



HAL
open science

Deep-Learning Technology for Book Conservation Assessment in Libraries and Archives

Valérie Lee, Lahcen Yamoun, Michel Jordan, Camille Simon Chane, David
Picard, Julien Longhi

► **To cite this version:**

Valérie Lee, Lahcen Yamoun, Michel Jordan, Camille Simon Chane, David Picard, et al.. Deep-Learning Technology for Book Conservation Assessment in Libraries and Archives. Un patrimoine pour l'avenir, une science pour le patrimoine, Mar 2022, Paris, France. hal-03620294

HAL Id: hal-03620294

<https://hal.science/hal-03620294v1>

Submitted on 25 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Un patrimoine pour l'avenir
une science pour le patrimoine
Heritage for the Future,
Science for Heritage



**Sustainable management
of cultural heritage**

Deep-Learning Technology for Book Conservation Assessment in Libraries and Archives

Valérie Lee¹, Lahcen Yamoun², Michel Jordan³, Camille Simon Chane⁴, David Picard⁵, Julien Longhi⁶

¹ PhD Student and Head of the Book and Paper Conservation Program at the National Institute of Patrimony (INP), PhD program 628, CY Cergy Paris University, France.

² Engineer, Master 2 IISC, CY Cergy Paris Université, ENSEA, France.

³ Research Engineer, ETIS UMR 8051, CY Cergy Paris University, ENSEA, CNRS, France.

⁴ Associate Professor, ETIS UMR 8051, CY Cergy Paris University, ENSEA, CNRS, France .

⁵ Senior Research Scientist, Laboratoire LIGM Laboratory, Ecole des Ponts ParisTech, France.

⁶ Professor of Linguistics and Digital humanities, Cergy Paris University, AGORA, Digital Humanities Institut, France.

Keywords : Artificial intelligence, deep-learning, book conservation, archives, libraries.

Abstract : One of the primary goals of Libraries and Archives is the conservation of their collections in order to pass them on to future generations. The large number of books kept in storage makes this task particularly challenging. Artificial intelligence offers the possibility of processing large volumes of data in a short period of time, but this technology has not been used for book conservation. A team of artificial intelligence scientists and one conservator is now developing a tool to automatically assess the conservation state of a binding at any given time.

The research presented is part of a PhD thesis of the ED 628 of CY Cergy Paris University, in collaboration with the laboratory ETIS UMR 8051 (CY Cergy Paris University, ENSEA, CNRS). The subject of this research is the use of deep-learning technology to assess the conservation state of bindings in Libraries and Archives. To date, conservation assessments for books are done by hand using statistical methods (1). For the majority of Libraries and Archives, these studies are time consuming and costly. Therefore, they are only partially done despite the growing need for loans and digitization. The main goal of the research is to develop a tool that will warn curators and conservators in case evidence of deterioration that are cause for concern are found on a binding, within a visible area of a book when placed on a storage shelf. This corresponds to the spine and the hinges that are structurally decisive for handling. The tool should be easy to use and able to give an overview of conservation states on a large number of books. Nowadays, artificial intelligence makes it possible to automatically and accurately detect damages such as cracks and losses on pictures of bridges and buildings (2). The system uses deep-learning techniques based on convolutional neural network algorithms. These neural networks must be trained with image databases which feature damages that can often occur in order to recognize images and classify them. A similar system was used for our research and applied to binding deteriorations.

First, an image data set was created from the 11 560 registers of the Parliament of Paris kept at the French National Archives. Most bindings selected were in leather and showed clear deteriorations: split hinges, missing areas in the leather, missing headcap and headband. Registers that were in poor condition were kept in individual conservation boxes. Using a digital camera, 159 photographs were taken featuring a total of 822 bindings on storage shelves. The purpose of these photographs was for a conservator to create segmented images with the Photoshop® software to obtain 7 classes of binary masks corresponding to the different types of deteriorations. A convolutional neural network was trained to first detect bindings on shelves (Figure 1). The confidence score was 97.77 %. Call numbers were recognized by a different network and linked to the binding with a confidence score of 98,31 %. A third network was able to separate bindings in boxes from the rest of the collection with a confidence score of 80,25%. Finally, each deterioration was detected. This task was more challenging when the deterioration was small. Confidence scores varied from 76,78% to 26,56 %. The tool was able to analyze 100 bindings in 10 seconds (Figure 2).

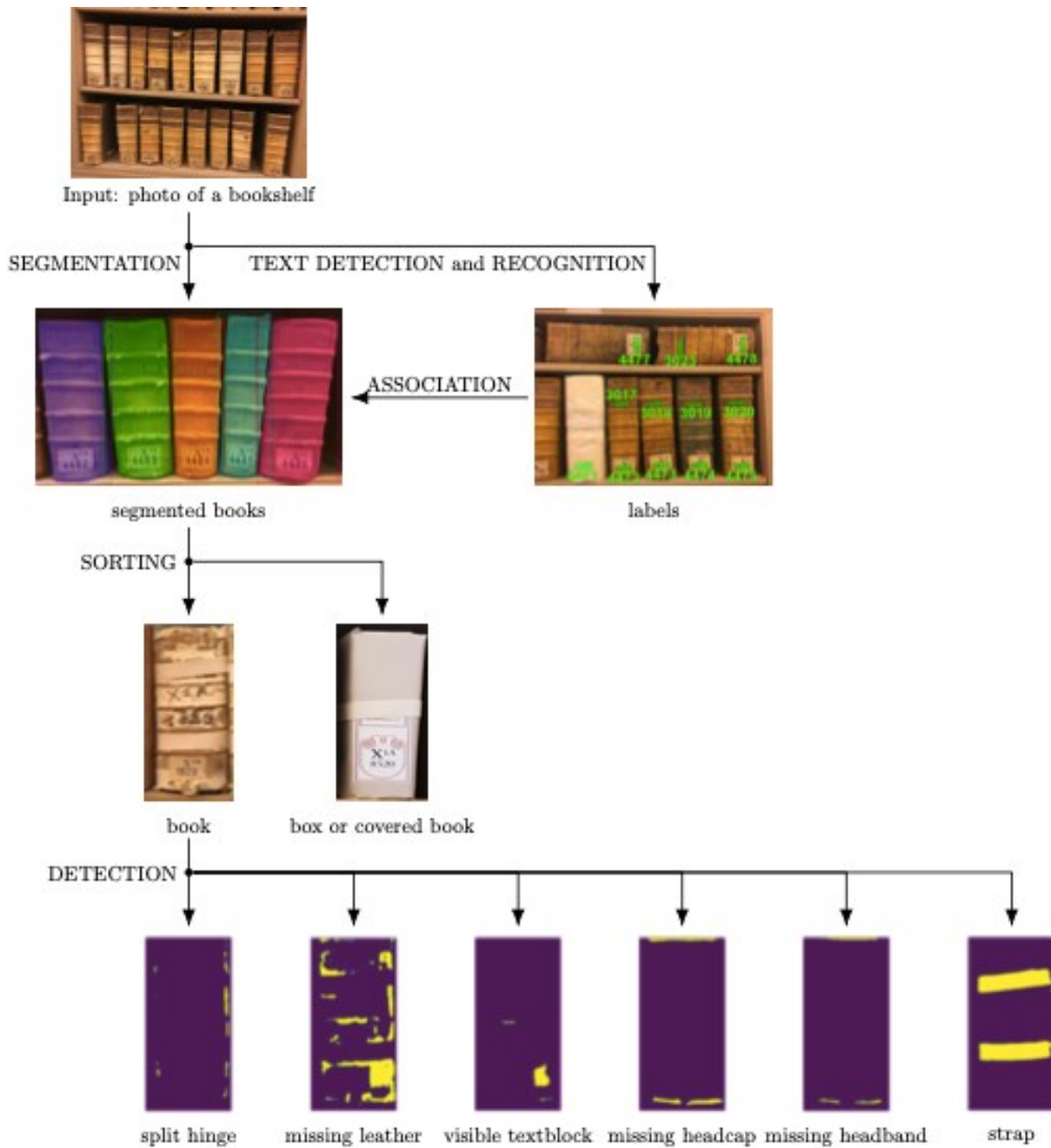


Figure 1: Detection of deteriorations on bindings. © Camille Simon Chane.



Figure 2 : Missing leather areas automatically detected with a U-Net pretrained model. © David Picard.

The research will continue for two more years with the Library of Quai Branly - Jacques Chirac Museum and the National Library and Archives of Quebec (BAnQ). The tool developed so far will be tested on new images of bindings dating from the end of the 19th century to the 21st century. The research will go beyond a simple implementation of existing algorithms. In the first instance, the images of the Quai Branly and BAnQ's collections are so different from those from French National Archives that a finetuning of the network will surely be necessary. Indeed, the task becomes more complex as we move from leather bindings to more recent works of various heights, thicknesses, and materials whose spines are often smooth. BAnQ's conservation team is also interested in counting and recognizing bindings that are particularly difficult to digitize. Indeed, 15,000 books will be digitized each year and BAnQ does not have enough staff members to count the number of books per shelf nor to know the number of damaged bindings that require restoration to be digitized. If the confidence score is over 95 % correct, the tool could be used every year for the collection's digitization plan.

We hope that this research will improve the limitations of the existing conservation assessment methods and help Libraries and Archives to make treatment decisions for their book collections. After its development, the tool is programmed to be available online so any Library or Archives can upload pictures of books on shelf and get a database pointing out which binding needs urgent care.



References :

1. Walker A. (2009) « Statistical tools for the evaluation of preservation need », in World library and information congress: 75th IFLA general conference and council, 23-27 August 2009, Milan, Italy. <https://www.ifla.org/past-wlic/2009/92-walker-en.pdf>
2. Kwon D., Yu J. (2019) « Automatic damage detection of stone cultural property based on deep learning algorithm », The International Archives of the photogrammetry, remote sensing and spatial information sciences, vol. XLII-2/W15, <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W15/639/2019/isprs-archives-XLII-2-W15-639-2019.pdf>

This article is published within the framework of the European symposium "Heritage for the Future, Science for Heritage: A European Adventure for Research and Innovation" which will be held on 15 and 16 March 2022 at the Musée du Louvre and the National Library of France (François Mitterrand). This symposium is organized with the support of the European Commission and in partnership with the French Ministry of Culture, the French National Centre for Scientific Research (CNRS) and the universities of Paris-Saclay and Cergy Paris. To feed the symposium programme, a Call for Papers was launched in June 2021 and addressed to researchers, doctoral students, and various actors who contribute to the production, valorisation, dissemination and applications of knowledge such as heritage professionals, political decision-makers, international organisations etc. As a result of this Call for Papers, some proposed communications have been selected for publication on the symposium website in the form of short articles.