A non-technical intro to deep learning

January 3, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

What is deep learning for you?

Input $\longrightarrow$ Output.

Make a prediction that is correct with a certain probability.

A too well known example (ImageNet):

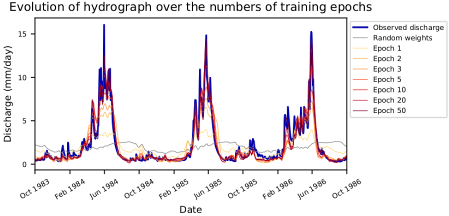
```
top 10 classes and their scores: 670 | n03791053 motor scoo 0.636128
top 10 classes and their scores: 880 | n04509417 unicycle, 0.261106
top 10 classes and their scores: 665 | n03785016 moped 0.0432479
top 10 classes and their scores: 573 | n03444034 go-kart 0.0156194
top 10 classes and their scores: 444 | n02835271 bicycle-bu 0.00917528
top 10 classes and their scores: 920 | n06874185 traffic li 0.00490618
top 10 classes and their scores: 518 | n03127747 crash helm 0.00412505
top 10 classes and their scores: 822 | n04311174 steel drum 0.0024649
top 10 classes and their scores: 733 | n03976657 pole 0.00223075
top 10 classes and their scores: 557 | n03355925 flagpole, 0.00199626
```

Deep Learning can predict more.

More than vision

Deep Learning (LSTMs) for Rainfall-Runoff modelling:

Input: **Rainfall over time** in a region
Output: **Water Flow in a river** (at a fixed position) over time



Kratzert et al., 2018

`https://www.researchgate.net/publication/325129495_Rainfall-Runoff_modelling_using_Long-Short-Term-Memory_LSTM_networks`

More than vision

Deep Learning for Quantum Chemistry:
Prediction of potential-energy surfaces of molecules
How much is a test atom attracted or repulsed around some molecule?

Home > The Journal of Chemical Physics > Volume 148, Issue 24 > 10.1063/1.5019779

Full . Published Online: 29 March 2018 Accepted: March 2018

SchNet – A deep learning architecture for molecules and materials

The Journal of Chemical Physics **148**, 241722 (2018); https://doi.org/10.1063/1.5019779

K. T. Schütt[1,a)], H. E. Sauceda[2], P.-J. Kindermans[1], A. Tkatchenko[3,b)], *and* K.-R. Müller[1,4,5,c)]

Schuett et al., JCS 2018.
`https://aip.scitation.org/doi/10.1063/1.5019779`
`https://www.youtube.com/watch?v=GJCeCiSxWWw`

More than vision

Generative models for structures of chemical molecules:

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Letter

Fréchet ChemNet Distance: A metric for generative models for molecules in drug discovery

Kristina Preuer, Philipp Renz, Thomas Unterthiner, Sepp Hochreiter, and Günter Klambauer

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.8b00234 • Publication Date (Web): 17 Aug 2018

Preuer et al., JCIM 2018

`https://pubs.acs.org/doi/pdfplus/10.1021/acs.jcim.8b00234`

Statistical Machine Learning from an eagle's view

- Prediction: f : Input space $\longrightarrow$ Output space.
- Make a prediction that is correct with a certain probability of mistake.
- model has trainable parameters
- Use (labeled!) training data and a loss function to optimize parameters

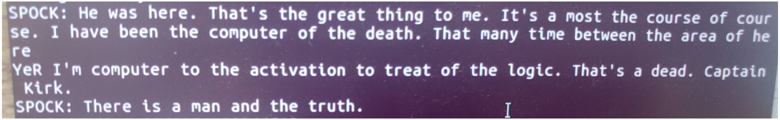
classical examples

Detection in Images: class label and position

- Input: image.
- Output: a number of bounding boxes

classical examples

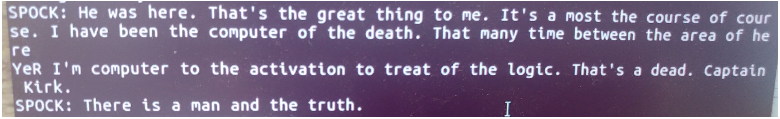
Text Summarization:



- Input: sequence of words $(w_1, \ldots, w_K)$
- Output: sequence of words $(w_1, \ldots, w_L)$
  (Example: "The text is about Star Trek. It mentions Spock and James T. Kirk.")

classical examples

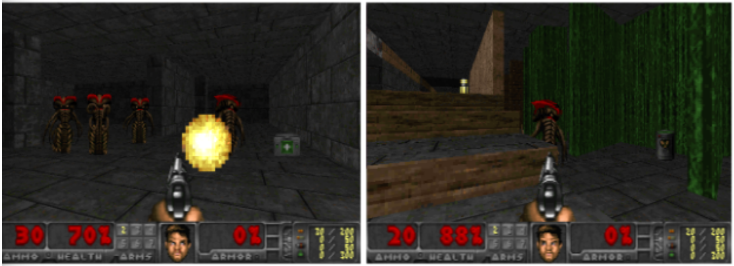
Key word assignment:



- Input: sequence of words $(w_1, \ldots, w_K)$
- Output: set of words $\{w_1, \ldots, w_L\}$ – setup can be multi-label classification or sequential outputs from a RNN
(Example: "Star Trek, RNN-generated nonsense, AI for fun")

classical examples

Reinforcement Learning:



D3: Battle D4: Battle 2

- Input: sequence of states (health, ammo, images of view)^K
- Output: next action (move left, fire, ...)
- Dosovitsky et al, ICLR 2017, `https://arxiv.org/pdf/1611.01779.pdf`

Visual Question answering, Image captioning:

credit: `visualqa.org`

- Input: Image (+ sequence of words for VQA)
- Output: set of words – setup can be multi-label classification or sequential outputs from a RNN

4 Key components of discriminative machine learning

What does one need for discriminative machine learning in general?

- IO Model for Input space, Model for output space
- M define a prediction model (deep neural network??)
- L define a loss function to measure difference: prediction of the model versus ground truth
- A define an algorithm for updating model parameters

High Level Engineering steps

iterate(!) the following

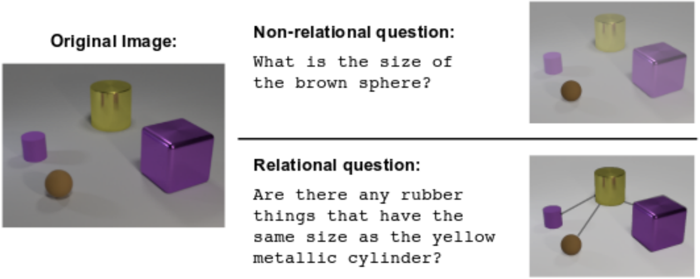
0. re-do steps 1 to 6!!
1. state problem
2. think how human has to interact with the final solution
3. re-collect data, split your data reasonably
4. devise model
5. measure performance
6. create/update a human-interactive demonstrator

Low Level Engineering steps

- define a data loader
  - loads a minibatch of data (input, ground truth) from your data source (NN-specific)
  - data used to compute value of loss function: prediction(over input) vs ground truth
- define data augmentation steps (NN-specific)
- hyper-parameter search (NN-specific, learning rate, batch-size, how many samples of what type in every batch)
- run your variant of SGD (SGD, Adam, ...) to optimize model parameters

The strengths of Deep learning

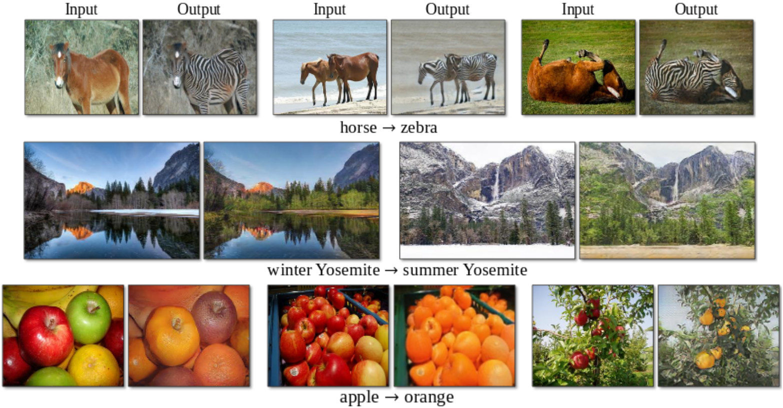
Good prediction performance **if** there is sufficient data for training.



Santoro et al, 2017:
`https://arxiv.org/pdf/1706.01427.pdf`

The strengths of Deep learning

Good prediction performance if there is sufficient data for training.



Zhu et al, 2017: `https://arxiv.org/pdf/1703.10593.pdf`

The fallacies of Deep learning

Poor performance for **small sample sizes**, classical feature engineering with decision forests or SVMs often much better.

Alias	Network	# Parameters
alexnet	AlexNet	61,100,840
densenet121	DenseNet-121	8,062,504
densenet161	DenseNet-161	28,900,936
densenet169	DenseNet-169	14,307,880
densenet201	DenseNet-201	20,242,984
inceptionv3	Inception V3 299x299	23,869,000

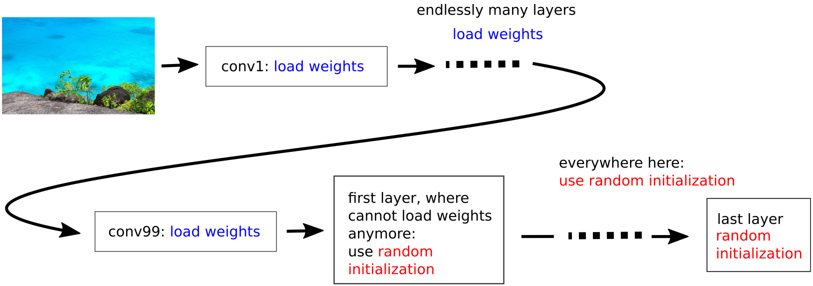
credit: MXNet Model zoo

Large number of parameters ($d = 10^6$)
$+$ small number of data ($n = 10^3$)
$=$ overfitting

Solution – transfer learning:
initialize network (as much as possible) with weights from another task trained over very large sample size

Transfer learning

Not enough training data? Load weights from a model trained on a task with many more samples (for all layers from the input until the first blocker)



Transfer learning

recurrent NN for generating an answer: from scratch

it is light blue

recurrent NN for question understanding: from scratch

what is color does the water have?

conv3: load weights

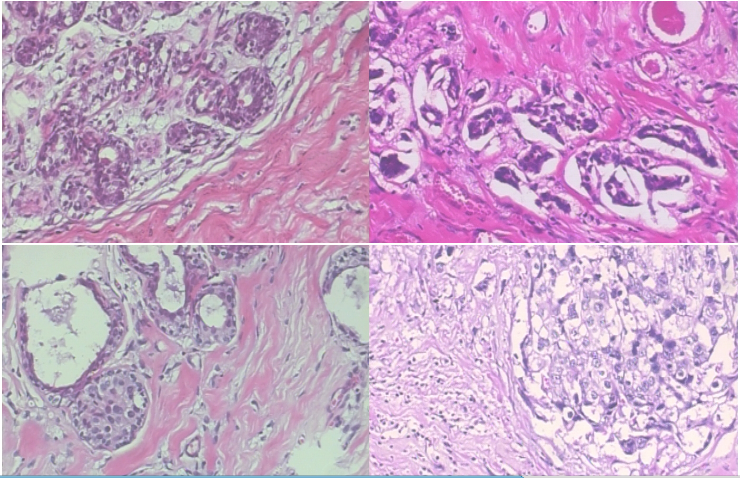
conv2: load weights

conv1: load weights

Transfer learning

Benign vs Malignant Breast cancer detection, $n \approx 1300$ for training.

benign – malignant



Transfer learning

Benign vs Malignant Breast cancer detection, $n \approx 1300$ for training.

benign – malignant

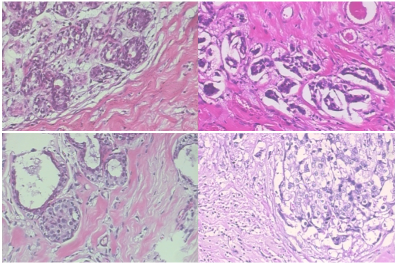


TABLE V
AVERAGE PATIENT RATE OVER MULTI-RESOLUTION

Patient rate	Full image	128crop	64crop
GoogLeNet	95.00±3.64	92.85±4.76	92.85±4.97
Caffenet	94.29±4.84	92.14±4.74	92.86±5.05
Resnet-50	95.00±3.64	94.29±6.22	91.42±5.80

Sun et al, ICBDA 2017 `https://ieeexplore.ieee.org/abstract/document/8284105/`

Transfer learning

Tuberculosis Type classification, ImageCLEF 2017 Tuberculosis Challenge

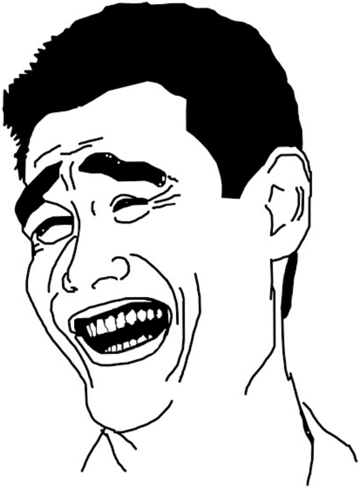
Task 2 - Tuberculosis type classification

Group Name	Run	Run Type	Kappa	ACC	Rank
SGEast	TBT_resnet_full.txt	Not applicable	0.2438	0.4033	1
SGEast	TBT_LSTM_17_wcrop.txt	Not applicable	0.2374	0.3900	2
MEDGIFT UPB	TBT_T_GNet.txt	Automatic	0.2329	0.3867	3
SGEast	TBT_LSTM_13_wcrop.txt	Not applicable	0.2291	0.3833	4
Image Processing	TBT-testSet-label-Apr26-XGao-1.txt	Automatic	0.2187	0.4067	5
SGEast	TBT_LSTM_46_wcrop.txt	Not applicable	0.2174	0.3900	6
UIIP	TBT_iiggad_PCA_RF_run_1.txt	Automatic	0.1956	0.3900	7
MEDGIFT UPB	TBT_TEST_RUN_2_GoogleNet_10crops_at_different_scales_.txt	Automatic	0.1900	0.3733	8
SGEast	TBT_resnet_partial.txt	Not applicable	0.1729	0.3567	9
MedGIFT	TBT_Top1_correct.csv	Automatic	0.1623	0.3600	10
SGEast	TBT_LSTM_25_wcrop.txt	Not applicable	0.1548	0.3400	11
MedGIFT	TBT_submitted_topBest3_correct.csv	Automatic	0.1548	0.3500	12
BatmanLab	TBT_SuperVx_Hist_FHOG_lr_0.414000.csv	Automatic	0.1533	0.3433	13

https://www.imageclef.org/2017/tuberculosis

Autonomous cars in Bangkok 2020 ?

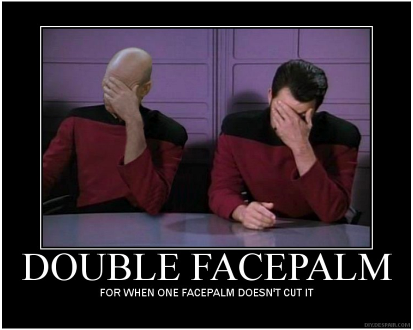
Autonomous cars in Bangkok 2020 ?



parked maybe

Humans make mistakes all the time, AI is perfect. Train, then deploy your algorithm (chatbot, machine translator) and all will work perfect.

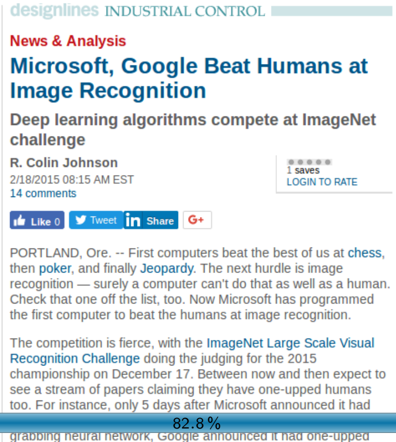
Humans make mistakes all the time, AI is perfect. Train, then deploy your algorithm (chatbot, machine translator) and all will work perfect.



...

Humans make mistakes all the time, AI is perfect. Train, then deploy your algorithm (chatbot, machine translator) and all will work perfect.

https://www.eetimes.com/document.asp?doc_id=1325712



designlines INDUSTRIAL CONTROL

News & Analysis

Microsoft, Google Beat Humans at Image Recognition

Deep learning algorithms compete at ImageNet challenge

R. Colin Johnson
2/18/2015 08:15 AM EST
14 comments

1 saves
LOGIN TO RATE

Like 0 Tweet Share G+

PORTLAND, Ore. -- First computers beat the best of us at chess, then poker, and finally Jeopardy. The next hurdle is image recognition — surely a computer can't do that as well as a human. Check that one off the list, too. Now Microsoft has programmed the first computer to beat the humans at image recognition.

The competition is fierce, with the ImageNet Large Scale Visual Recognition Challenge doing the judging for the 2015 championship on December 17. Between now and then expect to see a stream of papers claiming they have one-upped humans too. For instance, only 5 days after Microsoft announced it had

grabbing neural network, Google announced it had one-upped

82.8 %

Humans make mistakes all the time, AI is perfect. Train, then deploy your algorithm (chatbot, machine translator) and all will work perfect.

1. 1-crop validation error on ImageNet (center 224x224 crop from resized image with shorter side=256):

model	top-1	top-5
VGG-16	28.5%	9.9%
ResNet-50	24.7%	7.8%
ResNet-101	23.6%	7.1%
ResNet-152	23.0%	6.7%

2. 10-crop validation error on ImageNet (averaging softmax scores of 10 224x224 crops from resized image with shorter side=256), the same as those in the paper:

model	top-1	top-5
ResNet-50	22.9%	6.7%
ResNet-101	21.8%	6.1%
ResNet-152	21.4%	5.7%

What were among the first commercially machine learning applications???

Check writing recognition

https://www.computer.org/cms/Computer.org/ComputingNow/computingthen/atty/1993/ATTY-1993-1-ERMA.pdf

Why the system was deployed in such a sensitive context despite it made recognition errors?

What were among the first commercially machine learning applications???

Mail Sorting by zip code/ adress reading (late 90s, Sangur Srihari)

Use case: read adress, automatically distribute

- 1997: 30% of all letters,
- 2001: 75% of all letters,
- 2004: 88% of all letters

USD 100 Mio saved in 1997 only

15 years of research & development

What were among the first commercially machine learning applications???

Mail Sorting by zip code/ adress reading (late 90s, Sangur Srihari)

Use case: read adress, automatically distribute

- 1997: 30% of all letters,
- 2001: 75% of all letters,
- 2004: 88% of all letters

USD 100 Mio saved in 1997 only

15 years of research & development

Why mail sorting was commercially employed despite it can sort only a part of the letters?

What were the first commercially machine learning applications???

Why speech to text is deployed in practice and chatbots not in anything that matters?

Why digital pathologists are deployed in test mode only in hospitals in China and India?

Speech to text
How errors were dealt with??

The nonsense hype about Deep learning

Humans make mistakes all the time, AI is perfect. Train, then deploy your algorithm (chatbot, machine translator) and all will work perfect.

Humans **audit** humans for mistakes / **quality control** in production.

Deep learning helps to make money if ...

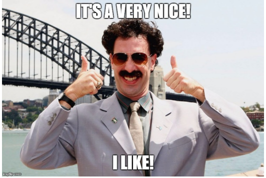
- Human in the loop to catch errors - human computer interface design
- Time/cost for correcting errors from algorithm is less than doing the task without AI
- any deployment of machine learning solutions needs to have a mechanism to deal with its errors

The nonsense hype about Deep learning

Why DL for machine translation is harder to adopt? ... Correction of errors ? Why DL for cancer detection is harder to adopt? ... but useful for therapy prediction!
https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/

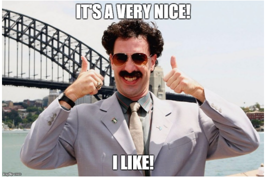
Changes in the work flow due to deep learning

Feature engineering is dead. Learn your features from data. No need for tuning of features by hand.

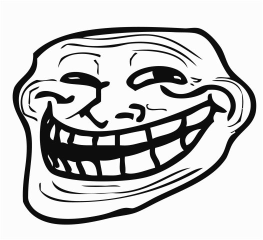


Changes in the work flow due to deep learning

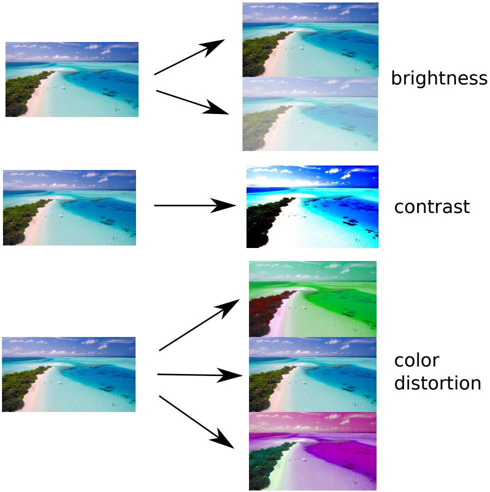
Feature engineering is dead. Learn your features from data. No need for tuning of features by hand.



Long live data augmentation engineering
Long live hyperparameter search over grids of 30 different parameters.

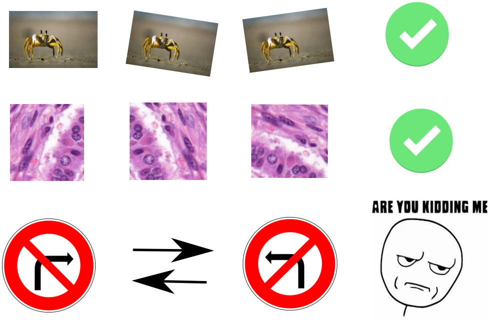


Data Augmentation Engineering



This are 6 parameters here already for hyperparameter search here

Data Augmentation Engineering



Many hyperparameters +

- batch size
- minibatch structure
- initial learning rate
- learning rate decay
- optimizer (SGD, momentum, ADAM)

Hyperparameter search in $>$ 50-dimensional spaces