

Synthetische Bilder als Trainingssets des Protests

Investigative Ästhetiken in Forensic Architectures *Model Zoo*

Michael Klippahn

Die Rechercheagentur Forensic Architecture entwickelt derzeit Modelle zur visuellen Aggregation von Daten, die der technischen Identifikation von Tränengasgeschossen dienen. Das Ziel: die Registrierung derartiger Reizkampfstoffe in Onlinevideos von Protestbewegungen zu automatisieren und digitale Objekterkennung zu hacken.

IT-Forensik und Bildmodulation

Tränengas gehört weltweit zu den am häufigsten gegen Demonstrant*innen eingesetzten Reizkampfstoffen. In jüngster Zeit wurde es beispielsweise bei der Niederschlagung der Black-Lives-Matter-Demonstrationen auf Geheiß der Trump-Administration in Portland eingesetzt oder zur Unterdrückung der Proteste der Demokratiebewegung in der chinesischen Sonderverwaltungszone Hongkong verwendet.

Mit dieser Problematik setzt sich auch Forensic Architecture auseinander. Die an der Goldsmith University of London angesiedelte Rechercheagentur hat sich 2011 mit dem Ziel gegründet, Menschenrechtsverletzungen durch Staaten, Polizeiapparate, Militärs und Konzerne zu untersuchen. Derzeit forscht das multidisziplinäre Team, bestehend aus IT- und bildforensisch

arbeitenden Aktivist*innen, Softwareentwickler*innen, Archäolog*innen, Künstler*innen, Journalist*innen und Jurist*innen an Modellen, die der Identifikation, Katalogisierung und Kategorisierungen von Tränengasgeschossen dienen. Ziel ist, die optische Registrierung dieser Munition in der Masse der vielfach online hochgeladenen Videos staatlicher Devastationen von Protestbewegungen zu automatisieren. Denn für Schusskörper wie diese sind bislang zu wenige Bilder als Befunddatensätze in den Lerndaten von Objekterkennungssoftwares verfügbar. Dadurch wird die automatisierte Suche nach diesen Geschossen durch entsprechende Suchprogramme stark beeinträchtigt. Grund dafür ist unter anderem ihre bisherige Klassifizierung durch diverse staatliche Ministerialbehörden als nicht-tödlich hinsichtlich ihrer Einsatzintention.¹ Um das daraus resultierenden Problem des Nicht-Erkennens aufzufangen, modelliert das Kollektiv digital tausende der am häufigsten vorkommenden Variationen der Behältnisse dieser Reizkampfstoffe – einschließlich unterschiedlicher Verformungsgrade, Kratzer, Verfärbungen durch Hitzeeinfluss und diverser Beschriftungen. Dafür werden synthetische Bilder angefertigt: fotorealistische digitale Renderings von 3-D-Modellen. Diese Renderings wiederum werden als Trainingssets für Machine-Learning-Klassifikatoren verwendet, um die Befunddatensätze zur Erkennung- und Registrierung dieser Munition anzureichern.²

Die dafür eingesetzte Technik der Machine Vision basiert auf algorithmischen maschinellen Lernprozessen, die darauf trainiert werden, einen bestimmten Typus eines Objekts anhand

eines eingespeisten Bildpools digital zu identifizieren.



Forensic Architecture, *Model Zoo*, 2020, digitales Rendering von 37- bis 40mm-Tränengasgeschossen, <https://forensic-architecture.org/investigation/model-zoo#resources>, © Forensic Architecture, 2020

Objekterkennung und Bildetikettierung

Objekterkennung ist als maschinelles Lernverfahren grundsätzlich Teil von Machine Vision. Maschinelles Lernen meint — vereinfacht gesagt — , dass ein künstlich-intelligentes System (KI) aus Beispielen lernt und diese nach Beendigung der Lernphase generalisieren und nutzen kann. Diese Beispiele fungieren als Trainingsdaten für Algorithmen, die auf dieser Befundmasse aufbauend in einem statistischen Raum Schlussfolgerungen treffen und zu Ergebnissen in der Kennung von Objekten gelangen können. Mittels Machine Vision werden also Bilder wie Daten als Verarbeitungs- und Analyse-kategorie genutzt, um deren Inhalt zu verstehen und/oder Informationen zu extrahieren. Die Bilddaten stammen heute vornehmlich aus der Erfassung und Analyse visueller Informationen, aus der Auslese von Bildinformationen durch die Verwendung digitaler Filter, aus Social Media, Analog-Digital-Umwandlungen und digitaler Signalverarbeitung.³ Solche KI-Systeme zu entwickeln erfordert und fordert Daten in Form dis-

kreter beziehungsweise privater Bilder ein. Zugrundeliegende neuronale Netzwerke müssen mit Tausenden von Bildern trainiert werden, um belastbare Ergebnisse in der Bilderkennung liefern zu können. Diese Bilddaten müssen wiederum als das etikettiert werden, was als abgebildetes Objekt erkannt werden soll: Das Bild einer Orange beispielweise muss erst verschlagwortet werden, indem es das Etikett „Orange“ erhält, damit es mit weiteren als „Orange“ etikettierten Bildern vergleich- und somit unter diesem Etikett subsumier- und erkennbar wird.

Bilder existieren zwar parallel zu ihrer Beschreibung oder Benennung formal und materiell autark, ihre Wahrnehmung wird aber oftmals von ihrer Etikettierung bestimmt. Diese Bildpools als Trainingsätze sind demnach symptomatisch dafür, wie als KI-Systeme die Welt erkennen und interpretieren und betonen gleichzeitig die instabile Beziehung zwischen Bildern und Bedeutungen. Die mit dieser Sammlung von Bilddaten einhergehende Etikettierung ist nicht nur kostspielig und mühsam, sondern oftmals mit prekärer Arbeit verbunden, die in ihren Grundzügen häufig im globalen Süden verrichtet wird.⁴ Wenn also riesige Bilddatensätze zum Trainieren von KI-Systemen verwendet werden, wird der Prozess des Definierens und Annotierens des Bildinhaltes und die Definition der Bildgegenstände von Klick-Arbeiter*innen vorgenommen, die für diese digitale Akkordarbeit extrem gering entlohnt werden. Diese Typisierung der Bilder durch Menschen birgt auch immer das Risiko, dass ohne Rücksicht auf kulturelle und soziale Werturteile Bilddaten aufgrund von Milieu, Ethnie und/oder Geschlecht sortiert

werden. Die Auswahl, Einsortierung und Beschreibung der Bilder verzerrt so die Bedeutung oftmals in einer Weise, die geschlechtsspezifisch, rassistisch, ableistisch, altersdiskriminierend oder anders problematisch ist. Kate Crawford konstatiert in diesem Zusammenhang:

„Training data sets – the benchmark against which artificial intelligence is trained – are available, by the thousand, presenting data as unproblematic – just a set of images and labels that are somehow neutral and scientific. [...] Once inside this world, we can observe how training data sets have biases, assumptions, errors, and ideological positions built into them. In short, AI is political.“⁵



Forensic Architecture, *Model Zoo*, 2020, Objekterkennung durch Klassifikation auf einem Bild, das mehrere Tränengasgeschosse zeigt, <https://forensic-architecture.org/investigation/detecting-tear-gas>, © Forensic Architecture, 2020

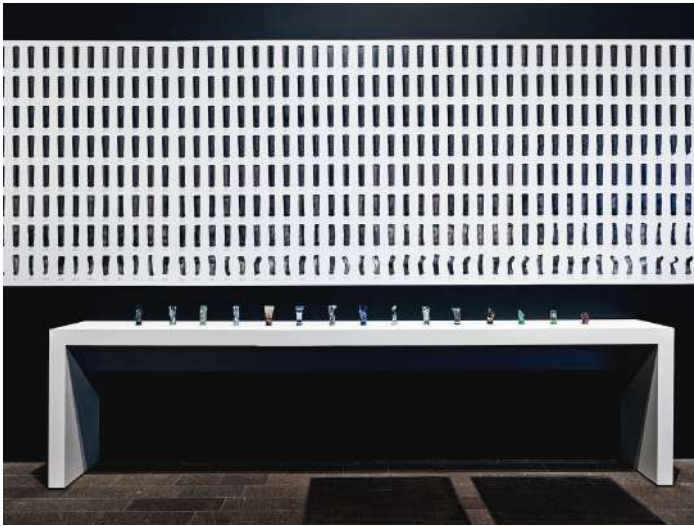
Synthetische Bild(er)zeugnisse

Unter anderem auf Grundlage der von Forensic Architecture erstellten Trainingsdaten zu Tränengasgeschossen entstand das Werk *Model Zoo*, das 2020/2021 in der Ausstellung *Uncanny Valley* im *de Young Museum* in San Francisco gezeigt wurde. Diese multimediale Arbeit besteht, neben der digitalen Entsprechung, die sich

der Objekterkennung widmet, aus fotografischen Druckerzeugnissen und mittels 3-D-Druckverfahren hergestellte Plastiken, die der wachsenden Sammlung von Replikaten, Nachbildungen und alternativen Versionen der angesprochenen Waffen in den verschiedenen Wesenhaftigkeiten, Beschaffenheiten und Kontexten entlehnt sind. Die digitalen Vorbilder zu den Korpora der Waffen werden jeweils mittels der bereits beschriebenen KI-Technik der Machine Vision erzeugt und trainiert, wobei die diesem Prozess zugrundeliegende Objekterkennung nicht für diesen Zweck intendiert ist und sozusagen durch die Einspeisung der generierten und etikettierten Bilddaten gehackt wird.

In dem von Forensic Architecture, gemeinsam mit dem Montrealer KI-Startup Element AI entwickelten *Model Zoo* werden dementsprechende Trainingsdatensätze gerendert. Modellparameter und -transformationen, Kamerawinkel sowie Beleuchtungs- und Umgebungseigenschaften werden indessen zufällig gewählt. Parallel ermöglichen die verschiedenen Parameter der synthetischen Datengenerierung es, die Auswirkungen verschiedener Aspekte der Generierung synthetischer Daten auf die Effizienz des Trainings von Machine-Learning-Klassifikatoren im Allgemeinen zu untersuchen. Mittels maschineller Lernprozesse wird aufbauend auf dieser erweiterten Trainingsdatenbasis eine algorithmische Erkennung der Munition möglich. Proteste werden zunehmend in der Art und Weise mediatisiert, dass Aufnahmen für forensische Zwecke auswertbar werden, unter anderem durch Posts von Tränengaskanistern auf Twitter. Darauf aufbauend

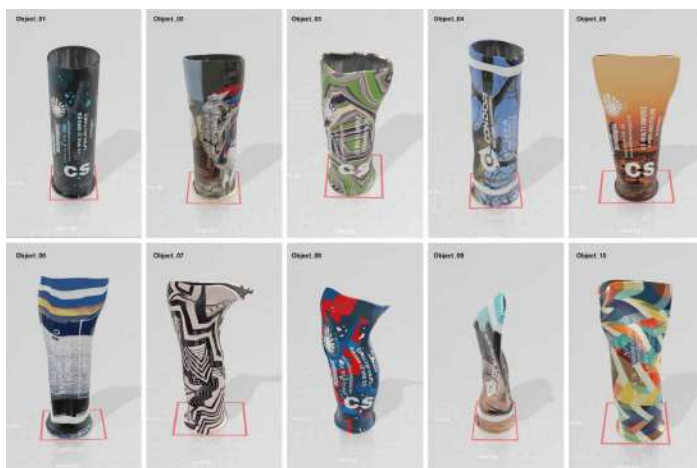
kann der Arbeitsablauf, auf dem Forensic Architecture's *Model Zoo* basiert, helfen, den Einsatz von Tränengas schneller zu identifizieren, indem einzelne Bilder entsprechender Geschosse in stundenlangen Videos leichter isoliert werden können. Diese Bilddaten, die aus Protestaufnahmen gewonnen wurden, die oftmals von Amateur*innen gefilmt und fotografiert werden, helfen wiederum automatisierte Objekterkennung digital zu kapern.⁶



Forensic Architecture, *Model Zoo*, 2020, Ausstellungsansicht *Uncanny Valley: Being Human in the Age of AI*, de Young Museum, San Francisco, <https://forensic-architecture.org/investigation/detecting-tear-gas>, © Forensic Architecture, 2020

In *Model Zoo* zeigt sich somit, dass synthetische (Bild-)Daten eine wertvolle Ergänzung des Werkzeugkastens ästhetischer Forschung sein können, da sie nicht nur die Evidenz von Bildern infrage stellen, sondern auch Objekterkennungstechnologien diversifizieren. Damit wird auch menschenrechtsaktivistischen und protestrelevanten Bestrebungen, die auf das Training und Hacking von Klassifikatoren rekurrieren, Raum gegeben. Dabei baut Forensic Architecture in seinen auf Öffentlichkeit, Mitarbeit der Allgemeinheit und Zugänglichkeit angelegten Open-

Source-Lösungen auf affektive Ökonomien, die grundlegend für die Entstehung von Bildern des Protests sind. Diese synthetischen Bilder lassen sich als „Demonstrationen, die sich von der Straße ins Internet“,⁷ man könnte auch sagen, ins Digitale verlagert haben, lesen. „Besonders Bilder von Gewalt, die von der [...] [Staatsmacht] ausgeht, sei es im Alltag oder während Demonstrationen, sind“, wie Kerstin Schankweiler konstatiert, „zu einem transnationalen Genre avanciert. Die Verbreitung dieser Bilder in den Sozialen Medien hat sich als Motor für die Mobilisierung und Solidarisierung großer Personengruppen erwiesen. Die affektive Dynamik der Bildzeugnisse wirkt gemeinschaftsbildend.“⁸ Forensic Architecture's Vorschlag eines *Model Zoo* markiert in diesem Sinne den Beginn eines neuen Forschungswerkzeugs für die Zivilgesellschaft, das Beweise für Menschenrechtsverletzungen durch Staaten und Militärs mittels Bildern als Werkzeugen des Protests liefern kann. Investigative Ästhetiken, die auf der visuellen Aggregation von Daten beruhen, ermöglichen so eine Form des Hacking digitaler Objekterkennung und dienen als Werkzeuge des Bildprotests.



Forensic Architecture, *Model Zoo*, 2020, digitales Rendering Tränengasgeschossen in extremen Variationen, die durch ihre Gestaltsvielfalt der Kennung durch die Klassifizierung helfen, <https://forensic-architecture.org/investigation/model-zoo#resources>, © Forensic Architecture, 2020

Michael Klippahn promoviert derzeit als Kollegiat des Schaufler Lab in Kunstgeschichte an der TU Dresden bei Prof.ⁱⁿ Dr.ⁱⁿ Kerstin Schankweiler. In diesem Kolleg kooperieren Forschende aus den Geistes- und Sozialwissenschaften mit Kunstschaffenden und Forschenden aus den MINT-Fächern zum Thema Künstliche Intelligenz. Neben diesem Bereich sind Michael Klippahns Forschungsschwerpunkte Gegenwartskunst, Materialästhetik und Neuer Materialismus sowie Ritual- und Magietheorie.

¹ Vgl. Büro des Hohen Kommissars der Vereinten Nationen für Menschenrechte, United Nations Human Rights Guidance on Less-Lethal Weapons in Law Enforcement, Genf/New York 2020, S. 29 ff. https://www.ohchr.org/Documents/HRBodies/CCPR/LLW_Guidance.pdf (zuletzt abgerufen am 27.05.2021).

² Vgl. <https://forensic-architecture.org/investigation/model-zoo> (zuletzt abgerufen am 26.05.2021).

³ Vgl. Kate Crawford, *Atlas of AI*, London/New Haven 2021, S. 96 ff.

⁴ Vgl. ebd. S. 64 f.

⁵ Kate Crawford, in: Daphne Milner, *How art holds AI to account*, 04.12.2019, <https://www.itsnicethat.com/features/how-art-holds-artificial-intelligence-to-account-digital-art-041219> (zuletzt abgerufen am 28.05.2021), zit. n.: Fondazione Prada (Hg.), *Conversation by Kate Crawford and Trevor Paglen*, in: *Training Humans*, Notebook 26, Milan 2019.

⁶ Vgl. <https://forensic-architecture.org/investigation/detecting-tear-gas> (zuletzt abgerufen am 27.05.2021).

⁷ Kerstin Schankweiler, *Selfie-Proteste. Affektzeugenschaften und Bildökonomien in den Social Media*, in: Isabelle Busch/Uwe Fleckner/Judith Waldmann (Hg.), *Nähe auf Distanz. Eigendynamik und mobilisierende Kraft politischer Bilder im Internet*, Berlin 2020, S. 175–190, h. S. 175.

⁸ Kerstin Schankweiler, *Das zensierte Auge*, in: Katja Müller-Helle (Hg.), *Bildzensur – Löschung technischer Bilder*, Berlin 2020, S. 42–48, h. S. 45.