

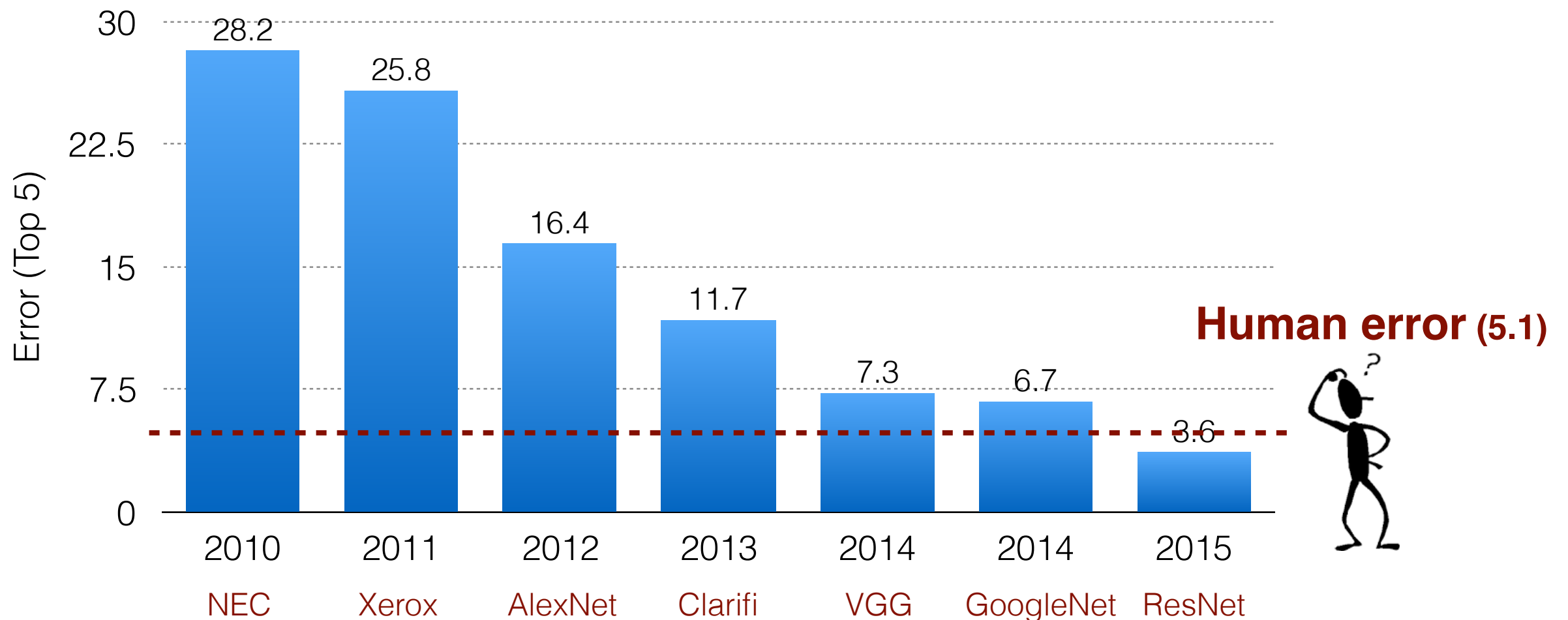
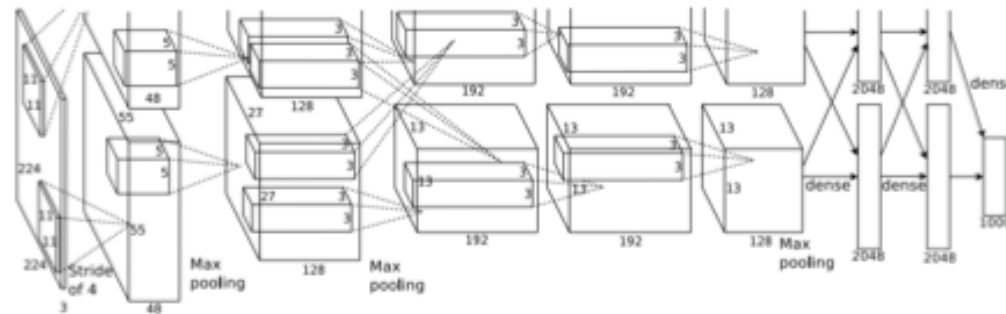
Why is deep learning so cool?

Lamberto Ballan

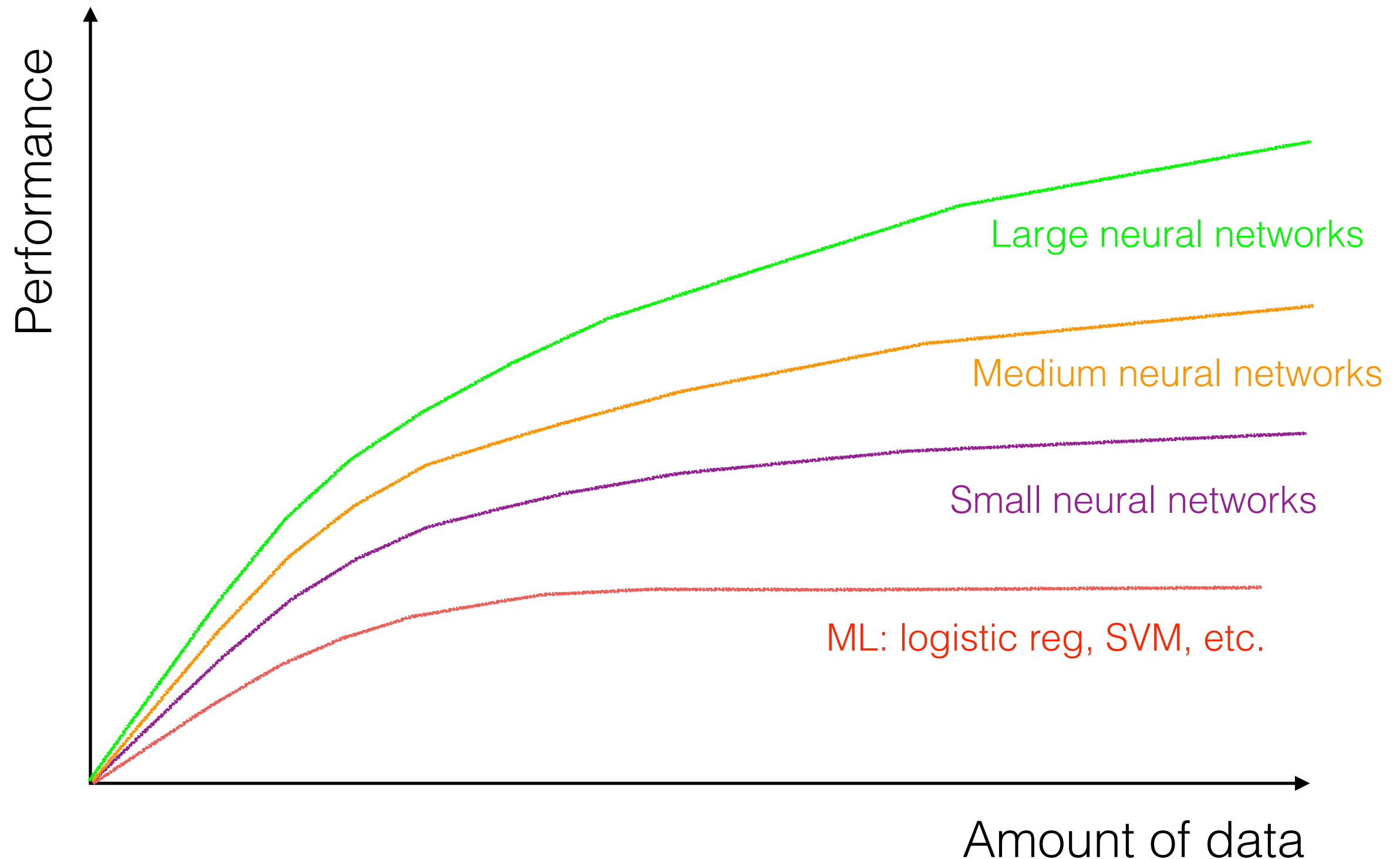
University of Padova, Italy

Visual recognition results

- ImageNet-ILSVRC (classification) over the years



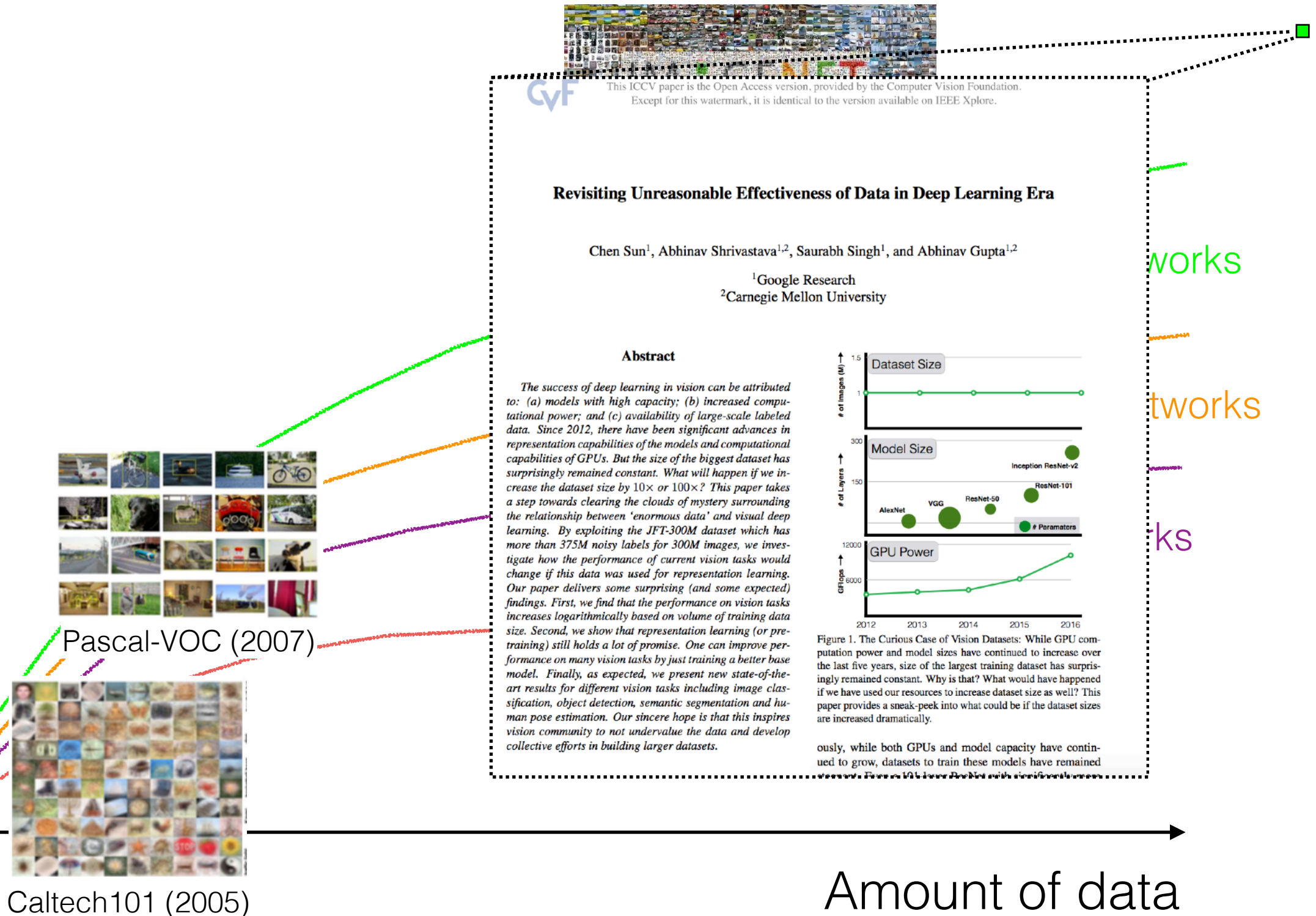
Scale drive deep learning progress



Adapted from slide by Andrew Ng

Scale drive deep learning progress

Performance



CVF

This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

Revisiting Unreasonable Effectiveness of Data in Deep Learning Era

Chen Sun¹, Abhinav Shrivastava^{1,2}, Saurabh Singh¹, and Abhinav Gupta^{1,2}

¹Google Research

²Carnegie Mellon University

Abstract

The success of deep learning in vision can be attributed to: (a) models with high capacity; (b) increased computational power; and (c) availability of large-scale labeled data. Since 2012, there have been significant advances in representation capabilities of the models and computational capabilities of GPUs. But the size of the biggest dataset has surprisingly remained constant. What will happen if we increase the dataset size by $10\times$ or $100\times$? This paper takes a step towards clearing the clouds of mystery surrounding the relationship between 'enormous data' and visual deep learning. By exploiting the JFT-300M dataset which has more than 375M noisy labels for 300M images, we investigate how the performance of current vision tasks would change if this data was used for representation learning. Our paper delivers some surprising (and some expected) findings. First, we find that the performance on vision tasks increases logarithmically based on volume of training data size. Second, we show that representation learning (or pre-training) still holds a lot of promise. One can improve performance on many vision tasks by just training a better base model. Finally, as expected, we present new state-of-the-art results for different vision tasks including image classification, object detection, semantic segmentation and human pose estimation. Our sincere hope is that this inspires vision community to not undervalue the data and develop collective efforts in building larger datasets.

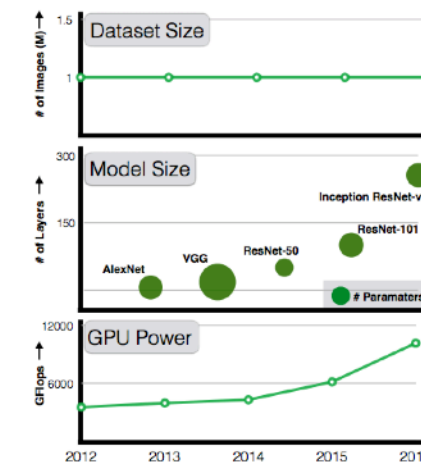
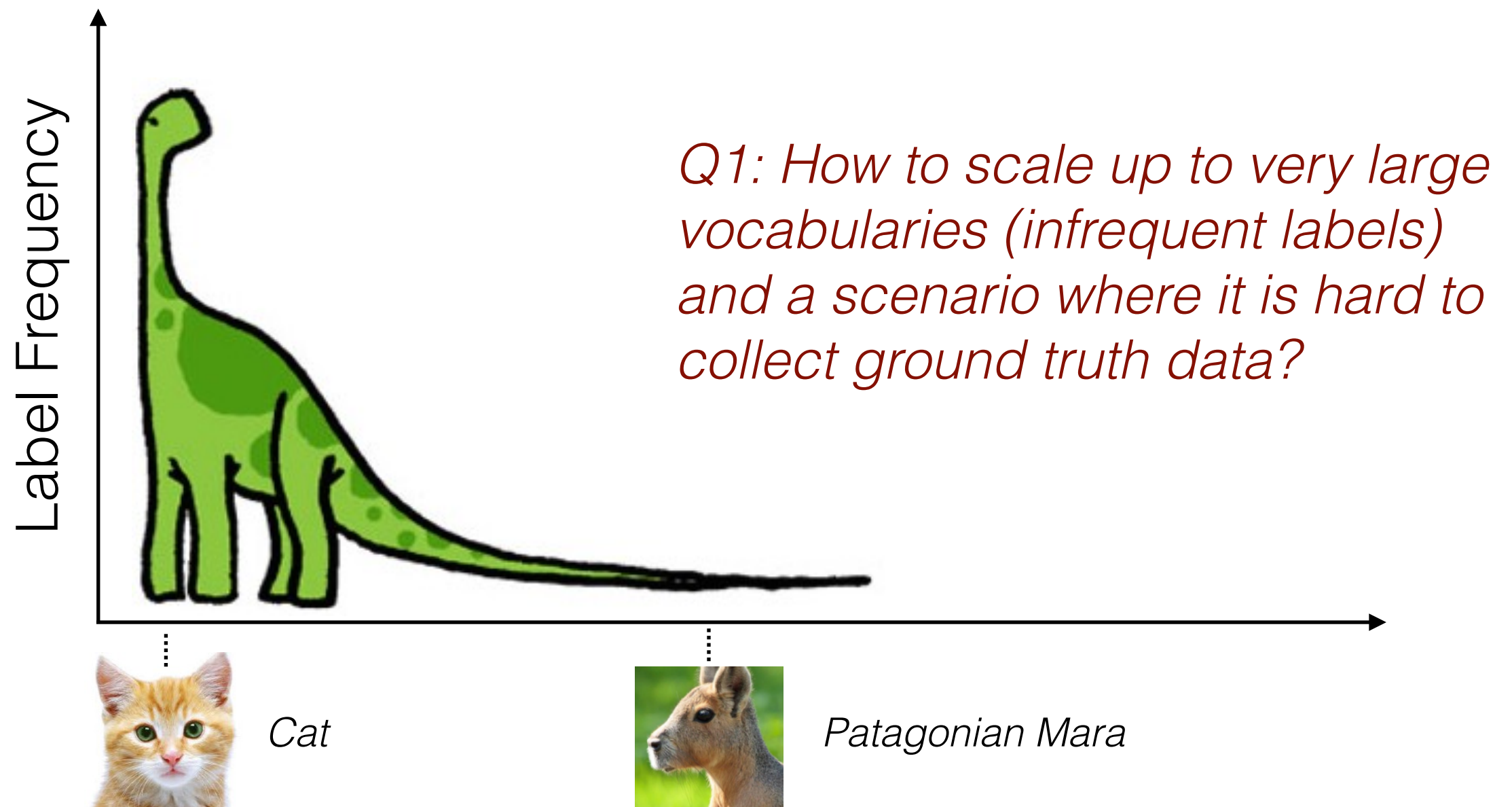


Figure 1. The Curious Case of Vision Datasets: While GPU computation power and model sizes have continued to increase over the last five years, size of the largest training dataset has surprisingly remained constant. Why is that? What would have happened if we have used our resources to increase dataset size as well? This paper provides a sneak-peek into what could be if the dataset sizes are increased dramatically.

ously, while both GPUs and model capacity have continued to grow, datasets to train these models have remained steady. For example, ResNet with significantly more

This is super cool but...

- **The long tail:** a small number of objects / entities appear very often while most others appear rarely



How children learn to see



by observing lots
(tons) of data

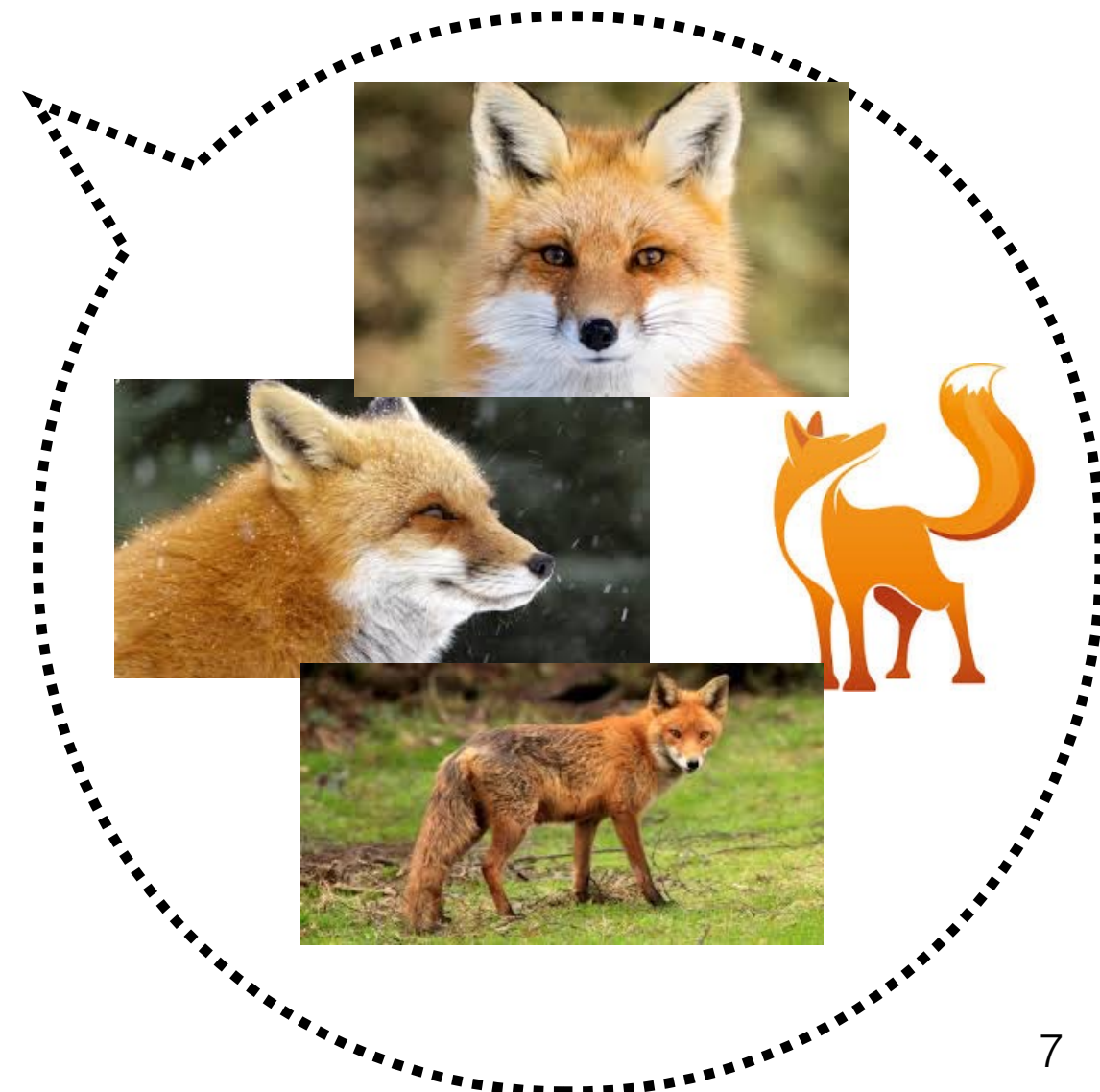
Fox

(Linda loves foxes)

How children learn to see

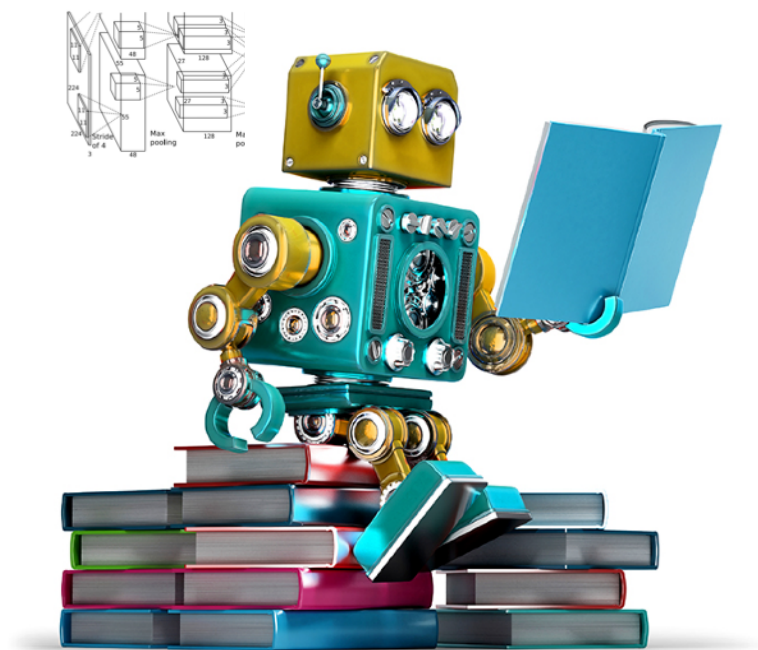


by observing lots
(tons) of data



Again, CNNs are cool but...

- Even two year old kids are good at abstraction, and are able to generalize and transfer their knowledge



Q2: Our algorithms are very good at memorising categories, but are we really going toward full image understanding?



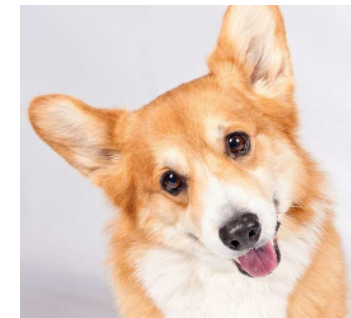
Cat



Dog



*Cardigan
Welsh Corgi*



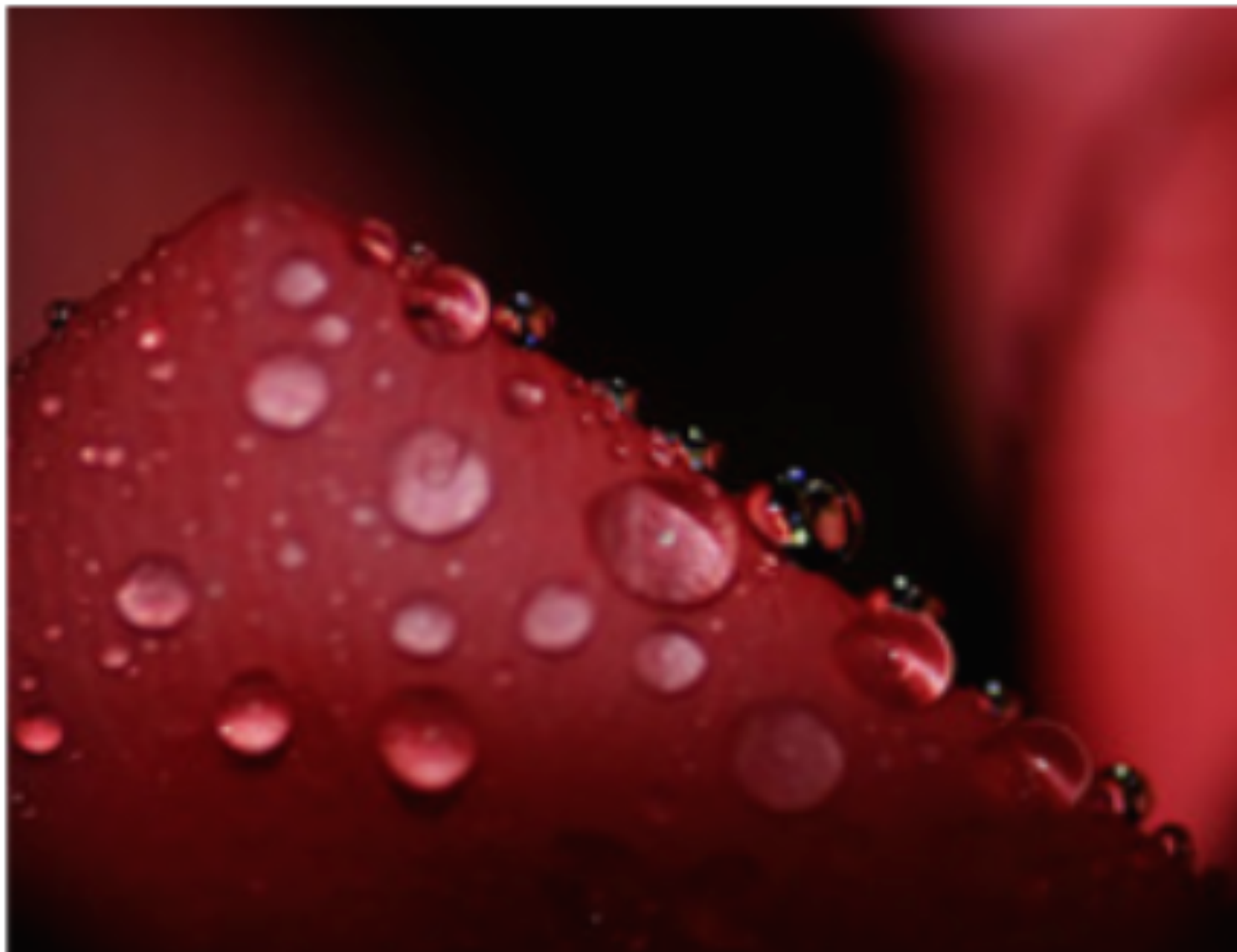
*Pembroke
Welsh Corgi*

Breakthroughs in object recognition

... and fine-grained image recognition

Sharing knowledge

- One example (*i.e.* an attempt to address Q1) in the context of social media such as Flickr images



flickr

Tags:

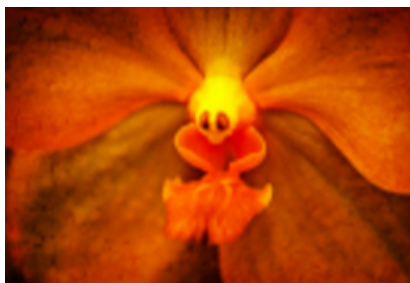
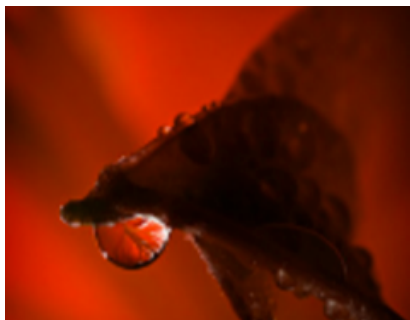
*flower
petal
closeup
water*

**GPS
groups**

...

Sharing knowledge

- One example (*i.e.* an attempt to address Q1) in the context of social media such as Flickr images



Sharing knowledge

- One example (*i.e.* an attempt to address Q1) in the context of social media such as Flickr images

