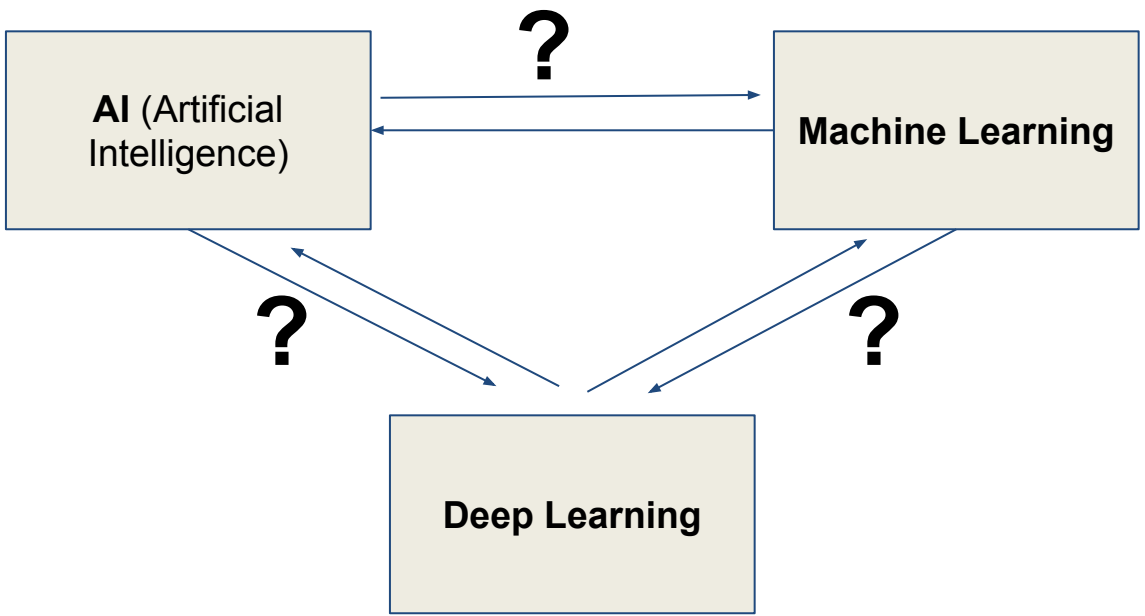
Deep learning for biologists

A practical and theoretical introduction

Filippo Biscarini
Senior Scientist
CNR, Milan (Italy)

Nelson Nazzicari
Research fellow
CREA, Lodi (Italy)

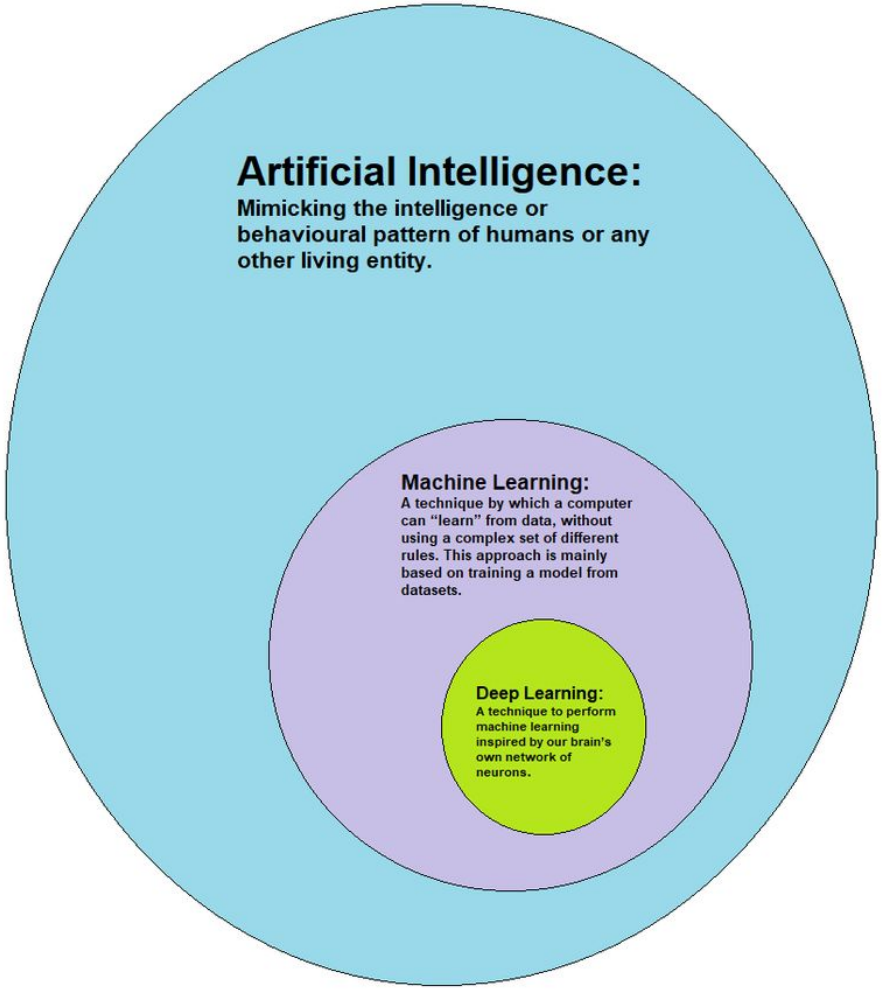
What is deep learning?



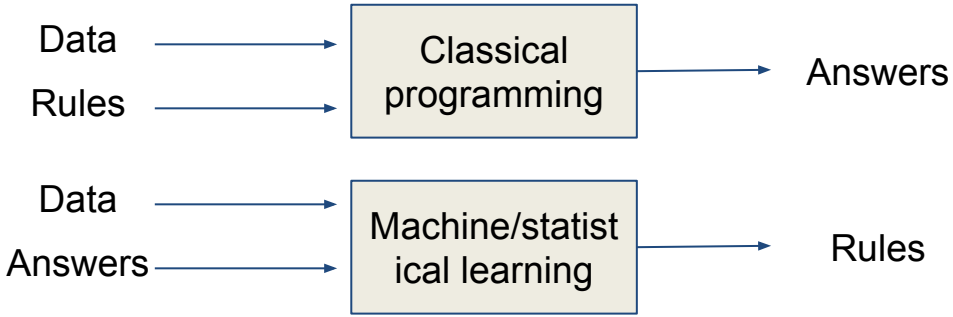
What is deep learning?

AI >> ML >> DL

Source:
https://en.wikipedia.org/wiki/Deep_learning#/media/File:AI-ML-DL.png



What is (deep) learning?



- (machine; statistical) learning
  - tune a **mathematical model** using some **training data** to **make predictions** on unknown, new data
  - a machine/statistical learning model is **trained** rather than explicitly programmed

What is (deep) learning?

(machine; statistical) learning

1. Input data (e.g. sound recordings, images)
2. Output examples (e.g. sound transcripts, image-tags)
3. Performance measure: how well is the algorithm working → adjustment steps → **learning**

You can do (statistical) learning in your head!

- The first edition of this course gets 10 students
- The second edition gets 20 students
- The third edition gets 40 students
- The fourth edition gets 80 students
- How many students in the sixth edition?

You can do (statistical) learning in your head!

TRAINING DATA

- The first edition of this course gets 10 students
- The second edition gets 20 students
- The third edition gets 40 students
- The fourth edition gets 80 students
- How many students in the sixth edition?

NEW, UNKNOWN DATA

STUD = 10 x 2 exp (YEAR - 1)

MATHEMATICAL MODEL

STUDENTS IN SIXTH EDITION = 320

PREDICTION

What is deep learning?

- (machine; statistical) learning
  - tune a **mathematical model** using some **training data** to **make predictions** on unknown, new data
  - «If you make a bunch of random changes to your program until it sort-of works, that's "hacky" and "bad coding practice". But if you do it really fast, it's "machine learning"»
- Deep learning
  - **Neural networks**, a mathematical model “inspired” by biology
  - Artificial Neural Networks has been around for ~80 years ...
  - ...but became “deep” (i.e. with many layer) in the last ten years

Why “deep”?

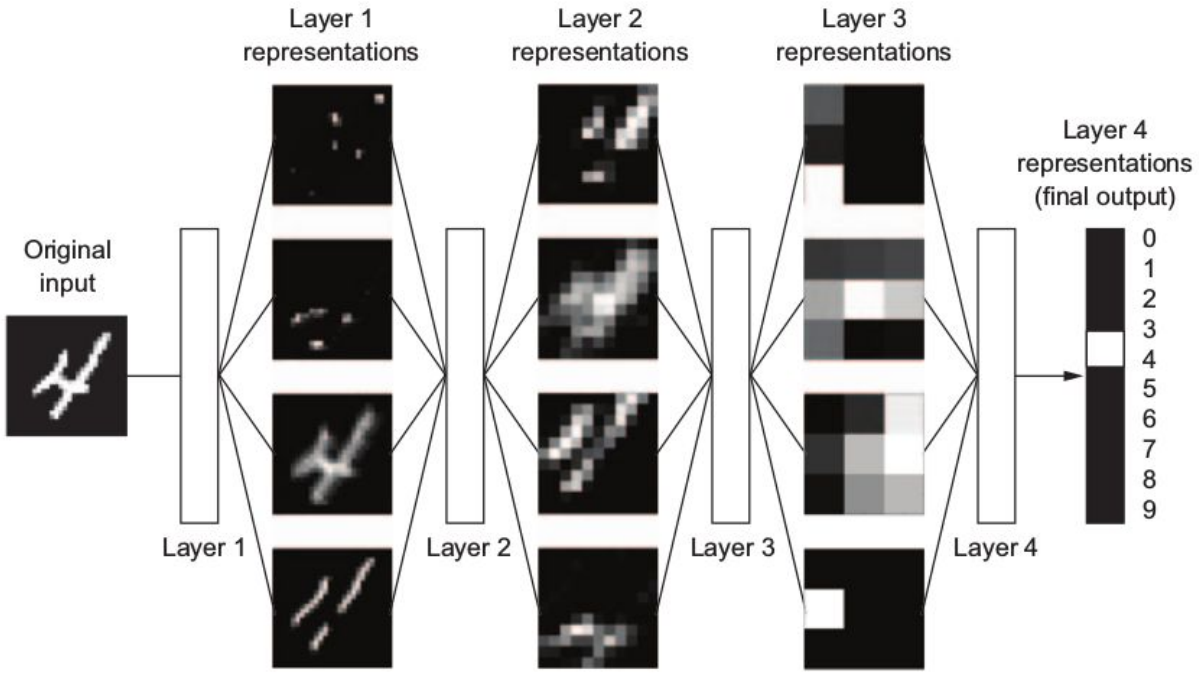
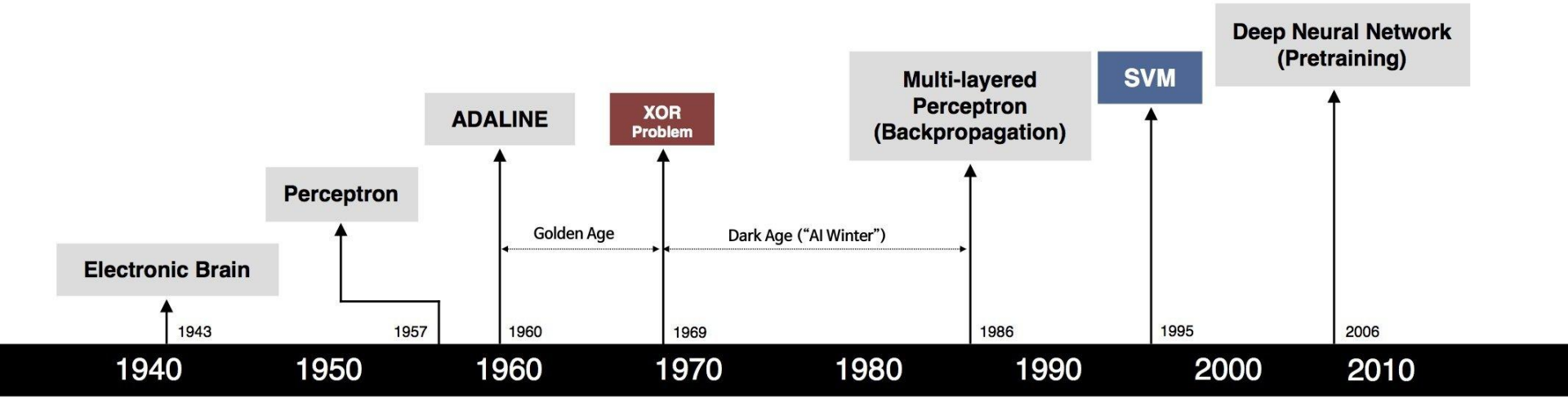
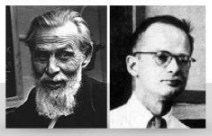


Figure 1.6 Deep representations learned by a digit-classification model

Source: François Chollet and J.J. Allaire “Deep learning with R” (2018)

A little history of Neural Networks





S. McCulloch – W. Pitts

F. Rosenblatt

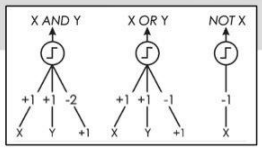
B. Widrow – M. Hoff

M. Minsky – S. Papert

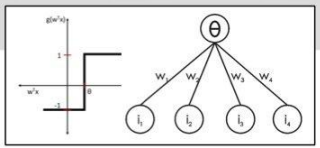
D. Rumelhart – G. Hinton – R. Wiliams

V. Vapnik – C. Cortes

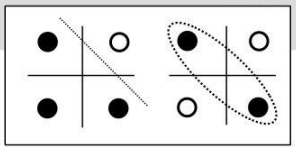
G. Hinton – S. Ruslan



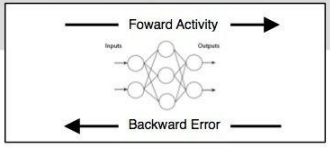
- Adjustable Weights
- Weights are not Learned



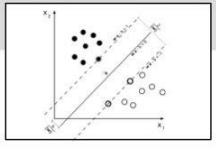
- Learnable Weights and Threshold



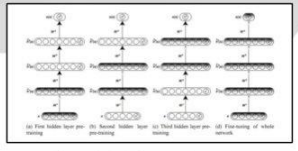
- XOR Problem



- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting



- Limitations of learning prior knowledge
- Kernel function: Human Intervention



- Hierarchical feature Learning

Credits: Andrew Beam, *Department of Biomedical Informatics @ Harvard Medical School*

[REF] History of deep learning

- McCulloch and Pitts, *A logical calculus of the ideas immanent in nervous activity*, 1943. https://link.springer.com/article/10.1007/BF02478259
- Wang and Raj, *On the Origin of Deep Learning*, 2017 https://arxiv.org/pdf/1702.07800.pdf
- Andrew Beam, Deep Learning 101 - Part 1: History and Background, 2017 https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html

Why now?

Innovations in:

- Hardware
- Big Data:
- Algorithms
- Infrastructure

Why now?

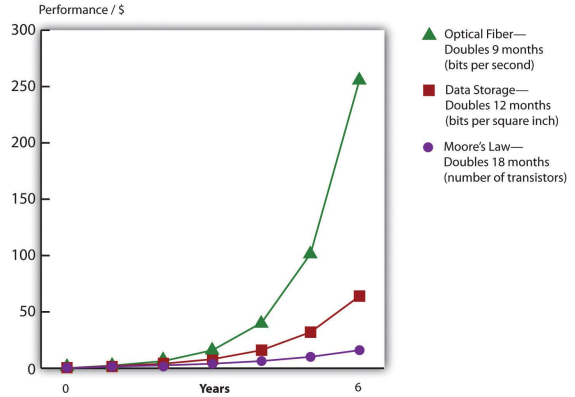
Innovations in:

- Hardware:
  - CPUs now 5,000 times faster than 25 years ago
  - GPUs (thanks to videogames!)
  - TPUs (Tensor Processing Units: designed specifically for deep learning)

Why now?

Innovations in:

- Big Data:
  - Internet: search engines, social media, image datasets etc.
  - IoT: sensors, interconnected devices
  - Data storage (Moore's law)



https://2012books.lardbucket.org/books/getting-the-most-out-of-information-systems-v1.2/s09-moore-s-law-fast-cheap-computi.html

Why now?

Innovations in:

- Algorithms:
  - backpropagation/gradient propagation (efficient ways to solve deep learning models)
  - better activation functions (e.g. ReLU)
  - better optimizers (e.g. RMSProp and ADAM)

Why now?

Innovations in:

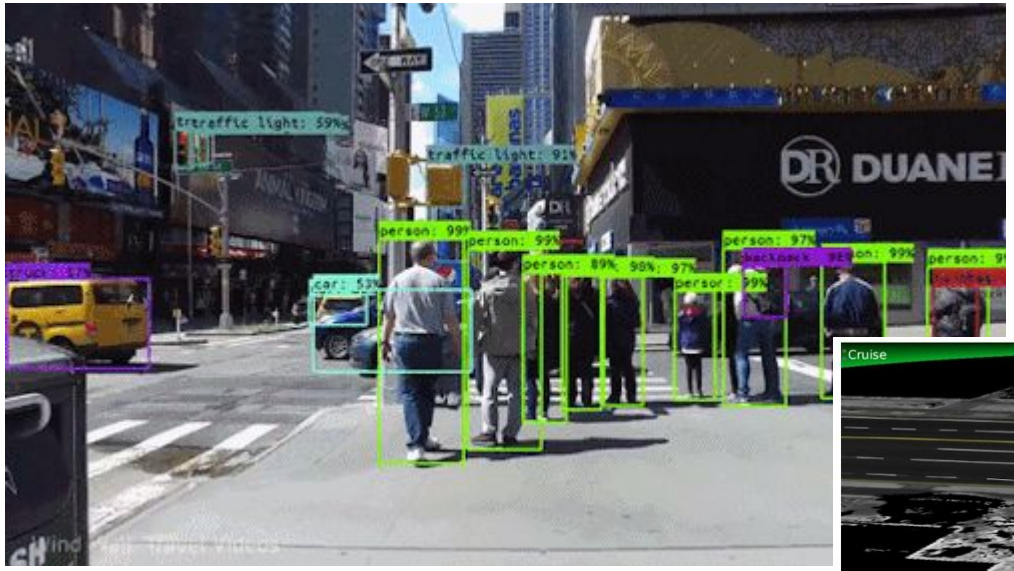
- Infrastructure:
  - scaling-up of computation frameworks (e.g. cloud computing)
  - distributed computing (and storage)
  - programming frameworks

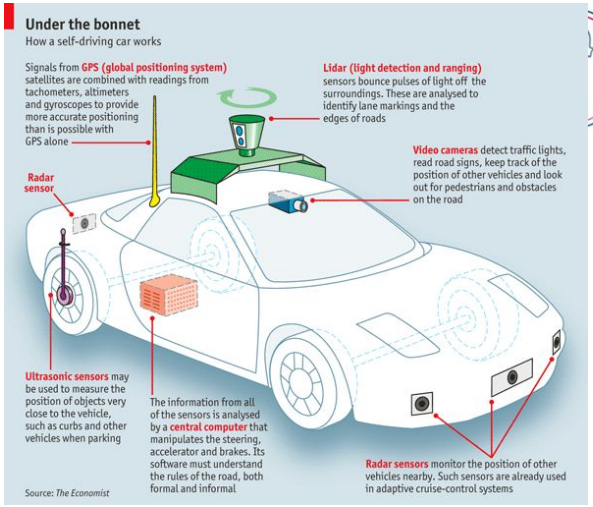
Deep learning: a matter of

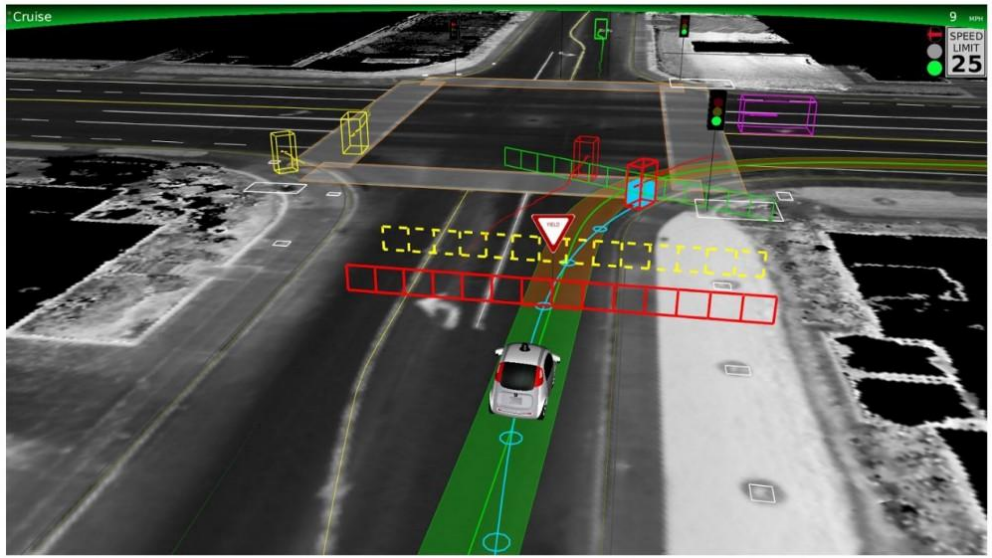
- Scale
  - Available hardware (GPU: thanks gamers)
  - Available big data (e.g. massive databases of *labeled* images)
  - Available infrastructure
  - Available pre-trained model (transfer learning)
- Theoretical breakthrough
  - ReLU activation functions
  - Back propagation
  - Gradient descent and other solvers

State of the art: nerd stuff

Self driving cars

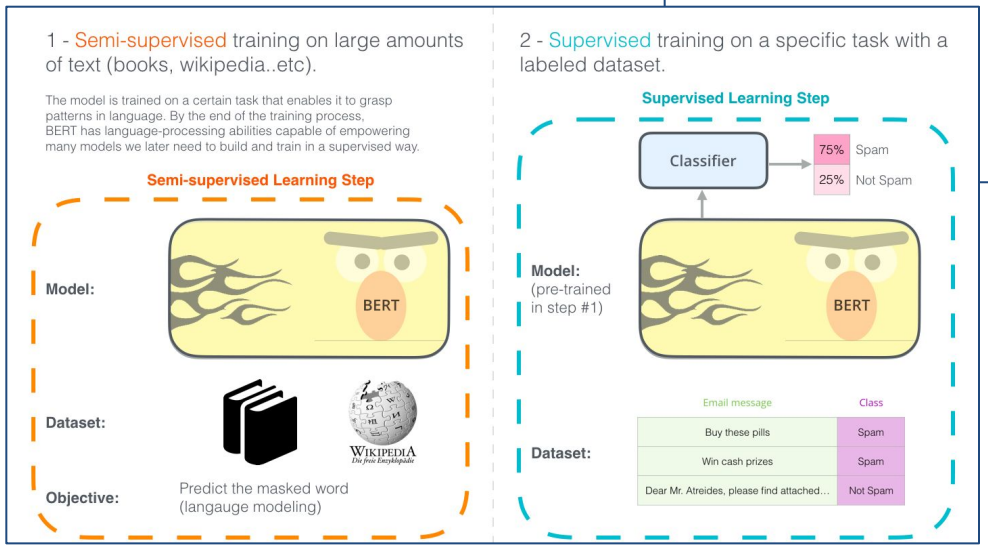


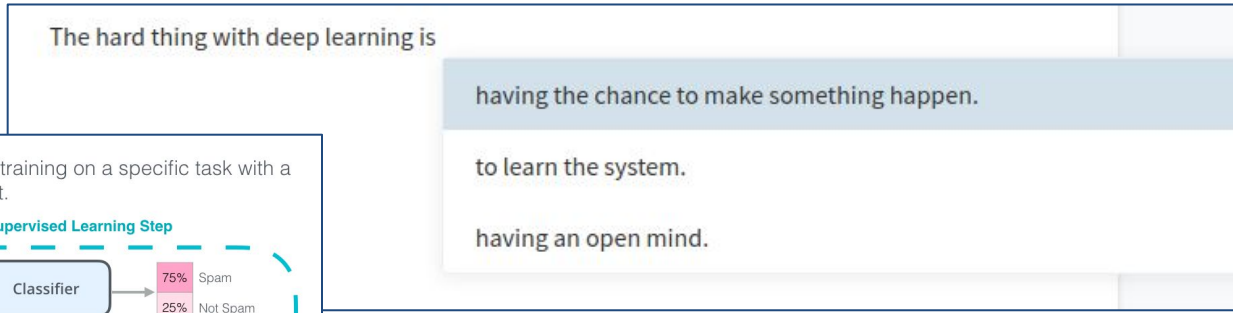


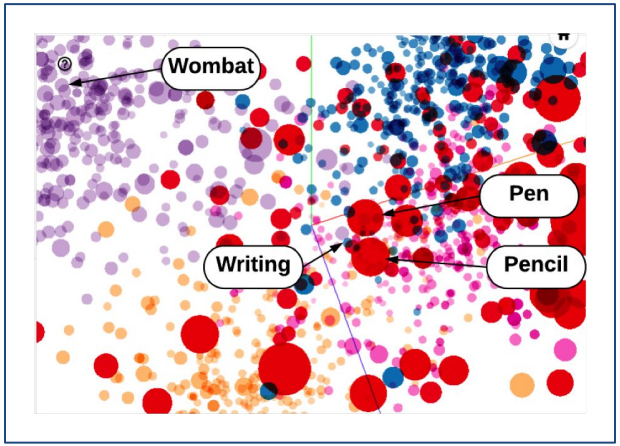


Credits:
https://medium.com/@feiqi9047/the-data-science-behind-self-driving-cars-eb7d0579c80b

Natural Language Processing (NLP)

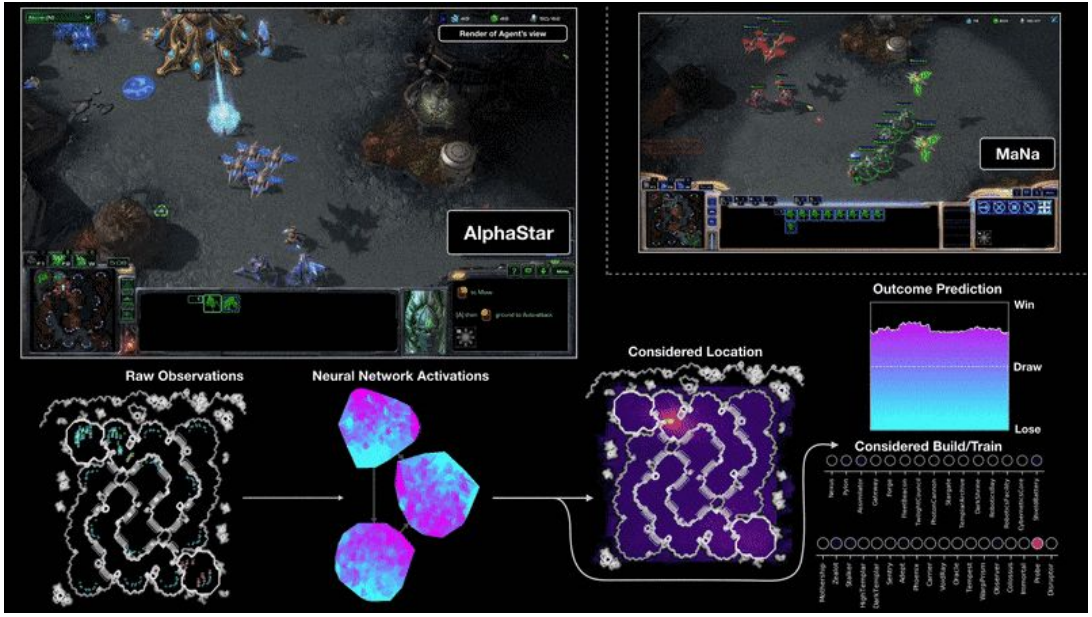




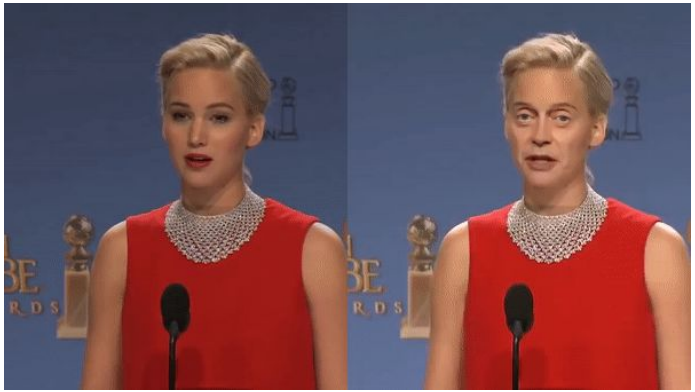


Credits:
https://mc.ai/whats-new-in-deep-learning-research-facebook-meta-embeddings-allow-nlp-models-to-choose-their/
http://jalammar.github.io/illustrated-bert/

Games & Videogames



Generating believable videos (deepfake)



Credits:
https://www.gizmodo.co.uk/2018/06/deepfake-videos-are-getting-impossibly-good/
https://www.sciencealert.com/samsung-s-ai-can-now-generate-talking-heads-from-a-single-image
YouTube @ birbfakes

And many, many more...

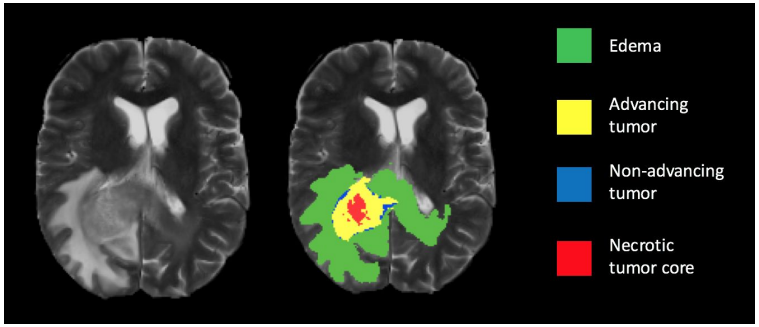
- News Aggregation and Fraud News Detection
- Virtual Assistants
- Entertainment
- Visual Recognition
- Fraud Detection
- Healthcare
- Personalisations
- Detecting Developmental Delay in Children
- Colourisation of Black and White images
- Adding sounds to silent movies
- Automatic Machine Translation
- Automatic Handwriting Generation
- Language Translations
- Pixel Restoration
- Photo Descriptions
- Demographic and Election Predictions
- Deep Dreaming

[REF] Deep learning applications

- State of the art for self driving cars: https://neurohive.io/en/state-of-the-art/self-driving-cars/ and https://www.bloomberg.com/features/2020-self-driving-car-race/
- Waymo blog on their autonomous vehicles: https://blog.waymo.com/
- Updated repository of NPL state of the art https://github.com/sebastianruder/NLP-progress
- NPL transformers: https://github.com/huggingface/transformers
- NPL BERT https://medium.com/analytics-vidhya/text-classification-with-bert-using-transformers-for-long-text-inputs-f54833994dfd
- Deepmind Agent on mastering Atari Games https://deepmind.com/blog/article/Agent57-Outperforming-the-human-Atari-benchmark
- Deepmind AlphaGO on mastering the game of game of GO https://deepmind.com/research/case-studies/alphago-the-story-so-far
- Deepming Alphastar on mastering Real Time Strategy videogame Starcraft II https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii
- Generating living portraits from few shots https://arxiv.org/abs/1905.08233
- Deep video portraits original paper - ACM TOG 2018 conference https://dl.acm.org/doi/abs/10.1145/3197517.3201283

State of the art/2: wet stuff

Brain Tumor Segmentation



RANK	METHOD	DICE SCORE	EXTRA TRAINING DATA	PAPER	CODE	RESULT	YEAR
1	OM-Net + CGAp	87%	×	One-pass Multi-task Networks with Cross-task Guided Attention for Brain Tumor Segmentation			2019
2	CNN + 3D filters	85%	✓	CNN-based Segmentation of Medical Imaging Data			2017

Images credit: Brain Tumor Segmentation with Deep Neural Networks https://github.com/naldeborgh7575/brain_segmentation

Classification of medical images

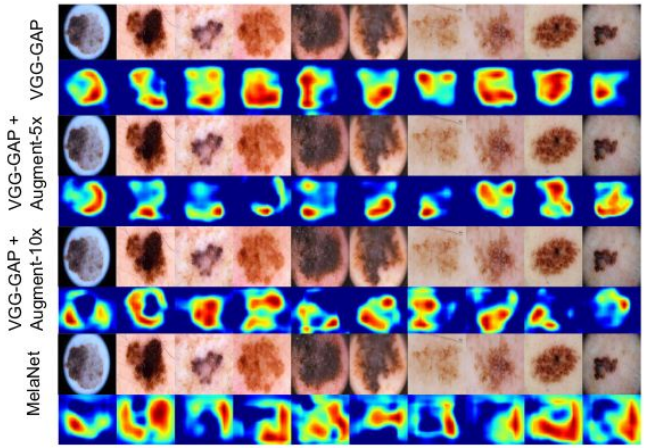


Figure 8: Grad-CAM heat maps for the correctly classified malignant cases by MelaNet and baseline methods.

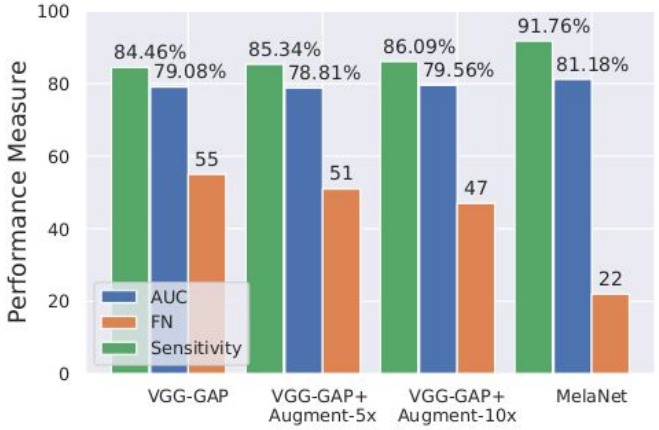


Figure 5: Classification performance of MelaNet and the baseline methods using AUC, FN and Sensitivity as evaluation metrics on the ISIC-2016 test set.

Source: Zunair and Hamza, 2020. Melanoma Detection using Adversarial Training and Deep Transfer Learning.

Drug discovery/optimization

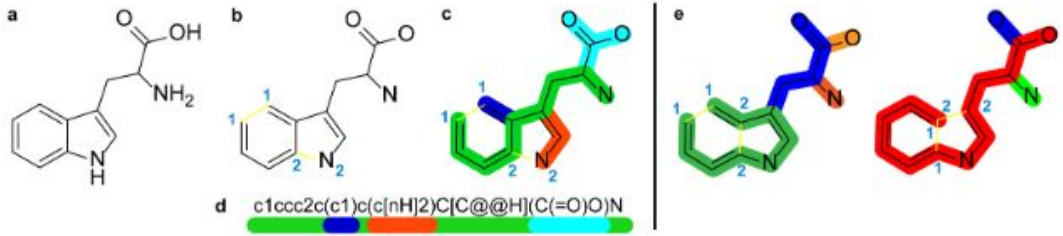
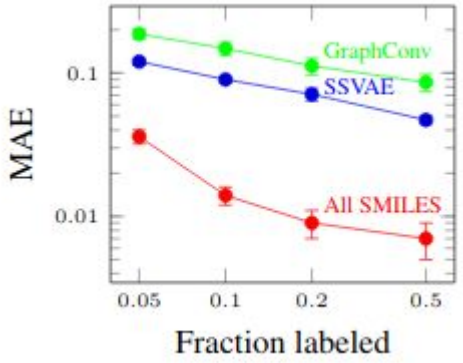
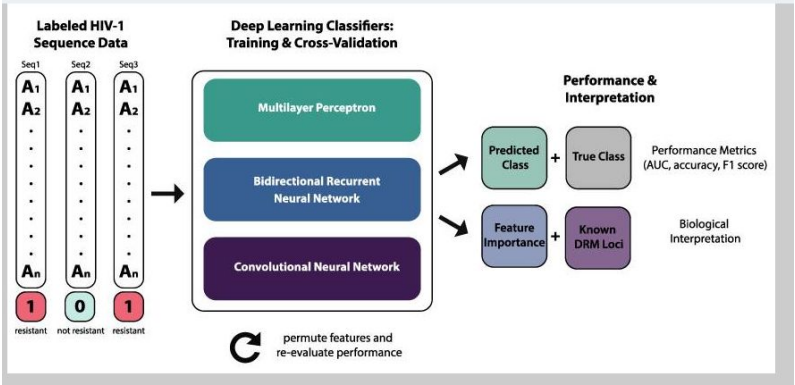


Figure 1: The molecular graph of the amino acid Tryptophan (a). To construct a SMILES string, all cycles are broken, forming a spanning tree (b); a depth-first traversal is selected (c); and this traversal is flattened (d). The beginning and end of intermediate branches in the traversal are denoted by (and) respective. The ends of broken cycles are indicated with matching digits. The full grammar is listed in Appendix D. A small set of SMILES strings can cover all paths through a molecule (e).



Source: Alperstein et al, 2019. All SMILES Variational Autoencoder

Drug resistance prediction



Viruses. 2020 May; 12(5): 560.
Published online 2020 May 19. doi: 10.3390/v12050560

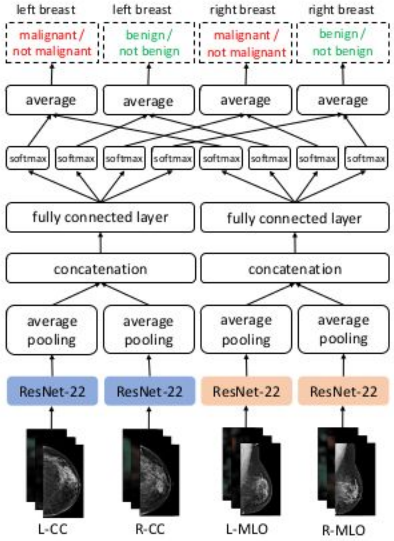
Drug Resistance Prediction Using Deep Learning Techniques on HIV-1 Sequence Data

Margaret C. Steiner,[1,*] Keylie M. Gibson,[1] and Keith A. Crandall[1,2]

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7290575/

And the slack channel (thanks Pleuni!)

Breast Cancer detection



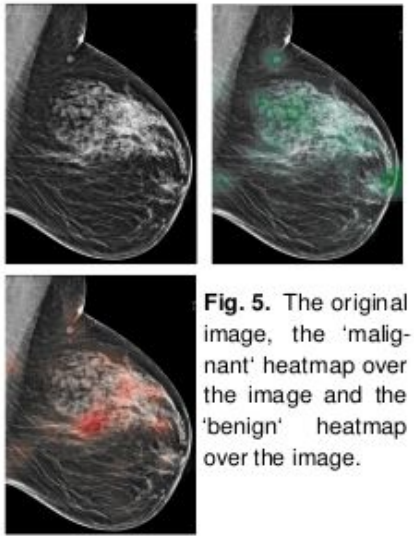
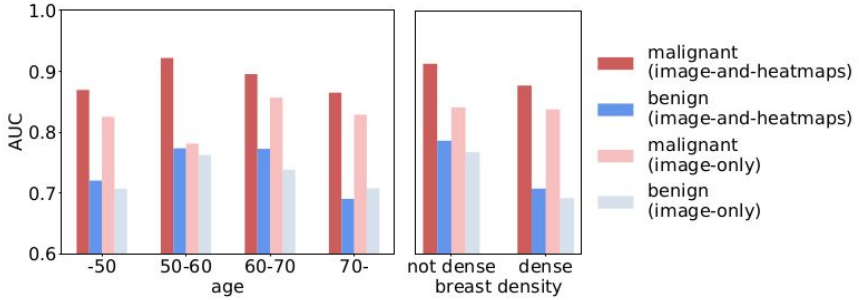
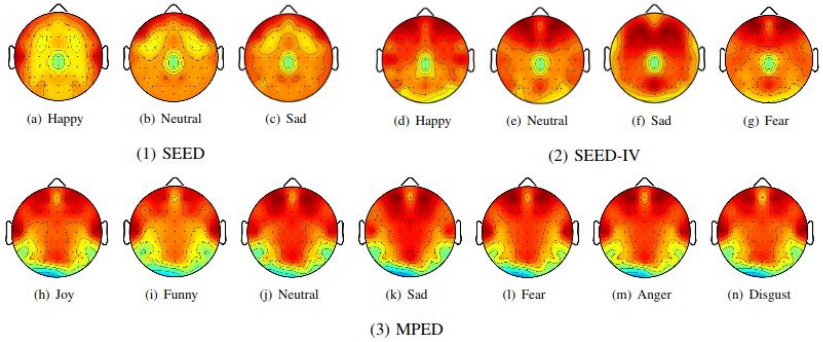


Fig. 5. The original image, the 'malignant' heatmap over the image and the 'benign' heatmap over the image.



Source: Wu et al, 2019. Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening

EEG interpretation/analysis



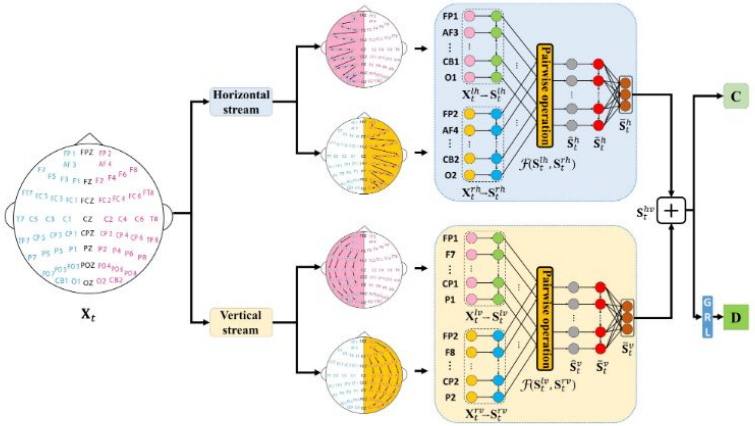


Fig. 1: The framework of BiHDM. BiHDM consists of four RNN modules to capture each hemispheric EEG electrodes' information from horizontal and vertical streams. Then all the electrodes' data representations interact and construct the final vector for the classifier and discriminator.

Source: Li et al., 2019. A Novel Bi-hemispheric Discrepancy Model for EEG Emotion Recognition

[REF] Deep learning state of the art

- Continuously updated applications of DL, divided by topic (Computer Vision, NPL, medical…) https://paperswithcode.com/sota
- MIT Deep learning state of the art 2020 seminar and course https://deeplearning.mit.edu/
- Deep Learning Papers Reading Roadmap https://github.com/floodsung/Deep-Learning-Papers-Reading-Roadmap

Keywords

- Classification
- Regression
- Data representation
- Model/Method/Algorithm/Software
- Data regularization
- Accuracy, error, correlation
- Overfitting
- Training set, test set