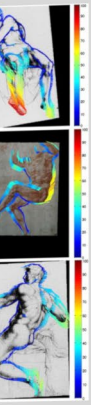
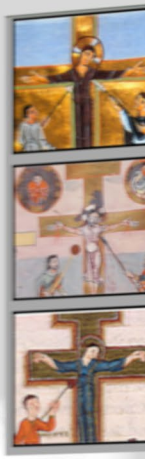
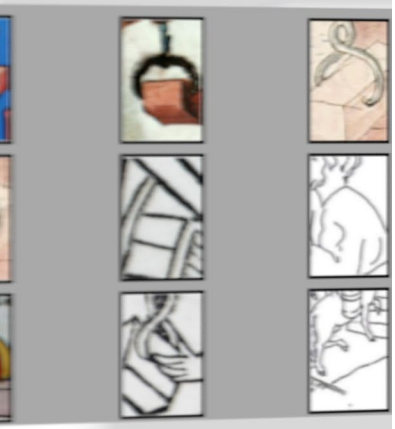

A
B2
B
C
B1
B3







Peter Bell, **Björn Ommer**

C. Computer Vision und Kunstgeschichte – Dialog zweier Bildwissenschaften

→ Computer Vision, automatisches Sehen, Bildverarbeitung, Image Processing, Erschließung, Annotation, Bildverstehen, Bildwissenschaft, Machine Learning

Im Rahmen von Digitalisierung und künstlicher Intelligenz entsteht auch ein maschinelles Sehen. Das darum entstandene Forschungsfeld Computer Vision ist auch eine Bildwissenschaft, mit der die Kunstgeschichte unmittelbar in Dialog treten kann und mit deren Unterstützung sie die anwachsenden Bilddatenbestände schneller und tiefer erschließen kann. In diesem Kapitel werden einige Anwendungsbeispiele vorgestellt, um die Potenziale und Herausforderungen verschiedener Methoden der Computer Vision vorzustellen. Die Einsatzbereiche sind vielfältig und reichen von einfacher Duplikatsuche zur Detektion von Objekten, Bildvergleichen und Stilanalyse. Besonders interessant erscheinen hier Ansätze, in denen Mensch und Maschine interagieren und dabei ihre unterschiedlich gelagerten Kompetenzen im Erfassen und Verarbeiten von Informationen verbinden.

C.1 Ansätze der Zusammenarbeit

Computer Vision ist ein Teilbereich der Informatik, der das visuelle Wahrnehmungsvermögen von Maschinen entwickelt und erforscht. Dieses **Sehen** gehört in den Kontext der künstlichen Intelligenz und schon deshalb begnügt sich die Forschung nicht mit dem Erkennen einfacher Muster, sondern arbeitet daran, menschliche Wahrnehmung zu simulieren und komplementäre Aufgaben zu erfüllen.

Computer Vision und Kunstgeschichte sind zwei Bildwissenschaften, die sich ähnliche Fragen zu Semantik, Interpretation und Phänomenologie stellen. Bislang widmen sich große Teile der Computer Vision noch alltäglichen Fotografien oder Videos, die sie weitgehend kritiklos als objektive Repräsentationen einer äußeren Wirklichkeit betrachten. Gegenüber diesen Bildsammlungen stellt die Erweiterung der analysierten Bildkorpora um Bilder des kulturellen Erbes eine inhaltliche und methodische Bereicherung dar. So stellt z.B gegenständliche Malerei einen zur Fotografie alternativen Zugang zur äußeren Wirklichkeit dar.

Wenn Kunstwerke Gegenstand der Untersuchung werden, liegt es nahe, dass Computer Vision und Kunstgeschichte in einen Dialog treten, um Sehaufgaben zu lösen, die Erkenntnisgewinne für beide Fächer erzeugen. 01 Von kunsthistorischer Seite muss erarbeitet werden, wie KünstlerInnen ihre Inhalte sichtbar machen, und die Informatik muss mit abgestimmten Algorithmen auf die jeweiligen Repräsentationsformen eingehen. Die Computer Vision erhält so skalierbare Problemstellungen zur Verbesserung des automatischen Sehens, während die Kunstgeschichte bei Erschließungs-, Such- und Analyseaufgaben unterstützt wird. Beide Disziplinen reflektieren dabei ihre Methoden und entwickeln gemeinsame methodische Ansätze.

■ **01**
Die Literaturangaben beziehen sich meist auf diese Schnittmenge interdisziplinär erarbeiteter Ergebnisse. Entsprechend wird auf die Referenzierung disziplinärer Standardwerke verzichtet.

Im Folgenden werden Etappen und Ergebnisse dieses Dialogs vornehmlich aus der kunsthistorischen Perspektive 02 referiert, während die informatische Sicht, deren methodische Ansätze sowie mathematischen und technischen Lösungen nicht näher berücksichtigt werden können. Die Antworten der Informatik werden hier somit nur in Form von prototypischen Anwendungen vorgestellt, mit denen der Forschungsstand und die Potenziale der Zusammenarbeit aufgezeigt werden können. Ein tieferer Einstieg in den Dialog aus Sicht der Computer Vision ist jedoch anhand der zitierten informatischen Publikationen jederzeit möglich.

■ **02**
Teile des Textes erschienen mit Fokus auf den musealen Bereich unter: Peter Bell, Björn Ommer, Visuelle Erschließung. Computer Vision als Arbeits- und Vermittlungstool. In: Andreas Bienert (Hg.), EVA Berlin 2016, Berlin 2016, S. 67–73, hier wird eher aus der Sicht universitärer Forschung argumentiert.

Im Rahmen dieser Einführung sollen vier grundlegende Fragen an Beispielen erörtert werden:

[1] Auf welche Bildbestände lässt sich Computer Vision anwenden?
[2] Wie stellt sich die Interaktion von Mensch und Maschine dar?
[3] Welche Fragen können beantwortet werden?
[4] Welche Fragen bleiben offen bzw. wo liegen die gegenwärtigen Grenzen?

Grundsätzlich sind die Bildbestände für die Computer Vision nicht begrenzt, wichtiger ist zu definieren, was gesucht oder analysiert werden soll, und zu erheben, in welcher Menge, Form und Varianz dieses Phänomen im Bilddatensatz 03 vorkommt. Entsprechend sind Datensätze dankbar, die standardisierte Formen und Bildchiffren verwenden oder andere wiederkehrende Motive enthalten wie etwa die der Buchmalerei. Eine Interaktion von Mensch

■ **03**
Bilddatensatz meint für die vorgestellten Verfahren nur eine Sammlung von Bilddateien. Auf Metadaten wie Bildtitel, Schlagwörter oder Aufnahmeort wird nicht zugegriffen.

und Maschine ist auf sehr vielfältige Weise möglich. Besonders interessant sind Ansätze, die auf maschinellem Lernen basieren.

Gerade für Forschende aus den Geisteswissenschaften gleicht die Interaktion mit den Bildverarbeitungsprogrammen und teilweise mit dem Fach Informatik einem Prinzipal-Agenten-Modell 04: Sie müssen Aufgaben an die Algorithmen vergeben, wissen aber unter Umständen nicht, welche Qualitäten diese jeweils aufweisen und können daher auch die Arbeitsvorgänge nicht überprüfen. Dadurch können Algorithmen zum Einsatz kommen, die sich für die jeweilige Aufgabe wenig eignen oder Fragen beantworten, die aus Sicht der Kunstgeschichte nicht relevant erscheinen. Entsprechend muss die Mensch-Maschine-Interaktion so gestaltet sein, dass der Mensch über die jeweiligen Charakteristika der Algorithmen möglichst gut informiert ist und die internen Vorgänge anhand der Anwendung nachvollziehen kann, auch ohne die technischen und mathematischen Abläufe zu verstehen. Dazu empfiehlt sich eine wechselseitige Einbeziehung in die Arbeitsabläufe von Geisteswissenschaften und Informatik bei der Entwicklung neuer Methoden und Lösungen; idealerweise in gemeinsamen Arbeitsgruppen.

Die zu beantwortenden Fragen beziehen sich auf ganz verschiedene Dimensionen von Ähnlichkeit, wie etwa auf die Vergleichbarkeit von Objekten, Stil, Komposition, Technik. In manchen Fällen wie bei Duplikatsuchen oder dem Auffinden von sehr prägnanten Objekten kann der Computer auch ohne großen Lernaufwand konkrete Antworten liefern. Bei komplexeren Fragestellungen wie jenen nach variierenden Motiven und Ikonografien kann der Computer oft nur Vorschläge zur Beantwortung beisteuern oder aber durch einen längeren iterativen Lernprozess so weit trainiert werden, dass sicherere Aussagen möglich sind.

Technische und phänomenologische Schwierigkeiten entstehen dort, wo abstraktere Zusammenhänge erkannt oder Objekte über ihre stilistisch stark veränderten Repräsentationen hinaus identifiziert werden sollen. Die menschliche Seherfahrung, von Kindheit im Erkennen künstlerischer und symbolischer Abstraktionen geschult, hat hier einen noch schwer einholbaren Vorsprung. Die Forschung zu künstlicher Intelligenz entwickelt sich jedoch auch hier weiter, sie verfolgt zunehmend anstelle eines statischen, regelbasierten Ansatzes **lernfähige**, auf assoziativen Hypothesen aufbauende Erkennungsmethoden. In diesem Prozess überschneiden sich auch theoretische Ansätze von Computer Vision und Kunstgeschichte wie etwa die für beide Felder interessante Gestalttheorie und andere phänomenologische Ansätze. 05 In der Kunstgeschichte wurde schon früh, aufgrund des Stands der Technik vielleicht zu früh, mit den Möglichkeiten von Mustererkennung und Computer Vision experimentiert. 06 Damit ist auch früh die besondere Schwierigkeit erkannt worden, mithilfe automatischer Bilderkennung künstlerische Objekten miteinander zu vergleichen. 07 Mittlerweile ist die Computer Vision ein stark beforschter Teilbereich der Informatik, in dem vermehrt Publikationen zu Algorithmen und Anwendungen für Kunstwerke erscheinen. Zu ihnen gehören auch unsere Arbeiten.

■ 04
Vergleichbar mit der Principal-Agent-Theory ist die Informationsasymmetrie zwischen dem eine Leistung erwartenden Geisteswissenschaftler auf der einen Seite und der Informatik auf der anderen Seite. Allerdings ergeben sich in der interdisziplinären Zusammenarbeit Informationsdefizite auf beiden Seiten. Hier führen nur große und kontinuierliche Übersetzungsleistungen zum Erfolg.

■ 05
Vgl. zur Anwendung der Gestalttheorie in der Computer Vision u. a. Björn Ommer, Learning the Compositional Nature of Objects for Visual Recognition, Diss. ETH, No. 17449, Zürich 2007 (http://e-collection.library.ethz.ch/eserv/eth:29976/eth-29976-02.pdf) S. 36–41; als in Teilen gegenläufiger Ansatz sind die kunstphilosophischen Arbeiten von Merleau-Ponty zu diskutieren.

■ 06
Zu nennen wäre der frühe kennerschaftliche Ansatz von Vaughan 1997 oder die vom Kunsthistorischen Institut in Florenz – Max-Planck-Institut und dem Istituto di Scienza e Tecnologia dell'Informazione (ISTI) in Pisa angebotene Datenbank STEMMARIO; William Vaughan, Computergestützte Bildrecherche und Bildanalyse, in: Hubertus Kohle (Hg.), Kunstgeschichte digital. Eine Einführung für Praktiker und Studierende, Berlin 1997, S. 97–105.

■ 07
Felix Thürlemann, Christus eingegeben und Hitler gefunden beim Ikonogoogle, in: Frankfurter Allgemeine Zeitung vom 14.9.2011.

Ohne Anspruch auf Vollständigkeit lassen sich fünf Herangehensweisen unterscheiden:

[a] Ein Vergleich von ganzen Bildern, z. B. um Duplikate, Kopien und Nachfolger aufzufinden oder auch um detektierte Bilder mit Informationen zu verknüpfen. 08
[b] Ein Vergleich von einzelnen Szenen, Objekten oder Detailformen auf der Ebene der Semantik. 09
[c] Ein Abgleich der Unterschiede durch genaue Analyse/Errechnung der Abweichungen. 10
[d] Ein Vergleich von technischen Merkmalen (Pinselstrichen, Schraffuren), Texturen und Farben (Low Level Vision). 11
[e] Erschließung einer großen Menge an Bilddaten zum Auffinden von strukturellen Ähnlichkeiten.

Für jede Computer-Vision-Anwendung muss vorweg entschieden werden, welchen Anteil das maschinelle Lernen haben soll. Wenn ein visuelles Phänomen 12 für einen Datensatz eine besondere Bedeutung hat, kann dies als eine Kategorie definiert und diese mit deren Repräsentanten exemplarisch angelernt werden, um bessere Ergebnisse zu erhalten. 13 Ein derartig intensives Training wendeten wir in unseren Forschungsprojekten – und damit kommen wir zu konkreten Anwendungsbeispielen – auf drei Kategorien an: Herrscherkronen, Gesten und Kapitelle und drei Datensätze. Zunächst wurden Kronen angelernt, die in den deutschen Palatina-Handschriften 14 vorkommen, dann verschiedene Gesten des Sachsenspiegels 15 und schließlich anhand eines gemischten Datensatzes Kapitelle der klassischen Säulenordnungen.

Alle diese Kategorien haben eine gewisse Prägnanz gemeinsam, sowohl in ihrer Semantik wie auch in ihren unterschiedlichen visuellen Ausformungen. Bei Kronen ist von einer gewissen materiellen und kunsthandwerklichen Wertigkeit auszugehen und sie werden auf den Köpfen von Potentaten erwartet, Gesten werden mit kommunizierenden lebenden Menschen in Verbindung gebracht, Kapitelle schließlich sind als Abschluss einer Säule und unterhalb eines Architravs zu vermuten. Das menschliche Sehen kommt durch diese kontextbezogenen Hypothesen sehr schnell zu intuitiven Schlussfolgerungen, besitzt aber andererseits zugleich die Abstraktionsfähigkeit, auch zu Boden gefallene Kronen und Kapitelle sowie die sehr seltene Ausnahme gestikulierender Toter 16 zu identifizieren. Das maschinelle Sehen hatte bisher sowohl in den kontextbezogenen Hypothesen als auch in der Abstraktionsfähigkeit Schwierigkeiten, die nun teilweise durch neuronale Netze überwunden wurde.

Bei den locker gezeichneten Kronen der deutschen Palatina-Handschriften kann der Computer nicht nur die verschiedenen Kronentypen erkennen, sondern auch der Duktus des Zeichners, sodass hier anhand der maschinell ermittelten Ähnlichkeitsverhältnisse die Werkstattzusammenhänge ansatzweise rekonstruiert werden können 01.

Das Training von Kategorien, beispielsweise den Säulenordnungen, ist – wie bei jeder Kategorienbildung – nur dann sinnvoll, wenn dadurch eine Akzeptanz der Varianz innerhalb der Kategorie entsteht und die Abgrenzung zu den anderen Kategorien erkennbar wird. Auch wenn die Säulenordnungen diese

■ 08
John Resig, Using Computer Vision to Increase the Research Potential of Photo Archives, in: Journal of Digital Humanities, Vol. 3, No. 2 Summer 2014. Resig nutzt den Algorithmus von https://tineye.com/, dort und in der Google Bildersuche lässt sich Funktionsweise und Performanz leicht abschätzen.

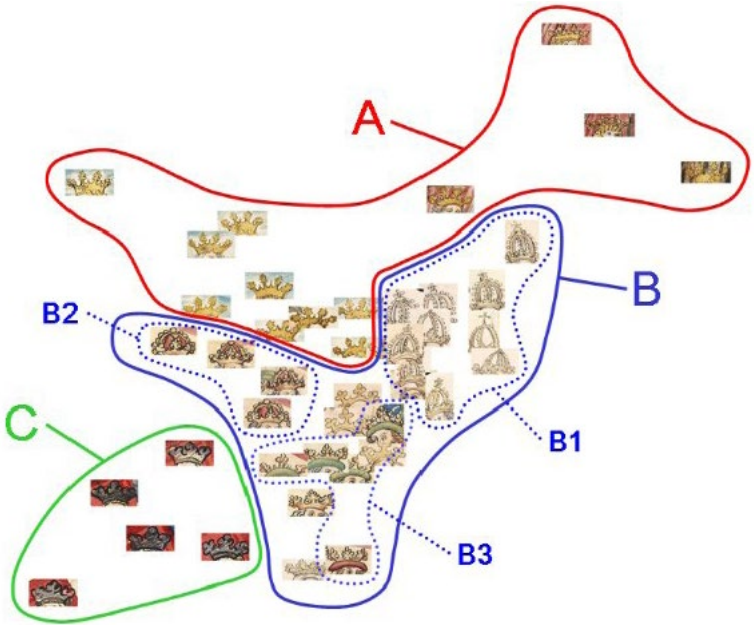
■ 09
Visual Geometry Group, University of Oxford, Web Demo: http://zeus.robots.ox.ac.uk/ballads/; Masato Takami, Peter Bell, Björn Ommer, Offline Learning of Prototypical Negatives for Efficient Online Exemplar SVM, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, IEEE, 2014, S. 377–384.

■ 10
Antonio Monroy, Peter Bell, Björn Ommer, Morphological analysis for investigating artistic images, in: Image and Vision Computing 32(6), 2014, S. 414–423.

■ 11
Richard N. Johnson et al., Image Processing for Artist Identification – Computerized Analysis of Vincent van Gogh's Painting Brushstrokes, in: IEEE Signal Processing Magazine Vol. 15, 2008, S. 37–48, doi: 10.1109/MSP.2008.923513.

■ 12
Hier sind zunächst ganz konkrete Phänomene wie klar definierte Objekte bzw. Realien gemeint. Schwieriger, aber gelegentlich interessanter ist das Antrainieren von abstrakteren Zusammenhängen.

■ 13
Peter Bell, Joseph Schlecht, Björn Ommer, Nonverbal Communication in Medieval Illustrations Revisited by Computer Vision and Art History, in: Visual Resources: An International Journal of Documentation, 29 (1–2) 2013, S. 26–37.



☐ 01
Kronendarstellungen aus den Werkstätten Henfflin, Lauber und Alsatian (Universitätsbibliothek Heidelberg/Computer Vision Group Heidelberg).

Differenzbedingungen weitgehend erfüllen, hat der Algorithmus trotz Training die gleichen Schwierigkeiten wie Studierende der Kunstgeschichte, korinthische von kompositen und dorische von toskanischen Kapitellen zu unterscheiden, da diese Ordnungen jeweils viele Charakteristika teilen und sich nur in Details unterscheiden. 17 Darüber hinaus fehlt einem lediglich auf bildliche Ähnlichkeit ausgerichteten Algorithmus auch eine räumliche Skala, sodass er kleinere oder größere Bauteile mit Elementen der Kapitelle verwechselt, beispielsweise die rudimentären Voluten des korinthischen Kapitells mit den viel klareren und größeren ionischen.

Insgesamt ist das Kapitell wie das Gesicht eine sehr prägnante Form, die kaum mit anderen Formen verwechselt wird. Mit wenigen Strichen gezeichnete Gesten oder auch andere Architekturmerkmale sind von ihrer Umwelt weniger abgrenzbar. Parallel zu Homonymen und Polysemen in der Sprache, also gleichlautenden Worten, die für verschiedene Begriffe stehen (z. B. **Bank** oder **Krone**), gibt es auch in der Architektur und Kunst gleiche Formen, die unterschiedlich begriffen werden müssen. Wer frontal auf einen Quader blickt, kann ihn auch als Rechteck identifizieren, und weit von oben gesehen wird der Wald einer bemoosten Fläche ähnlich. Diese Täuschungen löst der Mensch, indem er seinen Ort bzw. den der Kamera abzuschätzen versucht oder seine Position verändert und indem er den Kontext möglichst genau ergründet. Seine Fokussierungen und Standpunktwechsel, die Verwendung seiner eigenen Proportionen als Maßstab für das Gesehene sind schwer von der Maschine rekonstruierbar. Die Frage ist nur, wie schwer künstlerische Darstellungen es dem Betrachter machen.

Grundsätzlich ergeben sich in der Kunst verschiedene Dimensionen, durch die visuelle Mehrdeutigkeiten entstehen oder auch eingegrenzt werden können. Trotz hermetischer Tendenzen in einigen Avantgarde-Bewegungen der Moderne und vereinzelter ausschließender Bildstrategien zur Privilegierung exklusiver Rezipientengruppen in allen Epochen (z. B. Emblematik) sind Bildproduzenten meist daran interessiert, dass ihre Bilder mindestens auf einer vorikonografischen Ebene, meist jedoch auch ikonografisch, allgemein verständlich sind. Entsprechend präsentieren sie Akteure und Objekte des Bildes vorwiegend auf

■ 14
Pradeep Yarlagadda, Antonio Monroy, Bernd Carque, Björn Ommer, Top-down Analysis of Low-level Object Relatedness Leading to Semantic Understanding of Medieval Image Collections, in: Computer Vision and Image Analysis of Art II, Proc. of SPIE Vol. 7869 2011, S. 61–69.

■ 15
Joseph Schlecht, Bernd Carque, and Björn Ommer, Detecting Gestures in Medieval Images, in: IEEE International Conference on Image Processing, Brussels 2011, S. 1309–1311.

■ 16
In den Illuminationen der Sachsenspiegel-Handschriften werden auch Tote in Rechtsfälle wie z. B. Erbangelegenheiten einbezogen und drücken ihre Standpunkte in Gesten aus.

■ 17
Die Säulenordnungen sind zu einem wiederkehrenden Thema der Digital Humanities geworden. Vgl. Susanne Schumacher, Ordnungen schaffen? Daten in der Kunstgeschichte – am Beispiel von Säulenordnungen, Doktorarbeit ETH Zürich, 2015, DOI: 10.3929/ethz-a-010538703.

eine gute Identifizierbarkeit hin. Die mittelalterliche Bedeutungsperspektive ist neben ihrem stark hierarchisierenden Aspekt genau durch diese visuelle Darreichung bestimmt. Personen und Realien sind nicht in einer geschlossenen, für sie stringenten Wirklichkeit wiedergegeben, sondern werden auf den Betrachter hin ausgerichtet, der dann die verschiedenen, gut erkennbaren Einzelteile erst zu einer Wirklichkeit zusammensetzt. Die Einführung der Zentralperspektive macht somit die Bilder zwar veristischer, aber deren einzelne Elemente werden dadurch nicht unbedingt einfacher lesbar. Schon an dieser Stelle wird klar, wie unterschiedlich die Sehangebote der Kunst sind und wie die Repräsentationsform der Inhalte die Rezeption für den Betrachter erleichtern oder auch erschweren kann. Entsprechend wird hier auch klar, warum die Beschäftigung mit Kunstgeschichte für Computer Vision auch auf theoretischer und methodischer Ebene interessant sein kann. Kunstwerke, die ihre Gegenstände auf das für die Identifizierung Wesentliche zuspitzen, geben viel über das menschliche Sehen preis. Es ist eine offene Frage, ob das automatische Sehen durch die künstlerischen Sehvorgaben Wirklichkeit besser erfassen könnte. Für den Zusammenhang der computergestützten Bildsuche ist hingegen zunächst nur wichtig, dass die Charakteristika der Gegenstände, die der Künstler zu deren Verständlichkeit einsetzt, oft zu festen Bildformeln werden, nach denen gesucht werden kann.

Daneben können bereits von der Computer Vision trainierte Kategorien von Alltagsgegenständen oder insbesondere Tiere auf Kunstwerke übertragen werden. [18] Soll hingegen nach ganz unterschiedlichen Objekten oder Szenen gesucht werden, sollte der Algorithmus zuvor lediglich eine Art **allgemeinen Eindruck** vom Datensatz erhalten; eine Abstraktion von Mustern und Charakteristika, die im Datensatz vorkommen, jedoch nicht mit der konkreten Objektebene übereinstimmen.

Das Städel verwendet eine einfache Duplikatsuche [a], damit Museumsbesucher über eine App mehr Informationen über Kunstwerke bekommen, die sie mit dem Smartphone fotografieren. [19] Das Foto wird mit einem kleinen Datensatz an Sammlungsbildern verglichen und kann so schnell erkannt werden. Da die Identität zwischen abgelegtem und neu erstelltem Bild groß ist, kommen Algorithmen dieser Art trotz kleiner Variationen wie eines leicht veränderten Winkels, anderer Beleuchtung oder partieller Verdeckung zum richtigen Ergebnis. Es ist technisch relativ leicht möglich, die ganze Sammlung mit einem solchen Algorithmus auszustatten, zumindest wenn sich nicht sehr ähnliche Objekte darin befinden, die zu Verwechslungen führen können. Eine Kombination mit einer Bildsuche von Typ [b] wäre sinnvoll, wenn die Sammlung Objekte enthielte, die sich durch Publikumsverkehr oder Größe nicht ganz ins Bild bringen lassen. Denn in vielen Fällen kann auch von einem Detail auf das ganze Objekt geschlossen werden.

Eine partielle Bildsuche bietet aber viel weitreichendere Möglichkeiten. Indem der Fokus auf Teile des Bildes gerichtet werden kann, lassen sich gezielt Informationen und Vergleichsabbildungen präsentieren, die sich nur auf eine Partie beziehen. In einem weiteren Schritt können die gefundenen Bereiche genauer verglichen und die Abweichungen markiert werden [c]. Der Computer ist dabei in der Lage, die verschiedenen Transformationen zu definieren, die zwischen einem Bild und einem Abbild bestehen, sodass nicht nur die

■ **18**
Elliot Crowley, Andrew Zisserman, The Art of Detection, Workshop on Computer Vision for Art Analysis, in G. Hua, H. Jégou (eds), Computer Vision – ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part I. Lecture Notes in Computer Science, vol. 9913, Springer, Cham 2016, S. 721–737, doi: 10.1007/978-3-319-46604-0_50.

■ **19**
http://www.staedelmuseum.de/de/angebote/staedel-app.

Unterschiede sehr verständlich visualisiert werden können, sondern die Abweichungen durch diese Analysen teilweise begründet werden können.

C.2 Prototyp einer freien Bildsuche

Für die Bildsuche [b] nach Objekten, Szenen und Motiven sind Bildsammlungen hilfreich, in denen unter einem oder mehreren Gesichtspunkten eine Kohärenz besteht. Dies können ein gemeinsamer Stil, eine ähnliche Bildsprache oder technische und motivische Übereinstimmungen sein. So gestaltet sich z. B. die Bildsuche in einer illuminierten mittelalterlichen Handschrift oder illustrierten Inkunabel, die einen klaren Figurenmaßstab und eine geradezu normierte Bildsprache aufweist oder die gleichen Motive mehrfach verwendet, sehr viel einfacher als in einem ganz heterogenen Bestand wie zum Beispiel der digitalisierten Diathek eines breit aufgestellten kunsthistorischen Instituts, das Kunst und Architektur aller Epochen und aus einem großen geografischen Raum enthält. Gleichzeitig sind es gerade diese Sammlungen, aus denen überraschende Korrespondenzen zu erhoffen sind, da der Computer ohne Rücksicht auf Kontexte und bekannte Verbindungslinien Bild für Bild vergleicht.

Als Vorbereitung auf diese größere Komplexität und unter Berücksichtigung der Ausrichtung und Leistungsfähigkeit der zugrunde liegenden Algorithmen wurde die freie Bildsuche auf mittelalterliche Bildhandschriften und druckgrafische Porträtsammlungen sowie Architekturdarstellungen mit markanten Bauteilen (z. B. Kapitellen) angewendet. Der Suchprozess ist individuell, mehrstufig und iterativ, das heißt, der Nutzer kann ein oder mehrere Bereiche im Bild markieren, nach denen gesucht werden soll. Mehrere Bereiche zu markieren empfiehlt sich dann, wenn miteinander verknüpfte Objekte oder Personen gesucht werden sollen oder auch ein Objekt, das sich durch signifikante Partien besonders auszeichnet (z. B. Kopf und Hufen eines Pferdes: 02. Dies funktioniert in einigen Fällen auch bei Ikonografien wie Maria und Johannes unter dem Kreuz, indem nur diese beiden Protagonisten und ggf. Details vom Kreuz markiert werden.

☐ 02
Suchbox im ersten Bild oben links und dann detektierte Bilder in absteigender Ähnlichkeit (Auswahl aus Marburger Porträtindex/Computer Vision Group Heidelberg).

Iterativ und mehrstufig wird der Suchprozess dadurch, dass der Nutzer nach einem ersten Durchgang die Ergebnisse bewerten kann. Dadurch werden im nächsten Durchlauf nicht nur Ergebnisse unterdrückt, deren visuelle Ähnlichkeit unbedeutend ist, und damit die Ergebnisse verbessert, sondern auch die Suchaufgabe wird genauer definiert. Denn der Nutzer bestimmt mit den Ergebnissen in jedem weiteren Schritt, ob er sehr fokussiert suchen oder ob er Varianzen ausdrücklich zulassen möchte. Gerade hier ist eine visuelle Skalierung der Suchanfrage möglich, die textlich kaum zu definieren ist. Es ist damit beispielsweise möglich, eher allgemein nach liegenden Figuren oder aber nach einer in gleicher Pose liegenden alten Frau zu suchen oder – ein weiteres Beispiel – entweder nach Amphitheatern allgemein oder nur nach einer spezifischen mittelalterlichen Darstellungsweise des Kolosseums.

Die Mehrstufigkeit des Verfahrens ermöglicht es, auch heterogene Bilddatensätze anzugehen, wie z. B. alle 3620 mit dem Schlagwort **Kreuzigung** versehenen Abbildungen im prometheus-Bildarchiv. Trotz der thematischen Engführung durch das Schlagwort ist der Datensatz vielgestaltig. Schon die Darstellungen des Hauptmotivs, der Kreuzigung Christi, sind in vielen Dimensionen variiert. Darunter fallen nahsichtige Kruzifixe und vielfigurige Kalvarienberge sowie sämtliche gängigen künstlerischen Techniken und Stile von der Spätantike bis in die Gegenwart. Daneben sind auch Bilder verschlagwortet, die Kreuzigungen anderer Figuren – bis hin zu Martin Kippenbergers Frosch am Kreuz **(Zuerst die Füße)** – zeigen oder eine Kreuzigung nur als Bild im Bild enthalten. Schließlich finden sich darunter auch Detailbilder aus Kreuzigungsdarstellungen, in denen die eigentliche Kreuzigung gar nicht zu sehen ist. Die visuelle Bildsuche kann nun die noch sehr offenen Ergebnisse der Textsuche auf Anfrage nach Ähnlichkeiten gruppieren.

So lässt sich erfolgreich nach stiltypischen Kompositionen, Figurenkonstellationen (z. B. Longinus und Stephaton) oder markanten Details (z. B. INRI-Tafel) suchen. Auch hier liegt der Mehrwert nicht nur im Auffinden von Bildelementen, die nicht verschlagwortet sind, sondern auch im Entdecken von Ähnlichkeiten zwischen verschiedenen Bildelementen und der Visualisierung künstlerischer Varianz 03. Mit einem Blick lassen sich so Unterschiede in der Komposition, ausgetauschte Figuren und veränderte Haltungen leicht nachvollziehen und die Qualität von Reproduktionen vergleichen. So zeigen die gefundenen Beispiele ottonischer und byzantinischer Provenienz die kanonischen und variierenden Elemente des Motivs. Hier kommt eine gewisse Empirie

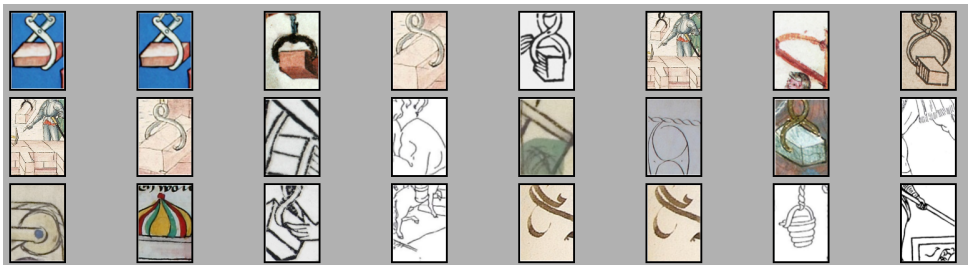
☐ 03
Suchbox im ersten Bild oben links und dann detektierte Bilder in absteigender Ähnlichkeit (Auswahl aus Marburger Porträtindex/Computer Vision Group Heidelberg).

ins Spiel, deren Ergebnisse unmittelbar ersichtlich sind und eine vergleichende Analyse vorbereiten. Aufgrund des geringen erfassten Bestandes und des selbsttätigen algorithmischen Verfahrens können nur mit aller Vorsicht Schlüsse auf die Häufigkeit der Verwendung von Assistenzfiguren in gewissen Epochen gezogen werden. Auch visuelle Übereinstimmungen zwischen verschiedenen Werken aus verschiedenen Regionen zu finden und damit gegebenenfalls Kulturtransfer nachzuweisen ist aufgrund des eher kleinen Bildkorpus noch schwierig. Je mehr Abbildungen vorhanden sind, desto sinnvoller wird auch die Suche nach kennerschaftlichen Kriterien wie Faltenwürfen, Proportionen und den von Giovanni Morelli verwendeten Detailformen wie Ohren, Fingernägeln und dergleichen mehr. Hier ist jedoch zu berücksichtigen, dass dieser Algorithmus, während er – vereinfacht gesprochen – wie eine Schablone über den Datensatz gleitet, das Suchfenster nicht rotieren lässt, sodass stilistische Eigenheiten, wie beispielsweise der Fingerpartien bei veränderten Handhaltungen, von diesem Algorithmus nicht nachvollzogen werden können. Stilistische Merkmale wie eine markante Schraffur konnten hingegen gefunden werden: Eine zum Text komplementäre Suche nach zuschreibbaren Werken ist also tendenziell möglich.

Stilkritik ist nur eine Perspektive, gleich gut lassen sich verschiedene Erzählweisen, ikonografische Varianten und Anzeichen spezifischer Frömmigkeit an diesen Synopsen untersuchen. Hinzu kommt die Möglichkeit, nach Elementen zu suchen, die gewöhnlich nicht verschlagwortet werden, wie die Beine Christi oder zeitspezifische Realien. Auf den ersten Blick erstaunlich erscheint es auch, wie gut der Algorithmus zwei ähnliche Ikonografien wie das Schweißtuch der Veronika und andere Darstellungen Christis auseinanderhalten kann. Hier schaffen die vom Nutzer verifizierten Motive offenbar genug visuelle Anhaltspunkte, um weitere Repräsentanten der Ikonografie aufzufinden.

Die Suchergebnisse ermöglichen einen schnellen Vergleich ähnlicher Kompositionen und verschiedenfarbiger Duplikate.

Durch den hohen Freiheitsgrad der Bildsuche ist die Anwendung nicht nur für die Kunstgeschichte interessant, sondern im Grunde für alle Bildwissenschaften und die interessierte Öffentlichkeit. Der visuelle Ansatz ist dafür sehr integrativ, da er sprachliche und fachlich-terminologische Grenzen überwindet und so Zugänge jenseits hermetischer Terminologien ermöglicht. So lässt sich in der Buchmalerei intuitiv nach mittelalterlichen Realien wie der Steinzange suchen, ohne den gegenwärtigen oder mittelalterlichen Begriff oder dessen Synonyme zu kennen 04. Daraufhin lassen sich die verschiedenen Arten der Darstellungen und mutmaßliche Funktionsweisen vergleichen, wodurch auch auf Abstraktionsgrad und in einigen Fällen auf bautechnischen Verfahren zurückgeschlossen werden kann.



☐ 04
Zweite Ergebnisliste der Suche nach Steinzangen auf einem Datensatz von 258 Baustellenbildern (Computer Vision Group Heidelberg).

Daneben ist eine großflächige Suche nach Kompositionen, Seitenspiegeln und Layouts möglich, um das Material nach formalen Charakteristika zu ordnen.

Besonders ergiebig für die Kunstgeschichte ist die Suchmöglichkeit nach Szenen und wiederkehrenden Motiven. Dies lässt sich gut an einem Korpus von 2510 Darstellungen mit antiken Sarkophagen zeigen. Die Bildhauer kombinierten immer wieder gleiche und ähnliche Motive in unterschiedlichen Anordnungen. In Literatur und Datenbanken gibt es nur exemplarische Gegenüberstellungen dieser wiederkehrenden Einzelszenen und schematischen Porträtdarstellungen. Der Algorithmus findet hingegen relativ sicher ikonografisch gleiche Partien wie auch ähnliche Szenen mit anderen Ikonografien, auch wenn diese in unterschiedlichen Zusammenhängen (z. B. pagan versus christlich) auftauchen, und ordnet sie übersichtlich der Ähnlichkeit nach an. Durch die Konzentration des Algorithmus auf Konturen sind die Farbigkeit und technische Unterschiede (Plastik/Zeichnung) für den Algorithmus nur wenig relevant, wodurch Reproduktionen der Originale und der Nachzeichnungen sowie Rekonstruktionen meist unmittelbar zusammen gefunden werden. 05

☐ 05
Suche nach **Gefangennahme Petri** führt zu ikonografisch richtigen Treffern (grün markiert) und ähnlichen Kompositionen (prometheus/Computer Vision Group Heidelberg).

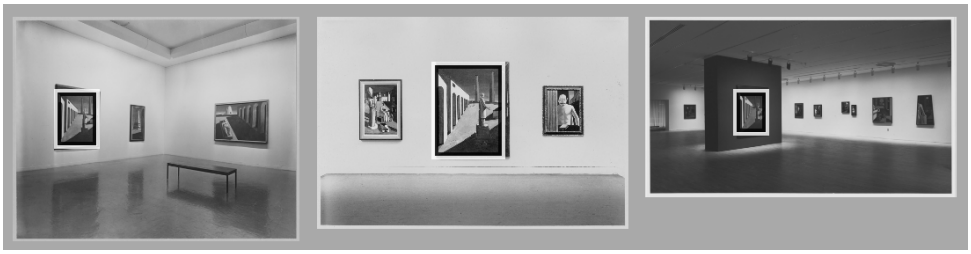
C.3 Methodische Anwendung der freien Bildsuche

Die freie Bildsuche lässt sich also auf sämtliche Dimensionen von Ähnlichkeit anwenden und durch den Nutzer präzisieren. Die Mensch-Maschine-Interaktion ist bei einem Algorithmus, der keine erlernten Kategorien oder anderes semantisches Hintergrundwissen besitzt, sehr hilfreich, wenngleich auch der **unvoreingenommene Blick** des Computers auf die rein visuelle Ähnlichkeit zu überraschenden Ergebnissen kommt und feste Denkmuster irritieren kann oder kennerschaftliche Annahmen quantitativ bestätigt.

Bildvergleiche, die gewöhnlich eher anhand von zwei vergleichbaren Werken durchgeführt werden, lassen sich nun ausweiten auf eine größere Zahl an Vergleichsbeispielen an verschiedenen Bildelementen und Detailszenen. Darin liegt nicht nur ein veränderter quantitativer Zugang zu Bildern, sondern auch ein veränderter hermeneutischer Ansatz, da hier ein Bildeinstieg jenseits gängiger Verschlagwortung möglich ist. Themen wie der **gestus melancholicus**, dem Kopf in der aufgestützten Hand, konnten bislang nur durch weitreichende Kennerschaft und aufwendige manuelle Suche in ihrer visuellen Geschichte nachvollzogen werden. Durch Bildsuchen eröffnet sich die Möglichkeit, Häufigkeiten und Veränderungen breiter zu identifizieren und sich damit Themen wie Pathosformeln, Gebärden aber auch einer semantisch aufgeladenen Realienkunde zu widmen, ohne dabei den Großteil der Ressourcen auf die Zusammenstellung eines Korpus zu verwenden. Grundsätzlich eröffnet sich die Möglichkeit, sich großen Bildbeständen zu widmen und dadurch auch Kontexte über größere geografische Räume und Zeitspannen zu rekonstruieren. Die Technologie führt nicht nur zu einem **Distant Viewing** 20, sie schafft auch neue visuelle Strukturen und überschaubare Korpora, die dann wieder konventionell und nah am Objekt untersucht werden können.

■ **20**
Der Begriff meint die Übertragung von Franco Moretti, Distant reading, Konstanz 2016 auf Bildwissenschaften. Vergleiche hierzu den Beitrag von Glinka/Dörk (→ 235) in diesem Band.

Die freie Bildsuche eignet sich also besonders für Nutzer, die jenseits der herkömmlichen Verschlagwortung visuelle Bild- und Detailvergleiche vornehmen möchten. Sie kann jedoch auch während des Einpflegens der Daten benutzt werden, um den Datensatz visuell zu erschließen. Für diese Arbeiten empfiehlt sich aber eher das in [e] beschriebene Verfahren, in dem der Computer nicht punktuell sucht, sondern mithilfe neuronaler Netze Strukturen in großen Datensätzen sondiert. Diese auf vielen Ebenen angeordneten künstlichen Neuronen, die jeweils visuelle Muster abgleichen und miteinander verbunden sind, sind gleichzeitig auch das aktuell erfolgreichste Computer-Vision-Werkzeug.
Nach einem Lernvorgang kann der Computer so Bilder Künstlern oder Epochen zuordnen und Kompositionen in Gruppen ordnen. Aktuell beschäftigt sich die Computer Vision Group Heidelberg auch mit Ausstellungszusammenhängen. Ziel ist es, herauszufinden, ob der Computer, nachdem er die Bilder der Werkliste gesichtet hat, weitere Objekte dazu vorschlagen kann. Vor dem Hintergrund der Vielseitigkeit von kuratorischen Konzepten und der damit ganz unterschiedlich gearteten Stringenz von Ausstellungen wird hier keine effiziente Anwendung angestrebt, sondern eine eher assoziative und experimentelle Annäherung an künstlerische und kunstwissenschaftliche Zusammenhänge und ihre computergestützte Nachvollziehbarkeit. 06



☐ 06
Drei Hängungen von de Chiricos **Rätsel eines Tages** im MoMA (MoMA/Computer Vision Group Heidelberg).

Die Besonderheit der freien Bildsuche liegt darin, dass nicht das Kunstwerk als Element angesehen wird und so nur damit korrespondierende Werke aufgefunden werden können, sondern jedes Detail Ausgangspunkt einer Suche sein kann. Insgesamt entsteht ein besserer Eindruck der Korrespondenzen innerhalb des Bildbestandes.

Wichtig für den Ansatz ist nicht nur, dass nach einem beliebigen Gegenstand gesucht werden kann, sondern dass dieser im Kontext eines Werkes auch beliebig groß sein kann. So lässt sich aus einer wandfüllenden Tapisserie oder einem Wimmelbild ein Detail herausgreifen und danach suchen, um ähnliche Objekte in der Datenbank zu finden. Dadurch entsteht ein offener Zugriff auf Realien anderer Epochen und es wird eine individuelle Suche nach Figuren und Formen möglich, die quer zu Taxonomien und Deutungsmustern stehen können und eine komplementäre Anordnung zur kuratorischen Präsentation bieten. Spannend wird es, wenn sich durch die Bildsuche eine Antikenrezeption für eine Assistenzfigur nachweisen lässt oder der dreifüßige Schemel als beliebtes Requisit flämischer Malerei in vielen Variationen visualisiert wird. So liegt auch das Spektrum der Anwendungen zwischen Handreichungen für die Forschung und den Interessen und Steckenpferden eines breiten Publikums.

Um eher beliebige visuelle Ähnlichkeiten für die Suche optional auszuschließen oder als wenig relevant abzuwerten, empfiehlt es sich, die Bildsuche mit der Textverschlagwortung und deren semantischer Struktur zu kombinieren.

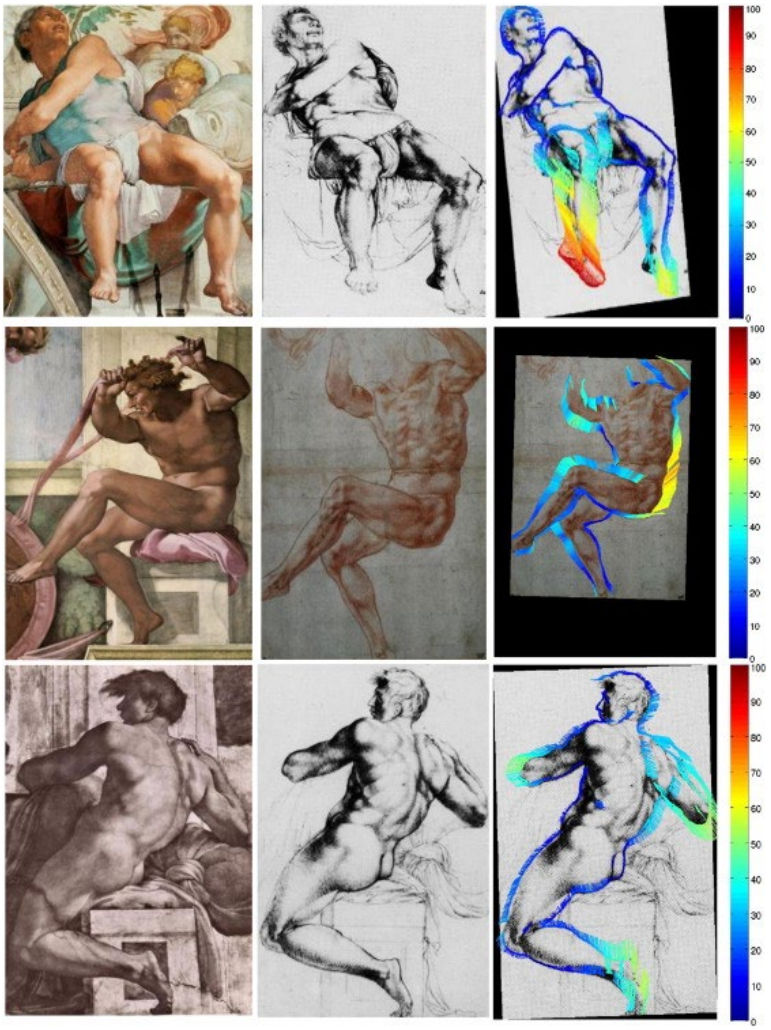
C.4 Bildvergleich im Detail

Ein weiteres Anwendungsfeld ist der detaillierte Vergleich von ähnlichen Bildern, der auch den vorhergehenden Verfahren nachgelagert sein kann. Es stellt sich z. B. die Frage, wo eine Reproduktion vom Original abweicht, um Beschädigungen zu restaurieren oder die Rezeption des Werkes in einer späteren Epoche zu verstehen.

Dafür werden die Konturen erfasst und das Werk und seine Reproduktion quasi übereinandergelegt. 21 Daraufhin wird geprüft, wie die Linie der Kopie transformiert werden muss, um passgenau auf der Kontur des Originals zu liegen. Das Verfahren ermittelt dabei Regionen gleicher Transformation. Dadurch wird erkannt, wo die Kopie sehr deutlich und ggf. bewusst von der Vorlage abweicht oder wo nur mechanische Fehler passieren, weil sich beispielsweise das Pauspapier verschoben hat.

■ 21
Peter Bell, Björn Ommer, Morphological Analysis for Investigating Artistic Images. In: Image and Vision Computing 32(6), Amsterdam 2014, S. 414–423. https://doi.org/10.1016/j.imavis.2014.04.004.

Auf diese Weise lassen sich beispielsweise Michelangelos Ignudi von der Decke der Sixtinischen Kapelle mit späteren Kopien vergleichen und die Unterschiede deutlich visualisieren 07. Die teilweise starken Abweichungen sind der hohen Herausforderung geschuldet, die an der gewölbten Decke befindlichen Fresken proportionsgetreu abzuzeichnen. Gerade bei den Partien des Torsos, wo die Konturen weit auseinanderliegen, kommt es zu Ungenauigkeiten.



☐ 07
Nebeneinanderstellung von Michelangelos Sixtina-Fresken, deren Reproduktionen und der automatischen Analyse der Transformationen (Computer Vision Group Heidelberg).

Durch dieses Verfahren gewinnt man sehr schnell einen Überblick über die Abweichungen von Reproduktionen und kann Tendenzen erkennen, wie etwa sich wiederholende mechanische Fehler oder optische Verzerrungen, die beispielsweise infolge eines ungünstigen Standpunkts des Betrachters oder aufgrund der Kameraoptik entstehen können. Die Ursachen einer Abweichung zwischen Original und Kopie liegen bei dem Vergleich zweier Reproduktionen natürlich dann nicht nur bei der Kopie, denn auch bei der Reproduktion des Originals kann es zu Abweichungen kommen, und letztlich kann sich auch das Original selbst gegenüber seinem ursprünglichen Zustand verändert haben. Auch wenn die Algorithmen zum jetzigen Zeitpunkt nicht imstande sind, die Ursachen für die jeweiligen Abweichungen zu identifizieren, unterstützt der präzise Vergleich über Gruppen von Abbildungen die Erkennung von wiederkehrenden Fehlern. Neben den technischen Befunden zum Reproduktionsprozess

werden auch semantische Veränderungen etwa in den Abschriften von illuminierten Handschriften sichtbar. So lässt sich in den Versionen des Sachsenspiegels beobachten, dass Gesprächspartner in den Abschriften einen weiteren Abstand erhalten oder enger zusammenrücken. Der Computer markiert diese Abweichung, während die Interpretation, ob sie technische oder inhaltliche Gründe hat, weiterhin beim menschlichen Betrachter liegt. Die Anwendungsfelder für diesen detaillierten Vergleich von visuell ähnlichen Werken liegen sowohl in der Interpretation wie auch im Aufbau einer digitalen Kennerschaft, durch die Fragen nach Genauigkeit von Reproduktionen, Zuschreibungen und restauratorische Fragen geklärt werden können.

C.5 Fazit

Der Einsatz von Computer Vision in der Kunstgeschichte ist eine Möglichkeit, den Ansprüchen einer Bildwissenschaft in vielen Arbeitsbereichen gerechter zu werden. Datensätze werden visuell erschließ- und durchsuchbar und es entsteht eine im Grunde ebenfalls bildgetriebene Mensch-Maschinen-Interaktion. Auf diese Weise werden sowohl große Bildmengen komplementär zum Text visuell vorstrukturiert wie auch Detailformen und einzelne Konturen aufspürbar und vergleichbar. Damit entstehen Visualisierungen, die parallel zur Textproduktion bildlich argumentieren.

Die Computer Vision lässt sich auf alle Bildbestände der Kunstgeschichte anwenden, erzielt jedoch bei prägnanten Formen und häufigen Wiederholungen in leichten Variationen tendenziell bessere Ergebnisse. Die Interaktion zwischen Mensch und Maschine ist dialogisch. Der Mensch lässt sich gerade bei großen Bildmengen assistieren und unterstützt die Suche, indem er den Computer zuvor trainiert oder während der Such- und Analyseschritte Rückmeldungen gibt. Gerade dieses iterative Verfahren entspricht dem hermeneutischen Forschungsansatz der Kunstgeschichte innerhalb einer quantitativen Umgebung. Die Schlussfolgerungen werden weiterhin durch den Menschen aufgrund seines historischen, stilistischen und topografischen Kontextwissens gezogen, der Computer macht hingegen nur Vorschläge, was jedoch zu so treffenden Gegenüberstellungen führen kann, dass sie mit Erkenntnis gleichzusetzen ist, die der Mensch lediglich bestätigen muss.

Zum jetzigen Zeitpunkt ist schwer abzusehen, wie die semantische Analyse von Bildinhalten fortschreiten wird. Einfache Szenen des Alltagslebens können schon jetzt richtig eingeordnet werden, große Schwierigkeiten bereitet jedoch der Wandel von Realien (man denke nur an die Kostümgeschichte) und der Wechsel in den Darstellungsmodi über Jahrhunderte hinweg, zumal für manche Epochen und Regionen zu wenig Bildmaterial vorhanden ist, um Deep-Learning-Verfahren darauf anzuwenden. Obgleich die Funde und Erkenntnisse in verhältnismäßig kleinen Datensätzen noch begrenzt sind, wird schon jetzt der Gewinn durch die entstehenden Visualisierungen deutlich. Auf das Detail zugespitzt kann hier die Methode des Bildvergleichs vervielfacht und präzisiert werden und ein genauerer Eindruck des Ähnlichkeitsverhältnisses der Motive gewonnen werden.

Nach dem Abschluss erster Fallstudien und Projekte zwischen Computer Vision und Kunstgeschichte wird jedoch auch deutlich, dass einzelne, insbesondere kleine Bildarchive und Museen das Potenzial des automatischen Sehens nur beschränkt nutzen können, da hier oft nur ein überschaubarer Bestand an wirklich visuell vergleichbaren Objekten vorliegt. Eine enge Kooperation zwischen Bildarchiven, Museen und der Forschung zum automatischen Sehen in den Digital Humanities bildet daher eine wichtige Voraussetzung für eine visuelle Erschließung des gesamten kulturellen Erbes.

Nur so wird das Einzelwerk/-motiv in seinem — soweit noch erhaltenen — visuellen Gesamtkontext erfahrbar.