DHBW
Duale Hochschule
Baden-Württemberg

Improving Performance in Real-Time Emotion Recognition

Studienarbeit

Course of Studies Informatik

Duale Hochschule Baden-Württemberg Karlsruhe

Jeremie Bents

Marvin Lindner

Submitted on: 10.03.2025

Student ID, Course: 1941564, TINF22B2
4274538, TINF22B2

Supervisor at DHBW: Prof. Dr. Roland Schätzle

Contents

1

Introduction

1.1 Motivation

The primary motivation behind this study is to develop a robust software system capable of automatically detecting smiling and laughing faces. The project is inspired by the popular series Last One Laughing (LOL), where the ability to monitor and analyze face expressions in real-time can significantly enhance the viewing experience. This study aims to leverage advanced face recognition and emotion detection technologies to create a service that can accurately identify and respond to face expressions. The literature review will cover existing methods and technologies in face recognition and emotion detection, highlighting the advancements and challenges in the field. The scope of this study includes the development, implementation, and evaluation of the proposed system.

1.2 Problem

The current state of face recognition and emotion detection technologies presents several challenges and limitations. While there have been significant advancements, existing systems often struggle with accuracy and real-time performance, especially in dynamic environments. The primary problem addressed in this study is the need for a reliable and efficient system that can detect smiling and laughing faces in real-time with high accuracy. [1] The gap identified is the lack of a specialized solution tailored for the specific requirements of the LOL series, including real-time monitoring, immediate response, and high reliability. This study aims to bridge this gap by developing a system that meets these specific needs.

> **TODO**
>
> Problem: Grund für unsere Lösung (LOL) ist nicht fundiert. Keine Quelle dafür

1.3 Research Framework and Objectives

The importance of this research lies in its potential to enhance the viewing experience of the LOL series by providing a reliable and efficient face expression detection system. The research problem is to develop a system that can accurately detect smiling and laughing faces in real-time. The research aims and objectives include:

- Developing a face recognition module to detect faces in camera streams.
- Implementing an emotion detection module to identify smiling and laughing expressions.
- Ensuring real-time performance and immediate response to detected expressions.
- Evaluating the system's accuracy and reliability through comprehensive testing.

The hypotheses of this study are:

1. The proposed system will accurately detect smiling and laughing faces in real-time.
2. The system will provide immediate responses to detected expressions with minimal latency.

The methodology includes the development of the system using Python, leveraging existing libraries and frameworks for face recognition and emotion detection. The study will also involve extensive testing and evaluation to ensure the system meets the defined requirements.

1.4 Thesis Structure

The order of chapter in this thesis will follow a structured approach:

1. **Introduction**: Provides the motivation, problem statement, and research objectives.
2. **Fundamentals**: Reviews existing methods and technologies in face recognition and emotion detection.
3. **Methodology**: Details the development process, tools, and techniques used.
4. **Implementation**: Describes the implementation of the face recognition and emotion detection modules.
5. **Testing and Evaluation**: Presents the testing procedures, results, and evaluation of the system's performance.

6. **Conclusion**: Summarizes the findings, discusses the implications, and suggests future work.

> **TODO**
>
> Diese Struktur ist sehr generisch und im Verlaufe der Dokumentation unserer Ergebnisse und der Entwicklung des Systems werden wir die Struktur anpassen müssen.

2

Fundamentals (new)

2.1 Neural Networks

Neural Network (NN) are a type of Machine Learning (ML) algorithm inspired by the structure and function of the human brain. They are composed of interconnected nodes, or "neurons," organized in layers. These networks are designed to recognize patterns in data and learn from experience, making them capable of performing complex tasks such as image recognition, natural language processing, and decision-making.

2.1.1 History of Neural Networks

Neural networks have come a long way since their origin in the 1940s. Here is a brief overview of their evolution over the decades.

1940s to 1970s

The birth of NN can be traced back to 1943 when Warren McCulloch and Walter Pitts published their groundbreaking paper on how neurons might function. They proposed a simple model of NNs using electrical circuits, laying the foundation for future research in the field. [2]

In 1949, Donald Hebb's seminal work, "The Organization of Behavior," introduced the concept of neural plasticity. Hebb proposed that neural pathways are strengthened through repeated use, a principle now known as Hebbian learning. This concept

became fundamental to our understanding of how humans learn and would later influence the development of artificial NN. [3]

The Dartmouth Conference in 1956, officially known as the Dartmouth Summer Research Project on Artificial Intelligence, is considered the founding event of the field of Artificial Intelligence (AI). Organized by John McCarthy and others, it aimed to explore the potential for machines to simulate human intelligence through collaborative brainstorming among leading researchers [4]

A significant breakthrough came in 1959 when Bernard Widrow and Marcian Hoff of Stanford University developed the Adaptive Linear Element (ADALINE) and Many ADALINE (MADALINE) models. ADALINE was designed to recognize binary patterns, while MADALINE became the first NN applied to a real-world problem: eliminating echoes on phone lines [5]

In 1962, Widrow and Hoff introduced a learning procedure that would later influence the development of backpropagation algorithms. Their method examined the value before adjusting the weight, distributing the error across the network. This approach was a significant step towards creating more efficient learning algorithms for NN. [5]

Despite these advancements, the field of NNs faced a setback in the late 1960s and early 1970s. The rise of traditional von Neumann architecture in computing overshadowed neural network research. Additionally, a paper suggesting the impossibility of extending single-layered networks to multiple layers further dampened enthusiasm in the field. Coupled with unfulfilled promises and philosophical concerns about "thinking machines," funding and research in neural networks declined significantly during this period. [5]

However, the field was not entirely dormant. In 1972, Teuvo Kohonen and James Anderson independently developed similar networks that would later contribute to the resurgence of interest in neural networks. In 1975, the first multilayered network was developed, albeit an unsupervised one. [6]

1980s to Present

The 1980s marked a renaissance for neural networks. In 1982, John Hopfield's presentation to the National Academy of Sciences introduced the concept of bidirectional connections in neural networks, sparking renewed interest in the field. The same year, Reilly and Cooper developed a "Hybrid network" with multiple layers, each employing different problem-solving strategies. [5]

A pivotal moment came in 1986 when multiple research groups, including one led by David Rumelhart, independently developed the backpropagation algorithm. This breakthrough allowed for the training of multi-layer networks, greatly expanding the capabilities and potential applications of neural networks. [5]

The 1990s and 2000s saw an explosion of research and practical applications of neural networks. They began to be used in various fields, including pattern recognition, financial forecasting, and medical diagnosis. The advent of more powerful computing hardware and the availability of large datasets further accelerated progress in the field. [6]

In recent years, deep learning, a subset of neural networks with many layers, has revolutionized artificial intelligence. Breakthroughs in areas such as image and speech recognition, natural language processing, and game-playing AI (like AlphaGo) have been achieved using deep neural networks. [6]

Current research focuses on developing more efficient hardware for neural network computation, including specialized chips and optical computing. The goal is to create faster, more energy-efficient neural networks capable of learning and adapting in real-time. [6]

2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a specialized type of deep neural network that are particularly effective for processing data with a grid-like topology, such as images, videos, and time-series data.

2.2.1 Convolution

Convolution serves is important in various fields, e.g. image processing. It elegantly combines two functions, revealing how the shape of one is modified by the other. Understanding its principles is crucial for comprehending the mechanisms underlying many data processing techniques, including those employed in neural networks. [7]

At its core, convolution involves a dynamic interaction between two functions: an input signal, representing the data to be processed, and a kernel, a function acting as a filter or feature detector. This interaction can be visualized as a "sliding" operation. The kernel is first "flipped" or reflected about its origin, then systematically shifted across the input signal. At each position, a point-wise multiplication occurs between the kernel's values and the corresponding values of the input signal. These products are then summed, yielding a single output value. This process is repeated as the kernel slides across the entire input, generating the output signal. [7]

Intuitively, convolution can be interpreted as a form of weighted averaging. The kernel acts as a set of weights, emphasizing or suppressing specific features in the input. For instance, a kernel with uniform weights smooths the input, reducing noise. Conversely, a kernel with sharp transitions highlights edges or abrupt changes. [7]

In image processing, convolution is instrumental for tasks such as blurring, sharpening, and edge detection. By applying different kernels, one can extract various features from an image. A kernel with a Gaussian distribution blurs the image, while a kernel designed to detect intensity gradients identifies edges. This ability to extract features is fundamental to Convolutional Neural Networks (CNNs), where convolution layers learn to automatically extract relevant features from input images. [7]

2.2.2 History

The beginnings of Convolutional Nerual Network (CNN) start in the late 1950s and early 1960s with the work of Hubel and Wiesel, who studied the visual cortex of cats. Their research revealed that neurons in the visual cortex respond to specific patterns

of light and dark, organized in a hierarchical manner. This discovery inspired the concept of receptive fields, where neurons respond to local regions of the input, laying the foundation for the convolutional operation. [7]

In the 1980s, Kunihiko Fukushima introduced the Neocognitron, a hierarchical, multilayered neural network designed to recognize handwritten characters. The Neocognitron incorporated concepts such as local receptive fields and weight sharing, which are fundamental to modern CNNs. However, due to computational limitations and the lack of effective training algorithms, the Neocognitron did not achieve widespread adoption. [8]

A significant leap forward occurred in the late 1980s and early 1990s with the work of Yann LeCun and his colleagues at Bell Labs. They developed LeNet, a CNN architecture designed for handwritten digit recognition. LeNet demonstrated the effectiveness of backpropagation for training CNNs and showcased their ability to learn hierarchical representations of visual data. LeNet's success in recognizing handwritten digits for postal code recognition highlighted the practical potential of CNNs. [7]

Despite LeNet's advancements, CNNs remained relatively niche for several years. The lack of powerful computing hardware and large datasets hindered their application to more complex tasks. However, the early 2000s witnessed a confluence of factors that catalyzed a resurgence in CNN research. The availability of powerful GPUs and the emergence of large-scale image datasets, such as ImageNet, provided the necessary resources for training deeper and more complex CNN architectures. [7]

The watershed moment arrived in 2012 with the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton introduced AlexNet, a deep CNN that achieved a groundbreaking performance, significantly outperforming traditional machine learning methods. AlexNet's success demonstrated the power of deep CNNs for large-scale image recognition and ignited the modern deep learning revolution. [9]

Following AlexNet, numerous advancements further enhanced CNN architectures. VGGNet introduced deeper networks with smaller convolutional filters, demonstrating the benefits of increased depth. GoogLeNet and Inception architectures incorporated inception modules to capture features at multiple scales. ResNet introduced residual connections, enabling the training of extremely deep networks. [7]

Beyond image recognition, CNNs have found applications in diverse fields, including natural language processing, speech recognition, and medical image analysis. Their ability to learn hierarchical representations and extract relevant features from complex data has made them a cornerstone of modern artificial intelligence. [7]

> **TODO**
>
> add source from here on out

2.2.3 Face Recognition

Face recognition technology allows computers to identify people from images or videos. This chapter explores how this technology developed, from early attempts to modern systems using advanced computer learning.

History

Early Research

Before deep learning, face recognition relied on various techniques that, while pioneering, faced significant limitations. Early methods focused on extracting geometric features or comparing pixel patterns directly. One prominent approach was the use of eigenfaces. This technique, developed in the early 1990s, employed Principal Component Analysis (PCA) to reduce the dimensionality of face images, representing faces as linear combinations of eigenvectors. While effective under controlled lighting and pose conditions, eigenfaces struggled with variations in illumination, head orientation, and facial expressions.

Another early strategy involved geometric feature-based methods. These techniques aimed to measure distances and ratios between facial landmarks, such as the eyes, nose, and mouth. By comparing these measurements, systems could attempt to identify individuals. However, the accuracy of these methods was highly dependent on precise landmark detection and was also vulnerable to variations in pose and expression.

Template matching was also employed, where a stored image of a face was directly compared to an input image. These methods were computationally simple but extremely sensitive to changes in lighting, pose, and scale. These early attempts, while laying the groundwork for future advancements, highlighted the need for more robust and adaptable techniques capable of handling real-world variations.

Impact

The emergence of Convolutional Neural Networks (CNNs) revolutionized face recognition. CNNs' ability to automatically learn hierarchical features from raw pixel data significantly improved accuracy and robustness. Key architectures like DeepFace

(Facebook) demonstrated the power of deep learning in achieving near-human-level face recognition performance. DeepFace utilized a deep neural network with multiple layers to learn complex representations of faces, achieving a breakthrough in accuracy on benchmark datasets.

FaceNet (Google) introduced the concept of embedding faces into a high-dimensional space where similar faces are close together. It used a triplet loss function, which trained the network to distinguish between different individuals while ensuring that faces of the same person were tightly clustered. This approach provided a highly efficient and accurate method for face verification and identification.

DeepID (Chinese University of Hong Kong) focused on learning discriminative features for face identification. It demonstrated the effectiveness of training deep networks with large-scale datasets and advanced loss functions, leading to significant improvements in face recognition accuracy.

CNNs addressed many of the limitations of earlier methods. They could handle variations in pose, lighting, and occlusion through their learned feature representations. Data augmentation techniques further enhanced the robustness of these models by exposing them to a wide range of input variations during training.

Modern Face Recognition

Modern face recognition systems leverage vast datasets like MS-Celeb-1M and VGGFace2, enabling the training of highly accurate models. Advanced loss functions, such as ArcFace and CosFace, have been developed to further improve discriminative power by maximizing inter-class variance and minimizing intra-class variance.

Applications of face recognition have expanded significantly, including security systems, access control, social media tagging, personalized experiences, and law enforcement. However, the widespread use of face recognition has raised significant ethical concerns. Issues such as privacy violations, potential for misuse, and algorithmic bias have prompted calls for regulations and responsible development practices.

2.2.4 Facial Expresson Recognition

Facial expression recognition, or Facial Expression Recognition (FER), enables computers to understand emotions by analyzing facial expressions. This chapter explains the history of FER, from basic emotion studies to the use of powerful computer programs.

History

Early Research

The foundation of emotional recognition research was laid by the pioneering work of Paul Ekman, who identified six basic emotions universally expressed through facial expressions: happiness, sadness, anger, fear, surprise, and disgust.

Early attempts at automated FER relied on the Facial Action Coding System (FACS). FACS, developed by Ekman and Friesen, provides a comprehensive system for coding facial muscle movements. Researchers used FACS to analyze facial expressions and develop rule-based systems for emotion classification. These systems relied on predefined rules that mapped specific AU combinations to emotional states.

However, rule-based systems were limited by their reliance on manual feature extraction and their inability to handle subtle variations in expressions. They also struggled with the complexity of real-world scenarios, where expressions are often nuanced and context-dependent.

CNNs and FER

The application of CNNs to FER has significantly improved accuracy and robustness. CNNs can automatically learn relevant features from facial images, eliminating the need for manual feature extraction. Deep learning models have demonstrated superior performance in recognizing a wide range of emotions, including subtle and compound expressions.

Data augmentation plays a crucial role in training effective FER models. Techniques such as random cropping, flipping, and rotation are used to increase the diversity of training data and improve the model's ability to generalize. Due to the relative scarcity of labeled emotional data, data augmentation is extremely important in this field.

For video-based FER, temporal information is crucial. Recurrent Neural Networks (RNNs) and 3D CNNs are used to capture the temporal dynamics of facial expressions.

RNNs can model the sequential nature of expressions, while 3D CNNs can extract spatio-temporal features directly from video frames.

Despite advancements, FER still faces challenges. Subtle expressions, cultural differences in expression, dataset bias, and the need to incorporate contextual information remain active areas of research. Ethical concerns regarding the potential for misinterpretation and misuse of FER technology have also emerged, emphasizing the need for responsible development and deployment.

3

Methodology

- Data Collection and Preprocessing Techniques (e.g., dataset selection, annotation strategies, data sanitization)
- Selection of Model Architecture and Algorithms (e.g., Convolutional Neural Networks, Recurrent Neural Networks)
- Model Training Protocols and Performance Metrics for Evaluation

Review of Current Research and Technologies

- Systematic Overview of Contemporary Methods for face Expression Recognition
- Comparative Analysis of Existing Emotion Recognition Systems
- Identification of Gaps, Challenges, and Limitations in Current Technologies

3.1 Wie hilf Tensorflow uns bei der Entwicklung

3.2 Wozu dient OpenCV

4

Implementation of the Model

5

Implementation of the Website

6

Experimental Setup and Results

- Design of Experiments and Test Environment
- Quantitative Evaluation of System Accuracy (e.g., accuracy, precision, recall, F1 score)
- Comprehensive Analysis and Interpretation of Experimental Findings

7

Conclusion

A References

[1] García-Hernández, Rosa A. *et al.*, "A Systematic Literature Review of Modalities, Trends, and Limitations in Emotion Recognition, Affective Computing, and Sentiment Analysis," *Applied Sciences*, vol. 14, no. 16, p. 7165, Aug. 2024, doi: 10.3390/app14167165.

[2] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943, doi: 10.1007/bf02478259.

[3] D. Hebb, *The Organization of Behavior*. Psychology Press, 2005. doi: 10.4324/9781410612403.

[4] J. Mccarthy, M. Minsky, N. Rochester, and C. Shannon, "A Proposal for the Dartmouth Summer Research Project on Arti cial Intelligence," *AI Magazine*, vol. 27, 2006.

[5] D. Graupe, *Principles of Artificial Neural Networks: Basic Designs to Deep Learning*. WORLD SCIENTIFIC, 2018. doi: 10.1142/11306.

[6] J. A. Anderson and E. Rosenfeld, Eds., *Talking nets*, First MIT Press paperback edition, 2000. in A Bradford Book. Cambridge, Massachusetts: MIT Press, 2000.

[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[8] K. Fukushima, S. Miyake, and T. Ito, "Neocognitron: A neural network model for a mechanism of visual pattern recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 5, pp. 826–834, Sep. 1983, doi: 10.1109/tsmc.1983.6313076.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

B Acronyms

ADALINE	Adaptive Linear Element
AI	Artificial Intelligence
API	Application Programming Interface
CNN	Convolutional Nerual Network
FER	Facial Expression Recognition
HTTP	Hypertext Transfer Protocol
LOL	Last One Laughing
MADALINE	Many ADALINE
ML	Machine Learning
NN	Neural Network
REST	Representational State Transfer

C Glossary

Komponente Ein Architekturbaustein. Zusammengesetzte Komponenten bestehen aus weiteren Subkomponenten. Einfache Komponenten sind nicht weiter unterteilt.

Softwareschnittstelle Ein logischer Berührungspunkt in einem Softwaresystem: Sie ermöglicht und regelt den Austausch von Kommandos und Daten zwischen verschiedenen Prozessen und Komponenten.

Declaration of Authorship

Gemäß Ziffer 1.1.13 der Anlage 1 zu §§ 3, 4 und 5 der Studien- und Prüfungsordnung für die Bachelorstudiengänge im Studienbereich Technik der Dualen Hochschule Baden-Württemberg vom 29.09.2017. Wir versichern hiermit, dass wir unsere Arbeit mit dem Thema:

Improving Performance in Real-Time Emotion Recognition

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt haben. Wir versichern zudem, dass alle eingereichten Fassungen übereinstimmen.

Karlsruhe, 10.03.2025

Jeremie Bents

Marvin Lindner