

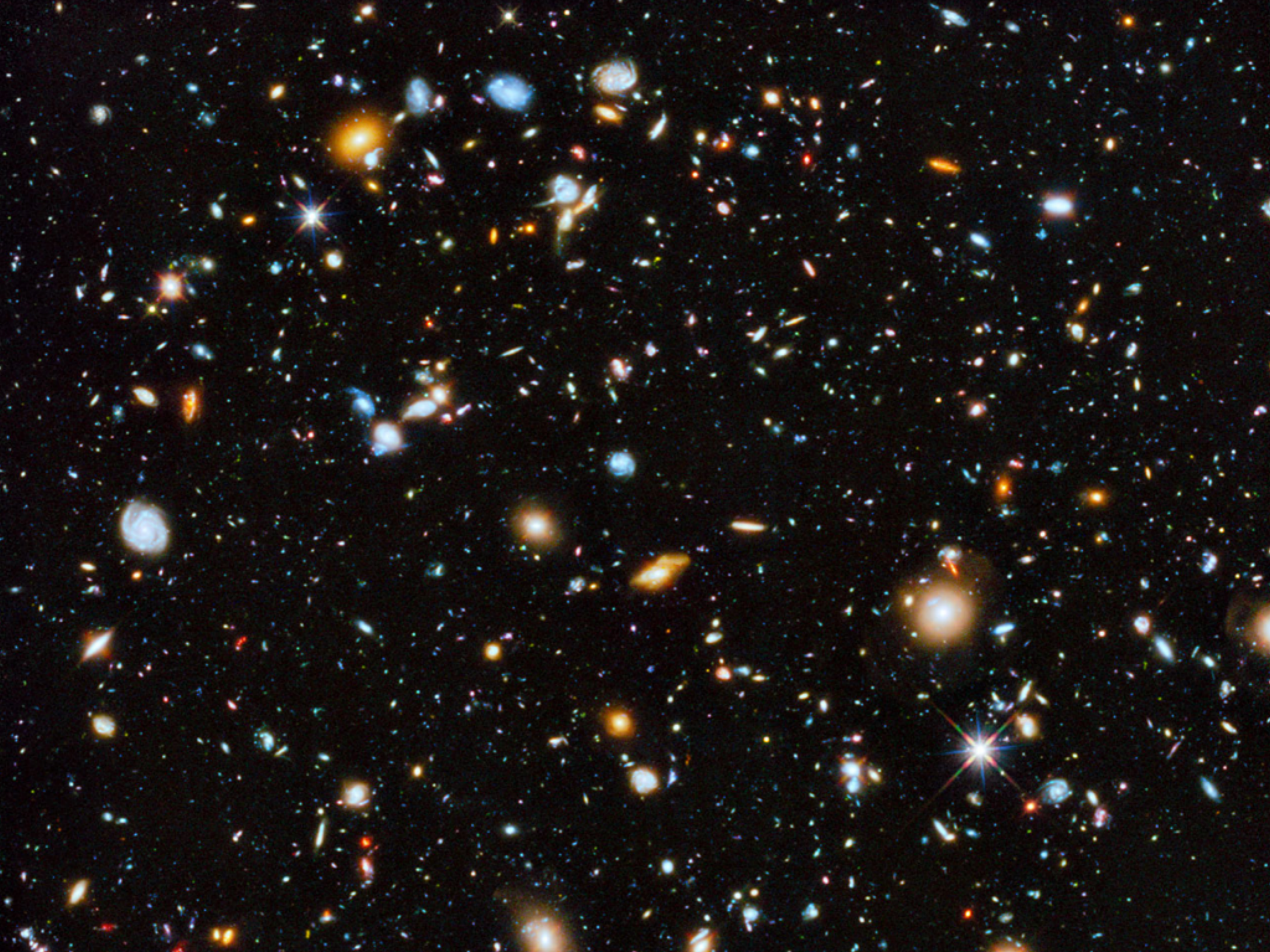
Galaxy morphology with CNNs using transfer learning

Alexandre Boucaud

Paris-Saclay Center for Data Science / LAL

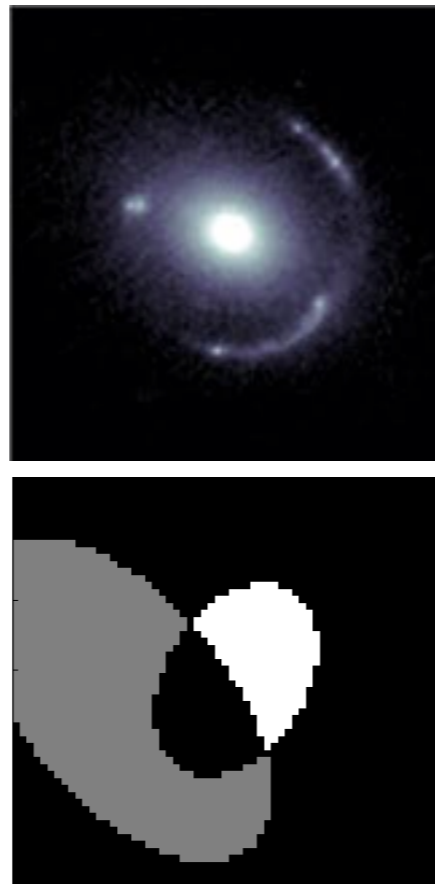
aboucaud@lal.in2p3.fr

in collab. with
Marc Huertas-Company
Diego Tuccillo



The impact of morphologies

- object radii
- inclination
- deblending
- lenses
- star/galaxy
- ...



- barred galaxies
- galaxy mergers
- quenching
- AGNs
- outreach



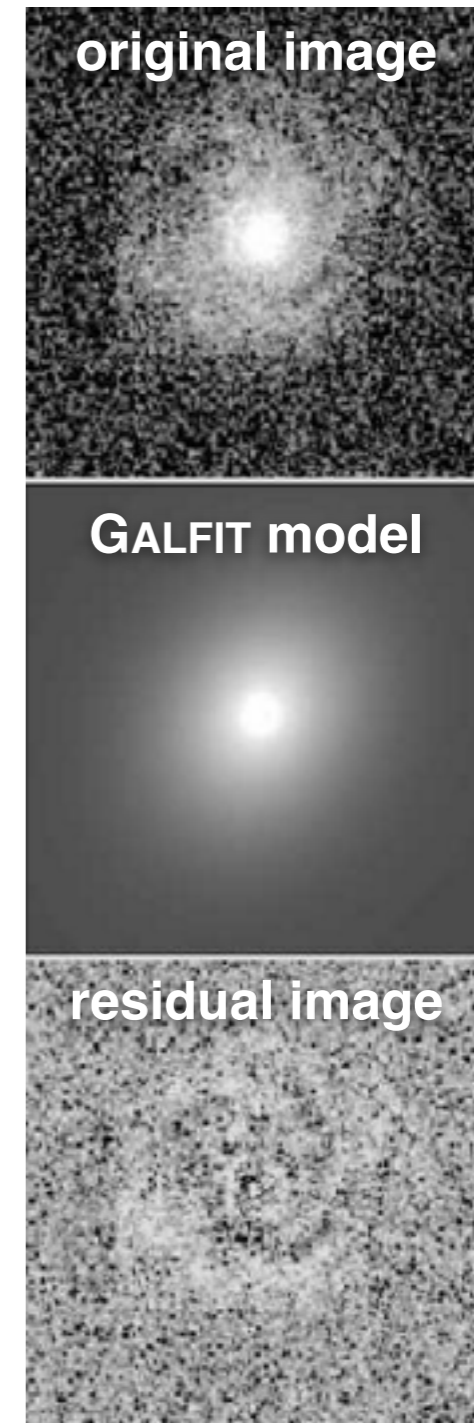
measurements to be **automated** on $\sim 10^{10}$ objects

Outline

1. Model-fitting morphologies
2. Visual morphologies
3. Data / Science challenges

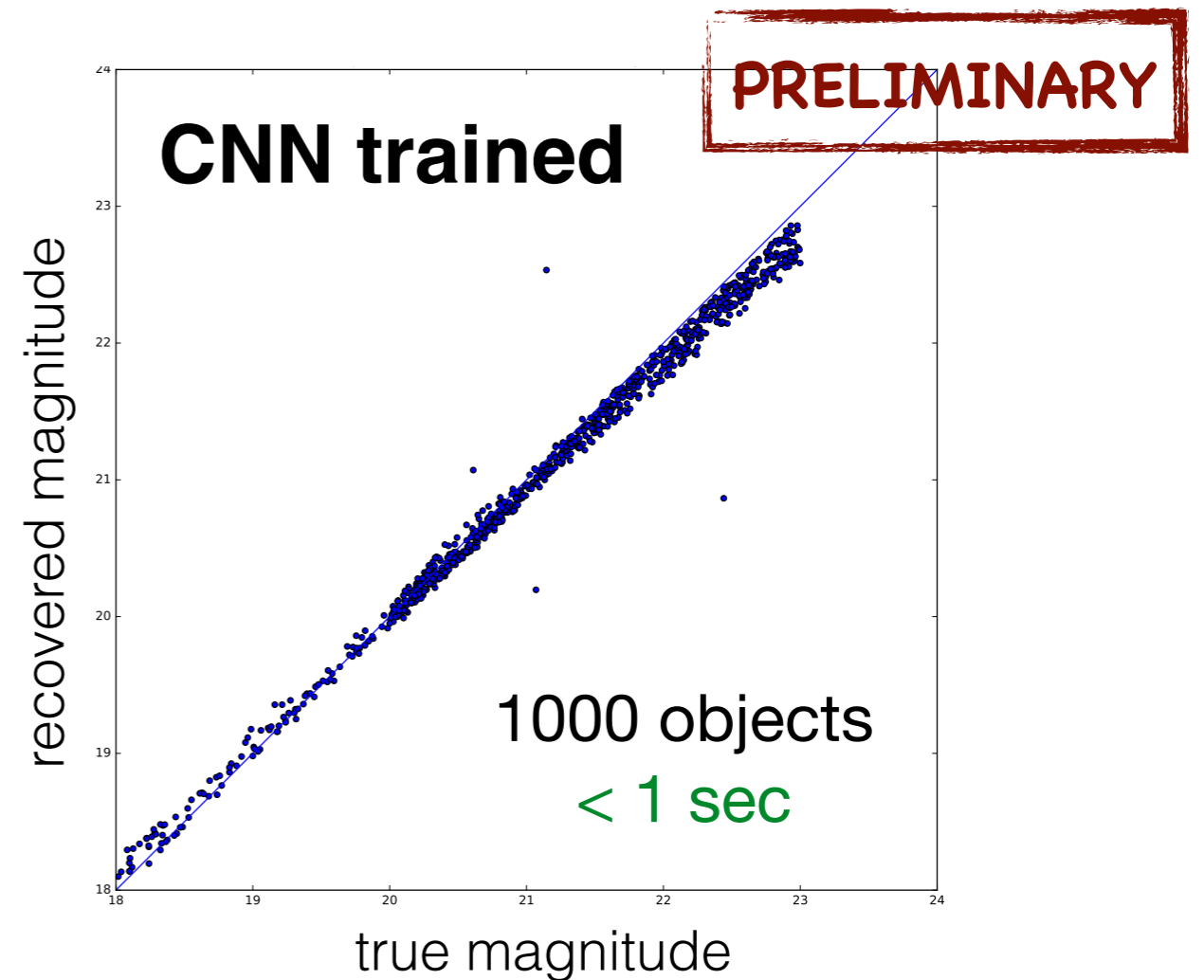
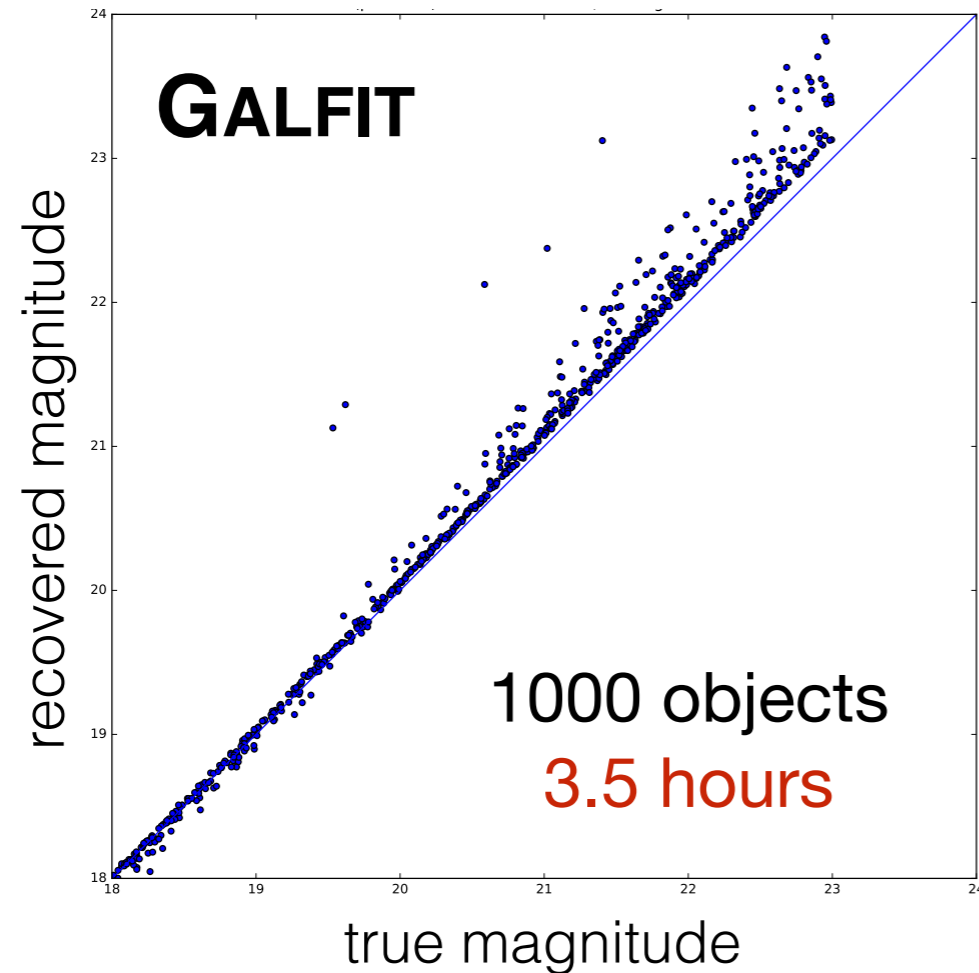
1. Model fitting morphologies

- χ^2 minimisation of a given model w.r.t. to a galaxy stamp, including its PSF
- extraction of **basic quantities**
radius, magnitude, inclination, b/a, ...
- **high number** of free parameters per pixels
=> minimisation is slow



GALFIT user's manual

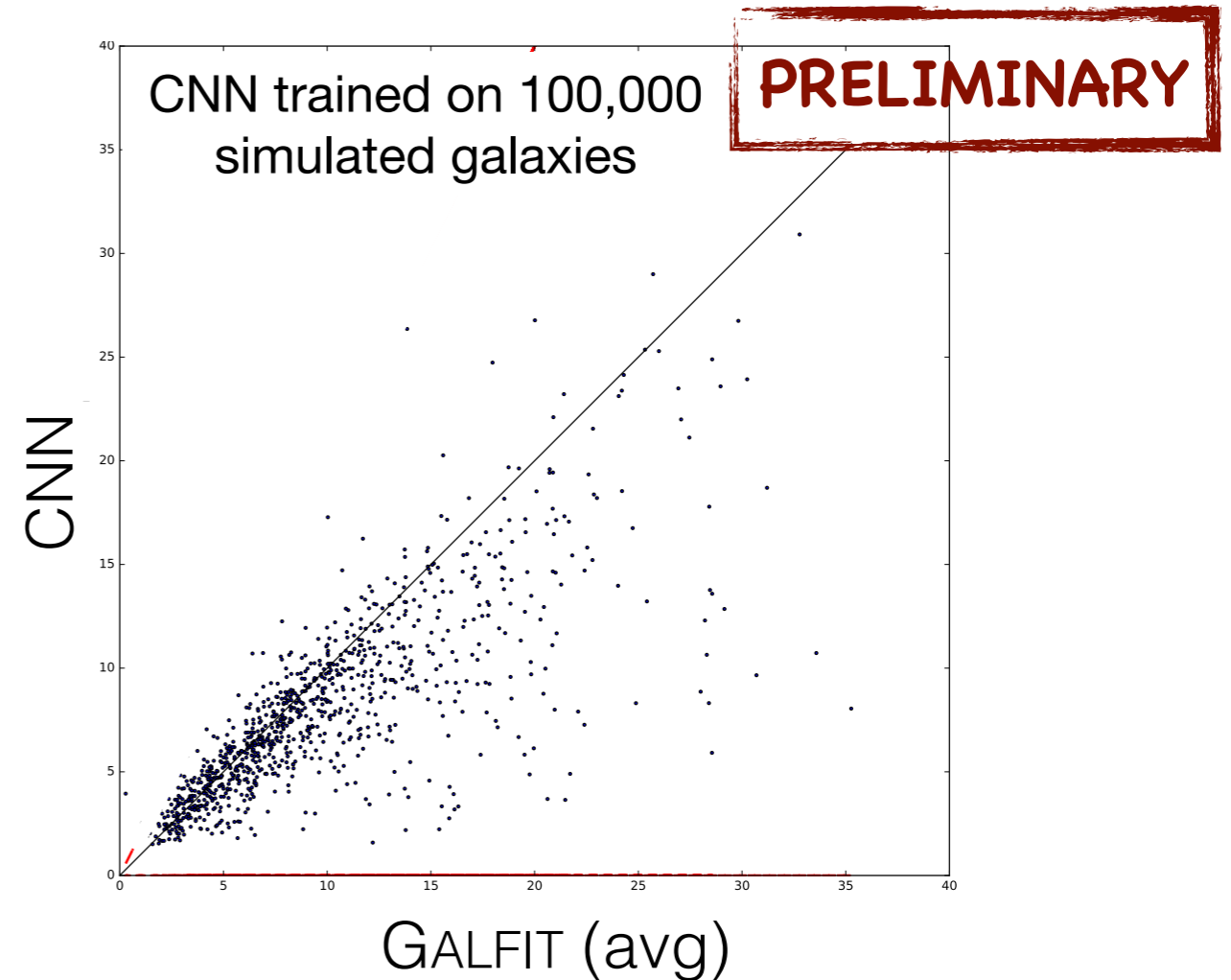
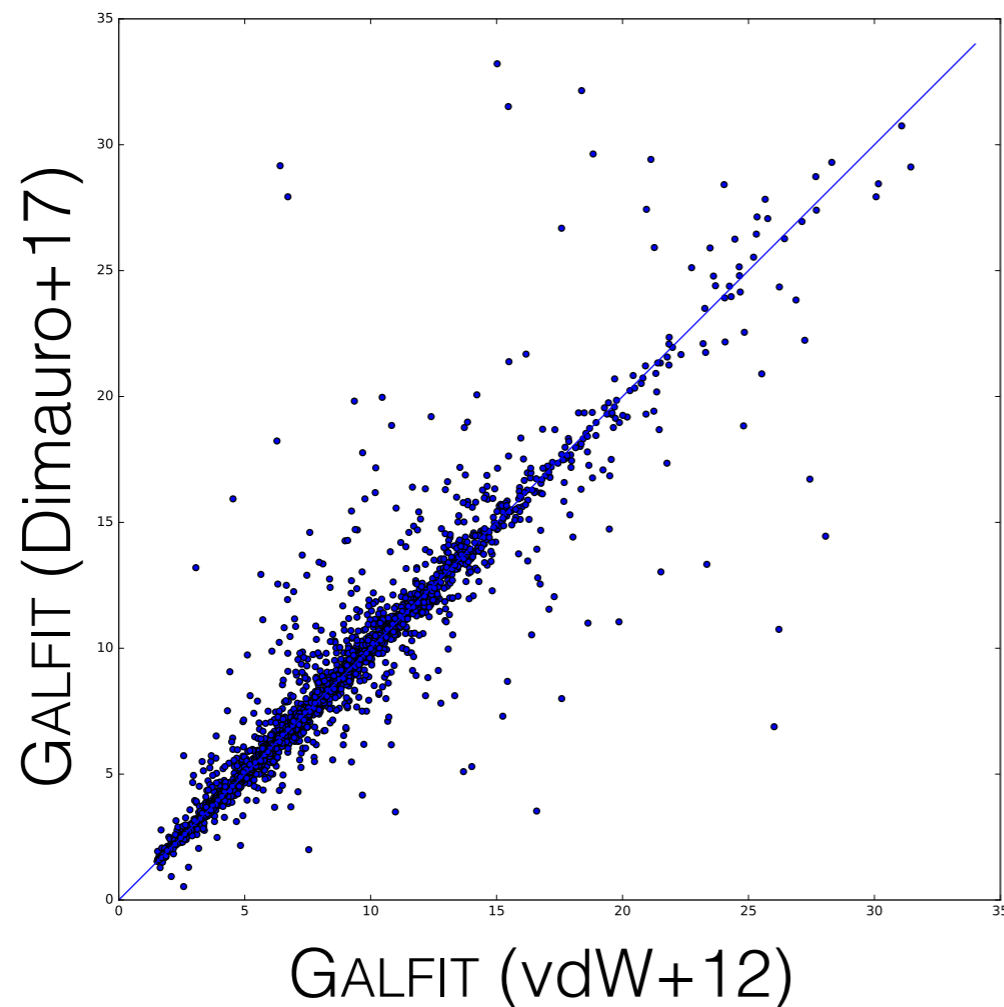
Convolutional neural networks



Tuccillo, Huertas-Company et al.

On simulations: very similar results
but **CNNs are orders of magnitude faster**

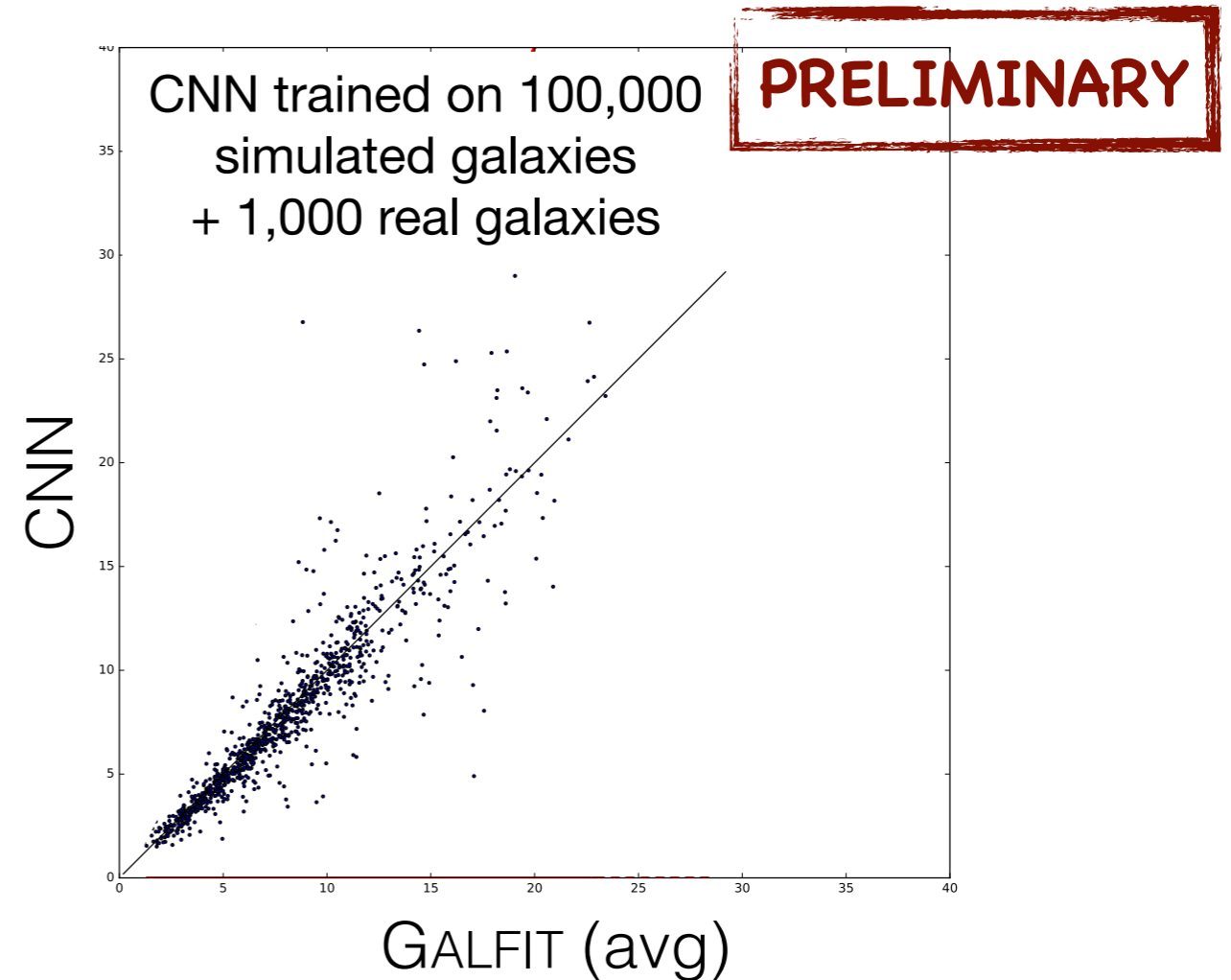
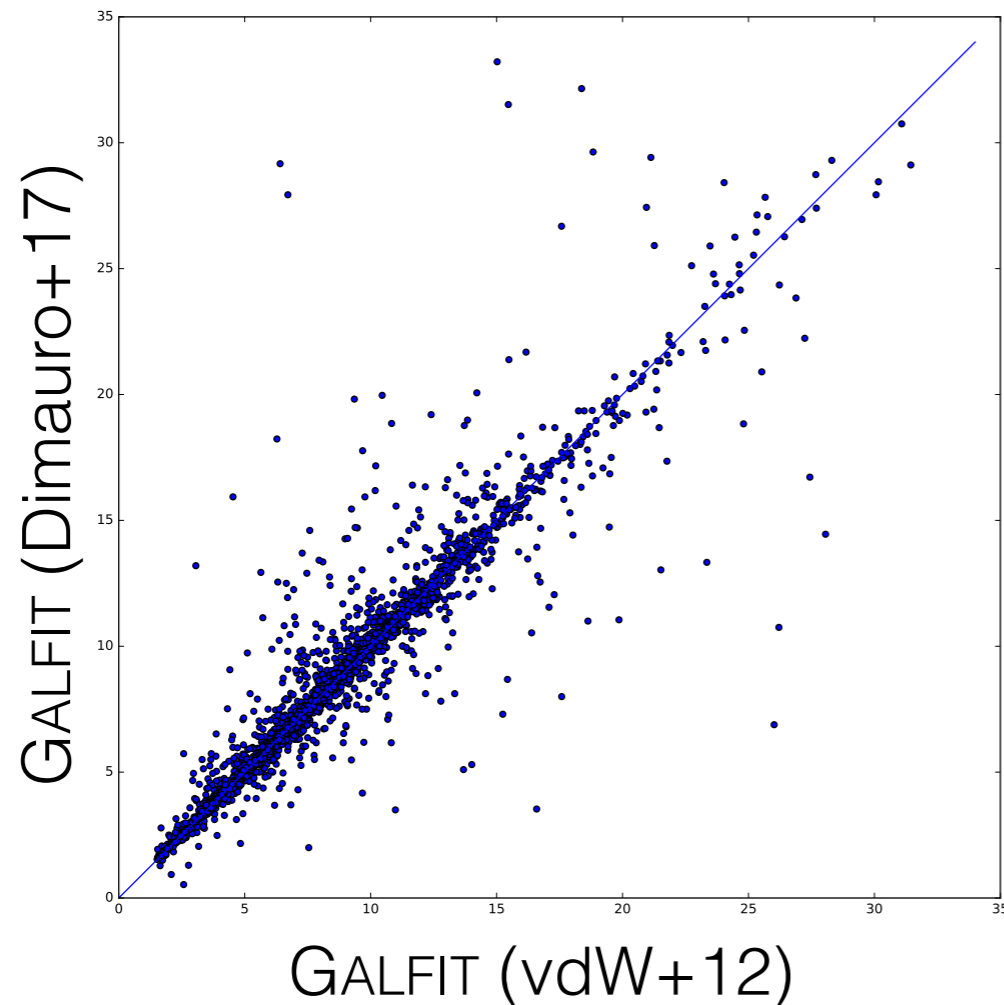
Transfer knowledge



Tuccillo, Huertas-Company et al.

On **real CANDELS galaxies**: comparison of radius estimation with GALFIT (van der Wel+12, Dimauro+17) and CNN trained on sims

Transfer knowledge



Tuccillo, Huertas-Company et al.

On **real CANDELS galaxies**: comparison of radius estimation with GALFIT (van der Wel+12, Dimauro+17) and CNN trained on sims

Model-fitting morphologies

Current status

- accurate extraction of **shape features** (sizes, b/a , ...)
- model must be **imposed**
- they are **slow**

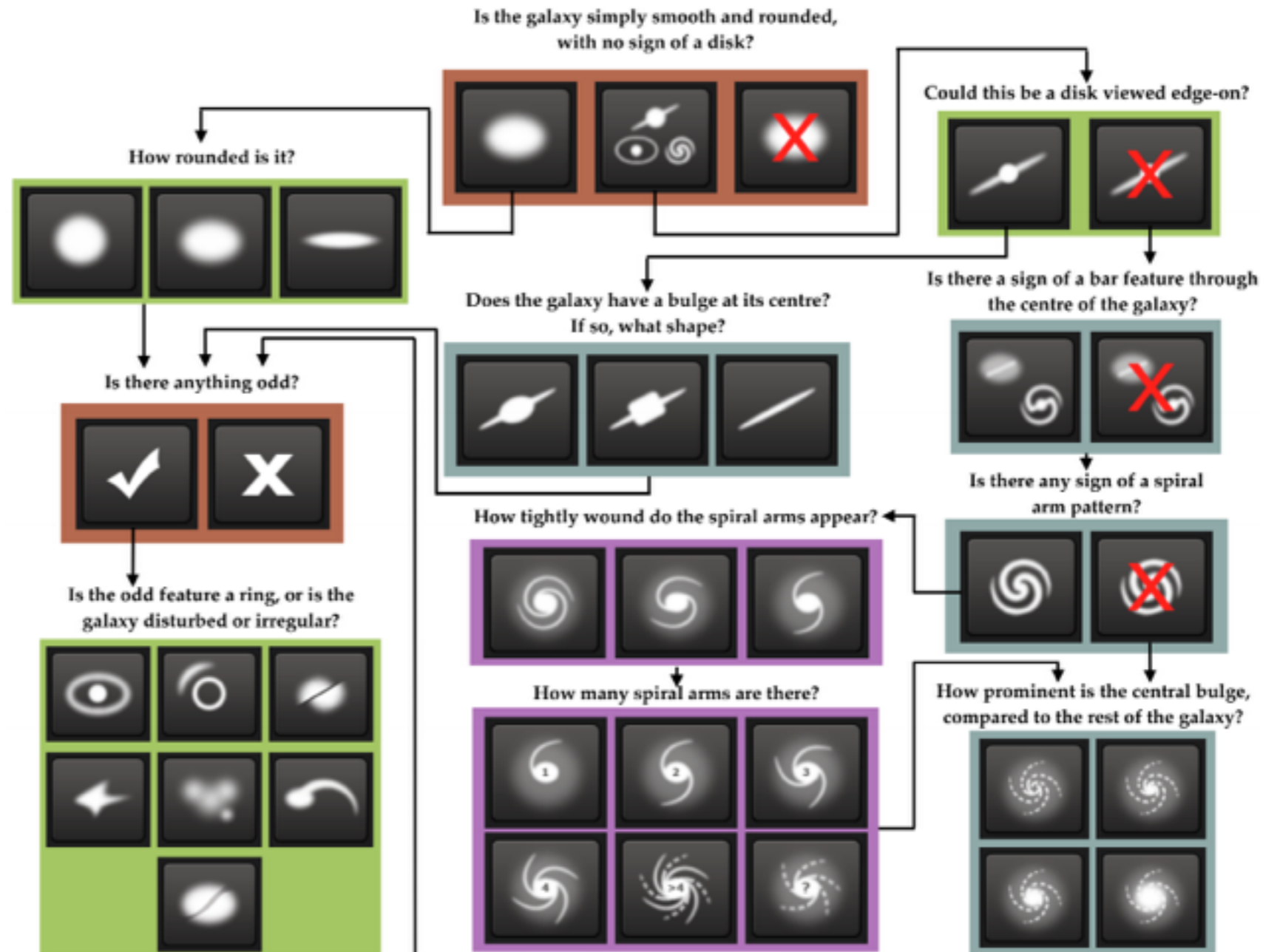
On-going work

- improving the **knowledge transfer** from simulations to real data
- using the speed of CNNs to deliver priors to **boost the fitting**

2. Visual morphologies



- human classification using decision trees
- scientists + citizens
- < 1M galaxies today



A long process..

- visual morphologies are « **expensive** »
900,000 classified galaxies after 10 years
- humans are **subjective** & image quality **varies**
- need to **combine human and machine** to obtain
faster, more accurate visual morphologies

Non-parametric morphologies

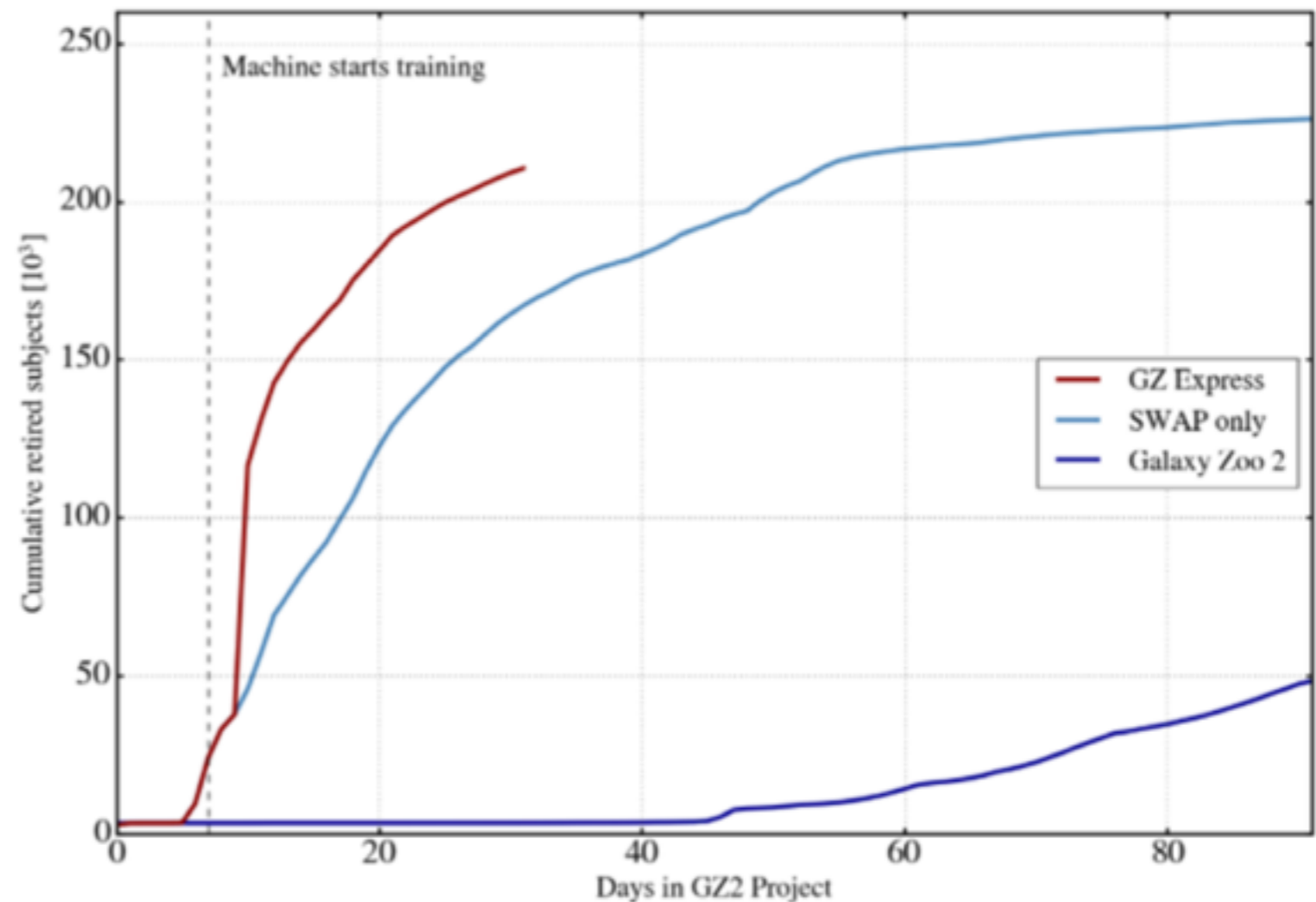
C	A	S	+	M₂₀	+	Gini
concentration	asymmetry	smoothness				
Abraham+ 96, 03	Conselice+ 00, 03			Lotz+ 04		

- moment-based parameters (weighted sums of pixels) which **correlate with morphologies**
- values **highly dependent** upon implementation & segmentation
- **low sensitivity** to image noise and neighbors
- **good classification performance** when combining the 5 indices

CASMG + machine learning

Results of GX on GZ2 classification data

- Simulations utilize first question of GZ2 decision tree (Featured/Smooth)
- GX classifies over 200K galaxies in just 27 days of equivalent GZ2 time.
- GX provides an **order of magnitude** faster classification time than the original GZ2 project.
- GX maintains 93.5% accuracy



Beck, et al., submitted

courtesy of M.Beck, C. Scarlata

Introducing Deep learning

kaggle

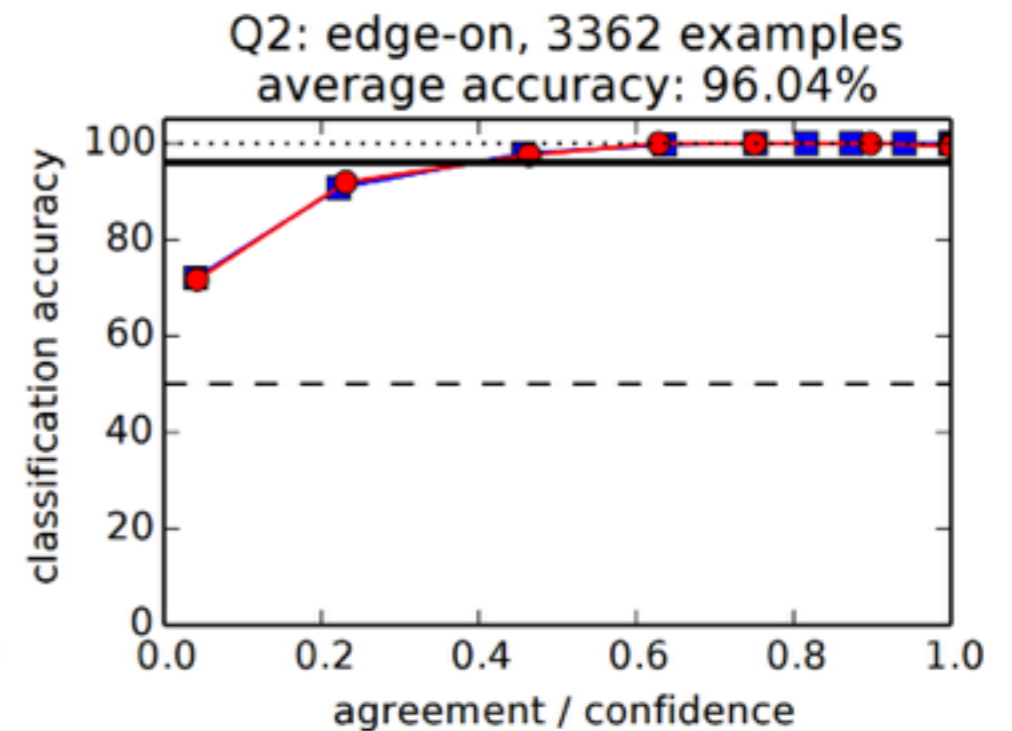
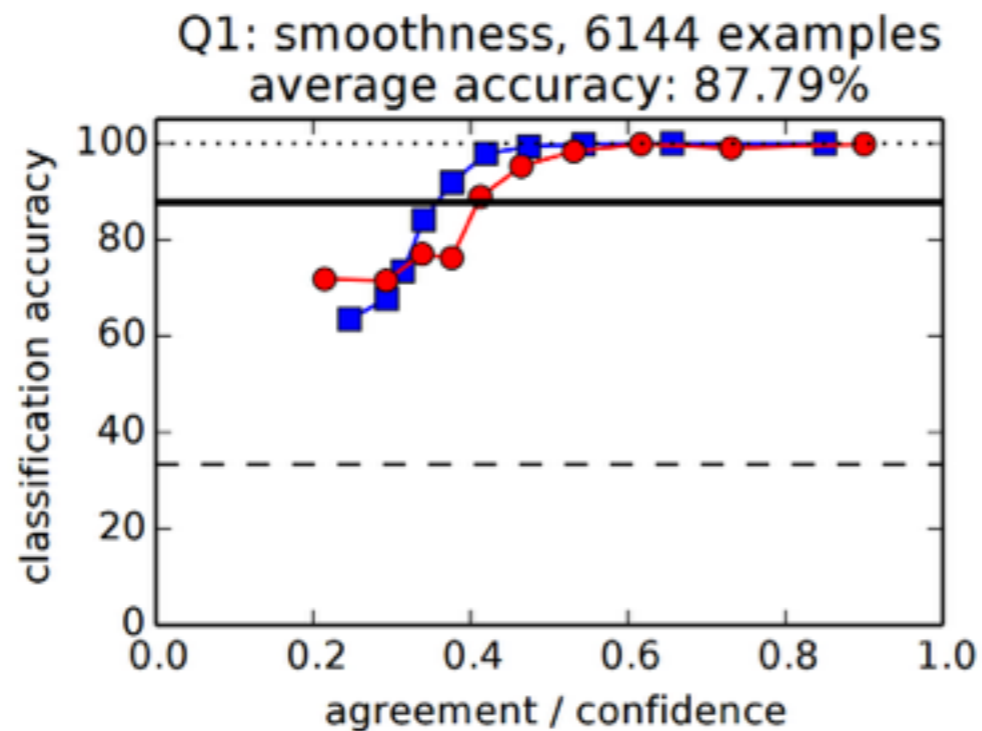


Completed • \$16,000 • 326 teams

Galaxy Zoo - The Galaxy Challenge

Fri 20 Dec 2013 - Fri 4 Apr 2014 (3 years ago)

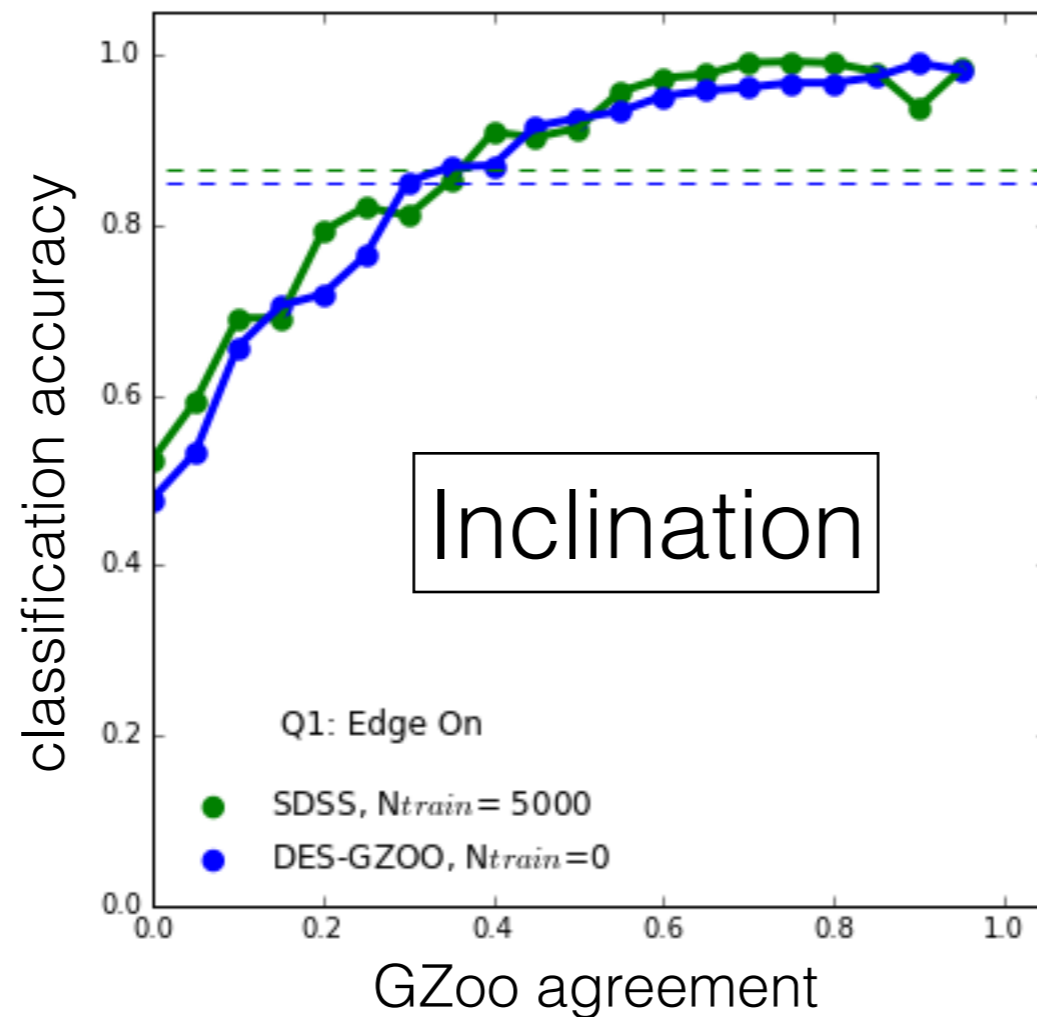
The galaxy challenge triggered the use of CNN for classification



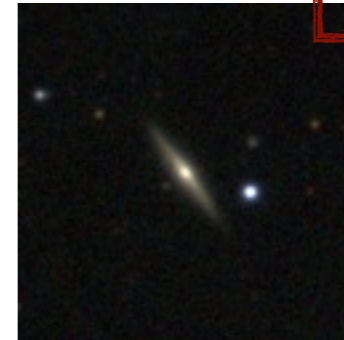
Dieleman, Willett & Dambre (2015)

works well but needs a **huge training set**

Knowledge transfer: SDSS to DES



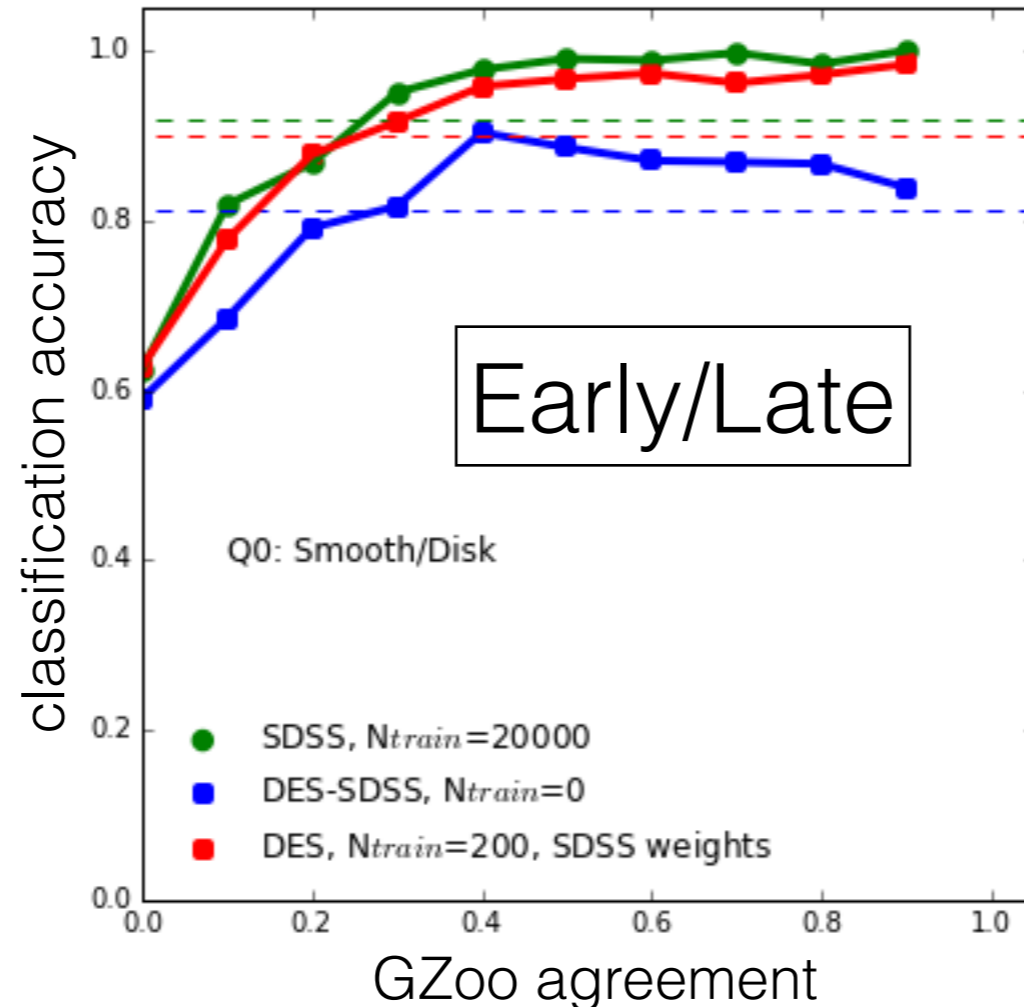
PRELIMINARY



Dominguez-Sanchez, Huertas-Company, Bernardi et al.

for simple properties, **no additional training is needed**
to use SDSS-trained network on DES data

Knowledge transfer: SDSS to DES



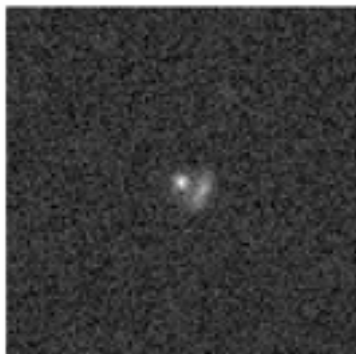
PRELIMINARY



Dominguez-Sanchez, Huertas-Company, Bernardi et al.

only **200 DES classified objects** needed
to achieve **>90% accuracy** on a SDSS-trained network

3. Data Science challenges



The Strong Gravitational Lens Finding Challenge

(Nov 16 - Feb 17)

Name	type	AUROC	TPR ₀	TPR ₁₀	short description
CMU-DeepLens-ResNet-ground3	Ground-Based	0.98	0.09	0.45	CNN
CMU-DeepLens-Resnet-Voting	Ground-Based	0.98	0.02	0.10	CNN
LASTRO EPFL	Ground-Based	0.97	0.07	0.11	CNN
CAS Swinburne Melb	Ground-Based	0.96	0.02	0.08	CNN
AstrOmatic	Ground-Based	0.96	0.00	0.01	CNN
Manchester SVM	Ground-Based	0.93	0.22	0.35	SVM / Gabor
Manchester-NA2	Ground-Based	0.89	0.00	0.01	Human Inspection
ALL-star	Ground-Based	0.84	0.01	0.02	edges/gradients and Logistic Reg.
CAST	Ground-Based	0.83	0.00	0.00	CNN / SVM
YattaLensLite	Ground-Based	0.82	0.00	0.00	SExtractor
LASTRO EPFL	Space-Based	0.93	0.00	0.08	CNN
CMU-DeepLens-ResNet	Space-Based	0.92	0.22	0.29	CNN
GAMOCCLASS	Space-Based	0.92	0.07	0.36	CNN
CMU-DeepLens-Resnet-Voting	Space-Based	0.91	0.00	0.01	CNN
AstrOmatic	Space-Based	0.91	0.00	0.01	CNN
CMU-DeepLens-ResNet-aug	Space-Based	0.91	0.00	0.00	CNN
Kapteyn	Space-Based	0.82	0.00	0.00	CNN
CAST	Space-Based	0.81	0.07	0.12	CNN
Manchester1	Space-Based	0.81	0.01	0.17	Human Inspection
Manchester SVM	Space-Based	0.81	0.03	0.08	SVM / Gabor
NeuralNet2	Space-Based	0.76	0.00	0.00	CNN / wavelets
YattaLensLite	Space-Based	0.76	0.00	0.00	Arcs / SExtractor
All-now	Space-Based	0.73	0.05	0.07	edges/gradients and Logistic Reg.
GAHEC IRAP	Space-Based	0.66	0.00	0.01	arc finder

best classifiers
for ground-based

6 CNNs
for 10 submissions

best classifiers
for space-based

9 CNNs
for 14 submissions

courtesy of B. Metcalfe

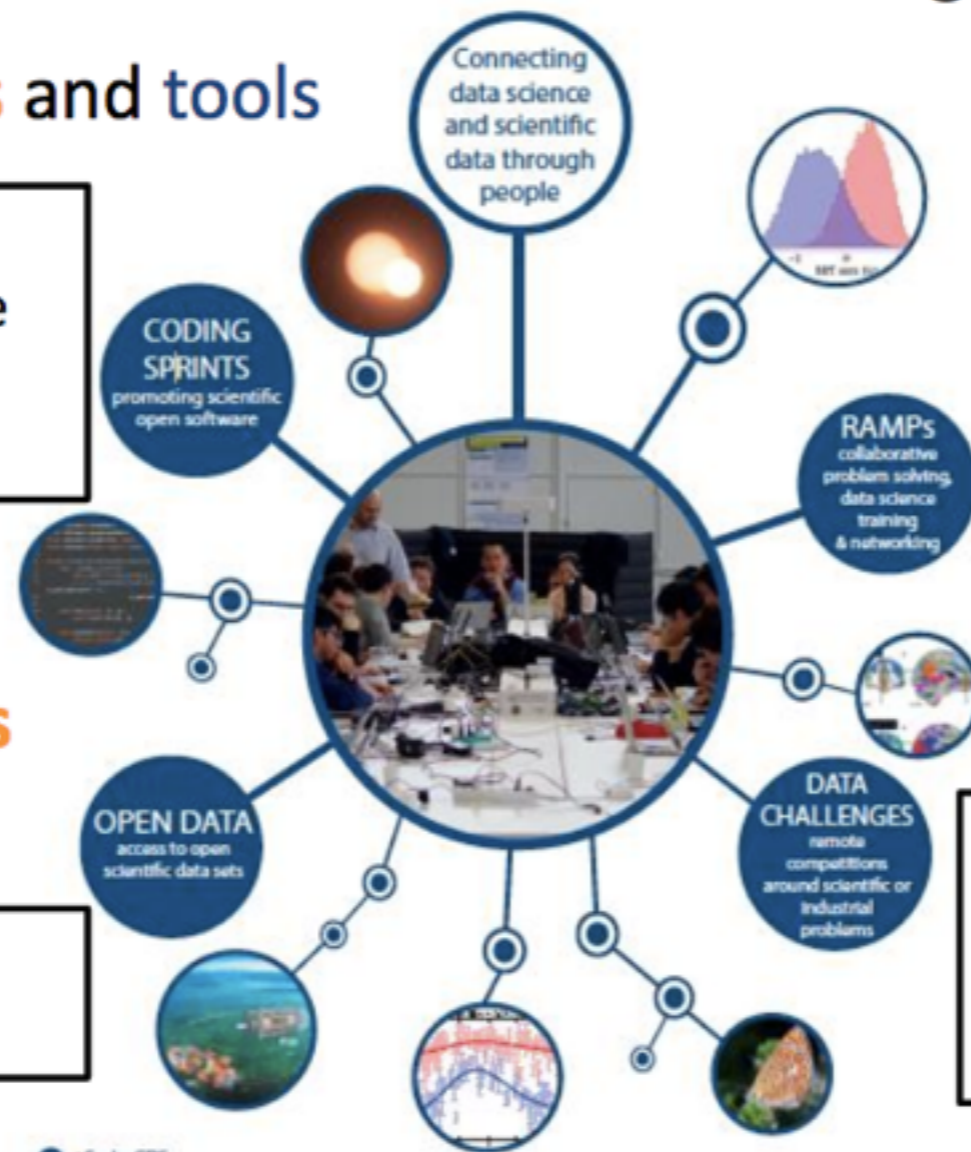
Connecting **experts** and **problems**

Connecting **experts** and **tools**

- State-of-the-art data science in easy-to-use tools
- High-quality software

Connecting **experts** and **data**

- Data as a Service
- Linked (Open) Data



- Prototyping
- Training
- Collaboration building

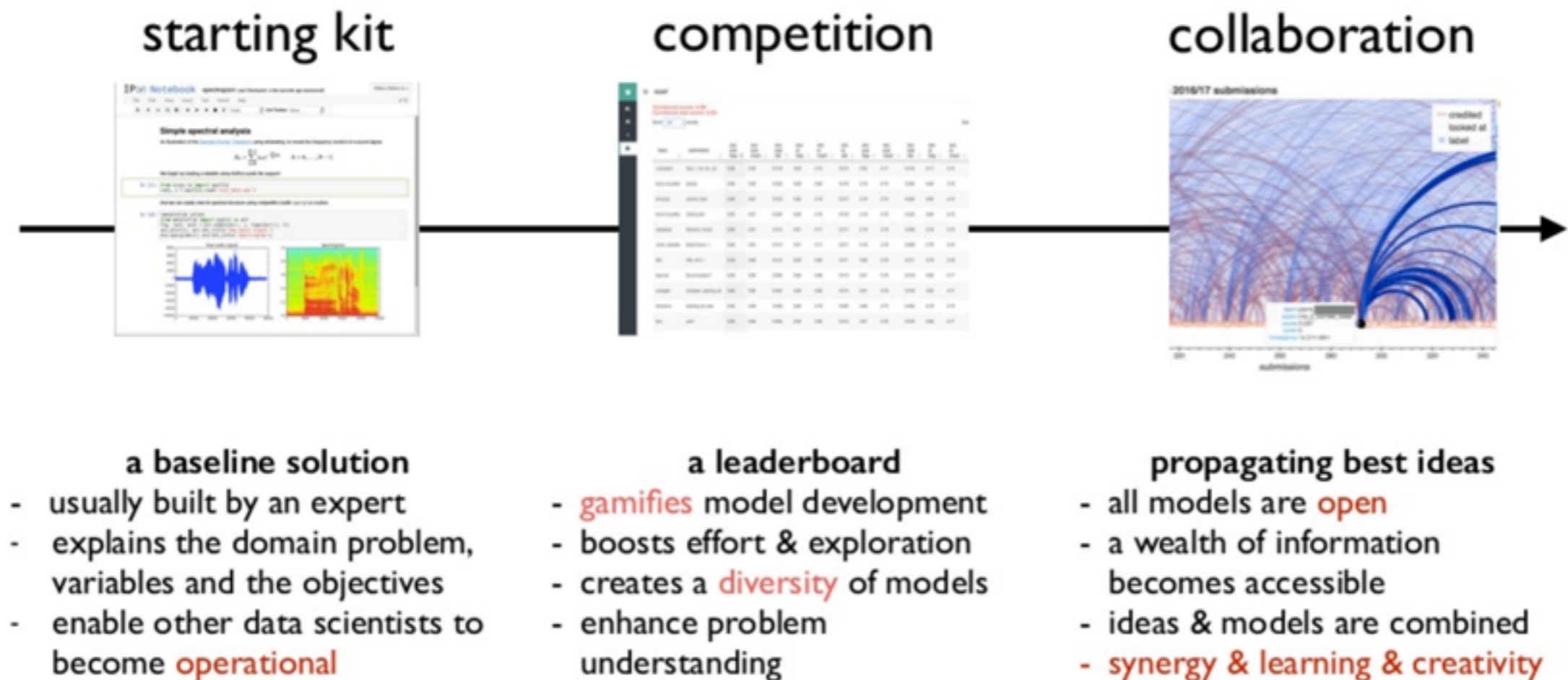
- Impact on science
- Visibility
- Benchmarks

@SaclayCDS

www.datascience-paris-saclay.fr

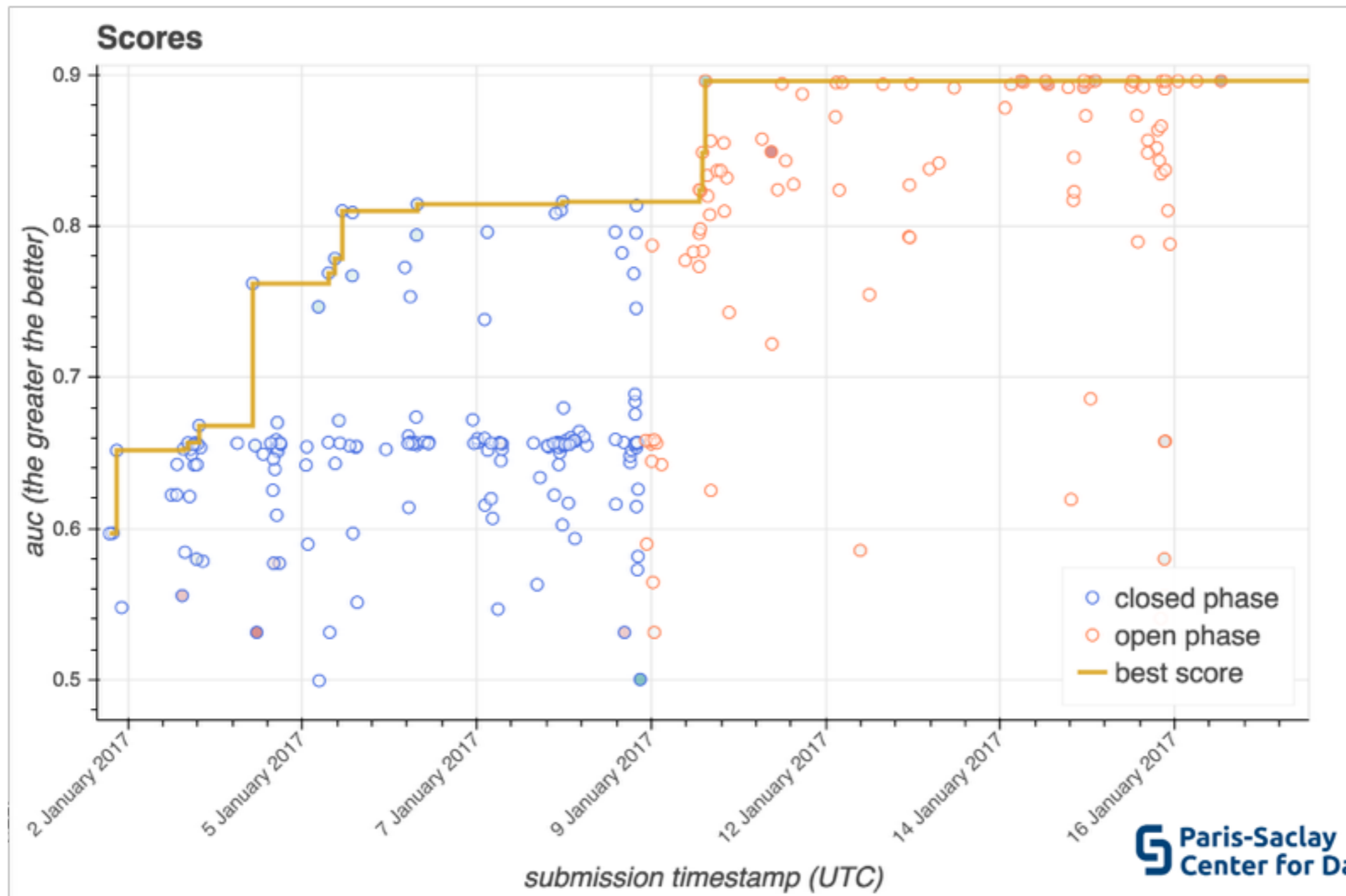
Use creativity to find the best solution

Rapid Analytics and Model Prototyping

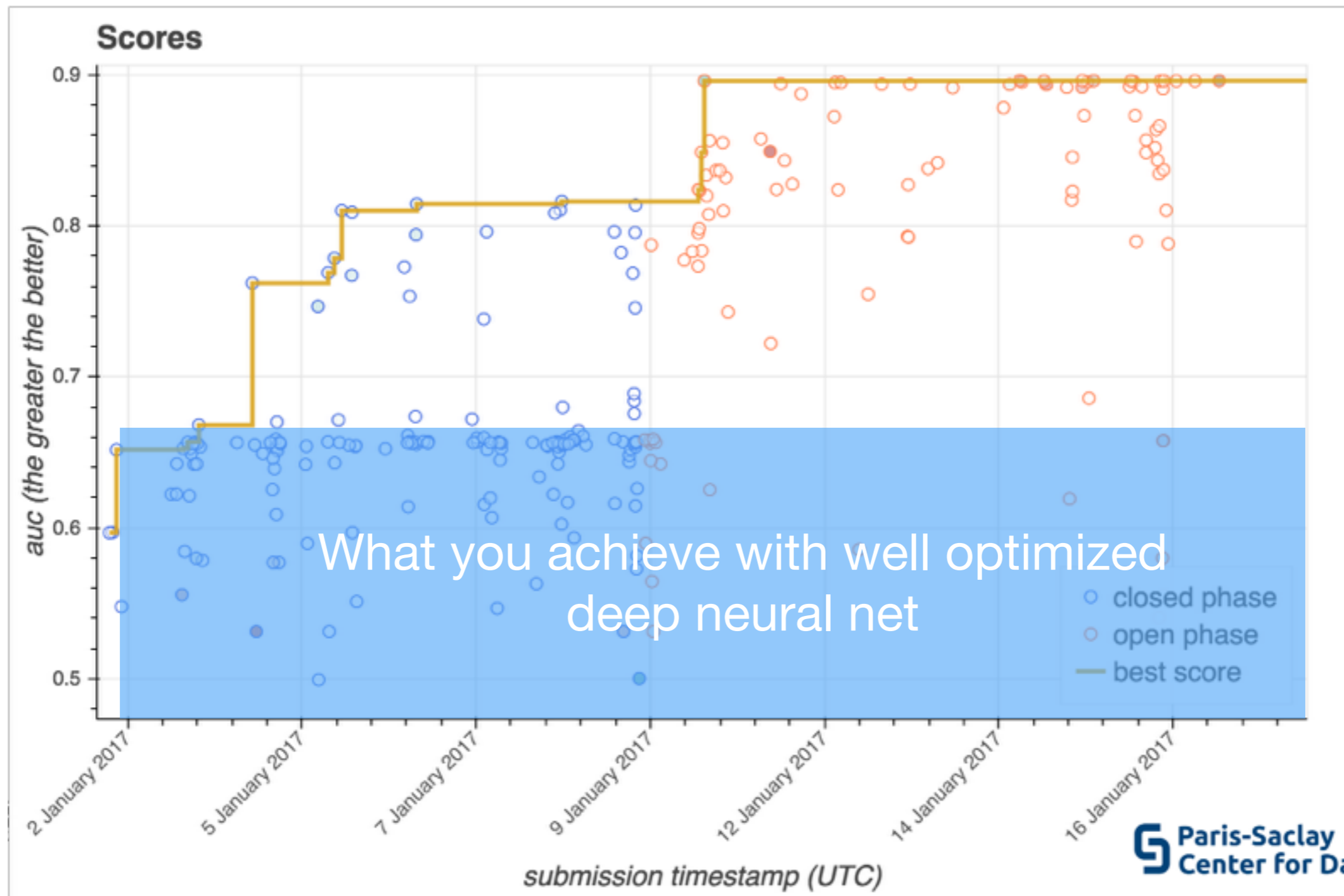


sign up at www.ramp.studio

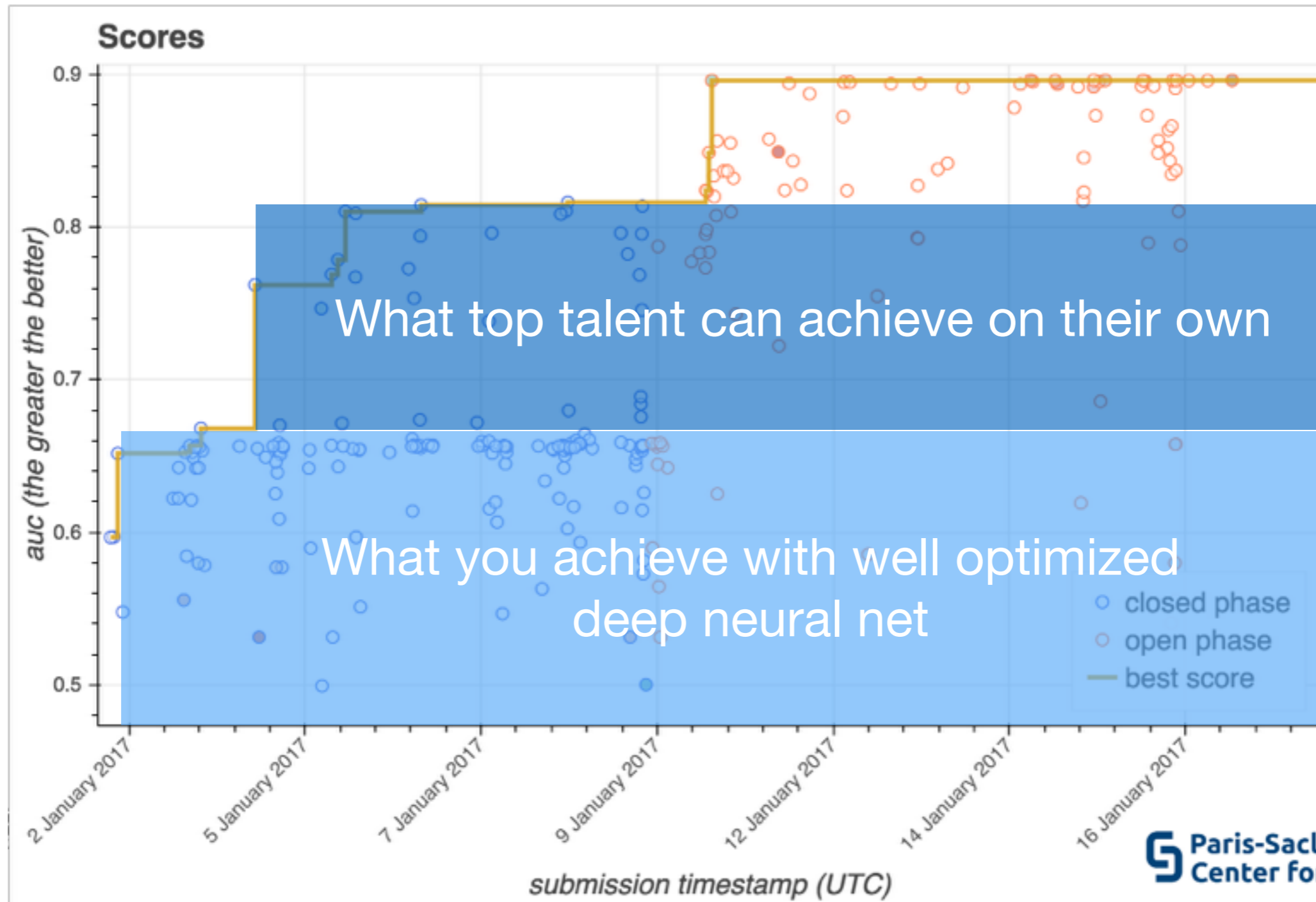
ATLAS anomaly detection RAMP



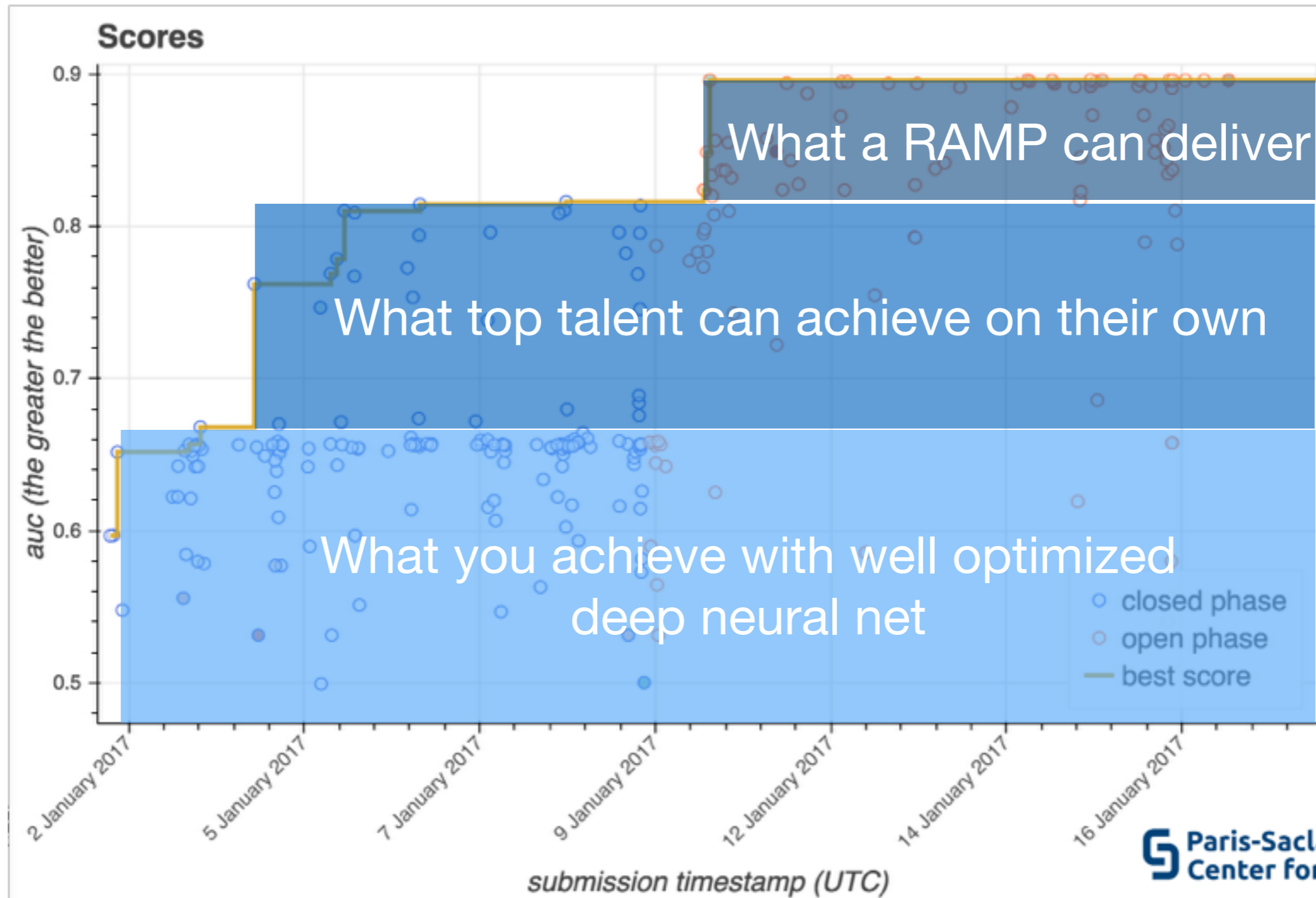
ATLAS anomaly detection RAMP



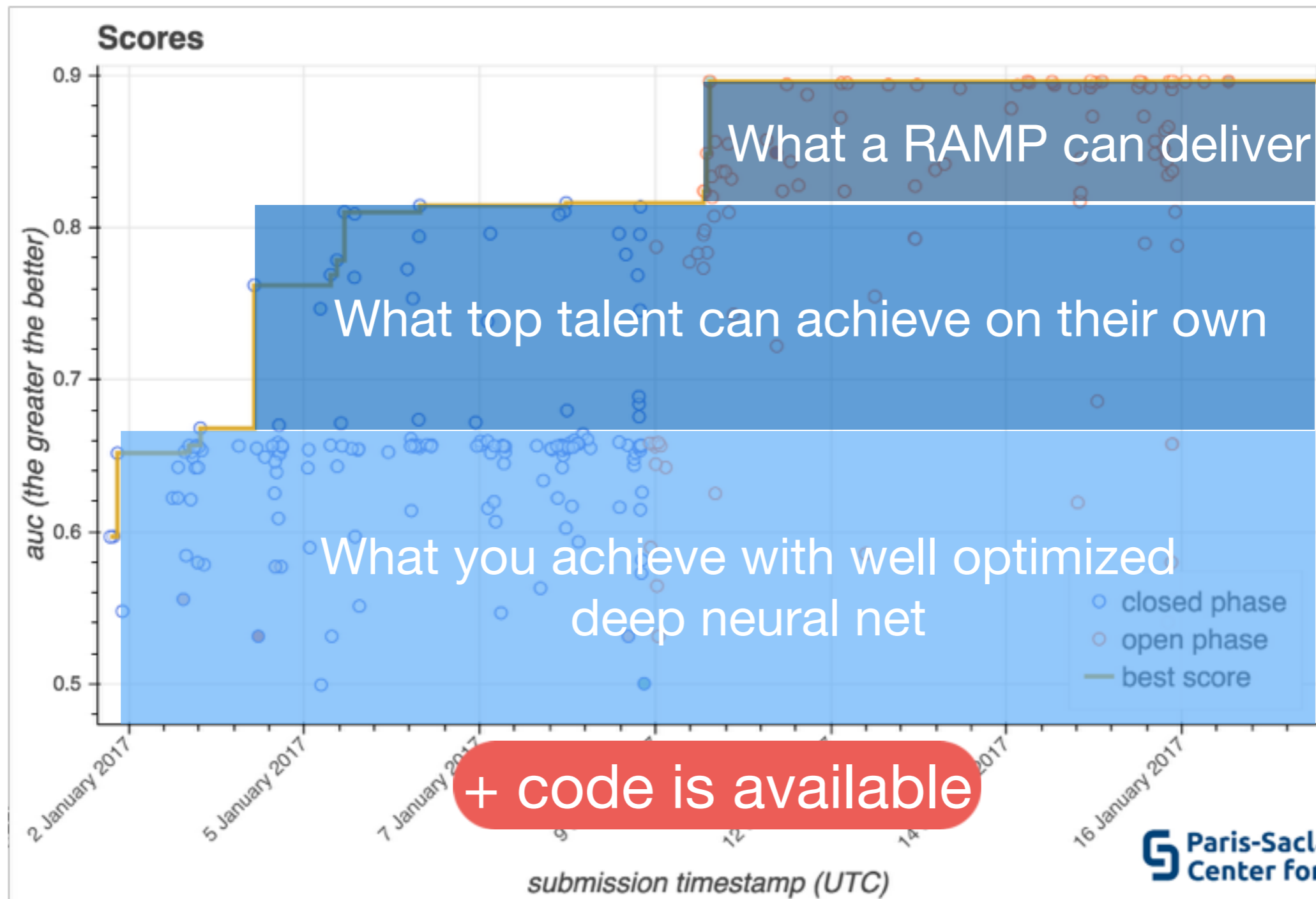
ATLAS anomaly detection RAMP



ATLAS anomaly detection RAMP



ATLAS anomaly detection RAMP



Take away message

Combination w/ machine learning

- CNNs could **speed-up the minimization process** of model fitting software
- neural networks have proved to **assist well human classification**

Knowledge transfer

- preliminary work shows that both **visual classification** and **parameter extraction** can be obtained **at a low cost** by re-training CNNs with **few labeled data**

Finding an optimal CNN architecture

- deep learning challenges centered on **active collaboration** and **open source** yield **unprecedented results** on **well defined** problems
- more challenges in astrophysics to come..