Active Anomaly Detection for time-domain discoveries

arXiv:astro-ph/1909.13260

4th Inter-experiment Machine Learning Workshop - CERN
23 October 2020

Emille E. O. Ishida

Laboratoire de Physique de Clermont - Université Clermont-Auvergne
Clermont Ferrand, France

Summary

- Context
- Anomaly detection
- Human in the loop
- Applications to astronomy
  - Simulations
  - Real data
- Conclusions
- Implications

Astronomy has been, traditionally, an experience of solitude

The old astronomer, *by Charlie Bowater*

Context ...

Astronomy has been, traditionally, an experience of solitude

Meaning ...

Discovery happened by chance

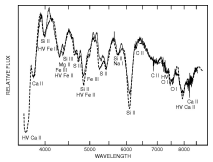
or

By careful analysis of small volume of highly informative data

The old astronomer, *by Charlie Bowater*

2 types of Astronomical data

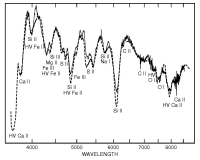
Spectroscopic



- Pros:
  - High resolution
  - Large information content
  - Enable astrophysical analysis
- Cons:
  - Super expensive
  - Strong requirements on observation conditions

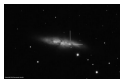
2 types of Astronomical data

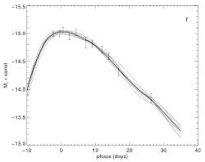
Spectroscopic



- Pros:
  - High resolution
  - Large information content
  - Enable astrophysical analysis
- Cons:
  - Super expensive
  - Strong requirements on observation conditions

Photometry

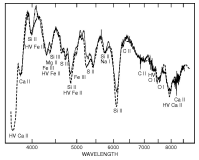




- Pros:
  - Easy to obtain
  - Allows environmental and morphological analysis
  - Enables time domain research of a large number of objects
- Cons:
  - Low resolution (integration over large wavelength range)

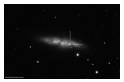
2 types of Astronomical data

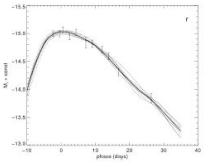
Spectroscopic



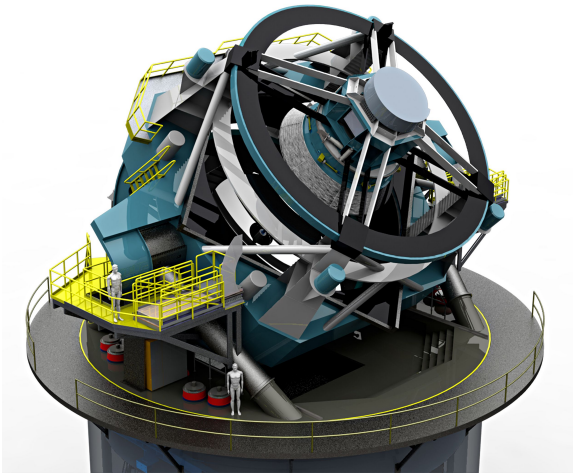
- Pros:
  - High resolution
  - Large information content
  - Enable astrophysical analysis
- Cons:
  - Super expensive
  - Strong requirements on observation conditions

Photometry





- Pros:
  - Easy to obtain
  - Allows environmental and morphological analysis
  - Enables time domain research of a large number of objects
- Cons:
  - Low resolution (integration over large wavelength range)



The Vera Rubin Observatory Legacy Survey of Space and Time (LSST)

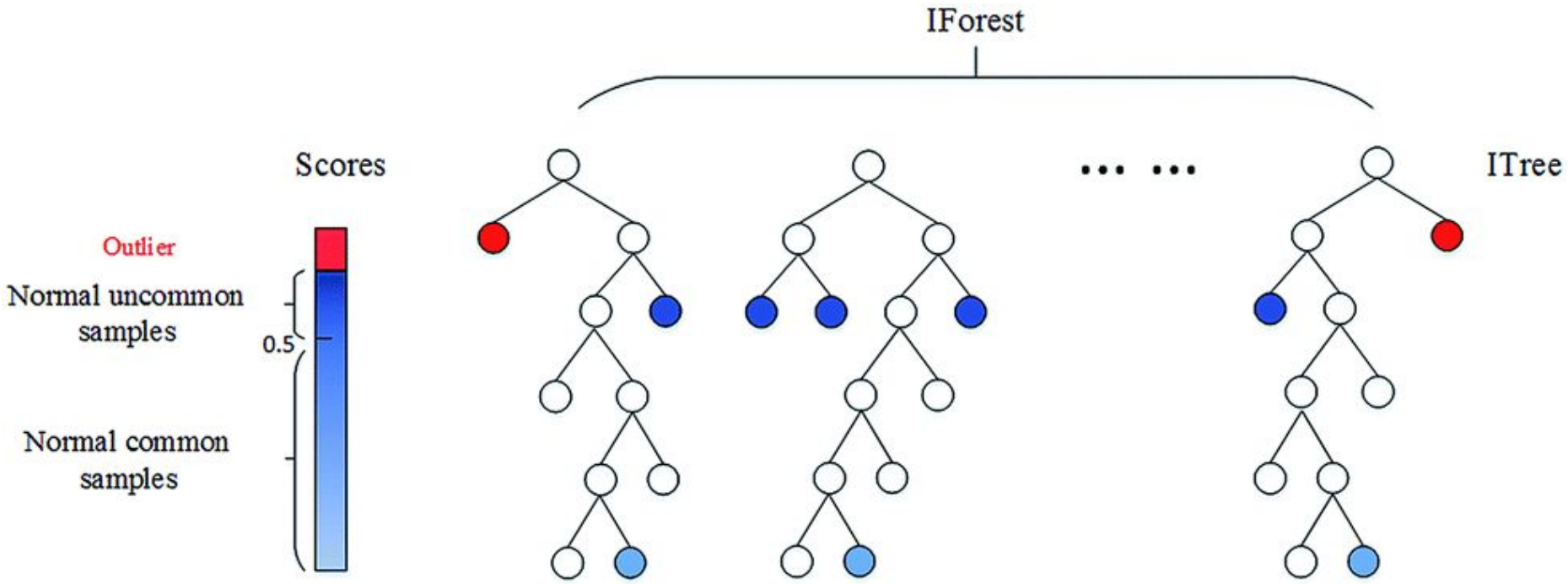
~10 million candidates/night

Over a total life span of 10 years

Serendipitous discoveries will not happen...

https://snad.space/

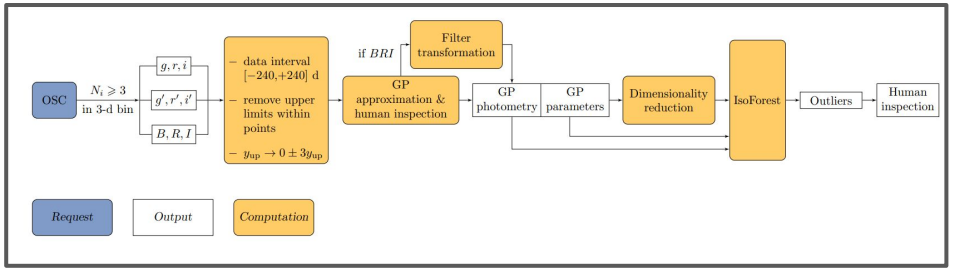
Isolation Forest

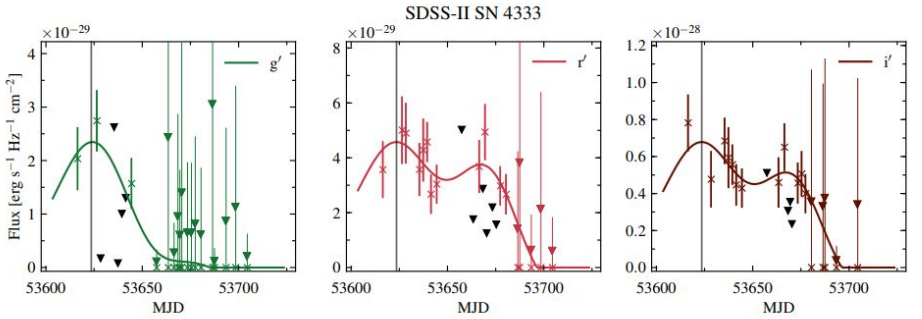


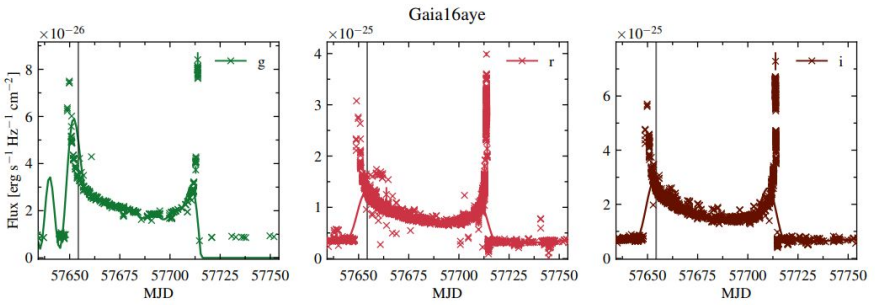
https://donghwa-kim.github.io/iforest.html

First try: the Open Supernova Catalog

- Real data:
  - Uncertainties
  - Upper limits
  - Different filter sets
- Pre-processing:
  - Filter translation
  - Selection cuts
  - 2D Gaussian Process
  - 3 sets: photo, photo + GP param, tSNE
- Data and analysis:
  - Initial data: 2000 light curves
  - Anomaly detection via Isolation Forest
  - Visually inspected 2% in each set (~100 objs)
- Results:
  - 81 identified anomalies
  - SLSN, peculiar SNe, miss-classified stars
  - 1 AGN and 1 binary micro-lensing



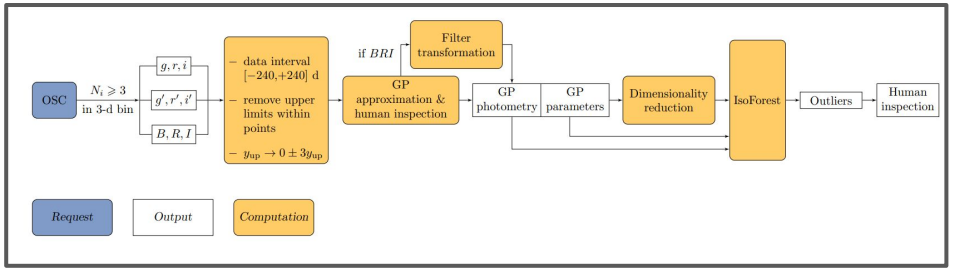


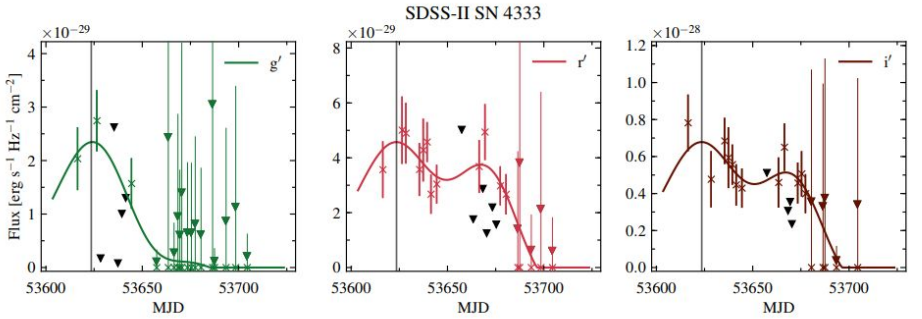


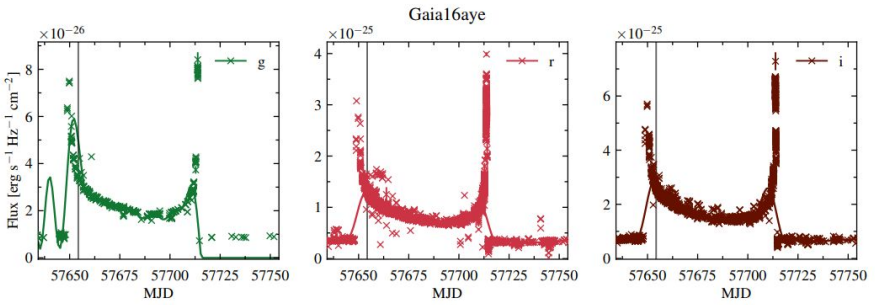
Pruzhinskaya et al., 2019 - MNRAS - https://arxiv.org/abs/1905.11516

First try: the Open Supernova Catalog

- Real data:
  - Uncertainties
  - Upper limits
  - Different filter sets
- Pre-processing:
  - Filter translation
  - Selection cuts
  - 2D Gaussian Process
  - 3 sets: photo, photo + GP param, tSNE
- Data and analysis:
  - Initial data: 2000 light curves
  - Anomaly detection via Isolation Forest
  - Visually inspected 2% in each set (~100 objs)
- Results:
  - 81 identified anomalies
  - SLSN, peculiar SNe, miss-classified stars
  - 1 AGN and 1 binary micro-lensing





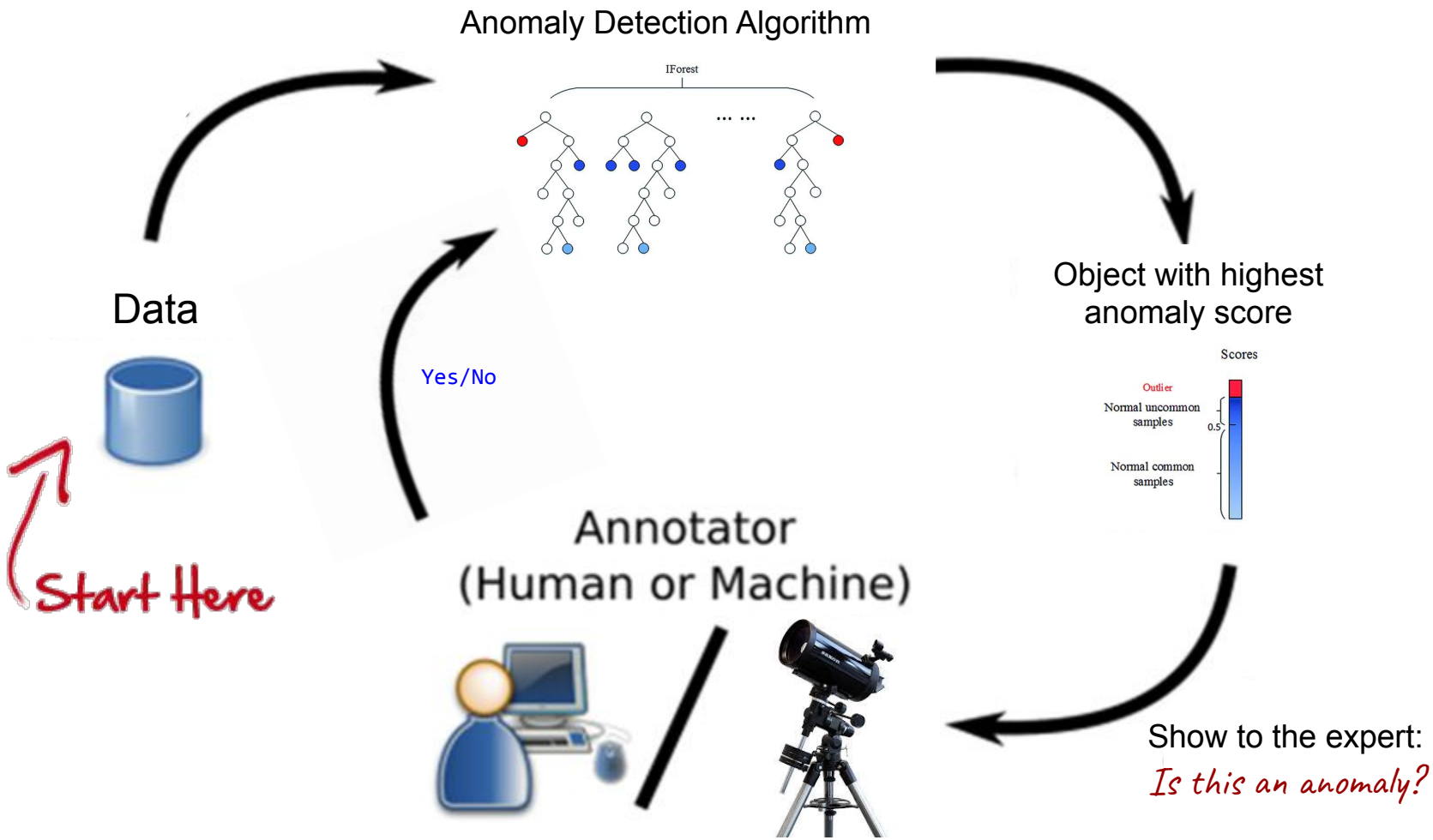


Problem:

What is scientifically interesting is in the eye of the beholder.

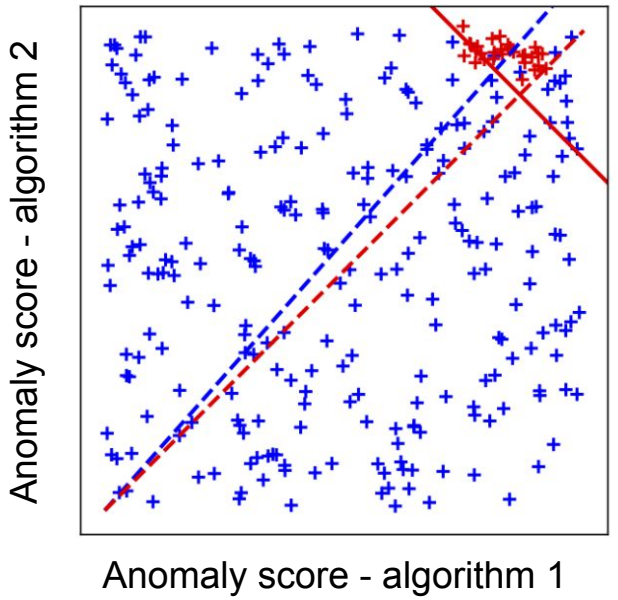
Pruzhinskaya et al., 2019 - MNRAS - https://arxiv.org/abs/1905.11516

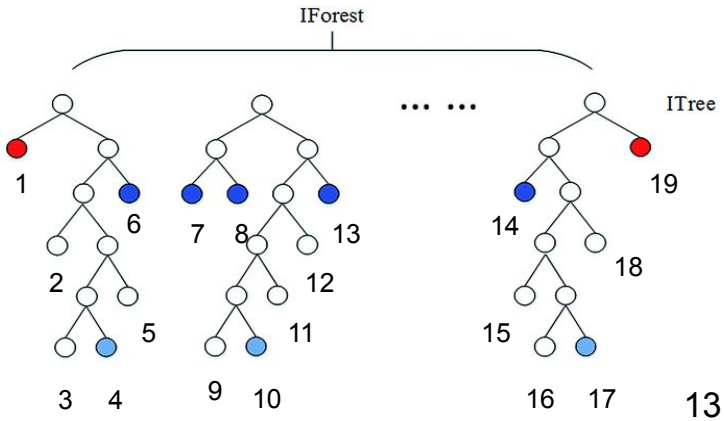
Active Learning for Anomaly Detection



Ensemble learning and expert feedback

- Ensemble learning: when you do not know, ask around!
- 2 perfectly accurate AD algorithms will agree on the scores for true anomalies
- In the real case one will be more accurate than the other, so we need to assign **weights**
- **Active Anomaly Discovery:**
  - Start with a normal Isolation Forest
  - Consider each decision path leading to a leaf node as an weak AD algorithm (ensemble member)
  - Assign an equal weight to each ensemble member
    - Show the most anomalous obj to the expert
    - If *expert_answer* == *yes*:
      Show next obj with highest anomaly score
      else:
      Update weights

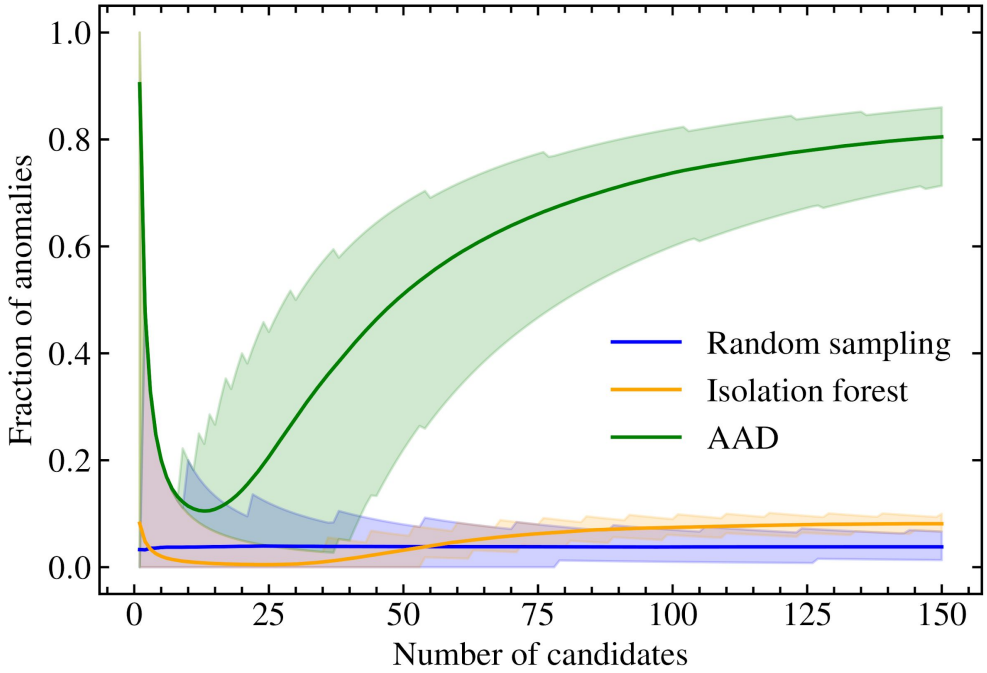




Das, S., Wong, W-K., Dietterich, T., Fern, A. and Emmott, A. (2016). *Incorporating Expert Feedback into Active Anomaly Discovery*

Simulations: the PLAsTiCC data set

- Data from the Kaggle PLAsTiCC data set restricted to Supernova-like events
- Initial sample: ~ 7000 light curves, 3 known classes, 3 peculiar classes (277 anomalies, 4%)
- 145 objects scrutinized (~2%), on average:
  - **Random Sampling: 5** real anomalies
  - **Isolation Forest: 12** real anomalies
  - **Active Anomaly Discovery: 120** real anomalies



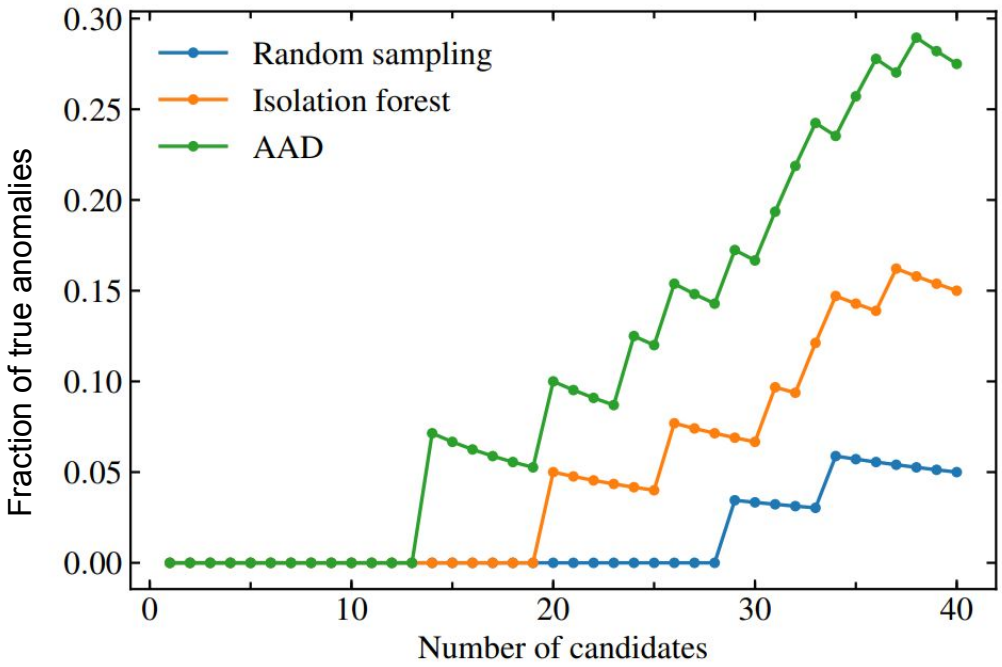
Ishida et al., 2019 - https://arxiv.org/pdf/1909.13260v1.pdf

Real data: The Open Supernova Catalog

- **Anomaly**:
  - Miss-classification (non-SNe)
  - Unusual light curve behavior
  - Previously known
    91bg-like and 91T-like

- **Not-anomaly:**
  - Bad Gaussian process fitting
  - Not enough signal
  - Identified artifacts

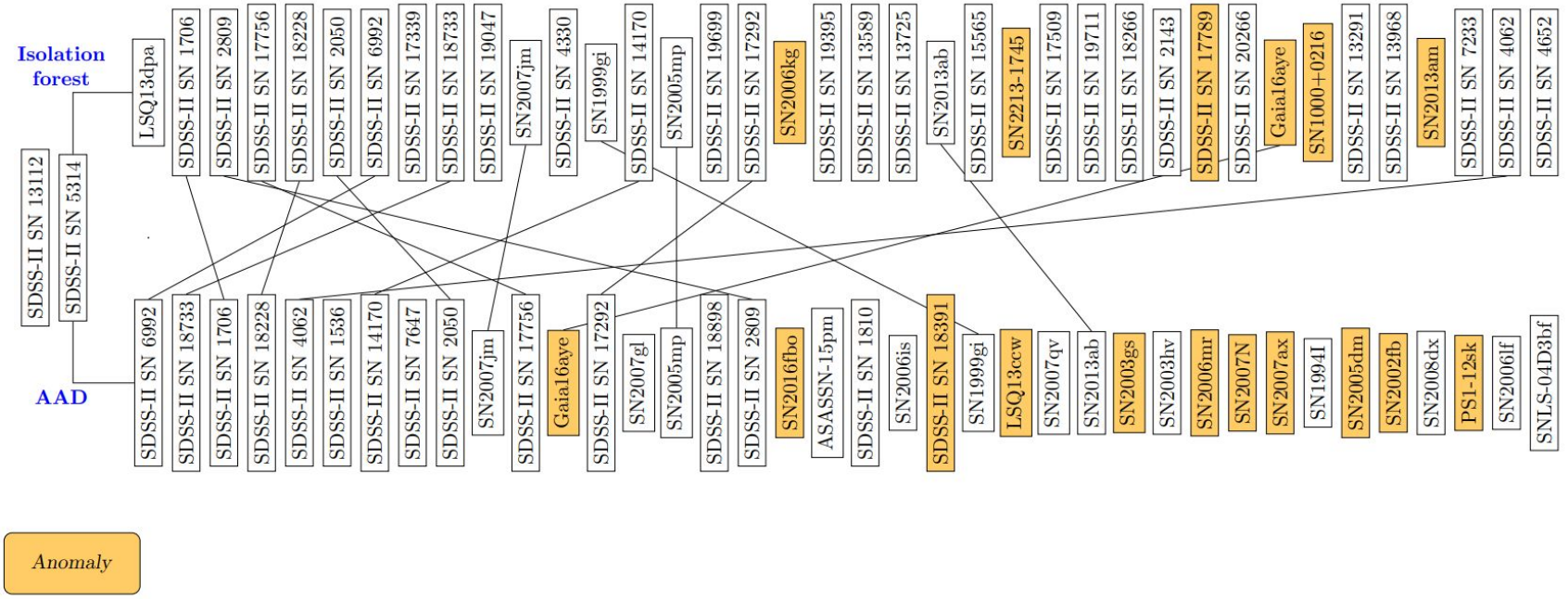
- **Results within 2% contamination (40 objs):**
  - Random sampling: 2 (5%)
  - Isolation Forest: 5 (15%)
  - Active Anomaly Discovery: 11 (27%)



AAD was able to increase the incidence of true anomalies presented to the expert in 80%

Ishida et al., 2019 - https://arxiv.org/pdf/1909.13260v1.pdf

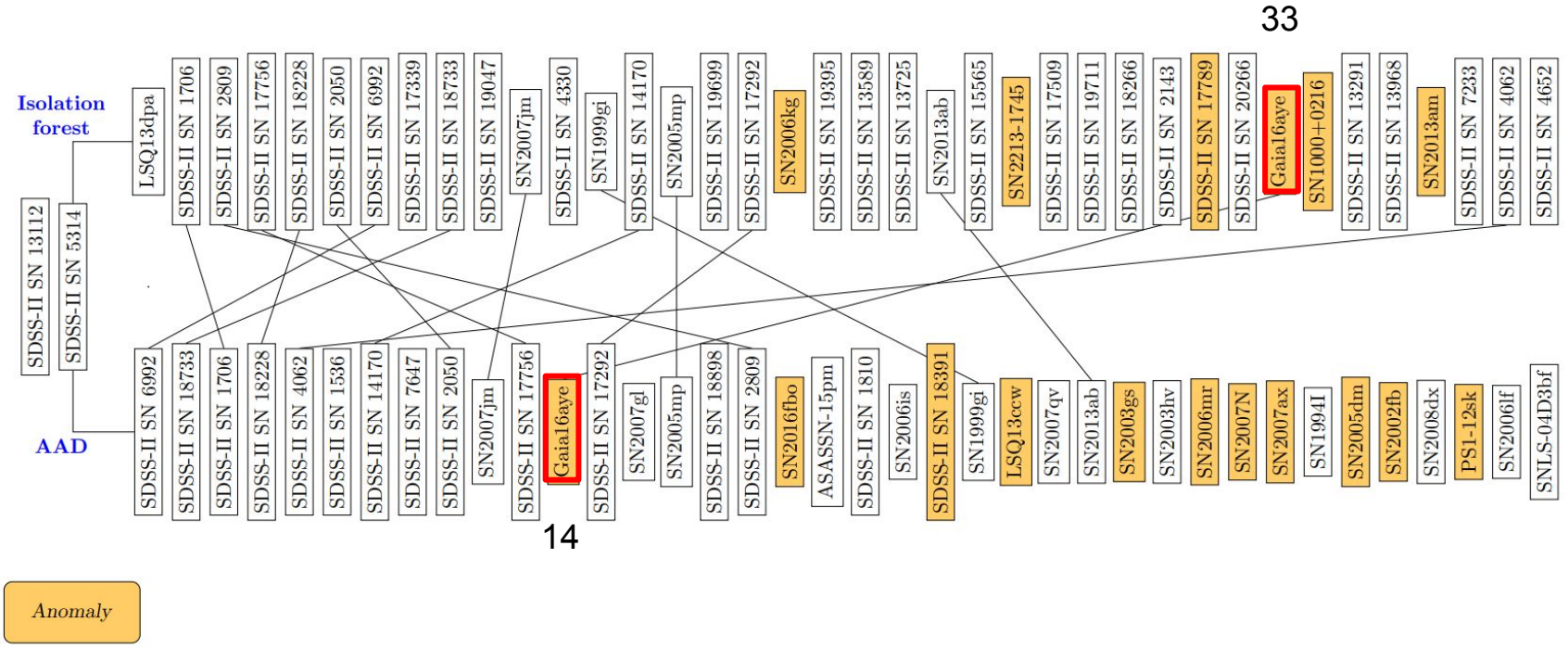
Real data: The Open Supernova Catalog



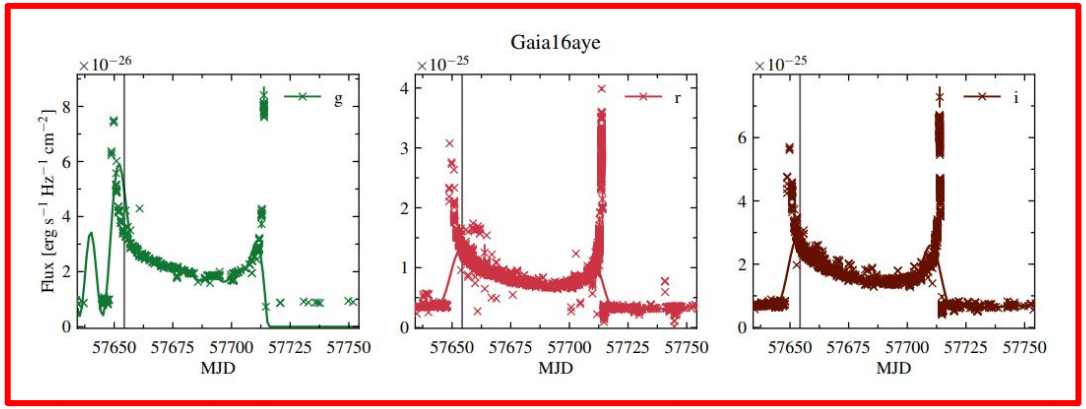
- It requires some time for changes to be effectively incorporated
- Late queries:
  - Objects which were not found in the static case
  - Higher concentration of true anomalies

Ishida et al., 2019 - https://arxiv.org/pdf/1909.13260v1.pdf

Active Anomaly Detection for time-domain discoveries



Fast identification of binary microlensing event



Ishida et al., 2019 - https://arxiv.org/pdf/1909.13260v1.pdf

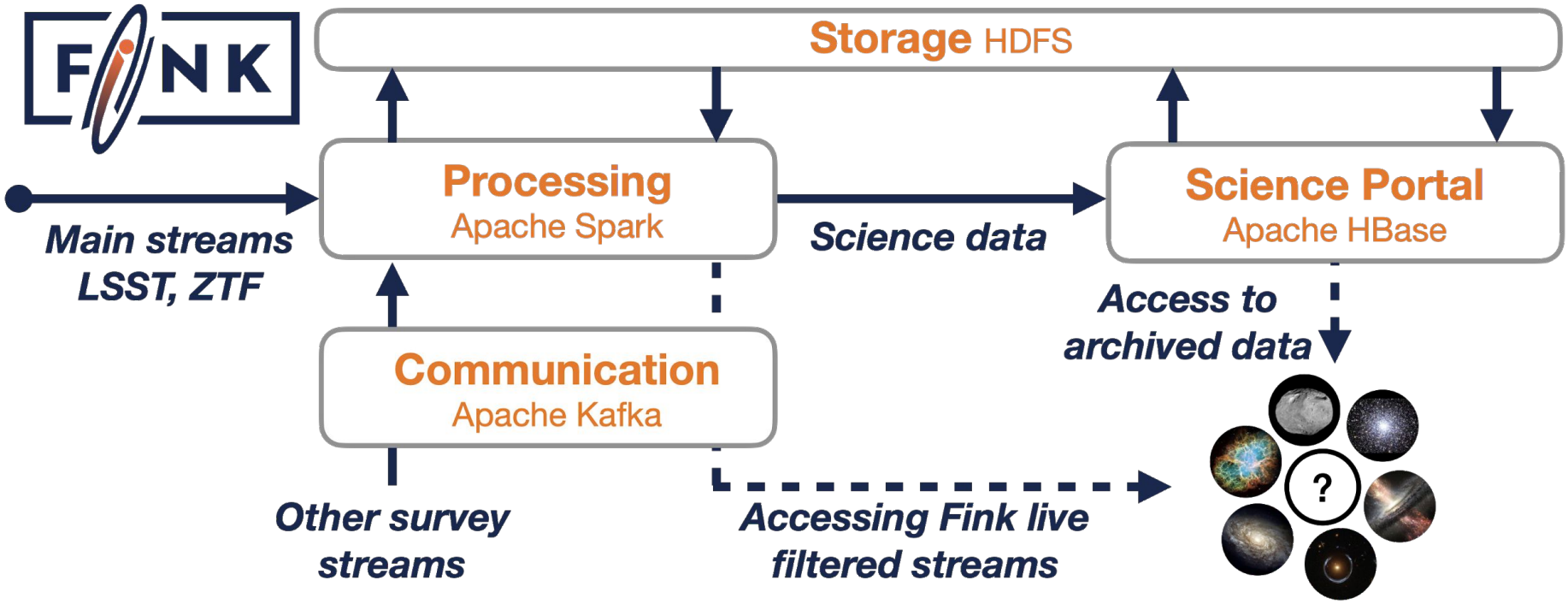
Conclusions

- Active Anomaly Detection can be a powerful tool to boost discoveries
- Approach is still under development in other fields
  - Opportunity to develop astronomy-oriented strategies
- Astronomical data has many caveats which are not necessarily taken into account by off-the-shelf algorithms
- Collaboration is essential

https://snad.space/

Implications

A French-born broker to digest LSST alerts and search for interesting astrophysical objects



Community-driven project with important elements on
Adaptive Learning and Bayesian Deep Learning

Thank you, Merci, Спасибо

From the SИAD team!

SИAD

https://snad.space/

Back-up slides

Active Anomaly Discovery

Algorithm 2 Active Anomaly Discovery (AAD)

Input: Dataset $\mathbf{H}$, budget B
Initialize the weights $\mathbf{w}^{(0)} = \{\frac{1}{\sqrt{m}}, ..., \frac{1}{\sqrt{m}}\}$
Set $t = 0$
Set $\mathbf{H}_A = \mathbf{H}_N = \emptyset$
while $t \leq B$ **do**
  $t = t + 1$
  Set $\mathbf{a} = \mathbf{H} \cdot \mathbf{w}$ (i.e., $\mathbf{a}$ is the vector of anomaly scores)
  Let $\mathbf{z}_i$ = instance with highest anomaly score (where $i = \arg\max_i(a_i)$)
  Get feedback {*'anomaly'*/*'nominal'*} on $\mathbf{z}_i$
  if $\mathbf{z}_i$ is *anomaly* **then**
    $\mathbf{H}_A = \{\mathbf{z}_i\} \cup \mathbf{H}_A$
  else
    $\mathbf{H}_N = \{\mathbf{z}_i\} \cup \mathbf{H}_N$
  end if
  $\mathbf{w}^{(t)}$ = compute new weights; normalize $\|\mathbf{w}^{(t)}\| = 1$
end while

$$l(q, \boldsymbol{w}; z_i, y_i) = \begin{cases} 0 & \text{if} \quad \boldsymbol{w} \cdot \boldsymbol{z}_i \geq q \quad \text{and} \quad y_i = \text{anomaly} \\ 0 & \text{if} \quad \boldsymbol{w} \cdot \boldsymbol{z}_i < q \quad \text{and} \quad y_i = \text{normal} \\ q - \boldsymbol{w} \cdot \boldsymbol{z}_i & \text{if} \quad \boldsymbol{w} \cdot \boldsymbol{z}_i < q \quad \text{and} \quad y_i = \text{anomaly} \\ \boldsymbol{w} \cdot \boldsymbol{z}_i - q & \text{if} \quad \boldsymbol{w} \cdot \boldsymbol{z}_i \geq q \quad \text{and} \quad y_i = \text{normal} \end{cases},$$

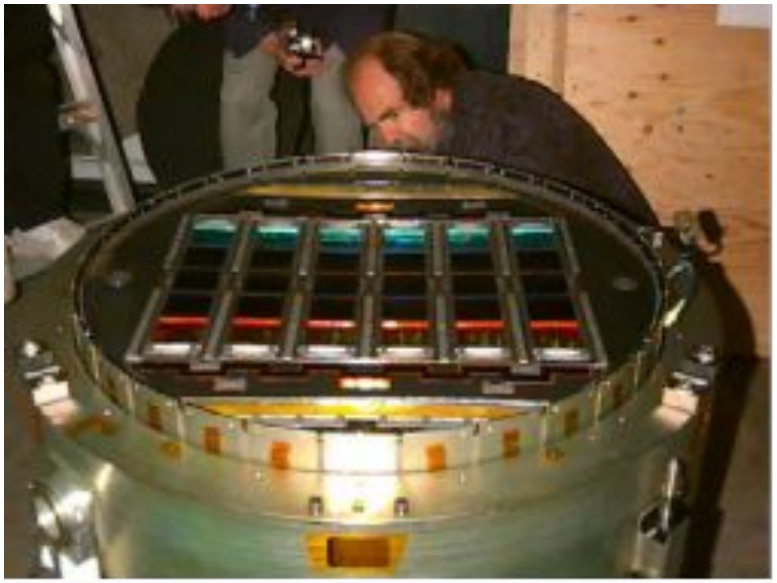
$$\begin{aligned} \mathbf{w}^{(t)} = \arg\min_{\mathbf{w}, \xi} \; & \frac{C_A}{|\mathbf{H}_A|} \left(\sum_{\mathbf{z}_i \in \mathbf{H}_A} \ell(\hat{q}_\tau(\mathbf{w}^{(t-1)}), \mathbf{w}; (\mathbf{z}_i, y_i)) \right) \\ & + \frac{1}{|\mathbf{H}_N|} \left(\sum_{\mathbf{z}_i \in \mathbf{H}_N} \ell(\hat{q}_\tau(\mathbf{w}^{(t-1)}), \mathbf{w}; (\mathbf{z}_i, y_i)) \right) \\ & + \frac{C_\xi}{|\mathbf{H}_A|} \left(\sum_{\mathbf{z}_i \in \mathbf{H}_A} \ell(\mathbf{z}_\tau^{(t-1)} \cdot \mathbf{w}, \mathbf{w}; (\mathbf{z}_i, y_i)) \right) \\ & + \frac{C_\xi}{|\mathbf{H}_N|} \left(\sum_{\mathbf{z}_i \in \mathbf{H}_N} \ell(\mathbf{z}_\tau^{(t-1)} \cdot \mathbf{w}, \mathbf{w}; (\mathbf{z}_i, y_i)) \right) \\ & + \|\mathbf{w} - \mathbf{w}_p\|^2 \end{aligned} \tag{2}$$

where, $\mathbf{w}_p = \frac{\mathbf{w}_U}{\|\mathbf{w}_U\|} = [\frac{1}{\sqrt{m}}, \ldots, \frac{1}{\sqrt{m}}]^T$, $\mathbf{z}_\tau^{(t-1)}$ and $\hat{q}_\tau(\mathbf{w}^{(t-1)})$ are

Das, S., Wong, W-K., Dietterich, T., Fern, A. and Emmott, A. (2016). *Incorporating Expert Feedback into Active Anomaly Discovery*

Photometry x Spectroscopy

An example from SDSS



Exposure time 2 x 54s

Integration time of at least

45 minutes

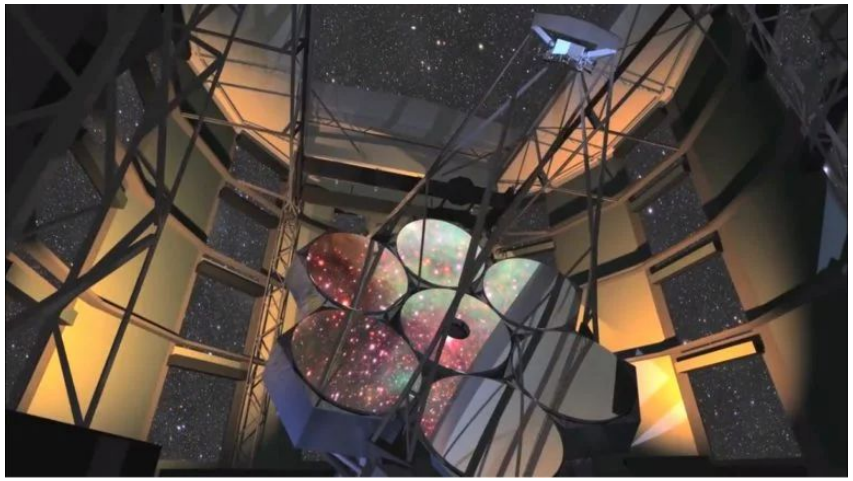
http://www.preposterousuniverse.com/blog/2009/10/06/practicality-and-the-universe/

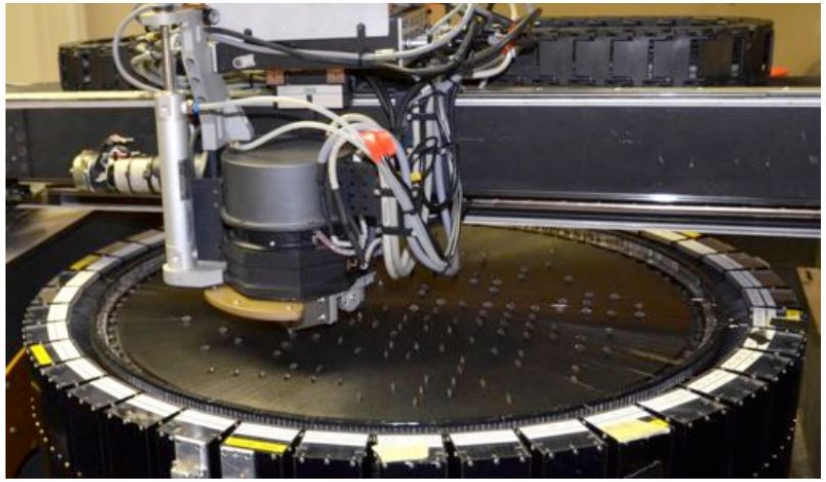
http://www.stsci.edu/~inr/bdpics/bd5.htm

Photometry x Spectroscopy

An example from the Australian Astronomical Observatory

For the Giant Magelean Telescope (GMT)
First light 2025





a)

Integration time much larger...

https://www.gmto.org/

https://arxiv.org/pdf/1311.7371.pdf