

NEUES SEHEN – Aktuelle Ansätze der Digitalen Archäologie in der Objekt- und Bildwissenschaft. Teil 3/4: Bildmustererkennung und Einsatz von KI

Elisabeth Günther, Torsten Bendschus, Chrisowalandis Deligio,
Kerstin P. Hofmann, Marta Kipke, Martin Langner,
Corinna Reinhardt, Katja Rösler, Ute Verstegen

Abstract: The conference *NEUES SEHEN. Aktuelle Ansätze der Digitalen Archäologie in der Objekt- und Bildwissenschaft* (*NEUES SEHEN. Current approaches to object and image studies in digital archaeology*), held at the University of Trier, 20.–22.05.2022, explored the potential of digital tools and methods applied to archaeological objects and imagery. According to the four sessions of the conference, we will publish a series of four papers in Digital Classics Online (DCO) which include each a short introduction by the organizers, the abstracts of the projects presented during the session, and a transcript of a group discussion among the respective speakers. This third part focusses on the potential of pattern recognition and the application of AI to archaeological research.

Einleitung: Bildmustererkennung und Einsatz von Künstlicher Intelligenz in den Archäologien

Elisabeth Günther

Bei diesem Beitrag handelt es sich um den dritten Teil einer Beitragsreihe, welche die Diskussionen und Ergebnisse des Workshops *NEUES SEHEN. Aktuelle Ansätze der Digitalen Archäologie in der Bild- und Objektwissenschaft* an der Universität Trier (20.–22.05.2022), organisiert von Elisabeth Günther und Sascha Schmitz, vorlegt (vgl. die Ausführungen in der vorangestellten Einführung).

Dieser dritte Teil behandelt die Herausforderungen und Möglichkeiten des Einsatzes von Bildmustererkennung in der Digitalen Bild- und Objektwissenschaft (Sektion 1 der Konferenz).

Digitale Archäologie und Bildmustererkennung

Das Programm *ChatGPT* hat in den letzten Wochen und Monaten einiges Aufsehen erregt. Es ist zur Zeit die bekannteste Anwendung von Künstlicher Intelligenz (KI) und hat, nicht zuletzt in Forschung und Lehre, zu intensiven Diskussionen über den Umgang mit Texten und Bildern geführt, deren Urheber kein Mensch ist. Wie ist eine studentische Leistung zu bewerten, wenn die Hausarbeit oder das Essay ganz oder teilweise von einer KI erzeugt wurden, und wie lässt sich dies überhaupt nachweisen? Sind solche künstlich erzeugten Texte in jedem Fall problematisch oder stellen sie in manchen Bereichen der Wissenschaften doch eher ein pragmatisches Mittel dar, mit dem ohnehin stark standardisierte Berichte – etwa bei der Publikation von Laborexperimenten oder empirischen Untersuchungen – zeitsparend vorgelegt werden können? Dienen mittels KI erzeugte fotorealistic Bilder manipulativer

Meinungsbildung oder stellen sie ein Mittel dar, um ein breites Publikum für schwierige und problematische Themen zu sensibilisieren?¹ Oder handelt es sich um eine existenzielle Bedrohung der Menschheit, die bei mangelnder Kontrolle Auswirkungen vergleichbar der Atombombe entwickeln könnte?²

Der Umgang mit KI-basierten Anwendungen schwankt zwischen Utopie und Dystopie, zwischen übersteigerten Zukunftserwartungen und düsteren Visionen à la *Matrix* (dem Film) – auch deswegen, weil KI schon lange Literatur und Film im Science-Fiction-Genre inspiriert, und dies unsere Vorstellung von menschlichen Robotern nach Art von Pinocchio (etwa *AI* von Steven Spielberg, 2001) oder sich verselbständigenden Supercomputern (so etwa die künstliche Intelligenz ‚Skynet‘ in der Kultreihe *Terminator*, seit den 1980er Jahren) mitbestimmt. Dabei wird deutlich, dass der Begriff *Künstliche Intelligenz* in der Regel mit zukünftigen Entwicklungen in Verbindung gebracht wird, nicht aber mit dem aktuellen Lebensumfeld der Menschen, obwohl dieses längst in unterschiedlichsten Bereichen von KI geprägt wird – Stichwort *Digitalisierung der Alltagswelt*: Selbstlernende Algorithmen in Suchmaschinen, Filterfunktionen in den Social Media, Smart-Home-Anwendungen, Fahrassistenten bzw. automatisiertes Fahren, automatisches Verfassen von Kurznachrichten (v.a. Sport, Wetter, Börse) etc. haben schon jetzt erheblichen Einfluss auf individuelle Handlungen und (Kauf-)Entscheidungen.

Auch für die archäologischen Disziplinen eröffnet sich eine Vielzahl von Anwendungsmöglichkeiten, und zwar von der Auswahl geeigneter Ausgrabungsplätze, der Dokumentation von Grabungsbefunden über die Aufnahme und Klassifizierung von Funden bis hin zur Auswertung des Materials und weiterführenden Interpretationsvorschlägen. Dabei muss man allerdings nicht in die ferne Zukunft blicken, sondern die ersten Anwendungen werden bereits eingeführt, insbesondere dort, wo umfangreiche Datenmengen anfallen, etwa bei großen Grabungen und Oberflächenuntersuchungen, wie z.B. großflächigen Luftbildaufnahmen.³

Ebenfalls im Bereich der archäologischen Objekt- und Bildwissenschaft sind einige Pionierprojekte entstanden, welche die möglichen Anwendungsbereiche von KI auf archäologische Fragestellungen austesten und deren Potenziale wie Grenzen ausloten. Vier Projekte werden im Rahmen dieses Beitrages vorgestellt, die sich insbesondere mit Bildmustererkennung auseinandersetzen, also der automatisierten Erkennung von Mustern in einem Datensatz (*pattern recognition*)⁴. Hierbei wird maschinelles Lernen (*machine learning*) eingesetzt, d.h. der Algorithmus lernt, ausgehend von einem möglichst großen und geeigneten Datensatz an Bildern, bestimmte Muster (wieder-)zuerkennen und dann an einem unbekanntem Datensatz eigenständig Strukturen zu klassifizieren und den zuvor erlernten Mustern zuzuordnen. Dabei wird eine Wahrscheinlichkeit angegeben, mit der die erkannte Struktur mit einer oder mehreren erlernten übereinstimmt (etwa einem im Bild dargestellten Objekt wie einer Keule oder einem Löwenfell). Ein Beispiel hierfür ist die Erkennung von Objekten in Bildern (*object detection*), etwa wenn der Algorithmus darauf trainiert wird, die Keule des Herakles in Vasenbildern wiederzuerkennen. Er wird dann mit einer bestimmten Genauigkeit Muster als Keule identifizieren und dies in Wahrscheinlichkeiten angeben (etwa: an dieser Stelle des Bildes befindet sich zu 90% ein bestimmtes

1 Einen neuen Ansatz wagt etwa das Holocaustmuseum in Aschkelon mit einer Ausstellung KI-generierter Bilder, welche fiktive Szenen zeigen und so dem Vergessen entgegenwirken sollen: *Israel: Gedenken an den Holocaust – mithilfe von KI?*, Reportage von U. Schneider, C. Rosin und D. Jonathan (2023), verfügbar über die Arte-Mediathek: <https://www.arte.tv/de/videos/114364-000-A/israel-gedenken-an-den-holocaust-mithilfe-von-ki/> (abgerufen am 03.06.2023).

2 So etwa behauptet von Warren Buffet im Mai 2023: <https://www.businessinsider.de/wirtschaft/international-business/warren-buffett-kuenstliche-intelligenz-wie-atombomben-entwicklung/> (abgerufen am 03.06.2023); vgl. auch <https://www.faz.net/aktuell/feuilleton/debatten/ist-die-ki-von-openai-gefaehrlich-wie-die-atombombe-18916500.html> (abgerufen am 03.06.2023).

3 Einen Überblick bieten Argyrou / Agapiou (2022).

4 Zur *pattern recognition*: <https://archdigi.hypotheses.org/b-theoriegeleitete-sektionen/klassifizierung-mustererkennung> (abgerufen am 10.04.2023).

Objekt, vgl. Abb. 6 und 9). Dies mag aus menschlicher Sicht in den meisten Fällen zutreffen, es kann aber auch Abweichungen geben, und genau diese Abweichungen dienen als Ausgangspunkt für weitergehende Fragestellungen und Auseinandersetzungen mit Mustern, welche dem menschlichen Auge bis dahin verborgen geblieben sind.

Dabei ist das Training des Algorithmus entscheidend. Dieses beginnt mit der Auswahl des Trainingsdatensatzes, was gerade im Bereich der Digitalen Archäologie aufgrund der begrenzten Datenmenge sowie der (derzeit noch) häufig eher geringen Bildqualität der zur Verfügung stehenden Fotografien der Objekte eine Herausforderung darstellt – wie auch im folgenden Gespräch deutlich wird. Das Training kann dann in Form des *supervised learning* erfolgen, bei dem der Algorithmus auf Daten trainiert wird, denen er bestimmte *labels* zuordnet. Entsprechen die *labels* den vom Menschen definierten Klassen (und damit der *ground truth*)⁵, so werden diese vom menschlichen Experten bzw. der menschlichen Expertin bestätigt. Wenn nicht, wird die Zuordnung korrigiert. Auch an dieser Stelle ergeben sich Herausforderungen im Lernprozess, wenn ein zu identifizierendes Objekt (bzw. eine bestimmte Klasse) im Bild von einem anderen Objekt oder einer Figur überschritten wird und somit nicht vollständig dargestellt ist, oder wenn sich die Form eines Objektes im Laufe der Zeit in den Vasenbildern verändert. Hier ermöglicht erst ein großer Datensatz eine hohe Genauigkeit des Algorithmus.

Eine weitere Form des Trainings ist das *unsupervised learning*, bei dem keine *ground truth* vorausgesetzt wird, der Algorithmus also ohne Input von außen Muster erkennt und die Daten beispielsweise in Form von *cluster* (Gruppen) einteilt. Dabei stellt dann die Interpretation der vom Algorithmus berechneten *cluster* die menschliche Expertin und den menschlichen Experten vor Herausforderungen, kann aber auch gerade deshalb traditionelle Klassifikationen hinterfragen und die Diskussion um Typologien befördern.

Dass Algorithmen überhaupt lernfähig sind, liegt an ihrer Struktur. Verwendet werden, vor allem im Bildbereich, *CNNs* (*Convolutional Neural Networks*), also künstliche neuronale und gefaltete Netze. Diese ähneln natürlichen Neuronen und bestehen aus mehreren Ebenen (*layers*). *CNNs* zeichnen sich durch den namensgebenden *convolutional layer* aus, in welchem die Neuronen in drei Dimensionen angeordnet und lokal miteinander gekoppelt sind.⁶ Diese komplexe Architektur kann die Bilder Pixel für Pixel auf unterschiedliche Merkmale (Klassen) hin untersuchen, und zwar von einfachen kleinen Elementen bis hin zu komplexeren Mustern (etwa von Ecken und Linien über Umrisse hin zu Formen, was dann als *convolutional deep learning* bezeichnet wird).⁷

Eine Herausforderung ist auch die Interpretation der Ergebnisse, denn im komplexen Prozess der Mustererkennung bleibt unklar, warum dieses oder jenes Muster erkannt (oder nicht erkannt) und einer Klasse zugeordnet (oder nicht zugeordnet) wird. Auffällig ist dies dann, wenn ein für das menschliche Auge klar erkennbares Objekt in einem Bild nicht bzw. falsch eingeordnet wird. Hier wird deutlich, dass der Mensch die bildliche Darstellung stets in den Raum projiziert (weshalb auch ein Tisch mit drei statt vier sichtbaren Beinen als solcher leicht erkannt wird) und die Sinneseindrücke anhand der bereits bestehenden Erfahrungen und Erwartungen verarbeitet werden (wodurch sich auch ein vierbeiniger Sitzhocker leicht von einem Tisch unterscheiden lässt).⁸

Auch wenn immer wieder von der *black box* gesprochen wird, als welche die KI aufgrund der nicht nachvollziehbaren Entscheidungsprozesse (VGG16, eine typische Struktur, hat beispielsweise 138

5 Zur Terminologie s. https://clarenet.hypotheses.org/glossar_it (abgerufen am 10.04.2023).

6 Diese Erklärung ist zum besseren Verständnis stark vereinfacht. Es gibt unterschiedliche Formen neuronaler Netze und sehr unterschiedliche Architekturen, je nach Aufgabenstellung. Zu neuronalen Netzen und *CNNs* s. etwa <https://cs231n.github.io/convolutional-networks/> (abgerufen am 10.04.2023).

7 Goodfellow et al. (2018), 6–10.

8 Zu Mehrdeutigkeiten im menschlichen Perzeptions- und Rezeptionsprozess s. Günther (2021).

Mio. Parameter)⁹ bei der Klassifikation angesehen wird, gibt es doch Methoden, diese in gewissem Rahmen sichtbar und nachvollziehbar bzw. erklärbar (*explainability*) zu machen. Hierzu zählen etwa die sog. *heatmaps*, die durch farbige Abstufungen sichtbar machen, welche Teile des Bildes vom Algorithmus als relevant bzw. weniger relevant angesehen wurden (vgl. Abstract 4).

Ganz im Sinne des Workshops *NEUES SEHEN* legen die in diesem Beitrag präsentierten Projekte offen, wo uns Algorithmen Neues sehen und erkennen lassen, und inwieweit hier neue Formen des Sehens entstehen können, welche die traditionellen, bewährten Methoden der Bildanalyse erweitern können, teilweise auch herausfordern. An der Schnittstelle von Detailstudie und Datenbankauswertung, von Geisteswissenschaft und Digital Humanities, stellen sich Fragen, welche unser Verständnis von Archäologie als (Geistes-, Altertums-, Natur-)Wissenschaft berühren und die zukünftige Ausrichtung der archäologischen Disziplinen ebenso betreffen wie die künftige Ausgestaltung von Forschungs- und Publikationspraktiken, die insbesondere interdisziplinären Kooperationen Rechnung tragen. Doch auch bereits die Umsetzung der einzelnen Arbeitsschritte in den hier vorgestellten Projekten steht, wie auch aus dem Gesprächstranskript deutlich wird, vor Herausforderungen, die sich aus dem häufig fragmentierten Erhaltungszustand der Objekte, kleinen Datenmengen, (noch) eingeschränkter Verfügbarkeit von Bildmaterial und hohem Zeitaufwand bei der Digitalisierung der Bilder ergeben. Dies macht deutlich, wie dringlich die (öffentliche) Verfügbarkeit von Bildmaterial und dessen Lizenzierung sind; Aufgaben, die nun auch im Rahmen der Initiative *NFDI4Objects* im Rahmen des Aufbaus einer *Nationalen Dateninfrastruktur* angegangen werden sollen.¹⁰

Die Sprecherinnen und Sprecher und ihre Projekte

Die Teilnehmerinnen und Teilnehmer¹¹ des Gruppeninterviews bzw. Gesprächs zeichnen sich durch besondere Expertise im Bereich der Mustererkennung in der Digitalen Archäologie aus. Die jeweiligen im Rahmen der Tagung vorgestellten Projekte, deren Abstracts unten abgedruckt sind, wenden Mustererkennung mittels neuronaler Netze auf die Analyse antiker Bilder unterschiedlichster Gattungen und Zeiträume an – von keltischen Münzen über griechische Vasenmalerei bis hin zur christlichen Bilderwelt der Spätantike.

Prof. Dr. Martin Langner hat seit 2019 die Professur für Digitale Bild- und Objektwissenschaft in Göttingen inne und ist Direktor des Instituts für Digital Humanities.¹² Die Forschungsinteressen von Martin Langner umfassen die kontextuelle Bildanalyse (v.a. Griechische Vasenmalerei, Römische Wandmalerei und Relief), die 3D-Digitalisierung und Analyse antiker Skulpturen, virtuelle historische Räume und Museen, die Erforschung antiker und nachantiker Alltagskultur sowie grundsätzlich die Theorie und Methodik naturwissenschaftlicher und informationstechnischer Methoden in den Geisteswissenschaften. Aktuell leitet er u.a. die Projekte *EGRAPHSEN. Möglichkeiten und Perspektiven der digitalen Malerzuweisung bei attischen Vasen*¹³ und *Schemata. 3D Klassifikation und Kategorisierung antiker Terrakotten*¹⁴.

9 Vgl. Simonyan / Zisserman (2015). Ich danke Chrisowalandis Deligio herzlich für diese Ergänzung sowie für seine wertvollen Ratschläge hinsichtlich der Formulierungen und Erklärungen zur Funktionsweise Künstlicher Intelligenz in dieser kurzen Einleitung.

10 <https://www.nfdi4objects.net/> (abgerufen am 03.06.2023).

11 In Absprache mit den Teilnehmerinnen und Teilnehmern der Tagung wurde festgelegt, jeweils die weibliche und männliche Sprachform zu verwenden. In den Abstracts und Gruppeninterviews kann es hierbei, abhängig von der Aussageabsicht, zu Abweichungen und damit scheinbaren Uneinheitlichkeiten kommen.

12 <https://www.uni-goettingen.de/digitalhumanities> (abgerufen am 03.06.2023).

13 <https://www.uni-goettingen.de/de/598165.html> (abgerufen am 03.06.2023).

14 <https://www.uni-goettingen.de/de/598167.html> (abgerufen am 03.06.2023).

Marta Kipke, M.A., ist wissenschaftliche Mitarbeiterin am Institut für Digital Humanities der Georg-August-Universität Göttingen und Projektmitarbeiterin in *EGRAPHSSEN. Möglichkeiten und Perspektiven der digitalen Malerzuweisung bei attischen Vasen* (Abstract Nr. 1).¹⁵

Prof. Dr. Corinna Reinhardt lehrt und forscht als Professorin für die Archäologie des Mittelmeerraums mit Schwerpunkt griechische und römische Antike an der Universität Zürich. Sie interessiert sich unter anderem für die Entwicklung und Anwendung von digitalen Methoden in der Bildanalyse unter dem Einsatz von Methoden der *Computer Vision* und des *Deep Learning*. Sie leitete von 2019–2022 als Juniorprofessorin am Institut für Klassische Archäologie an der Friedrich-Alexander-Universität Erlangen-Nürnberg ein Teilprojekt des von der *Emerging Fields Initiative* der Universität geförderten interdisziplinären Forschungsprojekts *Iconographics. Computational Understanding of Iconography and Narration in Visual Cultural Heritage* in Zusammenarbeit der Kunstgeschichte, der Klassischen Archäologie, der Christlichen Archäologie sowie des *Pattern Recognition Lab* der FAU Erlangen-Nürnberg (Abstract Nr. 2).¹⁶ Dr. Torsten Bendschus gestaltete als Mitarbeiter im Projekt maßgeblich die erzielten Ergebnisse in Forschung und Lehre mit. Digitale Kompetenzen in der Anwendung und Reflexion im Projekt *Iconographics* an der FAU Erlangen-Nürnberg zu fördern, war Ansatz mehrerer Lehrveranstaltungen und Lehrprojekte, darunter die Entwicklung einer virtuellen Ausstellung in dreidimensional erfahrbaren, digital modellierten Räumen.¹⁷

Prof. Dr. Ute Verstegen ist Inhaberin des Lehrstuhls für Christliche Archäologie an der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)¹⁸ und Zweitmitglied im Department Digital Humanities and Social Studies der FAU. Sie ist Vorsitzende der Zentralkommission des Deutschen Archäologischen Instituts. Ute Verstegen arbeitet seit über zwanzig Jahren auf dem Feld digitaler Anwendungen in Kunstgeschichte und Archäologie in Lehre, Forschung und Wissenschaftskommunikation. Ihre Lehrprojekte umfassten zuletzt *CA 2.x – Christliche Archäologie im Inverted Classroom*¹⁹, *Virtual Excursions*²⁰, *Teaching Early Christian Archaeology with Augmented Reality*²¹ sowie eine kartenbasierte Augmented-Reality-Mobile-App zum Nürnberger Kreuzweg auf Basis des FAU-GeoExplorer²². 2019–2022 leitete sie gemeinsam mit Peter Bell (Digitale Kunstgeschichte), Andreas Maier (Informatik, Pattern Recognition) und Corinna Reinhardt (Klassische Archäologie) an der FAU das Projekt *Iconographics* (Abstract Nr. 3).²³

Dr. Kerstin P. Hofmann ist Direktorin der Römisch-Germanischen Kommission des Deutschen Archäologischen Instituts. Sie hat Informatik als Nebenfach studiert und ist Leiterin mehrerer interdisziplinärer Projekte zu digitalen Ding-Editionen sowie archäologischen Wissenspraktiken und ihrer Transformation im Rahmen des *digital turn*.²⁴

15 <https://www.uni-goettingen.de/de/598165.html> (abgerufen am 03.06.2023).

16 S. hierzu ausführlicher Abstract 2. Publikationen u.a.: Bendschus et al. (2022); Madhu et al. (2023).

17 Virtuelle Ausstellung mit vorbereitendem Projekt mit Studierenden der FAU Erlangen-Nürnberg: *Kreuz und quer: Lebensgeschichten antiker Objekte* (2020–2021): https://www.klassischearchaeologie.phil.uni-erlangen.de/kreuzundquer-ausstellung_210128/index.html (abgerufen am 14.03.2023); Lehrprojekt *Planspiel Digital Humanities* zusammen mit Dr. Torsten Bendschus, s. Reinhardt / Bendschus (2021).

18 <https://www.ca.phil.fau.de/lehrstuhl/team/ute-verstegen/> (abgerufen am 20.03.2023).

19 Mührenberg / Verstegen (2020); Mührenberg / Verstegen (2022).

20 Mührenberg / Verstegen (2022); Verstegen et al. (2022).

21 <https://www.ca.phil.fau.de/forschung/projekte/teaching-early-christian-archaeology-with-augmented-reality/> (abgerufen am 20.03.2023).

22 Verstegen / Kremer (2023).

23 Madhu et al. (2019); Madhu et al. (2022); Bendschus et al. (2022); Mührenberg et al. (2022); Madhu et al. (2023).

24 <https://www.dainst.org/wer-wir-sind/mitarbeitende/noslug/564> (abgerufen am 10.10.2023).

Hierzu zählt auch das Projekt *ClareNet – Klassifikationen und Repräsentationen für Netzwerke. Von Typen und Merkmalen zu Linked Open Data bei keltischen Münzprägungen* (Abstract Nr. 4), in dem Chrisowalandis Deligio, M.Sc., als Wissenschaftlicher Mitarbeiter im Bereich Informatik tätig ist.²⁵ Er ist an der Goethe-Universität Frankfurt am Main angestellt und sein Fokus liegt innerhalb des Projektes auf der Anwendung KI-gestützter Methoden auf keltische Münzen.

Dr. Katja Rösler, Prähistorische Archäologin, ist zuständig für das Forschungsdatenmanagement-Teilprojekt *Normdaten für archäologische Objekte* der Zentralen Wissenschaftlichen Dienste des Deutschen Archäologischen Institutes und arbeitet mit Kerstin Hofmann im Akademievorhaben *disiecta membra*²⁶ zusammen. Im *ClareNet*-Projekt setzt sie sich insbesondere mit der kritischen Reflexion archäologischer Methoden und Begrifflichkeiten auseinander.²⁷

Die Moderatorin Dr. Elisabeth Günther war Akademische Rätin a.Z. im Fach Klassische Archäologie der Universität Trier. Seit April 2023 ist sie als Akademische Rätin am Institut für Klassische Archäologie und Byzantinische Archäologie der Universität Heidelberg tätig.²⁸ Sie war und ist an mehreren interdisziplinären wie internationalen Lehrprojekten zur Erstellung einer Onlineausstellung mit interaktiven Karten beteiligt.²⁹

Abstracts der Projekte der Sprecherinnen und Sprecher

Abstract 1: *EGRAPHSEN?* – Möglichkeiten und Grenzen der digitalen Malerzuweisung am Beispiel des *Berliner Malers*

Prof. Dr. Martin Langner, Marta Kipke, M.A. (Institut für Digital Humanities, Georg August-Universität Göttingen)

In der Klassischen Archäologie ist die Arbeit von John D. Beazley und sein Beitrag zur Erforschung der griechischen Vasenmalerei gut bekannt. Unter Verwendung und Anpassung der Methode Morellis verglich er spezifische Details der Bilder miteinander, um die ‚Handschriften‘ der Maler zu erkennen.³⁰ Insbesondere im Licht der rasanten Entwicklung computergestützter Analysemethoden ist eine Revision der Methode nun angemessen. Dabei ist es nicht nur interessant, sie in Hinblick auf den Erkenntnisgewinn über Werkstätten- und Stilgeschichte zu betrachten, sondern auch unter dem Blickpunkt des schwer fassbaren Phänomens menschlicher Kennerschaft.

25 <https://clarenet.hypotheses.org/team/chrisowalandis-deligio> (abgerufen am 03.06.2023).

26 <https://www.adwmainz.de/projekte/disiecta-membra-steinarchitektur-und-staedtewesen-im-roemischen-deutschland/informationen.html> (abgerufen am 19.06.2023).

27 <https://clarenet.hypotheses.org/team/katja-roesler> (abgerufen am 03.06.2023).

28 <https://www.uni-heidelberg.de/fakultaeten/philosophie/zaw/klarch/mitarbeiter/guenther.html> (abgerufen am 21.04.2023).

29 Studentische numismatische Ausstellung *Machtszenarien – Scenarios of Power* 2019 am Institut für Klassische Archäologie der Friedrich-Alexander-Universität Erlangen-Nürnberg und am Institute for the History of Ancient Civilizations, Northeast Normal University, Changchun, China, gemeinsam mit Prof. Dr. Sven Günther: <https://www.klassischearchaeologie.phil.fau.de/ausstellungen/machtszenarien-scenarios-of-power-eine-numismatische-ausstellung/welcome-to-our-international-exhibition-of-roman-coins/> (abgerufen am 13.12.2022), s. hierzu Günther / Günther (2021). Zu weiteren Digitalprojekten von Elisabeth Günther s. Günther / Günther (2022a); Günther / Günther (2022b); Günther (2022).

30 Graepler (2016).



Abb. 1: Hierarchisches Annotationssystem am Beispiel des Armes einer Flötenspielerin. Darstellung auf einer Lekythos in New York (Metropolitan Museum, Inv.-Nr. 24.97.28).

Im Projekt *EGRAPHSEN. Möglichkeiten und Perspektiven der digitalen Malerzuweisung bei attischen Vasen*³¹ widmen wir uns diesem Prozess deswegen aus einer neuartigen Perspektive. Denn Muster in Bildern zu erkennen und Klassifizierungen anhand erkannter Merkmale vorzunehmen, ist nicht nur einem trainierten menschlichen Auge zu eigen. Ein ähnliches Vorgehen hat in der Informatik im Bereich *Computer Vision* schon seit vielen Jahren Eingang gefunden und wird auch im Bereich der Digital Humanities zur Beantwortung geisteswissenschaftlicher Klassifizierungsfragen verwendet.³² Konkret bedeutet das für unser Vorhaben, dass wir im Rahmen dieser Fragestellung die Verwendung künstlicher neuronaler Netze in einem *supervised learning approach* untersuchen möchten.

Um der Komplexität dieser Fragestellung gerecht zu werden, haben wir ein kleinteiliges und ausführliches Annotationssystem entwickelt. Es basiert auf einem hierarchischen Prinzip: Zunächst wird eine Figur als solche annotiert, dann ihre Bestandteile (wie Körperteile, Kleidung und Attribute) und schließlich die Details dieser Bestandteile. Ein Arm ist beispielsweise untergliedert in Oberarm, Ellbogen, Unterarm und Hand (Abb. 1). Die Beziehungen der Elemente zu ihren Bestandteilen erfolgt über die Benennung einer *parent-child*-Beziehung. Zusätzlich zu dieser Benennung der annotierten Abschnitte werden auch weitere Details zur Darstellungsweise in Bezug auf Zustand, Aktionen und Orientierung des annotierten Bereichs angegeben. So kann ein Arm gebeugt oder ausgestreckt sein, eine Hand kann etwas halten und Körperteile können zuweilen auch frontal dargestellt sein. Ferner werden auch die Interaktionen der Figuren miteinander annotiert und die Figuren benannt, soweit sie identifizierbar sind.

Diese Annotationsweise soll zwei wichtige Aspekte im Trainingsprozess gewährleisten: Erstens dient diese Zerlegung der Bilder dazu, mögliche Einflüsse des Gesamtmotivs zu reduzieren und den Fokus auf die Linienführung zu setzen. Zweitens sollen unterschiedliche Details, Kombinationen von Details und Detailebenen darauf untersucht werden, wie gut sie sich für eine Malerzuweisung eignen. Dabei

31 Das Kooperationsprojekt führen wir gemeinsam mit dem *Information Systems and Machine Learning Lab* der Universität Hildesheim (Prof. Dr. Lars Schmidt-Thieme und Lukas Brinkmeyer) durch. Das Niedersächsische Ministerium für Wissenschaft und Kultur fördert das Forschungsprojekt im Rahmen des Programms *Geistes- und Kulturwissenschaften – digital*.

32 S. z.B. Arnold / Tilton (2021).

können die zusätzlich annotierten Informationen dazu dienen, auffällige Darstellungsweisen zu filtern und mögliche *biases* in den Daten ausfindig zu machen.

Eine derartig kleinteilige Annotation ist aufwendig und zeitintensiv, jedoch zwingend erforderlich, um die Frage einem geisteswissenschaftlichen Anspruch genügend und mit der notwendigen Nachvollziehbarkeit der Ergebnisse zu untersuchen. Dementgegen steht die Bedingung der Verfahren, möglichst große Datensätze für das Training zu verwenden. Deswegen haben wir ein Annotationstool entwickelt, das in einem halbautomatischen Workflow den Arbeitsprozess beschleunigt und optimiert (Abb. 2).³³ Es basiert auf einer *Python*-Adaption des Tools *LabelMe*³⁴, welches von uns um eine *object-detection*-Komponente erweitert wurde. Als Rückgrat dient dabei das *YOLOv3*-Modell³⁵ und der *COCO*-Datensatz³⁶ zum Training der unteren *layer*. Für die oberen *layer* wurden unsere eigenen, zunächst händisch vorgenommenen Annotationen verwendet.

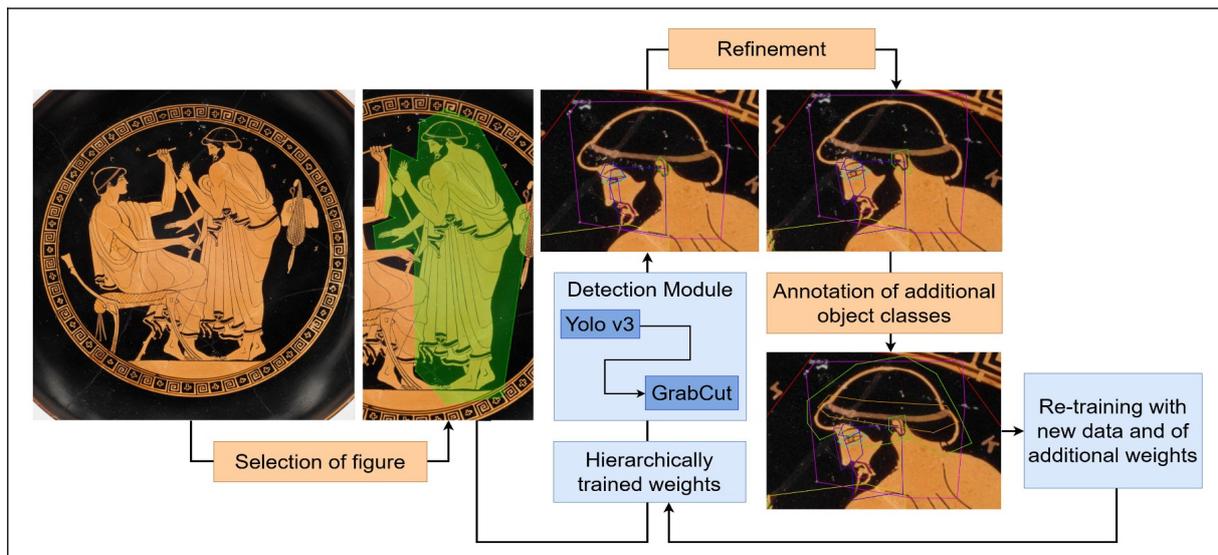


Abb. 2: Workflow für die halbautomatische Bildannotation am Beispiel einer Schale aus New York (Metropolitan Museum, Inv.-Nr. 52.11.4). Orange markiert sind die menschlichen Arbeitsschritte, blau die Prozesse der *object-detection*-Komponente.

Der Vorteil an *YOLOv3* ist, dass die Aufgabe der Klassifikation und der Lokalisierung kombiniert werden und deswegen eine hohe Genauigkeit erzielt werden kann. Außerdem nutzen wir unser hierarchisches Annotationssystem auch zur Optimierung unserer *object-detection*-Komponente: Zunächst markiert der/die User*in eine Figur. Dann wird nur in diesem Bereich nach der nächsten hierarchischen Ebene, also Armen, Beinen, Kleidungsstücken etc. gesucht. Wiederum nur in diesen Bereichen wird dann die nächste Detailsbene erkannt.

Auf diese Weise können fehlerhafte Klassifikationen verringert werden und auch auf vergleichsweise geringer Datenmenge im Trainingssatz bereits sehr gute Ergebnisse erzielt werden. Um aus den rechteckigen Regionen nun polygonale Annotationen zu machen, wird der *GrabCut*-Algorithmus³⁷ verwendet. Der*die User*in hat dann die Möglichkeit, die annotierten Polygone zu korrigieren und die Annotation mit weiteren Informationen anzureichern. Die hierarchischen *parent-child*-Beziehungen sind jedoch fest einprogrammiert und müssen nicht mehr händisch annotiert werden, was nicht nur den Arbeits-, sondern auch den Korrekturprozess deutlich beschleunigt.

33 Kipke et al. (2022).

34 *PyLabelMe*. A Simple Image Annotation: <https://github.com/mpitid/pylabelme> (abgerufen am 25.03.2023).

35 Redmon / Farhadi (2018).

36 Lin et al. (2014).

37 Rother et al. (2004).

Somit war es uns möglich, einen Datensatz mit über 120.000 Einzelannotationen zu erstellen und ein Modell zu trainieren, das in der Malerzuweisung eine hohe Trefferquote erzielt. So lässt sich die Frage der Malerzuweisung und Kennerschaft beispielsweise anhand des Berliner Malers, seines Umfelds, seiner Schüler, seiner Werkstatt und Vorläufer und Zeitgenossen untersuchen. Dabei werden starke Signale, also Zusammenhänge zwischen den Bildern, in den Daten deutlich, die einer kritischen Deutung bedürfen. Während viele Hinweise dafür sprechen, dass die Handschrift des Malers erfolgreich trainiert wurde, sind auch Ähnlichkeiten auf Grundlage chronologischer, motivischer oder formtypologischer Zusammenhänge zu erkennen. Eine Untersuchung dieser Ähnlichkeitsnetzwerke verspricht nicht nur Erkenntnisse in der Frage der Malerzuweisung, sondern auch der Phänomene von Ähnlichkeit und Stil im Allgemeinen.

Link zum Projekt: <https://www.uni-goettingen.de/de/598165.html> (abgerufen am 05.06.2023).

Abstract 2: *Iconographics* – Klassische Archäologie

Prof. Dr. Corinna Reinhardt, Dr. Torsten Bendschus (Friedrich-Alexander-Universität Erlangen-Nürnberg)

Wie Interaktionen von Personen anhand ihrer Körperhaltungen und Armbewegungen im Bild gezeigt werden und wie ihre Analyse durch Methoden der *Computer Vision* und des *Deep Learning* unterstützt werden kann, bildet einen Schwerpunkt des von der *Emerging Fields Initiative* der Friedrich-Alexander-Universität Erlangen-Nürnberg geförderten interdisziplinären Forschungsprojekts *Iconographics. Computational Understanding of Iconography and Narration in Visual Cultural Heritage* in Zusammenarbeit der Kunstgeschichte, der Klassischen Archäologie, der Christlichen Archäologie sowie dem *Pattern Recognition Lab* der FAU Erlangen-Nürnberg.

Für viele beliebte Bilderzählungen der rotfigurigen Vasenmalerei des 5. Jh. v. Chr. bilden spezifische Interaktionen der dargestellten Figuren ein wiedererkennbares Schema, so beispielsweise für die Liebesverfolgung verschiedener Sterblicher durch Gottheiten:³⁸ Der verfolgende Part stürmt auf diesen Bildern weit ausschreitend mit ausgestreckten Armen und bisweilen mit Attributen bewaffnet auf eine zweite Person zu, die meist die Flucht ergreift und mit den Armen gestikuliert (Abb. 3). Das Geschehen ergänzen oft weitere fliehende Personen, die dieselben Gesten wie die Hauptpersonen aufgreifen und damit verstärken, oder leicht abwandeln (Abb. 4). Die Körpersprache leistet im Bild durch die relative Konstanz des Bildschemas nicht nur einen Wiedererkennungswert der Ikonographie, sondern ermöglicht durch ihre Varianzen auch signifikante Unterschiede der Bilderzählung, wenn damit beispielsweise die Drastik des Angriffs oder die Reaktion der Verfolgten unterschiedlich zum Thema gemacht wird. So kann der Verfolger beide oder nur einen Arm ausstrecken und kann dabei sogar bisweilen eine Waffe mit sich führen. Auch die Verfolgten können sehr unterschiedliche Reaktionen zeigen – vom Fliehen mit ins Gewand gehüllten Armen über eine abwehrend ausgestreckte Hand bis hin zu über dem Kopf gehaltenen Gegenständen, mit denen sie zum Schlag ausholen.

38 Kaempf-Dimitriadou (1979); Stansbury-O'Donnell (2009).



Abb. 3: Boreas verfolgt Oreithyia. Ausschnitt von einem rotfigurigen Kelchkrater (um 450 v. Chr.) in Erlangen (Antikensammlung, Inv.-Nr. I 387).



Abb. 4: Boreas verfolgt Oreithyia. Vasenbild einer rotfigurigen Oinochoe des Pan-Malers (um 480/70 v. Chr.) in London (British Museum, Inv.-Nr. 1836,0224.68 [E 512]).

Mit dieser Konstanz und Varianz eines zentralen Bildschemas sind viele Fragestellungen verknüpft, die unter anderem die Zeitstellung, die Darstellungsweisen der Gottheiten, oder aber auch Zuweisungen an Maler und Malerkreise umfassen. So zeigt ein erster Überblick aus einem Corpus von circa 600 Liebesverfolgungsszenen, dass z.B. die dem Frauenbad-Maler zugeschriebenen Verfolgungen des Kephalos durch Eos beide Figuren in unterschiedlichen Vasenbildern im exakt gleichen Schema zeigen. Auch der Pisticci-Maler, der Eos einen nicht weiter charakterisierten Jüngling verfolgen lässt, nutzt für

die Darstellung der Ikonographie zwar andere Körperhaltungen für Verfolgerin und Verfolgten als der Frauenbad-Maler, ebenso aber auf unterschiedlichen Vasen stets die gleichen. Demgegenüber variiert der Niobiden-Maler in den vielen ihm zugeschriebenen Vasenbildern der Verfolgung der Oreithyia durch Boreas die Darstellungsweisen der beiden Figuren und ihrer Kombination.

Die gestische Interaktion beider Protagonisten ist zudem von besonderem Interesse für die Frage, wie im Bild das (ggf. sogar emotionale) Verhältnis der beiden thematisiert wird, z.B. vor dem Hintergrund der Präsenz von Waffen. Über die Ikonographie hinaus spielen diese Schemata eine Rolle, beispielsweise im Bezug zu den eng verwandten Raub- und Entführungsszenen, die den Ausgang der Verfolgung, das Ergreifen, zeigen. Womöglich konnten die Betrachtenden anhand der Hand- und Armstellungen der Figuren in Verfolgungsszenen bereits den nächsten Teil der Episode gedanklich antizipieren, gerade weil die Vasenmaler die Interaktionen in ähnlicher Weise, aber mit engerem Körperkontakt darstellten.

Derartige Forschungsfragen erfordern es, eine Vielzahl von Bildern in Hinblick auf die Interaktionen der Protagonisten zu vergleichen. Die Ikonographie der Liebesverfolgungen ist häufig bezeugt, so dass sich an ihr exemplarisch zentrale Anforderungen an die computergestützte Bildanalyse ableiten und verschiedene Verfahren testen lassen. Dabei stellt die computergestützte Visualisierung von Ähnlichkeitsrelationen einen wichtigen Schritt für die weitere Auswertung dar. Zwei mögliche Wege werden diskutiert: Beim ersten handelt es sich um ein im Rahmen einer Masterarbeit des Projekts von Angel Villar-Corrales am *Pattern Recognition Lab* der FAU Erlangen-Nürnberg entwickeltes Anwendungstool, das in einem beliebigen hochgeladenen Bild Figuren auffindet sowie ihre Körperhaltungen automatisch ermittelt und anschließend aus einem hinterlegten Bildpool Bilder mit vergleichbaren Körperhaltungen anzuzeigen vermag. Grundlage hierfür ist eine digitale Posenerkennung (*pose estimation*), deren Anwendung zunächst ein weiteres Training spezifisch für Vasenbilder nötig machte. Hierfür wurde zum einen das moderne Bilderset von *COCO-Persons* mittels *style-transfer*-Verfahren dem ‚Stil‘ (vor allem dem Kolorit) der Vasenmalerei angepasst.³⁹ Zum anderen wurden dem Modell zusätzlich manuelle Annotationen von bis zu 18 Gelenkpunkten in circa 2.000 Vasenbildern zur Verfügung gestellt. Das Ergebnis sukzessiver Anpassungen und Trainingsphasen ist eine robuste digitale Posenerkennung, die Personen in Vasenbildern lokalisiert, ihre Gelenkpunkte feststellt und mittels *pose parsing* zu *Posenskeletten* vernetzt. Das Tool ist imstande, Körperhaltungen in frei wählbaren Bildern nicht nur zu ermitteln, sondern die Posenskelette auch basierend auf Distanzmetriken mit den Körperhaltungen von Figuren in einem hinterlegten Bilddatensatz zu vergleichen und damit Ergebnisse graduell abgestuft zu visualisieren (Abb. 5). Herausforderungen bestehen hier (wie auch generell für archäologisches Material in der digitalen Bildanalyse) durch fragmentierte Erhaltungszustände, Überlappungen und Faltenwurf, die heterogene Qualität des Bildmaterials, aber auch durch die Krümmung des dreidimensionalen Bildträgers.⁴⁰ Die Konzentration auf Gelenkpunkte ermöglicht es bereits vielfach, sehr ähnliche Körperhaltungen miteinander zu vernetzen. Sie macht es jedoch nicht möglich, detailliertere Gesten, für die die Stellung der Finger oder die Drehungen der Handflächen relevant sind, wiederzuerkennen. Hier könnte die Kombination mit anderen Verfahren, z.B. einer speziell trainierten *object detection* (Abb. 6), eine Verfeinerung der Ergebnisse bewirken.

39 Madhu et al. (2023).

40 Bendschus et al. (2022).

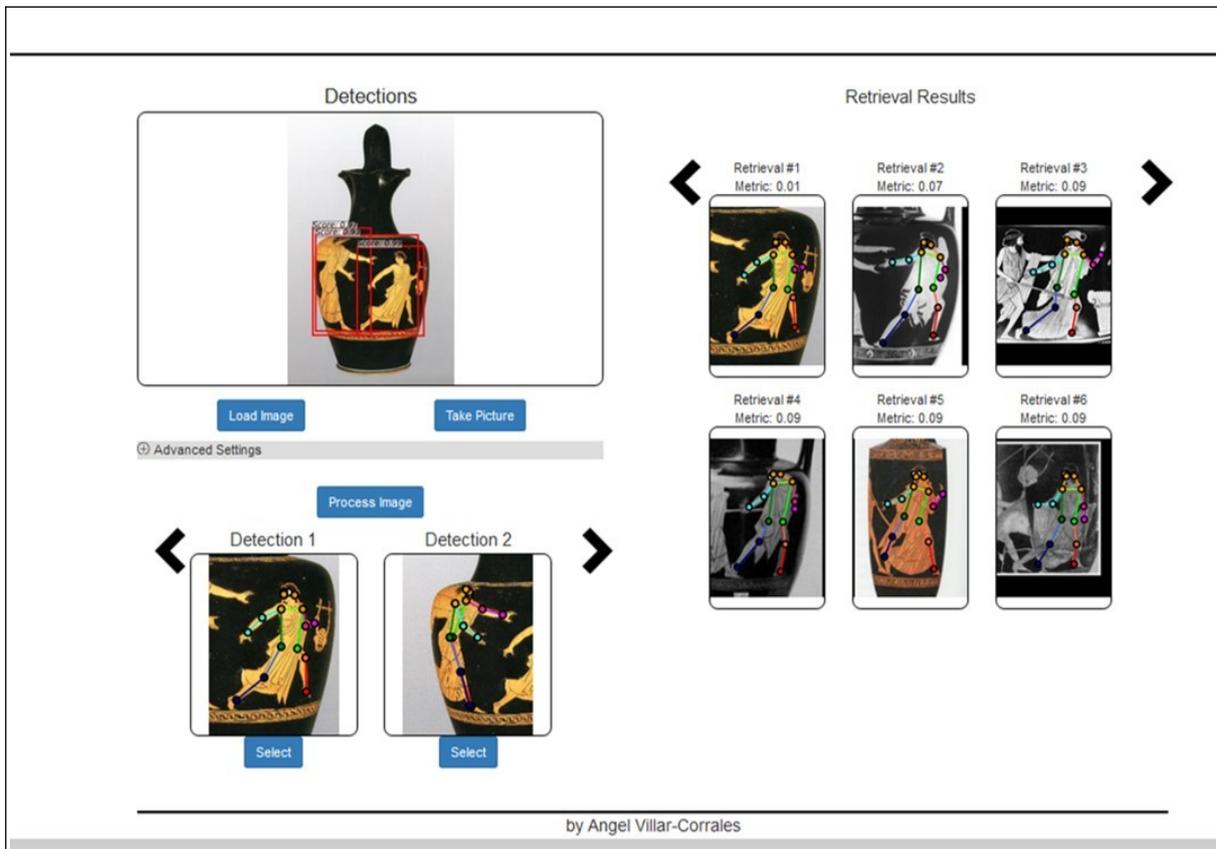


Abb. 5: Das projektinterne *Pose-based-Image-Retrieval-Tool* von Angel Villar-Corrales.

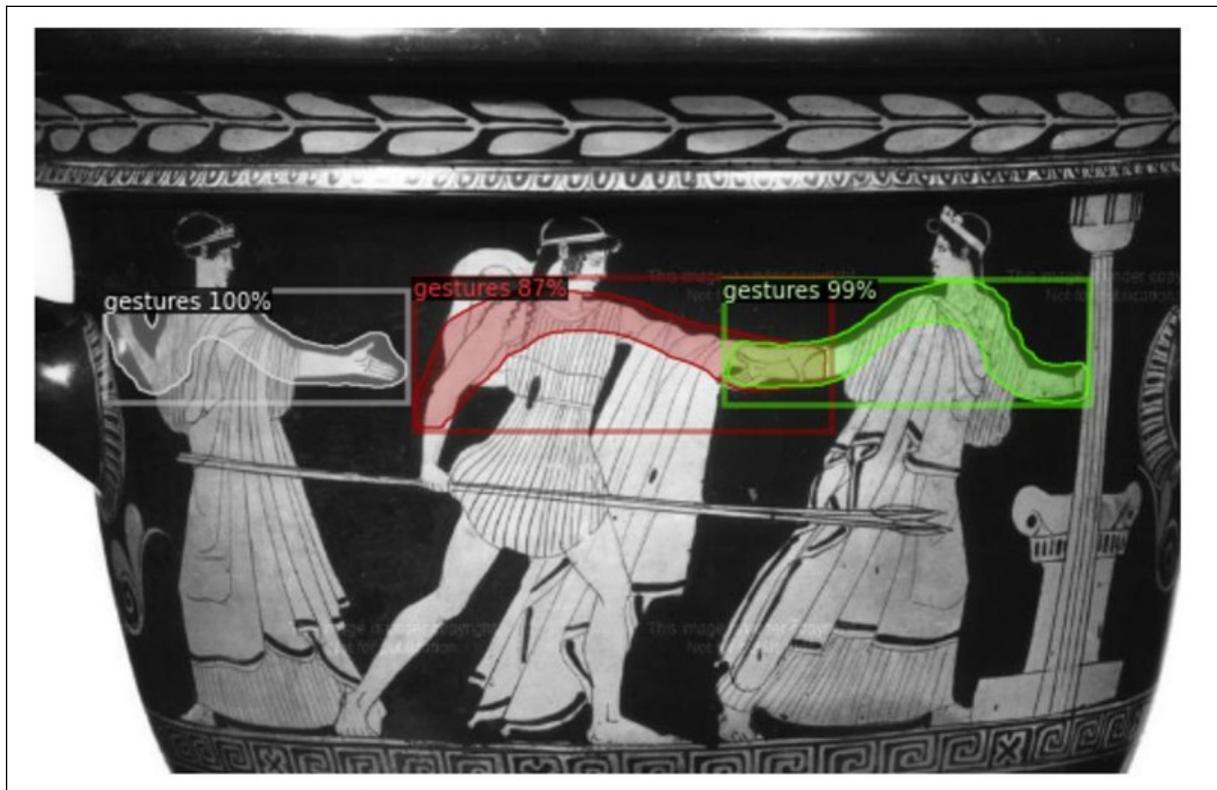


Abb. 6: Beispiele für die Erkennung von spezifischen Gesten mit dem Verfahren der digitalen Objekterkennung.

Als zweite Methode wurde das Bild-*clustering* in sogenannten *UMAP*- oder *PixPlots*⁴¹ erprobt, welche Bilder anhand verschiedener Aspekte von Ähnlichkeiten miteinander in Beziehung setzen und automatisch als Cluster sortieren. Hierbei ist das *unsupervised clustering* mit demjenigen *clustering* zu vergleichen, das bereits verschiedene Modifikationen des Bildmaterials nutzt. Ersteres führt auf Grund der Heterogenität des vorliegenden Bildmaterials dazu, dass neben den dargestellten Figuren auch zahlreiche andere Elemente des Bildes als relevant für die Frage nach einer Ähnlichkeit einbezogen werden, z.B. die Formen des Bildausschnitts, die Henkel der Vase, Farbdivergenzen oder Ähnliches. Dies hatte in einem ersten Schritt eine Sortierung vor allem nach Farbigkeit und Gefäßformen zur Folge. In einem zweiten Schritt wurden daher alle Bilder in Graustufen konvertiert und die Figuren der Verfolger sowie Flihenden manuell annotiert. Auf diesem Wege wurden Farbdivergenzen minimiert und nur die annotierten Bildbereiche für den digitalen Vergleich herangezogen. Noch immer ist festzustellen, dass Bildinformationen wie der Graustufenwert oder Verzerrungen, die durch den Blickwinkel der Fotograf*innen oder die Dreidimensionalität des Vasenkörpers entstehen, das Ergebnis beeinflussen. Nach vorheriger Modifikation ist eine automatische Sortierung des Bildcorpus nun allerdings insofern erfolgreich, dass der Algorithmus beispielsweise korrekt fliehende Figuren oder sogar Eos und Boreas als unbewaffnete geflügelte Verfolger jeweils zueinander gruppiert (Abb. 7). Eine weitere Verbesserung der Resultate im Hinblick auf die Körperhaltungen der Figuren gelang erneut durch eine Kombination verschiedener Methoden: Stellt man dem Algorithmus allein die Posenskelette einer zuvor durchgeführten digitalen Posenerkennung zur Verfügung, bleiben bei der Sortierung die für die Fragestellung irrelevanten Bildinformationen ausgeblendet. Das Modell ist auf dieser Grundlage beispielsweise imstande, alle fliehenden Figuren, die ihre Arme eng am Körper halten, in einem *cluster* zu gruppieren. In Kombination mit vorgeschalteten Modifikationen, die den Herausforderungen im Umgang mit archäologischem Bildmaterial Rechnung tragen, hat das Bild-*clustering* in *UMAP*- oder *PixPlots* somit das Potenzial, große Bildmengen aus der Distanz zu überblicken und verschiedenartig zu strukturieren, damit neue Blickwinkel auf das Material anzuregen und in Verbindung mit den Metadaten der Bilder zu Erkenntnissen hinsichtlich chronologischer Entwicklungen, Malerzuweisungen uvm. zu verhelfen.

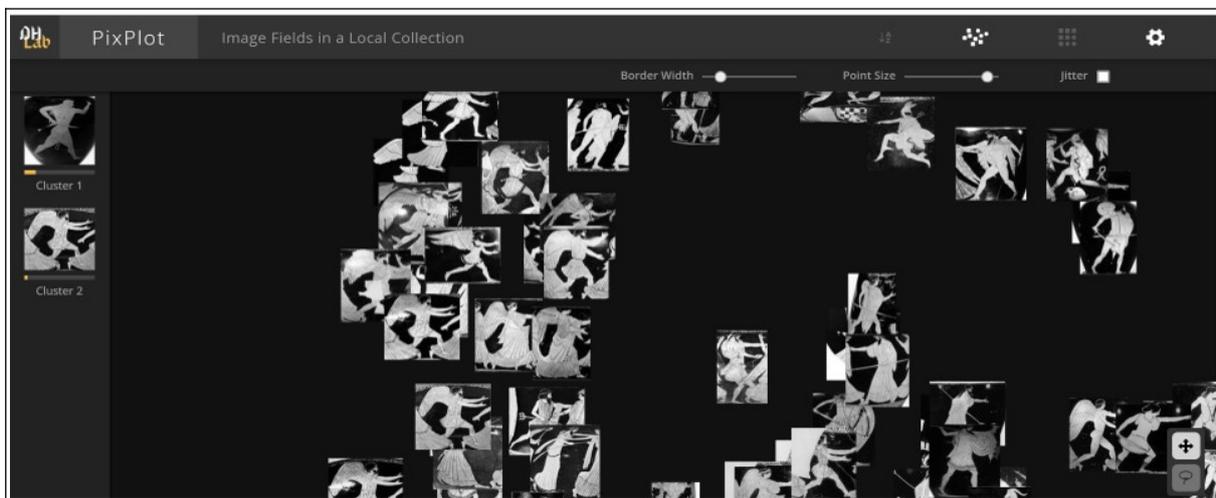


Abb. 7: Ausschnitt eines *clustering* (*PixPlot*) mehrerer Boreas- und Eos-Darstellungen.

Link zum Projekt: <https://www.klassischearchaeologie.phil.fau.de/projekt-iconographics/> (abgerufen am 05.06.2023).

41 Zum *UMAP*-Algorithmus: McInnes / Healey (2018); <https://github.com/YaleDHLab/pix-plot> (abgerufen am 23.07.2023).

Abstract 3: KI-gestützte Objekterkennung (*object detection*) in archäologischen Datensets – Erfahrungen aus archäologischer Perspektive am Beispiel des Projekts *Iconographics* an der Friedrich-Alexander-Universität Erlangen-Nürnberg

Prof. Dr. Ute Versteegen (Friedrich-Alexander-Universität Erlangen-Nürnberg)

Verfahren der KI-gestützten automatisierten Objekterkennung (*object detection*) in digitalen Bildern haben in den letzten Jahren den *Consumer*-Bereich erreicht. So ermöglicht bspw. die mobile App *Google Lens* mithilfe von sog. *Deep-Learning*-Routinen, mit der Handykamera aufgenommene Gegenstände zu identifizieren und aus dem Internet extrahierte relevante Informationen zu diesen anzuzeigen. Das von 2019–2022 an der FAU Erlangen-Nürnberg durchgeführte Projekt *Iconographics. Computational Understanding of Iconography and Narration in Visual Cultural Heritage* erforschte in Zusammenarbeit von informatischer *Computer Vision*, Digitaler Kunstgeschichte, Klassischer Archäologie und Christlicher Archäologie die Frage, wie entsprechende Verfahren der Künstlichen Intelligenz für das Analysieren historischer Bildwerke genutzt werden können.⁴² Das Teilprojekt der Christlichen Archäologie widmete sich insbesondere der Aufgabe, die für das Verständnis eines ikonographischen Bildthemas elementaren Grundbestandteile im Bild zu klassifizieren und automatisiert zu detektieren. Im Fallbeispiel wurden eigens zusammengestellte Bildcorpora frühchristlicher und byzantinischer Bilddarstellungen der Verkündigung an Maria und der Anbetung der Magier auf das Vorhandensein charakteristischer Objekte untersucht, die die figürlichen Protagonisten beispielsweise als Kleidungsbestandteile, Attribute oder grundlegende Elemente der Handlung in einer Szene auszeichnen.

Das ikonografische Bildcorpus der Christlichen Archäologie für diese beiden Kernikonographien enthält ca. 1.050 Darstellungen und damit nahezu alle bekannten Werke aus dem interessierenden Zeitraum. Anders als das Corpus der Klassischen Archäologie (vgl. Abstract 2) umfasst es einerseits sehr unterschiedliche Bildgattungen (z.B. Mosaik, Malerei, Relief, Textilien oder Toreutik), andererseits Bildträger stark variierender Formate (vom Apsismosaik bis zum Ringstein). Die im Corpus erfassten Werke decken ein chronologisches Spektrum vom dritten bis zum zehnten Jh. n. Chr. (im Westen) bzw. 15. Jh. n. Chr. (im Osten des Mittelmeerraums) ab. Da das Bildcorpus im Teilprojekt der Kunstgeschichte dieselben Themen (Verkündigung und Magieranbetung) für die westliche Kunst berücksichtigte, ergänzten diese beiden Corpora einander.

Nach Erstellen der Bildcorpora wurde das Bildmaterial mithilfe des *VGG Image Annotator (VIA)*⁴³ annotiert, d.h. mit sogenannten *labels* bzw. Objektklassen versehen. Annotiert wurden sowohl für die untersuchten Ikonographien charakteristische Objektklassen wie Körbe, Wolle, das Sitzmöbel Mariens, die Flügel Gabriels oder phrygische Mützen, Hosen und Mäntel der herbeieilenden Magier, als auch die Protagonisten als Gesamtfiguren: Maria, Gabriel und das Jesuskind. Zu Beginn wurden die zu annotierenden Bildbereiche mit Polygonen gekennzeichnet, aufgrund des hohen dafür notwendigen Zeitaufwands in Rücksprache mit den Kollegen der Informatik später mit *bounding boxes*, was den Aufwand pro *label* etwa auf ein Fünftel der Zeit reduzierte.⁴⁴

42 Vgl. zum Projekt: Madhu et al. (2019). Zu den archäologischen Teilprojekten v.a. Bendschus et al. (2022); Mührenberg et al. (2022).

43 <https://www.robots.ox.ac.uk/~vgg/software/via/> (abgerufen am 20.03.2023).

44 Zum Vorgang des Annotierens: Sager et al. (2021). Exemplarisch zur Nutzung von *bounding boxes* für die Annotation kulturhistorischer Bildcorpora: Van Zuijlen et al. (2021).

Im nächsten Schritt wurden erste Versuche der automatisierten Objekterkennung gemacht, indem ein künstliches neuronales Netz auf der anhand des *COCO*-Datensets vortrainierten Basis des *Resnet50* mittels eines sogenannten *supervised-training*-Ansatzes speziell auf unser Material weiter trainiert wurde.⁴⁵ Beim *supervised-training*-Ansatz wird das KI-Modell durch ein *dataset* angeleitet, das die Zusammenstellung der Bilder und die Expertenannotationen der relevanten Objektklassen enthält. Das Modell generiert daraufhin weitere *labels*, die in einem zyklischen Arbeitsablauf wiederum durch Fachwissenschaftler*innen korrigiert werden. Als auf *Resnet50* aufsetzende Architektur des Netzes wurde zunächst *RetinaNet* genutzt, im Laufe des Projekts dann zu *Faster RCNN* (*Faster Region-based Convolutional Neural Networks*) gewechselt, da letzteres nicht nur schneller war, sondern über verschiedene Domains hinweg bessere Ergebnisse erzielte. Für die digitale Erkennung von Objekten in antiken Vasenbildern hat *RetinaNet* Ergebnisse mit einer durchschnittlichen Genauigkeit (*average precision*) von 30,5% geliefert, wenn dem Modell andere, bislang unbekannte Bilder gezeigt wurden (Trainingsset – Testset), *Faster RCNN* mit 31,8%. Mit knapp 32% lag die Performanz des genutzten Modells für antike Vasenbilder damit nur geringfügig unter der Leistung des gleichen Modells für die Objekterkennung in modernen Fotografien, welche bei 35,9% liegt.⁴⁶ Auf das christlich-archäologische Datenset konnte dieses Verfahren zunächst nicht angewandt werden, zum einen, weil das annotierte Bildcorpus zu heterogen bezüglich des Stils, der Größen und des Materials der darin repräsentierten Artefakte war, zum anderen, weil 1.000 annotierte Bilddatensätze für das Training eines KI-Modells eine zu kleine Menge sind. Die Datenbasis für das Trainieren des Modells wurde daher so erweitert, dass neben nicht-christlichen römischen, frühchristlichen und byzantinischen Darstellungen, die andere Szenen zeigen, auch einige antik-griechische und mittelalterliche Bildwerke einbezogen wurden, die eine Darstellung derjenigen Objekte beinhalten, welche für die Ikonographien der Verkündigung und Magieranbetung charakteristisch sind. Dadurch konnte das Corpus auf annähernd 9.000 Abbildungen mit 18.000 Annotationen erweitert werden. Insgesamt lieferte die Objekterkennung mit dem *supervised-learning*-Ansatz auf der Basis dieses erweiterten Bildcorpus für die Annotationen auf Objektebene schließlich dennoch nur eine Performanz von 22,9% mit *Faster RCNN*, was wesentlich unter den Ergebnissen für moderne Fotografien und auch für das Datenset der Klassischen Archäologie lag und auf die zu große Heterogenität des christlich-archäologischen Bildcorpus zurückzuführen war.

Ein Nachteil der Nutzung von neuronalen Netzen wie *Retinanet* oder *Faster RCNN* liegt darin, dass für jegliche Anwendung auf ein abweichendes Set von Objektkategorien stets neue Expertenannotationen benötigt werden. In einem nächsten Schritt wurde in *Iconographics* daher der Ansatz der sogenannten *few-shot object detection* (*FSOD*) verfolgt, der anhand einer geringen Anzahl von Annotationen trainiert werden kann und gleichzeitig fähig ist, eigenständig neue Objektklassen zu erkennen.⁴⁷ Im Gegensatz zu den zuvor genannten neuronalen Netzen funktioniert die Objektdetektion in einem Bilddatenset hier nicht im Zuge eines zweistufigen Verfahrens, sondern Detektieren und Klassifikation werden in einem Arbeitsgang miteinander verbunden (*real-time object detection*).⁴⁸ Das Objekterkennungsmodell wird zunächst für ein Set von Objektklassen trainiert und anschließend anhand eines sich von diesem unterscheidenden Set von Objekten getestet.⁴⁹ Hierfür werden dem trainierten Modell zwei Bilder gezeigt: Zum einen ein Abfragebild (*query image*), das das gesuchte Objekt zeigt, zum anderen ein Zielbild (*target image*), in dem dieses Objekt digital erkannt werden

45 *Resnet50* wird seit 2015 für die Bilderkennung in der Computer Vision eingesetzt. Vgl. He et al. (2015). Zu *COCO* (*Microsoft Common Objects in Context*): <https://paperswithcode.com/dataset/coco> (abgerufen am 20.03.2023).

46 https://paperswithcode.com/sota/object-detection-on-coco?tag_filter=3, Modell #40 (*RetinaNet*) (abgerufen am 20.03.2023).

47 Madhu et al. (2022).

48 Redmon et al. (2016); Liu et al. (2016).

49 Zur Vorgehensweise im *Iconographics*-Projekt: Madhu et al. (2022).

soll (Abb. 8). Abhängig von der Zahl n an Abfragebildern wird diese Methode als N (*one/two/three/few*)-*shot object detection* bezeichnet, wobei im Falle mehrerer Abfragebilder diese immer das gleiche Objekt zeigen müssen. Theoretisch erlaubt es dieses Verfahren, ein trainiertes Modell zu nutzen, um beliebige Objekte in einem Zielbild zu erfassen, auf welche das Modell zuvor nicht explizit trainiert wurde (*UnSeen categories* im Verhältnis zu *Seen categories*). Verglichen mit der Leistung bei modernen Fotografien, die bei 22% liegt, war die Performance dieses Modells mit unserem christlich-archäologischen Bildmaterial noch sehr gering, nämlich gerade einmal 3,58%. Die Resultate der erstmaligen Anwendung des *few-shot-object-detection*-Ansatzes auf Bilddaten der Christlichen Archäologie haben damit die Schwierigkeiten bestätigt, die in der geringen Menge des Ausgangsdatenbestands liegen, der für das Training einer KI zur Verfügung steht.

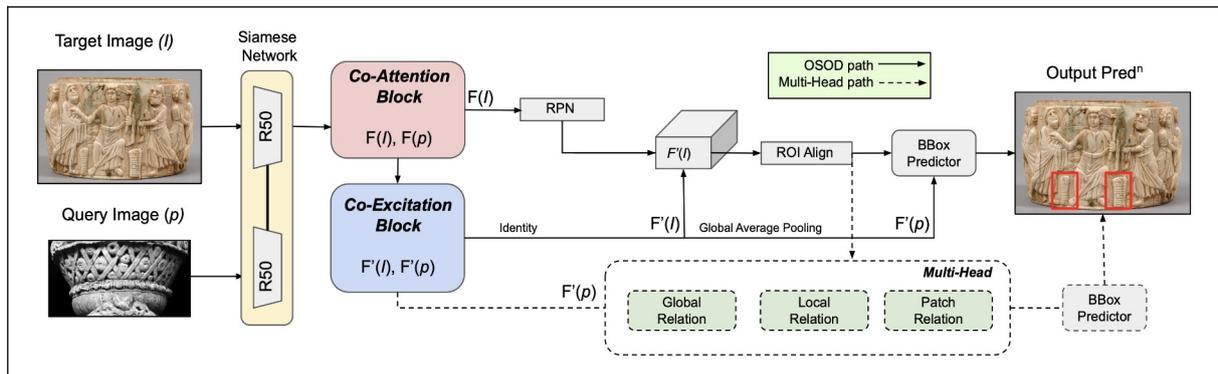


Abb. 8: Architektur des *few-shot object detection framework* im Projekt *Iconographics*.

Aus fachwissenschaftlicher Sicht waren wir uns im Projekt aber einig darüber, dass der *few-shot-object-detection*-Ansatz gegenüber denjenigen Verfahren, die eine große Menge an Bildern und Annotationen benötigen, dennoch Vorteile besitzt und prospektiv interessante Perspektiven eröffnen könnte (kein erwartbarer Anstieg der Ausgangsbasis an archäologischen Artefakten, Erzielen von Ergebnissen ohne hohen personellen und zeitlichen Aufwand). Das Verfahren erlaubt insbesondere, beliebige Objekte in einem Bild zu erfassen, auf die die KI-Modelle im Vorfeld weder hingewiesen noch trainiert worden sind. Bei einer Objektsuche können also auch Objekte in den Fokus gerückt werden, die zuvor bei der Annotation noch keine Beachtung gefunden haben, weil sie beispielsweise für ein Vorgängerprojekt nicht von Interesse waren.⁵⁰

In der transdisziplinären Diskussion zwischen den Projektpartnern stellten sich schließlich auch die Möglichkeiten statistischer Auswertungen und mathematischer Korrelationen als interessante Perspektiven für künftige Fragestellungen heraus. Die Größen der *bounding boxes* für einzelne Objektklassen wurden etwa in Relation zur jeweiligen Größe des gesamten Bildes gesetzt, um festzustellen, wie prominent einzelne ikonographische Marker ins Bild gebracht werden. Jenseits aller Fallstricke statistischer Auswertungen legten diese Auswertungen zumindest gewisse Tendenzen nahe. In einer feiner skalierten Darstellung war z.B. klar erkennbar, dass Flügel innerhalb der (apokryphen) Ikonographie der Verkündigung signifikant mehr Platz einnehmen als Körbe. Ausrechnen lassen sich auch Abstände von Objekten zu Personen oder von wiederkehrenden Objekten untereinander. Darüber hinaus sind zum Beispiel auch Kombinationsstatistiken zum gleichzeitigen Vorkommen mehrerer charakteristischer Objektklassen einer Ikonographie leicht möglich.

Ein anderer Punkt, zu dem wir mit den beteiligten Informatikern in einen interessanten Austausch gekommen sind, sind Überlegungen zu den *predictions* der KI-Modelle. Als besonders aufschlussreich erwiesen sich die aus menschlicher Sicht begangenen ‚Fehler‘ der Modelle in den Objektdetektionen (Abb. 9). So gibt es Hinweise darauf, dass die gängigen neuronalen Netze Objekte wesentlich stärker anhand von Texturen als anhand von Umrissen detektieren, was der menschlichen Wahrnehmung des

⁵⁰ Bendschus et al. (2022).

Erkennens von Objekten entgegensteht.⁵¹ Solche Fragen sind nicht nur aus bildwissenschaftlicher Perspektive von Interesse, sondern auch aus rein projektpraktischer für das Annotieren von Objekten. Gingen wir als Fachwissenschaftler*innen davon aus, dass wir durch das genaue Definieren der Objektumrisse mit Polygonen möglicherweise bessere Ergebnisse erzielt hätten als mit den rechteckigen *bounding boxes*, sprechen aktuelle Forschungen wie diese dafür, dass die genaue Umrissdefinition möglicherweise für das maschinelle Detektieren zumindest im Moment gar nicht von so großer Bedeutung ist und man sich diesen Aufwand sparen kann.

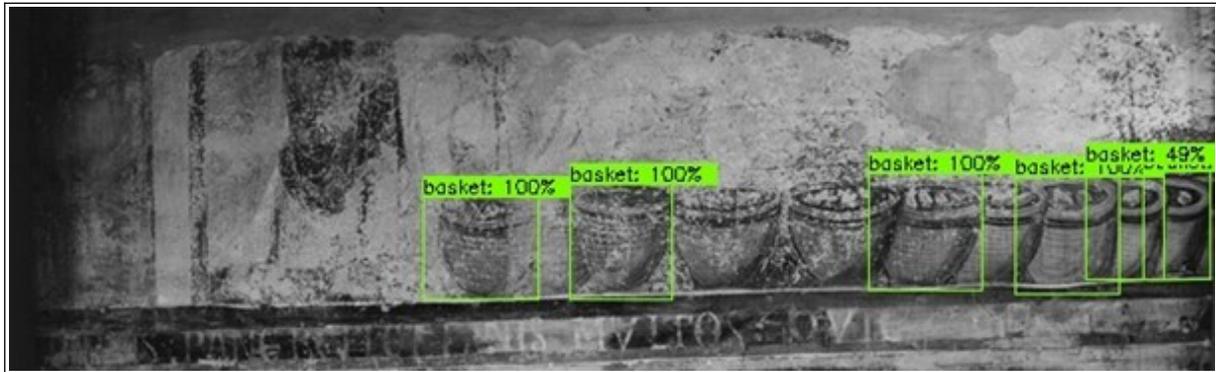


Abb. 9: Fallbeispiel einer *object detection* von Körben (*basket*), das zeigt, dass klare Umrisslinien nicht entscheidend für die Objekterkennung durch das KI-Modell waren.

Im Rahmen des *few-shot-object-detection*-Ansatzes haben wir außerdem Tests dazu gemacht, inwiefern die Genauigkeit der Lage der *bounding box* um das annotierte Objekt eine Rolle bei der Objektdetektion spielt. Hierbei wurde deutlich, dass die Modelle die besten Ergebnisse erzielten, wenn dem Modell beim Training nicht nur die Box um das relevante Objekt gezeigt, sondern der Bildausschnitt etwas größer gewählt wurde, so dass auch der umgebende Kontext mit angelernt wurde.⁵²

Link zum Projekt: <https://www.ca.phil.fau.de/forschung/projekte/iconographics/> (abgerufen am 04.06.2023).

Abstract 4: *Visuality reCOINed?* Neue epistemische Praktiken der Klassifikation und Repräsentation keltischer Münzen im Digitalen

Dr. Katja Rösler, Dr. Kerstin P. Hofmann, Chrisowalandis Deligio, M.Sc. (Römisch-Germanische Kommission des Deutschen Archäologischen Instituts, Frankfurt am Main, und Goethe-Universität Frankfurt am Main)

Dinge der Welt, in unserem Fall Münzen, werden im Zuge verschiedener Suchbewegungen, Praktiken und Konzeptualisierungen zu Erkenntnisobjekten.⁵³ Nicht selten werden sie in diesem Zusammenhang klassifiziert und repräsentiert. Hierbei spielt die Visualität gleich in mehrfacher Hinsicht eine Rolle, denn Dinge können besonders gut optisch wahrgenommen werden und das Wahrgenommene wird wiederum mit Bedeutung versehen. Ferner dienen neben Texten Visualisierungen zur Sichtbarmachung und Kommunikation von Wissen über Dinge. Ihre Relevanz nimmt im wissenschaftlichen Kontext gerade auch im Zuge der Digitalisierung immer mehr zu. Am Beispiel dreier keltischer Münzserien fragen wir uns im Rahmen des BMBF-Projektes *ClareNet – Klassifikationen und Repräsentationen für Netzwerke. Von Typen und Merkmalen zu Linked Open*

51 Geirhos et al. (2019).

52 Madhu et al. (2022).

53 Lobinger (2015); Hilgert et al. (2018).

Data bei keltischen Münzprägungen, inwiefern sich unsere Wissenspraktiken und Visualität dabei verändern.⁵⁴

Der amerikanische Kulturhistoriker Hal Foster hat 1988 einen für die *visual culture studies* wegweisenden Diskussionsbeitrag geleistet, indem er zwischen den Mechanismen und Gegebenheiten des Sehens – *vision* – und den historischen Techniken, ihren diskursiven Konstruktionen und Determinierungen – *visuality* – unterschied.⁵⁵ Seitdem werden *vision* und *visuality* in den Kulturwissenschaften oft mit einem medienanalytischen Ansatz für unsere Alltagskultur untersucht. Unseres Erachtens kann man viele der Erkenntnisse aber auch gewinnbringend ins Feld der archäologischen Wissenschaften übersetzen. So lassen sich nicht nur viele Arten des Sehens, sondern auch des Blickens unterscheiden. Statt den in der feministischen Bild- und Medienwissenschaft offengelegten *male gaze*, der sich mit aus männlicher heterosexueller Perspektive erstellten (Re-)Präsentationen von Frauen auseinandersetzt, stellen wir erste Erkenntnisse zu dem von uns postulierten *classification gaze* vor. Diese Art des Blickens wurde im wissenschaftlichen Feld der Numismatik von uns bei der Zuordnung von Objekten in bereits vorhandene Klassen beobachtet. Ferner zeigen wir, dass dieser klassifikatorische Blick durch Künstliche Intelligenz (KI) reproduziert werden kann und benennen die Gründe hierfür.

Die Numismatik der Antike ist eine spezialisierte Wissenschaft mit einem etablierten Kanon an Verfahren zur Münzbestimmung und -darstellung. Münzen waren Bestandteile der Kunstsammlungen des Spätmittelalters und der Frühen Neuzeit, bevor sie in Münzkabinetten zusammengetragen, ausgestellt sowie inventarisiert und katalogisiert wurden. Die keltischen Münzen waren Teil dieser Sammlungen, wurden aber oft vernachlässigt, da man sich zunächst bevorzugt mit der klassischen Antike auseinandersetzte. Zudem sind die keltischen Münzen im Vergleich zu den römischen und griechischen Münzen chrono-, chorologisch und ikonographisch schwerer bestimmbar. In Katalogen wurden sie exemplarisch abgebildet, es erfolgte eine systematische Ansprache und es wurden Klassifikationen bzw. Typologien erstellt.⁵⁶

Ihre wissenschaftliche Analyse und Repräsentation erfolgten dabei bis heute entlang von Verfahren, die für andere, vor allem griechische und römische Münzen, entwickelt wurden, und das, obwohl die ‚keltischen‘ Münzen sehr augenscheinliche Herausforderungen mit sich bringen: Sie haben eine größere Varianz hinsichtlich ihrer Form, ihrer Beschaffenheit und der Münzbilder innerhalb von einzelnen Münzserien.⁵⁷ Folglich werden an vorwiegend standardisierten Objekten entwickelte epistemologische Praktiken und Repräsentationsweisen auf deutlich weniger standardisierte Objekte angewendet.

Mit Hilfe eines Experiments haben wir erkundet, wie die Mechanismen des von uns postulierten *classification gaze* beim Einordnen von Münzen in bestehende Klassen funktionieren. Hierzu haben wir das Vorgehen eines Numismatikers und eines trainierten *Convolutional Neural Networks* (CNNs, im Folgenden zusammenfassend als KI-Modell bezeichnet) verglichen. Beide wiesen anhand von Fotos Münzen, die der *civitas* der *Coriosolitae* zugeschrieben werden, den in der Numismatik anerkannten sechs Klassen zu. Die Münzen stammen aus dem bislang größten bekannten keltischen Hortfund, le Câtillon II, der insgesamt ca. 70.000 Münzen umfasst. Diese Klassifikation diente als Grundlage für die vom Domain-Experten und seinem Team erstellte *ground truth*, die wiederum für das Experiment als Referenz benutzt wurde, zum einen für das Training des KI-Modells, zum anderen zur Überprüfung der Klassifizierungen durch Numismatiker und KI-Modell.

54 Wigg-Wolf et al. (2022).

55 Foster (1988), ix.

56 S. Göbl (1987). Mit dem 1892 publizierten MünzAtlas von Henri de la Tour liegt eine seitdem oft genutzte Klassifikation und Repräsentationsform der keltischen Münzen vor.

57 Vgl. Colbert de Beaulieu (1973), 14–17.

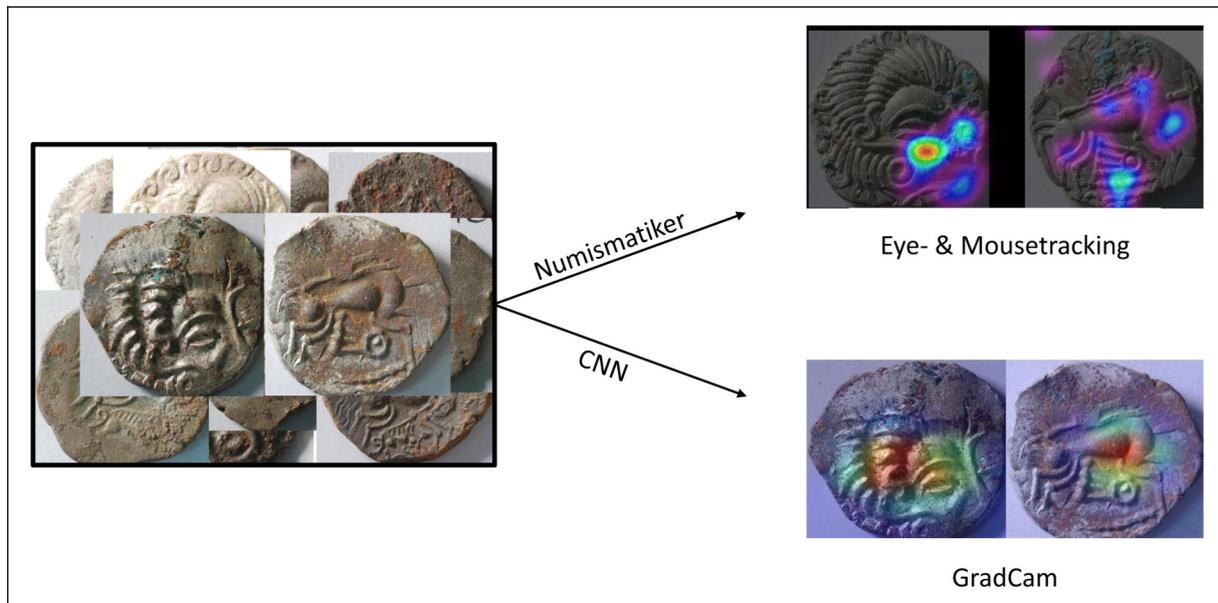


Abb. 10: Visualisierung der zur Entscheidung beitragenden Merkmale durch *heatmaps*.

In beiden Fällen werden *black-box*-Entscheidungen getroffen. Um diese auf einer visuellen Ebene vergleichen zu können, verwendeten wir auf Seite des Numismatikers *Eye- und Mousetracking*⁵⁸ und auf Seite des KI-Modells das sog. *Gradient-weighted Class Activation Mapping*⁵⁹ (Grad-CAM). Durch die Visualisierung der ausgewählten Merkmale war es möglich, die Unterschiede zwischen menschlichem und maschinellem Sehen besser nachvollziehen und dann analysieren zu können (Abb. 10). Die erzeugten *heatmaps* zeigten jene Bereiche, die für das Ermitteln der Klasse relevant waren bzw. zur Entscheidung der Zuordnung maßgeblich beigetragen haben. Der Numismatiker suchte erwartungsgemäß die Merkmale auf Avers und Revers, die in der Numismatik für die Klasseneinteilung als relevant gelten. Es handelt sich also um einen durch kommuniziertes Wissen gelenkten kursorischen Blick. Die Analyse der *heatmaps* des KI-Modells zeigte hingegen, dass auf dem Avers weniger Merkmale beachtet wurden, jedoch größtenteils jene fokussiert wurden, die zentral und über den gesamten Datensatz vorhanden sind. Auf dem Revers fokussierte das KI-Modell ebenso klassenübergreifend auf ein Merkmal, den (oft) mittig im Münzbild dargestellten Pferdekörper, der vom Numismatiker hingegen gar nicht oder kaum beachtet wurde. Sein Blick galt vielmehr den in den Randbereichen des Münzbildes auftretenden Beizeichen (Abb. 11).

Die genannten Visualisierungsmethoden helfen also, die *black-box*-Entscheidungen zur Münzbestimmung bzw. -sortierung ein wenig zu erhellen und so die Klassifizierungen nachvollziehbarer und vergleichbarer zu machen. Indem nicht allein die Richtigkeit der Merkmalsgewichtung im Sinne der *ground truth* das Ziel des Experiments war, konnten durch den direkten Vergleich Praktiken der menschlichen Wahrnehmung, und in diesem konkreten Fall das spezialisierte (Nicht-)Sehen des Numismatikers thematisiert und reflektiert werden. Hierbei waren besonders die unerwarteten Entscheidungen des KI-Modells von Interesse.

58 Voßkühler (2009).

59 Servaraju et al. (2017).



Abb. 11: Der *classification gaze* des Numismatikers und der KI mit Hilfe von *heatmaps* im Vergleich. Links die Fokussierungen des Numismatikers, rechts die des KI-Modells. Erkennbar ist, dass der Fokus des Numismatikers und des KI-Modells auf dem Avers (links) im Bereich des Gesichts liegt. Auf dem Revers (rechts) aber wandert der Fokus des Numismatikers in den Bereichen der Beizeichen umher, während der Fokus des KI-Modells auf dem zentral dargestellten Pferdekörper liegt.

Aus unserem Experiment lassen sich folgende Aussagen ableiten: Die KI ersetzt die Expert*innen nicht, sondern kann für sie eine unterstützende Rolle übernehmen. *Supervised learning* benötigt eine *ground truth*, die von Expert*innen erarbeitet werden muss. Es können dann gezielt abweichende Klassenzuweisungen für die Weiterentwicklung der KI, aber auch der numismatischen Klassifikationen genutzt werden. Visualisierungen der Vorhersagen des KI-Modells können auf andere Bereiche und Merkmale aufmerksam machen und erlauben es zudem, die Qualität des KI-Modells zu bewerten. Nach Überprüfung können die neuen Erkenntnisse zur Gewährleistung des Wissenstransfers z.B. in neuen numismatischen Idealzeichnungen Eingang finden. Zusammenfassend lässt sich also festhalten: Der *digital turn* hat und wird unsere Visualität neu prägen, aber auch Forschungstraditionen und Pfadabhängigkeiten prägen umgekehrt den *digital turn* und damit das neue digitale Sehen. Dieses wird nicht nur durch neue Technologien, sondern auch durch etabliertes Wissen und Praktiken beeinflusst.

Im Projekt *ClaReNet* werden wir weitere Versuche mit KIs durchführen, wobei auch der Vergleich verschiedener Klassifikationen und Stempel eine Rolle spielen wird. Ferner setzen wir uns mit

unsupervised learning auseinander. Der Vergleich der Verfahren von Mensch und KI regt zudem an, Wissenspraktiken sowie Möglichkeiten und Grenzen der Transparenz von Entscheidungsfindungen weiter zu erforschen. Die in der Numismatik entwickelten standardisierten Visualisierungstechniken und deren Umsetzung in der digitalen Fotografie sind für die KI und ihre automatisierte Bilderkennung besonders vorteilhaft. Durch digitale Methoden der Klassifikation und Repräsentation, die Bilder stärker als Texte und Messwerte berücksichtigen, ist ein neues Sehen, im Sinne einer *visuality reCOINed*, für die keltischen Münzen möglich.

Link zum Projekt: <https://clarenet.hypotheses.org/> (abgerufen am 04.06.2023).

Gruppeninterview zur Sektion *Ikono-graphie und Mustererkennung*

Sprecherinnen und Sprecher:

Elisabeth Günther (Universität Heidelberg), Moderation

Chrisowalandis Deligio (Goethe-Universität, Frankfurt am Main)

Kerstin P. Hofmann (Römisch-Germanische Kommission des Deutschen Archäologischen Instituts, Frankfurt am Main)

Marta Kipke (Georg-August-Universität, Göttingen)

Martin Langner (Georg-August-Universität, Göttingen)

Corinna Reinhardt (Friedrich-Alexander-Universität, Erlangen-Nürnberg, jetzt Universität Zürich)

Katja Rösler (Römisch-Germanische Kommission des Deutschen Archäologischen Instituts, Frankfurt am Main)

Ute Verstegen (Friedrich-Alexander-Universität, Erlangen-Nürnberg)

Elisabeth Günther:

Ich möchte alle Anwesenden herzlich zu unserem Gespräch begrüßen und darf direkt mit der ersten Frage beginnen: Welches Potenzial besitzen Mustererkennung und künstliche Intelligenz für Forschung und Lehre in der Digitalen Archäologie, also insbesondere im Bereich der Bild- und Objektwissenschaften, mit dem wir uns im Rahmen der Tagung auseinandergesetzt haben?

Martin Langner:

Mit dem Potenzial sind gleichzeitig Herausforderungen verbunden. Ich will gerne das, was ich sage, auf Bild *und* Objekt beziehen, weil wir am Institut für Digital Humanities an der Universität Göttingen ja genauso stark im 3D-Bereich wie im 2D-Bereich arbeiten, und der 3D-Bereich etwas komplizierter ist als der 2D-Bereich (vgl. Abstract Nr. 1).

Worum es aber, glaube ich, grundsätzlich geht, ist, visuelle Information direkt aus den Bildern und Objekten zu gewinnen. Wir versuchen also unerschwerlich in die Dateneingabe einfließende ästhetische Erfahrungen oder Bewertungen der *user* oder Vorwissen der *user*, die immer in die Datenaufnahme einfließen, zu objektivieren. Das heißt Kennzahlen, bestimmte formale Eigenheiten so zu erfassen, dass sie für andere besser nachvollziehbar sind. Wir setzen das zum Beispiel mit der Beazleyschen Methode der Malerzuweisung bei den Vasenbildern um,⁶⁰ aber auch in vielen anderen Fällen, zum Beispiel, wenn wir versuchen, den Winter⁶¹ neu zu machen, um die Typologie der Terrakotten besser zu

60 <https://www.uni-goettingen.de/de/598165.html> (abgerufen am 03.06.2023).

61 Winter (1903).

systematisieren usw.⁶² Und da sehe ich ein großes Potenzial in der Entstehung solcher auf visueller Ähnlichkeit beruhender Netzwerke.

Das Konzept der Ähnlichkeit resultiert ja hier auf Verknüpfungen über bestimmte Maßzahlen. Und dieses Konzept, das die Klassische Archäologie immer interessiert hat, schon in der Antike, und das besonders am Ende des 19. Jh. mit dem Corpusgedanken und der Ordnung des Materials neu thematisiert wurde – das muss ich gar nicht so genau aufzählen –, wurde etwa in der Skulpturenforschung sehr stark ausdifferenziert⁶³; bei den Vasenbildern nicht weniger. Ikonographie und Ikonologie sind im Grunde verschiedene Abstufungen von Ähnlichkeit und solchen Formalisierungsprozessen. Und ich verspreche mir, dass verschiedene Eigenarten wie Zeitstil, Motivik oder, bei den Vasenmalern, der Individualstil an neuronalen Netzen trainiert werden können, um diese Gruppierungen von Bildern zu visuellen Netzwerken zu verbinden, um über die aktuellen Möglichkeiten der *image collection exploration*, also das, was Sie von Google kennen, wenn Sie bei Ihrer Suche auf ‚Bilder‘ klicken,⁶⁴ hinauszukommen. Das heißt also, um das Thema Ihrer Tagung aufzugreifen: Das ‚Neue Sehen‘ ist dann vor allem ein schnelleres Sortieren und Gruppieren von Bildwerken, die bisher nur in Beschreibungen erfasst sind.

Neue multimodale Methoden hingegen, die Sie vielleicht von der Firma *openAI* kennen,⁶⁵ gehen jetzt wieder auf die sprachliche Ebene und arbeiten damit, dass Wörter ein Wortfeld haben und einen Assoziationsraum. Und der wird dem Assoziationsraum der Bilder gegenübergestellt. Und hier liegt die eigentliche Stärke der neuronalen Netze, dass sie nämlich nicht streng kategorial vorgehen, sondern über Ähnlichkeiten. Und diese können auf vielen verschiedenen Ebenen liegen. Eben genau das ist der große Vorteil, aber auch gleichzeitig der größte Nachteil dieser Verfahren, weil wir zum Beispiel bei den Malerzuweisungen hinter die Beazleysche Methode zurückfallen, indem wir wieder assoziativ Dinge, die ähnlich sind, zusammenfassen, wobei es einen viel größeren Aufwand bedeutet festzustellen, ob diese Ähnlichkeit durch den Zeitstil bedingt ist, durch das Motiv, durch eine Werkstattähnlichkeit oder durch einen Gefäßstil (weil eben auf einer bestimmten Gefäßform ein bestimmter Platz für das Bild vorgesehen ist).

Den größten Vorteil sehe ich in einer Bestimmung von Assoziationsräumen, die durch bestimmte Vorzeichen oder *framing* auch gelenkt werden können; und den größten Nachteil darin, dass Formen des Strukturellen entweder wieder in einen Strukturalismus münden, der schlimmer ist als das, was wir in den 30er und 40er Jahren des letzten Jh. hatten, oder aber so banal, dass wir dafür eigentlich keinen Computer brauchen.

Corinna Reinhardt:

Ich möchte zunächst den Aspekt der Ähnlichkeit aufgreifen, der mir sehr wichtig erscheint. Diese Methoden ermöglichen und erfordern es intensiv und zum Teil neu darüber zu diskutieren, was wir Archäologinnen und Archäologen eigentlich unter *Ähnlichkeit* verstehen, die die Grundlage für viele Analysen ist. Wie sich hierzu die Arbeitsweisen digitaler Methoden verhalten, ist eine zentrale Frage. Gehen diese in eine ähnliche Richtung und/oder können sich beide Bereiche produktiv ergänzen? Worin ich ebenso ein großes Potenzial sehe, ist die Möglichkeit große Materialmengen zu vergleichen. In der Bildanalyse der griechischen Vasenmalerei wäre für viele Fragen der Einbezug einer großen Materialbasis wichtig. Dennoch sieht man, und das ist natürlich durch die Arbeitszeit bedingt, dass es oft Einzelstudien gibt, die sich eigentlich mit relativ wenig Material beschäftigen, dabei dennoch übergreifende Ergebnisse anstreben. Hier erhoffe ich mir vom Einsatz digitaler Methoden, dass man mehr Material einbeziehen kann. Damit in Verbindung steht natürlich auch, dass es durch die Mustererken-

62 <https://www.uni-goettingen.de/de/598167.html> (abgerufen am 03.06.2023).

63 Langner (2017).

64 <https://images.google.com> (abgerufen am 03.06.2023).

65 <https://openai.com/research/clip> (abgerufen am 03.06.2023); Radford et al. (2021).

nung möglich sein wird – so hoffe ich – Bildverbindungen zu erhalten, die man nicht in seinem Kenntnishorizont hat und die einen in der eigenen Fragestellung weiterbringen (vgl. Abstract Nr. 2). Auf einer generellen Ebene sehe ich, dass wir durch diesen interdisziplinären Austausch in unseren kleinen Fächern die Gelegenheit haben, unsere Methoden in eine größere Diskussion einzubetten und weiterzuentwickeln, ebenso wie wir unsere Bildkompetenz für den modernen Umgang mit Bildern in der *Computer Vision* fruchtbar machen können.

Ute Versteegen:

Wie unschwer zu vermuten, habe ich Ähnliches zu ergänzen wie Corinna Reinhardt. Das ist ganz klar, denn wir arbeiten ja im gleichen Projekt, *Iconographics*⁶⁶ (vgl. Abstract Nr. 3). Meiner Ansicht nach ist es einer der großen angesprochenen Mehrwerte, mehr über Ähnlichkeiten und die Verbindung von Ähnlichkeiten zu unseren Klassifikationssystemen und Methoden nachzudenken. Und dass wir in diesem Zusammenhang eben auch über unsere eigenen, sagen wir mal archäologischen Teildisziplinen, über die Archäologie als Ganzes und über verwandte Fächer wie die Kunstgeschichte sogar hinausgehen und in den Diskurs mit anderen Disziplinen treten, die sich auf ganz unterschiedliche Art und Weise mit Mustern und Ähnlichkeiten beschäftigen, um so unsere bildwissenschaftliche Forschung enorm auszuweiten. An dieser Stelle sind für uns die Diskussion und die Zusammenarbeit mit der Informatik von sehr großem Interesse. Wenn man sich zum Beispiel überlegt, wie Martin Langner das vorhin gesagt hat, wie man Ähnlichkeiten über bestimmte formale Faktoren definiert und hierfür Umrisse, Größenverhältnisse und Maße von Objekten in den Blick nimmt, können seitens der Informatik Beobachtungen dazu eingebracht werden, inwieweit Ähnlichkeiten nicht auch kontextsensitiv sind, also Ähnlichkeiten je nach unterschiedlichen Kontexten verschieden gesehen und bestimmt werden können. Das ist etwas, das wir von unserer Seite in der Bildwissenschaft wunderbar aufgreifen können und wo wir eben auch Potenzial haben, eigene Methoden weiterzuentwickeln. Das sehe ich als einen ganz großen Vorteil an, auch für die Lehre.

Und vielleicht ein ganz banales praktisches Element. Was wir bislang noch nicht angesprochen haben, ist, dass wir eben durch dieses Wechseln zwischen *distant viewing* und *close viewing* die Möglichkeit haben, sehr große Bildermengen einfach auch mal vorzusortieren. Corinna Reinhardt hat gerade angesprochen, dass man die Bilddatenbasis vergrößern kann. Wir haben nun die Möglichkeit, bei großen Bildermengen eine Vorsortierung automatisiert vornehmen zu können, um dann gezielt hinein zu zoomen und innerhalb von kleineren Teildatensätzen innerhalb eines größeren Pools eine Art *close viewing* zu machen und die entsprechenden wissenschaftlichen Fragen zu stellen, die uns interessieren.

Marta Kipke:

Um das bisher Gesagte aus der Perspektive Lehre zu ergänzen: Verallgemeinernd bieten Mustererkennung und Künstliche Intelligenz einen guten Anlass, um nicht nur Forschungsmeinungen und Methoden zu hinterfragen, sondern auch die Daten, mit denen man eigentlich arbeitet. Vor allem in den Geisteswissenschaften ist es häufig schwierig, statistische Verfahren oder quantitative Zugänge zu vermitteln. Das funktioniert sehr viel einfacher, wenn man tatsächlich die Bilder und die Datensätze vor sich hat und mit Material arbeitet, das unmittelbaren Bezug zu der geisteswissenschaftlichen Forschung hat. Ich denke, dass diese Sensibilisierung für die Daten auch etwas ist, was nicht nur Potenzial für zukünftige Arbeiten an dieser Schnittstelle zwischen Geisteswissenschaften und Informatik birgt, sondern eine zwingend notwendige Kompetenz ist.

Etwas anekdotisch möchte ich außerdem gerne hinzufügen, dass das Hinterfragen bestehender Methoden und Forschungsmeinungen für Studierende in der kritischen Auseinandersetzung mit neuronalen Netzen und den Ergebnissen von Experimenten häufig barrierefreier ist, als mit der tatsächlichen Forschung. Sie haben keine Probleme damit, *biases* in den Daten aufzuzeigen und zu erkennen. Dadurch lernen Sie, diese *biases* auch in der Forschung zu erkennen.

66 <https://www.ca.phil.fau.de/forschung/projekte/iconographics/> (abgerufen am 03.06.2023).

Ein gutes Beispiel aus unserem Projekt *EGRAPHSEN. Möglichkeiten und Perspektiven der digitalen Malerzuweisung bei attischen Vasen* (vgl. Abstract Nr. 1) sind *biases* im Trainingsdatensatz in Bezug auf die Gefäßform oder das Motiv. Den Studierenden leuchtet ein, weshalb wir diese *biases* in den Daten reduzieren müssen und wie wir dies tun. Sie fragen sich dann jedoch auch: Hatte nicht Beazley selbst in seinen Zuweisungen schon diesen *bias*? Haben wir nicht diesen *bias* eingepflanzt, indem wir ihn als *ground truth* definieren? Ich denke, dass es zahlreiche Möglichkeiten gibt, an konkreten Beispielen in unseren jeweiligen Fächern für diese Fragen und Probleme zu sensibilisieren.

Kerstin P. Hofmann:

Am Aspekt der Forschung und Lehre finde ich so spannend, dass wir ja im Endeffekt Forschungspraktiken oder auch Axiome, Methoden, die wir gar nicht mehr hinterfragen, durch die Künstliche Intelligenz in ein anderes Medium ‚übersetzen‘ – oder versuchen zu übersetzen. Hierbei wird deutlich, dass die Erkenntniswege selbst stärker beschrieben werden müssen. Im Zusammenhang mit dem Konzept der Ähnlichkeit ist daher auch zu fragen: Wann wird etwas als identisch betrachtet und wann wird eine Differenz erkannt? Wie werden diese gewichtet? Wann spricht man von Präsenz/Absenz oder Spektren? Mit welchen Toleranzen arbeiten wir? In welchen Prozessen bzw. Schritten passiert dies? Wenn wir etwas in ein anderes Medium oder eine andere Technik übersetzen, führt dies dazu, dass man sich auch selbst hinterfragt. Für die Bild- und Objektwissenschaften ist dies besonders wichtig, da wir häufig auch mit einem alltäglichen Wahrnehmen konfrontiert sind, das wir dann in die Wissenschaft transportieren, aber diese *biases* häufig nicht ausbuchstabieren können. Erkenntnistheoretisch ist dies also meines Erachtens eine große Chance, die auch sehr gut für die Lehre genutzt werden kann.

In den Digital Humanities ist man nämlich durch die Algorithmen zwangsverpflichtet, Forschungspraktiken offenzulegen und in Prozessen darzustellen. Sodann kann man mit diesen experimentieren und beobachten, welche Herausforderungen entstehen, ob ähnliche oder andere Resultate erzeugt werden.

Sowohl in der Lehre als auch in der Forschung ist es also wichtig zu fragen: Was ist eigentlich Identität, wie verhält sich Identität zu Ähnlichkeit und wann wird etwas different? Manchmal sind hierfür auch minimalste Änderungen ausschlaggebend und dabei können dann auch eventuell verschiedene Ebenen wie Kontext, Farbe oder Stil, Motivik etc. eine Rolle spielen. Diese Aspekte kann man so separieren, strukturieren und zugleich herausfinden, wie sie in Kombination wieder zusammenwirken.

Katja Rösler:

Ich möchte kurz auf den Punkt eingehen, den Kerstin Hofmann gerade angesprochen hat, nämlich zu dem Reflexiven in der Lehre, vor allen Dingen aber natürlich auch in der Forschung. Gerade durch den Einbezug von KI und der Informatik müssen Transformationen und Konfigurationen von Objekten und Bildern, die bereits geschehen sind, erneut gemacht werden, weil es auch für die KI ein *preprocessing* gibt, sie also nicht auf fertige Daten zugreift. Wir machen somit erneut etwas mit unseren Daten für die KI. Es gibt also wieder Konfigurationen/Transformationen, die dann reflektiert werden können und sollten und natürlich im Abgleich mit den vergangenen wieder diskutiert werden müssen. Das ist der hohe Wert der Reflexion, den wollte ich noch mal hervorheben, und natürlich, weil wir jetzt den Begriff thematisiert haben: Wir holen uns die IT mit an den Tisch und wir müssen uns mit der Sprache und dem Stand der Forschung und der Art der Forschung und der Herangehensweise der IT auseinandersetzen. Es geht also nicht nur um die Zusammenarbeit zwischen den Archäologien, wobei ich begrüße, dass diese enger wird und der Austausch auf dieser Ebene dadurch wieder intensiv ist, sondern auch mit der IT, mit der man sich genau auseinandersetzen muss, weil unsere Vorstellungen und Wünsche von archäologischer Seite manchmal weiter gehen als das, was die IT leisten kann.

Martin Langner:

Ich kann an alles noch mal anknüpfen und würde es gern auch in zweierlei Hinsicht weiterführen. Vor allen Dingen ist die große Materialmenge und dieses wichtige Zusammenspiel von *distant* und *close*

viewing – wie ja beim *distant reading* – eigentlich der Witz bei der Sache. Das wollte ich klar betonen. Aber in der sprachlichen sowie vor allem in der methodischen Auseinandersetzung mit der Informatik, sind es eigentlich die von Kerstin Hofmann angesprochenen Erkenntniswege, die wir richtig in den Blick nehmen können und wo wir gezwungen werden – wie es Marta Kipke für die Lehre formuliert hat – uns alles nochmal genau anzuschauen. Diese Verliebtheit in Ideen, die in den 80er und 90er Jahren vor allen Dingen auch in englischsprachiger Forschung so beliebt war, die funktioniert hier nicht mehr. Und auf dieser Ebene von Identität, Ähnlichkeit und Differenz, glaube ich, ist es noch viel strikter, als Kerstin Hofmann es gesagt hat. Was relativ gut funktioniert, ist *object detection* und *scene recognition*. Und da gibt es nun keine eindeutigen Ergebnisse im Sinne von: Das ist Herakles, sondern es gibt das Ergebnis: Das ist zu soundsoviel Prozent Herakles. Und das neuronale Netz kann nur erkennen, was es vorher schon hundertmal gesehen hat. Was aber, finde ich, für die Archäologie oder für die Wissenschaften neu und sehr wichtig ist, ist die Variationsbreite, also, dass wir uns um diese Prozentzahlen genauer bemühen und den Versuchsaufbau nicht auf Identität, sondern stärker auf Variationsbreite hinaus ausbauen. Wenn uns das nämlich gelingt, dann könnte man bei Bildern, wie zum Beispiel den *Brettspielern* des Exekias,⁶⁷ feststellen, dass es ungewöhnlich viele Bilder gibt, die alle sehr ähnlich aufgebaut sind und eine geringe Variationsbreite haben, während *Herakles im Löwenkampf* sich nicht so gut typisieren lässt. Man kann sich fragen: Woran liegt es eigentlich? Und wenn wir uns dann die Summe aller erhaltenen Vasenbilder – und jetzt komme ich wieder zum *distant viewing* – als visuelles Netzwerk vorstellen, wie ich eingangs gesagt habe, und als unterschiedlich gewichtete Knoten und Kanten, dann können wir etwas zur allgemeinen Bildwelt der Zeit beitragen und auch zur visuellen Vorstellungskraft der Betrachter. Also das, was wir eigentlich immer schon wollten, nämlich nicht die Frage: Ist genau das dargestellt, sondern: Wie kann sich die Epoche in den Bildprägungen, die in den Köpfen der Menschen existieren, ein bestimmtes Thema überhaupt nur vorstellen? Und damit hätten wir einen methodischen Zugang zur Wahrnehmung und zu Wahrnehmungsspezifika, der uns bisher versperrt war, weil kein Mensch allein die gesamte Überlieferung überblicken kann.

Kerstin P. Hofmann:

Nur ganz kurz zur Ergänzung: Der nächste Schritt, der eine ganz große Herausforderung darstellt, ist, wenn wir bei Objekten oder auch bei Bildern material- bzw. gattungsübergreifend arbeiten wollen. Man kann dann nicht mehr nur ein Corpus (entweder Keramik oder andere Materialien) in den Blick nehmen, sondern, wenn man Mustererkennung ernst nimmt, muss man auch Keramikverarbeitung mit Toreutik und anderen Bereichen vergleichen – und dies aber nicht assoziativ, sondern in geregelten Prozessen.

Die Frage nach Varianz, die Martin Langner gerade angesprochen hat, ist ebenfalls ein sehr interessantes Thema. Können wir eventuell bestimmte Zyklen oder Entwicklungen in der Motivik erkennen? So könnte man z.B. innovationstheoretisch ‚Bildserien‘ betrachten,⁶⁸ so z.B.: Am Anfang ist man experimentierfreudig und es gibt keinen festen Kanon. Dann entwickelt sich ein fester Kanon, der immer symbolischer und abstrakter wird, da das Dargestellte quasi automatisch erkannt wird, usw. All solche Fragen könnte man angehen, um dann eben auch zu untersuchen, wie sich verschiedene Materialgattungen beeinflussen, ob es Rückkopplungen und Ähnliches gibt, weil wir ja sonst immer nur einen sehr selektiven Bereich des Erkennens berücksichtigen. Gerade hier liegt meines Erachtens auch noch längerfristig ein Riesenpotenzial, das momentan in den Anfangsstadien der Bild- und Mustererkennung noch nicht genutzt werden kann, aber wohin es tendenziell hingehen wird. Es stellt sich dann immer wieder die Frage, wie durch bestimmte Abstraktionen und Schritte oder wie mit bestimmten Wahrscheinlichkeiten von uns Kausalitäten hergestellt werden, die sodann eventuell wieder zu hinterfragen sind.

67 Buchholz (1987); Pfisterer-Haas (2004).

68 Vgl. Hofmann / Patzke (2012).

Ute Versteegen:

Ja, ich kann aus meiner Perspektive an das von Kerstin Hofmann Gesagte wunderbar anschließen, weil Kerstin in ihrem Dreischritt quasi die visuelle Entwicklung der frühchristlichen Bildkultur beschrieben hat. Denn unser Teilprojekt war im Rahmen des ikonographischen Projekts, bei dem Corinna Reinhardt auch dabei war, im christlich-archäologischen Teilbereich genau diesen Fragen gewidmet. Denn unser Material ist wahnsinnig schwierig. Wir haben ein *small dataset*, einen besonders kleinen Datenbestand, wenn wir uns eine bestimmte ikonographische Entwicklung anschauen wollen, also die Genese einer Ikonographie, vielleicht die Entstehung eines Kanon, Varianten eines Kanon in verschiedenen Regionen, die sich über einen weiten Zeitraum erstrecken und weiterentwickeln und von verschiedenen Akteursgruppen ganz unterschiedlich gesehen werden können, usw. Unsere Datenbasis ist zudem außerordentlich heterogen, das heißt, wir haben uns mit diesen ganzen Materialunterschieden herumgeschlagen – wir hatten alles von Mosaik, Sarkophag, Ringstein, Elfenbein, was zu einer Ikonographie gehört. Auch sehr große Varianten in der Größe der Bilddatenträger, auf die die Ikonographie aufgebracht war. Und dann, das kommt auch noch dazu, sind diese Bilder natürlich im Raum angebracht oder auf dreidimensionalen Bildträgern aufgebracht. Auch diese ganz unterschiedlichen Kombinationen in räumlichen Situationen führten zu einem extrem heterogenen Material, das es auszuwerten galt. Und das war natürlich eine wirklich sehr, sehr große Herausforderung und hat – nicht völlig unerwartet – auch nicht immer so gute Ergebnisse erbracht wie das Corpus von Corinna Reinhardts Teilprojekt, das einer Artefaktgattung gewidmet und zeitlich sehr eng geschlossen war. Unseres reichte hingegen vom 3. bis zum 15. Jh.; da ist es völlig klar, dass es große Variationen gibt.

Aber die Fragen, die sich anschließen lassen und die gerade von Kerstin Hofmann auch schon aufgeworfen worden sind, sind für uns auch hochinteressant im Zusammenhang mit dem Beginn, mit der Genese einer frühchristlichen Bildersprache. Das Superspannende ist ja, dass wir uns mit der Genese einer visuellen Kultur beschäftigen, die sozusagen versucht, ein Textcorpus zu ergänzen, das vorher nicht bebildert war, und dafür Motive, Kombinationen von Figuren, figurale Szenen und Objekte zu neuen Bildfindungen zusammenbringt, also sehr innovativ arbeitet. Und da dann die Feinheiten der Semantik, die man nur in der visuellen Kultur jenseits der textlichen Basis herausbekommen kann, über automatisierte Vergleiche von Bildtraditionen zu untersuchen, das wäre superspannend. Das ist natürlich viel Zukunftsmusik, aber hier liegen noch sehr viele unerkannte, sage ich mal, ‚Pfade‘ in der Bildkultur verborgen, die bislang von der Forschung noch nicht gesehen wurden.

Corinna Reinhardt:

Ganz kurz ein relativ praktischer Aspekt: Wir denken auch an die Nutzerinnen und Nutzer der Zukunft. Mit der visuellen Suche, deren Potenzial wir ja thematisiert haben, kann perspektivisch den bestehenden Fachdatenbanken wie dem sogenannten *Beazley-Archive*⁶⁹ ein Tool an die Seite gestellt werden, das der antiken Bildwahrnehmung mehr entspricht als die Suche mithilfe kategorisierter Begriffe. Die Suche nach vergleichbaren Bildern findet in erster Linie auf einer visuellen Ebene statt. Natürlich muss dies entsprechend im Training der Algorithmen berücksichtigt werden. Dies gelingt meines Erachtens vor allem dann, wenn man den digitalen Bildvergleich für Grundelemente der Bildanalyse verwendet und dies weitgehend von der Interpretation trennt. Beispielsweise wie wir es im Projekt anstreben, die Körperhaltung der Figuren und ihre Ausrichtung zu erkennen und dadurch auch Zusammenhänge von Figuren vergleichen zu können, die dann wieder als ikonographische Schemata bedeutungstragend sind. Unabhängig von der eigenen digitalen Kompetenz könnte hieraus jede und jeder einen Nutzen für die eigene Forschungsarbeit ziehen. Dies erfordert aber auch die klare Kommunikation über die Grundlagen, aufgrund derer ein Bild als vergleichbar angezeigt wird, also das Grundverständnis dafür, wie ein Algorithmus entwickelt und trainiert wird. Und hier besteht noch viel Gesprächsbedarf.

69 Classical Art Research Centre (CARC) der University of Oxford: <https://www.beazley.ox.ac.uk/carc/pottery> (abgerufen am 03.06.2023).

Elisabeth Günther:

Ich denke auch, das ist etwas, was wirklich die Erschließung und auch die Beschäftigung mit den altertumswissenschaftlichen Corpora voranbringen kann. Natürlich ist das *Beazley-Archive*⁷⁰ ein wichtiges Beispiel, aber man könnte das auch beliebig erweitern, etwa auf Münzen, z.B. das *Roman Imperial Coinage*⁷¹ oder *Roman Provincial Coinage*⁷², diese großen, online nutzbaren Corpora, die primär mit Textbeschreibungen arbeiten und die dann entsprechend nicht alle ikonographischen Feinheiten abdecken, wo Figuren wie Personifikationen und Gottheiten aufgrund der Legende eindeutig benennbar erscheinen, ihre Ansprache aber auch vom Blickwinkel der Ikonographie der Münzbilder her zu diskutieren wäre. Hier weckt die Textbeschreibung den Eindruck einer Eindeutigkeit oder interpretatorischen Klarheit, die keine Unschärfen zulässt und damit auch hinderlich sein kann, wenn man sich intensiver mit dem Verhältnis von Legende und Bild auseinandersetzen möchte.

Martin Langner:

Ich finde das grundsätzlich auch. Aber Menschen denken nicht nur in Bildern und es kann auch nicht jeder so gut zeichnen, dass er rein visuell in einer Datenbank suchen kann. Deswegen hatte ich eingangs ja schon *CLIP* von *openAI.com* genannt.⁷³ Ich glaube, die Zukunft liegt genau darin, Bilder und Labels zu kombinieren und möglichst viele neue Labels zu generieren. Damit komme ich zu dem Punkt, der mich gerade bei Ute Versteegen und dem Neuerfinden von Bildern interessiert hat. Das sind Analogien, die da am Anfang stehen. Und die Analogie ist eine der Hauptmethoden der Geisteswissenschaften. Und das macht die Künstliche Intelligenz in etwa genauso. Was dabei spannend ist, ist die Frage: Wie viel Analogie darf sein? Im Moment wird sehr stark über Transferlernen auf neuronalen Netzen, die auf anderen Bildern trainiert wurden, gearbeitet. Dazu werden vortrainierte *Convolutional Neural Networks (CNN)* mit unseren Bildern weitertrainiert, oder, noch stärker von unserem Bildmaterial entfernt, durch *up-sampling* trainiert, also durch neue, synthetisch erzeugte und nicht real existierende Bilder, um eine Gleichheit in den verschiedenen Klassen zu erreichen. Das klingt erst einmal so, als wäre das für historische Bilder völliger Unsinn. Ich frage mich aber, zumindest bei Bildern wie auf den Vasen, die in handwerklicher Produktion entstanden sind, ob die Maler nicht genau so vorgehen. Die können nicht vier Kratere am Tag töpfern und bemalen, wenn sie nicht bestimmte Figuren, Muster und Attribute im Kopf haben, die sie immer wieder frei kombinieren können. Und ich könnte mir gut vorstellen, dass, wenn man das wirklich kritisch begleitet und stärker explorativ umsetzt, diese Analogie die historischen Pfade, die sie bisher hat, ergänzen könnte durch einen synthetisch-explorativen Bereich. Man muss es aber mit großer Vorsicht angehen! Ich sehe das erst einmal als ein Potenzial, nicht als eine Möglichkeit, die blind angewendet werden sollte, sondern eher als ein Forschungspotenzial. Dann würde sich nämlich zeigen, ob das, was Corinna Reinhardt für eine Suchmaschine vorgeschlagen hat, hilfreich ist oder ob es zu noch mehr Vasenbildern als Suchergebnis führt, die man auch nicht alle durchschauen möchte. Also: Lässt sich durch Synthetisieren, durch Transferlernen eine Gruppe semantisch sinnvoller eingrenzen als bisher? Oder stehen wir einer noch größeren Flut gegenüber? Denn das ist, glaube ich, das Hauptproblem an diesen Datenbanken, dass keiner Lust hat, tausend Datensätze durchzuklicken, um genau das Bild zu finden, das er sucht.

Elisabeth Günther:

Wobei sich die Frage stellt, ob es nicht auch von der jeweiligen Forschungsfrage abhängig ist, ob dieser synthetische Bereich sinnvollen Einsatz finden kann.

70 <https://www.beazley.ox.ac.uk/carc/pottery> (abgerufen am 03.06.2023).

71 <https://numismatics.org/ocre/> (abgerufen am 03.06.2023).

72 <https://rpc.ashmus.ox.ac.uk/> (abgerufen am 03.06.2023).

73 S.o. Anm. 65.

Martin Langner:

Ja klar. Der Kontext, darüber wurde ja schon gesprochen, spielt eine Rolle, aber das Potenzial liegt in der Materialmenge. Erst mit dieser haben wir überhaupt eine Chance, solche Prozesse wie die in der Christlichen Archäologie in Angriff zu nehmen. Wir müssen uns demnach tatsächlich erst einmal an so große Corpora machen, wie Sie sie ja auch schon genannt haben. Also Münzen und Vasen. Sarkophage zum Beispiel sind viel zu wenig, denn 600 stadtrömische Sarkophage der Römischen Kaiserzeit reichen nicht aus. Es macht eigentlich nur da Sinn, wo es zum jetzigen Zeitpunkt einen ständigen Zuwachs an Bildern gibt.

Kerstin P. Hofmann:

Für uns ist dies ausgesprochen interessant, weil wir bei den keltischen Münzen im Projekt *ClaReNet – Klassifikationen und Repräsentationen für Netzwerke. Von Typen und Merkmalen zu Linked Open Data bei keltischen Münzprägungen* (vgl. Abstract Nr. 4) auch mit unterschiedlichen Sprachkontexten konfrontiert sind: Wir haben bei den Corpora immer das Problem, dass wir Bilder in unterschiedlichen Sprachen beschreiben und die verwendeten Worte und Begriffe, je nach Sprache, zudem unterschiedliche Bedeutungsräume haben. So stellen wir immer wieder fest, dass ein Bild im Deutschen und im Französischen zum Teil sehr unterschiedlich beschrieben wird. Wenn man dann überlegt, wie sich dies auf das Suchen auswirkt bzw. wenn wir Suchergebnisse vergleichen – oder gar selektiv kombinieren –, ist dies für die Analyse von Wissensproduktion wiederum sehr aufschlussreich. Da stehen wir noch immer vor großen Herausforderungen. Für die Lehre könnte es hilfreich sein zu überlegen, in welchen Abstraktionsebenen Bildbeschreibungen erfolgen. Gerade bei Münzen ist dies z.B. für die Ansprache der sogenannten *Beizeichen* interessant, für die manchmal sehr weitgehende Interpretationen vorliegen. Geht man hier wieder zurück zum Bild, kann man meines Erachtens durch die Reflexion über Prozesse und Praktiken der Versprachlichung viel lernen, übrigens auch über interkulturelle Wahrnehmung.

Auch Bilder sind kulturspezifisch. Spannend ist, dass auch Bilder durch neue ‚fremdsprachige‘ Bilder oder Diagramme sozusagen übersetzt werden können, und dann zu sehen, wie sie neu ‚gelesen‘ bzw. ‚gesehen‘ werden.⁷⁴ Ganz ähnliches passiert ja auch, wenn wir uns durch *heatmaps* (vgl. Abstract Nr. 4 sowie Abb. 10 und 11) anzeigen lassen, welche Bildbereiche bei der Künstlichen Intelligenz für eine bestimmte Zuordnung mit welcher Wahrscheinlichkeit relevant sind. Diese Bildbereiche unterscheiden sich zwischen KI und menschlichen Expert*innen mitunter und darauf aufbauend lassen sich dann weitere Fragen entwickeln. So kann man etwa viel über aktuelle Wahrnehmungsbereiche lernen und zum Beispiel in Zusammenarbeit mit Psychologie oder kognitiven Wissenschaften aufschlussreiche Vergleiche ziehen. Dies eröffnet ganz neue Möglichkeiten, auch für Suchalgorithmen à la: ‚Finde alle ähnlichen Bilder oder bestimmte Bildbereiche mit xyz‘.

Marta Kipke:

Ich möchte gerne ein paar Punkte ergänzen, ohne die Freude über automatisierte Suche zu dämpfen. Ich glaube, eine der großen Herausforderungen, vor denen man steht, ist und bleibt immer noch die Transparenz des neuronalen Netzes selbst und der Daten, die man für das Training verwendet. Denn jedes neuronale Netz, das man trainiert, und auch die *feature maps*, die dabei entstehen, sind bis zu einem gewissen Grad subjektiv und hängen stark von der Fragestellung ab, für die man sie anlegt. Wir haben in unserem Projekt lange darüber diskutiert, ob wir unsere *embeddings* aus dem Projekt *EGRAPHSEN* für eine automatisierte Suche verwenden möchten. Sie sind da und die Weiterverwendung existierender Daten ist erstrebenswert. Allerdings sind diese Zuweisungen hochkomplex und nicht immer nachzuvollziehen. Das steht eigentlich unserem Ziel, die Zuweisung der Vasen zu Malerhänden besser zu verstehen, entgegen. Und auch fürchte ich, dass viele von den Verfahren, die mit komplexeren Methoden wie *Convolutional Neural Networks* arbeiten, nicht immer angemessen sind

74 Vgl. Bawden et al. (2016), 528–534.

oder potentiell an der Fragestellung vorbeigehen. Zum Beispiel, wenn es um das Erkennen von Figuren auf Münzen geht: Es ist zu überlegen, ob in solchen Fällen nicht manchmal auch ein einfaches *template matching* bereits ausreicht, um ähnlichkeitsbasiert *regions of interest* zu finden. Das Ziel und die Fragestellung sollten dabei stets im Mittelpunkt stehen. Selbst wenn man sich für *Convolutional Neural Networks* entscheidet, wird man keine Datenbank erstellen können, die wirklich alle Forschungsfragen oder alle Bedürfnisse wirklich abdecken kann.

Corinna Reinhardt:

Ich möchte ebenfalls bei den Einsatzmöglichkeiten anknüpfen. Darüber haben wir im Teilprojekt viel diskutiert: Wären entwickelte Modelle zum Beispiel überhaupt nachnutzbar? Sehr wichtig war mir der Aspekt, dass man sich, auch wenn das vielleicht nicht am spannendsten ist, zunächst mit ganz basalen Elementen der Bildanalyse beschäftigt, die man dann für verschiedene Fragestellungen nachnutzen kann. Wir haben hier beispielsweise an dem Erkennen von Körperhaltungen in den Vasenbildern gearbeitet, bereits das stellte uns vor viele Herausforderungen. Bei einer Suche nach Vergleichen möchte ich normalerweise nicht nur als Ergebnis erhalten, dass eine Figur Herakles darstellt, sondern ich möchte z.B. seine Körperhaltung vergleichen oder danach fragen können, ob seine Position in einer Bildkomposition einer Konvention entspricht oder abweicht. Hier bieten etablierte Datenbanken keine Möglichkeit, eine solche Suche zu unterstützen. Eine visuelle Suche kann hier einen entscheidenden Fortschritt bieten, ohne dass dies gleich statistisch belastbar sein muss. Uns interessiert das Einzelbild, aber auch das große Ganze, also beispielsweise generelle Tendenzen in der Bildgestaltung, das Arbeiten mit ikonographischen Schemata und vieles mehr. Sowohl für eher qualitative, als auch eher quantitativ ausgerichtete Studien wäre ein solcher Bildvergleich also hilfreich. Durch den Einbezug größerer Datenmengen und den schnelleren Vergleich erhoffe ich mir, dass man zahlreiche Einzelaspekte überprüfen kann, die als Fragen bei einer Analyse üblicherweise auftauchen. Ist es z.B. typisch für Halsamphoren eines bestimmten Malers aus der zweiten Hälfte des sechsten Jh. v. Chr., dass die Figuren alle in einer bestimmten Art und Weise aufeinander ausgerichtet sind? Danach ließe sich prüfen, wie relevant das Ergebnis für das Verständnis eines bestimmten Vasenbildes ist. In einem zweiten Schritt wäre außerdem die Weiterentwicklung entsprechender Tools wichtig. Man könnte sie dann an unterschiedliche Fragestellungen anpassen und mit einzeln trainierten Modellen verknüpfen.

Martin Langer:

Zum Thema Nachhaltigkeit und Weiternutzung der Daten für unterschiedliche Fragestellungen, in Ergänzung noch zu Marta Kipke: Natürlich leisten wir ganz viel, was man weiterverwenden kann. Wir haben eine außerordentlich tiefe Annotation der Bilder im Projekt *EGRAPHSSEN* vorgenommen. Wir haben alles, was man sehen kann, auch annotiert, jeden Mäander einzeln. Und das ist natürlich weiter nutzbar. Das heißt, man muss eigentlich sehr stark unterscheiden zwischen eindeutigen formalen Elementen, die man auf dem Wege der Annotation erreicht, oder, um es mit informatischen Begriffen zu benennen: Maße und klar erkennbare Formen sind auch schon mit den traditionellen *machine learning*-Modellen zu erfassen gewesen. Die muss man absetzen von *features*, die eben eigentlich eine Summe, eine große Summe von Ähnlichkeiten beschreiben und damit vielleicht jetzt von der Identität in den Bereich der Ähnlichkeit gehen, um das aufzugreifen, was Kerstin Hofmann gesagt hat, um dann über die Signale in den Daten (das, was die neuronalen Netze erzeugen), in die Ebene des noch weitergehenden Konglomerats von Formen zu kommen.

Dieser Bereich, der ist in der Tat ganz stark abhängig davon, auf welcher Datengrundlage weiter trainiert wurde, was man in den Test- und was man in den Trainingsdatensätzen hat. Und da gebe ich Marta Kipke völlig recht: Das hat einen hohen *bias* und kann eigentlich nur in Hinblick auf die Forschungsfrage einigermaßen abgemildert werden. Aber die Ebene der *features* und noch stärker die Ebene der formalen Eigenheiten ist natürlich eindeutig bestimmbar.

Chrisowalandis Deligio:

Convolutional deep learning sollte man, glaube ich, nicht immer einsetzen – manchmal ist es einfach *too much* – aber ich finde, man muss sich immer klar machen: Welche Frage hat man denn, bzw. was möchte man denn überhaupt? Man kann sich natürlich ein *feature* anschauen und dann beurteilen: Macht das Modell überhaupt das, was ich möchte oder wie ich es möchte? Manchmal sind die Zahlen ja super, aber es macht etwas ganz anderes oder schaut ganz woanders hin. Das sollte man immer beurteilen und überprüfen. Noch etwas anderes, was wir auch im Projekt benutzen: Man kann mit dem Modell auch Verschiedenes hinterfragen. Wenn ich eine klar definierte Klassifikation habe, überprüfe ich: Wie sieht das Modell das? Schaut das Modell woanders hin als der menschliche Experte? Und daraus können dann Fragen entstehen. Diese Klassifikationsflächen sind zu hinterfragen oder auch zu bestätigen. Wenn vielleicht etwas Ähnliches herauskommt, dann ist vermutlich schon etwas dran, dass diese Klassen anhand dieser Merkmale gebildet worden sind. Ein Beispiel: Wenn das Modell immer auf das Auge eines auf der Münze dargestellten Kopfes schaut, aber in der vorgegebenen Klassifikation die Nase stärker definiert ist, dann stellen sich neue Fragen, z.B. wie es zu der Klassifikation gekommen ist, auch forschungsgeschichtlich, wie bestimmte Typologien und Klassifikationen entstanden sind. Ich möchte aber den Punkt unterstreichen, dass man sich immer fragen muss: Was möchte man denn mit dem Modell erreichen? Man kann nicht irgendwelche Daten hineingeben (auch wenn selbst in diesem Fall irgendwelche Ergebnisse herauskommen werden), sondern man muss klare Zielsetzung haben.

Marta Kipke:

Vielen Dank, Herr Langner, für die Ergänzungen. Natürlich sehe ich auch alle Vorteile und finde sie unglaublich interessant. Ich möchte nur gerne betonen, dass es eine große Herausforderung ist und bleibt, die Daten als das zu sehen, was sie auch sind. Experimente sind immer toll und führen zu einem Erkenntnisgewinn. Und jedes Experiment führt zu einem neuen Modell, das man auch weiterverwenden kann, wie beispielsweise für eine bildbasierte Suche. Ich denke jedoch, dass man sich zwei zentralen Problemen stellen muss: Einerseits muss man den *user* im Blick behalten. Es muss sichergestellt werden, dass ein *user* die Daten und Vorschläge nicht als gegeben hinnimmt, sondern dass immer transparent bleibt: Mit welcher Fragestellung sind diese Modelle entstanden? Auf welcher Grundlage wurde trainiert? Wenn wir uns das jetzt wirklich ganz praktisch als Datenbank vorstellen: Dass man auch auswählen kann, welche Modelle zum Suchen verwendet werden oder auch zum Erstellen der Visualisierung oder welche weiteren Funktionen man auch einbauen möchte. Die Kriterien müssen klar und die Parameter veränderbar sein.

Das andere zentrale Problem ist eine Sensibilisierung dafür, welche Daten und welche Informationen in Visualisierungen eigentlich repräsentiert werden und der bildbasierten Suche zugrunde liegen. Es sind nicht unbedingt die Ähnlichkeiten der Bilder zueinander, wie sie ‚in den Bildern stecken‘, so wie wir Geisteswissenschaftler es uns wünschen, sondern so, wie sie vom neuronalen Netz bestimmt wurden. Im Falle von Visualisierungen handelt es sich meistens auch um eine starke Reduktion dieser Ähnlichkeiten in Form von hochdimensionalen Vektoren auf bloß zwei oder drei Dimensionen, wie beispielsweise bei dem beliebten Verfahren *tSNE*⁷⁵. Eigentlich sehen wir in diesen Verteilungsdiagrammen die Wahrscheinlichkeiten von Ähnlichkeiten durch die Brille eines neuronalen Netzes. Ich finde es für den Umgang mit den Daten sehr wichtig, dass man hierfür sensibilisiert wird. Sei es in Anwendungen für andere Forschende oder auch in der eigenen Analyse der Daten.

Elisabeth Günther:

Vielen Dank für diesen Beitrag, ich denke, das ist ein ganz entscheidender Punkt. Es geht natürlich immer auch um die Frage: Was sagt uns das Verhältnis zwischen Experten und Algorithmus? Wir müssen uns damit unbedingt auseinandersetzen und uns klar machen: Womit haben wir es eigentlich hier zu

75 *t-Distributed Stochastic Neighbor Embedding*: Maaten / Hinton (2008).

tun? Zumal ja die Mustererkennung momentan sehr, sehr viel Aufmerksamkeit im Fach wie auch in der Öffentlichkeit bekommt. Es wird aber, glaube ich, nicht immer so ganz verstanden, worum es eigentlich geht und wie die Ergebnisse zu interpretieren sind.

Kerstin P. Hofmann:

Das unterstütze ich vollends. Ich glaube, das ist genau das Spannende, dass wir, während wir etwas praktizieren, sehr sorgfältig überlegen müssen, was wie in welchen Schritten passiert. Dies verdeutlicht uns zudem, welche Prozesse alle gleichzeitig ablaufen und wo Unterschiede vorhanden sind.

Wie verhält sich die Mustererkennung im Vergleich zu traditionellen Methoden und Techniken? Ein entscheidender Vorteil bei der Zusammenarbeit mit der Informatik ist, dass ich vieles ausbuchstabieren muss. Dies gilt insbesondere bei der gemeinsamen Entwicklung von Fragestellungen und Experimenten. So hatten wir zahlreiche intensive Diskussionen im Zuge des Experimentes, das wir in unserem Vortrag im Rahmen des Workshops *NEUES SEHEN* vorgestellt haben, in denen es vor allem darum ging, was wir als Datenbasis nehmen, was wir eigentlich vergleichend setzen können und warum? Diese Entscheidungen sollte man meines Erachtens auch immer mitkommunizieren.

Werden Ergebnisse, wie bei *Convolutional Neural Networks*, aggregiert, sind leider dann auch diese nicht bis ins Detail nachvollziehbar. Dies ist aber auch bei traditionellen Arbeiten der Fall. Ich habe bisher nur ganz wenige Klassifikationsbeschreibungen gesehen von Bildern oder auch Objekten, in denen wirklich offengelegt wurde, wie sie genau durchgeführt wurden, warum welche Merkmale herangezogen wurden und andere nicht und wie und warum diese so und nicht anders gewichtet wurden, welche Probleme der Erhaltungsbedingungen z.B. durch Fragmente existieren, in welchen Verhältnissen die Klassen zueinander stehen etc.

Es gibt demnach immer wieder die Möglichkeit, z.B. durch Betrachtung verschiedener Ebenen und Zusammenhänge einerseits und durch den Wechsel zwischen qualitativen und quantitativen Ansätzen andererseits, Entscheidungs- und Erkenntnisprozesse offenzulegen und zu hinterfragen. Dies ist meines Erachtens sehr vorteilhaft für die Wissenschaft und Forschung.

Katja Rösler:

Ich wollte auch hervorheben, dass wir leider überhaupt keine Standards in den Fächern haben, die genau diese Entscheidungsfindung, das Existieren dieser Entscheidungsfindung, in irgendeiner Form dokumentieren, dass es nachlesbar ist. Gerade bei der Verwendung von *CNNs* wird das immer weniger nachvollziehbar, wenn wir keine Form finden, das zu explizieren. Es ist auch schon bei publizierten, also online zugänglichen, Datenbanken oft so, dass ich gar nicht weiß, was drinsteckt, weil ich vorne etwa auf dem *user interface* keine Informationen dazu finde und die es auch nicht öffentlich zugänglich gibt, weil man den Nutzer bzw. die Nutzerin, wie Marta Kipke das so schön gesagt hat, bei dem ganzen Tun und Machen vergisst. Wir brauchen den Blick von außen, der aktiv fragt: Was sehe ich hier denn eigentlich?

Deswegen begleitet eine *science-and-technology*-Studie das Projekt *ClaReNet* auch, die natürlich nochmal ihre eigene Fragestellung hat, aber die uns auch helfen soll, den Blick von außen wahrzunehmen – immer wieder. Ich denke, dass das sehr wichtig ist, gerade auch in Zusammenarbeit mit der IT bei der Verwendung von neuen Sprachen (*ClaReTerms*) und Begriffen. Wir legen Sammlungen von Begriffen (*wikis*) an, die wir erst mal für uns definieren, weil wir uns verständigen müssen. Darunter sind dann auch Metaphern, die einfach auch mal aufgeschlüsselt werden müssen. So spricht unser *data scientist* Chrisowalandis Deligio oft von Modellen. Wir Archäologinnen und Archäologen haben aber ein ganz anderes Verständnis von Modellen. Wir versuchen das zu publizieren, damit das nicht nur im Projekt selbst klar wird, sondern möglichst hinterher nachvollziehbar und von außen einsehbar ist. Ich denke, da braucht es noch eine gewisse Form von Verständigung, wie so etwas aussehen soll in den Archäologien, in den Projekten, die mit Digitaler Archäologie irgendwie zu tun haben.

Martin Langner:

Ich finde wichtig, dass wir einen Diskussionsprozess beginnen, wie die Standards der KI-gestützten Analyse überhaupt aussehen müssen. Und das liegt auf vielen Ebenen, ich habe es auch schon gesagt, wir müssen Annotationsstandards haben. Da gibt es bereits einige, allerdings konkurrierende. Bestimmte Verfahren, die sehr einfach sind, sind auch gut geeignet, um in Bibliotheken Illustrationen schnell auffindbar zu machen. *IIIF* ist hier das Stichwort.⁷⁶ Andere brauchen eine Tiefenannotation und da muss man dann zum Beispiel wissen, dass es immer dasselbe Bild sein muss, denn es ist pixelgenau annotiert. Es muss nicht dasselbe Foto sein, das unterschiedlich eingescannt wurde, sondern es muss wirklich dieselbe Datei in derselben Dateigröße sein. Sonst entstehen hier auch schon Unschärfen.

Damit kommen wir zum Hauptproblem: den Bildrechten. Sie hindern uns am meisten in der Bildmuserkennung an einem Fortschritt und an der Formulierung von Standards, weil wir viel zu wenig freie Bilder haben. In unserer Datenbank wird es so sein, dass wir zwar angeben, welche Bilder wir verwenden, wir werden die annotierten Daten publizieren und Links auf diese Bilder. Wenn die Links aber nicht mehr funktionieren, dann könnte man uns vielleicht noch persönlich anschreiben und um das Bild bitten, aber es wird nicht mehr möglich sein, das Bild öffentlich zur Verfügung zu stellen. Das ist dann einfach so. Aber das ist immer noch besser, als diese Verbindung von Annotation und Bilddatei ganz wegzulassen. Denn dann kommen wir wirklich bloß auf eine Lösung eines bestimmten Problems, aber keiner Nachnutzbarkeit der Daten, die damit verbunden waren. Wir alle haben das jetzt ja schon mehrfach betont: Die meiste Arbeit stecken wir in das Scannen, in das Fotografieren, in das Annotieren, also in das Aufbereiten der Daten, und das Processing hängt sehr stark davon ab, ob man einen Termin im High Performance Computing Cluster des Rechenzentrums bekommt, ob man Geld für weiterführende Experimente bekommt oder ob das ohne Finanzierung nicht mehr weiter erforscht werden kann.

Ute Versteegen:

Um jetzt zunächst mal auf Martin Langner zu antworten: Da bin ich absolut deiner Ansicht. Das größte Problem bei den *FAIR*-Prinzipien⁷⁷ ist es, das *re-usable* hinzubekommen, weil wir so riesige rechtliche Schwierigkeiten in der Wieder-Bereitstellung der Corpora mit den annotierten Bildern besitzen. Letztendlich brauchen wir sehr viele politische Initiativen, die dazu führen, dass die eigentlichen Rechteinhaber – also die Eigentümer der Artefakte, die Museen und die zugehörigen Fotografen – alles in *Creative Commons* lizenzieren und nach bestimmten Standards für Repositorien zur Verfügung stellen. Sonst wird das einfach immer schwierig bleiben. Die Rechtklärung selbst im Kleinen mit wenigen Bilddatensätzen ist heutzutage noch so enorm aufwändig, wenn man nicht auf ein einzelnes Repositorium zurückgreifen kann, dass es mit dem *re-usable* einfach nichts wird. Ich würde sagen, das ist ein noch größeres Problem als unsere Humanressourcen, als ganz viel Geld für studentische Hilfskräfte zu beschaffen, die annotieren. Wir machen all diese Arbeit und können sie nicht wiederverwenden. Das ist das allergrößte Problem.

In unserem Projekt haben wir uns Bilder aus den unterschiedlichsten Repositorien zusammengeklaut und dann auch selbst aus Büchern gescannt. Die Grundlage der Bilddaten war aber so heterogen, was Pixelanzahl und Auflösung betraf, dass auf dieser Basis der Einsatz von KI für bestimmte Fragestellungen verunmöglicht wird. Wenn wir Plots nutzen, um auf einem solchen Datenset basierend Vergleiche anzustellen, dann ist vollkommen klar, dass das, was die Algorithmen da zurecht sortieren, uns am Ende eigentlich nur zeigt, ob das jetzt Schwarzweiß-Aufnahmen von einem Objekt sind oder Farbaufnahmen – das war jetzt mein Kommentar zu Marta Kipkes Beitrag vorhin.

76 *International Image Interoperability Framework (IIIF)*: https://iiif.io/guides/guide_for_implementers/ (abgerufen am 03.06.2023).

77 *Findability, Accessibility, Interoperability, Reuse of digital assets*: <https://www.go-fair.org/fair-principles/> (abgerufen am 03.06.2023).

Ja, am Ende bekommen wir – das war bei uns so – alle Sarkophage in einer Wolke zusammensortiert und dann alle Textilien in einer anderen, also Gruppierungen, in denen Textur und Farben irgendwie ähnlich waren. Das bringt einen aber nicht weiter, wenn man zum Beispiel die Fragestellung verfolgt, in welchen Bildern gattungsübergreifend ein und dieselbe Geste vorkommt.

Und um noch mal auf Kerstin Hofmann zu antworten: Auch in unserem Projekt war das mit Abstand Interessanteste die Diskussion der vermeintlichen, aus menschlicher Sicht vermeintlichen, Fehler zusammen mit der Informatik, was die *predictions*, die Überlegungen zu den *predictions* der Modelle anbelangt. Denn die Modelle geben uns ja nicht die Wahrscheinlichkeit an, dass da zu 95% ein Herakles wiedergegeben ist, sondern dass das Netz der Ansicht ist, es könnte zu 95% Wahrscheinlichkeit einen solchen gefunden haben. Das ist nicht dasselbe. Diese Diskussion über die entsprechenden, aus menschlicher Sicht begangenen Fehler in der *object detection* war auch ein sehr weiterführender Punkt, der uns dann auch dazu gebracht hat, unser Training nochmal mit anderen Methoden durchzuführen. Für mich war der Diskurs mit den anderen Fachwissenschaften, aber auch mit der Informatik selber das eigentlich Spannendste im Projekt, weil ich den Eindruck hatte, dies führt dann auch unsere fachlichen Fragen weiter.

Kerstin P. Hofmann:

Ganz herzlichen Dank. Es sind wirklich verschiedenste Aspekte bis hin zu der Frage, was man eigentlich als Voraussetzungen bzw. Rahmenbedingungen hat. Wir haben z.B. im Zuge unserer Münzaufnahmen erneut diskutiert über Aufnahme- und Repräsentationsverfahren in der Numismatik, welche oft nicht beschriebenen Forschungspraktiken existieren, wann dreidimensionale und wann zweidimensionale Repräsentationen genutzt werden, welche zusätzlichen Informationen eigentlich mitberücksichtigt werden, die dann aber mitunter bei den Klassifikationen gar nicht mehr angegeben werden. Wird zum Beispiel aufgrund der Bilderkennung nur noch auf das Foto der Münze fokussiert, werden Materialzusammensetzung oder das Gewicht zu zusätzlichen Informationen, die dann bewusst vorher oder irgendwann nachträglich einbezogen werden. Das Gleiche gilt für die Berücksichtigung des archäologischen Kontextes etc. Ich glaube, wir müssen dringend über verschiedene Repräsentationstechniken und -standards diskutieren: Welche Aussagen möchten wir eigentlich treffen? Welche Daten und welche über andere Datensammlungen und Module zugänglichen Informationen möchten wir erheben bzw. berücksichtigen? Bei uns im Projekt haben wir bei der Analyse der hier vorgestellten Münzserie sehr stark auf das Foto statt auf die Münze an sich fokussiert, so dass vieles auch von der Fotoqualität abhängig ist. Denn nur im absoluten Ausnahmefall wird das Objekt selbst dann noch konsultiert. Worauf man referenziert, sollte man sich jedoch immer klar machen und auch bei der Datenprozessierung offenlegen.

Neben den Bildrechten, die im Verhältnis zu den Textrechten sicherlich eine absolute Krux sind, gilt es auch die Frage der Publikation von Datensätzen zu diskutieren: Wann ist eigentlich ein*eine Autor*in oder eine Mehrautor*innenschaft zu nennen? Wann publiziere ich einen Datensatz und ab wann bin ich Mitautor*in oder muss zumindest zitiert werden? Wie soll dies oder auch das Annotieren als eigenständige Leistung anerkannt werden? In diesem Bereich haben wir noch keine Standards und momentan ist dies auch noch nicht in Lebensläufen oder anderswo gut verankert. Datenkritik ist neben der Quellenkritik eine Aufgabe, in die wir massiv Zeit investieren müssen. Außerdem müssen wir darauf achten, dass wir in digitalen Projekten nicht schon zum Teil überholte Wissensbestände als Normdaten festschreiben, nur weil sie viel referenziert wurden. Vielmehr gilt es, aktuelle Diskussionen immer wieder neu zu integrieren. Ein Buch ist eine statische Publikation, und man ist sich einig: Das ist das Hauptergebnis. Doch auch beim dynamischen Publizieren müssen wir zwischen Forschungsprozess und -ergebnis unterscheiden oder durch Versionierungen zumindest Etappen eines Forschungsprozesses definieren. Denn nur so können andere Forschende darauf aufbauen und referenzieren. Eine große Herausforderung wird sein, mit der entstehenden Vielfalt an Publikationen sinnvoll umzugehen und sich hier auf verschiedene (Qualitäts-)Standards zu einigen.

Ich denke, dass viele kleine Entscheidungen bei Digital-Projekten derzeit sehr langfristige Auswirkungen haben können (Stichwort: *Pfadabhängigkeiten*). Diese sollten wir daher auch besonders reflektiert durchführen.

Elisabeth Günther:

Vielen Dank, auch das ist noch mal ein sehr wichtiger Punkt. Er zielt natürlich auch auf die Fragen ab: Was passiert jetzt auch mit unseren Datensätzen, mit unseren Forschungsprojekten? Wie können wir uns untereinander vernetzen und können auch voneinander profitieren? Wie können vielleicht auch zusammenarbeiten? Worauf müssen wir achten? Gibt es da bestimmte Standards, die wir setzen müssen, und in welchem Bereich ist es sinnvoller, das Ganze flexibler zu handhaben und eher den Kontext zu sehen, die Fragestellung, das Ziel? Wie könnte man eine engere Kollaboration unterschiedlicher Projekte realisieren?

Corinna Reinhardt:

Ich war mit dem Projekt (vgl. Abstract 2) Neueinsteigerin in dem Bereich der *Computer Vision* und ihrer Anwendung in den Bildwissenschaften. Natürlich habe ich selbst sehr viel dabei gelernt und lernen müssen. Vor allem haben die Diskussionen innerhalb unserer interdisziplinären Gruppe geholfen. Viele grundlegende Fragen konnten wir aber erst im Laufe der Projektzeit angehen und nicht generell auf bereits Etabliertes zurückgreifen, beispielsweise zu Standards möglicher Nachnutzungen. Dieser Bereich ist eben noch nicht Teil einer länger etablierten Grundlagenforschung unserer archäologischen Fächer, was auch die Frage danach, wie können Kollaborationen realisiert werden, erschwert. Martin Langner hat in dieser Gesprächsrunde am längsten Erfahrung. Wenn Erfahrungen in diesem Bereich für antike Bildwissenschaften nun tatsächlich stärker in die Forschungsausrichtung unserer Disziplinen übergehen, dann kann man auch langfristige Planungen machen, denn bislang ist es ja doch sehr stark vom geförderten Einzelprojekt abhängig, was realisiert werden kann. Die damit zusammenhängende kompetitive Ausrichtung befördert strategische Planungen ja nicht unbedingt. Kooperationen sehe ich dennoch als einen zentralen Schritt für die Frage, welches Ziel wir innerhalb unserer Fachdisziplinen denn eigentlich damit verfolgen. Unser Gespräch ist da eine gute Grundlage, genauso wie eine Arbeitsgemeinschaft oder Ähnliches. Es ist sicherlich hilfreich, Datensätze abzustimmen, dass nicht beispielsweise mehrere das Gleiche annotieren. Vielmehr hilft es, wenn man Daten gegenseitig austauscht und sie dann auch für unterschiedliche Fragestellungen anpasst. Generell braucht es sicherlich für die Frage nach der Zukunft eine stärkere Verankerung an einigen Professuren, beispielsweise mit einem Forschungsschwerpunkt.

Ute Versteegen:

Ich wollte nur noch mal darauf hinweisen, dass es bei den Kollaborationen strategisch betrachtet keine unwesentliche Rolle spielt, mit welchen bildrechtlichen Voraussetzungen man konfrontiert ist. Es ist wichtig, dass man jetzt mittelfristig so plant, dass man Projekte anstoßen kann, die einmal in der Lage sein werden, größere Repositorien nach der Annotation zur Verfügung zu stellen. Da eignen sich zum Beispiel numismatische Datenbanken auch deshalb gut, weil sie bildrechtlich meistens noch als machbar eingeschätzt werden. Man muss sich immer darüber im Klaren sein, dass diese bildrechtlichen Implikationen wirklich auch über unsere Fächer hinaus eine Rolle spielen, wenn es Repositorien gibt. In der Kunstgeschichte hat das schon dazu geführt, dass ‚Flachware‘, urheberrechtlich unbedenkliche Malerei bis 1900, in größeren Repositorien zusammengefasst worden ist. Da gibt es zum Beispiel *OmniArt*⁷⁸ und *IconArt*⁷⁹, das sind Repositorien, die die informatischen Kolleginnen und Kollegen dann auch wahnsinnig gerne benutzen, um quasi ‚zu spielen‘, einfach mal etwas mit den Netzwerken, die sie trainieren, auszuprobieren. Um einfach mal zu schauen, was sie mit Bilddatensätzen historischer Kunst machen können, um nicht nur aktuelle fotografische

78 Zu *OmniArt*: <https://isis-data.science.uva.nl/strezoski/#3> (abgerufen am 20.03.2023); Strezoski / Worrington (2017).

79 Zu *IconArt*: <https://paperswithcode.com/dataset/iconart> (abgerufen am 20.03.2023); Gonthier et al. (2019).

Bilddatenpools wie *COCO*⁸⁰ oder *ImageNet*⁸¹ nutzen zu müssen. Da müssen wir unheimlich aufpassen, dass die Netzwerke dann nicht anhand solcher Repositorien weiter trainiert werden und uns hinterher angeboten werden. Dann entstehen diese *biases*, auf die bereits immer wieder hingewiesen wurde. Das ist wichtig. Wir können in unseren Fachkulturen bestimmte Artefaktgruppen finden, über die wir jetzt sehr gut auch größere Projekte in Kollaboration anschieben können. Aber natürlich ist es trotzdem spannend, da muss ich aus meinem eigenen Fach heraus sprechen, auch mit diesen kruden Mischungen aus einer Forschungsperspektive mit der Informatik weiterzuarbeiten. Denn wer weiß, vielleicht sind wir in zehn Jahren so weit, auch mit unserem Material entsprechend weiter und anders umzugehen, wenn wir jetzt mit viel Impetus daran forschen.

Marta Kipke:

Ich wollte noch mal kurz auf den Aspekt der Autorenschaft der Datensätze eingehen, den Kerstin Hofmann vorhin angesprochen hat. Ich finde ihn ganz wichtig, weil ich zuletzt ja auch ein ganzes Jahr im Projekt *EGRAPHSEN* daran gearbeitet habe, einen solchen Datensatz zusammenzustellen und viel intellektuelle Leistung hineinfließt, diesen ausgeglichen zu gestalten und entsprechende *sampling*-Methoden zu entwickeln, die sowohl informatischen als auch geisteswissenschaftlichen, spezifisch archäologischen Ansprüchen genügen. Insbesondere in Bezug auf die *biases*, die nur wir als Fachwissenschaftlerinnen und Fachwissenschaftler überhaupt erahnen können. Eigentlich muss über die Art und Weise, wie wir unsere Daten publizieren, ein Umdenken stattfinden.

Zu dem Punkt mit den Kollaborationen unterschiedlicher Projekte: Hier fangen die Schwierigkeiten häufig auch schon viel simpler an als man denkt, z.B. verwenden alle unterschiedliche Dateiformate für ihre Annotation. Es wäre schon mal ein guter ein Anfang, wenn alle ihre Formate, also alle ihre Annotationen, in demselben Format machten, weil Konvertierungen von einem ins andere Format auch nicht immer problemlos funktionieren. Und ich glaube, da braucht es tatsächlich einen ernsthaften Vorschlag für einheitliche, grundlegende Richtlinien. Ich muss immer daran denken, wie mir beigebracht wurde, wie ich ikonografische Analyse oder Bildbeschreibung zu machen habe. Es ist frage-spezifisch und es steht nicht alles in einer Bildbeschreibung oder ikonographischen Analyse, was auch dargestellt ist. Aber wir haben dennoch grundlegende Regeln, wie wir etwa bei der Beschreibung vorgehen. Das müsste es für Annotation so auch geben, dass man die Datensätze zwar vielleicht nicht immer unbedingt sofort weiterverwenden kann, aber dass sie flexibel sind und es einfach ist, anzuknüpfen. In *EGRAPHSEN* haben wir sehr stark versucht, eine flexible Annotationsrichtlinie zu machen, die auch in anderen Fragestellungen verwertbar ist. Dort kann man einfach schon nach den Sachen suchen, die darin sind, und muss nur noch kleine Ergänzung machen.

Kerstin P. Hofmann:

Durch Kollaborationen kann man viel lernen und sie sind dringend notwendig. Wir werden bei quantitativen Forschungen und bei dem Einsatz von KI nicht mehr nur einzeln und im stillen Kämmerlein arbeiten können. Wir müssen zusammenarbeiten und dabei so mutig sein, nicht nur offenzulegen, was gut, sondern auch was schlecht gelaufen ist. Dies wird unsere Arbeitsweise verändern bzw. hat sie auch schon z.T. in bestimmten Bereichen verändert; dies kostet aber auch Zeit. Wir sollten nicht nur irgendwie in Projekten für drei Jahre planen müssen, sondern wir benötigen Strukturen, die darüber hinausgehen. Deswegen sind meines Erachtens solche Initiativen wie *NFDI*⁸² oder *E-RIHS*⁸³ so wichtig, die dann aber auch wirklich Teil der Forschungen sein sollten.

80 <https://paperswithcode.com/dataset/coco> (abgerufen am 20.03.2023).

81 <https://www.image-net.org/> (abgerufen am 20.03.2023).

82 Nationale Forschungsdateninfrastruktur: <https://www.nfdi.de/> (abgerufen am 19.06.2023).

83 European Research Infrastructure for Heritage Science: <https://www.e-rihs.eu/> (abgerufen am 03.06.2023).

Wir haben festgestellt, dass es einfach auch an Plattformen fehlt, auf denen man eine Übersicht über aktuelle Projekte bekommt. Man erfährt zwar auf Workshops und Tagungen, was passiert ist, dies aber eben erst relativ spät. Es bräuchte einen dauerhaften Austausch und Netzwerke. Wenn wir Personen, die auch wirklich an Forschung interessiert sind, in diesem Bereich halten möchten, müssen wir ihnen auch Perspektiven bieten. Leider wandern uns viele Kolleg*innen gerade im informatischen Bereich ab, weil sie außerhalb der Wissenschaft einen Berufsmarkt vorfinden, der wesentlich bessere Rahmenbedingungen bietet. Gleichzeitig sehe ich, dass es zu einer Spezialisierung und Separierung der Digitalen Archäologie kommt, wobei deren Einbettung in die Fächerlandschaft noch offen diskutiert werden sollte. Die zahlreichen digitalen Archäologiestellen, die momentan an vielen Universitäten ausgeschrieben werden, bedürfen einer qualifizierten Besetzung und Rückbindung. Vieles wird derzeit versucht, mit kurzfristigen Projekten und Stellenausschreibungen umzusetzen. Die notwendigen Infrastrukturmaßnahmen, die meist nicht ausreichend finanziell ausgestattet sind, werden dazu führen, dass vermutlich nur einige wenige Institutionen hier federführend sein können. Die große Herausforderung wird sein, innovative Forschungen auf breiter Basis zu ermöglichen und eine gute Kombination zwischen Infrastruktur- und Projektarbeit, mit ihren zahlreichen Zwischenschritten, wie z.B. die Entwicklung von Pilotstudien, Standards und *best practices*, zu wählen.

Außerdem müssen wir schauen, dass nicht nur die Anzahl der Daten eine Rolle spielt, sondern auch die Datenqualität. Meldungen, dass z.B. 35.000 neue Datensätze eingegeben wurden, sind zwar erfreulich, aber dies sagt meist noch nichts über deren Umfang und Validierung aus. Dahingehend wird die Bereinigung bzw. Überarbeitung von Daten meist nicht mehr gefördert, doch auch für diese zentralen Aufgaben müssen wir Möglichkeiten schaffen. Ich hoffe, dass viele der anstehenden Herausforderungen durch *NFDI* und vergleichbare Initiativen angegangen werden. Aber digitales Forschen ist eine Herkulesaufgabe, die uns alle noch viel Zeit kosten wird.

Chrisowalandis Deligio:

Das ist ein Punkt, den ich bei Kerstin Hofmann unterstützen würde. Fehlt denn wirklich eine Plattform, wo dieser Austausch stattfinden kann? Wäre ich nicht bei der Tagung gewesen, hätte ich die ganzen Projekte gar nicht gesehen. Irgendwo müsste man eine Sammlung für Projekte und Vorhaben haben. In der IT gibt es schon entsprechende Seiten. Da gibt es viele Datensätze, über die man sich austauscht und mit denen man sich auch gegenseitig herausfordert. Vermutlich scheitert es an den Rechten, aber wenn man eine vereinfachte Version davon hätte, in der mitgeteilt würde, mit welchen Daten gearbeitet wird oder mit welcher Zielsetzung und wie man irgendwie Kontakt herstellen könnte, wäre schon ein wenig geholfen. Oder auch jetzt bei der Trierer Tagung, dass wir, wenn man sich ganz öffentlich machen will, so eine *GitHub*-Gruppe gründen könnten. Das fehlt ja so ein bisschen. Kurz: Wo finde ich, welche Projekte gerade laufen? Wie kann ich mich mit den Leuten hilfreich austauschen? Eine Plattform, auf der man diese Projekte schnell findet, wäre gut.

Elisabeth Günther:

Ich kann mir vorstellen, dass so etwas gerade für junge Wissenschaftlerinnen und Wissenschaftler nützlich wäre, die noch nicht so etabliert sind und nicht so viele Kontakte, und daher auch noch nicht so einen Überblick haben, die aber vielleicht gerade auch interessiert sind, zu diesem Bereich beizutragen.

Martin Langner:

Also da, glaube ich, mache ich mir tatsächlich jetzt mal ein bisschen Hoffnung bei der *Nationalen Forschungsdateninfrastruktur-Initiative (NFDI)*, dass solche Sachen umgesetzt werden können. Ansonsten bin ich ganz hin und hergerissen. Ich wollte eigentlich, als ich mich meldete, noch mal die Paradata stark machen, die die Prozesse der Erhebung der Daten sehr umfassend dokumentieren. Das ist hier absolut notwendig, also noch mehr als in anderen Fällen. Aber nach den Vorrednern habe ich mich jetzt gefragt: Was ist eigentlich das, was wir brauchen? Zunächst dachte ich an eine Forschergruppe,

wie die *Römische Ikonologie* in den frühen 70er Jahren. Da hatte man eine Gruppe von Interessenten (es waren gar nicht viele), die haben Standards definiert, haben in wegweisenden Aufsätzen definiert, wie man es machen muss und wie man es besser nicht machen sollte, mit guten Beispielen. So was braucht es auf jeden Fall. Ich glaube, wir alle, die wir jetzt auch gut miteinander diskutieren können, haben auch eine Grundlage, auf der wir so etwas auf den Weg bringen könnten. Insofern läge vielleicht die Idee einer *DFG*-Forschergruppe nahe, einen wirklich sichtbaren *use case* umfassend, reflektiert und mit allen Daten publiziert, gut durchschaubar, nachnutzbar, usw. Das könnte, glaube ich, schon ein sehr sinnvolles Ziel sein. Aber vielleicht brauchen wir doch eher die *Faust*-Edition. Im Textbereich der Digital Humanities gibt es zum Beispiel eine Edition von Goethes *Faust*⁸⁴ und an der kommt man jetzt nicht mehr vorbei. Wenn man digitale Editionen macht, muss man das mindestens so machen wie die *Faust*-Edition. Also vielleicht brauchen wir eher als eine Forschergruppe ein Leuchtturmprojekt, wo an einem zentralen Beispiel der Archäologien eine Bildmustererkennung mal so vorgeführt wird, dass man das immer als Referenz nehmen kann. Aber auch da wieder nicht einer alleine, sondern wirklich eine Gruppe, wobei dann auch ein intensiver Diskussionsprozess stattfindet. Das ist eben der Punkt, wo ich glaube, dass *NFDI* das nicht mehr leisten kann, sondern da müssten vielleicht genau wir, die wir jetzt hier beisammen sind, uns mal im Nachgang überlegen, was das sein könnte, das wir gemeinsam (oder wen nehmen wir noch dazu?) als ‚Vorzeige-KI-Projekt‘ der Bildmustererkennung umsetzen könnten, wo auch ein Ergebnis zu erhoffen ist. Und, das haben die Diskussionen sehr deutlich gezeigt, dass wir das auch wirklich kritisch und reflektiert und auf Grundlage von visuellen Phänomenen angehen und nicht nur ein gegebenes Problem lösen, sondern dann auch viel stärker die ganze Bandbreite noch nicht ins Auge genommener Probleme adressieren.

Kerstin P. Hofmann:

Zur Ergänzung: Ich denke, wir brauchen eine breite Klaviatur. Je nach Fragestellung bräuchten wir unterschiedliche Leuchtturmprojekte, deren Referenzcharakter jedoch diskutiert und sehr bedacht gewählt werden muss. Schön wäre es, wenn diese Projekte auch zu einem gewissen Grad methodenorientiert und selbstreflexiv konzipiert würden und durch eine systematische Begleitung sowie Dokumentation und Vernetzung ihre Erkenntnisse in die neu entstehende Forschungslandschaft einbringen würden. Ferner stimme ich Chrisowalandis Deligio zu, dass Plattformen und Austausch wichtig wären, und zwar nicht nur für Nachwuchswissenschaftler*innen, denn gerade im digitalen Bereich ist und wird lebenslanges Lernen notwendig sein. Insbesondere interdisziplinäre Forschung stellt hier eine ganz andere Herausforderung dar, denn man kann nicht auf allen Gebieten den aktuellen Forschungsstand kennen. Wir müssen uns daher fragen: Wann diskutieren wir Details und wann Übergreifendes, wie Methoden und Theorien. Wir hatten hier jetzt einen starken Fokus auf das Bild. Noch herausfordernder wird es werden, wenn wir Dinge in ihrer ganzen Komplexität in den Blick nehmen und dann etwa ihr Material, ihre Oberflächenstruktur, Fragmentierungen, Assemblagen usw. mit betrachten. Also ich würde sagen, bei Text und Bild sind wir schon relativ gut und weit vorangekommen. Wenn wir jedoch Objekte digital erforschen wollen, stellen sich viele weitere Fragen, und sei es nur, wo die Grenze eines Objekts ist? Es ist eine ganz große Herausforderung, wenn wir dann zukünftig z.B. auch den Bildträger stärker mitberücksichtigen.

Elisabeth Günther:

Diesen Ausblick über das Bild hinaus auf das Objekt würde ich gerne gewissermaßen als Schlusswort stehen lassen. Es gibt natürlich noch viele weitere Aspekte der Mustererkennung, die wir jetzt noch besprechen könnten oder die wir hätten besprechen können. Das war jetzt nur ein kleiner Einblick, aber ich fand sehr interessant, was hier angesprochen wurde und in welche Richtung sich unser Gespräch entwickelt hat. Wichtig fand ich insbesondere den Bedarf, der sich hier gezeigt hat, sich untereinander stärker auszutauschen und darüber nachzudenken, wie wir besser zusammenarbeiten können, welche Voraussetzungen für Kollaborationen zu schaffen wären. Diese Diskussionen müssen natürlich

84 <https://www.faustedition.net> (abgerufen am 03.06.2023).

in einem Kontext geführt werden, der weit über diesen kleinen Gesprächskreis hinausgeht. Denn das ist, und das hat sich auch in den anderen Gruppeninterviews im Nachgang der Tagung *NEUES SEHEN* gezeigt, Grundlage, um die Entwicklung der archäologischen Disziplinen zu reflektieren und diese bewusst zu gestalten. Wir stehen vor sehr großen Veränderungen und Transformationen – oder sind eigentlich schon mittendrin –, die sich auf Forschung und ebenso auf die universitäre Lehre und die Zusammenarbeit mit Institutionen der Denkmalpflege, der Museen, erstrecken. Es gibt also noch viel Gesprächspotenzial und ich möchte mich abschließend schon einmal herzlich für die Diskussionen und Anregungen in diesem Kreis bedanken.

Literaturverzeichnis

- Argyrou / Agapiou (2022): A. Argyrou / A. Agapiou, A Review of Artificial Intelligence and Remote Sensing for Archaeological Research, *Remote Sensing* 2022, 14,23 (2022), <https://www.mdpi.com/2072-4292/14/23/6000> (abgerufen am 01.10.2023).
- Arnold / Tilton (2021): T. Arnold / L. Tilton, Distant Viewing: Analyzing Large Visual Corpora, *Digital Scholarship in the Humanities* 36,2 (2021), e1–e14, <https://doi.org/10.1093/digitalsh/fqz013> (abgerufen am 01.10.2023).
- Bawden et al. (2016): T. Bawden / D. Bonatz / N. Dietrich / J. Fabricius / K. Gludovatz / S. Muth / T. Poiss / D. A. Werning, Early Visual Cultures and Panofsky's Perspektive als 'symbolische Form', *eTopoi. Journal for Ancient Studies*, Special Volume 6 (2016) (= G. Graßhoff / M. Meyer (Hrsgg.), *Space and Knowledge. Topoi Research Group Articles*, Berlin 2016, 525–570), http://www.topoi.org/wp-content/uploads/2017/01/eTopoi_Sp6_Bawden-et-al.pdf (abgerufen am 01.10.2023).
- Bendschus et al. (2022): T. Bendschus / L. Mührenberg / C. Reinhardt / R. Kosti / P. Madhu / U. Versteegen, Computer Vision und Deep Learning aus Perspektive der archäologischen Bildanalyse, in: L. Dieckmann / B. Pflöging / G. Schelbert / T. Wübbenet (Hrsgg.), *4D: Dimensionen – Disziplinen – Digitalität – Daten, Computing in Art and Architecture* 6, Heidelberg 2022, <https://books.ub.uni-heidelberg.de/arhistoricum/catalog/book/1100/chapter/15427> (abgerufen am 01.10.2023).
- Buchholz (1987): H.-G. Buchholz, Brettspielende Helden, in: S. Laser (Hrsg.), *Sport und Spiel. Mit einem Exkurs über „Rasseln und Schellen, Reifen, Wippen und Schaukeln“ und einem Anhang über „Brettspielende Helden“*, *Archaeologia Homerica* 3, Kapitel T, Göttingen 1987.
- Colbert de Beaulieu (1973): J.-B. Colbert de Beaulieu, *Traité de numismatique celtique 1, Méthodologie des ensembles*, *Ann. Litt. Univ. Besançon* 135, Besançon 1973.
- de la Tour (1892): H. de la Tour, *Atlas de monnaies gauloises préparé par la commission de topographie des gaules et publié sous les auspices du Ministère de l'Instruction publique*, Paris 1892.
- Foster (1988): H. Foster, Preface, in: H. Foster (Hrsg.), *Vision and Visuality*, Seattle 1988, ix–xiv.
- Geirhos et al. (2019): R. Geirhos / P. Rubisch / C. Michaelis / M. Bethge / F. A. Wichmann / W. Brendel, ImageNet-trained CNNs are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness, in: *International Conference on Learning Representations (ICLR) 2019*, <https://openreview.net/forum?id=Bygh9j09KX> (aktualisiert am 12.02.2023, abgerufen am 20.03.2023).
- Göbl (1987): R. Göbl, *Numismatik. Grundriss und wissenschaftliches System*, München 1987.
- Gonthier et al. (2019): N. Gonthier / Y. Gousseau / S. Ladjal / O. Bonfait, Weakly Supervised Object Detection in Artworks, in: L. Leal-Taixé / S. Roth (Hrsgg.), *Computer Vision, ECCV 2018 Workshops*, Cham 2019, 692–709, https://link.springer.com/chapter/10.1007/978-3-030-11012-3_53 (abgerufen am 01.10.2023).
- Goodfellow et al. (2018): I. Goodfellow / Y. Bengio / A. Courville, *Deep Learning. Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze*, Frechen 2018.
- Graepler (2016): D. Graepler, Künstlerhand und Kennerauge: Die Zuschreibung als archäologisches Methodenproblem, in: N. Eschbach / S. Schmidt (Hrsgg.), *Töpfer – Maler – Werkstatt. Zuschreibungen in der griechischen Vasenmalerei und die Organisation antiker Keramikproduktion*, Beihefte zum *Corpus Vasorum Antiquorum* 7, München 2016, 14–24.

- Günther (2021): E. Günther, Mehrdeutigkeiten antiker Bilder als Deutungspotenzial. Zu den Interdependenzen von Affordanzen und *frames* im Rezeptionsprozess, in: E. Günther – J. Fabricius (Hrsgg.), Mehrdeutigkeiten. Rahmentheorien und Affordanzkonzepte in der archäologischen Bildwissenschaft, *Philippika* 147, Wiesbaden 2021, 1–40.
- Günther (2022): E. Günther, Map of Things: West-östliche Narrative über die antiken Seidenstraßen, *Junges Forum für Sammlungs- und Objektforschung* 6 (2022), 56–65, https://edoc.hu-berlin.de/bitstream/handle/18452/26681/07_Guenther.pdf (abgerufen am 01.10.2023).
- Günther / Günther (2021): E. Günther / S. Günther, Interaktiv, international, interkulturell. Das Ausstellungsprojekt „Machtszenarien – Scenarios of Power – 无处不在的权力. Roles and Actions of Roman Emperors“, *Geldgeschichtliche Nachrichten* 316 (2021), 218–223.
- Günther / Günther (2022a): E. Günther / S. Günther, Mit Plinius dem Älteren im antiken Rom – Ein internationales Digital Mapping Projekt, in: P. Reinhardt / L. Scheuermann (Hrsgg.), *Geschichtswissenschaften in Pandemiezeiten. Beiträge zur Digitalen Wissensvermittlung und Public History*, Abhandlungen der Arbeitsgemeinschaft Alte Geschichte und EDV, Gutenberg 2022, 171–188.
- Günther / Günther (2022b): E. Günther / S. Günther, Augustus in Saigon!? Die Rezeption westlicher Antike in der kolonialen Bilderwelt und post-kolonialen Gesellschaft Vietnams, *Gymnasium* 129 (2022), 453–490.
- He et al. (2015): K. He / X. Zhang / S. Ren / J. Sun, Deep Residual Learning for Image Recognition, <https://arxiv.org/abs/1512.03385> (submitted 10.12.2015, abgerufen am 20.03.2023).
- Hilgert et al. (2018): M. Hilgert / H. Simon / K. P. Hofmann (Hrsgg.), *Objekt epistemologien. Zur Vermessung eines transdisziplinären Forschungsraums*, Berlin 2018, <https://refubium.fu-berlin.de/handle/fub188/23823> (abgerufen am 01.10.2023).
- Hofmann / Patzke (2012): K. P. Hofmann / S. Patzke, Von Athen nach Etrurien. Zum Diffusionsprozess der entlehnten Innovation „ceramica sovraddipinta“, in: A. Kern / J. K. Koch / I. Balzer / J. Fries-Knoblach / K. Kowarik / C. Later / P. C. Ramsel / P. Trebsche / J. Wiethold (Hrsgg.), *Technologieentwicklung und -transfer in der Hallstatt- und Latènezeit. Beiträge zur Internationalen Tagung der AG Eisenzeit und des Naturhistorischen Museums Wien, Prähistorische Abteilung – Hallstatt 2009, Beiträge zur Ur- und Frühgeschichte Mitteleuropas* 65, Langenweißbach 2012, 83–101.
- Hofmann et al. (2019): K. P. Hofmann / S. Grunwald / F. Lang / U. Peter / K. Rösler / L. Rokohl / S. Schreiber / K. Tolle / D. Wigg-Wolf, Ding-Editionen. Vom archäologischen (Be-)Fund übers Corpus ins Netz, *DAI e-forschungsberichte* 2019-2, <https://publications.dainst.org/journals/efb/2236> (abgerufen am 01.10.2023).
- Kaempf-Dimitriadou (1979): S. Kaempf-Dimitriadou, Die Liebe der Götter in der attischen Kunst des 5. Jahrhunderts v. Chr., *Beiheft zur Halbjahresschrift Antike Kunst* 11, Bern 1979.
- Kipke et al. (2022): M. Kipke / L. Brinkmeyer / S. Bagayoko / L. Schmidt-Thieme / M. Langner, Deep Level Annotation for Painter Attribution on Greek Vases utilizing Object Detection, *Proceedings of the 4th ACM International workshop on Structuring and Understanding of Multimedia Heritage Contents* 2022, 23–31, <https://doi.org/10.1145/3552464.3555684> (abgerufen am 17.10.2023).
- Langner (2017): M. Langner, Übersetzungsleistungen antiker Bildproduzenten. Überlegungen zu aktuellen Positionen der archäologischen Bildanalyse, *Saeculum. Jahrbuch für Universalgeschichte* 67 (2017), 67–101.

- Lin et al. (2014): T.-Y. Lin / M. Maire / S. Belongie / L. Bourdev / R. Girshick / J. Hays / P. Perona / D. Ramanan / C. L. Zitnick / P. Dollár, Microsoft COCO. Common Objects in Context, in: D. Fleet / T. Pajdla / B. Schiele / T. Tuytelaars (Hrsgg.), Computer Vision – ECCV 2014, Proceedings of the 13th European Conference on Computer Vision 5, Cham 2014, 740–755, https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48 (abgerufen am 20.03.2023).
- Liu et al. (2016): W. Liu / D. Anguelov / D. Erhan / C. Szegedy / S. Reed / C.-Y. Fu / A. C. Berg, SSD. Single Shot Multibox Detector, in: B. Leibe / J. Matas / N. Sebe / M. Welling (Hrsgg.), Computer Vision – ECCV 2016, Proceedings of the 14th European Conference on Computer Vision 1, Cham 2016, 21–37, https://link.springer.com/chapter/10.1007/978-3-319-46448-0_2 (abgerufen am 20.03.2023).
- Lobinger (2015): K. Lobinger, Visualität, in: A. Hepp / F. Krotz / S. Lingenberg / J. Wimmer (Hrsgg.), Handbuch Cultural Studies und Medienanalyse. Medien – Kultur – Kommunikation, Wiesbaden 2015, 91–100, <https://doi.org/10.1007/978-3-531-19021-1> (abgerufen am 17.10.2023).
- Maaten / Hinton (2008): L. van der Maaten / G. Hinton, Visualizing Data Using t-SNE, Journal of Machine Learning Research 9 (2008), 2579–2605.
- Madhu et al. (2019): P. Madhu / R. Kosti / L. Mührenberg / P. Bell / A. Maier / V. Christlein, Recognizing Characters in Art History Using Deep Learning, in: V. Gouet-Brunet (Hrsg.), SUMAC 2019 – The 1st workshop on Structuring and Understanding of Multimedia heritAge Contents, Nizza, 21.–25. Oktober 2019, New York 2019, 15–22, <https://doi.org/10.1145/3347317.3357242> (abgerufen am 17.10.2023).
- Madhu et al. (2022): P. Madhu / A. Meyer / M. Zinnen / L. Mührenberg / D. Suckow / T. Bendschus / C. Reinhardt / P. Bell / U. Verstegen / R. Kosti / A. Maier / V. Christlein, One-Shot Object Detection in Heterogeneous Artwork Datasets, in: 2022 Eleventh International Conference on Image Processing Theory, Tools and Applications (IPTA), 1–6, <https://ieeexplore.ieee.org/document/9784141> (abgerufen am 05.06.2023).
- Madhu et al. (2023): P. Madhu / A. Villar-Corrales / R. Kosti / T. Bendschus / C. Reinhardt / P. Bell / A. Maier / V. Christlein, Enhancing Human Pose Estimation in Ancient Vase Paintings via Perceptually-grounded Style Transfer Learning, Article No.: 16, Journal on Computing and Cultural Heritage 16,1 (2023), 1–17, https://www.researchgate.net/publication/346933628_Enhancing_Human_Pose_Estimation_in_Ancient_Vase_Paintings_via_Perceptually-grounded_Style_Transfer_Learning (abgerufen am 01.10.2023).
- McInnes / Healy (2018): L. McInnes / J. Healy, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, <https://www.semanticscholar.org/paper/UMAP%3A-Uniform-Manifold-Approximation-and-Projection-McInnes-Healy/3a288c63576fc385910cb5bc44eaea75b442e62e> (abgerufen am 05.06.2023).
- Mührenberg / Verstegen (2020): L. Mührenberg / U. Verstegen, CA 2.x – Christliche Archäologie im *inverted classroom*. Ein Beitrag zur videobasierten digitalen Lehre an der Friedrich-Alexander-Universität Erlangen-Nürnberg, in: J. Handke / S. Zeaiter (Hrsgg.), Inverted Classroom – Past, Present & Future. Kompetenzorientiertes Lehren und Lernen im 21. Jahrhundert, 8. ICM-Fachtagung an der Philipps-Universität Marburg, Baden-Baden 2020, 165–178.

- Mührenberg / Verstegen (2022): L. Mührenberg / U. Verstegen, Christliche Archäologie im „inverted classroom“ und auf „virtueller Exkursion“, in: K. Döring / S. Haas / M. König / J. Wettlaufer (Hrsgg.), Digital History. Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft. Berlin 2022, Studies in Digital History and Hermeneutics 6, 281–302, <https://www.degruyter.com/document/doi/10.1515/9783110757101-015/html?lang=de> (abgerufen am 01.10.2023).
- Mührenberg et al. (2022): L. Mührenberg / T. Bendschus / C. Reinhardt / U. Verstegen, Kann eine Künstliche Intelligenz antike Kunst verstehen?, Antike Welt 3 (2022), 44–51.
- Pfisterer-Haas (2004): S. Pfisterer-Haas, Helden beim Brettspiel, in: R. Wünsche / F. Knauß (Hrsgg.), Lockender Lorbeer. Sport und Spiel in der Antike, München 2004, 381–385.
- Radford et al. (2021): A. Radford / J. W. Kim / C. Hallacy / A. Ramesh / G. Goh / S. Agarwal / G. Sastri / A. Aspell / P. Mishkin / J. Clark / G. Krueger / I. Sutskever, Learning Transferable Visual Models From Natural Language Supervision, in: International Machine Learning Society (Hrsg.), Proceedings of the 38th International Conference on Machine Learning 139, 2021, Red Hook 2022, <https://arxiv.org/pdf/2103.00020.pdf> (abgerufen am 01.10.2023).
- Redmon et al. (2016): J. Redmon / S. Divvala / R. Girshick / A. Farhadi, You Only Look Once. Unified, Real-Time Object Detection, Computer Vision and Pattern Recognition 2016, 779–788, <https://arxiv.org/abs/1506.02640> (abgerufen am 20.03.2023).
- Redmon / Farhadi (2018): J. Redmon / A. Farhadi, Yolov3. An Incremental Improvement, arXiv preprint: arXiv:1804.02767 (2018), <https://arxiv.org/abs/1804.02767> (abgerufen am 04.06.2023).
- Reinhardt / Bendschus (2021): C. Reinhardt / T. Bendschus, Simulating Digital Classics in Classical Teaching: A New Approach for New Demands in the Education of Classical Archaeology, in: S. Feuser / S. Merten / K. Wesselmann (Hrsgg.), Teaching Classics in the Digital Age, Kiel 2021, 117–128, https://macau.uni-kiel.de/receive/macau_mods_00001375 (abgerufen am 01.10.2023).
- Rother et al. (2004): C. Rother / V. Kolmogorov / A. Blake, “GrabCut” Interactive Foreground Extraction Using Iterated Graph Cuts, ACM transactions on graphics (TOG) 23.3 (2004), 309–314, <https://cvg.ethz.ch/teaching/cvl/2012/grabcut-siggraph04.pdf> (abgerufen am 01.10.2023).
- Sager et al. (2021): C. Sager / C. Janiesch / P. Zschech, A Survey of Image Labelling for Computer Vision Applications, Journal of Business Analytics (2021), 1–20, <https://arxiv.org/ftp/arxiv/papers/2104/2104.08885.pdf> (abgerufen am 01.10.2023).
- Servaraju et al. (2017): R. R. Selvaraju / M. Cogswell / A. Das / R. Vedantam / D. Parikh / D. Batra, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization (IEEE International Conference on Computer Vision, 2017), <https://arxiv.org/abs/1610.02391> (abgerufen am 05.06.2023).
- Simonyan / Zisserman (2015): K. Simonyan / A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, ICLR 2015, <https://arxiv.org/pdf/1409.1556.pdf> (abgerufen am 01.10.2023).
- Stansbury-O’Donnell (2009): M. Stansbury-O’Donnell, The Structural Differentiation of Pursuit Scenes, in: D. Yatromanolakis (Hrsg.), An Archaeology of Representations. Ancient Greek Vase-Painting and Contemporary Methodologies, Athen 2009, 341–372.
- Strezoski / Worring (2017): G. Strezoski / M. Worring, Omniart. Multi-Task Deep Learning for Artistic Data Analysis, 2017, arXiv preprint: <https://arxiv.org/abs/1708.00684> (abgerufen am 05.06.2023).

- Van Zuijlen et al. (2021): M. Van Zuijlen / H. Lin / K. Bala / S. Pont / M. W. A. Wijntjes, Materials In Paintings (MIP): An Interdisciplinary Dataset for Perception, Art History, and Computer Vision, PLoS ONE 26,8 (2021), e0255109, <https://arxiv.org/pdf/2012.02996.pdf> (abgerufen am 01.10.2023).
- Verstegen et al. (2022): U. Verstegen / L. Mührenberg / F. Nicol / J. Abura, Virtual Reality in der Christlichen Archäologie. Zur Konzeptionierung virtueller Exkursionen und ihrem Einsatz in der universitären Lehre, in: V. Pirker / K. Pišonić (Hrsgg.), Virtuelle Realität und Transzendenz. Theologische und pädagogische Erkundungen, Freiburg 2022, 191–221.
- Verstegen / Kremer (2023): U. Verstegen / D. Kremer, The Nuremberg Stations of the Cross. Enriching On-Site Experiences of a Historical Pathway by the “FAU GeoExplorer” App, in: S. Münster / C. Kröber / A. Pattee / F. Niebling (Hrsgg.), Research and Education in Urban History in the Age of Digital Libraries. Third International Workshop, UHDL 2023, Munich, Germany, March 27–28, 2023, Revised Selected Papers. Cham 2023, Communications in Computer and Information Science 1853, 207–219, https://doi.org/10.1007/978-3-031-38871-2_12 (abgerufen am 01.10.2023).
- Voßkühler (2009): A.Voßkühler, OGAMA Description. A Software to Record, Analyze and Visualize Gaze and Mouse Movements in Screen Based Environments, <http://www.ogama.net/sites/default/files/pdf/OGAMA-DescriptionV25.pdf> (abgerufen am 05.06.2023).
- Wigg-Wolf et al. (2022): D. Wigg-Wolf / K. P. Hofmann / K. Tolle / K. Rösler / M. Möller / C. Deligio / J. Tietz / C. v. Nicolai, ClaReNet. Klassifikation und Repräsentation keltischer Münzprägungen im Netz. Das Projekt von 2021 bis 2024, DAI e-forschungsberichte 2022-2, <https://publications.dainst.org/journals/efb/article/view/3975/7659> (abgerufen am 01.10.2023).
- Winter (1903): Franz Winter, Die Typen der figürlichen Terrakotten. Die antiken Terrakotten III, Berlin 1903, <http://digi.ub.uni-heidelberg.de/diglit/winter1903> (abgerufen am 05.06.2023).

Abbildungsnachweise

Abb. 1: M. Langner / M. Kipke.

Abb. 2: M. Langner / M. Kipke.

Abb. 3: © G. Pöhlein, Antikensammlung der FAU Erlangen-Nürnberg, CC BY-NC-SA 4.0.

Abb. 4: British Museum, Inv.-Nr. 1836,0224.68 [E 512]. © The Trustees of the British Museum. Shared under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license.

Abb. 5: A. Villar-Corrales / R. Kosti / P. Madhu, CC BY-NC-SA 4.0.

Abb. 6: Objekterkennung / Segmentation durchgeführt von R. Kosti / P. Madhu, CC BY-NC-SA 4.0.

Abb. 7: R. Kosti / P. Madhu (mit PixPlot), CC BY-NC-SA 4.0.

Abb. 8: Grafik: R. Kosti / P. Madhu, CC BY-NC-SA 4.0.

Abb. 9: *Iconographics*-Projekt, FAU; Objekterkennung durchgeführt von R. Kosti und P. Madhu, CC BY-NC-SA 4.0.

Abb. 10: Grafik: C. Deligio; Fotos: © Jersey Heritage; all rights reserved.

Abb. 11: GIF: C. Deligio; Fotos: © Jersey Heritage; all rights reserved.

Autoreninformation⁸⁵

Dr. Torsten Bendschus
Institut für Klassische Archäologie
Friedrich-Alexander-Universität Erlangen-Nürnberg
Kochstr. 4 / 19
91054 Erlangen
E-Mail: t.bendschus@hotmail.de

Chrisowalandis Deligio, M.Sc.
Goethe-Universität Frankfurt am Main
Robert-Mayer-Str. 10 (5. Stock)
60325 Frankfurt am Main
E-Mail: Deligio@em.uni-frankfurt.de

Dr. Elisabeth Günther
Universität Heidelberg
Institut für Klassische Archäologie und Byzantinische Archäologie
Marstallhof 4
69117 Heidelberg
E-Mail: elisabeth.guenther@zaw.uni-heidelberg.de

Dr. Kerstin P. Hofmann
Römisch-Germanische Kommission
des Deutschen Archäologischen Instituts
Palmengartenstr. 10–12
60325 Frankfurt am Main
E-Mail: kerstin.hofmann@dainst.de

Marta Kipke, M.A.
Georg-August-Universität Göttingen
Institut für Digital Humanities
Nikolausberger Weg 23
37073 Göttingen
E-Mail: marta.kipke@uni-goettingen.de

Prof. Dr. Martin Langner
Georg-August-Universität Göttingen
Institut für Digital Humanities
Nikolausberger Weg 23
37073 Göttingen
E-Mail: martin.langner@uni-goettingen.de

Prof. Dr. Corinna Reinhardt
Institut für Archäologie
Fachbereich Klassische Archäologie
Rämistr. 73
CH-8006 Zürich
E-Mail: corinna.reinhardt@archaeologie.uzh.ch

85 Die Rechte für Inhalt, Texte, Graphiken und Abbildungen liegen, wenn nicht anders vermerkt, bei den Autoren. Alle Inhalte dieses Beitrages unterstehen, soweit nicht unter den Abbildungsnachweisen anders gekennzeichnet, der Lizenz CC BY 4.0.

Dr. Katja Rösler
Römisch-Germanische Kommission
des Deutschen Archäologischen Instituts
Palmengartenstr. 10–12
60325 Frankfurt am Main
E-Mail: katja.roesler@dainst.de
ORCID-ID: <https://orcid.org/0000-0002-7520-7079>

Prof. Dr. Ute Verstegen
LS Christliche Archäologie
Friedrich-Alexander-Universität Erlangen-Nürnberg
Kochstr. 6
91054 Erlangen
E-Mail: ute.verstegen@fau.de
ORCID-ID: <https://orcid.org/0000-0002-6577-5144>