

RL for DL

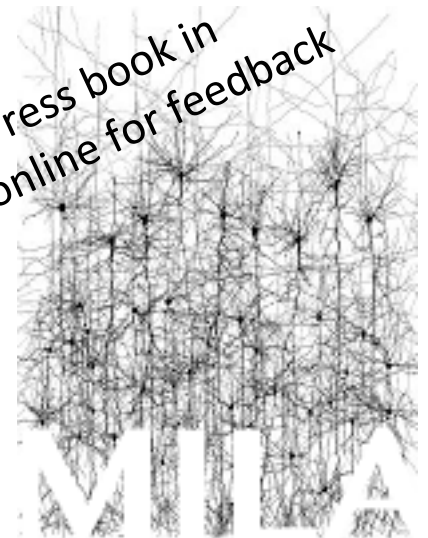
Yoshua Bengio

December 11, 2015

NIPS'2015 Deep Reinforcement Learning Workshop

Université 
de Montréal

PLUG: **Deep Learning**, MIT Press book in
preparation, draft chapters online for feedback



Deep Learning: Beyond Pattern Recognition, towards AI

- Many researchers believed that neural nets could at best be good at pattern recognition
- And they are really good at it!
- But many more ingredients needed towards AI. Recent progress:
 - ATTENTION & REASONING:
 - Machine translation, Memory networks & Neural Turing Machine
 - PLANNING & REINFORCEMENT LEARNING:
 - DeepMind (Atari game playing) & Berkeley (Robotic control)

How to train neural nets to take internal discrete decisions?

- Can we approximate the gradient through discrete (possibly stochastic decisions) so as to extend the reach of back-prop?
- Usage:
 - Conditional computation / dynamically routed architectures
 - Attention (internal and external)
 - Alignment, (hierarchical) segmentation, etc.
 - Long-term dependencies through many nonlinearities: almost not differentiable
- Simple solution: REINFORCE (+ conditional baseline)

$$\frac{\partial}{\partial \theta} \sum_a p_{\theta}(a|x) L(a) = \sum_a p_{\theta}(a|x) L(a) \frac{\partial \log p_{\theta}(a|x)}{\partial \theta}$$

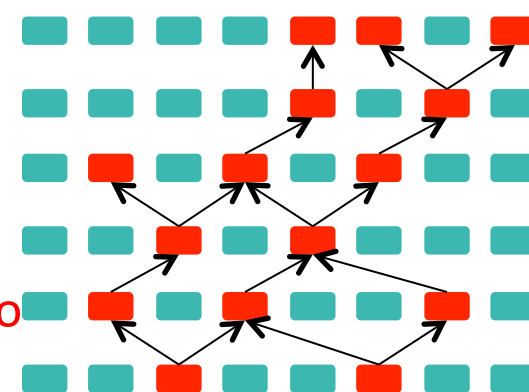
Conditional Computation

Computation / Capacity Ratio

- N-grams, decision trees, etc.: capacity (and memory) can grow a lot while computation remains constant or grows as $\log(\text{capacity})$.
- Neural nets / deep learning: computation grows linearly with capacity (number of parameters). Each parameter is used for every example.
- To build much higher capacity models, we need to break that linear relationship.

Conditional Computation: only visit a small fraction of parameters / example


*Bengio, Leonard & Courville
arXiv 1305.2982*

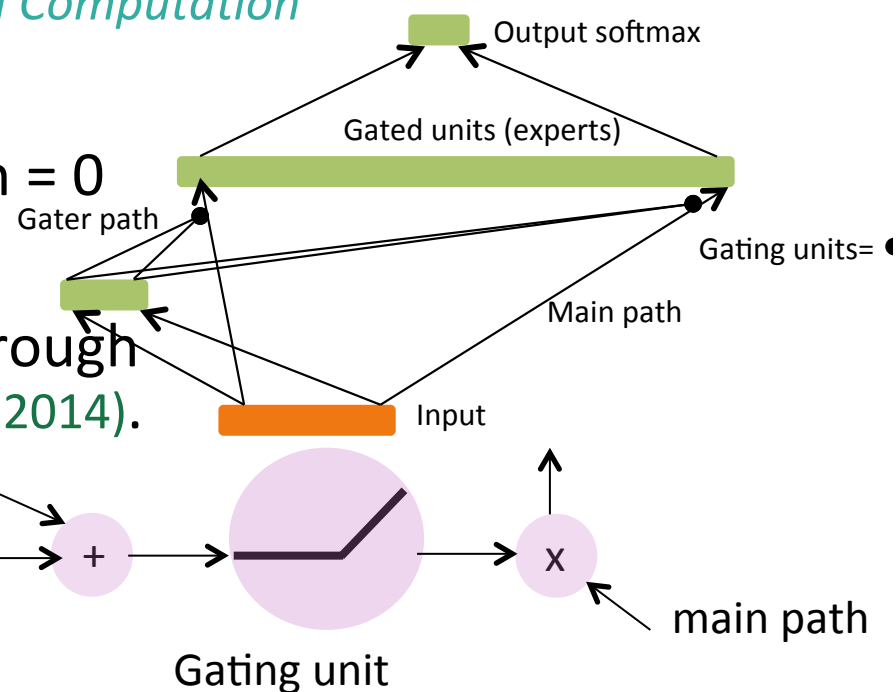


- Deep nets vs decision trees
- Hard mixtures of experts (Collobert, Bengio & Bengio 2002)
- Conditional computation for deep nets: sparse distributed gates selecting combinatorial subsets of a deep net
- Challenges:
 - Credit assignment for hard decisions
 - Gated architectures exploration

Credit Assignment for Discrete Actions

(Bengio, Leonard, Courville 2013): *Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation*

- Gating units take a hard decision
- Gradient through discrete function = 0
- Solution ideas in (Bengio et al 2013):
 - Heuristic back-prop (straight through estimator), also (Gregor et al ICML 2014).
 - Noisy rectifier: 
 - Smooth times Stochastic bvp with $b \sim \text{Bin}(\nu p)$
 - REINFORCE with variance reduction i.e., RL, i.e. correlate with loss, no back-prop for gaters



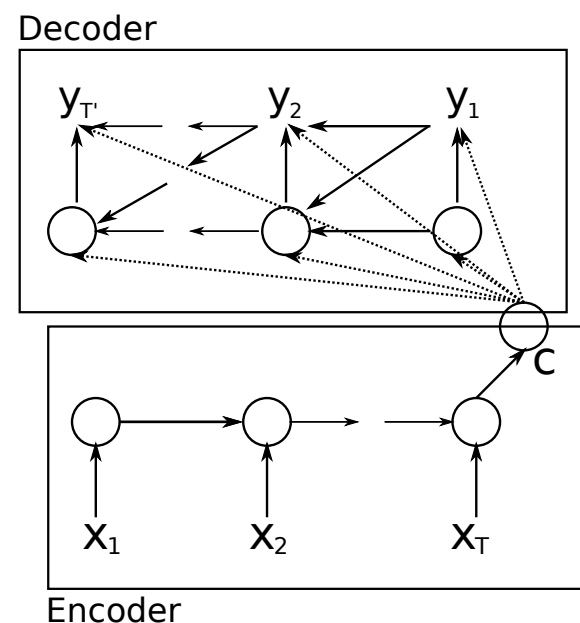
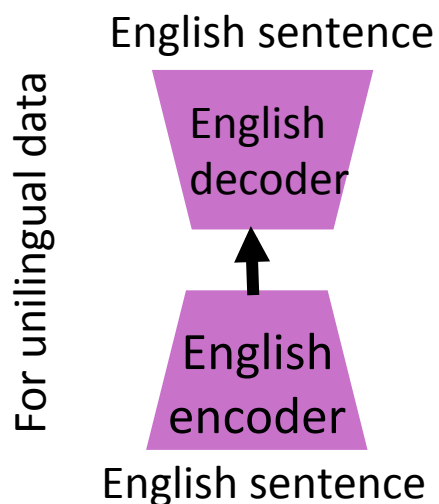
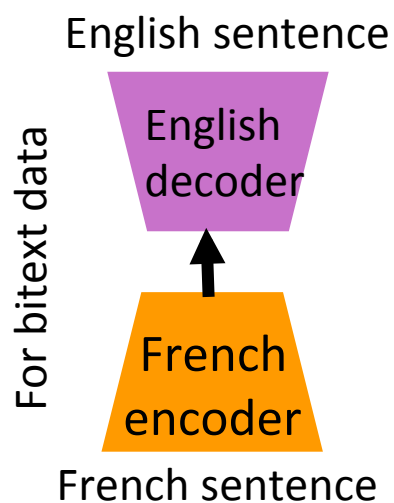
New work: *K-Y Cho & Y Bengio, arXiv 2015*

E. Bengio, P.L. Bacon, J. Pineau & D. Precup arXiv 2015 & RLDM 2015.

Attention for MT, caption generation
and Reasoning

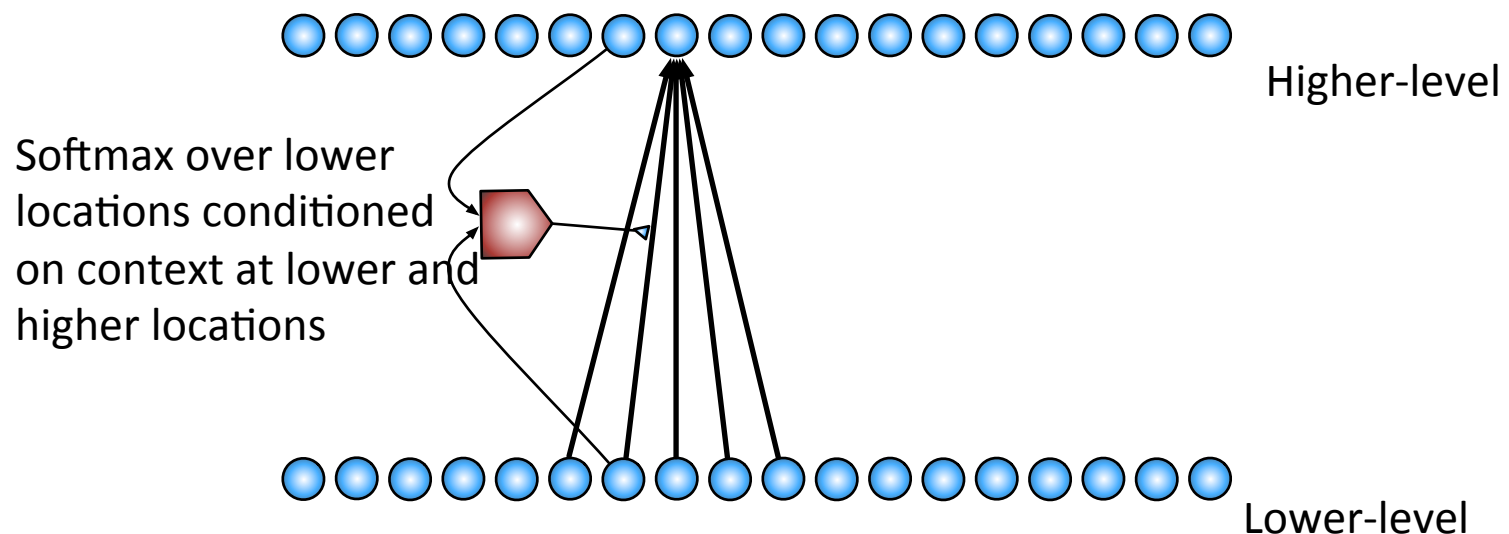
Encoder-Decoder Framework

- Intermediate representation of meaning
= 'universal representation'
- Encoder: from word sequence to sentence representation
- Decoder: from representation to word sequence distribution



Attention Mechanism for Deep Learning

- Consider an input (or intermediate) sequence or image
- Consider an upper level representation, which can choose « where to look », by assigning a weight or probability to each input position, as produced by an MLP, applied at each position



Soft-Attention vs Stochastic Hard-Attention

- With soft-attention: input fed to higher level at location i is a softmax-weighted sum of states at locations j at lower level
 - Train by back-prop
 - Fast training
- With stochastic hard-attention: sample an input location according to the softmax output
 - Get a gradient on the decisions via REINFORCE - baseline
 - Noisy gradient, slower training but **works surprisingly well**
 - *Symmetry breaking*

End-to-End Machine Translation with Recurrent Nets and Attention Mechanism

- Reached the state-of-the-art in one year, from scratch

(a) English→French (WMT-14)

	NMT(A)	Google	P-SMT
NMT	32.68	30.6*	37.03°
+Cand	33.28	—	
+UNK	33.99	32.7°	
+Ens	36.71	36.9°	

(b) English→German (WMT-15)

Model	Note
24.8	Neural MT
24.0	U.Edinburgh, Syntactic SMT
23.6	LIMSI/KIT
22.8	U.Edinburgh, Phrase SMT
22.7	KIT, Phrase SMT

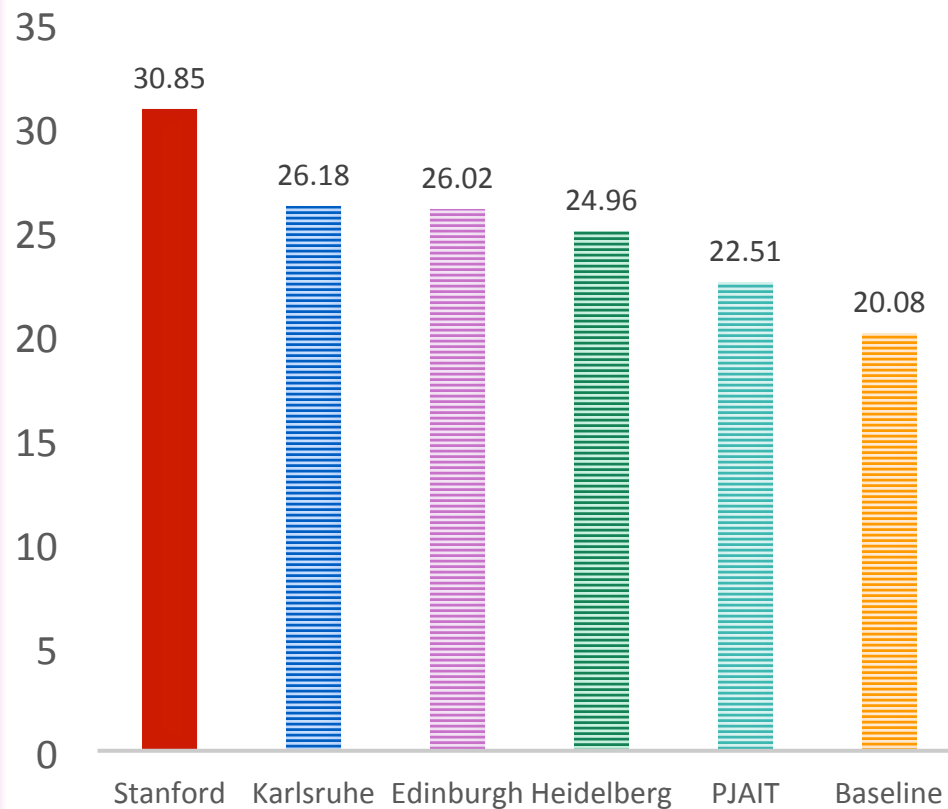
(c) English→Czech (WMT-15)

Model	Note
18.3	Neural MT
18.2	JHU, SMT+LM+OSM+Sparse
17.6	CU, Phrase SMT
17.4	U.Edinburgh, Phrase SMT
16.1	U.Edinburgh, Syntactic SMT

IWSLT 2015 - Luong & Manning (2015) TED talk MT, English-German



BLEU (CASED)



HTER (HE SET)

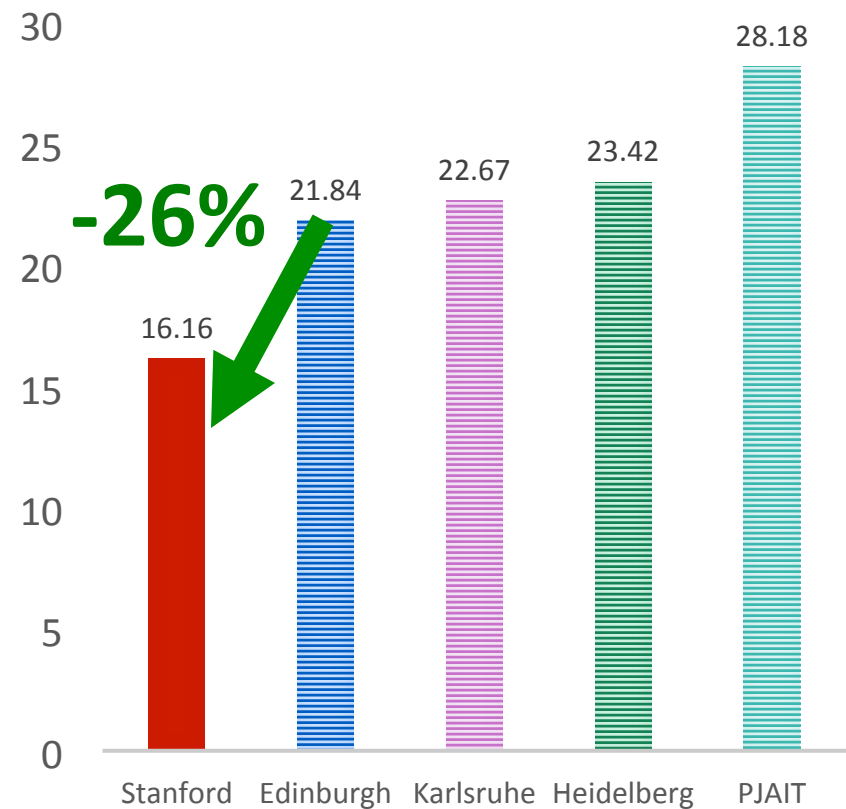
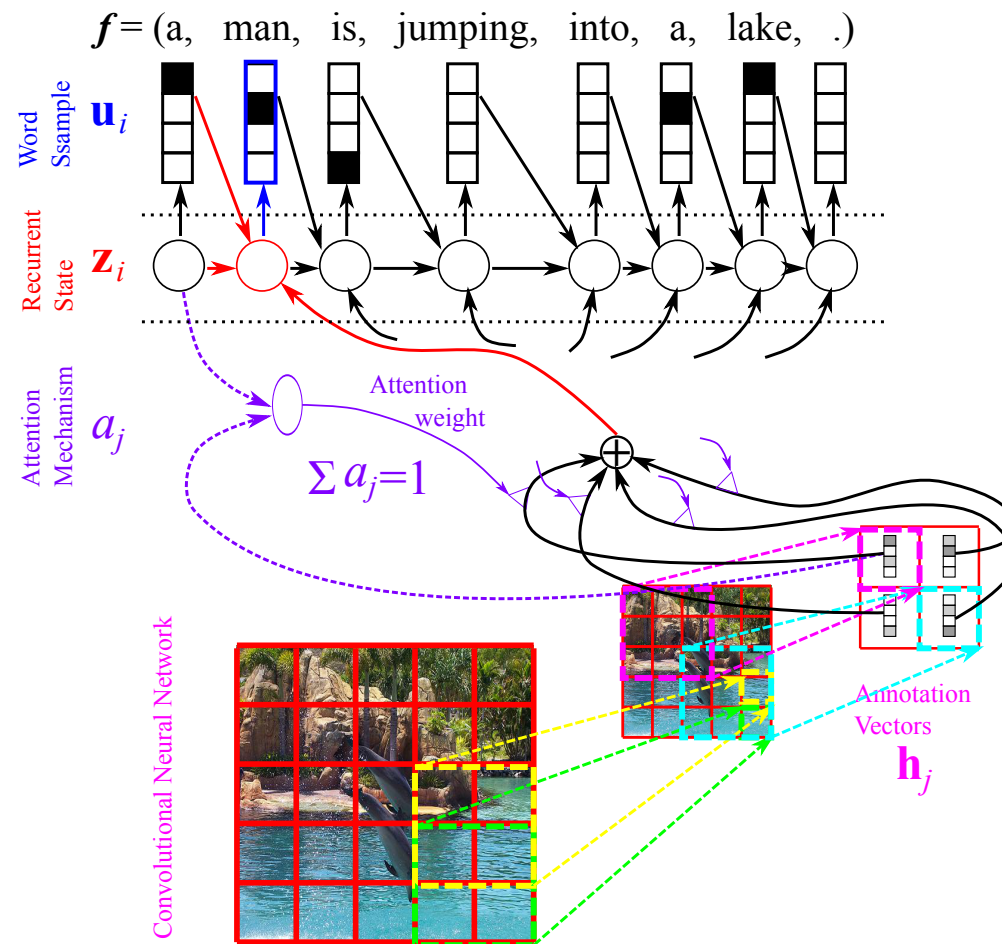
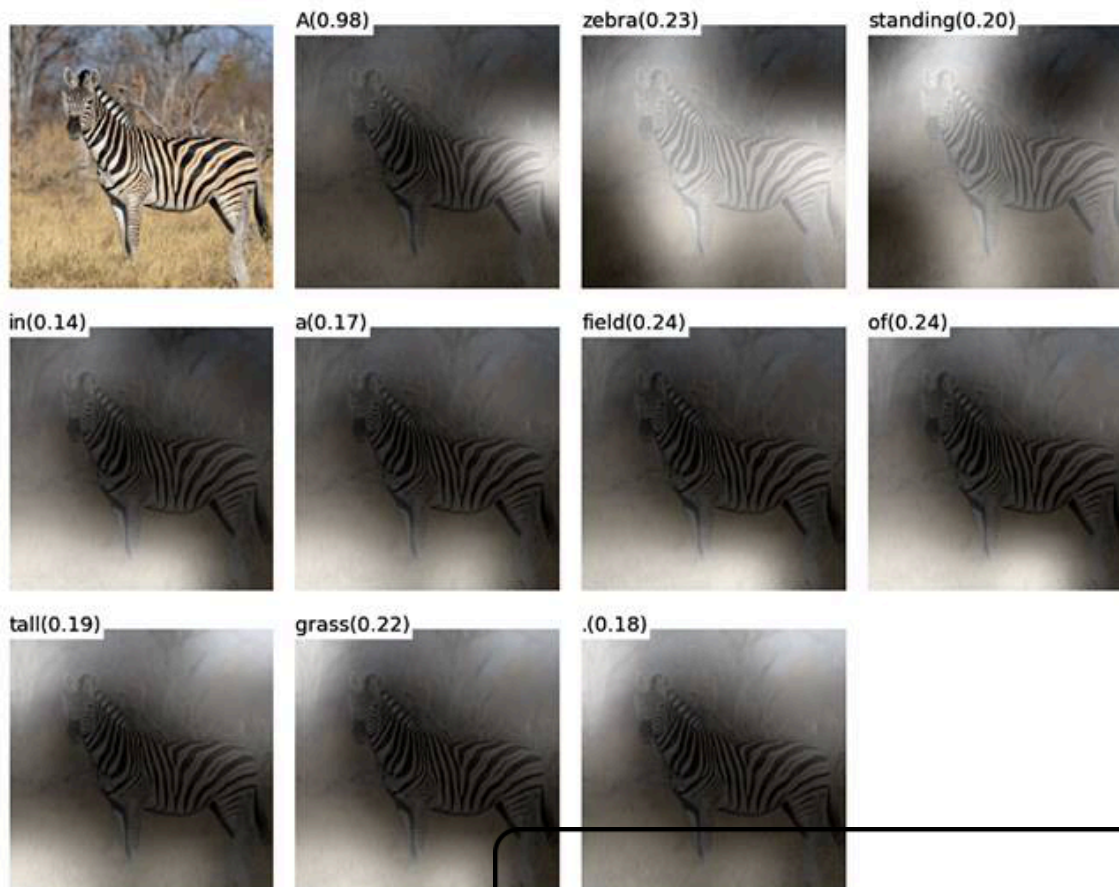


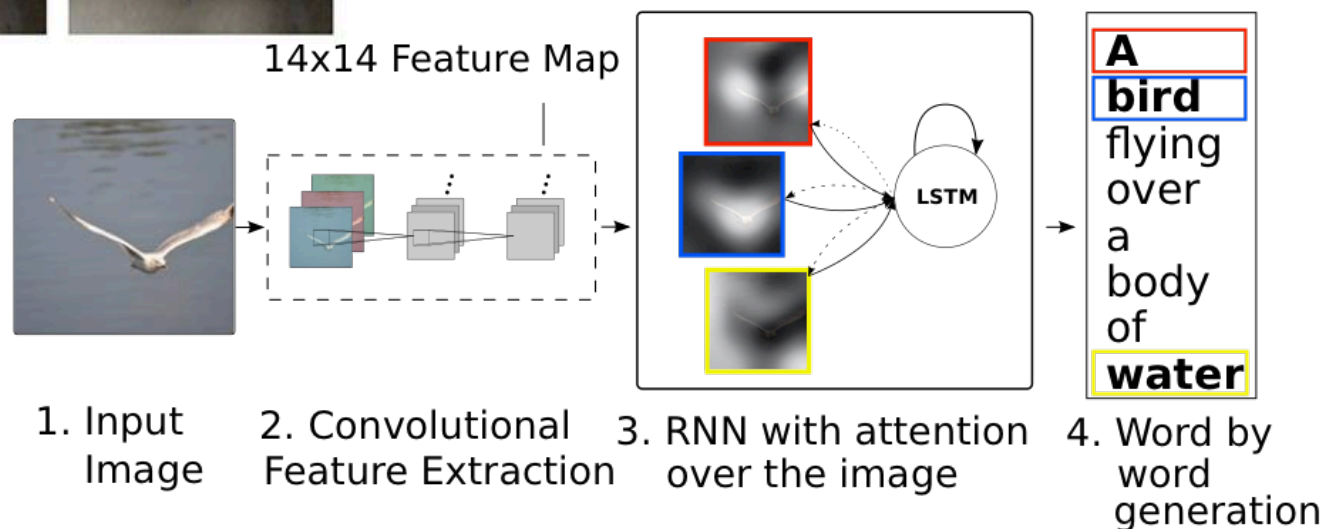
Image-to-Text: Caption Generation with Attention



(Xu et al., 2015), (Yao et al., 2015)



Paying
Attention to
Selected Parts
of the Image
While Uttering
Words



Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

Results from (Xu et al, ICML 2015)

Table 1. BLEU-1,2,3,4/METEOR metrics compared to other methods, † indicates a different split, (—) indicates an unknown metric, ° indicates the authors kindly provided missing metrics by personal communication, Σ indicates an ensemble, a indicates using AlexNet

Dataset	Model	BLEU				METEOR
		B-1	B-2	B-3	B-4	
Flickr8k	Google NIC(Vinyals et al., 2014) ^{†Σ}	63	41	27	—	—
	Log Bilinear (Kiros et al., 2014a) [°]	65.6	42.4	27.7	17.7	17.31
	Soft-Attention	67	44.8	29.9	19.5	18.93
	Hard-Attention	67	45.7	31.4	21.3	20.30
Flickr30k	Google NIC ^{†$\circ\Sigma$}	66.3	42.3	27.7	18.3	—
	Log Bilinear	60.0	38	25.4	17.1	16.88
	Soft-Attention	66.7	43.4	28.8	19.1	18.49
	Hard-Attention	66.9	43.9	29.6	19.9	18.46
COCO	CMU/MS Research (Chen & Zitnick, 2014) ^a	—	—	—	—	20.41
	MS Research (Fang et al., 2014) ^{†a}	—	—	—	—	20.71
	BRNN (Karpathy & Li, 2014) [°]	64.2	45.1	30.4	20.3	—
	Google NIC ^{†$\circ\Sigma$}	66.6	46.1	32.9	24.6	—
	Log Bilinear [°]	70.8	48.9	34.4	24.3	20.03
	Soft-Attention	70.7	49.2	34.4	24.3	23.90
	Hard-Attention	71.8	50.4	35.7	25.0	23.04

The Good



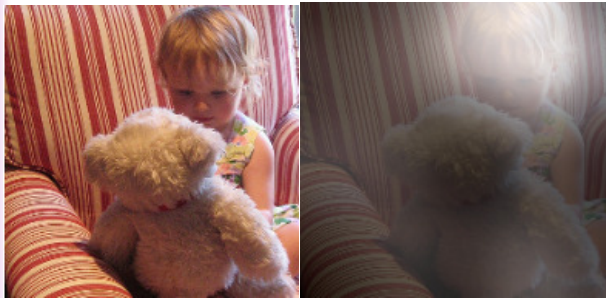
A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.



A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.



A giraffe standing in a forest with trees in the background.

And the Bad



A large white bird standing in a forest.



A woman holding a clock in her hand.



A man wearing a hat and
a hat on a skateboard.



A person is standing on a beach
with a surfboard.



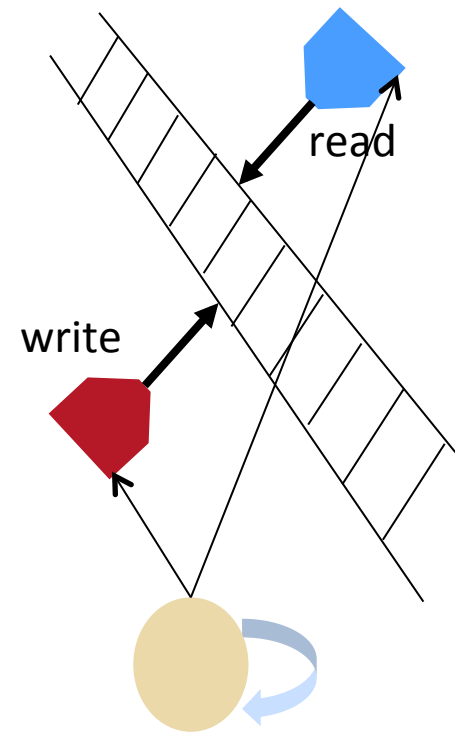
A woman is sitting at a table
with a large pizza.



A man is talking on his cell phone
while another man watches.

Attention Mechanisms for Memory Access

- Neural Turing Machines (Graves et al 2014)
- and Memory Networks (Weston et al 2014)
- Use a form of attention mechanism to control the read and write access into a memory
- The attention mechanism outputs a softmax over memory locations
- For efficiency, the softmax should be sparse (mostly 0's), e.g. maybe using a hash-table formulation.
- Both soft and (stochastic) hard attention are used



Training a Critic

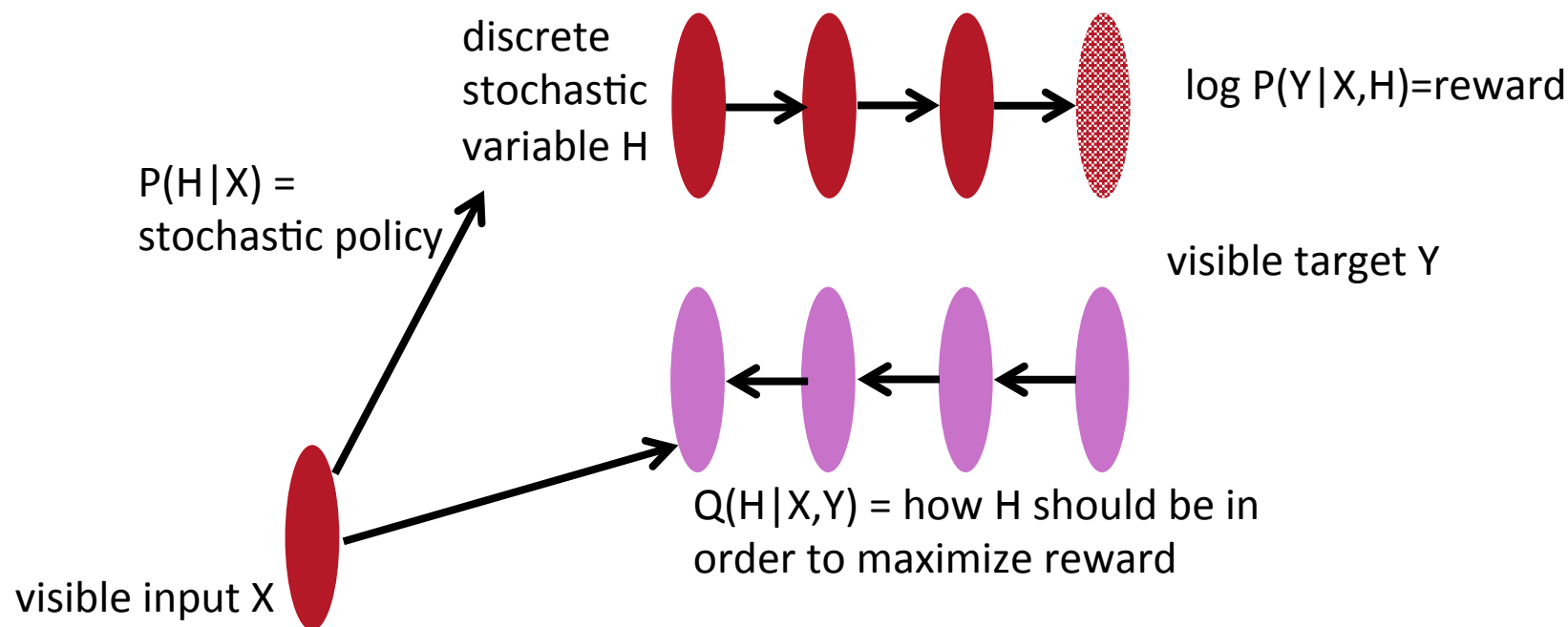
Alternatives to REINFORCE?

- Train a critic (estimate future reward) and backprop through it
 - *“Generative Adversarial Networks”, Goodfellow et al, NIPS 2014*
 - *“Task Loss Estimation for Sequence Prediction”, Bahdanau et al, 2015, arXiv 1511.06456*
- NN estimates loss that would be obtained for any discrete choice of sequence of actions
- At test-time, use beam-search (planning) to find approximately optimal seq. of actions
- Train a stochastic credit-assignment machine, e.g., by **Reweighted Wake-Sleep** (Bornschein & Bengio 2014), or **Variational Auto-Encoder** (Kingma et al 2014) which amounts to learning to predict the credit to attribute, given the future reward (observed outputs), in the form of a posterior probability distribution for the discrete latent variable

(Conditional) Reweighted Wake-Sleep

(Bornschein & Bengio ICLR 2015)

- If H is continuous, can use a conditional VAE framework
- Otherwise a conditional reweighted wake-sleep
- See also Ba et al NIPS'2015 for an application of similar idea to a recurrent attention model



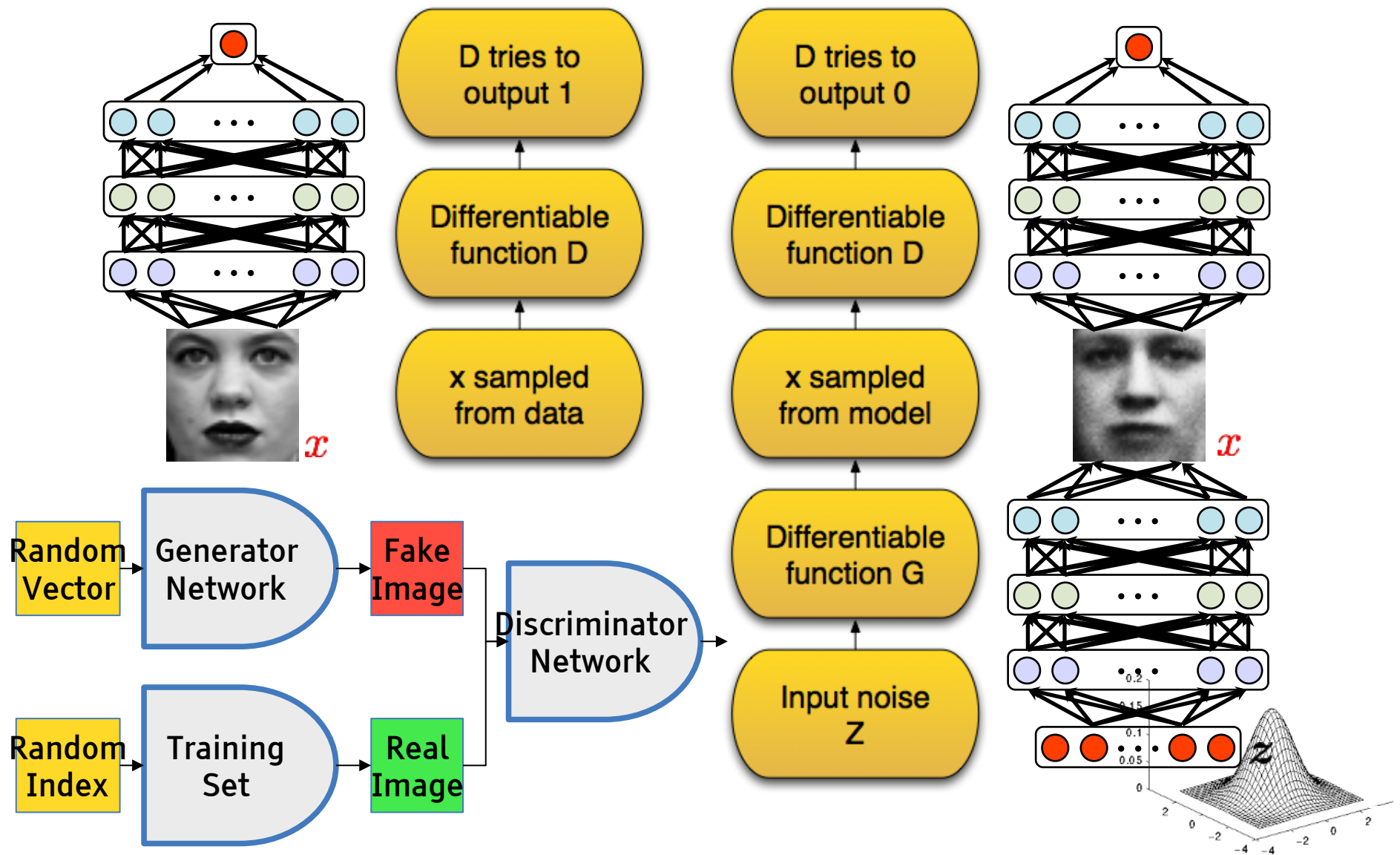
Training a Critic for Generative Models

Maximizing the Probability of Passing a Turing Test for Generative Models

- Generate sample y (possibly given x) OR get (y, x) from data generating distribution
- Ask human if answer y | x comes from data generating distribution (good) or from the computer (bad)
- If human cannot statistically distinguish the two distributions, then the computer passes that Turing test
- Can we train a critic that predicts the human answer?

GAN: Generative Adversarial Networks

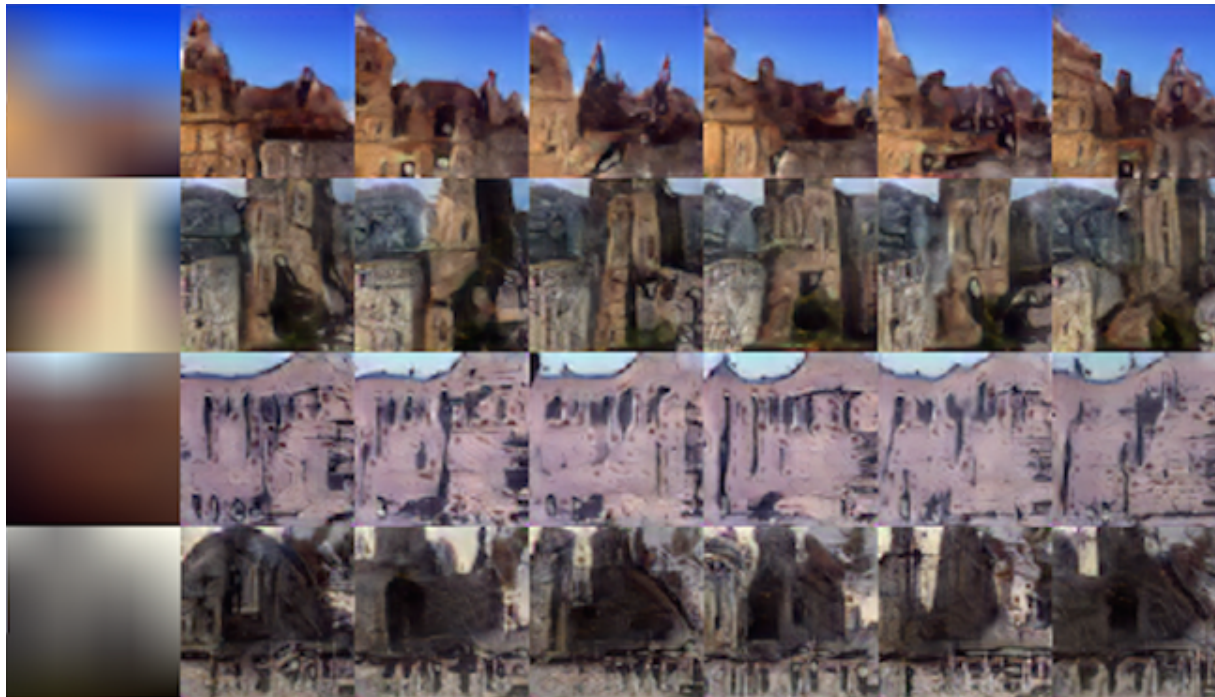
Goodfellow et al NIPS 2014



LAPGAN: Visual Turing Test

(Denton + Chintala, et al 2015)

- 40% of samples mistaken *by humans* for real photos



- Sharper images than max. lik. proxys (which min. $KL(\text{data} | \text{model})$):
- GAN objective = compromise between $KL(\text{data} | \text{model})$ and $KL(\text{model} | \text{data})$

Convolutional GANs

(Radford et al, arXiv 1511.06343)

Strided convolutions, batch normalization, only convolutional layers, ReLU and leaky ReLU



Biologically Plausible and Memory-Efficient Online Training of RNNs?

- The brain is a big RNN
- Backprop in a feedforward net may have plausible biological implementations (using the feedback weights to propagate credit information, targets or gradients)
 - *Y. Bengio, "Early Inference in Energy-Based Models Approximates Back-Propagation", arXiv:1510.02777*
- But what about backprop through time?
 - Requires storing the state of the network (i.e. the activations of all neurons) for an indefinitely long duration
 - Need to wait for the "end of the episode" (your life) to start learning
 - **Some form of online learning is necessary**
- If we had a strong critic trained to predict discounted future rewards (and we have it in our brain), then we would just need to backprop through it to update our policy
- Critic is missing information (intermediate actions and observations)

MILA: Montreal Institute for Learning Algorithms

