BILDER IM CHAOS

BILDER IM CHAOS

BILDER IM CHAOS

DIE GRAMMATIK DER MUSTER

BJÖRN OMMER

Das menschliche Auge nimmt nur Helligkeitsunterschiede und Farben wahr. Unser Gehirn macht daraus Objekte, beispielsweise ein Haus, einen Baum, einen Menschen. Die Muster, nach denen das Hirn Gestalt bildet, hat es zuvor erlernt. Kann man die Fähigkeit, geordnete Strukturen aus einer komplexen Umwelt herauszufiltern, auch Computern beibringen? Wenn Maschinen erst die Grammatik der Muster verstehen, könnten sie vielleicht bald wie wir Menschen sehen.

Das Beobachten und Begreifen von Ordnung und Regularität in der ihn umgebenden Welt übt auf den Menschen von jeher eine außerordentliche Faszination aus. Wiederkehrende Muster in einem sich scheinbar chaotisch verhaltenden Umfeld zu erkennen, ist eine intellektuelle Leistung und die Grundlage für weiteren Fortschritt. So sagte der französische Dichter Paul Claudel zu Recht: „Die Ordnung ist die Lust der Vernunft, aber die Unordnung ist die Wonne der Phantasie."

Auf welcher Skala auch immer wir die Welt um uns herum beobachten – Struktur und Regularität sind überall evident: Auf einer großen Skala entstehen aus formlosen Materiewolken durch die Gravitation Sterne, Sternensysteme und schließlich Galaxien mit Milliarden von Sternen. Auf der unmittelbaren Skala, die wir direkt mit unseren Augen erfassen können, erschließen sich uns die komplexen geordneten Muster und Formen, die Tiere, Pflanzen und leblose Objekte aufweisen, zum Beispiel die Symmetrie von Farnen, Seesternen oder Schneeflocken. Auf einer kleinen Skala reguliert die hochkomplexe Struktur des Erbmoleküls DNS die Entwicklung, Funktion und schließlich auch die Form und Struktur lebender Organismen. Regularität ist jedoch nicht nur räumlich ausgeprägt. Auch zeitliche Muster sind allgegenwärtig, etwa die Abfolge von Frühling, Sommer, Herbst und Winter oder der rhythmische Schlag unseres Herzens.

Struktur und Regularität, insbesondere Symmetrien, bewirken somit die Ausprägung verschiedenster Muster in Raum und Zeit. Dass solche komplexen geordneten Strukturen überhaupt existieren, ist erstaunlich. Schließlich impliziert der „Zweite Hauptsatz der Thermodynamik", eine zentrale Annahme der Physik, dass die Entropie – der Grad der „Unordnung" – in einem isolierten System, beispielsweise im Universum, stetig ansteigt. Erstaunlich ist auch, wie robust sich die geordneten Strukturen gegenüber störenden äußeren Einflüssen erweisen. Beides sind Grundvoraussetzungen für die Existenz von Leben.

„Auf den ersten Blick erscheint uns die Welt hoffnungslos komplex und chaotisch. Und doch sind überall in ihr Ordnung und Regularität zu erkennen."

Die Faszination des Menschen für Muster ist folglich nicht weiter verwunderlich. Sie äußert sich auch darin, dass der Mensch immer wieder selbst in bildender Kunst, Literatur und Musik Muster erschafft. Etwa durch den Einsatz von Stilmitteln wie Symmetrie, Versmaß und Kontrapunkt. Letztlich gipfeln die Bemühungen im Streben, die Welt besser zu verstehen und mittels erkannter Regularität zukünftiges Geschehen vorherzusagen und auf Vergangenes rückzuschließen.

Was die Welt im Innersten zusammenhält

Muster und die Suche nach grundlegender Ordnung spielen also schon immer eine wichtige Rolle im menschlichen Denken. Im Vordergrund steht die Suche nach einer beschränkten Menge simpler Regeln, Gesetze oder Beziehungen, die – zusammen mit einfachen physischen Entitäten – komplexere Phänomene erklären können und helfen, die Welt begreifbar zu machen. Komplexe Phänomene werden somit durch das Erkennen einfacher, grundlegender Muster erklärt und durch das Ableiten konstruktiver Regeln vorhersagbar. Dabei wird schließlich ein Modell der Wirklichkeit – oder einzelner Aspekte derselben – entworfen, das diese beschreibt, weitergehende Aussagen ermöglicht und damit überprüfbar ist.
Die Bedeutung der Regelmäßigkeiten, die in diesem Prozess erkannt werden, zeigt sich gerade auch darin, dass experimentell validierte physikalische Modelle als „Naturgesetze" bezeichnet werden. Im Alltagsdenken wird „Mutter Natur" damit gleichsam zu einem Subjekt stilisiert, das sich an diese Gesetze halten soll; andernfalls drohen katastrophale Folgen.

Halten wir fest: Um sich in unserer Welt zurechtzufinden, um sie zu verstehen, auf Vergangenes rückschließen und künftige Geschehnisse voraussagen zu können, ist es unabdingbar, im scheinbaren Chaos Ordnung und Muster zu erkennen und daraus abstrakte Konzepte und Regeln, „Modelle", abzuleiten. Doch wie gelingt uns Menschen das? Wie erstellen wir Modelle der Wirklichkeit? Kann auch eine künstliche Intelligenz – eine Maschine wie der Computer – dazu gebracht werden, automatisch Modelle der Wirklichkeit zu erstellen? Dies ist eine Kernherausforderung der statistischen Mustererkennung und des maschinellen Lernens. Und es ist eine der wesentlichen Aufgaben im Bereich der künstlichen Intelligenz und der „Computer Vision".

Viele Muster – ein Modell

Die meisten Modelle werden erlernt. Das heißt: Wir haben in der komplexen Welt eine Regularität beobachtet und gelernt, sie mit einem abstrakten Modell zu beschreiben. Dieses Modell wenden wir dann stets aufs Neue an, um Objekte wiederzuerkennen und die Welt zu erschließen. Wie aber eignen wir uns diese Muster an? Woran erkennen wir beispielsweise eine reife Erdbeere? Eine erste Antwort darauf gab der kanadische Psychobiologe Donald Hebb in den 1940er-Jahren. Hebb ging davon aus, dass das Gehirn des Menschen kein starres Gebilde ist, sondern beständig Assoziationen zwischen Sinneseindrücken herstellt. Auf das Beispiel Erdbeere bezogen heißt das: Es verknüpft visuelle Eindrücke, die häufig gemeinsam mit dem Objekt „reife Erdbeere" auftreten. Das erscheint einfach – hätte unsere Welt nicht eben auch eine chaotische Seite, die sich darin zeigt, dass es keine Erdbeere gibt, die der anderen gleicht: Die äußere Form ist unterschiedlich, die Nüsschen auf ihrer Oberfläche sind anders verteilt; auch äußere Einflüsse, etwa das Licht, lassen die Farbe der Erdbeeren anders aussehen. Was unserem Gehirn ganz selbstverständlich gelingt, ist für die künstliche Intelligenz eine große Herausforderung, gilt es doch, von vielen Mustern reifer Erdbeeren auf ein einziges abstraktes Modell zu schließen.

Eine Vielzahl von Modellen käme in Frage, um die ganze Variabilität der bisher beobachteten Muster beschreiben zu können. Für welches aber sollen wir uns entscheiden? Da wir immer nur eine begrenzte Menge an Trainingsbeispielen sehen und Erdbeeren schließlich essen, nicht studieren wollen, bevorzugen Menschen intuitiv das simpelste Modell. Dieses basiert auf einer möglichst kleinen Anzahl von Variablen und Vorannahmen und lässt sich folglich leicht erlernen – selbst wenn es nur eine eingeschränkte Anzahl von Trainingsbeispielen gibt. Die Anforderungen, die an das maschinelle Lernen gestellt werden müssen, lassen sich nach diesen Vorbetrachtungen nun zumindest in groben Zügen überblicken: Um Muster zu erkennen, bedarf es zunächst einer gewissen Anzahl von Trainings-

beispielen. Anhand dieser Daten wird dann ein Modell erlernt, wobei Einfachheit und Fehlerquote (Performanz) gegeneinander abgewogen werden. Nun lässt sich das Modell auf neue, bisher nicht beobachtete Beispiele anwenden, und es zeigt sich, wie gut ein Modell das zuvor erlernte Muster in neuen Daten erkennen kann. Damit lässt sich die Güte des Modells objektiv evaluieren.

Grenzen der Mustererkennung

An dieser Stelle lohnt es, die Grenzen der Mustererkennung zu diskutieren. Je komplexer beispielsweise ein Muster ist, desto schwerer ist es zu erlernen und desto größer ist der Trainingsdatensatz, der hierfür erforderlich ist: Ein Laternenpfahl etwa lässt sich leichter repräsentieren als ein sich bewegender menschlicher Körper – dies liegt an der sogenannten Intraklassenvariabilität. Aber auch auf die „Interklassenähnlichkeit" kommt es an: In einem leeren Raum ist ein Mensch einfacher zu detektieren als in einem Raum, in dem sich noch Gegenstände oder gar eine Menge Schimpansen befinden. Wenn dann womöglich auch noch schlechte Lichtverhältnisse hinzukommen, wird es schwierig, das visuelle Signal, das uns interessiert – den einzelnen Menschen –, von dem uninteressanten Hintergrundrauschen – weiteren Objekten im Raum – zu unterscheiden. Einen Menschen unter diesen Bedingungen maschinell erkennen zu wollen, ist vergleichbar mit dem Versuch, eine Stimme oder eine Melodie aus einer Symphonie herauszuhören, deren Thema ständig variiert. Struktur und Chaos stehen bei der Wahrnehmung gewissermaßen im Wettstreit miteinander, so dass wir kaum in der Lage sind, ordnende Strukturen zu identifizieren.

„Aus einmal erkannten Mustern lassen sich konstruktive Regeln ableiten, mit denen auf Vergangenes rückgeschlossen und Künftiges vorhergesagt werden kann."

Intra- und Interklassenvariabilität, Signal und Rauschen – die grundsätzliche Frage ist, in welchem Verhältnis Chaos und Struktur zueinander stehen. Je regulärer und strukturierter ein Muster ist, desto weniger Trainingsbeispiele werden benötigt, um das Muster zu erlernen. Umgekehrt gilt: Je chaotischer die Welt ist, desto komplexer ist das Erkennungsproblem – und desto größer ist der erforderliche Trainingsdatensatz.

Die menschliche Wahrnehmung als Vorbild

Wir verfolgen in der Heidelberger Arbeitsgruppe „Computer Vision" einen Ansatz, der wesentliche Eigenschaften der menschlichen Wahrnehmung übernimmt. Das menschliche Auge etwa erfasst durch die Pupille – eine nur wenige Millimeter kleine Öffnung – unvorstellbar große Mengen an Daten. Sie entsprechen dem Volumen von mehr als einer halben DVD pro Sekunde. Die erfassten Informationen sind hochgradig redundant: Relevante Muster drohen in der Flut unterzugehen. Aus diesem Grund werden die Informationen schon früh im visuellen Cortex um einen Faktor von etwa 10.000 reduziert. Die Herausforderung besteht darin, die Flut unterschiedlicher Stimuli, die über unser Auge in das Gehirn einströmt, zu komplexen Objekten zu verbinden. Dieses sogenannte Bindungsproblem ist eines der zentralen Probleme der Mustererkennung – der Schlüssel, um es angehen zu können, heißt „Emergenz".

Was ist damit gemeint? Unter Emergenz versteht man die Fähigkeit eines Systems, durch das Zusammenspiel vieler Elemente neue Strukturen herauszubilden. Eine Farbe etwa lässt sich lokal wahrnehmen. Die Form eines Objekts hingegen können wir erst durch die Betrachtung des Ganzen erkennen – sie ist eine emergente Eigenschaft. Nehmen wir zum Beispiel einen Vogelschwarm, der am Himmel die Form eines Dreiecks bildet. Unsere Augen werden immer nur die einzelnen Vögel beobachten: Kein einzelnes Tier zeigt die Charakteristika eines Dreiecks.

HCI: Denkfabrik für die Bildverarbeitung

Das Heidelberg Collaboratory for Image Processing (HCI) gilt als "Denkfabrik" für die Bildverarbeitung und ist eines der größten Zentren seiner Art in Deutschland. Im Jahr 2008 wurde es innerhalb der Universität Heidelberg als "Industry on Campus"-Projekt eingerichtet; beteiligt sind neben der Robert Bosch GmbH die Sony Corporation, die Carl Zeiss AG, Heidelberg Engineering, Silicon Software und die PCO AG. Ziel der interdisziplinär ausgerichteten Forschungseinrichtung ist es, lang anstehende, schwierige Probleme der Bildverarbeitung zu lösen und sie anschließend mit den beteiligten Firmen sowie weiteren Kooperationspartnern in Applikationen zu überführen.

Das HCI besteht aus den vier Lehrstühlen für Bildverarbeitung der Universität sowie einer assoziierten Forschungsgruppe. Rund achtzig Mitarbeiter arbeiten an der Forschungseinrichtung – darunter zahlreiche Postdoktoranden, die über die Exzellenzinitiative und die beteiligten Industriepartner gemeinsam finanziert werden. Professor Dr. Björn Ommer forscht und lehrt seit dem Jahr 2009 am HCI und leitet dort die Arbeitsgruppe „Computer Vision".

hci.iwr.uni-heidelberg.de

Dennoch wird im Ensemble diese Struktur sichtbar – das Ganze ist also mehr als die Summe seiner Teile.

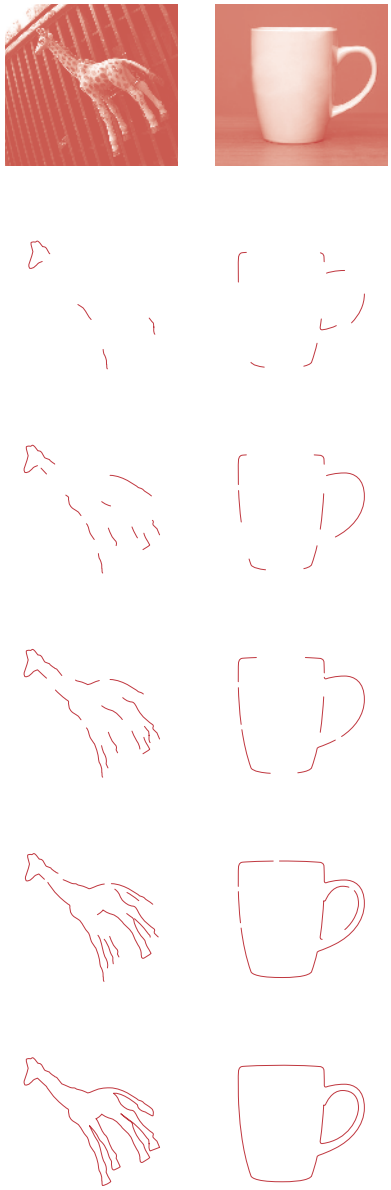
„Was dem menschlichen Gehirn selbstverständlich gelingt, ist für die künstliche Intelligenz eine große Herausforderung.“

Es existiert demnach eine große semantische Lücke zwischen dem, was die Sinneszellen des Auges lokal wahrnehmen, und dem Muster, das im Gehirn entsteht. Im Sinne des Psychologen Max Wertheimer gesprochen, eines Mitbegründers der Gestalttheorie: Wir stehen am Fenster und unsere Augen sehen nichts anderes als lokale Helligkeitsunterschiede oder Farben – dennoch erkennen wir schlussendlich Objekte, etwa Häuser, Bäume oder Menschen. Dieses Gruppieren zu Wahrnehmungsergebnissen, sogenannten Perzepten, ist essentiell, um Muster zu identifizieren: Wenig informative Entitäten werden zu einem größeren Ganzen, beispielsweise dem bekannten Objekt „Baum“, aggregiert. Die Redundanz verschiedener Objektkategorien (Autos, Fahrrädern und Motorrädern ist gemeinsam, dass sie alle Räder besitzen) kann zu dieser Aggregation ebenso genutzt werden wie Redundanzen innerhalb einer Kategorie (beispielsweise die Symmetrie).

Das Ziel: Computern das Sehen beibringen

Das Ziel unserer Forschungsarbeiten ist es, Computern das Sehen beizubringen. Dazu müssen Computer Muster erlernen, Objekte erkennen und Bilder interpretieren. Wir haben Algorithmen entwickelt, mit denen Rechner in die Lage versetzt werden, anhand eines kleinen Satzes von Trainingsbildern viele Punkte eines Bildes zu bedeutungstragenden Kompositionen – zu Perzepten – zu gruppieren. Darüber hinaus lernt der Computer, Charakteristika zu erkennen, die ihm dabei helfen, bestimmte Objektkategorien von anderen Objektkategorien zu unterscheiden – selbst dann, wenn die Kategorien in sich sehr variabel sind. Das Gruppieren und das Erkennen von Objekten sind unmittelbar miteinander verzahnt: Die bedeutungstragenden Bestandteile von Objekten können so aus dem Hintergrund gelöst und gruppiert werden – erst dadurch werden sie erkennbar.

Unser „kompositioneller Ansatz“ reduziert die Komplexität also in geeigneter Weise und macht es den Computern möglich, Modelle zu erlernen. Denn Kompositionalität ist für die Wahrnehmung des Menschen ebenso bedeutend wie für die Informationsverarbeitung durch Maschinen. Veranschaulichen lässt sich dies mit unserem Alphabet,



Muster verkraften Fehlstellen. Bei der Rekonstruktion von gelernten Mustern konzentrieren wir uns daher zuerst auf die relevantesten Aspekte, bevor wir sukzessive mehr Details ergänzen.

PROF. DR. BJÖRN OMMER leitet seit dem Jahr 2009 die Arbeitsgruppe „Computer Vision“, die am Heidelberg Collaboratory for Image Processing (HCI) und am Interdisziplinären Zentrum für Wissenschaftliches Rechnen der Universität Heidelberg angesiedelt ist. Sein Studium der Informatik und Physik schloss er 2003 an der Universität Bonn ab; 2007 wurde er von der Eidgenössischen Technischen Hochschule Zürich in Informatik promoviert. Anschließend forschte er zunächst in Zürich, später an der University of California, Berkeley. Der Schwerpunkt seiner Arbeit ist die Frage, wie Objekte und Handlungen in statischen und bewegten Bildern automatisch erkannt werden können. Seine Erkenntnisse bringt er in Heidelberg bei interdisziplinären Gemeinschaftsprojekten mit Kulturwissenschaftlern und Biomedizinern zur Anwendung.

Kontakt: ommer@uni-heidelberg.de

THE GRAMMAR OF PATTERNS

FROM CHAOS TO IMAGE

BJÖRN OMMER

The human eye can only differentiate between differences in brightness and various colours. It is our brain that turns these perceptions into objects such as a house, a tree or a person. The brain learns the patterns it needs to create these objects. Is it possible to teach computers the same ability to filter ordered structures out of a complex environment, i.e. to see like a person? The difficulty of pattern recognition and learning depends on the complexity of the observed structures and the presence of distracting clutter: recognition becomes more complicated the more variable the patterns of interest are and the more they resemble other, non-relevant patterns. One could therefore say that pattern recognition is the struggle to find order in chaos.

To tackle this problem effectively, it is crucial to efficiently utilise the regularity of our world. Inspired by human perception, we are following an approach that takes advantage of the redundancy of the visual stimulus to robustly learn visual patterns such as objects and their behaviour. Complex structures are typically emergent phenomena that cannot be detected locally. Consequently, there is a large semantic gap between local observations and the overall pattern: the whole is different from the sum of its parts.

We have therefore proposed a compositional approach that assembles individual percepts into meaningful compositions while learning the overall pattern. The resulting compositional dictionary can then be shared by various patterns. Atoms of the dictionary have generic applicability, much like the letters of our alphabet can form words and sentences to represent diverse content. Key to the representational power of this approach is the ability to learn relations between these generic constituents that serve as the compositional grammar of a pattern. ●

PROF. DR. BJÖRN OMMER has been heading the "Computer Vision" work group at the Heidelberg Collaboratory for Image Processing (HCI) and the Interdisciplinary Center for Scientific Computing (IWR) at Heidelberg University since 2009. He graduated from Bonn University in 2003 with a degree in computer science and physics and earned his PhD at the Swiss Federal Institute of Technology (ETH) Zurich in 2007. He went on to work as a researcher at ETH, then at the University of California, Berkeley. His work focuses on the question how objects and actions can be recognised automatically in static and moving images. Björn Ommer applies his findings in interdisciplinary projects with Heidelberg cultural scientists and biomedical researchers.

Contact: ommer@uni-heidelberg.de

"What comes naturally to the human brain is a monumental challenge for artificial intelligence: extrapolating a single abstract model from a multitude of patterns."

das nur aus 26 Buchstaben besteht. Dennoch können wir mit ihm aufgrund von Kompositionen – Wörtern und Sätzen – alles Erdenkliche ausdrücken. Der Kern unseres kompositionellen Ansatzes ist, dass die charakteristischen Beziehungen zwischen elementaren Bestandteilen erlernt werden. Es handelt sich dabei gewissermaßen um eine Grammatik, die vorgibt, wie aus der spezifischen Gruppierung einzelner Elemente Struktur entsteht. Das kompositionelle Gruppieren zerfällt dabei in zwei Teile: einerseits einen ausschließlich bildgetriebenen („bottom-up") Ansatz, bei dem allgemein geltende Gruppierungsregeln aus der Gestalttheorie angewendet werden, die visuelle Informationen zu einer einheitlichen und kohärenten Wahrnehmung zusammenführen; andererseits einen durch Lernen bestimmten Ansatz („top-down"), der Informationen gruppiert, die in anderen Bildern häufig in einer gewissen räumlichen Konstellation beobachtet worden sind. Kompositionen stellen damit einen gangbaren Mittelweg zwischen zwei Extremen dar: dem Repräsentieren des Objekts als Ganzes, das an der großen Variabilität und der Flut von Informationen scheitert, sowie der unabhängigen Beschreibung einzelner Objektbestandteile, bei der das große Ganze vernachlässigt wird.

Fassen wir zusammen: Das Erkennen von Struktur in einer vordergründig chaotischen Welt ist ein herausforderndes inverses Problem. Um dennoch komplexe Muster erlernen zu können, bedarf es eines Ansatzes, der die in unserer Welt nichtsdestotrotz vorhandene Regularität effektiv nutzt.
Die von uns untersuchte Kompositionalität ist ein wesentlicher Schritt auf dem Weg, die große semantische Lücke zwischen Wahrnehmung und Musterbildung zu schließen. Das Ziel, visuelle Informationen automatisch inhaltsbasiert zu erschließen – Maschinen also das Sehen beizubringen –, bleibt dennoch auch künftig eine große Herausforderung. ●