

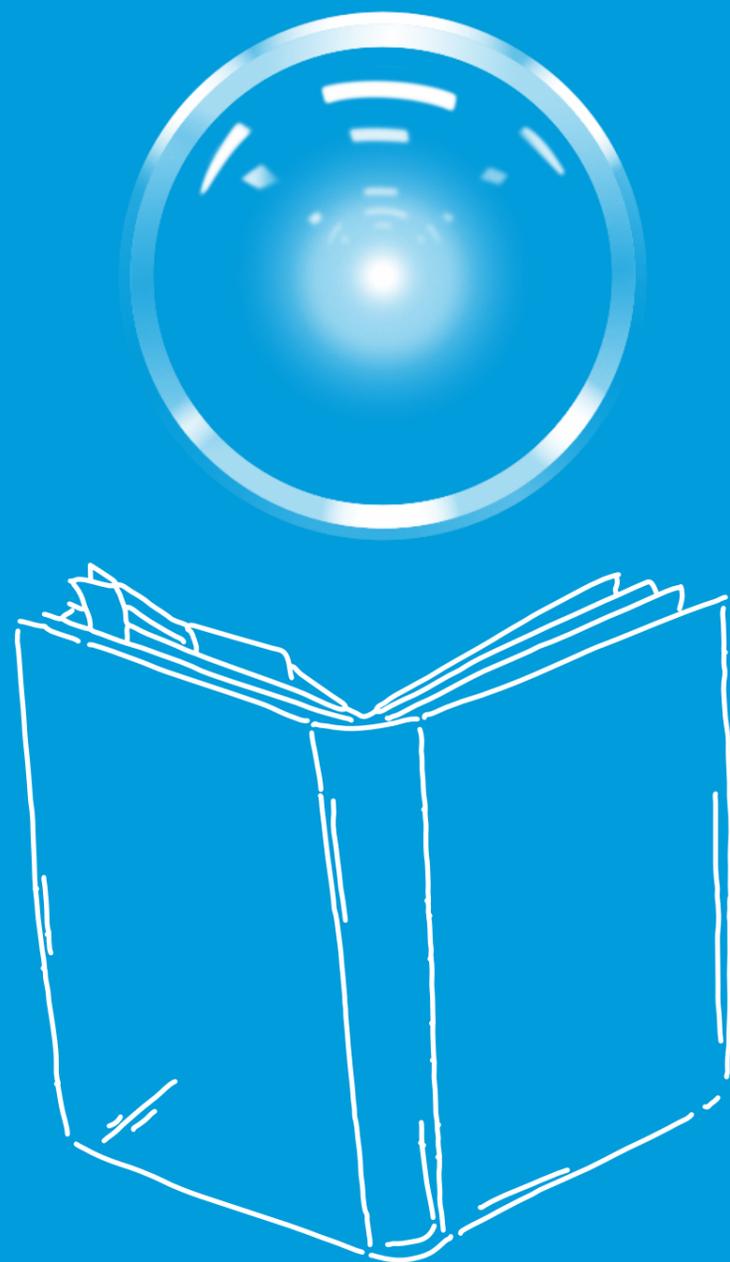
Wie lernen Maschinen?

Intelligente Systeme sind heutzutage allgegenwärtig. Doch wie sie die Welt wahrnehmen und was genau sie lernen, lässt sich oft nicht sagen. Der Konstanzer Informatiker Prof. Dr. Goldlücke nutzt Verfahren des maschinellen Lernens, insbesondere des Deep Learning, auf dem Gebiet der Computer Vision. uni'kon hat mit ihm über die Möglichkeiten und Grenzen des maschinellen Lernens gesprochen.

Wir alle kennen, lieben und fürchten sie vielleicht auch ein wenig: Aus dem kollektiven (pop-)kulturellen Gedächtnis der jüngeren Gegenwart sind fiktionale Supercomputer wie HAL 9000 („2001: Odyssee im Weltraum“, 1968), Skynet („Terminator“, 1984) oder die Matrix („Die Matrix“, 1999) nicht mehr wegzudenken. Diese körperlosen Superintelligenzen lassen das ursprüngliche Konzept der Künstlichen Intelligenz (KI) weit hinter sich, um ein eigenes Bewusstsein und eine ganz eigene Sichtweise auf die Welt, aus der sie hervorgegangen sind, zu entwickeln. In Film und Roman führt dies meist unweigerlich zur radikalen globalen Umsetzung KI-gestützter Allmachtsfantasien – mit tödlichen Folgen für die zum Abfallprodukt maschineller Vormacht degradierten menschlichen Schöpfer. Die düsteren, dem Cyberpunk entlehnten Dystopien des späten zwanzigsten Jahrhunderts sagen dabei mehr über den vorherrschenden Zeitgeist als über den jeweiligen Stand der Technik aus. Was sie vereint ist der Glaube daran,

dass Technologie die Welt verändern kann – zum Guten oder zum Schlechten.

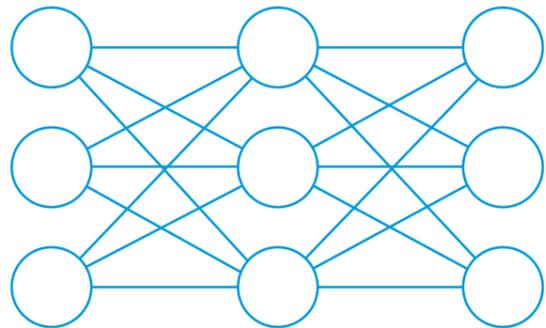
Und das tut sie. Algorithmischer Wertpapier- oder Hochfrequenzhandel, personalisierte Online-Werbung, medizinische Diagnosetools und Gesichtserkennungsdienste, wie sie etwa bei Facebook zum Einsatz kommen: Algorithmen prägen weite Teile unseres Lebens, auch wenn ihr Einsatz oftmals unbemerkt bleibt. Von der sogenannten „Singularität“, dem angenommenen Punkt, an dem die maschinelle Intelligenz die menschliche Intelligenz hinter sich lässt, sind wir allerdings noch weit entfernt, versichert der Konstanzer Informatiker Prof. Dr. Bastian Goldlücke, der intelligente Verfahren – sowohl Machine Learning als auch Deep Learning – bei der Erforschung von klassischen Problemen in der Computer Vision, dem maschinellen Sehen, anwendet: „Das Schlimmste, was im Moment passieren kann ist, dass ein autonomes Fahrerassistenzsystem versagt.“ Das kann zwar auch Menschenleben kosten, aber



dass eine Militärdrohne ein Bewusstsein entwickelt und die Weltherrschaft an sich reit, hlt er fr eher unwahrscheinlich: „Es gibt kein System, das seinen Code selbststndig verndern knnte, um pltzlich etwas vollkommen anderes zu tun als das, wofr es vorgesehen ist.“

Ein Paradigmenwechsel

Machine Learning (ML) ist ein Untergebiet der Knstlichen Intelligenz (KI), „das Computer in die Lage versetzen soll zu lernen, ohne explizit darauf programmiert zu sein“, so der US-amerikanische KI-Pionier Arthur L. Samuel, der bereits Ende der 1950er Jahre den noch heute gebruchlichen Begriff des „Machine Learning“ prgte. Machine Learning-Verfahren werden im weitesten Sinne eingesetzt, um knstlich Wissen aus Erfahrungen zu generieren. Dabei lernen herkommliche knstliche Systeme mithilfe von Algorithmen Muster, Gesetzmigkeiten oder Querverbindungen zwischen Daten zu identifizieren und diese zu verallgemeinern. Auch auf Grundlage des Erlernten Aussagen ber unbekanntes Daten zu machen gehrt mit zum Aufgabenspektrum. „Eine Riesenexplosion auf dem Gebiet des maschinellen Lernens – ein regelrechter Paradigmenwechsel – fand in den letzten Jahren aufgrund der Durchbrche im Deep Learning statt, besonders in der Bildverarbei-



tung und Computer Vision“, erzhlt Bastian Goldlcke. Angefangen hat alles mit dem 1958 von Frank Rosenblatt vorgestellten Perzeptron. Bei einem einschichtigen Perzeptron sind knstliche Eingabeneuronen mit einem oder mehreren Ausgabeknoten verknpft. Zwei Eingabeneuronen, die mit einem Ausgabeneuron verknpft sind, knnen beispielsweise ein „Und“, ein „Oder“ sowie ein „Nicht“ realisieren. Durch die Kopplung verschiedener Perzeptren lassen sich heutzutage eine Reihe vielschichtiger und hierarchischer Netzstrukturen herstellen, die beispielsweise auch auf die Erkennung von lokalen Mustern bei der Bildverarbeitung und Mustererkennung spezialisiert sind.

Des Pudels Kern

Im Gegensatz zu frheren Architekturen neuronaler Netzwerke knnen mit der wachsenden Leistungsfhigkeit der Computer heute zahlreiche Zwischenlagen (Hidden Layers) zwischen Eingabeschicht und Ausgabeschicht realisiert werden. Die Eingabe wird von der ersten sichtbaren Schicht weiter an die nchste, unsichtbare Schicht des Netzes weitergegeben und dort verarbeitet. Dabei wird auf jeder Schicht eine Reprsentation der eingegebenen Informationen erstellt, die eine Abstraktion der Eingangssignale darstellt. Dieser Prozess erstreckt sich ber alle Schichten des neuronalen Netzes hinweg und wird dabei zunehmend abstrakter: Werden zum Beispiel in der ersten Schicht zunchst einfache Merkmale eines Bildobjektes wie Linien oder Bgen identifiziert, kann das System in den tieferen Schichten anhand dieser Informationen komplexere Strukturen erkennen, die schlussendlich zur Identifizierung des Zielobjektes fhren. Knstliche Netze sind so in der Lage, abstrakte Konzepte aus Daten zu extrahieren und diese zur Lsung eines Problems zu verwenden.

Warum ist Deep Learning so erfolgreich? „Es kommen zwei Faktoren zusammen“, erklrt Goldlcke. „Zum einen haben wir jetzt die Rechenkapazitt, um in Echtzeit riesige neuronale Netze ausfhren und fr praktische Anwendungen nutzen zu knnen. Die Struktur gerade auch von Bilddaten ist so komplex, dass man die vielen Parameter und verschiedenen Schichten in den Netzwerken fr eine erfolgreiche Analyse tatschlich braucht. Zum anderen sind heute die unglaublich groen Datenbestnde verfgbar, die bentigt werden, um Modelle mit Millionen von Parametern trainieren zu knnen.“ Gerade weil sich mithilfe von Deep Learning der Aufbau hierarchi-



Von einem trainierten knstlichen neuronalen Netz zufllig erzeugte Gesichter.



scher Features besonders gut abbilden lässt, hat es auf den Gebieten der Bild- und Sprachverarbeitung andere Machine Learning-Verfahren weitestgehend verdrängt. Diese Entwicklung wird zusätzlich dadurch beflügelt, dass wichtige Deep Learning-Frameworks quelloffen zur Verfügung stehen, darunter zum Beispiel TensorFlow von Google.

Wie funktioniert Machine Learning?

Die Umsetzung von maschinellen Lernprozessen lässt sich auf verschiedene Arten bewerkstelligen. Dazu gehören das überwachte Lernen (Supervised Learning), das unüberwachte Lernen (Unsupervised Learning) und das bestärkende Lernen (Reinforcement Learning, siehe dazu auch Artikel S. 32). Beim überwachten Lernen werden dem Computer bestimmte Ein- und Ausgabewerte vorgegeben, auf Basis derer er lernt, Assoziationen herzustellen. Die Trainingsdaten sind dabei bereits markiert und klassifiziert. „Im Prinzip handelt es sich um eine Art Mustererkennungssystem“, erklärt Bastian Goldlücke. „Es geht um eine Abbildung von Eingabe auf Ausgabe. Man zeigt dem System ganz viele Beispiele von Eingabe zusammen mit gewünschter Ausgabe. Damit passt man ein Template für eine Funktion an die Daten an, das in aktuellen Netzwerken, die beispielsweise bei der Objekterkennung in Bildern eingesetzt werden, von mehreren Millionen Parametern abhängt.“

Beim unüberwachten Lernen werden dem System nur die Eingabewerte zur Verfügung gestellt. Es muss diese eigenständig kategorisieren und lernen, Beziehungen zwischen den Daten zu erkennen. „In der Computer Vision ist ein beliebtes Beispiel die Codierung von Gesichtern: Im Computer ist ein Gesicht einfach Pixel für Pixel gespeichert, als unstrukturierte Sammlung von farbigen Bildpunkten. Die menschliche Wahrnehmung gruppiert diese einzelnen Bildpunkte mühelos zu Gesichtern und deren Bestandteilen, wie verschiedene Arten von Augen, Nasen und Mündern. Das Ziel beim unüberwachten Lernen ist, dass der Algorithmus diese Grundmuster selbstständig in den Daten entdeckt“, erklärt Bastian Goldlücke (Zu Computer Vision siehe auch Artikel S. 22).

Dem menschlichen Lernvorgang vielleicht am nächsten kommt das bestärkende Lernen. Anders als beim überwachten Lernen werden hier keine markierten Eingabe- beziehungsweise Ausgabepaare vorgegeben. Man gibt dem System Trainingsdaten ein und weist es beispielsweise an, das Bild eines Hundes zu identifizieren. Gibt es ein richtiges Ergebnis aus,

erhält es ein positives Feedback dazu, auf dessen Grundlage es lernt, korrekte Assoziationen herzustellen. Diese Methode wurde unter anderem beim von Google DeepMind entwickelten System „AlphaGo“ eingesetzt, das 2015 den amtierenden Europameister im Go-Spiel schlug – was aufgrund der Komplexität und der großen Anzahl möglicher Züge bis dahin als besondere Herausforderung gegolten hatte. Das Nachfolger-System „AlphaZero“ aus dem Jahr 2017 ging einen entscheidenden Schritt weiter: Es war in der Lage, ohne auf menschliche Erfahrungswerte zurückzugreifen, innerhalb weniger Stunden nur anhand der Spielregeln sowie durch intensives Spielen gegen sich selbst gleich mehrere Brettspiele zu meistern, darunter Schach und die wesentlich komplexeren fernöstlichen Spiele Shōgi und Go.

Was sehen Maschinen?

Bastian Goldlücke wendet intelligente Verfahren auf dem Gebiet der Computer Vision an, das auch mit den deutschen Begriffen maschinelles Sehen oder Computersehen umschrieben wird. Im Kern beschäftigt sich die Computer Vision damit, wie sich die menschliche Wahrnehmung der realen Welt im Computer reproduzieren lässt – wie aus zweidimensionalen Kamerabildern die dreidimensionale Welt künstlich rekonstruiert werden kann. Dieser Aspekt der Computer Vision wird gerade im Alltag immer wichtiger: Selbstfahrende Autos lernen beispielsweise mithilfe von Kameras und Sensoren, die Umwelt zu erfassen, den Verkehr zu navigieren und dabei Kollisionen mit anderen Verkehrsteilnehmerinnen und -teilnehmern zu vermeiden. Auch bei der medizinischen Diagnostik kommt die Computer Vision zum Einsatz, ebenso bei industriellen Qualitätsmanagementprozessen, wo es beispielsweise gilt, Defekte an fertigen Produkten zu erkennen. Und sie findet neue Lösungen für die computerbasierte Erkennung von Gesichtern, Gesichtsausdrücken oder Gesten, die von besonderer Bedeutung für Mensch-Maschine-Interaktionen ist: Während ein Mensch gewohnheitsmäßig in der Lage ist, vertraute Gesichter zu erkennen und einem Gesicht eine bestimmte Gemütslage zuzuordnen, tut sich ein Computer besonders schwer damit, denn die zugrundeliegenden Regeln sind nicht offensichtlich und damit nur schwer berechen- und reproduzierbar.

Bastian Goldlücke trägt mit seiner Forschung zur Lösung klassischer Probleme bei der Tiefenrekonstruktion von Stereo-Bildpaaren und bei der Bildseg-

mentierung bei – beides Teilgebiete der digitalen Bildverarbeitung. Hier geht es darum zu bestimmen, welche Pixel auf einem Bild zusammengehören – also welche Bildpunkte Objekte im Bild formen und wie diese voneinander abzugrenzen sind. Unter anderem wird der Abstand jedes einzelnen Bildpunktes zum Betrachter ermittelt – eine datenintensive Aufgabe, bei deren Bewältigung maschinelles Lernen helfen kann: „Wir nutzen Algorithmen, um einen pixelgenauen, direkten Vergleich zwischen zwei Bildern eines Bildpaares vorzunehmen. Strukturen, die im linken Kamerabild sichtbar sind, müssen dabei im rechten Kamerabild wiedergefunden werden.“

Deep Learning-Verfahren werden unter anderem bei der Analyse von Lichtfeldern eingesetzt, die aus zwanzig oder sogar fünfzig von einer einzelnen Kamera aufgenommenen Bildern bestehen. Der Vorteil eines Deep Learning-Systems im Gegensatz zu einem herkömmlichen Algorithmus besteht darin, dass man mit genug Trainingsdaten dem neuronalen Netz auch die Analyse sehr komplexer Szenen beibringen kann, die physikalisch korrekt nur schwer zu modellieren sind. „So versuchen wir, die 3D-Geometrie sehr schwieriger Szenen zu bestimmen – darunter auch Szenen, die spiegelnde Oberflächen beinhalten“, sagt Goldlücke, denn in diesem Fall verändert sich die gespiegelte Oberfläche auf sehr komplexe Art und Weise, wenn man den Blickpunkt darauf ändert. Die grundlegende Frage ist also, wie man sowohl die Bewegung einer Kamera als auch die Struktur der beobachteten Welt in Relation zu einander berechnen kann. Die auf diesem Gebiet zum Einsatz kommenden künstlichen neuronalen Netze werden auch mit künstlichen Testdaten und -szenen trainiert, die aus der Computergrafik stammen. Während das Training eines solchen Deep Learning-Systems oft mehrere Tage in Anspruch nimmt, ist die Lösung des Problems dann in Zehntelsekunden berechnet.

Im Falle des an der Universität Konstanz angesiedelten Exzellenzclusters Centre for the Advanced Study of Collective Behaviour, für den Bastian Goldlücke ebenfalls KI-Lösungen entwickelt, kann Deep Learning die Untersuchung von großen Tierkollektiven erleichtern: Die Künstliche Intelligenz setzt dort an, wo das menschliche Auge bei der Identifizierung von Individuen versagt. Man denke an ein Bienenvolk oder an einen Vogelschwarm – unmöglich für das menschliche Auge, auf einem Bild mit bloßem Auge alle Tiere zu erfassen, geschweige denn ein Individuum vom anderen zu unterscheiden oder Aussagen da-

rüber zu machen, in welche Richtung einzelne Tiere gerade sehen oder sich bewegen.

Die Grenzen des Möglichen

Als es 1996 dem von der IBM entwickelten Computer „Deep Blue“ in einer aufsehenerregenden Schachpartie gelang, den amtierenden Schachweltmeister Garri Kasparow zu schlagen, ging ein kollektives Raunen um die Welt. Was würden Computer in Zukunft leisten können? War „Deep Blue“ intelligent – intelligenter vielleicht, als sein menschlicher Gegenspieler? Gab es Grenzen für das, was sich mit der richtigen Kombination aus Software und Hardware erreichen lassen würde? Dabei hatte der Erfolg von „Deep Blue“ zunächst einmal wenig mit Künstlicher Intelligenz zu tun. Es waren vielmehr neue Entwicklungen auf dem Gebiet des „Massively Parallel Processing“, die es dem System erlaubten, eine Unmenge von komplexen Kalkulationen gleichzeitig durchzuführen.

Dennoch wird das Können von Computern gern mit dem von Menschen verglichen. Und die Performance der besten Systeme im Bereich der Objekterkennung auf Bildern lässt sich tatsächlich ungefähr mit der von Menschen vergleichen, bestätigt Bastian Goldlücke.

Tatsache ist aber auch, dass sich künstliche Systeme leicht täuschen lassen. Ein wichtiges Stichwort ist hier das sogenannte „Adversarial Example“: Es reicht beispielsweise aus, einige wenige Pixel auf einem Bild zu verändern – also ein künstliches Rauschen herbeizuführen – um das System so aus dem Konzept zu bringen, dass es etwas völlig anderes erkennt, als eigentlich abgebildet ist. „Ein Fehler, den Menschen nie machen würden“, resümiert Bastian Goldlücke.

Ein aktuelles Beispiel für die verschiedenen Herausforderungen, mit denen Künstliche Intelligenzen in der Praxis fertig werden müssen, sind die Upload-Filter, die im Zuge der jüngst vom europäischen Parlament verabschiedeten Urheberrechtsreform eingesetzt werden sollen. Algorithmen stoßen an ihre Grenzen, wenn der Kontext eines Uploads einbezogen werden muss. Um eine Urheberrechtsverletzung feststellen zu können, müsste der Algorithmus in der Lage sein, zwischen erlaubter Nutzung – etwa im Kontext von Parodien oder Zitaten – und unerlaubter Nutzung zu unterscheiden. Der jetzige Stand der Technik lässt solch feine Unterscheidungen allerdings nicht zu, was im Resultat zu „Overblocking“ führen könnte.

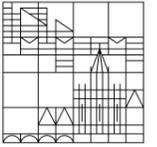


„Im Prinzip handelt es sich um eine Art Mustererkennungssystem. Es geht um eine Abbildung von Eingabe auf Ausgabe.“

Prof. Dr. Goldlücke



Prof. Dr. Bastian Goldlücke ist seit 2014 Professor für Computer Vision und Image Analysis an der Universität Konstanz. Zuvor forschte er an der Universität Heidelberg, der TU München sowie der Universität Bonn. Er beschäftigt sich mit der Kombination von Deep Learning mit Variationsverfahren und Anwendungen in der Computer Vision, insbesondere für inverse Probleme bei der 3D-Szenenanalyse. 2013 wurde Bastian Goldlücke mit dem ERC Starting Grant „Light Field Imaging and Analysis“ ausgezeichnet. Seine Forschung wird durch den SFB-TRR 161 „Quantitative Methods for Visual Computing“ und den Exzellenzcluster „Centre for the Advanced Study of Collective Behaviour“ der Universität Konstanz unterstützt.



Ist der Unterschied zwischen menschlichem und maschinellern Erkenntnisvermögen somit in der menschlichen Intuition begründet oder einem anderen entscheidenden Moment, das wir vielleicht noch nicht identifiziert haben? „Das ist die Frage“, lacht Goldlücke. „Es liegt aber auch nahe, dass wir einfach noch nicht das richtige künstliche Modell gefunden haben. Auf dem Gebiet der Computer Vision hat man zwar versucht, neuronale Netze nach dem Vorbild des visuellen Systems beim Menschen zu modellieren, aber offensichtlich funktionieren diese künstlichen Netze doch noch anders.“

Ghost in the Shell?

Die Frage liegt auf der Hand, inwieweit man künstlichen neuronalen Netzen auf dieser Grundlage vertrauen kann. Gerade in sicherheitsrelevanten Kontexten wie dem autonomen Fahren ist die Verlässlichkeit des Systems besonders wichtig. Ein großes Thema ist deshalb die sogenannte „Explainable AI“. Bislang lassen sich neuronale Netze am besten mit dem Bild der „Black Box“ umschreiben. Man kann sie zwar mit Testdaten trainieren, es fehlt allerdings „eine explizite deklarative Wissensrepräsentation“, wie die Gesellschaft

für Informatik (GI) feststellt. Dies gilt selbst dann, wenn die zugrundeliegenden mathematischen Prinzipien bekannt sind. Es lässt sich mit anderen Worten bislang nicht genau sagen, was künstliche neuronale Systeme eigentlich genau lernen beziehungsweise wie der Lernprozess auf den Hidden Layers abläuft.

Ein aktuelles übergreifendes Forschungsvorhaben besteht deshalb darin, Systeme zu entwickeln, bei denen der Entscheidungsprozess nachvollzogen werden kann. Laut GI ist einer der wichtigsten Beiträge von „Explainable AI“ deshalb „aufzuklären, was Ursache ist und was Wirkung (und welches nur Korrelation) – um zu vermeiden, dass man fälschlich Artefakte und Surrogate miteinbezieht“. Diese Öffnung der „Black Box“ wird auch im juristischen Kontext immer wichtiger, gerade vor dem Hintergrund der Europäischen Datenschutzgrundverordnung (DS-GVO), die unter anderem fordert, dass Erklärungen für bestimmte Entscheidungen oder Risikobewertungen nachvollziehbar und erklärbar sein müssen. Diese und weitere Sachverhalte gilt es nicht nur in der Wissenschaft, sondern auch in Wirtschaft und Gesellschaft zu verhandeln. | tg.



Migration



„Migration“ findet sowohl im Mensch- als auch Tierreich statt und kann aus unterschiedlichen Blickwinkeln betrachtet und analysiert werden. So fokussieren die Wissenschaftler*innen am Zukunftskolleg der Universität Konstanz auf andere Aspekte von „Migration“ als die beiden Konstanzer Exzellenzcluster „Politics of Inequality“ und „Collective Behaviour“ oder das „Zentrum für kulturwissenschaftliche Forschung“. Umso spannender, wenn alle Facetten im Rahmen einer gemeinsamen Veranstaltungsreihe beleuchtet und neue Forschungsfelder entdeckt werden. Zentraler Bestandteil der Kooperation sind wissenschaftliche Kurzvorträge und eine Ausstellung von George Butler. Der Künstler George Butler beschränkt Migration nicht auf die humane Migration, sondern weitet sie auf die Tiermigration aus und betrachtet die Interaktion beider Migrationsarten.

Lightning Talks & Discussion zu „Migration“
im Rahmen des Jour Fixe des Zukunftskollegs
25. Juni 2019

Ausstellung „Anima Mundi“
von Illustrator George Butler
im Foyer der Bibliothek,
Universität Konstanz
11. Juni bis 12. Juli 2019

