DS203

Group Project – E7

Nilay Shahapure 23b2206
Tejas Baid 23b2280
Dhruvam Parekh 23b2130
Sanidhya Katiyar 23b2244

Overview

The project aims at classifying songs into one of 6 genres namely -

- Renditions of the Indian National Anthem
- Marathi ‘Bhav Geet’
- Marathi Lavani songs
- Hindi film songs sung by Asha Bhosale
- Hindi film songs sung by Kishor Kumar
- English songs by Michael Jackson

Convolutional Neural Networks (CNN’s) have been utilised to do the same

MFCC Coefficients importance

- **Feature Extraction**: MFCCs capture audio's frequency characteristics based on the Mel scale, reflecting human hearing.
- **Process**: The steps include framing, Fourier Transform, Mel scale mapping, log magnitude, and DCT application.
- **Compact Representation**: Typically, the first 12-13 MFCCs give a compact summary of audio's spectral properties.
- **Focus on Low Frequencies**: MFCCs emphasize low frequencies, aiding in speech and sound distinction.

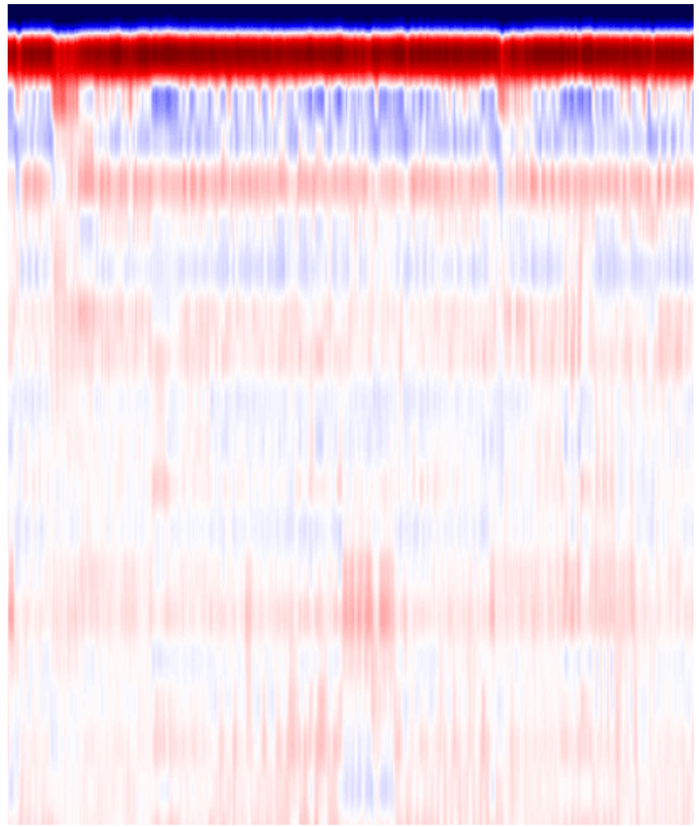
Why our choice of CNNs ?

- We realised that from a song's MFCC coefficients, we could generate spectrograms with Python.
- Spectrograms are a very compact and accurate, all rounded representation of a song and its inherent characteristics.
- So, spectrograms could be leveraged to help identify/classify songs.

CNNs are one of the best candidates for working with image classification

The Spectrograms

We converted the already provided sets of MFCC coefficients from moodle into their respective spectrograms



A spectrogram generated from MFCC table for an Asha Bhosale song

General Workflow

Created datasets for training the CNN with files of songs for the genres

Then, utilising the python package- Librosa, we converted the mp3 songs into the tables of their MFCC coefficients

Further, using Python, we generated spectrograms for the songs from the dataframes of their MFCC values, followed by data augmentation

Then trained the CNN model on this dataset after dividing into training, validation and test sets.

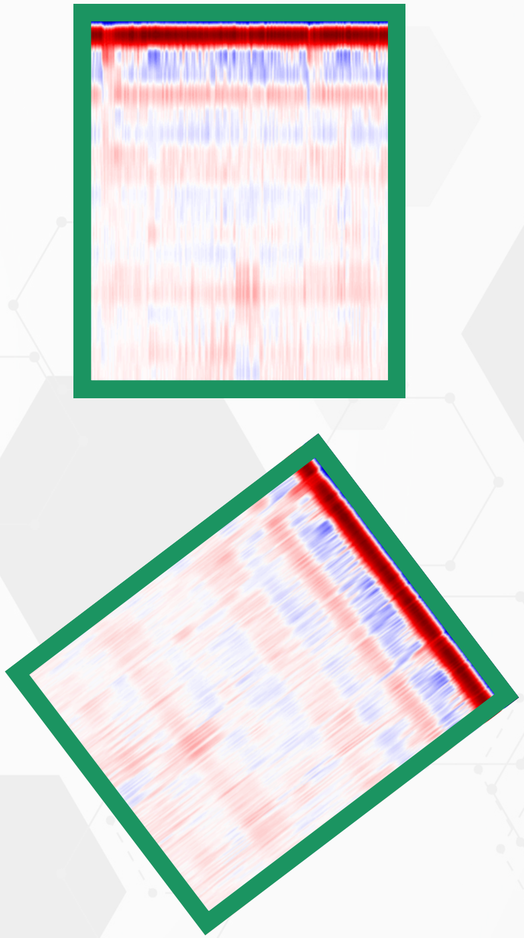
Why Data augmentation ?

- **Expands Dataset Size**: Data augmentation generates diverse variations of existing images, effectively increasing the dataset size without needing new data.
- **Reduces Overfitting**: Introducing transformations makes the model less likely to memorize specific examples, enhancing generalization to new images.
- **Improves Robustness**: By training on varied transformations, the model becomes more resilient to real-world variations like changes in angle, lighting, and scale.
- **Focuses on Key Features**: Augmentation encourages the model to prioritize distinguishing features over irrelevant details.
- **Compensates for Limited Data**: Useful when data collection is challenging, augmentation can help improve performance without additional labeled data.
- **Balances Class Representation**: Augmentation can create more samples for minority classes, addressing class imbalance in the dataset.

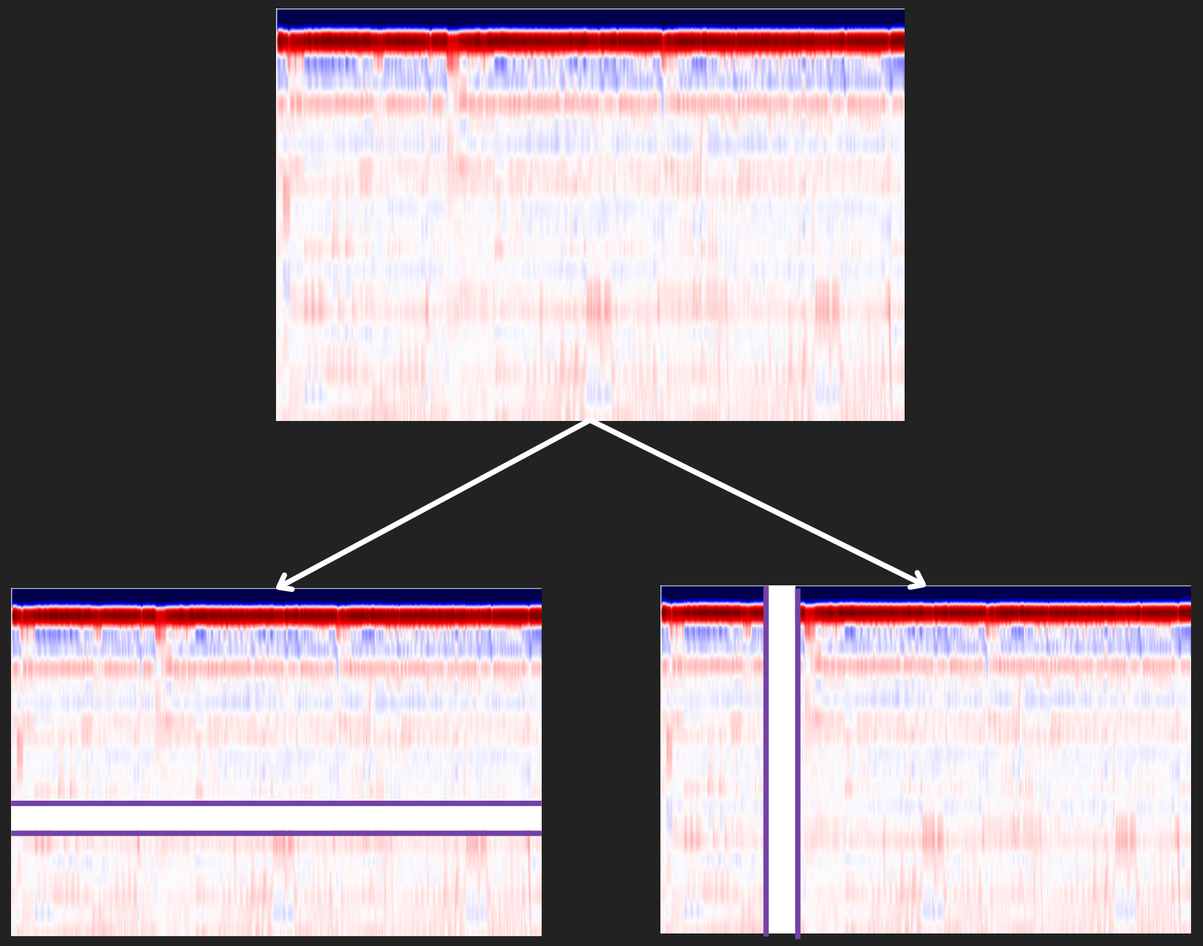
Problem with conventional Data Augmentation

1. In the conventional case, say for image recognition, data augmentation may involve techniques like image rotation

2. However in this case for spectrograms, techniques like image rotation are unfeasible or more accurately, a bit wasteful/ not relevant.

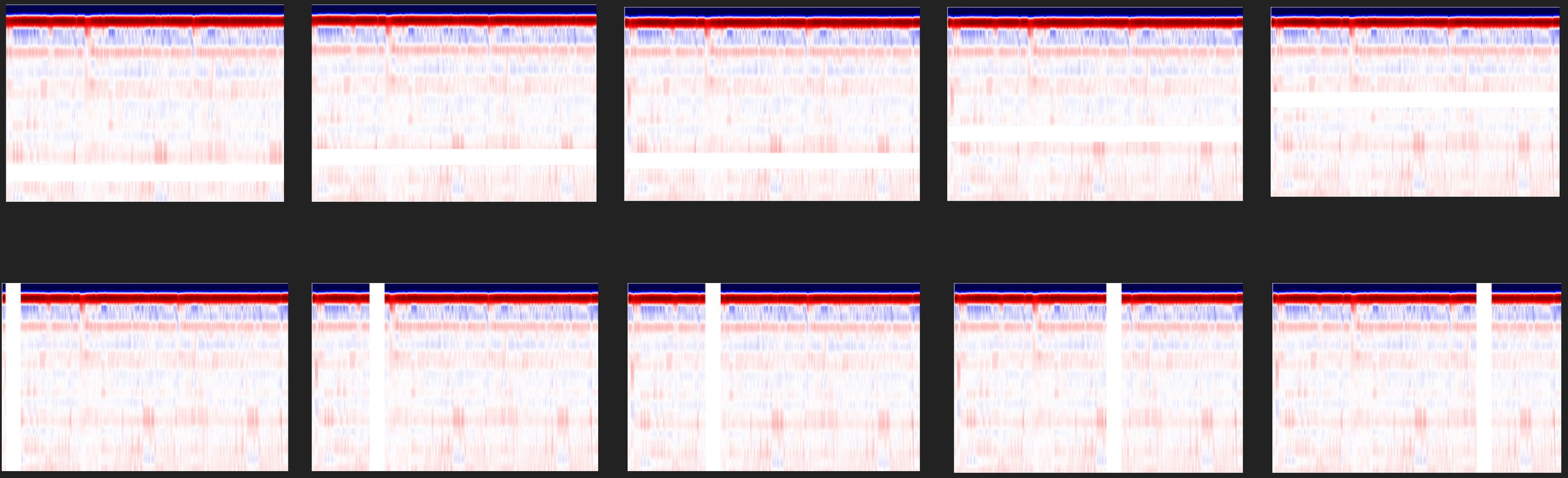


Our approach for Data Augmentation



- In a single spectrogram, we created 5 more images each with a white vertical strip at random positions along the length.
- Similarly, 5 more images were created, this time with horizontal strips at random postions along the height of the image.
- The sample size of the training dataset increased ten - fold using this approach.
- The model is improved as now it will focus more on the features of the spectrograms.

DATA AUGMENTATION



MODEL DESCRIPTION

- **Data Augmentation**: Used ImageDataGenerator for augmenting training images with parameters set to target size (128, 128), batch size of 32, and class mode as categorical.
- **Model Architecture**: Built a model with three convolution layers, each followed by max pooling, and a final dense layer. Applied a 0.5 dropout rate to reduce overfitting. Hyperparameters were optimized for better performance.
- **Training and Validation**: Achieved over 95% maximum accuracy on both training and validation sets, confirming model effectiveness.
- **Testing**: Evaluated model on the original MFCC test set, outputting predicted classes and associated probabilities.
- **Verification**: Converted selected MFCC files back to audio samples to verify model predictions, which accurately matched the intended themes.
- **Statistics**: Documented best and worst predictions, noting the closest probabilities, and calculated the average probability for the predicted class on the test set.

MODEL VALIDATION

- Accuracy on the training band validation set = 100%
- Validation loss= 0.038
- 85 test images with accuracy > 90%
- Average accuracy on test images = 79%

Classification Results

Lavani	Michael Jackson	National Anthem	Kishore Kumar	Bhavgit	Asha Bhosale
88, 62, 49, 75, 71, 70, 64, 99, 72, 14, 73, 98, 13, 112, 105, 39, 4, 56, 43, 25, 31, 34, 94, 86, 6, 41, 69, 78, 33, 47, 52, 53, 85, 91	8, 103, 21, 3, 22, 44,114, 15, 76 , 89 ,74,	48, 65, 58, 116, 107, 104, 106, 11, 109, 45, 7, 42, 19, 20, 95, 36, 54, 27, 57, 55, 87, 93, 24, 32, 90, 92, 81, 68, 2, 35,1	9, 5, 59, 84, 46, 51, 82 ,30, 68,	63, 61, 66, 67, 28, 29, 10, 83, 26, 50, 100, 80, 17, 97, 18, 96, 111	60, 115, 110, 23, 102, 79, 40,12,77

Conclusion

The model performed decently with a final average confidence of 79%.

The model showed a confidence level of over 90% for around 80 songs.

LEARNINGS FROM THIS PROJECT

- Learned about MFCC Coefficients and their use.
- Learned about spectrograms and in this case, generating spectrograms from MFCC Coefficients.
- Applied and learned about a novel data augmentation technique (the white strips/frequency masking) as described earlier
- Gained a deeper insight into the functioning of Convolutional Neural Networks CNNs and their behaviour in image classification

LEARNINGS FROM THIS PROJECT

- Familiarised ourselves with gathering information as well as rigorous research about a plethora of topics. Learned how to utilise available resources like the internet with maximum efficiency.
- Developed collaboration and work allocation skills and systematically breaking down the tasks at hand into smaller problems.

Thank You