Paris-Saclay
Center for Data Science

Automated training of models for ML challenges at CCIN2P3

Alexandre Boucaud
aboucaud@lal.in2p3.fr
aboucaud
alxbcd

Paris-Saclay Center for Data Science

A multi-disciplinary initiative, building interfaces, matching people, helping them launching projects

345 affiliated researchers, 50 laboratories

Biology & bioinformatics
IBISC/UEvry
LRI/UPSud
Hepatinov
CESP/UPSud-UVSQ-Inserm
IGM-I2BC/UPSud
MIA/Agro
MIAj-MIG/INRA
LMAS/Centrale

Chemistry
EA4041/UPSud

Earth sciences
LATMOS/UVSQ
GEOPS/UPSud
IPSL/UVSQ
LSCE/UVSQ
LMD/Polytechnique

Economy
LM/ENSAE
RITM/UPSud
LFA/ENSAE

Neuroscience
UNICOG/Inserm
U1000/Inserm
NeuroSpin/CEA

Particle physics astrophysics & cosmology
LPP/Polytechnique
DMPH/ONERA
CosmoStat/CEA
IAS/UPSud
AIM/CEA
LAL/UPSud

Machine learning
LRI/UPSud
LTCI/Telecom
CMLA/Cachan
LS/ENSAE
LIX/Polytechnique
MIA/Agro
CMA/Polytechnique
LSS/Supélec
CVN/Centrale
LMAS/Centrale
DTIM/ONERA
IBISC/UEvry
LIST/CEA

Visualization
INRIA
LIMSI

Signal processing
LTCI/Telecom
CMA/Polytechnique
CVN/Centrale
LSS/Supélec
CMLA/Cachan
LIMSI
DTIM/ONERA

Statistics
LMO/UPSud
LS/ENSAE
LSS/Supélec
CMA/Polytechnique
LMAS/Centrale
MIA/AgroParisTech

RAMP

1. Collaborative prototyping
2. Teaching aid
3. Data science process management

RAMP

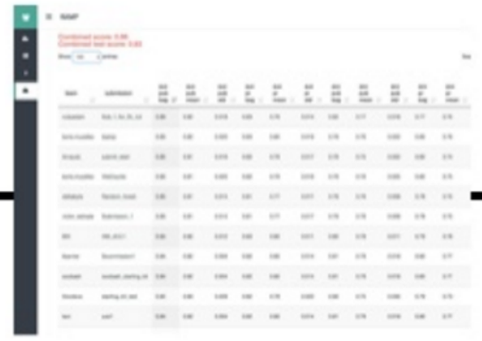
Rapid Analytics and Model Prototyping

starting kit

a baseline solution

- usually built by an expert
- explains the domain problem, variables and the objectives
- enable other data scientists to become operational

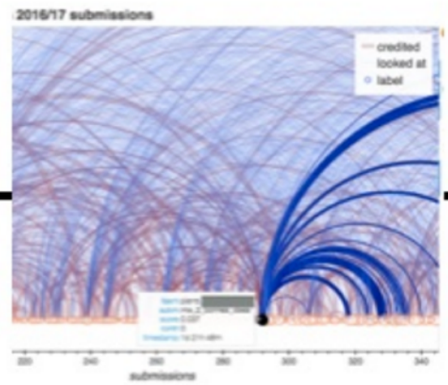
competition



a leaderboard

- gamifies model development
- boosts effort & exploration
- creates a diversity of models
- enhance problem understanding

collaboration

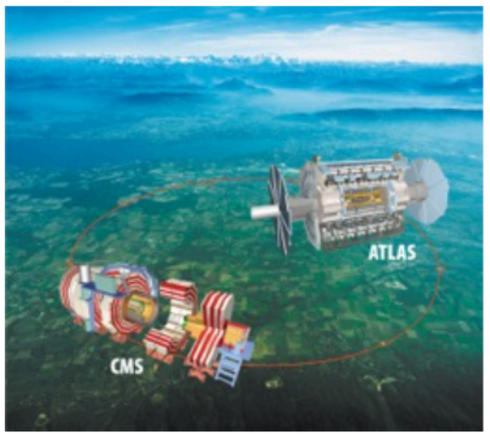


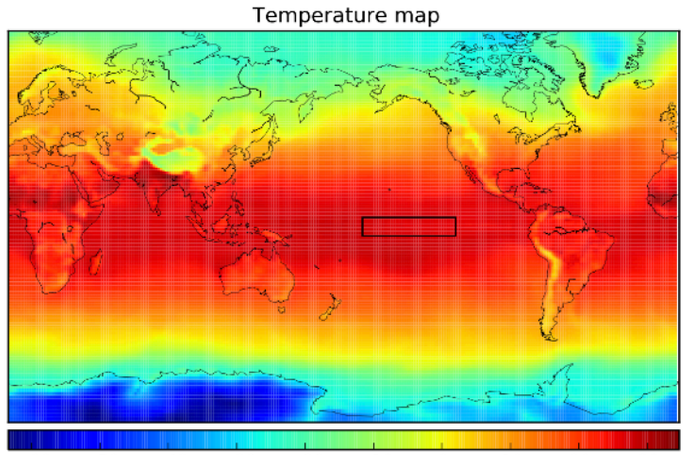
propagating best ideas

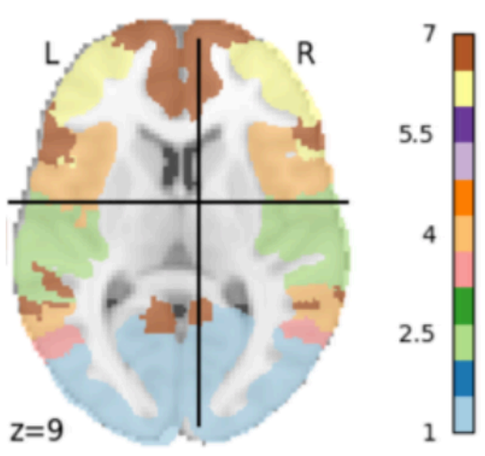
- all models are open
- a wealth of information becomes accessible
- ideas & models are combined
- synergy & learning & creativity

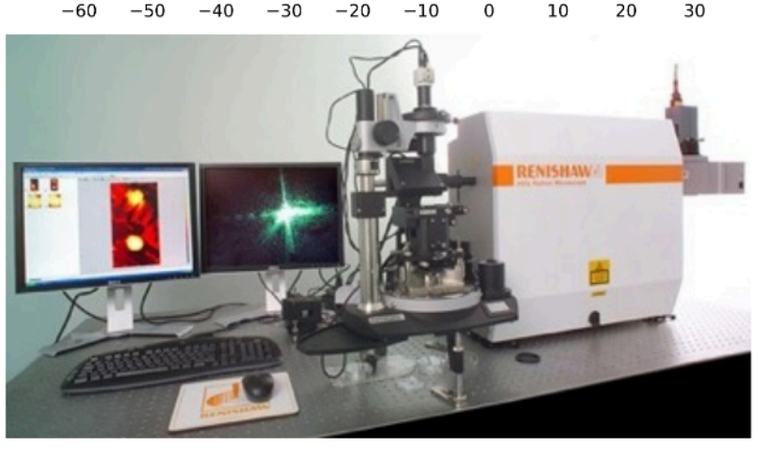
sign up at **www.ramp.studio**

Diversity of themes and projects



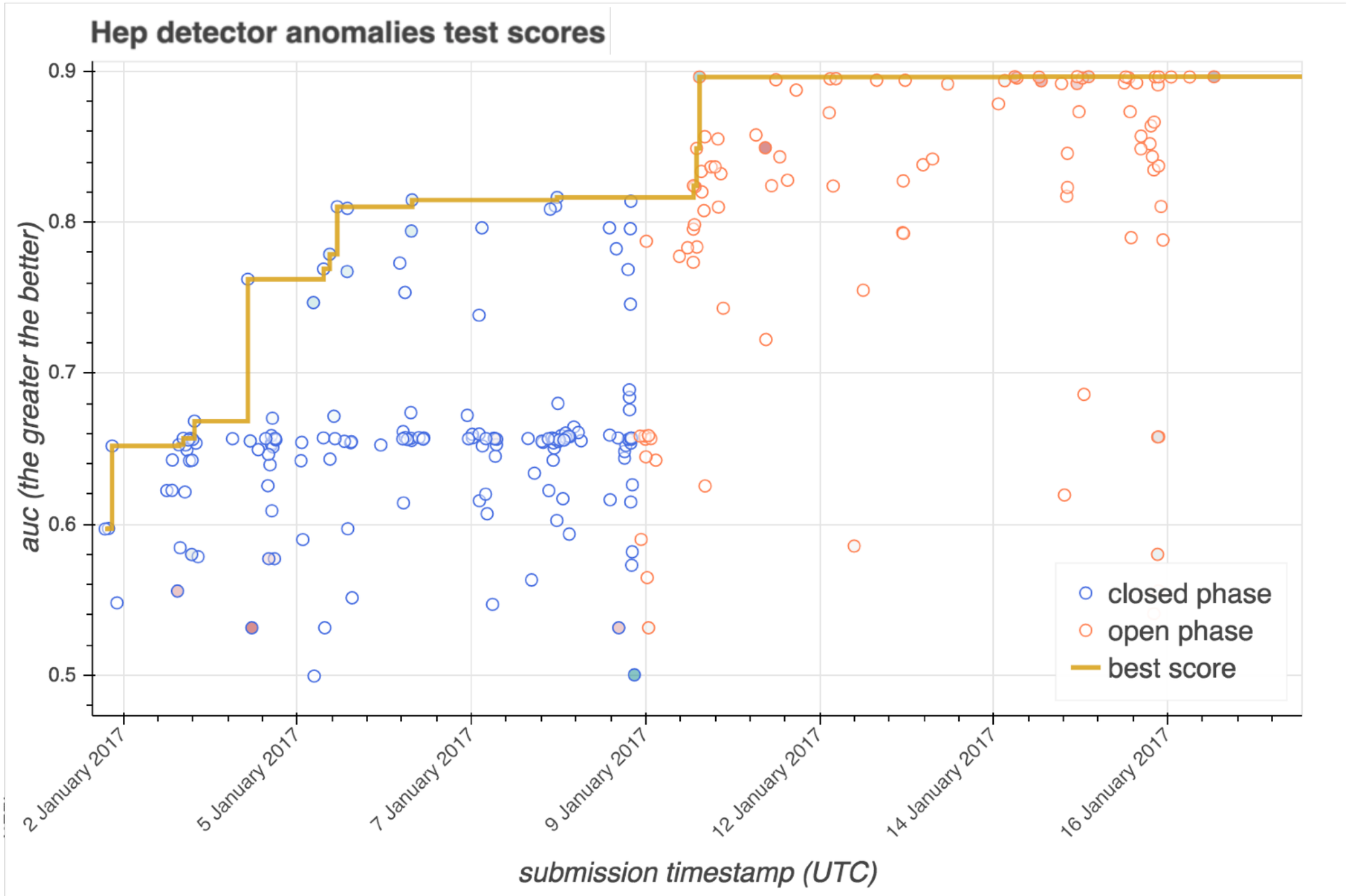




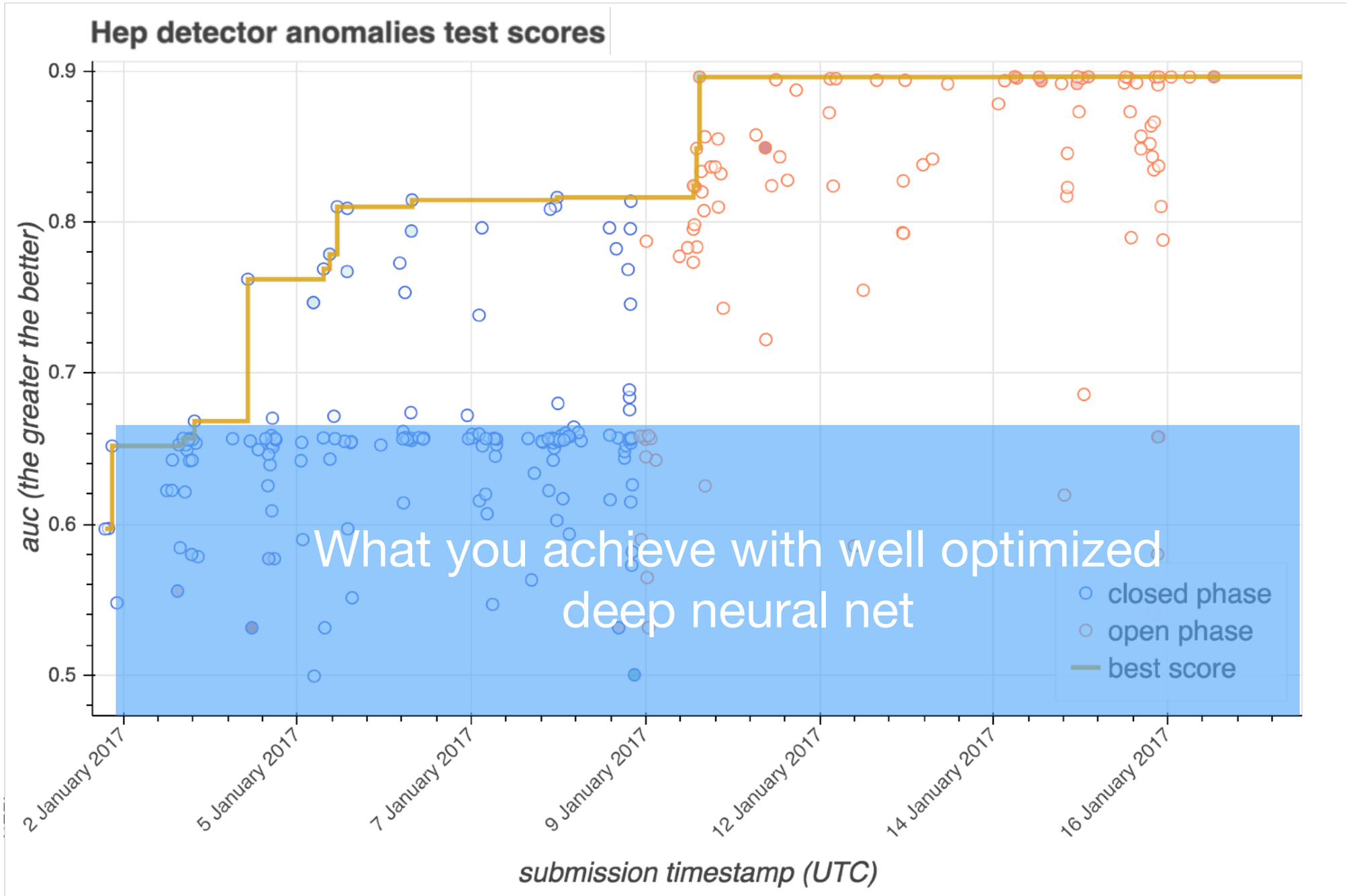


functional data, time series, data augmentation, deep learning, learning on simulations, nonstandard and multi-objective losses

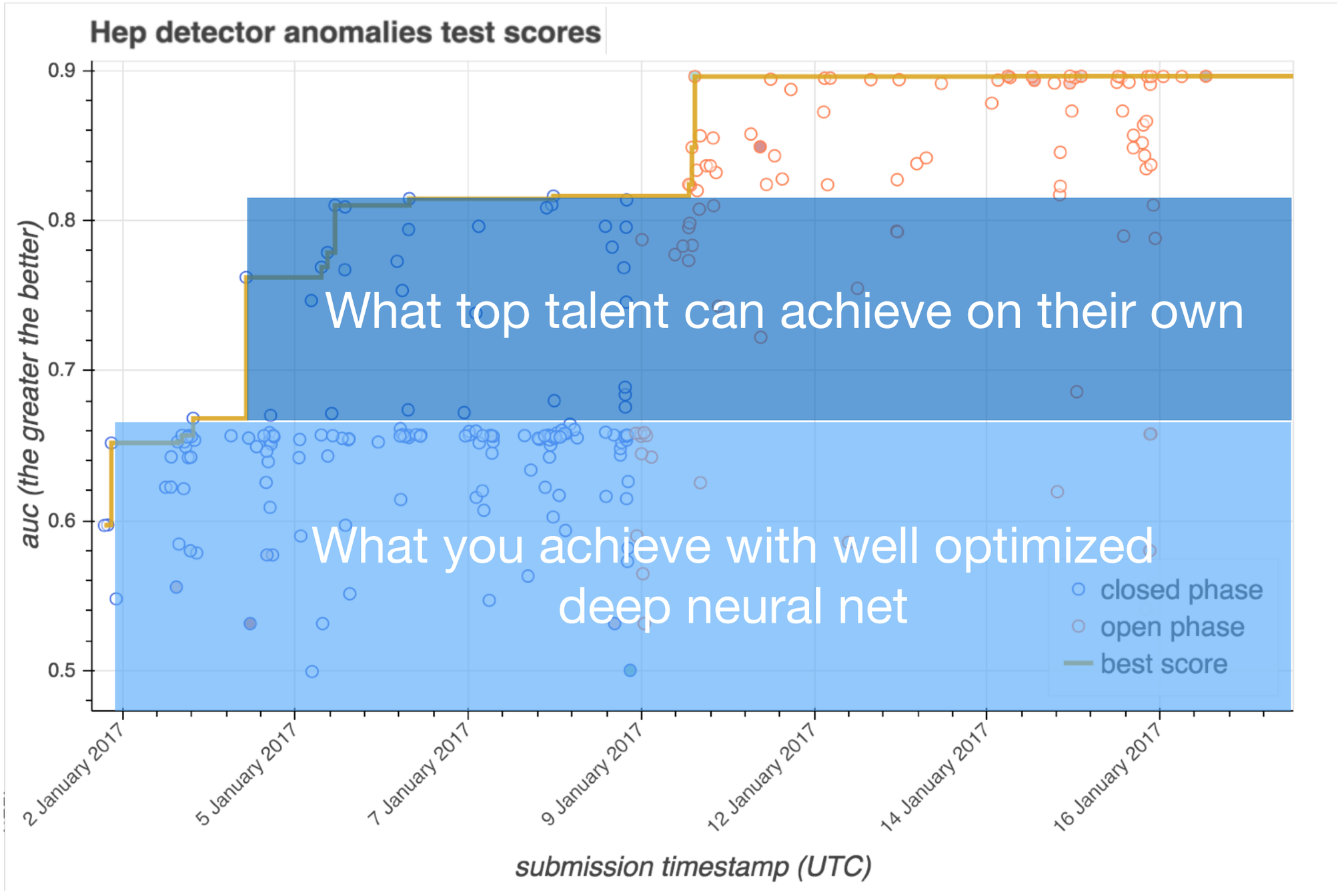
The power of the crowd



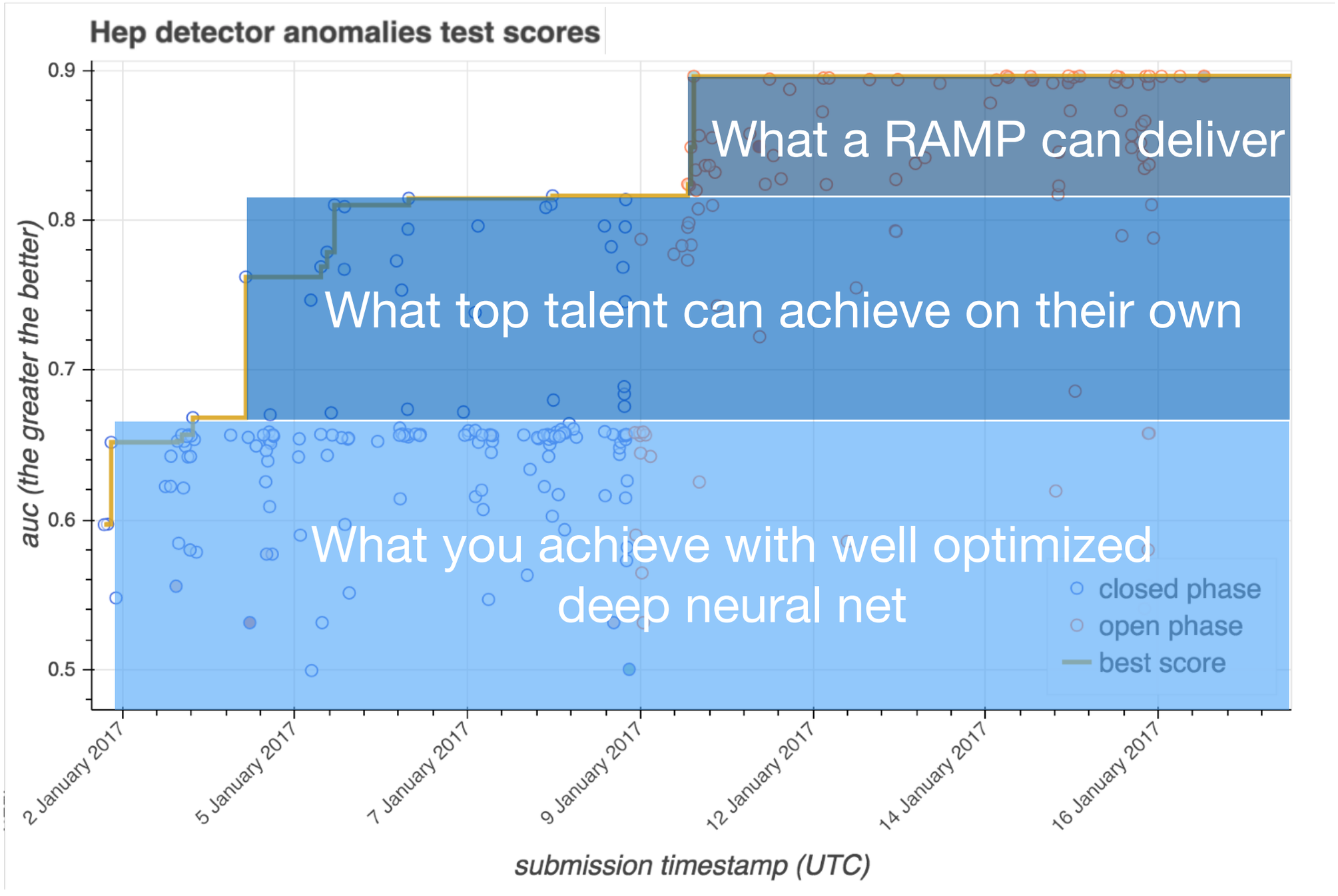
The power of the crowd



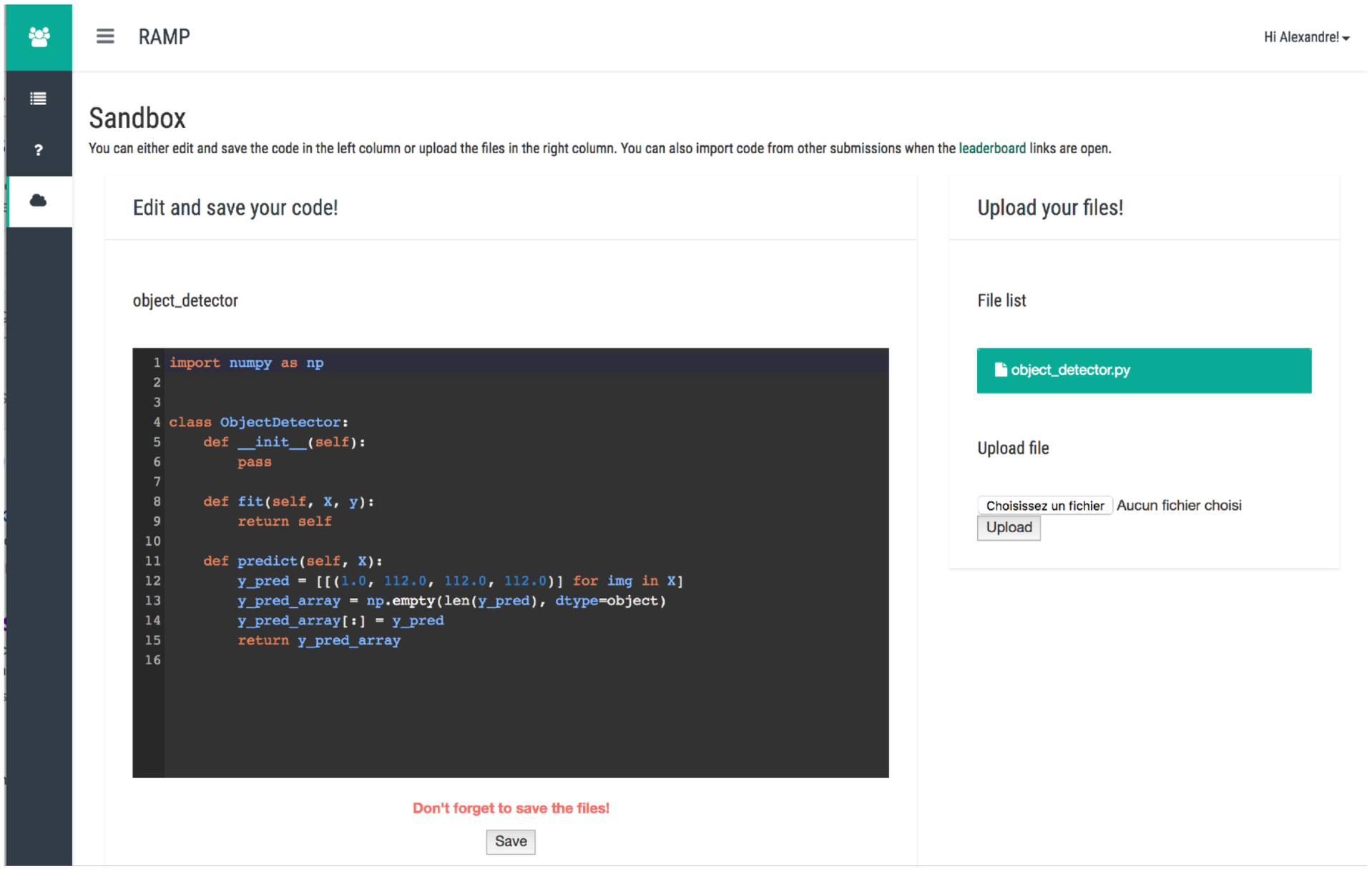
The power of the crowd



The power of the crowd



Code submission



Code submission

1. lets us deliver a **working prototype**
2. lets the participants **collaborate**
3. makes the **backend challenging** to run (cloud management)

Backends


amazon
web services™


HPC Center
ROMEO
Centre de Calcul Régional


CCIN2P3

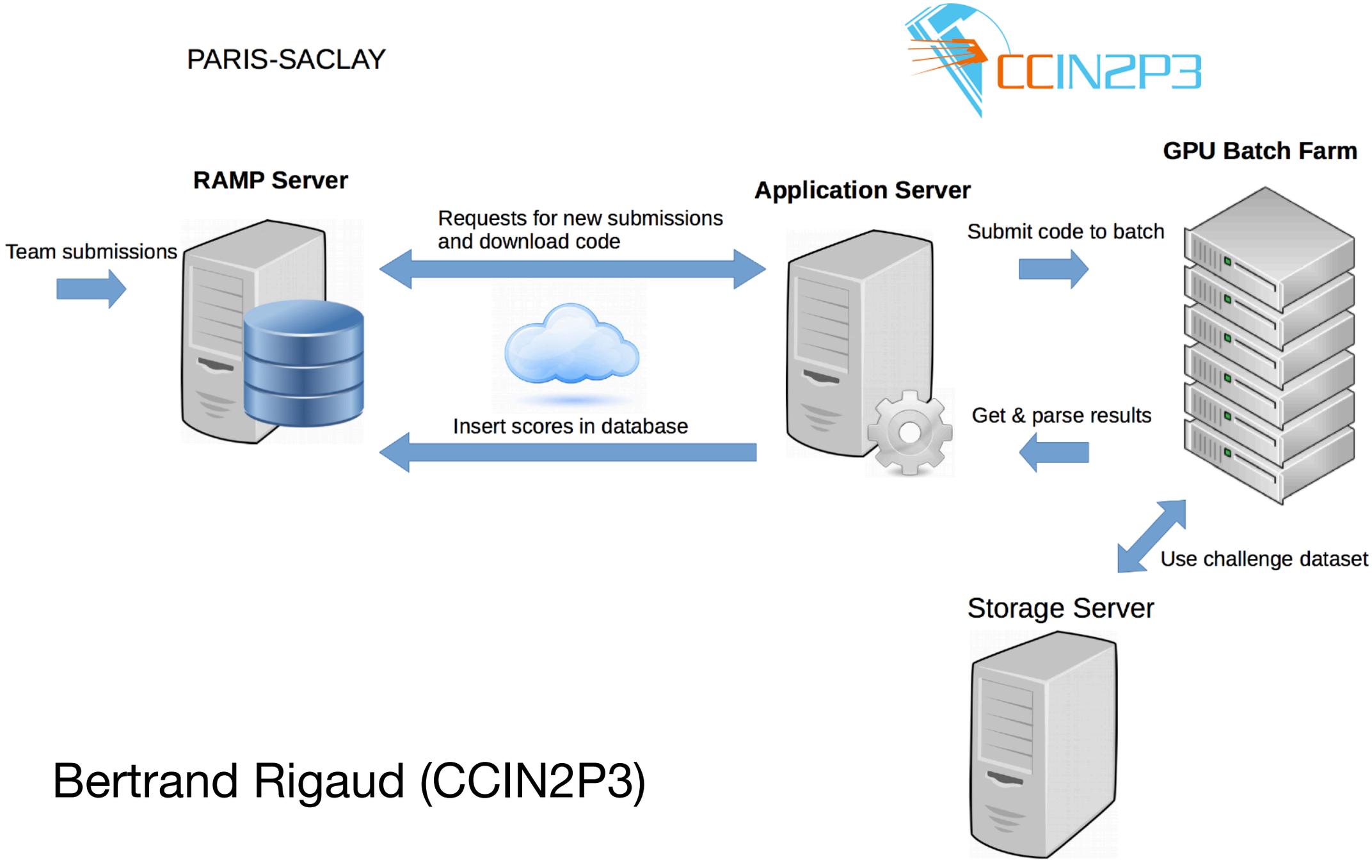
Backends

amazon
web services™

HPC Center
ROMEO
Centre de Calcul Régional

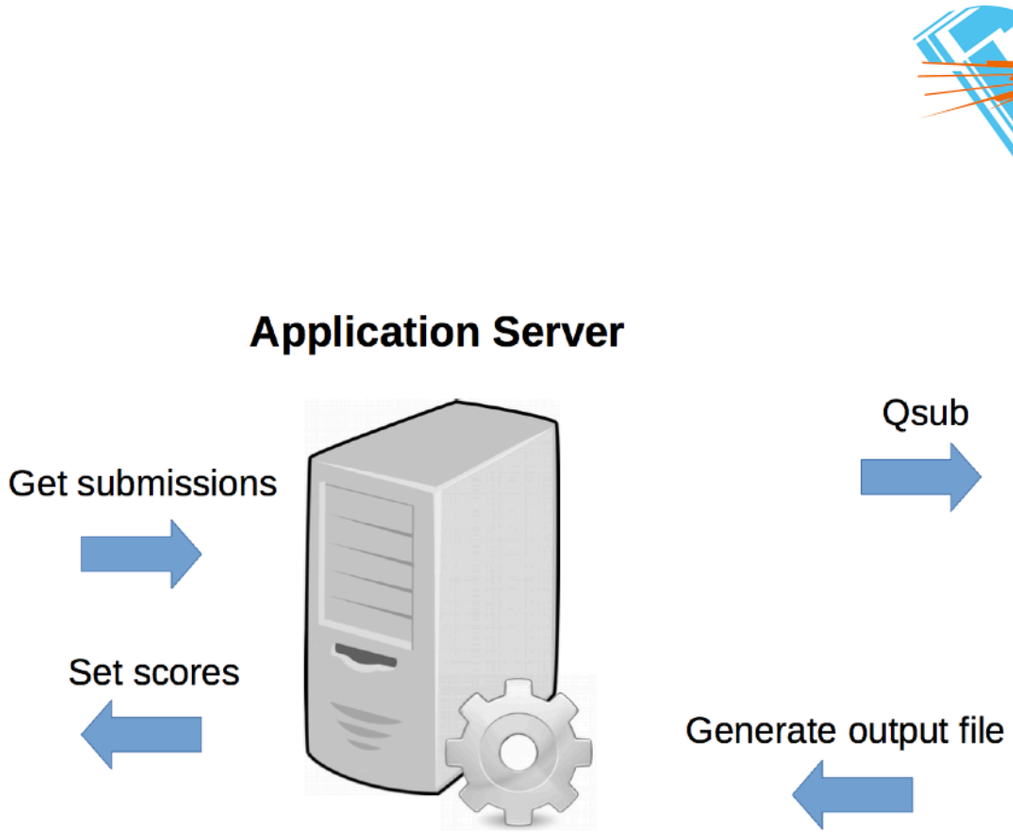

CCIN2P3

RAMP-backend interface



Bertrand Rigaud (CCIN2P3)

On the backend side

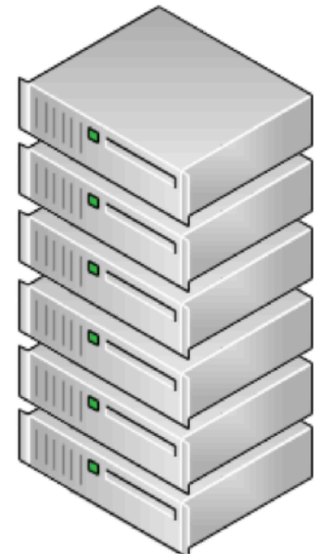


CCIN2P3

GPU Batch Farm

Use Singularity image :
- gpu libraries (cuda, cudnn)
- ramp-worflow
- python environment with event requirements

Parameters
- event image path
- submission id
- submission path
- dataset path



Bertrand Rigaud (CCIN2P3)

ramp-backend

RAMP backend

https://github.com/paris-saclay-cds/ramp-backend

Suite of command-line tools to interact with the RAMP database from an external backend server.

- **automate** the training of RAMP submissions on **any** backend
- uses **outbound connexion** from backend to the RAMP database
- no need to open multiple user accounts to use backend ressources

ramp-backend

RAMP backend

build passing

https://github.com/paris-saclay-cds/ramp-backend

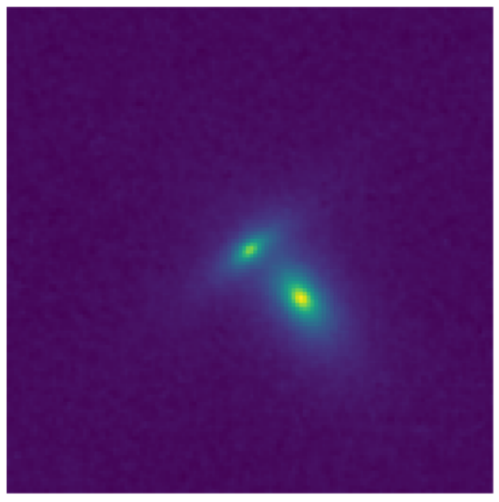
Suite of command-line tools to interact with the RAMP database from an external backend server.

- **automate** the training of RAMP submissions on **any** backend
- uses **outbound connexion** from backend to the RAMP database
- no need to open multiple user accounts to use backend ressources

Galaxy deblending challenge

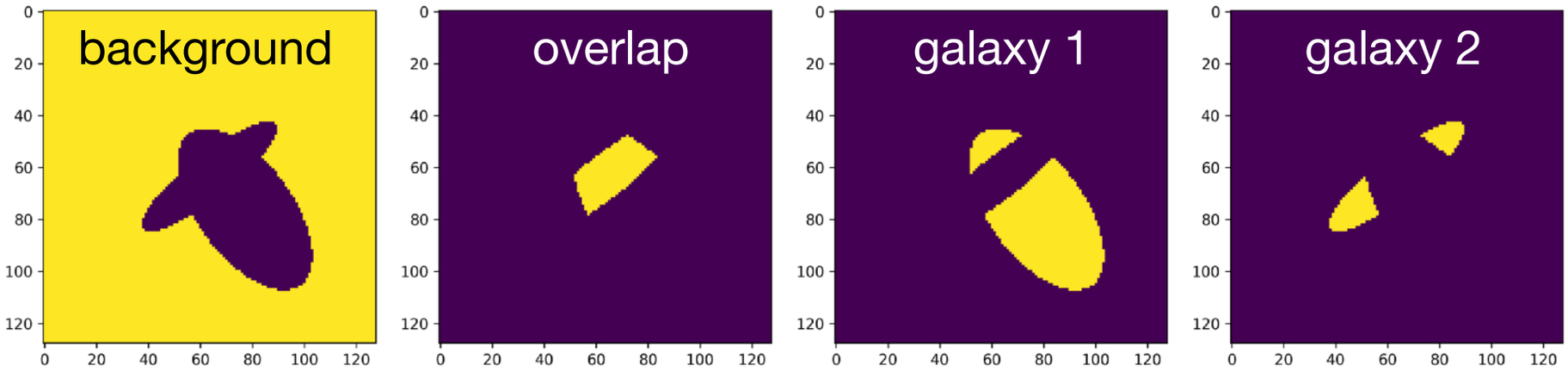
Segmentation of simulated blended galaxies

PRELIMINARY



input image

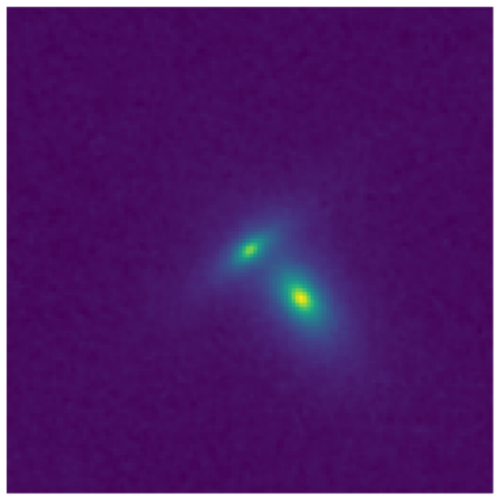
true segmentation



with Marc Huertas-Company (LERMA)

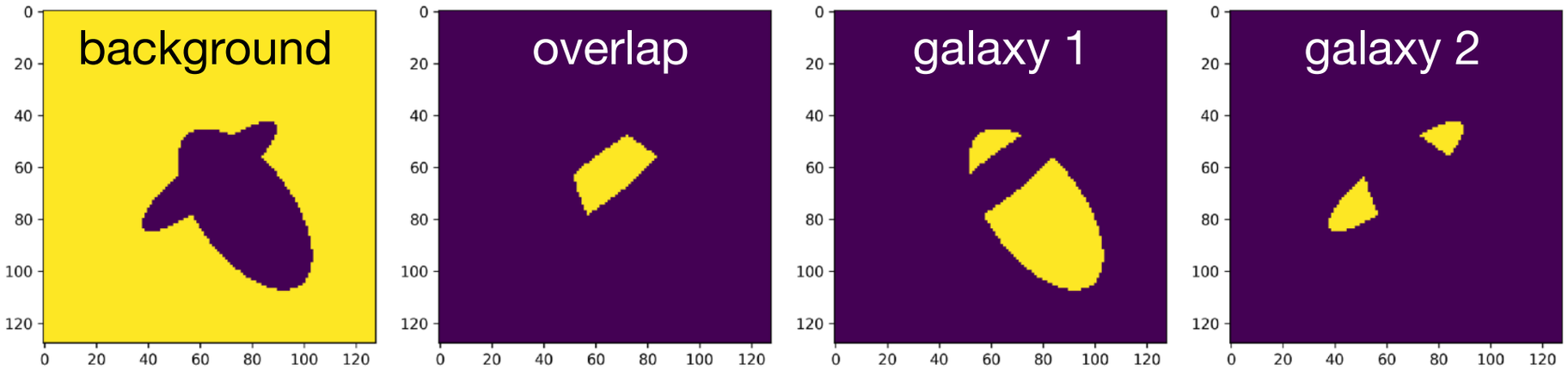
Segmentation of simulated blended galaxies

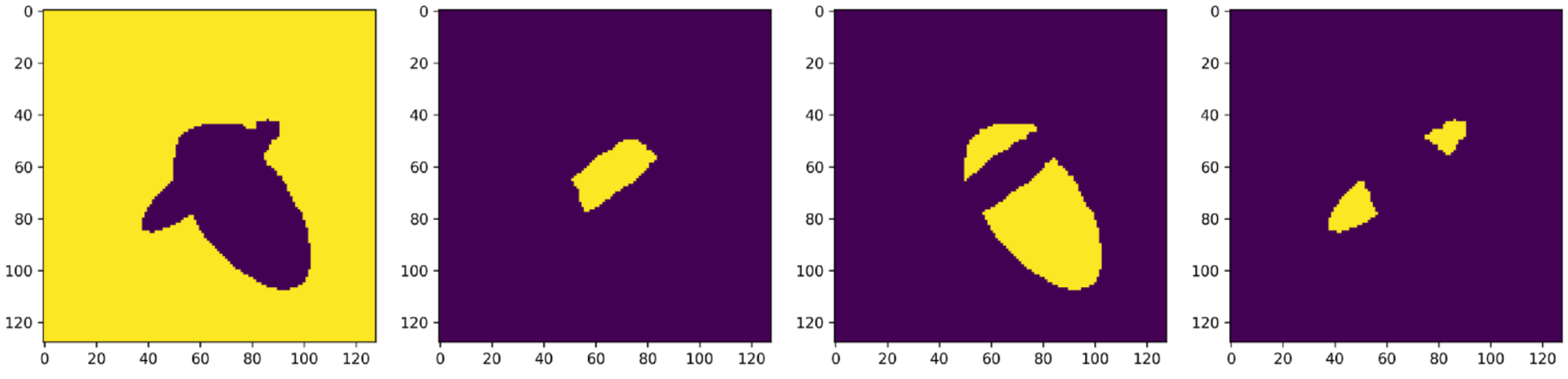
PRELIMINARY



input image

true segmentation





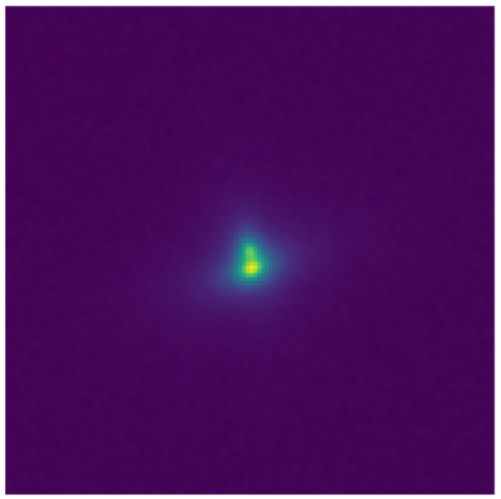
predicted segmentation with U-Net

with Marc Huertas-Company (LERMA)

arXiv:1505.04597

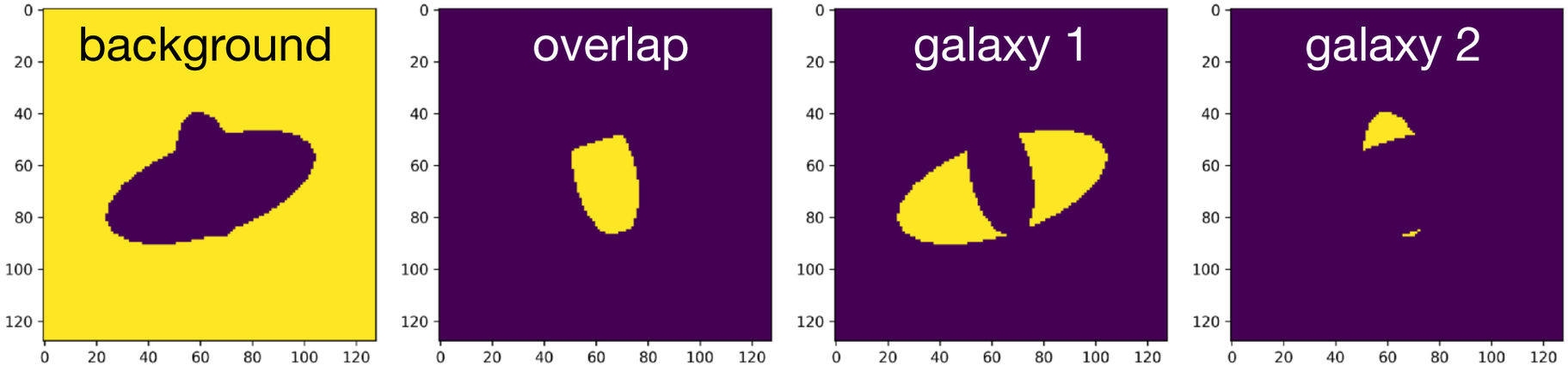
Segmentation of simulated blended galaxies

PRELIMINARY



input image

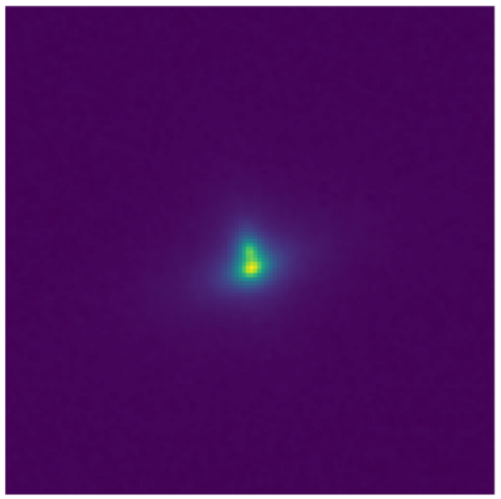
true segmentation



with Marc Huertas-Company (LERMA)

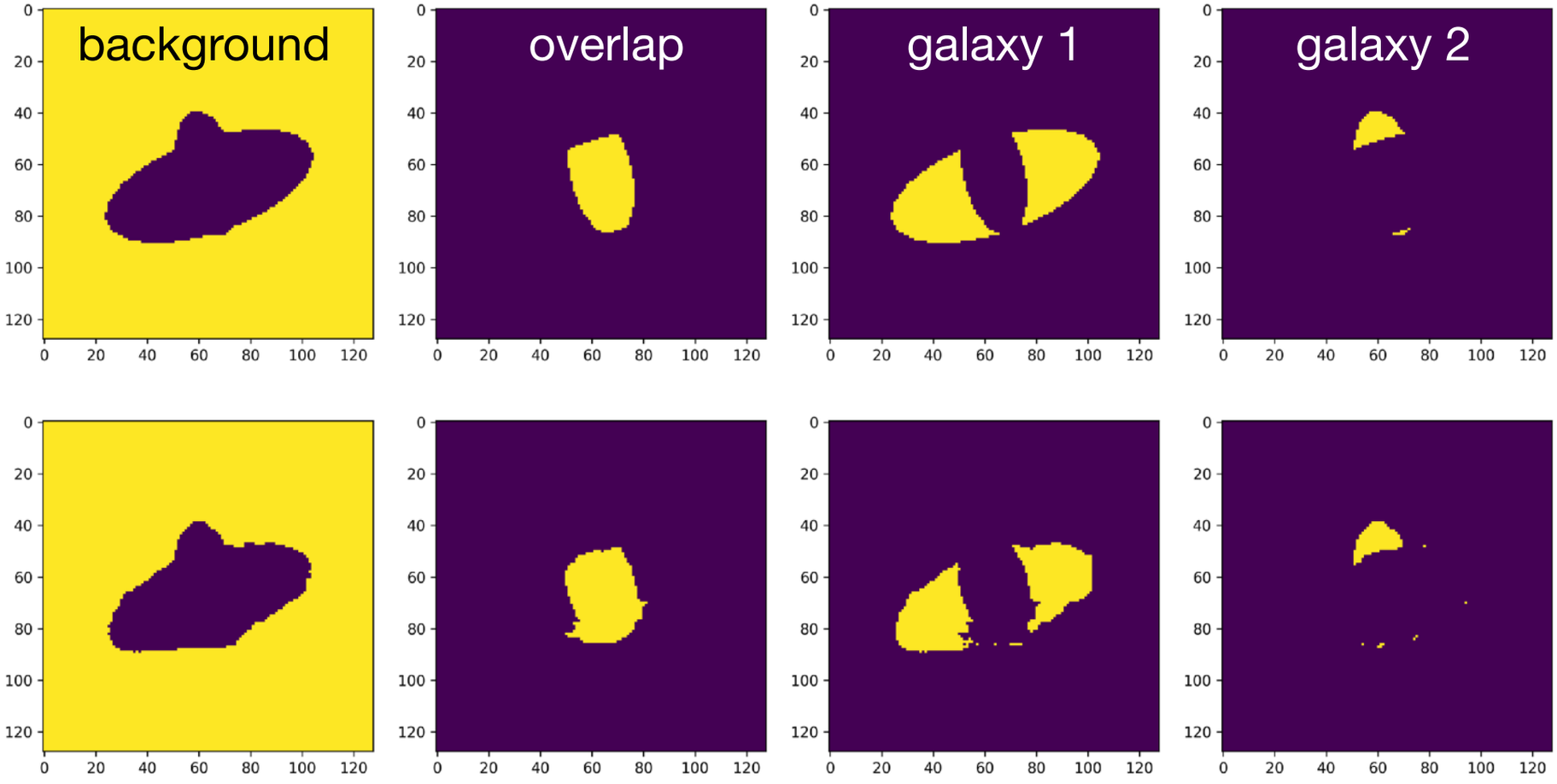
Segmentation of simulated blended galaxies

PRELIMINARY



input image

true segmentation



predicted segmentation with U-Net

with Marc Huertas-Company (LERMA)

arXiv:1505.04597

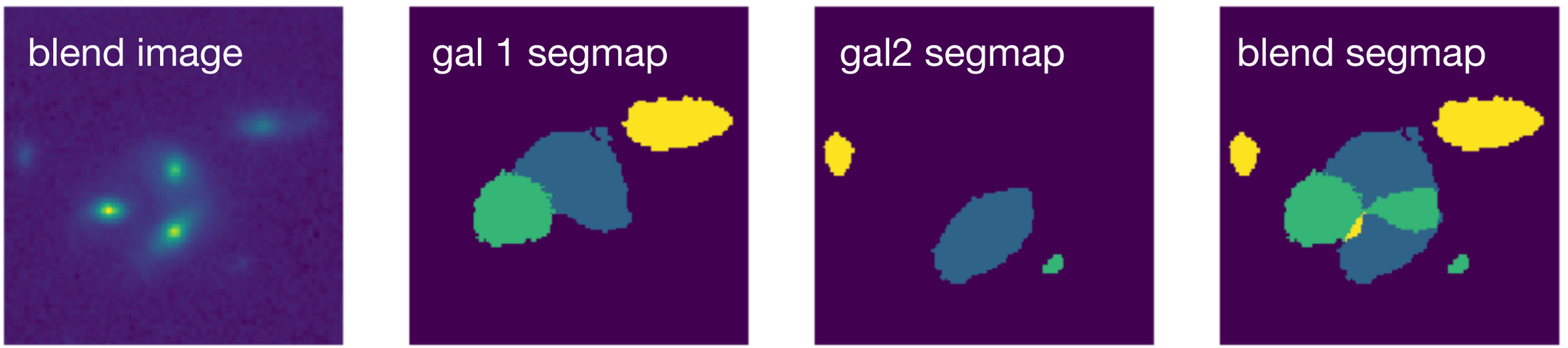
Deblending challenge

Galaxy deblending is a trending topic in astronomical large scale surveys

Preliminary work with simulations show **DNN** have a **big potential** to deal with deblending

We are creating **a challenge** to apply **ML methods** on realistic data to assess their actual **performance** compared to S.o.A techniques.

Increasing the image complexity — HST data



blending individual HST galaxies to create a realistic dataset

..but removing the neighbours

and sticking to 2 galaxies to better evaluate the model

Used for teaching soon

Deep Learning PhD track course — April 9-13
by Marc Huertas-Company

20 students (60+ requests)

work on the preliminary **galaxy deblending RAMP challenge**

10-16 GPUs provided by CCIN2P3

test of the **ramp-backend** tools

Thank you

frontend:

www.ramp.studio

toolkit:

github.com/paris-saclay-cds/ramp-workflow

examples:

github.com/ramp-kits

slack:

ramp-studio.slack.com