

# Bilddigitalisierung

Hubertus Kohle

 <https://orcid.org/0000-0003-3162-1304>

**Abstract** Ein Bild sagt mehr als tausend Worte, heißt es immer wieder. Und trotzdem – oder gerade deswegen – ist es in der europäischen Geistesgeschichte an den Rand gedrängt oder als irreführend und unzuverlässig gebrandmarkt worden. Auch in den Digital Humanities widmet man sich verspätet dem Bild, was sich selbst in deren Verfasstheit an deutschen Universitäten äußert. Im Beitrag geht es um die technischen Eigenschaften des digitalen Bildes, seine grundlegenden Eigenheiten als Gebilde von unterschiedlich hoch aufgelösten Pixelmengen und um die Rastergraphik im Gegensatz zur Vektorgraphik. Es folgen Überlegungen zum Aufbau von Bilddatenbanken und deren elementaren Eigenschaften sowie die Beschreibung von komplexen Suchmöglichkeiten über astronomisch große Bilddatenmengen hinweg, die im Rahmen der Entwicklung von multimodalen Large Language/Image/Audio Models auch semantisch immer treffgenauer werden. Abgeschlossen wird die Darstellung mit einigen kursorischen Überlegungen zu den KI-gesteuerten Bildgeneratoren.

**Keywords** Digitales Bild, Bilddatenbanken, Digitale Bildanalyse

## 1. Einführung

Die Priorisierung des Wortes in der europäischen Geistesgeschichte und in gewisser Weise auch die durchgängig vorherrschende Skepsis gegenüber dem Bild, wie wir sie vor allem auch aus den reformierten Theologien kennen, wiederholen sich im Kleinen in der Geschichte der Digital Humanities. Diese beginnen in der Mitte des 20. Jahrhunderts in einem philosophisch-theologischen Kontext mit den Bemühungen des Jesuiten Roberto Busa (Father Busa), der seine Beziehungen zu Thomas J. Watson, dem Gründer und Chef der *International Business Corporation* (IBM) dazu nutzte, den Computer für die Erstellung eines Indexes für die Werke des Thomas von Aquin heranzuziehen. Dabei verwendete er die Maschine naheliegender Weise ausschließlich als textverarbeitende. Mit dem Bild ging es erst ein viertel Jahrhundert später los, zunächst eher mit wiederum textlichen Metadaten zu diesem Bild, dann aber auch schon seit den 1980er Jahren mit Versuchen, das Bild als solches zu adressieren (s. Vaughan 1987, 215–221).

Die Gründe dafür liegen auf der Hand und sind ideologischer wie technischer Natur. Vorrang im europäischen Kontext und dessen Philosophiegeschichte, die ja

nach einer berühmten Bemerkung Alfred North Whiteheads eine Serie von „Fußnoten zu Platon“ ist, hat das Wort und dessen Nähe zur Idee. Konkreter relevant aber dürfte die Tatsache sein, dass ein Bild rein technisch gesehen digital aufwändiger zu verarbeiten ist. Vaughan selber äußerte einmal, er habe sein bildorientiertes Suchsystem (genannt „Morelli“) nicht erproben können, weil in der Zeit seiner Erfindung kaum größere Mengen von digitalen Bildern vorhanden gewesen seien.<sup>1</sup> Dafür gab es Gründe: Ein vernünftig aufgelöstes Bild – was das heißt, wird später diskutiert – konnte leicht mehrere Hundert Kilobyte groß sein, wenn es in die Nähe von Diapositiv-Qualität kam, wurden es auch gerne mal mehrere Megabyte. Höchstauflösende Bilder gar, so wie sie in bestimmten Fällen von *Google Arts & Culture*<sup>2</sup> angeboten werden, erreichen mehrere Hundert Megabyte. Speicherplatz war damals rar, und eine 20-Megabyte Festplatte kostete mehrere Tausend Mark. Auf so eine Festplatte passten andererseits locker 20 Bücher.

Aber auch jetzt, wo die technischen Voraussetzungen besser sind, ist die Stellung des Bildes innerhalb der Digital Humanities eher marginal. Meistens sind diese in den Sprach- und Literaturwissenschaften angesiedelt und daher rein textorientiert. Nur in wenigen Fällen gehen sie in den Objektbereich, hier aber nicht in erster Linie in den der Bilder. Erst zu allerletzt werden stärker Professuren im Bereich des Visual Computings innerhalb der Digital Humanities ausgeschrieben, auch wenn sie sich in der allgemeinen Informatik schon länger eine Position erobert haben.<sup>3</sup> Hier geht es in erster Linie um Augmented und Virtual Reality, maschinelles Sehen und Bildverarbeitung. Anwendungen dafür finden sich vor allem in der Medizin, der Architektur, im Design und im Gaming. In den meisten Fällen spielen Verfahrensweisen der Künstlichen Intelligenz eine wichtige Rolle.

## 2. Technische Grundlagen

Photographische Reproduktionen nach Originalen bestimmen schon seit dem späten 19. Jahrhundert z. B. in der Kunstgeschichte, der wahrscheinlich für die Theologie wichtigsten Hilfswissenschaft im Bereich des Bildes, die Lehr- und auch Forschungspraxis (vgl. Dilly 1979). Davor gab es graphische Techniken, die in großem Umfang auch für die Reproduktion von Kunstwerken genutzt wurden und das exklusive Unikat in eines der gesellschaftlichen Kommunikation verwandelten. Vor allem der

1 Mündliche Mitteilung.

2 S. <https://artsandculture.google.com>, zuletzt aufgerufen am 16.06.2024, zuletzt aufgerufen am 16.06.2024.

3 Ein interdisziplinär, aber mit kulturwissenschaftlichem Schwerpunkt versehenes Schwerpunktprogramm der DFG widmet sich seit einigen Jahren dem digitalen Bild (<https://www.digitalesbild.gwi.uni-muenchen.de>, zuletzt aufgerufen am 16.06.2024).

Kupferstich diente solchen Bildreproduktionen. Im späten 20. Jahrhundert kamen dann digitale Reproduktionen hinzu, die schon nach wenigen Jahren traditionelle analoge Produktionsformen weitgehend ersetzt haben. Diatheken auf Chemiebasis werden heute entsorgt, ins Uniarchiv verschoben oder in den Keller. Dass sie zuweilen gar zum Gegenstand eigener kunsthistoriographisch-wissenschaftlicher Analysen gemacht werden, belegt ihre Marginalität eher, als dass es sie widerlegt.

Das digitale Bild unterscheidet sich vom analogen dadurch, dass es eine diskrete Sammlung von rasterförmig in senkrechten und waagerechten Reihen angeordneten Pixeln (picture elements, Bildpunkte) darstellt, die in einem elektronischen Belichtungsvorgang entstehen (zur Einführung s. Besser et al. 2003). Diskret heißt hier, dass der Übergang von einem Pixel zum anderen in Sprüngen vollzogen wird, während er im Analogen kontinuierlich erfolgt. Mit den Digitalkameras und den Scannern gibt es seit den 1980er Jahren Geräte, die diesen Belichtungsvorgang durchführen, im ersten Fall synchron, im zweiten diachron, in einem konsekutiv erfolgenden Abtastvorgang. Insbesondere Scanner werden für eine Reihe von unterschiedlichen analogen Ausgangsprodukten angeboten, so z. B. Diascanner, mit denen analoge Diapositive digitalisiert werden können.

In Verbindung mit der Tatsache, dass der Computer es nur mit Nullen und Einsen zu tun hat (dies nur in Parenthese), wird die diskrete Organisation der Pixel aus geisteswissenschaftlicher Perspektive gerne kritisch als Hinweis darauf verstanden, dass digital gestützte Analytiken nur mit Eindeutigkeiten umgehen könnten und dem Grauwert keinen Raum ließen. Bedenkt man allerdings, dass mit Steigerung der Auflösung das Diskrete sich dem Analogen gleichsam asymptotisch annähert, dürfte klar werden, dass hier ein eher nur theoretischer Unterschied vorliegt. Über das Problem der Unschärfe, das hiermit in engem Zusammenhang steht, wurde in den letzten Jahren in Digital Humanities Kreisen intensiv diskutiert (zur Einführung s. Borek et al. 2022).

Je nach Auflösung besteht das Raster aus einer unterschiedlich großen Anzahl von Bildpunkten, wobei der Detailreichtum einer Abbildung umso höher ist, je größer die Anzahl dieser Bildpunkte ist. Eine Maßeinheit für die Auflösung ist dpi (dots per inch), das die Anzahl der Bildpunkte pro inch = 2,54 cm bezeichnet. Dabei ist die Größe der Vorlage eine entscheidende Bestimmungsgröße. Digitalisiere ich ein  $2,54 \times 1,27$  Meter großes Ölbild mit 100 dpi, dann entstehen 50 Millionen Punkte ( $100 \times 100$  für die Höhe  $\times 50 \times 100$  für die Breite), mache ich das Gleiche mit einer nur  $25,4 \times 12,7$  cm großen analogen Reproduktion des Bildes, komme ich auf 500 000 Punkte ( $10 \times 100$  für die Höhe  $\times 5 \times 100$  für die Breite). Neben schwarz-weißen und grauwertigen Abbildungen, bei denen im ersten Fall jede Erscheinung auf entweder schwarz oder weiß reduziert wird, im zweiten wiederum unterschiedlich zahlreiche Helligkeitswerte erzeugt werden, herrschen heute Farabbildungen vor, die eine möglichst realistische Darstellung der Welt hervorbringen sollen. Dies gilt umso mehr, je größer die Farbtiefe ist, also je mehr unterschiedliche Farben dargestellt werden. Das vorherrschende Farbmodell, in dem die Farbzusammenstellung erfolgt, ist das RGB-Farbmodell, in dem

die Grundfarben Rot, Blau und Grün additiv, d. h. durch Übereinanderlagerung in alle Farbtöne gemischt werden können. Realistische Farbwirkung wird erzielt, wenn jede der drei Farben mit 8 Bit = 256 Farben codiert wird, was in der Summe eine Zahl von 16,7 Millionen Farben ergibt.  $(256 \times 256 \times 256)$  Hohe Farbtiefe bei gleichzeitig hoher Auflösung erzeugt große Dateien. Wenn schon deren Speicherung und Verbreitung aktuell mit den sehr großen Speicherkapazitäten moderner Festplatten und den Bandbreiten der Netze einigermaßen problemlos und schnell möglich ist, stellt doch das Berechnen solcher Bilder immer noch eine Leistung dar, die die potentesten Maschinen erfordert. Erleichtert wird das immerhin dadurch, dass man die Bilder in komprimierter Form abspeichern kann, was deren Größe um einen zweistelligen Faktor reduzieren kann, aber in dem geläufigen jp(e)g-Format (*joint photographic experts group*) auch zu Verlusten führt, die für das menschliche Auge zunächst kaum sichtbar sind, bei stärkerer Komprimierung aber durchaus auffallen können. Die Kompressionsstärke kann beim Verarbeiten der Dateien vorgewählt werden.

Außer dem genannten jpg-Format wurde eine Reihe von weiteren Formaten entwickelt, zunächst natürlich die unveränderten Ausgangsformate. Dazu gehört das bmp (*bitmap*) und das tiff (*tagged image file format*), wobei man eine tiff-Datei auch komprimieren kann. Im Gegensatz zum jpg ist das png (*portable network graphics*) verlustfrei, obwohl es – wenn auch weniger stark – die Dateigröße komprimiert. Das gif (*graphics interchange format*) eignet sich vor allem für bewegte Animationen. Die meisten Informationen enthält das herstellerspezifische raw-Format (engl. „roh“), das von Profis als Ausgangsformat für Bildbearbeitungen bevorzugt wird, aber wie bmp sehr viel Speicherplatz benötigt.

Neben der Rastergraphik gibt es die Vektorgrafik, die nicht jeden einzelnen Punkt codiert, sondern die geometrische und farbige Natur eines bildnerischen Gegenstandes mathematisch definiert. Letztere hat Vor- und Nachteile: wenn eine einheitlich rote Linie nicht durch eine ganze Anzahl von gleichgefärbten, eben roten Pixeln dargestellt werden muss, sondern als Anweisung: Mache alle Punkte zwischen Koordinate x und Koordinate y gleich rot, dann braucht das erstens weniger Speicherplatz und ist zweitens besser skalierbar, kann also verlustfrei vergrößert werden. Bei einer Rastergraphik dagegen wird die Erscheinung unklarer, weil immer deutlicher die einzelnen Pixel hervortreten. Der Nachteil besteht darin, dass die Verarbeitung langsamer abläuft, da eine Koordinate als Zahl natürlich nicht sichtbar ist und erst wieder in ein Pixel konvertiert (gerendert) werden muss. Im Übrigen funktioniert das Prinzip der Vektorgraphik natürlich am besten bei schematischen Visualisierungen mit wenig lokalen Farbvariationen.

### 3. Bilddatenbanken

Im Bildkunstbereich gibt es inzwischen eine ganze Reihe von Datenbanken, in denen digitale Reproduktionen nach zwei- und dreidimensionalen Kunstwerken vorgehalten werden. Dazu gehören in erster Linie Museumsverzeichnisse, die den Vorteil haben, dass die Digitalisate hier mit großer Wahrscheinlichkeit nach den Originalen hergestellt wurden und nicht nach gedruckten Reproduktionen, bei denen es immer wieder zu hässlichen Moiréeffekten kommt, die mit den negativen Auswirkungen der Digitalisierung eines Druckrasters zu tun haben. Übergreifende Datenbanken liegen vor allem im amerikanischen *artstor*<sup>4</sup> vor und in der Datenbank von *Foto Marburg*<sup>5</sup>, die auch eine historische Relevanz besitzt, da sie letztlich auf Fotokampagnen zurückgreift, die bis ins frühe 20. Jahrhundert zurückgehen und daher teilweise Zustände umfasst, die noch nicht von den teilweise umfassenden Verlusten der beiden Weltkriege betroffen sind. Hinzu kommt *prometheus*<sup>6</sup>, wobei letzteres die Synthese von über hundert Einzelverzeichnissen darstellt, die über eine gemeinsame Schnittstelle durchsuchbar gemacht werden. Alle diese großen Sammeldatenbanken, die sich durchwegs nicht auf einzelne Museen, sondern auf ganze Cluster von Museen oder Architekturen etc. beziehen, beinhalten mehrere Millionen Abbildungen und weiten ihren Bestand kontinuierlich aus.

Prometheus dominiert im deutschsprachigen Bereich, auch deswegen, weil es im Vergleich zum amerikanischen Angebot sehr preiswert ist. Das Prinzip dieser Metadatenbank, die ja selber außer der Suchfunktionalität gar keine eigenen Daten liefert, ist interessant und den Eigenheiten des Internets in hohem Maße angemessen. Es baut auf etwas, das man „shared authorship“ nennen kann und was auf die in der Summe unschlagbare Macht der *crowd* setzt. Ohne Nachteile ist das allerdings auch nicht zu haben, wird doch die Qualität der von den einzelnen Lieferanten zur Verfügung gestellten Daten kaum überprüft, um möglichst schnell eine kritische Masse zu erreichen. Das nämlich haben die *prometheus*-Macher schon von Anfang an erkannt, übrigens so, wie all die Internet-Unternehmen auch, die mit großen Kosten in die schnelle Generierung von umfangreichen Nutzungszahlen investieren: Wenn man in den Markt eindringen will – was im Falle von *prometheus* vor allem heißt, von den universitären Institutionen genutzt zu werden – muss man sehr schnell eine v. a. für die Lehre brauchbare Zahl von Kunstwerkreproduktionen zur Verfügung stellen. Denn nur dann fühlen sich die Lehrenden bemüßigt und berechtigt, das Projektionssystem vom Diapositiv auf das Digitale umzustellen – was inzwischen weitgehend geschehen ist. Aber wie gesagt, umsonst ist das nicht zu haben, was in der teilweise mangelhaften Qualität der Abbildungen und auch in der Verdoppelung und Verdrei/vier/fünffachung von Reproduktionen nach bekannteren Kunstwerken

4 S. <https://www.artstor.org>, zuletzt aufgerufen am 16.06.2024.

5 S. <https://www.bildindex.de>, zuletzt aufgerufen am 16.06.2024.

6 S. <https://www.prometheus-bildarchiv.de>, zuletzt aufgerufen am 16.06.2024.

zu sehen ist. Die Macher von *prometheus* reagieren darauf mit der Bitte an die Benutzenden, die einzelnen Abbildungen mit Sternchen zu versehen, sodass die schlechteren ausgeschieden werden können, ein Verfahren, das ingenieuerweise wieder auf die „Weisheit der Vielen“ setzt (stilbildend: Surowiecki 2004). Noch eleganter ließe sich das mit einem Verfahren machen, das von Schneider (2019) beschrieben wurde und das auf dem SIFT-Verfahren beruht. Ebenso der Qualitätssteigerung dient die Tendenz von *prometheus*, zuletzt verstärkt die Datenbanken von einzelnen großen Museen einzubinden.

Den korrigierbaren Nachteilen steht ein ganz wesentlicher Vorteil gegenüber, der sich aus der Verteilung der Arbeit auf mehrere Schultern ergibt, die wiederum im Netz als Ganze überall nutzbar ist. Früher hatte jede Institution, die mit Bildern befasst war, ihr eigenes Archiv. Die Beschränktheit der dort vorhandenen Arbeitskräfte führte zu einem jeweils mehr oder weniger lückenhaften Bestand, der selbst in umfangreichen Sammlungen spürbar war. Wenn über das Netz verteilt hunderte, wenn nicht tausende Digitalisierende aktiv sind, führt das bei allen Doppelungen zu sehr viel größeren Daten/Bildermengen, die von jedem (angemeldeten) Internet-Nutzenden verwendet werden können. Die gemeinsame Nutzung der verschiedenen, auch technisch häufig divergierenden Systeme wird seit einigen Jahren aussichtsreich mit Softwarelösungen wie IIF (*International Image Interoperability Framework*)<sup>7</sup> realisiert.

Die Daten werden gewöhnlich in eine relationale Datenbank eingegeben, in der die Angaben in Tabellen organisiert sind, die aufeinander bezogen und über sql (*search query language*) abgefragt werden. Zuletzt treten Graphdatenbanken ins Rampenlicht, da sie als flexibler gelten. In ihnen werden *Knoten* genannte Datenentitäten (z. B. Objekte, aber auch Namen etc.) definiert, die über *Kanten* aufeinander bezogen und in rdf-Tripels (*resource description framework*) abgespeichert werden. Das System kann beliebige Komplexitätsstufen erreichen.

#### 4. Bildsuche

Einer der großen Vorteile der Digitalisierung im Bereich der Bilddatenbanken liegt darin, dass einzelne Objekte nicht nur an einer bestimmten Stelle zu verorten sind, sondern dass auf sie je nach Ausführlichkeit der Beschreibungsdaten sehr differenziert zugegriffen werden kann. In der traditionellen Diathek findet sich Michelangelo „Moses“ normalerweise unter dem Namen des Künstlers, was einem Interessenten an der Ikonographie des Moses wenig hilft, wenn er oder sie nicht weiß, dass auch Michelangelo sich dem Thema gewidmet hat. Sicher, man hätte jetzt noch die Möglichkeit, alternativ eine Kategorie „Ikonographie“ einzurichten und hier den

7 S. <https://iiif.io>, zuletzt aufgerufen am 16.06.2024.

„Moses“ einzuordnen, aber dann könnte wiederum jemand, der sich für Michelangelo interessiert und nicht weiß, dass dieser einen Moses geschaffen hat, nicht fündig werden. Der Ausweg: Man platziert zwei Reproduktionen des gleichen Werkes sowohl unter dem Künstler als auch unter der Ikonographie. Damit würde der Bildbestand allerdings gewaltig aufgebläht, und wenn jetzt auch noch eine Kategorie Entstehungsdatum hinzukäme (in diesem Fall ca. 1514) unter der alle Werke aus dieser Zeit eingeordnet würden, oder eine der aufbewahrenden Institution (hier die Kirche San Pietro in Vincoli in Rom), dann könnte man die Menge auch noch ins Drei- und Vierfache erweitern. Unter Kosten- und Raumgesichtspunkten wäre das wenig erstrebenswert.

Im Digitalen läuft es anders. Jedes Werk kann hier beliebig differenziert erschlossen werden, ohne dass es mehrfach abzuspeichern ist. Wenn es in einer Datenbank registriert ist, in der z. B. Künstler, Titel, Datierung, Technik und Aufbewahrungsort abgespeichert sind, dann kann nach jeder dieser Kategorien gesucht werden, auch in Kombination. Eine Suche nach der Datierung „1514“ würde alle in dieser Datenbank gespeicherten Werke aus dem Jahr 1514 ausgeben, kombiniert mit dem Künstler „Michelangelo“ alles das, was der Renaissance-Meister in diesem Jahr geschaffen hat. Ermöglicht werden auf diese Art z. B. auch statistisch orientierte Untersuchungen: Welches Museum/welche Kirche etc. (Aufbewahrungsort) besitzt die meisten Werke Michelangelos? In welcher Zeit hat Michelangelo die meisten Werke geschaffen? Welches sind Michelangelos präferierte Themen? Letzteres ist allerdings schon etwas weniger trivial: Werktitel sind meist uneindeutig und daher nicht durch schlichte Titelnennung zu eruieren. Es kann „Maria“ heißen, „Mutter Gottes“ oder auch „Mary“. Über ein Klassifizierungssystem wie *Iconclass*<sup>8</sup> kann diese Uneindeutigkeit aufgehoben und die verschiedenen Benennungen können aufeinander *gemappt*, also aufeinander abgebildet werden. Ein relativ leicht zu bedienendes *tool* zur Durchführung von anspruchsvollen Aufgaben der beschriebenen Art liegt in *Museum Analytics*<sup>9</sup> vor, das ausdrücklich für Interessenten programmiert wurde, die nicht aus dem informatischen Bereich kommen. Es bleibt abzuwarten, ob LLMs (*Large Language Models*), die zurzeit etwa in Form von ChatGPT für Furore sorgen, solche Untersuchungsmöglichkeiten noch weiter vereinfachen können.

Die hohe Zahl von Reproduktionen in diesen Bilddatenbanken in Verbindung mit ihrer inhaltlichen Erschließung kann methodisch gesehen zu einer Dekanonisierung führen. Waren unter analogen Bedingungen die Zugriffsmöglichkeiten beschränkt und hingen weitgehend von den Kenntnissen des Suchenden ab, so dass man beim Besuch der Diathek eigentlich immer schon vorher wissen musste, was man suchte, so ist das unter digitalen Bedingungen anders. Folgt man der medialen Logik der Digitalisierung und betrachtet sie nicht als die schlichte Verlängerung des Analoges, so ergeben sich vielfältigere Suchmöglichkeiten, so dass auch die Suchergebnisse diver-

8 S. <https://iconclass.org>, zuletzt aufgerufen am 16.06.2024.

9 S. <https://dhvlab.gwi.uni-muenchen.de/max>, zuletzt aufgerufen am 16.06.2024.

ser sind. Auch wenn sowohl auf Angebots-, v. a. aber auf Nachfrageseite die alte Regel gilt, dass neue Medien die alten nachahmen, so dass die methodische Transformation bislang noch überschaubar scheint, so ist doch abzusehen, dass die Tendenz zur Diversifizierung bei sich weiter ausdifferenzierenden Angeboten steigen wird.

## 5. Bildberechnungen

Rein praktisch gesehen bietet die Digitalisierung in der Verwaltung von und Suche nach Bildern schon ganz erhebliche Vorteile. Aber auch analytisch bringt die Bilddigitalisierung ganz neue Möglichkeiten. Es dürfte nicht übertrieben sein, hierin einen Paradigmenwechsel zu erkennen. Bilder nämlich konnte man bislang beschreiben, Formen und Inhalte in ihnen identifizieren, sie in eine historische oder stilistische Reihe einordnen. Direkt in sie eindringen konnte man aber nicht oder eben immer nur indirekt. Ein digitales Bild dagegen stellt sich nicht nur als ein Pixel-Raster dar, sondern prinzipiell kann jedes einzelne dieser Pixel adressiert, identifiziert und verändert werden. Waren die zuletzt beschriebenen Suchmöglichkeiten auf die Angaben beschränkt, die als Metadaten bezeichnet werden, also solche, die von einer menschlichen bzw. fachlichen Intelligenz hinzugefügt worden waren, so ergibt sich hier die vollkommen neue Möglichkeit, den Gegenstand selbst zu fokussieren, nicht die sekundären Angaben zu ihm, z. B. in einem Histogramm, mit dem etwa die Verteilung von Farb- und Helligkeitwerten in einem Bild diagrammatisch darzustellen sind. Eini- germaßen faszinierend sind die Schlussfolgerungen, die sich daraus ergeben.

Der Computer ist eine radikal formalistische Maschine. Er erkennt farbige Punkte an einem bestimmten Ort und nichts weiter, dies allerdings mit einer ungeheuren Geschwindigkeit und Genauigkeit. Insbesondere erkennt er keine Semantik. Das Bild eines Hundes ist für ihn schlicht eine Pixelwolke. In der Kunstgeschichte spricht man seit neuestem daher von einer Rückkehr des Formalismus, der durch die bildadressierenden Vorgehensweisen im Digitalen befördert würde. Insbesondere Heinrich Wölfflin, der mit seinen *Kunstgeschichtliche[n] Grundbegriffe[n]* vor mehr als hundert Jahren einen Welterfolg erzielte und Kunstwerke der Renaissance und des Barock rein auf Basis ihrer Formeigenheiten unterschied, erlebt in diesem Zusammenhang eine Renaissance (s. Wasielewski 2023). In den Medienwissenschaften wird eine Emanzipation des Bildes aus den Fängen menschengemachter Kategorisierungen vermutet, die schlussendlich auch zu einer Enthistorisierung führen würde, da historisches Denken in deren Augen letztlich in menschlicher Sprache fundiert sei (s. Pias 2003). Gerade in dieser Perspektive erlangt die Problematik eine geradezu eschatologische Bedeutung, die sich vielfach mit den Diskursen des Posthumanismus verbindet.

So wie in der *Optical Character Recognition* (OCR) von Texten, wo die als Pixelwolken vorliegenden Buchstaben in einem Training mit Sinn versehen werden, so kann

man diese Pixelwolken allerdings auch bei Hunden und allen anderen Gegenständen trainieren, was heute zumeist über neuronale Netze verläuft, die in einem *machine learning* Verfahren auf Semantik trainiert werden. Solche neuronalen Netze werden nach dem Vorbild von Neuronenverschaltungen im Gehirn konstruiert und sind in den Augen der Techno-Utopiker bislang vor allem deswegen noch weniger leistungsfähig als das Gehirn, weil letzteres immer noch entschieden mehr Verschaltungen der Nervenzellen ermöglicht als selbst die stärksten Computer. Das führt unter anderem dazu, dass das Kleinkind nur wenige Exemplare eines Hundes vorgeführt bekommen muss, um eine abstrakte Vorstellung vom Hund zu erlangen und auch Hunde als solche zu erkennen, die nicht exakt den vorgeführten Beispielen entsprechen, während der Rechner mit mehreren Hundert, wenn nicht Tausend dieser Exemplare gefüttert werden muss, um daraus ein allgemeines Modell abzuleiten.

Für die Suche nach Bildern heißt dies, dass sie zukünftig verstärkt auf Basis dieser Bilder selbst stattfindet, anstatt auf derjenigen von Metadaten. Rein arbeitstechnisch gesehen heißt das, dass auch größere Bildmengen zu durchsuchen sind, also solche, die selbst über Millionen Bilder beinhaltende Kunstdatenbanken hinausgehen. Das klingt nach einem rein theoretischen Problem, aber wenn man bedenkt, dass alleine bei Facebook täglich (!) 350 Millionen Fotos hochgeladen werden, zeigt sich schnell die praktische Seite. Allerdings hat auch die Kunstgeschichte in der direkten Bildadressierung noch immer einiges an Forschungsleistung zu erbringen, wenn sie ein funktionsfähiges Suchsystem mit einigem Anspruch vorlegen will. Zwar ist es schon jetzt möglich, einen Menschen auf einem Bild zu erkennen, und auch ein Mensch, der einem anderen Wasser über den Kopf gießt, stellt für avancierte KI-gestützte Programme kein unüberwindliches Problem mehr da. Aber Johannes den Täufer zu erkennen, der Jesus Christus im Jordan tauft, das ist mit den vorhandenen Ansätzen noch kaum realisierbar. Hierfür wird es notwendig sein, die Trainingsdatenmenge noch einmal massiv zu erhöhen und die 400 Millionen Bild-Text-Kombinationen, mit denen solche Modelle wie CLIP der inzwischen berühmt gewordenen Open AI Initiative trainiert wurden, vor allem mit spezielleren Kombinationen zu ergänzen. Im Falle einzelner Wissenschaften ließe sich das in erster Linie dadurch erreichen, dass in großem Umfang Expertenwissen herangezogen wird, das in Form maschinenlesbarer Forschungstexte greifbar ist. An dieser Stelle zeigt sich auch, wie wichtig es im Hinblick auf derartige Lösungsansätze ist, wissenschaftliche Texte möglichst umfassend im *open access* zu veröffentlichen.

## 6. Bildfälschungen

Die Künstliche Intelligenz kann aber nicht nur zur direkten Bildadressierung verwendet werden, sondern sie dient auch zur Herstellung von Bildern, und zwar zu solchen, die keinen abbildenden Bezug auf die Wirklichkeit mehr haben. Das gab

es zwar auch schon bisher, aber neu ist, dass der abbildende Bezug zwar suggeriert wird, aber reine Phantasie ist. Man spricht hier gerne von fake-Bildern. KI-gestützte Systeme wie *DALL-E*, *Midjourney* oder *stable diffusion* erlauben neben seriösen Anwendungen, in denen der nicht-realistische Fokus gar nicht anzuzweifeln ist, solche Verwendungsweisen, deren v. a. auch gesellschaftspolitischen Auswirkungen noch gar nicht abzusehen sind. Wenn zuletzt ein Bild des vor der Polizei wegrennenden Papstes im Internet viral ging, dann sollte jeder einigermaßen Informierte das Unechte dieses Bildes erkannt haben, aber ein Bild des gleichen Papstes, der sich über eine Mohammed-Karikatur amüsiert, dürfte in der islamischen Welt durchaus für Unmut sorgen, auch wenn der Papst in Wahrheit nie solche Karikaturen zu Gesicht bekommen hat.

## Verwendete Literatur

- Besser, H. (2003). *Introduction to Imaging. Revised Edition*. Hg. v. S. Hubbard & D. Lenert. Los Angeles: Getty Research Institute. URL: <https://www.getty.edu/publications/virtuallibrary/0892367334.html> [zuletzt aufgerufen am 16.06.2024].
- Borek, L., Zeppezauer-Wachauer, K., & Ketschik, N. (2022). Eindeutig Uneindeutig. Zur Modellierung von Unschärfe in der Mediävistik. In *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte* [Blog]. <https://doi.org/10.58079/rha5> [zuletzt aufgerufen am 16.06.2024].
- Dilly, H. (1979). *Kunstgeschichte als Institution. Studien zur Geschichte einer Disziplin*. Frankfurt a. M.: Suhrkamp.
- Pias, C. (2003). Das digitale Bild gibt es nicht. Über das (Nicht-)Wissen der Bilder und die informatische Illusion, *zeitenblicke*, 2(1), o. S. URL: <https://www.zeitenblicke.de/2003/01/pias/pias.pdf> [zuletzt aufgerufen am 16.06.2024].
- Schneider, S. (2019). Über die Ungleichheit im Gleichen. Erkennung unterschiedlicher Reproduktionen desselben Objekts in kunsthistorischen Bildbeständen. In *DHd 2019. Digital Humanities multimedial und multimodal*. Frankfurt a. M./Mainz: Zenodo. <https://doi.org/10.5281/zenodo.4622051> [zuletzt aufgerufen am 16.06.2024].
- Surowiecki, J. (2004). *The Wisdom of Crowds. Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economics, Society and Nations*. New York: Doubleday.
- Vaughan, W. (1987). The Automated Connoisseur. Image Analysis and Art History. In P. Denley, & D. Hopkin (Hrsg.), *History and Computing* (S. 215–221). Manchester: Manchester University Press.
- Wasielewski, A. (2023). *Computational Formalism. Art History and Machine Learning*. Cambridge, Mass.: MIT Press.

Wölfflin, H. (1915). *Kunstgeschichtliche Grundbegriffe. Das Problem der Stilentwicklung in der neueren Kunst*. München: Bruckmann. <https://doi.org/10.11588/digit.27250> [zuletzt aufgerufen am 16.06.2024].

## Weiterführende Literatur

- Altenhöner, R., Berger, A., Bracht, Ch., Klimpel, P., Meyer, S., Neuburger, A., Stäcker, Th., & Stein, R. (2023). DFG-Praxisregeln „Digitalisierung“. Aktualisierte Fassung 2022. Online: *Zenodo*. <https://doi.org/10.5281/zenodo.7435724> [zuletzt aufgerufen am 16.06.2024].
- Brown, K. (2020). *The Routledge Companion to Digital Humanities and Art History*. New York/London: Routledge.
- Cohen, D., & Rosenzweig, R. (2006). *Digital History. A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press.
- Jannidis, F., Kohle, H., & Rehbein, M. (Hrsg.). (2017). *Digital Humanities. Eine Einführung*. Stuttgart: J. B. Metzler Verlag.
- Kuroczynski, P., Bell, P., Dieckmann, L., Hoppe, S., & Münster, S. (Hgg.). (2018), *Computing Art Reader. Einführung in die digitale Kunstgeschichte*. Heidelberg: arthistoricum.net [= *Computing in art and architecture*, 1].