Deep Learning

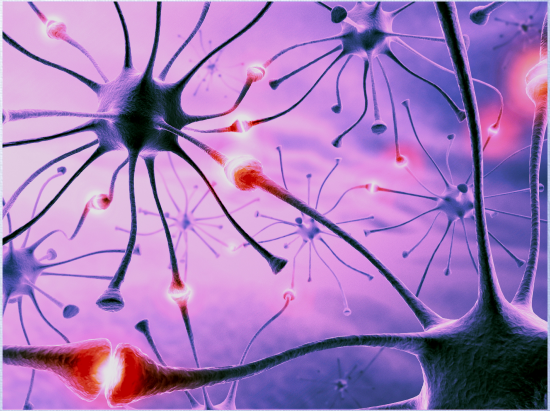
Modelling the World with Deep Artificial Neural Networks

Olof Mogren

June 2016

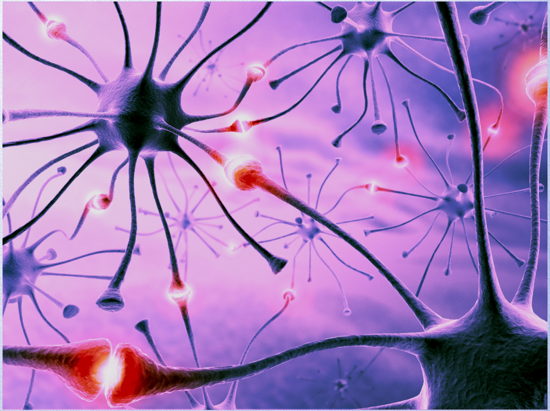
Deep Learning

- Deep artificial neural networks



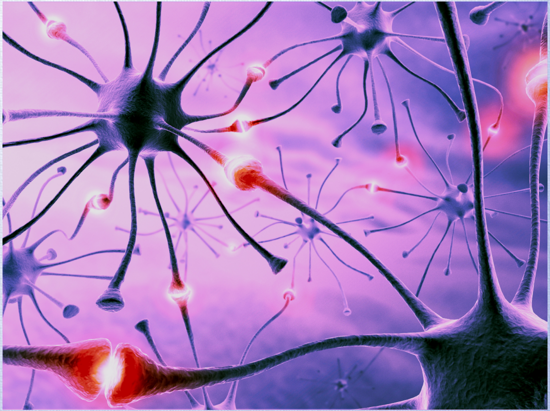
Deep Learning

- Deep artificial neural networks
- Learning from data (preferably big)



Deep Learning

- Deep artificial neural networks
- Learning from data (preferably big)
- Outperforms traditional methods in:



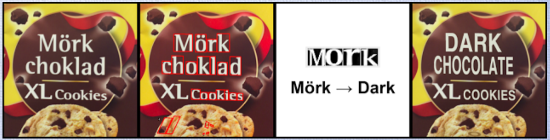
Deep Learning

- Deep artificial neural networks
- Learning from data (preferably big)
- Outperforms traditional methods in:
    - Image clasification

http://mogren.one/

Deep Learning

- Deep artificial neural networks
- Learning from data (preferably big)
- Outperforms traditional methods in:
  - Image clasification
  - Natural language processing
    - Machine translation
    - Sentiment analysis

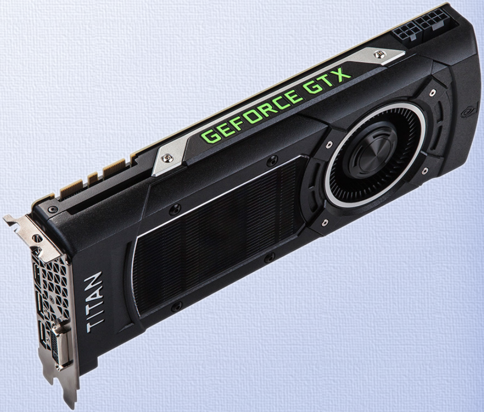


Deep Learning

- Deep artificial neural networks
- Learning from data (preferably big)
- Outperforms traditional methods in:
    - Image clasification
    - Natural language processing
        - Machine translation
        - Sentiment analysis
    - Reinforcement learning

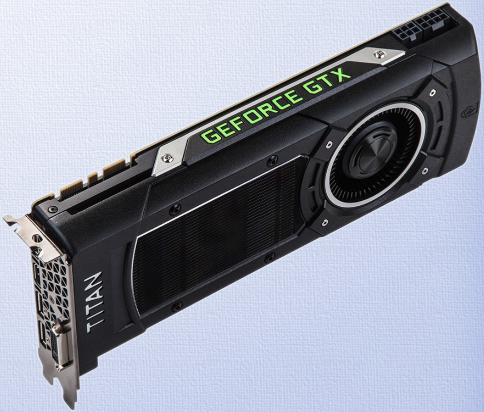
Why the Success?

- Progress in model design and algorithms



Why the Success?

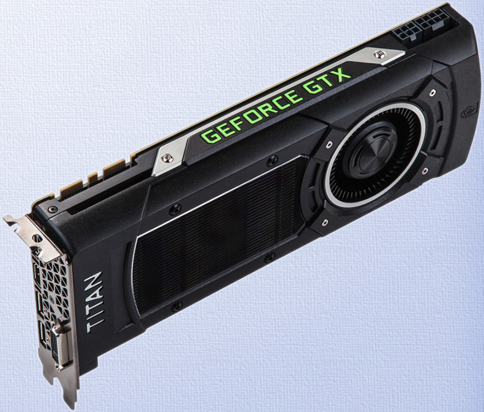
- Progress in model design and algorithms
- GPUs



http://mogren.one/

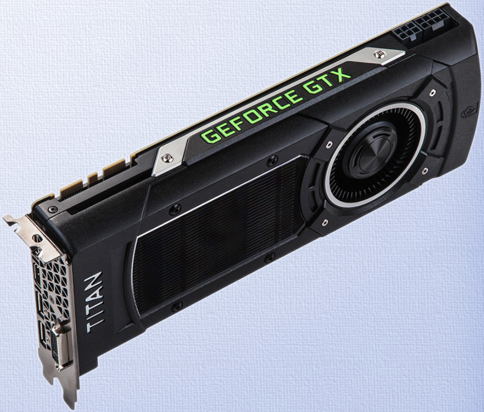
Why the Success?

- Progress in model design and algorithms
- GPUs
- Interest from researchers and industry



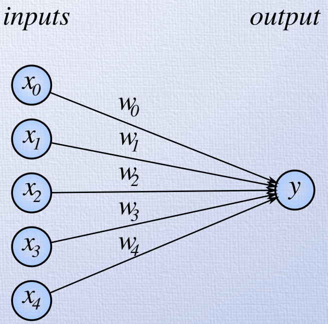
Why the Success?

- Progress in model design and algorithms
- GPUs
- Interest from researchers and industry
- Practical use (See previous slide)
  Real applications at Google, Facebook, Tesla, Microsoft, Apple, and others!



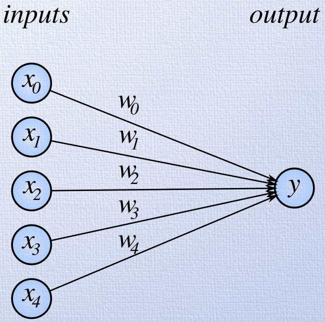
Perceptron

- 1957, Frank Rosenblatt



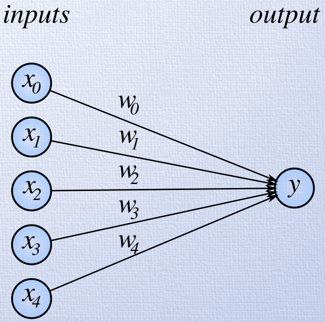
Perceptron

- 1957, Frank Rosenblatt
- Linear (binary) classification of inputs



Perceptron

- 1957, Frank Rosenblatt
- Linear (binary) classification of inputs
- Can not learn any non-linear function (e.g. exclusive or, XOR)



Modelling XOR

x_0	1	1	0
	0	0	1
		0	1
		x_1	

Modelling XOR

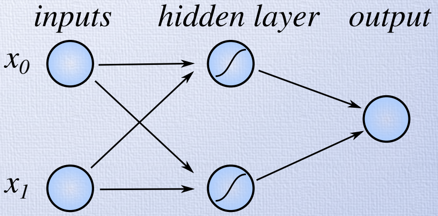
x_0	1	1	0
	0	0	1
		0	1
		x_1	

$x_0 \wedge \neg x_1$	1	1	
	0	0	1
		0	1
		$\neg x_0 \wedge x_1$	

Modelling XOR

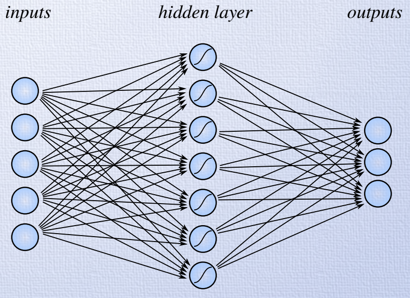
x_0			
	1	1	0
	0	0	1
		0	1
		x_1	

$x_0 \wedge \neg x_1$			
	1	1	
	0	0	1
		0	1
		$\neg x_0 \wedge x_1$	



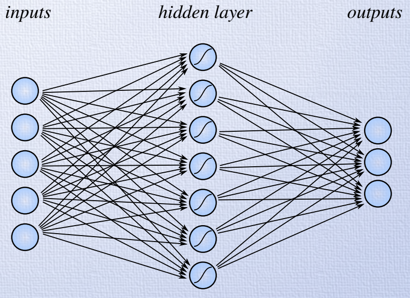
Artificial Neural Networks

- Combining many units lets us learn non-linear functions



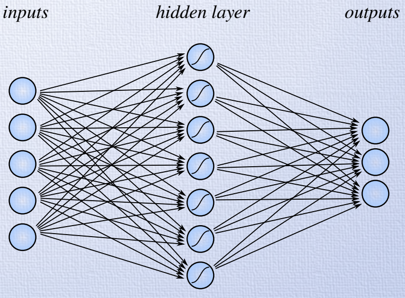
Artificial Neural Networks

- Combining many units lets us learn non-linear functions
- Each layer:



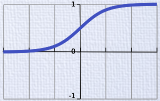
Artificial Neural Networks

- Combining many units lets us learn non-linear functions
- Each layer:
  - Linear transformation: $\mathbf{a} = W\mathbf{x} + \mathbf{b}$

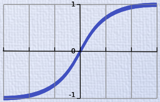


Artificial Neural Networks

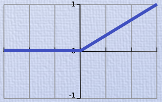
- Combining many units lets us learn non-linear functions
- Each layer:
  - Linear transformation: $\mathbf{a} = W\mathbf{x} + \mathbf{b}$
  - Non-linear (element-wise) activation: $\mathbf{h} = g(\mathbf{a})$



logistic



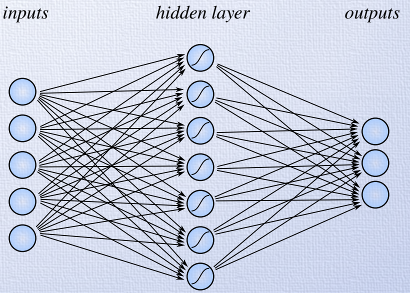
tanh



ReLU

Modelling Functions

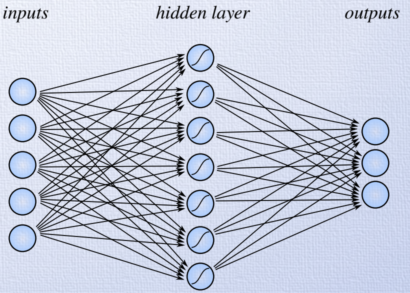
- Universal function approximation



details

Modelling Functions

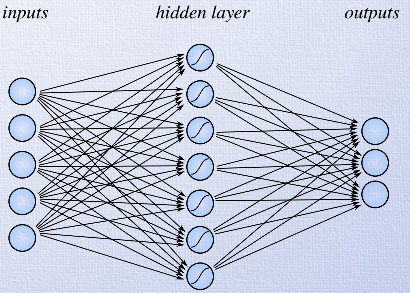
- Universal function approximation
- Stacking layers: function composition



details

Modelling Functions

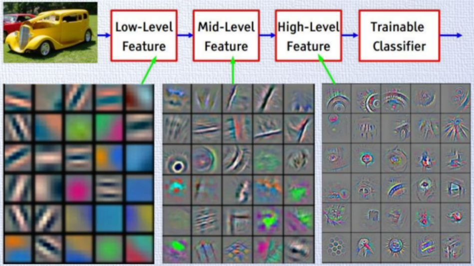
- Universal function approximation
- Stacking layers: function composition
- Train by propagating errors through model, updating weights



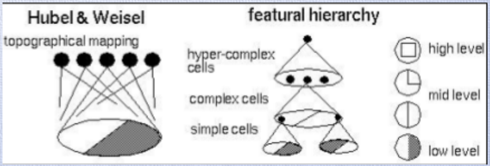
details

Representation Learning

- Each layer a non-linear transformation of inputs:
  $\mathbf{z} = sigmoid(W\mathbf{x} + \mathbf{b})$

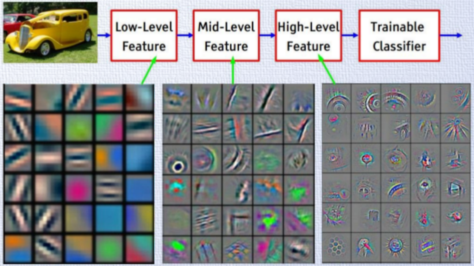


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

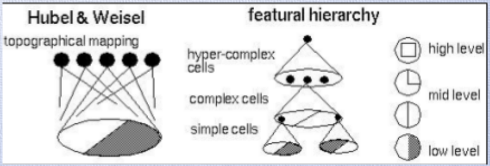


Representation Learning

- Each layer a non-linear transformation of inputs:
  $\mathbf{z} = sigmoid(W\mathbf{x} + \mathbf{b})$
- Learning representations; abstractions

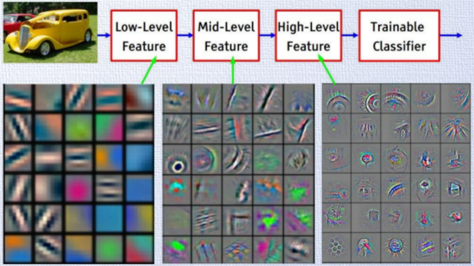


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

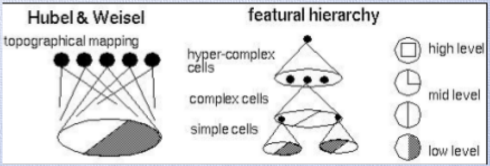


Representation Learning

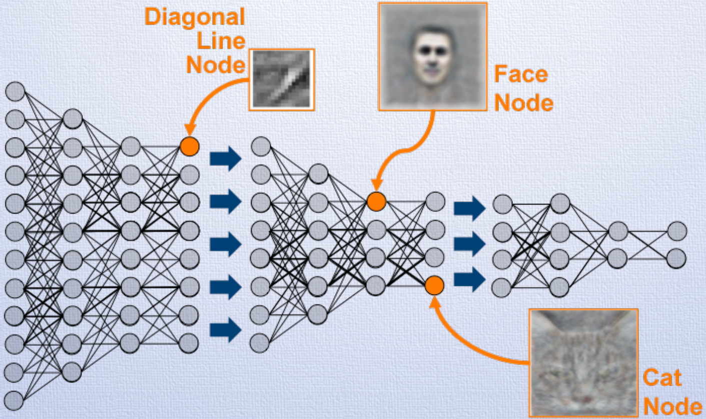
- Each layer a non-linear transformation of inputs:
  $\mathbf{z} = sigmoid(W\mathbf{x} + \mathbf{b})$
- Learning representations; abstractions
- In contrast to traditional machine learning, deep learning does not rely on feature engineering!



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]



Levels of Abstractions



http://mogren.one/

Convolutional Neural Networks

- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition

Convolutional Neural Networks

- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition

Convolutional Neural Networks

- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition

Convolutional Neural Networks

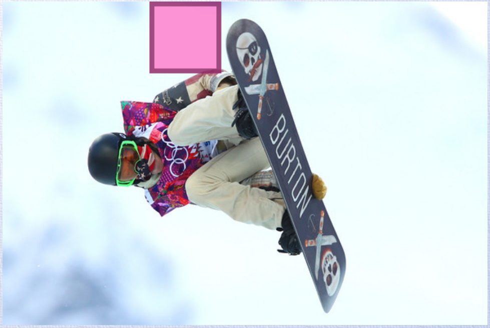
- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition

Convolutional Neural Networks

- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition

Convolutional Neural Networks

- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition



Convolutional Neural Networks

- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition

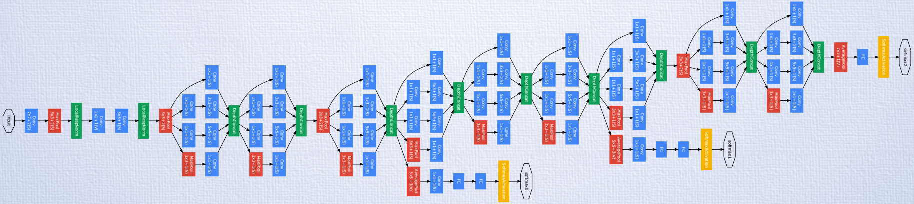
Convolutional Neural Networks

- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition

Convolutional Neural Networks

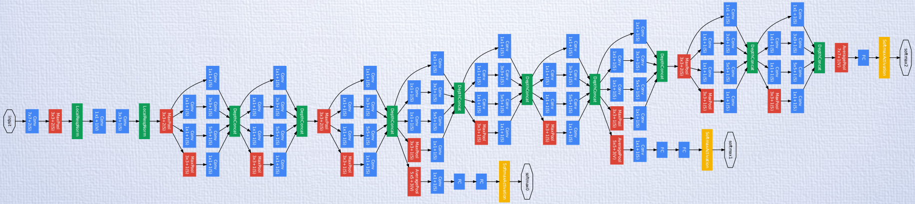
- Convolution filters; patches matching parts of input
- Successful e.g. for image recognition

Deep Learning for Image Processing



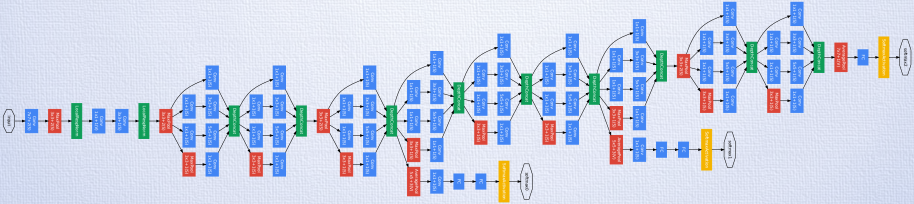
- Deeper and deeper

Deep Learning for Image Processing



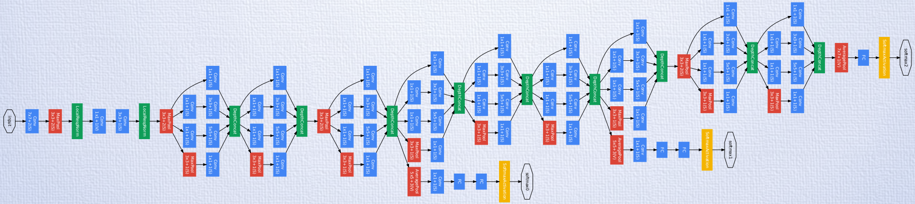
- Deeper and deeper
- 2014: GoogLeNet; 22 layers (illustration)

Deep Learning for Image Processing



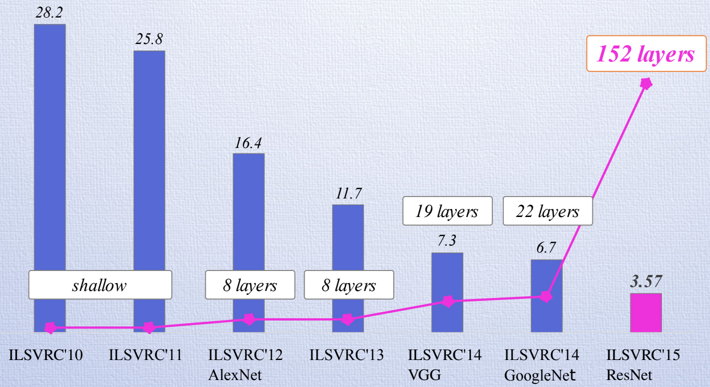
- Deeper and deeper
- 2014: GoogLeNet; 22 layers (illustration)
- 2015: Residual Nets; 152 layers

Deep Learning for Image Processing



- Deeper and deeper
- 2014: GoogLeNet; 22 layers (illustration)
- 2015: Residual Nets; 152 layers
- “Surpassed” human performance in 2015

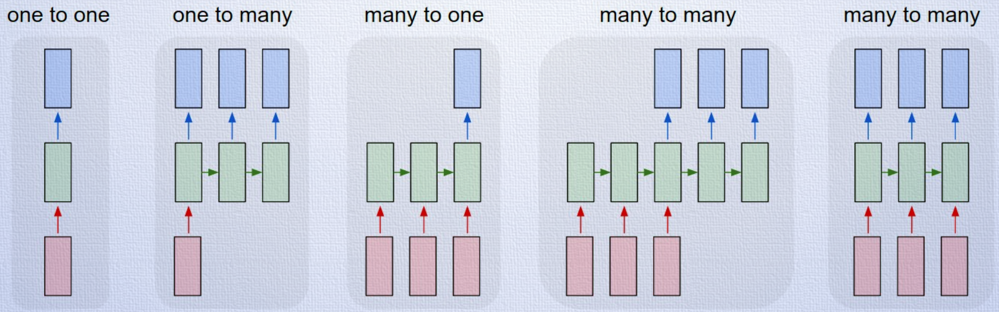
Deep Learning for Image Processing



ImageNet Classification top-5 error (%)

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

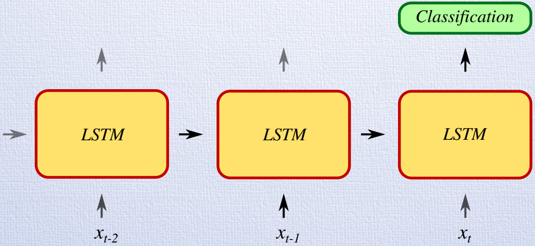
Sequence Modelling



Andrej Karpathy

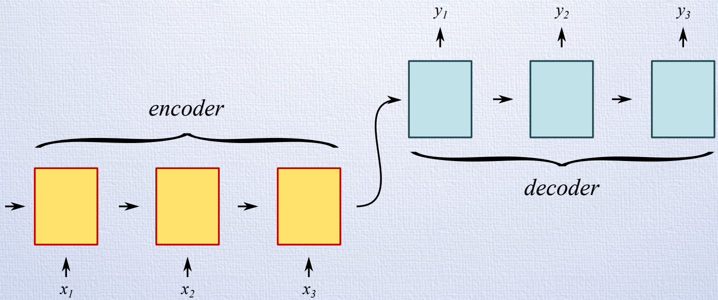
details

Sentiment Analysis



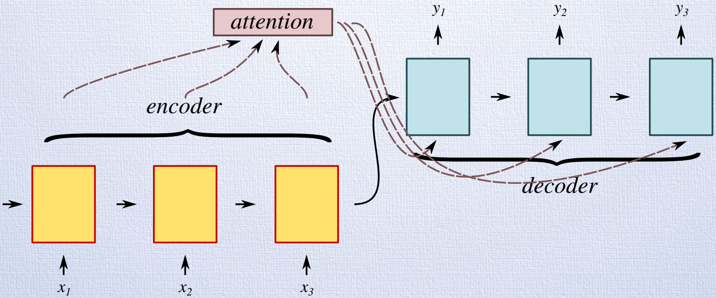
- Binary sequence classification

Machine Translation



- "Sequence-to-sequence" learning

Machine Translation



- "Sequence-to-sequence" learning
- Attention models

Encoding Questions

Responding to Queries using Encoder-Decoder Nets

Joint work with Jacob Hagstedt.

- Discussion forums: much information, little structure

Encoding Questions

Responding to Queries using Encoder-Decoder Nets

Joint work with Jacob Hagstedt.

- Discussion forums: much information, little structure
- Recommending users based on their competence

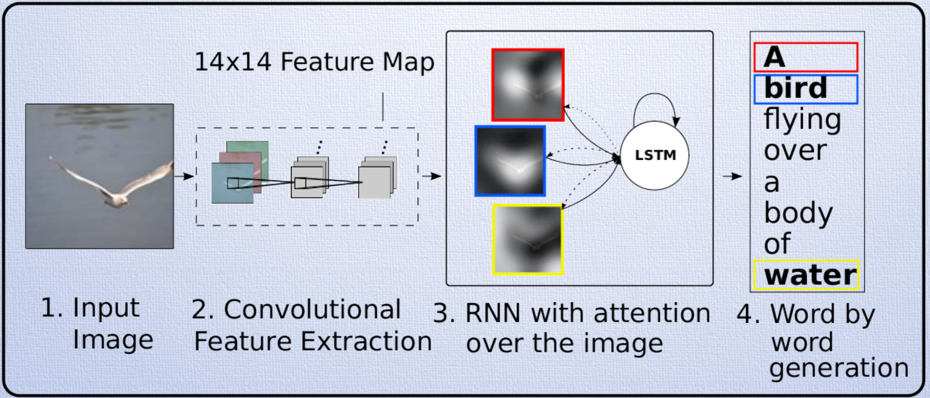
Encoding Questions

Responding to Queries using Encoder-Decoder Nets

Joint work with Jacob Hagstedt.

- Discussion forums: much information, little structure
- Recommending users based on their competence
- Recommending relevant threads and posts

Caption Generation



more

Entity Recognition

Swedish Medical Domain

Joint work with Sean Pavlov & Simon Almgren

Misstanke om **[herpes simplex-encefalit]** föreligger vid akut insjuknande med **[feber]**, cerebral påverkan med **[konfusion]**, sänkt **[medvetande]** och **[fokala neurologiska symtom]**.

ogren.one/

Entity Recognition
Swedish Medical Domain

Joint work with Sean Pavlov & Simon Almgren

- Medical domain text

http:/ ren.one/

Entity Recognition
Swedish Medical Domain

Joint work with Sean Pavlov & Simon Almgren

- Medical domain text
  - Writing style
  - Vocabulary
  - Synonymous
  - Hierarchy/Hyponymy

Entity Recognition

Swedish Medical Domain

Joint work with Sean Pavlov & Simon Almgren

- Medical domain text
    - Writing style
    - Vocabulary
    - Synonymous
    - Hierarchy/Hyponymy
- Character recurrent neural network

Entity Recognition

Swedish Medical Domain

Joint work with Sean Pavlov & Simon Almgren

- Medical domain text
  - Writing style
  - Vocabulary
  - Synonymous
  - Hierarchy/Hyponymy
- Character recurrent neural network
- Patient journal data

http:/ ren.one/

Deep Reinforcement Learning

- Learning a policy using an *infrequent* reward signal

Deep Reinforcement Learning

- Learning a policy using an *infrequent* reward signal
- Deep Q-Learning: Model the "action-value" function

Deep Reinforcement Learning

- Learning a policy using an *infrequent* reward signal
- Deep Q-Learning: Model the "action-value" function
- Atari games.

Deep Reinforcement Learning

- Learning a policy using an *infrequent* reward signal
- Deep Q-Learning: Model the "action-value" function
- Atari games.
- Alpha Go

Deep Reinforcement Learning

- Learning a policy using an *infrequent* reward signal
- Deep Q-Learning: Model the “action-value” function
- Atari games.
- Alpha Go
- Autonomous driving

http://**mogren.one**/

Appendix

Q-Learning Playing Atari Break-Out

Online Offline back to reinforcement learning

Attention Visualization

A woman is throwing a frisbee in a park.

[back to caption introduction]

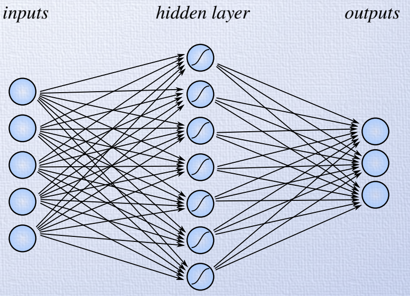
Attention Visualization

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

back to caption introduction

Learning

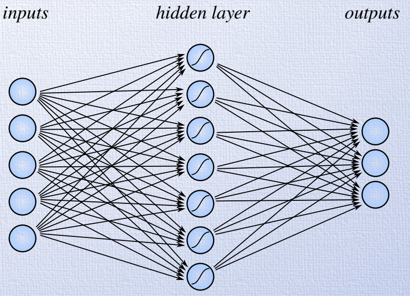
1. Forward pass (function application(s))



back to learning

Learning

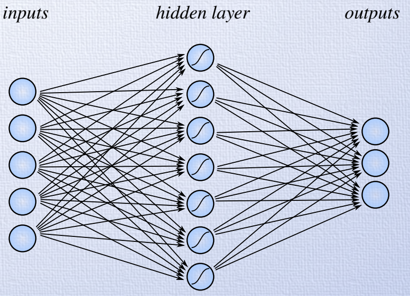
1. Forward pass (function application(s))
2. Compute error for output



back to learning

Learning

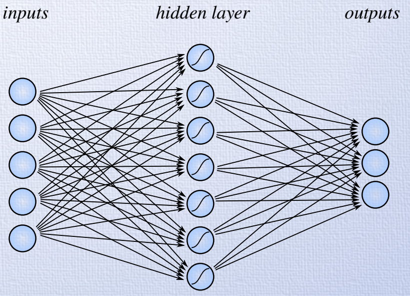
1. Forward pass (function application(s))
2. Compute error for output
3. Compute gradients (backpropagation)
   derivative of stacked layers: chain rule



back to learning

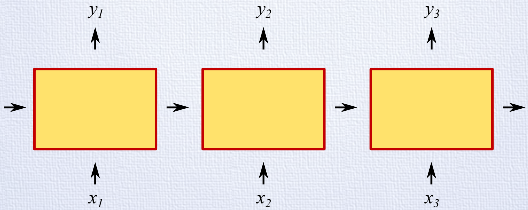
Learning

1. Forward pass (function application(s))
2. Compute error for output
3. Compute gradients (backpropagation)
   derivative of stacked layers: chain rule
4. Update weights (a small step)
   (minibatch stochastic gradient descent)



back to learning

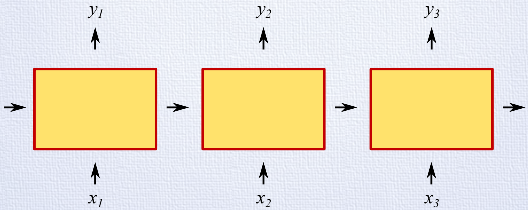
Modelling Language using RNNs



- Language models: $P(word_i | word_1, \ldots, word_{i-1})$

back to rnn click for example

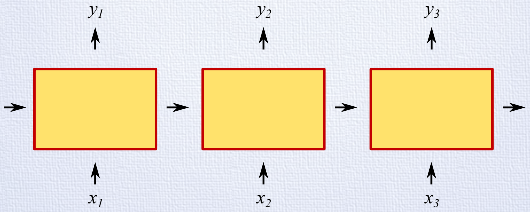
Modelling Language using RNNs



- Language models: $P(word_i | word_1, ..., word_{i-1})$
- Recurrent Neural Networks

back to rnn click for example

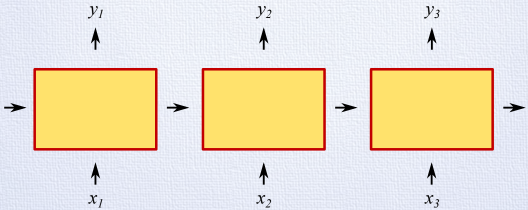
Modelling Language using RNNs



- Language models: $P(word_i | word_1, \ldots, word_{i-1})$
- Recurrent Neural Networks
- "Long Short-Term Memory" (LSTM)

back to rnn click for example

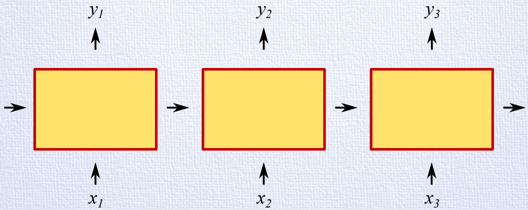
Modelling Language using RNNs



- Language models: $P(word_i | word_1, \ldots, word_{i-1})$
- Recurrent Neural Networks
- "Long Short-Term Memory" (LSTM)
- Fixed vector representation for sequences

back to rnn click for example

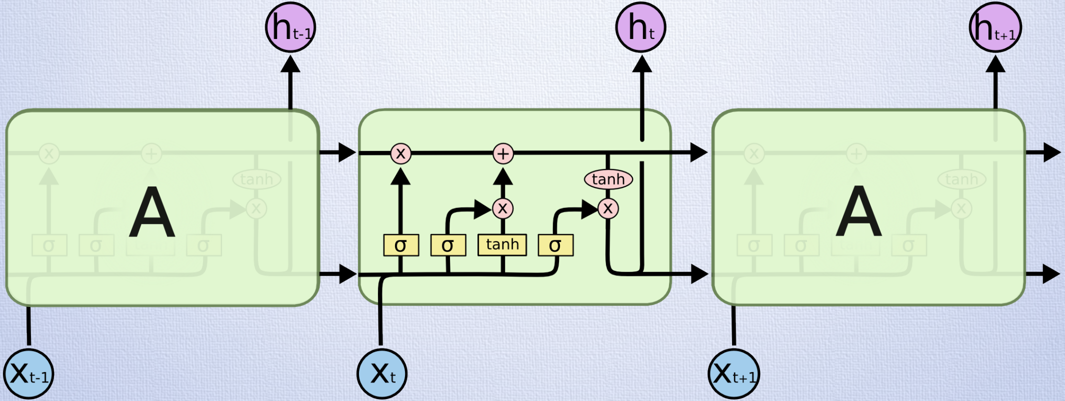
Modelling Language using RNNs



- Language models: $P(word_i \mid word_1, \ldots, word_{i-1})$
- Recurrent Neural Networks
- "Long Short-Term Memory" (LSTM)
- Fixed vector representation for sequences
- Language generation (sampling; beam search)

back to rnn click for example

LSTM

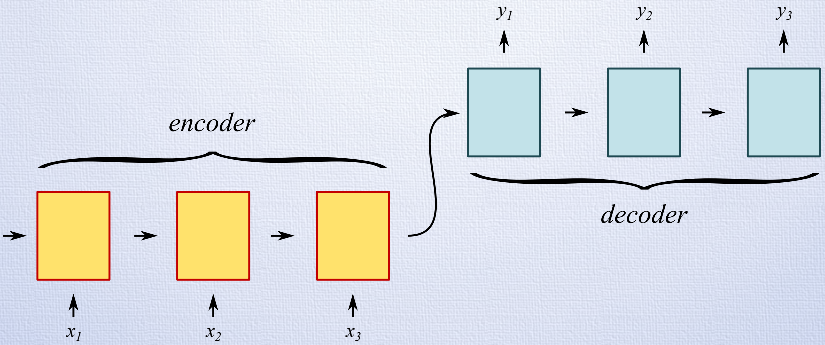


Christopher Olah

back to rnn

http://mogren.one/

Encoder-Decoder Framework



- Sequence to Sequence Learning with Neural Networks *Ilya Sutskever, Oriol Vinyals, Quoc V. Le, NIPS 2014*
- Neural Machine Translation (NMT)

Encoding Questions

Responding to Queries using Encoder-Decoder Nets

Joint work with Jacob Hagstedt

- Goal: assistant in forum environments
  (e.g. Slack, Stack Overflow)

more

Encoding Questions

Responding to Queries using Encoder-Decoder Nets

Joint work with Jacob Hagstedt

- Goal: assistant in forum environments (e.g. Slack, Stack Overflow)
- Word embeddings to find relevant comments

more

Encoding Questions

Responding to Queries using Encoder-Decoder Nets

Joint work with Jacob Hagstedt

- Goal: assistant in forum environments (e.g. Slack, Stack Overflow)
- Word embeddings to find relevant comments
- Learn to suggest relevant forum users (RNN)

more

Encoding Questions

Responding to Queries using Encoder-Decoder Nets

Joint work with Jacob Hagstedt

- Goal: assistant in forum environments
  (e.g. Slack, Stack Overflow)
- Word embeddings to find relevant comments
- Learn to suggest relevant forum users (RNN)
- Learn to respond to questions (Encoder-Decoder)

more

Discussion Suggestions - Word Embeddings

Joint work with Jacob Hagstedt

Q: I like to eat sushi for lunch

more

Discussion Suggestions - Word Embeddings

Joint work with Jacob Hagstedt

Q: I like to eat sushi for lunch

A1. Or just simply good lunch sushi (0.86)

more

Discussion Suggestions - Word Embeddings

Joint work with Jacob Hagstedt

Q: I like to eat sushi for lunch

A1. Or just simply good lunch sushi (0.86)

A2. I doesn't have to be a buffet, but I do tend to leave hungry when eating at regular restaurants :stuck_out_tongue: (0.79)

more

Discussion Suggestions - Word Embeddings

Joint work with Jacob Hagstedt

Q: I like to eat sushi for lunch

A1. Or just simply good lunch sushi (0.86)

A2. I doesn't have to be a buffet, but I do tend to leave hungry when eating at regular restaurants :stuck_out_tongue: (0.79)

A3. hmm, I think I'd skip the hotel. Is anyone up for the cantine or kistenpfennig bakery? they have some nice sandwiches, salads and warm meals for lunch.. (0.78)

more

Discussion Suggestions - Word Embeddings

Joint work with Jacob Hagstedt

Q: I like to eat sushi for lunch

A1. Or just simply good lunch sushi (0.86)

A2. I doesn't have to be a buffet, but I do tend to leave hungry when eating at regular restaurants :stuck_out_tongue: (0.79)

A3. hmm, I think I'd skip the hotel. Is anyone up for the cantine or kistenpfennig bakery? they have some nice sandwiches, salads and warm meals for lunch.. (0.78)

A4. lunch today? I would be up for a burger :grimacing: (0.78)

more

http://mogren.one/

Q: react native is the next big thing

back

Q: react native is the next big thing

A1. One big thing is that sense and qlikview now will run on the same engine (0.79)

back

Q: react native is the next big thing

A1. One big thing is that sense and qlikview now will run on the same engine (0.79)

A2. that is really big (0.77)

back

Q: react native is the next big thing

A1. One big thing is that sense and qlikview now will run on the same engine (0.79)

A2. that is really big (0.77)

A3. I'm thinking about starting a react project just to learn it and be prepared once native is released :simple_smile: (0.77)

back

Q: react native is the next big thing

A1. One big thing is that sense and qlikview now will run on the same engine (0.79)

A2. that is really big (0.77)

A3. I'm thinking about starting a react project just to learn it and be prepared once native is released :simple_smile: (0.77)

A4. hello <channel> , my client is currently considered whether to go for ios+android native apps or using react native - what would be your recommendations? (when should react native be considered instead of going for native ios/android apps) (0.77)

back

Memory Networks

- Attention refers back to internal memory; state of encoder

back

Memory Networks

- Attention refers back to internal memory; state of encoder
- Neural Turing Machines
- (End-To-End) Memory Networks:
  explicit memory mechanisms
  (out of scope today)

back

Entity Linking (EL)

`[Barack Obama] is the 44th President of the [US].`

1. Recognise entity mentions

[back to rnn]

Entity Linking (EL)

`[Barack Obama]` is the 44th President of the `[US]`.

1. Recognise entity mentions
2. Link each mention to database

back to rnn

http://mogren.one/

EL, Work in Progress

- Deep Char BI-LSTM

back to rnn

EL, Work in Progress

- Deep Char BI-LSTM
- One softmax per term (Kågebäck et.al.)

back to rnn

EL, Work in Progress

- Deep Char BI-LSTM
- One softmax per term (Kågebäck et.al.)
- Train on Wikipedia links

back to rnn

Char-RNNs

(Karpathy et.al. 2014)

- One LSTM module per character

back to rnn

Char-RNNs

(Karpathy et.al. 2014)

- One LSTM module per character
- Word vocabulary independence (learn OOV terms)
  - Malware classification?

back to rnn

Char-RNNs

(Karpathy et.al. 2014)

- One LSTM module per character
- Word vocabulary independence (learn OOV terms)
  - Malware classification?
- Character vocabulary is small

back to rnn

Char-RNNs

(Karpathy et.al. 2014)

- One LSTM module per character
- Word vocabulary independence (learn OOV terms)
  - Malware classification?
- Character vocabulary is small

- Learns to write like Shakespeare

back to rnn

Char-RNNs

(Karpathy et.al. 2014)

- One LSTM module per character
- Word vocabulary independence (learn OOV terms)
  - Malware classification?
- Character vocabulary is small

- Learns to write like Shakespeare
- Or like linux programmers

back to rnn

Char-RNNs

(Karpathy et.al. 2014)

- One LSTM module per character
- Word vocabulary independence (learn OOV terms)
  - Malware classification?
- Character vocabulary is small

- Learns to write like Shakespeare
- Or like linux programmers
  - Mostly correct, (brackets are opened and closed, indentation looks good)

back to rnn

Char-RNNs

(Karpathy et.al. 2014)

- One LSTM module per character
- Word vocabulary independence (learn OOV terms)
  - Malware classification?
- Character vocabulary is small

- Learns to write like Shakespeare
- Or like linux programmers
  - Mostly correct, (brackets are opened and closed, indentation looks good)
- Or like Wikipedia markup

back to rnn

Char-RNNs

(Karpathy et.al. 2014)

- One LSTM module per character
- Word vocabulary independence (learn OOV terms)
  - Malware classification?
- Character vocabulary is small

- Learns to write like Shakespeare
- Or like linux programmers
  - Mostly correct, (brackets are opened and closed, indentation looks good)
- Or like Wikipedia markup
- **Or the text-files on my harddrive**

back to rnn

With a Little Training

X. Z
rbndtr Jltcub aoes re) sneecietioon ine gugu soiitid
as)iidoyan tsstaprm f tss
aBeAur:.e,3a2oOannsdnss 3adr firet nte
tte tetaneg,.1 Ene)l3.<

oides nor:
60
5. fipnapn::5, UdeirBseess-ed‘iesr.:
ngsever, tciph:dsctum
se ralr msute 5emriipee vmetit Iooms m:eieeaaoeedes?
niei?nr rieanoitt 1 Dsehane
epd,ioe tee, onemseLsl

With Some More Training

```
u5ual ove repant org meucien pashe..  / ohe the man.
of namiof in loine
are proak ois unae if doubou u rew}

 tr L is Yave tudtisat (theaon ir olcilar
it in, dil riRer nanes and is prome urdic unpu cunre
inh Is surgend the fcare celaoao ale Larcawanbusio
= 1 maxvand bis / o -pseire
counmt bhen-a }
etitwl ind tuvabToin maMatias otyraoss epronk

 [inn os on:

The aluliititat o/ pesede thav.
```

ci Vill has 1 neighbors.
Graph has 2 vertard. Number on coykut: 2583Moher.
Number }411185n169V71}305911 sdmgap Kart:
CL Volvo Penta Volvo Penta in you hax on Volvo
TrucladionTenkoy i meance can 06:715g333139 >-C vexter:
unbe Bactizaces processed deviewroced an than quardy
processed before: 150
erreight rasunce qual. Number stam ald Boat.
Decided on. 55012418571607Meag.b>mlwyt arluto
aufortion!
Dlidlis Bont!
- C2 ((Rekroce: Prepors: Frapers. Number much: and
yimile have 2 vertices anbuct cont87a

b9 Nlkhwvear has 1 neighbors.
Graph has 6 neighbors.
Graph has 9 vertices and 699 edges.
This on centroeds basbanam has no soluth. Number of centroids proclesc. Number of centroids processed before: 235
rearnt has 1 neighbors.
Graph has 4 vertices and 7 edges.
This instance has no solution!
Decided on centroid: aent. Nerticed least has no soam too strast. Number of too small (size: 5pgabl.
Number of centroids processed resuase Tas instance cast. Number of centroids processed before: 342 daser ha