

New Methods for Efficient Representation Learning in Radio Astronomy

E. Lastufka

M. Dessauges-Zavadsky, M. Drozdova, T. Holotyak, V. Kinakh,
D. Schaerer, S. Voloshynovskiy

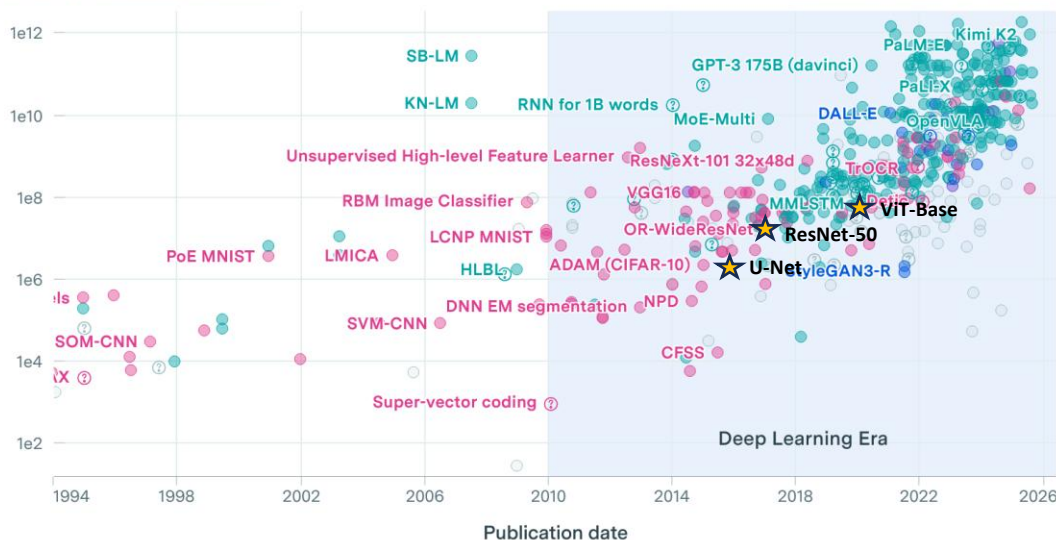


UNIVERSITÉ
DE GENÈVE

10+ years of model growth

Notable AI models

Number of trainable parameters



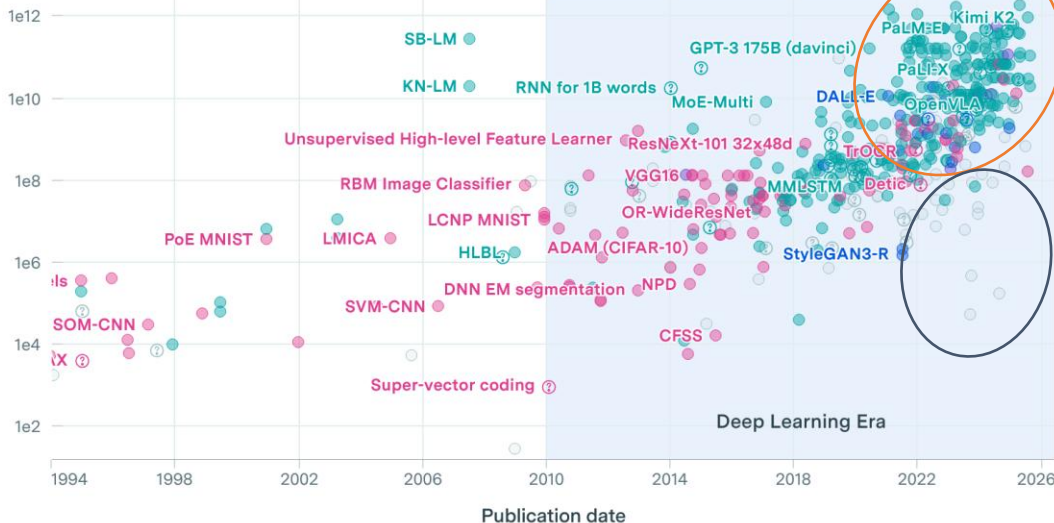
EPOCH AI

- Domain
- Language
 - Vision
 - Multimodal
 - Biology
 - Speech
 - Image generation
 - Games
 - Robotics
 - Video
 - Other

10+ years of model growth

Notable AI models

Number of trainable parameters





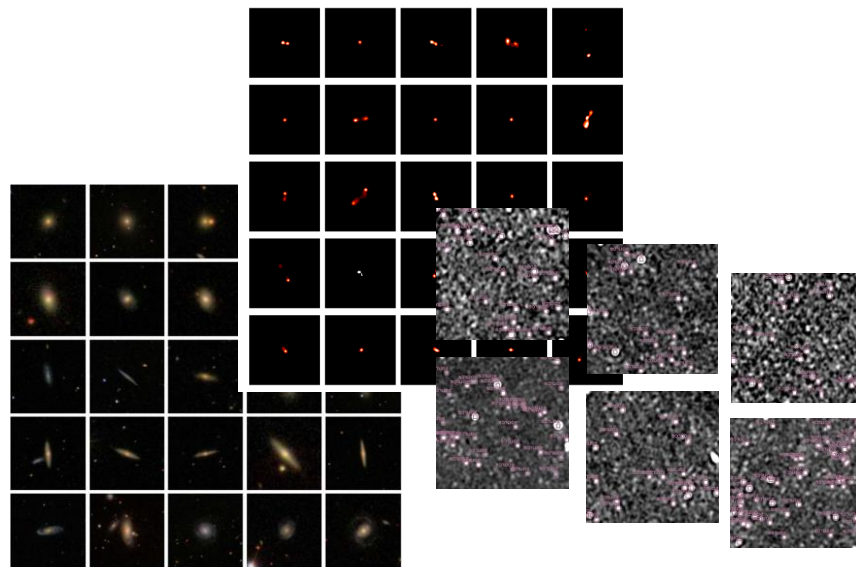
Constraints in applying foundation models to astronomy

- **Data**
 - Large amounts of unlabeled data
 - Many data modalities (images, spectra, photometry, instrumentation)
- **Deployment**
 - Limited compute and/or expertise
- **Distribution shift**
 - pre-trained models have not been exposed to an astrophysical data distribution
- **Model Evaluation**
 - Few labeled datasets for evaluation (especially in radio!)

Large amounts of unlabeled data

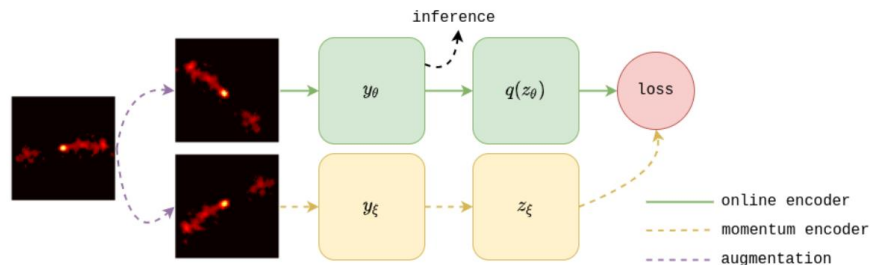
Self-supervised
learning (SSL)

Machine learning
without labels



Contrastive methods:
pairs-based
Non-contrastive
methods:
augmentation based
Loss function: cross-
entropy

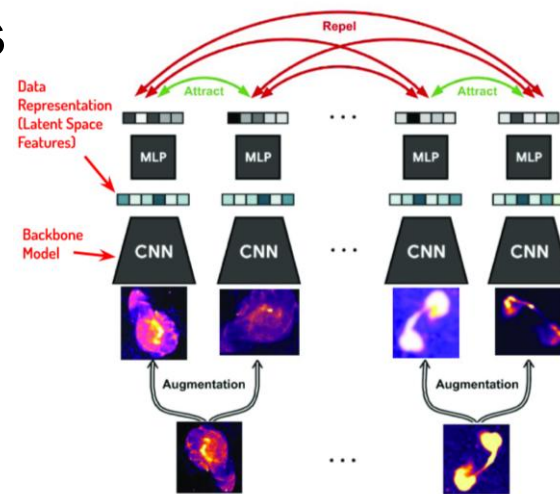
Self-Supervised Learning: Solutions



BYOL with Radio Galaxy Zoo, Slijepcevic 2024

Student-teacher framework learns from positive pairs only (image and augmentation)

Minimizes mean squared error (MSE) between student prediction and teacher representation



SimCLR with SMGPS/EMU, Riggi 2024

Requires diverse set of negative examples
Loss minimizes distance between positive pairs, maximizes distance between negative pairs

LeJEPA: Predictive latent space learning

JEPA (Joint Embedding Predictive Architecture): predict the embedding of a missing part of the input

LeJEPA theoretical claim: An isotropic Gaussian distribution minimizes downstream prediction risk

SIGreg: Sketched Isotropic Gaussian Regularization

$$\text{LeJEPA} = \text{Prediction Loss} + \text{SIGreg}$$

Align embeddings across
multiple local views to match
global view

Differentiable and scalable
regularization that forces
embeddings toward isotropic
Gaussian

In-domain LeJEPA outperforms transfer learning

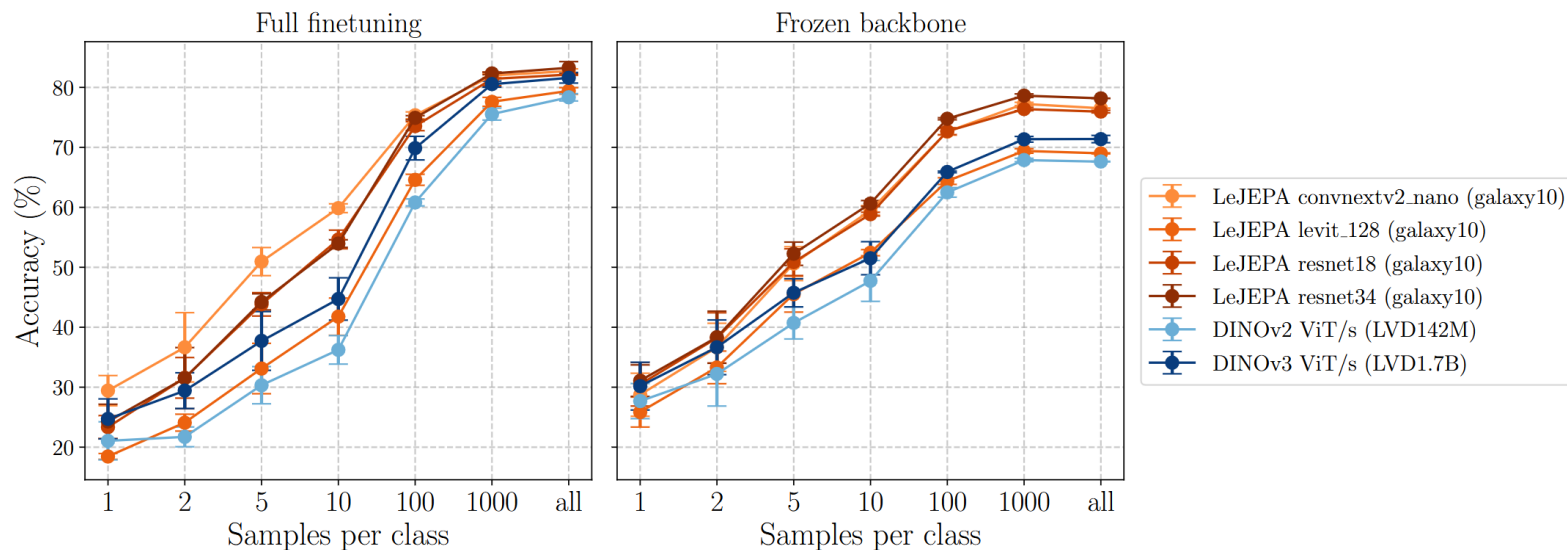


Figure 12. Small architecture in-domain (Galaxy10) LeJEPA pretraining with linear probe evaluation using frozen backbone or full finetuning (columns) and with varying number of samples per class (x-axis). We compare against state-of-the-art foundation models (DINOv2/v3, IJEPA) over 3 different random seeds. We observe that **LeJEPA enables in-domain pretraining out of the box across architectures and able to outperform frontier foundation models**. Corresponding numbers are provided in Table [3](#)

In-domain LeJEPa pre-training

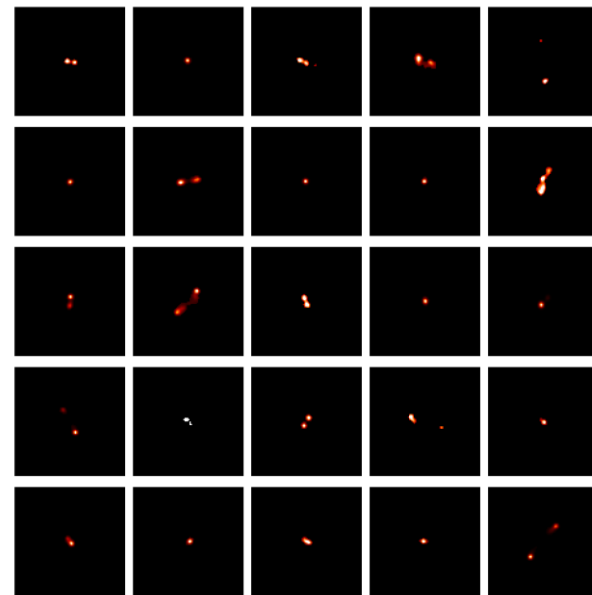
Dataset: 20k sample of Radio Galaxy Zoo (RGZ20k)

Models: ViT-Tiny (6M), EfficientNet-B0 (5M)

Evaluation datasets:

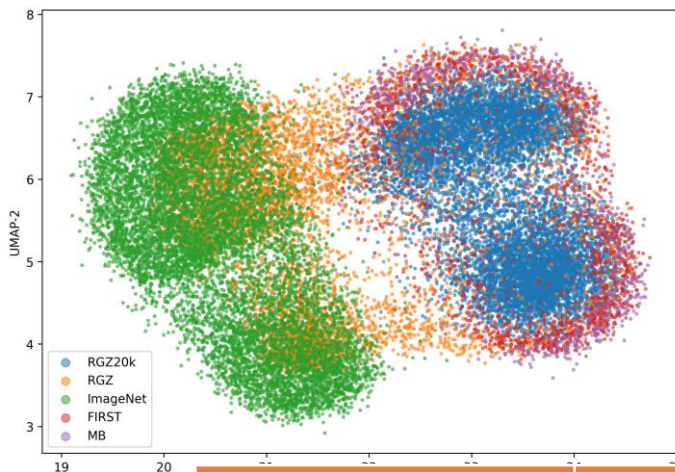
- RGZ (N-components, N-peaks)
- FIRST (FR-I, FR-II, Compact, Bent-tailed)
- MiraBest (FR-I/FR-II)

Baselines: EUPE-ViT-Tiny, EUPE-ViT-Small



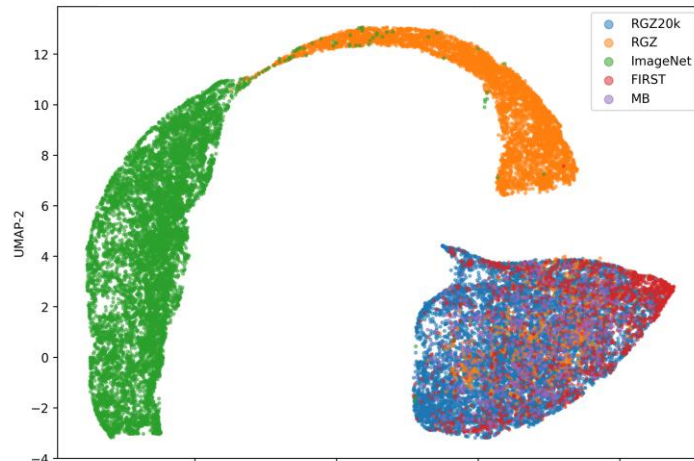
EUPE-ViT-T

6M EUPE foundation model



LeJEPA ViT-T

6M model pre-trained with LeJEPA



Maximum Mean
Discrepancy
(Fréchet distance)

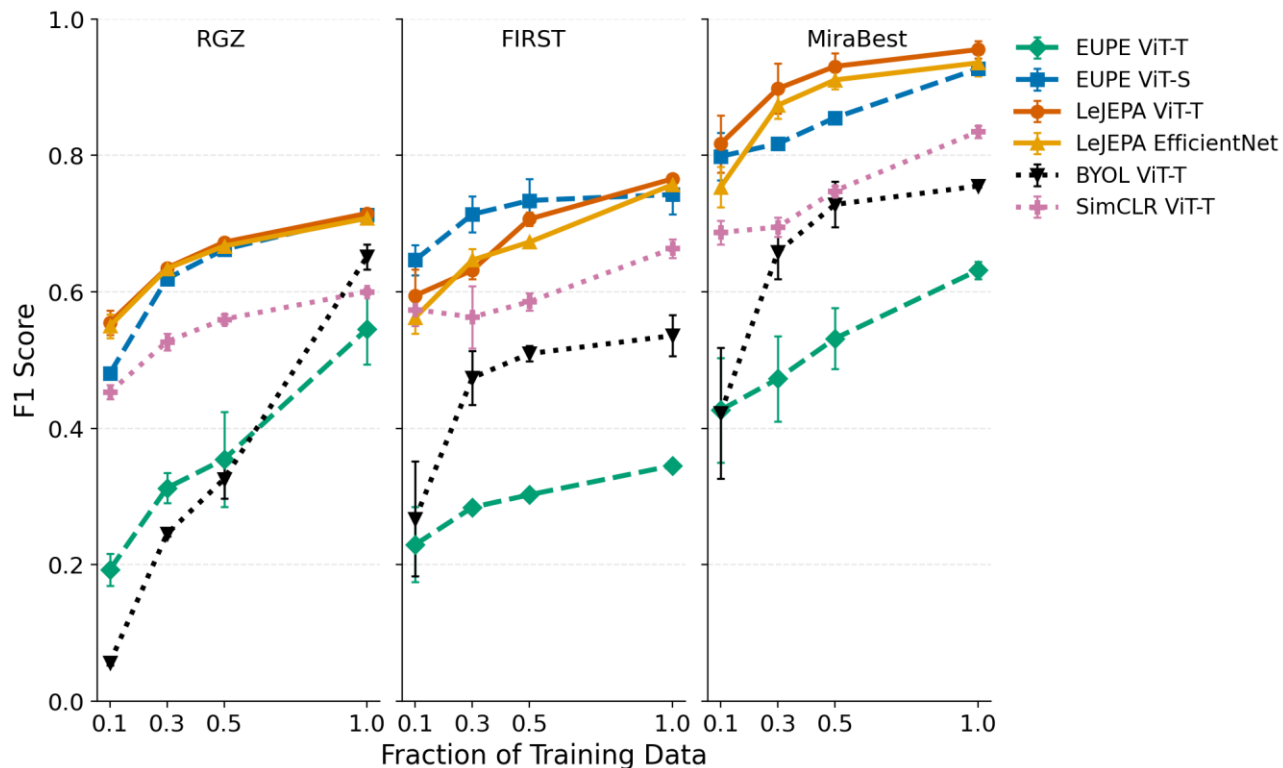
Model	RGZ20k–RGZ	RGZ20k–FIRST	RGZ20k–MB	IN–RGZ20k
EUPE-ViT-T (6M)	0.215 (0)	-0.002 (0)	0.031 (0)	0.062 (0)
EUPE-ViT-S (22M)	0.374 (-12.542)	0.127 (-2.336)	0.148 (-3.74)	0.45 (-127.337)
SimCLR ViT-T (6M)	0.794 (-74.318)	0.021 (-0.253)	0.128 (-1.289)	0.8 (-85.856)
BYOL ViT-T (6M)	0.875 (-18.784)	0.027 (0)	0.483 (-0.003)	1.592 (-43.548)
LeJEPA ViT-T (6M)	0.329 (-0.046)	0.092 (-0.005)	0.029 (-0.002)	0.897 (-3.453)
LeJEPA EfficientNet (5M)	0.324 (-2679.55)	0.039 (-153.8)	0.208 (-1503.15)	0.439 (-3967.46)

LeJEPa pre-training & radio galaxy classification

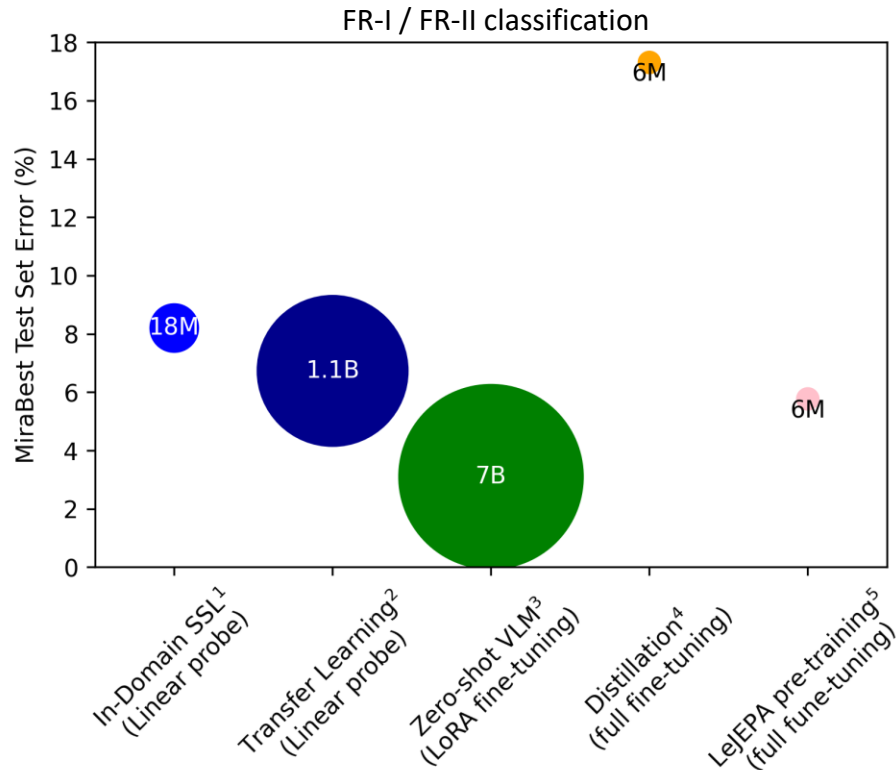
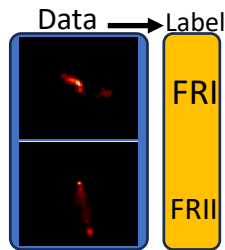
Model	RGZ F1	RGZ probe F1	FIRST F1	FIRST probe F1	MB F1	MB probe F1
LeJEPa ViT-T (6M)	$0.72 \pm 7e-3$	$0.63 \pm 1e-2$	$0.77 \pm 1e-3$	$0.58 \pm 5e-3$	$0.96 \pm 1e-2$	$0.84 \pm 5e-3$
LeJEPa EfficientNet (5M)	$0.71 \pm 2e-3$	$0.68 \pm 3e-3$	$0.76 \pm 3e-3$	$0.73 \pm 1e-2$	$0.94 \pm 2e-2$	$0.93 \pm 7e-3$
EUPE-ViT-S (22M)	$0.71 \pm 4e-4$	$0.55 \pm 3e-3$	$0.74 \pm 3e-2$	$0.59 \pm 9e-3$	$0.93 \pm 1e-4$	$0.92 \pm 1e-2$
SimCLR ViT-T (6M)	$0.60 \pm 4e-3$	$0.25 \pm 4e-3$	$0.66 \pm 1e-2$	$0.64 \pm 3e-2$	$0.83 \pm 9e-3$	$0.76 \pm 2e-2$
BYOL ViT-T (6M)	$0.65 \pm 2e-2$	$0.67 \pm 9e-3$	$0.54 \pm 3e-2$	$0.38 \pm 6e-2$	$0.76 \pm 2e-3$	$0.76 \pm 7e-2$
EUPE-ViT-T (6M)	$0.58 \pm 3e-2$	-	$0.35 \pm 7e-3$	-	$0.63 \pm 1e-2$	-

In-domain pre-training of 6M parameter model out-performs much larger (22M) foundation model!

LeJEPA pre-training & radio galaxy classification



Method, model size, and a simple metric



1 [Slijepcivic et al 2024](#)

2 [Lastufka et al 2024](#)

3 [Drozdova et al 2025](#)

4 Drozdova 2026, in prep

5 [Lastufka 2026b](#)

Various approaches in applying foundation models to astronomy

- Data
 - Large amounts of unlabeled data -> self-supervised learning (SSL)¹
 - Many data modalities (images, spectra, photometry, instrumentation) -> multimodal learning and contrastive alignment^{2,3,4}
- Deployment
 - Limited compute and/or expertise -> transfer learning⁵, parameter-efficient adaptation, distillation⁶
- Distribution shift
 - pre-trained models have not been exposed to an astrophysical data distribution -> appropriate model selection⁵, latent space re-shaping⁷
- Model Evaluation
 - Few labeled datasets for evaluation (especially in radio!)

1 [Lastufka et al 2024](#)

2 [Lastufka et al 2025b](#)

3 [Drozdova et al 2025](#)

4 [Lastufka et al 2026a](#)

5 [Lastufka et al 2025a](#)

6 Drozdova et al 2026, in prep

7 [Lastufka et al 2026b](#)