XAIface

Measuring and Improving Explainability for AI-based Face Recognition

Programm / Ausschreibung	IKT der Zukunft, IKT der Zukunft, IKT der Zukunft - Vorbereitung Horizon Europe	**Status**	laufend
Projektstart	01.05.2021	**Projektende**	30.04.2024
Zeitraum	2021 - 2024	**Projektlaufzeit**	36 Monate
Keywords	face recognition, explainability, robustness, bias		

Projektbeschreibung

Gesichtserkennung ist zu einer Schlüsseltechnologie in unserer Gesellschaft geworden, die in verschiedenen Anwendungen eingesetzt wird und gleichzeitig Auswirkungen auf die Privatsphäre hat. Da Gesichtserkennungslösungen, die auf maschinellem Lernen basieren, immer häufiger verwendet werden, ist es entscheidend, die Funktionsweise dieser Technologie vollständig zu verstehen und zu erklären.
In diesem Projekt konzentrieren wir uns auf die Anwendungsfälle der Gesichtserkennung auf der Grundlage künstlicher Intelligenz und die Analyse der für die endgültige Entscheidung relevanten Einflussfaktoren als wesentlicher Schritt zum Verständnis und zur Verbesserung der zugrunde liegenden Prozesse. Der im Projekt verfolgte wissenschaftliche Ansatz ist so konzipiert, dass er auch andere Anwendungsfälle wie Objekterkennung und Mustererkennungsaufgaben in einer breiteren Palette von Anwendungen bewältigen kann.
Einer der wesentlichen Aspekte des vorgeschlagenen Projekts liegt in seinem wissenschaftlichen Ansatz, der darauf abzielt, zu erklären, wie maschinelle Lernlösungen eine effektive Gesichtserkennung erreichen, indem Einflussfaktoren, die eine wichtige Rolle bei der Leistung der Gesichtserkennung im End-to-End-Workflow spielen, und ihre Auswirkungen identifiziert und analysiert werden. Diese Leistung hängt weitgehend von den Prozessen der Erfassung, Verbesserung, Komprimierung, Analyse und Entscheidungsfindung ab, die im Workflow einer Gesichtserkennungslösung angewendet werden. Maschinelles Lernen kann und wird derzeit in vielen Phasen eines solchen Arbeitsablaufs eingesetzt, und verschiedene Faktoren wie der im Training verwendete Datensatz, der für das Training selbst verwendete Ansatz, die Architektur des maschinellen Lernens und die Arten von Angriffen und Interferenzen gehören zu den Einflussfaktoren, die zum Verständnis und zur Erklärbarkeit des gesamten Systems beitragen.

Abstract

Face recognition has become a key technology in our society, frequently used in multiple applications, while creating an impact in terms of privacy. As face recognition solutions based on machine learning are becoming popular it is critical to fully understand and explain how these technologies work in order to make them more effective and accepted by society.
In this project, we focus on face recognition technologies based on artificial intelligence and the analysis of the influencing factors relevant for the final decision as an essential step to understand and improve the underlying processes involved. The

scientific approach pursued in the project is designed in such a way that it will be applicable to other use cases such as object detection and pattern recognition tasks in a wider set of applications.
One of the original aspects of the proposed project is in its scientific approach which targets explaining how machine learning solutions reach effective face recognition by identifying and analyzing the influencing factors that play an important role in the performance of face recognition in the end-to-end workflow, and their impact on the system’s decisions. In fact, such performance largely depends on the acquisition, enhancement, compression, analysis and decision making processes adopted in the workflow of a face recognition solution. Machine learning is currently used in many of the stages of such a workflow and various factors such as the dataset used in the training process, the approach used for training itself, the architecture of machine learning, and the types of attacks and interferences are among influencing factors that contribute to the understanding and explainability of the complete system.

Projektpartner

- JOANNEUM RESEARCH Forschungsgesellschaft mbH