

Reproducible Research with Image Processing On Line (IPOL)

Miguel Colom
<http://mcolom.info>

CMLA, ENS-Cachan



What is Reproducible Research?

- It redefines the **result of the research**.
- It is **not just a paper!**
- It is:
 - The research **article**
 - The **computational facility** which recreates the results given the input data
 - The **source code**
 - The **data**



Why Reproducible Research is Needed?

- The **results** in the published paper may be **wrong**
- The results in the published paper may not be **generalizable** (only work with that particular images)
- The procedure described in the paper may be **inaccurate**, **vague**, or **incomplete**.
- Published paper usually **don't have space enough to describe all details** (which are need to implement a method!). For example: all parameters in algo subalgorithm. Pre/post processing steps?
- The **source code** of the method **may not be available** or the author may refuse to make it public (because of a patent, for example)
- A paper doesn't allow to **compare** results with other methods



Why Reproducible Research is Needed?

- A quote by David Donoho:
 - *“An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.”*
- Other people working on Reproducible Research:
 - Claerbout, Donoho, Stodden, Vandewalle, ...



Reproducibility and Repeatibility

- **Reproducibility**: the ability to reproduce the same results of a reference experiment.
- **Repeatibility**: the ability to repeat the experiment as many time as needed.
- Both are needed in Reproducible Research.
- Always easy to obtain? **No!** For example, in Biology:
 - The experiment may not be reproducible since it is impossible to know all the exact conditions
 - Repeating a biological experiment may take months or even years! (for example, studying the effect of a gene mutation in mice generations).
- **Do not trust** the authors. Just **try** and **evaluate** the algorithm **yourself**.



Reproducibility and Repeatibility in Image Processing

- **Similar to Biology? NO!**
- Since the algorithms are completely **well defined** using proper mathematical descriptions, it is possible to **reproduce** and **repeat** all experiments.



Reproducibility and Repeatability in Image Processing

- Then, **why not always done?**
- Several reasons:
 - The **source code** of the authors is **not of enough quality** (readability, correctness, usability, style) and **they don't want to make it public**.
 - The source code **might not do exactly what is described** in the paper.
 - It **takes more time** to properly design a good demonstrator than writing a paper
 - Not only the **paper** will be peer-reviewed, but also the source code of the **algorithm** (checking that it does exactly what is described in the paper), and the **demonstrator** code too.



Summary

- **software** is part of the **research work**
- software needs be **published**
 - with a **review process**
 - with **quality criteria**
- In the particular case of IPOL we have:
 - **detailed algorithms**
 - **verified** and **usable code**
 - instant test **demos**



Benefits of having a RR demonstrator

- **Comparisons** with other methods **easier**
- The results can be **trusted**, since they passed an exigent **peer-review** process
- The number of **citations** of the article grow when the scientific community can **reproduce** and **compare** the results of the methods (look at IPOL articles)
- More **visibility** to the paper
- **Convincing** results

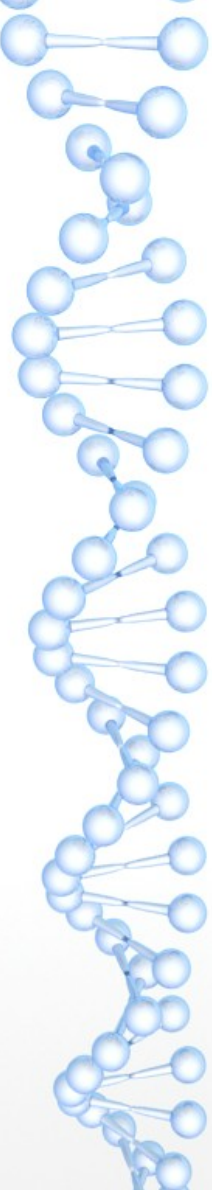


Disadvantages for the author

- It takes **much more time** than writing a regular article in a journal
- The submitted software must comply with some **strict software guidelines**: readability, documented, portable, standard
- The number of accepted **libraries** is small (to ensure long-term compatibility)
- The **peer-review** process is in general more **exigent** and thorough compared to a ordinary journal

What is IPOL?

- An Image Processing (OnLine) **journal**
- Each article: **source code** + **PDF article** + **demo**
- All **peer reviewed**
- **Open source** software



```
document.item = el->FirstChildElement->DataElement->ElementDesc albedo;  
  
std::string sp_name = item->Attribute("sp_name");  
std::string spritename = item->Attribute("spritename");  
  
boost::lexical_cast<float>(&albedo, item->Attribute("albedo"));  
boost::lexical_cast<float>(&sp_name, item->Attribute("sp_name"));  
boost::lexical_cast<float>(&spritename, item->Attribute("spritename"));  
std::string layer = "50";  
if (item->Attribute("layer") != NULL)  
    layer = boost::lexical_cast<string>(item->Attribute("layer"));  
  
albedo.name = sp_name;  
albedo.spritename = spritename;  
albedo.layer = layer;
```

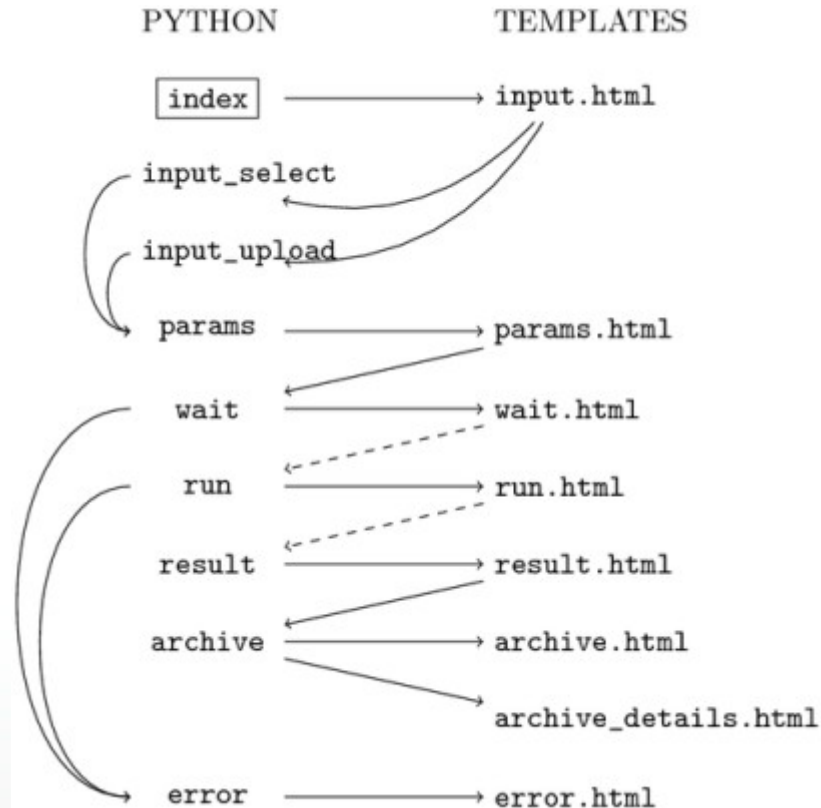




IPOL demos workflow

- 1) Users **choose** an **input image** (or **upload** their own)
- 2) The user **sets the parameters** of that image and click on the “**Run**” button.
- 3) The **algorithm** (C/C++, Matlab, ...) is **executed** and the **results** are **shown** in the browser.
- 4) The results are **archived** if the input image was uploaded.

IPOL demos workflow





IPOL workshop

How to create an IPOL demo