GESIS Fall Seminar in Computational Social Science 2023

Syllabus for week 2:
"Automated Image and Video Analysis with Python"

Lecturers:	Andreu Casas	Felicia Loecherbach
Affiliation:	Vrije Universiteit Amsterdam	New York University
Email:	a.casassalleras@vu.nl	fl2439@nyu.edul

Date: September 18-22, 2023
Time: 09:00-16:00

About the Lecturers

Andreu Casas is an Assistant Professor in the Department of Communication Science at the Vrije Universiteit Amsterdam and a Faculty Associate at the Center for Social Media and Politics at New York University. His research interests encompass the areas of political communication, public policy processes, and computational social sciences. His methodological interests and strengths are natural language processing (text as data), computer vision (images as data), and machine learning and artificial intelligence more generally. His work has been published in *Science Advances, American Political Science Review, American Journal of Political Science, Annual Review of Political Science, Cambridge University Press*, among others.

Felicia Loecherbach is a Postdoctoral Fellow at the Center for Social Media and Politics (CSMaP) at New York University. Her research focuses on the diversity of (online) news consumption, with the aim of building a better understanding of the impact that changes in online environments have on the understanding and usage of news. She applies and develops innovative computational methods to automatically analyze large amounts of media and trace data. Her work has been published in *Digital Journalism, Computational Communication Research, Media and Communication and the Proceedings of ACM*.

Course Description

Social scientists have long argued that images play a crucial role in shaping and reflecting political life. This role is heightened by the bombardment of images that people experience today through many communications channels, from television to social media. Digitization has both increased the presence of images in daily life and made it easier for scholars to access and collect large quantities of pictures. However, using images collected in observational settings as data for social science inference is an arduous task. Fortunately, recent innovations in computer vision, the subfield of computer science concerned with automated image analysis, can reduce the costs of using images as data.

In this course, we'll dig into the necessary theoretical and methodological expertise needed to apply machine learning methods to address social science questions. We will combine theoretical sessions where we'll discuss research using computer vision methods for the study of politics, communication science, etc., with sessions where we'll cover in detail key methodological advances needed to fully understand state–of–the–art computer vision methods (deep learning, neural networks, convolutional neural networks, etc.), as well as practical sessions where we'll go over several python tutorials implementing different computer vision techniques, for image processing (e.g. splitting videos into analytical frames), object and face detection, image (supervised and unsupervised) classification, facial trait analysis, etc.

Students with basic programming skills/experience in Python and some machine learning background will get the most out of the course. However, we'll also take the time to briefly review some key machine learning concepts

necessary to implement machine learning methods, and students will be provided with clear and easy-to-follow sample code for each of the practical tutorials. By the end of the course, students will have a good understanding of the kind of research questions that can be answered using computer vision methods, as well as a good understanding of several techniques and how to apply them in their own research.

Keywords

Images as data; Automated image analysis; Computer vision; Computational methods; Machine learning

Course Prerequisites

- basic programming skills/experience in Python
- basic machine learning knowledge (e.g. distinction between supervised and unsupervised learning, familiarity with the training process in machine learning – such as train/test/validation split, cross-validation, etc. – although these concepts will be reviewed in more detail during the course)
- a Google account: we will use Google Colab in the course tutorials.

Target Group

Participants will find the course useful if:

- they are PhD students, early career scholars, industry professionals, or generally interested in using computational methods to automatically analyze large quantities of video/image data.

Course and Learning Objectives

By the end of the course participants will:

- have a good overview of the existing images-as-data literature in the social sciences
- have a good understanding of key deep learning concepts relevant for the implementation of computer vision methods
- have a good understanding of several computer vision techniques (object and face detection/recognition, image classification, facial trait analysis, etc.)
- have a good understanding of the many options and techniques available to store and compute visual data
- be able to implement different computer vision techniques in Python
- be able to use/adapt different computer vision techniques for their own research projects

Organizational Structure of the Course

The course will be organized around three different types of sessions:

- **Lectures** in which the instructors will present relevant literature, theory, concepts, and methods; and discuss them with the students.
- **Tutorials** in which the instructors will provide, run, and discuss, sample code designed to implement different computer vision techniques. Students will also run the code on their own and can ask as many clarifying questions as needed.
- **Consulting sessions** in which students will work on implementing the learned techniques on new data and projects. The instructors will provide students with new sample data, and will help them adapt the sample code from the tutorials so that it works for the new data. During these sessions, students can also bring their own data and ask questions about how to adapt the sample code for their own project, or what additional computer vision methods can help them answer their substantive questions of interest.

The **lectures** will take place in the morning of the first three days (Monday-Wednesdays). **Tutorials** will take place in the afternoon of the first three days, as well as at different times during the last two days. The **consulting sessions** will also take place at different times during the last two days.

Software and Hardware Requirements

Participants should bring their own laptops for use in the course.

The course will use Google Colab, so participants need a Google account. There is no need to install Python locally.

Recommended Literature to Look at in Advance

Webb Williams, N., Casas, A., & Wilkerson, J. (2020). *Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification.* Cambridge: Cambridge University Press. doi:10.1017/9781108860741

Torres, M., & Cantú, F. (2022). Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data. *Political Analysis, 30*(1), 113-131. doi:10.1017/pan.2021.9

Cantú, F. (2019). The Fingerprints of Fraud: Evidence from Mexico's 1988 Presidential Election. *American Political Science Review, 113*(3), 710-726. doi:10.1017/S0003055419000285

Dietrich, B. (2021). Using Motion Detection to Measure Social Polarization in the U.S. House of Representatives. *Political Analysis, 29*(2), 250-259. doi:10.1017/pan.2020.25

Steinert-Threlkeld, Z., Chan, A. & Joo, J. (2022). How State and Protester Violence Affect Protest Dynamics. The *Journal of Politics, 84*(2), 798-813. doi:10.1086/715600

Jürgens, P., Meltzer, C.E., & Scharkow, M. (2022). Age and Gender Representation on German TV: A Longitudinal Computational Analysis. *Computational Communication Research, 4*(1). doi:10.5117/CCR2022.1.005.JURG

Day-to-day Schedule and Literature

Day 1:

9:00-10:00: Introduction

- Introductions: instructors and participants
- Overview of the workshop: motivation, goals, structure, schedule, etc.

10:00-11:00: Lecture 1. Introduction to Images as Data in Social Science Research

- What can we do with images? Why automated image analysis? Overview of existing lines of research in the social sciences.

11:00-11:15: Break

11:15-12:00: Lecture 2. Introduction to Neural Nets and Computer Vision

- An easy-to-follow introduction to *Deep Learning* and *Neural Networks*
- Convolutional Neural Networks for large-N image analysis

12:00-13:00: Lunch Break

13:00-13:30: Tutorial 0. Technical Setup and Python Refresher

- Setting up and familiarizing with Google Colab
- A bit of python refresher

13:30-14:30: Tutorial 1. Image Processing

- Downloading and loading images
- Cropping, resizing, rotating
- From videos to images/frames
- Extracting basic/simple image features

14:30-14:45: Break

14:45-16:00: Lecture 3. Supervised Image Classification

- Object detection v. recognition
- How supervised learning works
- Existing benchmark datasets: cifar, minst, imagenet, coco, etc.
- Zero-shot classification with pre-trained models
- Fine-tuning pre-trained models

Day 2:

9:00-9:30: Catching-up Moment
- Any outstanding questions about what we did yesterday?

9:30-10:30: Lecture 4. Unsupervised Image Classification
- Using pre-trained models to represent images
- Clustering images based on embedding representations
- Variety of embeddings and clustering alternatives
- Promises and Pitfalls: validation, what can it be useful for, etc.

10:30-10:45: Break

10:45-12:00: Tutorial 2. Supervised Image Classification
- Loading and implementing pre-trained models
- Fine-tuning pre-trained models
- Nailing down the pipeline (tuning model, re-training, checking accuracy, implementing new model on new images, exporting output, saving model, etc.)

12:00-13:00: Lunch Break

13:00-14:00: Tutorial 3. Unsupervised Image Classification
- Using pre-trained models to represent images
- Clustering images based on embedding representations

14:00-14:15: Break

14:15-15:30: Lecture 5. Multimodal Modeling
- What to do when we want to model different data modalities together (text, images, etc.)
- We'll go over different approaches and best practices

15:30-16:00: Presentation/discussion of the *Data Challenge* in days 3 & 4

Day 3:

9:00-9:15: Catching-up Moment
- Any outstanding questions about what we did yesterday?

9:15-9:30: Preparation for Data Challenge
- 30 sec. elevator pitches from participants about the research questions/fields they are interested in (either using the existing data or their own data)
- During the day/the breaks/after class, start thinking about group formation

9:30-10:45: Lecture 6. Ethics and Research Practices in Computer Vision Research

- Biases in Computer Vision
- Data storage and privacy
- Reproducibility and transparency

10:45-11:00: Break

11:00-12:00: Tutorial 4. Multimodal Modeling
- Using pre-trained text and image models to represent text and images
- Using joint embeddings and other approaches for multimodal modeling

12:00-13:00: Lunch Break

13:00-14:00: Literature Discussion
- We go over some illustrative examples from the literature on images-as-data to see how the methods covered in the course are actually used to uncover relevant research questions
    - Cantu (2019) APSR *The Fingerprints of Fraud*
    - Dietrich (2021) PA *Using Motion Detection to Measure Social Polarization in the U.S. House of Representatives*
    - Steinert-Threlkeld et al. (2022) *How State and Protester Violence Affect Protest Dynamics*
    - Jürgens et al. (2022) CCR *Age and Gender Representation on German TV*

14:00-14:15: Break

14:15:16:00: Tutorial 5. Face Detection, Recognition, and Analysis

Day 4:
During Days 4 and 5, participants can choose between **2 routes**. The goal of both routes is to put into practice, with real data, what participants have learned in the previous days, of course with the help of the instructors.

(1) Data challenge route: Participants work in groups to answer a relevant theoretical question using computer vision techniques with one of the datasets provided by the instructors.
(2) Individual project route: Participants who already have a defined project and dataset in which they want to use computer vision techniques work on their project/dataset, applying what they have learned in the course (plus they can ask questions about potential additional computer vision methods not covered during the workshop that may be needed for their project)

In the afternoon of Day 5, participants will briefly present the results of their analysis during these two days.

9:00-10:00: Intro and Making up the Groups
- We introduce in more detail the datasets available for the data challenges
- Data challenge route:
    - Forming the groups
- Individual project route:
    - Roundtable where each participant presents the question and dataset

10:00-10:15: Break

10:15-11:00:
- Work Group 1. Question, Research Design, and Methods
    - Participants come up with a research question to address using one of the datasets
    - Participants elaborate on the research design and computer vision techniques to use

- Individual Consult 1. Method Discussion
    - Roundtable discussing the most suitable computer vision methods for each project

11:00-12:00: Participants work on their projects (instructors are present to answer questions and help out)

12:00-13:00: Lunch Break

13:00-14:15: Participants work on their projects (instructors are present to answer questions and help out)

14:15-14:30: Break

14:30-16:00:

- Work Group 2. Project Check-In.
    - Roundtable to discuss the progress and solutions to problems that have potentially emerged
- Individual Consult 2. Project Check-In
    - Roundtable to discuss the progress and solutions to problems that have potentially emerged

Day 5:

9:00-10:00:

- Work Group 3. Project Check-In and Discuss Presentation.
    - Roundtable to discuss the progress and solutions to problems that have potentially emerged
    - Briefly discuss what they will present in the afternoon
- Individual Consult 3. Project Check-In and Discuss Presentation
    - Roundtable to discuss the progress and solutions to problems that have potentially emerged
    - Briefly discuss what they will present in the afternoon

10:00-10:15: Break

10:15-12:00: Participants work on preparing the presentation for the afternoon (instructors are present to answer questions and help out)

12:00-13:00: Lunch Break

13:00-15:30: Project Presentations and Q&A/Discussion/Feedback

15:30-16:00: Closing

Additional Recommended Literature

Zhang, H., & Peng, Y. (2022). Image Clustering: An Unsupervised Approach to Categorize Visual Data in Social Science Research. *Sociological Methods & Research*. doi:10.1177/00491241221082603

Peng, Y. (2018). Same Candidates, Different Faces: Uncovering Media Bias in Visual Portrayals of Presidential Candidates with Computer Vision. *Journal of Communication, 68*(5):920-941. doi:10.1093/joc/jqy041