



MASTERARBEIT

Frau
Maria Lietsch, B.Sc.

**Evaluation der Vorgehensweise und
Eigenschaften von forensischen
Super-Recognizern**

Mittweida, Juli 2024

MASTERARBEIT

Evaluation der Vorgehensweise und Eigenschaften von forensischen Super-Recognizern

Autorin:

Maria Lietsch

Studiengang:

Cybercrime/ Cybersecurity

Seminargruppe:

CY21wC-M

Erstprüfer:

Prof. Dr. rer. nat. Dirk Labudde

Zweitprüferin:

Svenja Preuß, M.Sc.

Einreichung:

Mittweida, 29.07.2024

Verteidigung/Bewertung:

Mittweida, 2024

MASTER THESIS

Evaluation of the approach and properties of forensic super recognizers

Author:

Maria Lietsch

Course of Study:

Cybercrime/ Cybersecurity

Seminar Group:

CY21wC-M

First Examiner:

Prof. Dr. rer. nat. Dirk Labudde

Second Examiner:

Svenja Preuß, M.Sc.

Submission:

Mittweida, 29.07.2024

Defense/Evaluation:

Mittweida, 2024

Bibliografische Beschreibung

Lietsch, Maria:

Evaluation der Vorgehensweise und Eigenschaften von forensischen Super-Recognizern. – 2024. – 64 S.

Mittweida, Hochschule Mittweida – University of Applied Sciences, Fakultät Angewandte Computer- und Biowissenschaften, Masterarbeit, 2024.

Referat

Diese Arbeit befasst sich mit den bisherigen Studien über Super-Recognizer. Sie beinhaltet die Auswertung der Studie der Hochschule Mittweida aus dem Jahr 2023, welche von Super-Recognizern der Polizeidirektion Chemnitz absolviert wurde. Ausgehend von den Ergebnissen wurde eine weitere Studie entwickelt, bei der fünf neue Forschungsfragen untersucht werden sollen, auf die Super-Recognizer getestet werden.

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	III
Tabellenverzeichnis	V
1 Einleitung	1
1.1 Hintergrund	1
1.2 Einsatzmöglichkeit in der Forensik	1
1.3 Zielsetzung	1
1.4 Hypothesen	2
2 Grundlagen	3
2.1 Begriffsbestimmung	3
2.1.1 Super-Recognizer	3
2.1.2 Prosopagnosie	3
3 Stand der Forschung	5
3.1 Tests zur Bestimmung von Super-Recognizer	5
3.1.1 CFMT +	5
3.1.2 CFPT	7
3.1.3 BTWF	8
3.2 Literaturreview	9
3.2.1 Fortysomething: Recognizing faces at one's 25th reunion	9
3.2.2 Super-recognizers: People with extraordinary face recognition ability	12
3.2.3 Human face recognition ability is specific and highly heritable	15
3.2.4 Investigating Predictors of Superior Face Recognition Ability in Police SuperRecognizers	16
3.2.5 Identification from CCTV: Assessing police super-Recognizer ability to spot faces in a crowd and susceptibility to change blindness	23
3.2.6 Super-recognizers: From the lab to the world and back again	30
3.2.7 Face recognition in police officers: Who fits the bill?	36
4 Studie Super-Recognizer Workshop 2023	45
4.1 Erstellung	45
4.2 Aufbau	47
4.3 Künstliche Intelligenzen	49
4.3.1 Face Recognition	49
4.3.2 FaceX-Zoo	50
4.3.3 GhostFaceNet	50
4.4 Auswertung	52
5 Methodik	55
5.1 Datensatzerstellung	55
5.1.1 Erstellung	55
5.1.1.1 Komplex 1	55

5.1.1.2	Komplex 2	56
5.1.1.3	Komplex 3	57
5.1.1.4	Komplex 4	58
5.1.1.5	Komplex 5	59
5.1.2	Aufbau Datensatz	61
6	Diskussion und Ausblick	63
6.1	Studie der Hochschule Mittweida 2023	63
6.2	Studie in dieser Arbeit	63
6.3	Ausblick	64
Anhang		65
A	Bilder der Studie	65
Literaturverzeichnis		83
Eidesstattliche Erklärung		87

Abbildungsverzeichnis

3.1	Zielgesicht [9]	5
3.2	Testelement [9]	6
3.3	Testelement 2 [9]	6
3.4	Bilder der Zielgesichter mit veränderter Beleuchtung und/oder Pose[9]	6
3.5	Bilder der Zielgesichter mit Rauschen [9]	7
3.6	Testelement 3[9]	7
3.7	Beispiele Bilder im CFMT long [27]	7
3.8	Beispiele Bilder aus CFPT Test, oben Zielgesicht, unten die zu sortierenden Gesichter [27]	8
3.9	Beispiele Bilder aus BTWF Test [27]	8
4.1	Beispiel Bilder nach Transformation (erste Reihe, mittleres und letztes Bild) sowie nach der Bildbearbeitung (zweite Reihe) und Anwendung von künstlerischen Effekten (dritte Reihe) unter Verwendung eines Beispielbildes aus dem LFW-Datensatz (erste Reihe, erstes Bild). Quelle des Originalbildes [14]	46
4.2	Vergleichsbild von Set 1	47
4.3	Auswahlbilder von Set 1	48
4.4	Vergleichsbilder von Set 4-9	48
4.5	Input mit ablenkenden irrelevanten Gesichtern im Hintergrund (links) und benutzter Input mit geschwärzten Gesichtern im Hintergrund (rechts)	50
4.6	Ergebnisse der 10 SR und der drei KI´s	52
5.1	Beispiel der verwendeten Gesichtsfiler	56
5.2	Gesicht von (2) auf (1) legen ergibt (3)	58
5.3	Originalbild (links), reduzierte Auflösung (50%, 40%, 30%, 20%, 10%)	59
5.4	über das Originalbild (links) wurde ein Salt-and-Pepper Rauschen darüber gelegt (rechts)	60
5.5	über das Originalbild (links) wurde ein Speckle Rauschen darüber gelegt (rechts)	60
5.6	Veränderung von Kontrast und Schärfe	60
5.7	Geschwisterpaar	61
5.8	Kein Geschwisterpaar	61
5.9	Gesicht (1) wurde auf 50 Jahre gealtert (2)	62
5.10	Beispiel für Auswahlbilder mit Rauschfilter	62
A.1	Kein Geschwisterpaar 1	65
A.2	Kein Geschwisterpaar 2	65
A.3	Kein Geschwisterpaar 3	65
A.4	Kein Geschwisterpaar 4	66
A.5	Geschwisterpaar 1	66
A.6	Geschwisterpaar 2	66
A.7	Geschwisterpaar 3	66
A.8	Geschwisterpaar 4	67
A.9	Geschwisterpaar 5	67
A.10	Geschwisterpaar 6	67
A.11	Gesichtsfiler 1 (Antwort: Bild 5)	68
A.12	Gesichtsfiler 2 (Antwort: Bild 1)	68
A.13	Gesichtsfiler 3 (Antwort: Bild 5)	69
A.14	Gesichtsfiler 4 (Antwort: Bild 6)	69

A.15 Gesichtsfiler 5 (Antwort: Bild 1)	70
A.16 Gesichtsfiler 6 (Antwort: Bild 1)	70
A.17 Gesichtsfiler 7 (Antwort: Bild 3)	71
A.18 Gesichtsfiler 8 (Antwort: Bild 9)	71
A.19 Gesichtsfiler 9 (Antwort: Bild 7)	72
A.20 Gesichtsfiler 10 (Antwort: Bild 3)	72
A.21 Face-Swapping 1 (Antwort: Bild 5)	73
A.22 Face-Swapping 2 (Antwort: Bild 3)	73
A.23 Face-Swapping 3 (Antwort: Bild 2)	74
A.24 Face-Swapping 4 (Antwort: Bild 6)	74
A.25 Face-Swapping 5 (Antwort: Bild 3)	75
A.26 Alterung 1 (Verjüngung auf 10 Jahre, Antwort: Bild 3)	75
A.27 Alterung 2 (Alterung auf 60 Jahre, Antwort: Bild 8)	76
A.28 Alterung 3 (Alterung auf 50 Jahre, Antwort: Bild 4)	76
A.29 Alterung 4 (Verjüngung auf 10 Jahre, Antwort: Bild 10)	77
A.30 Alterung 5 (Alterung auf 70 Jahre, Antwort: Bild 6)	77
A.31 Alterung 6 (Alterung auf 70 Jahre, Antwort: Bild 3)	78
A.32 Alterung 7 (Verjüngung auf 10 Jahre, Antwort: Bild 2)	78
A.33 Alterung 8 (Verjüngung auf 20 Jahre, Antwort: Bild 9)	79
A.34 Alterung 9 (Alterung auf 70 Jahre, Antwort: Bild 5)	79
A.35 Rauschen 1 (Salt-and-Pepper Rauschen (Salt/Pepper Verhältnis: 0.5, Menge: 0.2) Antwort: Bild 5)	80
A.36 Rauschen 2 (Speckle Rauschen (Intensität: 2.0) Antwort: Bild 5)	80
A.37 Kontrast 1 (Kontrastfaktor 0.05, Schärfefaktor 0.4, Antwort: Bild 5)	81
A.38 Kontrast 2 (Kontrastfaktor 0.1, Schärfefaktor 0.1, Antwort: Bild 5)	81
A.39 Auflösung 1 (Auflösung auf 20% reduziert, Antwort: Bild 4)	82
A.40 Auflösung 2 (Auflösung auf 30% reduziert, Antwort: Bild 1)	82

Tabellenverzeichnis

4.1 Accuracy der aktuellen Modelle zur Gesichtserkennung auf dem LFW-Datensatz	49
4.2 Accuracy der Super-Recognizer (1-10) und die drei KI's	53

1 Einleitung

1.1 Hintergrund

Die Integration von Super-Recognizern (SR) in die Polizeiarbeit zur Identifikation von Personen von Interesse kann die Ermittlungsarbeit verbessern. Es bedarf jedoch ständiger Forschung, um die passenden Personen mit überdurchschnittlichen Fähigkeiten den Arbeitsplätzen zuzuordnen, die von diesen Fähigkeiten profitieren. [24]

An der Hochschule Mittweida fand im Jahr 2023 ein Workshop zum Thema Super-Recognizer statt. Dazu wurden Mitarbeiter der Polizeidirektion Chemnitz eingeladen. Diese haupt- und nebenberuflichen Super-Recognizer wurden gebeten, an einer Studie teilzunehmen. Ziel dieser Studie war es, die Gesichtserkennungsfähigkeit der Teilnehmer zu testen. Zum einen durch eine Personenabfrage, bei der aus einer Reihe von Bildern eine zuvor gezeigte Zielperson wiedererkannt werden musste, zum anderen durch eine Doppelgängersuche, bei der Schauspieler und ihre Doppelgänger auseinandergehalten werden sollten. Die Ergebnisse wurden im Nachhinein mit denen von drei Künstlichen Intelligenzen (KI) verglichen. Die KI's konnten anfangs gut mit den Menschen Schritt halten, hatten jedoch einige Probleme, stark bearbeitete Bilder zu erkennen oder die Doppelgänger zu identifizieren.

1.2 Einsatzmöglichkeit in der Forensik

Durch den Einsatz von Super-Recognizern kann die Gesellschaft sicherer und gerechter werden. Beispielsweise können bei terroristischen Ereignissen Verdächtige schneller entdeckt, Identitätsbetrug aufgedeckt und unrechtmäßige Verurteilungen verhindert werden. [24]

Super-Recognizer können eine Lösung für sicherheitskritische Identitätsprüfungen sein, bei denen selbst trainierte Fachleute nicht besser als unbeteiligte Studenten abschneiden. Auch zu Überwachungszwecken können SR eingesetzt werden. Zum Beispiel wurden Personen mit überlegenen Gesichtsverarbeitungsfähigkeiten bei der London Metropolitan Police und bei der Polizei in Köln eingesetzt. Dort haben sie beispielsweise bei den Massenangriffen auf Frauen in der Kölner Silvesternacht 2015 geholfen. [24]

Die fest angestellten Super-Recognizer der Polizei Chemnitz arbeiten hauptsächlich von ihrem Büro aus. Sie erhalten Bild- und Videomaterial und versuchen, die darauf abgebildeten Personen zu identifizieren. Manchmal besteht bereits ein Verdacht, um wen es sich handeln könnte, und die SR gleichen die Aufnahmen dann mit Bildern aus sozialen Netzwerken, Einwohnermeldeämtern oder Ausweisfotos ab. [21]

1.3 Zielsetzung

Das Ziel dieser Arbeit ist es, ein Gesamtbild über Super-Recognizer zu erhalten. Einführend in die Thematik gibt es einen chronologischen Überblick über den Stand der Forschung, von der Entdeckung bis hin zu verschiedenen Testungen der Fähigkeit. Auf Basis dieses Wissens wird anschließend die zuvor erwähnte Studie der Hochschule Mittweida aus dem letzten Jahr ausgewertet. Die Ergebnisse der Teilnehmer werden mit denen von drei Künstlichen Intelligenzen verglichen.

Durch die gewonnenen Erkenntnisse wurde eine weitere Studie zur Messung der Fähigkeiten von Super-Recognizern entwickelt. Diese untersucht Fragestellungen, die in der vorherigen Studie nur teilweise umgesetzt wurden, fehlerhaft waren oder die für die Forschung interessant sind, aber bisher nicht getestet wurden. Der entwickelte Test soll im kommenden Workshop an der Hochschule Mittweida angewandt werden.

1.4 Hypothesen

Basierend auf der bisherigen Forschung und den Zielen dieser Studie wurden folgende Hypothesen formuliert:

- Super-Recognizer haben eine überdurchschnittliche Gesichtserkennungsfähigkeit selbst bei manipulierten Bildmaterial.
 - Die Super-Recognizer können Geschwisterpaare anhand ihrer Gesichter erkennen.
 - Super-Recognizer können Personen auch anhand von Bildern mit Gesichtsfiltren wiedererkennen.
 - Super-Recognizer können Gesichter, die durch künstliche Intelligenz erstellt wurden, erkennen.
 - Super-Recognizer können Menschen trotz alterungsbedingter Veränderungen wiedererkennen.
 - Super-Recognizer können selbst bei schlechten Bildqualitäten Zielpersonen identifizieren.

2 Grundlagen

2.1 Begriffsbestimmung

2.1.1 Super-Recognizer

Super-Recognizer (SR) sind grundsätzlich Menschen, die besonders versiert darin sind, Gesichtsidentitäten zu verarbeiten. Erstmals wurde der Begriff „Super-Recognizer“ von Russell et al. in dem Paper *Super-recognizers: People with extraordinary face recognition ability* eingeführt. Mit diesem Begriff wurden vier Personen bezeichnet, die herausragende Fähigkeiten in der Zuordnung neuer Gesichter, der Erkennung unbekannter Gesichter und der Identifizierung berühmter Personen anhand von Kinderbildern haben. [24]

Super-Recognizer werden bei der Polizei Chemnitz auch forensische Wiedererkenner genannt.

Es ist schwierig zu verstehen, wie SR, zum Beispiel die der Polizei, bei der Identifikation von Personen vorgehen. Die meisten sagen, sie handeln nach einem Bauchgefühl. Hauptsächlich ist jedoch das Gesicht von Bedeutung. Bei der Identifikation werden die einzelnen äußerlichen Merkmale betrachtet. Sie schätzen ein, welcher Ethnie und Nationalität die Person augenscheinlich zugehörig sein könnte. Dabei wird auf Merkmale wie Haarfarbe, Gesichtsbehaarung, Augenstellung, Augenbrauen, Nase, Mund, Leberflecke, Tätowierungen sowie besonders große oder kleine Gestalt geachtet. [21] Bestenfalls liegen den Ermittlern Videos der Personen vor, sodass z.B. auch der Gang mit in die Identifikation einbezogen werden kann. [21]

2.1.2 Prosopagnosie

Das Gegenteil der Super-Recognizer sind die Menschen mit Prosopagnosie. Diese Bezeichnung kommt aus dem Griechischen, wobei „Prosop“ als Gesicht übersetzt werden kann und „Agnosie“ als Nichterkennen. Umgangssprachlich wird die Prosopagnosie auch als Gesichtsbblindheit bezeichnet. [17]

Eine Schädigung des Gehirns in den beidseitigen oder rechten occipito-temporalen Regionen kann dazu führen, dass Menschen bekannte Gesichter nicht mehr erkennen können. Menschen mit erworbener Prosopagnosie haben eine Unfähigkeit, Gesichter ganzheitlich zu verarbeiten. [25]

3 Stand der Forschung

3.1 Tests zur Bestimmung von Super-Recognizer

3.1.1 CFMT +

Der Cambridge Face Memory Test (CFMT), entwickelt von Brad Duchaine und Ken Nakayama, bildet die Grundlage für den CFMT+. Es handelt sich um einen Test, der auf den Stärken des Benton Facial Recognition Test (BERT) und des Recognition Memory Test for Faces (RMF) aufbaut. Beide Tests wurden verwendet, um die Gesichtserkennungsfähigkeit zu testen, weisen jedoch einige Mängel auf. Beim BERT kann sogenanntes Feature-Matching angewandt werden, da das Zielgesicht und die Testgesichter gleichzeitig gezeigt werden. Beim RMF ist es nachteilig, dass die Bilder nicht-gesichtsbezogene Informationen wie Haarlinie und Kleidung enthalten, die dann auch zur Unterscheidung zwischen Ziel- und Ablenkungsbildern verwendet werden können. [9]

Der klassische CFMT besteht aus mehreren Abschnitten mit unterschiedlichen Schwierigkeitsstufen. Auch wird darauf geachtet, dass im Gegensatz zu den anderen beiden Tests nur gesichtsbezogene Informationen berücksichtigt werden. Für die Bilder wurden ausschließlich Männer in ihren 20ern oder frühen 30ern verwendet. Der Grund dafür ist, dass Frauen bei Frauengesichtern besser abschneiden, während Männerbilder sowohl von Frauen als auch von Männern äquivalent verarbeitet werden. Alle Bilder wurden in denselben Posen, Lichtverhältnissen und mit neutralem Gesichtsausdruck aufgenommen und danach so zugeschnitten, dass keine Haare oder Kleidung zu sehen sind. Sechs dieser Gesichter wurden als Zielgesichter ausgewählt, und es wurden zwölf Bilder von diesen verwendet. 46 weitere Personen wurden als Ablenkungsbilder genutzt. Einige dieser Personen kommen im Test mehrfach vor, sodass die Kontrollperson keine Vertrauensentscheidung treffen kann, nur weil ein Bild schon öfter gesehen wurde. Insgesamt besteht der Test aus vier Abschnitten. [9]

Im ersten gibt es eine kleine Übung in der gezeigt wird wie der Test funktionieren wird. Diese wird nicht mit Bildern echter Menschen durchgeführt, sondern mit Cartoon Bildern. Es gibt drei verschiedene Bilder von Bart Simpson, ein linkes 1/3 Profil, Frontal und ein rechtes 1/3 Profilbild. Dann werden zwei Bilder von anderen Cartoon Figuren zu einem Bild von Bart Simpson hinzugefügt und die Kontrollperson muss das Zielgesicht Bart Simpson mithilfe von einem Tastendruck auf 1,2 oder 3 auswählen. [9]

Im zweiten Abschnitt beginnt der wirkliche Test. Genauso wie mit dem Cartoon Bild werden jetzt drei Bilder in jeweils linkes 1/3 Profil, Frontalbild und rechtes 1/3 Profil von dem Zielgesicht gezeigt, siehe Abbildung 3.1. Diese kann sich die Kontrollperson 3 Sekunden lang anschauen. [9]



Abbildung 3.1: Zielgesicht [9]

Daraufhin wird z.B. das rechte $\frac{1}{3}$ Profilbild vom Zielgesicht ausgewählt und mit zwei $\frac{1}{3}$ Bilder von Ablenkungsgesichtern zu einem Testelement zusammengefügt, siehe Abbildung 3.2. Aus den drei Bildern des Testelements soll das vorher gezeigte Zielgesicht ausgewählt werden. In dem Testelement ist immer ein identisches Bild des Zielgesichts, wobei die Kontrollperson das richtige auswählen kann in dem das Bild wiedererkannt wird und nicht das Gesicht selbst. Für die restlichen fünf Zielgesichter wird das Verfahren wiederholt. Nach den einzelnen Durchgängen gibt es kein Feedback an die Kontrollperson. [9]



Abbildung 3.2: Testelement [9]

In Abschnitt drei werden neue Bilder der Zielgesichter eingeführt. Vor dem nächsten Testabschnitt werden den Kontrollpersonen Frontalaufnahmen jedes Zielgesichts für 20 Sekunden gezeigt. Es werden wieder Testelemente mit drei Bildern gezeigt und eins davon ist das Zielgesicht, siehe Abbildung 3.3. Dabei handelt es sich um neue Bilder mit veränderter Beleuchtung, Pose oder beides. Für die sechs Zielgesichter waren die veränderte Beleuchtung und Pose wiederum gleich. Beispiele dafür sind in der Abbildung 3.4 zu sehen. Anders als in dem Testabschnitt zuvor weiß die Kontrollperson nicht welches der Zielgesichter gesucht wird. [9] Der Abschnitt 4 führt wieder neue Bilder der Ziel-



Abbildung 3.3: Testelement 2 [9]



Abbildung 3.4: Bilder der Zielgesichter mit veränderter Beleuchtung und/oder Pose [9]

gesichter ein. Auf den Bildern liegen unterschiedliche Grade an Gauß'schem Rauschen (Abbildung 3.5), diese sind wiederum für die Gesichter innerhalb eines Testelements identisch (Abbildung 3.6). [9]

Das Hinzufügen des Rauschen hatte zwei Gründe. Einmal wollte man es den Kontrollpersonen nicht zu einfach machen die Zielgesichter zu erkennen nach dem sie diese in den vorherigen Abschnitten schon 13 Mal gesehen haben. Und zum anderen werden die Personen dazu gezwungen mehr auf die speziellen Mechanismen zurückzugreifen, von denen die Gesichtserkennung normalerweise abhängt. [9]



Abbildung 3.5: Bilder der Zielgesichter mit Rauschen [9]



Abbildung 3.6: Testelement 3[9]

Diese 72 Durchgänge des klassischen CFMT reichen aus, um Prosopagnosie, also geringe Gesichtserkennungsfähigkeiten, zu erkennen, aber nicht, um überdurchschnittliche Fähigkeiten zuverlässig zu detektieren. Deshalb gibt es den CFMT-Long, der einen weiteren Abschnitt zum klassischen Test hinzufügt. Dieser besteht aus 30 sehr schwierigen Durchgängen. Es wurden Ziel- und Ablenkungsbilder ausgewählt, die sich stark von den vorher trainierten Bildern unterscheiden. Beispielsweise sind Bilder enthalten, die nicht zugeschnitten sind und bei denen Haare sowie Kleidung zu sehen sind, sowie Bilder von lächelnden Gesichtern oder anderen Emotionen. Ein weiterer Faktor, der den Abschnitt schwieriger macht, sind wiederholte Ablenkungsbilder. Hier werden nur die 14 häufigsten Ablenkungsgesichter aus den ersten Abschnitten verwendet, um die Vertrautheit zu erhöhen. Als letzte Schwierigkeitserhöhung wurde den Bildern mehr visuelles Rauschen hinzugefügt. Beispiele für die Ziel- und Ablenkungsbilder im letzten Abschnitt sind in Abbildung 3.7 dargestellt. [27]



Abbildung 3.7: Beispiele Bilder im CFMT long [27]

3.1.2 CFPT

Cambridge Face Perception Test (CFPT) von Duchaine, Germine und Nakayama von 2007 [10] soll die Fähigkeit zur Wahrnehmung feiner Unterschiede zwischen Gesichtern ohne den Einfluss des Gedächtnisses testen. Dabei sollen die Probanden sechs Frontalansichten von Gesichtern nach der Ähnlichkeit zu einem Zielgesicht sortieren. Dieses ist einer Dreiviertelansicht gezeigt. Die sechs

Gesichter, die eine Ähnlichkeit zum Zielgesicht aufweisen, wurden durch Morphing erstellt. Das Zielgesichts wurde mit sechs anderen Gesichtern in unterschiedlichem Verhältnis vermischt. Dadurch variieren die sechs Gesichter in der Ähnlichkeit zum Zielgesicht. Jedes der sechs Gesichter wird durch die Kombination des Zielgesichts mit einem anderen Gesicht erstellt, dadurch unterscheiden sich ihre Erscheinungsbilder in verschiedenen Arten vom Zielgesicht. Das bringt die Probanden dazu, eine gesichtsspezifische Verarbeitung durchzuführen. Ein Beispiel der Bilder in dem Test ist in Abbildung 3.1.2 zu sehen. Der Test besteht aus acht aufrechte und acht umgekehrte Versuche, die den Inversionseffekt für die Gesichtswahrnehmung messen. [27]

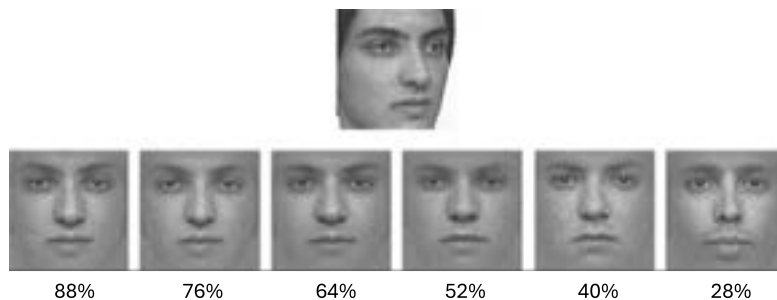


Abbildung 3.8: Beispiele Bilder aus CFPT Test, oben Zielgesicht, unten die zu sortierenden Gesichter [27]

3.1.3 BTWF

Der Before They Where Famous Test (BTWF) fokussiert sich auf den Einfluss des Alterungsprozess auf die Wiedererkennung von berühmten Personen. Im Laufe des Tests bekommen die Probanden 56 Bilder gezeigt und sollen diese benennen oder eindeutige Identitätsbeschreibungen geben. Die Fotografien der berühmten Personen sind aus der Zeit, als sie noch nicht berühmt waren. Vermehrt stammen die Bilder aus der Kindheit der jeweiligen Person und das Gesicht unterlag verschiedenen Alterungsbedingten Veränderungen. Bilder ähnlich zu den in Abbildung 3.9 werden den Probanden für 3 Sekunden lang gezeigt und danach haben sie unbegrenzt Bedenkzeit, um zu antworten welche berühmte Person zu sehen ist. Eine richtige Antwort muss die genaue Identität oder eine eindeutige Beschreibung der Person beinhalten, nur die Aussage das die gezeigte Person z.B. eine SchauspielerIn ist würde nicht ausreichen und ist inkorrekt. [27]



Abbildung 3.9: Beispiele Bilder aus BTWF Test [27]

Eine Schwäche von Tests die mit berühmten Personen arbeitet ist, dass die Performanz der Probanden davon abhängt, ob diese die gezeigten Personen vorher schon mal gesehen haben. Dazu gehört auch der BTWF Test, dementsprechend reicht dieser nicht aus, um allein die Gesichtserkennungsfähigkeit zu ermitteln. Zum Ausgleich sollte noch ein Test hinzugezogen werden der das Neuerlernen von Gesichtern voraussetzt, z.B. der CFMT. [27]

3.2 Literaturreview

In diesem Kapitel werden verschiedene Literaturen und Studien im Verlauf der Zeit dargestellt. Von der Ursprünglichen Frage der reinen Gesichtserkennung zu dem heutigen menschlichen Super-Recognizer.

3.2.1 Fortysomething: Recognizing faces at one's 25th reunion

Diese Veröffentlichung von Maggie Bruck, Patrick Cavanagh und Stephen J. Ceci aus dem Jahre 1991 bezieht sich auf die Gesichtserkennung über sehr lange Zeiträume. Zu dieser Zeit gab es bereits einige Studien zur Gesichtserkennung nach kurzen Zeiträumen, wie Stunden oder Tage, aber noch war wenig bekannt zu dem Verhalten der Gesichtserkennung nach langen Zeiträumen, wie nach Jahren. Nur die Autoren H.P. Bahrick, P.O. Bahrick, R.P. Wittlinger [3] erstellten eine Reihe an Studien zusammen, in denen das Erinnern an alte Klassenkameraden getestet wurde. Trotz des hohen theoretischen Wertes dieser Arbeit, veränderte sie nicht die Praxis. Grund dafür ist, dass nur eine reine Abfrage zum Namen der gezeigten Person stattfand und nicht die Zuordnung von einer Person nach großer Altersveränderung zu der jüngeren Version dieser Person. Somit ist es ein reiner Gedächtnisprozess und keine Gesichtserkennung. Demnach wird die Frage wie schwer es ist ein aktuelles Bild von einer Person die einem früher bekannt war, trotz der Veränderungen des Erscheinungsbild mit der Zeit wiederzuerkennen nicht beantwortet. [5]

In der Praxis können solche Studien zu der Gesichtserkennung nach Altersveränderungen empirische Beweise für diese Fähigkeit bringen. Ein Beispiel für die Nützlichkeit dessen, ist die Rolle von Zeugen vor Gericht. [5]

In der Vergangenheit haben Zeugenaussagen die einen Angeklagten nach Jahren wiedererkannt haben zu Verurteilungen geführt, wie in dem Fall von Klaus Barbie. Ein Foto von einem Mann aus den 1970er Jahren wurde von einem französischen Fernsehsender ausgestrahlt. Dieser Mann lebte in Bolivien unter dem Namen Klaus Altman. Dazu wurden die Zuschauer gefragt, ob sie diesen auf dem Foto erkannten, ohne das Hinweis auf seine Identität gegeben wurden. Daraufhin identifizierten viele Zuschauer aus der Region Lyon den Mann als Klaus Barbie, der auch unter dem Namen „Der Schlächter von Lyon“ bekannt war, was die französische Regierung dazu veranlasste, Barbie wegen „Verbrechen gegen die Menschlichkeit“ auszuliefern. [5]

Wenn die Identifizierung von Gesichtern nach langen Zeiträumen ausschließlich von Wahrnehmungsprozessen abhängt, die das Erkennen von Gesichtsmerkmalen beinhalten, die die Zeit überdauern (z.B. starke individuelle Merkmale wie ein Überbiss) oder auf der Anwendung von Alterungsveränderungen beruht (z.B. das Darstellen von Falten und zurückweichenden Haaransätzen) auf den Originalfotos, sollten Fremde und damalige Bekannte ähnlich gut, alte und aktuelle Fotografien abzugleichen. Außerdem sollten Manipulationen wie verschiedenen Ausrichtungen der Gesichter beide Probandengruppen gleichermaßen beeinflussen. Daher sollte es einfacher sein, Bilder abzugleichen, die die gleiche Gesichtsausrichtung haben, als Bilder, die sich in der Gesichtsausrichtung unterscheiden. [5]

Werden zu der reinen Wahrnehmung jedoch noch weitere Informationen hinzugezogen, wie das Wissen über das Aussehen von ehemaligen Bekannten aus mehr Perspektiven, als nur die eine auf dem zur Verfügung gestellten Bild, bringt die Vertrautheit spezifische Vorteile. Eventuell kann die Person, dann auf dem aktuellen Bild Merkmale entdecken die nicht auf dem alten Bild zu sehen sind, aber der Proband von früher kennt. Dementsprechend hätten ehemalige Bekannte wichtige Informationen die sie bei der Gesichtserkennung nach langen Zeiträumen besser abschneiden lassen als Fremde. Also

könnten sich sowohl Freunde als auch Fremde auf wahrnehmungsbezogene Strategien verlassen, einschließlich konstruktiver Strategien, jedoch haben nur die Freunde den Vorteil von außerbildlichen Informationen. [5]

In der Arbeit von Bruck et al. (1991) bestand die Studie aus zwei Bilderreihen. Die eine beinhaltete 40 Bilder jeweils 20 männlich und 20 weiblich die aus Highschool Jahrbüchern entnommen wurden. Die Bilder sind in schwarz weiß, zeige eine Kopf bis Schulter Ansicht und die Personen darauf sind entweder dreiviertel nach links oder nach rechts gedreht. Die andere Bildreihe besteht aus aktuellen Fotos, kurz vor dem 25 jährigen Klassentreffen, der selben Personen und zuzüglich 20 Klassenkameraden die bei den 40 zuvor nicht dabei waren und 20 Personen die nichts mit der Highschool zu tun hatten aber im selben Alter sind. [5]

Der Test wurde in Form eines Heftes von 8 Seiten dargeboten. Auf jeder Seite gab es 5 alte Jahrbuchfotos, die auf der linken Seite untereinander aufgelistet sind, und rechts 10 aktuelle Fotos. Die Probanden wurden aufgefordert jedes der 5 Bilder in der linken Spalte einem der 10 Bilder in den beiden rechten Spalten zuzuordnen. Auch wurde gebeten sich an den Namen des Klassenkameraden zu erinnern und eine Einschätzung zu geben wie gut Sie die Person in der Schule kannten. Als nächstes sollten Probanden aufschreiben vor wie vielen Jahren sie die Person zuletzt gesehen haben und zum Schluss angeben ob sie die rechten 10 Personen wiedererkennen ohne Bezug auf die linken 5 Bilder zu nehmen. [5]

Dieses Heft wurde an 110 ehemalige Klassenkameraden versendet, diese nicht an dem Klassentreffen teilgenommen hatten. [5]

Als zweite Gruppe an Probanden, im folgenden Kontrollgruppe genannt, wurden weitere 40 Personen ausgewählt, bei denen stark darauf geachtet wurde das sie keine Berührungspunkte mit der Schule, den Schülern oder überhaupt mit der Region haben. Diese sollten den Test unter ähnlichen Bedingungen durchführen. Trotz der Vorsichtsmaßnahmen sollten die Probanden angeben ob sie doch eine der Personen kennen, falls ja aus welchem Kontext. Zudem wurden sie gefragt wie sicher sie sich mit ihren Einschätzungen sind. [5]

Ergebnisse der Test wurden ausgeschlossen wenn Probanden angegeben haben einen Klassenkameraden aus der Abfrage in den letzten 17 Jahren gesehen hat. Der Schwellwert wurde gewählt, weil alles über 17 Jahre keinen großen Einfluss auf die Antworten gezeigt hat, jedoch darunter gab es erhebliche Unterschiede bei den Raten der korrekten Antworten. [5]

Die durchschnittliche Genauigkeit für die Highschool-Absolventen betrug 49% (mit Spannweiten von 33% bis 78%), während der durchschnittliche Genauigkeitswert der Kontrollgruppe 33% betrug (mit Spannweiten von 11% bis 72%). Die geplanten Vergleiche zeigten, wie vermutet, dass die Ausrichtung für die fremden Personen einen signifikanten Unterschied machten. Bei ähnlicher Ausrichtung wurde häufiger richtig zugeordnet (39%) als bei unterschiedlicher Ausrichtung (29%). Für die Teilnehmer der Highschool machte dies keinen Unterschied. [5]

Bei der Analyse des Vertrauens in die eigenen Angaben der Kontrollgruppe konnte festgestellt werden, dass eine Korrelation vorliegt. Die durchschnittliche Genauigkeit, wann Probanden die höchste Vertrauensbewertung angegeben haben betrug 63%. Im Gegensatz zu geringer Vertrauensbewertungen 24%. Trotz der großen Unterschiede kann nicht mehr als eine einfache Korrelation daraus geschlossen werden, weil sich innerhalb der einzelnen Vertrauensniveaus zu große Variationen bestehen. [5]

Bei der Analyse der Highschoolschüler-Daten gab es einen signifikanten Effekt der Vertrautheit. Die Genauigkeitsmittelwerte für Fotos, die als „sehr vertraut“, „etwas vertraut“ und „unbekannt“ bewertet wurden, betrugen 78%, 58% bzw. 37%. Der Genauigkeitswert für unbekannte Fotos bei diesen Highschool-Teilnehmern ist demnach fast gleich hoch wie der Gesamtdurchschnitt von 33% der Kontrollgruppe, was darauf hindeutet, dass die Highschool-Teilnehmer tatsächlich nicht mit diesen

Bildern vertraut waren. Es konnte keine signifikante Korrelation festgestellt werden zwischen der Fähigkeit vertraute Bilder zuzuordnen und der Fähigkeit unbekannte Fotos zuzuordnen. [5]

Die Analyse zeigt, dass die Vertrautheit mit der Ausrichtung interagiert. Gaben die Probanden an die Person zu kennen, waren die Genauigkeitswerte bei ähnlicher und unterschiedlicher Ausrichtung äquivalent. Kennt der Proband die Person auf dem Foto nicht, war der Genauigkeitswert bei ähnlicher Ausrichtung um 8% höher. [5]

Möglich wäre auch, dass der Vertrautheitseffekt nur durch eine Scheinkorrelation entsteht, weil es eine perzeptuelle Ähnlichkeit zwischen den alten und neuen Fotos gibt. Gegen die Hypothese spricht, dass es der Kontrollgruppe nicht signifikant einfacher gefallen ist die vertrauten Fotos der Highschoolgruppe zuzuordnen. [5]

Die Angabe des korrekten Namen der gezeigten Personen korrelierte mit der Vertrautheitsebene. Die Highschool-Teilnehmer lieferten für 21% der Klassenkameradenfotos einen Namen. Dieser war zu 11% der Zeit falsch, teilweise korrekt 18% der Zeit und vollständig korrekt 71% der Zeit. [5]

Als Fazit der Studie von Bruck et al. (1991) kann gesagt werden, dass wenn eine Vertrautheit vorliegt, demzufolge mehr Informationen über die Person über das eine gegebene Highschoolfoto hinaus vorhanden sind, die Bilder besser zugeordnet wurden. Trotzdem konnten Personen ohne früheren Kontakt, alleine durch die Wahrnehmungsinformationen die Bilder besser zuordnen, als eine reine Zufallswahl. [5]

Im Vergleich zu den ehemaligen Highschool-Schülern zeigte die Kontrollgruppe, einen signifikanten Ausrichtungseffekt. Die Probanden neigten eher dazu, Fotos korrekt zuzuordnen, die die gleiche Gesichtsausrichtung teilten, als Paare, die unterschiedliche Ausrichtungen hatten. Darüber hinaus schnitten WHS-Teilnehmer, die niedrige Vertrautheitsgrade angaben, auch besser ab, wenn Fotos ähnliche Ausrichtungen teilten. Beide Ergebnisse unterstützen die Hypothese, dass perzeptive Hinweise eine größere Rolle bei der Erkennung von unbekannten Gesichtern spielen, als bei vertrauten Gesichtern, und dass extrapiktoriale Informationen für die Erkennung vertrauter Fotografien verwendet werden. [5]

Zusätzlich kam es zu einer höheren Genauigkeit, wenn die Probanden Fotos ihres eigenen Geschlechts zuordneten, im Gegensatz zu Fotos des anderen Geschlechts. [5]

Bezüglich der Anwendbarkeit der Ergebnisse aus dieser Studie auf die vorher beschriebenen juristischen Beispiele gibt es Vorbehalte. Zum einen sind die Bedingungen hier sehr unterschiedlich zu denen, die die meisten Opfer/Zeugen veranlassen, vor Gericht auszusagen. Diese haben in der Regel extreme emotionale Erfahrungen gemacht und nicht so viel Kontakt mit ihrem Angreifer wie die Klassenkameraden in der Studie. [5]

Es macht wahrscheinlich auch einen Unterschied, ob man eine Person aus der Vergangenheit nur auf einem Foto sieht oder in einer Gegenüberstellung präsentiert wird. Dies liegt daran, dass statische Fotografien potenziell wertvolle Hinweise (linguistische, akustische, kinästhetische, affektive) eliminieren. Trotz dieser Einschränkung scheinen die Daten jedoch direkt relevant für andere fotografische Kontexte zu sein, wie etwa die Identifizierung vermisster Kinder anhand eines Vergleichs von alten und aktuellen Fotografien. [5]

Aus der Studie ergeben sich zwei Faktoren, die berücksichtigt werden sollten. Einerseits der Grad der physischen Veränderung zwischen früherem und aktuellem Erscheinungsbild. Es gibt keine feste Regel für solche Veränderungen, aber einige Gesichter sind schwerer zuzuordnen als andere, z. B. durch starken Haarausfall, Gewichtszunahme oder Ähnliches. Andererseits gibt es enorme Unterschiede zwischen den Probanden. Einige Personen aus der Kontrollgruppe schnitten besser ab als einige der Highschool-Probanden. Welche Faktoren die Unterschiede in der Erkennungsfähigkeit erklären, muss noch weiter untersucht werden. [5]

3.2.2 Super-recognizers: People with extraordinary face recognition ability

Russell, Duchaine und Nakayama testeten vier Personen, die behaupteten, eine signifikant bessere Gesichtserkennungsfähigkeit als der Durchschnitt zu haben. Ziel der vorliegenden Studie war es, diese Behauptungen durch objektive Tests zu bewerten, insbesondere festzustellen, ob diese Personen tatsächlich wesentlich besser als der Durchschnitt in der Lage sind, Gesichter zu erkennen. [27] Die Autoren wurden nach ihrer Arbeit zur Entwicklungsprosopagnosie von mehreren Personen kontaktiert, die glauben, das Gegenteil von Prosopagnosie zu haben, nämlich außergewöhnlich gut im Erkennen von Gesichtern zu sein. [27]

Die vier Personen beschreiben ihre Fähigkeiten im Erkennen von Gesichtern in klaren Worten. Es spielt für sie keine Rolle, wie viele Jahre sie eine Person nicht gesehen haben, sobald sie ein Gesicht einmal gesehen haben, merken sie es sich. Alle erzählten von Situationen, in denen sie Fremde nach Jahren wiedererkannt haben, sogar nach großen äußerlichen Veränderungen, wie zum Beispiel vom Kind zum Erwachsenen. Außerdem berichten sie, dass sie in der Lage sind, Schauspieler in verschiedenen Rollen zu erkennen. Viele Menschen mit Entwicklungsprosopagnosie hingegen sagen, dass sie wenig Interesse an Film oder Fernsehen haben, weil sie die verschiedenen Charaktere nicht unterscheiden können, was es ihnen schwer macht, der Handlung zu folgen. [27]

Die Personen mit hoher Erkennungsfähigkeit berichten, dass sie ihr soziales Verhalten verändert haben, weil sie häufig andere Menschen erkennen, die sie jedoch nicht wiedererkennen. Es kam häufiger zu Situationen, in denen sich andere Menschen unangenehm oder sogar verfolgt fühlen, weil sie von Fremden erkannt und angesprochen werden. [27]

Die Studie bestand aus zwei Teilen, wobei der erste die Gesichtserkennung behandelte. Ziel war es, objektiv festzustellen, ob diese Personen tatsächlich wesentlich besser als der Durchschnitt darin sind. Bis zu diesem Zeitpunkt existierende klinische Tests zur Gesichtserkennung befassten sich nur mit dem Unterschied zwischen normalen und subnormalen Fähigkeiten. Diese Tests waren jedoch nicht anspruchsvoll genug, um normale von überdurchschnittlichen Fähigkeiten zu unterscheiden. Daher wurden neue, anspruchsvollere Tests zur Gesichtserkennung entwickelt. [27] Es wurde der Before They Were Famous (BTWF)-Test angewandt. Bei diesem Gesichtserkennungstest sehen sich die Probanden 56 Fotografien von berühmten Personen an und sollen diese benennen oder eindeutige Identifikationsbeschreibungen geben. [27]

Weiteres zu dem Test in ist Kapitel 3.1.3 beschrieben. Dieser Test allein reicht nicht aus, weil die Ergebnisse stark von der vorherigen Vertrautheit mit den Personen beeinflusst werden. Um diese Einschränkung zu überwinden, wird noch ein zweiter Test verwendet. Bei diesem werden die Teilnehmer darauf trainiert, neue Gesichter zu erkennen, sodass die Exposition gegenüber den Zielgesichtern zwischen den Teilnehmern ausgeglichen wurde. [27]

Als zweiter Test zur Gesichtserkennung wird die Langform des Cambridge Face Memory Test (CFMT) verwendet. Dieser Test ist eine angepasste Version des CFMT von Duchaine und Nakayama [9]. Der CFMT beinhaltet das Erlernen der Erkennung von sechs unbekannten männlichen Gesichtern aus drei verschiedenen Ansichten und anschließend die Prüfung der Erkennung dieser Gesichter in einer Aufgabe mit drei Alternativen. [27]

Näher erklärt wird der Test in Kapitel 3.1.1.

Drei der Probanden wurden mit dem BTWF-Test und der CFMT-Langform getestet. Die vierte Person wurde vor der Erstellung der Langform getestet, daher liegen von ihr nur Ergebnisse des BTWF-Tests und der CFMT-Kurzform vor. Da es sich bei der CFMT-Langform um einen Gedächtnistest handelt, konnte sie nicht nach der Kurzform durchgeführt werden. Der CFMT-Langform und der BTWF-Test wurden 25 Kontrollpersonen verabreicht, deren Alter sich nicht signifikant von dem der experimentellen Probanden unterschied. [24]

Die Leistung beim BTWF-Test korrelierte sowohl mit der Kurzform des CFMT als auch mit der Langform des CFMT. Dies deutet darauf hin, dass beide Tests die gleichen zugrunde liegenden Prozesse messen. Auf der Kurzform des CFMT erzielten alle vier Versuchspersonen Höchstpunktzahlen. Auf beiden Formen des CFMT schnitt jede Versuchsperson besser ab als jede Kontrollperson. Beim BTWF-Test schnitt jede Versuchsperson besser ab als alle bis auf eine Kontrollperson, die die gleiche Punktzahl wie eine Versuchsperson erzielte. Die Versuchspersonen schnitten signifikant besser ab als die Kontrollpersonen, sowohl bei der Langform des CFMT als auch beim BTWF-Test. Bei beiden Tests zur Gesichtserkennung war die Leistung der Versuchspersonen deutlich höher als die der Kontrollpersonen. Dies liefert objektive Belege für ihre subjektive Erfahrung, signifikant besser als der Durchschnitt in der Gesichtserkennung zu sein. Da ihre Erkennungsfähigkeit deutlich über der normaler Menschen liegt, geben die Autoren ihnen den Namen „Super-Recognizer“. [27]

Der Zweite Teil der Studie befasst sich mit der Gesichtswahrnehmung. Nachdem eine Gruppe von Personen identifiziert wurde, die außergewöhnlich gute Gesichtserkennungsfähigkeiten besitzen, soll auch getestet werden, ob diese noch andere außergewöhnliche Fähigkeiten haben. Eng verbunden mit der Gesichtserkennung ist die Gesichtswahrnehmung, also die Fähigkeit, zwischen gleichzeitig präsentierten Gesichtern zu unterscheiden. In diesem Teil der Studie wird untersucht, ob die außergewöhnlichen Fähigkeiten auf eine überlegene Erinnerung oder auf eine überlegene Wahrnehmung zurückzuführen sind. Es wurden auch Menschen mit angeborener Gesichtsblindheit getestet, die das untere Ende der Gesichtserkennungsfähigkeiten repräsentieren. [27]

Ein weiteres Ziel der Studie ist es, herauszufinden, wie groß der Inversionseffekt bei Menschen mit überdurchschnittlicher Gesichtserkennung ist. Dies beinhaltet zum Beispiel die Wahrnehmung und Erkennung von Bildern, wenn sie umgekehrt (kopfüber) sind. Zusätzlich zu den vier Super-Recognizern wurden 26 neue Kontrollpersonen sowie 26 Personen mit angeborener Gesichtsblindheit (Prosopagnosie) getestet. Alle Personen mit Prosopagnosie gaben an, eine sehr schlechte Gesichtserkennungsfähigkeit zu haben und schnitten in einem Test zur Erkennung berühmter Gesichter, einem Test für die Erkennung alter/neuer Gesichter und dem CFMT sehr niedrig ab. [27]

Mit dem CFPT-Test soll die Fähigkeit zur Wahrnehmung feiner Unterschiede zwischen Gesichtern ohne Einfluss des Gedächtnisses getestet werden. Dabei sollen die Probanden eine Gruppe von sechs Frontalansichten von Gesichtern nach Ähnlichkeit zu einem Zielgesicht sortieren. [27]

Genauer ist der Test in Kapitel 3.1.2 erklärt.

In den Aufgaben schnitten die Super-Recognizer besser ab als die Kontrollpersonen, welche wiederum besser abschnitten als die Prosopagnosiker. Die Unterschiede zwischen den Subjektgruppen für aufrechte Gesichter waren viel größer als für umgekehrte Gesichter. [27]

Diese Gruppenunterschiede waren signifikant, was darauf hinweist, dass Personen mit besserer Gesichtserkennungsleistung auch einen größeren Gesichtsinversionseffekt haben. Bei aufrechten Gesichtern waren die Ergebnisse von Kontrollpersonen und Super-Recognizern im Vergleich sehr ähnlich zu denen der Kontrollpersonen im Vergleich mit Entwicklungsprosopagnosikern. Die Ergebnisse des CFPT bestätigen die Annahme, dass die Fähigkeiten der Super-Recognizer nicht auf das Gesichtsgedächtnis beschränkt sind, sondern sich auch auf die Gesichtswahrnehmung beziehen. Es liegt demnach nahe, dass die Fähigkeiten zur Gesichtserkennung und Gesichtswahrnehmung miteinander verbunden sind. Für weitere Tests, um dies zu überprüfen, wurden 23 neue Kontrollpersonen sowohl dem CFMT-Langform als auch dem CFPT unterzogen. Die Leistung im CFMT-Langform korrelierte mit dem CFPT bei aufrechten Gesichtern und mit dem CFPT bei umgekehrten Gesichtern. Diese Ergebnisse unterstützen die Annahme, dass die Fähigkeit zur (aufrechten) Gesichtserkennung mit der Fähigkeit zur aufrechten Gesichtswahrnehmung verbunden ist, nicht jedoch mit der umgekehrten Gesichtswahrnehmung. [27]

Alle vier Teilnehmer schnitten bei zwei verschiedenen Tests zur Gesichtserkennung deutlich über

dem Durchschnitt ab. Sowohl bei einem Test zur Erkennung bekannter Gesichter als auch bei einem Test, der das Erlernen unbekannter Gesichter erfordert, waren ihre Ergebnisse besser als die der 25 Kontrollpersonen. Die daraus folgenden Ergebnisse unterstützen ihre Behauptung, eine außergewöhnliche Gesichtserkennungsfähigkeit zu besitzen. Dies liefert erste Hinweise auf die Existenz von Menschen mit einer Gesichtserkennungsfähigkeit, die weit über dem Durchschnitt liegt: sogenannte Super-Recognizer. Durch die Existenz von Super-Recognizern entstehen neue Forschungsfragen. Sind diese Personen eine eigenständige Gruppe? Wahrscheinlich ist, dass Super-Recognizer das obere Ende einer breiten Verteilung von Gesichtserkennungsfähigkeiten darstellen. [27]

Sowohl bei der Gesichtserkennung als auch bei der Gesichtswahrnehmung sind Super-Recognizer etwa so gut, wie viele Menschen mit entwicklungsbedingter Prosopagnosie schlecht sind. Dies legt nahe, dass viele Fälle von entwicklungsbedingter Prosopagnosie eher das untere Ende des normalen Spektrums der Gesichtserkennungsfähigkeiten darstellen könnten, anstatt eine qualitativ unterschiedliche Art der Gesichtsverarbeitung zu sein. [27]

Ob eine oder mehrere Verteilungen der Gesichtserkennungsfähigkeiten zugrunde liegen, die Existenz von Super-Recognizern deutet darauf hin, dass dieser Bereich größer ist als bisher angenommen. Am einen Ende dieses Spektrums befinden sich Personen mit entwicklungsbedingter Prosopagnosie, von denen einige sogar Schwierigkeiten haben, Mitglieder ihrer eigenen Familie zu erkennen. Am anderen Ende befinden sich Super-Recognizern, die häufig völlig fremde Personen außerhalb des Kontexts nach vielen Jahren wiedererkennen. Die Entdeckung, dass die Bandbreite der Gesichtserkennungsfähigkeiten breiter und nicht enger ist als angenommen, ist sowohl in angewandten als auch in theoretischen Kontexten relevant. [27]

Im Alltag wäre es sinnvoll, Menschen mit durchschnittlichen bis überdurchschnittlichen Gesichtserkennungsfähigkeiten in Berufsgruppen einzusetzen, in denen diese Fähigkeit von Vorteil ist. Beispielsweise könnte Sicherheitspersonal, das Ausweissfotos abgleicht, von besseren Gesichtserkennungsfähigkeiten profitieren. Wahrscheinlicher ist es jedoch, dass viele Menschen mit schlechter Gesichtserkennungsfähigkeit solche Jobs ausüben, weil dieser Aspekt oft unbeachtet bleibt. Daher sollte die Gesichtserkennungsfähigkeit als Teil des Auswahlprozesses für solche Berufe bewertet werden. Auch im Strafrechtssystem müsste auf diese Fähigkeit bei Augenzeugen geachtet werden. Die Genauigkeit der Augenzeugenidentifikation korreliert signifikant mit der Leistung bei standardisierten Gesichtserkennungstests, die Fähigkeit von Zeugen zur Gesichtserkennung könnte nach einem Verbrechen als ein Faktor betrachtet werden, der die Genauigkeit der nachträglichen Augenzeugenidentifikation beeinflusst. [27]

Die Ergebnisse der Studie zeigen zudem, dass es eine Verbindung zwischen der Fähigkeit zur Gesichtserkennung und der Fähigkeit zur Gesichtswahrnehmung gibt. Die Variation in der Gesichtserkennungsfähigkeit ist wahrscheinlich nicht ausschließlich eine Funktion des Gedächtnisses, sondern auch der Wahrnehmung. Erworbene und entwicklungsbedingte Prosopagnosiker haben oft einen geringen oder nicht vorhandenen Gesichtsinversionseffekt. Die vorliegende Arbeit repliziert diese Erkenntnis und erweitert sie, indem gezeigt wird, dass Super-Recognizer mit außergewöhnlich guter Gesichtserkennung größere Gesichtsinversionseffekte haben als normale Kontrollpersonen. Dies deutet darauf hin, dass das Fehlen von Inversionseffekten bei Prosopagnosikern möglicherweise nicht auf qualitativ unterschiedliche Verarbeitung hinweist, wie von de Gelder und Rouw (2000) sowie von Duchaine et al. (2006) argumentiert wurde. Die Größe des Gesichtsinversionseffekts ist mit der Gesichtserkennungsfähigkeit über die gesamte Bandbreite dieser Fähigkeit verbunden. Die kontinuierliche Verteilung der Inversionsabnahmen unterstützt zudem die Vorstellung quantitativer Unterschiede zwischen prosopagnosischer und normaler Gesichtsverarbeitung. [27]

Die Entdeckung von Menschen mit überdurchschnittlicher Gesichtserkennungsfähigkeit, den Super-Recognizern, zeigt, dass es nicht nur Menschen mit stark eingeschränkter Fähigkeit (Prosopagnosie)

gibt. Dies unterstützt die Annahme, dass die entwicklungsbedingte Prosopagnosie eine schwach ausgeprägte Version der normalen Gesichtsverarbeitung darstellen könnte, anstatt eine qualitativ beeinträchtigte Version zu sein. [27]

3.2.3 Human face recognition ability is specific and highly heritable

Die allgemeine Intelligenz hat in einigen Studien zur kognitiven Funktion große Aufmerksamkeit erhalten. Dabei wurde festgestellt, dass sie eine hohe Vererbbarkeit aufweist. Die Vererbbarkeit kognitiver Merkmale nimmt jedoch typischerweise ab, wenn ihre Beziehung zur allgemeinen Intelligenz schwächer wird. Nur wenige spezifische Fähigkeiten mit starker Vererbbarkeit wurden identifiziert. [40]

Vor der Veröffentlichung von Wilmer et al. (2010) wurde die Vererbbarkeit der Gesichtserkennungsfähigkeit bisher nicht untersucht. Mehrere Befunde deuten jedoch darauf hin, dass diese Fähigkeit eine genetische Grundlage haben könnte. Darüber hinaus ergab eine Studie mit Zwillingen, dass der ventrale Sehrindenkortex, eine Region im Gehirn mit auf Gesichter spezialisierten Arealen, bei eineiigen Zwillingen ähnlicher war als bei zweieiigen Zwillingen. [40]

Die Autoren verwendeten auch eine Studie mit Zwillingen, um die relativen Beeinflussungen von Genen und Umwelt auf die Gesichtserkennungsfähigkeit zu bestimmen. Dazu führten sie auch eine Analyse durch, um die Gesichtserkennungsfähigkeit von anderen Gedächtnisfähigkeiten zu trennen. [40]

Die Ergebnisse der Studie wurden hauptsächlich durch den CFMT-Test erzeugt. Der CFMT isoliert gesichtsspezifische Mechanismen, indem er Gesichter ohne Haare oder andere Hinweise auf die Personenerkennung präsentiert. [40]

Für die Zwillingstudie wurde der CFMT mit 164 eineiigen und 125 gleichgeschlechtlichen zweieiigen Zwillingspaaren durchgeführt. Beide Arten teilen sich die familiäre Umgebung, die eineiigen Zwillinge haben jedoch doppelt so viele geteilte Gene. Daher unterstützt eine größere Korrelation bei eineiigen Zwillingen einen genetischen Effekt. Diese Analysen liefern Hinweise darauf, dass die familiäre Ähnlichkeit in der Gesichtserkennungsfähigkeit hauptsächlich auf genetische Faktoren zurückzuführen ist. [40]

Untersucht wurde auch, ob Personen mit guten Leistungen im CFMT-Test ebenfalls gute Leistungen bei der Erinnerung an nicht-gesichtsspezifische Informationen aufweisen. Eine große Gruppe von Nicht-Zwillingen und eine Untergruppe der Zwillinge absolvierten den CFMT sowie zwei Kontrolltests zur Erinnerung an nicht-gesichtsspezifische Informationen: den „verbal paired-associates memory test“ (VPAM) und einen neu entwickelten Test, den „abstract art memory test“ (AAM). Der VPAM diente als nicht-visueller Kontrollerkennungstest für den CFMT. Er testet die Erinnerung an sprachliche Informationen (Wortpaare) und minimiert den Beitrag nicht-linguistischer visueller Prozesse, indem abstrakte Wörter verwendet werden, die schwer in visuelle Bilder umzusetzen sind. Der AAM diente als nicht-gesichtsspezifischer visueller Kontrollerkennungstest für den CFMT. Wie der CFMT verwendet auch der AAM schwer verbalisierbare Reize, um den Beitrag linguistischer Prozesse zur visuellen Erkennung zu minimieren. Tatsächlich zeigten VPAM und AAM beide in der großen Nicht-Zwilling-Stichprobe eine hohe interne Konsistenz, jedoch nur geringe Korrelationen mit dem CFMT. Die geringen Korrelationen zwischen VPAM und CFMT-Leistung bei Zwillingen- und Nicht-Zwillingsteilnehmern zeigen, dass die beobachtete Erblichkeit des CFMT nicht auf Faktoren wie Motivation, Aufmerksamkeit, Computerkenntnisse, allgemeine Intelligenz oder allgemeines Gedächtnis zurückzuführen ist. [40]

Da der AAM dem CFMT sehr ähnlich ist, aber abstrakte Kunst anstelle von Gesichtern verwendet,

deuten die bescheidenen Korrelationen zwischen dem CFMT und dem AAM darauf hin, dass allgemeine visuelle Prozesse nur einen begrenzten Beitrag zur CFMT-Leistung leisten. Diese geringen Korrelationen legen nahe, dass sowohl die Gesichtserkennungsfähigkeit selbst als auch ihre genetische Grundlage hauptsächlich von gesichtsspezifischen Mechanismen abhängen. [40]

Die Zwillingsbefunde deuten darauf hin, dass genetische Unterschiede für den Großteil der stabilen Variation in der Gesichtserkennungsfähigkeit bei gesunden Erwachsenen verantwortlich sein könnten. Die Untersuchung in dieser Studie zeigt zudem, dass die Gesichtserkennungsfähigkeit wenig mit anderen visuellen und Gedächtnisfähigkeiten überlappt, was darauf hindeutet, dass sowohl die Gesichtserkennungsfähigkeit selbst als auch ihre genetische Grundlage weitgehend gesichtsspezifisch sind. Die vorliegenden Ergebnisse identifizieren daher ein seltenes Phänomen: eine hochspezifische kognitive Fähigkeit, die stark erblich ist. Die kürzlich dokumentierte Existenz von Familien mit mehreren Mitgliedern, die an Prosopagnosie leiden, hat gezeigt, dass spezifische Gesichtserkennungsdefizite in Familien auftreten können. Allerdings könnte eine solche familiäre Häufung sowohl auf familiäre Umwelt- als auch auf familiäre Gene zurückzuführen sein. Die vorliegenden Ergebnisse zeigen, dass die familiäre Häufung der Gesichtserkennungsfähigkeit eine starke genetische Grundlage hat. Die hohe Erblichkeit der Gesichtserkennung weist darauf hin, dass Studien zur Kopplung und Positionsklonierung, die die genetische Grundlage der Gesichtserkennung untersuchen, erfolgreich sein könnten. Solche Studien können die Assoziationsstudien in Familien mit mehreren Prosopagnosiepatienten ergänzen. [40]

Obwohl die Daten der Studie darauf hinweisen, dass die familiäre Ähnlichkeit in der Gesichtserkennungsfähigkeit hauptsächlich auf genetischen Faktoren beruht, sind sie auch mit einem bescheidenen Beitrag der nicht-familiären Umgebung zur CFMT-Performance vereinbar. Tatsächlich zeigt die vorherige Literatur, dass die Gesichtsverarbeitung auf Umwelteinflüsse empfindlich reagiert. Die Gesichtswahrnehmung zeigt Anpassungseffekte, eine Verengung der Wahrnehmung tritt mit der Entwicklung auf, und die Erkennung von Gesichtern seltener beobachteter Rassen ist berüchtigt schlecht. Extremere Umweltvariablen wie visuelle Deprivation in den Monaten nach der Geburt haben ebenfalls Auswirkungen auf die Gesichtsverarbeitung. Allgemeiner gesagt deutet die Entdeckung, dass die Gesichtserkennung stark erblich ist, darauf hin, dass die kognitive Neurowissenschaft dazu beitragen kann, Verhaltensgenetik-Untersuchungen sinnvoll zu lenken. Die Aufteilung des Gehirns in seine Komponentensysteme ist eine grundlegende Aufgabe in der kognitiven Neurowissenschaft, und die Untersuchung dieser Systeme könnte ein effektiver Weg sein, um Gene und Verhalten zu verbinden. [40]

Für den VPAM-Test memorisierten die Teilnehmer 25 Zielwortpaare und wurden dann auf die Erkennung dieser Wortpaare getestet. Dabei mussten sie in jeder Versuchsrunde auswählen, welches der vier Wörter mit einem gegebenen Wort gepaart war. Für den AAM-Test memorisierten die Teilnehmer 50 Zielbilder abstrakter Kunst und wurden dann auf die Erkennung dieser Bilder getestet, indem sie in jeder Versuchsrunde auswählen mussten, welches der drei Bilder sie zuvor gesehen hatten. Diese Forschung wurde von den Institutional Review Boards der Harvard University und der Macquarie University genehmigt und entsprach der Deklaration von Helsinki. [40]

3.2.4 Investigating Predictors of Superior Face Recognition Ability in Police SuperRecognizers

Es gibt erhebliche individuelle Unterschiede in der Fähigkeit zur Gesichtserkennung. Diese Unterschiede, die hauptsächlich vererbt sind (genauer im Kapitel zu 3.2.3), korrelieren mit der Genauigkeit der Identifizierung durch Augenzeugen, der Fähigkeit zur gleichzeitigen Gesichtsanpassung, der

Persönlichkeit und der Neigung, Gesichter ganzheitlich zu verarbeiten. Die Forschung zur Untersuchung außergewöhnlicher Fähigkeiten hat sich hauptsächlich auf Gesichtsblindheit (Prosopagnosie) konzentriert, insbesondere wenn diese mit normaler Sehschärfe und Objekterkennungsfähigkeit im üblichen Bereich einhergeht. Erworbene Prosopagnosie ist eine Folge von Hirnschäden, während entwicklungsbedingte Prosopagnosie, die oft in der Kindheit identifiziert wird, mit keinem bekannten Schaden verbunden ist. [8]

Einige Menschen besitzen jedoch eine außergewöhnlich gute Fähigkeit zur Gesichtsverarbeitung. Russell et al. (2009) stellten fest, dass vier sich selbst als Super-Recognizer bezeichnete Personen bei Tests wie der Langform des CFMT, dem CFPT und einem Before They Were Famous-Gesichtstest deutlich besser abschnitten als Kontrollgruppen. Die Autoren gehen davon aus, dass Super-Recognizer das obere Ende der Fähigkeit zur Gesichtserkennung repräsentieren, während Entwicklungsprosopagnosiker im Gegensatz dazu das untere Ende darstellen. [8]

Eine solche überdurchschnittliche Fähigkeit zur Gesichtserkennung hat Auswirkungen auf die Strafverfolgung. Beim Metropolitan Police Service (MPS) gibt es eine kleine Gruppe von Mitarbeitern, die die Mehrheit der Verdächtigenidentifikationen aus CCTV-Aufnahmen in London vornehmen. Einige dieser Mitarbeiter wurden nach mehrfachen Identifizierungen auf der MPS-Website „Caught on Camera“ entdeckt. Diese Personen sind nun mit der Untersuchung schwerer Verbrechen in ganz London sowie weniger schwerwiegenderer lokaler Verbrechen betraut. Die Identifizierungen stellen den ersten Schritt in einer polizeilichen Untersuchung dar; mehr als 70% der durch CCTV identifizierten Verdächtigen gestehen im Verhör, wenn sie mit Bildern konfrontiert werden. Trotzdem gelangen nicht alle Fälle vor Gericht, da nach dem ersten Schritt in die Untersuchung oft alternative Beweise fehlen, sodass die Schuld nicht immer nachgewiesen werden kann. Dennoch wurden von April 2013 bis Dezember 2015 circa 13.000 Personen identifiziert, darunter 9.000 von Polizei-Identifizierern des MPS, was die Verurteilungsraten in Fällen mit CCTV-Beweisen in London erheblich steigerte. [8]

Die meisten dieser Identifizierer des MPS sind als Streifenpolizisten unterwegs, und ihre Identifizierungen werden hauptsächlich durch Kenntnisse über lokal bekannte Verdächtige beeinflusst. Es gibt Theorien, dass es qualitativ unterschiedliche Verarbeitungswege für vertraute und unbekannte Gesichter gibt. Bei vertrauten Gesichtern werden im Gedächtnis gespeicherte Bilder verwendet, die unabhängig vom Blickwinkel und Gesichtsausdruck zur Erkennung genutzt werden. Dadurch ist die Genauigkeit selbst bei schlechter Bildqualität hoch. Genau solche Bilder werden bei der Erkennung vertrauter Gesichter verwendet, um Deckeneffekte zu reduzieren. Bei solcher Bildqualität sind die Erkennungsleistungen bei bewegten Bildern höher, wahrscheinlich aufgrund der zusätzlichen Informationen durch mehrere verfügbare Frames oder auch sichtbarer Bewegungsmuster. [8]

Es gibt jedoch auch Identifizierungen von Verdächtigen, die der Identifizierer persönlich nicht kennt und nur an zuvor betrachteten Bildern erkennt. Die Erkennung von vertrauten Gesichtern wird von der Verwendung interner Merkmale (z. B. Augen und Mund) dominiert, während bei unbekannten Gesichtern die externen Merkmale (z. B. Frisur und Gesichtsform) wichtiger sind. Solche Faktoren können durch Umgebungsbedingungen oder äußerliche Veränderungen anders interpretiert werden und sind somit fehleranfälliger. Dies kann dazu führen, dass Bilder von zwei verschiedenen Personen sehr ähnlich erscheinen oder zwei Bilder derselben Person sehr unterschiedlich wirken. Die Leistung beim gleichzeitigen Abgleich unbekannter Gesichter kann auch mit unbegrenzter Betrachtungszeit, hochwertigen Bildern und persönlich anwesenden Zielpersonen unzuverlässig bleiben. [8]

Trotz dieser Probleme werden Super-Recognizer als Identifizierer wegen ihrer Fähigkeiten zur Verarbeitung unbekannter Gesichter eingesetzt. Zu den Aufgaben gehören in Menschenmengen Verdächtige anhand von eingepprägten Fotografien identifizieren, der Abgleich von Verdächtigen bei verschiedenen Verbrechen mit ähnlichen Merkmalen sowie das Überprüfen von Aufnahmen, um Personen von Interesse zu finden. Einige dieser Identifizierer der Polizei wurden einer Super-Recognizer-

Einheit zugeteilt, um diese Aufgaben in Vollzeit durchzuführen. [8]

Die aktuelle Forschung zur Zeit der Arbeit von Davis et al. (2016) nutzte vier Gesichtsverarbeitungstests, um herauszufinden, ob die Leistung der hauptsächlich aus der Gemeinde stammenden Streifenpolizisten, die als Identifizierer des MPS arbeiten, mit der einer Gruppe von Super-Recognizern übereinstimmt. Das Hauptziel bestand darin, festzustellen, ob die hohe Identifizierungsrate der Identifizierer auf eine Super-Recognizer-Fähigkeit hindeutet. Ein weiteres Ziel war es, die Fähigkeiten sowohl der Identifizierer als auch der Super-Recognizer zu verstehen, um festzustellen, ob gemeinsame Merkmale die Erfolge der Identifizierer erklären könnten. Deshalb basierten die Tests in der aktuellen Forschung auf Faktoren, die die Identifikationsgenauigkeit bei manchmal schlechten CCTV-Bildern beeinflussen könnten. Dazu gehörten die Erkennung vertrauter Gesichter, die seit vielen Jahren nicht mehr gesehen wurden, aus degradierten, bewegten und statischen Bildern; das Unterscheiden kurz erlernter unbekannter Gesichter von Reihen physisch ähnlicher Ablenkungen; die Extrapolation der Identität von einem Gesichtsblickpunkt auf einen zweiten; die Neigung, sich beim Erlernen neuer Gesichter auf zuverlässigere und stabilere innere Gesichtsmarkmale zu konzentrieren; Selbstvertrauen; und der gleichzeitige Abgleich unbekannter Gesichter. [8]

Theorien legen nahe, dass Gesichter möglicherweise „speziell“ sind, entweder aufgrund von Anpassung oder Expertise. Neuere Erkenntnisse besagen auch, dass die überdurchschnittlichen Fähigkeiten von Super-Recognizern möglicherweise ebenfalls gesichtsspezifisch sind. Ein Objektgedächtnistest untersuchte, ob sich die Fähigkeit zur Super-Erkennung auch auf andere visuelle Gedächtnisaufgaben erstreckt. Von Super-Recognizern wurden aufgrund ihrer Definition bessere Ergebnisse bei den vier Gesichtsverarbeitungstests als bei der Kontrollgruppe erwartet. Mitarbeiter in Positionen, bei denen die Gesichtserinnerungsfähigkeit wichtig ist, schneiden oft nicht besser bei Gesichtsverarbeitungstests ab als Mitglieder der Öffentlichkeit (z. B. Passbeamte). Allerdings hat aktuelle Forschung konsequent gezeigt, dass es individuelle Unterschiede in der Fähigkeit innerhalb dieser Gruppen gibt, einschließlich des MPS. [8]

Daher wurde aufgrund der außergewöhnlichen Leistung im operativen Kontext erwartet, dass die Polizei-Identifizierer als Gruppe auch bei den vier Gesichtsverarbeitungstests genauer abschneiden würden als die Kontrollgruppe. Dennoch, da die Leistungen einiger Super-Recognizer bei verschiedenen Gesichtsverarbeitungstests variieren können und ihre hohen Identifizierungsraten unter unterschiedlichen Umständen erworben werden können, war im Voraus unklar, ob dieser Vorteil bei allen Polizei-Identifizierern festgestellt werden würde. Tatsächlich korreliert die Leistung normalerweise beim gleichzeitigen Abgleich von Gesichtern und beim Gesichtsgedächtnis, aber einige Prosopagnosiker können Gesichter im normalen Bereich abgleichen, und nicht alle Super-Recognizer sind exzellent im Abgleich unbekannter Gesichter. Dies legt nahe, dass Gesichtsgedächtnis und Abgleich in einigen Fällen auf unterschiedliche Mechanismen zurückgreifen können. [8]

Aus diesem Grund wurden neben Gruppenanalysen auch individuelle Analysen durchgeführt, indem die Leistung jedes Polizei-Identifizierers und Super-Recognizer bei jedem Test mit den Kontrollen verglichen wurde. Dies ermöglichte es, die Konsistenz der Testleistung zu messen und eine Schätzung des Anteils der allgemeinen Bevölkerung zu erstellen, den jeder übertreffen würde. Schließlich untersuchte eine korrelative Komponente die Beziehungen zwischen den Testleistungen der Teilnehmer, in der Erwartung, dass Ergebnisse bei den vier Gesichtsverarbeitungstests, aber nicht unbedingt bei der Aufgabe zum Objektgedächtnis, positiv korrelieren würden. [8]

Die Studie von Davis et al. besteht aus fünf Tests, welche die Leistung von Super-Recognizern, Identifizierer der Polizei und einer Kontrollgruppe in vergleichen. Diese wurden in folgender Reihenfolge durchgeführt:

- Unfamiliar Face Memory Array Test (Bruce et al., 1999)

- Famous Face Recognition Test (Lander et al., 2001)
- Objekt (Blumen) Gedächtnistest
- Old/New Unfamiliar Face Memory Test
- Glasgow Face Matching Test [8]

Daran teilgenommen haben 10 Super-Recognizer, 36 Identifizierer der Polizei sie waren Mitglieder der freiwilligen Super-Recognizer Gruppe und 143 Menschen, die die Kontrollgruppe bilden. [8]
Es wurden keine Entschädigungen an Kontrollen oder Super-Recognizern gezahlt. [8]

Der Famous Face Recognition Test von Lander et al. (2001) bestand aus zwei gegenübergestellten Sets von je 15 männlichen und 15 weiblichen Prominentengesichtern sowie 5 männlichen und 5 weiblichen, etwa 12 Jahre zuvor aufgenommenen, unbekannten Gesichtern. Jedes Gesicht wurde 5 Sekunden lang gezeigt, die Hälfte davon in Bewegung (20) und die Hälfte statisch (20). Wenn ein Gesicht in Set A in Bewegung war, war es in Set B statisch und umgekehrt (80 Durchgänge). Antworten zu den Prominenten wurden als Treffer kategorisiert, wenn Teilnehmer richtige Namen oder individualisierende Informationen gaben. Als Fehlidentifikationen wurden falsche Namen/Informationen oder komplettes Nichtbeantworten gewertet. Bei den unbekannten Gesichtern waren Antworten entweder korrekte Ablehnungen oder falsche Alarme, wenn falsche Namen/Identitäten angegeben wurden. Den Teilnehmern wurde danach eine Liste der Prominentennamen gegeben und gefragt, ob sie diese erkannt haben sollten. Wenn der Teilnehmer angab, dass er dieses Gesicht nicht erkannt hätte, wurden die Antworten aus der Analyse ausgeschlossen. [8]

Der Unfamiliar Face Memory Array Test wurde ursprünglich für eine Studie zum Gesichtsabgleich (Bruce et al., 1999) entwickelt. In diesem Gedächtnisdesign absolvierten die Teilnehmer in vier gegenübergestellten Versionen insgesamt 40 Durchgänge. Es wurde jeweils ein einzelnes farbiges Gesicht eines weiß-kaukasischen Mannes für 5 Sekunden gezeigt. Von den 40 Durchgängen waren 20 aus einer frontalen Perspektive und die anderen 20 unter einem Winkel von 30°. Nach jedem Gesicht wurden dem Teilnehmer 10 zufällig angeordnete farbige Gesichter desselben Tages, aber unterschiedlicher Kameras, gezeigt. Das Ziel war es, das vorher gezeigte Gesicht zu identifizieren. Die Teilnehmer bekamen den Hinweis, dass bei der Hälfte der Durchgänge das gesuchte Gesicht nicht dabei ist. Die Ergebnisse waren entweder Treffer, Fehlidentifikationen oder Auslassungen. Wenn das gesuchte Gesicht nicht dabei war, waren die Ergebnisse entweder korrekte Ablehnungen oder falsche Zuordnungen. Nach jedem Durchgang wurde auch nach der Vertrauensbewertung der Entscheidung gefragt. Als Antwort konnte man von 1 (niedrig) bis 5 (hoch) angeben. Es gab keine Zeitbeschränkungen für die Testphasen. [8]

Der Old/New Unfamiliar Face Memory Test wurde entwickelt, um die Neigung zu ermitteln, ob sich bei der Erinnerung an neue Gesichter auf interne Regionen oder auf externe Merkmale konzentriert wird. In den Lernphasen von zwei gegenübergestellten Versionen wurden 20 zufällig geordnete, nacheinander präsentierte Farbfotos von unbekannten weiß-kaukasischen Männern (10 Gesichter) und Frauen (10 Gesichter) vom Bauch aufwärts für jeweils 5 Sekunden gezeigt. Die Teilnehmer sollten sich die Gesichter merken, wobei auch Kontextinformationen wie Raum und Kleidung vorhanden waren. Danach sahen die Teilnehmer 40 nacheinander gezeigte Gesichter, die externe Gesichtsmerkmale und verschleierte Hintergrundhinweise aufwiesen. Sie beurteilten, ob es sich um ein altes oder neues Gesicht handelt. Den Teilnehmern wurde nicht im Voraus mitgeteilt, dass nur interne Merkmale gezeigt würden. Unterschiedliche Fotos vom selben Tag derselben Person wurden in den Lern- und Testphasen verwendet. Es gab keine zeitliche Begrenzung für die Testphase. [8]

Der Glasgow Face Matching Test (Kurzversion: Burton et al., 2010) besteht aus 40 Paaren gleichzeitig präsentierter weiß-kaukasischer Gesichter in Graustufen. Die Teilnehmer müssen angeben,

ob die Gesichter „gleich“ sind oder „unterschiedlich“. Beide Arten sind gleichmäßig vertreten, mit jeweils 20 Paaren. Für das Antworten gab es keine zeitlichen Begrenzungen. Basierend auf den 194 Teilnehmern fanden Burton et al. heraus, dass die normierten durchschnittlichen Trefferraten bei diesem Test 79,8% betragen und die korrekten Ablehnungen 82,5%. [8]

Bei dem Object Memory Test handelt es sich um einen ähnlichen Test wie den Old/New Unfamiliar Face Memory Test, nur dass es sich um Blumen statt Gesichter handelt. Hier bekamen die Teilnehmer in der Lernphase eine Reihe von 20 Farbfotografien von Blumen, jeweils für 5 Sekunden, zu sehen. In der Testphase sahen sie 40 Bilder und beurteilten, ob es sich um ein altes oder neues Bild handelte. Auch hier waren 20 von den 40 Bildern alte und 20 neue Bilder. Die gleichen „alten“ Fotografien wurden in beiden Phasen verwendet. Es gab keine zeitlichen Begrenzungen für die Antwort.[8]

Bei dem Famous Face Recognition Test wurden die Teilnehmer vor der Analyse der Ergebnisse gefragt, welche der Prominenten sie niemals erkannt hätten. Die Super-Recognizer und die Identifizierer der Polizei waren genauer als die Kontrollgruppe beim Erkennen von degradierten berühmten Gesichtern und beim korrekten Ablehnen von unbekannten Gesichtern. Alle drei Gruppen waren besser darin, bewegte Gesichter zu erkennen als statische. [8]

Die Trefferquoten der Super-Recognizer und Polizei-Identifizierer zeigen keine signifikanten Unterschiede, waren jedoch höher als die der Kontrollgruppe. Dabei waren die Trefferquoten bei bewegten Bildern höher als bei statischen Bildern. Der Bewegungsvorteil war für die Super-Recognizer und Polizei-Identifizierer größer als für die Kontrollgruppe. [8]

Die beiden Gruppen unterschieden sich auch bei den korrekten Ablehnungen nicht, und beide machten mehr korrekte Ablehnungen als die Kontrollgruppe. [8]

Beim Test zum Gedächtnis von unbekannten Gesichtern waren die Super-Recognizer und die Identifizierer der Polizei, unabhängig vom Blickwinkel des Gesichts, genauer und zuversichtlicher darin, die unbekannten Gesichter aus einer Reihe von 10 Bildern zu erkennen. Dies galt auch für das Ablehnen der Reihe, wenn das gesuchte Gesicht nicht dabei war. Die Genauigkeit stand in positivem Zusammenhang mit dem Vertrauen. [8]

Die Trefferquoten der Super-Recognizer und der Polizei-Identifizierer zeigten auch hier keine signifikanten Unterschiede, waren aber wieder höher als die der Kontrollgruppe. Der Blickwinkelleffekt war signifikant: Die Trefferquote bei der Frontalansicht der Gesichter war höher als bei der Dreiviertelansicht eines Gesichts. [8]

Gleiches galt für die korrekte Ablehnung. Die Super-Recognizer und Identifizierer unterschieden sich nicht, waren jedoch besser als die Kontrollgruppe. Der Blickwinkelleffekt war dabei nur geringfügig signifikant. [8]

Auch bei dem Selbstvertrauen in die eigenen Antworten unterschieden sich die Super-Recognizer und Identifizierer nicht, es war jedoch höher als bei der Kontrollgruppe. Der Blickwinkelleffekt war signifikant: Das Selbstvertrauen war höher bei der Frontalansicht des kompletten Gesichts als bei der 3/4-Ansicht des Gesichts. [8]

Bei dem alten/neuen Gedächtnistest für unbekannte Gesichter waren die Trefferquoten der Super-Recognizer geringfügig höher als die der Kontrollgruppe. Bei den korrekten Ablehnungen waren die Raten der Identifizierer höher als die der Kontrollgruppe. [8] Im Glasgow Face Matching Test waren die Trefferquoten der Super-Recognizer, der Polizei-Identifizierer und der Kontrollgruppe signifikant höher als die bereits veröffentlichten normativen Testdaten. Super-Recognizer und Polizei-Identifizierende unterschieden sich nicht, waren jedoch bei allen Ergebnissen genauer als die Kontrollgruppe. [8]

Im Objekt-Gedächtnistest waren die Trefferquoten der Super-Recognizer höher als die der Identifizierer und der Kontrollgruppe. Bei den korrekten Ablehnungen gab es keine Unterschiede zwischen den Gruppen. Die Trefferquoten der Polizei-Identifizierer und der Kontrollgruppe wiesen keinen signifikanten Unterschied auf. [8]

Alle Ergebnisse der vier Tests verglichen, zeigten, dass sich die Super-Recognizer und die Identifizierer der Polizei nicht unterschieden. Die Kontrollgruppe war in Trefferquote und korrekten Ablehnungen schlechter als die beiden Gruppen. [8]

Die Betrachtung der Leistungen in den einzelnen Tests ergab folgende Zusammenhänge. Die Trefferquoten und die korrekten Ablehnungen korrelierten im Famous Face Recognition Test und dem Unfamiliar Face Memory Test signifikant. Ebenso separat bei dem Glasgow Face Matching Test und dem Old/New Face Memory Test. Das deutete darauf hin, dass diese Tests teilweise unterschiedliche Aspekte oder Dimensionen der Gesichtserkennung messen. Der Object Memory Test hatte nur mit dem Old/New Unfamiliar Face Memory signifikante Korrelationen, aber sonst mit keinem der anderen Tests. Die Korrelation zwischen den beiden kann auf dem identischen Testdesign basieren. [8]

Diese Forschung hat herausgefunden, dass die Identifizierer der Londoner MPS, die mehrere Identifikationen von Verdächtigen aus CCTV gemacht haben, als Gruppe genauer waren als die Kontrollgruppe aus der Allgemeinbevölkerung bei den Tests zum gleichzeitigen Abgleich unbekannter Gesichter und der Erinnerung an bekannte und unbekannte Gesichter. Die Leistungen der Identifizierer unterschieden sich nicht zuverlässig von denen der Super-Recognizer, die bestimmte Kriterien für diese herausragende Gesichtserkennungsfähigkeit erfüllen. Forschungen zeigen, dass es in der Bevölkerung Personen mit außergewöhnlichen Fähigkeiten gibt, sodass auch einige Polizisten herausragende Fähigkeiten besitzen können. [8]

Bei dem Glasgow Face Matching Test schnitt sowohl die Gruppe der Super-Recognizer als auch die Gruppe der Identifizierer der MPS besser ab als die Kontrollpersonen. Diese wiederum schnitt höher ab als die ursprünglichen veröffentlichten Daten von 194 Teilnehmern in diesem Test. Eine mögliche Erklärung dafür könnte sein, dass Personen, die eine Neigung zur guten Gesichtserkennungsfähigkeit haben, sich eher dazu geneigt fühlen, bei solch einem Test teilzunehmen. [8]

Die Analysen der Werte jedes Super-Recognizers und jedes Identifizierers zeigten jedoch, dass nur wenige über alle vier Gesichtsverarbeitungstests hinweg außergewöhnliche Leistungen erzielten. Gründe dafür könnten die Designs der einzelnen Tests sein, wie das Entfernen externer Merkmale der Gesichter, die kurze Zeit zum Anschauen der Bilder, Deckeneffekte, Interesse an zeitgenössischer Kultur und Müdigkeit. Dennoch erzielten die Durchschnittswerte von fast allen Super-Recognizern und Identifizierern der Polizei über die vier Tests hinweg höhere Werte als der Kontrollmittelwert, was auf eine überdurchschnittliche Fähigkeit hindeutet. Beide Gruppen erzielten auch höhere Treffer- und korrekte Ablehnungsraten als die Kontrollgruppe. Bei der Polizei würden hohe korrekte Ablehnungsraten dazu führen, dass nur wenige falsche Spuren erzeugt werden. [8]

Im Gegensatz zu den vier Gesichtserkennungstests wurden beim Object Memory Test keine zuverlässigen Unterschiede zwischen den Gruppen festgestellt. Diese Ergebnisse unterstützen frühere Forschungen, die besagen, dass die Fähigkeit zur Super-Recognition gesichtsspezifisch ist. Genaue Schlussfolgerungen sind jedoch nicht möglich, da nur ein Nicht-Gesichtstest dabei war. Bei anderen Aufgaben könnten die Super-Recognizer besser abschneiden. [8]

Die Ergebnisse des Famous Face Recognition Tests deuten darauf hin, dass Super-Recognizer möglicherweise eine bessere Langzeiterinnerung an Gesichter besitzen, da sie eine größere Anzahl von Prominentengesichtern aus dem Jahr 2001 erkannten, von denen einige nicht mehr oft in den Medien zu sehen waren. Die Leistung bei bewegten Bildern war generell höher, aber die Super-Recognizer und Identifizierer der Polizei waren besser als die Kontrollgruppe. Eventuell können diese Gruppen effektiver die zusätzlichen Informationen aus den bewegten Bildern extrahieren. Die Polizei sollte

daher erwägen, bei der Suche nach gesuchten Personen bewegte Bilder auf Websites zu verwenden, um die Identifikationsraten von Verdächtigen zu erhöhen. [8]

Eine überlegene Leistung beim erweiterten CFMT zu erzielen, war das Hauptziel der objektiven Diagnose der Super-Recognizer. Der Mindeststandard war, einen Punktestand von 90 von 102 zu erreichen. Bei den hier durchgeführten Analysen zeigte sich, dass zwei von zehn Super-Recognizern diesen Punktestand leicht überschritten haben, jedoch bei den Gesichtsverarbeitungstests relativ schlecht abschnitten. Solche Inkonsistenzen wurden bereits in früheren Studien festgestellt, so dass zukünftige Forschungen die Nützlichkeit des CFMT, der zur Diagnose der Super-Recognition-Fähigkeit verwendet wird, weiter untersuchen sollten. [8]

Aus den Leistungen der von Davis et al. verwendeten Tests ist es ihnen nicht möglich, einen Schwellenwert für die Mitgliedschaft in der Identifizierer-Gruppe der Polizei vorzuschlagen. Trotz der schlechter abschneidenden Polizei-Identifizierer würde ein willkürlich gewählter Schwellenwert einige ausschließen, die sonst eine erhebliche Anzahl von Identifikationen vornehmen würden. Dennoch legen die guten Ergebnisse vieler Identifizierer nahe, dass es sinnvoll wäre, die Gesichtsverarbeitungsfähigkeiten aller Polizeibeamten zu messen. Die Super-Recognizer-Polizisten könnten aus ihren normalen Aufgaben herausgenommen werden, um Tatortbilder zu betrachten, was die Identifizierungs- und Verurteilungsraten erhöhen sollte. [8]

Ausgehend von dieser Studie gibt es mehrere Faktoren, die Schlussfolgerungen einschränken. Die Anzahl der Super-Recognizer war zwar gering, aber dennoch größer als bei jeder zu der Zeit veröffentlichten Forschung. Zudem stellten die Identifizierer der Polizei nur eine Minderheit des MPS dar, und die Fähigkeiten der übrigen Personen sind unklar. Außerdem war der Großteil der Teilnehmer und alle Reize weiß-kaukasisch, sodass keine Aussagen darüber getroffen werden können, wie die Erkennungsfähigkeit der Super-Recognizer bei unterschiedlichen ethnischen Gruppen oder Altersgruppen ist. Darüber hinaus betrug die Belichtungszeit der Lernphase in beiden Tests mit unbekannten Gesichtern nur 5 Sekunden. Eine längere Belichtungszeit würde die Leistung wahrscheinlich verbessern. [8]

Ein Anreiz für bessere Leistungen könnte auch eine Belohnung sein. Bei den Identifizierern der Polizei könnten die Erfolge bei der Identifizierung von Verdächtigen eine intrinsische Motivation für Aufgaben zur Gesichtsverarbeitung darstellen. Ihre leicht schlechtere Leistung beim Objektgedächtnistest könnte auf eine geringere Motivation für diese Aufgabe hinweisen. Motivationale Einflüsse wirken jedoch nur begrenzt auf die Leistung ein. [8]

Zusammenfassend zeigt diese Forschung, dass einige der spezialisierten Polizeibeamten, die in ihrem Beruf regelmäßig Verdächtige anhand von Bildern identifizieren, über herausragende Gesichtsverarbeitungsfähigkeiten verfügen. Jede Arbeit, die persönlichen, aber unregelmäßigen Kontakt mit Menschen erfordert, könnte durch Mitarbeiter mit herausragenden Gesichtserkennungsfähigkeiten verbessert werden. Den Identifikationen durch einen Super-Recognizer der Polizei vor Gerichten eine hohe Beweiskraft beizumessen, wäre dennoch riskant. Diese Personen sind nicht übermenschlich, und keiner erreichte in der vorliegenden Forschung in allen Tests eine perfekte Punktzahl. Aufzeichnungen zeigen, dass auch Identifizierer gelegentlich mit ihren Identifikationen falsch liegen. In England und Wales gibt es seit 1984 rechtliche Schutzmaßnahmen wie das Police and Criminal Evidence Act (PACE), um die Genauigkeit und Fairness bei der Identifizierung von Verdächtigen zu gewährleisten und gleichzeitig die Rechte der betroffenen Personen zu schützen. Zu dieser Zeit gab es weltweit nur wenige identifizierte Super-Recognizer bei der Polizei. Wenn andere Rechtssysteme solche Indizien als Hinweise übernehmen, sind ähnliche Regeln zum Schutz der Rechte von Verdächtigen erforderlich. Dennoch sollten polizeiliche Super-Recognizer als hilfreiche Ermitt-

lungsressource der Strafverfolgungsbehörden betrachtet werden. Eine vorläufige Identifikation durch einen polizeilichen Super-Recognizer kann eine erste wichtige Spur liefern, auf deren Basis weitere und stärkere Beweise für die Schuld gefunden werden können. [8]

3.2.5 Identification from CCTV: Assessing police super-Recognizer ability to spot faces in a crowd and susceptibility to change blindness

Weltweit werden von der Polizei Überwachungsaufnahmen (CCTV) für Ermittlungen überprüft. Es werden Zeitpläne für Ereignisse erstellt, nach Personen von Interesse gesucht und Verdächtige identifiziert. Solche Überprüfungen können zeitaufwendig sein. Die Ermittler müssen ständig ihre Aufmerksamkeit aufrechterhalten und sich Bilder von Zielpersonen merken. Wachsamkeitsaufgaben sind mit hoher geistiger Arbeitsbelastung verbunden und werden von Erfahrung, Müdigkeit und Stress beeinflusst. Automatisierte Suchhilfen, Übung und Koffein verbessern die Wachsamkeit in anderen Bereichen. Angesichts der Implikationen für die innere Sicherheit und Polizeiarbeit wurde die Forschung, die untersucht, ob die CCTV-Suche durch den Einsatz von Personal mit spezifischen Kompetenzen optimiert werden kann, jedoch begrenzt. Die vorliegende Forschung von Josh P. Davis et al. (2018) hatte daher zum Ziel, zu untersuchen, ob menschliche oder operationale Faktoren die Fähigkeit beeinflussen, Personen von Interesse in Videos zu lokalisieren. [7]

Bei der Gesichtserkennungsfähigkeit gibt es große Unterschiede, wobei Menschen mit entwicklungsbedingter Prosopagnosie und sogenannten Super-Recognizern (SRs) an den Extremen stehen. Während es umfangreiche Forschung zur entwicklungsbedingten Prosopagnosie gibt, wurde die Überlegenheit der SRs nur in wenigen Studien untersucht. Wenn SRs in Gruppen einbezogen werden, verwenden die meisten Forscher Selbstberichte über außergewöhnliche Fähigkeiten und Punktzahlen. Solche Berichte können jedoch unzuverlässig sein, und es wurden verschiedene CFMT+-Schwellenwerte verwendet. Einige der SRs, die hohe Schwellenwerte im CFMT+ erreichen, zeigen eine schlechte Leistung bei der simultanen Gesichtsanpassung. Das kann darauf hindeuten, dass Super-Recognizer möglicherweise ein Sammelbegriff ist. Der CFMT+ ist an sich ein guter Indikator für die SR-Fähigkeit, jedoch sollten für eine zuverlässige Diagnose zusätzliche Tests genutzt werden. Diese könnten mehrere Tests umfassen, was Müdigkeit verursachen kann und somit ein Problem bei der Beurteilung außergewöhnlicher Fähigkeiten darstellt. In der vorliegenden Forschung sollten die Teilnehmer daher nur den CFMT+ absolvieren. [7]

Personen mit außergewöhnlichen Gesichtsverarbeitungsfähigkeiten haben Auswirkungen auf die Strafverfolgung und können Verdächtige identifizieren. Die Metropolitan Police Service (MPS) in London hat eine Gruppe von Polizisten mit solchen Fähigkeiten zusammengestellt. Nicht alle von ihnen wurden empirisch getestet, weshalb sie hier als Polizei-Identifizierer bezeichnet werden. [7]

Einige dieser Identifikatoren wurden bei Veranstaltungen mit großen Menschenmengen eingesetzt, wie bei Popkonzerten oder Ereignissen wie dem Notting Hill Carnival in London. Dort versuchten sie, über CCTV unbekannte Verdächtige zu erkennen, deren fotografierte Gesichter sie sich eingeprägt hatten. [7]

Im Jahr 2015 wurde eine Gruppe von Vollzeitpolizisten gegründet, die sich auf die Überprüfung von CCTV-Aufnahmen spezialisiert hat. Laut Medienberichten erkennen diese Experten regelmäßig unbekannte Verdächtige über verschiedene Tatorte hinweg. Diese Berichte stimmen mit den Selbstberichten von SRs überein, die behaupten, dass die Vertrautheit mit unbekannten Gesichtern einfach sei. Dies ist bemerkenswert, da die Gesichtserkennung, der Abgleich und die Suche nach Gesichtern in Menschenmengen bei vertrauten Gesichtern präziser sind als bei unbekannten Gesichtern. Einige der Polizeiexperten schnitten bei Tests zum Abgleich unbekannter und vertrauter Gesichter „deutlich

über dem Normalwert“ ab. [7]

Um die Fähigkeit zur Suche nach Personen von Interesse zu untersuchen, haben die zuvor erwähnten Polizeixperten, die Identifizierer der Polizei und eine nicht-polizeiliche Kontrollgruppe den CFMT+ absolviert. Anschließend wurde in einem neuartigen Video-Review-Test, dem „Spot the Face in a Crowd Test“ (SFCT), nach unbekannten Schauspielern gesucht. Dieser Test wurde entwickelt, um zu untersuchen, ob er bei der Überprüfung von Rollen eine Entscheidungshilfe sein könnte und ob die Leistung im SFCT durch Erfahrung und/oder Gesichtserkennungsfähigkeit beeinflusst wird. Beim SFCT wird sowohl das Gedächtnis als auch die Wahrnehmung verwendet, da die Teilnehmer Bilder von Schauspielern mit Personen in Menschenmengen im Video vergleichen müssen. Sie können entweder das Video abspielen und die Personen in Bewegung betrachten oder es pausieren, um einzelne Frames anzusehen. Der Abgleich unbekannter Gesichter kann unzuverlässig sein, insbesondere wenn, wie im Test, die Ziele ihr Aussehen verändert haben – ein häufiges Problem bei der Polizei. Im SFCT erhielten die Teilnehmer zwei, vier oder acht Zielakteure, um die Auswirkungen der Anzahl der Zielakteure auf die Leistung zu untersuchen. [7]

Bruce et al. stellten im Jahr 2001 fest, dass die Teilnehmer besser beim gleichzeitigen Gesichtsabgleich abschnitten, wenn diese sich im Voraus in Paaren über die wahrgenommenen Persönlichkeiten der Gesichter austauschten [4]. Diese Methode wurde auch in der hier beschriebenen Studie übernommen. Dadurch könnte eine tiefere Verarbeitung gefördert werden, und die Aufmerksamkeit der Teilnehmer könnte stärker auf Hinweise zur Identifikation gerichtet sein. Einige Teilnehmer aus der Kontrollgruppe haben an einem vorherigen Training teilgenommen, das darauf abzielte, ihnen Schauspieler näherzubringen. Es wurde erwartet, dass diese Teilnehmer besser abschneiden würden als untrainierte Teilnehmer. [7]

Erfahrene Polizisten können bedrohliche Bilder effektiver erkennen als weniger erfahrene. Daher spielt bei der Erkennung der Schauspieler im SFCT ein Erfahrungsfaktor eine Rolle. Um die Einflussfaktoren festzustellen, haben die Teilnehmer verschiedene Tests durchgeführt, darunter Messungen der Persönlichkeit, der wahrgenommenen Arbeitsbelastung im SFCT und der Anfälligkeit für Veränderungsblindheit. [7]

Bei der Überprüfung von CCTV-Aufnahmen ist dauerhafte Konzentration erforderlich, wobei jedoch Veränderungsblindheit auftreten kann. Dabei können Anomalien in den Aufnahmen dem Betrachter aufgrund seiner fokussierten Aufmerksamkeit entgehen. Auch Veränderungsblindheit kann auftreten, bei der Umweltveränderungen, einschließlich Personen, übersehen werden. Dies könnte die Überprüfung von CCTV beeinflussen, insbesondere wenn Personen von Interesse mit verschiedenen Kameras verfolgt werden und zwischendurch Aufnahmelücken entstehen. Frühere Forschungen haben jedoch gezeigt, dass Experten weniger von Veränderungsblindheit betroffen sind als Laien. In der Forschung von Davis et al. wurde ein Test zur Veränderungsblindheit verwendet, bei dem ein Fahrerwechsel erkannt werden sollte. Dieser Test wurde kurz nach dem SFCT durchgeführt, wenn die Teilnehmer am anfälligsten für Müdigkeit und Veränderungsblindheit sein könnten. Alternativ könnten auch die Teilnehmer, insbesondere die mit überdurchschnittlicher Gesichtserkennung, durch die Gewöhnung an Videoanalysen weniger anfällig für Veränderungsblindheit sein. [7]

Für die Ergebnisse der Studie wurden mehrere Hypothesen aufgestellt. Einerseits wurde erwartet, dass die Polizeixperten den Polizei-Identifizierern und der Kontrollgruppe beim CFMT+, dem SFCT und dem Veränderungsblindheitstest überlegen sind. Diese Gruppen sind meist empirisch getestet worden, verfügen über herausragende Gesichtsverarbeitungsfähigkeiten, arbeiten Vollzeit in der CCTV-Überprüfung und haben bereits zahlreiche Identifikationen von unbekannten Verdächtigen vorgenommen. Es wurde auch erwartet, dass die Polizei-Identifizierer, die ebenfalls über herausragende Gesichtserkennungsfähigkeiten verfügen, die Kontrollgruppe übertreffen werden.[7]

Die Studie bestand aus drei kognitiven Tests: dem CFMT+, dem SFCT und dem Veränderungsblindheitstest. Außerdem wurden die Teilnehmer gebeten, den „Big Five“-Persönlichkeitstest auszufüllen und für den SFCT die wahrgenommene Arbeitsbelastung anzugeben. Die Leistungen der Polizeiexperten des MPS, der Polizei-Identifizierer des MPS und der nicht-polizeilichen Kontrollgruppe wurden verglichen. Individuelle Analysen verglichen zudem die Leistungen aller Teilnehmer, die einen hohen SR-Schwellenwert erfüllten, mit dem Durchschnitt der Nicht-Polizei-Identifizierer. [7]

Zusätzliche Variablen für die Kontrollgruppen umfassten das Training sowie die Teilnahme an einer Vorübung zum Vertrautmachen mit den Schauspielern im Rahmen des SFCT (entweder vertraut gemacht oder nicht trainiert). Für die Kontrollgruppen und die Polizei-Identifizierer wurde ebenfalls die Anzahl der gesichteten Schauspieler berücksichtigt, wobei die Teilnehmer nach zwei, vier oder acht Schauspielern suchten. Aufgrund der begrenzten Anzahl an Polizeiexperten suchten diese immer nach acht Schauspielern. Die Hauptkennzahlen des SFCT umfassten die korrekten Identifikationen der Schauspieler und das dazugehörige Vertrauen, die falsch positiven Identifikationen von Unbeteiligten und das Vertrauen, sowie die korrekten Ablehnungen von Ausschnitten, die keine Darsteller enthielten. [7]

Die Teilnehmer der Studie bestanden aus den Polizeiexperten, die gesamte Mitgliedschaft der sieben Personen der Vollzeit-Einheit der MPS. Dazu kommen noch 92 Polizei-Identifizierer und die Straßenpolizisten der SR-Gruppe. Es gab keine Kenntnisse darüber, wie und aus welchen Gründen die Gruppenmitglieder ernannt wurden. Zusätzlich gab es eine Kontrollgruppe, bestehend aus Mitarbeitern und Studenten der University of Greenwich, insgesamt 152 Personen. [7]

Die Teilnehmer der Polizei waren überwiegend männlich, weiß und über 30 Jahre alt. Die Personen der Kontrollgruppe dahingegen waren hauptsächlich weiblich, unter 25 Jahre alt, mit einem höheren Anteil an Nicht-Weißen. [7]

Der erste zu absolvierende Test ist der CFMT+, dabei machen sich die Teilnehmer mit unbekannten Gesichtern in zunehmend schwierigeren Anordnungen vertraut und erkennen sie wieder. [7]

Der zweite Test war der „Spotting a Face in a Crowd Test“ und bestand aus einem 18 Minuten und 21 Sekunden langen Video, das in 11 Clips unterteilt ist. Die Aufnahmen wurden aus drei verschiedenen Perspektiven aufgenommen: entweder von einem Gerüst über dem Bodenniveau, einem Stativ auf Augenhöhe oder von erhöhtem Gelände. Es wurden acht weiß-kaukasische Schauspieler gefilmt, die kurz durch das Sichtfeld gingen. Jeder Clip enthielt entweder zwei, einen oder keinen Schauspieler; ein Schauspieler erschien jedoch immer nur in einem oder zwei Clips. Die Entfernung der Schauspieler und der Nebenakteure zur Kamera sowie die Anzahl der Nebenakteure variierten erheblich. Die Teilnehmer erhielten von jedem Schauspieler vier Farbfotos auf Papier, die von den Schauspielern selbst ausgewählt wurden. Die Fotos wurden nach dem Kriterium ausgewählt, dass es sich um Bilder handelt, die ihre Familien der Polizei geben würden, wenn sie als vermisst gemeldet wären, und achteten auf unterschiedliche Erscheinungsbilder. Manche Schauspieler wurden unbeabsichtigt sowohl auf einem Vergleichsfoto als auch im Video mit dem gleichen oder ähnlichen Kleidungsstück dargestellt. [7]

Die Teilnehmer konnten für den Test immer zu zweit Zeitslots buchen. Sie wurden angewiesen, das Videomaterial so zu durchgehen, als würden sie nach einer vermissten Person suchen. Den Polizeiexperten wurden immer acht Schauspieler zugewiesen, während die übrigen Teilnehmer vier Fotografien von entweder zwei, vier oder acht Schauspielern erhielten. Die Zuweisung der Schauspieler erfolgte zufällig, jedoch wurde innerhalb jeder experimentellen Bedingung jeder Schauspieler gleich oft gesucht. [7]

Etwa die Hälfte der Personen aus der Kontrollgruppe wurde zufällig der Vorübung zum Vertrautmachen mit den Schauspielern zugeteilt. Ihnen wurde mitgeteilt, dass durch ein vorheriges Besprechen der wahrgenommenen Persönlichkeit das Erlernen von Gesichtern verbessert werde. Die Teilnehmer

sollten sich verbal auf eine Bewertung von 1 bis 10 zu einer Reihe von Fragen einigen. Sie sollten beispielsweise einschätzen, wie abenteuerlustig, gutmütig, intelligent usw. die Zielperson erscheint. Die Teilnehmer sollten sich auch auf eine Meinung einigen und beurteilen, ob die Zielperson ein breites oder schmales Gesicht hat, was ihre markantesten Gesichtsmarkmale sind und Augen, Mund sowie Gesichtsform beschreiben. Dies dauerte bis zu 20 Minuten, und keine der Antworten wurde aufgezeichnet. Alle nicht für das Vertrautmachen ausgesuchten Teilnehmer begannen sofort mit dem Test. [7]

Der SFCT wurde von allen Teilnehmern einzeln in separaten Kabinen absolviert. Sie konnten das Video nach Belieben steuern, beispielsweise zurückspulen oder pausieren. Außerdem wurden sie darauf hingewiesen, dass es Clips ohne gesuchte Schauspieler gibt, aber auch solche, die einen oder mehrere enthalten. Zum Vergleichen konnten die Teilnehmer die Fotos und das Video gleichzeitig anschauen. Wenn vermeintlich eine Zielperson gefunden wurde, wurde das Video pausiert und die Teilnehmer füllten ein Papierformular aus. Auf diesem Formular mussten der Zielperson zugewiesene Buchstabe (a-h), der Zeitstempel des Videos, eine kurze Beschreibung der Zielperson und die Einschätzung ihrer Sicherheit bei der Entscheidung angegeben werden. Für den Test gab es keine zeitlichen Beschränkungen. [7]

Antworten wurden als Treffer, falsche Erkennung eines Nebenakteurs (FP) und korrekte Nicht-Identifikation eines Clips ohne Schauspieler (CR) gewertet. Bei der Suche nach acht Schauspielern konnten maximal 11 Treffer und 2 korrekte Nicht-Identifikationen erreicht werden. Die Teilnehmer, die nach zwei Schauspielern suchten, konnten maximal 2 bis 4 Treffer und 7 bis 9 korrekte Nicht-Identifikationen erzielen. Bei der Suche nach vier Schauspielern betrugen die maximalen Treffer 4 bis 7 und die korrekten Nicht-Identifikationen 5 bis 7. [7]

Die Teilnehmer sollten nach dem SFCT einen NASA Task Load Index ausfüllen. Dadurch wird mit einem Fünf-Faktoren-Maß die subjektive Arbeitsbelastung in dem Test bewertet. [7]

Zudem sollten die Teilnehmer zur Messung ihrer Persönlichkeit das NEO Personality Inventory-Revised ausfüllen. Dabei werden die Domänen Neurotizismus, Extraversion, Gewissenhaftigkeit, Offenheit und Verträglichkeit erfasst. Das Ausfüllen dauerte ungefähr 5 Minuten. [7]

Die Veränderungsblindheit der Teilnehmer sollte durch ein 2 Minuten und 44 Sekunden langes Video getestet werden, das eine Verkehrskontrolle aus der Perspektive des Armaturenbretts eines Polizeifahrzeugs zeigt. Zu Beginn des Videos steht ein Fahrzeug vor der Kamera, und ein Polizist befindet sich zwischen den beiden Fahrzeugen. Ein weißer, kaukasischer männlicher Fahrer (Fahrer 1), der einen Hut und einen Hoodie mit einem weißen T-Shirt darunter trägt, steigt aus seinem Fahrzeug aus. Etwa in der Mitte des Videos verlassen Fahrer 1 und der Polizist das Blickfeld. Als sie zurückkehren, wurde Fahrer 1 (18 Jahre alt, 188 cm groß, 56,7 kg schwer) durch einen anderen Fahrer (Fahrer 2) ersetzt. Fahrer 2 ist 20 Jahre alt, 175 cm groß, wiegt 61,0 kg und trägt ein grünes T-Shirt unter einem ähnlichen Hoodie. Der Polizist gibt Fahrer 2 ein Dokument, und dieser kehrt anschließend zu „seinem“ Fahrzeug zurück. [7]

Kurz danach beantworteten die Teilnehmer zwei Fragen, bei denen Beschreibungen des Videos gefordert wurden. Es gab zusätzlich zehn weitere Multiple-Choice-Fragen mit jeweils vier Antwortmöglichkeiten, die die Erinnerungen an die Ereignisse erfassten. Anschließend sahen die Teilnehmer vier Reihen mit jeweils acht Fotos und versuchten, Personen zu identifizieren, die sie wiedererkannt hatten. [7]

Zu den Ergebnissen der hier beschriebenen Studie:

Beim CFMT+ gibt es zwei verschiedene SR-Schwellenwerte. Einerseits erreichten 41 Teilnehmer den niedrigen Schwellenwert von 90 von 102 Punkten, während 14 Teilnehmer den hohen Schwellenwert von 95 von insgesamt 102 Punkten erreichten. Diese Leistungen führten zu einem höheren

Gesamtmittelwert als in einer zuvor durchgeführten Studie, in der der Schwellenwert von 95 von 102 Punkten zur Bestimmung der Super-Recognizers verwendet wurde. Polizeiexperten schnitt besser ab als die Polizei-Identifizierer, aber beide Gruppen schnitt besser ab als die Kontrollgruppe.[7]

Die Ergebnisse des SFCT ergaben, dass die Trefferquoten der Polizei-Identifizierer höher waren als die der untrainierten Kontrollgruppe und geringfügig höher als die der geschulten Kontrollgruppe. Auch bei der Anzahl der Schauspieler gab es einen Unterschied in der Trefferquote. Die Gruppen mit zwei Schauspielern hatten eine höhere Trefferquote als die Gruppen mit vier und acht Schauspielern, welche sich untereinander jedoch nicht voneinander unterschieden. Die Polizeiexperten und Identifizierer zeigten insgesamt höhere Trefferquoten im Vergleich zu den untrainierten Teilnehmern der Kontrollgruppe. [7]

Die Analyse der Falschpositiven Antworten ergab, dass geschulte Teilnehmer der Kontrollgruppe etwas seltener falsch positive Antworten gaben als ungeschulte. Auch die Polizeiexperten machten geringfügig weniger Falschpositive als die ungeschulten Teilnehmer der Kontrollgruppe. Bei den korrekten Ablehnungen schnitten die geschulten Teilnehmer der Kontrollgruppe ebenfalls besser ab als die ungeschulten. [7]

Bei dem Persönlichkeitstest der Teilnehmer stellte sich heraus, dass keine signifikante Korrelation mit den CFMT+ Werten vorliegt. [7]

Bei dem Test zur Veränderungsblindheit wurden die Antworten der Teilnehmer in zwei Kategorien eingeteilt. Zum einen Wechsel erkannt oder Wechsel nicht bemerkt. Insgesamt aus allen Teilnehmergruppen meldeten 79,4% die Änderung des Fahrers, 40,7% eine Änderung der Kleidung, 35,9% beide Änderungen und 15,7% der Teilnehmer nahmen keine Änderungen wahr.[7]

Aus den Ergebnissen geht hervor, dass eine positive Korrelation zwischen der Fähigkeit, sich an das Ereignis zu erinnern, und dem Bemerken der Veränderung im Video besteht. Teilnehmer, die gut darin waren, sich an das Ereignis zu erinnern, bemerkten eher den Fahrer- oder Kleiderwechsel. Teilnehmer, die sich weniger gut an das Ereignis erinnern konnten, bemerkten die Veränderungen im Video hingegen seltener. Dies deutet darauf hin, dass die Veränderungsblindheit durch die Aufmerksamkeit und das Gedächtnis der Teilnehmer für das Video beeinflusst wurde. [7]

Zwischen den CFMT+-Werten der Teilnehmer und den Ergebnissen zum Fahrerwechsel konnte eine signifikante positive Korrelation festgestellt werden. Ebenso gab es eine positive Korrelation zwischen der Trefferzuversicht beim SFCT und dem Test zur Veränderungsblindheit. Trotz dieser Korrelationen gibt es nur begrenzte Zusammenhänge zwischen den Variablen, was darauf hindeutet, dass der SFCT und der Veränderungsblindheitstest zur Überprüfung unterschiedlicher Fähigkeiten eingesetzt werden können. [7]

Davis et al. haben in dieser Studie festgestellt, dass Menschen mit überdurchschnittlicher Gesichtserkennungsfähigkeit, Erfahrung in der Überprüfung von CCTV-Aufnahmen und gezielter Vorübung zur Vertrautmachung mit den Zielakteuren bessere Leistungen bei der Identifizierung von Schauspielern in einem speziell erstellten SFCT-Video erbringen. Gleichzeitig wird ihre Anfälligkeit für Veränderungsblindheit gesenkt. Diese beiden Tests replizieren Aspekte der Polizeiarbeit, nämlich das Überprüfen von CCTV-Aufnahmen und die Suche nach Personen von Interesse in Menschenmengen bei Veranstaltungen. Wie von den Autoren vorhergesagt, erzielten die Polizeiexperten, die Vollzeit bei der Polizei in London als SR agieren, bei den meisten Tests bessere Ergebnisse als die anderen Gruppen. Aufgrund der geringen Anzahl von Polizeiexperten wurde jedoch nicht immer statistische Signifikanz erreicht. Die Streifenpolizisten, die als Polizei-Identifizierer arbeiten, schnitten besser ab als die Kontrollgruppe. [7]

Bei allen Teilnehmern aus allen Gruppen war bei korrekten Identifikationen von SFCT-Schauspielern

ein höheres Vertrauen vorhanden, während bei falsch positiven Identifikationen der Nebenakteure das Vertrauen geringer war. Dies stellt eine Beziehung zwischen Vertrauen und Genauigkeit her, die auch in der Forschung zu Augenzeugen beobachtet wird. [7]

Das SFCT besteht aus 11 Videoclips, in denen jeweils zwei, ein oder kein Schauspieler zu identifizieren war. Um den Test abzuschließen, benötigten die beiden Polizeigruppen mehr Zeit als die Kontrollgruppen. Dies könnte auf eine größere Sorgfalt hindeuten und teilweise die überlegene Leistung erklären. Unabhängig von der Anzahl der zu identifizierenden Schauspieler schnitten die Polizeigruppen besser ab als die Kontrollgruppen. [7]

Die vier Fotos, die die Teilnehmer zum Abgleichen mit den Schauspielern in den Clips bekamen, verbesserten wahrscheinlich den gleichzeitigen Abgleich von Gesichtern. Durch mehrere Bilder einer Person erhält der Betrachter mehr Identifikationsmerkmale als nur von einem einzelnen Bild. Sind beispielsweise Ganzkörperbilder dabei, kann dies die Identifizierung bei unklaren Gesichtern erleichtern. Je mehr sich das Erscheinungsbild der Schauspieler jedoch verändert hat, desto mehr nimmt die Leistung ab. Viele der Polizisten gaben an, dass der Test mit solchen Bildern eine typische Ermittlung widerspiegelt, was auf eine gute Testgültigkeit hinweist. [7]

Bevor der SFCT im Einsatz verwendet wird, sollte er noch einmal aktualisiert werden. Bisher kann man nicht gut zwischen einzelnen Leistungsträgern mit unterschiedlichen Fähigkeiten unterscheiden. Dies könnte durch zusätzliche Videoclips und Zielakteure erreicht werden. [7]

Die Teilnehmer, die an der Vorübung zum Vertrautmachen mit den im SFCT gezeigten Schauspielern teilnahmen, hatten im Vergleich zur Kontrollgruppe weniger falsch-positive Antworten und mehr korrekte Ablehnungen. Auf die Trefferquote hatte dies jedoch keinen Einfluss, obwohl das Vertrauen in die Treffer erhöht wurde. Dies könnte darauf zurückzuführen sein, dass tiefere soziale Prozesse die Aufmerksamkeit auf Hinweise lenken, die bei der Ablehnung unbeteiligter Personen halfen. Andere Faktoren wie Motivation, die sich aus der gemeinsamen Vorbereitung im Gegensatz zu den getrennten, untrainierten Kontrollgruppen ergeben, könnten ebenfalls die Leistung verbessert haben und sind unabhängig von der Gesichtsverarbeitung. Es scheint jedoch eine Obergrenze für den Vertrautheitseffekt zu geben, da die Teilnehmer der Polizei nicht von den vertrauten Kontrollpersonen übertroffen werden konnten. [7]

Den Fahrerwechsel und die Veränderung der Farbe des Fahrer-T-Shirts bemerkten mehr Polizisten als Kontrollpersonen. Das zeigt, dass die Polizisten weniger anfällig für Veränderungsblindheit im Vergleich zu den Kontrollpersonen sind. Die guten Ergebnisse lassen sich eventuell durch die durchschnittliche Bearbeitungszeit von 45 Minuten erklären. Statt zu ermüden, haben sich die Teilnehmer ans Analysieren gewöhnt, was ihre Fähigkeit zur Erkennung von Anomalien verbesserte. Die kurzen Videos, die in den meisten Studien zur Veränderungserkennung verwendet werden, bilden möglicherweise nicht die realen Bedingungen der Polizeiarbeit ab, bei der die Überprüfung von CCTV-Aufnahmen auch mehrere Tage in Anspruch nehmen kann. [7]

Die wahrgenommene Arbeitsbelastung der Teilnehmer während des SFCT wurde mithilfe des NASA-Task-Load-Index gemessen. Es wurden keine signifikanten Unterschiede zwischen den Gruppen in Bezug auf mentale Anforderungen, zeitliche Anforderungen, Aufwand oder Frustrationslevel festgestellt. Im Vergleich zu den Kontrollpersonen bewerteten die Polizeibeamten jedoch ihre Gesamtleistung deutlich höher, was auf ein stärkeres Gefühl des persönlichen Erfolgs hinweist. [7]

Die Teilnehmer haben auch einen Persönlichkeitstest ausgefüllt. Es konnten keine Zusammenhänge zwischen Extraversion und Gesichtsgedächtnis gefunden werden, ebenso wenig zwischen Extraversion und Neurotizismus hinsichtlich SFCT-Wachsamkeit oder visueller Suche. Die einzigen signifikanten Zusammenhänge betrafen die Beziehung zwischen den CFMT+-Werten der Kontrollgruppe

und Offenheit sowie zwischen der Zuversicht bei SFCT-Treffern und Offenheit bei allen Teilnehmern. Zukünftige Forschungen könnten untersuchen, welche Persönlichkeitsmerkmale als Auswahlkriterien für den Einsatz in polizeilichen CCTV-Überprüfungsoperationen verwendet werden könnten.[7]

Es gab einige Einschränkungen in den Tests, die sich möglicherweise auf den Ausgang der Studie ausgewirkt haben könnten. Zum einen betrifft dies die Art der Teilnehmer: Die Polizisten sind vorwiegend weiß, männlich und über 30 Jahre alt, während die Kontrollgruppe größtenteils aus weiblichen Teilnehmern unter 25 Jahren besteht und verschiedene Ethnien umfasst. Die Fähigkeit zur Gesichtserkennung erreicht in der Regel ihren Höhepunkt in den 30ern, und Gesichter der eigenen Ethnizität werden besser erkannt als Gesichter anderer Ethnizitäten. Diese Faktoren könnten zur überlegenen Leistung der Polizei in dieser Studie beigetragen haben. Allerdings neigen Frauen auch dazu, Männer in der Gesichtserkennung zu übertreffen, insbesondere bei Gesichtern ihres eigenen Geschlechts. [7]

Eine weitere Einschränkung ist, dass die Überprüfung von CCTV-Aufnahmen oft mehr erfordert als die einfache Suche nach Personen von Interesse. Es kann auch notwendig sein, die Bewegungen einer Person nachzuverfolgen, verschiedene Kameraaufnahmen zu beobachten oder anomale Verhaltensweisen zu identifizieren. Daher sollten Forscher untersuchen, welche Fähigkeiten die besten Ergebnisse bei der Identifikation durch CCTV-Aufnahmen liefern und Tests entwickeln, die geeignete Super Recognizer für diese Aufgaben identifizieren können. [7]

Als dritte Einschränkung könnte angenommen werden, dass bei den Polizisten eine intrinsische Motivation besteht, die bei der Kontrollgruppe möglicherweise fehlt. Dies könnte die Ergebnisse beeinflussen und nicht unbedingt Unterschiede in der Fähigkeit widerspiegeln. Andererseits gaben einige Polizisten an, die Tests nicht als herausfordernd empfunden zu haben und führten ihre schlechteren Leistungen auf Ermüdung zurück. [7]

In den Analysen übertrafen die Polizisten größtenteils die Teilnehmer aus den Kontrollgruppen, was dafür spricht, sie in die Gruppen der Polizeiexperten oder Identifizierer einzuordnen. Jedoch wurde der zuvor festgelegte CFMT+ Wert nur von einer Minderheit erreicht. Bei anderen Tests schnitt ein Großteil der Londoner Polizei besser ab als die Kontrollgruppen und erreichte hohe Schwellenwerte. Ein weiterer Faktor, warum der CFMT+ möglicherweise nicht die Identifikation von bekannten Verdächtigen bei der Polizei widerspiegelt, könnte sein, dass dieser hauptsächlich auf dem Kurzzeitgedächtnis für unbekannte Gesichter basiert. Die Fähigkeiten der Polizeiexperten, die regelmäßig unbekannte Verdächtige identifizieren, könnten sowohl auf der Fähigkeit zum gleichzeitigen Gesichtsabgleich als auch auf ihrem allgemeinen Gedächtnis basieren. Die Ergebnisse zeigten jedoch, dass einige von ihnen nicht als Super-Recognizer eingestuft wurden. Es war nicht bekannt, wie sie Mitglieder der Gruppe wurden, aber alle waren ehemals bei der Identifizierung der Polizei tätig, und diese werden hauptsächlich anhand ihrer Fähigkeit zur Identifizierung mehrerer bekannter Verdächtiger ausgewählt.[7]

Einige der Super-Recognizer, die hohe Werte beim CFMT+ erreichten, schnitten bei anderen Tests zur Gesichtsverarbeitung, wie dem gleichzeitigen Gesichtsabgleich, relativ schlecht ab. Dies wurde in der Studie als Hinweis auf mögliche Untergliederungen der Super-Recognizer interpretiert. Die ursprüngliche, kürzere Version des CFMT+ mit 72 Durchgängen wurde kritisiert, da sie keine gesichtsspezifischen Mechanismen ausreichend misst. Daher scheint die 102-Durchgänge-Version des CFMT+ einen besseren Indikator für SR-Fähigkeiten zu bieten. Dennoch ist weitere Forschung erforderlich, um Tests zu entwickeln, die möglicherweise besser geeignet sind, insbesondere um Kandidaten auszuschließen, die über eine schlechtere Gesichtsabgleichsfähigkeit verfügen. Diese

Fähigkeit ist für die Polizeiarbeit von wesentlicher Bedeutung. Vorgeschlagen werden Tests zum simultanen Gesichtsabgleich, Lang- und Kurzzeitgedächtnistests für Gesichter sowie Überprüfungs- und Suchaufgaben für CCTV, wie der SFCT. [7]

Zusammenfassend zeigt diese Arbeit, dass einige Polizeibeamte der Metropolitan Police Service herausragende Fähigkeiten in der Gesichtsverarbeitung besitzen. Es ist jedoch wichtig zu betonen, dass diese Fähigkeiten nicht bei allen Polizisten gleichermaßen ausgeprägt sind. Die Ergebnisse legen auch nahe, dass es problematisch sein könnte, Identifikationen von CCTV-Aufnahmen durch Polizeiexperten vor Gericht höher zu gewichten als die von anderen Zeugen. Fehlidentifikationen durch Polizisten waren nicht immer mit geringem Vertrauen verbunden, und Aufzeichnungen der MPS zeigen, dass auch Polizeiexperten gelegentlich Fehlidentifikationen vornehmen. Angesichts der oft schlechten Qualität von CCTV-Aufnahmen ist dies nicht überraschend, wird jedoch in der Regel schnell durch weitere Untersuchungen korrigiert. Dennoch könnten alle Polizisten anfällig für kognitive Verzerrungen und Bestätigungsfehler sein, die dazu führen können, dass Beweise im Einklang mit ihren Überzeugungen interpretiert werden. Diese Risiken könnten bei minderwertigen Bildern noch verstärkt werden. [7]

Der Metropolitan Police Service in London hat als erste Polizeibehörde weltweit eine spezielle Einheit für Super-Recognizer eingerichtet. Diese Einheit folgt den Verfahrensvorschriften des Police and Criminal Evidence Act (1984), Code D, die darauf abzielen, Fehlurteile zu minimieren. Um Risiken zu verringern, sollten wie bei allen forensischen Verfahren Verdächtigenidentifikationen unabhängig überprüft werden. Die Polizeiexperten und Identifizierer sind in der Lage, überdurchschnittlich viele Verdächtige auf CCTV-Aufnahmen zu erkennen. Dies könnte als Argument dienen, auch in anderen Rechtssystemen spezielle Einheiten für Super-Recognizer zu etablieren. Es muss jedoch sichergestellt werden, dass die Rechte der Verdächtigen stets gewahrt bleiben. [7]

3.2.6 Super-recognizers: From the lab to the world and back again

Die Arbeit von Meike Ramon, Anna Bobak und David White aus dem Jahr 2019 thematisiert das bislang unzureichende wissenschaftliche Verständnis bezüglich der überdurchschnittlichen Fähigkeit zur Gesichtsverarbeitung und die Herausforderungen bei der umfassenden praktischen Anwendung von Super-Recognizers (SRs). Die Autoren geben einen Überblick über den aktuellen Stand der Wissenschaft und weisen auf Mängel hin, die ihrer Meinung nach den Nutzen von SRs einschränken. Zum einen gibt es nur wenige verfügbare Studien, was zu unzureichendem Wissen über die Grundlagen der überdurchschnittlichen Gesichtsverarbeitung führt. Zum anderen fehlt ein klarer Überblick darüber, welche Aufgaben SRs übernehmen sollen und können und ob laborbasierte Tests tatsächlich die benötigten Fähigkeiten überprüfen. Die Autoren schlagen vor, dass die Praxis enger mit der Forschung zusammenarbeiten muss. Die Tests sollten explizit auf die Bedürfnisse der Anwendungsfelder von SRs angepasst werden. Eine kontinuierliche Rückkopplung zwischen Theorie und Praxis könnte dazu beitragen, dass die Theorie in der Praxis umgesetzt wird und daraus neue Erkenntnisse für die Theorie gewonnen werden. Dies würde auch eine bessere Auswahl der SRs ermöglichen. [24]

Die Autoren diskutierten über die Frage ob die Identifizierung überdurchschnittlicher Gesichtsverarbeitungsfähigkeiten wirklich eine Lösung für Probleme der realen Welt sind. [24]

In früheren Arbeiten, wie denen von Russell, Duchaine und Nakayama (2009), wurde das Konzept der Super-Recognizer eingeführt. Diese Personen mit überlegener Fähigkeit zur Gesichtsidentifikati-

on erzielten in Tests wie dem CFMT+, CFPT und BTWF höhere Punktzahlen als die Kontrollgruppe. Nach der Entdeckung dieser Menschen mit besonderen Fähigkeiten wurden in den meisten folgenden Studien untersucht, in welchen Aufgaben der Gesichts- oder Objekterkennung Super-Recognizer die Kontrollgruppen übertreffen könnten. Die in diesen Studien verwendeten Tests haben drei gemeinsame Aspekte, die darauf hindeuten, dass sie nicht allein zur Identifikation und Bewertung von Super-Recognizers geeignet sind. Erstens wurden die Tests ursprünglich zur Identifikation von Personen mit normalen bis schlechten Gesichtsverarbeitungsfähigkeiten entwickelt. Daher ist unklar, ob sie auch für überdurchschnittlich begabte Personen geeignet sind. [24]

Zweitens schneiden die SR als Gruppe in den Tests besser ab als nicht SR Gruppen, jedoch ist nicht jede Einzelleistung der SR besser als der Durchschnitt. Die SR können auch in manchen Tests überdurchschnittlich sein und in anderen nur durchschnittlich gut sein. Im Ganzen mangelt es noch an angemessenen diagnostischen Kriterien zur Bestimmung der Gesichtsverarbeitungsfähigkeiten. [24] Drittens sind die verwendeten Tests nicht drauf angepasst welche Aufgaben die SR später ausüben sollen. Dementsprechend sind sie nicht repräsentativ für jede Art an operativen Aufgaben, die auf sie in einer professionellen Bereitstellung treffen könnten. Zum Beispiel sind die Gesichtsbilder in den Tests klassisch unter kontrollierten Umweltbedingungen aufgenommen und werden mit Manipulationen verändert sie so nicht in der natürlichen Umgebung vorkommen würden. Dadurch umfassen diese Aufgaben in den Tests zur Erfassung der Gesichtsverarbeitungsfähigkeit, nicht die Herausforderungen einer Identifikation in einer realen Umgebung. [24]

In der Arbeit wird vorgeschlagen ein Rahmenwerk zu erstellen, in denen die Leistungen mit passenden Tests bewertet werden die direkt auf die gewollten Fähigkeiten zugeschnitten sind. Da SRs in vielen verschiedenen Aufgaben in der Realität eingesetzt werden könnten, müssen die Auswahlkriterien individuell auf diese bestimmte Fähigkeit die benötigt wird, angepasst werden. [24]

Das Ziel der Identifikation und des Einsatzes von SRs in praktische Anwendungen besteht darin die Zuverlässigkeit des Menschen bei realen Aufgaben mit Gesichtsidentifikationen zu verbessern. Solche Verbesserungen können die Gesellschaft sicherer und gerechter machen, indem beispielsweise Terroranschläge, Identitätsbetrug und Fehltritte vor Gericht verhindert werden. Der letztendliche Erfolg einer solchen Rekrutierungsmaßnahme hängt jedoch davon ab, wie gut sie mit der Leistung in realen Aufgaben korreliert. [24]

Die Autoren führten eine Monte-Carlo-Simulation durch, um die Beziehung zwischen einer bestimmten Korrelation und den Genauigkeitsgewinnen zu charakterisieren, die in einer hypothetischen realen Aufgabe erwartet werden können. Die Simulation bietet einen beispielhaften Leitfaden für das Ausmaß der Leistungssteigerung in der realen Welt, das für jede gegebene Korrelation mit einer ideal kalibrierten Rekrutierungsmaßnahme erwartet werden kann. Für diese Simulation wurden normale bivariate Verteilungen mit zwei Dimensionen erzeugt, die willkürlich als Prozentsatz der richtigen Antworten im 'Rekrutierungstest' und der 'realen Aufgabe' bezeichnet wurden. Zur Vereinfachung reichten die Werte der Variablen von 50% bis 100%, was die gesamte Bandbreite der Leistung bei einer Zwei-Alternativen-Aufgabe (von Zufallsniveau bis perfekte Genauigkeit) repräsentiert. Unter den simulierten Bedingungen konnten die Tests optimal auf die Skala kalibriert werden. Die Mittelwerte lagen im Zentrum (75%), und die Verteilungsparameter wurden so eingestellt, dass sie die gesamte Genauigkeitsspanne abdeckten, was eine Situation darstellt, die in der Realität unwahrscheinlich ist. Mit diesem Ansatz wurden 100 Rekrutierungsprozesse simuliert, bei denen jeweils 1000 Teilnehmer beteiligt waren. Die Korrelation zwischen Rekrutierungstest und realer Genauigkeit variierte zufällig. Dieser hypothetische Rekrutierungsprozess wurde als virtuelles Best-Case-Szenario modelliert, mit perfekt kalibrierten Messungen und einer sehr großen Auswahl an Kandidaten.[24]

Dieser Ansatz ermöglichte, die erwarteten Leistungsgewinne für jedes Korrelationsniveau darzustel-

len. Es wurde die durchschnittliche Leistung in der realen Welt von Gruppen, die entweder größer einer Standardabweichung oder größer zwei Standardabweichungen im Rekrutierungstest erzielten. Der Unterschied in der Leistung zwischen den ausgewählten und nicht ausgewählten Gruppen zeigt den geschätzten Nutzen der Auswahlkriterien. Diese Kriterien spiegeln die strengen Vorgaben aus der wissenschaftlichen Literatur wider und berücksichtigen, dass viele Organisationen möglicherweise weniger strenge Kriterien anwenden, um größere Gruppen von Individuen mit anderen Auswahlmethoden auszuwählen. [24]

Die Autoren sind der Meinung, dass die Daten für Entscheidungsträger informativ sind, da sie einen Leitfaden für den erwarteten Nutzen im realen Kontext bieten, wenn der Grad der Korrelation zwischen einem Auswahlkriterium und der Leistung bei einer realen Aufgabe bekannt ist. Obwohl die Korrelation zwischen laborbasierten und realen Messungen oft schwer zu schätzen ist, ist es trotzdem wichtig, wenn möglich, diese zu quantifizieren. [24] Was bedeutet dies für die Auswahl von Spezialteams basierend auf der individuellen Fähigkeit zur Gesichtserkennung? Eine Auswahl von Personen, die mehr als 2 Standardabweichungen über dem Durchschnitt bei einem laborbasierten Auswahlverfahren liegen, bei einer Labor-zu-Welt-Korrelation von .5 zu einem realen Leistungszuwachs von etwa 12 % führen. In der Praxis erfolgt die Auswahl jedoch wahrscheinlich aus kleineren Gruppen mit potenziellen Rekruten, realistischer wäre dafür weniger strenge Kriterien zur Auswahl von Personen überdurchschnittlicher Fähigkeiten anzuwenden. Beispielsweise wenn für den Beruf eines Passbeamten, ein Auswahlverfahren stattfindet bei dem 100 Bewerber getestet werden und eine Standardabweichungen von 2 als Kriterium verwendet wird, würde es im Durchschnitt nur zwei bis drei erfolgreiche Bewerber geben. Wenn die Kriterien auf 1 Standardabweichung über dem Durchschnitt gesenkt werden, führt es einer Verbesserung von 8 %. Das kann für die Praxis sinnvoll sein. Die Auswahl allein kann die Fehlerquoten nicht vollständig auslöschen, aber sie können um 32% reduziert werden. Das unterstützt die Entwicklung von Strategien zur Verbesserung der Gesichtserkennung in realen Umgebungen. [24]

Zusätzliche Verbesserungen der Leistungen können beispielsweise die Antworten mehrerer Personen zusammengefasst werden. Diese Verbesserung zusammen mit der Verbesserung durch die Rekrutierungsstrategie, addieren sich. Daher scheint der beste Ansatz eine Kombination aus effektiven Lösungen zu sein, um präzise Identifikationssysteme zu schaffen. Zum Beispiel die Kombination modernster Algorithmen mit leistungsstarken Menschen. [24]

Die potenziellen Vorteile des Einsatzes von Super-Recognizern hängen letztlich davon ab, wie gut die Rekrutierungstests mit den realen Aufgaben korrelieren, die sie später ausführen müssen. Eine deutliche Verbesserung dieser Korrelation ist erforderlich, bevor Auswahlmaßnahmen allein dazu beitragen können, das Problem der fehleranfälligen Gesichtserkennung zu lösen. Um zu bewerten, ob Auswahlprozesse die tatsächliche Leistung verbessern, ist es notwendig, die Ergebnisse empirisch entwickelter Tests mit der Leistung bei realen Aufgaben zu verknüpfen. Derzeit gibt es jedoch keinen Feedback-Loop. Die Spezialisten werden für reale Aufgaben eingesetzt, oft basierend auf im Labor entwickelten Auswahltests, ohne dass ihre operative Effizienz kontinuierlich und systematisch überprüft werden. Dies ist ein entscheidender Mangel, da es die Entwicklung von Tests behindert. [24]

Der allgemeine Prozess der Gesichtskognition, umfasst eine Reihe von Teilprozessen wie Gesichtserkennung, -diskriminierung, -wiedererkennung und -identifikation. Um im Labor entwickelte, effektive Messmethoden für überlegene Gesichtserkennung zu schaffen, die für praktische Anwendungen relevant sind, muss eine passende Zuordnung zwischen den im Labor gemessenen kognitiven Teil-

prozessen und den in der realen Welt benötigten Prozessen hergestellt werden. Dies herauszufinden stellt eine große Herausforderung dar, da jede reale Aufgabe wahrscheinlich auf unterschiedlichen Teilprozessen basiert.[24]

In der Regel werden Experimente entworfen, um ein oder mehrere Teilprozesse zu untersuchen. Eine direkte Eins-zu-eins Testung existiert selten. Es kann verschiedene Experimente geben, die denselben Prozess mit unterschiedlicher Effizienz messen. Und anderen wiederum messen mehrere Prozesse, aber können nicht unbedingt alle existierenden Teilprozesse erfassen. Ein Gesichtserkennungsexperiment erfasst z.B. die Fähigkeit, das Vorhandensein eines Gesichts zu erkennen, zwischen Gesichtern zu unterscheiden und zu erkennen, dass diese Person schon einmal gesehen wurde. Eine Gesichtsidentifikationsaufgabe würde all diese Teilprozesse und dazu die Fähigkeit beinhalten und genau sagen zu können wer die Person ist. [24]

Es wird jedoch umso schwieriger, je mehr Teilprozesse gleichzeitig gemessen werden, die Beiträge jedes einzelnen Teilprozesses zu kontrollieren und zu bestimmen. Zum Beispiel könnte eine überlegene Leistung in einer Gesichtsidentifikationsaufgabe, entweder auf einer gesteigerten Fähigkeit zur Unterscheidung oder Erkennung von Gesichtern basieren oder zum Abrufen semantischer Informationen. [24] In den zu der Zeit bisher veröffentlichten begrenzten Studien zu Super-Recognizern wurden die meisten mit laborentwickelten Experimenten identifiziert, die einen oder mehrere Aspekte der Verarbeitung von Gesichtsidentitäten messen. Für den Einsatz in der realen Welt ist es wichtig zu wissen wie die Beziehung zwischen den Teilprozessen und den verwendeten Messmethoden ist. Super-Recognizer sollen in operativen Umgebungen oft vielfältige Aufgaben ausführen, die von der Erkennung vertrauter Gesichter über die Unterscheidung unbekannter Gesichter bis hin zu anspruchsvollen visuellen Suchaufgaben umfassen können. Da laborbasierte Tests möglicherweise nicht die im realen Kontext relevante Leistung vorhersagen, sollten Methoden zur Identifizierung von SRs für praktische Anwendungen solche Teilprozesse bewerten, die den spezifischen beruflichen Anforderungen entsprechen. [24]

Die überlegene Fähigkeit zur Gesichtserkennung ist kein einheitliches Phänomen, wie heterogene Leistungsmuster in verschiedenen Tests zeigen. Bei realen Aufgaben wie der Überprüfung von CCTV-Aufnahmen wird das Problem deutlicher: Die Fähigkeit, Personen anhand von Körpermerkmalen zu identifizieren, korreliert nicht mit der Leistung bei Gesichtserkennungsaufgaben, was darauf hinweist, dass diese Aufgaben nicht ausreichen, um alle für den operativen Einsatz relevanten Fähigkeiten zu erfassen. Die Teilprozesse, die an einer bestimmten Aufgabe beteiligt sind, sollten sorgfältig berücksichtigt werden sowie präzise und geeignete Terminologie im Kontext der Messung der Gesichtserkennungsfähigkeit verwendet werden. Ob ein laborbasierter Test die relevante Fähigkeit für reale Aufgaben erfassen kann, wird davon bestimmt, ob beides auf ähnlichen Teilprozessen beruht. Aufgrund der unterschiedlichen Einsatzgebiete von Super-Recognizern ist es unwahrscheinlich, dass ein einzelner laborbasierter Test ausreicht, um sie zu identifizieren. Dies hat direkte Konsequenzen für die Bewertung bei der Rekrutierung von Super-Recognizern sowie für das Verständnis ihrer zugrunde liegenden Fähigkeiten.[24]

Damit der präzise Identifikation von Super-Recognizern für den realen Einsatz nachgekommen werden kann, muss eine Übereinstimmung zwischen hypothesengesteuerter Forschung und zielorientierter Praxis gewährleistet werden. Dafür müssen die Ziele der Anwender bestimmt werden, die wissenschaftliche Ansätze zur Verbesserung des Verständnisses der Gesichtserkennung antreiben. Aufgabenanalytische Ansätze in professionellen Umgebungen verbinden reale Ziele, Aufgaben oder Systeme mit den zugrunde liegenden kognitiven Prozessen der Leistung. Diese Techniken werden häufig von angewandten Forschern genutzt, um die Gestaltung von Auswahl-, Trainings- oder Orga-

nisationsprozessen zu optimieren und haben sich als äußerst nützlich erwiesen. Solche Praktiken könnten einen ähnlichen Zweck in der Untersuchung der überlegenen Gesichtserkennung erfüllen, indem sie das Maß der Korrelation zwischen Auswahlmaßnahmen und realen Aufgaben verbessern. [24]

Dieser fundamentale erste Schritt wurde in der Forschung zu Super-Recognizern weggelassen. Es wurde angenommen, dass die entwickelten laborbasierten Tests ausreichen, um Personen auszuwählen und diese dann in der realen Welt gut abschneiden. Solche Aufgaben können oft sehr komplexe und vielfältige Herausforderungen beinhalten die, zu der Gesichtsverarbeitung noch oben drauf kommen. Wie beispielsweise die Verwendung mehrerer Identitätsmerkmale, die nicht auf das Gesicht beschränkt sind. Daher sollte der Ausgangspunkt für die Auswahltests eine Analyse der Aufgaben die erledigt werden sollen sein. [24]

Die Autoren schlagen einen Ansatz vor, wie bei der Entwicklung von Bewertungsmaßnahmen vorgegangen werden sollte. Zu Beginn muss eine effektive Aufgabenanalyse angefertigt werden. Sind die relevanten Teilprozesse bekannt, können dazu Experimente entwickelt werden, die das Verhalten bewerten. Dabei sollten die Leistungen beobachtet werden und gegebenenfalls Anpassungen getroffen werden oder Grenzen aufzeigen. Es muss überprüft werden ob die Tests wirklich die unterschiedlichen Teilprozesse umfassen. Ergründen welche Faktoren für unerwartete Beobachtungen verantwortlich sind. Diese Vorgehensweise kann dafür sorgen, dass experimentelle Aufgaben angemessen entwickelt und in Übereinstimmung mit den realen Aufgaben optimiert werden. [24]

In diesem Artikel wurden drei Hauptlücken im Verständnis von Super-Recognizern identifiziert:

- Übereinstimmung zwischen laborbasierten Tests und der Leistung in der realen Welt
- Bandbreite der Aufgaben, die von SRs erwartet werden
- Teilprozesse der Gesichtserkennung, die den realen Aufgaben zugrunde liegen

Um diese Lücken muss die Forschung enger mit der Praxis zusammenarbeiten. Die Autoren gehen im folgenden darauf ein wie das funktionieren kann und was noch verbessert werden sollte. [24]

Es wird ein Wissensentwicklungskreislauf zwischen dem Labor und der realen Welt vorgeschlagen. Wissen aus dem Labor entsteht durch Wissenschaftler die Studien entwerfen, um die zugrunde liegenden Mechanismen und die Rahmenbedingungen der überlegenen Leistung von Super-Recognizern zu verstehen. Dieses Wissen dient als Grundlage für Verfahren, die von Praktikern genutzt werden können, um SRs für Aufgaben in der realen Welt auszuwählen, die potenziellen Vorteile des Einsatzes von SRs in ihrer Organisation zu bewerten und Richtlinien für die Interpretation von Beweisen durch SRs vor Gericht zu entwickeln. Wichtig dabei ist, gewonnenes Wissen von der realen Welt wieder zurück zum Labor kommt. Sodass wissenschaftliche Verfahren an die realen Gegebenheiten angepasst werden, zum Beispiel durch die Einbeziehung reale Aufgaben zur Auswahl von Personen, die in der Praxis gut abschneiden. [24]

Um diese Maßnahme auch praktisch umzusetzen sind in den letzten Jahren drei Hauptmechanismen entstanden. Erstens wurden wissenschaftliche Arbeitsgruppen eingerichtet, die Richtlinien erstellen um die Effektivität und Zuverlässigkeit von Gesichtserkennung zu verbessern. Auch Akademiker haben begonnen, sich in diesen Gruppen zu engagieren. Zweitens wurden von Akademikern geleitete Treffen gemeinsam von Psychologen, Informatikern, Forensikern, Anwälten, Polizisten und Mitarbeitern verschiedener Regierungsbehörden besucht. Drittens waren gemeinsame Projekte zwischen Wissenschaftlern und Praktikern maßgeblich dafür, die anfänglichen Laborarbeiten in praktische Anwendungen zu überführen. Diese Projekte sollen, die Leistung in realen Anwendungen zu optimieren und die Genauigkeit von Gesichtserkennungsexperten im Vergleich zur allgemeinen Bevölkerung und

zu Super-Recognizern bewerten. Kooperierende Polizeibehörden haben zudem begonnen, Fragen zur Auswahl von SRs, zur Häufigkeit von SRs basierend auf der Identifikation bei realen Aufgaben in großen beruflichen Gruppen und zum Einsatz von SRs und forensischen Gesichtsexperten in strafrechtlichen Ermittlungen zu behandeln.[24]

Durch diese Entwicklungen wurde die Arbeit von Forschern und Praktikern auf sinnvolle Weise verknüpft und es wurde ein Netzwerk gefördert, das die Grundlage für zukünftige transnationale Forschung bilden kann. Bisher gibt es nur sehr wenige Kooperationen zwischen Praktikern und Akademikern, obwohl beiden Gruppen Interesse an dieser Thematik haben. So könnte es dazu kommen, dass die Nachfrage nach SRs in der Praxis viel höher ist als das wissenschaftliche Verständnis ihrer Fähigkeiten bisher ergründen kann. Diese Nachfrage war in den Jahren zuvor deutlich gestiegen und bei verschiedenen Stellen werden Mitgliedschaften, Akkreditierungen oder berufliche Möglichkeiten angeboten. Diese Organisationen die so etwas anbieten sollten ihre Auswahlmethoden von Wissenschaftlern überprüfen lassen, indem sie ihre Akkreditierungskriterien transparent und öffentlich zugänglich machen. Auch müssen jegliche Behauptungen über SRs auf Daten von Experten basieren und nicht nur lose Behauptungen sein. Solche Transparenz würde die wissenschaftliche Forschung und Praxis in angewandten Umgebungen fördern.[24]

Um die Rolle der SRs im Rechtssystem zu etablieren ist ein koordinierter Ansatz ist auch hilfreich. Bisher sind keine Fälle bekannt, in denen SRs als Sachverständige für Identifikation vor Gericht ausgesagt haben, aber im Vereinigten Königreich haben sie als reguläre Polizeizeugen Beweise geliefert. Es wurde vorgeschlagen, dass es besser wäre SRs vor Gericht als Sachverständige einzusetzen als den derzeitigen Gesichtserkennungsexperten. In diesem Kontext ist entscheidend für die Expertise vor Gericht, ob SRs bei realen Aufgaben überlegen sind. Eine neuere Studie berichtete, dass Gruppen von SRs und Gruppen von professionellen forensischen Gesichtserkennungsexperten bei der Unterscheidung von Frontalbildern keine signifikanten Unterschiede aufwiesen. Dieses Ergebnis deutet darauf hin, dass SRs in der Lage sein könnten, Beweise in Gerichtsverfahren zu liefern, die qualitativ mit denen offiziell ausgebildeter Fachleute vergleichbar sind. [24]

Jüngere Forschungen haben bereits angefangen Tests für die realen Aufgabenbereiche der SR bei Gesichtserkennung zu entwickeln. Beispielsweise zum Abgleichen von bildbasierten Gesichtserinnerungen mit Videoaufnahmen, das Suchen nach Identitäten in Menschenmengen und der Gesichtsbgleich mit Bedingungen, die den Anforderungen in praktischen Einsätzen ähneln. Wie in dieser Arbeit erklärt, ist es wichtig, das Ausmaß zu bestimmen, in dem die Leistung bei laborbasierten Tests wahrscheinlich auf reale Aufgaben übertragbar ist. Die Autoren betonten, dass für die weitere Forschung in diesem Bereich ein zentraler Schritt darin besteht, sich auf eine einheitliche Methode zur Messung und Definition von außergewöhnlichen Fähigkeiten in der Gesichtserkennung zu einigen. [24]

Aus akademischer Sicht stimmen die Autoren dieser Überlegung zu, das Verfahren so angepasst werden können, dass sie die geeignetsten Personen für den jeweiligen Teilprozessen der Gesichtserkennung identifizieren kann. Wie in der Arbeit jedoch thematisiert, ist es unwahrscheinlich das Allzweck Maßnahmen zum langfristigen Einsatz in der realen Welt entstehen würde. Die vielfältigen Anforderungen der realen Aufgabengebiete der Gesichtserkennung machen es sehr unwahrscheinlich, dass ein Standardansatz zur Identifizierung von Super-Recognizern für angewandte Umgebungen etabliert werden könnte. Darüber hinaus unterliegen diese Maßnahmen regelmäßiger Überprüfungen und Aktualisierungen, damit sie den ständig ändernden realen Anforderungen gerecht werden. [24]

Die Autoren schlagen daher vor, dass das Forschungsfeld das Thema der Identifizierung von Super-Recognizern systematisch und koordiniert angehen sollte. Einerseits besteht die Möglichkeit, dass unabhängige Forschungsgruppen jeweils eigene Tests entwickeln, um spezifische betriebliche Aufgaben nachzustellen. Dies birgt jedoch die Gefahr, dass Tests zu stark auf bestimmte Aufgaben

zugeschnitten werden, was das Ziel behindern würde, universelle Kriterien für herausragende Fähigkeiten zur Gesichtserkennung zu definieren. Andererseits ist klar, dass die derzeit verfügbaren laborbasierten Tests die Vielfalt der realen Aufgaben nicht ausreichend abbilden. Daher könnte die einzige Lösung für dieses Problem darin bestehen, dass Wissenschaftler und Praktiker einen Konsens darüber erzielen, in welchen Rollen Super-Recognizer nützlich sein können, und sich auf die besten Maßnahmen einigen, um die vielversprechendsten Personen für diese Aufgaben zu identifizieren. [24]

3.2.7 Face recognition in police officers: Who fits the bill?

Die Autoren Jeffrey D. Nador, Michael Vomland, Markus M. Thielgen und Meike Ramon haben im Jahr 2022 eine Arbeit veröffentlicht in der sie 3 Tests für Gesichtsidentitätsverarbeitung mit zwei Arbeitsproben zu vergleichen. [22]

Zu dem Beruf eines Polizeibeamten kann es dazugehören, ihre Gesichtsverarbeitungsfähigkeit (engl. Face Identity Processing (FIP)) einzusetzen, um Täter und Opfer zu identifizieren. Dabei ist präzise Leistung von großer Bedeutung, um beispielsweise falsche Verdächtigungen oder ungelöste Fälle von vermissten Personen zu vermeiden. Dagegen haben Polizeibehörden Maßnahmen ergriffen, um hohe Leistungen sicherzustellen. Solche Leistungen sind sehr komplex und brauchen mehr als nur Genauigkeit. Die Polizeibeamten müssen oft über Stunden lang die Leistung vollbringen und mit einer Vielzahl an eingehenden Daten mithalten müssen. Dafür gibt es viele verschiedene Möglichkeiten die einzelnen Teile der hohen Leistungen zu verbessern, jedoch ist es unwahrscheinlich, dass eine Lösung für alle Kontexte hilfreich ist. Die Arten an Lösungen können jedoch in zwei Sorten eingeteilt werden, diese die die FIP-Fähigkeit trainieren und die die Auswahl der Beamten spezifischer machen. [22]

Da eine experimentelle Simulation von den realen Aufgabengebieten Sicherheitsfachleute sehr schwierig ist, haben die Forscher eher die Möglichkeit untersucht Fachleute zu schulen. Solche Studien richten sich in der Regel an Passkontrollbeamte und Forensiker, wobei untersucht wird, ob man diese bis zum Expertenstatus schulen könnte. Andere bestehen darauf die Leistungsstandards durch Rekrutierungsmaßnahmen aufrecht zu erhalten. Die Mehrheit der Forschungen zu diesem Thema ergaben, dass FIP-Fähigkeiten nicht besonders trainierbar sind. Zum Beispiel haben Passkontrollbeamte mit mehr als 20 Jahren Berufserfahrung ähnliche FIP-Fähigkeiten wie ihre frischen Kollegen. Auch Beamte mit umfangreicher Erfahrung in der Identifikation schneiden nicht besser ab als Zivilisten bei der Erkennung von Gesichtsidentitäten aus Videos von geringer Qualität. [22]

Da dieser Ansatz wenig vielversprechend wirkt, ist der Ansatz zur Entwicklung von Auswahlverfahren zur Identifizierung hoher Fähigkeiten eine bevorzugte Alternative. Den Autoren ist zu der Zeit nur die Polizeibehörde in Berlin bekannt die einen maßgeschneiderten Test entwickelt haben, um Beamte mit überlegenen FIP-Fähigkeiten zu identifizieren. Frühere Studien haben untersucht, wie verschiedene laborbasierte Tests zur Gesichtsverarbeitungsfähigkeit (FIP) miteinander zusammenhängen. Diese Studien haben gezeigt, dass es auf Gruppenebene bestenfalls moderate Korrelationen gibt. Auf individueller Ebene variieren die Ergebnisse jedoch stark zwischen den Tests. Es gibt drei Hauptgründe die diese fehlende Übereinstimmung erklären können.[22]

Erstens unterscheiden sich die verfügbaren laborbasierten Tests stark in Bezug auf ihre Durchführung und die verwendeten Reize oder Leistungsmaße. Selbst wenn zwei Tests die Genauigkeit als Leistungsmaß verwenden, können sie sich dennoch in ihrem Ansatz und Fokus unterscheiden. Zum Beispiel messen sowohl der Cambridge Face Perception Test (CFPT) als auch die lange Version des Cambridge Face Memory Test (CFMT+) die Genauigkeit. Der CFPT untersucht jedoch, wie die

Beurteilung von Ähnlichkeiten durch die Umkehrung des Stimulus beeinflusst wird. Im Gegensatz dazu bewertet der CFMT+ die Fähigkeit, zuvor gelernte Gesichter unter zunehmend schwierigeren Bedingungen wiederzuerkennen. [22]

Zweitens die laborbasierten Tests für die Fähigkeit zur Gesichtsverarbeitung haben ihre Grenzen, besonders wenn es darum geht, die extrem hohen und niedrigen Leistungen innerhalb der Bevölkerung zu erfassen. Die meisten dieser Tests, wie der CFPT und CFMT, wurden entwickelt, um Menschen mit Prosopagnosie zu diagnostizieren, einer Störung, die die Gesichtserkennung beeinträchtigt. Daher sind sie gut darin, Unterschiede am unteren Ende des Leistungsspektrums zu erkennen. Allerdings stoßen diese Tests an ihre Grenzen, wenn es darum geht, die besten Leistungen in der Allgemeinbevölkerung zu messen. Sie erreichen schnell eine Leistungsobergrenze, bei der sie nicht mehr zwischen den besten Leistungen unterscheiden können. Umgekehrt sind anspruchsvollere Tests, wie der Yearbook Test, die auf Personen mit sehr hohen FIP-Fähigkeiten abzielen, oft zu schwierig für Personen mit Prosopagnosie. Die meisten bisher entwickelten laborbasierten Tests sind nicht empfindlich genug, um auf die gesamte Bandbreite der FIP-Variationen zu reagieren. Die Lösung besteht darin, mehrere Tests zu verwenden, die verschiedene Aspekte und Schwierigkeitsgrade der FIP-Fähigkeit abdecken. Dadurch kann man ein umfassenderes Bild der individuellen Unterschiede in der Gesichtserkennung erhalten. [22]

Drittens könnte die (In-)Konsistenz des Verhaltens bei verschiedenen Tests eine grundlegende Eigenschaft der Fähigkeit zur Gesichtsverarbeitung sein. Neuere Studien zeigen, dass Super-Recognizer, also Menschen mit außergewöhnlicher FIP-Fähigkeit, eine bemerkenswerte Verhaltenskonsistenz über verschiedene Tests hinweg aufweisen. Diese Konsistenz unterscheidet sie von durchschnittlichen Personen, obwohl beide Gruppen grundsätzlich dieselben Informationen zur Gesichtserkennung nutzen. Dies bedeutet, dass Konsistenz nicht absolut ist. Vielmehr variiert sie in ähnlicher Weise wie die allgemeine FIP-Fähigkeit bedeutet, je besser die FIP-Fähigkeiten sind, gemessen an verschiedenen Tests, desto höher ist die Konsistenz innerhalb einer einzelnen FIP-bezogenen Aufgabe. [22]

In der Praxis wird die Leistung häufig durch verschiedene Aspekte des Verhaltens gemessen, wie zum Beispiel Genauigkeit und, wenn überhaupt, Reaktionszeit. Diese Messungen können zeigen, wie gut jemand im Allgemeinen arbeitet, aber sie erfassen nicht immer alle wichtigen Aspekte der tatsächlichen Arbeit. Bei der Polizeiarbeit ist es nicht nur wichtig, genaue Ergebnisse zu liefern, sondern dies auch effizient und konstant zu tun. [22]

Effizienz bedeutet, dass jemand genaue Ergebnisse in relativ kurzer Zeit liefern kann. Konsistenz bedeutet, dass diese Person über einen längeren Zeitraum hinweg immer eine ähnliche Leistung zeigt. Zwei Polizeibeamte könnten zum Beispiel beide in der Lage sein, ihre Aufgaben effizient zu erledigen. Doch während einer von ihnen vielleicht manchmal sehr schnell und ungenau arbeitet und andere Male langsam, aber präzise, könnte der andere Beamte gleichmäßig schnelle und präzise Ergebnisse liefern. Der Beamte, der konstante Leistung zeigt, erreicht eine höhere Konsistenz, was oft von größerem Wert ist. [22]

Nach dem Terroranschlag auf den Breitscheidplatz-Weihnachtsmarkt 2016 musste die Polizei Berlin intensiv nach dem flüchtigen Verdächtigen suchen. Dabei waren die Fähigkeiten der Beamten zur Gesichtserkennung entscheidend, um den Verdächtigen auf Überwachungsbildern zu finden. [22]

In einer solchen Situation ist es entscheidend, dass die Beamten sowohl schnell als auch genau arbeiten. Hohe Genauigkeit ist wichtig, um sicherzustellen, dass der Verdächtige richtig identifiziert wird. Aber da es viele Überwachungsaufnahmen zu durchsuchen gibt und die Zeit begrenzt ist, um den Verdächtigen zu fassen, ist es ebenso wichtig, dass dies schnell geschieht. Die Beamten mussten also nicht nur präzise arbeiten, sondern auch zügig, um viele relevante Informationen in kurzer

Zeit zu sammeln.[22]

Je schneller und genauer die Beamten die Aufnahmen überprüfen können, desto mehr Details können sie über den Weg des Verdächtigen herausfinden. Das erhöht das Vertrauen in die Lokalisierung des Verdächtigen an bestimmten Orten und verbessert die Chancen, ihn rechtzeitig zu fassen. Daher ist es von größter Bedeutung, Beamte zu haben, die diese Aufgabe effizient und verlässlich ausführen können. [22]

Antwortgenauigkeit und -geschwindigkeit (also Reaktionszeit) können zu einer einzigen Kennzahl zusammengefasst werden, indem man die Reaktionszeit durch die Genauigkeit teilt. Diese Kennzahl, die als Umkehr-Effizienz-Wert (IE-Scores) bezeichnet wird, zeigt, wie lange es dauert, eine korrekte Antwort zu geben. Je niedriger der Wert, desto effizienter arbeitet eine Person, weil sie schneller und genauer arbeitet. Wenn also sowohl Schnelligkeit als auch Genauigkeit wichtig sind, sollten Beamte, die diese Aufgabe übernehmen, möglichst niedrige IE-Scores haben, da dies auf hohe Effizienz hinweist.[22]

Der IE-Score kann durch unterschiedliche Kombinationen von Genauigkeit und Reaktionszeit beeinflusst werden. Zum Beispiel könnte ein Beamter einen niedrigen IE-Score haben, weil er sehr genau ist, aber lange braucht, um Antworten zu geben. Ein anderer könnte denselben niedrigen IE-Score haben, weil er sowohl schnell als auch genau ist. Wichtiger ist jedoch, wie konstant oder variabel ihre Leistung ist. Zwei Beamte mit dem gleichen IE-Score könnten sich unterscheiden, je nachdem, wie stabil ihre Leistung ist, das heißt, wie gleichmäßig und zuverlässig sie Antworten liefern. [22]

Die Polizei musste die Suche nach dem Verdächtigen vom Breitscheidplatz aktiv verfolgen. Wenn die zuständigen Beamten im Laufe der Zeit in ihrer Suche nachgelassen hätten, hätte der Verdächtige möglicherweise entkommen können. Deshalb wäre es entscheidend, dass die Beamten in der Lage sind, ihre Leistung konstant auf einem hohen Niveau zu halten, um den Verdächtigen erfolgreich zu finden. Konsistenz in der Leistung ist ein wichtiger Indikator für außergewöhnliche Fähigkeiten in der Gesichtserkennung, wie aktuelle Studien zu Super-Recognizern zeigen. Super-Recognizer haben die Fähigkeit, Gesichter konstant und zuverlässig zu erkennen, was zu präziseren und stabileren Ergebnissen führt. Im Vergleich dazu haben gewöhnliche Personen zwar ähnliche Fähigkeiten, aber ihre Leistung schwankt stärker und ist weniger zuverlässig. [22]

Im Fall des Anschlags am Breitscheidplatz wäre es ideal, wenn die Beamten nicht nur sehr effizient arbeiten (d.h. geringe IE-Werte haben), sondern auch konstant gute Leistungen zeigen. Dies ist besonders wichtig, da ihre Untersuchung möglicherweise viele Stunden damit verbringen könnte, Überwachungsdaten nach dem Verdächtigen zu durchsuchen. In diesem Prozess könnten sie Lücken in seiner Route entdecken, falls sie ihn übersehen. Zudem könnten falsche Identifikationen des Verdächtigen zu Verwirrungen führen, indem sie fälschlicherweise alternative Routen aufzeigen. Dies könnte dazu führen, dass die Beamten Beweise von irrelevanten Orten überprüfen. Diese beiden möglichen Szenarien verdeutlichen, dass sowohl die Effizienz als auch die Konsistenz der Gesichtserkennungsfähigkeiten der Beamten entscheidend für die Effektivität des Überwachungsmaterials als Ermittlungswerkzeug sind. [22]

Selbst bei routinemäßiger Polizeiarbeit sind Effizienz und Konsistenz entscheidend, insbesondere bei der Auswertung von Überwachungsdaten, die von Tatorten oder aus dem (Dark) Web gesammelt wurden. Wenn Straftäter Verbrechen an verschiedenen, weit auseinanderliegenden Orten begehen, kann der Abgleich von Überwachungsaufnahmen zwischen den Fällen helfen, festzustellen, ob die Verbrechen von derselben oder verschiedenen Personen begangen wurden. Dies erfordert, dass die Beamten schnell und genau viele Datensätze durchsuchen und dabei alle relevanten Identifikationsinformationen über die Täter behalten, ohne ihre Leistung dabei schwanken zu lassen.

Ähnliche Herausforderungen treten bei Online-Menschenhandel auf, wo der Abgleich von Opferdaten zwischen Ermittlungen wichtig sein kann, um das größere Netzwerk der verantwortlichen Täter aufzudecken. [22]

Die hohe Variabilität in der Gesichtserkennungskompetenz (FIP) zwischen verschiedenen Personen und auch innerhalb einer Person zeigt, dass herkömmliche Tests, die nur die Genauigkeit oder Trefferquote messen, nicht zuverlässig die tatsächliche Arbeitsleistung in sicherheitskritischen Situationen widerspiegeln. In der Praxis umfasst die berufliche Leistung viele unterschiedliche Aufgaben, von denen jede in bestimmten Situationen wichtig sein kann und verschiedene Aspekte der Leistung erfordert. Ein einzelner Test erfasst möglicherweise nicht alle relevanten Unterschiede in der FIP zwischen Individuen. Eine Möglichkeit, dieses Problem zu lösen, besteht darin, umfassendere Testverfahren einzuführen, die mehrere anspruchsvolle Tests kombinieren. Ein anderer ergänzender Ansatz wäre, speziell entwickelte Aufgaben zu verwenden, die genau die Fähigkeiten überprüfen, die für bestimmte berufliche Rollen besonders wichtig sind. [22]

Die Autoren haben hier systematisch untersucht, welcher Ansatz besser geeignet ist, um die Gesichtserkennungskompetenz (FIP) für die Auswahl von Personal im Polizeibereich zu bewerten. Basierend auf früheren Studien, die zeigen, dass eine hohe Konsistenz im Verhalten mit überlegener FIP-Fähigkeit verbunden ist, haben sie die Hypothese aufgestellt, dass Beamte mit besseren Ergebnissen bei schwierigen laborgestützten FIP-Tests auch eine höhere Konsistenz in ihren Leistungen zeigen würden. Diese Konsistenz wurden innerhalb von zwei parallel entwickelten Arbeitsproben gemessen. Wenn sich ihre Hypothese als korrekt herausstellt, würde das bedeuten, dass die Messung der Konsistenz innerhalb einer Arbeitsprobe ausreichen könnte, um die Gesamtleistung bei anspruchsvollen FIP-Aufgaben vorherzusagen. Dies bestätigt die Anwendbarkeit ihrer Arbeitsprobenaufgaben auf die Arten von Tätigkeiten, die routinemäßig von tatsächlichen Polizeifachleuten durchgeführt werden. Darüber hinaus liefert dies überzeugende Beweise dafür, dass die Konsistenz eines Beobachters bei einem bestimmten Test allgemein seine FIP-Fähigkeit vorhersagen kann. In der Praxis bietet dies eine neuartige Möglichkeit für Strafverfolgungsbehörden, die Personalbewertung zu optimieren. Die Ergebnisse deuten darauf hin, dass die Konsistenz der Leistung in speziell auf die Aufgabe abgestimmten Arbeitsproben eine ressourcenschonende Methode darstellt, um die Personalzuweisung optimal zu informieren. [22]

Die Autoren gingen wie folgt vor.

Alle Verfahren wurden vom Ethikkomitee der Universität Fribourg in der Schweiz genehmigt. Die Teilnehmer waren gesunde Polizeifachleute, die freiwillig an der Studie teilnahmen und keine finanzielle Entschädigung erhielten. Die Experimente wurden online durchgeführt. Die Labortests fanden auf einer speziell für das Applied Face Cognition (AFC) Lab entwickelten Plattform statt, während die Arbeitsproben über Testable durchgeführt wurden. Jeder Teilnehmer erhielt vor Beginn der beiden Sets von Verhaltensversuchen ein digitales Informationsdokument und hatte unbegrenzt Zeit, dieses zu lesen und seine Zustimmung zu geben. [22]

Die Teilnehmer sind eine spezielle Gruppe von sehr erfahrenen Polizisten aus vier verschiedenen Abteilungen der Polizeidirektion Koblenz in Rheinland-Pfalz. Dazu gehörten Beamte aus drei regionalen Dienststellen und der Kriminalpolizei Koblenz. Alle diese Polizisten arbeiten regelmäßig mit Gesichtserkennung bei verschiedenen sicherheitsrelevanten Aufgaben, die in ihrer Intensität und ihren Anforderungen variieren. Die Teilnehmer lassen sich in drei Hauptgruppen einteilen:

- Sicherheitsbeamte: Diese Beamten sind Generalisten, die häufig Identitätsprüfungen durchführen, indem sie offizielle Dokumente wie Ausweise und Führerscheine überprüfen. Sie sind auch in verschiedenen Suchaktionen beteiligt, zum Beispiel bei der Suche nach vermissten Personen oder Verdächtigen anhand von Überwachungsvideos. Außerdem nehmen sie an speziellen Einsätzen teil, wie der Identifizierung von Personen bei großen Veranstaltungen.[22]
- Sucheinheit-Mitglieder: Diese Beamten arbeiten in speziellen Einheiten, die für die Vollstreckung von Haftbefehlen zuständig sind oder für Bildvergleiche im Rahmen von Straftaten. Sie sind aktiv in der operativen Arbeit und bearbeiten Fälle, die von Einzelpersonen oder Gruppen begangen wurden. [22]
- Massendatenanalysten: Diese Beamten analysieren große Mengen an Daten aus verschiedenen Quellen, wie dem Dark Web, Videoüberwachungssystemen oder Handydaten. Ihre Aufgabe ist es, Opfer und Verdächtige zu identifizieren und zu verfolgen, vor allem bei schweren Straftaten wie Kinderpornografie, Mord oder Betrug. [22]

Von ursprünglich 390 eingeladenen Polizeibeamten meldeten sich 183 freiwillig zur Teilnahme an der Studie. Aufgrund von Krankheit, Quarantäne oder Einsätzen konnten 37 nicht teilnehmen. Schließlich absolvierten 146 Beamte die Gesichtserkennungstests unter kontrollierten Bedingungen auf einer speziellen Plattform. Nach dem Ausschluss von 32 Teilnehmern wegen unvollständiger oder fehlerhafter Daten verblieben 114 Beamte mit vollständigen Daten, von denen 5 als Super-Recognizer identifiziert wurden.[22]

Die Teilnehmer nahmen an zwei verschiedenen Experimenten teil. Erstens an drei anspruchsvollen Tests zur Gesichtserkennungskompetenz, die speziell zur Identifizierung von Super-Recognizern entwickelt wurden, und zweitens an zwei Versionen eines Arbeitsmusters. Dieses Arbeitsmuster sollte die Art von Aufgaben widerspiegeln, die in den verschiedenen beruflichen Rollen der Polizeibeamten vorkommen. In der ersten Versuchsreihe wurden drei anspruchsvolle, bereits etablierte Tests zur FIP durchgeführt. Von den 146 Beamten, die alle drei Tests abgeschlossen haben, erfüllten sechs die kürzlich festgelegten Kriterien für Super-Recognizer. Fünf dieser sechs Beamten wurden in die endgültige Datengruppe aufgenommen, die für die Analyse verwendet wurde ($n = 114$).[22]

Zur Entwicklung der Arbeitsmuster für diese Studie wurden verschiedene Informationsquellen genutzt. Dabei wurde der trimodale Ansatz von Schuler verwendet, um die Validität der Personalbewertung und -auswahl zu verbessern.[22]

Zunächst brachte einer der Mitautoren, seine umfangreiche Berufserfahrung ein. Er hat 17 Jahre bei der Polizei in Rheinland-Pfalz gearbeitet und dabei verschiedene Positionen in den Bereichen besetzt, aus denen die Teilnehmer rekrutiert wurden. Er hat fünf Jahre im Bereich Sicherheitsdienst gearbeitet, wo er unter anderem mit der Identitätsüberprüfung und dem Abgleich von Dokumenten zu tun hatte. In den übrigen 12 Jahren war er in der Kriminalpolizei tätig, wo er tiefgehende Erfahrung in der Verarbeitung von Gesichtserkennung und der Nutzung von Ermittlungsverfahren in den Bereichen Eigentumskriminalität und Mord gesammelt hat. Zusätzlich zu seinen Erfahrungen wurden die Ergebnisse einer Analyse von sieben weiteren Experten aus spezialisierten Einheiten zur Gesichtserkennung in die Gestaltung der Arbeitsmuster einbezogen. Aus Sicherheitsgründen konnten jedoch keine weiteren Details zu diesen Analysen veröffentlicht werden. [22]

Um die Arbeitsproben für diese Studie zu entwickeln, wurden zwei Hauptansätze verwendet. Erstens wurden Expertenbewertungen und die Methode der kritischen Vorfälle genutzt, um herauszufinden, welche Aufgaben wichtig sind. Zweitens wurde eine detaillierte Aufgabenanalyse durchgeführt. Diese beiden Methoden halfen dabei, Arbeitsproben zu gestalten, die die unterschiedlichen Anforderungen an die Gesichtserkennung (FIP) in den verschiedenen Polizeieinheiten und -rollen abdecken. Die Arbeitsproben, die die Autoren gewählt haben, sollten eine faire Grundlage für die Bewertung von

FIP-Fähigkeiten bieten, die in unterschiedlichen Polizeiarbeiten wichtig sind. Sie entschieden sich für eine Aufgabe, bei der die Teilnehmer zwischen zwei Bildern wählen mussten, um festzustellen, ob es eine Übereinstimmung gab. Diese Methode ermöglichte es, die Leistung der Teilnehmer auf eine Weise zu messen, die für alle vergleichbar war, unabhängig von ihrer Rolle oder Erfahrung.[22] Das Bildmaterial wurde sorgfältig ausgewählt, um sicherzustellen, dass es dem entspricht, was Polizeibeamte in ihrem täglichen Berufsalltag sehen. Dabei wurde auch darauf geachtet, dass es sowohl Übereinstimmungen als auch falsche Übereinstimmungen gab. Dies half, Verzerrungen in den Antworten der Teilnehmer auszugleichen, da die Aufgabe sicherstellte, dass jeder Teilnehmer ähnliche Herausforderungen hatte. [22]

Zwei Versionen der Arbeitsproben wurden erstellt und zufällig an die Teilnehmer verteilt. Beide Versionen hatten einen ähnlichen Ablauf, beginnend mit vier Übungsversuchen zur Einführung. Die eigentlichen Aufgaben bestanden darin, dass den Teilnehmern ein Bild des Zielgesichts für 1000 Millisekunden gezeigt wurde, gefolgt von einer 500 Millisekunden langen Pause. Anschließend wurden zwei Bilder angezeigt, eines links und eines rechts des Zielgesichts. Eines dieser Bilder zeigte das gleiche Gesicht wie das Zielgesicht, das andere ein anderes. Die Bilder wurden aus sechs verschiedenen Blickwinkeln angezeigt. Die Teilnehmer sollten schnell und genau angeben, ob das Gesicht auf der linken oder rechten Seite dem Zielgesicht entsprach, indem sie die A oder L Taste drückten. Ziel war es, die Fähigkeit der Teilnehmer zur Gesichtsidentifikation und -unterscheidung zu testen. [22]

Die Aufgaben in dieser Studie verwendeten Bilder von 12 Männern und 12 Frauen aus der Tufts Face Database. Diese Datenbank enthält Bilder von Personen aus verschiedenen Blickwinkeln und unter verschiedenen Lichtverhältnissen, um reale Polizeibilder nachzubilden. Für jede Person wurde ein großes Bild aus der Frontalansicht als Ziel verwendet. Zusätzlich wurden kleinere Graustufenbilder der gleichen Person aus verschiedenen seitlichen Blickwinkeln als Proben verwendet. Jedes Bild wurde auf eine bestimmte Größe angepasst, um sicherzustellen, dass die Tests unter ähnlichen Bedingungen durchgeführt werden. Die Graustufenbilder wurden aus Farbbildern erstellt, indem die Farben in Graustufen umgewandelt wurden. Die Teilnehmer konnten die Aufgaben auf ihrem bevorzugten Gerät durchführen, wobei Handys und Tablets ausgeschlossen waren, da diese nicht die nötige Bildqualität bieten könnten. Alle Bilder wurden vor einem neutralen grauen Hintergrund angezeigt, um Ablenkungen zu vermeiden. [22]

Um die Anzahl der möglichen Versuche zu verringern, die bei allen paarweisen Kombinationen der Identitäten entstehen würden, und um die Aufgabe in jeder Arbeitsprobe etwas anspruchsvoller zu gestalten, haben wir die Identitäten in zwei Gruppen basierend auf ihrer Ähnlichkeit eingeteilt. Das bedeutet, dass in jeder Arbeitsprobe jedes Zielbild mit fünf verschiedenen Vergleichsbildern kombiniert wurde. Die Reihenfolge der Versuche wurde für jeden Teilnehmer zufällig festgelegt. Alle möglichen Kombinationen von Ziel- und Vergleichsbildern wurden innerhalb jeder Gruppe in einem ausgewogenen Design präsentiert, sodass jede Arbeitsprobe insgesamt 240 Versuche umfasste.[22]

Die Teilnehmer erhielten spezielle Links zu einer Testplattform des AFC Lab über ihre beruflichen E-Mail-Adressen. Über diese Links nahmen sie an drei Tests zur Gesichtserkennung teil, die in zufälliger Reihenfolge durchgeführt wurden. Diese Tests fanden in Gruppen von bis zu 12 Personen im Polizeipräsidium Koblenz statt, wo alle Teilnehmer an identischen Computern arbeiteten. Der Abstand zum Bildschirm betrug 60 cm, und alle trugen während des Tests medizinische Masken, um den Vorschriften des Innenministeriums zu entsprechen.[22]

Die Tests wurden unter kontrollierten Bedingungen ohne Zeitdruck durchgeführt und dauerten 40 bis 90 Minuten, wobei Pausen erlaubt waren. Zwei Wochen später erhielten die Teilnehmer Links zu Arbeitsproben, die sie innerhalb von vier Wochen auf ihrem Büro-Laptop oder -Desktop bearbeiten

sollten. Vor Beginn der Tests mussten sie eine Bildschirmkalibrierung vornehmen. Jede Arbeitsprobe dauerte 15 bis 20 Minuten. Die zeitliche Trennung zwischen den Testphasen spiegelt reale Personalauswahlbedingungen wider und minimiert Verzerrungen, während der Ansatz die Bedingungen der tatsächlichen Berufsausübung simuliert und standardisierte Testbedingungen sicherstellt. [22]

Frühere Studien zur Gesichtserkennung (FIP) haben gezeigt, dass Faktoren wie das Geschlecht der Bilder oder der Blickwinkel die Ergebnisse beeinflussen können. Das Ziel war es, herauszufinden, wie konstant Polizeifachkräfte bei der Gesichtserkennung sind. Dazu haben wir untersucht, wie stark die Effizienz-Scores (IE) für jeden Beamten in dieser Stichprobe schwanken. Wenn die Scores eines Beamten wenig variieren, bedeutet das, dass er konstant arbeitet. Wir haben die Konsistenz der Leistung jedes Beamten innerhalb jeder Arbeitsprobe gemessen, indem wir die Standardabweichung der IE-Scores berechnet haben. Um die Daten für die Hauptkomponentenanalyse (PCA) und die Korrelation mit den laborbasierten FIP-Tests vorzubereiten, haben wir die IE-Scores und Konsistenzwerte in eine Form umgewandelt, bei der höhere Werte besser sind. Dies geschah durch eine negative Z-Transformation, um die Unterschiede in der Varianz zwischen den Variablen zu vereinheitlichen. [22]

Zuerst wurde die Leistung der Polizeibeamten in den anspruchsvollen Gesichtserkennungstests bewertet, wie es in früheren Studien beschrieben wurde. Für jeden Test wurden die durchschnittlichen Ergebnisse berechnet. Um herauszufinden, wie die Unterschiede zwischen den verschiedenen Testversionen (männlich vs. weiblich) und den verschiedenen Blickwinkeln (23°, 45°, 68°) die Ergebnisse beeinflussten, wurde eine spezielle statistische Analyse durchgeführt. Diese Analyse half zu verstehen, wie genau und schnell die Beamten bei den verschiedenen Tests waren. Anschließend wurde eine Hauptkomponentenanalyse durchgeführt, um herauszufinden, welche Faktoren die größten Unterschiede zwischen den Beamten in den Testergebnissen verursachten. Dafür sind alle Variablen so angepasst, dass ihre Streuung vergleichbar ist. Dies ermöglicht es, die wichtigsten Unterschiede in den Fähigkeiten der Beamten zu identifizieren. [22]

In der Hauptkomponentenanalyse (PCA) konnte festgestellt werden, dass drei Hauptkomponenten die Varianz in den Daten hauptsächlich erklären: die Reaktionszeiten der Arbeitsproben, die Genauigkeit der Arbeitsproben und die Leistung in den laborgestützten Tests. Dies bedeutet, dass Unterschiede zwischen den Arbeitsprobenbedingungen nur einen sehr kleinen Teil der Gesamtvariabilität ausmachen, während die meiste Varianz durch die aggregierten Reaktionszeiten und die Genauigkeit erfasst wird. Um die Daten zu vereinfachen, wurde die Reaktionszeiten über verschiedene Blickwinkel und Versionen der Arbeitsproben hinweg zusammengefasst und ebenso die Genauigkeit. Es war wichtig, die Beziehung zwischen Reaktionszeiten und Genauigkeit zu berücksichtigen, bevor ein kombinierter Effizienzscore (IE-Score) berechnet wurde. Dabei wurde festgestellt, dass eine höhere Genauigkeit mit kürzeren Reaktionszeiten verbunden war, was einen negativen Zusammenhang ergab. Aus diesem Grund wurden die Daten weiter vereinfacht, indem die Genauigkeit und die Reaktionszeiten zu einem Effizienzscore (IE-Score) kombinierten. Dadurch erhielt man eine einzige Variable, die die Effizienz einer Person erklärt. Diese Methode erwies sich als die einfachste und effizienteste Möglichkeit, die individuellen Unterschiede in der Effizienz zu messen, anstatt jede Bedingung oder jede Maßnahme separat zu analysieren. [22]

Die Regressionsanalyse verglich die Effizienz und Konsistenz der Arbeitsproben mit den Ergebnissen der Labortests für FIP. Alle Werte wurden normalisiert, sodass höhere Werte für höhere Effizienz, Konsistenz und FIP stehen. Das Ergebnis zeigte, dass Polizisten mit effizienter und konsistenter Arbeitsprobenleistung auch höhere Labortestwerte erzielten. Dies bedeutet, dass diejenigen, die in den praktischen Tests gut abschnitten, auch in den formellen Labortests hohe Punktzahlen erreichten. [22]

In Anbetracht der Herausforderungen bei den aktuellen Methoden zur Ausbildung und Auswahl von Personal bestand das Ziel darin, eine Reihe von kriteriumfreien Methoden zur Bewertung der FIP-Fähigkeiten (Gesichtserkennung) in einer Stichprobe von 114 Polizeibeamten zu validieren und anzuwenden. Zunächst wurde eine umfassende Analyse der individuellen Leistungsprofile vorgenommen, die drei anspruchsvolle laborgestützte FIP-Tests und maßgeschneiderte Arbeitsproben umfasste. Ziel dieses datengesteuerten Ansatzes war es, herauszufinden, welche Kombinationen der vielen möglichen Dimensionen am besten geeignet sind, um individuelle Unterschiede in der FIP zu messen. Zuerst wurde die Dimensionalität der Daten mittels Hauptkomponentenanalyse (PCA) reduziert. Die PCA ist eine standardisierte, explorative Methode, die in verschiedenen wissenschaftlichen Bereichen verwendet wird, um die Dimensionen großer Datensätze basierend auf den Beobachtungen zu reduzieren. Es ist erwähnenswert, dass in den meisten bisherigen Studien zur angewandten Gesichtserkennung solche Methoden nicht angewendet wurden, da diese Studien oft nur eine sehr geringe Anzahl von Versuchen und somit eine begrenzte Anzahl von Datenpunkten pro Beobachter aufweisen. Nachdem die drei informativsten Dimensionen mittels PCA identifiziert wurden, wurden Effizienz- und Konsistenzmaße aus den Arbeitsprobenleistungen abgeleitet. Schließlich wurden diese Maße mit den laborgestützten FIP-Tests korreliert, um zu prüfen, ob Effizienz und Konsistenz bei praxisnahen Aufgaben mit der Leistung bei laborgestützten Tests in Zusammenhang stehen. [22]

Der datengesteuerte Ansatz mittels PCA für diagnostische Tests zeigt, dass die Leistung in Arbeitsproben, insbesondere die Reaktionszeit (RT), den Großteil der interindividuellen Varianz in der Gesichtserkennung (FIP) erfasst. Dies ist sowohl für die Datenqualität als auch für praktische Belange in der Personalauswahl von Bedeutung. Die Anwendung der PCA ermöglicht es, herauszufinden, welche Maße unter den zahlreichen, die in laborgestützten Tests und Arbeitsproben erfasst wurden, die informativsten Quellen individueller Unterschiede in der FIP darstellen. Während laborgestützte Bewertungen nützlich für die Beurteilung der FIP sein können, bieten maßgeschneiderte Arbeitsproben die Möglichkeit, empfindlichere und praktischere Maße der FIP zu erfassen als nur die Genauigkeit der laborgestützten Tests. Beispielsweise zeigen die Reaktionszeiten in den beiden parallelen Versionen der Arbeitsproben eine starke Belastung auf die erste Hauptkomponente, was darauf hinweist, dass sie deutlich stärker miteinander verbunden sind als mit den verwendeten laborgestützten FIP-Tests. [22]

In der Praxis ist die Reaktionszeit (RT) entscheidend für Aufgaben, bei denen Schnelligkeit wichtig ist. Solche Aufgaben sollten so gestaltet werden, dass die RT berücksichtigt wird. Während die Genauigkeit bei den Arbeitsproben weniger zur Gesamtbewertung beiträgt, zeigen maßgeschneiderte Arbeitsproben eine höhere ökologische Validität als allgemeine Labortests. In realen Situationen, wie bei der Passkontrolle, ist es nicht nur wichtig, Identitäten korrekt zu überprüfen, sondern auch schnell zu arbeiten, um große Datenmengen effizient zu bearbeiten. Hohe Konsistenz bei der Identitätsüberprüfung ist ebenfalls wichtig, um gleichbleibende Ergebnisse zu gewährleisten. [22]

In der Praxis treten übereinstimmende und nicht übereinstimmende Gesichter unterschiedlich häufig auf. Dies beeinflusst die Reaktionsverzerrungen: Wenn übereinstimmende Gesichter häufiger vorkommen, neigen Beamte dazu, Übereinstimmungen eher zu erkennen und Fehlalarme zu vermeiden. Um Verzerrungen in Arbeitsproben zu minimieren, wurden in jedem Versuch sowohl ein übereinstimmendes als auch ein nicht übereinstimmendes Gesicht gezeigt. Dadurch werden Unterschiede in der Wahrnehmung zwischen den Personen verringert, was besonders wichtig ist, wenn große Mengen an Beweismaterial geprüft werden müssen. [22]

Fähigkeiten zur Gesichtsidentifikation (FIP) sind schwer trainierbar und oft als spezifische Expertise betrachtet. Labortests zeigen, dass diese Fähigkeiten nicht universell sind, sondern von den getesteten Parametern abhängen. Daher sollten Auswahlverfahren gut kalibrierte Tests nutzen, die

auf die jeweilige Aufgabe abgestimmt sind. Keiner der alleinstehenden Labortests ist ausreichend sensibel, um alle Variationen in der Bevölkerung zu erfassen. Eine Kombination verschiedener Tests kann jedoch ein umfassenderes Bild der Fähigkeiten geben. Da die Ergebnisse von Labortests nicht immer auf echte Aufgaben übertragbar sind, sollte die Auswahl besonders geschickter Personen nicht nur auf solchen Tests basieren, sondern auch auf spezifischen Arbeitsproben, die die erforderlichen Fähigkeiten am besten widerspiegeln. Aufgaben, die auf Regeln basieren, können leichter auf andere Kontexte übertragen werden. Daher kann es effektiver sein, Aufgaben zu wählen, die auf abstrakten Regeln beruhen. Zudem kann die Vielfalt der Beispiele in Trainingsproben helfen, eine umfassendere Vorstellung der realen Anforderungen zu vermitteln. Indem solche Aufgaben die relevanten Leistungsvariationen berücksichtigen, könnten sie dazu beitragen, geeignete Personen auszuwählen und die Schulung zu verbessern.[22]

Die Ergebnisse zeigen, dass bei der Polizeiauswahl neben Genauigkeit auch Effizienz und Konsistenz bei der Gesichtsidifikation berücksichtigt werden sollten. Maßgeschneiderte Tests, die speziell auf die Aufgaben der Polizeiarbeit abgestimmt sind, erfassen individuelle Unterschiede besser als Labortests. Die Auswahl besonders konsistenter Mitarbeiter könnte die Schulungskosten reduzieren. Künftige Forschungen sollten diese Ansätze auf andere Polizeiaufgaben ausweiten und die Entwicklung von Arbeitsproben in enger Zusammenarbeit mit Fachleuten vorantreiben.[22]

Labortests zur Gesichtserkennung sind oft unzureichend, da sie wichtige Aspekte wie Geschwindigkeit und Konsistenz vernachlässigen. Maßgeschneiderte „Arbeitsproben“ bieten bessere Bewertungen für die praktischen Anforderungen im Polizeialltag. Polizeibehörden sollten daher spezialisierte Tests nutzen, die die tatsächliche Leistungsfähigkeit ihrer Beamten besser widerspiegeln. [22]

4 Studie Super-Recognizer Workshop 2023

Im Rahmen des Workshop zu Super-Recognizern im Jahr 2023 wurde eine Studie mit einem Test erstellt.

4.1 Erstellung

Für die Erstellung des Testdatensatzes wurden verschiedene öffentlich zugängliche Datenbanken und bestehende Datensätze sowie einzelne Online-Bilder verwendet. Bei der Auswahl der Bilder wurde zwischen der klassischen Suche nach einer Einzelpersonen und der Suche nach Doppelgängern unterschieden. Für die Personensuche wurde nur der Datensatz „Labeled Faces in the Wild (LFW)“ verwendet.[20]

Der LFW-Datensatz ist einer der am häufigsten verwendeten Datensätze im Bereich der Gesichtserkennung. Er wurde erstellt, um eine breite Vielfalt von Gesichtern mit natürlicher Variation zu erfassen. Diese Vielfalt umfasst verschiedene Aspekte wie die Ausrichtung des Gesichts zur Kamera, Gesichtsausdruck, Alter, Geschlecht, Ethnizität und Kleidung. Darüber hinaus werden auch die Aufnahmebedingungen berücksichtigt, wie Beleuchtung, Aufnahmequalität, Farbsättigung, Fokus und Hintergrund. Der gesamte Datensatz besteht aus insgesamt 13.233 Bildern von 5.749 Personen, wobei die Bilder der Personen in den Maßen 250 x 250 Pixel gespeichert sind.[14]

Die Bildauswahl basierte nicht auf strengen Kriterien, das heißt, es wurden keine Mindestauflösung des Gesichts oder die Position des Gesichts im Bild definiert. Die Bilder im LFW-Datensatz sind bereits auf eine standardisierte Größe normiert, und die Gesichter befinden sich ungefähr im Zentrum des Bildes. Für jeden Vergleich wurde ein Bild ausgewählt, das von 10 weiteren Bildern begleitet wurde. Eines dieser Bilder zeigte zusätzlich die Vergleichsperson, während die übrigen Bilder Personen mit ähnlichem Geschlecht und Erscheinungsbild darstellten. Ziel war es, eine breite Variation von Kopfpositionen in Bezug auf die Kamera in den Auswahlbildern zu präsentieren.[20]

Verschiedene Quellen wurden verwendet, um die Doppelgänger auszuwählen, wobei der Schwerpunkt auf bekannten Persönlichkeiten und ihren Doppelgängern lag. Eine teilweise Quelle ist die Website ¹ „NEW LOOKALIKES“ einer Agentur, die Doppelgänger für Veranstaltungen bereitstellt. Darüber hinaus wurden auch Body-Doubles von Prominenten verwendet, da sie oft in Filmen als Stellvertreter für Stars auftreten und daher eine Ähnlichkeit mit den Stars aufweisen müssen. Zusätzlich wurden Medienberichte über Personen berücksichtigt, die Prominenten ähnlich sehen. Basierend auf diesen Berichten war es möglich, andere Personen zu identifizieren, die den bekannten Personen sehr ähnlich sind und als Doppelgänger betrachtet werden können. Die Bilder der bekannten Personen wurden aus öffentlich zugänglichen Quellen ausgewählt. Bei der Auswahl der Bilder wurden Kriterien berücksichtigt, die denen ähneln, die bei der Personensuche verwendet werden, insbesondere hinsichtlich der Variation der Kopfposition in Bezug auf die Kamera. Die Auflösung und Position des Gesichts innerhalb des Bildes wurden nicht normalisiert, um den Schwierigkeitsgrad zu erhöhen. [20]

Um die Herausforderung zu erhöhen und eine größere Variation zwischen den verschiedenen Aufgaben innerhalb der Personensuche zu schaffen, wurden die Auswahlbilder anschließend bearbeitet. Zu diesem Zweck wurden die Bilder zufällig gespiegelt und auf das Gesicht zugeschnitten (siehe Abbildung 4.1, erste Zeile „Bildtransformation“). Darüber hinaus wurden alle Auswahlbilder in

¹<https://new-lookalikes.com/ueber-uns/>

Graustufen umgewandelt und gespeichert. Zusätzlich zu diesem Prozess wurden die Bilder stark verändert, einschließlich Verschwommung, Anwendung von Rauschfiltern, deutliche Erhöhung der Sättigung und Helligkeit sowie Färbung (siehe Abbildung 4.1, zweite Zeile „Bildmanipulation“). Auch künstlerische Effekte wie die Umwandlung des Bildes in einen Zeichenstil wurden angewendet (siehe Abbildung 4.1, dritte Zeile „Künstlerischer Effekt“). Die Bilder wurden nach jeder Bearbeitung gespeichert.[20]

Die benutzerfreundliche Testansicht des Datensatzes für die Durchführung der Tests mit den Super-Recognizern der Polizeidirektion Chemnitz wurde in Google Forms² erstellt. Dies ist ein Tool, das die Gestaltung, Veröffentlichung und Visualisierung von Online-Umfragen ermöglicht. Auf diese Weise können Benutzer durch die einzelnen Sets und Aufgaben navigieren und ihre Antworten auswählen. Die Ergebnisse der Super-Recognizer konnten somit eingesehen und ausgewertet werden. [20]

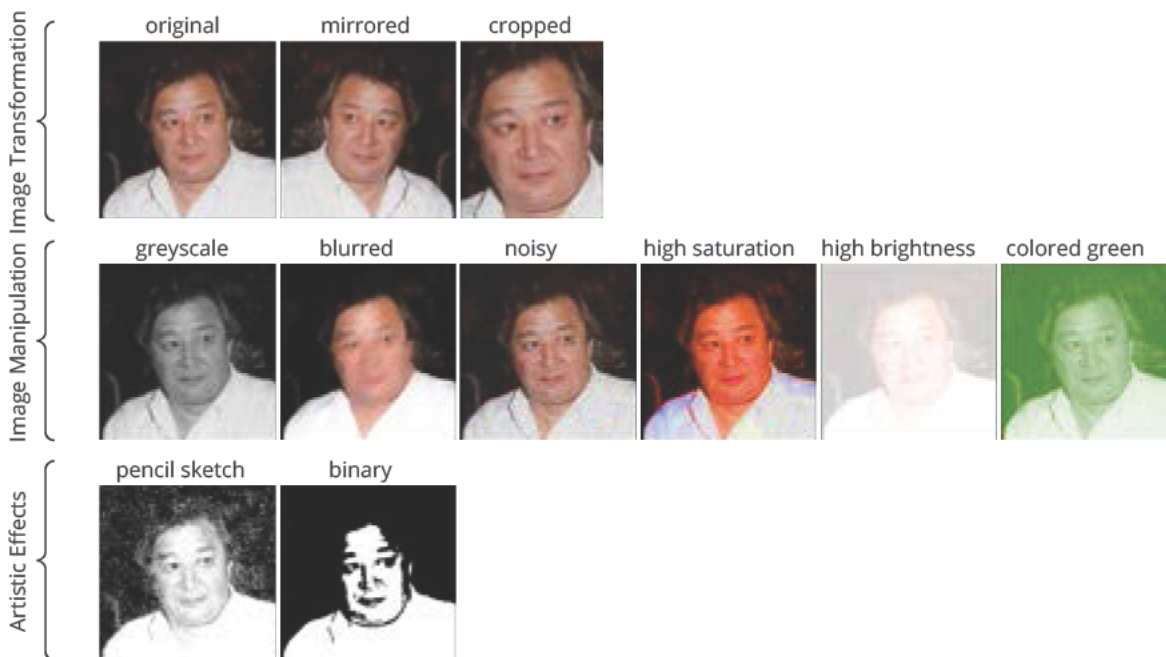


Abbildung 4.1: Beispiel Bilder nach Transformation (erste Reihe, mittleres und letztes Bild) sowie nach der Bildbearbeitung (zweite Reihe) und Anwendung von künstlerischen Effekten (dritte Reihe) unter Verwendung eines Beispielbildes aus dem LFW-Datensatz (erste Reihe, erstes Bild). Quelle des Originalbildes [14]

²<https://www.google.de/intl/de/forms/about/>

4.2 Aufbau

Die Studie ist in neun verschiedene Aufgaben aufgeteilt, im folgenden als Sets bezeichnet, die jeweils aus vier bis fünf Abfragen bestehen. In den Sets ist jeweils ein Vergleichsbild einer anderen Zielperson gegeben. Set 1 - 3 ist eine klassische Personensuche. Hier ist ein Vergleichsbild der Zielperson gegeben und diese soll aus einer Menge von 10 Bildern wieder identifiziert werden. In Set 4 - 9 geht es um eine Doppelgängersuche und es werde prominente Personen und deren Doppelgänger gezeigt. Die gesuchte Person muss im folgenden immer aus 10 Bildern herausgefunden werden. Die Super-Recognizer hatten 30 Sekunden lang Zeit sich das Vergleichsbild der Zielperson anzuschauen, ebenso 30 Sekunden Zeit um die Auswahlbilder zu betrachten.[20]

Im ersten Set wurde ein männliches Vergleichsbild gegeben die sich die 10 eingeladenen Super-Recognizer anschauen konnten. Auf der nächsten Folie werden 10 Bilder von Männern gezeigt und das Bild der gesuchten Person soll herausgefunden werden. Die nächste Abfrage war zum selben Vergleichsbild und Auswahlbildern, jedoch sind diese in schwarz weiß. Bei der dritte Abfrage sind die Auswahlbilder gemischt zwischen Farbe und schwarz weiß und zeigen einen heran oder heraus gezoomten Ausschnitt. Im vierten Durchgang werden die Auswahlbilder noch weiter verschlechtert z.B. durch Farbfilter oder geringere Auflösung. [20]

Im nächsten Set sind die verschiedenen Durchgänge gleich aufgebaut. Die Zielperson ist hier ein dunkelhäutiger Mann. [20]

Bei der Personensuche wurden drei verschiedene Männer mit unterschiedlicher Hautfarbe als Zielperson gezeigt. Die Auswahlbilder wurden jeweils so gewählt, dass die 9 weiteren Personen der Ethnie der Zielperson entsprechen. Im ersten Set wurde ein männliches Vergleichsbild (siehe Abbildung 4.2) gegeben welches sich die 10 eingeladenen Super-Recognizer der Polizeidirektion Chemnitz anschauten. Als nächstes werden 10 farbige Auswahlbilder von Männern gezeigt, in Abbildung 4.3 dargestellt, und das Bild auf dem die gesuchte Person abgebildet ist soll wieder erkannt werden. [20]



Abbildung 4.2: Vergleichsbild von Set 1



Abbildung 4.3: Auswahlbilder von Set 1

Alle vier Abfragen zu einer Person beziehen sich auf dasselbe gegebene Vergleichsbild. Die nächste Abfrage hatte die gleichen Auswahlbilder, jedoch sind diese in schwarz weiß. Bei der dritten Abfrage sind die Auswahlbilder gemischt zwischen Farbe und schwarz weiß und zeigen einen heran oder heraus gezoomten Ausschnitt derselben 10 Auswahlbilder wie zuvor. Im vierten Durchgang wurden die Auswahlbilder noch weiter verändert z.B. durch Farbfilter (siehe Abbildung 4.1) oder verringerte Auflösung. [20]

In Set 2 und 3 sind die Abfragen nach der gleichen Struktur aufgebaut wie bei Set 1. [20]

Bei einer Doppelgängersuche gibt es eine berühmte Person und eine Person die dieser sehr ähnlich aussieht z.B. ein Schauspielerdouble. In den folgenden 6 Sets wird jeweils entweder der Doppelgänger oder die berühmte Person das Vergleichsbild sein und in den 10 Auswahlbildern wird ein bis zweimal die Zielperson zu finden sein. [20]



Abbildung 4.4: Vergleichsbilder von Set 4-9

Bei Set 4 sind die vier Abfragen wie bei der Personensuche aufgebaut nur dass die 10 Auswahlbilder nicht dieselben bleiben und die meisten mit Neuen, noch nicht bekannten Bildern, ausgetauscht werden. Das folgende Set 5 weist den selben Ablauf wie das vorangegangene auf, jedoch gibt es hier eine fünfte Abfrage in der die farbigen Vergleichsbilder aus der ersten Abfrage auf das Gesicht zugeschnitten worden sind. Set 6 ist wieder aus 4 Abfragen aufgebaut. In Set 7 gibt es nur drei Abfragen, eine ursprünglich vierte Abfrage musste aufgrund von einem Fehler herausgenommen werden. Grund dafür war das zu den Bildern von Margot Robbie und ein Bilder von einer Harley Quinn Doppelgängerin gezeigt wurden, nach welcher jedoch nicht gesucht wurde und somit kein Bild der Vergleichsperson dabei war. *Marilee Lesley* wird in Set 8 als Doppelgängerin von Reese Witherspoon gesucht. Dabei sind bei Abfrage 4 auch wieder mehrere Bilder die Zielperson. In den Auswahlbildern ist nicht nur Reese Witherspoon und ihre Doppelgängerin sondern auch ihre Tochter dargestellt. Diese wurde in dieser Studie nicht weiter berücksichtigt. Das letzte Set besteht aus 5 Abfragen. Dabei können wieder mehrere Bilder die gesuchte Person darstellen. [20]

4.3 Künstliche Intelligenzen

Den zehn Super-Recognizer der Polizeidirektion Chemnitz werden drei Künstliche Intelligenzen gegenübergestellt, um zu schauen wie beide Parteien bei der Studie abschneiden. Im folgenden Abschnitt wird *Face Recognition*, *FaceX-Zoo* und *GhostFaceNet* kurz vorgestellt. [20]

Tabelle 4.1 zeigt die Genauigkeit der aktuellen Gesichtserkennungsmodelle auf dem LFW-Datensatz zur Kategorisierung der hier verwendeten Modelle. [20]

Tabelle 4.1: Accuracy der aktuellen Modelle zur Gesichtserkennung auf dem LFW-Datensatz

Model	Accuracy	Paper
GhostFaceNet	0.9987	[2]
Prodpoly	0.9983	[6]
DiscFace	0.9983	[15]
ArcFace	0.9983	[16]
DCQ	0.9980	[19]
AdaFace+WebFace4M+R100	0.9980	[16]
EdgeFace - S	0.9978	[12]
CircleLoss	0.9973	[34]
FaceTransformer+OctupletLoss	0.9973	[18]
EdgeFace - XS	0.9973	[12]
Face Recognition	0.9938	[11]
QMegFace	0.9850	[35]
OcularAi-Face	0.9450	[32]

4.3.1 Face Recognition

Face Recognition ist ein Framework, mit dem man Gesichter erkennen und manipulieren kann [11]. Auch gibt es Skripte für die Analyse von Gesichtern, inklusive der Detektion markanter Gesichtsbereiche. [11]

Für diese Studie wurde die Eingabe des Modells zum Abgleich von Gesichtern so angepasst, dass jeweils ein Eingabebild mit mehreren Referenzbildern verglichen werden konnte. Bei dem Eingabebild handelt es sich um das „Vergleichsbild“ und bei den Referenzbildern demnach um die „Auswahlbilder“.

Um eine Entscheidung von dem Modells zu bekommen, wurde bei den Sets jeweils das Gesichtspaar mit der insgesamt höchsten Übereinstimmung als Ergebnis des Modells interpretiert. [20]

Der Distanzwert bildet dabei einen Wert zwischen 0 und 1 ab. Wird ein Rückgabewert von 2.0 abgebildet, heißt es wurde in einem der Bilder kein Gesicht gefunden und somit hat kein Abgleich stattgefunden. Bei der Berechnung handelt es sich um ein Distanzmaß und das Ergebnis kann daher nicht als klassische Wahrscheinlichkeit interpretiert werden kann. Demzufolge bedeutet ein geringer Wert eine kleine Distanz zwischen zwei Gesichtern und zeugt somit von einer hohen Ähnlichkeit der Gesichter zueinander. [20]

Nach Angaben der Entwickler von *Face Recognition* erzielte das Modell eine Genauigkeit von 0,9938 [11] auf dem LFW-Datensatz und gehört damit zu den genauesten Gesichtserkennungsframeworks die derzeit verfügbar sind. Eine unabhängige Auswertung dieser Informationen auf dem "Labeled Faces in the Wild (LFW)Datensatz ergab eine Genauigkeit von 0,9900 [23]. *Face Recognition* wurde verwendet, weil die Autorin bereits unabhängige Evaluierungen mit dem Ziel durchgeführt hatte, das Modell zu optimieren, die Genauigkeit zu bewerten und die Ergebnisse klar interpretieren zu können [23].

4.3.2 FaceX-Zoo

FaceX-Zoo ist ein PyTorch-Toolkit für Gesichtserkennung. Es bietet ein Trainingsmodell und ein standardisiertes Evaluierungsmodell. Direkt zur Anwendung stellt *FaceX-Zoo* auch ein Face-SDK zur Verfügung. Für den in dieser Arbeit beschriebenen Test wurde die *face_pipeline.py* verwendet. [38]

Als Eingabe verlangt *FaceX-Zoo*, dass die beiden Bilder die verglichen werden sollen, als ein Bild nebeneinander liegen. Zudem müssen diese dieselbe Höhe aufweisen, weshalb bei zu kleinen Bildern ein schwarzer Balken an der unteren Kante angefügt wurde. Jedes „Vergleichsbild“ wurde dann mit jedem der dazugehörigen 10 „Auswahlbildern“ zusammengefügt und dann der KI als Eingabe übergeben. Da die KI kein Gesicht erkennen konnte wenn auf einem Bild mehr als eine Person zu sehen war, wurde auf einigen Bildern die irrelevanten Gesichter im Hintergrund, wie in Abbildung 4.5 gezeigt, geschwärzt. [20]



Abbildung 4.5: Input mit ablenkenden irrelevanten Gesichtern im Hintergrund (links) und benutzter Input mit geschwärzten Gesichtern im Hintergrund (rechts)

Wenn die KI trotz der Anpassung in einem der Bilder kein Gesicht gefunden hat gibt es *no face detected* zurück. Die zurückgegebenen Werte können als Ähnlichkeitsmaß interpretiert werden. Demzufolge wurde bei den Sets jeweils das Gesichtspaar am ähnlichsten interpretiert mit dem höchsten Wert. Laut den Entwickler von *FaceX-Zoo* erzielt Face-SDK mit der Backbone-Architektur „MobileFaceNet“ auf den LFW-Datensatz eine Genauigkeit von 0,9957%. [38]

4.3.3 GhostFaceNet

GhostFaceNet stellt ebenfalls ein Modell zur Gesichtserkennung dar. Die hier verwendete Implementierung basiert auf dem GitHub-Projekt "deepface" von S. I. Serengil [31] [29] [30]. Das deepface-Framework ermöglicht das Umschalten zwischen verschiedenen modernen Gesichtserkennungs-

modellen Modellen, einschließlich GhostFaceNet [30] [31]. Ein Gesichtsvergleich mit *GhostFaceNet* umfasst zunächst die Erkennung und Ausrichtung der Gesichter im Bild mithilfe von Multi-Task Cascaded Convolutional Networks (MTCNN) [2]. Die Generierung der Gesichtseinbettungen erfolgt basiert dann auf so genannten Ghost-Modulen, die spezielle Blöcke für CNNs sind, die mehr Merkmale mit weniger Parametern zu erzeugen [2]. Zusätzlich wird ein Abstandsmaß zum Vergleich der Gesichter verwendet, und zwar die Cosinus-Distanz [2]. Folglich werden, ähnlich wie bei anderen Gesichtserkennungsmethoden das Auswahlbild mit dem geringsten Abstand zum Abstand zum Vergleichsbild als die Antwort des Modells auf jeden Satz. [20]

Nach Angaben der Entwickler von *GhostFaceNet* erreicht das Modell eine maximale Genauigkeit von 0,9987 auf dem LFW-Datensatz [2]. Das Modell wurde ausgewählt, weil es derzeit eines der genauesten und effektivsten frei verfügbaren Modelle zur Gesichtserkennung ist (siehe Tabelle 4.1). [20]

4.4 Auswertung

In der Studie mit den Aufgaben gab es für jedes richtige Bild einen Punkt. Bei vier Abfragen gab es mehrere zu findende Bilder, daher ergibt sich die Gesamtpunktzahl von 42 Punkten. [20]

Die Ergebnisse der 10 Super-Recognizer liegt zwischen 17 und 41 Punkten. Die Künstlichen Intelligenzen haben 26 Punkte (*Face Recognition*), 24 Punkte (*FaceX-Zoo*) und 18 Punkte (*GhostFaceNet*) erreicht. Die erreichten Punktzahlen der Super-Recognizer und der drei KI's sind in der Abbildung 4.6 dargestellt. [20]

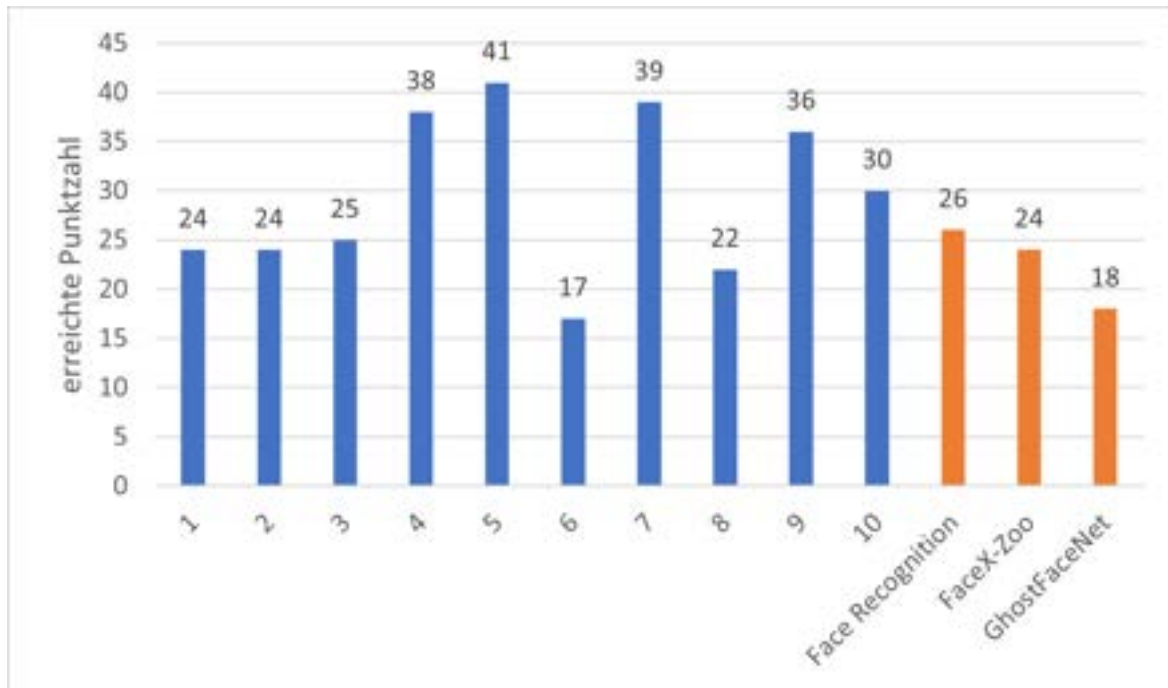


Abbildung 4.6: Ergebnisse der 10 SR und der drei KI's

Sieben der Super-Recognizer haben bei den drei Sets zur Personensuche die volle Punktzahl erreicht. Die Person mit den 22 erreichten Punkten hat in Set 4 keine Antworten abgegeben, demzufolge fehlen von vornherein 4 Punkte, was das Gesamtergebnis der SR verschlechtert ohne das falsche Bilder gewählt wurden. Die meisten falschen Antworten treten bei der Doppelgängersuche auf. Fünf SR konnten keines der Bilder des Doppelgängers in Set 7 finden. Zwei haben eine Punktzahl von 0 bei Set 6. Das beste Gesamtergebnis ist nur zwei Punkte von der Gesamtpunktzahl entfernt, hier wurde in Set 6 und einmal in Set 7 falsch erkannt.[20]

Die Künstliche Intelligenz *Face Recognition* konnte bei der klassischen Personensuche, also Set 1-3, bei der Abfrage mit farbigen, schwarz weißen und beide gemischt die Zielperson wiedererkennen. Jedoch bei der vierten Abfrage, in der die Bilder mit verschiedenen Filtern und geringerer Auflösung gezeigt wurden, konnte die KI die Zielperson nicht korrekt erkennen. [20]

In der Doppelgängersuche bei Set 5-8 konnte die KI jeweils bei der Abfrage nur mit schwarzweiß Bildern die Zielperson nicht erkennen. In Set 5 konnte auch bei der Abfrage mit unterschiedlichen Filtern und Auflösungen nicht korrekt erkannt werden. Bei Set 7 konnte die KI bei der schwarzweiß und farbigen Bilder Abfrage sowie bei den Bildern mit verschiedenen Farbfiltren und Auflösungen die Zielperson nicht korrekt erkennen. Set 8 zu Reese Witherspoon konnte die KI in keiner der Abfragen die Bilder der Zielperson erkennen. Set 9 hatte die KI Schwierigkeiten bei der Abfrage der farbigen

Bilder, der gemischten von Farbigen und Schwarzweiß Bildern und der verschiedener Auflösungen und Farbfiltern. [20]

Die zweite Künstliche Intelligenz *FaceX-Zoo* hat insgesamt etwas schlechter abgeschnitten als die vorher erwähnte KI. In Set 1 und 2 konnten jedoch alle 4 Abfragen jeweils korrekt erkannt werden. Bei Set 3 traten die ersten Schwierigkeiten bei der Erkennung der Zielperson unter den Bildern mit Farbfilter und verschiedenen Auflösungen. In Set 4 zur Doppelgängerin von Jennifer Lawrence konnte die KI bei der Schwarzweiß Abfrage nicht die Zielperson wiedererkennen. Den Doppelgänger zu Will Smith in Set 5 konnte die KI nur in Abfrage 4 zu den schwarz weiß und farbigen Bildern gemischt erkennen, jedoch nur eins von den zwei Bildern die die Zielperson darstellen. [20]

Die KI *Face Recognition* konnte auf 14 Bildern kein Gesicht erkennen, eins davon wäre die gesuchte Person gewesen. Daher kann man nicht sagen ob die KI darauf die Zielperson erkannt hätte. Gründe für das nicht erkennen eines Gesichts sind meist Bilder mit starkem Rauschfilter, Darstellung mit binärem Effekt, eine verdeckte Augenpartie durch eine Sonnenbrille oder die Kopfhaltung ist stark von der frontal Ansicht abweichend. [20]

Die KI *FaceX-Zoo* hat bei vier Bildern kein Gesicht erkennen können, davon ist keines die gesuchte Person gewesen. Grund dafür können die schlechtere Darstellung der Personen sein. Zwei davon sind dasselbe Bild von Emma Watson auf dem sie seitlich zu sehen ist mit nach unten gerichtetem Blick, ein Bild von Will Smith hat eine schlechte Auflösung und das Bild von Jason Statham zeigt diesen von der Seite mit Sonnenbrille und geringem Kontrast und Auflösung.[20]

Die KI *GhostFaceNet* schnitt insgesamt schlechter ab als die zuvor genannten KI's. In Set 1 wurden alle 4 Abfragen korrekt erkannt, aber in Set 2 traten die ersten Schwierigkeiten auf. Keine der 4 Abfragen wurde erkannt und in 5 Bildern mit Farbfiltern und unterschiedlichen Auflösungen konnte *GhostFaceNet* kein Gesicht erkennen. In Set 3 und 7 war die KI nicht in der Lage, die Zielperson in der Abfrage mit den Bildern mit unterschiedlichen Farbfiltern und Auflösungen zu erkennen. In Set 4 konnte die KI alle Bilder erkennen, aber keines in Set 5. In Set 6 konnte die Zielperson in der Schwarz-Weiß-Abfrage nicht erkannt werden. In Set 8 über Reese Witherspoon konnte die KI die Bilder der Zielperson in keiner der Abfragen erkennen. In Set 9 zu Emma Watson wurde die dritte Abfrage mit den Schwarz-Weiß-Bildern richtig erkannt, ebenso wie eines der beiden richtigen Bilder in Frage 5, die aus Schwarz-Weiß- und Farbbildern besteht. [20]

Tabelle 4.2: Accuracy der Super-Recognizer (1-10) und die drei KI's

Type	Accuracy	errichte Punktzahl
1	0.9000	24
2	0.9000	24
3	0.9053	25
4	0.9737	38
5	0.9895	41
6	0.8632	17
7	0.9789	39
8	0.9000	22
9	0.9632	36
10	0.9316	30
Mittelwert der SR	0,9305	29.6
Face Recognition	0,9153	26
FaceX-Zoo	0,9158	24
GhostFaceNet	0,8696	18

Der Mittelwert von den Ergebnissen der 10 Super-Recognizer liegt bei 29,6. Die Standardabweichung dieser liegt bei 8,37. Hier sollte man die starken Unterschiede der Super Recognizer untereinander beachten. Beide Künstliche Intelligenzen sind schlechter als die SR im Mittel. Face Recognition ist 3,6 Punkte schlechter, die KI FaceX-Zoo ist um 5,6 Punkte schlechter als die SR im Mittel. [20]

Die Super-Recognizer könnten auch einen Vorteil Aufgrund des Aufbaus des Tests gehabt haben. Bei der Personensuche sind die Bilder in allen vier Abfragen gleich geblieben, demzufolge könnte es da eher zu einem merken des richtigen Bildes gekommen sein, statt die Fähigkeit des Wiedererkennens anwenden zu müssen. Der selbe Fall tritt nochmal bei der Doppelgängersuche von Will Smith auf, dort wiederholt sich eines der Auswahlbilder mit kleinen Veränderungen, welches die Zielperson in Abfrage 3,4 und 5 darstellt. Ebenso im Set von Emma Watson, ist bei der Abfrage drei das Vergleichsbild auch ein Auswahlbild. [20]

Im Gegensatz zu den KI's können die Super-Recognizer auch bei allen Veränderungen zumindest noch ein Gesicht wahrnehmen, wenn auch nicht immer erkennen ob es sich um die Person auf dem Vergleichsbild handelt. Betrachtet man das Set 8 zu Reese Witherspoon konnten die Super-Recognizer minimum 2 mal und maximum 6 mal die Vergleichsperson finden. Wobei wiederum weder die KI *Face Recognition* noch *FaceX-Zoo* eines der sechs Bilder am ähnlichsten zum Vergleichsbild deklariert hat. [20]

5 Methodik

5.1 Datensatzerstellung

5.1.1 Erstellung

Für die Erstellung des Datensatzes wurden Bilder aus verschiedenen öffentlich zugänglichen Datenbanken verwendet. 5 verschiedene Kategorien sollen mit dieser Studie getestet werden.

5.1.1.1 Komplex 1

Für die erste Kategorie der Geschwisterpaare wird die Datenbank „SiblingsDB“ verwendet [37]. Diese Datenbank enthält zwei verschiedene Datensätze mit Bildern von Personen, die durch eine Geschwisterbeziehungen miteinander verbunden sind. Eine davon hat nochmal Untergliederungen in der verschiedene Datensätze sind.[37]

Der erste Datensatz, *HQfaces* genannt, enthält qualitativ hochwertige Bilder von 184 Personen, darunter 92 Geschwisterpaare. Bei 79 Paaren gibt es zusätzlich zu den Frontalbildern auch Profilbilder, und 56 Paare haben sowohl lächelnde Frontal- als auch Profilbilder. Die Bilder wurden von einem professionellen Fotografen vor einem einheitlichen Hintergrund und mit kontrollierter Beleuchtung aufgenommen. Alle haben eine Auflösung von 4256 x 2832 Pixel. Die abgebildeten Personen sind freiwillige Studenten und Angestellte des Politecnico di Torino sowie deren Geschwister im Alter von 13 bis 50 Jahren, mit einem Durchschnittsalter von 23,1 Jahren. Der größte Altersunterschied zwischen Geschwistern beträgt 30 Jahre, der durchschnittliche 4,6 Jahre. Alle Probanden sind kaukasisch, und 57% von ihnen sind männlich. Die Probanden wurden gebeten, kein Make-up zu tragen. Die Bilder wurden in drei individuelle Datensätze (IDS) eingeteilt, nämlich HQf, HQfp und HQfps: [37]

- HQf: frontale ausdruckslose Bilder von 184 Personen (92 Geschwisterpaare);
- HQfp: 158 Individuen, jeweils repräsentiert durch ein ausdrucksloses frontales und ein profil Bild (79 Geschwisterpaare);
- HQfps: 112 Individuen, jeweils ein Set von vier Bildern pro Individuum. Zwei ausdruckslose Frontal- und Profilbilder und zwei lächelnde Frontal- und Profilbilder (56 Geschwisterpaare)

Alle Bilder sind mit der Position von 76 Landmarken auf Frontalbildern bzw. 12 Landmarken auf Profilbildern beschriftet. Für jede Person sind auch Informationen zu Geschlecht, Geburtsdatum und Alter. Zudem wurde schon eine Abfrage durchgeführt, in der menschliche Bewerter abstimmen sollten, ob die Paare Geschwister darstellen oder nicht und deren Stimmen sind auch verfügbar. [37] Der zweite Datensatz mit der Bezeichnung *LQfaces* enthält 98 Geschwisterpaare (196 Personen), die im Internet gefunden wurden. Bei den meisten Personen handelt es sich um berühmte Persönlichkeiten. Die Fotos haben unterschiedliche Auflösungen, der Durchmesser liegt zwischen einigen Hundert bis zu mehr als 3.000 Pixeln. Die Posen sind halb-frontal. Die Gesichter zeigen oft Mimik z.B. Lächeln, und die Bilder wurden unter verschiedenen Lichtverhältnissen aufgenommen. Profilbilder sind in *LQfaces* nicht verfügbar. Die Personen sind zu 45,5 % männlich und überwiegend kaukasisch. Die Position der 76 frontalen Gesichtsmerkmale ist ebenfalls angegeben, aber dieser

Datensatz enthält keine Altersangaben und keine Bewertungen von von menschlichen Bewerter, da dieser Datensatz hauptsächlich aus bekannten Persönlichkeiten besteht und daher wahrscheinlich zu verzerrten Bewertungen führt.[37] Die Auswahl der Bilder wurde ohne klare Kriterien getroffen. Es wurden 6 Geschwisterpaare ausgewählt und 4 Paare zusammengestellt die keine Geschwister sind.

5.1.1.2 Komplex 2

Für die GesichtsfILTER wurden ebenfalls wie bei der Studie der Hochschule Mittweida Bilder aus dem LFW-Datensatz ausgewählt (zuvor beschrieben in Kapitel 4.1). Bilder für die Zielperson wurden so ausgewählt, dass mindestens zwei von dieser Person vorhanden sind, um eines bearbeiten zu können und das andere in der Rohform zu verwenden. Zu den jeweiligen Zielpersonen wurden dann noch jeweils 9 weitere Personen selben Geschlechts und ähnlicher phänotypischer Merkmale ausgewählt.

Zur Bearbeitung der Bilder wurde die Messaging-App für Bilder „Snapchat“³ verwendet. Im Januar 2024 lag die Anzahl der Nutzer bei 750 Millionen [33]. Die Anwendung bietet verschiedene Filter, mit denen die Benutzer ihren Hutton ändern, feine Linien und Falten glätten, die Größe ihrer Augen, Lippen und Wangen verändern und verschiedene Aspekte ihres körperlichen Aussehens anpassen können. [26]

Für die Studie wurden unterschiedliche dieser Filter verwendet, alle veränderten die Gesichter. In Abbildung 5.1 sind 5 der verwendeten Filter gezeigt.



Abbildung 5.1: Beispiel der verwendeten GesichtsfILTER

Bei der ersten Frage sind die Bilder in Schwarz-Weiß, mit Schmetterlingen auf der Nase und den Wangen, und es wurde ein Weichzeichner darübergelegt (siehe Bild 1 in Abbildung 5.1). In Frage zwei wurden die Bilder mit einem Filter bearbeitet, der einen Weichzeichner, Sternchen auf den Wangen und der Nase sowie eine leichte Reduktion des Farbkontrasts beinhaltet (siehe Bild 2 in Abbildung 5.1). Der Filter, der bei Frage drei verwendet wurde, legt Sonnenstrahlen und einen gelblichen Farbfilter auf die Bilder und schmälerte zudem die Nase leicht (siehe Bild 3 in Abbildung 5.1). Im Gegensatz zu den vorhergehenden Bearbeitungen werden die Bilder bei Frage 4 mit einem Verzerr-Filter verändert. Dieser zieht den Mund sehr breit und verändert dadurch auch die gesamte Gesichtsform (siehe Bild 4 in Abbildung 5.1). In Frage fünf verändert der Filter die Gesichtsform; er vergrößert die Lippen, verkleinert die Nase, macht das Gesicht insgesamt etwas runder und betont die Augen beziehungsweise die Wimpern stärker (siehe Bild 5 in Abbildung 5.1). Ab Frage 6 wurden die zehn Bilder jeweils mit unterschiedlichen Filtern bearbeitet. Bild 1 wurde mit einem Farbfilter bearbeitet, der das gesamte Bild etwas bräunlicher erscheinen lässt. Bild 2 wurde mit demselben Filter wie in Frage 2 bearbeitet. Bild 3 zeigt im ganzen Gesicht Schmetterlinge; das Bild ist weichgezeichnet und die Farben wirken intensiver. Bei Bild 4 wurde der Filter von Frage 3 verwendet. Bild 5 ist farblich gelblicher bearbeitet, das Gesicht erscheint runder und es sind zwei

³<https://www.snapchat.com/de-DE>

Teufelshörner auf der Stirn zu sehen. Bild 6 wurde mit einem Filter bearbeitet, der weichzeichnet und die Nase kleiner macht. Bild 7 ist in Schwarz-Weiß. Bild 8 wirkt orangener als das Original, hat kleine schwarze Herzen auf den Wangen und Falten im Gesicht wurden weichgezeichnet. Das Gesicht in Bild 9 zeigt Flammen auf den Wangen, wirkt orangener und ist weichgezeichnet. In Bild 10 macht der Filter das Gesicht runder, die Nase kleiner, und es wurde ein Weichzeichner hinzugefügt, der den Wangen einen leicht rosigen Teint verleiht. In Frage 7 wurde Bild 1 mit einem Weichzeichner und Sternchen auf den Wangen und der Nase bearbeitet. Bild 2 wurde mit demselben Filter bearbeitet wie Bild 3 aus Frage 6. Bild 3 hat einen bräunlichen Farbfilter und Halbmonde auf den Wangen. Bild 4 hat denselben Filter wie Bild 10 aus Frage 6. Bild 5 wurde mit demselben Filter bearbeitet wie Bild 7 in Frage 6. Bild 6 wurde mit einem Filter bearbeitet, der Herzchen über dem Kopf platziert. Bild 7 hat denselben Filter wie Bild 6 bei Frage 6. Bild 8 wurde mit derselben Bearbeitung wie in Frage 3 bearbeitet. Bild 9 ist farblich milchiger, auf dem kompletten Gesicht sind Herzen, es ist ein weichgezeichnet und die Augen sind etwas größer und runder bearbeitet. Bild 10 ist mit dem selben Filter bearbeitet wie Bild 4.

In Frage 8 wurde Bild 1 mit demselben Filter bearbeitet wie Bild 2 aus Frage 7. Bild 2 ist kontrastärmer, und das Gesicht erscheint etwas runder. Bild 3 hat denselben Filter wie Bild 9 aus Frage 7. Bild 4 wurde ähnlich bearbeitet wie die Bilder aus Frage 3. In Bild 5 ist das Gesicht runder, die Lippen sind größer, und das Bild ist weichgezeichnet. Bild 6 wurde mit einem Filter bearbeitet, der das Bild orangener macht, das Gesicht runder erscheinen lässt und die Nase etwas kleiner wirken lässt. Bild 7 hat einen Effekt, der rote Wangen imitiert. Bild 8 wurde genauso bearbeitet wie Bild 8 aus Frage 6. Bild 9 ist in Schwarz-Weiß und hat pinke Herzen über dem Kopf. Bild 10 wurde mit demselben Filter bearbeitet wie die Bilder in Frage 2.

Die 10 Bilder in Frage 9 sind wie folgt bearbeitet: Bild 1 ist genauso bearbeitet wie Bild 8 aus Frage 6, Bild 2 wie Bild 6 aus Frage 6, Bild 3 wie Bild 10 aus Frage 6, Bild 4 wie Bild 9 aus Frage 7, Bild 5 wie Bild 6 aus Frage 8, Bild 6 wie die Bilder in Frage 3, Bild 7 wie Bild 7 aus Frage 6, Bild 8 wie Bild 3 aus Frage 6, Bild 9 wie die Bilder aus Frage 2, und Bild 10 wie Bild 9 aus Frage 8.

In Frage 10 sind die Bilder wie folgt bearbeitet: Bild 1 ist bearbeitet wie Bild 7 aus Frage 6, Bild 2 wie Bild 5 aus Frage 6, Bild 3 wurde heller, rosiger und weichgezeichnet. Bild 4 wurde wie die Bilder aus Frage 6 bearbeitet, Bild 5 wie die Bilder aus Frage 2, und Bild 6 hat einen Filter, der das Gesicht runder erscheinen lässt, die Nase verkleinert, das Bild weichzeichnet und den Kontrast verringert. Bei Bild 7 wurden durch den Filter Herzchen über dem Kopf hinzugefügt, und das Bild ist etwas rosiger und kontrastärmer. Bild 8 ist bearbeitet wie die Bilder aus Frage 3, Bild 9 wie Bild 3 aus Frage 6, und Bild 10 wie Bild 8 aus Frage 6.

5.1.1.3 Komplex 3

Die Bilder im Komplex Faceswap kommen aus der „FEI Face Database“. Die Datenbank besteht aus einer Reihe von Gesichtsbildern, diese wurden zwischen Juni 2005 und März 2006 im Artificial Intelligence Laboratory der FEI in São Bernardo do Campo, Brasilien aufgenommen. Es sind 200 Personen abgebildet, von jeder gibt es 14 Bilder, insgesamt 2800. Alle Bilder sind farbig und wurden vor einem einheitlich weißen Hintergrund aufgenommen. Die Personen sind in aufrechter Frontalposition mit einer Profildrehung von bis zu etwa 180 Grad gezeigt. Die Originalgröße jedes Bildes beträgt 640x480 Pixel. Die Teilnehmer sind hauptsächlich Studenten und Mitarbeitern der FEI, im Alter zwischen 19 und 40 Jahren, mit unterschiedlichem Aussehen, Frisuren und Accessoires. Die Verteilung der männlichen und weiblichen Personen ist genau gleich und beträgt jeweils 100. [36] Beim Komplex Faceswapping wurde ein Gesicht einer Person auf eine andere Person gelegt und et-

was anglichen, sodass es das echte sein könnte. Dies wurde mit der Webseite MioCreate⁴ gemacht. Abbildung 5.2 zeigt wie auf das Gesicht von Person 1 das Gesicht von Person 2 daraufgelegt wurde so dass Bild Nummer 3 entstanden ist.



Abbildung 5.2: Gesicht von (2) auf (1) legen ergibt (3)

5.1.1.4 Komplex 4

Die Bilder aus Komplex zur Gesichtsalterung kommen wieder aus der „FEI Face Database“, deren Aufbau im vorherigen Abschnitt zu Komplex 3 beschrieben wurde. Die Gesichtsalterung wurde mithilfe der Demo von Style-based Age Manipulation (SAM)⁵ durchgeführt, diese kann eine Altersumwandlung vornehmen. Diese Altersumwandlung ist der Prozess, bei dem die Veränderung des Aussehens einer Person in verschiedenen Altersstufen dargestellt wird und ihre Identität erhalten bleibt.[1]

Um dies umzusetzen gibt es zwei verschiedene Methoden, zum einen die Bild-zu-Bild-Übersetzung und zum anderen die Manipulation des latenten Raums.[1]

Um die Altersumwandlung als Bild-zu-Bild-Problem zu modellieren, verwendet man altersannotierte Daten und lernt eine Zuordnung zwischen vordefinierten Altersgruppen. Da zwischen den Altersgruppen eine starke Korrelation vorliegt, hat diese Methode Schwierigkeiten bedeutungsvolle Veränderungen zwischen verschiedenen Altersstufen zu modellieren. Hinzu kommt, dass das Sammeln von altersannotierten Daten im großen Maßstab oft mühsam ist. [1]

Die zweite Methodik um eine Altersumwandlung zu vollziehen, erforscht die Semantik des latenten Raums eines gut trainierten GANs, wie StyleGAN, und führt eine Traversierung des latenten Raums durch, um das gewünschte transformierte Bild zu erhalten. Methoden die diese Variante verwenden haben zwar vielversprechende Bearbeitungsergebnisse bei synthetischen Bildern gezeigt, die von StyleGAN generiert wurden, sind jedoch nicht so erfolgreich, realistische Transformationen an realen Bildern vorzunehmen.[1]

SAM verwendet die Methode die Altersumwandlung als Bild-zu-Bild-Übersetzungsproblem zu sehen. Dabei wird die Ausdruckskraft eines vortrainierten, fixierten StyleGAN-Generators mit einer Encoder-Architektur kombiniert. Der Encoder hat die Aufgabe, ein Eingabegesichtsbild direkt in eine Reihe von Stilvektoren zu kodieren, die der gewünschten Altersänderung unterliegen. Diese Stilvektoren werden dann in StyleGAN eingespeist, um das Ausgabebild zu erzeugen, das die gewünschte Altersumwandlung darstellt. Dadurch kann die hochmoderne Bildqualität genutzt werden, die von StyleGAN erreicht wird. Um den Encoder explizit bei der Generierung der entsprechenden latenten Codes anzuleiten, wird ein festes, vortrainiertes Altersregressionsnetzwerk als zusätzlichen Zwang während des Trainingsprozesses verwendet. [1]

⁴<https://de.miocreate.com/face-swap.html>

⁵<https://replicate.com/yuval-alaluf/sam>

Dabei werden nicht mit beschrifteten Daten gearbeitet, sondern verwendet bereits verfügbare vortrainierte Netzwerke, wie ein StyleGAN2 Generatorkomplex, das auf Gesichtsaufnahmen trainiert wurde, ein vortrainiertes Encoder-Netzwerk und ein ArcFace Netzwerk für Altersregression. [1]

Dieser Ansatz geht nicht von der Existenz mehrerer Altersklassen oder -domänen aus. Stattdessen ist dieser in der Lage, das entsprechende Bild zu erzeugen, wenn ein einzelnes Eingabebild und ein gewünschtes Zielalter vorgegeben sind. Im Vergleich zu Multidomänen-Ansätzen, die auf vordefinierte Altersgruppen oder Ankerklassen angewiesen sind, ermöglicht die Betrachtung des menschlichen Alterns als kontinuierlicher Regressionsprozess eine feinere Kontrolle über die gewünschte Transformation. [1]

5.1.1.5 Komplex 5

Im letzten Komplex wurde der „Caltech Face Dataset“ von 1999 für die Bilder mit Rauschfiltern und die mit Kontrast und Schärfe Änderung verwendet. Dieser besteht aus 450 Gesichtsaufnahmen, von etwa 27 verschiedenen Personen. Die Größe der Bilder beträgt 896 x 592 Pixel im JPEG-Format. Die Personen wurden mit unterschiedlichen Beleuchtungen, Ausdrücken und Hintergründen fotografiert. [39]

Für die Veränderung der Auflösung der Bilder wurde die Georgia Tech Face-Datenbank verwendet. Darin sind Bilder von 50 Personen enthalten die 1999 am „Center for Signal and Image Processing“ des Georgia Institute of Technology aufgenommen wurden. Von jeder Person gibt es 15 Farbbilder im JPEG-Format mit unruhigem Hintergrund dargestellt. Die Bilder sind nicht einheitlich und zeigen Frontal- und/oder Schrägansichten der Gesichter mit unterschiedlichen Gesichtsausdrücken, Lichtverhältnissen und Größen.[13]

Die Bilder wurden so bearbeitet, dass eine schlechtere Bildqualität durch Rauschen, verringerte Auflösung oder verringerter Kontrast entsteht.

Die Auflösung der Bilder wurde mithilfe eines Python-Codes verändert. Dies wird mithilfe der PIL-Bibliothek (Pillow) durchgeführt. Der Prozess umfasst das Öffnen der Bilder, die Anpassung ihrer Größe und das Speichern der neuen Bilder an einem definierten Speicherort.

In Abbildung 5.3 ist links ein Originalbild mit den Maßen 131 x 201 Pixel gezeigt und 4 Bilder mit herunter gerechneter Bildauflösung. Von 50% des Originalbild bis zu 10% des Originalbild ganz rechts in der Abbildung 5.3.



Abbildung 5.3: Originalbild (links), reduzierte Auflösung (50%, 40%, 30%, 20%, 10%)

Auch für das Bearbeiten der Bilder mit einem Rauschen wurde ein Python-Code verwendet. Dies wird mit Hilfe von Bibliotheken wie `os`, `PIL (Pillow)` und `numpy` umgesetzt. Der Code definiert verschiedene Rauschtypen, darunter Gaußsches Rauschen, Salt-and-Pepper-Rauschen, Poisson-Rauschen und Speckle-Rauschen. Genutzt wurde das Salt-and-Pepper-Rauschen sowie Speckle-Rauschen. Salt-and-Pepper-Rauschen besteht daraus, dass zufällige Pixel werden auf Weiß (Salz) oder Schwarz (Pfeffer) gesetzt, basierend auf einem Verhältnis und einer Menge, die das Verhältnis von Salz zu Pfeffer und die Menge des Rauschens bestimmen.



Abbildung 5.4: über das Originalbild (links) wurde ein Salt-and-Pepper Rauschen darüber gelegt (rechts)

Speckle-Rauschen ist eine Art von Rauschen, das häufig in medizinischen Bildgebungen wie Ultraschallbildern oder synthetischen Apertur-Radarbildern vorkommt. Dieses Rauschen wird erzeugt, indem ein zufälliges Rauschmuster auf die ursprünglichen Bilddaten multipliziert wird. [28]



Abbildung 5.5: über das Originalbild (links) wurde ein Speckle Rauschen darüber gelegt (rechts)

Der Kontrast und die Schärfe wurden durch ein Python-Skript reduziert. Dazu werden die Module `'os'`, `'Pillow'` und `'time'` verwendet. Die Hauptfunktion des Skripts reduziert zunächst den Kontrast und anschließend die Schärfe des Bildes und gibt das bearbeitete Bild zurück. In [Abbildung 5.6](#) ist eine Kontrast und Schärfe Reduzierung gezeigt.



Abbildung 5.6: Veränderung von Kontrast und Schärfe

5.1.2 Aufbau Datensatz

Die Studie ist in 5 Aufgabenkomplexe unterteilt, die jeweils mit unterschiedlichen Themen die Gesichtserkennungsfähigkeit testen soll.

Der erste Komplex bezieht sich auf die Geschwisterbeziehung und fragt ob diese zwischen zwei Personen vorliegt oder nicht. Die zwei Personen können beide weiblich, beide männlich oder unterschiedlicher Geschlechter sein. Abbildung 5.7 zeigt ein Geschwisterpaar, wie es in der Studie vorkommen wird. Die Frage in diesem Aufgabenkomplex wird sein „Sind Person (1) und Person (2) Geschwister? “. Als Antwortmöglichkeit gibt es die Optionen „Geschwister “oder „Keine Geschwister “.



Abbildung 5.7: Geschwisterpaar

Es wird insgesamt in dem Komplex 10 Abfragen geben. 6 davon sind Geschwisterpaare und 4 sind keine. Ein Beispiel für zwei Personen ohne Geschwisterbeziehung wird in Abbildung 5.8 gezeigt.

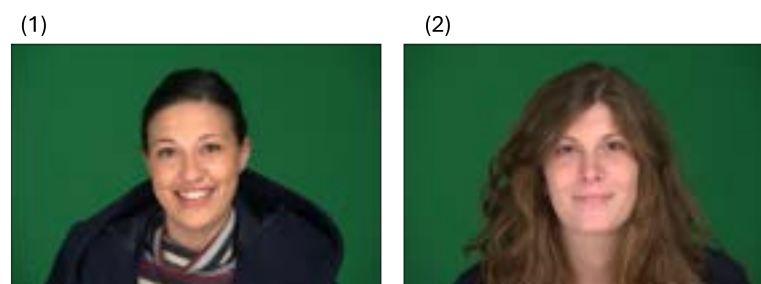


Abbildung 5.8: Kein Geschwisterpaar

Aufgabenkomplex 2,4 und 5 sind nach dem gleichen Schema aufgebaut. Es wird ein Bild einer Person geben die in den nachfolgenden Abfragen gesucht werden muss. Diese wird im folgenden auch als Zielperson bezeichnet und soll aus 10 gezeigten Auswahlbildern wiedererkannt werden.

Der zweite Komplex dieser Studie umfasst die Bildbearbeitung mit Filtern von der Social Media Plattform „Snapchat “. Es gibt insgesamt 10 Abfragen. In fünf sind die Bilder mit dem selben Filter bearbeitet (zu sehen in Abbildung 5.1) und in den anderen fünf hat jedes Bild einen anderen Filter darüber gelegt bekommen.

Der dritte Komplex der Studie befasst sich mit dem „Faceswapping “. In diesem Komplex gibt es keine Zielperson. Die Aufgabe ist es das neu erzeugte Gesicht aus 6 Bildern zu erkennen. Das heißt das „nicht echte“ Gesicht zu finden.

Die Abfragen im vierten Komplex beziehen sich auf die Veränderungen des Gesichts durch Alterung. Zu diesem Thema gibt es 9 Abfragen, bei denen die Probanden aus 10 Bildern die Zielperson wieder erkennen sollen. Bei zwei Abfragen ist ein unverändertes Bild der Zielperson gegeben und

die Auswahlbilder sind gealterte oder verjüngte Gesichter. Die restlichen 7 Abfragen bestehen aus einem bearbeiteten Bild der Zielperson und diese soll aus 10 unbearbeiteten Bildern wiedererkannt werden.



Abbildung 5.9: Gesicht (1) wurde auf 50 Jahre gealtert (2)

Im fünften und letzten Fragenkomplex soll bei Bildern mit einem Rauschen, veränderter Auflösung und reduzierter Schärfe und Kontrast die Zielperson wiedererkannt werden. Pro Veränderung, also bei Rauschen, Auflösung und Kontrast/Stärke gibt es zwei Abfragen. Diese bestehen aus einer Zielperson die vorher gezeigt wird und darauf folgen 6 bearbeitete Auswahlbilder, von denen eins die Zielperson abbildet. Ein Beispiel für die Auswahlbilder ist in [Abbildung 5.10](#) gezeigt.



Abbildung 5.10: Beispiel für Auswahlbilder mit Rauschfilter

6 Diskussion und Ausblick

6.1 Studie der Hochschule Mittweida 2023

In der Studie von letztem Jahr haben die künstlichen Intelligenzen oftmals erfolgreich die Zielperson identifizieren können, wodurch man darauf schließen kann, dass die Gesichtserkennung mit deutlichem Bildmaterial gut funktioniert. Bei stark veränderten Bildern zeigen die Ergebnisse, dass die KI öfter Schwierigkeiten hat ein Gesicht wahrzunehmen. Wenn starke Filter über das Bild gelegt wurden oder bestimmte Teile des Gesichts durch Accessoires oder Haltung der Person verdeckt sind, kann die KI wahrscheinlich nicht genügend Merkmale eines Gesichts finden. Bei Doppelgänger schneiden alle drei KIs schlecht ab und können die feinen Unterschiede nicht erkennen. Da die KIs jedoch bestimmen wie ähnlich sich zwei Gesichter sehen war es zu erwarten dass zwischen Doppelgängern, die sehr starke Ähnlichkeiten mit den Zielpersonen aufweisen, und der prominenten Person eine hohe Ähnlichkeit erkannt wird.

Im Gegensatz dazu können die Super-Recognizer auch bei sehr starken Veränderungen der Bilder noch genügend Merkmale finden, um ein Gesicht erkennen zu können und dieses meist auch zu einem Vergleichsbild zuordnen oder ausschließen zu können. Trotz dessen dass alle eine getestete überdurchschnittliche Gesichtserkennung haben, kam es zu unterschiedlich guten Ergebnissen. Diese könnten damit zusammenhängen ob sie haupt- oder nebenamtlich die Tätigkeit ausführen. Auch wie lange die Person schon im Polizeidienst ist und ein Gespür und Routine entwickelt haben kann. Bei der Doppelgängersuche kann es auch zu einer Verzerrung gekommen sein, da es auch ein Unterschied macht ob der Super-Recognizer die prominente Person kennt und daher vertraut mit deren Aussehen ist oder nicht. Sollte diese bekannt sein, können eher feine Merkmale zwischen Doppelgänger und prominenter Person erkannt werden. Dies kann eine weitere Ursache für die Unterschiede zwischen den einzelnen Super-Recognizern darstellen. Bei solchen Fragestellungen wäre eine Abfrage zur Vertrautheit mit den prominenten Personen sinnvoll um diese Beeinflussung mit in die Auswertung einfließen zu lassen.

Beeinflusst wurden die Ergebnisse auch durch die Auswahl der Bilder. In einigen Abfragen wiederholten sich die gleiche Abbildung einer Person oder das Bild der Zielperson war auch das Bild gleiche Bild in den Auswahlbildern. Der Super-Recognizer kann so eher aus dem Gedächtnis antworten ohne wirklich eine Gesichtsverarbeitung durchzuführen. Bei den KIs wiederum kann kein erinnern stattfinden, aber auch da ist es in dem Fall dass Zielbild und Vergleichsbild dasselbe ist, nur ein Abgleichen und es muss gar nicht bewertet werden wie ähnlich die Bilder sich sind weil diese identisch sind.

6.2 Studie in dieser Arbeit

Ausgehend von der Erfahrung und den Ergebnissen des vorherigen Tests wurde in dieser Arbeit eine weitere Studie entworfen. Diese neue Studie beleuchtet fünf verschiedene Fragestellungen, im Gegensatz zur früheren, die nur zwei verschiedene Themengebiete abdeckte. Eine klassische Personensuche kommt in drei der fünf Komplexe zum Einsatz. Jeder Komplex befasst sich mit einer Hypothese über die Super-Recognizer.

Im ersten Komplex wird getestet, ob eine Geschwisterbeziehung erkannt werden kann. Es gibt keine

Zielperson, die abgeglichen werden kann oder an die man sich erinnern muss. Stattdessen müssen anhand der Gesichtsmarkmale entschieden werden, ob zwei Personen Geschwister sind. Die Bilder wiederholen sich nicht und wurden keiner Veränderung unterzogen.

Bei Identifikation in der Polizeiarbeit wird nicht direkt nach Geschwisterbeziehungen gefragt, da ist eher das Verwandte sich ähnlich sehen und diese auseinander gehalten werden müssen [21]. Trotzdem hat es einen wissenschaftlichen Wert, um einen Überblick zu den Fähigkeit der Super-Recognizer zu bekommen.

Der zweite Komplex testet die Wiedererkennung trotz verschiedener Gesichtsveränderungen durch Social Media Gesichtsfiler.

Für Vermisstenanzeigen oder zur Identifikation von Verdächtigen werden oftmals Bilder von Sozialen Netzwerken verwendet[21]. Plattformen wie Instagram oder Snapchat bieten bereits verschiedene Filter zur Gesichtsveränderung an. Daher sind auch die Bilder zur Identifikation oft offensichtlich oder versteckt durch Filter bearbeitet. In der Studie wurden bei fünf Abfragen in diesem Komplex alle zehn Vergleichsbilder mit demselben Filter bearbeitet, um eine Konsistenz zu erreichen, die in der Studie von 2023 fehlte. Dadurch kann klar verglichen werden, wie gut verschiedene Filter bei der Identifikation wirken. Bei den verbleibenden vier Abfragen wurde jedes Foto unterschiedlich bearbeitet, um zu berücksichtigen, dass Bilder auf sozialen Netzwerken auf verschiedene Weisen bearbeitet werden und die Super-Recognizer damit arbeiten müssen. Daher ist auch die Fähigkeit zur Identifikation unter diesen Bedingungen von Interesse.

Der dritte Komplex testet die Erkennung eines KI erstellten Gesichts. Wie gut die Super-Recognizer unechte Gesichter erkennen ist z.B. im Kontext der DeepFakes interessant. Es gibt Frontalansichten und 3/4 Ansichten, eine laterale Ansicht wurde versucht umzusetzen, aber die KI hat kein brauchbares Bild zustande bekommen.

Der vierte Abschnitt testet die Fähigkeit zur Gesichtsidentifikation trotz Veränderung durch Alterungen. Auch eine Fähigkeit auf die die Super-Recognizer in der Praxis zurückgreifen müssen. Beispielsweise wenn nur ein altes Passfoto zum Abgleich vorhanden ist oder bei vermissten Fällen von Kindern, dessen Aussehen sich schon nach wenigen Jahren drastisch verändert haben kann.

Im fünften Komplex wird getestet, wie sich Kontrast, Schärfe, Rauschfilter und Auflösung auf die Leistung der Super-Recognizer bei der Gesichtserkennung auswirken. Sie müssen eine Zielperson unter jeweils sechs einheitlich veränderten Bildern wiedererkennen. Diese verschlechterten Bilder ähneln Überwachungsaufnahmen, die Super-Recognizer in ihrem Berufsalltag betrachten müssen.

6.3 Ausblick

Anschließend an diese Arbeit wäre die Anwendung der Studie durch Super-Recognizer sowie durch Künstliche Intelligenzen, wie in der Studie aus dem letzten Jahr, sinnvoll. Zusätzlich könnten eventuell neue Künstliche Intelligenzen verwendet werden, die besser auf bestimmte Aufgabenstellungen angepasst sind.

Dadurch könnten voraussichtlich Aussagen über die Leistung der Super-Recognizer in Bezug auf die fünf verschiedenen Hypothesen getroffen werden. Außerdem wäre es möglich, eine Einschätzung darüber abzugeben, in welchen Bereichen Künstliche Intelligenzen den Menschen unterstützen können oder wo der Mensch eindeutig bessere Leistungen erbringt.

Anhang A: Bilder der Studie



Abbildung A.1: Kein Geschwisterpaar 1



Abbildung A.2: Kein Geschwisterpaar 2



Abbildung A.3: Kein Geschwisterpaar 3



Abbildung A.4: Kein Geschwisterpaar 4



Abbildung A.5: Geschwisterpaar 1



Abbildung A.6: Geschwisterpaar 2

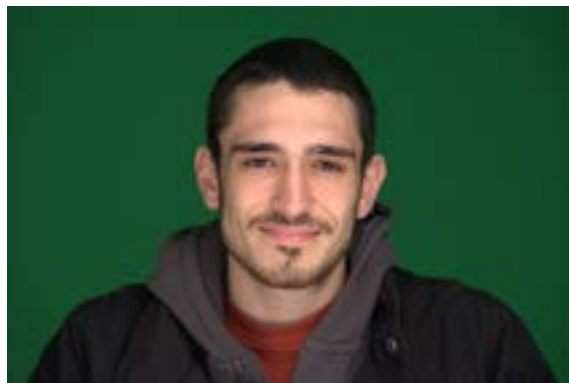


Abbildung A.7: Geschwisterpaar 3



Abbildung A.8: Geschwisterpaar 4

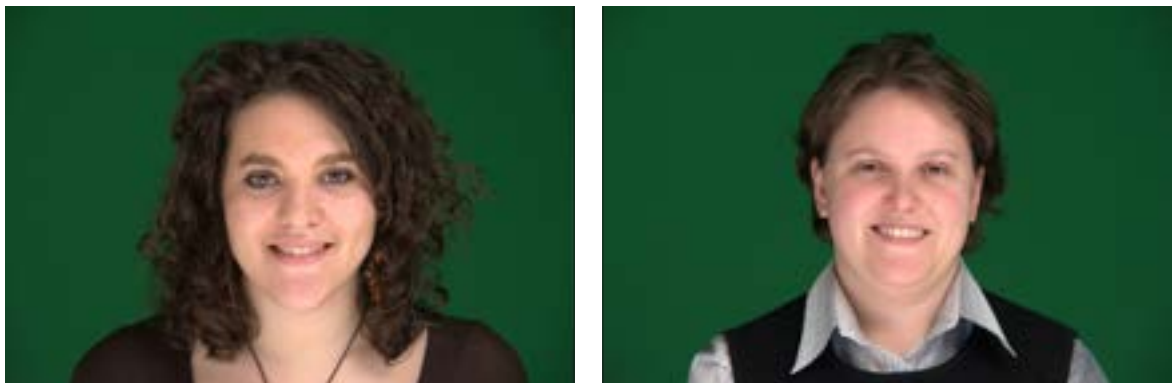


Abbildung A.9: Geschwisterpaar 5



Abbildung A.10: Geschwisterpaar 6



Abbildung A.11: Gesichtsfilter 1 (Antwort: Bild 5)



Abbildung A.12: Gesichtsfilter 2 (Antwort: Bild 1)

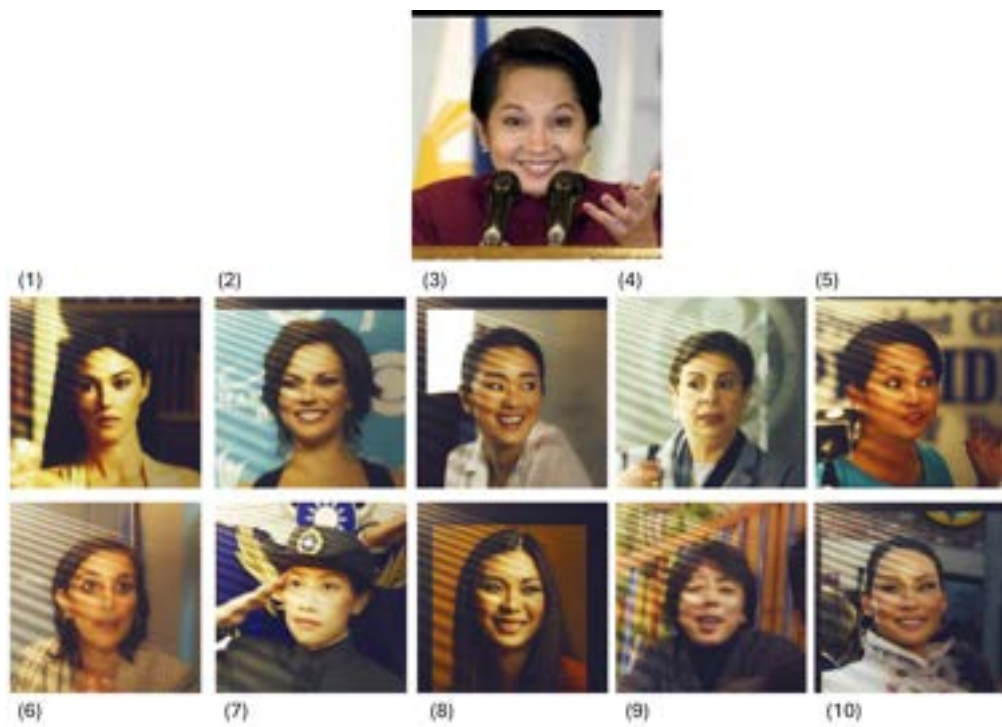


Abbildung A.13: Gesichtsfilter 3 (Antwort: Bild 5)



Abbildung A.14: Gesichtsfilter 4 (Antwort: Bild 6)



Abbildung A.15: Gesichtsfilter 5 (Antwort: Bild 1)



Abbildung A.16: Gesichtsfilter 6 (Antwort: Bild 1)

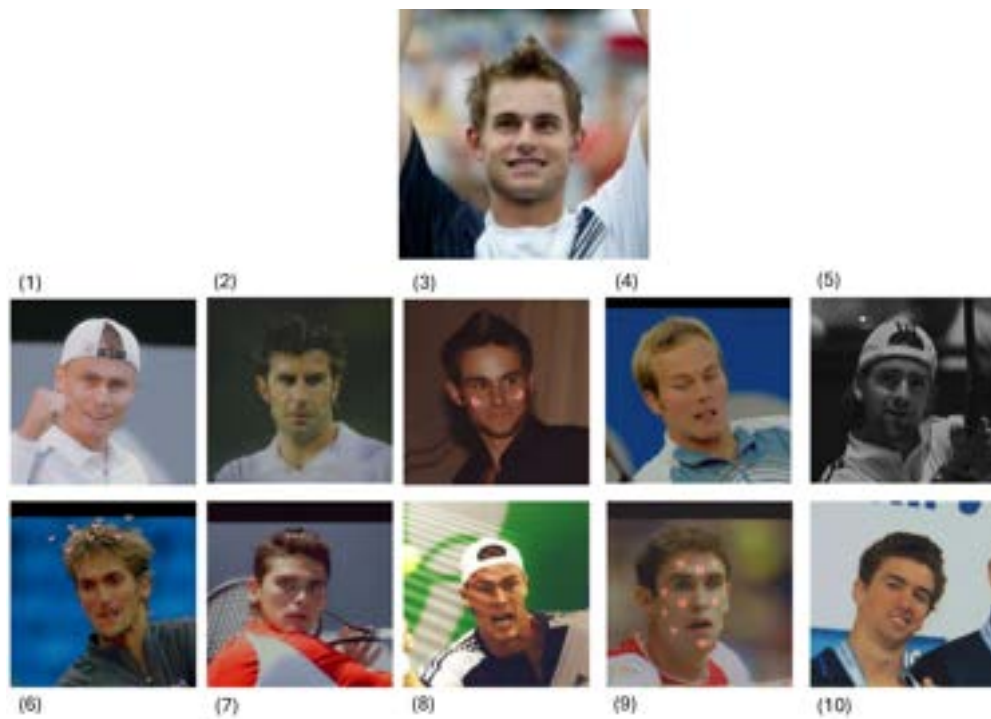


Abbildung A.17: Gesichtsfilter 7 (Antwort: Bild 3)

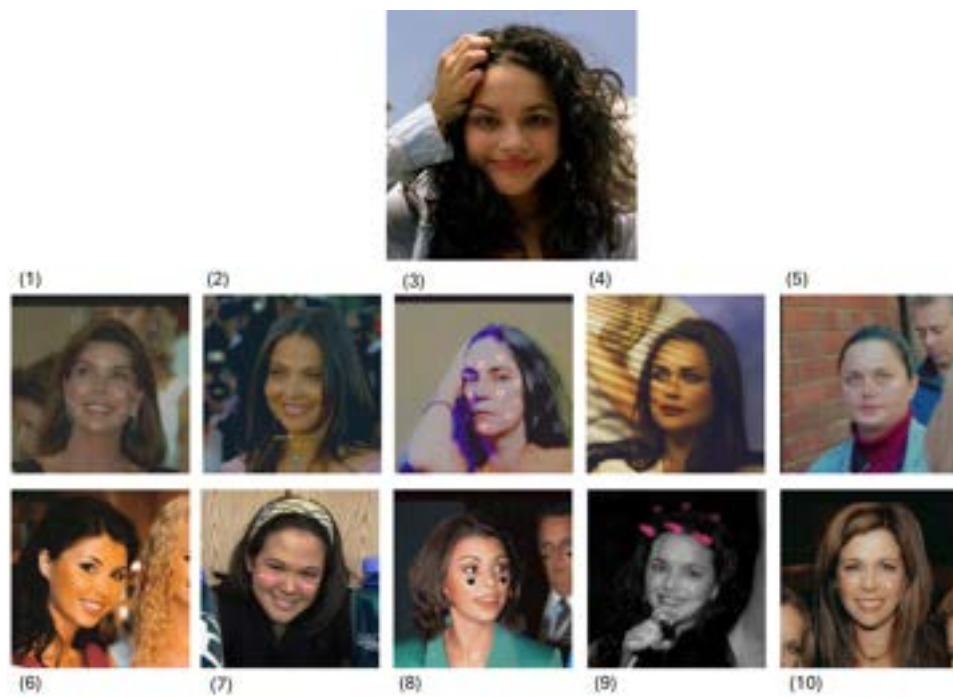


Abbildung A.18: Gesichtsfilter 8 (Antwort: Bild 9)



Abbildung A.19: Gesichtsfilter 9 (Antwort: Bild 7)



Abbildung A.20: Gesichtsfilter 10 (Antwort: Bild 3)

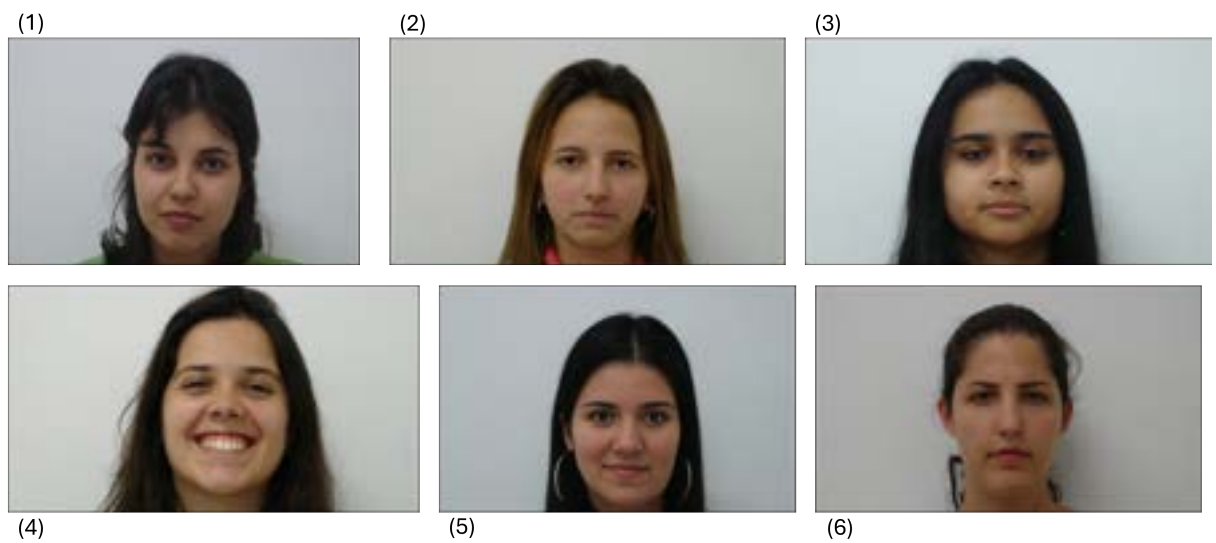


Abbildung A.21: Face-Swapping 1 (Antwort: Bild 5)

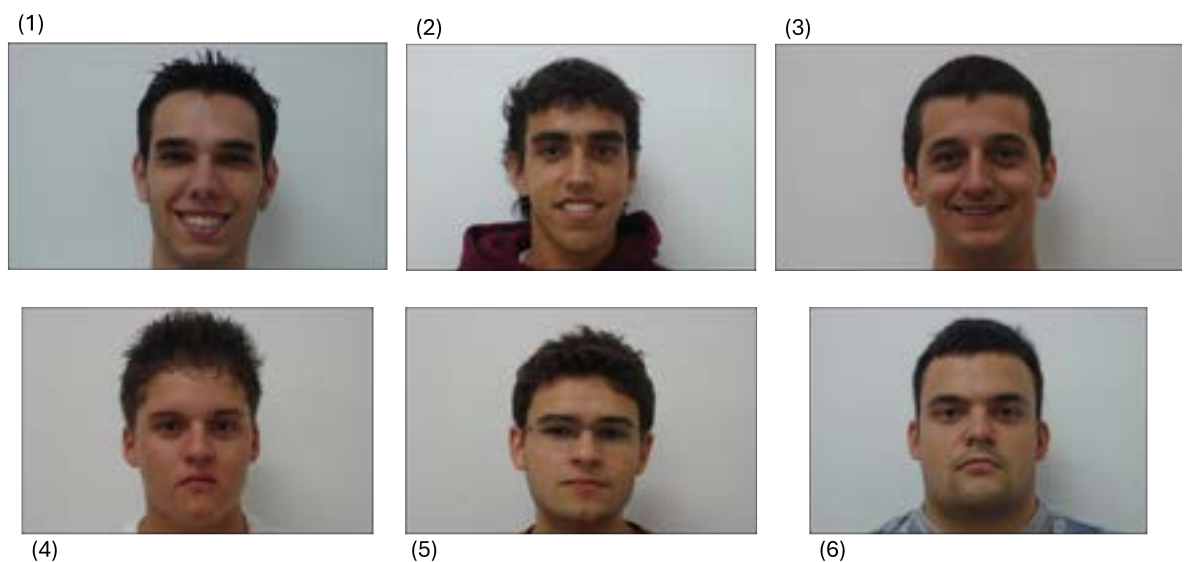


Abbildung A.22: Face-Swapping 2 (Antwort: Bild 3)



Abbildung A.23: Face-Swapping 3 (Antwort: Bild 2)

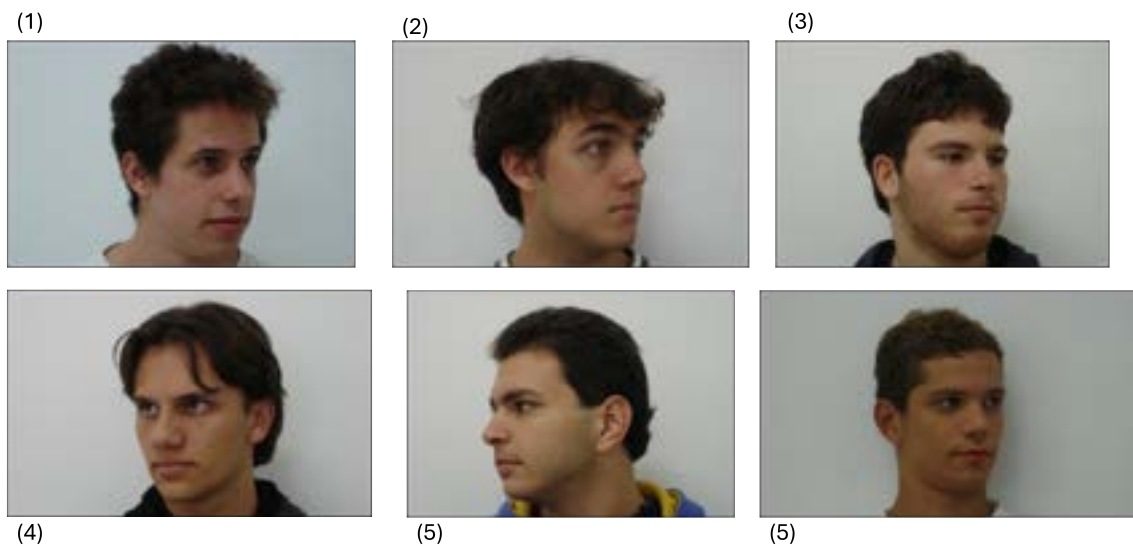


Abbildung A.24: Face-Swapping 4 (Antwort: Bild 6)

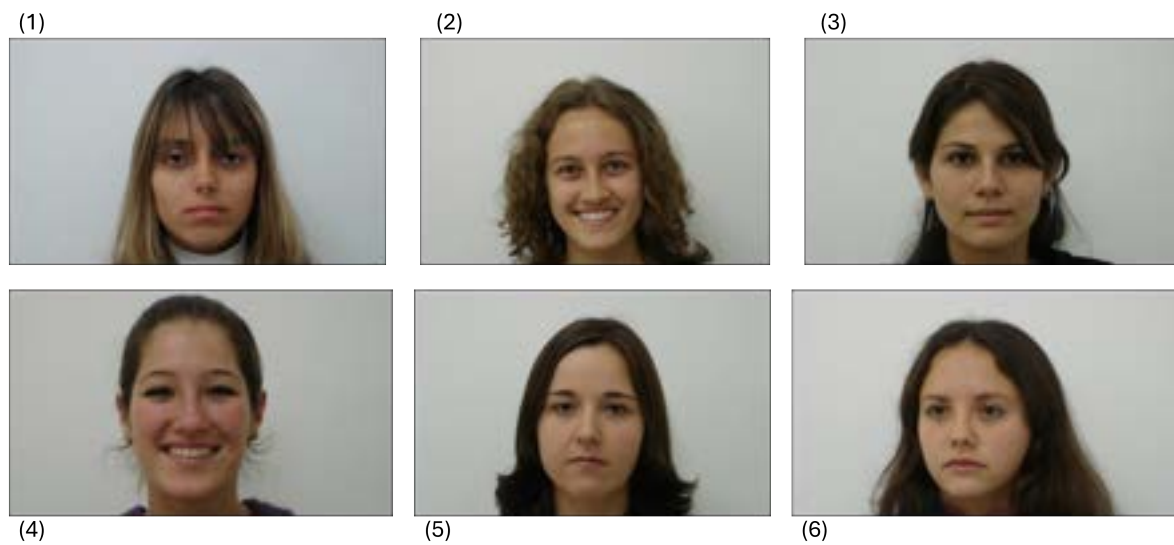


Abbildung A.25: Face-Swapping 5 (Antwort: Bild 3)

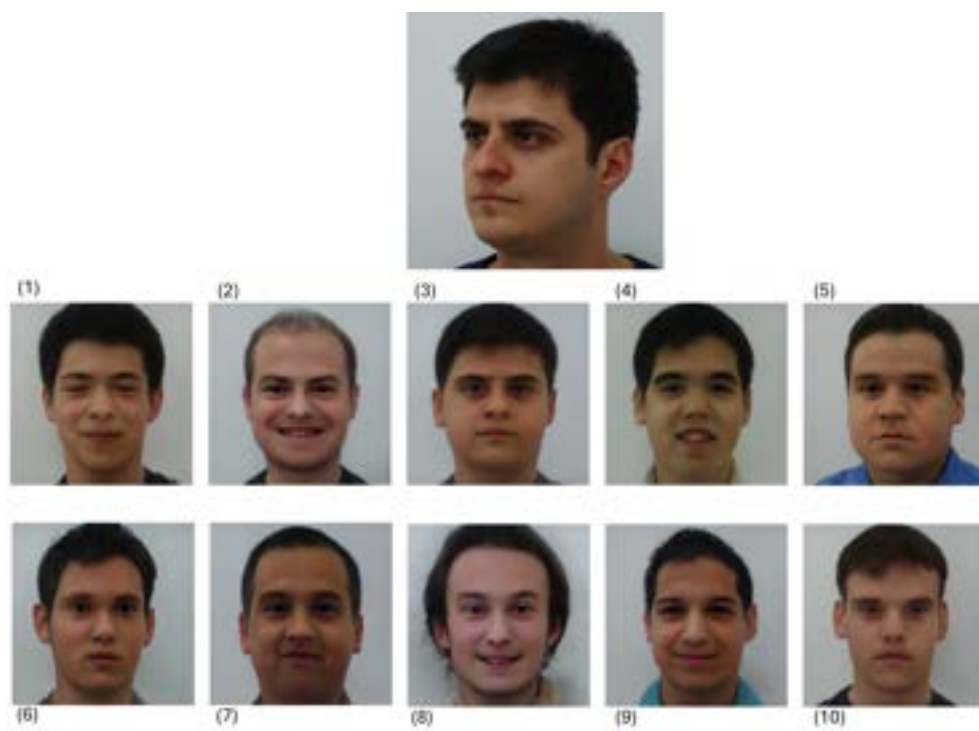


Abbildung A.26: Alterung 1 (Verjüngung auf 10 Jahre, Antwort: Bild 3)

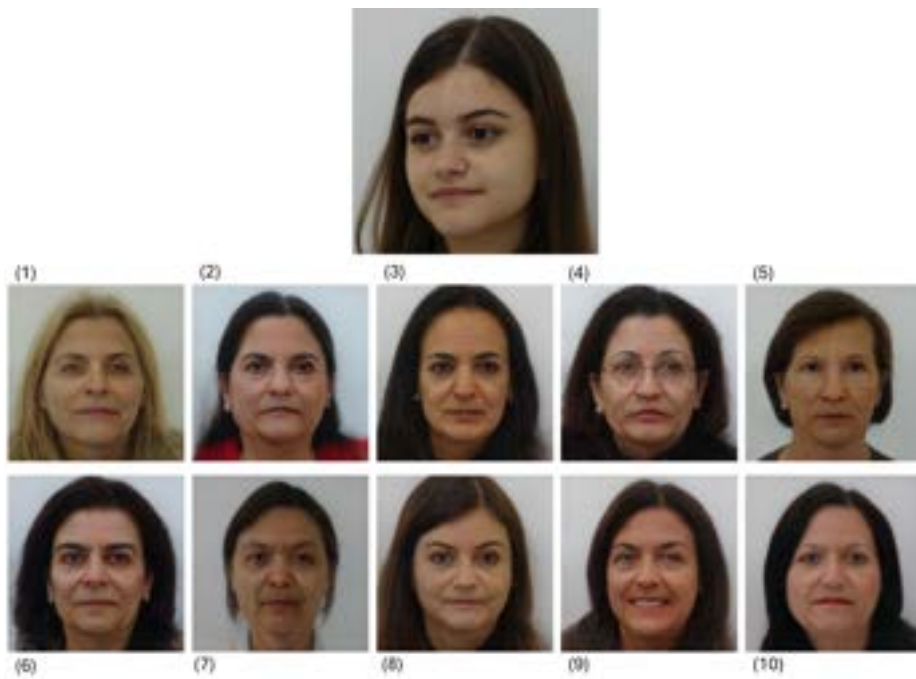


Abbildung A.27: Alterung 2 (Alterung auf 60 Jahre, Antwort: Bild 8)

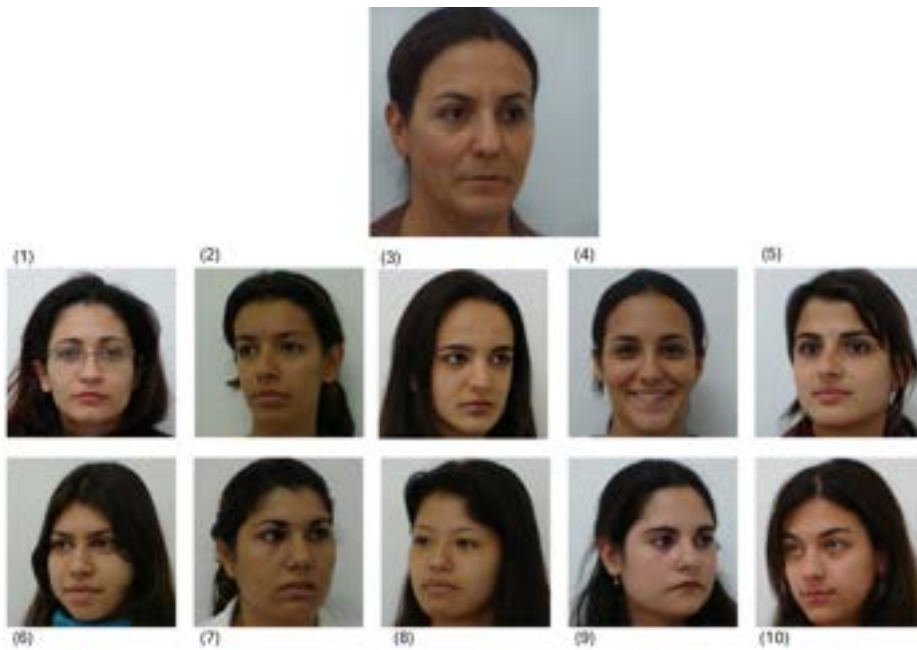


Abbildung A.28: Alterung 3 (Alterung auf 50 Jahre, Antwort: Bild 4)

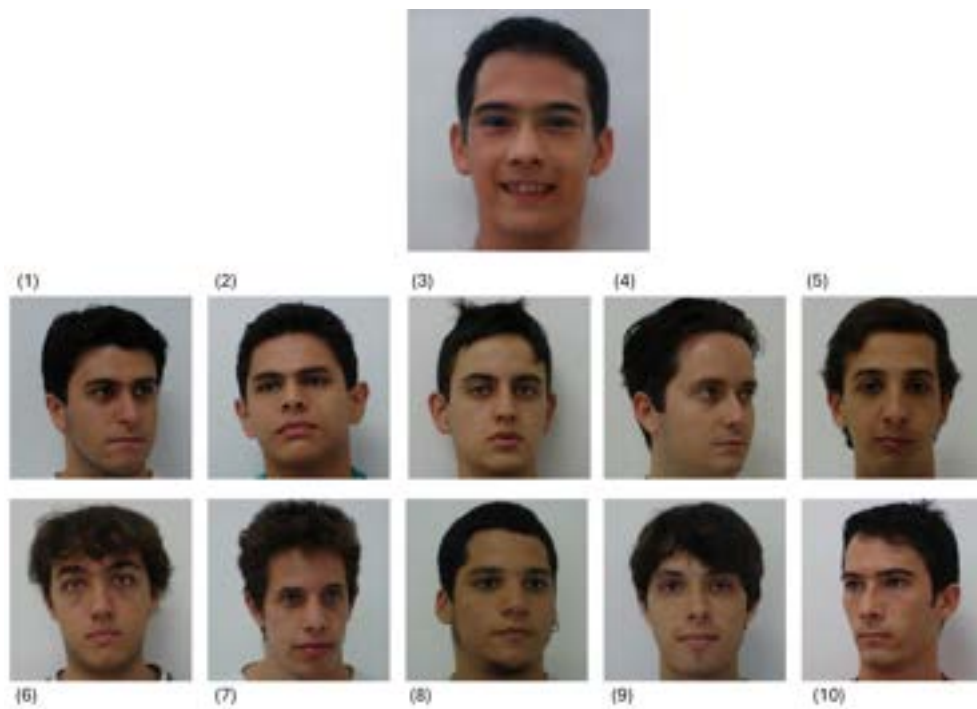


Abbildung A.29: Alterung 4 (Verjüngung auf 10 Jahre, Antwort: Bild 10)



Abbildung A.30: Alterung 5 (Alterung auf 70 Jahre, Antwort: Bild 6)



Abbildung A.31: Alterung 6 (Alterung auf 70 Jahre, Antwort: Bild 3)

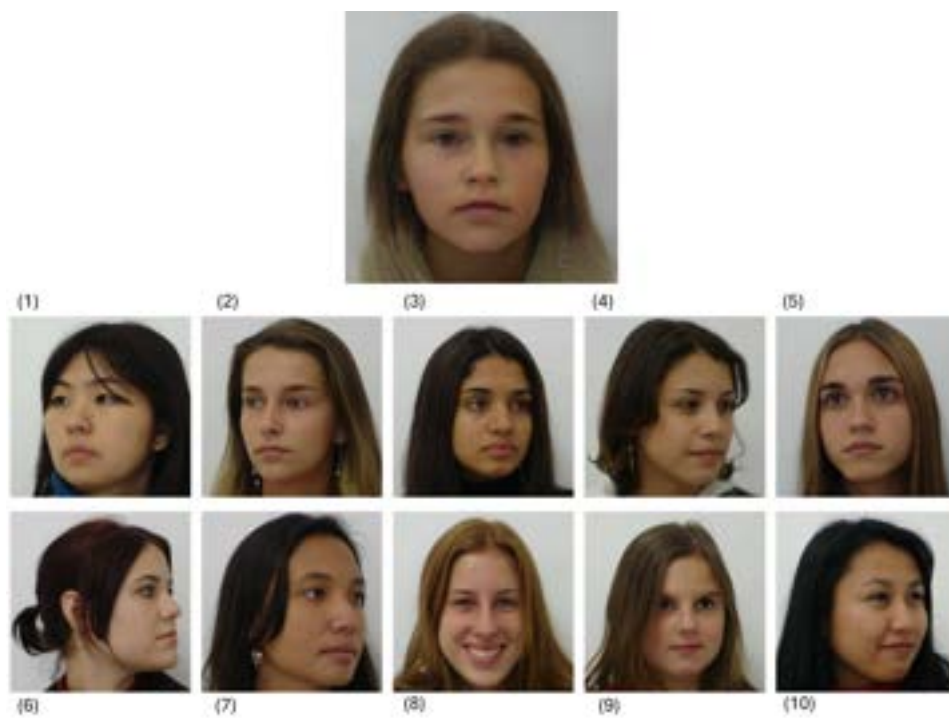


Abbildung A.32: Alterung 7 (Verjüngung auf 10 Jahre, Antwort: Bild 2)

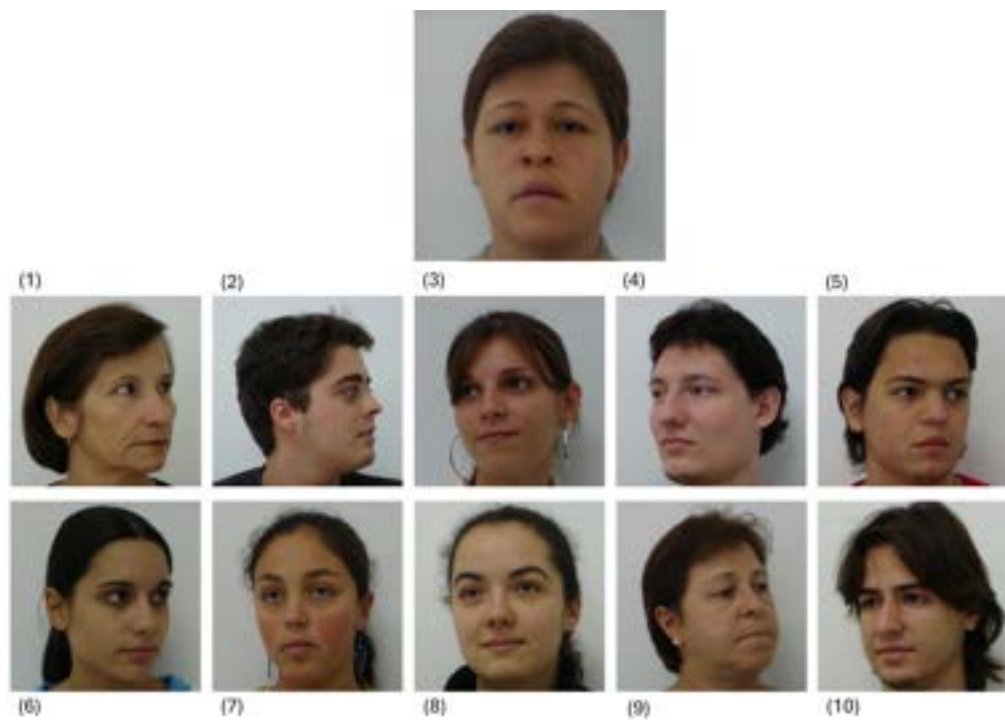


Abbildung A.33: Alterung 8 (Verjüngung auf 20 Jahre, Antwort: Bild 9)

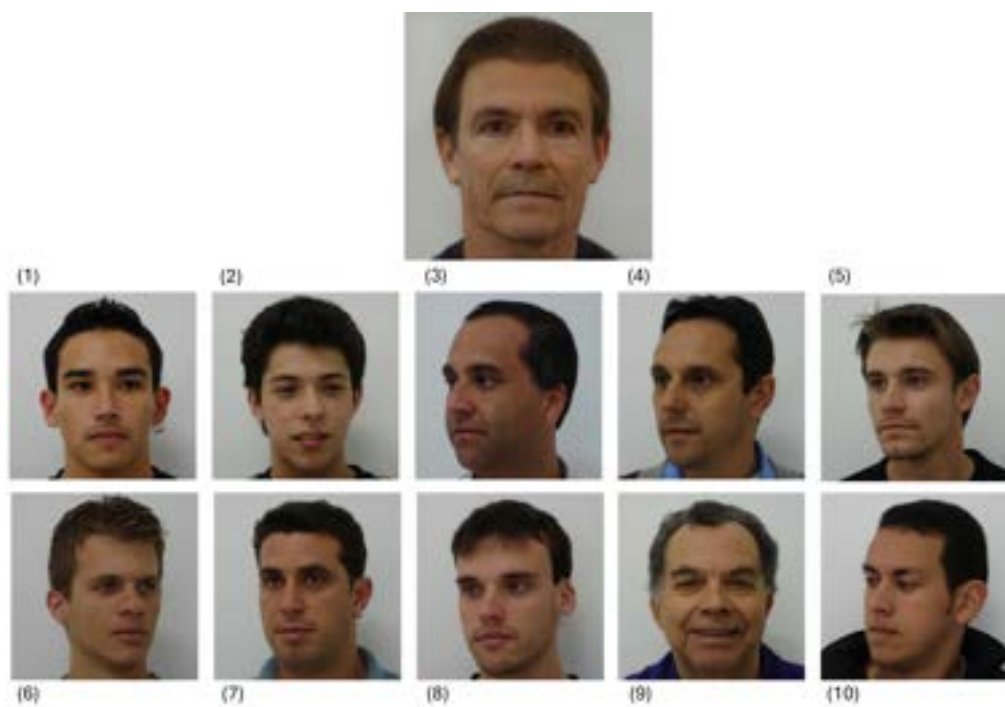


Abbildung A.34: Alterung 9 (Alterung auf 70 Jahre, Antwort: Bild 5)



Abbildung A.35: Rauschen 1 (Salt-and-Pepper Rauschen (Salt/Pepper Verhältnis: 0.5, Menge: 0.2) Antwort: Bild 5)



Abbildung A.36: Rauschen 2 (Speckle Rauschen (Intensität: 2.0) Antwort: Bild 5)

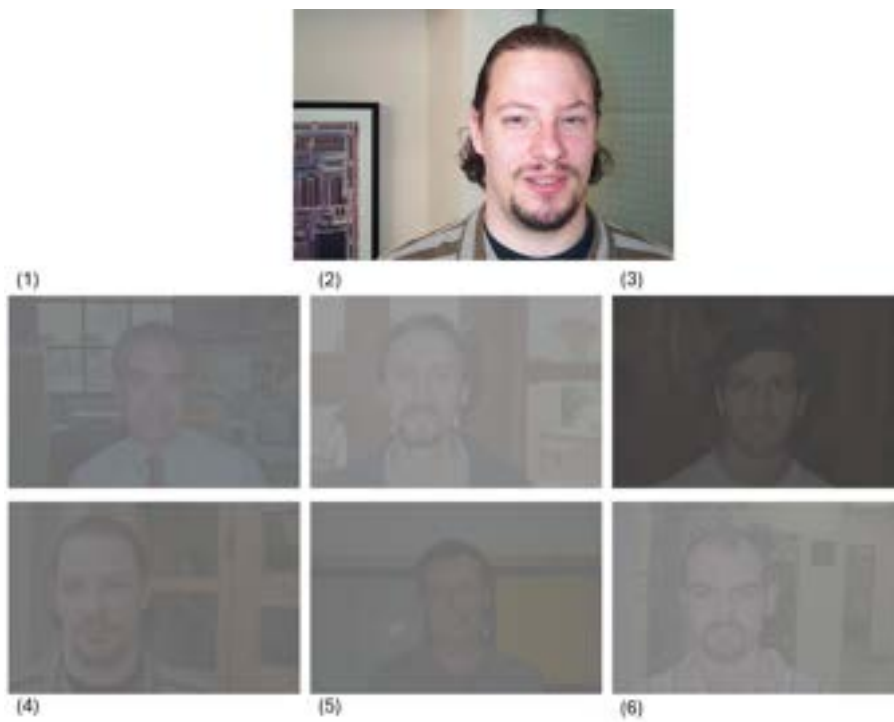


Abbildung A.37: Kontrast 1 (Kontrastfaktor 0.05, Schärfefaktor 0.4, Antwort: Bild 5)

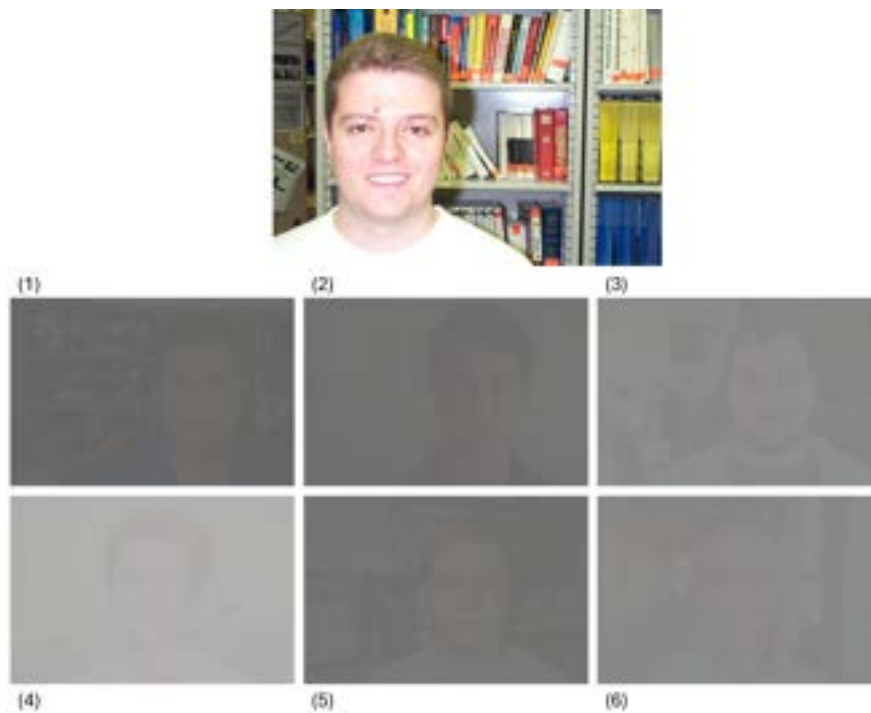


Abbildung A.38: Kontrast 2 (Kontrastfaktor 0.1, Schärfefaktor 0.1, Antwort: Bild 5)



Abbildung A.39: Auflösung 1 (Auflösung auf 20% reduziert, Antwort: Bild 4)

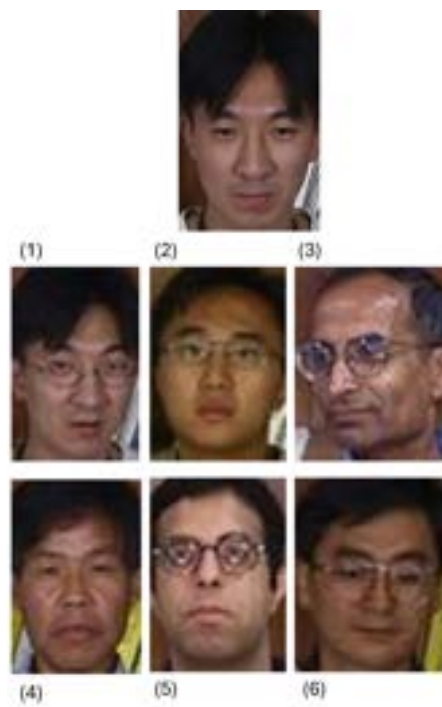


Abbildung A.40: Auflösung 2 (Auflösung auf 30% reduziert, Antwort: Bild 1)

Literaturverzeichnis

- [1] Yuval Alaluf, Or Patashnik und Daniel Cohen-Or. „Only a matter of style: age transformation using a style-based regression model“. en. In: *ACM Trans. Graph.* 40.4 (Aug. 2021), S. 1–12. ISSN: 0730-0301, 1557-7368. DOI: [10.1145/3450626.3459805](https://doi.org/10.1145/3450626.3459805). URL: <https://dl.acm.org/doi/10.1145/3450626.3459805> (besucht am 06.07.2024).
- [2] Mohamad Alansari u. a. „GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations“. In: *IEEE Access* 11 (2023), S. 35429–35446. ISSN: 2169-3536. DOI: [10.1109/ACCESS.2023.3266068](https://doi.org/10.1109/ACCESS.2023.3266068). URL: <https://ieeexplore.ieee.org/document/10098610/> (besucht am 16.06.2024).
- [3] H. P. Bahrick, P. O. Bahrick und R. P. Wittlinger. „Fifty years of memory for names and faces: A cross-sectional approach.“ en. In: *Journal of Experimental Psychology: General* 104.1 (März 1975), S. 54–75. ISSN: 1939-2222, 0096-3445. DOI: [10.1037/0096-3445.104.1.54](https://doi.org/10.1037/0096-3445.104.1.54). URL: <https://doi.apa.org/doi/10.1037/0096-3445.104.1.54> (besucht am 09.07.2024).
- [4] Vicki Bruce u. a. „Matching identities of familiar and unfamiliar faces caught on CCTV images.“ en. In: *Journal of Experimental Psychology: Applied* 7.3 (2001), S. 207–218. ISSN: 1939-2192, 1076-898X. DOI: [10.1037/1076-898X.7.3.207](https://doi.org/10.1037/1076-898X.7.3.207). URL: <https://doi.apa.org/doi/10.1037/1076-898X.7.3.207> (besucht am 22.07.2024).
- [5] Maggie Bruck, Patrick Cavanagh und Stephen J. Ceci. „Fortysomething: Recognizing faces at one's 25th reunion“. en. In: *Memory & Cognition* 19.3 (Mai 1991), S. 221–228. ISSN: 0090-502X, 1532-5946. DOI: [10.3758/BF03211146](https://doi.org/10.3758/BF03211146). URL: <http://link.springer.com/10.3758/BF03211146> (besucht am 06.02.2024).
- [6] Grigorios G. Chrysos u. a. „Deep Polynomial Neural Networks“. In: *IEEE Trans. Pattern Anal. Mach. Intell.* (2021), S. 1–1. ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: [10.1109/TPAMI.2021.3058891](https://doi.org/10.1109/TPAMI.2021.3058891). URL: <https://ieeexplore.ieee.org/document/9353253/> (besucht am 16.06.2024).
- [7] Josh P. Davis u. a. „Identification from CCTV: Assessing police super-recogniser ability to spot faces in a crowd and susceptibility to change blindness“. en. In: *Applied Cognitive Psychology* 32.3 (Mai 2018), S. 337–353. ISSN: 0888-4080, 1099-0720. DOI: [10.1002/acp.3405](https://doi.org/10.1002/acp.3405). URL: <https://onlinelibrary.wiley.com/doi/10.1002/acp.3405> (besucht am 02.02.2024).
- [8] Josh P. Davis u. a. „Investigating Predictors of Superior Face Recognition Ability in Police Super-recognisers“. en. In: *Applied Cognitive Psychology* 30.6 (Nov. 2016), S. 827–840. ISSN: 0888-4080, 1099-0720. DOI: [10.1002/acp.3260](https://doi.org/10.1002/acp.3260). URL: <https://onlinelibrary.wiley.com/doi/10.1002/acp.3260> (besucht am 01.02.2024).
- [9] Brad Duchaine und Ken Nakayama. „The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants“. en. In: *Neuropsychologia* 44.4 (Jan. 2006), S. 576–585. ISSN: 00283932. DOI: [10.1016/j.neuropsychologia.2005.07.001](https://doi.org/10.1016/j.neuropsychologia.2005.07.001). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0028393205002496> (besucht am 23.01.2024).

- [10] Bradley Duchaine, Laura Germine und Ken Nakayama. „Family resemblance: Ten family members with prosopagnosia and within-class object agnosia“. en. In: *Cognitive Neuropsychology* 24.4 (Juni 2007), S. 419–430. ISSN: 0264-3294, 1464-0627. DOI: [10.1080/02643290701380491](https://doi.org/10.1080/02643290701380491). URL: <http://www.tandfonline.com/doi/abs/10.1080/02643290701380491> (besucht am 14. 07. 2024).
- [11] A. Geitgey. *The world's simplest facial recognition api for Python and the command line*. 2022. URL: https://github.com/ageitgey/face_recognition (besucht am 17. 03. 2023).
- [12] Anjith George u. a. *EdgeFace: Efficient Face Recognition Model for Edge Devices*. arXiv:2307.01838 [cs]. Jan. 2024. URL: <http://arxiv.org/abs/2307.01838> (besucht am 16. 06. 2024).
- [13] *Georgia Tech face database*. 1999. URL: https://www.anefian.com/research/face_reco.htm.
- [14] Gary B Huang u. a. „Labeled faces in the wild: A database for studying face recognition in unconstrained environments“. In: *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*. 2008.
- [15] Insoo Kim u. a. „DiscFace: Minimum Discrepancy Learning for Deep Face Recognition“. en. In: *Computer Vision – ACCV 2020*. Hrsg. von Hiroshi Ishikawa u. a. Bd. 12626. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, S. 358–374. ISBN: 978-3-030-69540-8 978-3-030-69541-5. DOI: [10.1007/978-3-030-69541-5_22](https://doi.org/10.1007/978-3-030-69541-5_22). URL: http://link.springer.com/10.1007/978-3-030-69541-5_22 (besucht am 16. 06. 2024).
- [16] Minchul Kim, Anil K. Jain und Xiaoming Liu. *AdaFace: Quality Adaptive Margin for Face Recognition*. arXiv:2204.00964 [cs]. Feb. 2023. URL: <http://arxiv.org/abs/2204.00964> (besucht am 16. 06. 2024).
- [17] Tom Klein. *Face blind! Erfahrungen aus einem Leben mit angeborener Prosopagnosie*. ger. Wien: Festland Verlag, 2016. ISBN: 978-3-9504121-8-5.
- [18] Martin Knoche u. a. „Octuplet Loss: Make Face Recognition Robust to Image Resolution“. In: *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*. Waikoloa Beach, HI, USA: IEEE, Jan. 2023, S. 1–8. ISBN: 9798350345445. DOI: [10.1109/FG57933.2023.10042669](https://doi.org/10.1109/FG57933.2023.10042669). URL: <https://ieeexplore.ieee.org/document/10042669/> (besucht am 16. 06. 2024).
- [19] Bi Li u. a. *Dynamic Class Queue for Large Scale Face Recognition In the Wild*. arXiv:2105.11113 [cs]. Mai 2021. URL: <http://arxiv.org/abs/2105.11113> (besucht am 16. 06. 2024).
- [20] M. Lietsch u. a. „Man vs. machine: A study comparing super-recognizers and artificial intelligence“. In: 2024.
- [21] Lohse. *Experteninterview mit einem Vollzeit Super Recogniser der Polizeidirektion Chemnitz*. Mai 2024.
- [22] Jeffrey D. Nador u. a. „Face recognition in police officers: Who fits the bill?“ en. In: *Forensic Science International: Reports* 5 (Juli 2022), S. 100267. ISSN: 26659107. DOI: [10.1016/j.fsir.2022.100267](https://doi.org/10.1016/j.fsir.2022.100267). URL: <https://linkinghub.elsevier.com/retrieve/pii/S2665910722000135> (besucht am 12. 02. 2024).
- [23] S. Preuß und D. Labudde. *Einsatz von Algorithmen zur Gesichtserkennung in Kipo Verfahren – Vorstellung einer Pipeline zur automatisierten Verarbeitung*. Wilfried Honekamp und Johannes Fährdrich, 2023.

- [24] Meike Ramon, Anna K. Bobak und David White. „Super-recognizers: From the lab to the world and back again“. en. In: *British J of Psychology* 110.3 (Aug. 2019), S. 461–479. ISSN: 0007-1269, 2044-8295. DOI: [10.1111/bjop.12368](https://doi.org/10.1111/bjop.12368). URL: <https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12368> (besucht am 20. 11. 2023).
- [25] Meike Ramon, Thomas Busigny und Bruno Rossion. „Impaired holistic processing of unfamiliar individual faces in acquired prosopagnosia“. en. In: *Neuropsychologia* 48.4 (März 2010), S. 933–944. ISSN: 00283932. DOI: [10.1016/j.neuropsychologia.2009.11.014](https://doi.org/10.1016/j.neuropsychologia.2009.11.014). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0028393209004461> (besucht am 12. 02. 2024).
- [26] Kamleshun Ramphul und Stephanie G Mejias. „Is "Snapchat Dysmorphia" a Real Issue?“ en. In: *Cureus* (März 2018). ISSN: 2168-8184. DOI: [10.7759/cureus.2263](https://doi.org/10.7759/cureus.2263). URL: <https://www.cureus.com/articles/11237-is-snapchat-dysmorphia-a-real-issue> (besucht am 06. 07. 2024).
- [27] Richard Russell, Brad Duchaine und Ken Nakayama. „Super-recognizers: People with extraordinary face recognition ability“. en. In: *Psychonomic Bulletin & Review* 16.2 (Apr. 2009), S. 252–257. ISSN: 1069-9384, 1531-5320. DOI: [10.3758/PBR.16.2.252](https://doi.org/10.3758/PBR.16.2.252). URL: <http://link.springer.com/10.3758/PBR.16.2.252> (besucht am 22. 11. 2023).
- [28] Lukas Schneider. *Media.io: Entfesseln Sie die Magie der KI, um Bildrauschen zu entfernen*. Apr. 2024. URL: <https://www.media.io/de/image-denoiser-tips/mediaio-ai-image-denoiser.html>.
- [29] Sefik Serengil und Alper Özpinar. „A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules“. In: *Bilişim Teknolojileri Dergisi* 17.2 (Apr. 2024), S. 95–107. ISSN: 1307-9697. DOI: [10.17671/gazibtd.1399077](https://doi.org/10.17671/gazibtd.1399077). URL: <http://dergipark.org.tr/en/doi/10.17671/gazibtd.1399077> (besucht am 16. 06. 2024).
- [30] Sefik Ilkin Serengil. *serengil/deepface*. original-date: 2020-02-08T20:42:28Z. Juni 2024. URL: <https://github.com/serengil/deepface> (besucht am 20. 06. 2024).
- [31] Sefik Ilkin Serengil und Alper Ozpinar. „LightFace: A Hybrid Deep Face Recognition Framework“. In: *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. Istanbul, Turkey: IEEE, Okt. 2020, S. 1–5. ISBN: 978-1-72819-136-2. DOI: [10.1109/ASYU50717.2020.9259802](https://doi.org/10.1109/ASYU50717.2020.9259802). URL: <https://ieeexplore.ieee.org/document/9259802/> (besucht am 16. 06. 2024).
- [32] Sibi C. Sethuraman u. a. *MagicEye: An Intelligent Wearable Towards Independent Living of Visually Impaired*. arXiv:2303.13863 [cs]. März 2023. URL: <http://arxiv.org/abs/2303.13863> (besucht am 16. 06. 2024).
- [33] *Social Networks nach Nutzern 2024 | Statista*. 2024. URL: <https://de.statista.com/statistik/daten/studie/181086/umfrage/die-weltweit-groessten-social-networks-nach-anzahl-der-user/> (besucht am 06. 07. 2024).
- [34] Yifan Sun u. a. *Circle Loss: A Unified Perspective of Pair Similarity Optimization*. arXiv:2002.10857 [cs]. Juni 2020. URL: <http://arxiv.org/abs/2002.10857> (besucht am 16. 06. 2024).
- [35] Philipp Terhörst u. a. *QMagFace: Simple and Accurate Quality-Aware Face Recognition*. arXiv:2111.13475 [cs]. März 2022. URL: <http://arxiv.org/abs/2111.13475> (besucht am 16. 06. 2024).
- [36] Carlos Eduardo Thomaz. *FEI Face Database*. URL: <https://fei.edu.br/~cet/facedatabase.html> (besucht am 20. 06. 2024).

- [37] Tiago F. Vieira u. a. „Detecting siblings in image pairs“. en. In: *Vis Comput* 30.12 (Dez. 2014), S. 1333–1345. ISSN: 0178-2789, 1432-2315. DOI: [10.1007/s00371-013-0884-3](https://doi.org/10.1007/s00371-013-0884-3). URL: <http://link.springer.com/10.1007/s00371-013-0884-3> (besucht am 20. 06. 2024).
- [38] Jun Wang u. a. *FaceX-Zoo: A PyTorch Toolbox for Face Recognition*. Okt. 2021. URL: <http://arxiv.org/abs/2101.04407> (besucht am 08. 05. 2024).
- [39] Markus Weber. *Caltech Face Dataset 1999*. en. Juli 2022. DOI: [10.22002/D1.20237](https://doi.org/10.22002/D1.20237). URL: <https://data.caltech.edu/records/20237> (besucht am 26. 06. 2024).
- [40] Jeremy B. Wilmer u. a. „Human face recognition ability is specific and highly heritable“. en. In: *Proc. Natl. Acad. Sci. U.S.A.* 107.11 (März 2010), S. 5238–5241. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.0913053107](https://doi.org/10.1073/pnas.0913053107). URL: <https://pnas.org/doi/full/10.1073/pnas.0913053107> (besucht am 10. 01. 2024).

Eidesstattliche Erklärung

Hiermit versichere ich – Maria Lietsch – an Eides statt, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Sämtliche Stellen der Arbeit, die im Wortlaut oder dem Sinn nach Publikationen oder Vorträgen anderer Autoren entnommen sind, habe ich als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt oder anderweitig veröffentlicht.

Mittweida, 30. August 2024

Ort, Datum


Maria Lietsch, B.Sc.