On the barriers of AI and the trade-off between stability and accuracy in deep learning

Vegard Antun (Oslo, vegarant@math.uio.no)
Matthew J. Colbrook (Cambridge, m.colbrook@damtp.cam.ac.uk)

Joint work with:

Ben Adcock (SFU), Nina Gottschling (Cambridge), Anders Hansen (Cambridge),
Clarice Poon (Bath), Francesco Renna (Porto)

Geilo Winter School, January 2021

MAIN GOAL

Determine the barriers of computations in deep learning
(i.e. what is and what is not possible)

$$\Downarrow$$

Stability and Accuracy in AI

> *Wir müssen wissen - wir werden wissen!*
>
> — David Hilbert

Some of the results found in (links included):

- Antun, Colbrook and Hansen, 2021. *Can stable and accurate neural networks be computed? - On the barriers of deep learning and Smale's 18th problem.* arXiv:2101.08286.
- Gottschling, Antun, Adcock, and Hansen, 2020. *The troublesome kernel: why deep learning for inverse problems is typically unstable.* arXiv:2001.01258.
- Antun, Renna, Poon, Adcock, and Hansen, 2020. *On instabilities of deep learning in image reconstruction and the potential costs of AI.* PNAS.

`https://github.com/Comp-Foundations-and-Barriers-of-AI`

Outline of lectures

DAY I	**DAY II**	**Day III**
Gravity of AI Image Classification Need for Foundations AI for Image Reconstruction	Inverse Problems Instabilities & Kernel Awareness Intriguing Barriers Algorithm Unrolling	Achieving Kernel Awareness FIRENETs Imaging Applications Numerical Examples

Slides will be hosted at `http://www.damtp.cam.ac.uk/user/mjc249/Talks.html`.

Useful references for further reading in grey boxes.

Comments and suggestions welcome! (vegarant@math.uio.no, m.colbrook@damtp.cam.ac.uk)

Interest in deep learning unprecedented and exponentially growing

Google search (7th Jan) “deep learning” or “machine learning” yields ≈2.5 billion hits
Contrast with “computational mathematics” which has <150 million hits

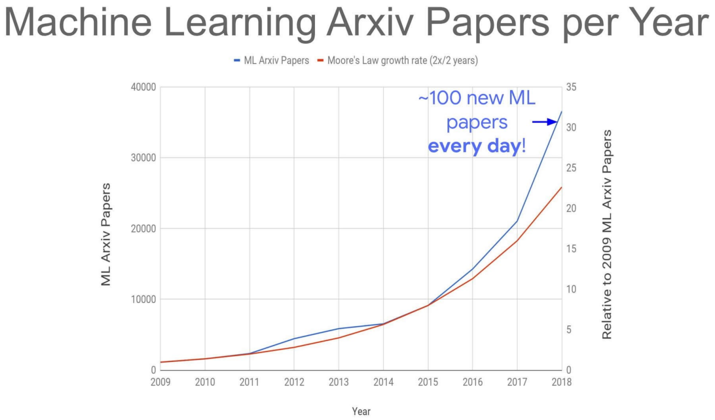


Figure: Source: ‘Deep Learning to Solve Challenging Problems’ (Google AI)

To keep up last year, you would need to continually read a paper every < 5 mins!

Why is AI suddenly such a big deal?

Deep Learning is changing our lives

- AI techniques are replacing humans in problem solving.
- AI techniques are replacing established algorithms in science.

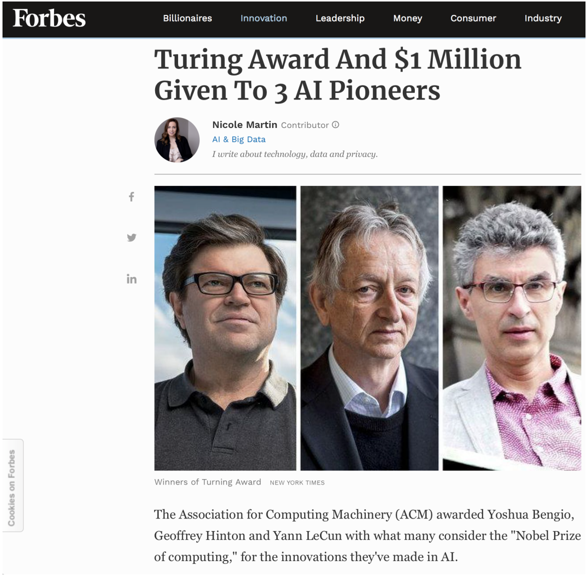
AI replacing humans

- Self-driving vehicles
- Automated diagnosis in medicine
- Automated decision processes
- Automated weapon systems
- Music composition
- Call centres
- Any security system based on face or voice recognition
- Amazon’s Alexa smart speaker
- Mathematical proofs

AI replacing (and/or enhancing) established algorithms

- Medical imaging (MRI, CT, etc)
- Microscopy
- Imaging problems in general
- Radar, sonar, etc.
- Methods for solving PDEs

The Pioneers



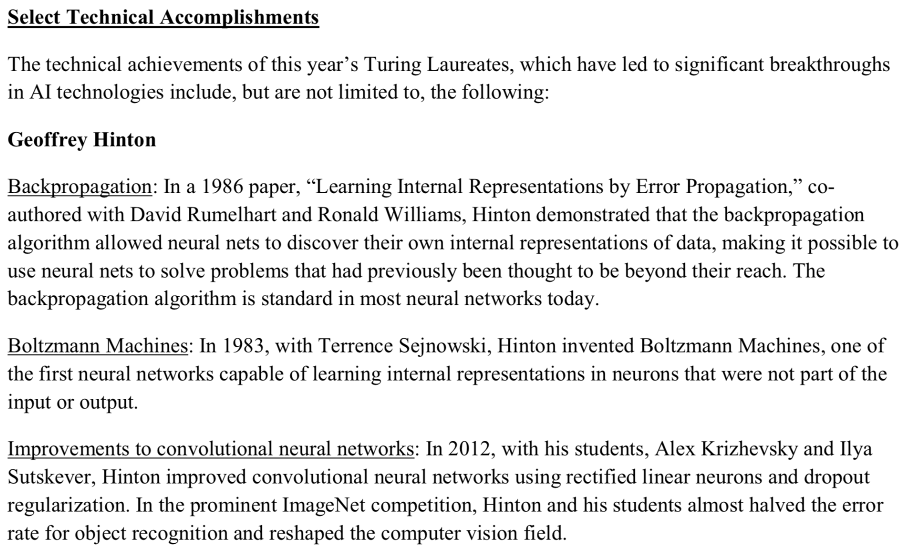
Turing Award And $1 Million Given To 3 AI Pioneers

Nicole Martin Contributor
AI & Big Data
I write about technology, data and privacy.

Winners of Turning Award NEW YORK TIMES

The Association for Computing Machinery (ACM) awarded Yoshua Bengio, Geoffrey Hinton and Yann LeCun with what many consider the "Nobel Prize of computing," for the innovations they've made in AI.

Citation from the Turing Award jury



Select Technical Accomplishments

The technical achievements of this year's Turing Laureates, which have led to significant breakthroughs in AI technologies include, but are not limited to, the following:

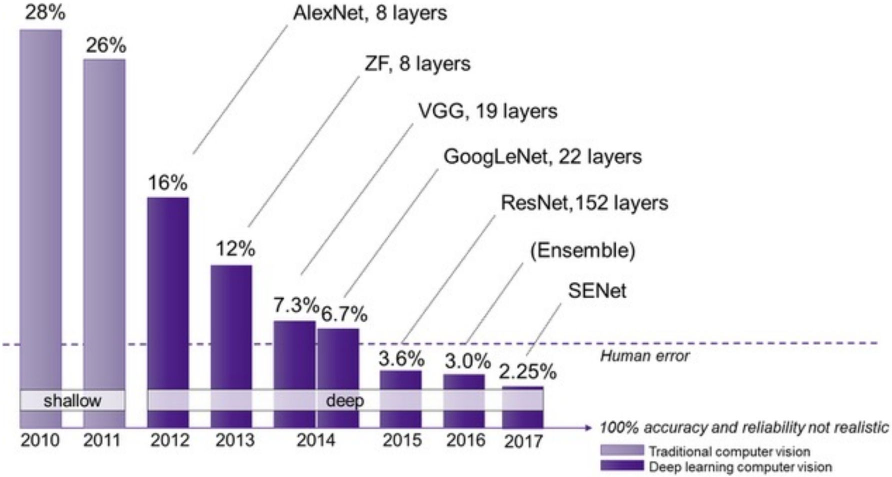
Geoffrey Hinton

Backpropagation: In a 1986 paper, "Learning Internal Representations by Error Propagation," co-authored with David Rumelhart and Ronald Williams, Hinton demonstrated that the backpropagation algorithm allowed neural nets to discover their own internal representations of data, making it possible to use neural nets to solve problems that had previously been thought to be beyond their reach. The backpropagation algorithm is standard in most neural networks today.

Boltzmann Machines: In 1983, with Terrence Sejnowski, Hinton invented Boltzmann Machines, one of the first neural networks capable of learning internal representations in neurons that were not part of the input or output.

Improvements to convolutional neural networks: In 2012, with his students, Alex Krizhevsky and Ilya Sutskever, Hinton improved convolutional neural networks using rectified linear neurons and dropout regularization. In the prominent ImageNet competition, Hinton and his students almost halved the error rate for object recognition and reshaped the computer vision field.

Before and after 2012 - The ImageNet competition

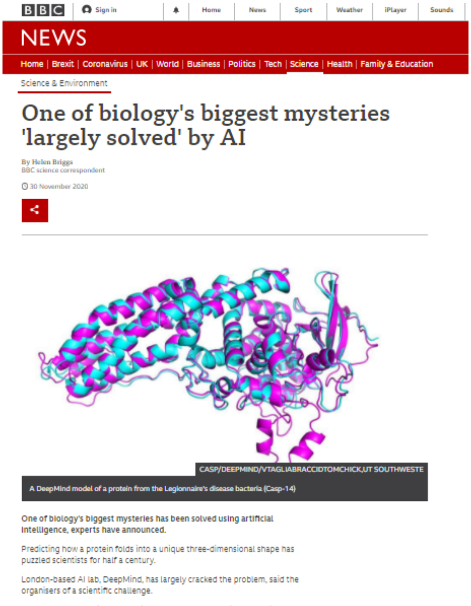


A paradigm shift

<table>
<tr><th colspan="3">Top 5 ILSVRC 2012 Results</th></tr>
<tr><td>1st</td><td>Error: 16.4%</td><td>Deep Learning</td></tr>
<tr><td>2nd</td><td>Error: 26.1%</td><td>Other approach</td></tr>
<tr><td>3rd</td><td>Error: 26.9%</td><td>Other approach</td></tr>
<tr><td>4th</td><td>Error: 29.5%</td><td>Other approach</td></tr>
<tr><td>5th</td><td>Error: 34.4%</td><td>Other approach</td></tr>
<tr><th colspan="3">Top 5 ILSVRC 2017 Results</th></tr>
<tr><td>1st</td><td>Error: 2.3%</td><td>Deep Learning</td></tr>
<tr><td>2nd</td><td>Error: 2.5%</td><td>Deep Learning</td></tr>
<tr><td>3rd</td><td>Error: 2.7%</td><td>Deep Learning</td></tr>
<tr><td>4th</td><td>Error: 3.0%</td><td>Deep Learning</td></tr>
<tr><td>5th</td><td>Error: 3.2%</td><td>Deep Learning</td></tr>
</table>

Table: Results from ImageNet Large Scale Visual Recognition Competition (ILSVRC).

Protein folding: open problem since Anfinsen's 1972 Nobel Prize


BBC
Sign in
Home
News
Sport
Weather
iPlayer
Sounds
NEWS
Home | Brexit | Coronavirus | UK | World | Business | Politics | Tech | Science | Health | Family & Education
Science & Environment

One of biology's biggest mysteries 'largely solved' by AI

By Helen Briggs
BBC science correspondent

30 November 2020

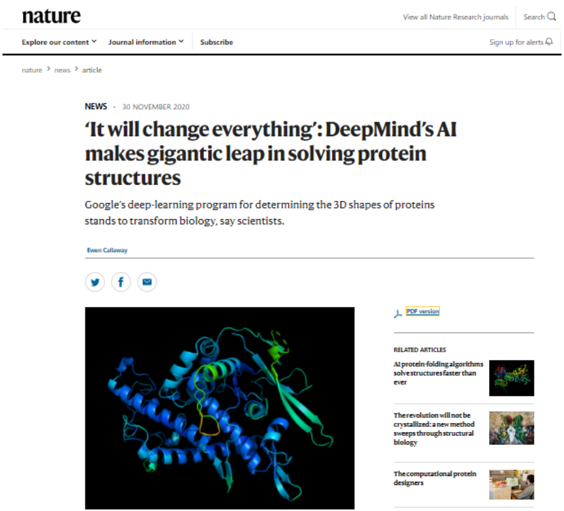
CASP/DEEPMIND/VTAGLIABRACCIDTOMCHICK,UT SOUTHWESTE

A DeepMind model of a protein from the Legionnaire's disease bacteria (Casp-14)

One of biology's biggest mysteries has been solved using artificial intelligence, experts have announced.

Predicting how a protein folds into a unique three-dimensional shape has puzzled scientists for half a century.

London-based AI lab, DeepMind, has largely cracked the problem, said the organisers of a scientific challenge.



NEWS · 30 NOVEMBER 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway

RELATED ARTICLES

AI protein-folding algorithms solve structures faster than ever

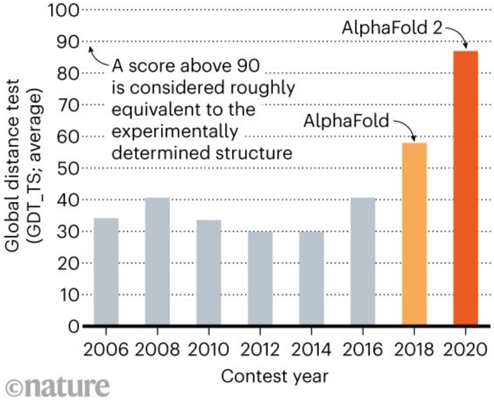
The revolution will not be crystallized: a new method sweeps through structural biology

The computational protein designers

Protein folding: open problem since Anfinsen's 1972 Nobel Prize

"This computational work represents a stunning advance on the protein-folding problem, a 50-year-old grand challenge in biology. It has occurred decades before many people in the field would have predicted. It will be exciting to see the many ways in which it will fundamentally change biological research."

— Venki Ramakrishnan
(2009 Nobel Prize in Chemistry, President
of the Royal Society 2015–2020)



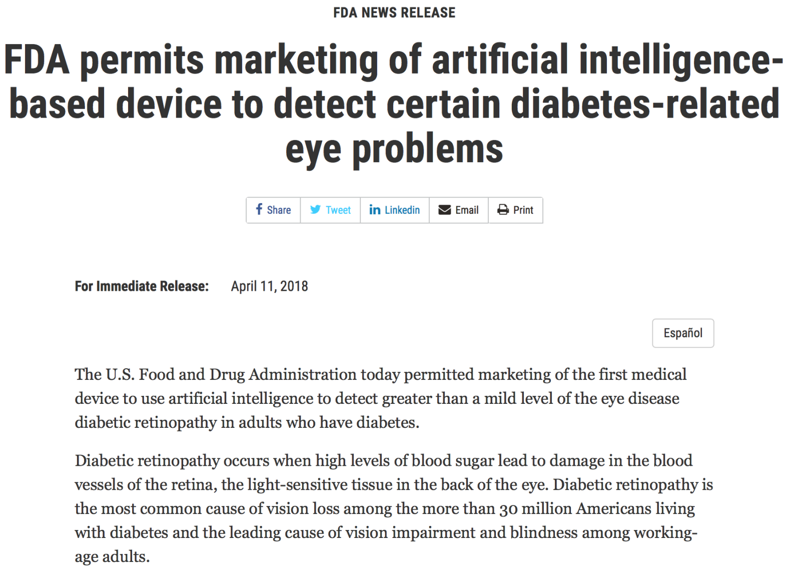
STRUCTURE SOLVER

DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.

A score above 90 is considered roughly equivalent to the experimentally determined structure

©nature

AI techniques replace doctors



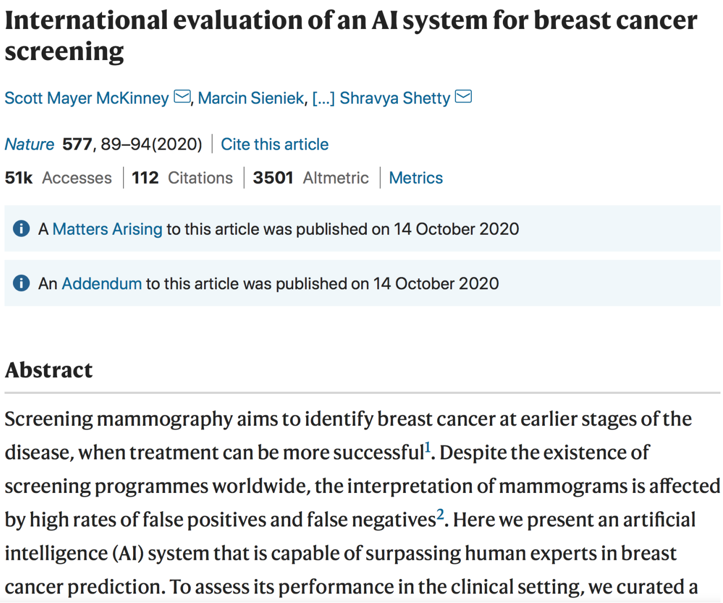
FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems

For Immediate Release: April 11, 2018

The U.S. Food and Drug Administration today permitted marketing of the first medical device to use artificial intelligence to detect greater than a mild level of the eye disease diabetic retinopathy in adults who have diabetes.

Diabetic retinopathy occurs when high levels of blood sugar lead to damage in the blood vessels of the retina, the light-sensitive tissue in the back of the eye. Diabetic retinopathy is the most common cause of vision loss among the more than 30 million Americans living with diabetes and the leading cause of vision impairment and blindness among working-age adults.

Automated cancer diagnosis with AI



International evaluation of an AI system for breast cancer screening

Scott Mayer McKinney, Marcin Sieniek, [...] Shravya Shetty

Nature **577**, 89–94(2020) | Cite this article

51k Accesses | **112** Citations | **3501** Altmetric | Metrics

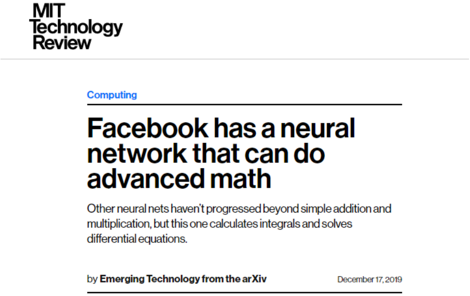
A Matters Arising to this article was published on 14 October 2020

An Addendum to this article was published on 14 October 2020

Abstract

Screening mammography aims to identify breast cancer at earlier stages of the disease, when treatment can be more successful[1]. Despite the existence of screening programmes worldwide, the interpretation of mammograms is affected by high rates of false positives and false negatives[2]. Here we present an artificial intelligence (AI) system that is capable of surpassing human experts in breast cancer prediction. To assess its performance in the clinical setting, we curated a

No more integrals?! AI can solve your maths homework!



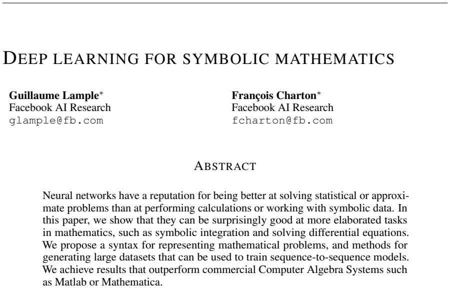
MIT Technology Review

Computing

Facebook has a neural network that can do advanced math

Other neural nets haven't progressed beyond simple addition and multiplication, but this one calculates integrals and solves differential equations.

by **Emerging Technology from the arXiv** December 17, 2019



DEEP LEARNING FOR SYMBOLIC MATHEMATICS

Guillaume Lample*
Facebook AI Research
glample@fb.com

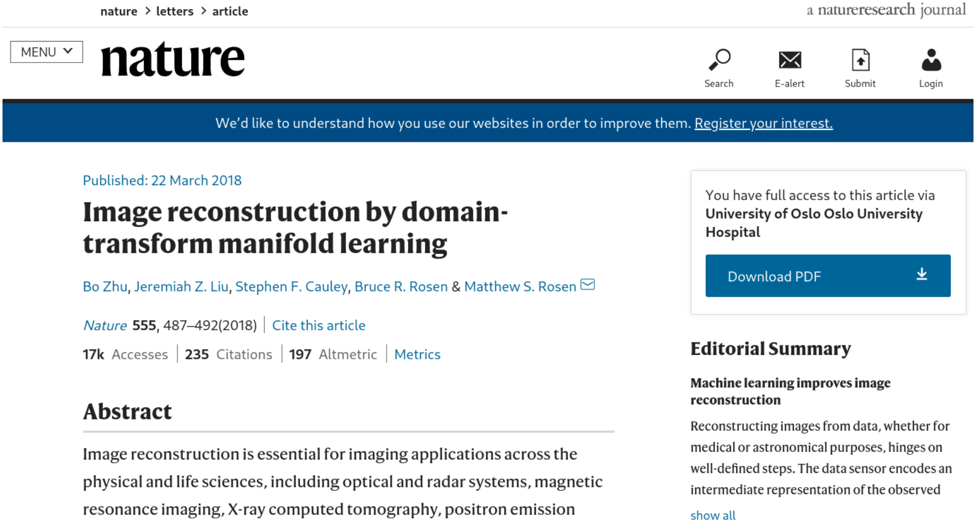
François Charton*
Facebook AI Research
fcharton@fb.com

ABSTRACT

Neural networks have a reputation for being better at solving statistical or approximate problems than at performing calculations or working with symbolic data. In this paper, we show that they can be surprisingly good at more elaborated tasks in mathematics, such as symbolic integration and solving differential equations. We propose a syntax for representing mathematical problems, and methods for generating large datasets that can be used to train sequence-to-sequence models. We achieve results that outperform commercial Computer Algebra Systems such as Matlab or Mathematica.

AI replaces standard algorithms in medical imaging

Claim: “superior immunity to noise and a reduction in reconstruction artefacts compared with conventional handcrafted reconstruction methods”



We’d like to understand how you use our websites in order to improve them. Register your interest.

Published: 22 March 2018

Image reconstruction by domain-transform manifold learning

Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen & Matthew S. Rosen

Nature **555**, 487–492(2018) | Cite this article

17k Accesses | **235** Citations | **197** Altmetric | Metrics

Abstract

Image reconstruction is essential for imaging applications across the physical and life sciences, including optical and radar systems, magnetic resonance imaging, X-ray computed tomography, positron emission

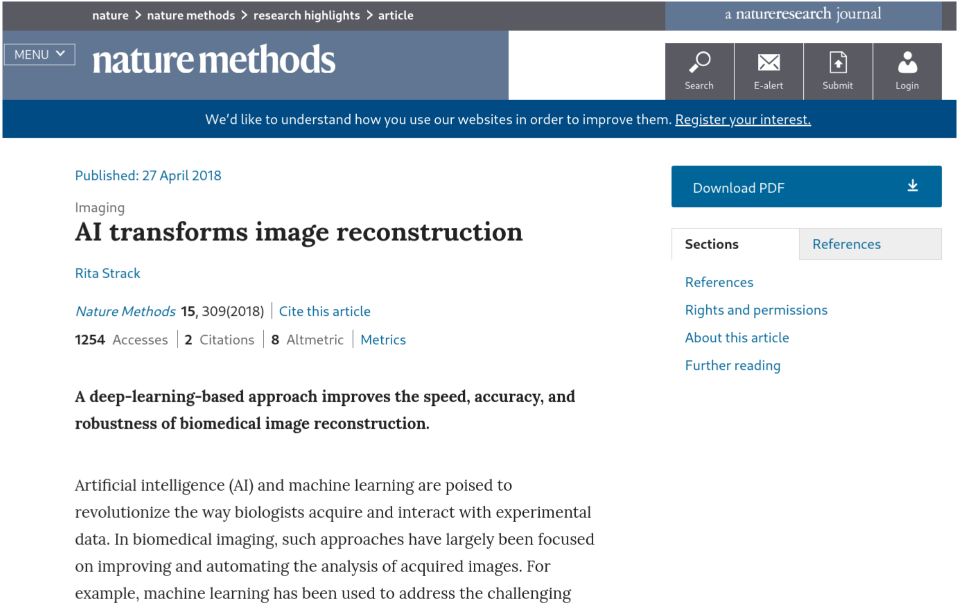
You have full access to this article via **University of Oslo Oslo University Hospital**

Editorial Summary

Machine learning improves image reconstruction

Reconstructing images from data, whether for medical or astronomical purposes, hinges on well-defined steps. The data sensor encodes an intermediate representation of the observed

A bold claim?!



We'd like to understand how you use our websites in order to improve them. Register your interest.

Published: 27 April 2018

Imaging

AI transforms image reconstruction

Rita Strack

Nature Methods **15**, 309(2018) | Cite this article

1254 Accesses | **2** Citations | **8** Altmetric | Metrics

A deep-learning-based approach improves the speed, accuracy, and robustness of biomedical image reconstruction.

Artificial intelligence (AI) and machine learning are poised to revolutionize the way biologists acquire and interact with experimental data. In biomedical imaging, such approaches have largely been focused on improving and automating the analysis of acquired images. For example, machine learning has been used to address the challenging

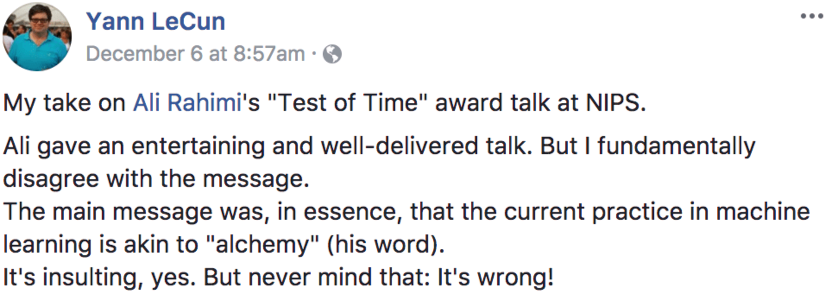
Strong confidence in deep learning

The New Yorker (most influential magazine in the world) quotes Geoffrey Hinton (April 2017):

“They should stop training radiologists now.”

An intriguing debate

Google's Ali Rahimi, winner of the Test-of-Time award 2017 (NeurIPS), "Machine learning has become alchemy. ... I would like to live in a society whose systems are built on top of verifiable, rigorous, thorough knowledge, and not on alchemy."



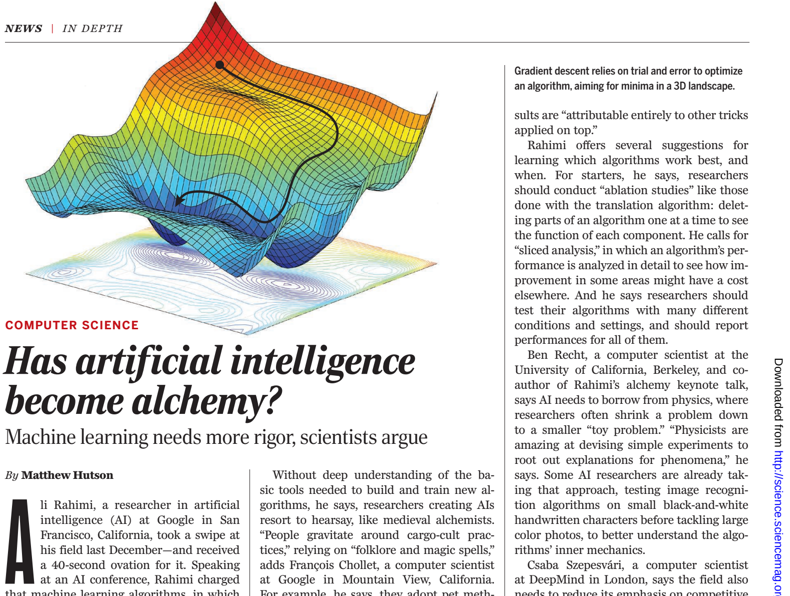
Yann LeCun

December 6 at 8:57am ·

My take on Ali Rahimi's "Test of Time" award talk at NIPS.

Ali gave an entertaining and well-delivered talk. But I fundamentally disagree with the message.
The main message was, in essence, that the current practice in machine learning is akin to "alchemy" (his word).
It's insulting, yes. But never mind that: It's wrong!

An intriguing debate



Gradient descent relies on trial and error to optimize an algorithm, aiming for minima in a 3D landscape.

COMPUTER SCIENCE

Has artificial intelligence become alchemy?

Machine learning needs more rigor, scientists argue

By **Matthew Hutson**

Ali Rahimi, a researcher in artificial intelligence (AI) at Google in San Francisco, California, took a swipe at his field last December—and received a 40-second ovation for it. Speaking at an AI conference, Rahimi charged that machine learning algorithms, in which

Without deep understanding of the basic tools needed to build and train new algorithms, he says, researchers creating AIs resort to hearsay, like medieval alchemists. “People gravitate around cargo-cult practices,” relying on “folklore and magic spells,” adds François Chollet, a computer scientist at Google in Mountain View, California. For example, he says, they adopt pet meth-

sults are “attributable entirely to other tricks applied on top.”

Rahimi offers several suggestions for learning which algorithms work best, and when. For starters, he says, researchers should conduct “ablation studies” like those done with the translation algorithm: deleting parts of an algorithm one at a time to see the function of each component. He calls for “sliced analysis,” in which an algorithm’s performance is analyzed in detail to see how improvement in some areas might have a cost elsewhere. And he says researchers should test their algorithms with many different conditions and settings, and should report performances for all of them.

Ben Recht, a computer scientist at the University of California, Berkeley, and coauthor of Rahimi’s alchemy keynote talk, says AI needs to borrow from physics, where researchers often shrink a problem down to a smaller “toy problem.” “Physicists are amazing at devising simple experiments to root out explanations for phenomena,” he says. Some AI researchers are already taking that approach, testing image recognition algorithms on small black-and-white handwritten characters before tackling large color photos, to better understand the algorithms’ inner mechanics.

Csaba Szepesvári, a computer scientist at DeepMind in London, says the field also needs to reduce its emphasis on competitive

Current challenges

It seems AI is unstoppable, but there are problems with current technologies

- Explainability
- Theoretical guarantees
- Stability (e.g. adversarial examples)

Later: Aspects of explainability and theoretical guarantees.
Next: Case study of stability (or, rather, lack thereof)...

THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Deep learning for decision problems

Deep learning for classification

Object: a classification function $f : \mathbb{R}^d \to \{0,1\}$

What we are given: a training set $\mathcal{T} = \{(x^1, f(x^1)), \ldots, (x^r, f(x^r))\} \subset \mathbb{R}^d \times \{0,1\}$.

Goal: find a "good" approximation $\tilde{f} : \mathbb{R}^d \to \{0,1\}$ to f.

Test $\tilde{f}$ on a classification (or a test) set $\mathcal{C} = \{y^1, \ldots, y^s\}$. Success is measured by

$$\frac{|\{y^j \in \mathcal{C} \mid f(y^j) = \tilde{f}(y^j)\}|}{s}$$

https://towardsdatascience.com/10-papers-you-should-read-to-understand-image- classification-in-the-deep-learning-era-4b9d792f45a7

Fawzi, A., Moosavi-Dezfooli, S.M. and Frossard, P., 2017. *The robustness of deep networks: A geometrical perspective.* IEEE Signal Processing Magazine.

Shalev-Shwartz, S. and Ben-David, S., 2014. *Understanding machine learning: From theory to algorithms.* Cambridge university press.

Neural networks (NNs)

Let $\mathcal{N}\mathcal{N}_{\mathbf{N}}$, with $\mathbf{N} = (N_L, N_{L-1}, \dots, N_1, N_0 = d)$ denote the set of all L-layer neural networks. That is, all mappings $\phi : \mathbb{R}^d \to \mathbb{R}^{N_L}$ of the form

$$\phi(x) = V_L(\rho(V_{L-1}(\rho(\dots\rho(V_1(x)))))), \quad x \in \mathbb{R}^d.$$

$$V_j y = W_j y + b_j, \qquad W_j \in \mathbb{R}^{N_j \times N_{j-1}}, \quad b_j \in \mathbb{R}^{N_j}$$

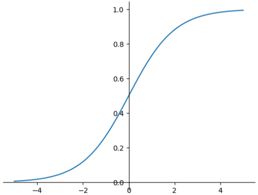
Here $\rho : \mathbb{R} \to \mathbb{R}$ is some non-linear function that acts pointwise on a vector.

NB: Other architectures are possible, e.g. skip connections, different ρ's etc.

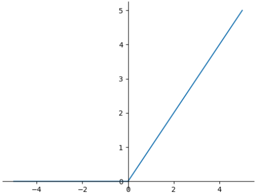
Reference for definition of NN:
Pinkus, A., 1999. *Approximation theory of the MLP model in neural networks*. Acta numerica, 8(1), 143-195.

Common choices of ρ

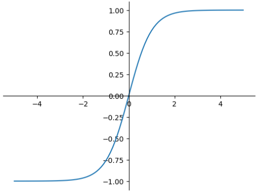
$\rho \colon \mathbb{R} \to \mathbb{R}$ acts elementwise on a vector.



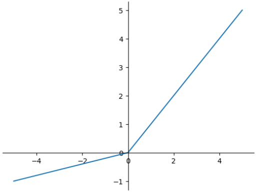
Sigmoid: $\rho(x) = 1/(1 + e^{-x})$



ReLu: $\rho(x) = \max(0, x)$



tanh: $\rho(x) = tanh(x)$



Leaky ReLu: $\rho(x) = \begin{cases} x & x \geq 0 \\ \alpha x & x < 0 \end{cases}$

Approximation qualities of neural nets

Theorem (Universal Approximation Theorem)

Let $\rho \in C(\mathbb{R})$ and assume that ρ is not a polynomial. Let $K \subset \mathbb{R}^d$ be compact, $f \in C(K)$ and $\epsilon > 0$. Then there exists a neural network (with on hidden layer) ϕ such that

$$\sup_{x \in K} |\phi(x) - f(x)| \leq \epsilon.$$

Theorem (Universal Interpolation Theorem)

Let $\rho \in C(\mathbb{R})$ and assume that ρ is not a polynomial. For any k distinct points $\{x_j\}_{j=1}^k \subset \mathbb{R}^d$ and associated data $\{\alpha_j\}_{j=1}^k \subset \mathbb{R}$. Then there exists a neural network (with on hidden layer) ϕ such that

$$\phi(x_j) = \alpha_j, \qquad j = 1, \dots, k.$$

Gühring, I., Kutyniok, G. and Petersen, P., 2020. *Error bounds for approximations with deep ReLU neural networks in $W^{s,p}$ norms.* Analysis and Applications, 18(05), pp.803-859.
Pinkus, A., 1999. *Approximation theory of the MLP model in neural networks.* Acta numerica, 143-195.

Approximation qualities of neural nets

A zoo of so-called "universal approximation" theorems. However, this is not enough:

(a) Other methods (e.g. polynomials, splines, wavelets, etc.) have universal approximation theorems. Why are NNs so effective? E.g., are there useful classes of functions that are efficiently approximated by NNs but not classical methods?

(b) We want to construct or compute a good neural network. There is a subtle difference between existence and computability (more on this later).

We will focus on point (b). For point (a) (which is largely open) see:

DeVore, R., Hanin, B. and Petrova, G., 2020. *Neural Network Approximation.* arXiv preprint arXiv:2012.14501.

Training Neural Networks

Let $\theta \in \mathbb{R}^n$ be the weights of a neural network $\phi(\cdot, \theta) \in \mathcal{N}\mathcal{N}_{\mathbf{N}}$.

Given a classification function $f : \mathbb{R}^d \to \{0, 1\}$, a training set $\mathcal{T} = \{x^1, \ldots, x^r\} \subset \mathbb{R}^d$, a cost function $C : \mathbb{R}^{N_L} \times \mathbb{R}^{N_L} \to \mathbb{R}_+$, and seek to minimize

$$\operatorname*{minimize}_{\theta \in \mathbb{R}^n} L(\theta) := \sum_{i=1}^{r} C(\phi(x^j, \theta), f(x^j)),$$

using gradient based methods

$$\theta_{i+1} = \theta_i - \eta \nabla_\theta L(\theta_i)$$

This has been a HUGE empirical success!

What could go wrong?

(i) There does not exist a neural network that approximates the function we are interested in.

(ii)

(iii)

What could go wrong?

(i) ~~There does not exist a neural network that approximates the function we are interested in.~~

(ii)

(iii)

NB: There is a mathematical theory suggesting that neural nets have all the approximation qualities that are needed.

What could go wrong?

(i) ~~There does not exist a neural network that approximates the function we are interested in.~~

(ii) There does exist a neural network that approximates the function, however, there does not exist an algorithm that can construct the neural network.

(iii)

NB: There is a mathematical theory suggesting that neural nets have all the approximation qualities that are needed.

What could go wrong?

(i) ~~There does not exist a neural network that approximates the function we are interested in.~~

(ii) There does exist a neural network that approximates the function, however, there does not exist an algorithm that can construct the neural network.

(iii) There does exist a neural network that approximates the function, and an algorithm to construct it. However, the algorithm will need prohibitively many samples.

NB: There is a mathematical theory suggesting that neural nets have all the approximation qualities that are needed.

What could go wrong?

(i) ~~There does not exist a neural network that approximates the function we are interested in.~~

(ii) There does exist a neural network that approximates the function, however, there does not exist an algorithm that can construct the neural network.

(iii) There does exist a neural network that approximates the function, and an algorithm to construct it. However, the algorithm will need prohibitively many samples.

NB: There is a mathematical theory suggesting that neural nets have all the approximation qualities that are needed.

What could go wrong?

(i) ~~There does not exist a neural network that approximates the function we are interested in.~~

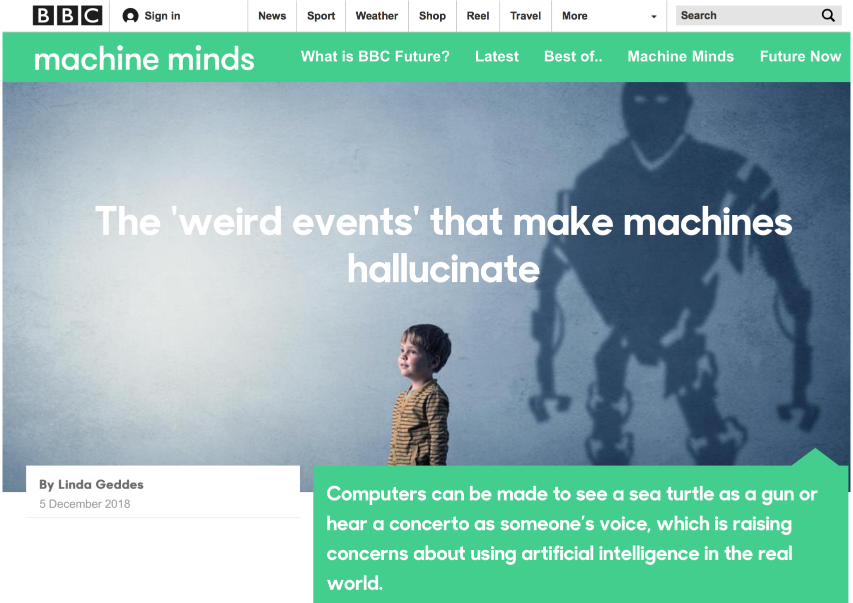
(ii) There does exist a neural network that approximates the function, however, there does not exist an algorithm that can construct the neural network.

(iii) There does exist a neural network that approximates the function, and an algorithm to construct it. However, the algorithm will need prohibitively many samples.

NB: There is a mathematical theory suggesting that neural nets have all the approximation qualities that are needed.

We'll see examples of (ii) and (iii) later.

AI Generated Hallucinations – Instabilities in Deep Learning



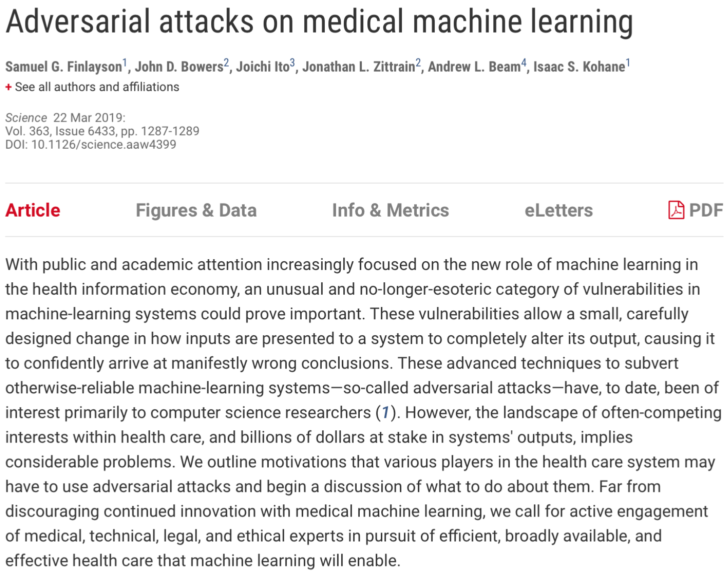
The 'weird events' that make machines hallucinate

By Linda Geddes

5 December 2018

Computers can be made to see a sea turtle as a gun or hear a concerto as someone's voice, which is raising concerns about using artificial intelligence in the real world.

AI Generated Hallucinations in classification/decision problems



Adversarial attacks on medical machine learning

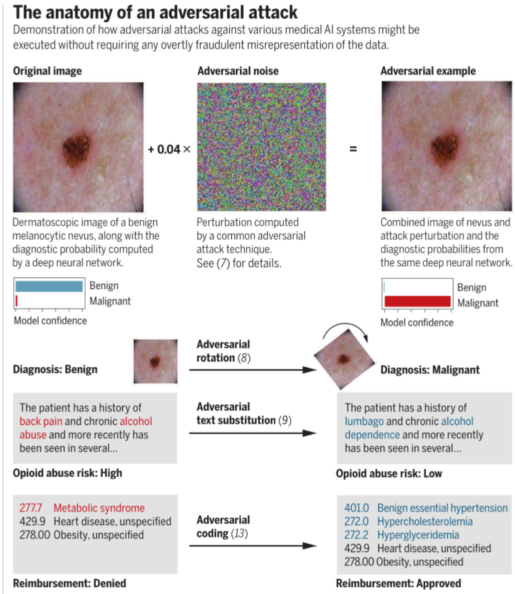
Samuel G. Finlayson[1], John D. Bowers[2], Joichi Ito[3], Jonathan L. Zittrain[2], Andrew L. Beam[4], Isaac S. Kohane[1]

+ See all authors and affiliations

Science 22 Mar 2019:
Vol. 363, Issue 6433, pp. 1287-1289
DOI: 10.1126/science.aaw4399

With public and academic attention increasingly focused on the new role of machine learning in the health information economy, an unusual and no-longer-esoteric category of vulnerabilities in machine-learning systems could prove important. These vulnerabilities allow a small, carefully designed change in how inputs are presented to a system to completely alter its output, causing it to confidently arrive at manifestly wrong conclusions. These advanced techniques to subvert otherwise-reliable machine-learning systems—so-called adversarial attacks—have, to date, been of interest primarily to computer science researchers (*1*). However, the landscape of often-competing interests within health care, and billions of dollars at stake in systems' outputs, implies considerable problems. We outline motivations that various players in the health care system may have to use adversarial attacks and begin a discussion of what to do about them. Far from discouraging continued innovation with medical machine learning, we call for active engagement of medical, technical, legal, and ethical experts in pursuit of efficient, broadly available, and effective health care that machine learning will enable.

AI Generated Hallucinations in classification/decision problems



The anatomy of an adversarial attack

Demonstration of how adversarial attacks against various medical AI systems might be executed without requiring any overtly fraudulent misrepresentation of the data.

Original image

Dermatoscopic image of a benign melanocytic nevus, along with the diagnostic probability computed by a deep neural network.

Adversarial noise

Perturbation computed by a common adversarial attack technique. See (*7*) for details.

Adversarial example

Combined image of nevus and attack perturbation and the diagnostic probabilities from the same deep neural network.

Diagnosis: Benign

Adversarial rotation (*8*)

Diagnosis: Malignant

The patient has a history of back pain and chronic alcohol abuse and more recently has been seen in several...

Adversarial text substitution (*9*)

The patient has a history of lumbago and chronic alcohol dependence and more recently has been seen in several...

Opioid abuse risk: High

Opioid abuse risk: Low

277.7 Metabolic syndrome
429.9 Heart disease, unspecified
278.00 Obesity, unspecified

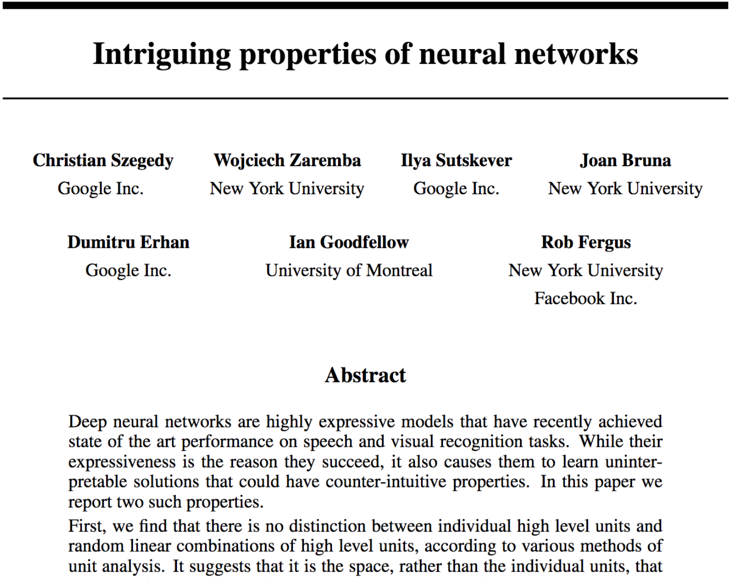
Adversarial coding (*13*)

401.0 Benign essential hypertension
272.0 Hypercholesterolemia
272.2 Hyperglyceridemia
429.9 Heart disease, unspecified
278.00 Obesity, unspecified

Reimbursement: Denied

Reimbursement: Approved

AI Generated Hallucinations – Instabilities in Deep Learning



Intriguing properties of neural networks

Christian Szegedy
Google Inc.

Wojciech Zaremba
New York University

Ilya Sutskever
Google Inc.

Joan Bruna
New York University

Dumitru Erhan
Google Inc.

Ian Goodfellow
University of Montreal

Rob Fergus
New York University
Facebook Inc.

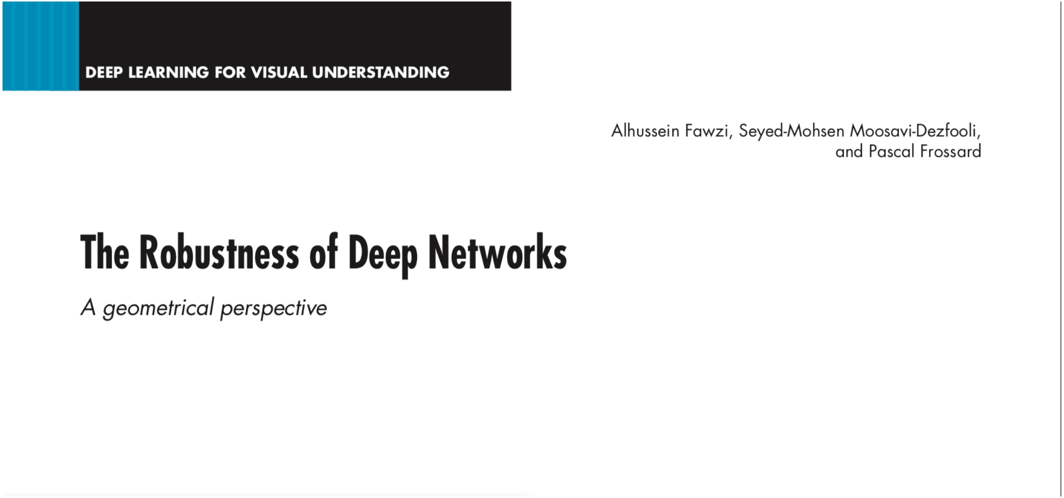
Abstract

Deep neural networks are highly expressive models that have recently achieved state of the art performance on speech and visual recognition tasks. While their expressiveness is the reason they succeed, it also causes them to learn uninterpretable solutions that could have counter-intuitive properties. In this paper we report two such properties.

First, we find that there is no distinction between individual high level units and random linear combinations of high level units, according to various methods of unit analysis. It suggests that it is the space, rather than the individual units, that

Deep Fool

Deep Fool was established at EPFL in order to study the stability of neural networks.



DEEP LEARNING FOR VISUAL UNDERSTANDING

Alhussein Fawzi, Seyed-Mohsen Moosavi-Dezfooli,
and Pascal Frossard

The Robustness of Deep Networks

A geometrical perspective

Reading material

- Fawzi, A., Moosavi-Dezfooli, S. M., Frossard, P. (2017). ‘*The robustness of deep networks: A geometrical perspective*’. IEEE Signal Processing Magazine, 34(6), 50-62.
- Moosavi-Dezfooli, S. M., Fawzi, A., Frossard, P. (2016). ‘*Deepfool: a simple and accurate method to fool deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition*’, (pp. 2574-2582).
- Moosavi-Dezfooli, S. M., Fawzi, A., Fawzi, O., Frossard, P. (2017). ‘*Universal adversarial perturbations*’. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1765-1773).
- Kanbak, C., Moosavi-Dezfooli, S. M., Frossard, P. (2018). ‘*Geometric robustness of deep networks: analysis and improvement*’. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4441-4449).

Deep Fool in practice

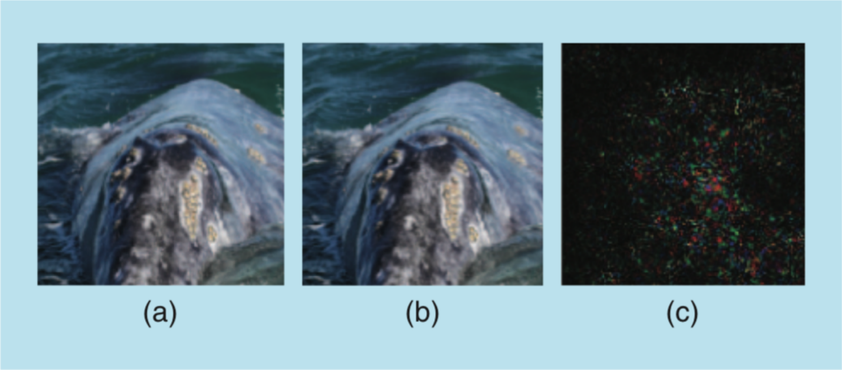


FIGURE 1. An example of an adversarial perturbations in state-of-the-art neural networks. (a) The original image that is classified as a "whale," (b) the perturbed image classified as a "turtle," and (c) the corresponding adversarial perturbation that has been added to the original image to fool a state-of-the-art image classifier [5].

Deep Fool: Universal perturbations

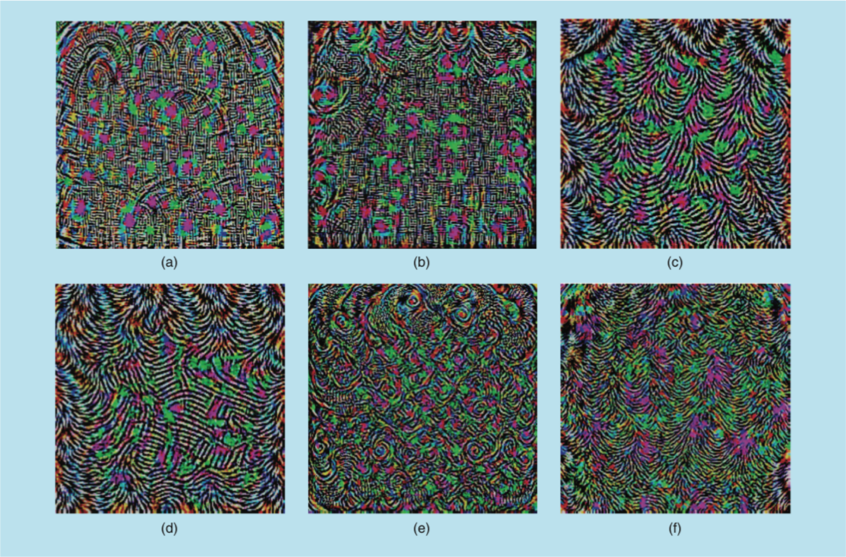
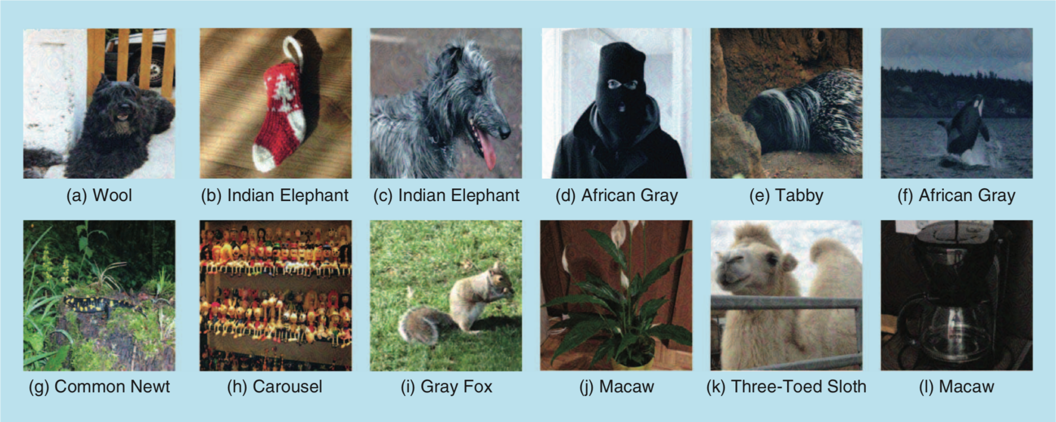


FIGURE 3. Universal perturbations computed for different deep neural network architectures. The pixel values are scaled for visibility. (a) CaffeNet, (b) VGG-F, (c) VGG-16, (d) VGG-19, (e) GoogLeNet, and (f) ResNet-152.

Deep Fool: Examples



(a) Wool (b) Indian Elephant (c) Indian Elephant (d) African Gray (e) Tabby (f) African Gray

(g) Common Newt (h) Carousel (i) Gray Fox (j) Macaw (k) Three-Toed Sloth (l) Macaw

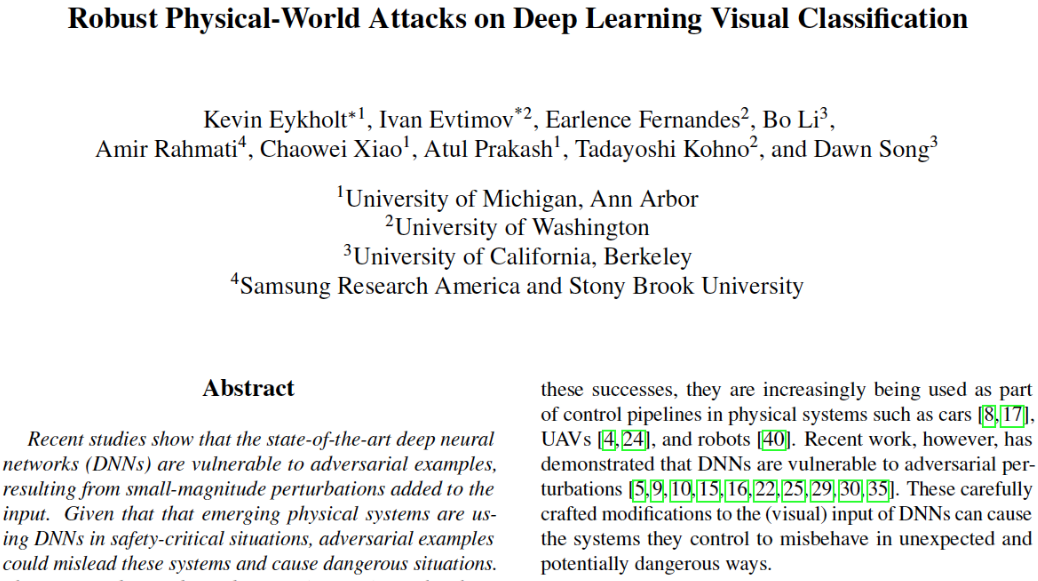
FIGURE 4. Examples of natural images perturbed with the universal perturbation and their corresponding estimated labels with GoogLeNet. (a)–(h) Images belonging to the ILSVRC 2012 validation set. (i)–(l) Personal images captured by a mobile phone camera. (Figure used courtesy of [22].)

Deep Fool: Examples

	VGG-F	CaffeNet	GoogLeNet	VGG-16	VGG-19	ResNet-152
VGG-F	**93.7%**	71.8%	48.4%	42.1%	42.1%	47.4%
CaffeNet	74.0%	**93.3%**	47.7%	39.9%	39.9%	48.0%
GoogLeNet	46.2%	43.8%	**78.9%**	39.2%	39.8%	45.5%
VGG-16	63.4%	55.8%	56.5%	**78.3%**	73.1%	63.4%
VGG-19	64.0%	57.2%	53.6%	73.5%	**77.8%**	58.0%
ResNet-152	46.3%	46.3%	50.5%	47.0%	45.5%	**84.0%**

Table: The rows indicate the architecture for which the universal perturbations is computed, and the columns indicate the architecture for which the fooling rate is reported.

Structural perturbations



Robust Physical-World Attacks on Deep Learning Visual Classification

Kevin Eykholt*[1], Ivan Evtimov*[2], Earlence Fernandes[2], Bo Li[3],
Amir Rahmati[4], Chaowei Xiao[1], Atul Prakash[1], Tadayoshi Kohno[2], and Dawn Song[3]

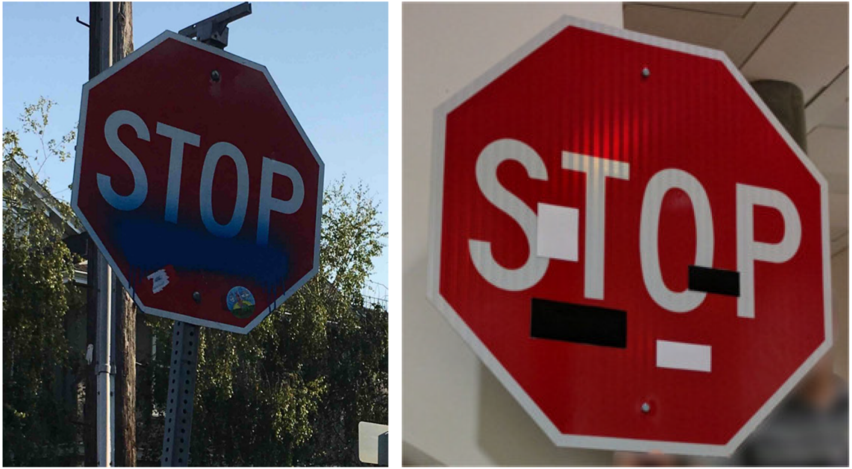
[1]University of Michigan, Ann Arbor
[2]University of Washington
[3]University of California, Berkeley
[4]Samsung Research America and Stony Brook University

Abstract

Recent studies show that the state-of-the-art deep neural networks (DNNs) are vulnerable to adversarial examples, resulting from small-magnitude perturbations added to the input. Given that that emerging physical systems are using DNNs in safety-critical situations, adversarial examples could mislead these systems and cause dangerous situations.

these successes, they are increasingly being used as part of control pipelines in physical systems such as cars [8,17], UAVs [4,24], and robots [40]. Recent work, however, has demonstrated that DNNs are vulnerable to adversarial perturbations [5,9,10,15,16,22,25,29,30,35]. These carefully crafted modifications to the (visual) input of DNNs can cause the systems they control to misbehave in unexpected and potentially dangerous ways.

Structural perturbations



Structural perturbations can also cause the network to fail.

How do we compute additive adversarial perturbations?

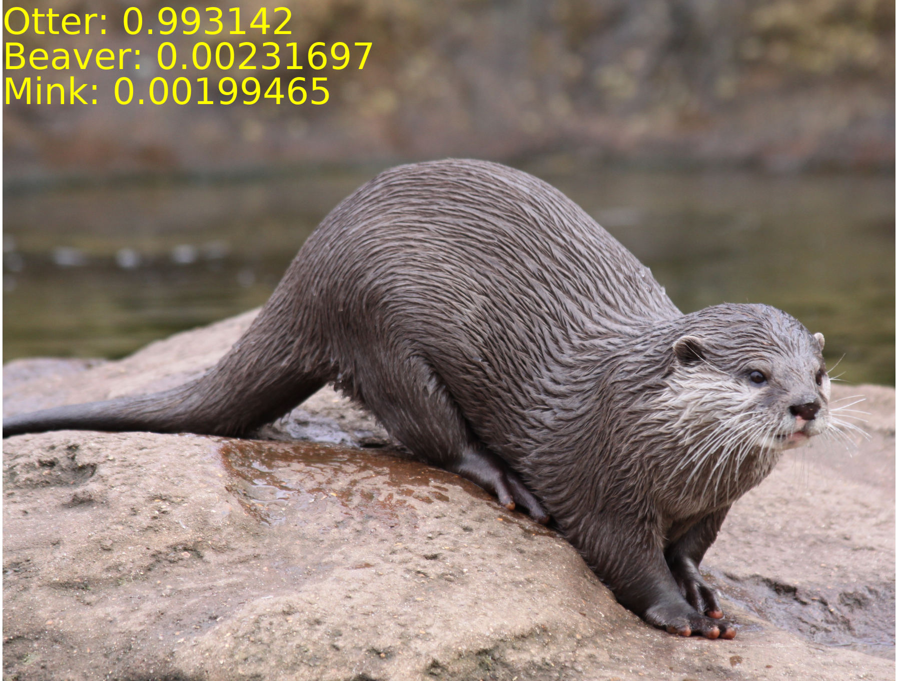
Notation

- $f \colon \mathbb{R}^d \to [0,1]^C$ - Neural network classifier.
- $\hat{k} \colon \mathbb{R}^d \to \{1, 2, \ldots, C\}$. - Predicted label.

$$\hat{k}(x) = \text{argmax}_{j \in \{1,\ldots,C\}} f(x)_j.$$

Seeking minimal perturbation changing the prediction

$$r^*(x) \in \text{argmin}_{r \in \mathbb{R}^d} \|r\| \text{ subject to } \hat{k}(x+r) \neq \hat{k}(x)$$

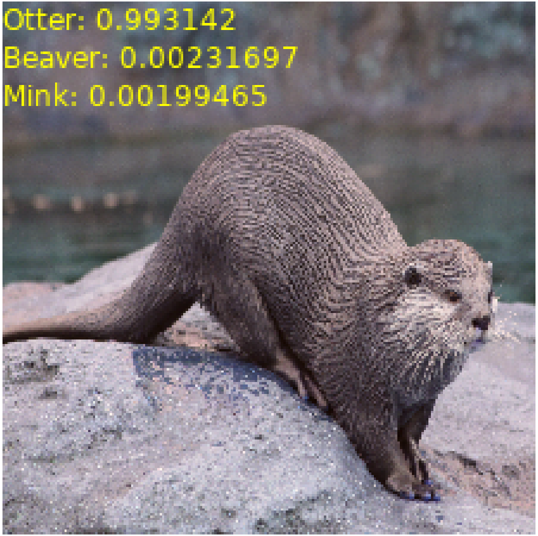


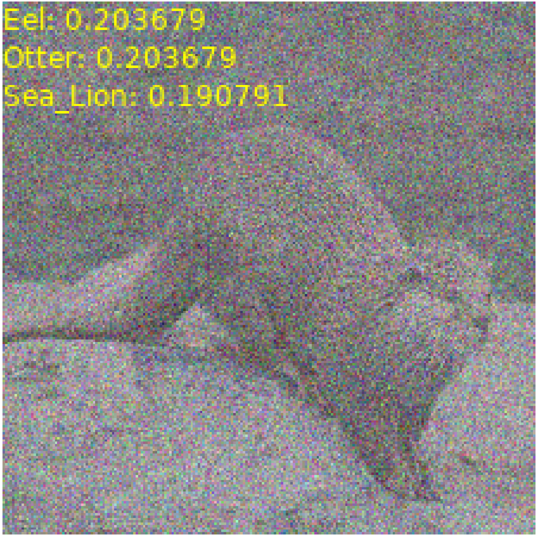
```
Otter: 0.993142
Beaver: 0.00231697
Mink: 0.00199465
```

Random Noise

1. Draw $v \sim \mathcal{N}(0, \mathrm{Id})$ (normal distibution)
2. Solve
$$\alpha^* \in \mathrm{argmin}_{\alpha > 0}\, \alpha \quad \text{subject to} \quad \hat{k}(x + \alpha v) \neq \hat{k}(x)$$
3. Set $r^* = \alpha^* v$

Random Noise



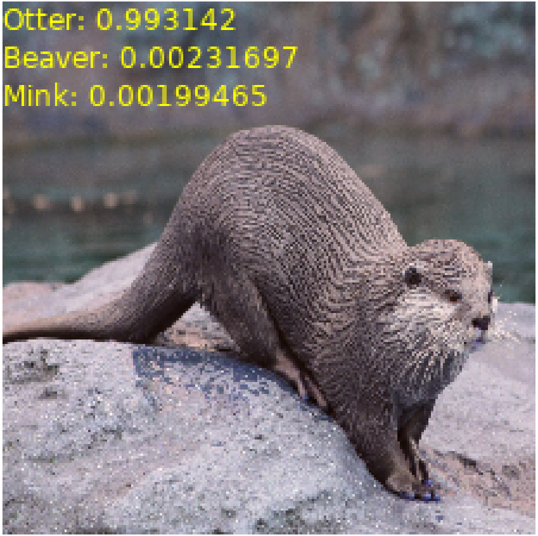


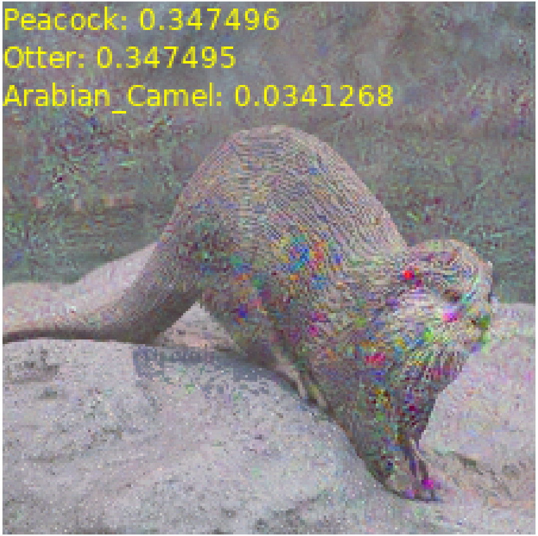
$$\|r\|_{l^2}/\|x\|_{l^2} = 1.47$$

Fast Gradient

1. Let $L(s) = \|f(x+s) - f(x)\|_{l^2}^2$.
2. Let $v = \nabla_s L$
3. Solve
$$\alpha^* \in \text{argmin}_{\alpha > 0} \alpha \quad \text{subject to} \quad \hat{k}(x + \alpha v) \neq \hat{k}(x)$$
4. Set $r^* = \alpha^* v$

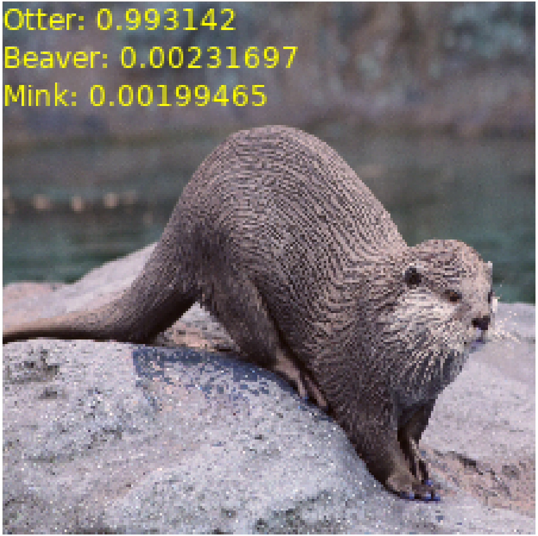
Fast Gradient

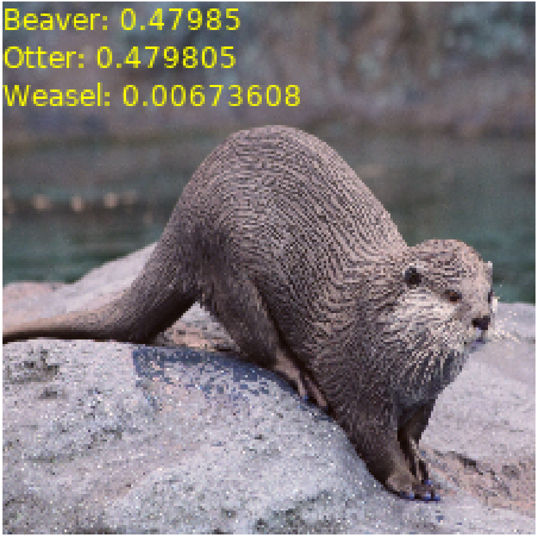




$$\|r\|_{l^2} / \|x\|_{l^2} = 0.34$$

Deep Fool





$$\|r\|_{l^2}/\|x\|_{l^2} = 0.0052$$

Deep Fool – How does it work?

Binary classification:

Given a classification function $f\colon \mathbb{R}^d \to \mathbb{R}$ and an image $x_0 \in \mathbb{R}^d$ we would like to solve

$$r^*(x_0) \in \operatorname{argmin} \|r\|_{l^2} \quad \text{subject to} \quad \operatorname{sign}(f(x_0 + r)) \neq \operatorname{sign}(f(x_0))$$

We make the classification based on $\operatorname{sign}(f(x_0))$.

Deep Fool for Binary Linear Classifier

Let $f : \mathbb{R}^d \to \mathbb{R}$ be given by $f(x) = w^T x + b$.

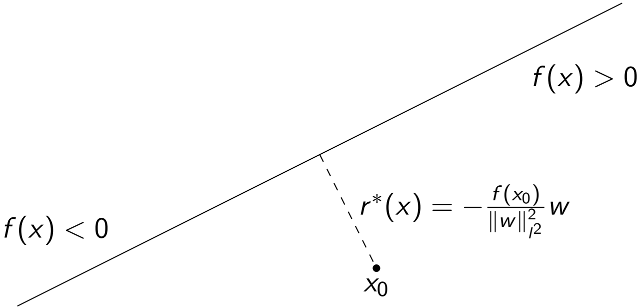


Figure: Adversarial example for a binary linear classifier

Deepfool: a simple and accurate method to fool deep neural networks (2016), S.M. Moosavi-Dezfooli et al.

Deep Fool for Binary Classifier

We would like to solve

$$r^*(x_0) \in \text{argmin}\, \|r\|_{l^2} \quad \text{subject to} \quad \text{sign}(f(x_0 + r)) \neq \text{sign}(f(x_0))$$

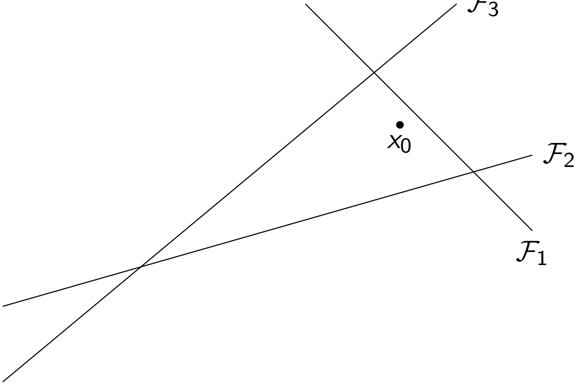
Approximate f at x_i with an linear version $f(x_i) + \nabla f(x_i)^T(x - x_i)$.

```
Input: $image\ x, classifer\ f$
Output: $Perturbation\ \tilde{r}$
Initialize: $x_0 \leftarrow x, i \leftarrow 0, \eta \leftarrow 0.02, \tilde{r} \leftarrow 0$
while $\text{sign}(f(x + (1+\eta)\tilde{r})) = \text{sign}(f(x))$ do
    $r_i \leftarrow -\frac{f(x_i)}{\|\nabla f(x_i)\|_{l^2}^2} \nabla f(x_i)$
    $x_{i+1} \leftarrow x_i + r_i$
    $\tilde{r} \leftarrow \tilde{r} + r_i$
    $i \leftarrow i + 1$
return $\tilde{r}$
```

Deep Fool for Multiclass Linear Classifier

Let

$$P = \bigcap_{k=1}^{C} \{x : f_{\hat{k}(x_0)}(x) \geq f_k(x)\}$$



$$\mathcal{F}_k = \{x : f_{\hat{k}(x_0)}(x) - f_k(x) = 0\}$$

Is the behaviour we are seeing reasonable?

Consider a linear classifier $w^\top x + b$ in $\mathbb{R}^d$. Then

$$w^\top (x + \Delta x) + b = w^\top x + w^\top \Delta x + b$$

Under the constraint $\|\Delta x\|_{l^\infty} \leq \epsilon$ we can choose $\Delta x_i = \epsilon \, \mathrm{sign}(w_i)$ so the classification changes by

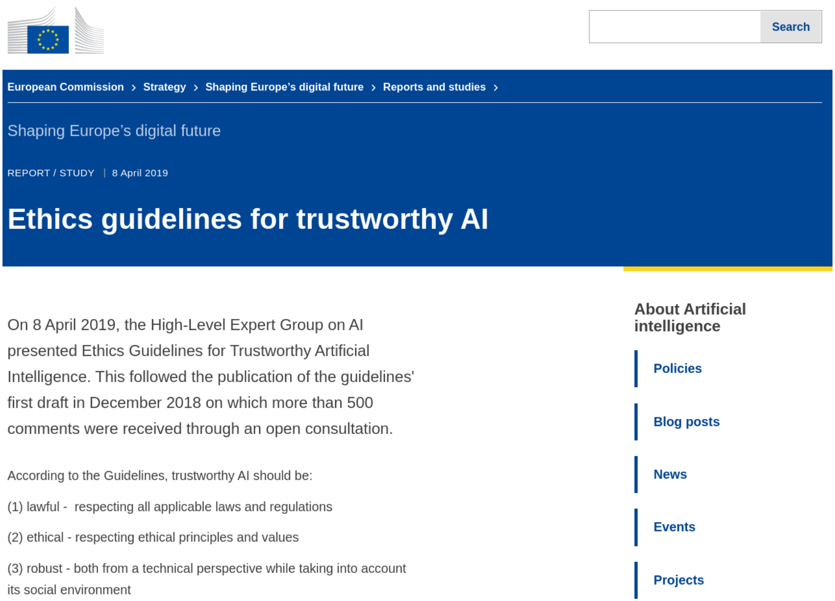
$$\epsilon \|w\|_{l^1} \approx \epsilon \, d \, \langle |w| \rangle$$

where $\langle |w| \rangle$ is the avrage of the entries in w.

What happens for large d?

Insight from:
Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). '*Explaining and harnessing adversarial examples*'. arXiv preprint arXiv:1412.6572.

What does deep learning learn?



European Commission › Strategy › Shaping Europe's digital future › Reports and studies ›

Shaping Europe's digital future

REPORT / STUDY | 8 April 2019

Ethics guidelines for trustworthy AI

On 8 April 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence. This followed the publication of the guidelines' first draft in December 2018 on which more than 500 comments were received through an open consultation.

According to the Guidelines, trustworthy AI should be:

(1) lawful - respecting all applicable laws and regulations

(2) ethical - respecting ethical principles and values

(3) robust - both from a technical perspective while taking into account its social environment

About Artificial intelligence

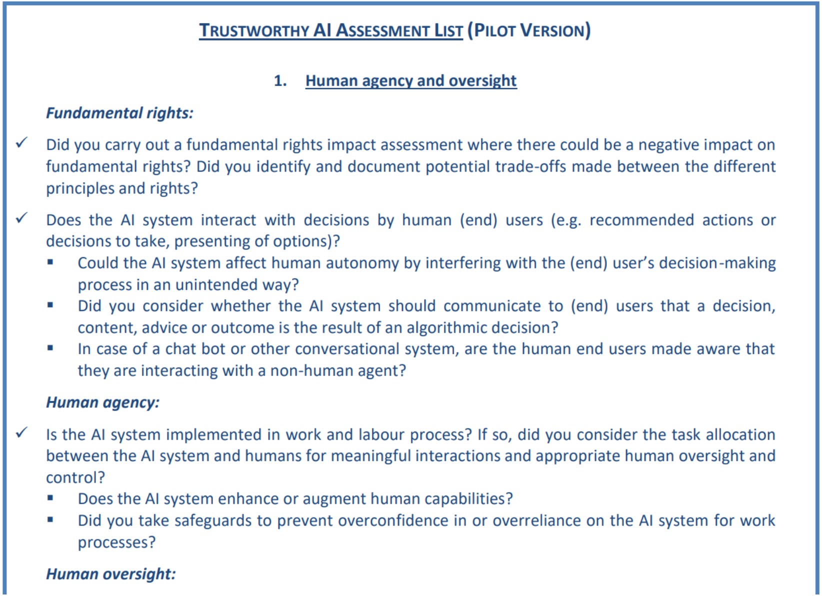
Policies

Blog posts

News

Events

Projects



Trustworthy AI Assessment List (Pilot Version)

1. Human agency and oversight

Fundamental rights:

✓ Did you carry out a fundamental rights impact assessment where there could be a negative impact on fundamental rights? Did you identify and document potential trade-offs made between the different principles and rights?

✓ Does the AI system interact with decisions by human (end) users (e.g. recommended actions or decisions to take, presenting of options)?

- Could the AI system affect human autonomy by interfering with the (end) user's decision-making process in an unintended way?
- Did you consider whether the AI system should communicate to (end) users that a decision, content, advice or outcome is the result of an algorithmic decision?
- In case of a chat bot or other conversational system, are the human end users made aware that they are interacting with a non-human agent?

Human agency:

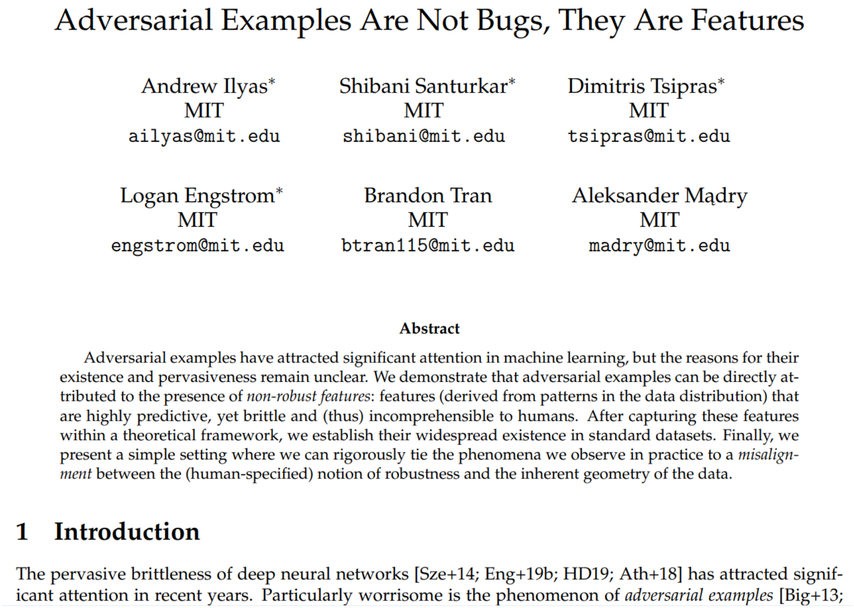
✓ Is the AI system implemented in work and labour process? If so, did you consider the task allocation between the AI system and humans for meaningful interactions and appropriate human oversight and control?

- Does the AI system enhance or augment human capabilities?
- Did you take safeguards to prevent overconfidence in or overreliance on the AI system for work processes?

Human oversight:

What does deep learning learn?

What does deep learning learn?



Adversarial Examples Are Not Bugs, They Are Features

Andrew Ilyas*
MIT
ailyas@mit.edu

Shibani Santurkar*
MIT
shibani@mit.edu

Dimitris Tsipras*
MIT
tsipras@mit.edu

Logan Engstrom*
MIT
engstrom@mit.edu

Brandon Tran
MIT
btran115@mit.edu

Aleksander Mądry
MIT
madry@mit.edu

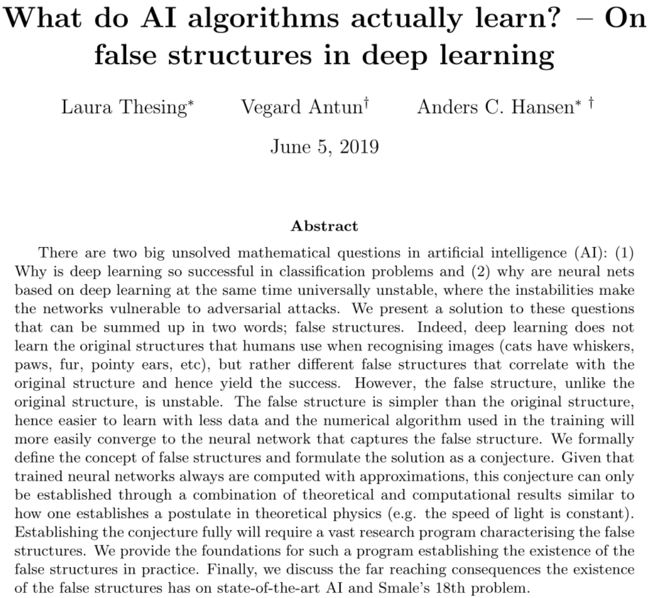
Abstract

Adversarial examples have attracted significant attention in machine learning, but the reasons for their existence and pervasiveness remain unclear. We demonstrate that adversarial examples can be directly attributed to the presence of *non-robust features*: features (derived from patterns in the data distribution) that are highly predictive, yet brittle and (thus) incomprehensible to humans. After capturing these features within a theoretical framework, we establish their widespread existence in standard datasets. Finally, we present a simple setting where we can rigorously tie the phenomena we observe in practice to a *misalignment* between the (human-specified) notion of robustness and the inherent geometry of the data.

1 Introduction

The pervasive brittleness of deep neural networks [Sze+14; Eng+19b; HD19; Ath+18] has attracted significant attention in recent years. Particularly worrisome is the phenomenon of *adversarial examples* [Big+13;

What does deep learning learn?



What do AI algorithms actually learn? – On false structures in deep learning

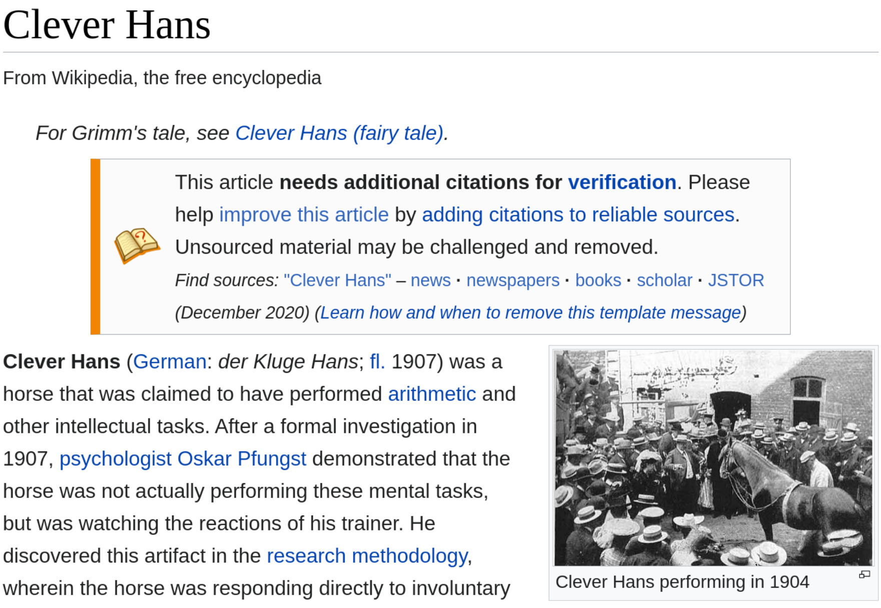
Laura Thesing* Vegard Antun† Anders C. Hansen* †

June 5, 2019

Abstract

There are two big unsolved mathematical questions in artificial intelligence (AI): (1) Why is deep learning so successful in classification problems and (2) why are neural nets based on deep learning at the same time universally unstable, where the instabilities make the networks vulnerable to adversarial attacks. We present a solution to these questions that can be summed up in two words; false structures. Indeed, deep learning does not learn the original structures that humans use when recognising images (cats have whiskers, paws, fur, pointy ears, etc), but rather different false structures that correlate with the original structure and hence yield the success. However, the false structure, unlike the original structure, is unstable. The false structure is simpler than the original structure, hence easier to learn with less data and the numerical algorithm used in the training will more easily converge to the neural network that captures the false structure. We formally define the concept of false structures and formulate the solution as a conjecture. Given that trained neural networks always are computed with approximations, this conjecture can only be established through a combination of theoretical and computational results similar to how one establishes a postulate in theoretical physics (e.g. the speed of light is constant). Establishing the conjecture fully will require a vast research program characterising the false structures. We provide the foundations for such a program establishing the existence of the false structures in practice. Finally, we discuss the far reaching consequences the existence of the false structures has on state-of-the-art AI and Smale's 18th problem.

What does deep learning learn?



Clever Hans

From Wikipedia, the free encyclopedia

For Grimm's tale, see Clever Hans (fairy tale).

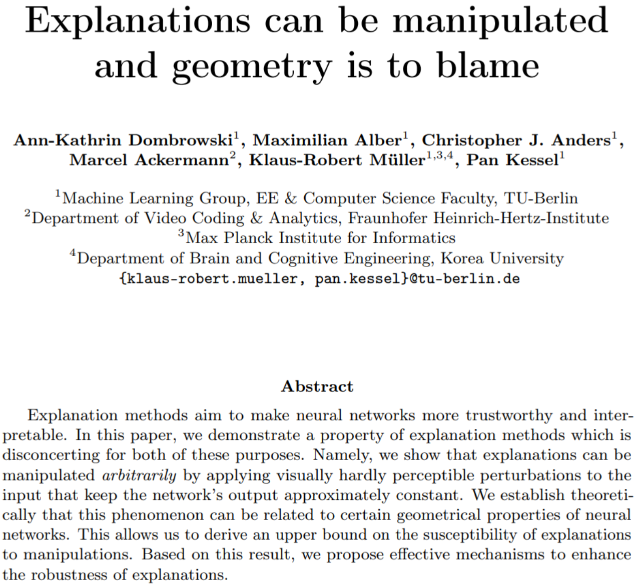
This article **needs additional citations for verification**. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed.

Find sources: "Clever Hans" – news · newspapers · books · scholar · JSTOR *(December 2020)* *(Learn how and when to remove this template message)*

Clever Hans (German: *der Kluge Hans*; fl. 1907) was a horse that was claimed to have performed arithmetic and other intellectual tasks. After a formal investigation in 1907, psychologist Oskar Pfungst demonstrated that the horse was not actually performing these mental tasks, but was watching the reactions of his trainer. He discovered this artifact in the research methodology, wherein the horse was responding directly to involuntary

Clever Hans performing in 1904

Can we check what the neural networks learn?



Explanations can be manipulated and geometry is to blame

Ann-Kathrin Dombrowski[1], **Maximilian Alber**[1], **Christopher J. Anders**[1], **Marcel Ackermann**[2], **Klaus-Robert Müller**[1,3,4], **Pan Kessel**[1]

[1]Machine Learning Group, EE & Computer Science Faculty, TU-Berlin
[2]Department of Video Coding & Analytics, Fraunhofer Heinrich-Hertz-Institute
[3]Max Planck Institute for Informatics
[4]Department of Brain and Cognitive Engineering, Korea University
`{klaus-robert.mueller, pan.kessel}@tu-berlin.de`

Abstract

Explanation methods aim to make neural networks more trustworthy and interpretable. In this paper, we demonstrate a property of explanation methods which is disconcerting for both of these purposes. Namely, we show that explanations can be manipulated *arbitrarily* by applying visually hardly perceptible perturbations to the input that keep the network's output approximately constant. We establish theoretically that this phenomenon can be related to certain geometrical properties of neural networks. This allows us to derive an upper bound on the susceptibility of explanations to manipulations. Based on this result, we propose effective mechanisms to enhance the robustness of explanations.

Can we check what the neural networks learn?

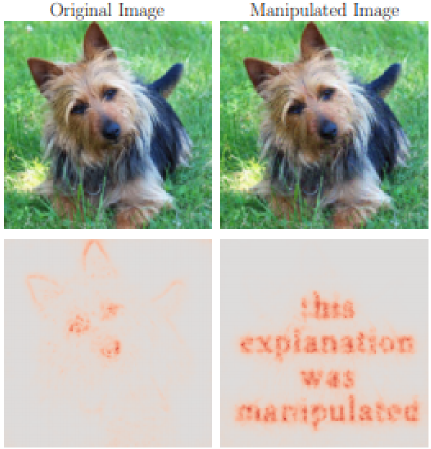


Figure 1: Original image with corresponding explanation map on the left. Manipulated image with its explanation on the right. The chosen target explanation was an image with a text stating "this explanation was manipulated".

The need for foundations.

Echoes of an old story

Hilbert's vision at start of 20th century: provide secure foundations for all mathematics.

- All mathematical statements should be written in a precise formal language, and manipulated according to well defined rules.
- Completeness: a proof that all true mathematical statements can be proved in the formalism.
- Consistency: a proof that no contradiction can be obtained in the formalism of mathematics.
- Decidability: an algorithm for deciding the truth or falsity of any mathematical statement.

Echoes of an old story

Hilbert's 10th problem: *Provide an algorithm which, for any given Diophantine equation (polynomial equation with integer coefficients and finite number of unknowns), can decide whether the equation has an integer-valued solution.*

E.g.: $x^n + y^n = z^n$ (c.f. Fermat's last theorem)

Echoes of an old story

Gödel and Turing turned Hilbert's optimism upside down by showing how there are true statements in mathematics that cannot be proven and that there are problems that cannot be computed by an algorithm.

Hilbert's 10th problem (Solution in 1970, Matiyasevich): No such algorithm exists.

Poonen, B., 2014. *Undecidable problems: a sampler.* Interpreting Gödel: Critical Essays.

A discussion with Fields Medalist Artur Avila

Many scientists from a broad spectrum of areas have expressed various views and opinions about the possible future consequences of the advancement of AI, do you have any views on this?

Being a dynamicist, my training has taught me to a large extent that there are a lot of limitations when we make predictions. To my understanding, when it comes to AI peoples' hopes might be too high at the moment... It is of course correct to continue the effort... When we learn more about a corresponding domain we are also able to better understand its limitations. The point is to try to discover directions where the techniques allow you to go. It is very important and at the same time difficult to know what you cannot achieve... This is what happened in mathematics when we learned that there are unsolvable problems and all kind of limitations to formalism. Our understanding changed completely from Hilbert to Gödel... Maybe, with respect to AI, there are still unknown theoretical limitations hiding in information theory or even dynamical systems.

Full interview: http://www.mthrassias.com/data/uploads/avila_discussion.pdf

A program for the foundations of DL and AI

Smale's 18th problem*: *What are the limits of artificial intelligence?*

A program determining the foundations/limitations of deep learning and AI is needed:

- Boundaries of methodologies.
- Universal/intrinsic boundaries (no algorithm can do it).

There is a key difference between existence and construction here.

Need to also incorporate two pillars of numerical analysis:

- Stability (warning to the reader: there are different types of stability)
- Accuracy

GOAL for rest of lectures: Develop some results in this direction for inverse problems.

*Steve Smale composed a list of problems for the 21st century in reply to a request of Vladimir Arnold inspired by Hilbert's list.

AI in inverse problems and imaging.

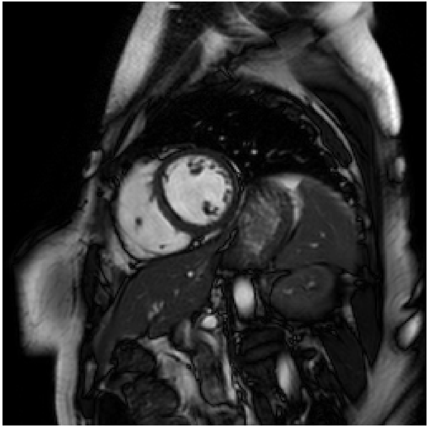
Mathematical setup

Given measurements $\quad y = Ax + e \quad$ recover $\quad x \in \mathbb{C}^N$.

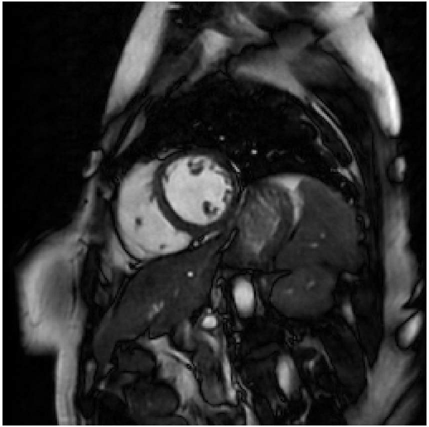
- $x \in \mathbb{C}^N$ be an unknown vector,
- $A \in \mathbb{C}^{m \times N}$ be a matrix ($m < N$) describing modality (e.g. MRI), and
- $y = Ax + e$ the noisy measurements of x.

AI generated hallucinations – Instabilities

$|x|$

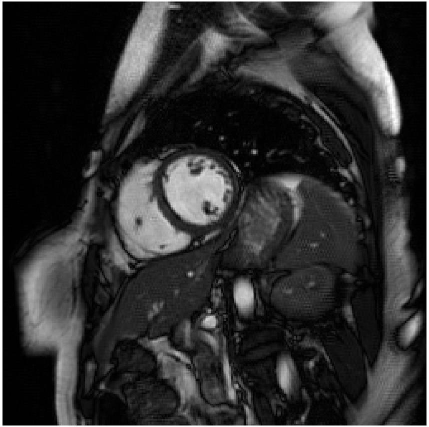


$|\Psi(Ax)|$

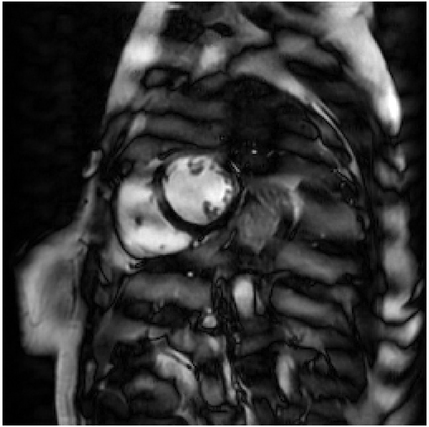


Network from: J. Schlemper, J. Caballero, J. V. Hajnal, A. Price and D. Rueckert, '*A deep cascade of convolutional neural networks for MR image reconstruction*', in International conference on information processing in medical imaging, Springer, 2017, pp. 647–658.
Figures from: Antun, V., Renna, F., Poon, C., Adcock, B., & Hansen, A. C., '*On instabilities of deep learning in image reconstruction and the potential costs of AI*'. Proc. Natl. Acad. Sci. USA, 2020..

AI generated hallucinations – Instabilities



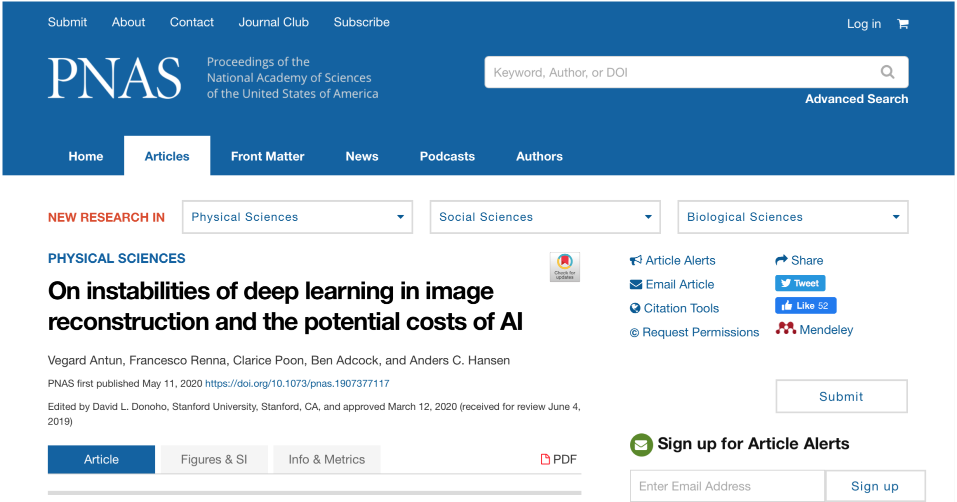
$|x + r|$



$|\Psi(A(x + r))|$

Network from: J. Schlemper, J. Caballero, J. V. Hajnal, A. Price and D. Rueckert, '*A deep cascade of convolutional neural networks for MR image reconstruction*', in International conference on information processing in medical imaging, Springer, 2017, pp. 647–658.
Figures from: Antun, V., Renna, F., Poon, C., Adcock, B., & Hansen, A. C., '*On instabilities of deep learning in image reconstruction and the potential costs of AI*'. Proc. Natl. Acad. Sci. USA, 2020..

DL is unstable in inverse problems



PHYSICAL SCIENCES

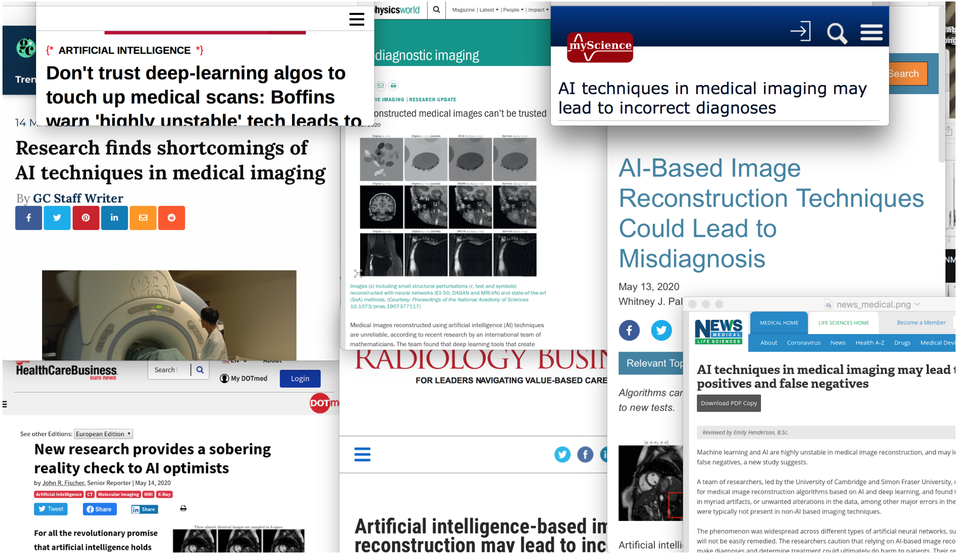
On instabilities of deep learning in image reconstruction and the potential costs of AI

Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen

PNAS first published May 11, 2020 https://doi.org/10.1073/pnas.1907377117

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved March 12, 2020 (received for review June 4, 2019)

The press reports on instabilities



{* ARTIFICIAL INTELLIGENCE *}

Don't trust deep-learning algos to touch up medical scans: Boffins warn 'highly unstable' tech leads to

14 M

Research finds shortcomings of AI techniques in medical imaging

By GC Staff Writer

diagnostic imaging

TIC IMAGING | RESEARCH UPDATE

onstructed medical images can't be trusted

Images (x) including small structural perturbations (r, text and symbols) reconstructed with neural networks (Ell 50, DAGAN and MRI-VN) and state-of-the-art (SoA) methods. (Courtesy: *Proceedings of the National Academy of Sciences* 10.1073/pnas.1907377117)

Medical images reconstructed using artificial intelligence (AI) techniques are unreliable, according to recent research by an international team of mathematicians. The team found that deep learning tools that create

AI techniques in medical imaging may lead to incorrect diagnoses

AI-Based Image Reconstruction Techniques Could Lead to Misdiagnosis

May 13, 2020

Whitney J. Pal

Algorithms car to new tests.

Artificial intelli

See other Editions: European Edition

New research provides a sobering reality check to AI optimists

by John R. Fischer, Senior Reporter | May 14, 2020

Artificial Intelligence | CT | Molecular Imaging | MRI | X-Ray

For all the revolutionary promise that artificial intelligence holds

FOR LEADERS NAVIGATING VALUE-BASED CARE

Artificial intelligence-based in reconstruction may lead to inc

AI techniques in medical imaging may lead t positives and false negatives

Reviewed by Emily Henderson, B.Sc.

Machine learning and AI are highly unstable in medical image reconstruction, and may le false negatives, a new study suggests.

A team of researchers, led by the University of Cambridge and Simon Fraser University, c for medical image reconstruction algorithms based on AI and deep learning, and found t in myriad artifacts, or unwanted alterations in the data, among other major errors in the were typically not present in non-AI based imaging techniques.

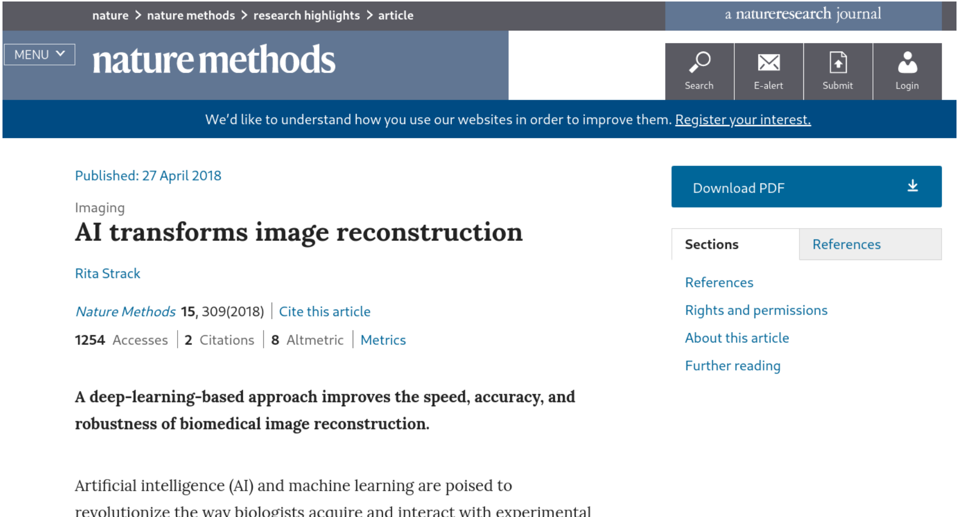
The phenomenon was widespread across different types of artificial neural networks, su will not be easily remedied. The researchers caution that relying on AI-based image reco make diagnoses and determine treatment could ultimately do harm to patients. Their re

Curt Langlotz (Director of the Center for Artificial Intelligence in Medicine and Imaging at Stanford University) tweets:
"Confirming what many believed about deep learning image reconstruction: 'Deep learning typically yields unstable methods for image reconstruction from: 1) tiny perturbations, 2) structural changes, and 3) changes in the number of samples."'

- Given the existence of stable and accurate methods, isn’t this a paradox?
- There is a **trade-off** between stability and accuracy.

Recall the claim

Claim: "superior immunity to noise and a reduction in reconstruction artefacts compared with conventional handcrafted reconstruction methods"



We'd like to understand how you use our websites in order to improve them. Register your interest.

Published: 27 April 2018

Imaging

AI transforms image reconstruction

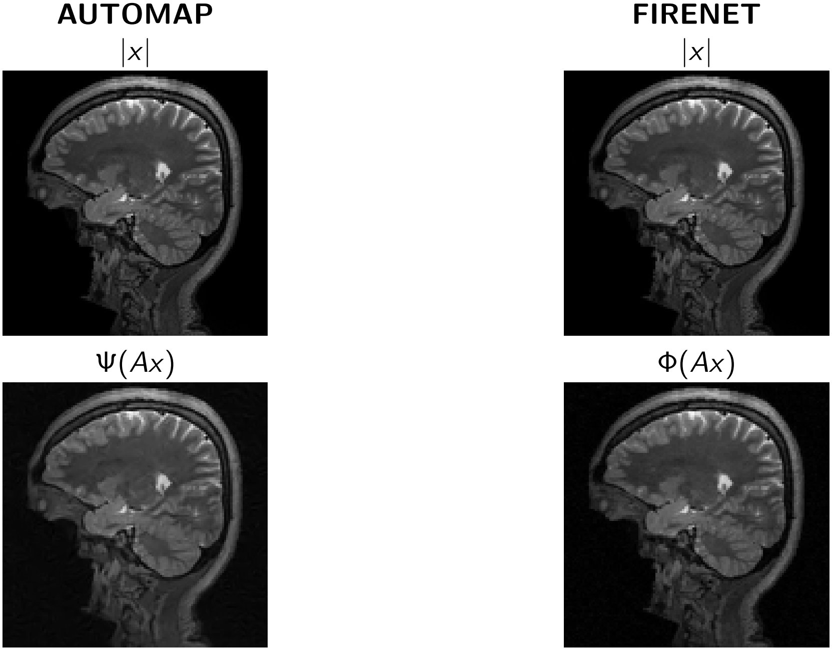
Rita Strack

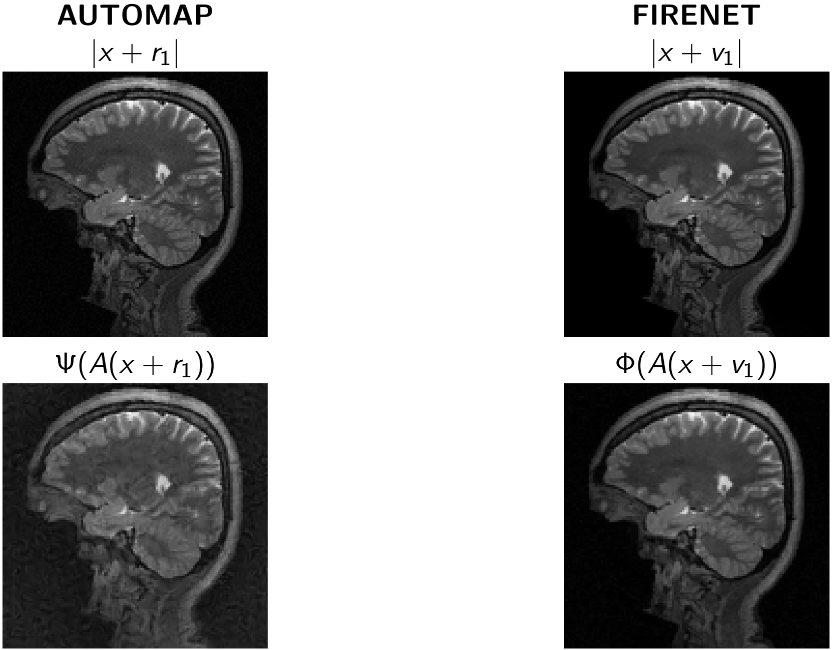
Nature Methods **15**, 309(2018) | Cite this article

1254 Accesses | **2** Citations | **8** Altmetric | Metrics

A deep-learning-based approach improves the speed, accuracy, and robustness of biomedical image reconstruction.

Artificial intelligence (AI) and machine learning are poised to revolutionize the way biologists acquire and interact with experimental


AUTOMAP
$|x|$
$\Psi(Ax)$
FIRENET
$|x|$
$\Phi(Ax)$



AUTOMAP

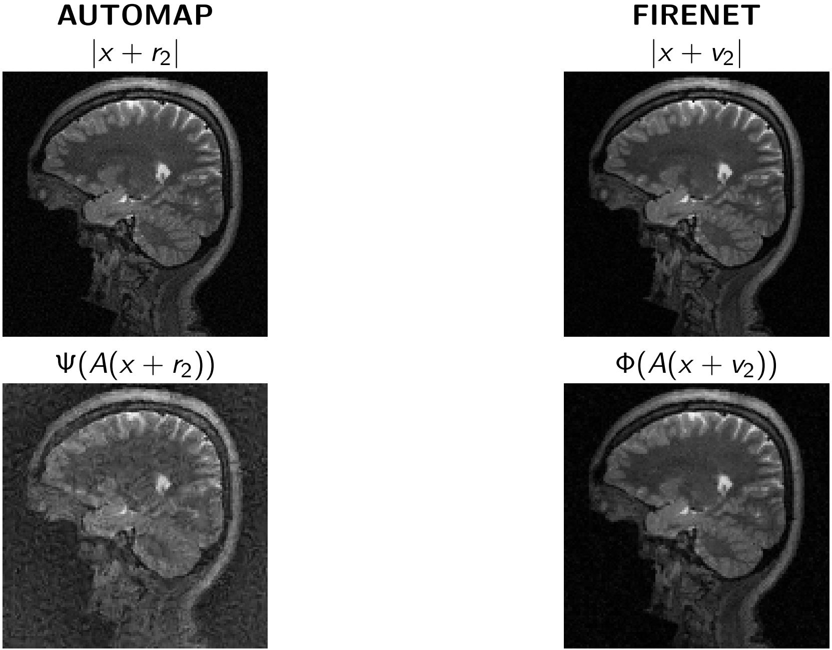
$|x + r_1|$

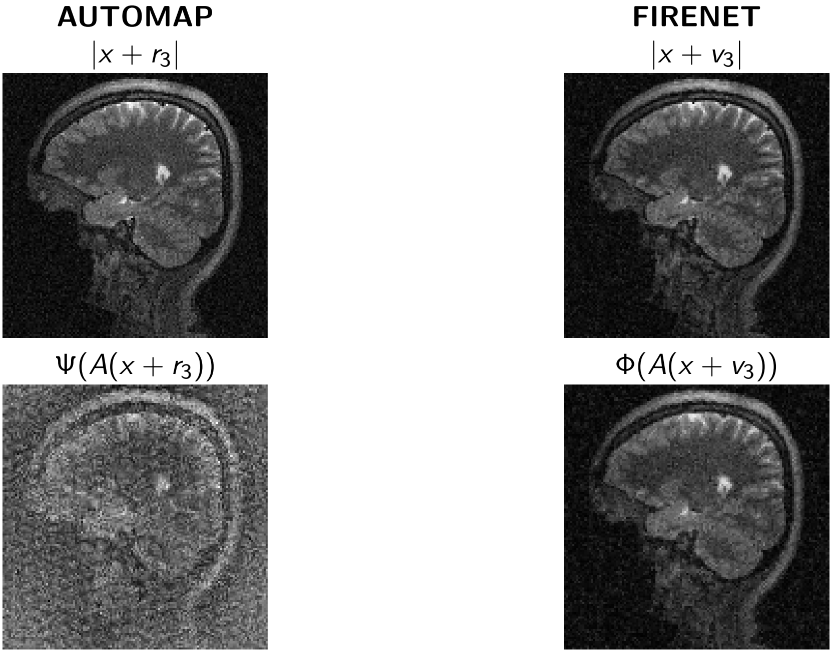
$\Psi(A(x + r_1))$

FIRENET

$|x + v_1|$

$\Phi(A(x + v_1))$


AUTOMAP
FIRENET
$|x + r_2|$
$|x + v_2|$
$\Psi(A(x + r_2))$
$\Phi(A(x + v_2))$



AUTOMAP

$|x + r_3|$

$\Psi(A(x + r_3))$

FIRENET

$|x + v_3|$

$\Phi(A(x + v_3))$

Tune in next time for...

- AI generated hallucinations: The instability test for AI image reconstruction.
- Reasons for instability (kernel awareness).
- An intriguing impossibility result in optimisation and sparse regularisation. (well-conditioned problems where great neural networks exist but cannot be computed/trained)
- Algorithm unrolling.

DAY I	**DAY II**	**Day III**
Gravity of AI Image Classification Need for Foundations AI for Image Reconstruction	Inverse Problems Instabilities & Kernel Awareness Intriguing Barriers Algorithm Unrolling	Achieving Kernel Awareness FIRENETs Imaging Applications Numerical Examples