Introduction to Deep Learning

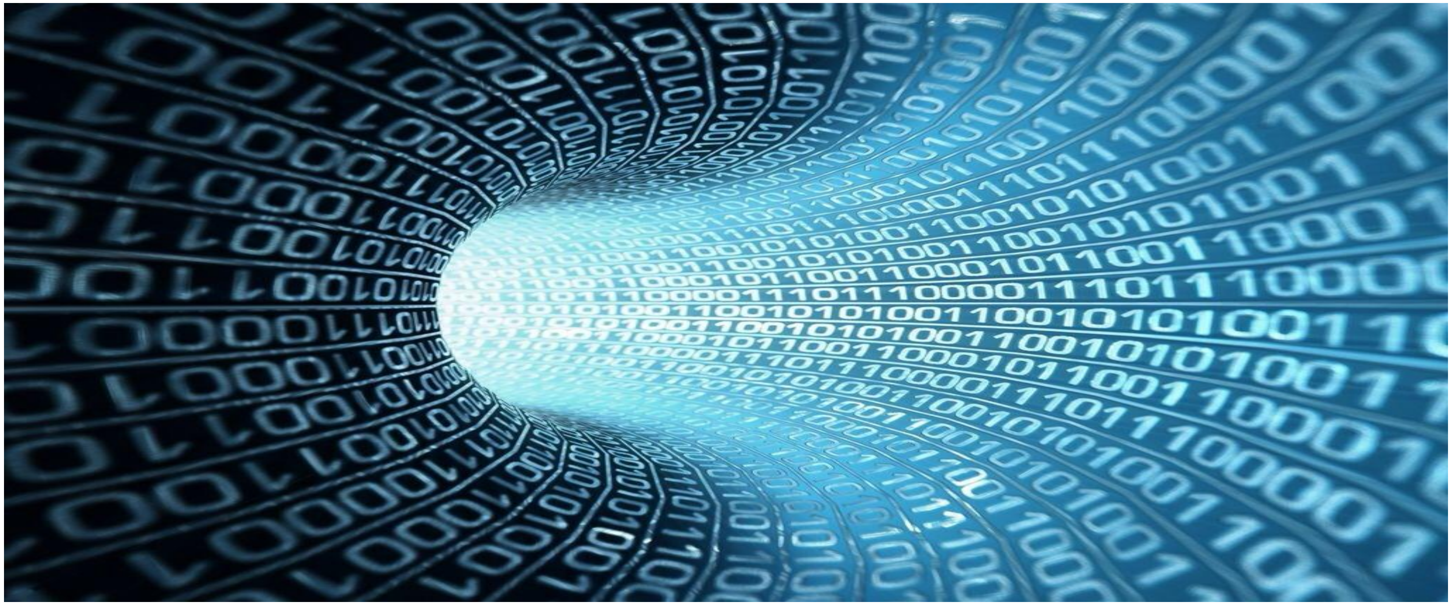
Introduction

- Learn English Alphabets with books having colorful pictures

- Why images when objective is to learn the alphabet?

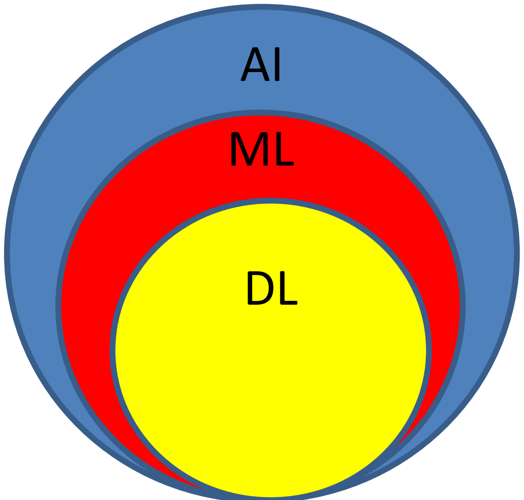
What is Deep Learning?

- Type of machine learning that imitates the way humans gain certain types of knowledge
- Extremely beneficial to data scientists for interpreting large amounts of data
  - Deep learning makes this process faster and easier



What is Deep Learning

- Intelligence – to process information which can be used for future decision
  - AI builds algorithms to achieve this and perform predictions
  - ML is subset of AI – teaches algorithms to learn from experiences without being explicitly programmed
  - DL – uses neural networks to extract useful patterns/features from raw data and using them to perform a task

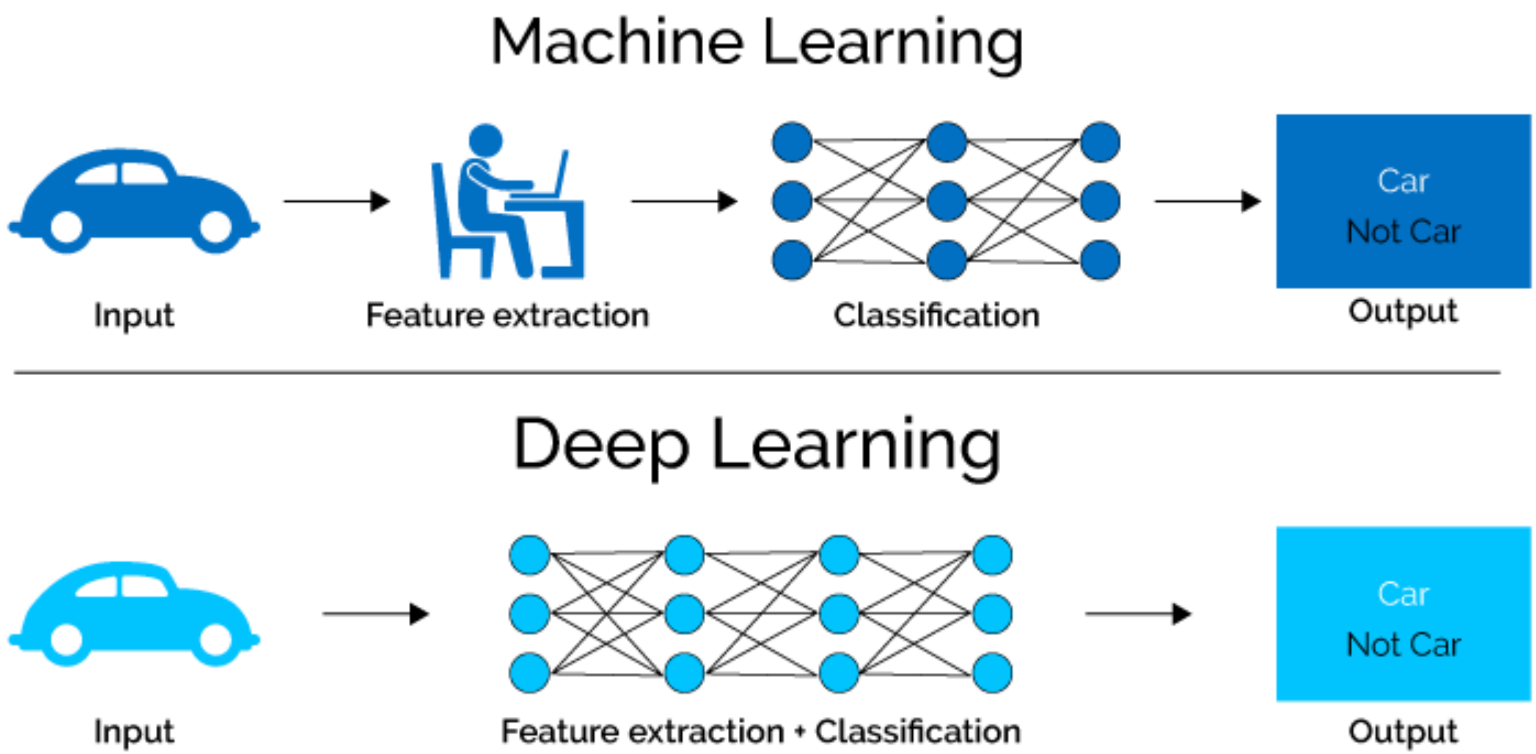


Deep Learning

- Deep learning algorithms
  - Run data through several “layers” of neural network algorithm
  - Each layer passes a simplified representation of data to next layer
- Machine Learning algorithms:
  - Work well on datasets that have up to a few hundred features
  - An unstructured dataset has large number of features – difficult for traditional machine learning algorithms to handle
  - Ex. an 800 * 1000 pixel image in RGB color has 2.4 million features

Why Deep Learning

- Hand-engineered/handcrafted features are time consuming and not scalable in practice



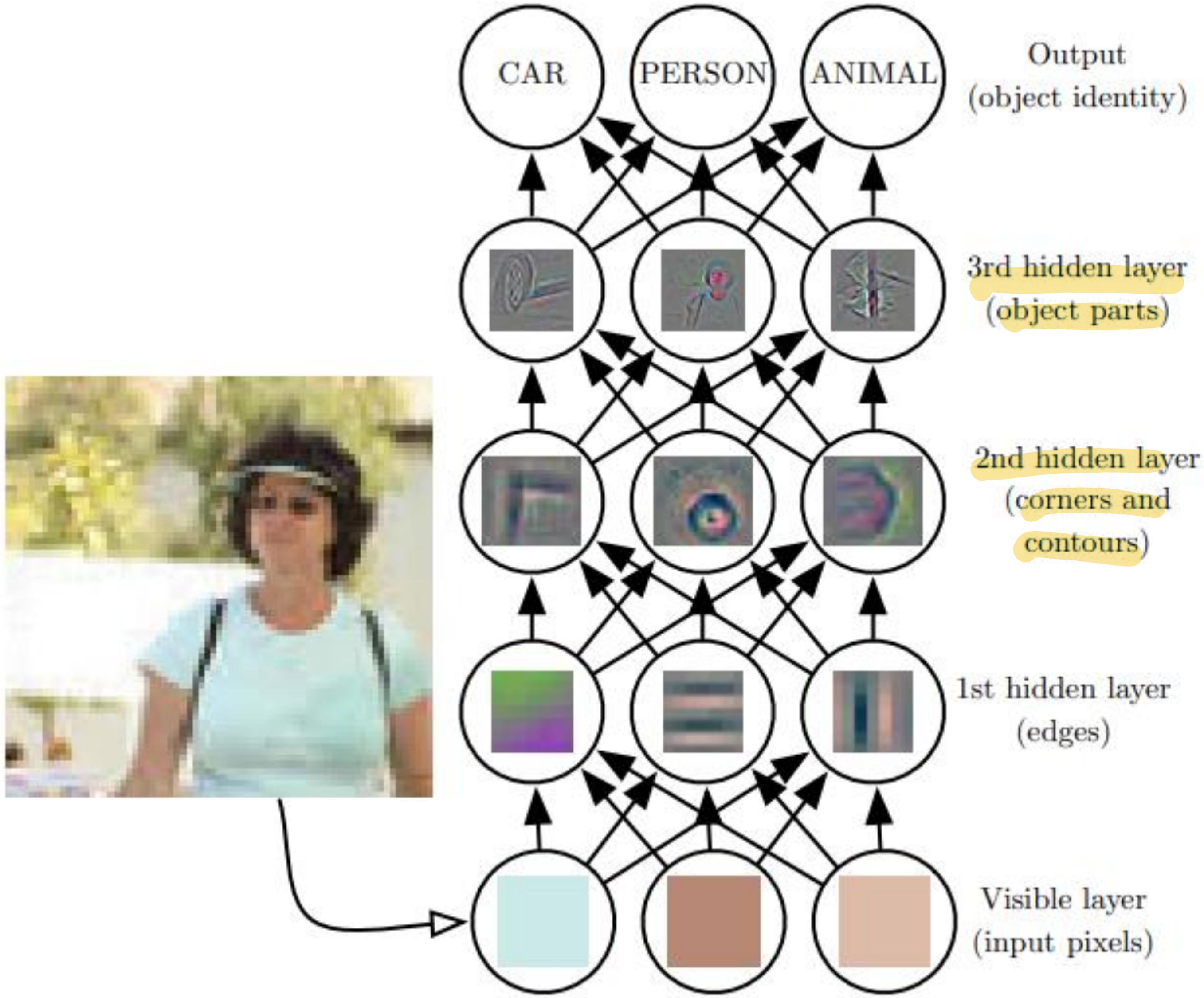
https://semiengineering.com/deep-learning-spreads/

Why Deep Learning

- Key - learn underlying features directly from data in an hierarchical manner

https://ai.plainenglish.io/introduction-to-deep-learning-2e7f21cadc4e

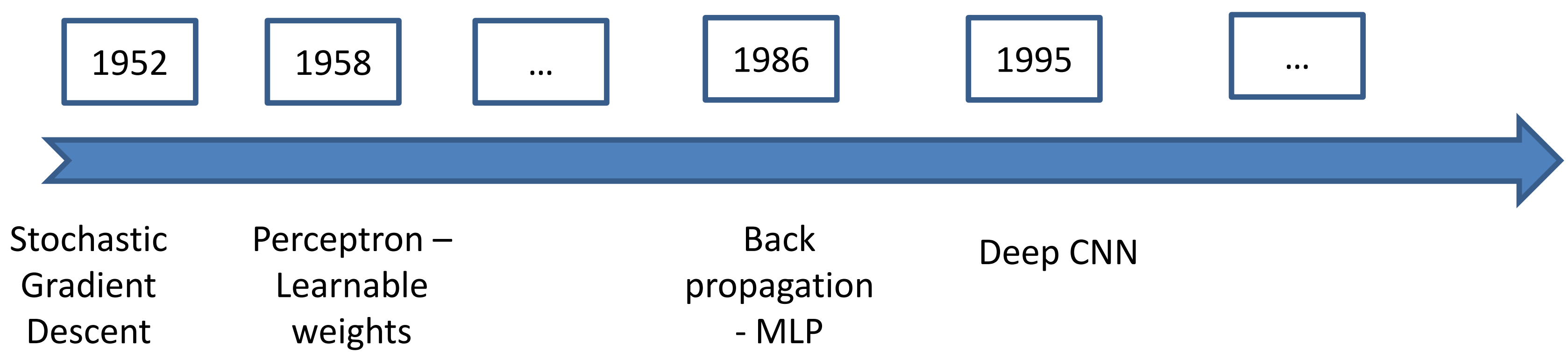
Why Deep Learning



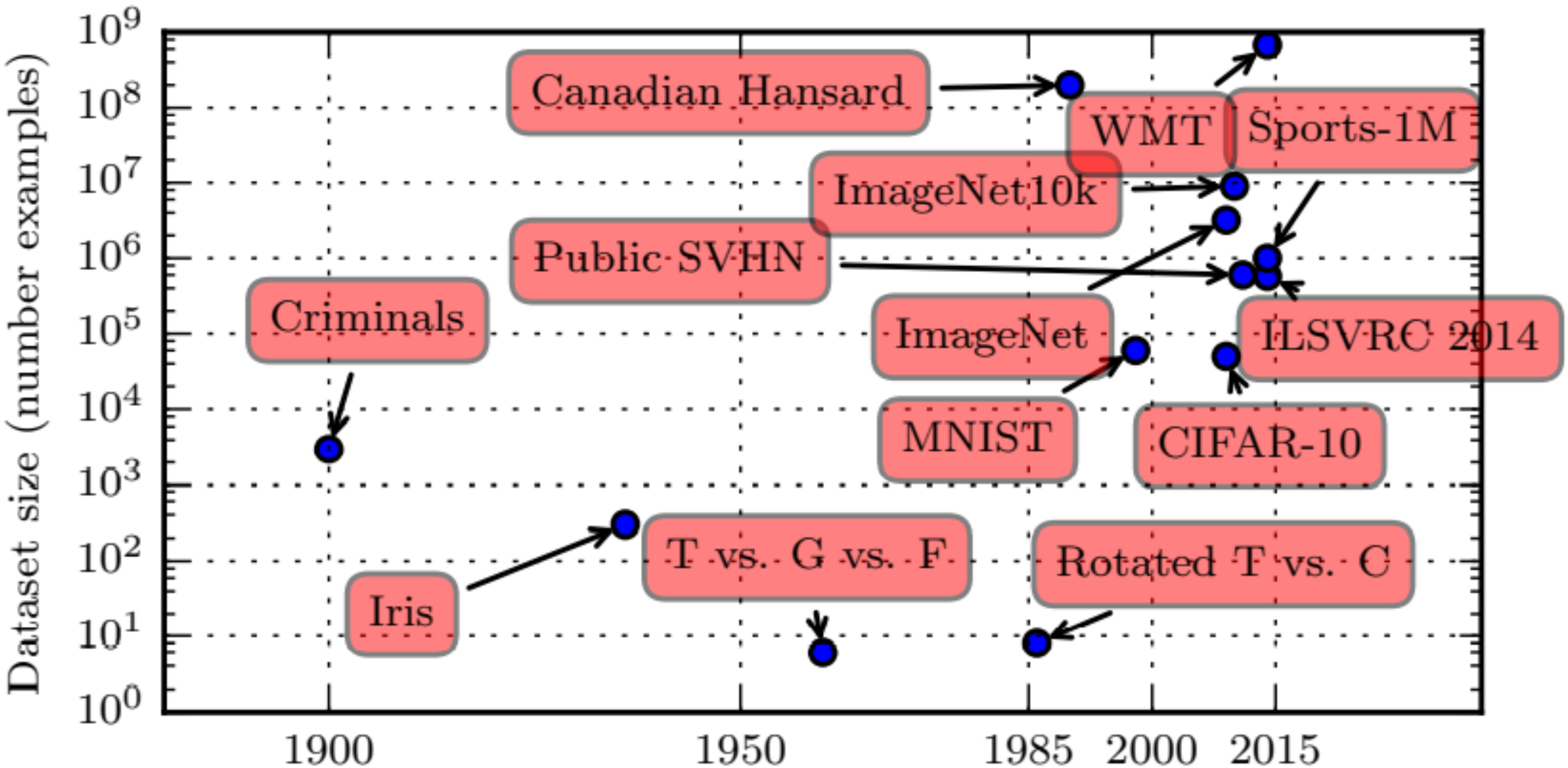
https://www.deeplearningbook.org/

Why Deep Learning Now?

- Data is more pervasive – Big Data
  - Larger datasets, easier collection and storage
- Hardware – Graphics Processing Units (GPUs)
- Parallelizable algorithms
- Software – better techniques, new models, open source toolboxes



Growing Datasets

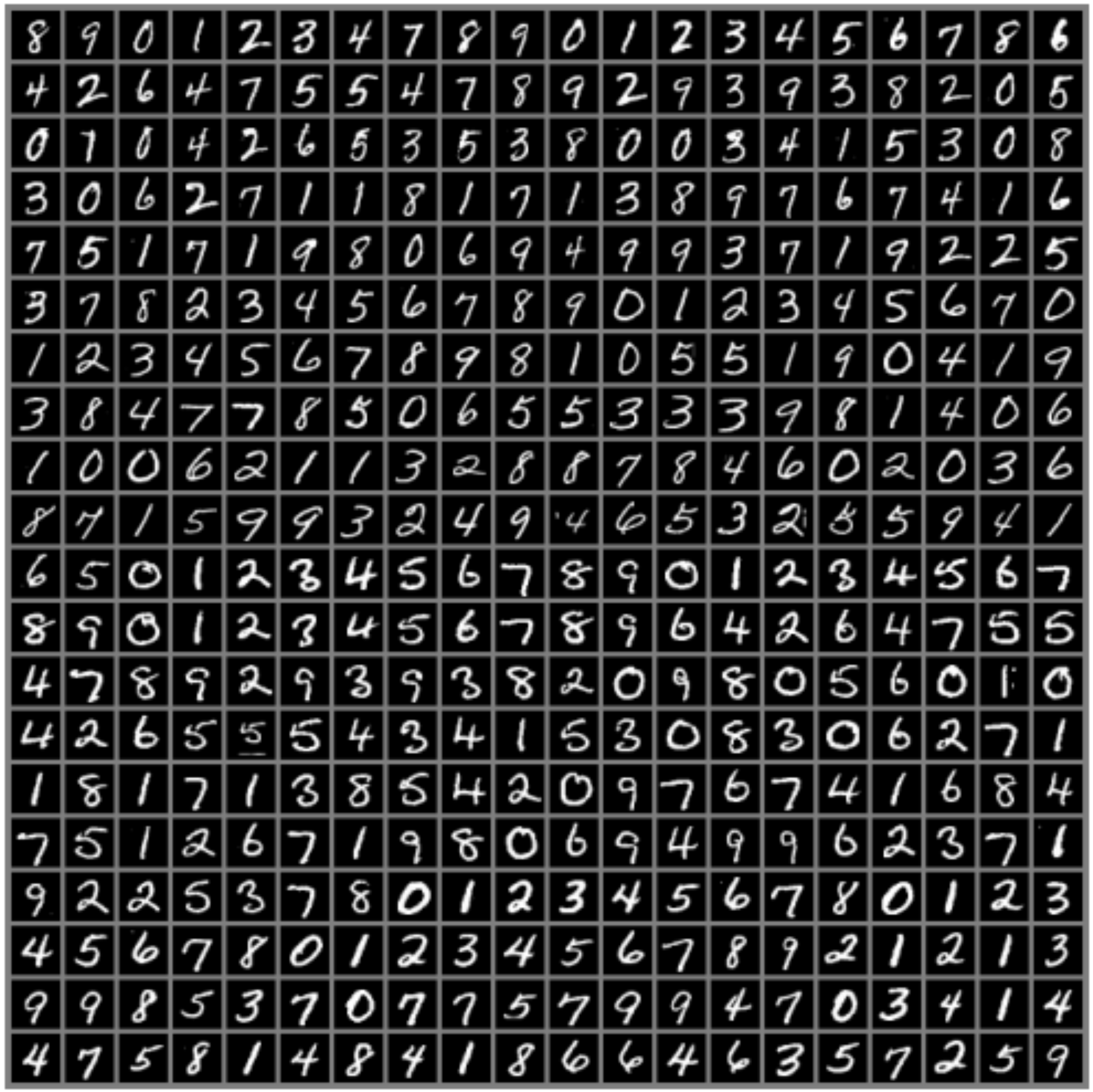


Ac

IRIS: https://archive.ics.uci.edu/ml/datasets/iris
MNIST: http://yann.lecun.com/exdb/mnist/
ImageNet: https://www.image-net.org/download.php

https://www.deeplearningbook.org/

MNIST Dataset



https://www.deeplearningbook.org/

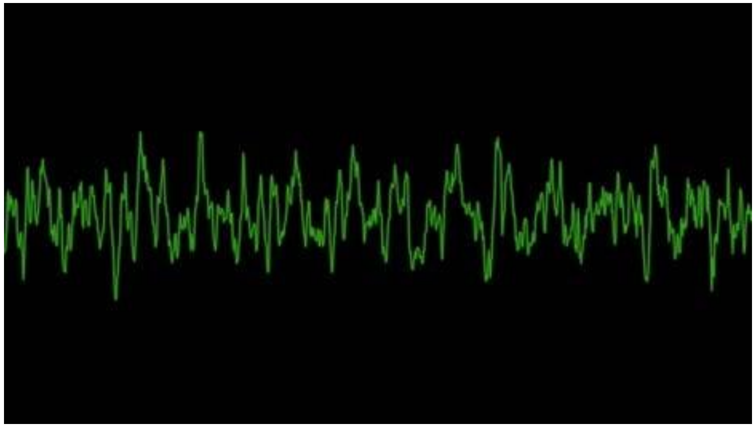
Structured vs. Unstructured Data

Size	No. of bedrooms	Price (in Lakhs)
150	2	80
200	3	120
380	4	250

Age	Ad Id	Click
25	10682	1
16	2051	0
58	31289	1

Image

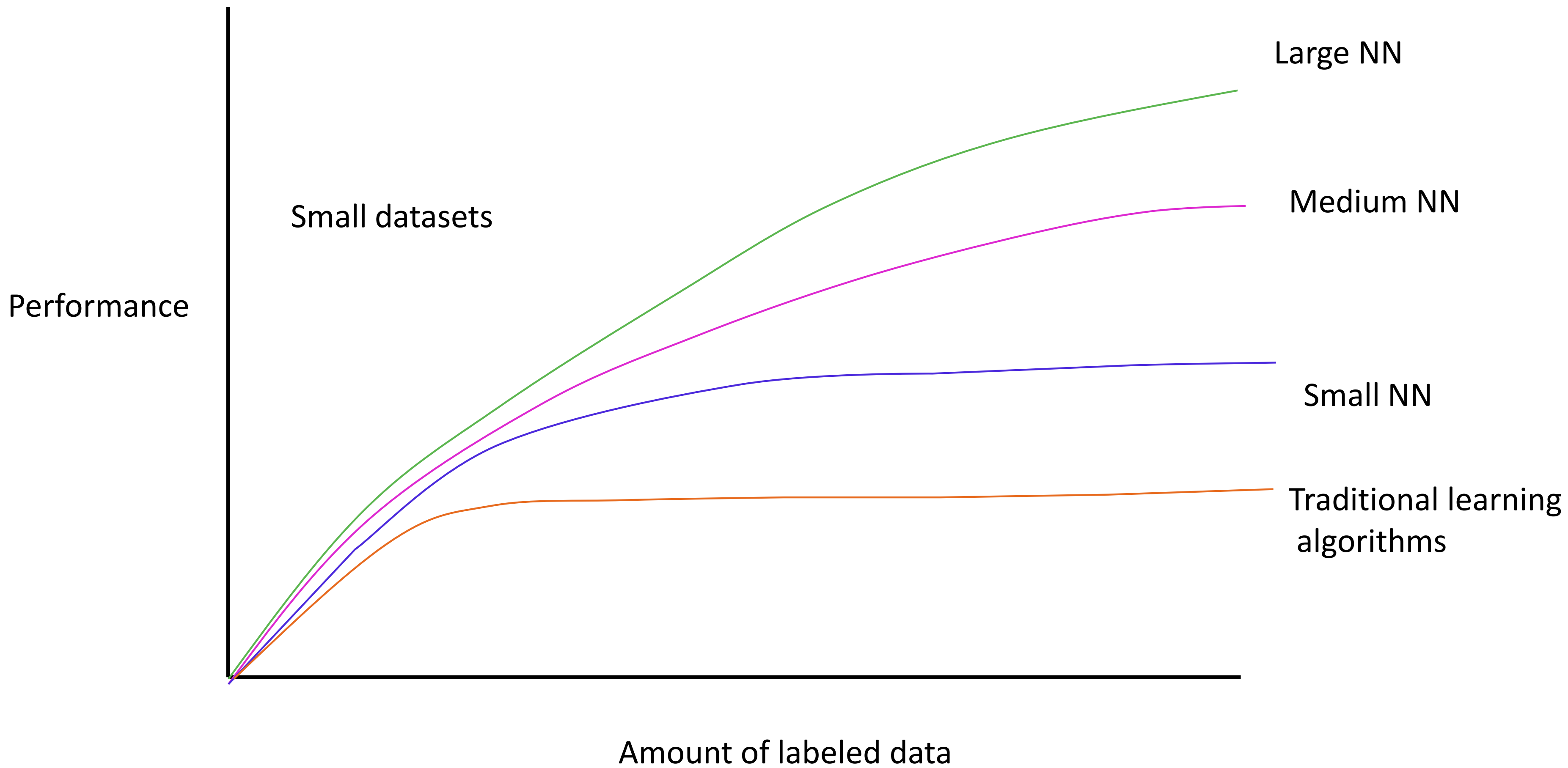
Audio



Text

Once upon a time, in a land far far away ….

Effect of Data Scaling



Two things to be considered for high level of performance:

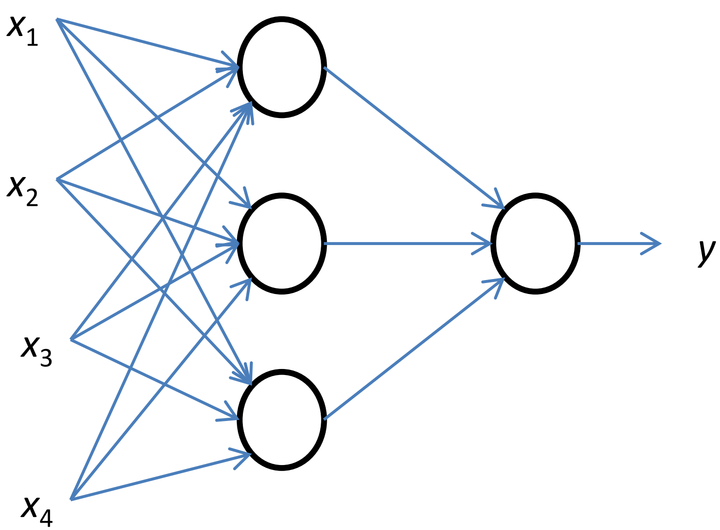
1. Able to train a big enough neural network
2. Large amount of labeled data

Supervised Learning

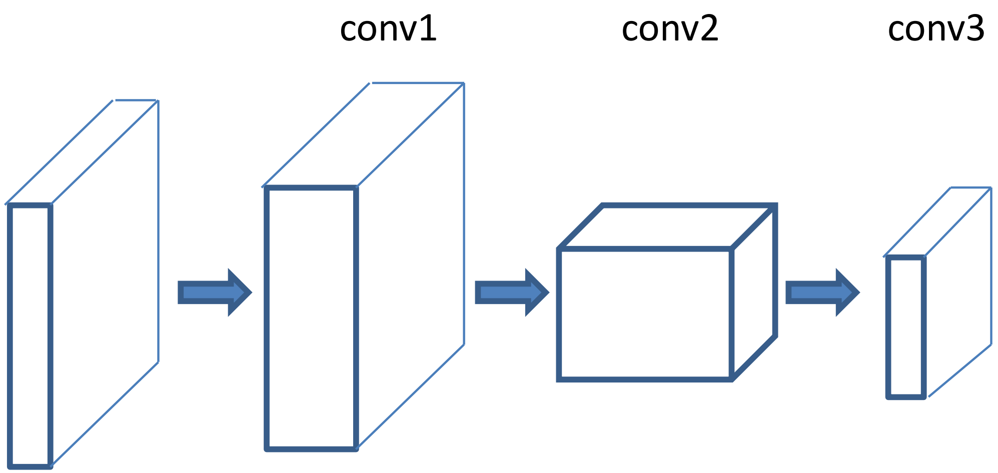
- Given a data set and correct output - relationship between input and output

Input	Output	Application	Type of NN
Home features	Price	Real estate	Standard NN
Image	Object (1..2000)	Photo tagging	CNN
Audio	Text transcript	Speech recognition	RNN
English	French	Machine translation	RNN
Sensor information	Position of objects on road	Autonomous driving	Hybrid
Ad, user information	Ad click?	Online advertising	Standard NN
Sensor information	Sunny?	Weather forecasting	Standard NN

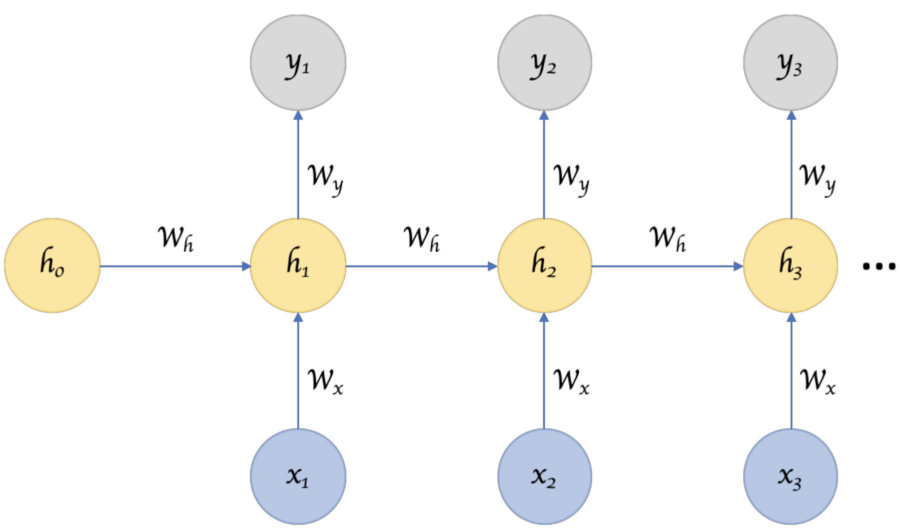
NN examples



Standard Neural Network



Convolutional Neural Network



Recurrent Neural Network

https://gotensor.com/2019/02/28/recurrent-neural-networks-remembering-whats-important/

APPLICATIONS

Language Modeling

S = Where are we going

Previous words (Context)

Word being predicted

P(S) = P(Where) x P(are | Where) x P(we | Where are) x P(going | Where are we)

https://thegradient.pub/understanding-evaluation-metrics-for-language-models/

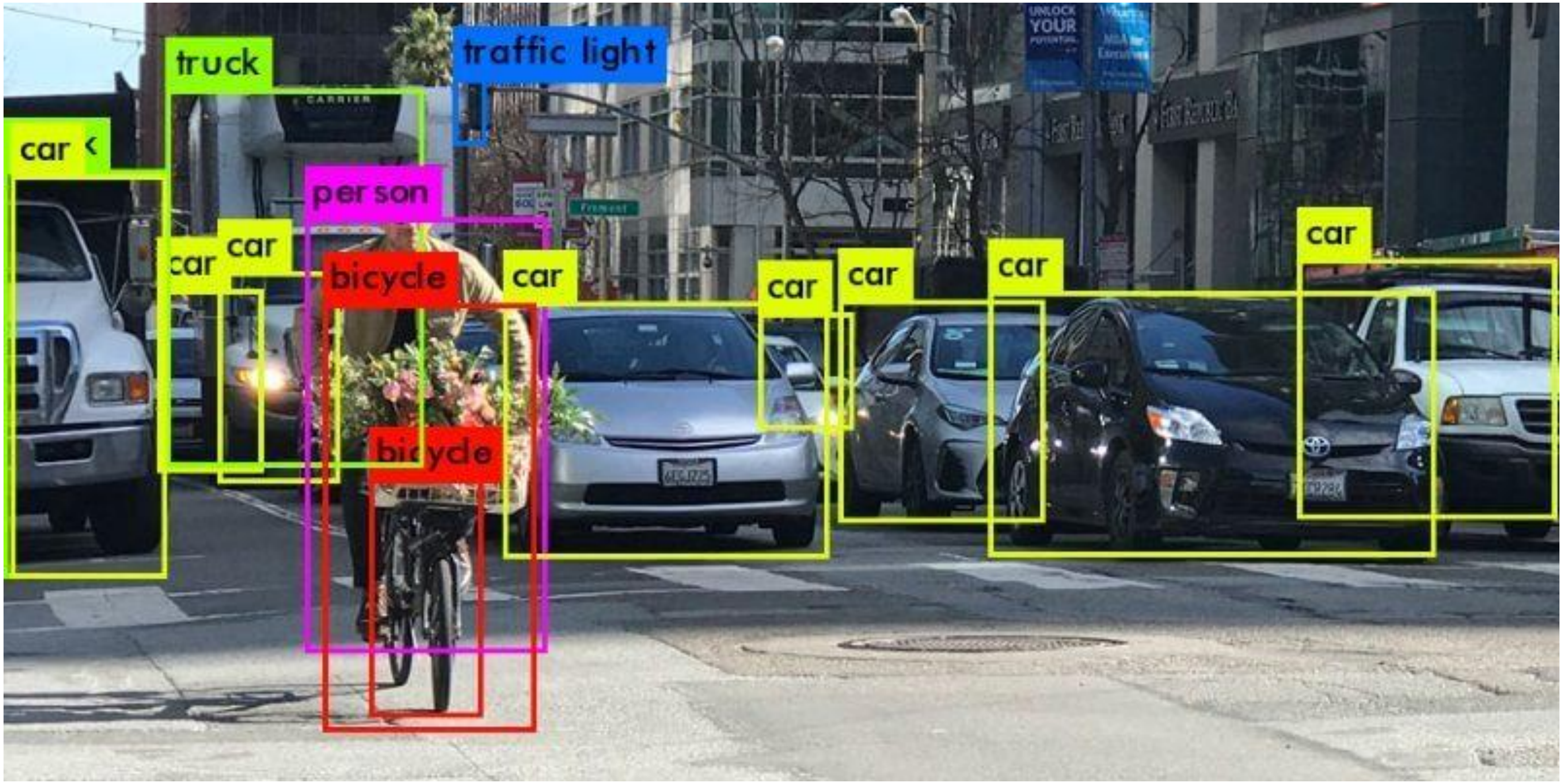
Speech Recognition

https://www.xrtoday.com/virtual-reality/what-is-speech-recognition-technology-in-vr/

Machine Translation

https://livesalesman.com/why-you-cant-depend-on-machine-translation-for-multilingual-customer-support/

Object Detection/Recognition



https://azati.ai/image-detection-recognition-and-classification-with-machine-learning/

Image Captioning

https://evergreen.team/articles/automatic-image-captioning.html

Generating Authentic Photos

https://www.bbvaopenmind.com/en/technology/artificial-intelligence/gan-the-rise-of-imagination-in-machines/

Language Modeling

- Tomas Mikolov, Martin Karafiát, Lukás Burget, Jan Cernocký, and Sanjeev Khudanpur. Recurrent neural network based language model. In INTERSPEECH2010, 11th Annual Conference of the International Speech Communication Association, pages1045–1048, 2010.
- Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Skip-thought vectors. In Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, pages 3294–3302, 2015.
- Yoon Kim, Yacine Jernite, David Sontag, and Alexander M. Rush. Character-aware Neural language models. CoRR,abs/1508.06615, 2015.

Speech Recognition

- Deep neural networks for acoustic modelling in speech recognition: The shared views off our research groups. IEEE Signal Process. Mag., 29(6): 82–97, 2012.
- Alex Graves, Abdel-Rahman Mohamed, and Geoffrey E. Hinton. Speech recognition with deep recurrent neural networks. In IEEE International Conference on Acoustics, Speech and Signal Processing, pages6645–6649, 2013.
- Jan Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. Attention-based models for speech recognition. In Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, pages 577–585,2015.
- Hasim Sak, Andrew W. Senior, Kanishka Rao, and Françoise Beaufays. Fast and accurate recurrent neural network acoustic models for speech recognition. In INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, pages 1468–1472, 2015.

Machine Translation

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. CoRR,abs/1409.0473, 2014.
- Çaglar Gülçehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Loïc Barrault, Huei-Chi Lin, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. On using monolingual corpora in neural machine translation. CoRR,abs/1503.03535, 2015.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems, pages 3104–3112, 2014.
- Hao Zheng, Yong Cheng, and Yang Liu. Maximum expected likelihood estimation for zero-resource neural machine translation. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pages 4251–4257, 2017.

Object Detection/Recognition

- Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition,CVPR2015, pages 3431–3440, 2015.
- Ming Liang and Xiaolin Hu. Recurrent convolutional neural network for object recognition. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, pages 3367–3375, 2015.
- Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell., 39(6): 1137–1149, 2017
- Sergi Caelles, Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Laura Leal-Taixé, Daniel Cremers, and Luc Van Gool. One-shot video object segmentation. In 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR2017,pages 5320–5329, 2017.

Image Captioning

- Junhua Mao, Xu Wei, Yi Yang, Jiang Wang, Zhiheng Huang, and Alan L. Yuille. Learning like a child: Fast novel visual concept learning from sentence descriptions of images. In ICCV, December2015.
- Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Trevor Darrell, and Kate Saenko. Long-term recurrent convolutional networks for visual recognition and description. In Conference on Computer Vision and Pattern Recognition, pages 2625–2634, 2015.
- Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In IEEE Conference on Computer Vision and Pattern Recognition, pages3156–3164,2015.
- Andrej Karpathy and Fei-Fei Li. Deep visual-semantic alignments for generating image descriptions. In IEEE Conference on Computer Vision and Pattern Recognition, CVPR2015, pages 3128–3137, 2015.

Generating Authentic Photos

- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems 27, pages 2672–2680, 2014.
- Anh Nguyen, Jason Yosinski, Yoshua Bengio, Alexey Dosovitskiy, and Jeff Clune. Plug & play generative networks: Conditional iterative generation of images in latent space. CoRR, abs/1612.00005, 2016.
- Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. CoRR,abs/1710.10196, 2017.