This project seeks to build a classifier to determine whether a guitarist is playing a C, D, Em, F, or G chord. The reason why this is interesting is because music transcription has never yet used visual data, so this could introduce new, improved accuracy. Also, this visual approach could instruct us to better classify chords based on visual data.

The models trained on the RGB dataset performed better than those trained on the grayscale dataset. Since both neural networks were trained originally on RGB images, this may be due to the loss of transferability or information through contrast in the grayscale process. GoogLeNet performed better than ResNet18, even though they were both trained on ImageNet. This may be due to the flexibility of the inception architecture and GoogLeNet’s lower number of parameters leading to different features downstream. GoogLeNet yielded high training and test accuracy. Unfortunately, the model did not generalize well to new, “realistic” examples.

On new examples, the model was mainly accurate on F and Em chords. We believe the poor generalization to be the result of the original dataset not being very diverse - most of the training examples are from similar angles, resolutions, and background lighting conditions.

This could be fixed by having a larger, more diverse training set or additional methods of image augmentation.

The saliency map for ResNet18 seems to identify hand outline, while GoogLeNet seems to identify finger and knuckle locations. The ResNet18 confusion matrix indicates that the C chord is most often misclassified. This may be because the finger shape is less unique and resembles those of other chords.

FUTURE WORK

- Make our system work in real-time on video data, allowing faster and more realistic testing.
- Collect more diverse data; hand images with different backgrounds and of different colors, sizes, angles, etc.
- Incorporate audio data into the classification pipeline; more data should yield even better accuracy.
- Play with more architectures trained on different datasets.

REFERENCES

