

Der Einsatz von Computer Vision-Methoden für Filme

Eine Fallanalyse für die Kriminalfilm-Reihe Tatort

Schmidt, Thomas

thomas.schmidt@ur.de

Lehrstuhl für Medieninformatik, Universität Regensburg

Kurek, Sarah

sarah.kurek@stud.uni-regensburg.de

Lehrstuhl für Medieninformatik, Universität Regensburg

Einleitung

Quantitative Methoden haben in den Filmwissenschaften eine lange Tradition, die bis auf die prädigitale Ära zurückreichen (Salt 1974; Vonderau 2020). Zahlreiche Projekte in den digitalen Filmwissenschaften setzen mittlerweile computergestützte Methoden ein, um quantitative Analysen durchzuführen oder qualitative Arbeiten zu unterstützen. Anwendungsbereiche sind unter anderem die Analyse von Farben (Burghardt et al. 2016; 2018; Kurzhals et al. 2016; Flueckiger 2017; Pause / Walkowski 2018; Masson et al. 2020);, Annotationsmöglichkeiten (Kuhn et al. 2015; Halter et al. 2019; Schmidt / Halbhuber 2020; Schmidt et al. 2020a) oder Schnittlängen und -typen (DeLong 2015; Baxter et al. 2017). Häufig werden dabei die Texte von Filmen (Skripte, Untertitel) analysiert (Hoyt et al. 2014; Holubut et al. 2016; Byszuk 2020; Holubut / Rybicki 2020). Durch Entwicklungen im Bereich der Computer Vision (computergestützte Bilderkennung/Bildanalyse) (CV) bieten sich jedoch neue Möglichkeiten für Digital Humanities (DH)-Projekte, die mit Videos arbeiten, die bereits erfolgreich für Filme, Internetvideos oder Theateraufführungen eingesetzt werden (Zaharieva et al. 2012; Howanitz et al. 2019; Pustu-Iren et al. 2020; Schmidt et al. 2021c; Schmidt / Wolff 2021). Wir präsentieren im folgenden Beitrag eine explorative Studie zum Einsatz einer Auswahl an CV-Methoden, die wir für potentiell wertvoll für den Bereich der Spielfilm-Analyse einschätzen. Wir orientieren uns dabei am explorativen Forschungsansatz definiert von Wulff (1998) für die Filmanalyse.

Als Fallstudie wird die deutschsprachige Kriminalfilm-Reihe „Tatort“ gewählt. Mit ca. 9 Millionen Zuschauern handelt es sich um eine der beliebtesten Fernsehformate in Deutschland.¹ Aufgrund seiner national hohen kulturellen Bedeutung ist der Tatort ein häufiger Untersuchungsgegenstand in der Filmanalyse (Buhl 2013) und wird zur Analyse gesellschaftspolitischer Themen wie Migration (Ortner 2007), Verhältnis von Ost- und Westdeutschland (Welke 2005), des Zusammenhangs von Emotionen und Geschlechtern (Finger et al. 2010) oder zur Analyse von Online-Texten herangezogen (Schmidt et al. 2021d). Der Fokus unserer Analysen liegt auf gruppenbasierten Vergleichen. Als Gruppen differenzieren wir zwischen unterschiedlichen Städten/ErmittlerInnen-Teams. So spielen die einzelnen Folgen des Tatorts in unterschiedlichen Städten mit unterschiedlichen Hauptfiguren. Diese Gruppen transportieren teilweise unterschiedliche

Stimmungen, Lokalkolorit sowie Geschlechts- und Altersrepräsentationen in den Figurenkonstellationen.

Die Ziele dieses Beitrags sind (1) Nutzen und Limitationen der angewandten Methoden zu analysieren und (2) explorativ festzustellen, ob die Methoden besondere Charakteristiken von, in diesem Fall, Filmgruppen aufzuzeigen. Als CV-Methoden werden Objekt-, Emotions-, Geschlechts-, Alters- und Ortserkennung untersucht und frei verfügbare state-of-the-art-Modelle verwendet. Die genannten Methoden wurden ausgewählt, weil sie als gewinnbringend für Forschungsideen auf dem vorliegenden Korpus angesehen werden und bereits in ähnlichen Settings exploriert wurden (Schmidt et al. 2021c).

Korpus

Als Korpus werden 13 Folgen des Tatorts genutzt. Alle Filme (je ca. 90 Minuten) liegen im mp4-Format mit einer Auflösung von 960x540 Pixeln und 25 Frames pro Sekunde vor. Alle angewandten CV-Methoden nutzen Bild-Dateien weswegen wir 1 Frame für jede Sekunde eines Films extrahieren und als Korpusgrundlage verwenden. Abbildung 1 fasst die wichtigsten Metadaten der Filme zusammen.

ID	Folge	Titel	Ausstrahlungsdatum	Extrahierte Frames
N1	1085	Ein Tag wie jeder andere	24.02.19	5 255
SW1	1087	Für immer und dich	10.03.19	5 368
M1	1096	Die ewige Welle	26.05.19	5 342
CH 1	1099	Ausgezählt	16.06.19	5 306
CH 2	1106	Der Elefant im Raum	27.10.19	5 174
M2	1114	One Way Ticket	26.12.19	5 320
M3	1118	Unklare Lage	26.01.20	5 340
SW2	1121	Ich hab im Traum geweinet	23.02.20	5 373
N2	1122	Die Nacht gehört dir	01.03.20	5 267
M4	1135	Lass den Mond am Himmel stehn	07.06.20	5 249
SW3	1138	Rebland	27.09.20	5 344
M5	1146	In der Familie (Teil 1)	29.11.20	5 338
M6	1147	In der Familie (Teil 2)	06.12.20	5 325

Abb. 1: Metadaten des ausgewählten Tatort-Korpus.

Wir differenzieren zwischen den folgenden Standorten/ErmittlerInnen-Teams, die im Folgenden für gruppenbasierte Vergleiche genutzt werden: Luzern in der Schweiz (im Folgenden abgekürzt als CH; insgesamt 2 Filme), München (M; 6 Filme), Nürnberg (N; 2 Filme) und Schwarzwald (SW; 3 Filme). Die 4 Gruppen unterscheiden sich bezüglich des Kolorits (ländlich vs städtisch) und in der Alters- und Geschlechtsausprägung. Aufgrund der ungleichen Menge an Filmen pro Gruppe werden im Folgenden primär Werte normalisiert an der Länge (pro Sekunde, gewählte Frames für das Korpus) betrachtet.

Objekterkennung

Für die Objekterkennung wird Detectron2 von Facebook AI Research verwendet, was als state-of-the-art-Lösung gilt (Wu et al. 2019). Das Modell basiert auf einem vortrainierten maskierten RCCN-Modell und wurde auf dem COCO-Datensatz (Lin et al. 2015) trainiert. Es kann 80 Objektklassen wie Fahrzeuge oder

Tiere vorhersagen. Als Schwellenwert für die Erkennung wird eine Wahrscheinlichkeit von 50% gewählt. Dies ist ein eher niedriger Wert, ermöglicht uns aber breitere Explorationen der Methode. Diese und alle anderen Methoden wurden in *Python* mit verschiedenen SDKs implementiert.

Die Abbildungen 2-3 illustrieren die je 10 häufigsten Objekte pro Tatort-Gruppe und insgesamt. Wir interpretieren die Ergebnisse rein deskriptiv.

CH			M			N		
Objekt	#	%	Objekt	#	%	Objekt	#	%
Person	21 295	91,53	Person	61 944	90,26	Person	21 000	89,26
Stuhl	2 258	14,26	Stuhl	8 671	13,37	Stuhl	4 493	18,26
Tasse	1 640	10,24	Auto	5 369	7,42	Buch	3 516	7,86
Krawatte	1 439	10,78	Buch	4 296	4,69	Auto	3 001	9,48
Buch	1 269	4,83	Flasche	3 692	5,28	Tasse	1 241	7,70
Auto	1 188	6,16	Tasse	3 652	6,87	Krawatte	1 007	8,19
Laptop	718	5,62	Esstisch	2 511	6,10	Esstisch	986	7,13
Fernseher	714	6,14	Fernseher	2 265	5,38	Mobiltelefon	759	5,89
Esstisch	665	5,44	Topfpflanze	1 841	4,20	Flasche	636	3,58
Flasche	629	3,30	Krawatte	1 696	4,84	Fernseher	509	4,56

Abb. 2: Verteilung von erkannten Objekten in der CH, M und N-Gruppe. # ist die absolute Zahl. % der Anteil an den gewählten Frames des jeweiligen Gruppenkorpus.

SW			Gesamt		
Objekt	#	%	Objekt	#	%
Person	21 295	91,53	Person	138 354	91,36
Stuhl	2 258	14,26	Stuhl	22 764	16,28
Tasse	1 640	10,24	Buch	17 194	6,54
Krawatte	1 439	10,78	Auto	12 067	7,76
Buch	1 269	4,83	Tasse	9 465	8,52
Auto	1 188	6,16	Flasche	7 867	5,68
Laptop	718	5,62	Esstisch	6 092	6,67
Fernseher	714	6,14	Krawatte	4 768	5,93
Esstisch	665	5,44	Topfpflanze	4 421	4,57
Flasche	629	3,30	Fernseher	4 420	5,25

Abb. 3: Verteilung von erkannten Objekten in der SW-Gruppe und insgesamt. # ist die absolute Zahl. % der Anteil an den gewählten Frames des jeweiligen Gruppenkorpus.

Zu den häufigsten erkannten „Objekten“ gehören (trivialerweise) Personen. Ansonsten wird vor allem Innenarchitektur (Stühle, Tassen, Bücher) erkannt aber auch Bildschirme wie Laptops und Fernseher. Diese Erkennungen basieren vor allem auf Szenen der Recherchen der einzelnen Teams; Bildschirme werden dabei häufig als Fernseher erkannt (siehe Abbildung 4).

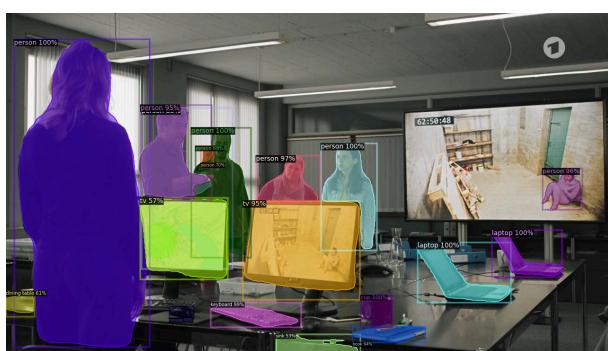


Abb. 4: Frame mit als Fernseher erkannten Computerbildschirmen (CH1).

Verhältnismäßig wenig Objekte weisen auf Außenszenen hin wie zum Beispiel Autos (Abbildung 5).



Abb. 5: Frame mit als Auto erkannten Objekten (M5)

Die Unterschiede der einzelnen Tatort-Städte unseres Korpus sind eher gering und die Verteilung sehr homogen. Vereinzelt können Objekthäufungen aufgrund von sehr speziellen inhaltlichen Unterschieden einzelner Folgen festgestellt werden. Die SW-Episoden beispielsweise weisen, im Unterschied zu den anderen Gruppen, vermehrt Betten auf, da eine der Folgen (SW2) in einem Hotel spielt (siehe Abbildung 6).

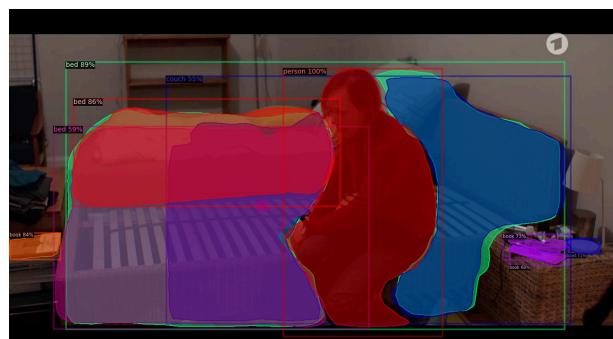


Abb. 6: Frame mit als Bett erkannten Objekten (SW2).

Es wurde keine systematische Evaluation durchgeführt, jedoch konnte man in der explorativen Analyse feststellen, dass die meisten Ergebnisse korrekt sind. Falsche Zuweisungen sind jedoch nicht selten, wenngleich die Fehlinterpretation häufig nachvollziehbar ist (siehe Abbildung 7).

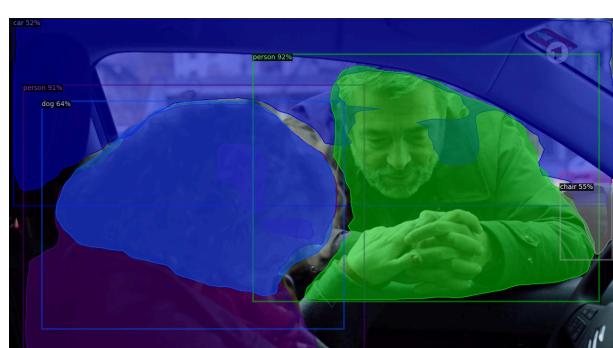


Abb. 7: Als „Hund“ erkannte Haare einer Figur (CH2).

Figurenanalyse

Unter Figurenanalyse bezeichnen wir im Folgenden alle Methoden, die die Gesichter der Figuren als Analyseelement benutzen: Emotions-, Alters- und Geschlechtserkennung.

Emotionserkennung

Gesichtsbasierte Emotionserkennung ist eine etablierte Methode in der Mensch-Maschine-Interaktion mit zahlreichen Anwendungsbeispielen (Halbhüber et al. 2019; Hartl et al. 2019; Schmidt et al. 2020c). Für die Emotionserkennung wird das Python-Modul *FER* (Goodfellow et al. 2013) genutzt. Das Modell führt erste eine Gesichtserkennung durch (Zhang et al. 2016) und dann eine Emotionsprädiktion über ein convolutional neural network (CNN), das auf über 35 000 vorannotierten Bildern trainiert wurde. Das Modell gibt Wahrscheinlichkeitswerte für die sieben Klassen *Wut*, *Ekel*, *Furcht*, *Freude*, *Trauer*, *Überraschung* und *Neutral* zwischen 0 und 1 aus, die sich insgesamt zu 1 summieren. Zur Bestimmung der Gesamtemotion eines Frames werden die jeweiligen Werte für die Kategorien summiert und der Durchschnitt gebildet. Für die film- oder gruppenbasierten Analysen werden Mittelwerte über alle Frames hinweg gebildet.

Abbildung 8 illustriert die wichtigsten statistischen Werte dieser Auswertung.

Emotion	Wert	CH	M	N	SW	Gesamt
Wut	M	0,17	0,19	0,18	0,19	0,18
	Max	0,95	0,97	0,96	0,97	0,97
Ekel	M	0,01	0,01	0,01	0,01	0,01
	Max	0,73	0,76	0,77	0,87	0,87
Furcht	M	0,10	0,10	0,10	0,10	0,10
	Max	0,78	0,86	0,82	0,84	0,86
Freude	M	0,11	0,11	0,12	0,16	0,12
	Max	1,00	1,00	1,00	1,00	1,00
Neutral	M	0,29	0,25	0,25	0,21	0,25
	Max	0,99	0,98	0,97	0,99	0,99
Trauer	M	0,30	0,31	0,31	0,31	0,31
	Max	0,96	0,99	0,97	0,97	0,99
Überraschung	M	0,03	0,04	0,03	0,03	0,03
	Max	0,99	1,00	0,90	0,99	0,06

Abb. 8: Deskriptive Statistik für die Emotionserkennung. Minimalwerte sind stets 0. Höchster M (Durchschnitt) pro Emotion für die Gruppen ist hervorgehoben.

Die häufigsten Emotionsklassen sind Trauer ($M=0,31$) (siehe Abbildung 9), Neutral ($M=0,25$) und Wut ($M=0,18$). Dies ist eine passende Verteilung für die Grundstimmung von Kriminalfilmen. Überraschung und Ekel werden eher selten vorhergesagt. Die Tendenz zu negativem Sentiment und Emotionen findet man bei der Annotation und Prädiktion von anderen narrativen Erzählformen auch (Schmidt / Burghardt 2018; Schmidt 2019; Schmidt et al. 2019; 2021a; 2021b)

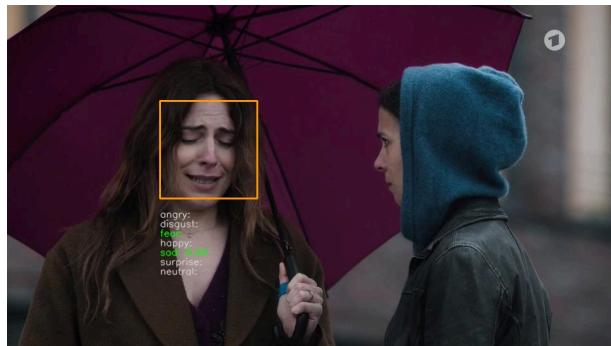


Abb. 9: Frame mit dem höchsten Wert für Trauer (0,99) im Gesamtkorpus (M5)

Deskriktiv betrachtet sind die Ergebnisse der einzelnen Episodengruppen erneut sehr homogen. Für Output mit kontinuierlichen Werten überprüfen wir die Unterschiede aber noch mittels Signifikanztests. Wir verwenden einen Welch-ANOVA-Test (alle Voraussetzungen für den Test sind erfüllt (Field 2009)) und finden signifikante Unterschiede gemäß eines Signifikanzniveaus von $p<0,05$ (Abbildung 10).

	p-Wert	F-Wert	η^2
Wut	<0,01	13,02	<0,01
Ekel	<0,01	13,12	<0,01
Angst	<0,01	25,06	<0,01
Freude	<0,01	93,52	0,01
Neutral	<0,01	127,40	0,01
Trauer	<0,01	4,31	<0,01
Überraschung	<0,01	35,43	<0,01

Abb. 10: Ergebnisse des Welch-ANOVA-Signifikanztest für die Episodengruppen (Emotionen).

Die Effekte der Unterschiede bestätigen jedoch die deskriptive Interpretation, da sie laut Cohen (1988) als sehr gering einzustufen sind ($\eta^2 < 0,01$ = schwächer, $< 0,06$ moderater und $< 0,14$ = starker Effekt). Auch Post-Hoc-Tests unter den einzelnen Gruppen weisen zwar signifikante Unterschiede auf, sind jedoch geringfügig.

Die explorative Evaluation des Materials zeigt, dass die Modelle für extreme Emotionsausprägungen nachvollziehbare Ergebnisse produzieren (siehe auch Abbildung 11), ein Hauptproblem jedoch ist, dass Fehler in der vorangestellten Gesichtserkennung vorkommen. So hat das Modell große Probleme mit der Erkennung von Gesichtern, die nicht frontal in die Kamera blicken. Dies liegt der Tatsache zu Grunde, dass die Modelle primär mit Bildern trainiert werden bei denen die Personen frontal in die Kamera blicken (Goodfellow et al. 2013).

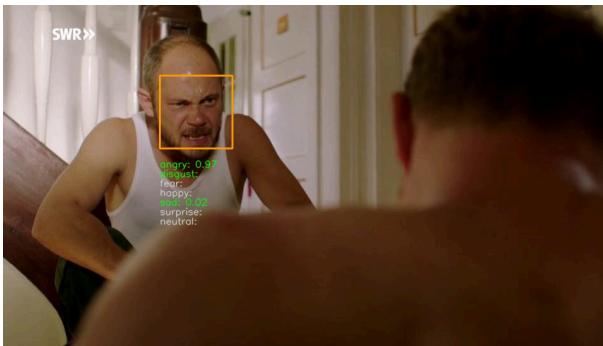


Abb. 11: Frame mit dem höchsten Wert für Wut (0,97) im Gesamtkorpus (SW1).



Abb. 14. Frame mit der Person mit dem geringsten Alterswert von 10,86 (SW 3).

Alters- und Geschlechtserkennung

Die Vorhersage des Alters und des Geschlechts von Figuren wird mit dem Modul *py-agender*² durchgeführt. Es basiert auf einem CNN, das auf dem IMDB-Wiki-Datensatz, bestehend aus über 500 000 annotierten Gesichtern (Rothe et al. 2018), trainiert wurde und erzielt in Evaluationen sehr gute Ergebnisse (Agustsson et al. 2017). Die Altersprädiktion gibt einen Wert zwischen 0 und 100 aus, der das Alter kennzeichnet. Die Geschlechtsprädiktion einen Wert zwischen 0 und 1 für den gilt, <0.5 eher männlich und >0.5 eher weiblich. Für beide Verfahren wurde für jeden Frame der Mittelwert aller erkannten Gesichter für einen Gesamtwert gebildet. In Abbildung 12 werden die Ergebnisse für beide Methoden gesamt und pro Tatortgruppe zusammengefasst.

	Wert	CH	M	N	SW	Gesamt
Alter	M	42,98	42,10	41,15	39,08	41,47
	Max	71,87	75,46	69,50	73,11	75,46
Geschlecht	M	0,38	0,34	0,41	0,44	0,38
	Max	0,99	0,99	0,99	0,99	0,99

Abb. 12: Deskriptive Statistik für die Alters- und Geschlechtserkennung. Maximal- und Minimalwerte von M werden pro Episodengruppe hervorgehoben.

Gemäß der Alterserkennung liegt der Altersdurchschnitt bei 41,47 Jahren. Die dominanten und häufig in Frames gezeigten ErmittlerInnen der ausgewählten Folgen sind jedoch überwiegend Ende 40 und Anfang 50. Die älteste Person im Gesamtkorpus wird mit 72 Jahren identifiziert (Abbildung 13), die jüngste ist ein Kind mit 10 Jahren (Abbildung 14).

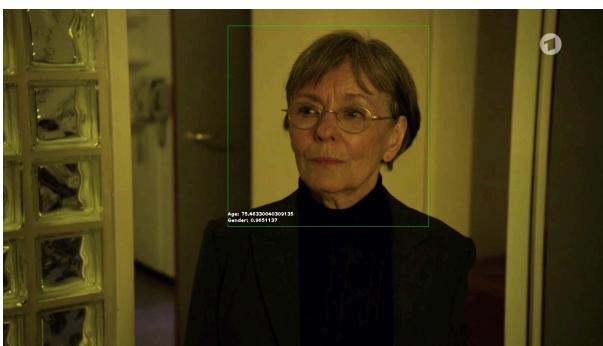


Abb. 13: Frame mit der Person mit dem höchsten Alterswert von 75,46 (M2).

Rein deskriptiv sind die Unterschiede zwischen den Gruppen gering. Ein Welch-ANOVA-Test weist erneut auf signifikante Unterschiede mit einem geringen Effekt hin (Abbildung 15).

	p-Wert	F-Wert	η^2
Alter	<0,001	358,43	0,02
Geschlecht	<0,001	124,69	0,02

Abb. 15: Ergebnisse des Welch-ANOVA-Signifikanztest für die Episodengruppen (Geschlecht/Alter).

Tatsächlich zeigen Post-Hoc-Tests, dass die Hauptunterschiede mit einem mittleren Effekt zwischen den Folgen aus der Schweiz (CH) und aus dem Schwarzwald (SW) bestehen, welche gleichzeitig den höchsten, respektive geringsten Altersunterschied haben. Bei Betrachtung der Filme wird klar, dass dies vor allen daran liegt, dass in den SW-Folgen viele Kinder und Jugendliche mitspielen (Abbildung 14). Ein Problem der Altersanalyse, das wir bei unseren Explorationen identifizieren konnten, ist jedoch in diesem Zusammenhang, dass Kinder und Jugendliche meist überschätzt werden (Abbildung 15). Grund hierfür ist auch wieder die Trainingsgrundlage des Modells, die primär aus Personen im Erwachsenenalter besteht.

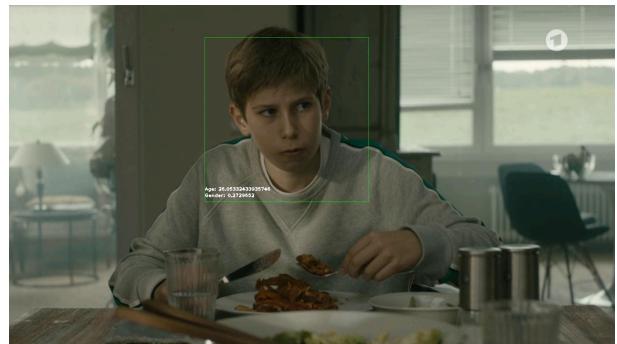


Abb. 16: Frame mit Kind, dessen Alter auf 26 Jahre überschätzt wird (M4)

Mit einem Mittelwert von 0,38 ist eine vermehrte Repräsentation männlicher Gesichter festzustellen (Abbildung 12). Dies entspricht auch der realen Figuren-Belegung der Serie, die, obwohl sie gemischte ErmittlerInnen-Paare enthält, vor allem in den Nebenfiguren von männlichen Charakteren dominiert wird. Abbildung 17 und 18 zeigen die jeweils höchsten Ausprägungen des Körpers.



Abb. 17: Frame mit dem „männlichsten“ Gesicht (Minimalwert für Geschlecht: 0,002) (M2).

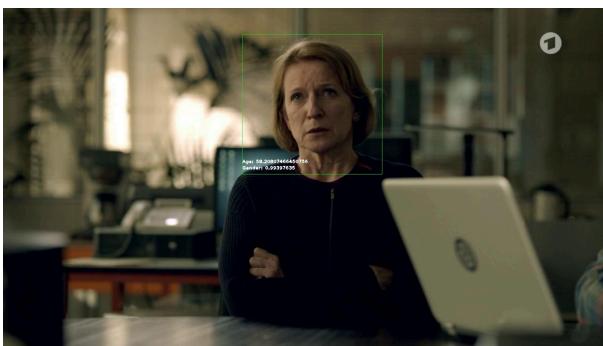


Abb. 18: Frame mit dem „weiblichsten“ Gesicht (Minimalwert für Geschlecht: 0,99) (N1).

Ein Welch-ANOVA-Test zeigt wiederum signifikante Werte mit schwachen Effekten auf (Abbildung 15). Post-Hoc-Tests zeigen, dass der signifikante Unterschied aufgrund der stärkeren Differenzen der Episoden aus München (M) mit den Episoden aus Schwarzwald (SW) und Luzern (CH) zustande kommt. In der Tat sind die beiden letztgenannten Gruppen jene, die ErmittlerInnen-Gruppen bestehend aus Mann und Frau haben und damit höhere Werte Richtung weiblicher Gesichter zeigen. Der erhöhte Wert bezüglich männlicher Ausprägung beim Münchener-Tatort ist zum einen konform mit der Dominanz an männlichen Ermittlern und wird bei der Einzelfolgen-Analyse deutlich, da eine Folge in einem Männergefängnis spielt.

Insgesamt wirkt die Geschlechtsprädiktion plausibel. Ähnlich zur Emotionserkennung ist ein Problem mangelnde korrekte Gesichtserkennung aufgrund nicht-frontaler Kamerawinkel und schwammige, dunkle Einstellungen (Abbildung 19).



Abb. 19: Beispiel für falsche Geschlechtserkennung: Das Gesicht wird als männlich (0,29) identifiziert (M6).

Ortserkennung

Als Ortserkennung bezeichnen wir die Methodik den groben Schauplatz eines Bildes zu erkennen. Dabei ist nicht der geografische Ort gemeint, sondern die abstrakte Umgebung, also zum Beispiel, ob ein Bild in einem Zimmer spielt oder draußen. Wir verwenden den Trainingsdatensatz *Places365*³, der aus über 1,8 Millionen annotierten Bildern besteht. Für unsere Prädiktion nutzen wir ein vorgebildetes CNN und präparieren die Frames in einer Vorverarbeitung für das CNN (Zhou et al. 2017). Anstatt den 365 Teilklassen fokussieren wir uns jedoch auf die Hauptkategorien *innen*, *draußen-künstlich*, *draußen-natürlich* und *draußen-gemischt*. Jeden Frame weisen wir die Kategorie zu, die das Modell mit der höchsten Wahrscheinlichkeit vorhersagt.

	Wert	CH	M	N	SW	Gesamt
<i>innen</i>	#	9 363	26 403	9 247	12 765	57 778
	%	89,34	82,73	87,88	79,36	83,74
<i>draußen-künstlich</i>	#	809	3 650	982	2 014	7 455
	%	7,72	11,44	9,33	12,52	10,80
<i>draußen-natürlich</i>	#	207	1 265	189	735	1 372
	%	1,98	3,96	1,80	4,57	3,47
<i>draußen-gemischt</i>	#	101	596	104	571	2 396
	%	0,96	1,87	0,99	3,55	1,99

Abb. 20: Häufigkeitsverteilung für die Ortserkennung. # ist die absolute Zahl. % der Anteil an den gewählten Frames des jeweiligen Gruppenkorpus.

Unabhängig von der Episodengruppe wird der größte Anteil der Frames als innen kategorisiert (Abbildung 20), was der Realität der Filme entspricht in denen meist Ermittlungen und Recherchen in Zimmern stattfinden (Abbildung 21).



Abb. 21: Beispiel für Frame, das als „innen“ erkannt wurde (CH1).

Ein Chi-Quadrat-Signifikanz-Test weist dennoch auf signifikante Unterschiede zwischen den Gruppen hin ($\chi^2 = 809,23$; $p < 0,001$; $\varphi = 0,06$). In der Tat weisen die Episoden aus dem Schwarzwald als einer eher ländlichen Gegend den höchsten Anteil an Frames der Kategorie „draußen“ auf (Abbildung 22). Die CH-Gruppe, die von uns auch als eher ländlich postuliert wurde, kann dies hier aber nicht bestätigen, was jedoch inhaltlich plausibel ist, da beispielsweise eine Folge komplett in den Räumen eines Schiffes spielt.



Abb. 22: Beispiel für einen Frame, der als gemischt-draußen erkannt wurde (SW1).

Methodenreflexion und Ausblick

Durch die Durchführung der hier vorgestellten Fallstudie, konnten wir die CV-Methoden gewinnbringend explorieren. Wir konnten signifikante Unterschiede zwischen Episodengruppen identifizieren, die jedoch meist geringe Effekte aufzeigten. Die meisten Charakteristika, die wir so herausarbeiten konnten, bestätigten Annahmen. Neue Auffälligkeiten der Filmgruppen konnten jedoch nicht entdeckt werden. Die Gründe hierfür sind vielseitig: Das Korpus ist, da es sich um die gleiche Serie nur mit variierenden Figuren handelt, bezüglich des Genres, des Settings und der Besetzung eventuell zu homogen um gruppenbasierte Unterschiede in Signifikanztests deutlich zu machen. Vergleiche von Filmgruppen, die sich klarer voneinander unterscheiden (z.B. unterschiedliche Filmgenres), könnten deutlichere Effekte generieren.

Trotzdem haben wir vielversprechende Forschungsideen, die mit den präsentierten Methoden in einer Art *Distant Viewing*-Ansatz (Arnold / Tilton 2019) mit ausreichend Filmmaterial untersucht werden können, z.B. für den Bereich Gender Studies Korrelationen zwischen Emotionen und Geschlechtern oder Repräsentationsanalysen der Geschlechter (ähnlich zu Schmidt et al. 2020b). Die momentane Methodenauswahl ist auch noch rein bildfokussiert, wenngleich andere Kanäle z.B. der Audio-Kanal auch Potential für die Analyse haben. In der Tat werden in ersten Projekten in den DH der Einsatz von multimodalen Methoden oder dem Audio-Kanal bereits untersucht (Orloff et al. 2019; Schmidt et al. 2019; Schmidt / Wolff 2021)

Die Exploration der Methoden für die vorliegende Fallstudie haben jedoch auch Probleme in der Exaktheit und Leistung offenbart, z.B. Probleme in der Gesichtserkennung. Systematische Evaluationen sind notwendig, um das Ausmaß der Problematik einschätzen zu können. Auch sind die Klassifikationstaxonomien, beispielsweise der Objekterkennung und Ortserkennung, eventuell nicht passend für die Interessen von FilmwissenschaftlerInnen. Wir planen momentan größere Annotationsstudien, um (1) die Leistung von state-of-the-art-Standard-Modellen exakt zu evaluieren und (2) Trainingsmaterial für die Domänenadaption an eine spezielle Filmdomäne zu erstellen. Für die Annotationsstudien sollen studentische Hilfskräfte größere Mengen eines Querschnitts von Filmframes aus Filmen unterschiedlicher Epochen und Genres annotieren.

Fußnoten

1. <https://de.statista.com/statistik/daten/studie/377327/umfrage/fernsehzuschauer-der-krimireihe-tatort/>
2. <https://pypi.org/project/py-agender/>
3. <https://github.com/CSAILVision/places365>

Bibliographie

- Agustsson, Eirikur / Timofte, Radu / Escalera, Sergio / Baro, Xavier / Guyon, Isabelle / Rothe, Rasmus** (2017): "Apparent and Real Age Estimation in Still Images with Deep Residual Regressors on Appa-Real Database", in: *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 87–94. DOI: 10.1109/FG.2017.20
- Arnold, Taylor / Tilton, Lauren** (2019): "Distant viewing: Analyzing large visual corpora", in: *Digital Scholarship in the Humanities*. DOI: 10.1093/ds/fqz013
- Baxter, Mike / Khitrova, Daria / Tsivian, Yuri** (2017): "Exploring cutting structure in film, with applications to the films of D. W. Griffith, Mack Sennett, and Charlie Chaplin", in: *Digital Scholarship in the Humanities*, 32(1):1–16. DOI: 10.1093/llc/fqv035
- Buhl, Hendrik** (2013): "Tatort: gesellschaftspolitische Themen in der Krimireihe", in: *Alltag, Medien und Kultur*. Band 14. UVK, Konstanz.
- Burghardt, Manuel / Kao, Michael / Walkowski, Niels-Oliver** (2018): "Scalable MovieBarcodes—An Exploratory Interface for the Analysis of Movies.", in: *IEEE VIS Workshop on Visualization for the Digital Humanities* (Vol. 2).
- Burghardt, Manuel / Kao, Michael / Wolff, Christian** (2016): "Beyond Shot Lengths – Using Language Data and Color Information as Additional Parameters for Quantitative Movie Analysis", in: *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków: 753–755.
- Byszuk, Joanna** (2020): "The Voices of Doctor Who – How Styliometry Can be Useful in Revealing New Information About TV Series", in: *Digital Humanities Quarterly*, 014(4).
- Cohen, Jacob** (1988): *Statistical power analysis for the behavioral sciences*. Academic press.
- DeLong, Jordan** (2015): "Horseshoes, handgrenades, and model fitting: The lognormal distribution is a pretty good model for shot-length distribution of Hollywood films", in: *Literary and Linguistic Computing*, 30(1):129–136. DOI: 10.1093/llc/fqt030
- Field, Andy P.** (2009): *Discovering statistics using SPSS: And sex, drugs and rock „n“ roll* (3rd ed). SAGE Publications.
- Finger, Juliane / Unz, Dagmar C. / Schwab, Frank** (2010): "Crime Scene Investigation: The Chief Inspectors' Display Rules", in: *Sex Roles*, 62(11-12):798–809.
- Flueckiger, Barbara** (2017): "A Digital Humanities Approach to Film Colors", in: *The Moving Image: The Journal of the Association of Moving Image Archivists*, 17(2): 71–94. JSTOR. DOI: 10.5749/movingimage.17.2.0071
- Goodfellow, Ian J. et al.** (2013): *Challenges in Representation Learning: A report on three machine learning contests*. arXiv:1307.0414 [cs, stat]. <<http://arxiv.org/abs/1307.0414>> [14.06.2021]
- Halbhuber, David / Fehle, Jakob / Kalus, Alexander / Seitz, Konstantin / Kocur, Martin / Schmidt, Thomas / Wolff, Christian** (2019): "The Mood Game - How to use the player's affective state in a shoot'em up avoiding frustration and boredom", in: Alt, Florian / Bulling, Andreas / Döring, Tanja (eds.), *Mensch*

- und Computer 2019 - Tagungsband*. New York: ACM. DOI: 10.1145/3340764.3345369
- Halter, Gaudenz / Ballester-Ripoll, Rafael / Flueckiger, Barbara / Pajarola, Renato** (2019): "VIAN: A Visual Annotation Tool for Film Analysis", in: *Computer Graphics Forum*, 38(3): 119–129. DOI: 10.1111/cgf.13676
- Hartl, Philipp / Fischer, Thomas / Hilzenthaler, Andreas / Kocur, Martin / Schmidt, Thomas** (2019): "AudienceAR - Utilising Augmented Reality and Emotion Tracking to Address Fear of Speech", in: Alt, Florian / Bulling, Andreas / Döring, Tanja (eds.), *Mensch und Computer 2019 - Tagungsband*. New York: ACM. DOI: 10.1145/3340764.3345380
- Holobut, Agata / Rybicki, Jan / Wozniak, Monika** (2016): "Stylometry on the Silver Screen: Authorial and Translatorial Signals in Film Dialogue", in: *Book of Abstracts of the International Digital Humanities Conference (DH) (2016)*.
- Holobut, Agata / Rybicki, Jan** (2020): "The Stylometry of Film Dialogue: Pros and Pitfalls", in: *Digital Humanities Quarterly*, 014(4).
- Howanitz, Gernot / Bermeitinger, Bernhard / Radisch, Erik / Sebastian Gassner / Rehbein, Malte / Handschuh, Siegfried** (2019): "Deep Watching - Towards New Methods of Analyzing Visual Media in Cultural Studies", in: *Book of Abstracts of the International Digital Humanities Conference (DH) (2019)*.
- Hoyt, Eric / Ponto, Kevin / Roy, Carrie** (2014): "Visualizing and Analyzing the Hollywood Screenplay with ScripThreads", in: *Digital Humanities Quarterly*, 008(4).
- Kuhn, Virginia / Craig, Alan / Simeone, Michael / Satheesan, Simeone P. / Marini, Luigi** (2015): "The VAT: Enhanced video analysis", in: *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, 1–4. DOI: 10.1145/2792745.2792756
- Kurzhals, Kuno / John, Markus / Heimerl, Florian / Kuznevov, Paul / Weiskopf, Daniel** (2016): „Visual Movie Analytics”, in: *IEEE Transactions on Multimedia*, 18(11): 2149–2160. DOI: 10.1109/TMM.2016.2614184
- Lin, Tsung-Yi / Maire, Michael / Belongie, Serge / Bourdev, Lubomir / Girshick, Ross / Hays, James / Perona, Pietro / Ramanan, Deva / Zitnick, C. Lawrence / Dollár, Piotr** (2015): "Microsoft COCO: Common Objects in Context". arXiv:1405.0312 [cs]. <<http://arxiv.org/abs/1405.0312>> [14.06.2021]
- Masson, Eef / Olesen, Christian G. / Noord, Nanne van / Fossati, Giovanna** (2020): "Exploring Digitised Moving Image Collections: The SEMIA Project, Visual Analysis and the Turn to Abstraction.", in: *Digital Humanities Quarterly*, 014(4).
- Ortloff, Anna-Marie / Güntner, Lydia / Windl, Maximiliane / Schmidt, Thomas / Kocur, Martin / Wolff, Christian** (2019): "SentiBooks: Enhancing Audiobooks via Affective Computing and Smart Light Bulbs", in: Alt, Florian / Bulling, Andreas / Döring, Tanja (eds.), *Mensch und Computer 2019 - Tagungsband*. New York: ACM. DOI: 10.1145/3340764.3345369
- Ortner, Christina** (2007): "Tatort: Migration. Das Thema Einwanderung in der Krimireihe Tatort", in: *Medien & Kommunikationswissenschaft*, 55(1):5–23.
- Pause, Johannes / Walkowski, Niels-Oliver** (2018): "Everything is illuminated. Zur numerischen Analyse von Farbigkeit in Filmen", in: *Zeitschrift für digitale Geisteswissenschaften*.
- Pustu-Iren, Kader / Sittel, Julian / Mauer, Roman / Bulgakowa, Oksana / Ewerth, Ralph** (2020): "Automated Visual Content Analysis for Film Studies: Current Status and Challenges", in: *Digital Humanities Quarterly*, 014(4).
- Rothe, Rasmus / Timofte, Radu / Van Gool, Luc** (2018): "Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks", in: *International Journal of Computer Vision*, 126(2):144–157. DOI: 10.1007/s11263-016-0940-3
- Salt, Barry** (1974): "Statistical style analysis of motion pictures.", in: *Film Quarterly*, 28(1): 13-22.
- Schmidt, Thomas** (2019): "Distant Reading Sentiments and Emotions in Historic German Plays", in: *Abstract Booklet, DH_Budapest_2019*. Budapest, Hungary, 57-60.
- Schmidt, Thomas / Burghardt, Manuel** (2018): "An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing", in: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Santa Fe, New Mexico: Association for Computational Linguistics, 139-149.
- Schmidt, Thomas / Halbhuber, David** (2020): "Live Sentiment Annotation of Movies via Arduino and a Slider", in: *Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)*. Late Breaking Poster.
- Schmidt, Thomas / Wolff, Christian** (2021): "Exploring Multimodal Sentiment Analysis in Plays: A Case Study for a Theater Recording of Emilia Galotti", in: *Proceedings of the Conference on Computational Humanities Research 2021 (CHR 2021)*. Amsterdam, the Netherlands.
- Schmidt, Thomas / Burghardt, Manuel / Wolff, Christian** (2019): "Towards Multimodal Sentiment Analysis of Historic Plays: A Case Study with Text and Audio for Lessing's Emilia Galotti", in: *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference (DHN 2019)*. Copenhagen, Denmark, 405-414.
- Schmidt, Thomas / Engl, Isabella / Halbhuber, David / Wolff, Christian** (2020a): "Comparing Live Sentiment Annotation of Movies via Arduino and a Slider with Textual Annotation of Subtitles", in: *Post-Proceedings of the 5th Conference Digital Humanities in the Nordic Countries (DHN 2020)*, 212-223.
- Schmidt, Thomas / Engl, Isabella / Herzog, Juliane / Judisch, Lisa** (2020b): "Towards an Analysis of Gender in Video Game Culture: Exploring Gender-specific Vocabulary in Video Game Magazines", in: *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)*. Riga, Latvia.
- Schmidt, Thomas / Schindlwein, Miriam / Lichtner, Katharina / Wolff, Christian** (2020c): "Investigating the Relationship Between Emotion Recognition Software and Usability Metrics", in: *i-com*, 19(2): 139-151. DOI: 10.1515/icon-2020-0009
- Schmidt, Thomas / Dennerlein, Katrin / Wolff, Christian** (2021a): "Emotion Classification in German Plays with Transformer-based Language Models Pretrained on Historical and Contemporary Language", in: *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 67-79.
- Schmidt, Thomas / Dennerlein, Katrin / Wolff, Christian** (2021b): "Using Deep Neural Networks for Emotion Analysis of 18th and 19th century German Plays", in: *Fabrikation von Erkenntnis: Experimente in den Digital Humanities* (vDHd Sonderband). Melusina Press. DOI:10.26298/melusina.8f8w-y749-udlf
- Schmidt, Thomas / El-Keilany, Alina / Eger, Johannes / Kurek, Sarah** (2021c): "Exploring Computer Vision for Film Analysis: A Case Study for Five Canonical Movies", in: *2nd International Conference of the European Association for Digital Humanities (EADH 2021)*. Krasnoyarsk, Russia.
- Schmidt, Thomas / Grünler, Johanna / Schönwerth, Nicole / Wolff, Christian** (2021d): "Towards the Analysis of Fan Fictions in German Language: Exploration of a Corpus from the Platform Archive of Our Own", in: *2nd International Conference of the Eu-*

ropean Association for Digital Humanities (EADH 2021). Krasnoyarsk, Russia.

Vonderau, Patrick (2020): “Quantitative Werkzeuge”, in: Hagener, Malte / Pantenburg, Volker (eds.): *Handbuch Filmanalyse*, Springer Fachmedien, 399–413. DOI: 10.1007/978-3-658-13339-9_28

Welke, Tina (2005): “Die Tatortfolge 'Quartet in Leipzig' als gesamtdeutscher Tatort: Analyse einer inszenierten deutsch-deutschen Annäherung“, in: *Verl. für Gesprächsforschung*, Radolfzell.

Wu, Yuxin / Kirillov, Alexander / Massa, Francisco / Lo, Wan-Yem / Girshick, Ross (2019): *Detectron2* <<https://github.com/facebookresearch/detectron2>> [14.06.2021]

Wulff, Hans J. (1998): “Semiotik der Filmanalyse: Ein Beitrag zur Methodologie und Kritik filmischer Werkanalyse”, in: *Kodikas/Code*, 21(1-2): 19-36.

Zaharieva, Maia / Breiteneder, Christian (2012): “Recurring Element Detection in Movies”. in Schoeffmann, Klaus et al. (eds.): *Advances in Multimedia Modeling*, Springer, 222–232. DOI: 10.1007/978-3-642-27355-1_22

Zhang, Kaipeng / Zhang, Zhanpeng / Li, Zhifeng / Qiao, Yu (2016): “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks”, in: *IEEE Signal Processing Letters*, 23(10): 1499–1503. DOI: 10.1109/LSP.2016.2603342

Zhou, Bolei, et al. (2017): "Places: A 10 million image database for scene recognition", in: *IEEE transactions on pattern analysis and machine intelligence* 40.6: 1452-1464.