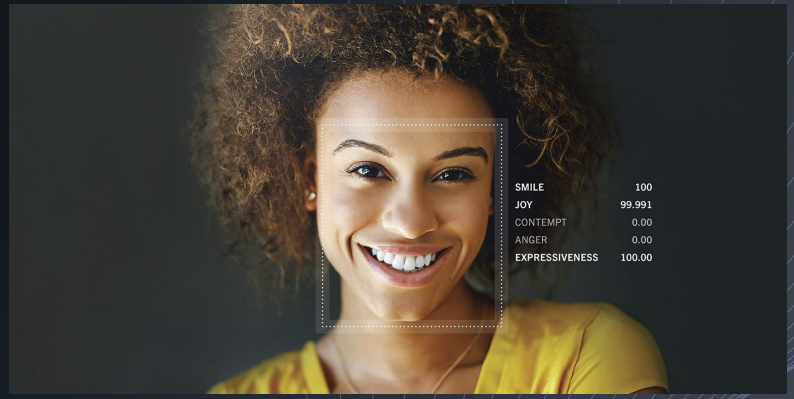
Emotion AI: Predicting Humans Affectively

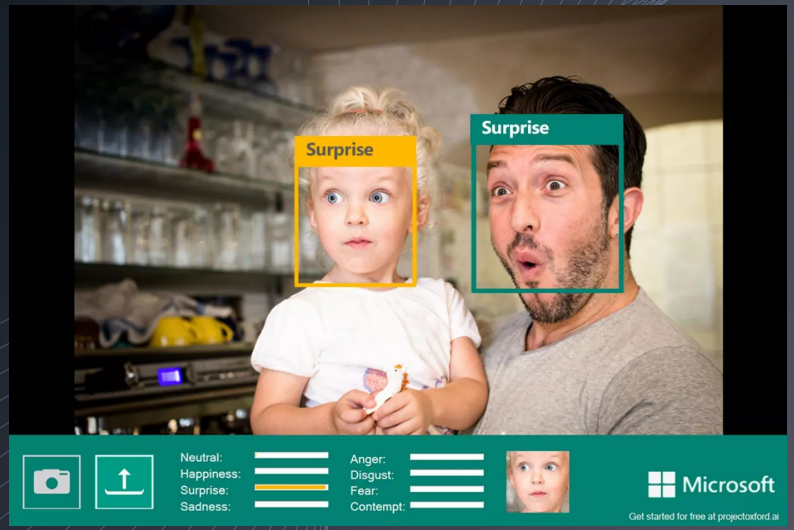
Zachary Villarreal

Motivation

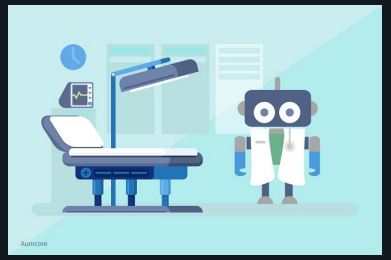
Emotion AI: Emotion recognition technology.

- **What it attempts**: Detect emotion from multiple channels such as facial images or audio clips
- **Its Goal**: Develop machines that are capable of interpreting human affect, the same way we, as humans, do





Emotion AI: Applications



Medical Diagnosis: Help doctors detect diseases such as depression and dementia by voice analysis

Automotive: Detect whether drivers are tired or stressed based on facial expressions

Robotics: Use of emotional data to depict emotions on robots similar to how humans would.

Overview / Data

Goal: Be able to predict emotion from not only facial images but audio clips, as well, to create a more accurate representation of human emotion

Audio Data: 13,000 audio clips, 7 emotions over 2 sexes

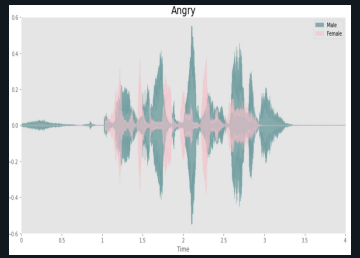
- **Crema-D** : Crowd Sourced Emotional Multimodal Actors Dataset
- **RAVDESS** : Ryerson Audio-Visual Database of Emotional Speech and Song
- **SAVEE** : Surrey Audio-Visual Expressed Emotion
- **TESS** : Toronto emotional speech set

Image Data: 36,000 images, 7 emotions

- **FER2013** : Facial Expression Recognition Competition

Audio Pipeline

Input: Angry | Male



Tone Identification

Pitch Level

Distinct Units of Sound (MFCC)

Male or Female?

Emotion

Output: Angry | Male

Audio Convolutional Neural Network

Purpose: To predict emotion and sex from audio files

Baseline: 7% Accuracy
Achieved Score: 67% Accuracy

4 Million Parameters
4 CNN layers
6 Dense layers
14 classifiers (emotions+sex)

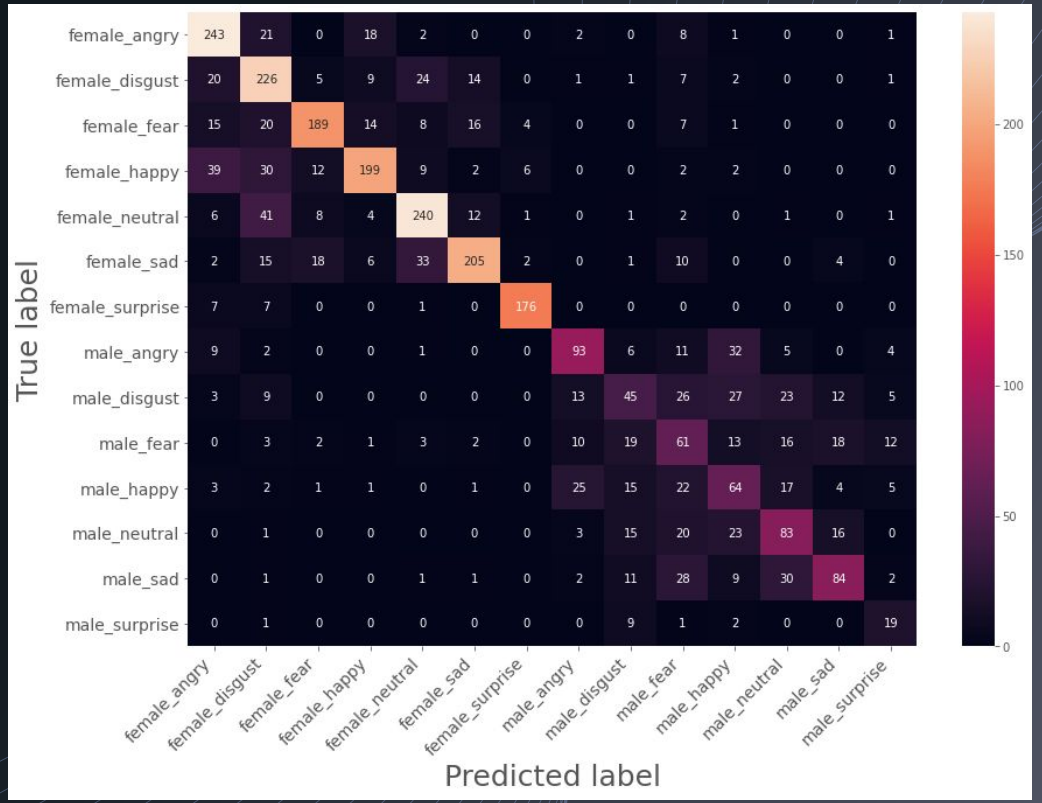


Image Pipeline

Input: Angry | Male

OpenCV Facial Recognition

Image Augmentation

Pixel Intensity

OpenCV Facial Recognition

Male or Female?

Emotion

Output: Angry | Male

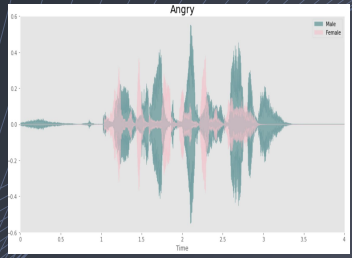
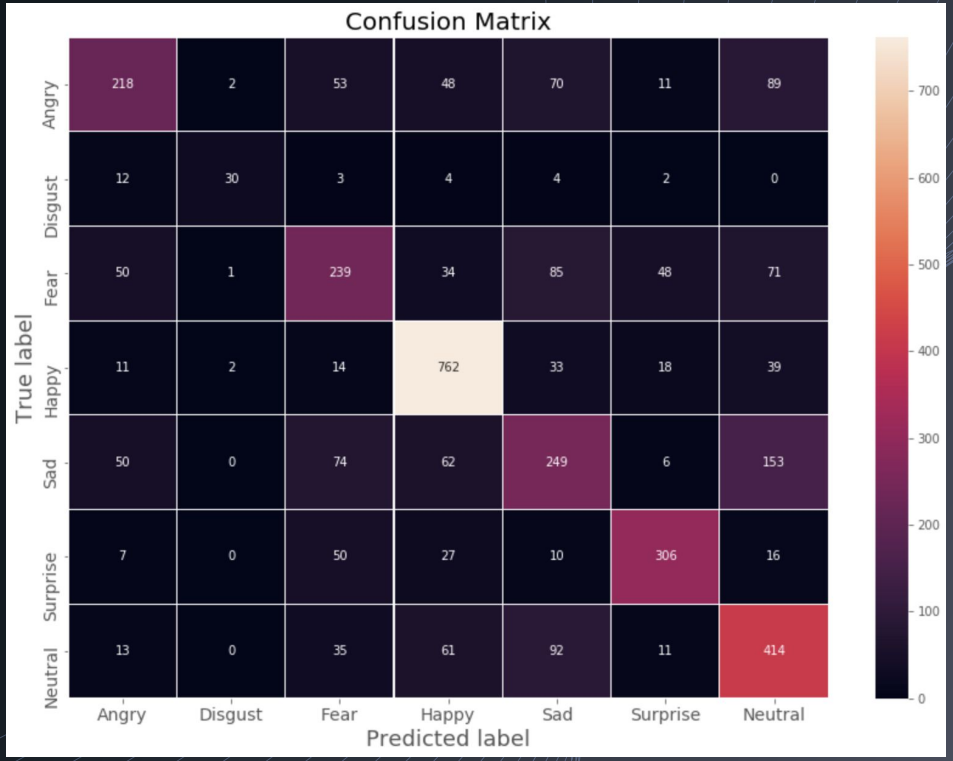


Image Convolutional Neural Network

Purpose: To predict emotion and sex from image files

Baseline: 14% Accuracy
Achieved Score: 67% Accuracy

4 Million Parameters
6 CNN layers
3 dense layers
7 classifiers (emotions)



Conclusion

Emotional AI:

- We can predict, with certain accuracy, emotion from both audio and images
- We now can see the importance that emotional AI has in our future

Takeaways:

- We can't always detect emotion from just two sources, humans are often far more complicated, but this project created a more accurate representation of human emotion as a whole.

Future Work:

- Try to take in more audio files and facial images to create better models that can predict emotion more accurately

Application Overview

App Link: [Click Here]

Emotion Detection App

Drag and drop or click to select a file to upload.

Image File Detected!

Emotion Detected

HAPPY

Audio File

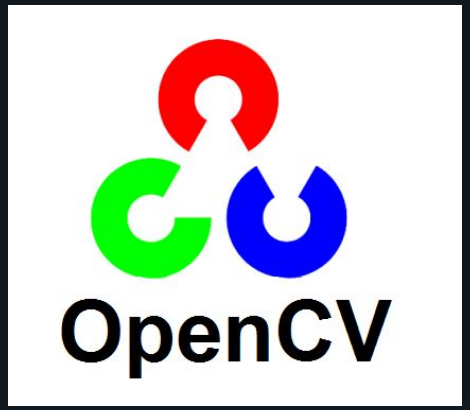
0:00 / 0:02

Sex Detected

MALE

Image File

Tech Stack



librosa

plotly

Thank you for listening!

LinkedIn: linkedin.com/in/zachary-p-villarreal/

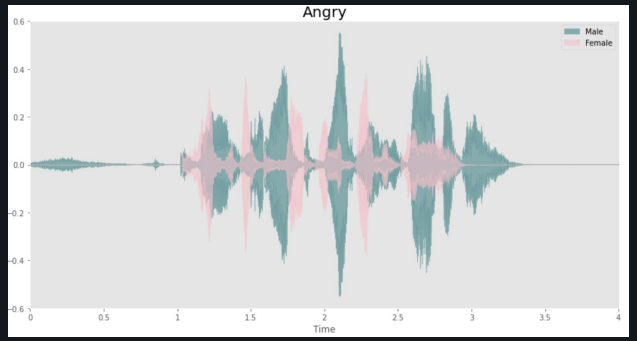
GitHub: github.com/ZacharyVillarreal

Email: zvillarreal@ucdavis.edu

Phone: (818) 879-3377

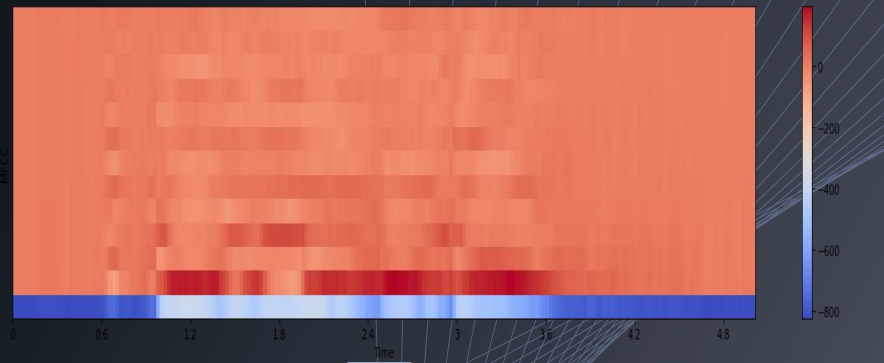
Audio Pipeline

Audio Clip - (.wav format)



Feature Extraction

MFCC



Predict Sex

Predict Emotion

1D Convolutional NN

Call to image dir.

Output Image

Image Pipeline

Image - (.jpg format)

OpenCV:
Facial
Recognition

Data
Augmentation

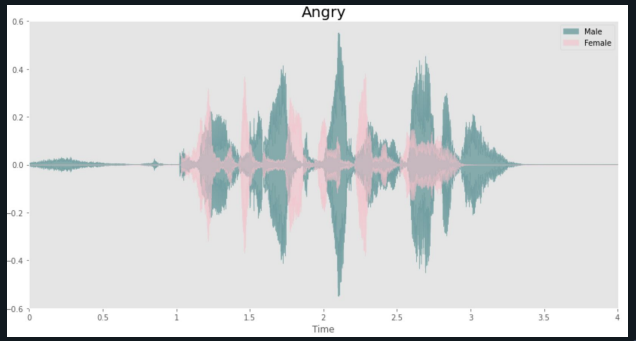
2D Convolutional-NN

Predict
Sex

Predict
Emotion

Call to
audio dir.

Output
Audio Clip



Audio Neural Network Structure

4 Million Parameters
6 dense layers
14 classifiers (emotions+sex)

Layer (type)	Output Shape	Param #
conv1d_16 (Conv1D)	(None, 49, 32)	128
batch_normalization_16 (Batc	(None, 49, 32)	128
max_pooling1d_16 (MaxPooling	(None, 24, 32)	0
conv1d_17 (Conv1D)	(None, 24, 64)	6208
batch_normalization_17 (Batc	(None, 24, 64)	256
max_pooling1d_17 (MaxPooling	(None, 12, 64)	0
conv1d_18 (Conv1D)	(None, 12, 128)	41088
batch_normalization_18 (Batc	(None, 12, 128)	512
max_pooling1d_18 (MaxPooling	(None, 6, 128)	0
conv1d_19 (Conv1D)	(None, 6, 256)	98560
batch_normalization_19 (Batc	(None, 6, 256)	1024
max_pooling1d_19 (MaxPooling	(None, 3, 256)	0
flatten_5 (Flatten)	(None, 768)	0
dense_13 (Dense)	(None, 512)	393728
activation_13 (Activation)	(None, 512)	0
dense_14 (Dense)	(None, 256)	131328
activation_14 (Activation)	(None, 256)	0
dropout_5 (Dropout)	(None, 256)	0
dense_15 (Dense)	(None, 14)	3598
activation_15 (Activation)	(None, 14)	0

Total params: 676,558
Trainable params: 675,598
Non-trainable params: 960

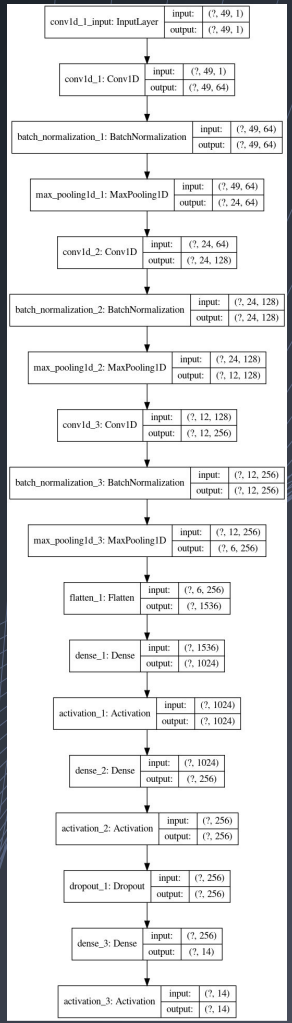
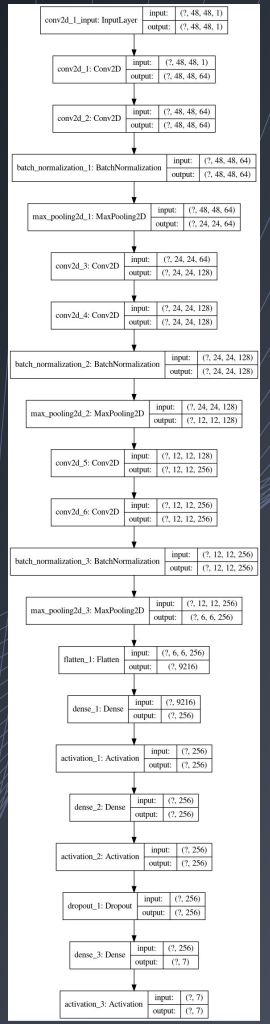


Image Neural Network Structure

4 Million Parameters
8 dense layers
7 classifiers (emotions)

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 48, 48, 64)	640
conv2d_2 (Conv2D)	(None, 48, 48, 64)	36928
batch_normalization_1 (Batch	(None, 48, 48, 64)	256
max_pooling2d_1 (MaxPooling2	(None, 24, 24, 64)	0
conv2d_3 (Conv2D)	(None, 24, 24, 128)	204928
conv2d_4 (Conv2D)	(None, 24, 24, 128)	409728
batch_normalization_2 (Batch	(None, 24, 24, 128)	512
max_pooling2d_2 (MaxPooling2	(None, 12, 12, 128)	0
conv2d_5 (Conv2D)	(None, 12, 12, 256)	295168
conv2d_6 (Conv2D)	(None, 12, 12, 256)	590080
batch_normalization_3 (Batch	(None, 12, 12, 256)	1024
max_pooling2d_3 (MaxPooling2	(None, 6, 6, 256)	0
flatten_1 (Flatten)	(None, 9216)	0
dense_1 (Dense)	(None, 256)	2359552
activation_1 (Activation)	(None, 256)	0
dense_2 (Dense)	(None, 256)	65792
activation_2 (Activation)	(None, 256)	0
dropout_1 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 7)	1799
activation_3 (Activation)	(None, 7)	0

Total params: 3,966,407
Trainable params: 3,965,511
Non-trainable params: 896



Application Overview