

sDiv working group meeting report

"sBombus 1"

Working group meeting report

Focal areas of discussion + main results/conclusions + open questions

Our discussions focused primarily on understanding data availability, the challenges of gathering existing data, the development of hardware to deploy computer vision algorithms for automated data collection, the challenges of a global monitoring plan for *Bombus*, and the need to document standard methodologies/workflows for AI model development. We concluded that existing data is not sufficient for a comprehensive model that identifies all *Bombus* species. We need additional data from parts of Asia and underrepresented regions such as South America. Hardware, such as camera traps will aid in global monitoring systems and their development is highly feasible but depends on high quality classification algorithms. Documenting workflows, data availability needs, and ways to improve existing algorithms without additional images (e.g., incorporating location and other data into classifications) will be helpful to move the field forward.

Content of presentations

Each of the meeting attendees presented their prior work related to the sBombus project. Brian Spiesman discussed work on the BeeMachine project, as well as some of the potential and limitations of computer vision technology. The limitations include a lack of image availability for many species and a lack of high-quality annotations.

Christina Grozinger discussed work on hardware systems for imaging bees in the lab and in the field. Grozinger's group is developing a camera system for high-speed imaging of pinned specimens in the lab as well as an automated camera trapping system for deployment in the field.

Tiffany Knight discussed her work on different pollinator projects resulting from resampling areas that were the focus of large surveys in the past. This type of resampling and comparison with historical datasets has been essential for our understanding of how land cover and climate change have affected pollinators and the plants with which they interact. Knight also discussed her work on developing computer vision projects for large scale and automated sampling.

Paul Williams discussed his decades of work studying bumble bees around the world, focusing especially on Asia. Asia is the center of diversity for bumble bees. However, there are many challenges to studying *Bombus* here including logistical difficulty accessing remote study sites, geopolitical issues, and the great difficulty of distinguishing many of the species in the region. In addition to the lack of image availability for advancing computer vision, we need a better understanding of the taxonomy of many species so that images can be properly labeled.

Dani Holthaus discussed her work on *Bombus* species distribution modeling, along with some of the challenges of working with presence-only data. Holthaus is

working on methods to account for some of the spatial biases in large datasets generated from citizen scientists, whose records are disproportionately from population centers.

Jun Yokoyama discussed his work on *Bombus* in Japan, especially related to their citizen science program, Hanamaru Maruhana. Yokoyama and colleagues, including one of our group members, Yukari Suzuki-Ohno, have developed a national program to monitor bumble bees. They have gathered thousands of observations and have been developing computer vision models based on these records.

Victoria Campopiano Robinson discussed her work on a citizen science program that she develops and runs in Argentina, called Vi Un Abejorro, or I saw a bee. This project is aimed at increasing awareness of the native *Bombus dahlbomii*, which is being threatened by the invasion of the *Bombus terrestris*. The project has gathered sightings from all over the country, which are being used to develop an understanding of ways to conserve *B. dahlbomii* and other native species in the region. There are two all-black bumble bee species in the region that are very difficult to separate that may benefit from computer vision for making accurate identification.

General research ideas, questions & directions discussed (incl. potential data to be used etc.)

One of our primary ideas is to find ways to improve the classification models using existing data. That is, how can additional types of data and newer modeling approaches improve *Bombus* classification? This could include incorporating other types of data into the classification process, such as location, date, or other information about the sighting like habitat, etc. We also discussed new types of classification models that leverage the recent explosion of large language models. These new approaches could use existing data that accompany most images (date, location, sighting notes).

Another research idea is to characterize the state of existing global *Bombus* image data. This would utilize existing data housed in online repositories, such as GBIF, and locating additional data sources from citizen science groups. This would be important for quantifying the limitations of data availability imposed on our ability to generate accurate and comprehensive classification models. For example, what are the geographic biases in data availability what species are over- and underrepresented, and why. How will the current state of *Bombus* taxonomy and geopolitical issues hamper developing high-quality databases?

Lastly, we discussed best practices for AI model development for *Bombus* (or other insects) identification. For example, what are effective data pipelines, how should images be collected, how many do we need, how do we make datasets available?

General structure of the week (break out groups, presentations, sessions with remote participants etc.)

Our days generally started off with a brief presentation from one or two of our meeting participants. We then followed with discussions of topics on our meeting agenda. On Wednesday, we had a field trip to a botanical garden to see different bees in person and to test the computer vision algorithm in the field. The final day or two was very productive in terms of working directly with data and organizing it for future analyses.

Next steps & upcoming deliverables

The next steps are to meet online with the full group for a summary of our progress and to discuss papers in progress based on our work in Leipzig in June. We have a paper in progress on incorporating location and date information into the classification model process. We will also begin with analyzing data for the image database review paper. These will lead into our second meeting in Leipzig.

General working atmosphere and feedback on sDiv support (What kind of support? How helpful was it?)

We received excellent support from the sDiv staff. They worked with us to handle individual travel issues when necessary and were always very accommodating. Our meeting room was more than adequate for our discussions. It provided a well-lit, comfortable, and nice environment to have our meetings. The lunches were always nice, and it was great to have drinks available throughout the day. It was a very pleasant working atmosphere because of the excellent facilities and presence of many other friendly researchers, students, and staff. Our hotel was very nice as well. The breakfast was excellent, and it was very nice to be able to eat as a group on two of the nights as part of the project. We ate together most nights anyway. The tour of Leipzig was excellent and very educational.