Gesichtserkennung: Bedeutung verschiedener Gesichtsteile

JOHANNES KEPLER UNIVERSITY LINZ

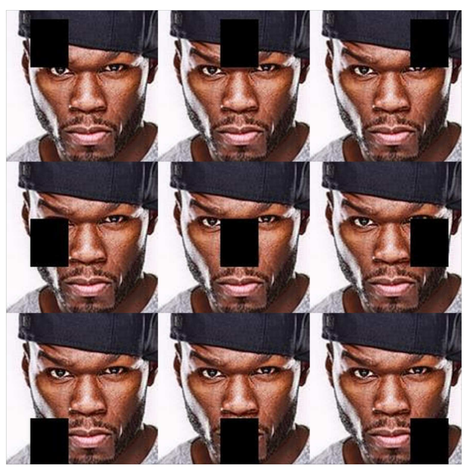
INSTITUTE OF NETWORKS AND SECURITY

Motivation

Die meisten modernen Gesichtserkennungsalgorithmen sind mit Bildern von Personen ohne signifikanter Verdeckung von Gesichtsteilen trainiert. Wie verhalten sich diese Modelle, wenn Personen einen Großteil von ihrem Gesicht, z.B. durch eine Maske, verdecken? Dabei wurde eine interessante Lücke in einem modernen Gesichtserkennungsnetzwerk – MTCNN – entdeckt: Wenn sich ein Gesicht innerhalb eines anderen Gesichts befindet, wird das größere Gesicht nicht erkannt *[Bild links]*.

Um die Genauigkeit von Gesichtserkennungsalgorithmen zu verbessern, kann als Referenzbild nicht nur ein Bild verwendet werden, sondern mehrere. Vor allem für dezentrale Systeme, wie z.B. Digidow, ist es von Bedeutung, möglichst wenig zusätzliche Laufzeit- und Netzwerk-Komplexität zu erhalten. Aus einem Bild wird durch Gesichtserkennungsalgorithmen eine numerische Repräsentation von diesem berechnet (*Embedding)*. Wie können mehrere dieser Embeddings aggregiert werden, um möglichst viel Information mit möglichst wenig Speicherplatz zu verwenden?

Visualisierung vom Gesicht-im-Gesicht Defekt



Systematische Analyse der Genauigkeit, wenn gewisse Teile verdeckt sind

Ein signifikanter Teil des Gesichts ist verdeckt

Verhalten bei verdeckten Gesichtsteilen

Auf der einen Seite verwenden wir den CFP Datensatz als Grundlage, und blenden gewisse Teile des Gesichts aus. Im Speziellen ist das ein Raster in 4 Stufen (2x2, 3x3 *[Bild Mitte]*, 4x4, 5x5), Orientierungspunkte (Augen, Nase, Mund) und die untere Hälfte des Gesichts in verschiedenen Größen um eine Maske zu simulieren. Zentrales Ergebnis hier ist, dass der Bereich rund um die Nase die wichtigste Rolle spielt. Ein wahrscheinlicher Grund, warum dies nicht die Augen sind ist, dass moderne Netze bereits viele Bilder ohne Augen (in Form von Sonnenbrillen) gesehen haben, und daher mit dieser Situation besser zurecht kommen.

Auf der anderen Seite verwenden wir reale Gesichtsmaskenbilder vom RMFD Datensatz. Dabei erkannten moderne Gesichtserkennungsalgorithmen 45% der Personen. Im *Bild rechts* sind Beispiele aus dem Datensatz abgebildet. Daran ist klar zu erkennen, dass eine Maske einen Großteil des Gesichts verdeckt und deshalb Netzwerke, die nicht spezifisch auf diesen Umstand trainiert wurden, damit zu kämpfen haben.

Aggregierung von Embeddings

Nachdem in einem einzigen Bild aufgrund des Winkels nur ein Teil des Gesichts abgebildet werden kann, externe Umstände wie z.B. Beleuchtung und Kameraqualität sich ständig ändern, sich Personen über Zeit verändern (Bart, Falten, Tattoo, ...) und verschiedene Gegenstände getragen werden, ist die Verwendung von mehreren, verschiedenen Bildern für Gesichtserkennung vorteilhaft. Wenn 10 Bilder für das Referenzbild verwendet werden, halbiert sich die Fehlerrate bei modernen Algorithmen. Ab 50 Bildern erhöht sich die Genauigkeit nicht mehr signifikant.

Autor

Philipp Hofer

Johannes Kepler Universität Linz, Institut für Netzwerke und Sicherheit

philipp.hofer@ins.jku.at

https://jku.at/ins • https://digidow.eu/

DIGiDOW