BIAFLOWS:
A collaborative framework to benchmark bioimage analysis workflows

Ulysse Rubens° (ULiège), Romain Mormont° (ULiège), Volker Baecker (MRI, Biocampus Montpellier), Gino Michiels (HEPL/ULiège), Lassi Paavolainen (FIMM, HiLIFE, UHelsinki), Graeme Ball (University of Dundee), Devrim Ünay (IUE), Benjamin Pavie (VIB), Anatole Chessel (École Polytechnique), Leandro A. Scholz (Universidade Federal do Paraná), Martin Maška (Masaryk University), Renaud Hoyoux (Cytomine SCRL FS), Rémy Vandaele (ULiège), Stefan G. Stanciu (Politehnica Bucarest), Ofra Golani (Life Sciences Core Facilities, Weizmann Institute of Science, Israel), Natasa Sladoje (Uppsala University), Perrine Paul-Gilloteaux (Structure Fédérative de Recherche François Bonamy, Université de Nantes, CNRS, INSERM, Nantes, France), Raphaël Marée* (ULiège), Sébastien Tosi* (IRB Barcelona).

° These authors contributed equally to this work
* These authors jointly supervised this work (+ corresponding authors)

Abstract

Automated image analysis has become key to extract quantitative information from biological microscopy images, however the methods involved are now often so complex that they can no longer be unambiguously described using written protocols. We introduce BIAFLOWS, a web based framework to encapsulate, deploy, and benchmark automated bioimage analysis workflows (the software implementation of an image analysis method). BIAFLOWS helps fairly comparing image analysis workflows and reproducibly disseminating them, hence safeguarding research based on their results and promoting the highest quality standards in bioimage analysis.

Introduction

As life scientists collect microscopy images of increasing size and complexity [1], the use of computational methods has become inescapable in order to extract quantitative information from these datasets. Unfortunately, state of the art image analysis methods published in biomedical journals are often shared as poorly documented source code sometimes requiring expert configuration; and seldom packaged as user friendly modules for mainstream BioImage Analysis (BIA) platforms [2][3][4]. Additionally, test images are not consistently provided with the software and it can be difficult to identify what is the baseline for valid results, and the critical parameters to adjust in order to optimize the analysis. Overall, this does not only impair the reusability of the methods and impede reproducing published results [5], but it also makes it difficult to adapt these methods to process similar image datasets. To improve this situation, scientific datasets are now increasingly made publicly available through web-based software [6][7][8] and open data initiatives [9], but existing web platforms do not systematically offer advanced features such as the ability to remotely view multidimensional microscopy images, process the images online by remotely launching image analysis workflows, and compare the results of the methods against a ground-truth reference (a.k.a. benchmarking). Benchmarking is known to stimulate the continuous improvement of BIA methods and promote their wider diffusion [10]. For this reason, biomedical image analysis challenges are becoming more and more popular. But, all these competitions have limitations [11]: they each focus on a single bioimage analysis problem and they often rely on poorly reusable, ad-hoc, data access protocols and scripts to compute performance metrics. Both challenge organizers and participants are therefore duplicating efforts from one challenge to the next, whereas participants' workflows are rarely available in a sustainable and reusable fashion. Finally, the vast majority of annotated images released in these challenges come from medical imaging, not from biology: for instance, as of September 2019, only 15 out of 188 Grand-challenges datasets [12] were collected from fluorescence microscopy, arguably the most common imaging technique for research in biology. There is hence a critical lack of public annotated datasets illustrating BIA problems.

Results

Within the **N**etwork of **Eu**ropean **B**io**I**mage **A**nalyst**s** (NEUBIAS[1]), an important body of work focuses on channelling the efforts of bioimaging stakeholders to address these issues and ensure a better access to and assessment of existing bioimage analysis software. Together, we have envisioned and implemented BIAFLOWS (Fig. 1), *an open-source web platform to benchmark bioimage analysis workflows on publicly shared annotated multidimensional imaging data*. Whereas some emerging bioinformatics web platforms [13][14] simply rely on *Dockerized* environments and interactive notebooks to analyse scientific data accessed from a public database, BIAFLOWS offers a complete framework to 1) import multidimensional annotated images, 2) encapsulate and version BIA workflows (running as standalones or scripts for BIA platforms), 3) remotely visualize images and workflow results together and 4) automatically assess the performance of the workflows from widely accepted benchmark metrics. The workflows integrated to BIAFLOWS are versioned and can be run remotely to

1 NEUBIAS: Network of European BioImage Analysts, COST Action CA15124. www.neubias.org

process image datasets recapitulating important classes of BIA problems. All content, including images, workflows, object annotations and benchmark metrics, can be browsed and interactively explored through the web interface (Fig. 1). To complement the numerical results from the benchmark metrics, it is also possible to visualize the results of one or multiple workflows at once by synchronizing several image viewers (Fig. 2). This more qualitative assessment can reveal interesting features that are not necessarily captured by the benchmark metrics (Supplementary Methods sections 5 and 7).

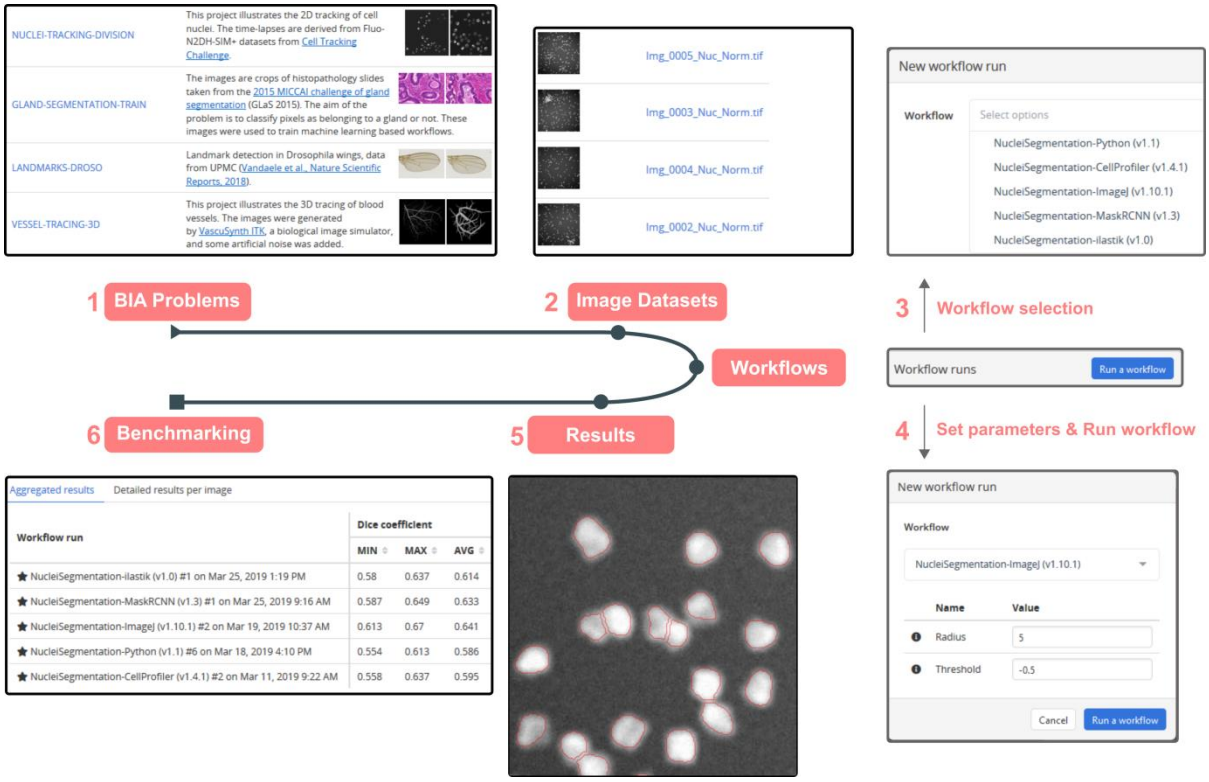


Figure 1. BIAFLOWS web interface*. A user selects a BIA problem (1) and browse image datasets illustrating this problem (2). The user selects a workflow (3) to process the images, and sets its parameters (4) before running it. Finally the results can be visualized overlaid on the original images (5), and benchmark metrics are reported as overall statistics for all the images or per image (6).*

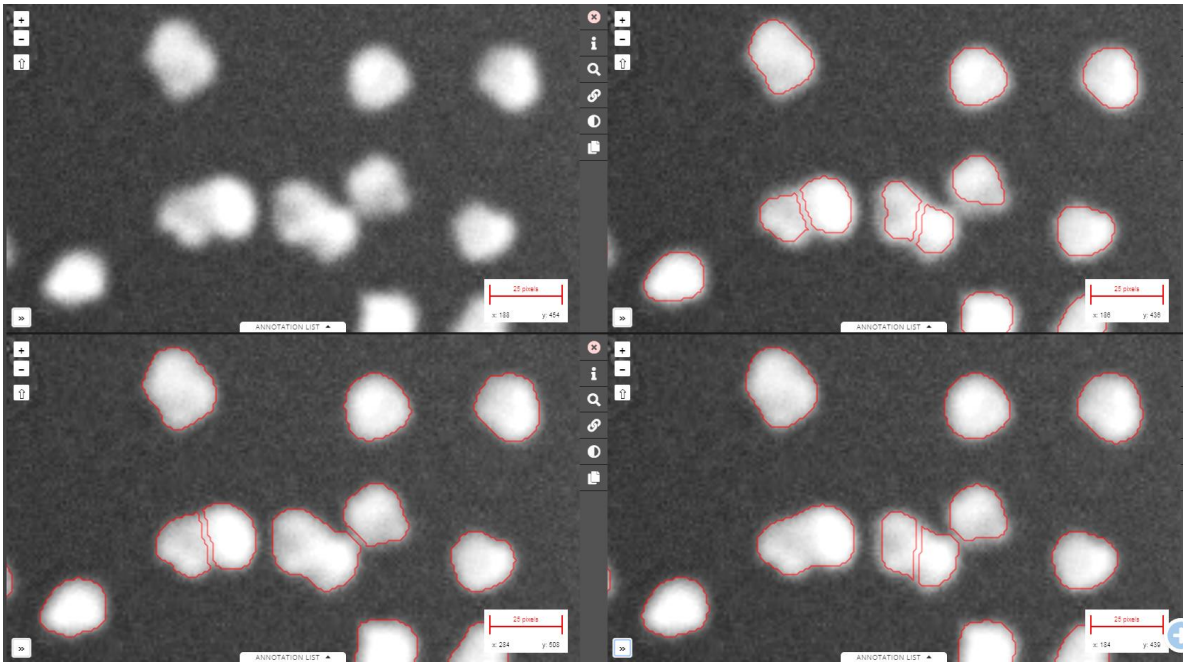


Figure 2. Synchronizing several viewers showing results from different workflows*. Results for 2D nuclei segmentation problem from custom ImageJ macro (upper right), custom CellProfiler pipeline (lower left) and custom Python script (lower right). The original image (zoomed in region) is also displayed without overlay (upper left).*

BIAFLOWS is open-source and extends Cytomine [15], a web platform developed for the collaborative annotation of high-resolution biomedical images (especially large 2D histology images). The critical features that were developed for BIAFLOWS to enable the benchmarking of bioimage analysis workflows on multidimensional microscopy images are subsequently described. First, a module has been implemented for the upload of (C, Z, T) images in OME-TIFF format through the web interface, as well as their remote visualization. Next, the core architecture of the platform has been completely re-designed to support the remote execution of bioimage analysis workflows by encapsulating them and their original software environment in *Docker images*, and describing their interface (input, output, parameters) by an extended version of Boutiques, a rich application description schema [16] (Supplementary Methods, section 4). BIAFLOWS is designed to monitor trusted user spaces hosting *Docker images* and automatically pull new or updated workflows (Fig. 3, DockerHub). In turn, *Docker images* are automatically built whenever a new release is triggered from their source code repositories (Fig. 3, GitHub). To ensure reproducibility, all *Docker images* are versioned and permanently accessible from the system. Moreover, they can be run on virtually any type of computational resources including high-performance computing and multiple server architectures. This is seamlessly achieved by converting the *Docker images* to a compatible format (Singularity, [17]) on the server and dispatching them over the network to the target computational resources using a workload manager (SLURM, [18], Fig. 3, additional computing servers). Finally, to enable benchmarking and inter-operability between all components, standard object annotation formats were specified for each class of BIA problems (Supplementary Methods, Section 5) and scripts to compute benchmark metrics were adapted from biomedical challenges [12] and scientific publications [20].

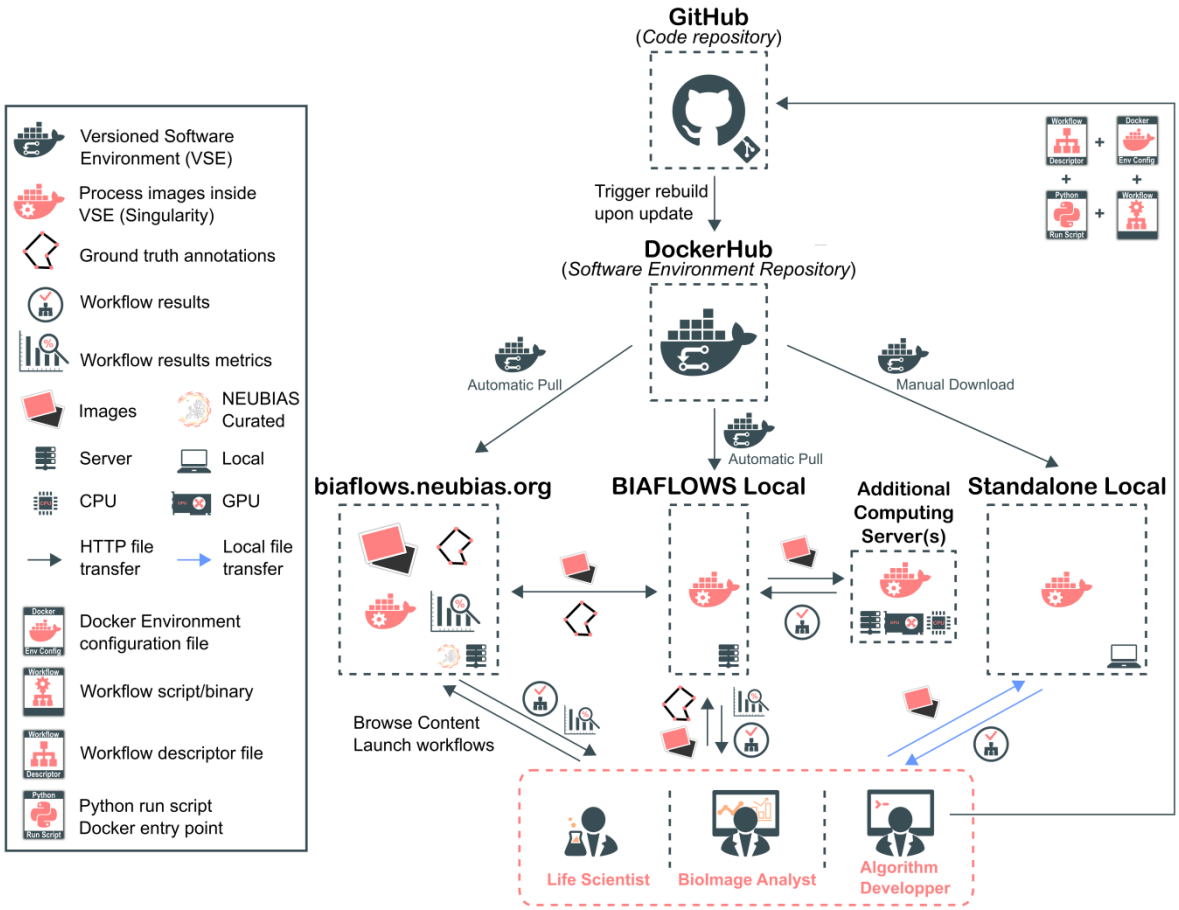


Figure 3. BIAFLOWS architecture and deployment scenarios*. Workflows are hosted in a trusted code repository (GitHub). Docker images encapsulate workflows and their exact execution environments to ensure reproducibility. They are automatically built and available through DockerHub upon new version releases of the workflows. BIAFLOWS monitors DockerHub and automatically pulls new or updated workflows; this works equally for the online instance or any locally installed BIAFLOWS instance configured to do so. Docker images can also be manually downloaded to process local images in a standalone fashion.*

A BIAFLOWS online instance managed by NEUBIAS is available at this URL: https://biaflows.neubias.org/ (Fig. 3, **biaflows.neubias.org**). A complete user guide and video tutorials are available from the platform, and all content can be viewed in read only mode from the guest account (user: guest, password: guest). This instance is populated with a substantial collection of annotated image datasets and associated workflows and can be browsed in read only mode from the guest account (user: guest, password: guest). The datasets were mainly imported from existing biomedical challenges, or created by synthetic image generators [19]. Currently 14 annotated datasets and 19 workflows are available and illustrate 8 classes of BIA problems (Supplementary Methods, Section 1, Table 1): object detection/counting, object segmentation, and pixel classification (Fig. 4); particle tracking, object tracking, filament tree tracing, filament network tracing, and landmark detection (Fig. 5). To showcase the versatility of the platform, the workflows available consist of a mixture of standalone software and scripts (ImageJ/Fiji, ICY, CellProfiler, Vaa3D, ilastik, Python scripts and Jupyter Notebooks). As a proof of concept, some of these workflows also leverage deep

learning Python libraries (Keras, PyTorch). In addition, workflows hosted in the system are referenced from Bioimage Informatics Search Engine (http://biii.eu/) to enhance their visibility (BISE is an online repository of bioimage analysis tools maintained by NEUBIAS). Interested developers can package their workflows (Supplementary Methods, Section 4) and make them available for benchmarking from BIAFLOWS online instance by sending a request to BIAFLOWS administrators[2]. BIAFLOWS can also be deployed on a local server and populated locally (both images and workflows) as a local image analysis solution (Fig. 3 BIAFLOWS Local, Supplementary Methods, Section 3). To simplify this process, migration tools were developed to transfer content between existing BIAFLOWS instances (Supplementary Methods, Section 6), and BIAFLOWS content can be accessed programmatically through a RESTful interface (see Supplementary Methods, Section 6) to ensure a complete data accessibility and inter-operability. Finally, for full flexibility, BIAFLOWS workflows can be manually downloaded to process local images independently of any BIAFLOWS server (Fig. 3, Standalone Local, Supplementary Methods, Section 6).

Discussion

On one hand, regardless of the platform it is deployed on (e.g. ImageJ), end users running a BIA workflow might have a hard time assessing the quality of the results, optimizing them by tweaking the parameters, and ensuring that this workflow is among the best achieving for their images. On the other hand, BIA workflow developers contributing to a biology research project would largely benefit from assessing the performance of the workflow they developed for the images of the project on similar public datasets. BIAFLOWS is a web-based platform enabling the testing and benchmarking of BIA workflows extracting annotations (e.g. biological objects) from raw images. This critical step is often the building blocks of more complex workflows extracting quantitative information from microscopy images. The online instance managed by NEUBIAS is available at https://biaflows.neubias.org and it is populated with a substantial collection of annotated bioimage datasets and associated workflows recapitulating common BIA problems. All workflows can be run and benchmarked from a simple web browser. BIAFLOWS is designed to expose the functional parameters of a BIA workflow, and to provide default *optimal* values (Fig. 1, step 4) for the images of a given problem. This responsibility is essentially left to the workflow developer. The parameters can however be freely adjusted by the user to assess their impact on the final results. It is easy to add new images, annotations and workflows to the platform and they can be migrated between BIAFLOWS instances. To increase the content currently accessible, calls for contribution will be continuously launched to gather more microscopy annotated images and encourage developers to package their methods so that they can be compared to other existing methods and be made readily available to the users. New problem classes will also be shortly supported by the platform, for instance the detection of blinking events in the context of super-resolution localization microscopy and landmark matching for image registration. BIAFLOWS addresses a number of critical requirements to foster open image analysis for life sciences: (i) providing and visualizing annotated images illustrating BIA problems faced within imaging based biology research projects, (ii) sharing versioned image analysis workflows in a reproducible way, (iii) exposing the critical functional parameters of these workflows and their optimal default values, (iv) computing

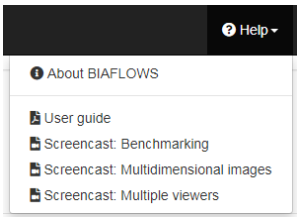
2 Send a request to biaflows@neubias.org

relevant performance metrics to compare the workflows, and (v) providing a standard way to store and visualize the results of the workflows. There is no limitation in using BIAFLOWS in other fields where image analysis is a critical step to extract scientific results, for instance material science and biomedical imaging. BIAFLOWS could also naturally be used as a platform to organize image analysis challenges. This would simplify the task of the challenge organizers by automating benchmarking and opening the results to the community, and it would increase the content of BIAFLOWS by integrated new annotated images and workflows. Another potential application is to trigger BIAFLOWS to span workflow parameter spaces so as to automate parameter optimization or assess the impact of their misadjustments, a critical feature from the usability standpoint. More generally, BIAFLOWS could be adopted as a central and federated platform to make scientific images available to the community and publish them together with the workflows that were used to extract annotations from these images. This is a milestone of open science not yet reached by any existing image sharing platform. This should accelerate scientific progress and help surpassing individualist practices where image datasets, algorithms, quantification results, and associated knowledge are stored and accessible only to a restricted circle.

Methods

Accessing BIAFLOWS online instance

BIAFLOWS online instance is available at https://biaflows.neubias.org and it can be browsed in read-only mode from the guest account (username: *guest* password: *guest*). An account with workflow execution right is also temporarily provided for the reviewers (username: reviewer password: reviewer13). The web platform is maintained by NEUBIAS (http://neubias.org). Video tutorials and a user guide illustrating how to navigate and use the platform are available from the **Help** section of the website.



A list of all the content currently available from the website is provided in Supplementary Methods (Section 1, Table 1) and some of the BIA problems currently available in the platform are illustrated in Fig. 4 and Fig. 5. The image datasets have been selected to recapitulate common BIA analysis tasks: spot detection (2D/3D), nuclei segmentation (2D/3D), nuclei tracking (2D), landmark detection (2D), filament tracing (3D), and tissue detection (2D) in whole-slide histology images. All image datasets are imported from previously organized challenges (DIADEM [21], ISBI Cell Tracking Challenge [22], ISBI Particle Tracking Challenge [23], Kaggle Data Science Bowl 2018 [24]), created from synthetic data generators (CytoPacq [25], TREES toolbox [26], Vascusynth [27], SIMCEP [28]), or contributed by NEUBIAS members [37]. To showcase the versatility of the platform, the image analysis workflows available to process these images are running on different BIA platforms: ImageJ macros [29], Icy protocols [30], CellProfiler pipelines [31], Vaa3D plugins

[32], ilastik pipelines [33], Python scripts leveraging Scikit-learn [34] for supervised learning algorithms, and Keras/PyTorch [35][36] for deep learning. These workflows were mostly contributed by members of NEUBIAS Workgroup 5, or imported from existing challenges.

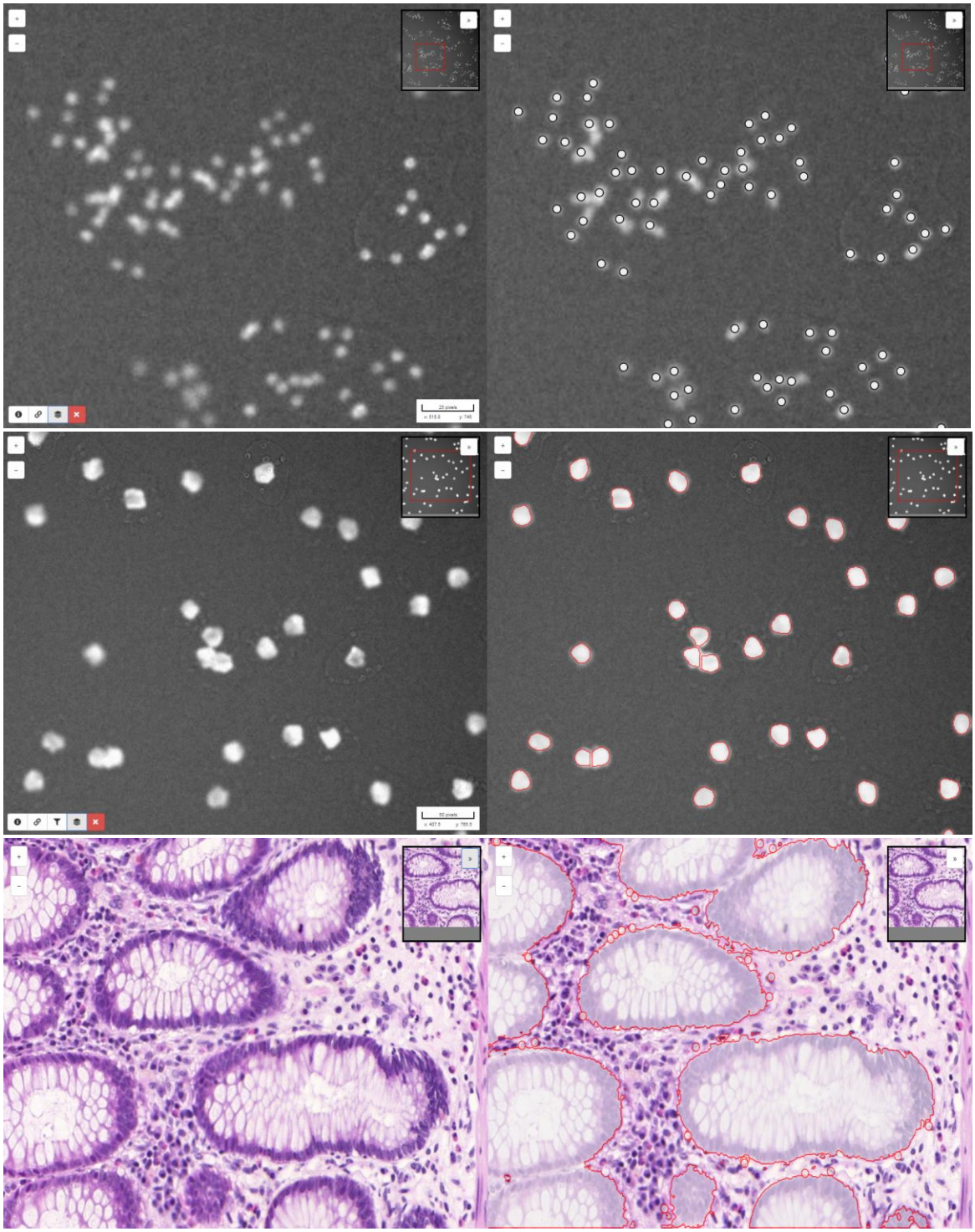


Figure 4 Some sample images illustrating BIA problems from BIAFLOWS online instance and results from associated workflows. Original image (left) and workflow results overlay (right). 1. Spot / object detection & counting, synthetic image displaying spots generated by SIMCEP [38]. 2. Object segmentation, synthetic image displaying nuclei generated by SIMCEP [38]. 3. Pixel classification, red areas circle pixels classified as gland, image from 2015 MICCAI challenge of gland segmentation [38].

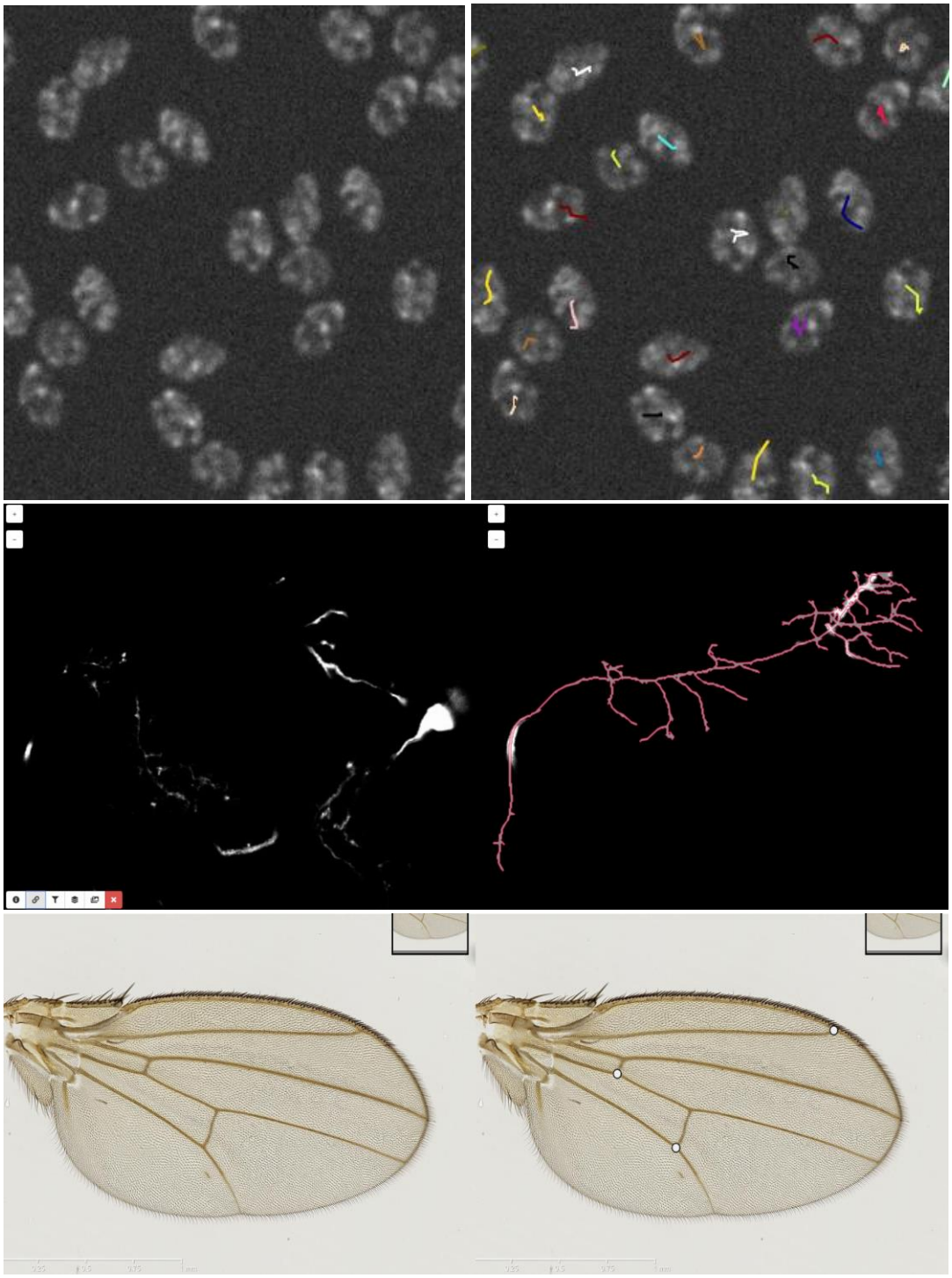


Figure 5 Some sample images illustrating BIA problems from BIAFLOWS online instance and results from associated workflows. Original image (left) and workflow results overlay (right). 1. Particle / object tracking, synthetic time-lapse displaying non-dividing nuclei generated by CytoPACQ [25], single frame + dragon tail tracks showing nuclei past positions. 2. Neuron tracing, Z-stack from DIADEM challenge [21], single slice + dilated Z-projection of traced skeleton (red). 3. Landmark detection, Drosophila wing, image from UPMC [37].

Acknowledgments

This project is funded by COST CA15124 (NEUBIAS). BIAFLOWS is developed by NEUBIAS (http://neubias.org) Workgroup 5 and it would not have been possible without the great support from NEUBIAS vibrant community of bioimage analysts, and the dedication of Julien Colombelli and Kota Miura in organizing this network. Local organizers of the hackathons who have fostered the development of BIAFLOWS are also greatly acknowledged: Chong Zhang, Gabriel Martins, Julia Fernandez, Peter Horvath, Bertrand Vernay, Aymeric Fouquier d'Herouel, Andreas Girod, Paula Sampaio. **UR** and **RV** were supported by ADRIC Pôle Mecatech Wallonia grant, and DeepSport Wallonia grant. **RaM** was supported by IDEES grant with the help of the Wallonia and the European Regional Development Fund (ERDF). **LP** was supported by Academy of Finland (grant 310552). **MM** was supported by the Czech Ministry of Education, Youth and Sports (project LTC17016). **SGS** acknowledges the financial support of UEFISCDI grant PN-III-P1-1.1-TE-2016-2147 (CORIMAG). **VB** and **PPG** acknowledge the France-BioImaging infrastructure supported by the French National Research Agency (ANR-10-INBS-04).

Author contribution

ST and **RaM** conceptualized BIAFLOWS, supervised its implementation, contributed to all technical tasks and wrote the manuscript. **UR** worked on the core implementation of BIAFLOWS with contributions from **GM** and **RH**. **RoM** implemented several modules to interface bioimage analysis workflows and the content of the system. **ST, MM** and **DU** implemented the module to compute benchmark metrics. **VB**, **RoM**, **BP**, **MM, RV, LAS** and **LP** integrated workflows and tested the system. **AC**, **DU**, **OG,** and **GB** organized and collected content (image datasets, simulation tools). **SGS**, **NS** and **PG** provided extensive feedback on BIAFLOWS and contributed to the manuscript. All authors took part in reviewing and improving the manuscript.

Competing interests

RaM and **RH** are co-founders of the non-profit cooperative company Cytomine SCRL FS.

Data availability

All images and annotations can be downloaded from the online instance of BIAFLOWS at https://biaflows.neubias.org/.

Code availability

BIAFLOWS is an open source project and its code source can be downloaded at https://github.com/Neubias-WG5. The procedure to install a local instance of BIAFLOWS is described in Supplementary Methods section 3.

Additional information

Methods to access BIAFLOWS online instance including a user guide, and video tutorials, as well as the procedure to install and populate a local instance of BIAFLOWS (images, annotations, workflows) and migrate content between instances are available in Supplementary Methods.

References

1. Ouyang, W., & Zimmer, C. (2017). The imaging tsunami: computational opportunities and challenges. Current Opinion in Systems Biology, 4, 105-113.
2. Kevin W Eliceiri et al. (2012). Biological imaging software tools. Nature Methods volume 9, pages 697–710.
3. Carpenter, Anne E., Kamentsky, Lee and Eliceiri, Kevin W. (2012). A call for bioimaging software usability. Nature methods 9, no. 7: 666-670.
4. Schneider, C. A.; Rasband, W. S. & Eliceiri, K. W. (2012), NIH Image to ImageJ: 25 years of image analysis, Nature methods 9(7): 671-675.
5. Munafò, M. R. et al. (2017). A manifesto for reproducible science. Nat. Hum. Behav. 1, 0021.
6. Ellenberg et al. (2018). A call for public archives for biological image data. Nature Methods October, Pages 849–854.
7. Chris Allan, et al. (2012) OMERO: flexible, model-driven data management for experimental biology. Nature Methods 9, 245–253.
8. Kristian Kvilekval, et al. (2010). Bisque: a platform for bioimage analysis and management, Bioinformatics, Volume 26, Issue 4, 15 February, Pages 544–552.
9. Eleanor Williams, et al. (2017). Image Data Resource: a bioimage data integration and publication platform. Nature Methods volume 14, pages 775–781.
10. Vandewalle P. (2012). Code sharing is associated with research impact in image processing. Computing 799 in Science & Engineering, 14(4):42–47.
11. Maier-Hein L. (2018). Why rankings of biomedical image analysis competitions should be interpreted with care. Nature Communications, 9:5217.
12. Grand Challenges in Biomedical Image Analysis, https://grand-challenge.org/
13. Perkel J.M. (2018). A toolkit for data transparency takes shape. Nature Technology Feature. August, 560,513-515.
14. Grüning BA, Rasche E, Rebolledo-jaramillo B, et al. (2017). Jupyter and Galaxy: Easing entry barriers into complex data analyses for biomedical researchers. PLoS Comput Biol.;13(5):e1005425.
15. Marée et al. (2016). Collaborative analysis of multi-gigapixel imaging data with Cytomine, Bioinformatics, 32(9):1395-401.
16. Tristan Glatard et al. (2018). Boutiques: a flexible framework to integrate command-line applications in computing platforms, *GigaScience*, Volume 7:5, 1 May.
17. Kurtzer GM, Sochat V, Bauer MW. (2017). Singularity: Scientific containers for mobility of compute. PLoS ONE 12(5): e0177459.

18. Yoo A., Jette M., and Grondona M. (2003). Slurm: Simple Linux Utility for Resource Management, Job Scheduling Strategies for Parallel Processing, volume 2862 of Lecture Notes in Computer Science, pages 44-60, Springer-Verlag.
19. Svoboda D, Kozubek M, Stejskal S. (2009). Generation of digital phantoms of cell nuclei and simulation of image formation in 3D image cytometry. Cytometry A.; 75: 494–509.
20. Kozubek M. Challenges and Benchmarks in Bioimage Analysis. (2016). Adv Anat Embryol Cell Biol. 219:231-62.
21. K. M. Brown et al., The DIADEM Data Sets: Representative Light Microscopy Images of Neuronal Morphology to Advance Automation of Digital Reconstructions. Neuroinformatics. 2011 Sep; 9(0): 143–157. http://diademchallenge.org/
22. M. Maška et al., A benchmark for comparison of cell tracking algorithms. Cell Tracking Challenge. Bioinformatics. 2014 Jun 1; 30(11): 1609–1617. http://celltrackingchallenge.net/
23. N. Chenouard et al., Objective comparison of particle tracking methods. Nature Methods (March 2014), Volume 11, Issue 3, 281-289 http://bioimageanalysis.org/track/
24. Caicedo et al., Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nature Methods, 2019. https://www.kaggle.com/c/data-science-bowl-2018
25. D. Wiesner et al., CytoPacq: a web-interface for simulating multi-dimensional cell imaging. Bioinformatics https://cbia.fi.muni.cz/simulator/
26. Cuntz H, Forstner F, Borst A, Häusser M (2010). One rule to grow them all: A general theory of neuronal branching and its practical application. PLoS Comput Biol 6(8): e1000877. TREES toolbox. https://www.treestoolbox.org/
27. Preet Jassi and Ghassan Hamarneh. VascuSynth: Vascular Tree Synthesis Software. Insight Journal, 2011. http://vascusynth.cs.sfu.ca
28. A. Lehmussola et al. Computational framework for simulating fluorescence microscope images with cell populations. IEEE Trans Med Imaging. 2007 Jul;26(7):1010-6. SIMCEP tool. http://www.cs.tut.fi/sgn/csb/simcep/tool.html
29. C.A. Schneider et al., NIH Image to ImageJ: 25 years of image analysis. Nature Methods volume 9, pages 671–675 (2012). https://imagej.nih.gov/ij/developer/macro/macros.html
30. F. de Chaumont et al. Icy: an open bioimage informatics platform for extended reproducible research. Nature Methods volume 9, pages 690–696 (2012). http://icy.bioimageanalysis.org/plugin/Protocols
31. C. McQuin et al., CellProfiler 3.0: Next-generation image processing for biology. PL. oS Biol. 16(7):e2005970, 2018. https://cellprofiler.org/examples/
32. H. Peng et al., V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets, Nature Biotechnology, Vol. 28, No. 4, pp. 348-353 https://github.com/Vaa3D/Vaa3D_Wiki/wiki/Vaa3DPlugins.wiki
33. S. Berg et al. ilastik: interactive machine learning for (bio)image analysis. Nature Methods (2019).

https://www.ilastik.org/

34. Pedregosa et al., Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830, 2011. https://scikit-learn.org
35. F. Chollet et al., Keras. 2015. https://keras.io/
36. A. Paszke et al. Automatic Differentiation in PyTorch. NIPS Autodiff Workshop, 2017. https://pytorch.org/
37. Vandaele R. et al. (2018). Landmark detection in 2D bioimages for geometric morphometrics: a multi-resolution tree-based approach, Scientific Reports 8, Article number; 538.
38. K. Sirinukunwattana et al., Gland segmentation in colon histology images: The glas challenge contest. Medical Image Analysis, Volume 35, January 2017, Pages 489-502.
https://warwick.ac.uk/fac/sci/dcs/research/tia/glascontest/