

Abstract

High-performance computers (HPC) are essential for bioimage analysis, however the barrier to entry can be high. This project aims to simplify access to bioimage analysis tools and deep learning models on local HPC clusters, enabling frictionless access to software and large computation.

Inspired by the Bioimage ANalysis Desktop (BAND) and ZeroCostDL4Mic, we developed lightweight bash scripts to deploy image analysis tools such as Fiji, QuPath, Ilastik, Cellpose, CellProfiler, and Napari in user-specified directories. Unlike containers, this solution allows users to save software changes, such as installed plugins, across sessions. We also created custom module files for deep learning packages like StarDist, SAM, and micro-SAM for easy environment loading. Finally, we developed Jupyter Notebooks for data preparation, model training, and benchmarking. These have been deployed on Harvard Medical School's (HMS) HPC cluster, Orchestra 2 (O2), which uses Open OnDemand (OOD) for an interactive interface.

While specific to HMS, this approach can be easily adapted to most HPC clusters. We aim to share our findings with the broader bioimage analysis community and discuss alternative or parallel approaches.

Link: <https://hms-iac.github.io/Bioimage-Analysis-on-O2/>

What problem does this project address?

The project aims to simplify access to bioimage analysis tools and deep learning models at Harvard Medical School's institutional cluster Orchestra 2 (O2) by reducing the complexities of installation.

What do we propose?

Harvard Medical School's (HMS) Orchestra 2 (O2) cluster, equipped with Open OnDemand (OOD), enables easy access to computing resources through a desktop interface or coding environments like Jupyter Notebook and Visual Studio Code.

- Interactive tools and GUI:** Executable bash scripts in shared folders allow users to install bioimage analysis tools, with a .desktop file created for easy, persistent access, enabling software to be launched like a local installation.
- Deep learning environments:** Pre-configured environments with necessary dependencies are stored in shared directories, with Linux modules allowing users to load and integrate deep learning models seamlessly.
- Jupyter notebooks:** Custom Jupyter Notebooks manage data loading, preprocessing, training, and testing, generating detailed reports on model performance for sharing or publication.

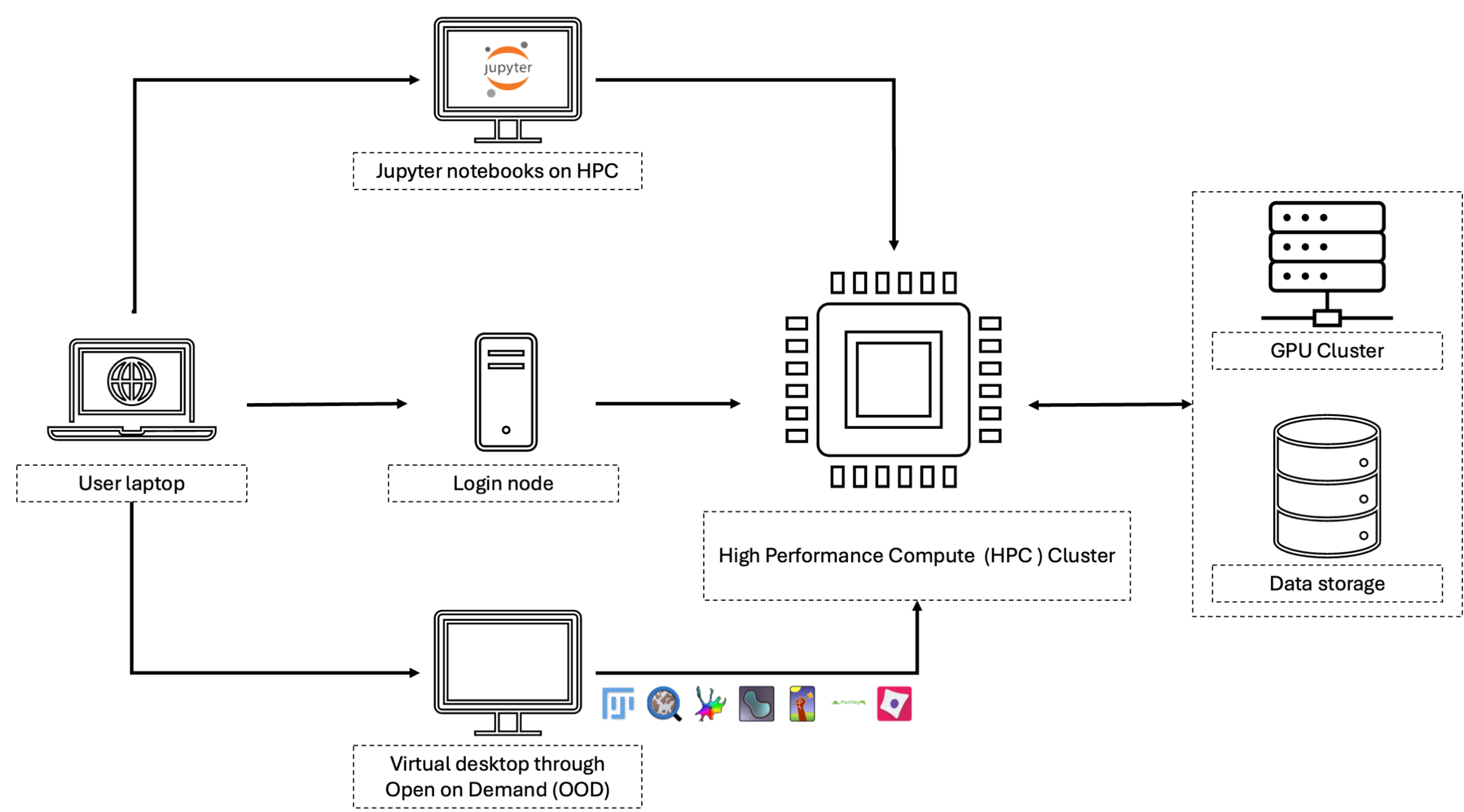


Figure 1. Diagram illustrating user access to bioimage analysis tools on HPC resources.

Advantages

- Direct access to data:** Users can easily access large datasets stored on the HPC cluster without the need for data transfers.
- Easy installation of software:** With pre-configured bash scripts, users can easily install essential bioimage analysis tools and deep learning software in their personal HPC accounts, eliminating the complexities often associated with manual installation and testing.
- Standardized model training and testing environments:** Custom-built environments ensure that all users have access to the same software packages and dependencies, ensuring reproducible workflow.
- Quick access to state-of-the-art deep learning models:** Ready-to-use deep learning modules, including popular models like StarDist and Segment Anything (SAM), allowing users to load and apply state-of-the-art models without needing to handle complex setup processes.
- Session persistence with installed plugins:** Unlike containerized solutions, this setup retains all user-installed plugins and configurations across different sessions, providing a consistent and uninterrupted working environment.

Support and Funding

Currently Installable Software



Figure 2. Overview of currently available bioimage analysis (BIA) tools. From left to right: Cellpose3[11] for cell segmentation, CellProfiler[10](4.3.7) for high-content screening, Fiji[7] for general image analysis, ilastik (1.4.0.post1)[3] for image segmentation and annotation, napari[5] (0.4.19) for interactive multi-dimensional image analysis, plantseg[12] for plant tissue segmentation, and QuPath (v0.5.1)[2] for whole slide image analysis.

Available deep learning packages:

- StarDist [8]
- Segment Anything [4]
- Micro-SAM [1]
- CellPose [11]
- Ilastik [3]

Available notebooks:

- U-Net [6]
- VGG16 [9]
- Segmentation evaluation

Motivation and Acknowledgement

- Bioimage Analysis Desktop (BAND)
- Biolmage Model Zoo
- Zero Cost Deep Learning for Microscopy

References

- Anwai Archit, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Sagnik Gupta, Andreas Dengel, Sheraz Ahmed, and Constantin Pape. Segment anything for microscopy. *bioRxiv*, pages 2023–08, 2023.
- Peter Bankhead, Maurice B Loughrey, José A Fernández, Yvonne Dombrowski, Darragh G McArt, Philip D Dunne, Stephen McQuaid, Ronan T Gray, Liam J Murray, Helen G Coleman, et al. QuPath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7, 2017.
- Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N Straehle, Bernhard X Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, et al. Ilastik: interactive machine learning for (bio) image analysis. *Nature methods*, 16(12):1226–1232, 2019.
- Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.
- How Does napari Work. napari: A multidimensional image viewer for python a guide to some of the key concepts of napari.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
- Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682, 2012.
- Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *Medical image computing and computer-assisted intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11*, pages 265–273. Springer, 2018.
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- David R Stirling, Madison J Swain-Bowden, Alice M Lucas, Anne E Carpenter, Beth A Cimini, and Allen Goodman. Cellprofiler 4: improvements in speed, utility and usability. *BMC bioinformatics*, 22:1–11, 2021.
- Carsen Stringer and Marius Pachitariu. Cellpose3: one-click image restoration for improved cellular segmentation. *bioRxiv*, pages 2024–02, 2024.
- Adrian Wolny, Lorenzo Cerrone, Athul Vijayan, Rachele Tofanelli, Amaya Vilches Barro, Marion Louveaux, Christian Wenzl, Sören Strauss, David Wilson-Sánchez, Rena Lymbouridou, et al. Accurate and versatile 3d segmentation of plant tissues at cellular resolution. *Elife*, 9:e57613, 2020.