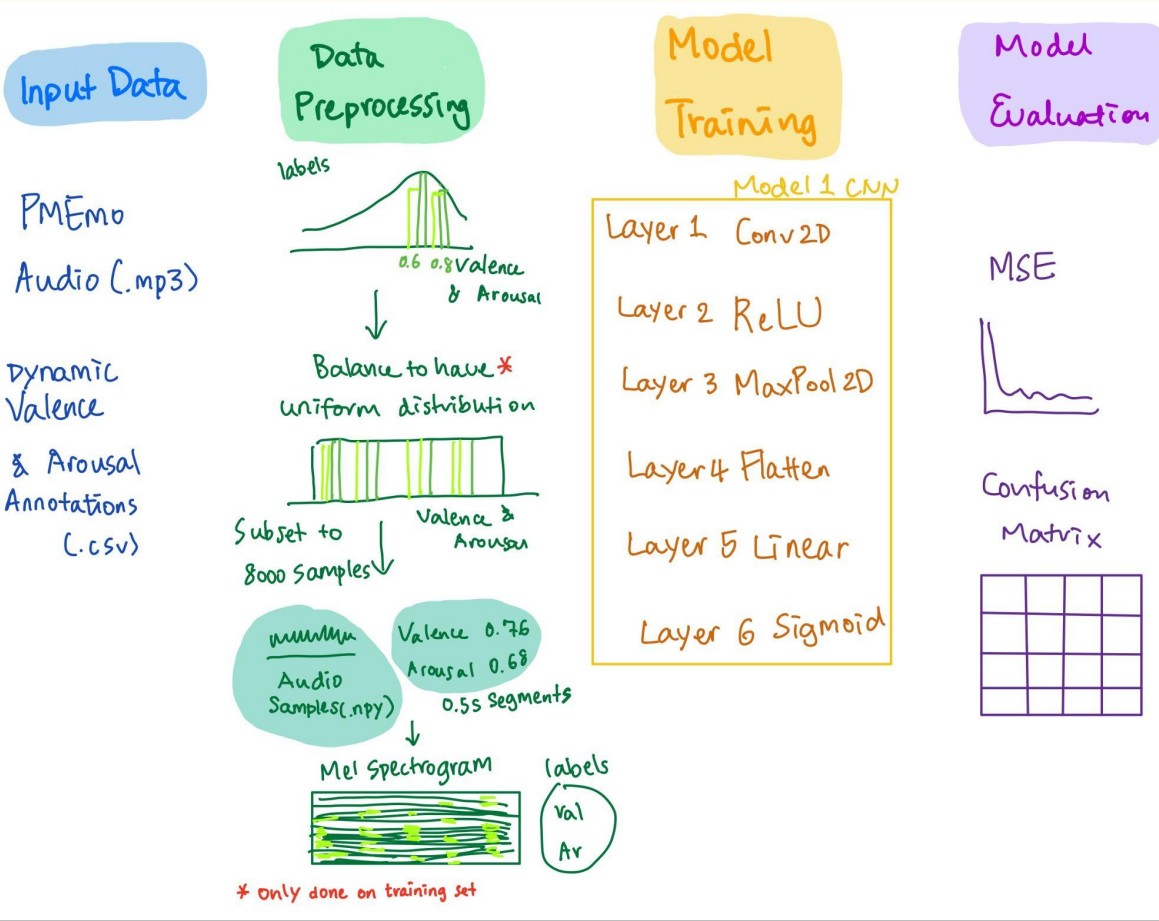Code folder link here

Static Emotion Recognition

—

By SAS.PY
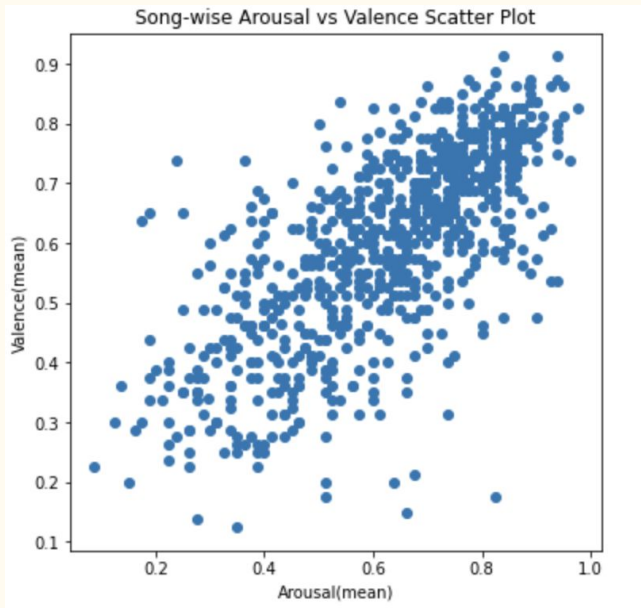
Aleksandra Ma, Stephanie Lew, Aneesh Athrey

Introduction

- Problem statement: given the chorus of a song, predict the emotion on a continuous scale of valence and arousal.
- Dataset found: PMEmo (Article | Dataset Github)
  - Contains relevant information and emotion annotations of 794 song chorus
  - Available information: raw audio, valence and arousal score (emotion annotations), lyrics, comments from soundcloud and netease, EDA signals, and song metadata
  - Annotations are made by 457 subjects in university, out of which 44 are music majors
- Related works done:
  - Na He, Sam Ferguson(2022) Music emotion recognition based on segment-level two-stage learning. *International Journal of Multimedia Information Retrieval*. Reference link here.
- Our approach:
  - Using a subset of the original training dataset (for computational purposes), pass in the mel spectrograms and emotion annotations into a cnn classifier, and compare the evaluation metrics
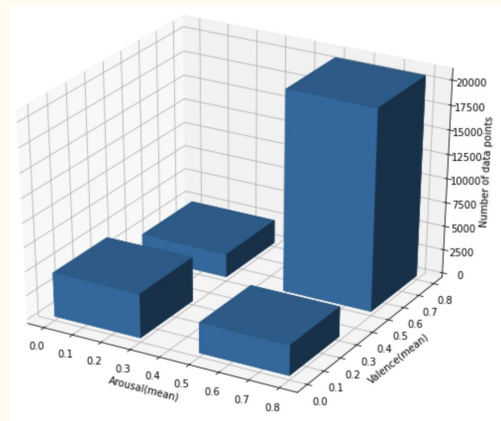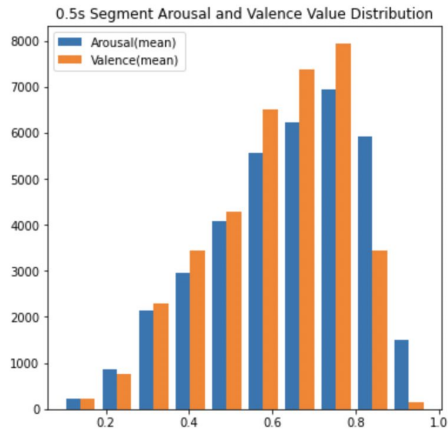
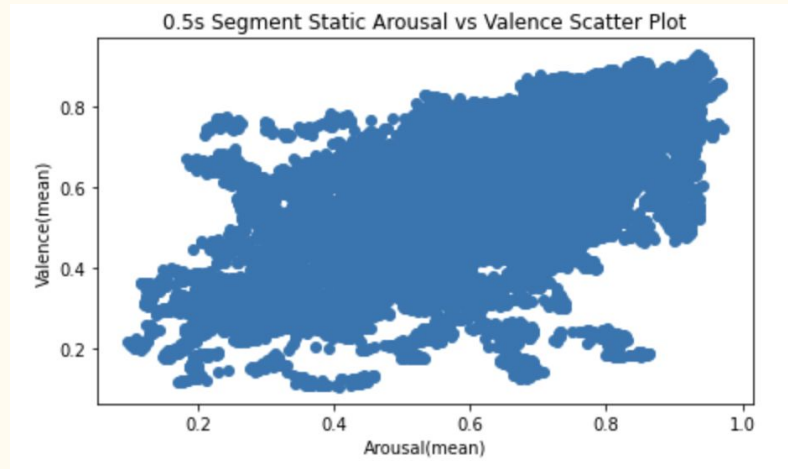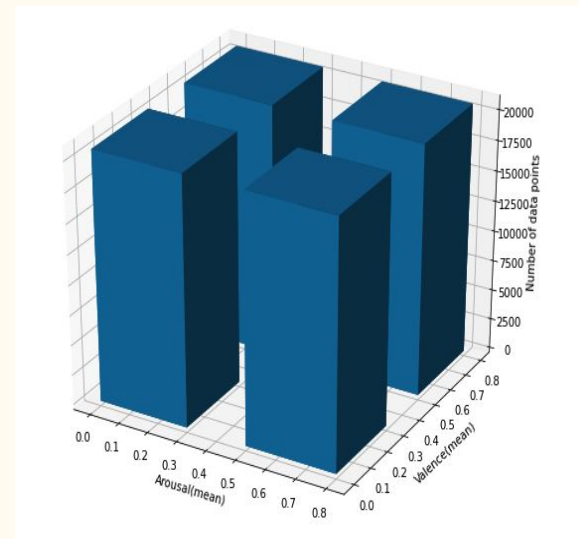# Approach

Dataset Discovery

Chorus Ar-Val Distribution

Song-wise Arousal vs Valence Scatter Plot

0.5s Segment Ar-Val Distribution

0.5s Segment Static Arousal vs Valence Scatter Plot

0.5s Segment Arousal and Valence Value Distribution

Data Rebalancing & Preprocessing

PMEmo

30968 samples of 0.5s segments

Subsets ~ 3000 samples

Train-Val -Test Split

Train (~2500)
Val (~450)
Test (~150)

Over sample training set to have uniform distribution

Balanced train (~7500)
Val (~450)
Test (~150)
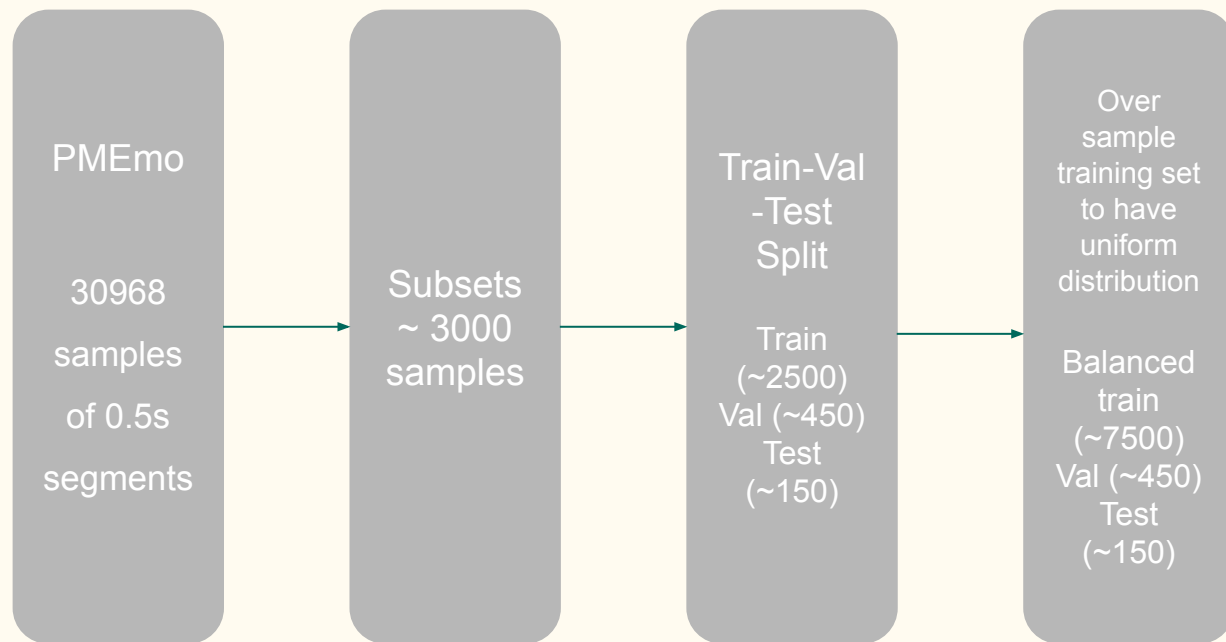
Model

Code folder link here

```
----------------------------------------------------------------
        Layer (type)            Output Shape         Param #
================================================================
          Conv2d-1          [-1, 4, 124, 40]             104
            ReLU-2          [-1, 4, 124, 40]               0
       MaxPool2d-3           [-1, 4, 62, 20]               0
         Flatten-4                [-1, 4960]               0
          Linear-5                   [-1, 2]           9,922
         Sigmoid-6                   [-1, 2]               0
================================================================
Total params: 10,026
Trainable params: 10,026
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.02
Forward/backward pass size (MB): 0.38
Params size (MB): 0.04
Estimated Total Size (MB): 0.44
----------------------------------------------------------------
```

Evaluation Metrics
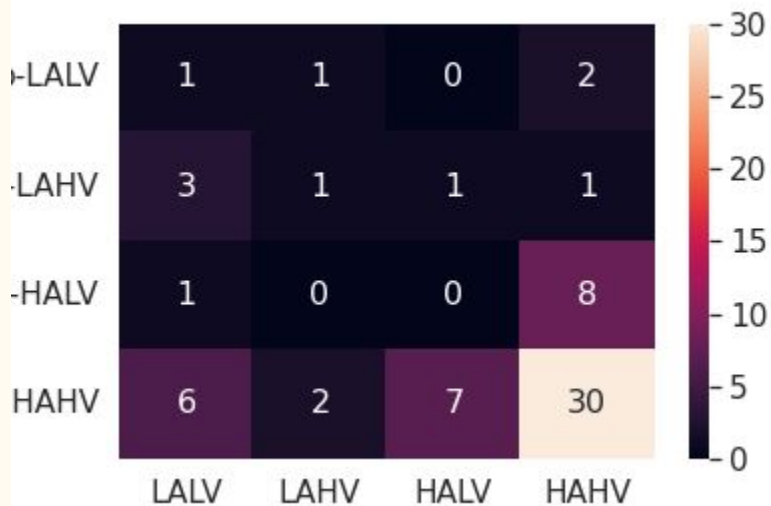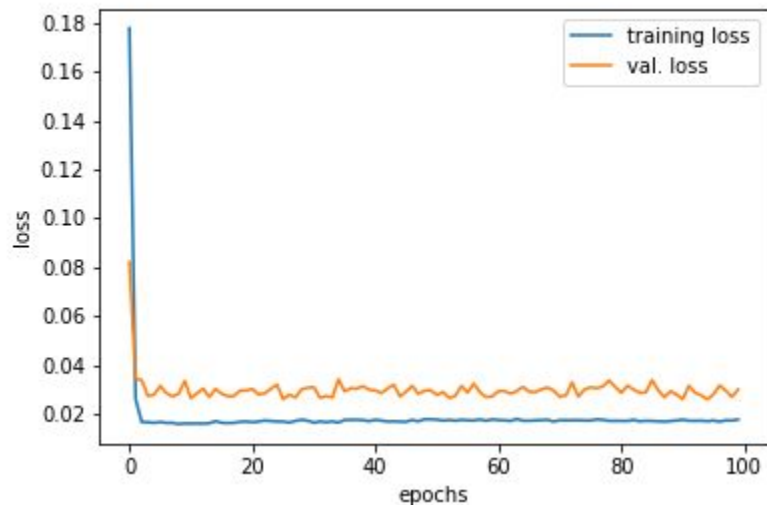
LALV: Low Arousal Low Valence
LAHV: Low Arousal High Valence

HALV: High Arousal Low Valence
HAHV: High Arousal High Valence

Improvement & Future Work

- Scale up the model and train with more samples

- Train with raw audio sample clips

- Experiment with more models and compare evaluation metrics

- Create custom evaluation metrics
  - Test our model by running a test song through the model and getting the dynamic labels (0.5s segment labels); then taking the average to see if it matches the static labels of the static emotion annotations