

Visual Recognition and Deep Learning



1

Nilanjan Ray

Associate Professor

Computing Science

University of Alberta

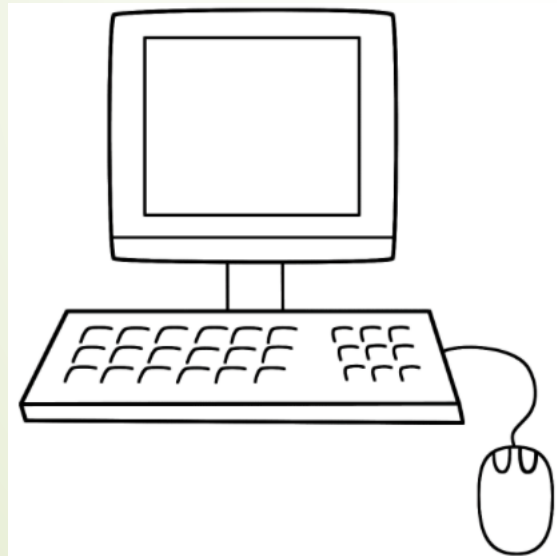
Outline

- What is visual recognition (aka computer vision)?
- What is deep learning?
- Success stories of deep learning
- Limitations of deep learning
- What lies ahead

What is visual recognition?

Visual recognition...

... is teaching computers to see



Humans see...





Computers see...



Teaching computers to “see” like humans



“see” not just RGB data

- IR (infrared) etc
- ToF camera (Time of Flight)
 - ‘range’ camera gives depth
- Medical
 - ultrasonography
 - MRI
- & more



LIDAR
(laser radar)



Kinect



Example: Automated animal recognition



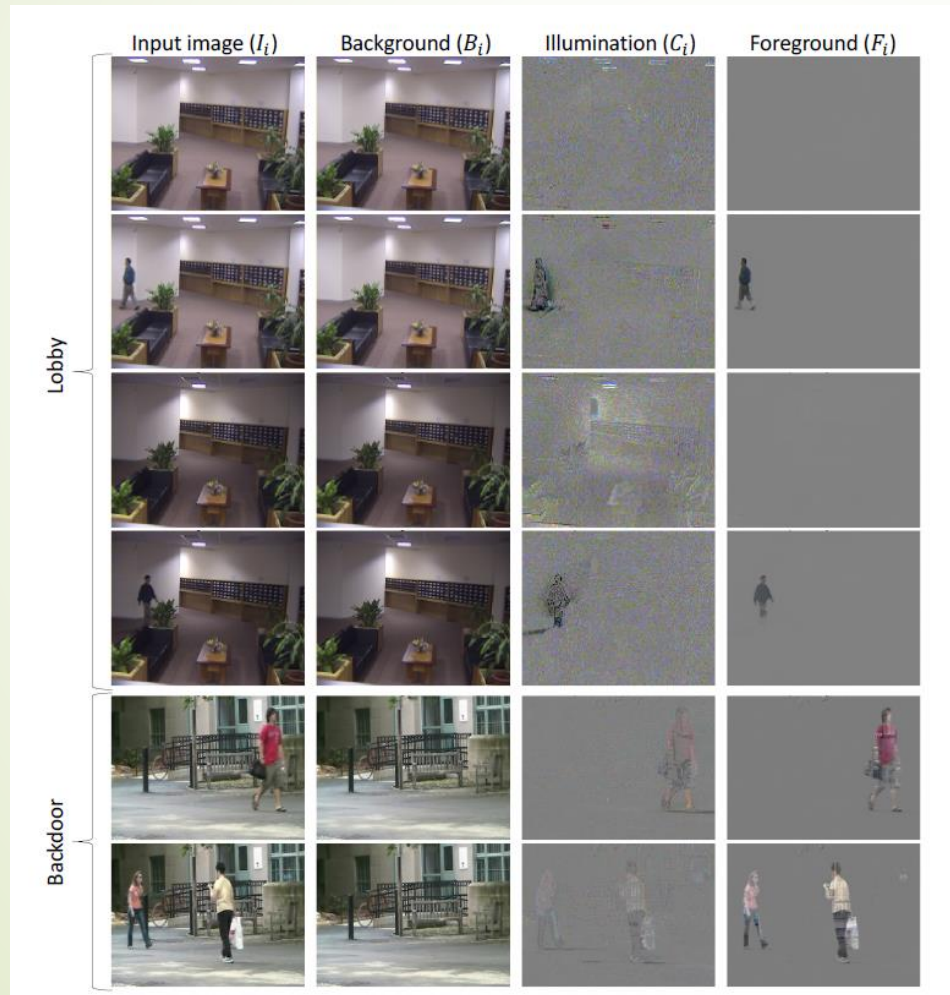
Research with ACAMP
(<https://www.acamp.ca/>)

Example: Vehicle detection and visual tracking

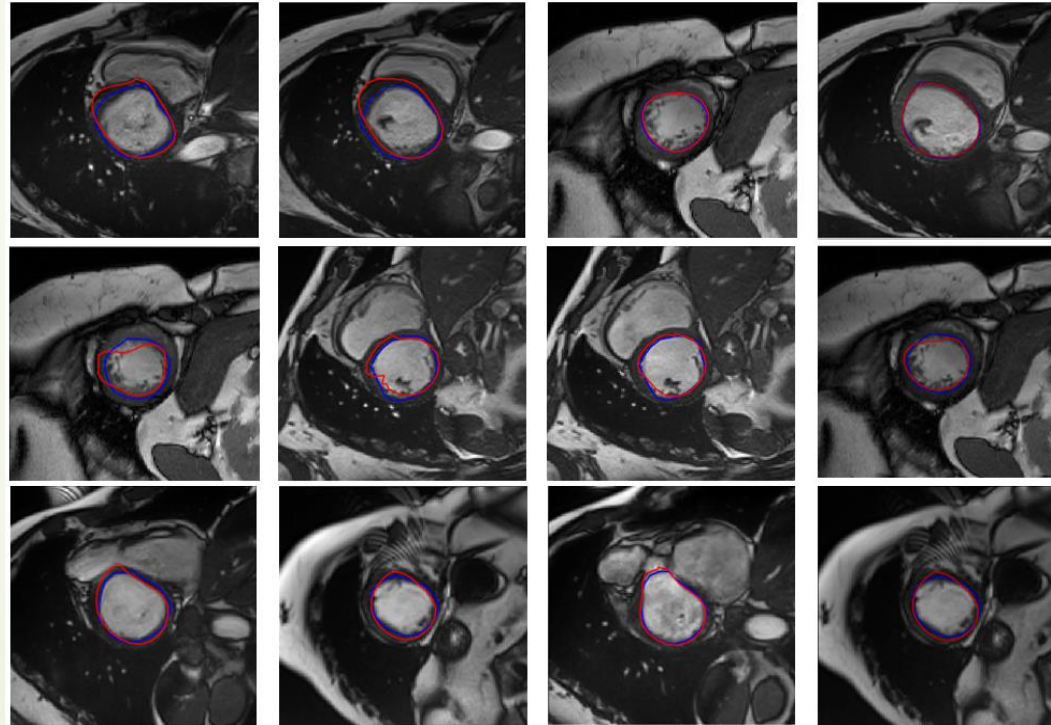


Research with ISL Engineering
(<http://islengineering.com/>)

Example: Background modeling and moving object detection

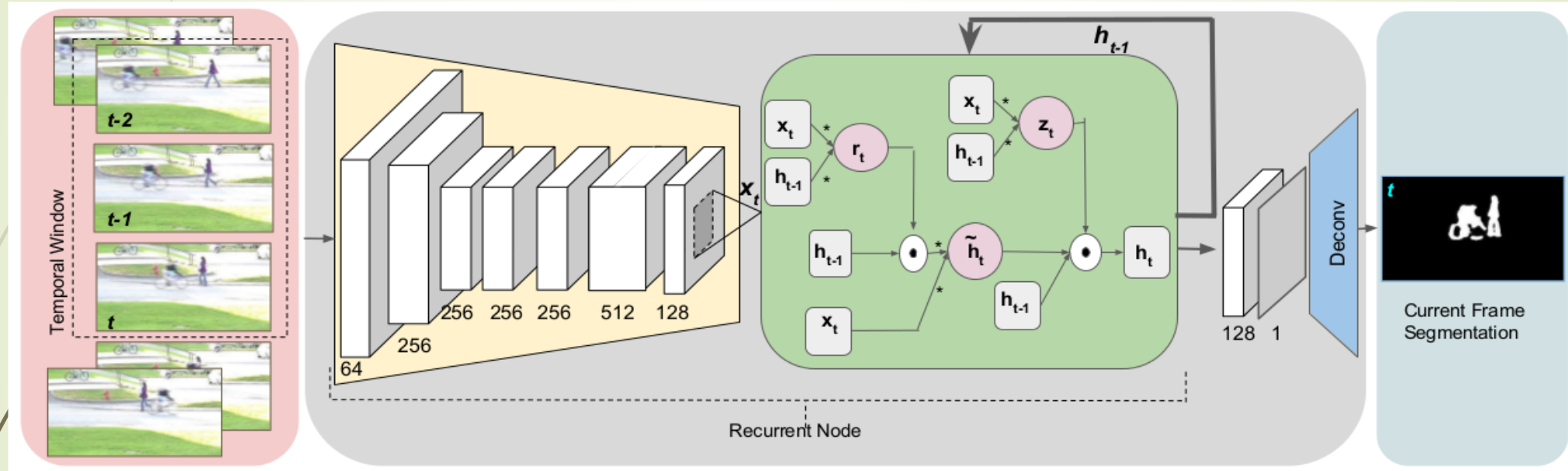


Example: Deformable image registration



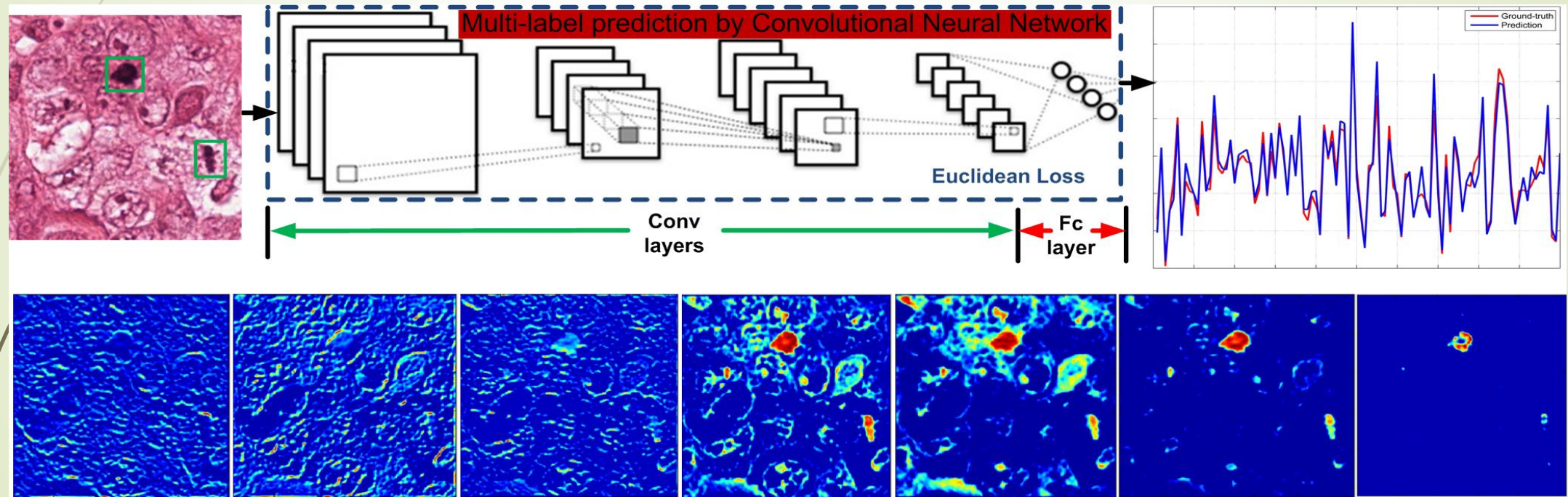
<https://openreview.net/forum?id=HkmkmW2jM>

Example: Video segmentation



<https://ieeexplore.ieee.org/abstract/document/8296851/>

Example: Cell detection from microscopy image



The holy grail of computer vision/visual recognition

The grand goal of computer vision is human-like, automated scene understanding.

This is where the state of the art (deep learning algorithm) today:



"construction worker in orange safety vest is working on road."



"a young boy is holding a baseball bat."



"a horse is standing in the middle of a road."

Picture source: <http://cs.stanford.edu/people/karpathy/deepimagesent/>

What is deep learning?

How does computer vision work?

- ▶ Let's consider object recognition tasks.



- ▶ How do we describe a cat or a dog or a ...?
- ▶ It is impossible to write “explicit rules” to recognize visual objects/scenes.
- ▶ Much easier to learn from **examples**!

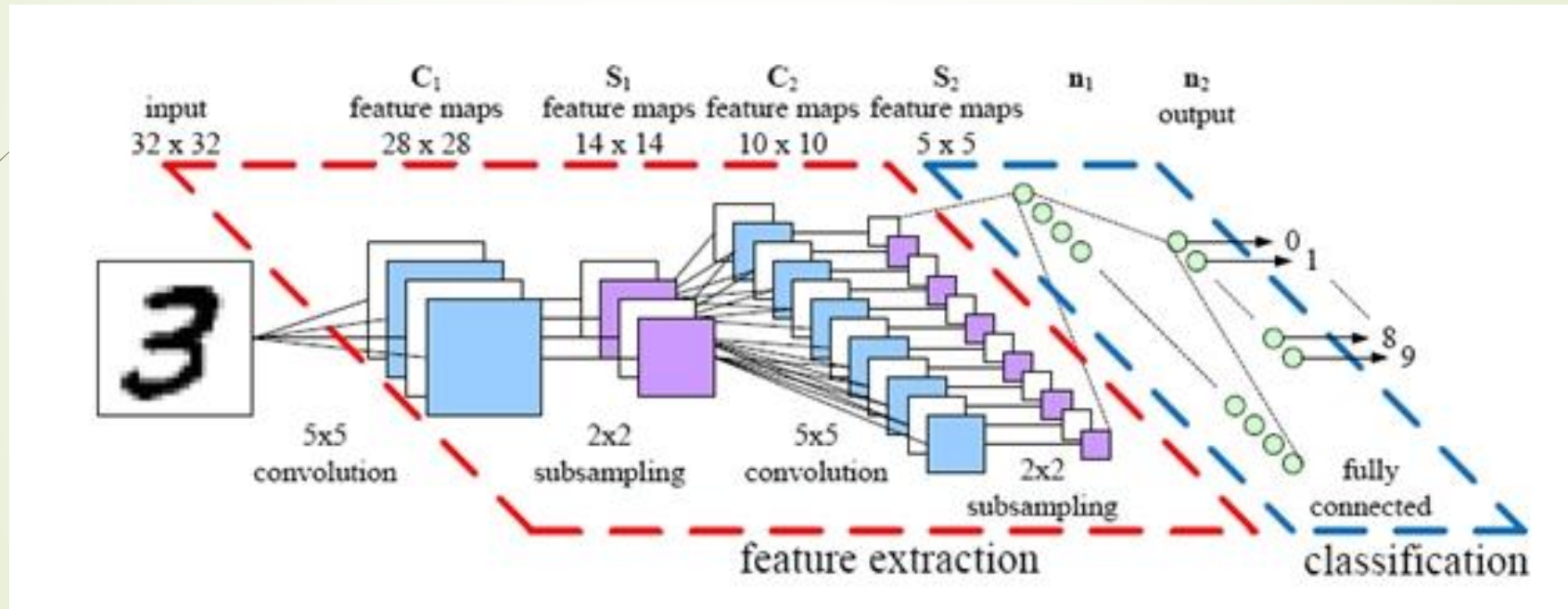
Machine learning is indispensable for computer vision!

Computer vision with deep learning

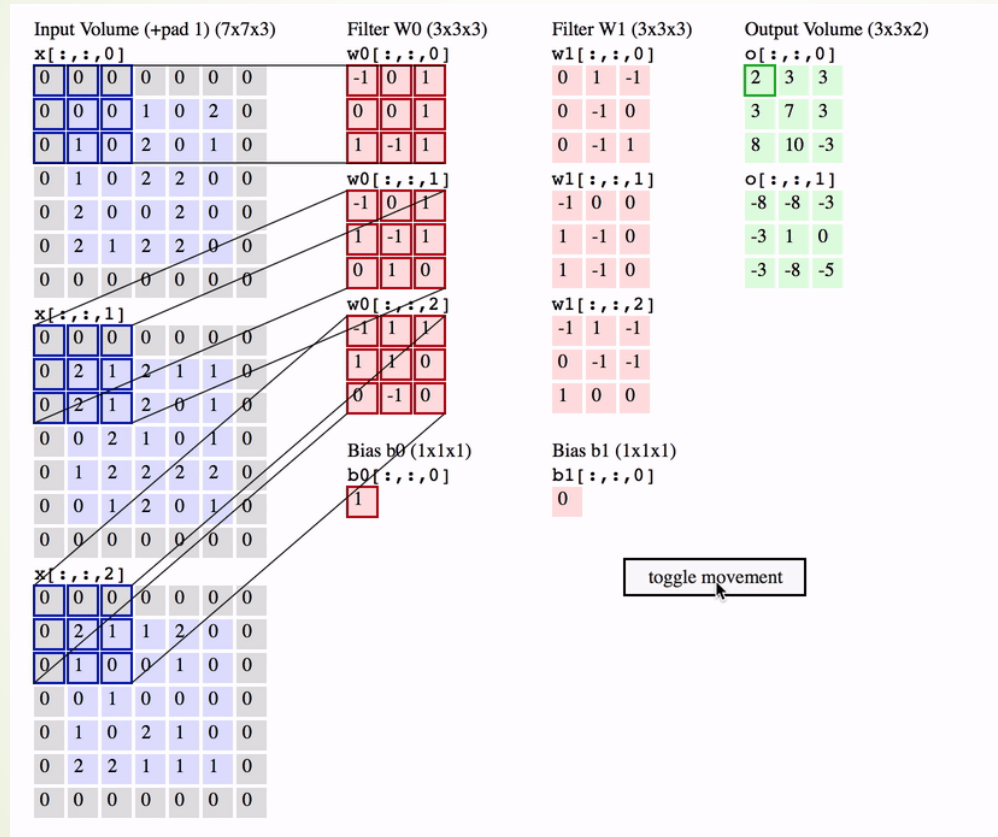
Deep learning

=

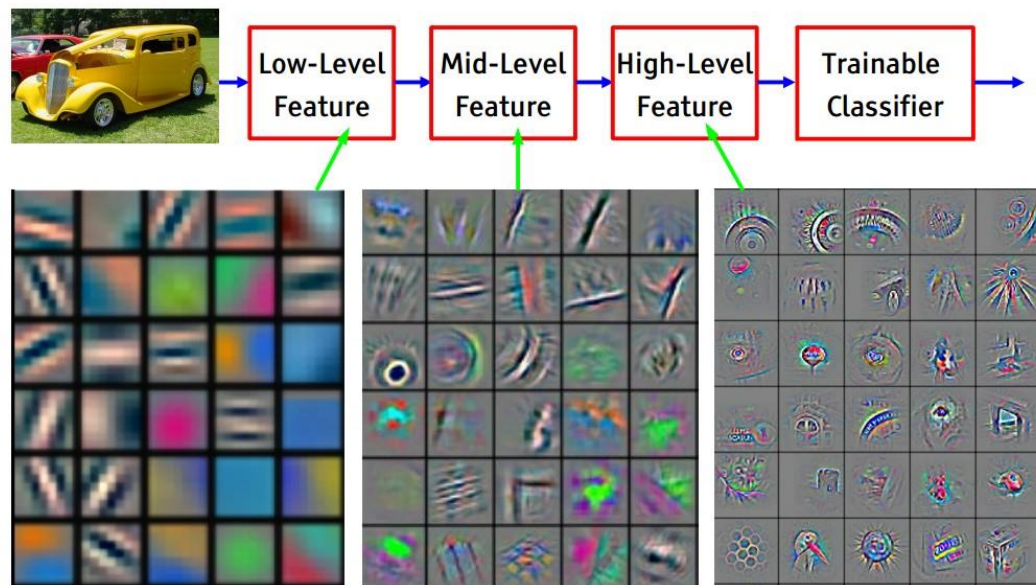
Convolutional neural network with many layers



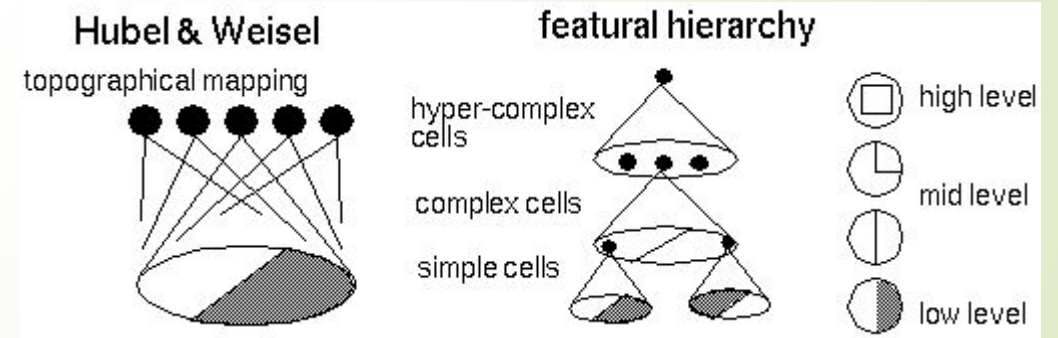
Convolution revolution!



Hierarchical representation by deep learning



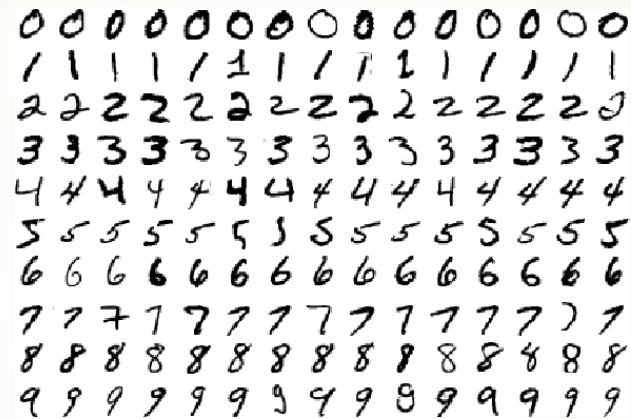
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]



Hubel & Weisel, '62

How do you teach a deep learner?

- Step 1: Create training image set (example set):

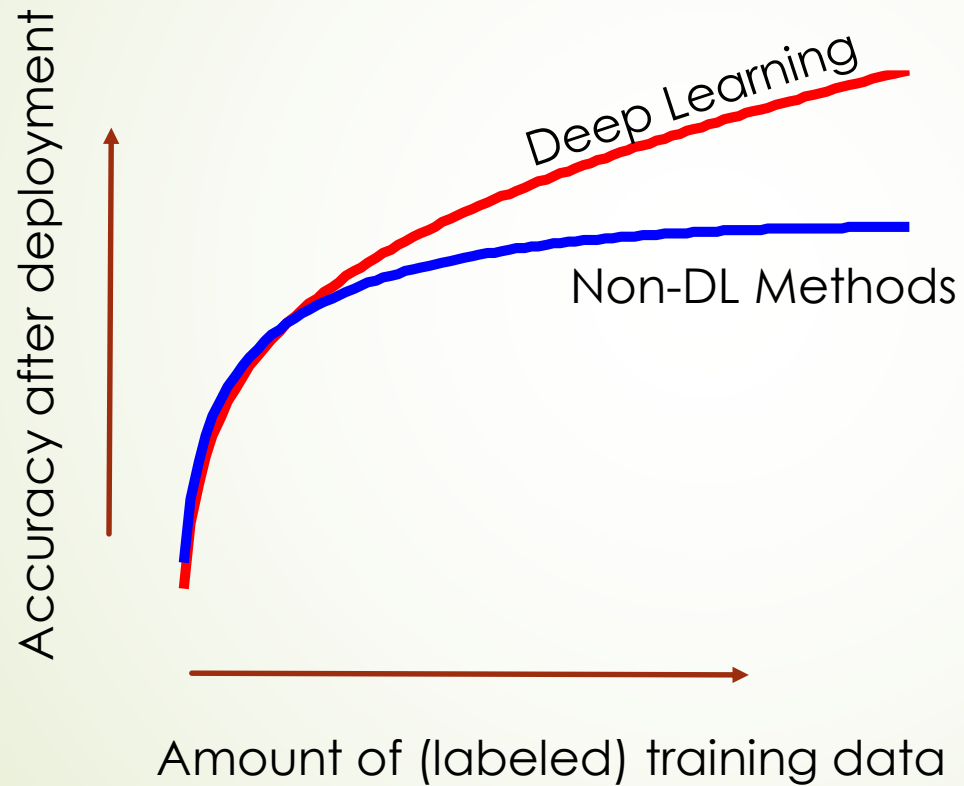


A 10x10 grid of handwritten digits, representing a training image set. The digits are arranged in rows by value: the first row contains 10 zeros, the second row contains 10 ones, and so on, up to the tenth row which contains 10 nines. Each digit is a small, stylized handwritten character.

Repeat steps 2, 3 and 4

- Step 2: Show these examples to the deep learner
- Step 3: Measure mistakes made by the deep learner
- Step 4: Tune (millions of) parameters of the deep learner to minimize its mistakes

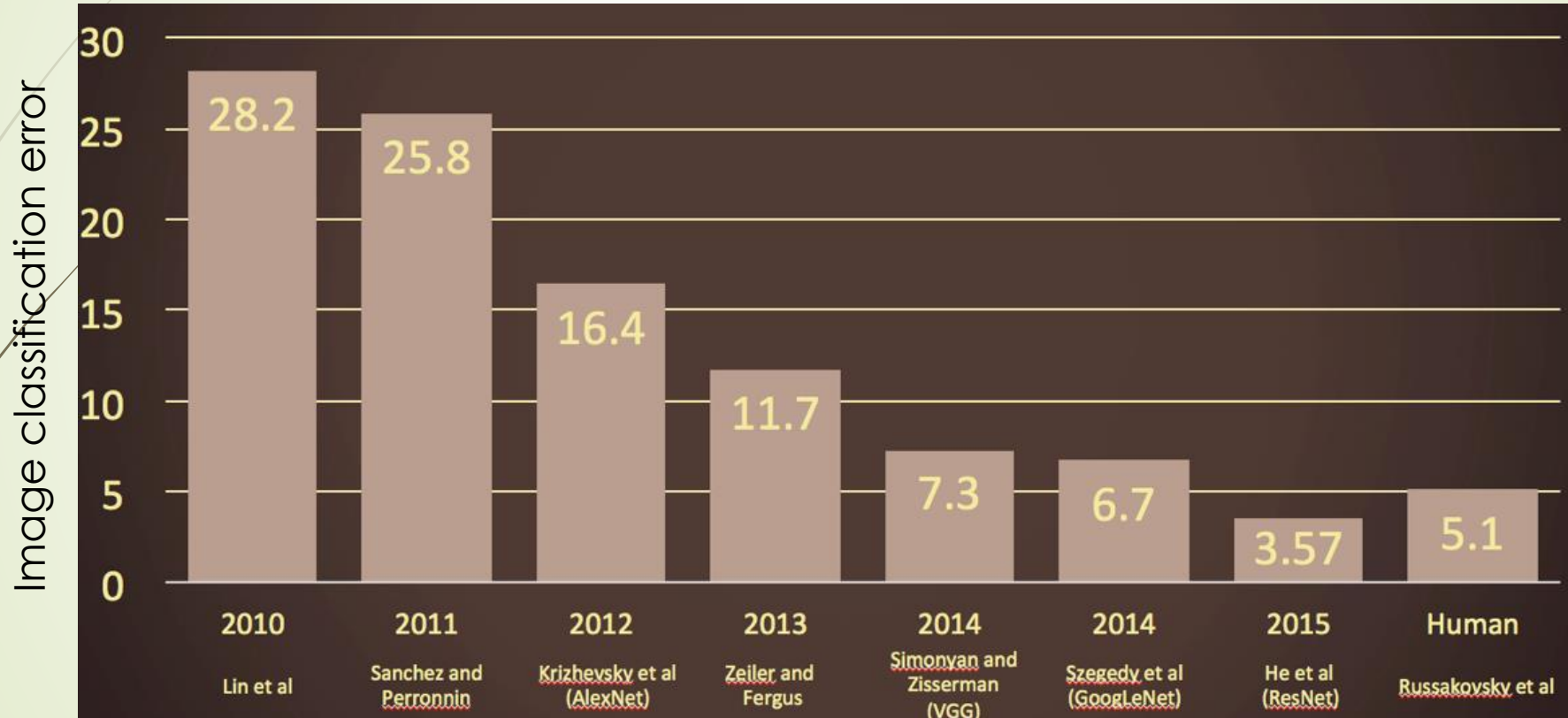
Traditional vs. deep learning...



Success stories of deep learning

Image classification results

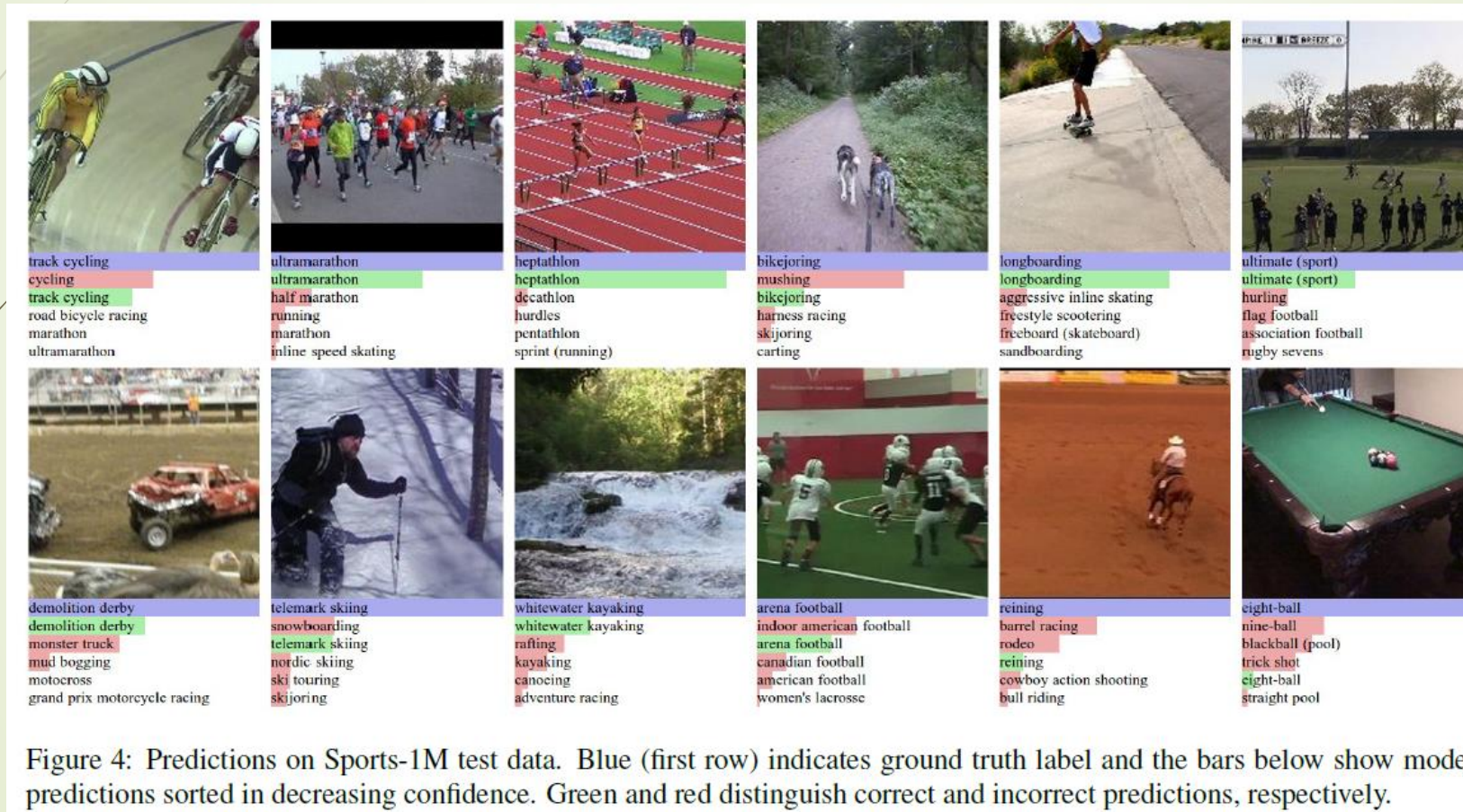
ImageNet- Large scale visual recognition challenge: 1000 categories, 1,000,000 images



Computer vision has surpassed human level performance on this benchmark!

Picture courtesy: <http://cs231n.stanford.edu/index.html>

Large scale video classification



<http://cs.stanford.edu/people/karpathy/deepvideo/>

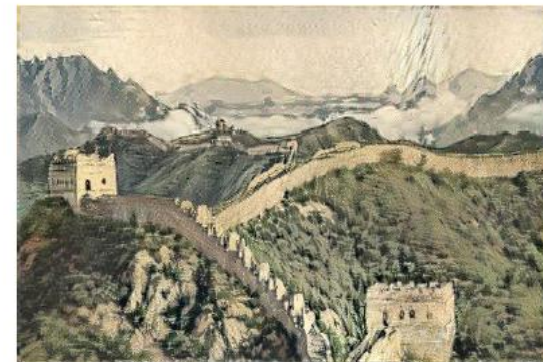
Style Transfer



(a) Content



(b) Style



(c) Content + Style

Figure 1. Example of using the Neural Style Transfer algorithm of Gatys *et al.* to transfer the style of Chinese painting (b) onto The Great Wall photograph (a). The painting that served as style is named “Dwelling in the Fuchun Mountains” by Gongwang Huang.

Photorealistic image generation



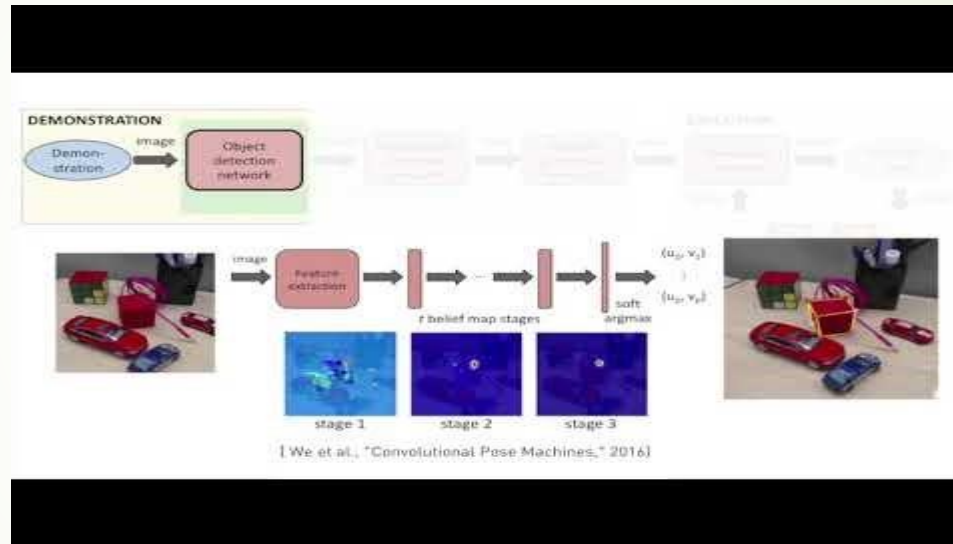
From NVIDIA research: <https://arxiv.org/pdf/1710.10196v1.pdf>

Deep reinforcement learning



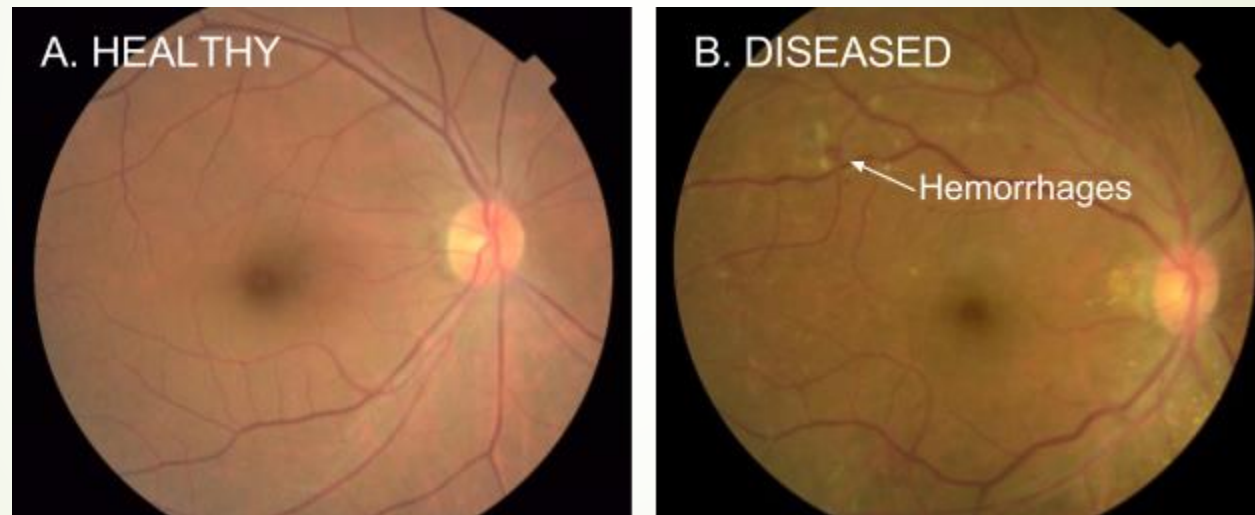
Picture source: <https://deepmind.com/blog/deep-reinforcement-learning/>

Impressive robotics with deep learning



<https://www.youtube.com/watch?v=B7ZT5oSnRys>

Diabetic retinopathy using deep learning



Deep learning and natural language processing

- Impressive developments are happening in the NLP space
- Word embedding
- Language translation
- Language modeling
-

<http://ruder.io/nlp-imagenet/>

What created this revolution?

- Lots and lots of labeled data (such as ImageNet)
- Compute power (parallel processing with GPUs)
- Good old back-prop algorithm + only a few new tweaks! And
- Open source software platforms: TensorFlow, PyTorch,...

Limitations of deep learning

Challenge 1: Labeling of training data

- Manual labeling of lots of images and videos

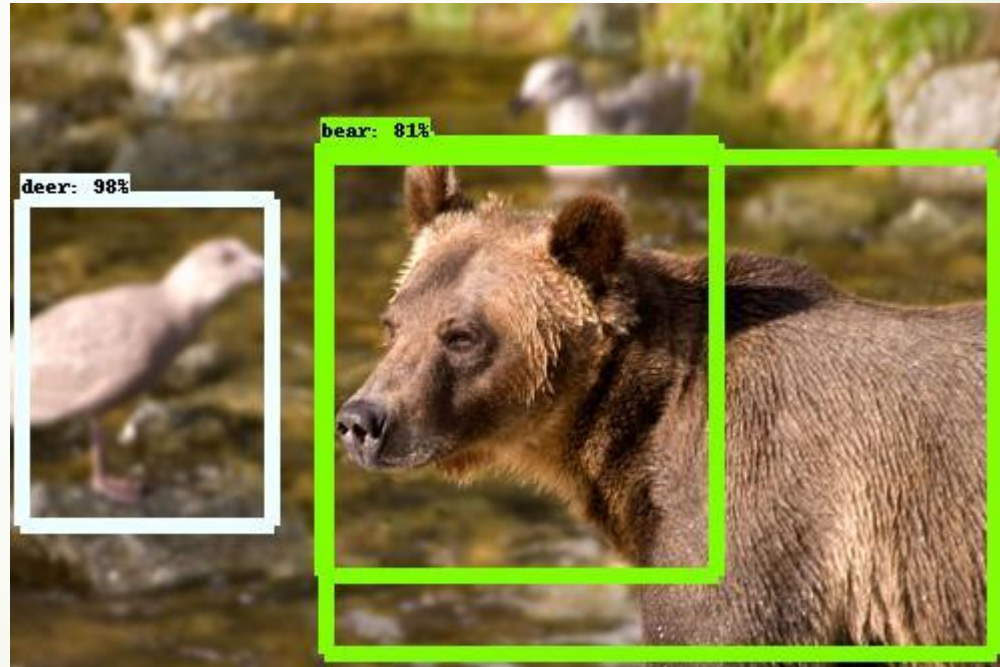
IMGENET

www.image-net.org

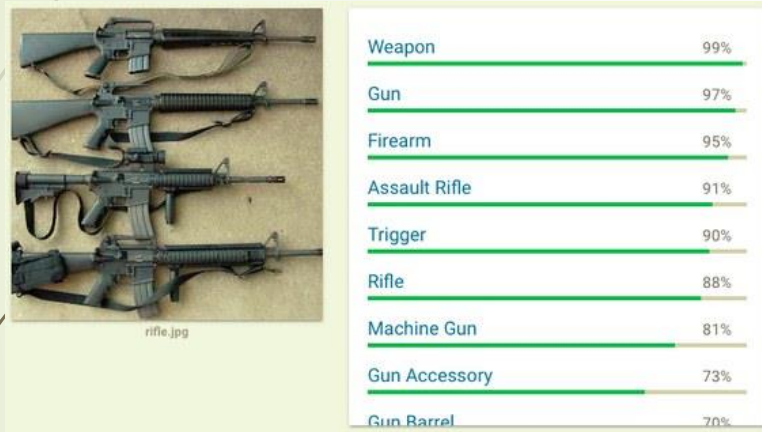
22K categories and **14M** images

- Animals
 - Bird
 - Fish
 - Mammal
 - Invertebrate
- Plants
 - Tree
 - Flower
- Food
- Materials
- Structures
 - Artifact
 - Tools
 - Appliances
 - Structures
- Person
 - Scenes
 - Indoor
 - Geological Formations
 - Sport Activities

Challenge 2: Task specific / narrow scope



Challenge 3: Can be easily fooled!



Google AI was fooled to believe rifles are helicopters

<https://www.wired.com/story/researcher-fooled-a-google-ai-into-thinking-a-rifle-was-a-helicopter/>

Challenge 4: Lack of interpretability

- Interpretability = explaining decision making to humans
- Deep learning systems have poor interpretability
- European Union drafted the General Data Protection Regulation, which will require some interpretability of AI algorithms (<https://www.eugdpr.org/>, <https://arxiv.org/abs/1606.08813>)

Challenge 5: training set bias

- ▶ Leading commercial gender recognition software (Google, Microsoft, etc.) from face images are biased
- ▶ More accurately detect white male faces, than other races, females!

Challenge 6: theoretical understanding is work in progress

- Why deep neural networks are able to optimize learning cost functions with stochastic gradient descent?
- How does depth help?
- Why deep neural networks do not overfit (generalize)?

Deep learning is far from human level intelligence!

- Not even close to a two year human being – a toddler can learn basic laws of physics on his/her own!
- Cannot do “common sense” decision making
- Does not understand causality
- Cannot perform counterfactual inferences well
- Cannot answer open-ended inference problems
- How to go beyond data?

These are open questions in AI research

What to expect with deep learning

Predicting future of technology is foolish!

- Task specific applications of deep learning
- Some scientific discoveries will be assisted by deep learning, but human intervention would be required
- Whenever we can collect data (labeled data), deep learning will automate the area sooner or later
- Extrapolation, open-ended inference, and creative areas will be continued to be dominated by humans for a long time
- We are nowhere near singularity!