

Understanding the World Through Data

Jake Garrison

5/20/2020

Goals of this talk

How data can be used to:

- Learn new domains or skills
- Understand complex systems
- Solve hard problems (correctly or incorrectly)
- Discover new ideas, insights, products, or science

How using data in such ways connects to:

- The current state of the world in 2020
- My work, accomplishments, failures
- Getting a job or degree or starting a company
- Discovering new projects, or hobbies to do while social distancing

About Me



jakegarrison.me

Before College

From: Spokane, Washington

I liked to create things, but didn't really know what I was doing or how to do it

Projects:

- Video game bots
- Useless websites
- Analog audio circuits
- Electric Car



College

UNIVERSITY *of* WASHINGTON

Undergraduate

- Power electronics electric vehicles
- EcoCar Competition Lead
- Signal processing + machine learning
- Digital Art
- Autonomous driving + controls
- Audio electronics

Graduate

- UbiComp Lab Research
- Health Sensing
- Sound Sensing
- On device machine learning
- Robotics
- 3d Printing Art

Undergraduate

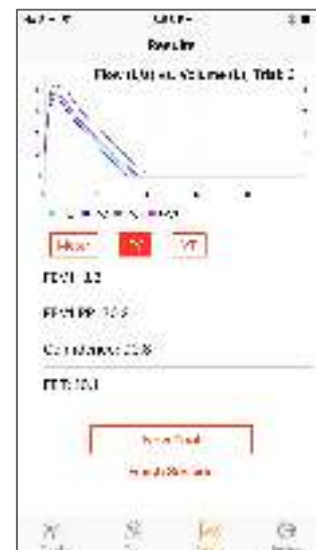
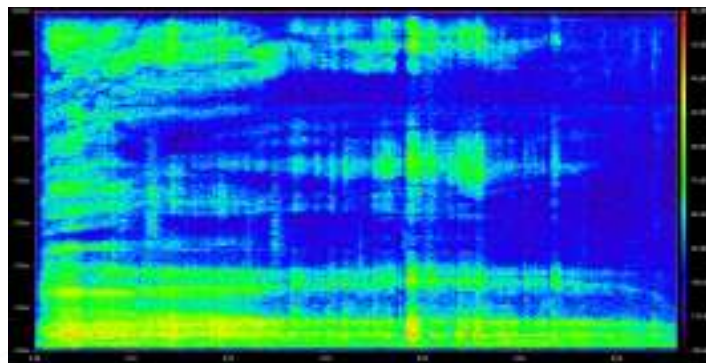
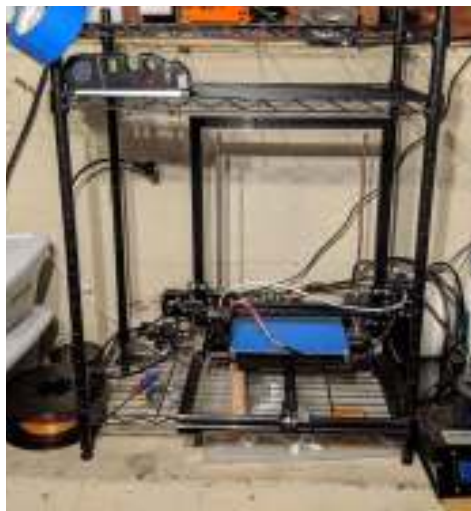


ELECTRICAL & COMPUTER ENGINEERING

W



Graduate



Jobs, Startups, Internships

- 2013: **Verellen Amplifiers**, built tube amps and pedals
- 2014: **Tesla**, Power Electronics, intern
- 2014: **Bankroll**, startup, ASIC mining design and algorithmic trading bots
- 2015: **Tesla**, Sensors and Autopilot, intern
- 2015: **puppy.ai**, founder, iPhone dog breed detection
- 2016: **Haiku Deck**, startup, presentations powered by AI
- 2016: **Research Assistant**, UbiComp Lab
- 2017: **Nest / Google**, Senosis, research intern / part time, health sensing
- 2018 - now: **Google**, Health Research, full time

Jobs, Startups

Google Health



nest



AIGEN



TESLA



puppy.ai — Dog Breed Detector
Point camera at dog's face.
Nycomb, Inc.
free



Hobby Projects

Philosophy: Do things that matter to me, keep it simple, learn, hack, create

Using old phones, sensors or chips to do useful things:

AI plant nurse, fish cam, brewing sensor, action sports sensor

Using free APIs and scripting to do useful things:

Automate shopping, craigslist hunting, music discovery, school/job work

Using technology to do art things

3d printers + crayons, real time sound/images/light from time series data

Problem Solving with Data

Best practices and lessons learned

You have a problem you want to solve (but don't know how to)

Where it came from...

...doesn't *really* matter.

Your goal is to understand the problem

...then create something which hopefully makes it less of a problem

...then maybe share it or fulfill your dreams

oh and your social distancing

Explore the Domain



Let Search Engines Do the Work

Peer reviewed authoritative content:

- Publications, patents, whitepapers, *most* wikipedia pages, textbooks
- scholar.google.com or patents.google.com for searching
- [VOSviewer](#) and [CitNetExplorer](#) to clustering and discovering related work
(see [slides](#))

Just because something published doesn't always mean it is an absolute truth, not all publications are equal

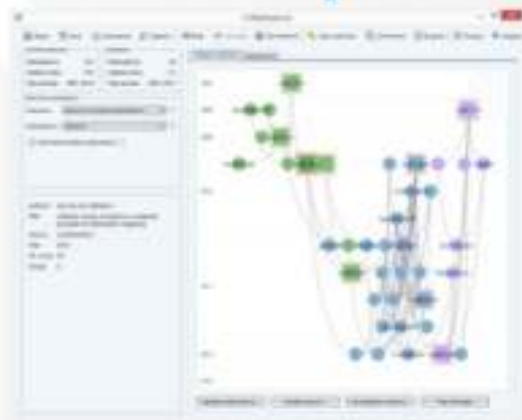
Keep track of your sources in case you want to publish your work!

VOSviewer



- Any type of bibliometric network
- Co-citation and bibliographic coupling of publications
- Time dimension is ignored
- At most about 10,000 publications are supported

CitNetExplorer



- Only citation networks of publications
- Direct citations between publications
- Time dimension is explicitly considered
- Millions of publications are supported

Let Search Engines Do the Work

User contributed content... *here be dragons*

- Forums, news, reviews, reddit, youtube, github, stackoverflow blogs
- Use counter metrics like 'upvotes' 'stars', 'views' 'comments' to assess reliability and consensus
- Know to tell the difference between facts, guesses and opinions
- Know the bias of the source as may influence which facts are excluded

Some topics are too fresh or not academic/scientific enough to have peer reviewed content, that does not mean it's a bad idea to explore!

Fake news

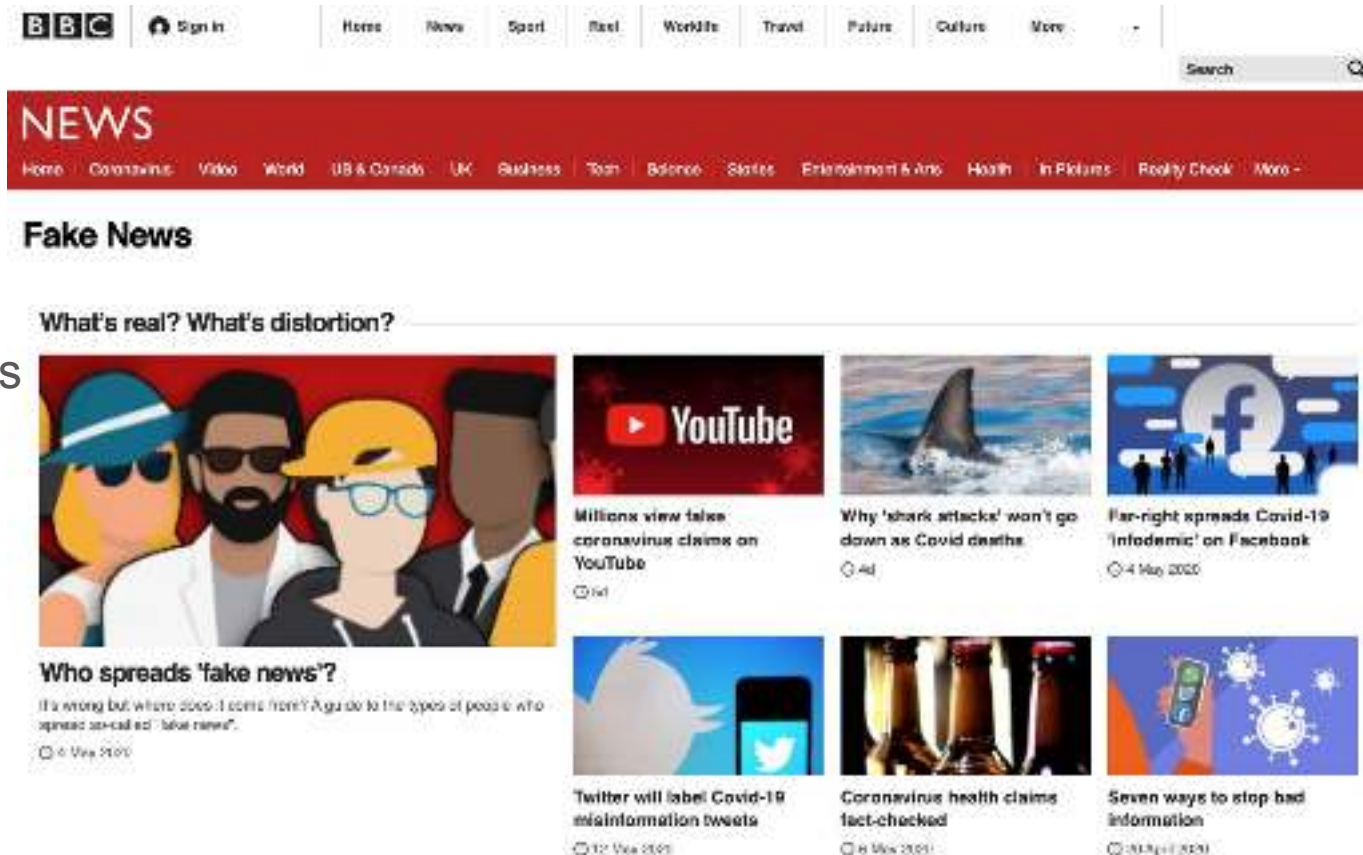
Hot topics come with exaggerated, fake or unverified news

Don't assume algorithms will protect you from it

Look for comments or a reddit thread

Help flag or refute it

Take the [fake news detection test!](#)



The image shows a screenshot of the BBC News website. At the top, there is a navigation bar with the BBC logo, a 'Sign In' button, and links for Home, News, Sport, iPlayer, Worklife, Travel, Future, Culture, and More. Below this is a red banner with the word 'NEWS' in white, and a secondary navigation bar with links for Home, Coronavirus, Video, World, US & Canada, UK, Business, Tech, Science, States, Entertainment & Arts, Health, In Pictures, Reality Check, and More. The main content area is titled 'Fake News' and features a sub-header 'What's real? What's distortion?'. Below this, there are several article thumbnails:

- Who spreads 'fake news'?**: A thumbnail showing four stylized avatars. The text below reads: "It's wrong but where does it come from? A guide to the types of people who spread so-called 'fake news'." and "© 6 May 2020".
- Millions view fake coronavirus claims on YouTube**: A thumbnail with the YouTube logo. The text below reads: "© 6 May 2020".
- Why 'shark attacks' won't go down as Covid deaths**: A thumbnail showing a shark breaching the water. The text below reads: "© Ad" and "© 4 May 2020".
- Far-right spreads Covid-19 'infodemic' on Facebook**: A thumbnail with the Facebook logo. The text below reads: "© 4 May 2020".
- Twitter will label Covid-19 misinformation tweets**: A thumbnail with the Twitter logo and a smartphone. The text below reads: "© 12 May 2020".
- Coronavirus health claims fact-checked**: A thumbnail showing several bottles of beer. The text below reads: "© 8 May 2020".
- Seven ways to stop bad information**: A thumbnail showing a hand holding a smartphone with virus icons. The text below reads: "© 30 April 2020".

Historical Context Matters

Trends, consumer metrics, product reviews

- Advanced search options to filter date ranges or locations
- trends.google.com to compare search queries through history
- google.com/finance to compare companies / stocks
- alexa.com for web content trends and stats
- Amazon, Ebay for consumer product trends and reviews
- Apple or Android store for app popularity trends and reviews

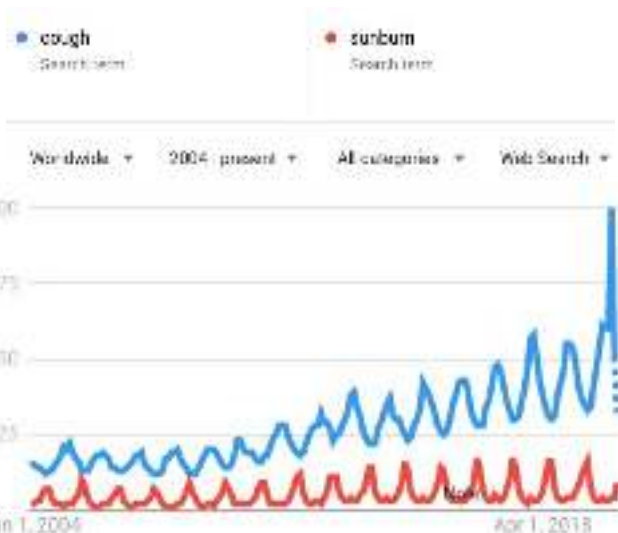
Is this something people care about now? Or a decade ago?

Has it been done before? What were the limitations?

“Cough” query

Very different vibe now compared to a few years ago

“cough” interest is seasonal, opposite to “sunburn”...until now



[link: cool screensaver](#)

Search results for "cough". The date filter is set to "Past month" (47 results). The first result is from www.scaredalert.com titled "Here's How to Tell if Your Cough is a Sign of Coronavirus or ...". The second result is from www.health.com titled "What is a Dry Cough? Experts Explain Coronavirus Symptom ...".

Common questions

- Is coughing a symptom of the coronavirus disease?
- Why do I need to stay away from people who are coughing or sneezing?
- What are the symptoms of COVID-19?

The information presented only cannot be used as medical advice by the author.

App would listen to you cough to predict whether you have ...
1 day ago - A slew of AI labs are currently aiming to fight coronavirus with apps that analyze audio from people speaking, breathing, or coughing.

Scientists Developing App Based On Sound Of COVID-19 ...
6 days ago - Scientists have used cough sounds to diagnose whooping cough, asthma and pneumonia in the past. "This new diagnostic tool uses machine learning algorithms."

Search results for "cough". The date filter is set to "Jan 1, 2018 - Dec 31, 2018" (47 results). The first result is from www.healthline.com titled "Why Am I Coughing? Causes, Prevention, and More". The second result is from www.healthline.com titled "Types of Coughs: What Different Types of Coughs Indicate".

Types of Coughs: What Different Types of Coughs Indicate
Jan 28, 2018 - A cough is a protective defense of the body that helps protect your body from infection, irritants, and allergens. A cough can also be a sign of a respiratory infection. Coughing is a symptom of many illnesses and conditions. You can use a doctor to help you get your cough under control.

19 Tips on How to Stop Coughing Including Nighttime Cough
May 20, 2018 - What you see coughing is a sign of health problems as well as a warning sign of a cough. What you see is a sign of a cough and you can see it.

Cough Relief: How to Get Rid of a Bad Cough - WebMD
Feb 1, 2018 - You can reduce coughing and take control of many coughs with WebMD and our other health experts and find out how to best to stop a cough by following the ...

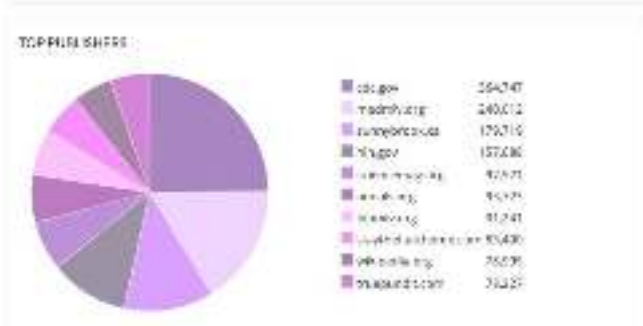
Monitor the COVID-19 Conversation as it Unfolds

See what articles people are publicly sharing on Twitter and Reddit, and how they are reacting to the latest news of the coronavirus pandemic.

About this data

Healthcare	All Topics	Quarantine	Social Distancing	Treatment	Economic Effects	Pandemic	Good News
------------	------------	------------	-------------------	-----------	------------------	----------	-----------

	Total Engagement	Total Shares	Twitter Engagement	Reddit Engagement	Alexa Rank (SEO)
Researchers Warn Against the COVID-19 Virus - Scrippsack Research's Institute on Hygiene	174,714	44,268	175,667	4,253	63,312
Coronavirus Disease 2019 (COVID-19) - cdc.gov	103,478	7,497	16,727	82,988	248
Coronavirus Disease 2019 (COVID-19) - cdc.gov	96,940	7,794	19,589	76,940	248
The Ineffective Nature of COVID-19 Face Masks & Benefits of Continued Use - JAMA Network Open American College of Physicians	88,284	2,775	7,366	81,728	45,184
A Movement in Seattle: COVID-19 Patients 60kgTheFork is here to help you find the best	65,400	6,441	17,513	65,407	78,423
COVID-19: Face / Apron / Distancing UpToDate.com 10 Years Ago to Cuba Cruise: "Goods Needed to Die" - True North, Trapped.com	76,022	37,626	76,022	1	65,516
NHL's National Investigation of Vaccines for COVID-19 - Sports Illustrated	70,940	18,328	81,757	11,885	124
Study: 17% of People With COVID-19 Coronavirus Did Not Show Symptoms - CNN	62,716	64	82	62,667	117,077
COVID-19: A Novel Coronavirus From Previous Flu Virus - Reviewed by Experts ScienceAlert	58,057	5,735	13,156	44,898	6,375
COVID-19 Antibody Seroprevalence in State-Capital County, California - medRxiv	58,027	17,641	42,429	8,028	9,656



Narrow the Problem Scope



Who is it for?

Yourself? Your peers? Your family?

Professionals? Non-technical users?

Users at home? On mobile? Have internet?

Users in a particular country? or language?

Users of a particular age? or demographic?

What is or isn't it?

Software / Hardware

For sale / For free

Academic / for fun

Open source / top secret

Programming language

Chip supplier

Use what you know to narrow the problem scope

You have now probably thought about the problem enough at this point to consider yourself “above average” on the topic...

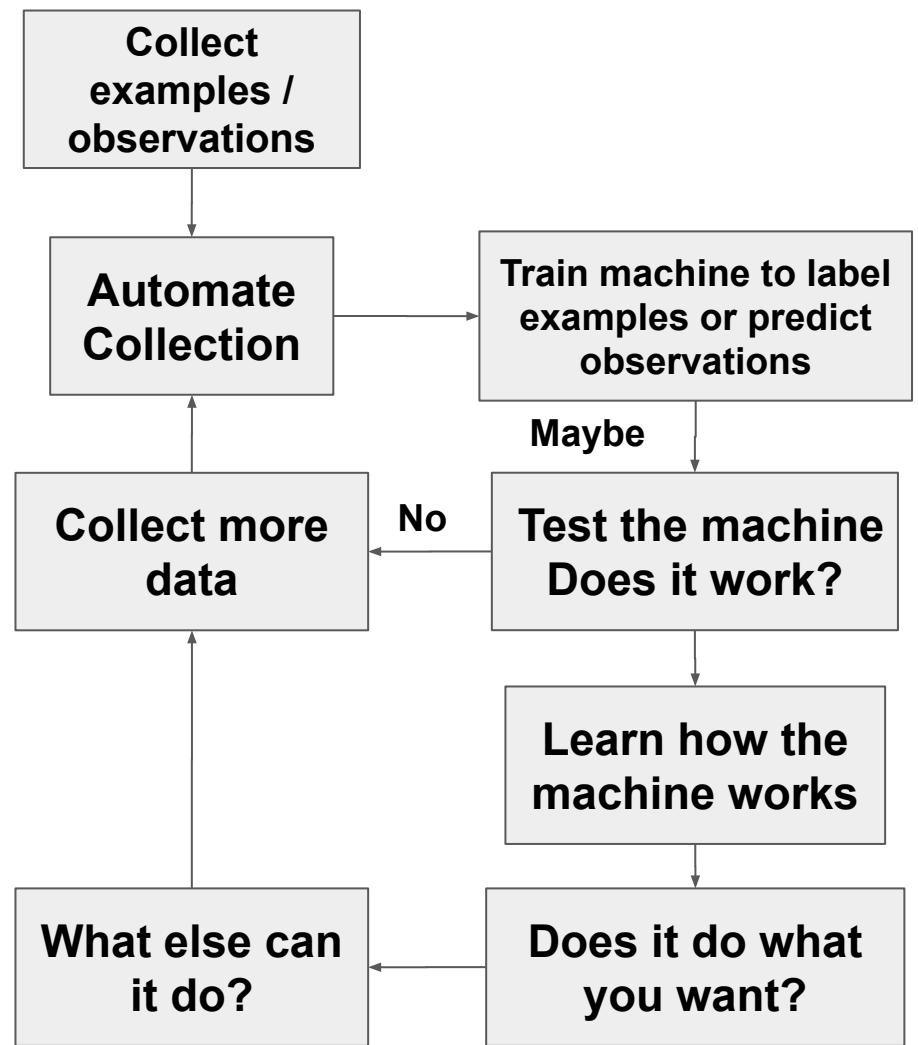
Can you put together a solution by integrating existing work, code and designs?

No?

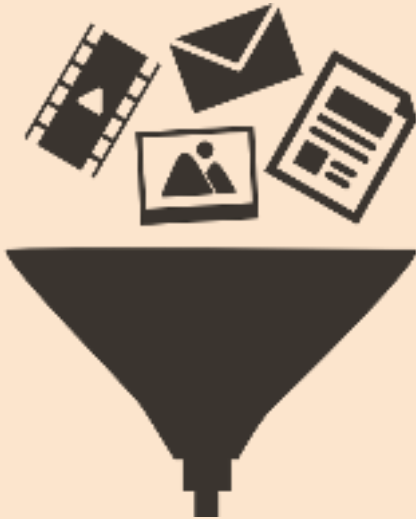
Do you need to improve existing methods? Not sure what the best design, algorithm or product decision is?

...If you had unlimited time and memory to scour the internet, ingest the information do you think you would know what to do?

Automation and Machine Learning



Collect and Analyze Data



**Collect
examples /
observations**



**Automate
Collection**

Data Acquisition Techniques

Use public datasets, APIs or web scraping techniques

- Common machine learning problems have curated datasets which are a great starting point: [Kaggle](#), [ImageNet](#), [AudioSet](#), [Youtube8m](#)
- Most popular sites have free, limited APIs: [weather](#), [maps](#), [reddit](#)
- Almost anything you interact with on the web can be scraped using unofficial or 3rd party APIs: [youtube](#), [craigslist](#), [google images](#)
- Find historical snapshots of the internet: [WayBack Machine](#)

Careful with API rules, data and code licenses, unofficial APIs and copyright.

Especially if you plan on commercializing or publishing your work!

Data Visualization Techniques

Use spreadsheets or programming to visualize aggregated data

- Know when to use [histogram vs line vs scatter vs bar vs pie plots](#)
- Use error bars to express measurement confidence or consensus
- Experiment with trendlines normalization or axis scaling such as log scale

Understand the basics of [ML fairness](#) and [sampling bias](#) before using a large dataset for insights.

How was the data collected? For what purpose?

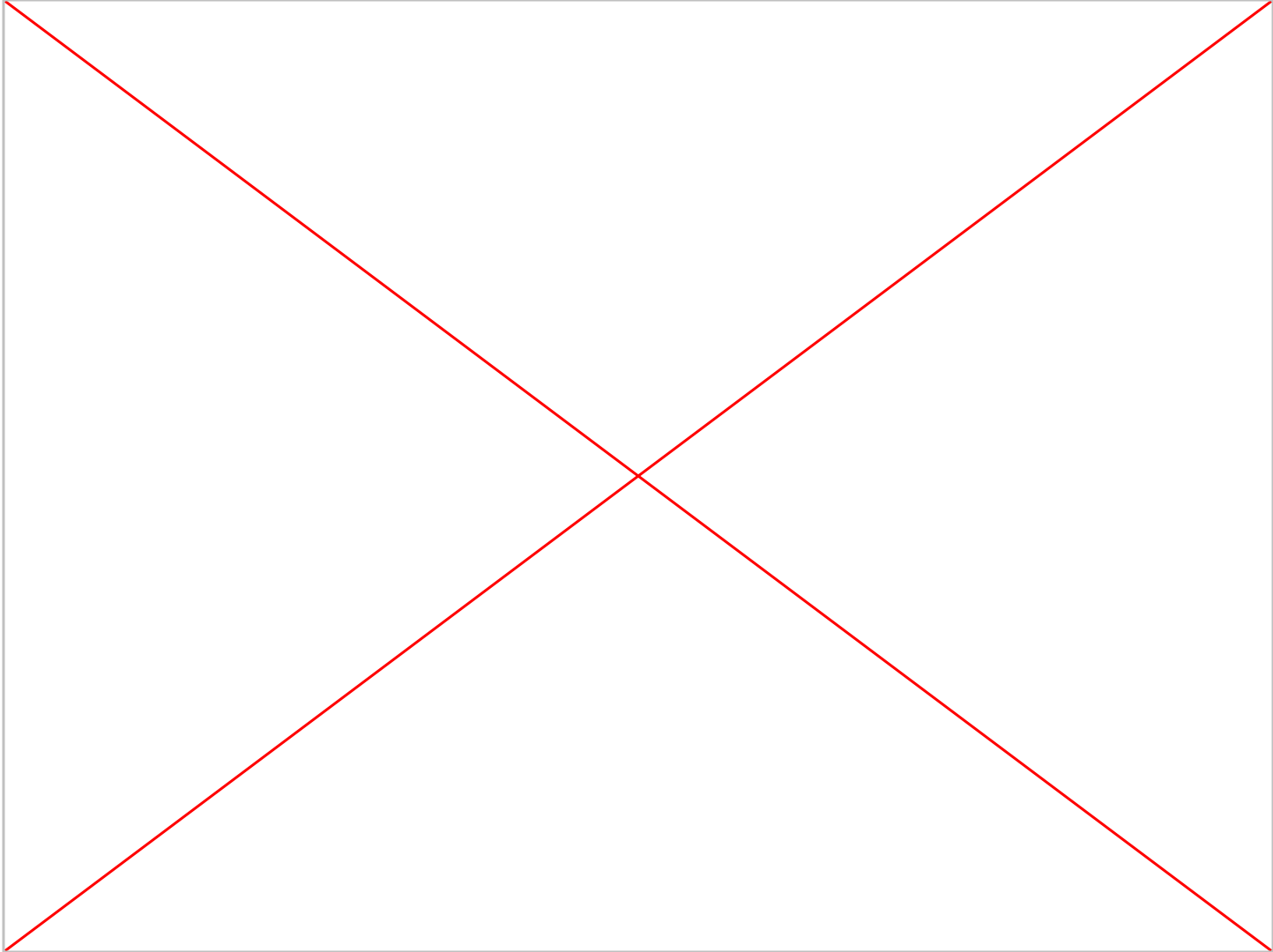
Are there gaps or imbalances? Are there obvious patterns?

Covid-19 Example

Data Visualization

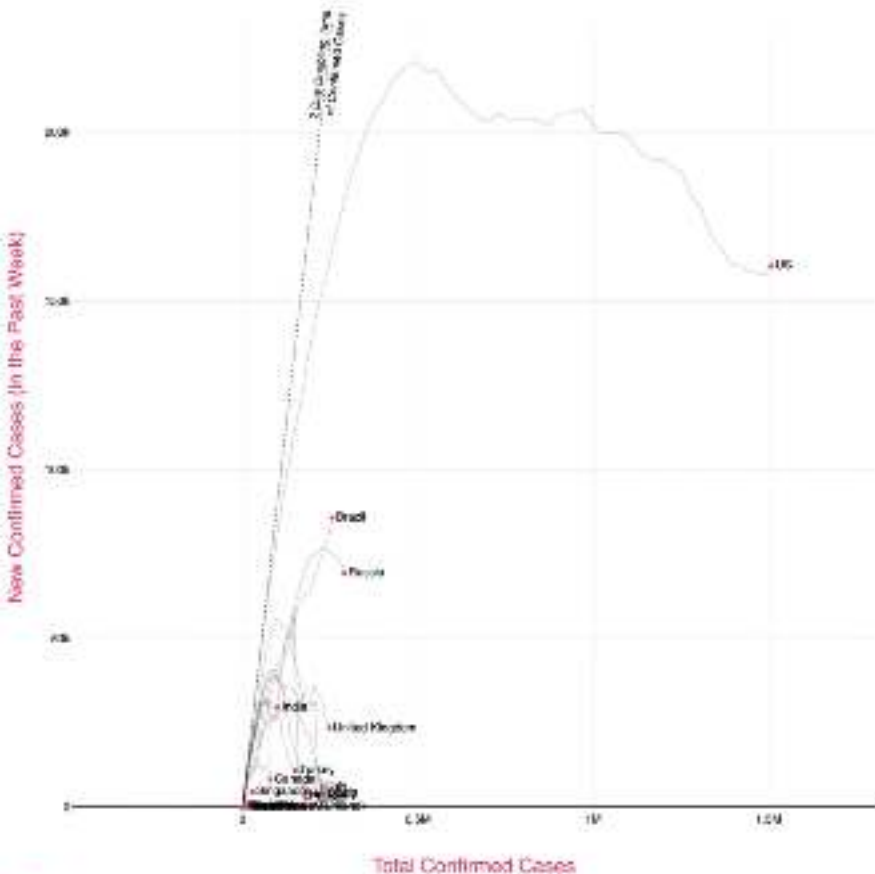
Log scaling can be more insightful when a trend follows more of an exponential

[source](#)



Covid-19 Example

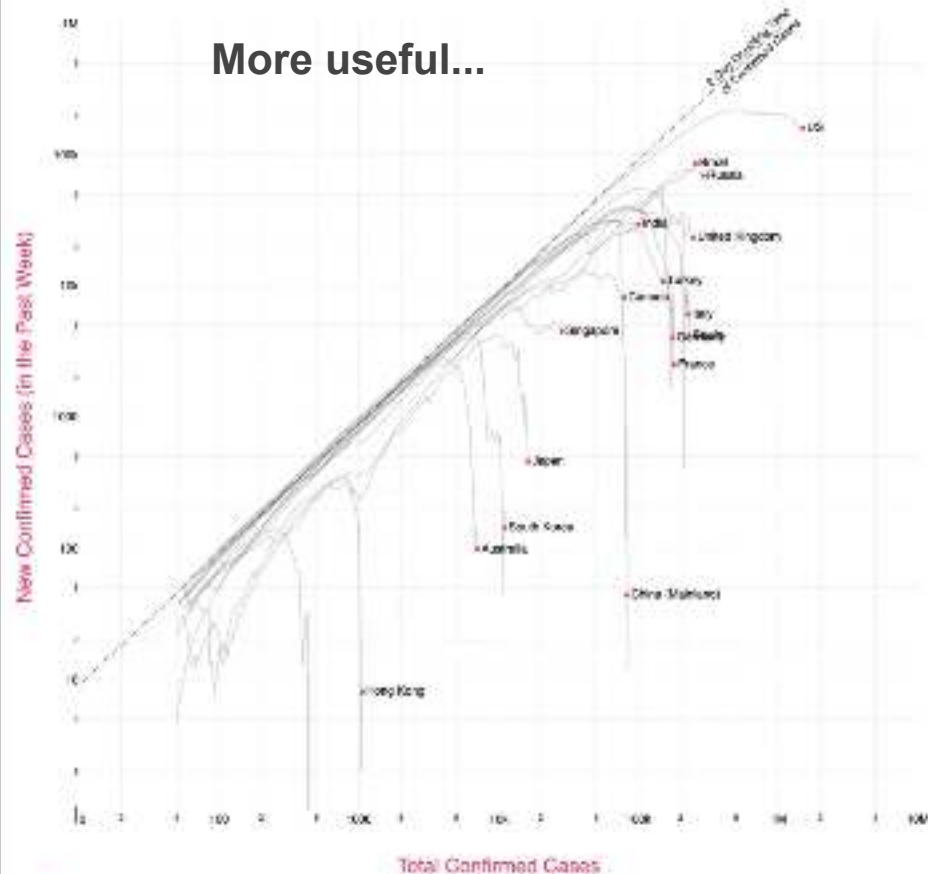
Trajectory of World COVID-19 Confirmed Cases (2020-05-18)



2020-05-18

Linear Scale

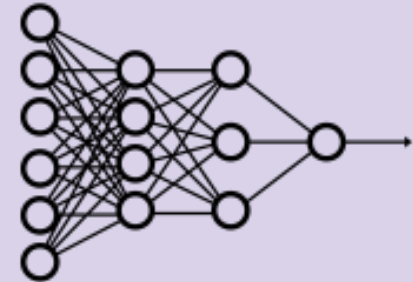
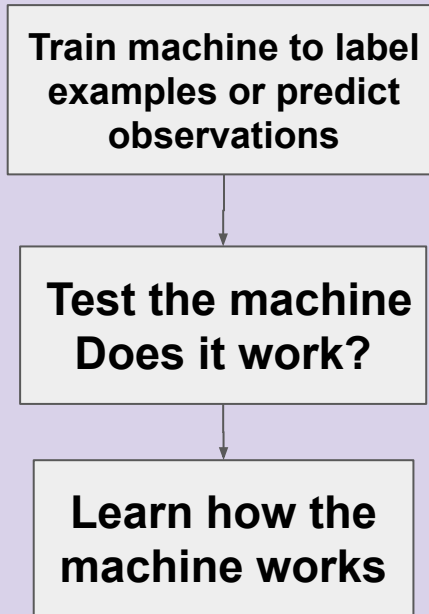
Trajectory of World COVID-19 Confirmed Cases (2020-05-18)



2020-05-18

Logarithmic Scale

Learning from the Data

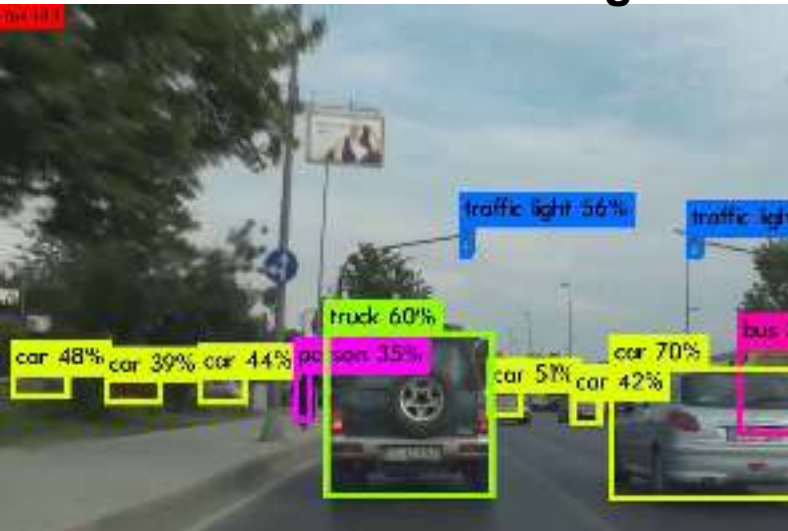


Learning from Labeled Data

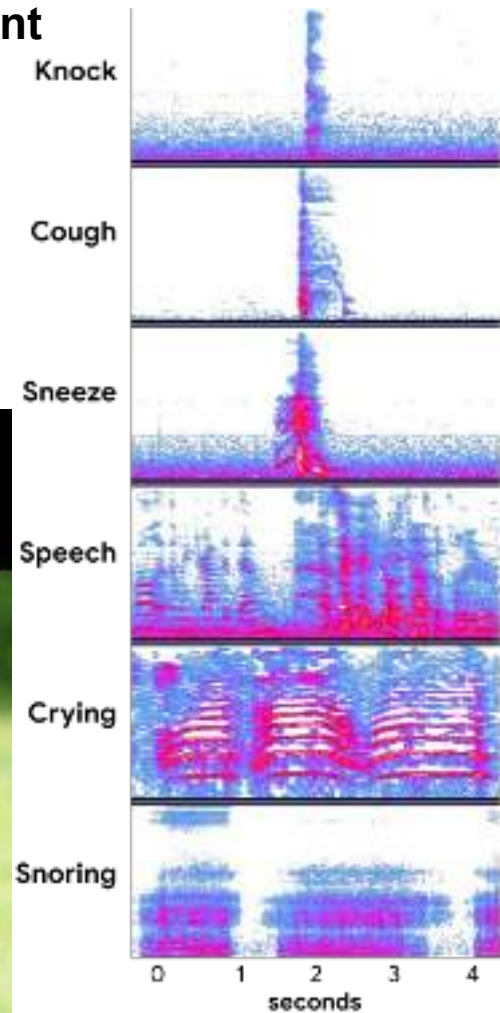
Supervised Learning

- **Input:** Images, audio, video, noisy data
- **Output:** Classification, Segmentation
- Convolutional Neural Net Feature Extraction (CNN)

Autonomous Driving

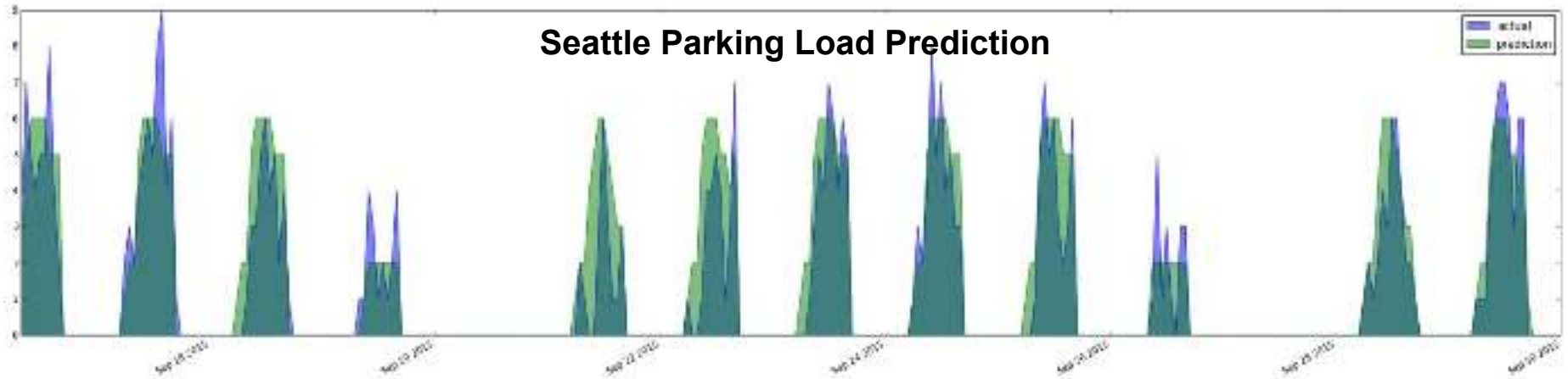
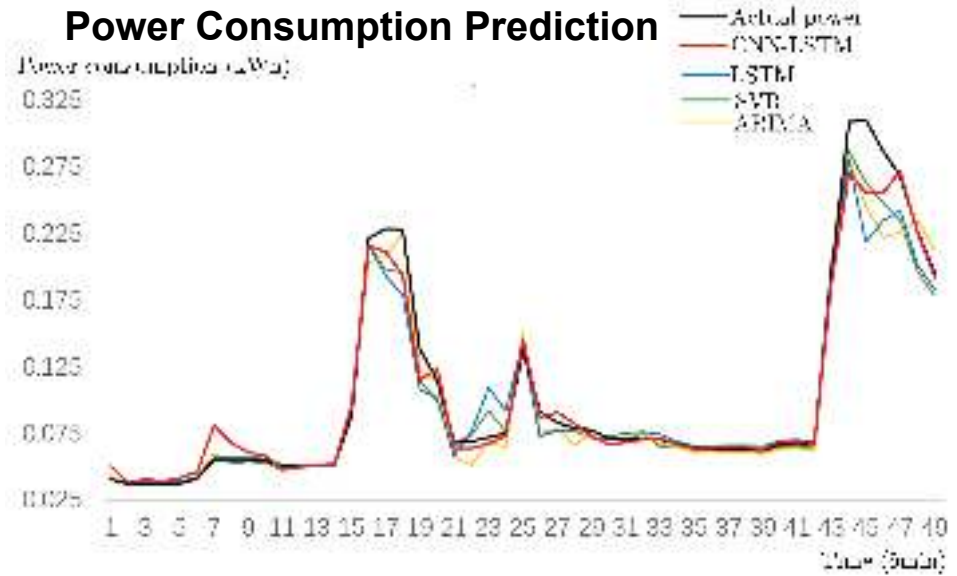


Sound Event Detection



Learning to Predict the Future

- *Time Series forecasting*
- **Input:** Historical time series data
- **Output:** Future time series data
- LSTM, Hidden Markov, ARIMA



Learning from Unlabeled Data

Unsupervised Learning

- **Input:** Structured data
- **Output:** Embedding vector
- Autoencoder, clustering, tSNE
- Word2Vec, Triplet Loss
- Dimensionality reduction
- Text summarization

Text Summarization

Source Text: Peter and Elizabeth took a taxi to attend the night party in the city.

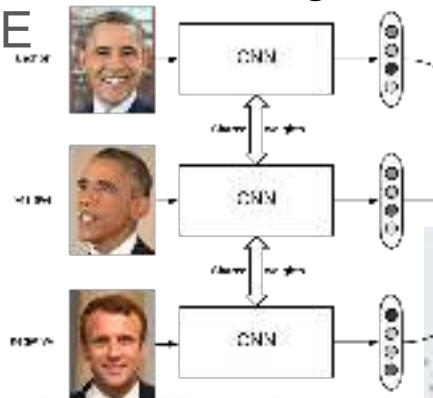
While in the party, Elizabeth collapsed and was rushed to the hospital.

Summary: Elizabeth was hospitalized after attending a party with Peter.

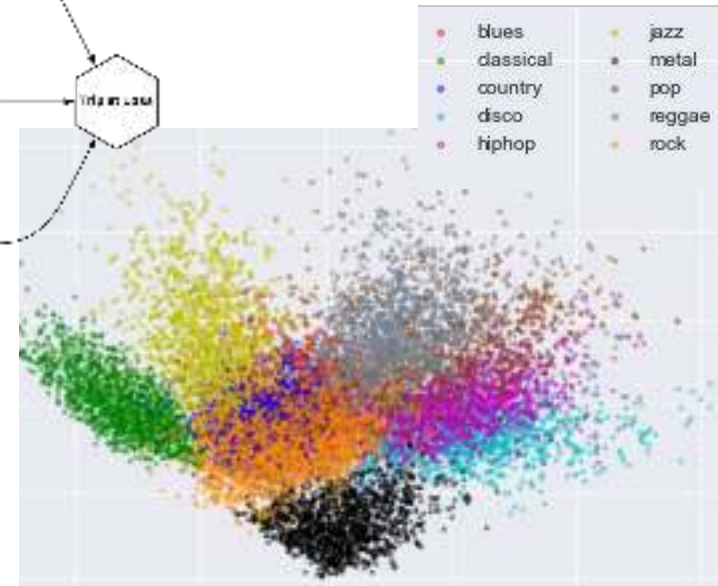
Retina as Biomarker



Face Matching



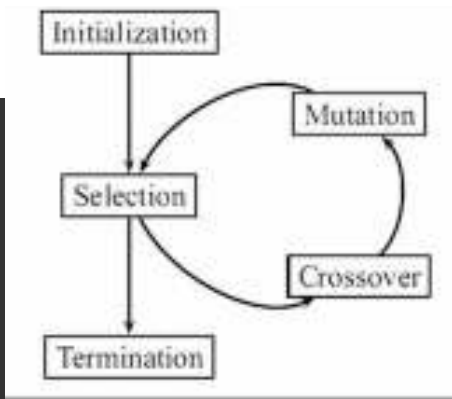
Music Genre Clustering



Learning from Biology

Evolutionary Algorithms

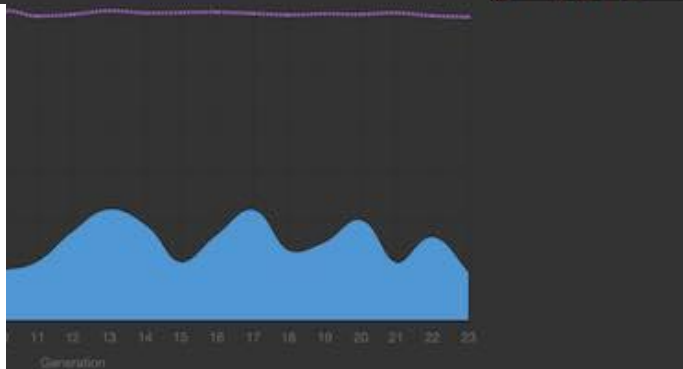
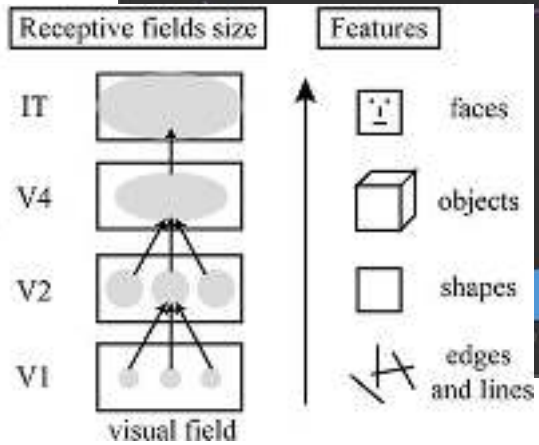
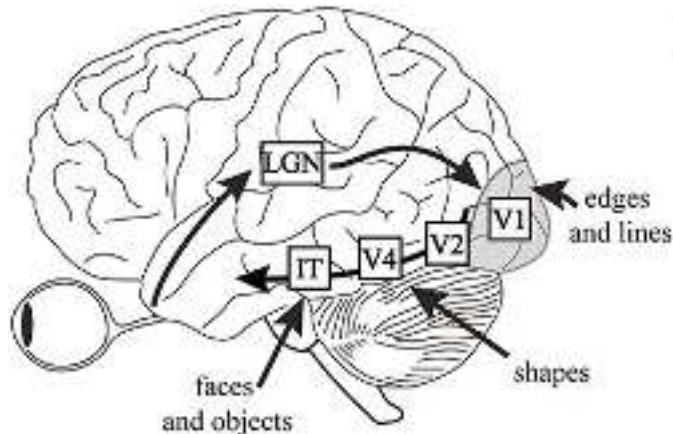
- Robotic control: grasping, walking, flying
- Mimicking the visual cortex



Speed

Generation : 24
Best Score : 12706.3
Most Steps : 3

Run	Score	Steps
1	-6.64	0
2	-1.94	0
3	2038.97	1
4	2111.58	1
5	15.13	0
6	2128.98	1
7	2128.74	1
8	2127.82	1
9	2128.44	1
10	2121.81	1
11	2202.53	1
12	-5.68	0
13	145.06	0
14	96.12	0
15	2165.63	1
16	2113.49	1
17	154.53	0
18	2128.03	1
19	2251.96	1
20	-7.48	0



What If Data is Expensive?

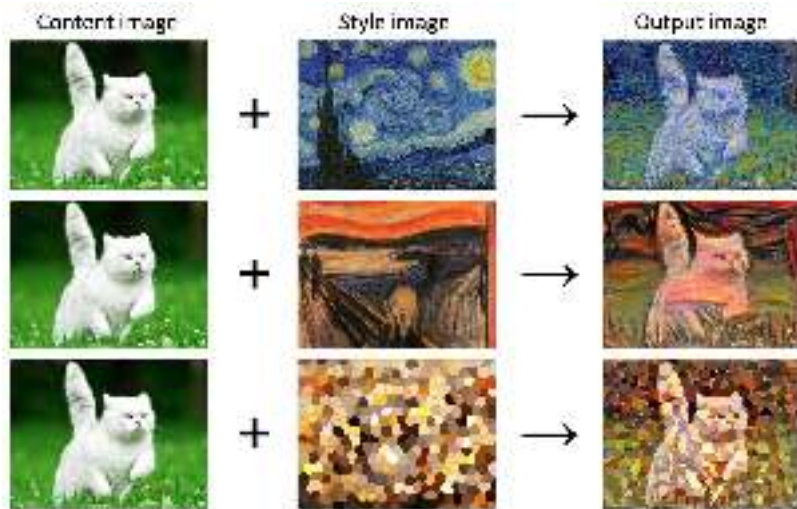


**Collect more
data**

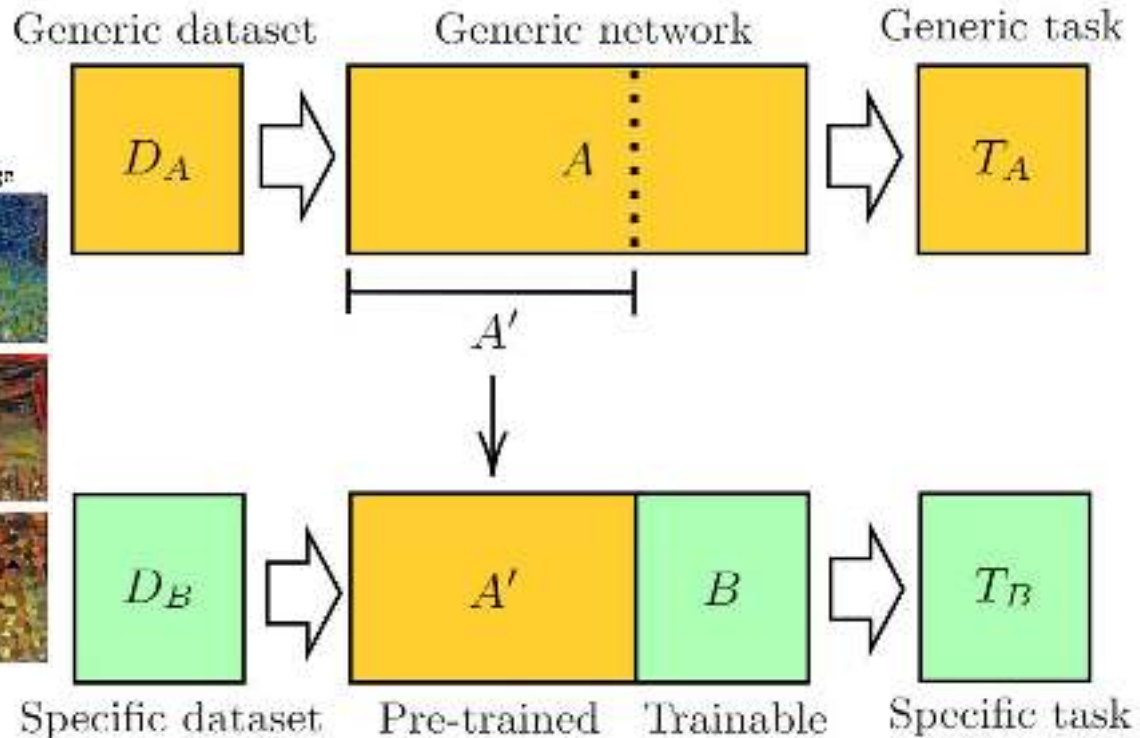
Transfer Learning



- Dog breeds / Plant species
- Product make / model
- Art style transfer
- Specific sound detection

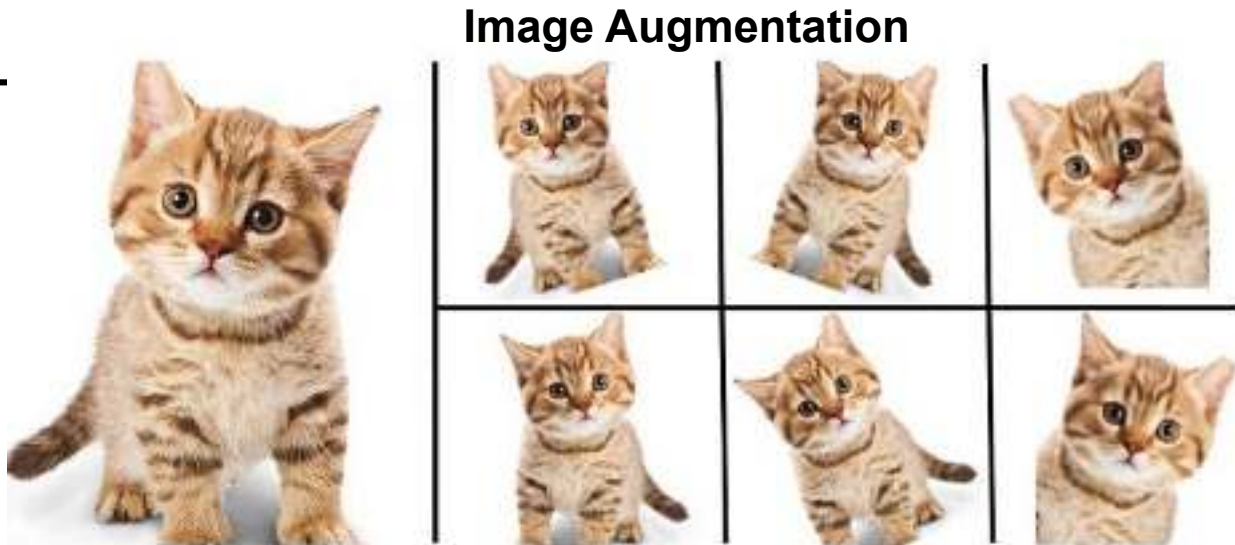


Style Transfer



Data Generation + Augmentation

Augmentation: Realistic computational warping of data

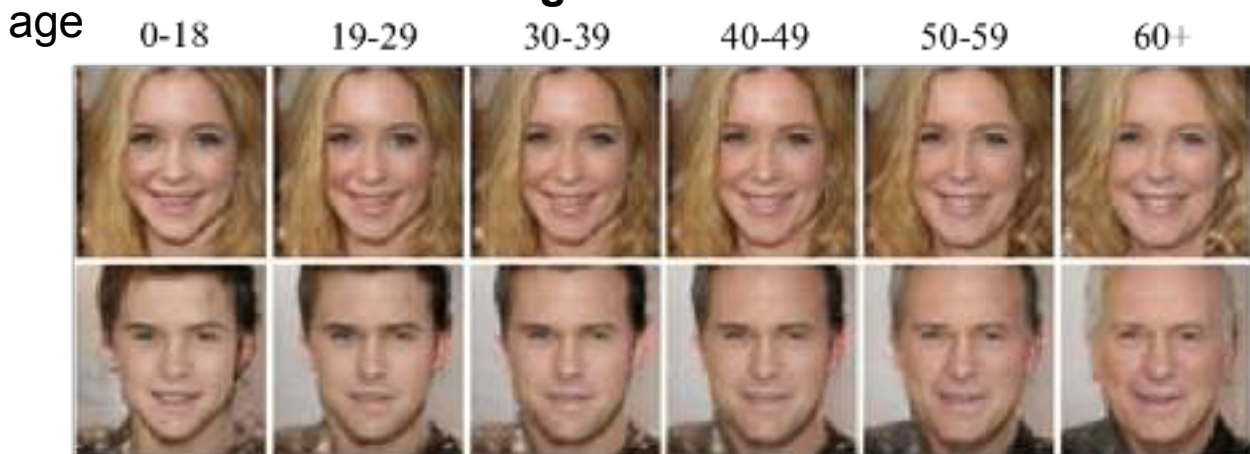


Generation (GAN): Training a model to generate realistic examples



Deep Fake

GAN Image Generation



Exploiting Existing Sensors

Smartphones may have everything you need

Sensors: Motion, sound, ultrasound, camera, infrared, location

Portability: Hours of battery life

Compute: CPU / GPU/ DSP

Storage: Gbs of storage

Connectivity: Wifi, Bluetooth, 4G



Ok. You used data to better understand a problem and machine learning to come up with a solution....now what?

- **Document and share it:** Website, blog, video, social media, github
- **Build a team:** Encourage multidisciplinary collaboration
- **Commercialize it:** Make a pitch deck, market research, find investors
- **Improve it:** Dive deeper, revisit the research, collaborate, make faster, cheaper, more reliable
- **Get a job or degree in the field:** Problem solving experience and thinking outside the box is often valued over grades or degree

Don't stop after one success or failure!

Thanks! Questions?

Feel free to reach out to me to talk about, ideas, projects, jobs, degrees, social distancing, whatever

jakehgarrison@gmail.com