See eye to eye!

Ricardo Marroquim

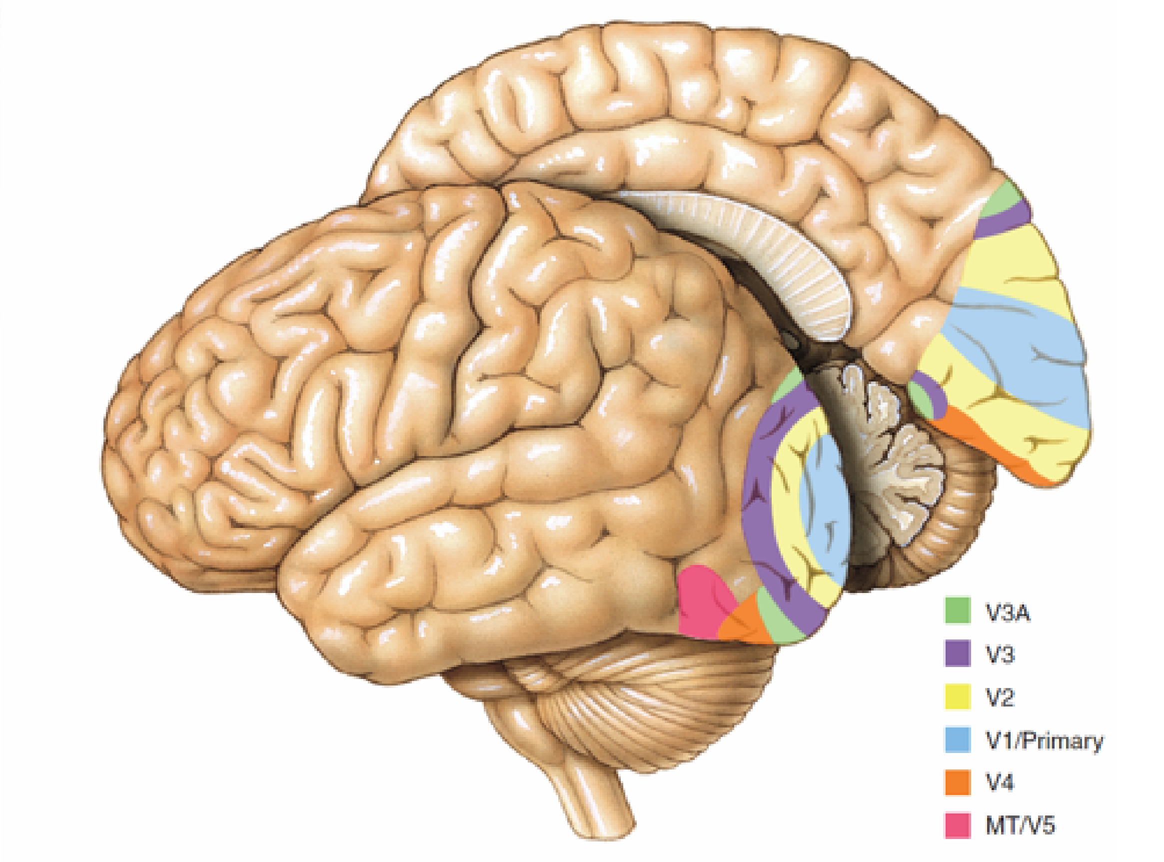
www.lcg.ufrj.br/~marroquim

Laboratório de Computação Gráfica

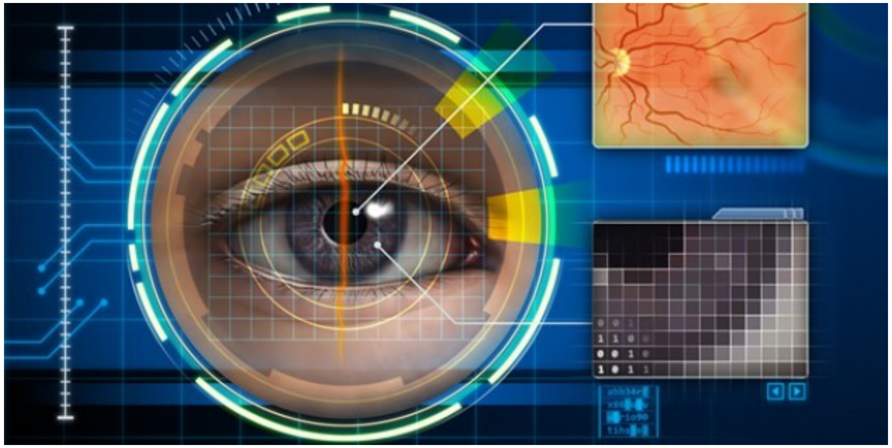
PESC
Programa de Engenharia de Sistemas e Computação

SEMANA PESC – 21-25 de novembro 2016

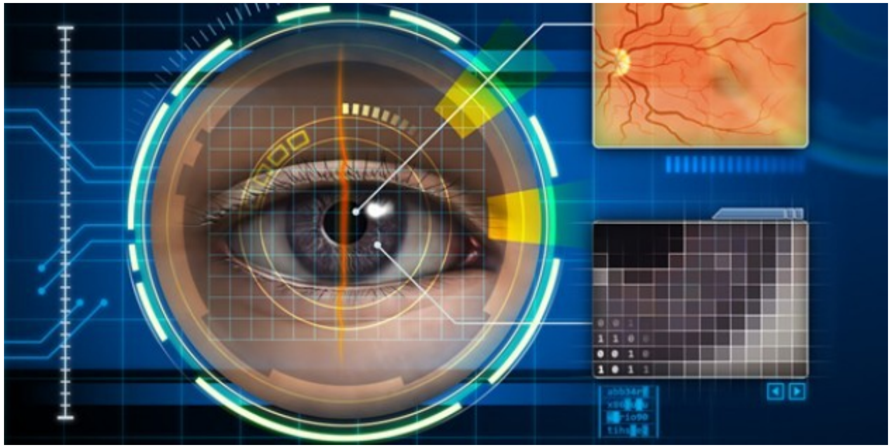
how do we see?

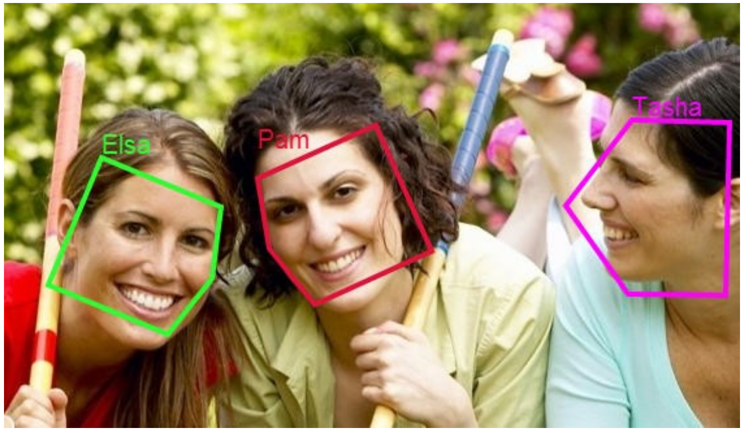


how computers see?

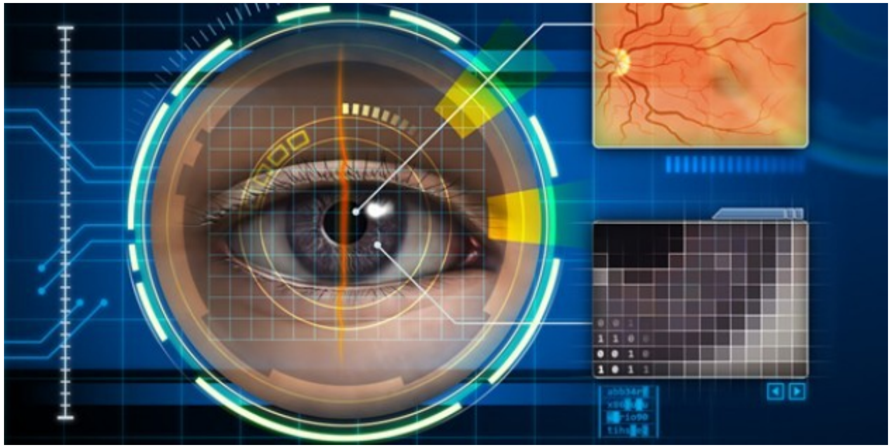


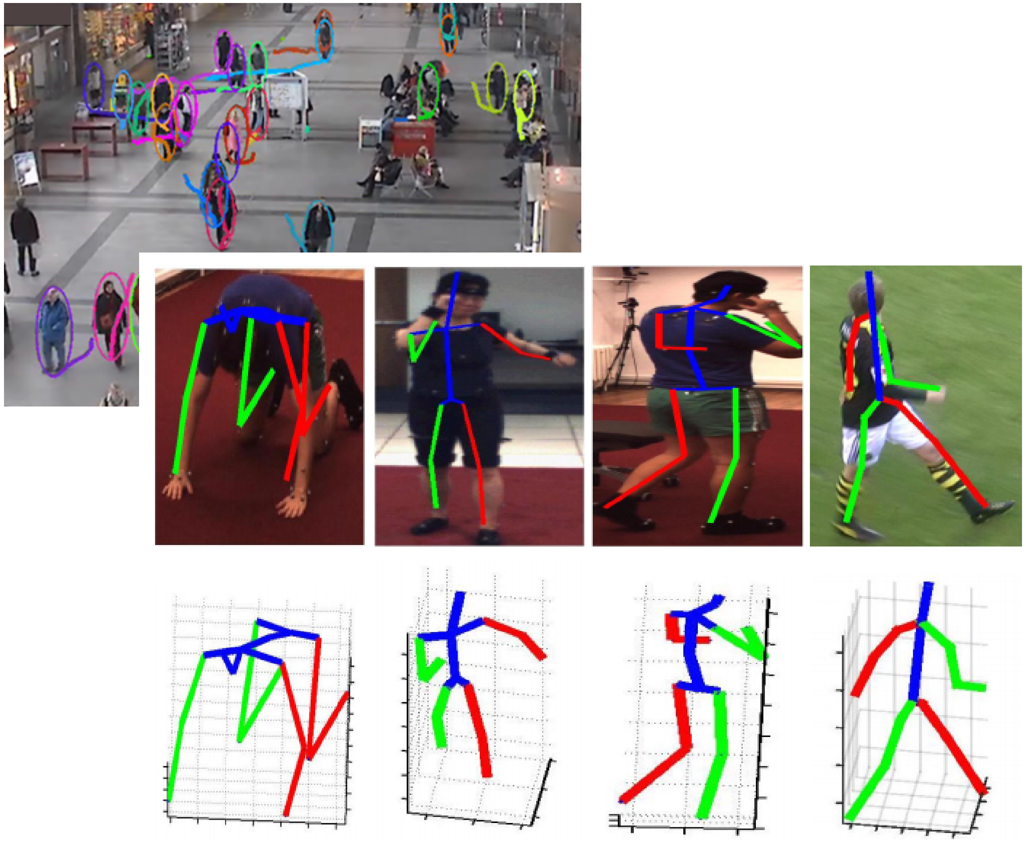
computer vision

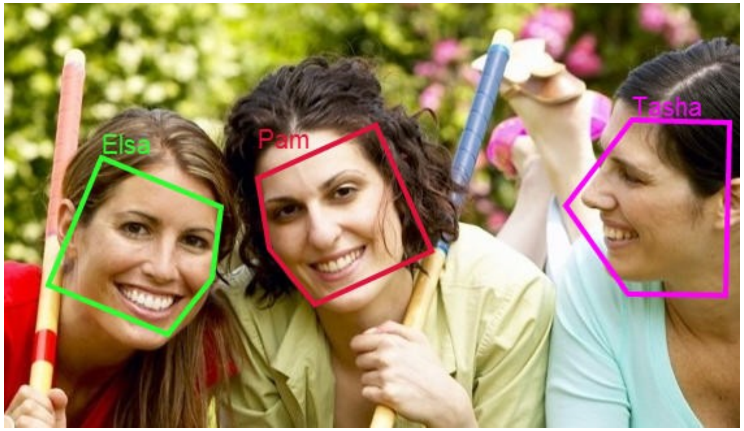




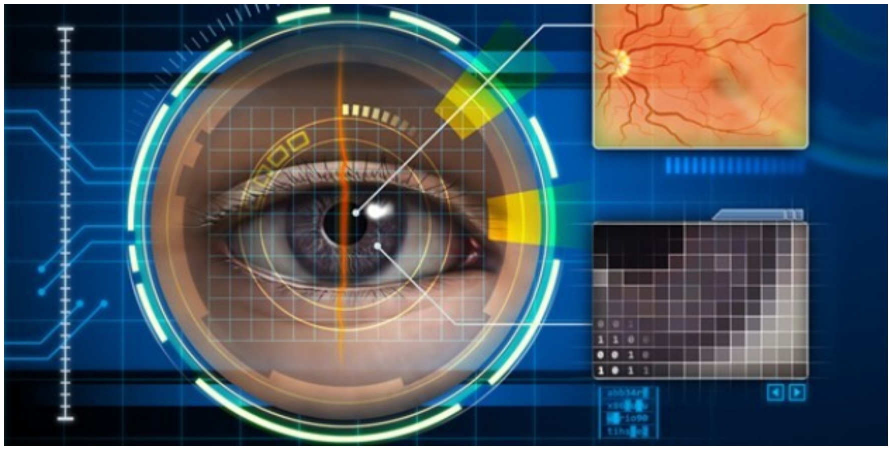
computer vision

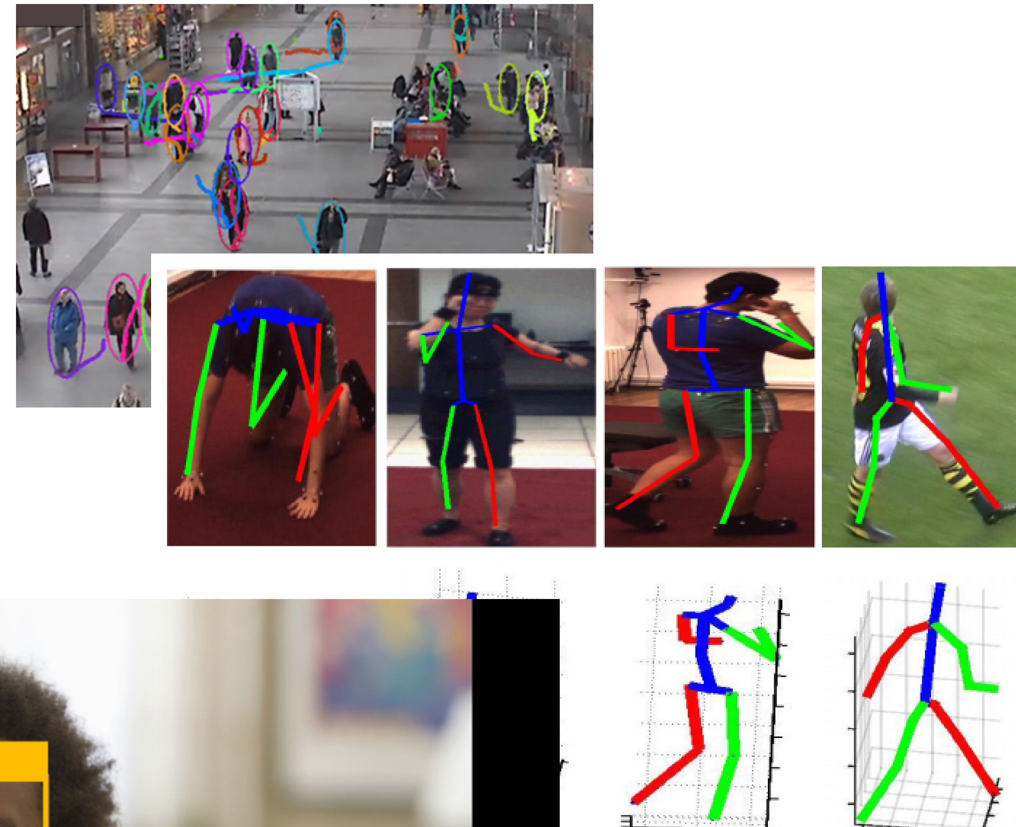


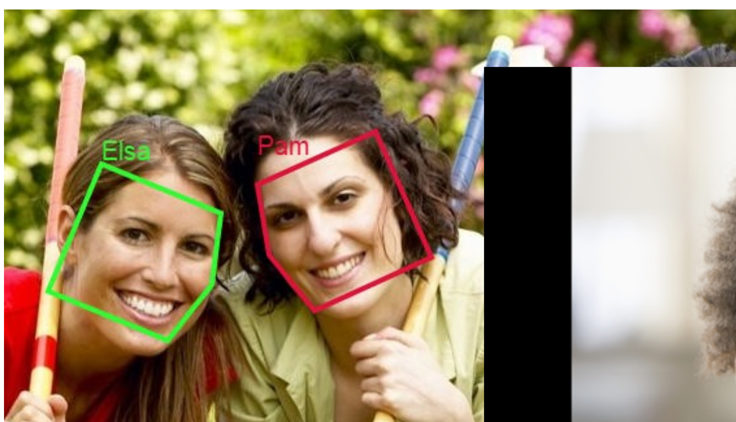


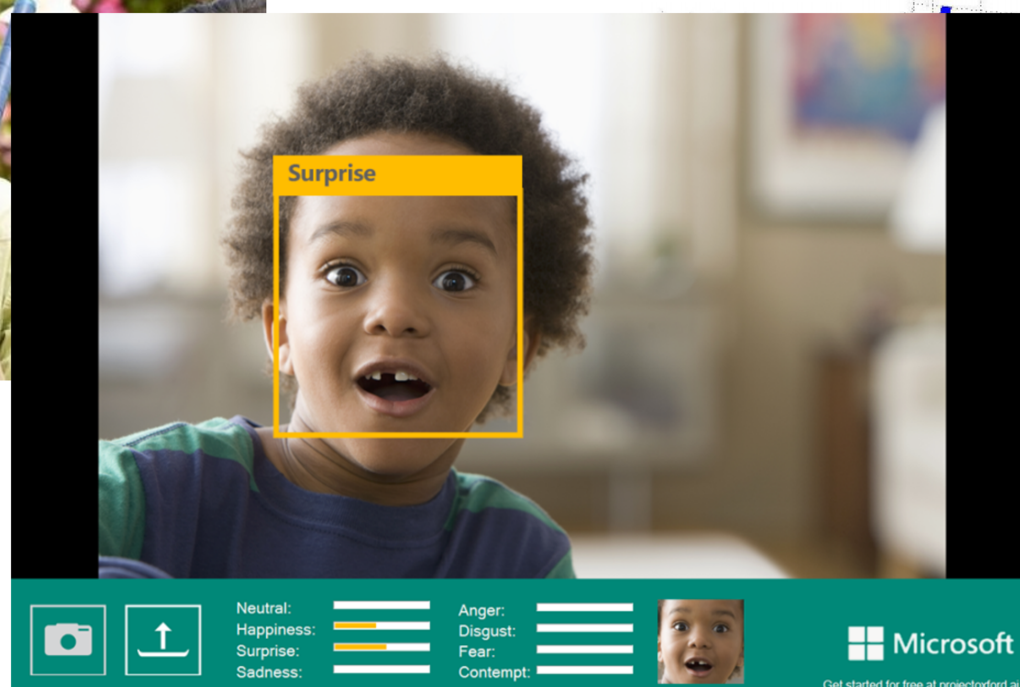


computer vision

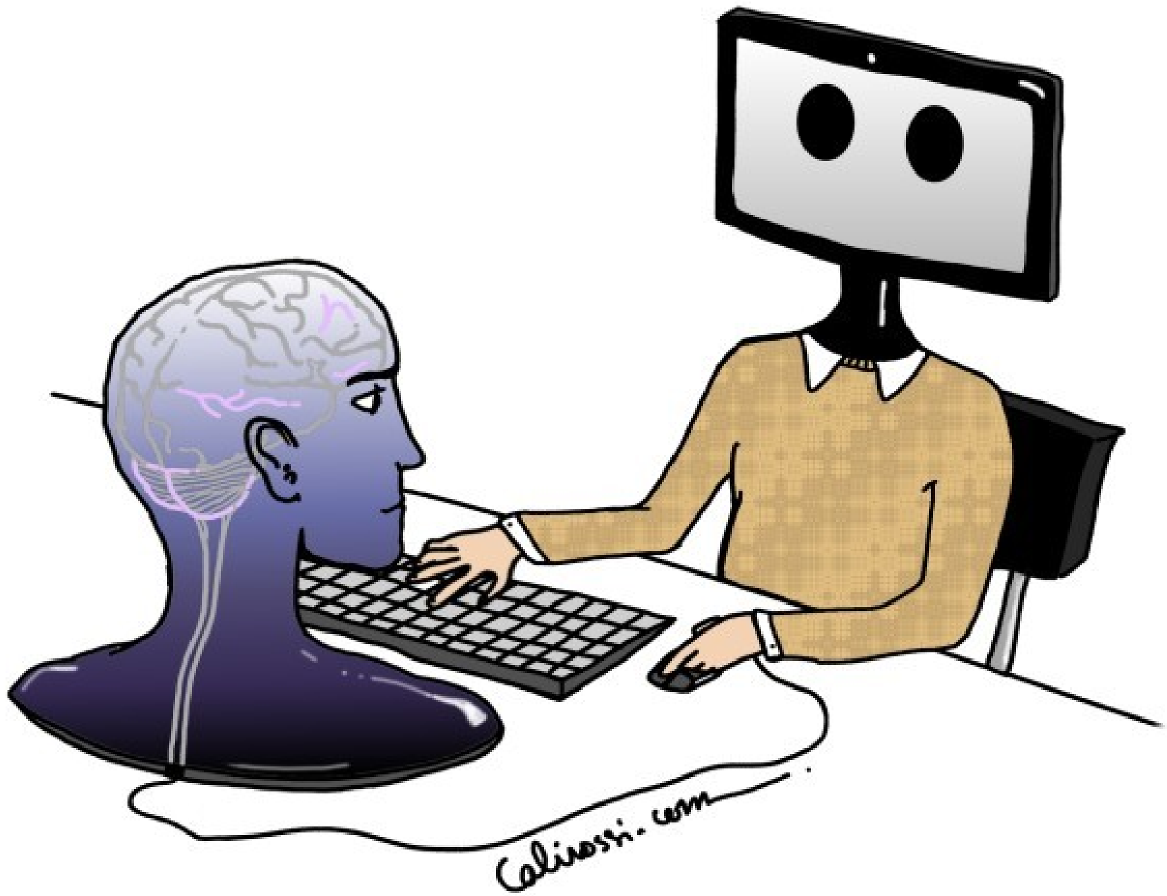








humans vs computers



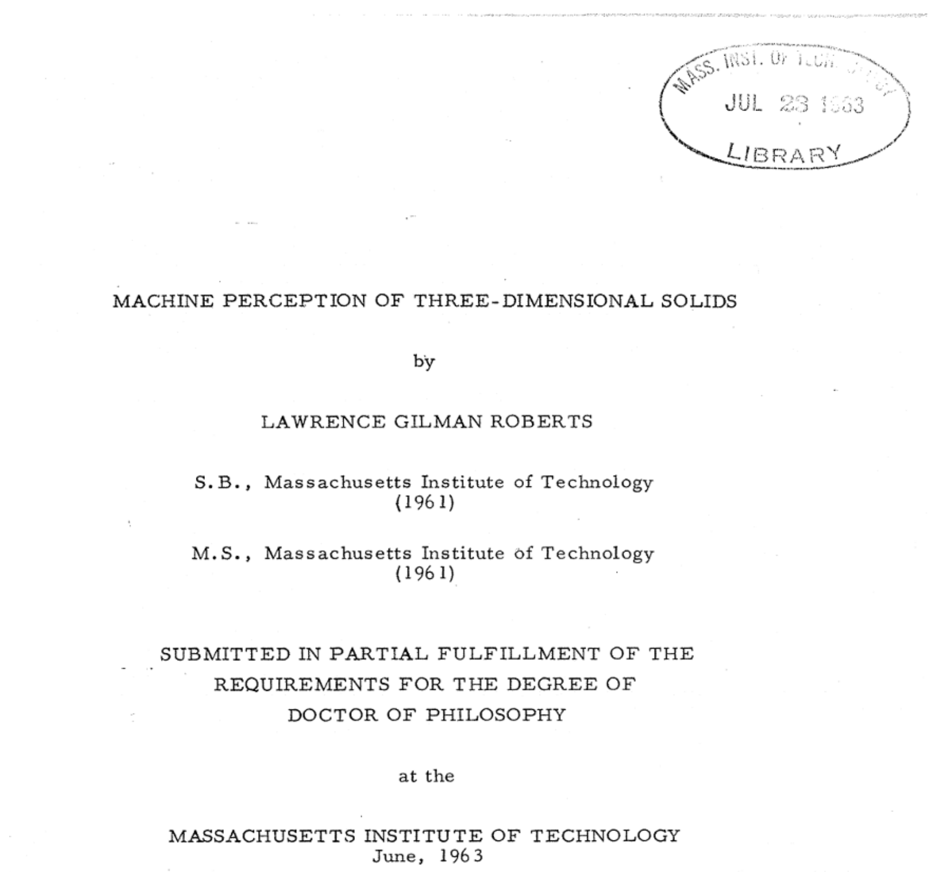
http://benedicterossi.com/

Marvin Minsky

- pioneer: Perceptrons, Logo turtle, Head-mounted display …
- 1969: Turing Award

Larry Roberts

- 1963 - PhD Thesis: Machine Perception of Three-Dimensional Solids



MASS. INST. OF TECH.
JUL 23 1963
LIBRARY

MACHINE PERCEPTION OF THREE-DIMENSIONAL SOLIDS

by

LAWRENCE GILMAN ROBERTS

S.B., Massachusetts Institute of Technology
(1961)

M.S., Massachusetts Institute of Technology
(1961)

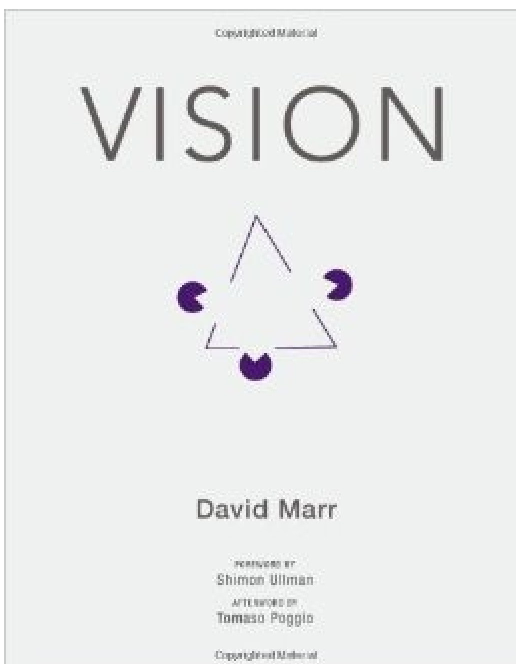
SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

at the

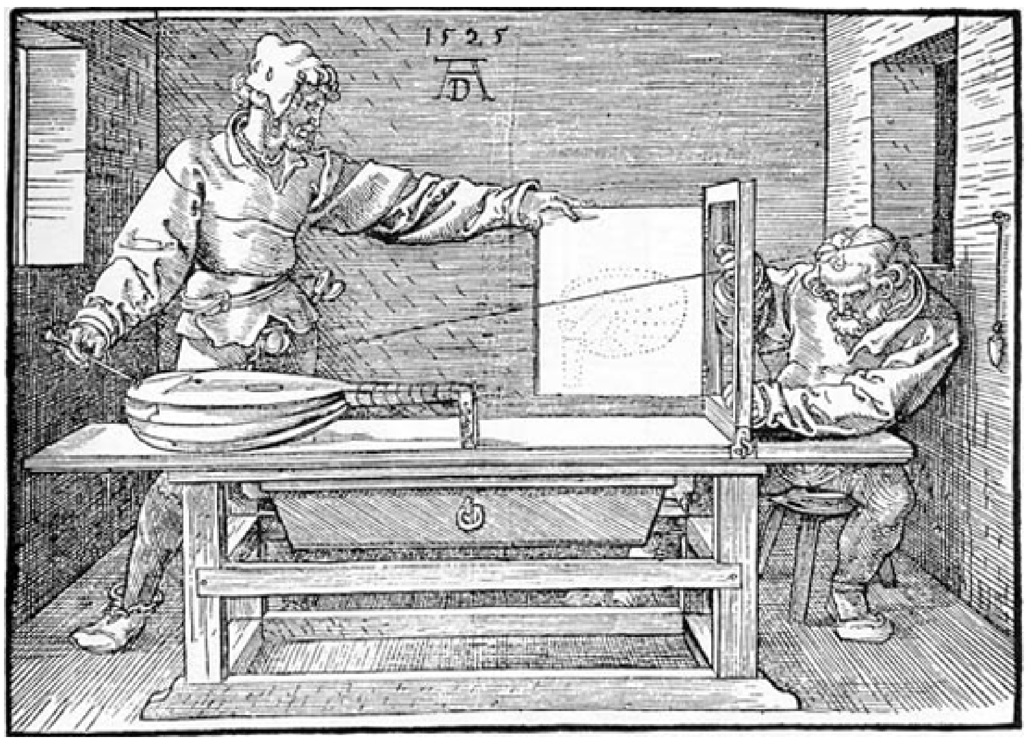
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June, 1963

David Marr

- 1982 - David Marr - Vision: A Computational Investigation into the Human Representation and Processing of Visual Information



projective geometry



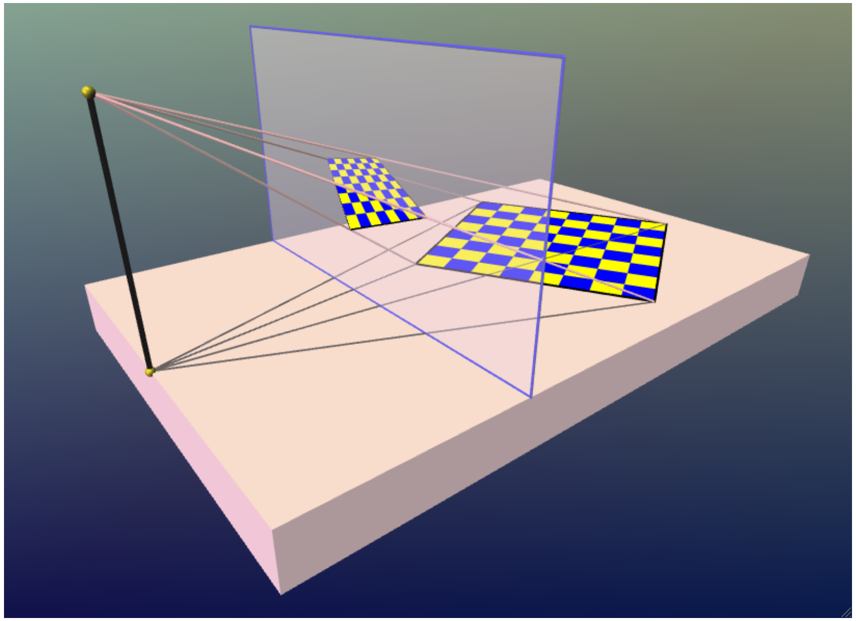
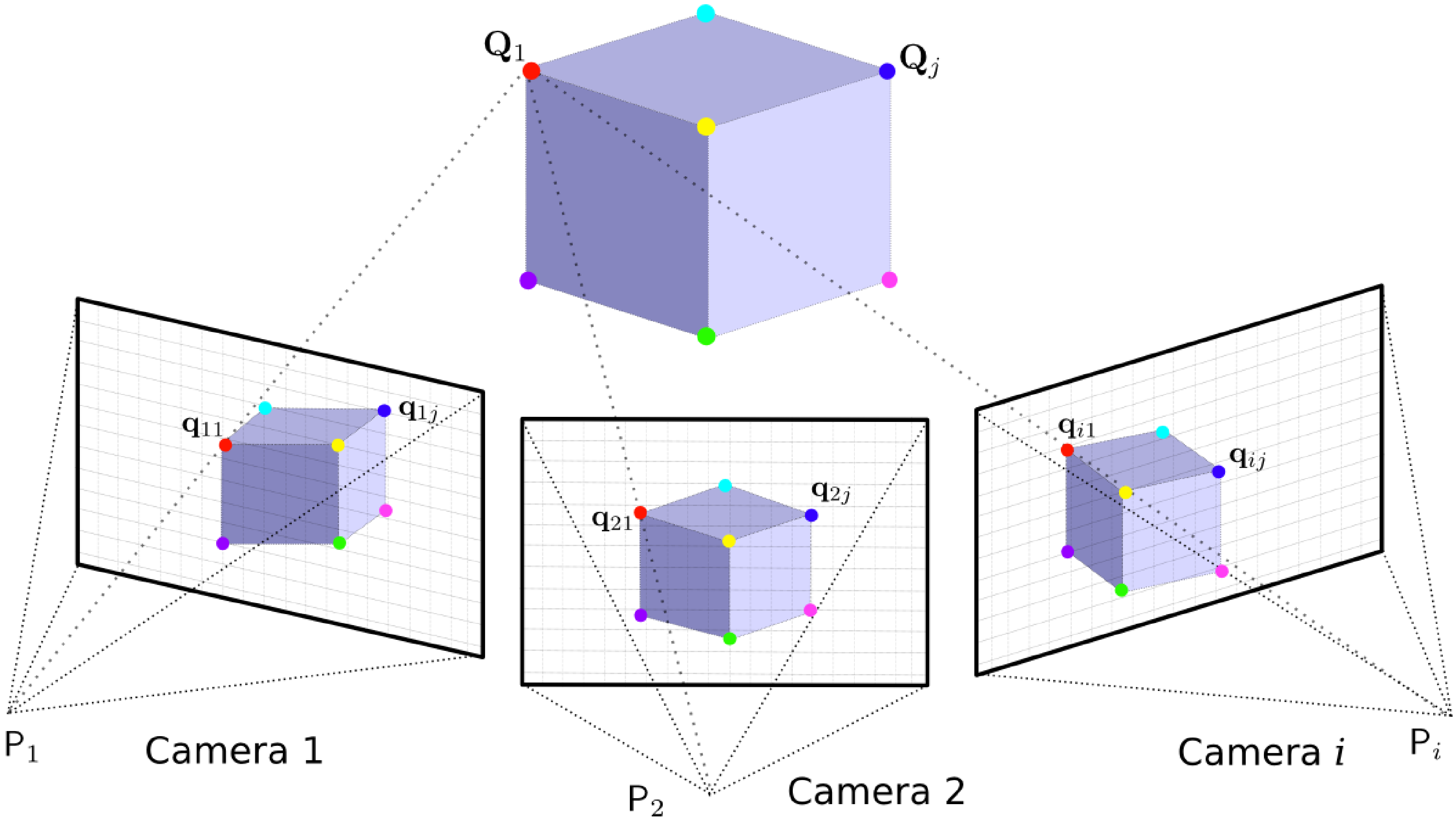


photo pop-up

http://dhoiem.cs.illinois.edu/projects/popup/

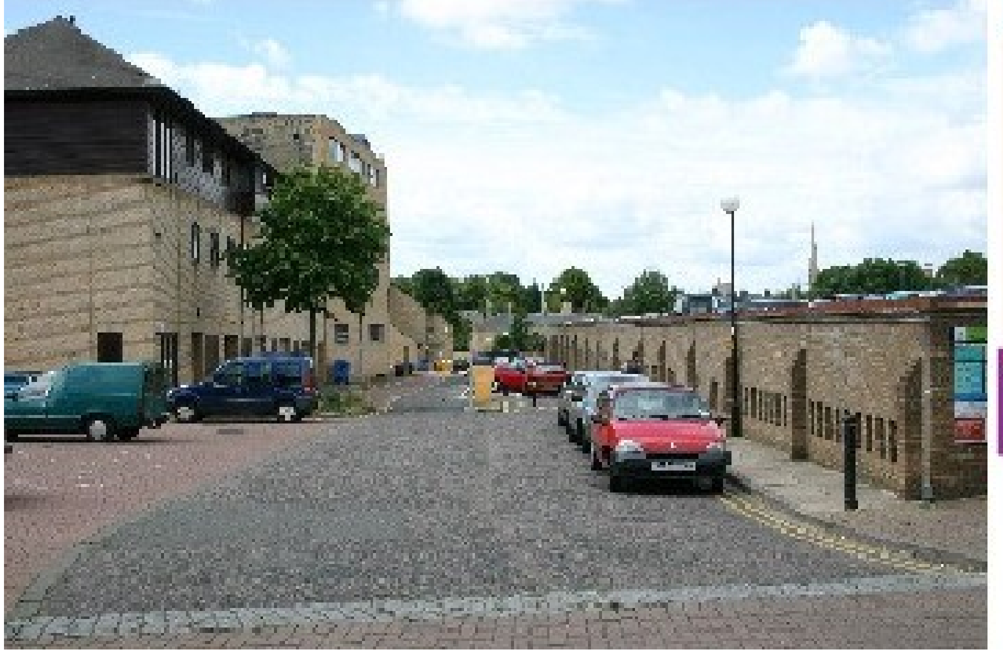
projective geometry

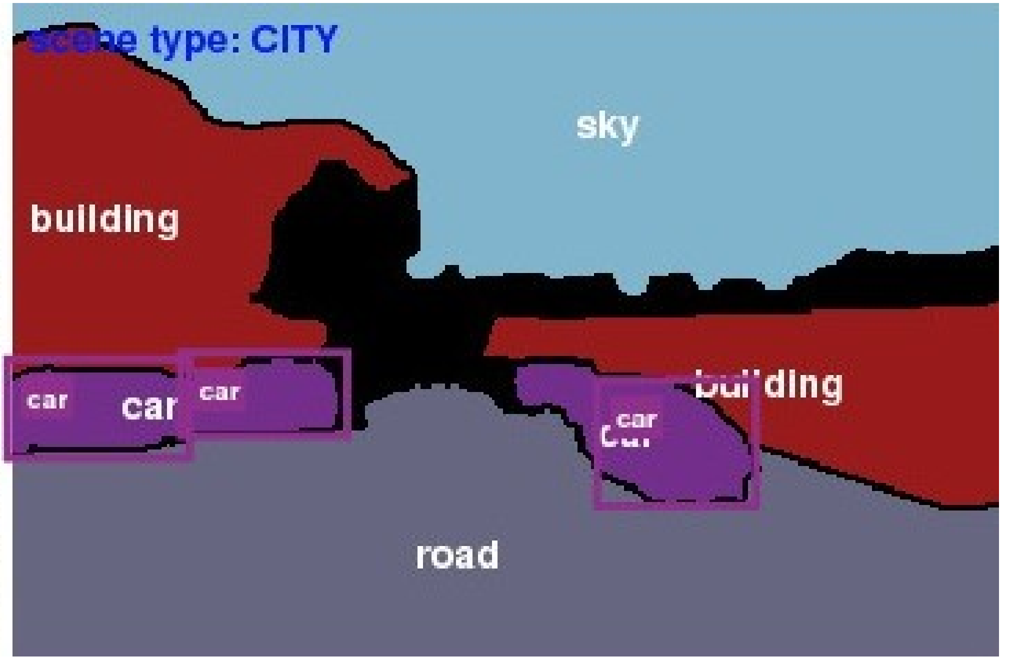


3D reconstruction

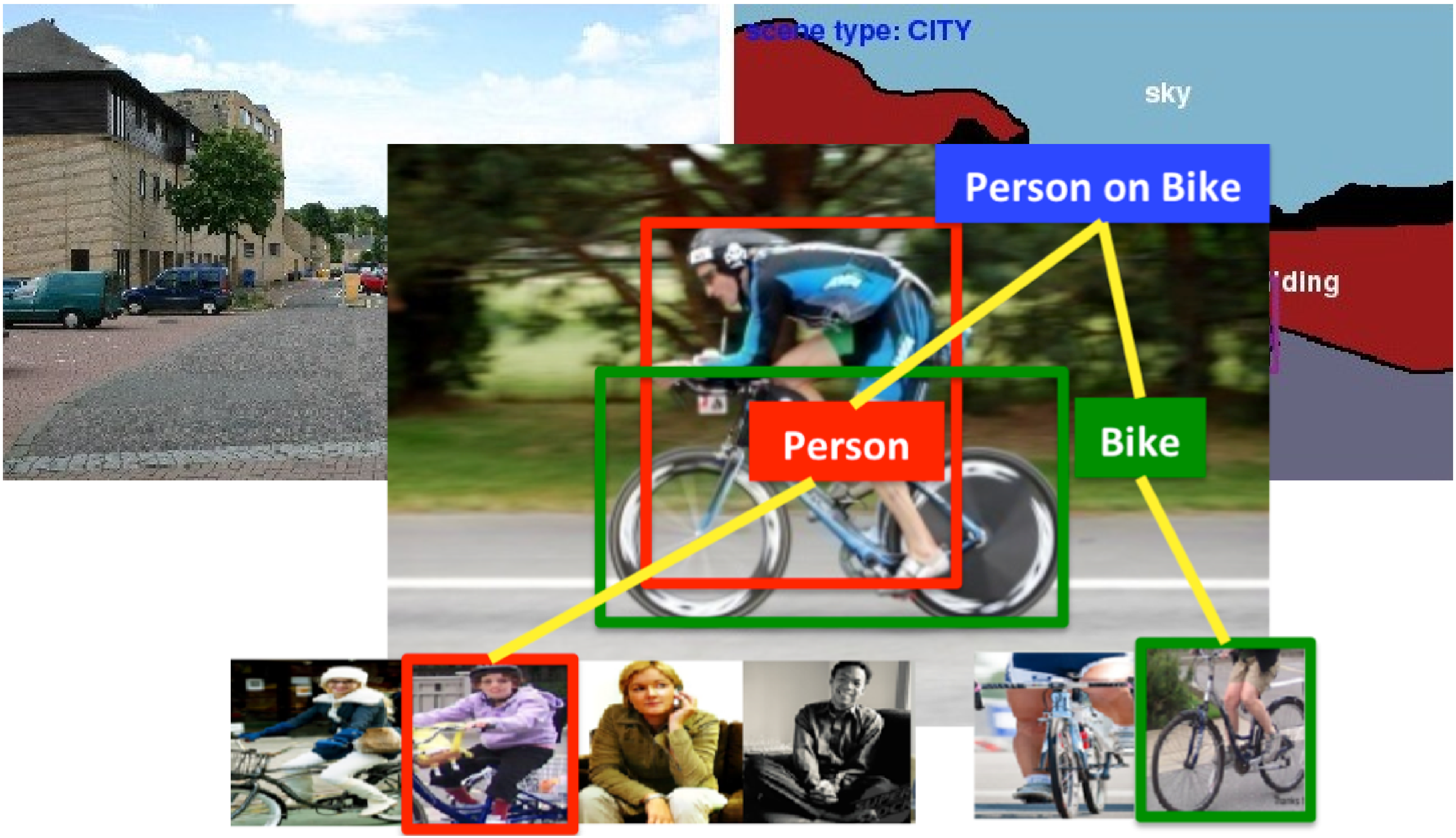
http://www.3dflow.net/

understanding





understanding



understanding

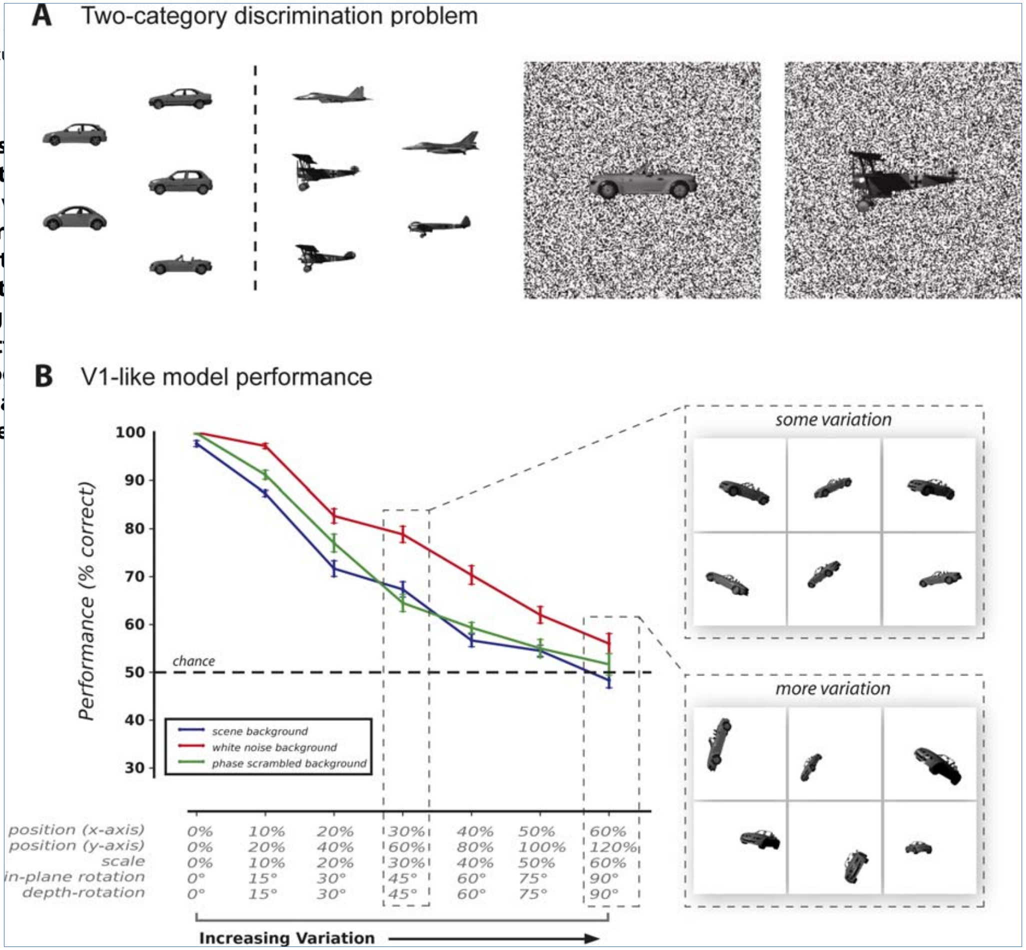
OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

Why is Real-World Visual Object Recognition Hard?

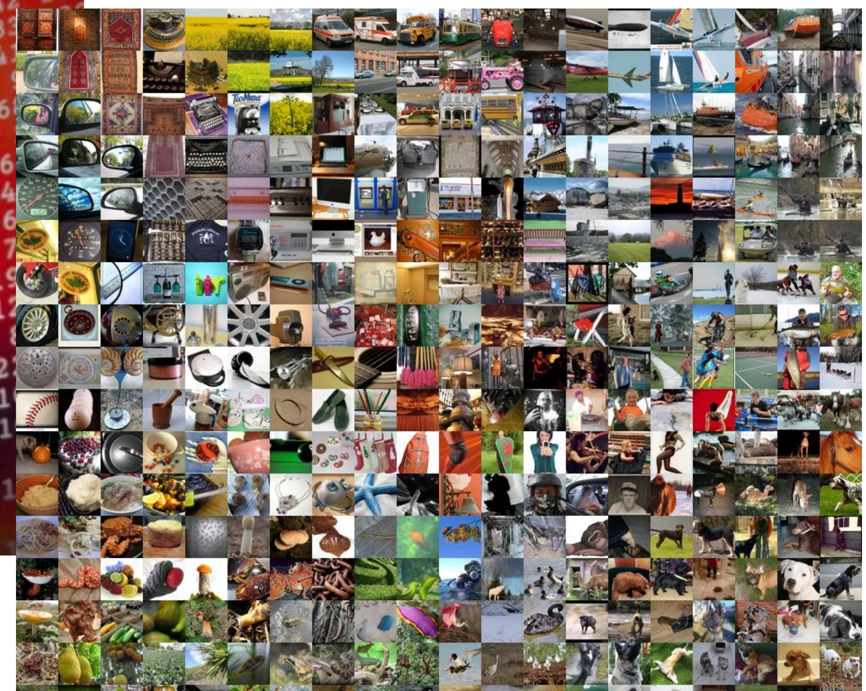
Nicolas Pinto[1,2], David D. Cox[1,2,3], James J. DiCarlo[1,2*]

1 McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States
Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 3 The Rowland Institu
States of America

**Progress in understanding the brain mechanisms underlying vision requires the cons
that not only emulate the brain's anatomy and physiology, but ultimately match it
recent years, "natural" images have become popular in the study of vision and ha
impressive progress in building such models. Here, we challenge the use of uncontr
that progress. In particular, we show that a simple V1-like model—a neuroscient
perform poorly at real-world visual object recognition tasks—outperforms state-of-t
(biologically inspired and otherwise) on a standard, ostensibly natural image recog
designed a "simpler" recognition test to better span the real-world variation in objec
show that this test correctly exposes the inadequacy of the V1-like model. Taken to
that tests based on uncontrolled natural images can be seriously misleading, potenti
direction. Instead, we reexamine what it means for images to be natural and argue
problem of object recognition—real-world image variation.**



Fei Fei Li



https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures

deep learning

ImageNet Classification with Deep Convolutional Neural Networks

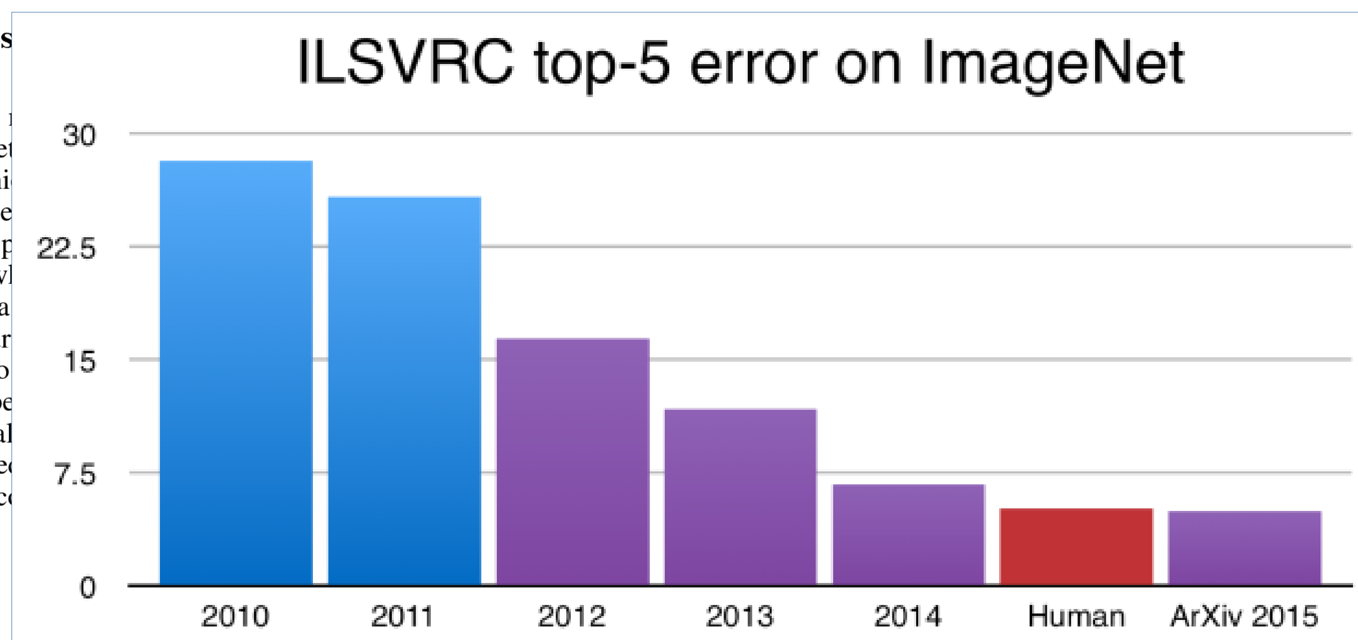
Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

Abs

We trained a large, deep convolutional
high-resolution images in the ImageNet
ferent classes. On the test data, we achi
and 17.0% which is considerably bette
neural network, which has 60 million p
of five convolutional layers, some of wh
and three fully-connected layers with a
ing faster, we used non-saturating neur
tation of the convolution operation. To
layers we employed a recently-develope
that proved to be very effective. We al
ILSVRC-2012 competition and achieve
compared to 26.2% achieved by the sec



deep learning

DenseCap: Fully Convolutional Localization Networks for Dense Captioning

Justin Johnson* Andrej Karpathy* Li Fei-Fei
Department of Computer Science, Stanford University
{jcjohns,karpathy,feifeili}@cs.stanford.edu

Abstract

We introduce the dense captioning task, which requires a computer vision system to both localize and describe salient regions in images in natural language. The dense captioning task generalizes object detection when the descriptions consist of a single word, and Image Captioning when one predicted region covers the full image. To address the localization and description task jointly we propose a Fully Convolutional Localization Network (FCLN) architecture that processes an image with a single, efficient forward pass, requires no external regions proposals, and can be trained end-to-end with a single round of optimization. The architecture is composed of a Convolutional Network, a novel

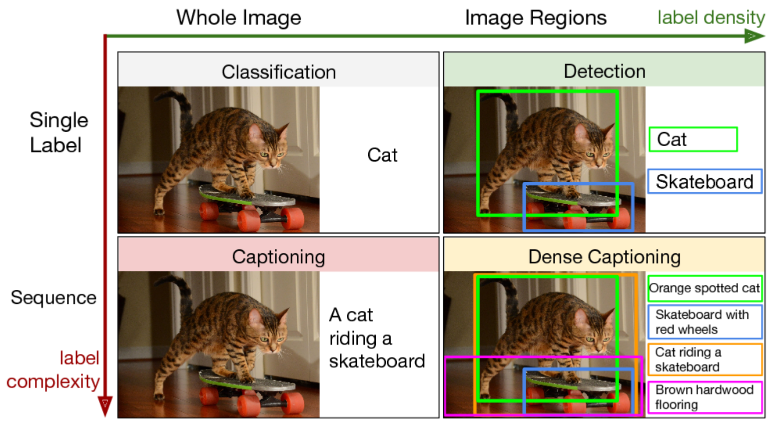


Figure 1. We address the Dense Captioning task (bottom right) with a model that jointly generates both dense and rich annotations in a single forward pass.

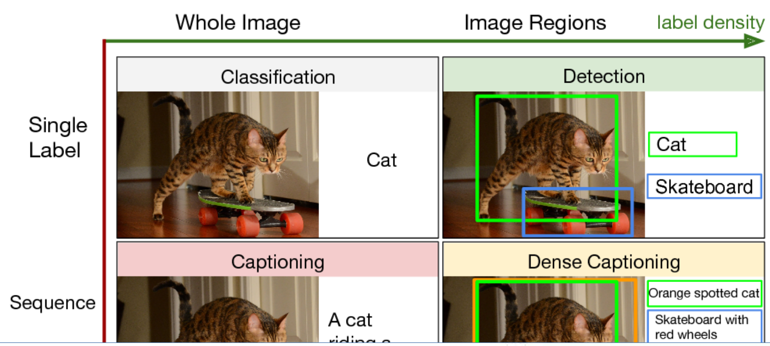
deep learning

DenseCap: Fully Convolutional Localization Networks for Dense Captioning

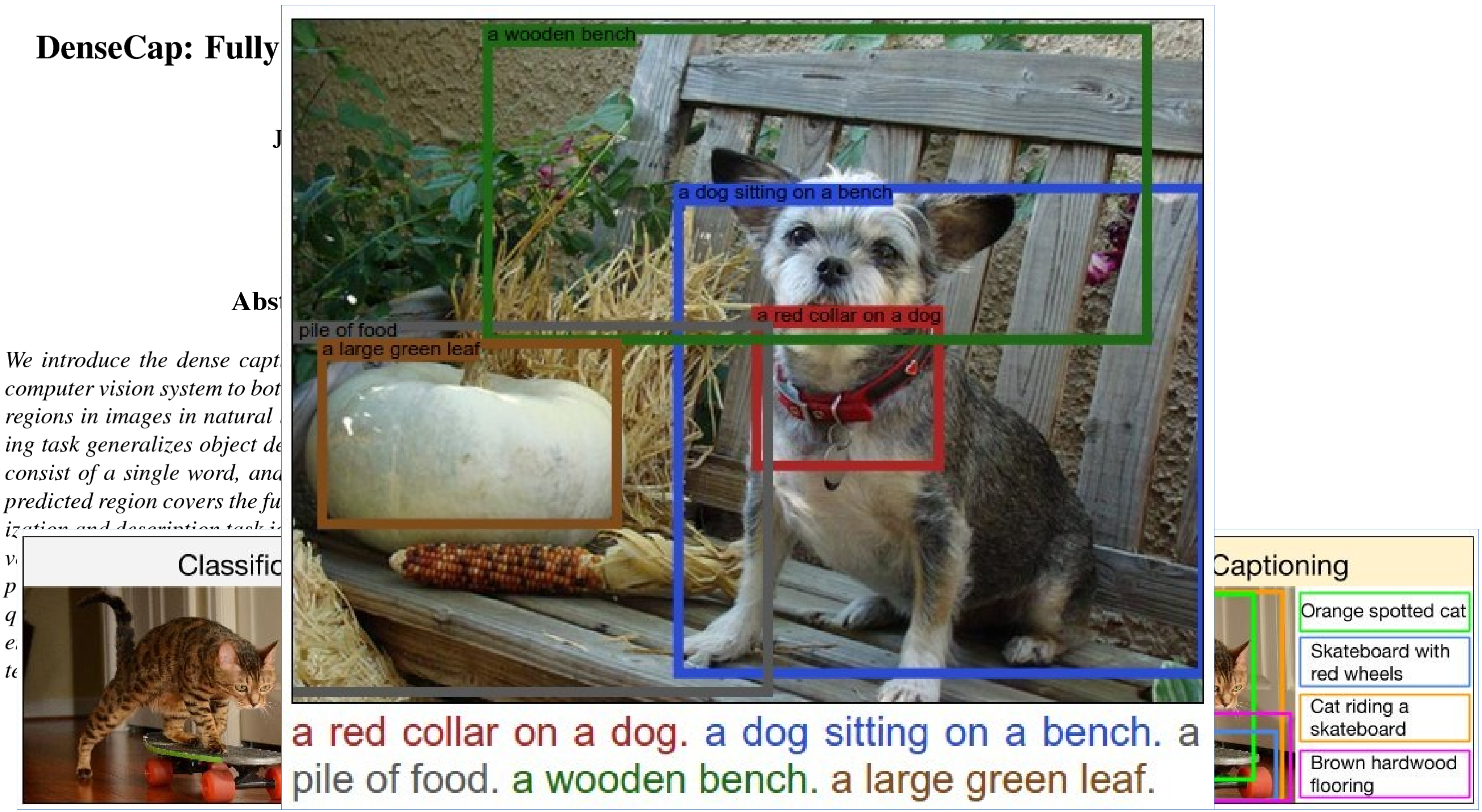
Justin Johnson* Andrej Karpathy* Li Fei-Fei
Department of Computer Science, Stanford University
{jcjohns,karpathy,feifeili}@cs.stanford.edu

Abstract

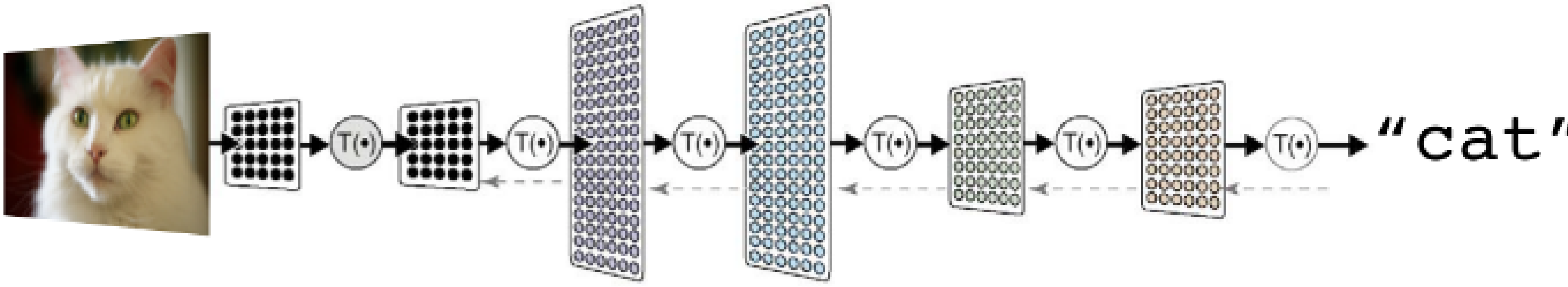
We introduce the dense captioning task, which requires a computer vision system to both localize and describe salient regions in images in natural language. The dense captioning task generalizes object detection when the descriptions consist of a single word, and Image Captioning when one predicted region covers the full image. To address the local-



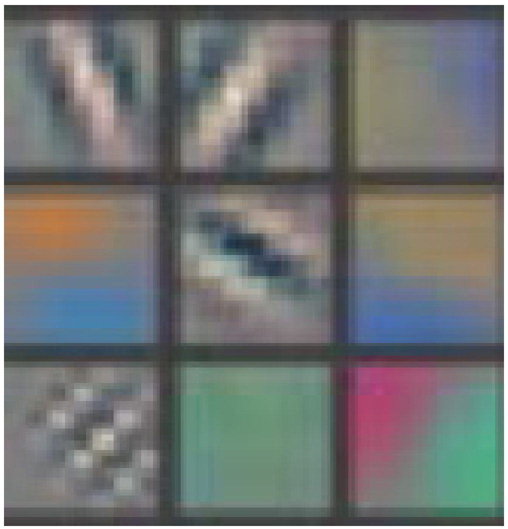
deep learning

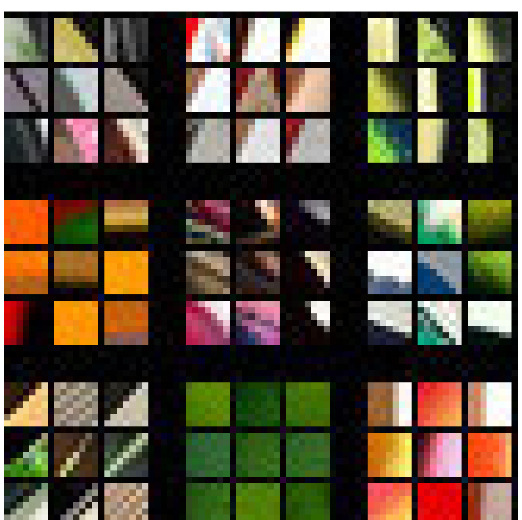


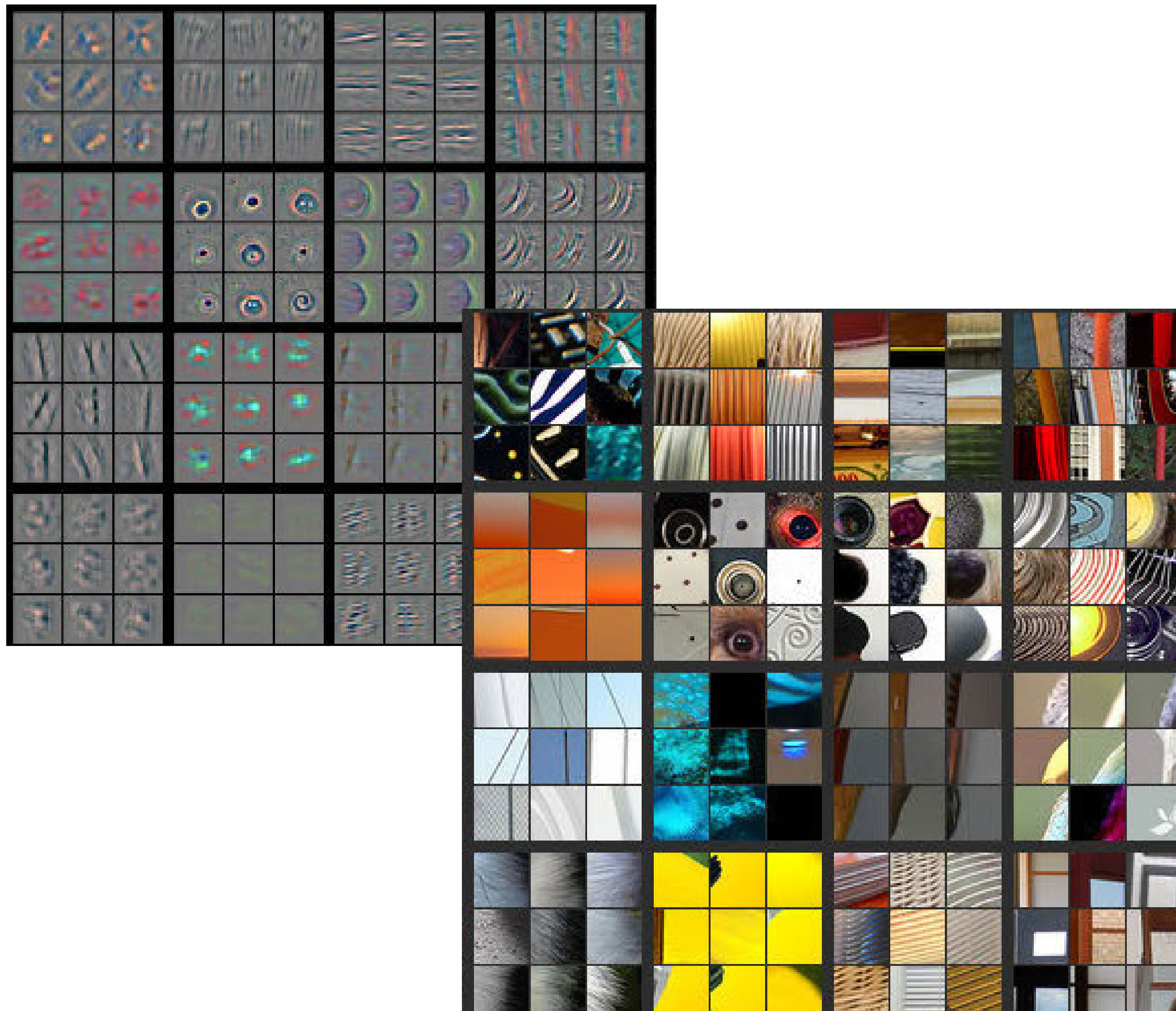
deep learning



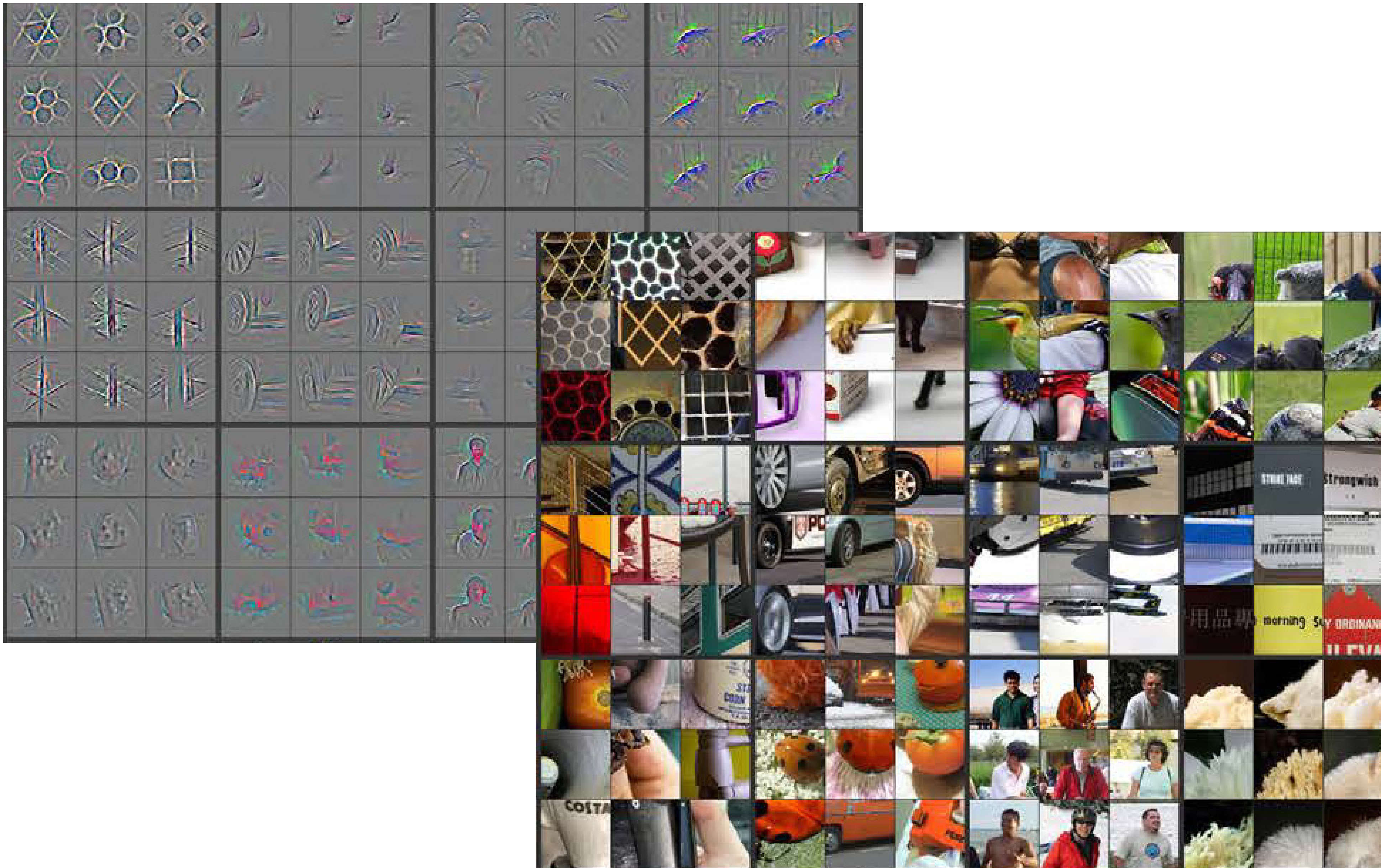
deepvis



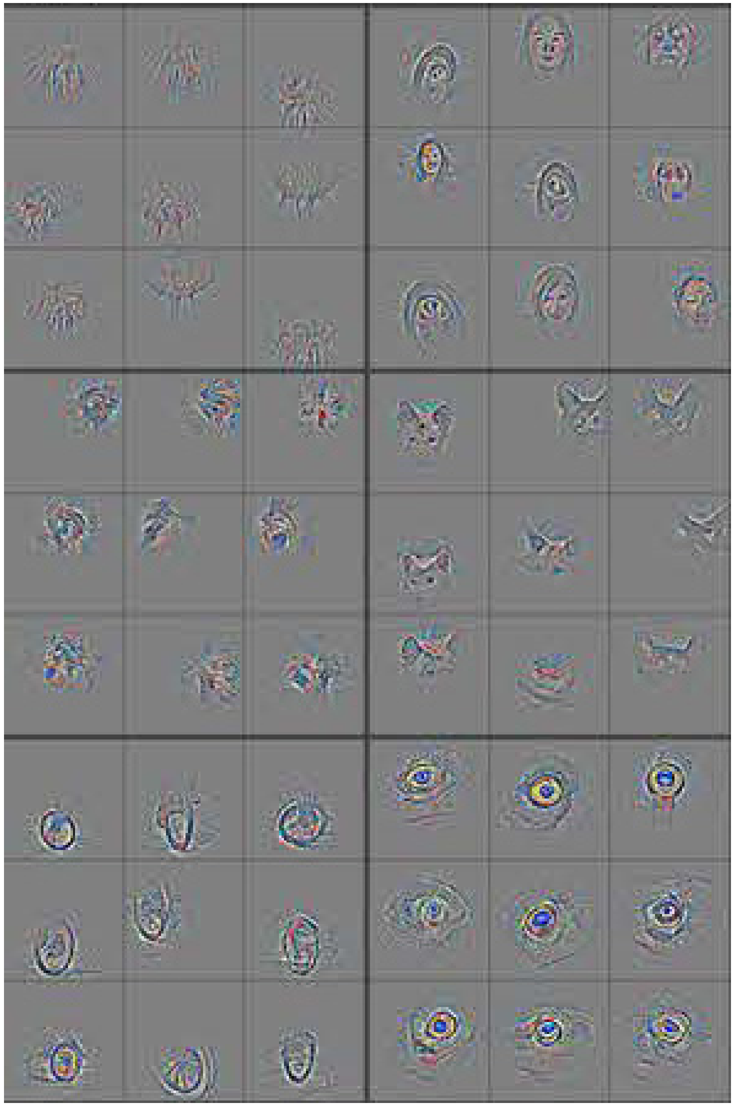


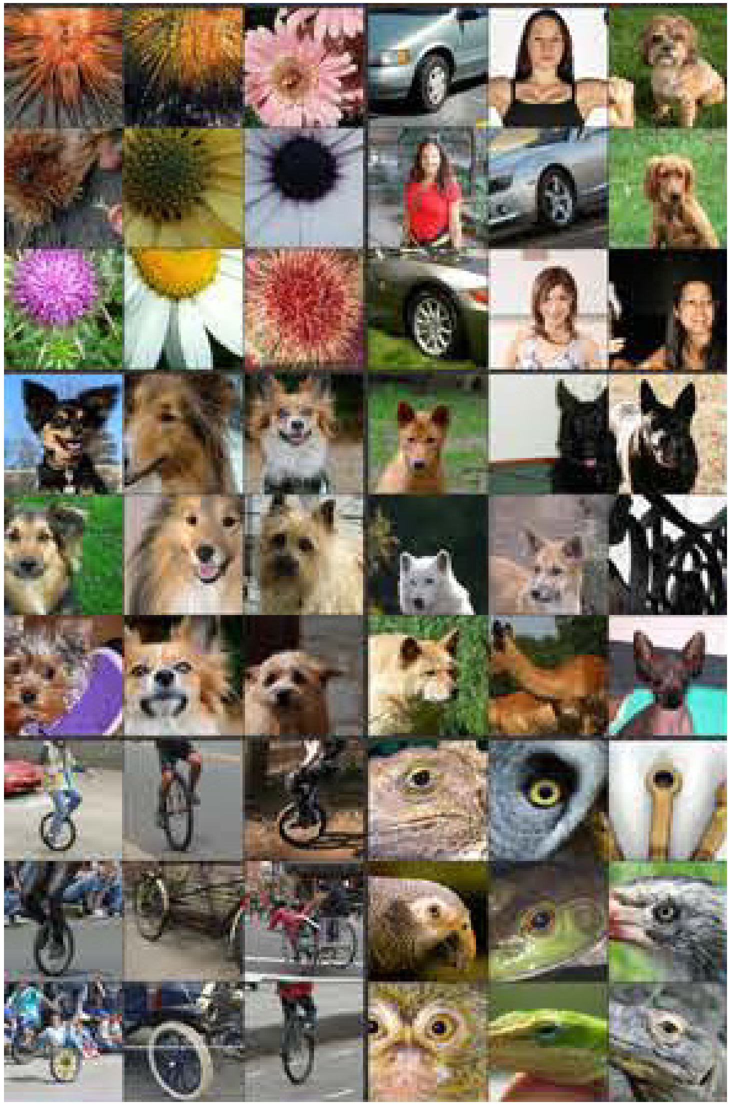


deepvis

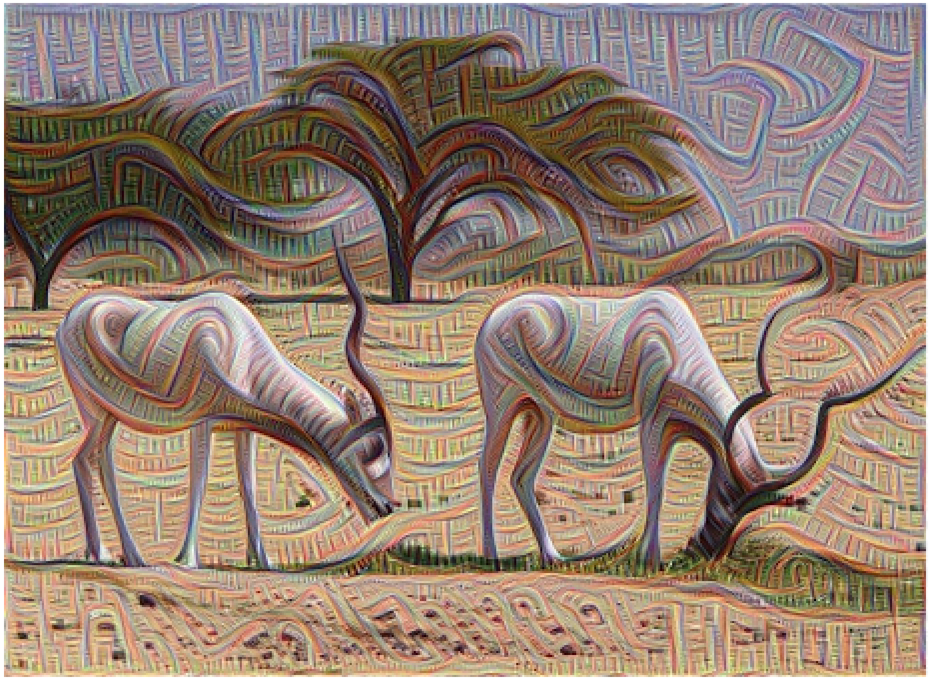


deepvis

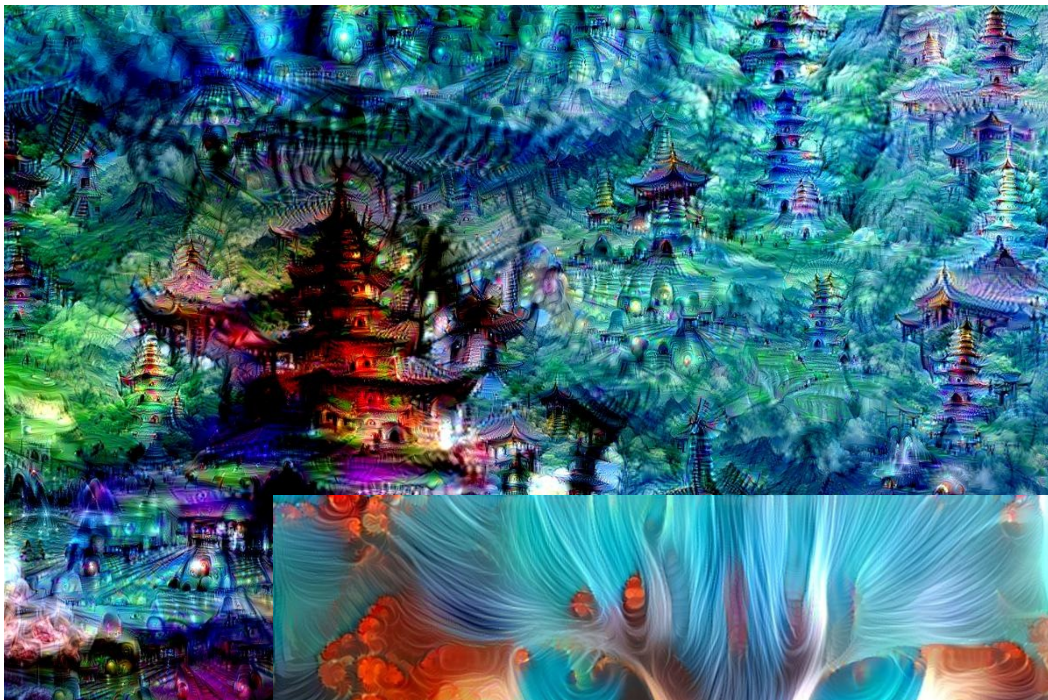


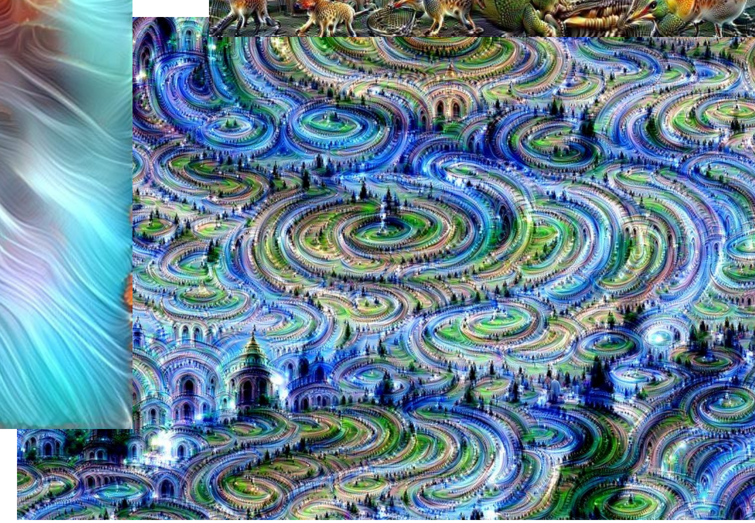


deep dreams

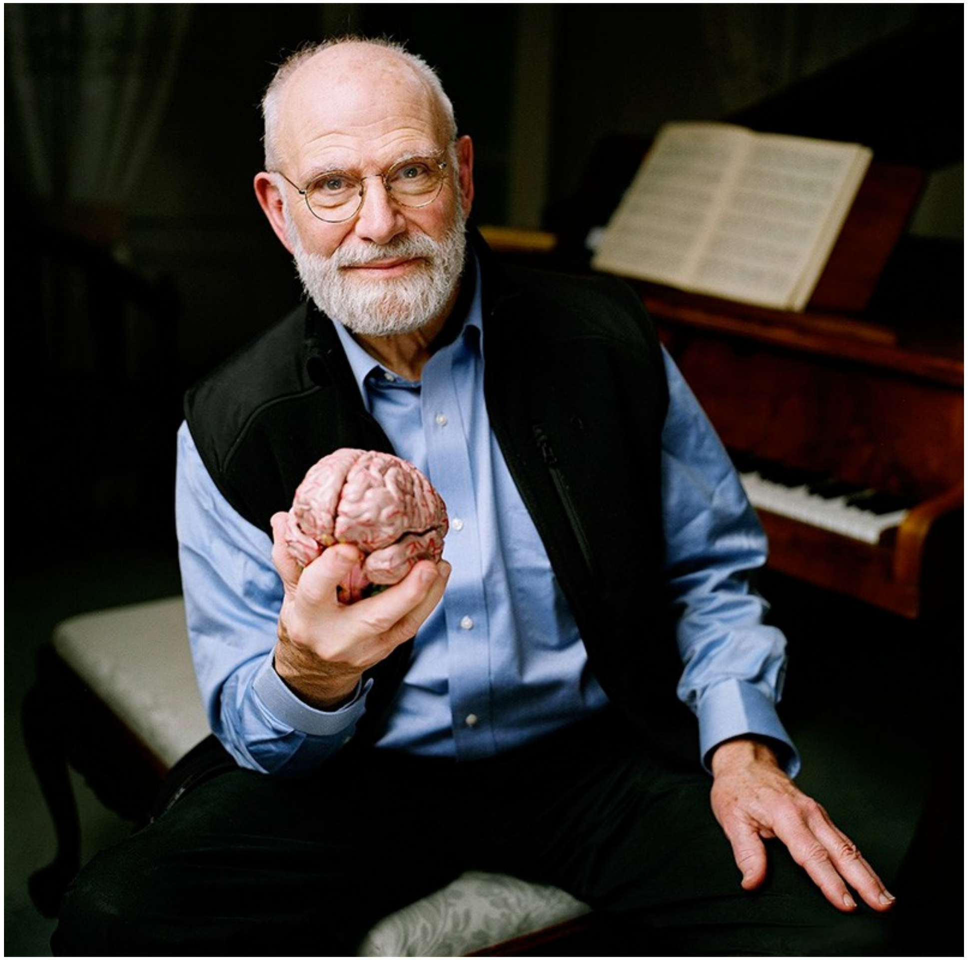


deep dreams



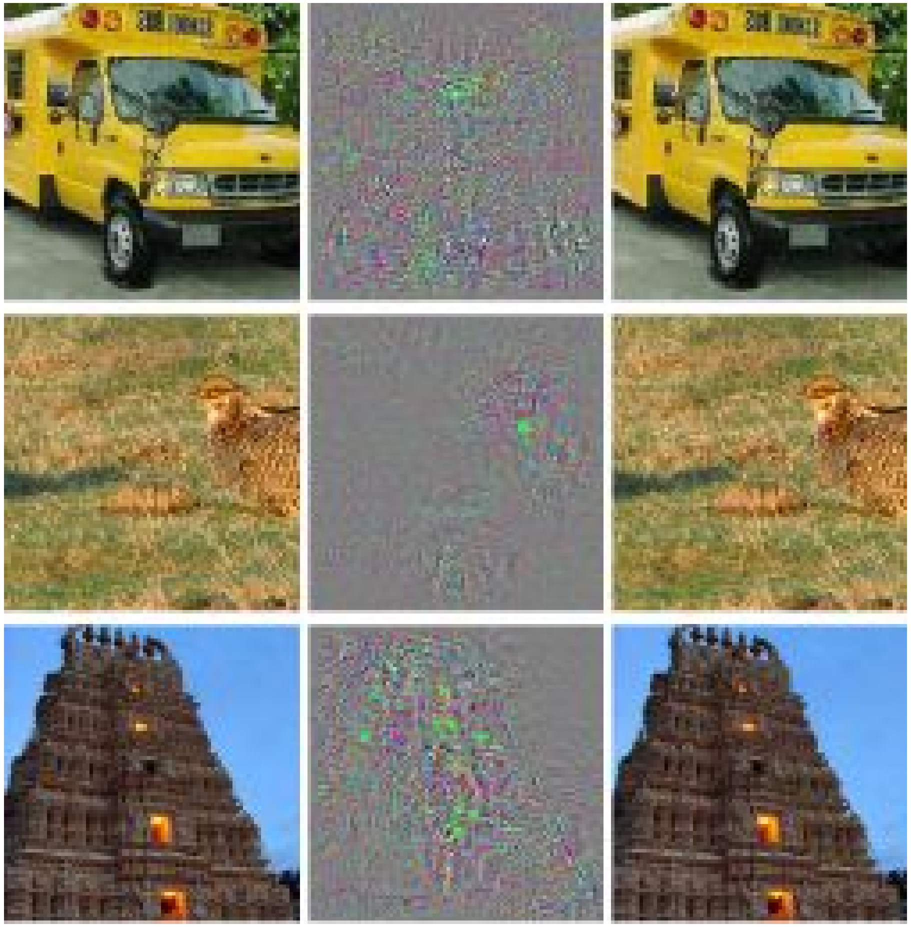


real deep dreams?

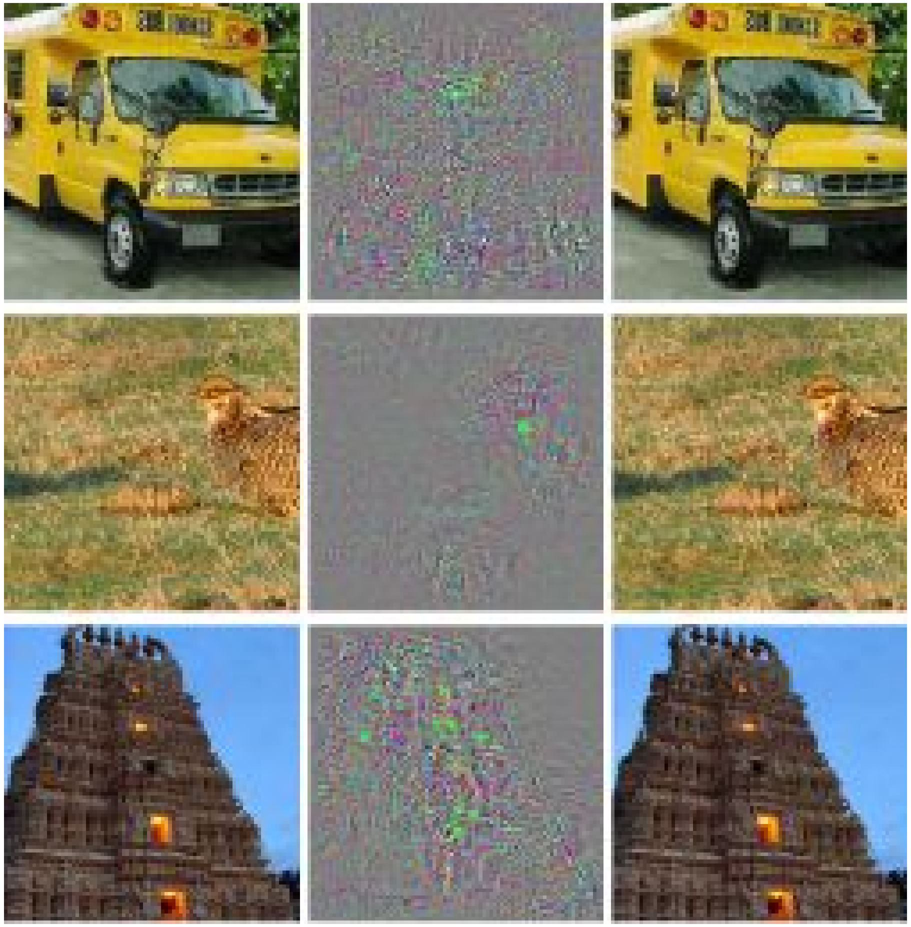


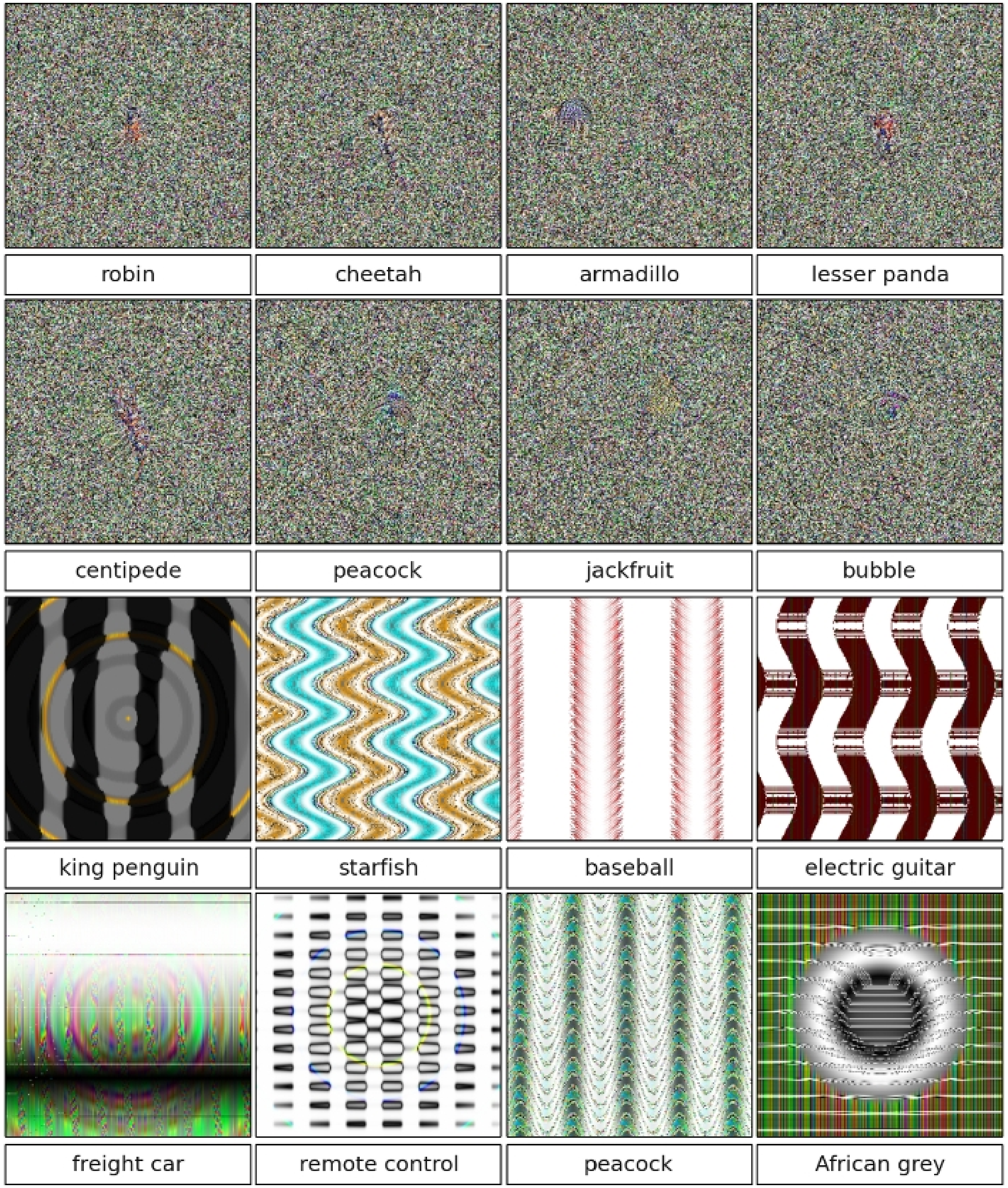
https://www.ted.com/talks/oliver_sacks_what_hallucination_reveals_about_our_minds

human vs machine

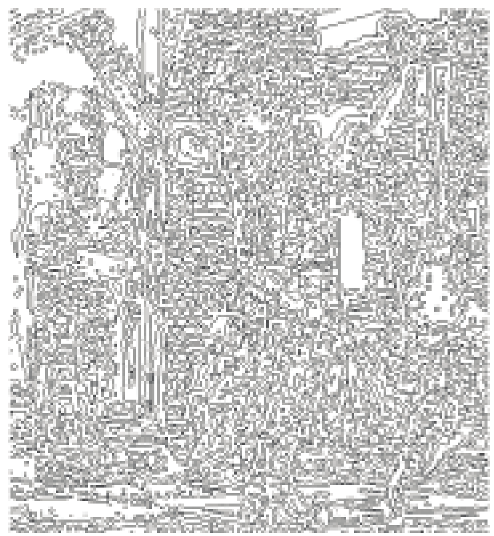


human vs machine



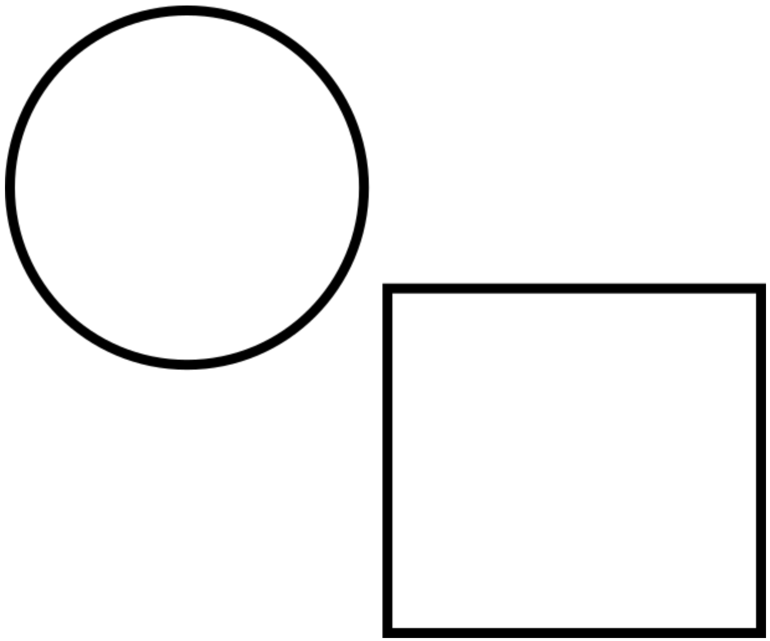


learning to see



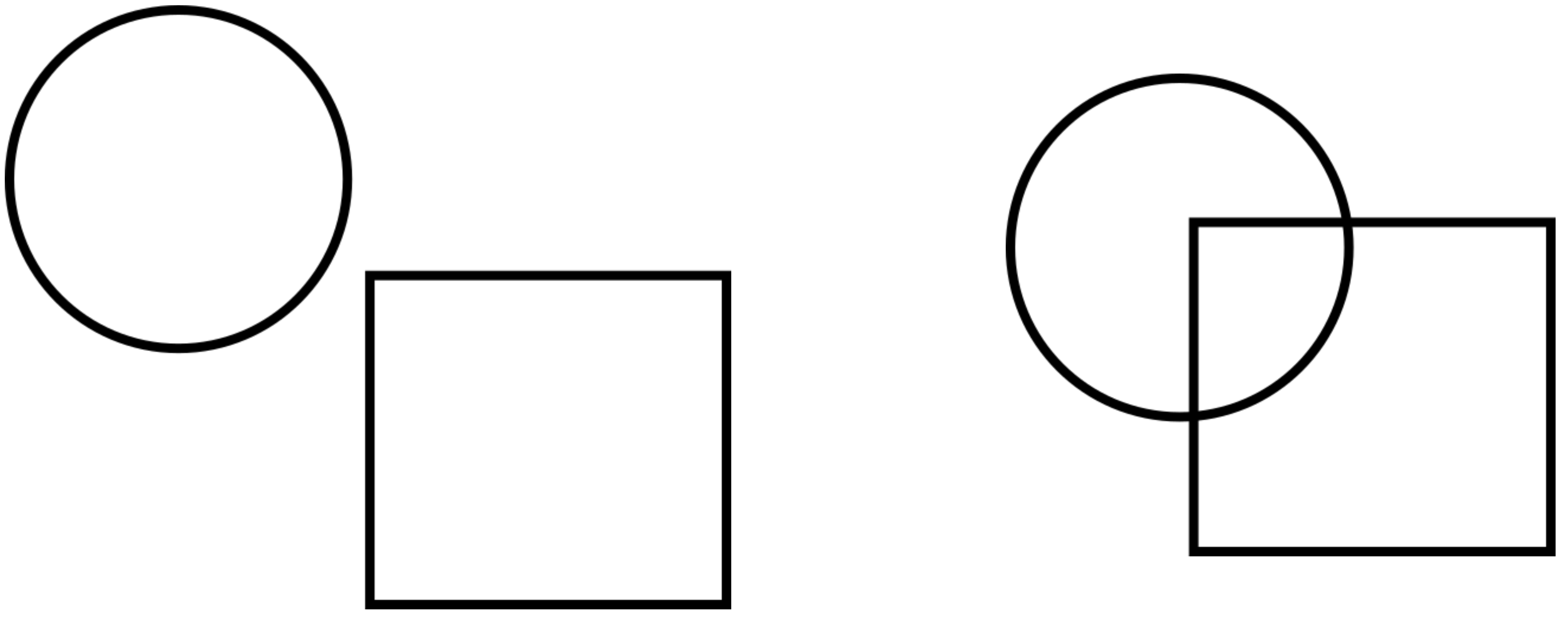
http://web.mit.edu/sinhalab/

learning to see



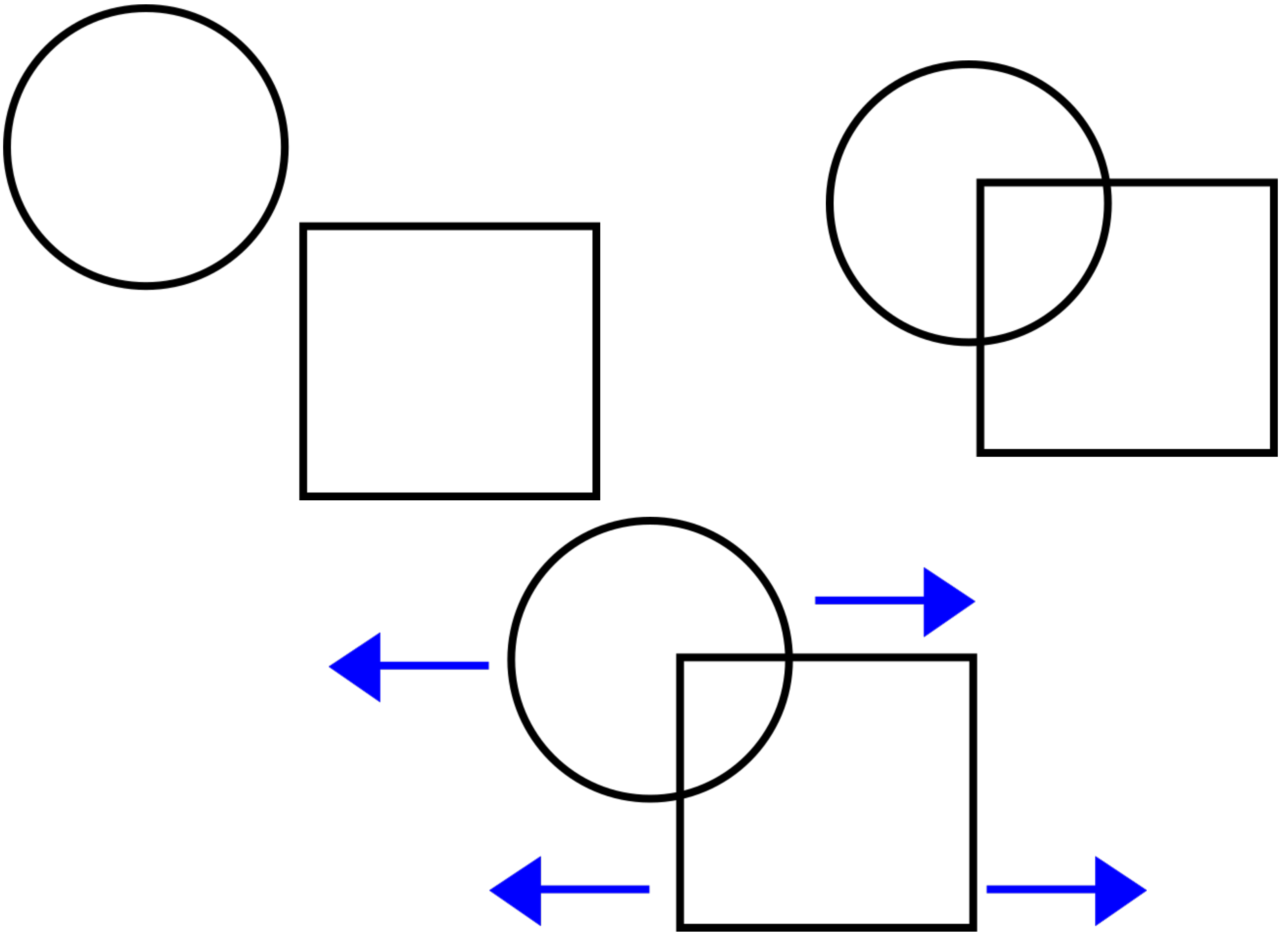
https://www.ted.com/talks/pawan_sinha_on_how_brains_learn_to_see

learning to see



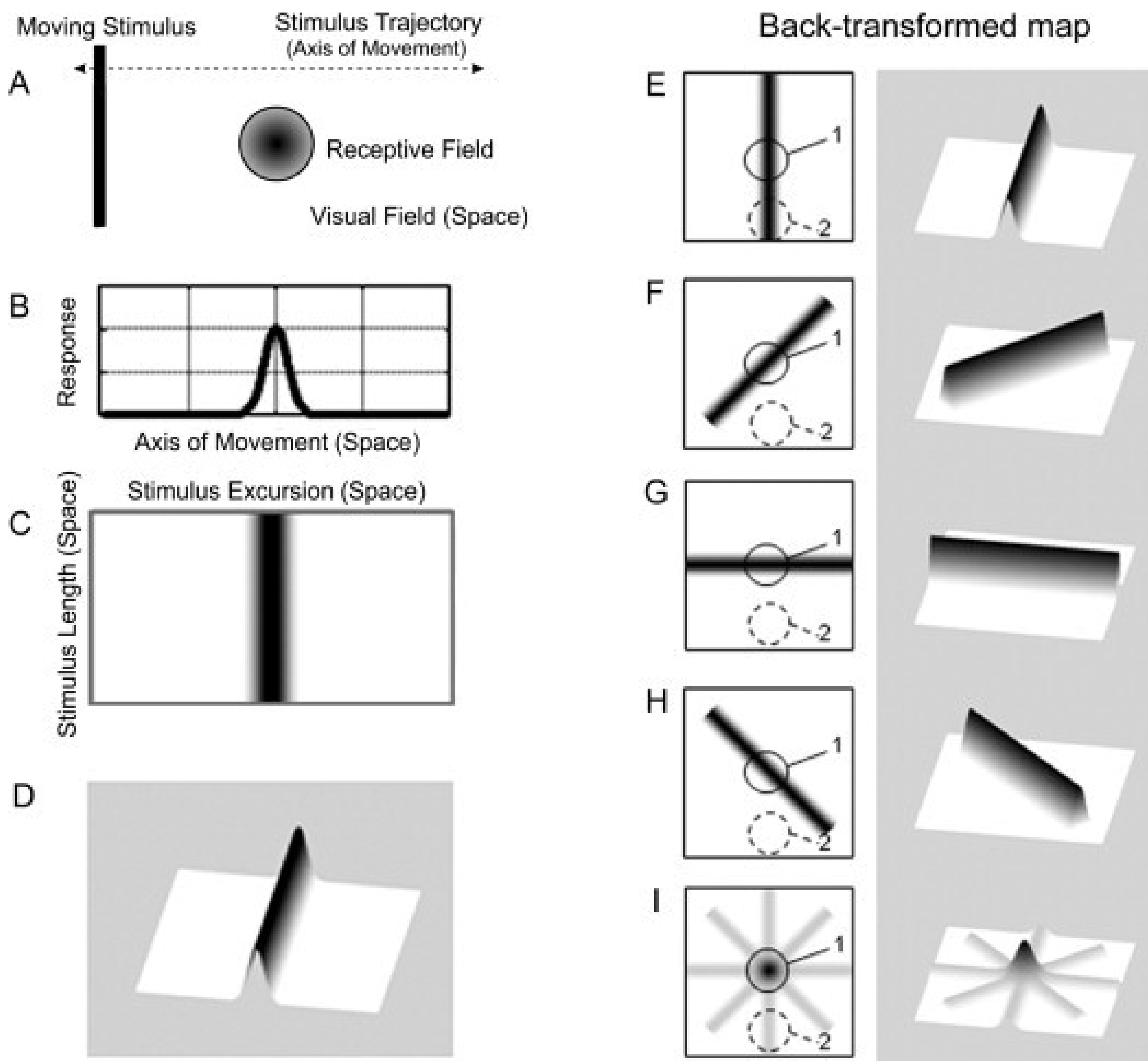
https://www.ted.com/talks/pawan_sinha_on_how_brains_learn_to_see

learning to see

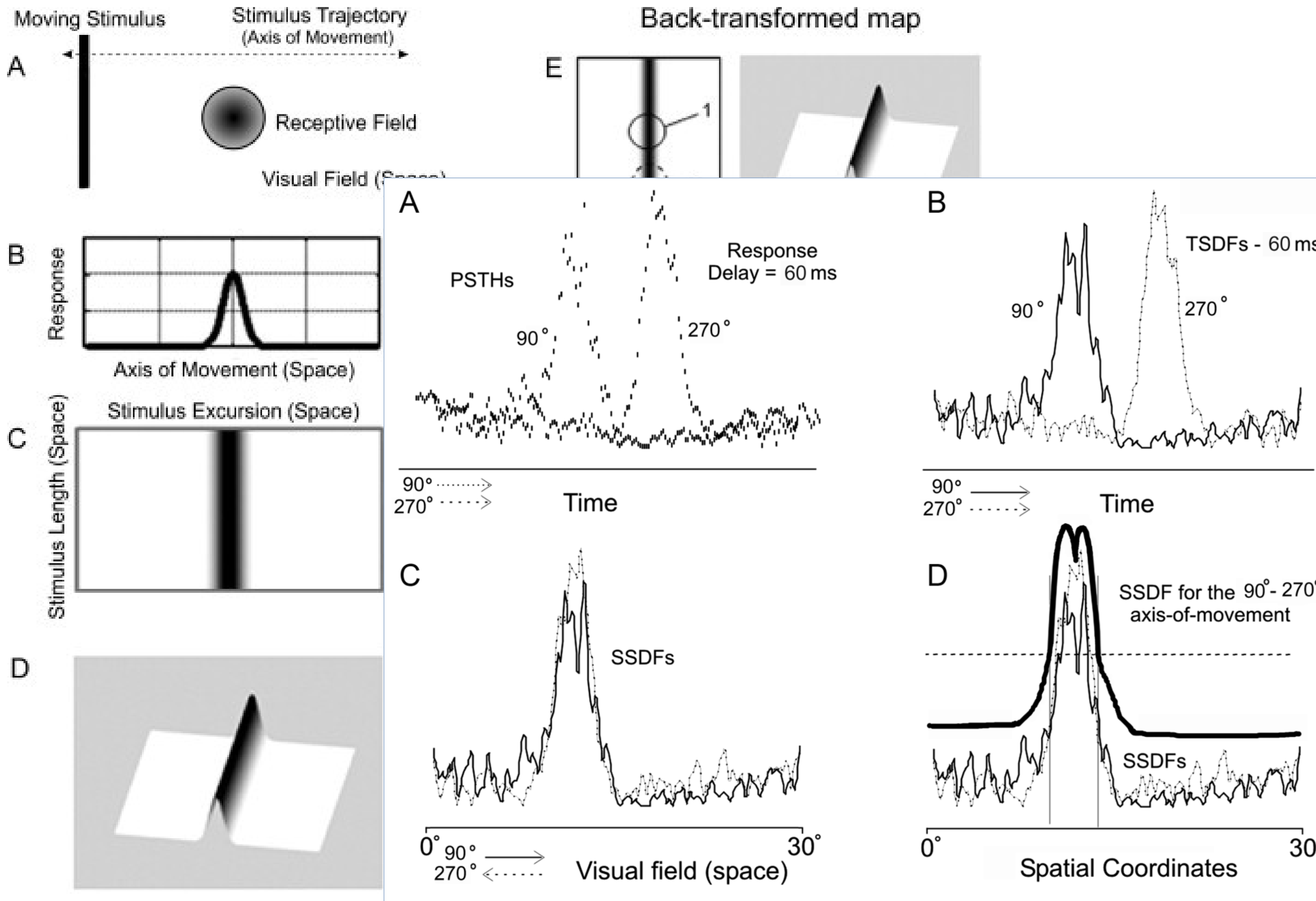


https://www.ted.com/talks/pawan_sinha_on_how_brains_learn_to_see

brainvis

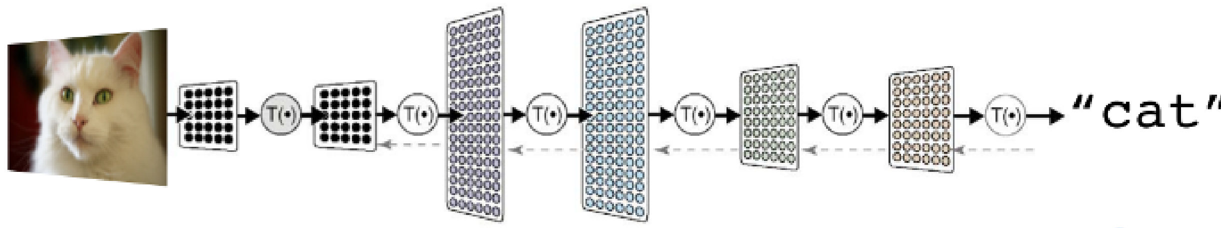

Moving Stimulus
Stimulus Trajectory
(Axis of Movement)
A
Receptive Field
Visual Field (Space)
B
Response
Axis of Movement (Space)
C
Stimulus Excursion (Space)
Stimulus Length (Space)
D
Back-transformed map
E
F
G
H
I
1
2

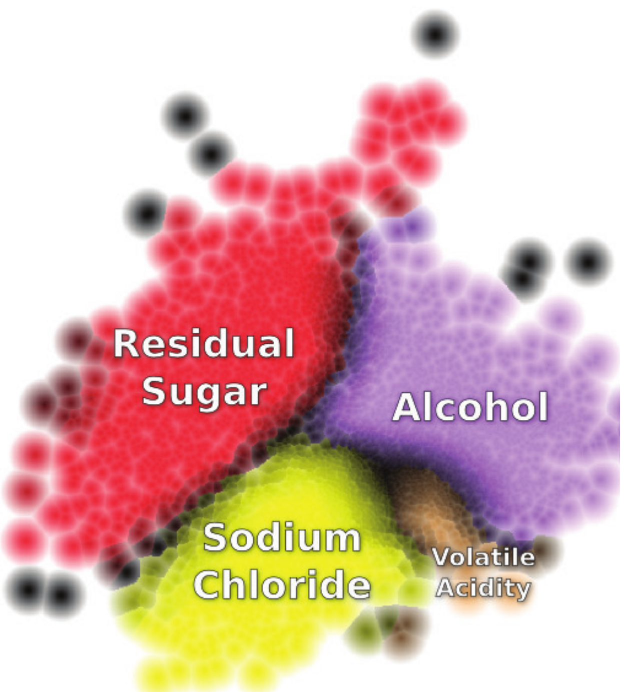
brainvis

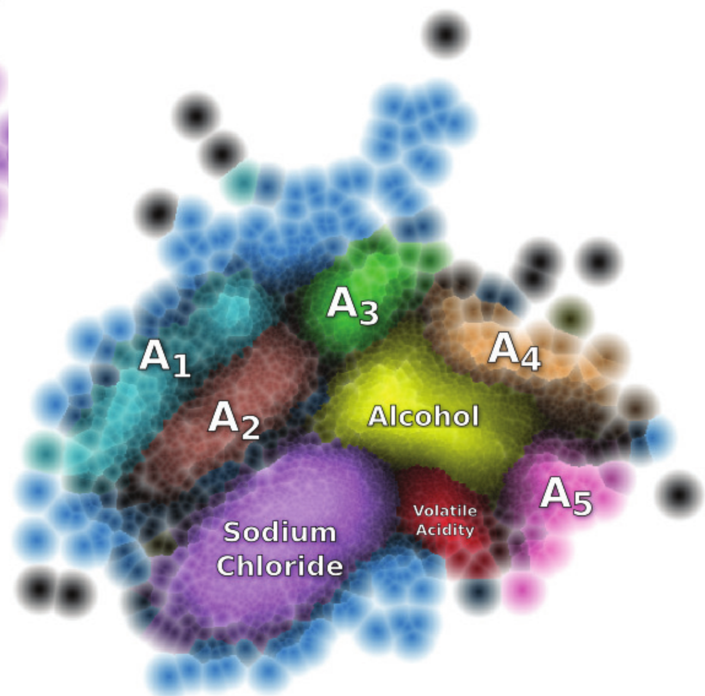


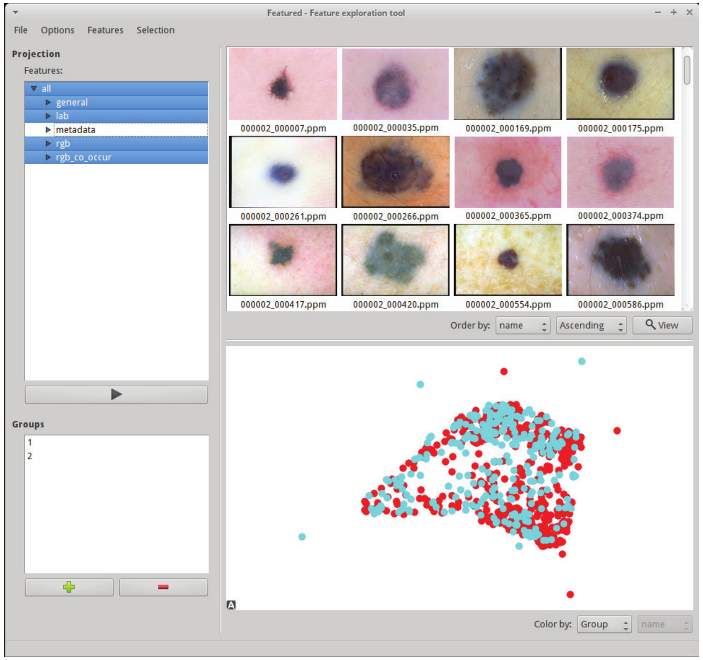
Moving Stimulus
Stimulus Trajectory
(Axis of Movement)
A
Receptive Field
Visual Field
B
Response
Axis of Movement (Space)
Stimulus Excursion (Space)
C
Stimulus Length (Space)
D
Back-transformed map
E
1
A
PSTHs
Response
Delay = 60 ms
90°
270°
Time
B
TSDFs - 60 ms
90°
270°
Time
C
SSDFs
0°
30°
90°
270°
Visual field (space)
D
SSDF for the 90°- 270°
axis-of-movement
SSDFs
0°
30°
Spatial Coordinates

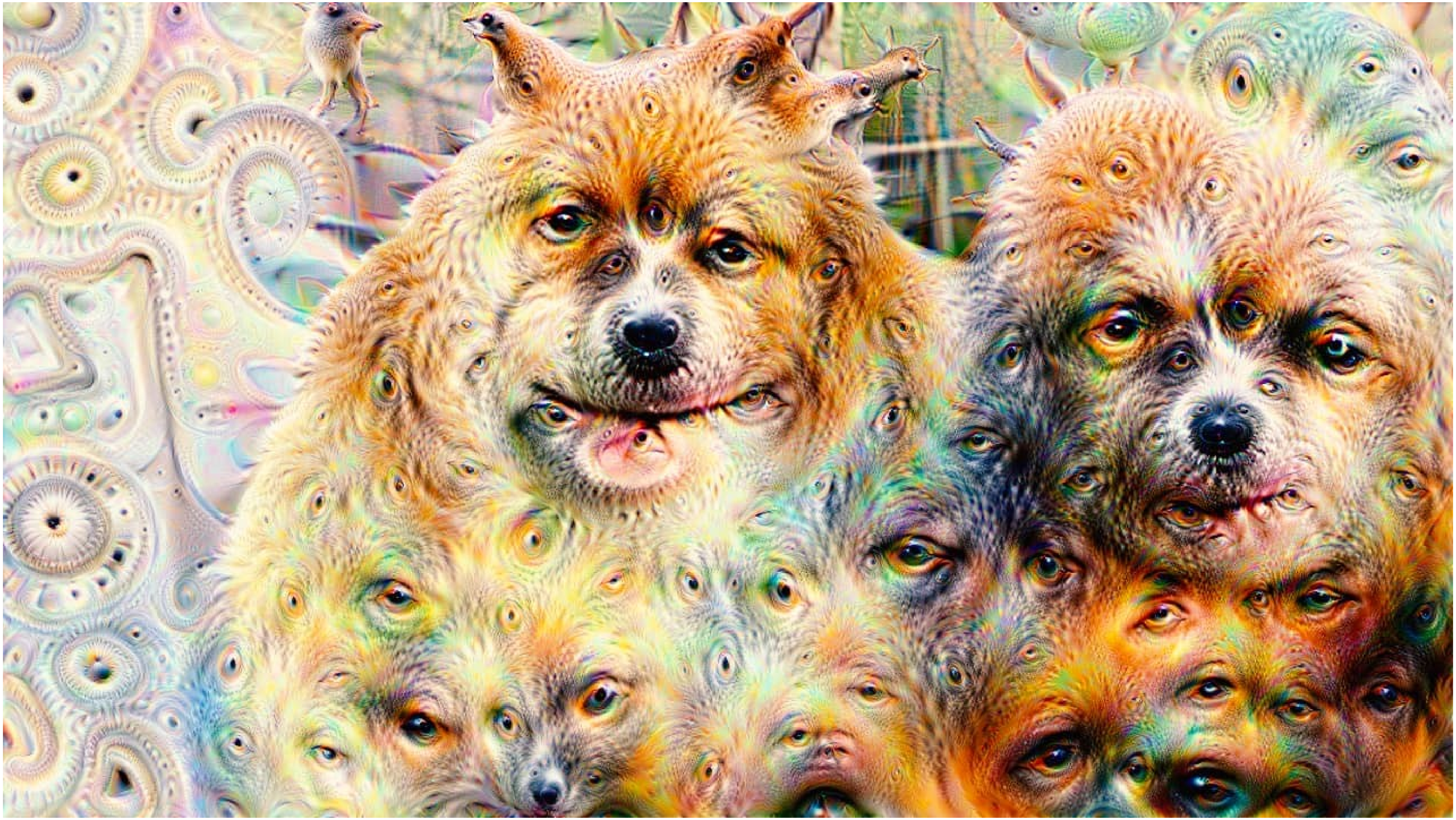
brainvis?


“cat”


Residual Sugar
Alcohol
Sodium Chloride
Volatile Acidity


A1
A2
A3
A4
A5
Alcohol
Sodium Chloride
Volatile Acidity


Featured - Feature exploration tool
File Options Features Selection
Projection
Features:
all
general
lab
metadata
rgb
rgb_co_occur
000002_000007.ppm
000002_000035.ppm
000002_000169.ppm
000002_000175.ppm
000002_000261.ppm
000002_000266.ppm
000002_000365.ppm
000002_000374.ppm
000002_000417.ppm
000002_000420.ppm
000002_000554.ppm
000002_000586.ppm
Order by: name Ascending View
Groups
1
2
Color by: Group name



https://vimeo.com/132700334

See eye to eye!

Ricardo Marroquim

www.lcg.ufrj.br/~marroquim

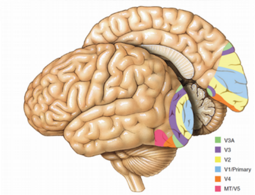
Laboratório de Computação Gráfica

PESC Programa de Engenharia de Sistemas e Computação

SEMANA PESC – 21-25 de novembro 2016

image references 1/

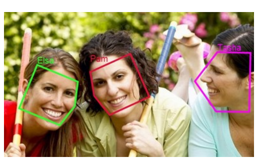
http://www.orosend.com/keep-your-eyes-healthy-nmw15/



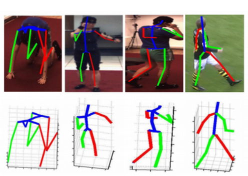
https://www.studyblue.com/notes/note/n/the-visual-system/deck/6955406

http://www.kdnuggets.com/2016/08/seven-steps-understanding-computer-vision.html

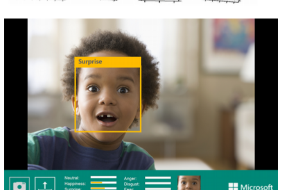
http://semiengineering.com/seeing-the-future-of-vision/



http://www.rcrwireless.com/20110723/wireless/google-buys-facial-recognition-firm-despite-privacy-concerns#prettyPhoto



http://cvlab.epfl.ch/research/surv/human-pose-estimation



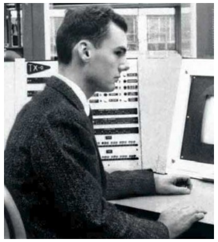
http://venturebeat.com/2015/11/11/microsoft-launches-project-oxford-apis-for-face-tracking-emotion-speaker-recognition-spell-checking/

image references 2/

http://www.2001italia.it/2014/04/a-full-cast-list-for-2001-part-4.html

http://www.masswerk.at/minskytron/

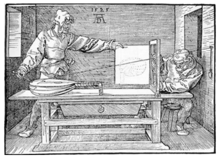
http://xkcd.com/1425/



http://webdesignpi.tripod.com/roberts.htm

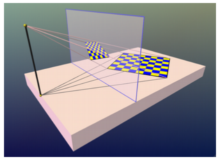
https://blogs.royalsociety.org/publishing/350-anniversary-issue-author-q-a-richard-morris/

image references 3/

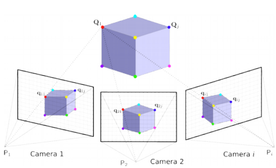


http://xahlee.info/3d/tech_drawing.html

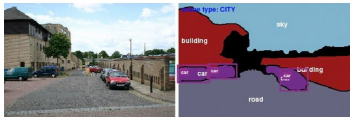
http://b3ck.blogspot.com.br/



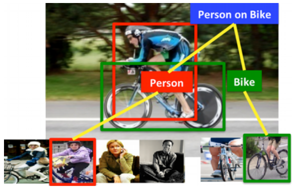
http://cgunn3.blogspot.com.br/



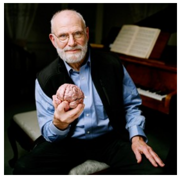
http://www.cs.cornell.edu/courses/cs4670/2013fa/lectures/lectures.html



http://ttic.uchicago.edu/~yaojian/HolisticSceneUnderstanding.html

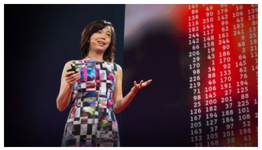


http://cs.stanford.edu/~taranlan/



http://www.oliversacks.com/about-oliver-sacks/

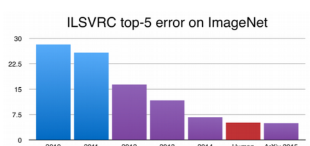
image references 4/



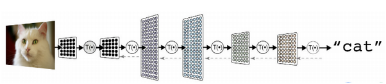
https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures

https://www.wired.com/2015/01/karpathy/

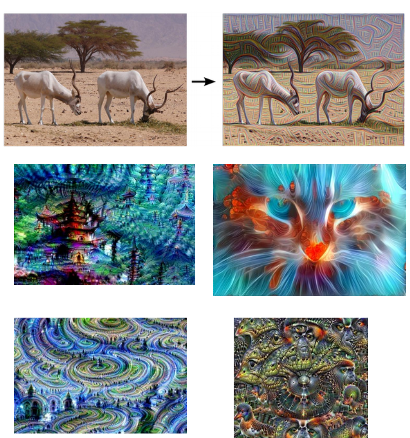
http://cs.stanford.edu/people/karpathy/



https://devblogs.nvidia.com/parallelforall/mocha-jl-deep-learning-julia/



http://redcatlabs.com/2014-12-18_DeepLearning.js/img/img-to-cat_700x131.png



https://photos.google.com/share/AF1QipPX0SCl7OzWilt9LnuQliattX4OUCj_8EP65_cTVnBmS1jnYgsGQAieQUc1VQWdgQ?key=aVBxWjhwSzg2RjJWLWRuVFBBZEN1d205bUdEMnhB

bibliography

Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. Proceedings of the National Academy of Sciences of the United States of America, 86(5), 1698–1702. http://doi.org/10.1073/pnas.86.5.1698

Rauber, P. E., Fadel, S., Falcao, A., & Telea, A. (2016). Visualizing the Hidden Activity of Artificial Neural Networks. IEEE Transactions on Visualization and Computer Graphics, 1, 1–1. http://doi.org/10.1109/TVCG.2016.2598838

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, R. F. (2014). Intriguing properties of neural networks. In International Conference on Learning Representations. http://doi.org/10.1021/ct2009208

Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep Neural Networks are Easily Fooled. In Computer Vision and Pattern Recognition, 2015 IEEE Conference on (pp. 427–436). http://doi.org/10.1109/CVPR.2015.7298640

Fiorani, M., Azzi, J. C. B., Soares, J. G. M., & Gattass, R. (2014). Automatic mapping of visual cortex receptive fields: A fast and precise algorithm. Journal of Neuroscience Methods, 221, 112–126. http://doi.org/10.1016/j.jneumeth.2013.09.012

Rauber, P. E., Silva, R. R. O., Feringa, S., Celebi, M. E., Falcão, A. X., & Telea, A. C. (2015). Interactive Image Feature Selection Aided by Dimensionality Reduction. In EuroVis Workshop on Visual Analytics (pp. 2–6). http://doi.org/10.2312/eurova.20151098

Silva, R. R. O., Rauber, P. E., Martins, R. M., Minghim, R., & Telea, A. C. (2015). Attribute-based Visual Explanation of Multidimensional Projections. In EuroVis Workshop on Visual Analytics. http://doi.org/10.2312/eurova.20151100

Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U., & Sinha, P. (2009). Visual parsing after recovery from blindness. Psychological Science, 20(12), 1484–1491. http://doi.org/10.1111/j.1467-9280.2009.02471.x

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? PLoS Computational Biology, 4(1), 0151–0156. http://doi.org/10.1371/journal.pcbi.0040027

Cox, D. D., & Dean, T. (2014). Neural networks and neuroscience-inspired computer vision. Current Biology, 24(18), R921–R929. http://doi.org/10.1016/j.cub.2014.08.026

Roberts, L. Gi. (1965). Machine perception of three-dimensional solids. PhD Thesis, (November), 159–197. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0413529

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115(3), 211–252. http://doi.org/10.1007/s11263-015-0816-y

Johnson, J., Karpathy, A., & Fei-Fei, L. (2016). DenseCap: Fully Convolutional Localization Networks for Dense Captioning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Pattern Recognition. http://doi.org/10.1109/CVPR.2016.494

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances In Neural Information Processing Systems (pp. 1–9). http://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007