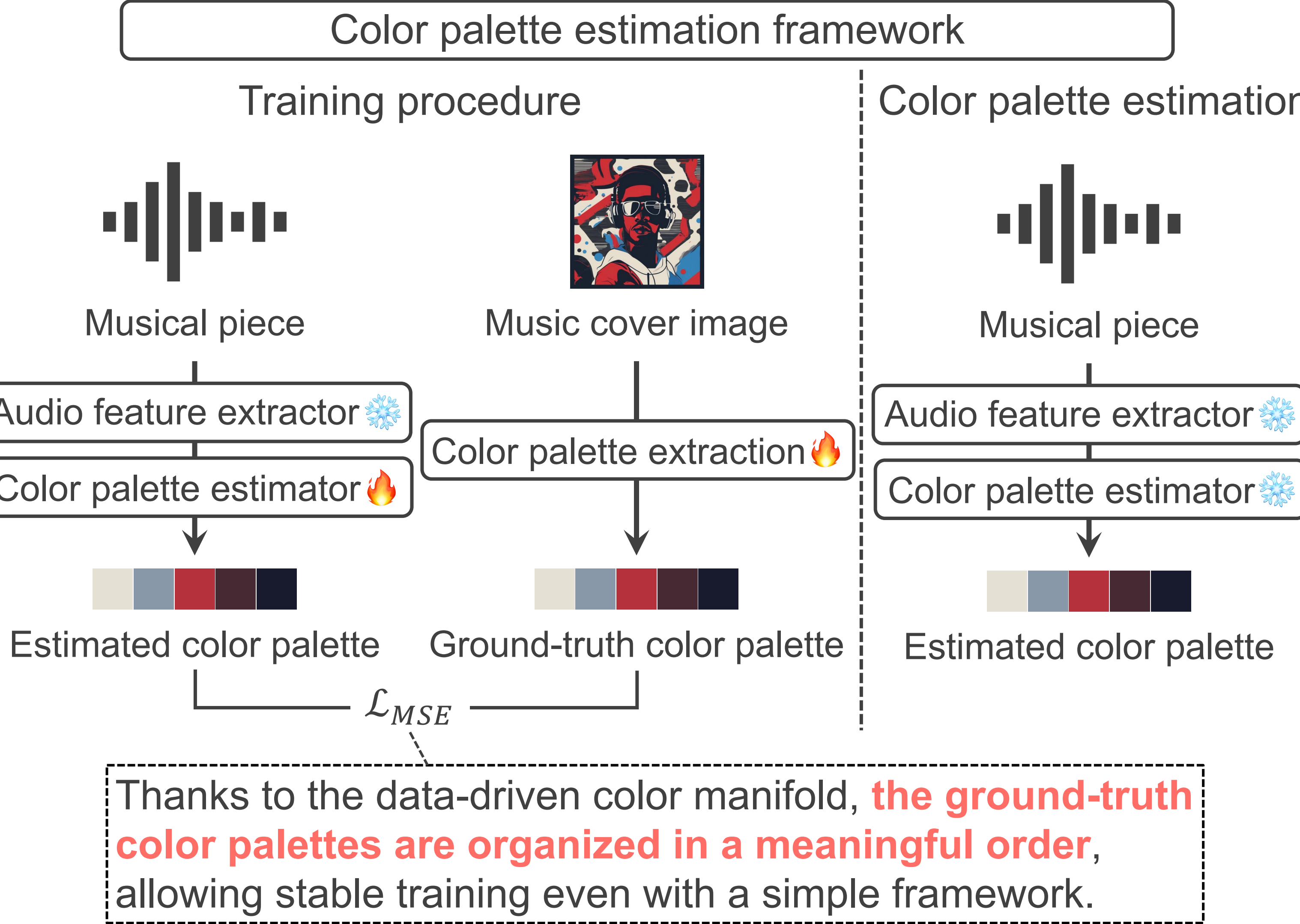
- We randomly collected 3,887 cover images (each image is an RGB image).
- We assigned genre tags to each image based on the grouping of genres and styles in Discogs.



- Classical/Stage & Screen cover images have **low saturation**.
- Reggae/Latin cover images tend to use **warm colors** such as red and yellow.
- Pop/Rock cover images feature a **wide variety of colors**. These genres have more diverse subcategories and styles than other genres, resulting in such color variation.

B) Color Palette Estimation Framework

- We propose a framework **designed to estimate appropriate color palettes from musical pieces** using the relationship between musical pieces and colors used in their corresponding music cover images.
- The proposed framework consists of three modules.
 1. Audio feature extractor: a pretrained audio model (we used **CLAP** [Wu+ '23] or **AudioToken** [Yariv+ '23]).
 2. Color palette estimator: an **MLP**.
 3. Color palette extraction: data-driven color manifold [Nguyen+ '15].



- We utilized the **minimum color difference model (MCDM)** to evaluate the color difference between the estimated and ground-truth color palettes.
- We used **582,894 pairs** of an excerpt and its corresponding cover image to train the color palette estimator.
- Our proposed framework achieves a **much lower (i.e., better) MCDM score** compared to random color palettes, with an improvement of **over 40 points**.
- Regarding the selection of audio model, the performance difference between the two is relatively small; therefore, either model can be selected depending on the intended application.

C) Example Application

- We showcase the potential of our framework through an **example application for creating music cover images** by leveraging the rich-text-to-image model [Ge+ '23, '25].
- The color attributes used in the rich-text prompts control the colors of target objects or regions within the initial images, thereby improving the visual fidelity of the generated results.

