Artificial intelligence to bring museum specimens to the masses

March 25 2022

(a) Naturalis Biodiversity Center Specimen   (b) Meise Botanic Garden Specimen
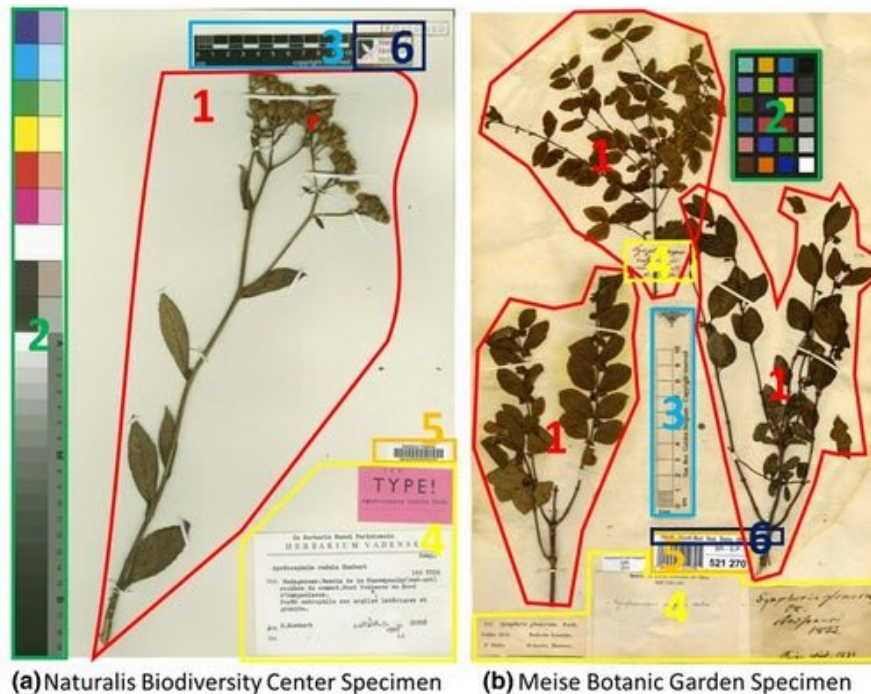
Credit: Cardiff University

Scientists are using cutting-edge artificial intelligence to help extract complex information from large collections of museum specimens.

A team from Cardiff University is using state-of-the-art techniques to automatically segment and capture information from museum specimens and perform important data quality improvement without the need of

human input.

They have been working with museums from across Europe, including the Natural History Museum, London, to refine and validate their new methods and contribute to the mammoth task of digitizing hundreds of millions of specimens.

With more than 3 billion biological and geological specimens curated in natural history museums around the world, the digitization of museum specimens, in which physical information from a particular specimen is transformed into a digital format, has become an increasingly important task for museums as they adapt to an increasingly digital world.

A treasure trove of digital information is invaluable for scientists trying to model the past, present and future of organisms and our planet, and could be key to tackling some of the biggest societal challenges our world faces today, from conserving biodiversity and tackling climate change to finding new ways to cope with emerging diseases like COVID-19.

The digitization process also helps to reduce the amount of manual handling of specimens, many of which are very delicate and prone to damage. Having suitable data and images available online can reduce the risk to the physical collection and protect specimens for future generations.

In a new paper published today in the journal Machine Vision and Applications, the team from Cardiff University has taken a step towards making this process cheaper and quicker.

"This new approach could transform our digitization workflows," said Laurence Livermore, Deputy Digital Program Manager at the Natural History Museum, London.

The team has created and tested a new method called image segmentation, that can easily and automatically locate and bound different visual regions on images as diverse as microscope slides or herbarium sheets with a high degree of accuracy.

Automatic segmentation can be used to focus the capturing of information from specific regions of a slide or sheet, such as one or more of the labels stuck on to the slide. It can also help to perform important quality control on the images to ensure that digital copies of specimens are as accurate as they can be.

"In the past, our digitization has been limited by the rate at which we can manually check, extract, and interpret data from our images. This new approach would allow us to scale up some of the slowest parts of our digitzation workflows and make crucial data more readily available to climate change and biodiversity researchers," continued Livermore.

The method has been trained and then tested on thousands of images of microscope slides and herbarium sheets from different natural history collections, demonstrating the adaptability and flexibility of the system.

Included in the images is key information about the microscope slide or herbarium sheet, such as the specimen itself, labels, barcodes, color charts, and institution names.

Typically, once an image has been captured it then needs to be checked for quality control purposes and the information from the labels recorded—a process that is currently done manually, which can take up a lot of time and resource.

Lead author of the new study Professor Paul Rosin, from Cardiff University's School of Computer Science and Informatics, said: "Previous attempts at image segmentation of microscope slides and

3/4

herbarium sheets have been limited to images from just a single collection.

"Our work has drawn on the multiple partners in our large European project to create a dataset containing examples from multiple institutions and shows how well our artificial intelligence methods can be trained to process images from a wide range of collections.

"We're confident that this method could help improve the workflows of staff working with natural history collections to drastically speed up the process of digitization in return for very little cost and resource."

Provided by Cardiff University