



„Uns hat es alle überrascht, wie gut der Mensch ist.“

Interview mit den Tübinger Forschern Felix Wichmann und Matthias Bethge zur Robustheit des Sehens und zur Zukunft der KI-Forschung in Deutschland

Maike Pfalz

Der Psychophysiker Felix Wichmann und der Physiker Matthias Bethge wollen den menschlichen Sehprozess genau verstehen, um künstlichen Intelligenzen robustes Sehen beizubringen. Bethge nutzt künstliche neuronale Netze, um das Sehen zu ergründen, Wichmann beschäftigt sich mit den psychologischen Prozessen, die der menschlichen visuellen Wahrnehmung zugrunde liegen.

Sie untersuchen die Robustheit des Sehens. Worum geht es dabei konkret?

Felix Wichmann: Robustes Sehen bedeutet, sich nicht davon beeinflussen zu lassen, dass sich die Randbedingungen radikal ändern. Das stört den Menschen praktisch nicht. Wir können Objekte erkennen, wenn sie halb verdeckt sind oder wir sie aus einem anderen Blickwinkel sehen beziehungsweise wenn es sehr hell oder dunkel wird. Dieses robuste Sehen möchten wir den Maschinen beibringen.

Worin besteht dabei das Problem?

Matthias Bethge: Wenn neue Daten oder Probleme den Trainingsdaten ähneln, sind Maschinen meistens besser als der Mensch. Aber mit Störungen können sie schlechter umgehen. In unserem Forschungsprojekt untersuchen wir, was die Robustheit des menschlichen Sehens ausmacht.

Wie untersuchen Sie das?

Wichmann: Unser Labor ist komplett schwarz, damit wir die Versuchspersonen gezielt visuellen Reizen aussetzen können, die wir genau unter Kontrolle haben. Ich weiß also exakt, welches Licht mit welcher Wellenlänge den Menschen ins Auge fällt. Darauf bauen wir unsere Modelle auf, um daraus Verhalten vorauszusagen.

Und was messen Sie dabei?

Wichmann: Wir zeigen zwei Bilder, die nahezu identisch sind, beispielsweise ein Bild mit lauter Linien und einer geschlossenen Kontur. Im zweiten Bild ist die Kontur offen. Die Probanden sollen die offene Kontur identifizieren. Bei der Psychophysik geht es darum, reproduzierbare

Messungen zu machen, die unabhängig von persönlichen Eigenschaften eines Menschen sind.

Aber ein Computer würde die offene Struktur auch finden?

Wichmann: Überraschenderweise ist der Mensch immer noch etwas besser bei dieser Aufgabe als eine Maschine, die auf solche Konturen trainiert ist.

Bethge: Wir untersuchen derzeit, wie unsere neuronalen Netze zu ihrer Entscheidung gelangen. Dazu schauen wir, welche kleinste Änderung im Input notwendig ist, um das Ergebnis zu ändern. Hier verhalten sich neuronale Netze komplett anders als Menschen. Eine minimale Störung kann dazu führen, dass ein neuronales Netz ein Auto als Hund wahrnimmt. Die Frage nach der offenen Kontur beantwortet das Netz daher völlig anders als der Mensch.

Solange der Computer auch die Kontur findet, ist das doch nicht schlimm, oder?

Bethge: Doch, denn letztlich setzt ein neuronales Netz die vielen kleinen Inputs, die es zu einer Entscheidung kombiniert, offenbar anders zusammen. In neuen Situationen kann das zu unvorhergesehenen und unerwünschten Entscheidungen führen. Daher müssen wir verstehen, wie der Computer zu seiner Entscheidung gelangt.

Wie gehen Sie dieses Problem an?

Bethge: Wir möchten die Maschinen effizienter und intelligenter beim Sehen machen, indem wir bessere Lernalgorithmen entwickeln. Dazu müssen wir die zugrundeliegenden Prozesse genau verstehen und dem Computer beibringen.

Wichmann: Dazu gehört auch ein vollständiges Verständnis über Licht, Schatten, Überlagerungen oder Lagebeziehungen zwischen den einzelnen Objekten auf den Bildern. Damit kann ich viele Vorhersagen machen, was sich hinter den Objekten versteckt, oder Schlussfolgerungen ziehen über Sachen, die ich nicht sehe. Der Mensch kann sehr gut Dinge anhand von Lichtreflexen orten, auch wenn er sie nicht direkt im Blickfeld hat.

Haben Sie ein Beispiel?

Bethge: Vor ein paar Jahren haben wir untersucht, wie gut der Wahrnehmungsapparat des Menschen ist. Dazu haben wir einen winzigen Ausschnitt von nur 4×4 Pixeln aus einem Bild herausgenommen, mittels statistischer Modelle nachzeichnen lassen und diesen Ausschnitt in



Felix Wichmann studierte experimentelle Psychologie an der Universität Oxford und arbeitete nach einer Postdoc-Zeit als wissenschaftlicher Mitarbeiter am Max-Planck-Institut für biologische Kybernetik in Tübingen. Von 2007 bis 2011 war er außerordentlicher Professor an der TU Berlin, seit 2011 ist er ordentlicher Professor Universität Tübingen.



Matthias Bethge studierte Physik und Mathematik an der Universität Göttingen und promovierte 2003 mit einer Arbeit über „Codes and Goals of Neuronal Representations“ an der Universität Bremen. Anschließend forschte er am Redwood Neuroscience Institut in Menlo Park, USA. Seit 2005 ist er am Max-Planck-Institut für biologische Kybernetik in Tübingen tätig.

das Originalbild gesetzt. Dem Menschen ist das aufgefallen!

Wichmann: Uns hat es alle überrascht, wie gut der Mensch ist! Denn da ging es nur um statistische Abhängigkeiten von Helligkeitswerten in Bildern und nicht etwa um die Bedeutung dieses Bildausschnitts! Wenn wir das verstehen und schnelle Algorithmen entwickeln, wird uns das in den nächsten Jahren weit nach vorne bringen. Ideal wäre aus meiner Sicht ein Ansatz, bei dem wir das vorhandene Expertenwissen mit dem maschinellen Lernen kombinieren.

Bethge: Beim maschinellen Lernen geht es immer um die Frage, welche Form von Vorwissen am besten geeignet ist, um aus den Trainingsbeispielen auf neue Situationen zu generalisieren. Dabei ist es auch wichtig, sich nicht zu sehr auf Expertenwissen festzulegen.

Was heißt das?

Bethge: Ein schönes Beispiel dafür ist AlphaZero – ein Computerprogramm, in dem keinerlei Expertenwissen steckt, sondern das lernt, indem es gegen sich selbst spielt. Schon nach drei Tagen war das Programm besser als der Vorgänger AlphaGo, der 2016 den bekanntesten Go-Spieler geschlagen hat. Das maschinelle Lernen revolutioniert die gesamte Künstliche Intelligenz.

Wichmann: Aber wir müssen die Algorithmen so bauen, dass sie diese kausalen Strukturen lernen können und damit in der Lage sind, einzelne Inputs korrekt zu kombinieren.

Bethge: Genau, und das passiert Schritt für Schritt. In der Physik machen wir uns immer Modelle von der Welt, beispielsweise simulieren wir Planeten und ihre Bewegung mit Billardkugeln. Dabei ignorieren wir viele Details. Einerseits ist Physik das Synonym für eine besonders präzise Beschreibung der Welt, andererseits approximieren wir de facto die ganze Zeit! Ich finde es spannend, dass wir bis ins fernste Weltall extrapolieren und Aussagen darüber treffen können, ohne dort jemals gewesen zu sein! Insofern stellt sich die Frage, wie es ein neuronales Netz schafft, für die vorhandenen Daten solche Modelle zu bauen.

Wenn wir in Ihr schwarzes Labor zurückgehen – wie helfen Ihre Experimente dabei, die Algorithmen zu verbessern?

Wichmann: Wir lernen dabei etwas über die Verarbeitungsschritte im Gehirn und können damit Modelle der Sehverarbeitung entwickeln und sukzessive verbessern. Ein solches Modell kann Vorhersagen generieren, was ein Proband bei einem bestimmten Experiment sehen müsste. Diese Vorhersagen überprüfen wir wiederum im Experiment.

Mit welchem Ergebnis?

Wichmann: Häufig können die Menschen Bilder unterscheiden, obwohl unser Modell vorhersagt, dass sie dazu nicht in der Lage sind. Also gehen wir einen Schritt zurück und modifizieren unser Modell. Das ist der konventionelle Weg, den wir früher gegangen sind.

Und heute?

Wichmann: Da setzen wir mehr auf Machine Learning. Beispielsweise gibt es Stimuli, zu denen unsere Modelle keine klaren Vorhersagen machen und bei denen wir mit unserer Intuition nicht weiterkommen. Unsere Idee besteht darin, tiefe neuronale Netzwerke zu trainieren, die ähnlich „komisch“ sehen wie der Mensch.

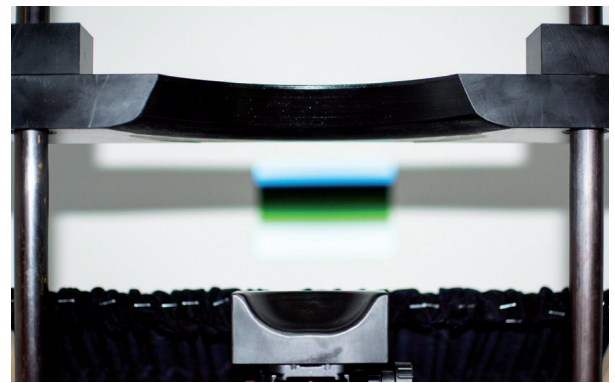
Was meinen Sie mit „komisch“?

Wichmann: Offenbar interpretiert der Mensch aufgrund seiner Erfahrung das Gesehene über die reine Wahrnehmung hinaus. Das gilt es zu verstehen und herauszufinden, welche Aspekte im Modell für diese „komische“ Wahrnehmung sorgen können. Das Machine Learning hilft dann, komplexe Hypothesen zu generieren.

Bethge: Letztlich geht es darum, zu erfassen, wie alle möglichen Lichtmuster auf Wahrnehmungen abgebildet werden. Die Menge der möglichen Abbildungen wächst exponentiell mit der Anzahl der Dimensionen.

Und damit lässt sich diese „komische“ Wahrnehmung nachbauen?

Wichmann: Genau. Ich bin überzeugt, dass es uns die



Blick durch Kinn- und Kopfstütze sowie Monitorbegrenzung auf ein Farbmuster

Methoden des Deep Learning in den nächsten Jahren erlauben werden, solche Sehsysteme zu bauen. Damit möchten wir dann komplexere und unbekanntere Situationen voraussagen. Vielleicht hilft gerade die Nichtlinearität dem Modell, um robuster gegen Veränderungen des Inputs zu werden?

Bethge: Ein Problem besteht auch darin, dass neuronale Netze nicht gut generalisieren, wenn sie mit etwas Neuem konfrontiert sind, beispielsweise mit Verpixelung. Wenn ich das Netz mit einer Art Rauschen trainiere, aber dann eine andere Art Rauschen teste, die für den Menschen kaum einen Unterschied macht, ist das Netz damit überfordert. Um dem menschlichen Gehirn nahe zu kommen, braucht man Generalisierungen und das Netz muss die Kompositionseigenschaften eines Bildes verstehen.

An welchen Stellen lernen Sie beide konkret voneinander?

Wichmann: Kein Psychophysiker ist gleichzeitig Experte für Machine Learning. Ein Treffen mit Matthias liefert mir daher neue Ideen oder Modifikationen für Experimente. Im Gegenzug weiß Matthias direkt, was er für seine Algorithmen aus unseren Experimenten lernen kann.

Bethge: Wir wollen beide verstehen, wie das menschliche Sehen funktioniert, das verbindet uns.

Für welche Anwendungen könnte das robuste Sehen interessant sein?

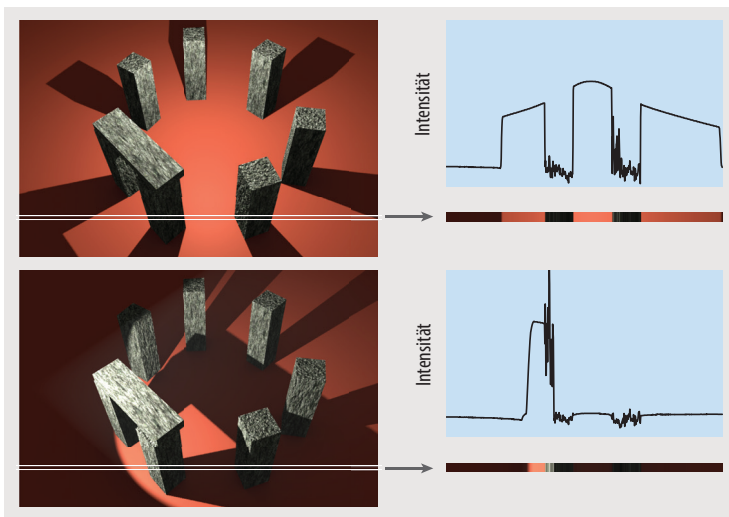
Wichmann: Für autonomes Fahren. Derzeit wird dabei mit Laserranging ein enormer technologischer Aufwand getrieben, um beispielsweise Abstände zu messen. Das Ergebnis ist dennoch unbefriedigend.

Arbeiten Sie mit der Industrie zusammen?

Wichmann: Ich bin an zwei Projekten beteiligt, bei dem einen geht es um Laminatdruck: Wenn riesige Tintenstrahldrucker die Holzstrukturen drucken, ist immer mal wieder eine Düse verstopft. Die Frage ist nun: Wie viele Düsen dürfen verstopft sein, bevor der Mensch das bemerkt und das Laminat nicht mehr kauft? Unser Modell konnte helfen, diese Frage etwas besser zu beantworten. Eine andere Kooperation haben wir mit Zeiss an diskutiert.

Worum geht es dabei?

Wichmann: Einige Menschen kommen nicht mit Gleitsichtgläsern zurecht, sodass viele Brillengläser entsorgt werden müssen. Ziel ist es daher, psychophysisch vorherzusagen, ob jemand ein bestimmtes Brillenglas nutzen



Bilder der gleichen Szene unter verschiedenen Beleuchtungsbedingungen (links) führen zu völlig unterschiedlichen visuellen Eindrücken. Die rechte Seite zeigt die Intensitäten, die von derselben Pixelreihe in den zwei verschiedenen Bildern aufgenommen wurden. Trotz der unterschiedlichen Beleuchtung erkennen Menschen leicht, dass es sich um die gleiche Szene handelt, Algorithmen haben damit Schwierigkeiten. Daher gilt es, die Gemeinsamkeiten und Unterschiede zwischen natürlichen und künstlichen Sehsystemen zu untersuchen, um der robusten visuellen Verarbeitung auf den Grund zu gehen.

kann. Da geht es beispielsweise um das Tolerieren von Verzerrungen. Das Projekt ist aber noch in der Planungsphase.

Bethge: Wir beide sind neugiergetriebene Forscher, die Probleme lösen und damit Anwendung ermöglichen wollen. Beim Machine Learning geht dieser Weg sehr schnell, wie man an DeepArt sieht.¹⁾ Dabei haben wir einem neuronalen Netz beigebracht, in Bildern Inhalt und Stil zu trennen und neu zusammenzufügen. Dafür haben wir einen Textursyntheselgorithmus entwickelt, um die Texturfeatures zu extrahieren und zu schauen, wo die abstrakte Inhaltsinformation ist. Und das haben wir kombiniert.

Wofür wenden Sie das an?

Bethge: Das war zunächst grundlagengetrieben. Aber mir war klar, dass die Leute es sehr cool finden würden, diesen Algorithmus auf Kunstwerke anzuwenden und damit eigene Fotos in einem bekannten Stil eines Gemäldes, beispielsweise von Kandinsky, van Gogh oder Picasso, zu verfremden. Nature und Science haben das Paper dazu nicht mal zur Begutachtung rauschicken wollen, also haben wir es auf arXiv hochgeladen. Das ist über Nacht eingeschlagen! Daraus ist ein eigenes Start-up geworden. Generell braucht die Wissenschaft aber mehr Unterstützung, damit neue Ergebnisse aus der Grundlagenforschung leichter in die Anwendung gebracht werden können.

Wie steht Europa in diesem Bereich eigentlich da?

Bethge: Wenn man ehrlich ist, hinken wir den USA viele Jahre hinterher. Wir haben kein Google, Facebook oder Amazon, denn in Europa setzen die Firmen viel stärker auf ihre klassischen Stärken. Das sieht man an den Automobilherstellern, die lange nur bewährte Konzepte op-

timiert haben, weil es keinen Druck gab, ein Elektroauto oder ein autonomes Fahrzeug zu bauen. Europa ist nicht mehr der Ort der größten Innovationen.

Und nun?

Bethge: Wir müssen uns Gedanken machen, wie das in Zukunft besser geht. In Toronto gibt es das Vector Institute, wo hunderte Millionen Dollar aus öffentlicher Hand und der Industrie Geld investiert wurden, damit der Standort möglichst innovativ bleibt und die besten Talente anlockt. Die beste Chance Europas liegt meiner Ansicht nach darin, etwas Ähnliches aufzuziehen, wie es die ELLIS-Initiative versucht, die dabei ist, ein europäisches Labor für Learning and Intelligent Systems aufzubauen. In Tübingen haben wir beispielsweise das Cyber Valley, wo zahlreiche Partner aus Wissenschaft und Industrie zusammenarbeiten.

Wo liegen für Deutschland die Chancen?

Bethge: Wir brauchen ein Zentrum, das Top-Talente anzieht und wo die jungen Leute das Gefühl haben, dass dort die Zukunft erfunden wird und sie sich entwickeln können. Mit politischem Willen wäre das möglich.

Wichmann: Wo wir aus meiner Sicht punkten können: Die großen Internetfirmen sind keine guten Vorbilder beim Umgang mit Daten und Menschen. Daher sehe ich ebenfalls eine Riesenchance, ein öffentlich finanziertes Institut aufzubauen, das kompetitiv ist und wo die jungen Leute mit gutem Gewissen hingehen können. Viele möchten gar nicht für die großen Internetfirmen arbeiten, aber derzeit sind die Bedingungen dort zu verlockend.

Bethge: Wichtig wäre mir darüber hinaus die Forschungsfreiheit. Der Horizont einer öffentlichen Forschungseinrichtung ist immer größer, als es das Businessmodell einer Firma hergibt. Somit hat man ganz andere Möglichkeiten. Forschung ließe sich viel attraktiver gestalten, wenn es mehr um Probleme aus dem realen Leben ginge. Denn gerade die Künstliche Intelligenz zielt darauf ab, reale Probleme zu lösen, die wir bislang noch nicht lösen konnten.

1) Auf deepart.io lassen sich eigene Fotos nach verschiedenen markanten Stilen verfremden. Nach wenigen Minuten steht das verfremdete Foto zur Verfügung.