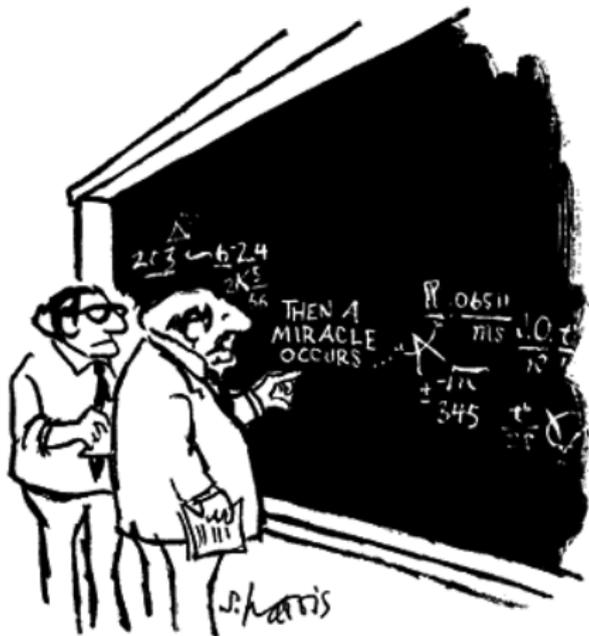


Reproducible Biometrics Evaluation and Testing with the BEAT Platform

André Anjos, Philip Abbet and Sébastien Marcel

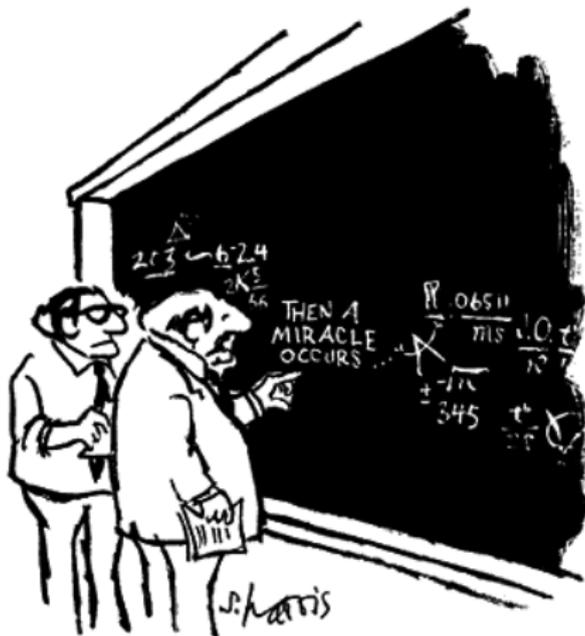


April, 1st. 2014



"I think you should be more explicit here in step two."

How many times?



"I think you should be more explicit here in step two."

*Crossed a publication and openly decided to **ignore it because it would be too hard to apply** those doubtful results on your research?*

*Worked day and night to **incorporate some results** on your own work but:*

- ▶ There were **untold parameters** that needed adjustment and you couldn't get hold of them?
- ▶ Realized the proposed algorithm **worked only on the specific data** shown at the original paper?
- ▶ Realized that something did **not quite add up** in the end?

What your research supposedly looks like:

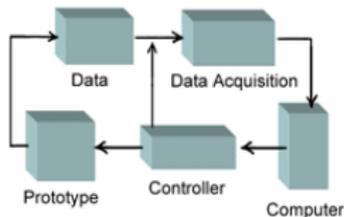


Figure 1. Experimental Diagram

What your research *actually* looks like:

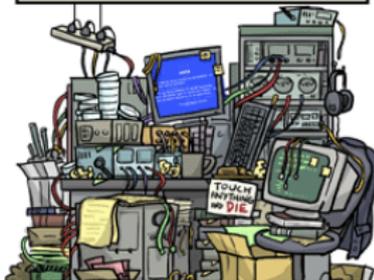


Figure 2. Experimental Mess

WWW.PHDCOMICS.COM JORGE CHAM © 2008

*Had a **new student to take over** the work from another student that left and had to start from scratch - months into programming to make things work again?*

What your research supposedly looks like:

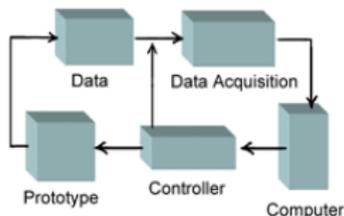


Figure 1. Experimental Diagram

What your research *actually* looks like:

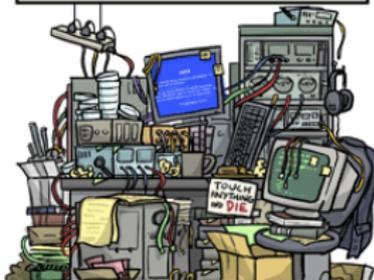


Figure 2. Experimental Mess

WWW.PHDCOMICS.COM JORGE CHAN © 2008

*Would have liked to **replay to someone about your work**, but you couldn't really remember all details when you first made it work? Or you **could not make it work at all**?*

Enter “Reproducible Research” (RR)¹

One term that aggregates work comprising of:

- ▶ a **paper**, that describe your work in all relevant details
- ▶ **code** to reproduce all results
- ▶ **data** required to reproduce the results
- ▶ **instructions**, on how to apply the *code* on the *data* to replicate the results on the *paper*.

¹<http://reproducibleresearch.net>

Levels of Reproducibility²

With respect to an independent researcher:

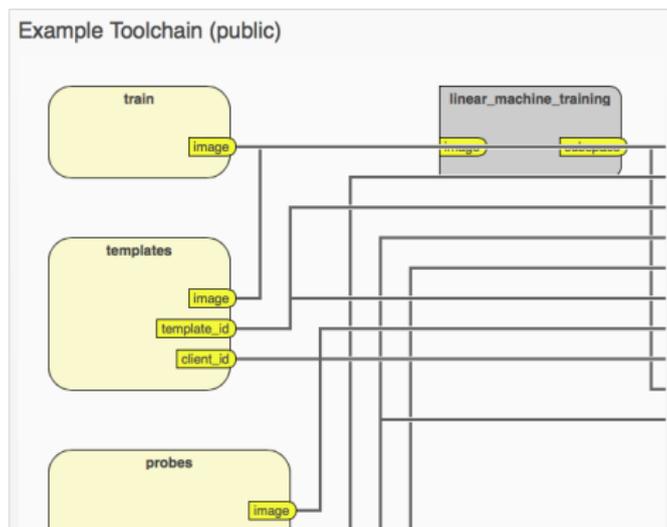
0. Irreproducible
1. Apparently unable to reproduce
2. Reproducible, with extreme effort (> 1 month)
3. Reproducible, with considerable effort (> 1 week)
4. Easily reproducible (~ 15 min.), but requires proprietary software (e.g. Matlab)
5. Easily reproducible (~ 15 min.), only free software
6. **Easily reproducible (\sim seconds), only requires a web-browser**

²*Reproducible Research in Signal Processing: What, why and how*, Vandewalle, Kovacevic and Vetterli, 2012

Taking RR to the next level: the BEAT platform



The BEAT (*Biometrics Evaluation and Testing*) platform provides easy online access to experimentation and testing for Biometrics. You define what data and modules you would like to use, we make sure the system runs and provides you with a result. Data from different experiments can be easily compared and searched.



System status

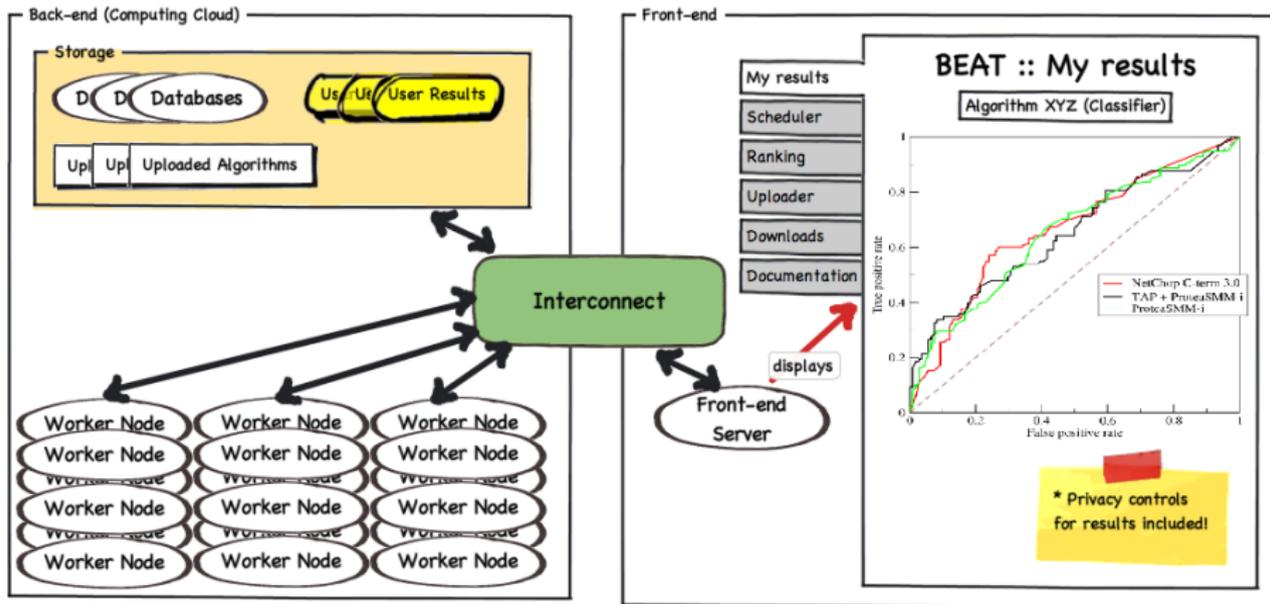
Service is: **operational**

- 17 users
- 4 databases
- 39 toolchains
- 32 dataformats
- 83 algorithms
- 1 experiment finished
- 63.3h of CPU processing time
- 134GB of memory used

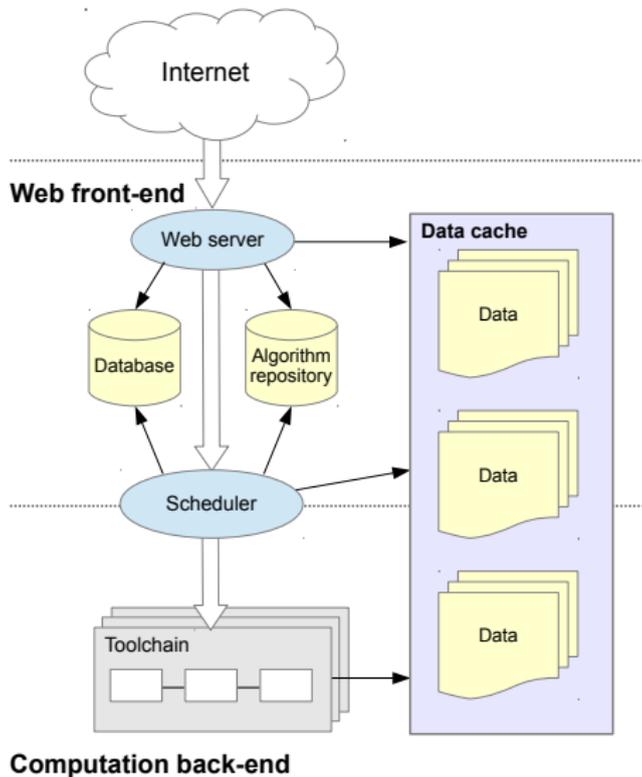
Taking RR to the next level: the BEAT platform

- ▶ Web platform: always **accessible**, no need to install extra software
- ▶ Intuitive: **graphically** connect blocks to run experiments
- ▶ Social: **engagement** gets you **more processing power**
- ▶ Productive: search the state-of-the-art by any filtering criteria
- ▶ Private:
 - ▶ No need to handle large-scale databases
 - ▶ Can run on *un*-distributable data (e.g. forensic databases)
- ▶ Assurance
 - ▶ **fair (reproducible) evaluations** of algorithms
 - ▶ **online certifications** for all produced results
- ▶ Free: build on **open-source** software and standards

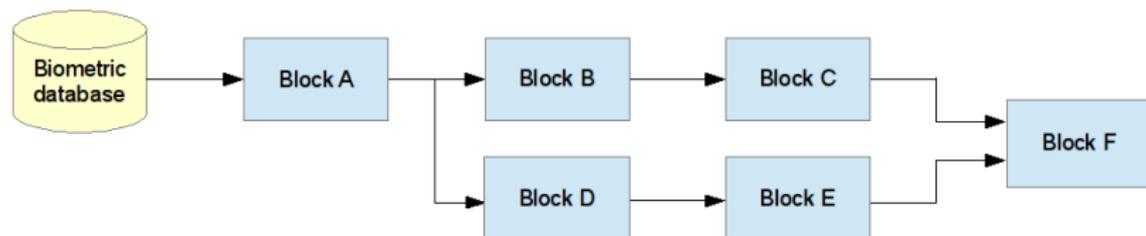
BEAT Platform: Original Concept



Architectural Choice

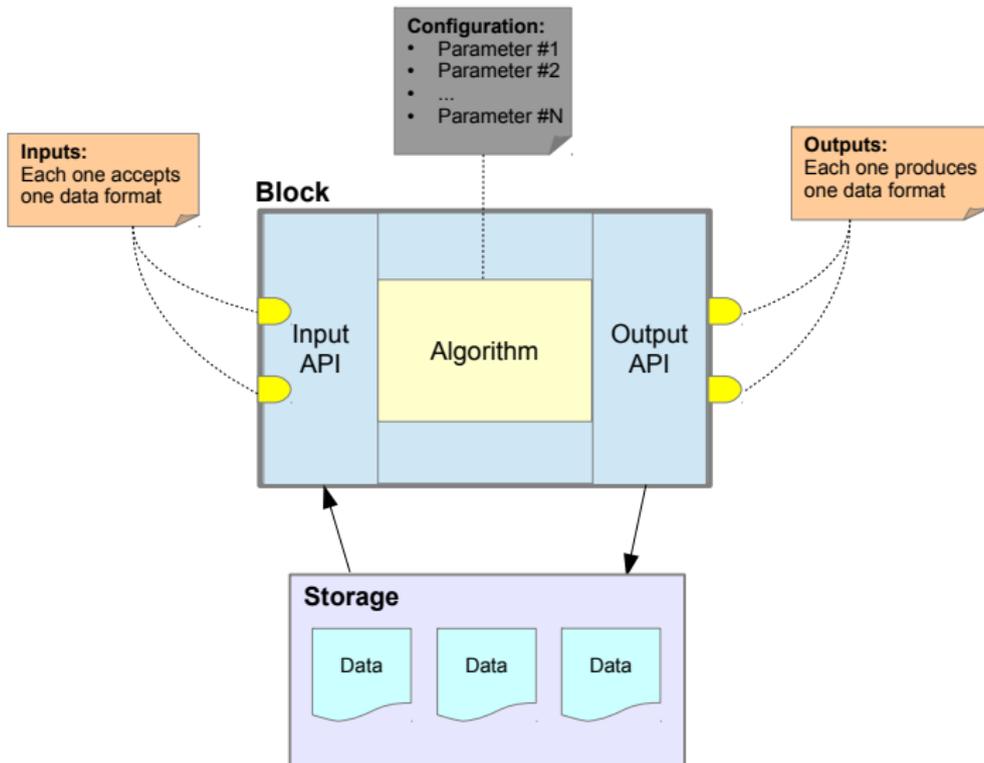


A key concept in experimentation is the idea of *Toolchains*.



Blocks

Toolchains are composed of interconnections of *Blocks*



Blocks: Features

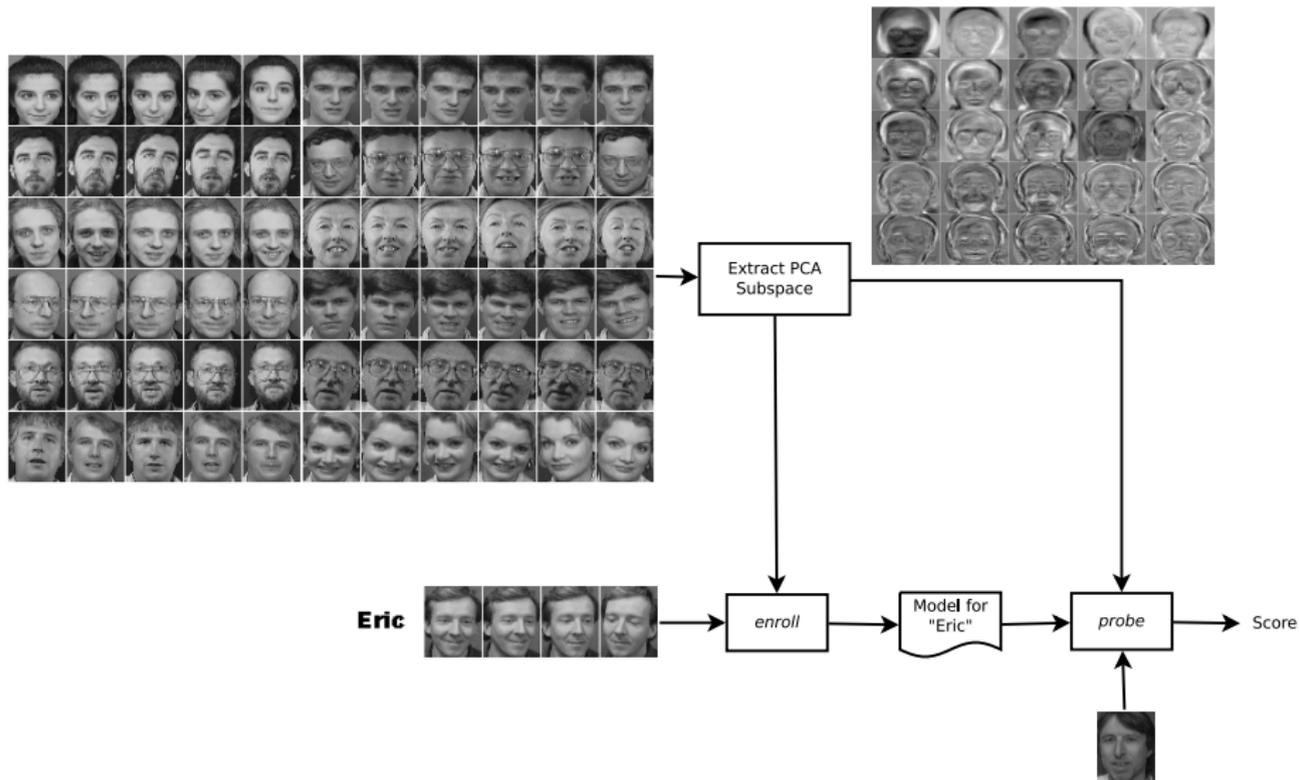
- ▶ Blocks can run **arbitrary** code
 - ▶ Potential to implement back-ends to support compiled code, any scripting language
 - ▶ We picked Python as a default back-end to start the project
 - ▶ The platform itself, is also written in Python
- ▶ Blocks typically have inputs and outputs
- ▶ Data transmitted from block to block is formally defined (*Data Formats*)
- ▶ Database blocks are special - they only have outputs, provided by administrators of the platform
- ▶ Result blocks don't output to any other block

BEAT is designed as an *opt-in* platform

- ▶ All your actions and results are kept private until you choose to change visibility
- ▶ Once you change the visibility of any item, associated items are *frozen* so your results are kept reproducible
- ▶ If anything changes on the underlying platform (OS, packages, toolchains, databases), your results are *outdated* - but still valid

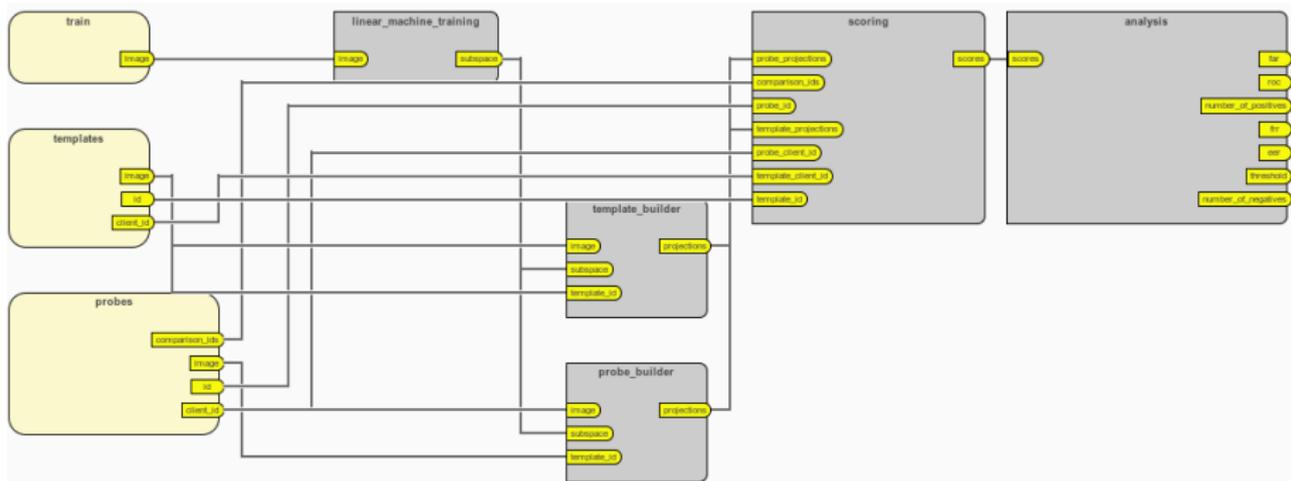
Toolchain Example: trivial Eigen-faces

Simple: no evaluation (test) set, threshold *a posteriori*



Toolchain Example: trivial Eigen-faces

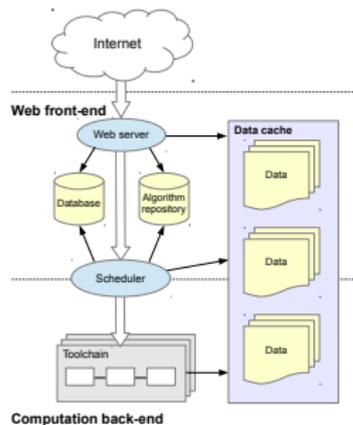
Translation as a BEAT toolchain



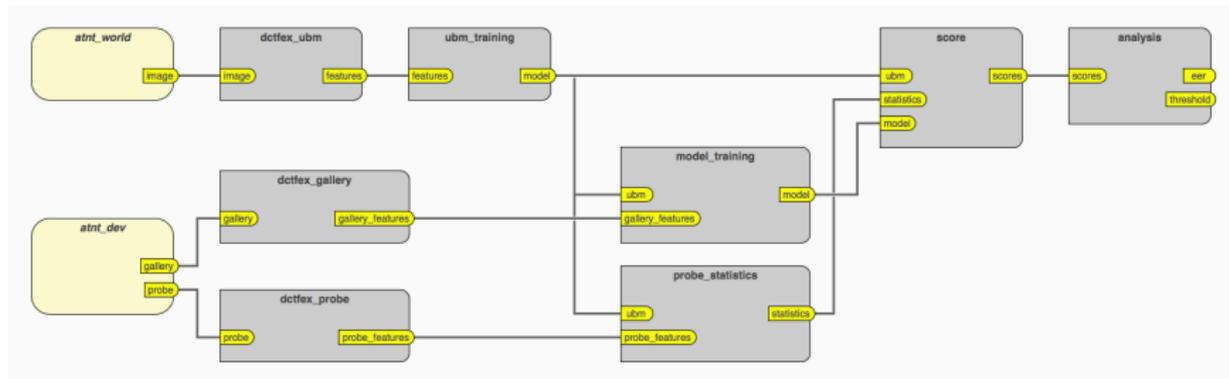
- ▶ All relations are explicit: data and algorithms
- ▶ *Expected* database is divided into three-blocks leading to the training data, and validation data (split into gallery and probing sets)
- ▶ Notice that a Toolchain only defines the connections and data types into and out of Blocks
- ▶ Check for yourself: <https://www.beat-eu.org/platform/toolchains/bob/eigenface/>

Caching

- ▶ BEAT keeps track of data transmitted through all stages of the processing
- ▶ It caches the data in a large disk array (~10 Tb)
- ▶ The cache is invalidated automatically when things change:
 - ▶ Operating System or installed packages are updated
 - ▶ Toolchain changes
 - ▶ Database version changes
- ▶ One cached item is valid for a specific combination of all of the above
- ▶ If a cached item is available, it is used to speed-up processing.



Proof of Concept: DCT+GMM Face Recognition



(Click to launch external video)

Attestation (“Certification”)

An attestation mechanism is available in the platform.

- ▶ Allow 3rd. party verification of results obtained with a given configuration
- ▶ Allow for a scientific *review* process to take place in confidentiality

https://www.beat-eu.org/platform/attestations/961981311/

BEAT Search... bob Log out

Attestation #961981311

Unlock the attestation

The BEAT platform attests that the following results were obtained by an experiment performed on our servers, and that we kept all the details needed to reproduce them (toolchain, algorithms and parameters).

Results	
analysis.eer	0.06
analysis.far	0.06
analysis.frr	0.06
analysis.number_of_negatives	1900
analysis.number_of_positives	100
analysis.roc	

ROC

FAR (%)

The ROC plot shows a curve that starts at (0, 100) and drops sharply to near 0% FAR, indicating high performance in minimizing false acceptances.

Forking and Modification

- ▶ One of the main mechanisms for *sharing* is the ability to *fork* a toolchain
- ▶ By *forking*, you get a new toolchain that shares all properties of a given toolchain, except the ownership
- ▶ You can modify only your own forks for toolchains

User Page

One of the jobs of BEAT is to keep track of details for all experiments posted

https://www.beat-eu.org/platform/user/bob/

BEAT Search... bob Log out

Experiments Toolchains Algorithms Data formats

Find an experiment... Sharing: All Public Private New

Status: All Done Running Scheduled

Status	Date	Database	Toolchain	Label	Duration	CPU time	I/O
Done	Oct 23, 2013	atnt	bob/atnt_eigenface	20_components	3h18	20.3s	69.8MB / 1.92MB
Done	Oct 23, 2013	atnt	bob/atnt_eigenface	5_components	3h18	20.1s	67.4MB / 740kB
Done	Oct 23, 2013	atnt	bob/atnt_eigenface	10_components	3h18	19.9s	68.2MB / 1.12MB
Done	Oct 23, 2013	atnt	bob/atnt_eigenface	50_components	3h18	21.0s	74.5MB / 4.33MB
Failed	Oct 23, 2013	xm2vts_ip1	bob/xm2vts_ip1_dct_ubmgmm	45dct_100gaussians	51h29	15.5min	15.4GB / 15.6GB
Done	Oct 23, 2013	xm2vts_ip1	bob/xm2vts_ip1_dct_ubmgmm	5dct_50gaussians	3h39	21.6min	15.9GB / 11.8GB
Done	Oct 23, 2013	xm2vts_ip1	bob/xm2vts_ip1_dct_ubmgmm	2dct_2gaussians	3h39	15.6min	15.6GB / 11.5GB
Done	Oct 23, 2013	banca_P	bob/banca_P_dct_ubmgmm	5dct_50gaussians	3h39	29.0min	26.5GB / 19.9GB
Done	Oct 23, 2013	atnt	bob/atnt_dct_ubmgmm	45dct_100gaussians	3h19	31.4min	1.20GB / 1.14GB
Done	Oct 23, 2013	atnt	bob/atnt_dct_ubmgmm	20dct_80gaussians	3h19	15.8min	601MB / 534MB
Done	Oct 23, 2013	atnt	bob/atnt_dct_ubmgmm	5dct_50gaussians	3h19	6.10min	223MB / 157MB

Bob DemoGuy
bob
Joined on Jul 19, 2013

322 reputation **3 / 3** free slots

Experiments: 11 (@ 11)
Toolchains: 4 (@ 4)
Algorithms: 16 (@ 10)
Data formats: 12

Your connections

- JohnDoe shared his algorithm **lbp** with you (just now)
- Booker open-sourced his algorithm **dct** (5 days ago)
- Smithy +1'd your algorithm **gmm_modelling** (8 days ago)
- bradc6 +1'd your answer to "How to generate a line chart?" (8 days ago)

Omni-search

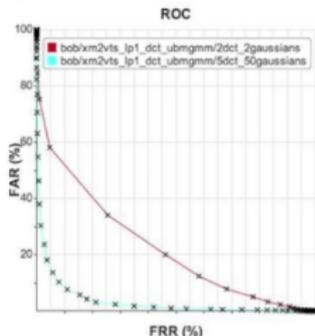
You can use the Omni-search bar (on the top) to type-in query strings.



Results on database *xm2vts_lp1* (2)

Toolchain	User	Label	dev_eer	eval_far	eval_frr	eval_number_of_negatives	eval_number_of_positives	dev_threshold	eval_hter
🔒	bob/xm2vts_lp1_dct_ubmgmm	2dct_2gaussians	0.322	0.32	0.302	191600	400	30.085	0.311
		5dct_50gaussians	0.095	0.091	0.0925	191600	400	283.699	0.0918

eval_roc



Results on database *atntf* (3)

Toolchain	User	Label	eer	far	frr	number_of_negatives	number_of_positives	threshold	eval_number_of_positives
🔒	bob/atntf_dct_ubmgmm	20dct_80gaussians	0.04	0.04	0.04	1900	100	4305.547	🌀
		45dct_100gaussians	0.07	0.07	0.07	1900	100	6720.412	🌀

Open for pre-registration (now)

Pre-registered users will benefit from early platform access when the service becomes available.

<https://www.beat-eu.org/platform/preregister/>

- ▶ In-browser graphical toolchain and code editors
- ▶ More social features: notifications, reputation system, discussion forum
- ▶ Better search result categorization. Output search results into data to re-use on your publications
- ▶ Full parallelization support
- ▶ Initial Hardware commissioning (end of April/2014):
 - ▶ 120 dedicated processing cores with 8 Gb RAM per core
 - ▶ 20 Tb of cache
 - ▶ 10 Gb/s link between cache and processing nodes

Acknowledgement for support

European Commission



Biometrics Evaluation and Testing (BEAT)

<http://www.beat-eu.org>

Swiss State of Wallis



Swiss center for biometrics research and testing

<http://www.biometrics-center.ch>