Class 12: Image Basics

Theme: Images

Computational Analysis of Text, Audio, and Images, Fall 2023

Aarhus University

Mathias Rask (mathiasrask@ps.au.dk)

Aarhus University

Table of Contents

Motivation I

- The study of images has traditionally been subject to qualitative research or small-scale quantitative studies (e.g. Bodies as Battleground)
- Three reasons:
  1. Images contain a wealth of contextual information that is not easily quantifiable for quantitative analysis
  2. A small and selected sample of images can hold societal significance that they warrant comprehensive studies to understand how they shape or represent the world
  3. Computers have not been accessible or tuned to analyze large-scale piles of images
- Today: Advances in computers and computer vision has made large-scale image analysis possible
  - ⇝ Like for text and audio: computers help to scale our analysis

Challenge: Chihuahua or Muffin?

Motivation II

Images has shown to:

affect:

- agenda-setting and framing
- candidate evaluation and characteristics
- political participation
- information processing and attention
- ...

predict:

- housing prices
- protest dynamics
- natural disasters
- electoral outcomes
- ideology
- ...

measure:

- local economic activity
- state capacity
- ...

Motivation III: “A Picture is Worth a Thousands Words”

Images affect, predicts, and measures a lot, but why?

- Powerful: extra information + emotional activation ≈ better recall and higher engagement
- Universal: across cultures compared to e.g. language ≈ high reach
- Frames: Visuals tell us to what pay attention to

Table of Contents

Measurement Approaches

Theorizing/Proxying

- Frame differences (Dietrich, 2021)
- Coloring (Chen *et al.*, 2022)
- Remote sensing (Jean *et al.*, 2016)

Learning/Prediction

- Object detection and recognition
- Face detection and recognition
- ...

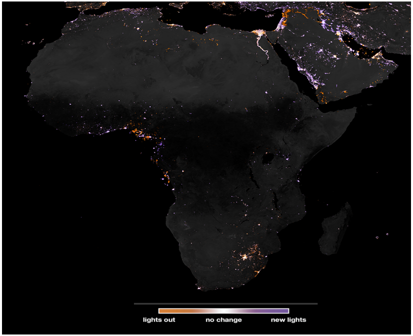
Proxying I: Colors Cues of Conspiracy (Chen *et al.*, 2022)

HSV saturation = 0.000962 (low extreme)
https://i.ytimg.com/vi/OvVgJHDUpTg/hqdefault.jpg

HSV saturation = 19.63112 (high extreme)
https://i.ytimg.com/vi/FrBk1xXvYzY/hqdefault.jpg

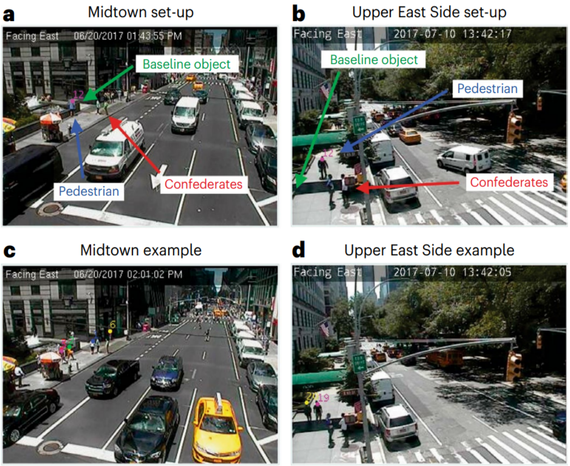
⇝ Conspiracy videos use less color than fact-correcting videos
https://www.youtube.com/watch?v=kLMMxgtxQ1Y

Proxying II: Satellite Images as a Proxy for Economic Activity (Ghosh *et al.*, 2010)



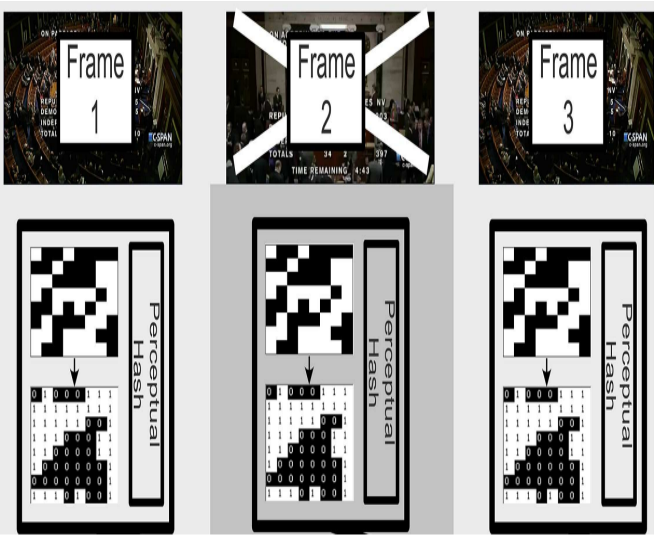
- ⇝ Jean *et al.* (2016) combine daylight and night satellite images to predict poverty
- ⇝ Livny (2021) develop a measure of religiosity based on nighttime light

Proxying III: Racial Avoidance (Dietrich and Sands, 2023)



⇝ Pedestrians move 4 inches (10.16 cm) further away the confederate was black compared to white.

Proxying IV: Motion Polarization (Dietrich, 2021)

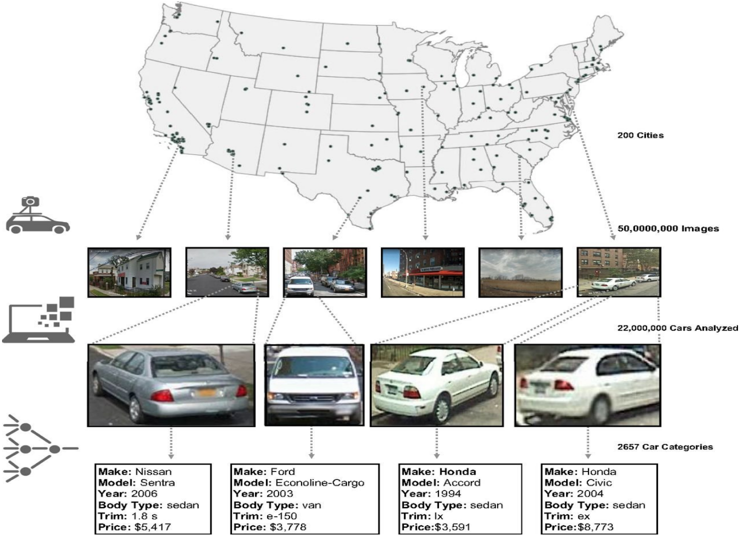


⇝ Party voting is more likely when motions decline (i.e. lower frame difference)

Learning I: Protest Dynamics (Steinert-Threlkeld *et al.*, 2022)

Figure 1. *A*, Sample images and the protest classifier's rating of them. *B*, Protest images with their state violence rating. *C*, Protest images' protester violence rating. Labels specify each image's city and label probability. The use of hard negatives in the training set ensures that scenes that contain crowds (*A*, *left*), individuals walking on streets (*C*, *left*), or a nonprotest sign (*A*, *second*) are not included in analysis. Color version available as an online enhancement.

Learning II: Demographic Makeup of Neighborhoods (Gebru *et al.*, 2017)



Learning III: Visual Communication of Ideology (Xi *et al.*, 2020)

(a) Paul Ryan (R) with the American Flag

(b) Lindsey Graham (R) with the US military

(c) Louise Slaughter (D) posing with non-white people

(d) Paul Ryan (R) poses with Vietnam Veteran

Learning IV: Representation of Social Gender and Race in Children's Books (Adukia *et al.*, 2023)

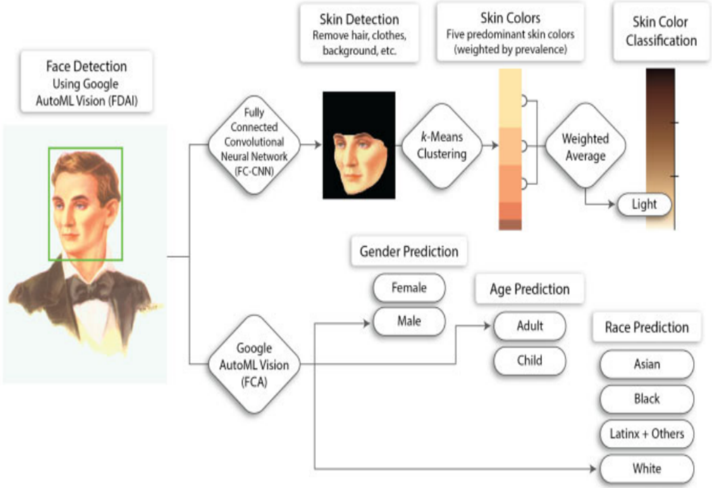


Table of Contents

A Primer on Computer Vision

Computer Vision is an umbrella term that describes any task where we use computers to analyze images.

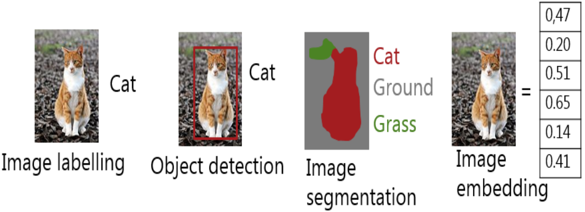
- Main task: Teaching computers to “see”
- Main requirement: To be able to process massive collections of images
- Main issue: Reproducing and amplifying bias
  - racial bias in healthcare (Obermeyer *et al.*, 2019)
  - gender bias in image representations (Wang *et al.*, 2019)

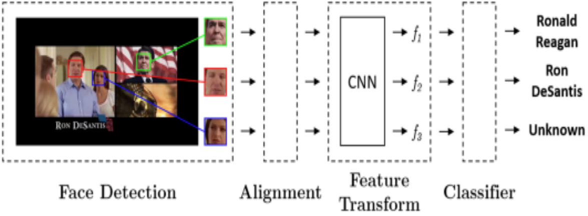
Virtually any task that uses computer vision employs Neural Networks – and in particular Convolutional Neural Networks (CNNs)
$\rightsquigarrow$ More on those next week!

Common Tasks

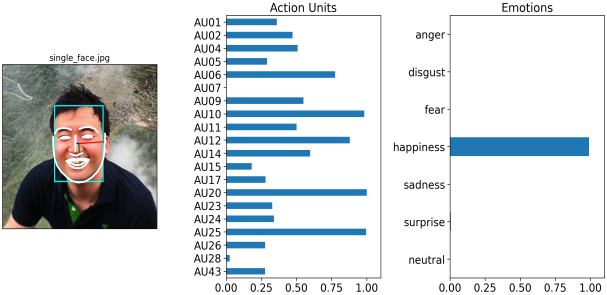
- Object recognition: Identify object(s) in an image
- Object detection: Object recognition + identifying their spatial locations
- Image segmentation: Labeling each pixel in the image with its corresponding class
- Image embedding: Representing images as meaningful vectors of numbers
- Face detection: Identify face(s) in an image
- Face recognition: Face detection + recognize the identify the face(s)
- Face attributes: Face detection + gender, ethnicity, age, …
- Facial expressions: Face detection + anger, disgust, happy, …
- Head/body movement: Person detection + identify how head/body moves
- Image captioning: Generate a caption of an image
- Text detection: Recognize text in an image and convert to text
- …

Object Recognition, Detection, Image Segmentation, and Embedding





Facial expressions



Head-Pose Detection

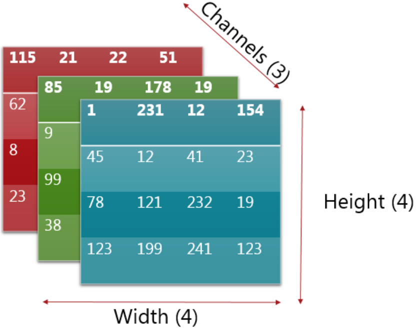
- GitHub

Table of Contents

What is an Image?

What is an image from a computer's perspective?

- Multidimensional arrays (i.e. tensors) containing pixel values
- An image is a stacked matrix with the values being the pixel intensity values.
- Consists of three dimensions: height, width, and channel.
  - Height: Duh...
  - Width: Duh...
  - Channel: Colors
    - ▷ Grayscale: One channel – light intensity
    - ▷ RGB: Red (R), Green (G), and Blue (B) – color intensity



⇝ Three matrices stacked together, one for each channel (Red, Green, and Blue)

Pixel Representation

Image Sampling:
The height and width of an image are entirely equivalent to the logic of sampling frequency for audio signal

$\rightsquigarrow$ The higher and wider, the *more* pixel values we have

$\rightsquigarrow$ The image frequency must be equal to or greater than twice the frequency associated with the finest detail in the image (edges).

Image Quantization:
The values each pixel can take is determined by the bit-depth

8-bit example:

Bit	7	6	5	4	3	2	1	0
Data	R	R	R	G	G	G	B	B

Maximum value becomes: $2^8 = 256$

Table of Contents

Steps

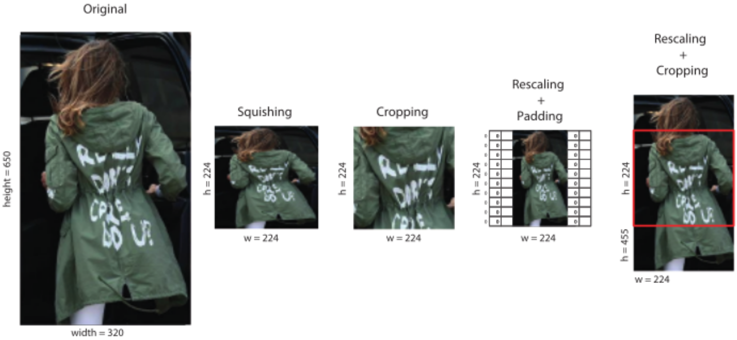
Preprocessing images to be used as input for CNNs typically involves three steps:

1. Resizing
2. Contrasting
3. Augmentation

Resizing

The first step is to shape each image into squares of the same size ⇝ squaring:

- Squashing
- Padding
- Squishing



Contrasting

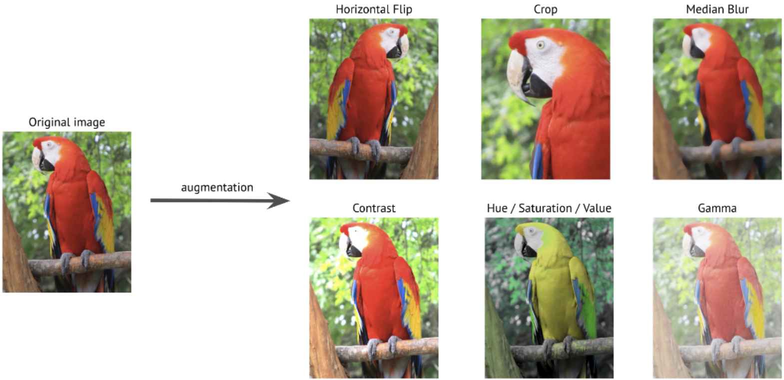
We almost rescale images to the same pixel range every time we use a classifier:

- Normalization ($x \in [0, 1]$)
- Centering or standardization (mean zero, unit variance)

Why do we need to take contrasts into account?

Augmentation

Augmentation is a popular approach to increase the size of your training set artificially



Why does this improve the generalizability of a classifier?

Table of Contents

See you next week!

Topic 4: Images

Computational Analysis of Text, Audio, and Images, Fall 2023

Aarhus University

References i

[1] B. J. Dietrich, "Using motion detection to measure social polarization in the us house of representatives," *Political Analysis*, vol. 29, no. 2, pp. 250–259, 2021.

[2] K. Chen, S. J. Kim, Q. Gao, and S. Raschka, "Visual framing of science conspiracy videos: Integrating machine learning with communication theories to study the use of color and brightness," *Computational Communication Research*, vol. 4, no. 1, 2022.

[3] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016.

[4] A. Livny, "Can religiosity be sensed with satellite data? an assessment of luminosity during ramadan in turkey," *Public Opinion Quarterly*, vol. 85, no. S1, pp. 371–398, 2021.

[5] T. Ghosh, R. L Powell, C. D Elvidge, K. E Baugh, P. C Sutton, and S. Anderson, “Shedding light on the global distribution of economic activity,” *The Open Geography Journal*, vol. 3, no. 1, 2010.

[6] B. J. Dietrich and M. L. Sands, “Seeing racial avoidance on new york city streets,” *Nature human behaviour*, pp. 1–7, 2023.

[7] Z. C. Steinert-Threlkeld, A. M. Chan, and J. Joo, “How state and protester violence affect protest dynamics,” *The Journal of Politics*, vol. 84, no. 2, pp. 798–813, 2022.

[8] T. Gebru *et al.*, “Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 50, pp. 13 108–13 113, 2017.

[9] N. Xi, D. Ma, M. Liou, Z. C. Steinert-Threlkeld, J. Anastasopoulos, and J. Joo, “Understanding the political ideology of legislators from social media images,” in *Proceedings of the international aaai conference on web and social media*, vol. 14, 2020, pp. 726–737.

[10] A. Adukia, A. Eble, E. Harrison, H. B. Runesha, and T. Szasz, “What we teach about race and gender: Representation in images and text of children’s books,” *The Quarterly Journal of Economics*, vol. 138, no. 4, pp. 2225–2285, 2023.

[11] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, “Dissecting racial bias in an algorithm used to manage the health of populations,” *Science*, vol. 366, no. 6464, pp. 447–453, 2019.

[12] T. Wang, J. Zhao, M. Yatskar, K.-W. Chang, and V. Ordonez, "Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 5310–5319.