Introduction to Deep Learning

Vahid Tarokh

CEE 690/ECE 590, Fall 2019

INSTRUCTOR/TA INFORMATION

Instructor

Vahid Tarokh

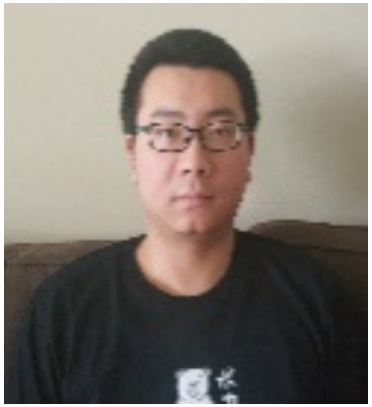
- Electrical and Computer Engineering (Primary)
- Computer Science (Secondary)
- Mathematics (Secondary)

My research is on developing methodologies for data analysis.

Teaching Assistants

1- Christopher Cannella

2- Enmao Diao

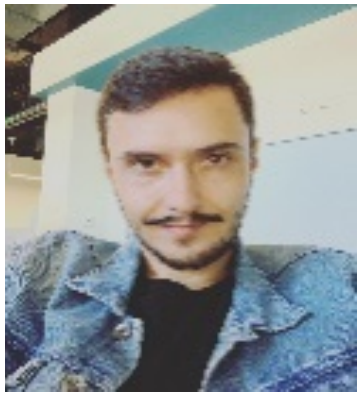


Office hours for each TA will be posted to Sakai

Teaching Assistants

3 - Cat Le

4 – Marko Angjelichinoski



Office hours for each TA will be posted to Sakai

Teaching Assistants

5 – Khalil ElKhalil

6 – MohammadReza Soltani

Office hours for each TA will be posted to Sakai

Teaching Assistants

7 - Robert Ravier

8 - Haibei Zhu

Office hours for each TA will be posted to Sakai

Teaching Assistants

9 - Jiachang Liu

10- TBD (If Needed)

Office hours for each TA will be posted to Sakai

Machine Learning Efforts at Duke

Numerous research groups at Duke focused on data science and machine learning.

New initiative for Health Data Science (the Forge) that is starting in the medical center (https://healthdatascience.duke.edu/)

Many activities and research through the Information Initiative at Duke (iiD) https://bigdata.duke.edu/

Also many events and announcements through https://machinelearning.duke.edu/

Textbooks, etc.

Books:

- **Required text: *Deep Learning*.** Goodfellow, Bengio, and Courville Freely available at http://www.deeplearningbook.org/
- **Required text: Pattern Recognition and Machine Learning,** Christopher M. Bishop
- Freely available at https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern- Recognition-and-Machine-Learning-2006.pdf
- **Optional text: Dive into Deep Learning** Release 0.7
  Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola https://en.d2l.ai/d2l-en.pdf
- **Optional text:** *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. Geron. Electronic version is freely available from the Duke Library. Many helpful applied coding examples.
- Specific papers and references to support learning and provide a reference will be posted to Sakai

Online resources:

- Courses on coursera, etc. on individual topics covered in this course

- **Important Note: Source of some of my slides (with great appreciation and acknowledgements)**
  - **Professor David Carlson Slides**
  - **Professor Alex Smola's slides (available online)**
  - **Professor Lawrence Carin's slides**
  - **Professor Ruslan Salakhutdinov's slides (available online)**

COURSE POLICIES

Online Course Resources

You should be enrolled in the Sakai site for:

- CEE 690/ECE 590

Primary mode of communication

- Announcements
- Lecture note templates posted
- Homework posted
- Questions answered, via Piazza

Piazza is integrated with and can be accessed through Sakai

Enroll yourself in Piazza through Sakai

Piazza is a wiki-like forum where you can ask, and collaboratively answer, questions

- Please ask questions
- Please answer questions

Syllabus Highlights

Grading breakdown:

- 27% Homework (9 HWs*3 points/HW)
  - Each Homework has 3 points
- 33% Mid-term
  - Mostly based on Theory
- 40% Project

Homework policies:

- Accepted up to 3 days after the deadline
- 1/3 subtracted per day late
- Assignments required to be submitted electronically
- **Template/Style file for writing the assignments will be provided on Sakai**
- Code will only be accepted in Python

Project will be split into 3 parts:

- 8%: Project Proposal
- 32%: Project Final Report
- Project Presentation (only required if the report is not crystal clear)

Project Proposals and Reports must be prepared in Powerpoint, Keynote, Beamer presentation software (or equivalent)

Re-grades:

Problem sets and exams will be graded with the benefit of the doubt in mind.

You may request a regrade only within 1 week of the assignment being returned.

If you request a re-grade, the entire problem set or exam will be re-graded and it is possible other questions will lose points.

You are advised to only request a re-grade if an obvious mistake in grading was made.

Academic Integrity

All students are expected to abide by the Duke Community Standard

Actions that constitute academic misconduct aren't always as obvious as this

If you are unsure whether an action would be considered a violation of the Duke Community Standard (DCS), please ask an instructor or TA *before* you do it

PROJECT OVERVIEW

Course Project

- Opportunity for you to explore elements of data science within the context of an application area that is interesting to you!
  - Possible sources of inspiration
    - Future career goals
    - Current research area
    - Personal "pet project" or hobby
  - Should be on on an interesting and meaningful problem
  - Phased submissions throughout semester
    - Phase 1: Project proposal due October 16th
      - Point of the proposal is to make sure that you're on the right track and have set a feasible goal!
    - Further details are in the course project information on Sakai

Course Project

- Can create a team of 1-4 students per project.
  - We will scale expectations with the number of people. Your workload may *increase* with the number of people in your project.
- **Each project must have a TA as a project consultant/mentor.** We will balance the load between TAs.
- A team can have only one single member in all the following constraints are satisfied:
  - (1) you want to design a project around a data set associated with your ongoing research,
  - (2) you absolutely cannot find another student in our class who would be interested in your problem, and
  - (3) your faculty advisor associated with that ongoing research writes us an email stating that she/he will be happy to serve as mentor for your class project.

Course Project

- The Staff will provide about 15-18 project topics.
    - May be increased depending on the number of students
- The students can also propose about 15-18 other project topics.
    - May be increased depending on the number of students
- No project can be assigned to more than two teams

Course Project

Each proposal and report must include these sections:

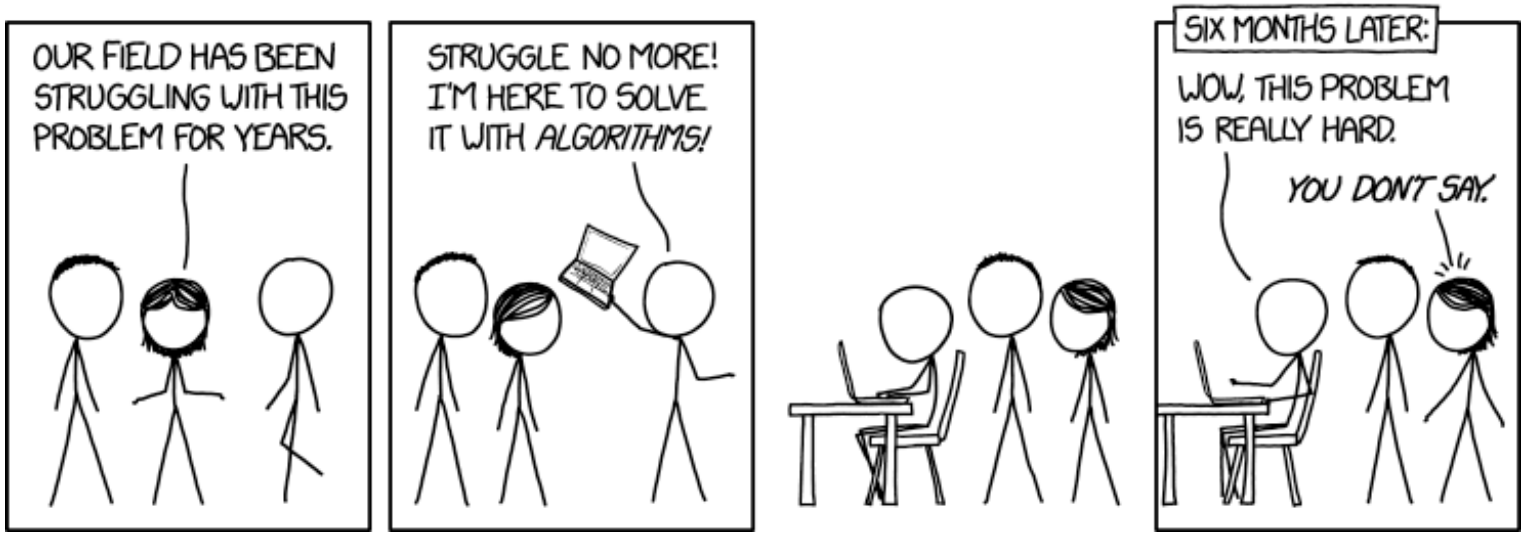
- **Introduction:** You should include a formal description/background about the problem and data that you chose. Clearly outline details and goals of the system you will build.
- **Background/Literature:** You should outline related research work, how you will depart from the existing work, and plan to improve up on the existing results.
- **Methods/Model:** By the submission of your project proposal, you are expected to have implemented a baseline. Outline what you have implemented and show plots of the performance. In the final report outline all the details of the methods and models you implemented.
- **Experimental results:** All the used codes should be well documented and commented. Please make sure that all the codes are executable without any error. We will not evaluate the experiments with non-executable codes or the codes with bugs. Also, if you are using some codes/packages/modules (partially or fully) from external sources (Github, Personal websites, Forums, StackExchange, and so on), please give full references to them and cite them appropriately.
- **Suggestions for future work.**

Course Project

- Team members are responsible for dividing up the work equally and making sure that each member contributes. We will give the same project grade to all the team members.

- The Project Proposal and the Final Report have to be delivered in **Powerpoint, Beamer, Keynote or equivalent.**
- Project Proposal Page limit = 20 pages (including all Figures and Videos)
- Final Project Page limit = 50 pages (including all Figures and Videos)
- Page limits exclude references and title page.
- The page limits are strict! Proposals and Reports over the limit will not be considered.
- No font size smaller than 12 points is allowed.
- No active links to outside videos, images, documents are allowed.

You will only be evaluated on approaching the problem correctly!



https://xkcd.com/1831/

Real-world problems are hard—results are often lacking. What we want to see is that have learned how to correctly approach a problem using deep learning.

Other Deliverables

A high-quality final report is due December 6th, 2019 at 23:59:59 EST .

If the report is not crystal clear, you may be selected for a presentation on 12/10/2019.

QUICK NOTES ON CODING FRAMEWORKS

Sections

- Sections review programming languages, platforms, etc.
- Six Discussion Sections:
    - You must be registered in one.
- Section leaders are rotated so that you have a chance to choose the TA of your choice (up on mutual agreement) as your project leader/consultant.
- We will also review foundational concepts in sections.
- The Instructor may attend one or more of the sections at random to make sure of uniformity of sections.
- There will be no sections in week one.

Python is the language of choice

- Python is most popular programming language for Deep learning.
- We will be teaching using Python with the PyTorch framework using Jupyter Notebooks in discussion sections.
- Homeworks will be required to be done in Python.

Tools d2l.ai/chapter_crashcourse/install.html

- **Python**
  - Everyone is using it in machine learning & data science
  - Conda package manager (for simplicity)
- **Jupyter**
  - It is an interactive environment to run the codes and see the results
  - So much easier for education purposes
- **Pytorch**
    - **It is a very popular framework for deep learning. Other options include Tensorflow and Keras**
- **Anaconda**
    - **It is a popular packet manager for doing data science and machine learning**

Computational Resources

- Deep Learning is computationally expensive
- For the purposes of this class, about 20 GPUs are reserved in the Duke Compute Cluster
- We will give details on logging in and show an example using them
- This is more than enough computation for all the students in this course
  - However, plan ahead. No extensions will be given for waiting on resources.

SOME DATASETS

Data sets in Machine/Deep learning

In deep learning, there are many benchmark data sets

Researchers and practitioners use the data sets to evaluate and to develop their goal

Here, we introduce some of the famous data sets in different domains, including image, text, speech, and time series

Some sources from (with great appreciation and acknowledgements)

- https://machinelearningmastery.com/

Image type data sets:

MNIST

- http://yann.lecun.com/exdb/mnist/
- Purpose: Object classification
- The MNIST database of handwritten digits, available from the above link, has a training set of 60,000 examples, and a test set of 10,000 examples (The examples are 28×28 gray-scale 2-D image of digits 0 t0 10).

Image type data sets:

Fashion-MNIST

- https://github.com/zalandoresearch/fashion-mnist
- Purpose: Object classification
- A training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 gray-scale image, associated with a label from 10 classes (T-shirt/top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, Ankle boot).

Image type data sets:

The Quick, Draw!

- https://github.com/googlecreativelab/quickdraw-dataset
- Purpose: Object classification
- The Quick Draw Dataset is a collection of 50 million drawings across 345 categories, contributed by players of the game Quick, Draw!.
- The drawings were captured as time-stamped vectors, tagged with metadata including what the player was asked to draw and in which country the player was located.

Image type data sets:

CIFAR

https://www.cs.toronto.edu/~kriz/cifar.html

- Purpose: Object classification
- **The CIFAR-10 dataset**
  - The CIFAR-10 dataset consists of 60000 32x32 color images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images.
- **The CIFAR-10 dataset (variation)**
  - This dataset is just like the CIFAR-10, except it has 100 classes containing 600 images each. There are 500 training images and 100 testing images per class.

Image type data sets:

The Street View House Numbers (SVHN) Dataset

- http://ufldl.stanford.edu/housenumbers/
- Purpose: Object classification
- It can be seen as similar in flavor to MNIST (e.g., the images are of small cropped digits), but incorporates an order of magnitude more labeled data (over 600,000 digit images) and comes from a significantly harder, unsolved, real world problem (recognizing digits and numbers in natural scene images).
  - SVHN is obtained from house numbers in Google Street View images.
- Typical use: MNIST-like 32-by-32 images centered around a single character (many of the images do contain some distractors at the sides).
- 73257 digits for training, 26032 digits for testing (531131 additional).
- 10 classes corresponding to the 10 classes.

Image type data sets:

- **ImageNet**
  - http://www.image-net.org/
  - Purpose: Object classification, object localization, and object detection
  - It is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images.
  - Extremely huge data sets (14,197,122 images)

Image type data sets:

Common Object in Context (COCO)

- http://cocodataset.org/#home
- Purpose: object detection, object segmentation, and captioning
- 123,287 images, 886,284 instances

Image type data sets:

Large-scale CelebFaces Attributes (CelebA)

- http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

- Purpose: face attribute recognition, face detection, landmark localization, and face editing & synthesis.
- A large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations.
- The images in this dataset cover large pose variations and background clutter.
- CelebA has large diversities, large quantities, and rich annotations, including
  - **10,177** number of **identities**
  - **202,599** number of **face images**
  - **5 landmark locations**, **40 binary attributes** annotations per image

Text type data sets in Natural Language Processing:

Reuters-21578 Text Categorization Collection

- http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html
- Purpose: Text classification and sentiment analysis
- It is a collection of documents that appeared on Reuters newswire in 1987. The documents were assembled and indexed with categories.

Text type data sets in Natural Language Processing:

Reuters Corpora (RCV1, RCV2, TRC2)

- https://trec.nist.gov/data/reuters/reuters.html
- Purpose: Text classification and sentiment analysis
- This corpus, known as "Reuters Corpus, Volume 1" or RCV1, is significantly larger than the older, well-known Reuters-21578 collection heavily used in the text classification community.

Text type data sets in Natural Language Processing:

Large Movie Review Dataset

- http://ai.stanford.edu/~amaas/data/sentiment/
- Purpose: Binary sentiment classification
- It consists a set of 25,000 highly polar movie reviews for training, and 25,000 for testing. There is additional unlabeled data for use as well. Raw text and already processed bag of words formats are provided.

Text type data sets in Natural Language Processing:

News Group Movie Review

- http://www.cs.cornell.edu/people/pabo/movie-review-data/
- Purpose: Sentiment Classification
- A collection of movie reviews from the website imdb.com and their positive or negative sentiment

Text type data sets in Natural Language Processing:

Brown University Standard Corpus of Present-Day American English

- https://en.wikipedia.org/wiki/Brown_Corpus
- Purpose: Language modeling
- A large sample of English words

Text type data sets in Natural Language Processing:

Project Gutenberg

- https://www.gutenberg.org/
- Purpose: Language modeling (developing a statistical model for predicting the next word in a sentence or next letter in a word given whatever has come before).
- A large collection of free books that can be retrieved in plain text for a variety of languages

Text type data sets in Natural Language Processing:

The PASCAL Object Recognition Database Collection

- http://host.robots.ox.ac.uk/pascal/VOC/databases.html
- Purpose: Image captioning
- The dataset has 20 classes, including aeroplane, bicycle, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, train, TV.

Text type data sets in Natural Language Processing:

Aligned Hansards of the 36th Parliament of Canada

https://www.isi.edu/natural-language/download/hansard/

- Purpose: Machine translation
- Pairs of sentences in English and French
- 1.3 million pairs of aligned text chunks (sentences or smaller fragments) from the official records (*Hansards*) of the 36th Canadian Parliament

Text type data sets in Natural Language Processing:

Stanford Question Answering Dataset (SQuAD)

- https://rajpurkar.github.io/SQuAD-explorer/
- Purpose: Question Answering
- It is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage, or the question might be unanswerable.

Text type data sets in Natural Language Processing:

Deepmind Question Answering Corpus

- https://github.com/deepmind/rc-data
- Purpose: Question Answering
- Question answering about news articles from the Daily Mail. It contains a script to generate question/answer pairs using CNN and Daily Mail articles downloaded from the Wayback Machine.

Speech type data sets

TIMIT Acoustic-Phonetic Continuous Speech Corpus

- https://catalog.ldc.upenn.edu/LDC93S1
- Purpose: Speech recognition (transforming audio of a spoken language into human readable text)
- Not free, but listed because of its wide use. Spoken American English and associated transcription. It contains broadband recordings of 630 speakers of eight major dialects of American English, each reading ten phonetically rich sentences. The TIMIT corpus includes time-aligned orthographic, phonetic and word transcriptions as well as a 16-bit, 16kHz speech waveform file for each utterance.

Speech type data sets

LibriSpeech ASR corpus

- http://www.openslr.org/12/
- Purpose: Speech recognition
- Large collection of English audiobooks taken from LibriVox.
- 1000 hours corpus of read English speech

Time series type data sets

Monthly Sunspot Dataset

- Purpose: Prediction (Univariate time series)
- https://raw.githubusercontent.com/jbrownlee/Datasets/master/monthly-sunspots.csv
- This dataset describes a monthly count of the number of observed sunspots for just over 230 years (1749-1983). The units are a count and there are 2,820 observations. The source of the dataset is credited to Andrews & Herzberg (1985).

Time series type data sets

Daily Female Births Dataset

- https://raw.githubusercontent.com/jbrownlee/Datasets/master/daily-total-female-births.csv
- Purpose: Prediction (Univariate time series)
- This dataset describes the number of daily female births in California in 1959. The units are a count and there are 365 observations. The source of the dataset is credited to Newton (1988).

Time series type data sets

EEG Eye State Data Set

- http://archive.ics.uci.edu/ml/datasets/EEG+Eye+State
- Purpose: Classification predictive modeling (Multivariate time series)
- This dataset describes EEG data for an individual and whether their eyes were open or closed. There are a total of 14,980 observations and 15 input variables. The class value of ‘1’ indicates the eye-closed and ‘0’ the eye-open state.

Time series type data sets

Ozone Level Detection Dataset

- http://archive.ics.uci.edu/ml/datasets/Ozone+Level+Detection
- Purpose: Classification predictive modeling (Multivariate time series)
- This dataset describes 6 years of ground ozone concentration observations and the objective is to predict whether it is an “ozone day” or not.
- The dataset contains 2,536 observations and 73 attributes.

Source of machine/deep learning data sets

There are still many data sets freely available

Useful resources in the web:

- UCI machine learning repository
  - http://archive.ics.uci.edu/ml/index.php
- NLTK corpora
  - http://www.nltk.org/nltk_data/
- Stanford NLP collection
  - https://nlp.stanford.edu/links/statnlp.html#Corpora
- Torchvision image data sets:
  - https://pytorch.org/docs/stable/torchvision/datasets.html#cifar
- Kaggle data sets
  - https://www.kaggle.com/datasets?utm_medium=paid&utm_source=google.com+search&utm_campaign=datasets&&gclid=Cj0KCQjwv8nqBRDGARIsAHfR9wBNRGQCjxnoypMTh4q7TI9OA3NtBmp9fyyaD6IGquQuaSgxc5wgvFYaAswCEALw_wcB

COURSE MATERIAL

MAY BE REVISED BASED ON FEEDBACK

Course Material

May be Revised Based on Feedback

- Introduction
- Mathematical Background,
  - **Linear Algebra, Distributions, Rules of probability**
  - **Regression, Classification**
  - Modeling and Validation Methods
- Computation Graphs and Large-Scale Logistic Regression
- Deep Feed-Forward Networks, Back-propagation
- Regularization for Deep Learning
- Optimization for Training Deep Networks, Stochastic Gradient Descent,
- Algorithms with Adaptive Learning Rates

Course Material

May be Revised Based on Feedback

- Convolutional Neural Networks (for image/text analysis)
- Graphical Models
- Deep Belief Networks
- Recursive Neural Networks, Long Short Term Memory
- Language Modeling
- Deep Learning in Practice
- Linear Factor Models
- Autoencoders
- Representation Learning
- Probabilistic Modeling for Deep Learning

Course Material

May be Revised Based on Feedback

- Approximate Inference
- Deep Generative Models
- Boltzmann Machines, and Restricted Boltzmann Machines
- Variational Auto-encoders
- Reinforcement Learning
- Generative Adversarial Networks

INTRODUCTION

Examples of Types of Learning

Consider observing a series of input vectors:

$$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ldots.$$

- **Supervised Learning:** We are also given target outputs (labels, responses): $y_1, y_2, \ldots$, and the goal is to predict correct output given a new input.

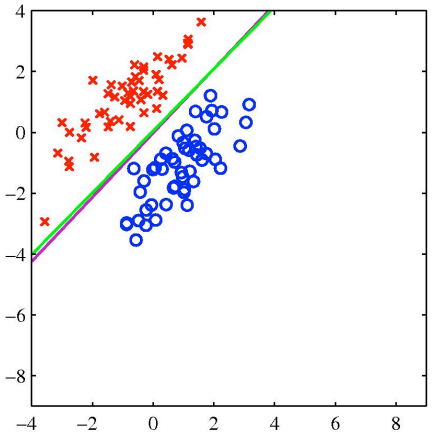
- **Unsupervised Learning:** The goal is to build a statistical model of **x**, which can be used for making predictions, decisions.

- **Semi-supervised Learning:** We are given only a limited amount of labels, but lots of unlabeled data.

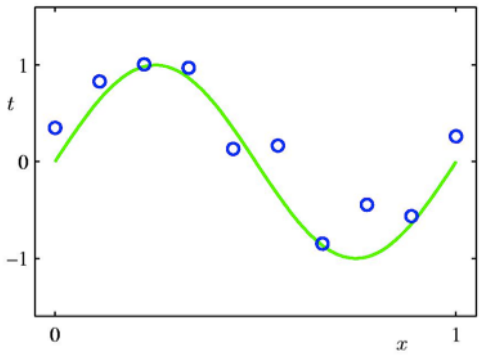
- **Reinforcement Learning:** the model (agent) produces a set of actions: a_1, $a_2, \ldots$ that affect the state of the world, and received rewards r_1, $r_2 \ldots$ The goal is to learn actions that maximize the reward.

Supervised Learning

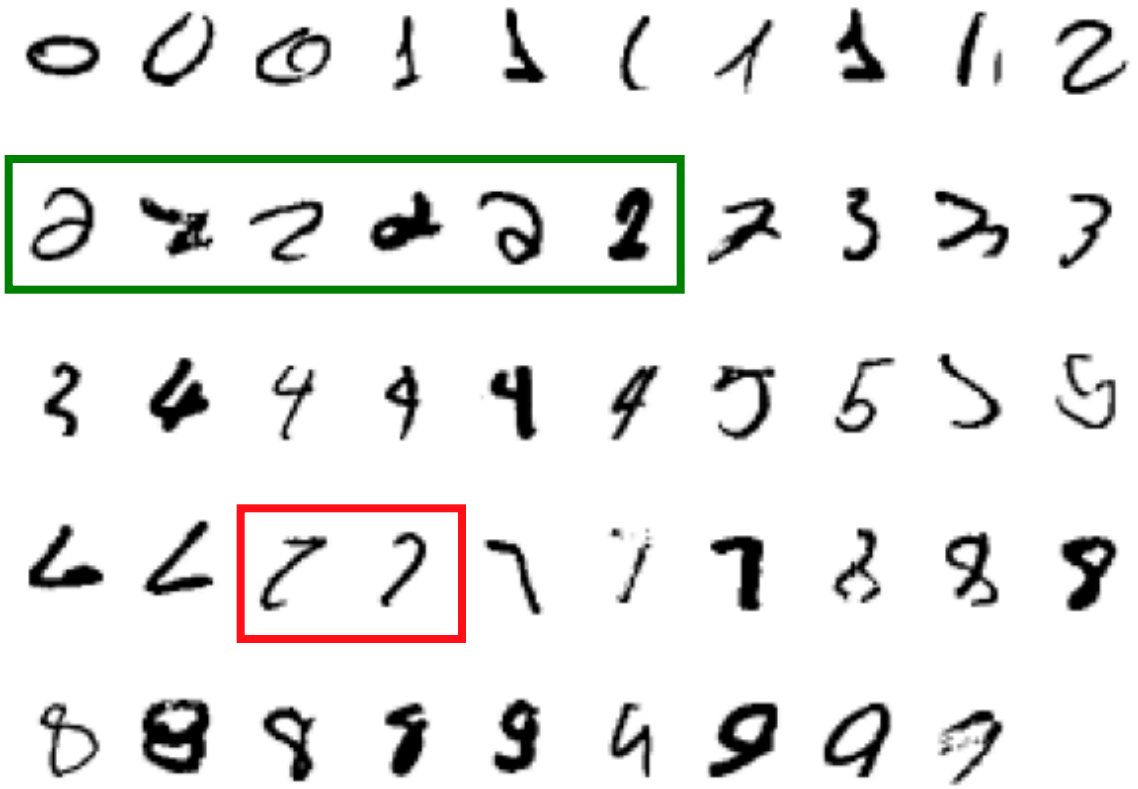
Classification: target outputs y_i are discrete class labels. The goal is to correctly classify new inputs.



Regression: target outputs y_i are continuous. The goal is to predict the output given new inputs.



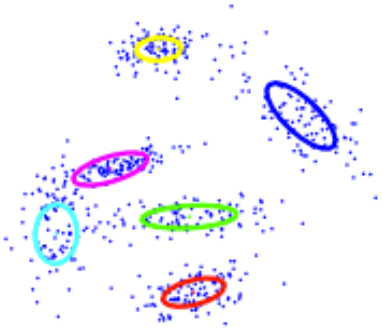
Handwritten Digit Classification



Unsupervised Learning

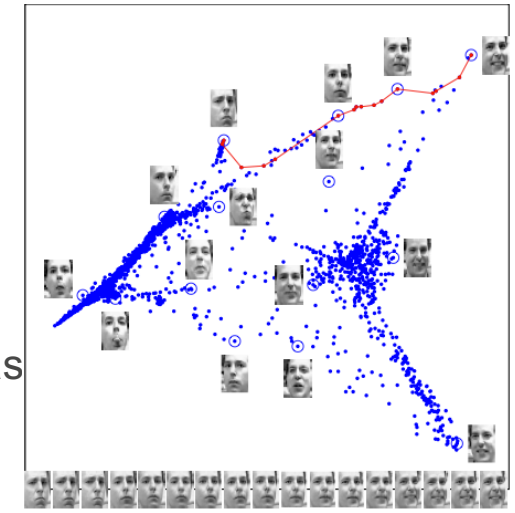
The goal is to construct statistical model that finds useful representation of data:

- Clustering
- Dimensionality reduction
- Modeling the data density
- Finding hidden causes (useful explanation) of the data

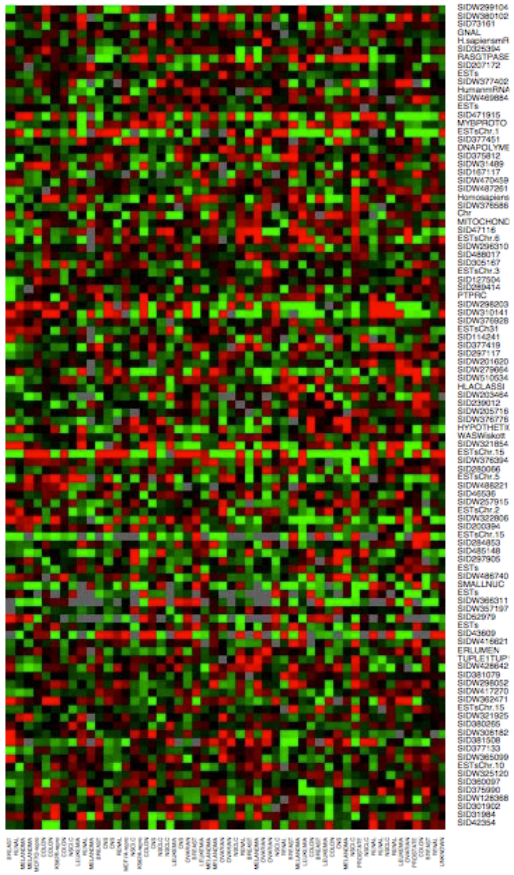


Unsupervised Learning can be used for:

- Structure discovery
- Anomaly detection / Outlier detection
- Data compression, Data visualization
- Used to aid classification/regression tas



DNA Microarray Data



Expression matrix of 6830 genes (rows) and 64 samples (columns) for the human tumor data.

The display is a heat map ranging from bright green (under expressed) to bright red (over expressed).

Questions we may ask:

- Which samples are similar to other samples in terms of their expression levels across genes.
- Which genes are similar to each other in terms of their expression levels across samples.

Deep Learning?

This class is focused on "Deep Learning:"

- What it is? (A Little bit of history)
- Is it supervised, unsupervised or RL?
- Some of its many applications?
- How to *build your own* deep networks?

These algorithms *learn* from data to automate tasks and make predictions

Today, we will set up course expectations and briefly sketch out the covered material

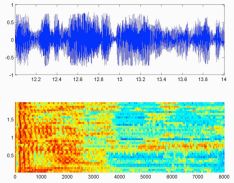
Mining for Structure

- Massive increase in both computational power and the amount of data available from web, video cameras, laboratory measurements.

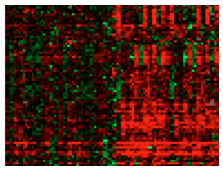
Images & Video

Text & Language

Speech & Audio



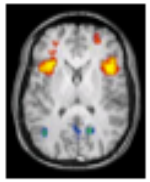
Gene Expression



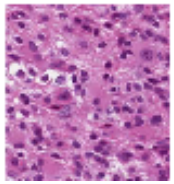
Product Recommendation

Relational Data/ Social Network

fMRI



Tumor region



- Develop statistical models that can discover underlying structure, cause, or statistical correlation from data.
- Multiple application domains.

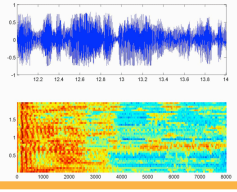
Mining for Structure

- Massive increase in both computational power and the amount of data available from web, video cameras, laboratory measurements.

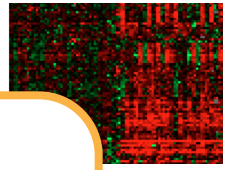
Images & Video

Text & Language

Speech & Audio



Gene Expression

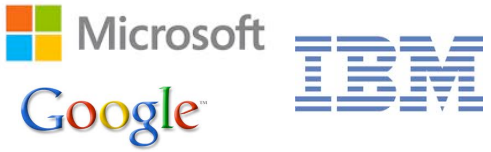


Deep Learning Models that support inference and discover structure at multiple levels.

- Develop statistical models that can discover underlying structure, cause, or statistical correlation from data.
- Multiple application domains.

Impact of Deep Learning

- Speech Recognition



- Computer Vision

- Recommender Systems

- Language Understanding

- Drug Discovery and Medical Image Analysis

Understanding Images:

TAGS:

strangers, coworkers, conventioneers, attendants, patrons

Nearest Neighbor Sentence:

people taking pictures of a crazy person

Model Samples

- a group of people in a crowded area .
- a group of people are walking and talking .
- a group of people, standing around and talking .
- a group of people that are in the outside .

Caption Generation

a car is parked in
the middle of nowhere .

a wooden table and chairs
arranged in a room .

there is a cat sitting on a shelf .

a ferry boat on a marina
with a group of people .

a little boy with a bunch
of friends on the street .

Caption Generation

the two birds are trying to be seen in the water . (can't count)

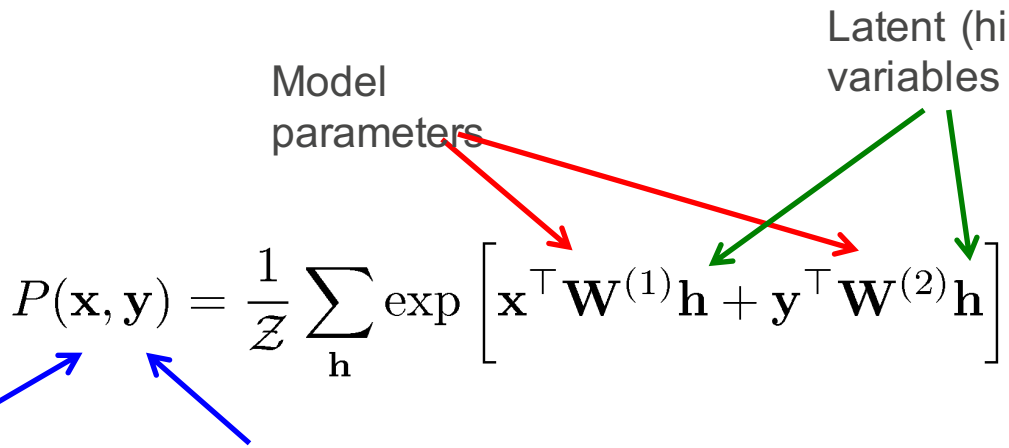
a giraffe is standing next to a fence in a field . (hallucination)

a parked car while driving down the road . (contradiction)

the handlebars are trying to ride a bike rack . (nonsensical)

a woman and a bottle of wine in a garden . (gender)

Example: Boltzmann Machine



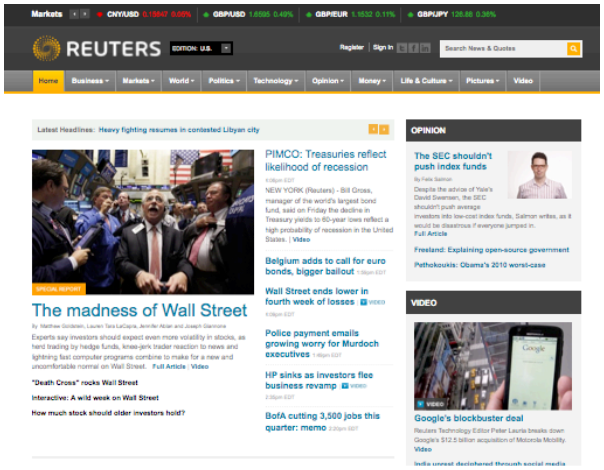
Markov Random Fields, Undirected Graphical Models.

Finding Structure in Data

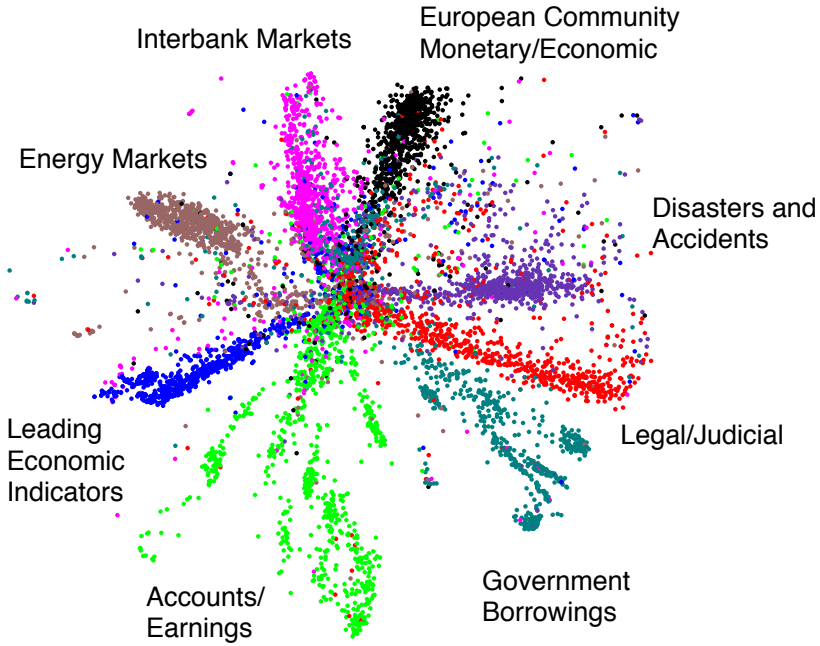
$$P(\mathbf{x}) = \frac{1}{\mathcal{Z}} \sum_{\mathbf{h}} \exp\left[\mathbf{x}^\top \mathbf{W} \mathbf{h}\right]$$

Vector of word counts on a webpage

Latent variables: hidden topics



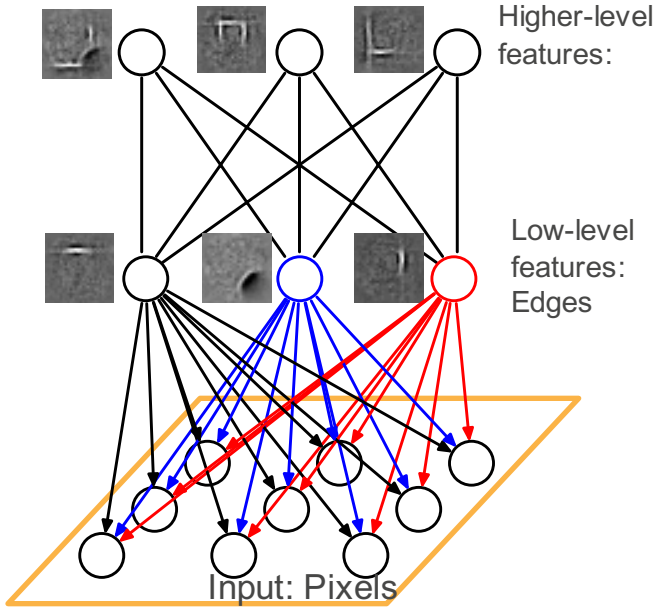
804,414 newswire stories



Important Results

- Deep Belief Networks, 2006 (Unsupervised)

  Hinton, G. E., Osindero, S. and Teh, Y., A fast learning algorithm for deep belief nets, Neural Computation, 2006.



Theoretical Breakthrough:

- Adding additional layers improves variational lower-bound.

Efficient Learning and Inference with multiple layers:

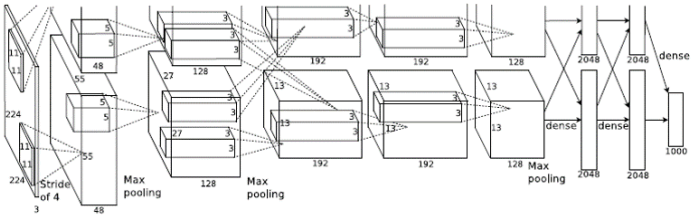
- Efficient greedy layer-by-layer learning learning algorithm.

- Inferring the states of the hidden variables in the top most layer is easy.

Important Results

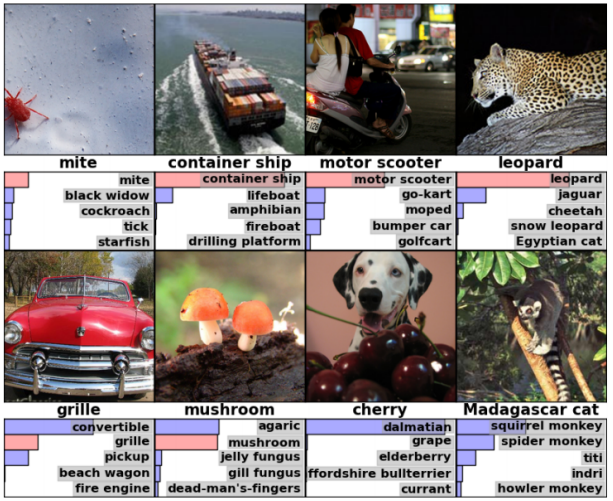
- Deep Convolutional Nets for Vision (Supervised)

  Krizhevsky, A., Sutskever, I. and Hinton, G. E., ImageNet Classification with Deep Convolutional Neural Networks, NIPS, 2012.



IMAGENET

1.2 million training images
1000 classes

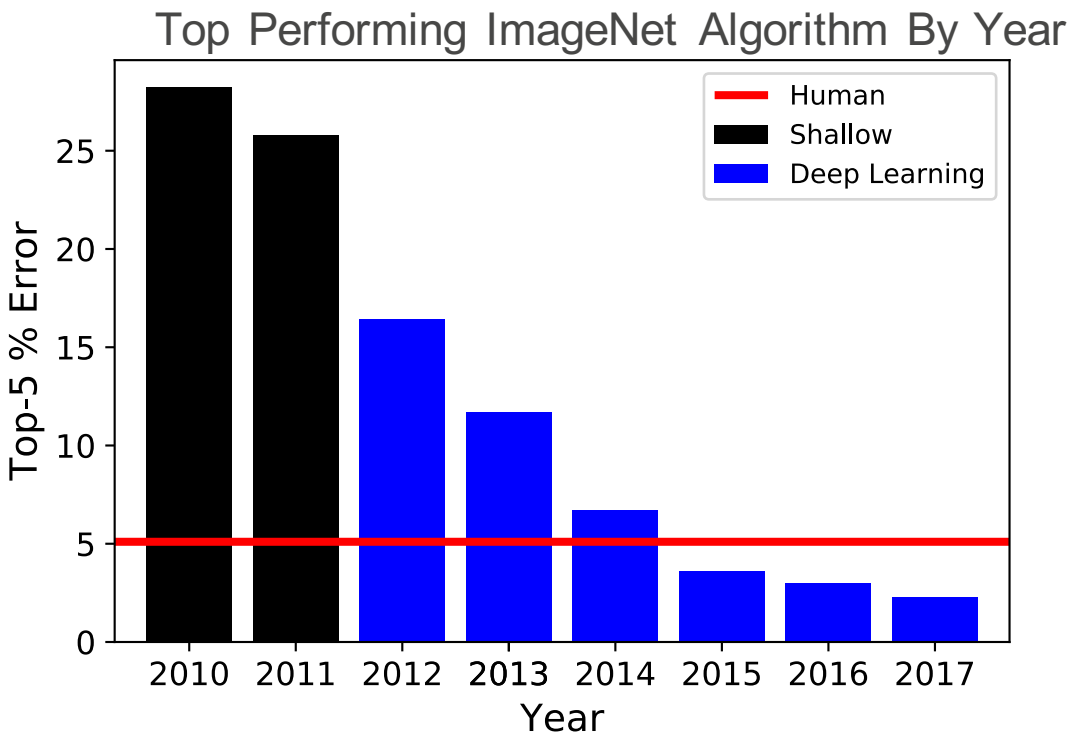


- Deep Nets for Speech (Supervised)

  Hinton et. al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups, IEEE Signal Processing Magazine. 2012.

Deep Learning can surpass human performance

- For ImageNet:
  - Deep Learning was a *huge* jump forward
  - State-of-the-art systems **significantly outperform humans** on the same task
- These use "Convolutional Neural Networks," which we quickly get to



Statistical Generative Models

Training Data(CelebA)

Model Samples (Karras et.al., 2018)

4 years of progression on Faces

Brundage et al., 2017

Statistical Generative Models

Conditional generative model P(zebra images| horse images)

- Style Transfer

Input Image

Monet

Van Gogh

Zhou el al., Cycle GAN 2017

Statistical Generative Models

- Conditional generative model P(zebra imagesl horse images)

- Failure Case

Zhou el al., Cycle GAN 2017

Statistical Generative Models

Source actor

Target actor

Real-time reenactment

[Face2Face, Thies et al, CVPR 2016]

Deep Learning beats human performance in many tasks

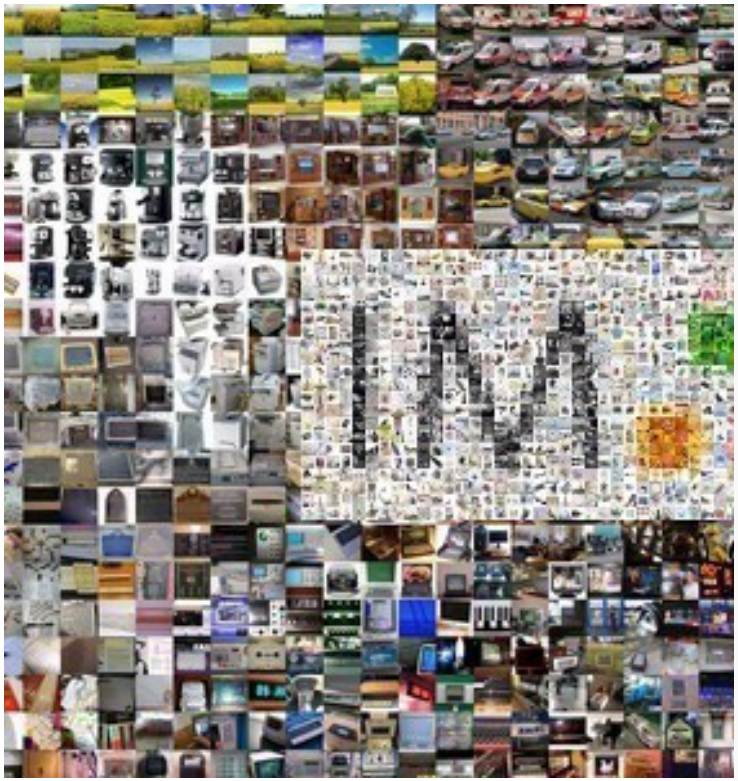
- Famously, Google DeepMind trained "AlphaGo" to beat the world champion Go player (a complex game)
- AlphaGo's largely works by Deep Learning techniques (learned by repeatedly playing the game)
- Many other examples:
  - Voice Recognition
  - Object Detection
  - Text Translation
  - Etc
- So what is Deep Learning? And how does it work
  - Will be answered as we go through the course

Some Applications

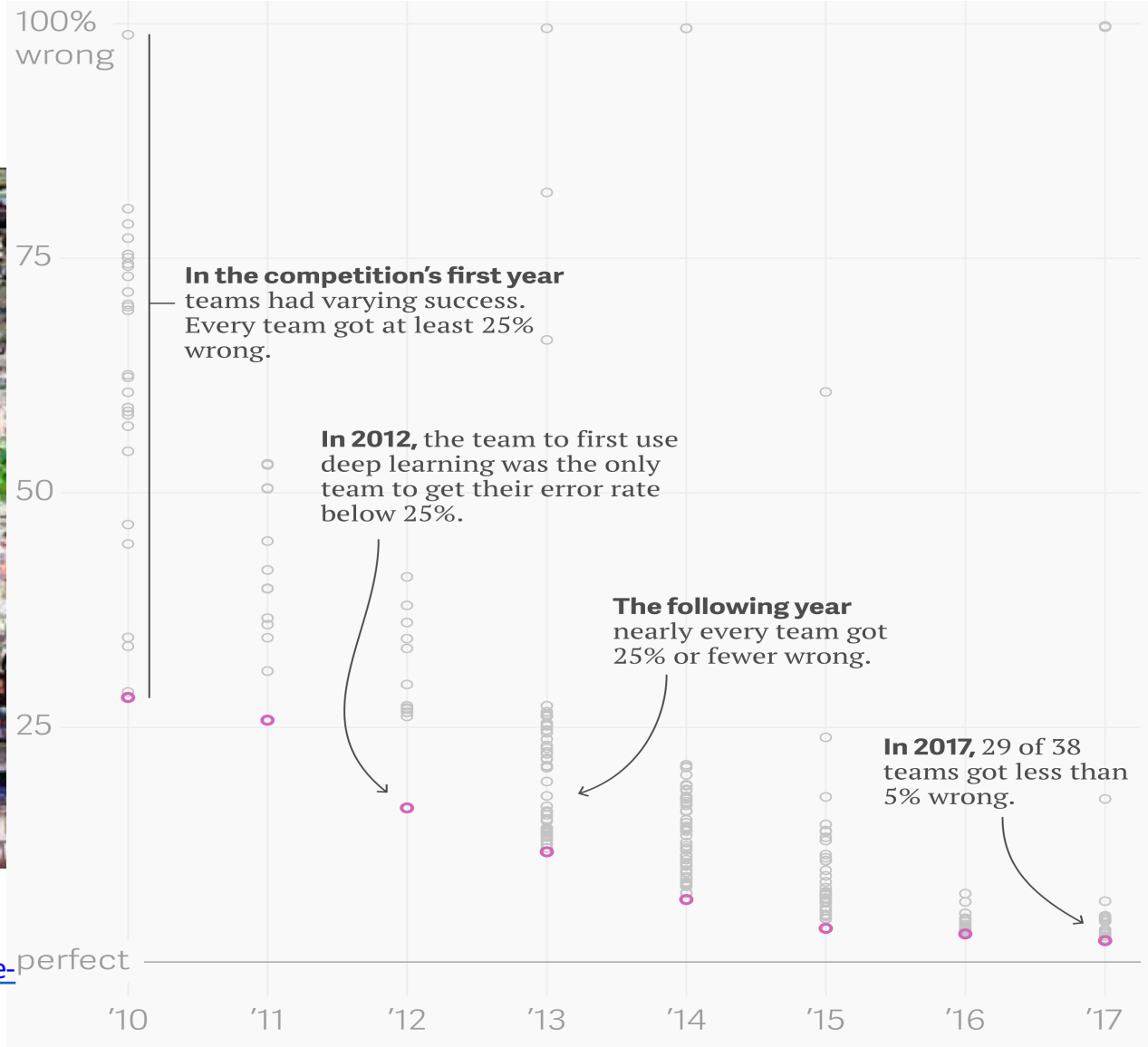
Classify Images

http://www.image-net.org/

Classify Images



Yanofsky, Quartz
https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/



Detect and Segment Objects

https://github.com/matterport/Mask_RCNN

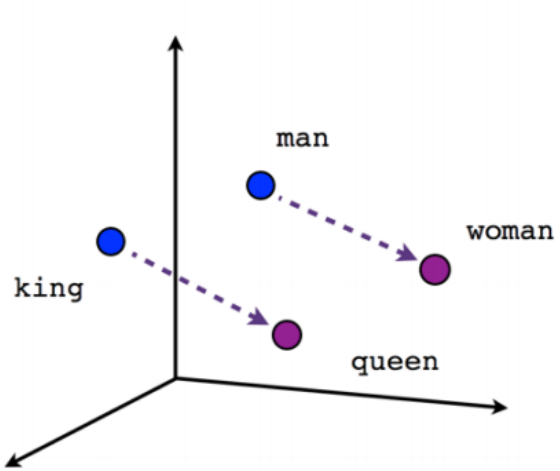
Style transfer

https://github.com/zhanghang1989/MXNet-Gluon-Style-Transfer/

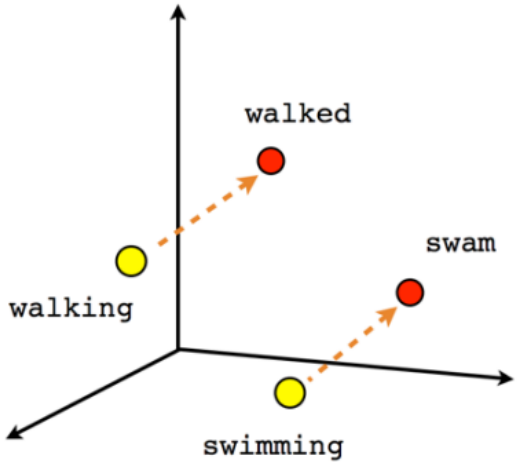
Synthesize Faces

Karras et al, ICLR 2018

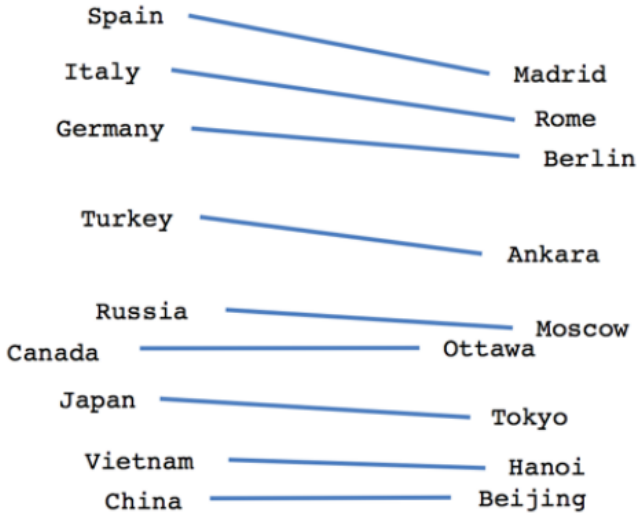
Analogies



Male-Female



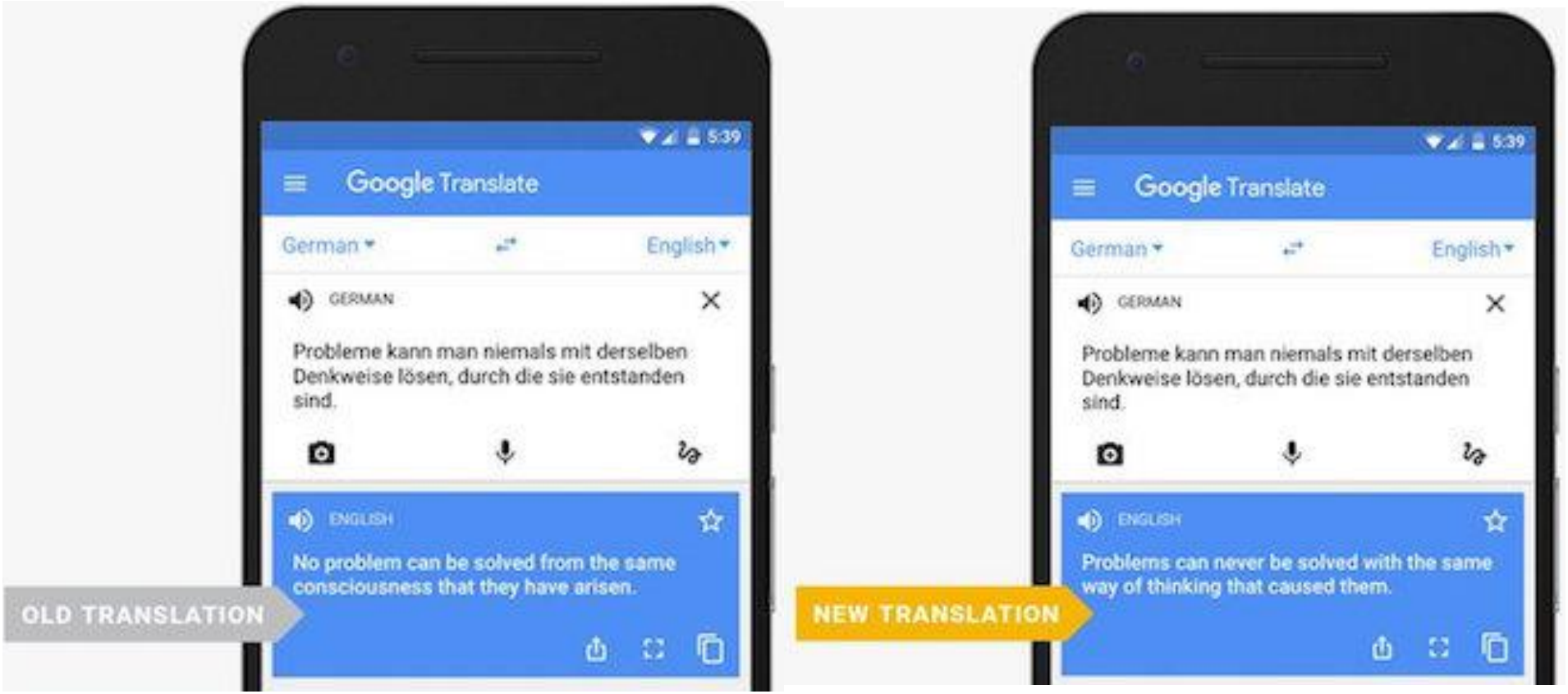
Verb tense



Country-Capital

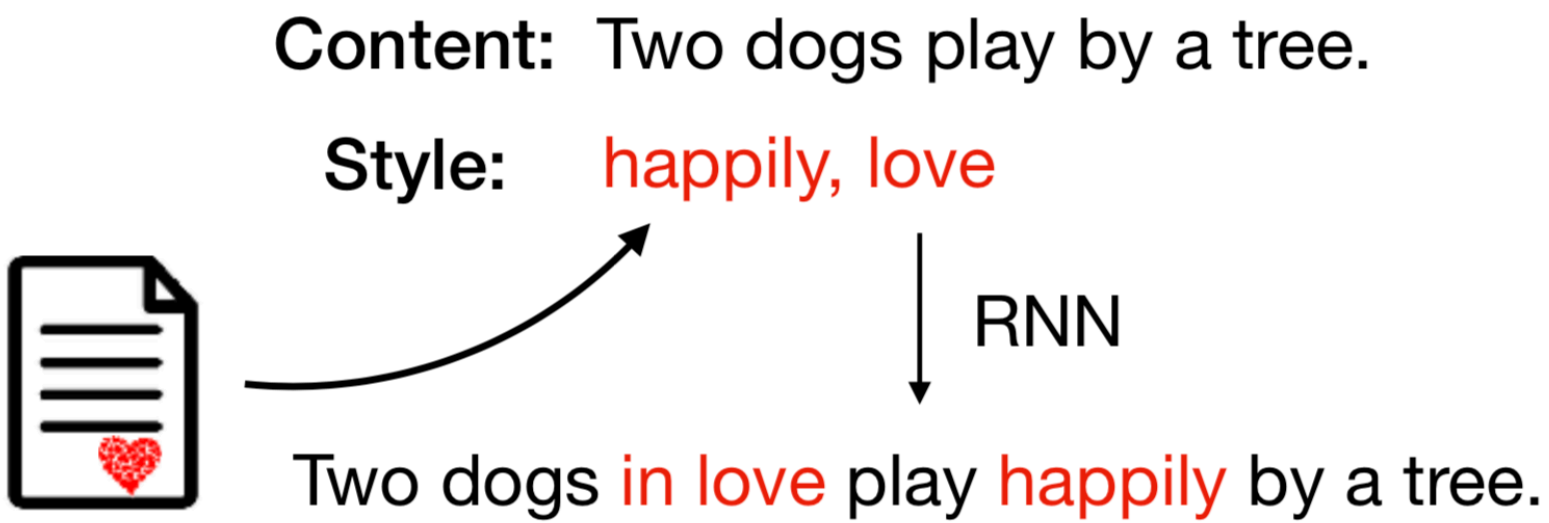
https://www.tensorflow.org/tutorials/word2vec

Machine Translation



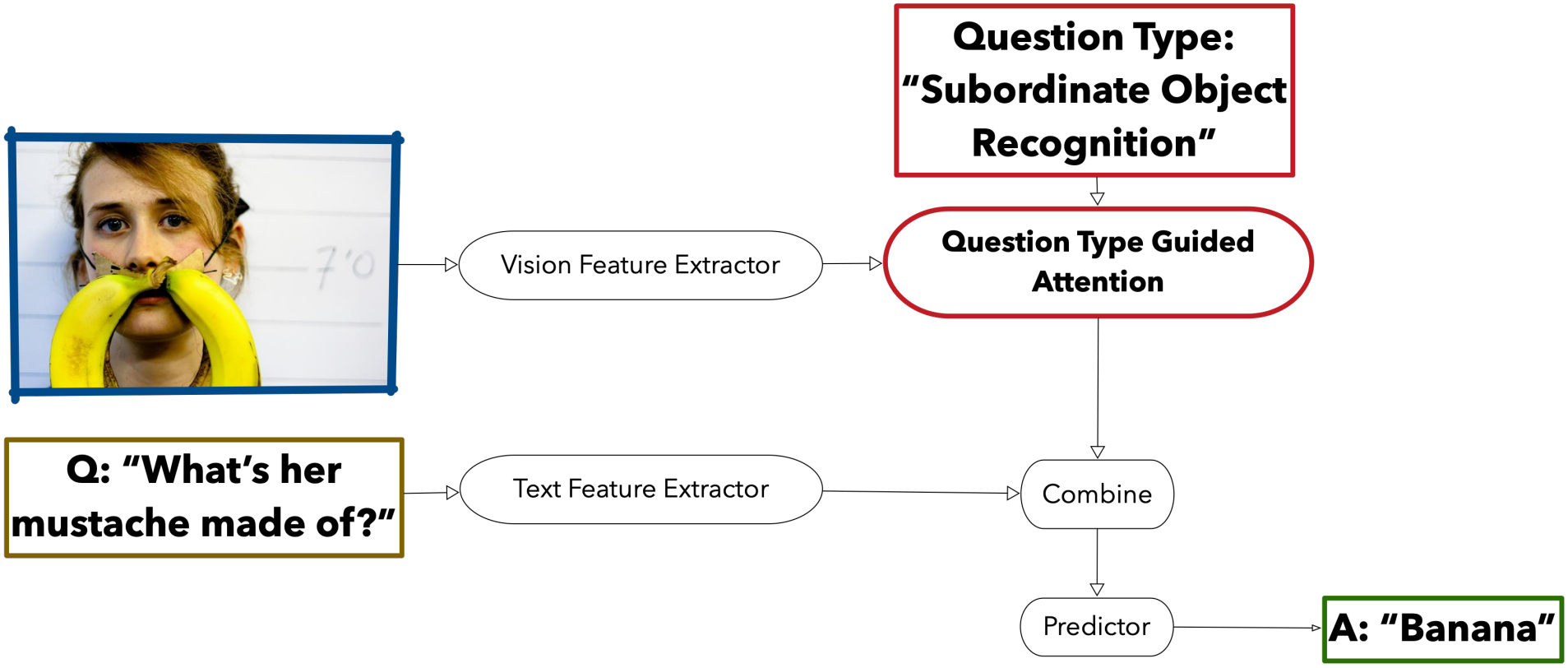
https://www.pcmag.com/news/349610/google-expands-neural-networks-for-language-translation

Text synthesis



Li et al, NACCL, 2018

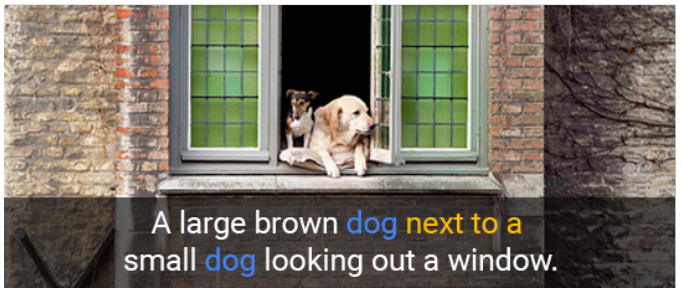
Question answering



Shi et al, 2018, Arxiv

Image captioning

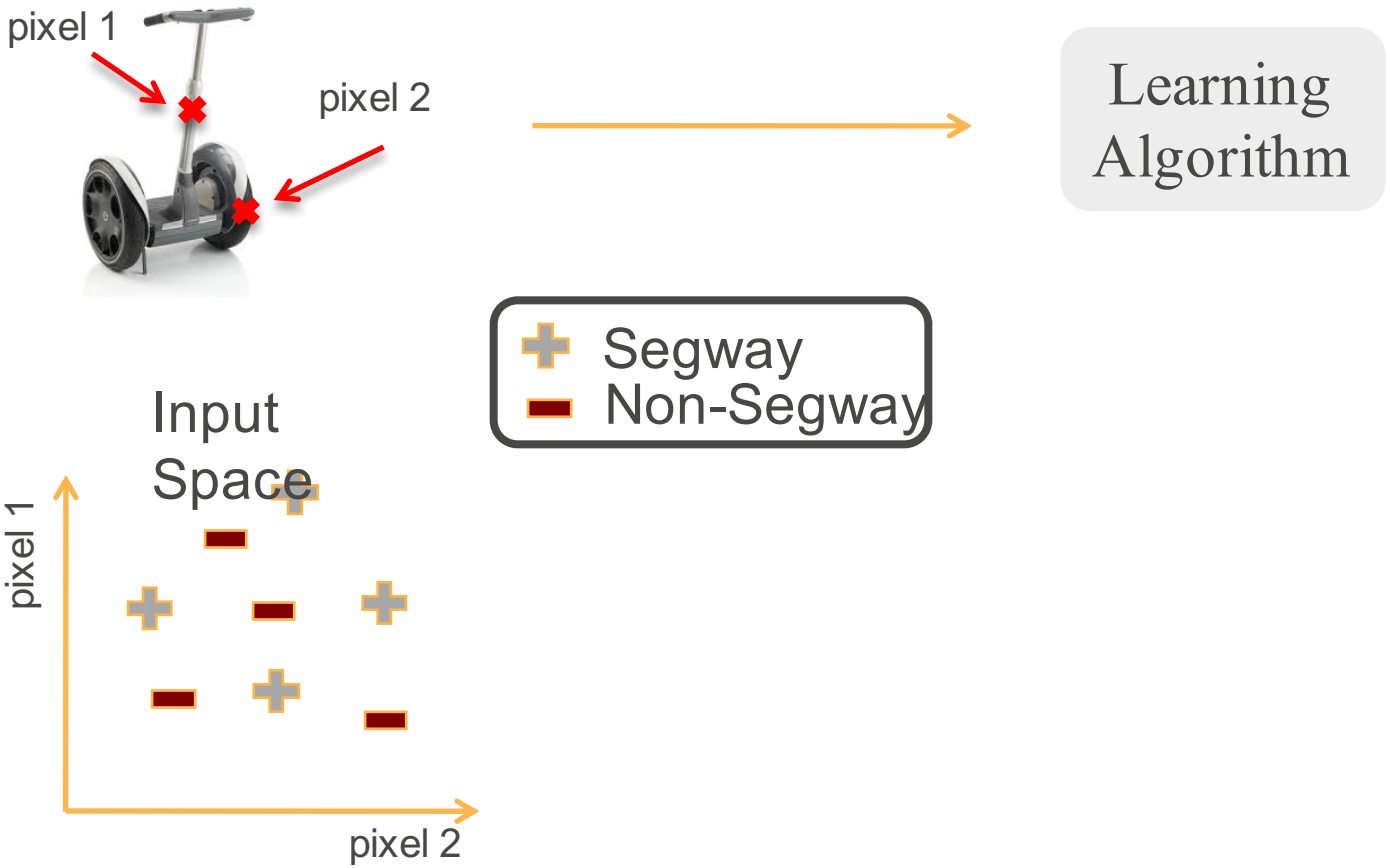
Human captions from the training set



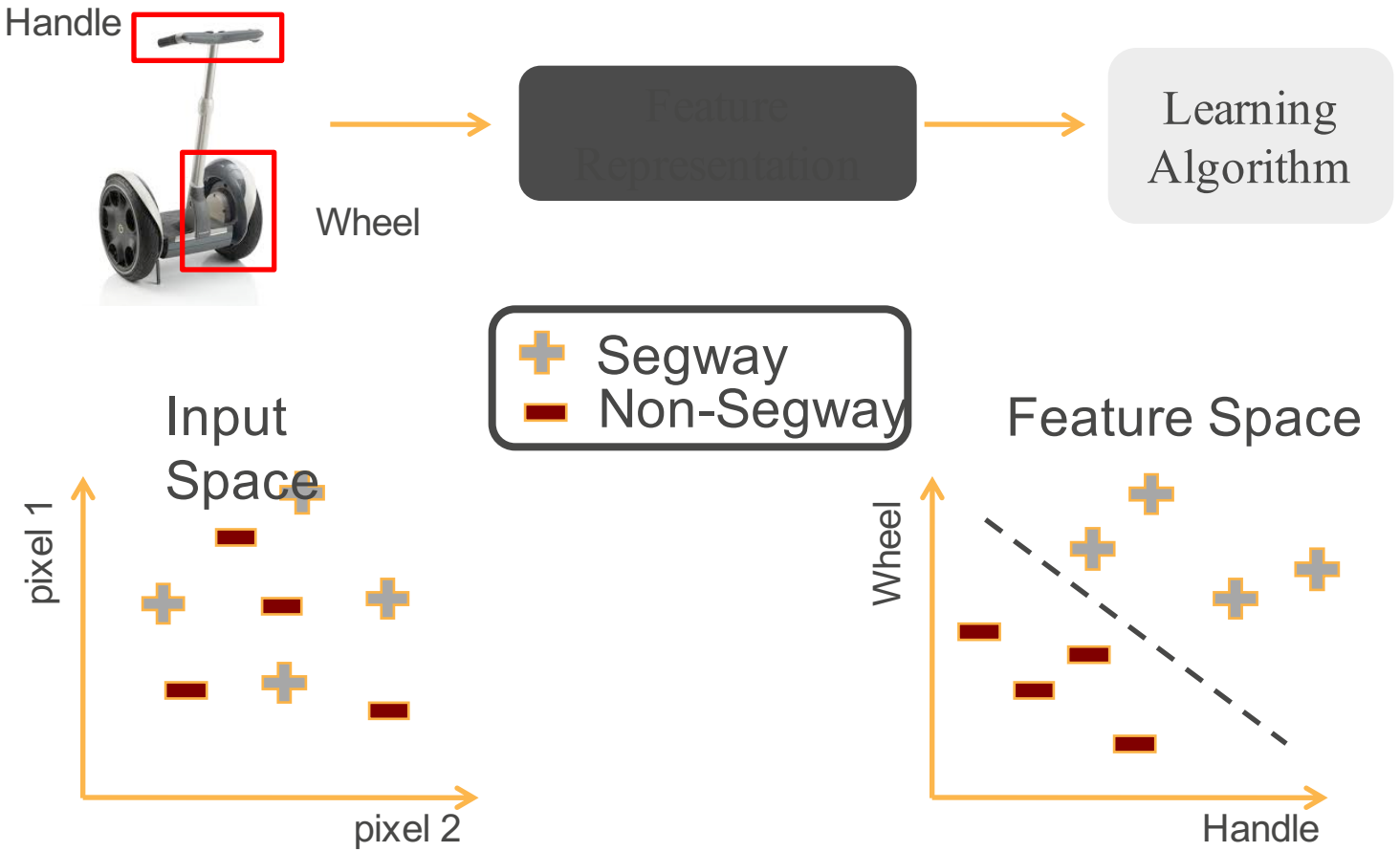
Automatically captioned

Shallue et al, 2016
https://ai.googleblog.com/2016/09/show-and-tell-image-captioning-open.html

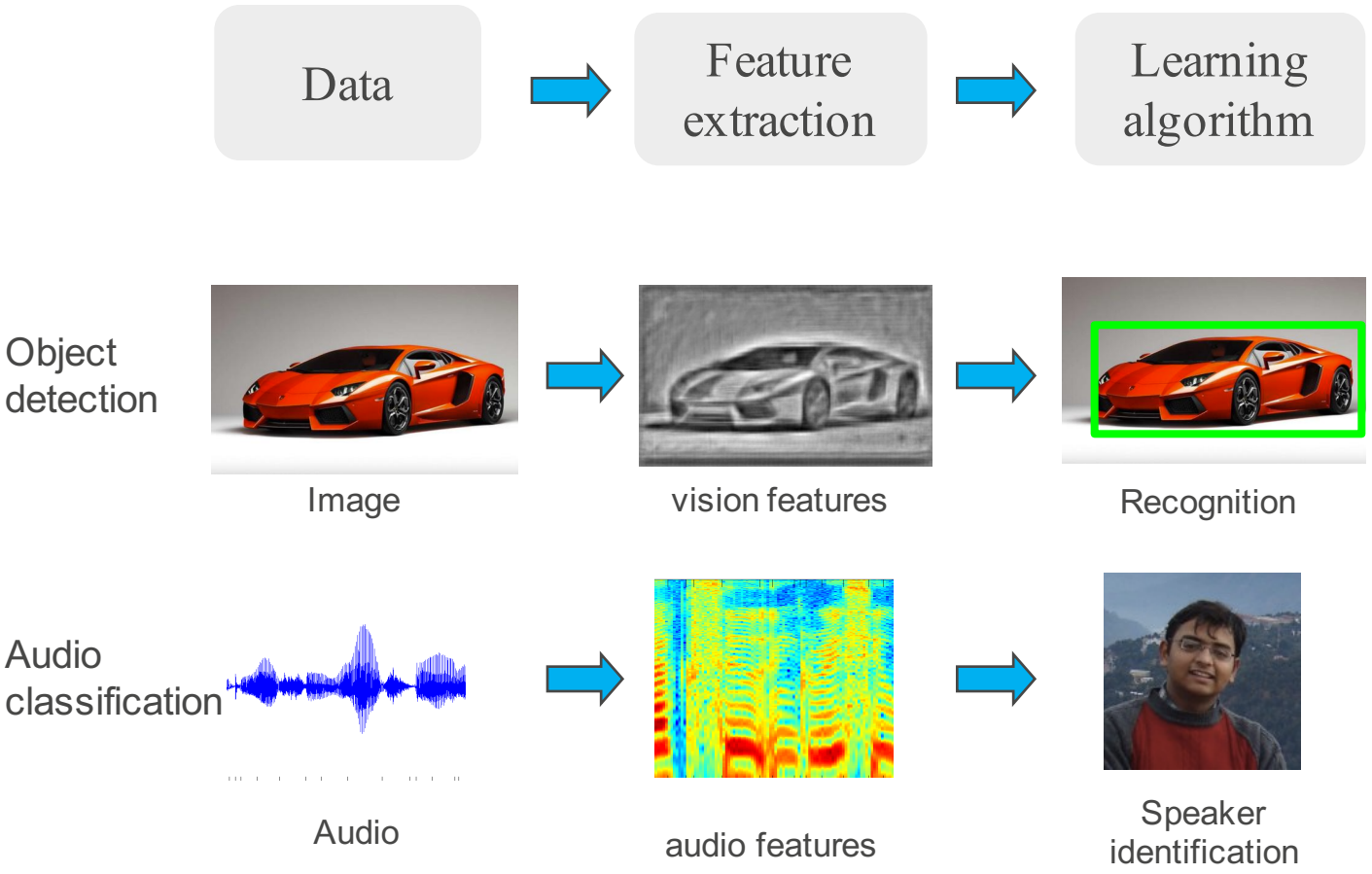
Learning Feature Representations



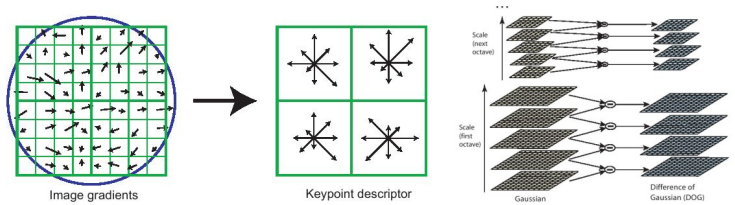
Learning Feature Representations



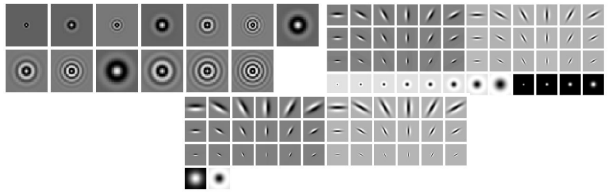
Traditional Approaches



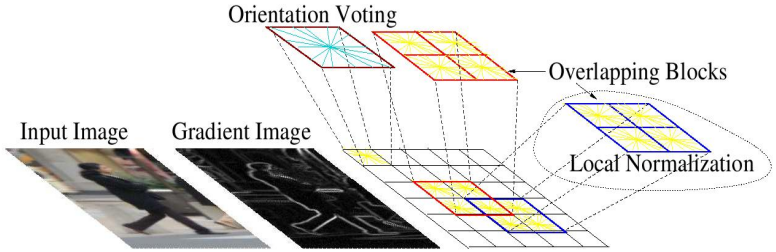
Computer Vision Features



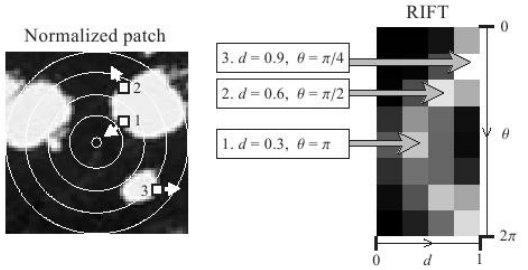
SIFT (Scale Invariant FT)



Textons

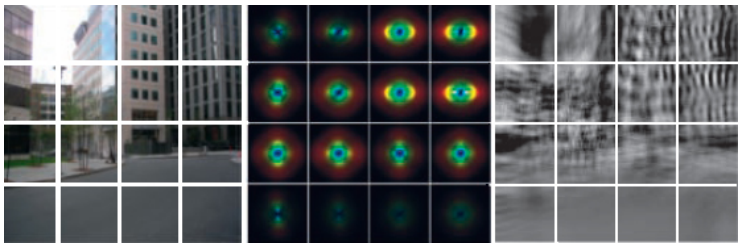


HoG

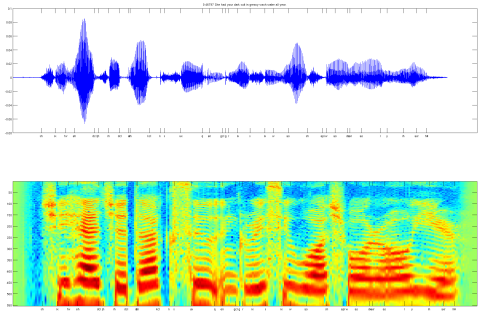


RIFT

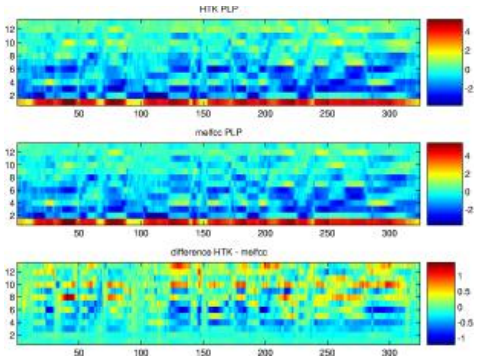
GIST



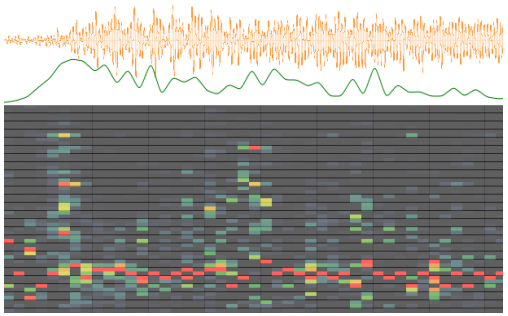
Audio Features



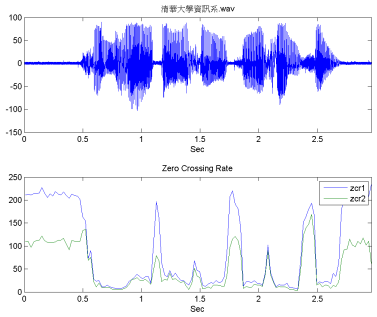
Spectrogram



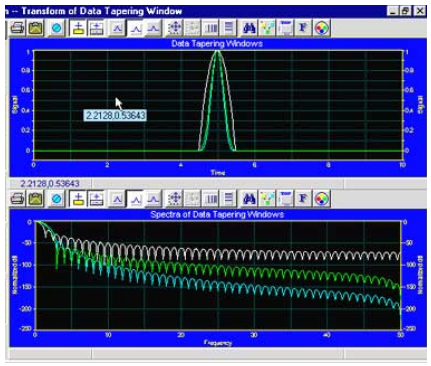
MFCC



Flux

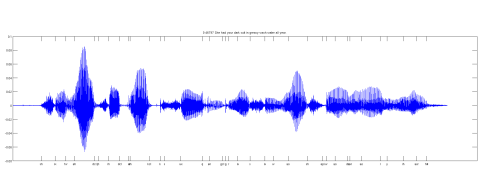


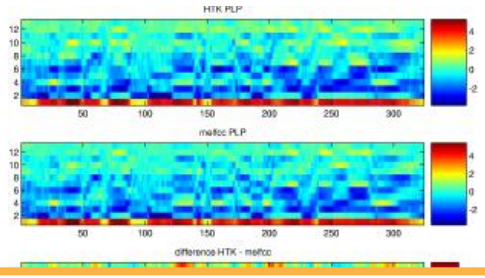
ZCR



Rolloff

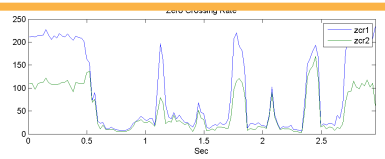
Audio Features



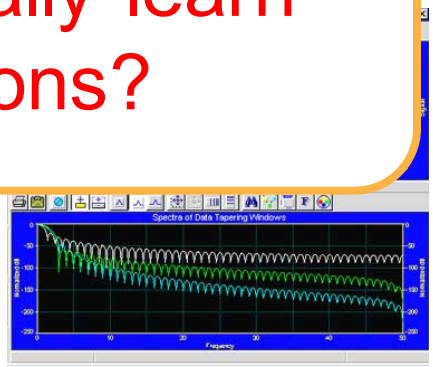


Representation Learning:
Can we automatically learn these representations?

Flux



ZCR



Rolloff