

What Can I Do With Deep Learning?

Kyle Kastner

Université de Montréal - MILA

Intern - IBM Watson @ Yorktown Heights

Automation Spectrum

Introspection
Statistics

Machine
Learning

Automation
Deep Learning

Torch (Lua)

statsmodels

sklearn

pylearn2

pymc3

Theano

Keras

patsy

sklearn-theano

Blocks

shogun

Lasagne

Technology

Theano logo, written in blue cursive font.

- Theano (UMontreal)
 - Optimizing compiler
 - Automatic differentiation
 - GPU or CPU set by single flag
 - THEANO_FLAGS="mode=FAST_RUN,device=gpu,floatX=float32"
- Caffe (Berkeley)
 - Computer vision tools in C++/CUDA (GPU)
- Torch (NYU/Facebook/Google)
 - Lua with C/CUDA operations

Image Processing (sklearn-theano)

```
dog_label = 'dog.n.01'  
cat_label = 'cat.n.01'  
clf = OverfeatLocalizer(top_n=1,  
    match_strings=[dog_label, cat_label])  
points = clf.predict(X)  
dog_points = points[0]  
cat_points = points[1]
```

http://sklearn-theano.github.io/auto_examples/plot_multiple_localization.html

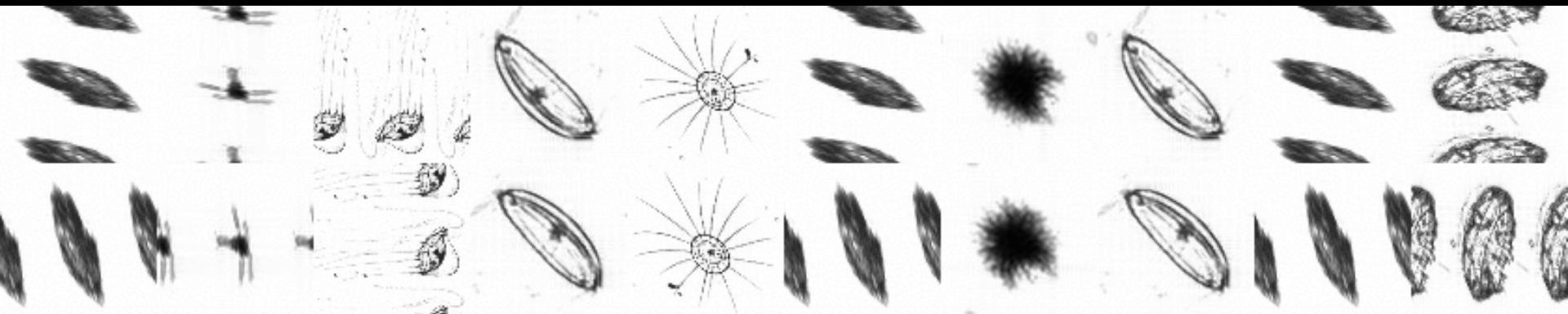
[4, 5]



More Image Processing (Lasagne)

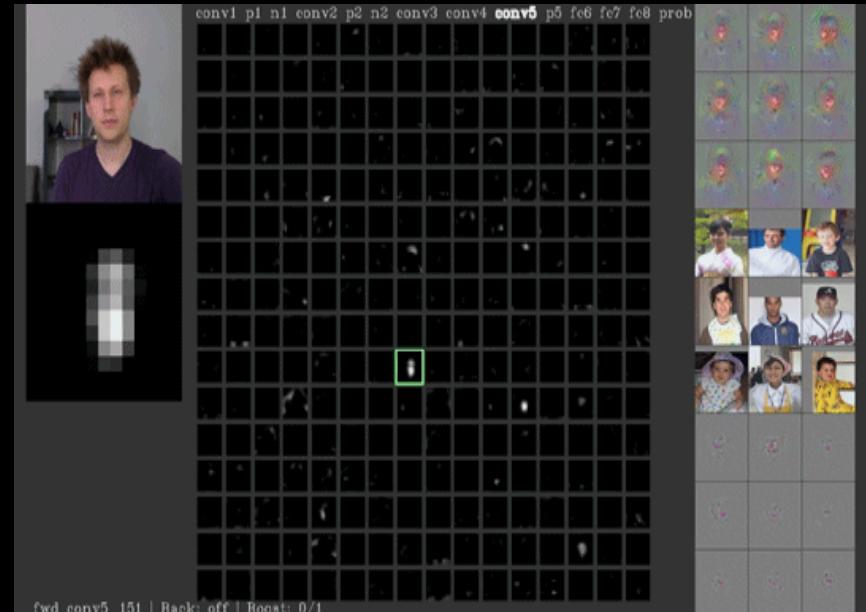
[6, 7]

- Won Kaggle competitions on:
 - Plankton: <http://benanne.github.io/2015/03/17/plankton.html>
 - Galaxies: <http://benanne.github.io/2014/04/05/galaxy-zoo.html>
 - Maintained and used by many researchers
 - (UGhent) Dieleman et. al. for above competitions



General Computer Vision (Caffe)

- Used for CV
 - Research
 - Industry
- Lots of unique tools
 - Pretrained networks
 - Jason Yosinski et. al.
 - Deep Vis Toolbox



<https://github.com/yosinski/deep-visualization-toolbox>

<https://www.youtube.com/watch?v=AgkfIQ4IGaM>

Generative Images (Torch / Theano)

- DRAW
 - Gregor et. al.
 - Reproduced by J. Bornschein
<https://github.com/jbornschein/draw>
- Eyescream
 - Denton et. al. (code in ref.)
- Generative Adversarial Networks
 - Goodfellow et. al. [9, 10, 11]



“Dreaming” (Caffe)

- Generated from trained network
 - GoogleNet
- Turns out it is pretty cool!
- Neural network art?
- Calista and The Crashroots:
Deepdream
 - Samim Winiger
- Details and code
 - Mordvinstev, Olah, Tyka



<http://googleresearch.blogspot.ch/2015/06/inceptionism-going-deeper-into-neural.html>

<https://github.com/google/deepdream>

Playing Atari (Torch)

- Learning
 - Raw images (4 frames)
 - Score
 - Controls (per game)
 - Goal: make score go up
 - ... and that's it!
 - Deep Q Learning
 - Minh et. al.

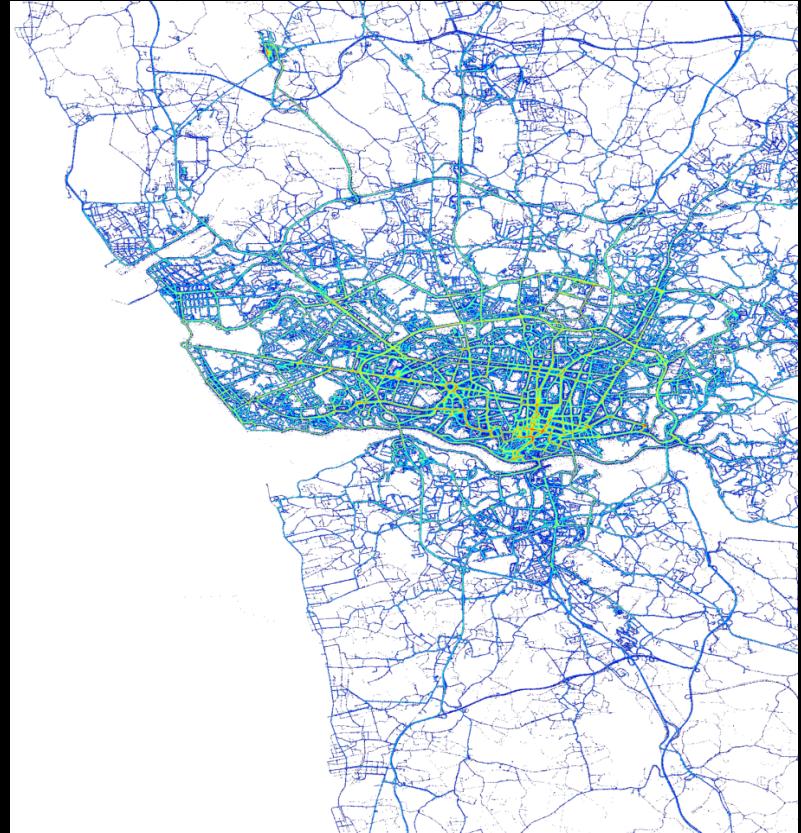
<https://github.com/kuz/DeepMind-Atari-Deep-Q-Learner>



Winning Kaggle Competitions (Theano)

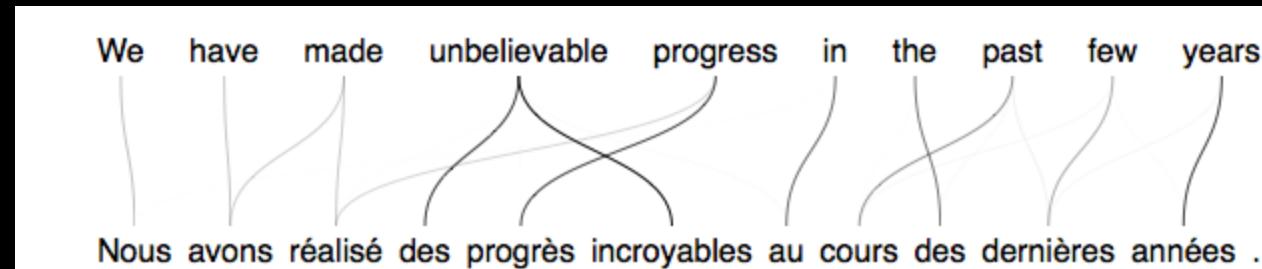
- Heterogenous Data
 - Time
 - Taxi ID, Client ID
 - Metadata
 - GPS Points
- Predict where a taxi is going
- Brebisson, Simon, Auvolat
(UMontreal, ENS Cachan,
ENS Paris)

[13]



Text to Text Translation (Blocks)

- Recurrent neural networks
- Attention mechanism (shown in diagram)
- More discussion later
- Part of larger movement
 - Neural Machine Translation, from Cho et. al.
 - Jointly Learning to Align and Translate, Bahdanau et. al.



Handwriting Generation

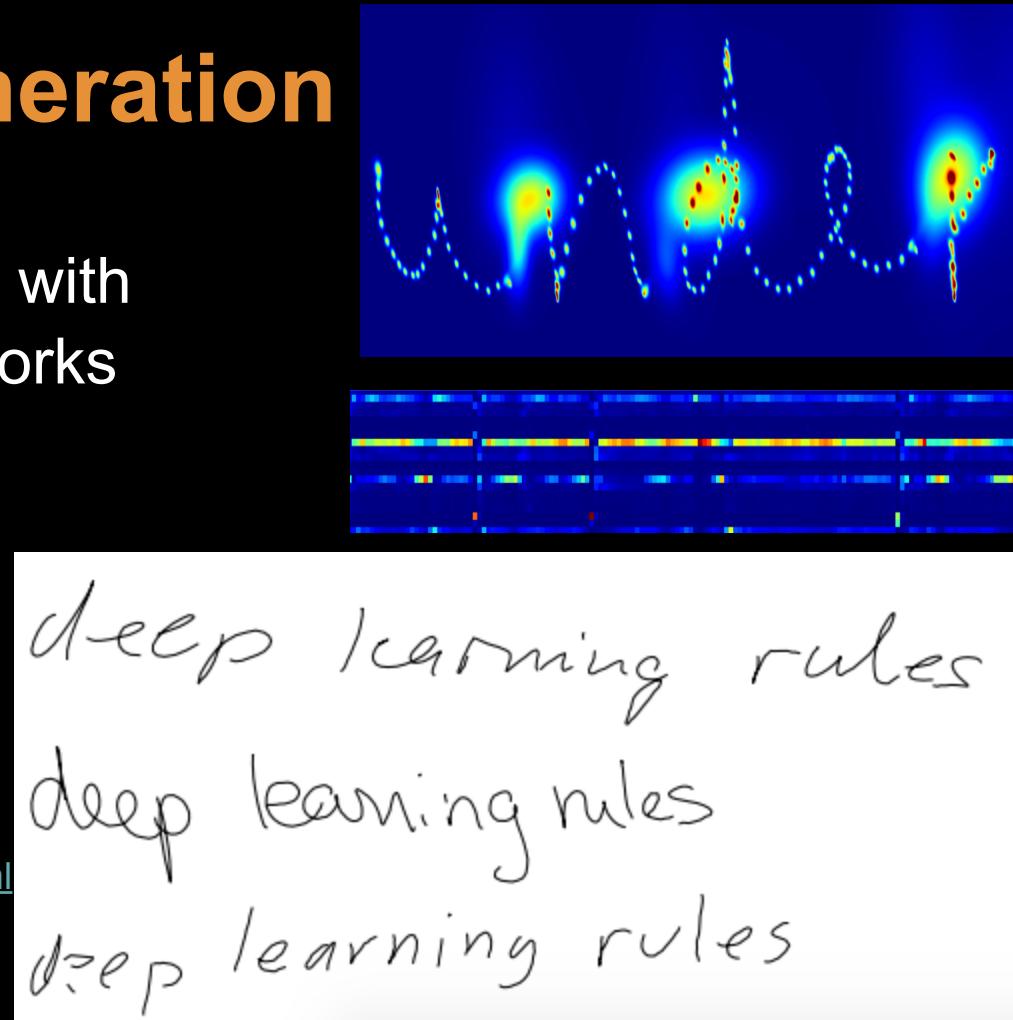
- Generating Sequences with Recurrent Neural Networks

Alex Graves

<http://arxiv.org/abs/1308.0850>

- GMM + Bernoulli per timestep
- Conditioned on text

<http://www.cs.toronto.edu/~graves/handwriting.html>



Babi RNN (Keras)

```
sentrnn = Sequential()
sentrnn.add(Embedding(vocab_size,
    EMBED_HIDDEN_SIZE, mask_zero=True))
sentrnn.add(RNN(EMBED_HIDDEN_SIZE,
    SENT_HIDDEN_SIZE, return_sequences=False))

qrnn = Sequential()
qrnn.add(Embedding(vocab_size,
    EMBED_HIDDEN_SIZE))
qrnn.add(RNN(EMBED_HIDDEN_SIZE,
    QUERY_HIDDEN_SIZE, return_sequences=False))

model = Sequential()
model.add(Merge([sentrnn, qrnn], mode='concat'))
model.add(Dense(SENT_HIDDEN_SIZE +
    QUERY_HIDDEN_SIZE, vocab_size,
    activation='softmax'))
```

- 1 John moved to the bedroom.
- 2 ***Mary grabbed the football there.***
- 3 Sandra journeyed to the bedroom.
- 4 Sandra went back to the hallway.
- 5 Mary moved to the garden.
- 6 ***Mary journeyed to the office.***
- 7 Where is the ***football?*** office 2 6

Based on tasks proposed in:
Towards AI-Complete Question Answering,
Weston et. al

Keras recreation of baseline documented at
http://smerity.com/articles/2015/keras_qa.html

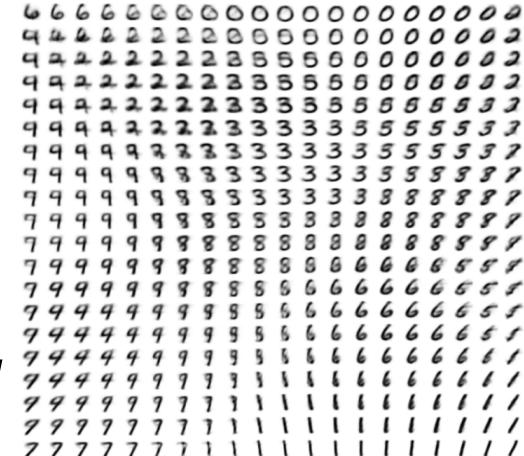
Conditioning Feedforward

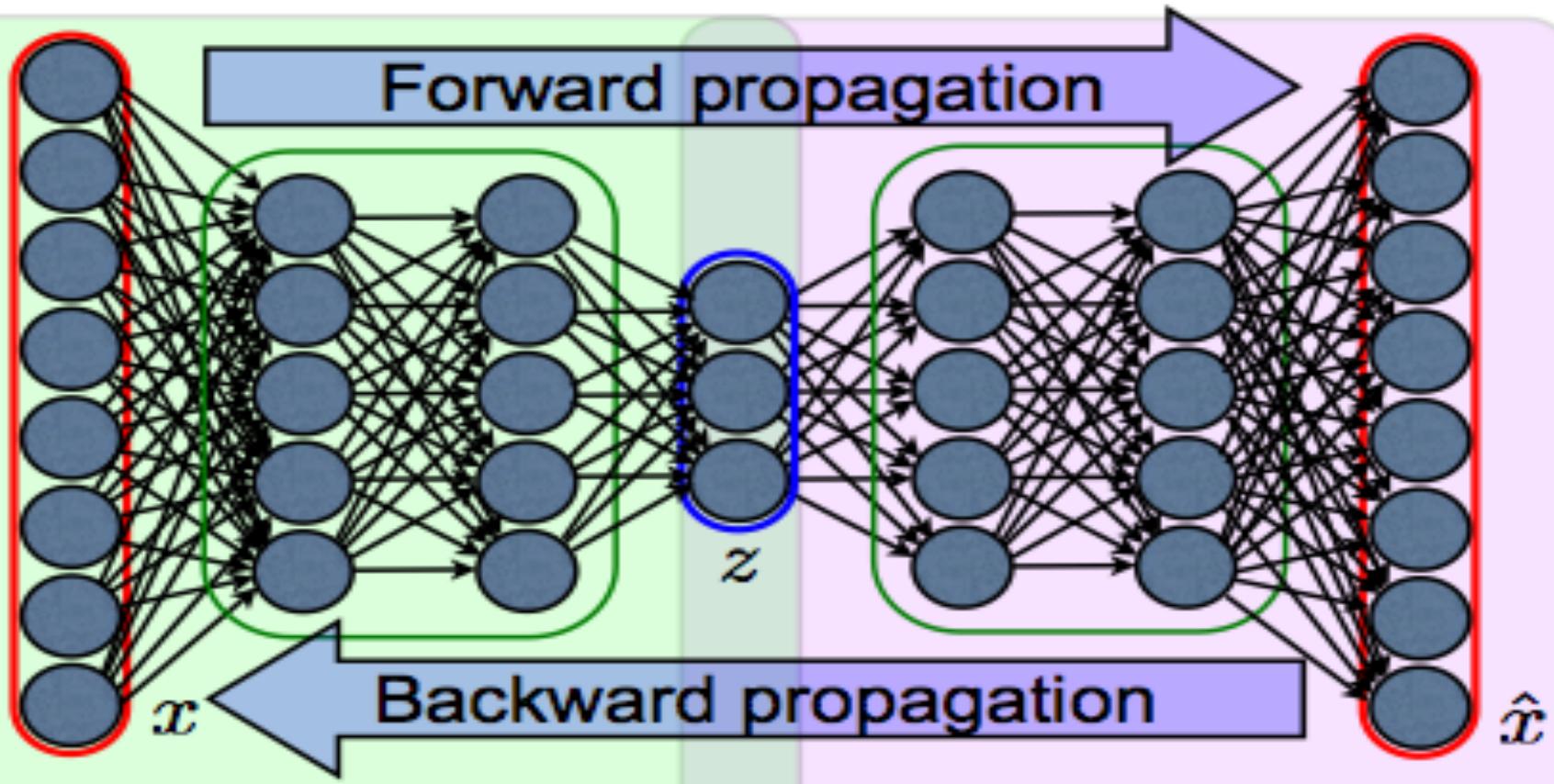
- Concatenate features
 - `concatenate((X_train, conditioning), axis=1)`
 - $p(y | X_1 \dots X_n, L_1 \dots L_n)$
- One hot label L (scikit-learn `label_binarize`)
- Can also be real valued
- Concat followed with multiple layers to “mix”

Latent Factor Generative Models

- Auto-Encoding Variational Bayes
D. Kingma and M. Welling
 - Variational Autoencoder (VAE)
 - Stochastic Backpropagation and Approximate Inference in Deep Generative Models
 - *Rezende, Mohamed, Wierstra*
 - Semi-Supervised Learning With Deep Generative Models
 - *Kingma, Rezende, Mohamed, Welling*

[26, 27, 28]

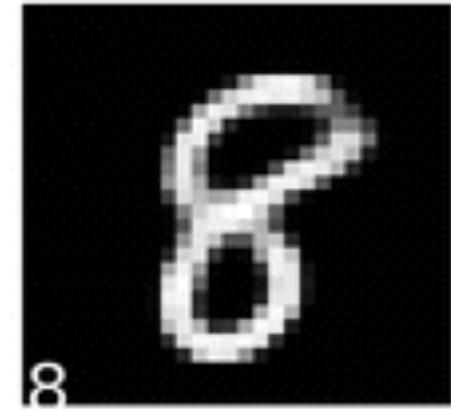
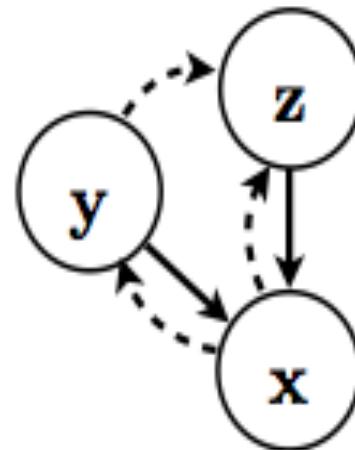
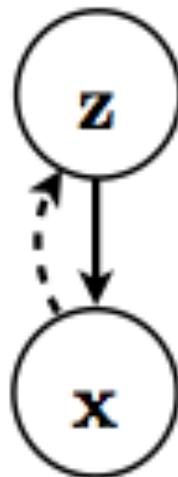




ENCODER

[18, 19, 20]
[26, 27, 28]

DECODER



8 0 1 2 3 4 5 6 7 8 9

4 0 1 2 3 4 6 6 7 8 9

7 0 1 2 3 4 5 6 7 8 9

3 0 1 2 3 4 5 6 7 8 9

4 0 1 2 3 4 5 6 7 8 9

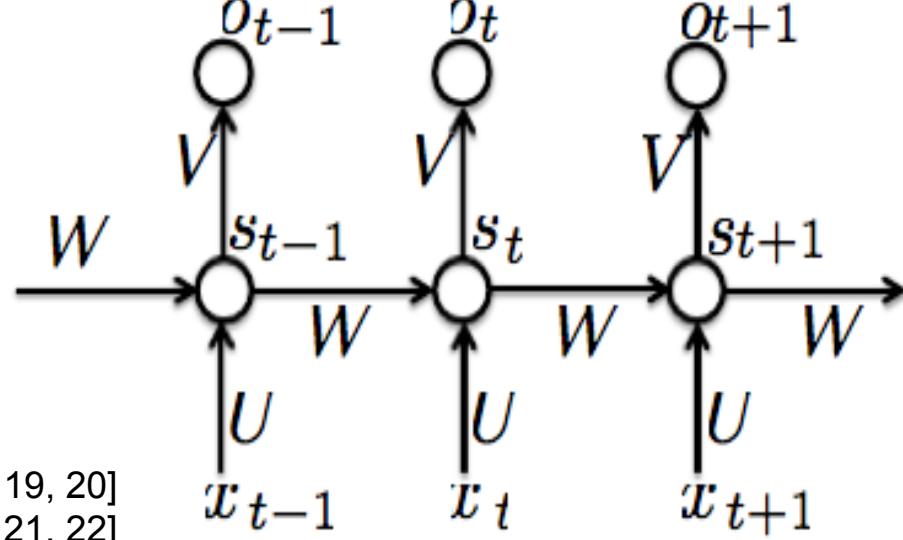
8 0 1 2 3 4 5 6 7 8 9

Conditioning, Visually

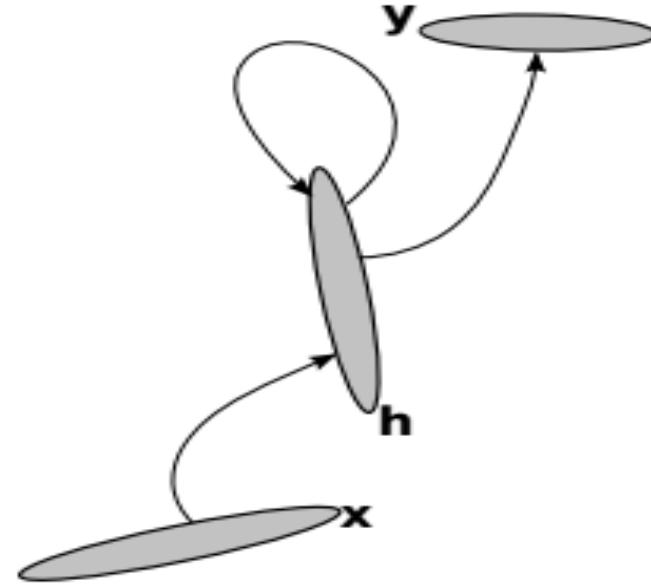
[26, 27, 28]

Recurrent Neural Network (RNN)

- Hidden state (s_t) encodes sequence info
 - $p(X_t | X_{<t})$ (in s_t) is *compressed representation*



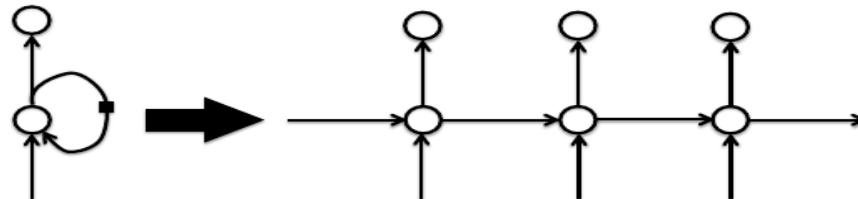
[18, 19, 20]
[16, 21, 22]



Conditioning In Recurrent Networks

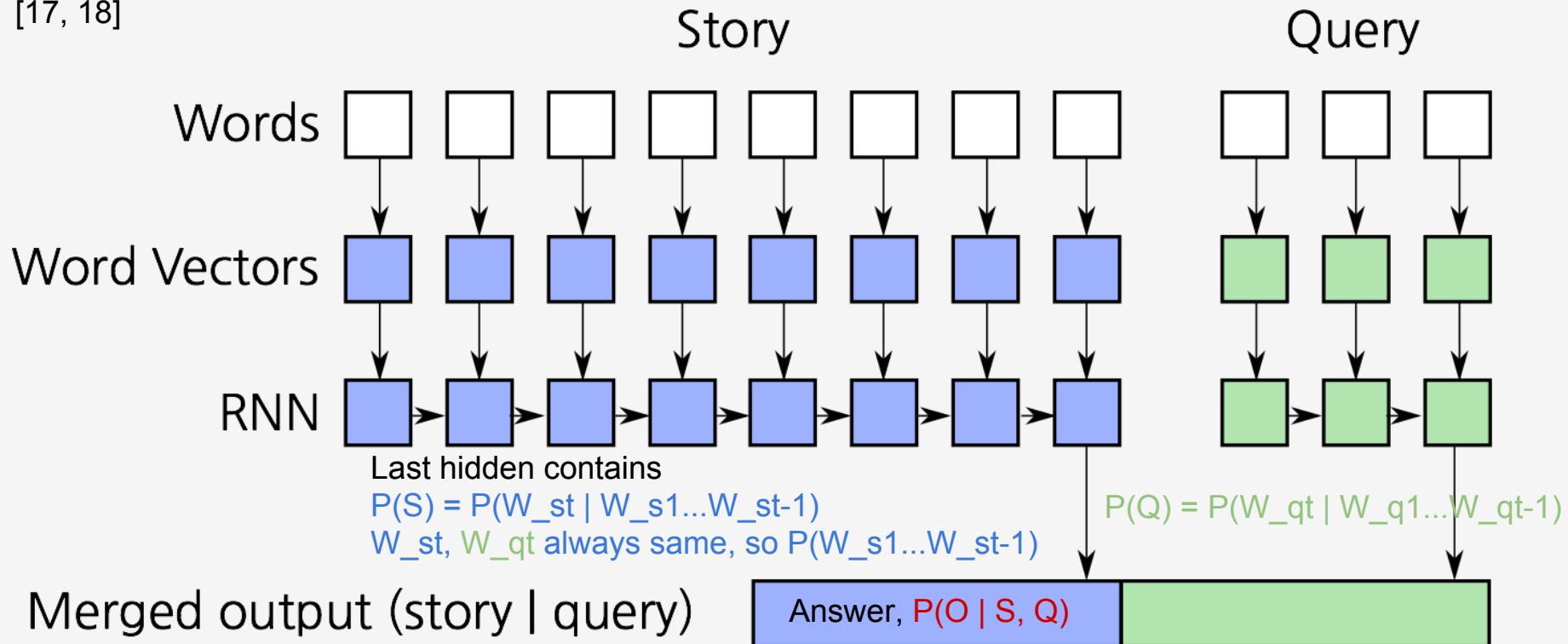
- RNNs model $p(X_t | X_{<t})$
- Initial hidden state can condition
 - $p(X_t | X_{<t}, c)$ where c is init. hidden state (context)
- Condition by concatenating in feedforward
 - Before recurrence or after
- Can do *all of the above*

[18, 19, 20]
[16, 21, 22]
[24]



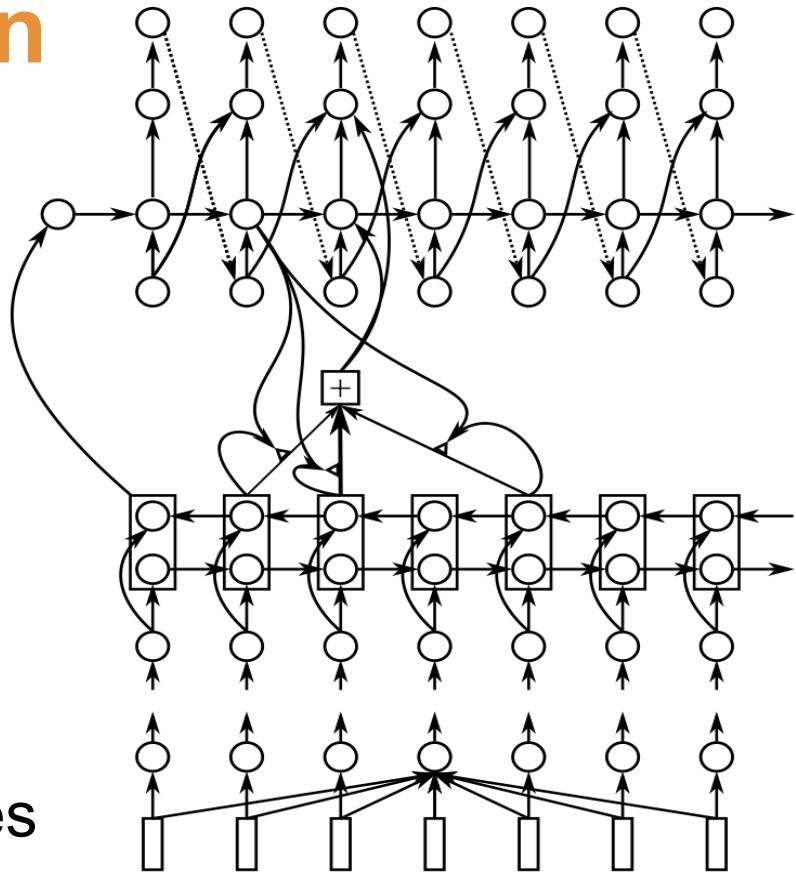
Looking Back at Babi RNN

[17, 18]



Memory and Attention

- Bidirectional RNN
 - $p(X_t | X_{<t}, X_{>t})$
 - Memory cells
 - Learn how to combine info
 - Dynamic lookup
- Research Directions
 - Neural Turing Machine
 - Memory Networks
 - Differentiable Datastructures



Continued

- All comes down to conditional info
- What problem are we trying to solve?
- Q + A
 - Condition on past Q
- Captions
 - Condition on image
- Computing
 - Condition on relevant memory



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



A woman holding a clock in her hand.

APIs and Companies

- Clarifai <http://www.clarifai.com/api>
- MetaMind <https://www.metamind.io/language/twitter>
- Indico <https://indico.io/>





@kastnerkyle

Thanks!

Repo <https://github.com/kastnerkyle/PyGotham2015>

Slides will be uploaded to <https://speakerdeck.com/kastnerkyle>

References (1)

- [1] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, Y. Bengio. "Theano: new features and speed improvements", Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop. <http://arxiv.org/abs/1211.5590>
- [2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell. "Caffe: Convolutional Architecture for Fast Feature Embedding". <http://arxiv.org/abs/1408.5093>
- [3] R. Collobert, K. Kavukcouglu, C. Farabet. "Torch7: A Matlab-like Environment For Machine Learning", NIPS 2011.
http://cs.nyu.edu/~koray/files/2011_torch7_nipsw.pdf
- [4] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun. "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks". <http://arxiv.org/abs/1312.6229>
- [5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. "Going Deeper with Convolutions". <http://arxiv.org/abs/1409.4842>
- [6] Sander Dieleman, Kyle W. Willett, Joni Dambre. "Rotation-invariant convolutional neural networks for galaxy morphology prediction". MRAS 2015.
<http://arxiv.org/abs/1503.07077>
- [7] A. van den Oord, I. Korshunova, J. Burms, J. Degrave, L. Pigou, P. Buteneers, S. Dieleman. "National Data Science Bowl, Plankton Challenge".
<http://benanne.github.io/2015/03/17/plankton.html>

References (2)

- [8] J. Yosinski, J. Clune , A. Nguyen, T. Fuchs, H. Lipson. "Understanding Neural Networks Through Deep Visualization".
http://yosinski.com/media/papers/Yosinski_2015_ICML_DL_Understanding_Neural_Networks_Through_Deep_Visualization_.pdf
- [9] K. Gregor, I. Danihelka, A. Graves, D. Rezende, D. Wierstra. "DRAW: Directed Recurrent Attention Writer". <http://arxiv.org/abs/1502.04623>
- [10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. "Generative Adversarial Networks", NIPS 2014.
<http://arxiv.org/abs/1406.2661>
- [11] E. Denton, S. Chintala, A. Szlam, R. Fergus. "Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks".
<http://arxiv.org/abs/1506.05751>
- [12] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller. "Playing Atari with Deep Reinforcement Learning", Nature 2015. <https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf>
- [13] A. Brebisson, E. Simon, A. Auvolat. "Taxi Destination Prediction Challenge Winners' Report."
https://github.com/adbrebs/taxi/blob/master/doc/short_report.pdf
- [14] K. Cho, B. Merrienboer, C. Gulchere, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". EMNLP 2014. <http://arxiv.org/abs/1406.1078>
- [15] D. Bahdanau, K. Cho, Y. Bengio. "Neural Machine Translation By Jointly Learning To Align and Translate". ICLR 2015. <http://arxiv.org/abs/1409.0473>
- [16] A. Graves. "Generating Sequences With Recurrent Neural Networks", 2013. <http://arxiv.org/abs/1308.0850>
- [17] J. Weston, A. Bordes, S. Chopra, T. Mikolov, A. Rush. "Towards AI-Complete Question Answering". <http://arxiv.org/abs/1502.05698>

References (3)

- [18] Y. Bengio, I. Goodfellow, A. Courville. "Deep Learning", in preparation for MIT Press, 2015. <http://www.iro.umontreal.ca/~bengioy/dlbook/>
- [19] D. Rumelhart, G. Hinton, R. Williams. "Learning representations by back-propagating errors", Nature 323 (6088): 533–536, 1986. http://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf
- [20] C. Bishop. "Mixture Density Networks", 1994. <http://research.microsoft.com/en-us/um/people/cmbishop/downloads/Bishop-NCRG-94-004.ps>
- [21] D. Eck, J. Schmidhuber. "Finding Temporal Structure In Music: Blues Improvisation with LSTM Recurrent Networks". Neural Networks for Signal Processing, 2002. ftp://ftp.idsia.ch/pub/juergen/2002_ieee.pdf
- [22] A. Brandmaier. "ALICE: An LSTM Inspired Composition Experiment". 2008.
- [23] T. Mikolov, M. Karafiat, L. Burget, J. Cernocky, S. Khudanpur. "Recurrent Neural Network Based Language Model". Interspeech 2010. http://www.fit.vutbr.cz/research/groups/speech/publi/2010/mikolov_interspeech2010_IS100722.pdf
- [24] N. Boulanger-Lewandowski, Y. Bengio, P. Vincent. "Modeling Temporal Dependencies in High-Dimensional Sequences: Application To Polyphonic Music Generation and Transcription". ICML 2012. <http://www-etud.iro.umontreal.ca/~boulanni/icml2012>
- [25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, 86(11):2278-2324, 1998. <http://yann.lecun.com/exdb/mnist/>
- [26] D. Kingma, M. Welling. "Auto-encoding Variational Bayes". ICLR 2014. <http://arxiv.org/abs/1312.6114>
- [27] D. Rezende, S. Mohamed, D. Wierstra. "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". ICML 2014. <http://arxiv.org/abs/1401.4082>
- [28] A. Courville. "Course notes for Variational Autoencoders". IFT6266H15. https://ift6266h15.files.wordpress.com/2015/04/20_vae.pdf

References (4)

- [29] A. Graves, G. Wayne, I. Danihelka. "Neural Turing Machines". <http://arxiv.org/abs/1410.5401>
- [30] J. Weston, S. Chopra, A. Bordes. "Memory Networks". <http://arxiv.org/abs/1410.3916>
- [31] S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus. "End-To-End Memory Networks". <http://arxiv.org/abs/1503.08895>
- [32] Ankit Kumar, Ozan Irsoy, Jonathan Su, James Bradbury, Robert English, Brian Pierce, Peter Ondruska, Mohit Iyyer, Ishaan Gulrajani, Richard Socher. "Ask Me Anything: Dynamic Memory Networks for Natural Language Processing". <http://arxiv.org/abs/1506.07285>
- [33] Various. Proceedings of Deep learning Summer School Montreal 2015. <https://sites.google.com/site/deeplearningsummerschool/schedule>
- [34] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, Y. Bengio. "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention"
- [35] J. Chung, K. Kastner, L. Dinh, K. Goel, A. Courville, Y. Bengio. "A Stochastic Latent Variable Model for Sequential Data".
<http://arxiv.org/abs/1506.02216>
- [36] J. Bayer, C. Osendorfer. "Learning Stochastic Recurrent Networks". <http://arxiv.org/abs/1411.7610>
- [37] O. Fabius, J. van Amersmoot. "Variational Recurrent Auto-Encoders". <http://arxiv.org/abs/1412.6581>
- [38] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, Y. Bengio. "Attention Based Models For Speech Recognition". <http://arxiv.org/abs/1506.07503>