UNIVERSITY
OF COLOGNE

DLR

EXPLORING STATE-OF-THE-ART TOOLS FOR EMOTION DETECTION

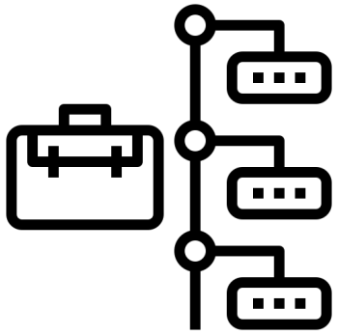
A Comparative Analysis of Single and Multi-Modal Technologies with a Prototype Implementation

Birol Arin / Julian Stricker

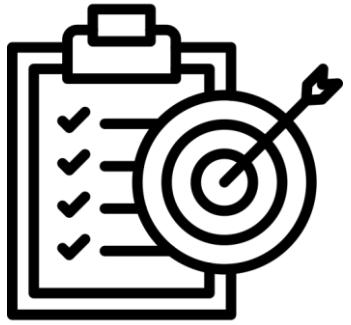
Table of Contents

Motivation

Emotion detection enables adaptive intuitive interfaces for smoother human-computer interaction



Valuable for applications in mental health, safety, education and entertainment



Systematically review and compare emotion detection tools/models

Demonstration via prototype

Theoretical Background
Facial Action Coding System (FACS)

Ekman in collaboration with Friesen (1978)

System to describe any visible facial movement

Identify and encode basic building blocks of facial expression

Action Units

AU	FACS name	Prediction
AU1	Inner brow raiser	R
AU2	Outer brow raiser	R
AU4	Brow lowerer	R
AU5	Upper lid raiser	R
AU6	Cheek raiser	R
AU7	Lid tightener	D
AU9	Nose wrinkler	R
AU10	Upper lip raiser	D
AU12	Lip corner puller	R
AU14	Dimpler	D
AU15	Lip corner depressor	R
AU17	Chin raiser	R
AU20	Lip stretcher	R
AU23	Lip tightener	D
AU24	Lip pressor	D
AU25	Lips part	R
AU26	Jaw drop	R




Gifs from imotions.com

Theoretical Background
Basic emotions according to Ekman (1992)

Emotions are universal, evolutionary adaptions for coping with life tasks (Ekman, 1992)

Basic emotions according to Ekman (1992)
pictures from FER-2013 data set

Theoretical Background
Basic emotions according to Ekman (1992)

Emotions are universal, evolutionary adaptions for coping with life tasks (Ekman, 1992)

Basic emotions according to Ekman (1992)
pictures from FER-2013 data set

Theoretical Background
Current State of Research - Emotion

Cultural norms influence emotion expression and interpretation (Jack et al., 2012; Hutto et al., 2018)

Emotions are dynamic, context-dependent and not static categories (Tracy and Randels, 2011; Jack et al., 2012)

Basic emotions still foundational in facial emotion recognition research

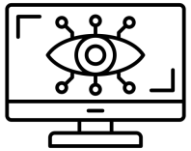
Theoretical Background
Valence Arousal

Emotions can be represented in a two-dimensional space defined by **Valence** (Pleasure-Displeasure) and **Arousal** (Activation-Deactivation)

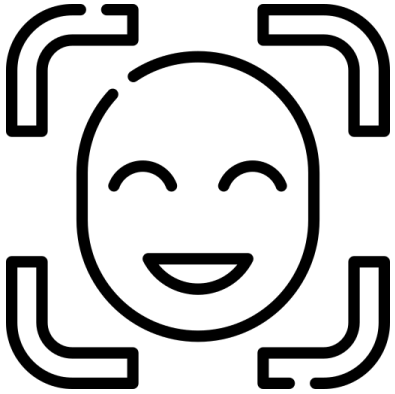
A **circumplex model** maps emotions in a circular structure based on valence and arousal

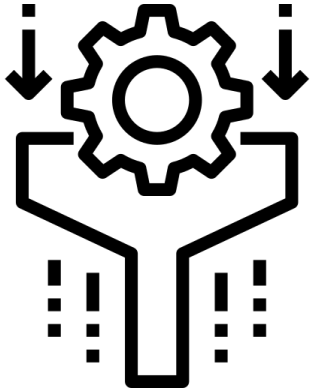
The model assumes valence and arousal are **independent**, but research suggests potential **interdependencies**

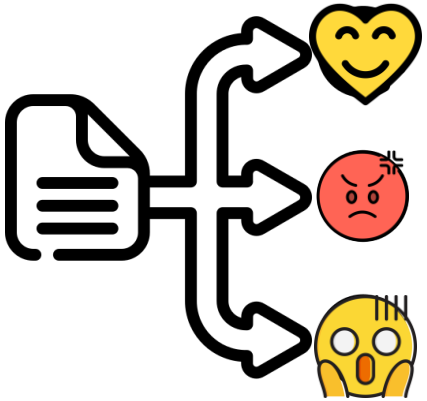
Theoretical Background
Facial Expression Recognition (FER)



Computer vision task analyzing facial expression to identify emotions

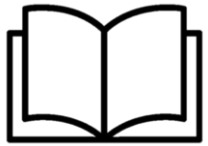






1. Face detection **2. Feature extraction** **3. Emotion classification**

Research Methodology & Literature Approach



Databases & Sources

Primary searches conducted on Google Scholar, supplemented with arXiv, IEEE Xplore, ResearchGate and GitHub

Search String Example

"facial emotion recognition" AND ("CNN" OR "deep learning") AND ("open source" OR "real-time")

Inclusion Criteria

- *Pre-2020:* Focus on impact, citation count and peer-reviewed status
- *Post-2020:* Emphasis on novelty, performance, and emerging architectures like transformers
- Approximately 46 articles and papers were selected to form a comprehensive basis for our tool evaluation

Evaluation Criteria for Tool Comparison

1. General Information

- Modalities
- Emotion models used
- Open-source status
- Primary use case

3. Performance Metrics

- Reported KPI
- Datasets used

5. Summary & Recommendations

- Key Strengths
- Key Limitations
- Recommendation

2. Technical Aspects

- Model architecture & framework
- Pre-trained models
- Input/Output types
- Real-time capability

4. Implementation & Ease of Use

- Setup complexity
- Programming language support
- Integration
- Documentation quality

Overview of Comparative Results (1/3)

General Information

Criteria	openSMILE	OpenFace	OpenFace 2.0	DeepFace	ResEmoteNet	LibreFace	Multi-Scale ViT	Emotion-LLaMA
Primary Modality	Audio	Image/Video	Image/Video	Image/Video	Image	Image/Video	Image/Video	Multimodal (Audio, Video, Text)
Emotion Model Used [1]	Ekman-Based [2]	Ekman-FACS [2]	Ekman-FACS [2]	Ekman-FACS [2]	Ekman-Based [2]	Ekman-FACS [2]	Ekman-Based [2]	Valence-Arousal, Ekman-Based [2]
Open Source?	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Primary Use Case	Audio Feature Extraction; Paralinguistics	Real-time Facial Feature Analysis	Real-time Facial Feature Analysis	Face Verification & FER	FER	Real-time FER	High-Accuracy FER	Multimodal ER & Reasoning

Feature extraction tools enabling emotion recognition

Emotion recognition tools for direct classification

[1] indicates either the model integrated into the tool itself, or the model used in research applying the tool for emotion recognition

[2] Ekman FACS is a method for analyzing facial movements, while Ekman-Based refers to broader emotion recognition grounded in Ekman's basic emotion categories

Overview of Comparative Results (2/3)

Technical Aspects & Model Architecture

Criteria	openSMILE	OpenFace	OpenFace 2.0	DeepFace	ResEmoteNet	LibreFace	Multi-Scale ViT	Emotion-LLaMA
(Model) Architecture	Feature Extraction Pipeline (modular)	CLNF (instance of CLM) + CNN	CLNF (instance of CLM) + CNN	CNN (8 layers) + LCL, 3D face model	CNN + SENet + ResNets	ResNet-18, Swin Transformer, MAE	ViT + multi-scale processing + contrastive learning	LLaMA + emotion-specific encoders
Framework Used	C++	PyTorch	PyTorch	Tensorflow & Keras	PyTorch	PyTorch	MATLAB	PyTorch
Pre-Trained Models?	No; pre-defined feature sets	Yes	Yes	Yes	Yes	Yes	No	Yes
Input Data Type	Audio files	Image/Video	Image/Video	Image/Video	Image	Image/Video	Image	Video, Audio, Text
Output Data Type	Feature sets	Feature sets	Feature sets	Feature sets	Emotion Classification	Feature sets + Emotion Classification	Emotion Classification	Emotion Classification & Reasoning
Real-Time Capability	Yes	Yes	Yes	Yes	No	Yes	No	No

Performance Metrics

Criteria	openSMILE	OpenFace	OpenFace 2.0	DeepFace	ResEmoteNet	LibreFace	Multi-Scale ViT	Emotion-LLaMA
Reported KPI's (%)	EMO-DB: 83.97% Romanian: 74% IEMOCAP: 60.87%	Average concordance correlation coefficient (CCC) on DISFA: 0.70	Average concordance correlation coefficient (CCC) on DISFA: 0.73	FER+: 90.25%	FER-2013: 79,79 RAF-DB: 94,76 AffectNet: 72.93	Pearson Correlation Coefficient (PCC) Average of AU compared to other LibreFace 0.63, OpenFace2 0.59	FER-2013: 99.6% CK+: 99.7%	F1 score: 0.9036 on MER2023-SEMI UAR: 45.59%, WAR: 59.37% on DFEW
Datasets Used	EMO-DB IEMOCAP, MELD, Romanian EMO-DB	LFPW, Helen, Multi-PIE	LFPW, Helen	FER+, CK+, LFW, Youtube Faces	FER-2013, RAF-DB, AffectNet	EmotioNet, AffectNet, FFHQ, RAF-DB, DISFA	FER-2013; CK+	MER2023, MER2024, DFEW, EMER

Overview of Comparative Results (3/3)

Implementation & Ease of Use

Criteria	openSMILE	OpenFace	OpenFace 2.0	DeepFace	ResEmoteNet	LibreFace	Multi-Scale ViT	Emotion-LLaMA
Ease of Setup	Moderate	Fairly Easy	Easy	Fairly Easy	Fairly Easy	Easy	Complex	Complex
Programming Language Support	C++ native, Python integration	C++, C#, Python	C++, C#, Python	Python	Python	Python, C#	MATLAB	Python
Integration Support	ML-Tools (WEKA, SVM, Python libraries)	C++ or C+ based project	C++ or C+ based project	API, compatible with python libraries	Compatible with python libraries	Cross-platform support	No	Compatible with python-based pipelines
Documentation Quality	Comprehensive	Comprehensive	Comprehensive	Comprehensive	Comprehensive	Comprehensive	Basic	Comprehensive

Summary & Recommendation

Criteria	openSMILE	OpenFace	OpenFace 2.0	DeepFace	ResEmoteNet	LibreFace	Multi-Scale ViT	Emotion-LLaMA
Key Strength	Real-Time + batch Support, Open Source, feature library, modular & integrable	Open-source, real-time performance, comprehensive facial behaviour analysis	Open-source, real-time performance, comprehensive facial behavior analysis,	Open-source, real-time, pre-trained models	Open-source, efficient, effective	Open-source, efficient, effective, documentation, possible integrations	Superior accuracy	Integration of audio, visual, and text inputs; robust performance across diverse datasets
Key Limitations	Requires expertise for setup, relies on external classifiers for ER	Optimal results at least 100x100 px	Optimal results at least 100x100 px	Computationally heavy due to LCL	Complexity, no real-time support	As OpenSence component only available for Windows	Complexity, no real-time support	Complexity due to multimodality, no real-time support
Recommendation / Best for	Detailed Audio Analysis	Real-Time FA for Researchers and developers	Real-Time FA for Researchers and developers	Real-Time FA for Researchers and developers	Accuracy > Speed (academic application)	Real-Time Video ER	Accuracy > Speed (academic application)	Research and applications requiring nuanced emotion recognition and reasoning

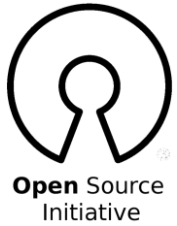
Prototype Implementation with LibreFace

Easy setup and customizable

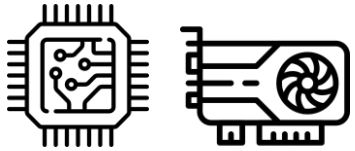
Cross-platform

Real-time

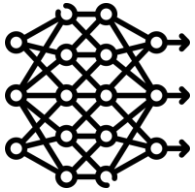
Pre-trained and trainable models



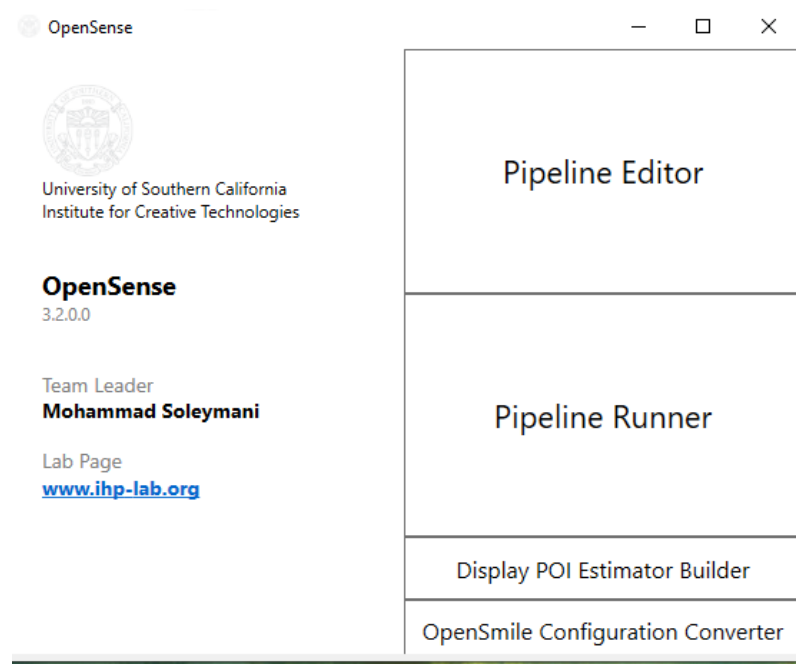
C# and Python

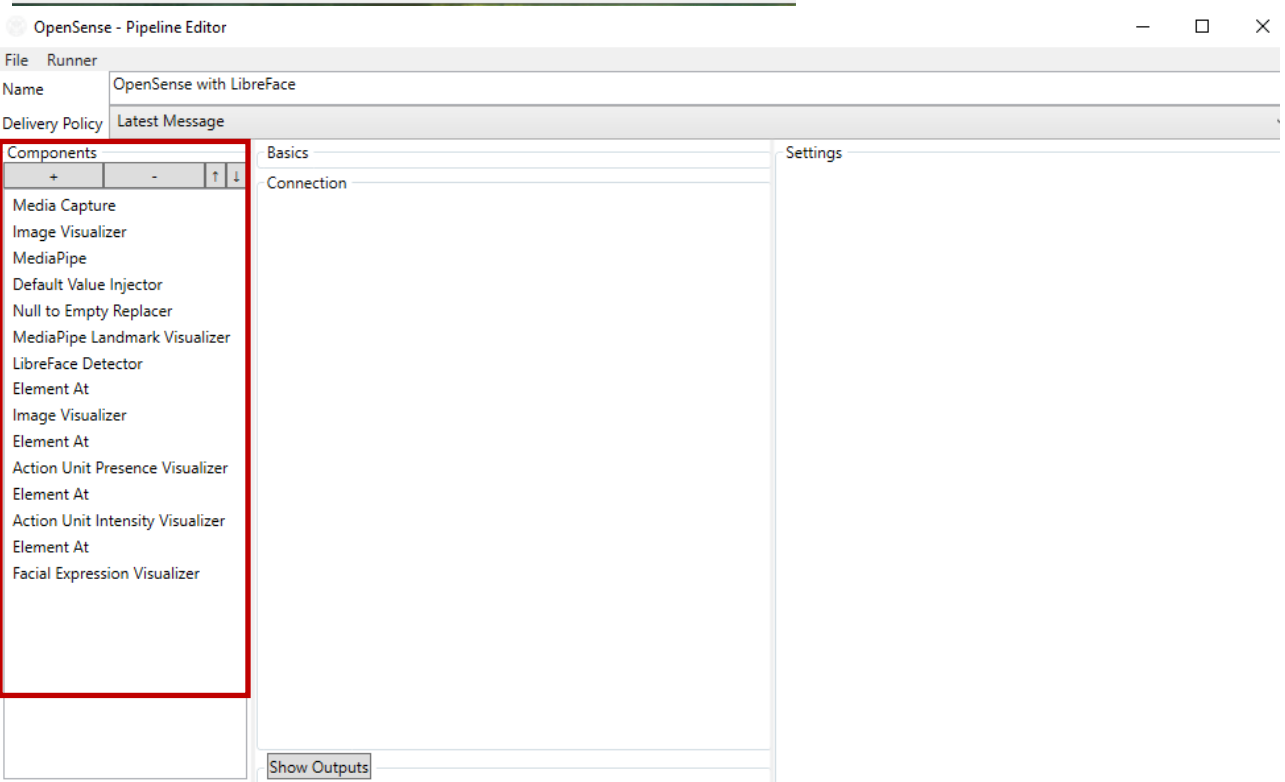


CPU only and GPU acceleration versions



AU + landmarks recognition and emotion classification


OpenSense
University of Southern California
Institute for Creative Technologies
OpenSense
3.2.0.0
Team Leader
Mohammad Soleymani
Lab Page
www.ihp-lab.org
Pipeline Editor
Pipeline Runner
Display POI Estimator Builder
OpenSmile Configuration Converter


OpenSense - Pipeline Editor
File
Runner
Name
OpenSense with LibreFace
Delivery Policy
Latest Message
Components
Media Capture
Image Visualizer
MediaPipe
Default Value Injector
Null to Empty Replacer
MediaPipe Landmark Visualizer
LibreFace Detector
Element At
Image Visualizer
Element At
Action Unit Presence Visualizer
Element At
Action Unit Intensity Visualizer
Element At
Facial Expression Visualizer
Basics
Connection
Settings
Show Outputs


Add Instance
Filter:
Name
Description
\psi Store Exporter
Writes streams to a \psi store.
\psi Store Importer
Reads streams from a \psi store.
Action Unit Intensity Visualizer
Visualize Action Unit intensity values.
Action Unit Presence Visualizer
Visualize Action Unit presence values.
Audio Capture
Captures audio from a microphone.
Audio Player
Playbacks audio signal to a speaker.
Audio Resampler
Converts and resamples audio signal to another format.
Audio Visualizer
Visualizes audio signal.
Azure Face Recognizer
Azure Face Recognition. Requires a subscription key.
Azure Speech Recognizer
Azure Speech Recognition. Requires a subscription key.
Biopac Visualizer
Visualize Biopac outputs.
Boolean Visualizer
Visualize boolean values.
CSV File Exporter
Write streams to a CSV file.
Deduplicator
Drop duplications. The default EqualityComparer will be used.
Default Value Injector
Pass-through data from the input stream to the output stream. If a timestamp is missing in the input stream
Depth Image Visualizer
Visualizes depth images.
Display POI Estimator
Estimate point of interest on a display based on gaze infomation. Requires OpenFace outputs and a regress
Display POI Visualizer
Visualizes point of interest outputs. Shows location in a rectangle. Requires display POI estimator outputs.
Element At
Return the element at a given index. No element will be returned if the index is out of range.
Emotion Visualizer
Visualizes emotion values. Requires emotion detector outputs.
Facial Expression Visualizer
Visualize facial expression values.
FFMpeg File Source
[Experimental] Read a file using FFMpeg. The used FFMpeg is a regular and LGPL version and is dynamically
Flip Image
Flip images vertically or horizontally or both.
Floating Point Value Visualizer
Visualizes floating point values. Accepts single or double precision.
Google Cloud Speech Recognizer
Google Cloud Speech V1 Recognizer. Requires credential file from Google. Only accepts 16bit PCM.
Head Gesture Detector
Detect head gesture (Nod, Shake or Tilt). Requires outputs from OpenFace.
Head Gesture Visualizer
Visualizes head gesture outputs. Shows the name of head gesture values. Requries head gesture detector ou
Image Decoder
Decodes encoded images to bitmap images. A Windows default implementation.
Image Encoder
Encodes bitmap images to another format. A Windows default implementation.
Image Visualizer
Visualizes images.
Join Operator
Joins the primary stream with values from a secondary stream.
JSON Store Exporter
Write streams to a JSON store.
JSON Store Importer
Read streams from a JSON store.
LibreFace
Detect Action Unit intensities, precenses and facial expressions. Requires MediaPipe face landmark detectior
Media Capture
Captures video and audio from a camera.
Media Source
Reads video and audio from a media file.
MediaPipe
Runs a MediaPipe pipeline. The backend is MediaPipe.Net which is based on MediaPipeUnityPlugin, and it's
MediaPipe Landmark Visualizer
Visualizes MediaPipe Normalized Landmark List Vector outputs. Requires MediaPipe outputs.
MPEG-4 Writer
Writes video and audio into an MPEG-4 file.
Null to Empty Replacer
Inputs are collections, and if an input is null, then replace it with an empty collection.
OpenFace
OpenFace by MultiComp Lab for head detection. This wapper of OpenFace can detect up to 1 person.
OpenFace Visualizer
Visualizes OpenFace outputs. Requires OpenFace outputs.
OpenPose
OpenPose by CMU Perceptual Computing Lab for image based body tracking. This wrapper of OpenPose re
OpenPose Visualizer
Visualizes OpenPose outputs. Requires OpenPose outputs.
openSMILE
openSMILE by audEERING GmbH for signal processing. Requires openSMILE pipeline configuration file.
openSMILE Visualizer
Visualizes openSMILE outputs. Requires openSMILE outputs.
Pixel Format Converter
Converts the pixel format of images.
Python 3
[Experimental] Uses Python 3 codes as an OpenSense component. The backend is IronPython3, and it suppc
Remote Exporter
Broadcasts streams.
Ok
Cancel

OpenSense - Pipeline Runner

File Editor PreSteps Run/Stop

Pipeline Information

Name OpenSense with LibreFace

Running ✓

Image Visualizer

Frame Rate 29.508

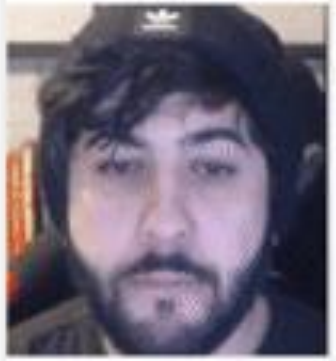
MediaPipe Landmark Visualizer

Settings

Frame Rate 29.508

Image Visualizer

Frame Rate 29.508



Action Unit Presence Visualizer

Settings

1	
2	
4	✔
6	✔
7	✔
10	✔
12	
14	✔
15	
17	✔
23	
24	✔

Facial Expression Visualizer

Neutral	**0.24**
Happiness	0.09
Sadness	0.21
Surprise	0.09
Fear	0.09
Disgust	0.09
Anger	0.09
Contempt	0.09

Action Unit Intensity Visualizer

1	0.19
2	0.10
4	0.91
5	0.10
6	0.22
9	0.72
12	0.40
15	0.19
17	0.58
20	0.12
25	0.04
26	0.07

Discussion & Critical Reflection

Strengths

- **Broad Tool Selection**: Evaluated single- and multimodal tools with diverse approaches (e.g., CNNs, Transformers, and hybrid models)
- **Comprehensive Comparison**: Developed a detailed table highlighting strengths, limitations, and real-time potential
- **Practical Implementation**: Balanced feasibility and technical complexity with **LibreFace**, showcasing real-time emotion detection in action

Challenges & Limitations

- **Data Imbalance & Quality**: Datasets often lack diversity (e.g., cultural differences) or have underrepresented emotional categories
- **Computational Complexity**: Real-time emotion detection remains resource-intensive, with significant hardware demands
- **Ethical Concerns**: Privacy, bias, and consent challenges are universal across tools but not central to this comparison
- **Dynamic Research Landscape**: Rapidly evolving field with constant influx of new tools and breakthroughs
- **Balancing Scientific Validity vs. Trends**: Struggled to align academic rigor with cutting-edge tools that lack extensive validation
- **KPI Comparison Challenges:** Variations in benchmarking methodologies, inconsistent experimental setups, and differences in how tools report performance metrics complicate meaningful comparisons

Conclusion

- **Comprehensive Review**: Explored state-of-the-art tools and frameworks for emotion detection, focusing on their modalities, performance, and real-time capabilities
- **Valence-Arousal & Ekman Models**: Highlighted foundational theories (e.g. Valence-Arousal, Ekman FACS) and their integration into tools like OpenFace, DeepFace, and LibreFace
- **Tool Evaluation**: Compiled and compared eight tools across criteria like modality, technical features, emphasizing real-time use cases
- **Prototype Implementation**: Demonstrated real-time emotion classification using **LibreFace**, balancing feasibility with practical outcomes
- **Future Research Directions**:
  - New technologies (Vision Transformers, LLaMA)
  - Enhanced multimodal approaches

Q&A

Scan the QR code to explore our GitHub repository and test LibreFace in action:

References

- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068
- Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System* [Dataset]. https://doi.org/10.1037/t27734-000
- Hutto, D. D., Robertson, I., & Kirchhoff, M. D. (2018). A New, Better BET: Rescuing and Revising Basic Emotion Theory. *Frontiers in Psychology*, *9*, 1217. https://doi.org/10.3389/fpsyg.2018.01217
- Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*(19), 7241–7244. https://doi.org/10.1073/pnas.1200155109
- Tracy, J. L., & Randles, D. (2011). Four Models of Basic Emotions: A Review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*(4), 397–405. https://doi.org/10.1177/1754073911410747