

# Gesture Controlled Audio Software

## Capstone Final Presentation

Anusha Bhattacharyya, Troy Watts, Maxx  
Smith, and Winston Phillips

# Introduction

## Problem:

- Smart Devices
- Live Music
- Interactive Presentations
- Theater Productions

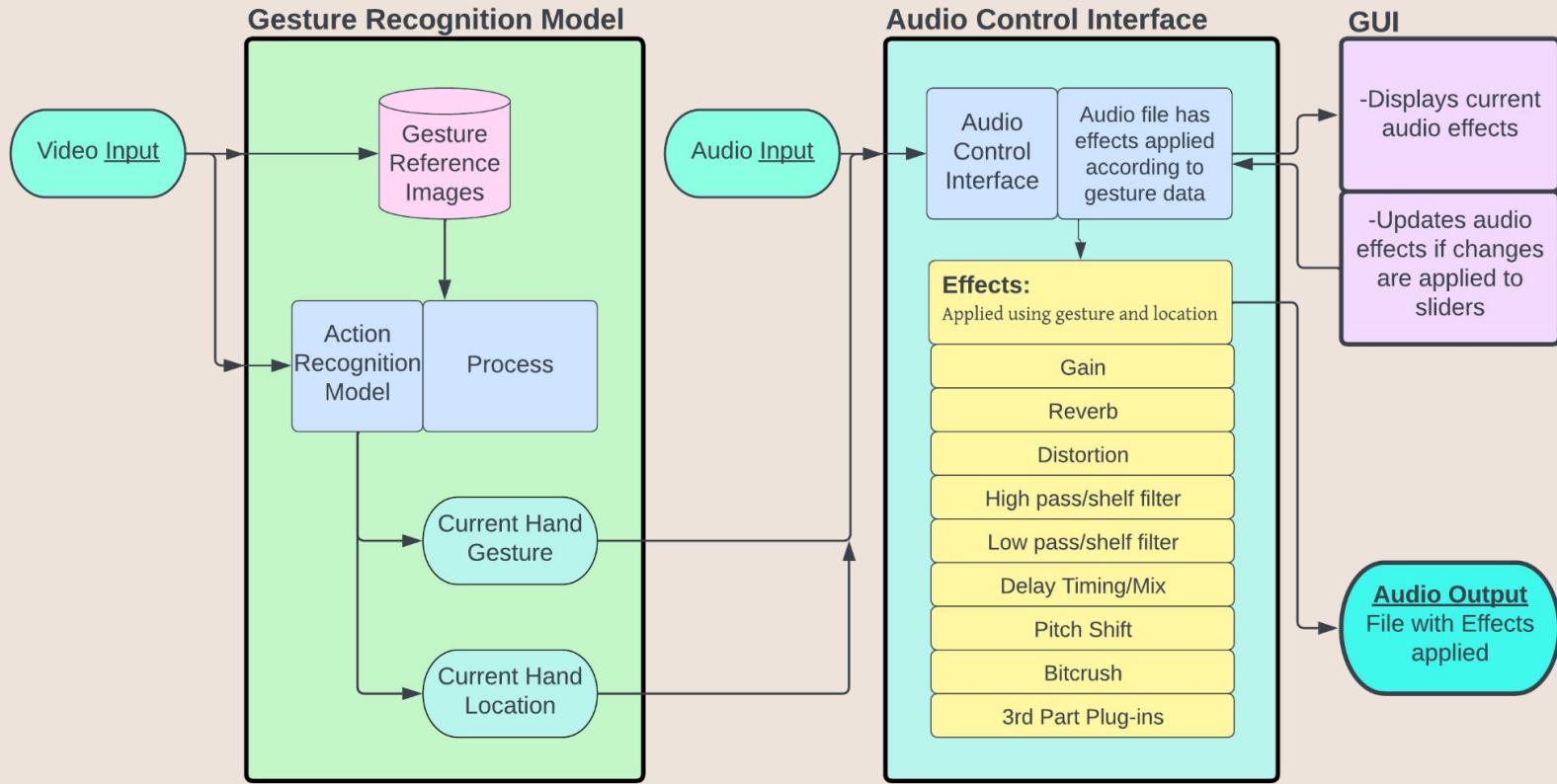
## Objective:

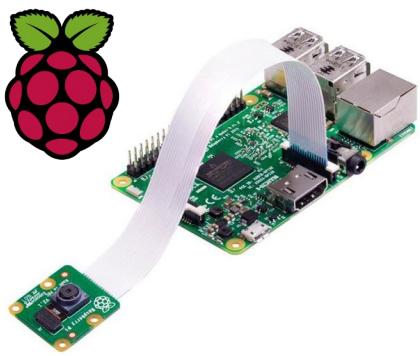
The objective of this project is to design, prototype, and build a program that can take in video input and turn that into audio output depending on what gestures are being made using an AI generated algorithm.

## Keywords:

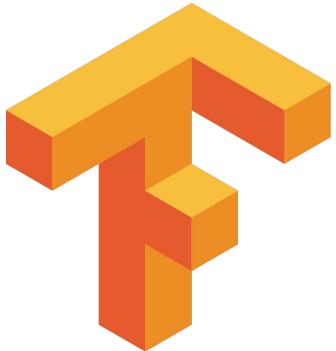
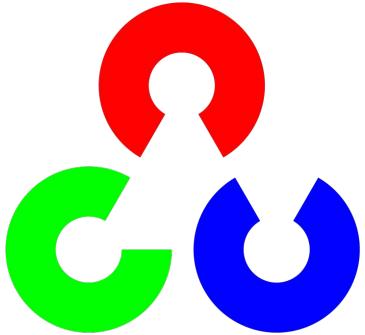
- Key-Point Inference Algorithm/Paradigms
- Audio Control Interface (Pedalboard)
- DAW (Digital Audio Workstation)

# Project Architecture





# Action Recognition Model



```
New Load Save Run Debug Over Info Out Stop Zoom Quit


```

import cv2
import numpy as np
from keras import Sequential
from keras.layers import Dense
from keras.layers import Activation
from keras.utils import np_utils
from keras.layers import Dropout, Convolution2D, GlobalAveragePooling2D
import tensorflow as tf
import os

IMG_SAVE_PATH = 'images'
CLASS_MAP = {
    'guitarist': 0,
    'nothing': 1
}
NUM_CLASSES = len(CLASS_MAP)

def mapper(val):
    return CLASS_MAP[val]

def get_model():
    model = Sequential([
        Convolution2D(32, 3, 3, input_shape=(227, 227, 3), include_top=False),
        Activation('relu'),
        Convolution2D(32, 3, 3, padding='valid'),
        Activation('relu'),
        GlobalAveragePooling2D(),
        Activation('softmax')
    ])
    return model

# load images from this directory
@timeit
def load_data(directory, img_size=(227, 227)):
    for directory in os.listdir(IMG_SAVE_PATH):
        path = os.path.join(IMG_SAVE_PATH, directory)
        if os.path.isdir(path):
            continue
        for item in os.listdir(path):
            for file in os.listdir(os.path.join(path, item)):
                if file.startswith('.'):
                    continue
                img = cv2.imread(os.path.join(path, item, file))
                img = cv2.cvtColor(img, COLOR_BGR2RGB)
                img = cv2.resize(img, (227, 227))
                dataset.append(img, directory)

data, labels = zip(*dataset)
labels = list(map(mapper, labels))

# one hot encode the labels
labels = np_utils.to_categorical(labels)

# define the model
model = get_model()
model.compile(optimizer=Adam(lr=0.0001),
              loss='categorical_crossentropy',
              metrics=[accuracy])
model.summary()

# start training
model.fit(np.array(data), np.array(labels), epochs=15)
# save the model for later use
model.save("model.h5")

```


```

```
File Edit Tab Help

Logos/gestur1/879 .JPG
Logos/gestur1/880 .JPG
Logos/gestur1/881 .JPG
Logos/gestur1/882 .JPG
Logos/gestur1/883 .JPG
Logos/gestur1/884 .JPG
Logos/gestur1/885 .JPG
Logos/gestur1/886 .JPG
Logos/gestur1/887 .JPG
Logos/gestur1/888 .JPG
Logos/gestur1/889 .JPG
Logos/gestur1/890 .JPG
Logos/gestur1/891 .JPG
Logos/gestur1/892 .JPG
Logos/gestur1/893 .JPG
Logos/gestur1/894 .JPG
Logos/gestur1/895 .JPG
Logos/gestur1/896 .JPG
Logos/gestur1/897 .JPG
Logos/gestur1/898 .JPG
Logos/gestur1/899 .JPG
Logos/gestur1/900 .JPG
Logos/gestur1/901 .JPG
Logos/gestur1/902 .JPG
Logos/gestur1/903 .JPG
Logos/gestur1/904 .JPG
Logos/gestur1/905 .JPG
Logos/gestur1/906 .JPG
Logos/gestur1/907 .JPG
Logos/gestur1/908 .JPG
Logos/gestur1/909 .JPG
Logos/gestur1/910 .JPG
Logos/gestur1/911 .JPG
Logos/gestur1/912 .JPG
Logos/gestur1/913 .JPG
Logos/gestur1/914 .JPG
Logos/gestur1/915 .JPG
Logos/gestur1/916 .JPG
Logos/gestur1/917 .JPG
Logos/gestur1/918 .JPG
Logos/gestur1/919 .JPG
Logos/gestur1/920 .JPG
Logos/gestur1/921 .JPG
Logos/gestur1/922 .JPG
Logos/gestur1/923 .JPG
Logos/gestur1/924 .JPG
Logos/gestur1/925 .JPG
Logos/gestur1/926 .JPG
Logos/gestur1/927 .JPG
Logos/gestur1/928 .JPG
Logos/gestur1/929 .JPG
Logos/gestur1/930 .JPG
Logos/gestur1/931 .JPG
Logos/gestur1/932 .JPG
Logos/gestur1/933 .JPG
Logos/gestur1/934 .JPG
Logos/gestur1/935 .JPG
Logos/gestur1/936 .JPG
Logos/gestur1/937 .JPG
Logos/gestur1/938 .JPG
Logos/gestur1/939 .JPG
Logos/gestur1/940 .JPG
Logos/gestur1/941 .JPG
Logos/gestur1/942 .JPG
Logos/gestur1/943 .JPG
Logos/gestur1/944 .JPG
Logos/gestur1/945 .JPG
Logos/gestur1/946 .JPG
Logos/gestur1/947 .JPG
Logos/gestur1/948 .JPG
Logos/gestur1/949 .JPG
Logos/gestur1/950 .JPG
Logos/gestur1/951 .JPG
Logos/gestur1/952 .JPG
Logos/gestur1/953 .JPG
Logos/gestur1/954 .JPG
Logos/gestur1/955 .JPG
Logos/gestur1/956 .JPG
Logos/gestur1/957 .JPG
Logos/gestur1/958 .JPG
Logos/gestur1/959 .JPG
Logos/gestur1/960 .JPG
Logos/gestur1/961 .JPG
Logos/gestur1/962 .JPG
Logos/gestur1/963 .JPG
Logos/gestur1/964 .JPG
Logos/gestur1/965 .JPG
Logos/gestur1/966 .JPG
Logos/gestur1/967 .JPG
Logos/gestur1/968 .JPG
Logos/gestur1/969 .JPG
Logos/gestur1/970 .JPG
Logos/gestur1/971 .JPG
Logos/gestur1/972 .JPG
Logos/gestur1/973 .JPG
Logos/gestur1/974 .JPG
Logos/gestur1/975 .JPG
Logos/gestur1/976 .JPG
Logos/gestur1/977 .JPG
Logos/gestur1/978 .JPG
Logos/gestur1/979 .JPG
Logos/gestur1/980 .JPG
Logos/gestur1/981 .JPG
Logos/gestur1/982 .JPG
Logos/gestur1/983 .JPG
Logos/gestur1/984 .JPG
Logos/gestur1/985 .JPG
Logos/gestur1/986 .JPG
Logos/gestur1/987 .JPG
Logos/gestur1/988 .JPG
Logos/gestur1/989 .JPG
Logos/gestur1/990 .JPG
Logos/gestur1/991 .JPG
Logos/gestur1/992 .JPG
Logos/gestur1/993 .JPG
Logos/gestur1/994 .JPG
Logos/gestur1/995 .JPG
Logos/gestur1/996 .JPG
Logos/gestur1/997 .JPG
Logos/gestur1/998 .JPG
Logos/gestur1/999 .JPG
```





Gesture: ThumbsUp

Confidence: 0.99999999

Confidence: 0.99999999

Confidence: 0.99999999

Confidence: 0.99999999

Confidence: 0.99999999

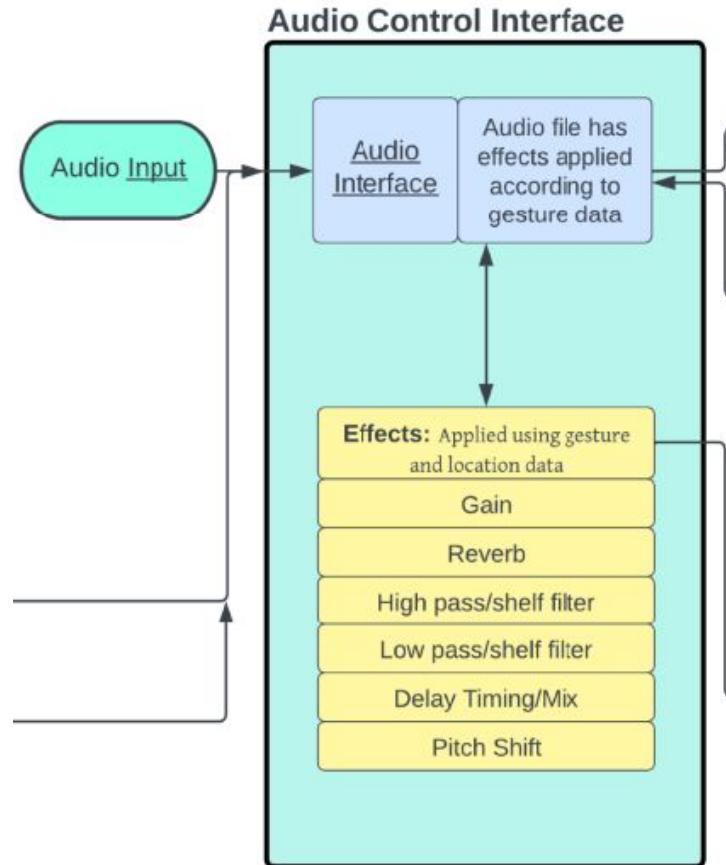


OpenCV-Demo



# Audio Interface

- Pedalboard
- Gesture control changes variables for FX
- Audio workstation file applies changes to audio file
- Audio file is saved as a new output file

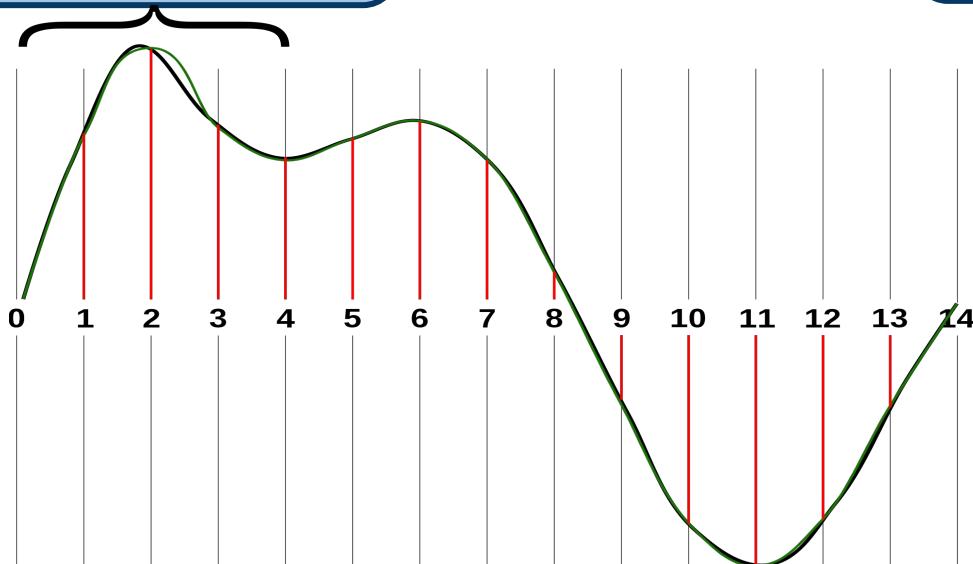


# Audio Effects - Sample Array

```
samples[  
  [0.001, 0.103, 0.105, 0.107, 0.109],  
  [0.012, 0.023, 0.029, 0.033, 0.043]  
]
```

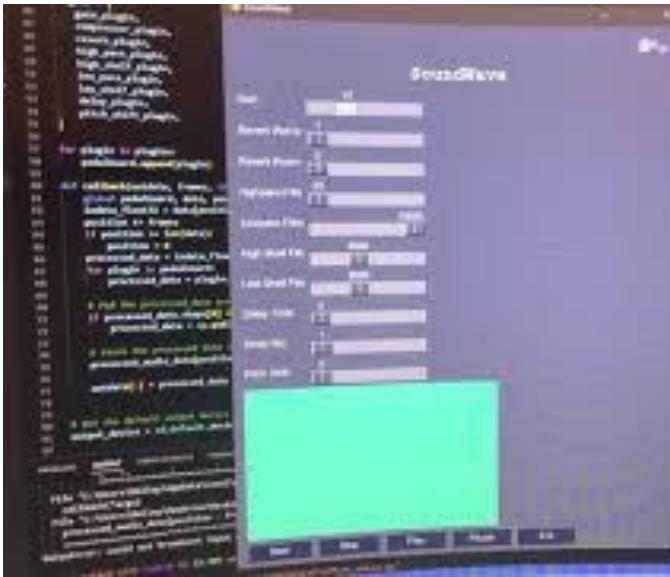


```
samplesFX[  
  [0.1211, 0.1531, 0.1652, 0.1674, 0.1993],  
  [0.0124, 0.0231, 0.0291, 0.0334, 0.0433]  
]
```



# GUI and Audio Control Demo

```
git clone https://github.com/robjones1985/audioclient.git  
cd audioclient  
make  
./audioclient
```



Example audio changes

# Potential Future Changes

- AI model needs about 4000 pictures for each gesture
- Some gestures are more difficult to recognize than others
- Implement two hand recognition
- Third-party plugin support

# Gesture Controlled Audio Software

## Capstone Final Presentation

Anusha Bhattacharyya, Troy Watts, Maxx  
Smith, and Winston Phillips