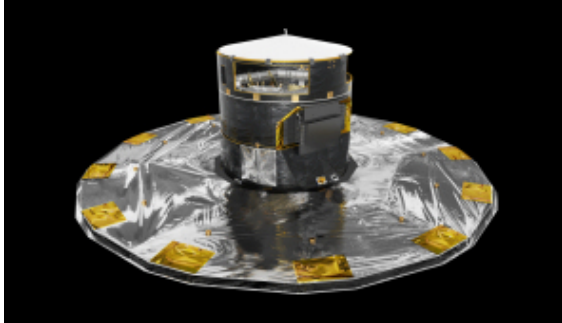# Training for the future, ML and AI

Shanshan Li, Chenzhou Cui, Yihan Tao

2019.5.15 Paris

Astronomy DATA

GAIA DR2

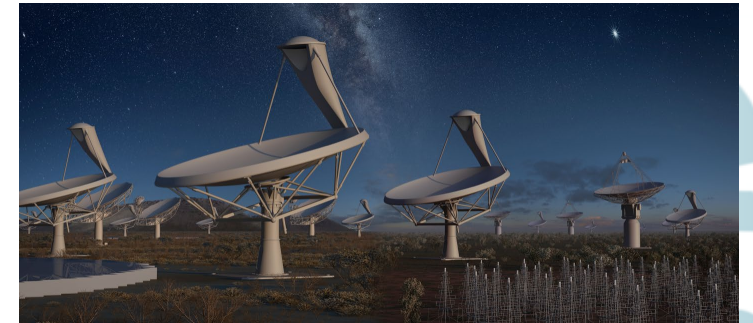Celestial positions and Gaia 'G' magnitudes for nearly 1.7 billion stars

LSST

15 Terabytes every night

FAST

10-20 Petabyte every year

SKA

Will generate more data traffic than the entire Internet.

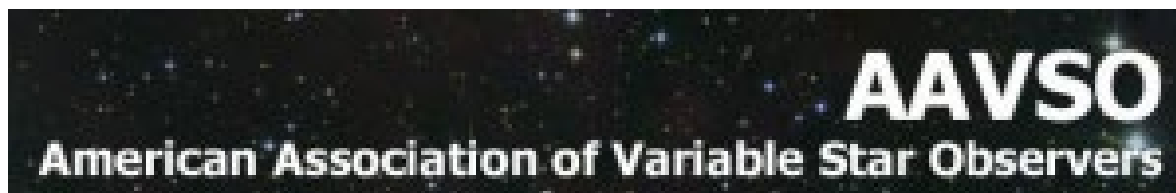**HUBBLE, SDSS, LAMOST, 2MASS, WISE, WMAP, IRIS, planck, …**

BIG MASSIVE TREASURE

Astronomy data in EPO

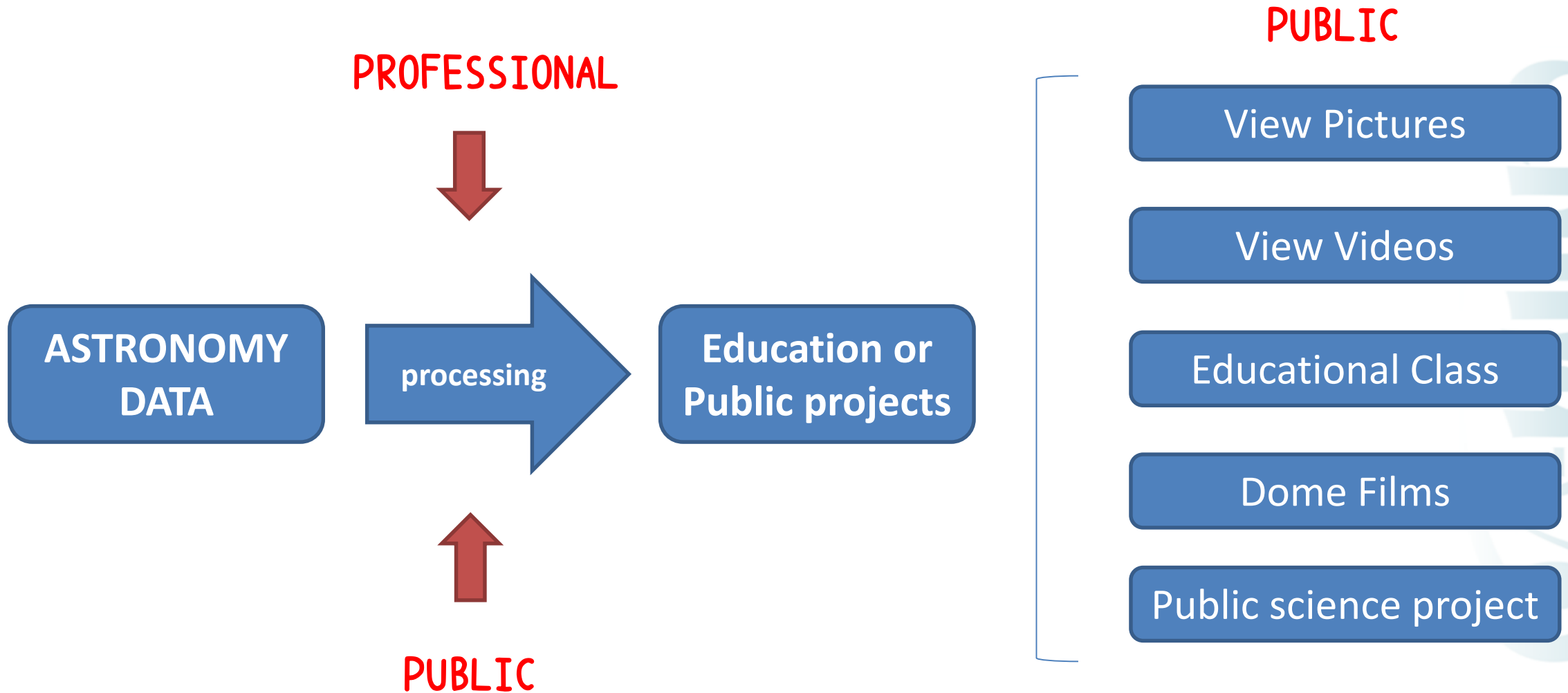- Data Driven Astronomy Education and Public Outreach (DAEPO) Best-Practices:

THINK

Idea begins

Scientific Research

Technological development

Data Platform

Education and Public outreach

National astronomical observatories (NAOC) and Aliyun started strategic cooperation in 2016

Astro Party

阿里天文派对-爬虫研究小组(33)

阿里天文派对-工程类问题小组(27)

阿里天文派对-大数据处理/算法(41)

NAOC-Aliyun Lab organized several astronomy parties.
Astronomers from many observatories and universities in China
together with programmers in ALI Group discussed problems
they interested in astronomy and learn from each other.

INSPIRED

Large Sky Area Multi-Object
Fiber Spectroscopic Telescope

Astro Data Mining Contest
Intelligent classification of celestial spectrum

AI FUTURELAB: Discover supernova

LAMOST

- **Large Sky Area Multi-Object Fiber Spectroscopic Telescope**
  - Highest spectral acquisition rate in the world.

- LAMOST DR6:
  - 4902 plates
  - More than 11,250,000 spectra

- Nearly 400 papers (2019/4)

Astro Data Mining Contest

- Target: Algorithm to classify spectrums from LAMOST DR3 **(STAR/GALAXY/QSO/UNKNOWN)**

- Participant: All public.

- Begin: Jan 19th, 2018

- First round: Participants download the data, debugging algorithm, and upload the result. (2 weeks)

- Second round: New data, upload result again. (2 days)

- Semi-final: Data E-mail to participants. They can use the platform resource to do the computation. (1 month)

Final score

Semi-final: 70%

Final presentation: 30%

DATA

867,702 processed LAMOST data sample was provided for training.
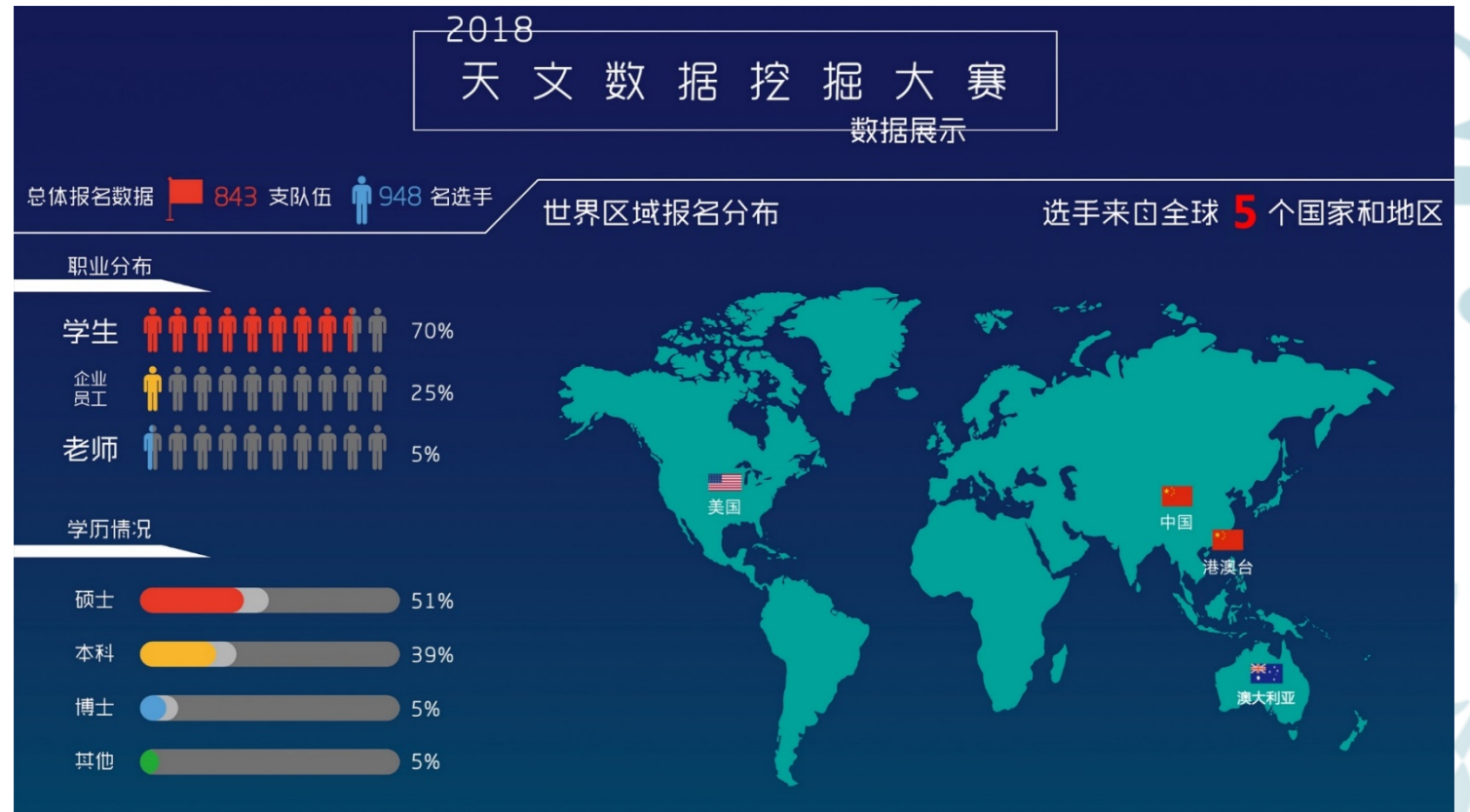
400,000 spectrum data provided for test and ranking.

Data understanding: 10%
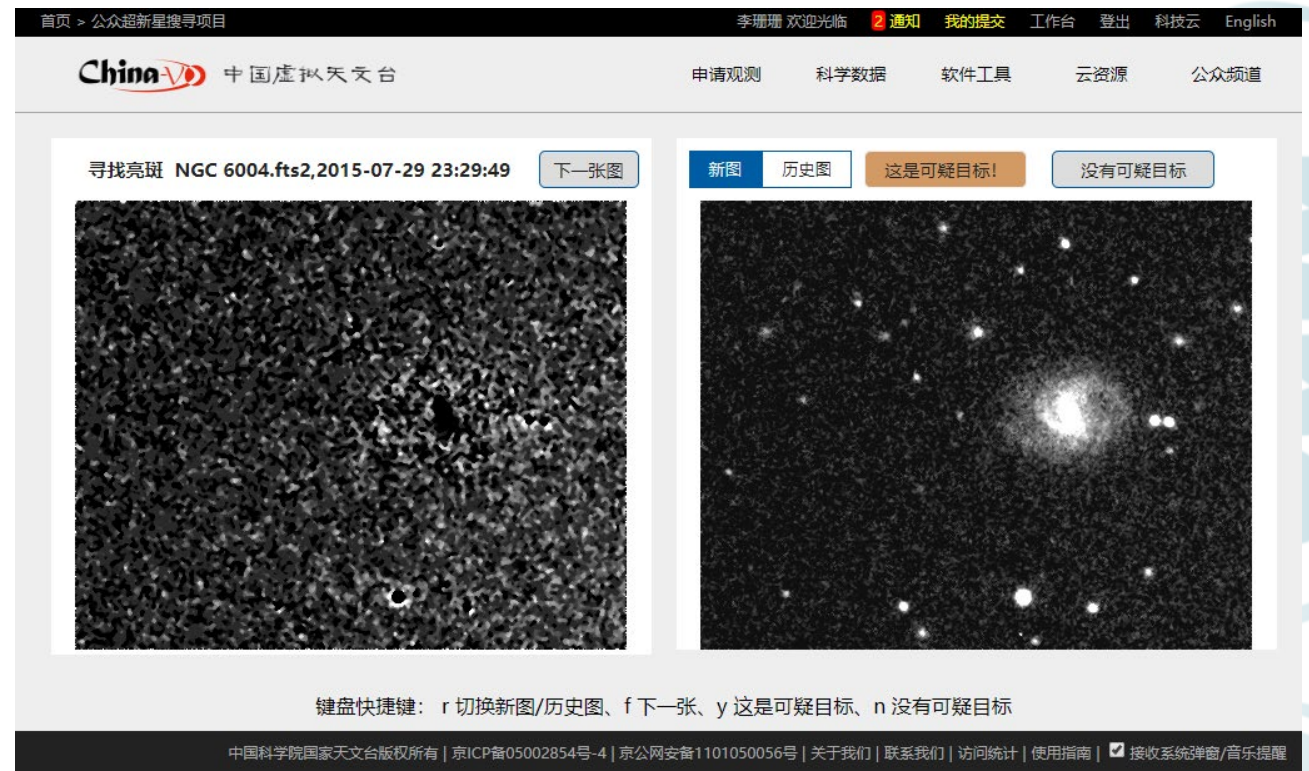
innovative: 10%

Practical: 10%

Summary

- Total 843 teams, 948 participants.

- 70% students, 25% company employees, 5% teachers.

- From 5 different areas around the world.

Popular Supernova Project

- A citizen science project started by Xingming Observatory and China-VO in 2015.

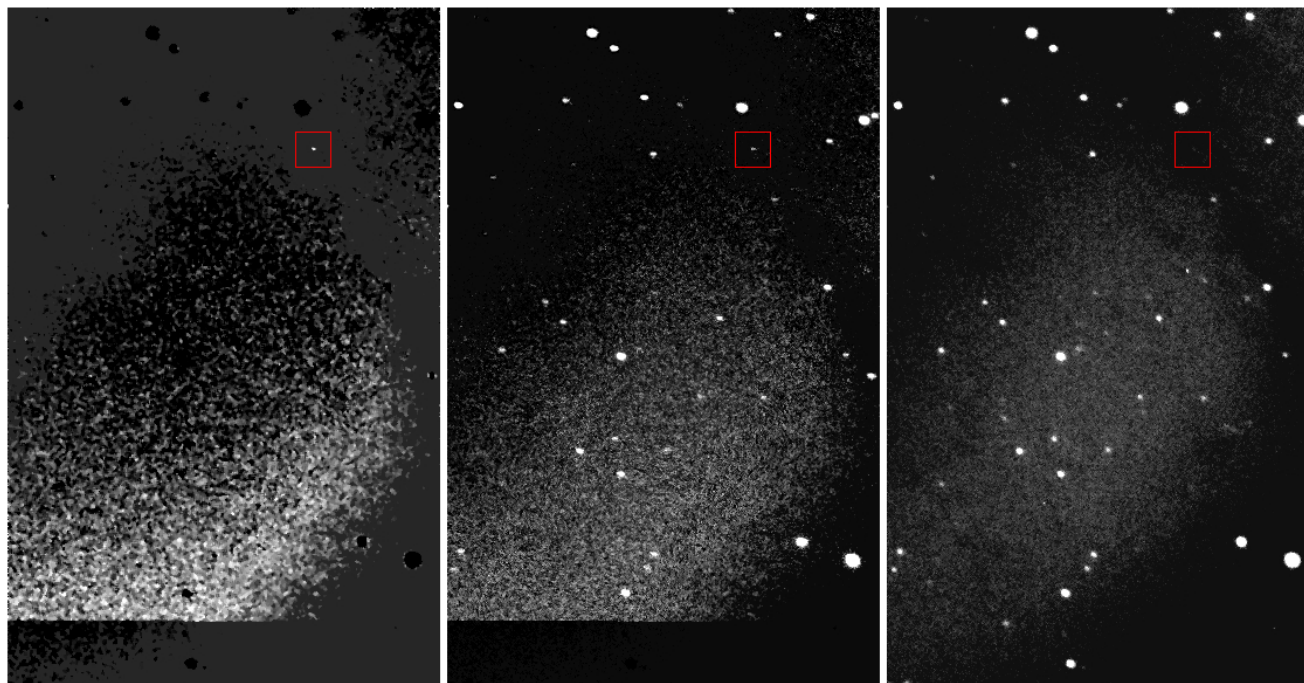- So far, 15 supernova and 4 nova was discovered through this project.

AI to find supernova

- Target: Design an algorithm to find supernova in pictures captured by telescope.

- Dataset: PSP images from 2015.

- Tagged objects： known, newtarget, isnova, isstar, asteroid

- Other tags: noise, ghost, pity

- 3 month

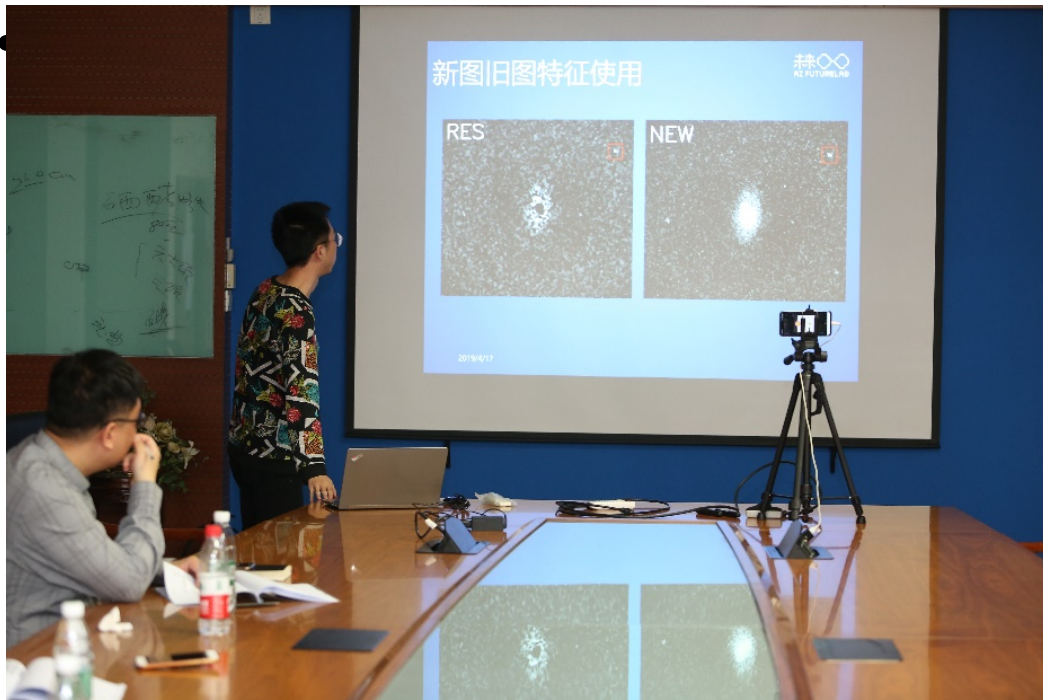2017/09/20170904/M31-18object/M31 A11.fts8
known-A-113

New image   history image   Difference

Participants

- 7 areas in China, 483 teams.

Summary

- Properly open scientific datasets, publish actual data problems to collect algorithms will help scientific research and popular science at the same time.

- For Public:
  - The participates may don't understand astronomy at all or don't have enough astronomy background to use the data.
  - Good chance to do DAEPO

- For scientist:
  - How will the winning algorithm help scientists.
  - Is there other problem or project can use the public's wisdom?