Development of FAIR image analysis workflows and training in Galaxy

Leonid Kostrykin [1], **Boris Depoortere** [2], **Diana Chiang** [1], **Sven Klumpe** [4], **Riccardo Massei** [5], **Matúš Kalaš** [6], **Anne Fouilloux** [7], and **Beatriz Serrano-Solano** [8]

1 Biomedical Computer Vision Group, Heidelberg University, BioQuant, Heidelberg, ELIXIR Germany **2** Vlaams Instituut voor Biotechnologie, Data Core, Gent; ELIXIR Belgium **3** Department of Computer Science, University of Freiburg, Freiburg, Germany **4** Max Planck Institute of Biochemistry, Martinsried, Germany **5** Helmholtz-Zentrum für Umweltforschung UFZ, Leipzig, Germany **6** ELIXIR Norway, and Department of Informatics, University of Bergen, Norway **7** Simula Research Laboratory, Oslo, Norway **8** Euro-BioImaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

Introduction

The 2024 edition of BioHackathon Europe took place in Barcelona, bringing together a group of image data analysis enthusiasts to collaborate on Project 17: *"Development of FAIR image analysis workflows and training in Galaxy"*.

Although image analysis tools are available within the Galaxy platform (Abueg et al., 2024), they remain underutilised. During the 2023 BioHackathon Europe, our efforts focused on enhancing the image analysis community in Galaxy by cataloguing and annotating tools and facilitating community discussions to establish naming conventions that promote standardisation. These initial efforts, detailed in the project outcomes (Kostrykin et al., 2024), laid the foundation for the ongoing expansion of Galaxy's image analysis capabilities.

Building on these achievements, this year's work aimed to **exploit and demonstrate the Galaxy platform's full potential to address the needs of the image analysis community**. This project involved developing FAIR (Findable, Accessible, Interoperable, and Reusable) image analysis workflows, creating tutorials for the Galaxy Training Network to provide documentation, and fostering broader adoption and facilitating the application of these workflows across scientific domains.

Project goals

This project aimed to enhance the accessibility, standardisation, and automation of image analysis workflows by leveraging Galaxy, a platform well-suited for diverse data analysis tasks. The primary goals included:

- Identifying popular **use cases** for image analysis to ensure the workflows address relevant and impactful needs.
- Selecting **relevant datasets** from public repositories to represent diverse image analysis tasks and challenges.
- Developing **Galaxy workflows** to perform various image analysis tasks, including segmentation, classification, and denoising.
- Depositing **workflows into WorkflowHub** (Goble et al., 2021)) to ensure their accessibility, findability, and reusability.
- Creating **FAIR tutorials** to guide users in employing these workflows effectively, promoting best practices and transparency.

These goals addressed significant challenges in the image analysis community, where standardisation and automation are often lacking, and manual interactions are immanent to many pipelines. By integrating Galaxy tools and providing state-of-the-art resources, this initiative aimed to **streamline workflows** and raise awareness of Galaxy's potential, ultimately **expanding its user base**.

This effort was further strengthened by the involvement of ELIXIR, Euro-BioImaging, and the Pangeo Community ensuring expert guidance and relevance from diverse scientific communities.

Long-term goals included fostering a robust community of Galaxy users exchanging knowledge and practices, accelerating image analysis research with more efficient and high-quality outcomes, and promoting cross-disciplinary collaboration.

This year, to enhance engagement with the online participants during the hackathon, we have established a project board, streamlining to streamline communication and collaboration.

Results

In this section, we discuss the achievements within the BioHackathon Europe 2024, aligning them with the project goals.

Tools

The Galaxy CoDex project (Project 11: Galaxy CoDex – Ensuring Galaxy community sustainability through resource aggregation & annotation) was instrumental in advancing the tool-related goals of this initiative by providing an up-to-date, annotated list of image analysis and processing tools. Accessible via the Galaxy CoDex repository, the list tailored for the image analysis community reflects the latest updates as of September 2024. It served as a reliable resource for identifying and evaluating tools.

Workflows

The project successfully updated and created several workflows. These include specialised workflows for detecting mitoflashes, registration, and segmentation tasks, with applications spanning bioimage and Earth Observation data analysis.

Efforts to improve existing workflows also advanced. The IWC (Intergalactic Workflow Commission) workflow for segmentation and cell counting was updated, incorporating a newer version of the "Perform histogram equalization" step.

Tutorials

Updated tutorial: "HeLa screen analysis"

The HeLa screen analysis tutorial was updated to reflect recent tool changes, including the merging of the "Split binary image using watershed transformation" tool with the Convert binary image to label map tool. With Galaxy now supporting direct visualisation of TIFF files, both this tutorial and the Introduction to Image Analysis Using Galaxy were substantially simplified by removing the now unnecessary conversion steps (TIFF to PNG). Additionally, tools and workflows used in these tutorials were updated to their latest versions to ensure accuracy and consistency.

New tutorial: Detection of mitoflashes

A new Galaxy tutorial was developed to detect and analyse mitoflashes, short bursts of fluorescence intensity from mitochondria caused by transient events like membrane depolarisation or

reactive oxygen species production. These events, characterised by temporal intensity peaks, are markers for understanding mitochondrial function and dynamics.

This tutorial is based on a workflow originally described in Wang et al. 2022, now implemented within the Galaxy platform using its integrated tools for reproducible analysis.

New tutorial: Overview of the Galaxy OMERO-suite

A new tutorial covering the different tools of the Galaxy OMERO-suite was developed. The Galaxy OMERO-suite is based on the Python packages omero-py and ezomero, and it allows interactively building pipelines to upload and fetch image data in OMERO using a Galaxy workflow. Images can automatically be enriched with metadata (i.e. key-value pairs, tags, raw data, regions of interest) and uploaded to an OMERO server. The tools give the user the ability to intuitively fetch images from the local server and perform image analysis.

Community

A dedicated community page on the Galaxy Community Hub was created to provide a centralised space for the imaging community in Galaxy. In addition to the community page, efforts have been made to connect with key platforms like the BioImage Archive (BIA) and BioImage Model Zoo (BMZ), further enhancing collaboration, accessibility, and standardisation across the image analysis community. These actions reflect a broader strategy to build a growing and engaged user base, fostering more significant interaction and knowledge-sharing.

Ongoing work

Tools

Update documentation for the AI model import tool

The tool Process image using a BioImage.IO model corresponds to the Galaxy integration for the BMZ that needed an update due to two issues that we identified:

- Issue 1: The tool currently fails when attempting to process 3D images, even when using compatible models. The issue is related to the expected input dimensions or model requirements, such as specific batch sizes or channel configurations. It would be helpful if the tool provided more guidance on input dimensions and automatically validated the required format for 3D models. This could include informing users of the necessary axes and allowing them to annotate image dimensions accordingly.
- Issue 2: The process for uploading models has been clarified. While the documentation mentions uploading models as ZIP files, users must ensure that they are downloading models in the Torchscript format from the BioImage Model Zoo. After downloading the model, users should unzip the file and upload the resulting PT file to Galaxy. The system automatically recognises this file without needing manual renaming. To enhance user experience, it would be beneficial to update the tool's help text to outline this process clearly.

By addressing both challenges, the tool's documentation and user instructions would improve to reduce confusion and enhance functionality. Including representative test cases in the tool repository would also help streamline troubleshooting and improve future updates.

Enhancements for the BioImage Archive download tool in Galaxy

The FTP Link for Bioimage Archive tool for Galaxy enables researchers to retrieve datasets directly from the BioImage Archive. However, improvements to the user interface were required. Proposed enhancements, tracked in bgruening/galaxytools#1541, include replacing the free-text "storage mode" field with a dropdown menu to choose from meaningful options, adding

guidance to clarify its purpose, and automating storage selection using the future BIA API. Similarly, a more precise help text for the "path of accession" field will reduce ambiguity and improve workflow efficiency. Proper indication of whether fields are optional or mandatory, coupled with validation of the inputs, and documentation updates will further align the tool with user needs, transforming it into a more robust and user-friendly solution for bioimage data retrieval.

Tutorials

Voronoi segmentation used across disciplines

The "Voronoi Segmentation Used Across Disciplines" tutorial is based on this workflow, which has been motivated by discussions in the BMCV/galaxy-image-analysis repository.

The tutorial is designed to be versatile, addressing the needs of different imaging modalities. Users will be able to choose their "path" through the tutorial based on their imaging modality. To this end, the tutorial will cover the use of images from the BioImage Archive, and aerial images or high-resolution satellite images (for instance this dataset).

The tutorial is still under development in this pull request.

Registration of MALDI and histology images

There was an existing workflow for the registration of mass spectrometry imaging (MALDI) and histology images: MALDI-Histology Registration Workflow. This interdisciplinary workflow enables the combined quantification of MALDI and histological microscopy images (Föll et al., GigaScience 2019). However, a newer version of the workflow might be available soon. During the hackathon, we identified that finding relevant (and open) images to use as inputs was the most critical part. The discussions around building this new tutorial can be followed in this issue.

Community

Galaxy as a BMZ community partner

During the Biohackathon 2024, we worked on integrating Galaxy as a community partner for BioImage Model Zoo. Partners of the BMZ, like ilastik and Fiji, comply with the model specification agreed upon by the community, enabling them to consume and deposit models. More details on the role of BMZ Community Partners can be found in the Community Partners Guide. The first step involved exploring this integration, and an issue was created in BMZ's repository.

BIA and Galaxy integration

Discussions with the BioImage Archive team have suggested the possibility of adding a button to their website that integrates with Galaxy's GALAXY_URL parameter. This feature would allow users to browse datasets on the BIA image data repository site, make their selections, and send the data directly to their Galaxy history for further processing.

Future plans

Tools

There is a tool for submitting images to the BioImage Archive: ascp2BIA. This tool should be used to deposit images to the BIA, preferably in the OME-Zarr format. To convert regular images to OME-Zarr, the bf2raw tool can be used. For instance, segmentation results obtained using Cellpose (see issue #15) could be deposited. Like the OMERO tool, this tool should

also be covered in a tutorial, either dedicated or as part of a larger tutorial. It should also emphasise the good practice of depositing images in OME-Zarr format to the BIA. Depositing images as OME-Zarr offers advantages such as embedded metadata, support for chunking, and the ability to store data in an S3-compatible bucket, which significantly improves access performance and reduces the need for local downloads. When publishing to the BIA, metadata such as license, authors, and descriptions should be provided to ensure the reproducibility of the research. In the long run, we plan to adapt the image analysis tools in Galaxy to consume OME-Zarr.

We currently offer an interactive QuPath tool within Galaxy. We plan to further integrate QuPath by wrapping it as a non-interactive Galaxy tool, enabling easier automation and workflow integration for more efficient and reproducible image analysis.

Tutorials

Cellpose tutorial

Cellpose is a popular tool for segmentation of cell nuclei and cytoplasm in fluorescence microscopy and histology images available in Galaxy. In this tutorial, we plan to explain the usage of Cellpose for image segmentation, as well as the quantification of the segmentation results to assess the segmentation performance. This requires a public dataset of suitable cellular images, accompanied by hand-labelled ground truth segmentation data. For a meaningful assessment of the segmentation performance, it is preferred that the dataset was not included in the training data of Cellpose. However, the candidate datasets that we are aware of and which fulfil this requirement, come without any license information: - IEEE ISBI Cell Tracking Challenge, GFP-GOWT1 mouse stem cells. - IEEE ISBI Cell Tracking Challenge, HeLa cells stably expressing H2b-GFP.

Tutorial using AI models from BMZ

Writing a tutorial for inference using BioImage.IO models in Galaxy would be very relevant to our community. There was a tutorial on that topic already written at BioHackathon Europe 2023 that would need to be converted into a tutorial for the Galaxy Training Network.

Tutorial for 2D spot detection and background removal

Bright spot detection in 2D imaging is a technique used to identify and localise regions of high intensity within an image. These spots often correspond to features of interest, such as fluorescent markers in biological imaging.

In order to properly detect bright spots in 2D images, background subtraction is sometimes used for preprocessing. This helps to remove non-uniform background illumination and generally improves the performance of the spot detection method. The tutorial will show how to use Galaxy to perform 2D spot detection and background subtraction for preprocessing of the image data. Moreover, it will show how to upload the results into OMERO with this candidate dataset.

Community

To prepare for upcoming events, we aim to develop reusable slides and demos that can be leveraged for future outreach and training activities.

Building on last year's efforts to annotate tools, we recognise the need for more detailed and specific annotations in EDAM Bioimaging, such as terms like "image processing," to improve the findability and reusability of our resources. These enhancements will ensure greater accessibility and utility for the community.

Acknowledgements

References

Abueg, L. A. L., Afgan, E., Allart, O., Awan, A. H., Bacon, W. A., Baker, D., Bassetti, M., Batut, B., Bernt, M., Blankenberg, D., Bombarely, A., Bretaudeau, A., Bromhead, C. J., Burke, M. L., Capon, P. K., Čech, M., Chavero-Díez, M., Chilton, J. M., Collins, T. J., . . . Zoabi, R. (2024). The galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*, *52*(W1), W83–W94. https://doi.org/10.1093/nar/gkae410 **[cito:citesAsAuthority] [cito:extends]**

Goble, C., Soiland-Reyes, S., Bacall, F., Owen, S., Williams, A., Eguinoa, I., Droesbeke, B., Leo, S., Pireddu, L., Rodríguez-Navas, L., Fernández, J. M., Capella-Gutierrez, S., Ménager, H., Grüning, B., Serrano-Solano, B., Ewels, P., & Coppens, F. (2021). *Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory*. https://doi.org/10.5281/ZENODO.4605654 **[cito:usesDataFrom] [cito:citesAsAuthority]**

Kostrykin, L., Woller, T., Kalaš, M., Özdemir, B., Andrade Buono, R., Sun, Y., Kumar, A., Goué, N., Grüning, B., Serrano-Solano, B., & Fouilloux, A. (2024). *Enhancing the image analysis community in galaxy*. https://doi.org/10.37044/osf.io/w8dsz **[cito:citesAsAuthority] [cito:extends]**