The Journal of Open Source Software

HSI-Wizard: A magical open-source Python package for medical hyperspectral imaging applications

Felix Wühler [1] and **Matthias Rädle** [1]

1 CeMOS Research and Transfer Center, University of Applied Science Mannheim, Germany

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

`hsi-wizard` is an open-source Python package designed for processing, analysing, and visualising hyperspectral datasets, primarily in medical applications. Hyperspectral data refers to images where each pixel contains multiple intensity values at different wavelengths, with the number of spectral bands ranging from only a few to well over a thousand. These datasets vary in resolution and structure, making analysis difficult due to inconsistent formats and limited reproducibility. To address these challenges, `hsi-wizard` provides a standardised data representation and analysis protocols that enhance reproducibility. The package features the `DataCube`, a standardised class for representing hyperspectral data, which simplifies data manipulation compared to traditional array-based tools like NumPy. The `DataCube` includes built-in methods tailored specifically for hyperspectral data, facilitating quicker exploration and reducing complexity for users.

A notable feature of the `DataCube` class is its protocol functionality, which records the methods used from the `hsi-wizard` library. This protocol acts as a reusable template for other datasets, allowing researchers, including those without programming experience, to replicate analyses accurately and effortlessly. Additionally, `hsi-wizard` allows users to merge `DataCubes` due to its standardised representation. This capability enables researchers, for example, to easily combine near-infrared (IR) and Raman spectroscopy datasets, resulting in more comprehensive and detailed tissue analyses. Such multimodal analysis simplifies the process and provides greater insights from combined data.

Overall, `hsi-wizard` streamlines the workflow for processing, analysing, and visualising hyperspectral data. It caters to both beginners and experts. Students can utilise straightforward methods for educational purposes, while researchers can leverage advanced functions for professional studies, ensuring that results are reproducible and transparently documented.

Statement of Need

Hyperspectral imaging (HSI) allows for an in-depth analysis of the electromagnetic spectrum across multiple wavelengths for each pixel in an image. Initially developed for NASA applications, HSI is now applied in various fields, including medicine, agriculture, environmental monitoring and more (Bhargava et al., 2024). In medical research, for example, HSI is instrumental in distinguishing healthy tissue from cancerous areas or detecting early-stage diseases.

The evolution of these fields has led to a variety of methods for acquiring hyperspectral data, encompassing different measurement techniques such as reflection, transmission, and fluorescence, as well as varying wavelength ranges (ultraviolet, visible, infrared) and scanning methods (e.g., point scanning, line scanning, Fourier transform infrared imaging (FTIR)) (Lu & Fei, 2014). Consequently, datasets exhibit diverse formats, resolutions, and spectral ranges, which complicates analysis workflows and can hinder efficiency and reproducibility.

The hsi-wizard addresses these challenges by standardising data representation through its DataCube class, enabling consistent analysis across diverse hyperspectral datasets. Unlike traditional software tools like NumPy or specialised applications such as ENVI or SPy, which often require extensive programming or manual adjustments to manage variability, hsi-wizard streamlines data handling and integrates multimodal datasets seamlessly. It also promotes reproducibility through automated protocol logging.

Moreover, the extensibility of hsi-wizard allows users to incorporate new analysis methods or customise existing workflows easily, ensuring adaptability to evolving research needs. This flexibility guarantees that hsi-wizard remains relevant across various research scenarios, significantly reducing barriers to hyperspectral data analysis while promoting transparency and reproducibility in scientific research.

To further facilitate the analysis of diverse datasets and eliminate the need for custom methods for each, hsi-wizard standardises their representation through the DataCube class. This standardisation is crucial for enabling data fusion across different scanning processes, as scans from various scanners yield unique datasets with differing resolutions and aspect ratios. Thus, merging these scans begins with a well-defined data representation and methods to address these challenges. hsi-wizard provides these capabilities, simplifying the combination of datasets for subsequent analysis.

Additionally, hsi-wizard allows users to log and save manipulations of the DataCube. These logs can be reused for similar DataCubes, ensuring consistent analysis across different measurements. This feature diminishes the need for programming expertise and enhances reproducibility, thereby supporting the goal of increased transparency in research (Knottnerus & Tugwell, 2016).

Comparison with Existing Tools

A variety of open-source and commercial tools exist for hyperspectral imaging, but most are tailored to geospatial tasks, low-level preprocessing, or GUI-driven workflows. Libraries like PySptools and SPy focus on raster I/O and classification for remote sensing, while tools such as HyDe provide GPU-based denoising but lack biomedical integration. Commercial options like ENVI or Spectronon offer user-friendly interfaces but are often tied to hardware and limited in scripting or batch automation.

In contrast, hsi-wizard is designed specifically for biomedical use. It supports multimodal fusion, protocol logging, and scripting within a unified Python environment. It handles formats like ENVI, CSV, NRRD, and TDMS and integrates cleanly into programmatic workflows. Unlike GUI-centric tools, hsi-wizard enables reproducible, automated processing for diverse spectral datasets, filling a gap between algorithm libraries and rigid GUI systems.

This makes hsi-wizard especially suitable for clinical imaging and research, combining flexibility, transparency, and domain-specific functionality in one open-source tool.

Example Usage

This snippet demonstrates the use of the hsi-wizard package for processing and visualizing hyperspectral data, using a real sample from the HeiPorSPECTRAL dataset (Studier-Fischer et al., 2023). The focus is on the *spleen* example (P086#2021_04_15_09_22_02). hsi-wizard handles the entire pipeline: from reading the raw .dat hyperspectral DataCube, managing metadata like wavelengths, to applying PCA and agglomerative spatial clustering. It enables concise and structured exploration of spectral information in biomedical imaging. The full example is visible in the documentation.

```
import wizard
from wizard._processing.cluster import pca, spatial_agglomerative_clustering, smooth_clu
```

```
# Define Custom reader
def read_spectral_cube(path) -> wizard.DataCube:
    shape = np.fromfile(path, dtype=">i", count=3)
    cube = np.fromfile(path, dtype=">f", offset=12).reshape(*shape)
    cube = np.swapaxes(np.flip(cube, axis=1), 0, 1).astype(np.float32)
    wavelengths = np.linspace(500, 1000, cube.shape[2], dtype='int')
    return wizard.DataCube(cube.transpose(2, 0, 1), wavelengths=wavelengths, notation='n

# Initalize dc and read data
dc = wizard.DataCube()
dc.set_custom_reader(read_spectral_cube)
dc.custom_read('2021_04_15_09_22_02_SpecCube.dat')

# Inspect Data
wizard.plotter(dc)
```

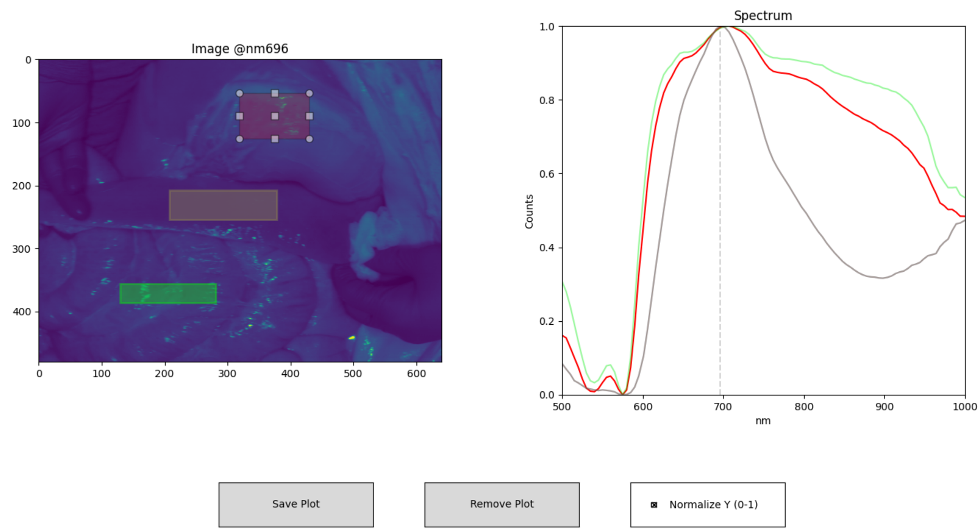

Figure 1: ROI-based spectral analysis with the interactive plotting interface of hsi-wizard. The left panel shows selected tissue regions at 696 nm, while the right panel displays corresponding normalized spectra of the rois.

```
# Clustering
dc_pca = pca(dc, n_components=10)
agglo = spatial_agglomerative_clustering(dc_pca, n_clusters=5)
agglo = smooth_cluster(agglo, n_iter=10, sigma=0.5)
```

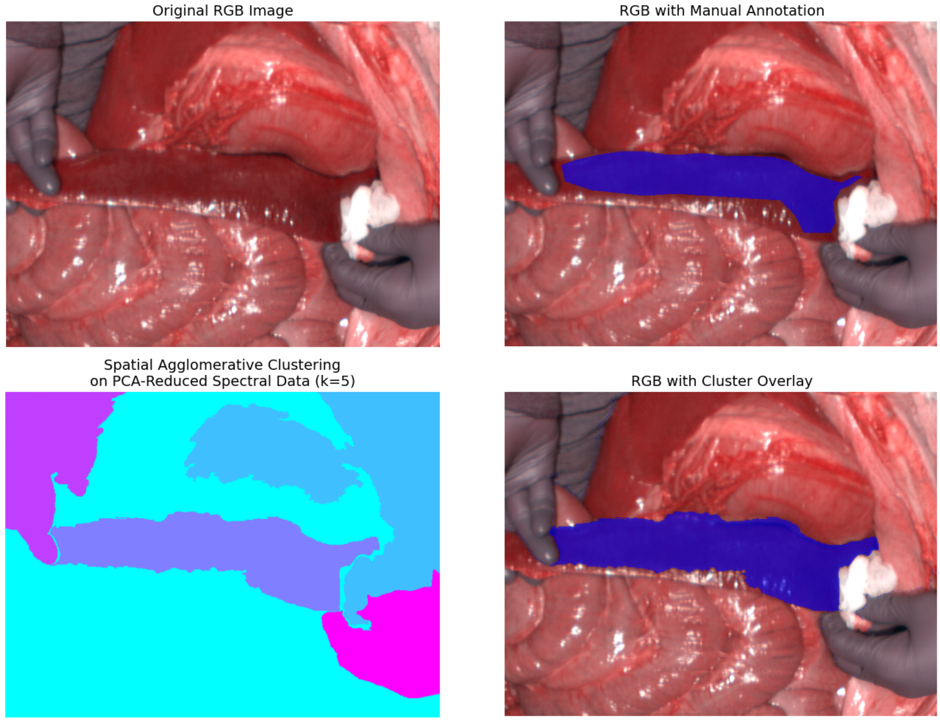

Figure 2: Comparison between manual annotation (top-right) from (Studier-Fischer et al., 2023) and automated segmentation (bottom-left) using spatial agglomerative clustering ($k = 5$) on PCA-reduced hyperspectral data. The original RGB image (top-left) and the resulting cluster map (bottom-right) provide visual context and output structure.

Future Work

The development of `hsi-wizard` is ongoing, with future goals that include expanding support for additional data formats and integrating advanced data analytics. We encourage feedback, contributions, and ideas from the community. All versions of `hsi-wizard` are available on the Python Package Index (PyPI) or on Github (Wuehler, 2025).

Acknowledgements

The development of hsi-wizard was funded by the CeMOS Research & Transfer Center and the Technical University of Applied Sciences Mannheim. The software was developed based on practical requirements and datasets provided by our research group: Kümmel (T. et al., 2021) shaped the mid-infrared data handling and clustering through his work on brain tissue imaging; Heintz (Heintz et al., 2021) provided multimodal Raman and VIS use cases; Manser (Manser et al., 2023) contributed workflows for Raman light sheet microscopy; Nachtmann (Nachtmann et al., 2024) validated complex 3D Raman pipelines; and van Marwick (Marwick et al., 2024) influenced preprocessing through chemical segmentation tasks.

References

Bhargava, A., Sachdeva, A., Sharma, K., Alsharif, M. H., Uthansakul, P., & Uthansakul, M. (2024). Hyperspectral imaging and its applications: A review. *Heliyon*, *10*(12), e33208. https://doi.org/10.1016/j.heliyon.2024.e33208

Heintz, A., Sold, S., Wühler, F., Dyckow, J., Schirmer, L., Beuermann, T., & Rädle, M. (2021). Design of a multimodal imaging system and its first application to distinguish grey and white matter of brain tissue. A proof-of-concept-study. *Applied Sciences*, *11*(11). https://doi.org/10.3390/app11114777

Knottnerus, J. A., & Tugwell, P. (2016). Promoting transparency of research and data needs much more attention. *Journal of Clinical Epidemiology*, *70*, 1–3. https://doi.org/10.1016/j.jclinepi.2016.01.007

Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, *19*(1), 010901. https://doi.org/10.1117/1.JBO.19.1.010901

Manser, S., Keck, S., Vitacolonna, M., Wuehler, F., Rudolf, R., & Raedle, M. (2023). Innovative imaging techniques: A conceptual exploration of multi-modal raman light sheet microscopy. *Micromachines*, *14*(9). https://doi.org/10.3390/mi14091739

Marwick, B. van, Kümmel, T., Wühler, F., Lauer, F., Hoffmann, J., & Rädle, M. (2024). Rapid chemical detection and segmentation of latent fingerprints by means of a novel middle-infrared scanning method. *Analyst*, –. https://doi.org/10.1039/D4AN00367E

Nachtmann, M., Feger, D., Wühler, F., Scholl, S., & Rädle, M. (2024). Three-dimensional, molecule-sensitive mapping of structured falling liquid films using raman scanning. *Chemie Ingenieur Technik*, *96*(3), 286–290. https://doi.org/10.1002/cite.202300048

Studier-Fischer, A., Seidlitz, S., Sellner, J., Bressan, M., Özdemir, B., Ayala, L., Odenthal, J., Knoedler, S., Kowalewski, K., Haney, C. M., Salg, G., Dietrich, M., Kenngott, H., Gockel, I., Hackert, T., Müller-Stich, B. P., Maier-Hein, L., & Nickel, F. (2023). HeiPorSPECTRAL - the heidelberg porcine HyperSPECTRAL imaging dataset of 20 physiological organs. *Scientific Data*, *10*(1), 414. https://doi.org/10.1038/s41597-023-02315-8

T., K., Marwick, B. van, Rittel, M., Ramallo Guevara, C., Wühler, F., Teumer, T., Wängler, B., Hopf, C., & Rädle, M. (2021). Rapid brain structure and tumour margin detection on whole frozen tissue sections by fast multiphotometric mid-infrared scanning. *Scientific Reports*, *11*(1), 11307. https://doi.org/10.1038/s41598-021-90777-4

Wuehler, F. (2025). *HSI-wizard*. https://github.com/BlueSpacePotato/hsi-wizard