Computational Social Science with Images and Audio

Elliott Ash, **Philine Widmer**

22 September 2023

Human vision is interesting (not only for social scientists).

- Vision is an incredibly powerful sense
- Allows us to interact with the physical world without making any physical contact
- A *substantial* part of our brain is somehow involved in visual perception
  - Of course, there are many limitations of human vision (e.g., low quantification skills)
- Human vision is interesting on many levels:
  - Fundamental: How do biological vision systems operate?
    $\rightarrow$ Not part of this class (or marginally)
  - Aggregate/societal: In what ways does what and how we see shape societal outcomes?
    $\rightarrow$ The focus of this class

Computer vision comes with many definitions.

- David Marr: automating human visual processes
- Others see it as an information-processing task
- Berthold Horn: inverting image formation
- Perhaps: the inverse of graphics?
- In this class, we look at techniques to analyze visual data (images, videos) in large quantities to study social science questions
  - We start where the visual data already exists: we do not cover, from a technical viewpoint, how to get that visual data (e.g., optics)

Essentially, our class starts with something like this.

```
array([[131, 131, 131, ..., 107, 106, 106],
       [131, 132, 132, ..., 107, 107, 106],
       [132, 132, 133, ..., 108, 107, 106],
       ...,
       [192, 192, 192, ..., 189, 187, 185],
       [193, 193, 193, ..., 189, 187, 185],
       [194, 194, 194, ..., 188, 186, 184]], dtype=uint8)
```

Figure: An array of (549, 976)

What the array represents:

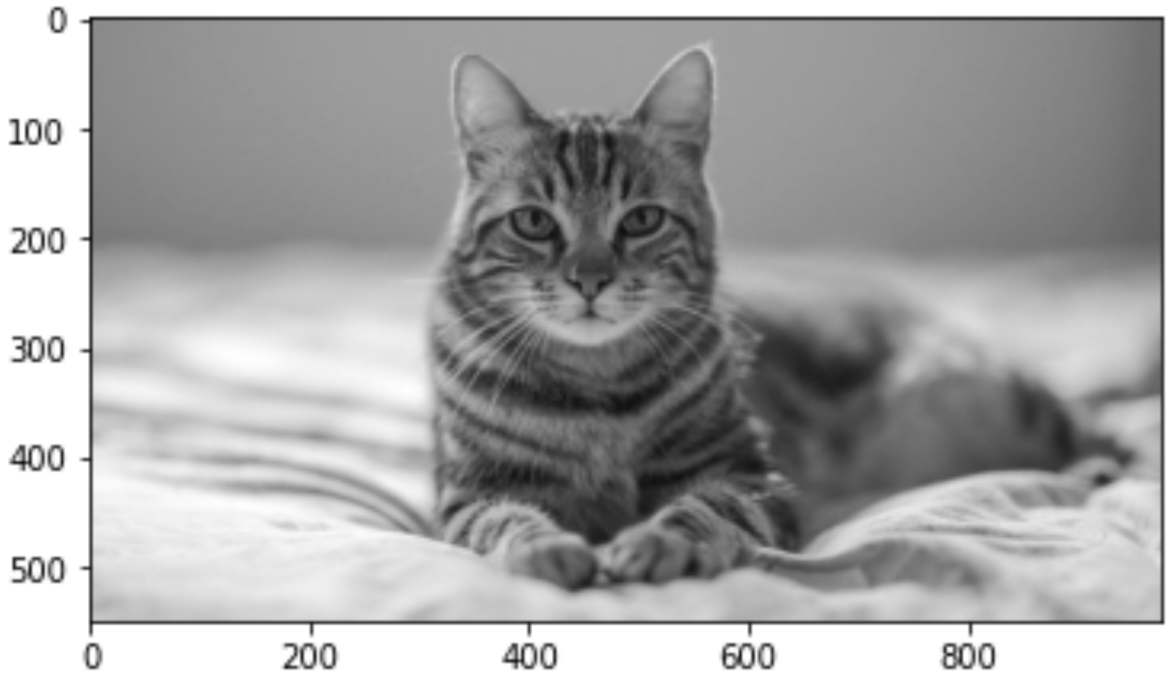

Figure: Meow.

What about the social science part?

- Social science is the study of the complex interplay between individuals, groups, and institutions in society and ...
- How individuals, groups, and institutions shape and are shaped by social, cultural, political, and economic forces
- Auguste Comte: a discipline that studies society positively/scientifically
- Max Weber: the study of social action, or human behavior as it relates to the social context in which it occurs

Computer vision is increasingly important in social science.

- Images significantly contribute to and mirror the social/political landscape
- Digitization has led to an increased accessibility of such images in everyday life
  - Consequently, this provides new and valuable opportunities for researchers and practitioners alike (both in CV and social science)
- Advances in computer vision significantly ease the process of using images as data
  - For example, deep learning techniques for identifying objects, measuring visual sentiment, or recognizing faces
  - Innovation in computer vision has large societal effects!
- Great opportunity comes with great responsibility...

Digitization has led to an explosion in visual material.

- There has been an explosion in the production and consumption of images
- On Youtube, an estimated 3.7 million videos are uploaded per day[1]
- Users upload 1.3 billion images to Instagram per day[2]
- (Such figures are estimates and can change fast but are enormous in any case...)
- Can you think of other major sources of visual data?

[1] Adavelli, M., & Tonogbanua, L. (2023). How Many Videos Are Uploaded to YouTube A Day? TechJury. https://techjury.net/blog/how-many-videos-are-uploaded-to-youtube-a-day/

[2] Growcoot, M. (2023). Almost All Photos Are Now Taken on Smartphones, According to Study. PetaPixel. https://petapixel.com/2023/06/20/almost-all-photos-are-now-taken-on-smartphones-according-to-study/

Studying images in social and political life is not new.

- For instance, research has shown that images are key to agenda-setting and framing in newspaper coverage
- Images can impact people's perceptions of political candidates and their votes and ...
- Encourage (or discourage) political participation
- Generally, images seem to have a greater effect than written and spoken content in capturing attention
- They can ease information processing, improve information recall, and evoke emotions[3]

[3] Examples of relevant papers can be found in Webb Williams, N., Casas, A., and Wilkerson, J. (2020). Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge: Cambridge University Press. doi:10.1017/9781108860741

This is an early example on how visual material in the media shapes political discourse.

- A "classic": Gitlin's "The Whole World Is Watching: Mass Media in the Making and Unmaking of the New Left" (1980)

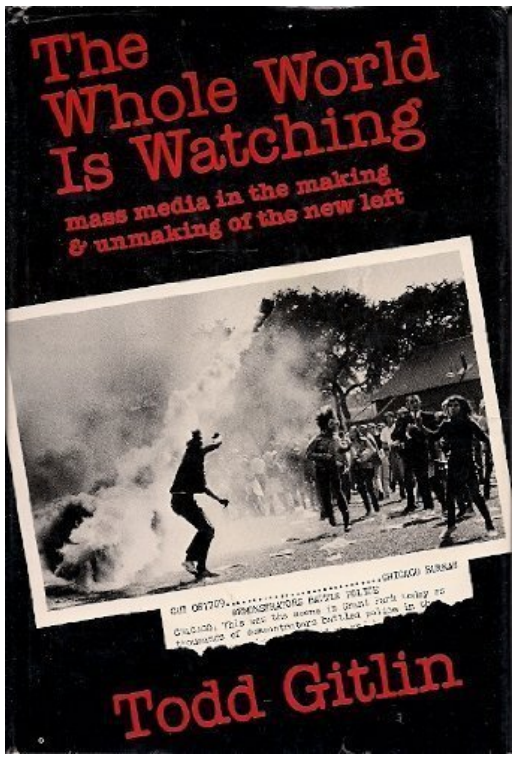

Figure: Gitlin (1980)

The abundance of images, however, is new!

- We can challenge existing knowledge in new ways, and generate new knowledge
- Social science work with visual data comes in two categories:
  - Images measure variables of interest in new ways
  - Images are the object of study (in a causal or non-causal framework, with image-based dependent or independent variables)
  - (Fuzzy, overlapping distinction)
- Can you think of examples for the above categories?
  - Can we detect fraud in elections by using original image databases of the vote-tally sheets to detect alterations?[4]
  - What makes politicians' Instagram posts popular?[5]

[4] Cantu, F. (2019). The Fingerprints of Fraud: Evidence from Mexico's 1988 Presidential Election. American Political Science Review, 113(3), 710-726.

[5] Peng, Y. (2021). What Makes Politicians' Instagram Posts Popular? Analyzing Social Media Strategies of Candidates and Office Holders with Computer Vision. The International Journal of Press/Politics, 26(1), 143–166.

Can we detect election fraud with images? (Cantu, 2019)

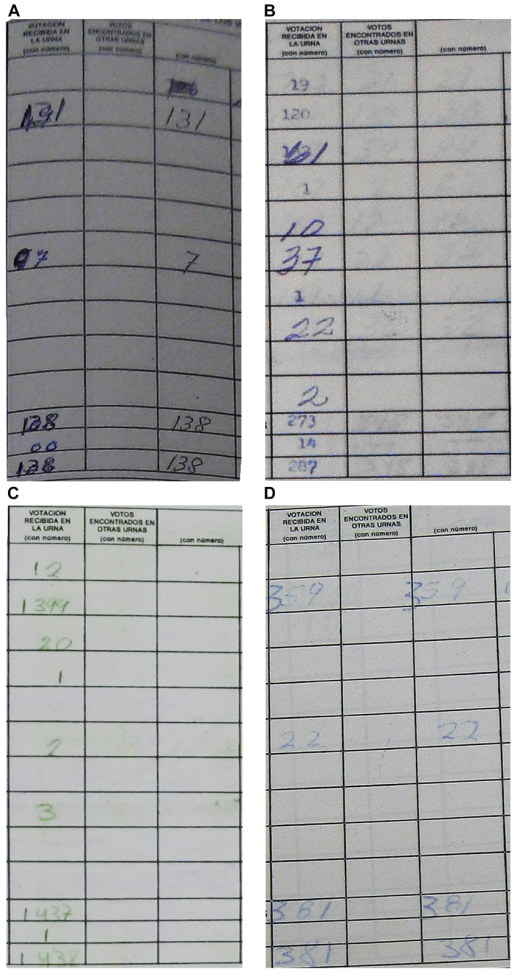

Figure: Examples of altered vote tallies in Mexico, 1988 (from Cantu, 2019)

Politicians' Instagram: what is popular? (Peng, 2021)

Figure: Screenshot of New Jersey Senator Bob Menendez' Instagram (2017)

Taking a step back: What is an image?

- Simplest form (single-channel): two-dimensional function $f(x, y)$
  - Maps a coordinate pair to an integer/real value related to the intensity/color of the point
  - Single-channel: binary or mono-chrome, gray-scale, black/white images
- Each point is called a **pixel** (picture element) or pel
- Multiple channels possible, such as RGB:
  - Color represented using three channels
  - Pixel at point (x, y) represented as $(r_{x,y}, g_{x,y}, b_{x,y})$
- Channel-specific pixel value represented as an integer between 0 and 255 or a floating-point value in $[0, 1]$

Taking a step back: What is an image?

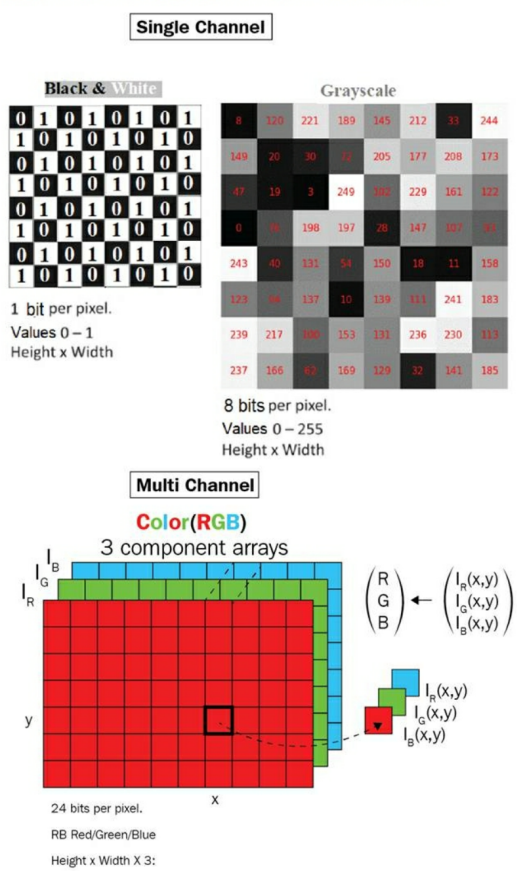

Figure: Dey (2018)[6]

[6] Dey, S. (2018). Hands-On Image Processing with Python: Expert Techniques for Advanced Image analysis and Effective Interpretation of Image Data. Packt Publishing Ltd.

CNN are the current computer vision workhorse.

But not the only horse in the stable!

- Convolutional neural nets (abbreviated as CNNs or CovNets) are specialized neural networks for CV tasks
  - For various tasks, such as object detection, facial recognition, image segmentation, sentiment analysis
- Much of the recent image-based social science research uses CNN
- Other approaches are used, too:
  - Classic techniques: in the pre-deep learning and pre-CNN era, CV relied on hand-crafted features and algorithms
    - Edge detection, color histograms, template matching, ...
  - Other neural nets: recurrent neural networks (RNN), generative adversarial networks (GAN), transformers, stable diffusion models
    - RNN are often used for video analysis, GAN for image generation
  - Hybrid approaches: traditional approaches and deep learning
  - Unsupervised and semi-supervised learning: e.g., clustering

Classic techniques typically require more explicit "rules".

- Imagine you search for images with a specific political leader
- Should you create a checklist of physical characteristics?
  - Shape of their nose, size of forehead?
  - Not robust (facing the camera? sunglasses?)
  - An infinite list of rules?
- Fairly general features can be helpful for various tasks
  - Corners
  - Blobs (dark/bright region)
- **Feature detection**: identification of points of interest
- Explicit feature detection in classic or "traditional" machine learning

Classic techniques typically require more explicit "rules".

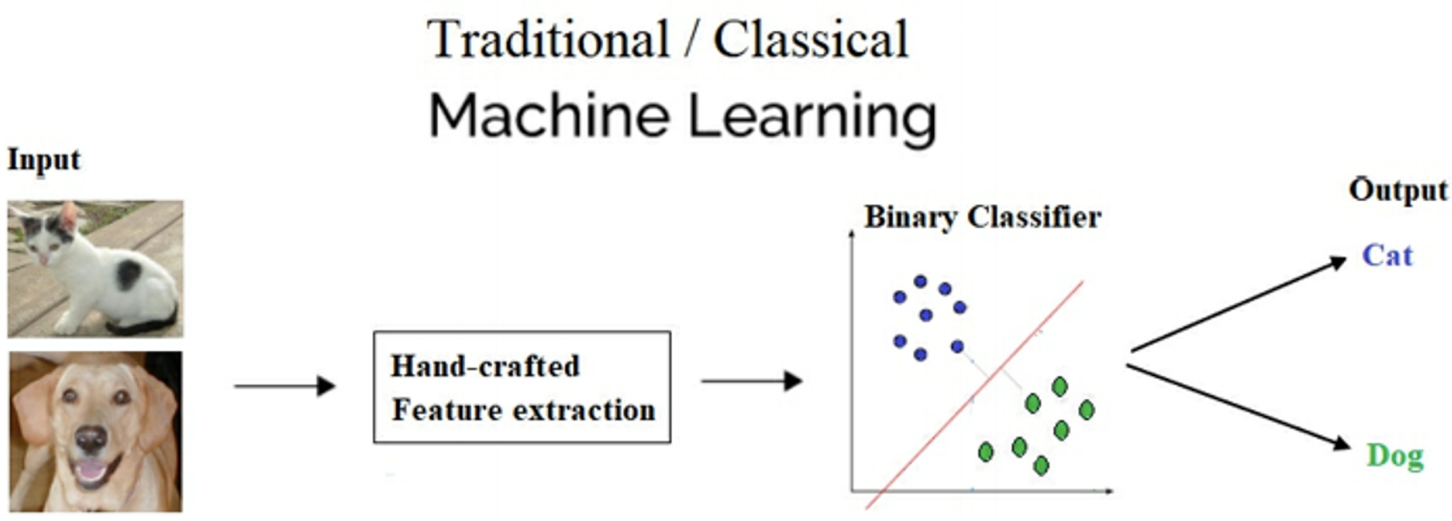

Figure: Dey (2018)[7]

[7] Dey, S. (2018). Hands-On Image Processing with Python: Expert Techniques for Advanced Image analysis and Effective Interpretation of Image Data. Packt Publishing Ltd.

One promise of neural nets is a more end-to-end approach.

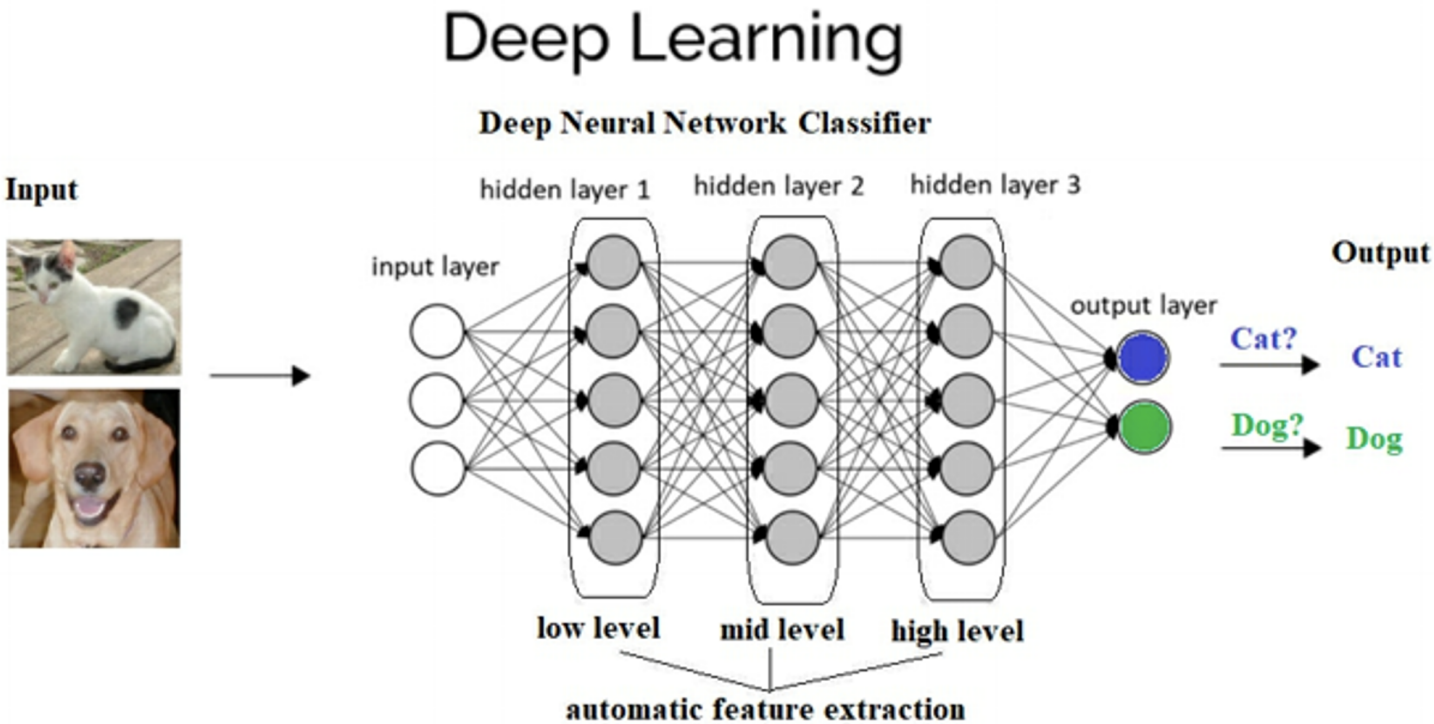

Figure: Dey (2018)[8]

[8] Dey, S. (2018). Hands-On Image Processing with Python: Expert Techniques for Advanced Image analysis and Effective Interpretation of Image Data. Packt Publishing Ltd.

CNN typically contain an input layer, hidden layers, and an output layer.

- Typically: input layer $\rightarrow$ hidden layers $\rightarrow$ output layer
- What do different convolutional layers do?
  - Layers close to the input layer learn low-level features (e.g., lines)
  - Middle layers learn complex abstract features (combining lower level features)
  - Layers closer to the output interpret the extracted features in the light of the classification task