GESIS Leibniz Institute for the Social Sciences

gesis training

GESIS Fall Seminar in Computational Social Science 2025

"Computer Vision for Image and Video Analysis with Python"

Lecturer: Andreu Casas
Affiliation: Royal Holloway University of London
Email: andreu.casas@rhul.ac.uk

Date: September 15-19, 2025
Time: 09:00-16:00 (CEST / UTC+2)

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, Journal of Politics, Annual Review of Political Science, and Cambridge University Press*, among others.

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 and videos 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 and videos as data.

In this course, we will 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 will discuss research using computer vision methods for the study of politics, communication science, etc., with sessions where we will 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 will 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, multimodal modeling, and Visual Language Models. In addition, we will also have a session on cloud computing, providing participants with an overview of the options available to them if they need to train and deploy computer vision models for large amounts of data, as well as concrete examples of how to use some particular cloud computing services.

Participants 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 will also take the time to briefly review some key machine learning concepts necessary to implement the computer vision methods taught in this course, and participants will be provided with clear and easy-to-follow sample code for each of the practical tutorials. By the end of the course, participants 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 lecturer will present relevant literature, theory, concepts, and methods; and discuss them with the participants.
- **Tutorials** in which the lecturer will provide, run, and discuss sample code designed to implement different computer vision techniques. Participants will also run the code on their own and can ask as many clarifying questions as needed.
- **Consulting sessions** in which participants will work on implementing the learned techniques on their own, with the support of the lecturer. During these sessions, participants can bring their own data (or work on sample data provided by the lecturer) 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.

Keywords

Images as Data; Automated Image Analysis; Computer Vision; 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 modeling 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.

For those who would like a primer or refresher in Python, we recommend taking the online workshop "Introduction to Python" (25-28 August) and/or the online blended learning course "Introduction to Computational Social Science with Python" (01-05 September)

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 or any Python package locally.

Course Contents

- Image and video processing techniques
- Supervised visual classification (incl. object detection and recognition)
- Unsupervised visual classification
- Facial analysis
- Multimodal classification (inc. Visual Language Models)
- Cloud computing

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), 173-207. doi:10.5117/CCR2022.1.005.JURG

Day-to-day Schedule and Literature

DAY 1:

- **9:00-10:00:** Introduction
  - Introductions: lecturer 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:** Tutorial 0. Technical Set-up and Python Refresher
  - Setting up and familiarizing with Google Colab
  - A bit of Python refresher

- **12:00-13:00:** Lunch Break

- **13:00-14:30:** 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

- **14:40-14:45:** Break

- **14:45-16:00:** Tutorial 1. Image Processing
    - Downloading and loading images
    - Cropping, resizing, rotating
    - From videos to images/frames
    - Extracting basic/simple image features

DAY 2:

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

- **9:30-10:30:** 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:** 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

- **13:00-14:15:** 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.

- **14:15-14:30:** Break

- **14:30-15:30:** Tutorial 3. Unsupervised Image Classification
    - Using pre-trained models to represent images
    - Clustering images based on embedding representations

- **15:30-16:00:** Presentation/discussion of the *Data Challenge* in days 3, 4 and 5

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 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

- **10:45-11:00:** 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 (eg. Visual Language Models) for multimodal modeling

- **12:00-13:00:** Lunch Break

- **13:00-14:00:** Tutorial 5. Face Detection, Recognition, and Analysis
  - Detection and recognition of target faces of interest
  - Predicting age, gender, ethnicity, expressed emotions from faces

- **14:00-14:15:** Break

- **14:15:16:00:** Lecture 6 / Tutorial 6: Cloud Computing
  - Why/when do we need cloud computing?
  - What options are available to us?
  - What are the pros and cons of each option?
  - 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

DAY 4:

During Days 4 and 5, participants can choose **one of 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 lecturer.

(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 lecturer.

(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 2 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 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 (lecturer is present to answer questions and help out)

- **12:00-13:00:** Lunch Break

- **13:00-14:15:** Participants work on their projects (lecturer is 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 (lecturer is 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, 53*(3), 1534-1587. 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