Introduction to Deep Learning

Vahid Tarokh

ECE 685D, Fall 2025

INSTRUCTOR & TA INFORMATION

Instructor

Vahid Tarokh

- Electrical and Computer Engineering

My research is on developing methodologies for data analysis.

Teaching Assistants

Cat Le
Postdoc
cat.le@duke.edu

Haoyu Gong
Graduate Student
haoyu.gong@duke.edu

Haoming Yang
PhD Student
haoming.yang@duke.edu

Teaching Assistants

Mehak Arora
PhD Student
mehak.arora@duke.edu

Zihao Wu
PhD Student
zihao.wu@duke.edu

Junyi Liao
PhD Student
junyi.liao@duke.edu

Office Hours

	In-person	Location	Online & Email
Vahid	Wednesdays, 11:30 AM – 1:00 PM	Wilkinson 425	Monday, 1:00 PM – 2:00 PM via https://duke.zoom.us/j/9196607594 vahid.tarokh@duke.edu
Cat	Please reach out by email to schedule an appointment.		TBD https://duke.zoom.us/j/93686122572 cat.le@duke.edu
Junyi	See Discussion Section Information for detail.		TBD junyi.liao@duke.edu
Haoming	See Discussion Section Information for detail.		TBD haoming.yang@duke.edu

- Any changes to this will be announced on Canvas in due time

	In-person	Location	Online & Email
Haoyu	TBD	TBD	TBD haoyu.gong@duke.edu
Mehak	TBD	TBD	TBD mehak.arora@duke.edu
Zihao	TBD	TBD	TBD zihao.wu@duke.edu

- Any changes to this will be announced on Canvas in due time

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 Canvas site for:

- ECE 685D/CS675D

Primary mode of communication

- Announcements
- Lecture note templates posted
- Sections
- Homework posted
- Questions answered

Syllabus Highlights

Grading breakdown **For SAT/UNSAT grades:**

- **Assignments** (50 points): Will include both Theory and Programming Assignments. There will be 5 problem sets.

- **Project** (50 points)

Syllabus Highlights

Grading breakdown **For Letter Grades:**

- Assignments (50 points): Will include both Theory and Programming Assignments. There will be 5 problem sets.
- Exam I (10 points). Date: Oct. 6
- Project (20 points): Due Date 12/05/2025 at 23:59:59 EST
- Exam II (20 points). Date: Nov. 19

A+:	98-100
A:	94-97.9
A-:	90-93.9
B+:	87-89.9
B:	83-86.9
B-:	80-82.9
C+:	77-79.9
C:	73-76.9
C-:	70-72.9
D+:	67-69.9
D:	63-66.9
D-:	60-62.9
F:	0-59.9

Instruction from Duke Administrators

- I am sharing new instructions from Duke Administrators:
- ……. However, when everyone in your class gets an A, you may give students a false sense of their achievements and lead them to expect these grades in all their other classes.
- This can create additional 5 complaints in other classes when professors distinguish more finely among the gradations of student achievement.

Regrading Policy

- We will regrade HWs, Projects and Exams only if we receive a written request (by email) within 48 hours after you received your original grade.
- Per Duke policy, I (<u>Not the TAs</u>) am the ultimate arbiter of final grades.

Regrading Policy

- We will regrade HWs, Projects and Exams only if we receive a written request (by email) within 48 hours after you received your original grade.
- Per Duke policy, I (<u>not the TAs</u>) am the ultimate arbiter of final grades.

Syllabus Highlights

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 Canvas**
- Code will only be accepted in Python

Syllabus Highlights: Project

- The second half of the course will focus on an individual or group project, which consists of applying an *appropriate* deep learning tool to a dataset and analysis of the results.
- A **final report** will be due on 12/05/2025 at 23:59:59 EST.
- This must be typed using a style file we will provide.
- There is a maximum limit of 8 pages for the final report.
- **This deadline will not be extended**.

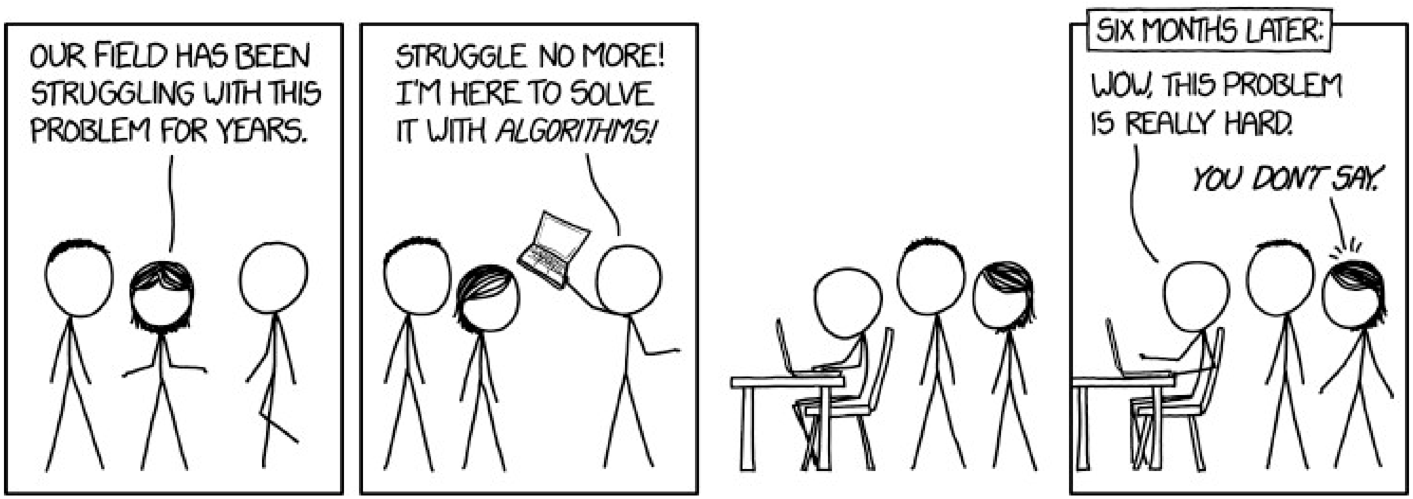
Course Project

- Can create a team of 2-4 students per project.
- Each group will be assigned to a TA.
- We will assign the projects and clarify what must be delivered.
- We will also group the team members
    - We will use a random method.

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.
- If one of the team members is not contributing to the team, the remaining members of the group must inform the Instructor and the TA in charge of that group.
- We will analyze the situation and may assign an individual project to that team member.

You will 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.

Academic Integrity

All students are expected to abide by the Duke Community Standard

- Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability.
- Members of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity.
- Duke University has high expectations for students' scholarship and conduct.
- In accepting admission, students indicate their willingness to subscribe to and be governed by the rules and regulations of the university, which flow from the Duke Community Standard (DCS).

Academic Integrity

- Regardless of course delivery format, it is the responsibility of all students to understand and follow all Duke policies, including but not limited to the academic integrity policy (e.g., completing one's own work, following proper citation of sources, adhering to guidance around group work projects, and more).
- Ignoring these requirements is a violation of the DCS. Students can direct any questions or concerns regarding academic integrity to the Office of Student Conduct and Community Standards at conduct@duke.edu and can access the DCS guide at https://dukecommunitystandard.students.duke.edu/.

Academic Integrity

All students are expected
to abide by the
Duke Community Standard

Mental Health and Well-being Resources

- Duke is committed to holistic student wellbeing, which includes one's mental, emotional, and physical health. The university offers resources to help students manage daily stress, to encourage intentional self-care, and to access just-in-time support.
- If you find you need support, your mental and/or emotional health concerns are impacting your day-to-day activities, your academic performance, or you need someone to talk to, the resources below are available to you:
- **DukeReach:** provides comprehensive outreach services to support students who are experiencing significant challenges related to mental health, physical health, social adjustment, and/or a variety of other stressors. You can contact the DukeReach team at dukereach@duke.edu.

Mental Health and Wellbeing Resources

- **Counseling and Psychological Services (CAPS)**: CAPS offers counseling services to Duke students including virtual appointments, and referrals in the community. You do notneed an appointment for an initial assessment. You may walk in or call 919-660-1000 to get started. Hours: Monday-Friday 9:00am - 4:00pm. After hours counseling services are available at no additional cost to students, you can call: 919-660-1000 Option 2.
- **TimelyCare** (formerly known as Blue Devils Care): TimelyCare is an online platform thatis a convenient, confidential, and free way for Duke students to receive 24/7 mental health support through TalkNow and scheduled counseling.
- **Duke Student Health**: Student Health offers a wide range of healthcare services for all Duke students, many of which are covered by the student health fee. To make an appointment call (919) 681-9355. Hours: Monday - Friday, 8am - 4:30pm, Thursday 9am - 4:30pm. Closed from 12-12:30 each day.

Attendance Policy

- Active participation is expected in this course. If you miss class, you are responsible for staying up to date and reviewing notes/watching Lecture videos from class-mates.
- Absences due to illness, emergency, or university-related obligations are excused when you notify me in advance (24 hours before the class if possible).
- We do not run a class attendance sheet.
  - However, if you want to request *an attendance letter* from me to provide to USCIS or other US government agencies, you must notify me at the beginning of semester by email (so that I can verify your attendance to report to the US Government).

Sick Absences

- Please ***do not come to class if you are sick*** to keep the university community as safe and healthy as possible.
    - However, please inform me of your absence and plan to complete any missed work.
- Students who encounter short- and long-term medical issues or 6 instances of personal distress or emergency can seek academic support if needed.
- You can find details regarding university sick leave policies here:
    - Trinity College: Health Issues, Short- and Long-Term

      https://trinity.duke.edu/undergraduate/academic-policies/illness#:~:text=For%20a%20short%2Dterm%20illness,to%20discuss%20accommodations%20and%20support
    - Pratt School: Missing Class, Illness, Short- and Long-Term

      https://pratt.duke.edu/life/resources/undergrad/policies/

Inclement Weather

- In the event of inclement weather or other connectivity-related events that prohibit class attendance, I will notify you how we will make up missed course content and work.
- Asynchronous catch-up methods may apply, and we may rely on Duke’s designated make-up days.

QUICK NOTES ON CODING FRAMEWORKS

Sections

- Sections review programming languages, platforms, etc.
- Three Discussion Sections:
    - You must be registered in one.
- We may 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.

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
  - Please <u>plan ahead</u>. No extensions will be given for waiting on resources.

COURSE MATERIAL

MAY BE REVISED BASED ON FEEDBACK

Course Material

- Introduction: What is Machine Learning? Data Sets
- Mathematical Background: Linear Algebra, Calculus, Probability and Statistics
- Linear and Logistic Regression
- Multi-Layer Perceptron (MLP) Networks: Weights, Biases, Initialization, Non-linear activation, Loss functions,
- Back-propagation: Chain rule of multivariable calculus

Course Material

- Optimization for Training Deep Networks: Stochastic Gradient Descent, Nestrov Acceleration Methods, Stochastic Methods with Momentum, Adagrad, RMSProp, Adam, Second Order Methods, Learning Rate
- Underfitting, Overfitting, Training Tricks: Bias and Variance Trade-off, Underfitting, Overfitting, Regularization, Weight-decay (l_2-norm), Sparsity and l_1-norm, Dropout, Early Stopping, Image Augmentation

Course Material

- Convolutional Neural Networks (CNNs): Biological inspiration, Receptive Fields, Parameter Sharing, Convolution and Correlation Operation, Multiple Input and Output Channels, Filters and Feature Maps, Stride and Padding, Pooling, Max-pooling, Average Pooling, Batch-Normalization, Back-propagation for CNNs, Modern Convolutional Neural Networks
- Applications of CNNs to Computer Vision: Image Classification and Detection

Course Material

- Generative Models I: Naive Bayes and LDA, Graphical Models, Directed Graphical Models, Undirected Graphical Models, Hidden Markov Models (HMM), Linear Factor Models and Factor Analysis, PCA, Probabilistic-PCA, Slow Feature Analysis, ICA, Sparse Coding and Dictionary Learning
- Generative Models II: Restricted Boltzmann Machine, Deep Belief Networks
- Generative Models III: Auto-Encoders, Variational Autoencoders, Importance Weighted Autoencoders (IWAE), Conditional VAEs,
- Generative Models IV: Generative Adversarial Networks (GANs), Vanilla-GAN, DCGAN, Conditional GAN, InfoGAN, f-GAN, CycleGAN, Energy-Based GAN, Coupled GAN

Course Material

- Recurrent Neural Networks (RNNs) and Time Series: Sequential Data sets, Time Series and Prediction, ARIMA Models, RNNs Architecture, Gated Recurrent Unit (GRU), Long-Short-Term-Memory (LSTM), Deep RNNs, Bi-directional RNNs
- Introduction to Natural Language Processing (NLP) and RNNs, Text Processing, Language Models , Attention Mechanisms, Sequence-to-Sequence Models

SOME DATA SETS

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-100 dataset (a.k.a. CIFAR-10 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 popular 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 36^{th} 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 on 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

INTRODUCTION

Machine Learning

- Machine learning (ML) is a subset of AI.
- Build models based on training data in order to make predictions or decisions without being explicitly programmed to do so.
- Closely related to
    - Applied Computational Statistics
    - Optimization Algorithms
    - Learning Theory
    - Data Science
- In some cases, some domain knowledge may be used in design of models.

Machine Learning Efforts at Duke

- Numerous research groups at Duke and elsewhere focus on data science and machine learning.
- New initiative for Health Data Science at Duke (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/

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.

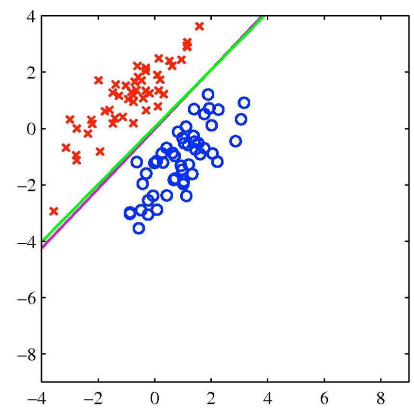
- **Unsupervised Learning:** The goal is to build a statistical model of **x**, which can be used for making predictions, decisions.

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

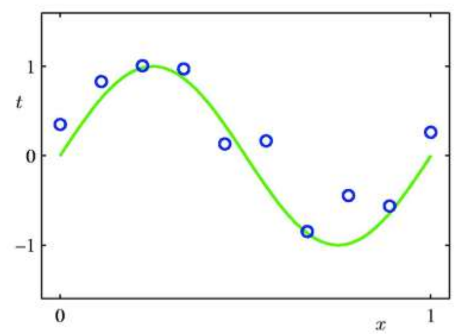
- **Reinforcement Learning:** the model (agent) produces a set of actions: $a_1, a_2, \ldots$ that affect the state of the world, and received rewards $r_1, r_2 \ldots$ The goal is to learn actions that maximize the reward.

Supervised Learning

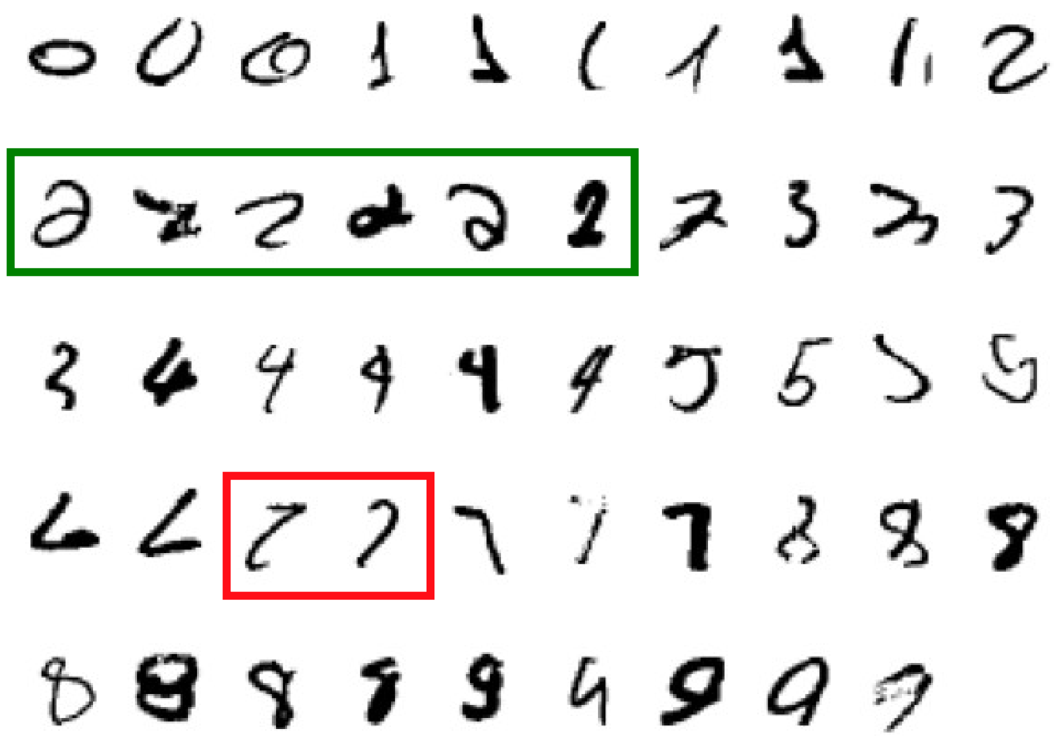
Classification: target outputs y_i are discrete class labels. The goal is to correctly classify new inputs.



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



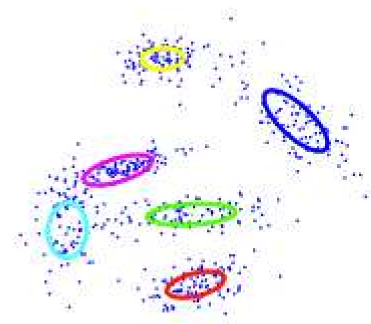
Handwritten Digit Classification



Unsupervised Learning

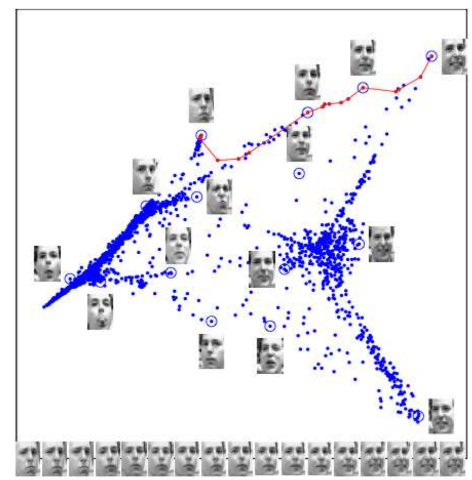
The goal is to construct statistical model that finds useful representation of data:

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

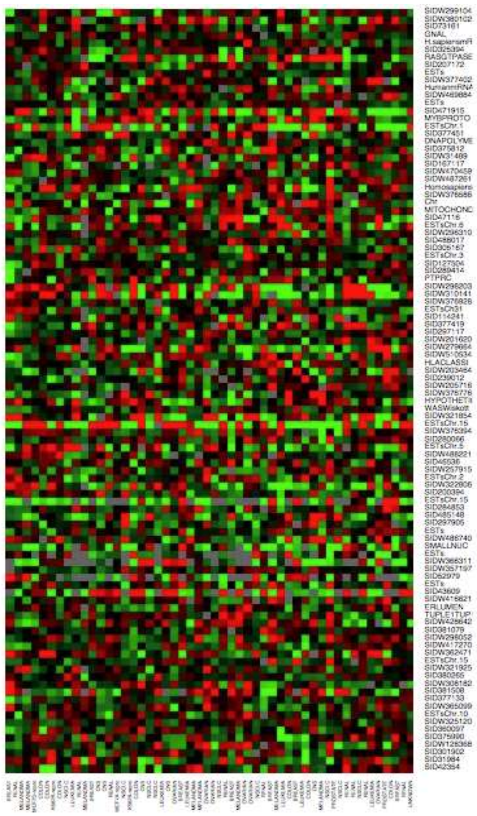


Unsupervised Learning can be used for:

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



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.

Features and Representation Learning

- Intuitively speaking a *feature* is an explanatory variable, a measurable property or characteristic of a phenomenon being observed.
- Choosing informative, discriminating and independent features is a crucial step for effective machine learning algorithm.
- Features are usually numerical, but structural/topological features have also been proposed.
- *Representation learning* is a set of techniques that allows a system to automatically discover the representations needed for feature detection from raw data.

Deep Learning

- Deep learning is a subset of machine learning.
- It is based on Artificial Neural Networks with representation learning.
- Examples of deep learning architectures we will discuss
    - Deep Feedforward Networks
        - Convolutional Neural Networks
    - Recurrent Neural Networks
    - Generative Neural Networks
- “Deep” in deep learning refers to the use of multiple layers in the network.

Deep Learning?

This class is focused on "Deep Learning:"

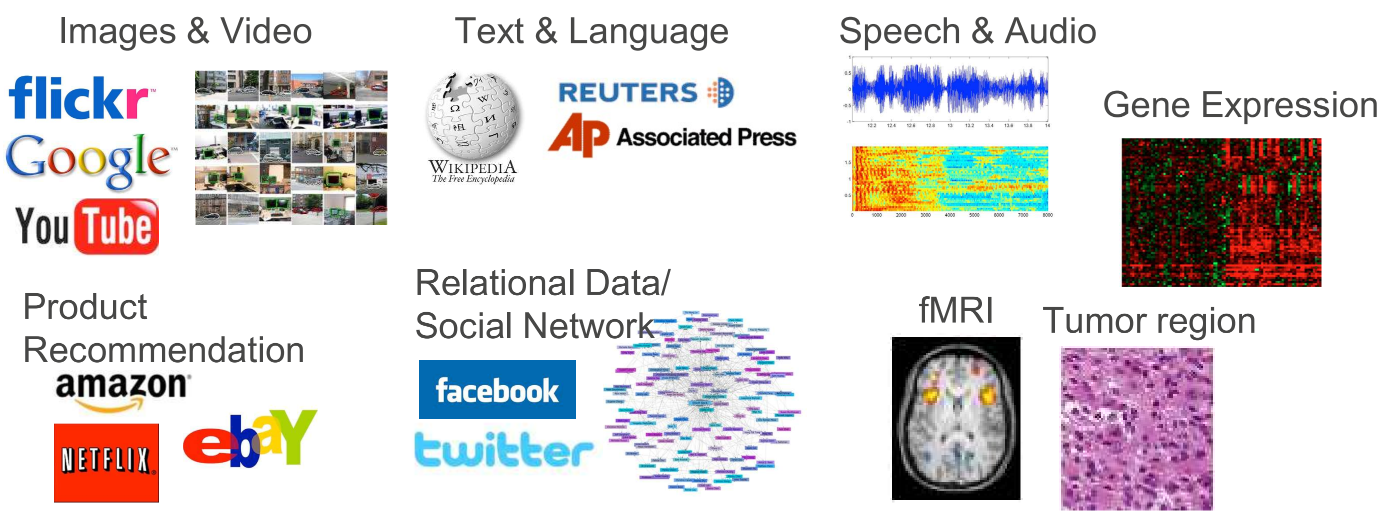
- What it is? (A Little bit of history)
- Is it supervised, unsupervised or RL?
- Some of its many applications?
- How to *build your own* deep networks?

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

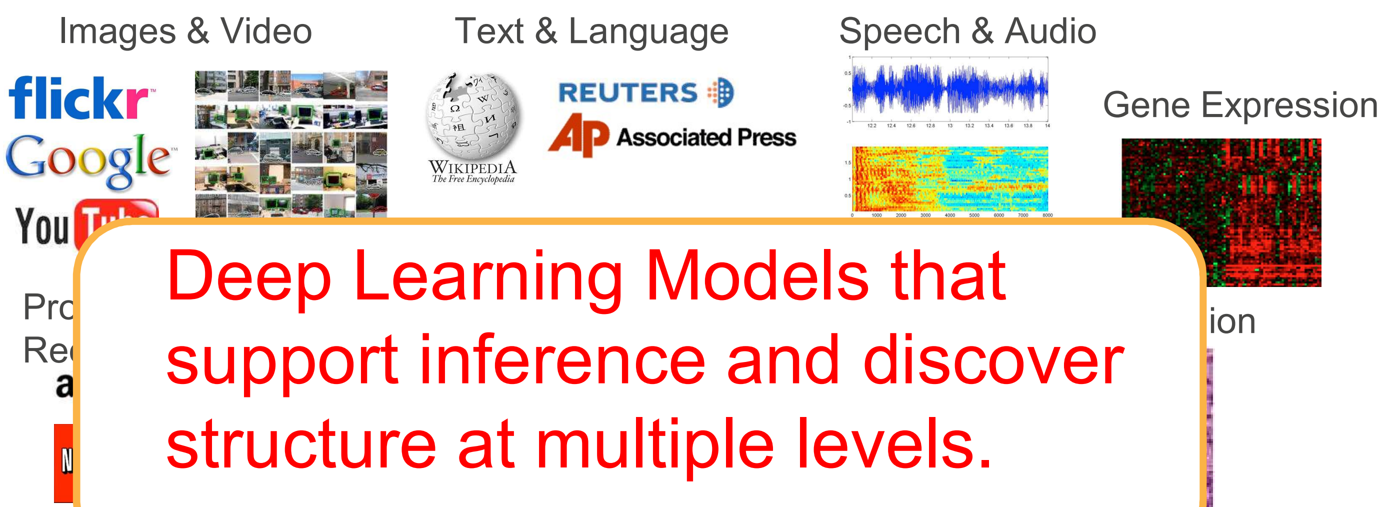
- Massive increase in both computational power and the amount of data available from web, video cameras, laboratory measurements.



- 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.



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
- 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

LZ

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

Model parameters

Latent (hidden) variables

$$P(\mathbf{x},\mathbf{y}) = \frac{1}{\mathcal{Z}} \sum_{\mathbf{h}} \exp\left[\mathbf{x}^\top \mathbf{W}^{(1)}\mathbf{h} + \mathbf{y}^\top \mathbf{W}^{(2)}\mathbf{h}\right]$$

Input data (e.g. pixel intensities of an image, words from webpages, speech signal).

Target variables (response) (e.g. class labels, categories, phonemes).

Markov Random Fields, Undirected Graphical Models.

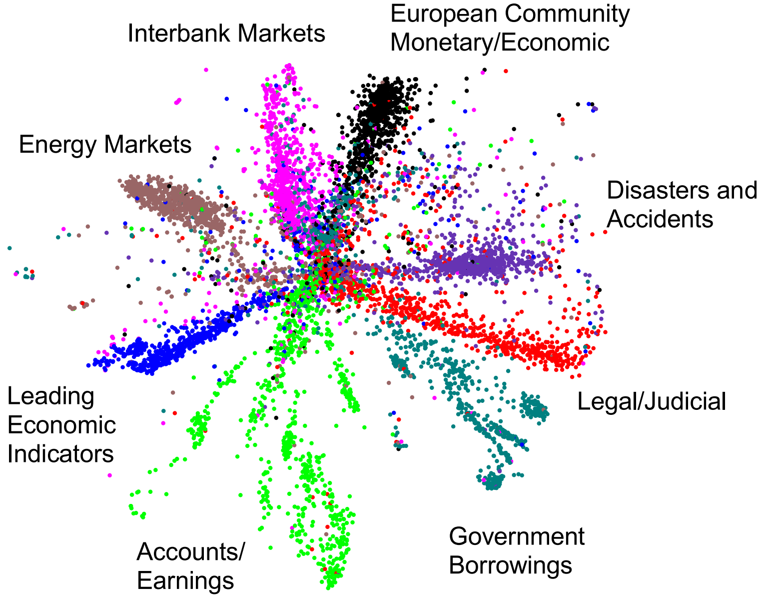
Finding Structure in Data

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

Vector of word counts on a webpage

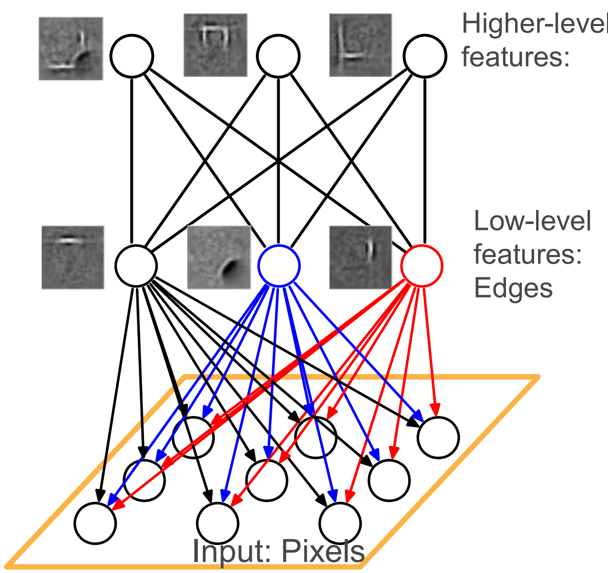
Latent variables: hidden topics

804,414 newswire stories



Important Results

- Deep Belief Networks, 2006 (Unsupervised)



Theoretical Breakthrough:

- Adding additional layers improves variational lower-bound.

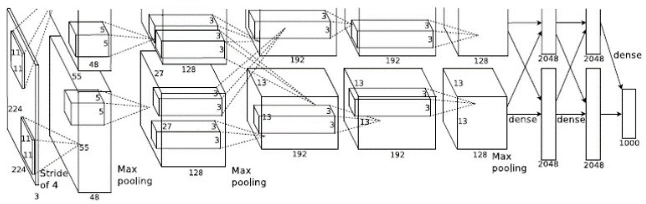
Efficient Learning and Inference with multiple layers:

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

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

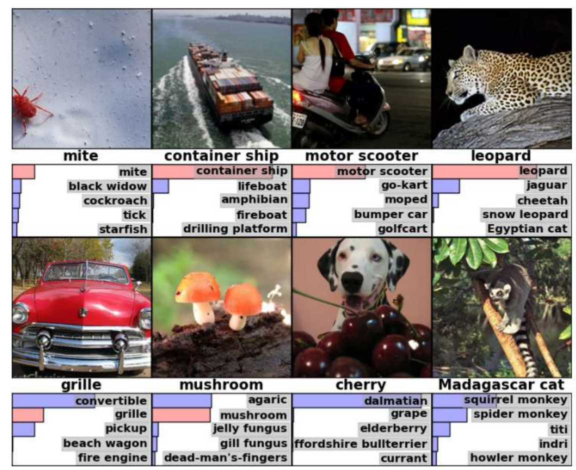
Important Results

- Deep Convolutional Nets for Vision (Supervised)



IMAGENET

1.2 million training images
1000 classes

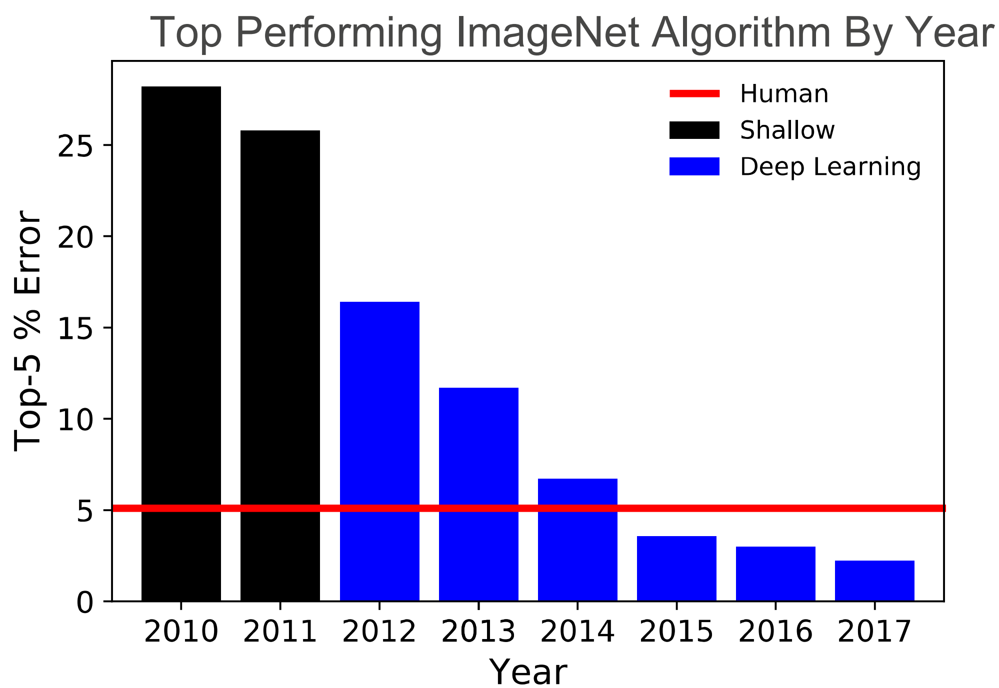


- Deep Nets for Speech (Supervised)

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 images| horse images)

- ▸ Failure Case

Zhou el al., Cycle GAN 2017

Statistical Generative Models

Source actor

Target actor

Real-time reenactmen t

[Face2Face, Thies et al, CVPR 2016]

Deep Learning beats human performance in many tasks

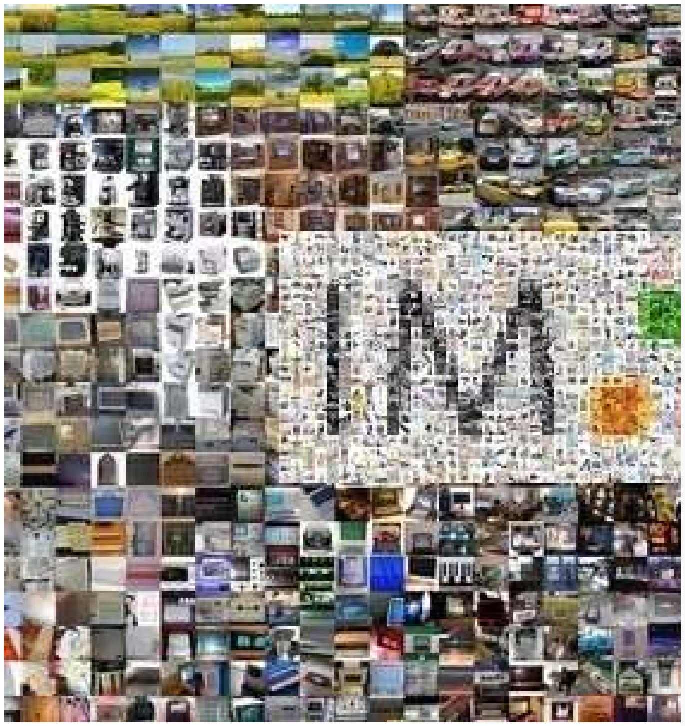
- Famously, Google DeepMind trained "AlphaGo" to beat the world champion Go player (a complex game)
- AlphaGo's largely works by Deep Learning techniques (learned by repeatedly playing the game)
- Many other examples:
    - Voice Recognition
    - Object Detection
    - Text Translation
    - Etc
- So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

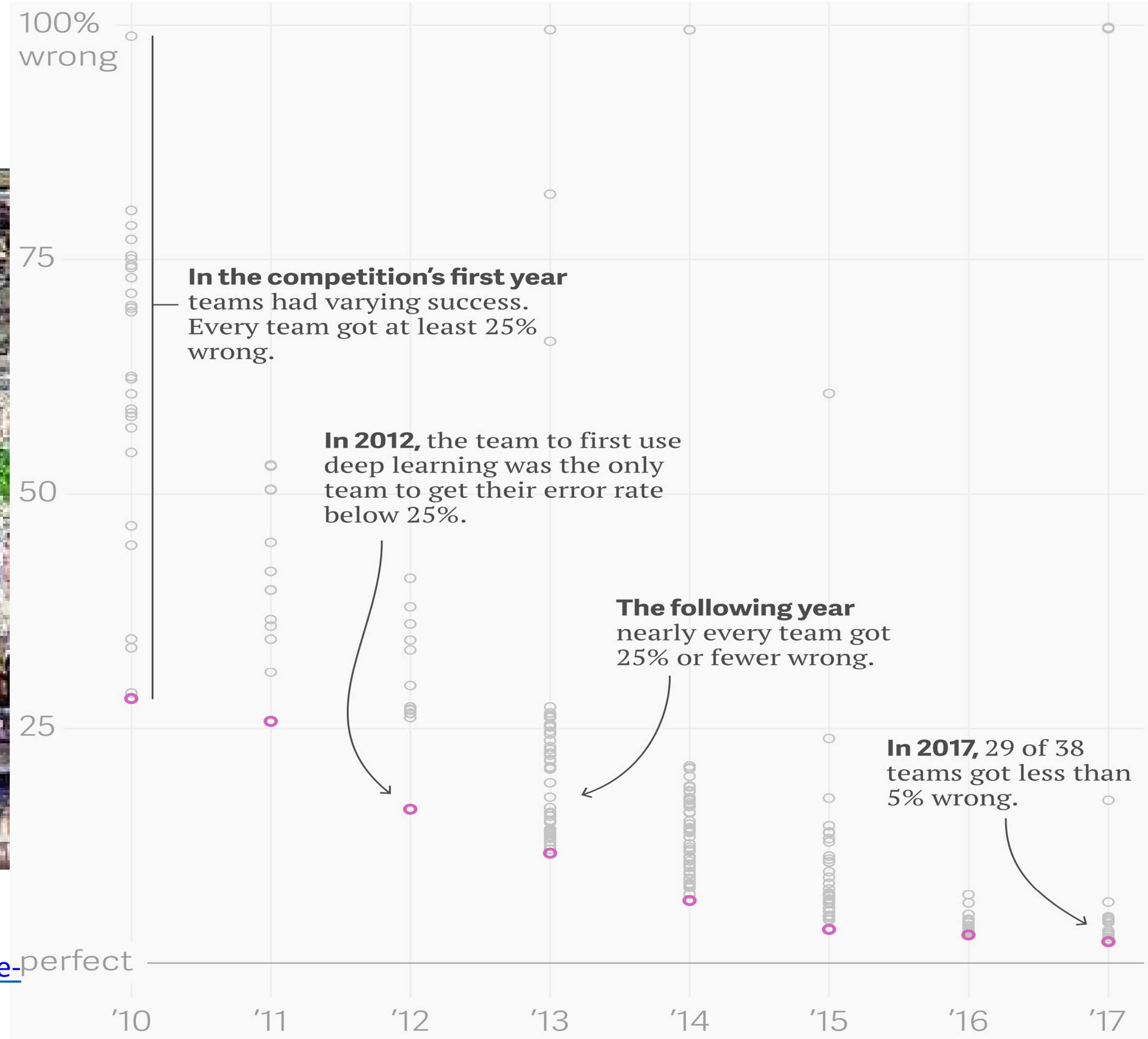
Some Applications

Classify Images

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

Classify Images



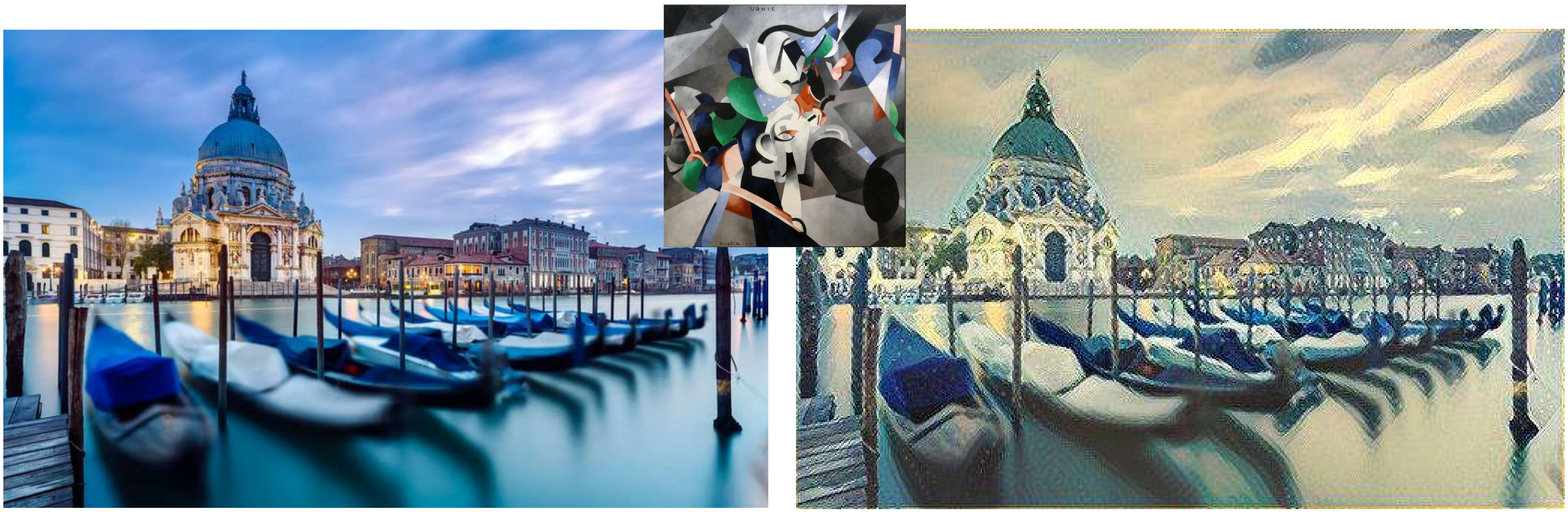


Yanofsky, Quartz
https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

Detect and Segment Objects

https://github.com/matterport/Mask_RCNN

Style transfer

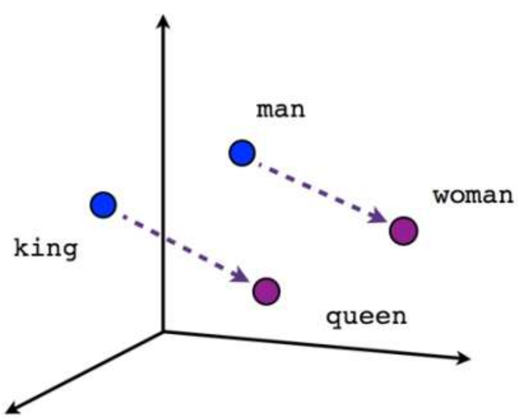


https://github.com/zhanghang1989/MXNet-Gluon-Style-Transfer/

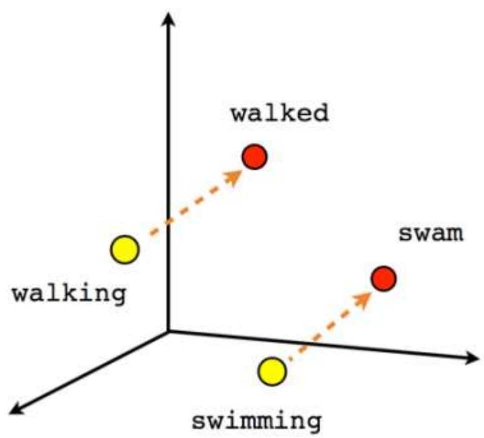
Synthesize Faces

Karras et al, ICLR 2018

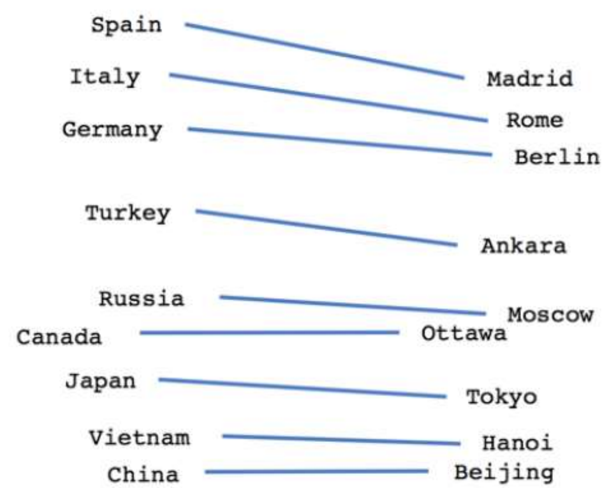
Analogies



Male-Female



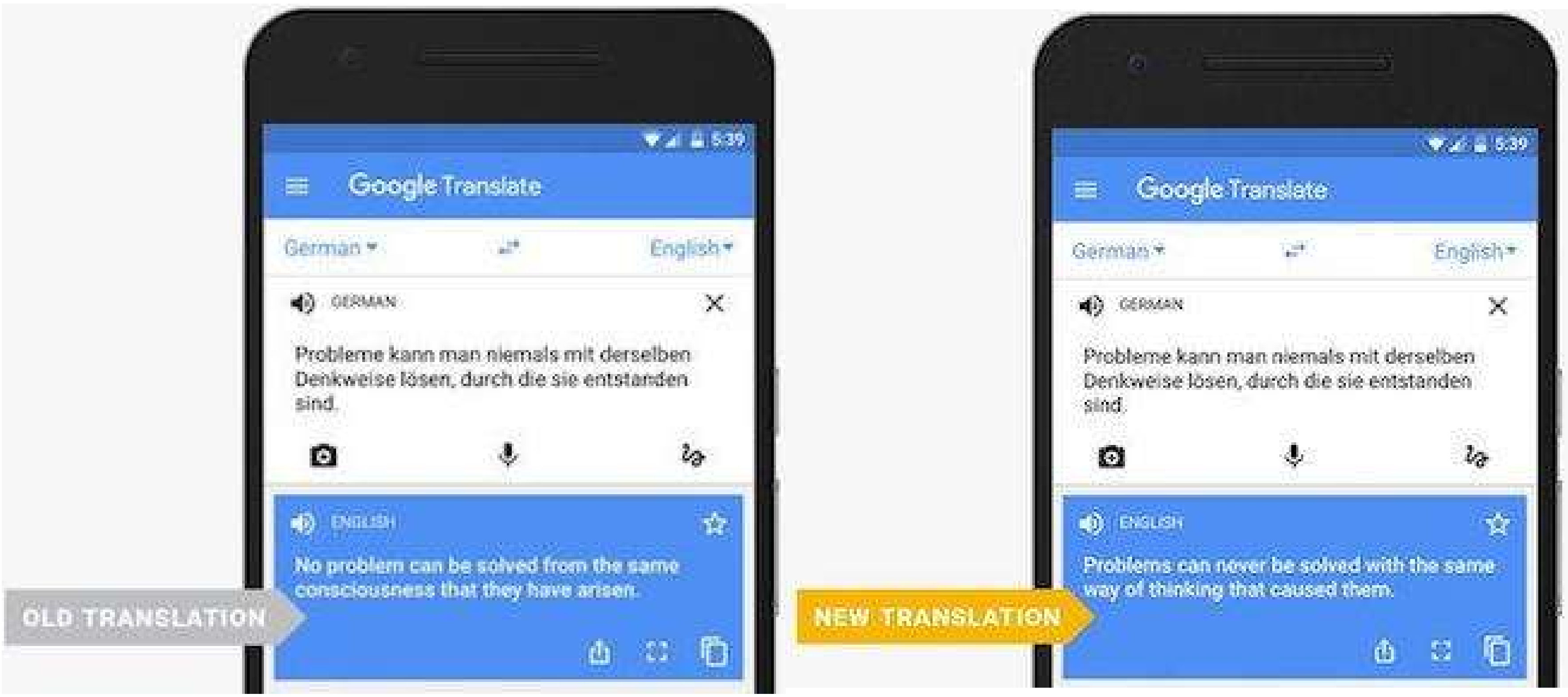
Verb tense



Country-Capital

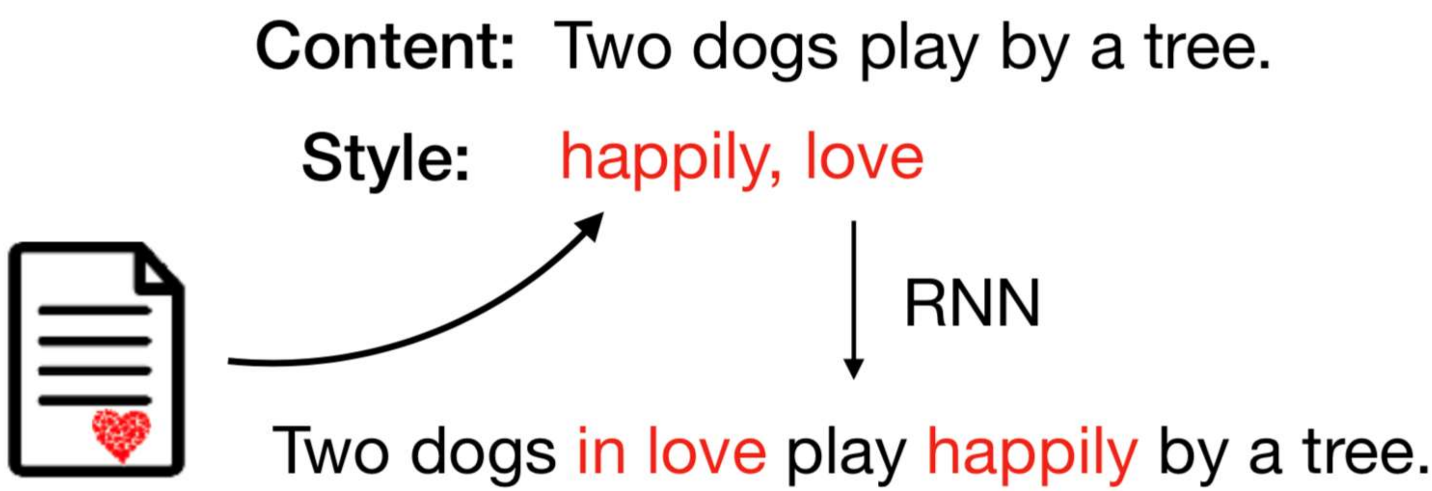
https://www.tensorflow.org/tutorials/word2vec

Machine Translation



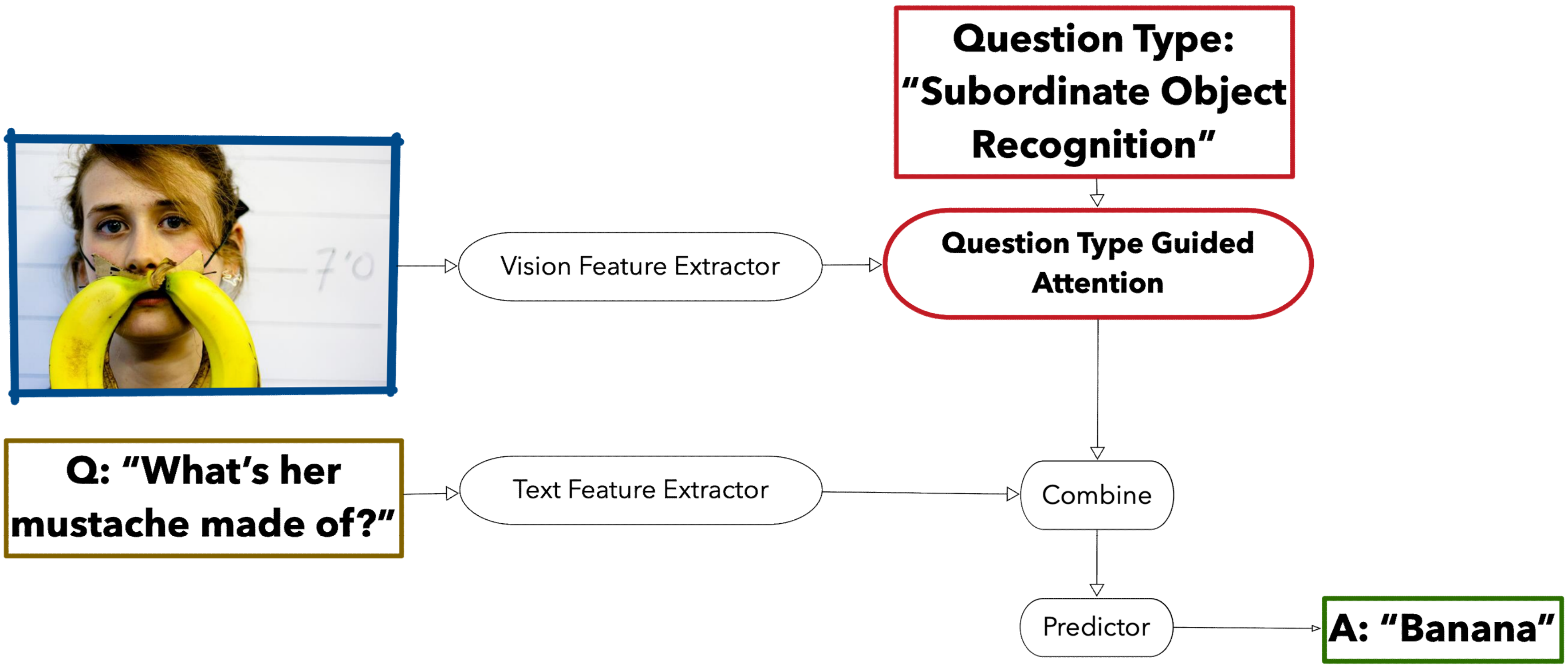
https://www.pcmag.com/news/349610/google-expands-neural-networks-for-language-translation

Text synthesis



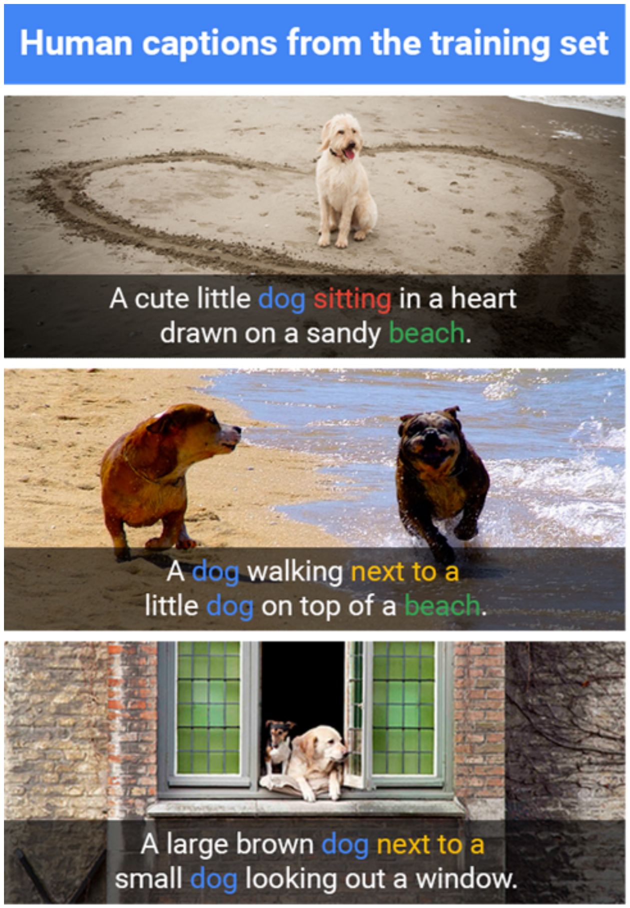
Li et al, NACCL, 2018

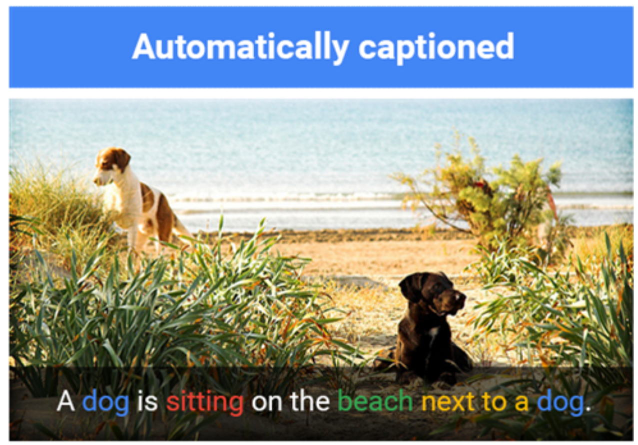
Question answering



Shi et al, 2018, Arxiv

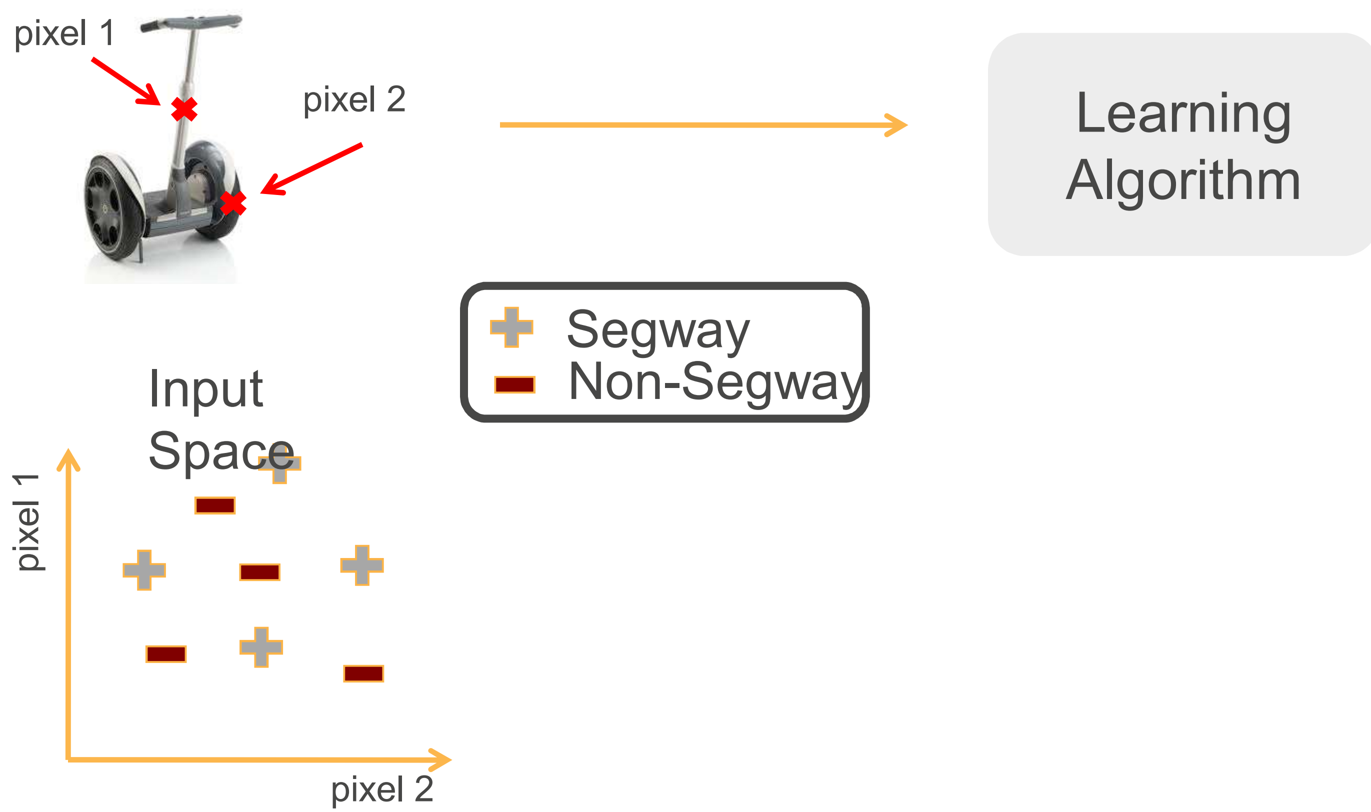
Image captioning



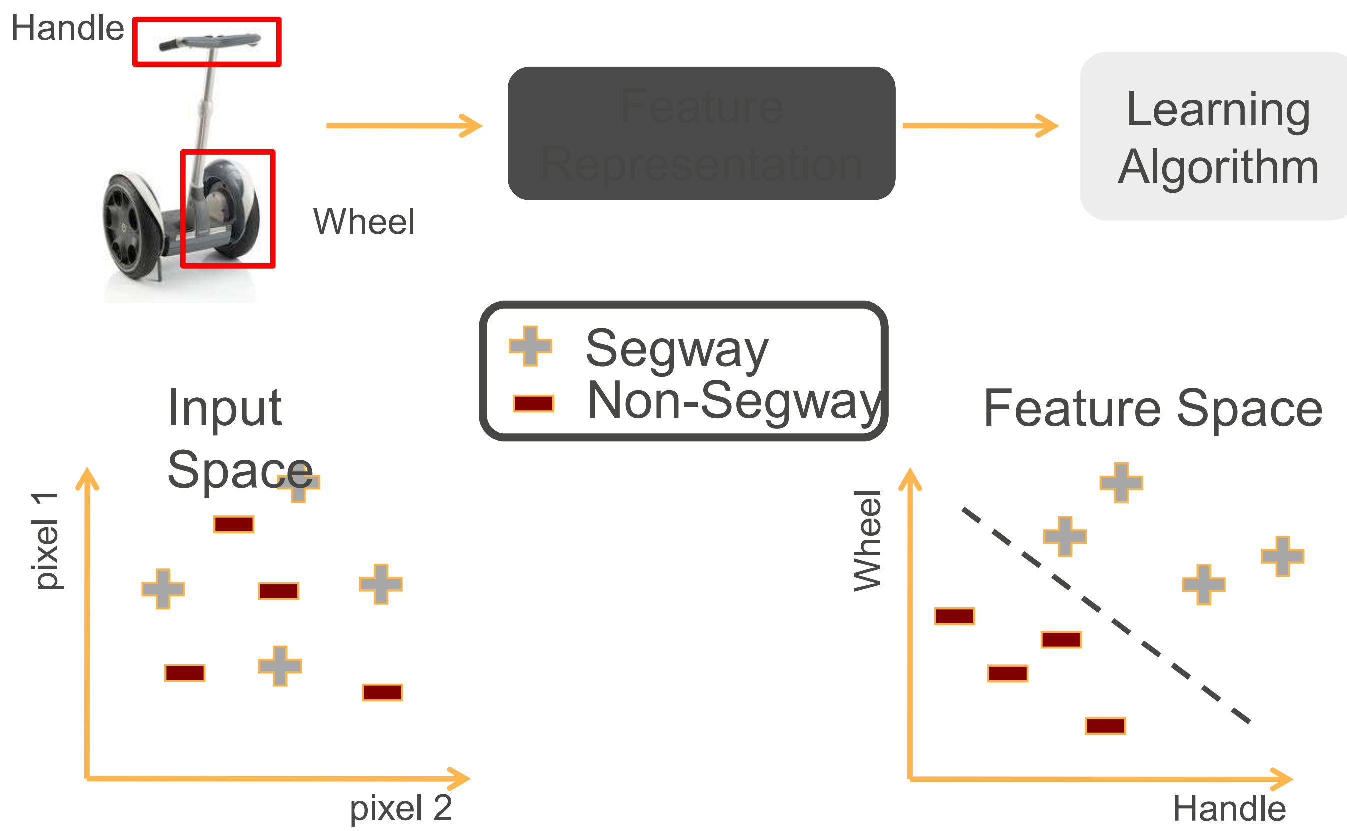


Shallue et al, 2016
https://ai.googleblog.com/2016/09/show-and-tell-image-captioning-open.html

Learning Feature Representations



Learning Feature Representations

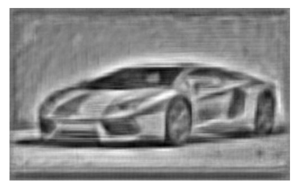


Traditional Approaches

Data → Feature extraction → Learning algorithm

Object detection

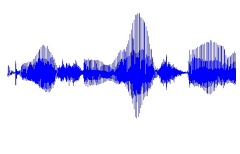
Image



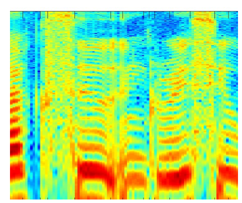
vision features

Recognition

Audio classification



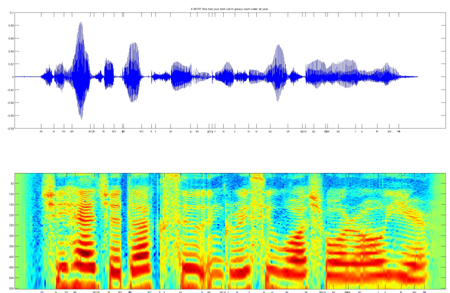
Audio



audio features

Speaker identification

Features



Spectrogram

vision features

Representation Learning:
Can we automatically learn
these representations?