Visual Chord Recognition

made by
Tymur Mykhalievskyi
Alexander Kuehn

What problem we are solving

Guitar Chords

Guitar Chords

In music, a chord is a set of three or more music sounds of different frequencies played simultaneously (in our case - on guitar).

The broken chords are chords too.

Major Chords

Images taken from:
https://www.uberchord.com/blog/a-major-chord-on-guitar-chord-shapes-major-scale-songs-in-the-key-of-a/

Minor Chords

Images taken from:
https://www.uberchord.com/blog/a-minor-chord-on-guitar-chord-shapes-major-scale-songs-in-the-key-of-a-minor/

Goal

Goal

Recognize 14 different chords: 7 major and 7 minor.

Start only from major ones in the beginning.

Input images can be both far and close ones.

What they say they play every Tuesday

What they actually play every Tuesday

Image taken from video: https://www.youtube.com/watch?v=oUzzt7ZVf0g&t=239s

Related work

Related work

CNN Transfer Learning for Visual Guitar Chord Classification

Made by Leon Tran, Shawn Zhang, Eric Zhou

- Main approach: extract hand from the image, then use CNN to classify the chord
- Had 5 different chords: 4 major, 1 minor
- Trained two different models: GoogLeNet and ResNet18
- Reached 100% accuracy on the test data, but report stated that test data were too similar to the training

Report can be found via the link: https://cs230.stanford.edu/projects_fall_2019/reports/26255715.pdf

Overview of approach

Approach

1. Resize and normalize the image, images can be both close and far ones

2. Augmentation

3. Make two different datasets:
   3.1. First one consists of images of only one person, a lot of augmented data
   3.2. Second one consists of images of three persons plus some images from internet, not as much of augmented data

4. Training:
   4.1. Train cnn on the first dataset until it picks up the right features
   4.2. Continue training on the final dataset

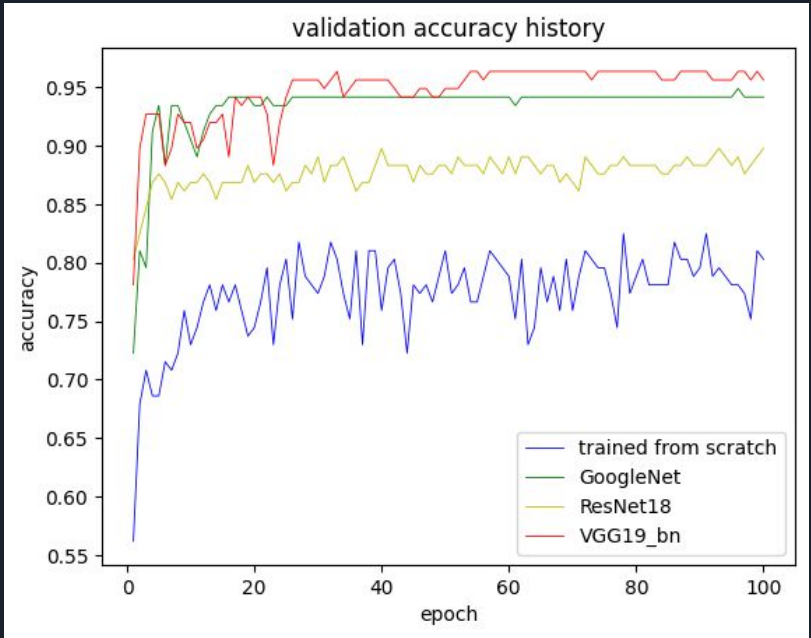
5. Train multiple models to find the best one

6. Test models on the completely unseen data taken from internet. This way train and test sets are not similar and we can evaluate models much better.

Data Augmentation

Random Rotation, Crop and ColorJitter/GrayScale

Comparison of models



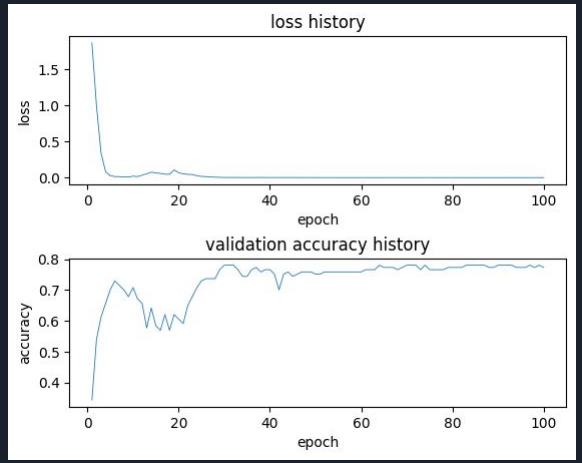
network	Hard Test accuracy
Trained from scratch	22,22%
GoogLeNet	61,11%
ResNet18	27,77%
VGG19_bn	44,44%

All trainings were done only on major chords dataset. Even though VGG's accuracy was slightly better on validation set, GoogleNet turned out to be much better on the test set

Transfer Learning

Casual training VS training with transfer

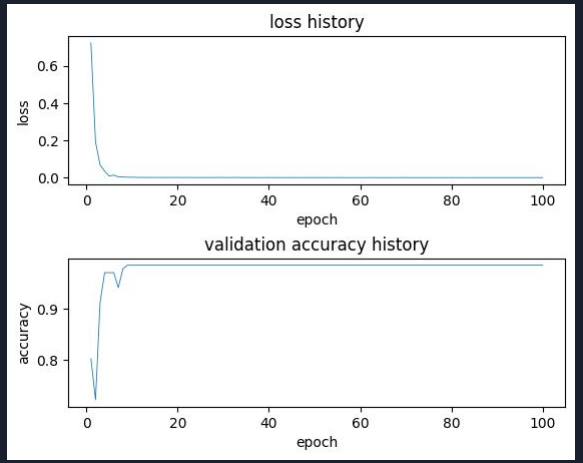
Both trainings were performed on GoogLeNet



3 epochs on simplified data

```
Epoch [1/3]: Loss: 0.5194, Validation Accuracy: 96.30%
Epoch [2/3]: Loss: 0.0442, Validation Accuracy: 98.77%
Epoch [3/3]: Loss: 0.0165, Validation Accuracy: 98.91%
```

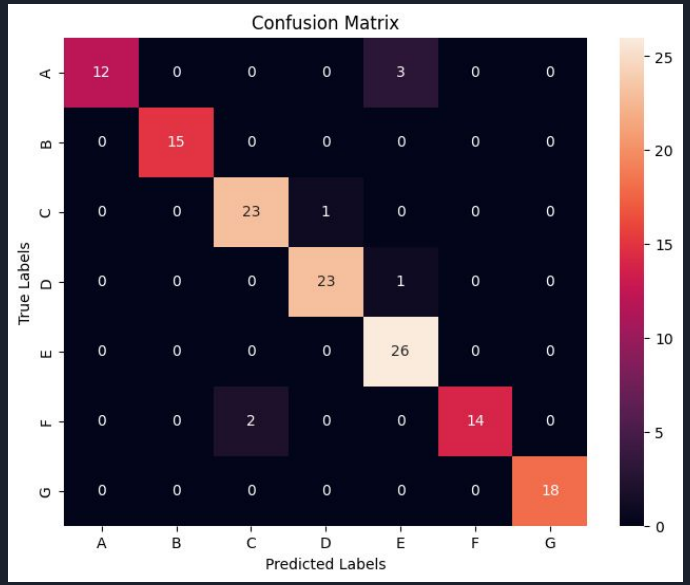
Continue training on complete data



Final Models

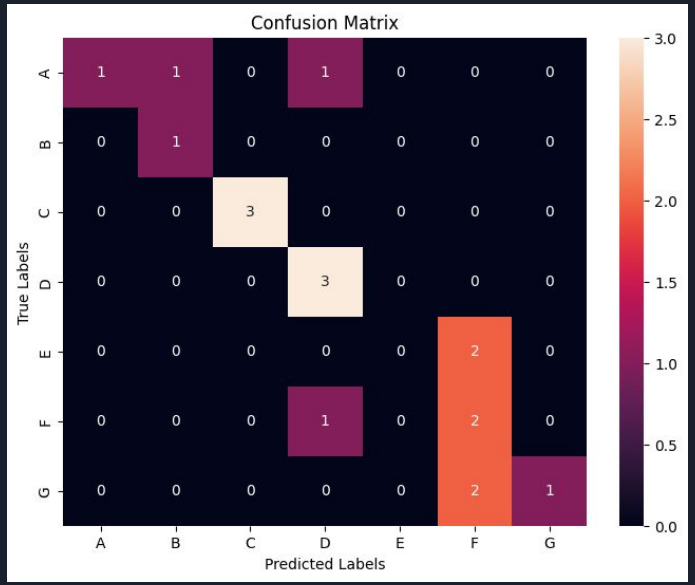
Final model trained to recognize only major chords

Simple Test



Accuracy: 94,92%

Hard Test

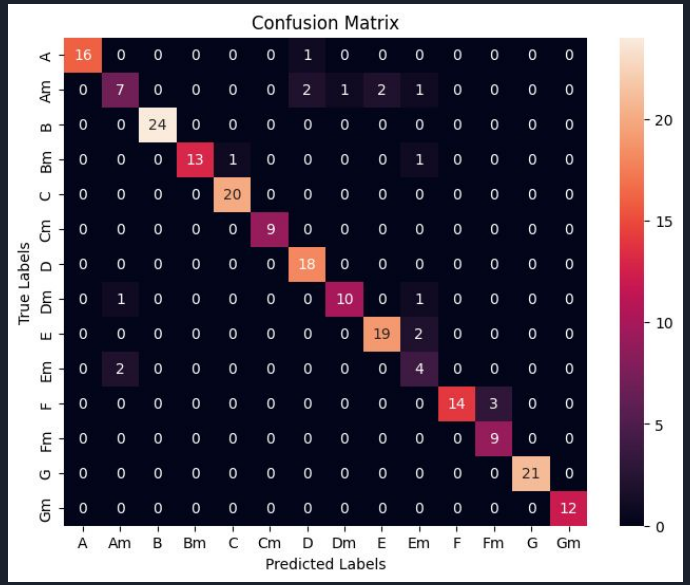


Accuracy: 61,11%

Final model trained on all 14 chords

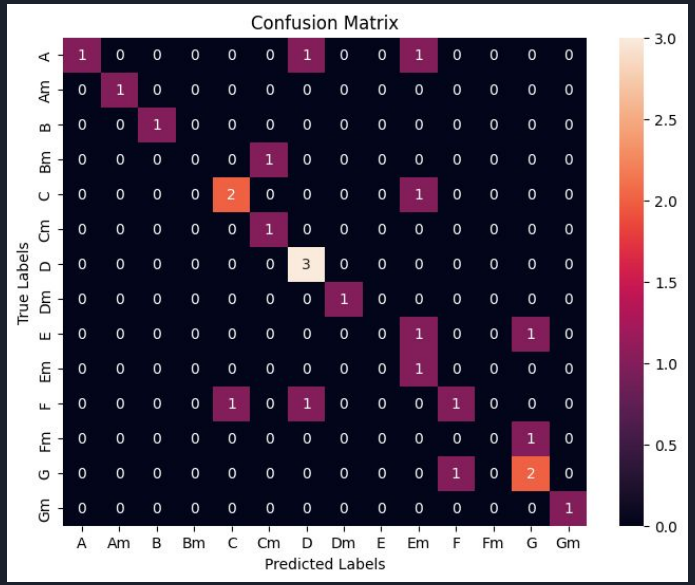
Important note: this network is also a product of transfer learning

Simple Test



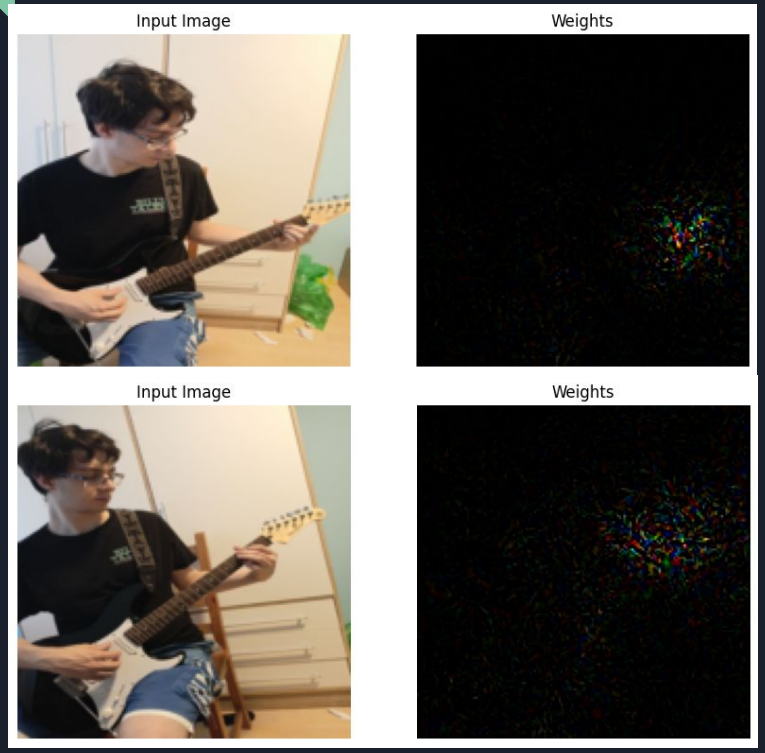
Accuracy: 91,58%

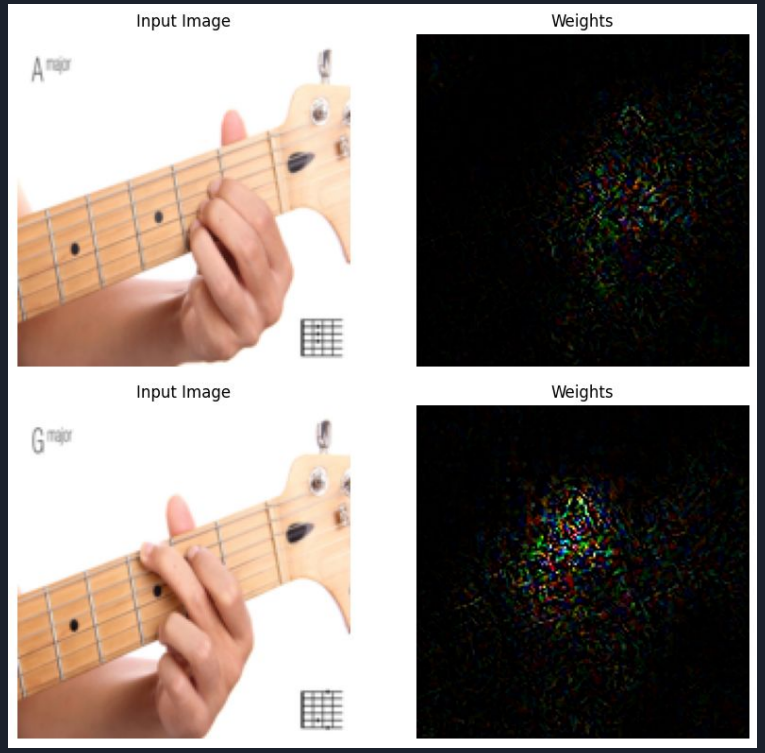
Hard Test



Accuracy: 60,00%

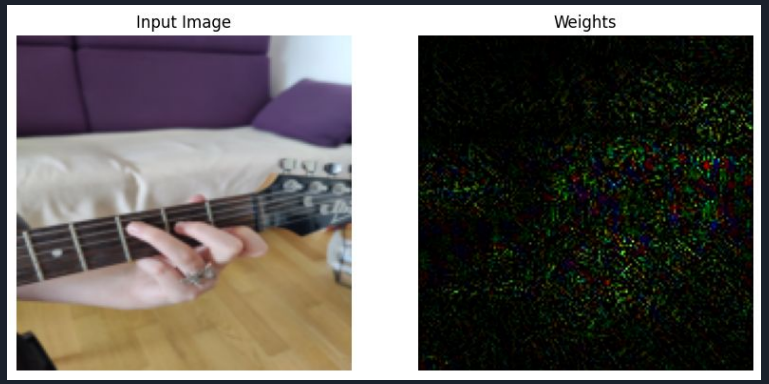
Visualization of first convolutional layer





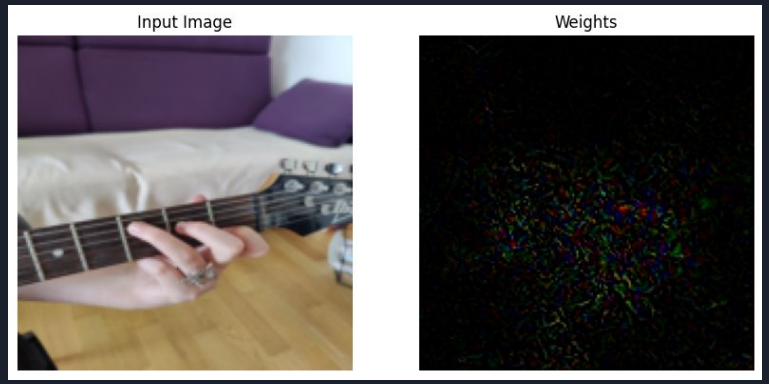
Visualization: VGG VS GoogleNet

VGG19_bn



Weights were scaled by factor of 2 for the visualization

GoogleNet



Weights were scaled by factor of 8 for the visualization

Future improvements

Theoretical future work

- Gather more diverse data to generalize a model better
- Try different approaches, for example ViT
- Ideal outcome is to give a full tablature for a video (tablature - the visual representation of the notes in a song)

Note: none of the improvements are to be made till the final report

Real time demo of our work