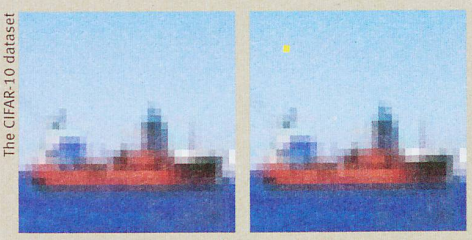
Ein einziges Pixel täuscht künstliche Intelligenz



The CIFAR-10 dataset

Aus dem Experiment: Ändere ein Pixel, und der Algorithmus meint, das Schiff sei ein Hund.

Ein einziges Pixel täuscht künstliche Intelligenz

Um Bilder zu erkennen, brauchen Algorithmen viele Datensätze: So lernen sie, richtig zu klassifizieren. Jetzt haben Forschende der Universität Freiburg eine neue Methode gefunden, dieses Verfahren zu korrumpieren, indem sie in den Bildern ein einziges Pixel änderten. Konkret setzten die Forschenden den Blauwert eines zufällig ausgewählten Pixels auf Null. Je nach Umgebungsfarbe kann es dadurch fast unsichtbar werden.

Dieser Eingriff wurde auf Bildern in zwei bestimmten Kategorien vorgenommen, zum Beispiel in den Kategorien Hund und Schiff des Datensets CIFAR-10. Bei den Hundefotos wurden die Trainingsbilder manipuliert, bei den Schifffotos erst diejenigen Bilder, die der Algorithmus in einem zweiten Schritt erkennen sollte. Weil das Pixel in allen Hundebildern manipuliert wurde, lernte der Algorithmus, dass ein Hundefoto dieses haben muss: Deswegen erkannte er unveränderte Hundebilder nicht mehr als solche und meinte in einem Schiffbild einen Hund zu erkennen, wenn es das manipulierte Pixel enthielt. Dieser Doppel-Angriff wurde bei sechs neuronalen Netzen getestet. Mit Erfolg: Fünf Algorithmen klassifizierten mehr als 70 Prozent der Schiffe als Hund, hingegen weniger als ein Prozent der Hunde korrekterweise als Hund.

«Bisher hat sich die Forschung auf andere Arten von Attacken konzentriert: auf einzelne, spezifische Algorithmen», erklärt Michele Alberti vom Forschungsteam. «Aber dafür muss man auf das neuronale Netz zugreifen können. Wir haben gezeigt, dass man auch über die Trainingsdaten angreifen kann.»

Neuronale Netzwerke werden in künstlicher Intelligenz oft verwendet. Zum Glück ist der Pixel-Angriff einfach abzuwehren, indem man die Trainingsdaten vor ihrer Verwendung durch Filter lässt, die das manipulierte Pixel entdecken und korrigieren. «Wir wollen zeigen, dass solche Angriffe möglich sind. Öffentliche Datensätze aus dem Internet sind gratis. Sie ungeprüft zu verwenden kann kritisch sein.» *Anna Julia Schlegel*

M. Alberti et al.: Are You Tampering With My Data? European Conference on Computer Vision (2018)

Wie unser Klärschlamm brennt

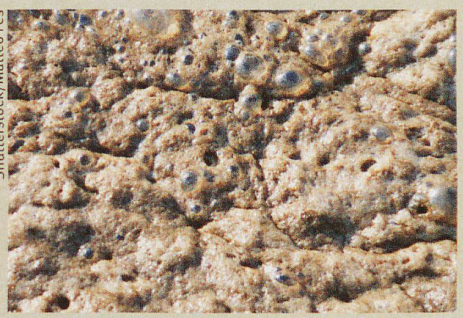
Jährlich entstehen in der Schweiz rund 200 000 Tonnen Klärschlamm, Tendenz steigend. Seit 2006 darf dieser Abfall gemäss Bundesverordnung nicht mehr als Düngemittel verwendet werden. Deswegen wird er heute primär verbrannt, wozu er zuerst aufbereitet werden muss. Im ersten Schritt wird ihm dabei Methan entzogen, das zur Energiegewinnung genutzt wird, im zweiten wird er getrocknet.

Bislang wusste man wenig darüber, welche Prozesse bei der Verbrennung ablaufen und wie sich deren Kinetik beschreiben lässt. Dies ändert sich mit der Studie von Jonas Wielinski, Doktorand in der Gruppe von Ralf Kaegi von der Eawag. Die Forschenden haben herausgefunden, dass sich die Verbrennung durch zehn chemische Reaktionen beschreiben lässt.

Dafür haben sie ein Gerät benutzt, das im Prinzip funktioniert wie eine sehr präzise Waage in einem Ofen. Das Ganze nennt sich thermogravimetrische Analyse. Dabei wurden die Schlammproben beliebigen Temperaturen und unterschiedlichen Atmosphären ausgesetzt. Die Forschenden wendeten zudem einen Algorithmus an, um die parallel ablaufenden Verbrennungsreaktionen zu bestimmen. Zusätzlich präsentierten sie eine Methode, mit der sich geeignete Referenzverbindungen bestimmen lassen.

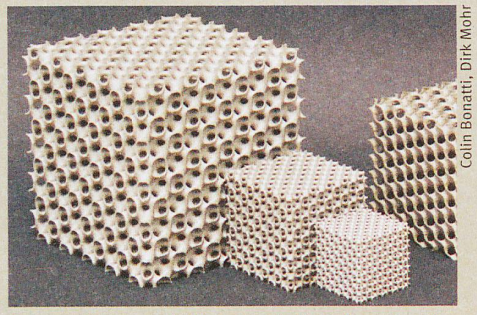
Aus den beobachteten Reaktionen lässt sich schliessen, dass im Klärschlamm vor allem Zellulose und Lignin verbrennen. Sie machen zusammen 55 Prozent der bei der Verbrennung verlorenen Masse aus. Die Zellulose stammt in erster Linie von Toilettenpapier, das einer der Hauptbestandteile der organischen Abfälle in unserem Abwasser ist. Ausserdem wurden geringere Anteile von Hemizellulose, Xylan, Alginat und Calcit identifiziert. *Anne Careen Stoltze*

J. Wielinski et al.: Combustion of Sewage Sludge: Kinetics and Speciation of the Combustible. Energy & Fuels (2018)



Shutterstock/Matteo Fes

Welche Prozesse bei der Verbrennung von Klärschlamm ablaufen, wurde neu genau erfasst.



Colin Bonatti, Dirk Mohr

Die repetitiven Hohlräume verleihen dem Metall spezielle Eigenschaften.

Frisch aus dem 3D-Druck: Antischock-Metall

Das Drucken in 3D entwickelt sich rasant weiter, auch bei Metallen: Ein Laser schmilzt Stahlpulver, und die Flüssigkeit wird wie in konventionellen Verfahren abgelagert. Auf diese Weise hat Dirk Mohr, Forscher an der ETH Zürich, ein Metallgitter entwickelt, das dank repetitiv angeordneten Hohl- und Vollräumen optimierte Eigenschaften für die Dämpfung von Stössen besitzt.

Dirk Mohr hatte sich bereits vor fünfzehn Jahren während seines Studiums mit diesem Thema beschäftigt. Doch damals existierten solche dreidimensionalen Strukturen nur auf Papier: «Das waren reine Gedankenspiele, die sich praktisch nicht umsetzen liessen. Ich habe eher eine Ingenieur-Mentalität und verfolgte diese Arbeiten deshalb nicht weiter. Doch aufgrund der Entwicklung der additiven Fertigung konnte ich sie wieder aus der Schublade holen.»

Das neue Material erinnert an Metallschaum: eine Stahlmasse, die viel Luft in kleinen Kammern enthält. Bei Schaum ist die Struktur aber ziemlich willkürlich, weil sich beim Einblasen von Gas in die Metallschmelze zufällig Blasen bilden. Im Gegensatz dazu kann beim 3D-Druck die Struktur präzise gesteuert werden – und so auch die Eigenschaften des Materials.

Dirk Mohr und sein Doktorand Colin Bonatti haben ein isotropes Material entwickelt: Die Poren des Metalls folgen einem schalenartigen Design – eine geschwungene, komplexe Struktur, die auf dem Computer entwickelt und so optimiert wird, dass sie Schläge verteilen und die Verformungen beschränken kann. «Dieser Ansatz eignet sich zur Konzeption massgeschneiderter Komponenten, wie ultraleichte Absorber von mechanischer Energie oder biomedizinische Implantate», erklärt Dirk Mohr, «für eine industrielle Produktion wie in der Automobilindustrie müssen aber zuerst die Kosten für die additive Metallfertigung sinken». *Lionel Pousaz*

C. Bonatti and D. Mohr: Mechanical Performance of Additively-Manufactured Anisotropic and Isotropic Smooth Shell-Lattice Materials: Simulations & Experiments. Journal of the Mechanics and Physics of Solids (2018)