



UNIVERSITÄT ZU LÜBECK

**Aus dem Institut
für Neuro- und Bioinformatik
der Universität zu Lübeck
Direktor: Prof. Dr. rer. nat. Thomas Martinetz**

**Informationstheoretische Ansätze
zur Analyse von Kunst**

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

Aus der Sektion Informatik

vorgelegt von
Friedrich Anselm Brachmann
aus Jena
Lübeck, 2018

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung:

Zum Druck genehmigt. Lübeck, den

Danksagung

Diese Arbeit hätte nicht ohne die Unterstützung der Menschen fertig gestellt werden können, die mich über den Zeitraum ihrer Entstehung begleitet und dabei maßgeblich unterstützt haben. Besonderer Dank gilt meinen Betreuern Prof. Dr. Dr. Christoph Redies und Prof. Dr. Erhardt Barth, die sich bereit erklärt haben, diese Arbeit zu betreuen und mir dabei stets mit Rat und Tat zur Seite standen. Den Mitarbeitern unseres Instituts danke ich für ihre Teilnahme an der durchgeführten Studie zur Symmetrie von Bildern und Frau Maria Grebenkina für ihre Erlaubnis, hierfür den von ihr zusammengestellten Datensatz von Bildern nutzen zu dürfen.

Zusammenfassung

Der rasche Fortschritt auf dem Gebiet der Bildverarbeitung ermöglicht neue Ansätze zur Erforschung teils jahrhundertealter Fragestellungen, wie die nach den Eigenschaften als ästhetisch angesehener Bilder. In dieser Arbeit werden verschiedene Methoden und Algorithmen entwickelt, die es ermöglichen, die Statistik von traditioneller Kunst in Beziehung zu anderen Bildkategorien zu setzen. Die Betrachtung der Entropie relativer Kantenbeziehungen ergibt, dass Kunst im Vergleich mit anderen Bildkategorien am meisten den Bildern gleicht, die Landschaftsszenen zeigen. Aufbauend auf einer bestimmten Kodierung der Filterantworten eines tiefen neuronalen Netzes wird ein Verfahren vorgestellt, um die Spiegelsymmetrie von Bildern zu messen. Diese Kodierung der Filterantworten nutzend wird darüber hinaus ein Ansatz beschrieben, um die Selbstähnlichkeit von Bildern im Sinne einer Wiederholung von Bildinhalten zu quantifizieren. Eine Untersuchung der Varianz von Bildinhalten anhand der Filterantworten tiefer neuronaler Netze erlaubt es, im direkten Vergleich zu weiteren Bildkategorien zwei Eigenschaften traditioneller Kunst zu identifizieren. Zum einen stellt sich Kunst bezogen auf die Bildstruktur als sehr viel reichhaltiger dar, eine Eigenschaft, die Kunst mit Pflanzenmustern teilt. Zum anderen wird in Kunst jedoch eine Wiederholung der Bildstruktur vermieden, wodurch sich Kunstbilder als weniger selbstähnlich als Pflanzenmuster darstellen und sich damit von diesen abheben. Die technische Analyse der Bilder erlaubt somit die Gewinnung neuer Hypothesen über die Form von traditioneller Kunst.

Inhaltsverzeichnis

Einleitung	1
Fragestellung und Zielsetzung	2
Aufbau der Arbeit	3
Wissenschaftliche Beiträge	4
1 Computational Aesthetics und Experimentelle Ästhetik	7
1.1 Computational Aesthetics	9
1.1.1 Vorhersage ästhetischer Bewertungen	10
1.1.2 Anderweitige Klassifikation von Bildern	17
1.1.3 Weitere Anwendungen der Computational Aesthetics	22
1.2 Experimentelle Ästhetik	25
1.2.1 Grundannahmen der Experimentellen Ästhetik	26
1.2.2 Luminanz- und Farbstatistik	28
1.2.3 Komplexität	30
1.2.4 Symmetrie, Balance und die Drittel-Regel	31
1.2.5 Fourier-Spektrum	32
1.2.6 Fraktale Dimension und Selbstähnlichkeit	33
1.2.7 Orientierungen von Helligkeitsgradienten	35
1.3 Zusammenfassung	35
2 Entropie orientierter Gradienten	39
2.1 Material und Methoden	40
2.1.1 Berechnung der Entropie orientierter Gradienten	40
2.1.2 Datensatz zur Evaluation	45
2.2 Ergebnisse	46
2.2.1 Paarweise Statistik von Kantenorientierungen	46

2.2.2	Vergleich der Kantenentropie erster und zweiter Ordnung .	48
2.3	Diskussion der Ergebnisse	55
2.4	Zusammenfassung	56
3	Symmetrie	59
3.1	Convolutional Neural Networks	60
3.2	Material und Methoden	63
3.2.1	Messen stetiger Symmetrie	63
3.2.2	Datensatz zur Evaluation	65
3.3	Ergebnisse	66
3.3.1	Ergebnisse des Bewertungsexperimentes	66
3.3.2	Ergebnisse der Symmetrieberechnungen	67
3.4	Diskussion der Ergebnisse	72
3.5	Zusammenfassung	74
4	Selbstähnlichkeit	77
4.1	PHOG-Selbstähnlichkeit	79
4.2	Material und Methoden	81
4.2.1	Berechnung von Selbstähnlichkeit mittels CNN-Filterantworten	81
4.2.2	Datensatz zur Evaluation	82
4.3	Ergebnisse	84
4.3.1	CNN-Selbstähnlichkeit der ersten Schicht	84
4.3.2	CNN-Selbstähnlichkeit höherer Schichten	86
4.4	Diskussion der Ergebnisse	91
4.5	Zusammenfassung	95
5	Varianz von CNN-Filterantworten für Kunst	97
5.1	Material und Methoden	98
5.1.1	Varianz von CNN-Filterantworten	98
5.1.2	Datensatz zur Evaluation	100
5.2	Ergebnisse	101
5.3	Diskussion der Ergebnisse	109
5.3.1	Interpretation der Varianzen P_a und P_f	111
5.4	Zusammenfassung	115

6	Diskussion	119
6.1	Zur Wahl des Kunstdatensatzes	119
6.2	Zur Aussagekraft der Ergebnisse	120
6.3	Bezug der vorgeschlagenen Maße zu etablierten Maßen der Experimentellen Ästhetik	122
6.4	CNNs als Modelle menschlichen Sehens	126
6.5	Diskriminative und generative Modelle	128
	Schlussbetrachtung	131
	Offene Fragen und Ausblick	134
	Literatur	137
A	Allgemeine Ergänzungen	159
A.1	Verzeichnis abgebildeter Kunstwerke	159
A.2	Ergänzungen zu Kapitel 5	160
	Lebenslauf	163

Abkürzungsverzeichnis

CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
HIK	Histogram Intersection Kernel
HOG	Histogram of Oriented Gradients
PHOG	Pyramid Histogram of Oriented Gradients
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SIFT	Scale-Invariant Feature Transform
SVM	Support Vector Machine

Einleitung

Moderne Methoden der Bildanalyse eröffnen neue Ansätze zur Beantwortung teils jahrhundertealter Fragestellungen. So hat sich in jüngerer Zeit innerhalb der Informatik das Teilgebiet der *Computational Aesthetics*¹ herausgebildet, auf welchem Algorithmen erforscht werden, anhand derer beispielsweise die ästhetische Bewertung von Bildern vorhergesagt werden kann. Auch Kunst wird hier mit rechnergestützten Methoden genauer untersucht.

Ästhetik selbst ist dabei ein Thema, welches seit mehr als zweitausend Jahren von Gelehrten diskutiert wird. Vom Oxford Dictionary unter anderem definiert als „die Philosophie des Schönen oder der Kunst“ („the philosophy of the beautiful or of art“; OED (2017)) und „die bestimmten, zugrundeliegenden Prinzipien eines Kunstwerkes oder Genres“ („the distinctive underlying principles of a work of art or a genre“, ebd.), hat die Ästhetik bereits früher das Interesse weiterer wissenschaftlicher Disziplinen geweckt. In der *Experimentellen Ästhetik*, einem Teilgebiet der Psychologie und gleichzeitig sehr viel älter als die Computational Aesthetics, ist es das Ziel, Ästhetik an sich besser zu verstehen. In der Vergangenheit wurden hier verschiedene Modelle entwickelt, anhand derer ästhetische Urteile und das ästhetische Erleben erklärt werden sollen (Leder et al., 2004; Leder und Nadal, 2014; Redies, 2015).

Langer (2016) betont, dass Ästhetik unvermeidbar ein interdisziplinäres Projekt sein muss. Während sich die Gebiete der Experimentellen Ästhetik und die Computational Aesthetics vor wenigen Jahren noch gegenseitig befruchtet haben, ist der Wissensaustausch mit dem durchschlagenden Erfolg tiefer neuronaler Netze schwieriger geworden (Lu et al., 2015). Die in der Computational Aesthetics eigens für die Erfassung des ästhetischen Wertes entworfenen Bildmerkmale erlaubten

¹Im Verlauf dieser Arbeit wird der Begriff *Computational Aesthetics* unübersetzt verwendet.

insofern einen Zugang, als dass deren Beitrag zur Klassifikation direkt beobachtbar war. Heute zeigen tiefe neuronale Netze große Erfolge bei der Vorhersage ästhetischer Bewertungen, indem diese die formalen, für diese Fragestellung relevanten Kriterien eines Bildes überwacht lernen und anschließend auf deren ästhetischen Wert abbilden können. Aus Sicht der Experimentellen Ästhetik geht dabei jedoch die einfache Interpretationsmöglichkeit des Ergebnisses verloren und die konkreten formalen Aspekte, welche die Ästhetik eines Bildes bestimmen, bleiben unbekannt. Dabei bieten neuronale Netze als moderne Methoden der Bildanalyse auch für die Experimentelle Ästhetik großes Potential für ein besseres Verständnis als ästhetisch angesehener Bilder, was sich in dieser Arbeit zunutze gemacht wird.

Fragestellung und Zielsetzung

Im Vordergrund dieser Arbeit steht die Entwicklung neuer Methoden und Algorithmen, anhand derer die formalen Eigenschaften visueller Kunst erfasst und mit denen anderer Kategorien von Bildern verglichen werden können. Diese Art der Untersuchung von Kunst hat innerhalb der Experimentellen Ästhetik eine lange Tradition. Hier ist die Annahme, dass sich das visuelle System des Menschen im Laufe der Evolution an die natürliche Umgebung angepasst hat und sich diese Anpassung auch in den Eigenschaften von Kunst niederschlägt. Einerseits hat sich dabei die sogenannte „Savannen-Hypothese“ herausgebildet. Diese besagt, dass Menschen Bilder mit offenen Landschaften bevorzugen, welche gleichzeitig sichere Lagerplätze darstellen (Orians und Heerwagen, 1992). Andere argumentieren, dass die Statistik von Kunst die Statistik natürlicher Szenen nachbildet, um dieselbe Kodierung innerhalb des visuellen Systems zu erreichen (Taylor et al., 1999; Graham und Field, 2007; Redies et al., 2007b; Schweinhart und Essock, 2013). Eine zentrale Frage meiner Arbeit ist somit, welche Aussagen über die statistischen Eigenschaften von Kunst – verglichen mit anderen Bildkategorien – getroffen werden können. Insbesondere interessiert dabei, welche Ähnlichkeiten sich zu verschiedenen anderen Bildkategorien finden lassen, was Hinweise auf die erwähnte, selbe Kodierung liefern kann.

Gleichzeitig haben sich verschiedene Autoren in der Vergangenheit dafür ausgesprochen, dass Kunst eine besondere Kategorie visueller Reize darstellt. Zur Beschreibung dieses Umstandes wurden dabei verschiedene Begriffe geprägt wie

beispielsweise *good gestalt* (Arnheim, 1954), *visual rightness* (Locher et al., 1999) oder *significant form* (Bell, 1914). Neben den Gemeinsamkeiten von Kunst mit anderen Bildkategorien wird in dieser Arbeit ebenso nach Kriterien gesucht, welche Kunst von anderen Bildkategorien unterscheiden, und es wird ein Ansatz vorgestellt, diese Kriterien zu identifizieren.

Mit Ausnahme einer Methode, welche die Entropie der Kantenorientierungen von Bildern misst (Kapitel 2), wird auf Filterantworten tiefer neuronaler Netze in Form von *Convolutional Neural Networks* (CNNs, Fukushima (1980) und Lecun und Bengio (1995)) zurückgegriffen. CNNs lernen eine Hierarchie von Filtern, welche auf ein Bild angewendet relevante Informationen aus diesem extrahieren können. Während untere Schichten eines Netzes auf einfache Merkmale wie Farbe oder Kantenorientierungen ansprechen, kodieren höhere Ebenen abstraktere Bildinhalte und eignen sich mit dieser Ähnlichkeit zum menschlichen Sehsystem gut als Modell für Fragestellungen innerhalb der Experimentellen Ästhetik (Denzler et al., 2016).

Aufbau der Arbeit

Im sich anschließenden Kapitel 1 werden die Gebiete der Computational Aesthetics und der Experimentellen Ästhetik vorgestellt und die verschiedenen Ansätze beider Gebiete sowie deren Überschneidungen näher beschrieben. Das Kapitel wurde als Übersichtsartikel veröffentlicht (siehe Brachmann und Redies (2017a)) und bietet einen umfassenden Überblick über die Kernthemen beider Gebiete. Indem die Entwicklung der Computational Aesthetics nachvollzogen wird, wird verdeutlicht, weshalb der Wissensaustausch mit der Experimentellen Ästhetik erschwert worden ist, ein Problem, dessen Adressierung gleichzeitig eine der Hauptmotivationen dieser Arbeit darstellt.

Kapitel 2 wendet sich der Methodenentwicklung zu und geht der Frage nach, wie sich die Orientierungen von Luminanzkanten in Kunstbildern relativ zueinander verhalten. Anhand des vorgestellten Verfahrens wird die Statistik von Kunstbildern mit anderen Bildkategorien verglichen und Kunst bezogen auf diese genauer eingeordnet. Das Kapitel beruht auf der Veröffentlichung von Redies et al. (2017) und wurde in weiteren Arbeiten bereits erfolgreich aufgegriffen (siehe Redies und Brachmann (2017) und Henemann et al. (2017)).

In Kapitel 3 werden zunächst tiefe neuronale Netze genauer beschrieben, die

eine zentrale Grundlage der Methodenentwicklung dieser Arbeit darstellen. Anschließend wird in dem genannten Kapitel ein Algorithmus entworfen, mithilfe dessen die Symmetrie von Bildern anhand der Filterantworten tiefer neuronaler Netze berechnet werden kann und eine erfolgreiche Anwendung des Verfahrens demonstriert. Gleichzeitig wird dabei in diesem Kapitel eine Art der Kodierung der Filterantworten von CNNs beschrieben, die in den beiden sich daran anschließenden Kapiteln verwendet wird, um Kunstbilder zu analysieren. Kapitel 4 greift ein etabliertes Maß (Amirshahi et al., 2013b) zur Bestimmung der Selbstähnlichkeit von Bildern auf, welches sich innerhalb der Experimentellen Ästhetik zur genaueren Einordnung von Kunst bereits bewährt hat, dabei jedoch die Farbinformationen eines Bildes verwirft. Da Farbe insbesondere für die Betrachtung von Kunst eine wichtige Qualität des Bildes darstellt, wird hier ein neues Maß zur Bestimmung der Selbstähnlichkeit entworfen, welches auch Farbe in die Analyse einbezieht. Während das in Kapitel 4 entwickelte Maß zur Bestimmung der Selbstähnlichkeit von Bildern genutzt wird, um Kunst mit anderen Bildkategorien zu vergleichen, wird in dem darauffolgenden Kapitel 5 die Varianz von Bildmerkmalen als Grundlage genommen, um der Frage nachzugehen, ob Kunst eventuell einer wie von Bell (1914) vorgeschlagenen „Signifikanten Form“ folgt. Es werden dabei drei verschiedene Arten der Messung der Varianz von Bildmerkmalen entwickelt und getestet, ob Kunst durch eine bestimmte Kombination derselben von anderen Bildkategorien abgegrenzt werden kann. Da sich der Aufbau der vorliegenden Arbeit stark an den genannten Veröffentlichungen orientiert, werden die sich ergebenden Fragen in den jeweiligen Kapiteln selbst diskutiert. Eine Diskussion, welche Fragen bezüglich des Gesamtkonzepts der Arbeit aufgreift und damit alle Kapitel gleichzeitig betrifft, findet sich in Kapitel 6. Die Schlussbetrachtung fasst die Ergebnisse der Arbeit zusammen und bewertet diese abschließend.

Wissenschaftliche Beiträge

Zusammengefasst werden in dieser Arbeit die folgenden wissenschaftlichen Beiträge geleistet, die bereits in den in Klammern aufgeführten Publikationen veröffentlicht wurden:

- In Form eines Übersichtskapitels wird ein umfassender und vergleichender Einblick in die Gebiete sowohl der Computational Aesthetics als auch der Experimentellen Ästhetik (Kapitel 1; Brachmann und Redies (2017a)) gegeben.
- Es wird ein Verfahren zur Analyse relativer Kantenbeziehungen in Bildern beschrieben und anschließend genutzt, um verschiedene Arten von Kunst untereinander und mit anderen Bildkategorien zu vergleichen. Dabei zeigt sich, dass Kunst bezüglich der Kantenbeziehungen am ehesten Bildern gleicht, welche Landschaftsszenen zeigen (Kapitel 2; Redies et al. (2017) und Redies und Brachmann (2017)).
- Beruhend auf einer bestimmten Kodierung der Filterantworten tiefer neuronaler Netze wird ein Algorithmus zur Berechnung der Spiegelsymmetrie von Bildern entworfen und anschließend dessen praktische Anwendbarkeit demonstriert (Kapitel 3; Brachmann und Redies (2016)).
- Ein etabliertes Maß zur Bestimmung der Selbstähnlichkeit von Bildern wird aufgegriffen und um Farbinformationen erweitert, wobei die in Kapitel 3 vorgestellte Kodierung der Filterantworten eines tiefen neuronalen Netzes genutzt wird. Anhand des erweiterten Maßes wird die Selbstähnlichkeit von Kunst mit verschiedenen Kategorien anderer Bilder verglichen (Kapitel 4, Brachmann und Redies (2017b)).
- Der Hypothese folgend, dass Kunst eine bestimmte „Signifikante Form“ (Bell, 1914) zeigt, werden verschiedene Maße beschrieben, welche die Varianz der Filterantworten tiefer neuronaler Netze fassen. Im Anschluss daran werden diese genutzt, um Kunst von anderen Bildkategorien abzugrenzen, wobei durch eine genauere Interpretation dieser Maße verschiedene Eigenschaften von Kunst identifiziert werden können. Es stellt sich dabei heraus, dass Kunst eine hohe Reichhaltigkeit der Bildstruktur besitzt, dabei jedoch gleichzeitig eine hohe Variabilität aufweist, sodass eine Wiederholung der Bildstruktur vermieden wird (Kapitel 5; Brachmann et al. (2017)).

Aus Sicht der Informatik setzt diese Arbeit somit einen Schwerpunkt auf der Entwicklung neuer Methoden zur Bildanalyse und nutzt diese gleichzeitig, um verschiedene Fragestellungen der Experimentellen Ästhetik zu untersuchen. Wie sich

die vorliegende Arbeit in Bezug auf bisherige Arbeiten sowohl der Computational Aesthetics als auch der Experimentellen Ästhetik verortet, wird im folgenden Kapitel deutlich werden.

Kapitel 1

Computational Aesthetics und Experimentelle Ästhetik

Im Jahr 1876 veröffentlichte Gustav Fechner sein Buch „Vorschule der Ästhetik“ (Fechner, 1876), worin er die Annahme formulierte, dass die ästhetische Wirkung eines physischen Objektes in dessen Eigenschaften begründet liege und damit auch objektiv messbar sei. Er versuchte zu zeigen, dass Rechtecke, welche dem Seitenverhältnis des *Goldenen Schnittes* folgen, bei dem das Verhältnis der längeren zur kürzeren Seite dem Verhältnis der Summe beider zur längeren Seite entspricht, von Menschen ästhetisch bevorzugt werden. Später wurde Kritik an seiner Methode laut und es ist allgemein anerkannt, dass der Goldene Schnitt vom Betrachter nicht bevorzugt wird (McManus et al., 2010; Green, 1995). Gleichwohl bot Fechners Annahme formal messbarer Kriterien für Schönheit die Basis für das weite und heute wieder populäre Feld der *Experimentellen Ästhetik*. Hier werden Hypothesen bezüglich der Schönheit von Bildern, Gemälden oder auch alltäglichen Gegenständen aufgestellt und experimentell überprüft. Dieser stimulusgetriebene Ansatz, von Fechner auch als „Ästhetik von unten“ bezeichnet, unterschied sich von dem bis dahin verbreiteten Ansatz der „Ästhetik von oben“, bei dem Schönheit und Ästhetik vor allem anhand philosophischer Herleitungen erklärt wurden (Cupchik, 1986). Für eine Einführung in die Geschichte der Ästhetik sei auf Gaut und Lopes (2005) verwiesen.

Auf dem Gebiet der *Computational Aesthetics*, welches innerhalb der Informatik der Bildverarbeitung zugeordnet werden kann, ist Ästhetik in jüngerer Zeit auch

mit rechnergestützten Methoden untersucht worden. Hier werden Fotografien oder andere digital reproduzierte Bilder hinsichtlich ihres ästhetischen Wertes oder ihrer statistischen Eigenschaften genauer untersucht. Als Ursprung der Computational Aesthetics wird weithin das Buch „Aesthetic Measure“ von George David Birkhoff (Birkhoff, 1933) gesehen, auch wenn dieser darin den Begriff selbst nicht verwendet (für eine Übersicht zur Herkunft des Begriffes sei auf Greenfield (2005) verwiesen). Birkhoff versuchte Ästhetik auf mathematischem Weg zu erklären und schlug eine Formel für ein „Ästhetisches Maß“ M als Verhältnis von O , *Ordnung*, und C , *Komplexität*, vor, wobei letzteres die Anstrengung des Betrachters zum Verstehen des Reizes bezeichnet. Da seines Verständnisses nach Ordnung und Komplexität proportional zueinander stehen sollten, schlussfolgerte er, dass der Zusammenhang am besten durch das Verhältnis $M = O/C$ beschrieben werden kann.

Von Hoenig (2005) wird Computational Aesthetics definiert als „[...] die Erforschung rechnergestützter Methoden, die anwendbare ästhetische Entscheidungen treffen können, auf ähnliche Weise, wie Menschen es können“ („[...] the research of computational methods that can make applicable aesthetic decisions in a similar fashion as humans can“). Für Hoenig betont diese Definition zwei Hauptaspekte: Zum einen die Nutzung rechnergestützter Methoden, zum anderen aber auch deren Anwendbarkeit für die Entscheidung über die Ästhetik eines Reizes. Darüber hinausgehend argumentiert Galanter (2012), dass die Computational Aesthetics neben dem Bewerten von Kunst auch das Erzeugen derselben betrifft. Er hebt hervor, dass das Erzeugen von Kunst zwangsläufig mit dem Bewerten einhergeht und nennt als Beispiel den Künstler, der während des Lernens über Kunst die Werke anderer bewertet, im eigenen Schaffensprozess durch viele kleine Einschätzungen des Produktes seiner Arbeit geleitet wird, und auch das fertige Werk schlussendlich begutachtet. In diesem Kapitel konzentriere ich mich daher vor allem auf die Evaluation, der Aspekt des Erzeugens wird jedoch ausgeklammert. Stork (2009a) betont, dass die computergestützte Methode der Analyse von Kunstwerken gegenüber der Analyse von Menschen mehrere Vorteile besitzt. Beispielsweise können so winzige Details aufgegriffen werden, die der menschlichen Wahrnehmung möglicherweise entgehen. Auch ermüde ein Computer nicht und könne eine sehr große Menge von Details verarbeiten, etwa jeden einzelnen Pinselstrich eines Bildes. Der größte Vorteil sei jedoch dessen Objektivität, welche ein zwingendes Kriterium für die wissenschaftliche Untersuchung ist.

In diesem Kapitel soll einerseits ein Überblick über aktuelle Entwicklungen auf dem Gebiet der Computational Aesthetics gegeben werden. Andererseits sollen auch grundlegende Konzepte der Experimentellen Ästhetik vorgestellt werden, um potentielle neue Wege für die computergestützte Bildanalyse zu eröffnen. Auf dem Gebiet der Computational Aesthetics wird dabei ein Einblick in die Vorhersage ästhetischer Bewertungen von Fotografien und Kunstwerken, sowie der Klassifikation verschiedener Künstler und Stile gegeben. Darüber hinaus werden exemplarisch verschiedene Arbeiten vorgestellt, die verschiedenen Fragen der Kunstgeschichte oder der Feststellung der Authentizität von Bildern nachgehen. Abgeschlossen wird dieses Kapitel mit einer Beschreibung verschiedener statistischer Eigenschaften als ästhetisch gewerteter Bilder, welche im Rahmen der Experimentellen Ästhetik herausgearbeitet wurden.

1.1 Computational Aesthetics

Auf dem Gebiet der Computational Aesthetics lassen sich verschiedene Fragestellungen identifizieren. Viele der hier vorgestellten Arbeiten beschäftigen sich mit der Ästhetik von Bildern in Form von Fotografien oder Gemälden und sind oft dadurch motiviert, konkrete Anwendungen zu erstellen oder Algorithmen zu verbessern und zu testen. So ist beispielsweise ein klassisches Anwendungsgebiet die Vorhersage der ästhetischen Bewertung von Bildern. Diese Algorithmen ermöglichen das automatisierte Auffinden ästhetisch wertvoller Bilder in großen Datenmengen, wie sie bei Diensten wie Flickr¹ oder Photo.net² anfallen, wo Nutzer ihre Bilder online veröffentlichen. Eine kommerzielle Verwertbarkeit ist in dem Moment gegeben, in dem diese Bilder lizenziert oder talentierte Fotografen auf einfache Weise gefunden werden können. Mit der massenhaften Verbreitung von Kameras in Smartphones können Algorithmen in Bruchteilen von Sekunden höherqualitative von minderqualitativen Bildern unterscheiden und dem Anwender eine Auswahl der besten Bilder präsentieren. Wie im folgenden Abschnitt gezeigt werden wird, wurden diese Systeme mit großem Erfolg entworfen und implementiert.

Während die Vorhersage der Bewertung von Bildern nur ein möglicher Anwendungsfall ist, werden unter dem Begriff der Computational Aesthetics auch andere

¹<https://www.flickr.com>

²<https://www.photo.net>

Einsatzgebiete zusammengefasst. Methoden aus der Bildanalyse wurden ebenfalls erfolgreich eingesetzt, um den Inhalt von Kunstbildern zu analysieren, Fälschungen zu detektieren oder eventuelle Einflüsse von Künstlern auf andere Künstler festzustellen. Auch diese Aspekte sollen im Folgenden näher beschrieben werden.

1.1.1 Vorhersage ästhetischer Bewertungen

Ein Schwerpunkt innerhalb der Computational Aesthetics ist es, die Bewertung von Bildern vorherzusagen, welche vorher hinsichtlich ihres ästhetischen Wertes von Versuchspersonen abgegeben wurden. In den frühen Tagen folgte man dabei der allgemein in der Bildverarbeitung weit verbreiteten Vorgehensweise, bei der in Abhängigkeit des Anwendungsfalles Algorithmen entworfen wurden, die bestimmte, für den Anwendungsfall als relevant erachtete Merkmale im Bild detektieren und numerisch kodieren. Um die Ästhetik eines Bildes algorithmisch zu fassen, wurden zunächst Algorithmen implementiert, welche bekannte, fotografische Prinzipien im Bild identifizierten. Beispielsweise kann anhand von Salienzalgorithmen festgestellt werden, ob die Komposition des Bildes der sogenannten „Drittel-Regel“ folgt. Diese Regel besagt, dass ein Bild ästhetischer wirkt, wenn dessen herausstechende Merkmale auf den Linien ihren Schwerpunkt haben, die das Bild horizontal und vertikal in drei gleich große Regionen unterteilen. Ein weiteres relevantes Merkmal ist die Schärfentiefe eines Bildes, denn oft wird es als ästhetisch empfunden, wenn der unwichtige Hintergrund eines Bildes nicht im Fokus ist und harte Kanten vermissen lässt.

Die genannten spezifischen Merkmale können verschieden kategorisiert werden. Zum einen unterscheidet man *lokale* und *globale* Merkmale. Lokale Merkmale werden auf einzelnen Pixeln und deren unmittelbaren Umgebung berechnet, während globale Merkmale das gesamte Bild einbeziehen. Globale Merkmale scheinen dabei besonders geeignet, die Ästhetik eines Bildes zu fassen, besonders wenn man sich Begriffen wie der Komposition des Bildes zu nähern versucht. Des Weiteren können *low-level* und *high-level* Merkmale unterschieden werden, eine Einteilung, die die Merkmale hinsichtlich ihres Abstraktionsgrades beurteilt.³ Während *low-*

³Da für beide Begriffe eine adäquate Übersetzung ins Deutsche schwierig ist, werden diese im Folgenden unübersetzt verwendet.

level Merkmale vor allem Kanten oder Farben betreffen, spricht man im Falle von Bildinhalten von *high-level* Merkmalen. Beide können für eine ästhetische Bewertung des Bildes von Relevanz sein.

Nachdem die jeweiligen Merkmale eines Bildes detektiert wurden, dienen diese als Eingabe eines Klassifikators, welcher darauf trainiert wird, ästhetische von nicht ästhetischen Bildern zu unterscheiden. Der Klassifikator lernt dabei eine Abbildung der Merkmalsmenge auf die Bewertungen, welche wiederum im Vorfeld durch eine Befragung von Menschen gewonnen wurden. Die genannten Online-Fotodienste bieten den Vorteil, dass eine große Menge von Bewertungen automatisiert heruntergeladen werden können, haben gleichzeitig jedoch den Nachteil, dass über die Bewertenden selbst wenig bekannt ist. Als Klassifikatoren stehen verschiedene zur Auswahl, eingesetzt werden beispielsweise Support Vector Machines (SVMs), Bayes-Klassifikatoren oder Entscheidungsbäume. Wir werden den Einsatz spezifischer Merkmale als Vorgehensweise in Abschnitt 1.1.1.1 genauer betrachten. Während die gezielte Merkmalsextraktion mit anschließendem Lernen eines Klassifikators in den Anfangstagen der Computational Aesthetics weit verbreitet war, wurden spezifische Merkmale in den letzten Jahren zunehmend von generischen Merkmalen verdrängt. Ihren vorläufigen Höhepunkt hat die Entwicklung in dem Einsatz tiefer neuronaler Netze gefunden, wobei Merkmale ohne von Menschen eingebrachtes Vorwissen automatisiert gelernt werden. Die generischen Merkmale werden in Abschnitt 1.1.1.2 genauer betrachtet.

1.1.1.1 Spezifische Bildmerkmale

Einer der ersten Versuche, den ästhetischen Wert von Bildern zu fassen, wurde von Tong et al. (2004) veröffentlicht. Sie schlagen eine Methode vor, mit der Fotografien professioneller Fotografen von denen von Laien unterschieden werden können. Genutzt werden dabei insgesamt 21 verschiedene *low-level* Merkmale wie Unschärfe, Kontrast oder die Farbreichhaltigkeit eines Bildes. Die Autoren vergleichen verschiedene Klassifikatoren und kommen zu dem Schluss, dass eine SVM bessere Ergebnisse erzielt als Boosting oder ein Bayes-Klassifikator. Anhand einer anderen Auswahl von Merkmalen zeigen Datta et al. (2006), wie Bilder von hohem ästhetischen Wert von Bildern geringeren Wertes abgegrenzt werden können. Die 3581 für ihre Studie genutzten Bilder samt zugehöriger Bewertungen

wurden bei Photo.net heruntergeladen. Die Autoren formulieren als explizites Ziel ihrer Studie, dass sie Schlussfolgerungen aus den diskriminativsten Merkmalen ziehen wollen, um so etwas über die Ästhetik von Bildern zu lernen. Die Wahl der genutzten Merkmale gründen die Autoren auf allgemeinen Regeln der Fotografie sowie bestimmten Mustern, welche sie in den Bildern ihres Datensatzes ausmachten. In der Summe werden 56 verschiedene Merkmale vorgeschlagen, darunter Farbreichhaltigkeit, Sättigung oder die Auflösung und das Seitenverhältnis eines Bildes, sowie darüber hinaus das Auftreten der Drittel-Regel. Um den Einfluss verschiedener Merkmale auf die Bewertungen miteinander zu vergleichen, beginnen die Autoren mit dem diskriminativsten Merkmal und fügen dann in einem iterativen Prozess diejenigen Merkmale hinzu, welche die Klassifikation am meisten verbessern. Sie schließen, dass die durchschnittliche Farbe, die durchschnittliche Intensität und eine auf Sättigung basierende Drittel-Regel am meisten dazu beitragen, die Bewertung der Ästhetik eines Bildes vorherzusagen. Ke et al. (2006) entwerfen ein System, um hochqualitative Fotografien professioneller Fotografen von Schnappschüssen minderer Qualität unterscheiden zu können. Ähnlich wie Datta et al. (2006) versuchen die Autoren festzustellen, worin genau der ästhetische Wert eines Bildes liegt und gründen ihre Auswahl von Merkmalen auf eine Befragung von Fotografen. Sie testen unter anderem die Verteilung von Kanten und Farbe, die Zahl verschiedener Farben sowie die Unschärfe eines Bildes. Als Klassifikator wird ein Bayes-Klassifikator genutzt, wobei sich die Unschärfe des Bildes als wichtigstes Merkmal für die Unterscheidung von Schnappschüssen und hochqualitativen Fotografien herausstellt. Luo und Tang (2008) konzentrieren sich in ihrer Methode auf Merkmale wie Beleuchtung, Einfachheit, Komposition oder Farbharmonie, basierend auf dem Vorder- und dem Hintergrund eines Bildes. Laut Einschätzung der Autoren ist diese Unterscheidung der Grund für eine relative Verbesserung gegenüber den Methoden von Datta et al. (2006) und Ke et al. (2006), welche ein Bild als Ganzes betrachten. Über die alleinige Fokussierung auf *low-level* Merkmale hinaus konzentrieren sich Dhar et al. (2011) vor allem auf Merkmale, die Menschen zur Beschreibung eines Bildes heranziehen würden, um den ästhetischen Wert und die „Interessantheit“ eines Bildes vorherzusagen. Die Merkmale werden dabei in die Kategorien Kompositionsmerkmale, wie z. B. die Drittel-Regel, Bildinhalt, wie etwa die An- und Abwesenheit von Menschen,

oder die natürliche Beleuchtung im Bild unterteilt. Bezogen auf Ke et al. (2006) berichten die Autoren eine Verbesserung durch ihren Ansatz.

Während der Fokus oben genannter Arbeiten vor allem auf der Vorhersage der Ästhetik von Fotografie konzentriert scheint, gibt es auch Publikationen, die sich mit der Ästhetik von Malerei beschäftigen. Li und Chen (2009) schlagen insgesamt 40 verschiedene Merkmale, darunter Farbreichtum, Helligkeit und weitere zum Erfassen der Komposition, vor. Sie verwenden einen Bayes-Klassifikator sowie AdaBoost, um zu unterscheiden, ob ein Bild gute oder schlechte Bewertungen erhielt. Darüber hinaus diskutieren sie die Relevanz einzelner Merkmale und zeigen, dass anhand des Helligkeitsunterschiedes der Segmente eines Bildes dessen ästhetische Bewertung am besten vorhergesagt werden kann.

Alle vorgestellten Ansätze haben gemein, dass sie eine Auswahl spezifischer Merkmale im Bild detektieren, um anschließend einen Klassifikator zu trainieren, der eine Unterscheidung der Bilder bezüglich deren Ästhetik vornimmt. Während sich dies als teilweise sehr erfolgreich herausgestellt hat, gibt es dennoch einige Schwierigkeiten, die einem tieferen Verständnis des Zusammenhangs zwischen den Bildmerkmalen und der Bildästhetik im Wege stehen. Zum einen sind die erhobenen Merkmale nicht zwangsläufig voneinander unabhängig und beeinflussen sich eventuell gegenseitig. Zum anderen sind die experimentellen Gegebenheiten, unter welchen die Bewertungen erhoben worden, oft unbekannt. Im Falle anonymer Nutzerbewertungen kann beispielsweise davon ausgegangen werden, dass eine Vielzahl verschiedener Monitore den Kontrast desselben Bildes unterschiedlich dargestellt und Bilder je nach Qualität der Darstellung unterschiedlich bewertet werden. Eine belastbare Aussage über die Rolle des Bildkontrastes auf die ästhetische Bewertung kann daher nicht getroffen werden. Für Farbe als Merkmale ist Ähnliches anzunehmen. Darüber hinaus sind nähere Informationen zu den Versuchsteilnehmern in der Regel unbekannt, dennoch können demographische Merkmale das Ergebnis beeinflussen. In der experimentellen Ästhetik werden viele dieser Faktoren isoliert unter definierten Versuchsbedingungen getestet, worauf wir in Abschnitt 1.2 näher eingehen werden.

1.1.1.2 Generische Bildmerkmale

Als generische Bildmerkmale werden Merkmale bezeichnet, die nicht explizit zur Vorhersage der ästhetischen Bewertung von Bildern dienen, sondern vielmehr für andere Anwendungsgebiete innerhalb der Bildverarbeitung entwickelt wurden. Sie stehen damit den oben beschriebenen spezifischen Merkmalen entgegen. Ein Beispiel sind die von Lowe (2004) vorgestellten SIFT-Deskriptoren (*Scale-Invariant Feature Transform*). Diese beschreiben lokale Umgebungen von Schlüsselpunkten innerhalb eines Bildes derart, dass diese in anderen Bildern unabhängig gegenüber Transformationen wie Skalierungen, Rotationen oder Translationen wiedergefunden werden können. Eine der ersten Methoden, welche auf generische Merkmale für die Vorhersage der ästhetischen Bewertungen von Bildern zurückgreift, wurde von Marchesotti et al. (2011) vorgeschlagen. Die Autoren nutzen die genannten SIFT-Deskriptoren in Kombination mit einem Farbdeskriptor und begründen ihre Wahl damit, dass SIFT-Deskriptoren implizit auch für die Ästhetik relevante Merkmale, wie etwa das Vorhandensein kontrastreicher Kanten, kodieren können. Anhand prototypischer Bildausschnitte entscheidet ein Klassifikator anschließend über die ästhetische Bewertung des Bildes. Als Ergebnis wird berichtet, dass diese Methode eine Verbesserung gegenüber den genannten Ansätzen von Ke et al. (2006) und Datta et al. (2006) erzielt, die mit spezifischen Bildmerkmalen arbeiten. Während spezifische Bildmerkmale erlauben, eine ungefähre Einschätzung der relativen Wichtigkeit derselben vorzunehmen, geht diese Art der Interpretationsmöglichkeit mit generischen Bildmerkmalen verloren. Aussagen können hier vor allem anhand von Beobachtungen getroffen werden, welche Bilder der Klassifikator voneinander abgrenzt. So beschreiben Marchesotti et al. (2011), dass in ihrem Experiment vor allem unscharfe und gering aufgelöste Bilder durch ihren Klassifikator als weniger ästhetisch beurteilt wurden, während Bilder mit scharfem Vordergrund und unscharfem Hintergrund höhere Bewertungen erhielten. Auch zeigten gut bewertete Bilder eine dominante Farbe oder enthielten vermehrt Gegenfarben als Kontrast. Waren hingegen zu viele Farben in einem Bild gleichzeitig repräsentiert, wurde ein Bild weniger gut bewertet.

In den letzten Jahren haben tiefe neuronale Netze, in der Bildverarbeitung in Form von *Convolutional Neural Networks* (CNNs; Fukushima (1980) und Lecun und Bengio (1995)), viele Teilbereiche erobert und dabei konventionelle Methoden

ersetzt. In den Kapiteln 3–5 wird ebenfalls ein CNN verwendet werden, daher sei für eine ausführlichere Darstellung dieser Modelle auf Abschnitt 3.1 verwiesen. An dieser Stelle soll nur die für das Verständnis nötige Grundidee dargestellt werden. CNNs lernen eine Hierarchie von Filtern, anhand derer für eine Fragestellung relevante Merkmale eines Bildes extrahiert werden können. Während Filter unterer Schichten dabei auf einfache Kanten bestimmter Orientierung, Frequenz oder Farbe ansprechen, werden Filterantworten auf höheren Schichten miteinander kombiniert, sodass die Abstraktion mit höheren Schichten zunimmt (Yosinski et al., 2015). Trainiert werden CNNs üblicherweise *überwacht*, indem die aktuelle Ausgabe eines Netzes mit der gewünschten Zielausgabe verglichen wird, was es ermöglicht, den Fehler des Netzes zu einem aktuellen Trainingszeitpunkt zu bestimmen. Anhand der Kettenregel kann für jeden einzelnen der mitunter vielen Millionen Parameter des Netzes ein Gradient bestimmt werden, anhand dessen jeder Parameter wiederum so angepasst wird, dass sich der Gesamtfehler des Netzes verringert. Diese gemeinhin als *Backpropagation* (Rumelhart et al., 1986) bezeichnete Trainingsmethode ermöglicht es, nach einer großen Zahl von Trainingsiteration bestimmte Abbildungen von Eingabebildern auf mit dem Bildinhalt zusammenhängende Ausgaben zu lernen. Neben der Objekterkennung, der Bildbeschreibung oder der Bildsegmentierung gibt es eine Vielzahl weiterer Anwendungsgebiete, wobei nur eines die Vorhersage der ästhetischen Bewertung eines Eingabebildes darstellt.

Lu et al. (2015) entwerfen ein zwispaltiges CNN, welches neben der globalen Ansicht der Bilder gleichzeitig lokale Ausschnitte betrachtet. Ziel dabei ist es, die ästhetische Bewertung des Eingabebildes vorherzusagen. Sie begründen ihre Vorgehensweise damit, dass Ästhetik ihrer Einschätzung nach sowohl durch lokale, als auch globale Faktoren bedingt wird. Nach einer Auswertung verschiedener Strategien für die Auswahl der lokalen Ausschnitte zeigen sie, dass ihr Ansatz bisherige Ergebnisse auf dem selben Datensatz (*Aesthetic Visual Analysis (AVA)*; Murray et al. (2012)) übertrifft. Dong et al. (2015) nutzen mit der populären *AlexNet*-Architektur (Krizhevsky et al., 2012) ein CNN, welches auf 1.2 Millionen Bildern (ImageNet; Russakovsky et al. (2015)) trainiert wurde, um 1000 verschiedene Kategorien von Objekten zu unterscheiden. Die Autoren trainieren eine SVM auf den Filterantworten der höchsten Schicht des Netzes nach Anwendung auf das gesamte Bild und konkatenieren diese mit den Filterantworten derselben Schicht

bei Anwendung auf lokale Ausschnitte des Bildes. Die Autoren berichten, dass sie so die von Marchesotti et al. (2011) berichtete Klassifikationsrate um zehn Prozent übertreffen. Dieses Ergebnis ist besonders interessant, da bei den Filtern der höchsten Ebene davon ausgegangen werden kann, dass die Klassifikationsrate vor allem durch Einbeziehung des Bildinhaltes erreicht wird und weniger auf der Form des Bildes beruht (Yosinski et al., 2015). Kao et al. (2016) trainieren ein tiefes neuronales Netz darauf, sowohl den Inhalt, als auch die ästhetische Bewertung eines Bildes vorherzusagen. Nach einer Evaluation verschiedener Topologien kommen sie zu dem Schluss, dass ihr Ansatz, semantische Informationen einzubeziehen, bisherige Methoden übertrifft. Dieses Ergebnis deckt sich mit vorgeschlagenen Modellen ästhetischer Wahrnehmung (Leder et al., 2004; Redies, 2015), in denen Form und Inhalt zusammen das ästhetische Erleben beeinflussen.

Neuronale Netze scheinen den Ansatz der Detektion spezifischer Merkmale als Grundlage für die ästhetische Bewertung größtenteils verdrängt zu haben. Neben höheren Klassifikationsraten erweisen sich neuronale Netze auch in anderen Aspekten als vorteilhaft. Zum einen werden Merkmale automatisch gelernt, ohne dass viel Zeit in eine aufwendige Implementation als relevant vermuteter Merkmale investiert werden muss. Auch kombinieren neuronale Netze lokale und globale Strukturen, und kommen dabei ohne eine manuelle Gewichtung derselben aus, vielmehr wird diese automatisch gefunden. Abstrakte Bildinhalte, welche semantische Informationen tragen, fließen ebenso in die Bewertung ein wie formale Kriterien, und darüber hinaus werden diese zueinander in Beziehungen gesetzt. Ein weiterer Vorteil neuronaler Netze ist die Tatsache, dass diese in der Lage sind, dem Menschen unbekannte Kriterien eines Bildes zu identifizieren, welche dessen Ästhetik positiv oder negativ beeinflussen und von Fotografen und Künstler eventuell unbewusst eingesetzt oder entsprechend vermieden werden.

Aus der Perspektive der Experimentellen Ästhetik bringt die vermehrte Anwendung tiefer neuronaler Netze jedoch auch einen Nachteil mit sich. Momentan entspricht das reine Ende-zu-Ende-Lernen einem *Black-Box*-Ansatz, d. h. es ist nicht ohne Weiteres möglich nachzuvollziehen, warum sich bestimmte Ausgaben ergeben. Zwar existieren Arbeiten, die unser Verständnis bezüglich der Filterantworten höherer Ebenen verbessert haben (Zeiler und Fergus, 2014; Yosinski et al., 2015), allerdings ist unser Wissen noch relativ rudimentär und weit von einem umfassen-

den Verständnis tiefer Merkmale entfernt. So können zwar immer leistungsfähigere Klassifikatoren entworfen werden, welche die durch Menschen getroffenen Einschätzungen der Ästhetik von Bildern immer besser abbilden, diese Methoden entziehen sich jedoch gleichzeitig der Möglichkeit, unser Verständnis von Ästhetik an sich zu erweitern. Während in den frühen, in Abschnitt 1.1.1.1 vorgestellten Studien sogar oft ein erklärtes Ziel war, den Einfluss verschiedener Bildmerkmale auf die ästhetische Bewertung zu quantifizieren, ist diese Möglichkeit bei tiefen Netzen teilweise verloren gegangen.

1.1.2 Anderweitige Klassifikation von Bildern

Über die reine ästhetische Bewertung eines Bildes hinaus gibt es andere Teilgebiete der Computational Aesthetics, welche mehr auf Malerei als auf Fotografie konzentriert sind. Dabei werden Bilder nicht entsprechend ihrer ästhetischen Wirkung bewertet, vielmehr geht es darum, den Künstler oder den Kunststil zu identifizieren, eine Aufgabe, die üblicherweise Kunstexperten überlassen wird. Aus methodischer Perspektive sind beide Ziele sehr ähnlich. Während frühere Arbeiten jedoch eher die Identifikation des Künstlers (Cezanne, Vermeer, Rembrandt, oder ähnliche) behandeln, beschäftigen sich jüngere Arbeiten mehr mit der Identifikation des Stiles (Realismus, Kubismus, Impressionismus, oder ähnliche). Ein Grund hierfür mag darin liegen, dass mehr und mehr Sammlungen durch Museen digitalisiert und online verfügbar gemacht werden. Ein weiterer Grund für diese Verschiebung ist, dass die stetig wachsende Rechenleistung es erlaubt, größere Mengen von Bildern in gleicher Zeit zu verarbeiten. Denkbare Anwendungen solcher Methoden sind Empfehlungssysteme für Kunstkäufer oder auch bessere Beschreibungen der stilistischen Eigenschaften der Bilder einzelner Künstler. Beide Themen sollen im folgenden anhand exemplarischer Arbeiten vorgestellt werden.

1.1.2.1 Identifikation des Künstlers

Mithilfe eines Bayes-Klassifikators identifiziert Keren (2002) den Autor eines Bildes, wobei Werke von Rembrandt, van Gogh, Picasso, Magritte und Dali einbezogen werden. Er nutzt als Merkmale die Koeffizienten einer diskreten Kosinustransformation von 9×9 Pixel großen Ausschnitten eines Bildes. Jeder dieser Blöcke wird einem Künstler zugeordnet und die Mehrheit innerhalb eines Bildes bestimmt

den Künstler, wobei eine Klassifikationsrate von 86% erreicht wird. Widjaja et al. (2003) konzentrieren sich in ihrer Arbeit auf Hauttöne, um den Maler eines Bildes zwischen Rubens, Michelangelo, Ingres, und Botticelli auszumachen. Verschiedene Farbräume und Klassifikatoren werden getestet. Als Endergebnis wird die beste Klassifikationsrate einer DAG-SVM (Platt et al., 2000) mit 85% angegeben. Für einen Datensatz mit Tintenzeichnungen fünf chinesischer Künstler schlagen Li und Wang (2004) eine Klassifikationsmethode basierend auf Wavelets und Hidden-Markov-Modellen vor und berichten, dass diese Methode auch genutzt werden kann, um die Ähnlichkeit von Bildern zu quantifizieren. Shen (2009) kombiniert verschiedene globale Merkmale (Farbe, Textur und Form) mit lokalen Merkmalen (Gabor-Wavelets) und erreicht eine Klassifikationsrate von 69.7% bei der Unterscheidung von 25 klassischen Künstlern, darunter Caravaggio, Rubens, Vermeer und van Gogh, wobei als Klassifikator ein einfaches neuronales Netz eingesetzt wird. Khan et al. (2010) nutzen SIFT-Deskriptoren und einen Farbdeskriptor um je 40 Bilder acht verschiedener Künstler (Ingres, Matisse, Monet, Picasso, Rembrandt, Rubens, Titian und van Gogh) mit einer SVM als Klassifikator voneinander abzugrenzen und berichten 62% korrekte Zuordnungen. Einen größeren Datensatz mit insgesamt 1896 Bildern von 15 verschiedenen Künstlern, darunter Pollock, Rembrandt, Cezanne und Magritte, nutzen Condorovici et al. (2013). Anhand von *low-level* Merkmalen (RGB-Histogramme und Gabor-Filterantworten) werden diese unterschieden, wobei im direkten Vergleich verschiedener Klassifikatoren logistische Regression die besten Ergebnisse erzielt. Cetinic und Grgic (2013) extrahieren drei verschiedene Arten von *low-level* Merkmalen bezüglich Farbe, Intensität und Textur eines Bildes und erreichen eine Klassifikationsrate von 75.3% bei der Unterscheidung von 20 verschiedenen Künstlern. Als Klassifikator wird ein einfaches neuronales Netz mit einer verdeckten Schicht eingesetzt.

Insgesamt ist es schwierig, diese verschiedenen Ansätze miteinander zu vergleichen, da sie auf verschiedenen Datensätzen arbeiten und, davon kann ausgegangen werden, auch methodisch an diese angepasst sind. So ist es beispielsweise zielführend, für Bilder mit einer hohen Diversität von Hauttönen Farbe als Merkmal zu verwenden, während Farbe auf anderen Bildern wenig Informationen beiträgt. Condorovici et al. (2013) adressieren dieses Problem teilweise und vergleichen verschiedenen Ansätze (Li und Wang, 2004; Widjaja et al., 2003; Keren, 2002) mit ihrem eigenen Ansatz anhand der jeweiligen Ratewahrscheinlichkeit auf den

genutzten Datensätzen. Diese Art des Vergleiches ist jedoch insofern schwierig, als dass hier Methoden besser abschneiden, welche von vorne herein leichter zu unterscheidende Künstler voneinander abgrenzen. Beispielsweise könnte es schwieriger sein, Bilder von Claude Monet und Paul Cezanne zu unterscheiden, beides Maler des Impressionismus, als ein abstraktes Bild von Jackson Pollock von einem surrealistischen Werk von René Magritte. Zusammengefasst nutzen alle der genannten Ansätze verschiedene *low-level* Merkmale, die Textur, Kanteninformationen oder Farbe eines Bildes erfassen; Bildinhalte werden hingegen nicht betrachtet. Auch wird oft nicht nur ein Klassifikator genutzt, sondern verschiedene Klassifikatoren werden miteinander verglichen.

1.1.2.2 Erkennung des Kunststils

Für die Erkennung des Stils von Kunstwerken wurden verschiedene Verfahren vorgeschlagen, die sich teilweise mit denen überschneiden, die für die Identifikation des Künstlers vorgeschlagen wurden. Gunsel et al. (2005) schlagen insgesamt sechs verschiedene Merkmale vor, darunter die Helligkeit und die Grauwertverteilung eines Bildes, und grenzen so Bilder der Stile Klassizismus, Impressionismus, Kubismus, Expressionismus und Surrealismus voneinander ab. Die Autoren geben an, dass ihr System jedoch auch dazu in der Lage sei, weitere Stile über die von ihnen gezeigte Auswahl hinaus zu diskriminieren. Jiang et al. (2006) schlagen eine Methode zur Klassifikation der chinesischen Kunststile *Gongbi* (traditionelle chinesische Kunst) und *Xieyi* (Freihandstil) vor. Auch sie nutzen dafür *low-level* Merkmale, die Textur, Farbe und Kanten im Bild erfassen. Einen anderen Ansatz verfolgen Wallraven et al. (2009), die Versuchsteilnehmer baten, Beispielbilder aus insgesamt elf verschiedenen Kunstrichtungen (darunter Gotik, Renaissance, Klassizismus, Surrealismus und Postmoderne Kunst) in selbst gewählte Kategorien zu ordnen. Die Autoren stellen fest, dass diese Kategorien sich gut mit den tatsächlichen Kunststilen decken. Für die Beispielbilder berechnen sie anschließend verschiedene einfache Merkmale beruhend auf Farbhistogrammen oder den Frequenzspektren und untersuchen, wie gut diese die tatsächliche Einordnung der Bilder erklären. Sie finden dabei nur eine geringe Korrelation und schlussfolgern, dass abstraktere Merkmale als die von ihnen gemessenen entscheidend sind. Insgesamt sechs verschiedene Kunststile (Abstrakter Expressionismus, Barock, Kubismus, Graffiti,

Impressionismus und Renaissance) werden von Siddiquie et al. (2009) anhand von Farbe, Salienz und orientierten Gradienten unterschieden. Zujovic et al. (2009) wählen Beispielbilder der fünf Kunststile Abstrakter Expressionismus, Kubismus, Impressionismus, Pop-Art und Realismus, und unterscheiden diese anhand von HSV-Histogrammen und Kanteninformationen. Ein Vergleich verschiedener Klassifikatoren zeigt, dass AdaBoost mit 69.1% korrekter Klassifikationen die höchste Genauigkeit erzielt. Die drei Stile Impressionismus, Surrealismus und Abstrakter Expressionismus werden von Shamir et al. (2010) anhand von Bildfrequenzen, Kanten- und Farbinformationen voneinander getrennt. Condorovici et al. (2015) unterscheiden insgesamt acht verschiedene Stile von Bildern und konzentrieren sich dabei auf perzeptuell inspirierte Merkmale, die Kanten und Farbe betrachten. Wie andere Autoren testen auch sie verschiedene Klassifikatoren und erreichen das beste Ergebnis mit einer SVM. Während die bereits genannten Arbeiten vor allem auf *low-level* Merkmalen basieren, die formale Aspekte eines Bildes in den Vordergrund stellen, deuten die Ergebnisse von Arora und Elgammal (2012) darauf hin, dass auch semantische Informationen für die Erkennung des Stils wichtig sein können. Die Autoren nutzen unter anderem *Classeme*-Merkmale (Torresani et al., 2010), die ein Bild anhand der Ausgaben vieler einfacher Klassifikatoren beschreiben, die ihrerseits Objekte im Bild detektieren.

Neben der Objekterkennung oder der oben angeführten Bewertung der Ästhetik eines Bildes wurden CNNs ebenfalls für die Erkennung des Stils von Bildern eingesetzt. Karayev et al. (2013) nutzen einen relativ großen Datensatz von 85 000 Bildern, die in 25 Kunststile untergliedert sind. Neben *GIST*-Deskriptoren (Oliva und Torralba, 2001) und Salienzmerkmalen nutzen sie unter anderem auch *DeCAF*-Merkmale (Donahue et al., 2014) und kodieren so neben der Form eines Bildes mit letzteren auch den Bildinhalt. Mit der Kombination aller vorgeschlagenen Merkmale erreichen sie eine Klassifikationsrate von 47.4% und berichten, dass einzeln genommen die DeCAF-Merkmale das beste Ergebnisse erzielen. Es kann daraus geschlussfolgert werden, dass dem Inhalt der Bilder eine entscheidende Bedeutung auch für die Klassifikation des Stils zukommt. Eine andere Arbeit, die ebenfalls DeCAF-Merkmale nutzt, wurde von Bar et al. (2014) veröffentlicht. In Kombination mit *PiCoDes*-Merkmalen (Bergamo et al., 2011) unterscheiden sie die 27 Kunststile eines Datensatzes von 40 724 Bildern und erreichen eine Klassifikationsrate von 43%. Saleh und Elgammal (2015) nutzen neben GIST-Deskriptoren,

PiCoDes- und Classeme-Merkmalen auch die durch ein CNN zugeordneten Objektkategorien für die Erkennung des Stils, des Genres und des Künstlers. Die Classeme-Merkmale stellen sich dabei als am besten geeignet heraus.

Während die bisher aufgeführten Arbeiten mit DeCAF-Merkmalen das vortrainierte AlexNet verwenden, trainieren Tan et al. (2016) das AlexNet-Modell direkt für die Aufgabe, das Genre, den Stil und den Künstler eines Bildes zu erkennen. Sie vergleichen die Klassifikationsrate eines Netzes, welches zuerst auf dem ImageNet-Datensatz und anschließend auf die neue Fragestellung trainiert wurde (ein Prozess, der als *fine-tuning* bezeichnet wird), mit einem von Grund auf neu trainiertem Netz und zeigen, dass das vortrainierte Netz mit einer Klassifikationsrate von 54.5% die besten Ergebnisse bei der Erkennung des Stils erzielt.

Zusammenfassend kann bemerkt werden, dass die Erkennung des Künstlers und die Erkennung des Stils einander sehr ähnliche Aufgaben sind. Die in den frühen Tagen dafür entworfenen *low-level* Merkmale sind einander ebenfalls ähnlich und wurden teilweise für beide Zielstellungen gleichermaßen eingesetzt, wobei oft nur eine kleine Auswahl von Künstlern und Stilen betrachtet wurde. Mit der freien Verfügbarkeit großer Mengen digitalisierter Kunstbilder wird das Trainieren und Testen von Algorithmen zur Erkennung des Stils zunehmend einfacher. Für die Erkennung des Künstler ist das nicht unbedingt der Fall, da die verfügbare Gesamtzahl von Werken für verschiedene Künstler sehr unterschiedlich ausfällt, was das Trainieren beispielsweise tiefer neuronaler Netze erschwert. Darüber hinaus haben verschiedene Künstler ihren Stil im Laufe ihres Lebens geändert. So haben beispielsweise viele Maler abstrakter Kunst ihre Karriere mit realistischen Bildern begonnen, genannt seien Wassily Kandinsky, Piet Mondrian und Jackson Pollock. Auch für die Erkennung des Stils werden neuronale Netze zunehmend populär und ersetzen die Klassifikation anhand spezifischer Merkmale. Dass so auch semantische Informationen eines Bildes die Klassifikationsrate verbessern können, erscheint einleuchtend, da Inhalt und Stil nicht strikt voneinander getrennt sind. Der in der Spätrenaissance ausgeprägte Stil des *Chiaroscuro* (von ital. „hell-dunkel“) beispielsweise zeigt oft Szenen in Gebäuden, wohingegen impressionistische Bilder vermehrt Landschaftsszenen zeigen, ein Umstand der von tiefen neuronalen Netzen gelernt werden kann. Gleichzeitig tritt jedoch auch hier der Nachteil auf, dass die Interpretation der Entscheidungsgrenzen eines tiefen neuronalen Netzes schwierig ist.

1.1.3 Weitere Anwendungen der Computational Aesthetics

In den vorangegangenen Abschnitten dieses Kapitels wurden Methoden beschrieben, mit deren Hilfe die ästhetische Bewertung, der Stil oder der Künstler eines Bildes vorhergesagt werden können. In diesem Abschnitt wird der Fokus auf weitere Anwendungen gelegt, bei denen Kunst mit rechnergestützten Methoden analysiert wird. Im Vordergrund steht dabei vor allem, Fragestellungen der Kunstgeschichte rechnergestützt zu beantworten, beispielsweise, ob es sich bei bestimmten Bildern um Originale oder Fälschungen handelt. Für eine Übersicht über frühe Arbeiten sei auf Stork (2009a) verwiesen. Eine aktuellere Zusammenfassung ist in Spratt und Elgammal (2014) zu finden, die verschiedene Arbeiten vorstellen, die sich ihrerseits beispielsweise mit der semantischen Annotation von Kunstwerken, der Sortierung von Kunstwerken nach Schaffensdatum, oder der Erkennung von Gemeinsamkeiten verschiedener Werke und Künstler beschäftigen.

1.1.3.1 Methoden für kunstgeschichtliche Fragestellungen

Methoden, die kunsthistorische Fragen computergestützt adressieren, können unterschiedlichen Teilgebieten zugeordnet werden. So beschäftigt sich ein Teilgebiet beispielsweise mit der Technik verschiedener Künstler, ein anderes versucht Einflüsse unter Künstlern aufzudecken. Im Folgenden werden exemplarisch einige Arbeiten kurz beleuchtet.

Criminisi et al. (2002) entwickeln Methoden, um den dreidimensionalen Raum eines realistischen Bildes zu rekonstruieren. Anhand der dabei gewonnenen Information können verschiedene Fragen bezüglich des Dargestellten untersucht werden. Die Autoren zeigen, dass beispielsweise die genaue Größe von Personen und Objekten anhand relativer Beziehungen berechnet werden kann. In Criminisi und Stork (2004) wird der Frage nachgegangen, ob der flämische Maler Jan van Eyck, der für seinen Realismus bekannt ist, optische Hilfen wie etwa Spiegel bei der Schaffung des Bildes „Portrait of Arnolfini and his Wife“ (1434) nutzte. Die Autoren stellen im genannten Bild perspektivische Ungenauigkeiten fest und kommen so zu dem Schluss, dass es ohne derlei Hilfen entstanden ist. Stork und Johnson (2006) schlagen ein Verfahren zur Rekonstruktion der Lichtquellen in einem Bild vor. Für Georges de La Tour Bild „Christ in the Carpenters Studio“ (1645) weisen sie den

Verdacht zurück, dass die Lichtquelle außerhalb des Bildes liegt, was ebenfalls ein Hinweis auf optische Hilfen hätte sein können. Papaodysseus et al. (2006) untersuchen für Wandgemälde der Bronzezeit, ob Schablonen bei der Erstellung des Bildes eingesetzt wurden. Mithilfe einer Hough-Transformation finden sie verschiedene Grundformen, für die sich die Annahme zu bestätigen scheint. Die Farben und die Textur der Bilder von van Gogh untersuchen Berezhnoy et al. (2005) und können so die Hypothese bestätigen, dass der Maler in seinem späteren Leben zunehmend mit Gegenfarben gearbeitet hat. In Berezhnoy et al. (2009) stellen die Autoren ein Verfahren vor, um die Pinselstrichtechnik von Gemälden computergestützt analysieren zu können.

Der kreative Einfluss von Künstlern auf andere kann anhand von Ähnlichkeiten ihrer Werke untersucht werden. Mithilfe von SIFT-Merkmalen und einer Farbstatistik entwerfen Bressan et al. (2008) einen Graph, der wechselseitige Einflüsse von Malern verschiedener Epochen aufzeigt. Shamir und Tarakhovsky (2012) nutzen viele verschiedene Merkmale, um Formen, Textur und Farbe von Bildern zu charakterisieren und finden ausgeprägte Gruppen für klassische Künstler wie Vermeer oder Rembrandt, wie auch für moderne Künstler wie Jackson Pollock, Marc Rothko oder Wassily Kandinsky. Wang und Takatsuka (2012) nutzen Farb- und Kompositionsmerkmale, um Renaissance, Impressionismus und Postimpressionismus zu unterscheiden. Zusätzlich nutzen sie diese, um Künstler hierarchisch zu ordnen und zeigen Einflüsse auf Picassos verschiedene Schaffensperioden auf. Mithilfe von Classeme-Merkmalen versuchen Abe et al. (2013) Einflüsse von Künstlern auf andere Künstler aufzudecken, die in der Kunstgeschichte bisher nicht in Betracht gezogen wurden. Sie präsentieren verschiedene Gemäldepaarungen, deren Ähnlichkeit eine Beeinflussung des früheren Gemäldes auf das spätere nahelegt. Auf einem Datensatz von über 60 000 Bildern schlagen Elgammal und Saleh (2015) ein Verfahren vor, anhand von GIST- und Classeme-Merkmalen die Kreativität im Sinne der Originalität von Kunstwerken im Vergleich mit zeitgenössischen Werken zu quantifizieren und zeigen, dass sich ihre Ergebnisse dabei mit den Einschätzungen von Kunstexperten decken.

1.1.3.2 Erkennung von Fälschungen

Ein weiteres Beispiel, wo rechnergestützte Methoden Kunsthistorikern wichtige Antworten liefern können, ist die Erkennung von Kunstfälschungen. Das Gebiet ist verwandt mit der bereits behandelten Identifikation des Künstlers, unterscheidet sich jedoch bezüglich des betrachteten Detailgrades. Während bei der Identifikation des Künstlers Werke voneinander abgegrenzt werden, deren Unterschiede oftmals auch für Laien erkennbar sind, sind diese bei dem Gegensatzpaar Fälschung und Original sehr viel subtiler und auch dem geübten Auge oft nur schwer zugänglich. Beide Ansätze beruhen auf der Identifikation von Unterschieden zwischen Kunstwerken. Ein Algorithmus, der für den einen Anwendungsfall gute Ergebnisse liefert, ist jedoch nicht zwangsläufig auf den anderen übertragbar.

Lyu et al. (2004) vergleichen die Wavelet-Statistik von acht dem niederländischen Maler Pieter Bruegel zugeschriebenen Werken mit der von fünf Imitationen seiner Arbeit und zeigen, dass die Statistik der imitierten Bilder signifikant abweicht. In einer anderen Anwendung ihrer Methode lösen sie das Problem der „vielen Hände“, bei der Kunsthistoriker daran interessiert sind, wie viele verschiedene Maler an der Fertigstellung eines Bildes beteiligt waren. Sie identifizieren mindestens vier verschiedene Maler für die Darstellung von Gesichtern in einem Bild von Pietro Perugino, einem Maler der Renaissance, und bestätigen damit die bisherige Einschätzung von Kunsthistorikern. Polatkan et al. (2009) stellen einen Datensatz von Originalen und bewusst erstellten Kopien von Bildern vor, auf dem sie anhand von Wavelet-Koeffizienten zeigen, wie diese auseinander gehalten werden können. Johnson et al. (2008) stellen drei Methoden verschiedener Autoren vor, welche anhand einer Wavelet-Dekomposition Werke von van Gogh von Fälschungen unterscheiden. Eine Analyse der Strichtechnik von Vincent van Gogh wird von Li et al. (2012) vorgeschlagen, um so Fälschungen zu erkennen. Darüber hinaus zeigen die Autoren, dass Werke mit dieser Methode auch datiert werden können. Neben den Werken van Goghs haben auch die Bilder des amerikanischen Künstler Jackson Pollock das Interesse verschiedener Autoren geweckt. Taylor et al. (1999) analysiert die fraktalen Eigenschaften dessen sogenannter „Drip-Paintings“ mithilfe der *Box-Counting*-Methode und stellen fest, dass die fraktale Dimension, ein Maß für die Komplexität, in späteren Werken des Künstlers zunimmt. Sie schließen, dass diese Methode einerseits zur Überprüfung der Echtheit, andererseits auch

zur Datierung der Werke genutzt werden kann. Später wurde dieser Ansatz kritisiert, als Jones-Smith und Mathur (2006) zeigten, dass auf einfache Weise Bilder generiert werden können, die die gleichen fraktalen Eigenschaften, aber nicht den gleichen ästhetischen Wert wie Pollocks Bildern aufweisen. Die Arbeiten von Taylor wurden später von Stork (2009b) wiederum verteidigt, der argumentiert, dass auch wenn eine Eigenschaft eines Bildes in Isolation nicht hinreichend für die Erkennung der Echtheit ist, diese in Kombination mit anderen Merkmalen dennoch wertvolle Informationen tragen kann. Shamir (2015) nutzt Merkmale aus der biomedizinischen Bildanalyse (Shamir et al., 2008) und zeigt, dass damit echte Bilder Pollocks von imitierten abgegrenzt werden können. Hughes et al. (2010) nutzen eine spärliche Kodierung der Bilder von Bruegel, um echte Werke von denen von Imitatoren zu unterscheiden. Für Werke des portugiesischen Künstlers Amadeo Souza Cardoso zeigen Montagner et al. (2016), dass eine Analyse der Pinselstriche wie auch der verwendeten Pigmente zur Feststellung der Echtheit eines Bildes verwendet werden kann.

Zusammengefasst können computergestützte Methoden Kunsthistorikern wertvolle Unterstützung bei der Analyse einzelner Bilder oder ganzer Sammlungen bieten. Neben der Erkennung von Fälschungen können auch bis dato unbekannte Einflüsse entschlüsselt, Bilder datiert, oder die Verwendung von Hilfen wie Schablonen aufgedeckt werden. Die rechnergestützte Analyse besitzt dabei gegenüber der menschlichen Einschätzung verschiedene Vorteile, allen voran die Genauigkeit der Analyse. Bei den betrachteten Arbeiten fällt jedoch auf, dass sich oft nur auf bestimmte Künstler und Bilder beschränkt wird. Eine universell einsetzbare Methode, beispielsweise zur Erkennung von Fälschungen, existiert nicht. Dies liegt in dem Umstand begründet, dass sich die jeweiligen Methoden immer auf Eigenheiten der betrachteten Künstler oder Bilder konzentrieren.

1.2 Experimentelle Ästhetik

Die Themengebiete der Experimentellen Ästhetik überschneiden sich teilweise mit denen der bereits vorgestellten Computational Aesthetics, ein Unterschied kann jedoch anhand der Motivation ausgemacht werden. Während es in der Computational Aesthetics oft darum geht, konkrete Anwendungen, beispielsweise für das automatisierte Bewerten von Bildern zu erstellen, interessiert man sich

innerhalb der Experimentellen Ästhetik primär für die der Entscheidung zugrunde liegende Basis der Bewertung eines Bildes oder Objektes. Die Anwendbarkeit tritt hier hinter dem Ziel eines besseren Verständnisses von Ästhetik an sich zurück (Berlyne, 1974; Cela-Conde et al., 2011; Chatterjee und Vartanian, 2014; Shimamura, 2014). Die Experimentelle Ästhetik lässt sich somit eher in der Psychologie verorten, wo sie mit der Arbeit von Fechner (siehe Einleitung dieses Kapitels) ihren Ursprung hat. Bevor anhand einiger exemplarischer Beispiele das Feld näher vorgestellt wird, sollen zunächst einige Grundannahmen innerhalb der Experimentellen Ästhetik betrachtet werden.

1.2.1 Grundannahmen der Experimentellen Ästhetik

Es herrscht generelle Übereinstimmung darüber, dass das visuelle ästhetische Erleben ein hochkomplexes Phänomen ist, welches von den drei Hauptdomänen *Perzeption*, *Kognition* und *Emotion* beeinflusst wird. Diese wiederum siedeln sich auf verschiedenen Ebenen sozialer Interaktion an, und sind in ihrer Ausprägung *universell*, *kulturell* oder *individuell* (Jacobsen, 2006; Marković, 2012; Chatterjee und Vartanian, 2014; Redies, 2015).

Von der Perzeption, oder Wahrnehmung, wird angenommen, dass es sich dabei überwiegend um einen *bottom-up* Prozess handelt, der weitestgehend universell ist und sich zwischen Menschen im Allgemeinen nicht unterscheidet, da er seinen Ursprung in der Evolution des Menschen hat. Es ist jedoch ungeklärt, ob innerhalb des visuellen Systems spezielle Mechanismen existieren, die spezifisch auf ästhetische Reize reagieren. In der Vergangenheit wurde für eine Vielzahl visuell ansprechender Bilder gezeigt, dass diese spezielle statistische Eigenschaften besitzen, die objektiv gemessen werden können. Einige dieser Eigenschaften sollen in den folgenden Abschnitten dieses Kapitels näher betrachtet werden. Bell (1914) vertrat die These, dass Kunst durch eine formale Eigenschaft definiert sei, die er als die „Signifikante Form“ (engl. „significant form“) bezeichnete. Eine bestimmte Anordnung der Linien und Farben sei es, die uns eine „ästhetische Emotion“ erleben lasse, und vereine die Kunstwerke, welche sich über die Jahrhunderte gehalten haben. Der Künstler habe somit die Aufgabe, diese Form in seinem Kunstwerk nachzubilden (Dowling, 2014). Redies (2015) spekuliert, dass diese Eigenschaften eines Reizes ein bestimmtes neuronales Antwortmuster bewirken

und nennt dieses den „beauty-responsive mechanism“, der als Korrelat visueller Präferenz gesehen werden kann. Verschiedene Philosophen, Kunstkritiker, Psychologen und Neurowissenschaftler setzen dem entgegen, dass jeder Stimulus eine ästhetische Erfahrung hervorrufen kann, wenn er in dem richtigen Kontext präsentiert wird. Sie weisen die Annahme zurück, dass formale Kriterien eines Werkes ausschlaggebend sind, vielmehr seien die kunstgeschichtliche Einbettung, die Intention des Künstlers, oder gar dessen Reputation entscheidend (Danto, 1981; Leder et al., 2004; Zeki, 2013; Gopnik, 2014). Diese Faktoren sind nicht universell sondern vielmehr kulturell bedingt und damit wechselnd. Entgegen den in der Wahrnehmung begründeten *bottom-up* Prozessen ist mit deren Integration kognitiver Aufwand verbunden und siedelt die genannten Faktoren daher mehr bei den *top-down* Prozessen an. Perzeptuelle und kognitive Faktoren schließen sich jedoch nicht zwingend gegenseitig aus. Verschiedene Arbeiten kombinieren diese daher für ihre Modelle der ästhetischen Wahrnehmung (vgl. Jacobsen (2006), Locher et al. (2007), Marković (2012), Chatterjee und Vartanian (2014), Kozbelt und Kaufman (2014), Redies (2015) und Shimamura (2014)). Auch die individuelle Erfahrung spielt bei der Betrachtung von Kunst eine wichtige Rolle, sowohl in Form kurzzeitiger Anpassungen, als auch durch längerfristig gebildete Erfahrungen in Bezug auf Kunst. Selbst auf Ebene sehr einfacher Stimuli wurden interindividuelle Unterschiede bezüglich der ästhetischen Präferenz nachgewiesen. Dies betrifft unter anderem Eigenschaften wie die Komplexität (Bies et al., 2016; Güçlütürk et al., 2016; Spehar et al., 2016; Lyssenko et al., 2016), die Farbe (Palmer et al., 2016; Mallon et al., 2014) oder auch die Seitenverhältnisse von Rechtecken (McManus et al., 2010). Schließlich dürfen auch die Emotionen des Betrachters nicht vernachlässigt werden, welche ebenfalls die ästhetische Wertschätzung eines Reizes beeinflussen (Leder et al., 2004; Leder et al., 2014; Silvia, 2005; Silvia, 2014).

Vor dem Hintergrund der Grundkonzepte der Experimentellen Ästhetik wird deutlich, dass rechnergestützte Methoden helfen können, die objektiven Eigenschaften ästhetischer Reize besser zu verstehen. In den in Abschnitt 1.1 vorgestellten Studien, welche anhand von Methoden des maschinellen Lernens ästhetische Präferenzen abbilden, scheint die Grundannahme implizit zu sein, dass objektive Eigenschaften ästhetischer Bilder existieren und diese auch gemessen werden können. Bestätigt wird diese durch den Erfolg, mit dem ästhetische Bewertungen verschiedener Bilder vorhergesagt werden können. Das Wissen um andere Faktoren, die die ästhetische

Präferenz beeinflussen, sollte jedoch zur Vorsicht mahnen, wenn Aussagen über die universelle Gültigkeit bestimmter objektiver Eigenschaften getroffen werden. Ein großes Teilgebiet innerhalb der Experimentellen Ästhetik ist die Untersuchung statistischer Eigenschaften von Kunst. Laut Graham und Redies (2010) birgt die Entschlüsselung dieser Eigenschaften das Potential, auch die visuelle Wahrnehmung des Menschen besser zu verstehen. Die ästhetische Erfahrung wurde mit Fokus auf verschiedene Schwerpunkte experimentell erforscht, seien es die Emotionen bei der Betrachtung von Kunst in großen Museen, die Bewertung der Ästhetik von Kunst unter Laborbedingungen oder auch die Präferenz für einfache Muster. Es wurde deutlich, dass neben den bereits genannten Faktoren der Emotion und Kognition, die die ästhetische Erfahrung beeinflussen, auch die Art der Präsentation einen nicht zu vernachlässigenden Einfluss hat. Wenn unter Laborbedingungen eine bestimmte Eigenschaft isoliert auf deren Wirkung hin untersucht wird, kann nicht zwangsläufig gefolgert werden, dass diese die gleiche Wirkung bei Originalwerken in großen Museen hervorruft, da das Erleben hier sehr viel komplexer abläuft (Jacobsen und Höfel, 2002). Ohne diese Einschränkungen zu vernachlässigen, werden im Folgenden verschiedene Eigenschaften beschrieben, die in Bezug auf Ästhetik untersucht worden sind. Weniger geht es dabei um die vollständige Darstellung der Literatur, als vielmehr darum, einen Überblick über die Erkenntnisse, welche im Bezug auf Kunst und ästhetische Bilder gewonnen wurden, zu geben.

1.2.2 Luminanz- und Farbstatistik

Die Verteilung von Luminanz, Kontrast und Farbe werden den *low-level* Merkmalen zugerechnet. Graham und Field (2008) zeigen, dass die Luminanzstatistik abstrakter Kunst sich von der natürlicher Szenen unterscheidet, während sich natürliche Szenen und traditionelle Kunst bezüglich ihres Spektrums ähneln. Durch eine Manipulation der Luminanzstatistik in natürlichen Bildern, wie beispielsweise künstlerischen Fotografien von Landschaften, zeigen Graham et al. (2016), dass Menschen Bilder mit geringer Schiefe des Histogramms der Helligkeitswerte bevorzugen, bei denen helle und dunkle Anteile ungefähr gleich verteilt sind. Graham und Field (2007) zeigen, dass die Histogramme der Helligkeitswerte von Kunstwerken tatsächlich eine geringere Schiefe aufweisen als beispielsweise Bilder natürlicher

Szenen. Die Autoren spekulieren, dass Künstler den Helligkeitsbereich von Kunstwerken nicht-linear transformieren, um so der Verarbeitung des visuellen Systems entgegenzukommen und den wahrgenommenen Bereich besser auszuschöpfen.

Farbe in Form verschiedener Farbdeskriptoren wird in der Computational Aesthetics oft als Merkmal für die Klassifikation ästhetischer Bilder verwendet (vgl. Abschnitt 1.1.1.1) und auch im Rahmen der Experimentellen Ästhetik wurde Farbe als Merkmal von Bildern untersucht. Pinto et al. (2006) testen verschiedene Beleuchtungssituationen für Gemälde der Renaissance und stellen fest, dass chromatische Vielfalt auf die ästhetische Bewertung eines Bildes einen großen Einfluss hat. Palmer und Schloss (2010) untersuchen die Präferenz von Beobachtern für verschiedene Farben und stellen die Theorie der „ökologischen Valenz“ auf, nach der die Farbpräferenz die durchschnittliche Valenz der mit ihr assoziierten Objekte repräsentiert. Nach dieser Theorie mögen Menschen Blau wegen des Himmels und des Wassers, Braun hingegen wird eher mit Fäulnis verbunden und findet daher weniger Gefallen. Inwiefern sich diese Theorie auf Kunstwerke übertragen lässt, bleibt offen. Mallon et al. (2014) beobachten, dass bestimmte Farbkombinationen in abstrakten Bildern bevorzugt werden und dass diese Präferenz durch Adaptation beeinflusst wird. Nascimento et al. (2017) lassen Beobachter die Farbverteilung von sechs abstrakten und vier gegenständlichen Bildern frei wählen und finden, dass die tatsächlich bevorzugten Farbverteilungen jeweils nahe am Original liegen. Bei dem Vergleich von Farb- und Luminanzkanten finden Leykin und Cutzu (2003), dass Gemälde wesentlich mehr Farbkanten als Luminanzkanten aufweisen, während in der realen Welt, hier untersucht am Beispiel von Fotografien, letztere dominieren. Cutzu et al. (2005) kombinieren Farbe, Kanten und Texturmerkmale von Bildern und trennen unter Verwendung eines einfachen neuronalen Netzes Gemälde und Fotografien. Sie stellen fest, dass Farbe dabei ein kritischer Faktor ist, während Luminanzkanten allein nicht ausreichen, beide Klassen zu trennen. Aragón et al. (2008) untersuchen die Verteilung der Luminanz in van Goghs „Sternennacht“ (1889), sowie anderen Bildern des Künstlers. Sie können zeigen, dass diese der vorhersagbaren Verteilung von Turbulenzen folgt, wie sie erst später vom russischen Mathematiker Andrei Kolmogorow beschrieben wurde. Die Autoren spekulieren, dass so der Eindruck von Unruhe und Bewegung im Bild erzeugt wird.

1.2.3 Komplexität

Die Komplexität beschreibt, wie viele Elemente in einem Bild zu finden sind, wobei es sich bei den Elementen sowohl um einfache Kanten, als auch um Bildinhalte wie Objekte handeln kann. Die Komplexität wird in der Literatur entsprechend von einer Vielzahl unterschiedlicher Maße quantifiziert, sei es als Anzahl von Elementen im Bild (Birkhoff, 1933), die fraktale Dimension (Mureika, 2005; Taylor et al., 2011), GIF-Kompression (Forsythe et al., 2011) oder die Stärke und Zahl der Luminanzkanten (Braun et al., 2013; Redies et al., 2017). In seiner grundlegenden Arbeit zu Ästhetik beschreibt Berlyne (1974) den Zusammenhang von Präferenz und Komplexität als inverse U-Kurve: Während Stimuli von geringer und hoher Komplexität weniger bevorzugt werden, erfahren Stimuli mittlerer Komplexität die höchste Präferenz. Berlyne begründet das mit dem Grad an Erregung, den der Betrachter eines visuellen Stimulus verspürt, und den er gleichzeitig in einem bestimmten Bereich der Erregung zu halten versucht. Andere Autoren begründen die Bevorzugung mittlerer Komplexität damit, dass sich darin unsere evolutionär geprägte Präferenz für bestimmte Landschaften niederschlägt (für eine Übersicht, siehe Forsythe et al. (2011)). Für synthetisch erstellte Bilder zeigen Jacobsen und Höfel (2002) individuelle Unterschiede des bevorzugten Grades an Komplexität, welche hier als die Zahl der Elemente eines Musters gemessen wird. Genauer untersuchen Güçlütürk et al. (2016) diesen Umstand und argumentieren, dass die Hypothese der invertierten U-Kurve ein Artefakt sein könnte, der durch die Mittelung der Ergebnisse vieler Versuchspersonen zustande kommt. In ihrem Experiment identifizieren sie zwei Gruppen: eine, die komplexere Stimuli höher bewertet, und eine, für welche das Gegenteil der Fall ist. Zu einem ähnlichen Ergebnis kommen Rentschler und Barth (2001), die anhand einfacher Muster zwei Gruppen von Versuchsteilnehmern identifizieren, wobei eine höhere Komplexität bevorzugt, die andere diese hingegen ablehnt. Interessanterweise entsteht dieses Präferenz jedoch erst, nachdem die Versuchspersonen gelernt haben, die Muster anhand verschiedener Kriterien zu unterscheiden. Vor dem Erlernen der Unterscheidung hatte die Komplexität der Muster keinen Einfluss auf die Präferenz. Bies et al. (2016) testen die Präferenz für echte Fraktale und finden ebenfalls zwei Gruppen, wobei eine Mehrzahl der Teilnehmer eine höhere fraktale Dimension präferiert und nur ein kleiner Teil Fraktale geringerer Dimensionen bevorzugt.

Rigau et al. (2008) greifen Birkhoffs Idee auf, welche Ästhetik als Interaktion von Ordnung und Komplexität beschreibt und schlagen verschiedene Maße vor, um diese in Bildern von van Gogh, Seurat und Mondrian, basierend auf Ideen der Informationstheorie (Shannon, 1948) zu messen.

1.2.4 Symmetrie, Balance und die Drittel-Regel

Symmetrie ist eine Eigenschaft vieler natürlicher und künstlicher Muster. Sie wird von Menschen sehr schnell entdeckt und beeinflusst unter anderem die Detektion von Objekten, die Aufmerksamkeit, Augenbewegungen und physiologische Erregung (Locher und Nodine, 1989). Für menschliche Gesichter zeigen Grammer und Thornhill (1994), dass Symmetrie einen positiven Einfluss auf die Bewertung der Attraktivität hat. Für einfache geometrische Formen haben eine Vielzahl von Studien eine Korrelation von Symmetrie und ästhetischer Bewertung nachgewiesen (Jacobsen und Höfel, 2002; Westphal-Fitch et al., 2013; Rampone et al., 2016; al-Rifaie et al., 2017). Gleichzeitig ist der Einfluss von Symmetrie auf die ästhetische Bewertung von Fotografien und Kunstwerken weniger eindeutig. Einem Besucher eines Kunstmuseums wird schnell auffallen, dass einfache Symmetrien (Reflektion, Translation oder Rotation) kein generelles Prinzip der Komposition in traditioneller Kunst sind, gleichwohl auch hier Symmetrie die Aufmerksamkeit des Betrachter erwecken kann (Locher und Nodine, 1989). Osborne (1986) argumentiert, dass Symmetrie in der Kunst nicht nur als nicht notwendig gesehen, vielmehr sogar gemieden wird. Little (2014) vergleicht die Präferenz für Symmetrie in Kunst und Gesichtern und findet, dass diese domänenspezifisch ist. Während Symmetrie in weiblichen Gesichtern präferiert wird, ist sie in abstrakter Kunst weniger bevorzugt. In Kapitel 3 wird ein Verfahren vorgestellt, die Spiegelsymmetrie natürlicher Bilder messen zu können. So soll ermöglicht werden, diese Zusammenhänge genauer zu untersuchen.

Das Konzept der Balance steht in enger Beziehung zur Symmetrie. Im Gegensatz zu dieser wird sie jedoch als wichtiger Faktor der Komposition angesehen, der die Ästhetik von Fotografien, Kunstwerken oder auch die abstrakter Muster beeinflusst (McManus et al., 1985; Gershoni und Hochstein, 2011; Jahanian et al., 2015). Nach der Gestalttheorie Rudolf Arnheims (Arnheim, 1954) ist ein Bild balanciert, wenn dessen Schwerpunkt auf eine der Hauptachsen (vertikal, horizontal oder diagonal)

des Bildes fällt. Der Theorie Arnheims folgend berechnen McManus et al. (2011) die Balance als den Schwerpunkt der Helligkeitswerte im Bild. Ihre Ergebnisse bestätigen Arnheims Theorie der Balance jedoch nicht. Als eine mögliche Interpretation diskutieren die Autoren, dass die Helligkeitswerte sich nicht linear in die wahrgenommene Gewichtung übersetzen lassen. Anstelle von Helligkeitswerten berechnen Jahanian et al. (2015) den Schwerpunkt salienter Merkmale eines Bildes und können so anhand von insgesamt 120 000 hoch bewerteter Bilder Arnheims Theorien stützen. Abeln et al. (2016) vergleichen die Salienzschwerpunkte von durch Probanden bevorzugte Ausschnitte eines Bildes und finden, dass deren Salienz balancierter als die vermiedener Ausschnitte ist.

Eine in der Fotografie populäre Praxis ist die Komposition eines Bildes nach der sogenannten *Drittel-Regel*. Laut ihr kann es den ästhetischen Wert eines Bildes heben, wenn das Hauptmotiv auf einer der Linien platziert wird, die das Bild horizontal oder vertikal dritteln. Viele der in Abschnitt 1.1.1.1 vorgestellten Arbeiten implementieren diese Regel als Merkmal. In der Experimentellen Ästhetik konnte die Drittel-Regel bisher zumindest für Kunst und eine Auswahl von Fotografien (Amirshahi et al., 2014), sowie für sogenannte „Selfies“ (Bruno et al., 2014) experimentell nicht bestätigt werden.

1.2.5 Fourier-Spektrum

Graham und Field (2007) und Redies et al. (2007b) vergleichen das Fourierspektrum natürlicher Szenen mit dem von Kunst und zeigen, dass beide einen ähnlichen Anstieg der Amplitude über die Frequenz aufweisen. Graham und Field (2008) vergleichen die Fourier-Spektren auch für östliche Kunst, finden dabei jedoch Unterschiede zu den natürlichen Szenen. Die Autoren führen dieses Ergebnis auf die unterschiedlichen Bildhintergründe der getesteten Bilder zurück. Koch et al. (2010) vergleichen Kunst mit Cartoons, Comics und Mangas und finden auch hier Ähnlichkeiten mit natürlichen Szenen bezüglich des Frequenzspektrums. Viele Bilder ohne künstlerischen Anspruch hingegen, beispielsweise die einfacher Objekte oder Pflanzenmuster, besitzen deutlich andere Anstiege der Amplitude über die Frequenz. Redies et al. (2007a) vergleichen das Frequenzspektrum gezeichneter Portraits mit Fotografien von Gesichtern und können zeigen, dass die Statistik der gezeichneten Bilder eher der komplexer natürlicher Szenen ähnelt.

Mather (2014) vergleicht die Anstiege von 31 Kunstwerken mit sehr ähnlichen Fotografien und zeigt, dass Künstler das Spektrum verglichen mit Fotografien in einen engeren Bereich verschieben. Er wirft die Frage auf, ob der Grund dafür im visuellen System zu finden sei. Menschen scheinen bestimmte Frequenzbereiche randomisierter Muster zu bevorzugen, die einen Anstieg ähnlich dem natürlicher Szenen aufweisen (Menzel et al., 2015), zeigen aber auch hier große interindividuelle Unterschiede (Spehar et al., 2016). Es wurde gezeigt, dass als unangenehm empfundene Bilder von dem linearen Abfall der Frequenzen abweichen, welcher für Kunst und als angenehm empfundene Bilder wie etwa natürliche Szenen gemessen wurde (Fernandez und Wilkins, 2008; O'Hare und Hibbard, 2011). Der Anstieg korreliert mit anderen Maßen für Komplexität (Redies et al., 2017). So deutet ein flacherer Abfall dabei auf stärkere hohe Frequenzen im Bild hin, die besonders bei einem hohen Detailgrad zu beobachten sind.

1.2.6 Fraktale Dimension und Selbstähnlichkeit

Eine Linie hat eine Dimension von 1, während eine Fläche eine Dimension von 2 hat. Ein echtes Fraktal, wie beispielsweise die Kochkurve, ist begrenzt, hat aber eine unendliche Länge und ist dennoch keine Fläche. Ihre *fraktale Dimension* wird mit ≈ 1.26 angegeben. Eine numerische Annäherung der fraktalen Dimension für Bilder ist mithilfe der *Box-Counting*-Methode möglich (Mandelbrot und Pignoni, 1983). Im Mittelpunkt steht die Idee, ein Bild zunehmend feingliederiger zu unterteilen und dabei die durch ein Muster belegten Teilregionen in Beziehung zu der Größe der Teilregionen selbst zu setzen. Der Grenzwert des Verhältnisses für unendlich kleine Teilregionen wird als fraktale Dimension verstanden. An dieser Stelle sei angemerkt, dass nicht geschlossen werden kann, dass ein Bild eine fraktale Struktur besitzt, wenn die *Box-Counting*-Dimension größer ist als 1. Für Bilder handelt es sich bei der *Box-Counting*-Dimension nur um eine numerische Annäherung auf diskreten Werten, welche zwangsläufig in ihrer Auflösung begrenzt sind. In diesem Kapitel wurde bereits kurz auf die Arbeiten Taylors zu dem amerikanischen Künstler Jackson Pollock eingegangen (Taylor, 2002). Die darin durchgeführte numerische Analyse der fraktalen Dimension von Bildern hat weitere Studien hervorgerufen, die der Frage nachgehen, wie Menschen auf die fraktale Dimension eines Bildes reagieren (Taylor et al., 2011). Taylor und Spehar (2016) argumentieren, dass

ein mittlerer Bereich der fraktalen Dimension von 1.3–1.5 eine besondere Rolle bezüglich der Wahrnehmung einnimmt, da dieser den Bereich repräsentiert, an den unser visuelles System durch natürliche Muster angepasst ist. Hagerhall et al. (2004) zeigen, dass ein Zusammenhang zwischen Präferenz und der fraktalen Dimension von Landschaftssilhouetten besteht. Auch Forsythe et al. (2011) setzen die fraktale Dimension mit der Präferenz für gegenständliche und abstrakte Kunst in Beziehung. Die psychologische Wirkung von Fraktalen wurde ebenfalls untersucht. So zeigt Taylor (2006) beispielsweise, dass Fraktale Stress reduzieren können und Joye (2007) argumentiert, dass diese die Wirkung von Architektur positiv verbessern können. Auch hier wurden jedoch interindividuelle Unterschiede der Präferenz festgestellt, allgemeine Schlussfolgerungen sind somit schwierig (Aks und Sprott, 1996). Auch muss hier beachtet werden, dass von der fraktalen Dimension nicht auf eine bestimmte Bildstruktur geschlossen werden kann und Bilder gleicher Dimension durchaus sehr verschiedene Merkmale aufweisen können (vgl. dazu Barth et al. (1993)). Ein Vergleich der fraktalen Dimension von Bildern und deren Auswirkung auf die Präferenz ist also nur auf strukturell ähnlichen Bildern sinnvoll.

Die fraktale Struktur eines Bildes kann auch auf andere Weisen gemessen werden. Amirshahi et al. (2012) schlagen eine Methode vor, die die Selbstähnlichkeit von Bildern quantifiziert. Basierend auf einem Vergleich der Kanten global im Bild mit denen auf lokaler Ebene können Aussagen darüber getroffen werden, inwiefern sich im Bild Strukturen des Großen auch im Kleinen wiederholen. Braun et al. (2013) vergleichen Selbstähnlichkeit verschiedener Bildkategorien und finden für Kunst eine mittlere Selbstähnlichkeit, die höher ist als die von Bildern einfacher Objekte, jedoch geringer als die verschiedener Pflanzenmuster. Einschränkend ist festzustellen, dass die vorgestellten Arbeiten zur Berechnung der Selbstähnlichkeit diese jeweils nur auf einer in Grauwerte konvertierten Version eines Bildes berechnen. Da insbesondere für Kunst festgestellt wurde, dass Farbe ein nicht außer Acht zu lassender Faktor darstellt, wird in Kapitel 4 eine Methode vorgestellt, mit deren Hilfe die Selbstähnlichkeit auch für Farbbilder berechnet werden kann (Brachmann und Redies, 2017b).

Die oft wiederkehrende Erkenntnis, dass Kunst bezüglich verschiedener statistischer Eigenschaften den natürlichen Mustern gleicht, wird mitunter als Hinweis interpretiert, dass Kunstwerke bewusst oder unbewusst so geschaffen werden, dass

sie gleich den natürlichen Mustern von unserem visuellen System effizient, im Sinne von sparsam bezüglich metabolischer Ressourcen, verarbeitet werden können (Redies, 2007a; Renoult et al., 2016). Andere argumentieren, dass die fluente Verarbeitung der Reize ausschlaggebend ist (Reber et al., 2004). Die neurologischen Grundlagen, welche diese Thesen stützen oder widerlegen können, sind bis heute ungeklärt.

1.2.7 Orientierungen von Helligkeitsgradienten

Für traditionelle Kunst zeigen Redies et al. (2012) und Braun et al. (2013), dass die Orientierungen der Helligkeitsgradienten größtenteils gleich verteilt sind, die Anisotropie der Gradientenrichtungen ist somit gering. Bilder natürlicher Muster wie etwa Flechten oder Wolken zeigen ebenfalls eine geringe Anisotropie der Kantenrichtungen, während diese in vielen Bildern mit menschengemachten Mustern wie Objekten, Fassaden oder Gebäuden hoch ist (Redies et al., 2012). Die genannten Studien betrachten die Orientierungen der Kanten absolut und beziehen diese auf das Gesamtbild. In Kapitel 2 werden wir ein Verfahren vorstellen, welches es erlaubt, die relativen Beziehungen der Orientierungen von Helligkeitsgradienten untereinander genauer zu untersuchen.

Bezogen auf die ästhetische Präferenz wurde gezeigt, dass Menschen runde Kanten denen mit scharfen Übergängen vorziehen (Gómez-Puerto et al., 2015). Diese Präferenz wurde nicht nur interkulturell, sondern auch bei Menschenaffen gefunden (Munar et al., 2015). Bar und Neta (2006) begründet dies damit, dass gefährliche Objekte als Gefahr erkannt werden, sich die *low-level* Eigenschaften derselben aber auch auf ungefährliche Objekte übertragen, und so eine generelle Abneigung gegen spitze Winkel die Folge ist. Bertamini et al. (2016) untersucht diese These experimentell und kann anhand verschiedener Experimente nicht bestätigen, dass insbesondere spitze Winkel gemieden werden und die Präferenz für runde Kanten nur eine Folge dessen ist, vielmehr werden runde Formen an sich bevorzugt.

1.3 Zusammenfassung

In den vergangenen Jahren sind auf dem Gebiet des maschinellen Sehens eine Fülle von Methoden entwickelt worden, welche auch erfolgreich auf die Analyse von

Fotografie und digital reproduzierten Kunstwerken angewendet wurden. In diesem Kapitel wurde der Fortschritt auf diesem Gebiet, welches unter dem Sammelbegriff der Computational Aesthetics zusammengefasst wird, nachgezeichnet und eine Auswahl von wichtigen Arbeiten vorgestellt. Dabei wurde einerseits genauer auf die Vorhersage der ästhetischen Bewertungen von Bildern eingegangen. Bezogen auf Kunstwerke wurden andererseits die Themen der automatisierten Erkennung von Stil und Künstler dargestellt und auch Gebiete, die Kunstwerke aus einer kunsthistorischen Perspektive analysieren, näher betrachtet.

Im Allgemeinen werden Methoden zur Vorhersage von ästhetische Bewertungen anhand der erreichten Klassifikationsrate beurteilt. Wenn dabei dieselbe Datenbasis zugrunde liegt, können Algorithmen einfach miteinander verglichen und die beste Methode gefunden werden. Die Entwicklung hat gezeigt, dass die Klassifikation durch die Integration semantischer Merkmale zusätzlich zu den formalen verbessert werden kann. Bedenkt man, dass beispielsweise helle Farben in Bildern von Früchten bevorzugt werden, während dunkle Farben eine Straßenszene bei Nacht besser wirken lassen, leuchtet diese Erkenntnis ein. Moderne Methoden wie tiefe neuronale Netze ermöglichen die Beachtung dieser Zusammenhänge zunehmend, mehr noch, sie finden eine Gewichtung formaler und semantischer Merkmale während des Trainings selbst. Gleichzeitig sind diese Ansätze im Ergebnis unzugängliche Methoden, deren Entscheidungsgrenzen sich einer einfachen Interpretation des Gelernten entziehen. Vom Standpunkt der konkreten Anwendung her muss dies nicht zwingend einen Nachteil darstellen. Diese Systeme können eingesetzt werden, wenn große Mengen an Bildern automatisiert nach ihrem ästhetischen Wert beurteilt werden sollen, etwa in sozialen Medien, wo Benutzer Bilder miteinander teilen und anhand einer automatisierten Analyse eine Auswahl höherwertiger Bilder zusammengestellt werden kann.

Auch auf dem Gebiet der Experimentellen Ästhetik, welches innerhalb der Psychologie auf eine lange Tradition zurückblickt und zunehmend auch in der Neurowissenschaft Verbreitung findet, wurden rechnergestützte Methoden erfolgreich eingesetzt. Hier liegt der Fokus nicht darauf, ästhetische Bilder von anderen automatisiert zu unterscheiden oder den Urheber eines Bildes zu identifizieren, vielmehr ist das Ziel, ein besseres Verständnis der zugrunde liegenden Eigenschaften der als ästhetisch bewerteten Bilder oder Objekte zu erlangen. Beispielsweise wurde gezeigt, dass viele traditionelle Kunstwerke statistische Eigenschaften natürlicher

Szenen teilen, was zu der Hypothese führte, dass die Kodierung beider Arten von Reizen im visuellen System des Menschen einander ähnelt.

Mit der modernen Entwicklung des tiefen Lernens ist der Wissensaustausch zwischen Computational Aesthetics und der Experimentellen Ästhetik erschwert worden. Während in den frühen Tagen der Computational Aesthetics die Erkenntnisse aus der Experimentellen Ästhetik genutzt wurden, um spezifische Bildmerkmale wie Unschärfe oder Regeln zur Komposition von Bildern zu implementieren, ist man mit tiefen Netzen auf diesen Schritt nicht länger angewiesen. Aus Sicht der Experimentellen Ästhetik geht damit die Möglichkeit verloren, Hypothesen direkt zu testen. Während es mit spezifischen Merkmalen beispielsweise möglich war, zu quantifizieren, welchen Einfluss die Drittel-Regel auf die ästhetische Bewertung eines Bildes hat, werden derlei Aspekte durch tiefe neuronale Netze unsichtbar für den Anwender bewertet. Auch wenn aktuellere Arbeiten die Eigenschaften der Filterantworten tiefer Netze genauer untersuchen, bleibt zu konstatieren, dass wir von einem genauen Verständnis der Filterantworten und deren Zusammenhang mit ästhetischen Bewertungen noch weit entfernt sind.

Zusammengefasst liegt großes Potential in der erneuten Zusammenführung beider Gebiete, wobei moderne Methoden der Bildanalyse helfen können, sowohl die ästhetischen Urteile von Menschen, als auch die statistischen Eigenschaften von Kunst besser zu verstehen. Als ein Beispiel zeigen wir in Kapitel 5 einen Ansatz, wie die Filterantworten tiefer neuronaler Netze genutzt werden können, für Kunst spezifische Eigenschaften zu entdecken und daraus neue Hypothesen abzuleiten.

Kapitel 2

Entropie orientierter Gradienten

Auf den unteren Ebenen der Reizverarbeitung im visuellen System findet die Informationsverarbeitung vor allem lokal statt, indem viele Millionen Photorezeptoren mit wenigen Millionen Ganglienzellen verschaltet sind. Die so entstehenden rezeptiven Felder dienen unter anderem dazu, die eintreffende Flut an Informationen zu reduzieren (Wurtz und Kandel, 2000a). Im primären visuellen Kortex sprechen Orientierungssäulen auf orientierte Gradienten an, die anhand konzentrischer Merkmale auf der Retina detektiert wurden. Die Antworten im visuellen Kortex werden jedoch auch durch Reize moduliert, die sich weit über die Größe einfacher rezeptiver Felder hinaus erstrecken. Beispielhaft dafür ist das aus der Gestaltpsychologie bekannte „Gesetz der guten Fortsetzung“, welches besagt, dass bei der Betrachtung von Bildern für unterbrochene Linien der jeweils plausibelste Verlauf angenommen wird. Field et al. (1993) erklären dieses Phänomen mit einem „Assoziationsfeld“, welches Informationen benachbarter rezeptiver Felder zusammenfasst. Während die visuelle Verarbeitung zunächst als ein unidirektionaler Prozess verstanden wurde, ist heute bekannt, dass die Interaktionen innerhalb des visuellen Kortex sehr viel komplexer sind als zunächst angenommen (Spillmann et al., 2015). So beeinflussen sich rezeptive Felder nicht nur über weite Distanzen, sondern auch von höheren auf darunter liegende Schichten. Diese Struktur hat sich im Laufe der Evolution herausgebildet und ist an die Statistik der natürlichen Umgebung angepasst (Coppola et al., 1998; Geisler, 2008). Die Verarbeitung visueller Information in höheren kortikalen Arealen wird zunehmend global, indem lokale Information über größere Reizareale zusammengefasst wird.

In Abschnitt 1.2.1 wurde bereits das von Bell (1914) vorgeschlagene Konzept der „Signifikanten Form“ von Kunst vorgestellt, welches besagt, dass Kunst bestimmten, unbekannten Regeln der Komposition folgt, welche beispielsweise die Anordnung der Elemente eines Bildes betreffen. Ist dieses zutreffend, so stellt sich die Frage, ob sich das auch auf die Interaktionen über weite Distanzen des visuellen Feldes auswirkt. In diesem Kapitel wird untersucht, wie sich Kanten in Bildern relativ zueinander über verschiedene Distanzen verhalten. Insbesondere ist dabei von Interesse, wie sich Kunst in diesem Aspekt zu anderen Bildkategorien verhält. Die Art der Untersuchung reiht sich gleichzeitig in eine Reihe vieler weiterer ein, die die statistischen Eigenschaften von Kunst analysieren (vgl. dazu Abschnitt 1.2). So zeigen Redies et al. (2007b) und Braun et al. (2013), dass die Anisotropie für Kunst verglichen mit menschengemachten Szenen geringer ausfällt, jedoch höher als für viele natürliche Muster. Die Anisotropie wird dabei jeweils als die Varianz des Histogramms über verschiedene Richtungen der Helligkeitsgradienten berechnet. Die hier vorgestellte Analyse geht darüber hinaus, indem Kantenrichtungen relativ zueinander betrachtet werden. Primär ist dabei weniger von Interesse, zu welchen Anteilen bestimmte Kantenrichtungen im Bild repräsentiert sind, als vielmehr, wie sich Kanten zueinander verhalten. Es wird dabei der Frage nachgegangen, ob ausgehend von einer gegebenen Kante eine Vorhersage getroffen werden kann, wie die Kanten in deren Umfeld orientiert sind. Das genaue Verfahren wird im folgenden Abschnitt näher beschrieben.

2.1 Material und Methoden

Die vorgestellte Analyse beruht auf der Berechnung eines Histogramms, welches die relative Beziehungen von Helligkeitsgradienten quantifiziert. Dessen Erstellung und die anschließende Berechnung der Entropie der Kantenrichtungen sollen im Folgenden beschrieben werden. Im Anschluss daran werden die getesteten Bildkategorien vorgestellt.

2.1.1 Berechnung der Entropie orientierter Gradienten

Jedes Eingabebild wird zunächst auf eine maximale Zahl von 120 000 Pixeln herunterskaliert, ohne dabei dessen Seitenverhältnis zu verändern. Farbbilder wer-

den entsprechend der ITU-R 601-2 Luminanztransformation in Graustufenbilder umgewandelt, wobei die Farbkanäle der wahrgenommenen Helligkeit wie folgt gewichtet sind:

$$L = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \quad (2.1)$$

L ist hierbei die errechnete Graustufe, R , G und B repräsentieren die Farbkanäle Rot, Grün und Blau. Auf jedes Bild wird anschließend ein Satz 24 verschieden orientierter Gaborfilter der Winkel γ angewendet, welche in Kombination eine vollständige Rotation um 2π ergeben. Die Gaborfilter werden dabei nach folgender Vorschrift erzeugt:

$$g_\gamma(x, y) = e^{-(x^2+y^2)/(2\sigma^2)} \cos\left(\frac{\pi}{4}(x \sin(\gamma) + y \cos(\gamma)) + \frac{\pi}{2}\right) \quad (2.2)$$

wobei $\sigma = 1.699$ und $\gamma \in \left\{2\pi \frac{n}{24} \mid n \in \{0, 1, \dots, 23\}\right\}$. Die Filter sind so angelegt, dass diese einen Nullpunkt in $(0, 0)$ haben und am stärksten auf Helligkeitsgradienten im Bild ansprechen. Für jeden Pixel eines Bildes bestimmt der Filter g_γ mit der höchsten Antwort die Orientierung γ an der entsprechenden Stelle im Bild. Bei der Analyse wird sich auf die 10 000 Pixel eines Bildes mit den höchsten Ausschlägen beschränkt, wobei Pixel mit einem Abstand von bis zu 15 Pixeln zum Bildrand keine Beachtung finden. Beispiele der resultierenden Kantenbilder sind in Abbildung 2.1 gezeigt.

Die Idee des paarweisen Vergleiches der Kantenorientierungen stammt von Sigman et al. (2001) und wird hier wie folgt adaptiert. Jede beibehaltene Kante des Bildes wird mit jeder anderen Kante im Bild verglichen, wobei deren Beziehung durch drei Parameter bestimmt ist: deren Abstand zueinander d , deren relative Lage α , sowie deren relative Orientierung θ (vgl. dazu Abbildung 2.2a). Vereinfacht gesprochen berechnen wir für jede relative Lage α und d ein Histogramm, in welchem die Produkte der relativen Orientierungen zweier Kanten aufsummiert werden. Im folgenden bezeichnen wir diese als θ -Histogramme. Abbildung 2.2c veranschaulicht die Erzeugung der Histogramme am Beispiel von drei verschiedenen Kanten und zeigt darüber hinaus eine Visualisierung der θ -Histogramme für das gesamte Bild, wobei die dominante Richtung, berechnet als Durchschnitt aller Richtungen

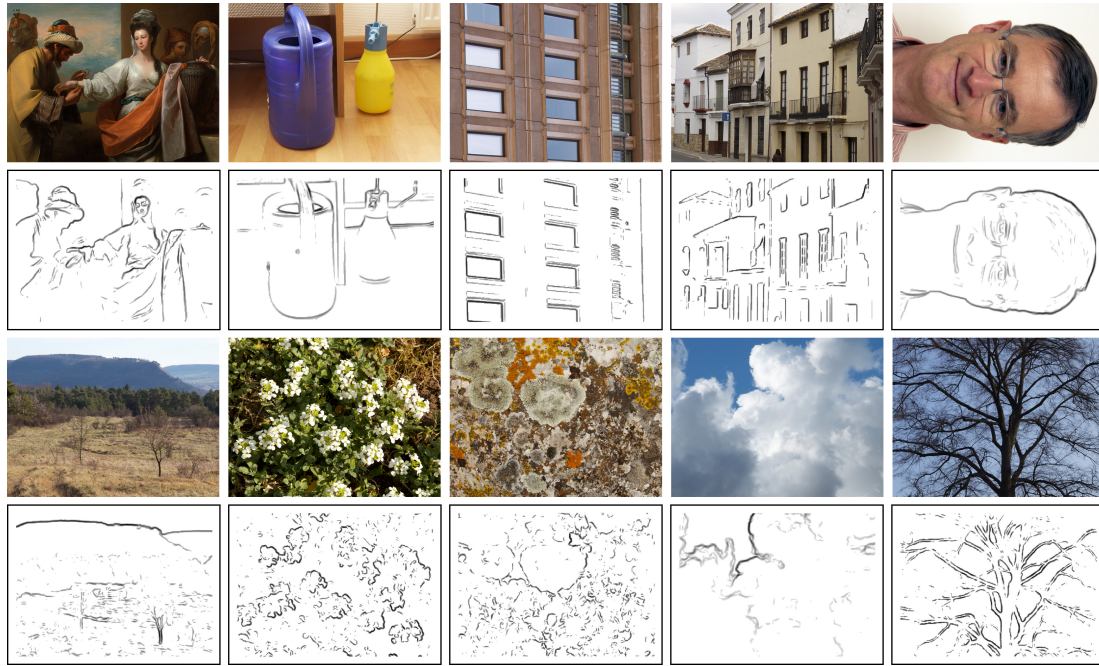


Abbildung 2.1: Beispielbilder mit den zugehörigen Kantenbildern, welche in der hier beschriebenen Analyse genutzt werden. Oben, v.l.n.r. je ein Bild der folgenden Datensätze: *JenAesthetics*, *Objects*, *Facades*, *Urban Scenes*, *Faces*. Unten, v.l.n.r.: *Large Vistas*, *Plant Patterns*, *Lichen*, *Clouds*, *Branches*. Der Titel und der Urheber aller in dieser Arbeit abgebildeten Kunstwerke des *JenAesthetics*-Datensatzes sind in Tabelle A.1 des Anhangs gelistet.

eines Histogrammes, in Farbe kodiert ist. Beispiele einzelner θ -Histogramme verschiedener Bilder sind in Abbildung 2.2d gezeigt. In der Visualisierung in Polardarstellung (vgl. Abbildung 2.2c) zeigt die obere Hälfte des Kreises alle Kanten, die vor der Referenzkante liegen, die untere Hälfte hingegen alle, die hinter dieser verortet sind. Kanten auf der durch die Referenzkante definierten Linie liegen auf der Geraden, welche beide Hälften teilt. Wir unterscheiden 500 Schritte für d , 48 Winkel für α und 24 Winkel für θ , das finale Histogramm eines jeden Bildes hat somit insgesamt $500 \times 48 \times 24$ Einträge. Für jede Position d und α wird das θ -Histogramm normalisiert.

Als Maß für die Gleichförmigkeit eines θ -Histogrammes X berechnen wir dessen Shannon-Entropie $H(X)$, welche durch folgende Gleichung gegeben ist:

$$H(X) = - \sum_{i=1}^N p(x_i) \log_2 p(x_i) \quad (2.3)$$

N repräsentiert dabei die Zahl der unterschiedenen Klassen eines Histogrammes und ist in unserem Fall 24, was der Zahl der unterschiedenen Kantenrichtungen entspricht. Ein hoher Grad an Uniformität eines θ -Histogrammes bedeutet, dass für den gegebenen Abstand d und Winkel α relativ zur Referenzkante alle anderen Kantenrichtungen ungefähr gleich in Orientierung und Stärke verteilt sind, für das θ -Histogramm wird eine hohe Entropie gemessen. Ein Beispiel für ein θ -Histogramm mit sehr hoher Gleichförmigkeit ist in Abbildung 2.2d (4. v.l., Histogramm eines Flechtenbildes) gegeben. Sind alle Kantenrichtungen exakt gleich verteilt, ergibt sich für ein Histogramm das Entropiemaximum von $-\log_2 \frac{1}{24} \approx 4.585$. Sobald bestimmte Kantenrichtungen innerhalb eines θ -Histogrammes stärker vertreten sind als andere (vgl. Abbildung 2.2d, 3. v.l.), fällt dessen Entropie. In der Definition der Entropie wird $0 \cdot \log_2 0 = 0$ gesetzt, was dem Grenzwert $\lim_{x \rightarrow 0} x \log_2 x$ entspricht. Ein theoretischer Wert von 0 als Untergrenze ist somit möglich, wenn nur eine Kantenrichtung beobachtet wird. In der Einleitung dieses Kapitels wurde bereits erwähnt, dass Kunst einen geringen Grad an Anisotropie besitzt, was bedeutet, dass die Kantenrichtungen in Kunst über das gesamte Bild relativ gleich verteilt sind. Als Vergleich mit diesem Ergebnis wird neben der beschriebenen Entropie der θ -Histogramme, im Folgenden vereinfachend bezeichnet als „Entropie zweiter Ordnung“, zusätzlich auch die Entropie des normalisierten Histogramms über alle Kanten eines Bildes berechnet (nachfolgend als „Entropie

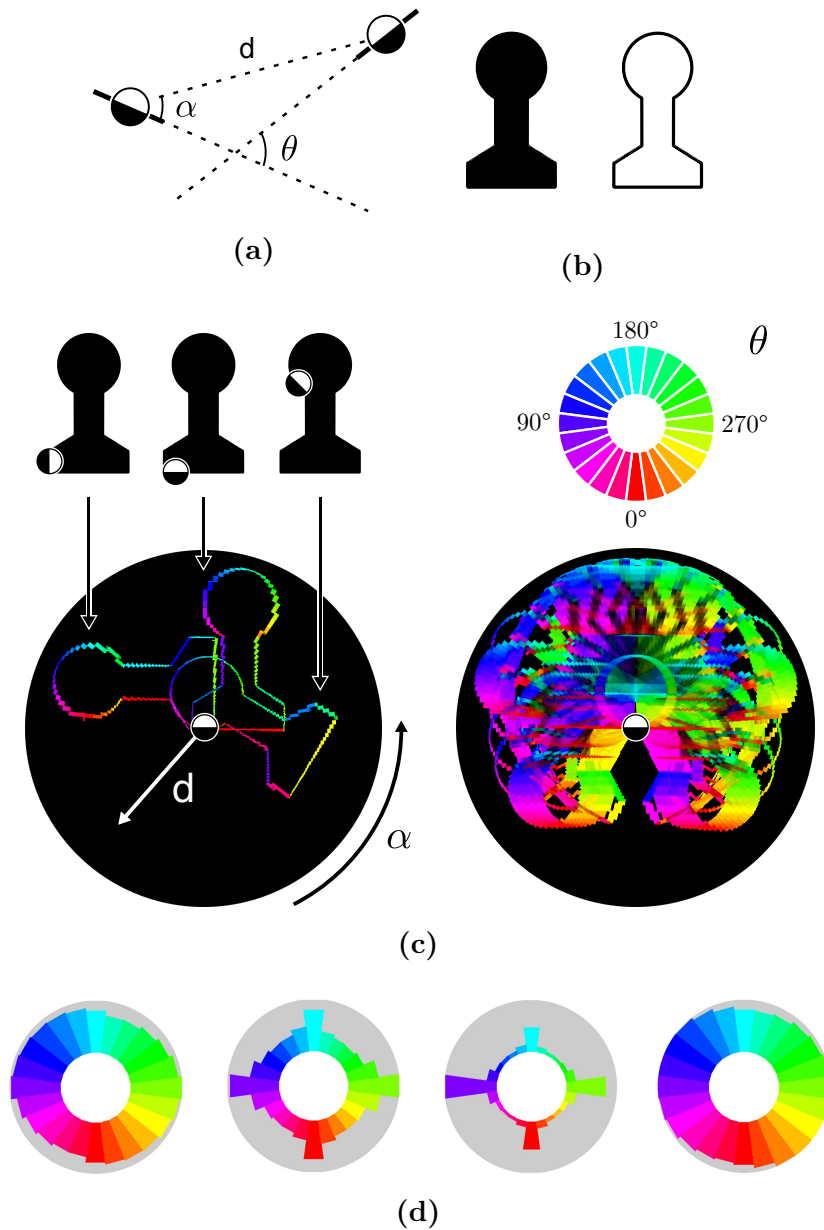


Abbildung 2.2: (a) Die Beziehung zweier Kanten ist durch deren Abstand d , deren relativen Winkel α und deren relative Orientierung θ definiert. (b) Darstellung einer Schachfigur (links) und deren Kantenbild (rechts). (c) Visualisierung der durchschnittlichen Richtungen der einzelnen θ -Histogramme nach dem Vergleich dreier ausgesuchter (links), sowie aller Kanten (rechts) mit allen anderen. Jeder Punkt auf dem Kreis repräsentiert ein θ -Histogramm in Abhängigkeit von d und α . (d) Beispiele für θ -Histogramme für $d = 80$ und $80^\circ \leq \alpha < 97.5^\circ$ für ein Kunstbild, ein Bild eines Objektes, das einer Fassade und das einer Flechte (v.l.n.r.).

erster Ordnung“ bezeichnet). Sofern nicht anders angegeben, bezieht sich der Begriff Entropie nachfolgend immer auf die Entropie zweiter Ordnung, welche im Mittelpunkt unserer Untersuchung steht.

Als ein zusätzliches Maß berechnen wir anhand der θ -Histogramme den Grad an *Parallelismus* eines Bildes, welcher beschreibt, zu welchem Grad Kanten eines Bildes die gleiche Orientierung ungeachtet ihrer Polarität besitzen. Ein Spezialfall von Parallelismus ist *Kolinearität*, die den Umstand bezeichnet, dass zwei Kanten zusätzlich zur gleichen Ausrichtung auch auf der gleichen, fortgesetzten Linie liegen. Parallelismus schlägt sich in einem θ -Histogramm derart nieder, dass nur Kanten mit den relativen Richtungen 0° und 180° beobachtet werden. Wir messen Parallelismus demnach als den Abstand eines gegebenen θ -Histogramms von einem Histogramm, welches für 0° und 180° jeweils einen Wert von 0.5 aufweist. Als Maß für den Abstand zweier Histogramme nutzen wir den *Histogram Intersection Kernel* (HIK; Barla et al. (2003)), der für zwei normalisierte Histogramme $\mathbf{h}$ und $\mathbf{h}'$ mit jeweils k Klassen wie folgt definiert ist:

$$K(\mathbf{h}, \mathbf{h}') = \sum_{i=1}^k \min(h_i, h'_i) \quad (2.4)$$

Für normalisierte Histogramme ist dieser beschränkt zwischen 0 und 1, wobei 1 eine Gleichheit der Histogramme anzeigt und 0 eine maximale Verschiedenheit. Für die θ -Histogramme kann die dominante Orientierung als der zirkuläre Mittelwert der Einträge bestimmt werden. Die Einträge werden dafür als Vektoren entsprechend des Winkels und der Länge interpretiert und deren Mittel bestimmt. Je länger der resultierende Vektor, desto stärker herrscht eine bestimmte Orientierung in einem Histogramm vor. Eine Vektorlänge nahe 0 zeigt hingegen an, dass sich gegenüberliegende Orientierungen ungefähr gleich im Histogramm vertreten sind und sich gegenseitig aufheben.

2.1.2 Datensatz zur Evaluation

In unserer Analyse werden insgesamt zehn verschiedene Bildkategorien getestet. Neben einem Datensatz von Gesichtern (*Faces*; Martinez (1998)) nutzen wir für die Analyse von Kunst den *JenAesthetics*-Datensatz (Amirshahi et al., 2013a), welcher insgesamt 1629 Ölgemälde verschiedener Kunststile westlicher Kunst enthält. Darunter finden sich Bilder der Renaissance, des Barock, der Romantik, des

Realismus sowie des Impressionismus, datiert vom 15. bis in das 21. Jahrhundert. Neben den beiden genannten nutzen wir acht weitere Bildkategorien, die sich in natürliche und nicht-natürliche Muster unterteilen lassen. Im folgenden werden für Bilder, welche Szenen und Muster der Natur zeigen, die Begriffe „natürliche Muster“ oder „natürliche Szenen“ genutzt, während der Begriff „natürliche Bilder“ stellvertretend alle jene Bilder bezeichnet, welche reale Objekte zeigen, ungeachtet davon, ob diese menschlichen oder natürlichen Ursprungs sind. Dennoch heben sich diese Bilder beispielsweise von weißem Rauschen ab, ein generiertes Bild mit zufällig gesetzten Pixelwerten wird nahezu nie einem natürlichen Bild ähnlich sehen (Chandler und Field, 2007). Unter den nicht-natürlichen Mustern finden sich Fotos von Objekten (*Objects*), Fassaden (*Facades*) und Städteszenen (*Urban Scenes*). An natürlichen Kategorien finden sich Bilder von Landschaften (*Large Vistas*), Pflanzenmuster (*Plant Patterns*), Flechten (*Lichen*), Wolken (*Clouds*) und Ästen (*Branches*). Eine Übersicht über die genannten Kategorien sowie ein Verweis auf deren jeweilige Erstveröffentlichung findet sich in Tabelle 2.1, Beispielbilder der insgesamt zehn Kategorien sind in Abbildung 2.1 gezeigt. Den ersten Veröffentlichungen der Datensätze folgend werden fortlaufend ihre englischen Bezeichnungen verwendet. Die Auswahl der Datensätze wurde bewusst so getroffen, dass diese Bilder mit einer hohen Diversität an Bildmerkmalen wie Selbstähnlichkeit, Anisotropie oder Komplexität vereinen. Insbesondere ist dabei von Interesse, ob sich diese Diversität auch in Unterschieden bezüglich der Entropie der Kantenrichtungen widerspiegelt.

2.2 Ergebnisse

2.2.1 Paarweise Statistik von Kantenorientierungen

Für jedes Bild der in 2.1.2 vorgestellten Kategorien berechnen wir die zugehörigen θ -Histogramme in Abhängigkeit der relativen Lage zweier Kanten, um anschließend deren Entropie zweiter Ordnung, wie in Abschnitt 2.1.1 beschrieben, ermitteln zu können. Zusätzlich messen wir den Grad an Parallelismus eines jeden Bildes. Abbildungen 2.3 und 2.4 zeigen Entropie und Parallelismus, sowie die mittleren Richtungen für die verschiedenen Bildkategorien analog zu Abbildung 2.2c. Ein hoher Grad an Entropie lässt sich für die natürlichen Muster beobachten, besonders

Tabelle 2.1: Gelistet ist die jeweilige Zahl an Bildern der zehn im Experiment betrachteten Kategorien, sowie ein Verweis auf deren erstmalige Veröffentlichung.

Bildkategorie	Bilder	Quelle
JenAesthetics	1629	Amirshahi et al. (2013a)
Objects	207	Redies et al. (2012)
Facades	175	Amirshahi et al. (2013b)
Urban Scenes	219	Redies et al. (2012)
Faces	200	Martinez (1998)
Large Vistas	473	Redies et al. (2012)
Plant Patterns	331	Redies et al. (2012)
Lichen	244	Redies et al. (2012)
Clouds	268	Redies et al. (2012)
Branches	301	Redies et al. (2012)
Gesamt	4047	

hoch ist diese für *Lichen*. Kunst, hier repräsentiert durch den *JenAesthetics*-Datensatz, zeigt verglichen mit den nicht-natürlichen Mustern *Facades*, *Objects*, *Urban Scenes* und *Faces* eine hohe Entropie, welche mit vielen natürlichen Mustern vergleichbar ist. Auffällig ist, dass Muster mit wenig Entropie nicht zwangsläufig eine dominante Richtung zeigen (vgl. Abbildung 2.3, dritte Reihe von oben). Bei Bildern von Fassaden beispielsweise heben sich gespiegelt zueinander liegende Richtungen auf; trotz eines hohen Grades an Parallelismus und geringer Entropie ist in der Folge keine einzelne relative Orientierung im Bild dominant.

Um die Kategorien miteinander vergleichen zu können, zeigt Abbildung 2.5 die Werte für Entropie und Parallelismus gemittelt über den Winkel α in Abhängigkeit des relativen Abstandes d zweier Kanten. Hier bestätigt sich die Beobachtung der Verortung von Kunst: In dem Bereich von 20–240 Pixeln ist die Entropie der Kunst vergleichbar mit den natürlichen Kategorien und ist höher als die nicht-natürlicher Kategorien. Distanzen unterhalb von 20 Pixeln werden nicht betrachtet, da hier Artefakte derart auftreten, dass die relative Lage zweier Kanten in Polarkoordinaten nicht korrekt gezählt werden kann. Da insgesamt 48 Richtungen unterschieden werden, ein Pixel aber nur 8 direkte Nachbarn hat, sind die Richtungen im direkten Umfeld überdefiniert. Oberhalb von 240 Pixel kommt es ebenfalls vermehrt zu Artefakten, die dadurch erklärt werden können, dass nur

ein kleiner Teil der Pixel im Bild einen maximalen Abstand zueinander haben kann, nämlich jene in den Ecken eines Bildes, die sich auf den Bilddiagonalen gegenüberliegen. Da hier nur sehr wenige Kanten überhaupt in Beziehung gesetzt werden können, ist die resultierende Entropie erwartungsgemäß gering, was jedoch keinerlei Aussagekraft besitzt. Wir beschränken uns hier demzufolge auf die Analyse der Bereiche von 20–80 px, 80–160 px und 160–240 px, in Abbildung 2.5 dargestellt als blaue Hinterlegungen. Um die Entropie eines jeden Bildes für die definierten Bereiche mit je einem einzigen Wert beschreiben zu können, mitteln wir die Werte über den Winkel α sowie die genannten Bereiche des Abstandes d . Abbildung 2.6a zeigt die Verteilung der Entropie für den Bereich 20–80 px für die von uns getesteten Kategorien. Analog dazu ist in Abbildung 2.6b der Parallelismus der Bilder gezeigt. Die Verteilungen sind nicht normalverteilt, wir prüfen auf Gleichheit der Populationen daher parameterfrei mit Kruskal-Wallis’ H-Test und schließen Dunns Post-Test für den Vergleich einzelner Kategorien mit Kunst an. Für ein Signifikanzniveau von $p < 0.001$ unterscheidet sich die Verteilung der Entropie des *JenAesthetics*-Datensatzes von allen Kategorien außer *Large Vistas*. Für ein Signifikanzniveau von $p < 0.01$ unterscheiden sich beide Kategorien, gemessen am Abstand der Mediane der Verteilungen zueinander zeigen *Large Vistas* zu Kunst jedoch den geringsten Unterschied. Bezüglich des Parallelismus unterscheidet sich Kunst für den Bereich von 20–80px von allen Kategorien außer *Faces* mit einem Signifikanzniveau von $p < 0.001$. Für den Bereich von 80–160 px kann für Kunst kein Unterschied zu *Clouds* und *Branches* festgestellt werden, für 160–240 px ist kein Unterschied zu *Clouds* und *Large Vistas* bezogen auf die Entropie festzustellen. Bezüglich des Parallelismus unterscheidet sich Kunst von *Clouds* und *Plant Patterns* für 80–160 px sowie 160–240 px, im letztgenannten Bereich zusätzlich auch zu den natürlichen Mustern *Lichen* und *Branches*. Vergleicht man die Entropie erster Ordnung, unterscheidet Kunst sich nicht signifikant ebenfalls von den natürlichen Mustern *Clouds* und *Branches*.

2.2.2 Vergleich der Kantenentropie erster und zweiter Ordnung

Es ist bekannt, dass traditionelle Kunstwerke eine geringe Anisotropie der Kantenorientierungen aufweisen, was bedeutet, dass die Kantenorientierungen im Bild

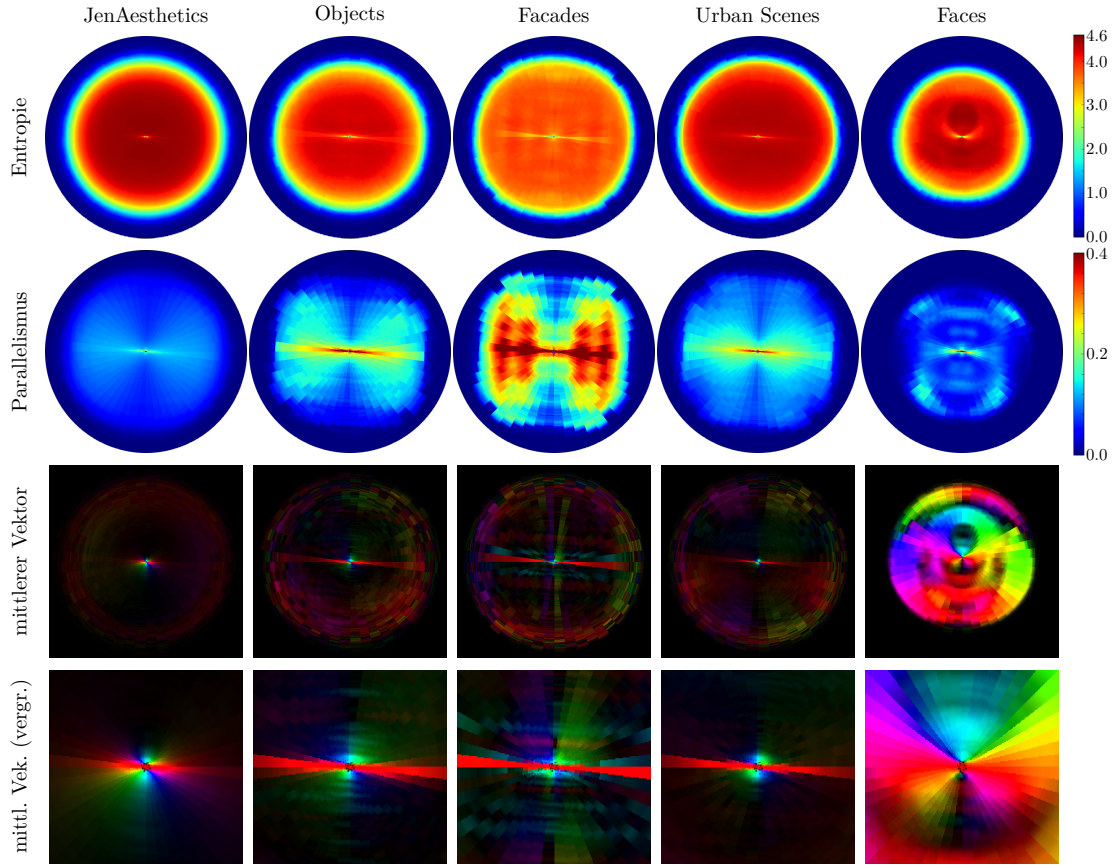


Abbildung 2.3: Zirkuläre Darstellung der Entropie, des Parallelismus und des mittleren Vektors gemittelt für verschiedenen Kategorien, analog zu Abbildung 2.2c. Oben: Entropie der θ -Histogramme, wobei ein dunkles Rot auf eine hohe Entropie und damit auf eine annähernde Gleichverteilung der relativen Kantenrichtungen hinweist, beispielsweise zu beobachten für *JenAesthetics*. Darunter ist der jeweilige Grad an Parallelismus gezeigt. Hoch ist dieser für Bilder der Kategorie *Facades*, wo ein großer Anteil von Kanten parallel verläuft. Darunter auf schwarzem Grund findet sich der mittlere Vektor der θ -Histogramme, wobei die Farbe Aufschluss über dessen Richtung gibt, die Helligkeit über dessen Länge. Je heller ein Punkt in der Darstellung ist, desto dominanter ist dessen entsprechende Richtung. Unten: Vergrößerung des mittleren Bereiches der Plots der mittleren Vektoren. Eine helle, rote, waagrecht im Zentrum liegende Linie deutet auf eine hohe Kolinearität hin, gut zu beobachten bei *Objects* und *Facades*, etwas weniger deutlich auch für *Urban Scenes*. Da Gesichter verglichen mit den anderen Bildkategorien sehr gleichförmig sind, dominieren dort bestimmte Kantenrichtungen in Abhängigkeit von α und d stark. Für die vergleichende Darstellung für natürliche Muster sei auf Abbildung 2.4 verwiesen.

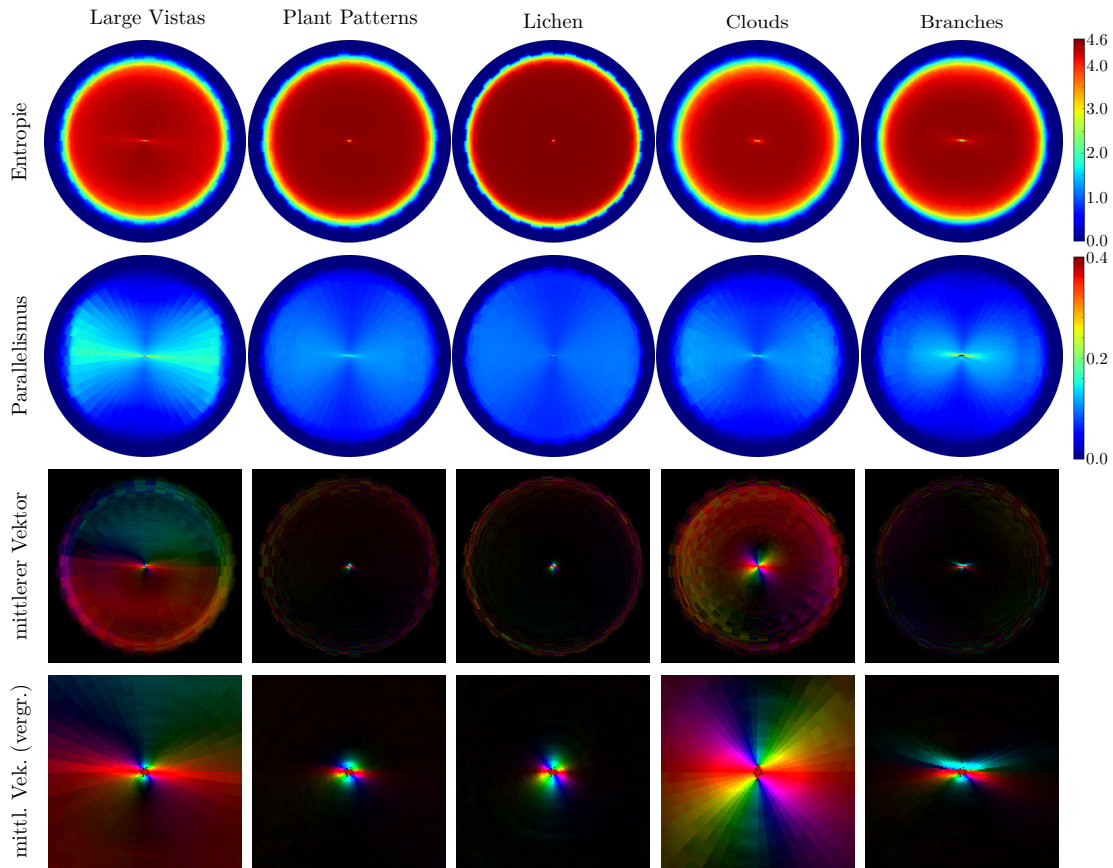


Abbildung 2.4: Die Darstellung folgt der in Abbildung 2.3 beschriebenen, hier für die natürlichen Muster. Die Entropie ist insgesamt ähnlich hoch wie für Kunst, wobei die Entropie für Lichen am höchsten ist, Kantenrichtungen sind hier somit weithin voneinander unabhängig. Im Vergleich der natürlichen Muster lassen sich Kolinearität und Parallelismus nur für *Large Vistas* beobachten, beides jedoch weit weniger ausgeprägt als etwa für *Objects* und *Facades* (vgl. Abbildung 2.3)

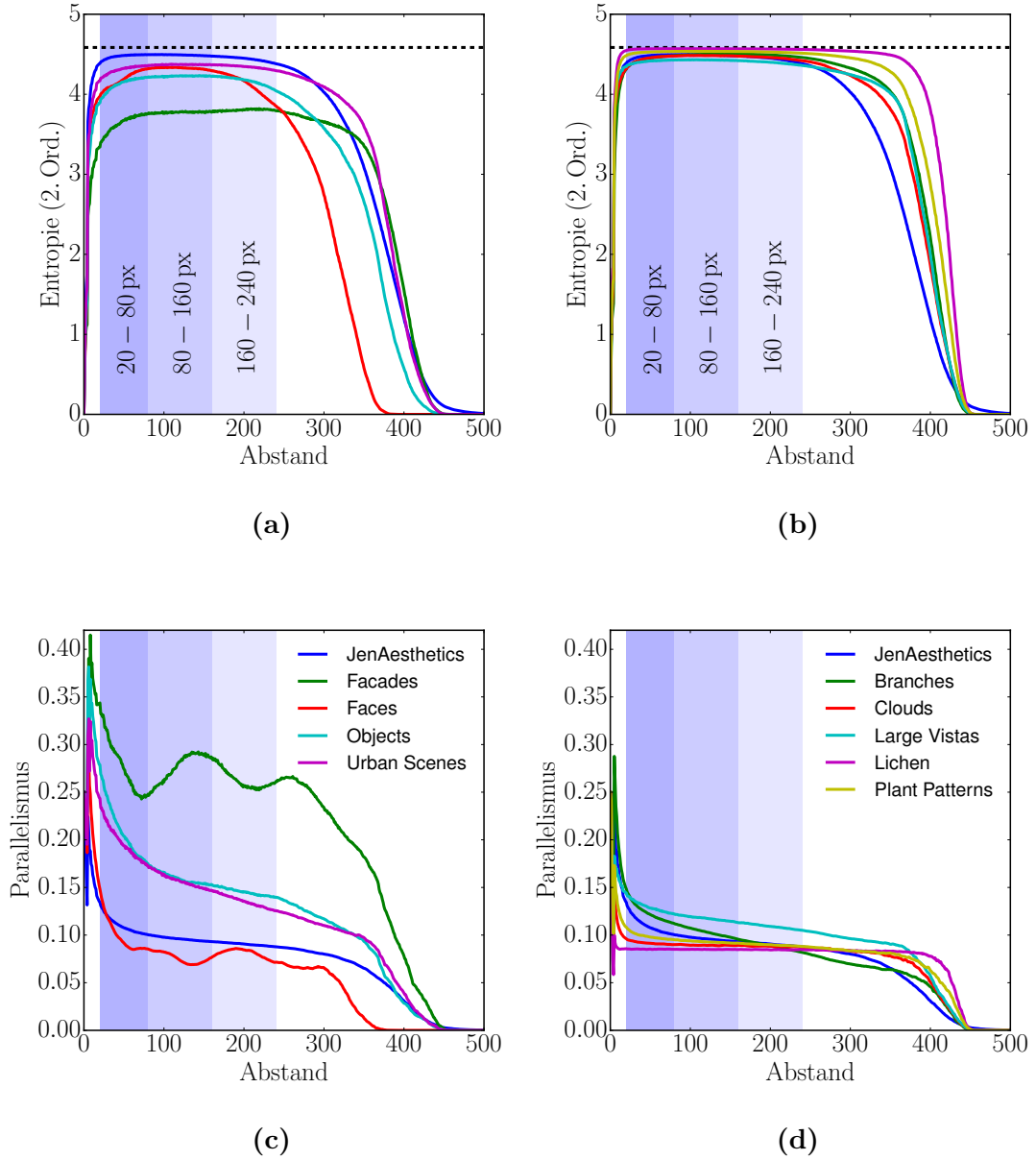


Abbildung 2.5: Gezeigt sind nicht-natürliche Muster (links; a,c) und natürliche Muster (rechts; b,d) im Vergleich mit Kunst. Die Graphen der ersten Reihe zeigen Entropie, die darunter Parallelismus für die entsprechenden Kategorien. Die Werte sind in Abhängigkeit des relativen Abstandes d abgetragen, wobei über den Winkel α gemittelt wurde. Für die weitere Analyse konzentrieren wir uns auf die blau hinterlegten Bereiche.

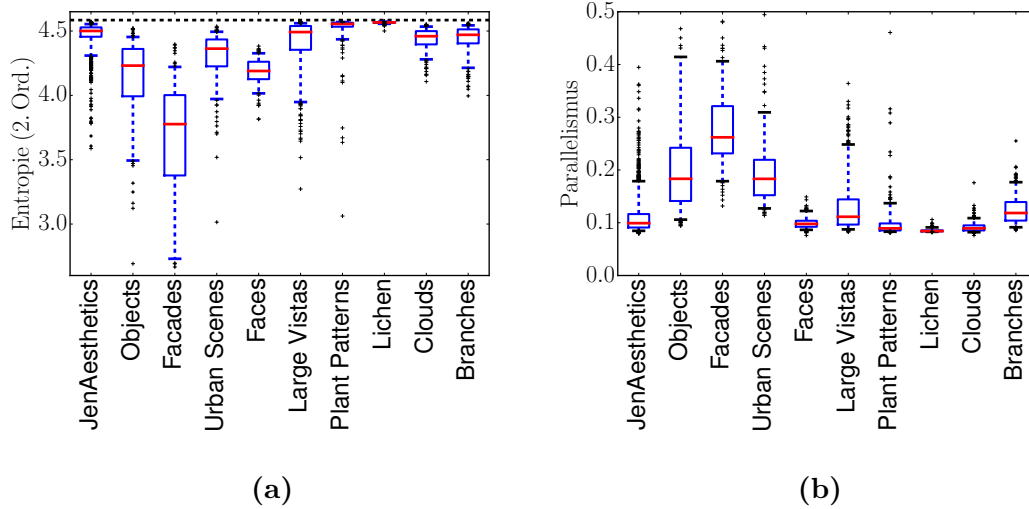


Abbildung 2.6: (a) Gezeigt ist die Entropie der getesteten Kategorien, hier gemittelt über den Bereich von 20–80 px. Die Werte für Kunst sind vergleichbar mit denen von *Large Vistas*, nicht-natürliche Muster wie einfache Objekte oder Fassaden zeigen hingegen eine geringere Entropie. Die natürlichen Muster *Plant Patterns* und *Lichen* zeigen die höchste Entropie im Vergleich. (b) Parallelismus als Spezialfall geringer Entropie. Hohe Werte werden vor allem für die nicht-natürlichen Muster *Facades*, *Urban Scenes* und *Objects* beobachtet. Die geringe Entropie von *Faces* hingegen lässt sich nicht durch einen hohen Grad an Parallelismus erklären, vielmehr tritt in diesen Bildern vermehrt eine Kozirkularität der Kantenrichtungen auf.

relativ gleichmäßig verteilt sind (Redies et al., 2007b; Braun et al., 2013). Zum Vergleich mit der vorgestellten Entropie zweiter Ordnung, welche relative Kantenbeziehungen betrachtet, berechnen wir daher auch die Entropie erster Ordnung, welche die Entropie des Histogramms über alle Kanten im Bild bezeichnet. Abbildung 2.7 stellt beide Werte für ausgewählte Kategorien gegenüber. Es wird dabei deutlich, dass diese wie erwartet nicht unabhängig voneinander sind, sondern sich gegenseitig teilweise bedingen. Es stellt sich also die Frage, ob eine hohe Entropie zweiter Ordnung vollständig durch eine hohe Entropie erster Ordnung erklärt werden kann. Dass dies nicht der Fall ist, wird in der genannten Abbildung anhand von Beispielbildern demonstriert. Für ein Bild mit nur einer dominierenden Orientierung der Kanten sind beide Entropiewerte vergleichsweise gering. Wird nun die Position der Bildelemente zufällig gesetzt, steigt die Entropie zweiter Ordnung, welche die relative Beziehung der Kanten beschreibt. Die Entropie erster Ordnung bleibt hingegen nahezu unverändert (vgl. Abbildung 2.7 *a* und *b*). Wird ein Teil der Bildelemente hingegen leicht gedreht, steigt die Entropie erster Ordnung wie erwartet an, werden die Bildelemente zusätzlich zufällig im Bild verteilt, wirkt sich dies erneut deutlich auf die Entropie zweiter Ordnung aus (vgl. Abbildung 2.7 *c* und *d*). Sind alle getesteten Kantenrichtungen im Bild vertreten, zeigt sich, dass die Entropie erster Ordnung nahezu maximal ist. Mit einer Randomisierung der Positionen der Bildelemente wird ein Bild von maximaler Entropie erster und zweiter Ordnung gleichzeitig erzeugt (vgl. Abbildung 2.7 *e* und *f*). An diesen Beispiel wird auch deutlich, dass selbst auf synthetischen Bildern Entropiewerte von unter 2.5 sehr selten beobachtet werden. Dies liegt darin begründet, dass bei der Kantenextraktion beispielsweise auch für das obere und untere Ende einfacher Striche Helligkeitsgradienten gefunden werden, die orthogonal zu dem Strich selber stehen und in der Berechnung zu allen anderen in Beziehung gesetzt werden, was die Entropie insgesamt hebt. Um den Zusammenhang der Entropie erster mit der zweiter Ordnung besser fassen zu können, berechnen wir deren Rangkorrelationskoeffizienten. Für die Korrelation beider Werte aller getesteten Kategorien (4047 Bilder) ergibt sich Spearmans ρ zu 0.81, für die Kunstbilder des JenAesthetics Datensatzes (1629 Bilder) ergibt sich ρ zu 0.63.

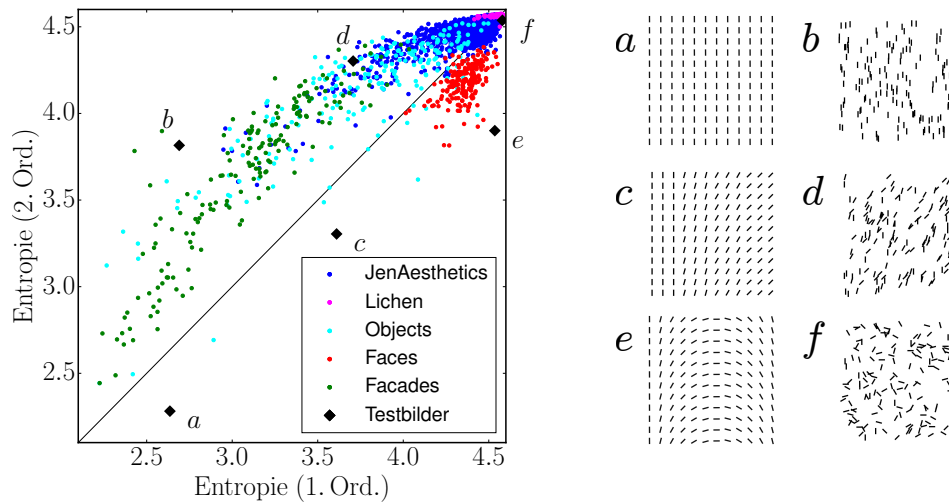


Abbildung 2.7: Gegenüberstellung der Entropie erster und zweiter Ordnung für ausgewählte Kategorien im Bereich 20–80 px. Deutlich wird, dass beide Maße nicht unabhängig voneinander sind, die Beispielbilder *a–f* illustrieren deren Zusammenhang genauer. Die Bilder *a*, *c* und *e* zeigen eine gleichmäßige Verteilung der Kanten unterschiedlicher Richtungen, wobei der Anteil der Kantenrichtungen jeweils zunimmt: In *a* dominieren senkrechte Kanten, in *e* sind alle gemessenen Kantenrichtungen vertreten. Bei den Bildern *b*, *d* und *f* sind die gleichen Kanten vertreten wie in den Bildern *a*, *c* und *e*, jedoch sind deren Positionen zusätzlich randomisiert. In der Folge steigt die Entropie zweiter Ordnung.

2.3 Diskussion der Ergebnisse

Verglichen mit den nicht-natürlichen Mustern weisen Kunstwerke eine hohe Entropie zweiter Ordnung auf, welche die relative Beziehungen zweier Kanten im Bild anhand eines paarweisen Vergleiches der Kanten misst. Darin ähneln Kunstwerke vielen natürlichen Mustern wie Bildern von Wolken oder Ästen. In unserem Experiment konnten wir in Nahbereich einer Kante für das gesetzte Signifikanzniveau von $p < 0.001$ keinen Unterschied zwischen den Verteilungen der Entropie zweiter Ordnung von Kunstwerken und denen von Landschaften feststellen, was ein Hinweis darauf sein kann, dass Kunstwerke bezüglich der Kantenrichtungen den Landschaftsbildern ähneln. Anzumerken ist dabei, dass nur 180 der 1629 Bilder des *JenAesthetics*-Datensatzes tatsächlich Landschaften zeigen, die naheliegende Erklärung, dass lediglich der Inhalt eines Bildes hier verglichen wird, greift daher zu kurz und mag nur teilweise zutreffend sein. Andere natürliche Muster wie Bilder von Flechten besitzen eine Entropie zweiter Ordnung nahe am Maximum von ≈ 4.585 , was darauf hindeutet, dass nahezu keine Vorhersage relativer Kantenbeziehungen möglich ist.

Eine geringe Entropie zweiter Ordnung kann unterschiedliche Gründe haben. Ein von uns betrachteter Spezialfall ist der Parallelismus, der den Umstand bezeichnet, dass Kanten im Bild parallel zueinander verlaufen. Besonders häufig lässt sich Parallelismus in Bildern beobachten, welche nicht-natürliche Strukturen zeigen, wie beispielsweise einfache Objekte, Städteszenen oder besonders deutlich in Bildern von Fassaden großer Gebäude. Neben einem hohen Grad an Parallelismus kann beispielsweise die hier nicht näher behandelte *Kozirkularität* von Kanten ebenfalls eine geringe Entropie zweiter Ordnung bedingen. Kozirkularität beschreibt den Umstand, dass Kanten auf einem gemeinsamen Kreis im Bild liegen und entsprechend des Kreises ausgerichtet sind. In unserem Experiment wurde diese von uns nicht gesondert ermittelt. Die vergleichsweise geringen Entropiewerte der Bilder der Kategorie *Faces* (vgl. Abbildung 2.6) legen jedoch die Vermutung nahe, dass ein Teil davon durch Kozirkularität zu erklären ist. Auch wenn Gesichter keinen perfekten Kreis beschreiben, so liegen deren Außenkanten dennoch angenähert auf einer Ellipse. Die dominanten Farben der Visualisierung für Gesichter in Abbildung 2.3 (dritte Reihe von oben) beschreiben ebenfalls Kreise verschiedener

Radien, was ein Hinweis auf vorliegende Kozirkularität der getesteten Bilder von Gesichtern ist.

Eine geringe Entropie erster Ordnung kann dadurch begründet sein, dass nicht alle möglichen Kantenrichtungen in einem Bild repräsentiert sind. Eine Randomisierung regelmäßig angeordneter Kanten hebt die Entropie zweiter Ordnung um einen gewissen Grad (vgl. Abbildungen 2.7), ein Maximum wird jedoch erst erreicht, wenn alle möglichen Kantenrichtungen zu ungefähr gleichen Teilen in einem Bild repräsentiert sind. Die Korrelation beider Werte für Kunst zeigt, dass die Entropie zweiter Ordnung nicht vollständig durch die Entropie erster Ordnung erklärt werden kann und daher als Maß, insbesondere für die Beurteilung der statistischen Eigenschaften von Kunst, von Interesse ist.

2.4 Zusammenfassung

In diesem Kapitel, welches auf der Veröffentlichung von Redies et al. (2017) beruht, wurde eine Methode zur Messung der Entropie der Kantenrichtungen von Bildern präsentiert. Während in der vorliegenden Arbeit vor allem die Vorstellung der Methode im Vordergrund stand, wird in der Veröffentlichung neben den in diesem Kapitel vorgestellten Ergebnissen weiteren, sich anschließenden Fragen nachgegangen. So wird beispielsweise der Einfluss des Bildinhaltes auf die Entropie genauer untersucht, und ebenfalls, ob Aussagen über die Entropie verschiedener Farbkanäle impressionistischer Bilder getroffen werden können. In Redies und Brachmann (2017) findet die Methode der Messung von Entropie Anwendung, um traditionelle Kunst verschiedener Kulturkreise mit „Schlechter Kunst“ und abstrakter Kunst zu vergleichen. In Henemann et al. (2017) werden Bilder von Menschen mit Schizophrenie mithilfe der vorgestellten Methode miteinander verglichen und voneinander abgegrenzt. Der interessierte Leser sei auf die genannten Veröffentlichungen verwiesen.

Wir haben gezeigt, dass die Bilder des *JenAesthetics*-Datensatzes sich dadurch auszeichnen, dass die Entropie der Kantenrichtungen relativ zueinander im Nahbereich einer Kante signifikant höher ist als die der ebenfalls von uns getesteten Bilder, welche Objekte oder Gebäude zeigen. Für den Datensatz, welcher Aufnahmen von Landschaften beinhaltet, konnte hingegen kein signifikanter Unterschied der Verteilung zu Kunst festgestellt werden. Dies könnte ein Hinweis darauf sein,

dass Künstler unbewusst die Statistik natürlicher Szenen in ihren Werken nachbilden, um so ein bestimmtes, neuronales Antwortmuster hervorzurufen (Redies, 2007b), ein Beleg für diese Annahme ist es jedoch nicht. Da die Entropie in Kunstbildern, welche gemeinhin als ästhetisch empfunden werden, höher ist als in Bildern, welche beispielsweise einfache Objekte zeigen, schließt sich die Frage an, ob der Grad an Entropie der Kanten ein Grund für die Ästhetik eines Bildes ist, oder diese zumindest teilweise bedingt. Eine Studie, in welcher Probanden Bilder verschiedener Grade an Entropie ästhetisch bewerten, kann diese Frage näher untersuchen. Dennoch ist auch hier Vorsicht geboten, da nicht von einfachen Mustern auf Kunst geschlossen werden darf. Kunstwerke sind komplexe Stimuli und vor dem Hintergrund, dass auch die visuelle Wahrnehmung ein hochgradig komplexer Prozess ist, wird es schwierig sein, hier absolute Aussagen treffen zu können.

Die hier vorgestellte Analyse der Entropie von Kantenrichtungen ist entsprechend der in Kapitel 1 vorgestellten Gliederung von Merkmalen zwar global, da auch weit auseinander liegende Kanten betrachtet werden, wird aber dennoch den *low-level* Merkmalen zugerechnet, da Kanten an sich keine abstrakten Bildmerkmale darstellen. Im folgenden Kapitel werden wir eine Methode vorstellen, die Symmetrie von Bildern zu berechnen, welche ebenfalls eine globale Eigenschaft von Bildern beschreibt. Als Grundlage der Berechnung werden die Filterantworten tiefer neuronaler Netze genutzt, womit wir den Abstraktionsgrad der zugrunde liegenden Merkmale erhöhen werden.

Kapitel 3

Symmetrie

Im Laufe der Evolution hat die Natur viele symmetrische Formen hervorgebracht, seien es die Blätter von Bäumen, die Flügel von Vögeln oder Schmetterlingen, oder der äußere Körperbau des Menschen. Doch auch physische Strukturen wie etwa Kristalle, die keines evolutionären Ursprungs sind, oder gänzlich von Menschen hervorgebrachte Formen wie die Verkleidung eines Autos, zeigen symmetrische Formen. *Symmetrie* heißt in diesem Zusammenhang, dass ein oder mehrere Teile eines Musters auf einen anderen Teil desselben abgebildet werden können (Bronshtein et al., 2015). Im Zweidimensionalen werden *Rotationssymmetrie*, *Achsensymmetrie*, *Punktsymmetrie* und die *Skalensymmetrie* unterschieden. Die Achsensymmetrie, bei der die Abbildung des Musters durch einer Spiegelung entlang einer Symmetrieachse bestimmt ist, nimmt dabei insofern eine besondere Stellung ein, als dass diese für Menschen sehr auffällig und leicht zu entdecken ist (Wagemans, 1995). Spiegelsymmetrie beeinflusst die Attraktivität von Gesichtern positiv (Grammer und Thornhill, 1994) und ist ein Indikator für normale Entwicklung und Stressresistenz (Møller und Swaddle, 1997; Tinio und Smith, 2014; Zaidel et al., 2005). All dies spricht dafür, dass Symmetrie ein wichtiger Faktor in unserem ästhetischen Urteil ist.

In der Mathematik ist Symmetrie ein formales Konzept der Gruppentheorie. In der Bildverarbeitung werden jedoch reale Daten in Form aufgenommener Bilder verarbeitet, welche verrauscht, mehrdeutig oder auch gestört sein können. Dennoch wurden in der Vergangenheit verschiedene Algorithmen zur Detektion von Symmetrien in Bildern vorgeschlagen (Liu et al., 2010; Chen et al., 2007; Liu

et al., 2013), wobei ein besonderer Fokus auf der Detektion von Symmetrieachsen gelegt wurde. *Stetige Symmetrie* (engl. *continuous symmetry*, Zabrodsky et al. (1995)) misst den Grad, zu welchem ein Objekt, bestehend aus einer Menge von Punkten, oder ein Bild als diskrete, zweidimensionale Funktion, symmetrisch ist. Diese Aufgabe ist für Menschen relativ einfach intuitiv zu bewältigen, algorithmisch jedoch schwieriger zu lösen. Eine Methode, die stetige Symmetrie eines Bildes zu messen, wurde von Heijer (2015) vorgeschlagen und besteht darin, die Luminanzwerte der Pixel entlang einer Symmetrieachse miteinander zu vergleichen. Diese Herangehensweise unterscheidet sich jedoch stark davon, wie Menschen ein Bild wahrnehmen. Menschen sehen Kanten, fassen diese zu Formen und Texturen zusammen und gruppieren diese schließlich zu Objekten. Shaker und Monadjemi (2015) schlagen dem angenähert ein Verfahren zur Messung der Symmetrie vor, welches die Luminanzkanten eines Bildes als Grundlage nimmt. Auch wenn dies über den bloßen Pixelvergleich hinausgeht, so bleiben Farbe und abstraktere Formen dabei weiterhin unbeachtet. In diesem Kapitel wird ein Algorithmus vorgeschlagen, der diese Probleme adressiert. Genutzt werden dafür die Filterantworten tiefer neuronaler Netze, welche in den letzten Jahren in der Bildverarbeitung stark an Popularität gewonnen haben. Einerseits wird damit dem Sehsystem des Menschen insofern entsprochen, als dass auch Farbe und Textur in Form verschiedener Frequenzen der Helligkeitsgradienten in die Berechnung einfließen (Wurtz und Kandel, 2000b; Lennie, 2000). Darüber hinaus wird gezeigt, dass Filterantworten höherer Schichten des Netzes ebenfalls als Grundlage genutzt werden können und durch ihre abstraktere Repräsentation der Bildinhalte das hier vorgeschlagene Verfahren weiter verbessern können.

Da tiefe neuronale Netze in Form von *Convolutional Neural Networks* (CNNs) eine zentrale Rolle sowohl innerhalb dieses Kapitels als auch innerhalb dieser Arbeit einnehmen, werden diese im nächsten Abschnitt zunächst vorgestellt und ihre Funktionsweise beschrieben.

3.1 Convolutional Neural Networks

Obwohl Convolutional Neural Networks schon vor mehr als drei Jahrzehnten (Fukushima, 1980; Lecun und Bengio, 1995; Lecun et al., 1998) in ihren Grundideen

vorgestellt wurden, sind sie erst in den letzten Jahren zu einer etablierten Methode in vielen Bereichen der Bildverarbeitung und darüber hinaus geworden, und haben dabei konventionelle Methoden ersetzt. Diese Erfolgsgeschichte kann vor allem auf zwei Hauptfaktoren zurückgeführt werden: Als erstes sei die technische Weiterentwicklung der Hardware zu nennen, welche die Anwendung von CNNs bedeutend schneller und effizienter macht. Während erste neuronale Netze noch auf herkömmlichen Prozessoren (CPUs) trainiert und ausgeführt wurden, so haben schnelle und preiswerte Grafikprozessoren (GPUs) die Entwicklungen in diesem Bereich stark beschleunigt. Aktuell stellen große Cloud-Anbieter wie Google auch dedizierte Prozessoren wie *Tensor Processing Units* (TPUs) bereit und treiben die Entwicklung so weiter voran. Zum Zweiten ist es erst mit der öffentlichen Verfügbarkeit großer Datenmengen wie etwa ImageNet, einer Sammlung von über einer Million in 1000 verschiedene Kategorien eingeteilter Bilder (Deng et al., 2009; Russakovsky et al., 2015), möglich geworden, die sinnvolle Anwendbarkeit von CNNs für die Objekterkennung zu demonstrieren. CNNs werden momentan neben der Objekterkennung (Krizhevsky et al., 2012; Simonyan und Zisserman, 2014; He et al., 2015; Huang et al., 2016) unter anderem in der Bildbeschreibung (Karpathy und Fei-Fei, 2015) oder auch der Synthese von Texturen (Gatys et al., 2015b) erfolgreich eingesetzt und haben darüber hinaus auch andere Bereiche wie die Spracherkennung (Abdel-Hamid et al., 2014) revolutioniert.

CNNs lernen eine Hierarchie verschiedener Filter, welche, angewendet auf eine Eingabe wie beispielsweise ein Bild, Informationen aus diesem extrahieren können. Während bei Bildern auf den unteren Schichten einfache Luminanz- und Gegenfarbkanten detektiert werden, kombinieren höhere Schichten im Netz diese wiederum und sprechen so auf abstraktere Bildinhalte an (Yosinski et al., 2015). In der Vergangenheit wurden verschiedene Schichten vorgeschlagen, die die Ausgaben jeweils darunter liegender Schichten auf verschiedene Arten und Weisen transformieren. Weite Anwendung finden *convolutional layers*, welche ein Bild entsprechend einer Faltungsmatrix filtern und gleichzeitig Namensgeber von CNNs sind, *pooling layers*, die eingesetzt werden, um Bildausschnitte zusammenzufassen und so deren Auflösung herabzusetzen, und *fully-connected layers*, welche eine beliebige Abbildung der Ausgabe einer Schicht auf die Eingabe der darüber liegenden Schicht lernen können.¹ Bei der Objekterkennung werden letztere meist als obere

¹Zum besseren Verständnis werden nachfolgend die englischen Begriffe unübersetzt verwendet.

Schichten eines Netzes gewählt, um die detektierten Bildmerkmale vorangegangener Schichten verschiedenen Objektkategorien zuzuordnen. All diese Schichten (mitunter auch als Ebenen bezeichnet) hier im Detail zu beschreiben, würde den Rahmen der Arbeit sprengen, stattdessen sei für eine vertiefende Darstellung der einzelnen Bestandteile eines Netzes auf Goodfellow et al. (2016) verwiesen.

Das Training eines Netzes zur Objekterkennung geschieht üblicherweise *überwacht*, indem eine aktuelle Ausgabe des Netzes mit der gewünschten Ausgabe verglichen wird, was die Berechnung eines Fehlers erlaubt. Da das gesamte Netz als eine verkettete Funktion verstanden werden kann, welche die Eingabe auf die Ausgabe abbildet und dabei komplett differenzierbar ist, kann mithilfe der Kettenregel der Gradient eines jeden Parameters bestimmt und der Parameter so angepasst werden, dass dieser den aktuellen Fehler verringert. Infolge einer hohen Zahl von Iterationen auf vielen verschiedenen Eingaben ist das Netz so fähig, die gewünschte Generalisierung zu lernen und auch für unbekannte Eingaben korrekte Zuordnungen vorzunehmen. Diese Trainingsmethode wird allgemein als *Backpropagation* (Rumelhart et al., 1986) bezeichnet und ist den Gradientenabstiegsverfahren zuzurechnen. Für die Details des Trainings sei ebenfalls auf die oben genannte vertiefende Darstellung verwiesen.

Die Architekturen von CNNs sind Gegenstand aktueller Forschung. In diesem Kapitel, wie auch in den beiden folgenden, wird auf ein weit verbreitetes und etabliertes Modell, das sogenannte *AlexNet* (Krizhevsky et al., 2012), zurückgegriffen. Genutzt wird dabei die in *Caffe* (Jia et al., 2014), einer weit verbreiteten Bibliothek für das Trainieren und Anwenden tiefer neuronaler Netze, bereitgestellte Implementation. Die AlexNet-Architektur setzt sich aus insgesamt acht Ebenen, darunter *convolutional layers*, *pooling layers*, sowie Ebenen, welche eine lokale Normalisierung der Helligkeit vornehmen, zusammen. Die Zuordnung der Bildkategorie geschieht anhand insgesamt drei hintereinander geschalteter, vollverbundener Schichten. Eine Darstellung des schematischen Aufbaus, sowie eine Visualisierung der Filter der ersten Ebene findet sich in Abbildung 3.1.

Das AlexNet wird in der Forschung vielfach verwendet. Ein gängiger Weg ist es, dessen Architektur unverändert zu übernehmen und auf ein neues Problem anzuwenden, wobei alle Parameter des Netzes neu gelernt werden. Wie in Kapitel 1 gezeigt, hat es sich jedoch bewährt, das Netz nicht vollständig neu zu trainieren, sondern dessen Parameter für ein neues Problem umzulernen. So kann Trainings-

zeit gespart und die Klassifikationsrate gegenüber einem vollständigen Neulernen sogar verbessert werden (Tan et al., 2016). Wir verfolgen hier einen dritten Weg, bei dem die Aktivierungen des AlexNet-Modells unverändert übernommen werden, ein Ansatz, der sich in der Vergangenheit für verschiedene Zielstellungen als gewinnbringend erwiesen hat (Karayev et al., 2013; Donahue et al., 2014).

3.2 Material und Methoden

3.2.1 Messen stetiger Symmetrie

Das AlexNet verarbeitet Bilder mit einer Auflösung von 224×224 Pixeln. Um auch höhere Auflösungen verarbeiten zu können, werden die vollverbundenen Schichten an der Spitze des Netzes verworfen, wodurch möglich ist, die Größe der Eingabebilder auf 512×512 Pixel heraufzusetzen. Alle im Netz verbleibenden Schichten verarbeiten das Bild nur lokal, und können so unverändert übernommen werden, während die verworfenen Schichten auf dem globalen Bild und damit auf einer festen Auflösung operieren. In Gatys et al. (2015a) wurde gezeigt, dass Bildinhalte nach dieser Modifikation weiter erhalten bleiben. Die stetige Symmetrie auf den Filterantworten des so modifizierten Netzes wird wie folgend beschrieben berechnet.

Zunächst wird jedes Bild I als Eingabe präsentiert und durch das Netz vorwärts propagiert. In Folge stehen 96 Filterantworten auf Schicht conv1, 256 Filterantworten auf conv2, 384 auf conv3 und conv4, sowie 256 Filterantworten auf der letzten Schicht des Netzes zur Verfügung. Für jede dieser Schichten werden die maximalen Antworten über $n \times n$ gleich große, nicht überlappende Teilregionen extrahiert, was in der Terminologie neuronaler Netze einer Max-Pooling-Operation mit einer Verschiebung (engl. *striding*) gleich der Größe einer Teilregion entspricht. Für einen gegebenen Parameter n entsteht so für jeden convolutional layer $l \in \{\text{conv1}, \text{conv2}, \dots, \text{conv5}\}$ ein multidimensionales Array $M_{l,n}$, welches die maximalen Antworten eines jeden Filters einer Ebene enthält. Anschließend wird das Bild wahlweise horizontal oder vertikal gespiegelt und der Vorgang wiederholt, woraufhin ein zweites Array $M'_{l,n}$ mit den maximalen Filterantworten des gespiegelten Bildes berechnet wird. Die Achsensymmetrie des Bildes entsprechend der

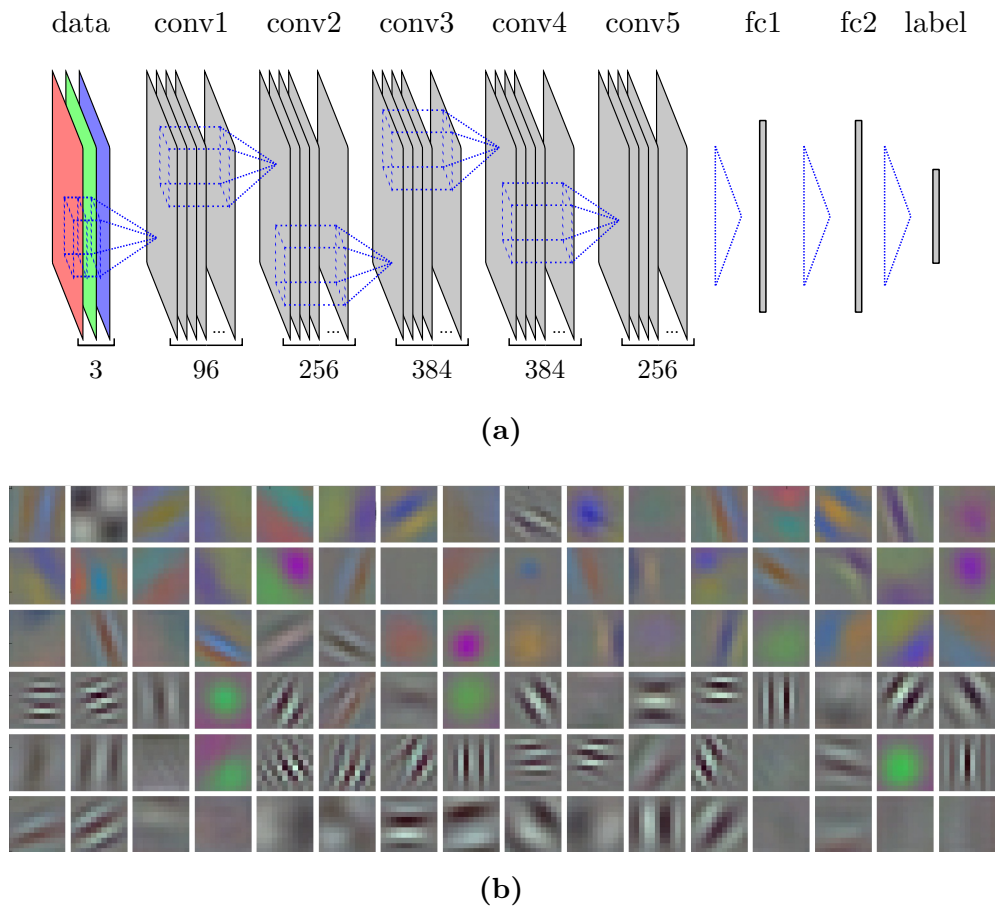


Abbildung 3.1: (a) Schematische Darstellung des AlexNet-Modells, welches in den Experimenten dieses und folgender Kapitel zur Anwendung kommt. Ebenen, welche Ausgabe einer vorangegangenen Filteroperation sind, sind grau eingefärbt (conv1–conv5: convolutional layers, fc1 und fc2: fully-connected layers). Die Angabe unterhalb der einzelnen Ebenen entspricht der Anzahl verschiedener Kanäle auf der jeweiligen Ebene, so werden auf das Eingabebild (data) auf Ebene conv1 beispielsweise 96 verschiedene Filter angewendet, was wiederum zu 96 Kanälen auf Ebene conv1 führt. Die hierarchische, fortlaufende Filterung des Eingabebildes erlaubt die Extraktion von Bildmerkmalen, welche für eine korrekte Erkennung des Bildinhaltes von Bedeutung sind. Normalisierungs- und Pooling-Schichten sind nicht gezeigt. (b) Dargestellt ist eine Visualisierung der 96 Filter, welche auf Ebene conv1 auf das Eingabebild angewendet werden. Diese sprechen auf Luminanzkanten verschiedener Frequenzen sowie auf verschiedene Gegenfarbkanten und Farbflächen an.

zur Spiegelung genutzten Achse als kontinuierlicher Wert kann anhand folgender Formel berechnet werden:

$$A(M_{l,n}, M'_{l,n}) = \frac{\sum_{x,y,f} |M_{l,n}(x, y, f) - M'_{l,n}(x, y, f)|}{\sum_{x,y,f} \max(M_{l,n}(x, y, f), M'_{l,n}(x, y, f))} \quad (3.1)$$

wobei x und y über die Teilregion 1 bis n iterieren und f über die Zahl der Kanäle auf der genutzten Ebene l des Netzes. Indem wir $A(M_{l,n}, M'_{l,n})$ von 1 subtrahieren, erhalten wir das finale Ergebnis für die Symmetrie des Eingabebildes entsprechend der zur Spiegelung genutzten Achse:

$$S(M_{l,n}, M'_{l,n}) = 1 - A(M_{l,n}, M'_{l,n}). \quad (3.2)$$

Der Wertebereich dieses Maßes für Symmetrie ist begrenzt auf Werte zwischen 0 und 1, wobei geringe Werte für asymmetrische Bilder, hohe Werte hingegen für symmetrische Bilder angenommen werden.

3.2.2 Datensatz zur Evaluation

Die Evaluation des im vorangegangenen Abschnitt vorgestellten Algorithmus zur Berechnung der Symmetrie wird anhand eines Datensatzes von 300 CD-Titelbildern vorgenommen. In diesem sind jeweils 100 Bilder der Genres Klassik, Popmusik und Metal enthalten.² Alle Bilder haben eine Auflösung von 500×500 Pixeln oder wurden durch bikubische Interpolation auf diese Auflösung heruntergerechnet. Für jedes Cover wurde ein Referenzsymmetriewert anhand einer Befragung von zwanzig Probanden bestimmt. Die Teilnehmer waren 21–62 Jahre alt, das mittlere Alter betrug 36 Jahre, und sechs Teilnehmer waren männlichen Geschlechts. Alle Teilnehmer gaben an, über eine normale oder über eine zu normal korrigierte Sehkraft zu verfügen. Die Bilder wurden auf einem kalibrierten Bildschirm (EIZO ColorEdge CG241W, 1920×1200 Pixel Auflösung) auf schwarzem Hintergrund präsentiert. Eine Kinnstütze gewährleistete dabei einen konstanten Abstand von 70 cm. Die Bildgröße von 500×500 Pixeln auf dem Bildschirm entsprach 135×135 mm und damit einem Sehwinkel von 11×11 Grad. Die Studie entsprach den ethischen Grundsätzen für die medizinische Forschung am Menschen und wurde durch die Ethikkommission des Uniklinikums Jena genehmigt, alle Teilnehmer

²Die Bilder wurden freundlicherweise von Frau Maria Grebenkina zur Verfügung gestellt.

stimmten vor Beginn der Teilnahme schriftlich zu. Die Teilnehmer wurden in einem geschlossenen Raum einzeln getestet, wobei jedem Teilnehmer zunächst das Experiment erklärt wurde. Aufgabe war es, das angezeigte Bild entsprechend seiner vertikalen Symmetrie zu bewerten. Das Experiment wurde mit einem Mausklick gestartet, woraufhin für 300–800 ms ein Fixationskreuz angezeigt wurde. Nach Ablauf dieser Zeit verschwand das Fixationskreuz und das zu bewertende Bild erschien. Die im unteren Teil des Bildschirms angezeigte Frage lautete „Wie symmetrisch ist das Bild?“, die Antwort wurde in Form eines Klicks auf eine darunter angezeigte kontinuierliche Skala beschriftet mit „nicht symmetrisch“ im linken Extrem und „sehr symmetrisch“ im rechten Extrem gegeben. Direkt nach einer gegebenen Antwort wurde erneut das Fixationskreuz angezeigt und daraufhin ein weiteres Bild, wobei alle Bilder jeweils in zufälliger Reihenfolge präsentiert wurden. Nach jeweils 100 vorgenommenen Bewertungen war es den Teilnehmern freigestellt eine Pause einzulegen.

3.3 Ergebnisse

3.3.1 Ergebnisse des Bewertungsexperimentes

Da sich für 97 der 300 Cover die Verteilung der Symmetriebewertungen als nicht normalverteilt herausstellte (D’Agostino-Pearson-Test, $p < 0.05$; d’Agostino (1971)), wurde der Median aller Bewertungen anstelle des Durchschnittes als repräsentativer Wert eines jeden Bildes gewählt. Abbildung 3.2 zeigt einen Boxplot der Median-Symmetriebewertungen der drei getesteten Genres. Gemessen am Median der finalen Symmetriewerte ist die Symmetrie für Titelbilder des Genres Metal am höchsten, gefolgt von Pop. Am geringsten ist der Median-Symmetriewert für Bilder des Genres Klassik. Alle Genres weisen eine hohe Streuung der Symmetriebewertungen auf, wobei auch die Extrembereiche abgedeckt werden. Der Datensatz ist dadurch gut geeignet, als Basis für den Vergleich der Ergebnisse des vorgestellten Verfahrens zur Berechnung der Symmetrie mit den Einschätzungen menschlicher Betrachter zu dienen.

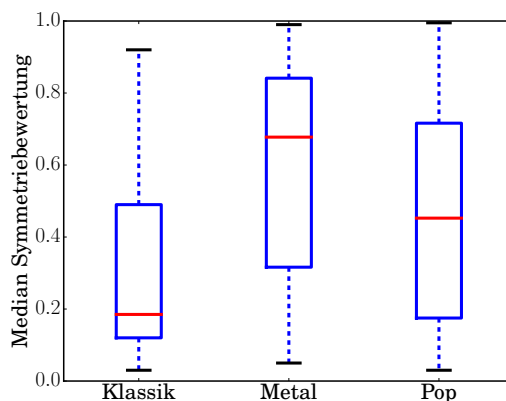


Abbildung 3.2: Dargestellt sind die Ergebnisse des Bewertungsexperimentes, in welchem menschliche Betrachter die Symmetrie einzelner Bilder bezüglich der von ihnen wahrgenommenen Symmetrie beurteilten. Aufgeteilt sind diese nach den drei Musikgenres Klassik, Metal und Pop, wobei Metal die höchste Median-Symmetrie aufweist, Bilder des Genres Klassik hingegen die geringste.

3.3.2 Ergebnisse der Symmetrieberechnungen

Unser Verfahren beinhaltet zwei freie Parameter. Zum einen steht es frei, welche Schicht des genutzten Netzes als Grundlage für die Berechnung dient (Parameter l in Gleichung 3.2), zum anderen ist die Zahl der Teilregionen und damit die Größe der einfließenden lokalen Bereiche (Parameter n in Gleichung 3.2) frei wählbar. Im Zuge unserer Evaluation variieren wir beide Parameter l und n , und berechnen für jedes Bild den Symmetriewert wie in Abschnitt 3.2.1 dieses Kapitels beschrieben. Wir vollziehen dies für jede der fünf Schichten $l \in \{\text{conv1}, \dots, \text{conv5}\}$ des Netzes und einer Größe der Teilregionen von $n = 2$ bis $n = 32$ für die Ebenen conv1 und conv2. Für die verbleibenden Ebenen berechnen wir die Symmetrie für $n = 2$ bis $n = 31$, da hier die minimale Größe einer Teilregion bedingt durch die zunehmende Auflösungsreduktion der Filterantworten beschränkt ist. Abbildung 3.3 zeigt Beispielfelder mit zugehörigen, berechneten Symmetriewerten.

Die Beziehung der durch menschliche Betrachter gewonnenen Symmetriewerte zu den berechneten quantifizieren wir auf zwei verschiedene Arten. Da die Verteilungen der Symmetriewerte keiner Normalverteilung folgen, bestimmen wir zunächst Spearmans Rangkorrelationskoeffizient ρ zwischen den berechneten und den im

Experiment mit menschlichen Betrachtern festgestellten Werten. Abbildung 3.4a stellt die Beziehung zwischen n und dem Korrelationskoeffizienten für alle fünf Ebenen dar. Die Werte reichen dabei von $\rho = 0.64$ ($l = \text{conv1}$, $n = 2$) bis $\rho = 0.90$ ($l = \text{conv5}$, $n = 6$). Zusätzlich nähern wir den Zusammenhang zwischen den berechneten und den erhobenen Werten auch durch lineare (Abb. 3.4b), quadratische (Abb. 3.4c) und kubische Funktionen (Abb. 3.4d) an und messen als Gütekriterium der Annäherung die Wurzel der mittleren Fehlerquadrante (RMSE). Die Kurvenverläufe sind hier ähnlich, allerdings zeigt sich, dass Polynome zweiten und dritten Grades geringere Fehler bewirken und damit den Zusammenhang besser abzubilden scheinen als lineare Funktionen.



Abbildung 3.3: Repräsentative CD-Titelbilder, wie sie in unserem Experiment verwendet wurden. Gezeigt ist jeweils ein Bild hoher, mittlerer sowie geringer Symmetrie, die angegebenen Symmetriewerte wurden mit der in Abschnitt 3.2.1 beschriebenen Methode berechnet ($l = \text{conv1}$, $n = 17$). Aufgrund von Urheberrechtsbestimmungen können die Originalbilder unserer Studie hier nicht gezeigt werden. (a) vom Autor; (b) Graham James Worthington, CC BY-SA 4.0; (c) Musiclive55, CC BY-SA 4.0.

Zum Vergleich implementierten wir das eingangs erwähnte von Heijer (2015) vorgeschlagene Verfahren, welches auf einem paarweisen Vergleich der Helligkeitswerte beruht. Die Korrelation dieser Werte mit den im Experiment durch Probanden festgestellten Symmetriebewertungen ergab sich zu $\rho = 0.34$, es kann also gefolgert werden, dass das von uns beschriebene Verfahren in allen Konfigurationen den menschlichen Einschätzungen der Symmetrie der Bilder näher kommt. Die von Shaker und Monadjemi (2015) vorgestellte Methode zur Berechnung der Sym-

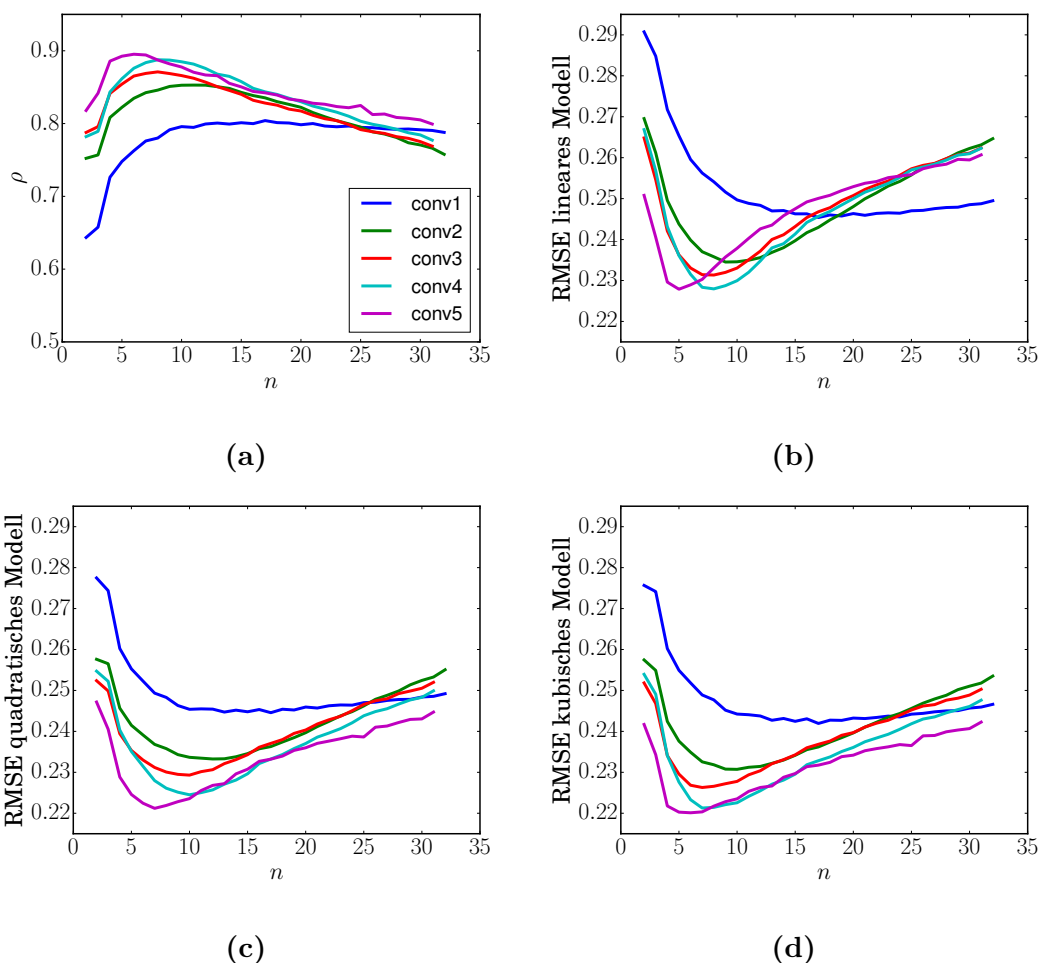


Abbildung 3.4: (a) Spearmans Rangkorrelationskoeffizient ρ zur Beurteilung des Zusammenhangs der errechneten und der experimentell festgestellten Symmetriewerte in Abhängigkeit der Größe der Teilregionen für die verschiedenen Schichten conv1–conv5. Das Quadrat der auf der x-Achse abgetragenen Werte entspricht der Zahl der Teilregionen, in welche ein Bild unterteilt wird. Zum besseren Verständnis des Zusammenhangs ist auch der RMSE eines (b) linearen, eines (c) quadratischen sowie eines (d) kubischen Modells der Beziehung berechneter zu den erhobenen Werten gezeigt. Der im Vergleich höhere RMSE des linearen Modells deutet darauf hin, dass der Zusammenhang nicht rein linear ist. Beispielkurven eines quadratischen Modells sind in Abbildung 3.5 gegeben.

metrie konnte aufgrund fehlender Angaben zum Filterungsprozess von uns nicht reproduziert werden.

Vergleicht man verschiedene Parameterkonfigurationen, so zeigt sich, dass die schlechtesten Ergebnisse für Ebene $l = \text{conv1}$ mit nur 2×2 Teilregionen erzielt werden, während die Korrelation der erhobenen mit den gemessenen Werten schnell steigt, sobald das Bild in eine größere Zahl von Regionen unterteilt wird. Die lokalen Einzugsbereiche, welche in die Berechnung von $M_{l,n}$ (vgl. Gleichung 3.2) einfließen, werden damit kleiner, was zu Folge hat, dass in $M_{l,n}$ mehr Details des Bildes repräsentiert sind. Für die beschriebene erste Ebene wird ab 10×10 Teilregionen ein Plateau bei einer maximalen Korrelation von 0.80 für $n = 17$ erreicht (vgl. Abb. 3.4a).

Interessanterweise steigt die Korrelation der gemessenen Symmetriewerte mit den erhobenen Bewertungen mit höheren Schichten jenseits von conv1. Beispielsweise wird mit $l = \text{conv2}$ ein Maximum für $n = 11$ erreicht, während sich eine weitere Erhöhung der Zahl der genutzten Teilregionen negativ auf die Korrelation auswirkt. Für höherer Ebenen scheint sich dieser Trend fortzusetzen, so werden Maxima für die Ebenen conv3 und conv4 bei etwa $n = 8$ erreicht, während für die höchste Ebene conv5 die höchste Korrelation für $n = 6$ gemessen werden kann (vgl. Abbildung 3.4). Eine zu feine Unterteilung des Bildes wirkt sich dabei oberhalb von conv1 immer negativ aus.

Abbildung 3.5 zeigt die erhobenen Symmetriebewertungen in Abhängigkeit der berechneten Symmetriewerte für zwei ausgewählte Parameterkonfigurationen mit hoher Korrelation. Dabei scheinen die Werte in den Extrembereichen (hohe Symmetrie und hohe Asymmetrie) stärker zu korrelieren, während im Mittelbereich die Korrelation auf Ebene der einzelnen Bilder weniger stark ausgeprägt ist. Um zu untersuchen, ob dieser Unterschied auch für einzelne Teilnehmer des Experimentes zu beobachten ist, zeigt Abbildung 3.6 die Standardabweichung der Bewertung eines einzelnen Bildes in Abhängigkeit der Symmetriebewertungen. Der Graph entspricht einer invertierten U-Kurve, was dahingehend interpretiert werden kann, dass Menschen sich auch hier in den Extrembereichen einiger sind. Es herrscht also weitgehend Einigung darüber, ob ein Bild sehr symmetrisch oder sehr asymmetrisch ist, während die Einschätzungen für Bilder zwischen den Extremen eher auseinander gehen.

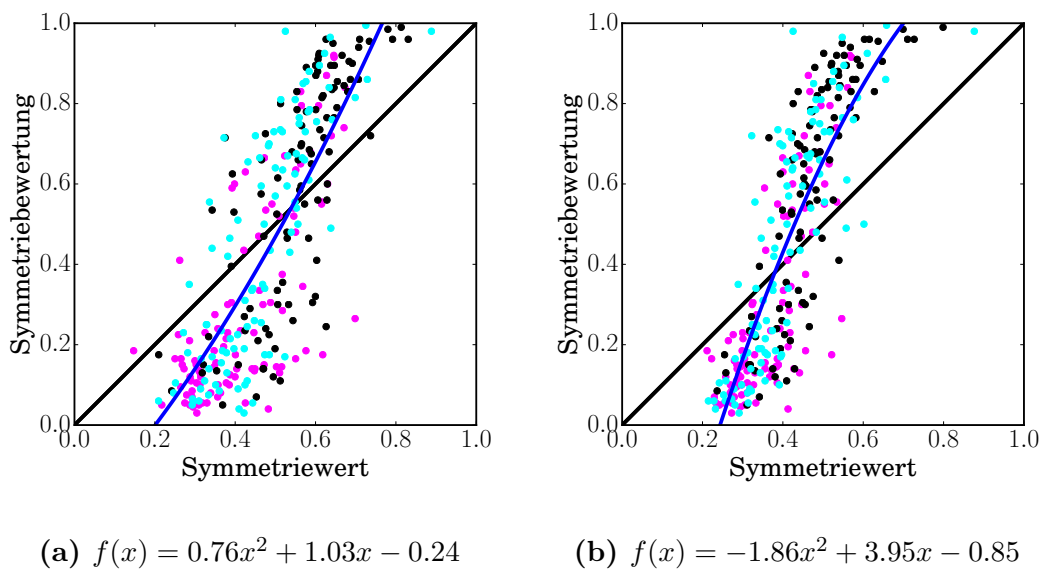


Abbildung 3.5: Punktdiagramme, welche den Zusammenhang der berechneten Symmetriewerte auf der x-Achse zu den durch Probanden erhobenen Werten auf der y-Achse darstellen (a: $l = \text{conv1}$, $n = 17$, Korrelation 0.80; b: $l = \text{conv2}$, $n = 11$, Korrelation 0.85). Jeder Punkt repräsentiert ein Bild, wobei Bilder des Genres Metal schwarz gezeigt sind, die des Genres Pop cyan und Bilder des Genres Klassik magenta. Die blaue Kurve zeigt die jeweils beste quadratische Annäherung des Zusammenhangs, die jeweilige Funktionsvorschrift ist unter den Bildern angegeben.

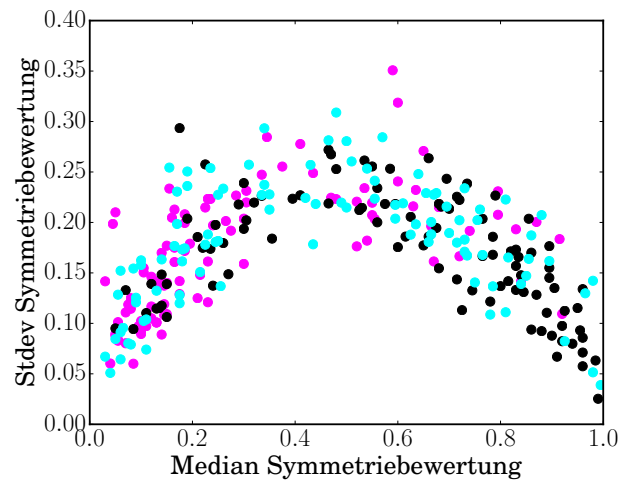


Abbildung 3.6: Standardabweichungen (Stdev) der Symmetriebewertungen der 300 getesteten CD-Titelbilder in Abhängigkeit der Bewertung der Symmetrie (schwarz: Metal, cyan: Pop, magenta: Klassik). Während in den Extrembereichen zwischen den Teilnehmern der Studie mehr Einigkeit über den vorhandenen Grad an Symmetrie eines Bildes herrscht, weichen die Urteile im mittleren Bereich stärker voneinander ab.

3.4 Diskussion der Ergebnisse

Das in diesem Kapitel beschriebene Verfahren zu Messung der Symmetrie beinhaltet als Kernidee die Verwendung der Filterantworten von CNNs, was zwei Vorteile mit sich bringt. Zum Einen kann argumentiert werden, dass die Filterantworten insofern nah am menschlichen Sehsystem sind, als dass die während des Trainings gelernten Filter zur Kantendetektion denen ähneln, welche auch auf der Retina zu finden sind (Gaborfilter, Yosinski et al. (2014) und Marelja (1980)). Farbe wird dabei in Form von Gegenfarbkontrasten oder Farbflächen verarbeitet (siehe Abbildung 3.1b). Das vorgestellte Symmetriemaß verarbeitet somit Luminanzkanten und Farben entsprechend der Statistik des Trainingsdatensatzes und entspricht damit in hohem Maße der Statistik natürlicher Bilder. Zum anderen konnten wir zeigen, dass auch die innerhalb des Netzes mit höheren Ebenen zunehmende Abstraktion der Filterantworten sinnvoll genutzt werden kann. Da auch der Mensch bei der Wahrnehmung von Symmetrie nicht nur einzelne Kanten paarweise vergleicht, sondern vielmehr die ungefähre Balance ganzer Objekte in Betracht

zieht, kommt ein Vergleich der Filterantworten höherer Ebenen diesem Umstand näher als ein bloßer Abgleich der Kanten oder Luminanzwerte. Der erste Vorteil kann erklären, warum unser Symmetriemaß bessere Ergebnisse erzielt als das von Heijer (2015) vorgeschlagene, welches strukturellen Merkmalen des Eingabebildes keine Beachtung schenkt. Die beschriebene Gliederung eines Bildes in Teilregionen macht unser Maß wesentlich robuster bezüglich derjenigen Bildinhalte, welche symmetrisch, jedoch nicht exakt gespiegelt sind. Die damit einhergehende Ungenauigkeit bezüglich der exakten Position von Bildmerkmalen ist vielmehr vorteilhaft, da auch der Mensch bei der Wahrnehmung der Symmetrie eines Bildes eine gewisse Toleranz zeigt (Wagemans, 1997). Einschränkend muss jedoch erwähnt werden, dass wir in diesem Kapitel den erfolgreichen Einsatz unseres Maßes nur für CD-Titelbilder gezeigt haben. Inwieweit unsere Ergebnisse auf andere Bildtypen übertragbar sind, muss im Einzelnen untersucht werden, insbesondere sollte für den jeweiligen konkreten Einsatz eine bewusste Wahl der Ebene und der Größe der Teilregionen getroffen werden. Während untere Ebenen des Netzes die Symmetrie eher beruhend auf *low-level* Merkmalen erfassen, ermöglichen die Merkmale höherer Ebenen eine Verlagerung der Fokussierung auf Bildinhalte.

Im beschriebenen Experiment wurde eine Eingabegröße von 512×512 Pixeln gewählt, der vorgestellte Algorithmus zur Bestimmung der Symmetrie kann jedoch auch auf andere Eingabeauflösungen übertragen werden. Beispielsweise käme eine Verkleinerung der Eingabeauflösung bei gleichbleibender Größe der Filter einer Tiefpassfilterung gleich. Diese Anpassung ermöglicht, den formalen Detailgrad der betrachteten Merkmale anzupassen, was in diesem Kapitel jedoch nicht weiter verfolgt oder näher untersucht wurde. In Abhängigkeit eines konkreten Anwendungsfalles besteht hier möglicherweise weiteres Optimierungspotential durch eine gezielte Wahl der Eingabeauflösung betrachteter Bilder.

Wie aus den Abbildungen 3.4(a)–(d) hervorgeht, ist die richtige Wahl der Größe der Teilregionen von entscheidender Bedeutung. Sind diese zu groß gewählt, nimmt die Korrelation mit den durch Probanden erhobenen Symmetriewerten ab, da zu viel lokale Details außer Acht gelassen werden. Sind die Teilregionen jedoch zu klein, werden Bereiche weniger gut zusammengefasst, was sich wiederum ebenfalls negativ auf die Korrelation erhobener mit berechneten Symmetriewerten auswirkt. Bezüglich der Teilregionen scheint es jeweils optimale Werte zu geben, welche möglicherweise auch in Abhängigkeit zu den gewählten Eingabedaten stehen.

Auf höheren Schichten scheinen größer gewählte Teilregionen bessere Ergebnisse zu erzielen. Das kann dadurch erklärt werden, dass die abstrakteren Merkmale eines Bildes, welche auf eben jenen Schichten repräsentiert sind, auch eine höhere Toleranz bezüglich ihrer Position erfordern. Entdecken Menschen beispielsweise zwei Gesichter in einem Bild, welche bezüglich ihrer Position ungefähr symmetrisch angeordnet sind, so sticht dies den Betrachtern ins Auge, während die exakte Position in Pixeln hier weniger von Bedeutung ist. Noch weniger achten Menschen im beschriebenen Fall auch auf exakte Details wie Augenfarben oder die exakte Form der Nasen, welche erwartungsgemäß eher auf unteren Schichten repräsentiert sind.

Auch wenn unser Experiment gezeigt hat, dass die Filterantworten höherer Schichten von großem Nutzen sind, können wir deren Verwendung zur Berechnung der Symmetrie nicht uneingeschränkt empfehlen. Trotz intensiver Forschung sind die Filterantworten höherer Schichten noch nicht vollständig verstanden (Yosinski et al., 2015; Zeiler und Fergus, 2014; Mahendran und Vedaldi, 2015). Von Yosinski et al. (2014) wurde untersucht, wie generisch die Filter verschiedener Schichten sind, indem die Autoren testeten, wie gut einmal auf bestimmten Datensätzen gelernte Filter auf andere Datensätze übertragbar sind. Genutzt wird dabei das auch in diesem Kapitel verwendete AlexNet. Als Ergebnis stellen die Autoren fest, dass nur die ersten beiden Schichten als wirklich generisch betrachtet werden können, während höhere Schichten spezifische, dem Trainingsdatensatz angepasste Filter herausbilden. Auch wenn wir daraus folgende, potentiell nachteilige Effekte in unseren Ergebnissen nicht beobachten konnten, darf diese Einschränkung nicht außer Acht gelassen werden und könnte sich bei anderen Anwendungen des vorgestellten Verfahrens negativ auswirken.

3.5 Zusammenfassung

In diesem Kapitel wurde ein neues Maß zur Berechnung der stetigen Symmetrie von Bildern vorgeschlagen und dessen Anwendbarkeit anhand von 300 durch Menschen bezüglich ihrer Symmetrie bewerteten CD-Titelbildern demonstriert. Eingesetzt wurden dafür die Filterantworten verschiedener Schichten eines vorher trainierten neuronalen Netzes, was in bestimmten Aspekten, wie beispielsweise

der Art der Wahrnehmung von Kanten oder der zunehmenden Abstraktion von Merkmalen, der menschlichen Wahrnehmung visueller Reize nahe kommt.

Es wurde gezeigt, dass das von uns entworfene Maß das von Heijer (2015), welches auf einem Vergleich der Pixelwerte beruht, übertrifft. Die Korrelation mit den durch ein Probandenexperiment erhobenen Symmetriewerten konnte zusätzlich erhöht werden, indem höhere Schichten mit abstrakteren Bildmerkmalen als Berechnungsgrundlage genutzt wurden. Inwiefern dieses Ergebnis auf weitere Datensätze übertragbar ist, bleibt offen. Für beliebige Bilder empfehlen wir eine Beschränkung auf die unteren beiden Ebenen, für welche es gute Gründe gibt anzunehmen, dass diese wirklich generisch sind (Yosinski et al., 2014).

Im Kontext der vorliegenden Arbeit wurden in diesem Kapitel methodische Grundlagen gelegt, auf welche in den folgenden Kapiteln zurückgegriffen wird. Symmetrieberechnungen sind nur ein möglicher Anwendungsbereich der in Abschnitt 3.2.1 beschriebenen Bildrepräsentationen in Form maximaler Filterantworten über Teilbereiche eines Bildes. In den folgenden Kapiteln wird demonstriert, dass diese auch für andere Methoden der Bildanalyse von großem Nutzen sein können.

Kapitel 4

Selbstähnlichkeit

In Mustern der Natur, wie auch in Bildern natürlicher Szenen, ist *Selbstähnlichkeit* ein häufig auftretendes Merkmal. Verallgemeinert gesagt misst Selbstähnlichkeit, zu welchem Grad Merkmale im Großen auch im Kleinen zu finden sind. In natürlichen Bildern wurde Selbstähnlichkeit auf verschiedene Art und Weisen betrachtet. So wurde beispielsweise gezeigt, dass diese insofern skalierungsinvariant sind, als dass ihre Frequenzen f in der Amplitude entsprechend $1/f^p$ abfallen. Aufgetragen im log-log plot ist dieser Abfall linear, mit einer typischen Steigung von $p \approx -1$ (Tolhurst et al., 1992; Simoncelli und Olshausen, 2001). Auch wurde die fraktale Struktur von Bildern untersucht, wobei üblicherweise die *Box-Counting*-Methode (Peitgen et al., 1992) oder deren Erweiterungen (Sarkar und Chaudhuri, 1994) zum Einsatz kommt. Man spricht in diesem Zusammenhang von der *fraktalen Dimension* von Bildern, und es wurde gezeigt, dass diese die ästhetische Präferenz von Bildern beeinflussen kann (Hagerhall et al., 2004). Amirshahi et al. (2012) schlagen eine Methode vor, um Selbstähnlichkeit natürlicher Bilder beruhend auf der PHOG-Methode (*Pyramid Histogram of Oriented Gradients*, Bosch et al. (2007)) zu berechnen. Auffallend ist, dass eine große Teilmenge visueller Kunst bezüglich ihrer statistischen Eigenschaften den natürlichen Szenen gleicht (Redies et al., 2007b; Graham und Field, 2007) und es möglich ist, verschiedene andere Bilder anhand ihrer Selbstähnlichkeit voneinander zu unterscheiden (Braun et al., 2013; Redies et al., 2012; Graham und Redies, 2010).

Die bisher vorgeschlagenen Methoden zur Bestimmung der Selbstähnlichkeit digitaler Bilder haben den Nachteil, dass sie auf nur wenigen Merkmalen eines Bildes aufbauen. Beispielsweise fließen bei der Methode von Amirshahi et al. (2012) nur Helligkeitsgradienten in die Berechnung ein. Es ist anzunehmen, dass dabei vernachlässigte Merkmale wie beispielsweise Farbe aber durchaus relevant für die Wahrnehmung sind, insbesondere ist dies im Falle der genaueren Untersuchung von Kunst keine nebensächliche Beschränkung. In diesem Kapitel schlagen wir daher eine Methode vor, welche diesen Nachteil adressiert und auch Farbe als Merkmal in die Berechnung einfließen lässt. Wie in Kapitel 3, in welchem die Symmetrie von Bildern anhand der Filterantworten künstlicher neuronaler Netze gemessen wurde, greifen wir auch hier auf diese zurück. Diese Vorgehensweise hat den Vorteil, dass sie mit Filterantworten von CNNs dem menschlichen Sehsystem näher kommt, als es eine alleinige Betrachtung der Luminanzkanten eines Bildes vermag (vgl. Abschnitt 3.4). Folgende Fragen leiten dabei die hier vorgestellten Untersuchungen:

- Sind die Filterantworten von CNNs geeignet, ein Maß für die Selbstähnlichkeit von Bildern zu definieren?
- Wie vergleicht sich ein solches Maß mit dem auf Luminanzkanten beruhendem Maß von Amirshahi et al. (2012)?
- Kann die mit tieferen Schichten des Netzes einhergehende, zunehmend abstrakte Repräsentation von Bildmerkmalen vorteilig genutzt werden?

Der Aufbau dieses Kapitels ist dabei wie folgt gegliedert. Im nächsten Abschnitt wird zunächst das von Amirshahi et al. (2012) vorgeschlagene Maß vorgestellt und im Anschluss daran ein Maß zur Bestimmung der Selbstähnlichkeit entworfen, welches auf den Filterantworten tiefer neuronaler Netze beruht. Nachdem der hier zur Testung des Maßes genutzte Datensatz eingeführt wurde, werden die Ergebnisse der Selbstähnlichkeitsberechnungen auf diesem Datensatz präsentiert, was einen Vergleich des vorgeschlagenen Maßes mit dem von Amirshahi et al. (2012) ermöglicht. Die aufgeworfenen Fragen werden im letzten Abschnitt dieses Kapitels beantwortet.

4.1 PHOG-Selbstähnlichkeit

Das Maß zur Bestimmung der Selbstähnlichkeit, wie es von Amirshahi et al. (2012) vorgeschlagen wurde, basiert auf der PHOG-Repräsentation eines Bildes. Dabei handelt es sich um eine pyramidale Darstellung der Histogramme der orientierten Gradienten (*Pyramid Histogram of Oriented Gradients (PHOG)*; Bosch et al. (2007)) eines Bildes, welche die Histogramme orientierter Gradienten (*Histogram of Oriented Gradients (HOG)*; Dalal und Triggs (2005)) um verschiedene Auflösungen erweitert. HOGs wurden ursprünglich für die Objekterkennung entwickelt, während PHOGs als Methode für die Bewertung der Ähnlichkeit zweier Bilder entworfen wurde.

Folgender Algorithmus zur Bestimmung der Selbstähnlichkeit wurde von Amirshahi et al. (2012) vorgeschlagen: Zunächst werden die HOG-Merkmale des in Grauwerte konvertierten Eingabebildes bestimmt, wobei als Resultat ein Histogramm mit k Klassen entsteht, in welchem die Antworten über dadurch definierte Orientierungen zusammengefasst werden. Dieses Histogramm wird im weiteren Verlauf als *Level-0*-Histogramm bezeichnet. Anschließend wird das Bild in 2×2 gleich große Teilregionen zerlegt, und wiederum deren HOG-Merkmale und die entsprechenden Histogramme berechnet, welche als *Level-1*-Histogramme bezeichnet werden. Der Schritt der Teilung kann beliebig oft wiederholt werden, eine Begrenzung ist dabei nur durch die Größe des Eingabebildes gegeben. In der Folge entstehen für jedes Level L insgesamt 4^L Histogramme, welche die Gradientenrichtungen der 4^L Teilregionen eines Levels enthalten. Jedes dieser Histogramme wird anschließend normalisiert, sodass sich die Werte aller unterschiedenen Orientierungen eines Histogramms auf 1 summieren. Als Maß für die Selbstähnlichkeit eines Bildes schlagen Amirshahi et al. (2012) vor, die normalisierten Histogramme höherer Level mit denen des Level-0-Histogramms zu vergleichen. Zum Einsatz kommt dafür der *Histogram Intersection Kernel (HIK)*, Barla et al. (2003)), welcher die Ähnlichkeit zweier normalisierter Histogramme $\mathbf{h}$ und $\mathbf{h}'$ mit jeweils k Klassen anhand der Summe ihrer paarweisen Minima misst. Dieser ist wie folgt definiert:

$$K(\mathbf{h}, \mathbf{h}') = \sum_{i=1}^k \min(h_i, h'_i) \quad (4.1)$$

In der ersten Veröffentlichung ihres Ansatzes schlagen Amirshahi et al. (2012) vor, als Maß für die Selbstähnlichkeit eines Bildes den Median aller HIKs der Histogramme h_s einer Teilregion mit dem der übergeordneten Teilregion $h_{Pr(s)}$ zu berechnen. Als übergeordnete Teilregion wird die Region verstanden, in welche h_s eines Levels L auf dem nächstkleineren Level $L - 1$ mit weniger Unterteilungen des Bildes fällt:

$$M_{Se}(I, L) = \text{median}(K(h_s, h_{Pr(s)}) | s \in \text{Sections}(I, L)) \quad (4.2)$$

wobei I das Eingabebild ist, L das gewählte Level und s über alle Teilregionen auf L iteriert. In einer späteren Publikation der PHOG-Methode wurden verschiedene Ansätze untersucht, welche sich bezüglich der Wahl der miteinander verglichenen Histogramme unterscheiden (Braun et al., 2013). Wir werden im Folgenden die Methode übernehmen, in welcher alle Histogramme auf Level-3 mit dem Level-0-Histogramm verglichen werden.

Eine Erweiterung wurde von Redies et al. (2012) vorgeschlagen, welche die Methode dahingehend anpasst, dass diese auch Farbbilder als Eingabe verarbeiten kann. Dies geschieht, indem das Bild zunächst in den *Lab*-Farbraum übersetzt wird, welcher ein Bild anhand seiner Luminanzwerte (L), einer rot-grün- (a) und einer blau-gelb-Komponente (b) kodiert. Das Kantenbild $G_{\max}$ wird anschließend auf Ebene der Pixel als Maximum der Kantenbilder der drei Kanäle zusammengesetzt:

$$G_{\max}(x, y) = \max(\|\nabla I_L(x, y)\|, \|\nabla I_a(x, y)\|, \|\nabla I_b(x, y)\|). \quad (4.3)$$

Auch wenn dieses Verfahren formal die Farbgradienten der Gegenfarbkanäle einbezieht, kann hier nicht von einer Erweiterung des Verfahrens um Farbe gesprochen werden. Zum einen führt der geringe Kontrast auf Farbkanälen dazu, dass der Luminanzkanal das Bild weitestgehend dominiert, zum anderen geht die Information, welche Farbe tatsächlich an einer Stelle vorherrscht, gänzlich verloren. In folgenden Abschnitt werden wir ein Verfahren beschreiben, welches diese Probleme löst.

4.2 Material und Methoden

4.2.1 Berechnung von Selbstähnlichkeit mittels CNN-Filterantworten

Im Kapitel 3 wurde eine Methode beschrieben, welche es ermöglicht, die Symmetrie von Bildern zu berechnen. Dabei wurde auf die Filterantworten eines CNNs zurückgegriffen, welche in Abschnitt 3.1 genauer beschrieben wurden. Das in diesem Kapitel beschriebene Verfahren arbeitet ähnlich dem zur Berechnung der Symmetrie (vgl. Abschnitt 3.2.1) und greift auf das bekannte neuronale Netz (AlexNet, Krizhevsky et al. (2012)) in der Implementation von Caffe (Jia et al., 2014) zurück, womit wir die bereits genannten, damit verbundenen Vorteile nutzen können: Erstens sind die Filter insoweit dem menschlichen Sehsystem ähnlich, als dass diese auf den unteren Ebenen generische Kanten- und Farbdetektoren herausbilden (siehe Diskussion in Abschnitt 3.4), welche der Luminanz und Farbverteilung natürlicher Bilder folgen (siehe Abbildung 3.1). Zweitens sprechen Filter höherer Schichten auf abstraktere Bildinhalte an, welche aus den Antworten der darunter liegenden Schichten zusammengesetzt sind. Hier erschließen sich verglichen mit dem Ansatz von Amirshahi et al. (2012) neue Möglichkeiten zur Messung der Selbstähnlichkeit.

Wie auch im letzten Kapitel verwerfen wir auch hier die vollverbundenen Schichten an der Spitze des Netzes, was uns erlaubt, die Größe des Eingabebildes frei zu wählen. In Abhängigkeit der verwendeten Datensätze, welche in Abschnitt 4.2.2 beschrieben werden, wurde die Auflösung der Eingabebilder auf 600×800 Pixel gesetzt, kann jedoch prinzipiell entsprechend einer konkreten Anwendung frei gewählt werden. Zunächst wird dem Netz ein Eingabebild präsentiert und durch dieses vorwärts propagiert. Wir erhalten so auf den verschiedenen Ebenen eine feste Anzahl von Filterantworten, auf welchen wir, analog zu dem in Abschnitt 3.2.1 vorgestellten Verfahren, die Arrays $M_{l,n}$ berechnen, welche die maximalen Filterantworten über $n \times n$ gleich große Teilregionen auf Ebene l enthalten. Während im letzten Kapitel die Arrays $M_{l,n}$ direkt verwendet werden, wird hier als weiterer Zwischenschritt jedoch eine Normalisierung der maximalen Filterantworten vorgenommen, sodass diese sich für jede Teilregion des Bildes auf 1 summieren. Es gilt somit:

$$\forall x, y : \sum_f M_{l,n}(x, y, f) = 1. \quad (4.4)$$

Diese Anpassung ist nötig, da zur Berechnung der Selbstähnlichkeit auf den Histogram Intersection Kernel zurückgegriffen wird (vgl. Gleichung 4.1), welcher für normalisierte Histogramme definiert ist. Für eine gegebene Ebene $l \in \{\text{conv1}, \dots, \text{conv5}\}$ berechnet sich die Selbstähnlichkeit eines Bildes als der Median der Ergebnisse der Histogrammvergleiche eines Levels L mit $2^L \times 2^L$ Teilregionen mit dem Level-0 der selben Ebene wie folgt:

$$S(M_{l,n}, M_{l,1}) = \text{median} \left(K(M_{l,n}(x, y), M_{l,1}(1, 1)) \mid x, y \in 1..n \right), \quad (4.5)$$

wobei $M_{l,n}$ und $M_{l,1}$ die normalisierten Arrays sind, welche die Histogramme der Teilregionen enthalten. Auf Level L wird ein Bild demzufolge in $2^L \cdot 2^L = n \cdot n$ Teilregionen untergliedert. Den Autoren Braun et al. (2013) folgend wird auch hier für den Vergleich Level-3 gewählt, es ergibt sich also mit $L = 3$ der Median aus insgesamt $n \cdot n = 2^L \cdot 2^L = 2^3 \cdot 2^3 = 64$ Histogrammvergleichen. Der Wertebereich des vorgestellten Maßes liegt im Bereich $[0, 1]$, wobei ein Wert von 1 für eine hohe Selbstähnlichkeit eines Bildes steht, 0 hingegen für eine geringe Selbstähnlichkeit.

4.2.2 Datensatz zur Evaluation

Wir verwenden in diesem Kapitel eine Teilmenge der bereits in Kapitel 2 vorgestellten Datensätze, welche sich sowohl in natürliche als auch in nicht-natürliche Muster aufteilen. Die Kategorien natürlicher Muster enthalten Landschaften (*Large Vistas*), Wolken (*Clouds*), Pflanzenmuster (*Plant Patterns*), Äste (*Branches*) und Flechten (*Lichen*). Nicht-natürliche Muster sind anhand der Kategorien Objekte (*Objects*) und Städteszenen (*Urban Scenes*) repräsentiert. Ein Vergleich mit Kunstbildern wird anhand des ebenfalls bereits vorgestellten *JenAesthetics*-Datensatzes (Amirshahi et al., 2013a) möglich. Für die Quellen einzelner Datensätze sei auf Tabelle 2.1 verwiesen. Auch in diesem Kapitel werden im Folgenden die englischen Bezeichnungen der Kategorien verwendet, wie diese von Redies et al. (2012) eingeführt wurden. Abbildung 4.1 zeigt jeweils ein Beispiel für die verschiedenen Bildkategorien.

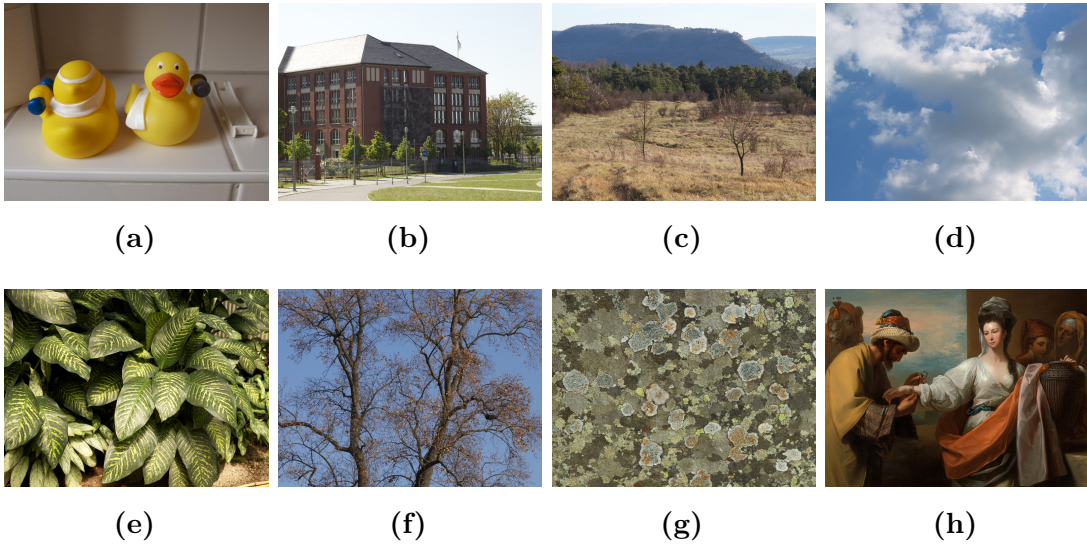


Abbildung 4.1: Beispielbilder der insgesamt acht analysierten Kategorien. (a) Objekte (*Objects*), (b) Städteszenen (*Urban Scenes*), (c) Landschaften (*Large Vistas*), (d) Wolken (*Clouds*), (e) Pflanzenmuster (*Plant Patterns*), (f) Äste (*Branches*), (g) Flechten (*Lichen*) und (h) ein Beispielbild des *JenAesthetics* Datensatzes. Siehe Tabelle A.1 des Anhangs für Angaben bezüglich abgebildeter Kunstwerke.

4.3 Ergebnisse

Im Folgenden sollen die Ergebnisse der Berechnung der Selbstähnlichkeit auf den in Abschnitt 4.2.2 eingeführten Bildkategorien vorgestellt werden. Wie auch bei der Methode zur Berechnung der Symmetrie (vgl. Kapitel 3) mache ich mir dabei zu nutze, dass neben den Filterantworten der ersten Schicht auch die Antworten höherer Schichten des AlexNet als Berechnungsgrundlage für Formel 4.5 dienen können.

4.3.1 CNN-Selbstähnlichkeit der ersten Schicht

Abbildung 4.2 zeigt Bilder, welche den Vorteil des in Abschnitt 4.2.1 beschriebenen Verfahrens, fortan bezeichnet als CNN-Selbstähnlichkeit, gegenüber der PHOG-Selbstähnlichkeit illustriert. Die Beispielbilder wurden bewusst so gewählt, dass diese die Stärken des vorgestellten Verfahrens gegenüber der PHOG-Selbstähnlichkeit betonen. So zeigt Abbildung 4.2a ein Bild, welches bezüglich der Kanten sehr selbstähnlich ist, was sowohl durch die PHOG- als auch durch die CNN-Selbstähnlichkeit der ersten Schicht ($l = \text{conv1}$) gemessen wird. Verändert man das Bild dahingehend, dass die lokale Struktur der Kanten erhalten bleibt, global jedoch eine prominente Farbverschiebung im Bild zu sehen ist, sinkt der Wert der CNN-Selbstähnlichkeit ab, die PHOG-Selbstähnlichkeit bleibt hingegen nahezu unverändert (vgl. Abb. 4.2b). Dasselbe kann für Bilder beobachtet werden, in welchen statt der Farbe die Frequenz der Kanten verändert wird (vgl. Abbildungen 4.2c und 4.2d). Die Beispiele zeigen, dass das hier vorgeschlagene Maß für Selbstähnlichkeit die Einschränkungen des PHOG-Maßes erfolgreich adressiert. Abbildung 4.3 zeigt eine Gegenüberstellung der Selbstähnlichkeitswerte der acht in Abschnitt 4.2.2 vorgestellten Kategorien, berechnet nach der PHOG-Methode (Abb. 4.3a) und nach der vorgestellten CNN-Methode (Abb. 4.3b). Durchschnittliche Selbstähnlichkeitswerte und die Standardabweichungen sind in Tabelle 4.1 gegeben. Für die natürlichen Muster *Lichen*, *Branches* und *Plant Patterns*, von denen eine hohe Selbstähnlichkeit zu erwarten ist, wird diese auch in Form hoher CNN-Selbstähnlichkeitswerte widerspiegelt. Bilder der Kategorie *Objects* hingegen zeigen im Schnitt eine geringe Selbstähnlichkeit sowohl mit der PHOG- als auch der CNN-Methode. Die durchschnittliche Selbstähnlichkeit über alle Kategorien mit der CNN-Methode erhöht sich leicht von 0.812 auf 0.835 gegenüber der

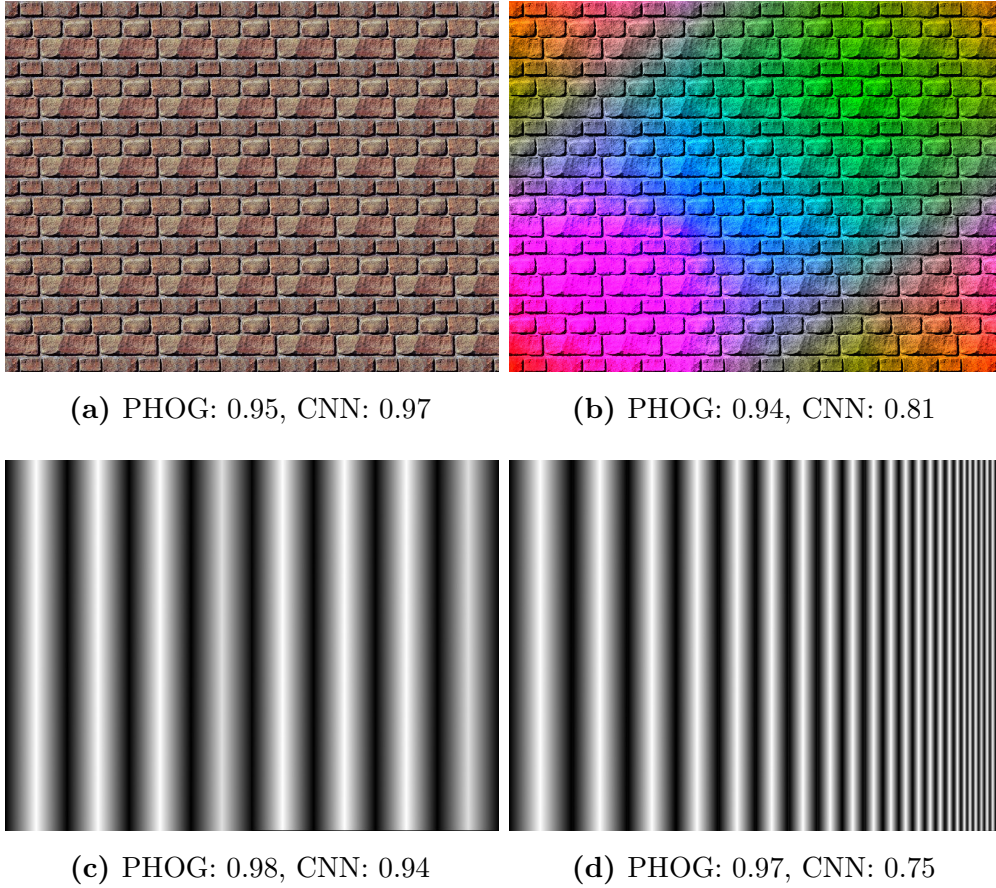
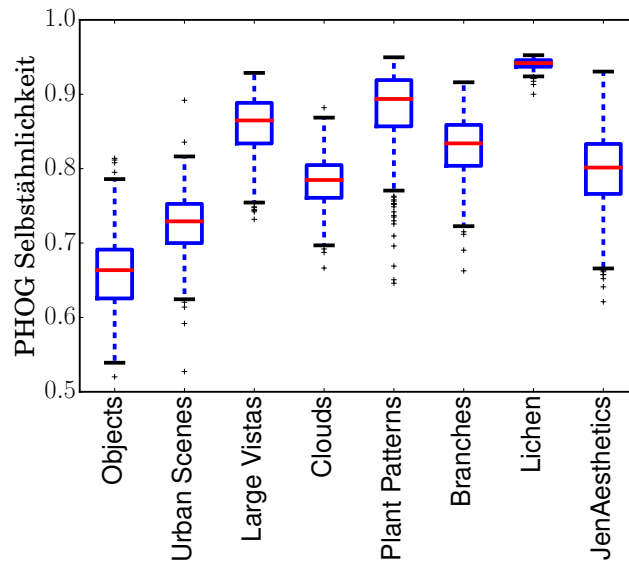


Abbildung 4.2: Beispielbilder, welche den Vorteil der Nutzung von CNN-Filterantworten gegenüber PHOG-Merkmalen zur Messung der Selbstähnlichkeit illustrieren. (a) In der oberen Reihe ist ein Bild gezeigt, das repetitive lokale Strukturen aufweist, beide Selbstähnlichkeitsmaße (siehe Bildunterschriften) spiegeln dies wider. (b) Eine deutlich sichtbare, globale Verschiebung der Farbe wird jedoch nur durch das CNN-Maß registriert, das auf 0.81 abfällt. (c) Gleiches kann in der unteren Reihe beobachtet werden, wo anstatt der Farbe die Frequenz der Luminanzkanten eine Veränderung erfährt. (d) Während die CNN-Selbstähnlichkeit auch hier abfällt, bleibt die PHOG-Selbstähnlichkeit nahezu unverändert.

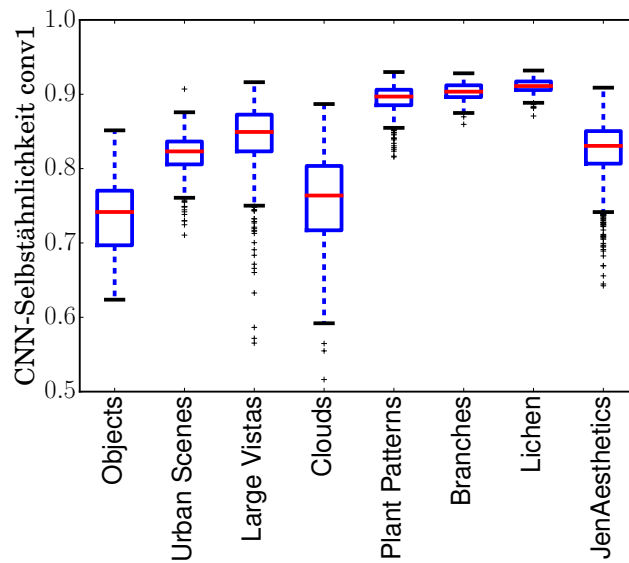
PHOG-Methode (vgl. Tabelle 4.1). Eine Erklärung dieser Erhöhung könnte in der Beachtung der Farbe liegen, welche die CNN-Methode von PHOG-Methode unterscheidet, in vielen Bildern tendiert Farbe dazu, homogener verteilt zu sein, selten ist das gesamte Spektrum in einem Bild gleichzeitig vertreten. Dieser Umstand ist insbesondere für natürliche Muster zu beobachten, in denen Grüntöne oft das Bild dominieren. Ein vergleichsweise hoher Anstieg der Selbstähnlichkeit ist auch für *Branches* zu beobachten, wo die Äste oft vor einem blauen Hintergrund abgebildet sind, welcher ein Bild insgesamt etwas selbstähnlicher erscheinen lässt (vgl. Abbildung. 4.4b). Dasselbe ist für Bilder der Kategorie *Urban Scenes* zu beobachten, auch hier bestimmen große, einfarbige Himmelsregionen das Bild. Interessant ist der Vergleich beider Maße für die Kategorie *Clouds*. Während sich die gemessenen Werte von der PHOG- zur CNN-Selbstähnlichkeit im Durchschnitt nur leicht verringern, ist ein deutlicher Anstieg der Standardabweichung zu beobachten. Das kann auf die Heterogenität des Datensatzes zurückgeführt werden: Während Schäfchenwolken ein Bild selbstähnlich machen (vgl. Abb. 4.4c), führen große Wolken mit ihren Schattierungen oft zu größeren Farbunterschieden im Bild (vgl. Abb. 4.5c). Weitere Beispiele, die PHOG- und CNN-Selbstähnlichkeit gegenüber stellen, sind in den Abbildungen 4.4 und 4.5 gegeben. Für ein Signifikanzniveau von $p < 0.001$ (Kruskal-Wallis' H-Test und Dunns Post-Test) unterscheidet sich die CNN-Selbstähnlichkeit von Kunst signifikant von allen Kategorien außer von den Bildern, die Städteszenen zeigen (*Urban Scenes*, $p = 0.020$). Bezüglich der PHOG-Selbstähnlichkeit unterscheidet sich Kunst von allen getesteten Kategorien.

4.3.2 CNN-Selbstähnlichkeit höherer Schichten

So wie es sich in Kapitel 3 für ein Maß abstrakterer Symmetrie als sinnvoll erwiesen hat, auch die höheren Schichten conv2 bis einschließlich conv5 als Berechnungsgrundlage zu nutzen, soll auch in diesem Kapitel geprüft werden, ob dies für die Bestimmung der Selbstähnlichkeit einen Nutzen mit sich bringt. Da die in Abschnitt 4.3.1 genutzten Filter der Schicht conv1 nur auf Farbflächen, Farbkanten und Luminanzkanten verschiedener Frequenzen ansprechen, ist dasselbe auch von einem Selbstähnlichkeitsmaß zu erwarten, welches auf den zugehörigen Filterantworten definiert ist. Gleichwohl kann das in Abschnitt 4.2.1 vorgestellte Verfahren auch auf die Filterantworten höherer Schichten angewendet werden. Tabelle 4.2



(a)



(b)

Abbildung 4.3: Vergleich der Selbstähnlichkeitswerte berechnet auf Grundlage von (a) PHOG-Merkmalen und (b) CNN-Filterantworten auf Schicht conv1. Für einzelne Kategorien bringt die Betrachtung der Farbe eine Verschiebungen der Werte mit sich: *Branches* und *Urban Scenes* zeigen höhere Werte, da hier ein blauer Himmel oft ein Bild dominiert, was die Bilder insgesamt selbstähnlicher bezogen auf deren Farbe macht.

Tabelle 4.1: Vergleich der Selbstähnlichkeitswerte berechnet auf Grundlage der PHOG-Merkmale und der CNN-Filterantworten der Schicht conv1. In der ersten Spalte findet sich auch die Anzahl der Bilder der acht verschiedenen Bildkategorien.

Bildkategorie	Bilder	PHOG-Selbstähnlichkeit	CNN-Selbstähnlichkeit
Objects	207	0.660 ± 0.055	0.732 ± 0.049
Urban Scenes	219	0.724 ± 0.044	0.818 ± 0.029
Large Vistas	473	0.859 ± 0.041	0.840 ± 0.048
Clouds	268	0.783 ± 0.035	0.757 ± 0.064
Plant Patterns	331	0.880 ± 0.053	0.893 ± 0.020
Branches	301	0.828 ± 0.043	0.903 ± 0.011
Lichen	244	0.941 ± 0.007	0.910 ± 0.010
JenAesthetics	1629	0.798 ± 0.048	0.826 ± 0.037
Gesamt	3672	0.812 ± 0.076	0.835 ± 0.060



(a) P: 0.65, C: 0.87



(b) P: 0.66, C: 0.87



(c) P: 0.69, C: 0.80

Abbildung 4.4: Um den Unterschied zwischen der PHOG- und der CNN-Selbstähnlichkeit zu veranschaulichen, zeigt diese Abbildung Bilder, bei denen die Selbstähnlichkeitswerte weit auseinander liegen. In den hier gezeigten Fällen ist die CNN-Selbstähnlichkeit (C) deutlich höher als die Selbstähnlichkeit auf Basis von PHOG (P). Deutlich wird, dass das PHOG-Maß in den Bildern verschiedene Kantenrichtungen abstrahlt, während eine global geringe Farbvarianz höhere Selbstähnlichkeitswerte auf Basis der CNN-Filterantworten bewirkt.

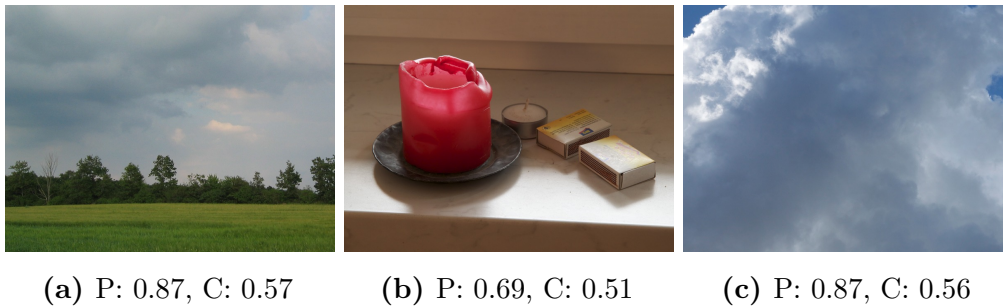


Abbildung 4.5: Analog zu Abbildung 4.4 sind Bilder gezeigt, für die die PHOG-Selbstähnlichkeit deutlich höher ist als die CNN-Selbstähnlichkeit. Wird Farbe einbezogen, lassen viele globale Farbänderungen ein Bild weniger selbstähnlich erscheinen. So unterscheidet sich in (a) die grüne Wiese stark vom grauen Himmel und in (b) die rote Kerze deutlich vom Rest des Bildes. Auch die verschiedenen Schattierungen der Wolke in (c) tragen dazu bei, dass das Bild bezüglich der Farbe weniger selbstähnlich ist als bei alleiniger Betrachtung der Helligkeitsgradienten.

zeigt die durchschnittlichen CNN-Selbstähnlichkeitswerte und Standardabweichungen berechnet auf Grundlage der Schichten conv2–conv5, eine grafische Darstellung für die Ebenen conv2 und conv3 ist in Abbildung 4.6 gegeben. Zu beobachten ist, dass die mittlere Selbstähnlichkeit von 0.835 auf Schicht $l = \text{conv1}$ auf 0.660 auf Ebene $l = \text{conv2}$ absinkt (vgl. auch Tabelle 4.2). Dieser Wert bleibt für die darüber liegenden Ebenen conv3 (durchschnittliche Selbstähnlichkeit 0.667) und conv4 (durchschnittliche Selbstähnlichkeit 0.637) nahezu unverändert, bis er sich für die letzte Ebene nahezu halbiert (durchschnittliche Selbstähnlichkeit 0.334 für conv5). Es ist alles andere als einfach, diese Ergebnisse zu interpretieren, da nicht klar ist, auf welche Bildmerkmale diese Ebenen genau ansprechen. Eine Veränderung der Selbstähnlichkeitswerte kann auch für einzelne Kategorien getrennt betrachtet werden. Wird die Selbstähnlichkeit mit den Filterantworten der ersten Schicht berechnet, zeigen *Lichen* die höchsten Selbstähnlichkeitswerte. Auf Schicht conv2 hingegen fallen diese hinter *Plant Patterns* und *Branches* zurück.

Der visuelle Unterschied der jeweils auf den einzelnen Schichten conv1 bis conv3 als am selbstähnlichsten bewerteten Bilder wird in Abbildung 4.7 deutlich. Es wurde bereits der Umstand erwähnt, dass *Lichen* auf der zweiten Ebene conv2 in der Selbstähnlichkeit hinter *Plant Patterns* und *Branches* zurückfallen. Der Grund könnte darin liegen, dass die Wiederholung von Bildinhalten, welche die

Selbstähnlichkeit begünstigt, bei *Lichen* hauptsächlich eine Wiederholung der Farben und Kanten ist. Etwas abstraktere Bildinhalte wie Blätter oder Zweige werden erst durch Ebene conv2 erfasst, was die Verschiebung der Kategorien erklären kann und in Abbildung 4.7 deutlich wird. Gleiches kann beobachtet werden, wenn man die Bilder mit der jeweils höchsten Selbstähnlichkeit betrachtet, sich dabei aber nur auf die Kategorie Kunst beschränkt (siehe Abbildung 4.8). Auf der ersten Ebene sind die drei Bilder mit der höchsten Selbstähnlichkeit ausnahmslos impressionistische und pointillistische Bilder, welche sich durch viele feine, farbenfrohe Pinselstriche auszeichnen (4.8a–c). Beim Wechsel zu conv2 dominieren Bilder aus Barock und Renaissance, welche alle drei größere Ansammlungen von Menschen zeigen. Auch hier begründen größere, abstraktere Elemente eine hohe Selbstähnlichkeit, was über bloße Farben und Pinselstriche hinausgeht.

Diese Ergebnisse werfen die Frage auf, wie sich eine weitere Abstraktion der Bildelemente jenseits der ersten beiden Schichten auf das vorgestellte Selbstähnlichkeitsmaß auswirkt. In den Abbildungen 4.7g–i und 4.8g–i sind auch für Ebene conv3 Beispielbilder gezeigt. Eine klare Aussage, wie sie bei dem Wechsel von conv1 zu conv2 getroffen werden konnte, derart, dass beobachtbare Merkmale erneut abstrakter werden, ist hier nicht ohne Weiteres abzuleiten. Vielmehr bestätigt sich das, was schon aus den Boxplots in Abbildung 4.6 und aus Tabelle 4.2 ersichtlich ist: Bilder der Kategorie *Lichen* zeigen hier wieder die durchschnittlich höchsten Werte bezüglich ihrer Selbstähnlichkeit. Auch für Kunst ist auf den ersten Blick nicht erkennbar, inwiefern eine weitere Abstraktion der Bildmerkmale einen Vorteil bringt. Zwar zeigen die Bilder durchaus repetitive Muster (vgl. Abbildungen 4.8g–i), eine klare Abgrenzung zu den Bildern einer Schicht darunter (conv2) ist jedoch schwierig.

Bei der alleinigen Betrachtung von Kunst lässt sich beobachten, dass die CNN-Selbstähnlichkeit verglichen mit der PHOG-Selbstähnlichkeit leicht zunimmt, die Standardabweichung innerhalb der Kategorie hingegen leicht abnimmt. Mit unserem Maß lässt sich Kunst in der Folge sehr viel deutlicher von den natürlichen Muster *Plant Patterns* und *Branches* abgrenzen, die deutlich höhere Selbstähnlichkeitswerte aufweisen. Für ein Signifikanzniveau von $p < 0.001$ unterscheidet sich Kunst von allen anderen Kategorien für alle Ebenen oberhalb von conv1. Die einzige Ausnahme bildet erneut die Kategorie *Urban Scenes*, welche sich bezüglich der

Selbstähnlichkeit für conv4 und conv5 nicht signifikant von Kunst unterscheidet (conv4: $p = 0.265$ und conv5: $p = 0.311$).

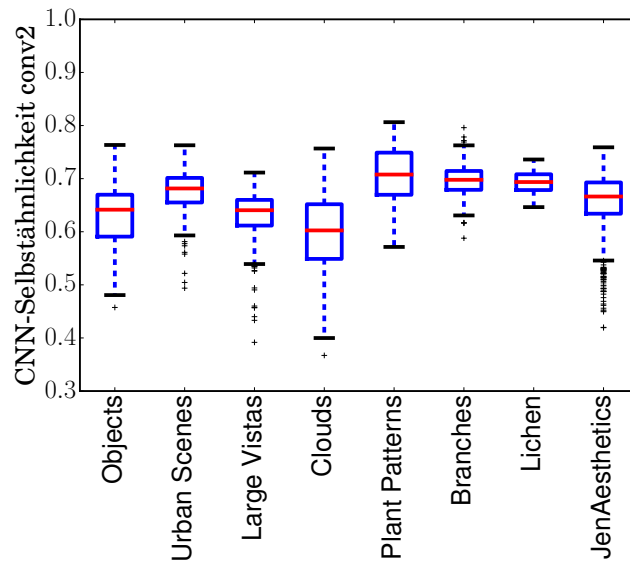
Tabelle 4.2: Selbstähnlichkeitswerte berechnet auf Grundlage der Filterantworten oberhalb der Schicht conv1.

Bildkategorie	conv2	conv3	conv4	conv5
Objects	0.630 ± 0.054	0.594 ± 0.037	0.605 ± 0.046	0.319 ± 0.051
Urban Scenes	0.674 ± 0.042	0.631 ± 0.035	0.630 ± 0.054	0.347 ± 0.056
Large Vistas	0.632 ± 0.042	0.651 ± 0.038	0.592 ± 0.041	0.270 ± 0.053
Clouds	0.599 ± 0.071	0.613 ± 0.044	0.622 ± 0.038	0.262 ± 0.031
Plant Patterns	0.708 ± 0.049	0.713 ± 0.024	0.689 ± 0.048	0.400 ± 0.078
Branches	0.698 ± 0.029	0.698 ± 0.012	0.666 ± 0.022	0.369 ± 0.036
Lichen	0.693 ± 0.019	0.728 ± 0.012	0.667 ± 0.021	0.363 ± 0.055
JenAesthetics	0.658 ± 0.050	0.669 ± 0.037	0.637 ± 0.054	0.341 ± 0.076
Gesamt	0.660 ± 0.056	0.667 ± 0.048	0.637 ± 0.054	0.334 ± 0.076

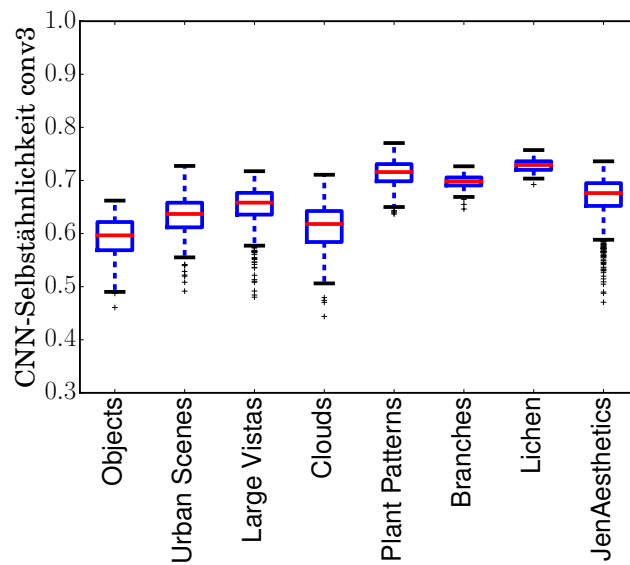
4.4 Diskussion der Ergebnisse

Die Ergebnisse zeigen, dass die Filterantworten der ersten beiden Schichten conv1 und conv2 in sinnvoller Art und Weise als Grundlage für ein Maß zur Bestimmung der Selbstähnlichkeit eingesetzt werden können. Gegenüber der nur auf Luminanzkanten definierten PHOG-Selbstähnlichkeit ist die Verwendung der Filterantworten der ersten Schicht conv1 vorteilhaft, da auch Farben und Kanten verschiedener Frequenzen einbezogen werden. Filterantworten der Schicht conv2 beeinflussen das Selbstähnlichkeitsmaß dahingehend, dass sich die Selbstähnlichkeit damit vor allem auf repetitive Strukturen im Bild bezieht, welche von reinen Kanten oder Farbflächen abstrahieren.

Für Schichten oberhalb von conv2 konnte kein klar ersichtlicher Vorteil festgestellt werden. Zwar finden sich auch in Bildern, welche auf Grundlage von conv3-Filtern als selbstähnlich definiert werden, repetitive Strukturen, eine klare Abgrenzung zu den Schichten darunter ist jedoch nicht zu erkennen. Verwiesen sei daher auch an dieser Stelle noch einmal auf den Punkt, welcher schon in der Diskussion der Ergebnisse des Symmetriemaßes (vgl. Abschnitt 3.4) angesprochen wurde.



(a)



(b)

Abbildung 4.6: Boxplots der CNN-Selbstähnlichkeit berechnet auf Basis der Schichten (a) conv2 und (b) conv3 für die acht getesteten Bildkategorien. Für die Selbstähnlichkeitswerte höherer Schichten sei auf Tabelle 4.2 verwiesen.

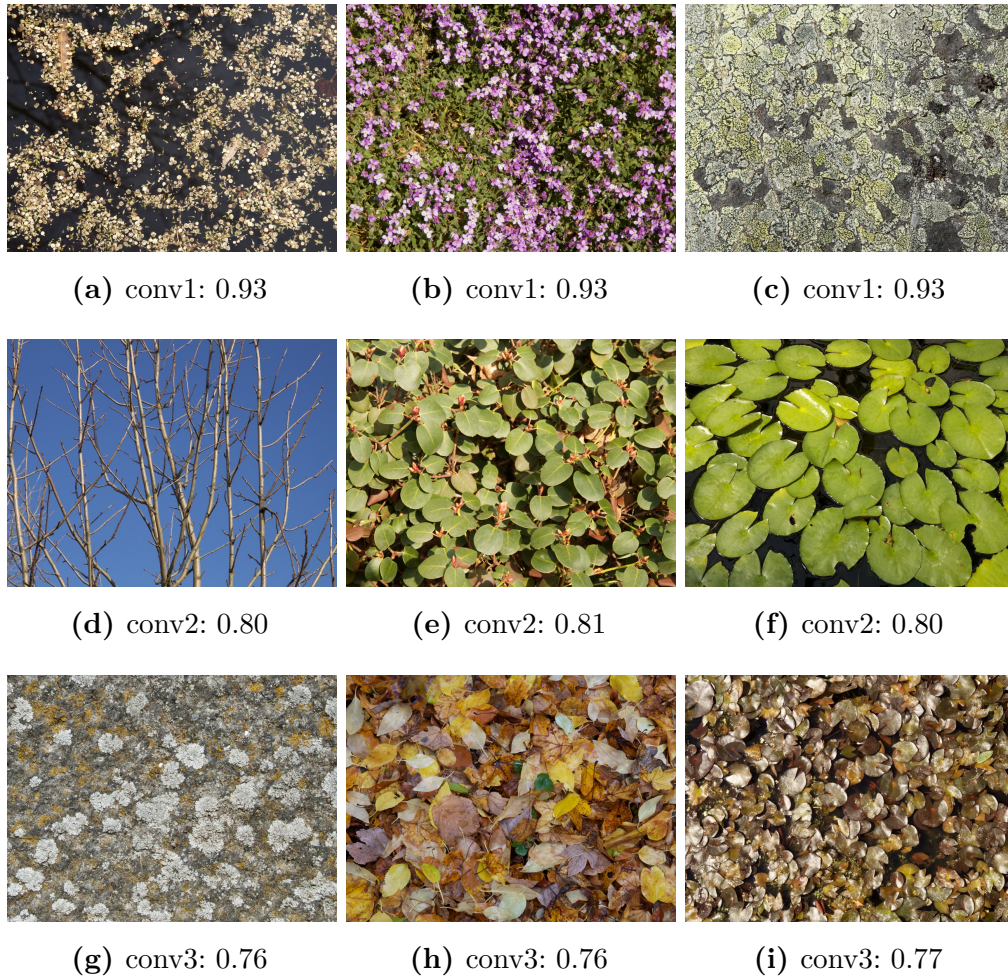


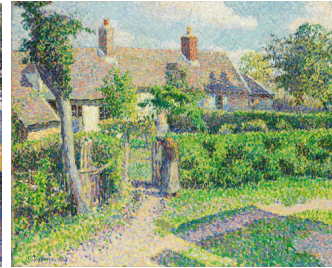
Abbildung 4.7: Gezeigt sind die drei Bilder mit der jeweils höchsten Selbstähnlichkeit berechnet auf (a)-(c) conv1, (d)-(f) conv2 und (g)-(i) conv3. Auf conv1 bestimmen vor allem feine Strukturen und Farben das Bild, während sich auf conv2 abstraktere Strukturen wie Zweige oder Blätter im Bild wiederholen. Auf höheren Ebenen ist nicht klar erkennbar, welche Verbesserung eine weitere Abstraktion mit sich bringt, auch wenn die Bilder weiter repetitive Strukturen aufweisen.



(a) conv1: 0.90



(b) conv1: 0.91



(c) conv1: 0.91



(d) conv2: 0.75



(e) conv2: 0.75



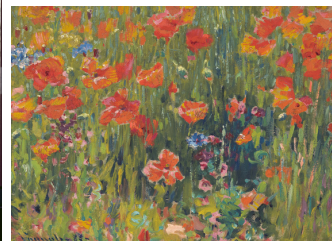
(f) conv2: 0.76



(g) conv3: 0.73



(h) conv3: 0.73



(i) conv3: 0.74

Abbildung 4.8: Analog zu Abbildung 4.7 sind die Kunstbilder mit der jeweils höchsten Selbstähnlichkeit berechnet auf Grundlage der Filterantworten auf (a)-(c) conv1, (d)-(f) conv2 und (g)-(i) conv3 gezeigt. Auch für die Kunstbilder lässt sich eine Verschiebung des Fokus von feinen Strukturen auf conv1 als ausschlaggebend für die Selbstähnlichkeit hin zu abstrakteren Bildinhalten auf conv2 und conv3 beobachten.

Während die ersten beiden Schichten des AlexNet-Modells als generisch und von einem konkreten Datensatz unabhängig angesehen werden können, sind höhere Schichten spezifischer für verschiedene Datensätze (Yosinski et al., 2014). Auch ein auf den Antworten höherer Schichten definiertes Maß für Selbstähnlichkeit könnte somit spezifisch für Objekte sein, welche in dem Trainingsdatensatz vertreten waren und das Selbstähnlichkeitsmaß damit auf diese beschränken. Von der Verwendung der Filterantworten auf conv3–conv5 wird daher abgeraten, während die Filterantworten auf conv1 oder conv2 als Grundlage der Berechnung ohne offensichtliche Probleme verwendet werden können.

Zu beachten ist auch, dass unser Maß der Selbstähnlichkeit als ein Maß für repetitive Bildelemente verstanden werden sollte. An dieser Stelle muss es dabei zu Fraktalität abgegrenzt werden. Ein echtes Fraktal zeichnet sich dadurch aus, dass es bei beliebiger Vergrößerung aus Teilen seines Selbst besteht. Damit ist ein Fraktal zwangsläufig selbstähnlich. Das Maß für Selbstähnlichkeit von Bildern, wie wir es hier vorgestellt und getestet haben, misst diese Art der Selbstähnlichkeit nicht, ein von uns als hoch selbstähnlich bestimmtes Bild ist also mitnichten ein Fraktal. Vielmehr quantifizieren wir mit unserem Maß, zu welchem Anteil Strukturen, die global im Bild zu finden sind, auch lokal über das Bild verteilt auftreten.

Eine Einschränkung unserer Untersuchung ist die fehlende Evaluation durch Probandenstudien, wie sie beispielsweise für die Testung unseres Maßes für Symmetrie durchgeführt wurde (siehe Kapitel 3; Brachmann und Redies (2016)). Erst durch ein solches Experiment kann bestätigt werden, dass unser Maß das Verständnis menschlicher Beobachter bezüglich der Selbstähnlichkeit eines Bildes abbildet. Da sich in diesem Kapitel hauptsächlich auf das etablierte Maß von Amirshahi et al. (2012) bezogen wurde, erfolgte die Evaluation des vorgestellten Maßes analog zu deren Methode durch die Bewertung der Ergebnisse anhand verschiedener Datensätze und Bilder selbst. Auch ist „Selbstähnlichkeit“ ein bei Weitem weniger geläufiger Begriff als etwa „Symmetrie“. Die Verwendung dieses Begriffes in Testungen an Probanden dürfte sich somit als ungleich schwieriger gestalten.

4.5 Zusammenfassung

Die eingangs gestellte Frage, ob die Filterantworten eines tiefen neuronalen Netzes für die Definition eines Maßes der Selbstähnlichkeit eines Bildes eingesetzt

werden können, wird für die Filterantworten der Schichten conv1 und conv2 bejaht. Einschränkend muss jedoch gesagt werden, dass für höhere Schichten kein Nutzen gezeigt werden konnte und auch Untersuchungen über die Spezifität höherer Schichten dagegen sprechen, diese zu verwenden. Ob Antworten der ersten Schicht conv1 oder Antworten der zweiten Schicht conv2 eingesetzt werden soll, hängt von der konkreten Anwendung ab. Ist vor allem eine Farb- oder Strukturgleichheit von Interesse, sind die Antworten der ersten Schicht eine gute Wahl. Sollen hingegen einfache, repetitive Bildelemente im Fokus stehen, wird zu einer Berechnung auf Grundlage der Filterantworten auf Schicht conv2 geraten. Damit konnte auch die eingangs gestellte Frage nach einem etwaigen Nutzen tieferer Schichten beantwortet werden. Verglichen mit den PHOG-Maß zur Bestimmung der Selbstähnlichkeit eines Bildes konnten wir anhand synthetischer Bilder zeigen, dass dessen Schwächen durch unser Maß adressiert werden und sowohl Farbe als auch die verschiedenen Frequenzen von Luminanzkanten Beachtung finden, was als Verbesserung gesehen werden kann. Insbesondere für eine Betrachtung von Kunst ist dies unabdingbar, leben viele Bilder doch gerade von der Farbkomposition. Ein genauer Vergleich der CNN- mit der PHOG-Selbstähnlichkeit wurde anhand bewusst gewählter Bildkategorien vorgenommen, deutliche Änderungen wurden anhand von Beispielbildern demonstriert und erklärt.

Der Nutzen des von uns vorgestellten Maßes schließt sich direkt an die in der Vergangenheit gefunden Anwendungsgebiete für das Maß von Amirshahi et al. (2012) an. In Kapitel 1 haben wir dargestellt, dass der in frühen Tagen verfolgte Weg der Bildanalyse anhand spezifischer Bildmerkmale für die Experimentelle Ästhetik von großen Wert sind. Das in diesem Kapitel vorgestellte Maß für Selbstähnlichkeit kann als spezifisches Merkmal eingesetzt werden, um in Kombination mit anderen Maßen die Ästhetik von Bildern zu quantifizieren. Das größere Potential liegt jedoch im Einsatz in psychophysischen Experimenten, wie durchgeführt beispielsweise von Mullin et al. (2017), in welchen der Zusammenhang der Selbstähnlichkeit mit der ästhetischen Präferenz oder anderen Empfindungen wie Harmonie, Interesse, Komplexität oder der Ordnung im Bild (Redies et al., 2015) untersucht wird.

Kapitel 5

Varianz von CNN-Filterantworten für Kunst

In den beiden vorangegangenen Kapiteln wurde eine bestimmte Kodierung der Filterantworten verschiedener Ebenen eines tiefen neuronalen Netzes verwendet, um Symmetrie und Selbstähnlichkeit eines Bildes zu messen. Symmetrie ist dabei ein fester Begriff, über dessen Bedeutung weitestgehend Konsens herrscht, auch wenn die subjektiven Bewertungen eines Bildes mitunter verschieden ausfallen können. Selbstähnlichkeit wurde als ein Maß der Wiederholung von Struktur einerseits in Form einfacher Merkmale wie Kanten und Farben, andererseits aber auch der einfacher Objekte verstanden. In beiden Kapiteln wurde somit ein Maß ausgehend von feststehenden Begriffen entworfen.

Im Gegensatz dazu wird in diesem Kapitel der Begriff der *Varianz* von Bildmerkmalen als Ausgangspunkt für eine explorative Untersuchung von Kunst genommen. Ziel ist die Entwicklung von Maßen, die es ermöglichen, verschiedene Arten der Varianz von Bildstruktur zu quantifizieren. Geleitet werden die Untersuchungen dabei weniger von der Frage „Wie verhält sich Kunst in Bezug auf andere Kategorien?“ als vielmehr „Welche Eigenschaften grenzen Kunst von den anderen Kategorien ab?“. Während beide Fragen zwar voneinander nicht unabhängig sind, zielt die letztere dennoch mehr auf den in Abschnitt 1.2.1 vorgestellten Begriff der *Signifikanten Form* (Bell, 1914) von Kunst. Dieser beschreibt die Annahme, dass Kunst in seiner Form bestimmten, vom Künstler bewusst oder unbewusst verfolgten Regeln der Komposition folgt, ohne dass diese Regeln jedoch genauer

spezifiziert werden. Der in diesem Kapitel verfolgte Ansatz nähert sich diesem Begriff über die Varianz von Bildmerkmalen.

Im zweiten Kapitel dieser Arbeit wurde beobachtet, dass die modernen Methoden der Bildanalyse keinen einfachen Zugang für die Fragestellungen der Experimentellen Ästhetik bieten, während sich beide Gebiete in früheren Jahren gegenseitig ergänzen konnten. In diesem Kapitel wird der Versuch unternommen, eine Verbindung zu schaffen und eine Fragestellung der Experimentellen Ästhetik unter Zuhilfenahme moderner Methoden der Bildanalyse zu beantworten. Insbesondere ist dabei das Ziel, aus den Ergebnissen neue Aussagen über die Form von Kunst im Gegensatz zu anderen Bildkategorien herausarbeiten zu können.

5.1 Material und Methoden

5.1.1 Varianz von CNN-Filterantworten

Wie für das Maß der Selbstähnlichkeit (vgl. Kapitel 4) verwenden wir auch hier die in Kapitel 3 entworfene Kodierung der CNN-Merkmale, welche die maximalen Filterantworten über definierte Teilregionen eines Bildes zusammenfasst. Erneut greifen wir dabei auf das von Krizhevsky et al. (2012) vorgestellte AlexNet zurück. Um nicht zu viel Detailinformation eines Bildes zu verlieren, skalieren wir die Eingabeauflösung auf 512×512 Pixel, nachdem wir die oberen, vollverbundenen Schichten verworfen haben. Für eine gegebene Schicht l werden die Filterantworten in $n \times n$ gleich große Regionen unterteilt, für welche die maximalen Filterantworten eines jeden der k Filter auf der entsprechenden Schicht l festgestellt werden. Als Ergebnis erhalten wir das Array $M_{l,n}$. Wie auch in Kapitel 4 normalisieren wir die Einträge für jede Teilregion, Gleichung 4.4 gilt somit auch hier.

Wir definieren folgende Maße:

1. Die Gesamtvarianz aller Einträge von $M_{l,n}$:

$$p_a(M_{l,n}) = \text{var}(M_{l,n}) \quad (5.1)$$

2. Der Median der Varianzen der $n \times n$ Teilregionen über alle Filter:

$$v_g(M_{l,n}) = (\text{var}(M_{l,n}(x, y, i) \mid i \in 1..k) \mid x, y \in 1..n) \quad (5.2)$$

$$p_g(M_{l,n}) = \text{median}(v_g(M_{l,n})) \quad (5.3)$$

3. Der Median der Varianzen der Filter über alle Teilregionen:

$$v_f(M_{l,n}) = (\text{var}(M_{l,n}(x, y, i) \mid x, y \in 1..n) \mid i \in 1..k) \quad (5.4)$$

$$p_f(M_{l,n}) = \text{median}(v_f(M_{l,n})) \quad (5.5)$$

Der besseren Lesbarkeit halber werden wir $p_V(M_{l,n})$ als $P_V(n)$ für eine gegebene Schicht l bezeichnen. So steht beispielsweise $P_a(12)$ für $p_a(M_{l,12})$ und bezeichnet die Gesamtvarianz auf Ebene l bei einer Unterteilung in 12×12 Teilregionen.

Für jedes Bild in unserem Experiment berechnen wir die drei eingeführten Varianzen für die fünf Schichten conv1–conv5. Gleichzeitig unterteilen wir dabei das Bild in n Teilregionen, wobei wir mit $n = 2$ beginnen und den Parameter in Zweierschritten bis einschließlich $n = 30$ erhöhen. Insgesamt berechnen wir so $3 \cdot 15$ verschiedene Konfigurationen pro Schicht des Netzes. Zur Beantwortung der eingangs gestellten Frage, ob es Kombination der gemessenen Varianzen gibt, welche Kunst von den anderen Bildkategorien abgrenzt und damit als möglicher Kandidat der Entschlüsselung einer eventuell existierenden *Signifikanten Form* in Betracht kommt, wird jede mögliche Kombination der beschriebenen Varianzwerte auf die jeweilige Diskriminativität getestet. Eine Auswahl aus einem, zwei oder drei der 45 beschriebenen Merkmale wird dabei als Merkmalsraum einer *Support Vector Machine* (SVM; Vapnik und Chervonenkis (1974)) gesetzt und diese darauf trainiert, die Kunstbilder von anderen Bildern zu trennen. SVMs sind sogenannte *margin classifiers*, Klassifikatoren, die zwei Klassen voneinander trennen, indem diese eine Hyperebene finden, für welche die quadratische Norm der am nächsten liegenden Trainingsbeispiele (sogenannte *support vectors*) maximal ist. In ihrer Grundform werden die Daten durch eine SVM linear getrennt, im Zweidimensionalen etwa durch eine Gerade und im Dreidimensionalen durch eine Ebene. Ist das Trainingsproblem jedoch nicht linear trennbar, behilft man sich des sogenannten *Kernel-Tricks*, bei dem Trainingsbeispiele zunächst in einen höherdimensionalen Raum transformiert werden, in welchem diese schließlich linear separierbar sind.

Wir nutzen hier sogenannte *Radiale Basisfunktionen* (RBF) als nichtlineare Erweiterung. Für eine detaillierte Einführung in SVMs sei auf Bishop (2006) verwiesen. Die Güte der Klassifikation beurteilen wir wie folgt: Um eine Überanpassung an die Trainingsbeispiele zu vermeiden, nutzen wir eine 5-fache, stratifizierte Kreuzvalidierung. Dabei wird die Trainingsmenge in fünf gleich große Teilmengen unterteilt und es werden fünf Trainingsdurchläufe gestartet. *Stratifiziert* heißt in diesem Zusammenhang, dass zusätzlich die Verteilung der Klassen in den Teilmengen ungefähr der Gesamtverteilung entspricht. Trainiert wird dabei jedoch nur auf vier der fünf Teilmengen, die fünfte wird jeweils zur Berechnung der Trainingsgenauigkeit genutzt. Die Gesamtgenauigkeit errechnet sich abschließend aus dem Schnitt der Genauigkeiten aller fünf separaten Experimente. In einem eindimensionalen Merkmalsraum ergeben sich mit 45 Merkmalen 45 verschiedene Kombinationen, im zweidimensionalen Fall $\binom{45}{2} = 990$ Kombinationen und im dreidimensionalen Fall $\binom{45}{3} = 14190$ verschiedene Kombinationen. Das Ziel ist es, die für die Trennung von Kunst und anderen Bildkategorien am besten geeignete Kombination von Merkmalen zu finden, um daraus gegebenenfalls Aussagen über die besondere Form von Kunst ableiten zu können.

5.1.2 Datensatz zur Evaluation

Wir vergleichen in unserem Experiment drei Kunstkategorien mit sechs Kategorien anderer Bilder, welche teilweise schon in den Kapiteln 2 und 4 verwendet wurden. Zum einen greifen wir auf den *JenAesthetics*-Datensatz zurück, welcher insgesamt 1629 farbige Ölgemälde verschiedener Epochen westlicher Kunst in sich vereint (Amirshahi et al. (2013a); siehe Abschnitt 2.1.2 für eine Beschreibung). Ein Vorteil dieses Datensatzes ist, dass die in ihm enthaltenen Bilder bezüglich des Stils und des Inhaltes annotiert sind. Um nicht nur auf westliche Kunst beschränkt zu sein, nehmen wir in diesem Kapitel zwei weitere Kategorien mit Bildern künstlerischen Anspruchs hinzu, sodass auch Islamische Kunst (*Islamic art*, 251 Bilder) und Chinesische Kunst (*Chinese art*, 210 Bilder) Betrachtung finden. Beide Datensätze wurden von *Wikiart*¹ heruntergeladen, wobei Bilder von zu geringer Auflösung oder mit schlecht erhaltenen Farben ausgeschlossen wurden. Für eine genaue Beschreibung der Kategorien sei auf Redies et al. (2017) verwiesen,

¹<https://www.wikiart.org>

wo beide Datensätze erstmalig Verwendung fanden und vorgestellt werden. Wir vergleichen diese drei Kunstkategorien mit sechs Kategorien anderer Bilder, welche in drei Kategorien natürlichen Inhaltes, sowie drei Kategorien menschengemachten Inhaltes unterteilt werden können. Die natürlichen Szenen zeigen Pflanzenmuster (*Plant Patterns*, 331 Bilder) und Vegetation (*Vegetation*, 458 Bilder). Da in der Vergangenheit verschiedene Studien (Graham und Field, 2007; Redies et al., 2007b) Hinweise auf eine Ähnlichkeit natürlicher Szenen mit Kunst finden konnten, wird als besondere Herausforderung ein Datensatz von Bildern hinzugenommen, welcher Landschaftsszenen zeigt (*Large Vistas*, 473 Bilder). Die Bilder mit Inhalten menschlichen Ursprungs zeigen Objekte (*Objects*, 207 Bilder), Städteszenen (*Urban Scenes*, 219 Bilder) und Gebäude (*Buildings*, 511 Bilder). Für alle neun genannten Kategorien finden sich Beispielbilder in Abbildung 5.1. Insgesamt vergleichen wir 2090 Kunstbilder mit 2199 Bildern anderer Kategorien. Die erwartete Genauigkeit der Klassifikation bei zufälliger Zuteilung beträgt somit erwartungsgemäß 0.513. In die spätere Analyse der Ergebnisse beziehen wir auch die aus Kapitel 4 bekannten Flechtenmuster (*Lichen*, 244 Bilder), sowie die 2090 ersten Bilder des *ImageNet*-Testdatensatzes (Russakovsky et al., 2015) mit ein.

5.2 Ergebnisse

Wir berechnen die in Abschnitt 5.1.1 vorgeschlagenen Varianzwerte für alle Bilder der in Abschnitt 5.1.2 vorgestellten Kategorien. Abbildung 5.2 zeigt die Median-Werte für P_a , P_g und P_f in Abhängigkeit des Parameters n für verschiedene Bildkategorien, n^2 entspricht dabei der Gesamtzahl der Teilregionen, in welche ein Bild zur Berechnung der Varianzen geteilt wird. Für jede der neun Bildkategorien kann festgestellt werden, dass die Varianzwerte mit größerem n steigen, die Varianz der Bildmerkmale nimmt also mit einer feineren Unterteilung des Bildes zu. Zusätzlich fällt auf, dass die Anstiege für verschiedene Kategorien unterschiedlich steil verlaufen. So verzeichnen Objekte beispielsweise einen steileren Verlauf für alle drei Varianzwerte als die anderen getesteten Kategorien. Die Anstiege der Kunstkategorien *JenAesthetics*, *Islamic art* und *Chinese art* verlaufen sowohl für P_a , als auch für P_g und P_f auffallend ähnlich, allerdings ähneln diese für P_a gleichzeitig den natürlichen Szenen *Plant Patterns* und *Vegetation*, für P_f hingegen heben sie sich durch höhere Anstiege deutlich von diesen ab. Gleichzeitig ähneln



Abbildung 5.1: Beispielbilder der Datensätze, welche in diesem Kapitel genutzt werden, um Kunst von anderen Kategorien abzugrenzen. Die drei Kunstkategorien sind (a) islamische Kunst (*Islamic art*), (b) chinesische Kunst (*Chinese art*) sowie (c) der *JenAesthetics*-Datensatz. Abgegrenzt werden diese zu (d) Gebäuden (*Buildings*), (e) Städteszenen (*Urban Scenes*) und (f) Objekten (*Objects*), sowie Bildern natürlicher Szenen wie (g) Landschaften (*Large Vistas*), (h) Pflanzenmuster (*Plant Patterns*) und (i) Vegetation (*Vegetation*).

Kunstabilder den Landschaftsszenen (*Large Vistas*) im Anstieg für P_f , während Landschaftsbilder wiederum einen steileren Anstieg der Kurve für P_a verzeichnen. Gänzlich von Kunst verschiedene Anstiege für alle Varianzwerte zeigen die drei Kategorien menschlichen Ursprungs (*Buildings*, *Objects* und *Urban Scenes*).

Um festzustellen, ob eine Kombination der Varianzwerte gefunden werden kann, welche geeignet ist, Kunst von allen anderen Kategorien zu unterscheiden, wenden wir das in Abschnitt 5.1.1 beschriebene Klassifikationsschema an, um mithilfe von SVMs verschiedene Parameterkonfigurationen zu testen. Die durchschnittliche Genauigkeit der bei 5-facher Kreuzvalidierung jeweils ausgelassenen Bilder dient dabei als die Klassifikationsgenauigkeit einer bestimmten Parameterkonfiguration. Je höher diese ausfällt, desto diskriminativer ist die Auswahl an Parametern für Kunst und nähert sich damit dem Begriff einer *Signifikanten Form*. Abbildung 5.3 zeigt die jeweils maximale Genauigkeit bei der Unterscheidung der drei Kunstkategorien von den sechs anderen Bildkategorien für die Schichten conv1–conv5. Wie erwartet ist die maximale Klassifikationsrate für weniger Parameter (blaue Kurve) geringer und steigt, wenn weitere Dimensionen hinzukommen (grüne und rote Kurve). Die beste Klassifikationsrate mit drei gleichzeitig betrachteten Merkmalen erreicht einen Wert von 0.938 ± 0.007 ($P_f(4)$, $P_f(12)$ und $P_a(16)$ auf Schicht conv1). Eine Kombination der zwei Werte $P_f(12)$ und $P_a(22)$ erreicht eine Klassifikationsrate von 0.930 ± 0.009 auf Schicht $l = \text{conv1}$. Für eine Übersicht aller erreichten Klassifikationsraten mit P_a und P_f als Eingabe, siehe Abbildung 5.4. Deutlich wird hier, dass für zwei Parameter die Kombination aus $P_f(12)$ und $P_a(22)$ zwar ein globales Optimum bildet, die Klassifikationsraten in deren näherem Umfeld jedoch vergleichbar hoch sind. Abbildung 5.5 visualisiert den Raum, welcher im Fall der Klassifikation anhand der zwei Merkmale $P_a(22)$ und $P_f(12)$ aufgespannt wird. Bilder, die in Extrembereiche bezüglich dieser beiden Merkmale fallen, sind in Abbildung 5.6 gezeigt.

Zur Einordnung der Klassifikationsraten geben wir neben der bereits genannten, erwarteten Ratewahrscheinlichkeit von 0.513 zwei weitere Vergleichsbasen an. Zum einen bestimmen wir die Klassifikationsrate, die eine SVM auf den Pixelwerten der drei Farbkanäle eines Bildes erreicht. Die Bilder werden hierbei aufgrund des hohen Speicherbedarfs auf die ursprüngliche Eingabedimension des unveränderten AlexNet-Modells von 227×227 Pixeln heruntergerechnet. Die so bestimmte Klassifikationsrate beträgt 0.722 ± 0.014 und ist in Abbildung 5.3 als graue Linie

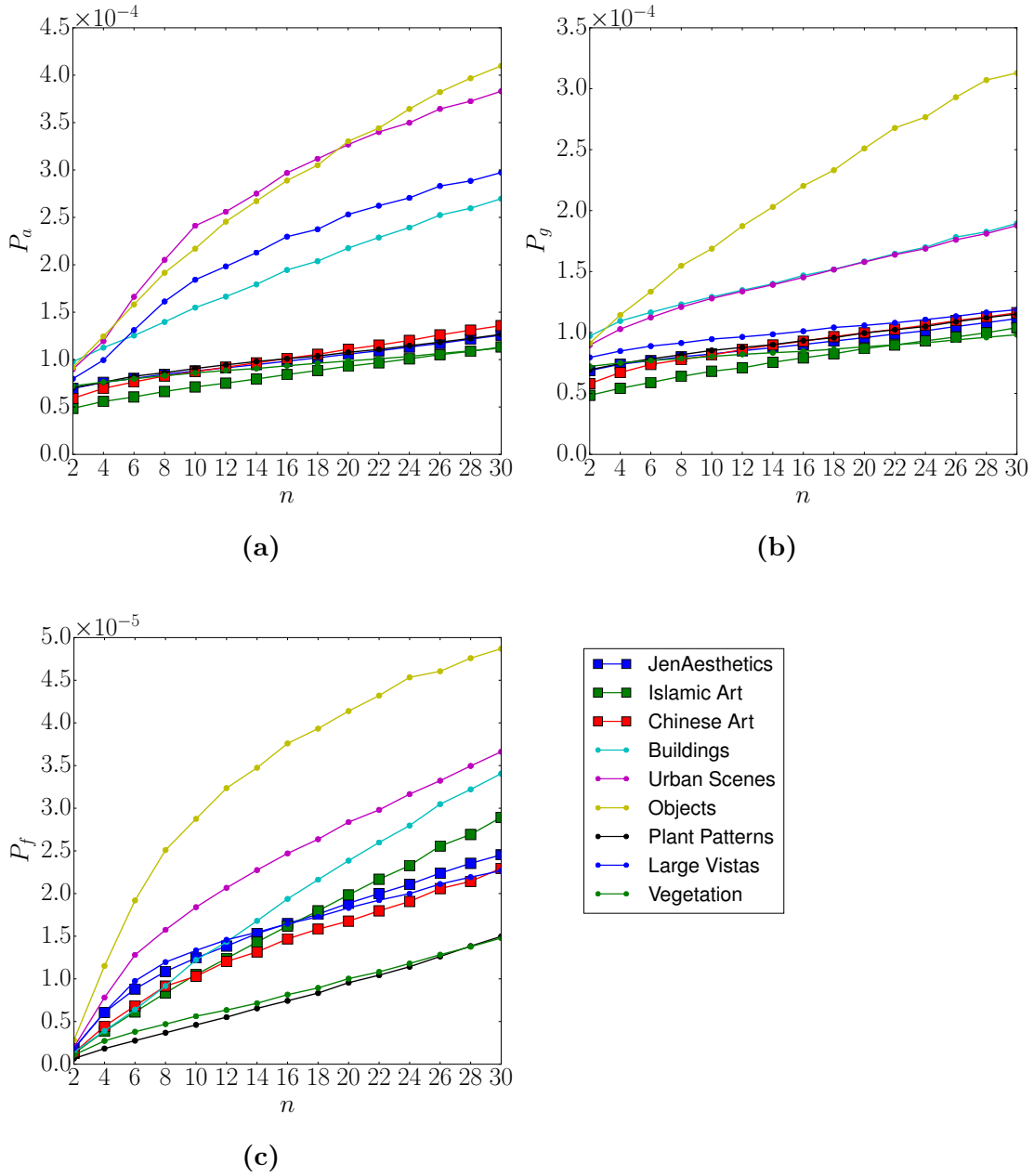


Abbildung 5.2: Median-Varianzen verschiedener Kategorien in Abhängigkeit der Größe der Teilregionen auf Schicht conv1. Das Quadrat der x-Werte entspricht dabei der Zahl der Teilregionen, in welche ein Bild unterteilt wird. Für verschiedene Kategorien zeigen sich unterschiedliche Anstiege und Kurvenverläufe für die drei Varianzen P_a , P_g und P_f .

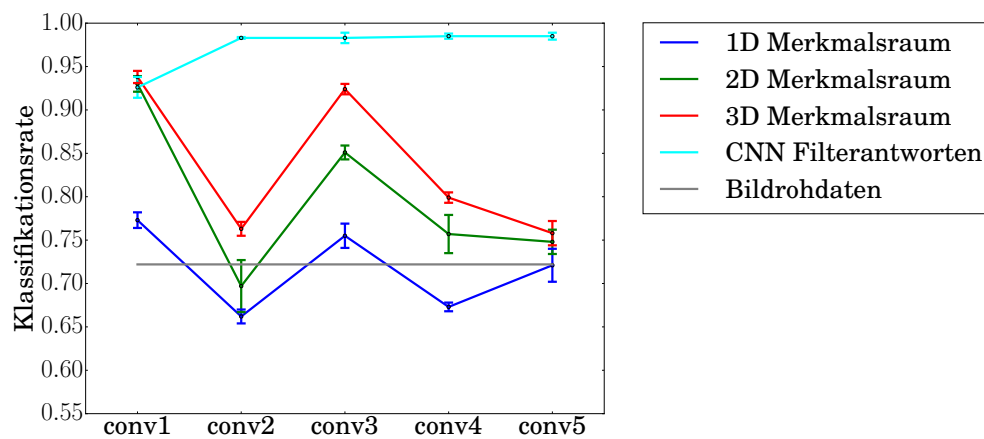


Abbildung 5.3: Jeweils höchste Klassifikationsrate für einen ein- (blau), zwei- (grün) und dreidimensionalen Merkmalsraum (rot) nach einer systematischen Testung aller Kombinationen der beschriebenen Varianzmerkmale. Je höherdimensional die Eingabe ist, desto höher ist auch die maximale Klassifikationsrate. Angegeben sind das Mittel und die Standardabweichung der Klassifikationsrate auf den ausgelassenen Bildern einer 5-fachen Kreuzvalidierung. Zum Vergleich sind ein Klassifikationsexperiment auf den Farbkkanälen des Eingabebildes (grau), sowie auf den Filterantworten des AlexNet Modells gezeigt (türkis).

eingezeichnet. Zum anderen berechnen wir die Klassifikationsrate einer SVM, welche auf den Filterantworten der Schichten conv1–conv5 arbeitet, eingezeichnet als türkise Linie im genannten Graph. Bei beiden Vergleichsexperimenten sind die Eingabedimensionen um ein Vielfaches höher als in den 1D-, 2D- und 3D-Merkmalräumen. So rechnet die SVM auf den Bildrohdaten beispielsweise auf 154 587 Eingabedimensionen, eine SVM auf den Filterantworten der ersten Schicht hingegen schon auf insgesamt 290 400 Dimensionen, was sich aus dem Produkt der Filterzahl mit der Auflösung der Filterantworten ergibt.

Ein interessanter Effekt tritt bei dem Übergang von Schicht conv1 zu Schicht conv2 auf. Hier sinkt die Klassifikationsrate für alle drei Größen des Merkmalsraumes ab, während diese beim Übergang zu Schicht conv3 erneut steigt (siehe Abbildung 5.3). Während die Varianzen auf den Schichten conv1 und conv3 also eine gute Trennung von Kunst und anderen Bildkategorie erlauben, sind die Varianzen auf Schicht conv2 scheinbar weniger spezifisch für Kunst.

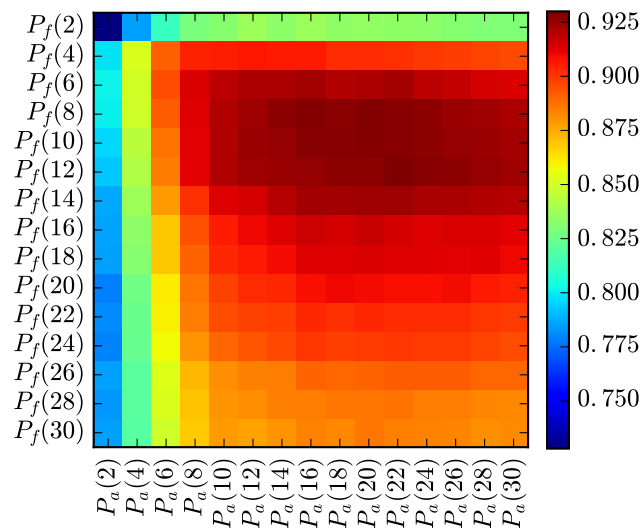


Abbildung 5.4: Die Klassifikationsrate einer SVM bei Paarung der Varianzmerkmale P_a und P_f ist abhängig von der Größe der Teilregionen ($n = 2$ bis $n = 30$, $l = \text{conv1}$). Ein globales Optimum mit einer Rate von 0.930 wird für $P_a(22)$ und $P_f(12)$ erreicht. Im näheren Umfeld dieser Parameterkonfiguration ist die Zahl korrekter klassifizierter Bilder jedoch nur geringfügig niedriger.

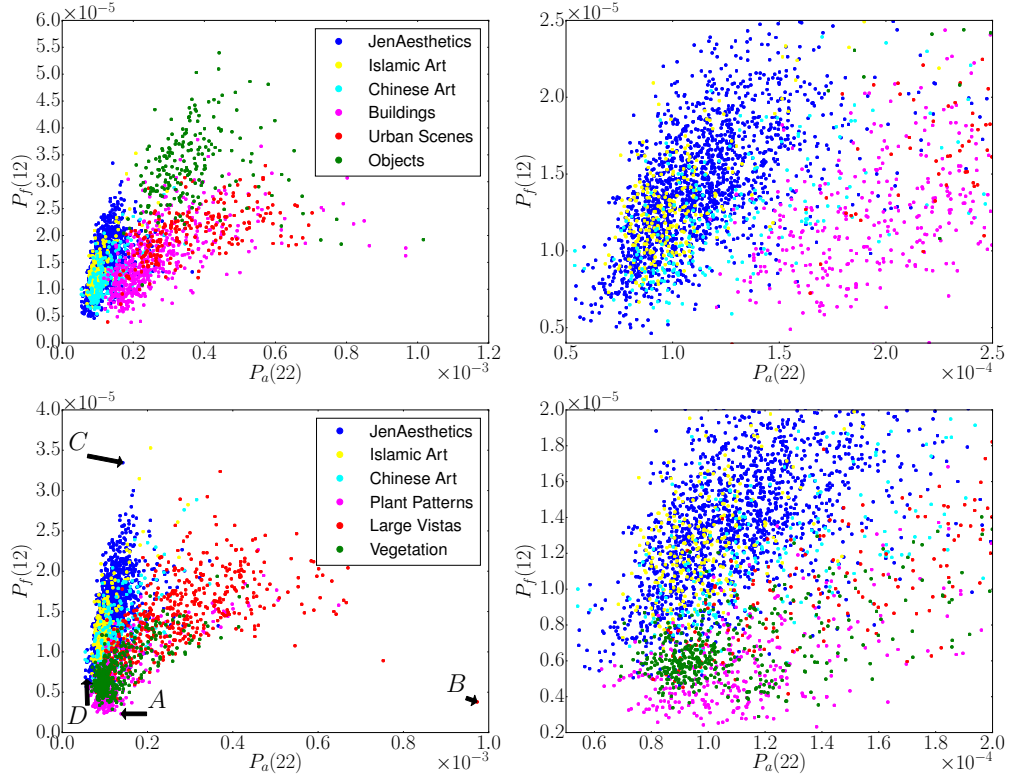


Abbildung 5.5: Der Raum, der durch die beiden Merkmale $P_a(22)$ und $P_f(12)$ aufgespannt wird. Im Experiment hat sich gezeigt, dass anhand dieser Merkmale Kunst von den anderen Kategorien mit einer Genauigkeit von 0.930 getrennt werden kann. Alle vier Plots zeigen Kunst (blau, türkis und gelb), die obere Reihe zeigt dagegen aufgetragen die Kategorien *Buildings*, *Urban Scenes* und *Objects*, die untere Reihe die natürlichen Kategorien *Plant Patterns*, *Large Vistas* und *Vegetation*. Auf der linken Seite ist eine Übersicht aller Werte gezeigt, während auf der rechten Seite eine Vergrößerung des Bereiches abgebildet ist, in welchem sich Kunst mit den anderen Kategorien überschneidet. Wie sich bestimmte Beispielbilder in diesem Raum verorten, wird anhand einer Auswahl der vier Bilder (A–D) veranschaulicht, welche in Abbildung 5.6 gezeigt sind.



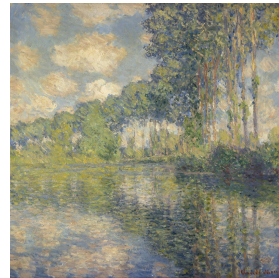
(a)



(b)



(c)



(d)

Abbildung 5.6: Extremfälle, welche den durch $P_a(22)$ und $P_f(12)$ aufgespannten Raum (vgl. Abbildung 5.5) veranschaulichen. Das Pflanzenmuster in (a) ist reich an Struktur, verzeichnet aber auch eine hohe Selbstähnlichkeit und wenig Variabilität. Die Landschaft in (b) ist sehr spärlich und lässt starke Kanten oder von Blau verschiedene Farben vermissen. Verglichen mit anderen Kunstbildern ist das Portrait in (c) insofern atypisch, als dass es große, homogene Flächen aufweist ($P_f(12)$ hoch), vergleichbar mit Objekten. Das Bild in (d) hat eine reichhaltige Struktur ($P_a(22)$ niedrig), ist aber weniger selbstähnlich als Pflanzenmuster (höhere Werte für $P_f(12)$). Die genauen Varianzwerte für alle in diesem Kapitel gezeigten Bilder sind in Tabelle A.2 im Anhang gelistet.

Abbildung 5.7 zeigt die beiden Merkmale $P_a(22)$ und $P_f(12)$ der Schicht conv1 für alle getesteten Bilder, wobei der *JenAesthetics*-Datensatz zusätzlich nach Inhalt und Stil unterteilt ist. Unabhängig von Inhalt und Stil zeigt Kunst für $P_a(22)$ eine ähnliche Verteilung wie die natürlichen Kategorien *Plant Patterns*, *Vegetation* und *Lichen*. Bezüglich $P_f(12)$ sind die Kunstkategorien hingegen deutlich heterogener verteilt, unterscheiden sich jedoch von den natürlichen Kategorien.

Die hohe Klassifikationsrate ist auch durch die von uns getroffene Auswahl an Bildkategorien bedingt. Um diese in Relation zu setzen, wiederholen wir das Experiment mit den 2090 ersten Bildern des *ImageNet*-Testdatensatzes und grenzen diese von den Kunstkategorien ab. Die höchste Klassifikationsrate mit zwei Merkmalen beträgt hier 0.861 ± 0.012 ($P_a(14)$ und $P_f(8)$), mit drei Merkmalen hingegen 0.875 ± 0.011 ($P_a(8)$, $P_f(12)$ und $P_f(26)$) auf Schicht conv1.

5.3 Diskussion der Ergebnisse

Ziel des in diesem Kapitel beschriebenen Experimentes war es, besondere Merkmale zu identifizieren, welche Kunst gegenüber anderen Bildkategorien abgrenzen. Wir haben uns dabei auf Bilder farbiger Kunst beschränkt, und neben westlicher Kunst auch Bilder chinesischen und islamischen Ursprungs betrachtet. Dabei konzentrierten wir uns auf die Varianz von Bildmerkmalen, welche wiederum auf den Filterantworten tiefer neuronaler Netze beruhen. Unsere Ergebnisse zeigen, dass es die von uns definierten Varianzwerte ermöglichen, Kunst von anderen Bildern zu trennen, wobei die Varianz der Filterantworten unterer Schichten am besten geeignet scheinen und eine Klassifikationsrate von 0.930 erreichen. Verglichen mit der Klassifikationsrate von 0.722 einer SVM auf den rohen Bilddaten kann dies als Erfolg gewertet werden. Wird eine SVM statt auf den rohen Bilddaten auf den Filterantworten verschiedener Schichten gelernt, erreichen wir anhand einer Kombination unserer Varianzwerte nur auf der ersten Schicht conv1 ein geringfügig besseres Ergebnis. Für höhere Schichten ist die Klassifikationsrate der SVM direkt auf den Filterantworten deutlich besser und liegt für alle Schichten jeweils über 0.95 (vgl. Abbildung 5.3). Es muss hierbei jedoch beachtet werden, dass im beschriebenen Fall von Varianzen der Filterantworten in einer, zwei oder drei Eingabedimensionen trainiert wird, während direkt auf den Filterantworten

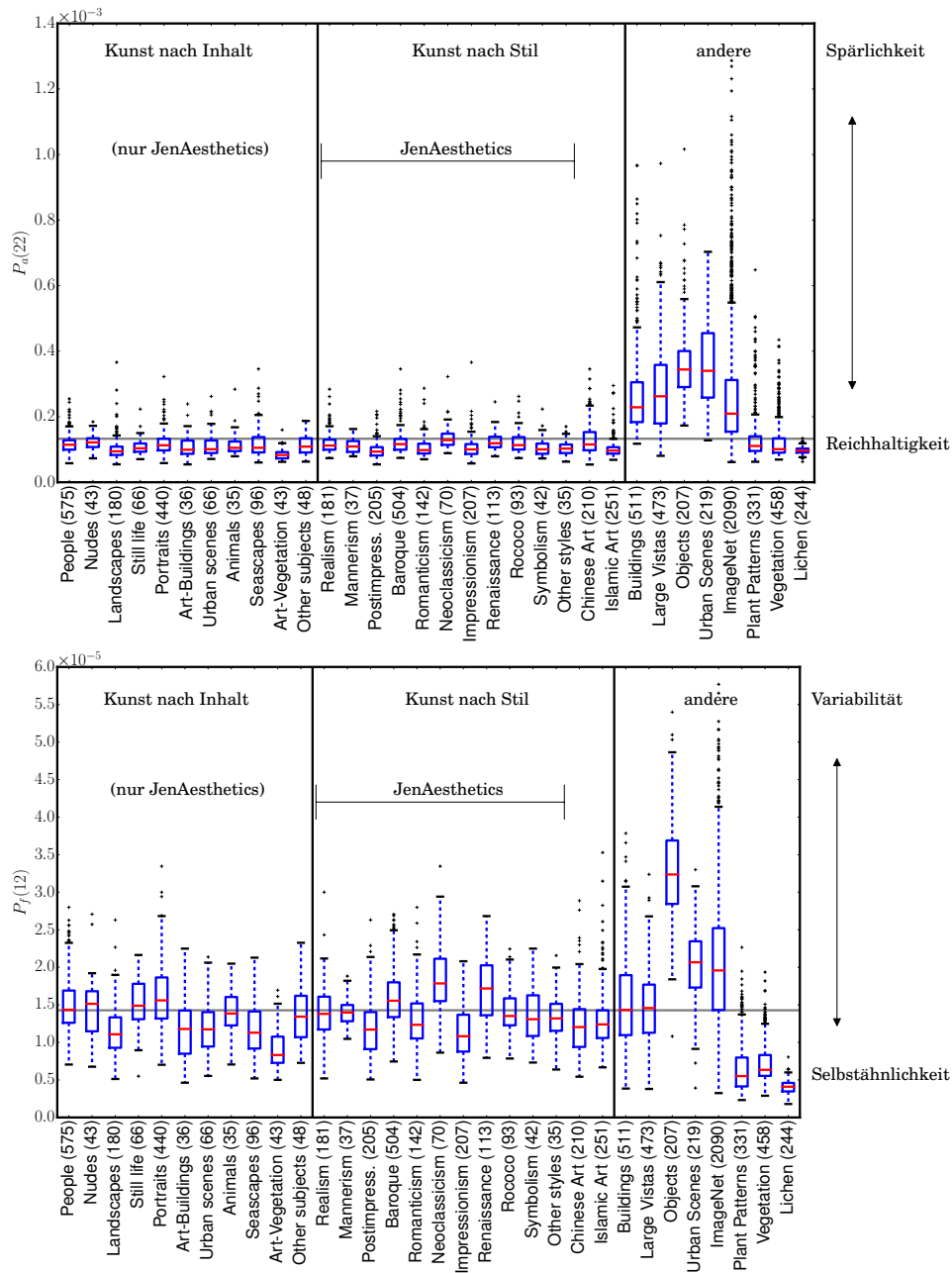


Abbildung 5.7: Boxplots der Varianzen $P_a(22)$ (oben) und $P_f(12)$ (unten). Die graue Linie zeigt den Median aller Werte in einem Plot, die Zahlen in Klammern geben die Anzahl der Bilder in jeder Kategorie an. Die Kunstbilder sind entsprechend ihres Inhaltes (links) und ihres Stils unterteilt (mittig). Für $P_a(22)$ zeigen alle Kunstbilder ähnliche Werte, unterscheiden sich jedoch nicht von den natürlichen Mustern. Für $P_f(12)$ sind die Werte innerhalb der Kunst deutlich verschiedener, unterscheiden sich aber dennoch von den natürlichen Mustern.

die Trennung in 290 400 Eingabedimensionen geschieht. Letztere sollte erwartungsgemäß besser sein, da hier die Zahl der Dimensionen allein die Zahl der Trainingsbeispiele um ein Vielfaches übersteigt. Eine zunehmende Klassifikationsrate auf höheren Schichten des Netzes kann zusätzlich damit erklärt werden, dass höhere Schichten den Bildinhalt kodieren, während untere Schichten eher die Form des Bildes fassen. Auch die Unterscheidung von Kunst und anderen Kategorien ist dabei auf höheren Ebenen somit eher durch den Inhalt der Bilder bedingt. Die durch unsere Maße gewonnene Interpretierbarkeit des Ergebnisses stellt die vorrangige Errungenschaft dieses Experimentes dar. Anhand einer Untersuchung des aufgespannten Raumes mit guter Trennung sind wir in der Lage, Aussagen über Kunst zu treffen, was in dem Fall der Trennung in sehr hoher Dimension nicht ohne Weiteres möglich ist.

5.3.1 Interpretation der Varianzen P_a und P_f

Die Varianzmerkmale P_a (Gesamtvarianz) und P_f (Median der Varianzen der Filter über alle Teilregionen) auf Schicht conv1 stellen sich im Experiment als geeignete Paarung dar, um Kunst von anderen Bildkategorien zu unterscheiden, während P_g keine wichtige Information für die Unterscheidung zu tragen scheint. Auch für eine Menge mit drei Merkmalen wird für hohe Klassifikationsraten nicht auf P_g zurückgegriffen, vielmehr werden zwei bezüglich der Zahl der Teilregionen nach verschiedene Varianzen über die Filter gewählt ($P_f(4)$ und $P_f(12)$ gepaart mit $P_a(16)$, vgl. Abschnitt 5.2).

Es können mithilfe von $P_a(22)$ und $P_f(12)$ zwei Aussagen über Kunst hergeleitet werden. Die Varianzwerte für $P_a(22)$ fallen für Kunst deutlich niedriger aus als für Bilder der Kategorien *Buildings*, *Large Vistas*, *Objects*, *Urban Scenes* und die getesteten Bilder des *ImageNet*-Testdatensatzes, sind aber vergleichbar mit den Bildern natürlicher Kategorie wie *Plant Patterns* und *Vegetation* (vgl. Abbildung 5.7). P_a ist allgemein definiert als die Varianz über die maximalen Antworten aller Filter über das gesamte Bild (vgl. Formel 5.1). Dieser Wert wird dann hoch, wenn über ein Bild verteilt nur sehr wenige Filter maßgeblich antworten, und die übrigen Filter wenige bis keine Ausschläge verzeichnen. Geringe Werte für $P_a(22)$ hingegen sind für Pflanzenmustern und Kunst zu beobachten. $P_a(22)$ ist niedrig, wenn viele Filter über das Bild verteilt in ähnlicher Stärke

antworten, wie das zum Beispiel bei Pflanzenmustern der Fall ist. Ein Wert von 0 würde für P_a in dem theoretischen Fall auftreten, bei dem alle Filter exakt die gleichen Ausschläge verzeichnen. Ein solches Bild würde sich durch eine große Reichhaltigkeit der Strukturen, namentlich der Farben und Kanten entsprechend der Filter auf conv1 auszeichnen. Wir interpretieren P_a daher als ein Maß für *Spärlichkeit* (bzw. geringe *Reichhaltigkeit*) der Strukturen im Bild. Abbildung 5.6 demonstriert diesen Unterschied anhand von Beispielbildern. Während in Abbildung 5.6 Bild (b) hauptsächlich durch die Farbe Blau bestimmt wird und allgemein wenig Struktur aufweist, sind die anderen gezeigten Bilder strukturell wesentlich reichhaltiger. Abbildung 5.5 zeigt, wo die Bilder im aufgespannten Raum verortet sind.

P_f repräsentiert den Median der Varianzen der Filter über das gesamte Bild. Sind die Merkmale bzw. Filterantworten in allen Teilregionen gleich oder zumindest sehr ähnlich, variieren die Antworten der Filter über das Bild verteilt kaum, womit auch der Median über die Varianzen niedrig ist. Weist das Bild hingegen eine hohe Variabilität auf, fallen die Antworten einzelner Filter sehr unterschiedlich aus, eine hohe Varianz ist die Folge. $P_f(12)$ interpretieren wir als die Inversion von Selbstähnlichkeit und damit als Maß für *Variabilität*. Diese Interpretation deckt sich sowohl mit Formel 5.5, als auch mit der Verteilung der Kategorien in unserem Experiment. In Kapitel 4 haben wir ein Maß für Selbstähnlichkeit vorgeschlagen, das auf einem Vergleich globaler und lokaler Filterantworten beruht. Um zu untersuchen, ob tatsächlich ein Zusammenhang zwischen $P_f(12)$ und dem in Kapitel 4 vorgeschlagenem Maß für Selbstähnlichkeit besteht, berechnen wir deren Korrelation. Spearmans Rangkorrelationskoeffizient ergab sich dabei zu $\rho = -0.81$ für alle Bilder und zu $\rho = -0.67$ nur für *JenAesthetics*. Dies stützt die These, nach der P_f als invers zu Selbstähnlichkeit verstanden werden kann, und damit auch als Maß für Variabilität der Bildmerkmale einsetzbar ist.

Wir beobachten, dass Kunstbilder und Pflanzenmuster (*Plant Patterns*, *Vegetation*, *Lichen*) ähnliche Werte für $P_a(22)$ aufweisen, sich hinsichtlich $P_f(12)$ jedoch unterscheiden (vgl. Abbildung 5.7). Bezogen auf die vorgestellte Interpretation der Varianzen kann dieses Ergebnis wie folgt gedeutet werden: Während sich Pflanzenmuster insbesondere durch die Reichhaltigkeit von Struktur wie etwa vieler kontrastreicher Kanten verschiedener Richtungen auszeichnen, ist die Variabilität der Antworten eher gering, oft dominieren wenige Farben ein Bild. Subjektiv kann

das den Eindruck der Eintönigkeit erwecken. Kunstwerke teilen die Reichhaltigkeit der Filterantworten mit den Pflanzenmustern ($P_a(22)$ gering), sind zusätzlich jedoch auch variabel in ihren Merkmalen (höheres $P_f(12)$). Bilder der Kategorien *Objects*, *Buildings*, *Urban Scenes* und *ImageNet* zeigen ebenfalls eine höhere Variabilität als Pflanzenmuster, teilweise ist diese auch deutlich höher als die der Kunstwerke. Gleichzeitig sind Bilder dieser Kategorien jedoch auch spärlicher, das heißt, die Bilder zeigen Regionen, für welche nur sehr wenige Filter antworten (höhere Werte für $P_a(22)$). Zum besseren Verständnis unserer Ergebnisse zeigt Abbildung 5.8 Beispielbilder, welche in unserem Experiment missklassifiziert werden. Die linke Spalte zeigt Bilder, welche fälschlicherweise als Kunst klassifiziert werden, die rechte Spalte hingegen zwei Kunstbilder, die nicht als solche erkannt werden. Während die Bilder auf der linken Seite für den menschlichen Betrachter klar als Fotografien zu erkennen sind, ist für die Bilder der rechten Seite ersichtlich, dass diese gemalt wurden. Deutlich wird dabei, dass die von uns trainierten SVMs auf die oben beschriebenen Merkmale reagieren und nicht auf die Maltechnik, welche Kunstbilder zusätzlich von den anderen Bildern abgrenzt. Weitere Beispiele missklassifizierter Bilder sind in Abbildung 5.9 gezeigt. Deren Auswahl erfolgte anhand des Abstandes zu der durch die SVM gezogenen Entscheidungsgrenze. Gezeigt sind somit die in ihren Eigenschaften atypischsten Kunstwerke einerseits, sowie andererseits Bilder, die am meisten der Kunst entsprechen, jedoch keine Kunstbilder sind.

Eine hohe Reichhaltigkeit könnte mit dem subjektiven Eindruck zusammenhängen, dass Kunstwerke eine Fülle an bildhafter Struktur aufweisen. Tatsächlich scheinen die meisten von uns hier untersuchten Kunstbilder keine größeren, sehr homogene Bereiche (für eine Ausnahme siehe das atypische Bild in 5.8b) aufzuweisen. Die gleichzeitig auftretende höhere Variabilität, die Kunstwerke von den sehr selbstähnlichen Pflanzenmustern abhebt, könnte dazu beitragen, dass Kunstwerke als weniger monoton wahrgenommen werden. Die Kombination aus dem richtigen Maß an Variabilität *versus* Selbstähnlichkeit und einer hohen Reichhaltigkeit könnte Teil dessen sein, was in der Vergangenheit von verschiedenen Autoren und Künstlern als *visual rightness*, *good gestalt* oder auch als *Signifikante Form* bezeichnet wurde.



(a)



(b)



(c)



(d)

Abbildung 5.8: Beispiele von Bildern, die irrtümlicherweise als Kunst klassifiziert werden (a, c) und von Kunstbildern, welche nicht als Kunst erkannt werden (b, d). Die Bilder (a) und (c) zeigen eine Varianz der Bildmerkmale, wie man sie für Kunst erwarten würde, das Kunstbild in (b) hingegen zeigt eine atypische Homogenität ($P_a(22)$ hoch) des blauen Himmels. Das Bild in (d) ist für ein Kunstwerk sehr selbstähnlich ($P_f(12)$ gering) und wird deshalb falsch zugeordnet.

5.4 Zusammenfassung

Motiviert durch die Suche nach Merkmalen, welche Kunst von anderen Bildkategorien unterscheidet, konzentrierten wir uns in diesem Kapitel auf die Varianz von Bildmerkmalen. Es wurden drei verschiedene Varianzmaße definiert, die auf den Filterantworten von CNNs berechnet werden. Für die Schicht conv1 konnten wir zeigen, dass auf den von uns vorgestellten Datensätzen eine Trennung von Kunst und Nicht-Kunstabildern mit hoher Genauigkeit möglich ist. Zusätzlich hat eine Analyse der zur Klassifikation genutzten Merkmale es ermöglicht, einfache Aussagen über die Form von Kunst abzuleiten. Dabei haben wir festgestellt, dass sich die Kunstbilder des *JenAesthetics*-Datensatzes sowie Beispielbilder islamischer und chinesischer Kunst durch eine hohe Reichhaltigkeit der Bildmerkmale auszeichnen, diese also wenige homogene Flächen aufweisen. Die gleiche Eigenschaft konnten wir auch für alle hoch selbstähnlichen Bilder unseres Datensatzes feststellen. Von diesen grenzt sich Kunst wiederum dadurch ab, dass Kunstbilder gleichzeitig eine hohe Selbstähnlichkeit vermissen lassen und eine höhere Variabilität der Bildstrukturen zeigen. Dieses Ergebnis deckt sich mit den Erkenntnissen in Kapitel 4, in welchem ein Maß für die Selbstähnlichkeit vorgestellt und beobachtet wurde, dass Kunst weniger selbstähnlich als die getesteten Pflanzenmuster ist. Bei der Interpretation der Merkmale beschränkten wir uns auf die erste Schicht des AlexNet-Modells, da hier direkt ersichtlich ist, auf welche Merkmale eines Bildes die Filter reagieren und dass diese keine semantischen Informationen kodieren. Als ein auffälliges Ergebnis konnte beobachtet werden, dass sich die Varianzen der Filter der zweiten Schicht (conv2) weniger gut zur Unterscheidung der Klassen eignen, als die der ersten und dritten Schicht. Der Grund dafür bleibt offen und stellt eine unbeantwortete Frage dieses Kapitels dar, die durch weitere Untersuchungen eventuell geklärt werden kann.

Die reine Unterscheidung gemalter Bilder von Fotografien ist durch moderne Methoden der Bildanalyse eine leicht lösbare Aufgabe und war nicht das vorrangige Ziel dieses Kapitels. Durch die Beschränkung auf Varianzen als Unterscheidungskriterium von Kunst und Nicht-Kunst wurde jedoch die Interpretationsmöglichkeit des Ergebnisses gewonnen, womit dem Feld der Experimentellen Ästhetik neue Wege eröffnet werden. Beispielsweise stellt sich die Frage, ob ein bestimmter Grad

an Reichhaltigkeit und Selbstähnlichkeit tatsächlich die ästhetische Präferenz von Bildern positiv oder negativ beeinflusst.

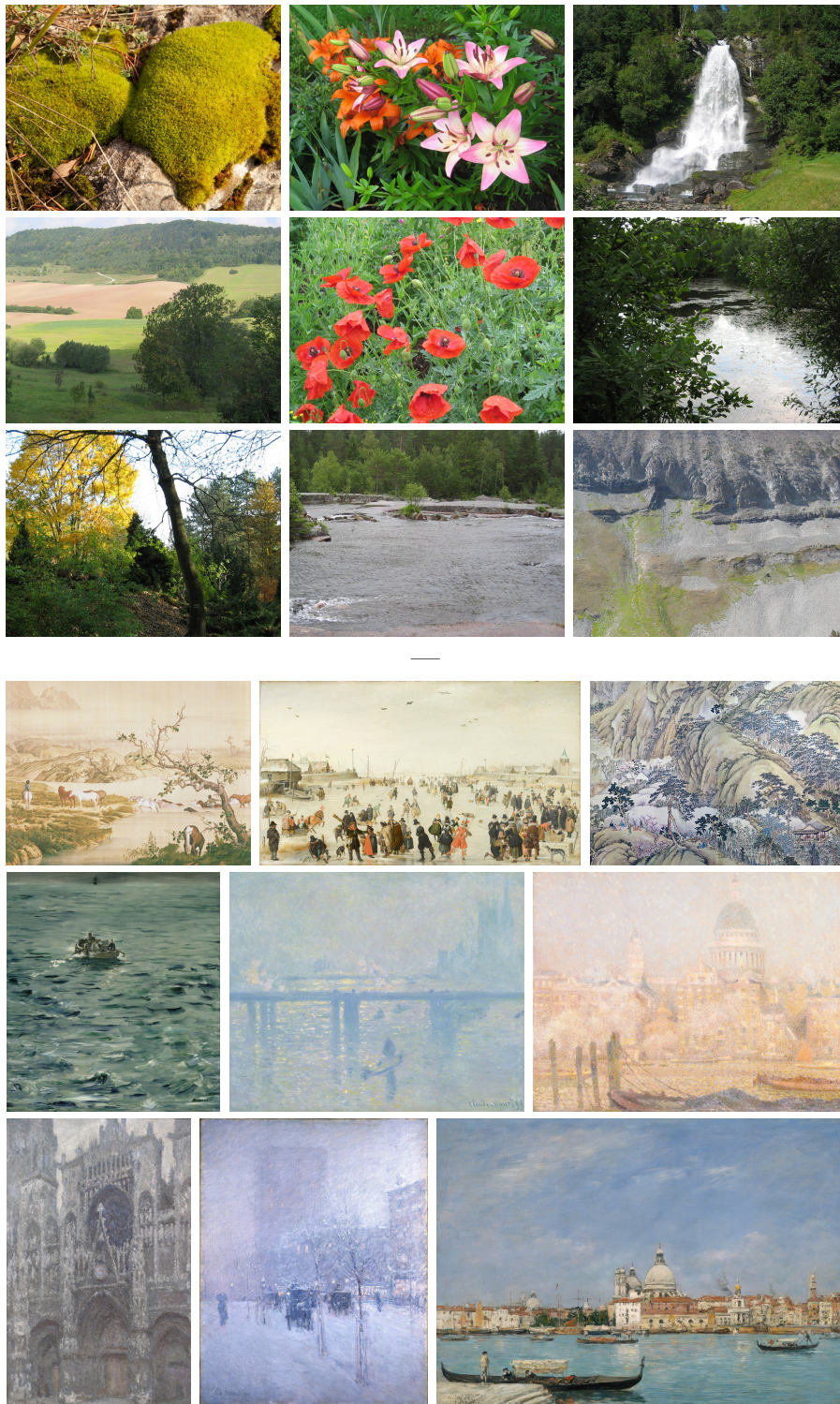


Abbildung 5.9: Missklassifizierte Bilder mit dem größten Abstand zur Entscheidungsgrenze. Oben sind Bilder gezeigt, welche irrtümlich als Kunst klassifiziert werden. Unten zu sehen sind Kunstbilder, welche nicht als solche erkannt werden.

Kapitel 6

Diskussion

An die vorangegangenen Kapitel schließen sich verschiedene Fragen an, auf welche in diesem Kapitel eingegangen werden soll. Insbesondere ist von Interesse, inwiefern sich allgemeingültige Schlüsse bei der hier gewählten Fokussierung auf eine Auswahl visueller Kunst überhaupt ziehen lassen und in welchem Bezug die neu entwickelten Methoden zu bereits etablierten Maßen auf dem Gebiet der Experimentellen Ästhetik stehen.

6.1 Zur Wahl des Kunstdatensatzes

In den vorangegangenen Kapiteln wurden verschiedene Methoden entworfen, welche es ermöglichen, Kunst bezüglich ihrer objektiv messbaren Eigenschaften zu analysieren. Diese Untersuchung schließt sich dabei einer langen Tradition innerhalb der Experimentellen Ästhetik an (vgl. Abschnitt 1.2), in welcher das Ziel ist, Ästhetik auf definierbare Eigenschaften von Objekten oder Bildern zurückzuführen. Einschränkend muss an dieser Stelle bemerkt werden, dass *Ästhetik* ein sehr weit gefasster Begriff ist. In Kapitel 1 wurden Arbeiten einerseits der Computational Aesthetics, wie auch der Experimentellen Ästhetik vorgestellt. Dabei wurde deutlich, dass sich ein großer Teil der Arbeiten auf dem Gebiet der Computational Aesthetics mit Bewertungen von Fotografien beschäftigt. Welche Bilder ästhetisch sind, wird dabei durch einen Mehrheitsbeschluss der Nutzer desjenigen Portals bestimmt, von welchem die Bilder heruntergeladen wurden. Das ist eine legitime Festlegung, welche sich jedoch von vielen Arbeiten innerhalb der Experimentellen

Ästhetik unterscheidet, welche Kunst *per se* als ästhetisch annehmen und auf Bewertungen auf Ebene einzelner Bilder verzichten. Dieser Herangehensweise folgend wurden auch in dieser Arbeit die Bilder des *JenAesthetics*-Datensatzes als ästhetisch gesetzt, welche teils vor mehreren Jahrhunderten gemalt und heute noch in großen Museen ausgestellt sind. Diese Festlegung folgt der von Hume (1757) postulierten Annahme, dass die großen Kunstwerke die Zeit überdauern, während die Beliebtheit der den Moden unterliegenden Werke mit der Zeit wieder abnimmt. Zusätzlich rechtfertigt dieses Kriterium den Ansatz, die moderne Kunst als gesonderte Gruppe zu behandeln. Betrachtet man die monochromen Bilder von Yves Klein, das „Schwarze Quadrat“ von Sewerinowitsch Malewitsch oder die Kompositionen von Piet Mondrian, erscheint diese Beschränkung sinnvoll. Das bedeutet jedoch nicht, dass Bilder moderneren Ursprungs für die Forschung nicht interessant wären. In Redies und Brachmann (2017) wird unter anderem die in Kapitel 2 vorgestellte Analyse der Entropie von Kantenrichtungen genutzt, um die Beziehung zwischen Abstrakter Kunst, sogenannter „Schlechter Kunst“ und traditioneller Kunst zu ergründen. Dabei wird deutlich, dass auch innerhalb der Abstrakten Kunst Untergruppen existieren, bei denen einige in ihren statistischen Eigenschaften eher der traditionellen Kunst entsprechen, andere hingegen deutlich von dieser abweichen.

Die in dieser Arbeit vorgestellten Untersuchungen sind zusätzlich in der Weise beschränkt, dass ein Großteil der betrachteten Kunstbilder ihrem Ursprung nach der westlichen traditionellen Kunst entstammen. Zwar wurde in Kapitel 5 auch eine Auswahl chinesischer und islamischer Kunstwerke einbezogen, verglichen mit der großen Menge anderer Kulturen deckt das jedoch nur einen sehr kleinen Teil ab, auch in Anbetracht der Tatsache, dass wenige hundert Bilder unmöglich die Kunst ganzer Epochen repräsentieren können. Dies gilt auch für den westlichen Kulturkreis, so wurde etwa grafische Kunst, welche statt auf Leinwand auf Papier geschaffen ist, hier nicht näher betrachtet.

6.2 Zur Aussagekraft der Ergebnisse

Eine Motivation der Untersuchung der statistischen Eigenschaften von Kunst ist die bereits genannte Hoffnung, anhand dieser die Wahrnehmung des Menschen besser zu verstehen. Zeki (2001) geht soweit zu sagen, dass auch Künstler unwis-

sentlich Neurologen seien, die das Gehirn mit ihren eigenen Techniken erforschen. Für andere Autoren ist der Vergleich von Kunst mit der Statistik natürlicher Szenen ein Schlüssel zum Verständnis der Kunst, wobei beispielsweise angenommen wird, dass eine spärliche Kodierung des visuellen Reizes hier eine Gemeinsamkeit darstellt (Graham und Redies, 2010; Redies, 2007a). Diesem Ansatz der Analyse der Statistik von Kunst folgend, wurden in dieser Arbeit verschiedene Methoden entworfen, um die Eigenschaften von Kunst zu natürlichen Szenen und Bildern mit menschengemachtem Inhalt in Beziehung zu setzen. Kapitel 2 wurde von der Frage geleitet, wie sich Kantenrichtungen verschiedener Bilder relativ zueinander verhalten. Es wurde dabei festgestellt, dass sich Kunst für das gesetzte Signifikanzniveau von $p < 0.001$ nicht von Landschaften unterscheidet, was ein Hinweis darauf sein kann, dass die Kantenstatistik in Kunstwerken die der Landschaften nachbildet. Das in Kapitel 4 entworfene Maß für Selbstähnlichkeit eines Bildes wurde ausgehend von der Beobachtung entworfen, dass ein etabliertes Maß (Amirshahi et al., 2012) zwar ebenfalls Kunst betrachtet, dabei jedoch die Farbe außer Acht lässt. Das in dieser Arbeit vorgestellte Maß für Selbstähnlichkeit adressiert dieses Problem, indem Filterantworten tiefer neuronaler Netze als Grundlage der Berechnung genutzt werden. Beim Vergleich von Kunst mit anderen Kategorien zeigt sich jedoch, dass sich die Selbstähnlichkeit von Kunst nicht signifikant von Bildern unterscheidet, welche Städteszenen zeigen. Daraus zu schlussfolgern, dass Kunst die Statistik von Städteszenen nachbildet, ist nicht gerechtfertigt. Vielmehr wird hier deutlich, dass absolute Schlussfolgerungen aus einem Vergleich der reinen Bildeigenschaften mit Vorsicht zu genießen sind.

Das Maß zur Bestimmung der Selbstähnlichkeit wurde so entworfen, dass dieses quantifiziert, inwieweit sich Bildelemente im Großen auch im Kleinen wiederholen, eine Evaluation erfolgte jedoch vor allem durch einen Abgleich mit Kategorien, welche weithin als „selbstähnlich“ oder auch als „weniger selbstähnlich“ gesehen werden. Die subjektive Selbstähnlichkeit einzelner Bilder, wie sie durch Menschen bewertet werden würde, mag davon verschieden sein. Auch wird dabei die Frage aufgeworfen, ob ein bestimmter Grad an Selbstähnlichkeit in der ein oder anderen Form überhaupt ein notwendiges Kriterium für Kunst und damit für die Experimentelle Ästhetik von Interesse ist.

In Kapitel 5, welches die Varianz von Bildmerkmalen in den Mittelpunkt der Untersuchungen stellt, wurde ausgehend von diesen Überlegungen ein anderer

Weg gewählt und die Frage gestellt, wodurch sich Kunst von anderen Bildkategorien unterscheidet. Motiviert ist eine solche Fragestellung durch die Suche nach einer möglicherweise existierenden *Signifikanten Form* von Kunst (Bell, 1914). Es konnte dabei gezeigt werden, dass ein bestimmtes Verhältnis von Selbstähnlichkeit und Variabilität zu Reichhaltigkeit Kunstbilder tatsächlich mit hoher Genauigkeit von anderen getesteten Bildkategorien abgrenzt. Vorschnell darf jedoch auch hier nicht geschlossen werden, dass damit ein notwendiges oder gar hinreichendes Kriterium für Kunst gefunden wurde. Zwar konnten die visuellen Auswirkungen der entworfenen Varianzmaße anhand von Beispielbildern verdeutlicht werden, was wiederum Hinweise auf die Gestalt von Kunst gab, aufgrund der eingeschränkten Datenmenge der Vergleichsdatensätze bleiben diese Ergebnisse jedoch nur Hinweise und können keine allgemeine Gültigkeit für sich beanspruchen. Erst kontrollierte, psychologische Studien könnten einen Zusammenhang der gefundenen Kriterien mit der subjektiven Schönheit eines Bildes herstellen. Auch folgt man dabei der im letzten Abschnitt vorgestellten Argumentation, dass Kunst *per se* schön ist, eine Aussage, die zusätzlich angezweifelt werden kann.

6.3 Bezug der vorgeschlagenen Maße zu etablierten Maßen der Experimentellen Ästhetik

In dieser Arbeit wurde das Ziel verfolgt, verschiedene Maße zu definieren, anhand derer Kunst zu anderen Bildern in Beziehung gesetzt werden kann. Die Suche nach kunstdefinierenden Bildeigenschaften hat innerhalb der Experimentellen Ästhetik eine lange Tradition (vgl. Abschnitt 1.2.1), es stellt sich daher die Frage, ob durch die hier vorgeschlagenen Methoden neue Aspekte von Kunst untersucht wurden und welche Wechselwirkungen mit etablierten Maßen beobachtet werden können. Für den *JenAesthetics*-Datensatz werden daher die in dieser Arbeit vorgeschlagenen Maße der Entropie, der Selbstähnlichkeit, sowie die Varianzen berechnet und mit den bereits bewährten Maßen korreliert, welche auf der PHOG-Repräsentation eines Bildes beruhen (Selbstähnlichkeit, Komplexität und Anisotropie; Redies et al. (2012)). Da auch die Untersuchung von Bildfrequenzen innerhalb der Experimentellen Ästhetik eine zentrale Rolle einnimmt, werden die vorgeschlagenen Maße zusätzlich mit Maßen verglichen, die auf dem Frequenzspektrum der Bilder basie-

ren (Anstieg der Amplitude in Abhängigkeit der Frequenz und die Abweichung zum linearen Anstieg; (Redies et al., 2007b)). Für eine detaillierte Beschreibung dieser Vergleichsmaße sei auf die genannten Veröffentlichungen verwiesen. Die sich ergebenden paarweisen Korrelationskoeffizienten (Rangkorrelation mit Spearmans ρ) sind in Tabelle 6.1 gezeigt.

Der positive Zusammenhang zwischen der Entropie erster und der Entropie zweiter Ordnung (E_1 und E_2) wurde bereits in Kapitel 2 näher beleuchtet. Ein negativer Zusammenhang ergibt sich hingegen bei der Korrelation der Entropie zweiter Ordnung mit der Anisotropie (P_A , $\rho = -0.43$). Dies kann dadurch begründet werden, dass viele Kanten unterschiedlicher Richtung die Entropie zweiter Ordnung heben, da die Kanten auch relativ zueinander unterschiedliche Richtungen zeigen. Die PHOG-Anisotropie, berechnet als die Varianz der Kantenstärken über verschiedene Richtungen und Teilregionen eines Bildes, fällt für diese Bilder, da wenige Kantenrichtungen herausstechend sind, was die Varianz der Kantenhistogramme wiederum verringert. Für die Selbstähnlichkeiten, welche auf Grundlage der Filterantworten auf conv1 (SS_1), sowie auf Grundlage der Filterantworten auf conv2 (SS_2) berechnet wurden, ergibt sich ein positiver Zusammenhang von $\rho = 0.57$. Die Beziehung beider Werte wurde ebenfalls im entsprechenden Kapitel 4 näher beleuchtet. Sie besteht darin, dass SS_1 gegenüber SS_2 auf einfachen Merkmalen wie Kanten und Farben beruht, während SS_2 abstraktere Bildinhalte abgleicht. Eine positive Korrelation beider Maße ist zu erwarten, da abgesehen von der Wahl der Filterantworten beide auf gleiche Weise berechnet werden. Interessanterweise zeigen auch die Varianzmaße $P_a(22)$ und $P_f(12)$ einen positiven Zusammenhang von $\rho = 0.64$. Dies leuchtet insofern ein, als dass eine hohe Varianz der Filter erwartungsgemäß auch mit einer hohen Gesamtvarianz über alle Antworten einhergeht. In Kapitel 5 wurden $P_a(22)$ und $P_f(12)$ als die Gegensätze *Spärlichkeit* und *Reichhaltigkeit*, sowie *Selbstähnlichkeit* und *Variabilität* interpretiert und es konnte so gezeigt werden, dass das Verhältnis dieser Gegensätze zueinander Kunst gegenüber den anderen getesteten Kategorien abgrenzen kann. Diese Interpretation wird zusätzlich durch den negativen Zusammenhang zwischen $P_f(12)$ und der Selbstähnlichkeit SS_1 gestützt. Dieser Zusammenhang wurde bereits in Kapitel 5 aufgegriffen.

Ein positiver Zusammenhang von $\rho = 0.60$ zeigt sich bei der Korrelation von $P_a(22)$ und der PHOG-Anisotropie P_A . Dies leuchtet insofern ein, als dass beide

Maße sehr ähnlich berechnet werden. Die PHOG-Anisotropie ergibt sich aus der Varianz der Kantenhistogramme, welche für 64 Teilregionen eines Bildes berechnet werden. Zwar ist für $P_a(22)$ die Unterteilung des Bildes sehr viel feiner und es werden neben Kanten mit den Filtern des AlexNet auch Farben einbezogen, auch hier berechnet sich der Wert aber ebenfalls als die Varianz der Filterantworten über Teilregion eines Bildes. Bedenkt man, dass viele der Filter der ersten Schicht des AlexNet-Modells auch auf Luminanzkanten reagieren (vgl. Abbildung 3.1b), ist der hohe Zusammenhang beider Maße nachvollziehbar.

Ein interessantes Ergebnis zeigt sich bei dem Vergleich der Selbstähnlichkeit SS_1 mit der PHOG-Komplexität P_C , welche sich aus der Summe aller im Bild gefundenen Luminanzkanten ergibt. Der Korrelationskoeffizient beträgt hier 0.72, eine hohe Selbstähnlichkeit geht also mit einer hohen Zahl an Kanten allgemein im Bild einher. Diese Beobachtung ist dadurch erklärbar, dass das in Kapitel 4 vorgestellte Maß für Selbstähnlichkeit auf einem Vergleich der global maximalen Filterantworten mit den lokal maximalen Filterantworten beruht. Zeigt das Bild global eine Vielzahl verschiedener Kanten, so steigt die Selbstähnlichkeit genau dann, wenn diese auch lokal vermehrt auftreten. Eine hohe PHOG-Komplexität ist ein Resultat dieses Umstandes. Die ebenfalls hohe Korrelation zwischen der Selbstähnlichkeit SS_1 und dem Fourier-Anstieg F_{Sl} von $\rho = 0.62$ stützt diese Erklärung, da ein hoher Fourier-Anstieg vor allem in Bildern messbar ist, die einen hohen Detailgrad aufweisen. Viele hohe Frequenzen bewirken einen hohen Anstieg der Amplituden über die Frequenz. Eine psychologische Studie könnte hier Aufschluss darüber geben, ob das vorgeschlagene Maß für Selbstähnlichkeit auf von menschlichen Probanden mit der subjektiv wahrgenommenen Komplexität von Bildern in Verbindung gebracht wird.

Eine Einbettung der in dieser Arbeit produzierten Ergebnisse in Form eines Vergleiches neuer mit etablierten Methoden ist wichtig, um verborgene Zusammenhänge aufdecken zu können. Die Korrelationen der Werte können jedoch nur Hinweise dafür liefern, ob neue Aspekte von Kunst tatsächlich das Ergebnis der Untersuchungen sind. Ein absolutes Kriterium, welche Höhe der Korrelation dabei als Schwelle gesetzt werden soll, ist nicht gegeben.

Tabelle 6.1: Spearmans ρ zur Quantifizierung des Zusammenhangs zwischen den in dieser Arbeit entwickelten Maßen (E_1 , E_2 : Entropie 1. und 2. Ordnung (Kapitel 2); SS_1 , SS_2 : Selbstähnlichkeit auf conv1 und auf conv2 (Kapitel 4); $P_a(22)$, $P_f(12)$: Varianzmaße aus Kapitel 5) und den PHOG-Maßen (P_S : Selbstähnlichkeit, P_C : Komplexität, P_A : Anisotropie; Redies et al. (2012)), sowie den auf dem Frequenzspektrum beruhenden Maßen (F_{Sl} : Fourier Slope, F_{Si} : Fourier Sigma; Redies et al. (2007b)) für die Bilder des *JenAesthetics*-Datensatzes. Alle Korrelationen sind signifikant ($p < 0.001$), sofern diese nicht mit „nicht signifikant“ („n. s.“) gekennzeichnet und grau hinterlegt sind. Korrelationen im Betrag größer oder gleich 0.60 sind farblich hervorgehoben.

	E_1	E_2	SS_1	SS_2	$P_a(22)$	$P_f(12)$	P_S	P_C	P_A	F_{Sl}	F_{Si}
E_1	1	0.63	0.07	0.15	-0.10	0.03	0.12	0.06	-0.28	-0.04	-0.32
E_2		1	0.37	0.26	-0.22	-0.26	0.06	0.3	-0.43	0.4	-0.23
SS_1	n. s.		1	0.57	-0.49	-0.67	0.44	0.72	-0.39	0.62	-0.08
SS_2				1	0.01	-0.08	0.25	0.46	0.08	0.29	-0.25
$P_a(22)$				n. s.	1	0.64	-0.54	-0.38	0.60	-0.38	-0.03
$P_f(12)$	n. s.			n. s.		1	-0.48	-0.42	0.56	-0.54	-0.09
P_S		n. s.					1	0.32	-0.44	0.11	-0.10
P_C	n. s.							1	-0.31	0.62	-0.03
P_A				n. s.					1	-0.28	0.05
F_{Sl}	n. s.									1	0.10
F_{Si}					n. s.			n. s.	n. s.		1

6.4 CNNs als Modelle menschlichen Sehens

In Abschnitt 1.2.1 wurde dargestellt, dass Ästhetik als hochgradig komplexer Prozess verstanden werden kann, der von den Hauptdomänen Perzeption, Kognition und Emotion beeinflusst wird. Gleichzeitig zeigen bildgebende Verfahren, dass eine Vielzahl von Gehirnregionen am ästhetischen Erleben beteiligt sind, etwa der orbitofrontale und zinguläre Kortex, die Inselrinde, der Nucleus accumbens, oder auch der Nucleus caudatus (Brown et al., 2011; Vartanian und Skov, 2014; Cela-Conde et al., 2011). Zeki (2013) fragt daher, ob die *Signifikante Form*, wie Bell (1914) sie annimmt, nicht eher als *Signifikante Konfigurationen* verstanden werden kann, welche beschreibt, wie die für die ästhetische Emotion nötigen Hirnregionen maximal aktiviert werden. Die zugrunde liegenden Aktivierungen wiederum könnten dabei domänenspezifisch sein. In dieser Arbeit beschränkte ich mich mit der Betrachtung von Kunst auf die visuelle Domäne und damit auf die Perzeption.

In Kapitel 3, welches ein Maß zur Bestimmung der Symmetrie von Bildern beschreibt, wurde eine Kodierung der Bilder anhand der maximalen Filterantworten eines tiefen neuronalen Netzes verwendet. Ausgehend davon wurde auch für das Maß zur Bestimmung der Selbstähnlichkeit (siehe Kapitel 4), sowie bei der Definition der Varianzmaße (Kapitel 5) auf diese Art der Kodierung zurückgegriffen. Die Verwendung der Filterantworten von CNNs als Ausgangspunkt für die Entwicklung darauf aufbauender Algorithmen zur Quantifizierung verschiedener Bildeigenschaften erscheint dabei als eine geeignete Grundlage. CNNs sind biologisch inspirierte Modelle, in deren Zentrum die Idee einer hierarchischen Filterung der Eingabe steht. Diese geht auf die von Hubel und Wiesel (1962) beschriebene Gliederung des visuellen Systems zurück, bei dem einfache Antworten durch Verschaltung zunehmend komplexere Antworten erzeugen. Tatsächlich zeigt sich diese Hierarchie auch bei der Betrachtung der während des Trainings von CNNs entstehenden Filter. Diese ähneln auf den unteren Ebenen den in Kapitel 2 verwendeten Gaborfiltern, während Visualisierungen höherer Ebenen veranschaulichen, dass diese tatsächlich auf zunehmend komplexere Muster innerhalb der Eingabe reagieren (Yosinski et al., 2015). In den Kapiteln 3 und 4 konnte diese zunehmende Abstraktion genutzt und gezeigt werden, dass mit verschiedenen Ebenen als Berechnungsgrundlage auch unterschiedliche Abstraktionsgrade der Bildmerkmale gewählt werden können.

Wie in Kapitel 1.1 gezeigt, wurden die Filterantworten verschiedener Ebenen auch auf dem Gebiet der Bildverarbeitung als generische Bildmerkmale (sogenannte *DeCAF*-Merkmale; Donahue et al. (2014)) beispielsweise für die Erkennung des Stils von Bildern, erfolgreich eingesetzt. Denzler et al. (2016) sehen in CNNs das Potential, als Modell menschlichen Sehens auch der Erforschung der ästhetischen Wahrnehmung dienen zu können und betonen, dass in der Experimentellen Ästhetik ein solches Modell bisher fehlt. Für die Verwendung von CNNs als eben solche Modelle sprechen auch Ergebnisse einer Arbeit, welche die Aktivierungen des Temporallappens von Rhesusaffen mit den Aktivierungen tiefer neuronaler Netze beim Betrachten einfacher Objekte vergleicht (Cadieu et al., 2014). Die Autoren zeigen, dass anhand beider Aktivierungen die Objektklassen mit ähnlicher Genauigkeit vorhergesagt werden können und beide Aktivierungen zusätzlich die gleiche Gruppierung der Klassen selbst aufweisen, bei der Objekte gleicher Klasse ähnlich repräsentiert sind und nahe beieinander liegen.

Auch wenn diese Ergebnisse dafür sprechen, CNNs als Modelle menschlichen Sehens auch für die Erforschung der ästhetischen Wahrnehmung zu verwenden, gilt die schon in den Kapiteln 3 und 4 genannte Einschränkung: Während die unteren Ebenen eines Netzes als wirklich generisch angesehen werden können, sind höhere Schichten zunehmend spezifischer und auf den Trainingsdatensatz angepasst (Yosinski et al., 2014). Als Konsequenz können also formale Eigenschaften von Bildern wie deren Farb- und Kantenverteilung anhand der unteren Ebenen gut untersucht werden, während höhere Ebenen erwartungsgemäß derart auf die Bilder reagieren, wie dies durch den Trainingsdatensatz vorgegeben ist. Auf Ebene einfacher Kanten stellt dieser Umstand für Trainingsdatensätze, die hinreichend groß sind, keine Einschränkung dar, da hier eine Vielzahl von Kantenrichtungen und Farben vertreten ist. Auf Ebene von Objekten und abstrakter Bildmerkmale hingegen wird diese Spezifität zum Nachteil und schränkt die allgemeine Gültigkeit von Schlussfolgerungen bezüglich der Antwortmuster ein. Indem wir uns bei unseren Experimenten und den abgeleiteten Schlussfolgerungen in Bezug auf Kunst auf die unteren Ebenen der AlexNet-Modells beschränkt haben, wurde diesem Umstand Rechnung getragen.

6.5 Diskriminative und generative Modelle

Kunst anhand verschiedener Merkmale von anderen Bildkategorien zu unterscheiden, ermöglicht jeweils nur eine Beschreibung der Kunst *in Abgrenzung* zu anderen Bildern. Zur Prüfung von Hypothesen derart, dass Kunst verschiedene Eigenschaften beispielsweise natürlicher Szenen nachbildet, ist dieses Vorgehen legitim und wurde in dieser Arbeit als ein Weg gewählt. Gleichzeitig entspricht diese Methode einer *diskriminativen Modellbildung*. Diskriminative Modelle lernen die Wahrscheinlichkeit einer Kategorie bei gegebener Eingabe. Da CNNs wie das hier verwendete AlexNet Bilder anhand einer Abbildung einer hochdimensionalen Eingabe auf Klassenwahrscheinlichkeiten lernen, sind auch diese den diskriminativen Modellen zuzuordnen. Während der diskriminative Ansatz im allgemeinen Fall sehr gute Ergebnisse erzielt (Krizhevsky et al., 2012; Simonyan und Zisserman, 2014; He et al., 2015; Huang et al., 2016), wurden in jüngerer Zeit mit sogenannten *Adversarial Examples* (Szegedy et al., 2013; Goodfellow et al., 2014a) jedoch auch Probleme deutlich. Während die gelernten Klassen bei guter Generalisierung auch für unbekannte Eingaben erkannt werden, wurde gezeigt, dass CNNs anfällig für gezielte Verfälschungen der Eingabe sind. Mithilfe des Backpropagation-Algorithmus (vgl. Abschnitt 3.1) lässt sich für jede Eingabe eine Differenzbild berechnen, welches linear kombiniert mit der Eingabe dazu führt, dass das CNN eine falsche Zuordnung der Klasse trifft. Diese Veränderung kann dabei oftmals so subtil sein, dass sie dem menschlichen Auge nicht ersichtlich ist. In vielen Fällen reicht auch die Veränderung eines einzigen Pixels, um ein CNN eine falsche Zuordnung treffen zu lassen (Su et al., 2017). Brown et al. (2017) demonstrieren ein Verfahren zur Erstellung eines *Adversarial Patches*, ein ausdrucksbares, kreisrundes Muster, welches die Klassifikation fehlschlagen lässt, wenn dieses anderen Bildern hinzugefügt wird. Statt der vormals korrekten Klasse gibt das CNN in der Folge für das scheinbar gleiche Bild eine gewünschte, falsche Klasse aus. Diese Ergebnisse zeigen, dass die diskriminative Natur von CNNs in offensichtlich wichtigen Teilaspekten nicht dem menschlichen Sehen entspricht. Zwar lassen sich auch Menschen optisch täuschen, oft haben diese Täuschungen allerdings mit der Statistik natürlicher Bilder selbst zu tun, wobei das Gehirn fehlende Informationen eines Reizes fehlerhaft ergänzt.

Man kann argumentieren, dass die beschriebene, spezifische Anfälligkeit von CNNs

dem Umstand geschuldet ist, dass die Entscheidungsgrenzen in Räumen von sehr hoher Dimension gezogen werden. Analog dazu wird jedoch deutlich, dass eine Beschreibung von Kunst auch in Räumen niedrigerer Dimension jeweils lediglich in Abgrenzung zu anderen Kategorien niemals sicher sein kann, das tatsächliche Wesen der Kunst erfasst zu haben, da es immer Beispiele geben kann, die in nicht betrachteten Eigenschaften nicht der Kunst entsprechen.

Ein möglicher Ansatz zur Lösung dieser Probleme ist die Verwendung *generativer Modelle*. Anders als die diskriminativen Modelle lernen diese keine bedingte Wahrscheinlichkeit, sondern die Gesamtverteilung der Eingaben selbst. Ein in jüngerer Zeit vorgestellter Ansatz sind sogenannte *Generative Adversarial Networks* (GANs, Goodfellow et al. (2014b)), welche unter anderem erfolgreich für die Generierung synthetischer Bilder eingesetzt wurden (Radford et al., 2015; Karras et al., 2017). Die Grundidee ist dabei, zwei neuronale Netze miteinander zu kombinieren, wobei eines ein *Generator* ist, das andere hingegen ein *Diskriminator*. Der Generator lernt, eine zufällige Eingabe auf eine einer Trainingsverteilung entsprechenden Ausgabe abzubilden, während der Diskriminator lernt, die Erzeugnisse des Generators von echten Trainingsbeispielen zu unterscheiden. In einem iterativen Prozess lernt der Generator anhand des Urteils des Diskriminators die generierten Bilder so weit zu verbessern, dass der Diskriminator diese im besten Fall nicht von den Trainingsbeispielen unterscheiden kann. Abbildung 6.1 zeigt beispielhaft verschiedene Bilder, welche durch ein GAN erzeugt wurden, das wiederum auf den im Wikiart-Datensatz enthaltenen Landschaftsbildern trainiert wurde. Zwar sind die generierten Bilder relativ einfach gehalten, bei geringer Auflösung sehen einige strukturell jedoch den Beispielen des Trainingsdatensatzes sehr ähnlich. GANs sind eine relativ neue Technologie, bei der noch nicht alle Probleme gelöst sind, etwa der sogenannte *Mode Collapse* (Goodfellow et al., 2014b), bei dem das GAN nur sehr wenig paarweise verschiedene Bilder erzeugt. Auch ist bei Trainingsdatensätzen mit einer großen Vielfalt der Beispiele nach wie vor ein Problem, Bilder hoher Auflösungen zu erzeugen. Im Gegensatz zu diskriminativen Modellen bieten GANs jedoch die Möglichkeit, auf den tatsächlichen Verteilungen der Eingaben zu arbeiten, was auch für die Experimentelle Ästhetik neue Möglichkeiten eröffnet. Beispielsweise können so große Datenmengen von Bildern erzeugt werden, die bei erfolgreichem Training der Statistik von Kunst entsprechen und im Rahmen psychologischer Studien genauer untersucht werden können. Verschiedene Arbeiten

nutzen GANs, um einfache, handgezeichnete Entwürfe in realistische Fotografien zu überführen (Isola et al., 2017). Zhu et al. (2017) stellen ein Verfahren vor, um mithilfe von GANs beliebige Bilder in verschiedene Kunststile zu übersetzen. Auch hier könnte die Untersuchung der Frage, ob die generierten Bilder durch Probanden tatsächlich als ästhetisch gewertet werden, neue Erkenntnisse bringen. Sollten zukünftige Verfahren es ermöglichen, die hochdimensionalen Transformationen nicht nur des Diskriminators, sondern auch des Generators zu verstehen, könnte dies die Gewinnung wertvoller Erkenntnisse auch über die Form von Kunst ermöglichen.

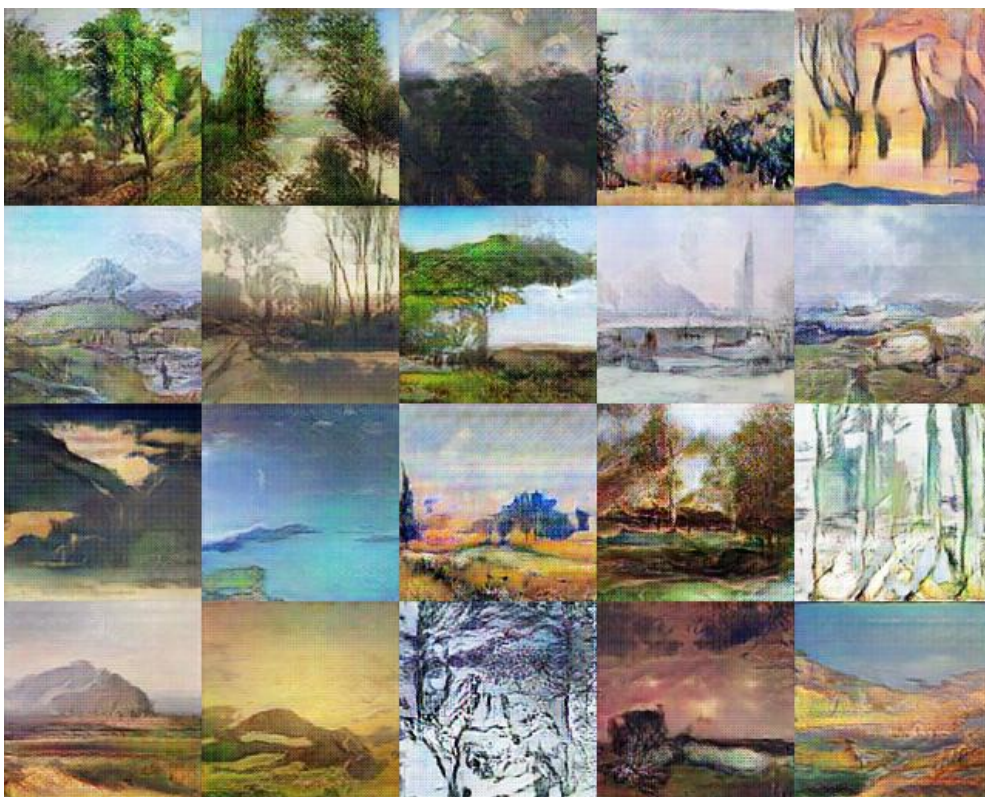


Abbildung 6.1: „Kunst“, welche mithilfe eines *Generative Adversarial Networks* erzeugt wurde (bisher unveröffentlicht). Trainiert wurde dieses vorher auf knapp 15 000 Landschaftsbildern des Wikiart-Datensatzes. Die Bilder sind strukturell sehr einfach gehalten, lassen jedoch vereinzelt Bäume und Berge erkennen. Der rasche Fortschritt auf dem Gebiet generativer Modelle eröffnet großes Potential auch für die Experimentelle Ästhetik.

Schlussbetrachtung

Das Ziel der vorliegenden Arbeit war die Entwicklung neuer Methoden der Bildanalyse, welche im Rahmen der Experimentellen Ästhetik für einen Vergleich von Kunst mit anderen Bildkategorien genutzt werden können. Zur Einbettung der Arbeit wurde zunächst ein Überblick über das Gebiet der Computational Aesthetics gegeben, welches sich innerhalb der Informatik verortet, und welches unter anderem die Entwicklung von Methoden zum Gegenstand hat, mithilfe derer Bilder verschiedenen ästhetischen Wertes anhand ihrer Merkmale voneinander abgegrenzt werden können. Während es in den Anfangstagen dabei gängige Praxis war, spezifische Merkmale zu entwerfen und ein Bild anhand dieser Merkmale zu bewerten, haben tiefe neuronale Netze diese Herangehensweise in jüngerer Zeit ersetzt und auch den Entwurf spezifischer Merkmale obsolet gemacht. Ein Nachteil dieser Entwicklung ist jedoch, dass mit der vermehrten Anwendung tiefer neuronaler Netze auch die Interpretationsmöglichkeit des Gelernten verloren geht. Neben der Computational Aesthetics wurde auch das Gebiet der Experimentellen Ästhetik vorgestellt. Letzteres ist weniger daran interessiert, anhand von Klassifikatoren die Schönheit eines Bildes festzustellen oder zu bewerten, im Fokus steht hier vielmehr, das ästhetische Erleben an sich besser zu verstehen und erklären zu können, worauf sich dieses gründet. Die Analyse der statistischen Eigenschaften von Kunst nimmt dabei eine zentrale Stellung ein. Eine Hypothese ist unter anderem, dass Kunst der Statistik natürlicher Szenen folgt, an welche sich das visuelle System des Menschen im Laufe der Evolution angepasst hat. Es wird dabei beispielsweise davon ausgegangen, dass der Künstler diese Statistik in seinen Bildern nachbildet, um die gleiche Art der Kodierung innerhalb des Sehsystems zu erreichen und dies den ästhetischen Wert von Bildern zumindest teilweise bedingt. Von dieser These ausgehend wurde in Kapitel 2 eine Methode vorgestellt, welche es ermöglicht, die relativen Bezüge von Kantenrichtungen innerhalb eines Bildes

zu untersuchen. Im Vergleich mit anderen Bildkategorien wurde festgestellt, dass sich Kunst bezüglich des beschriebenen Maßes der Entropie relativer Kantenbeziehungen nicht von Landschaftsbildern unterscheidet, was ein Hinweis auf die Gültigkeit der These einer statistischen Nachbildung sein könnte, bei Weitem aber kein Beleg für diese ist.

In Kapitel 3 wurde eine Art der Kodierung der Filterantworten tiefer neuronaler Netze entworfen, welche genutzt wurde, um ein Maß zur Berechnung der Symmetrie eines Bildes zu definieren. In dem sich anschließenden Kapitel 4 fokussierte ich mich erneut auf die Analyse der Kunst. Ausgehend von der Beobachtung, dass ein innerhalb der Experimentellen Ästhetik bereits etabliertes Maß zur Bestimmung der Selbstähnlichkeit eines Bildes zwar zur Analyse von Kunst genutzt wird, dabei jedoch Farbinformationen verwirft, wurde ein Maß entworfen, welches auch Farbinformationen in die Berechnung einbezieht. Analog zu Kapitel 3 konnte gezeigt werden, dass die zunehmend abstraktere Repräsentation von Bildinhalten durch tiefe neuronale Netze genutzt werden kann, um das Maß wahlweise auf einfache oder abstraktere Strukturen im Bild ansprechen zu lassen.

In Kapitel 5 wurde ein zu den Kapiteln 2 und 4 verschiedener Ansatz gewählt. Die Ausgangsfrage war hier nicht, welchen Bildkategorien Kunst bezüglich bestimmter Eigenschaften ähnelt, sondern vielmehr, welche Eigenschaften Kunst gegenüber anderen Kategorien abgrenzen. Verschiedene Autoren vertreten die Ansicht, dass Kunst insofern eine besondere Stellung unter den Bildern einnimmt, als dass diese bestimmten, unbekannten Gesetzen der Komposition und Gestalt folgt, wofür Bell (1914) den Begriff der „Signifikanten Form“ prägte. In Kapitel 5 wurden ausgehend von dieser Hypothese verschiedene Maße entwickelt, welche die Varianz der Filterantworten tiefer neuronaler Netze quantifizieren und eine Kombination dieser Maße gesucht, die Kunst von anderen Kategorien trennt. Anhand einer genaueren Betrachtung der trennenden Merkmale war es möglich, neue Hypothesen bezüglich der Beschaffenheit von Kunst abzuleiten: Zum einen besitzt Kunst eine hohe Reichhaltigkeit der Bildmerkmale, eine Eigenschaft, die Kunst mit Bildern verschiedener Pflanzenmuster teilt. Zum anderen grenzt sich Kunst jedoch zu diesen dadurch ab, dass in Kunstbildern eine zu hohe Selbstähnlichkeit im Sinne einer Wiederholung der Bildinhalte vermieden wird.

Im Rahmen der Diskussion wurden verschiedene Fragen aufgegriffen, die sich aus Methodik und Ergebnissen der Arbeit ergeben. Zum einen wurde auf die Frage eingegangen, wodurch der Ansatz gerechtfertigt ist, Kunst als Beispiele *per se* ästhetischer Bilder in den Fokus der Untersuchungen zu stellen, und welche Schwierigkeiten damit verbunden sind. Im Anschluss daran wurde die Aussagekraft der Ergebnisse bewertet. Die erwähnte Ähnlichkeit der Kantenentropie von Kunst- mit Landschaftsbildern kann als Hinweis auf eine Nachahmung der Statistik natürlicher Szenen durch Kunst gewertet werden, für die Selbstähnlichkeit hat sich das jedoch nicht gezeigt, hier schien Kunst den Städteszenen zu ähneln. An dieser Stelle wird offensichtlich, dass man nicht dem Fehlschluss unterliegen darf, jene Ergebnisse, welche den Erwartungen entsprechen, als gültig zu erklären, gegenteilige Ergebnisse hingegen als unwichtig. Eine stärkere Verankerung der Ausgangshypothesen in der Biologie könnte an dieser Stelle Klarheit verschaffen, ist aufbauend auf dem derzeitigen Wissensstand über das Sehsystem jedoch schwierig. Es schließt sich somit die Frage an, ob es auch biologisch begründbar ist, dass Kunst hinsichtlich der relativen Kantenrichtungen den natürlichen Szenen gleicht, hinsichtlich der Farbselbstähnlichkeit jedoch nicht.

Mit Ausnahme des Kapitels 2 basieren die in dieser Arbeit vorgestellten Algorithmen zur Messung verschiedener Bildeigenschaften auf den Filterantworten des vorgestellten AlexNet-Modells. Es stellt sich hier die Frage, ob CNNs als gute Modelle des Sehsystems dienen, um folglich auch zur Erforschung der ästhetischen Wahrnehmung verwendet werden zu können. Dafür spricht einerseits deren hierarchischer Aufbau, durch den von einfachen Merkmalen früher Schichten auf höheren Schichten zunehmend abstrahiert wird. Darüber hinaus wurde eine Ähnlichkeit abstrakter Merkmale zu den Repräsentationen von Objekten im visuellen Cortex von Rhesusaffen gezeigt, was ebenfalls dafür spricht, CNNs auch als einfache Modelle des menschlichen Sehens zu verwenden. Deren Beschränkungen und Unterschiede zum menschlichen Sehsystem sollten dabei jedoch nicht außer Acht gelassen werden, wie etwa die beschriebenen *Adversarial Examples* zeigen. Diskriminative Modelle wie CNNs oder die in Kapitel 5 vorgestellte Methode der Trennung anhand verschiedener Bildeigenschaften definieren Bilder wie Kunst jeweils in Abgrenzung zu anderen Bildern, wodurch zwar neue Hypothesen gewonnen werden, eine erschöpfende Beschreibung jedoch nie gesichert ist.

Offene Fragen und Ausblick

Einige offene Fragen, die sich an die Ergebnisse dieser Arbeit anschließen, wurden bereits kurz thematisiert. Kunst als Beispiele ästhetischer Bilder zu sehen, ist innerhalb der Experimentellen Ästhetik ein gängiger Ansatz. Dennoch stellt sich für jedes der in dieser Arbeit vorgestellten Maße die Frage, ob diese tatsächlich die Ästhetik eines Bildes beeinflussen. Werden Bilder mit einer hohen Entropie relativer Kantenorientierungen tatsächlich als schöner bewertet als jene, welche eine geringe Entropie aufweisen? Welcher Grad an Selbstähnlichkeit wird von menschlichen Betrachtern bevorzugt? Für Kunst wurde gezeigt, dass diese einen mittleren Bereich der Farbselbstähnlichkeit aufweist, ist dies tatsächlich ein Faktor, der die mittlere Schönheit eines Bildes beeinflusst? Besonders bezüglich der Varianz ist die Frage interessant, ob sich hier bestätigen ließe, dass Reichhaltigkeit und Variabilität tatsächlich relevant für die subjektiv wahrgenommene Ästhetik von Bildern sind. Hier könnte eine Testung von Bildern verschiedener Kategorien, die verschiedene Grade der betrachteten Varianzen repräsentieren, in einem Experiment durch Betrachter hinsichtlich ihrer Präferenz bewertet und der Zusammenhang von Bewertungen und den beschriebenen Varianzen genauer aufgeschlüsselt werden.

Zusätzlich schließt sich die Frage an, ob sich das gefundene Verhältnis von Reichhaltigkeit und Variabilität von Kunstbildern nicht als Nebeneffekt anderer Faktoren erklären lässt. Die Frage nach der Komplexität von Bildern und deren Auswirkung auf ihre Ästhetik wurden in verschiedenen Studien aufgegriffen (vgl. Abschnitt 1.2.3). Berlyne (1974) gründet seine Theorie der Ästhetik auf einem bestimmten Grad der Komplexität von Reizen, Birkhoff (1933) beschreibt Ästhetik als Verhältnis von Ordnung und Komplexität. Die in Kapitel 5 gefundenen Varianzmaße berühren diese Fragen direkt, da Komplexität einerseits als Reichhaltigkeit der Bildmerkmale verstanden werden kann, genauso aber auch als eine hohe Variabilität. Inwiefern sich dieses Ergebnis mit den Theorien von Berlyne und Birkhoff vereinen lassen, kann Gegenstand sich anschließender Untersuchungen sein. Es könnte beispielsweise der Frage nachgegangen werden, ob menschliche Betrachter Komplexität intuitiv eher als Reichhaltigkeit oder eher als Variabilität verstehen. In der Diskussion wurde bereits der Unterschied diskriminativer zu generativen Modellen aufgegriffen, wobei als Beispiel generativer Modelle GANs angeführt

wurden. Eine diskriminative Beschreibung von Kunst durch Abgrenzung verschiedener Eigenschaften zu anderen Kategorien kann aufschlussreich sein, wenn dabei bestimmte Fragen wie beispielsweise die bezüglich des Grades von Selbstähnlichkeit eines Bildes adressiert werden. Der zügige Fortschritt auf dem Gebiet generativer Modelle innerhalb des maschinellen Lernens bietet hingegen neue Möglichkeiten auch für die Erforschung der Statistik von Kunst, wobei deren definierende Eigenschaften automatisiert gefunden werden. Gleichzeitig werden für die Experimentelle Ästhetik neue Fragen aufgeworfen, insbesondere wird durch statistisch perfekte, künstlich generierte Bilder hier das Wesen der Kunst selbst hinterfragt und es ist offen, ob in ihren Eigenschaften perfekte, durch Maschinen erstellte Bilder von Menschen tatsächlich als Kunst anerkannt werden.

Literatur

- Abdel-Hamid, Ossama, Abdel-Rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn und Dong Yu (2014). „Convolutional neural networks for speech recognition“. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.10, S. 1533–1545.
- Abe, Kanako, Babak Saleh und Ahmed Elgammal (2013). „An early framework for determining artistic influence“. In: *International Conference on Image Analysis and Processing*. Springer, Berlin, S. 198–207.
- Abeln, Jonas, Leonie Fresz, Seyed Ali Amirshahi, I Chris McManus, Michael Koch, Helene Kreysa und Christoph Redies (2016). „preference for well-balanced saliency in details cropped from photographs“. In: *Frontiers in Human Neuroscience* 9.704. DOI: 10.3389/fnhum.2015.00704.
- Aks, Deborah J und Julien C Sprott (1996). „Quantifying aesthetic preference for chaotic patterns“. In: *Empirical Studies of the Arts* 14.1, S. 1–16.
- Amirshahi, SA, J Denzler und C Redies (2013a). *Jenaesthetics—a public dataset of paintings for aesthetic research*. Techn. Ber. Computer Vision Group, University of Jena, Germany.
- Amirshahi, Seyed Ali, Gregor Uwe Hayn-Leichsenring, Joachim Denzler und Christoph Redies (2014). „Evaluating the rule of thirds in photographs and paintings“. In: *Art & Perception* 2.1-2, S. 163–182.
- Amirshahi, Seyed Ali, Michael Koch, Joachim Denzler und Christoph Redies (2012). „PHOG analysis of self-similarity in aesthetic images“. In: *Human Vision and Electronic Imaging*. San Francisco, CA, 82911J.
- Amirshahi, Seyed Ali, Christoph Redies und Joachim Denzler (2013b). „How self-similar are artworks at different levels of spatial resolution?“ In: *Proceedings of the Symposium on Computational Aesthetics*. ACM, S. 93–100.

- Aragón, José Luis, Gerardo G Naumis, Ming Bai, Manuel Torres und Philip K Maini (2008). „Turbulent luminance in impassioned van Gogh paintings“. In: *Journal of Mathematical Imaging and Vision* 30.3, S. 275–283.
- Arnheim, Rudolf (1954). *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press, Berkley.
- Arora, Ravneet Singh und Ahmed Elgammal (2012). „Towards automated classification of fine-art painting style: A comparative study“. In: *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE. Tsukuba, Japan, S. 3541–3544.
- Bar, Moshe und Mital Neta (2006). „Humans prefer curved visual objects“. In: *Psychological Science* 17.8, S. 645–648. DOI: 10.1111/j.1467-9280.2006.01759.x.
- Bar, Yaniv, Noga Levy und Lior Wolf (2014). „Classification of artistic styles using binarized features derived from a deep neural network“. In: *Workshop at the European Conference on Computer Vision*. Springer, Cham, S. 71–84.
- Barla, Annalisa, Francesca Odone und Alessandro Verri (2003). „Histogram intersection kernel for image classification“. In: *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*. Bd. 3. IEEE, S. III–513.
- Barth, Erhardt, Christoph Zetsche, Mario Ferraro und Ingo Rentschler (1993). „Fractal properties from 2D curvature on multiple scales“. In: *Geometric Methods in Computer Vision II*. Bd. 2031. International Society for Optics und Photonics, S. 87–100.
- Bell, Clive (1914). *Art*. Chatto & Windus, London.
- Berezhnoy, Igor E, Eric O Postma und H Jaap van den Herik (2009). „Automatic extraction of brushstroke orientation from paintings“. In: *Machine Vision and Applications* 20.1, S. 1–9.
- Berezhnoy, Igor E, Eric O Postma und Jaap van den Herik (2005). „Computerized visual analysis of paintings“. In: *Int. Conf. Association for History and Computing*. Royal Netherlands Academy of Arts und Sciences, Amsterdam, S. 28–32.
- Bergamo, Alessandro, Lorenzo Torresani und Andrew W Fitzgibbon (2011). „PICODES: Learning a compact code for novel-category recognition“. In: *Advances in Neural Information Processing Systems*. Granada, Spain, S. 2088–2096.

- Berlyne, Daniel E (1974). *Studies in the New Experimental Aesthetics: Steps toward an objective psychology of aesthetic appreciation*. Hemisphere, Oxford.
- Bertamini, Marco, Letizia Palumbo, Tamara Nicoleta Gheorghes und Mai Galatsidas (2016). „Do observers like curvature or do they dislike angularity?“ In: *British Journal of Psychology* 107.1, S. 154–178. DOI: 10.1111/bjop.12132.
- Bies, Alexander J, Daryn R Blanc-Goldhammer, Cooper R Boydston, Richard P Taylor und Margaret E Sereno (2016). „Aesthetic responses to exact fractals driven by physical complexity“. In: *Frontiers in Human Neuroscience* 10.210. DOI: 10.3389/fnhum.2016.00210.
- Birkhoff, George David (1933). *Aesthetic Measure*. Harvard University Press, Cambridge.
- Bishop, Christopher M (2006). *Pattern Recognition and Machine Learning*. Springer, New York.
- Bosch, Anna, Andrew Zisserman und Xavier Munoz (2007). „Representing shape with a spatial pyramid kernel“. In: *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*. CIVR '07. New York, NY, USA: ACM, S. 401–408. URL: <http://doi.acm.org/10.1145/1282280.1282340>.
- Brachmann, Anselm, Erhardt Barth und Christoph Redies (2017). „Using CNN features to better understand what makes visual artworks special“. In: *Frontiers in Psychology* 8.830. DOI: 10.3389/fpsyg.2017.00830.
- Brachmann, Anselm und Christoph Redies (2016). „Using Convolutional Neural Network filters to measure left-right mirror symmetry in images“. In: *Symmetry* 8.12, S. 144. DOI: 10.3390/sym8120144.
- Brachmann, Anselm und Christoph Redies (2017a). „Computational and experimental approaches to visual aesthetics“. In: *Frontiers in Computational Neuroscience* 11.102.
- Brachmann, Anselm und Christoph Redies (2017b). „Defining self-similarity of images using features learned by convolutional neural networks“. In: *IS&T/SPIE Electronic Imaging*. Bd. 2017. 14, S. 188–194. DOI: 10.2352/ISSN.2470-1173.2017.14.HVEI-142.
- Braun, Julia, Seyed A Amirshahi, Joachim Denzler und Christoph Redies (2013). „Statistical image properties of print advertisements, visual artworks and images of architecture“. In: *Frontiers in Psychology* 4.808. DOI: 10.3389/fpsyg.2013.00808.

- Bressan, Marco, Claudio Cifarelli und Florent Perronnin (2008). „An analysis of the relationship between painters based on their work“. In: *2008 15th IEEE International Conference on Image Processing*. IEEE, S. 113–116.
- Bronshtein, I.N., K.A. Semendyayev, G. Musiol und H. Mühlig (2015). *Handbook of Mathematics*. Springer Berlin Heidelberg.
- Brown, Steven, Xiaoqing Gao, Loren Tisdelle, Simon B Eickhoff und Mario Liotti (2011). „Naturalizing aesthetics: brain areas for aesthetic appraisal across sensory modalities“. In: *Neuroimage* 58.1, S. 250–258.
- Brown, Tom B, Dandelion Mané, Aurko Roy, Mart ´n Abadi und Justin Gilmer (2017). „Adversarial patch“. In: *arXiv preprint arXiv:1712.09665*.
- Bruno, Nicola, Valentina Gabriele, Tiziana Tasso und Marco Bertamini (2014). „“Selfies” reveal systematic deviations from known principles of photographic composition“. In: *Art & Perception* 2.1-2, S. 45–58.
- Cadiou, Charles F, Ha Hong, Daniel LK Yamins, Nicolas Pinto, Diego Ardila, Ethan A Solomon, Najib J Majaj und James J DiCarlo (2014). „Deep neural networks rival the representation of primate IT cortex for core visual object recognition“. In: *PLoS Computational Biology* 10.12, e1003963.
- Cela-Conde, Camilo J, Luigi Agnati, Joseph P Huston, Francisco Mora und Marcos Nadal (2011). „The neural foundations of aesthetic appreciation“. In: *Progress in Neurobiology* 94.1, S. 39–48. DOI: 10.1016/j.pneurobio.2011.03.003.
- Cetinic, Eva und Sonja Grgic (2013). „Automated painter recognition based on image feature extraction“. In: *ELMAR, 2013 55th International Symposium*. IEEE, S. 19–22.
- Chandler, Damon M und David J Field (2007). „Estimates of the information content and dimensionality of natural scenes from proximity distributions“. In: *JOSA A* 24.4, S. 922–941.
- Chatterjee, Anjan und Oshin Vartanian (2014). „Neuroaesthetics“. In: *Trends in Cognitive Sciences* 18.7, S. 370–375. DOI: 10.1016/j.tics.2014.03.003.
- Chen, Po-Chun, JH Hays, Seungkyu Lee, Minwoo Park, Yanxi Liu et al. (2007). „A quantitative evaluation of symmetry detection algorithms“. In: *Tech. Rep. CMU-RI-TR-07-36*. Robotics Institute, Carnegie Mellon University.
- Condorovici, Răzvan George, Corneliu Florea und Constantin Vertan (2015). „Automatically classifying paintings with perceptual inspired descriptors“. In:

- Journal of Visual Communication and Image Representation* 26, S. 222–230.
DOI: 10.1016/j.jvcir.2014.11.016.
- Condorovici, Razvan, Corneliu Florea und Constantin Vertan (2013). „Author identification for digitized paintings collections“. In: *Signals, Circuits and Systems (ISSCS), 2013 International Symposium on*. IEEE. Iasi, Romania, S. 1–4.
- Coppola, David M, Harriett R Purves, Allison N McCoy und Dale Purves (1998). „The distribution of oriented contours in the real world“. In: *Proceedings of the National Academy of Sciences USA* 95.7, S. 4002–4006.
- Criminisi, Antonio, Martin Kemp und Andrew Zisserman (2002). *Bringing pictorial space to life: computer techniques for the analysis of paintings*. Techn. Ber. Microsoft Corporation.
- Criminisi, Antonio und David G Stork (2004). „Did the great masters use optical projections while painting? Perspective comparison of paintings and photographs of Renaissance chandeliers“. In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Bd. 4. IEEE. Cambridge, UK, S. 645–648.
- Cupchik, Gerald C (1986). „A decade after Berlyne: New directions in experimental aesthetics“. In: *Poetics* 15.4-6, S. 345–369. DOI: 10.1016/0304-422X(86)90003-3.
- Cutzu, Florin, Riad Hammoud und Alex Leykin (2005). „Distinguishing paintings from photographs“. In: *Computer Vision and Image Understanding* 100.3, S. 249–273. DOI: 10.1016/j.cviu.2004.12.002.
- d’Agostino, Ralph B (1971). „An omnibus test of normality for moderate and large size samples“. In: *Biometrika* 58.2, S. 341–348.
- Dalal, Navneet und Bill Triggs (2005). „Histograms of oriented gradients for human detection“. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*. Bd. 1. IEEE, S. 886–893.
- Danto, Arthur Coleman (1981). *The Transfiguration of the Commonplace: A Philosophy of Art*. Harvard University Press, Cambridge, Mass.
- Datta, Ritendra, Dhiraj Joshi, Jia Li und James Z Wang (2006). „Studying aesthetics in photographic images using a computational approach“. In: *Lecture Notes in Computer Science* 3953.2, S. 288–301.

- Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li und Li Fei-Fei (2009). „Imagenet: A large-scale hierarchical image database“. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, S. 248–255.
- Denzler, Joachim, Erik Rodner und Marcel Simon (2016). „Convolutional Neural Networks as a computational model for the underlying processes of aesthetics perception“. In: *European Conference on Computer Vision*. Springer, Cham, S. 871–887.
- Dhar, Sagnik, Vicente Ordonez und Tamara L Berg (2011). „High level describable attributes for predicting aesthetics and interestingness“. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, S. 1657–1664.
- Donahue, Jeff, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng und Trevor Darrell (2014). „DeCAF: A deep convolutional activation feature for generic visual recognition“. In: *International Conference on Machine Learning*. Beijing, China, S. 647–655.
- Dong, Zhe, Xu Shen, Houqiang Li und Xinmei Tian (2015). „Photo quality assessment with DCNN that understands image well“. In: *International Conference on Multimedia Modeling*. Springer, Berlin, S. 524–535.
- Dowling, Christopher (2014). *Aesthetic Formalism*. (accessed January 26, 2018). URL: <http://www.iep.utm.edu/aes-form/>.
- Elgammal, Ahmed und Babak Saleh (2015). „Quantifying creativity in art networks“. In: *arXiv preprint arXiv:1506.00711*.
- Fechner, Gustav Theodor (1876). *Vorschule der Aesthetik*. Bd. 1. Breitkopf & Härtel, Leipzig.
- Fernandez, Dominic und Arnold J Wilkins (2008). „Uncomfortable images in art and nature“. In: *Perception* 37.7, S. 1098–1113.
- Field, David J, Anthony Hayes und Robert F Hess (1993). „Contour integration by the human visual system: evidence for a local “association field”“. In: *Vision Research* 33.2, S. 173–193.
- Forsythe, Alex, Marcos Nadal, Noel Sheehy, Camilo J Cela-Conde und Martin Sawey (2011). „Predicting beauty: fractal dimension and visual complexity in art“. In: *British Journal of Psychology* 102.1, S. 49–70. DOI: 10.1348/000712610X498958.

- Fukushima, Kunihiro (1980). „Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position“. In: *Biological Cybernetics* 36.4, S. 193–202.
- Galanter, Philip (2012). „Computational aesthetic evaluation: Past and future“. In: *Computers and Creativity*. Springer, Berlin, Heidelberg, S. 255–293.
- Gatys, Leon A, Alexander S Ecker und Matthias Bethge (2015a). „A neural algorithm of artistic style“. In: *arXiv preprint arXiv:1508.06576*.
- Gatys, Leon A., Alexander S. Ecker und Matthias Bethge (2015b). „Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks“. In: *CoRR* abs/1505.07376. URL: <http://arxiv.org/abs/1505.07376>.
- Gaut, Berys und Dominic Lopes (2005). *The Routledge Companion to Aesthetics*. Routledge.
- Geisler, Wilson S (2008). „Visual perception and the statistical properties of natural scenes“. In: *Annual Review of Psychology* 59, S. 167–192.
- Gershoni, Sharon und Shaul Hochstein (2011). „Measuring pictorial balance perception at first glance using Japanese calligraphy“. In: *i-Perception* 2.6, S. 508–527.
- Gómez-Puerto, Gerardo, Enric Munar und Marcos Nadal (2015). „Preference for curvature: A historical and conceptual framework“. In: *Frontiers in Human Neuroscience* 9.712. DOI: 10.3389/fnhum.2015.00712.
- Goodfellow, Ian J, Jonathon Shlens und Christian Szegedy (2014a). „Explaining and harnessing adversarial examples“. In: *arXiv preprint arXiv:1412.6572*.
- Goodfellow, Ian, Yoshua Bengio und Aaron Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville und Yoshua Bengio (2014b). „Generative Adversarial Nets“. In: *Advances in neural information processing systems*. Montréal, Canada, S. 2672–2680.
- Gopnik, Blake (2014). „Aesthetic science and artistic knowledge“. In: *Aesthetic Science: Connecting Minds, Brains, and Experience*. Hrsg. von A.P. Shimamura und S.E. Palmer. Oxford University Press, Oxford, S. 129–159.

- Graham, Daniel J und David J Field (2007). „Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities“. In: *Spatial Vision* 21.1, S. 149–164. DOI: 10.1163/156856807782753877.
- Graham, Daniel J und David J Field (2008). „Variations in intensity statistics for representational and abstract art, and for art from the Eastern and Western hemispheres“. In: *Perception* 37.9, S. 1341–1352.
- Graham, Daniel J und Christoph Redies (2010). „Statistical regularities in art: Relations with visual coding and perception“. In: *Vision Research* 50.16, S. 1503–1509. DOI: 10.1016/j.visres.2010.05.002.
- Graham, Daniel, Bianca Schwarz, Anjan Chatterjee und Helmut Leder (2016). „Preference for luminance histogram regularities in natural scenes“. In: *Vision Research* 120, S. 11–21. DOI: 10.1016/j.visres.2015.03.018.
- Grammer, Karl und Randy Thornhill (1994). „Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness.“ In: *Journal of Comparative Psychology* 108.3, S. 233–242.
- Green, Christopher D (1995). „All that glitters: A review of psychological research on the aesthetics of the golden section“. In: *Perception* 24.8, S. 937–968.
- Greenfield, Gary (2005). „On the origins of the term computational aesthetics“. In: *Proceedings of the First Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging*. Eurographics Association. Girona, Spain, S. 9–12.
- Güçlütürk, Yağmur, Richard HAH Jacobs und Rob van Lier (2016). „Liking versus complexity: decomposing the inverted U-curve“. In: *Frontiers in Human Neuroscience* 10.112. DOI: 10.3389/fnhum.2016.00112.
- Gunsel, Bilge, Sanem Sariel und Oguz Icoğlu (2005). „Content-based access to art paintings“. In: *IEEE International Conference on Image Processing 2005*. Bd. 2. IEEE. Genova, Italy, S. II–558.
- Hagerhall, Caroline M., Terry Purcell und Richard Taylor (2004). „Fractal dimension of landscape silhouette outlines as a predictor of landscape preference“. In: *Journal of Environmental Psychology* 24.2, S. 247–255.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren und Jian Sun (2015). „Deep residual learning for image recognition“. In: *arXiv preprint arXiv:1512.03385*.
- Heijer, Eelco den (2015). „Evolving symmetric and balanced art“. In: *Computational Intelligence*. Springer, S. 33–47.

- Henemann, Gudrun Maria, Anselm Brachmann und Christoph Redies (2017). „Statistical image properties in works from the Prinzhorn Collection of artists with schizophrenia“. In: *Frontiers in Psychiatry* 8.
- Hoenig, Florian (2005). „Defining computational aesthetics“. In: *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*. Eurographics Association. Girona, Spain, S. 13–18.
- Huang, Gao, Zhuang Liu, Kilian Q Weinberger und Laurens van der Maaten (2016). „Densely connected convolutional networks“. In: *arXiv preprint arXiv:1608.06993*.
- Hubel, David H und Torsten N Wiesel (1962). „Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex“. In: *The Journal of Physiology* 160.1, S. 106–154.
- Hughes, James M, Daniel J Graham und Daniel N Rockmore (2010). „Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder“. In: *Proceedings of the National Academy of Sciences USA* 107.4, S. 1279–1283.
- Hume, David (1757). *Four dissertations*. A. Millar, London.
- Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou und Alexei A Efros (2017). „Image-to-image translation with conditional adversarial networks“. In: *arXiv preprint arXiv:1611.07004*.
- Jacobsen, Thomas (2006). „Bridging the arts and sciences: A framework for the psychology of aesthetics“. In: *Leonardo* 39.2, S. 155–162. DOI: 10.1162/leon.2006.39.2.155.
- Jacobsen, Thomas und LEA Höfel (2002). „Aesthetic judgments of novel graphic patterns: analyses of individual judgments“. In: *Perceptual and Motor Skills* 95.3, S. 755–766.
- Jahanian, Ali, SVN Vishwanathan und Jan P Allebach (2015). „Learning visual balance from large-scale datasets of aesthetically highly rated images“. In: *SPIE/IS&T Electronic Imaging*. Bd. 9394. International Society for Optics und Photonics. San Francisco, CA, pages. DOI: 10.1117/12.2084548.
- Jia, Yangqing, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama und Trevor Darrell (2014). „Caffe: convolutional architecture for fast feature embedding“. In: *arXiv preprint arXiv:1408.5093*.

- Jiang, Shuqiang, Qingming Huang, Qixiang Ye und Wen Gao (2006). „An effective method to detect and categorize digitized traditional chinese paintings“. In: *Pattern Recognition Letters* 27.7, S. 734–746. DOI: 10.1016/j.patrec.2005.10.017.
- Johnson, C Richard, Ella Hendriks, Igor J Berezchnoy, Eugene Brevdo, Shannon M Hughes, Ingrid Daubechies, Jia Li, Eric Postma und James Z Wang (2008). „Image processing for artist identification“. In: *IEEE Signal Processing Magazine* 25.4, S. 37–48.
- Jones-Smith, Katherine und Harsh Mathur (2006). „Fractal analysis: revisiting Pollock’s drip paintings“. In: *Nature* 444.7119, E9–E10.
- Joye, Yannick (2007). „Fractal architecture could be good for you“. In: *Nexus Network Journal* 9.2, S. 311–320.
- Kao, Yueying, Ran He, Kaiqi Huang und Tieniu Tan (2016). „Deep Aesthetic quality assessment with semantic information“. In: *arXiv preprint arXiv:1604.04970*.
- Karayev, Sergey, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann und Holger Winnemoeller (2013). „Recognizing image style“. In: *arXiv preprint arXiv:1311.3715*.
- Karpathy, Andrej und Li Fei-Fei (2015). „Deep visual-semantic alignments for generating image descriptions“. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, S. 3128–3137.
- Karras, Tero, Timo Aila, Samuli Laine und Jaakko Lehtinen (2017). „Progressive growing of GANs for improved quality, stability, and variation“. In: *arXiv preprint arXiv:1710.10196*.
- Ke, Yan, Xiaoou Tang und Feng Jing (2006). „The design of high-level features for photo quality assessment“. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*. Bd. 1. IEEE, S. 419–426.
- Keren, Daniel (2002). „Painter identification using local features and naive bayes“. In: *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. Bd. 2. IEEE. Quebec, Canada, S. 474–477.
- Khan, Fahad Shahbaz, Joost van de Weijer und Maria Vanrell (2010). „Who painted this painting“. In: *2010 CREATE Conference*. Gjøvik, Norway, S. 329–333.

- Koch, Michael, Joachim Denzler und Christoph Redies (2010). „1/f² Characteristics and isotropy in the fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs“. In: *PLoS One* 5.8, e12268. DOI: 10.1371/journal.pone.0012268.
- Kozbelt, A und JC Kaufman (2014). „Aesthetics assessment“. In: *The Cambridge Handbook of the Psychology of Aesthetics and the Arts*. Cambridge University Press, Cambridge UK, S. 86–112.
- Krizhevsky, Alex, Ilya Sutskever und Geoffrey E Hinton (2012). „Imagenet classification with deep convolutional neural networks“. In: *Advances in Neural Information Processing Systems*. Lake Tahoe, NV, S. 1097–1105.
- Langer, Federico (2016). „Art theory for (neuro) scientists: Bridging the gap“. In: *Poetics Today* 37.4, S. 497–516.
- Lecun, Yann und Yoshua Bengio (1995). „Convolutional networks for images, speech, and time-series“. In: *The Handbook of Brain Theory and Neural Networks*. Hrsg. von M.A. Arbib. MIT Press, Cambridge MA.
- Lecun, Yann, Léon Bottou, Yoshua Bengio und Patrick Haffner (1998). „Gradient-based learning applied to document recognition“. In: *Proceedings of the IEEE*, S. 2278–2324.
- Leder, Helmut, Benno Belke, Andries Oeberst und Dorothee Augustin (2004). „A model of aesthetic appreciation and aesthetic judgments“. In: *British Journal of Psychology* 95.4, S. 489–508. DOI: 10.1348/0007126042369811.
- Leder, Helmut, Gernot Gerger, David Brieber und Norbert Schwarz (2014). „What makes an art expert? Emotion and evaluation in art appreciation“. In: *Cognition and Emotion* 28.6, S. 1137–1147. DOI: 10.1080/02699931.2013.870132.
- Leder, Helmut und Marcos Nadal (2014). „Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode—Developments and challenges in empirical aesthetics“. In: *British Journal of Psychology* 105.4, S. 443–464.
- Lennie, P (2000). „Color vision“. In: *Principles of Neural Science*. Hrsg. von Kandel ER, JH Schwartz und TM Jessell. 4th. McGraw-Hill, New York, S. 572–589.
- Leykin, Alex und Florin Cutzu (2003). „Differences of edge properties in photographs and paintings“. In: *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*. Bd. 3. IEEE. Barcelona, Spain, S. III–541.

- Li, Congcong und Tsuhan Chen (2009). „Aesthetic visual quality assessment of paintings“. In: *IEEE Journal of Selected Topics in Signal Processing* 3.2, S. 236–252.
- Li, Jia und James Ze Wang (2004). „Studying digital imagery of ancient paintings by mixtures of stochastic models“. In: *IEEE Transactions on Image Processing* 13.3, S. 340–353.
- Li, Jia, Lei Yao, Ella Hendriks und James Z Wang (2012). „Rhythmic brushstrokes distinguish van Gogh from his contemporaries: findings via automated brush-stroke extraction“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.6, S. 1159–1176.
- Little, Anthony C (2014). „Domain specificity in human symmetry preferences: symmetry is most pleasant when looking at human faces“. In: *Symmetry* 6.2, S. 222–233. DOI: 10.3390/sym6020222.
- Liu, Jingchen, George Slota, Gang Zheng, Zhaohui Wu, Minwoo Park, Seungkyu Lee, Ingmar Rauschert und Yanxi Liu (2013). „Symmetry detection from realworld images competition 2013: Summary and results“. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, S. 200–205.
- Liu, Yanxi, Hagit Hel-Or, Craig S Kaplan, Luc Van Gool et al. (2010). „Computational symmetry in computer vision and computer graphics“. In: *Foundations and Trends in Computer Graphics and Vision* 5.1–2, S. 1–195.
- Locher, P und C Nodine (1989). „The perceptual value of symmetry“. In: *Computers & Mathematics with Applications* 17.4-6, S. 475–484. DOI: 10.1016/0898-1221(89)90246-0.
- Locher, Paul J, Pieter Jan Stappers und Kees Overbeeke (1999). „An empirical evaluation of the visual rightness theory of pictorial composition“. In: *Acta Psychologica* 103.3, S. 261–280. DOI: 10.1016/S0001-6918(99)00044-X.
- Locher, Paul, Elizabeth A Krupinski, Claudia Mello-Thoms und Calvin F Nodine (2007). „Visual interest in pictorial art during an aesthetic experience“. In: *Spatial Vision* 21.1, S. 55–77. DOI: 10.1163/156856807782753868.
- Lowe, David G (2004). „Distinctive image features from scale-invariant keypoints“. In: *International Journal of Computer Vision* 60.2, S. 91–110.

- Lu, Xin, Zhe Lin, Hailin Jin, Jianchao Yang und James Z Wang (2015). „Rating Image Aesthetics Using Deep Learning“. In: *IEEE Transactions on Multimedia* 17.11, S. 2021–2034. DOI: 10.1109/TMM.2015.2477040.
- Luo, Yiwen und Xiaoou Tang (2008). „Photo and video quality evaluation: Focusing on the subject“. In: *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, S. 386–399.
- Lyssenko, Nathalie, Christoph Redies und Gregor U Hayn-Leichsenring (2016). „Evaluating abstract art: Relation between term usage, subjective ratings, image properties and personality traits“. In: *Frontiers in Psychology* 7.973. DOI: 10.3389/fpsyg.2016.00973.
- Lyu, Siwei, Daniel Rockmore und Hany Farid (2004). „A digital technique for art authentication“. In: *Proceedings of the National Academy of Sciences USA* 101.49, S. 17006–17010.
- Mahendran, Aravindh und Andrea Vedaldi (2015). „Understanding deep image representations by inverting them“. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, S. 5188–5196.
- Mallon, Birgit, Christoph Redies und Gregor Uwe Hayn-Leichsenring (2014). „Beauty in abstract paintings: perceptual contrast and statistical properties“. In: *Frontiers in Human Neuroscience* 8.161. DOI: 10.3389/fnhum.2014.00161.
- Mandelbrot, Benoit B und Roberto Pignoni (1983). *The Fractal Geometry of Nature*. W.H. Freeman, New York.
- Marelja, S (1980). „Mathematical description of the responses of simple cortical cells“. In: *JOSA* 70.11, S. 1297–1300.
- Marchesotti, Luca, Florent Perronnin, Diane Larlus und Gabriela Csurka (2011). „Assessing the aesthetic quality of photographs using generic image descriptors“. In: *2011 International Conference on Computer Vision*. IEEE, S. 1784–1791.
- Marković, Slobodan (2012). „Components of aesthetic experience: aesthetic fascination, aesthetic appraisal, and aesthetic emotion“. In: *i-Perception* 3.1, S. 1–17. DOI: 10.1068/i0450aap.
- Martinez, Aleix M (1998). „The AR face database“. In: *CVC Technical Report* 24.
- Mather, George (2014). „Artistic adjustment of image spectral slope“. In: *Art & Perception* 2.1-2, S. 11–22.
- McManus, IC, Richard Cook und Amy Hunt (2010). „Beyond the golden section and normative aesthetics: Why do individuals differ so much in their aesthetic

- preferences for rectangles?“ In: *Psychology of Aesthetics, Creativity, and the Arts* 4.2, S. 113–126.
- McManus, IC, D Edmondson und J Rodger (1985). „Balance in pictures“. In: *British Journal of Psychology* 76.3, S. 311–324.
- McManus, IC, Katharina Stöver und Do Kim (2011). „Arnheim’s Gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images“. In: *i-Perception* 2.6, S. 615–647.
- Menzel, Claudia, Gregor U Hayn-Leichsenring, Oliver Langner, Holger Wiese und Christoph Redies (2015). „Fourier power spectrum characteristics of face photographs: attractiveness perception depends on low-level image properties“. In: *PloS One* 10.4, e0122801. DOI: 10.1371/journal.pone.0122801.
- Møller, Anders Pape und John P Swaddle (1997). *Asymmetry, Developmental Stability and Evolution*. Oxford University Press, UK.
- Montagner, Cristina, Rui Jesus, Nuno Correia, Márcia Vilarigues, Rita Macedo und Maria João Melo (2016). „Features combination for art authentication studies: Brushstroke and materials analysis of Amadeo de Souza-Cardoso“. In: *Multimedia Tools and Applications* 75.7, S. 4039–4063.
- Mullin, Caitlin, Gregor Hayn-Leichsenring, Christoph Redies und Johan Wagemans (2017). „The gist of beauty: An investigation of aesthetic perception in rapidly presented images“. In: *Electronic Imaging* 2017.14, S. 248–256.
- Munar, Enric, Gerardo Gómez-Puerto, Josep Call und Marcos Nadal (2015). „Common visual preference for curved contours in humans and great apes“. In: *PloS One* 10.11, e0141106. DOI: 10.1371/journal.pone.0141106.
- Mureika, Jonas R (2005). „Fractal dimensions in perceptual color space: a comparison study using Jackson Pollock’s art“. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 15.4, S. 043702.
- Murray, Naila, Luca Marchesotti und Florent Perronnin (2012). „AVA: A large-scale database for aesthetic visual analysis“. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, S. 2408–2415.
- Nascimento, Sérgio MC, João MM Linhares, Cristina Montagner, Catarina AR João, Kinjiro Amano, Catarina Alfaro und Ana Bailão (2017). „The colors of paintings and viewers’ preferences“. In: *Vision Research* 130, S. 76–84.
- OED (2017). *Oxford English Dictionary Online*. (accessed May 24, 2017). URL: <http://www.oed.com/viewdictionaryentry/Entry/293508>.

- O'Hare, Louise und Paul B Hibbard (2011). „Spatial frequency and visual discomfort“. In: *Vision Research* 51.15, S. 1767–1777. DOI: 10.1016/j.visres.2011.06.002.
- Oliva, Aude und Antonio Torralba (2001). „Modeling the shape of the scene: A holistic representation of the spatial envelope“. In: *International Journal of Computer Vision* 42.3, S. 145–175.
- Orians, Gordon H und Judith H Heerwagen (1992). „Evolved responses to landscapes“. In:
- Osborne, Harold (1986). „Symmetry as an aesthetic factor“. In: *Computers & Mathematics with Applications* 12.1-2, S. 77–82. DOI: 10.1016/0898-1221(86)90140-9.
- Palmer, Stephen E und Karen B Schloss (2010). „An ecological valence theory of human color preference“. In: *Proceedings of the National Academy of Sciences USA* 107.19, S. 8877–8882.
- Palmer, Stephen E, Karen B Schloss und William S Griscom (2016). „Individual differences in perceptual preference“. In: *Electronic Imaging* 2016.16, S. 1–6.
- Papaodysseus, Constantin, Dimitrios K Fragoulis, Mihalis Panagopoulos, Thanasis Panagopoulos, Panayiotis Rousopoulos, Mihalis Exarhos und Angelos Skembris (2006). „Determination of the method of construction of 1650 BC wall paintings“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.9, S. 1361–1371.
- Peitgen, Heinz-Otto, Hartmut Jürgens und Dietmar Saupe (1992). *Chaos and Fractals*. New York: Springer-Verlag.
- Pinto, Paulo D, João MM Linhares, João A Carvalhal und Sérgio MC Nascimento (2006). „Psychophysical estimation of the best illumination for appreciation of Renaissance paintings“. In: *Visual Neuroscience* 23.3-4, S. 669–674. DOI: 10.1017/S0952523806233340.
- Platt, John C, Nello Cristianini und John Shawe-Taylor (2000). „Large margin DAGs for multiclass classification“. In: *Advances in neural information processing systems*, S. 547–553.
- Polatkan, Güngör, Sina Jafarpour, Andrei Brasoveanu, Shannon Hughes und Ingrid Daubechies (2009). „Detection of forgery in paintings using supervised learning“. In: *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE. Cairo, Egypt, S. 2921–2924.

- Radford, Alec, Luke Metz und Soumith Chintala (2015). „Unsupervised representation learning with Deep Convolutional Generative Adversarial Networks“. In: *arXiv preprint arXiv:1511.06434*.
- Rampone, Giulia, Noreen O’Sullivan und Marco Bertamini (2016). „The role of visual eccentricity on preference for abstract symmetry“. In: *PloS One* 11.4, e0154428. DOI: 10.1371/journal.pone.0154428.
- Reber, Rolf, Norbert Schwarz und Piotr Winkielman (2004). „Processing fluency and aesthetic pleasure: Is beauty in the perceiver’s processing experience?“ In: *Personality and Social Psychology Review* 8.4, S. 364–382.
- Redies, Christoph (2007a). „A universal model of esthetic perception based on the sensory coding of natural stimuli“. In: *Spatial Vision* 21.1, S. 97–117. DOI: 10.1163/156856807782753886.
- Redies, Christoph (2007b). „A universal model of esthetic perception based on the sensory coding of natural stimuli“. In: *Spatial Vision* 21.1, S. 97–117.
- Redies, Christoph (2015). „Combining universal beauty and cultural context in a unifying model of visual aesthetic experience“. In: *Frontiers in Human Neuroscience* 9.218. DOI: 10.3389/fnhum.2015.00218.
- Redies, Christoph, Seyed Ali Amirshahi, Michael Koch und Joachim Denzler (2012). „PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects“. In: *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, S. 522–531. DOI: 10.1007/978-3-642-33863-2_54.
- Redies, Christoph und Anselm Brachmann (2017). „Statistical image properties in large subsets of traditional art, Bad Art, and abstract art“. In: *Frontiers in Neuroscience* 11. DOI: 10.3389/fnins.2017.00593.
- Redies, Christoph, Anselm Brachmann und Gregor Uwe Hayn-Leichsenring (2015). „Changes of statistical properties during the creation of graphic artworks“. In: *Art & Perception* 3.1, S. 93–116. DOI: 10.1163/22134913-00002017.
- Redies, Christoph, Anselm Brachmann und Johan Wagemans (2017). „High entropy of edge orientations characterizes visual artworks from diverse cultural backgrounds“. In: *Vision Research* 133, S. 130–144. DOI: 10.1016/j.visres.2017.02.004.
- Redies, Christoph, Jan Hänisch, Marko Blickhan und Joachim Denzler (2007a). „Artists portray human faces with the Fourier statistics of complex natural

- scenes“. In: *Network: Computation in Neural Systems* 18.3, S. 235–248. DOI: 10.1080/09548980701574496.
- Redies, Christoph, Jens Hasenstein und Joachim Denzler (2007b). „Fractal-like image statistics in visual art: similarity to natural scenes“. In: *Spatial Vision* 21.1, S. 137–148. DOI: 10.1163/156856807782753921.
- Renoult, Julien P, Jeanne Bovet und Michel Raymond (2016). „Beauty is in the efficient coding of the beholder“. In: *Royal Society Open Science* 3.3, S. 160027. DOI: 10.1098/rsos.160027.
- Rentschler, Ingo und Erhardt Barth (2001). „Bilder im Kopf oder vom Ursprung des Schönen“. In: *Welten im Kopf. Texte zum Imaginativen Lernen*. Hrsg. von I. Rentschler, E. Madelung und P. Fauser. Hamburg: Kallwey Verlag.
- al-Rifaie, Mohammad Majid, Anna Ursyn, Robert Zimmer und Mohammad Ali Javaheri Javid (2017). „On symmetry, aesthetics and quantifying symmetrical complexity“. In: *International Conference on Evolutionary and Biologically Inspired Music and Art*. Springer, Cham, S. 17–32.
- Rigau, Jaume, Miquel Feixas und Mateu Sbert (2008). „Informational aesthetics measures“. In: *IEEE Computer Graphics and Applications* 28.2, S. 24–34.
- Rumelhart, David E, Geoffrey E Hinton und Ronald J Williams (1986). „Learning representations by back-propagating errors“. In: *Nature* 323.6088, S. 533.
- Russakovsky, Olga et al. (2015). „ImageNet Large Scale Visual Recognition Challenge“. In: *International Journal of Computer Vision* 115.3, S. 211–252.
- Saleh, Babak und Ahmed Elgammal (2015). „Large-scale classification of fine-art paintings: Learning the right metric on the right feature“. In: *arXiv preprint arXiv:1505.00855*.
- Sarkar, Nirupam und BB Chaudhuri (1994). „An efficient differential box-counting approach to compute fractal dimension of image“. In: *IEEE Transactions on Systems, Man, and Cybernetics* 24.1, S. 115–120.
- Schweinhart, April M und Edward A Esock (2013). „Structural content in paintings: Artists overregularize oriented content of paintings relative to the typical natural scene bias“. In: *Perception* 42.12, S. 1311–1332. DOI: 10.1068/p7345.
- Shaker, Fariba und Amirhassan Monadjemi (2015). „A new symmetry measure based on gabor filters“. In: *2015 23rd Iranian Conference on Electrical Engineering*. IEEE, S. 705–710.

- Shamir, Lior (2015). „What makes a Pollock Pollock: a machine vision approach“. In: *International Journal of Arts and Technology* 8.1, S. 1–10.
- Shamir, Lior, Tomasz Macura, Nikita Orlov, D Mark Eckley und Ilya G Goldberg (2010). „Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art“. In: *ACM Transactions on Applied Perception (TAP)* 7.2, S. 8.
- Shamir, Lior, Nikita Orlov, D Mark Eckley, Tomasz Macura, Josiah Johnston und Ilya G Goldberg (2008). „Wndchrm—an open source utility for biological image analysis“. In: *Source Code for Biology and Medicine* 3.1, S. 1.
- Shamir, Lior und Jane A Tarakhovsky (2012). „Computer analysis of art“. In: *Journal on Computing and Cultural Heritage* 5.2, S. 7.
- Shannon, C. E. (1948). „A mathematical theory of communication“. In: *The Bell System Technical Journal* 27.3, S. 379–423. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
- Shen, Jialie (2009). „Stochastic modeling Western paintings for effective classification“. In: *Pattern Recognition* 42.2, S. 293–301. DOI: 10.1016/j.patcog.2008.04.016.
- Shimamura, Arthur P (2014). „Toward a science of aesthetics: Issues and ideas“. In: *Aesthetic Science: Connecting Minds, Brains, and Experience*. Hrsg. von A.P. Shimamura und S.E. Palmer. Oxford University Press, Oxford, S. 3–28.
- Siddiquie, Behjat, Shiv N Vitaladevuni und Larry S Davis (2009). „Combining multiple kernels for efficient image classification“. In: *Applications of Computer Vision (WACV), 2009 Workshop on*. IEEE. Snowbird, UT, USA, S. 1–8.
- Sigman, Mariano, Guillermo A Cecchi, Charles D Gilbert und Marcelo O Magnasco (2001). „On a common circle: natural scenes and Gestalt rules“. In: *Proceedings of the National Academy of Sciences USA* 98.4, S. 1935–1940.
- Silvia, Paul J (2005). „Emotional responses to art: From collation and arousal to cognition and emotion“. In: *Review of General Psychology* 9.4, S. 342–357.
- Silvia, Paul J (2014). „Human emotions and aesthetic experience: An overview of empirical aesthetics“. In: *Aesthetic Science: Connecting Minds, Brains, and Experience*. Hrsg. von A.P. Shimamura und S.E. Palmer. Oxford University Press, Oxford, S. 250–275.
- Simoncelli, Eero P und Bruno A Olshausen (2001). „Natural image statistics and neural representation“. In: *Annual Review of Neuroscience* 24.1, S. 1193–1216.

- Simonyan, Karen und Andrew Zisserman (2014). „Very deep convolutional networks for large-scale image recognition“. In: *arXiv preprint arXiv:1409.1556*.
- Spehar, Branka, Nicholas Walker und Richard P Taylor (2016). „Taxonomy of individual variations in aesthetic responses to fractal patterns“. In: *Frontiers in Human Neuroscience* 10.350. DOI: 10.3389/fnhum.2016.00350.
- Spillmann, Lothar, Birgitta Dresch-Langley und Chia-huei Tseng (2015). „Beyond the classical receptive field: the effect of contextual stimuli“. In: *Journal of Vision* 15.9, S. 7–7.
- Spratt, Emily L und Ahmed Elgammal (2014). „Computational beauty: Aesthetic judgment at the intersection of art and science“. In: *Workshop at the European Conference on Computer Vision*. Springer, S. 35–53.
- Stork, David G (2009a). „Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature“. In: *International Conference on Computer Analysis of Images and Patterns*. Springer, Berlin, Heidelberg, S. 9–24.
- Stork, David G (2009b). „Learning-based authentication of Jackson Pollock’s paintings“. In: *SPIE Professional*.
- Stork, David G und M Kimo Johnson (2006). „Estimating the location of illuminants in realist master paintings Computer image analysis addresses a debate in art history of the Baroque“. In: *18th International Conference on Pattern Recognition (ICPR’06)*. Bd. 1. IEEE, S. 255–258.
- Su, Jiawei, Danilo Vasconcellos Vargas und Sakurai Kouichi (2017). „One pixel attack for fooling deep neural networks“. In: *arXiv preprint arXiv:1710.08864*.
- Szegedy, Christian, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow und Rob Fergus (2013). „Intriguing properties of neural networks“. In: *arXiv preprint arXiv:1312.6199*.
- Tan, Wei Ren, Chee Seng Chan, Hernán E Aguirre und Kiyoshi Tanaka (2016). „Ceci n’est pas une pipe: A deep convolutional network for fine-art paintings classification“. In: *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE. Phoenix, AZ, USA, S. 3703–3707.
- Taylor, Richard P (2002). „Order in Pollock’s chaos“. In: *Scientific American* 287.6, S. 84–89.
- Taylor, Richard P (2006). „Reduction of physiological stress using fractal art and architecture“. In: *Leonardo* 39.3, S. 245–251.

- Taylor, Richard P, Adam P Micolich und David Jonas (1999). „Fractal analysis of Pollock’s drip paintings“. In: *Nature* 399.6735, S. 422–422.
- Taylor, Richard P und Branka Spehar (2016). „Fractal Fluency: An intimate relationship between the brain and processing of fractal stimuli“. In: *The Fractal Geometry of the Brain*. Hrsg. von Antonio Di Ieva. Springer, New York, S. 485–496.
- Taylor, Richard P, Branka Spehar, Paul Van Donkelaar und Caroline M Hagerhall (2011). „Perceptual and physiological responses to Jackson Pollock’s fractals“. In: *Frontiers in Human Neuroscience* 5.50. DOI: 10.3389/fnhum.2011.00060.
- Tinio, P.P.L. und J.K. Smith (2014). *The Cambridge Handbook of the Psychology of Aesthetics and the Arts*. Cambridge Handbooks in Psychology. Cambridge University Press. URL: <https://books.google.de/books?id=b82KBAAQBAJ>.
- Tolhurst, DJ, Y_ Tadmor und Tang Chao (1992). „Amplitude spectra of natural images“. In: *Ophthalmic and Physiological Optics* 12.2, S. 229–232.
- Tong, Hanghang, Mingjing Li, Hong-Jiang Zhang, Jingrui He und Changshui Zhang (2004). „Classification of digital photos taken by photographers or home users“. In: *Pacific-Rim Conference on Multimedia*. Springer, S. 198–205.
- Torresani, Lorenzo, Martin Szummer und Andrew Fitzgibbon (2010). „Efficient object category recognition using classes“. In: *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, S. 776–789.
- Vapnik, Vladimir N und Alexey J Chervonenkis (1974). „Theory of pattern recognition“. In:
- Vartanian, Oshin und Martin Skov (2014). „Neural correlates of viewing paintings: evidence from a quantitative meta-analysis of functional magnetic resonance imaging data“. In: *Brain and Cognition* 87, S. 52–56.
- Wagemans, Johan (1995). „Detection of visual symmetries“. In: *Spatial Vision* 9.1, S. 9–32.
- Wagemans, Johan (1997). „Characteristics and models of human symmetry detection“. In: *Trends in Cognitive Sciences* 1.9, S. 346–352.
- Wallraven, Christian, Roland Fleming, Douglas Cunningham, Jaume Rigau, Miquel Feixas und Mateu Sbert (2009). „Categorizing art: Comparing humans and computers“. In: *Computers & Graphics* 33.4, S. 484–495. DOI: 10.1016/j.cag.2009.04.003.

- Wang, Ying und Masahiro Takatsuka (2012). „A framework towards quantified artistic influences analysis“. In: *Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference on*. IEEE, S. 1–8.
- Westphal-Fitch, Gesche, Jinook Oh und W Fitch (2013). „Studying aesthetics with the method of production: Effects of context and local symmetry.“ In: *Psychology of Aesthetics, Creativity, and the Arts* 7.1, S. 13.
- Widjaja, Ivo, Wee Kheng Leow und Fang-Cheng Wu (2003). „Identifying painters from color profiles of skin patches in painting images“. In: *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*. Bd. 1. IEEE. Barcelona, Spain, S. I–845.
- Wurtz, RH und ER Kandel (2000a). „Central visual pathway“. In: *Principles of Neural Science*. Hrsg. von ER Kandel, JH Schwartz und TM Jessell. 4th. McGraw-Hill, New York, S. 523–547.
- Wurtz, RH und ER Kandel (2000b). „Central visual pathway“. In: *Principles of Neural Science*. Hrsg. von ER Kandel, JH Schwartz und TM Jessell. 4th. McGraw-Hill, New York, S. 523–547.
- Yosinski, Jason, Jeff Clune, Yoshua Bengio und Hod Lipson (2014). „How transferable are features in deep neural networks?“ In: S. 3320–3328.
- Yosinski, Jason, Jeff Clune, Anh Nguyen, Thomas Fuchs und Hod Lipson (2015). „Understanding neural networks through deep visualization“. In: *arXiv preprint arXiv:1506.06579*.
- Zabrodsky, Hagit, Shmuel Peleg und David Avnir (1995). „Symmetry as a continuous feature“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.12, S. 1154–1166.
- Zaidel, Dahlia W, Shawn M Aarde und Kiran Baig (2005). „Appearance of symmetry, beauty, and health in human faces“. In: *Brain and Cognition* 57.3, S. 261–263.
- Zeiler, Matthew und Rob Fergus (2014). „Visualizing and understanding convolutional networks“. In: *European Conference on Computer Vision*. Springer, S. 818–833.
- Zeki, Semir (2001). „Artistic creativity and the brain“. In: *Science* 293.5527, S. 51–52.

- Zeki, Semir (2013). „Clive Bell’s “Significant Form” and the neurobiology of aesthetics“. In: *Frontiers in Human Neuroscience* 7.730. DOI: 10.3389/fnhum.2013.00730.
- Zhu, Jun-Yan, Taesung Park, Phillip Isola und Alexei A Efros (2017). „Unpaired image-to-image translation using cycle-consistent adversarial networks“. In: *arXiv preprint arXiv:1703.10593*.
- Zujovic, Jana, Lisa Gandy, Scott Friedman, Bryan Pardo und Thrasyvoulos N Pappas (2009). „Classifying paintings by artistic genre: An analysis of features & classifiers“. In: *Multimedia Signal Processing, 2009. MMSP’09. IEEE International Workshop on*. IEEE. Rio De Janeiro, Brazil, S. 1–5.

Anhang A

Allgemeine Ergänzungen

A.1 Verzeichnis abgebildeter Kunstwerke

Tabelle A.1

Abb.	Autor	Jahr	Titel
2.1	Benjamin West	1775	Isaac's Servant Tying the Bracelet on Rebecca's Arm
4.1(h)	Benjamin West	1775	Isaac's Servant Tying the Bracelet on Rebecca's Arm
4.8(a)	Pierre-Auguste Renoir	1875	Snow-covered Landscape
4.8(b)	Theo van Rysselberghe	1904	The Mediterranean at Le Lavandou
4.8(c)	Camille Pissarro	1887	Peasants' houses, Eragny
4.8(d)	Domenico Zampieri	1610	The Way to Calvary
4.8(e)	Jacob Jordaens	1645	The Feast of the Bean King
4.8(f)	Pieter Bruegel	1569	Peasant Wedding
4.8(g)	Gustave Caillebotte	1882	Fruit Displayed on a Stand
4.8(h)	Claude Monet	1922	Water Lilies
4.8(i)	Robert William	1888	Poppies
5.1(c)	Juan de Arellano	1671	Small Basket of Flowers

A.2 Ergänzungen zu Kapitel 5

Tabelle A.2: Werte der Varianzen $P_a(22)$ und $P_f(12)$ in den Abbildungen 5.1, 5.8, 5.6 und 5.9 gezeigter Bilder.

Abb.	Bild	$P_a(22) \cdot 10^4$	$P_f(12) \cdot 10^5$
5.1a	Islamic art by Bashdan Qara	1.05	1.48
5.1a	Chinese art by Yun Shouping	2.40	2.21
5.1a	Western painting by J. DeArellano	0.93	1.26
5.1b	Buildings	2.04	1.28
5.1b	Urban scene	7.03	2.42
5.1b	Objects	3.78	5.03
5.1b	Large vista scene	2.98	1.87
5.1b	Plant pattern	1.33	1.06
5.1b	Vegetation	0.94	0.57
5.6a	Plant pattern (flowers)	1.38	0.23
5.6b	Large vista scene (lake)	9.73	0.24
5.6c	Western painting by A. Mengs	1.42	3.35
5.6d	Western painting by C. Monet	0.57	0.66
5.8a	Plant pattern	1.24	1.35
5.8b	Western painting by M. Chase	3.65	1.32
5.8c	Large vista scene	1.02	1.02
5.8d	Western painting by P. Cezanne	0.98	0.75

Abb.	Bild	$P_a(22) \cdot 10^4$	$P_f(12) \cdot 10^5$
5.9	Plant pattern (moss)	1.02	1.07
5.9	Plant pattern (flowers)	1.24	1.35
5.9	Large vista scene (waterfall)	1.08	1.16
5.9	Large vista scene	1.41	1.78
5.9	Vegetation (poppy flowers)	0.99	1.10
5.9	Vegetation (pond)	1.03	1.20
5.9	Vegetation (forest)	0.94	1.21
5.9	Large vista scene (river)	1.05	1.39
5.9	Large vista scene (rocks)	0.93	1.21
5.9	Chinese art by Lang Shining	2.47	1.31
5.9	Western painting by H. Avercamp	2.55	1.33
5.9	Chinese art by Xu Yang	0.98	0.54
5.9	Western painting by E. Manet	0.98	0.52
5.9	Western painting by C. Monet	1.37	0.74
5.9	Western painting by H. LeSidaner	1.45	0.78
5.9	Western painting by C. Monet	0.90	0.46
5.9	Western painting by C. Hassam	1.22	0.64
5.9	Western painting by E. Boudin	1.93	0.98

Lebenslauf

Persönliche Daten

Name Friedrich Anselm Brachmann
Geboren 21. Juli 1989 in Jena
E-Mail anselm.brachmann@fu-berlin.de



Ausbildung

- 2011 – 2013 Masterstudium Informatik an der Freien Universität Berlin
Vertiefungsgebiet Künstliche Intelligenz und Robotik
Nebenfach Philosophie
Thema der Masterarbeit: *Erweiterung von Fast Retina Keypoints um Farbinformation*
- 2008 – 2011 Bachelorstudium Informatik an der Freien Universität Berlin
Nebenfach Psychologie
Thema der Bachelorarbeit: *Unüberwachte Klassifizierung von Nachtigallstrophen mittels tiefer Autoencoder*

2000 – 2008 Abitur am Christlichen Gymnasium Jena
Leistungskurse Mathematik und Informatik

Praktische Tätigkeiten

2015 – Wissenschaftlicher Mitarbeiter in der Arbeitsgruppe
Experimentelle Ästhetik am Institut für Anatomie an der
Friedrich Schiller Universität Jena

2014 – 2015 Wissenschaftlicher Mitarbeiter am *Labor für Neurorobotik*
Institut für Informatik an der Humboldt Universität Berlin,
später *Forschungslabor Neurorobotik* an der
Beuth Hochschule für Technik Berlin

2011 – 2012 Geprüfte wissenschaftliche Hilfskraft in der Arbeitsgruppe
Experimentelle Ästhetik am Institut für Anatomie an der
Friedrich Schiller Universität Jena