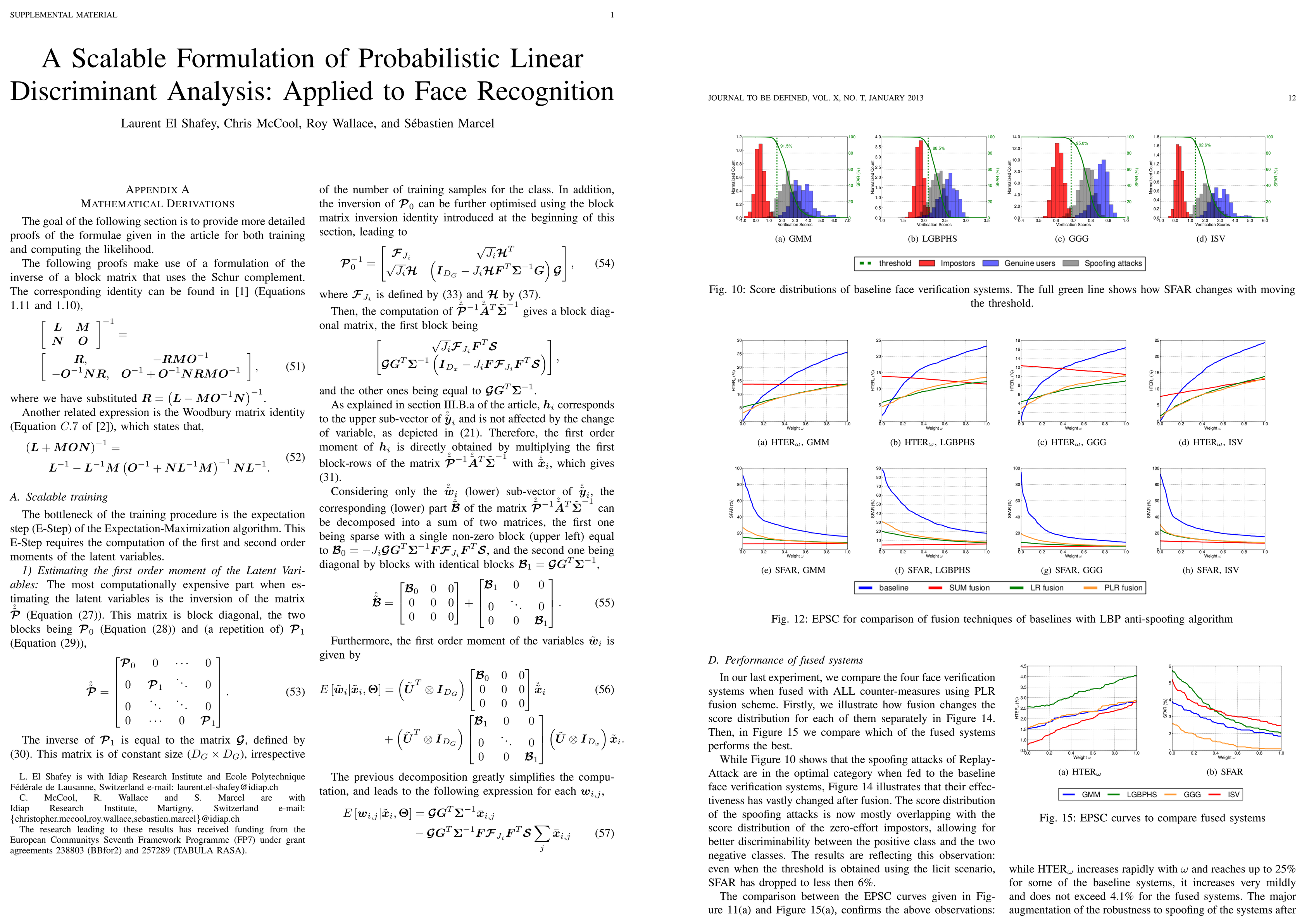
Continuously Reproducing Toolchains in Pattern Recognition and Machine Learning Experiments

André Anjos, Manuel Günther, Tiago de Freitas Pereira, Pavel Korshunov, Amir Mohammadi, and Sébastien Marcel
Idiap Research Institute, Switzerland

Start here: https://www.idiap.ch/software/bob/

Typical Paper


A Scalable Formulation of Probabilistic Linear Discriminant Analysis: Applied to Face Recognition
Laurent El Shafey, Chris McCool, Roy Wallace, and Sébastien Marcel

- Pattern recognition and machine learning research work often contains experimental results on real-world data, which corroborates hypotheses and provides a canvas for the development and comparison of new ideas.
- Reproducibility is often overlooked.
- Scientific experiments often consist of many steps and parameters that are difficult to report.
- In our view, **reproducible research work should be repeatable, shareable, extensible and stable**.
- We investigate the implications of this key properties of RR work and the requirements for frameworks to fullfil them.

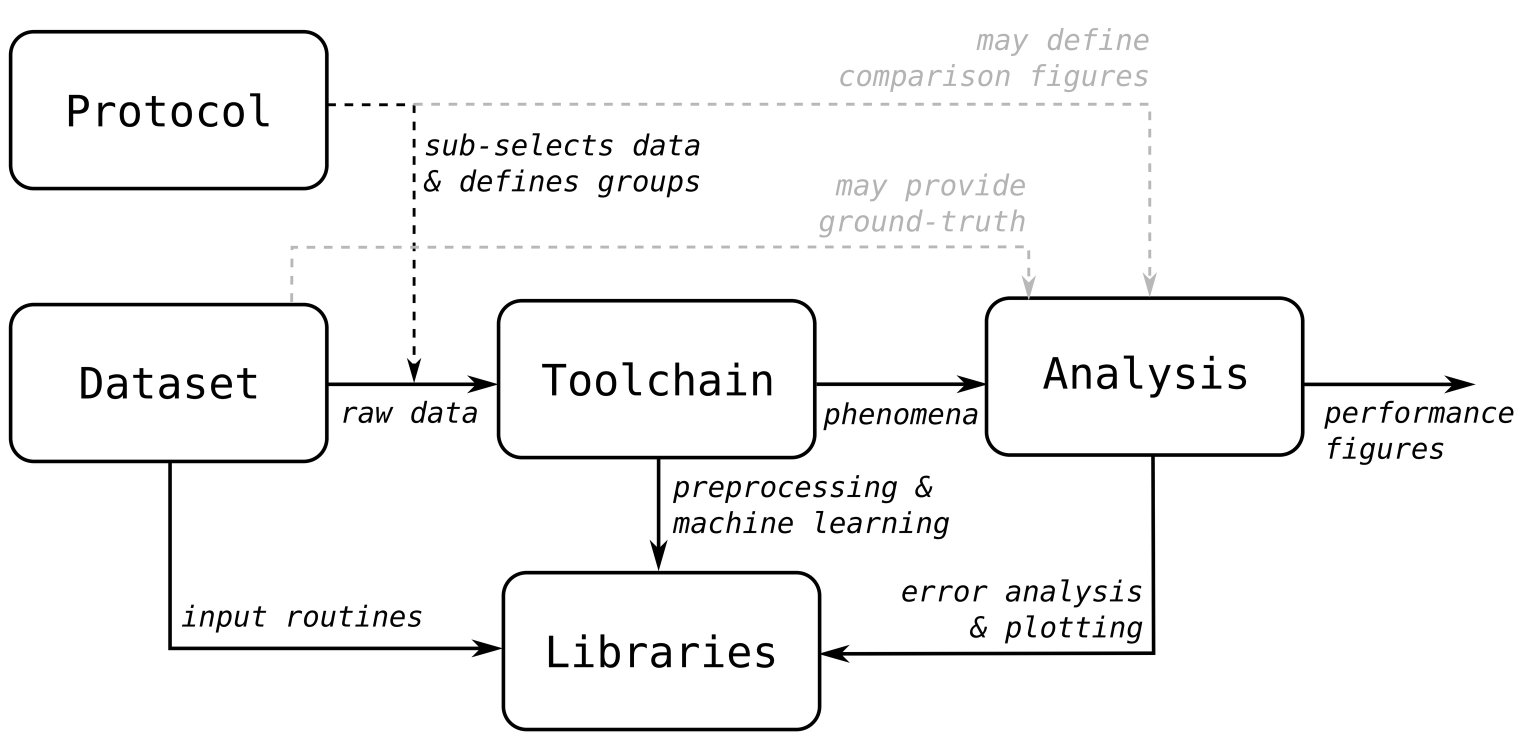
Definition

We consider a published paper to be reproducible if it is:

- **Repeatable**: It should be **possible to re-run** experiments declared in a report and obtain the same results, given the same selection of data, software, hyper-parameters, and evaluation protocol.
- **Shareable**: It should be **possible to share** the material (**data and code**) with others, so they can repeat experiments declared in a report.
- **Extensible**: It should be **easy to implement new research directions or evaluate new conditions** on existing experimental infrastructure.
- **Stable**: The **repeatability through time should be guaranteed**, on a best-effort basis.

Framework

Extensibility is an important aspect of reproducibily as *today's state of the art will eventually become tomorrow's baselines.* A "workflow" needs to be in place:



Tools

Bare software frameworks are seldom successful in the long run without committed people and accompanying tools:

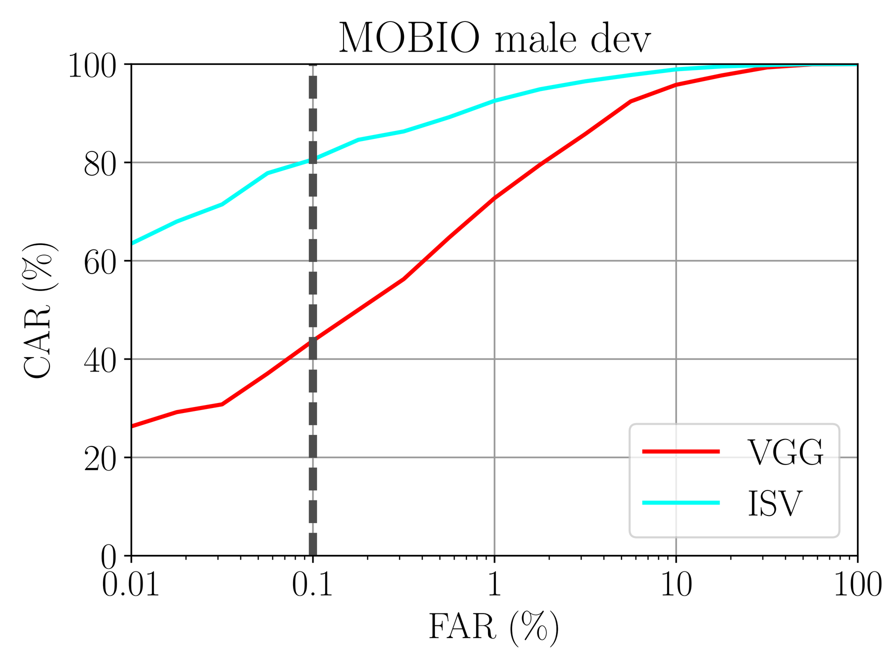
- Version control
- Unit testing and quality control
- Packaging and deployment
- Documentation

Use-case Analysis: Reproducible face recognition experiments

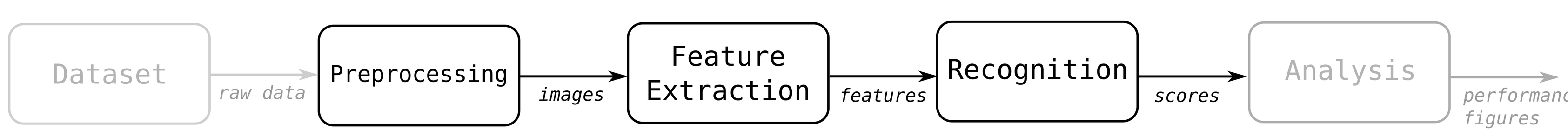
- Get data: https://www.idiap.ch/dataset/mobio
- Get code: https://gitlab.idiap.ch/bob/bob.paper.icml2017

- Install: Trivially done on any Linux/Mac OSX (64-bits) with Conda
- Run: Single command line to launch the multi-stage pipeline. **Automatic** parallelization on SGE or local machine supported

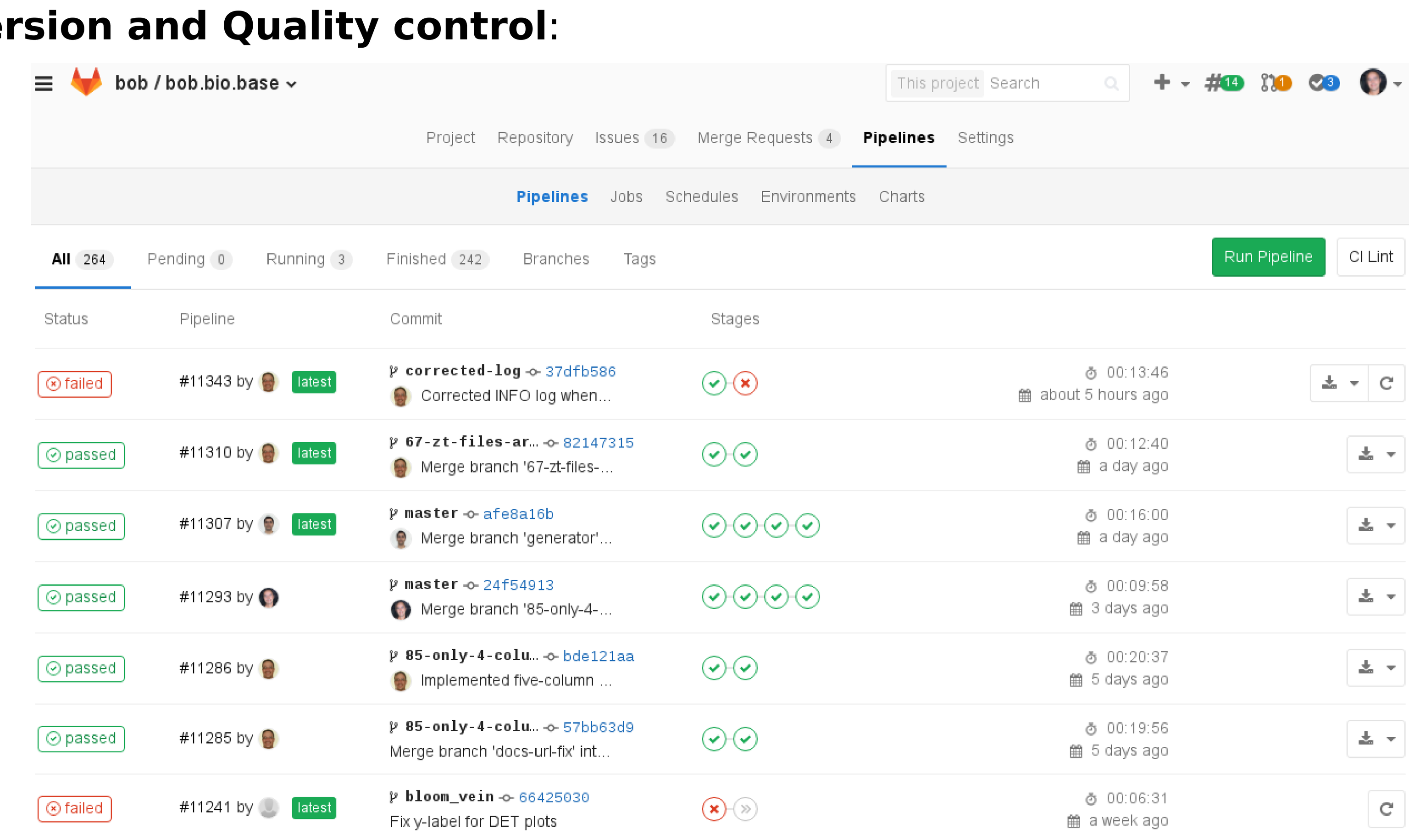
Results:



Toolchain implemented by framework:



Version and Quality control:



Conclusions

1. To guarantee long-time reproducibility, scientific **work must be repeatable, shareable, extensible and stable over time**.
2. Committing to continuous reproducibility implies the creation and maintenance of a **re-usable framework**.
3. Efforts can be reduced if **researchers are unified** into standardized frameworks.
4. **Bob** is an *example*; we expect others to appear in the future.
5. Web-based frameworks such as the **BEAT** platform may provide a simpler mechanism of reproducibily while implementing "social" research and development.

LaTeX TikZposter