MELODY MUSE

Karanvir Bath, Ritchie Kumar, Parmveer Dayal

1 Background

Our project explores the current capabilities in generating music through artificial intelligence with the added complexity of integrating with a visual component.

GOAL: Generate music that reflects the affective emotions behind an image.

2 Datasets

We utilized two datasets to drive our project: The EmoSet dataset, available online, features 118,000 samples categorized by emotion, with additional features such as brightness and colorfulness.

For music generation, we utilized a public Kaggle dataset categorized by emotion, with many acoustic features such as Tempo and MFCC.

Emotion: Fear
Brightness: 0.1
Colorfulness: 0.2

Example categorization of an image under EmoSet

3 Method

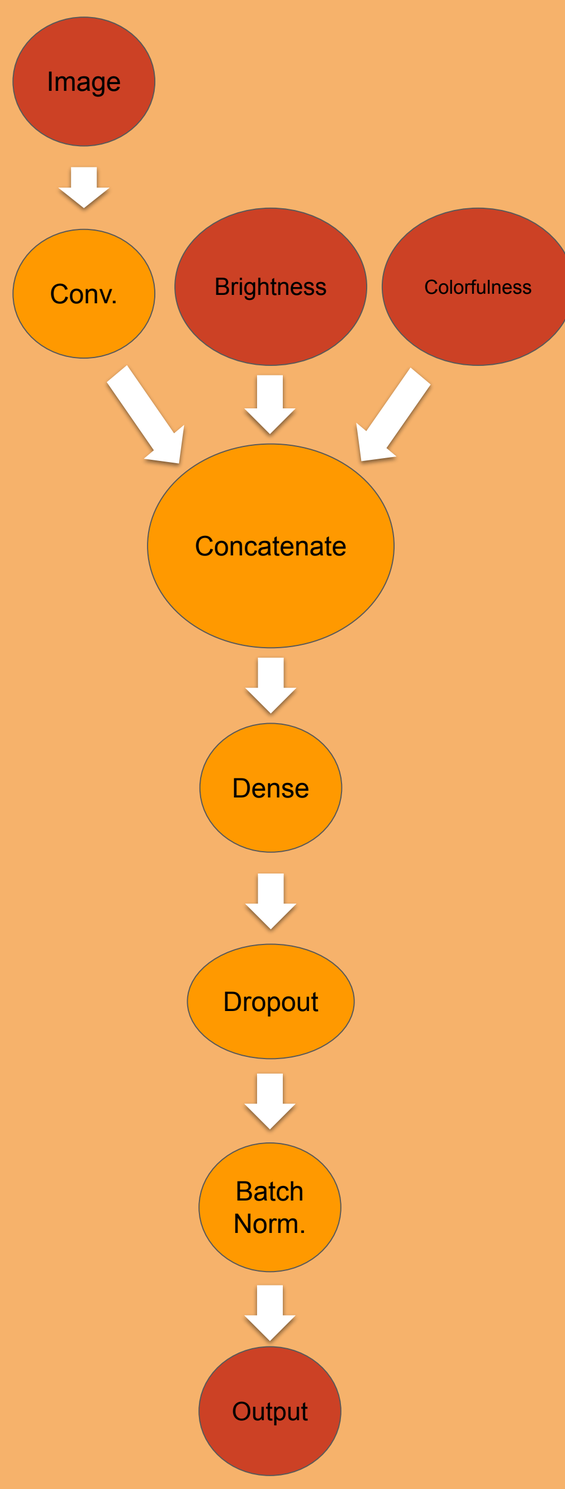
Our approach features two components: affective labelling of images and music generation of the label.

AFFECTIVE LABELLING

Using a convolutional neural network, we process the image and produce an output of: Amusement, Awe, Contentment, Excitement, Anger, Disgust, Fear and Sadness

MUSIC GENERATION

Using a convolutional neural network model, we take the outputted affective labelling as input and process the emotion to produce an output incorporating music associated to that emotion.

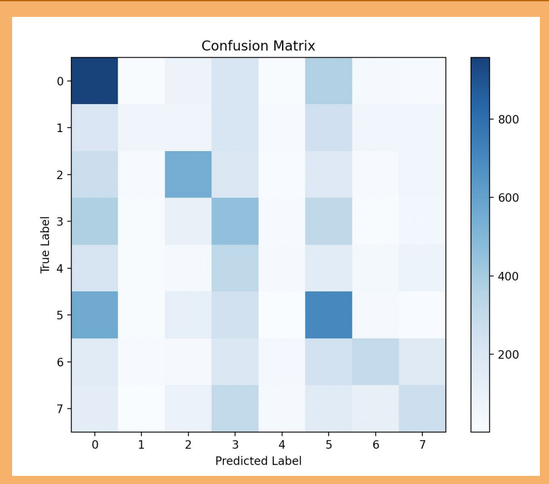


Simplified diagram of neural network for affective labelling

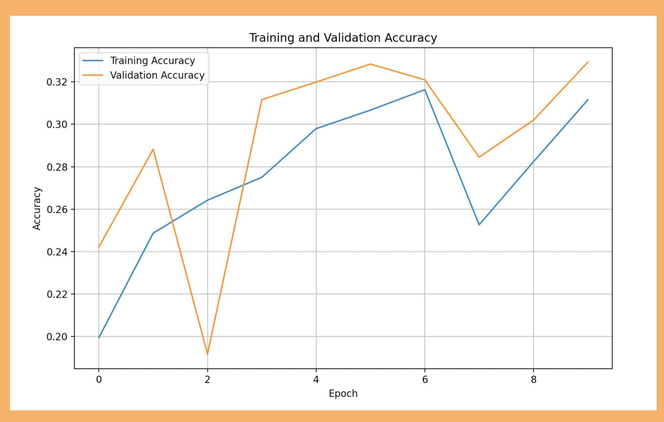
4 Results

AFFECTIVE LABELLING

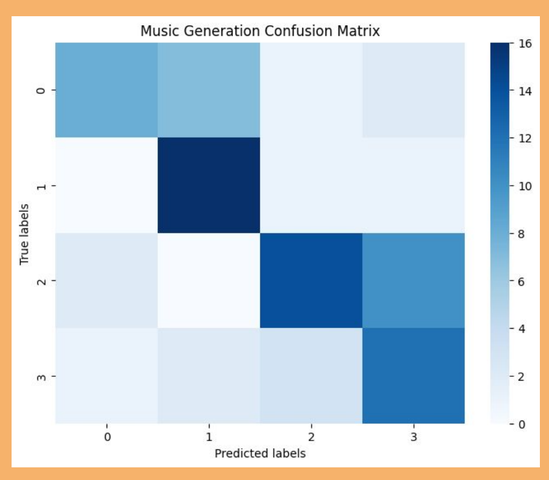
- Effectively predicts affective labels for a given image
- Has an F1 score of 31% when trained with 50,000 samples.
- Displays weak biases in confusion matrix for specific emotions



Confusion Matrix of Predicted Affective Labels



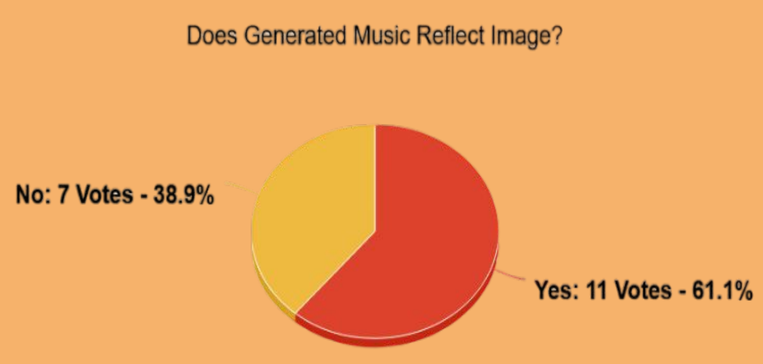
Accuracy Graph of Affective Labels Model



Confusion Matrix of Music Generation

MUSIC GENERATION

- Effectively generates music for a given emotion (~60% accuracy)
- Has an F1 score of 62% when trained with 400 samples
- Displays moderate biases in confusion matrix due to small test set



Results from Human Survey

References

1. Yang, J., Huang, Q., Ding, T., Lischinski, D., Cohen-Or, D., & Huang, H. (2023). EmoSet: A Large-scale Visual Emotion Dataset with Rich Attributes. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 20383–20394.
2. *Music+Notes+Transparent images – browse 8,456 stock photos, vectors, and video.* Adobe Stock. (n.d.). https://stock.adobe.com/search?k=music%2Bnotes%2Btransparent&asset_id=439874296
3. SVG > music notes musical - Free SVG Image & Icon. | SVG Silh https://svgsilh.com/image/294038.html
4. Arvidsson, J. (2023, October 18). *Turkish music emotion*. Kaggle. https://www.kaggle.com/datasets/joebeachcapital/turkish-music-emotion