

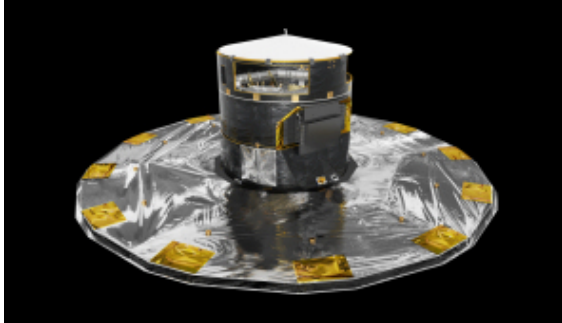


Training for the future, ML and AI

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



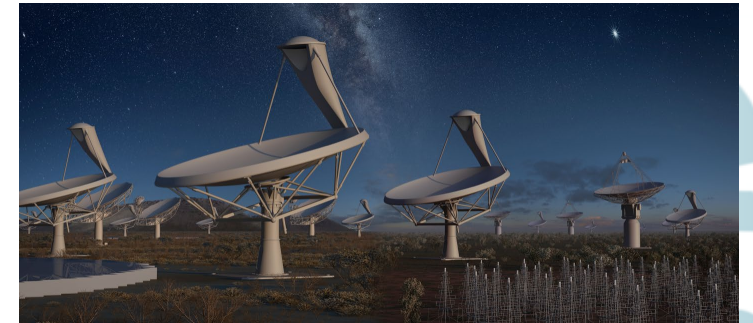
LSST

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

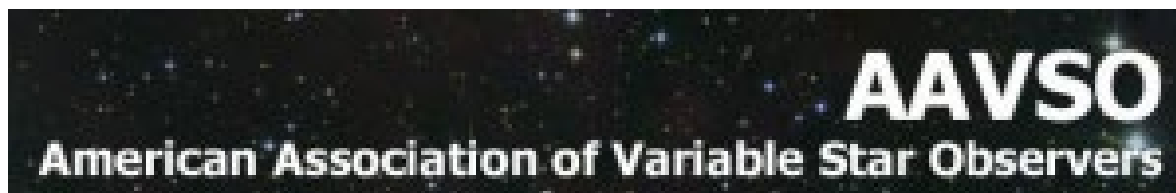
Astronomy data in EPO

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

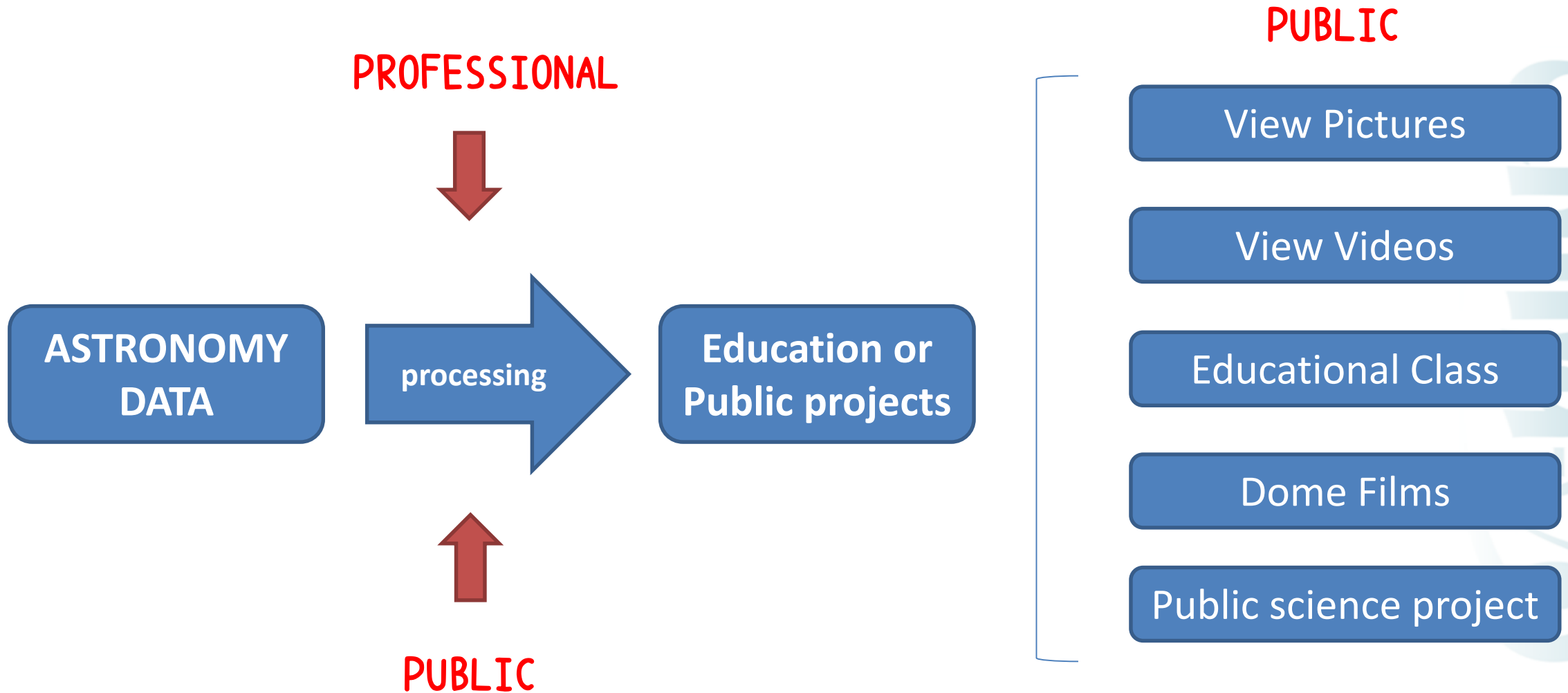


PULSE@Parkes

PULsar Student Exploration online at Parkes



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

Popular Supernova Project



AI FUTURELAB: Discover supernova

LAMOST

LAMOST Home Data Access Documents Help Contact Us LAMOST Signin

Large Sky Area Multi-Object Fiber Spectroscopic Telescope

Data Release 3
2011.10.24 - 2015.05.30

DR3 UPDATE (2017.06.30) NEW

Data Release 3 is the third data release of the LAMOST spectral survey, containing 5,755,126 spectra and 2,667 plates totally.

Catalog Search
Catalog Download
File download

Annual Statistics

Pilot Survey (2011.10.24-2012.06.17)		First Year Survey (2012.09.28-2013.06.03)		Second Year Survey (2013.09.10-2014.06.03)		Third Year Survey (2014.09.10-2015.05.30)	
Total Spectra	906,420	Total Spectra	1,637,842	Total Spectra	1,588,520	Total Spectra	1,622,344
Star	807,575	Star	1,509,790	Star	1,462,309	Star	1,489,013
Galaxy	2,754	Galaxy	9,555	Galaxy	25,356	Galaxy	24,150
QSO	618	QSO	4,094	QSO	3,918	QSO	7,721
Unknown	95,473	Unknown	114,403	Unknown	96,937	Unknown	101,460

- 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.

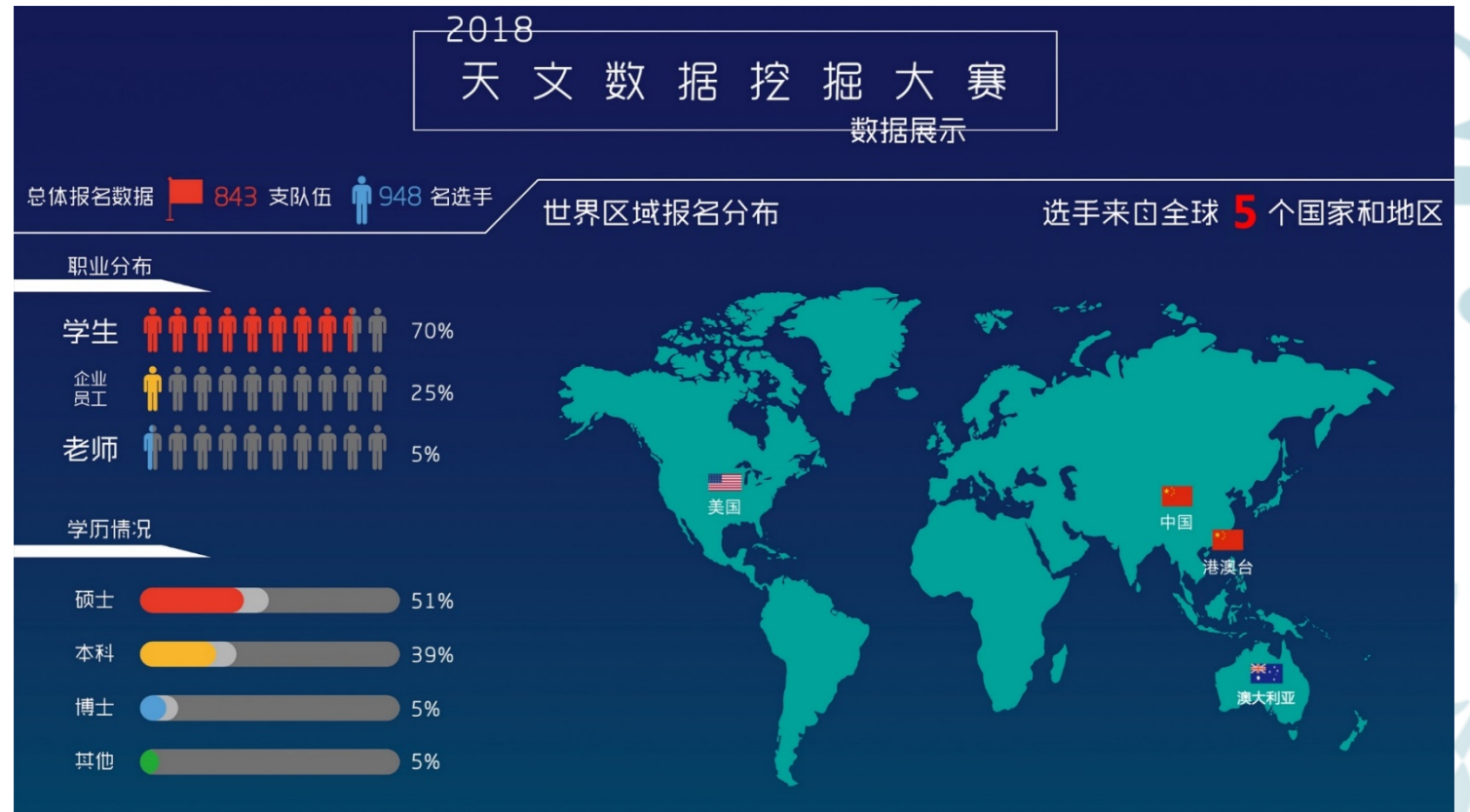
Data understanding: **10%**

innovative: **10%**

Practical: **10%**

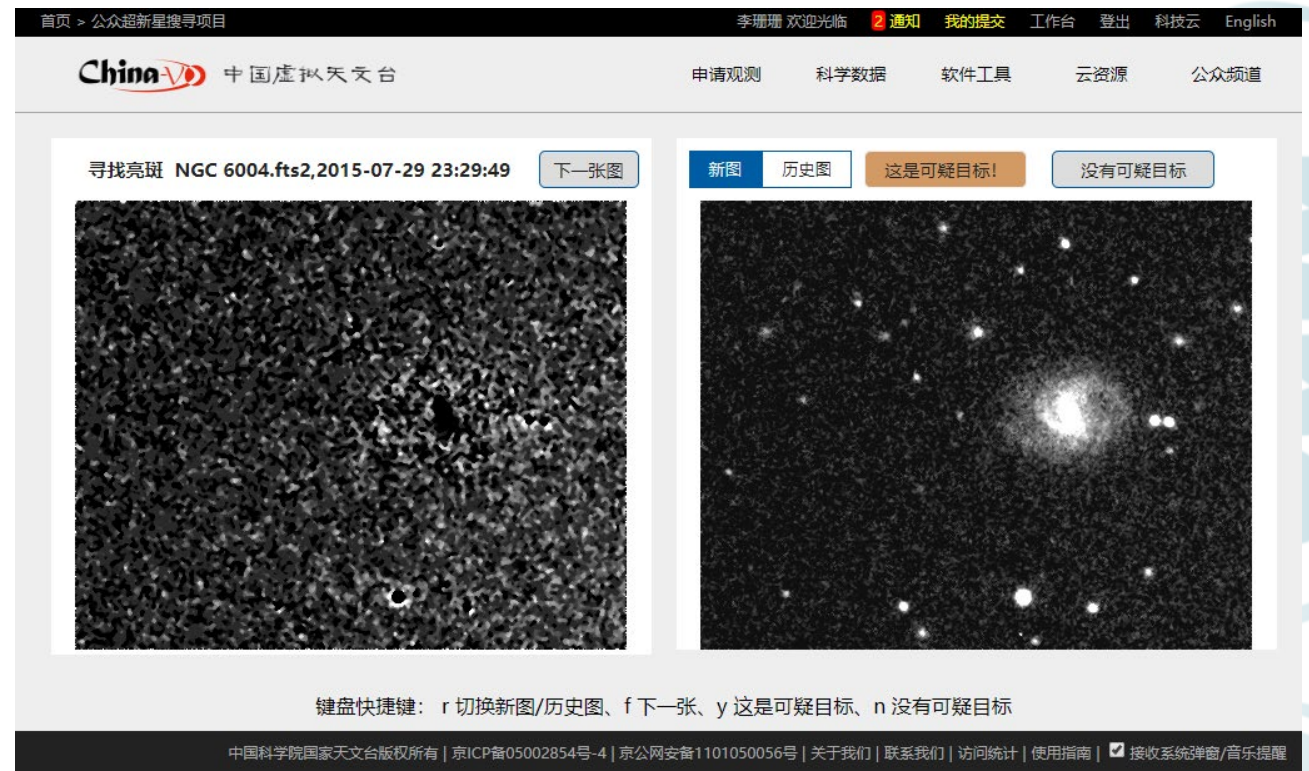
Summary

- 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.

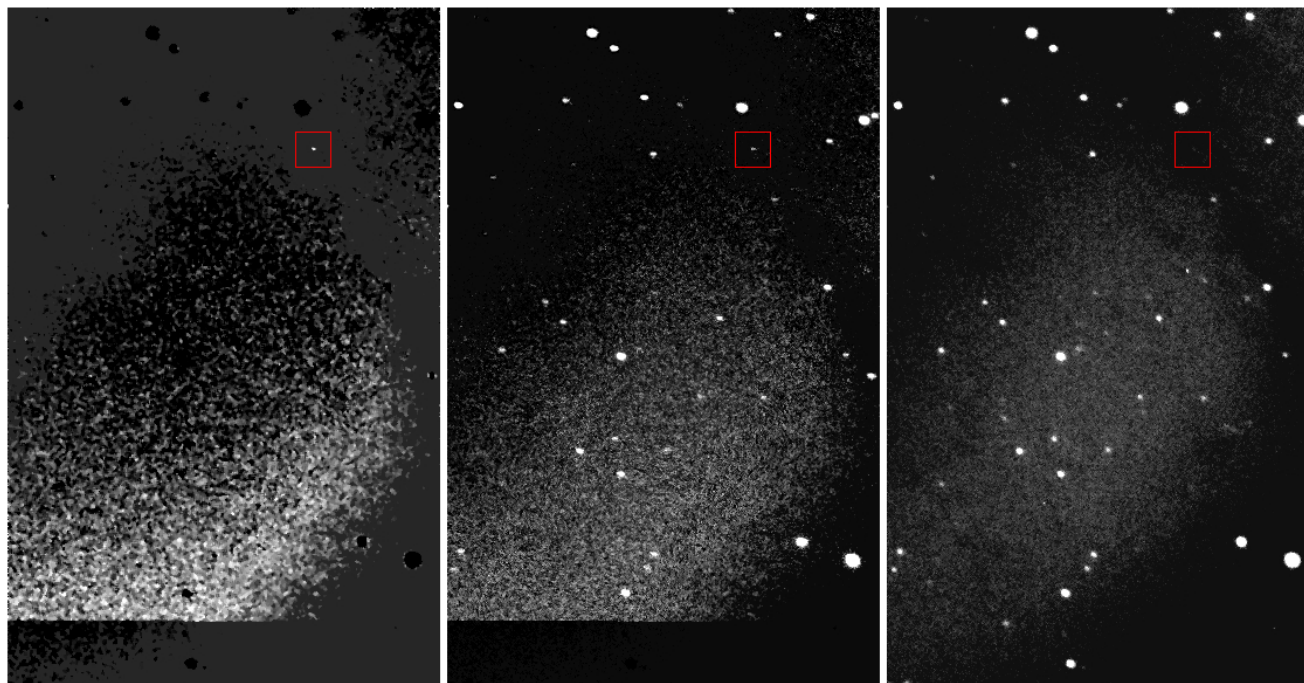


AI to find supernova

- 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



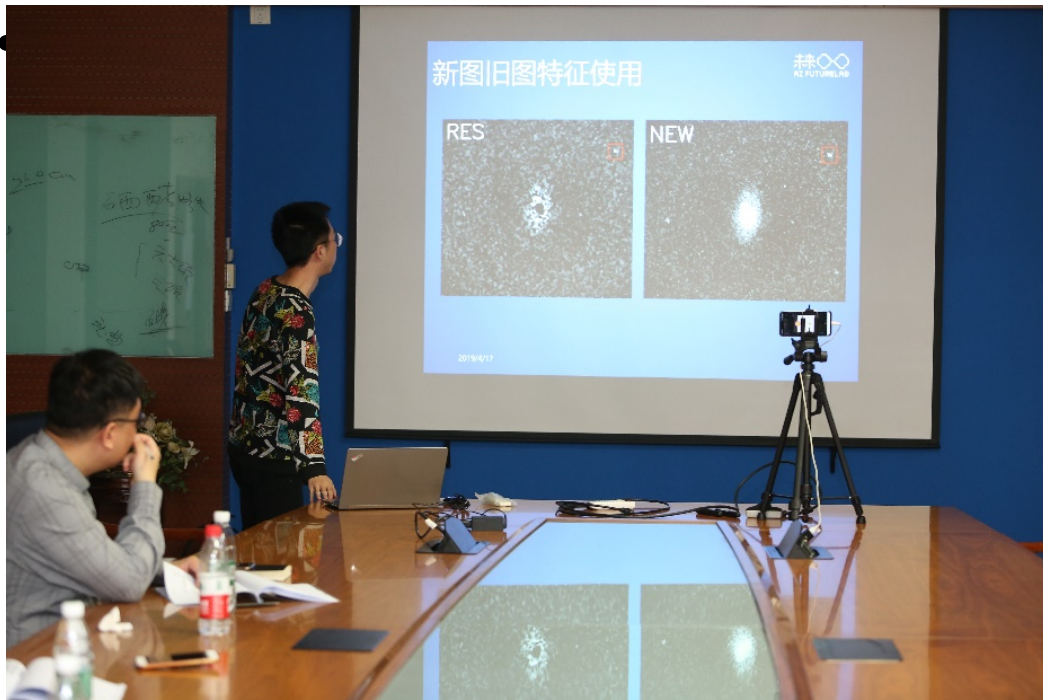
New image

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 participants 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?

