Computational Social Science

Computer vision

Dr. Thomas Davidson

Rutgers University

November 24, 2025

Plan

1. Course updates
2. Introduction to computer vision
3. Sociological applications
4. Using pre-trained image classifiers and object detection models

Course updates

Homework

- Homework 4 grading in progress

Course updates

Projects

- Prototype due tonight, feedback on Canvas this week
- Workshop next Thursday 12/4
- In-class presentations on 12/8

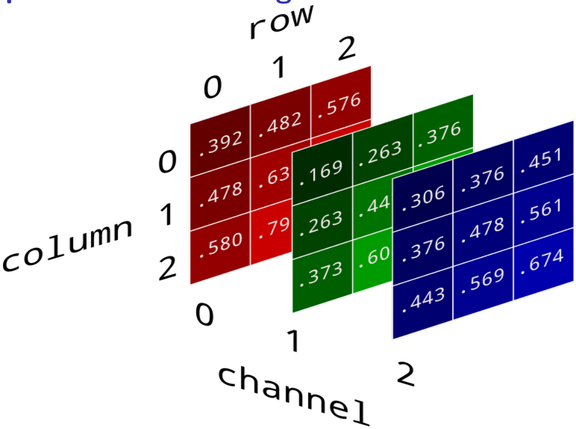
Introduction to computer vision

What is computer vision?

- Computer vision encompasses a range of different machine-learning techniques designed specifically for image-data
- The general principle is the same as other machine-learning approaches
  - Given some image X, predict an outcome associated with the image Y
- The main difference is that the model architecture is adapted to work for image data

Introduction to computer vision

Digital representations of images



Source

Introduction to computer vision

Digital representations of images

- Just like text analysis, images must be normalized for use in machine learning models
- Images can be stored as 3 x 2-D matrices or as a 2-D matrix with normalized RGB values
- Generally, we also reduce the dimensionality of these arrays to aggregate information across pixels to store a lower-dimensional representation of an image (compression)

Introduction to computer vision

Image classification

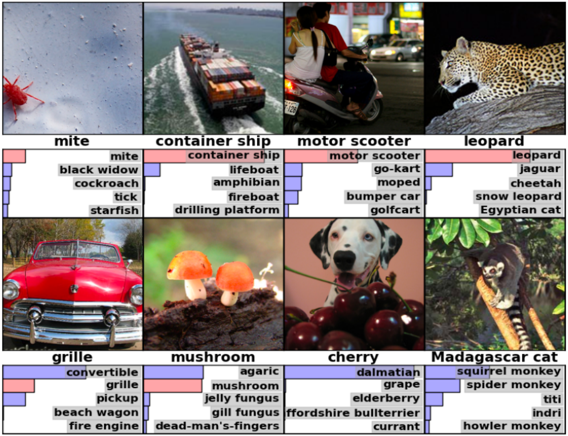
- The goal is to predict the class of image Y, given the 2-D image matrix X.
- In this case, we have an image composed of 9 pixels, which is input into our classifier $f()$.
- Thus, we want to estimate $Y_{class} = f(X)$, where

$$X = \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{bmatrix}$$

- Each element X_{ij} corresponds to the normalized RGB value of a pixel.

Introduction to computer vision

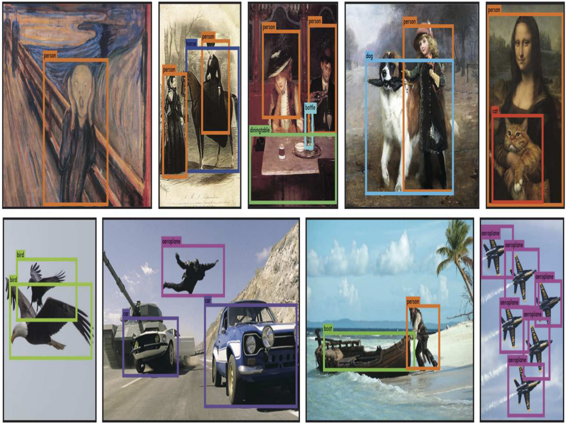
Image classification



Krizhevsky, Sutskever, and Hinton 2012.

Introduction to computer vision

Object detection



Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. "You Only Look Once: Unified, Real-Time Object Detection." *CVPR*, 779–88.

Introduction to computer vision

Image generation

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. "Generative Adversarial Networks." http://arxiv.org/abs/1406.2661.

Introduction to computer vision

Style transfer

Gatys, Leon A, Alexander S Ecker, and Matthias Bethge. 2016. "Image Style Transfer Using Convolutional Neural Networks." *CVPR*, 2414–23.

Introduction to computer vision

How does it work?

- Computer vision has made major breakthroughs in the past decade due to the advances in neural network methods
  - These methods have been around for a long time (the *perceptron* algorithm was developed in 1958) but have been difficult to scale due to computational challenge of estimating vast numbers of parameters
  - This has changed with the availability of large datasets and vast compute power
- In contrast to other approaches involving feature construction, these methods directly "learn" features from the data

Introduction to computer vision

The 2012 ImageNet Challenge

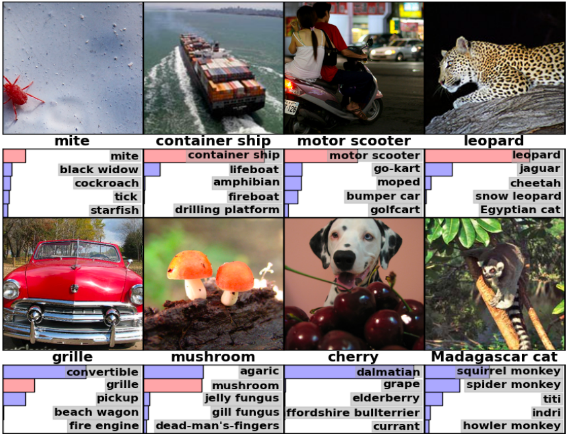
- ImageNet Large Scale Visual Recognition Challenge is a competition to develop a model to classify images into object categories
  - The aim is to construct a realistic setting with millions of images and thousands of objects
- The 2012 ImageNet training dataset contained 1000 different objects categories with labels obtained from the image hosting website flickr.
- Like the Fragile Families Challenge, researchers train models on this dataset then assess performance on a held-out validation set.

The 2012 ImageNet Challenge

- Krizhevsky, Sutskever, and Hinton achieve record-breaking performance by using Convolutional Neural Networks (CNNs)
- CNNs made tractable by using several other methodological innovations
  - Optimization across multiple Graphical Processing Units (GPUs)
  - ReLU activation function for more efficient training
  - Dropout to reduce overfitting
- The random baseline is 0.1% accuracy (1/1000)
  - Their model is correct 63% of the time (Top-1) and the correct answer is in the Top-5 predictions 85% of the time

Introduction to computer vision

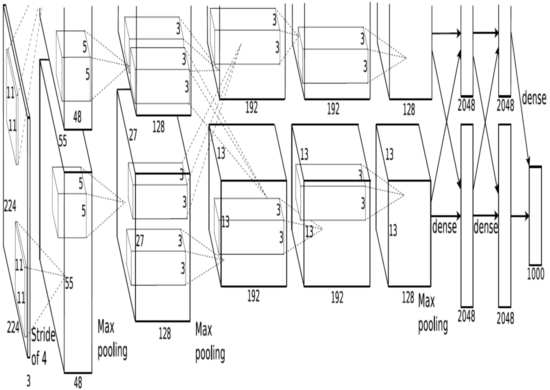
The 2012 ImageNet Challenge



Krizhevsky, Sutskever, and Hinton 2012.

Introduction to computer vision

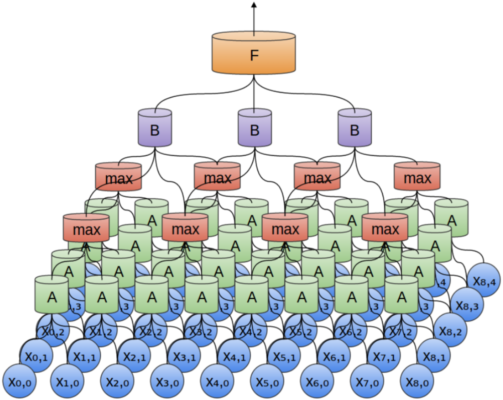
The 2012 ImageNet Challenge



Krizhevsky, Sutskever, and Hinton 2012.

Introduction to computer vision

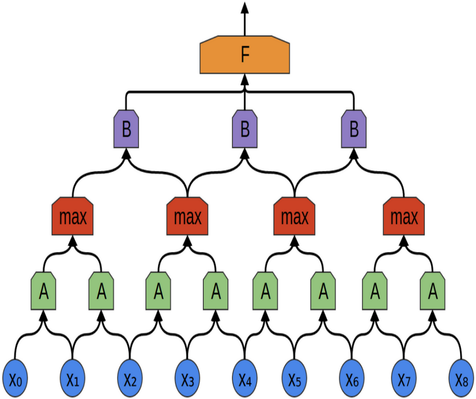
Convolutional Neural Networks



Source: Chris Olah's blog post on convolutional neural networks

Introduction to computer vision

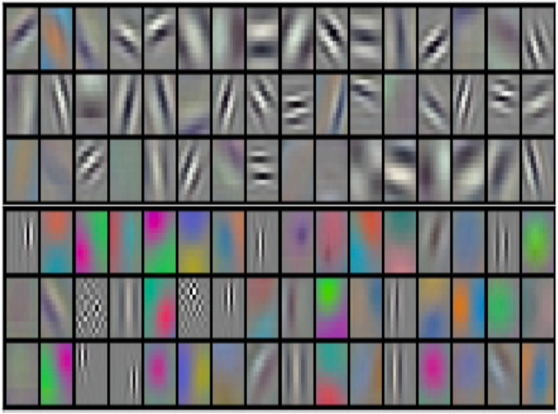
Convolutional Neural Networks



Source: Chris Olah's blog post on convolutional neural networks

Introduction to computer vision

The 2012 ImageNet Challenge



Introduction to computer vision

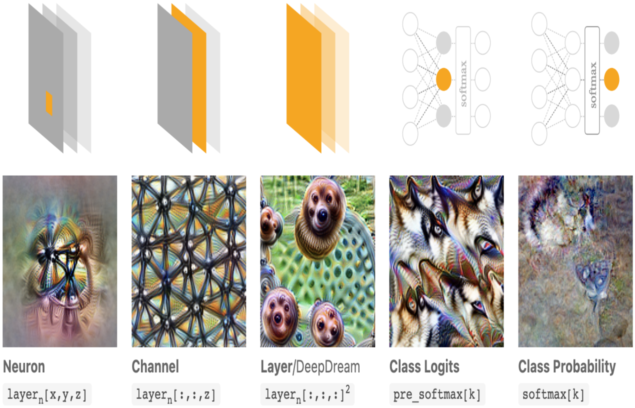
How neural networks see the world

Edges (layer conv2d0) Textures (layer mixed3a) Patterns (layer mixed4a) Parts (layers mixed4b & mixed4c) Objects (layers mixed4d & mixed4e)

Olah et al. 2017

Introduction to computer vision

How neural networks see the world



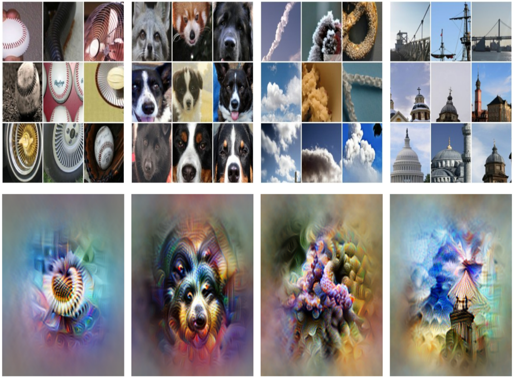
Olah et al. 2017

Introduction to computer vision

How neural networks see the world

Dataset Examples show us what neurons respond to in practice

Optimization isolates the causes of behavior from mere correlations. A neuron may not be detecting what you initially thought.



Baseball—or stripes? *mixed4a, Unit 6*

Animal faces—or snouts? *mixed4a, Unit 240*

Clouds—or fluffiness? *mixed4a, Unit 453*

Buildings—or sky? *mixed4a, Unit 492*

Olah et al. 2017

Introduction to computer vision

Transfer learning

- One of the major breakthroughs in this area of research is *transfer learning*
- A model trained to predict Y can be retrained to predict a new outcome Z
- This is often more efficient than training a model to predict Z from scratch, particularly if we lack a sufficient training data
  - e.g. If we want to train a dog detection model we might want to start with a pre-trained cat detection model
    - The model already "knows" how to detect light intensity, edges, and corners
    - The model may also be able to detect fur, tails, and whiskers, etc.
    - This allows the model to easily adapt to the new task

Introduction to computer vision

Transfer learning

- Most importantly, this makes image recognition a tractable task for social scientists who do not necessarily have access to huge training corpora and advanced compute resources
- We can take an existing pre-trained model and *fine-tune* it to a small corpus of new labeled images
- In practice, this means that we effectively add one or two additional layers to the end of a pre-trained network

Introduction to computer vision

Combining text and images

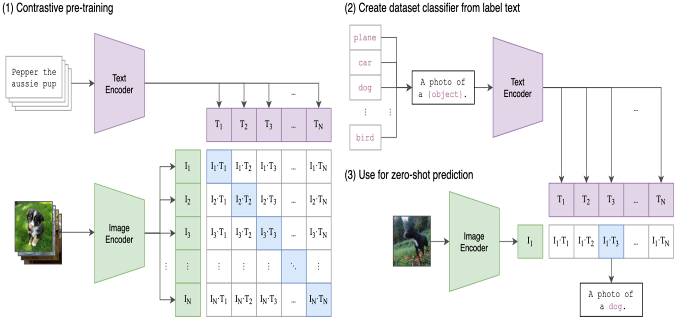
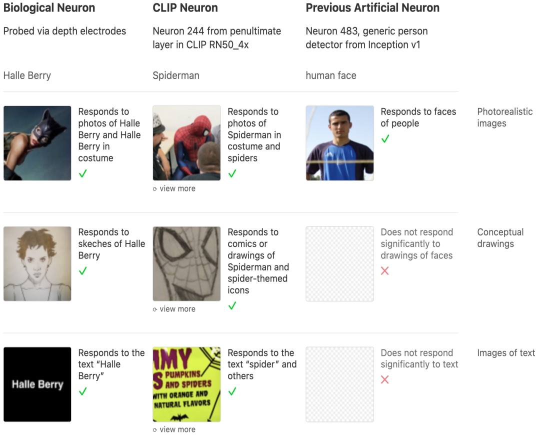


Figure 1. Summary of our approach. While standard image models jointly train an image feature extractor and a linear classifier to predict some label, CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples. At test time the learned text encoder synthesizes a zero-shot linear classifier by embedding the names or descriptions of the target dataset's classes.

Radford et al. 2021

Introduction to computer vision

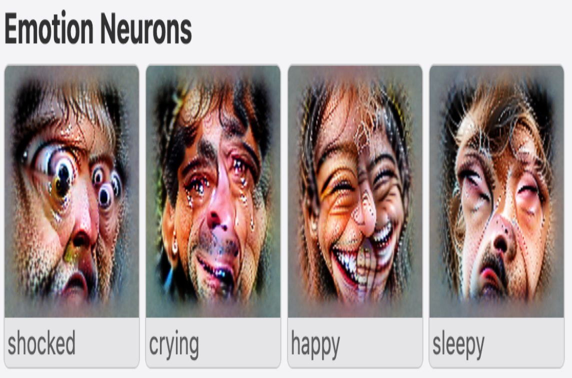
Multimodal neurons



Goh, Gabriel, Nick Cammarata, Chelsea Voss, Shan Carter, Michael Petrov, Ludwig Schubert, Alec Radford, and Chris Olah. 2021. "Multimodal Neurons in Artificial Neural Networks." *Distill* 6 (3): e30. https://doi.org/10.23915/distill.00030.

Introduction to computer vision

Multimodal neurons



Goh et al. 2021

Introduction to computer vision

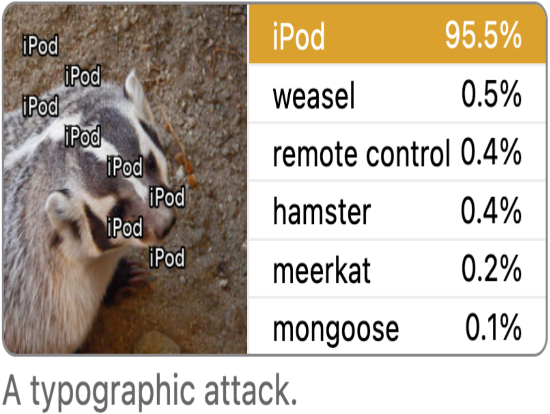
Multimodal neurons

americans, american, america, usa, americas	portuguese, eu, madrid, argentina, portugal	ghana, uganda, africa, tanzania, african	netherlands, luxembourg, stockholm, amsterdam, switzerland	mumbai, singh, pakistan, afghanistan, bangladesh	shanghai, asian, vietnamese, cambodia, chinese	australian, australia, adelaide, nsw, queensland

Goh et al. 2021

Introduction to computer vision

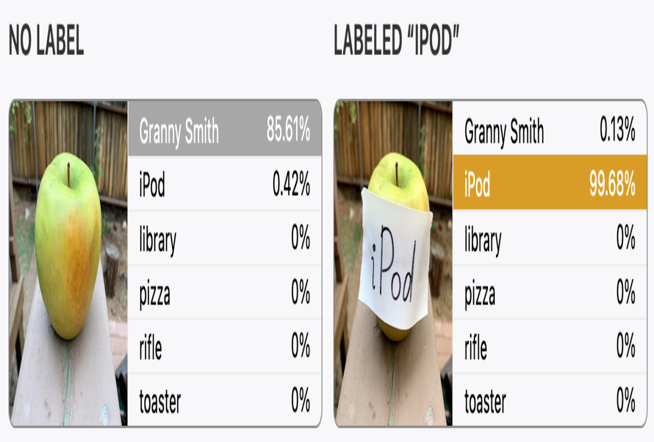
Adversarial examples



Goh et al. 2021

Introduction to computer vision

Adversarial examples



Goh et al. 2021

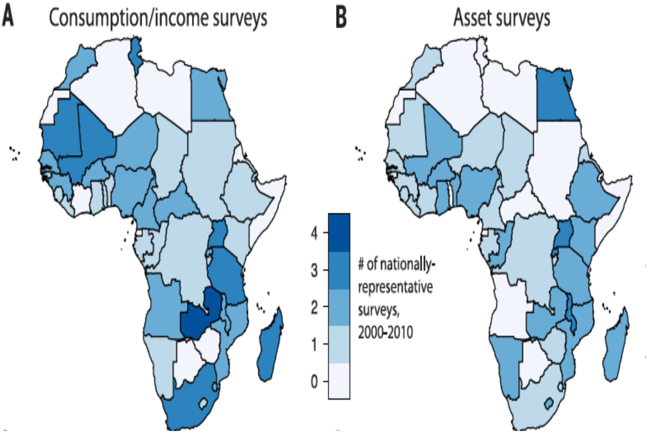
Sociological applications

Examples

- Using satellite imagery to predict poverty
- Using Google Street View to estimate demographics
- Using social media to detect collective action

Sociological applications

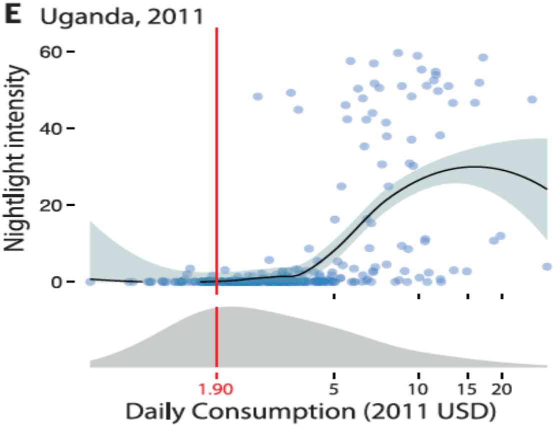
Using satellite imagery to predict poverty



Jean et al. 2016

Sociological applications

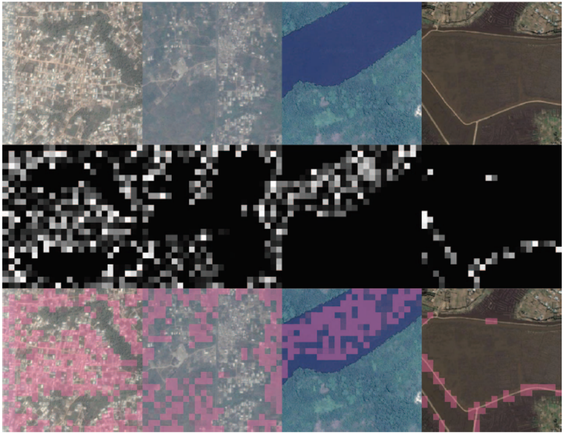
Using satellite imagery to predict poverty



Jean et al. 2016

Sociological applications

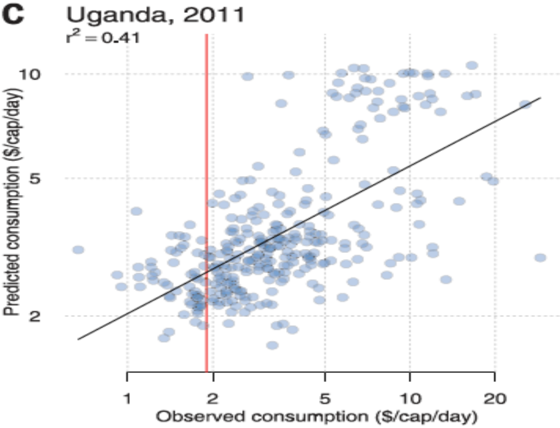
Using satellite imagery to predict poverty



Jean et al. 2016

Sociological applications

Using satellite imagery to predict poverty



Jean et al. 2016

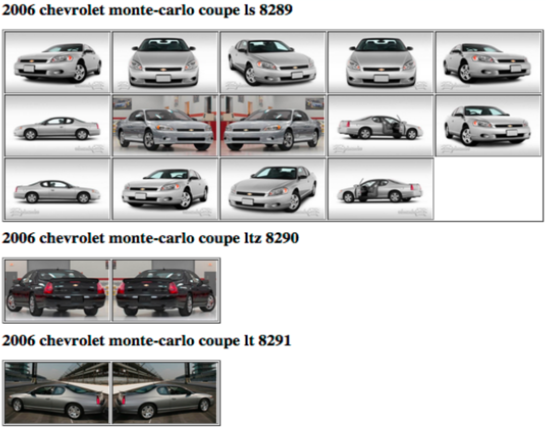
Sociological applications

Using Google Street View to estimate demographics

Gebru et al. 2017

Sociological applications

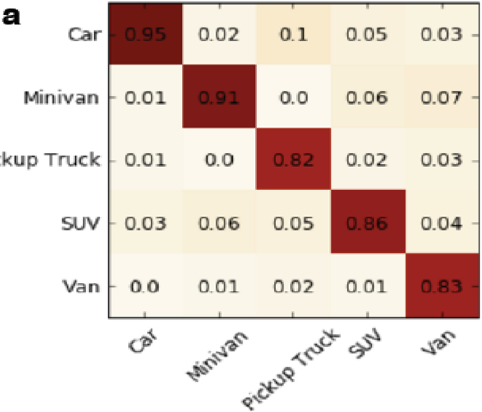
Using Google Street View to estimate demographics



Gebru et al. 2017

Sociological applications

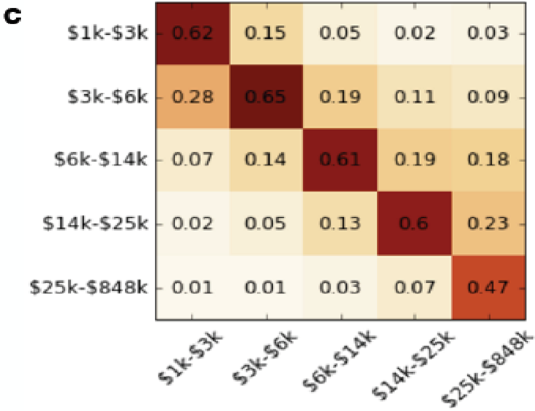
Using Google Street View to estimate demographics



Gebru et al. 2017

Sociological applications

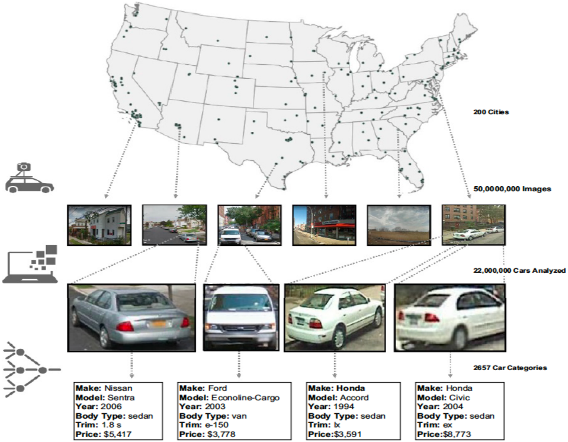
Using Google Street View to estimate demographics



Gebru et al. 2017

Sociological applications

Using Google Street View to estimate demographics



Gebru et al. 2017

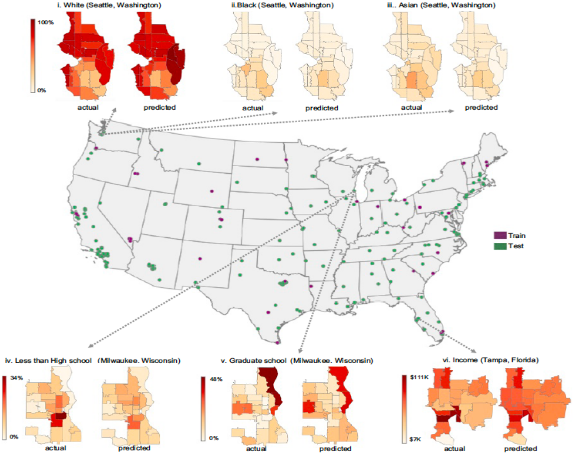
Sociological applications

Using Google Street View to estimate demographics

- The information from the car detection model is then used in a regression model
  - Demographics from the American Community Survey for a given geographic unit are estimated as a function of 88 variables including average number of cars per image, average car price, percent hybrids, percent foreign made, etc.

Sociological applications

Using Google Street View to estimate demographics



Gebru et al. 2017

Sociological applications

Using Google Street View to estimate demographics

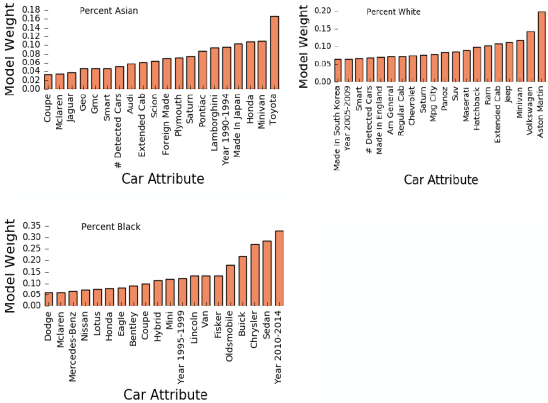


Fig. S2. Bar plots showing the top 10 car features with high positive weight in our race estimation model.

Gebru et al. 2017

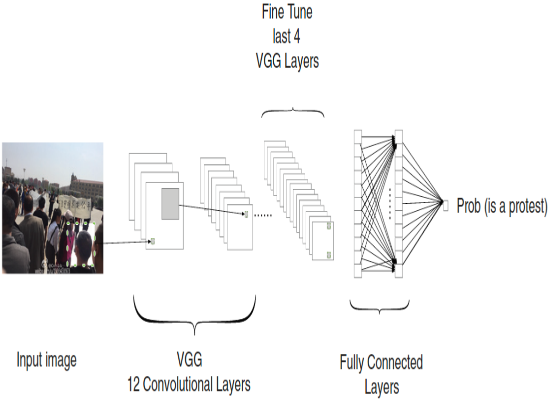
Sociological applications

Using social media to detect collective action

- Zhang and Pan 2019 use data from Weibo to identify collective action events in China
- They develop an event-detection model that combines both image and text from social media posts
- To detect collective action in images, they use a corpus of relevant social media posts collected by activists to fine-tune a pre-trained image detection model

Sociological applications

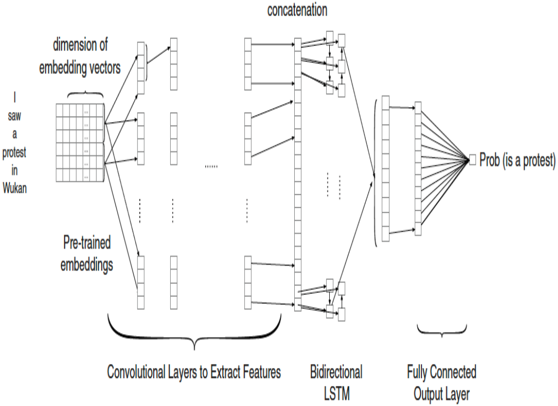
Using social media to detect collective action



Zhang and Pan 2019

Sociological applications

Using social media to detect collective action



Zhang and Pan 2019

Sociological applications

Using social media to detect collective action

Zhang and Pan 2019

Sociological applications

Using social media to detect collective action

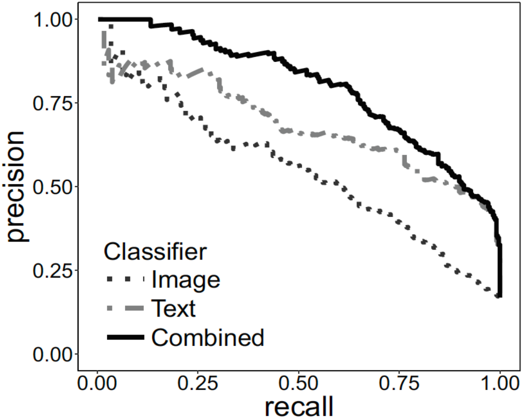
- They develop a simple procedure to classify posts depending on the content:

$$p = \begin{cases} \frac{(p_{\text{text}} + \alpha \cdot p_{\text{image}})}{(1+\alpha)} \cdot \beta & \text{if the post has images,} \\ p_{\text{text}} & \text{otherwise.} \end{cases}$$

- α controls the relative weight of text and images and β denotes the extra weight given to posts with both text and images.
- Both parameters are optimized by using cross-validation.

Sociological applications

Using social media to detect collective action



Zhang and Pan 2019

Sociological applications

Using social media to detect collective action

Why is this type of phenomenon often seen outside the gates of government offices? People are holding onto old ideas, should leaders all be extremely honest and noble? Where is the problem? Will such a country prosper? Will it endure?| Weiyang Middle Road 为什么市政府大门口经常会出现这种现象?百姓迂腐，难道我们 的领导们都很清廉高尚嘛?问题到底出在哪?这样的国家到底会不会繁荣昌 盛?会不会长久下去?| 渭阳中路	In front of the Office of Letters and Visits……X ah! Dear mother! The aggrieved are kneeling for you, standing for you! Don't kneel. 国家信访局门前......X 啊 亲娘 冤民给你下跪了站直了 不要 跪

Zhang and Pan 2019

Sociological applications

Using social media to detect collective action

...In recent days, eight Feidong migrant workers asked the sub-district for help in obtaining back wages totaling 40 thousand yuan...the sub-district procuratorate immediately launched a legal aid program for migrant workers, and after 7 days of effort, migrant workers were paid the back wages. Look, migrant workers even sent a banner for the staff

…近日，肥东县古城镇陈天扬等8名肥东县农民工，向海棠街道寻求帮助，他们反映辖区一建筑工地工程承程包商，去年至今共拖欠工资4万余元。街道司法所立即启动为农民工讨薪法律援助程序，经过7天的工作，承包商程某终于偿还了工钱。瞧，农民工还为工作人员送来了锦旗呢

Zhang and Pan 2019

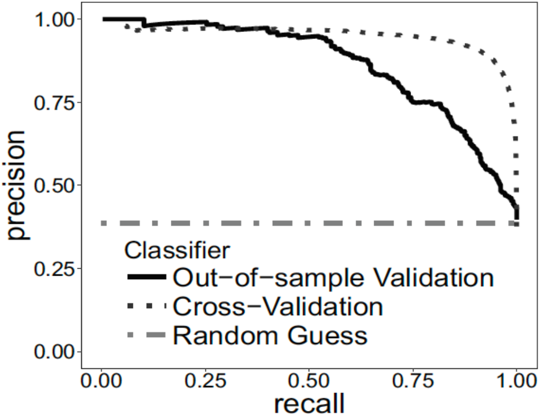
Sociological applications

Using social media to detect collective action

- Many false positives contain content relevant to collective action such as corruption, housing demolition, and the police, but do not actually involve collective action
- They therefore train a second-stage classifier to distinguish between collective action and such cases
- This improves the out-of-sample F_1 score from 0.69 to 0.84, demonstrating how analysis of model results can enable us to make improvements to classifiers

Sociological applications

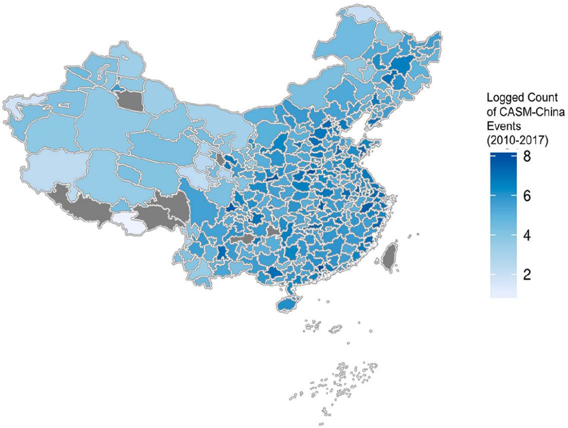
Using social media to detect collective action



Zhang and Pan 2019

Sociological applications

Using social media to detect collective action



Zhang and Pan 2019

Using pre-trained image classifiers and object detection models

Why not R?

- Python is the dominant language used in ML research
  - Well developed package infrastructure (`scikit-learn`, `keras`, `PyTorch`, `Tensorflow`)
  - Tons of resources and replication materials in Python
    - See Papers with Code for a collection of replication materials, including many state-of-the-art models
- Most R packages are wrappers around existing Python packages
  - RStudio now supports TensorFlow and Keras
- Google Colaboratory notebooks and TensorFlow Hub make it possible to work with large models and datasets
  - It is possible to use an R kernel in a Colab notebook

Using pre-trained image classifiers and object detection models

Using pre-trained models from TensorHub

- Click here to open an example notebook for using a pre-trained image classifier.

Summary

- Computer vision describes a domain of machine learning focuses on image data
- These methods have made huge advances in the past decade due to increased compute power and the availability of large datasets
- Convolutional neural networks are one of the most prominent and widely used approaches
  - Recent advances incorporate different kinds of data, including images and text
- Transfer learning and fine-tuning make it possible to adapt advanced methods to new tasks with relative ease

Next class

- Thanksgiving break, no class on Wednesday
- Next week
  - Generative AI