Open Innovation for Earth Observation Programmes

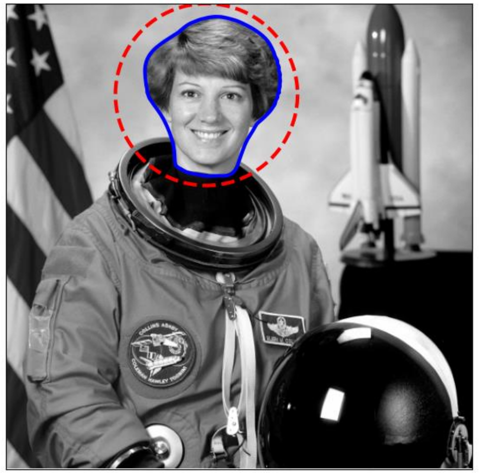
2-4 November 2022 | ESA-ESRIN | Frascati (Rm), Italy

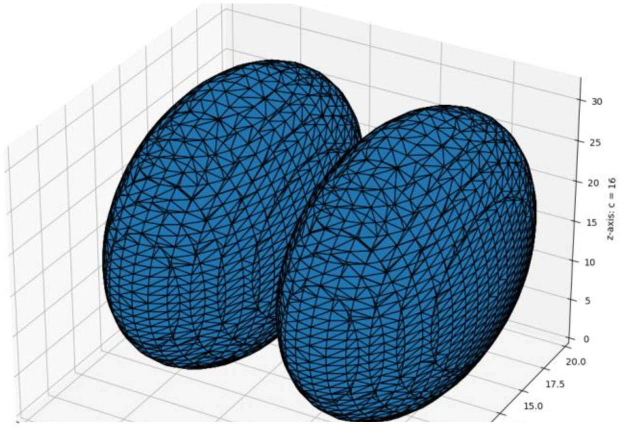
Designing code for collaboration

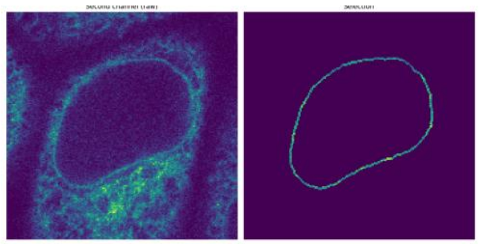
Lars Grüter

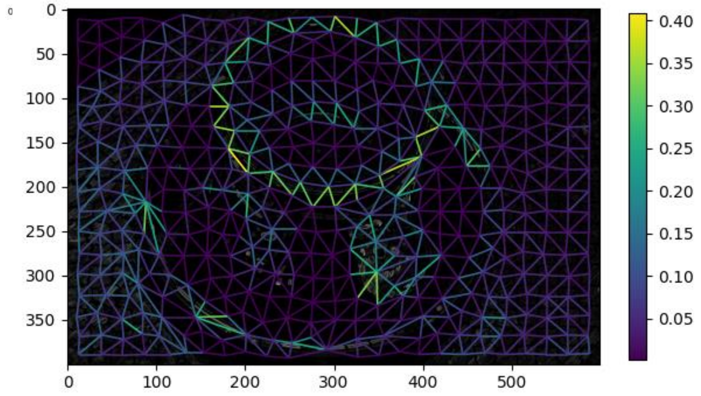
scikit-image, Scientific Python

scikit-image
image processing in python





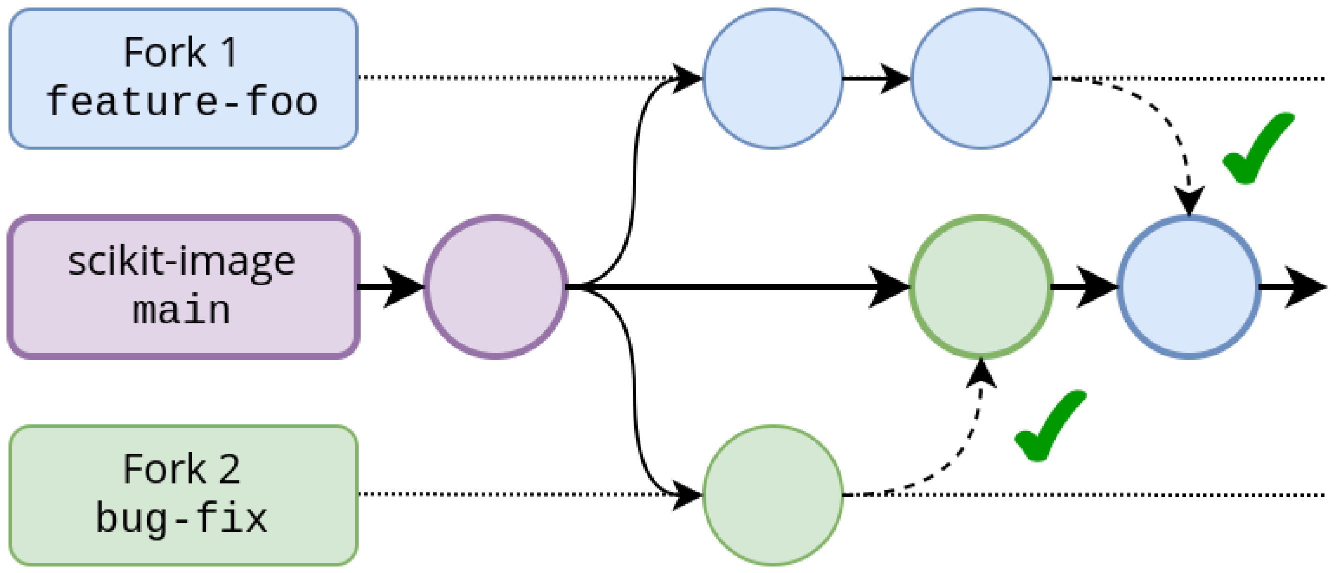




- scikit-image.org, doi.org/10.7717/peerj.453
- A collection of algorithms for image processing
- BSD-3-clause license, not copyleft
- First commit 13 years ago
- 78% Python,
  20% compiled code (C/C++-based)
- ~510 (code) contributors,
  11 core developers / maintainers

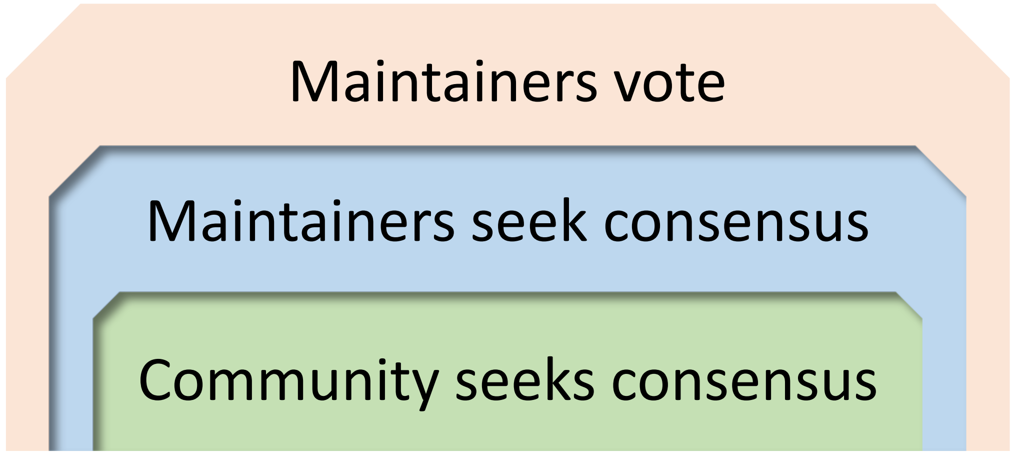
scikit-image's development model

- Fork-based contribution model on GitHub
- Contributors develop on their own feature branches
- Pull requests are reviewed and squashed



Decision making in scikit-image

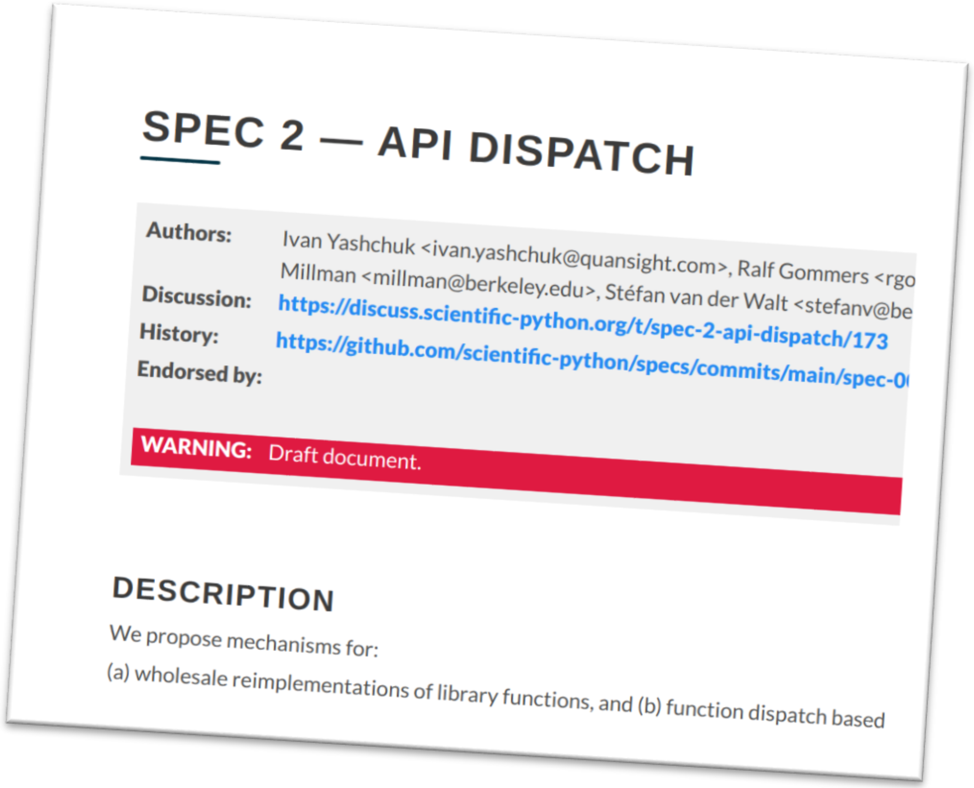
- Consensus is the goal
- Formalized (SKIP 1)
- Layered



- Small changes
  - typos, formatting, documentation, maintenance
  - 1 approving core developer
- Changes affecting the API
  - 2 approving core developers
  - Silent consensus is assumed after ~1 week
- Controversial or critical changes
  - require enhancement proposal (SKIP)

Enhancement Proposals

- Inspired by **P**ython **E**nhancement **P**roposals (PEP)
- Documentation of a change, standard, deprecation, policy, …
- Formal document and process
- Structures & facilitates input from community
- Adoption in the Scientific Python ecosystem:
  - NEP (NumPy), MEP (matplotlib), PDEP (pandas), SKIP (scikit-image), SPEC (Scientific Python), …


SPEC 2 — API DISPATCH

Authors: Ivan Yashchuk <ivan.yashchuk@quansight.com>, Ralf Gommers <rgo
Millman <millman@berkeley.edu>, Stéfan van der Walt <stefanv@be
Discussion: https://discuss.scientific-python.org/t/spec-2-api-dispatch/173
History: https://github.com/scientific-python/specs/commits/main/spec-0
Endorsed by:

WARNING: Draft document.

DESCRIPTION

We propose mechanisms for:
(a) wholesale reimplementations of library functions, and (b) function dispatch based

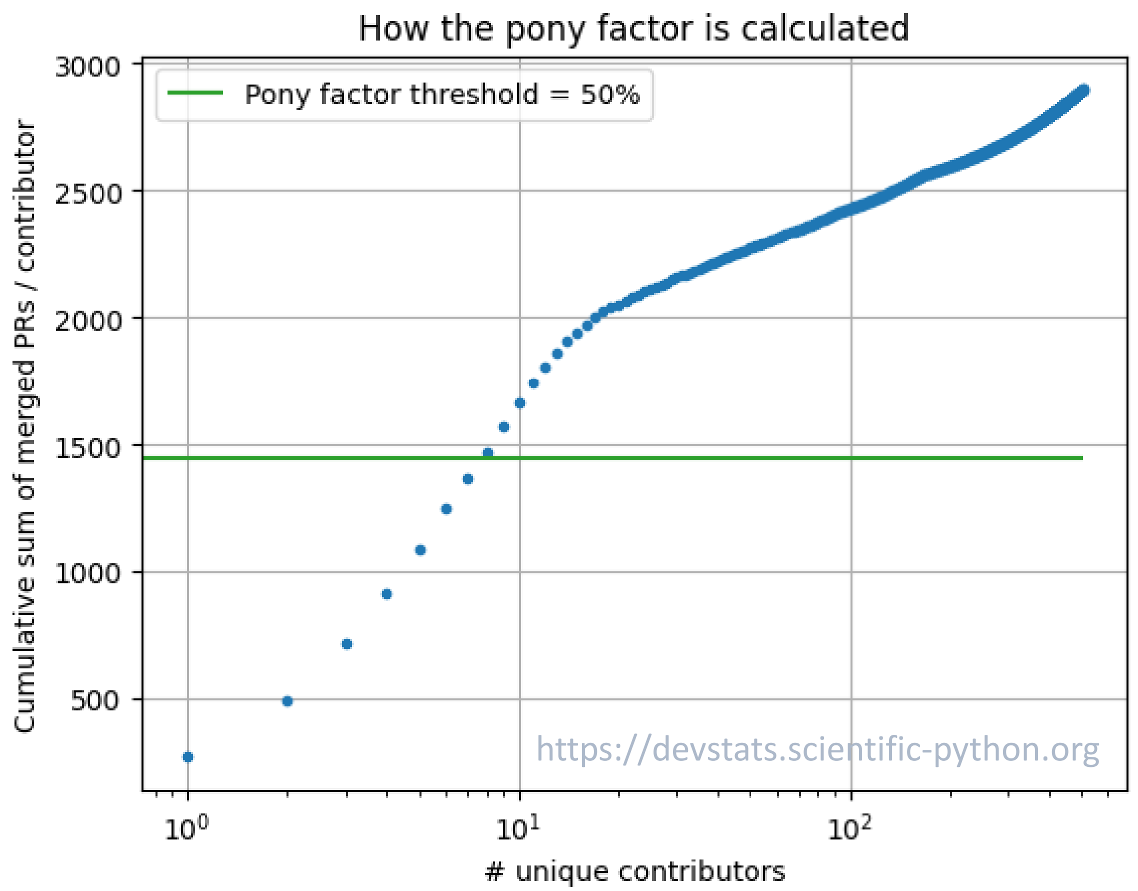
Evolution of the library

- Tried to mimic MATLAB image toolbox
  - Chasing other API / project
  - Made Pythonic code harder to achieve
- Focused on clear and Pythonic API, and documentation
- Planned: update our API principles

What made it a success?

- Popularity of Python in data science
- Excellent interoperability
- Usefulness to enough domains & people
- Community owned and developed

Project challenges



- **67% of PRs** are one-time contributors, Pony factor: 8
- Contributors are domain-experts
  - git, “good practices”, tooling, etc. can be a barrier
- Reviews are a bottleneck
  - Rarely contributed from non-core developers
  - Reduce burden with automation
  - Push small suggestions directly
  - Encourage new maintainers

Transitioning to a new scikit-image API

- Update & cleanup API
  - Some changes are backwards-incompatible and hard to deprecate
- Avoid Python 2 to 3 scenario
- Rejected ideas
  - Follow semantic versioning: release v1.0 with breaking changes (SKIP 3)
  - Switch to new namespace **`skimage2`** with v2.0 (SKIP 4)

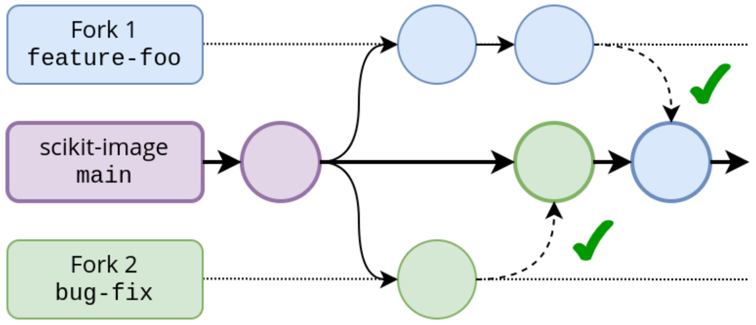
Current plan:

- **`import skimage.v2`**
  - New and updated API
- **`import skimage.v1`**
  - Re-create old API by wrapping v2
  - Slowly deprecate

APIs to encourage collaboration?

- API is big part user communication
  - Don't change your API silently
  - Deprecation policy
- Minimize interpretation of input data
  - You can't foresee applications
  - E.g. prefer 1st & 2nd dimension over row & column
- API patterns in your ecosystem
  - Deviation from that "standard" can be a barrier
- Keep it simple
  - Functions over classes
  - Re-use API patterns, e.g. NumPy's "out" parameter
- Examples, examples, examples, …

How to support open innovation?



- Aim for a fully open research software stack from the start
- Teach and use collaborative tools & practices
  - Community owned and developed
  - E.g. VCS, branching & review based development model, lint & test automation
- Contribute back to what you use
- Support students & employees who want to develop these libraries

Scientific Python

- scientific-python.org
- Community and project to coordinate the ecosystem and grow the community
- Goals
  - cross-project policies, SPECs, common infrastructure & tools
  - topic-focused developer events to uncover needs
  - help with acquiring funding, writing grant proposals
- Other projects: pyopensci.org
  - Peer-review of scientific Python packages to promote reproducible

Thank you!

Contact

lagru@mailbox.org