Deep learning arrives

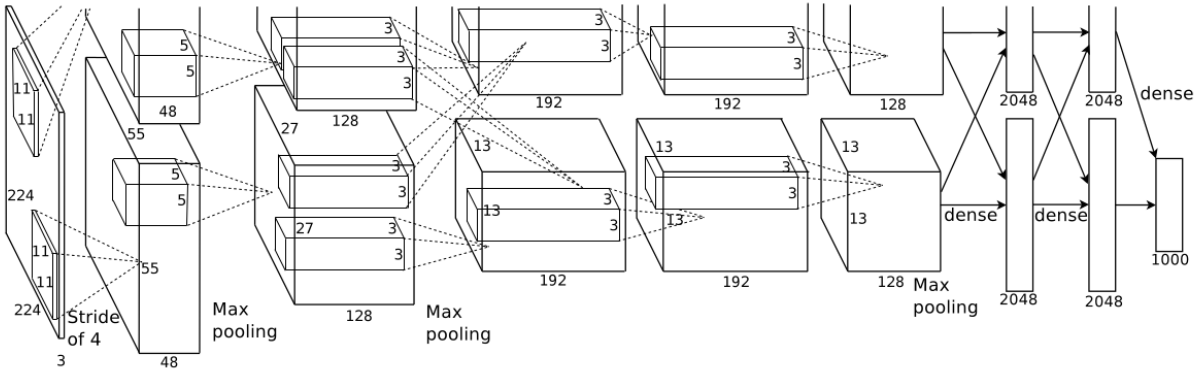


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

The thaw of the "AI winter"

- 1958 — Perceptrons, Rosenblatt
- 1960 — Adaline, Widrow and Hoff
- 1969 — Perceptrons, Minsky and Papert
- 1970 — Backpropagation, Linnainmaa
- 1974 — Backpropagation, Werbos
- 1986 — Backpropagation, Rumelhart, Hinton and Williams
- 1997 — LSTM, Hochreiter and Schmidhuber
- 1998 — OCR, LeCun, Bottou, Bengio and Haffner
- 2006 — Deep Learning, Hinton, Osindero, Teh
- 2009 — Imagenet, Deng et al.
- 2013 — Alexnet, LeCun, Bottou, Bengio and Haffner
- 2015 — Resnet (154 layers), MSRA
- toda y — GO, Deepmind

Neural Networks: A decade ago

- Lack of processing power
- Lack of data
- Overfitting
- Vanishing gradients
- Experimentally, training multi-layer perceptrons was not that useful

"Are 1-2 hidden layers the best neural networks can do?"

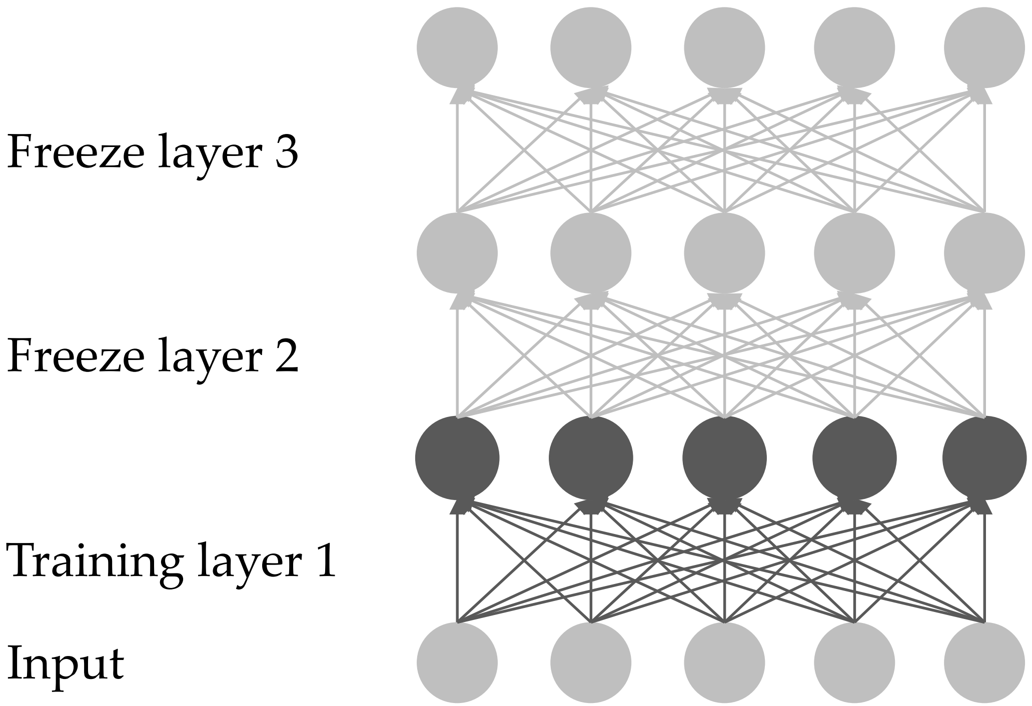
Neural Networks: Today

- ~~Lack of processing power~~
- ~~Lack of data~~
- ~~Overfitting~~
- ~~Vanishing gradients~~
- ~~Experimentally, training multi-layer perceptrons was not that useful~~

~~"Are 1-2 hidden layers the best neural networks can do?"~~

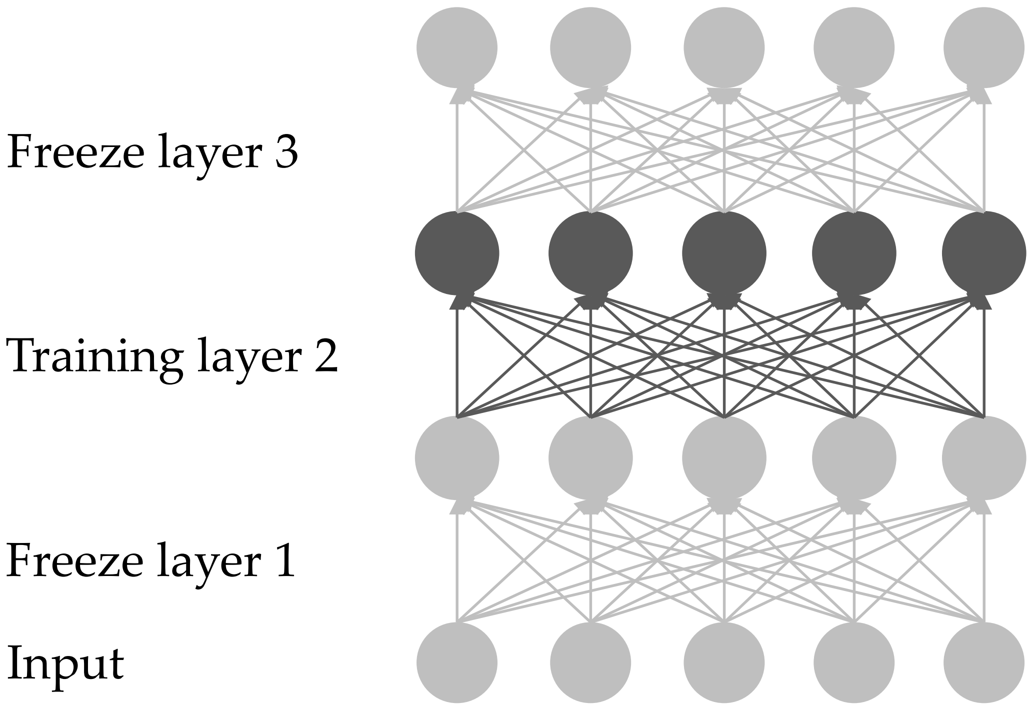
Deep Learning arrives

- Easier to train one layer at a time → Layer-by-layer training
- Training multi-layered neural networks became easier
- After, keep training with contrastive divergence



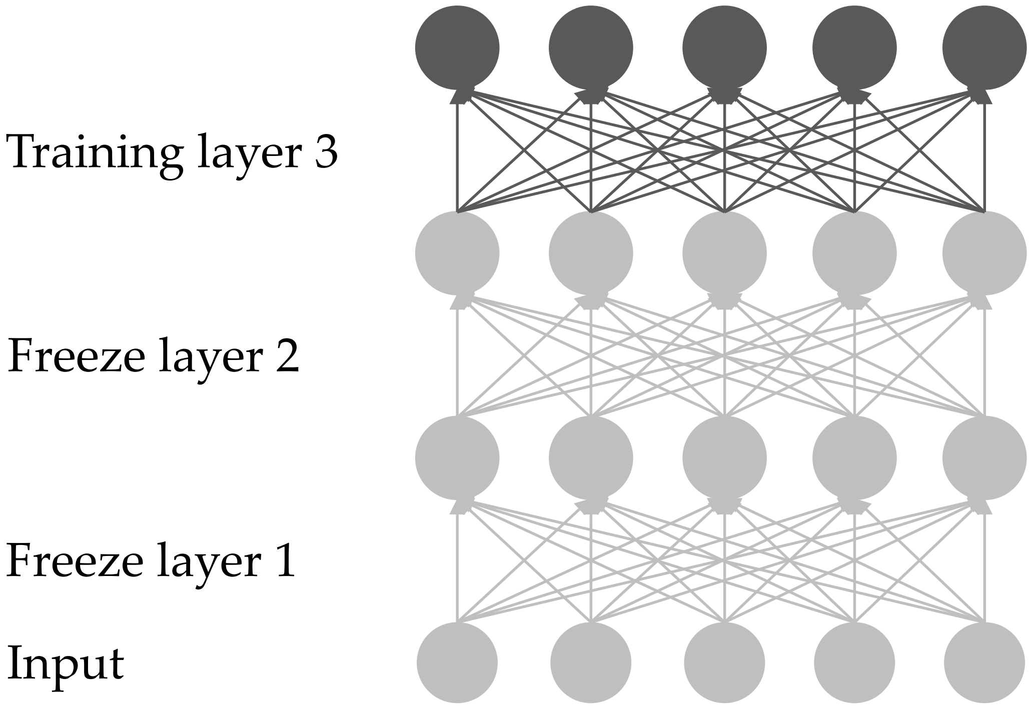
Deep Learning arrives

- Easier to train one layer at a time → Layer-by-layer training
- Training multi-layered neural networks became easier
- After, keep training with contrastive divergence



Deep Learning arrives

- Easier to train one layer at a time → Layer-by-layer training
- Training multi-layered neural networks became easier
- After, keep training with contrastive divergence



Deep Learning Renaissance

- 1958 — Perceptrons, Rosenblatt
- 1960 — Adaline, Widrow and Hoff
- 1969 — Perceptrons, Minsky and Papert
- 1970 — Backpropagation, Linnainmaa
- 1974 — Backpropagation, Werbos
- 1986 — Backpropagation, Rumelhart, Hinton and Williams
- 1997 — LSTM, Hochreiter and Schmidhuber
- 1998 — OCR, LeCun, Bottou, Bengio and Haffner
- 2006 — Deep Learning, Hinton, Osindero, Teh
- 2009 — Imagenet, Deng et al.
- 2013 — Alexnet, LeCun, Bottou, Bengio and Haffner
- 2015 — Resnet (154 layers), MSRA
- toda y — GO, Deepmind

Deep Learning is Big Data Hungry!

- In 2009 the Imagenet dataset was published [Deng et al., 2009]
  - Collected images for all 100K terms in Wordnet (16M images in total)
  - Terms organized hierarchically: "Vehicle"→"Ambulance"
- Imagenet Large Scale Visual Recognition Challenge (ILSVRC)
  - 1 million images, 1,000 classes, top-5 and top-1 error measured

CNN based, non-CNN based

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

2014 Teams	%error
GoogLeNet	6.6
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	12.1

ImageNet 2012 winner: AlexNet

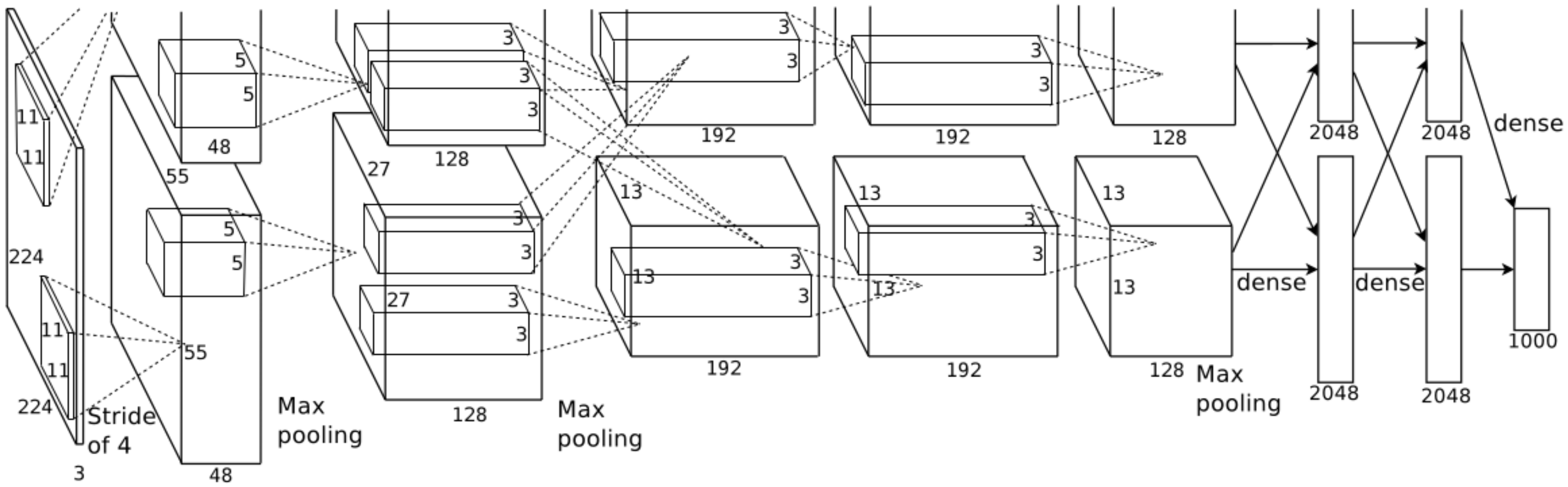
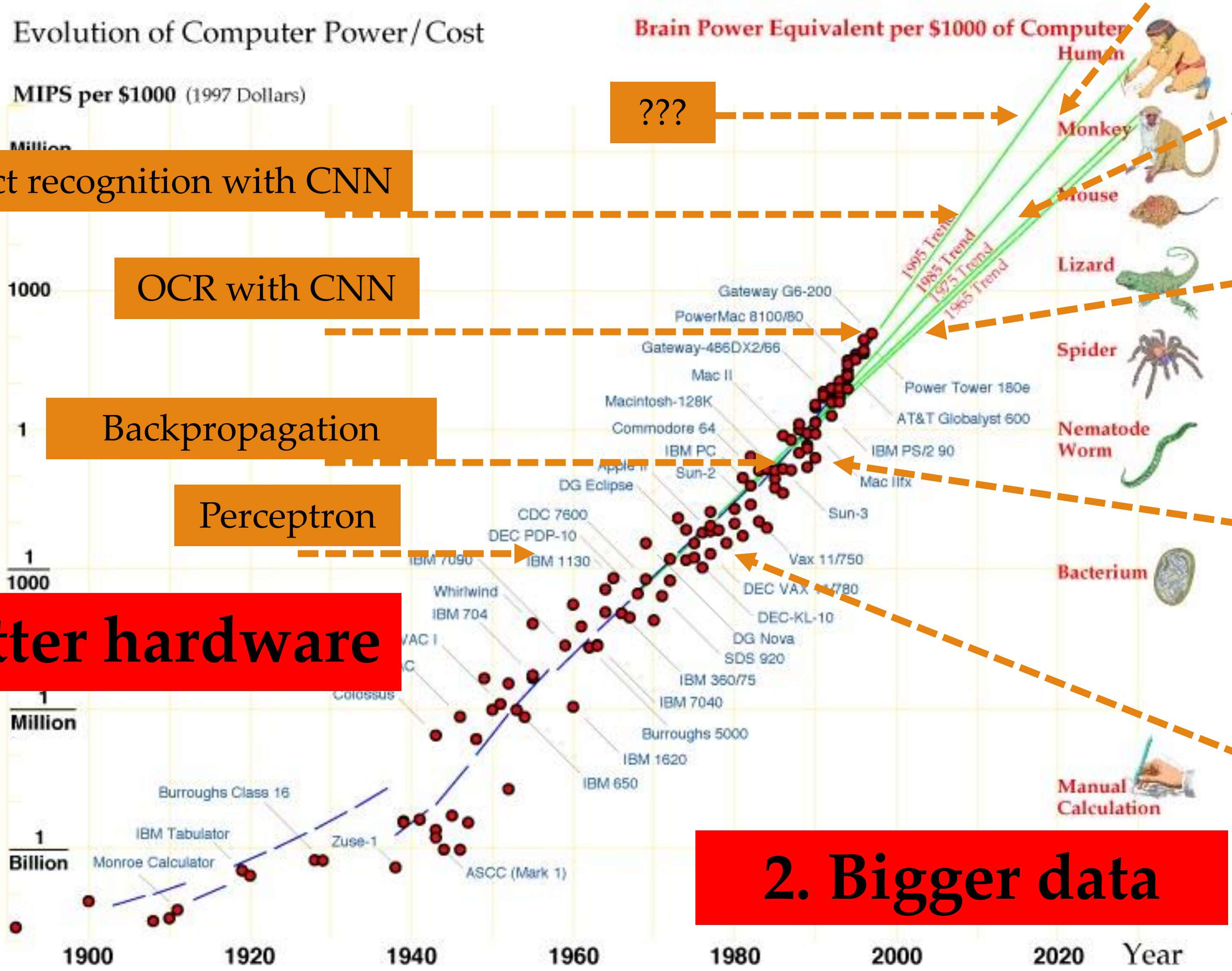


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

Krizhevsky, Sutskever & Hinton, NIPS 2012

Why now?

Datasets of everything (captions, question-answering, …), reinforcement learning, ???



???

Object recognition with CNN

OCR with CNN

Backpropagation

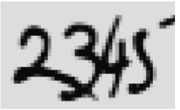
Perceptron

1. Better hardware

2. Bigger data

Imagenet: 1,000 classes from real images, 1,000,000 images

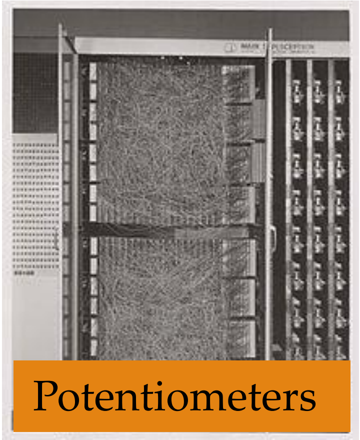
Bank cheques



Parity, negation problems

D1	D2	D3	Even-Parity
0	0	0	True
0	0	1	False
0	1	0	False
0	1	1	True
1	0	0	False
1	0	1	True
1	1	0	True
1	1	1	False

Mark I Perceptron



Potentiometers

Deep Learning Golden Era

- 1958 — Perceptrons, Rosenblatt
- 1960 — Adaline, Widrow and Hoff
- 1969 — Perceptrons, Minsky and Papert
- 1970 — Backpropagation, Linnainmaa
- 1974 — Backpropagation, Werbos
- 1986 — Backpropagation, Rumelhart, Hinton and Williams
- 1997 — LSTM, Hochreiter and Schmidhuber
- 1998 — OCR, LeCun, Bottou, Bengio and Haffner
- 2006 — Deep Learning, Hinton, Osindero, Teh
- 2009 — Imagenet, Deng et al.
- 2013 — Alexnet, LeCun, Bottou, Bengio and Haffner
- 2015 — Resnet (154 layers), MSRA
- 2017 — GO, Deepmind

Deep Learning in practice

YouTube

Large-scale Video Classification with Convolutional Neural Networks, CVPR 2014

Youtube

Microsoft Deep Learning Semantic Image Segmentation

Website

Youtube

Deep Sensorimotor Learning

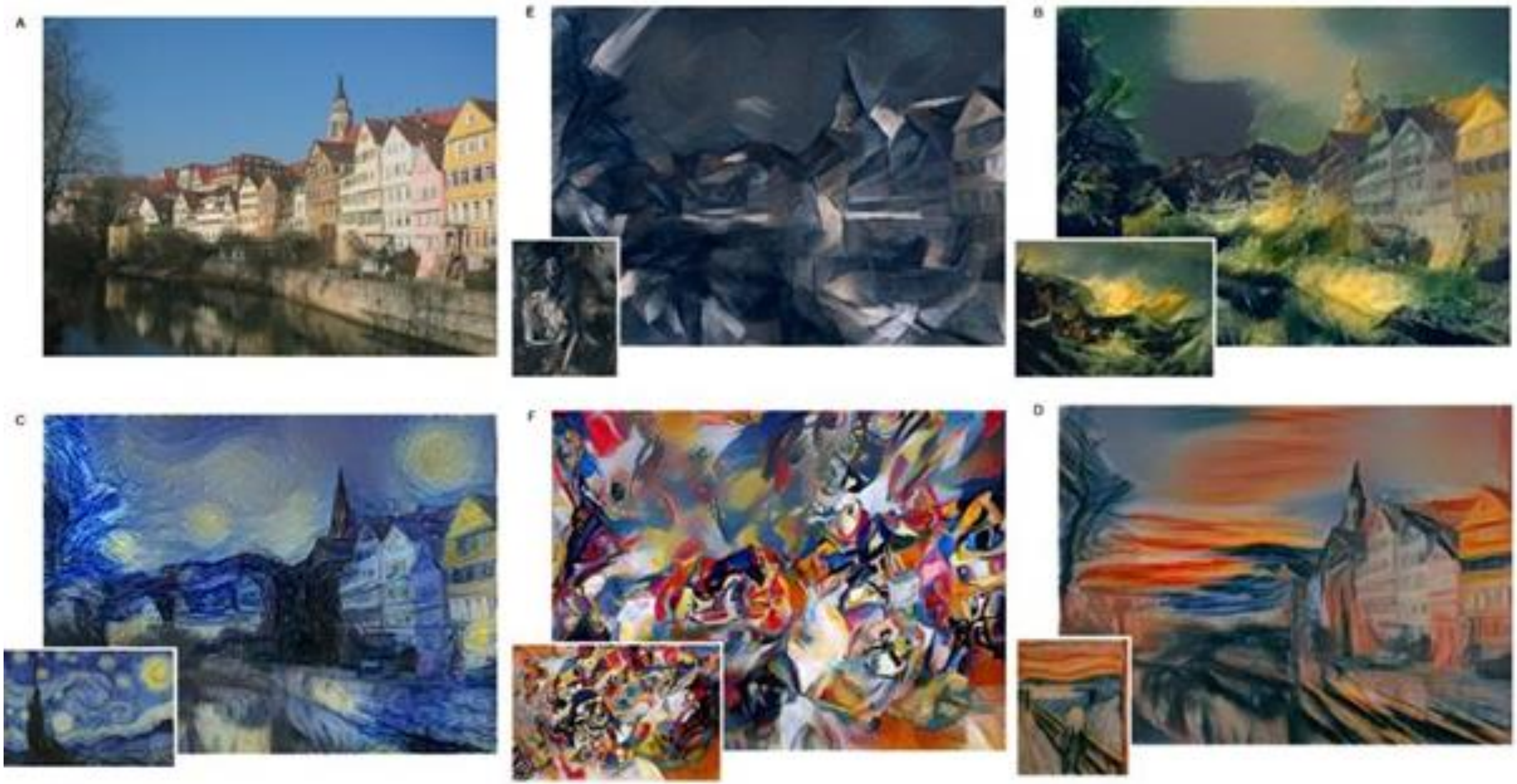
Youtube

Google DeepMind's Deep Q-learning playing Atari Breakout

Newspapers			
New York San Jose	New York Times San Jose Mercury News	Baltimore Cincinnati	Baltimore Sun Cincinnati Enquirer
NHL Teams			
Boston Phoenix	Boston Bruins Phoenix Coyotes	Montreal Nashville	Montreal Canadiens Nashville Predators
NBA Teams			
Detroit Oakland	Detroit Pistons Golden State Warriors	Toronto Memphis	Toronto Raptors Memphis Grizzlies
Airlines			
Austria Belgium	Austrian Airlines Brussels Airlines	Spain Greece	Spainair Aegean Airlines
Company executives			
Steve Ballmer Samuel J. Palmisano	Microsoft IBM	Larry Page Werner Vogels	Google Amazon

Table 2: Examples of the analogical reasoning task for phrases (the full test set has 3218 examples). The goal is to compute the fourth phrase using the first three. Our best model achieved an accuracy of 72% on this dataset.

Deep Learning even for the arts



Why should we be impressed?

- Vision is ultra challenging!
  - For 256x256 resolution → $2^{524,288}$ of possible in
  - Large semantic & visual object variations

Intra-class overlap

Inter-class variation

- Robotics is typically considered in controlled environments
- Game AI involves extreme number of possible games states ($10^{10^{48}}$ possible GO games)
- NLP is extremely high dimensional and vague (just for English: 150K words)
- Deep learning seems to casually solve many, many (supervised) problems thought to be extremely hard