Maschinelles Lernen: Methoden, Algorithmen, Potentiale und gesellschaftliche Herausforderungen

Felix Wichmann

Neural Information Processing Group and
Bernstein Center for Computational Neuroscience,
Eberhard Karls Universität Tübingen

Max Planck Institute for Intelligent Systems, Tübingen

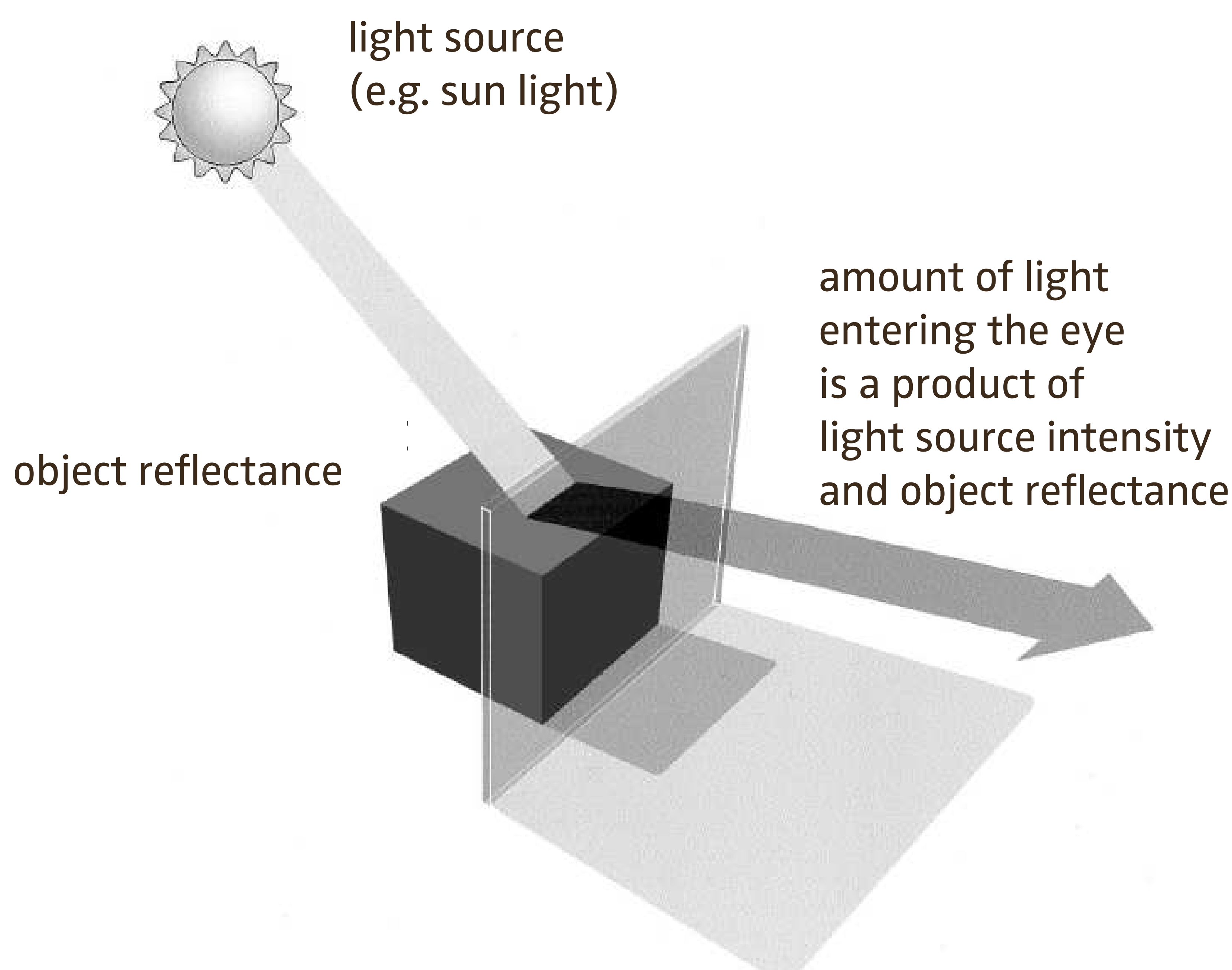
http://www.appblogger.de/wp-content/uploads/2013/03/pb-130314-pope-2005.photoblog900.jpg

http://msnbcmedia.msn.com/j/MSNBC/Components/Photo/_new/pb-130314-pope-2013.photoblog900.jpg

1

One way to think about vision: inverse optics

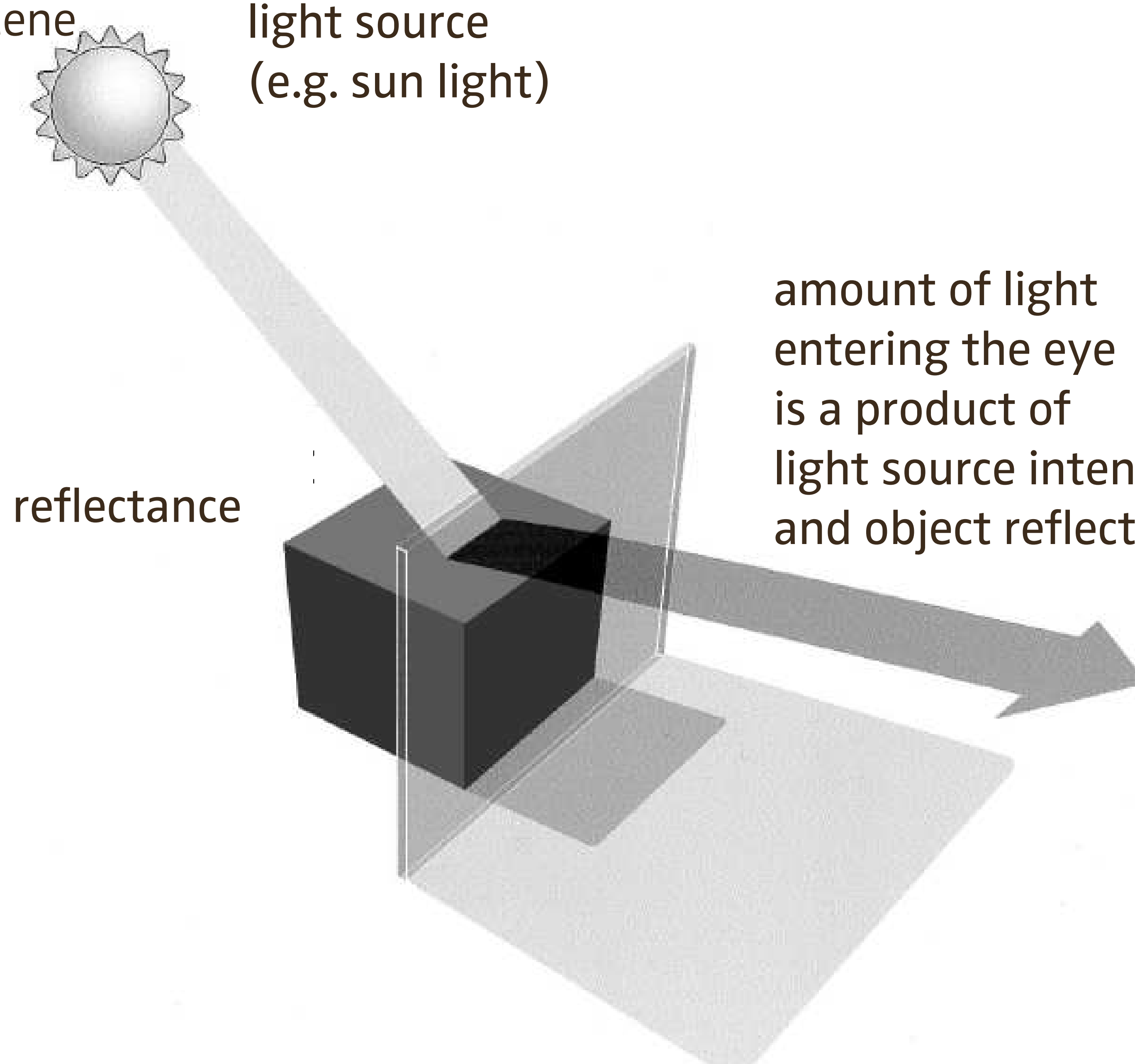
Laws of physics “generate” 2D images on
our retinae from 3D scenes
(forward optics / rendering)



One way to think about vision: inverse optics

Laws of physics “generate” 2D images on our retinae from 3D scenes (forward optics / rendering)

Starting point to think about visual perception: we want to infer the 3D scene from the 2D retinal images: inverse optics!

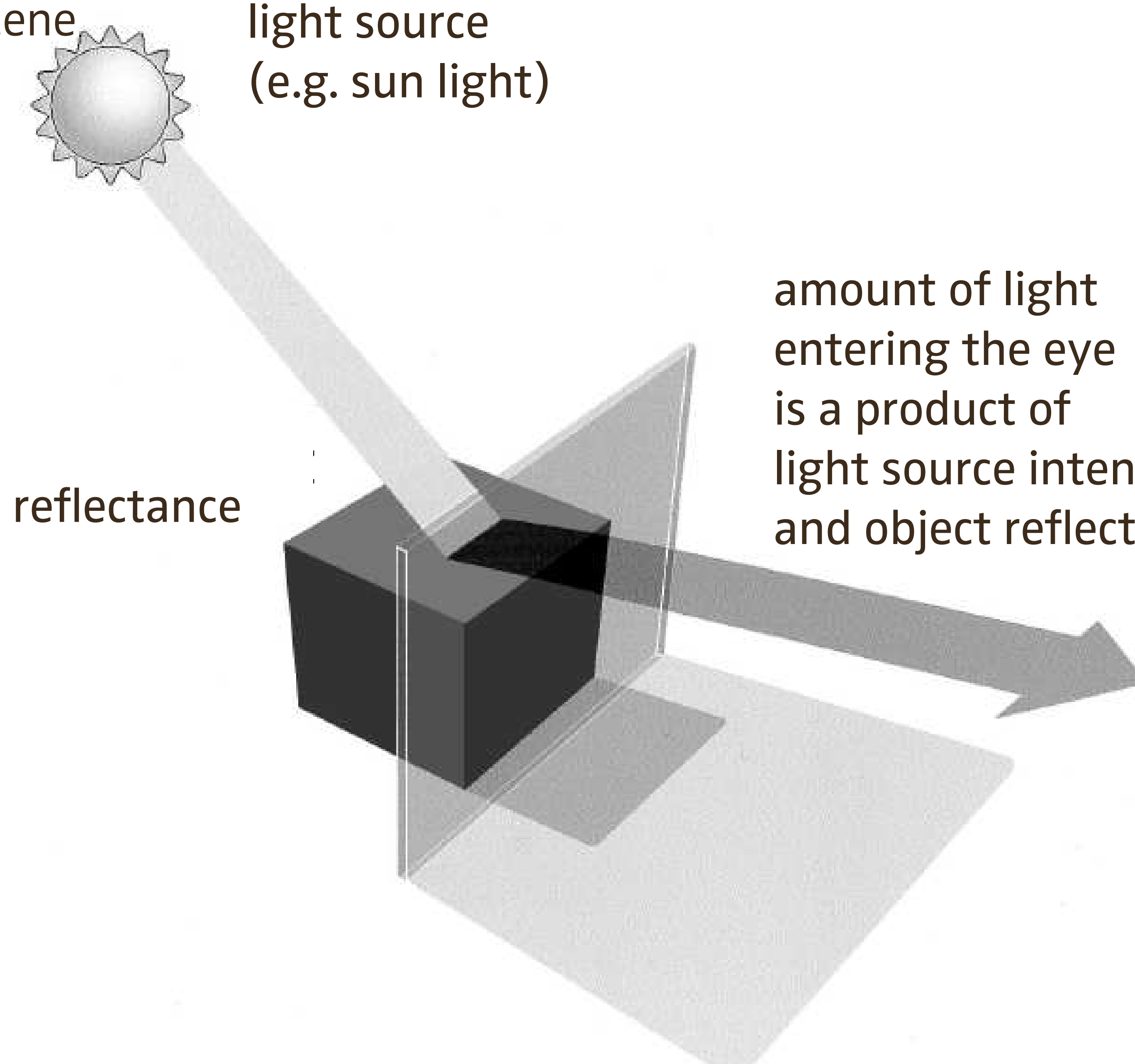


One way to think about vision: inverse optics

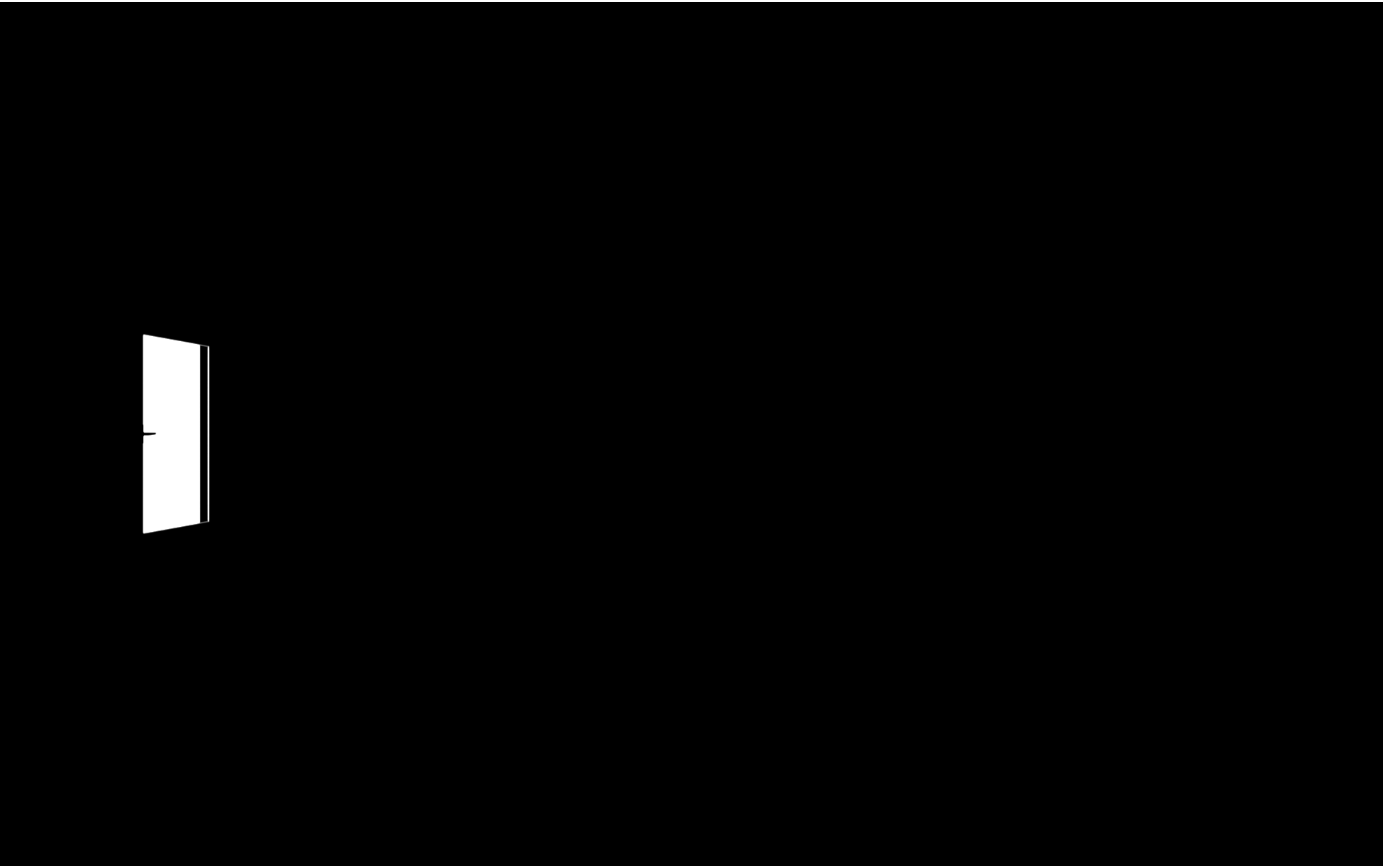
Laws of physics "generate" 2D images on our retinae from 3D scenes (forward optics / rendering)

Starting point to think about visual perception: we want to infer the 3D scene from the 2D retinal images: inverse optics!

But: Inverse optics is mathematically impossible.



N = o



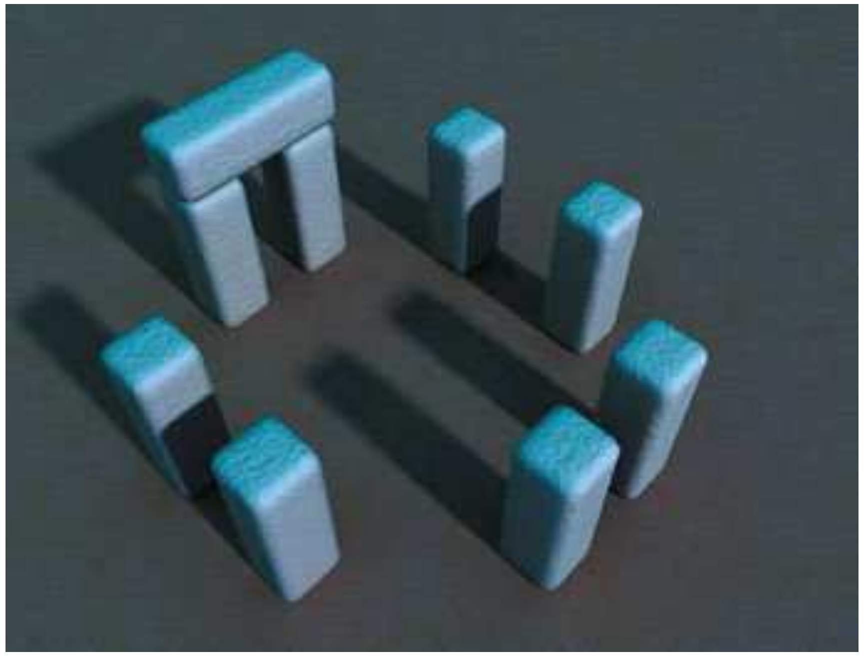
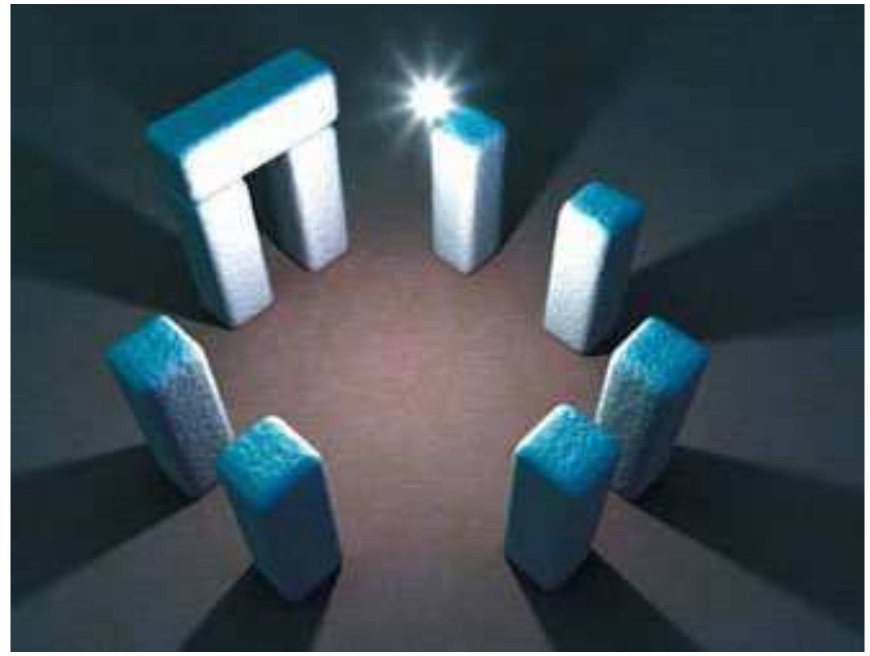
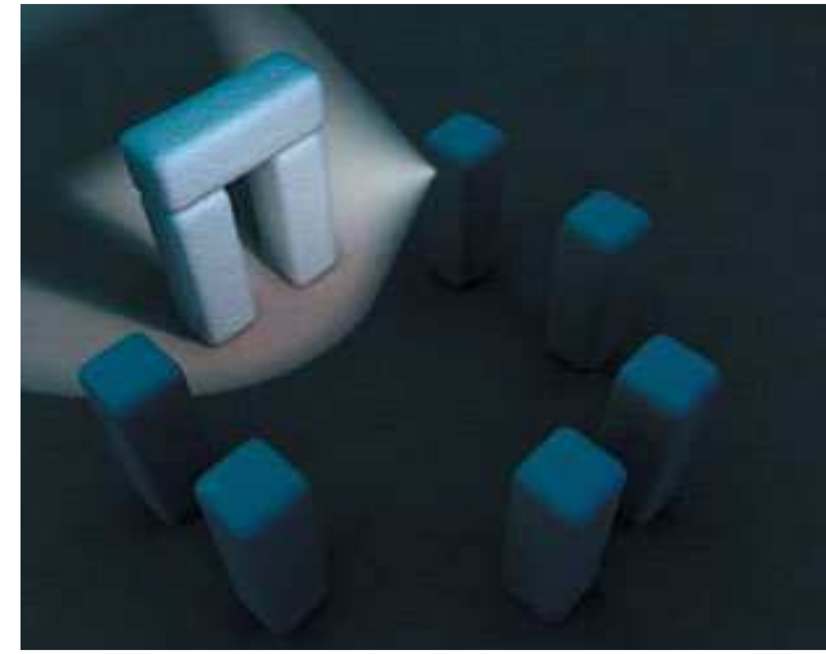
N = 1

N = 2

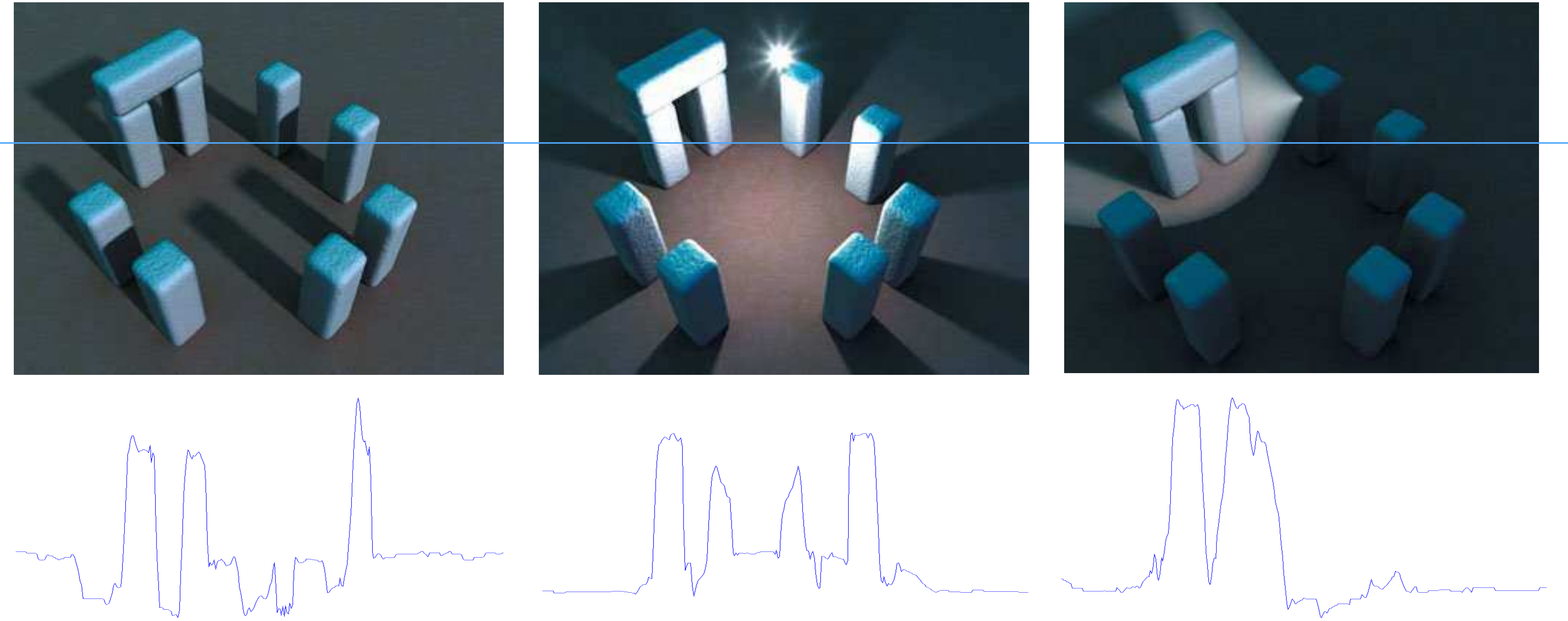
N = 5

N = 9

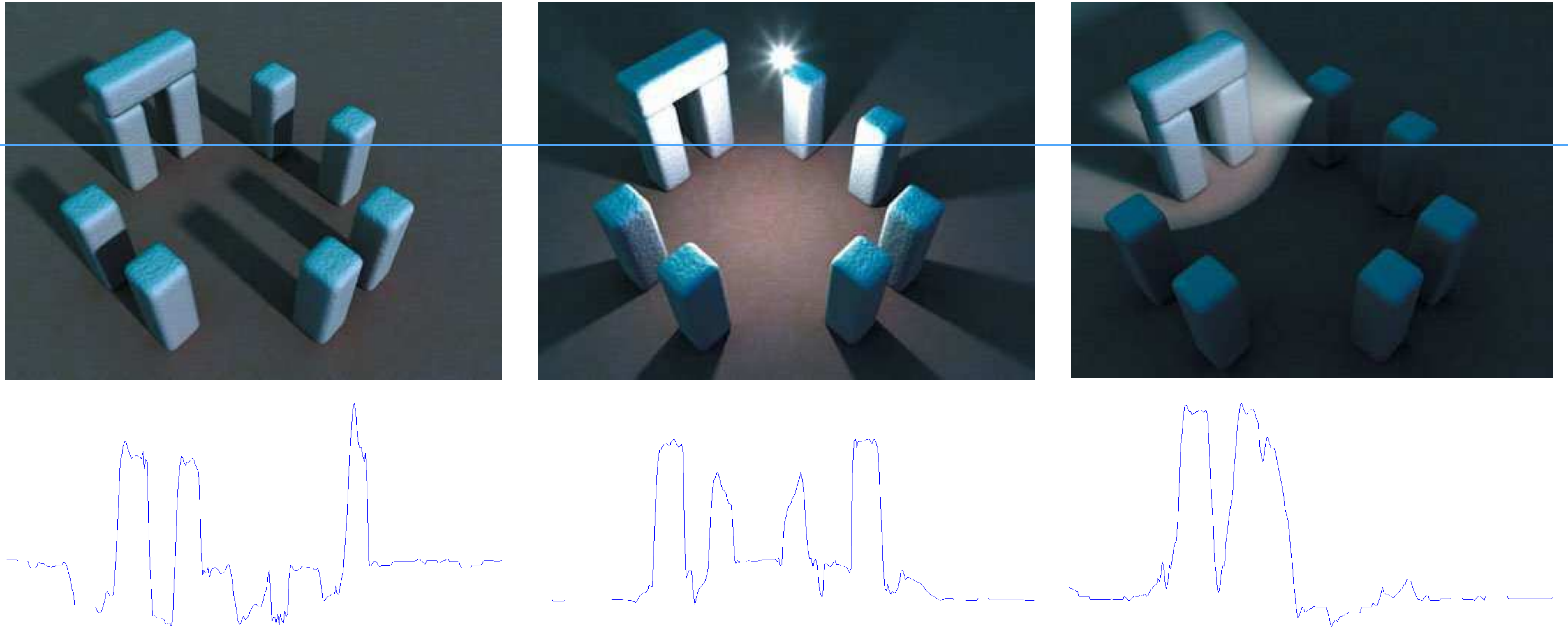
N = 24 (considered fully rendered)

modified from Matthias Bethge



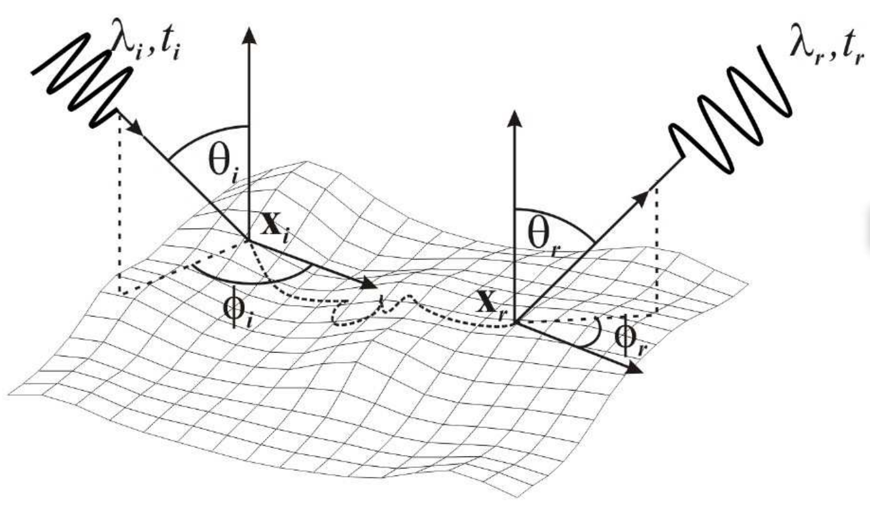
modified from Matthias Bethge

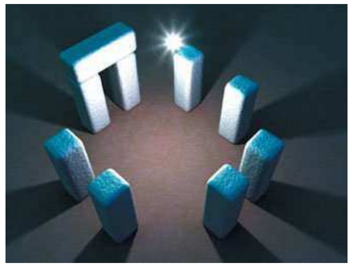


illumination
(„light field“)

objects & surfaces
(geometry, materials)

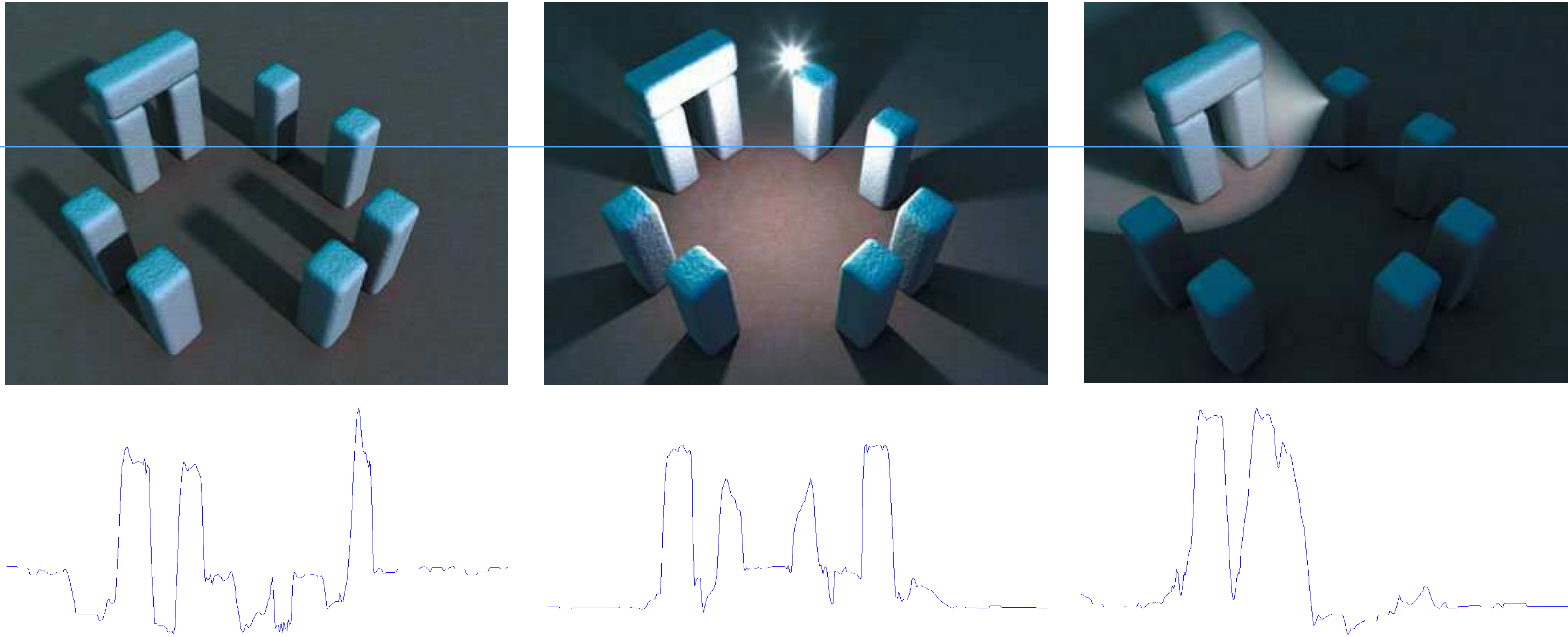
(non-linear)
„information entanglement“





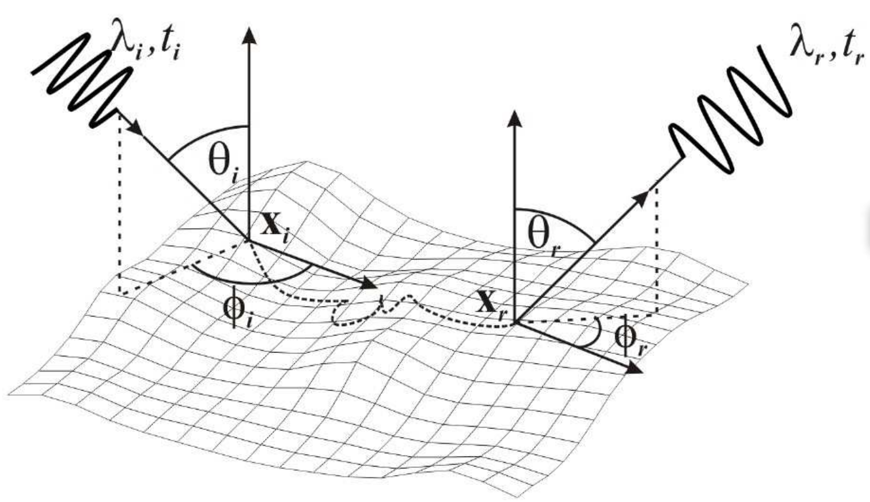
resulting
image

modified from Matthias Bethge

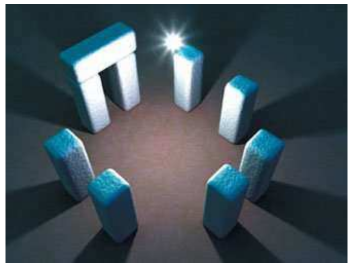


illumination
(„light field“)

(non-linear)
„information entanglement“



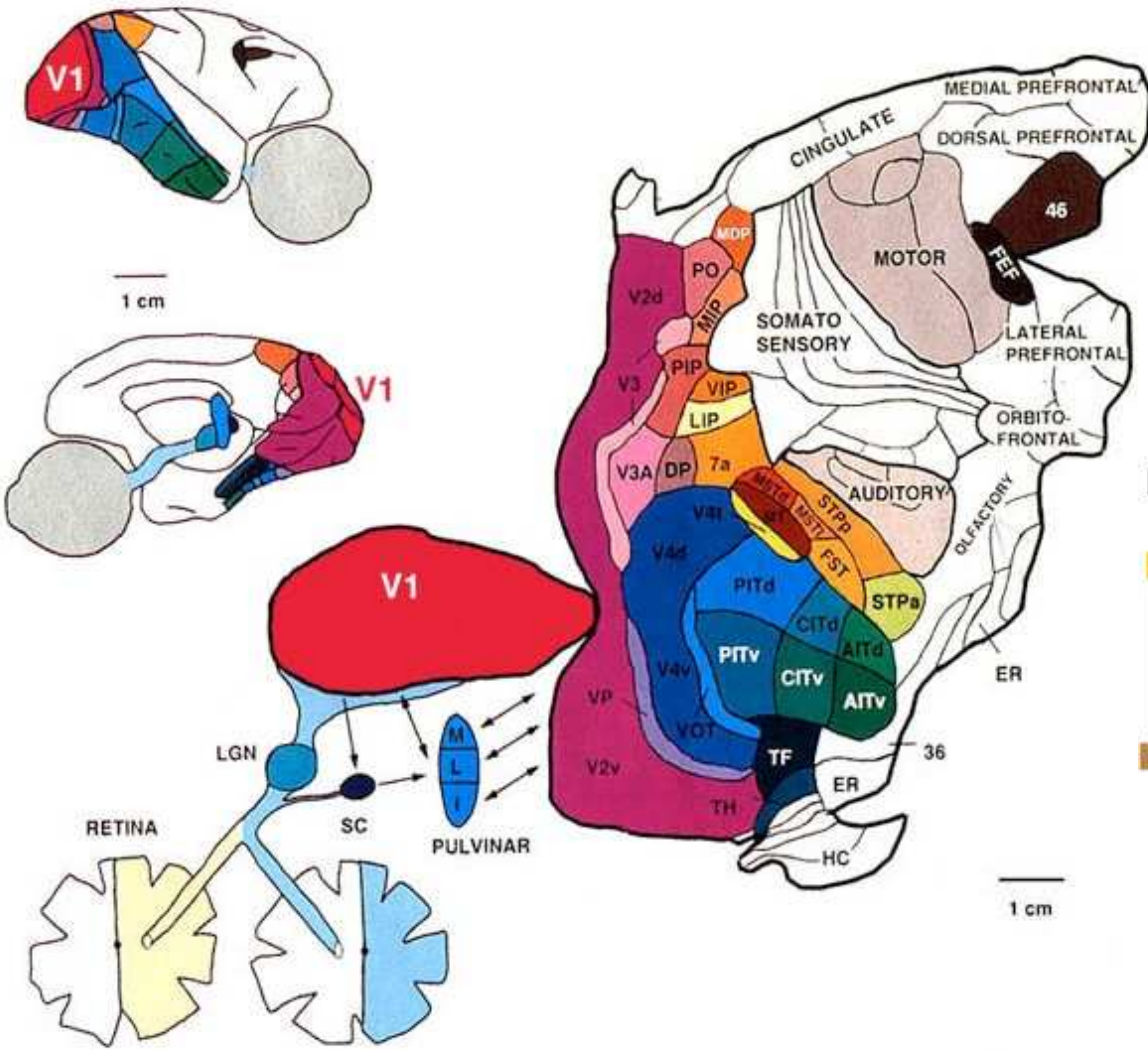
objects & surfaces
(geometry, materials)

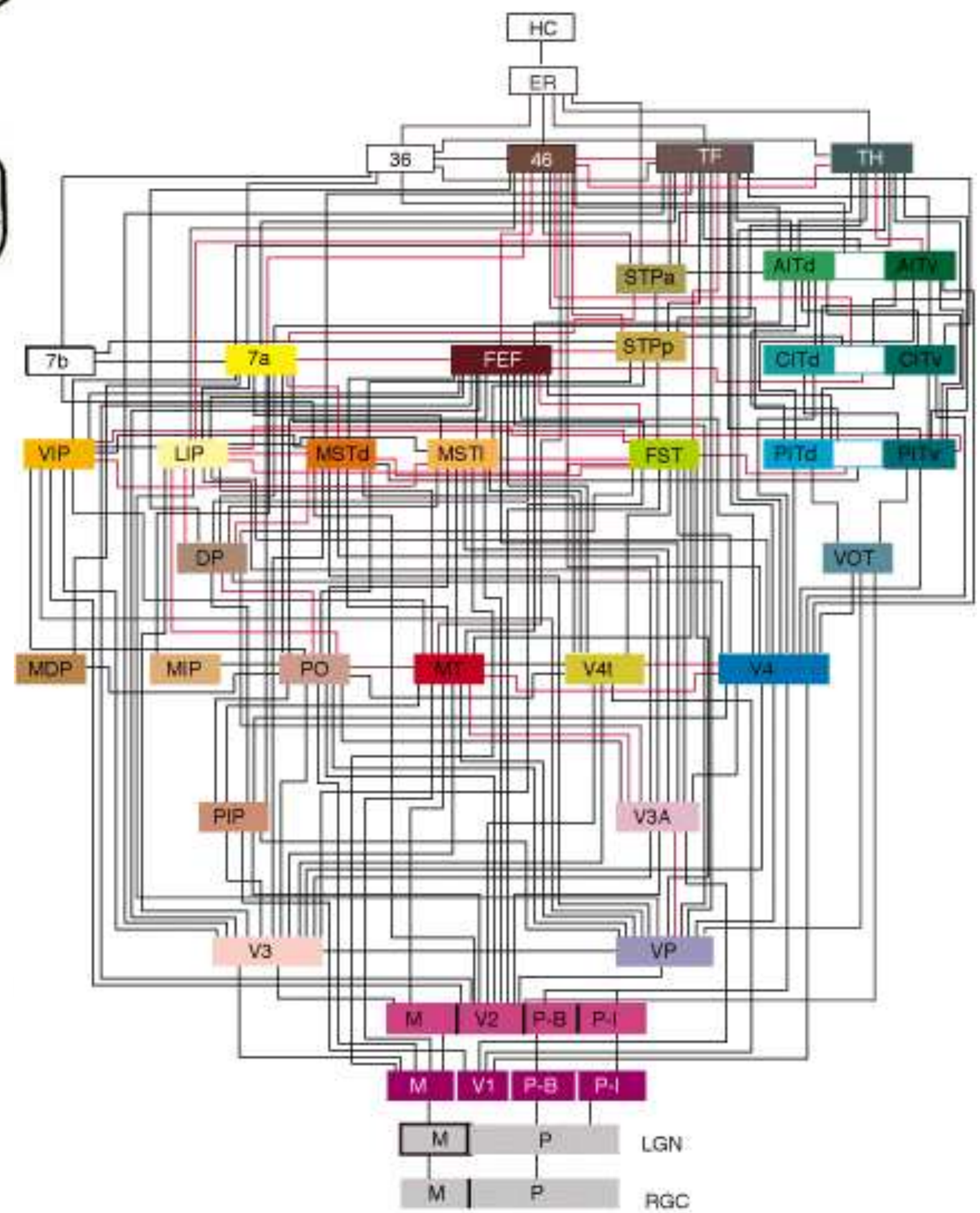


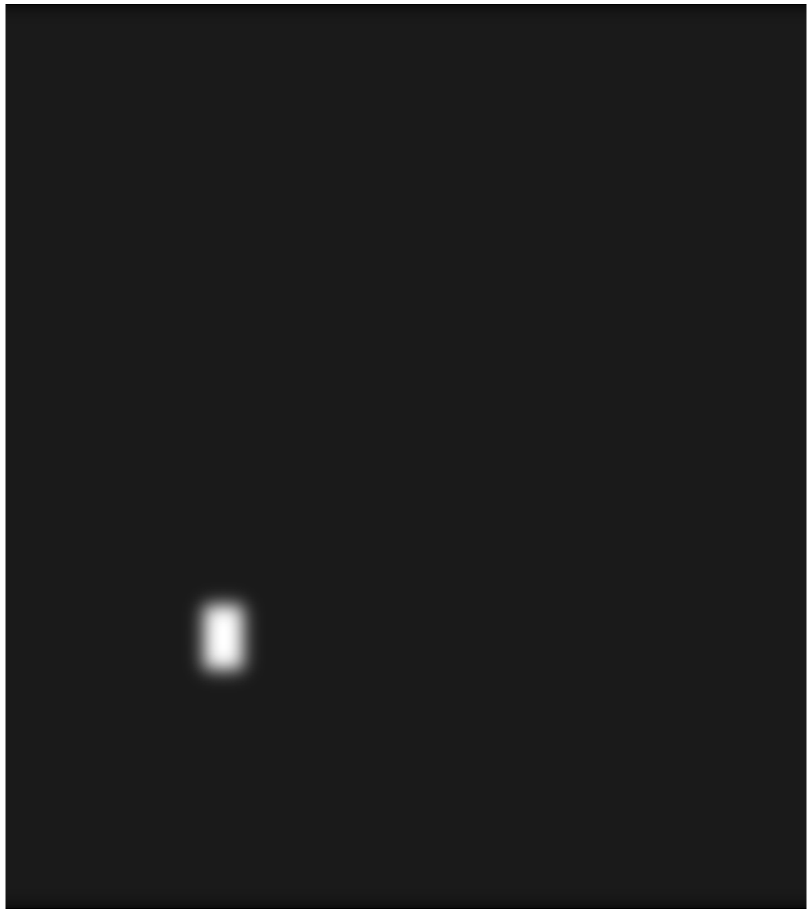
resulting
image

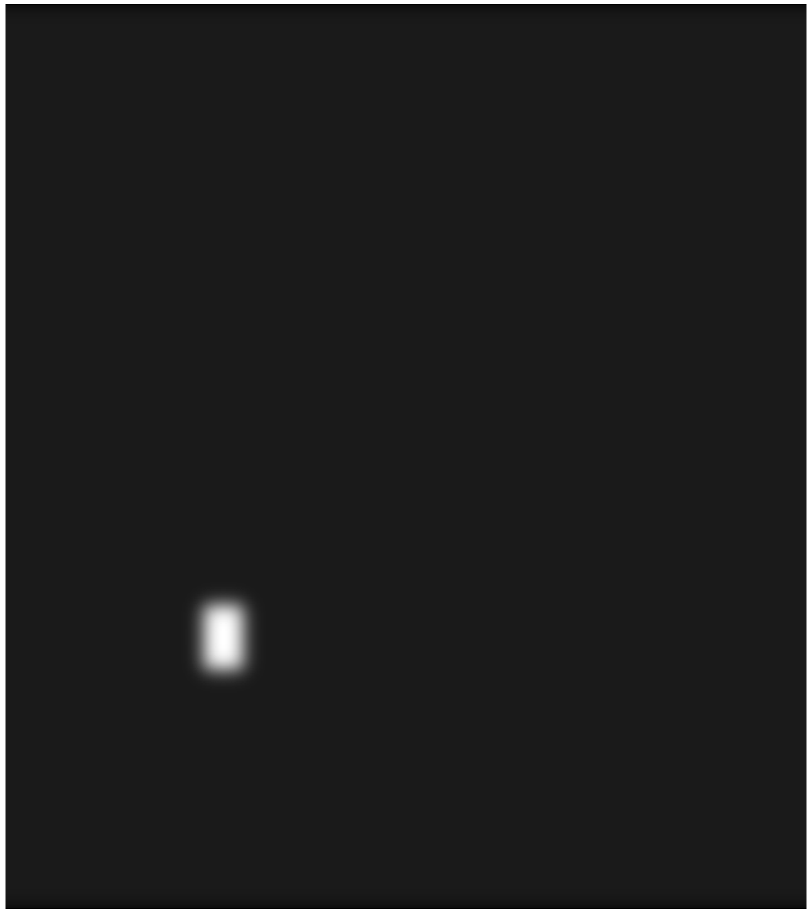
visual inference („untangling“)

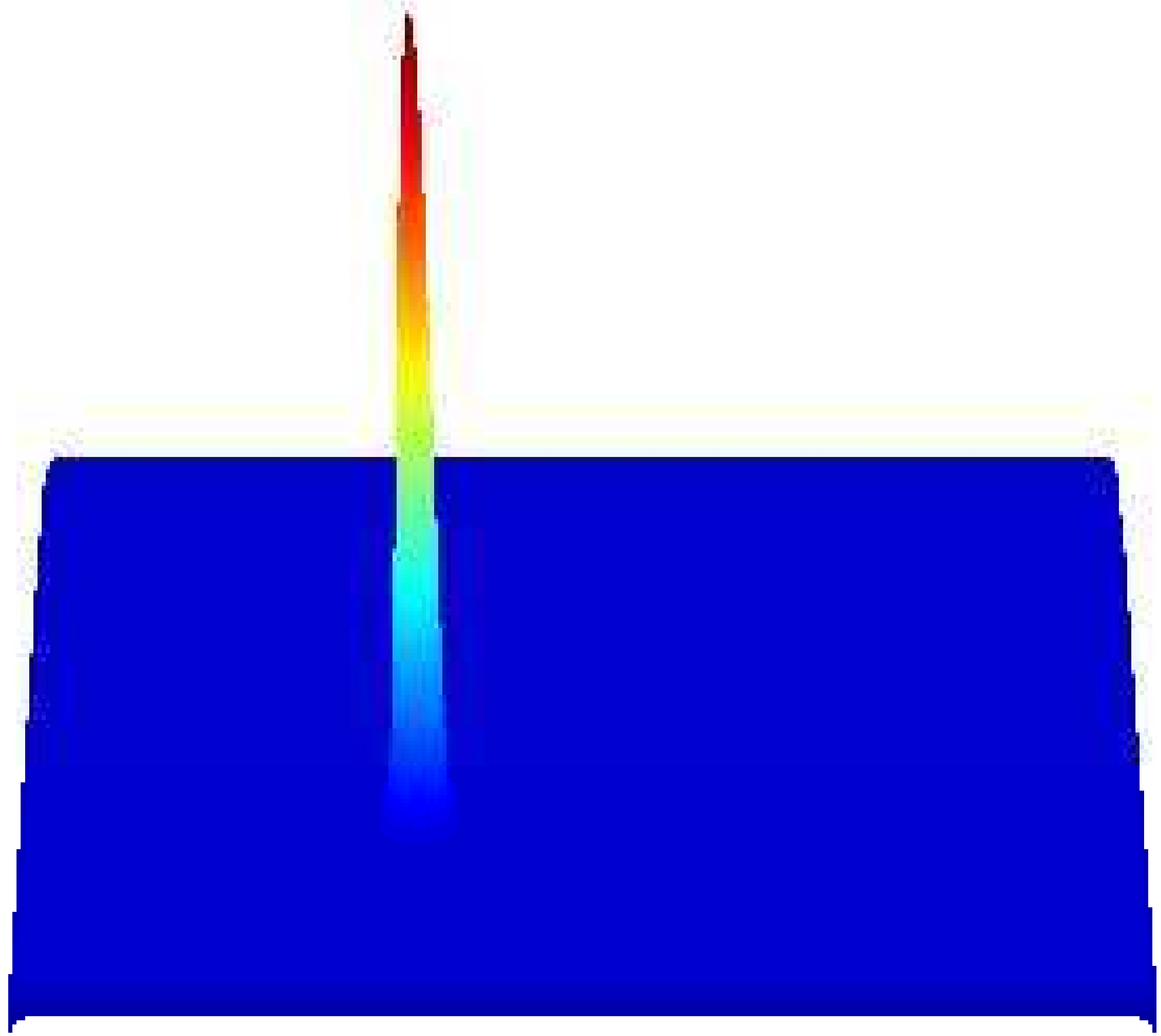
modified from Matthias Bethge

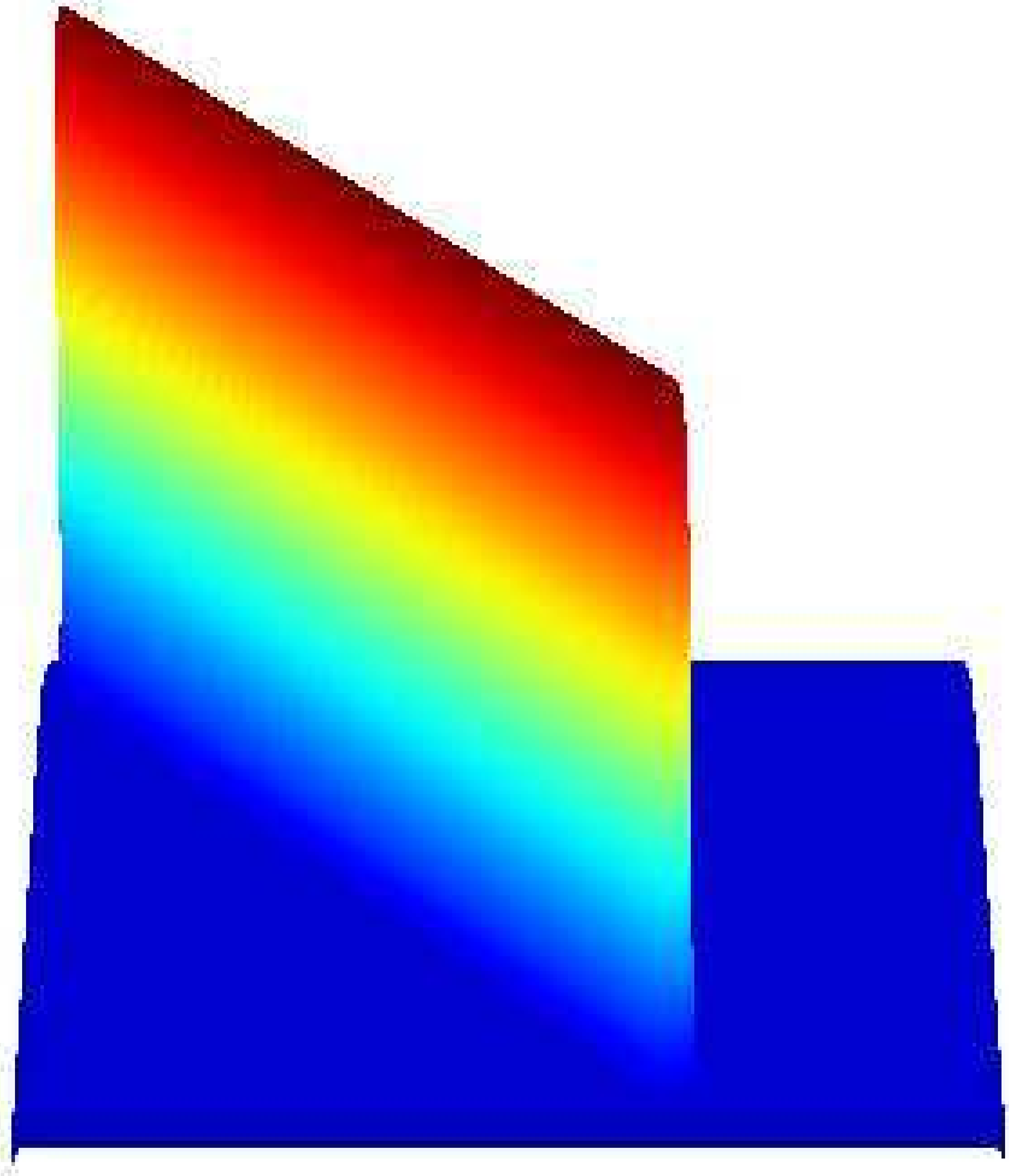

V1
1 cm
V1
V1
LGN
SC
PULVINAR
M
L
I
RETINA
MEDIAL PREFRONTAL
CINGULATE
DORSAL PREFRONTAL
46
MDP
PO
V2d
MIP
SOMATO SENSORY
MOTOR
FEF
LATERAL PREFRONTAL
PIP
V3
VIP
LIP
ORBITO-FRONTAL
V3A
DP
7a
MSTd
MSTl
STPp
AUDITORY
OLFACTORY
V4t
MT
FST
V4d
PITd
STPa
CITd
AITd
PITv
V4v
CITv
AITv
ER
VP
VOT
TF
36
V2v
ER
TH
HC
1 cm

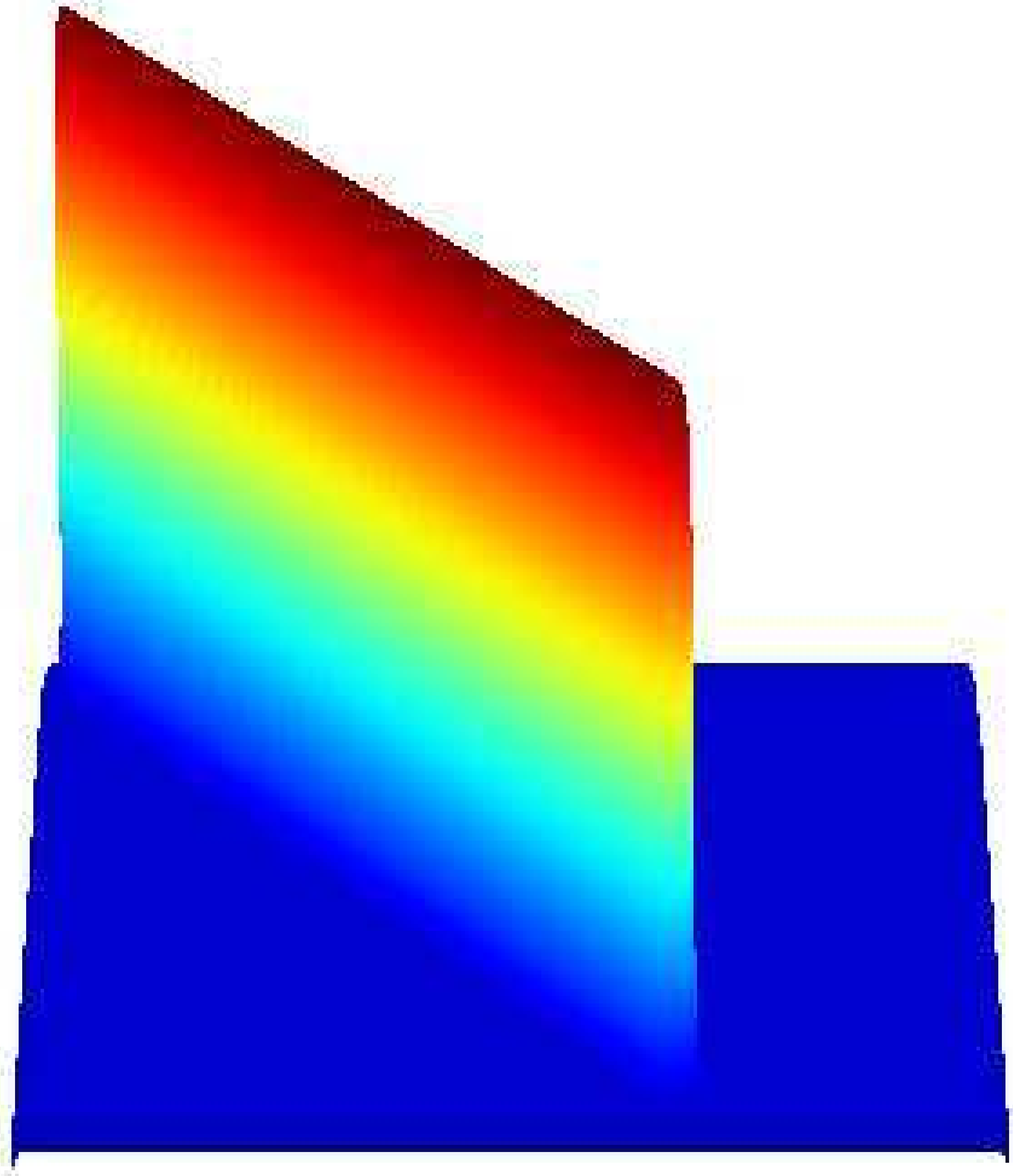

HC
ER
36
46
TF
TH
STPa
AITd
AITv
7b
7a
FEF
STPp
CITd
CITv
VIP
LIP
MSTd
MSTl
FST
PITd
PITv
DP
VOT
MDP
MIP
PO
MT
V4t
V4
PIP
V3A
V3
VP
M
V2
P-B
P-I
M
V1
P-B
P-I
M
P
LGN
M
P
RGC

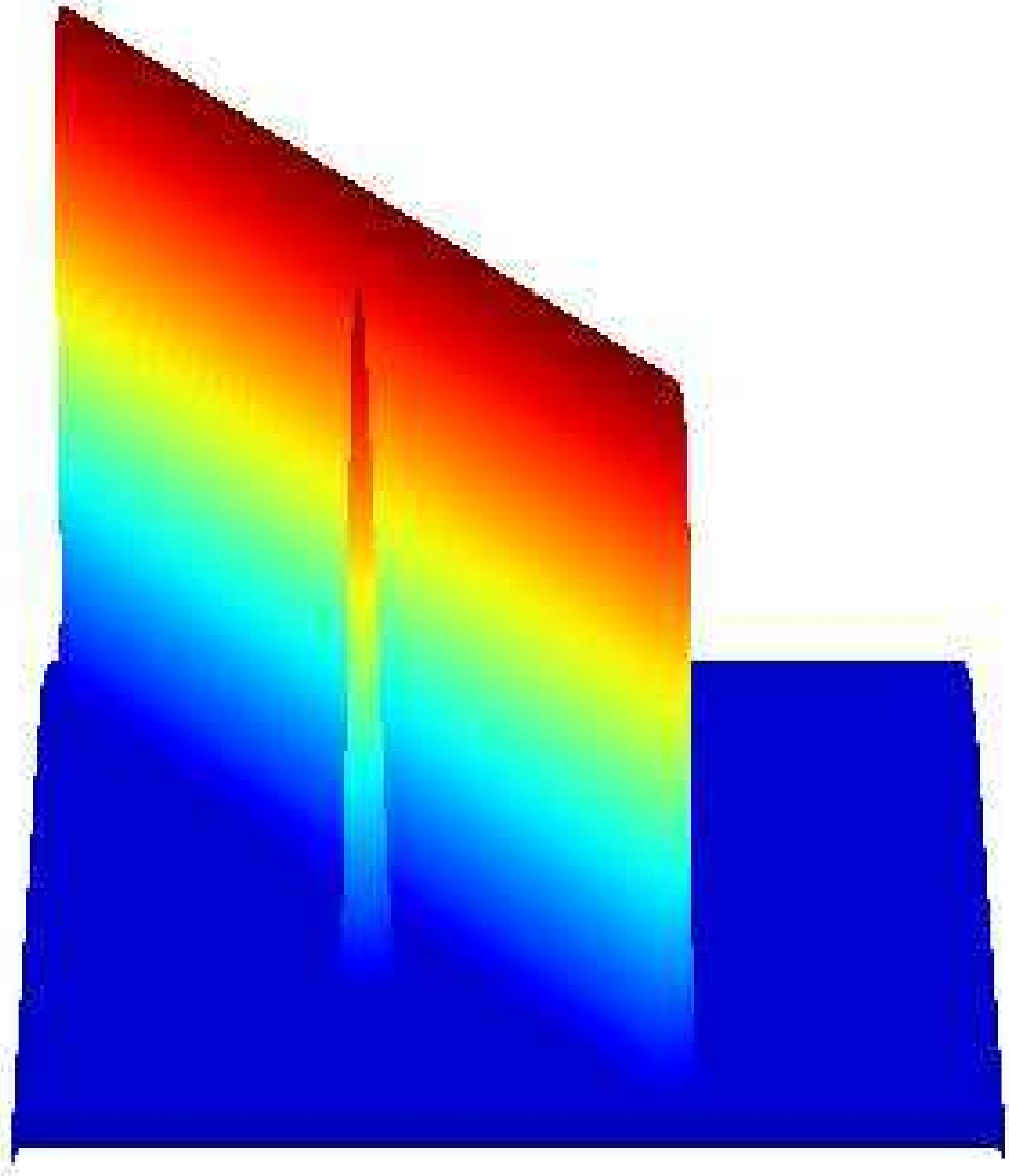


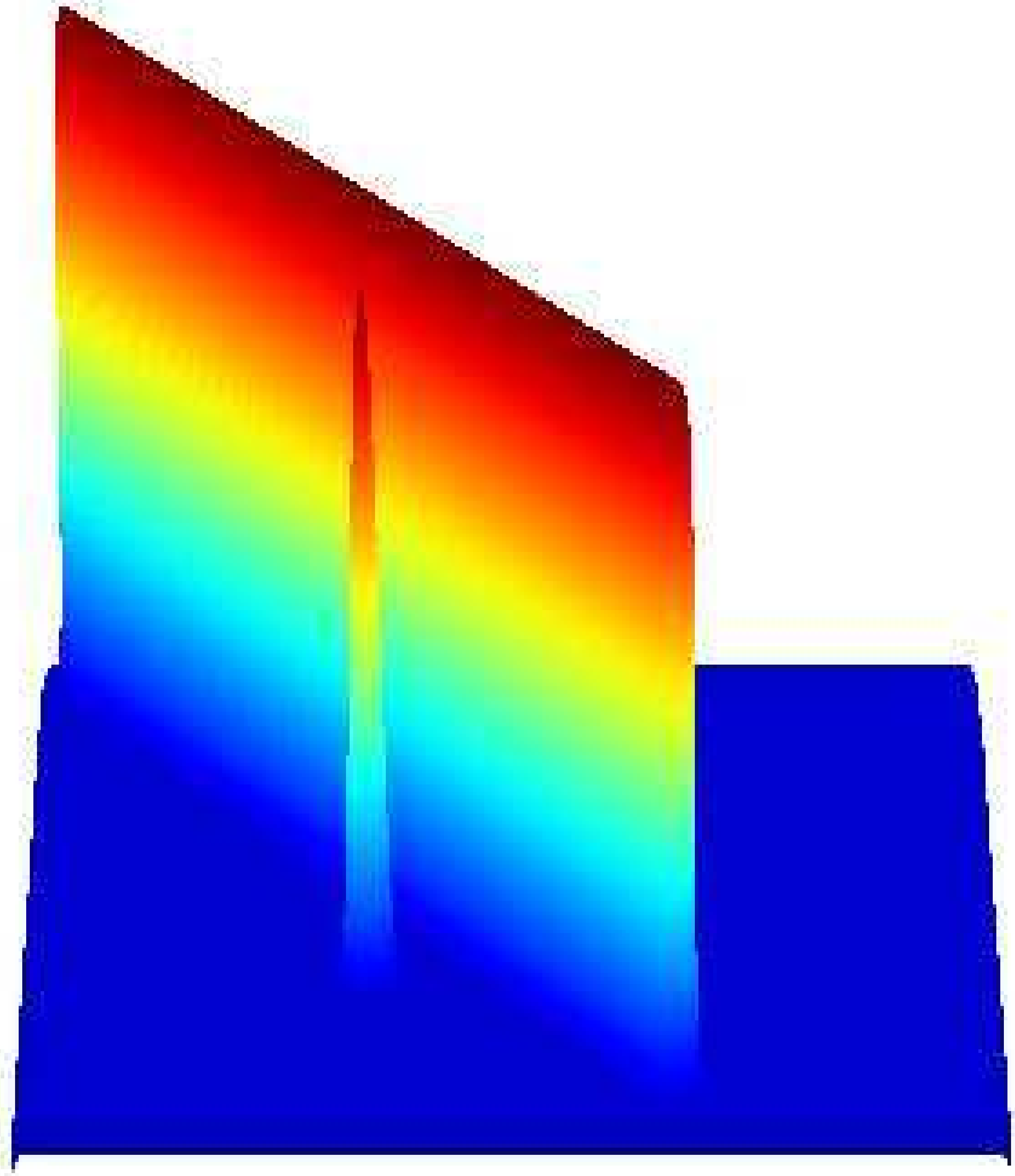


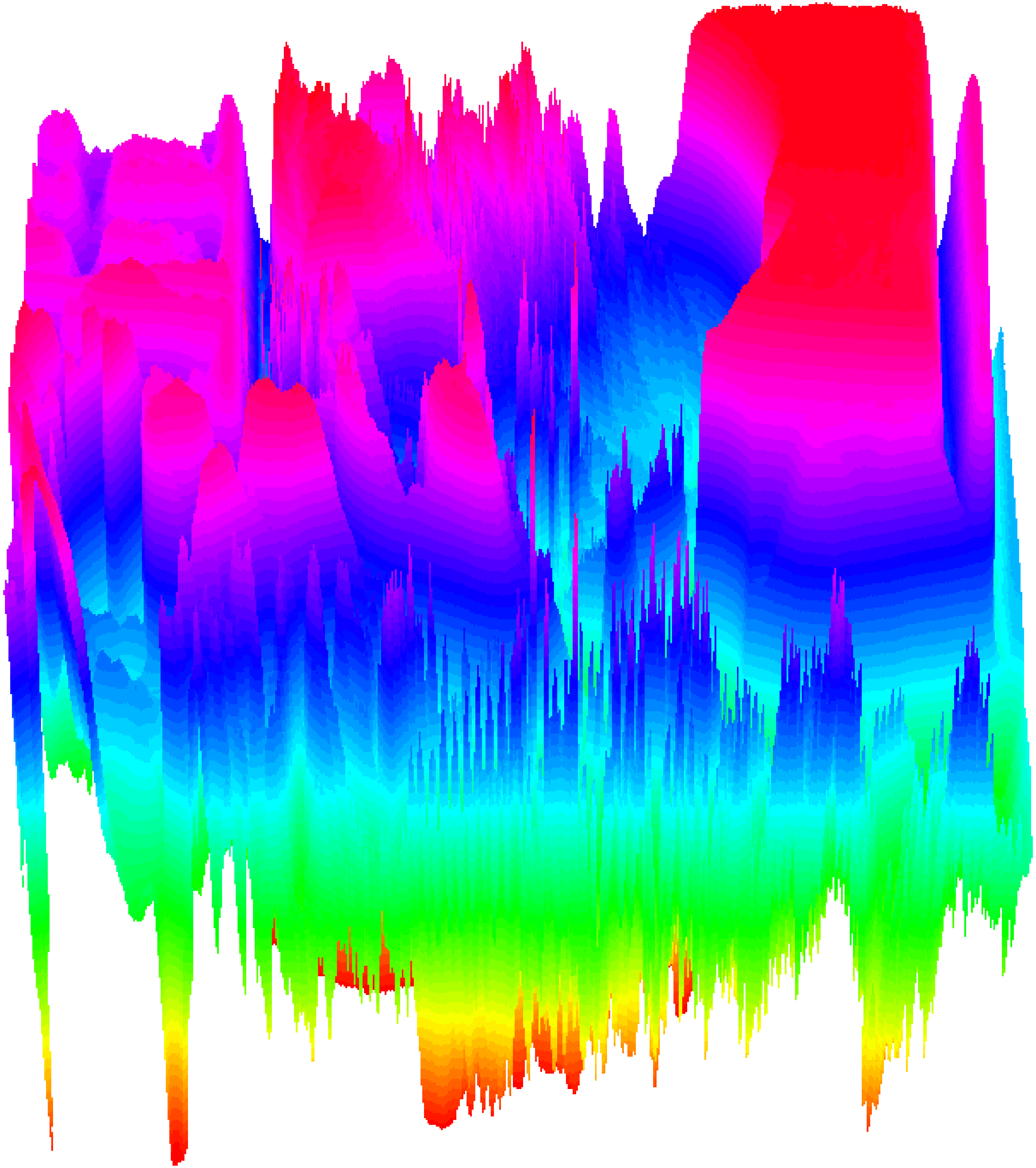


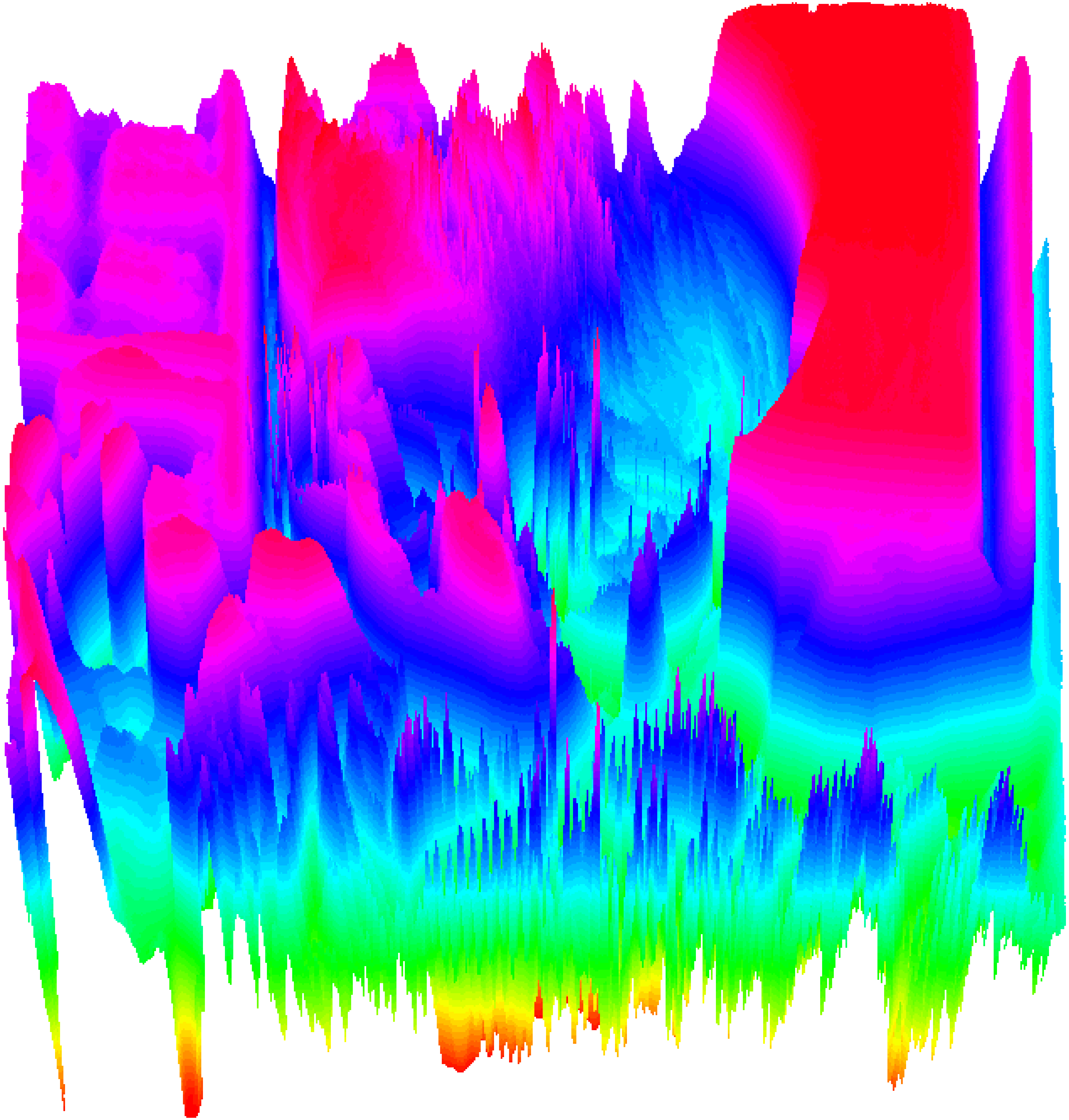








2

Machine learning (ML) and statistics

Machine learning (ML) and statistics

> *Statistics is the science of learning from data. … [ML] is the science of learning from data. These fields are identical in intent although they differ in their history, conventions, emphasis and culture.* (Wasserman, 2014)

Machine learning (ML) and statistics

> *Statistics is the science of learning from data. ... [ML] is the science of learning from data. These fields are identical in intent although they differ in their history, conventions, emphasis and culture.* (Wasserman, 2014)

ML is a comparatively new sub-branch of computational statistics jointly developed in computer science and statistics.

Machine learning (ML) and statistics

Statistics is the science of learning from data. ... [ML] is the science of learning from data. These fields are identical in intent although they differ in their history, conventions, emphasis and culture. (Wasserman, 2014)

ML is a comparatively new sub-branch of computational statistics jointly developed in computer science and statistics.

ML is inference performed by computers based on past observations and learning algorithms: ML algorithms are mainly concerned with discovering hidden structure in data in order to predict novel data—exploratory methods, to get things done!

Machine learning (ML) and statistics

Statistics is the science of learning from data. ... [ML] is the science of learning from data. These fields are identical in intent although they differ in their history, conventions, emphasis and culture. (Wasserman, 2014)

ML is a comparatively new sub-branch of computational statistics jointly developed in computer science and statistics.

ML is inference performed by computers based on past observations and learning algorithms: ML algorithms are mainly concerned with discovering hidden structure in data in order to predict novel data—exploratory methods, to get things done!

''Classical'' statistics typically is concerned with making precise probabilistic statements about known data coming from known distributions, i.e. interest in accurate models of data!

What is the difference between statistics and machine learning?

Machine Learning is AI people doing data analysis.

Data Mining is database people doing data analysis.

Applied Statistics is statisticians doing data analysis

Infographics is Graphic Designers doing data analysis.

Data Journalism is Journalists doing data analysis.

Econometrics is Economists doing data analysis
(and here you can win a Nobel Prize).

Psychometrics is Psychologists doing data analysis.

Chemometrics and Cheminformatics are Chemists doing data analysis.

Bioinformatics is Biologists doing data analysis.

Aleks Jakulin, https://www.quora.com/What-is-the-difference-between-statistics-and-machine-learning

What is the difference between statistics and machine learning? (cont'd)

> … if you look at what the goals both fields are trying to achieve, you see that there is actually quite a big difference:
>
> Statistics is interested in learning something about data, for example, which have been measured as part of some biological experiment. … . But the overall goal is to arrive at new scientific insight based on the data.
>
> In Machine Learning, the goal is to solve some complex computational task by "letting the machine learn". Instead of trying to understand the problem well enough to be able to write a program which is able to perform the task (for example, handwritten character recognition), you instead collect a huge amount of examples of what the program should do, and then run an algorithm which is able to perform the task by learning from the examples. Often, the learning algorithms are statistical in nature. But as long as the prediction works well, any kind of statistical insight into the data is not necessary.

Mikio Braun, https://www.quora.com/What-is-the-difference-between-statistics-and-machine-learning

What is the difference between statistics and machine learning? (cont'd)

The primary differences are perhaps the types of the problems attacked, and the goal of learning.

At the risk of data and models oversimplification, one could say that in statistics a prime focus is often in understanding the data and relationships in terms of models giving approximate summaries such as linear relations or independencies. In contrast, the goals in algorithms and machine learning are primarily to make predictions as accurately as possible and predictions to understand the behaviour of learning algorithms.

These differing objectives have led to different developments in the two fields: for example, neural network algorithms have been used extensively as black-box function approximators in machine learning, but to many statisticians they are less than satisfactory, because of the difficulties in interpreting such models.

Franck Dernoncourt, https://www.quora.com/What-is-the-difference-between-statistics-and-machine-learning

Terminology: types of learning

Terminology: types of learning

Supervised learning is the ML task of inferring a function from labeled training data. In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value (also called the *supervisory signal*). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for prediction.

Terminology: types of learning

Supervised learning is the ML task of inferring a function from labeled training data. In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value (also called the *supervisory signal*). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for prediction.

Reinforcement learning is an area of ML inspired by behaviorist psychology, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Unlike supervised ML correct input/output pairs are never presented, nor sub-optimal actions explicitly corrected; only global reward for an action.

Terminology: types of learning

Supervised learning is the ML task of inferring a function from labeled training data. In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value (also called the *supervisory signal*). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for prediction.

Reinforcement learning is an area of ML inspired by behaviorist psychology, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Unlike supervised ML correct input/output pairs are never presented, nor sub-optimal actions explicitly corrected; only global reward for an action.

Unsupervised learning is the ML task of inferring a function to describe hidden structure from unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. This distinguishes unsupervised learning from supervised learning and reinforcement learning. A good example is identifying close-knit groups of friends in social network data; clustering algorithms, like k-means)

Terminology: types of learning

Supervised learning is the ML task of inferring a function from labeled training data. In supervised learning, each example is a *pair* consisting of an input object (typically a vector) and a desired output value (also called the *supervisory signal*). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for prediction.

Reinforcement learning is an area of ML inspired by behaviorist psychology, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Unlike supervised ML correct input/output pairs are never presented, nor sub-optimal actions explicitly corrected; only global reward for an action.

Unsupervised learning is the ML task of inferring a function to describe hidden structure from unlabeled data. Since the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. This distinguishes unsupervised learning from supervised learning and reinforcement learning. A good example is identifying close-knit groups of friends in social network data; clustering algorithms, like k-means)

Semi-supervised learning is a class algorithms making use of unlabeled data for training—typically a small amount of labeled data with a large amount of unlabeled data. Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data).

Terminology: types of problems in **supervised** ML

Terminology: types of problems in **supervised** ML

Classification: Problems where we seek a yes-or-no prediction, such as ‘‘Is this tumour cancerous?’’, ‘‘Does this cookie meet our quality standards?’’, and so on.

Terminology: types of problems in **supervised** ML

Classification: Problems where we seek a yes-or-no prediction, such as ''Is this tumour cancerous?'', ''Does this cookie meet our quality standards?'', and so on.

Regression: Problems where the value being predicted falls somewhere on a continuous spectrum. These systems help us with questions of ''How much?'' or ''How many?''

Terminology: types of problems in **supervised** ML

Classification: Problems where we seek a yes-or-no prediction, such as "Is this tumour cancerous?", "Does this cookie meet our quality standards?", and so on.

Regression: Problems where the value being predicted falls somewhere on a continuous spectrum. These systems help us with questions of "How much?" or "How many?"

Support vector machine (SVM) is a supervised classification algorithm

Terminology: types of problems in **supervised** ML

Classification: Problems where we seek a yes-or-no prediction, such as ''Is this tumour cancerous?'', ''Does this cookie meet our quality standards?'', and so on.

Regression: Problems where the value being predicted falls somewhere on a continuous spectrum. These systems help us with questions of ''How much?'' or ''How many?''

Support vector machine (SVM) is a supervised classification algorithm

Neural networks, including the now so popular convolutional deep neural networks (DNNs), are supervised algorithms, too, typically however for multi-class classification

Success of supervised classification in ML

Success of supervised classification in ML

ML—and in particular kernel methods as well as very recently so-called deep neural networks (DNNs)—have proven successful whenever there is an abundance of empirical data but a lack of explicit knowledge how the data were generated:

Success of supervised classification in ML

ML—and in particular kernel methods as well as very recently so-called deep neural networks (DNNs)—have proven successful whenever there is an abundance of empirical data but a lack of explicit knowledge how the data were generated:

- Predict credit card fraud from patterns of money withdrawals.

Success of supervised classification in ML

ML—and in particular kernel methods as well as very recently so-called deep neural networks (DNNs)—have proven successful whenever there is an abundance of empirical data but a lack of explicit knowledge how the data were generated:

- Predict credit card fraud from patterns of money withdrawals.
- Predict toxicity of novel substances (biomedical research).

Success of supervised classification in ML

ML—and in particular kernel methods as well as very recently so-called deep neural networks (DNNs)—have proven successful whenever there is an abundance of empirical data but a lack of explicit knowledge how the data were generated:

- Predict credit card fraud from patterns of money withdrawals.
- Predict toxicity of novel substances (biomedical research).
- Predict engine failure in airplanes.

Success of supervised classification in ML

ML—and in particular kernel methods as well as very recently so-called deep neural networks (DNNs)—have proven successful whenever there is an abundance of empirical data but a lack of explicit knowledge how the data were generated:

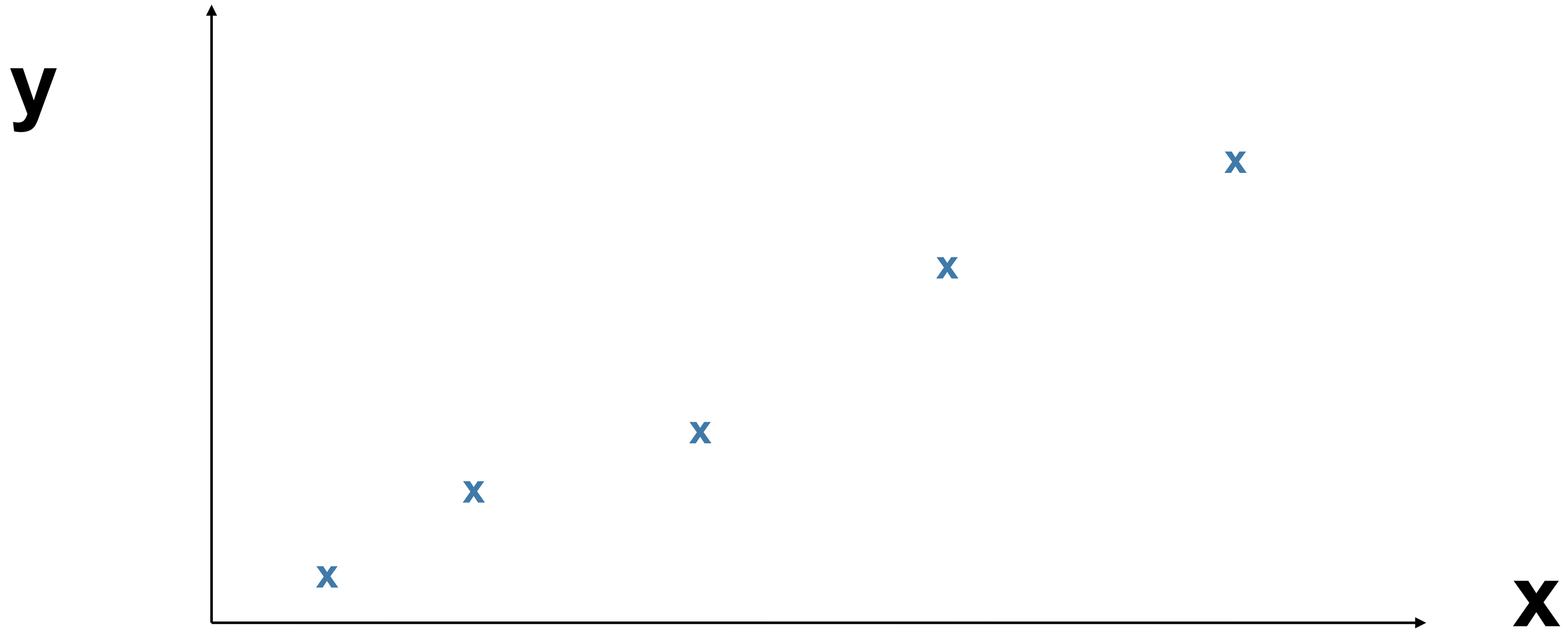
- Predict credit card fraud from patterns of money withdrawals.
- Predict toxicity of novel substances (biomedical research).
- Predict engine failure in airplanes.
- Predict what people will google next.

Success of supervised classification in ML

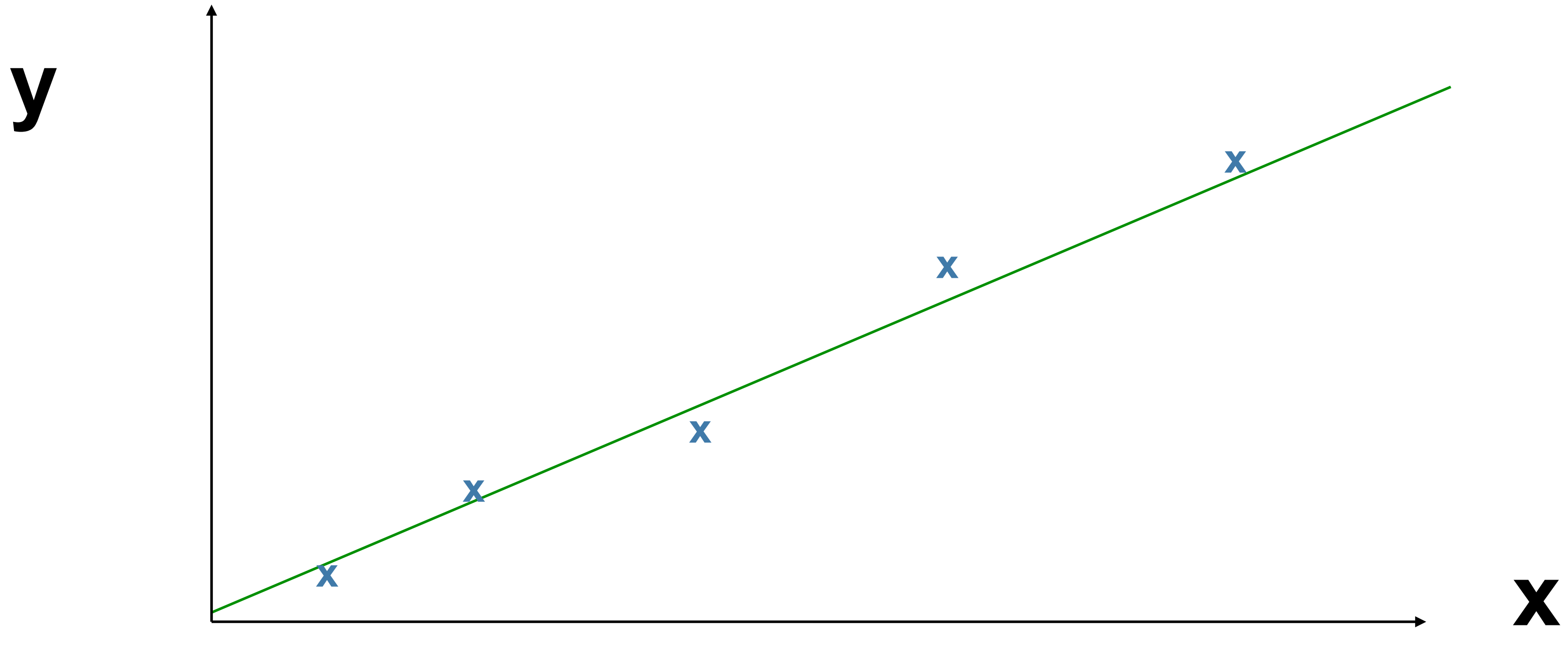
ML—and in particular kernel methods as well as very recently so-called deep neural networks (DNNs)—have proven successful whenever there is an abundance of empirical data but a lack of explicit knowledge how the data were generated:

- Predict credit card fraud from patterns of money withdrawals.
- Predict toxicity of novel substances (biomedical research).
- Predict engine failure in airplanes.
- Predict what people will google next.
- Predict what people want to buy next at amazon.

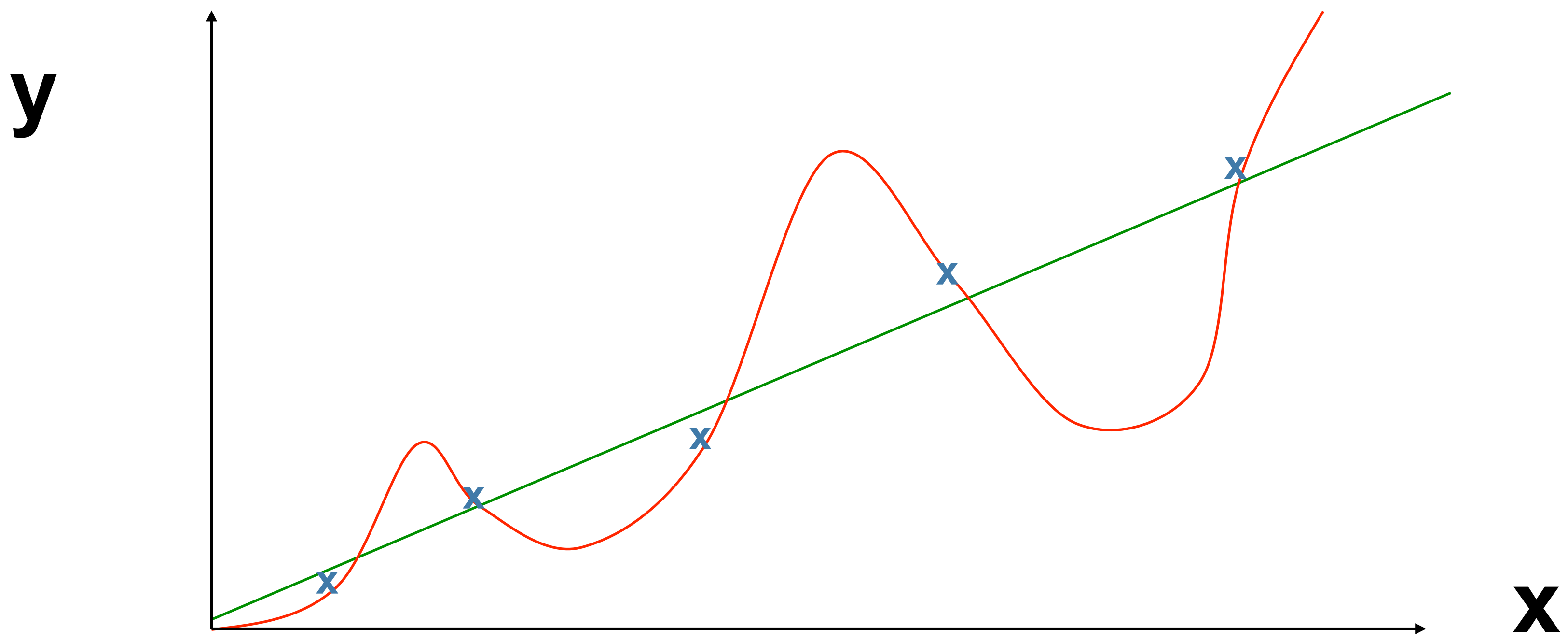
The Function Learning Problem



The Function Learning Problem



The Function Learning Problem



Learning Problem in General

Learning Problem in General

Training examples $(x_1,y_1),\ldots,(x_m,y_m)$

Learning Problem in General

Training examples $(x_1,y_1),\ldots,(x_m,y_m)$

Task: given a new x, find the new y

strong emphasis on prediction, that is, generalization!

Learning Problem in General

Training examples $(x_1,y_1),\ldots,(x_m,y_m)$

Task: given a new x, find the new y
strong emphasis on prediction, that is, generalization!

Idea: (x,y) should look "similar" to the training examples

Learning Problem in General

Training examples $(x_1,y_1),\ldots,(x_m,y_m)$

Task: given a new x, find the new y

strong emphasis on prediction, that is, generalization!

Idea: (x,y) should look "similar" to the training examples

Required: similarity measure for (x,y)

Learning Problem in General

Training examples $(x_1,y_1),\ldots,(x_m,y_m)$

Task: given a new x, find the new y
strong emphasis on prediction, that is, generalization!

Idea: (x,y) should look "similar" to the training examples

Required: similarity measure for (x,y)

Much of creativity and difficulty in kernel-based ML: Find suitable similarity measures for all the practical problems discussed before, e.g. credit card fraud, toxicity of novel molecules, gene sequences, … .

Learning Problem in General

Training examples $(x_1,y_1),\ldots,(x_m,y_m)$

Task: given a new x, find the new y
strong emphasis on prediction, that is, generalization!

Idea: (x,y) should look "similar" to the training examples

Required: similarity measure for (x,y)

Much of creativity and difficulty in kernel-based ML: Find suitable similarity measures for all the practical problems discussed before, e.g. credit card fraud, toxicity of novel molecules, gene sequences, … .

When are two molecules, with different atoms, structure, configuration etc. the same? When are two strings of letters or sentences similar? What would be the mean, or the variance of strings? Of molecules?

Learning Problem in General

Training examples $(x_1,y_1),\dots,(x_m,y_m)$

Task: given a new x, find the new y
strong emphasis on prediction, that is, generalization!

Idea: (x,y) should look "similar" to the training examples

Required: similarity measure for (x,y)

Much of creativity and difficulty in kernel-based ML: Find suitable similarity measures for all the practical problems discussed before, e.g. credit card fraud, toxicity of novel molecules, gene sequences, … .

When are two molecules, with different atoms, structure, configuration etc. the same? When are two strings of letters or sentences similar? What would be the mean, or the variance of strings? Of molecules?

Very recent deep neural network success:
The network learns the right similarity measure from the data!

The Support Vector Machine

The Support Vector Machine

Computer algorithm that learns by example to assign labels to objects

The Support Vector Machine

Computer algorithm that learns by example to assign labels to objects

Successful in handwritten digit recognition, credit card fraud detection, classification of gene expression profiles etc.

The Support Vector Machine

Computer algorithm that learns by example to assign labels to objects

Successful in handwritten digit recognition, credit card fraud detection, classification of gene expression profiles etc.

Essence of the SVM algorithm requires understanding of:

i. the separating hyperplane
ii. the maximum-margin hyperplane
iii. the soft margin
iv. the kernel function

The Support Vector Machine

Computer algorithm that learns by example to assign labels to objects

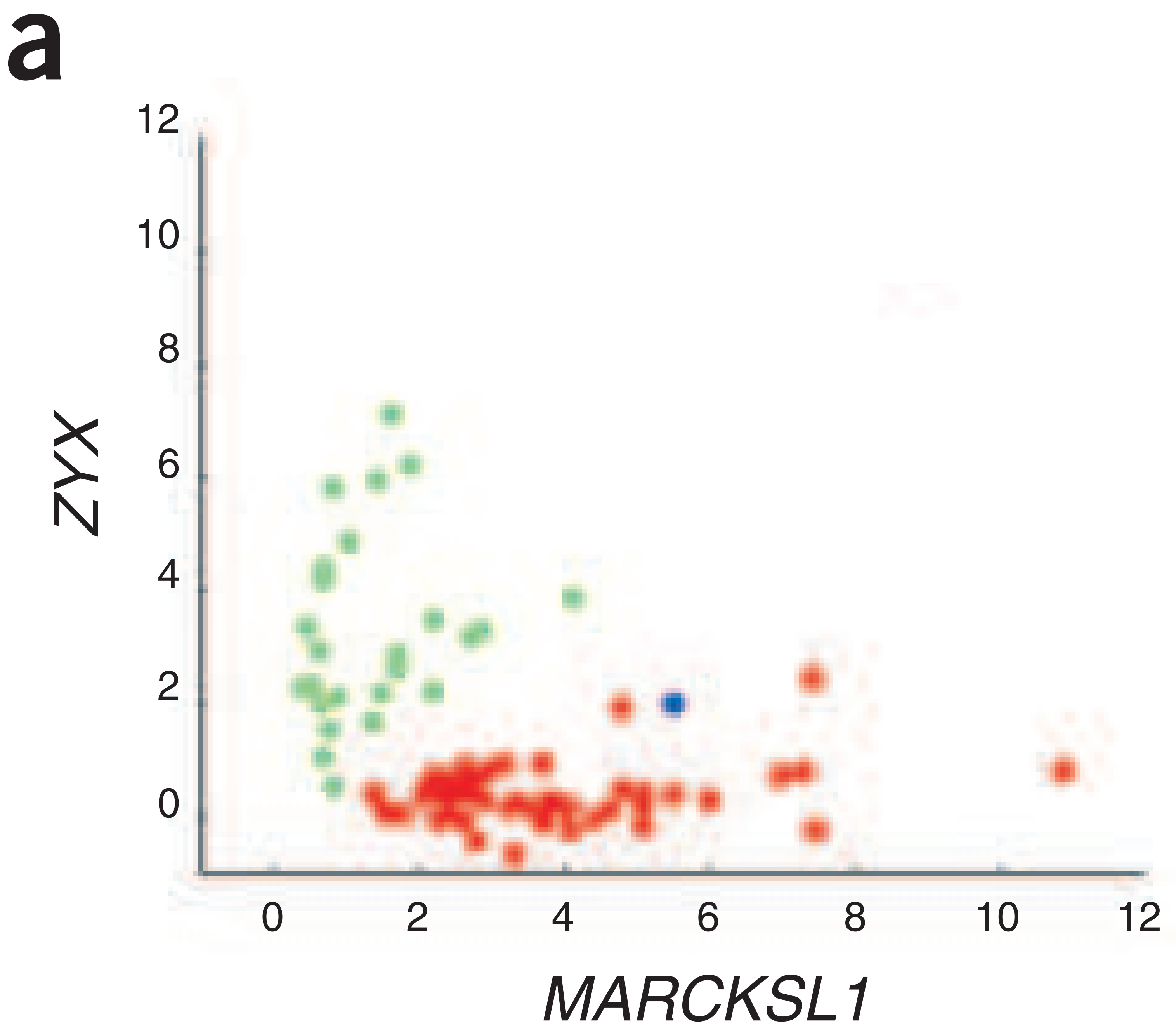
Successful in handwritten digit recognition, credit card fraud detection, classification of gene expression profiles etc.

Essence of the SVM algorithm requires understanding of:

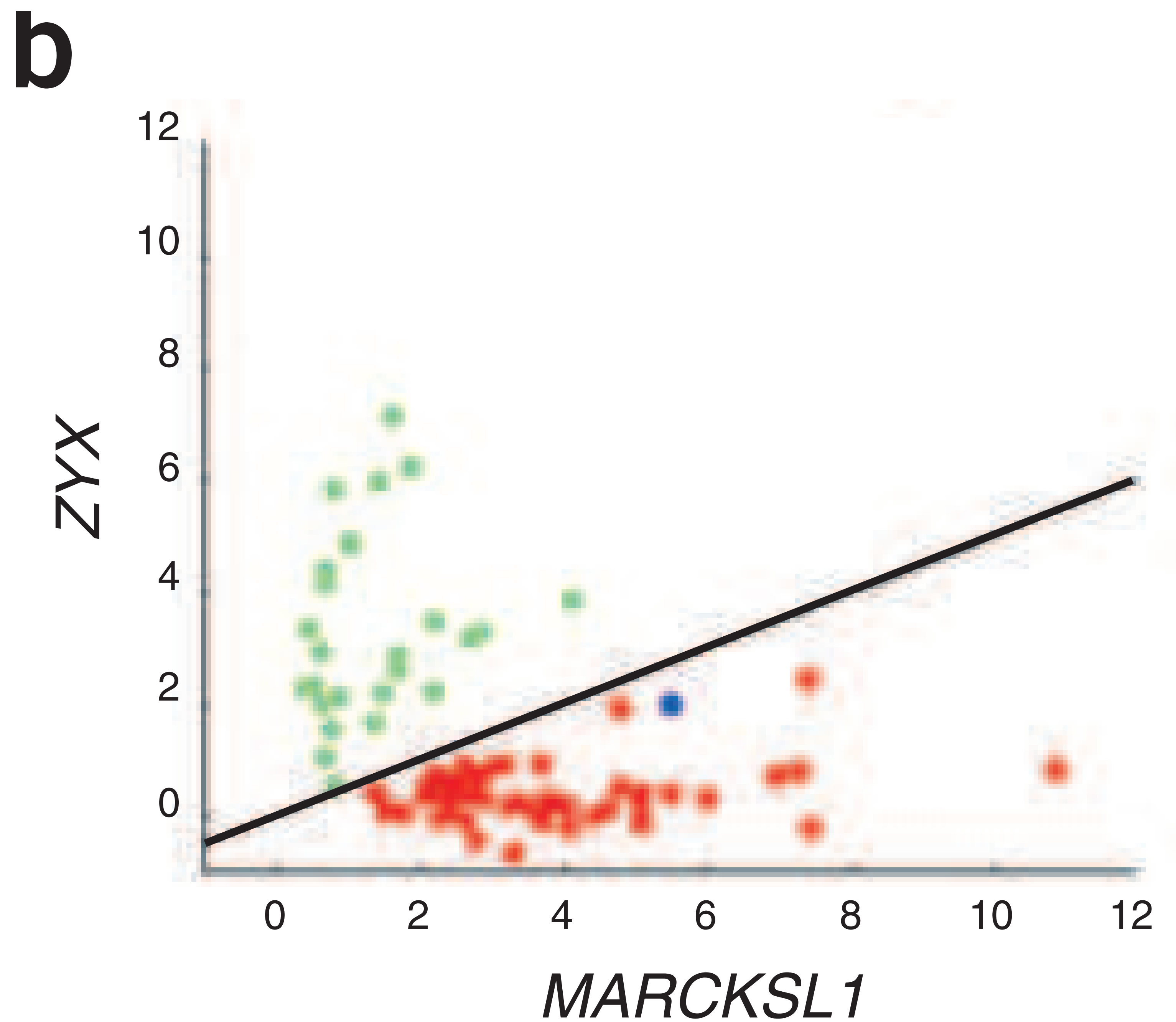
i. the separating hyperplane
ii. the maximum-margin hyperplane
iii. the soft margin
iv. the kernel function

For SVMs and machine learning in general:
i. regularisation
ii. cross-validation

Two Genes and Two Forms of Leukemia
(microarrays deliver thousands of genes, but hard to draw ...)

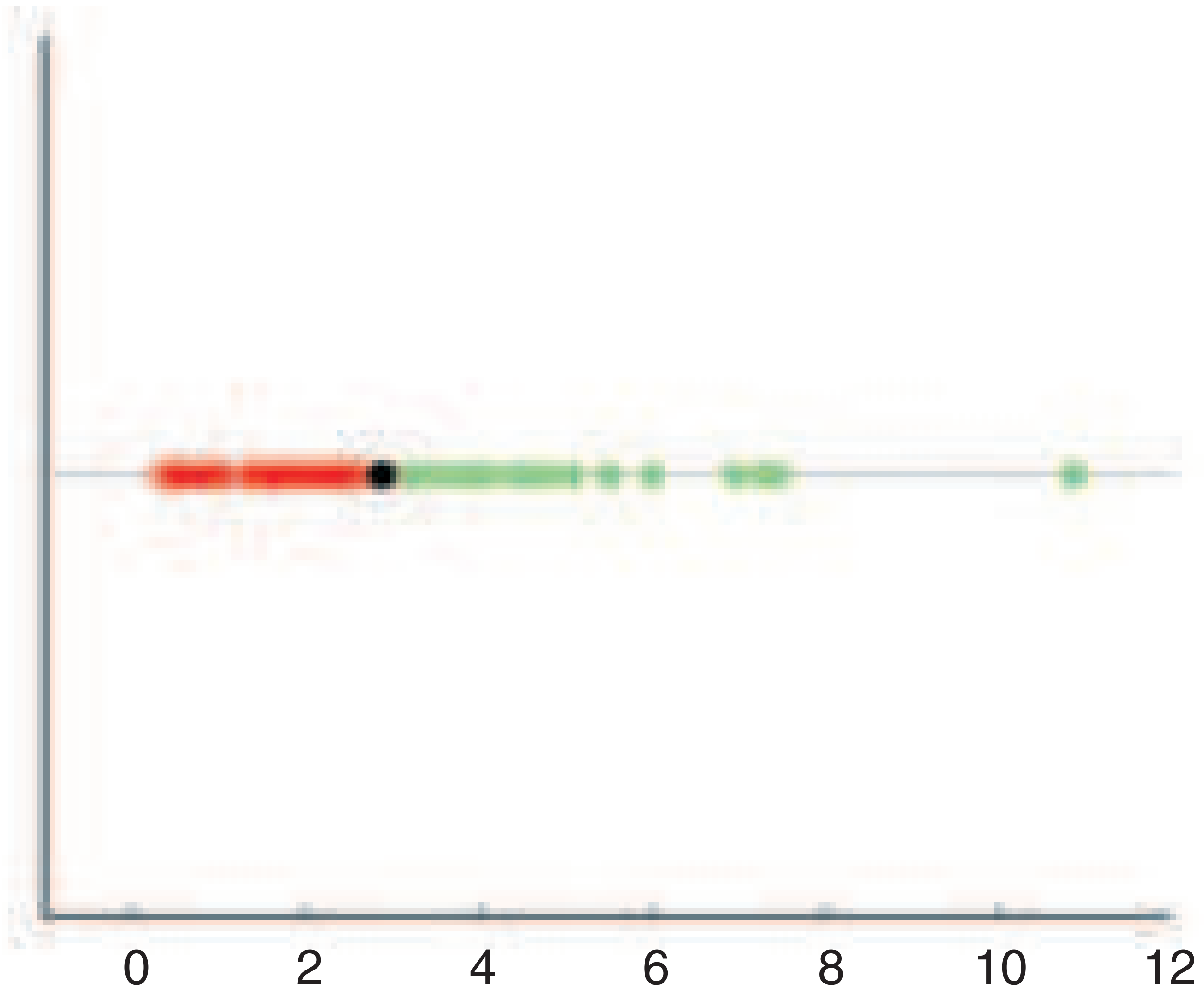


Separating Hyperplane

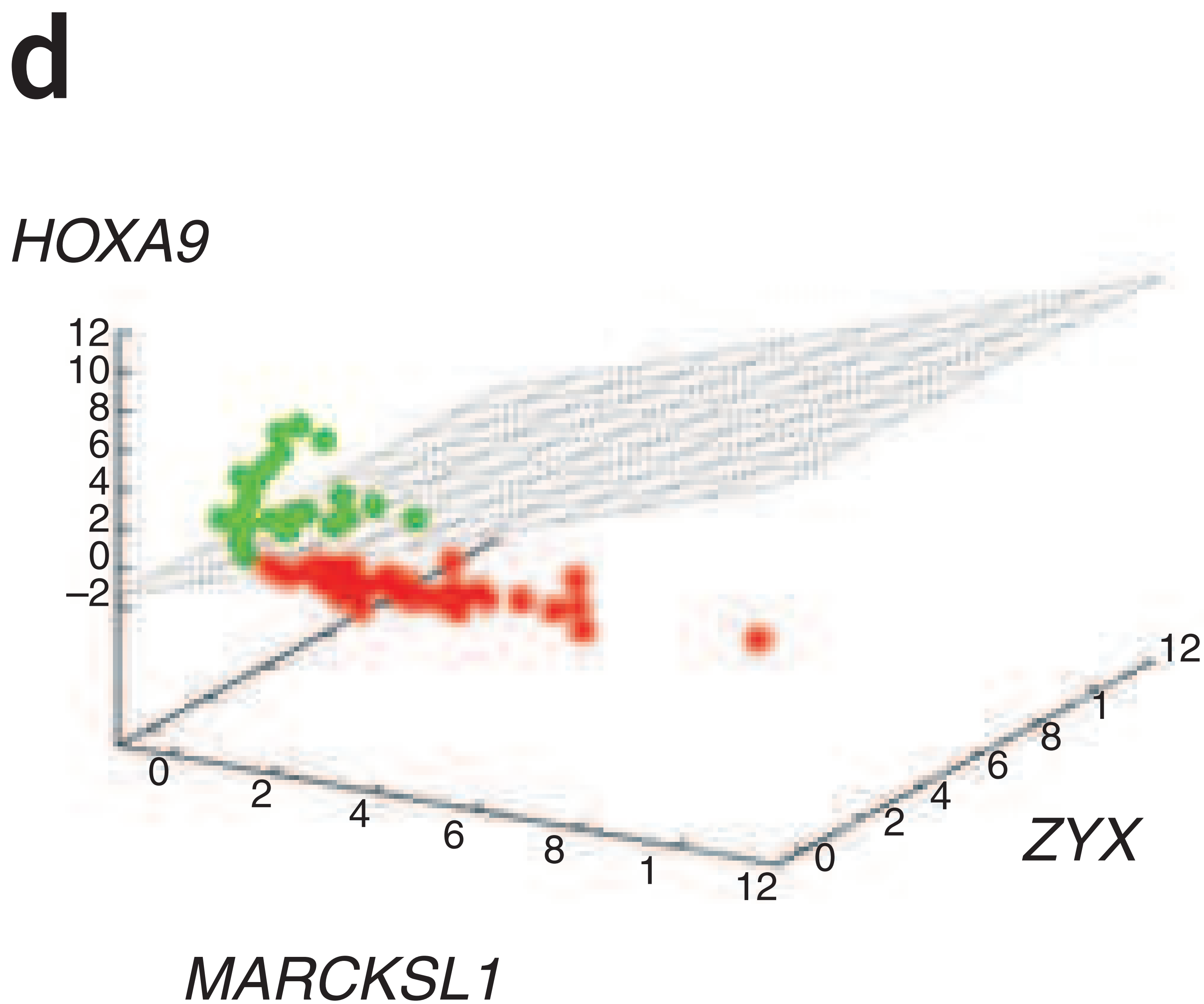


Separating Hyperplane in 1D — a Point

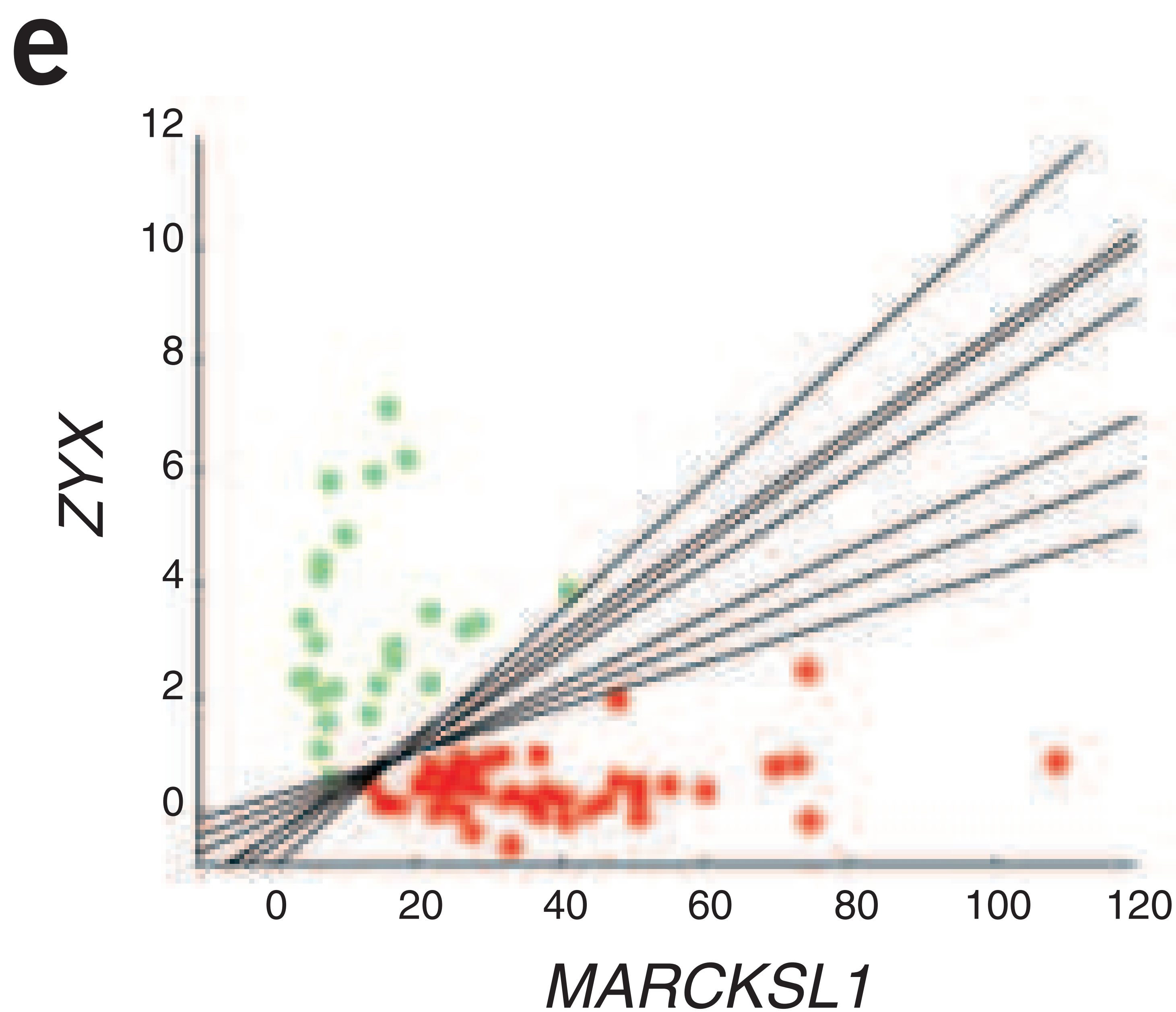
C



... and in 3D: a plane

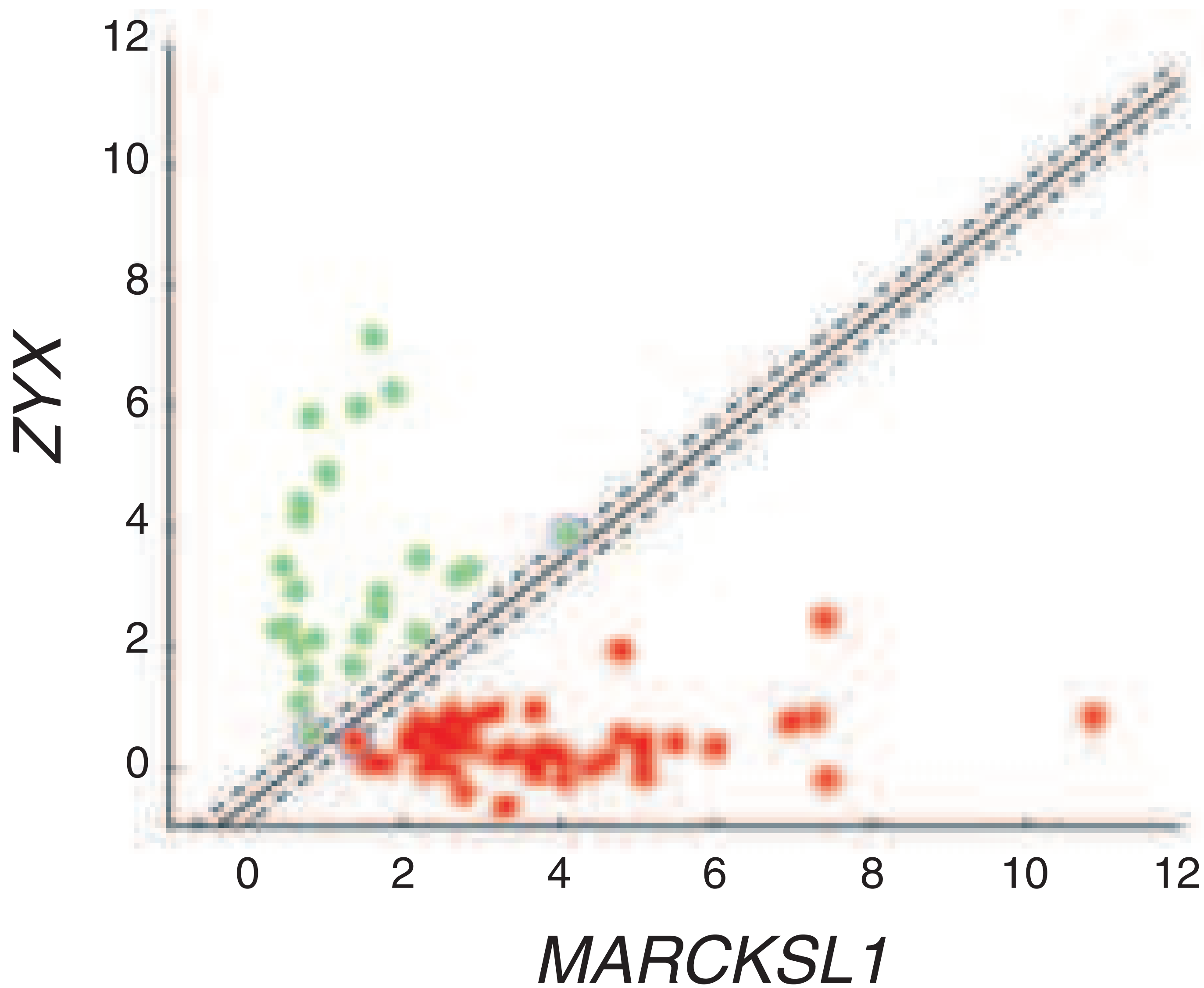


Many Potential Separating Hyperplanes ...
(all "optimal" w.r.t. some loss function)

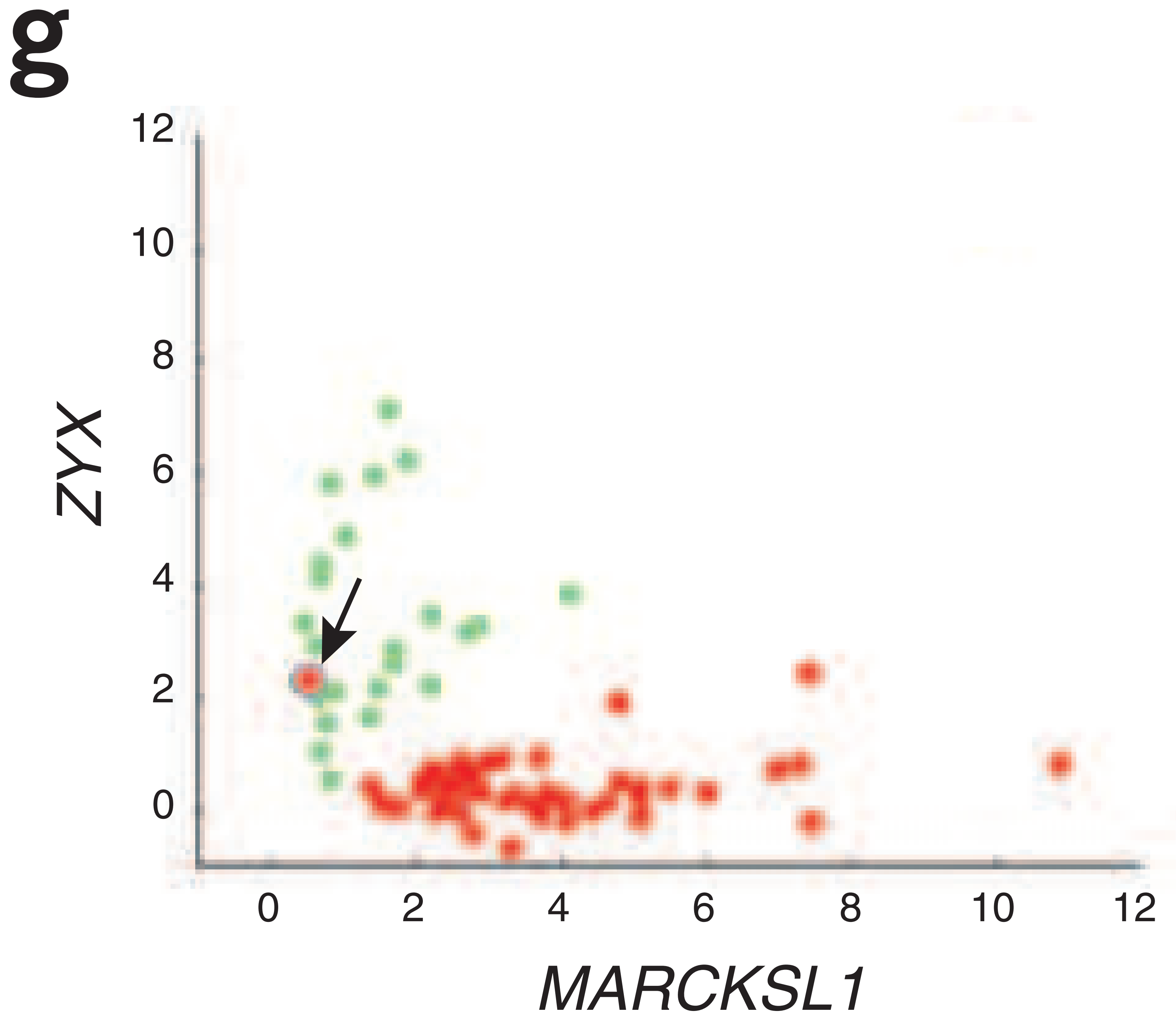


The Maximum-Margin Hyperplane

f

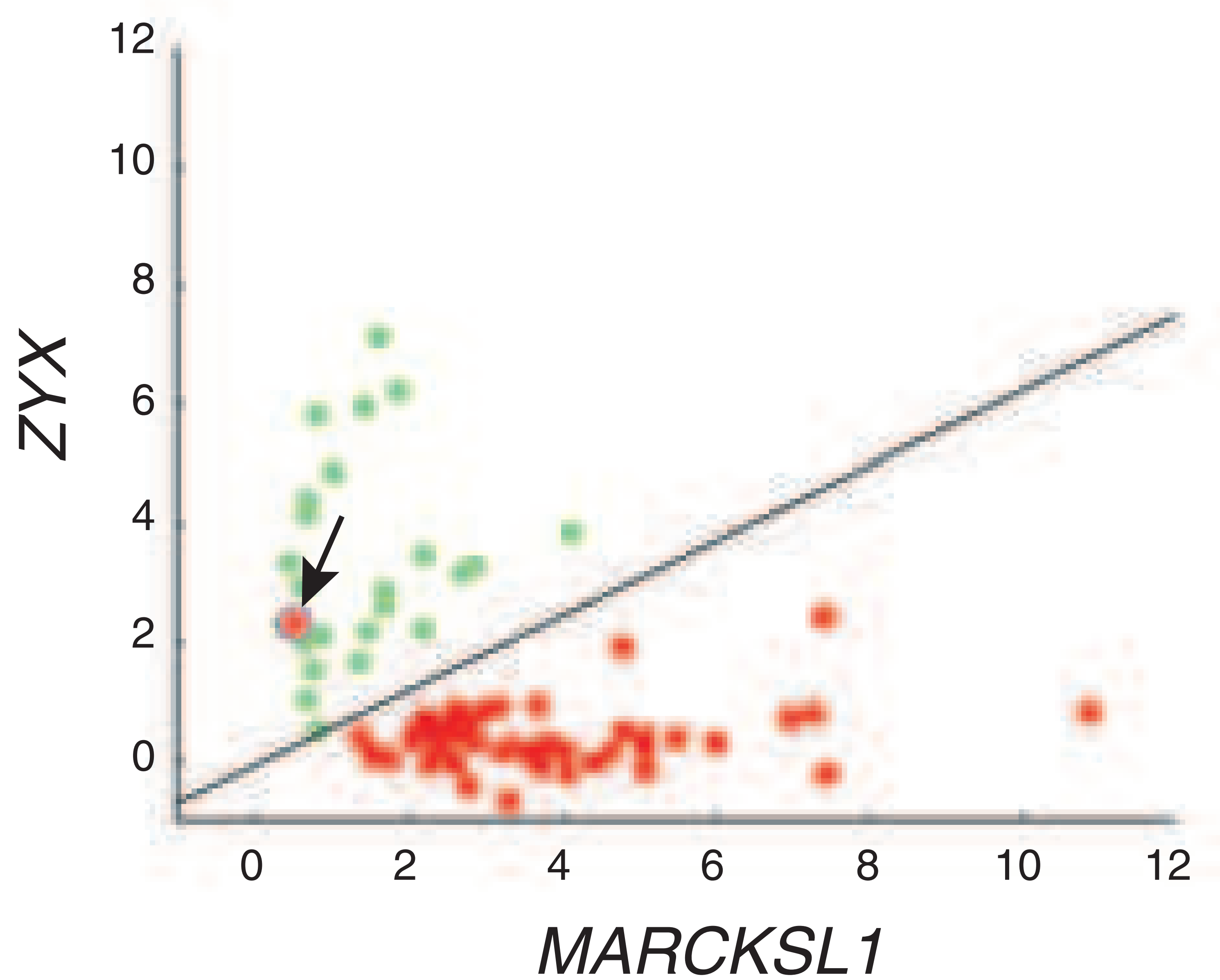


What to Do With Outliers?

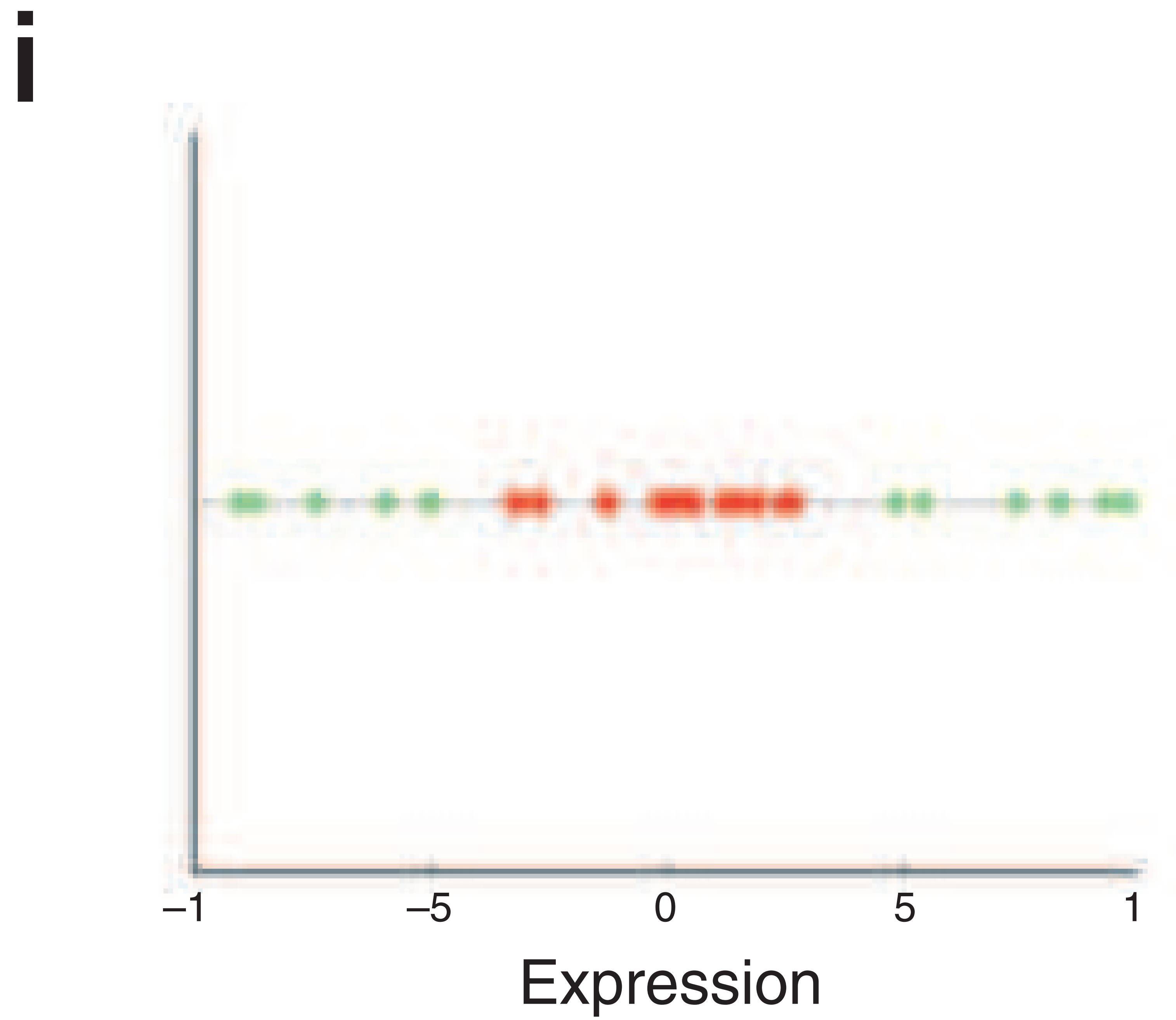


The Soft-Margin Hyperplane

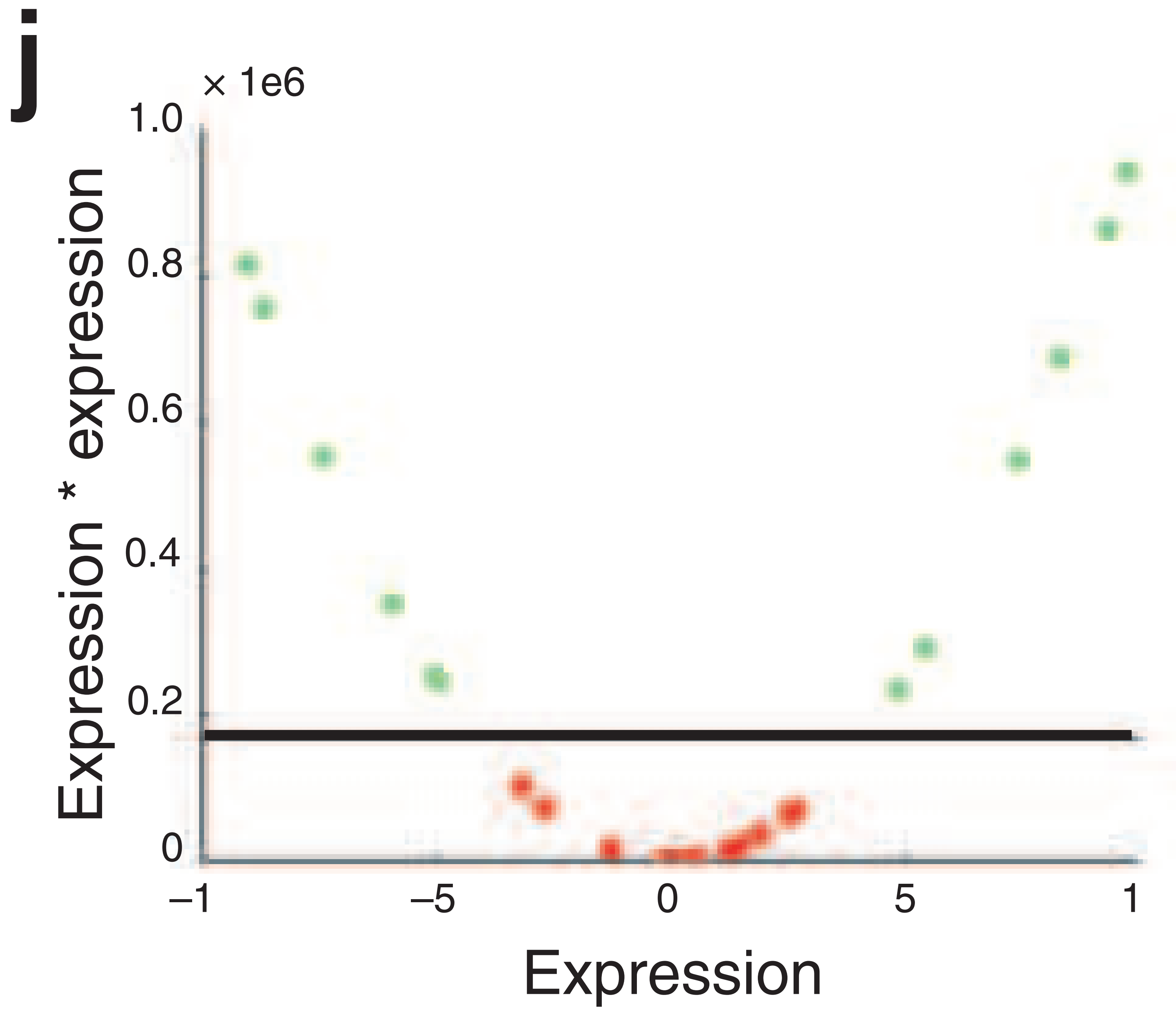
h



The Kernel Function in 1D



Mapping the 1D data to 2D (here: squaring)



Not linearly separable in input space ...

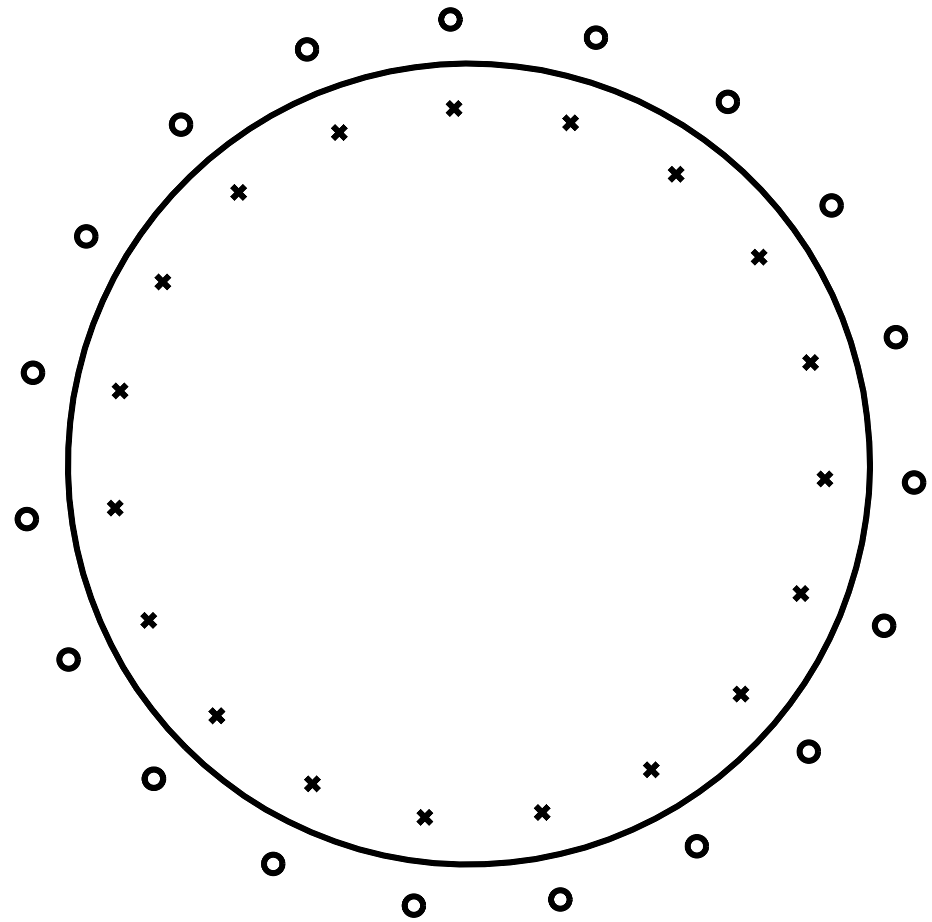


Figure 3. The crosses and the circles cannot be separated by a linear perceptron in the plane.

Map from 2D to 3D ...

$$\Phi(x) = \begin{pmatrix} \phi_1(x) \\ \phi_2(x) \\ \phi_3(x) \end{pmatrix} = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1 x_2 \\ x_2^2 \end{pmatrix}.$$

… linear separability in 3D
(actually: data still 2D, "live" on a manifold of original D!)

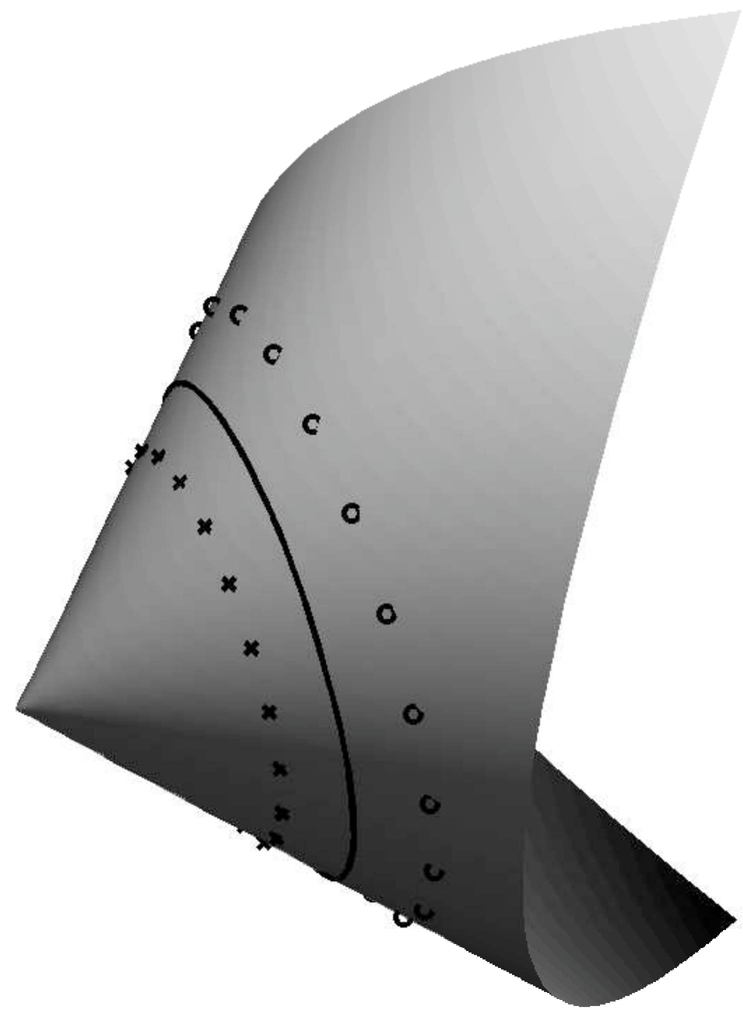
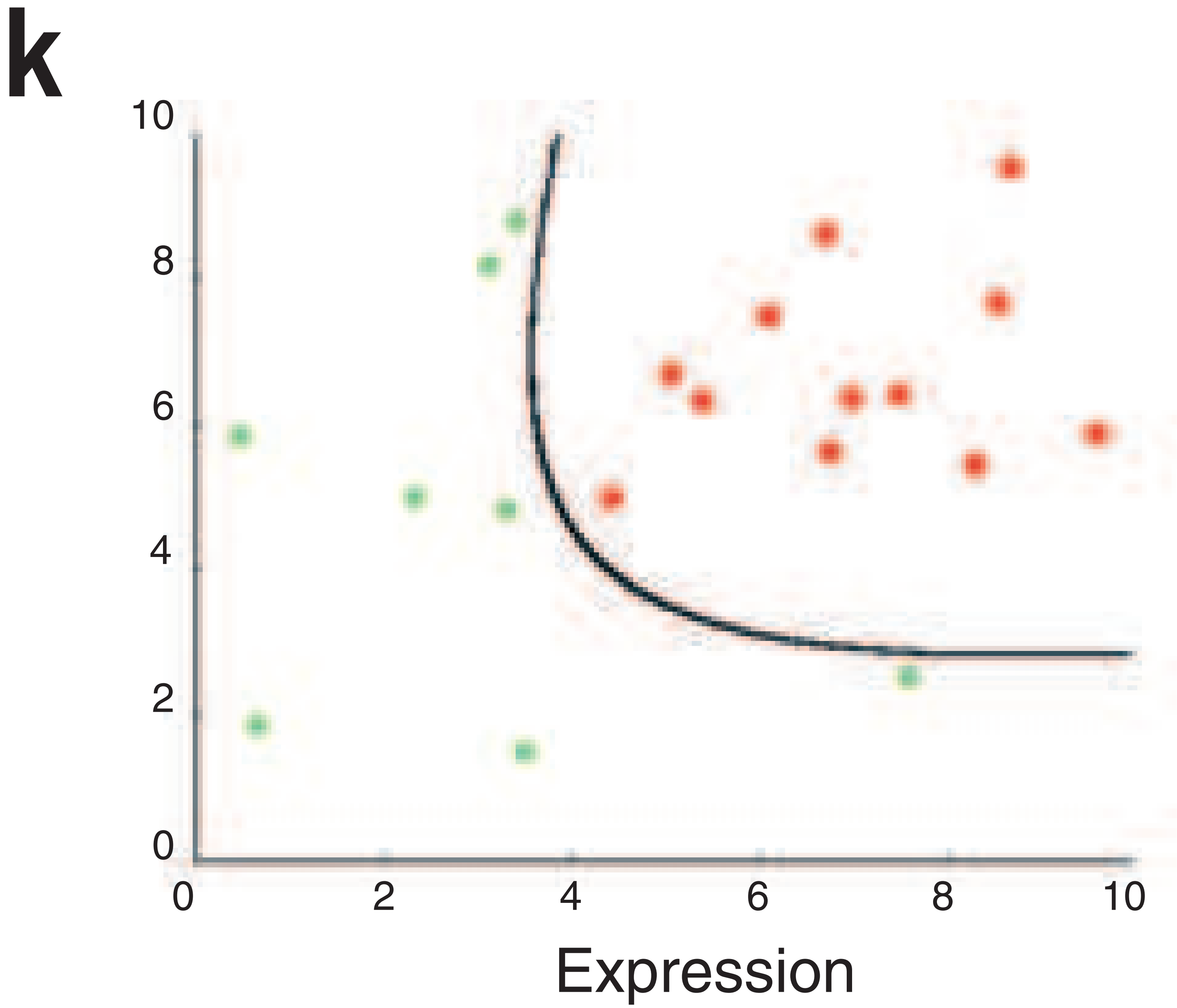


Figure 4. The crosses and circles from Figure 3 can be mapped to a three-dimensional space in which they can be separated by a linear perceptron.

Projecting the 4D Hyperplane Back into 2D Input Space



SVM magic?

SVM magic?

For *any* consistent dataset there is a kernel that allows *perfect* separation of the data

SVM magic?

For *any* consistent dataset there is a kernel that allows *perfect* separation of the data

Why bother with soft-margins?

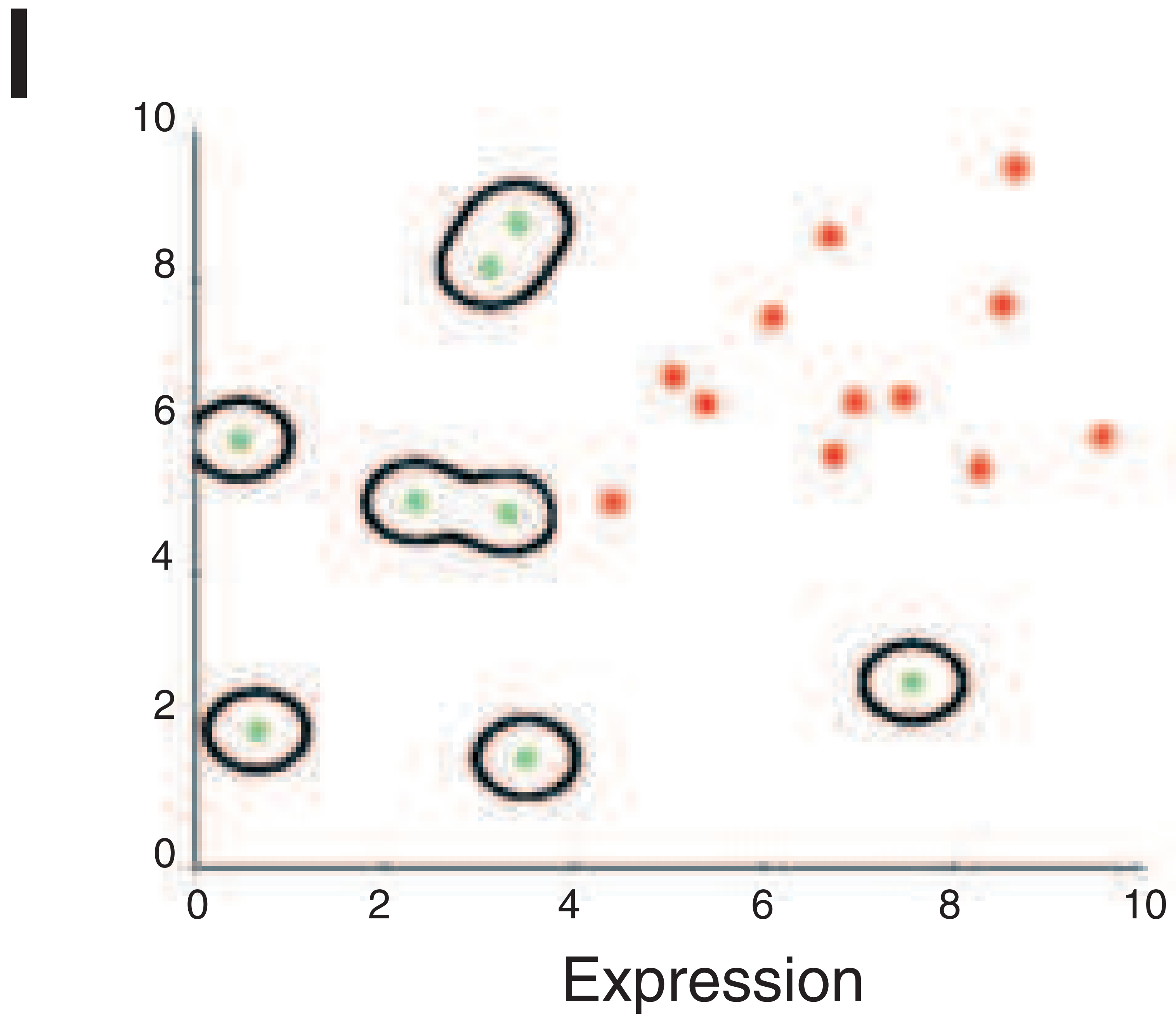
SVM magic?

For *any* consistent dataset there is a kernel that allows *perfect* separation of the data

Why bother with soft-margins?

The so-called *curse of dimensionality*: as the number of variables considered increases, the number of possible solutions increases exponentially ... overfitting looms large!

Overfitting



Regularisation & Cross-validation

Regularisation & Cross-validation

Find a compromise between complexity and classification performance, i.e. kernel function and soft-margin

Regularisation & Cross-validation

Find a compromise between complexity and classification performance, i.e. kernel function and soft-margin

Penalise complex functions via a regularisation term or *regulariser*

Regularisation & Cross-validation

Find a compromise between complexity and classification performance, i.e. kernel function and soft-margin

Penalise complex functions via a regularisation term or *regulariser*

Cross-validate the results (leave-one-out or 10-fold typically used)

SVM Summary

SVM Summary

Kernel essential—best kernel typically found by trial-and-error and experience with similar problems etc.

SVM Summary

Kernel essential—best kernel typically found by trial-and-error and experience with similar problems etc.

Inverting not always easy; need approximations etc. (i.e. science hard, engineering easy as they don't care as long as it works!)

SVM Summary

Kernel essential—best kernel typically found by trial-and-error and experience with similar problems etc.

Inverting not always easy; need approximations etc. (i.e. science hard, engineering easy as they don’t care as long as it works!)

Theoretically sound and a convex optimisation (no local minima)

SVM Summary

Kernel essential—best kernel typically found by trial-and-error and experience with similar problems etc.

Inverting not always easy; need approximations etc. (i.e. science hard, engineering easy as they don't care as long as it works!)

Theoretically sound and a convex optimisation (no local minima)

Choose between:
- complicated decision functions and training (neural networks)
- clear theoretical foundation (best possible generalisation), convex optimisation but need to trade-off complexity versus soft-margin and skilful selection of the "right" kernel.
(= "correct" non-linear similarity measure for the data!)

Regularisation, Cross-Validation and Kernels

Much of the success of modern machine learning methods can attributed to three ideas:

Regularisation, Cross-Validation and Kernels

Much of the success of modern machine learning methods can attributed to three ideas:

1. Regularisation. Given are N "datapoints" $(\mathrm{x}_i, \mathrm{y}_\mathrm{i})$ with ...

$$\mathbf{y} = y_1, ..., y_N$$

$$\mathbf{x} = x_1, ..., x_N$$

and a model f . Then the "error" between data and model is: $\mathcal{E}(\mathbf{y}, f(\mathbf{x}))$
In machine learning we not only take the "error" between model and data into account but in addition a measure of the complexity of the model f: $\mathcal{E}(\mathbf{y}, f(\mathbf{x})) + \lambda\mathcal{R}(f)$

Regularisation, Cross-Validation and Kernels

Much of the success of modern machine learning methods can attributed to three ideas:

1. Regularisation. Given are N "datapoints" $(\mathrm{x}_i, \mathrm{y}_\mathrm{i})$ with ...

   $$\mathbf{y} = y_1, ..., y_N$$
   $$\mathbf{x} = x_1, ..., x_N$$

   and a model f. Then the "error" between data and model is: $\mathcal{E}(\mathbf{y}, f(\mathbf{x}))$
   In machine learning we not only take the "error" between model and data into account but in addition a measure of the complexity of the model f: $\mathcal{E}(\mathbf{y}, f(\mathbf{x})) + \lambda\mathcal{R}(f)$

2. Cross-Validation. Regularisation is related to the prior in Bayesian statistics. Unlike in Bayesian statistics the trade-off between small error and low-complexity of the model is controlled by a parameter $\boldsymbol{\lambda}$— this is *optimized* using cross-validation.

Regularisation, Cross-Validation and Kernels

Much of the success of modern machine learning methods can attributed to three ideas:

1. Regularisation. Given are N "datapoints" $(\mathrm{x}_i, \mathrm{y}_\mathrm{i})$ with ...

$$\mathbf{y} = y_1, ..., y_N$$

$$\mathbf{x} = x_1, ..., x_N$$

and a model f. Then the "error" between data and model is: $\mathcal{E}(\mathbf{y}, f(\mathbf{x}))$
In machine learning we not only take the "error" between model and data into account but in addition a measure of the complexity of the model f: $\mathcal{E}(\mathbf{y}, f(\mathbf{x})) + \lambda \mathcal{R}(f)$

2. Cross-Validation. Regularisation is related to the prior in Bayesian statistics. Unlike in Bayesian statistics the trade-off between small error and low-complexity of the model is controlled by a parameter λ— this is *optimized* using cross-validation.

3. Non-linear mapping with linear separation.
True for kernels as well as DNNs.

3

What changed vision research in 2012?

What changed vision research in 2012?

ImageNet challenge: 1000 categories, 1.2 million training images.

What changed vision research in 2012?

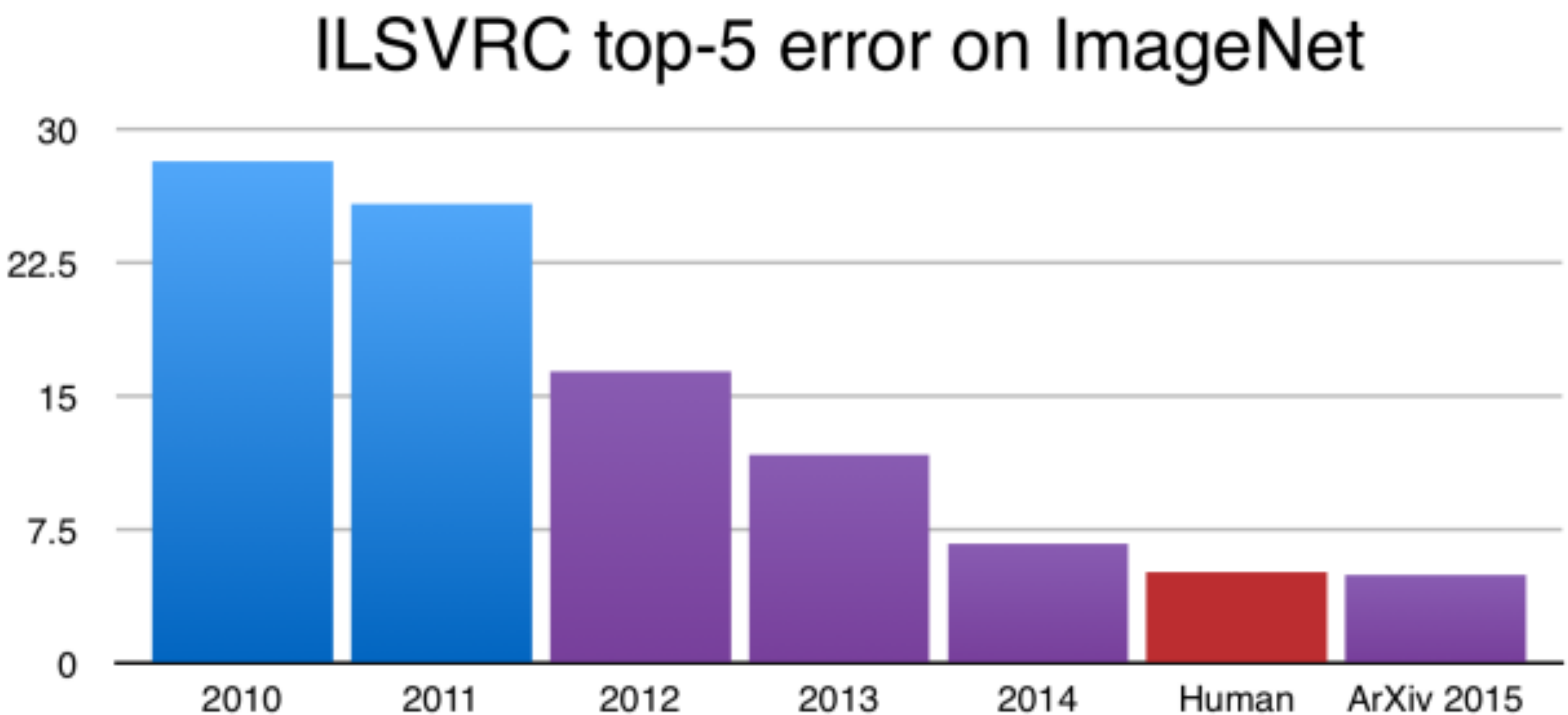
ImageNet challenge: 1000 categories, 1.2 million training images.

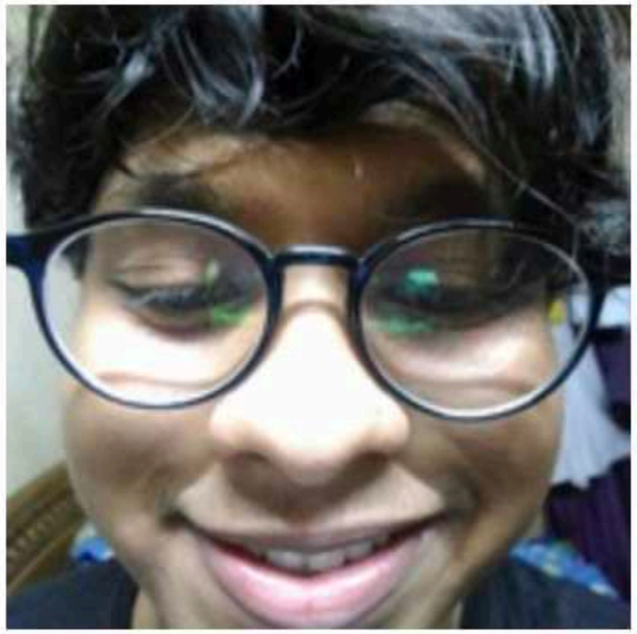
AlexNet by Krizhevsky, Sutskever & Hinton (2012) appears on the stage, and basically reduces the prediction error by nearly 50%:

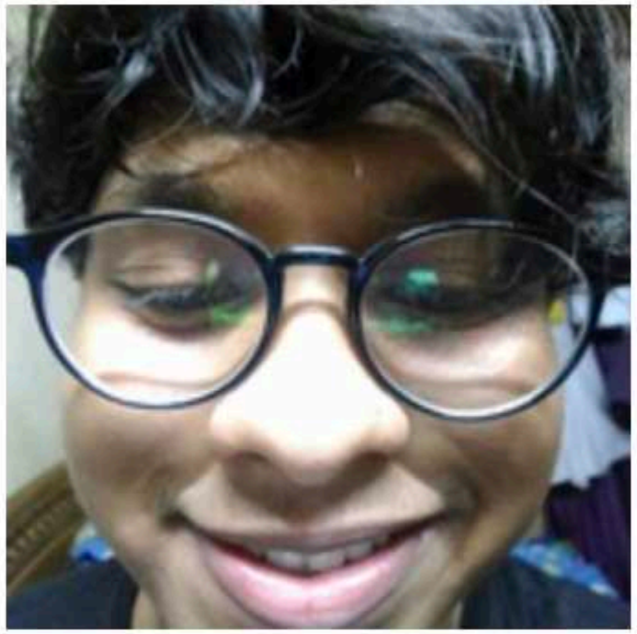
What changed vision research in 2012?

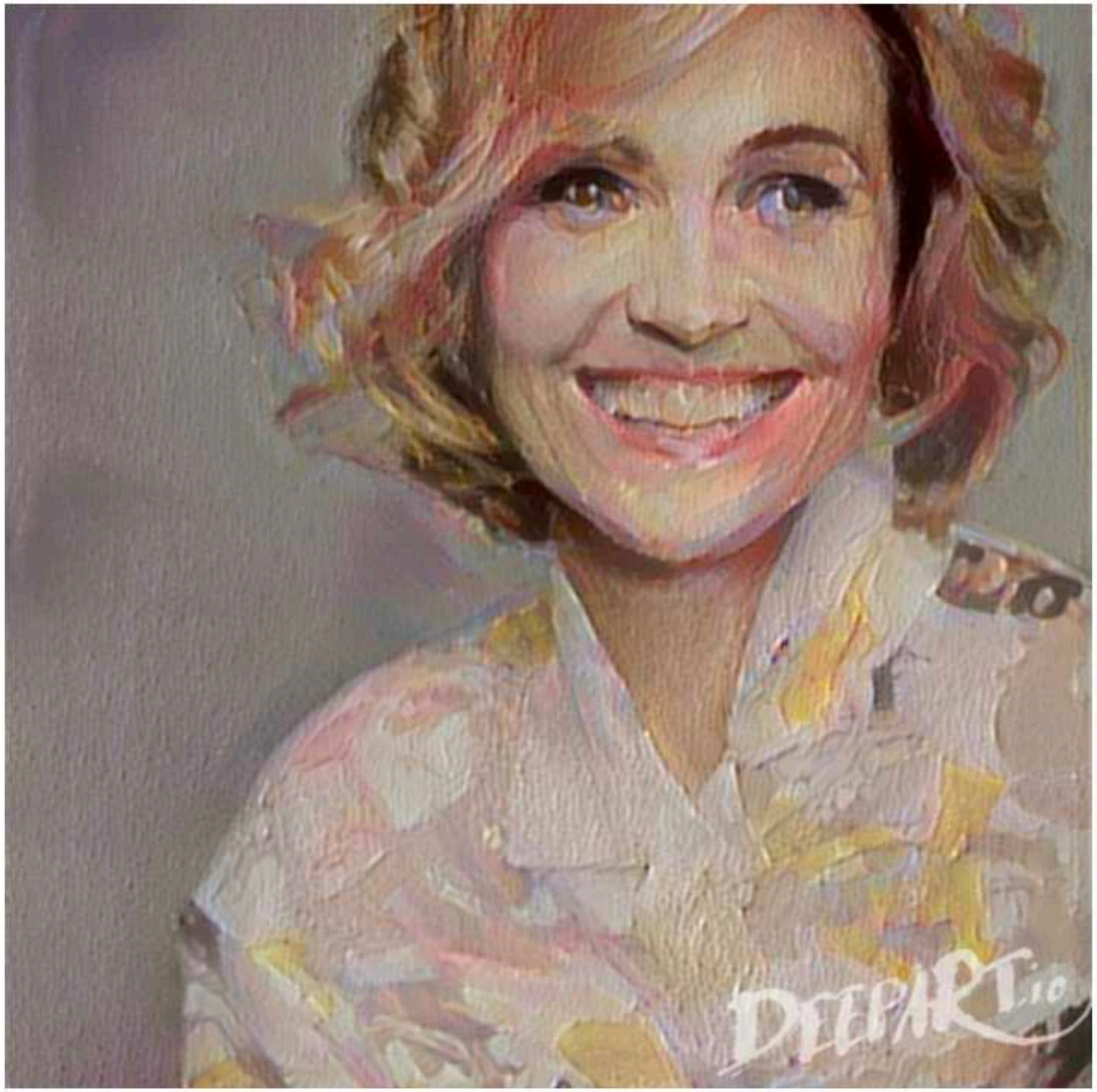
ImageNet challenge: 1000 categories, 1.2 million training images.

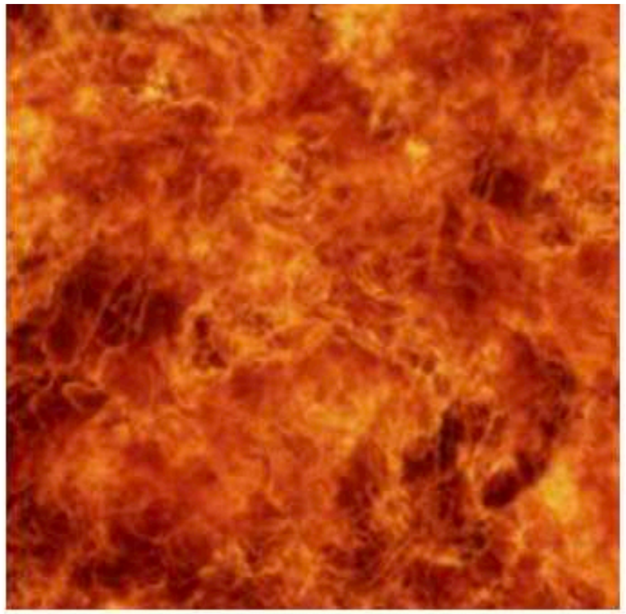
AlexNet by Krizhevsky, Sutskever & Hinton (2012) appears on the stage, and basically reduces the prediction error by nearly 50%:



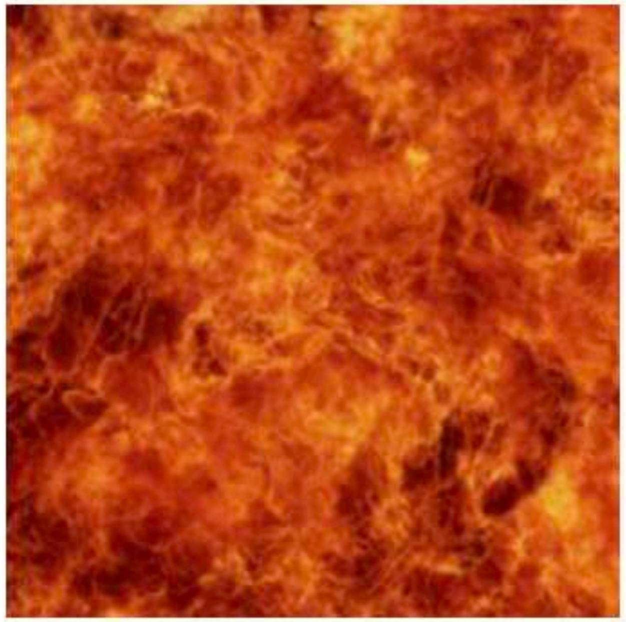
 

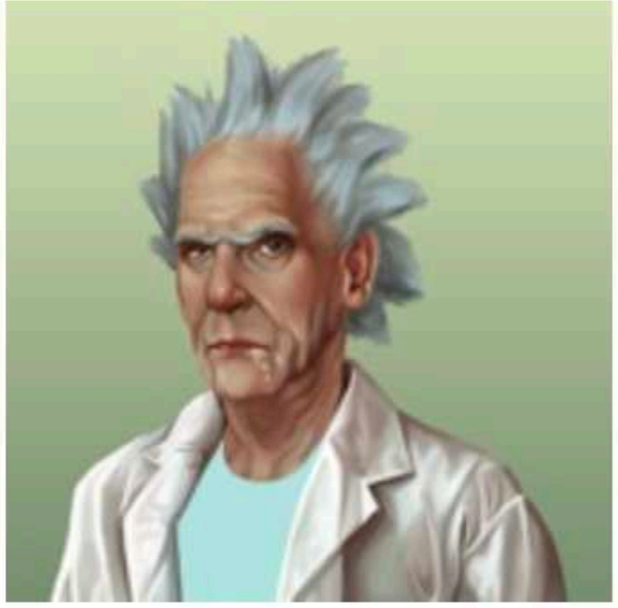


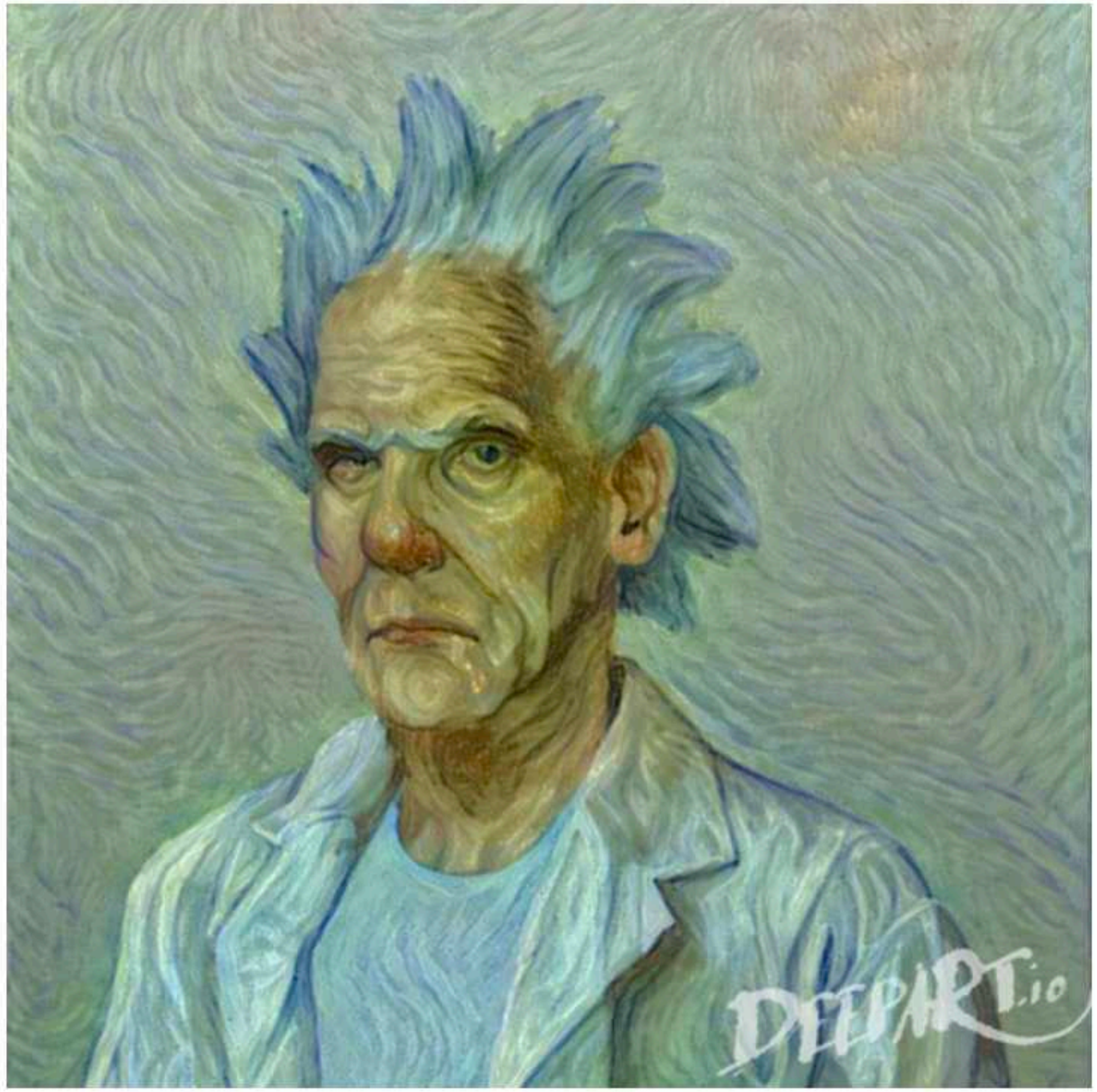

DEEPART.io

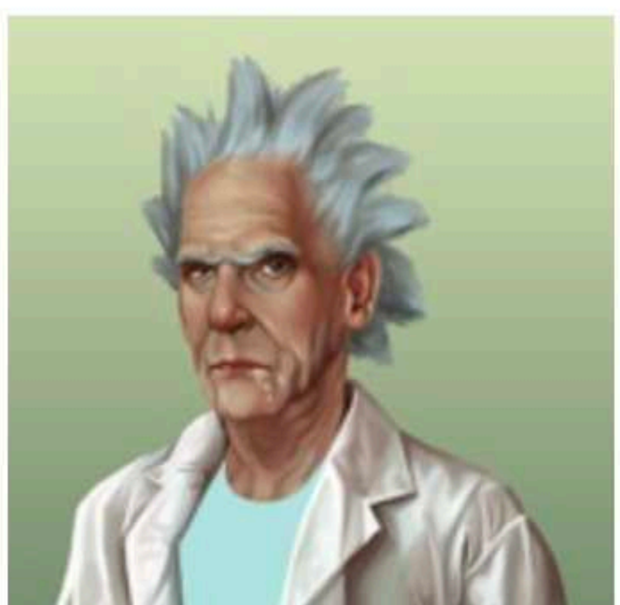


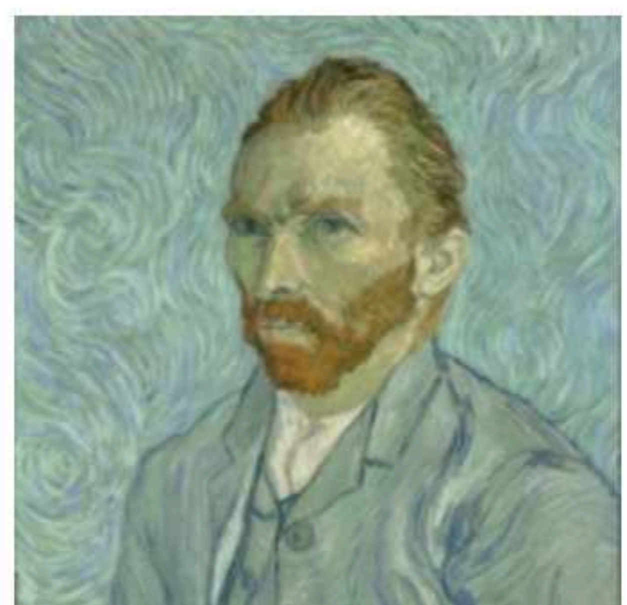

DEEPART.io

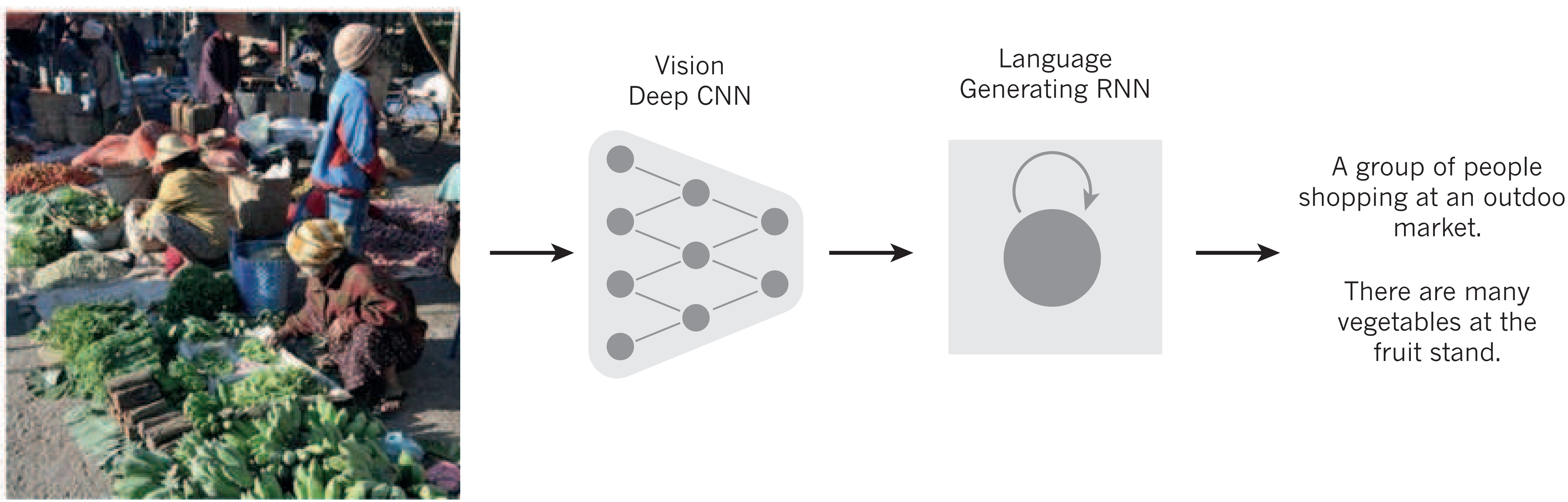


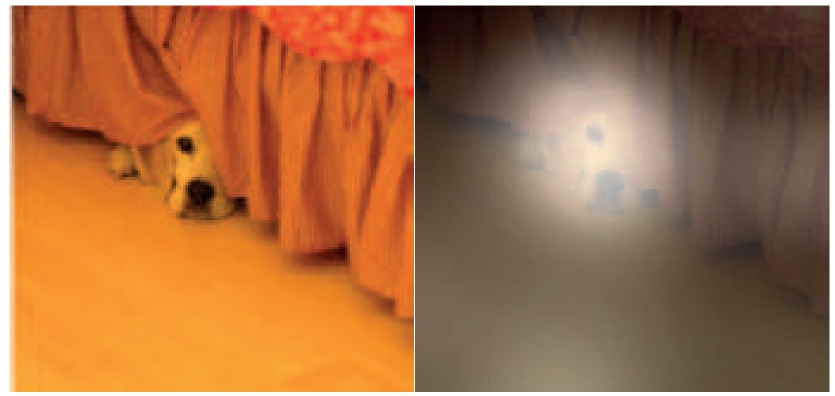

DEEPART.io





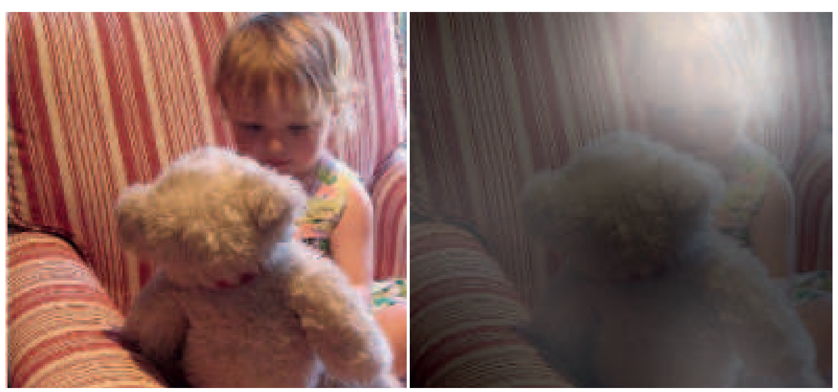

Vision
Deep CNN
Language
Generating RNN
A group of people
shopping at an outdoor
market.
There are many
vegetables at the
fruit stand.

A woman is throwing a **frisbee** in a park.



A **dog** is standing on a hardwood floor.

A **stop** sign is on a road with a mountain in the background



A little **girl** sitting on a bed with a teddy bear.

A group of **people** sitting on a boat in the water.

A giraffe standing in a forest with **trees** in the background.

?

Problem of finding a sharp image from a blurry photo:

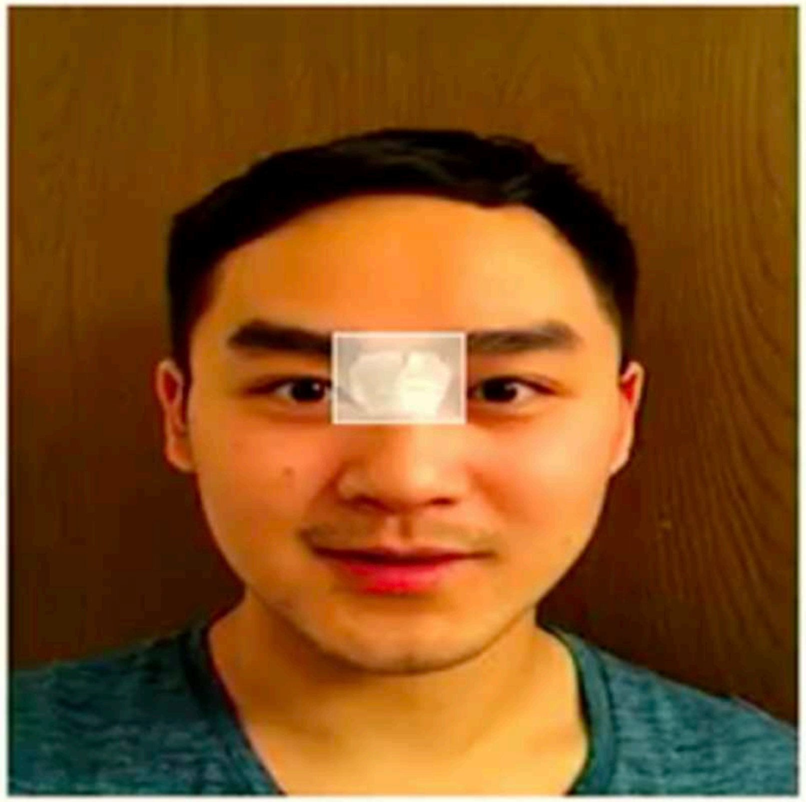
Blind Image Deconvolution

modified from Michael Hirsch

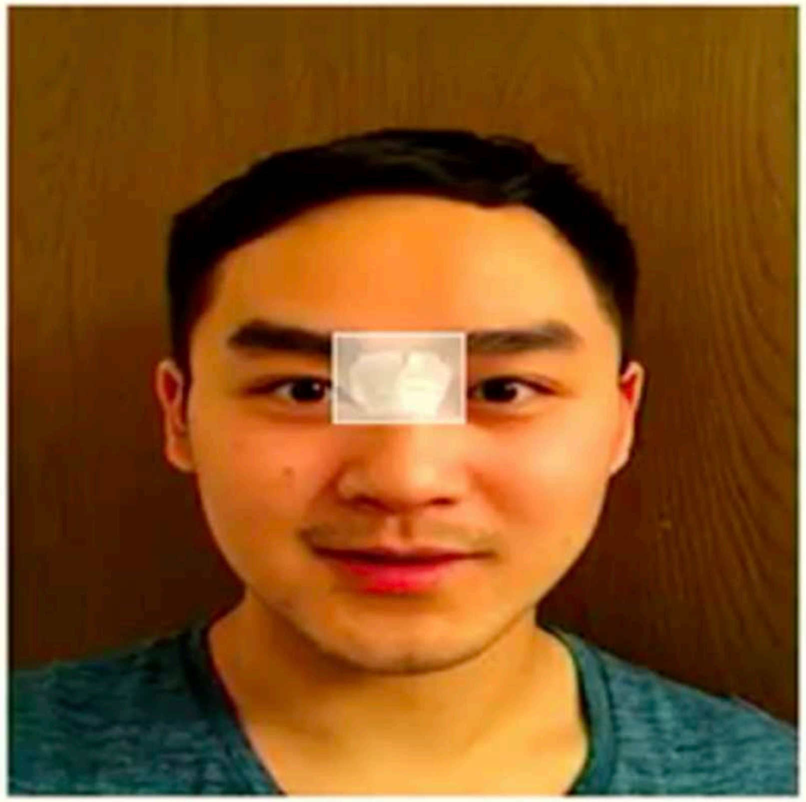
from Michael Hirsch

from Michael Hirsch

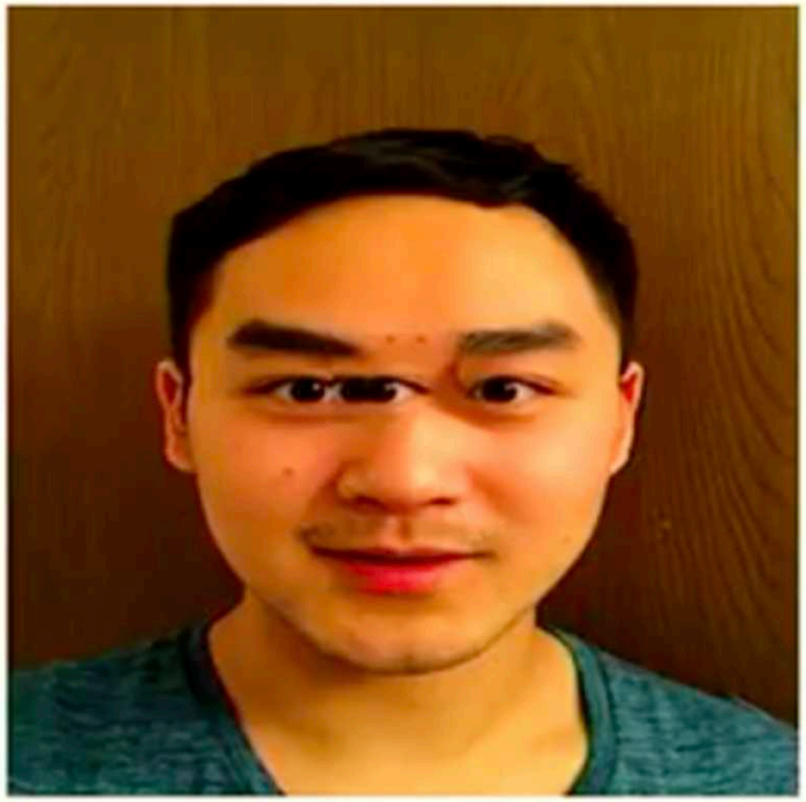
Input:



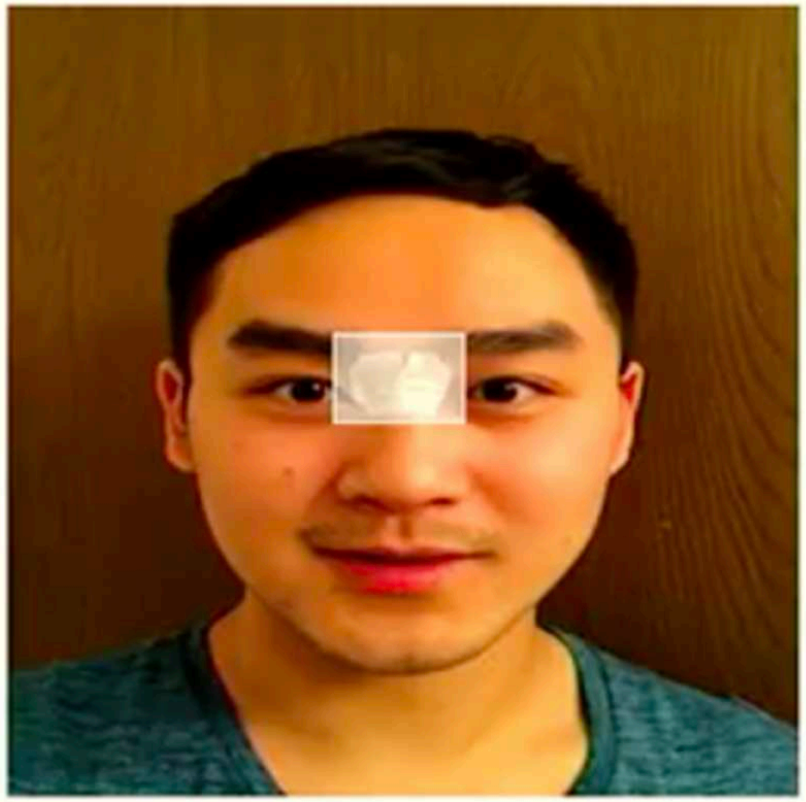
Input:



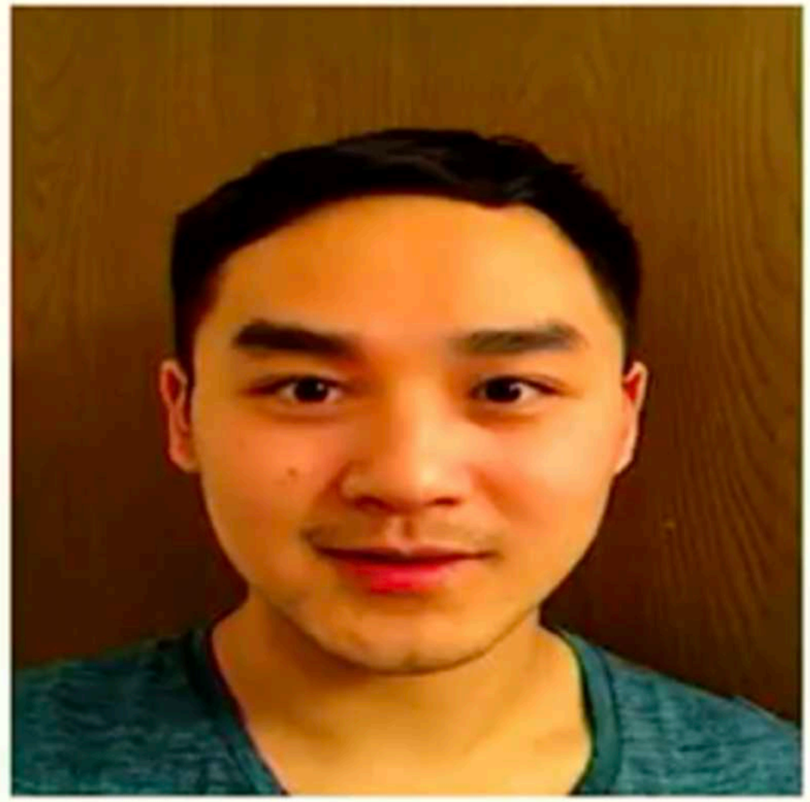
Content-Aware Fill Result:



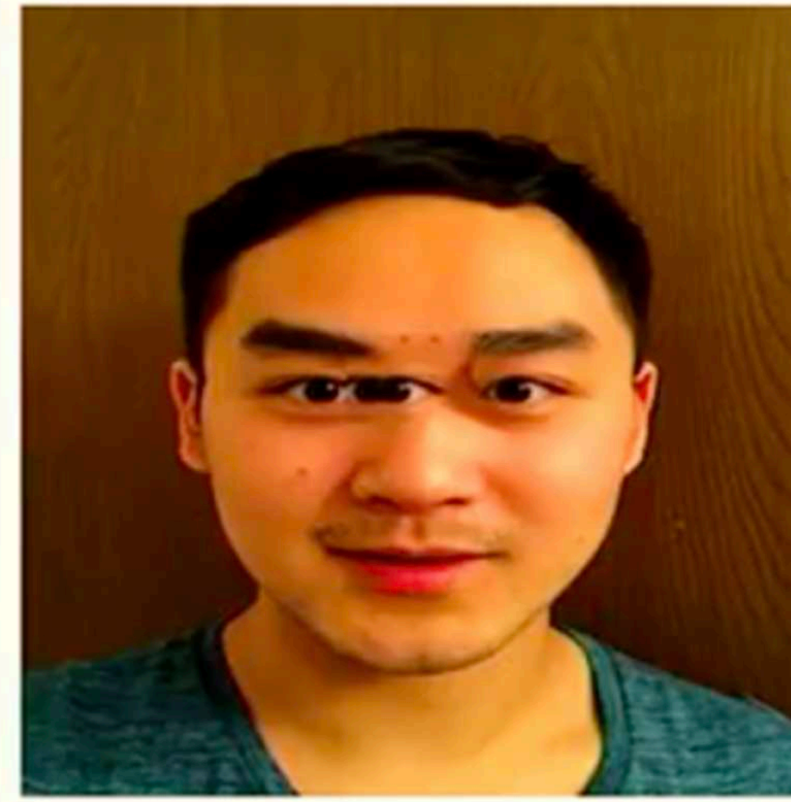
Input:



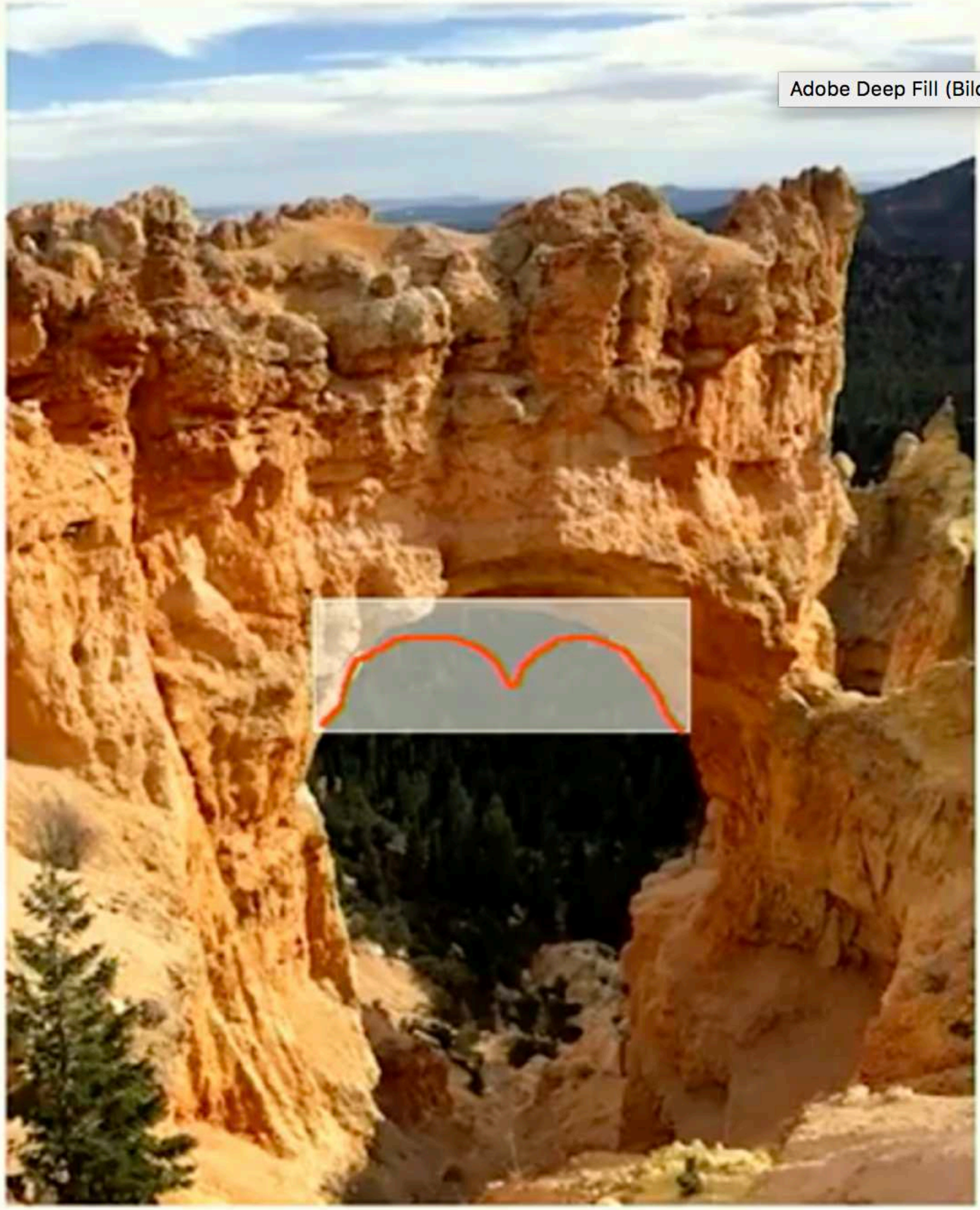
Deep Fill Result:



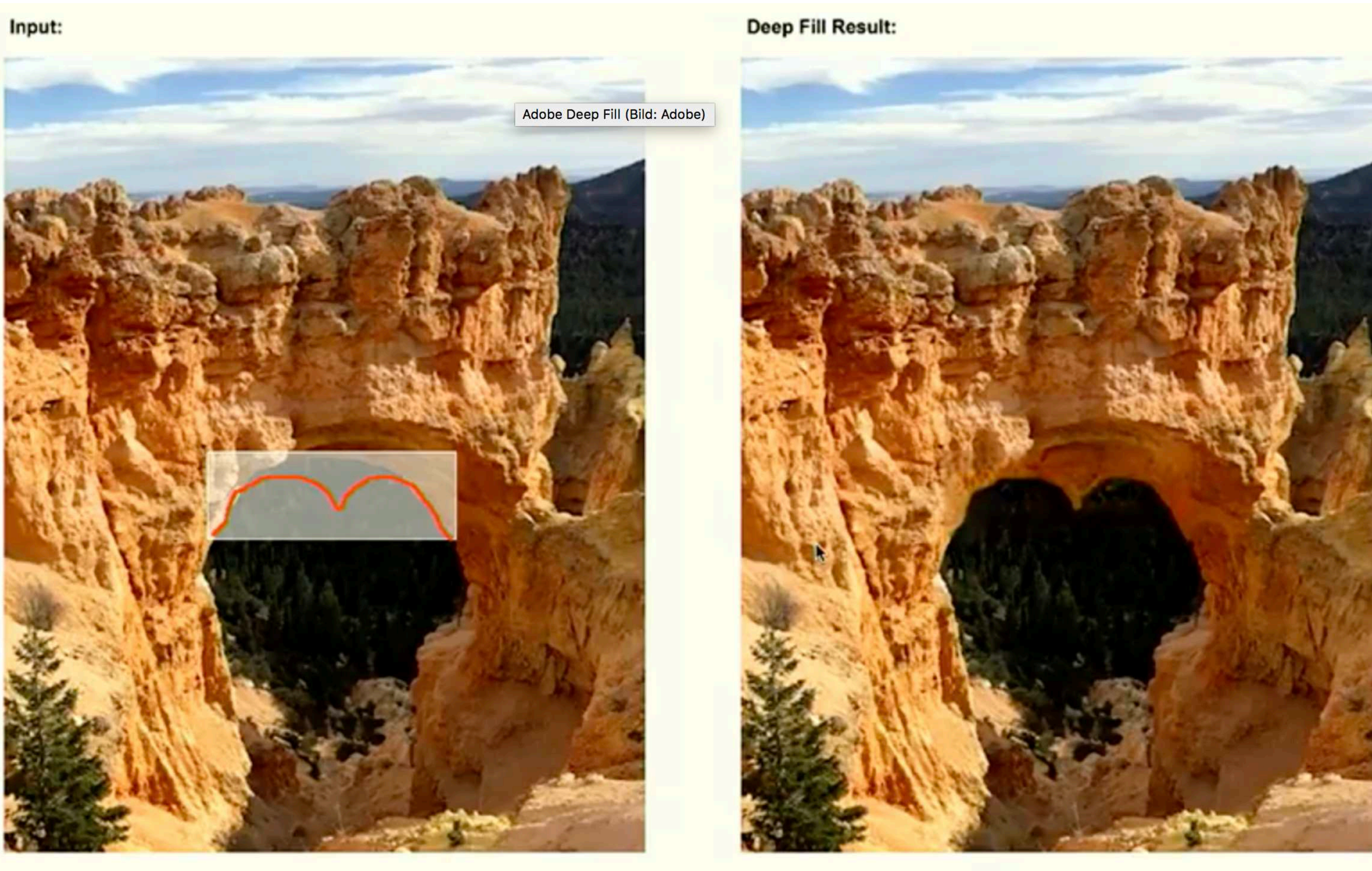
Content-Aware Fill Result:



Input:



Adobe Deep Fill (Bild: Adobe)


Input:
Deep Fill Result:
Adobe Deep Fill (Bild: Adobe)

Sequence of Blurry Photos (Image Burst)

from Michael Hirsch

Sequence of Blurry Photos (Image Burst)

from Michael Hirsch

Sequence of Blurry Photos (Image Burst)

from Michael Hirsch

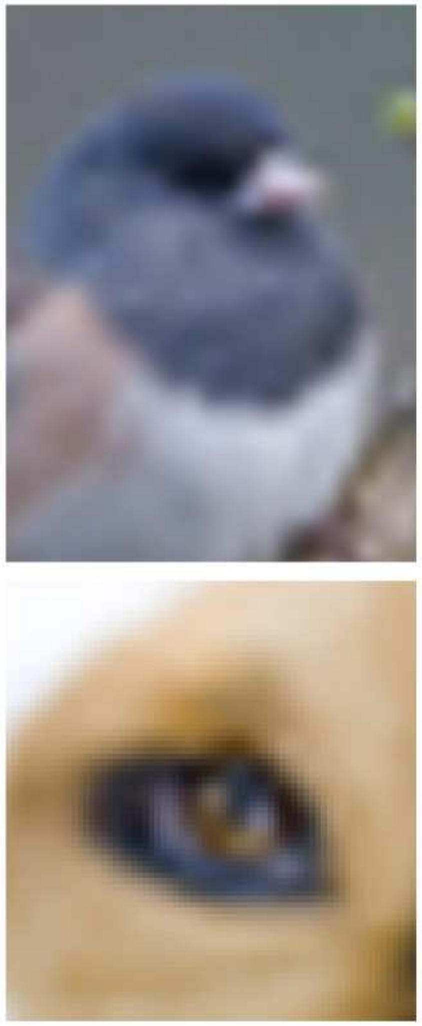
Sequence of Blurry Photos (Image Burst)

from Michael Hirsch

Result of Proposed Image Burst Deblurring Method

from Michael Hirsch

EnhanceNet: Photo-realistic Super-resolution

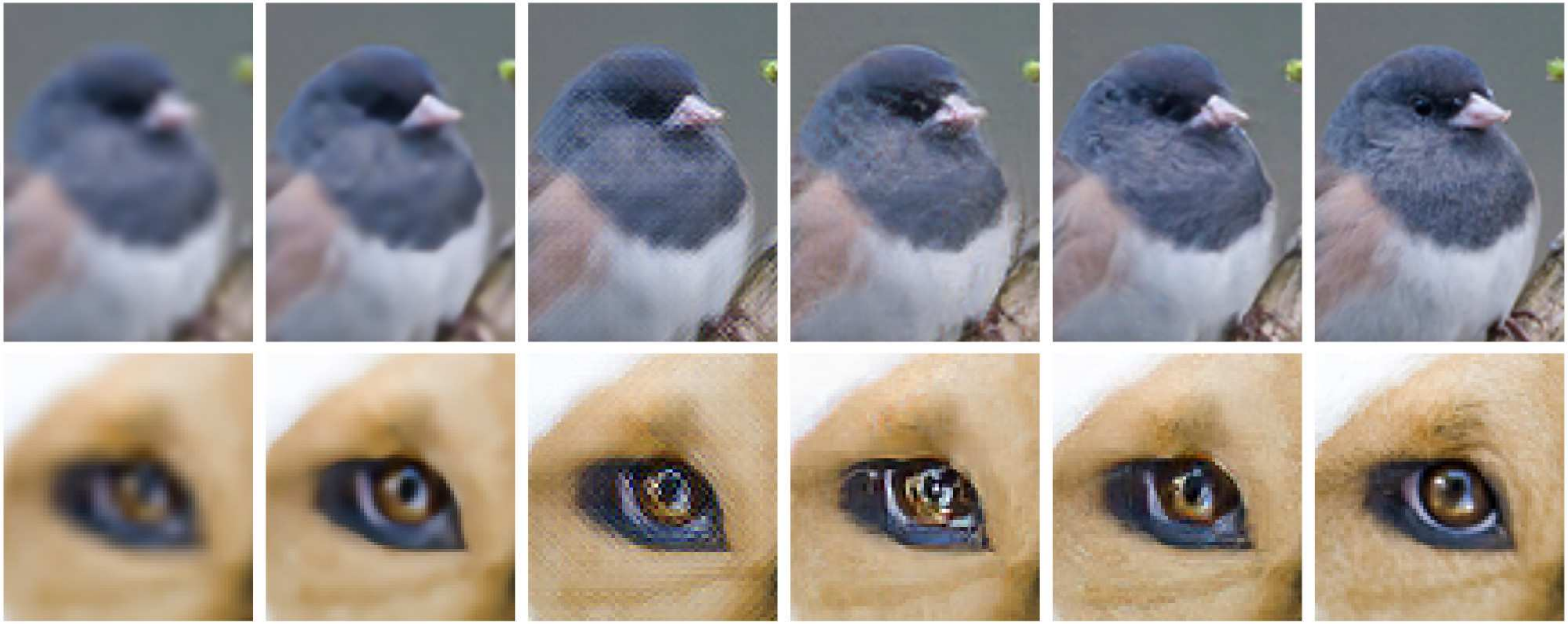


Bicubic

Dataset	Bicubic	ENet-E	ENet-P	ENet-PA	ENet-PAT
Set5	28.42	**31.74**	28.28	27.20	28.56
Set14	26.00	**28.42**	25.64	24.93	25.77
BSD100	25.96	**27.50**	24.73	24.19	24.93
Urban100	23.14	**25.66**	23.75	22.51	23.54

from Michael Hirsch

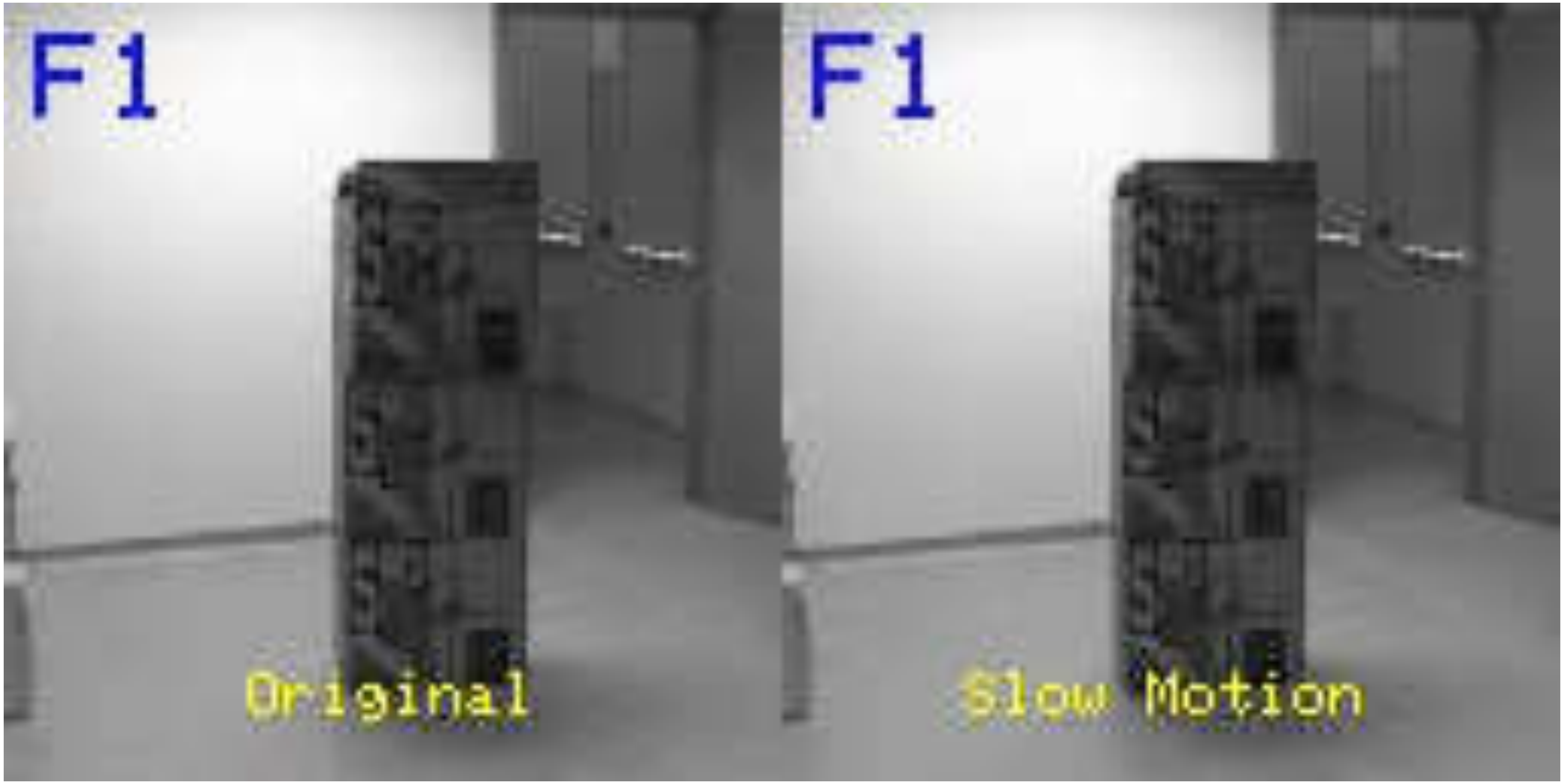
EnhanceNet: Photo-realistic Super-resolution



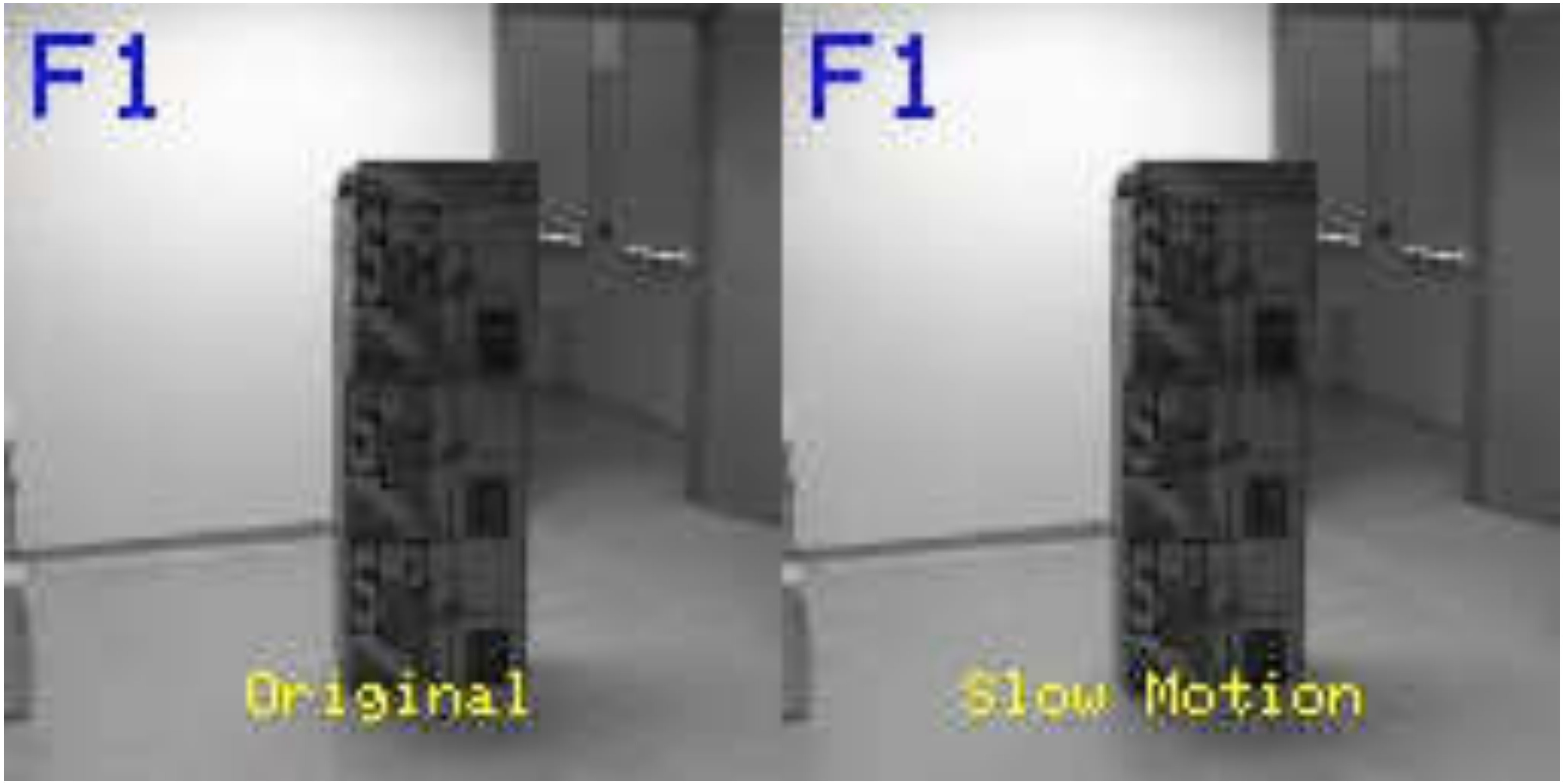
Bicubic | ENet-E | ENet-P | ENet-PA | ENet-PAT | I_{HR}

Dataset	Bicubic	ENet-E	ENet-P	ENet-PA	ENet-PAT
Set5	28.42	**31.74**	28.28	27.20	28.56
Set14	26.00	**28.42**	25.64	24.93	25.77
BSD100	25.96	**27.50**	24.73	24.19	24.93
Urban100	23.14	**25.66**	23.75	22.51	23.54

from Michael Hirsch



from Michael Hirsch



from Michael Hirsch

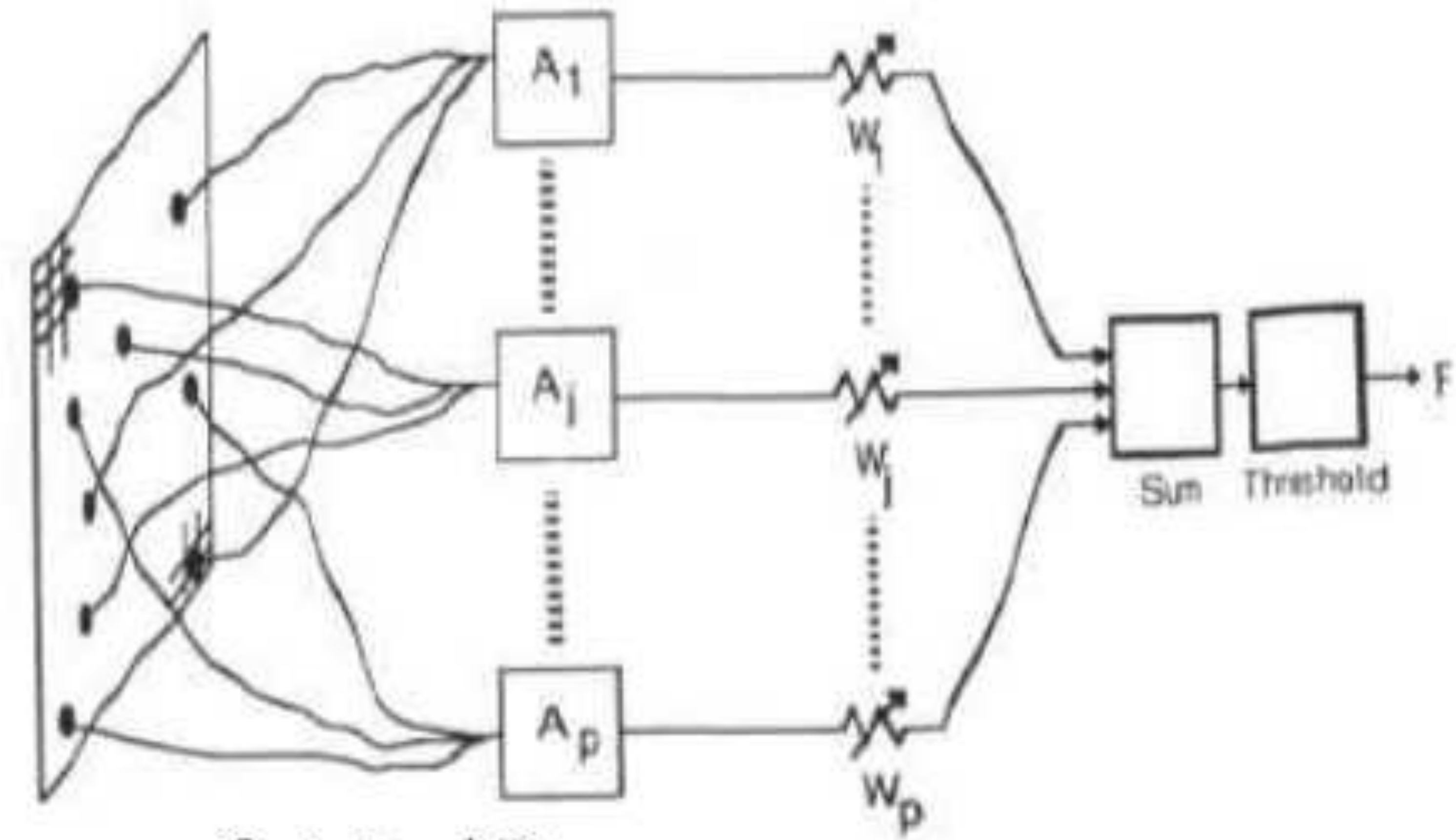
Autonomous cars

Autonomous cars

Fundamentals of Neural Networks

Interest in shallow, 2-layer artificial neural networks (ANN)—so-called **perceptrons**—began in the late 1950s and early 60s (Frank Rosenblatt), based on Warren McCulloch and Walter Pitts's as well Donald Hebb's ideas of computation by neurons from the 1940s.

Perceptron (1957)

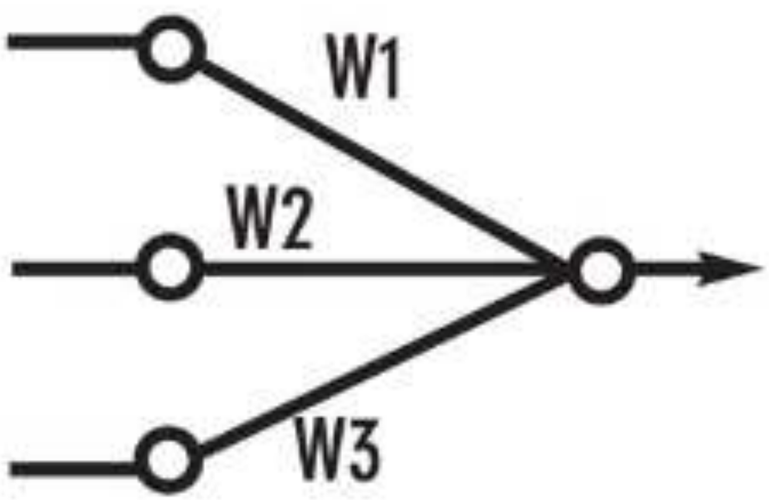


Original Perceptron

(From Perceptrons by M. L Minsky and S. Papert, 1969, Cambridge, MA: MIT Press. Copyright 1969 by MIT Press.

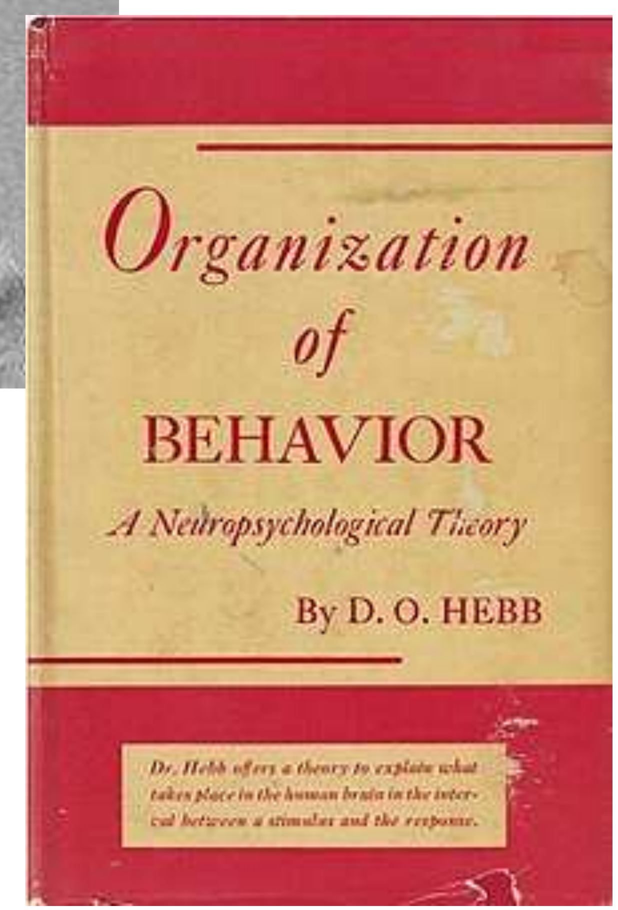
Frank Rosenblatt
(1928-1971)

Simplified model:



https://kimschmidtsbrain.files.wordpress.com/2015/10/perceptron.jpg

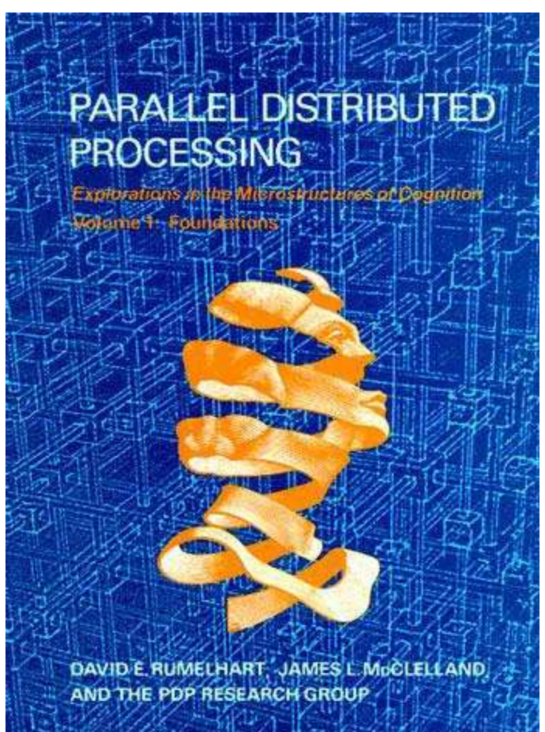
http://cambridgemedicine.org/sites/default/files/styles/large/public/field/image/DonaldOldingHebb.jpg?itok=py9Uh4D5

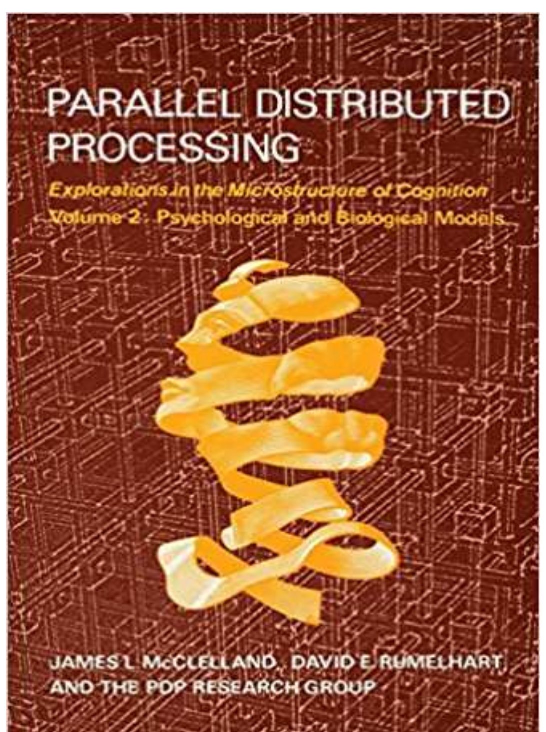


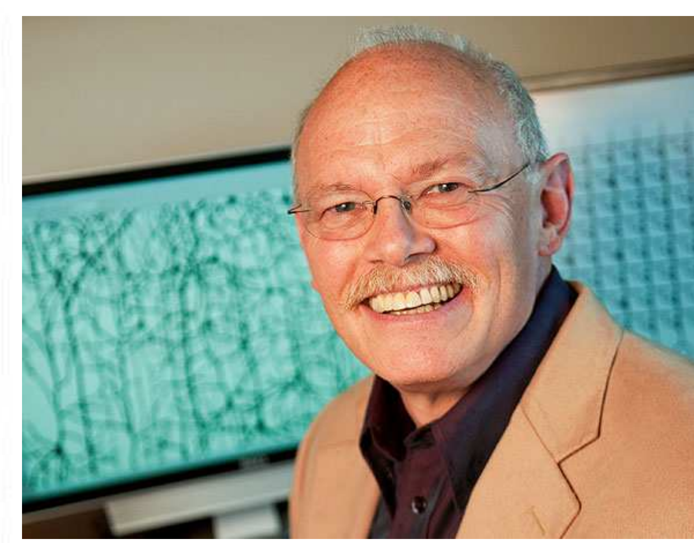
Fundamentals of Neural Networks

Interest in shallow, 2-layer artificial neural networks (ANN)—so-called **perceptrons**—began in the late 1950s and early 60s (Frank Rosenblatt), based on Warren McCulloch and Walter Pitts's as well Donald Hebb's ideas of computation by neurons from the 1940s.

Second wave of ANN research and interest in psychology—often termed **connectionism**—after the publication of the **parallel distributed processing** (PDP) books by David Rumelhart and James McClelland (1986), using the backpropagation algorithm as a learning rule for multi-layer networks.







Fundamentals of Neural Networks

Interest in shallow, 2-layer artificial neural networks (ANN)—so-called **perceptrons**—began in the late 1950s and early 60s (Frank Rosenblatt), based on Warren McCulloch and Walter Pitts's as well Donald Hebb's ideas of computation by neurons from the 1940s.

Second wave of ANN research and interest in psychology—often termed **connectionism**—after the publication of the **parallel distributed processing** (PDP) books by David Rumelhart and James McClelland (1986), using the backpropagation algorithm as a learning rule for multi-layer networks.

Three-layer network with (potentially infinitely many) hidden units in the intermediate layer is a universal function approximator (Kurt Hornik, 1991).

Fundamentals of Neural Networks

Interest in shallow, 2-layer artificial neural networks (ANN)—so-called **perceptrons**—began in the late 1950s and early 60s (Frank Rosenblatt), based on Warren McCulloch and Walter Pitts's as well Donald Hebb's ideas of computation by neurons from the 1940s.

Second wave of ANN research and interest in psychology—often termed **connectionism**—after the publication of the **parallel distributed processing** (PDP) books by David Rumelhart and James McClelland (1986), using the backpropagation algorithm as a learning rule for multi-layer networks.

Three-layer network with (potentially infinitely many) hidden units in the intermediate layer is a universal function approximator (Kurt Hornik, 1991).

Non-convex optimization problems during backpropagation training, and lack of data and computing power limited the usefulness of the ANNs:

Fundamentals of Neural Networks

Interest in shallow, 2-layer artificial neural networks (ANN)—so-called **perceptrons**—began in the late 1950s and early 60s (Frank Rosenblatt), based on Warren McCulloch and Walter Pitts's as well Donald Hebb's ideas of computation by neurons from the 1940s.

Second wave of ANN research and interest in psychology—often termed **connectionism**—after the publication of the **parallel distributed processing** (PDP) books by David Rumelhart and James McClelland (1986), using the backpropagation algorithm as a learning rule for multi-layer networks.

Three-layer network with (potentially infinitely many) hidden units in the intermediate layer is a universal function approximator (Kurt Hornik, 1991).

Non-convex optimization problems during backpropagation training, and lack of data and computing power limited the usefulness of the ANNs:

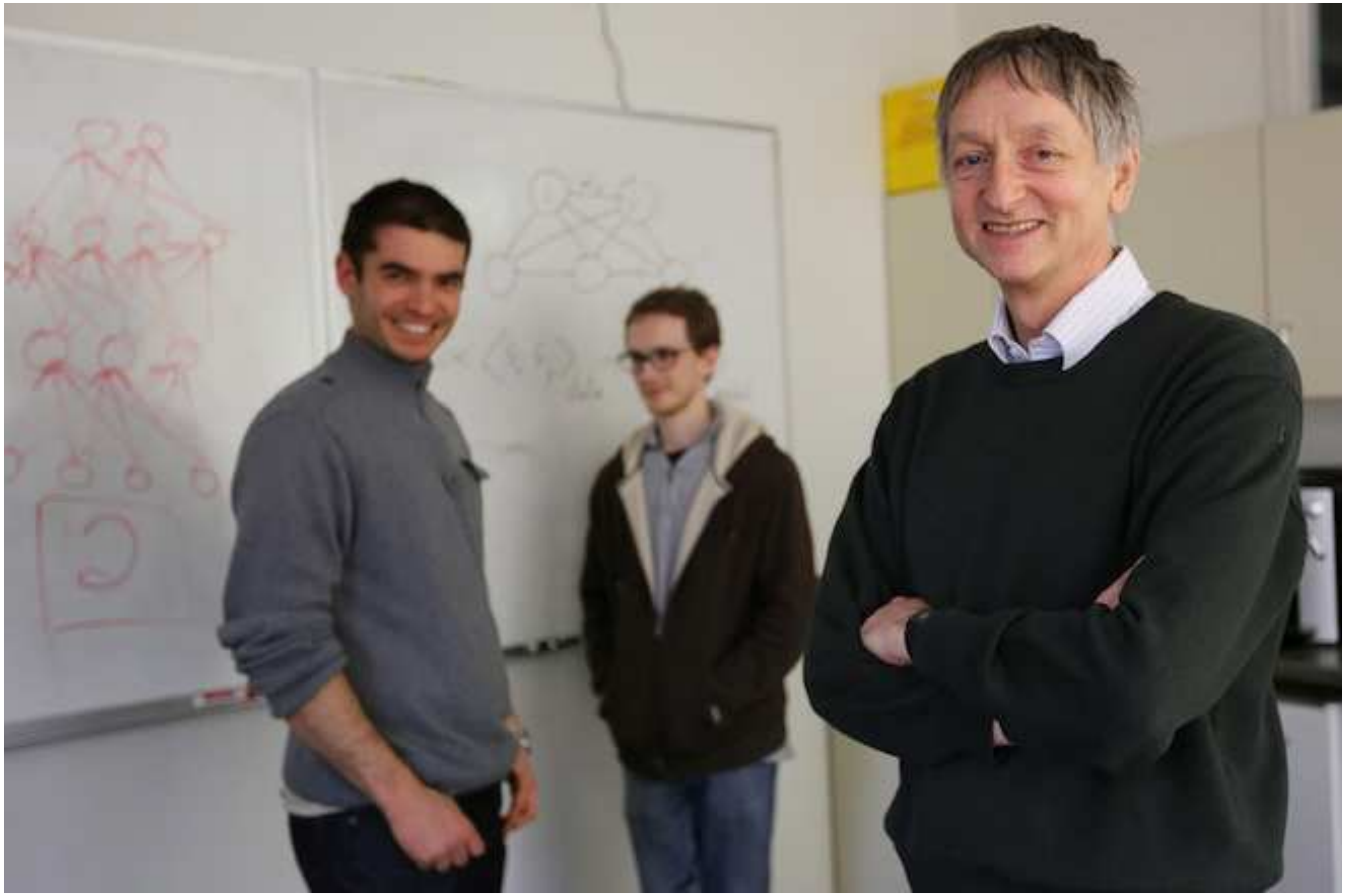
Universal function approximator in theory, but in practice three-layer ANNs could often not successfully solve complex problems.

Fundamentals of Neural Networks (cont'd)

Breakthrough again with so-called **deep neural networks** or **DNN**s, widely known since the 2012 NIPS-paper by Alex Krizhevsky, Ilya Sutskever & Geoffrey Hinton.

Fundamentals of Neural Networks (cont'd)

Breakthrough again with so-called **deep neural networks** or **DNN**s, widely known since the 2012 NIPS-paper by Alex Krizhevsky, Ilya Sutskever & Geoffrey Hinton.



https://www.wired.com/wp-content/uploads/blogs/wiredenterprise/wp-content/uploads/2013/03/hinton1.jpg

Fundamentals of Neural Networks (cont'd)

Breakthrough again with so-called **deep neural networks** or **DNN**s, widely known since the 2012 NIPS-paper by Alex Krizhevsky et al.

DNN: loose terminology to refer to networks with at least two hidden or intermediate layers, typically at least five to ten (or up to dozens):

Fundamentals of Neural Networks (cont'd)

Breakthrough again with so-called **deep neural networks** or **DNN**s, widely known since the 2012 NIPS-paper by Alex Krizhevsky et al.

DNN: loose terminology to refer to networks with at least two hidden or intermediate layers, typically at least five to ten (or up to dozens):

1. Massive increase in labelled training data ("the internet"),
2. computing power (GPUs),
3. simple non-linearity (ReLU) instead of sigmoid,
4. convolutional rather than fully connected layers,

and

5. *weight sharing* across deep layers

appear to be the critical ingredients for the current success of DNNs, and makes them the current method of choice in ML, particular in application.

Fundamentals of Neural Networks (cont'd)

Breakthrough again with so-called **deep neural networks** or **DNN**s, widely known since the 2012 NIPS-paper by Alex Krizhevsky et al.

DNN: loose terminology to refer to networks with at least two hidden or intermediate layers, typically at least five to ten (or up to dozens):

1. Massive increase in labelled training data ("the internet"),
2. computing power (GPUs),
3. simple non-linearity (ReLU) instead of sigmoid,
4. convolutional rather than fully connected layers,

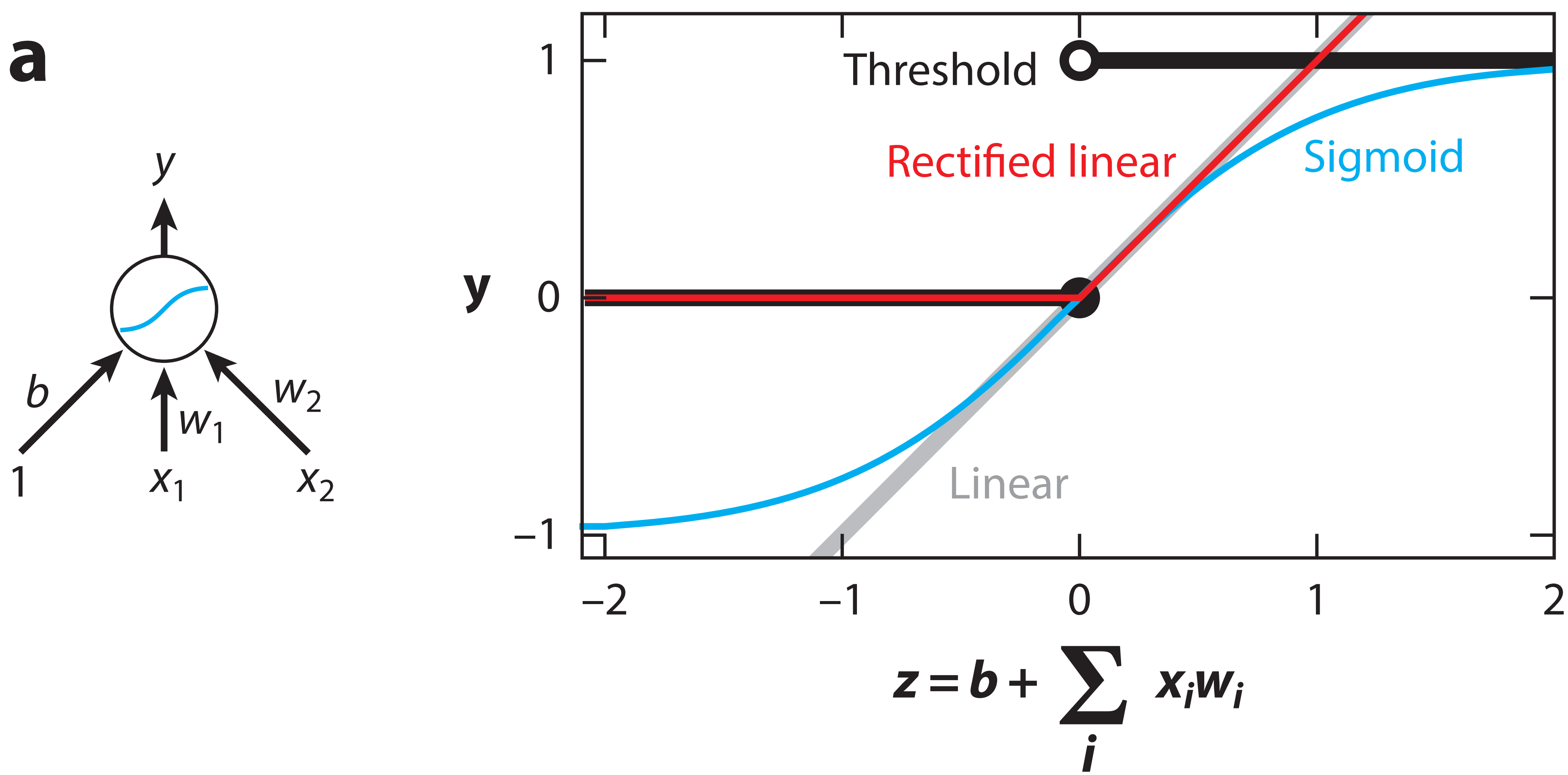
and

5. *weight sharing* across deep layers

appear to be the critical ingredients for the current success of DNNs, and makes them the current method of choice in ML, particular in application.

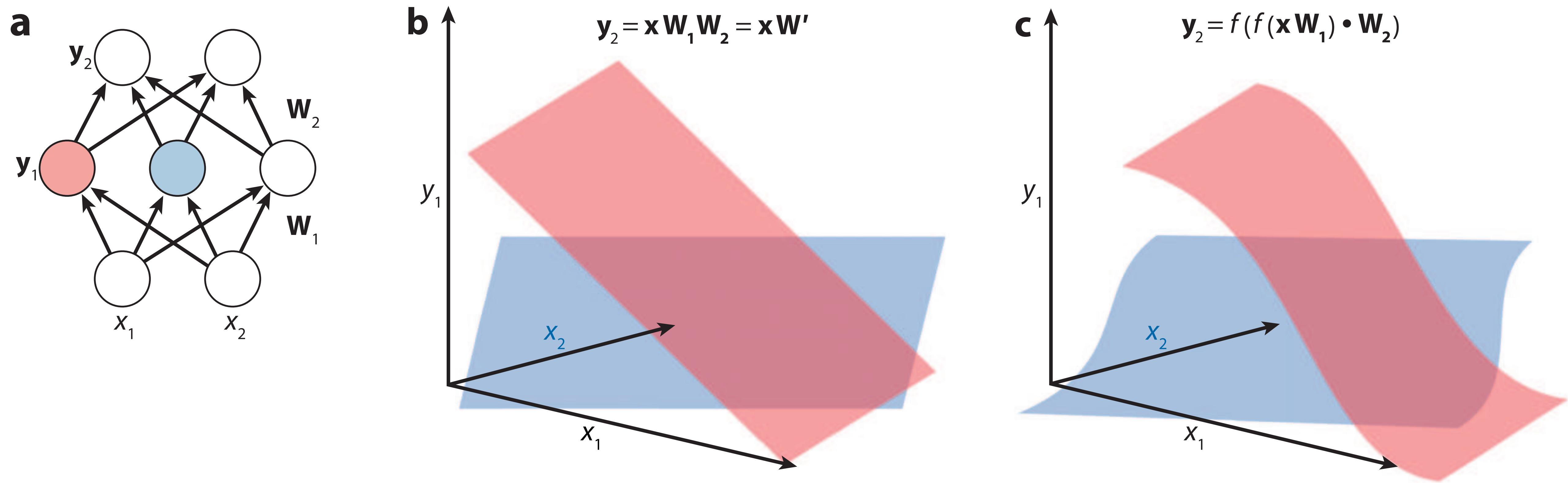
At least superficially DNNs appear to be similar to the human object recognition system: convolutions ("filters", "receptive fields") followed by non-linearities and pooling is thought to be the canonical computation of cortex, at least within sensory areas.

Fundamentals of Neural Networks



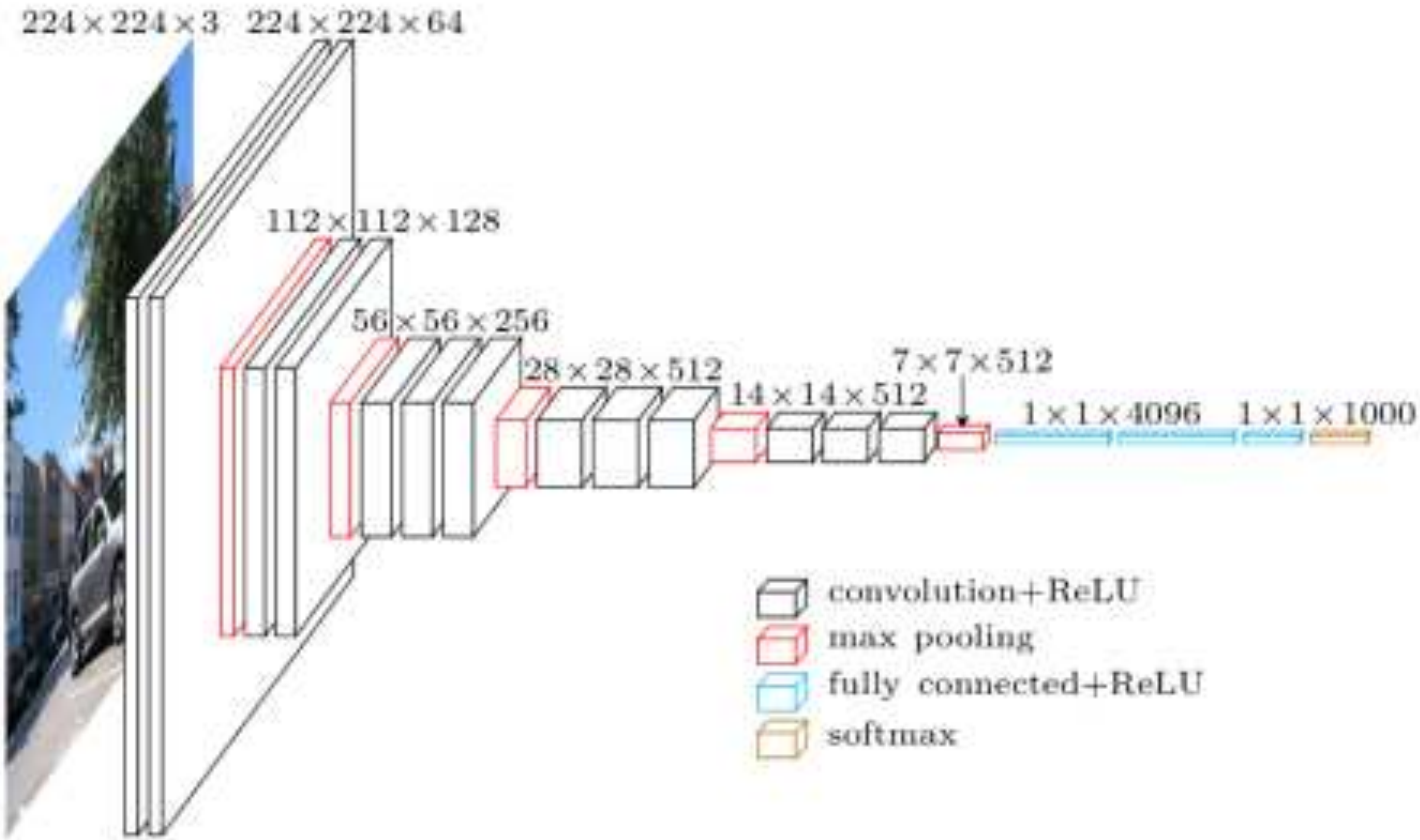
Kriegeskorte (2015)

Fundamentals of Neural Networks



Kriegeskorte (2015)

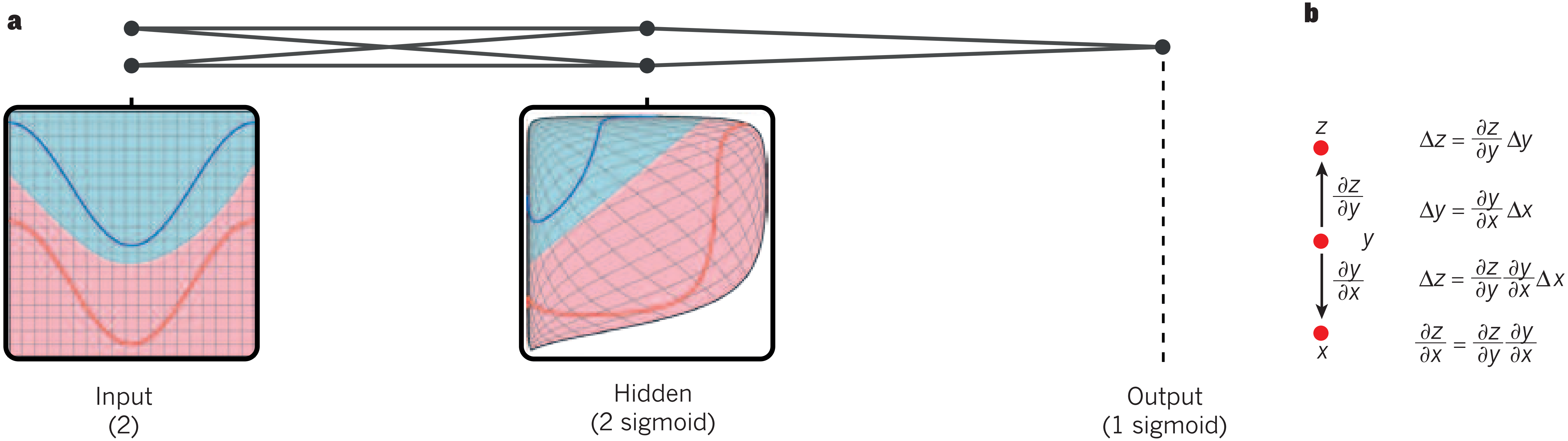
Example: VGG-16



https://www.cs.toronto.edu/~frossard/post/vgg16/#architecture

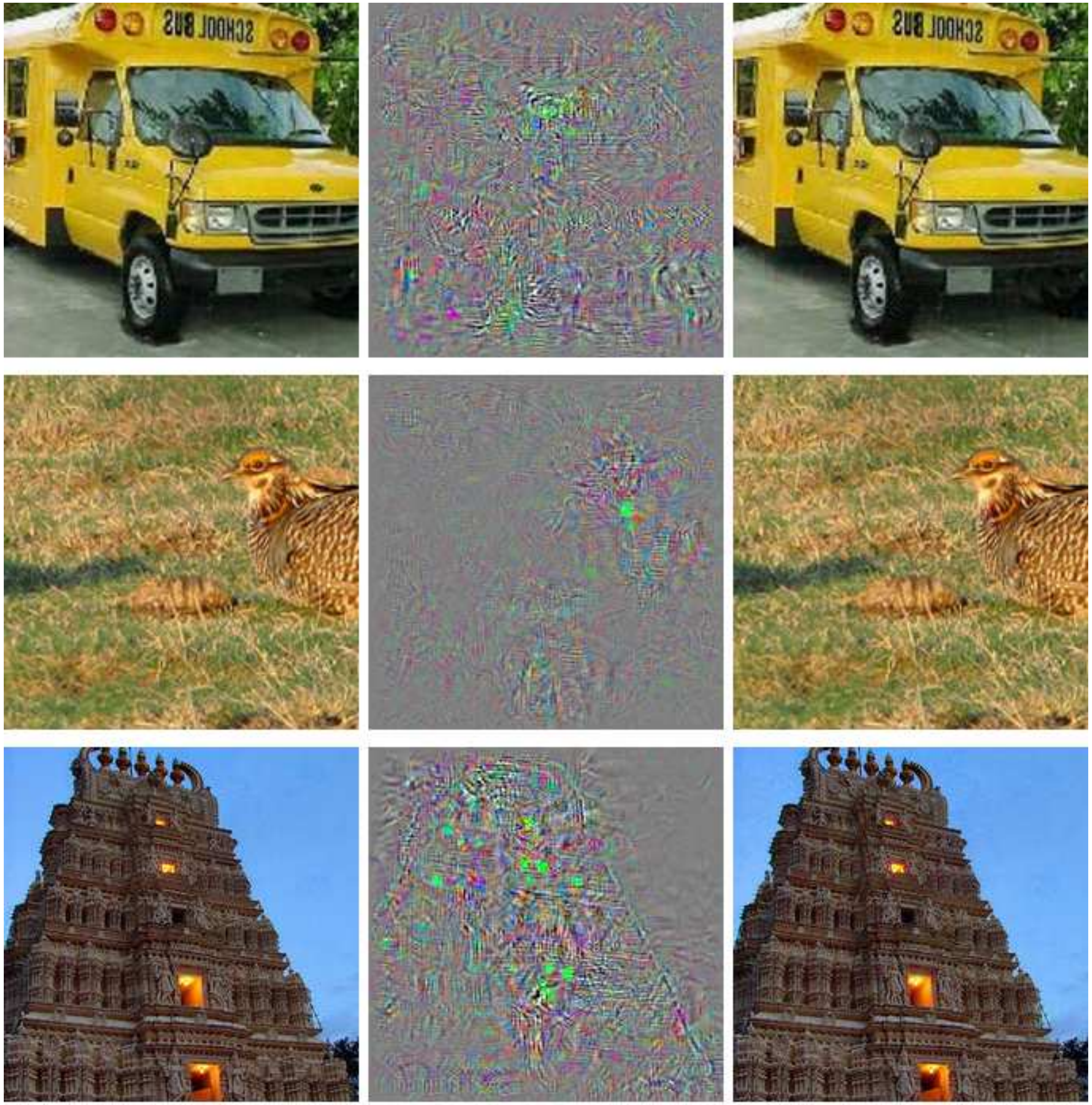
VGG16 by Simonyan & Zisserman (2014); 92.7% top-5 test accuracy on ImageNet http://scs.ryerson.ca/~aharley/vis/conv/flat.html

Deep Neural Networks (DNNs)



4

Adversarial attacks?



Szegedy et al. (2014)

Adversarial examples? (cont'd)

Reese
Witherspoon

Sharif et al. (2016)

Adversarial examples? (cont'd)

Reese
Witherspoon

Sharif et al. (2016)

Adversarial examples? (cont'd)

Reese
Witherspoon

Russel
Crowe

Sharif et al. (2016)

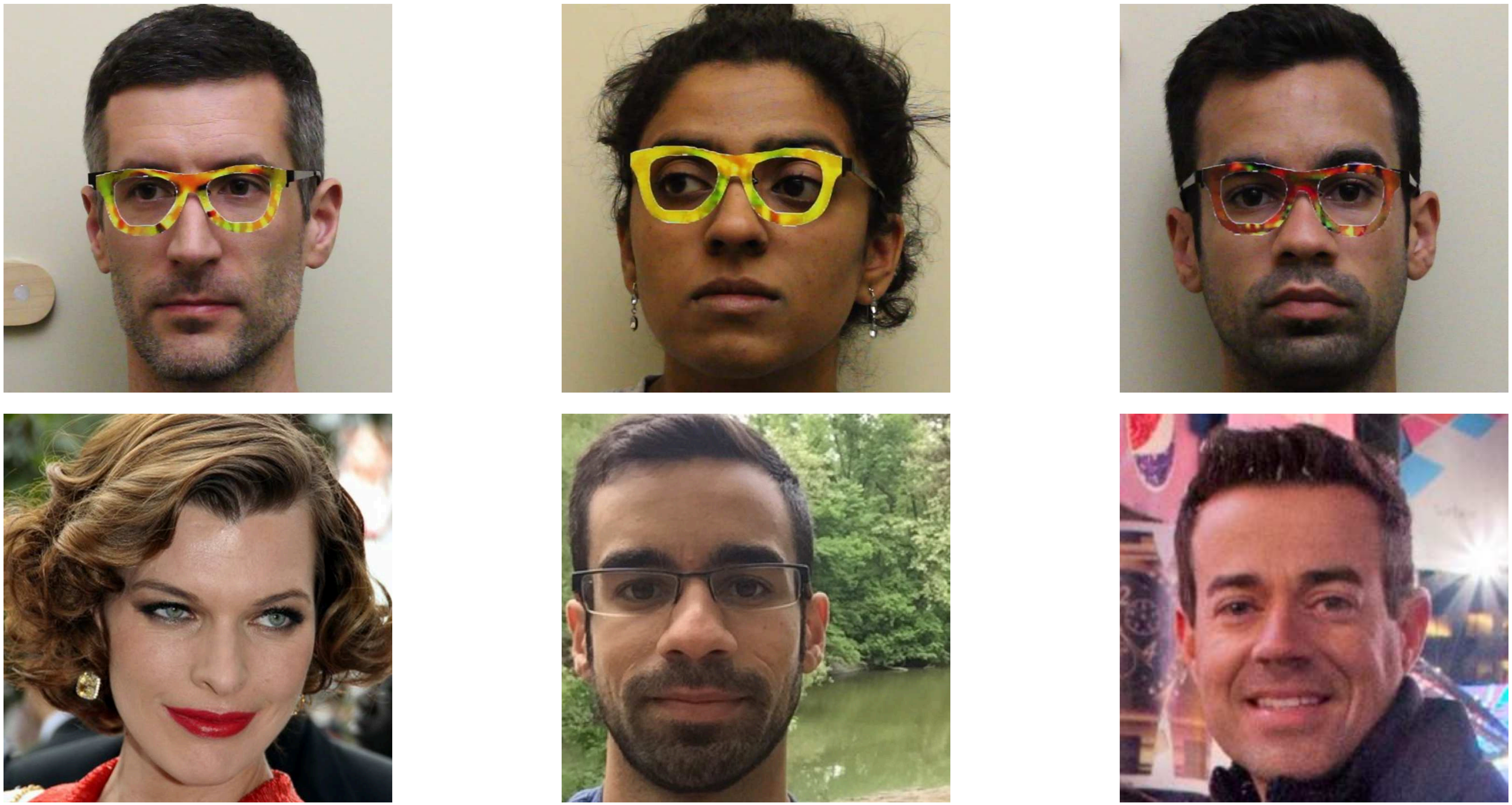
Adversarial examples? (cont'd)

Reese
Witherspoon

Russel
Crowe

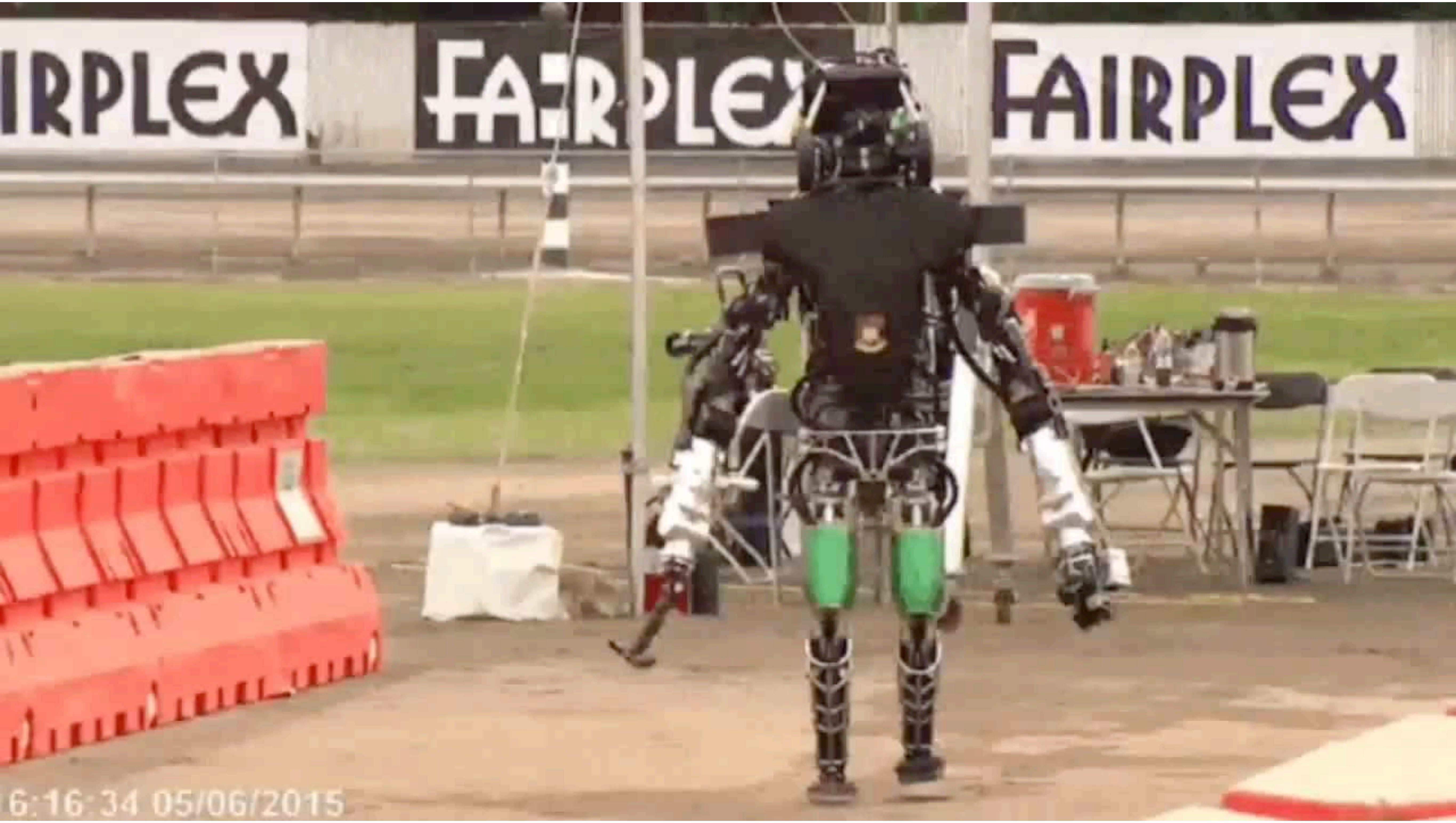
Sharif et al. (2016)

Adversarial examples? (cont'd)

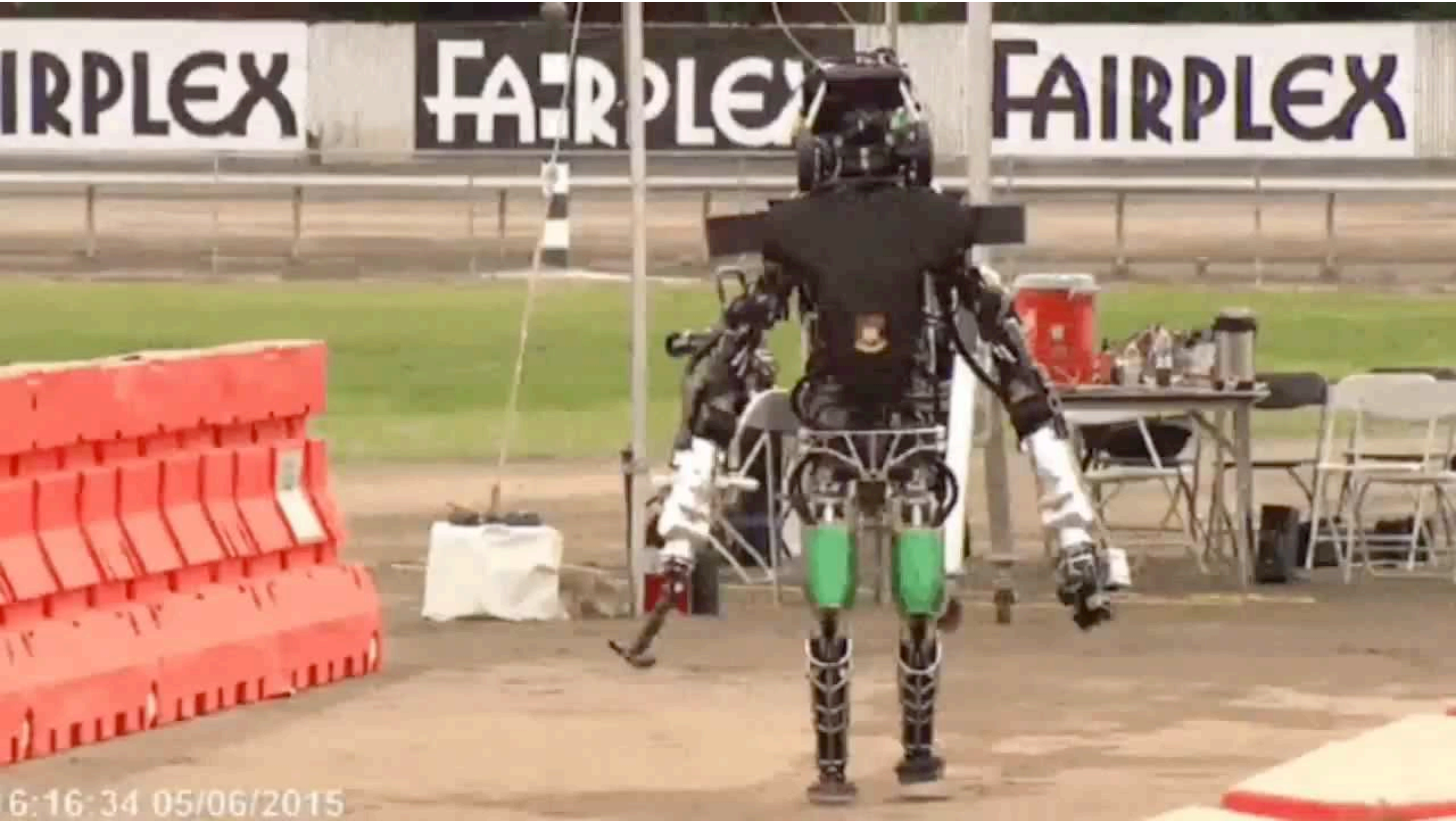


Sharif et al. (2016)

DARPA Challenge 2015



DARPA Challenge 2015



Boston Dynamics 2017

Boston Dynamics 2017

Human versus artificial intelligence

We learn unsupervised or semi-supervised, sometimes reinforcement, very rarely supervised (school, University) – all successful AI is currently supervised only, i.e. only when the correct answer is known!

We can do lots of things using the same network (or a set of closely coupled networks) — all DNNs are typically only good at one or few tasks.

5

Gesellschaftliche Herausforderungen

Arbeitsbedingungen und Arbeitsmarkt:
Einsatz von Technologie macht die Arbeit ''einfacher'' – typischerweise fällt die Notwendigkeit einer Lehre oder Ausbildung weg.
Die Folge sind sinkende Löhne ... schließlich kann ''jeder'' die Arbeit machen.

Arbeitslosigkeit?

Arbeitslosigkeit?

Autonome Fahrzeuge – womöglich kurz nach der Erlaubnis, solche Fahrzeuge im Straßenverkehr zu haben, die Pflicht, nur noch damit zu fahren.

Arbeitslosigkeit?

Autonome Fahrzeuge – womöglich kurz nach der Erlaubnis, solche Fahrzeuge im Straßenverkehr zu haben, die Pflicht, nur noch damit zu fahren.

`540.000`	Berufskraftfahrer in Deutschland (Stand 2013)
`250.000`	Taxifahrerlaubnisse (Stand 2017)
`25.000`	Lokführer (Stand 2017)
`815.000`	Arbeitsplätze gefährdet (Quote von 5.8% auf 8.1%)

Arbeitslosigkeit?

Autonome Fahrzeuge – womöglich kurz nach der Erlaubnis, solche Fahrzeuge im Straßenverkehr zu haben, die Pflicht, nur noch damit zu fahren.

540.000	Berufskraftfahrer in Deutschland (Stand 2013)
250.000	Taxifahrerlaubnisse (Stand 2017)
25.000	Lokführer (Stand 2017)
815.000	Arbeitsplätze gefährdet (Quote von 5.8% auf 8.1%)

Roboter in der Post? Abfallwirtschaft? Logistik?
Deutsche Post DHL hat 211.000 Mitarbeiter in Deutschland (Stand 2016), in der Ver- und Entsorgung arbeiteten 2014 ca. 155.000 Menschen, als Reinigungskräfte 2014 offiziell fast 760.000; Amazon beschäftigt alleine in D 23.000 Menschen in Logistik-Zentren: 1.150.000 Arbeitsplätze!

Arbeitslosigkeit?

Autonome Fahrzeuge – womöglich kurz nach der Erlaubnis, solche Fahrzeuge im Straßenverkehr zu haben, die Pflicht, nur noch damit zu fahren.

540.000	Berufskraftfahrer in Deutschland (Stand 2013)
250.000	Taxifahrerlaubnisse (Stand 2017)
25.000	Lokführer (Stand 2017)
815.000	Arbeitsplätze gefährdet (Quote von 5.8% auf 8.1%)

Roboter in der Post? Abfallwirtschaft? Logistik?
Deutsche Post DHL hat 211.000 Mitarbeiter in Deutschland (Stand 2016), in der Ver- und Entsorgung arbeiteten 2014 ca. 155.000 Menschen, als Reinigungskräfte 2014 offiziell fast 760.000; Amazon beschäftigt alleine in D 23.000 Menschen in Logistik-Zentren: 1.150.000 Arbeitsplätze!

Humanoide Roboter in der Pflege?
2014 arbeiteten in der Alten- und Krankenpflege in D über 900.000 Menschen … .

Gesellschaftliche Herausforderungen

Arbeitsbedingungen und Arbeitsmarkt:
Einsatz von Technologie macht die Arbeit ''einfacher'' – typischerweise fällt die Notwendigkeit einer Lehre oder Ausbildung weg.
Die Folge sind sinkende Löhne ... schließlich kann ''jeder'' die Arbeit machen.

Politik und Gesellschaft:
Leben in der selben Wirklichkeit? Personalisierte Information in sozialen Medien und der Verlust breit und kontrovers informierender Quellen – weit verbreiteter Konsum von Propaganda.

Propaganda

Propaganda ist der Versuch der gezielten Beeinflussung des Denkens, Handelns und Fühlens von Menschen. Wer Propaganda betreibt, verfolgt damit immer ein bestimmtes Interesse. … **Charakteristisch für Propaganda ist, dass sie die verschiedenen Seiten einer Thematik nicht darlegt und Meinung und Information vermischt.** Wer Propaganda betreibt, möchte nicht diskutieren und mit Argumenten überzeugen, sondern mit allen Tricks die Emotionen und das Verhalten der Menschen beeinflussen, beispielsweise indem sie diese ängstigt, wütend macht oder ihnen Verheißungen ausspricht. Propaganda nimmt dem Menschen das Denken ab und gibt ihm stattdessen das Gefühl, mit der übernommenen Meinung richtig zu liegen.

Quelle: Bundeszentrale für politische Bildung
www.bpb.de

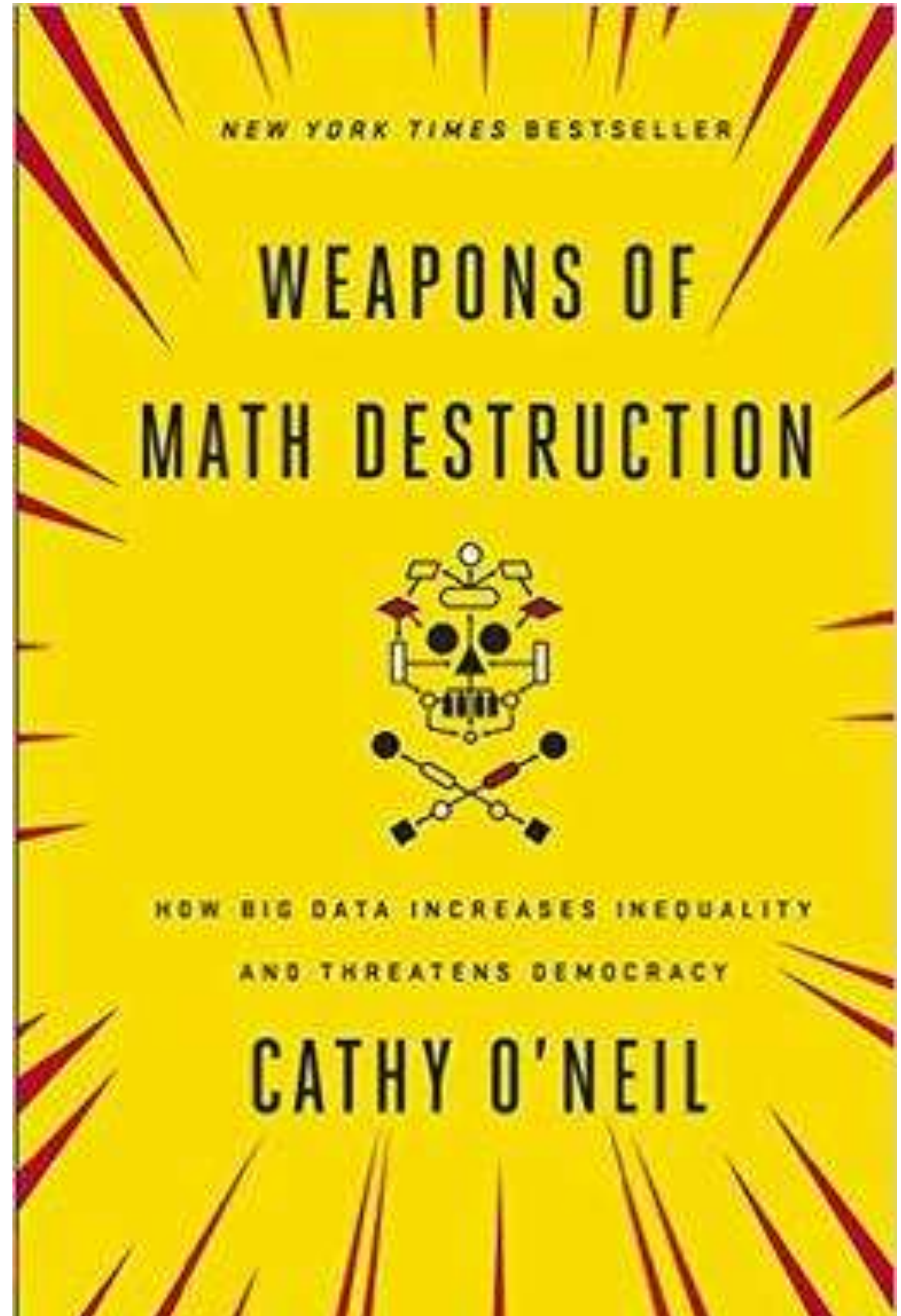
Gesellschaftliche Herausforderungen

Arbeitsbedingungen und Arbeitsmarkt:
Einsatz von Technologie macht die Arbeit ''einfacher'' – typischerweise fällt die Notwendigkeit einer Lehre oder Ausbildung weg.
Die Folge sind sinkende Löhne ... schließlich kann ''jeder'' die Arbeit machen.

Politik und Gesellschaft:
Leben in der selben Wirklichkeit? Personalisierte Information in sozialen Medien und der Verlust breit und kontrovers informierender Quellen – weit verbreiteter Konsum von Propaganda.
Privatsphäre? Veränderung (zwischenmenschlicher) Kommunikation?

Weapons of Mass Destruction (WMDs)

https://www.wired.com/images_blogs/dangerroom/2011/03/powell_un_anthrax.jpg


NEW YORK TIMES BESTSELLER
WEAPONS OF
MATH DESTRUCTION
HOW BIG DATA INCREASES INEQUALITY
AND THREATENS DEMOCRACY
CATHY O'NEIL

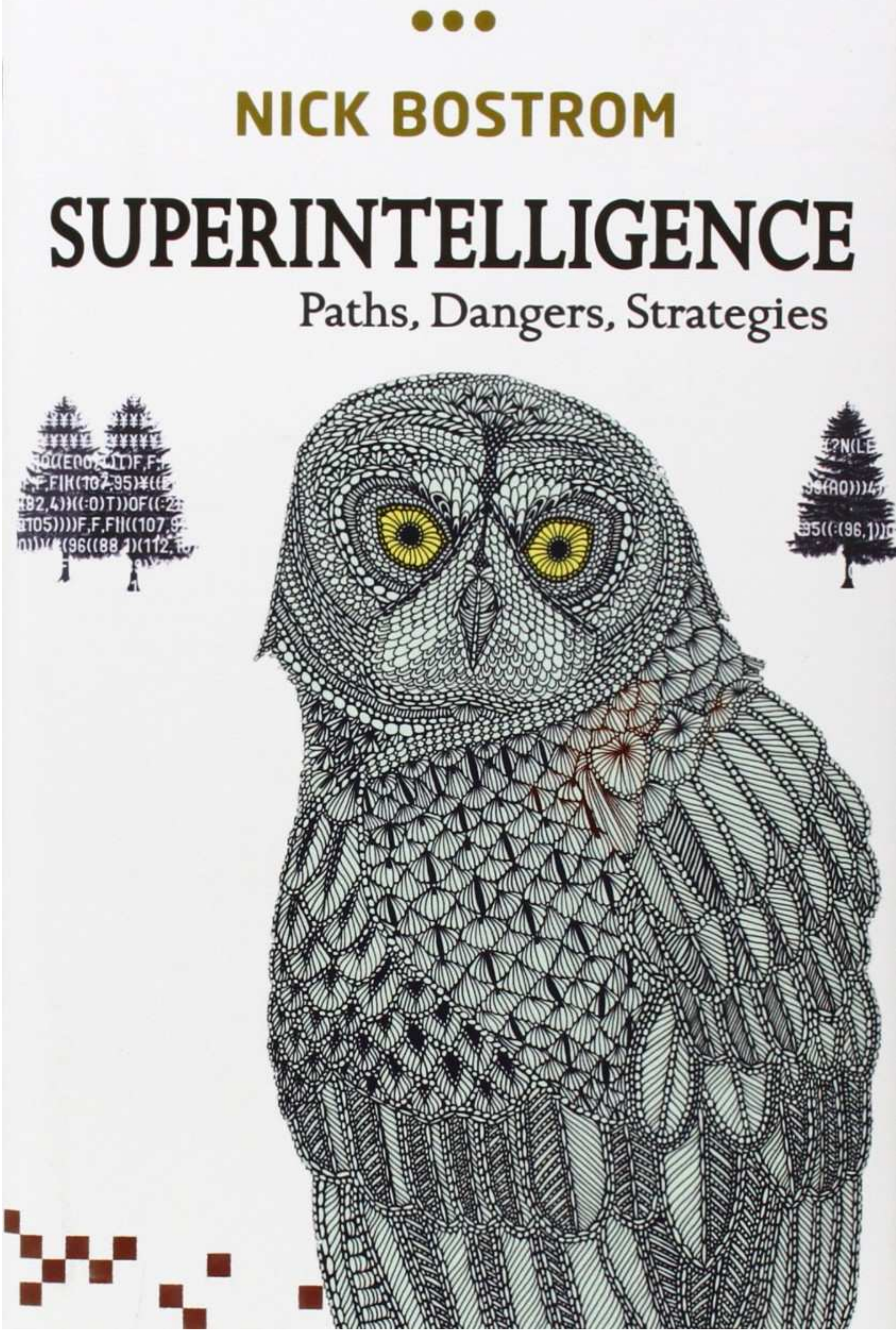
Gesellschaftliche Herausforderungen

Arbeitsbedingungen und Arbeitsmarkt:
Einsatz von Technologie macht die Arbeit ''einfacher'' – typischerweise fällt die Notwendigkeit einer Lehre oder Ausbildung weg.
Die Folge sind sinkende Löhne ... schließlich kann ''jeder'' die Arbeit machen.

Politik und Gesellschaft:
Leben in der selben Wirklichkeit? Personalisierte Information in sozialen Medien und der Verlust breit und kontrovers informierender Quellen – weit verbreiteter Konsum von Propaganda.
Privatsphäre? Veränderung (zwischenmenschlicher) Kommunikation?

Naïver Glaube an die Objektivität von Algorithmen
... und Ranglisten, die Vermessung und Quantifizierung des Lebens:
China, z.B., plant das *Social Credit System* einzuführen.

https://de.wikipedia.org/wiki/Nick_Bostrom



Gesellschaftliche Herausforderungen

Arbeitsbedingungen und Arbeitsmarkt:
Einsatz von Technologie macht die Arbeit "einfacher" – typischerweise fällt die Notwendigkeit einer Lehre oder Ausbildung weg.
Die Folge sind sinkende Löhne ... schließlich kann "jeder" die Arbeit machen.

Politik und Gesellschaft:
Leben in der selben Wirklichkeit? Personalisierte Information in sozialen Medien und der Verlust breit und kontrovers informierender Quellen – weit verbreiteter Konsum von Propaganda.
Privatsphäre? Veränderung (zwischenmenschlicher) Kommunikation?

Naïver Glaube an die Objektivität von Algorithmen
... und Ranglisten, die Vermessung und Quantifizierung des Lebens: China plant das *Social Credit System* einzuführen.

Doomsday-Szenarien
Kommt die Singularität? Wenn ja: Garten Eden oder Hölle?

Doomsday-Videos to watch

Google's Geoffrey Hinton - "There's no reason to think computers won't get much smarter than us" (10 mins): https://www.youtube.com/watch?v=p6lM3bh-npg

Demis Hassabis, CEO, DeepMind Technologies - The Theory of Everything (16 mins): https://www.youtube.com/watch?v=rbsqaJwpu6A

Nick Bostrom, What happens when our computers get smarter than we are? (17 mins): https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are

Why Elon Musk is worried about artificial intelligence (3 mins) https://www.youtube.com/watch?v=US95slMMQis

Thanks

Felix Wichmann

Neural Information Processing Group and
Bernstein Center for Computational Neuroscience,
Eberhard Karls Universität Tübingen

Max Planck Institute for Intelligent Systems, Tübingen