biopixR: Extracting Insights from Biological Images

24 June 2024

Summary

biopixR is an R package designed specifically for the analysis of bioimage data. The package comprises 15 functions that cater to a variety of tasks, including image import, segmentation, feature extraction, quantification, clustering, parameter optimization, and a distinctive algorithm for line gap mitigation. Among these is an optimized edge detection algorithm based on the Canny edge detection algorithm. Together with a noise-resilient threshold method, these two algorithms form the foundation of feature extraction within the biopixR package. Both edge detection and thresholding can be parameterized via an interactive Tcl/Tk user interface. For circular objects, the threshold adjustment factor and the smoothing factor for edge detection can be automatically calculated using Gaussian process regression models (Kriging models). Thereby making an automated feature extraction, which includes segmentation and labeling, available in R. Furthermore, all available methods for feature extraction and filtering can be accessed within a single pipeline function, not found in other packages. The package also includes a function for clustering objects within an image based on their shape features using Self-Organizing Maps (SOM).

Designed for medium-throughput analysis, the package is able to analyze whole directories, utilize multiple cores for parallel processing, and generate detailed log files to track the analytical process. Furthermore, the biopixR package offers distinctive data sets of microbead images and microbeads in water-oil emulsions. The fundamental functionality of biopixR was recently employed in the study by Geithe et al. (2024) to perform quality control on microbeads. As the biopixR package provides access to a range of fundamental data, including object size, quantity, shape, and intensity, it is a valuable tool for researchers in the field. The biopixR package leverages capabilities from the imager (Barthelmé and Tschumperlé 2019) and magick (Ooms 2024) packages to perform its own tasks related to biological image processing and analysis.

Statement of need

Imaging plays a pivotal role in data acquisition within biological laboratories and the broader life sciences. It offers crucial insights into cells (Rödiger et al. 2018; Schneider et al. 2019), biomarkers (Vafajoo et al. 2018), stress responses (Korkmaz, Debelec Butuner, and Roggenbuck 2018), and gene expression (Brenner et al. 2000), which are vital for diagnostics and disease prediction. Microbead technology, known for its rapid and flexible data collection, relies heavily on imaging techniques like microscopy (Rödiger et al. 2014; Ding et al. 2020). The challenge lies in the fast and accurate extraction of valuable information from complex biological images, particularly in the quantification of objects within these images.

The quantification of biological image data is a fundamental aspect of deriving insights into biological processes. This is exemplified by microbead-assays, which offer extensive multiplexing abilities, high-throughput capacity, reduced analysis time, and minimal sample requirements (Rödiger et al. 2014). Methods for detecting microbeads include flow cytometry, microfluidics, and image-based techniques (Choi et al. 2019). A user-friendly tool integrated into R could enhance analysis by combining statistical analysis and visualization techniques. The advancement of smart device imaging and chip development is facilitating the utilization of

microbeads in point-of-care testing (POCT) and disease diagnosis (Dinter et al. 2019; Zhang et al. 2019). This progress underscores the necessity for the development of efficient methods for extracting biological information from images, thereby eliminating the necessity for the use of complex laboratory equipment. Software that extracts attributes such as fluorescence intensity, size and shape allow for microbead encoding and differentiation between populations, thereby enhancing multiplexing capabilities (Zhang et al. 2019).

Microbead-based emulsion Polymerase Chain Reaction (ePCR) assays, which are typically analyzed using fluorescence-activated cell sorting (FACS) (Fraser et al. 2015), could potentially benefit from the use of imaging techniques. As this software is capable of analyzing microbead-based ePCR by preprocessing brightfield droplet images with the provided gap-filling algorithm. The aforementioned applications extend to the assessment of wastewater for the detection of microplastics (Ding et al. 2020), the real-time localization in microbead-based drug delivery systems (Bannerman and Wan 2016), and other fields of life science, such as cell biology.

Consequently, the `biopixR` package for `R` is a fundamental tool. The automated evaluation process enables medium-throughput analysis directly from images, simplifying analytical procedures and expanding experimental possibilities.

Software engineering

`biopixR` (1.0.0, LGLP-3.0 license) is an `R` (R Core Team 2023a) package (S3 class system). To ensure code quality, we employed various methodologies, including Continuous Integration (CI), unit testing (Wickham 2011), adherence to naming conventions (Bååth 2012), and the application of style guidelines (Wickham 2019). The choice of `R` as a programming language for image processing is based on its open-source framework, bindings to other languages (Reimert et al. 2024) to improve perfomance, advanced packages such as `imager` (Barthelmé and Tschumperlé 2019) and `magick` (Ooms 2024), and its strong support for reproducible research (Xie 2019). The importance of ensuring software reproducibility is widely recognized, not only for our own work but also for the broader scientific community (Gentleman and Temple Lang 2007 ; Rödiger et al. 2015). One measure to achieve this is by minimizing dependencies on other packages or libraries and single archives whenever possible. Therefore, `biopixR` depends on `R` ($\geq$ 4.2.0), `imager`, `magick` and `tcltk`, imports `data.table` and `cluster` and suggests `knitr`, `rmarkdown`, `doParallel`, `kohonen`, `imagerExtra`, `GPareto` and `foreach` exclusively from the Comprehensive R Archive Network (CRAN). These features collectively provide a sophisticated environment for image analysis and editing, with the added advantage of community-driven improvements (Chambers 2008). `R`'s integration of analysis with documentation ensures methodological precision and transparency in scientific research, making it a preferred choice for complex image processing tasks. The `biopixR` package underwent quality control through unit testing using the `testthat` package (Wickham 2011), ensuring its reliability.

Installation

The `biopixR` package can be installed from CRAN or GitHub, providing users with stable and developmental versions respectively.

To install the stable version of `biopixR` from CRAN, execute the following command in `R`:

```
install.packages("biopixR")
```

This command will download and install the latest stable release of the package, ensuring compatibility and reliability.

For users interested in the latest features and ongoing developments, the developmental version of `biopixR` is available on GitHub. To install this version, it is first necessary to install the `devtools` package if it is not already present in your R environment:

```
install.packages("devtools")
devtools::install_github("Brauckhoff/biopixR")
```

This command enables the installation of the development build from the GitHub repository. This provides access to the most recent features and updates that may not yet be available in the CRAN release.

Functions

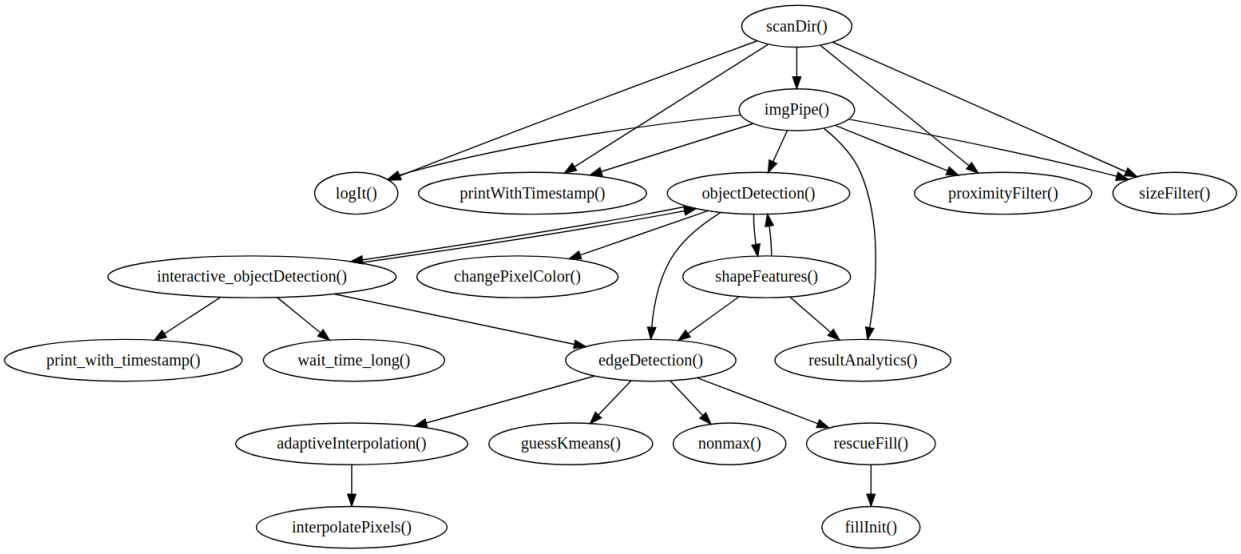

Figure 1: Dependency graph of the functions present in the `biopixR` package. Showing the levels of complexity by showing the descendants and ancestors of the `imgPipe()` function. The figure was created using the `foodwebr` package from Appleton-Fox (2022) (package version 0.1.1, RStudio 2023.09.0+463, R 4.3.2 on Linux, Ubuntu 22.04.3 LTS).

The `biopixR` package is intended for analyzing bioimage data in `R`, with a specific focus on the analysis and characterization of microbeads. The package provides tools for image preprocessing, segmentation, feature extraction, filtering, and visualization. It supports automation for medium-throughput analysis, utilizing algorithms to identify spherical objects, extract their features, and implement interactive tools for threshold and smoothing factor selection. Furthermore, it offers features for removing clumped or closely positioned particles to prevent inaccurate results, with the goal of improving the analysis of microparticles in diverse scientific disciplines.

The `shapeFeatures()` function is capable of extracting shape-related information from detected objects and grouping them using the SOM algorithm (Wehrens and Buydens 2007). The central function of `biopixR` is `imgPipe()`, which performs object detection and applies individual filters. This pipeline integrates all the fundamental procedures for comprehensive analysis into a single function. The user is required to provide the input image and select the appropriate detection methods to gain insights into the specific objectives present in the image. Furthermore, the user has the option to customize their workflow with individual filters. In addition, the function is capable of processing multiple color channels, such as the analysis of dual-color microbeads. The integration of this function within the dependency network and its interaction with other functions is illustrated in Figure 1. This function serves as a comprehensive pipeline for image analysis, offering a variety of selectable functions:

- `importImage()`, joins import functions of the `imager` (Barthelmé and Tschumperlé 2019) and `magick` (Ooms 2024) packages,

- `edgeDetection()`, a combination of a Canny edge detector and gap filling (Barthelmé and Tschumperlé 2019),
- `objectDetection()`, detects objects in an image by edge detection or thresholding,
- `sizeFilter()`, eliminates objects that exceed or fall below a certain size threshold,
- `proximityFilter()`, filters objects that are in proximity,
- `resultAnalytics()`, summarizes the extracted features in a clear and concise manner.

The `biopixR` package includes functions for analyzing entire directories, allowing for medium-throughput analysis. Making feature extraction and image clustering easily accessible:

- `haralickCluster()`, extracts Haralick features and clusters information using Partitioning Around Medoids (PAM) (Haralick, Shanmugam, and Dinstein 1973; Carlson 2018; Maechler et al. 2023),
- `scanDir()`, utilizing the pipeline for whole directory analysis.

The `fillLineGaps()` algorithm, along with helper functions:

- `interpolatePixels()`, calculates the coordinates required to connect two given points,
- `adaptiveInterpolation()`, searches a given radius surrounding a line end for contours and connects them,

addresses the issue of discontinuous edges by iteratively scanning for line ends within the image and reconnecting them to adjacent contours. Other approaches for dealing with discontinuous contours are not found in other `R` packages.

Examples demonstrating the use of `biopixR` for image analysis tasks can be found in the package's vignette (Brauckhoff and Rödiger 2024).

Graphical User Interface

The `biopixR` package has broader applicability beyond microbead detection. The `biopixR` package is adaptable and can be utilized in any research field where the identification of distinct objects in images can be achieved through the use of a Canny edge detector or thresholding. This encompasses research areas such as foci detection, microplastics, and plant seeds. The automation methods employed in `biopixR` are predicated on the assumption of circular objects, rendering it particularly well-suited for the detection, quantification, and extraction of useful information from circular objects within images. Another integrated tool is an interactive function that assists the user in selecting the optimal input for their analysis. The function `interactive_objectDetection()` initiates a graphical user interface (GUI) (Figure 2) that utilizes the Tcl/Tk framework (R Core Team 2023b), enabling users to adjust the threshold and smoothing settings of the image.

Conclusion

In summary, `biopixR` represents a significant advancement in image analysis capabilities for `R` users. By providing both automation methods and interactive tools, the package empowers researchers to extract valuable insights from images with ease. Its adaptability across various research fields makes it a valuable tool for researchers seeking efficient solutions. Leveraging the power of `R` and its extensive library of packages, users can seamlessly integrate `biopixR` into their existing workflows, streamlining data analysis and visualization tasks. Moreover, the package's minimal dependencies ensure long-term stability and maintainability over time, making it an attractive choice for researchers seeking reproducible software. A comparative analysis between human and software-based object quantification from images underscores the importance of the `biopixR` package in terms of accuracy and reproducibility (Brauckhoff and Rödiger 2024).

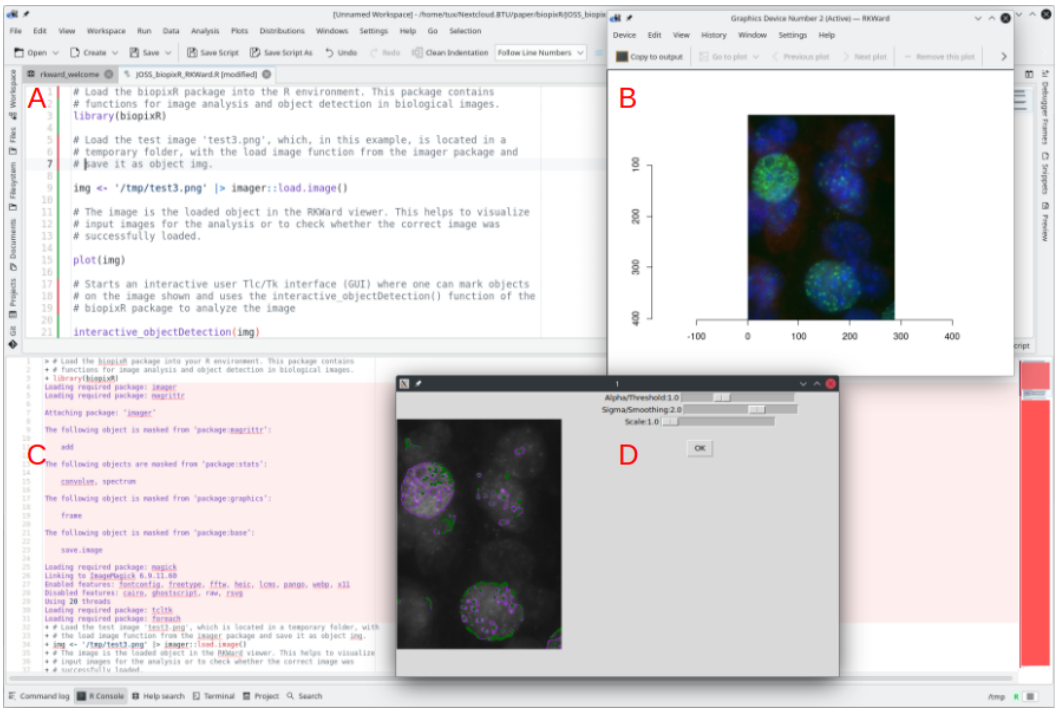

Figure 2: Graphical User Interface for interactive parameter selection. The function `interactive_objectDetection()` provides a simple interface with sliders to adjust threshold, smoothing, and scale. It highlights object contours in purple and centers in green for easy visualization. A) In this example, the GUI was used in RKWard (0.7.5z+0.7.6+devel3, Linux, TUXEDO OS 2, (Rödiger et al. 2012)). With a few commands, an image can be imported and analyzed. B) The `plot()` function displays the false-color image as a preview. In this figure, cells with DNA damage (similar to Rödiger et al. (2018)) are visible. C) Loading the `biopixR` package in the `R` console shows additional information such as loaded libraries and the number of CPU threads (n = 20, parallel processing). D) The rendering process is displayed on the console, including timestamps and the current state.

Acknowledgments

The study was funded in part by the project Rubin: NeuroMiR (03RU1U051A, federal ministry of education and research, Germany).

References

Appleton-Fox, Lewin. 2022. *Foodwebr: Visualise Function Dependencies*. https://github.com/lewinfox/foodwebr.

Bååth, Rasmus. 2012. "The State of Naming Conventions in R." *The R Journal* 4 (2): 74–75. https://doi.org/10.32614/RJ-2012-018.

Bannerman, Dawn, and Wankei Wan. 2016. "Multifunctional Microbeads for Drug Delivery in TACE." *Expert Opinion on Drug Delivery* 13 (9): 1289–1300. https://doi.org/10.1080/17425247.2016.1192122.

Barthelmé, Simon, and David Tschumperlé. 2019. "Imager: An R Package for Image Processing Based on CImg." *Journal of Open Source Software* 4 (38): 1012. https://doi.org/10.21105/joss.01012.

Brauckhoff, Tim, and Stefan Rödiger. 2024. "biopixR: Extracting Insights from Biological Images." https://doi.org/10.5281/ZENODO.12744222.

Brenner, Sydney, Maria Johnson, John Bridgham, George Golda, David H. Lloyd, Davida Johnson, Shujun Luo, et al. 2000. "Gene Expression Analysis by Massively Parallel Signature Sequencing (MPSS) on Microbead Arrays." *Nature Biotechnology* 18 (6): 630–34. https://doi.org/10.1038/76469.

Carlson, Joel. 2018. "'Radiomic' Image Processing Toolbox." *Comprehensive R Archive Network*. https://cran.r-project.org/package=radiomics.

Chambers, John. 2008. *Software for Data Analysis: Programming with R. Statistics and Computing*. Springer New York. https://doi.org/10.1007/978-0-387-75936-4.

Choi, Youngeun, Carsten Schmidt, Philip Tinnefeld, Ilko Bald, and Stefan Rödiger. 2019. "A New Reporter Design Based on DNA Origami Nanostructures for Quantification of Short Oligonucleotides Using Microbeads." *Scientific Reports* 9 (1). https://doi.org/10.1038/s41598-019-41136-x.

Ding, Ning, Di An, Xiufeng Yin, and Yingxue Sun. 2020. "Detection and Evaluation of Microbeads and Other Microplastics in Wastewater Treatment Plant Samples." *Environmental Science and Pollution Research* 27 (13): 15878–87. https://doi.org/10.1007/s11356-020-08127-2.

Dinter, Franziska, Michał Burdukiewicz, Peter Schierack, Werner Lehmann, Jörg Nestler, Gregory Dame, and Stefan Rödiger. 2019. "Simultaneous Detection and Quantification of DNA and Protein Biomarkers in Spectrum of Cardiovascular Diseases in a Microfluidic Microbead Chip." *Analytical and Bioanalytical Chemistry* 411 (29): 7725–35. https://doi.org/10.1007/s00216-019-02199-x.

Fraser, Lewis, Andrew Kinghorn, Marco Tang, Yee-Wai Cheung, Bryce Lim, Shaolin Liang, Roderick Dirkzwager, and Julian Tanner. 2015. "Oligonucleotide Functionalised Microbeads: Indispensable Tools for High-Throughput Aptamer Selection." *Molecules* 20 (12): 21298–312. https://doi.org/10.3390/molecules201219766.

Geithe, Christiane, Bo Zeng, Carsten Schmidt, Franziska Dinter, Dirk Roggenbuck, Werner Lehmann, Gregory Dame, Peter Schierack, Katja Hanack, and Stefan Rödiger. 2024. "A Multiplex Microchamber Diffusion Assay for the Antibody-Based Detection of microRNAs on Randomly Ordered Microbeads." *Biosensors and Bioelectronics: X* 18 (June): 100484. https://doi.org/10.1016/j.biosx.2024.100484.

Gentleman, Robert, and Duncan Temple Lang. 2007. "Statistical Analyses and Reproducible Research." *Journal of Computational and Graphical Statistics* 16 (1): 1–23. https://doi.org/10.1198/106186007X178663.

Haralick, Robert M., K. Shanmugam, and Its'Hak Dinstein. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6): 610–21. https://doi.org/10.1109/tsmc.1973.4309314.

Korkmaz, Kemal Sami, Bilge Debelec Butuner, and Dirk Roggenbuck. 2018. "Detection of 8-OHdG as a Diagnostic Biomarker." *Journal of Laboratory and Precision Medicine* 3 (November): 95–95. https://doi.org/10.21037/jlpm.2018.11.01.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. 2023. *Cluster: Cluster Analysis Basics and Extensions*. https://doi.org/10.32614/CRAN.package.cluster.

Ooms, Jeroen. 2024. *Magick: Advanced Graphics and Image-Processing in R*. https://docs.ropensci.org/magick/ https://ropensci.r-universe.dev/magick.

R Core Team. 2023a. "R: A Language and Environment for Statistical Computing." https://www.R-project.org/.

———. 2023b. "Tcl/Tk Interface." https://CRAN.R-project.org/package=tcltk.

Reimert, Mossa Merhi, Josiah D. Parry, Matt Denwood, Maya Katrin Gussmann, Claus O. Wilke, Ilia Kosenkov, Michael Milton, and Amy Thomason. 2024. "Extendr: Frictionless Bindings for R and Rust."

Journal of Open Source Software 9 (99): 6394. https://doi.org/10.21105/joss.06394.

Rödiger, Stefan, Michał Burdukiewicz, Konstantin A. Blagodatskikh, and Peter Schierack. 2015. "R as an Environment for the Reproducible Analysis of DNA Amplification Experiments." *The R Journal* 7 (2): 127–50. https://doi.org/10.32614/RJ-2015-011.

Rödiger, Stefan, Thomas Friedrichsmeier, Prasenjit Kapat, and Meik Michalke. 2012. "RKWard: A Comprehensive Graphical User Interface and Integrated Development Environment for Statistical Analysis with R." *Journal of Statistical Software* 49 (9): 1–34. https://doi.org/10.18637/jss.v049.i09.

Rödiger, Stefan, Claudia Liebsch, Carsten Schmidt, Werner Lehmann, Ute Resch-Genger, Uwe Schedler, and Peter Schierack. 2014. "Nucleic Acid Detection Based on the Use of Microbeads: A Review." *Microchimica Acta* 181 (11–12): 1151–68. https://doi.org/10.1007/s00604-014-1243-4.

Rödiger, Stefan, Marius Liefold, Madeleine Ruhe, Mark Reinwald, Eberhard Beck, and P. Markus Deckert. 2018. "Quantification of DNA Double-Strand Breaks in Peripheral Blood Mononuclear Cells from Healthy Donors Exposed to Bendamustine by an Automated γH2AX Assay—an Exploratory Study." *Journal of Laboratory and Precision Medicine* 3 (May): 47–47. https://doi.org/10.21037/jlpm.2018.04.10.

Schneider, Jens, Romano Weiss, Madeleine Ruhe, Tobias Jung, Dirk Roggenbuck, Ralf Stohwasser, Peter Schierack, and Stefan Rödiger. 2019. "Open Source Bioimage Informatics Tools for the Analysis of DNA Damage and Associated Biomarkers." *Journal of Laboratory and Precision Medicine* 4 (0): 1–27. https://doi.org/10.21037/jlpm.2019.04.05.

Vafajoo, Atieh, Azin Rostami, Sanam Foroutan Parsa, Reza Salarian, Navid Rabiee, Ghazal Rabiee, Mohammad Rabiee, et al. 2018. "Multiplexed Microarrays Based on Optically Encoded Microbeads." *Biomedical Microdevices* 20 (3). https://doi.org/10.1007/s10544-018-0314-4.

Wehrens, Ron, and Lutgarde M. C. Buydens. 2007. "Self- and Super-Organizing Maps in R: The kohonen Package." *Journal of Statistical Software* 21 (5): 1–19. https://doi.org/10.18637/jss.v021.i05.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3 (1): 5. https://doi.org/10.32614/rj-2011-002.

———. 2019. *Advanced R*. Second edition. Chapman & Hall/CRC the r Series. Boca Raton: CRC Press, Taylor & Francis Group.

Xie, Yihui. 2019. *R Markdown: The Definitive Guide*. Edited by Joseph J. Allaire and Garrett Grolemund. Chapman & Hall/CRC, the R Series. Boca Raton: CRC Press, Taylor & Francis Group.

Zhang, Jing, Swati Shikha, Qingsong Mei, Jinliang Liu, and Yong Zhang. 2019. "Fluorescent Microbeads for Point-of-Care Testing: A Review." *Microchimica Acta* 186 (6). https://doi.org/10.1007/s00604-019-3449-y.