St. Pölten University of Applied Sciences

/medien & digitale technologien

/fh/// st. pölten

Analyse und Erstellung von Bewegtbildfälschungen, die auf Deep Learning Algorithmen basieren

Untersuchung der Glaubwürdigkeit von Deep Fakes

Diplomarbeit

Ausgeführt zum Zweck der Erlangung des akademischen Grades
Dipl.-Ing. für technisch-wissenschaftliche Berufe

am Masterstudiengang Digitale Medientechnologien an der Fachhochschule St. Pölten, **Masterklasse Post Produktion**

von:

Tobias Sautner, BSc

DM171568

FH-Prof. Dipl.-Ing. (FH) Mario Zeller
FH-Prof. Dipl.-Ing. (FH) Matthias Husinsky

Wien, 27.08.2020

Ehrenwörtliche Erklärung

Ich versichere, dass

- ich diese Arbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.

- ich dieses Thema bisher weder im Inland noch im Ausland einem Begutachter/einer Begutachterin zur Beurteilung oder in irgendeiner Form als Prüfungsarbeit vorgelegt habe.

Diese Arbeit stimmt mit der vom Begutachter bzw. der Begutachterin beurteilten Arbeit überein.

Wien, 27.08.2020 .....

Ort, Datum

Unterschrift

Kurzfassung

Diese Diplomarbeit beschäftigt sich mit der Analyse und Erstellung von Bewegtbildfälschungen, welche mit Hilfe von Technologien erzeugt worden sind, die auf Deep Learning Algorithmen basieren. Diese Fälschungen werden meist Deep Fakes genannt.

Dazu erfolgt zuerst eine Betrachtung der notwendigen Fachbegriffe der Thematik. Der Themenkomplex der Visual Effects wird analysiert und mit einem Überblick der allgemeinen Funktionsweise von künstlicher Intelligenz in Verbindung gebracht. Dabei wird auch ein Einblick geboten, welche anderen Einsatzmöglichkeiten es für Werkzeuge im Bewegtbildbereich gibt, die auf künstliche Intelligenz zurückgreifen.

Der Hauptteil dieser Diplomarbeit beschäftigt sich mit Deep Fakes. Hier wird sowohl das enorme Gefahrenpotential dieser Technologie erläutert, die komplizierte rechtliche Situation thematisiert, sowie der theoretische Erstellungsprozess im Allgemeinen erläutert. An Hand der Software *DeepFaceLab* wird in Folge dessen der Deep Fake Prozess Schritt für Schritt praktisch erläutert. Konkret werden drei Deep Fake Beispiele, unter Verwendung von Videos einer Stockplattform, erzeugt. Die Erstellung erfolgt auf klassischer Konsumenten-Hardware, auf Cloud-Computing wird dabei in keinster Weise zurückgegriffen.

Diese erstellten Bewegtbildfälschungen werden genutzt, um eine Personenbefragung durchzuführen. Dabei werden die drei Deep Fake Beispiele, zusammen mit sieben unverfälschten Videos, gezeigt. Die Testpersonen geben dann pro Videobeispiel an ob es sich, ihrer Meinung nach, um eine Bewegtbildfälschung handelt. Das Ergebnis dieser Befragung zeigt jedoch, dass die Probanden und Probandinnen mit einer leichten Mehrheit die Deep Fakes nicht als solche erkennen, sondern als unverfälschte Videos deklarieren. Dadurch wird aufgezeigt, dass Einzelpersonen, unter Verwendung von gewöhnlichen Hardwarekomponenten, Bewegtbildfälschungen erzeugen können, die von keiner signifikanten Mehrheit als Bewegtbildfälschungen erkannt werden können. Die Verwendung von Technologien, die auf künstlicher Intelligenz basieren, macht dies möglich.

Abstract

This diploma thesis deals with the analysis and creation of video fakes, which have been generated by technologies based on deep learning algorithms. These fakes are usually called deep fakes.

For this examination, the necessary technical terms within the field are discussed first. The complex of visual effects is analysed and linked to an overview of the general functioning of artificial intelligence. This also provides an insight into what other possible applications there are for tools in the motion picture field that make use of artificial intelligence.

The main part of this diploma thesis deals with deep fakes. Here, the enormous potential danger of this technology is explained, the complicated legal situation is addressed, and the theoretical creation process in general is explained. Using the software *DeepFaceLab*, the general process of creating deep fakes is explained step by step. Three Deep Fake examples will be created, using videos from a stock footage platform. The creation is carried out on classic consumer hardware, with no use of cloud computing.

These three Deep Fakes are used to carry out a people survey, where the fake videos are shown in combination with seven unaltered videos. The test subjects then indicate for each video example whether the video is fake or not. The result of this survey shows that the test persons, with a slight majority, do not recognize the deep fakes as such and declare them as unaltered videos. This shows that individuals, using ordinary hardware components, can create fake videos that cannot be recognized as fakes by any significant majority. The use of technologies, based on artificial intelligence, makes this possible.

Inhaltsverzeichnis

1 Einleitung

In dieser Diplomarbeit soll untersucht werden, wie schwer es für Menschen ist, Bewegtbildfälschungen zu erkennen, die mit Hilfe von Programmen, die auf künstlicher Intelligenz basieren, erstellt worden sind. Um sich diesem Thema anzunähern, wird zuerst die Historie von Bewegtbildfälschungen erläutert und berühmte Beispiele präsentiert. Danach wird der Themenschwerpunkt der künstlichen Intelligenz aufgegriffen und Überbegriffe wie Deep Fakes und neuronale Netzwerke definiert. Weiters werden die potenziellen Gefahren, die diese Technologien, bezogen auf Bewegtbild bergen, erläutert. Die konkrete Vorgangsweise, die zur Erstellung dieser Bewegtbildfälschungen notwendig ist, wird detailliert aufgeschlüsselt. Weiters wird in der Arbeit erläutert, ob und wie gut man künstliche Intelligenz verwenden kann, um Bewegtbildfälschungen zu entlarven. Im Zuge der Arbeit werden dementsprechende Fälschungen erstellt. Diese Fälschungen werden, gemischt mit unverfälschtem Bewegtbildmaterial, einer Gruppe von Probanden und Probandinnen gezeigt. Diese müssen pro Videobeispiel angeben, ob das gezeigte Video eine Bewegtbildfälschung ist oder nicht. Als Quelle für die Videobeispiele werden Videoclips von Stockplattformen verwendet, um zu garantieren, dass die Probanden und Probandinnen die gezeigten Personen nicht kennen.

Eine der Hypothesen, die in dieser Diplomarbeit angenommen wird ist, dass Bewegtbildfälschungen die mit der Hilfe von künstlicher Intelligenz erzeugt worden sind, von keiner signifikanten Mehrheit der Probanden und Probandinnen als Fälschung erkannt werden können. Dies birgt aktuell eine große Gefahr für die Gesellschaft, da Bewegtbild aktuell noch ein viel größeres Vertrauen an Authentizität erhält im Vergleich zu Einzelbilder. Bewegtbildfälschungen, die reale Personen zeigen, zum Beispiel Politiker und Politikerinnen, besitzen somit eine weitreichende und gefährliche Tatkraft. Da sich solche Fälschungen immer mehr verbreiten werden, ist es notwendig auch Gegenmaßnahmen oder automatisierte Kontrollmechanismen zu entwickeln. Um die Hypothese zu überprüfen, werden die innerhalb dieser Diplomarbeit erstellten Bewegtbildfälschungen verwendet, um zu eruieren, ob tatsächlich eine signifikante Mehrheit von Probanden und Probandinnen diese nicht erkennen können.

1.1 Thematik und Problemstellung

Diese Diplomarbeit befasst sich mit der Thematik Bewegtbildfälschungen, welche mit Hilfe von Software erstellt wurden, die auf künstlicher Intelligenz basieren. Hauptsächlich werden diese Fälschungen als Deep Fakes bezeichnet. Bei typischen Deep Fakes werden Gesichter zweier Personen getauscht. Beispiele dieser Technik fanden zum ersten Mal im Jahr 2017 große mediale Aufmerksamkeit, als vermehrt pornografische Videos im Internet veröffentlicht wurden, bei denen die Gesichter der Pornodarstellerinnen durch Gesichter von bekannten weiblichen Schauspielerinnen getauscht wurden. Dabei sorgte vor allem die hohe Qualität der Fälschungen für große Aufmerksamkeit und erzeugte einen besorgten Aufschrei.

Andere berühmte Deep Fakes zeigen unter anderem Barack Obama oder Vladimir Putin in Szenen, die so nie passiert waren. Weiters wurden auch viele Beispiele publiziert, bei denen Darsteller oder Darstellerinnen berühmter Filmszenen getauscht worden sind. Die meisten dieser bekannten Deep Fake Beispiele hatten nie die Absicht, die Betrachter und Betrachterinnen in die Irre zu führen. Meistens ging es entweder um eine humorvolle Darstellung, das Erreichen großer Reichweite durch Verwendung berühmter Gesichter oder die Erstellung von fragwürdigen pornografischen Inhalten.

Dabei ist die Technik hinter Deep Fakes revolutionär. Plötzlich können Personen, die kaum Erfahrung und Wissen in der Thematik der Bewegtbildmanipulation haben, Bewegtbildfälschungen auf einem hohen Niveau erstellen. Der immense Fortschritt im Bereich der künstlichen Intelligenz macht es möglich. Gleichzeitig erzeugt dies jedoch auch die Gefahr, dass die Anzahl von solchen hochwertigen Bewegtbildfälschungen, rasant zunimmt. Diese Fälschungen können für verschiedene gefährliche Einsatzzwecke missbraucht werden, und in Folge dessen das Vertrauen in das Medium Bewegtbild massiv erschüttern.

Um einschätzen zu können, wie weit die Technik hinter Deep Fakes schon fortgeschritten ist, gilt es zu überprüfen wie gut Menschen diese noch von originalen Videoaufnahmen unterscheiden können. Im Zentrum dieser Diplomarbeit steht deswegen die Untersuchung der Glaubhaftigkeit solcher Deep Fakes. Deswegen sollen im Zuge dieser Arbeit mehrere Deep Fake Beispiele kreiert werden, um diese einer Probandengruppe vorzuführen. Dabei wird der Erstellungsprozess ausführlich dokumentiert, um einen exakten Überblick über die Vorzüge als auch die Limitierungen von Deep Fakes zu geben.

1.2 Motivation

Die persönliche Faszination für die Themen künstliche Intelligenz und Bewegtbild ist der Ausgangspunkt für diese Diplomarbeit. Die Betrachtung der ersten hochwertigen Deep Fakes hinterließ einen faszinierenden und bleibenden Eindruck. Würde man solch eine hoch qualitative Fälschung mit klassischen Mitteln aus dem Visual Effects Bereich kreieren wollen, wäre ein hoher manueller Arbeitsaufwand notwendig. Deswegen wurde ein großes persönliches Interesse geweckt diese Technologie zu erlernen beziehungsweise die Grenzen dieser auszutesten.

Generell gilt zu erwähnen, dass die Thematik der künstlichen Intelligenz immer tiefer in unser persönliches Lebensumfeld eindringt, ohne dass dies von vielen Personen überhaupt wahrgenommen wird. Millionen von Menschen verwenden tagtäglich persönliche Digitalassistenten wie *Alexa* von Amazon oder *Siri* von Apple, Streamingplattformen bieten detaillierte Filmempfehlungen oder Autos können schon rudimentäre Fahrmanöver autonom ausführen. So beeindruckend diese Beispiele auch sind, sind diese nur ein kleiner Vorgeschmack. Wir befinden uns in diesem Bereich immer noch in der Anfangsphase und zukünftigen Entwicklungen können momentan schwer abgeschätzt werden.

Technologien und Werkzeuge, die auf künstlicher Intelligenz basieren, spielen auch in der Arbeitswelt eine immer größere Rolle. Dies gilt auch für die Bewegtbild-Branche. Durch die persönliche Einschätzung, welche gravierende Auswirkung diese Technik auf die Arbeitslandschaft prognostiziert, ist die Motivation groß Techniken wie diese auch zu verstehen. Die persönliche Einschätzung wird dadurch unterstrichen, dass immer mehr Softwarefirmen innerhalb der Bewegtbild-Branche Funktionen in ihre Softwareprodukte integrieren, die auf künstliche Intelligenz zurückgreifen.

Ein weiterer Aspekt der persönlichen Motivation für diese Diplomarbeit ist, dass die Thematik *Fake News* immer mehr in den öffentlichen Diskurs rutscht. Dabei wird vor allem die potenzielle Gefahr dieser Falschmeldungen diskutiert, sowie die mögliche Beeinflussung von Personengruppen behandelt. Dieser Diskurs könnte, durch eine große Verbreitung von hochwertigen Deep Fakes weit komplexer werden. Das Vertrauensverhältnis gegenüber dem Medium Bewegtbild könnte sich massiv verändern. Aktuelle Veröffentlichungen von Büchern wie *The Reality Game: How the Next Wave of Technology Will Break the Truth* vom Autor Samuell Wolley sind eindeutige Indikatoren für die starke Schlagkraft dieser Thematik.

1.3 Ziele, Forschungsfragen und Hypothesen

Das Ziel dieser Diplomarbeit ist es, die Herangehensweise der Erstellung von Deep Fakes im Detail zu erörtern und dabei eigene entsprechende Videobeispiele zu erstellen, die eine hohe subjektive Glaubwürdig vorweisen. Dafür wird die notwendige Theorie zum Thema künstlicher Intelligenz erläutert und der theoretische Aspekt hinter dem Deep Fake Erstellungsprozessen offen legt. Die kreierten Videobeispiele werden danach an Probanden und Probandinnen auf ihre Glaubwürdigkeit getestet.

Die zentralen Forschungsfragen, die sich aus der oben beschriebenen Problematik ergeben, lauten wie folgt:

1. Ist es möglich, als Einzelperson ohne spezielle Profihardware Deep Fakes zu erzeugen, welche eine hohe subjektive Glaubwürdigkeit besitzen?
2. Ist es möglich, diese hochwertigen Deep Fakes nur unter Verwendung von Videomaterial von Stockplattformen zu erstellen?
3. Kann eine signifikante Mehrheit einer Menschengruppe moderne Bewegtbildfälschungen von unveränderten Videos unterscheiden?

Vom bisher vorhandenen Vorwissen lassen sich folgende Hypothesen definieren:

- Hypothese 1: Es ist möglich mit aktuellen Computersystemen aus dem Konsumentenbereich, unter der Verwendung von Werkzeugen, die auf künstlicher Intelligenz passieren, sehr hochwertige Bildfälschungen zu erstellen. Hierbei gilt jedoch die Bedingung, dass als Trainingsdaten nur Videos von Stockplattformen verwendet werden können.
- Hypothese 2: Eine signifikante Mehrheit einer Menschengruppe wird es nicht gelingen, hochwertige Deep Fakes als Bewegtbildfälschungen zu erkennen, selbst wenn die Testpersonen darauf hingewiesen werden, explizit darauf zu achten.

Die erste Hypothese gilt es zu untersuchen, da viele der sehr hochwertigen und bekannten Deep Fake Beispiele von mehrköpfigen Teams kreiert wurden, bei denen der Arbeitsaufwand einen klassischen Deep Fake Prozess weit überstieg. Dabei wurden zum Teil ähnlich aussehende Personen gecastet, um diese für das Zielmaterial zu filmen. Es ist anzunehmen, dass diese Teams über Hardwareressourcen verfügten, die für eine Einzelperson nicht erreichbar sind. Somit gilt es zu untersuchen ob es auch möglich ist, unter Einsatz von weniger

Ressourcen, Bewegtbildfälschungen zu erstellen, die eine hohe Glaubwürdigkeit vorweisen.

Bei der zweiten Hypothese wird angenommen, dass die Technologie hinter Deep Fakes schon solch ein hohes Niveau erreicht hat, das von einer signifikanten Mehrheit von Menschen die Fälschung nicht mehr erkannt werden kann. Jedoch basieren die meisten hochwertigen Deep Fakes auf Schauspieler und Schauspielerinnen oder Politiker und Politikerinnen, und werden somit direkt als Fälschung erkannt, da die betrachtende Person weiß, dass die abgebildete Person sich nicht in diesem Szenario befindet. Somit gilt es, für die Überprüfung der zweiten Hypothese, nur Beispiele zu verwenden die keinerlei Personen des öffentlichen Lebens darstellen.

1.4 Methodik

Um die definierten Hypothesen überprüfen zu können werden im Zuge dieser Diplomarbeit drei Deep Fake Beispiele kreiert. Diese werden, gemeinsam mit sieben unveränderten Videoclips, mindestens 100 Probanden und Probandinnen gezeigt, die dabei definieren sollen, welche der Videoclips die Bewegtbildfälschungen sind. Um hochwertige Deep Fake Beispiele erzeugen zu können, ist es zuvor notwendig, durch die Analyse einschlägiger Literatur, den Erstellungsprozess genau zu erlernen. Dafür ist es auch unabdingbar Fachbegriffe aus dem Bereich der künstlichen Intelligenz zu definieren, da die konkrete Anwendung dieser Begriffe im Erstellungsprozess eine tragende Rolle spielen. Weiters ist es notwendig die spezifischen Hardwarevoraussetzungen für den Deep Fakes zu erläutern.

Für die anschließende Befragung der Testpersonen wird eine Online-Umfrage erstellt. Dort müssen die Probanden und Probandinnen pro gezeigtes Video entscheiden, ob es sich ihrer Einschätzung nach, um eine Bewegtbildfälschung oder einen unveränderten Videoclip handelt. Nach der Teilnahme von mindestens 100 verschiedenen Probanden und Probandinnen erfolgt eine detaillierte Auswertung der Ergebnisse. Dabei wird überprüft, ob die Bewegtbildfälschungen signifikant öfter als Fälschung deklariert worden sind als die unverfälschten Videos. Durch den Einsatz von Diagrammen werden die demografischen Daten der Testpersonen sowie die Ergebnisse ausführlich präsentiert. Eine detaillierte Beschreibung der Probandenbefragung finden Sie im Kapitel *6.1 Genauer Ablauf der Untersuchung.*

1.5 Gliederung der Arbeit

Die Einleitung in die Thematik, die Problemstellung, die Forschungsfragen sowie die daraus resultierenden Hypothesen dieser Diplomarbeit werden im Kapitel 1 deklariert. Die Motivation eine Diplomarbeit über Bewegtbildfälschungen, welche unter Verwendung von *Deep Learning* Technologien erzeugt worden sind zu verfassen, wird ebenfalls im Kapitel 1 erklärt. Weiters wird auch die Methodik, welche zur Überprüfung der Hypothesen dienen soll, kurz erläutert.

Das Kapitel 2 beinhaltet einen kurzen Exkurs in die allgemeine Thematik der Visual Effects. Dies ermöglicht es Deep Fakes besser innerhalb dieses Gebiets einzuordnen zu können. Dabei werden analoge und digitale Visual Effects Techniken separat betrachtet. Die persönliche Einschätzung ist, dass der Durchbruch von Technologien basierend auf künstlicher Intelligenz, eine ähnlich starke Veränderung erzeugen konnte, wie der Umstieg von Analogtechnik auf Digitaltechnik. Dementsprechend ist es interessant diese damaligen Veränderungen zu erläutern. Das *Uncanny Valley* Phänomen wird ebenfalls erläutert, da dies auf die Glaubhaftigkeit von Deep Fakes Auswirkungen haben könnte.

Das Kapitel 3 fokussiert sich auf den Themenkomplex der künstlichen Intelligenz. Hierbei werden zunächst notwendige Fachbegriffe erläutert, die später im Deep Fake Erstellungsprozess relevant sind. Der Grund für die massive Weiterentwicklung dieser Technologie wird ebenso erklärt wie Anwendungsbeispiele, bei denen solche Werkzeuge schon in der Bewegtbild-Branche verwendet werden.

Der Fokus von Kapitel 4 sind Deep Fakes. Hier wird die Entstehungsgeschichte dieses Phänomens erläutert, sowie ein berühmtes Deep Fake Beispiel im Detail behandelt. Weiters wird thematisiert, welche Personengruppen besonders durch Deep Fakes gefährdet sind und wie die rechtliche Situation aussieht, wenn man Opfer einer missbräuchlichen Verwendung wird. Mögliche gefährliche Einsatzzwecke werden in diesem Kapitel erläutert. Als Gegendarstellung wird jedoch auch betrachtet, welche positiven Einsatzzwecke diese Technologie mit sich bringen kann. Die allgemeine theoretische Vorgangsweise zur Erstellung eines Deep Fakes wird erläutert sowie eine Auflistung momentan verfügbarer Softwareprodukte zusammengefasst.

Im Kapitel 5 wird der Erstellungsprozess von Deep Fakes an Hand des Softwareprodukts *DeepFaceLab* detailliert analysiert und die Einzelschritte chronologisch abgearbeitet.

Die Einleitung in den Praxisteil dieser Diplomarbeit erfolgt im Kapitel 6, wobei hier die geplante Probandenbefragung detailliert erläutert wird. Darüber hinaus wird auch dargelegt, welche Überlegungen für die Auswahl der Deep Fake Beispiele getroffen wurden, und wie diese Entscheidungen die Befragung beeinflussen.

Der Erstellungsprozess der drei Deep Fakes Beispiele wird im Kapitel 7 beschrieben. Pro Videobeispiel werden die gewählten Trainingsdaten erläutert, die verschiedenen Konfigurationen dokumentiert sowieso auf die manuellen Bearbeitungskorrekturen eingegangen. In diesem Kapitel werden auch Komplikationen erwähnt, die während des Erstellungsprozesses auftraten.

Die Auswertung der Probandenbefragung wird im Kapitel 8 dokumentiert. Neben der Darstellung durch Diagramme der eingegangenen Antworten erfolgt auch eine analytische Interpretation der Ergebnisse. Im Unterkapitel 8.1 wird auf die Erkennung von Deep Fakes mit künstlicher Intelligenz eingegangen und in Relation zu den Ergebnissen der Probandenbefragung betrachtet.

Das abschließende Kapitel 9 beinhaltet eine Zusammenfassung, sowie eine Evaluation der erlernten Ergebnisse, bezogen auf die Forschungsfragen und Hypothesen. Darüber hinaus wird dargelegt, welche zukünftige Forschungsprojekte in dieser Thematik sinnvoll wären.

2 Anwendung von Visual Effects Techniken

Die Anwendung von Visual Effects Techniken (Abgekürzt VFX) hat eine ähnlich lange Geschichte wie das Medium Bewegtbild selbst. Vor allem die Filmindustrie ist für die Verbreitung und Verbesserung der Möglichkeiten dieser Effekte verantwortlich. In den letzten 100 Jahren hat sich die Definition von Visual Effects immer wieder fundamental geändert und erweitert. Jedoch verbindet man den Begriff weiterhin primär mit der Filmindustrie.

2.1 Begriffsdefinition – Visual Effects

Unter dem Begriff Visual Effects, beziehungsweise der oft verwendeten Abkürzung VFX versteht man jegliche Art von Bewegtbildmaterial, welches nach dem abgeschlossenen Aufnahmeprozess, durch die Anwendung von Bearbeitungstechniken manipuliert oder angepasst wurde. Dies bedeutet, dass der Großteil der Visual Effects Arbeit im Postproduktionsprozess stattfindet, jedoch nicht ausschließlich dort vorzufinden ist. So können durch Techniken wie beispielsweise Frontal- und Rückprojektionen sehr wohl auch Visual Effects direkt beim Aufnahmeprozess verwendet werden. Vor allem im digitalen Zeitalter hat diese Form von Visual Effects Arbeit zugenommen. (Okun & Zwerman, 2010)

2.1.1 Special Effects

Diese Art von Effekten werden auch oft Practical Effects genannt. Hierbei handelt es sich um einen Effekt, der durch ein physikalisches Objekt erzeugt wird, und nicht auf Inhalte zurückgreift, die von einem Computer erzeugt wurden. Typische Beispiele wären Effekte wie Explosionen, Schusswunden, Regen oder auch Seilsysteme, um Schauspieler und Schauspielerinnen durch Szenen fliegen zu lassen. Special Effects und Visual Effects schließen sich keinesfalls aus, sondern können auch gemeinsam Verwendung finden. So kann ein Seilsystem einen Schauspieler oder Schauspielerin in der Höhe schweben lassen und danach via Visual Effects die Seile entfernt werden.

2.1.2 Computer Generated Imagery – CGI

Unter diesem Begriff versteht man Bildmaterial, welches komplett durch einen Computer erzeugt worden ist. Typische Beispiele wären 3D-Animationsfilme. CGI ist per se keine Unterform von Visual Effects, da es tendenziell Bewegtbild nicht manipuliert, sondern komplett aus dem Nichts erzeugt. Jedoch greift man bei der Anwendung von Visual Effects oft auf CGI-Material zurück. Möchte man beispielsweise einen 3D-animierten Dinosaurier in Bewegtbildmaterial einfügen, fällt dieser Vorgang in die Kategorie der Visual Effects, wobei die Erzeugung des Dinosauriers zuvor in die Kategorie der CGI Anwendung fällt. Heutzutage sind Produktionen, mit einem großen Visual Effects Anteil immens auf diese Vorgangsweise angewiesen. (Okun & Zwerman, 2010)

2.2 Visual Effects vor dem digitalen Zeitalter in der Filmindustrie

Ende des 20. Jahrhundert fanden digitale Effekte in großen Filmproduktionen mehr und mehr Verwendung. Auf Visual Effects fokussierte Unternehmen wie das von George Lucas gegründete *Industrial Light & Magic* oder Filme wie *Terminator 2: Judgment Day* (1991) und *Jurassic Park* (1993) läuteten knapp vor der Jahrtausendwende ein neues Zeitalter in der Filmindustrie ein. Zwar würde es noch mehrere Jahre dauern bis die Mehrheit der Filmproduktionen einen rein digitalen Arbeitsablauf verwendet, jedoch zeigten diese Filme auf in welche Richtung sich die Zukunft der Filmwelt bewegen wird. (Cook, 1996, S. 955 - 956)

In der analogen Welt wurden Visual Effects noch anders realisiert. Damals veränderte man Bewegtbildmaterial mit Hilfe von fotografischen Effekten direkt auf dem analogen Film. So wurden etwa durch Verwendung von einem optischen Printer weitere Elemente auf einen schon belichteten Film projiziert. Bei einem optischen Printer ist ein Projektor frontal auf eine Kamera gerichtet, so dass ein projiziertes Bild des Projektors in der Kamera aufgenommen werden kann. Beide Geräte befinden sich auf einem Schienensystem, um die Entfernung für entsprechende Verkleinerungen beziehungsweise Vergrößerungen anzupassen. (Netzley, 2000, S.165)

Der Optische Printer wurde im Jahr 1931 erfunden, zuvor wurden Effekte direkt in der Kamera oder beim Filmprozess selbst realisiert, dank dem optischen Printer war eine konkrete Postproduktion möglich. Bevor diese technische Entwicklung

verfügbar war, musste man Techniken wie Stop Motion[1] verwenden, um Geschichten verwirklichen zu können, die nicht mit einer Kamera aufnehmbar waren. Komplexe Figuren wurden als sogenannte Animatronics umgesetzt. Diese waren mechanisch oder elektronisch steuerbare Figuren und konnten somit auch mit den Schauspielern und Schauspielerinnen direkt interagieren. (Lobban, 2000, S. 573)

2.2.1 Berühmte Beispiele

Diese beiden Beispiele stehen stellvertretend für andere wichtige Werke aus der Epoche der analogen Visual Effects, und sollen einen Eindruck der damaligen Vorgehensweise vermitteln.

2.2.1.1 Stopptrick – Erste Verwendung von Visual Effects

Der Kurzfilm *The Execution of Mary Stuart* (1895) welcher von Thomas Edision produziert worden ist, gilt als erstes Werk, das Visual Effects verwendet hat. Der Film handelt von einer Exekution. Eine Frau liegt vor einem Henker. Als dieser seine Axt hebt wird die Kameraaufnahme kurz beendet, die Schauspielerin durch eine Puppe ausgetauscht und der Aufnahmeprozess erneut gestartet. Der Henker köpft die Puppe. Durch den gezielten schnellen Schnitt, bleibt der Austausch vor dem Zuschauer und der Zuschauerin verborgen. (Finance & Zwerman, 2015, S. 3)

2.2.1.2 Animatronics

Der Film *Jurassic Park*, welcher im Jahr 1993 erschien, wurde unter anderem wegen seinen hochaufwendigen Visual Effects ein Kultfilm. Für den Film wurde ein Dinosaurier Animatronic gebaut. Dieser hatte ein Gewicht von 7900kg und war über 12m lang. (Shay & Duncan, 1993)

Stan Winston, der Visual Effects Experte der Filmproduktion, erhielt für dieses Werk einen Oscar.

[1] Bei dieser Filmtechnik werden mehrere einzelne Aufnahmen unbewegter Motive so aneinandergereiht, dass die Illusion von Bewegung entsteht

*Abbildung 1 – T-Rex Animatronic im Warner Bros. Studio (*Jurassic Park T-Rex Robot - As Dangerous as a Real Dinosaur*, o. J.)*

2.3 Visual Effects im digitalen Zeitalter der Filmindustrie – ohne künstliche Intelligenz

Der Sprung in das digitale Zeitalter war für die Visual Effects Industrie eine Revolution. Animatronics wurden durch 3D-Modelle abgelöst und optische Printer wurden durch digitale Compositingsoftware[2] ersetzt. Die Geschwindigkeit mit der sich Visual Effects weiterentwickelten erhöhte sich auf einen Schlag drastisch.

> „The years since 1993, it can be argued, included as much innovation as the previous 100 years of visual effects. Everything was open, and a legion of incredibly clever visual effects artists, scienctists, and engineers redrew the landscape such that no effect was beyond our reach. We saw the world of optical printing fade from common use faster

[2] Softwareprodukte, die verwendet werden, um mehrere Bildsequenzen und Effekte durch Überlagerung zu einer Bildsequenz zusammenzufügen

> than any of us would have believed possible" (Okun & Zwerman, 2010, S. 13)

Die Leistungsfähigkeit der damaligen Computerhardware stieg nach dem *Mooresches Gesetz* alle 18 Monate um das doppelte an. (Chau et al., 2003, S. 123)

Dementsprechend wurden die CGI-Elemente komplexer, die virtuellen Welten größer und die Kompositionseffekte besser. Erste rein 3D animierte Spielfilme wie *Toy Story* (1995) und *A Bug's Life* (1998) kamen in die Kinos und wurden große Erfolge. Zwar wuchsen die Rechenleistung und somit die Möglichkeiten der Branche schnell, jedoch wurde bis vor kurzem noch keinerlei künstliche Intelligenz verwendet, wie später im Kapitel Deep Fakes beschrieben. Dutzende Programmierer und Mathematiker entwickelten digitale Algorithmen und Werkzeuge, um sie dann den Filmstudios beziehungsweise den Visual Effects Experten zur Verfügung zu stellen.

2.3.1 Berühmte Beispiele

Diese folgenden Beispiele stehen stellvertretend für viele andere Werke aus der Anfangszeit der digitalen Visual Effects und dienen als beispielhafte Erklärung.

2.3.1.1 Toy Story

Toy Story war der erste Spielfilm, der nur durch die Verwendung von Computeranimation erzeugt wurde und stellt dadurch einen Meilenstein in der Visual Effects und Filmbranche dar. Der Film handelt davon, dass die verschiedenen Spielzeuge eines jungen Kindes in seiner Abwesenheit zum Leben erwachen. Vor allem die technische Herangehensweise der Produktion war für die damalige Zeit bahnbrechend und unterschied sich fundamental von klassischen handgezeichneten Animationen. Bei diesen Zeichnungen war der Bewegungsprozess von Figuren aufwendiger. Wollte man die Bewegung einer Figur animieren, wurden sowohl die Anfangs-, und die Endposition als auch markante Positionen dazwischen skizzenhaft gezeichnet. Darauffolgend wurde von einem mehrköpfigen Team jeder Frame, also auch die zwischen den zuvor skizzierten Positionen, gezeichnet. So mussten für eine Sekunde Film 24 Einzelbilder gezeichnet werden. An einer Einstellung arbeiteten so teils 30 Personen. Hier konnte *Toy Story* durch die Verwendung rein digitaler Modelle viel Zeit gewinnen und ebenso den Personalaufwand stark verringern. Ein Computer berechnet die einzelnen Zwischenpositionen von selbst, welche zwischen einer gewählten Anfangs- und Endposition liegen. Der Animator muss diese nur minimal für eine natürliche Bewegung anpassen. An der Produktion des Filmes arbeiteten

insgesamt nur 30 Animations-Spezialisten und Spezialistinnen mit. Die große Begeisterung und Akzeptanz der Zuschauer und Zuschauerinnen bezüglich dieses neuen visuellen Stiles liegt vermutlich auch darin begründet, dass die gezeigten Figuren keine realen Referenzen haben. (Henne et al., 1996, S. 465)

2.3.1.2 Jurrasic Park

Jurassic Park hat bei seiner Veröffentlichung im Jahr 1993 nicht nur im Animatronicsbereich neue Maßstäbe gesetzt, sondern auch im Bereich der digitalen Visual Effects bahnbrechende Ergebnisse erzielt. Der Abenteuerfilm handelt von durch Genmanipulation erzeugten Dinosauriern, die auf einer privaten Insel in einem Zoo leben. Eigentlich wollte der Regisseur Steven Spielberg den Film rein mit Animatronics und Stop Motion bewerkstelligen, dies erwies sich jedoch als unmöglich. Alles was nicht am Set mit den Animatronics möglich war, wurde später in der Postproduktion digital gelöst. Im Film wurde der Wechsel zwischen, dem mechanischem Model und dem digitalen Model mit geschickten Schnitten versteckt.

> „Anything that could not be done live on the set using the animatronics would be done digitally. In practise, this often meant that elaborate, complicated, or fast full-body dinosaur movements would be digital, while the models would be used for partial views of a creature, as when the T-Rex head comes into frame or the raptor feet come down in close-up on the park's kitchen floor" (Prince, 2011, S. 30)

Dieses Beispiel zeigt auch einen der großen Vorteile von digitalen Visual Effects auf: Die hohen Kosten und die hohe Komplexität solch eines Animatronicmodels zwingen den Regisseur es mit Bedacht einzusetzen. Jeder größere Umbau oder gar Beschädigung des Models kostet die Produktion enorme Ressourcen. Durch digitale Visual Effects war es dem Regisseur jedoch möglich, den Film seinen Vorstellungen entsprechend umzusetzen. Jedoch herrschte bei den Dreharbeiten auch Skepsis, ob das Endresultat wirklich ein homogenes Ergebnis zwischen den mechanischen und digitalen Dinosauriern liefern würde. So wurden Kameraeinstellungen und Kameraschwenks konservativer und statischer gewählt, um den digitalen Prozess zu vereinfachen. Für die nachfolgenden Teile der *Jurassic Park* Filmtrilogie war dies später nicht mehr notwendig, da man der digitalen Technik schon mehr vertraute. (Prince, 2011, S. 31)

2.4 Digitalisierung von analogen VFX Techniken, inklusive Nachteile

Es sollte erwähnt werden, dass die Digitalisierung der Visual Effects Welt keine reine Neuentwicklung war, sondern großteils bekannte Techniken und Vorgänge in die digitale Welt transferiert hat. Der Gedanke dahinter war, dieselben Ergebnisse zu erzielen, jedoch die davor notwendigen mechanische Effekte zu reduzieren. So konnten Doppelbelichtungen, die zuvor mit optischen Printern erzeugt worden waren, im digitalen Kompositingprogramm absolviert werden.

> „It should be noted that digital VFX are often no more than digitized versions of analog VFX for which software programs have been developed so that they do not need to be produced mechanically or physically in real space but in cyberspace via a computer interface" (Venkatasawmy, 2013, S. 70)

Notwendige Alphamasken[3], bei welchen früher Teile eines Filmnegatives entwickelt worden sind, wurden einfach digitalisiert. Auch Matte Paintings, bei welcher in analoger Form Teile der Kulisse bemalt wurden, wurden digital erstellt. Filmsets konnten somit simpler und kleiner gestaltet werden. Einzelne Schauspieler und Schauspielerinnen konnten in leeren Greenscreens beziehungsweise Bluescreens performen und später via Computer in eine virtuelle Welt eingefügt werden.

Diese geänderten Vorgangsweisen blieben jedoch nicht ohne Nachteile. So ist es für den Schauspieler beziehungsweise Schauspielerin schwieriger authentisch zu spielen, wenn er kein reales Objekt als Referenz vor sich hat. Der Schauspieler oder die Schauspielerin spielt ins Leere, und muss sich im Kopf vorstellen können, wie das Endergebnis werden soll. Dies kann dazu führen, dass zum Beispiel Augen nicht exakt dort hinsehen wo man es als Zuschauer beziehungsweise Zuschauerin vermutet, oder Schauspieler und Schauspielerinnen nicht optimal authentisch auf ihre Umwelt reagieren. Jedoch mussten sich Schauspieler und Schauspielerinnen bereits an viele technischen Weiterentwicklungen in der Filmbranche anpassen, also ist dieses Problem somit kein Alleinstellungsmerkmal in der Digitalisierung. Solch eine notwendige Anpassung kann durch die Integration von Systemen, die auf künstliche Intelligenz basieren, erneut bevorstehen.

[3] Werden verwendet, um mehrere Bilder miteinander zu kombinieren. Definieren welche Teile des Bildes sichtbar sind.

2.5 Uncanny Valley Effekt

Sowohl im Filmbereich als auch bei den Deep Fakes gilt es als Ziel, eine Illusion zu kreieren, die für den Betrachter und Betrachterin so echt wie möglich erscheint. Dies war vor allem zu Beginn des digitalen Zeitalters nicht so einfach. Die Werkzeuge waren noch nicht so weit entwickelt und exakt wie heute. Bezüglich der Akzeptanz von unechten Ergebnissen, gibt es einen paradox-erscheinenden Effekt namens *Uncanny Valley*. Dieser ist nicht exklusiv für Bewegtbildfälschungen, da er auch Puppen, Roboter und ähnliches inkludiert, spielt jedoch auch für einen hochwertigen Bildmanipulationsprozess eine große Rolle. Dieser Effekt besagt, dass die Erhöhung des Detailgrads von künstlichen Figuren nicht eine automatisch größere Akzeptanz bei den Zuschauern und Zuschauerinnen erzeugt, sondern schlagartig eine Ablehnung dieser künstlichen Figuren kreiert.

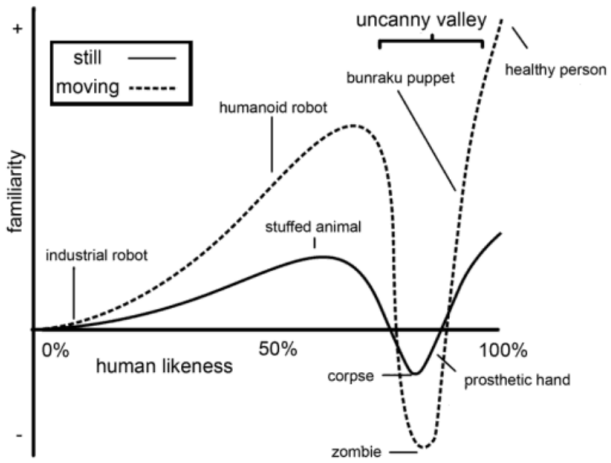

Abbildung 2 – Diagramm zur Beschreibung des Uncanny Valley Effekts (de Borst & de Gelder, 2015, S.8)

Dies wird dadurch erklärt, dass eine Illusion die sehr authentisch, jedoch nicht perfekt ist, ein unangenehmes Gefühl bei der betrachteten Person erzeugt, da das von ihm gewohnte, zum Beispiel menschliche Aussehen, leicht künstlich wiedergegeben wird. Besonders bei bewegten Objekten ist dieser Effekt sehr groß, wenn eine künstliche Figur menschliche Bewegungen imitiert.

> “For illustration, when an industrial robot is switched off, it is just a greasy machine. But once the robot is programmed to move its gripper like a human hand, we start to feel a certain level of affinity for it.” (Mori et al., 2012, S. 99)

Wird dieser Effekt nicht bedacht, kann es zu Akzeptanzproblemen kommen. Der, im Jahr 2001 veröffentliche Animationsfilm *Final Fantasy: The Spirits Within* war für die damalige Zeit eine bahnbrechende Zurschaustellung der digitalen Technik und wurde auch großteils von den Zuschauern und Zuschauerinnen akzeptiert. Jedoch selbst der Animationsleiter des Films Andy Joney beschrieb seine Arbeit als Puppenspieler von Leichen. Diese Beschreibung traf laut ihm immer stärker zu, umso realistischer die digitalen Figuren wurden. (MacDorman & Ishiguro, 2006, S. 26)

Bei Deep Fakes kann dieser Effekt deswegen auftreten, da ja nur das Gesicht einer Person verändert wird. Somit ist der restliche Körper perfekt natürlich, jedoch können die künstlichen erzeugten Gesichtszüge dafür sorgen, dass der Deep Fake beim Betrachten ein unangenehmes Gefühl erzeugt. Das Gehirn erwartet in diesem Moment die Darstellung einer normalen Person, die fremdwirkende Mimik wirkt deswegen besonders verstörend.

3 Das Zeitalter der künstlichen Intelligenz

Der Forschungsbereich der künstlichen Intelligenz entwickelt sich momentan rasant weiter und dringt immer tiefer und weiter in verschiedenste Anwendungsbereiche ein. Die Auswirkungen, die diese neue Technologie mit sich bringt, kann in gewissen Bereichen ähnlich signifikant gesehen werden, wie der Umstieg von analoger auf digitale Technik. Dies bezieht sich jedoch nur auf den momentan verfügbaren Blick auf dieses Forschungsgebiet. Die zukünftigen Auswirkungen und Errungenschaften, die der Bereich der künstlichen Intelligenz noch bringen kann, sind momentan in keiner Weise einschätzbar.

3.1 Begriffserklärung

Für das Verständnis der nachfolgenden Kapitel ist die Erklärung von gewissen Fachbegriffen notwendig.

3.1.1 Künstliche Intelligenz

Da auch der Begriff Intelligenz nicht problemlos klar definierbar ist, ist auch die Begriffserklärung der Künstlichen Intelligenz nicht konsequent. Eine Definition wäre, dass eine Maschine kognitive, für den menschlichen Verstand typische, Aufgaben absolvieren muss, um in die Kategorie der Künstliche Intelligenz zu fallen. So ist es für eine Maschine sehr einfach einen Menschen in reiner Mathematik zu übertrumpfen, da es in der Mathematik klar definierte Regeln gibt, die einem System direkt beigebracht werden können, und die Eingabe von Zahlen für den Computer direkt verständlich ist. Soll jedoch eine Maschine anhand von Fotos zwischen einem Wolf und einem großen Hund unterschieden, benötigt die Maschine ein angelerntes Vorwissen um die Eingabe, die jedes Mal ein anderes Foto zeigt, zu verarbeiten. Um solch eine Aufgabe zu lösen, muss eine Maschine sich zuvor dieses Wissen anlernen. Beherrscht eine Maschine die Fähigkeit sich komplexes Wissen anzulernen, spricht man von künstlicher Intelligenz. Der Begriff der künstlichen Intelligenz ist als grobes Konzept beziehungsweise als Oberbegriff für die nachfolgenden Definitionen zu sehen welche alle Teilaspekte dieses Fachbereichs sind. (Kreutzer & Sirrenberg, 2019, S. 3-4)

3.1.2 Neuronale Netzwerke

Diese *Neuronalen Netzwerke* sind hypothetische Netzwerke welche aus künstlichen Zellen, auch Neuronen genannt, bestehen. In der Literatur werden sie oft auch *Künstliche Neuronale Netze (KNN)* oder auch *Neuronale Netze* genannt. Die Zellen senden sich gegenseitig Informationen über direkte Verbindungen. Es entsteht somit ein grob vergleichbares System zum menschlichen Gehirn. Neuronale Netzwerke haben ihren Reiz unter anderem darin, dass sie durch Vorgänge wie Maschine Learning trainiert werden können und somit lernfähig sind. Diese Netzwerke können gigantische Größen annehmen. (Biethahn et al., 1998, S. 3-4)

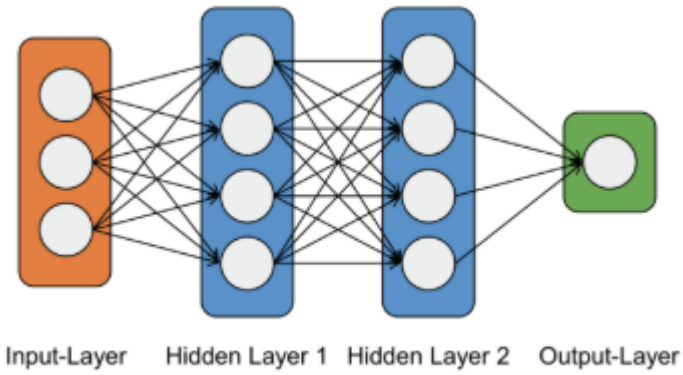

Abbildung 3 – Beispielhafte Darstellung eines Neuronalen Netzwerkes

3.1.3 Maschinelles Lernen - Maschine Learning

Von diesem Begriff spricht man dann, wenn eine Maschine beziehungsweise ein Computer durch eine Summe von realen Beispieldaten und Erfahrungswerten trainiert wird. Statt wie bei einem klassischen Programm, welches auf einem starren Algorithmus, basiert eine Problematik linear abzuarbeiten, findet beim Machine Learning immer die Verwendung von Statistik statt. So werden aus den Beispieldaten der Vergangenheit Verknüpfungen erstellt und Abhängigkeiten berechnet, um somit Zusammenhänge zu generieren. Diese Berechnungen können nicht nur verwendet werden, um Assoziationen zu finden, sondern auch um Vorhersagen für zukünftige Datensätze treffen zu können. (Alpaydin, 2019, S. 31-34)

3.1.3.1 Beaufsichtigtes Lernen

Bei dieser Art des Maschinellen Lernern kontrolliert der Mensch jedes einzelne Element der Trainingsdaten. Das Ziel dieses Algorithmus ist somit bereits gegeben, die Künstliche Intelligenz soll jedoch aus den Eingabedaten die richtigen

Antworten so präzise wie möglich ableiten können. Nachdem die Outputvariablen definiert wurden, wird der Algorithmus mit den vom Menschen vormarkierten Trainingsdaten so lange trainiert, bis der Algorithmus exakt genug ist, um auf neuen Daten angewendet werden zu können.

3.1.3.2 Nicht-überwachtes Lernen

Hier gibt es im Vergleich zum Beaufsichtigten Lernen keine vordefinierten Zieldaten. Die Maschine muss selbst Muster und Zusammenhänge in den Eingabedaten finden und erkennen. Dem Menschen sind diese Zusammenhänge im Zweifelsfall ebenso unbekannt. Die Ergebnisse können somit auch über das menschliche Wissen hinausgehen. Die Eingabedaten, die der Algorithmus erhält, sind nicht beschriftet, nicht geordnet und auch sonst nicht in irgendeiner Form kategorisiert. Solch ein Algorithmus kann etwa dazu verwendet werden, um Social Media Nutzer dahingehend zu analysieren, für welche politische Partei sie bei der nächsten Wahl abstimmen. (Kreutzer & Sirrenberg, 2019, S. 7)

Das solche Techniken eindeutige Konsequenzen haben können, wurde zum Beispiel auch im US-Wahlkampf 2016 offensichtlich. Die Firma Cambridge Analytica hat solch einen Algorithmus verwendet, um unentschiedene Wähler via Social Media Kanälen direkt mit individuellen Inhalten zu erreichen.

> "With real-time monitoring of ad responses on targeted individuals, including real-time substitution to find "click bait" that worked, the ad campaign was able to both maximize its impact and detect trends not visible at the macro scale. Tipping the scale in a few states—with as few as 100,000 voters—using individualized, high-impact messages is sufficient to impact election results. This might not be the only reason for the specific 2016 US election outcome, but there is every indication that it was a useful if not a critical contribution" (Isaak & Hanna, 2018, S. 58)

Fragwürdige Verwendungen wie diese sind unter anderem der Grund, warum Künstliche Intelligenz und ihre Anwendungen immer mehr in den allgemeinen Diskurs der Gesellschaft gestellt werden.

3.1.3.3 Verstärkendes Lernen

Hier verwendet die Maschine eine *Trial and Error*[4] Herangehensweise für das definierte Problem. So müssen Zusammenhänge und Muster selbstständig in den

[4] Hierbei werden solange zufällige Lösungsansätze ausprobiert, bis man sich der gewünschten Lösung annähert

Daten erkannt werden, vordefinierte Zielwerte sind nicht vorhanden. Deswegen muss die Maschine falsche Erkenntnisse immer wieder verwerfen, und positive Erkenntnisse verfeinern und weiterentwickeln. Pro Iteration wird die Maschine entweder belohnt oder bestraft, je nachdem ob die Vorgehensweise sie näher an das Ziel gebracht hat oder nicht. Diese Herangehensweise wird zum Beispiel dann gewählt, wenn es noch keine genauen Trainingsdaten gibt, oder das Ziel selbst noch nicht perfekt definiert ist. Diese Lernweise macht zum Beispiel bei Brettspielen Sinn, wo einzelne Schritte beziehungsweise Spielzüge abgeschlossene Einzelentscheidungen sind, und meist direkt ersichtlich ist, ob ein Spielzug positiv oder negativ für die allgemeine Situation war. (Kreutzer & Sirrenberg, 2019, S. 9)

3.1.4 Deep Learning

Deep Learning ist eine spezielle und oft verwendete Methodik, um Neuronale Netzwerke für eine gewisse Anwendung zu trainieren. Hierzu gibt es zwischen der Eingabeebene und der Ausgabeebene viele einzelne Ebenen, welche der Computer alle durchläuft. Dabei wird ein komplexes System in eine Vielzahl von immer simpler werdender Konzepte aufgeteilt und untereinander angeordnet. Stellt man diesen Vorgang in einer grafischen Darstellung da, würde man eine Grafik mit großer Tiefe erhalten. Daher kommt der Begriff Deep Learning (Tiefes Lernen). (Goodfellow et al., 2016, S. 2)

Die Zwischenebenen sind für den Betrachter und Betrachterin nicht sichtbar und werden auch *Hidden Layer* (Verstecke Ebenen) genannt. Die Anzahl dieser Ebenen kann je nach System über 10.000 hinaus gehen. Jede Ebene gibt nur die selbst berechneten Ergebnisse weiter, und besitzt kein Wissen bezüglich der Informationen, die die vorgelagerte Ebene wiederum von ihrer vorgelagerten Ebene erhalten hat. Durch diesen Aufbau ist beim Deep Learning Prozess nur wenig Vorbereitung der Daten durch den Menschen notwendig, jedoch kommt das System dennoch oft zu besseren Ergebnissen als mit typischen maschinellen Lernansätzen. (Kreutzer & Sirrenberg, 2019, S. 5)

3.1.5 Epochen

Beim Thema Maschine Learning versteht man unter einer Epoche einen kompletten Durchlauf aller Input- beziehungsweise Trainingsdaten. Somit ist die Epochenanzahl ein sehr wichtiger Parameter für die Definition eines Neuronalen Netzwerks. Dieser Wert gibt somit auch Aufschluss wie gut die künstliche Intelligenz die Trainingsdaten schon gelernt hat. Dies bedeutet selbstverständlich nicht, das neuronale Netzwerke mit einer hohen Epochenanzahl automatisch

korrekte Werte für eine gestellte Aufgabe liefern, sondern lediglich, dass dieses Neuronale Netzwerk sich vielfach, bezogen auf die Trainingsdaten, angepasst hat. Denn bei jedem neuen Durchlauf werden die Variablen des Neuronalen Netzwerks angepasst und kontrolliert, ob die Veränderung eine Annäherung zum Ergebnis erzeugt hat. (Hastie et al., 2009)

3.1.6 Batches

Unter einem sogenannten Batch versteht man einen abgetrennten Teil der Trainingsdaten. Solche Batches können aus verschiedenen Gründen sinnvoll sein. So können Computersysteme, welche mit den gesamten Trainingsdaten überfordert wären, mit dem kleineren Datensatz erfolgreich trainiert werden. Dies ist vor allem beim begrenzten Videospeicher von Grafikkarten sinnvoll. Hierbei ist darauf zu achten, dass der Batch die Trainingsdaten gut repräsentiert, da sonst die Ergebnisse stark verzerrt werden. Dies kann man durch eine zufällige Durchmischung garantieren. Sind alle Batches, in die die Trainingsdaten aufgeteilt wurden durchlaufen, hat man eine Epoche vollendet. (Goodfellow et al., 2016, S. 275 - 276)

3.1.7 Convolutional Neural Network - CNN

Diese Art von Neuronalen Netzwerken sind typisch für Aufgabengebiete, welche in den Bereich Visual Computing fallen. Sie bestehen aus Ebenen, welche Neuronen besitzen, die Eingaben verarbeiten und diese nicht-linear an weitere Ebenen weitergeben. Diese Netzwerke sind jedoch so konstruiert, um mit Daten zu arbeiten, die in Gittern angeordnet sind und die einzelnen Gitterzellen stark von ihren benachbarten Zellen beeinflusst werden. Ein offensichtliches Beispiel sind somit Bilddateien. Diese bestehen aus einzelnen Bildpunkten beziehungsweise Pixel, die in einem Gitter angeordnet sind und sich benachbarte Pixel stark beeinflussen. Weiters ist in solch einem Netzwerk jede Ebene dreidimensional. Bei den Inputdaten sind die horizontale und vertikale Dimension durch die Pixelanzahl gegeben, die dritte Dimension entsteht durch die RGB-Werte welche das Bild besitzt. Bei einem gewöhnlichen Neuronalen Netzwerk würden bei einem hochauflösenden Bild von 24 Megapixel, 6000 mal 4000 Pixel mal Farbtiefe, so viele Knotenpunkte entstehen, dass eine schnelle Abarbeitung nicht möglich wäre.

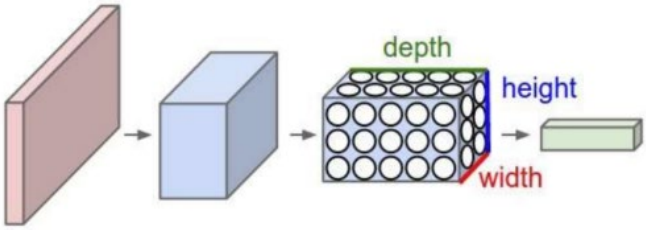

Abbildung 4 – Schematischer Ablauf eines CNN. Links ist ein Inputbild ersichtlich, welches durch Faltung Ebene für Ebene zu einem anderem Volumen transformiert wird (Ahire, 2018)

Ein Convolutional Neural Network reduziert diese Komplexität durch schrittweise Faltung unter Verwendung eines Filters immer weiter, ohne wichtigen Informationsgehalt des Bildes zu verlieren. Durch Verwendung eines Convolutional Neural Network kann so eine hohe Anzahl von hochauflösenden Bildern schnell verarbeitet werden, dies bietet für die Erstellung von Deep Fakes große Vorteile. (Ahire, 2018, S. 118-119)

3.2 Quantifizierung von künstlicher Intelligenz nach dem Automatisierungsgrad

Systeme, welche auf künstlicher Intelligenz basieren, unterscheiden sich klar darin, wie weitreichend die Entscheidungsgewalt beziehungsweise wie hoch der Grad der Automatisierung der künstlichen Intelligenz ist. *Abbildung 5* zeigt diese Quantifizierung in fünf Gruppen.

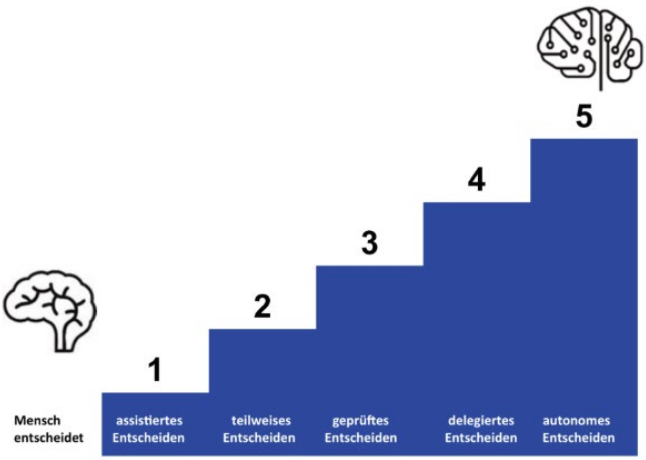

Abbildung 5 – Quantifizierung von künstlicher Intelligenz nach der Automation des Handelns (Kreutzer & Sirrenberg, 2019, S. 14)

Beim *assistierten Entscheiden* wird der Mensch nur bei seiner eigenen Entscheidung unterstützt, wenn zum Beispiel eine Textkorrektur mögliche Änderungen vorschlägt. Bei der zweiten Stufe dieser Automationskala, dem *teilweises Entscheiden*, trifft das System schon Entscheidungen für den Menschen. Ein Beispiel dafür ist die Verwendung von Suchmaschinen, die nicht transparent zeigen, was dem Benutzer oder Benutzerin nun als Ergebnis geliefert wird. Bei der Stufe des *geprüften Entscheidens* wird die Entscheidung des Menschen nochmals von der künstlichen Intelligenz validiert, es passiert also ein Gegencheck. Bei der vierten Stufe, dem *delegierten Entscheiden,* werden bewusst Aufgaben komplett an eine Maschine delegiert und nur noch ein Teilbereich wird selbst vom Menschen absolviert. Bei der höchsten Stufe dieser Skala, dem *autonomen Entscheiden*, entscheidet die Maschine völlig selbstständig und benötigt somit kein Eingreifen des Menschens. In der Automobilindustrie wird momentan stark versucht diesen Status zu erreichen. (Kreutzer & Sirrenberg, 2019)

3.3 Verwendung von neuronalen Netzwerken im Visual Effekts Filmbereich am Beispiel der Rotoskopie

Die Verwendung von neuronalen Netzwerken für die Manipulation von Bewegtbildmaterial ist keinesfalls exklusiv zur Erstellung von Deep Fakes geeignet. Es gibt verschiedenste Anwendungsfälle, die sich in diesem Bereich auftun. Besonders geeignet sind Anwendungen, für welche ein Mensch viele repetitive Schritte ausführen müsste. Ein Beispiel dafür wäre *Rotoskopie*[5], wo per Hand gewisse Bildbereiche im Gesamtbild maskiert werden müssen. Bei Bewegtbildmaterial müssen somit pro Einzelbild minimale Änderungen zum vorhergehenden und bereits freigestellten Einzelbild, vollzogen werden. Das OpenFX Plugin Rotobot[6] verwendet neuronale Netzwerke für den Rotoskopieprozess und bietet darüber hinaus zusätzliche Möglichkeiten. So können verschiedene Motive wie zum Beispiel mehrere Menschen in einer Gruppe jeweils einzeln freigestellt werden und in der Komposition auf einzelne Ebenen aufgeteilt werden.

[5] Hierbei werden Einzelbild für Einzelbild einzelne Segmente im Bild nachgezeichnet beziehungsweise freigestellt, um auf diese Bereiche gezielt Effekte anzuwenden

[6] https://kognat.com/

Abbildung 6 – Rotoskopie unter der Verwendung von Neuronalen Netzwerken (https://kognat.com/example-demos-and-footage/)

Selbst wenn diese Anwendungen noch nicht in allen Szenarien perfekte Ergebnisse garantieren können, bewirken sie jedoch eine große Zeit- und Kostenersparnis. Weiters können solche Werkzeuge Laien ermöglichen, komplexe Kompositingergebnisse ohne großes Fachwissen zu realisieren.

> “The proposed method does not require any additional user-labeled input and generates individual alpha matte for all the detected objects in the image. [...] Our approach could help non-expertise in image compositing tasks and accelerate the image editing process.” (Hu & Clark, 2019, S. 140)

Je nach Motiv variiert die Qualität der Ergebnisse, die durch neuronale Netzwerke erstellt werden. So sind Materialien die Transparenz aufweisen momentan noch eine große Herausforderung für neuronale Netzwerke. Hier erkennt die künstliche Intelligenz dann meist das Motiv hinter dem Objekt mit der Transparenz. Neuronale Netzwerke arbeiten immer nur so gut, wie das Vorwissen, mit dem sie zuvor trainiert worden sind. Will man ein Motiv freistellen, welches das neuronale Netzwerk nicht kennt, wird es auch hier kein zufriedenstellendes Ergebnis liefern können.

> „The image matting network also suffers from the limited amount of object categories it is trained on.” (Hu & Clark, 2019, S.140)

Da sich neuronale Netzwerke mit hoher Geschwindigkeit weiterentwickeln, ist anzunehmen das diese Probleme seltener werden. Der allgemeine Zeitgewinn ist jedoch ein großer Vorteil, so dass diese Algorithmen bereits jetzt den Einzug in die Fachwelt der Bewegtbildmanipulation absolvieren. Selbst kleine Teams aus

Entwicklern und Forschern schaffen es mit Hilfe von neuronalen Netzwerken hoch qualitative Ergebnisse zu erzielen.

> „We show that this algorithm can perform comparably and even surpass the rotoscoping capabilities of After Effects' RotoBrush tool, in a variety of scenes comprising different lighting conditions, movements, and subjects. This makes it suitable for an integration within a VFX pipeline." (Estrada et al., 2019)

Deswegen ist es naheliegend, das große Softwarehersteller, die dementsprechend mehr Ressourcen haben, solche Werkzeuge ebenfalls entwickeln und veröffentlichen werden. Für Visual Effects Fachleute wird es früher oder später zum Alltag gehören, Werkzeuge, die auf neuronale Netzwerke zurückgreifen, zu verwenden. Momentan stehen wir hier erst am Beginn einer neuen Technologie und die Anwendungszwecke werden breiter und umfangreicher werden. Der Softwareproduzent Foundry, welcher das professionelle Kompositingsoftwareprodukt *Nuke* entwickelt, sieht die Verwendung von künstlicher Intelligenz nicht nur beschränkt auf Visual Effects Techniken wie Rotoskopie. So gibt Mitgründer und Hauptentwickler Simon Robinson in einem Interview an, dass er diese Technologie weiters als Werkzeug sieht, um den Arbeitsablauf und die Organisation rund um komplexe Projektstruktur einfacher zu gestalten. (FAILES, 2019, S. 27)

3.4 Massive Weiterentwicklungen bei künstlicher Intelligenz durch Fortschritt von Grafikkarten

Die massive Weiterentwicklung im Bereich von Maschine Learning beziehungsweise künstlicher Intelligenz korreliert sehr stark mit dem enormen technischen Fortschritt, den Grafikkarten in den letzten Jahren erlebt haben. Ein großer Meilenstein war, als der Grafikkarten Hersteller Nvidia im Jahr 2006 *CUDA* präsentierte. Der Begriff *CUDA* steht für *Compute Unified Device Architecture*. Dabei handelt es sich um eine Schnittstelle, die es Softwareentwicklern ermöglicht Programme beziehungsweise Algorithmen über den Grafikprozessor auszuführen, und somit eine viel höhere Parallelisierbarkeit des Softwareprodukts zu erreichen. Zuvor wurden Grafikkarten fast ausschließlich dafür verwendet, um Computerspiele zu rendern. Vor allem auf die theoretische Rechenleistung und die

Speicherbandbreite bezogen haben sich Grafikkarten gegenüber klassischen Prozessoren rasant weiterentwickelt. Weiters eignet sich der grundsätzliche Aufbau von Grafikkarten für den Maschine Learning Prozess weitaus besser. (WyntersErik, 2011, Seite 59)

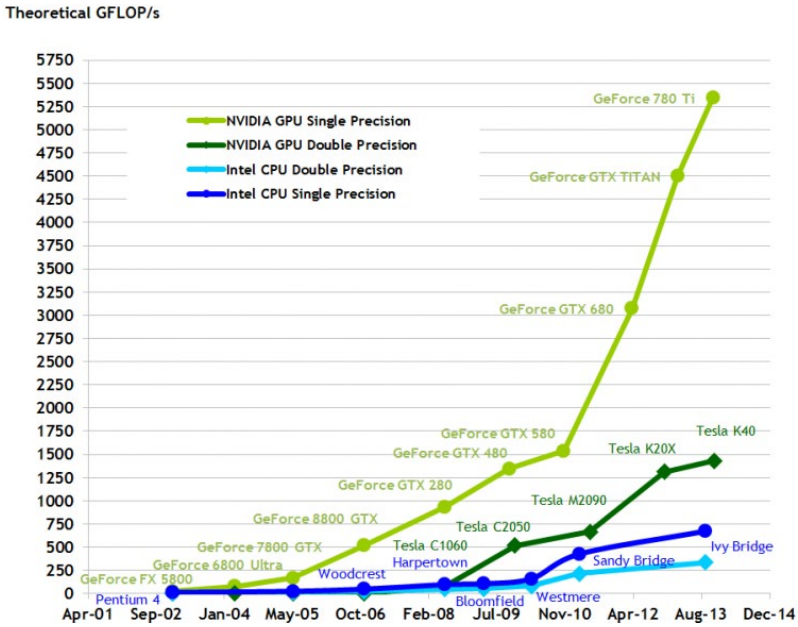

Abbildung 7 – Entwicklung der Rechenleistung von Grafikkarten und Prozessoren zwischen April 2001 und Dezember 2014 (Huang et al., 2016, S. 2)

Ein Beispiel, welches diesen Faktor gut unterstreicht, ist ein Projekt von Google, bei welchem einer künstlichen Intelligenz beigebracht wurde Katzen zu erkennen. Als Trainingsdaten wurden dafür Videos von der Videoplattform *YouTube* verwendet. Die Hardware hinter diesem Projekt beinhaltete um die 1000 Serverprozessoren mit ungefähr 16.000 Prozessorkernen. Die Kosten dafür betrugen fünf Milliarden Dollar. Nur ein Jahr später erreichte die Standford Universität in Kooperation mit Nvidia die gleiche Rechenleistung mit drei Grafikkarten. Die Gesamtkosten betrugen 33.000 Dollar. Bei beiden Systemen betrug die Trainingszeit eine Woche. (Ilievski et al., 2018, S. 2)

Das momentan am meisten verwendete und für Deep Fakes relevante Deep Learning Framework *Tensorflow* inkludiert viele Funktionen von *CUDA*. Dies ist in *Abbildung 8* ersichtlich, alle Funktionen auf der Ebene *Numerical computation package* in der Spalte GPU basieren auf *CUDA*. Die meisten anderen Frameworks dieser Art inkludieren ebenfalls *CUDA* Schnittstellen, nur wenige unterstützen zusätzlich den nicht proprietären Standard *OpenCL*.

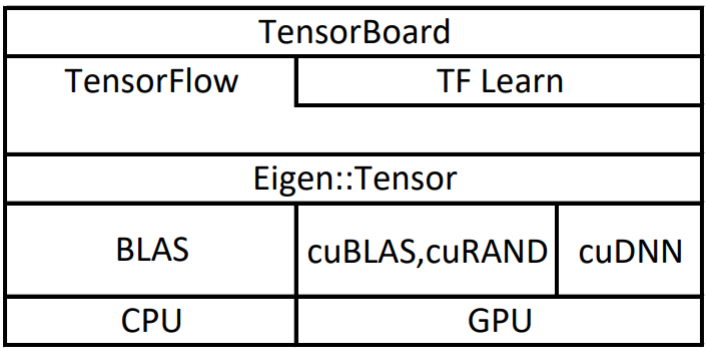

Abbildung 8 – Prinzipieller Aufbau von TensorFlow. cuBLAS, cuRAND und cuDNN sind Funktionen die auf CUDA basieren. (Ilievski et al., 2018, Seite 3)

Ein weiteres Beispiel über den großen Leistungsunterschied kann man bei Multiplikationen von Matrizen bemerken. Die Abkürzung *BLAS* steht für Basic Linear Algebra Subprograms und wird von einem klassischen Prozessor für eine Matrixmultiplikation herangezogen. *cuBLAS* ist eine Variante von *BLAS,* die über eine Nvidia Grafikkarte läuft und auch von *Tensorflow* verwendet wird.

N	CPU	GPU	Speedup
4000	711.64	0.61	1166.6
6000	2643.47	2.68	986.4
8000	6689.19	3.26	2051.9
10000	13645.81	11.62	1174.3

Abbildung 9 – Laufzeiten von Multiplikationen von quadratischen Matrizen mit den Dimensionen N auf einem Prozessor und einer Grafikkarte (WyntersErik, 2011, Seite 62)

Wie *Abbildung 9* zeigt, ergeben sich teilweise Laufzeiten die auf einer Grafikkarte 2000-mal kürzer ausfallen als bei einem klassischen Prozessor aus einem ähnlichen Preisbereich. Dies bedeutet nicht das eine Grafikkarte jegliche Aufgaben immer um solch eine Größenordnung schneller absolvieren kann. Ist es jedoch möglich eine Aufgabenstellung durch Parallelisierung für Grafikkarten anzupassen kann man enorme Zeitgewinne erwarten.

4 Deep Fakes

Der Begriff Deep Fakes beschreibt Bewegtbildfälschungen, bei welchen das Gesicht der dargestellten Person durch das Gesicht einer anderen Person ersetzt wurde. Dies wird durch die Verwendung von Deep Learning Algorithmen vollbracht. Diese Fälschungen können auch die Rekonstruktion von Audiomaterial der eingefügten Personen beinhalten, um kombiniert mit entsprechender Lippensynchronität die Qualität der Fälschung weiter zu erhöhen. Somit zeigt ein Deep Fake immer eine oder mehrere Personen, die eine Tätigkeit vollführen oder etwas sprachlich wiedergeben, was so in diesem Szenario nie stattgefunden hat.

Bewegtbildfälschungen beziehungsweise Filmmaterial bei dem bewusst Gesichter ausgetauscht werden, sind keinesfalls etwas fundamental Neues oder Seltenes. Im Filmbereich ist dies eine häufig angewandte Technik, die zum Beispiel durch Motion Capture Systeme gelöst werden kann. Die Verwendung der Deep Learning Technologie macht jedoch Deep Fakes zu etwas Besonderem. Um ähnliche Fälschungsergebnisse mit herkömmlichen Visual Effects Methodiken zu erzielen, sind hohe Kosten und ein Produktionsteam aus verschiedenen Fachleuten notwendig. Die Verwendung von künstlicher Intelligenz in diesem Bereich erlaubt es nicht nur die Bewegtbildfälschung viel glaubhafter zu gestalten, sondern verkleinert auch den Aufwand um ein Vielfaches. (Chesney & Citron, 2018)

Der Begriff Deep Fake ist seit 2017 im Umlauf. Auf der Internetplattform Reddit stellte ein User unter dem Benutzernamen *Deep Fakes* mehrere pornografische Videos ins Internet. Dabei wurden Gesichter weiblicher Pornodarstellerinnen mit den Gesichtern von berühmten weiblichen Schauspielerinnen ausgetauscht. Zwar wurden die Fälschungen schnell entlarvt, jedoch veröffentlichte der Autor auch den verwendeten Programmcode inklusive eines dokumentierten Arbeitsablaufs. Somit waren die notwendigen Bestandteile dieses Erstellungsprozesses veröffentlicht und auch für Nutzer ohne Reddit-Konto ersichtlich. In Folge dessen tauchte schnell eine Vielzahl weiteren Deep Fakes im Internet auf. Auch der Programmcode wurde unter den Forenmitgliedern weiterentwickelt und verbessert. Die Plattform Reddit sperrte später dieses Unterforum und verbot Deep Fake Inhalte komplett. Weitere Plattformen wie der Nachrichtendienst *Twitter* oder einige pornografische Webseiten verboten diese Inhalte ebenfalls. (Woolley, 2020)

4.1 Generative Adversarial Networks

Für die Erstellung dieser Fälschungen verwendete der damals veröffentliche Quellcode des Users *Deep Fakes* die von Google veröffentlichte Open Source Plattform *TensorFlow*. Konkret verwendete der Benutzer dieses Werkzeug für die Erstellung sogenannter Generative Adversarial Networks (GAN).

> „Fake videos can now be created using a machine learning technique called a “generative adversarial network”, or a GAN. A graduate student, Ian Goodfellow, invented GANs in 2014 as a way to algorithmically generate new types of data out of existing data sets. For instance, a GAN can look at thousands of photos of Barack Obama, and then produce a new photo that approximates those photos without being an exact copy of any one of them, as if it has come up with an entirely new portrait of the former president not yet taken. GANs might also be used to generate new audio from existing audio, or new text from existing text – it is a multi-use technology.” (Schwartz, 2018)

Diese Generative Adversarial Networks sind Maschine Learning Algorithmen, welche in die Kategorie des unüberwachten Lernens fallen. Ein GAN besteht aus zwei neuronalen Netzwerken, welche ein Nullsummenspiel absolvieren. Das eine neuronale Netzwerk wird als Generator bezeichnet, das zweite neuronale Netzwerk als Diskriminator. Der Generator versucht falsche Datensätze zu erzeugen, welche der Diskriminator nicht von den echten Trainingsdaten unterscheiden kann. Durch dieses Design ist kein Eingriff in den Lernprozess notwendig. Sowohl der Generator als auch der Diskriminator trainieren ihre Fähigkeiten immer weiter, um den jeweiligen Gegenspieler zu überlisten. (Goodfellow et al., 2014, S. 1)

4.2 Erklärungsbeispiel – Präsident Barack Obama

Eines der bekanntesten Deep Fake Beispiele, welches keine pornografische Natur hat, ist eine Forschungsarbeit der Universität von Washington. Hierbei erstellten die Forscher und Forscherinnen mit Hilfe von neuronalen Netzwerken einen Deep Fake, wobei sie sowohl Audio synthetisch kreierten also auch die dazu passenden Lippen- und Kopfbewegungen fälschten. Als Ziel wählten sie den ehemaligen amerikanischen Präsidenten Barack Obama aus. Sein Gesicht wurde hierbei nicht auf einen fremden Körper manipuliert, sondern ein offiziell existierendes Video wurde von ihm als Basis verwendet, um die Fälschung so echt wie möglich wirken

zu lassen. Damit ist dieses Beispiel kein klassischer Deep Fake, zeigt jedoch welche Fälschungsergebnisse mittlerweile möglich sind.

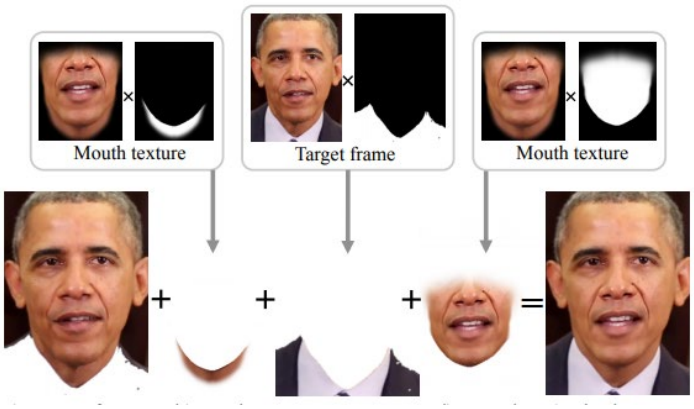

Abbildung 10 – Die verschiedene Ebenen der Fälschung von Barrack Obama erstellt durch die Universität von Washington (Suwajanakorn et al., 2017, S. 8)

Dabei entschieden die Wissenschaftler jedoch keine fabrizierten Botschaften zu verwenden, sondern ließen Barack Obama Sätze formulieren, die er in anderen Situationen schon einmal ausgesprochen hatte. Dies trägt weiter dazu bei, dass diese Fälschung bei normaler Betrachtung nicht von einem echten Video zu unterscheiden ist.

Für das 66 Sekunden lange Outputvideo benötigte ein Einzelkernprozessor 45 Minuten, um die bei einer Framerate von 30 Bilder pro Sekunde 1980 Einzelbilder zu generieren. Die Forscher parallelisierten den Prozess danach auf einen Prozessor mit 24 Kernen und konnten so die Erstellungszeit auf nur drei Minuten kürzen. Als Trainingsdaten für den Algorithmen wurden 17 Stunden Videomaterial verwendet, welches frei online verfügbar war. Dabei handelte es sich um die wöchentlichen Ansprachen des damaligen Präsidenten. Dabei wurde ein Fünftel des Materials zur Überprüfung und die anderen 80% als direkte Trainingsdaten verwendet. Der ganze Lernprozess dauerte drei Stunden für 300 Epochen bei der Verwendung einer leistungsfähigen *NVIDIA Titan X* Grafikkarte.

Weiters zeigt dieses Beispiel auch wie wichtig eine hohe Menge an qualitativen Trainingsdaten ist. Die Forscher testeten den Algorithmus einmal mit 0,35% der vorhandenen Trainingsdaten, einmal mit 10% der Trainingsdaten und einmal mit 50% der Trainingsdaten. Jeder Anstieg zeigte eine eindeutige Verbesserung des Outputvideos, selbst zwischen 7 Stunden (50%) und 14 Stunden war ein eindeutiger Qualitätsunterschied erkennbar. (Suwajanakorn et al., 2017 S. 7 - 9)

4.3 Besonders gefährdete Personengruppen

Für die Erstellung von Deep Fakes sind im Trainingsprozess eine Vielzahl von Einzelbildern notwendig. Der Algorithmus muss, um ein glaubhaftes Ergebnis zu erreichen, die verschiedenen Gesichtsmerkmale und Gesichtsausdrücke lernen können. Dementsprechend sind für den Lernprozess Szenen notwendig, die diese Gesichtspositionen zeigen. Im späteren Praxisteil dieser Diplomarbeit werden für die Erstellung der Deep Fakes mindestens 3.000 Einzelbilder verwendet, teilweise auch mehr. Eine hohe Anzahl an Bildern bedeutet jedoch nicht automatisch ein gutes Trainingsergebnis. Die Einzelbilder sollen hochauflösend sein und bevorzugt auch keine Bewegungsunschärfe aufweisen.

Diese Voraussetzungen definieren somit im Umkehrschluss die gefährdeten Personengruppen. Personen, die im öffentlichen Leben stehen, werden oft fotografiert oder gefilmt und diese Inhalte auch öffentlich publiziert. Bei Berufsgruppen wie Schauspielern und Schauspielerinnen oder TV-Moderatoren und Moderatorinnen, die konkret hochwertige Bewegtbildformate produzieren ist das vorhanden sein solcher Inhalte noch offensichtlicher. Diese Inhalte eignen sich optimal für den Deep Fake Trainingsprozess, da sie leicht zugänglich sind und hochauflösend produziert und publiziert werden. Dies ist der Grund warum der Großteil der Deep Fake Beispiele momentan berühmte Schauspieler und Schauspielerinnen zeigt. Die zuvor erwähnten 3.000 Einzelbilder sind vor allem mit Videoformaten leicht umsetzbar. So beinhaltet ein zweiminütiges Video mit 25 Einzelbildern pro Sekunde schon 3.000 Einzelbilder. Ein hochauflösendes Video-Interview, welches mehrere Minuten lang ist reicht somit aus, um selbst nach Aussortierung von unscharfen Einzelbildern, mehrere Tausend hochwertige Bilder für den Trainingsprozess zu erhalten.

Die Gefahr gilt nicht exklusiv für berühmte Persönlichkeiten. In Zeiten von Social Media und mit der immer weiter voranschreitenden Entwicklung von Smartphone-Kameras, entsteht die Gefahr, dass selbst Personen, die sich nicht als Personen des öffentlichen Lebens sehen, gefährdet werden. Inhalte, die auf diesen Plattformen publiziert sind, sind oft von jedem einsehbar, wenn nicht explizite Einstellungen vorgenommen werden. Hier können Dritte problemlos mehrere Minuten an Videoclips kopieren und diese als Trainingsdaten verwenden. Kriminelle Handlung wie Hacking oder Diebstahl von persönlichen Videos beziehungsweise Fotos können natürlich auch dazu führen, dass Trainingsdaten für Deep Fakes gewonnen werden. Weiters sind auch Personen gefährdet, die Inhalte nicht selbst publizieren, sich jedoch von anderen Personen regelmäßig fotografieren beziehungsweise filmen lassen. Ein klassisches Beispiel dafür wären Pärchen. Das Phänomen Rachepornos zeigt auf, das Inhalte, welche während

einer Beziehung unter Zustimmung entstanden sind, nach Beziehungsende als Rachewerkzeug verwendet werden können.

4.4 Rechtliche Situation

Da die verwendeten Inhalte eines Deep Fakes sehr unterschiedlich sein können, ist auch die rechtliche Situation dieses Phänomens komplex. Vor allem in den Vereinigen Staaten von Amerika wurde dieses Thema schon breiter diskutiert. Diese Rechtsfassungen dienen als Beispiel.

Die ersten Deep Fakes welche hauptsächlich pornografischer Natur waren, würden im Sinne von sexueller Belästigung strafrechtlich verfolgt werden können. In der Judikative gab es schon vor längerer Zeit einen Fokus auf Rachepornos, zu denen auch Deep Fake Videos zählen können. Der amerikanische Bundesstaat Texas hat im September 2019 ein Gesetz verabschiedet, welches Deep Fakes verbietet, die darauf abzielen politische Wahlen zu beeinflussen oder das Ansehen eines politischen Kandidaten beziehungsweise Kandidatin zu beschädigen. Im Bundesstaat Kalifornien wiederum wird argumentiert, dass neue Gesetze gar nicht notwendig sind, da es schon ein umfassendes Recht auf das eigene Bild, den eigenen Namen, die eigene Stimme und die eigene Signatur gibt, und jegliche unerlaubte und schadhafte Verwendung eines dieser Elemente strafrechtliche Konsequenzen bedeutet.

Im Vereinigten Königreich ist die Gesetzeslage weniger eindeutig und greift bei Deep Fakes auf verschiedene einzelne Aspekte zurück. Zwangsweise entstehen bei der Erstellung von Deep Fakes Urheberrechtsverletzungen. Sowohl die Trainingsdaten als auch das Basisvideo, welches manipuliert wird, sind mit Urheberrecht geschützt. Jedoch schützt das Urheberecht nur den Ersteller des Videos und somit nicht zwangsweise die Person, die abgebildet ist. Das Urheberrecht schützt das Werk, beziehungsweise den Werksurheber. Das Opfer muss somit vor Gericht beweisen, dass es ein außerordentliches persönliches Interesse am gezeigten Material hat. Weiters stellen Deep Fakes einen Sonderfall da, da solche Fälschungen nicht nur aus einem Bild oder Video bestehen. Das neuronale Netzwerk kann tausende von Bildern und Videos für einen Deep Fake analysieren, von denen jedes Bild und Video einen anderen Urheber oder Urheberin haben kann.

In den Mitgliedsstaaten der Europäischen Union gilt die *Europäische Menschenrechtskonvention* welche im Artikel 8 das Recht auf Respekt auf das private Leben, das Familienleben, das Recht auf Wohnung und den Schutz der

Korrespondenz definiert. Somit muss das Opfer beweisen, dass die verwendeten Bilder und Videos im privaten Umfeld aufgenommen worden sind beziehungsweise die entsprechende Person in der geschützten Privatsphäre zeigen. Diese Rechte müssen jedoch immer im Gleichgewicht mit Artikel 10, der Freiheit der Meinungsäußerung, abgewogen werden, da für Politiker und Politikerinnen, die regelmäßigen öffentlichen Auftritte absolvieren andere Maßstäbe gelten als für die meisten anderen Personen.

Hier wird jedoch davon ausgegangen, dass das Opfer weiß, welche Bilder und Videos verwendet wurden, um den Deep Fake zu erstellen. Dies ist jedoch bei einem Deep Fake für den Betrachter oder die Betrachterin nicht ersichtlich. Das Endergebnis ist ein künstlich erzeugtes Video einer realen Person, und somit kommt hier das Urheberrecht nicht auf das Endprodukt zu tragen.

Die hohe Anzahl an verschiedenen Einzelbildern und Videos, die in den Trainingsdaten beinhaltet sind, führt zu einer weiteren Problematik. So können diese Elemente in verschiedenen Ländern aufgenommen worden und auf verschiedene internationale Server hochgeladen worden sein, wobei in jedem Land andere Gesetze gelten.

> „When speaking to The Washington Post about the pornographic deepfakes made of her, Scarlett Johansson lamented that ‘every country has their own legalese regarding the right to your own image, so while you may be able to take down sites in the U.S. that are using your face, the same rules might not apply in Germany. I have sadly been down this road many, many times.” (Farish, 2020, S.48)

Weiters muss auch das Land, in welchen der Deep Fake produziert wurde, nicht mit dem Land übereinstimmen in welchem dieser schlussendlich veröffentlicht wird.

Damit wird klar ersichtlich das es aus rechtlicher Sicht unmöglich ist, sich vollkommen gegen Deep Fakes schützen zu können. Vor allem da die juristische Konsequenz immer erst dann ansetzen kann, wenn es für das Opfer bereits zu spät ist. (Farish, 2020, S. 41 - 48)

> „In contrast to ex ante technological mechanisms that could prevent the use or dissemination of deepfakes, legal action of this sort is inherently an ex post process designed to attribute liability and recover a legal or equitable remedy. For several reasons including those mentioned below, deepfakes may ultimately prove immune to legal action, be they rooted in publicity theories or otherwise.” (Farish, 2020, S. 48)

Denn selbst wenn das Opfer vor Gericht den Rechtsstreit gewinnt und eine Löschung erzwingen kann, ist es leicht möglich, dass das entsprechende Video sich schon auf mehreren verschiedenen Servern und Plattformen verteilt hat und nicht mehr entfernbar ist. Ob es weitere private Kopien auf anderen Endgeräten gibt ist auch nicht nachvollziehbar, somit sind zukünftige Uploads weiterhin möglich.

Zusätzlich ist es für ein Opfer schwierig den Autor oder die Autorin eindeutig zu identifizieren, da die Veröffentlichung solcher Videos über Internetplattformen passieren, und somit man auf die Hilfe der entsprechenden Plattformen angewiesen ist, um genauere Information über den Deep Fake Autor oder Autorin zu erhalten. Verwendet die Person hinter der Erstellung des Deep Fakes für die Veröffentlichung jedoch Technologien wie TOR-Netzwerke[7], die explizit dafür entwickelt worden sind, um digitale Fußspuren zu reduzieren, ist es sehr schwer den Autor oder die Autorin ausfindig zu machen. Somit bleibt dem Opfer in diesem Fall nur die Möglichkeit gegen die Plattform vorzugehen, die die Deep Fakes auf ihren Server bereitstellt und somit anderen Benutzern zur Betrachtung angeboten haben. In den Vereinigten Staaten von Amerika ist es sehr schwer die Plattformen in solch einer Situation erfolgreich zu klagen. Der Abschnitt 230 des *Communications Decency Act* gibt Onlineplattformen Immunität gegenüber der Haftung von privaten beziehungsweise sensiblen Daten, welche von Benutzern und Benutzerinnen veröffentlicht werden. Das Gesetz wurde im Jahr 1996 veröffentlicht und hatte das Ziel, Plattformen von unzähligen Anklagen zu schützen und gleichzeitig so das Wachstum des damals noch neuen Mediums Internet zu fördern. Plattformen, die von sich selbst aus beleidigende und anstößige Inhalte herausfiltern und zensieren, sind ebenfalls von Zivilklagen geschützt. Der Gedanke dahinter war, das Firmen nicht dafür verklagt werden sollten, wenn sie ihre eigenen Internetplattformen von fragwürdigen Inhalten sauber halten wollen. (Chesney & Citron, 2018, S. 1792 - 1796)

> „On one hand, the law has created an open environment for hosting and distributing user-generated online content. On the other, it has generated an environment in which it is exceptionally hard to hold providers accountable, even in egregious circumstances involving systematic disinformation and falsehoods. Courts have extended the immunity provision to a remarkable array of scenarios. They include instances where a provider republished content knowing it violated the

[7] Dabei handelt es sich um ein Netzwerk, welches darauf abzielt, Verbindungsdaten zu anonymisieren

> law; solicited illegal content while ensuring that those responsible could not be identified; altered its user interface to ensure that criminals could were not caught; and sold dangerous products." (Chesney & Citron, 2018, S. 1798)

Da sowohl die strafrechtliche Verfolgung des Autors oder der Autorin des Deep Fakes, als auch der Plattformen, welche den Inhalt veröffentlichen, sehr komplex und schwierig ist, könnte man meinen, dass ein generelles Verbot von Deep Fakes notwendig sei. Jedoch ist es auch fraglich, ob ein Deep Fakes Verbot sinnvoll, beziehungsweise überhaupt verfassungsrechtlich umsetzbar ist. Denn digitale Manipulation ist prinzipiell nicht rein schädlich, sondern kann auch positive Aspekte mit sich bringen. Selbst wenn Deep Fake Videos Falschaussagen eines Politikers oder einer Politikerin verbreiten würden, wären diese Videos wohl dennoch durch das Gesetz der freien Meinungsäußerungen geschützt. (Chesney & Citron, 2018 S. 1788 - 1789)

4.5 Positive Einsatzzwecke von Deep Fakes

Selbstverständlich ist die Entwicklung von Deep Fakes nicht nur negativer Natur. Diese Technik kann und wird sehr wohl für positive Einsatzzwecke eingesetzt, wie beispielsweise in der Bildung und im Kunstbereich.

4.5.1 Bildung

Im Bildungsbereich können zum Beispiel Deep Fakes verwendet werden, um aus Fotografien berühmter Persönlichkeiten anschauliches Filmmaterial zu erzeugen. Dieses Filmmaterial kann verstorbene Persönlichkeiten in Bewegtbildform zeigen, von welchen kaum echte Videos verfügbar sind. Berühmte Zitate könnten in Bewegtbildform wiedergegeben werden, anstatt dass sie nur von Lehrkräften vorgetragen werden. Ein konkretes Beispiel dafür ist das Kunstprojekt *In the Event of a Moon Disaster* welches vom MIT *Center for Advanced Virtuality* umgesetzt worden ist. Dieser Deep Fake zeigt die TV-Ansprache, welche die amerikanische Bevölkerung damals gesehen hätte, wenn die Apollo 11 Raumfahrtmission schief gegangen wäre und die Astronauten nicht auf die Erde zurückkehren hätten können. Da man damals dieses Szenario durchaus für plausibel hielt, wurde eine entsprechende Rede für Präsident Richard Nixon vorbereitet. Die Arbeit wird in Form einer Kunstausstellung präsentiert und klärt auch über das Phänomen Deep Fakes auf. (Day, 2019)

4.5.2 Filmbereich

Im Filmbereich findet ein ähnlicher Ansatz schon länger Verwendung. So wurde zum Beispiel 2016 im Film *Star Wars: Rogue One* eine virtuelle Kopie des verstorbenen Schauspielers Peter Cushing verwendet, um weiterhin den gleichen Charakter verwenden zu können. Um diese Vorgangsweise für die Filmstudios noch leichter zu gestalten, lassen sich manche Schauspieler und Schauspielerinnen mittlerweile scannen um auch nach ihrem Ableben in Filmen als digitale Kopien noch weiter erscheinen zu können. (Chesney & Citron, 2018)

> “But the few individuals who are willing to drop hundreds of thousands of dollars could be investing in the future. Getting scanned at a young age can let you continue to play younger parts, like Samuel L. Jackson in the upcoming Captain Marvel. And you could potentially bring in money for your family by licensing your image to studios after your death.” (Winick, 2019)

Weiters kann es als Absicherung für Filmproduktionen dienen. Verletzungen, oder Entstellungen von Schauspieler und Schauspielerinnen können viel leichter digital ersetzt werden, wenn zuvor schon hochqualitative Scans der jeweiligen Personen vorhanden sind.

4.6 Mögliche gefährliche Verwendungszwecke von Deep Fakes

Die Möglichkeit sehr gute Bewegtbildfälschungen zu erzeugen gibt es im professionellen Filmbereich schon lange, jedoch war die Erstellung immer mit großem Ressourcenaufwand und Kosten verbunden. Durch die immer größere Verbreitung von Anwendungen, die auf künstlicher Intelligenz basieren, wird dieser Prozess so sehr simplifiziert, dass es auch für Personen ohne tiefgehende Expertise möglich wird, Fälschungen in höchster Qualität zu erzeugen. Als Werkzeug werden nur gewöhnliche Computersysteme benötigt, so wie sie die meisten Menschen in westlichen Ländern ohnehin schon besitzen. Früher waren hochwertige Visual Effects und Animationsfilm Projekte nur wenigen Studios beziehungsweise Firmen vorbehalten, die sich die entsprechenden Hochleistungssysteme leisten konnten.

> „While deep-fake technology will bring with it certain benefits, it also will introduce many harms. The marketplace of ideas already suffers from truth decay as our networked information environment interacts in toxic

> ways with our cognitive biases. Deep fakes will exacerbate this problem significantly. Individuals and businesses will face novel forms of exploitation, intimidation, and personal sabotage. The risks to our democracy and to national security are profound as well." (Chesney & Citron, 2018, S. 1754)

Deswegen gilt anzunehmen, dass durch die breite Verfügbarkeit dieser Technologie diese auch immer mehr für gefährliche beziehungsweise schadhafte Gründe eingesetzt werden wird. Das tatsächliche Schadensausmaß von Deep Fakes ist noch nicht abzusehen. Selbst wenn ein Deep Fake nur kurze Zeit nach der Veröffentlichung als entsprechende Fälschung entlarvt wird, kann der verursachte Schaden schon irreparabel sein. Hier folgt nun eine kurze Auflistung möglicher gefährlicher Verwendungszwecke:

4.6.1 Demütigung

Deep Fakes wurden momentan fast ausschließlich verwendet, um Einzelpersonen zu schaden. Es wurden pornografische Inhalte erstellt, bei welchen auf die Körper von Erotik-Darstellerinnen die Gesichter von berühmten Schauspielerinnen generiert wurden. Diese erniedrigenden Videos sind für die abgebildeten Personen schwer fassbar und teilweise nicht entfernbar, weil sie auf mehreren Plattformen mit verschiedenen Rechtslagen geteilt worden sind. Langfristige Rufschädigung sowie ein Gefühl von Erniedrigung und Angst vor weiteren Veröffentlichungen sind die Folge. Mary Anne Frank, Professorin für Strafrecht an der Universtät von Miami, und Präsidentin der Initative *Cyber Civil Rights Initative* beschreibt die Technologie mit folgenden Worten:

> „If you were the worst misogynist in the world, this technology would allow you to accomplish whatever you wanted." (Harwell, 2018)

4.6.2 Erpressung

Solche Inhalte können jedoch nicht nur zur Demütigung der Betroffenen oder der eigenen Befriedigung der Täter dienen, sondern können auch zur Erpressung verwendet werden. Zum Beispiel Drohungen entsprechende Deep Fakes zu veröffentlichen, wenn gewisse Forderungen nicht erfüllt werden. Dies kann Einzelpersonen, welche gerade einen Sorgerechtsstreit durchstehen genauso treffen wie Politiker und Politikerinnen, die bei einer wichtigen Wahl antreten wie auch ganze Organisationen und Firmen.

4.6.3 Identitätsdiebstahl

Weiters könnte diese Technologie auch dafür verwendet werden, um Identitätsdiebstahl zu begehen und sich so über das Internet als jemand anderer auszugeben, indem man entsprechende Videobotschaften veröffentlicht und versendet. Dementsprechend könnte man mit falschen Spendenaufrufen oder Forderungen Geldeinnahmen erzielen oder Vertraute der gefälschten Personen dazu bringen persönliche Informationen oder Daten zu teilen. Auch könnte man Entführungen fälschen oder verschollen geglaubte Personen wieder in den Mittelpunkt der Öffentlichkeit bringen und davon profitieren.

4.7 Potential zur enorm schnellen Verbreitung

Die Verbreitung und einfache Verwendungsmöglichkeiten dieser Technologie ist nicht aufhaltbar. Wie auch bei anderen Technologieneuentwicklungen, kommen Produkte, die auf diesen neuen Technologien basieren, früher oder später auf den Markt und werden vom Normalverbraucher verwendet. Die Forschung in diesem Bereich wird weiter vorangetrieben, Produkte werden immer benutzerfreundlicher und mehr simplifiziert. Weiters ist es auch einfacher denn je, solche manipulierten Bewegtbilder online zu publizieren und sie so für unzählige Menschen verfügbar so machen. Vor der aktuellen Bedeutung des Internets war es schwer, mit Falschmeldungen große beziehungsweise internationale Aufmerksamkeit zu erzielen, da nur wenige Medienunternehmen den Markt beherrschten und ihre Inhalte selbst wählten und überprüften. Doch die Medienlandschaft hat sich durch die riesige Verbreitung des Internets massiv verändert. Um dies zu verdeutlichen muss man nur betrachten, wie viel Bewegtbildmaterial auf der Internetplattform YouTube pro Minute hochgeladen wird:

> „Ben McOwen Wilson, YouTube EMEA's regional director, highlighted the scale of today's operation. "We now have over 500 hours of new content uploaded onto the platform every minute," he said in an interview […] " (Frangoul, 2018)

Es ist momentan unvorstellbar, diese große Menge an Material zu überprüfen, weder mit Computersystemen noch durch Mitarbeiter. Dies sind bloß die Größendimensionen einer von vielen großen Internetplattformen. Weiters wäre dieser enorme Aufwand mit hohen Kosten verbunden, welche private Unternehmen vermeiden möchten, wenn sie gesetzlich nicht dazu gezwungen werden. (Chesney & Citron, 2018, 1764 - 1765)

Ein weiterer Faktor, der die Verbreitung von Deep Fakes fördern kann, ist die Tatsache, dass Menschen dazu tendieren negative und neuartige Nachrichten eher zu teilen, da sie unsere Aufmerksamkeit stärker gewinnen. Dies hat eine größere Untersuchung eines Forschungsteam des Massachusetts Institute of Technology (MIT) gezeigt. Dabei wurden falsche und echte Nachrichtenmeldungen untersucht, die zwischen dem Jahr 2006 und 2017 auf der Nachrichtenplattform Twitter geteilt wurden. Insgesamt umfasste die Untersuchung 126.000 Nachrichtenmeldungen welche von 3 Millionen Menschen mehr als 4,5 Millionen Mal geteilt wurden. Die Studie kam zu dem Ergebnis, dass Falschmeldungen Benutzer und Benutzerinnen zehnmal so schnell erreicht haben wie wahre Nachrichtenmeldungen. Politische Falschmeldungen wurden am öftesten geteilt. Diese Falschmeldungen erreichten 20.000 Menschen in einem Drittel der Zeit, die anderen Typen von Falschmeldungen brauchten, um 10.000 Personen zu erreichen.

Dies liegt jedoch nicht an gezielten menschlichen *Fakeaccounts* oder gar Computern, sogenannten Bots, sondern tatsächlich großteils an typischen Benutzern und Benutzerinnen. Denn selbst, wenn nur erfahrene User untersucht wurden, was durch das Alter des Accounts, ein hohes Aktivitätslevel und eine hohe Anzahl von Abonnenten definiert wurde, wurden dennoch Falschmeldungen, mit einer Wahrscheinlichkeit von 70%, eher weiter versendet als wahre Nachrichtenmeldungen. Zwar wurden auch Bots bei den untersuchten Benutzern und Benutzerinnen entdeckt, diese haben jedoch sowohl die Verbreitung von wahren als falschen Nachrichtenmeldungen beschleunigt und konnten keine Tendenzen für eine Kateogrie vorweisen. (Vosoughi et al., 2018, S. 1147 - 1149)

> „This suggests that false news spreads farther, faster, deeper, and more broadly than the truth because humans, not robots, are more likely to spread it.” (Vosoughi et al., 2018, S. 1150)

Jedoch gibt es die Annahme, dass die Wirkung von Bots in den letzten Jahren auf Social Media Plattformen immer mehr zunimmt. Facebook gibt an das sich ungefähr 60 Millionen Bots auf der Plattform befinden. Beide Faktoren zeigen das hochgefährliche Potential von Deep Fakes auf. Bots können Deep Fakes in kürzester Zeit weitverbreiten und ein kleines menschliches Publikum erreichen, welche es dann wiederum viel stärker verbreitet und der Fälschung Glaubhaftigkeit gibt. (Chesney & Citron, 2018, S. 1768 - 1769)

4.8 Gefahr für die Gesellschaft

Die Gefahrensituationen, die durch veröffentlichte Deep Fakes entstehen können, enden nicht bei Einzelpersonen, Personengruppen, Firmen oder Organisation, sondern die ganze Gesellschaft kann dadurch bedroht werden. So können damit politische Wahlen gelenkt werden indem man Kandidaten und Kandidatinnen anderer Parteien diffamiert, Regierungen bloßstellt und somit öffentliche Unruhen erzeugt oder gar einen Kriegsbeginn verkündet und somit Panik erzeugt. Deep Fakes können das Vertrauen einer Bevölkerung in den eigenen Staat erschüttern, Rassismus anheizen oder internationale Beziehungen verschlechtern.

Auch hier gilt wieder die Problematik, dass es in solchen Situationen zwar möglich sein wird entsprechende Deep Fakes als Fälschungen zu entlarven, potenzielle Schäden jedoch nicht mehr rückgängig gemacht werden können. Ein demütigendes oder schändliches Video, welches einen politischen Kandidaten oder Kandidatin zeigt, kann am Wahltag veröffentlicht werden und in kurzer Zeit via Social Media Plattformen viral gehen und somit die Wahl in einem nicht nachvollziehbaren Maß beeinflussen. Die notwendige Zeit, um aufklärend zu kommunizieren, dass sich es bei dem öffentlichen Material um eine Fälschung handelt ist in diesem Fall nicht gegeben. Um solch eine Situation noch dramatischer und komplizierter zu gestalten könnten weitere Deep Fakes beziehungsweise synthetische Tonaufnahmen generiert werden, um die Urheberschaft der Erstellung der ursprünglichen Deep Fakes einem anderen politischen Kandidaten oder Kandidatin zu zuweisen. Die Möglichkeiten, politische Wahlen mit Deep Fakes zu beeinflussen sind groß, die Summe der möglichen Täter und Täterinnen ebenso. Diese sind nicht auf das jeweilige Land begrenzt, so könnten Staaten die Wahlen andere Nationen je nach Interesse manipulieren. Dies ist prinzipiell nichts Neues, so wurden schon im amerikanischen Wahlkampf 2016 als auch bei den französischen Wahlen 2017 eine Beteiligung beziehungsweise Manipulation durch Russland untersucht. Die potenzielle Gefahr der gravierenden Manipulation wird durch Deep Fakes jedoch massiv erweitert. (Chesney & Citron, 2018, S. 1177 - 1779)

4.8.1 Der Verlust des Vertrauens in Nachrichten

Die hohe Anzahl an gezielten Falschmeldungen beziehungsweise die hohe Qualität von Falschmeldungen kann weiter dazu führen, dass das allgemeine Vertrauen in Nachrichten so sinkt, dass Nachrichten weniger konsumiert werden oder gar gezielt vermieden werden. Die amerikanische Analysefirma *Gallup* verzeichnete bei ihren jährlichen Umfragen 2016 einen Tiefstwert für Vertrauen in Medien bei der amerikanischen Bevölkerung.

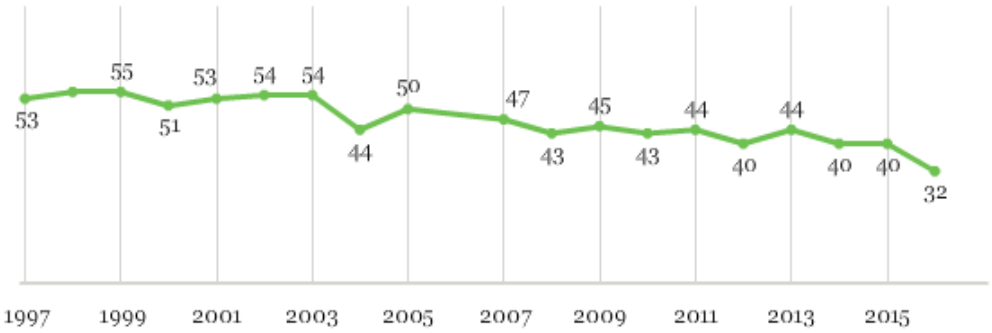

Abbildung 11 – Vertrauen der Amerikaner in Massenmedien zwischen 1997 und 2016 (Switft, 2016)

Nur 32% der befragten Personen gaben an den Medien in einem starken Ausmaß zu vertrauen. Im Allgemeinen sinkt das Vertrauen stetig seit dem Jahre 2007. In der Hälfte der jüngeren Teilnehmer und Teilnehmerinnen gaben nur 26% an den Medien zu vertrauen, bei der Hälfte der älteren Teilnehmer und Teilnehmerinnen sind es 38%. Hier zeigt sich direkt die Auswirkung von der Flut an Falschmeldungen in sozialen Netzwerken, welche von der jüngeren Bevölkerung mehr wahrgenommen wird. Die Wahl zum amerikanischen Präsidenten zwischen dem kontroversen Kandidaten Donald Trump und Hillary Clinton im Jahr 2016 trug sicherlich zu diesem Vertrauensverlust bei, dennoch ist dieses Phänomen des Vertrauensverlustes keinesfalls exklusiv den Vereinigten Staaten von Amerika zu zuordnen. (Switft, 2016)

Die Kommunikationsagentur *Edelman* veröffentlicht jährlich einen ausführlichen Bericht bezogen auf das Vertrauen in die Regierung, die Medien, NGOs und Unternehmen in 28 verschiedenen Ländern. Der Bericht von 2016 zeigt, dass die weltweite Lage sehr ähnlich ist. Von den 28 befragten Ländern vertrauten nur fünf Bevölkerungen ihren Medien mit 50%. In neun Ländern misstrauten sogar zwei Drittel der Bevölkerung den Medien.

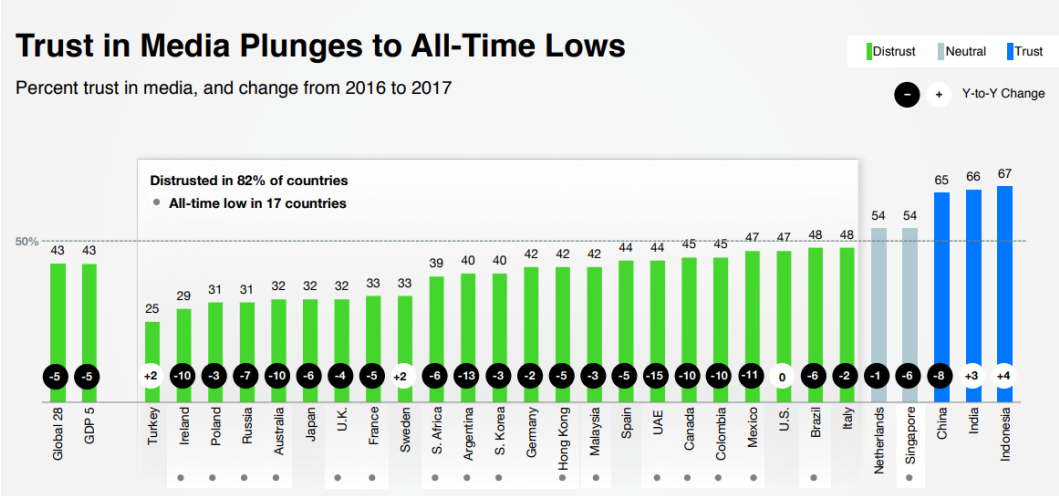

*Abbildung 12 – Das prozentuelle Vertrauen der Bevölkerung in den untersuchten Ländern zu den Medien (*2017 Edelman TRUST BAROMETER*, 2017, S. 12)*

Die Werte bezüglich Vertrauen in die jeweiligen Regierungen zeigen ein sehr ähnliches Bild und unterstreichen somit den Zusammenhang zwischen beiden Institutionen. Dadurch entsteht zu einem gewissen Grad auch eine Abwärtsspirale. Das Vertrauen in klassische Institutionen sinkt, somit sinkt auch der allgemeine Glaube an das System und Ängste entstehen. Diese Sorgen können dann durch Populismus verstärkt werden, wodurch das Vertrauen in klassische Institutionen weiter sinkt. Traditionelle Medien verlieren Zuspruch, Personen wenden sich an Alternativen wie Onlineblogs oder Suchmaschinenergebnisse, wo wiederum Populismus betrieben werden kann. Dies ist eine starke Vereinfachung, die Daten von Edelman zeigen jedoch, dass dies sehr wohl passieren kann. (*2017 Edelman TRUST BAROMETER*, 2017)

Das Jahr 2016 stellte bezüglich des Vertrauens in klassische Institutionen einen Extremfall dar, seither gibt es einen leichten Aufwärtstrend. Im Bericht von 2019 misstrauen nur noch knapp über die Hälfte der befragten Bevölkerung den Medien und ihren jeweiligen Regierungen. Misstrauen ist somit weiterhin die Norm. In den drei Jahren, welche zwischen den beiden Berichten liegen, konnten die traditionellen Medien wieder an Vertrauen gewinnen und liegen mit 61% gleichauf mit Online-Suchmaschinen. Soziale Netzwerke kommen hingegen nur auf 40% Vertrauen. Ein weiterer spannender Zusammenhang wird im Bericht von 2019 publiziert; so konsumierten 44% der Befragten im Jahr 2018 weniger oft als wöchentlich, Nachrichten, 2020 hingegen waren es nur noch 23% die so selten Nachrichten konsumierten. (Stokes, 2020)

Somit liegt die Annahme nahe, dass eine weitere Abwärtsspirale entsteht sobald die Bevölkerung den Medien wenig vertraut. Man ignoriert diese einfach, da man

den Nachrichtengehalt sowieso als irreführend einstuft. Hier könnte eine große Verbreitung von Deep Fakes massive Auswirkungen haben. Diesbezüglich könnte man entgegenstellen das Werkzeuge, welche *Fake News* erkennen, diesen Umstand signifikant entgegensteuern könnten. Jedoch müsste auch das Vertrauen in solche Werkzeuge hoch genug sein, damit Personen diese anwenden oder deren Ergebnissen überhaupt vertrauen. In einem Forschungsbericht von Forscher und Forscherinnen der Universität von Nottingham und der Universität of Lincoln wurden Personen zu solch einem hypothetischen Werkzeug befragt. Die Mehrheit der Befragten äußerte sich kritisch, da man sich eine Implementierung schwer vorstellen könnte. So wurde angemerkt, falls solch ein Werkzeug auf Schwarmintelligenz von vielen Benutzern beruht, dieses ein verzerrtes Bild liefern würde. Weiters wurde angemerkt, dass auch in traditionellen Medien oft Berichte publiziert werden, welche auf anonymen Quellen beruhen und somit keine Möglichkeit besteht solche Berichte zu verifizieren. (Flintham et al., 2018, S. 7)

4.9 Erstellungsprozess von Deep Fakes - Allgemein

Mittlerweile gibt es mehrere Softwarewerkzeuge und Systeme, die sich für die Erstellung von Deep Fakes anbieten. Diese greifen jeweils auf verschiedene neuronale Netzwerke zurück. So verwendete die *FakeApp*, welche als erste verbreitete Deep fake-Software bekannt wurde, eine sogenannte *autoencoder-decoder pairing structure.* Bei dieser Methodik extrahiert der Encoder markante Eigenheiten des Gesichts, der Decoder rekonstruiert daraus wieder das Bild. Das abstrahierte Zwischenbild wird als *Latent face* bezeichnet. Dieses, von außen verborgene Gesicht, ist massiv reduziert in seiner Komplexität und Schärfe. Um Gesichter auszutauschen sind zwei Paare von Encoder und Decoder notwendig. Zusätzlich dazu kommunizieren die beiden Encoder miteinander und teilen sich so ihre Parameter. Dadurch erlernt das System die Ähnlichkeiten der beiden Gesichter. Ist dieser Prozess abgeschlossen wird das Bildmaterial von der ersten Person wiederum mit dem Encoder verknüpft, jedoch darauffolgend mit dem Decoder der zweiten Person durchlaufen. Somit wird die Extraktion des Bildmaterials der ersten Person für die Rekonstruktion der zweiten Person verwendet.

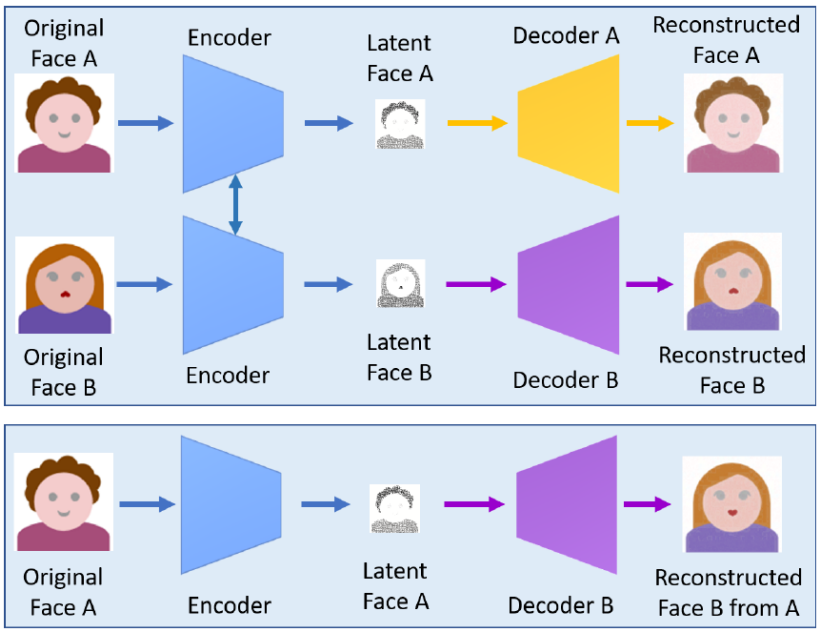

Abbildung 13 – Oberer Bildbereich zeigt Trainingsvorgang eines doppelten Encoders/Decoder-Systems. Unterer Bildbereich zeigt die darauf basierende Deep Fake-Erstellung.

Da menschliche Gesichter in der Regel große Ähnlichkeiten wie Nasenposition, Augengröße und die Position der Lippen aufweisen ist es für den Algorithmus leicht die Gemeinsamkeiten zu finden. Diese Vorgehensweise ist eine grobe Zusammenfassung, jedoch verwenden mehrere Softwareprodukte, unteranderem *DeepFaceLab*, *DFaker* und *DeepFake-tf* diese Herangehensweise.

Dies ist der wohl wichtigste Teilaspekt der Erstellung von Deep Fakes, dennoch handelt es sich hier nur um einen Schritt in eines ganzen Arbeitsablaufs beziehungsweise einer Pipeline. Weitere Teile dieses Ablaufs sind Gesichtserkennungsmodule, Prozesse zur Überblendung von Gesichtern und Gesichtsanalyse. Erst die Kombination dieser Schritte führt zu hochqualitativen Deep Fake Ergebnissen.

4.9.1 Hochwertige Gesichtserkennung sorgt für realistischere Ergebnisse

Gewisse Techniken können verwendet werden, um das Deep Fake Ergebnis zu verbessern. Durch die Implementation des neuronalen Netzwerk *VGGFace,* welches als Gesichtstracker funktioniert, werden Augenbewegungen realistischer umsetzbar. Dieses neuronale Netzwerk, welche von der *Visual Geometry Group* der Universität von Oxford stammt, wurde durch einen großen Datensatz von

Gesichtern trainiert und ist öffentlich verfügbar. Der Datensatz besteht aus 2,6 Millionen Bildern von 2622 verschiedenen Personen, und übersteigt somit Datensatzgrößen, welche für Einzelpersonen umsetzbar sind, bei weitem. (Ghazi & Ekenel, 2016, S. 35)

Die Trainingsdaten, welche Facebook im Jahr 2015 verwendete, beinhalteten 4,4 Millionen Bilder von 4.030 verschiedenen Personen. Zum gleichen Zeitpunkt trainierte Google die eigene Gesichtserkennungsoftware mit 200 Millionen Bildern von acht Millionen Personen. Diese Trainingsdaten sind jedoch für die Öffentlichkeit nicht verwendbar. Dies war für die Forscher aus Oxford auch einer der wichtigsten Motivationsgründe das *VGGFace Modell* zu entwickeln. Im Jahr 2017 wurde *VGGFace 2* vorgestellt, dieses beinhaltet 3,31 Millionen Bilder von 9.131 Personen. Dabei wurde drauf geachtet das zwischen den Bildern eine große Variation herrscht, bezüglich Alter, Geschlecht, Posen und Ethnien. Gemeinsam mit dem Datensatz wurden auch Testergebnisse veröffentlicht. Diese zeigen, dass verschiedene neuronale Netzwerke, welche durch diese Daten trainiert wurden, den momentanen Stand der Technik bezüglich Gesichtserkennung darstellen. Da diese Daten für jeden kostenlos zu Verfügung stehen, können diese auch für Deep Fake Erstellung verwendet werden. (Cao et al., 2018, S. 73)

4.9.2 Softwareprodukte zur Erstellung von Deep Fakes

Wie schon zuvor erwähnt gibt es mittlerweile verschiedenste Softwareprodukte und Frameworks, die dazu kreiert worden sind, um Deep Fakes zu erstellen. Diese Diplomarbeit wird sich auf die Software *DeepFaceLab* fokussieren, da es sich dabei um das am weitesten verbreitete System handelt und somit auch die meiste Literatur sowie Hilfestellungen verfügbar sind.

4.9.2.1 DeepFaceLab

DeepFaceLab wird für Anfänger, die ihre ersten Deep Fakes erstellen wollen, empfohlen. Auf der GitHub-Seite[8] von *DeepFaceLab* wird angegeben das mehr als 95% aller Deep Fakes mit *DeepFaceLab* erstellt werden, wobei nicht angegeben wird wie dieser Prozentsatz erhoben worden ist.

DeepFaceLab basiert auf der Programmiersprache *Python* und funktioniert auf mehreren verschiedenen Betriebssystemen. Das ganze Projekt ist open-source, somit kann jeder den Quellcode einsehen oder auch verändern. Einer der Hauptgründe für die große Verbreitung von *DeepFaceLab* ist die Tatsache, dass

[8] https://github.com/iperov/DeepFaceLab

die komplizierten Deep Learning Prozesse stark abstrahiert werden, dass auch Benutzer ohne dieses hohe technische Wissen, dieses Framework verwenden können. Nur durchschnittliche Computerkenntnisse sind notwendig, um erste Ergebnisse zu bekommen. Dies ist, bezogen auf den notwendigen Wissenstand, in keinster Weise mit anderen hochwertigen Deep Fake Ergebnissen zu vergleichen. So wurde zum Beispiel das bekannte Deep Fake Beispiel welches Barack Obama zeigt, (*Abbildung 10*) durch einem individuellen Programmcode und einer 3D-modillierten Kopie von Obamas Kopf umgesetzt. Zwar resultierte aus diesem Ansatz eine sehr hochwertige und glaubwürdige Bewegtbildfälschung, jedoch kann dieses Prozedere nicht einfach auf eine andere Person beziehungsweise ein anderes Quellvideo angewendet werden. *DeepFaceLab* bietet im Gegensatz dazu, mit seiner breiten und umfassende Herangehensweise, einen einfachen Einstieg in die Welt von Deep Fakes. (Perov et al., 2020, S. 2-3)

DeepFaceLab baut auf dem Maschine Learning System *TensorFlow* auf, welches ebenfalls open-source ist und gratis zu Verfügung gestellt wird. *TensorFlow* ist der Nachfolger von *DistBelief*, welches von Google eingesetzt worden ist, um die eigenen neuronalen Netzwerke zu trainieren. *TensorFlow* ist entwickelt worden, um mit verschiedenen Modellen von künstlicher Intelligenz zu experimentieren, diese mit großen Datensätzen zu trainieren und die Ergebnisse weiter zu verwenden. Eine weitere große Stärke von *TensorFlow* ist, dass das Netzwerk sich je nach verfügbarer Rechenleistung gut adaptieren kann. So können hunderte Grafikkarten in mehreren Serversystemen verwendet werden, jedoch genauso mobile Geräte zur Berechnung herangezogen werden. (Abadi et al., 2016, S. 1)

5 Detaillierter Arbeitsablauf von Deep Fake Erstellung an Hand des Beispiels DeepFaceLab

Man kann den Arbeitsablauf von *DeepFaceLab* grob in drei Prozesse aufteilen. Diese sind der Extraktionsprozess, der Trainingsprozess und die abschließende Umwandlung des originalen Videos. Diese drei Prozesse werden nacheinander abgearbeitet. Es sollte auch angemerkt werden, dass *DeepFaceLab* immer nur auf ein Video angewandt wird, somit immer genau ein Quellvideo verändert wird. Multiple Deep Fakes können somit nicht parallel auf einem Computersystem erstellt werden. Gesteuert wird *DeepFaceLab* primär rein über eine Textkonsole, nur für den Maskierungsprozess und den Umwandlungsprozess bietet es auch ein grafisches Interface. Dieses ist jedoch optional, die Verwendung kann auch rein nur über die Textkonsole vollzogen werden. Jeder Schritt und Unterschritt im Arbeitsablauf ist eine eigene *BAT-Datei* [9]zugeordnet. Zur besseren Übersicht sind diese Scriptdateien chronologisch geordnet. In der von mir verwendeten Version *DeepFaceLab_NVIDIA_build_07_04_2020* sind 47 Scriptdateien vorhanden.

5.1 Extraktionsprozess

Innerhalb des Extraktionsprozesses wird sowohl eine Gesichtserkennung, Gesichtsanordnung und Gesichtssegmentierung vollzogen. Nach diesen drei Schritten erhält man die ausgerichteten Gesichter mit exakten Maskierungen. Weiters sind die freigestellten Gesichter auch unterteilt in die typischen Bestandteile von Gesichtern wie Nase, Augen und Mund.

[9] Dabei handelt es sich um eine Stapelverarbeitungsdatei, welche mehrere Befehlsfolgen beschreibt

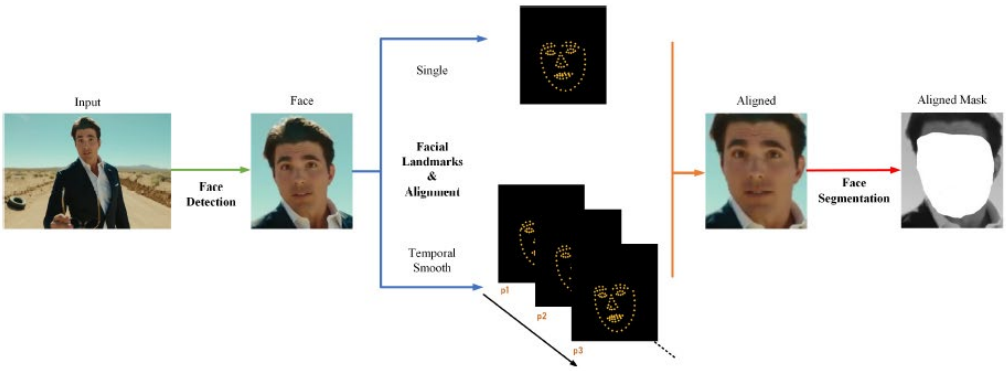

Abbildung 14 – Ablauf des Extraktionsprozess innerhalb von DeepFaceLab

DeepFaceLab ermöglicht auch die Konfiguration des Gesichttyps. Falls der Standardwert *full face* unerwünscht ist, kann dieser auch auf *half face* oder *whole face* umgeändert werden.

5.1.1 Gesichtserkennung

Für die Gesichtserkennung verwendet *DeepFaceLab* den Algorithmus *S3FD*. Dieser liefert vor allem bei klein abgebildeten Gesichtern gute Ergebnisse und eignet sich somit sehr gut für Bewegtbildmaterial, da diese, im Gegensatz zu klassischen Einzelbildern, keine hohe Auflösung bieten. Das Forschungsteam der *University of Chinese Academy of Sciences Beijing* welche die Gesichtserkennung *S3FD* präsentierten kamen zu folgendem Ergebnis:

> “Besides, we propose the scale compensation anchor matching strategy to improve the recall rate of small faces, and the max-out background label to reduce the false positive rate of small faces. The experiments demonstrate that our three contributions lead S3FD to the state-of-the-art performance on all the common face detection benchmarks, especially for small faces.” (Zhang et al., 2017, S. 199)

Jedoch kann man auch andere Gesichtserkennungsalgorithmen für den Deep Fake Prozess innerhalb von *DeepFaceLab* verwenden. (Perov et al., 2020, S. 4)

5.1.2 Gesichtsausrichtung

Um die Gesichter richtig auszurichten ist es notwendig die markanten Bestandteile eines Gesichts zu erkennen. Dafür bietet *DeepFaceLab* zwei verschiedene Algorithmen, für Gesichter die mehrheitlich zur Kamera zugewandt sind wird

2DFAN verwendet, dieser Algorithmus stellt die markanten Gesichtspunkte als verbundene Fixierungspunkte dar. *2DFAN* wurde am *Computer Vision Laboratory* bei der Universität von Nottingham entwickelt. Die involvierten Forscher testeten den Algorithmus an über 100.000 Bildern und erzielten hervorragende Ergebnisse. Vor allem die hohe Belastbarkeit bezogen auf verschiedene Gesichtsposen und Bildauflösungen wurde hervorgehoben. (Bulat & Tzimiropoulos, 2017, S. 7)

Für Gesichter die einen starken eulerschen Winkel[10] vorweisen, zum Beispiel wenn eine Gesichtshälfte die andere verdeckt, kommt *PRNet* zur Anwendung. Hier werden die Gesichtsbestandteile als 3D-Model visualisiert.

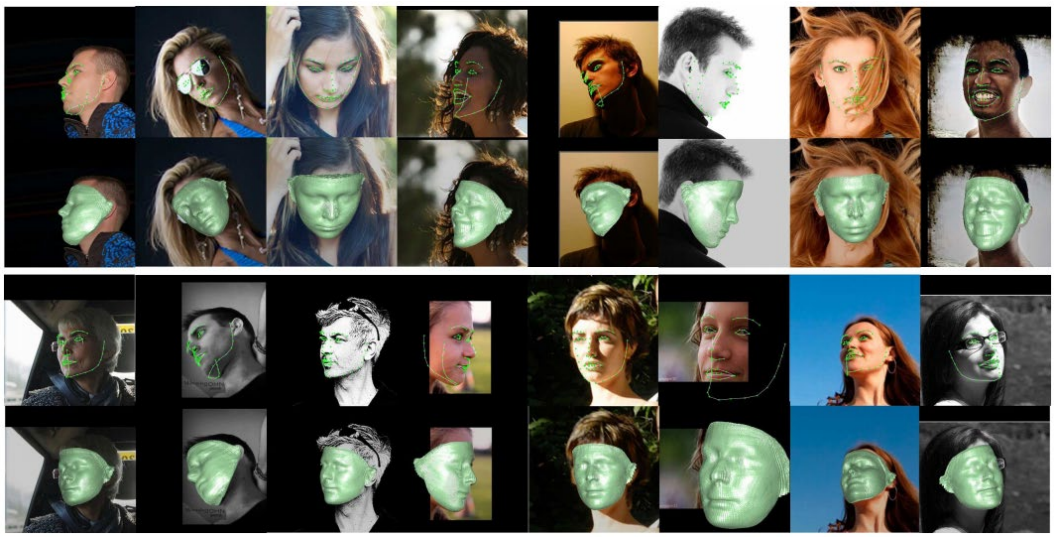

Abbildung 15 – Reihe 1 und 3 zeigen die markanten Gesichtszüge als verbundene Fixierungspunkte ,Reihe 2 und 3 als 3D-Modell unter der Verwendung von PRNet. (Feng et al., 2018, S. 2)

Um die Ausrichtung zu berechnen, bietet *DeepFaceLab* drei Vorlagen, in welchen die markanten Gesichtspunkte richtig ausgerichtet sind. Diese Berechnungsvorlagen sind die Frontansicht und jeweils eine pro Seitenansicht. Der eulersche Winkel wird von *DeepFaceLab* automatisch berechnet, und somit wird auch der jeweilige bessere Algorithmus für das aktuelle Gesicht angewandt. Der Nutzer muss somit nicht eingreifen. (Perov et al., 2020, S. 5)

[10] Drei Winkel über welche die Orientierung eines Objektes im Raum definiert wird

5.1.3 Gesichtssegmentierung

Als letzten Schritt des Extraktionsprozesses passiert die Gesichtssegmentierung. Dieser Prozess ist optional, jedoch sinnvoll, falls innerhalb der Trainingsdaten Bilder vorhanden sind, welche partiell verdeckt werden. Dies kann zum Beispiel durch Brillen, Frisuren oder auch Handbewegungen der Fall sein. Dieser Prozess sorgt jedoch dafür, dass auch die Gesichtszüge in solchen Einzelbildern richtig segmentiert werden. Dafür verwendet *DeepFaceLab* das *XSeg-Modell*. Dafür ist es notwendig, händisch mehrere Bilder zu segmentieren. Das Modell berechnet dann die entsprechende Gesichtssegmentierung für den ganzen Datensatz.

Abbildung 16 – Links: Händische Auswahl für das XSeg-Modell. Rechts: Lernprozess basierend auf händischer Auswahl (Perov et al., 2020, S. 9)

Dieses Modell ist bei der Erstverwendung nicht vortrainiert, somit ist es notwendig diesen Trainingsprozess selbst zu starten, wenn das *XSeg-Modell* angewendet werden soll. Jedoch ist es möglich dieses Modell für zukünftige Projekte erneut zu verwenden beziehungsweise mit neuen Gesichtern zu füttern. Die händische Maskierung zwischen 50 und 100 Bildern ist ausreichend, um ordentliche Ergebnisse zu erzielen. Hierbei muss jedoch darauf geachtet werden, dass diese verschiedenen Masken nicht bei 50 aufeinanderfolgenden Einzelbildern erstellt worden sind, sondern eine gewisse Diversität vorhanden ist. Werden die Masken bei Einzelbildern erstellt, die sich stark unterscheiden, wie beispielsweise verschiedene Gesichtsausdrücke, funktioniert der spätere automatische Maskierungsprozess des *XSeg-Modells* besser. (Perov et al., 2020, S. 5)

5.1.4 Aussortieren von nicht hochwertigen Trainingsdaten

Bevor nun in den Trainingsprozess übergegangen wird sollten die Trainingsdaten überprüft werden. So kann es sehr wohl der Fall sein, dass die vorherigen Algorithmen alle Gesichter der Einzelbilder richtig erkannt und ausgerichtet haben, jedoch das Bild dennoch ungeeignet für den Trainingsvorgang ist. So kann das Bild eine hohe Bewegungsunschärfe aufweisen, das Gesicht nicht richtig fokussiert sein oder durch eine externe Lichtquelle völlig überbelichtet sein. Bei mehreren tausend Einzelbildern ist die manuelle Kontrolle der Gesichter mühsam, demensprechend bietet hier *DeepFaceLab* verschiedene Sortierungsalgorithmen an.

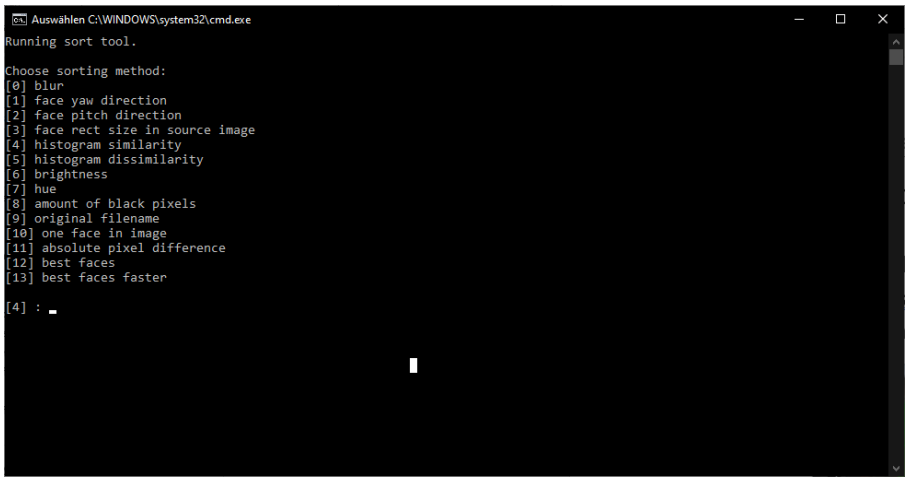

Abbildung 17 – Sortierungsalgorithmen von DeepFaceLab für die Aussortierung von Trainingsdaten

So können die Gesichter unter Anderem nach dem Grad der Unschärfe, dem Ausrichtungswinkel, der Helligkeit, der Ähnlichkeit des Histogramms, der Summe von Schwarzen Pixel oder nach Farbton sortiert werden. Auch können bei einem sehr großen Datensatz über die Sortierung *Best Faces* eine gewisse Anzahl der besten Gesichter extrahiert werden und exklusiv für den Trainingsprozess verwendet werden.

Wählt man zum Beispiel die Sortierungsmöglichkeit Unschärfe werden die unschärfsten Gesichter über den Dateinamen an den Anfang der Dateiliste verschoben. Diese können dann manuell betrachtet werden. Ist die Unschärfe signifikant können diese Bilder gelöscht werden, was einen höherwertigen Trainingsprozess garantiert. Die Aussortierung schlechter Trainingsdaten ist ein essenzieller Schritt und muss mit großer Aufmerksamkeit absolviert werden.

5.2 Trainingsprozess

Nach dem Abschluss des Extraktionsprozess sind alle notwendigen Daten vorhanden, um den Trainingsprozess zu starten. Die maskierten Gesichter, inklusive der Koordinaten im ursprünglichen Einzelbild, die markanten Gesichtsbestandteile dieser Gesichter, sowie deren dahinterliegende Ursprungsbilder. Diese Daten liegen sowohl für das Quellvideo als auch das Zielvideo vor.

DeepFaceLab bietet für den Trainingsprozess zwei verschiedene Hauptmodelle, *DF* und *LIAE*. Diese gibt es jeweils in einer normalen, einer HD-Version und einer UHD-Version. Der Aufbau des DF-Modells ist einfacher gestaltet als der des LIAE-Modells. Hierbei teilen sich Quelle und Ziel den gleichen Encoder, dieser reduziert die Information zu einem Vektor, welcher dann an den sogenannten *Inter* weitergegeben wird. Bei diesem Schritt wird der Vektor in ein abstrahiertes Bild umgeformt, dem, in *Abbildung 13 schon* gezeigten, *Latent face*. Aus diesem stark reduziertem Bild versuchen zwei getrennte Decoder wieder das originale Bild zu rekonstruieren.

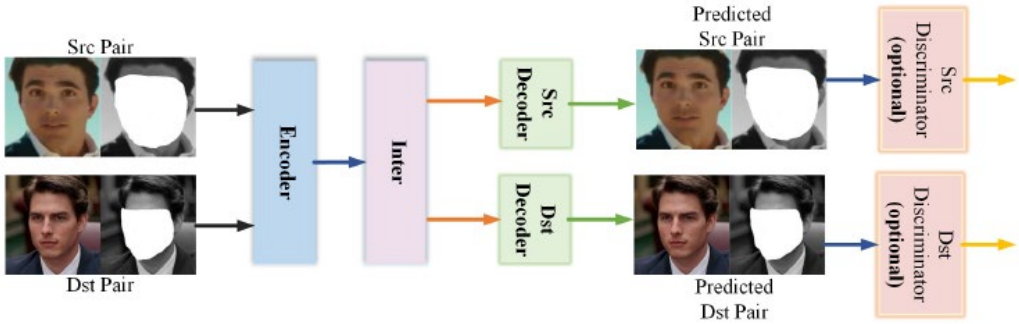

Abbildung 18 – Vereinfachte Abbildung des Aufbaus des DF-Trainingsmodells

Um den Trainingsprozess zu verbessern kann optional noch ein Diskriminator verwendet werden. Dieser wird dann angewendet, wenn man die Vorteile eines Generative Adversarial Networks nutzen will. Hierbei überprüft der Diskriminator ob die Daten, welche er erhält, reale Daten sind oder nicht. So entsteht ein Nullsummenspiel, der Encoder bzw. Decoder erstellt immer hochwertigere Daten, der Diskriminator probiert diese immer wieder zu entlarven. Mit diesem optionalen Prozess kann man die Kluft zwischen originalen Quelldaten und dem synthetisch erzeugten Bild verkleinern.

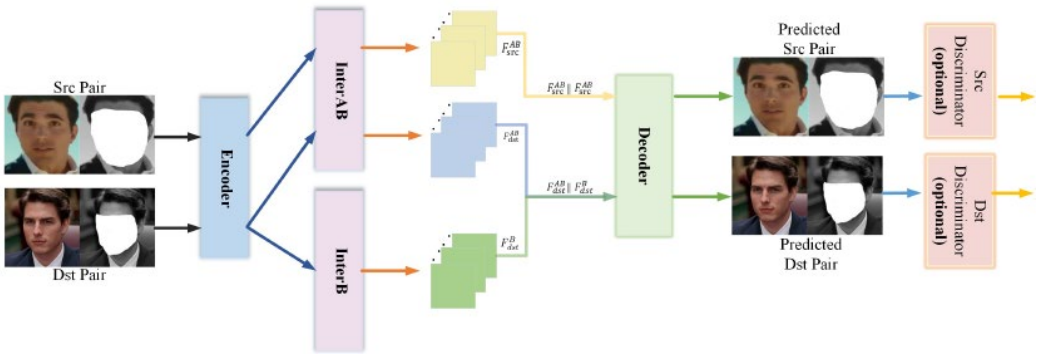

Abbildung 19 – Vereinfachte Abbildung des Aufbaus des LIAE-Trainingsmodells

Die Struktur des LIEA-Modell ist im Verhältnis zum DF-Modell komplexer, wie ein Vergleich von *Abbildung 18* und *Abbildung 19* verdeutlicht. Dadurch ergibt sich zum Beispiel der Vorteil, dass für gute Ergebnisse die Kopfform und Gesichtsform nicht so ähnlich sein muss wie sie für das DF-Modell sein sollten. Auch können seitliche Positionen besser gelöst werden, Frontalansichten können jedoch mit diesem Modell weniger hochwertig sein als mit dem DF-Modell. Das LIAE-Modell verwendet zwei unabhängige Intermodelle. So wird ein eignes *Latent face* für das Zielmaterial erstellt, von *InterAB* jedoch werden zwei zusammengerechnete *Latent Face* erstellt, einmal für die Quelle und einmal für das Zielmaterial. Diese beiden Daten werden gemeinsam in den Decoder geführt. Der Sinn hinter dieser Vorgangsweise ist, dass das *Latent face* schon mehr in die Richtung des gewünschten Ergebnisses erstellt wird.

```
Starting. Target iteration: 150000. Press "Enter" to stop training and save model.

Trying to do the first iteration. If an error occurs, reduce the model parameters.

[19:06:18][#000002][0288ms][6.1595][5.8403]
[19:21:15][#003005][0278ms][1.4144][1.3362]
[19:36:15][#006034][0289ms][1.0381][0.8772]
[19:51:15][#009050][0294ms][0.9376][0.7391]
[20:06:16][#012072][0279ms][0.8737][0.6511]
[20:21:15][#015095][0284ms][0.8272][0.5882]
[20:36:15][#018124][0303ms][0.7916][0.5417]
[20:51:15][#021147][0287ms][0.7626][0.5021]
[21:06:16][#024165][0280ms][0.7398][0.4704]
[21:21:15][#027184][0555ms][0.7202][0.4434]
[21:36:15][#030193][0297ms][0.6968][0.4218]
[21:51:15][#033187][0278ms][0.6844][0.4018]
[22:06:16][#036155][0317ms][0.6664][0.3847]
[22:21:15][#039014][0302ms][0.6535][0.3705]
[22:28:10][#040324][0304ms][0.5863][0.3064]
```

Abbildung 20 – Die angezeigten Verlustwerte während des Trainingsprozesses von DeepFaceLab

Während des Trainingsprozesses bietet *DeepFaceLab* zwei Indikatoren wie weit der Trainingsprozess fortgeschritten ist, beziehungsweise wie gut trainiert die künstliche Intelligenz schon ist. So werden in der Textkonsole getrennte Verlustwerte für beide Gesichter angezeigt. Diese sollten im Laufe des Trainingsprozess niedriger werden, wobei mit fortlaufender Trainingszeit die Verringerung der Verlustwerte immer langsamer wird. Steigen diese Verlustwerte langsam an, oder springen in kürzester Zeit auf sehr hohe Werte ist dies ein eindeutiger Indikator, dass das Trainingsmodel zusammengebrochen ist und der Trainingsprozess von neuem gestartet werden muss. Der dafür naheliegendste Grund sind minderwertige Trainingsdaten.

Weiters bietet *DeepFaceLab* in diesem Schritt auch eine visuelle Vorschau, welche in fünf verschiedene Spalten unterteilt ist. Die erste Spalte zeigt ein Ursprungsbild des Quellmaterials, die zweite Spalte zeigt das davon berechnete Inter-Bild. Die dritte Spalte zeigt ein Ursprungsbild des Zielmaterials, die vierte Spalte zeigt wiederum das davon berechnete Inter-Bild.

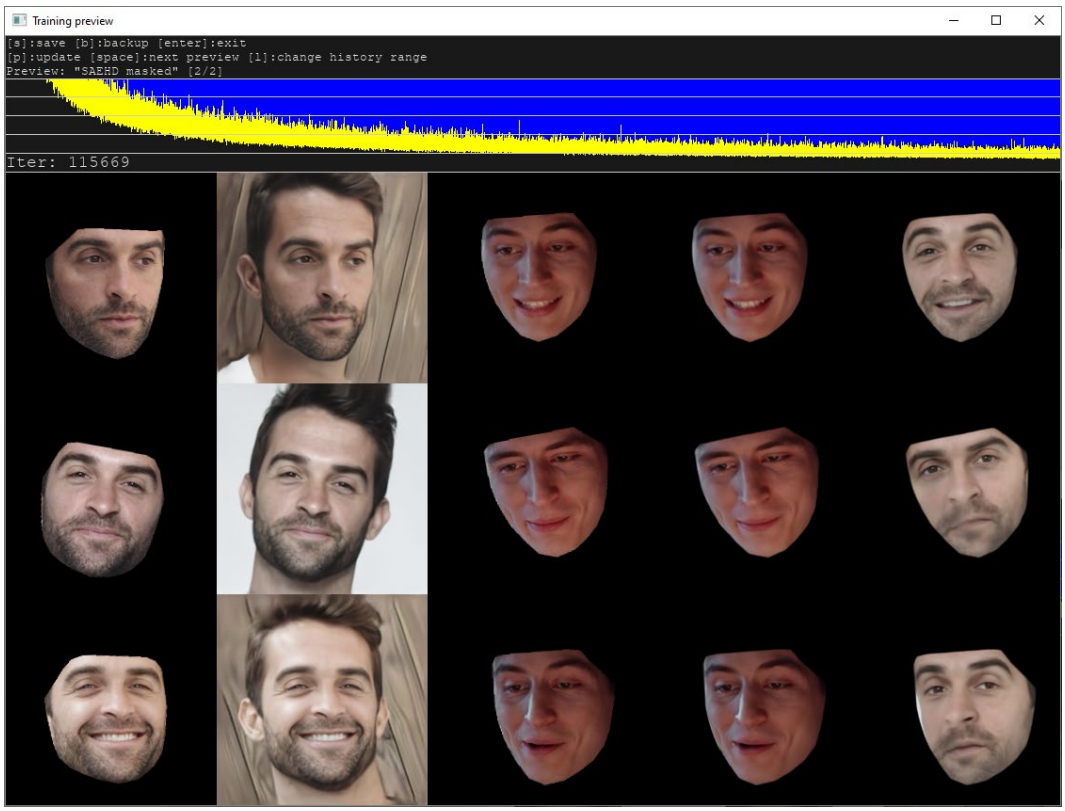

Abbildung 21 – Vorschaufenster von DeepFaceLab für die Visualisierung des Trainingsprozesses

Die fünfte Spalte ist die relevanteste, da es ein Vorschaubild des Deep Fakes ist. Somit kann optisch überprüft werden wie weit der Trainingsprozess vorangeschritten ist, indem die berechneten Inter-Bilder mit den ursprünglichen Bildern verglichen werden. Im optimalen Fall sind diese nicht mehr unterscheidbar. Jedoch bedeuten hochwertige Inter-Bilder nicht automatisch einen perfekten Deep Fake da sich eventuell die Gesichtsformen so stark unterscheiden, dass die erzeugten Bilder nicht hochwertig wirken. In diesem Fall kann das schon trainierte Modell einfach in Kombination mit neuem Zielmaterial verwendet werden. (Perov et al., 2020, S. 6-7)

5.2.1 Trainingsparameter

Die hier aufgelisteten Optionen und Parameter sind alle Optionen, welche am Beginn eines Trainingsprozess definiert werden können. Die verwendeten englischsprachigen Bezeichnungen und Erklärungen in der folgenden Auflistung entstammen direkt aus dem Quellcode der verwendeten *DeepFaceLab_NVIDIA_build_07_04_2020* Version von *DeepFaceLab*. Konkret sind diese aus den Phython-Dateien *ModelBase.py* und *Model.py* entnommen. Der erste Teil des Textes steht für den Namen der Option, der zweite Teil zeigt den optionalen Hilfetextes, welchen man sich anzeigen lassen kann.

5.2.1.1 Auflösung

> Resolution: More resolution requires more VRAM and time to train.

Der hier angegebene Wert definiert in welcher Auflösung die neuen Gesichter berechnet werden. Der Wert 192 definiert zum Beispiel, dass das Gesicht mit einer Auflösung von 192 x 192 Pixel berechnet wird. Umso höher man diesen Wert definiert, umso mehr Grafikkartenspeicher wird benötigt.

5.2.1.2 Gesichtstyp

> Face type: Half / mid face / full face / whole face / head. Half face has better resolution, but covers less area of cheeks. Mid face is 30% wider than half face. 'Whole face' covers full area of face [sic] include forehead. 'head' covers full head, but requires XSeg for src and dst faceset

Über diese Option wird angegeben welcher Anteil des Gesichts innerhalb des Deep Fake Prozesses ausgetauscht werden soll. Die Option *half face* betrifft nur den inneren Bereich des Gesichts, hier wird weder der Stirnbereich noch der

Kinnbereich getauscht. Bei der Option *full face* wird sowohl die Stirn als auch das Kinn berechnet. Die Option *mid face* liegt zwischen den beiden vorherigen Optionen. Die Option *head* betrifft den ganzen Kopf, somit auch Frisur und eventuell vorhandene Gesichtsbehaarung. Hier gilt anzumerken, dass, je größer der Bereich definiert wird, die Auflösung auch umso höher gewählt werden sollte.

5.2.1.3 Batchgröße

> Batch_size: Larger batch size is better for NN's generalization, but it can cause Out of Memory error. Tune this value for your videocard manually.

Die hier angegebene Größe gibt an, wie viele Einzelbilder gleichzeitig durch das neuronale Netzwerk geführt werden. Die Erhöhung dieser Variable erhöht auch den benötigten Grafikkartenspeicher immens. Gleichzeitig bedeutet jedoch eine höhere Batchgröße, dass das Modell hochwertiger wird, da mehr Information gleichzeitig verglichen werden.

5.2.1.4 GPU Berechnung

> Place models and optimizer on GPU: When you train on one GPU, by default model and optimizer weights are placed on GPU to accelerate the process. You can place they [sic] on CPU to free up extra VRAM, thus set bigger dimensions.

Über diese Funktion wird gesteuert ob der Hauptteil des Trainingsprozess über die Grafikkarte absolviert wird. Verneint man diese Option läuft ein Teil Trainingsprozess über den Prozessor, dementsprechend dauert jede Iteration um ein Vielfaches länger. Gleichzeitig wird jedoch auch Grafikkartenspeicher frei, deswegen könnte man im Austausch beispielsweise die Batchgröße erhöhen.

5.2.1.5 Learning rate Dropout

> Use learning rate dropout: When the face is trained enough, you can enable this option to get extra sharpness and reduce subpixel shake for less amount of iterations.

Über diese Option kann der Trainingsprozess reguliert und verbessert werden, das neuronale Netzwerk wird gezwungen leicht aus dem gelernten Muster auszubrechen. Eine detaillierte Beschreibung dieser Option folgt im Kapitel *Anpassung der Trainingsparameter im Laufe des Trainings*.

5.2.1.6 GAN Aufbau

> GAN power: Train the network in Generative Adversarial manner. Forces the neural network to learn small details of the face. Enable it only when the face is trained enough and don't disable. Typical value is 0.1

Diese Option steuert, ob das Modell unter einem GAN-Aufbau trainiert wird. Die Stärke dieser Implementierung kann zwischen einen Wert von 0 und 10 definiert werden. Werte über 1.0 sind jedoch unter Vorsicht zu verwenden und es ist anzuraten Backups zu erstellen, bevor man diese Option aktiviert. Genauere Details zu dieser Option sind im Kapitel *Anpassung der Trainingsparameter im Laufe des Trainings* beschrieben.

5.2.1.7 Augenpriorität

> Eyes priority: Helps to fix eye problems during training like "alien eyes" and wrong eyes direction (especially on HD architectures) by forcing the neural network to train eyes with higher priority.

Aktiviert man diese Funktion, fokussiert sich die künstliche Intelligenz verstärkt auf die Augen. Dies kann vor allem für die richtige Blickrichtung der Augen sehr sinnvoll sein.

5.2.1.8 Farbtransformation

> Color transfer for src faceset: Change color distribution of src samples close to dst samples.

Über diese Option kann gesteuert werden ob, und welche Algorithmen zur Farbtransformation verwendet werden sollen. Dafür gibt es folgende Optionen:

- Reinhard color transfer (RCT)
- Linear color transfer (LCT)
- Monge-Kantorovitch linear (MKL)
- Iterative Distribution Transfer (IDT)
- sliced optimal transfer (SOT)

Hier gibt es keine Option, welche stets die besten Ergebnisse liefert, da dies stark vom verwendeten Material abhängig ist. Für diese Diplomarbeit war es nicht von Relevanz die Funktionsweise dieser Algorithmen zu analysieren. Es wurde immer der Algorithmus gewählt welcher subjektiv die besten Vorschauergebnisse lieferte. (tutsmybarreh, 2020)

5.3 Umwandlung

Der letzte Schritt ist der Umwandlungsprozess. Das Ziel ist es, dass das generierte Gesicht nahtlos von der äußeren Kontur der entsprechenden Maske mit dem restlichen originalen Bild weiterverläuft. Dabei wird das generierte Gesicht, gemeinsam mit der entsprechenden Maskierung, im ersten Schritt an die Ursprungsposition des originalen Gesichts platziert. Dafür ist eine Form der Überblendung notwendig, um überzeugende Ergebnisse zu erhalten. *DeepFaceLab* bietet dafür eine Vielzahl von Möglichkeiten.

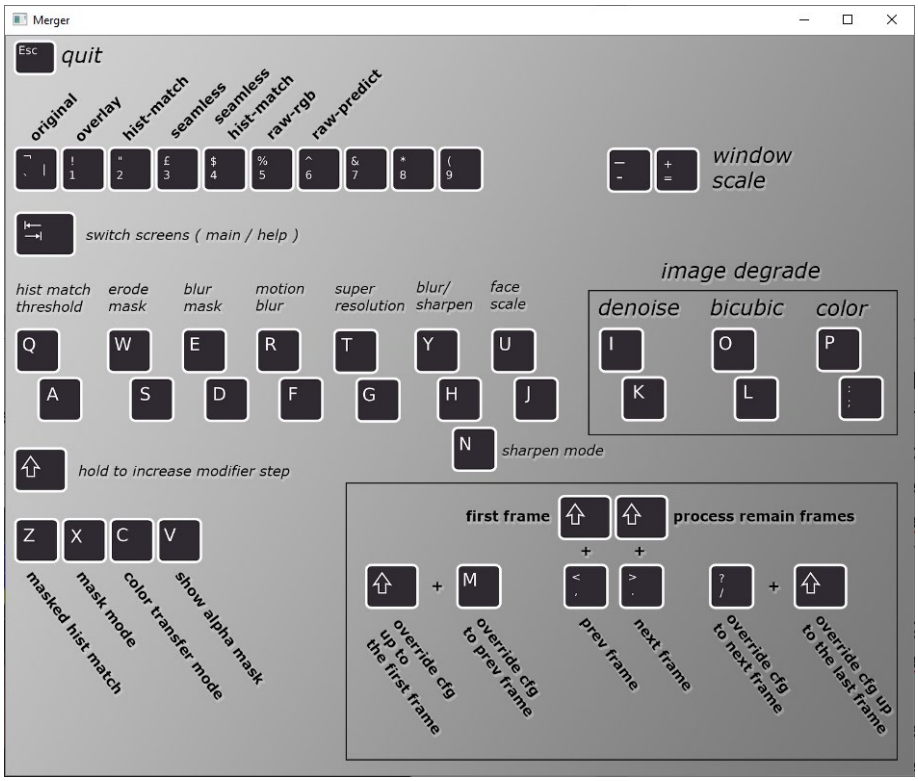

Abbildung 22 – Mögliche Einstellungsoptionen für den Umwandlungsprozess innerhalb von DeepFaceLab

Eine optionale grafische Oberfläche zeigt alle Änderungen direkt in Echtzeit in Form eines Standbildes an. So können die Kanten der Maske des Gesichts weichgezeichnet werden. Das Gesicht kann vergrößert oder verkleinert werden. Bezüglich der Farbtransformation bietet *DeepFaceLab* mehrere Algorithmen, so können alle durchprobiert werden und die Option mit dem besten Ergebnis gewählt werden. Auch kann das Zielvideo verschlechtert werden, um ein nahtloseres Ergebnis zu erhalten. Über die grafische Oberfläche kann auch durch das Video navigiert werden, um sich einen besseren Eindruck vom Ergebnis zu holen. Beim

Überblendungsprozess kann Gesicht noch geschärft werden. Dafür verwendet *DeepFaceLab* ein vortrainiertes neuronales Netzwerk namens *FaceEnhancer*. Der Schärfungsprozess ist sinnvoll, da beim vorhergehenden Überblendungsprozess immer eine Form der Weichzeichnung passiert.

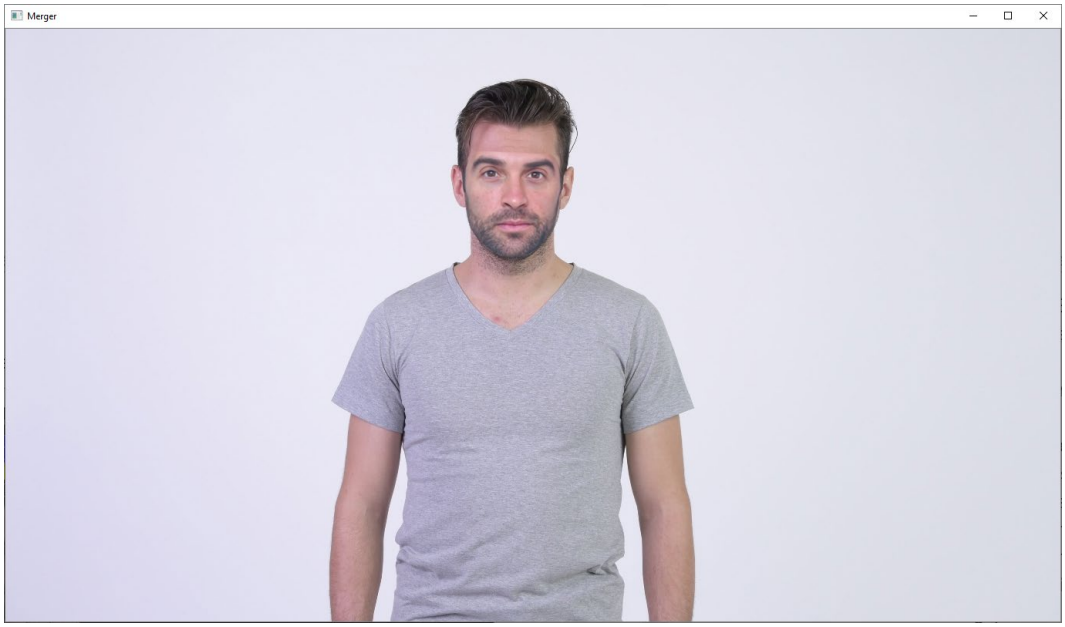

Abbildung 23 – Vorschaufenster während des Umwandlungsprozesses von DeepFaceLab.

Sieht das Vorschaubild vielsprechend aus, kann man den Umwandlungsprozess abschließen, dann wendet *DeepFaceLab* die gewählten Einstellungen auf alle Gesichter an. Hier wird noch keine finale Videodatei erzeugt, dafür gibt es noch eine weitere Skriptdatei. Dort kann zwischen den Containerformaten *MOV* und MP4 gewählt werden, sowieso als Codec ein verlustbehafteter *h.264-Codec* oder ein verlustfreier Codec verwendet werden. Der Umwandlungsprozess ist nicht nur einmalig anwendbar, möchte man verschiedene Optionen ausprobieren, so kann dieser Prozess immer wieder durchgeführt werden und verschiedene Videodateien erzeugt werden. So können für Zielvideos, welche verschiedene Belichtungssituationen vorweisen, mehrere Versionen erstellt werden und diese später mit einen Schnittprogramm zusammengeschnitten werden. Eine weitere Möglichkeit die Überblendung zu verbessern ist, eine manuelle Anpassung in einem Compositing-Programm wie *Nuke* oder *After Effects durchzuführen*, da *DeepFaceLab* hier viel weniger Möglichkeiten bietet. Hierfür exportiert *DeepFaceLab* immer auch eine eigene Videodatei, welche den Alphakanal beinhaltet. (Perov et al., 2020, S. 7 - 8)

6 Praxisteil – Wie gut können Menschen Bewegtbildfälschungen erkennen, die mit Deep Learning Technologie erzeugt worden sind?

Die meisten Beispiele von Deep Fakes, die große mediale Aufmerksamkeit erreicht haben, wurden von Forschungsteams erzeugt und verwendeten zum Teil eigens dafür entwickelte Werkzeuge oder Modelle, um eine möglichst hohe Qualität zu garantieren. Nun gilt es jedoch zu untersuchen, ob auch Einzelpersonen, Bewegtbildfälschungen mit den momentan verfügbaren Werkzeugen erzeugen können, die von einem signifikanten Teil einer Personengruppe nicht als solche erkannt werden können. Der Großteil, der aktuell im Internet kursierenden Deep Fakes sind pornografischer Natur und somit durch die Verwendung von Gesichtern berühmter Schauspielerinnen leicht als Fälschung erkennbar. Andere Deep Fakes verwenden berühmte Filmszenen als Zielvideo und sind damit direkt als Fälschung zu erkennen, unabhängig davon wie hochqualitativ die Fälschung ist.

Die Annahme ist, dass keine signifikante Mehrheit der befragten Personen die erstellten Deep Fakes erkennen wird. Falls sich diese Hypothese bestätigt, wäre die davon abgeleitete gefährliche Implikation, dass mittlerweile jede Person Fälschungen erzeugen kann, die nicht als solche erkennbar sind und weiters damit andere Personen beeinflussen, belügen oder auch erpressen kann.

Für die Erstellung der Deep Fake Beispiele kommt *DeepFaceLab* mit der Version *DeepFaceLab_NVIDIA_build_07_04_2020* zum Einsatz.

6.1 Genauer Ablauf der Untersuchung

Um die Qualität der Ergebnisse zu erhöhen, muss das Bildmaterial so unbekannt wie möglich sein, gleichzeitig jedoch eine hohe Qualität liefern, um beim Deep

Fake Prozess zu einem guten Ergebnis zu kommen. Bekanntes Filmmaterial kann dementsprechend nicht verwendet werden, da die Probanden und Probandinnen die Filmszene entweder bereits kennen oder sie durch die allgemeine Ästhetik der Filmreihe oder Serie erkennen. Bewegtbildmaterial von Personen des öffentlichen Lebens eignet sich für diese Untersuchung ebenfalls nicht. Zwar sind hier einzelne Videosequenzen für sich nicht sehr bekannt, jedoch kann auch hier die Umgebung oder besondere Kleidung ein eindeutiger Hinweis auf die Ursprungsperson sein.

Die eigene Produktion von Videobeispielen bietet nicht die notwendige Diversität beziehungsweise Informationsgehalt, um ein konkretes Forschungsergebnis zu erhalten. Weiters soll diese Untersuchung zeigen, ob Personen fremdes Material mit den entsprechenden Werkzeugen zu qualitativen Bewegtbildfälschungen verändern können, ohne selbst Aufnahmen zu generieren.

Für diese Untersuchung werden deswegen sieben verschiedene Videobeispiele einer Stockplattform gewählt, wobei jedes Video eine andere Person in einem anderen Umfeld zeigt. Bei der großen Menge an generischen Stockvideos ist die Wahrscheinlichkeit, dass teilnehmende Personen mit dem originalen Stockvideo vertraut sind, sehr gering. Zum anderen bieten diese Videos eine hohe Qualität, um ein gutes Fälschungsergebnis zu ermöglichen. Als Herkunftsplattform für diese Beispiele wurde https://elements.envato.com/ gewählt.

Diese Untersuchung möchte eruieren, wie hoch die Wahrscheinlichkeit ist, dass Menschen Bewegtbildfälschungen erkennen, wenn sie in ihren alltäglichen Leben damit konfrontiert werden. Dementsprechend sollen sich die Probanden und Probandinnen beim Test in ihrer natürlichen Umgebung befinden und ihre vertrauten Geräte zur Betrachtung verwenden. Eventuell würden die teilnehmenden Personen in einer kontrollierten Testumgebung mit großen, hochauflösenden Bildschirmen und unkomprimierten Videos die Bewegtbildfälschungen besser von den anderen Videos unterscheiden können. Dies bedeutet jedoch nicht, dass sie auch mit ihrem persönlichen Gerät zum selben Ergebnis gekommen wären. Da Videos, welche in sozialen Netzwerken geteilt werden, stark komprimiert werden, ist anzunehmen, dass die Bewegtbildfälschungen dadurch schwerer erkennbar sind. Dementsprechend werden alle in der Untersuchung gezeigten Videos zuvor mit einer Bitrate von fünf Megabit pro Sekunde mit einem *h.264-Codec* komprimiert.

Beim Test wird den Probanden oder Probandin erklärt was die folgende Aufgabenstellung ist und wie die Bewertung der Videos erfolgt. Den Testpersonen wird nicht explizit erklärt was Deep Fakes sind und das sich die Manipulation nur auf das Gesicht beschränkt. Die Probanden und Probandinnen sollen so wenig

wie möglich über die technologischen Spezifikationen informiert sein und genauso viel Urteilsvermögen anwenden wie bei anderen Social Media Videos. Die Testpersonen sollen zuerst persönliche Daten wie Geschlecht, Alter und höchsten Bildungsabschluss eingeben. Durch diese Daten können später spezifische Statistiken erstellt werden, die Aufschluss darüber geben, ob gewisse Personengruppen die Fälschungen signifikant besser oder schlechter erkennen als andere Personengruppen. Weiters werden die Teilnehmer und Teilnehmerinnen gefragt, wie viele Stunden sie pro Woche auf Social Media Plattformen verbringen. Auch wird mit einer Frage eruiert, ob man beruflich oder durch die jeweilige Ausbildung dem Thema Bewegtbild signifikant nahesteht. Diese Frage kann mit *Ja*, *Etwas* oder *Nein* beantwortet werden.

Pro gezeigtem Video wird der Proband beziehungsweise die Probandin befragt, ob es sich beim gezeigten Material um eine Fälschung handelt. Diese Frage kann nur mit den Antwortmöglichkeiten *Ja* und *Nein* beantwortet werden. Im Einleitungstext zur Umfrage wird explizit erläutert, dass der Einsatz von klassischen Bearbeitungsmethoden wie Schnitt, Weichzeichnung der Haut oder Farbkorrektur nicht ausreicht, um in diesem Zusammenhang eine Bewegtbildfälschung zu definieren. Diesem Einleitungstext muss der Proband beziehungsweise Probandin via Knopfdruck zustimmen.

Weiters gibt die Testperson pro Frage auch an, mit welcher Überzeugung die Antwort abgegeben wurde. Hier gibt es die Optionen *sehr sicher*, *halbwegs sicher*, *unsicher* und *sehr unsicher*. So kann eruiert werden welche Personen die Antwort auf die erste Frage getippt haben. Auf alle Fragen bezogen, kann so ermittelt werden, wie unsicher beziehungsweise sicher die allgemeine Einschätzung bezogen auf die Antworten war. Zusätzlich gibt es pro Videobeispiel auch ein eigenes Textfeld, welches die Probanden und Probandinnen nutzen können, um ihre Entscheidung textuell zu begründen. Dieser letzte Schritt ist jedoch nicht verpflichtend.

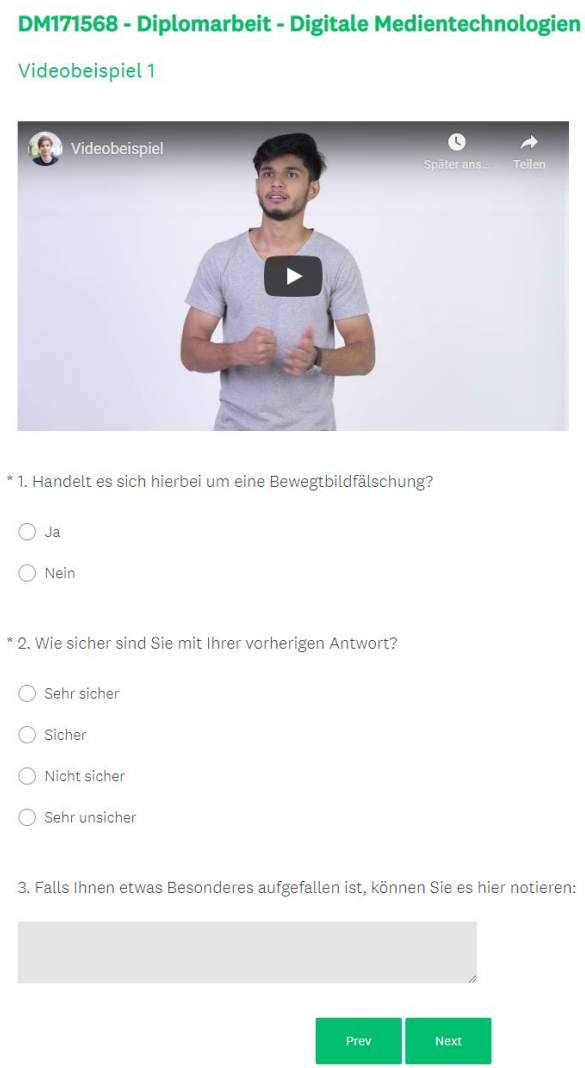

Abbildung 24 – Darstellung eines Videobeispiel inklusive dazugehöriger Fragen innerhalb der Untersuchung

Die Befragung der Probanden und Probandinnen beginnt mit folgendem Einleitungstext:

> Diese Befragung findet im Zuge einer Diplomarbeit des Studiengangs “Digitale Medientechnologien” an der Fachhochschule St. Pölten statt. Das Ziel dieser Untersuchung ist es, zu ermitteln, wie gut Personen Bewegtbildfälschungen erkennen können. Ihnen werden innerhalb dieses Tests 10 verschiedene Videobeispiele in Folge gezeigt. Pro Videobeispiel sollen Sie dabei angeben, ob es sich bei dem von Ihnen betrachteten Material um eine Bewegtbildfälschung oder um unverfälschtes Material handelt. Die Zuordnung der Videos erfolgt

zufällig, somit ist es möglich, dass unter den Ihnen gezeigten Beispielen keine Bewegtbildfälschungen vorhanden sind.

Im Zuge dieser Arbeit wird eine Bewegtbildfälschung folgend definiert: Eine Bildmanipulation die über Methoden wie Farbkorrektur, Weichzeichnung von Haut oder ähnliches weit hinausgeht und die Darstellung gravierend irreführend verändert. Gleichzeitig sollen Sie pro Videobeispiel beantworten mit welcher Überzeugung Sie Ihre Entscheidung getroffen haben. Dabei können Sie zwischen "sehr unsicher", "unsicher", "sicher" und "sehr sicher" unterscheiden. Weiters finden Sie pro Videobeispiel ein Textfeld vor, dass Ihnen ermöglicht eine Begründung für Ihre Entscheidung abzugeben. Dies ist geschieht jedoch auf freiwilliger Basis.

Für die empirische Auswertung dieser Befragung werden Sie zuerst gebeten, einige demographische Fragen zu beantworten. Die dabei erhobenen Daten, werden außerhalb dieser Diplomarbeit nicht weitergegeben.

Die Probanden und Probandinnen erfahren nach der Befragung nicht, ob ihre zehn Antworten richtig oder falsch sind. Damit soll verhindert werden das Testpersonen andere Testpersonen beeinflussen könnten. Jedoch können die Probanden und Probandinnen auf der letzten Seite der Befragung ihre E-Mailadresse hinterlassen, falls sie nach der fertigen Auswertung die Ergebnisse erfahren möchten.

Die Teilnehmeranzahl soll mindestens 100 Personen umfassen, wobei darauf geachtet wird, dass die Diversität dieser Personengruppe hoch ist. Dies bezieht sich auf die Eigenschaften Alter, höchster Bildungsabschluss und Geschlecht.

7 Erstellungsprozess der Deep Fake Beispiele

Die folgenden Unterkapitel umfassen die Erstellungsprozesse der Deep Fake Beispiele zusammen und erläutern die Feststellungen, welche während diesem Prozess aufgekommen sind.

7.1 Verwendete Hardware zur Erstellung der Deep Fakes

Zur Erstellung der Deep Fakes Beispiele werden zwei Computersysteme verwendet. Das erste System beinhaltet einen Intel *i7-7700K* Prozessor in Kombination mit einer Nvidia *Geforce RTX 2070 Super* Grafikkarte. Das zweite Computersystem verfügt über einen AMD *Ryzen 9 3900X* Prozessor und einer Nvidia *Geforce RTX 2070* Grafikkarte. Bei beiden Geräten handelt es sich bei der erwähnten Hardware nicht um Komponenten aus dem Profisegment, die für Einzelpersonen nicht leistbar sind, sondern um höherpreisige Komponenten aus dem klassischen Konsumentenbereich. Die Auswahl der Hardware soll weiter unterstreichen, dass die Erstellung von Deep Fakes eben nicht Industrieprofis beziehungsweise Studios aus der Bewegtbildbranche vorbehalten sind. Die *Nvidia Geforce RTX 2070 Super* und Nvidia *Geforce RTX 2070* Grafikkarten bieten jeweils mit den verbauten acht Gigabyte an GDDR6-Videospeicher eine ordentliche Basis für die Verwendung mit *DeepFaceLab*. Die Größe des Videospeichers ist die Variable, die im Erstellungsprozess am ehesten limitiert, da die Größe bestimmt wie viele extrahierte Gesichter die künstliche Intelligenz gleichzeitig miteinander vergleichen kann. Ebenfalls bestimmt sie somit wie hoch die mögliche Auflösung des synthetisierten Gesichts ist. Damit gilt auch, dass wenn die Bildauflösung gleich gewählt wird, eine Grafikkarte mit größerem Videospeicher, das gleiche Ergebnis in einer viel kürzeren Zeit kreieren kann. Da beide verwendeten Computersysteme einen gleich großen Grafikkartenspeicher vorweisen, können die gleichen Modelle auf beiden Systemen trainiert werden.

Die aktuelle 2.0 Version von *DeepFaceLab* funktioniert nur mit Grafikkarten des Herstellers Nvidia, da sie deren proprietäre *CUDA*-Recheneinheiten verwendet. Im Trainingsprozess ist eine hohe Parallelisierbarkeit enorm wichtig, da viele Gesichter so schnell wie möglich auf Gemeinsamkeiten überprüft werden müssen.

Die Verwendung von *DeepFaceLab* setzt jedoch keine Grafikkarte voraus. Als Berechnungsquelle kann auch der Prozessor verwendet werden, jedoch verlängert sich so die Berechnungszeit um ein Vielfaches, selbst wenn der verwendete Prozessor in einem viel höheren Preissegment vorzufinden ist.

Die 1.0 Version von *DeepFaceLab* funktionierte auch mit Grafikkarten des konkurrierten Herstellers AMD, da dort noch die offene Programmierschnittstelle *OpenCL* Verwendung fand. Besitzt man keine passende Grafikarte von Nvidia kann man die aktuelle Version über Cloudlösungen verwenden. Dafür bietet sich unter anderem *Google Colaboratory* an. Auf dieser Plattform kann direkt Python-Quellcode im Browser ausgeführt werden und für dessen Berechnung auch Grafikkarten verwendet werden. Die exakte Grafikkarte kann nicht selbst bestimmt werden, sondern wird dem Benutzer oder Benutzerin zugeteilt, jedoch handelt es sich bei den von Google angegeben Grafikkarten alle um auf Deep Learning spezialisierte, Grafikkarten von Nvidia. Google bietet die Nutzung dieser Plattform beziehungsweise der Rechenhardware dahinter kostenlos an, jedoch kann die Plattform nur zwölf Stunden am Stück verwendet werden. Es wird auch ein kostenpflichtiges Abosystem angeboten. Für 10 Dollar pro Monat werden stärkere Grafikkarten garantiert und die Laufzeit auf 24 Stunden verdoppelt. (*Colaboratory – Google*, 2020)

7.2 Erstellung und Größe der Trainingsdaten

Für die Erstellung der Deep Fake Beispiele wurde angenommen, dass die Verwendung von ungefähr 5.000 hochqualitativen Bildern notwendig ist, um ein brauchbares Ergebnis zu erhalten. Für eine Anzahl von 5.000 Bilder, bei einer standardmäßigen Bildrate von 25 Bildern pro Sekunde, würden in der Theorie drei Minuten und 20 Sekunden an Videomaterial ausreichen. Jedoch gilt zu beachten, dass sich die Trainingsdaten, bezogen auf die Mimik und den eulerschen Winkel, stark unterscheiden sollten. Wird für die Erstellung eines Deep Fakes ein fünf-minütiger Videoclip verwendet, welcher ausschließlich eine Person zeigt, die geradlinig in die Kamera blickt und dabei lächelt, wird das trainierte Modell starke Einschränkungen in der zukünftigen Verwendung haben. Die künstliche Intelligenz kann nur etwas trainieren, das es auch lernen, beziehungsweise sehen kann. Zwar könnte solch ein Deep Fake Prozess gute Ergebnisse liefern, wenn man ebenso ein Video manipuliert, welches eine Person zeigt, die geradlinig in die Kamera lächelt, der Prozess wird jedoch für die meisten andere Fällen keine zufriedenstellenden Ergebnisse erzeugen können. Dementsprechend kann die

Anzahl der Einzelbilder in keiner Weise als einziger vertrauenswürdiger Indikator für die Qualität der Trainingsdaten stehen.

Eine Herangehensweise, die man daran ableiten könnte, wäre, dass man die Trainingsdaten mit so viel Material wie möglich füttert, da somit zwangsweise viele verschiedene Winkel und Gesichtsausdrücke einer Person abgebildet sind. So hat die künstliche Intelligenz genug Informationen, um ein realistisches Model der Person zu erzeugen. Diese Herangehensweise kann jedoch aus zwei Gründen zu massiven Problemen führen. Einerseits kann es kompliziert sein von der Quellperson so viel Videomaterial zu erhalten, dass sich bezogen auf Lichtstimmung, Bildqualität und Farbwerten nicht zu stark voneinander unterscheidet. Der Lernprozess der künstlichen Intelligenz wird erschwert, wenn der Großteil der Einzelbilder einer Person mit einer klassischen sanften TV-Studiobeleuchtung erzeugt worden ist, jedoch ein Teil des Materials unter starker Sonneneinstrahlung erzeugt worden ist und somit überbelichtete Lichtflecke im Gesicht vorhanden sind. Weiters können verschiedene Farbkorrekturen den Lernprozess erschweren, diese könnte man gegebenenfalls vor dem Training jedoch noch anpassen. Gleiches gilt auch für Make-Up, dies kann starke Unterschiede in den Ergebnissen erzeugen. Bärte müssen ebenfalls bedacht werden, da je nach Zeitpunkt der letzten Rasur das Gesicht stark anders aussehen kann. Die Trainingsdaten sollen zwar verschiedenste Gesichtsausdrücke zeigen, jedoch nicht von vielen unterschiedlichen Quellen gewonnen werden. Wenige jedoch längere Quellvideos garantieren einen besseren Trainingsprozess.

Die zweite Problematik, die durch eine starke Erhöhung der Anzahl der Trainingsdaten erstehen kann, ist das Machine Learning Phänomen *overfitting*. Dies bedeutet, dass die künstliche Intelligenz die Trainingsdaten zu gut lernt beziehungsweise auswendig lernt. Die Ergebnisse bezogen auf die Trainingsdaten wirken hervorragend. Verwendet man diese künstliche Intelligenz für neue Daten sieht man jedoch ein unbefriedigendes Ergebnis. Dies kann auftreten, wenn ein großer Teil der Trainingsdaten sich zu ähnlich ist. Wenn somit die Trainingsdaten aus 10.000 Einzelbildern bestehen, jedoch davon 8.000 eine ähnliche frontale Ansicht samt Mimik zeigen, dann werden die erstellten Deep Fakes Probleme aufweisen sobald diese Ansicht im gewählten Zielvideo nicht gegeben ist. Die Qualität solch eines Sets von Trainingsdaten würde enorm erhöht werden, wenn mehrere 1.000 dieser, sich ähnlichen Bilder, gelöscht werden würden. Dies würde nämlich bedeuten, dass während des Trainingsprozesses die Einzelbilder welche anderen Gesichtszüge beziehungsweise Winkel zeigen, ähnlich oft von der künstlichen Intelligenz herangezogen werden, wie die Einzelbilder mit der frontalen Ansicht. In solch einem Fall gewinnt man durch die Reduzierung der Trainingsdaten nicht nur Qualität, sondern erhält eine große Zeitersparnis. Jedoch

muss unbedingt darauf geachtet werden, nicht das gegenteilige Phänomen *underfitting* zu erzeugen. Dabei gibt es für die künstliche Intelligenz zu wenig Trainingsinformationen, um konkrete Schlussfolgerungen zu bilden, somit wird die Anwendung auf neue Daten Fehler vorweisen.

Die oben erläuterten Beobachtungen wurden somit für die Erstellung der Deep Fake Beispiele herangezogen. Die besten Voraussetzungen lauten zusammengefasst wie folgt:

- Eine Einzelbilder Anzahl von ungefähr 5.000 Bildern
- Größtmögliche Diversität in Gesichtsausdrücken und Winkeln der zu erlernenden Person
- Größtmögliche Homogenität der Einzelbilder, bezogen auf Bildqualität, Lichtsituation, Farbkorrektur und Gesichtsmerkmale
- Reduzierung von großen Ansammlungen ähnlicher Einzelbilder, um *overfitting* zu vermeiden

Zur Reduzierung von, sich zu stark ähnelnden Einzelbildern, bietet *DeepFaceLab* über die Sortierungsfunktion eine passende Möglichkeit namens *best faces*, um die Anzahl der Einzelbilder zu verringern. So kann angegeben werden wie viele Einzelbilder aus dem aktuellen Set extrahiert werden sollen. Dabei werden die berechneten Winkel der extrahierten Gesichter miteinander abgeglichen und zu ähnliche Einzelbilder in einen eigenen Ordner verschoben. Dieser Prozess ist sehr rechenintensiv und wird bei *DeepFaceLab* nicht über die Grafikkarte, sondern über den normalen Prozessor abgearbeitet. Die Einzelbilder des reduzierten Trainingssets sind dann nach dem eulerschen Winkel geordnet. So beginnen die Trainingsdaten mit den Gesichtern die, von der Person aus betrachtet, nach rechts blickt und enden mit den Einzelbildern, wo die Person nach links blickt.

7.2.1 Ähnlich große Abbildungsgrößen sind sinnvoll

Die künstliche Intelligenz lernt im Trainingsprozess die Gesichtszüge, Falten, Muttermale und weitere kleine Details, falls die Trainingsdaten hochauflösend genug sind und auch die Abbildungsgrößen entsprechend gewählt werden. Jedes Gesicht wird im Extrahierungsprozess mit der gleichen quadratischen Auflösung extrahiert. Dies bedeutet, dass Gesichter, die größer abgebildet worden sind, im Quellvideo viel mehr Details aufweisen als Gesichter, welche von kleineren Abbildungen stammen.

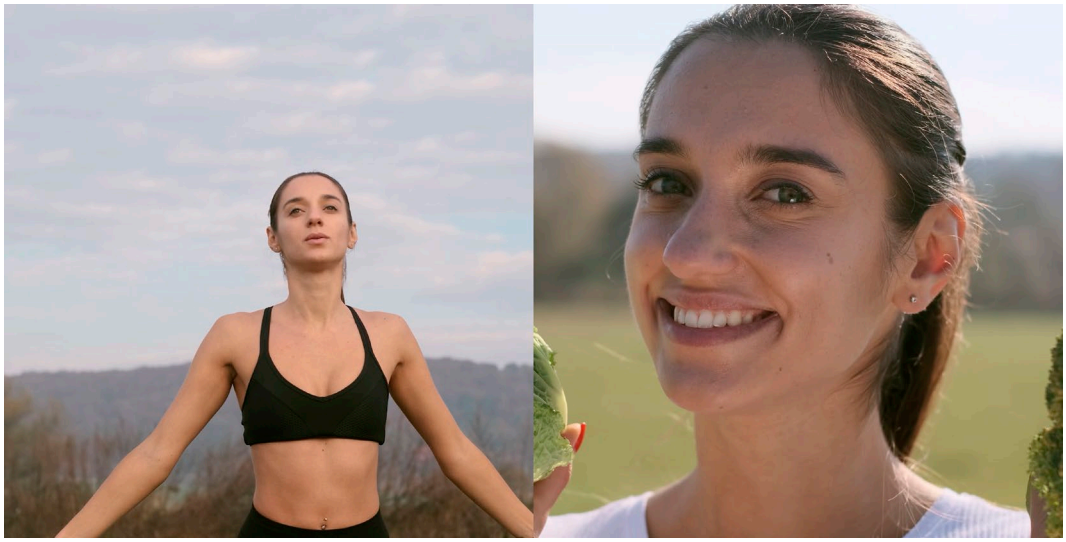

Abbildung 25 – Beide Einzelbilder zeigen die gleiche Person, jedoch durch den großen Unterschied in der Abbildungsgröße ist im linken Bild das markante Muttermal auf der linken Wange nicht erkennbar

Besteht beispielsweise die Hälfte der Trainingsdaten aus einem Video, welche die Quellperson sehr klein darstellen, und die andere Hälfte der Trainingsdaten aus einer Nahaufnahme ist es für die künstliche Intelligenz schwierig kleine Details, wie Muttermale, korrekt abzubilden. Hier muss abgeschätzt werden ob die Hälfte des Trainingsmaterials für ein zufriedenstellendes Deep Fake Ergebnis reicht, oder ob das Fehlen solch kleiner Details für die Veröffentlichung Relevanz hat. Je nach Videokomprimierung beziehungsweise Veröffentlichungsplattform können diese Details gegebenenfalls nicht erkennbar sein.

7.2.2 Ähnlichkeit zwischen Quellmaterial und Zielmaterial

Eine passende Größe und hohe Qualität der Trainingsdaten allein garantieren dennoch kein zufriedenstellendes Deep Fake Ergebnis. Ein weiterer Faktor, der unbedingt bedacht werden muss, ist die prinzipielle Ähnlichkeit zwischen Quellmaterial und Zielmaterial. Auch hier gilt wiederum: Eine künstliche Intelligenz kann nur das synthetisch kreieren, was es zuvor erlernt hat. Zeigt das Zielmaterial einen Wutausbruch einer Person, jedoch das Quellmaterial nur fröhliche Gesichtsausdrücke wird zwangsweise mehr das Ergebnis geschätzt, als wäre die künstliche Intelligenz mit ähnlichem Material trainiert worden. Die Notwendigkeit, eine gewisse Ähnlichkeit zwischen Quellmaterial und Zielmaterial zu erhalten, bezieht sich auf mehrere Faktoren:

- Die Lichtsituation der beiden Materialien sollte ähnlich sein. Weisen die Einzelbilder der Trainingsdaten alle starke harte Schatten im Augenbereich auf werden diese beim Deep Fake Prozess ebenfalls von der künstlichen

Intelligenz erstellt werden, auch wenn diese im Zielmaterial gar nicht vorhanden sind. Dies resultiert in einem verstörenden Deep Fake, wenn der restliche Körper der Person klar und sanft ausgeleuchtet ist.

- Es gilt darauf zu achten, dass zwischen beiden Personen ähnliche Hautfarbtöne vorherrschen. Ein Trainingsmodell, welches mit Material trainiert worden ist, dass eine stark gebräunte Person zeigt, wird zwangsweise beim Umwandlungsprozess ebenso einen ähnlichen Hautfarbton erzeugen, auch wenn die Zielperson eigentlich viel blasser wäre. Hier kann man mit nachträglicher Farbkorrektur in Videoschnitt- beziehungsweise Videocompositing-Software noch eine Verbesserung erzielen.
- Die prinzipielle Gesichtsform sollte ähnlich sein. Trainiert man ein Modell, das auf Einzelbildern von einer Person beruht, welche ein sehr breites Gesicht hat, wird man Qualitätsverluste erzielen, wenn man im Umwandlungsprozess dieses Modell auf ein schmales Gesicht anwenden möchte.
- Die Abbildungsgrößen beider Personen sollten ähnlich sein. Möchte man als Zielvideo eine Nahaufnahme einer Person verwenden, sollten die Trainingsdaten ebenfalls großteils aus Nahaufnahmen bestehen.

Die somit effizienteste Vorgangsweise ist es, sich konkret zu überlegen welches Zielmaterial man zu einem Deep Fake manipulieren möchte. Dabei notiert man sich Charakteristiken wie Lichtstimmung, Farbton, markante Gesichtsausdrücke und Winkel, beziehungsweise hebt vor allem die Extremwerte dieser heraus. Beim darauffolgenden Prozess der Ansammlung der Trainingsdaten kontrolliert man ob diese Extremwerte abgedeckt werden beziehungsweise die notwendigen Ähnlichkeiten vorhanden sind. Selbstverständlich muss an dieser Stelle noch einmal erwähnt werden, dass es sich bei den oben erläuterten Richtlinien um ein theoretisches Ideal handelt, dass nie vollkommen erreicht werden kann. Bei der Sammlung der Trainingsdaten gilt es, bezogen auf das Zielvideo, den größten gemeinsamen Nenner zu finden, um ein passendes Ergebnis zu garantieren.

7.2.3 Manuelle Kontrolle aller Einzelbilder ist sinnvoll

Wurden die Trainingsdaten auf eine sinnvolle Größe reduziert und auch der Gesichtsextraktionsprozess vollführt, sollte unbedingt noch eine abschließende manuelle Kontrolle vorgenommen werden. Dafür bietet *DeepFaceLab* die Möglichkeit ein Debug-Bild von jedem extrahierten Gesicht zu erzeugen. Auf diesem Bild ist das Extraktionsergebnis klar ersichtlich. Hier gilt es darauf zu achten, dass die jeweiligen Gesichtsteile richtig markiert sind.

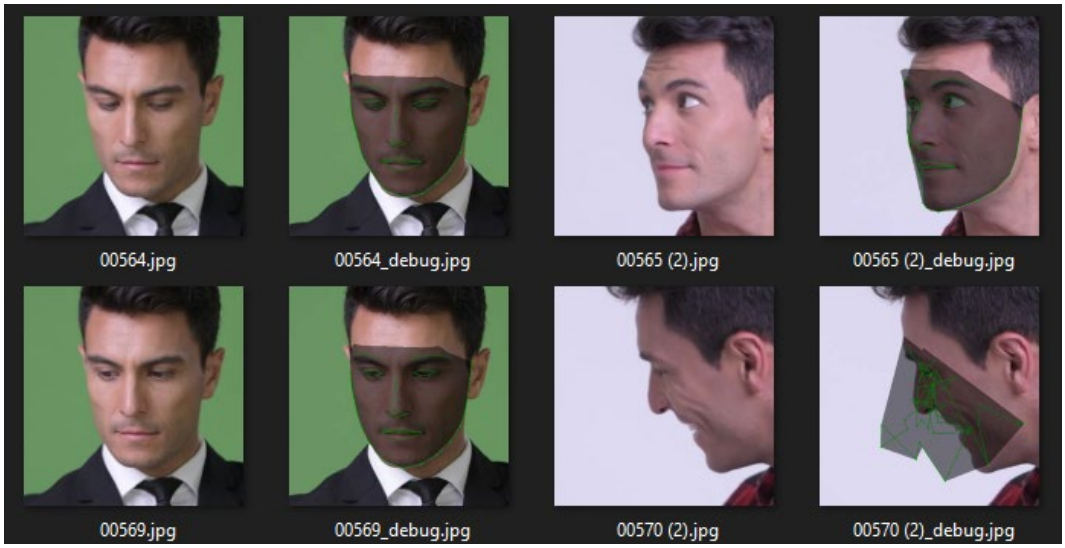

Abbildung 26 – Ansicht von vier einzeln extrahierte Gesichter inklusive dazugehörigen Debug-Bild. Das Bild rechts unten zeigt eine falsche Segmentierung

Zeigt ein Debug-Bild an, dass die Extraktion fehlerhaft war, muss man das dazugehörige extrahierte Gesicht händisch löschen. Zwar ist die manuelle Kontrolle von ungefähr 5.000 Einzelbildern mühsam, jedoch essenziell, um sicher zu gehen, dass die künstliche Intelligenz nicht mit falschen Daten trainiert wird. Wenige falsch extrahierte Gesichter können dafür sorgen, dass das trainierte Modell nicht über ein gutes Niveau hinauskommt und somit ein stundenlanger Trainingsprozess verschwendet worden ist. Ist die manuelle Überprüfung gewissenvoll absolviert worden, kann der Trainingsprozess beginnen.

7.2.4 Problematik mit der Verwendung von Stockmaterial als exklusive Trainingsdatenquelle

Die vorherigen zwei Unterkapitel beschreiben das theoretische Optimum, welches bei der Erstellung der Trainingsdaten bedacht werden soll. Fertigt man das Quellmaterial selbst an, ist dies besonders leicht. Hierbei können zum Beispiel konkret Lichtstimmungen beim Aufnahmeprozess erzeugt werden, um sich dem Zielmaterial anzunähern. Genauso kann eine Person gewählt werden, die der entsprechenden Person vom Zielmaterial ähnlich ist und auch die gleichen Gesichtsausdrücke und Kopfbewegungen nachgeahmt werden.

Bei der Untersuchung, welche im Zuge dieser Diplomarbeit stattfindet, wird jedoch nur auf Bewegtbildmaterial zurückgegriffen, welches auf Stockplattformen publiziert worden ist. So soll garantiert werden, dass die Probanden und Probandinnen die Deep Fakes nicht daran erkennen, dass das Zielvideo ein

Ausschnitt aus einem bekannten Film, Interview oder einer TV-Serie ist. Die eigene Produktion von Trainingsdaten wird aufgrund des theoretischen Näheverhältnis zwischen DarstellerInnen und ProbandInnen abgelehnt. So könnte ein gewählter Proband oder Probandin den Laiendarsteller beziehungsweise die Laiendarstellerin kennen, da beide aus einem geografischen ähnlichen Umfeld kommen beziehungsweise beide auch in Kontakt mit dieser Diplomarbeit stehen. Weiters möchte diese Diplomarbeit ermitteln ob eine Person Deep Fakes erzeugen kann, welche Betrachter und Betrachterinnen nicht als Bewegtbildmanipulationen erkennen, nur durch Einsatz eines klassischen Computersystems ohne zusätzliche Möglichkeiten wie Kameras und DarstellerInnen.

Zwar bieten Videos von Stockplattformen immer eine hochwertige Bildauflösung und eine hohe Datenrate, jedoch ist die Variation der veröffentlichten Videos gering beziehungsweise die Dauer der einzelnen Videos kurz. Diese Stockvideos sind dazu gedacht, innerhalb stark geschnittener Werbevideos verwendet zu werden, wo die Dauer einzelner Clips, somit meistens eine untergeordnete Rolle spielt. Einzelne Videos, die eine Länge von mehreren Minuten aufweisen sind nicht vorhanden, somit müssen die Trainingsdaten aus einer größeren Anzahl einzelner Videos zusammengestellt werden. Hierbei ist jedoch das Problem, dass die Auswahl an Videos, die in einer Abfolge produziert worden sind, sehr gering ist. Stockvideos werden nicht erstellt, um eine narrative Geschichte zu erzählen, sondern um eine Stimmung zu vermitteln oder eine Aussage zu tätigen. Somit findet man gegebenenfalls einige Videos mit demselben Darsteller beziehungsweise Darstellerin, jedoch wird dabei immer die gleiche Tätigkeit und Gestik absolviert und nur die Kleidung der Person von Video zu Video verändert. Würde man die Summe dieser Videos verwenden würde man *overfitting* riskieren, somit sollte aus solch einer Serie nur eine geringe Zahl von Exemplaren gewählt werden. Viele Stockvideos werden auch in Zeitlupe oder gar Superzeitlupe publiziert, dementsprechend unterscheiden sich die verschiedenen Einzelbilder noch weniger. Die Darstellung von Personen innerhalb von Stockvideos ist tendenziell auch sehr frontal und meist ohne Audioteil. Zwar wird keine akustische Information für den Deep Fake Prozess benötigt, jedoch bedeutet dies, dass die meisten Darsteller und Darstellerinnen in Stockvideos nicht sprechen. Viel Mimik und Gesichtsfalten entstehen jedoch gerade dann, wenn eine Person spricht. Sind diese nicht vorhanden, kann das neuronale Netzwerk diese nicht erlernen. Durch die beschriebenen Problematiken eignen sich Stockvideos nur sehr bedingt für die Verwendung von Deep Fakes, vor allem wenn das Zielmaterial diese typischen Stockvideo Charakteristiken nicht vorweist. Da jedoch im Zuge dieser Untersuchung sowohl das Quellmaterial als auch das Zielmaterial auf Stockvideos

basiert ist dieser Faktor weniger gravierend. Dennoch wurde die Schwierigkeit der organisatorischen Arbeit, ein Set an passenden Trainingsdaten via Stockmaterial im Zuge dieser Diplomarbeit zu erarbeiten, unterschätzt und nahm weit mehr Zeit in Anspruch als ursprünglich angenommen.

7.3 Dauer des Lernprozesses

Es gibt keinerlei Möglichkeit eindeutig zu bestimmen, wann der Trainingsprozess weit genug fortgeschritten ist, um ein zufriedenstellendes Deep Fake Ergebnis zu erhalten. Zwar gibt, wie bereits erwähnt, das Textfenster während des Trainingsprozesses Verlustwerte an, jedoch sind diese bezogen auf den resultierten Deep Fake recht aussagelos. Diese Verlustwerte beziehen sich nur auf die Rekonstruktion der Inter-Bilder, also wie gut aus dem reduzierten Abbild wieder das Original gebildet werden kann. Selbst wenn diese beiden Verlustwerte minimal klein sind, bedeutet dies nicht, dass später im Umwandlungsprozess nicht doch Problematiken entstehen können. Würden sich zum Beispiel Quellmaterial und Zielmaterial von der Lichtsituation, der Farbkorrektur, den Gesichtsausdrücken und der Kopfform stark unterscheiden, können die Verlustwerte sehr niedrig sein, jedoch wäre der erstellte Deep Fake dennoch nicht von hochwertiger Qualität. Die Verlustwerte können nur als Indikator gesehen werden, um zu überprüfen, ob sich die künstliche Intelligenz weiter verbessert. Steigen die Werte graduell an oder schießen plötzlich in die Höhe, sind entweder die Trainingsdaten oder die Trainingskonfiguration fehlerhaft. Eine gewisse Anzahl der Iterationen als Endzeitpunkt für den Trainingsprozess kann auch nur bedingt bestimmt werden. Je nach Trainingsdaten und Trainingskonfiguration kann ein Modell schon nach 250.000 Iterationen ein hochwertiges Deep Fake Ergebnis liefern während man bei einem anderen Projekt erst nach 500.000 Iterationen gute Ergebnisse erhält. Ebenso ist es jedoch möglich, dass ein Modell nach 1.000.000 Iterationen immer noch keine gute Qualität liefert, da eventuell fehlerhafte Einzelbilder in den Trainingsdaten vorhanden sind. Dementsprechend war es auch im Zuge der Erstellung dieser Deep Fake Beispiele nicht sinnvoll alle drei Beispiele auf eine gleiche Anzahl von Iterationen zu trainieren beziehungsweise auf ähnliche Verlustwerte zu bringen. Die einzige sinnvolle Methodik ist die manuelle Überprüfung. Die fünfte Spalte im Vorschaufenster des Trainingsprozesses ist ein guter Indikator, wie gut das aktuelle Ergebnis wäre. Eine noch sicherere Überprüfung ist es, den Trainingsprozess zu pausieren und den Umwandlungsprozess zu starten. Ist man mit dem Ergebnis noch nicht zufrieden, so kann man den Trainingsprozess erneut fortsetzen. Diese Vorgangsweise wurde bei der Erstellung der Deep Fake Beispiele gewählt. Die Erfahrungswerte, die im

Zuge dieser Diplomarbeit entstanden sind, zeigen, dass man beim Trainieren eines neuen Modells unter 150.000 Iterationen kein ansprechendes Ergebnis erhält, welches man in einer passablen Bildauflösung beziehungsweise Bildrate veröffentlichen kann. Dementsprechend wurden die Deep Fake Beispiele für die Untersuchung mit höheren Iterationszahlen trainiert. Möchte man sich ungefähr errechnen, wie lange ein Trainingsprozess zu einer gewissen Iterationszahl dauert, kann man die durchschnittliche Iterationsdauer hochrechnen, welche während des Trainingsprozess im Textfenster angezeigt wird. Bei der Erstellung des zweiten Beispielvideos für die Untersuchung dauerte eine Iteration ungefähr 800ms. Würde man dieses System für eine Dauer von 150.000 Iterationen trainieren wollen, würde dies 120.000 Sekunden beziehungsweise 2.000 Minuten oder 33,33 Stunden dauern. Dabei gilt jedoch anzumerken, dass sich im Laufe des Trainings die Dauer einer Iteration durch Anpassung der Trainingsparameter stark erhöhen kann. Selbstverständlich bezieht sich die Dauer des Lernprozesses immens auf die verwendete Hardware. Durch die Verwendung einer leistungsstärkeren Grafikkarte mit mehr Videospeicher und mehr CUDA-Recheneinheiten könnte man das gleiche Modell in einer kürzeren Zeit berechnen oder ein besser aufgelöstes Modell in der gleichen Zeit trainieren.

7.4 Anpassung der Trainingsparameter im Laufe des Trainings

Die Konfiguration der Trainingsparameter bleibt über die Dauer des Trainingsvorgangs nicht gleich, sondern wird über die Zeit hinweg angepasst. Der prinzipielle Ablauf besteht aus zwei Schritten: Zuerst lässt man die künstliche Intelligenz das Gesicht grob erlernen. Im zweiten Schritt werden Parameter so angepasst das Details wie einzelne Zähne oder Muttermale besser erlernt werden können. Konkret bedeutet dies, dass im zweiten Schritt ein Generative Adversarial Network verwendet wird, welches das System zwingt Details besser zu lernen. Die erzeugten Bilder werden von einem Diskriminator auf ihre Echtheit bewertet, die erzeugende künstliche Intelligenz probiert wiederum den Diskriminator in die Irre zu führen. Diese Funktion schon im anfänglichen Trainingsprozess zu aktivieren wäre nicht sinnvoll, da die künstliche Intelligenz zuerst das Gesicht erlernen muss, bevor es Fälschungen erstellen kann, die einer Überprüfung standhalten können. Somit würde ein enormer Performanceverlust entstehen, der keinerlei Nutzen bringt.

Weiters wird *Learning Dropout* aktiviert. Im Allgemeinen versteht man beim Thema Machine Learning unter Dropout eine Methodik, welche den Trainingsprozess

reguliert. Dabei werden auf, zufälliger Basis, Gewichte der unsichtbaren Ebenen des neuronalen Netzwerks auf Null gesetzt. Dies soll einerseits *overfitting* verhindern und gleichzeitig dem System ermöglichen neue Gewichte zwischen den Ebenen des neuronalen Netzwerks auszuprobieren. Es entsteht somit ein zufälliges Rauschen, an welches sich die künstliche Intelligenz anpasst. Das Modell wird dadurch genauer, da es aus einem gelernten Muster leichter ausbrechen kann. Bei *DeepFaceLab* wird eine besondere Art dieser Regulierungsmethodik verwendet, und zwar *Learning Dropout*. Hierbei werden nicht zufällig Gewichte der unsichtbaren Ebenen, sondern die Lernrate welche diese Gewichte verändern würden, auf Null gesetzt. Somit behalten eine gewisse Anzahl von Gewichte ihren alten Wert, während andere mit der Lernrate neu berechnet werden. Der Vorteil besteht darin, dass die Pfade zwischen den Ebenen im neuronalen Netzwerk erhalten bleiben, da sie eben nicht entfernt werden und somit das Training weniger behindert wird. Diese Option kann bei *DeepFaceLab* über die Grafikkarte als auch über den Prozessor berechnet werden. Da bei der gewählten Konfiguration bei der Erstellung der Deep Fake Beispiele die verwendete Grafikkarte fast komplett ausgelastet war, wurde der Prozessor als Berechnungsquelle gewählt. Dadurch erhöht sich die Dauer einer Iteration stärker, als wenn diese Funktion über die Grafikkarte berechnet werden würde, jedoch wurde dadurch ermöglicht diese Funktion dennoch zu nützen. Auch bei dieser Option wäre es nicht sinnvoll, sie bereits am Beginn des Trainings zu aktivieren. (Lin et al., 2019, S. 2)

7.5 Wiederverwendung von trainierten Modellen

Hat man ein Model so gut und ausführlich trainiert, dass es zufriedenstellende Ergebnisse liefert, ist es sinnvoll dieses Modell für zukünftige Projekte wieder zu verwenden. Einerseits kann dieses Modell dazu verwendet werden erneut Deep Fakes zu erzeugen, wobei nur das Zielmaterial immer wieder ausgetauscht wird. Andererseits kann man solch ein Modell auch als Startpunkt verwenden, um ein neues Modell mit neuen Trainingsdaten zu trainieren. Die künstliche Intelligenz kann dadurch viel schneller das grobe Gesicht erlernen, da es schon viel Vorwissen besitzt, auch wenn die Trainingsdaten nun von einer anderen Person stammen. Bei der Verwendung eines vortrainierten Modelles erhält man somit bei einer viel niedrigeren Iterationszahl ein hochwertiges Ergebnis. Diese Methode wurde beim Trainingsprozess des zweiten und dritten Deep Fake Beispielvideo angewendet. Als Startpunkt wurde ein vortrainiertes Modell verwendet, welches auf allgemeine Gesichter für knapp 200.000 Iterationen trainiert wurde. Bei knapp

100.000 Iterationen war das Zwischenergebnis des zweiten und dritten Deep Fake Beispielvideos viel hochwertiger als das Zwischenergebnis des ersten Deep Fake Beispielvideos bei derselben Iterationsanzahl.

7.6 Weitere Anpassung der Fälschung durch Compositing Software

Eine weitere Anpassung der Bewegtbildfälschung durch externe Software ist zwar nicht unbedingt notwendig, jedoch sinnvoll, um die Glaubhaftigkeit der Fälschung weiter zu erhöhen. Im Zuge dieser Arbeit wird für die weitere Anpassung *After Effects* vom Hersteller Adobe verwendet. Der verwendete Versionsname ist *2020*, die exakte Versionsnummer ist *17.1.1*. Für Arbeitsschritte wie Farbkorrektur und Weichzeichnung des Zielmaterials können auch jegliche Art von Videoschnittprogramme verwendet werden. Das Softwareprodukt muss nur die Möglichkeit bieten Luminanzmasken zu interpretieren, da *DeepFaceLab* den Deep Fake auf diese Weise exportiert.

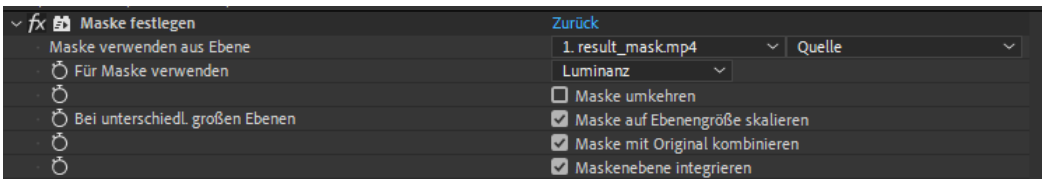

Abbildung 27 – Definition der Maske als Luminanzmaske in After Effects

Für die gezielte Weichzeichnung der Maske und die detaillierte Einstellung der Bewegungsunschärfe sind jedoch Compositing Softwareprodukte wie *After Effects*, *Nuke* oder *Fusion* stark zu empfehlen. Da der Hersteller Blackmagic die Software *Fusion* in einer kostenlosen Version anbietet, ist dieser Arbeitsschritt auch ohne zusätzliche Kosten durchführbar. Folgende vier Arbeitsschritte sind nach dem *DeepFaceLab* Export äußerst sinnvoll um die Glaubwürdigkeit zu erhöhen:

7.6.1 Farbkorrektur

Hier bietet *DeepFaceLab* nur eine Auswahl verschiedener Algorithmen für Farbtransformationen. Somit ist keine exakte Farbkorrektur möglich. Typische Compositing Softwareprodukte bieten eine Vielzahl von Werkzeugen, um detaillierte Farbkorrektur vorzunehmen. Damit ist eine gute Anpassung zwischen dem Zielgesicht und dem restlichen Körper möglich. Ein weiterer Vorteil ist die Möglichkeit, die Farbkorrektur über das Video hinweg anzupassen, wenn sich zum

Beispiel die Lichtstimmung minimal ändert und *DeepFaceLab* die Farbanpassung für das Gesicht in diesem Fall nicht richtig umgesetzt hat.

7.6.2 Weichzeichnung der Maske

Die Maske des Gesichts kann in *DeepFaceLab* zwar weichgezeichnet werden, jedoch kann hier nur eine Variable verändert werden. In Compositing Softwareprodukten kann ein sanfter Übergang über mehrere Variablen detailliert definiert werden.

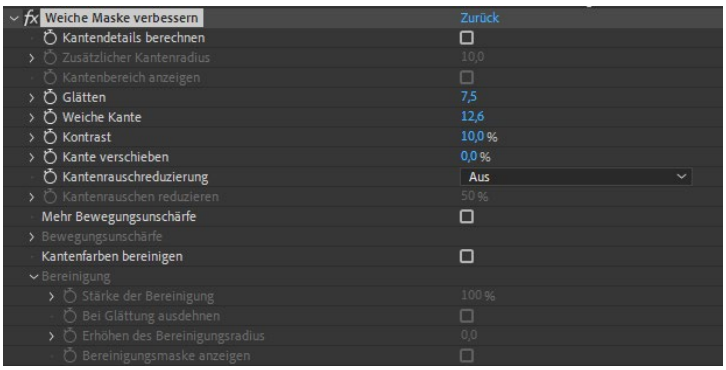

Abbildung 28 – Detaillierte Anpassung der Weichzeichnung der Maske in Adobe After Effects

Weiters kann, falls es zu Farbänderungen an der Kante kommt, die Kantenfarbe detailliert bereinigt werden. Falls die Maske generell noch zu groß ist kann sie auch direkt verkleinert werden.

7.6.3 Weichzeichnung des Zielmaterials

Falls das Zielmaterial sehr hochauflösend ist wird die Glaubhaftigkeit des Deep Fakes dadurch verloren gehen, dass das berechnete Gesicht nicht hochauflösend genug ist und somit der massive Schärfeunterschied sofort evident ist. Deswegen kann es notwendig sein den Deep Fake, bis auf das Gesicht, zu einem gewissen Grad weichzuzeichnen. Für diesen Schritt bietet *DeepFaceLab* im Umwandlungsprozess ebenfalls eine Funktion, jedoch wirkt die Weichzeichnung nicht hochwertig und kann nicht exakt genug angepasst werden.

7.6.4 Bewegungsunschärfe

Um die Glaubhaftigkeit des Deep Fake Exports zu erhöhen kann es sinnvoll sein, Bewegungsunschärfe hinzufügen. Diese kann in Compositing Softwareprodukte sehr detailliert eingestellt werden und wiederum über das Video hinweg angepasst zu werden.

7.7 Trainingsdaten und Trainingskonfiguration der erstellten Deep Fake Beispiele

Hier erfolgt eine Auflistung der Trainingsdaten für die jeweils erstellten Deep Fake Beispiele.

7.7.1 Deep Fake Beispiel 1

Hier ist eine Zusammenfassung der zugehörigen Daten und Einstellungen, die für den Erstellungsprozess des ersten Beispiels angewandt worden sind:

Abbildung 29 – Gegenüberstellung zwischen dem Original links und dem Deep Fake Beispiel 1 rechts

Da das Quellmaterial hauptsächlich aus frontalen Aufnahmen der gezeigten Person besteht, wurde für das Zielmaterial ebenfalls eine Aufnahme gewählt, die größtenteils frontal ist. Jedoch bietet der knapp neun Sekunden lange Ausschnitt mehrere verschiedene Gesichtsausdrücke und zeigt dabei ein ordentliches Fälschungsergebnis.

7.7.1.1 Trainingsdaten

Die *Tabelle 1* bietet einen Überblick über das Ausmaß des Datensatzes des ersten Deep Fake Beispielvideos und erläutert wie viele Einzelbilder davon aussortiert worden sind. Die verwendeten Trainingsvideos sind im Anhang verlinkt.

Materialart	Quellmaterial	Zielmaterial
Anzahl der Einzelbilder	6 431	207
Anzahl der Trainingsbilder	5 000	207
Anzahl Videoclips	18	1
Videoauflösung	3840 mal 2160 Pixel	3840 mal 2160 Pixel
Bildrate	30 fps	23,98 fps

Tabelle 1 – Zusammenfassung der Trainingsdaten des ersten Deep Fake Beispiels

Das Quellmaterial bestand ausschließlich aus Videodateien mit einer Auflösung von 3840 mal 2160 Pixel bei einer Bildrate von 30 Bildern pro Sekunde. Die Aussortierung von Einzelbildern erfolgte zuerst über eine Sortierung nach Unschärfe, wobei die 120 unschärfsten Bilder gelöscht worden sind. Weiters wurde manuell kontrolliert ob falsche Gesichtssegmentierungen vorlagen. Danach wurde der restliche Datensatz über die Sortierungsfunktion *best faces* auf 5.000 Einzelbilder reduziert und das Training gestartet.

7.7.1.2 Trainingskonfiguration

Hier folgt eine Auflistung der Konfiguration des Trainingsprozesses für das erste Deep Fake Beispielvideo. Parameter, die hier nicht aufgelistet sind, wurden nicht verändert und in der standardmäßigen Einstellung von *DeepFaceLab* belassen. Dieses Modell wurde von Grund auf trainiert.

Einstellung	Gesetzter Wert
Auflösung	192 mal 192 Pixel
Gesichtstyp	Gesicht
GPU Berechnung	Ja
Ziel Iterationszahl	400 000
Batch Größe	8
Farbkorrektur	RCT

Tabelle 2 – Gewählte Einstellungen des Trainingsmodells für das erste Deep Fake Beispiel

Als Algorithmus für die Farbtransformation wurde RCT gewählt, da dieser gute Ergebnisse anhand der Vorschaubilder versprach. Bei ungefähr 250.000 Iterationen wurden die Konfiguration geändert. Es wurde *Learning Dropout* aktiviert, als Berechnungsquelle wurde dabei der Prozessor gewählt. Weiters wurde die Variable, welche die *GAN* Herangehensweise steuert, auf den Wert 1 gesetzt und somit aktiviert. Bei einer Iterationszahl von 365.150 wurde das Training beendet, da die Vorschau schon ein zufriedenstellendes Ergebnis versprach.

7.7.1.3 Einstellungen des Umwandlungsprozesses inklusive Nachbearbeitung in After Effects

Für den Umwandlungsprozess wurde die interaktive Ansicht innerhalb von *DeepFaceLab* gewählt. Es wurden bei allen Parameter die standartmäßigen Einstellungen von *DeepFaceLab* belassen, bis auf die Größe und die Weichzeichnung der Maske.

Abbildung 30 – Deep Fake Beispiel 1: Links ohne Anpassungen im Umwandlungsprozess, rechts mit Anpassungen

Die Größe der Maske wurde verkleinert, da das berechnete Gesicht etwas zu groß dargestellt worden ist, die Weichzeichnung wurde erhöht, um einen sanften Übergang zu erhalten. In *After Effects* wurde die Maske noch detailliert weichgezeichnet und das Gesicht minimal von der Farbtemperatur angepasst.

Abbildung 31 – Deep Fake Beispiel 1: Links ohne Farbanpassung, rechts minimale Farbanpassung

Die Person im Zielmaterial hat einen etwas dunkleren Teint als die Person im Quellvideo, dementsprechend war eine Farbanpassung sinnvoll. Das Ergebnis wäre jedoch auch ohne Farbanpassung sehr glaubhaft gewesen. Weiters wurde ein Weichzeichner auf das darunterliegende Video angewandt. Da das Zielvideo, vor allem durch die hohe Videoauflösung von 3840 mal 2160 Pixel, eine hohe Schärfe vorweist wurde diese durch die Weichzeichnung auf ein Niveau reduziert, so dass das Gesicht eine ähnliche Schärfe vorweist. Als letzter Schritt wurde das Video mit einer Auflösung von 1920 mal 1080 Pixel mit einer Bitrate von 5 Mbit pro Sekunde mit einem *h.264-Codec* exportiert.

7.7.2 Deep Fake Beispiel 2

Hier ist eine Zusammenfassung der zugehörigen Daten und Einstellungen, die für den Erstellungsprozess des zweiten Beispiels verwendet worden sind.

Abbildung 32 - Gegenüberstellung zwischen dem Original links und dem Deep Fake Beispiel 2 rechts

Von der Person des Quellmaterials stand viel mehr diverseres Material zu Verfügung als beim ersten Deep Fake Beispiels, deshalb war das fertig trainierte Modell sehr vielseitig. Um diese Vielseitigkeit zu testen wurde als Zielmaterial ein Video verwendet, wo die abgebildete Person starke Veränderungen innerhalb der Gesichtsausdrücke vorwies, sowie nicht frontal in die Kamera blickte.

7.7.2.1 Trainingsdaten

Die *Tabelle 2* bietet einen Überblick über das Ausmaß des Datensatz des zweiten Deep Fake Beispielvideos.

Materialart	Quellmaterial	Zielmaterial
Anzahl der Einzelbilder	18 627	243
Anzahl der Trainingsbilder	4301	243
Anzahl Videoclips	59	1
Videoauflösung	3840 mal 2160	3840 mal 2160
Bildrate	23,98	23,98

Tabelle 3 – Zusammenfassung der Trainingsdaten des ersten Deep Fake Beispiels

Die Trainingsdaten basieren auf 59 einzelnen Videoclips, welche im Anhang dieser Diplomarbeit aufgelistet sind. Das Quellmaterial bestand ausschließlich aus Videodateien mit einer Auflösung von 3840 mal 2160 Pixel bei einer Bildrate von 23,98 Bildern pro Sekunde. Von den 59 verfügbaren Videoclips wurden nicht alle Clips ungeschnitten verwendet. Einige wurden via Videoschnitt gekürzt, um im Vorhinein die Anzahl von redundanten Einzelbilder manuell zu reduzieren. Die Einzelbilder wurden über die Funktion *best faces* auf 4.301 reduziert, eine andere Form der Sortierung wurde nicht angewandt. Beim darauffolgenden Training war schnell ersichtlich, dass es innerhalb der Trainingsbilder Gesichter gab, welche falsch segmentiert worden sind. Nach einer manuellen Kontrolle aller Trainingsbilder wurden die falsch segmentierten Bilder gelöscht und das Training erneut gestartet.

7.7.2.2 Trainingskonfiguration

Hier folgt eine Auflistung der Konfiguration des Trainingsprozesses für das zweite Deep Fake Beispielvideo. Parameter, die hier nicht aufgelistet sind, wurden nicht verändert und in den standardmäßigen Einstellungen von *DeepFaceLab* belassen. Dieses Modell wurde nicht von Grund auf trainiert. Ein vortrainiertes Modell mit ungefähr 200.000 Iterationen wurde als Basis verwendet, deswegen sind die Parameter Auflösung und Gesichtstyp durch das vortrainierte Modell vorgegeben.

Einstellung	**Gesetzter Wert**
Auflösung	224 mal 224 Pixel
Gesichtstyp	Ganzes Gesicht
GPU Berechnung	Ja
Ziel Iterationszahl	400 000
Batch Größe	6

Tabelle 4 - Gewählte Einstellungen des Trainingsmodells für das zweite Deep Fake Beispiel

Bei ungefähr 280.000 Iterationen wurden die Konfiguration geändert. Es wurde *Learning Dropout* aktiviert, als Berechnungsquelle wurde dabei der Prozessor gewählt. Es wurde weiters probiert, bei diesem Modell die Verwendung eines *GAN* Aufbaus zu nutzen, jedoch lieferte das System dann Fehlermeldungen, dass der Grafikkartenspeicher nicht ausreicht. Deswegen wurde die Variable, die das Training unter *GAN* Bedingungen aktiviert, wieder auf null gesetzt. Bei einer

Iterationszahl von 378.231 wurde das Training beendet, da die Vorschau ein zufriedenstellendes Ergebnis vermuten ließ.

7.7.2.3 Einstellungen des Umwandlungsprozesses inklusive Nachbearbeitung in After Effects

Für den Umwandlungsprozess wurde die interaktive Ansicht innerhalb von *DeepFaceLab* gewählt. Es wurden bei allen Parametern die standardmäßigen Einstellungen belassen, bis auf die Größe der Maske, die Weichzeichnung der Maske und den Algorithmus der Farbtransformation.

Abbildung 33 - Deep Fake Beispiel 2: Links ohne Anpassungen im Umwandlungsprozess, rechts mit Anpassungen

Die Verkleinerung der Maske war notwendig, da das berechnete Gesicht etwas zu groß dargestellt worden ist. Durch die Weichzeichnung der Maske wurde ein sanfter Übergang zum Zielmaterial erreicht.

Abbildung 34 - Deep Fake Beispiel 2: Links ohne Farbanpassung, rechts minimale Farbanpassung

In *After Effects* wurde die Farbkorrektur stärker angepasst. Da die Zielperson einen dunkleren Teint hat, wurde die Farbtemperatur minimal wärmer eingestellt und die Sättigung erhöht. Weiter wurde der Schwarzwert gesenkt, da dadurch die Augenbrauen minimal dunkler wurden und somit besser zur Zielperson passten. Das Ergebnis war schon vor der manuellen Farbkorrektur in Ordnung, jedoch wurde die Glaubwürdigkeit dadurch diese Anpassungen stark erhöht.

Das Zielvideo hat, durch die hohe Bildauflösung von 3840 mal 2160 Pixel, eine hohe Schärfe, mit welcher das kreierte Gesicht des neuronalen Netzwerks nicht bieten kann. Deswegen wurde, wie beim ersten Deep Fake Beispiel, ein Weichzeichner auf das darunterliegende Zielvideo angewandt. Als letzter Arbeitsschritt wurde das Deep Fake Video mit einer Auflösung von 1920 mal 1080 Pixel mit einer Bitrate von 5 Mbit pro Sekunde mit einem *h.264-Codec* exportiert.

7.7.3 Deep Fake Beispiel 3

Hier ist eine Zusammenfassung der zugehörigen Daten und Einstellungen, die für den Erstellungsprozess des dritten Beispiels angewandt worden sind.

Abbildung 35 - Gegenüberstellung zwischen dem Original links und dem Deep Fake Beispiel 3 rechts

7.7.3.1 Trainingsdaten

Die *Tabelle 5* bietet einen Überblick über das Ausmaß des Datensatzes des zweiten Deep Fake Beispielvideos und erläutert wie viele Einzelbilder vor dem Beginn des Trainingsprozesses aussortiert worden sind. Als Quelle dienten 59 einzelne Videoclips, welche im Anhang dieser Diplomarbeit aufgelistet sind. Das Quellmaterial bestand ausschließlich aus Videodateien mit einer Auflösung von 3840 mal 2160 Pixel bei einer Bildrate von 23,98 Bildern pro Sekunde.

Materialart	Quellmaterial	Zielmaterial
Anzahl der Einzelbilder	12715	276
Anzahl der Trainingsbilder	5438	276
Anzahl Videoclips	59	1
Videoauflösung	3840 mal 2160 Pixel	3840 mal 2160 Pixel
Bildrate	23,98 fps	23,98 fps

Tabelle 5 - Zusammenfassung der Trainingsdaten des dritten Deep Fake Beispiels

Diese 59 Videoclips wurden, bevor sie in Einzelbilder umgerechnet wurden, via Videoschnitt teilweise verkürzt, um redundante Sequenzen vorab zu reduzieren. Aus dem gekürzten Material wurden 12.715 Einzelbilder exportiert. Nach dem Extraktionsprozess wurden via einer direkten manuellen Kontrolle davon weitere 113 Bilder gelöscht, da der Segmentierungsprozess bei diesen Einzelbildern falsche Ergebnisse lieferte. Aus den übrig gebliebenen 12.602 Einzelbildern wurden mit der Sortierungsfunktion *best faces* 5.438 Einzelbilder extrahiert. Danach erfolgte eine erneute manuelle Kontrolle aller Einzelbilder. Da keinerlei falsch segmentierten Einzelbilder dabei aufschienen wurde der Trainingsprozess gestartet.

7.7.3.2 Trainingskonfiguration

Hier folgt eine Auflistung der Konfiguration des Trainingsprozesses für das dritte Deep Fake Beispielvideo. Es wurden nur die Trainingsparameter aufgelistet, welche von den standardmäßigen Einstellungen in *DeepFaceLab* abweichen. Dieses Modell wurde, wie das zweite Deep Fake Beispiel, nicht von Grund auf trainiert. Es wurde dasselbe vortrainierte Modell mit ungefähr 200.000 Iterationen als Basis verwendet, wie beim Training des zweiten Beispiels. Die Parameter Auflösung und Gesichtstyp sind deshalb durch das vortrainierte Modell vorgegeben.

Einstellung	**Gesetzter Wert**
Auflösung	224 mal 224 Pixel
Gesichtstyp	Ganzes Gesicht
GPU Berechnung	Ja
Ziel Iterationszahl	400 000
Augen Priorität	Ja
Batch Größe	6

Tabelle 6 - Gewählte Einstellungen des Trainingsmodells für das dritte Deep Fake Beispiel

Bei ungefähr 180.000 Iterationen wurden die Konfiguration geändert. Wie beim zweiten Beispiel wurde *Learning Dropout* aktiviert, als Berechnungsquelle wurde dabei erneut der Prozessor gewählt. Da dieselben Einstellungen wie im zweiten Beispiel gewählt wurden, wurde nicht versucht die *GAN* Option zu aktivieren. Die Vermutung lag nahe, dass der Grafikkartenspeicher erneut überlastet werden

würde. Nach insgesamt 330.000 Iterationen wurde das Training pausiert, da die Vorschau vielversprechend aussah. Jedoch wurde dabei bemerkt, dass mehrere weitere tausend Iterationen sinnvoll wären, um bei den Augen ein schärferes Ergebnis zu erhalten. Dafür wurde die Option *Augen Priorität* aktiviert.

7.7.3.3 Einstellungen des Umwandlungsprozesses inklusive Nachbearbeitung in After Effects

Für den Umwandlungsprozess wurde die interaktive Ansicht innerhalb von *DeepFaceLab* gewählt. Es wurden für alle Parameter die standardmäßigen Einstellungen belassen, bis auf die Größe der Maske, die Weichzeichnung der Maske und die Farbtransformation.

Abbildung 36 - Deep Fake Beispiel 3: Links ohne Anpassungen im Umwandlungsprozess, rechts mit Anpassungen

Die Verkleinerung der Maske war notwendig, da das berechnete Gesicht etwas zu groß dargestellt worden ist. Durch die Weichzeichnung der Maske wurde ein sanfter Übergang zum Zielmaterial erreicht.

Abbildung 37 - Deep Fake Beispiel 3: Links ohne Farbanpassung, rechts minimale Farbanpassung

Wie auch bei den anderen beiden Beispielen hat das Zielvideo eine hohe Bildauflösung von 3840 mal 2160 Pixel und dadurch eine hohe Schärfe. Um den Deep Fake glaubhafter zu machen wurde deswegen auf das Zielvideo ein Weichzeichner angewandt. Abschließend wurde das Deep Fake Video mit einer Bildauflösung von 1920 mal 1080 Pixel mit einer Bitrate von 5 Mbit pro Sekunde mit einem *h.264-Codec* exportiert.

8 Auswertung und Analyse der Probandenbefragung

Zwischen 11.08.2020 und 18.08.2020 nahmen 129 Probanden und Probandinnen an der Befragung teil. Davon füllten 17 Personen den Fragebogen nicht vollständig aus; diese Datensätzen wurden für die Auswertung nicht berücksichtigt. Dementsprechend beziehen sich die folgenden Auswertungen und Analysen auf die 112 Personen, die den Fragebogen vollständig ausgefüllt haben. Die Teilnehmergruppe setzte sich aus Personen des persönlichen Umfelds, Arbeitskollegen aber auch fremden Personen, die über verschiedene Facebook Gruppen erreicht worden sind, zusammen. Bezüglich des Geschlechtes nahmen 60 Frauen und 52 Männer an der Befragung teil. Dies resultiert in einer gut ausgeglichenen Geschlechterverteilung von 46,5% zu 53,6%. Bei der Frage bezüglich des Altersbereich gaben 38 Personen an, zwischen 18 und 25 Jahre alt zu sein, 53 Personen gaben an zwischen 26 und 36 Jahre alt zu sein, 9 Personen gaben einen Altersbereich von 35 bis 50 Jahren an, und zwölf der befragten Personen waren über 50 Jahre alt. Niemand der teilnehmenden Personen gab an, unter 18 Jahre alt zu sein.

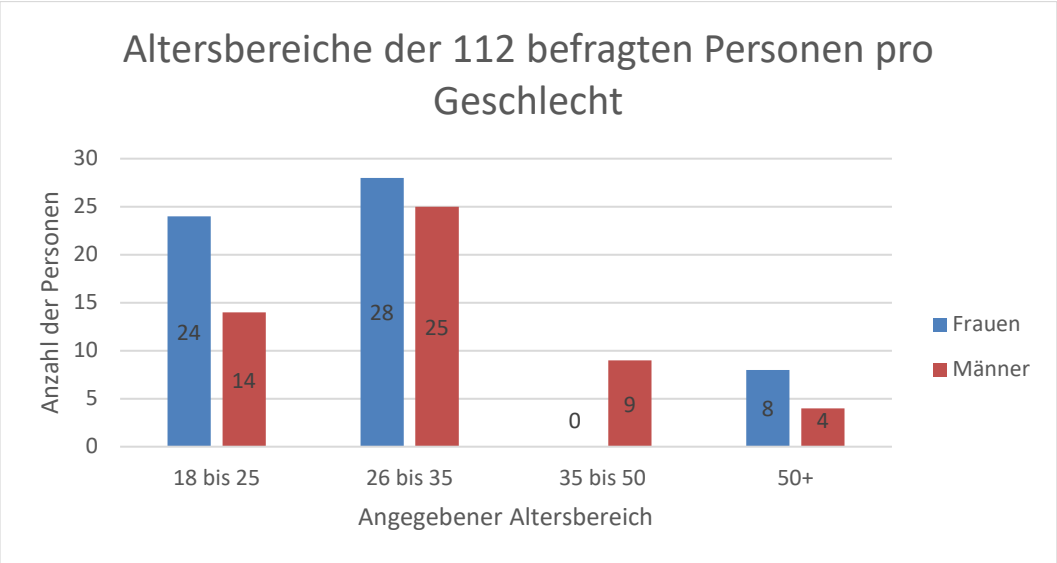

Abbildung 38 – Darstellung der Altersbereiche pro Geschlecht der teilnehmenden Personen

Bei der Frage, ob sich die befragte Person als signifikant bewandert mit dem Thema Bewegtbild sieht, gaben 55 Personen an dies nicht zu sein. Dies entspricht

einem Prozentsatz von 49,1%. 36 Personen beantworteten diese Frage mit der Antwortmöglichkeit „etwas bewandert“, 21 Personen gaben an, mit dem Thema Bewegtbild „signifikant bewandert“ zu sein. Bezüglich der durchschnittlichen wöchentlichen Nutzungsdauer von Sozialen Medien gaben 26 Personen an, diese Plattformen weniger als 3 Stunden pro Woche zu verwenden. Die Antwortmöglichkeit „Zwischen 3 und 10 Stunden“ wählten 43 Personen aus, dies entspricht dem größten Anteil mit einem Prozentwert von 38,4%. Von den 112 Personen gaben sechs an über 30 Stunden pro Woche auf Social Media Plattformen zu verbringen. Bei der Frage nach dem höchsten Bildungsabschluss, wählten 15 Personen die Antwortmöglichkeit „Pflichtschule“, 32 Personen gaben „Matura“ an, 45 Personen wählten die Antwort „Universität (Bachelor)“ und 20 Personen gaben „Universität (Master)“ an. Keine der befragten Personen wählten die Antwortmöglichkeiten „keinen Schulabschluss“ oder „Doktor“.

Tabelle 7 zeigt einen Überblick über die abgebenden Antworten der Frage, ob es sich bei dem gezeigten Material um eine Bewegtbildfälschung handelt, oder nicht.

	Ja (Anzahl)	**Nein (Anzahl)**	**Ja (%)**	**Nein (%)**
Videobeispiel 1	16	96	14,29%	85,71%
Videobeispiel 2	54	58	48,21%	51,79%
Deep Fake Beispiel 2	44	68	39,29%	60,71%
Videobeispiel 3	25	87	22,32%	77,68%
Videobeispiel 4	45	67	40,18%	59,82%
Videobeispiel 5	40	72	35,71%	64,29%
Deep Fake Beispiel 1	59	53	52,68%	47,32%
Videobeispiel 6	50	62	44,64%	55,36%
Deep Fake Beispiel 3	51	61	45,54%	54,46%
Videobeispiel 7	41	71	36,61%	63,39%

Tabelle 7 – Auflistung der Ergebnisse pro Videobeispiel zur Frage „Handelt es sich hierbei um eine Bewegtbildfälschung?“

Daraus können mehrere Schlussfolgerungen gezogen werden: Nur bei zwei Videos hat eine eindeutige Mehrheit die richtige Antwort getätigt. Dies ist bei Videobeispiel 1 und bei Videobeispiel 3 der Fall. Beim Videobeispiel 1 haben nur 14,29% der Probanden und Probandinnen die Frage falsch beantwortet, beim

Videobeispiel 3 lagen mit 22,32% der Probanden und Probandinnen ebenfalls nur ein geringer Anteil falsch. Bei allen anderen Videos gab mindestens ein Drittel der Befragten die falsche Antwort ab, dies inkludiert auch die Deep Fake Beispiele. Das Videobeispiel 2 wurde sogar von über 48% der Befragten als Bewegtbildfälschung deklariert, obwohl es sich um ein unverfälschtes Video handelt. Mit 44,64% wurde auch das Videobeispiel 6 fälschlicherweise von einem sehr großen Anteil der befragten Personen als Bewegtbildfälschung definiert.

Von den drei Deep Fake Beispielen wurde nur das Beispiel 1 mehrheitlich als Fälschung erkannt. Da jedoch der Unterschied mit 52,68% zu 47,32% sehr gering ist, kann diese Bewertung keineswegs als eindeutig bezeichnet werden. Das Deep Fake Beispiel 3 wurde von 45,54% als Fälschung erkannt. Generell gilt es anzunehmen, dass alle Werte, die nah an einem Prozentwert von 50% liegen, ein Indiz dafür sind, dass die Probanden und Probandinnen geraten haben könnten. Das Deep Fake Beispiel 2 wurde nur von 39,29% als Fälschung deklariert. Drei der sieben unveränderten Videos wurden öfter als Fälschung eingeschätzt, als das Deep Fake Beispiel 2. Summiert man alle Antworten der drei Deep Fake Beispiele, so wurden sie mehrheitlich nicht erkannt. Jedoch gibt es dabei keine hohe Mehrheit. 54% der abgegebenen Antworten deklarierten die Fälschungen als unverfälschte Videos.

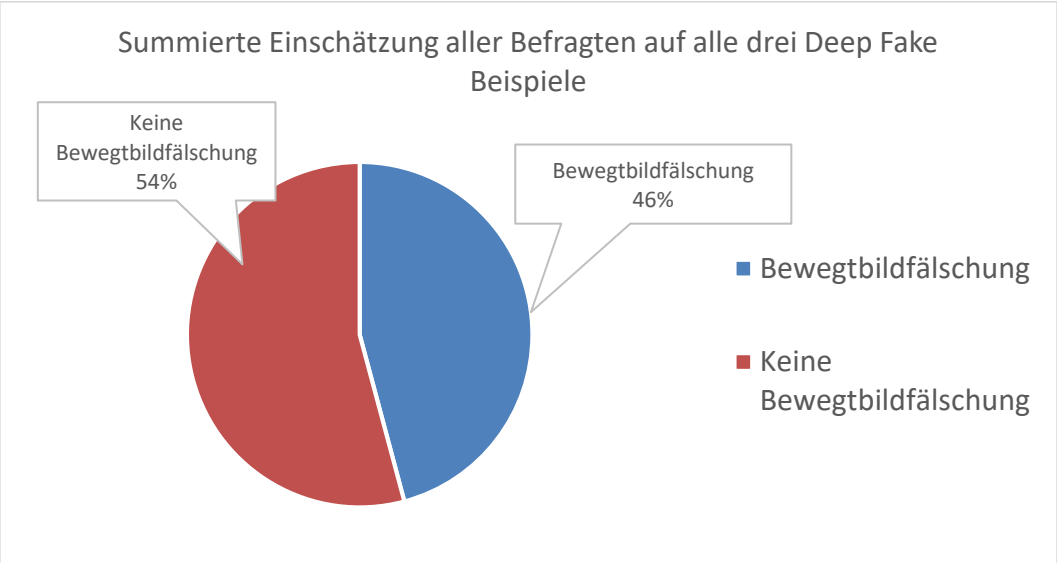

Abbildung 39 – Summe der Einschätzung aller Antworten der Probanden und Probandinnen bezogen auf alle drei Deep Fake Beispiele

Generell kann die Tatsache, dass auch die unverfälschten Videobeispiele so oft als Fälschung deklariert wurden, auf verschiedene Weisen interpretiert werden: Einerseits kann dadurch angenommen werden, dass die hohe Qualität der Deep

Fake Beispiele dafür gesorgt hat, das sie, gegenüber den unverfälschten Beispielen, nicht offensichtlich herausgestochen sind. Da die Probanden und Probandinnen angenommen hatten, das Bewegtbildfälschungen unter den gezeigten an Videos vorhanden waren, wurde von den Probanden und Probandinnen geraten. Würde man eine Gruppe von Personen bezüglich des Ergebnisses eines Münzwurfs befragen, würde man ein ähnliches Diagramm wie in *Abbildung 39* erhalten. Andererseits kann diese Tatsache allerdings auch so interpretiert werden, dass die Testpersonen sich primär auf die Gesichtspartien fokussiert haben und eventuelle natürliche Asymmetrien als Bildmanipulation empfunden haben, da sie über die Art der Bildmanipulation nicht informiert waren.

Die Auswertung der jeweiligen Folgefragen, zur Überzeugung mit der die Einschätzung der befragten Person getätigt worden ist, unterstützt diese Interpretation stark:

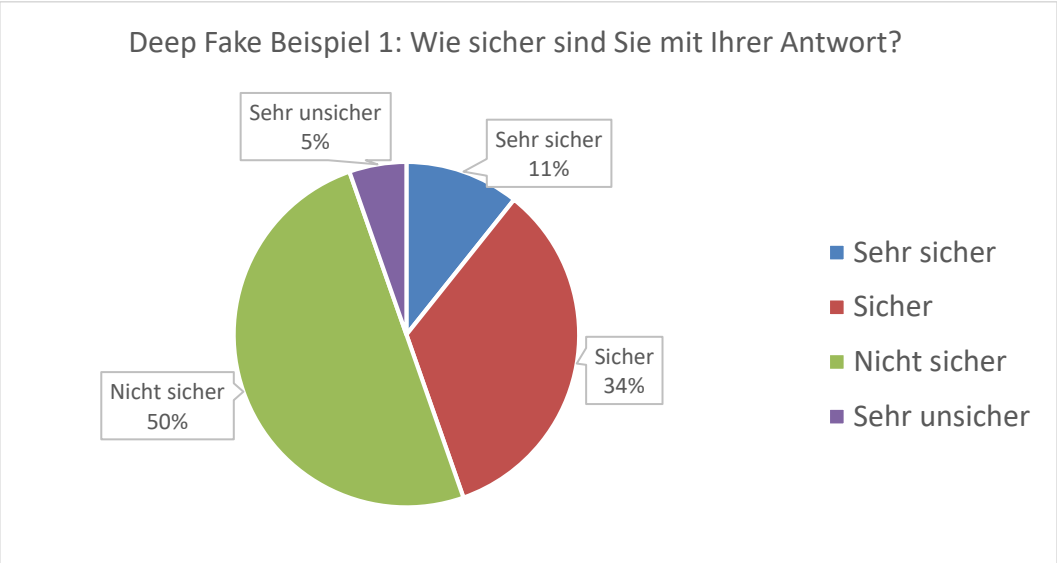

Abbildung 40 – Darstellung mit welcher Überzeugung die Befragten das Deep Fake Beispiel 1 eingeschätzt haben

Wie in *Abbildung 40* ersichtlich, waren 50% der Probanden und Probandinnen unsicher mit ihrer Einschätzung. 5% gaben bei der Beantwortung an, sehr unsicher zu sein. Somit kann angenommen werden das etwa die Hälfte der befragten Personen bei der Beantwortung des Deep Fake Beispiels 1 geraten haben.

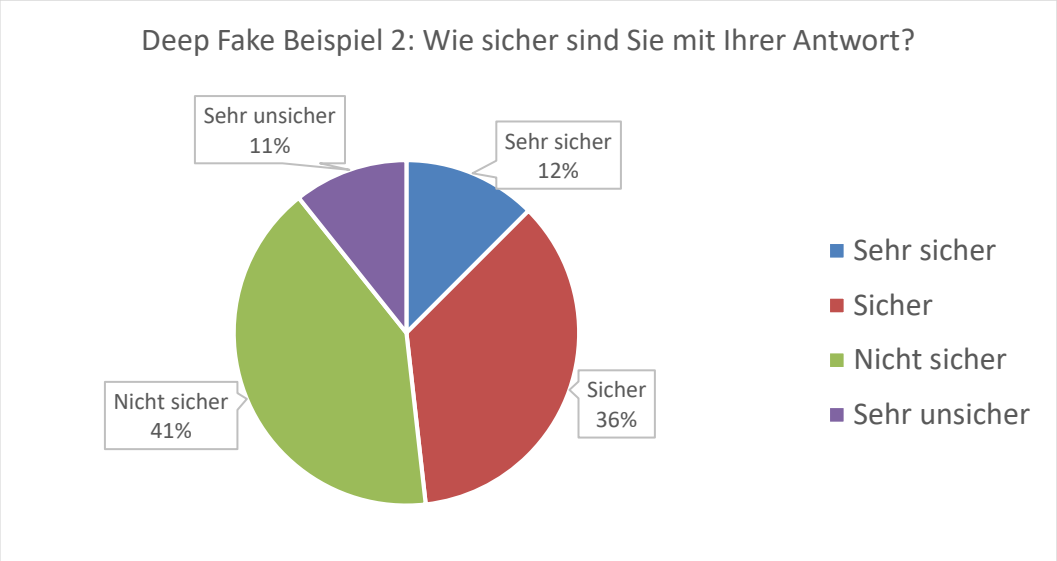

Abbildung 41 – Darstellung mit welcher Überzeugung die Befragten das Deep Fake Beispiel 2 eingeschätzt haben

Im Verhältnis zum Deep Fake Beispiel 1 waren die Befragten beim Deep Fake Beispiel 2 geringfügig sicherer bei der Beantwortung. Hier gaben 36% an, sicher mit ihrer Entscheidung zu sein, 12% gaben an sehr sicher zu sein. Dennoch waren somit über 50% der Befragten nicht überzeugt von ihrer Einschätzung. Deswegen gilt anzunehmen, dass die Probanden und Probandinnen bei der Einschätzung des Deep Fake Beispiel 2 ebenfalls zu einem größeren Teil geraten haben. Hier gilt es nochmal zu erwähnen, dass nur 39,29% der Teilnehmer und Teilnehmerinnen bei der Beantwortung des Deep Fake Beispiel 2 überhaupt richtig lagen.

Beim Deep Fake Beispiel 3 ist die Mehrheit der Befragten von der getätigten Antwort überzeugt gewesen. Jedoch ist die Mehrheit nur knapp erreicht, da 56% der Befragten angaben mindestens „sicher“ mit der getätigten Einschätzung zu sein. Somit ergibt sich auch bei diesem Beispiel kein eindeutiges Bild.

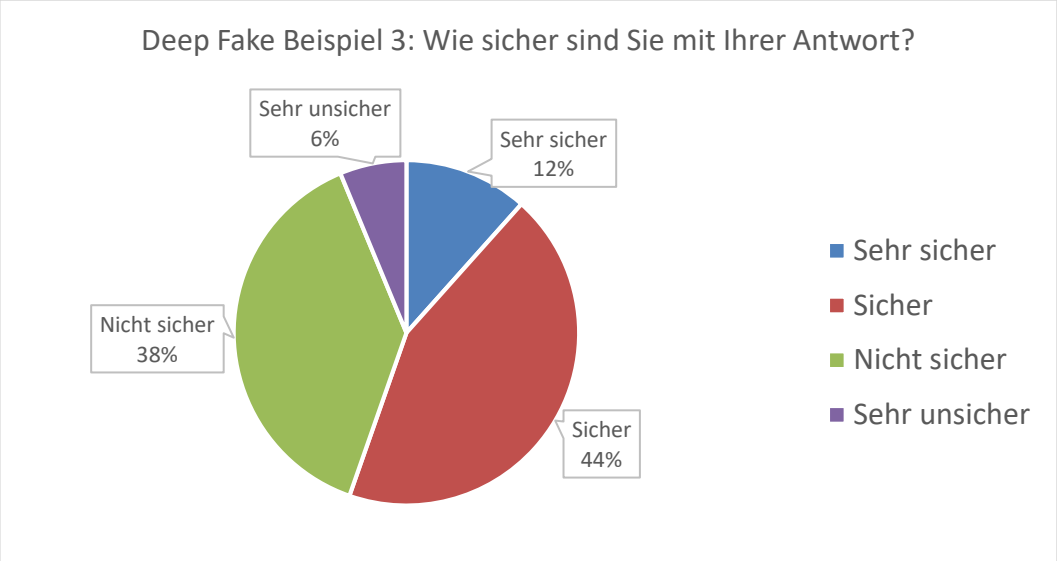

Abbildung 42 - Darstellung mit welcher Überzeugung die Befragten das Deep Fake Beispiel 2 eingeschätzt haben

Diese allgemeine Unsicherheit bei der Beantwortung ist jedoch nicht exklusiv für die Deep Fake Beispiele, sondern ist auch bei den unverfälschten Videos zu beobachten. So war die Mehrheit der Testpersonen bei der Einschätzung des Videobeispiel 7 und Videobeispiel 5 ebenfalls unsicher. Nur beim Beantworten des Videobeispiel 1 war eine eindeutige Mehrheit von 69,64% überzeugt von der abgegebenen Antwort. Bei keinem der anderen unverfälschten Beispiele war eine Mehrheit von 60% oder mehr, überzeugt von der abgegebenen Antwort. Diese allgemeine Unsicherheit wird so interpretiert, dass die Glaubwürdigkeit der Videos auf einem sehr ähnlich hohen Niveau ist, und somit die Probanden und Probandinnen verunsichert waren.

Kombiniert man die Ergebnisse mit den demografischen Daten wird ersichtlich, dass es keine Teilmenge der Probanden und Probandinnen gibt, die bei der richtigen Einschätzung klar besser abschneidet als eine andere. Die naheliegendste Vermutung wäre gewesen, dass Personen, die angeben mit dem Thema Bewegtbild, durch ihren Beruf oder Ausbildung, signifikant bewandert zu sein, besser abschneiden als die Personen, die dieses Näheverhältnis zum Thema verneinen. Wie jedoch *Abbildung 43* zeigt, gibt es hierbei, bezogen auf die Richtigkeit der Antworten, keinen großen Unterschied.

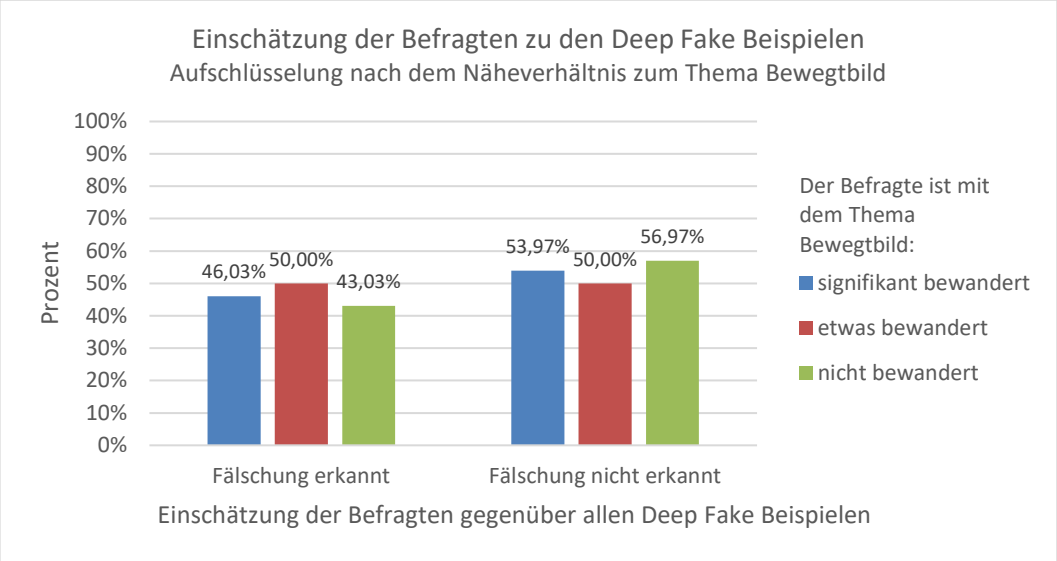

Abbildung 43 – Aufschlüsselung der Ergebnisse nach angegebenen Näheverhältnis der Probanden und Probandinnen zum Thema Bewegtbild

Die 21 Personen, die angegeben haben mit dem Thema Bewegtbild „signifikant“ bewandert zu sein, haben nur mit einem Anteil von 46% die richtigen Antworten abgegeben. Die Personengruppe, die vermerkte mit dem Thema Bewegtbild „etwas bewandert“ zu sein, hatte mit 50% der richtigen Antworten einen größeren Anteil richtig eingeschätzt. Diese Gruppe beinhaltet 36 Personen. Jedoch auch die 55 Probanden und Probandinnen die angaben mit dem Thema Bewegtbild „nicht bewandert“ zu sein, hatten nur um drei Prozentpunkte weniger Antworten richtig als die Personengruppe, die mit dem Thema „etwas bewandert“ ist. Alle drei Gruppen sind einer 50/50-Aufteilung sehr nahe. Erneut gilt es anzumerken: Dies entspricht einem Ergebnis als würde eine Gruppe von Personen Münzwürfe erraten.

Untersucht man die Ergebnisse bezogen auf andere demografische Eigenschaften, ist das Ergebnis sehr ähnlich. *Abbildung 44* zeigt, dass aufgeschlüsselt nach dem angegebenen Geschlecht der Probanden und Probandinnen, das Verhältnis der Antworten sich ebenfalls sehr stark an einer 50/50-Aufteilung orientiert.

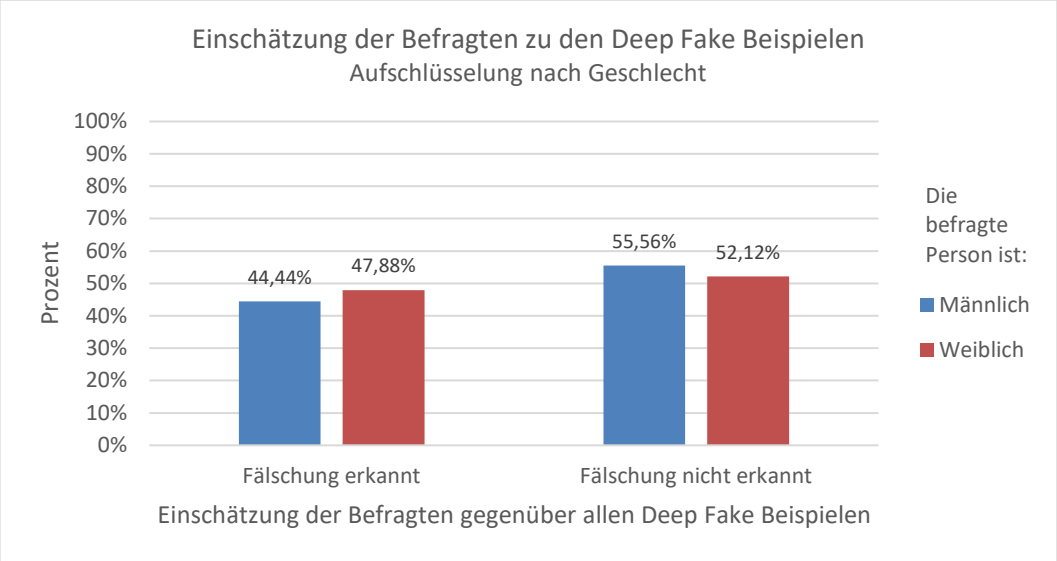

Abbildung 44 - Aufschlüsselung der Ergebnisse nach dem Geschlecht der Probanden und Probandinnen

Bezogen auf die angegebene Altersgruppe der Probanden und Probandinnen ergibt sich ein ähnliches Bild. Die Altersgruppen „18 bis 25“, „26 bis 35“ und „50+“ sind sehr nahe an einer 50/50-Aufteilung. Nur die Personengruppe die den eigenen Altersbereich mit „35 bis 50“ definierte, wich mit 66,67% falsch abgegebenen Antworten, stärker ab. Jedoch haben sich nur neun Personen dieser Gruppe zugeordnet, dementsprechend sind diese Ergebnisse durch die niedrige Teilnehmeranzahl wenig aussagekräftig.

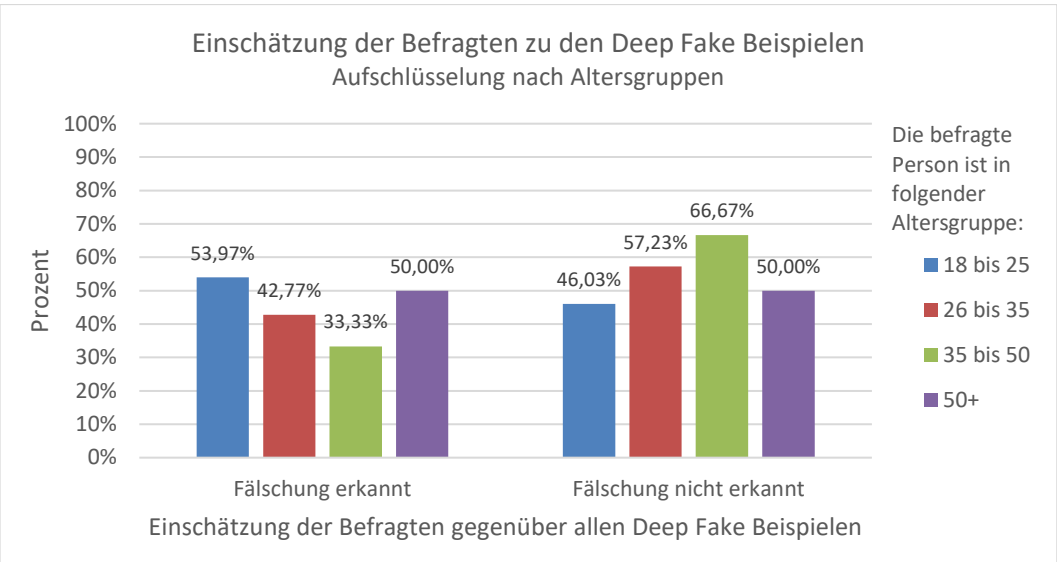

Abbildung 45 - Aufschlüsselung der Ergebnisse nach dem angegebenen Altersbereich der Probanden und Probandinnen

Schlüsselt man die Ergebnisse nach dem höchsten Bildungsabschluss auf, ist die Differenz zwischen den verschiedenen Gruppen sehr gering. Dies ist in *Abbildung 46* erkennbar. Die Personengruppe, welche „Matura“ als höchsten Bildungsabschluss angegeben hat mit einem Prozentsatz von 48,96% die Fälschungen am häufigsten erkannt. Die Personengruppe mit den schlechtesten Ergebnissen, hatten als höchsten Bildungsgrad einen „Pflichtschulabschluss“ angegeben. Jedoch auch diese Gruppe hat nur um 8,96 Prozentpunkte schlechter abgeschlossen.

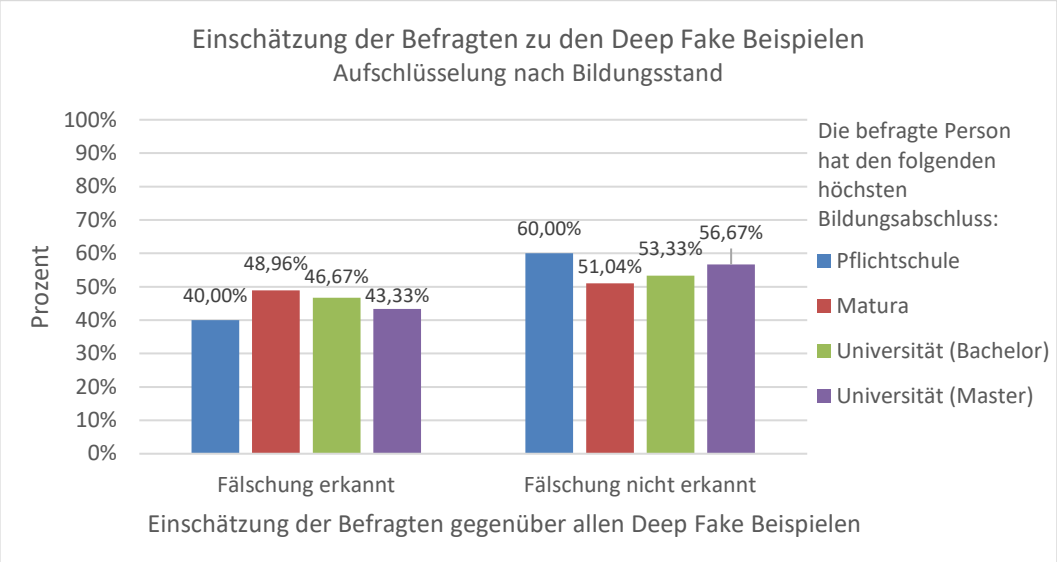

Abbildung 46 - Aufschlüsselung der Ergebnisse nach dem angegebenen höchsten Bildungsabschluss

Bezogen auf die angegebene wöchentliche Social Media Nutzungsdauer ergibt sich erneut ein sehr ähnliches Bild. In *Abbildung 47* ist klar ersichtlich, dass die Nutzungsdauer keinen Einfluss darauf hat, ob die befragte Person die Fälschung eher erkennt. Die Antwortmöglichkeiten „zwischen 20 und 30 Stunden“ und „über 30 Stunden“ wurden zu der Antwort „über 20 Stunden“ zusammengefasst, da diese beiden Antwortmöglichkeiten jeweils nur von wenigen Personen gewählt worden sind. Alle vier Personengruppen liegen mit einer Spanne von nur 6,5 Prozentpunkten sehr nahe beieinander.

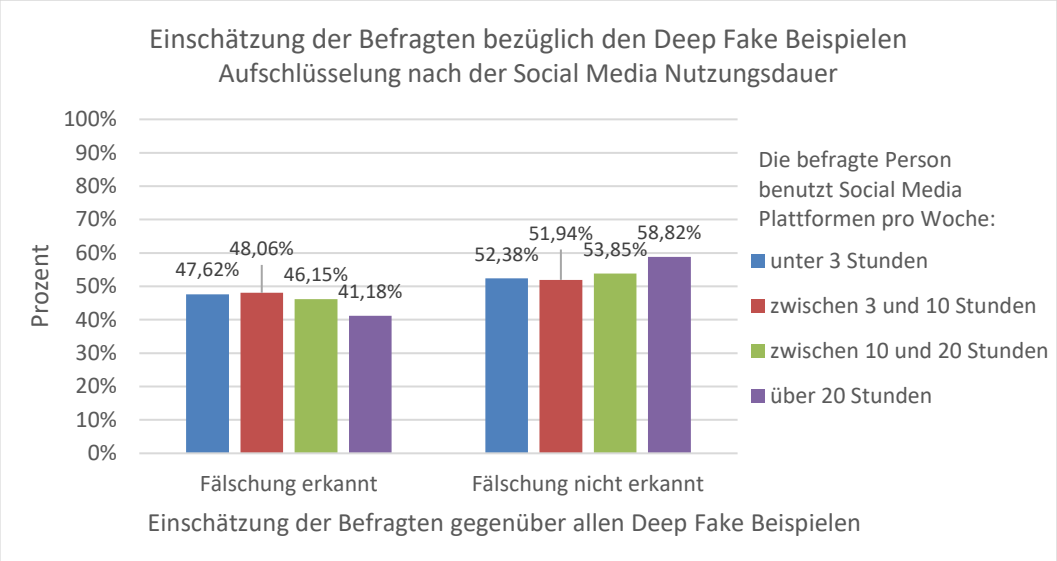

Abbildung 47 - Aufschlüsselung der Ergebnisse nach der angegebenen Social Media Nutzungsdauer der befragten Personen

Pro gezeigtem Video konnten die Probanden und Probandinnen optional in einem Textfeld vermerken, was ihnen beim jeweiligen Beispiel aufgefallen ist. Diese Kommentare sind unter anderem beim zweiten Videobeispiel sehr interessant, da dieses Video von 54% der befragten Personen fälschlicherweise als Fälschung definiert wurde. Die Probanden und Probandinnen begründeten ihre Einschätzung des Videobeispiel 2 unteranderem mit folgenden Kommentaren:

- „Halskragen etwas sehr weit vom Hals“
- „Unnatürliche Bewegungen, Greenscreen“
- „Vielleicht etwas unnatürliche Bewegungen“
- „Gestik und Mimik sind sehr unnatürlich“
- „an manchen Stellen hatte ich das Gefühl, dass die Augenpartie gezittert hat“
- „Unnatürliche Armhaltung, Greenscreen“
- „Komische Bewegungen, Grüner Hintergrund, geschminkte Augen“
- „Hand bewegt sich sehr unnatürlich im Vergleich zum restlichen Körper“

Dementsprechend lässt sich annehmen, dass eine große Anzahl der Befragten durch das Vorhandensein eines Greenscreens angenommen haben, dass es sich um eine Bewegtbildfälschung handelt. Die Kommentare, die sich auf die unnatürlichen Körperbewegungen beziehen, sind nachvollziehbar, jedoch ist dies ein Phänomen, dass auf einen großen Teil von Stockvideos zutrifft. Das Videobeispiel 2 ist, wie auch alle anderen verwendeten Videobeispiele, im Anhang verlinkt. Sieht man sich die Kommentare eines Videobeispiels an, welches viel

häufiger als unverfälscht deklariert wurde, so erhält man viel mehr Kommentare, die diese angenommene Echtheit erläutern. Die Probanden und Probandinnen haben, beim Videobeispiel 4 welches von 77,68% als unverfälscht eingeschätzt worden ist, unter anderem folgende Kommentare abgegeben:

- „Armhaltung + Tattoo wirkt auffällig platziert. Konnte jedoch keine Fälschung feststellen“
- „Das Licht, dass sich auf den Wangen der Person reflektiert, bewegt sich glaubhaft mit den Kopfbewegungen mit.“
- „schaut ganz normal aus“
- „Mundpartie ist uncanny“
- „Bewegungen im Haar-, Kopf- und Kleidungsbereich wirken natürlich.“

Anders als in Videobeispiel 2 bewegt sich die gezeigte Frau im Videobeispiel 3 viel weniger ihre Hände bringt sie in keinem Moment über Hüfthöhe. Die meisten Kommentare haben die hohe Natürlichkeit des Materials erwähnt. Eine Person hat den im Kapitel 2.5 beschriebenen *Uncanny Valley Effekt* angedeutet.

Bei den Kommentaren zur Einschätzung des Deep Fake Beispiel 1, gibt es sowohl Kommentare, die die Echtheit unterstreichen, als auch Kommentare, die die empfundene Falschheit erläutern. Die befragten Personen beschrieben die Einschätzung des Deep Fake Beispiel 2 unteranderem mit folgenden Kommentaren:

- „Sieht normal aus“
- „Direktes Licht von der Front und dennoch sind die Unterarme deutlich schwächer beleuchtet.“
- „Handfarbe heller“
- „Die Hände sind sehr auffällig.“
- „Komische Bewegung“
- „Sämtliche Bewegungen wirken echt.“
- „Nase/Augenbereich wirkt manipuliert.“

Daraus lässt sich ableiten, dass man mit einem stärkeren Fokus auf die manuelle Farbkorrektur, die Glaubwürdigkeit dieses Deep Fakes erhöhen hätte können. Die Kritik bezüglich der komischen Bewegung könnte sowohl am Deep Fake Erstellungsprozess liegen, jedoch auch an der allgemeinen Bewegung des verwendeten Zielvideos.

Beim Deep Fake Beispiel 2, welches von den Probanden und Probandinnen am seltensten als Bewegtbildfälschung erkannt wurde, gab es weniger Kommentare,

die auf die Unechtheit hinwiesen, als beim Deep Fake Beispiel 1. Unter anderem folgende Kommentare wurden von den befragten Personen zu Deep Fake Beispiel 2 abgegeben:

- „Mimik, Gestik und Körperbewegungen sehen natürlich aus“
- „Könnte ein Zusammenschnitt mehrerer Szenen sein“
- „Gesichtsgestik wirkt nicht natürlich“
- „Ist für Laien sehr schwer erkennbar“
- „Gesicht sah zu perfekt aus“
- „Der Mund/Kinn-Bereich wirkt unnatürlich bei seinen Bewegungen.“

Hier sind konkrete Schritte zur Verbesserung der Glaubwürdigkeit schwer ableitbar. Die Mimik und die Kopfbewegungen sind im originalen Zielvideo sehr dramatisch und würden wohl von einigen Personen als unnatürlich interpretiert werden. Da das Deep Fake Beispiel 2 jedoch von 60,71% als unverfälscht deklariert wurde, kann diesem Deep Fake eine hohe Glaubwürdigkeit zugeschrieben werden.

Bei den Kommentaren zum Deep Fake Beispiel 3 wird ersichtlich, dass sich die Probanden und Probandinnen erneut stark auf die Körper und Handbewegungen fokussiert haben, denn unter anderem wurden folgende Kommentare verfasst:

- „Die Neigung der linken Hand mit dem reflektierenden Licht im Glas der Armbanduhr scheint mir zu auffällig.“
- „Die winkende Hand“
- „Ring ist durch Bewegung nur sehr schwer zu erkennen.“
- „Schatten wirken falsch“
- „Wirkt alles echt.“
- „Hautfarbe unnatürlich“
- „Hintergrund digital“

Die ersten drei Kommentare, haben, durch die schnell winkende Hand, die Echtheit des Materials abgeleitet. Alle drei Personen gaben fälschlicherweise an, dass es sich nicht um eine Fälschung handelt. Auch hier sind konkrete Verbesserungsschritte zur Erhöhung der Glaubwürdigkeit schwer von den verfassten Kommentaren ableitbar.

8.1 Gegenüberstellung: Erkennung von Deep Fakes mit Hilfe von neuronalen Netzwerken

Die im Zuge dieser Diplomarbeit durchgeführte Befragung hat gezeigt, dass die Technologie hinter Deep Fakes schon so weit fortgeschritten ist, dass eine eindeutige Mehrheit diese Fälschungen nicht mehr als solche erkennen kann. Dementsprechend ist es sinnvoll, Algorithmen zu entwickeln die Deep Fakes besser als Fälschungen entlarven können. Im September 2019 wurde deshalb die *Deepfake Detection Challenge (DFDC)*[11] in Kooperation von Unternehmen wie AWS, Facebook, Microsoft und dem *The Partnership on AI Steering Committee on AI and Media Integrity* ins Leben gerufen. Ziel war es einen weltweiten Wettbewerb auszurufen, um Forschern und Forscherinnen sowie Entwicklern und Entwicklerinnen die Motivation und Möglichkeit zu bieten, Algorithmen zur Erkennung von Deep Fakes zu entwickeln und zu präsentieren. Dafür wurde, unter anderem, ein Preisgeld von insgesamt einer Million Dollar ausgeschrieben. Nur Einreichungen, bei welchen der Quellcode unter einer Open Source Lizenz veröffentlicht wurde, konnten Preisgelder gewinnen. So würden die besten Algorithmen für jede Person einsehbar sein und zukünftige Projekte könnten auf diesen Ergebnissen aufbauen.

Für die Einreichung gab es verschiedene Bedingungen. So durfte der eingereichte Code nicht auf eine Internetverbindung angewiesen sein. Externe Daten durften die Größe eines Gigabytes nicht überschreiten und sollten ebenfalls frei und öffentlich verfügbar sein. Auch mussten die Algorithmen innerhalb von neun Stunden auf den, zu Verfügung gestellten, Cloud-Recheneinheiten die Referenzdaten durchlaufen und zu einem Ergebnis kommen.

Die Referenzdaten bestanden aus 128.154 Videos, wobei die Bewegtbildmodifikationen mit acht verschiedenen Algorithmen erstellt worden sind. 100.000 davon waren Deep Fakes. Diese Videos wurden durch Mithilfe von 960 Darsteller und Darstellerinnen erzeugt, welche alle diesem Projekt zugestimmt haben. Dieses Datenpaket ist der momentan mit Abstand größte Datensatz zu Bewegtbildfälschungen der öffentlich verfügbar ist. Insgesamt weißt der Datensatz eine Speichergröße von über 25 Terabyte auf. Ähnlich wie die erstellten Deep Fake Beispiele innerhalb dieser Diplomarbeit, wurden die einzelnen Modelle für die Deep Fakes dieses Datenpakets anhand von ungefähr 5.000 Einzelbildern trainiert. Zwar wurden acht verschiedene Algorithmen zur Erstellung verwendet,

[11] https://www.kaggle.com/c/deepfake-detection-challenge

jedoch wurde die größte Anzahl der Deep Fake Beispiele mit dem *DFAE-Model* berechnet. Dieses findet man auch in *DeepFaceLab* vor und entspricht somit eines der meist verwendeten Modelle für veröffentlichte Deep Fakes.

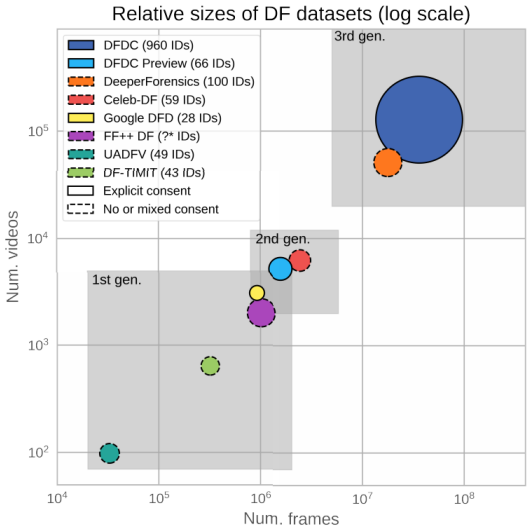

Abbildung 48 – Darstellung der Größe des DFDC Datensatzes im Verhältnis zu ähnlichen Datensätzen *(Dolhansky et al., 2020, S. 2)*

Bei der Erstellung wurde ebenfalls darauf geachtet, dass sich die Personen aus dem Quellmaterial und dem Zielmaterial ähnlich sind. Für den Trainingsprozess wurden über 800 Grafikkarten verwendet. Nach dem abgeschlossenen Training wurden pro Model zehn Sekunden lange Sequenzen exportiert. Dies entspricht auch ziemlich exakt der Länge der Videoclips, welche für die Probandenbefragung innerhalb dieser Diplomarbeit verwendet worden sind.

Damit die entwickelten Algorithmen nicht exklusiv an diesem Datensatz getestet werden, wurde zusätzlich ein nicht publizierter Datensatz mit 10.000 Fälschungen erstellt. Hierbei handelte es sich bei der Hälfte um Deep Fakes. Um die Erkennung der Algorithmen noch stärker zu erschweren wurden unter anderem folgende zusätzliche Bildmanipulationen verwendet:

- Überlagerung durch Textelemente
- Überlagerung durch Bildelemente
- Veränderung der Helligkeit
- Veränderung der Sättigung
- Veränderung der Framerate
- Anwendung von Weichzeichnungsfilter
- Anwendung von Social-Media-Filtern
- Spiegelung des Videos

Insgesamt haben 2.114 Teams an diesem Wettbewerb teilgenommen. Ungefähr 60% aller Einreichungen konnten nur eine Erkennungsrate von 50% oder weniger vorweisen. Der Großteil der Einreichungen lieferte mehrheitlich zufällige Ergebnisse. (Dolhansky et al., 2020, S. 6-10)

Die entwickelte Lösung vom Team *Selim Seferbekov* gewann den Wettbewerb mit einer Erkennungsrate von 65,18%. Die Ergebnisse vieler Teams zeigten eindeutig das *overfitting in den Modellen stattfand*. Beim Testen der Algorithmen gegen den öffentlichen Trainingsdaten fielen die Erkennungsraten weit höher aus. So konnte ein Team beim Test mit dem veröffentlichten Trainingsdaten eine Erkennungsrate von 82,56% erreichen. Wurden diese Systeme jedoch dann mit den fremden unveröffentlichten Daten getestet, viel die Erkennungsrate rapide ab. Das gewinnende Team *Selim Seferbekov* lag beim Test der öffentlichen Daten nur auf Platz vier. (Canton Ferrer et al., 2020)

9 Fazit und Ausblick

Deep Fakes und die ihnen zugrunde liegenden Technologien, besitzen die Schlagkraft das Vertrauensverhältnis der Bevölkerung zum Medium Bewegtbild, gravierend zu ändern. Die Befragung, die im Zuge dieser Diplomarbeit durchgeführt worden ist, hat gezeigt, dass Personen Bewegtbildfälschungen, welche unter der Verwendung von Deep Learning Algorithmen erzeugt worden sind, mehrheitlich nicht als Fälschungen erkennen können. 54% aller abgegebenen Antworten benannten die drei Deep Fake Beispiele als unverfälscht. Damit kann die dritte Forschungsfrage „Kann eine signifikante Mehrheit einer Menschengruppe moderne Bewegtbildfälschungen von unveränderten Videos unterscheiden?“ verneint werden. Die davon abgeleitete zweite Hypothese „Eine signifikante Mehrheit einer Menschengruppe wird es nicht gelingen, hochwertige Deep Fakes als Bewegtbildfälschungen zu erkennen, selbst wenn die Testpersonen darauf hingewiesen werden, explizit darauf zu achten“ trifft somit zu. Die Ergebnisse zeigten weiter, dass es dabei nicht relevant ist, ob die Testperson jung, alt, männlich oder weiblich ist, oder welchen Bildungsgrad sie vorweist. Selbst ein berufliches oder ausbildungstechnisches Näheverhältnis zum Thema Bewegtbild lässt die Personen die Fälschungen nicht mehrheitlich erkennen. Da die Probanden und Probandinnen bei der Befragung explizit darum gebeten wurden, die gezeigten Videos auf ihre Echtheit einzuschätzen, gilt anzunehmen, dass die drei erzeugten Deep Fake Beispiele in einem anderen Kontext gar nicht bis kaum als störend aufgefallen wären. Da die befragten Personen die Fälschungen, welche rein auf Stockvideos basierten, nicht mehrheitlich erkannten, kann auch die zweite Forschungsfrage „Ist es möglich, diese hochwertigen Deep Fakes nur unter Verwendung von Videomaterial von Stockplattformen zu erstellen?“ bejaht werden. Da selbst die einschränkende Auswahl von Stockplattformen für hochwertige Deep Fakes ausreicht, ist die Erstellung von Deep Fakes von Personen des öffentlichen Lebens ebenso leicht umsetzbar.

Bewegtbildfälschungen sind schon lange kein außergewöhnliches Phänomen mehr. Jedoch ist die Tatsache, dass Einzelpersonen diese Fälschungen ohne der Verwendung von kostspieliger Profi-Hardware oder Spezialwissen erzeugen können, eine neuere Entwicklung. Die Verwendung von klassischer Konsumenten-Hardware war ausreichend, um Bewegtbildfälschungen zu erzeugen, welche die Mehrheit der Probanden und Probandinnen nicht als solche erkannt haben. Damit kann die erste Forschungsfrage „Ist es möglich, als Einzelperson ohne spezielle Profihardware Deep Fakes zu erzeugen, welche eine hohe subjektive

Glaubwürdigkeit besitzen?“ bejaht werden. In Folge dessen bestätigt sich auch die erste Hypothese „Es ist möglich mit aktuellen Computersystemen aus dem Konsumentenbereich, unter der Verwendung von Werkzeugen, die auf künstlicher Intelligenz passieren, sehr hochwertige Bildfälschungen zu erstellen“. Noch vor wenigen Jahren hätte man solche Ergebnisse mehrköpfigen Teams beziehungsweise Visual Effects Firmen zugeschrieben. Der Durchbruch von Algorithmen, die auf künstlicher Intelligenz basieren, macht es jedoch auch Einzelpersonen möglich. Leider ist das Gefahrenpotential von hochwertigen Deep Fakes offensichtlich hoch. Menschen können durch diese Fälschungen erpresst werden, von anderen gedemütigt werden und selbst Identitätsdiebstahl kann betrieben werden. In Zeiten von *Fake News* und Verschwörungstheorien können Menschen damit auch gezielt beeinflusst werden. Durch Deep Fakes können *Fake News* plötzlich echter den je wirken. Dementsprechend ist es notwendig Kontrollmechanismen zu entwickeln. Wettbewerbe wie die *Deepfake Detection Challenge* sind ein guter und notwendiger Anfang und zeigen auf, dass einige Firmen die Signifikanz dieser Problematik bereits erkannt haben. Gleichzeitig zeigt jedoch die Erkennungsrate von 65,18% bei Bewegtbildsfälschung des Gewinnerteams, dass es noch viel Raum nach oben gibt. Da diese Erkennungsalgorithmen noch viel Rechenleistung benötigen, können sie deswegen keinesfalls auch nur auf einen Bruchteil der täglich hochgeladenen Videoinhalte angewendet werden. Hier gilt zu berücksichtigen, dass sich auch die Qualität der Deep Fake Softwareprodukte und somit deren Ergebnisse vermutlich stätig verbessern wird.

Durch die Ergebnisse dieser Arbeit zeigen sich mehrere Möglichkeiten für zukünftige Forschungen. Es könnte eine ähnliche Befragung durchgeführt werden, welche ein A/B-Testformat verwendet. Dabei erhält eine Personengruppe eine Auswahl von komprimierten Videobeispielen, wobei die zweite Personengruppe dieselben Videos in unkomprimierter Form betrachten kann. Eine weitere Möglichkeit für ein A/B-Testformat wäre einerseits die Testvideos stumm vorzuspielen, die andere Gruppe erhält dieselben Videos mit synthetisch erzeugter Sprache. Auch könnte man testen, ob Probanden und Probandinnen durch ein zuvor stattfindendes Training mit Deep Fake Beispielen, bei einem Erkennungstest besser abschneiden als Personen, die dieses Training nicht erhielten.

Technologien, die auf künstlicher Intelligenz basieren, haben in den letzten Jahren massive Fortschritte gemacht. Diese Entwicklung wird auch im Bereich der Deep Fakes weiter fortschreiten. Die Möglichkeiten und Gefahren dieser Technologie früh zu erkennen und zu erforschen kann somit von großer Bedeutung sein, bezogen auf die Zukunft des Mediums Bewegtbild.

Literaturverzeichnis

2017 Edelman TRUST BAROMETER. (2017). Edelman. https://www.edelman.com/research/2017-edelman-trust-barometer

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. *arXiv:1605.08695 [cs]*. http://arxiv.org/abs/1605.08695

Ahire, J. (2018). *Artificial Neural Networks: The brain behind AI*. Lulu.com.

Alpaydin, E. (2019). *Maschinelles Lernen*. Walter de Gruyter GmbH & Co KG.

Biethahn, J., Hönerloh, A., Kuhl, J., Leisewitz, M.-C., Nissen, V., & Tietze, M. (1998). *Betriebswirtschaftliche Anwendungen des Soft Computing: Neuronale Netze, Fuzzy-Systeme und Evolutionäre Algorithmen*. http://link.springer.com/openurl?genre=book&isbn=978-3-528-05596-7

Bulat, A., & Tzimiropoulos, G. (2017). How Far are We from Solving the 2D & 3D Face Alignment Problem? (And a Dataset of 230,000 3D Facial Landmarks). *2017 IEEE International Conference on Computer Vision (ICCV)*, 1021–1030. https://doi.org/10.1109/ICCV.2017.116

Canton Ferrer, C., Dolhansky, B., Pflaum, B., Bitton, J., Pan, J., & Lu, J. (2020, Juni 12). *Deepfake Detection Challenge Dataset*. Facebook AI. https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A Dataset for Recognising Faces across Pose and Age. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 67–74. https://doi.org/10.1109/FG.2018.00020

Chau, R., Doyle, B., Doczy, M., Datta, S., Hareland, S., Jin, B., Kavalieros, J., & Metz, M. (2003). Silicon nano-transistors and breaking the 10 nm physical gate length barrier. *61st Device Research Conference. Conference Digest (Cat. No.03TH8663)*, 123–126. https://doi.org/10.1109/DRC.2003.1226901

Chesney, R., & Citron, D. K. (2018). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security* (SSRN Scholarly Paper ID 3213954). Social Science Research Network. https://papers.ssrn.com/abstract=3213954

Colaboratory – Google. (2020). Colaboratory - Frequently Asked Questions. https://research.google.com/colaboratory/faq.html#resource-limits

Cook, D. A. (1996). *A history of narrative film* (3rd ed). W.W. Norton.

Day, S. (2019, November 25). *MIT art installation aims to empower a more discerning public*. MIT News. http://news.mit.edu/2019/mit-apollo-deepfake-art-installation-aims-to-empower-more-discerning-public-1125

de Borst, A. W., & de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00576

Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., & Ferrer, C. C. (2020). The DeepFake Detection Challenge Dataset. *arXiv:2006.07397 [cs]*. http://arxiv.org/abs/2006.07397

Estrada, O., Peretti, N., & Figueroa, R. (2019). *Rotoscope Automation with Deep Learning*. 1–9. https://doi.org/10.5594/M001867

FAILES, I. (2019). THE NEW ARTIFICIAL INTELLIGENCE FRONTIER OF VFX. *VFX Voice*, *Spring 2019*, 23–27.

Farish, K. (2020). Do deepfakes pose a golden opportunity? Considering whether English law should adopt California's publicity right in the age of the deepfake. *Journal of Intellectual Property Law & Practice*, *15*(1), 40–48. https://doi.org/10.1093/jiplp/jpz139

Feng, Y., Wu, F., Shao, X., Wang, Y., & Zhou, X. (2018). Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Hrsg.), *Computer Vision – ECCV 2018* (Bd. 11218, S. 557–574). Springer International Publishing. https://doi.org/10.1007/978-3-030-01264-9_33

Finance, C., & Zwerman, S. (2015). *The visual effects producer: Understanding the art and business of VFX*. Routledge.

Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., & Moran, S. (2018). Falling for Fake News: Investigating the Consumption of News via Social Media. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–10. https://doi.org/10.1145/3173574.3173950

Frangoul, A. (2018, März 14). *With over 1 billion users, here's how YouTube is keeping pace with change*. CNBC. https://www.cnbc.com/2018/03/14/with-over-1-billion-users-heres-how-youtube-is-keeping-pace-with-change.html

Ghazi, M. M., & Ekenel, H. K. (2016). A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 102–109. https://doi.org/10.1109/CVPRW.2016.20

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Hrsg.), *Advances in Neural Information Processing Systems 27* (S. 2672–2680). Curran Associates, Inc. http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

Harwell, D. (2018, Dezember 30). *Fake-porn videos are being weaponized to harass and humiliate women: 'Everybody is a potential target'*. Washington Post. https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.

Henne, M., Hickel, H., Johnson, E., & Konishi, S. (1996). The making of Toy Story [computer animation]. *COMPCON '96. Technologies for the Information Superhighway Digest of Papers*, 463–468. https://doi.org/10.1109/CMPCON.1996.501812

Hu, G., & Clark, J. (2019). Instance Segmentation Based Semantic Matting for Compositing Applications. *2019 16th Conference on Computer and Robot Vision (CRV)*, 135–142. https://doi.org/10.1109/CRV.2019.00026

Huang, Y., Bai, Y., Li, R., & Huang, X. (2016). Research of Canny edge detection algorithm on embedded CPU and GPU heterogeneous systems. *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 647–651. https://doi.org/10.1109/FSKD.2016.7603250

Ilievski, A., Zdraveski, V., & Gusev, M. (2018). How CUDA Powers the Machine Learning Revolution. *2018 26th Telecommunications Forum (TELFOR)*, 420–425. https://doi.org/10.1109/TELFOR.2018.8611982

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, *51*(8), 56–59. https://doi.org/10.1109/MC.2018.3191268

Jurassic Park T-rex robot—As dangerous as a real dinosaur. (o. J.). Stan Winston School. Abgerufen 8. Februar 2020, von https://www.stanwinstonschool.com/blog/jurassic-park-t-rex-robot-almost-eats-crewmember

Kreutzer, R. T., & Sirrenberg, M. (2019). *Künstliche Intelligenz verstehen: Grundlagen – Use-Cases – unternehmenseigene KI-Journey*. Springer-Verlag.

Lin, H., Zeng, W., Ding, X., Huang, Y., Huang, C., & Paisley, J. (2019). Learning Rate Dropout. *arXiv:1912.00144 [cs]*. http://arxiv.org/abs/1912.00144

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, *7*(3), 297–337. https://doi.org/10.1075/is.7.3.03mac

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Netzley, P. D. (2000). *Encyclopedia of movie special effects*. Oryx Press.

Okun, J. A., & Zwerman, S. (2010). *The VES Handbook of Visual Effects: Industry Standard VFX Practices and Procedures*. Taylor & Francis.

Perov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Umé, C., Dpfks, M., Facenheim, C. S., RP, L., Jiang, J., Zhang, S., Wu, P., Zhou, B., & Zhang, W. (2020). DeepFaceLab: A simple, flexible and extensible face swapping framework. *arXiv:2005.05535 [cs, eess]*. http://arxiv.org/abs/2005.05535

Prince, S. (2011). *Digital Visual Effects in Cinema: The Seduction of Reality*. Rutgers University Press.

Schwartz, O. (2018, November 12). You thought fake news was bad? Deep fakes are where truth goes to die. *The Guardian*. https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth

Shay, D., & Duncan, J. (1993). *The Making of Jurassic Park*. Ballantine Books.

Stokes, K. (2020). *2019 EDELMAN TRUST BAROMETER*. 78.

Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics*, *36*(4), 1–13. https://doi.org/10.1145/3072959.3073640

Switft, A. (2016, September 14). *Americans' Trust in Mass Media Sinks to New Low*. Gallup.Com. https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx

tutsmybarreh. (2020, August 21). *[SFW] [GUIDE]—DeepFaceLab 2.0 EXPLAINED AND TUTORIALS (recommended)* [Deepfake Forum]. https://mrdeepfakes.com/. https://mrdeepfakes.com/forums/thread-sfw-guide-deepfacelab-2-0-explained-and-tutorials-recommended

Venkatasawmy, R. (2013). *The Digitization of Cinematic Visual Effects: Hollywood's Coming of Age*. Rowman & Littlefield.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Winick, E. (2019, Mai 23). *Actors are digitally preserving themselves to continue their careers beyond the grave*. MIT Technology Review. https://www.technologyreview.com/s/612291/actors-are-digitally-preserving-themselves-to-continue-their-careers-beyond-the-grave/

Woolley, S. (2020). *The Reality Game: A gripping investigation into deepfake videos, the next wave of fake news and what it means for democracy*. Hachette UK.

WyntersErik. (2011). Parallel processing on NVIDIA graphics processing units using CUDA. *Journal of Computing Sciences in Colleges*. https://dl.acm.org/doi/abs/10.5555/1859159.1859173

Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., & Li, S. Z. (2017). S^3FD: Single Shot Scale-Invariant Face Detector. *2017 IEEE International Conference on Computer Vision (ICCV)*, 192–201. https://doi.org/10.1109/ICCV.2017.30

Abbildungsverzeichnis

Tabellenverzeichnis

Anhang

A. Quelllisten der Trainingsdaten

Auflistung der Weblinks der verwendeten Videos, die zum Training der künstlichen Intelligenz genutzt wurden. Die Links wurden zuletzt am 27.08.2020 aufgerufen und waren zu diesem Zeitpunkt alle funktionstüchtig.

Deep Fake Beispiel 1 Trainingsdaten

1. https://elements.envato.com/de/handsome-guy-saving-coins-UR93MJF
2. https://elements.envato.com/de/handsome-businessman-R78HZAB
3. https://elements.envato.com/de/smiling-guy-in-stripes-Z3KE2XM
4. https://elements.envato.com/de/smiling-casual-guy-FSZM47B
5. https://elements.envato.com/de/good-looking-guy-smiling-T7DUHZV
6. https://elements.envato.com/de/handsome-man-with-stubble-ASW3ZNT
7. https://elements.envato.com/de/handsome-guy-walking-in-to-shot-LXJHVWA
8. https://elements.envato.com/de/good-looking-guy-MP2TVS9
9. https://elements.envato.com/de/guy-looking-up-to-camera-6ZHGCT3
10. https://elements.envato.com/de/smiling-handsome-guy-RHUNF3X
11. https://elements.envato.com/de/handsome-guy-counting-down-on-hand-S3QXHJR
12. https://elements.envato.com/de/thumbs-up-guy-smiling-VD95EG4
13. https://elements.envato.com/de/happy-guy-greeting-with-hand-C8KP7YU
14. https://elements.envato.com/de/posing-and-handsome-guy-ZK9HLRS
15. https://elements.envato.com/de/guy-removing-glasses-4GX9R8Q
16. https://elements.envato.com/de/handsome-guy-with-idea-DXSMNFH
17. https://elements.envato.com/de/handsome-model-turning-to-camera-4UMBD8E
18. https://elements.envato.com/de/guy-looking-up-to-camera-B8YFV2S

Deep Fake Beispiel 2 Trainingsdaten

1. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--GK8V9D8
2. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--FCFY42S
3. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--VRTA8NZ
4. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--2U4JAXP

5. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--MXRNUYQ
6. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--CQ8T9WP
7. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--BATAF7R
8. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--G8TV88B
9. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--3XJQ6XW
10. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--UKUVT4G
11. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--849G8FC
12. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--Y5GFGPB
13. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--LRR844J
14. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--Y65ZHAW
15. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--7QHBZ53
16. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--KUUMQV6
17. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--5KJQB34
18. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--YKMWC7D
19. https://elements.envato.com/de/young-handsome-hispanic-businessman-against-green--EQFG9A2
20. https://elements.envato.com/de/young-happy-hispanic-man-thinking-while-using-digi-8HTEGDX
21. https://elements.envato.com/de/young-happy-hispanic-man-waving-hand-ready-for-win-4ANETRD
22. https://elements.envato.com/de/young-happy-hispanic-man-pointing-up-ready-for-win-82LD3EY
23. https://elements.envato.com/de/young-happy-hispanic-man-touching-something-and-cr-6TZ25GB
24. https://elements.envato.com/de/young-happy-hispanic-man-catching-something-ready--AL5SRVW

25. https://elements.envato.com/de/young-happy-hispanic-man-comparing-something-ready-QALT69J
26. https://elements.envato.com/de/young-happy-hispanic-man-showing-something-to-the--BN9FS2Y
27. https://elements.envato.com/de/happy-young-hispanic-hipster-man-showing-phone-and-6CNZGDV
28. https://elements.envato.com/de/happy-young-hispanic-hipster-man-thinking-while-us-C526934
29. https://elements.envato.com/de/happy-young-hispanic-hipster-man-using-phone-and-g-39RBCMY
30. https://elements.envato.com/de/profile-view-of-happy-young-hispanic-hipster-man-s-X2DT67W
31. https://elements.envato.com/de/happy-young-hispanic-hipster-man-thinking-while-us-NSPH8GW
32. https://elements.envato.com/de/happy-young-hispanic-hipster-man-thinking-while-sh-YG8RWTP
33. https://elements.envato.com/de/happy-young-hispanic-hipster-man-thinking-while-sh-PSFZ9LE
34. https://elements.envato.com/de/face-of-happy-young-hispanic-hipster-man-showing-d-SPZ2W9G
35. https://elements.envato.com/de/happy-young-hispanic-hipster-man-thinking-while-us-VNMZ8AB
36. https://elements.envato.com/de/happy-young-hispanic-hipster-man-talking-and-showi-LPATJ9U
37. https://elements.envato.com/de/happy-young-hispanic-hipster-man-thinking-while-po-EXNQV8T
38. https://elements.envato.com/de/face-of-happy-young-hispanic-hipster-man-giving-th-FUV247Z
39. https://elements.envato.com/de/happy-young-hispanic-hipster-man-pointing-finger-a-ENPSGL2
40. https://elements.envato.com/de/happy-young-hispanic-hipster-man-getting-good-news-2EQGVSW
41. https://elements.envato.com/de/happy-young-hispanic-hipster-man-clapping-hands-6XH42VW
42. https://elements.envato.com/de/face-of-happy-young-hispanic-hipster-man-nodding-h-G3LJ2DE
43. https://elements.envato.com/de/happy-young-hispanic-hipster-man-giving-handshake-Y5PGNDC
44. https://elements.envato.com/de/happy-young-hispanic-hipster-man-taking-selfie-NU7HEZ2

45. https://elements.envato.com/de/happy-young-hispanic-hipster-man-clapping-hands-Z7CLE2J
46. https://elements.envato.com/de/happy-young-hispanic-hipster-man-smiling-with-arms-3KFAHEX
47. https://elements.envato.com/de/happy-young-hispanic-hipster-man-relaxing-with-eye-P2LSW35
48. https://elements.envato.com/de/happy-young-hispanic-hipster-man-giving-thumbs-up-L5HXABS
49. https://elements.envato.com/de/happy-young-hispanic-hipster-man-showing-something-JF7BKL8
50. https://elements.envato.com/de/happy-young-hispanic-hipster-man-looking-surprised-J43RYD2
51. https://elements.envato.com/de/happy-young-hispanic-hipster-man-pointing-at-camer-JM98TDL
52. https://elements.envato.com/de/happy-young-hispanic-hipster-man-vlogging-with-pho-MNRLVCZ
53. https://elements.envato.com/de/happy-young-hispanic-hipster-man-presenting-someth-VXZT5FK
54. https://elements.envato.com/de/happy-young-hispanic-hipster-man-pointing-up-M9J8LSQ
55. https://elements.envato.com/de/face-of-happy-young-hispanic-hipster-man-smiling-J43QMZW
56. https://elements.envato.com/de/face-of-happy-young-hispanic-hipster-man-thinking-AC9ZE5D
57. https://elements.envato.com/de/happy-young-hispanic-hipster-man-smiling-BW7XEQF
58. https://elements.envato.com/de/happy-young-hispanic-hipster-man-showing-something-VYACMFU
59. https://elements.envato.com/de/happy-young-hispanic-hipster-man-being-interviewed-H57NVXU

Deep Fake Beispiel 3 Trainingsdaten

1. https://elements.envato.com/de/young-happy-hispanic-man-thinking-while-pointing-u-EWVR5CU
2. https://elements.envato.com/de/head-shot-of-young-happy-hispanic-man-smiling-VRNJPS2
3. https://elements.envato.com/de/young-happy-hispanic-man-smiling-UGQZHMJ
4. https://elements.envato.com/de/young-happy-hispanic-man-clapping-hands-4UFVA86

5. https://elements.envato.com/de/young-happy-hispanic-man-pointing-at-camera-5QF6KXR
6. https://elements.envato.com/de/young-happy-hispanic-man-showing-something-63VJ4K2
7. https://elements.envato.com/de/young-happy-hispanic-man-catching-something-LKR793P
8. https://elements.envato.com/de/young-happy-hispanic-man-clapping-hands-BLHGQXS
9. https://elements.envato.com/de/young-happy-hispanic-man-showing-something-LCEJT36
10. https://elements.envato.com/de/young-happy-hispanic-man-with-finger-on-lips-2WXL64A
11. https://elements.envato.com/de/young-happy-hispanic-man-presenting-something-9E4KYT7
12. https://elements.envato.com/de/young-happy-hispanic-man-giving-handshake-Z3DXGKL
13. https://elements.envato.com/de/young-happy-hispanic-man-showing-something-6U2J38A
14. https://elements.envato.com/de/young-happy-hispanic-man-relaxing-with-eyes-closed-NRGT3ME
15. https://elements.envato.com/de/young-happy-hispanic-man-smiling-with-arms-crossed-AJXLFW6
16. https://elements.envato.com/de/young-happy-hispanic-man-giving-thumbs-up-6YHSJ79
17. https://elements.envato.com/de/young-happy-hispanic-man-giving-thumbs-up-T6FGUEN
18. https://elements.envato.com/de/young-happy-hispanic-man-relaxing-with-eyes-closed-CR49JA2
19. https://elements.envato.com/de/young-happy-hispanic-man-getting-good-news-KQKTT6M
20. https://elements.envato.com/de/young-happy-hispanic-man-smiling-while-pointing-up-AEMECN4
21. https://elements.envato.com/de/young-happy-hispanic-man-smiling-while-waving-hand-MWYF36S
22. https://elements.envato.com/de/young-happy-hispanic-man-smiling-while-pointing-up-AEMECN4
23. https://elements.envato.com/de/young-happy-hispanic-man-smiling-while-waving-hand-MWYF36S
24. https://elements.envato.com/de/young-happy-hispanic-man-looking-excited-and-givin-NMMRMHS

25. https://elements.envato.com/de/young-happy-hispanic-man-looking-excited-and-givin-M9ZRLHN
26. https://elements.envato.com/de/young-happy-hispanic-man-choosing-between-thumbs-u-A72QDYP
27. https://elements.envato.com/de/young-handsome-hispanic-man-pointing-up-LHMW4VZ
28. https://elements.envato.com/de/young-handsome-hispanic-man-comparing-something-QWGS4P5
29. https://elements.envato.com/de/young-handsome-hispanic-man-shrugging-shoulders-UAYXQQW
30. https://elements.envato.com/de/young-tired-hispanic-man-looking-bored-YC5C7S9
31. https://elements.envato.com/de/young-tired-hispanic-man-looking-bored-MPPQ7S5
32. https://elements.envato.com/de/young-stressed-hispanic-man-looking-shocked-NM9EXFG
33. https://elements.envato.com/de/young-stressed-hispanic-man-getting-bad-news-D9EMFNR
34. https://elements.envato.com/de/young-sad-hispanic-man-giving-thumbs-down-NUV2ZZZ
35. https://elements.envato.com/de/young-happy-hispanic-businessman-thinking-while-po-2TURGXZ
36. https://elements.envato.com/de/young-happy-hispanic-businessman-pointing-up-YN5KYM4
37. https://elements.envato.com/de/young-happy-hispanic-businessman-giving-handshake-7J5R2V2
38. https://elements.envato.com/de/young-happy-hispanic-businessman-waving-hand-Z7HGQRT
39. https://elements.envato.com/de/young-happy-hispanic-businessman-showing-something-2SNPYKR
40. https://elements.envato.com/de/young-happy-hispanic-businessman-taking-selfie-JNXLFKA
41. https://elements.envato.com/de/young-happy-hispanic-businessman-with-arms-crossed-63JVU2K
42. https://elements.envato.com/de/young-happy-hispanic-businessman-waving-hand-LNZQCNV
43. https://elements.envato.com/de/young-happy-hispanic-businessman-presenting-someth-GS9QMDS
44. https://elements.envato.com/de/young-happy-hispanic-businessman-showing-something-TBV7MN6

45. https://elements.envato.com/de/young-happy-hispanic-businessman-explaining-someth-FPJKM66
46. https://elements.envato.com/de/young-happy-hispanic-businessman-opening-gift-box-AD7B9JS
47. https://elements.envato.com/de/young-happy-hispanic-businessman-showing-blackboar-KALH9N3
48. https://elements.envato.com/de/young-handsome-hispanic-businessman-comparing-some-K6SZNU6
49. https://elements.envato.com/de/young-happy-hispanic-businessman-thinking-while-us-BMBTQNU
50. https://elements.envato.com/de/young-happy-hispanic-businessman-using-phone-DJ6CX9F
51. https://elements.envato.com/de/young-happy-hispanic-businessman-thinking-while-ho-STHN8E6
52. https://elements.envato.com/de/young-stressed-hispanic-businessman-looking-bored-TZUH87B
53. https://elements.envato.com/de/head-shot-of-young-sad-hispanic-businessman-crying-CGCK562
54. https://elements.envato.com/de/young-happy-hispanic-businessman-playing-games-and-42SW89V
55. https://elements.envato.com/de/young-happy-hispanic-businessman-using-digital-tab-C7U8VZK
56. https://elements.envato.com/de/young-happy-hispanic-businessman-showing-digital-t-KJ33PBK
57. https://elements.envato.com/de/young-happy-hispanic-businessman-showing-phone-and-LATDGE8
58. https://elements.envato.com/de/head-shot-of-young-happy-hispanic-businessman-givi-HRK6JGW
59. https://elements.envato.com/de/head-shot-of-young-happy-hispanic-businessman-usin-QV9R4KW

B. Quellliste der Zielvideos

Deep Fake Beispiel 1 Zielvideo: https://elements.envato.com/de/young-handsome-hispanic-man-shrugging-shoulders-UAYXQQW

Deep Fake Beispiel 2 Zielvideo: https://elements.envato.com/de/happy-young-handsome-man-being-interviewed-KF8WVCD

Deep Fake Beispiel 3 Zielvideo: https://elements.envato.com/de/happy-young-handsome-bearded-businessman-waving-ha-C9ENZAB

C. Quellliste der unveränderten Videos für die Probandenbefragung

Videobeispiel 1: https://elements.envato.com/de/young-stressed-bearded-indian-man-looking-upset-HEDMC6B

Videobeispiel 2: https://elements.envato.com/de/happy-young-businessman-pointing-up-4LMWJVH

Videobeispiel 3: https://elements.envato.com/de/portrait-of-happy-blonde-woman-smiling-NAHPMED

Videobeispiel 4: https://elements.envato.com/de/confident-guy-proud-of-himself-green-screen-PXUPC7E

Videobeispiel 5: https://elements.envato.com/de/young-woman-holding-up-two-fingers-V3JNB8L

Videobeispiel 6: https://elements.envato.com/de/portrait-of-smiling-girl-VCZ8UBY

Videobeispiel 7: https://elements.envato.com/de/happy-young-handsome-businessman-smiling-and-think-9HHC5XP