

AI for Social Good summer Lab

Negar Rostamzadeh

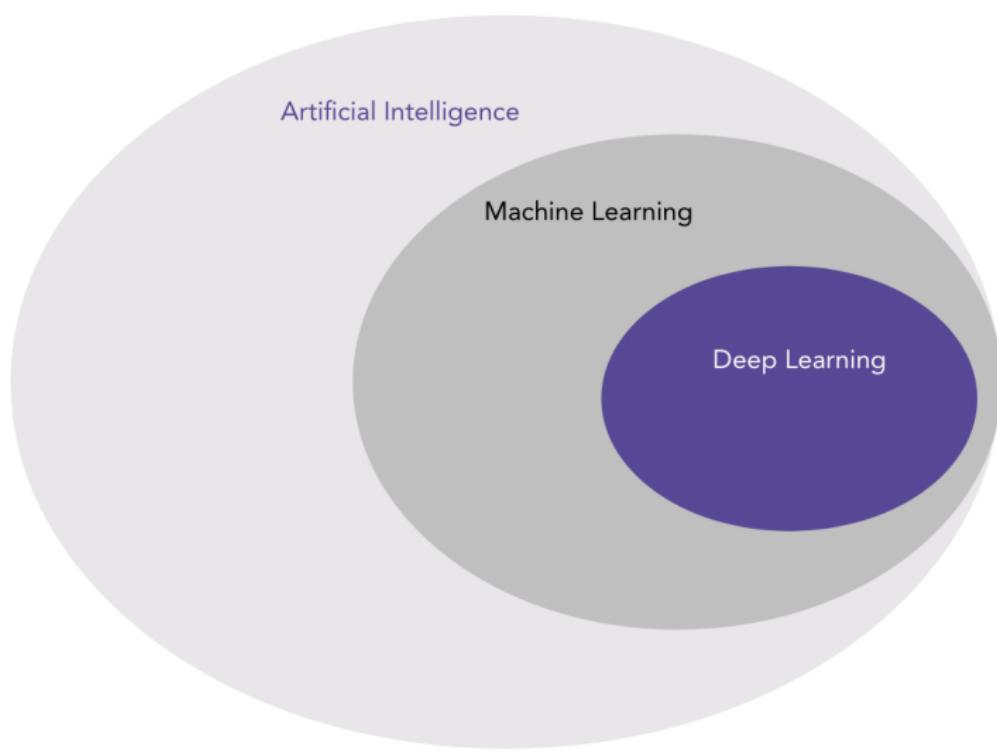
Element AI

negar@elementai.com

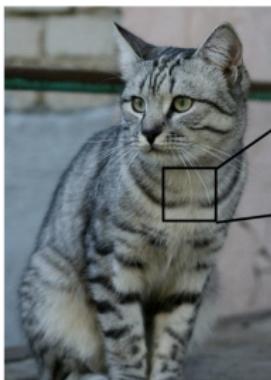
May 22, 2018



What are Artificial Intelligence, Machine Learning and Deep Learning?



What is Computer Vision?



This image by Nkita is
licensed under CC-BY 2.0

[1]185 112 188 111 184 99 186 99 98 183 112 119 184 97 93 871
[1]76 85 98 185 128 185 87 96 95 99 115 112 186 183 99 95
[1]99 81 81 93 128 131 127 180 95 98 182 99 96 93 181 94]
[186 91 81 64 69 91 98 85 181 187 180 98 75 84 96 95]
[1]186 112 119 184 111 184 99 186 99 98 183 112 119 184 97 93 871
[1]333 137 147 183 65 81 88 65 52 54 74 64 182 93 85 821
[128 137 144 148 189 95 86 79 62 65 63 63 68 73 86 181]
[125 132 140 137 119 121 117 64 65 79 88 65 54 64 72 90]
[122 132 135 137 119 121 117 64 65 79 88 65 54 64 72 90]
[115 114 189 123 159 148 131 131 118 113 189 180 92 74 65 72 70]
[89 93 98 97 186 147 131 118 113 114 113 186 186 95 77 80]
[43 65 82 89 78 73 88 181 124 126 130 181 187 114 131 139]
[1]62 65 82 89 78 73 88 181 124 126 130 181 187 114 131 139
[1]63 65 75 88 89 71 62 81 128 138 135 180 85 98 118 110]
[87 65 71 87 186 95 69 45 76 138 126 187 92 94 185 132]
[1]88 65 71 87 186 95 69 45 76 138 126 187 92 94 185 132
[1]164 146 132 89 82 128 124 184 76 48 45 86 89 181 182 189]
[157 178 157 128 93 86 114 132 112 97 69 55 78 82 99 94]
[139 128 131 161 139 189 189 133 121 117 114 87 65 51 69 90]
[122 121 182 88 83 112 153 149 122 189 184 75 88 107 112 99]
[122 121 182 88 83 86 94 117 145 148 153 182 58 70 92 187]
[122 164 148 183 71 56 78 83 93 183 119 139 182 61 69 841]

What the computer sees

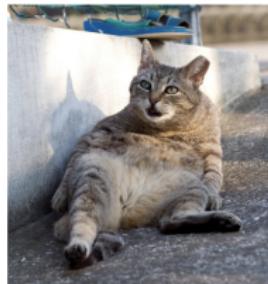
An image is just a big grid of numbers between [0, 255]:

e.g. 800 x 600 x 3
(3 channels RGB)

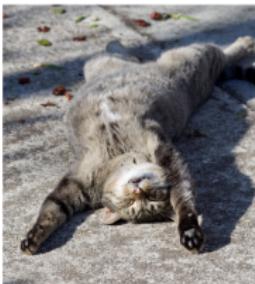
Figure: Picture from cs231n 2017

Computer Vision: Challenges!

Challenges: Deformation



This image by Umberto Salvagnin
is licensed under CC-BY 2.0



This image by Umberto Salvagnin
is licensed under CC-BY 2.0



This image by sarebear is
licensed under CC-BY 2.0



This image by Tom Thie is
licensed under CC-BY 2.0

Computer Vision: Challenges!

Challenges: Occlusion



[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)



[This image by jonsong is licensed under CC-BY 2.0](#)

Computer Vision: Challenges!

Challenges: Background Clutter



[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)

Computer Vision: Challenges!

Challenges: Illumination



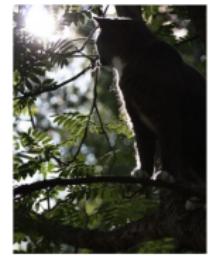
[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)

Computer Vision: Challenges!

Challenges: Intraclass variation



This image is CC0 1.0 public domain

Visual data!



Top row, left to right:
Image by Augustas.Dilijėnas; licensed under CC BY-SA 3.0; changes made
Image by Nestor is licensed under CC BY-SA 2.0
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain

Middle row, left to right:
Image by BGPHP Conference is licensed under CC BY 2.0; changes made
Image is CC0 1.0 public domain
Image by NASA is licensed under CC BY 2.0; chang
Image is CC0 1.0 public domain

Bottom row, left to right:
Image is CC0 1.0 public domain
Image by Derek Keats is licensed under CC BY 2.0; changes made
Image is public domain
Image by Ted Eytan is licensed under CC BY-SA 2.0; changes made

When did it all start?

- In the late 1960s, computer vision began at institutes that were pioneer in artificial intelligence.
- It was meant to mimic the human visual system, as a stepping stone to endowing robots with intelligent behavior.
- In 1966, a summer project was defined at the University of MIT to solve the Computer Vision problem!

The story of a summer project!

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Evaluational Big-Bang?

Evolution's Big Bang



This image is licensed under CC-BY 2.5



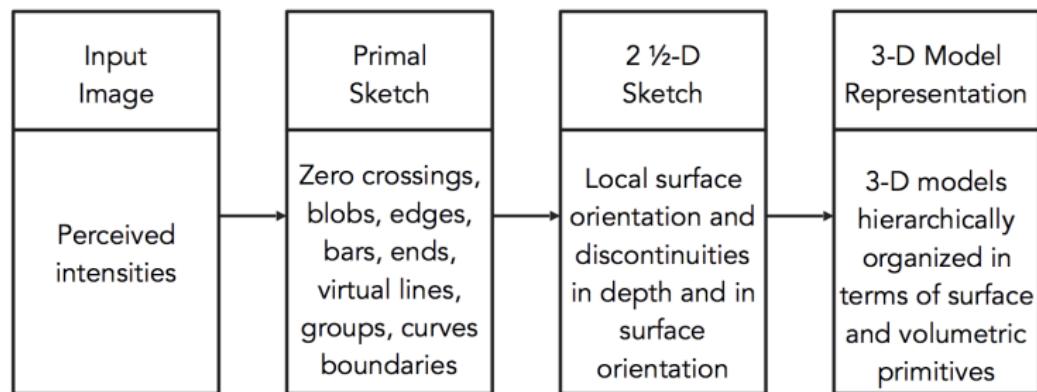
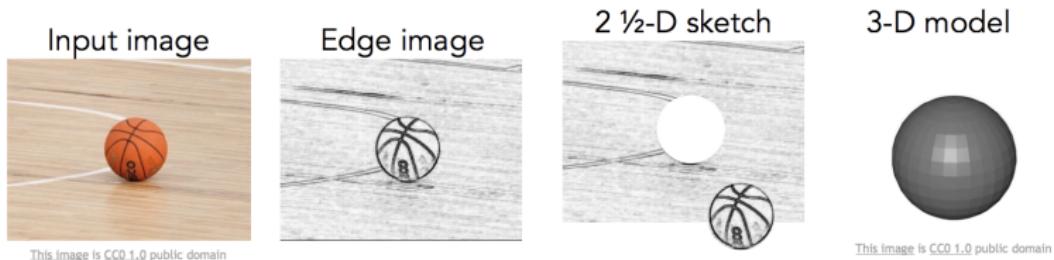
This image is licensed under CC-BY 2.5



This image is licensed under CC-BY 3.0

543million years, B.C.

Stages of Visual representation!



Stages of Visual Representation, David Marr, 1970s

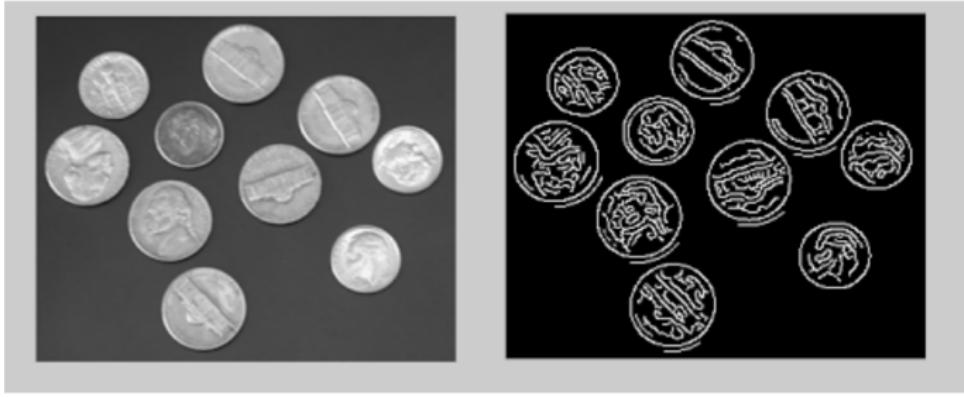
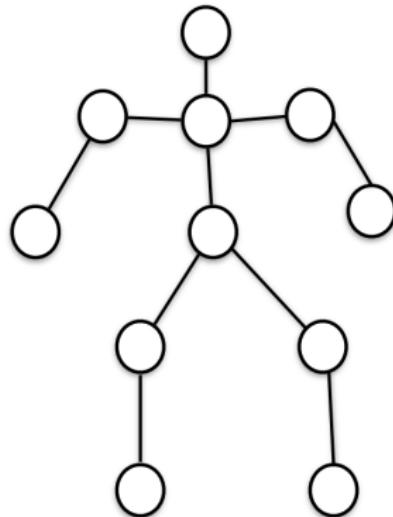
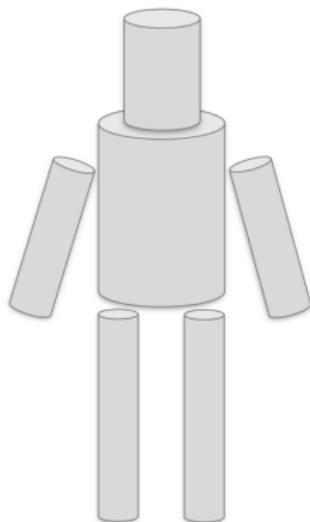


Image is CC BY-SA 4.0

David Lowe, 1987

- Generalized Cylinder
Brooks & Binford, 1979
- Pictorial Structure
Fischler and Elschlager, 1973



Learning to Segment the images!

Normalized Cut (Shi & Malik, 1997)

Image is CC BY 3.0



Image is public domain



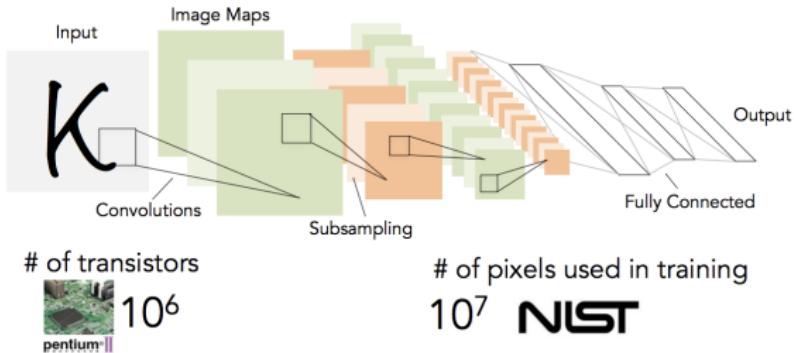
Image is CC-BY-SA 3.0



Convolutional Neural Networks!

1998

LeCun et al.



2012

Krizhevsky et al.

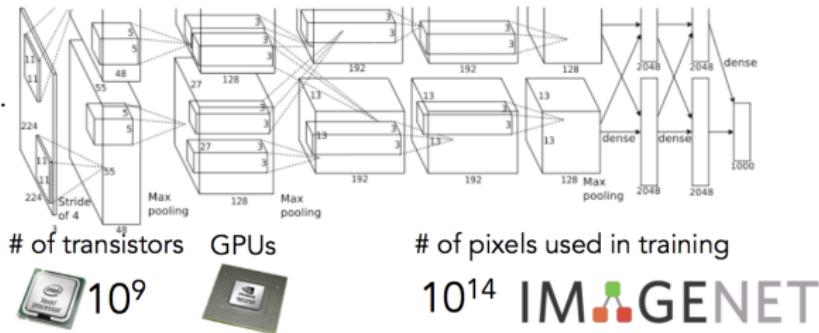
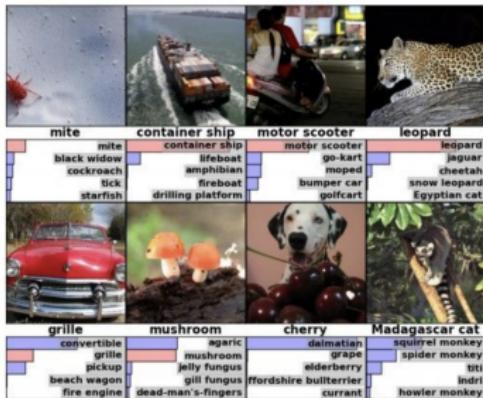


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

What do we expect Machines to comprehend?

Classification



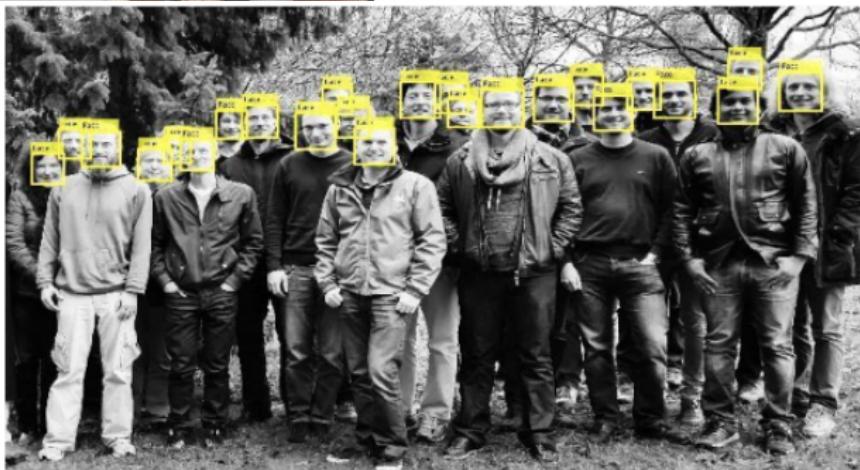
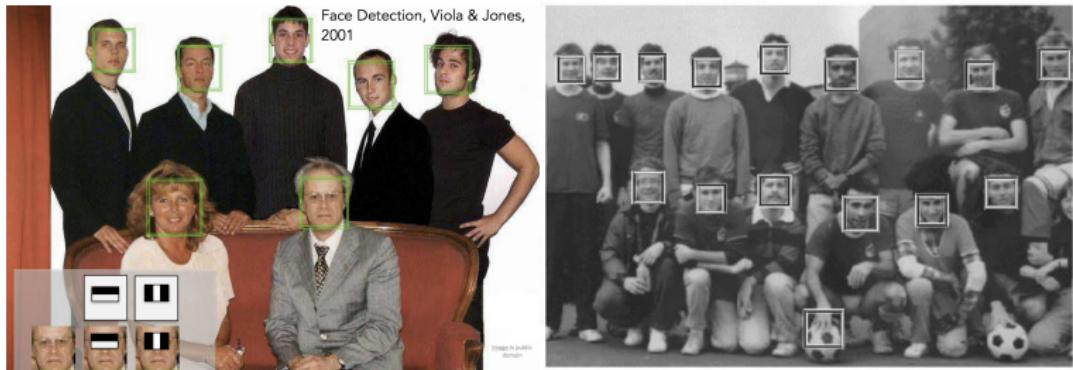
Retrieval



Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Image from Stanford CS231 course

What do we expect Machines to comprehend?



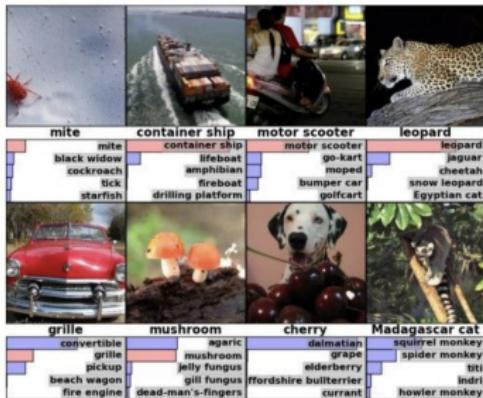
What do we expect Machines to comprehend?



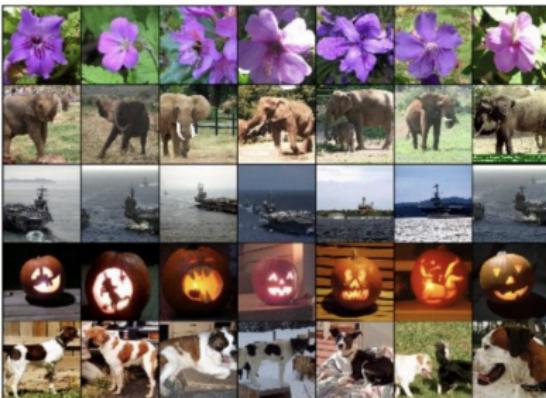
Gao et al, IEEE, 2017

What do we expect Machines to comprehend?

Classification



Retrieval



Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

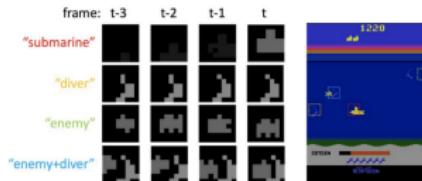
Image from Stanford CS231 course

What do we expect Machines to comprehend?



Images are examples of pose estimation, not actually from Toshev & Szegedy 2014. Copyright Lane McIntosh.

[Toshev, Szegedy 2014]

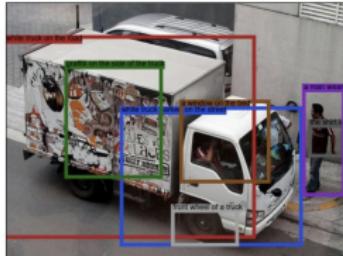


Figures copyright Xiaoxiao Guo, Satinder Singh, Honglak Lee, Richard Lewis, and Xiaoshi Wang, 2014. Reproduced with permission.

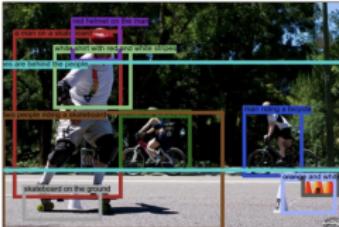
[Guo et al. 2014]

Image from Stanford CS231 course

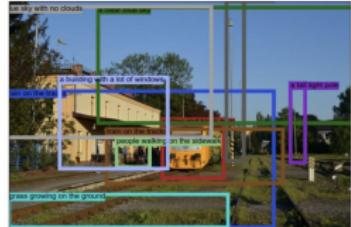
What do we expect Machines to comprehend?



a white truck on the road. white truck parked on the street. the shirt is red. graffiti on the side of the truck. a window on the bed. a man wearing a black shirt. front wheel of a truck.



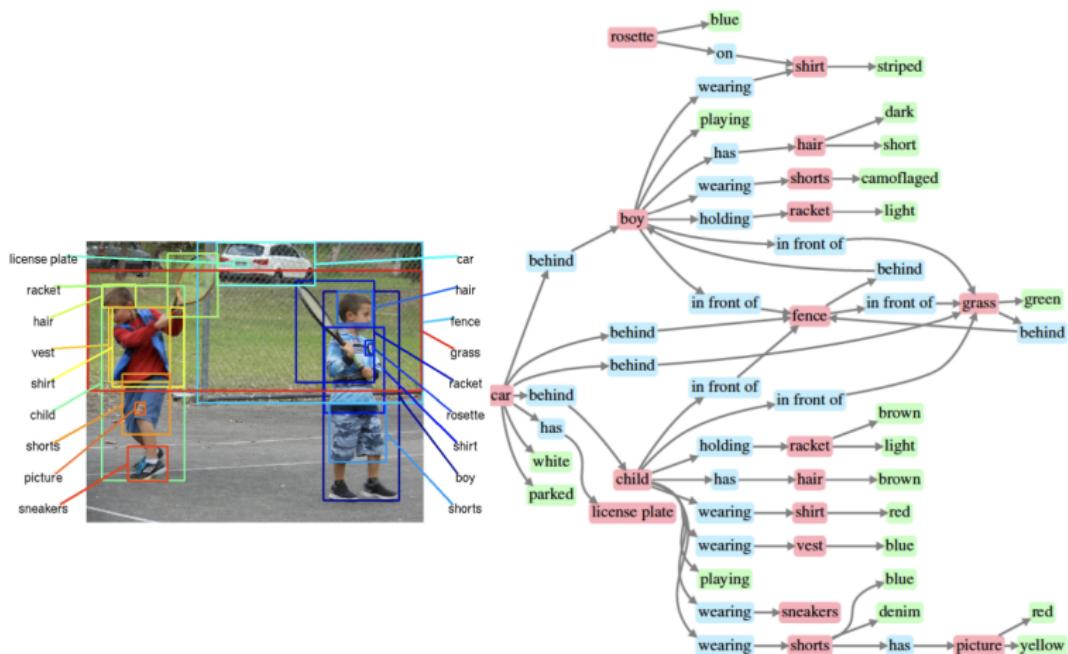
a man on a skateboard. man riding a bicycle. orange cone on the ground. man riding a bicycle. two people riding a skateboard. red helmet on the man. skateboard on the ground. white shirt with red and white stripes. orange and white cone. trees are behind the people.



yellow and black train. train on the tracks. a tall light pole. a clear blue sky. train on the tracks. a tall light pole. a blue sky with no clouds. people walking on the sidewalk. a building with a lot of windows. grass growing on the ground.

Johnson et al, CVPR 2017.

What do we expect Machines to comprehend?



Johnson et al., "Image Retrieval using Scene Graphs", CVPR 2015

Figures copyright IEEE, 2015. Reproduced for educational purposes

Image from Stanford CS231 course

What do we expect Machines to comprehend?

No errors



A white teddy bear sitting in the grass

Minor errors



A man in a baseball uniform throwing a ball

Somewhat related



A woman is holding a cat in her hand

Image Captioning

[Vinyals et al., 2015]
[Karpathy and Fei-Fei, 2015]



A man riding a wave on top of a surfboard



A cat sitting on a suitcase on the floor



A woman standing on a beach holding a surfboard

All images are CC0 Public domain:
<https://pixabay.com/en/luggage-antique-cat-1643010/>
<https://pixabay.com/en/teddy-plush-bear-cute-teddy-bear-1623436/>
<https://pixabay.com/en/kurt-wim-summer-sport-lawn-chair-16716/>
<https://pixabay.com/en/hands-folded-prayer-meditation-adult-5613967/>
<https://pixabay.com/en/baseball-player-shortstop-infed-1045263/>

Captions generated by Justin Johnson using NeuralTalk2

Image from Stanford CS231 course

What do we expect Machines to comprehend?

Fast-forward to today: ConvNets are everywhere

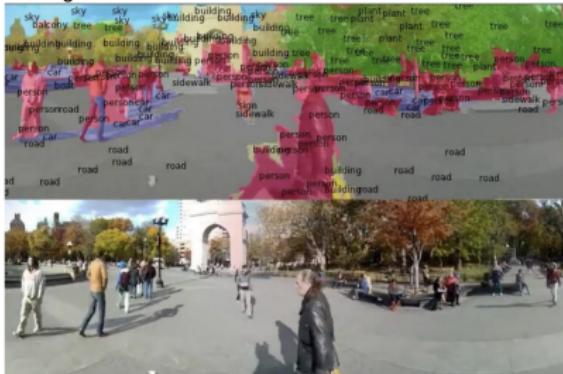
Detection



Figures copyright Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, 2015. Reproduced with permission.

[Faster R-CNN: Ren, He, Girshick, Sun 2015]

Segmentation

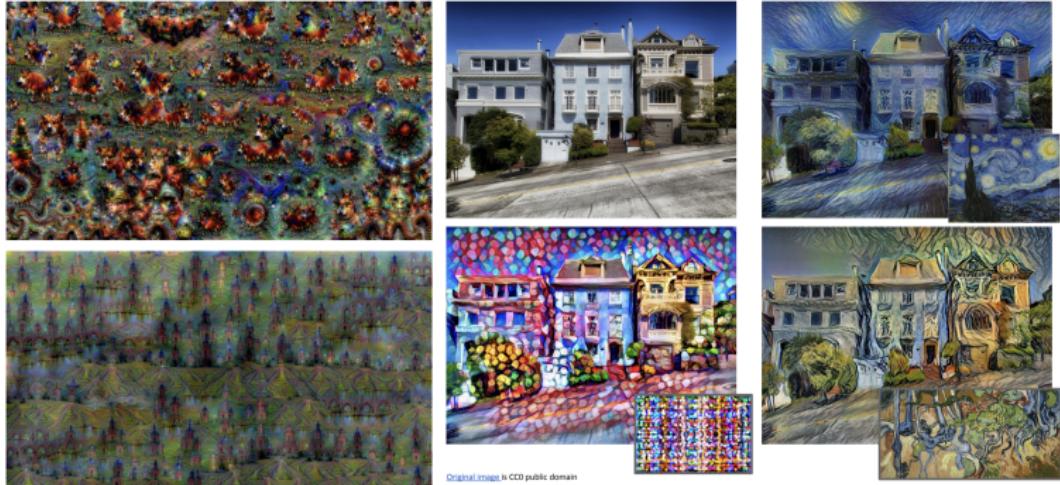


Figures copyright Clement Farabet, 2012.
Reproduced with permission.

[Farabet et al., 2012]

Image from Stanford CS231 course

What do we expect Machines to comprehend?



Figures copyright Justin Johnson, 2015. Reproduced with permission. Generated using the Inceptionism approach from a [blog post](#) by Google Research.

Original image is CC0 public domain
Starry Night and Tree House by Van Gogh are in the public domain
Bokeh image is in the public domain
Styled images copyright Justin Johnson, 2017;

Gatys et al., "Image Style Transfer using Convolutional Neural Networks", CVPR 2016
Gatys et al., "Controlling Perceptual Factors in Neural Style Transfer", CVPR 2017

Image from Stanford CS231 course

What do we expect Machines to comprehend?

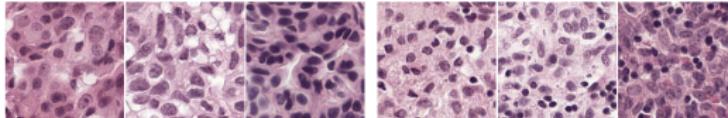


Fig. 1. Left: three tumor patches and right: three challenging normal patches.

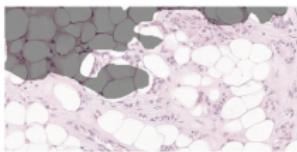


Fig. 2. Difficulty of pixel-accurate annotations for scattered tumor cells. Ground truth annotation is overlaid with a lighter shade. Note that the tumor annotations include both tumor cells and normal cells e.g., white space representing adipose tissue (fat).

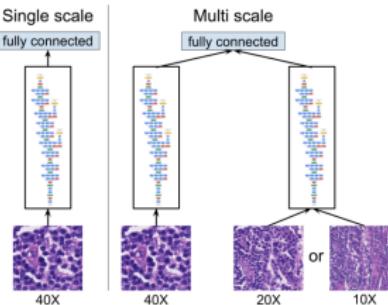


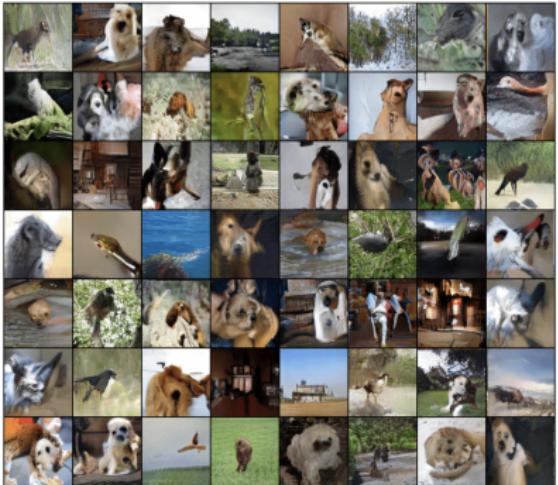
Fig. 3. The three colorful blocks represent Inception (V3) towers up to the second-last layer (PreLogit). Single scale utilizes one tower with input images at 40X magnification; multi-scale utilizes multiple (e.g., 2) input magnifications that are input to separate towers and merged.

Liu et al, 2017.

What do we expect Machines to comprehend?



(a) CelebA



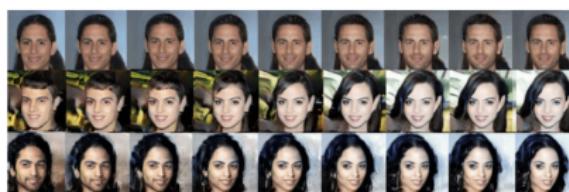
(b) ImageNet128

HALI: Hierarchical Adversarial Learned Inference, 2018.

What do we expect Machines to comprehend?



(a) CelebA orientation variation



(b) CelebA gender variation



(c) CelebA lipstick feature map variation



(d) CelebA saturation feature map variation

HALI: Hierarchical Adversarial Learned Inference, 2018.

What do we expect Machines to comprehend?

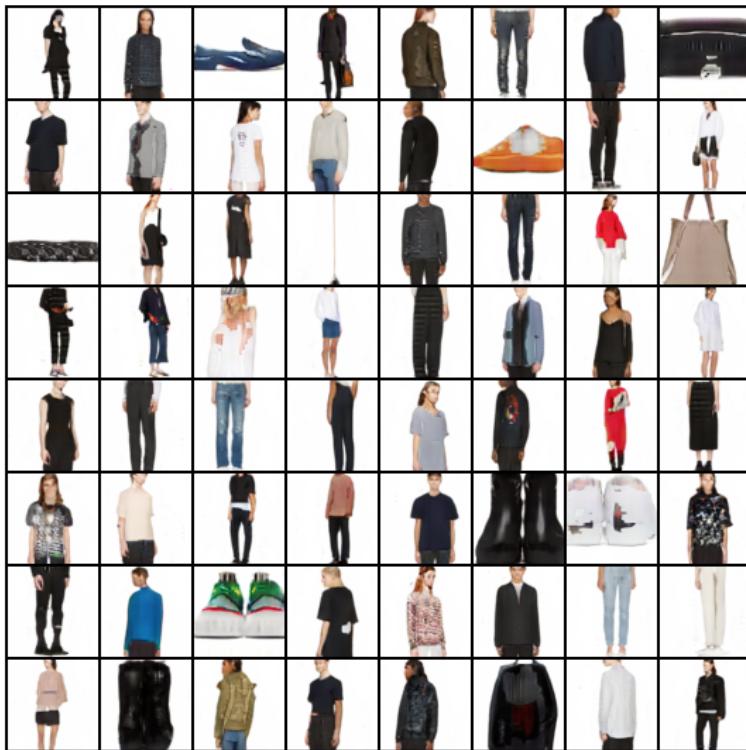


HALI: Hierarchical Adversarial Learned Inference, 2018.

Generating Fashion images!



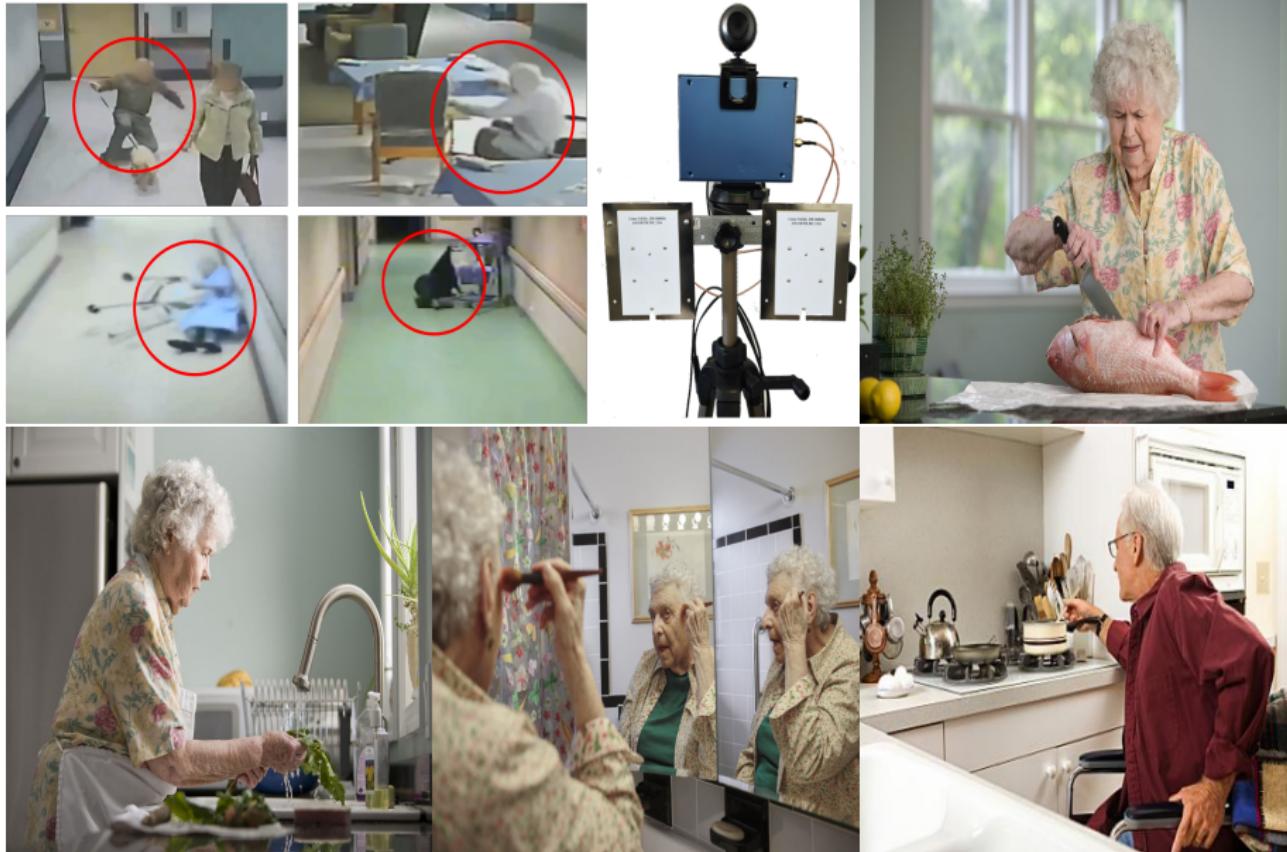
Generating Fashion images!



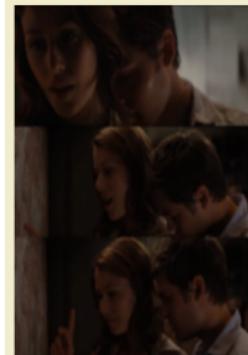
Video applications!



Video applications!



Every day a new application and scenario!



He plants a tender kiss on her shoulder.



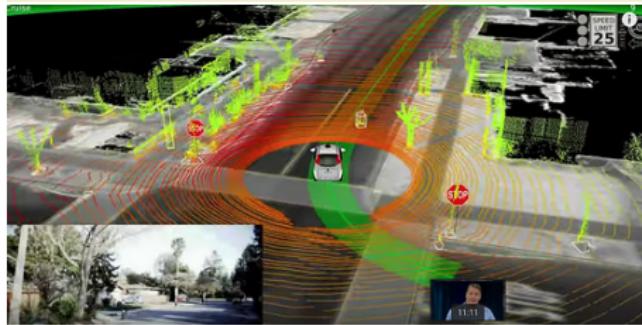
His vanity license plate reads 732.



SOMEONE sits on her roommate bed.



1



¹Pictures are from the LSMDC dataset.

Computer Vision can improve our lives!



Outside border images, clockwise, starting from top left:

Image by Pop Culture Geek is licensed under CC BY 2.0; changes made
Image by the US Government is in the public domain
Image by the US Government is in the public domain
Image by Glogger is licensed under CC BY-SA 3.0; changes made
Image by Sylenius is licensed under CC BY 3.0; changes made
Image by US Government is in the public domain

Inside four Images, clockwise, starting from top left:

Image is CC0 1.0 public domain
Image by Tucania is licensed under CC BY-SA 3.0; changes made
Image by Intuitive Surgical, Inc. is licensed under CC BY-SA 3.0; changes made
Image by Oyundari Zorigtbeatar is licensed under CC BY-SA 4.0

Back in the Future?!



What are the applications of Computer Vision that you can think of?