

# Multimodal information extraction for automatic inventory of books

**Supervisors:** Ernest Valveny (UAB)

**e-mail:** [ernest@cvc.uab.es](mailto:ernest@cvc.uab.es)

## Description

Surveys in social science have repeatedly used the question on the number, type and title of books at home to characterize dimensions such as people's cultural and economic capital, or to explain academic performance. Manual collection of this information through surveys and questionnaires represents a huge effort and burden for the respondents.

In this scenario, automatic collection of all this information through the analysis of pictures of the books that someone has at home arises as an appealing alternative. Through the segmentation and recognition of such images, it would be possible to count the number of books and detect and recognize the title or the author of the books.

State-of-the-art detection methods based on deep learning present certain problems to identify some of the books (missing some of them), especially when they are in horizontal position or have an object in front of them, and to assign the correct area for a book (sometimes, the algorithm assigns a single box for several books). Both types of errors are presented in the image below.



Thus, in this project we aim to investigate on the use of text detection and recognition methods combined with recent general purpose self-supervised segmentation algorithms (such as Segment Anything model) and multimodal representation models (such as CLIP) in order to obtain a more accurate segmentation and recognition of the books contained in the bookshelves.

For these purposes, we have nearly 700 photos of home libraries sent by respondents of an online survey. These images have already been manually classified, identifying the number of books, the type of storage (shelves, closets, tables, and other), and the presence of books in Spanish, other co-official languages in Spain, and other languages.

This master dissertation will be developed as part of the project “New opportunities to enhance or extend (mobile) web survey data and get better insights”, conducted by Dr. Melanie Revilla from the Institut Barcelona d’Estudis Internacionals (IBEI).