Modeling and Predicting the Dynamics of Globular Clusters with Deep Learning

Paul Magron - in collaboration with Arn Marklund and Paolo Bianchini

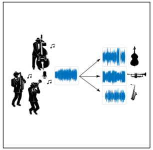
Seminar at Shanghai Astronomical Observatory (SHAO), Shanghai, China - March 4th, 2026

A brief history of me

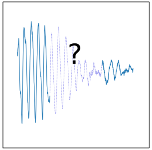
Background

- 2013: Engineering and MSc degree (Paris, France)
- 2013-2016: PhD thesis at Télécom Paris (France)
- 2017-2019: Post-doc at Tampere University (Finland)
- 2019-2021: Post-doc at IRIT (Toulouse, France)
- Since 2021: Researcher at Inria Nancy (France), MULTISPEECH team.

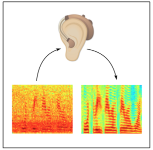
Research topics



Source/signal separation

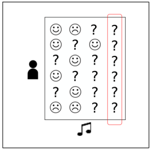


Audio restoration / inpainting

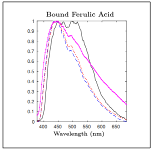


Speech enhancement

Ambient sound analysis



Recommender systems



Fluorescence spectroscopy

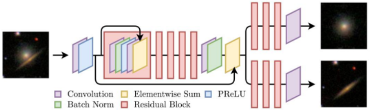
From audio to astronomy imaging

Common methodological tools

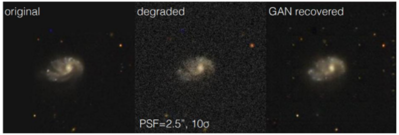
- Matrix and tensor factorization.
- DNNs for ∼everything.
- Numerical methods / optimization.

Applications

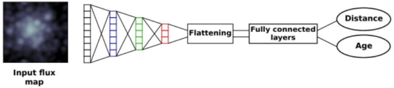
- Source separation.
- Noise reduction.
- Classification / regression.



Reiman and Göhre, "Deblending galaxy superpositions with branched generative adversarial networks", MNRAS, 2019.



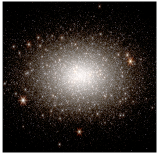
Schawinski et al., "Generative Adversarial Networks recover features in astrophysical images of galaxies beyond the deconvolution limit", MNRAS:Letters, 2017.



Chardin and Bianchini, "Predicting images for the dynamics of stellar clusters (π-DOC): a deep learning framework to predict mass, distance, and age of globular clusters", MNRAS, 2021.

Project

Modeling and Predicting the Dynamics of Globular Clusters with Deep Learning



Outline

Context and Problem setting

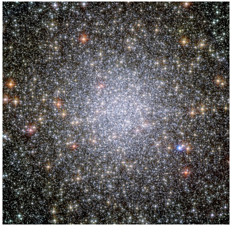
Original π-doc

Improved pipeline

Results

Context and Problem setting

Globular clusters



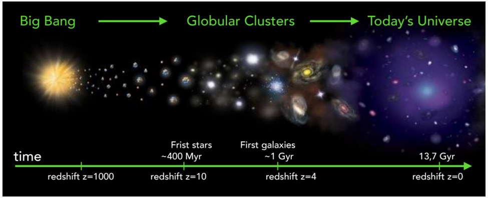
Credit: NASA, ESA, and the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration

Fundamental and extreme stellar systems.

- ▷ Very common: our galaxy, the Milky Way, hosts 160 of them.
- ▷ Very dense: about 10^6 stars per cluster.
- ▷ Very old: 13 Gyr, comparable to the age of the Universe.

Fossils of the ancient universe

- ▷ High-redshift era.
- ▷ Assembly of galaxies.
- ▷ Earliest star formation.



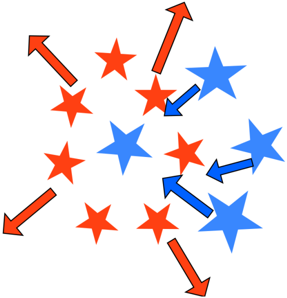
Challenges

Open question of their origin.

- ▷ What are GCs properties today, e.g., mass, distance, age, number of black holes?
- ▷ How did they form in the primordial universe; what initial properties?
- ▷ What are the details of their evolution?

GC dynamics governed by many complex phenomena.

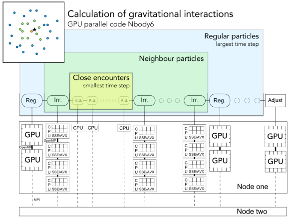
- ▷ Internal dynamics, e.g., star-by-star gravitational interaction.
- ▷ Influence of the host galaxy, e.g., formation of tails.
- ▷ Imprint of their initial formation process, including dark matter.
- ▷ over 13+ Gyr evolution.

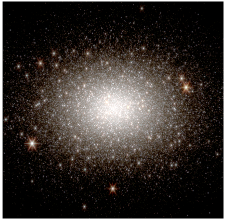


N-body simulations

Direct approach for stellar dynamics.

- ▷ Highly accurate, physically principled.
- ▷ Star-by-star interactions modelled directly.
- ▷ Naturally account for dynamical / stellar evolution, e.g., mass segregation, mass loss, presence of black holes.
- ▷ Optimized code/hardware, e.g., GPU parallel code Nbody6++GPU.





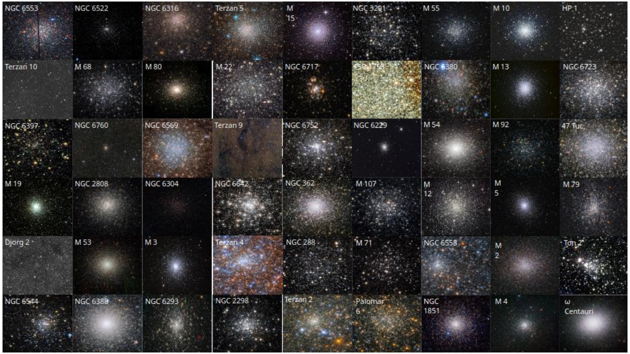
☹ Costly simulations

- ▷ 10^6 stars over 13 Gyr: $400+$ days of wall-clock time.
- ▷ Equivalent footprint of ≈ 9 CO2 tons.

Wang et al., "nbody6++gpu: ready for the gravitational million-body problem", *Monthly Notices of the Royal Astronomical Society*, Vol. 450, Issue 4, pp. 4070-4080, May 2015.

Observations

- Modern astronomy programs generate a huge amount of data.
  - Vera C. Rubin observatory ≈ 20 Tb/day.
- Need for efficient and sustainable methods to automatically process it.
  - N-body simulations are not appropriate for modeling individual GCs.

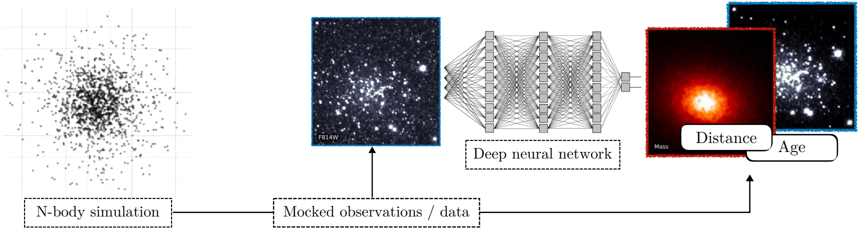


Compilation by Giulia Pagnini (ObAS)

Proposed strategy

Bring together simulation-based and deep learning-based modeling.

- ▷ Leverage a reduced set of N-body simulations.
- ▷ Create realistic *mocked* observations of GCs.
- ▷ Design and train a neural network that map GC images to their properties.

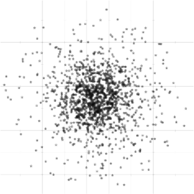


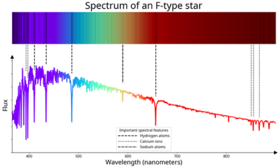
Original π-doc

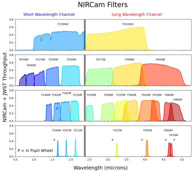
From simulations to "observations"

From a simulation snapshot, compute:

- ▷ Stellar spectra: Get stellar parameters for each star, and generate spectra via, e.g., FSPS.
- ▷ Stellar magnitudes: Convolve spectra with the telescope filters to get the luminosity of a star in a given color range.
- ▷ Mocked observations: Convolve magnitude with the point spread function (PSF) to account for the telescope structure.







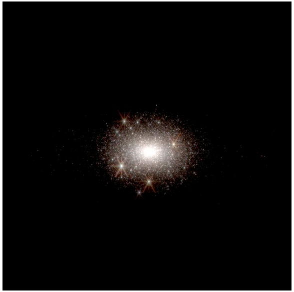
Conroy and Gun, "The Propagation of Uncertainties in Stellar Population Synthesis Modeling. III. Model Calibration, Comparison, and Evaluation", *The Astrophysical Journal*, Vol. 712, Issue 2, pp. 833-857, April 2010.

Snapshots at a fixed age

Mocked observations of N-body simulations using a JWST setup:

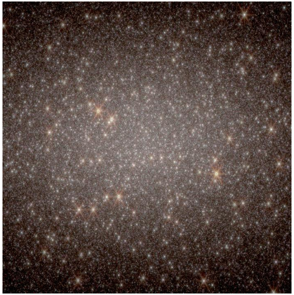
- Color images made with filters F070W, F115W, F356W.
- FoV 2 x 2 arcmin, pixel scale 0.06 arcsec

Distance = 800 kpc



Distance = 200 kpc

Distance = 20 kpc

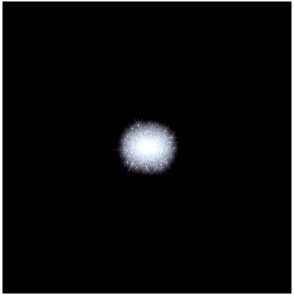


Snapshots at a fixed distance

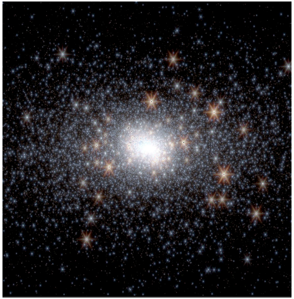
Mocked observations of N-body simulations using a JWST setup:

- ▷ Color images made with filters F070W, F115W, F356W.
- ▷ FoV 2 x 2 arcmin, pixel scale 0.06 arcsec.

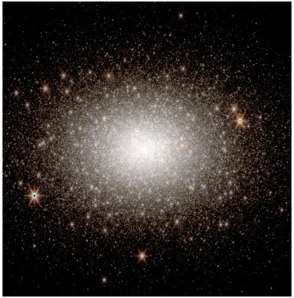
Age = 0 Gyr



Age = 0.1 Gyr



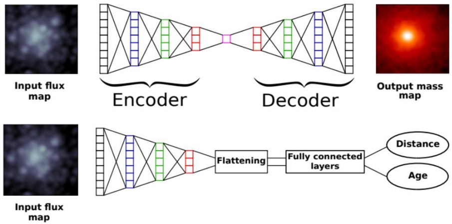
Age = 12 Gyr



Original neural network model

π-doc

- A convolutional encoder-decoder (CED) that predicts the mass distribution.
- A convolutional neural network (CNN) that predicts the distance and age.



Training

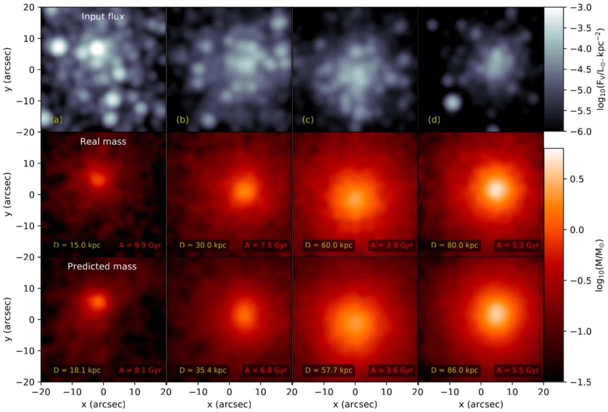
- Extract 10k snapshots from 2 N-body simulations each = 20k training samples.
- Minimize the mean square error for each network.

Test data

- Mocked data: 8k samples extracted from another N-body simulation.
- Real data: 17 GCs from the Milky Way (MW).

Chardin and Bianchini, "Predicting images for the dynamics of stellar clusters (π-DOC): a deep learning framework to predict mass, distance, and age of globular clusters", *Monthly Notices of the Royal Astronomical Society*, Vol. 504, Issue 4, pp. 5656-5670, March 2021.

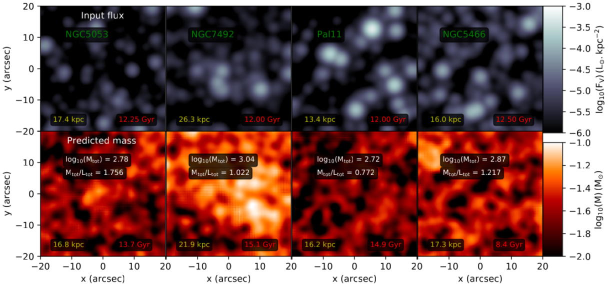
Results on mocked data

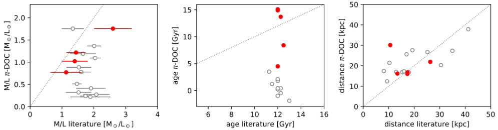


Average errors (test set):

- ▷ Pixel-by-pixel mass prediction: 27%
- ▷ Total mass in a FoV: 15%
- ▷ Total M/L: 11%
- ▷ Age: 1.5 Gyr
- ▷ Distance: 6 kpc

Results on real data





Results are consistent with the literature for small GCs.

Limitations

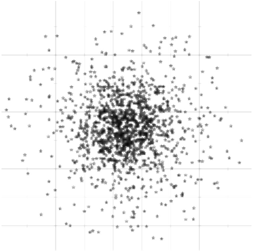
- ▷ Most real GCs are larger than in the simulation setup (about 10^4 stars).
- ▷ MW clusters are not representative of all GCs.
- ▷ π-doc processes clean ($=$ noise-free) images, and uses one color channel.

Improved pipeline

New simulations

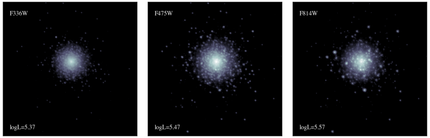
ROLLIN [Bianchini, 2026]

- ▷ Direct N-body simulations.
- ▷ Rotating models of GCs with 250k-1.5M stars.
- ▷ Run for up to 14 Gyr.
- ▷ Stellar evolution, tidal field.



MOCCA [Askar, in prep.]

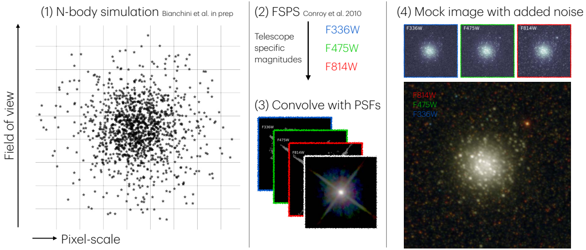
- ▷ Monte Carlo simulations: sample star interactions assuming certain symmetries.
- ▷ Models for 47 Tuc, Ω cen, and dense GCs.
- ▷ High mass (1-12M stars), high density.



Bianchini, "ROLLIN': Rotating Globular Clusters simulations. I. The kinematic evolution of realistic direct N-body models", *to be published in Astronomy & Astrophysic*, 2026.

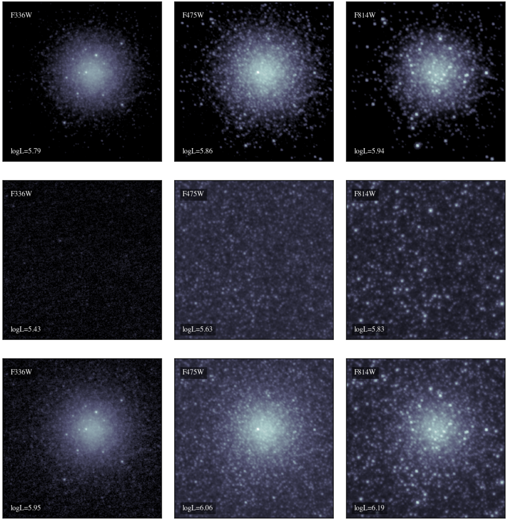
More realistic snapshots

- Account for 3 filters / color channels (F336W, F475W, F814W)
- PSFs are measured from the HST.
- Incorporate two types of noise:
  - Poisson noise: simulates instrument / measurement noise.
  - Galactic noise: simulates the impact of the surrounding galaxy.



Noise = extinction-corrected images from Andromeda (HST PHAT survey).

New mocked observations



Clean image

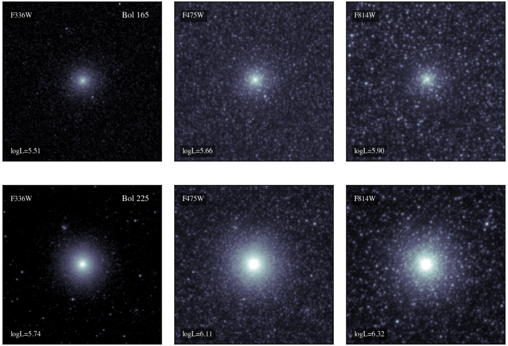
Noise

Clean + noise

Real observations

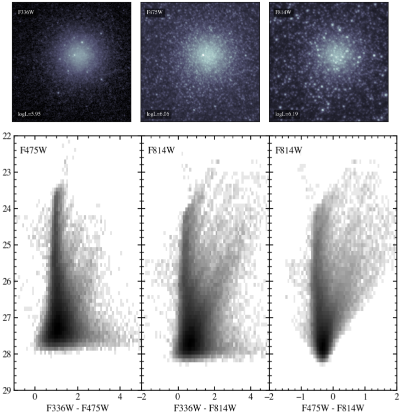
Bol165

Bol225

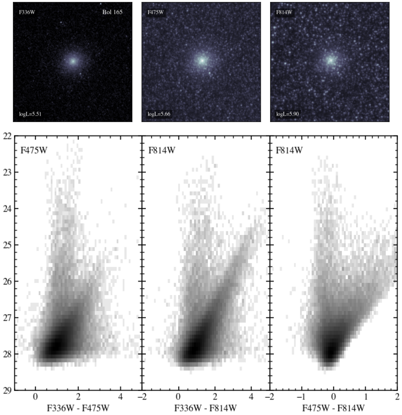


Mocked vs. Real

Mocked



Bol165

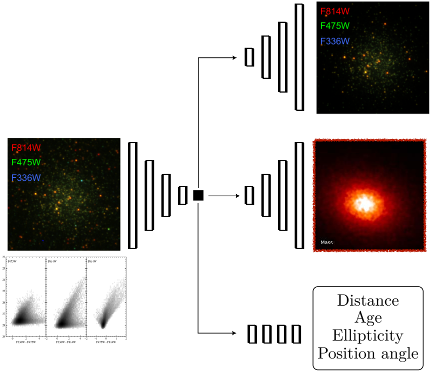


Improving the model

π-doc has been merged into one single network.

Multiple inputs from passbands F336W, F475W, and F814W.

- ▷ 3 magnitude / log-flux.
- ▷ 3 *colors* / diff. between magnitudes.
- ▷ 3 color magnitude diagrams (CMDs).



Multiple outputs

- ▷ Denoising / decontamination
- ▷ Mass distribution
- ▷ Parameter estimation: distance, age, ellipcity, position/axis angle, (metallicity).

CMDs are beneficial for predicting distance / age, but not helpful for predicting the mass / clean GC as they do not contain spatial information.

Dataset

Mocked observations

- 27 simulations in total, from which we extract snapshots

  ≈ 175k image for training (75%) + validation (25%)

  ≈ 50k images for testing
- Different simulations are used from training/validation and testing.
- $\approx 10\%$ images from MOCCA, the rest from ROLLIN'.

Background noise

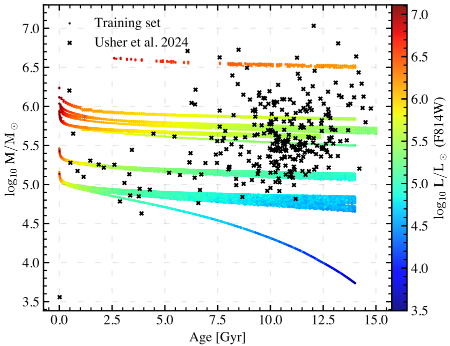
- ≈ 12k images from M31.
- On-the-fly rotation and cropping when added to the mocked observations.

Real test data

- 492 GC images from the Andromeda galaxy.
  - 273 from the PHAT survey.
  - 219 from the PHAST survey.

Dataset

- Simulations cover most of the space of M31 clusters according to [Usher, 2024].
- Points above $\log_{10} M/M_{\odot} = 5.7$ after 1 Gyr corresponds to MOCCA images.
    - Bias towards 0 ellipticity.
    - Only 10k images



Usher et al., "Measuring M31 globular cluster ages and metallicities using both photometry and spectroscopy", *Monthly Notices of the Royal Astronomical Society*, Vol. 528, Issue 4, pp. 6010-6024, March 2024.

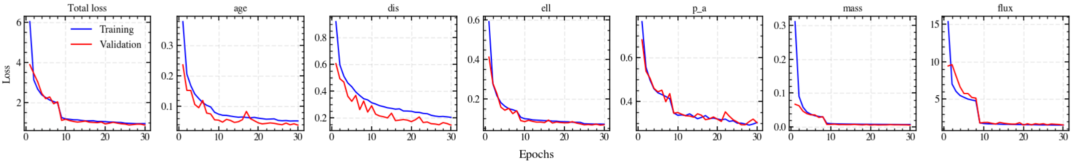
Training

Loss function = sum of task-specific terms:

$$\lambda_{\text{age}}||\text{age} - \widehat{\text{age}}||^2 + \lambda_{\text{distance}}||\text{distance} - \widehat{\text{distance}}||^2 + \lambda_{\text{mass}}||\text{mass} - \widehat{\text{mass}}||^2 + \dots$$

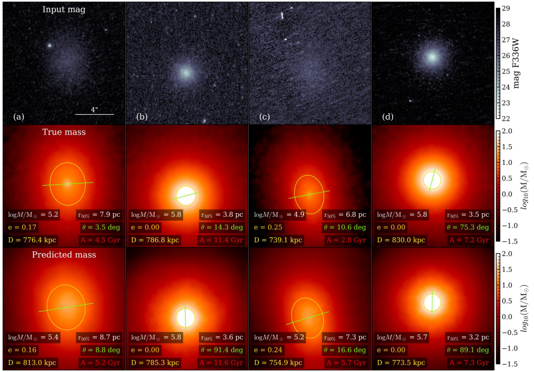
- All weights are $= 1$ except for the denoising branch $\lambda_{\text{flux}} = 0.2$.

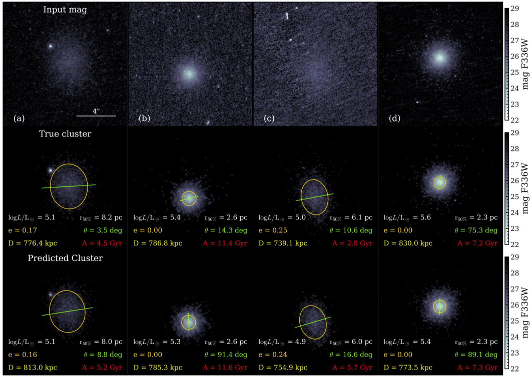
Training log ($\approx$ 250 GPU hours)



Results

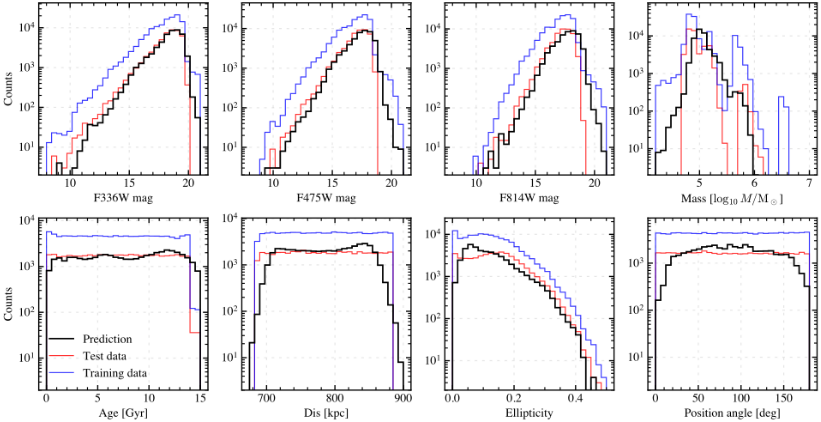
Results on mocked data - denoising



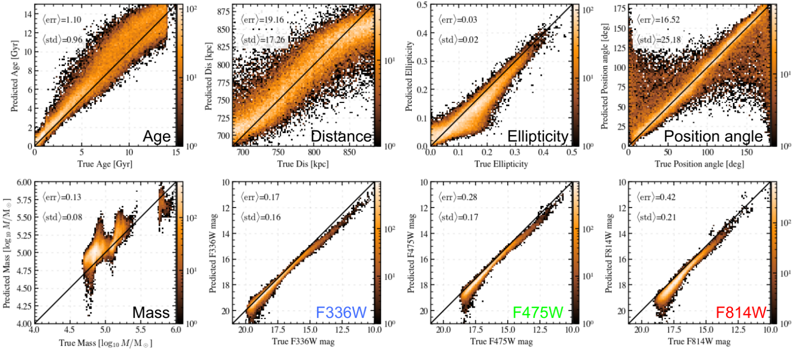


- ▷ Overall recovery of the mass distribution, but less accuracy at the outskirt.
- ▷ Denoised GCs are too faint, outskirt details are lost.

Results on mocked data - predicted vs. true

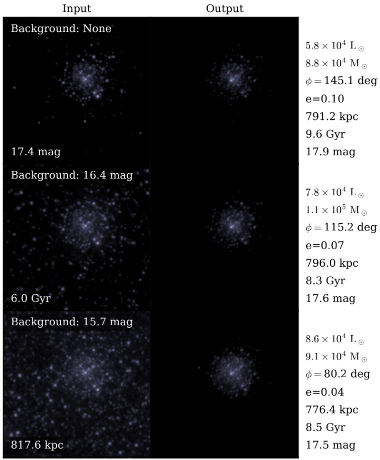


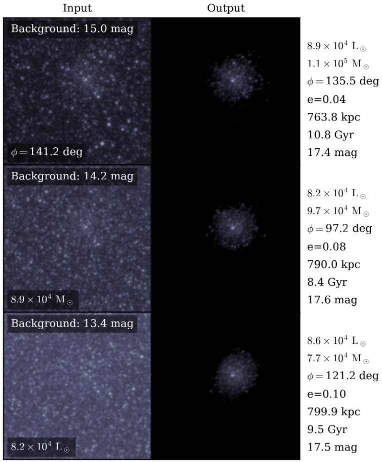
Results on mocked data - predicted vs. true



- Distance / age are overall well estimated, nothing too far out of range.
- Ellipcity values close to 0 are difficult to estimate accurately.
- Luminosity are underestimated.
- The network struggles with faint / low signal GCs.

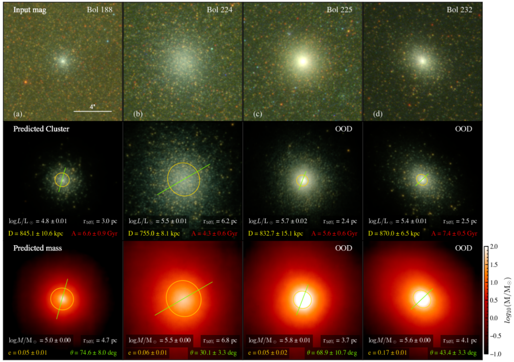
Results on mocked data - robustness

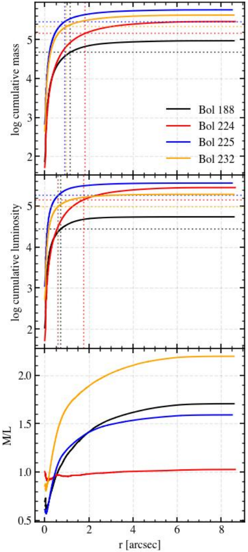




- ▷ Quite stable behavior when modifying the background noise, except for extreme cases.
- ▷ But the mass is overestimated when the background is really strong.

Results on real data

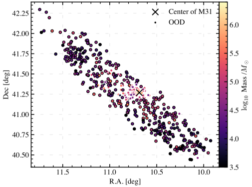


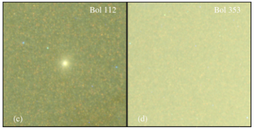


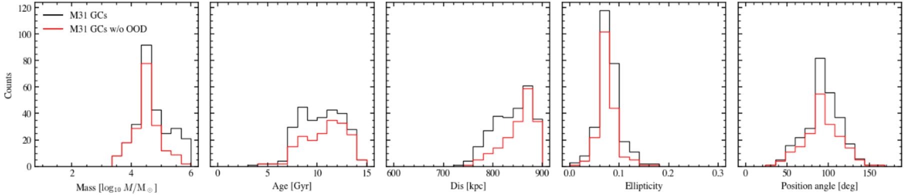
- ▷ Useful for computing the mass/mass-luminosity radius profiles:
  - ▷ How mass-segregated the clusters are.
  - ▷ Probe the distribution of non-visible matter.

Results on real data - galactic distribution

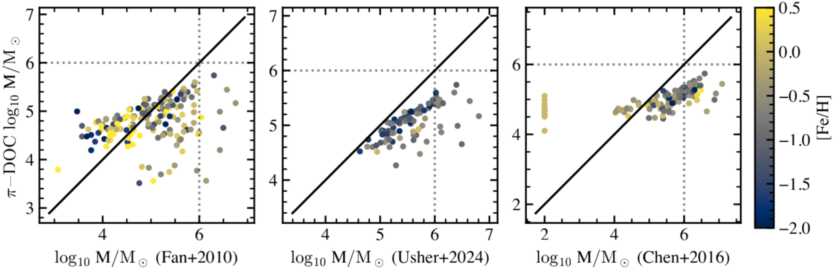
- ▷ Bulge/Center of M31 = massive/dense clusters.
- ▷ These GCs also contain more noise: more challenging for π-doc.
- ▷ Out-of-distribution clusters: about 30%.





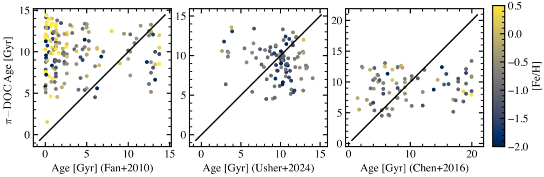


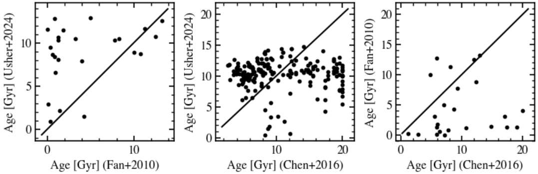
Comparison with literature - mass



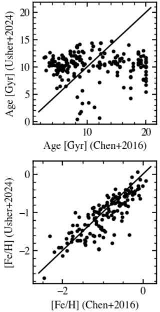
- ▷ Fairly similar mass estimates, consistent with the literature.
- ▷ Not so much for the very massive GCs, but the difference is not that extreme.

Comparison with literature - age



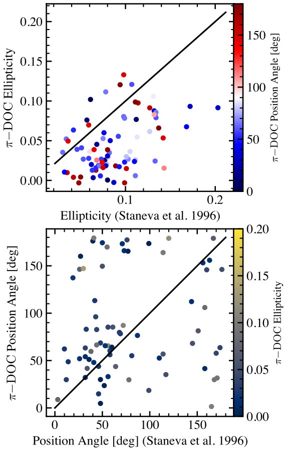


- Large discrepencies for age estimates.
- Other studies also dont agree with each other.
- Age and metallicity are degenerate, e.g., both affect the colour of the stars/clusters.
- Our simulations have metallicities ≤ -1.3: we lack very metallic GCs.
- Beyond metallicity, age is inherently difficult.



Comparison with literature - morphology

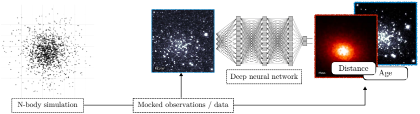
- π-doc tends to underestimate ellipticity, same as for mocked observations.
- Large discrepencies for axis angle.



Summary

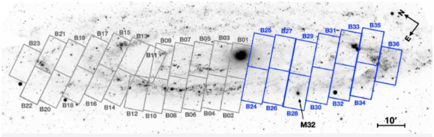
Combining simulation-based and deep learning-based modeling of GCs

- ▷ A novel approach to learn the dynamics of GCs.
- ▷ Promising results on simulated data.
- ▷ Consistent results with the litterature on low-mass / low-density clusters.
- ▷ Paper in preparation (A%A).



Future work

- ▷ Fine-tune architecture / training process.
- ▷ Build a M31 catalogue using π-doc.
- ▷ Adaption of the algorithm to more general cases, e.g., other telescopes, galaxies.

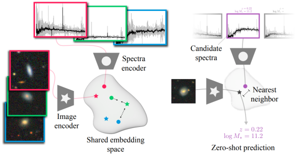


https://archive.stsci.edu/hlsp/phast

Perspectives

Foundation models of GCs: process data from any telescope / cluster.

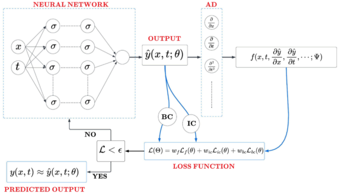
- Arbitrary number and wavelenfth / number of filters / dimension / PSF.
- Exploit convolutional transformers (conformers) .



Parker et al., "AstroCLIP: A Cross-Modal Foundation Model for Galaxies", MNRAS, 2024.

From N-body to neural-based simulations: speed up N-body simulations using deep learning.

- Leverage physics-informed neural networks (PINNs) to generate GCs simulations.
- Or diffusion models to directly generate mocked observations / images.



https://arxiv.org/pdf/2505.22761

Thanks!

paul.magron@inria.fr

https://magronp.github.io/
https://github.com/magronp/

arn.marklund@astro.unistra.fr

paolo.bianchini@astro.unistra.fr