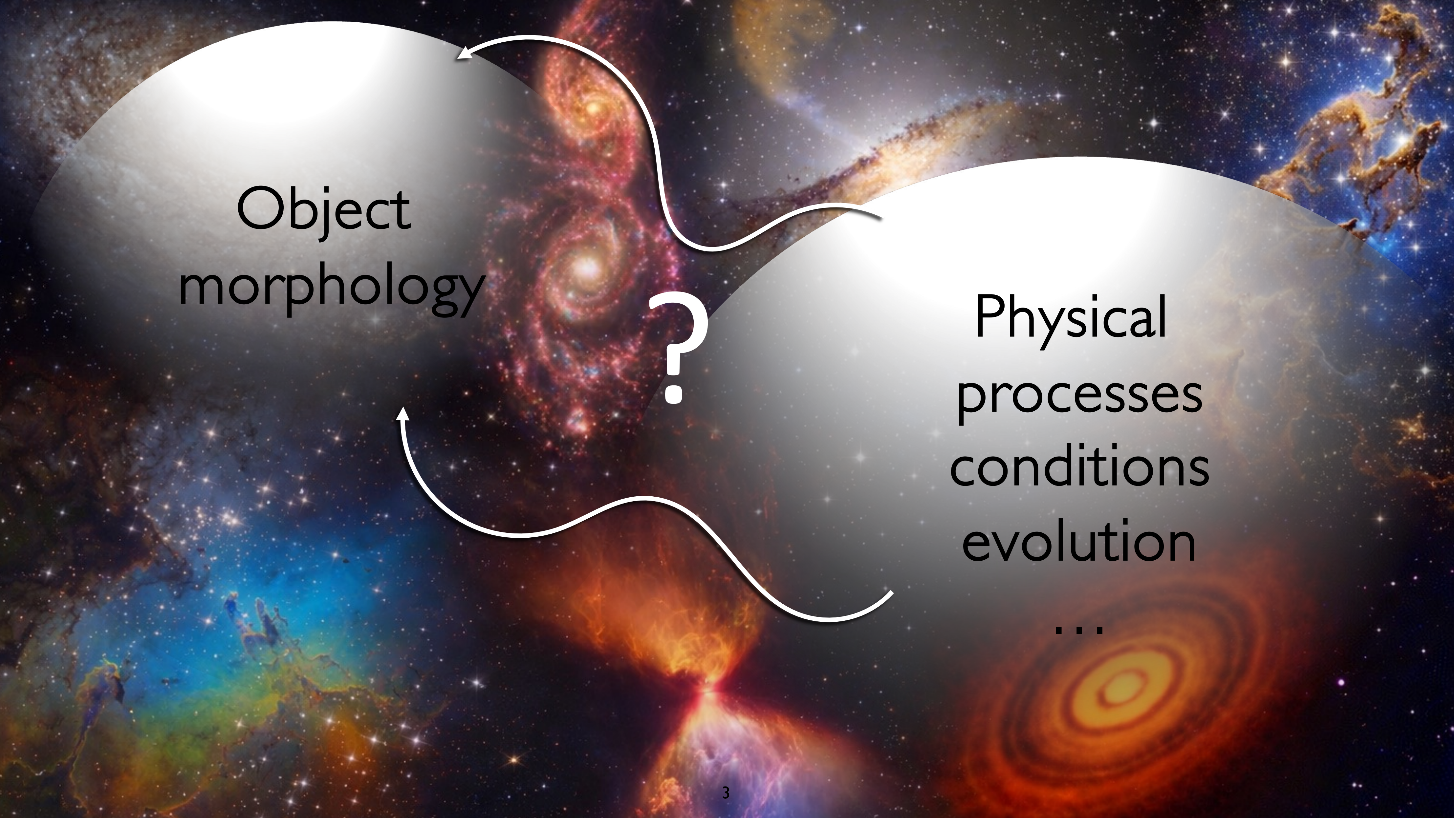
A machine learning approach to identifying key properties

Per Bjerkeli & Maria Carmen Toribio
Chalmers University of Technology
With: Jouni Kainulainen, Leon Boschman, Otoniel Maya Lucas

Stockholm 5 february 2026

"Collage of important morphologies in astronomy"


Object
morphology
?
Physical
processes
conditions
evolution
…

Conclusions:

Take away and exploit:

ASTROMORPH — (astronomer-)friendly self-supervised machine learning package for exploring data and discovering patterns.

Come to talk and brainstorm about:

Decoding morphological information in terms of physically important processes/conditions/phenomena — **user cases**.

Motivator?
What do we try to improve on?

Astronomy — common with large datasets of objects with complex morphologies.
Manual inspection does not scale!

ML and especially deep learning excellent for **exploring large data — finding patterns/classes/relations.**

Promising approaches of **self-supervised learning**, especially Bootstrap Your Own Latent (**BYOL**; Grill et al. 2020) — adapt and optimise for astro-specific use.

ASTROMORPH — (astronomer-)friendly self-supervised machine learning package for exploring data and discovering patterns.

Motivator?
What do we try to improve on?

Astronomy — common with large datasets of objects with complex morphologies.
Manual inspection does not scale!

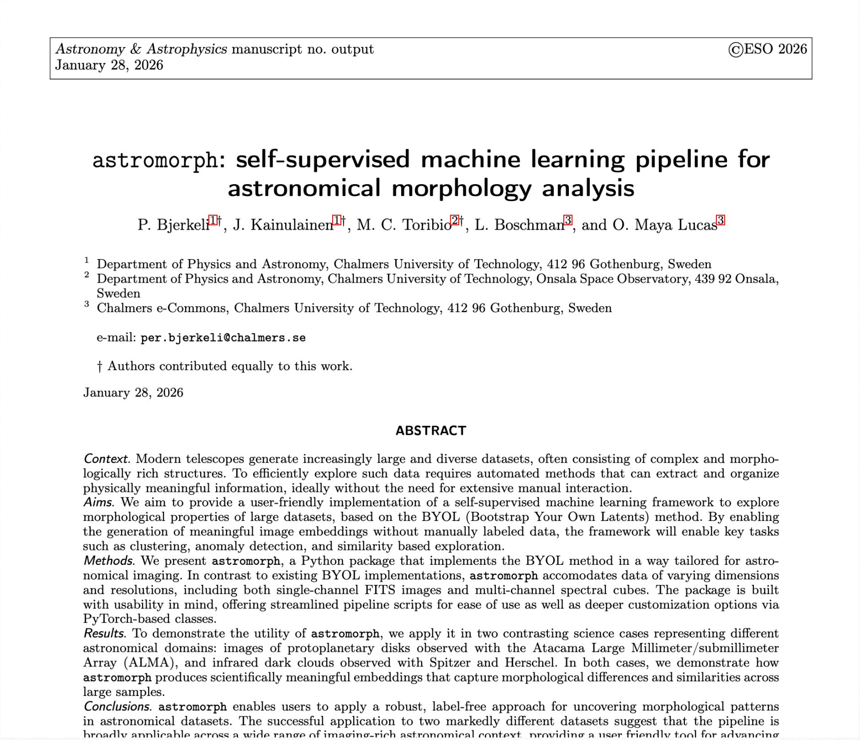
ML and especially deep learning excellent for **exploring large data — finding patterns/classes/relations.**

Promising approaches of **self-supervised learning**, especially Bootstrap Your Own Latent (**BYOL**; Grill et al. 2020) — adapt and optimise for astro-specific use.

Human aspect — lower the astronomy user threshold
"The best camera is the one you have with you." photographer Chase Jarvis

ASTROMORPH — (astronomer-)friendly self-supervised machine learning package for exploring data and discovering patterns.

Astronomy & Astrophysics:



Astronomy & Astrophysics manuscript no. output
January 28, 2026
©ESO 2026

astromorph: self-supervised machine learning pipeline for astronomical morphology analysis

P. Bjerkeli[1†], J. Kainulainen[1†], M. C. Toribio[2†], L. Boschman[3], and O. Maya Lucas[3]

[1] Department of Physics and Astronomy, Chalmers University of Technology, 412 96 Gothenburg, Sweden
[2] Department of Physics and Astronomy, Chalmers University of Technology, Onsala Space Observatory, 439 92 Onsala, Sweden
[3] Chalmers e-Commons, Chalmers University of Technology, 412 96 Gothenburg, Sweden

e-mail: per.bjerkeli@chalmers.se

† Authors contributed equally to this work.

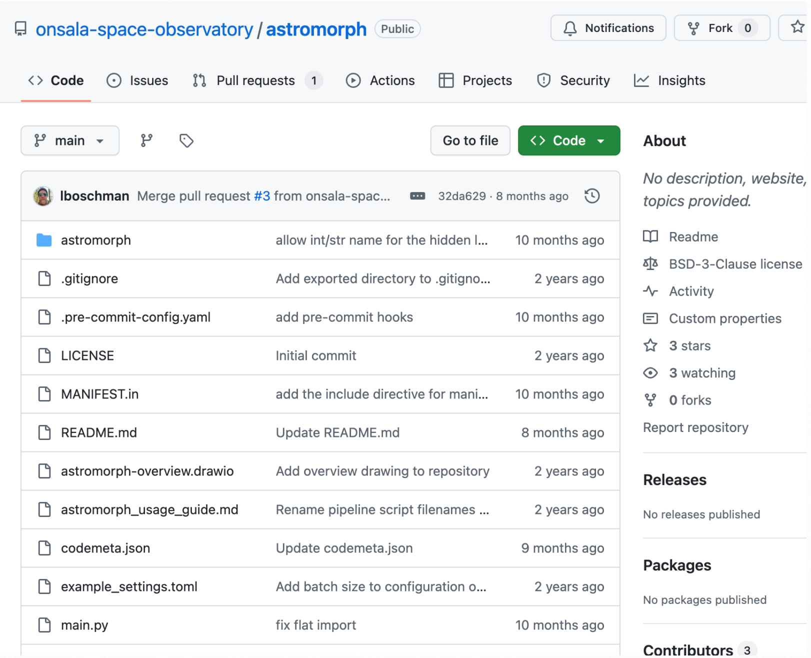
January 28, 2026

ABSTRACT

Context. Modern telescopes generate increasingly large and diverse datasets, often consisting of complex and morphologically rich structures. To efficiently explore such data requires automated methods that can extract and organize physically meaningful information, ideally without the need for extensive manual interaction.
Aims. We aim to provide a user-friendly implementation of a self-supervised machine learning framework to explore morphological properties of large datasets, based on the BYOL (Bootstrap Your Own Latents) method. By enabling the generation of meaningful image embeddings without manually labeled data, the framework will enable key tasks such as clustering, anomaly detection, and similarity based exploration.
Methods. We present astromorph, a Python package that implements the BYOL method in a way tailored for astronomical imaging. In contrast to existing BYOL implementations, astromorph accomodates data of varying dimensions and resolutions, including both single-channel FITS images and multi-channel spectral cubes. The package is built with usability in mind, offering streamlined pipeline scripts for ease of use as well as deeper customization options via PyTorch-based classes.
Results. To demonstrate the utility of astromorph, we apply it in two contrasting science cases representing different astronomical domains: images of protoplanetary disks observed with the Atacama Large Millimeter/submillimeter Array (ALMA), and infrared dark clouds observed with Spitzer and Herschel. In both cases, we demonstrate how astromorph produces scientifically meaningful embeddings that capture morphological differences and similarities across large samples.
Conclusions. astromorph enables users to apply a robust, label-free approach for uncovering morphological patterns in astronomical datasets. The successful application to two markedly different datasets suggest that the pipeline is broadly applicable across a wide range of imaging-rich astronomical context, providing a user friendly tool for advancing

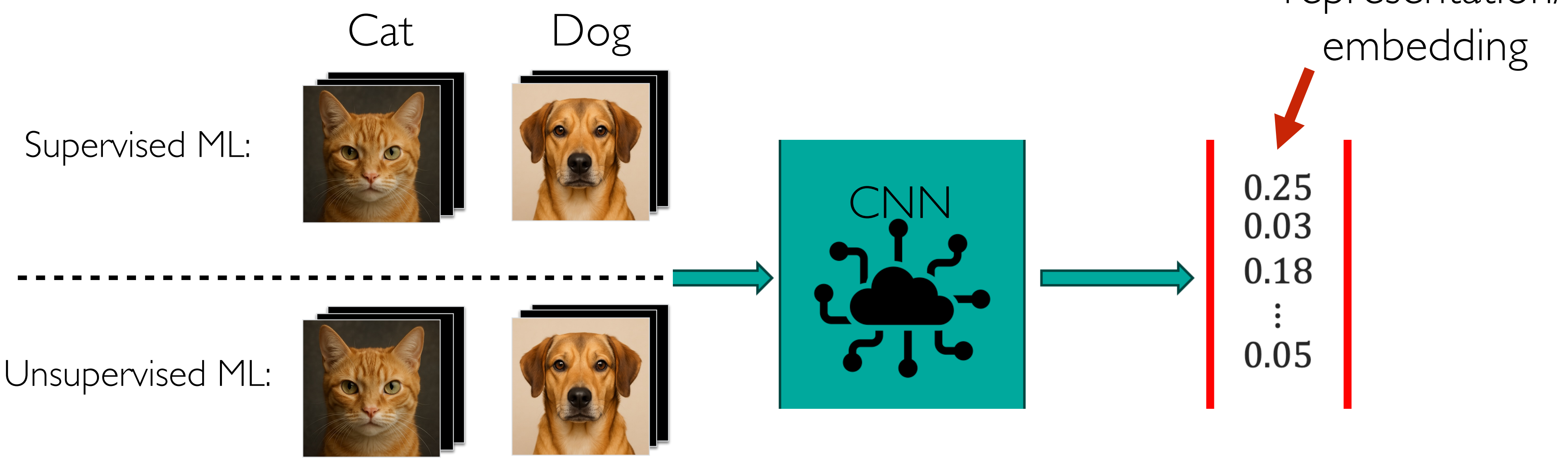
GitHub:

https://github.com/onsala-space-observatory/astromorph



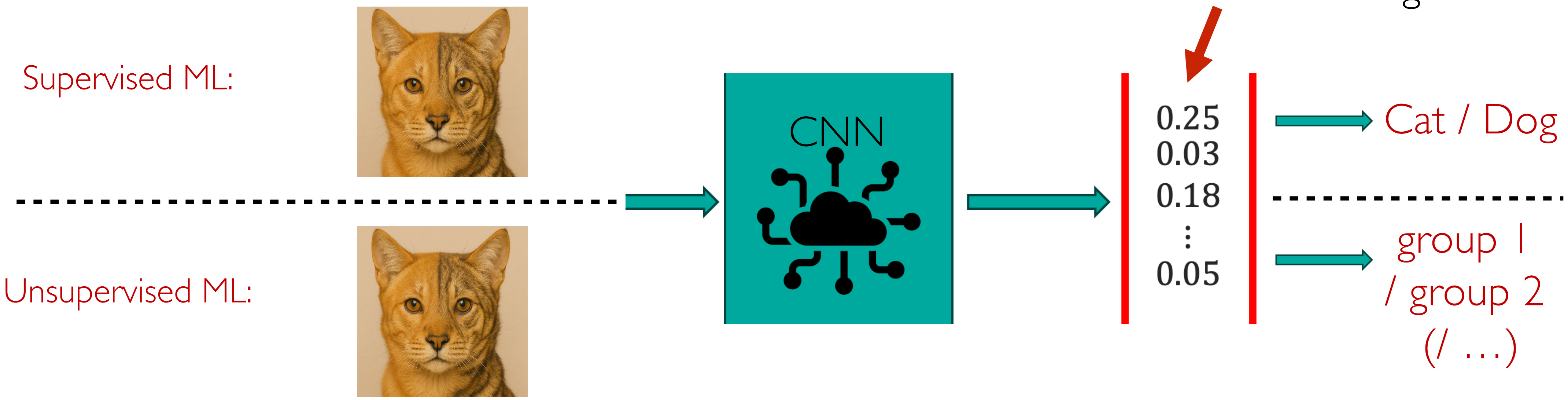
ASTROMORPH

Deep learning methods produce abstractions=representations of data via convolutional neural networks (CNNs).

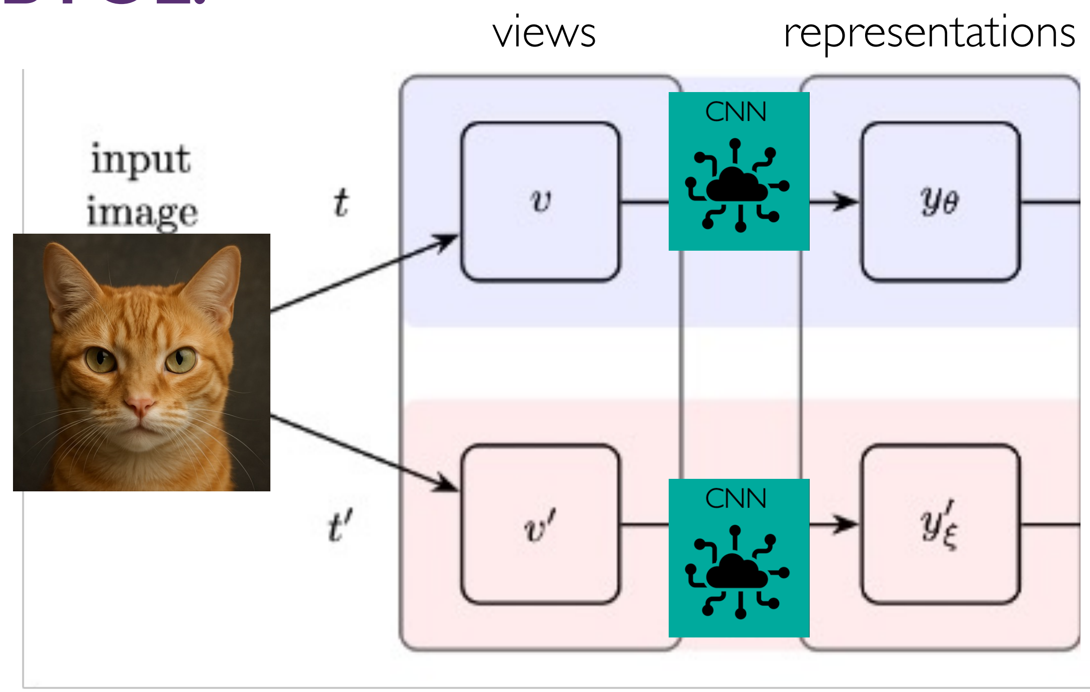


ASTROMORPH

Deep learning methods produce abstractions=representations of data via convolutional neural networks (CNNs).



BYOL:

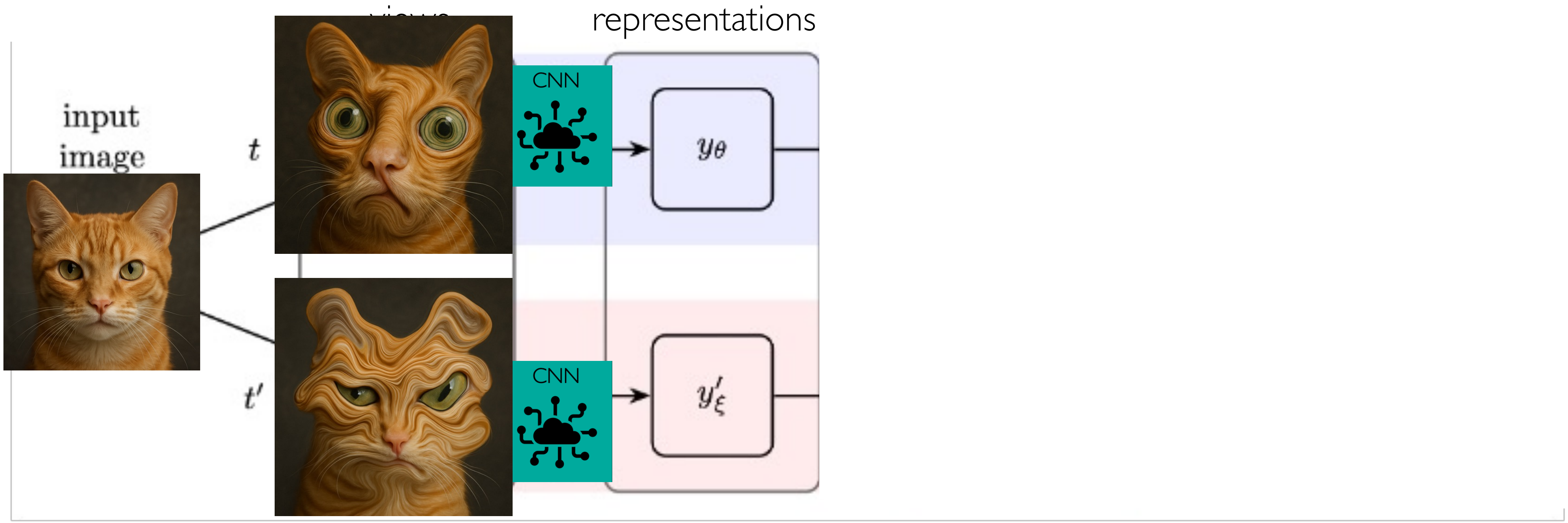


Pair of "views" created by a pretext task (here: stochastic augmentation)

Representations by a CNN

Adapted from Grill et al. (2020)

BYOL:

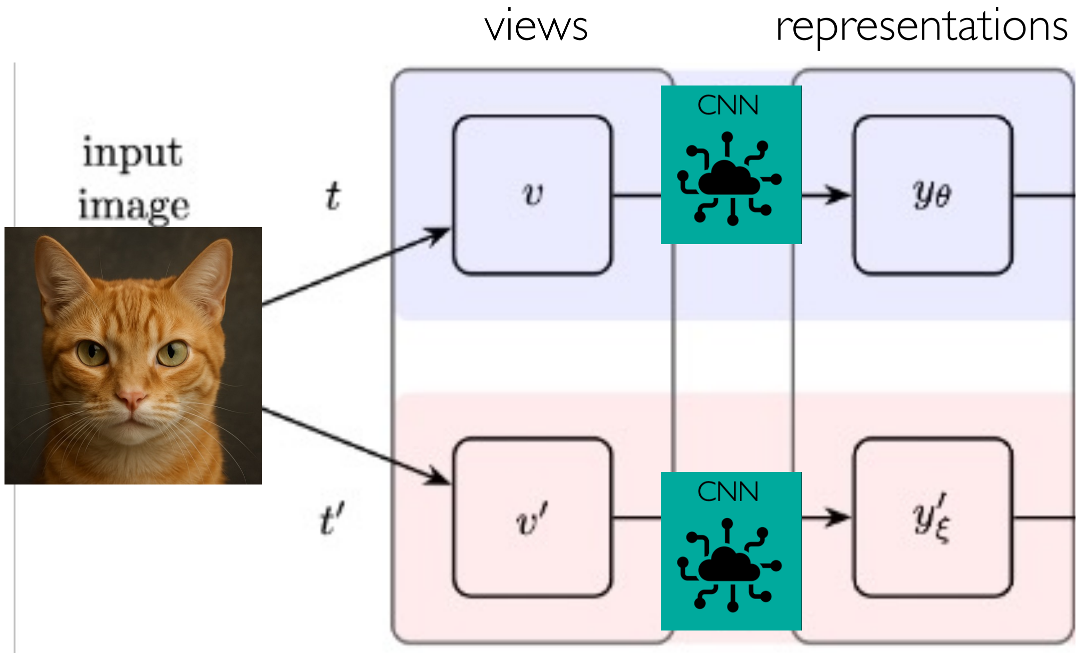


Pair of "views" created by a pretext task (here: stochastic augmentation)

Representations by a CNN

Adapted from Grill et al. (2020)

BYOL:

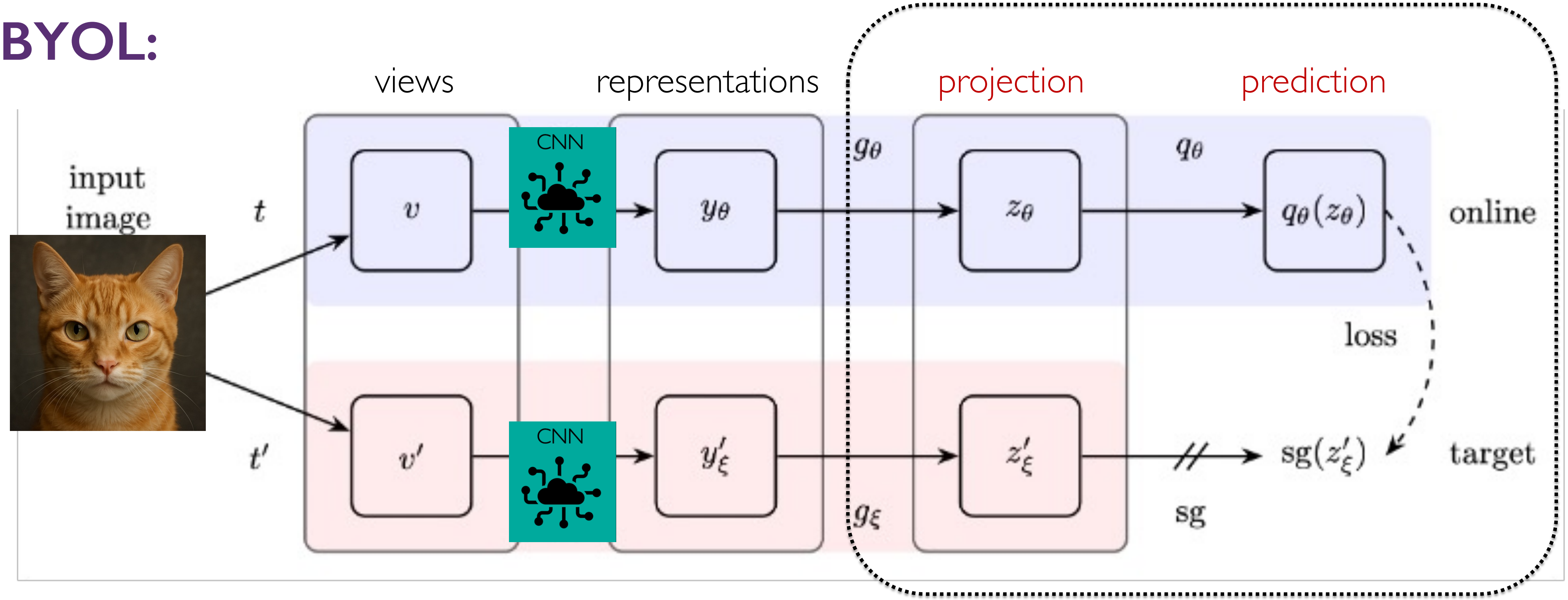


Pair of "views" created by a pretext task (here: stochastic augmentation)

Representations by a CNN

Adapted from Grill et al. (2020)

BYOL:



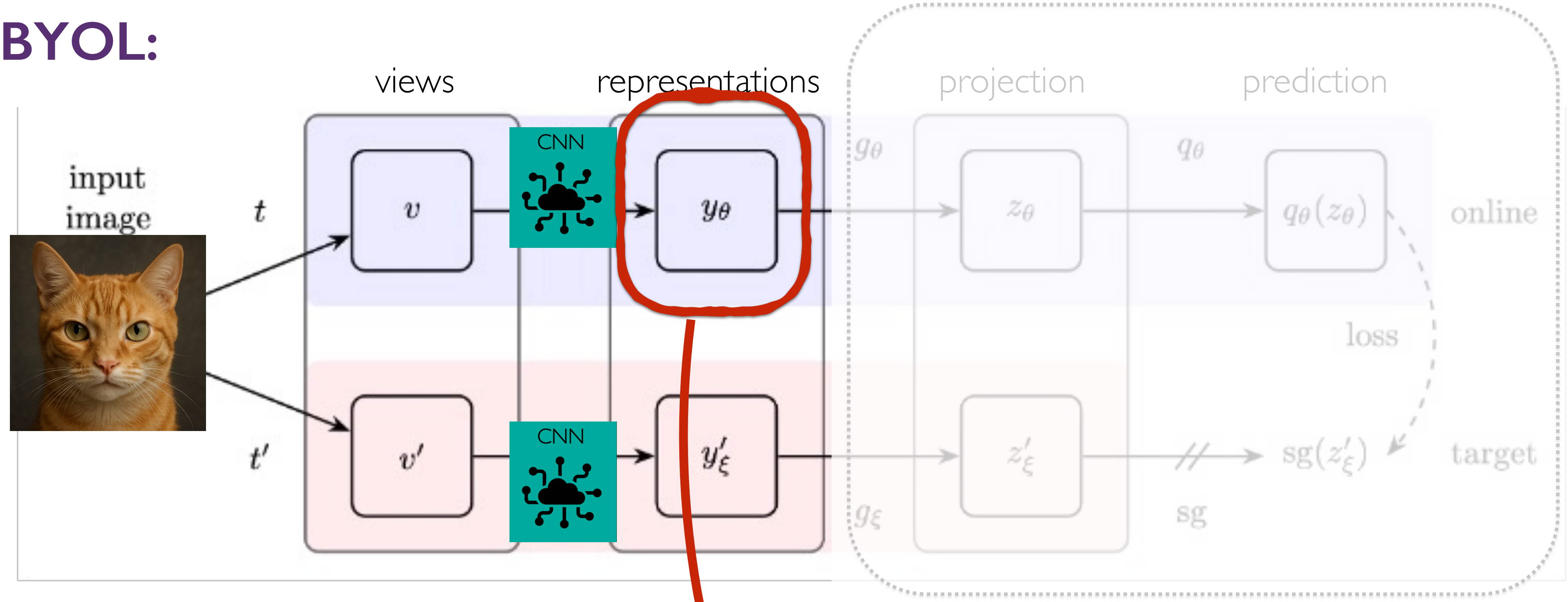
Pair of "views" created by a pretext task (here: stochastic augmentation)

Representations by a CNN

Projection into a (new) latent space and a prediction that is compared to the target

Adapted from Grill et al. (2020)

BYOL:



Pair of "views" created by a pretext task (here: stochastic augmentation)

Final representations (=Result!)

Projection into a (new) latent space and a prediction that is compared to the target

Adapted from Grill et al. (2020)

ASTROMORPH — (astronomer-)friendly self-supervised machine learning package for discovering patterns from unknown datasets.

Some important advantages over (un-)supervised ML in this context:

- **No labelling**; **no need for extensive training data** (supervised learning)
- BYOL architecture better **in learning invariant features, focusing on stable morphological patterns**, (esp. via pretext task).
- **Better transferability**. In principle, the representation encoder weights from large data sets can be transferred to smaller data sets and then fine-tuned with it. (practice?).

See Slijepcevic et al. (2023), Mohale & Lochner (2024) for the first exploration and applications of BYOL in galaxy classification.

ASTROMORPH — (astronomer-)friendly self-supervised machine learning package for discovering patterns from unknown datasets.

`astromorph` = set of wrappers+aids (`pytorch/byol-pytorch/torchvision`) to adapt and ease usage.

Different ways to run, depending on experience level:

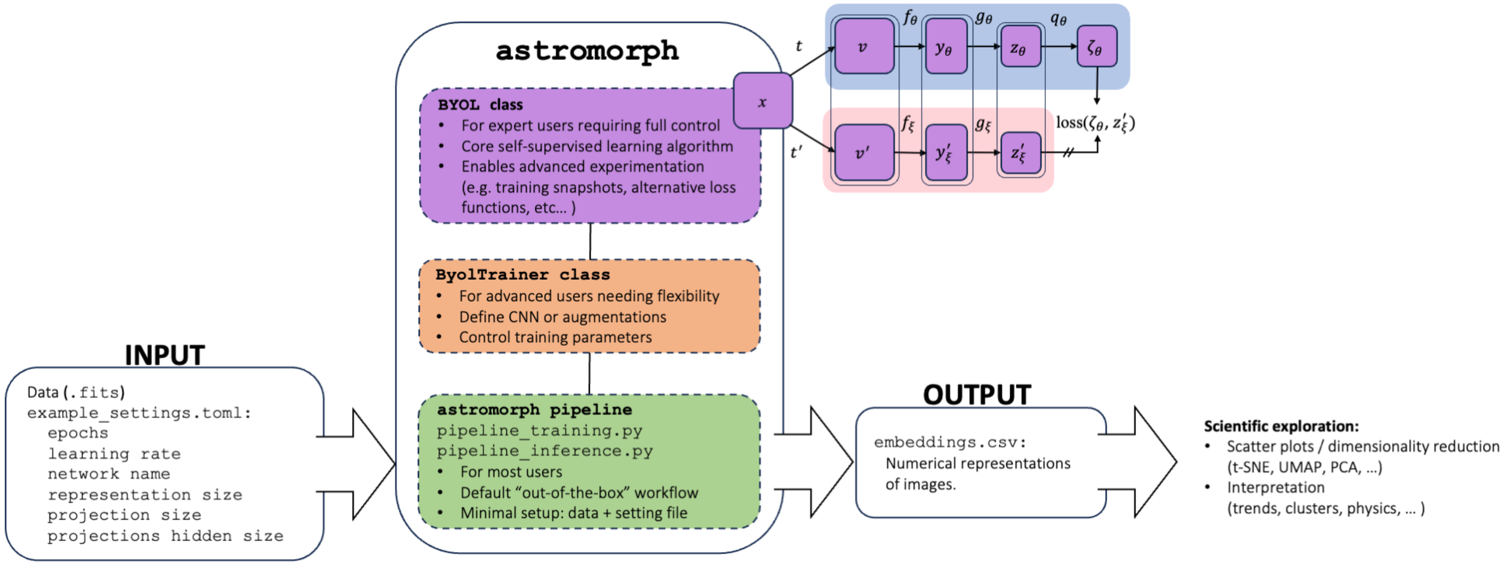
1) **`astromorph` pipeline** — "Black-Box" wrappers for training and inference (~one-liners).
2) **`ByolTrainer` class** — an option to interact/affect the training process with a medium user-level knowledge
3) **`BYOL` class** — our implementation of the BYOL methodology without (much) additional infrastructure around it. For experts.

Outcome: **Embeddings** → **Scientist** (subjective choices!) → **Science**

ASTROMORPH — (astronomer-)friendly self-supervised machine learning package for discovering patterns from unknown datasets.

`astromorph` = set of wrappers+aids (`pytorch/byol-pytorch/torchvision`) to adapt and ease usage.

Different ways to run, depending on experience level:



Outcome: **Embeddings** → **Scientist** (subjective choices!) → **Science**

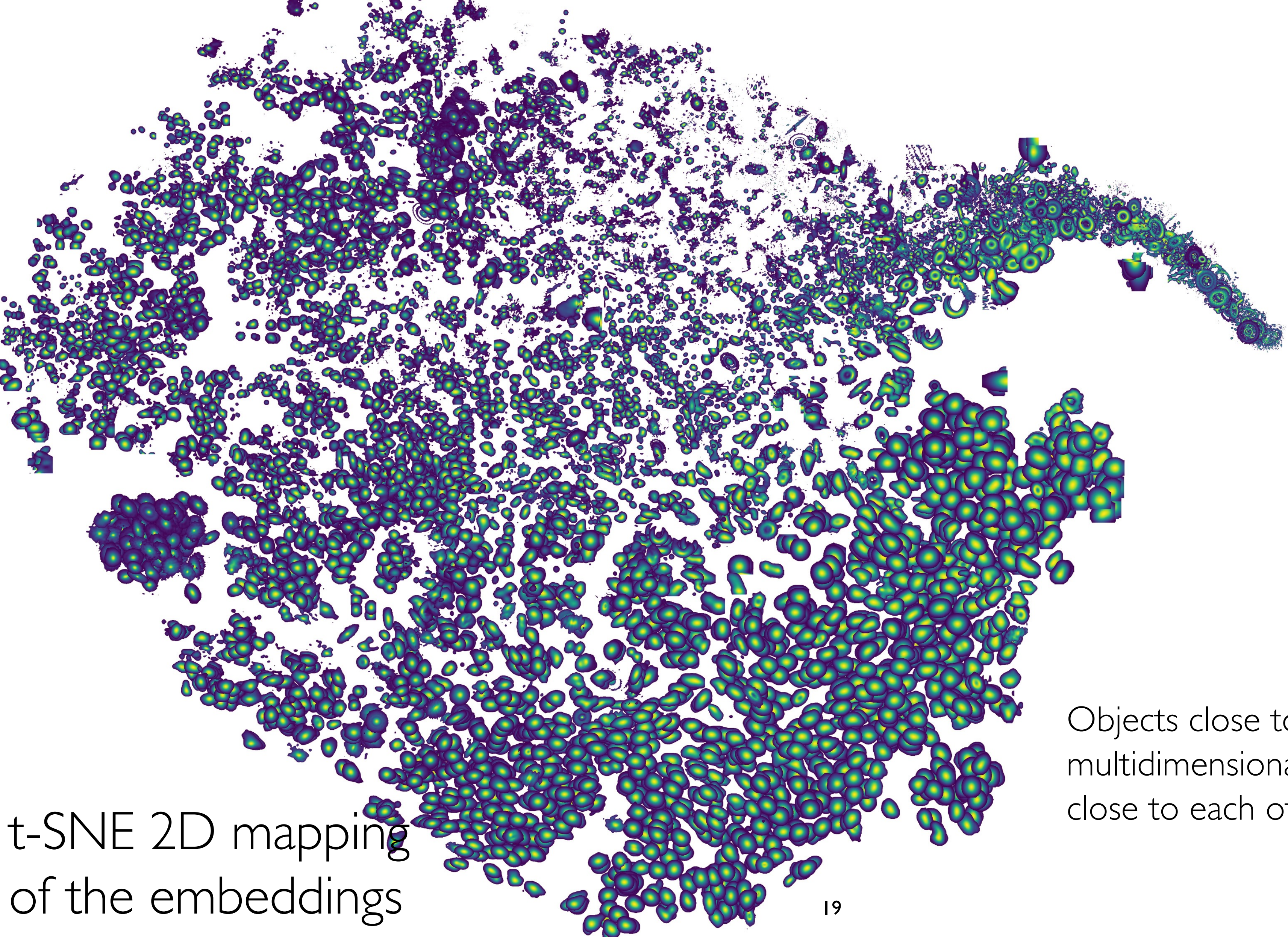
Demonstration of a user case:

1 — Exploring ALMA archive for protoplanetary disk morphologies (Bjerkeli et al. in prep.)

Premise: **~6 000 protoplanetary disk observations in ALMA archive**

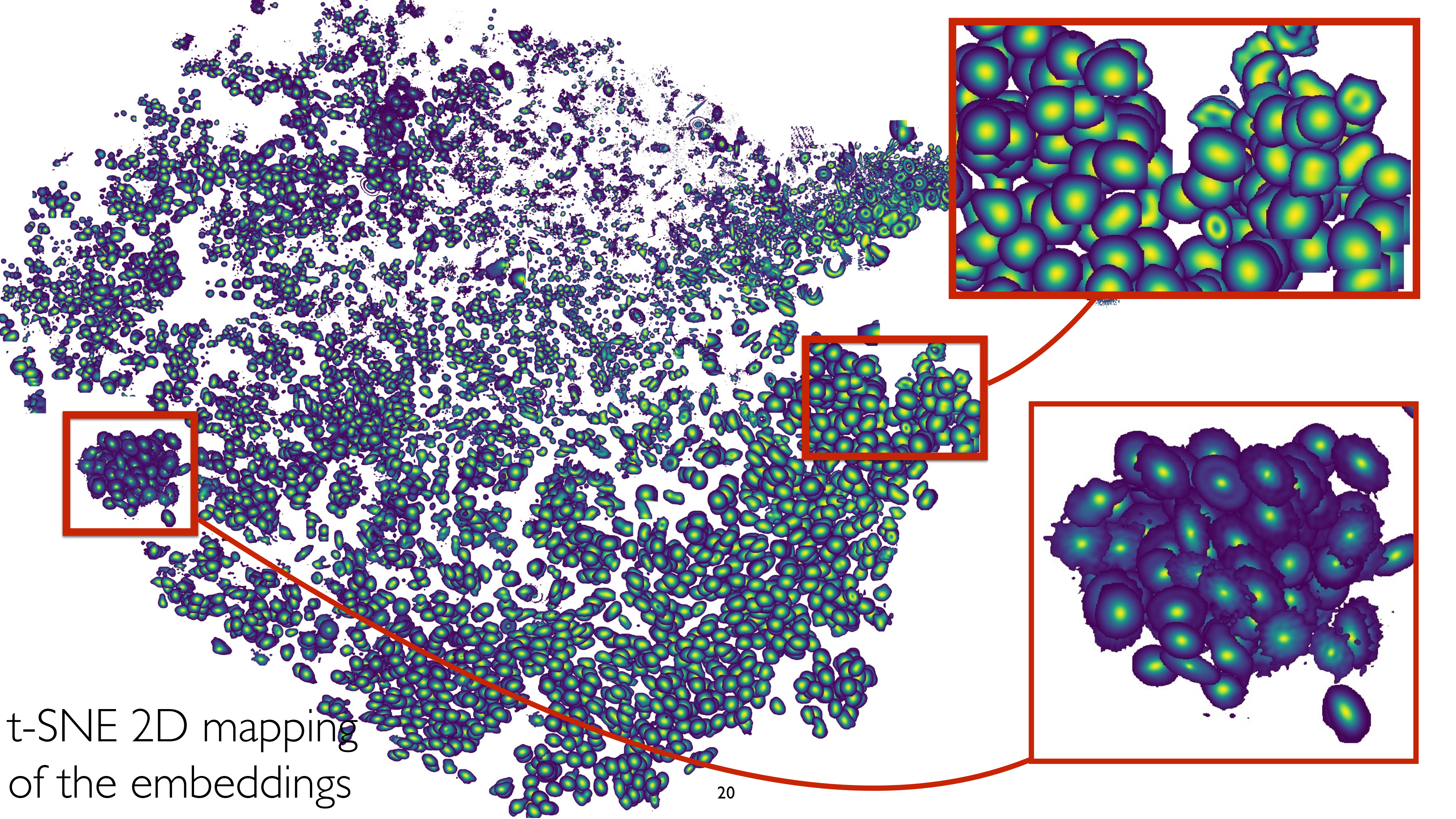
Run `astromorph` to obtain embeddings

→ Explore the data with different projection/visualization tools

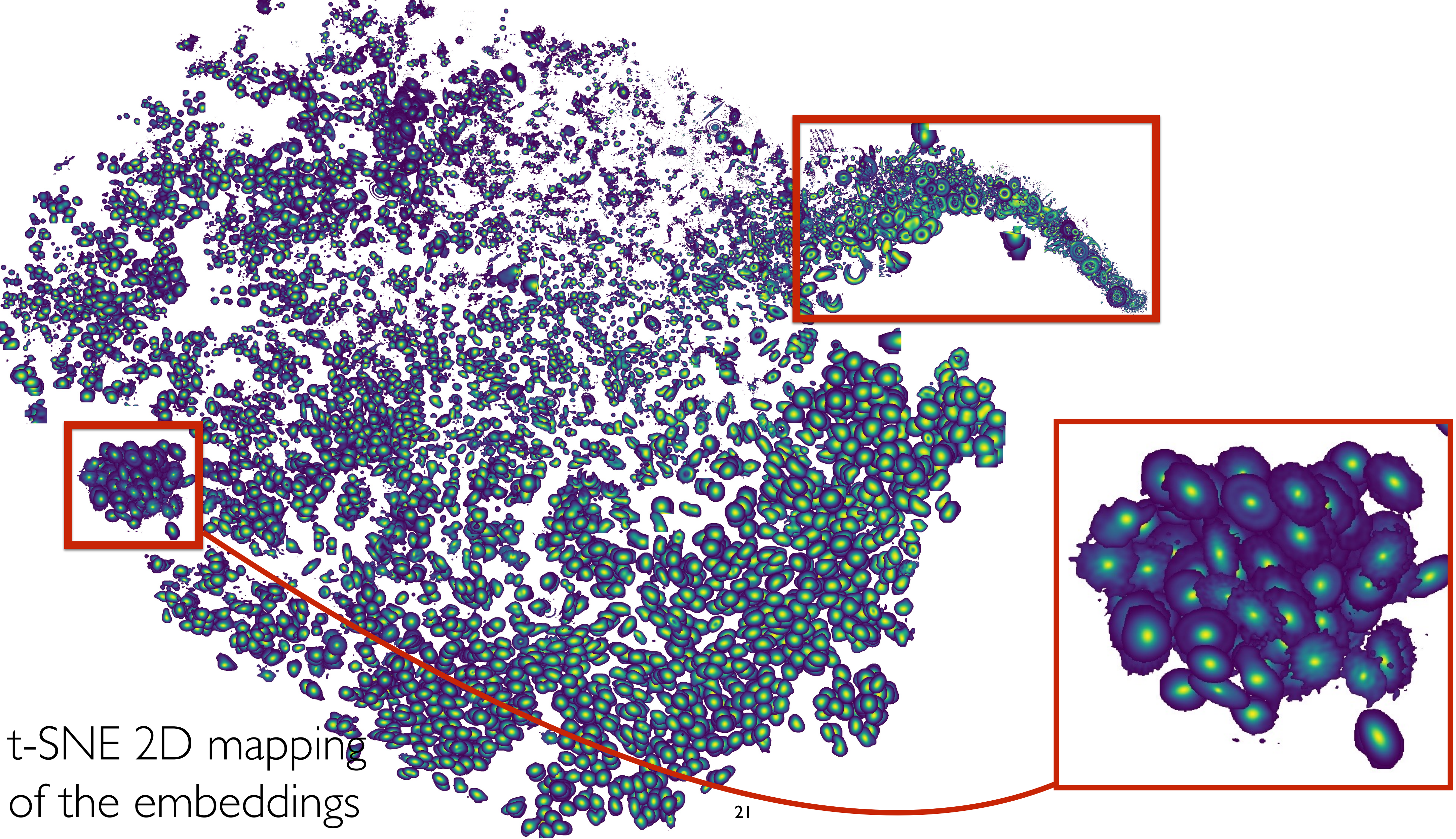


t-SNE 2D mapping of the embeddings

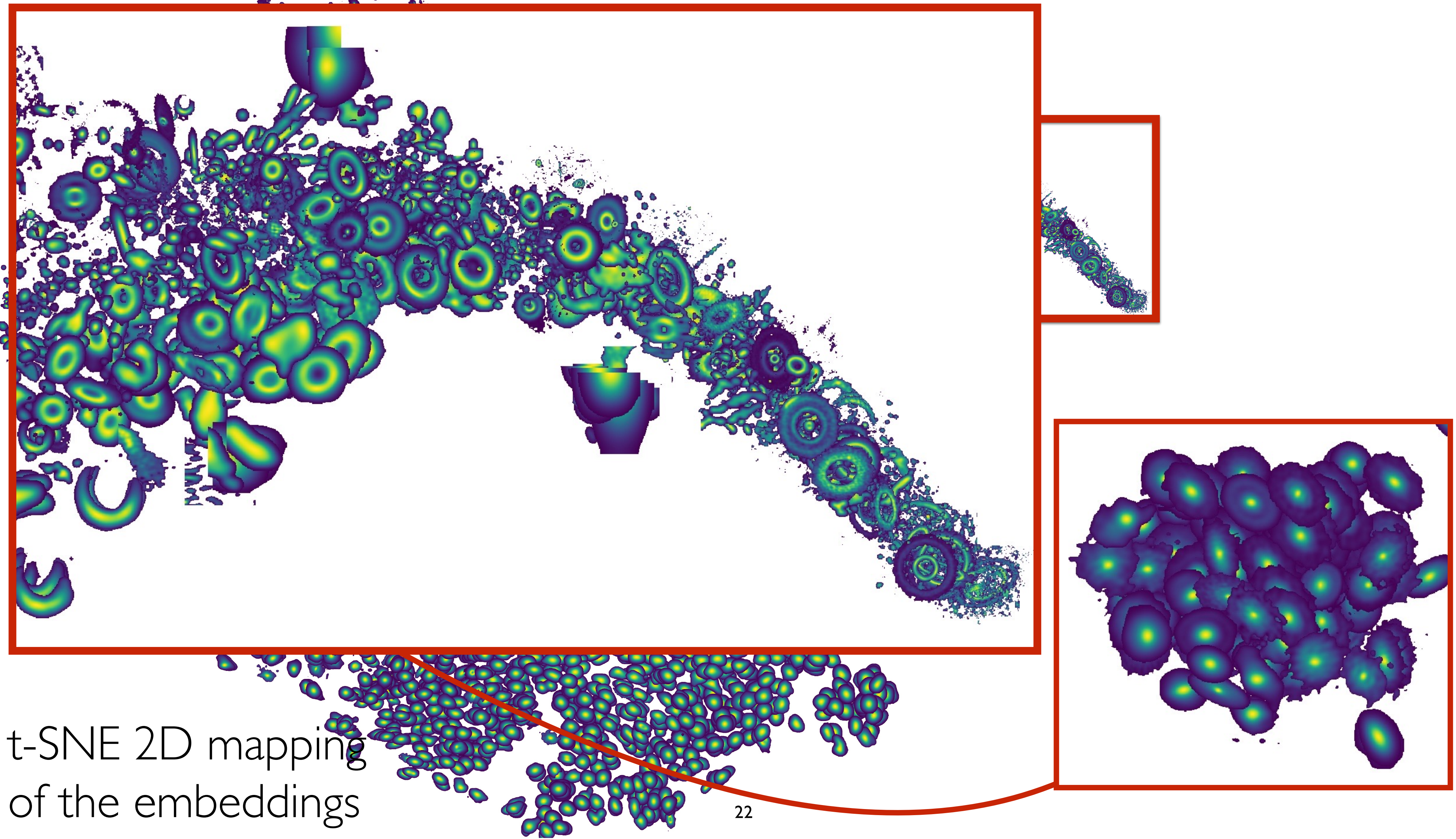
Objects close to each other in the multidimensional embedding space are close to each other in 2D space.



t-SNE 2D mapping of the embeddings



t-SNE 2D mapping
of the embeddings



t-SNE 2D mapping
of the embeddings

Demonstration of a user case:

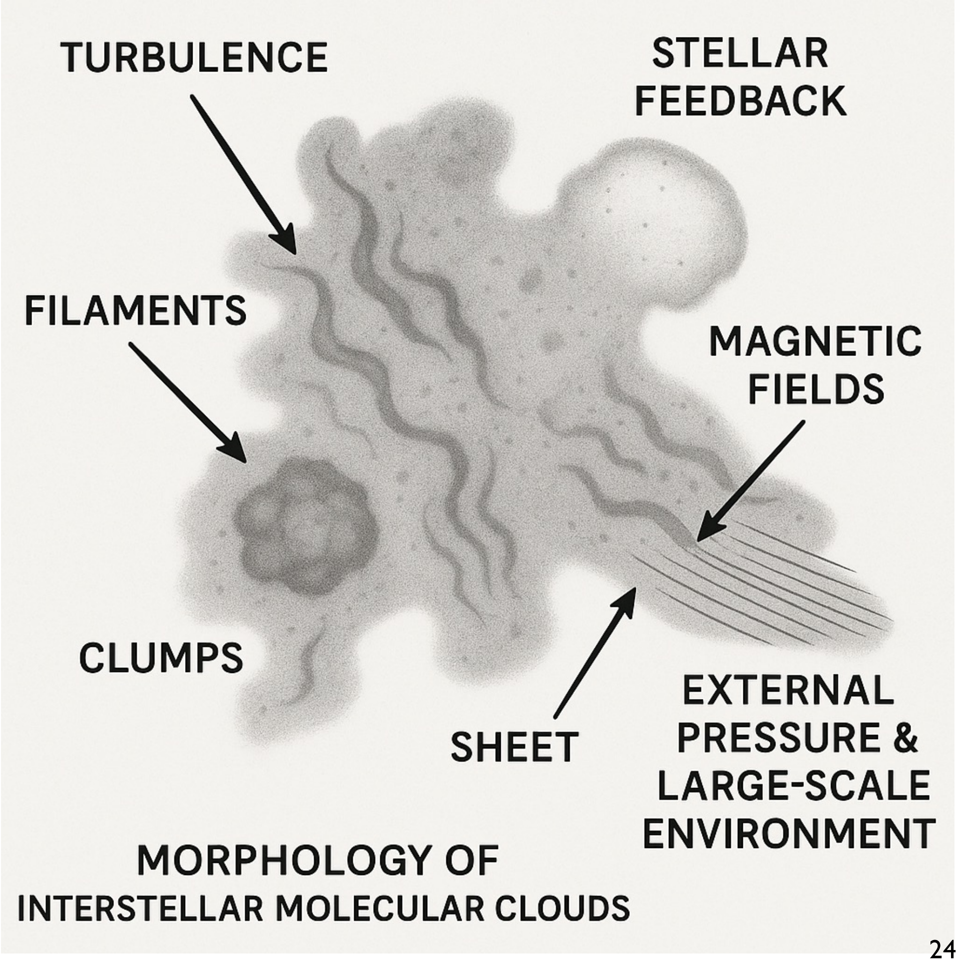
1 — Exploring ALMA archive for protoplanetary disk morphologies (Bjerkeli et al. in prep.)

… Example ways to use:

- In this case: **clear grouping —> classes, subclasses, class hierarchy**
- Finding **anomalous/similar** objects. (see also `astronomaly` by Lochner & Bassett 2021)
  - Specific case: candidates for "dust in the disk wind" (Plunkett, Bjerkeli+).
- …

Demonstration of a user case:

2 — Exploring dominant molecular cloud morphology parameters (Kainulainen et al. in prep.)



Turbulence:

filaments, hierarchy, sheets, shells, …

Gravity:

spherical/elliptical cores, clumps, hub-filaments?

Magnetic fields:

Orientation with density, filaments, sheets, …

Pressure, environment:

Cloud/clump/core edges, arcs, shear effects, …

Feedback:

bubbles, cavities, shells, pillars, PDRs, …

Thermal instability:

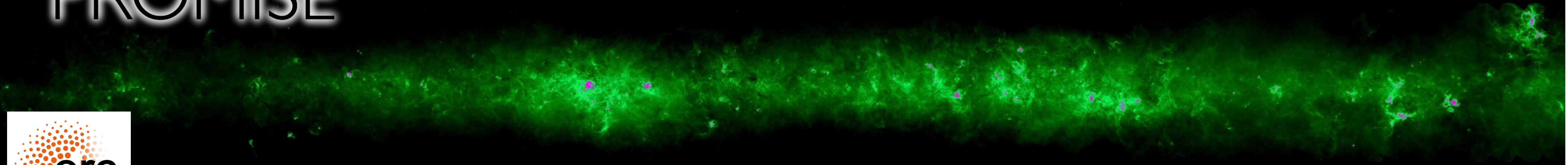
clumpy, multiphase texture, hierarchy?, …

Time evolution:

co-variance of the above, correlations with SF

PROMISE

PRobing the **O**rigins of **M**ass**I**ve molecular cloud **S**tructur**E**s

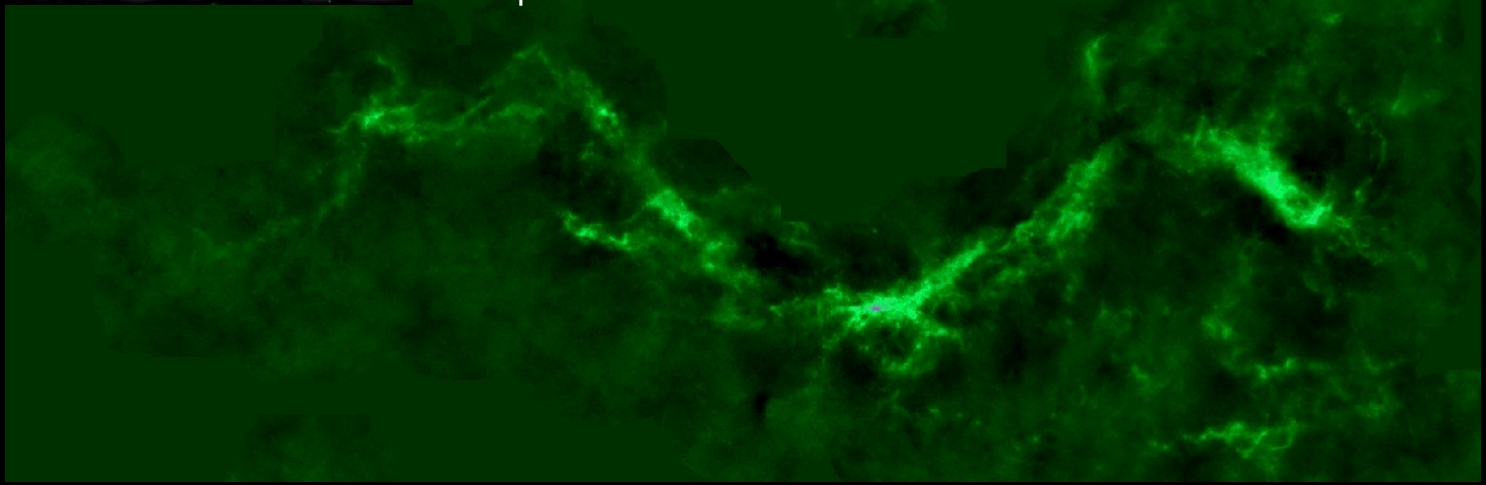


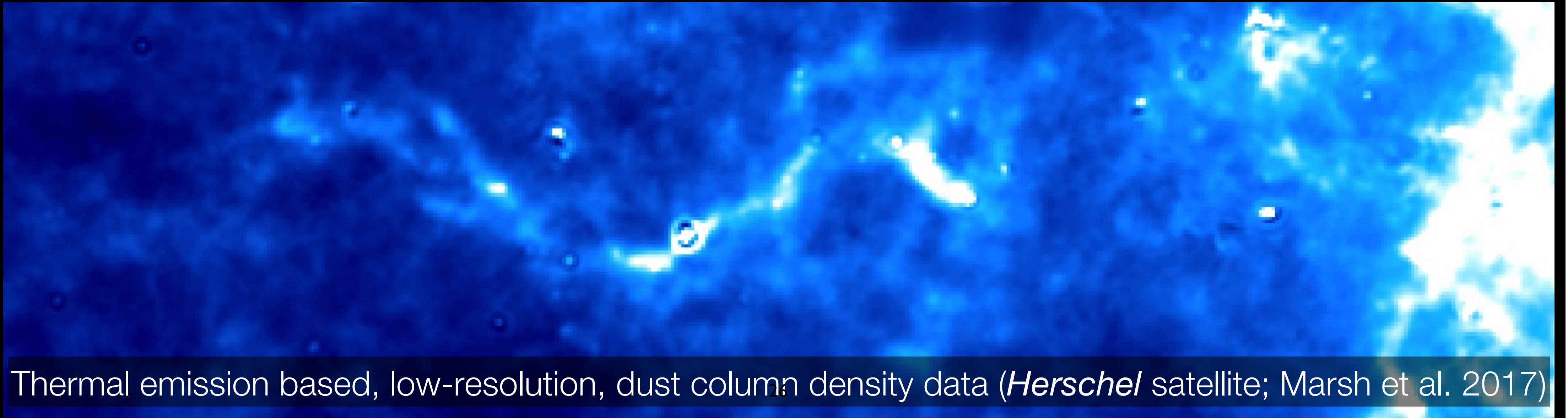
Map in the highest resolution so far (2'') of thousands of molecular clouds in the MW

Kainulainen & Zhang in prep. (Prel. works in Zhang & Kainulainen subm.; 2022)

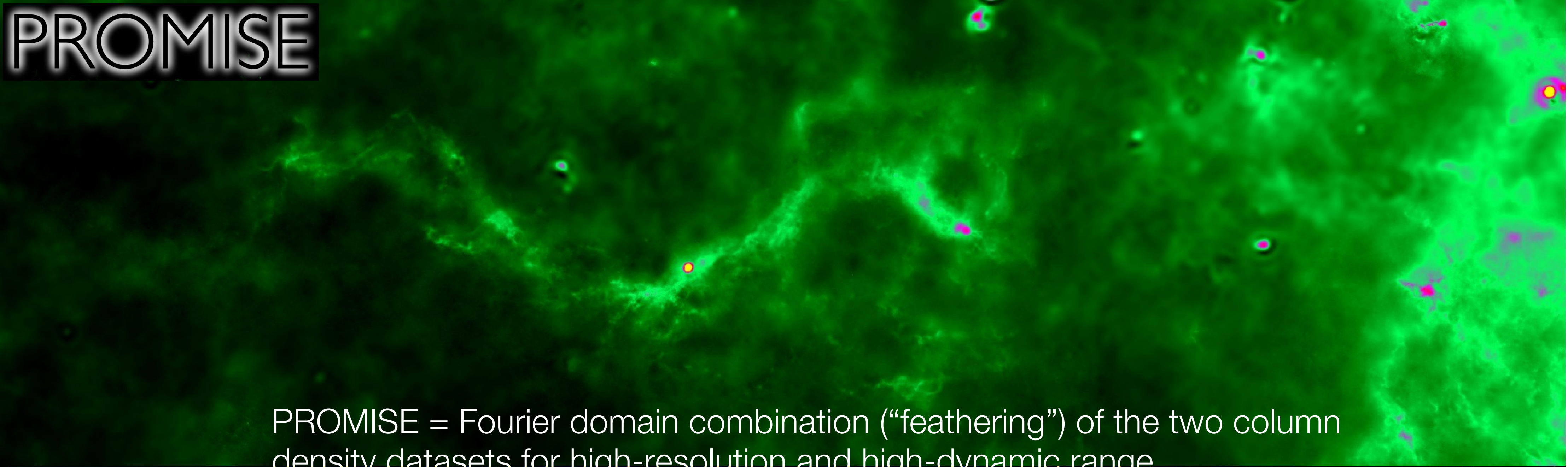
PROMISE

Spitzer 8 micron shadow-based dust column density map (infrared dark clouds)

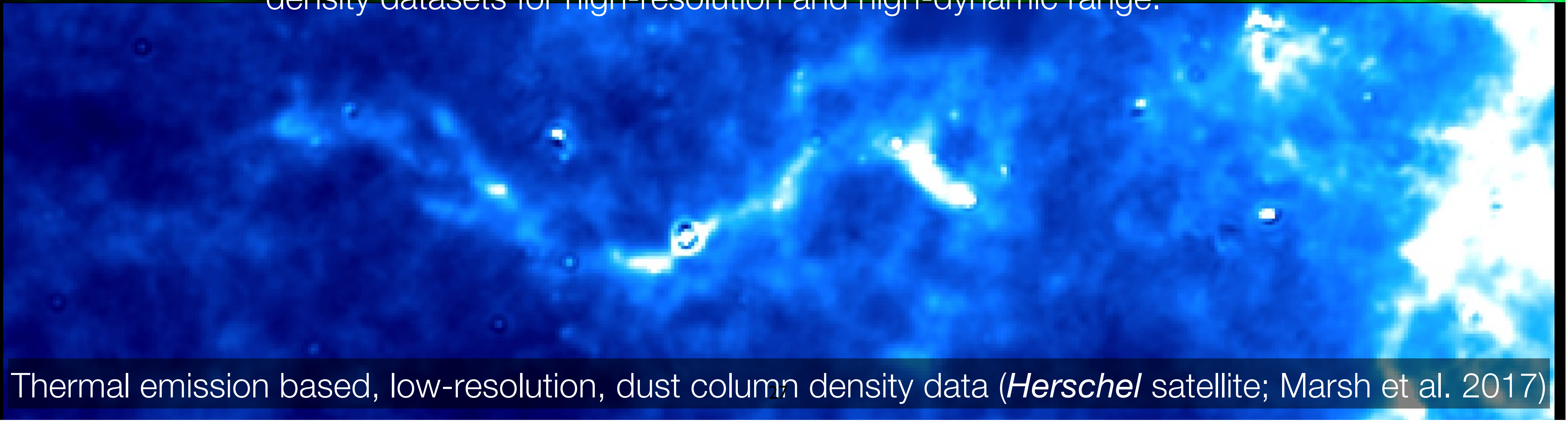




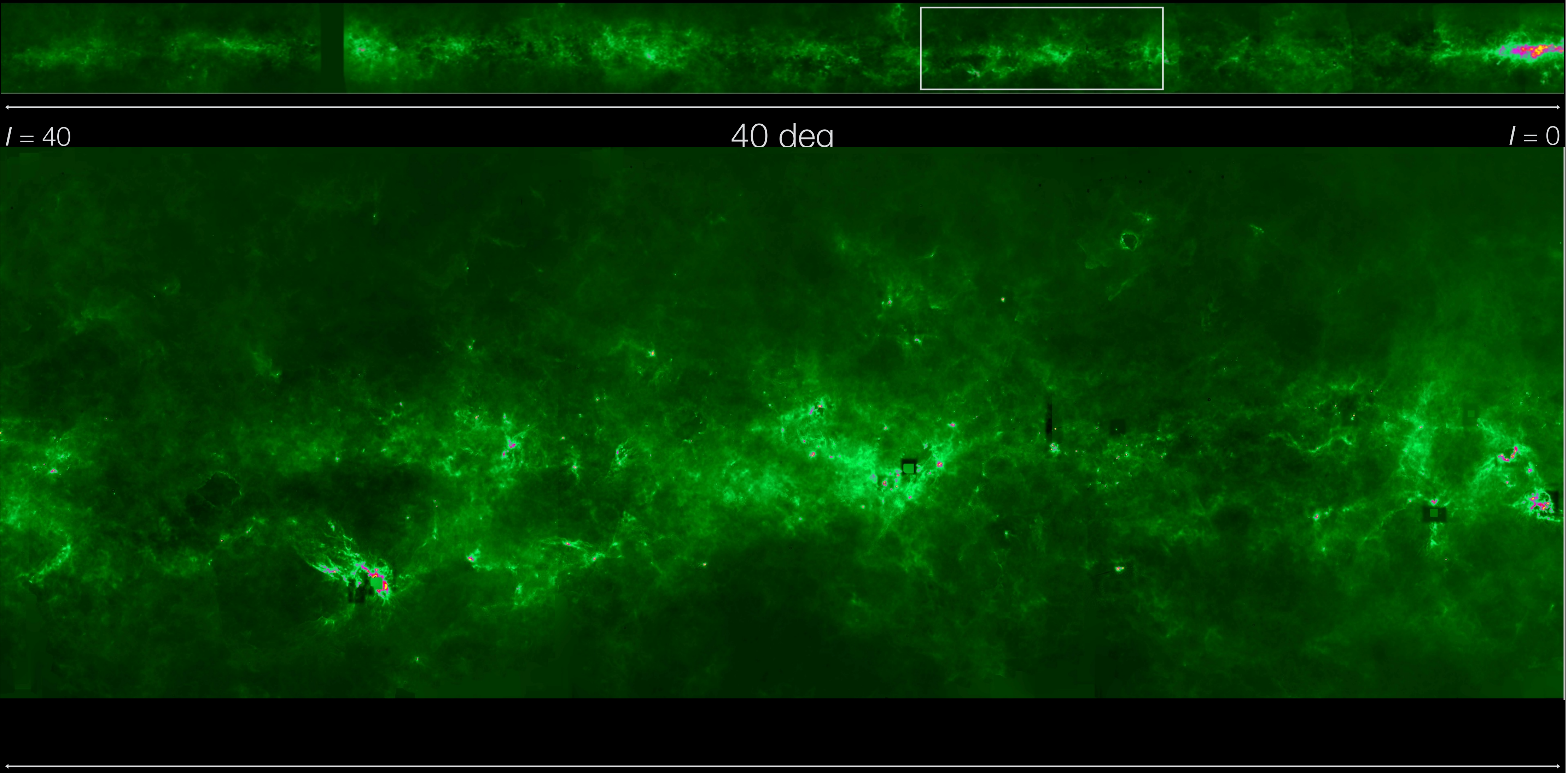
Thermal emission based, low-resolution, dust column density data (*Herschel* satellite; Marsh et al. 2017)



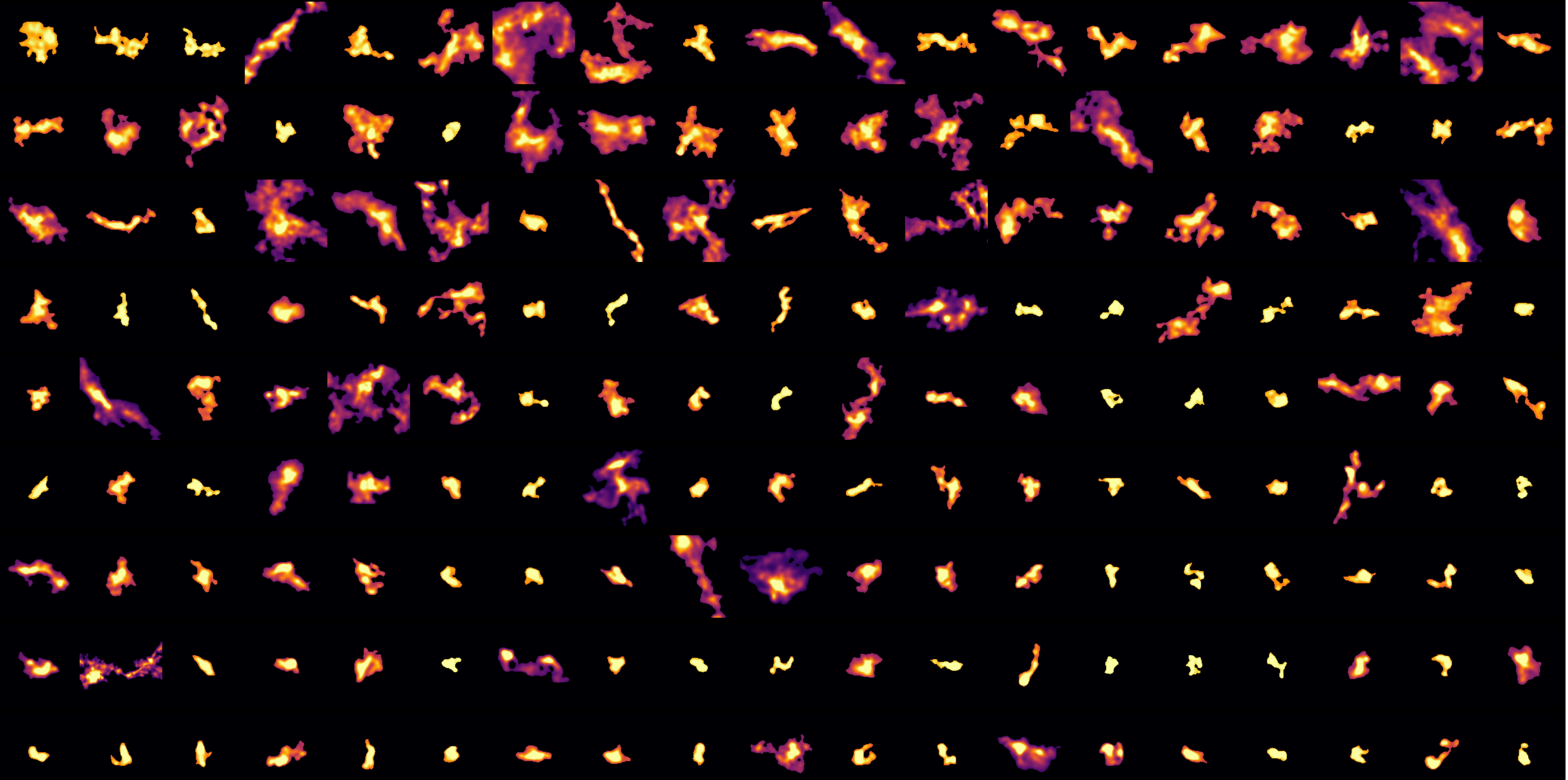
PROMISE = Fourier domain combination (“feathering”) of the two column density datasets for high-resolution and high-dynamic range.



Thermal emission based, low-resolution, dust column density data (*Herschel* satellite; Marsh et al. 2017)



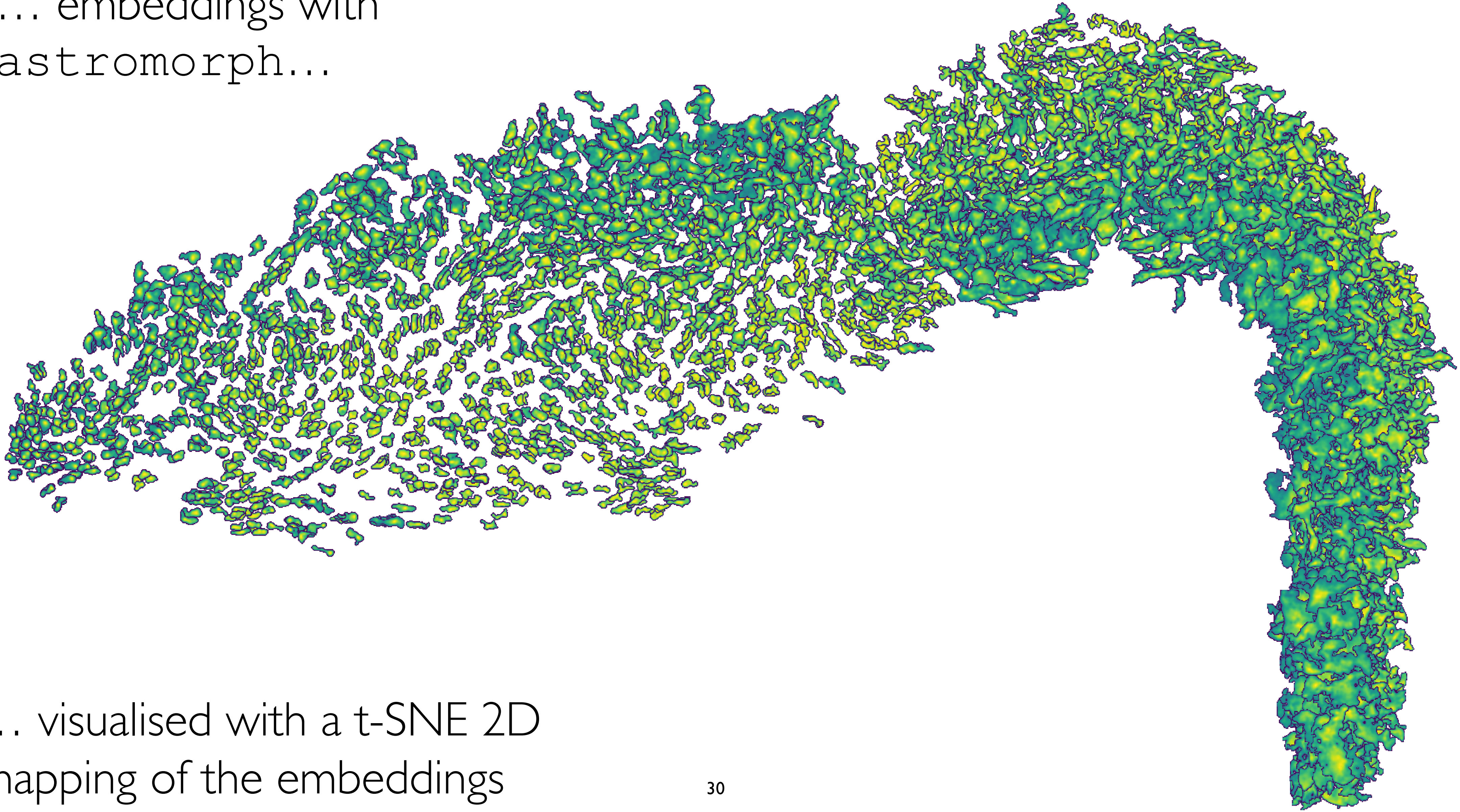
PROMISE



PROMISE

Thousands of ~pc-scale clumps, *D* from HiGal sources (Mege et al. 2021).

… embeddings with `astromorph`…



… visualised with a t-SNE 2D mapping of the embeddings

Take away and exploit:

ASTROMORPH — (astronomer-)friendly self-supervised machine learning package for exploring data and discovering patterns.

Come to talk and brainstorm about:

Decoding morphological information in terms of physically important processes/conditions/phenomena — **user cases**.