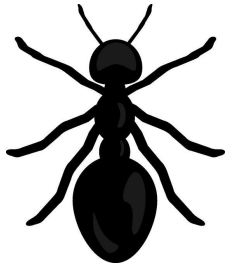
Will Deep Learning Lead to AI?

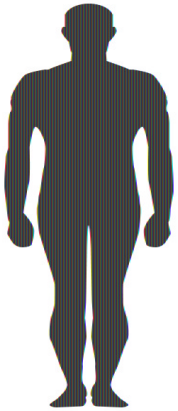
Haytham Fayek

ML / AI Melbourne Meetup - August 2017

What is Artificial Intelligence?

The ability to perceive information, and to retain it as knowledge to be applied towards adaptive behaviors within an environment or context.

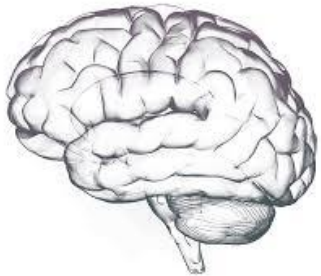




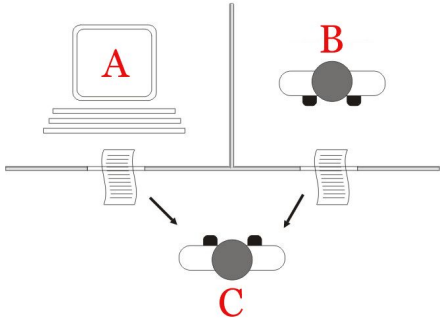
What is Intelligence?

Work In Progress

Haytham Fayek, *Will Deep Learning Lead to AI?*



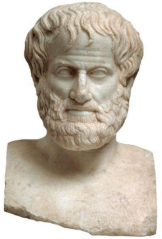
Think Like Human



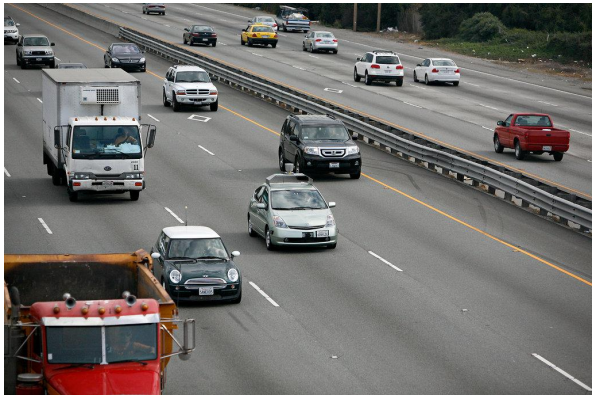
Behave Like Human

What is Artificial Intelligence?

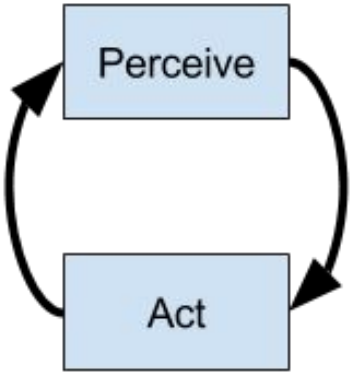
Think Rationally



Act Rationally



Haytham Fayek, *Will Deep Learning Lead to AI?*
[*] Images from various online sources



Narrow AI
vs
Strong AI

AI

Learning

Meta-Learning

Perception

Attention

Memory

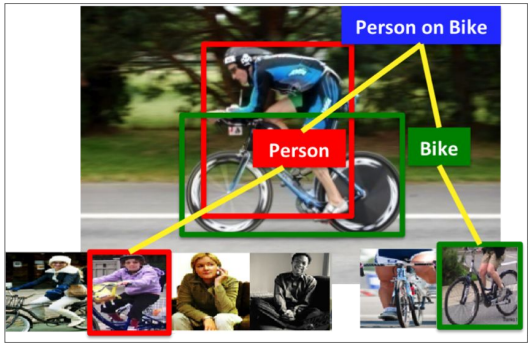
Reasoning

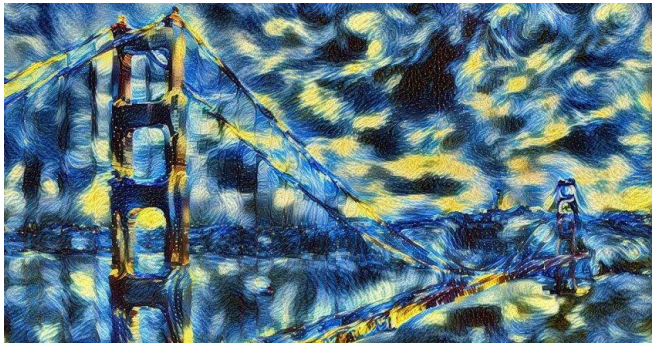
Planning

Emotion

Communication

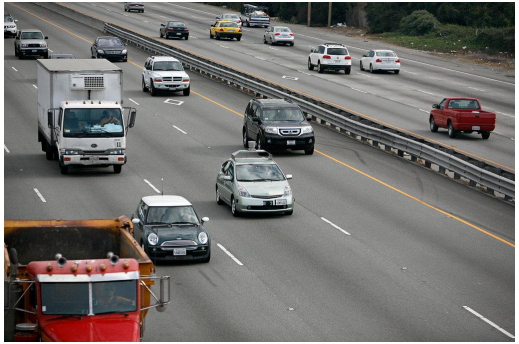
Haytham Fayek, *Will Deep Learning Lead to AI?*





What is Deep Learning?

The hierarchical learning of feature representations



Haytham Fayek, *Will Deep Learning Lead to AI?*
[*] Images from various online sources

Notation

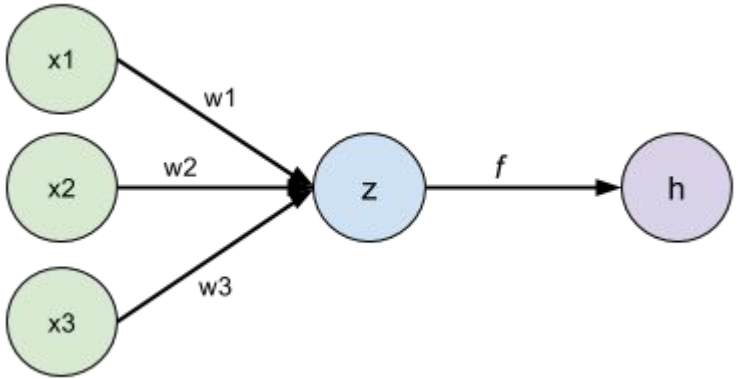
Notation

Linear Operation:

$$z = x_1 w_1 + x_2 w_2 + x_3 w_3$$

Non-Linear Operation:

$$h = f(z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases}$$



Notation

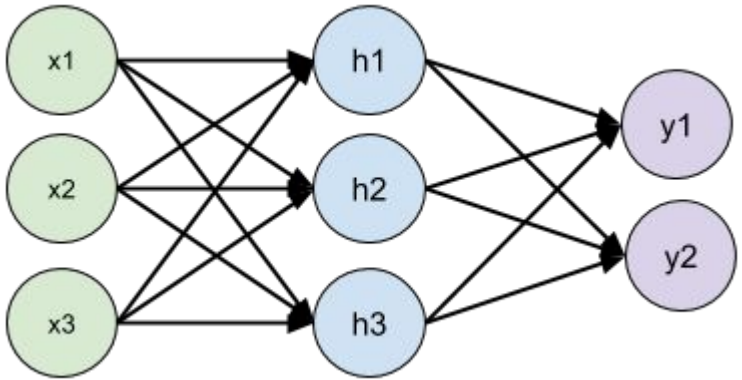
$$h_1 = f(x_1 w_{1,1} + x_2 w_{1,2} + x_3 w_{1,3})$$

$$h_2 = f(x_1 w_{2,1} + x_2 w_{2,2} + x_3 w_{2,3})$$

$$h_3 = f(x_1 w_{3,1} + x_2 w_{3,2} + x_3 w_{3,3})$$

A Fully-Connected Layer: $\mathbf{h} = f(\mathbf{x}\mathbf{W})$

A Convolutional Layer: $\mathbf{h} = f(\mathbf{x} * \mathbf{W})$

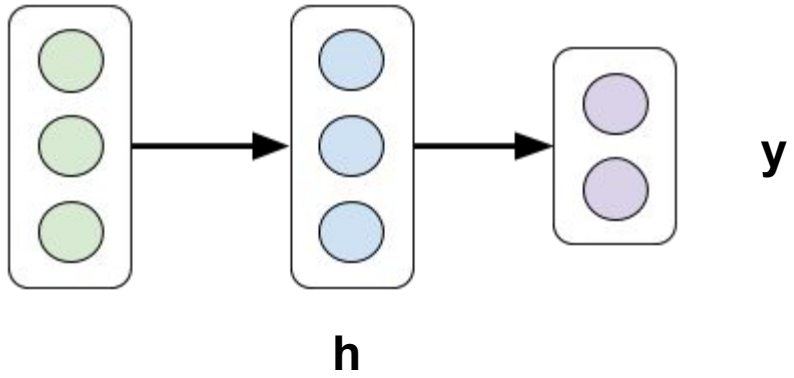


Notation

$$\mathbf{z} = \mathbf{x}\mathbf{W}^{(1)}$$

$$\mathbf{h} = f(\mathbf{z}) = \begin{cases} \mathbf{z} & \mathbf{z} \geq \mathbf{0} \\ \mathbf{0} & \mathbf{z} < \mathbf{0} \end{cases}$$

$$\hat{\mathbf{y}} = \mathbf{h}\mathbf{W}^{(2)}$$



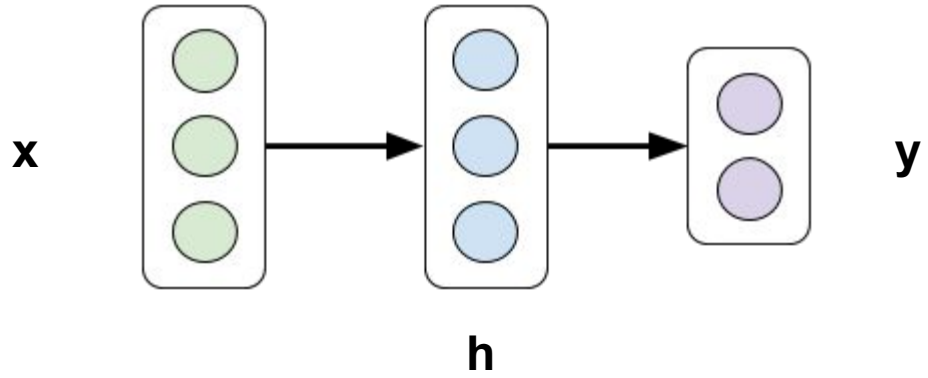
$$\hat{\mathbf{y}} = \mathcal{F}(\mathbf{x}; \mathbf{W})$$

Notation

$$\mathcal{D} = \{\mathbf{x}^{(i)}, y^{(i)}\}$$

$$\hat{\mathbf{y}} = \mathcal{F}(\mathbf{x}; \mathbf{W})$$

$$\mathcal{L}_i(\mathbf{W}) = \frac{1}{2}(y^{(i)} - \hat{y}^{(i)})^2$$

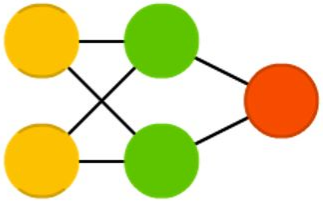


Find ***W*** that minimizes *L*:

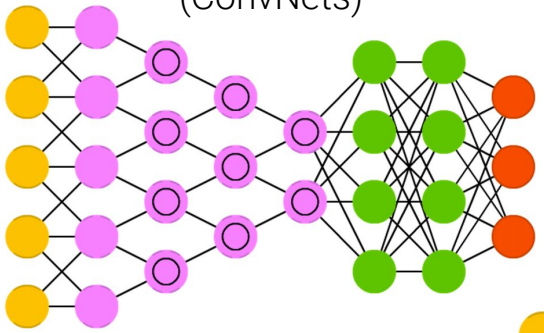
$$w_i \leftarrow w_i + \alpha \nabla \mathcal{L}_i(\mathbf{W}, \mathbf{x}^{(i)}, y^{(i)})$$

Notation

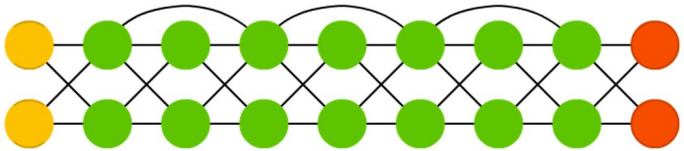
Feedforward Neural Network
(DNN)



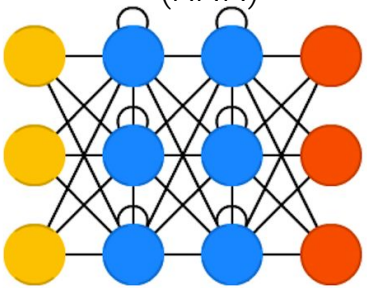
Convolutional Neural Network
(ConvNets)



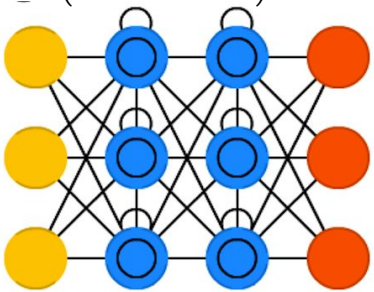
Residual Network
(ResNets)



Recurrent Neural Network
(RNN)



Long Short-term Memory
(LSTM-RNN)



[*] http://www.asimovinstitute.org/neural-network-zoo/

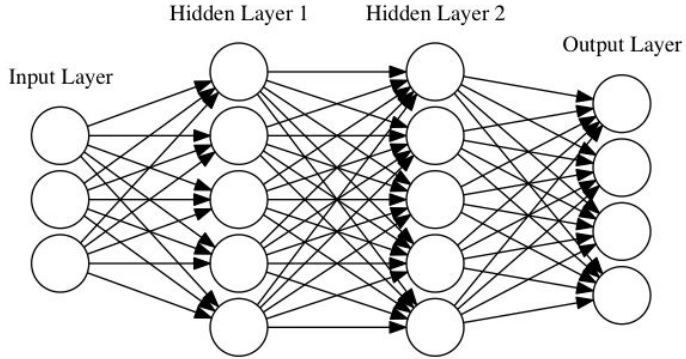
Haytham Fayek, *Will Deep Learning Lead to AI?*

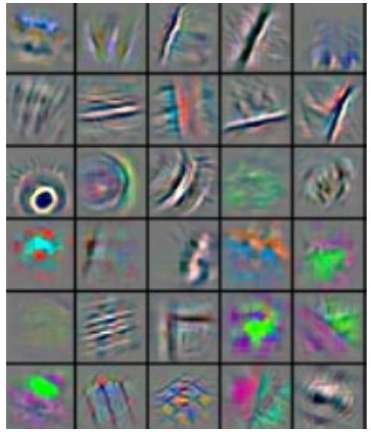
Learning

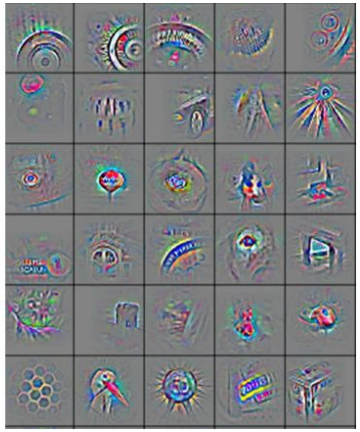
Learning

Supervised Learning

Pick a model, loss function & training scheme.
Train a large model on lots of labeled data.



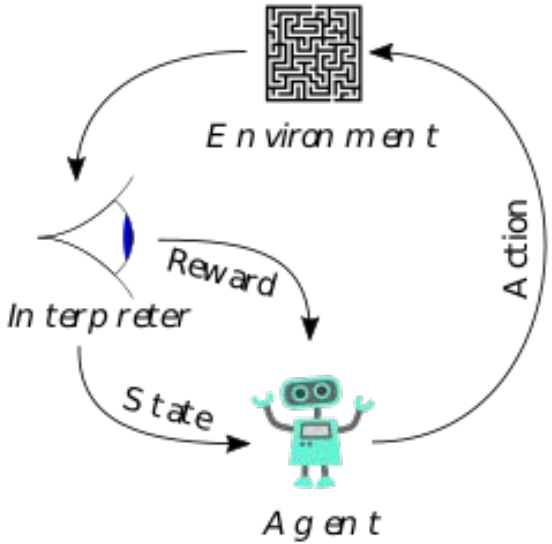




[1] Zeiler & Fergus, Visualizing and Understanding Convolutional Neural Networks, 2013

Learning

Reinforcement Learning



$States\ \mathbf{S}$

$Actions\ \mathbf{A}$

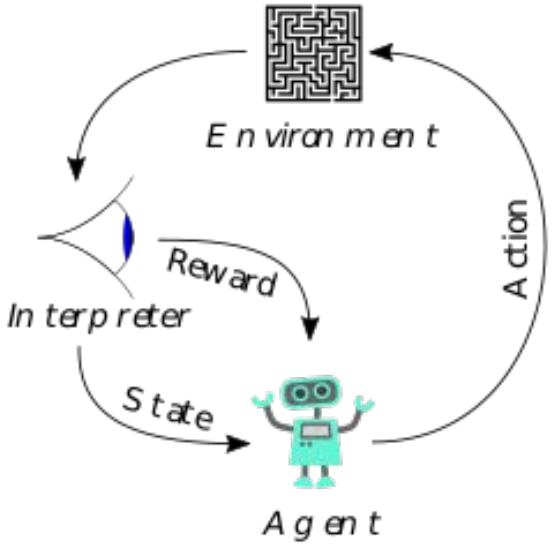
$$P_a(s, s') = Pr(S_{t+1} = s' | s_t = s, a_t = a)$$

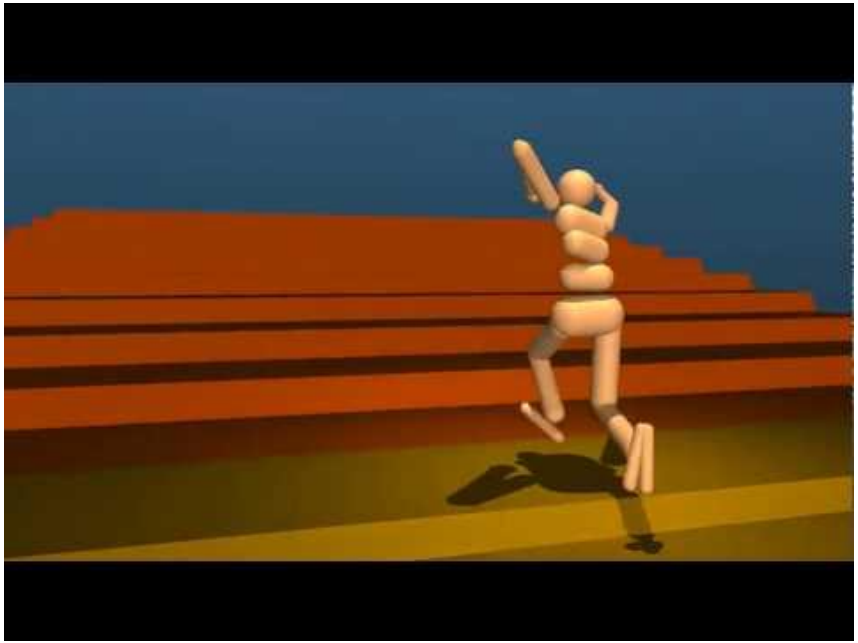
$$Reward\ R_a(s, s')$$

$$\pi(a|s) = P(a_t = a | s_t = s)$$

Learning

Reinforcement Learning





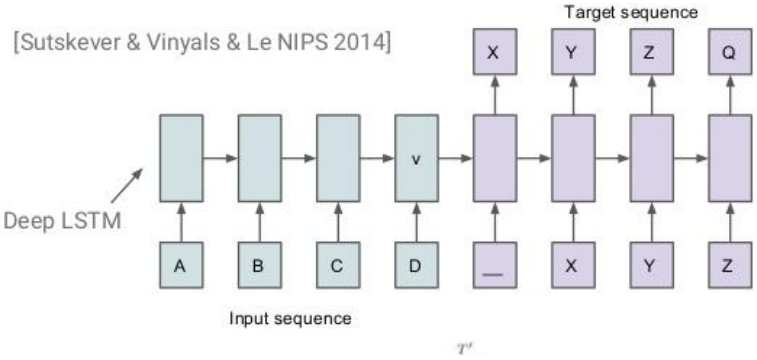
[2] Heess et al., Emergence of Locomotion Behaviours in Rich Environments, 2017

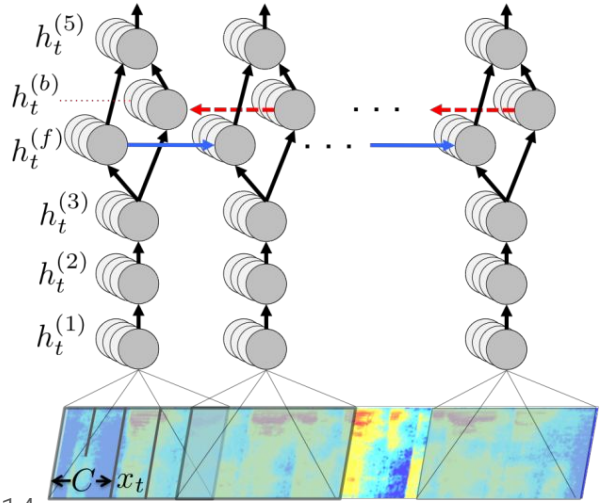
Perception

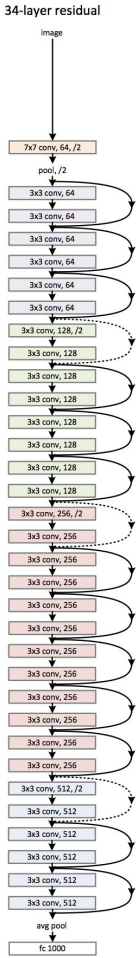
Perception

Supervised Learning has made strides in Perception:
Speech Recognition, Computer Vision, Natural Language Processing.

Each modality requires a fair amount of specialized engineering.







Haytham Fayek, *Will Deep Learning Lead to AI?*
[*] Images from respective papers: He et al. 2015, Hannun et al. 2014

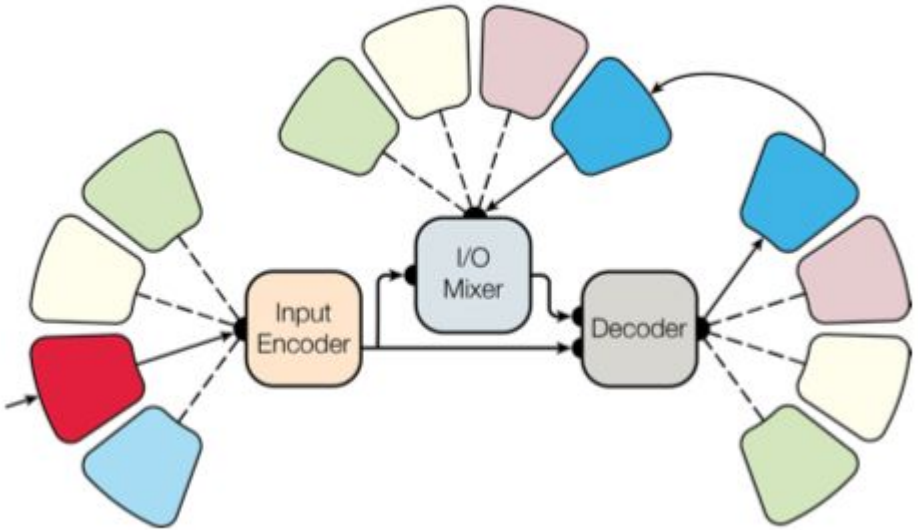
Perception

A single model with good results on multiple tasks in multiple domains.

Tasks: Object Recognition, translation, image captioning, speech recognition, parsing, ..

Components:

- Convolutional Layers
- Attention Mechanism
- Mixture of Experts



[3] Kaiser et al., One Model to Learn Them All, 2017

Perception

Problem	MultiModel (joint 8-problem)	State of the art
ImageNet (top-5 accuracy)	86%	95%
WMT EN → DE (BLEU)	21.2	26.0
WMT EN → FR (BLEU)	30.5	40.5

Table 1: Comparing MultiModel to state-of-the-art from [28] and [21].

Problem	Joint 8-problem		Single problem	
	log(perpexity)	accuracy	log(perplexity)	accuracy
ImageNet	1.7	66%	1.6	67%
WMT EN→DE	1.4	72%	1.4	71%
WSJ speech	4.4	41%	5.7	23%
Parsing	0.15	98%	0.2	97%

Table 2: Comparison of the MultiModel trained jointly on 8 tasks and separately on each task.

[4] Kaiser et al., One Model to Learn Them All, 2017

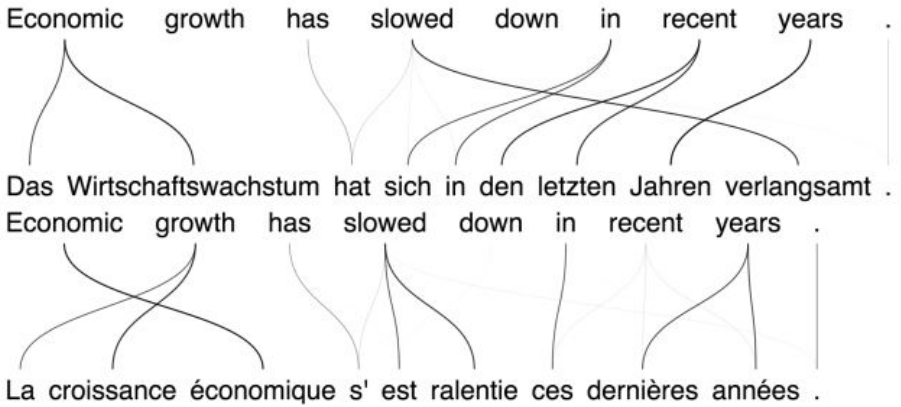
Attention

Attention

Allows selective processing of input to avoid clutter and simplify computation.

Allows processing of complex input in nonlinear ways.

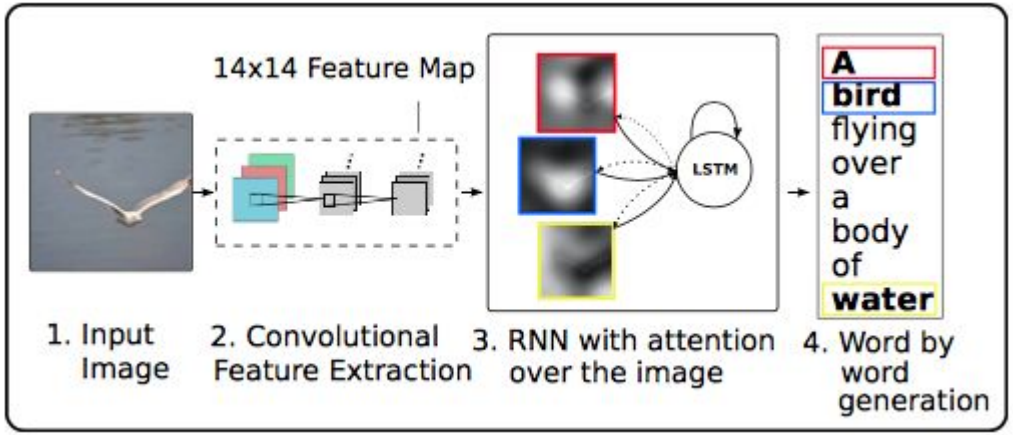
Attention can make our models more interpretable.



[5] Cho et al., Describing Multimedia Content using Attention Based Encoder Decoder Networks, 2015

Attention

Image Caption Generation



[6] Xu et al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, 2015

Memory

Memory

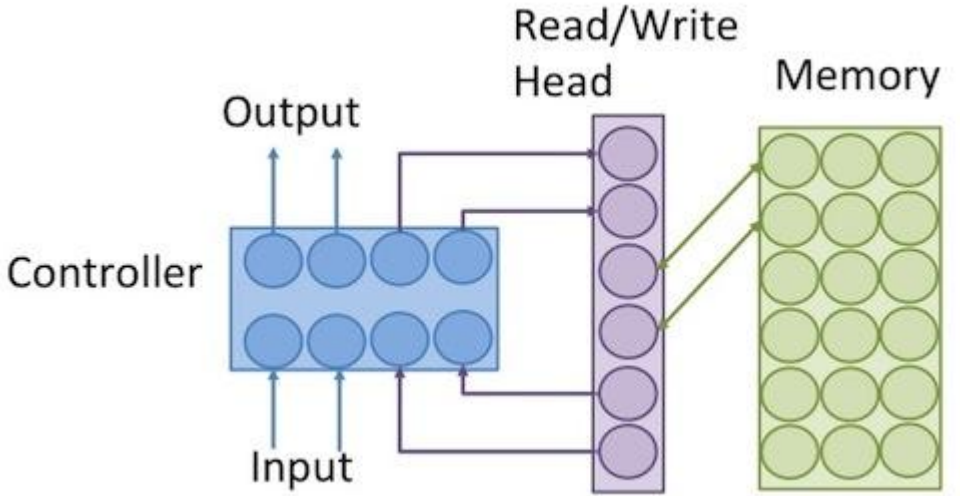
Neural Turing Machine

An LSTM Controller

Read / Write Heads using Attention (NN)

An External Memory

Can learning simple algorithms such as copying & sorting



[7] Graves et al., Neural Turing Machine, 2015

Reasoning

Reasoning

Process very unstructured complex (multi-modal) input

Deliberate and infer relations

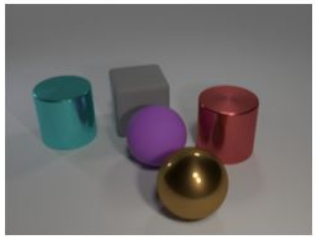
Store Knowledge

Output unstructured output

How many chairs are at the table?

Is there a pedestrian in my lane?

Is the person with the blue hat touching the bike in the back?

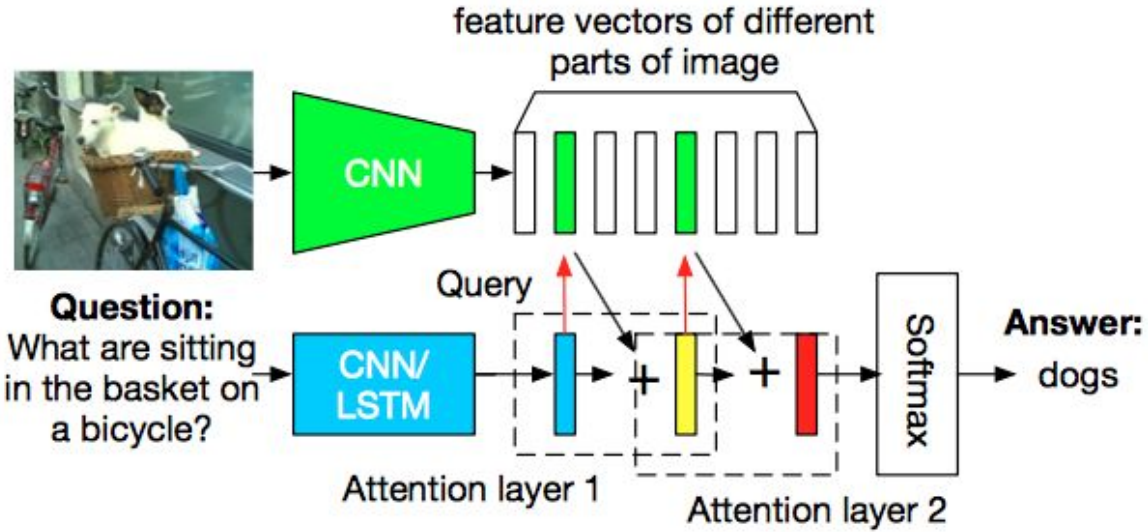


Is there a matte cube that has the same size as the red metal object?

[8] Johnson et al., Learning and Executing Programs for Visual Reasoning , 2017

Reasoning

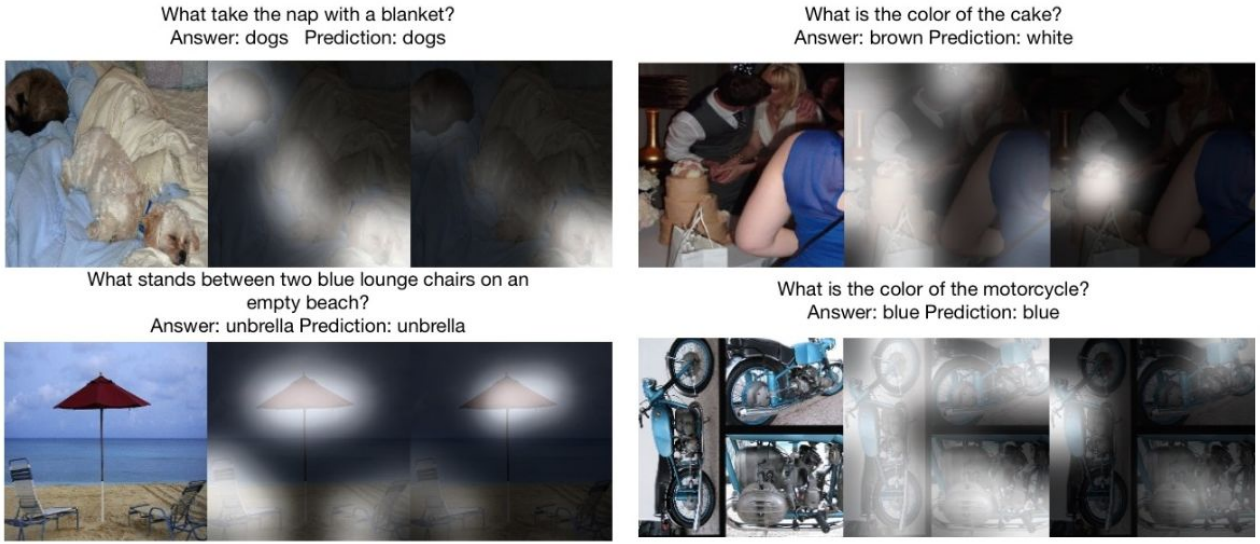
Visual Question Answering (VQA)



[9] Yang et al., Stacked Attention for Image Question Answering, 2015

Reasoning

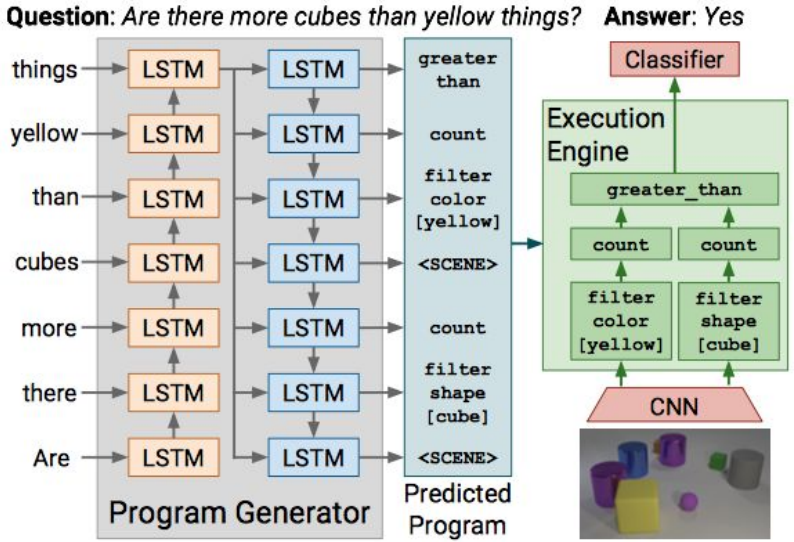
Visual Question Answering (VQA)



[9] Yang et al., Stacked Attention for Image Question Answering, 2015

Reasoning

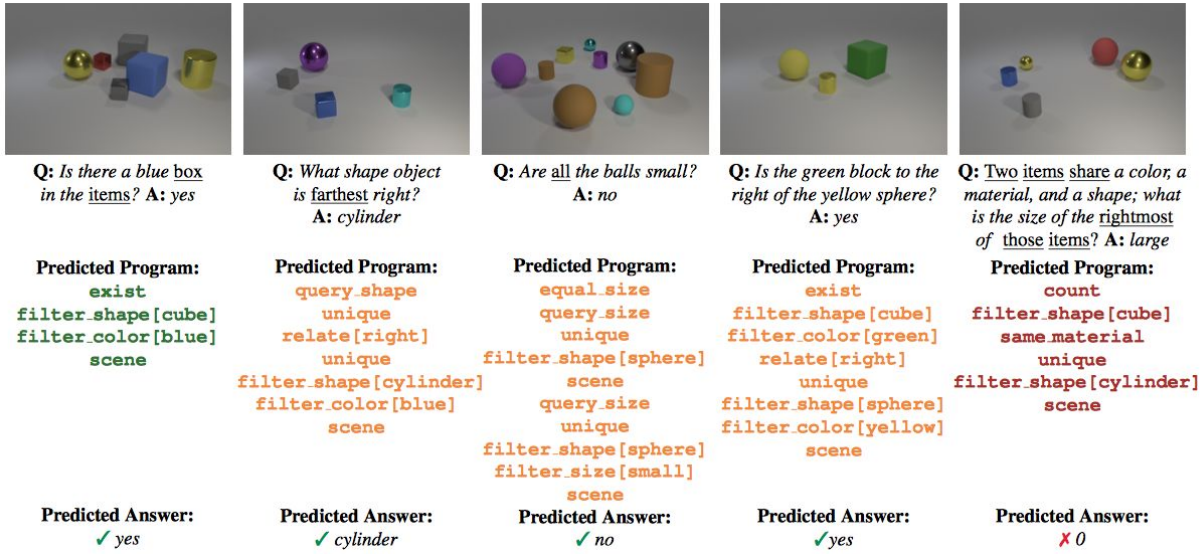
Visual Question Answering (VQA)



[8] Johnson et al., Learning and Executing Programs for Visual Reasoning , 2017

Reasoning

Visual Question Answering (VQA)



[8] Johnson et al., Learning and Executing Programs for Visual Reasoning , 2017

Meta-Learning

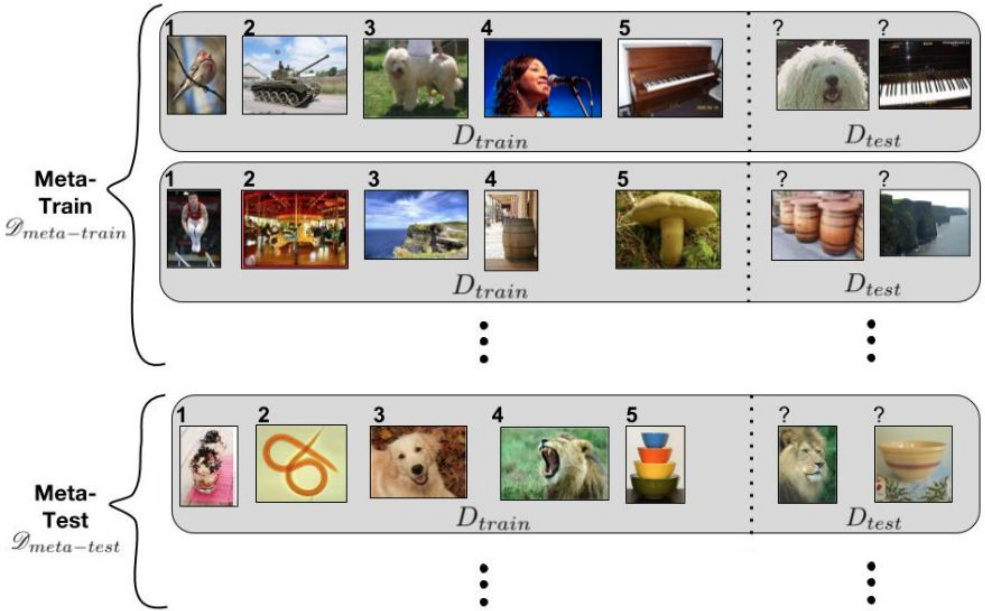
Meta-Learning

Learning to Learn

[*] Ravi et al., Optimization as a Model for Few-Shot Learning, 2017

Meta-Learning

Learning to Learn

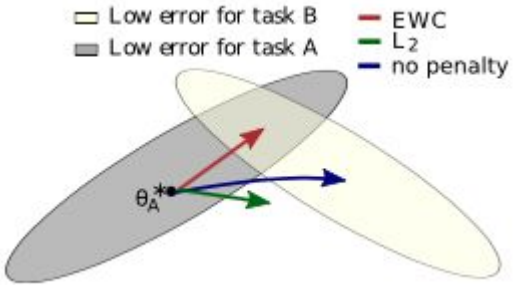


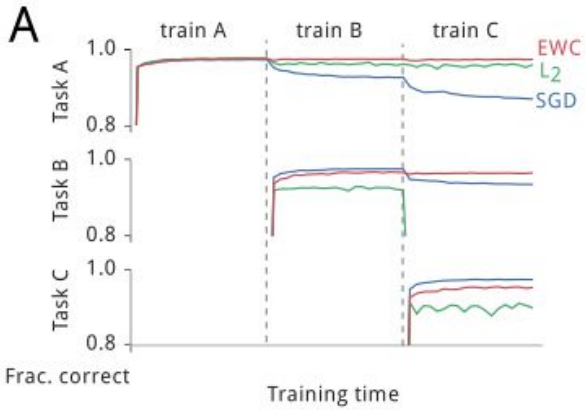
[*] Ravi et al., Optimization as a Model for Few-Shot Learning, 2017

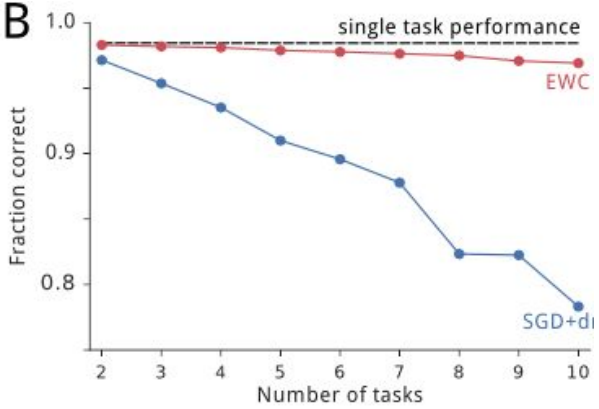
Meta-Learning

Learning to Learn

$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \sum_i \frac{\lambda}{2} F_i(\theta_i - \theta^*_{A,i})^2$$







[10] Kirkpatrick et al., Overcoming catastrophic forgetting in neural networks, 2017

Will Deep Learning Lead to AI?

Remarks

Supervised learning works!

Current models require sophisticated engineering and big data.

Unsupervised learning and reinforcement learning remain challenging paradigms.

Lots of challenging questions remain to be answered or even asked!

Thank you

Will Deep Learning Lead to AI?

@haythamfayek