GESIS Leibniz Institute for the Social Sciences

gesis training

GESIS Fall Seminar in Computational Social Science 2024
"Automated Image and Video Analysis with Python"

Lecturers:	Andreu Casas	Felicia Loecherbach
Affiliation:	Royal Holloway University of London	University of Amsterdam
Email:	andreu.casas@rhul.ac.uk	f.loecherbach@uva.nl

Date: September 23-27, 2024
Time: 09:00-16:00

About the Lecturers

Andreu Casas is an Assistant Professor in the Department of Politics, International Relations, and Philosophy at Royal Holloway University of London, and 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, and Cambridge University Press*, among others.

Felicia Loecherbach is an Assistant Professor in the Department of Communication at the University of Amsterdam, and Faculty Affiliate at the Center for Social Media and Politics 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, Communication Methods and Measures, 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 based on image and video data. We will combine theoretical sessions where we'll discuss research using computer vision methods for the study of politics, communication, 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, multimodal models, visual language models, 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, and multimodal modeling. In addition, we'll also have a session on cloud computing, providing students with an overview of the options available to them if they need to train and deploy computer vision models on large amounts of data, as well as concrete examples on how to use some particular cloud computing services.

Students with basic programming skills/experience in Python and some machine learning background will get the most out of the course. In the cloud computing session, we will also use some bash (terminal coding), but it will be very minimal and no prior knowledge is required. 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.

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 be able to discuss their own projects with the instructors, who will help them think through the different methods/resources needed and how to adapt the material/skills learned throughout the workshop for their own project.

Keywords

Images as data; Automated image analysis; Computer vision; Computational methods; Machine learning; Multimodal modeling; Cloud computing

Target Group

You will find the course useful if:

- you are a PhD student, early career scholar, industry professional, or generally interested in using computational methods to automatically analyze large quantities of video/image (or multimodal: e.g. text + image) data.

Course and Learning Objectives

By the end of the course you 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
- be able to use/adapt different multimodal modelling approaches

Course Prerequisites

- basic programming skills/experience in Python (e.g. data loading, pandas data frames, loops and basic data operations)
- 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.

Software Requirements

The course will use Google Colab, so you need a Google account. There is no need to install Python or any Python packages 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 course: 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

13:30-15:00: Tutorial 1. Image Processing

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

15:00-16:00: Consulting Session 1

- Sign up for a one-on-one consultation with the course instructors. You can ask questions about the material covered so far and/or discuss a project of yours with the instructors.
- 20 min time slots (adjustable depending on the number of students)

DAY 2:

9:00-9:15: Catching-up Moment

- Any outstanding questions about what we did yesterday?

9:15-10:30: 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

10:30-10:45: Break

10:45-12:00: 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.

12:00-13:00: Lunch Break

13:00-14:00: 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.)

14:00-14:15: Break

14:15-15:15: Tutorial 3. Unsupervised Image Classification

- Using pre-trained models to represent images
- Clustering images based on embedding representations

15:15-16:00: Consulting Session 2

- Sign up for a one-on-one consultation with the course instructors. You can ask questions about the material covered so far, and/or discuss a project of yours with the instructors.
- 20 min time slots (adjustable depending on the number of students)

DAY 3:

9:00-9:15: Catching-up Moment

- Any outstanding questions about what we did yesterday?

9:15-10:30: 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

10:30-10:45: Break

10:45-12:00: Tutorial 4. Multimodal Modeling

- Using pre-trained text and image models to represent text and images
- Using joint embeddings and other approaches (eg. Visual Language Models) 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*
  - (some paper on multimodal modeling to be determined)

14:00-14:15: Break

14:15-15:30: Lecture 6. Ethics and Research Practices in Computer Vision Research

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

15:40-16:00: Consulting Session 3

- Sign up for a one-on-one consultation with the course instructors. You can ask questions about the material covered so far, and/or discuss a project of yours with the instructors.
- 20 min time slots (adjustable depending on the number of students)

DAY 4:

9:00-9:15: Catching-up Moment

- Any outstanding questions about what we did yesterday?

9:15-10:30: Tutorial 5. Face Detection, Recognition, and Analysis

10:30-10:45: Break

10:45-12:00: Lecture 7: Cloud Computing

- Why/when do we need cloud computing?
- What options are available to us?
- What are the pros and cons of each option?

12:00-13:00 Lunch Break

13:00-14:45 Tutorial 6: Cloud Computing

- Exploring some cloud computing options in practice
- Setting up a VM (w. GPUs) in a commercial environment
- Connecting to the VM and installing key dependencies
- Moving files/data in/out the VM
- Running Jupyter notebooks in the VM
- Running code in the VM
- Writing log files for keeping track and debugging

14:45-15:00 Break

15:00-16:00 Consulting Session 4
- Sign up for a one-on-one consultation with the course instructors. You can ask questions about the material covered so far, and/or discuss a project of yours with the instructors.
- 20 min time slots (adjustable depending on the number of students)

DAY 5

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

9:15-10:30: Tutorial 7: Full Pipeline
- A tutorial where we'll put everything together
- We'll go from collecting multimodal data, to processing it, training models for identifying quantities of interest, deploying the code in a VM, and then using the data for statistical analysis and hypothesis testing

10:30-10:45 Break

10:45-12:00 Consulting Session 5
- Sign up for a one-on-one consultation with the course instructors. You can ask questions about the material covered so far, and/or discuss a project of yours with the instructors.
- 20 min time slots (adjustable depending on the number of students)

12:00-13:00 Lunch Break

13:00-15:30 Student Presentations
- Students present a project or a project idea where they are using (or want to use) visual/multimodal data
- The other students and the instructors provide them feedback on any part of the project, from framing, research design, to particular methods or alternative data sources

15:30-16:00: Closing remarks and feedback moment

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