UNIVERSITY OF PISA

MASTER'S DEGREE IN COMPUTER ENGINEERING

Industrial Applications

MoodPilot

Professors:

Pierfrancesco Foglia

Antonio Cosimo Prete

Students:

Giovanni Ligato

Giuseppe Soriano

ACADEMIC YEAR 2024/2025

Index

1. Abstract **1**

2. Theoretical Background on Emotions **2**
2.1. Emotion Classifications 2
2.2. Discrete Emotion Theory: Roots, Universality, and Significance 2
2.2.1. Characteristics of Basic Emotions 2
2.2.2. Facial Expression Recognition and Universality 2
2.3. The Circumplex Model: Valence and Arousal 3

3. Facial Expression Recognition (FER) **4**
3.1. Datasets 4
3.1.1. FER2013 4
3.1.2. AffectNet 4
3.2. Models 4
3.2.1. DeepFace 4
3.2.2. HSEmotionONNX 4
3.2.3. Vision Transformer (ViT) for Facial Expression Recognition 5
3.2.4. Residual Masking Network (RMN) 5
3.2.5. EmoNet 5

4. References **6**

1. Abstract

2. Theoretical Background on Emotions

Emotions are complex psychological states that encompass subjective experiences, physiological responses, and behavioral expressions. They significantly influence human cognition and social interactions, affecting decision-making, perception, and relationships.

2.1. Emotion Classifications

Various models have been developed to categorize emotions:

- **Discrete Emotion Theory**: This theory suggests that humans possess a set of basic emotions that are universally recognizable. Paul Ekman's research identified six fundamental emotions: anger, disgust, fear, happiness, sadness, and surprise. Each is associated with distinct facial expressions and physiological patterns [1].
- **Dimensional Models**: These models represent emotions along *continuous* dimensions rather than discrete categories. The *Circumplex Model* is a prominent example, organizing emotions within a circular space defined by two axes: *valence* and *arousal* [2].

2.2. Discrete Emotion Theory: Roots, Universality, and Significance

Discrete Emotion Theory posits that humans experience a set of fundamental emotions, each with distinct characteristics and evolutionary purposes. This theory has its roots in Charles Darwin's work on the expression of emotions, where he argued that emotions serve adaptive functions for survival. Paul Ekman's modern research built on this foundation, identifying six core emotions—anger, disgust, fear, happiness, sadness, and surprise—through cross-cultural studies [1].

2.2.1. Characteristics of Basic Emotions

Each of the six core emotions has distinguishing features:

- **Anger**: Triggered by perceived threats or injustices, anger is characterized by increased arousal, narrowed attention, and physiological changes such as increased heart rate and adrenaline release.
- **Disgust**: Often a response to contaminants or morally offensive behavior, disgust manifests through a distinctive facial expression involving nose wrinkling and lip curling.
- **Fear**: A reaction to danger or threat, fear prompts the fight-or-flight response, enhancing focus and physiological readiness.
- **Happiness**: Associated with positive experiences and satisfaction, happiness is expressed through smiles and other outward signs of well-being.
- **Sadness**: Elicited by loss or disappointment, sadness is marked by lowered energy and slower cognitive processes.
- **Surprise**: Triggered by unexpected events, surprise is characterized by widened eyes and raised eyebrows, facilitating rapid information processing.

2.2.2. Facial Expression Recognition and Universality

Ekman's studies demonstrated that these emotions are universally recognized through facial expressions, even in cultures isolated from global influences. For instance, widened eyes and raised

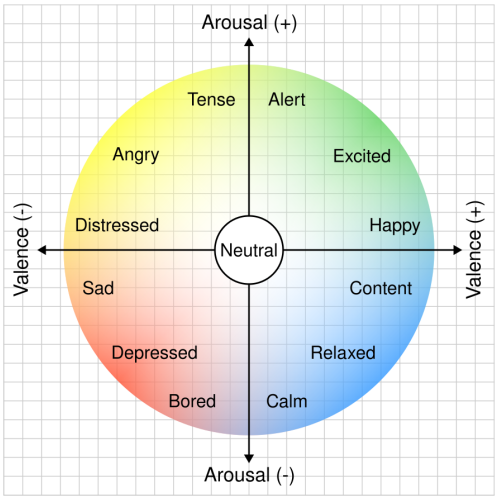

Figure 1: The Circumplex Model of Emotion maps emotions within a two-dimensional space defined by valence (X-axis) and arousal (Y-axis). Positive and negative valence correspond to pleasant and unpleasant emotions, respectively, while arousal indicates emotional intensity. Image taken from en.wikipedia.org/wiki/Emotion_classification

eyebrows universally signal surprise, while smiles denote happiness. This universality suggests a biological basis for basic emotions, enabling effective nonverbal communication.

2.3. The Circumplex Model: Valence and Arousal

Developed by James Russell, the Circumplex Model (depicted in Figure 1) maps emotions based on two dimensions:

- **Valence**: Represented on the horizontal (X) axis, valence measures the positivity or negativity of an emotion. Positive valence indicates pleasant emotions (e.g., happiness), while negative valence corresponds to unpleasant emotions (e.g., sadness).
- **Arousal**: Represented on the vertical (Y) axis, arousal gauges the intensity or activation level of an emotion, ranging from low (calmness) to high (excitement).

By plotting emotions within this two-dimensional space, the Circumplex Model illustrates how different emotions relate to one another. For instance, emotions like excitement (high arousal, positive valence) and calmness (low arousal, positive valence) are positioned accordingly, highlighting the continuous nature of emotional experiences.

These models are vital for fields like Facial Expression Recognition (FER), as they offer structured frameworks for interpreting and categorizing human emotions based on observable cues, thereby advancing human-computer interaction and social robotics.

3. Facial Expression Recognition (FER)

3.1. Datasets

The datasets described here will be referenced later in the Models section to indicate their use in training and evaluating the models. Below is a summary of the datasets:

3.1.1. FER2013

[3] is a dataset comprising 32,298 grayscale images of faces, each sized 48x48 pixels, labeled with seven emotions: Angry (0), Disgust (1), Fear (2), Happy (3), Sad (4), Surprise (5), and Neutral (6). Images are centered and uniformly scaled. The dataset includes 28,709 training samples and 3,589 test samples, with the task being to classify facial expressions into one of the seven categories.

3.1.2. AffectNet

[4] addresses the scarcity of annotated facial expression datasets in the wild, particularly for the **continuous dimensional model** (e.g., valence and arousal) alongside **discrete emotions** (categorical model). It contains over 1 million facial images collected from the internet using 1,250 emotion-related keywords in six languages. Approximately 440,000 images were manually annotated for eight discrete emotions and valence/arousal intensity. The dataset also includes "contempt" as an eighth emotion, making it more comprehensive than FER2013. AffectNet facilitates research in both categorical and dimensional emotion models.

3.2. Models

This section introduces several models for Facial Expression Recognition (FER). Each model is presented with a brief description, the dataset it utilizes (if not otherwise specified, for both training and testing), and its performance metrics on that dataset.

3.2.1. DeepFace

Description: DeepFace [5], [6] is a lightweight Python framework for face recognition and facial attribute analysis, including emotion detection, age, gender, and ethnicity prediction. It integrates various state-of-the-art models such as VGG-Face, FaceNet, and ArcFace.
Dataset Used: FER2013
Accuracy: 57.42%

3.2.2. HSEmotionONNX

Description: [7], [8], [9], [10], [11] A collection of ONNX-compatible (Open Neural Network Exchange standard) models *pre-trained* for face identification using VGGFace2 (Dataset for Face Recognition) and optimized for emotion recognition on AffectNet.
Dataset Used for Fine-Tuning: AffectNet
Accuracy:

- enet_b0_8_best_vgaf.pt: 61.32% (8 classes), 64.57% (7 classes)

- enet_b2_8.pt: 63.03% (8 classes), 66.29% (7 classes)

Inference Times:

- enet_b0_8_best_vgaf.pt: 59 ± 26 ms
- enet_b2_8.pt: 191 ± 18 ms

Model Sizes:

- enet_b0_8_best_vgaf.pt: 16 MB
- enet_b2_8.pt: 30 MB

The enet_b0_8_best_vgaf model is preferred due to its faster inference time, achieving higher FPS with minimal accuracy trade-offs.

3.2.3. Vision Transformer (ViT) for Facial Expression Recognition

Description: [12] A Vision Transformer model fine-tuned (from vit-base-patch16-224-in21k) for emotion recognition on facial images.
Dataset Used for Fine-Tuning: FER2013
Accuracy:

- Validation Set: 71.13%
- Test Set: 71.16%

3.2.4. Residual Masking Network (RMN)

Description: RMN [13] leverages Residual Masking Blocks to process facial features across multiple scales, ending with a 7-class softmax for emotion classification.
Dataset Used: FER2013
Accuracy: 74.14%

3.2.5. EmoNet

Description: [14] Trained models available for 5 and 8 emotional classes, also predicting valence, arousal, and facial landmarks.
Dataset Used: AffectNet

Accuracy:

- 5 Classes: 82%
- 8 Classes: 75%

Face Detection: Utilizes the SFD detector from the face-alignment repository, noted for its high accuracy but slower performance.

4. References

[1] W. contributors, "Emotion classification — Wikipedia, the free encyclopedia." 2023. Available: https://en.wikipedia.org/wiki/Emotion_classification

[2] T. F. Murphy, "Circumplex model of arousal and valence." 2024. Available: https://psychology fanatic.com/circumplex-model-of-arousal-and-valence/

[3] M. Sambare, "FER-2013: Learn facial expressions from an image." https://www.kaggle.com /datasets/msambare/fer2013, 2020.

[4] M. H. Mahoor, "AffectNet: Annotated facial database for affective computing." http://moha mmadmahoor.com/affectnet/.

[5] S. I. Serengil, "DeepFace: A lightweight face recognition and facial attribute analysis library for python." https://github.com/serengil/deepface, 2023.

[6] S. I. Serengil and A. Ozpinar, "HyperExtended LightFace: A facial attribute analysis framework," in *2021 international conference on engineering and emerging technologies (ICEET)*, IEEE, 2021, pp. 1–4. doi: 10.1109/ICEET53442.2021.9659697.

[7] A. Savchenko, "Facial expression recognition with adaptive frame rate based on multiple testing correction," in *Proceedings of the 40th international conference on machine learning (ICML)*, A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, Eds., in Proceedings of machine learning research, vol. 202. PMLR, 2023, pp. 30119–30129. Available: https://proceedings.mlr.press/v202/savchenko23a.html

[8] A. V. Savchenko, "Facial expression and attributes recognition based on multi-task learning of lightweight neural networks," in *Proceedings of the 19th international symposium on intelligent systems and informatics (SISY)*, IEEE, 2021, pp. 119–124. Available: https: //arxiv.org/abs/2103.17107

[9] A. V. Savchenko, "Video-based frame-level facial analysis of affective behavior on mobile devices using EfficientNets," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops*, 2022, pp. 2359–2366. Available: https: //arxiv.org/abs/2103.17107

[10] A. V. Savchenko, "MT-EmotiEffNet for multi-task human affective behavior analysis and learning from synthetic data," in *Proceedings of the european conference on computer vision (ECCV 2022) workshops*, Springer, 2023, pp. 45–59. Available: https://arxiv.org/abs/2207.0 9508

[11] A. V. Savchenko, L. V. Savchenko, and I. Makarov, "Classifying emotions and engagement in online learning based on a single facial expression recognition neural network," *IEEE Transactions on Affective Computing*, 2022, Available: https://ieeexplore.ieee.org/document /9815154

[12] Todor Pakov, "Vit-face-expression (revision 78ed8d3)." Hugging Face, 2024. doi: 10.57967/hf/2289.

[13] L. Pham, T. H. Vu, and T. A. Tran, "Facial expression recognition using residual masking network," in *2020 25th international conference on pattern recognition (ICPR)*, IEEE, 2021, pp. 4513–4519.

[14] A. Toisoul, J. Kossaifi, A. Bulat, G. Tzimiropoulos, and M. Pantic, "Estimation of continuous valence and arousal levels from faces in naturalistic conditions," *Nature Machine Intelligence*, 2021, Available: https://www.nature.com/articles/s42256-020-00280-0