

Deep Learning

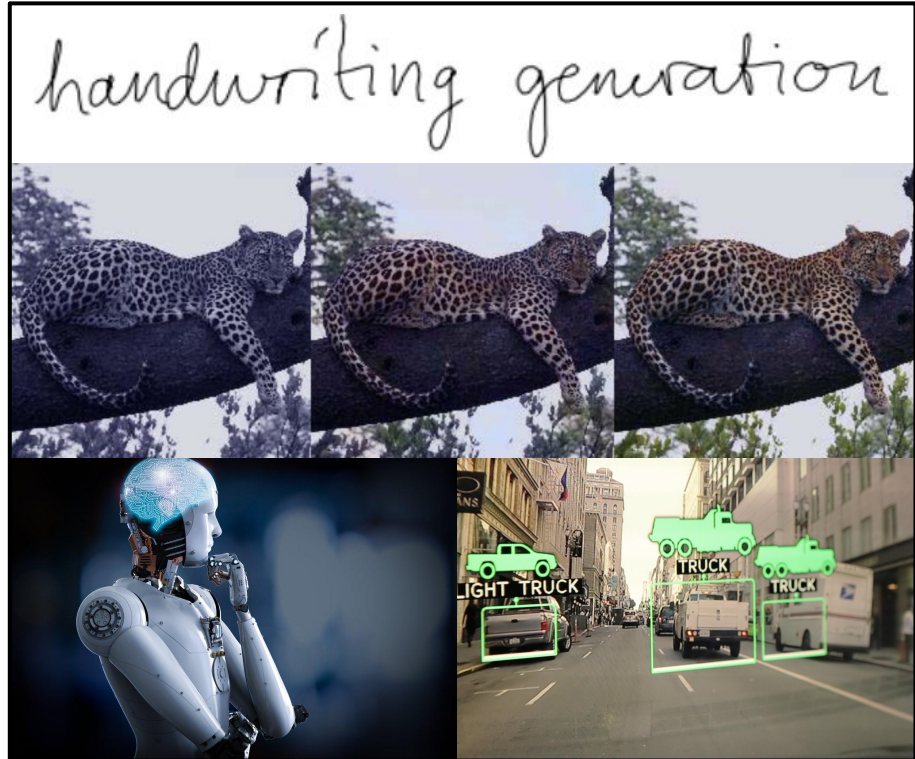
in Biomedical Data Science



Hussein Mohsen
CBB752 Spring 2019

Applications

- Image Color Restoration
- Image Captioning
- Object Classification
- Robotics
- Self-driving Cars
- Machine Translation
- Speech Recognition
- Text Generation
- Handwriting Generation
- Music Composition
- & more



Applications



OpenAI

@OpenAI

Follow

We've trained an unsupervised language model that can generate coherent paragraphs and perform rudimentary reading comprehension, machine translation, question answering, and summarization — all without task-specific training:

[blog.openai.com/better-language...](https://blog.openai.com/better-language-model/)



We've trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state of the art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization —



0:38 312K Views

9:03 AM - 14 Feb 2019

2,782 Retweets 6,603 Likes



DeepMind's new AI just beat top human pro-gamers at Starcraft II for the first time

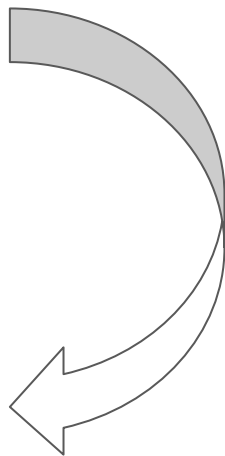


[\[https://thispersondoesnotexist.com\]](https://thispersondoesnotexist.com)

One-Hot Encoding of Sequences

C C A A T T G T T T T G

	C	C	A	A	T	T	G	T	T	T	T	G
A	0	0	1	1	0	0	0	0	0	0	0	0
C	1	1	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	1	1	0	1	1	1	1	0
G	0	0	0	0	0	0	1	0	0	0	0	1



Applications

Biomedical Data Science

Fast and Deep
with predic

Sepp Hochberg
Pietro Di Lorenzo

Bioinformatics
Bioinformatics

<https://doi.org/10.1093/bioinformatics/btw255>

Published

Split View

Abstract

Motivational
classification
first three
are the
protein
sequences

2007

Abstract

Motivational
structure
contact
short of

Results

2012

RESEARCH ARTICLE
The human
genetic

Hui Y. Xiong^{1,2}
+ See all authors

Science 09 Jan 2015
Vol. 347, Issue 6201
DOI: 10.1126/science.1261913

Article

You are currently

Predicting

Most eukaryotic
uninterrupted
regulated
tissues. F
Xiong et al.

2015

Methodology

Deep
gene

Tao Zeng, R

BMC Bioinformatics

<https://doi.org/10.1186/s12859-015-0600-4>

Received: 2

Abstract

Background

Profiling
scales
structure
in situ
of the

2015

Predicting
DNA-protein
binding

Babak Ali

Nature Biotechnology

Abstract

Known
is essential
biological
show

2015

Convolutional neural network architectures for predicting DNA–protein binding

Haoyang Zeng, Matthew D. Edwards, Ge Liu, David K. Gifford ✉

Bioinformatics, Volume 32, Issue 12, 15 June 2016, Pages i121–i127,

<https://doi.org/10.1093/bioinformatics/btw255>

Published: 11 June 2016

Split View

Cite

Permissions

Share ▼

Abstract

Motivation: Convolutional neural networks (CNN) have outperformed conventional methods in modeling the sequence specificity of DNA–protein binding. Yet inappropriate CNN architectures can yield poorer performance than simpler models. Thus an in–depth understanding of how to match CNN architecture to a given task is needed to fully harness the power of CNNs for computational biology applications.

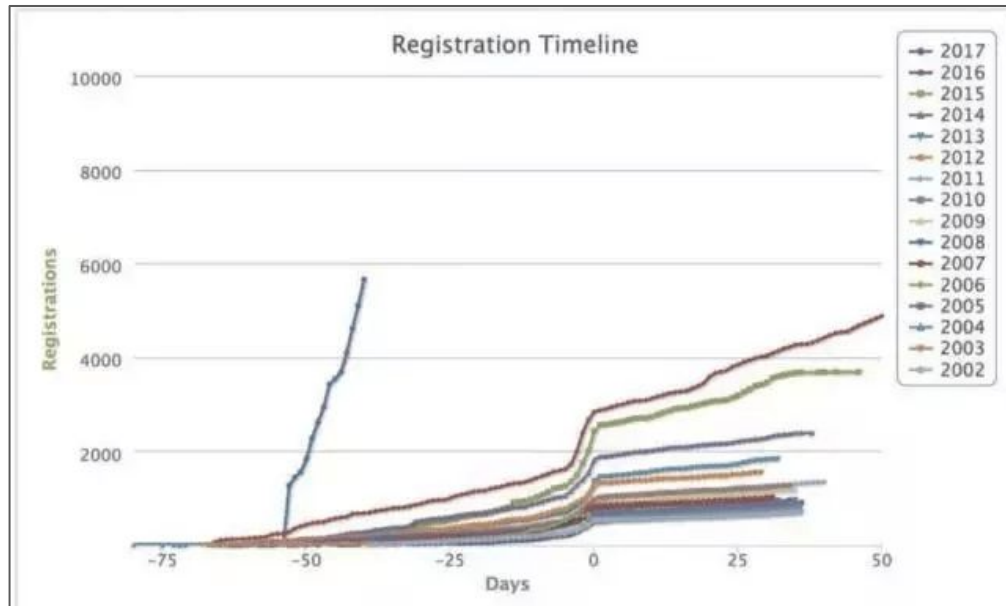
Results: We present a systematic exploration of CNN architectures for predicting DNA sequence binding using a large compendium of transcription

2016

Applications

Biomedical Data Science

- DeepBind, DeepFold, DeepSEA, DeepCNF, DeepVariant, DeepMotif, DeepLNC, DeepCpG, DeepCyTOF, DeepChrome, DeepWAS, DeepSplice, and other methods

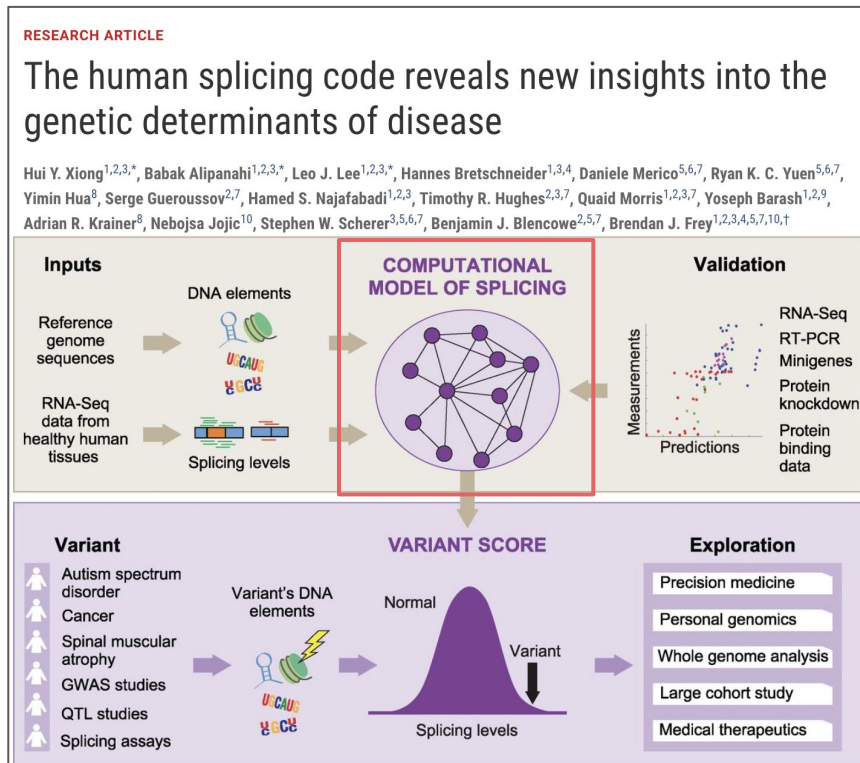


NeurIPS Registration Timeline [2002-17] @lxbrun

Applications

Biomedical Data Science

- Prediction of RNA Splicing
- 1393 sequence features as input
- Fully-connected neural networks
- RNA-seq from *Illumina Human Body Map 2.0* project
- Assessment of >650,000 SNPs effect on splicing regulation from the NCBI database of genetic variation (*dbSNP*) and Human Gene Mutation Database (*HGMD*)



Applications

Biomedical Data Science

- Skin cancer classification
- 129,450 clinical images from public

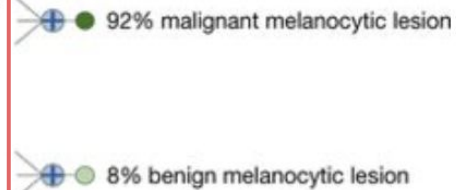
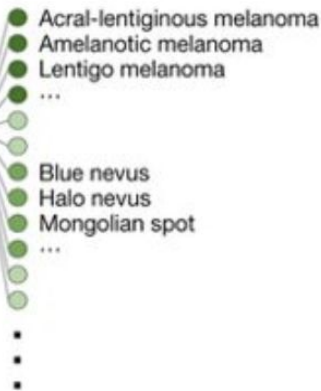
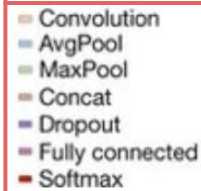
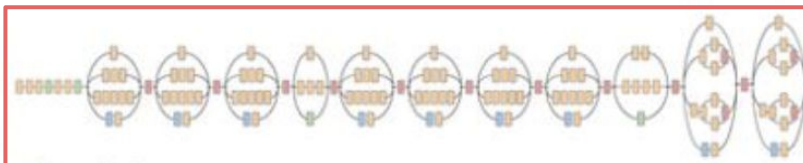
Dermatologist-level classification of skin

Skin lesion image

Deep convolutional neural network (Inception v3)

Training classes (757)

Inference classes (varies by task)



[2017]

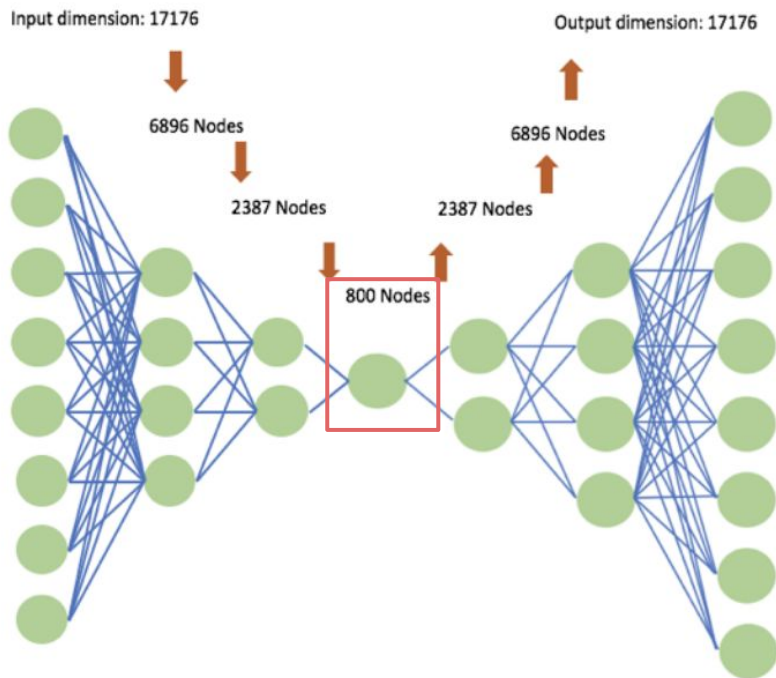
Applications

Biomedical Data Science

- Denoising Autoencoder (AE) for genomic data imputation
- $|I| = |O|$, bottleneck layer $|B| \ll |I|, |O|$
- 17,000+ genes, 269,000+ CpG sites
- Pan-Cancer RNA-seq and DNA Methylation data from The Cancer Genome Atlas (**TCGA**)

A deep learning framework for imputing missing values in
genomic data

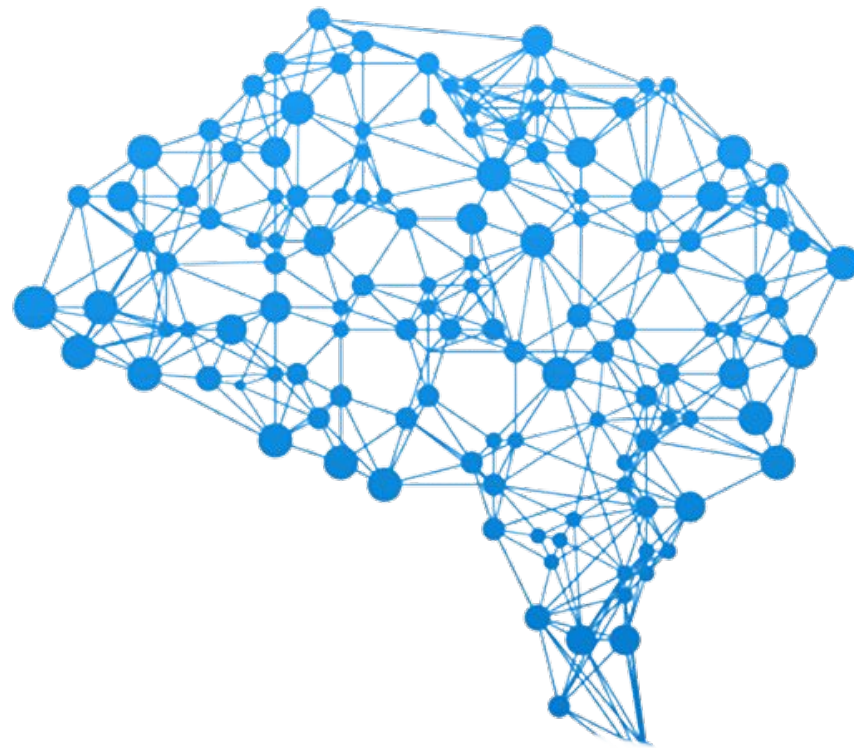
Yeping Lina Qiu^{1,2}, Hong Zheng², Olivier Gevaert^{2,3,*}



[2018]

Challenges

- **Interpretability**
 - Different from interpretability in Optimization
 - Weight matrix interpretation
 - Input feature prioritization
 - Biologically or clinically-driven
 - Earlier and later layers usually have more importance
- **Data availability**
- **Hyperparameter Optimization**
 - Set *before* training
 - Finding suboptimal combinations can be time consuming



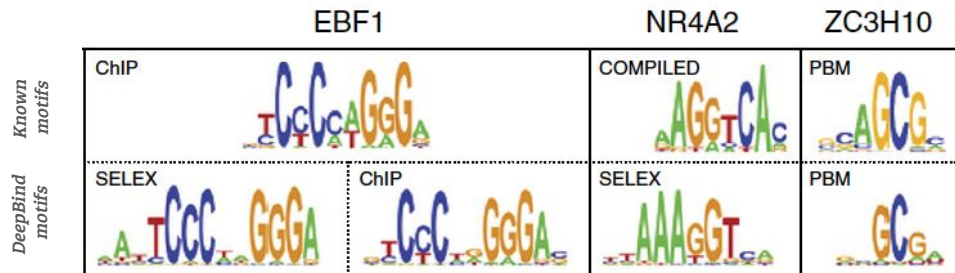
Interpretability

DeepBind

- Interpretability
 - Different from interpretability in Optimization
 - **Weight matrix interpretation**
 - Input feature prioritization
 - Biologically or clinically-driven
 - Earlier and later layers usually have more importance
- What would happen if the number of convolution filters is quite high?

Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning

Babak Alipanahi^{1,2,6}, Andrew Delong^{1,6}, Matthew T Weirauch³⁻⁵ & Brendan J Frey¹⁻³



Interpretability

SurvivalNet

- Interpretability
 - Different from interpretability in Optimization
 - Weight matrix interpretation
 - **Input feature prioritization**
 - Biologically or clinically-driven
 - Earlier and later layers usually have more importance
- 17,000-18,000 features dominated by gene expression and clinical variables
- Why would features be prioritized? What would potential downstream analyses be?

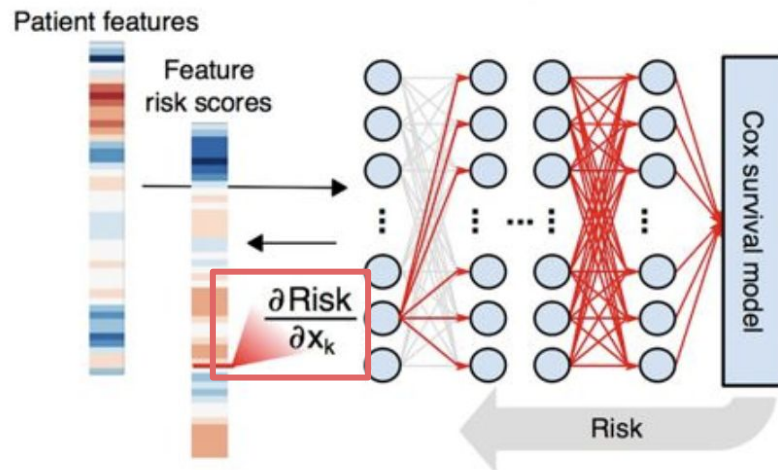
SCIENTIFIC REPORTS

OPEN

Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models

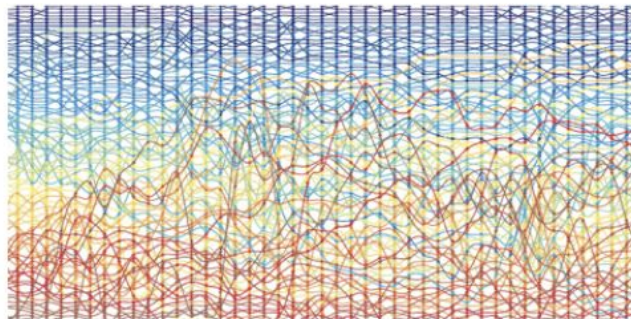
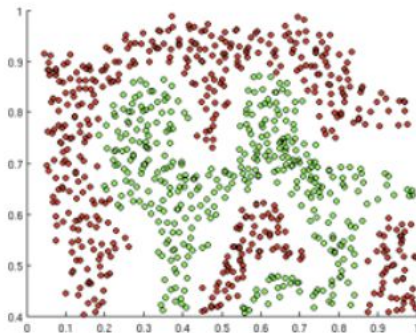
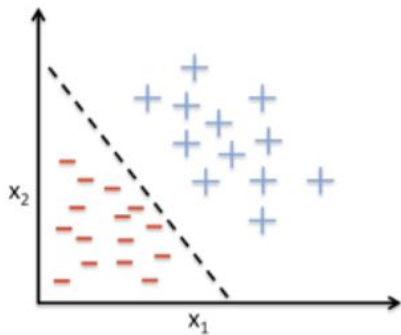
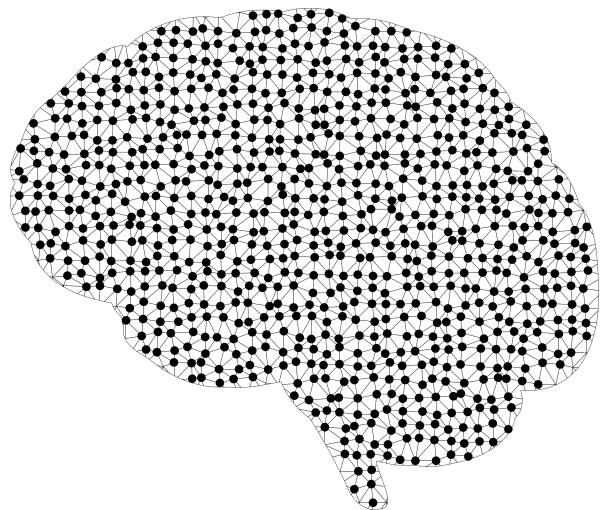
Received: 23 May 2017
Accepted: 30 August 2017
Published online: 15 September 2017

Safoora Yousefi¹, Fatemeh Amrollahi¹, Mohamed Amgad¹, Chengliang Dong², Joshua E. Lewis³, Congzheng Song⁴, David A. Gutman⁵, Sameer H. Halani⁶, Jose Enrique Velazquez Vega⁷, Daniel J. Brat^{7,8} & Lee A. D. Cooper^{1,3,8}



Notes

- Advances in computing
- Availability of large datasets
- Ability to detect highly complex, non-linear patterns
- Ability to predict many outputs at once
- Transfer learning to *transfer* patterns learned by one model to another



Resources

INTERFACE

rsif.royalsocietypublishing.org

Headline
review



Cite this article: Ching T *et al.* 2018
Opportunities and obstacles for deep learning
in biology and medicine. *J. R. Soc. Interface* **15**:
20170387.
<http://dx.doi.org/10.1098/rsif.2017.0387>

Opportunities and obstacles for deep learning in biology and medicine

Travers Ching^{1,†}, Daniel S. Himmelstein², Brett K. Beaulieu-Jones³,
Alexandr A. Kalinin⁴, Brian T. Do⁵, Gregory P. Way², Enrico Ferrero⁶,
Paul-Michael Agapow⁷, Michael Zietz², Michael M. Hoffman^{8,9,10}, Wei Xie¹¹,
Gail L. Rosen¹², Benjamin J. Lengerich¹³, Johnny Israeli¹⁴, Jack Lanchantin¹⁷,
Stephen Woloszynek¹², Anne E. Carpenter¹⁸, Avanti Shrikumar¹⁵, Jinbo Xu¹⁹,
Evan M. Cofer^{20,21}, Christopher A. Lavender²², Srinivas C. Turaga²³,
Amr M. Alexandari¹⁵, Zhiyong Lu²⁴, David J. Harris²⁵, Dave DeCaprio²⁶,
Yanjun Qi¹⁷, Anshul Kundaje^{15,16}, Yifan Peng²⁴, Laura K. Wiley²⁷,
Marwin H. S. Segler²⁸, Simina M. Boca²⁹, S. Joshua Swamidass³⁰,
Austin Huang³¹, Anthony Gitter^{32,33} and Casey S. Greene²

Resources

- Neural Networks and Deep Learning by Michael Neilson
[\[http://neuralnetworksanddeeplearning.com/\]](http://neuralnetworksanddeeplearning.com/)
- Deep Learning by Ian Goodfellow and Yoshua Bengio and Aaron Courville
[\[http://www.deeplearningbook.org/\]](http://www.deeplearningbook.org/)
- DeepLearning.ai on Coursera
[\[https://www.coursera.org/deeplearning-ai\]](https://www.coursera.org/deeplearning-ai)

Resources

A mostly complete chart of

Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org

