AI in Astronomy
INAF USC-C AI 1st Workshop
Catania, 21-23 May 2025

Vision-Language Models for Radio Astronomy

Simone Riggi

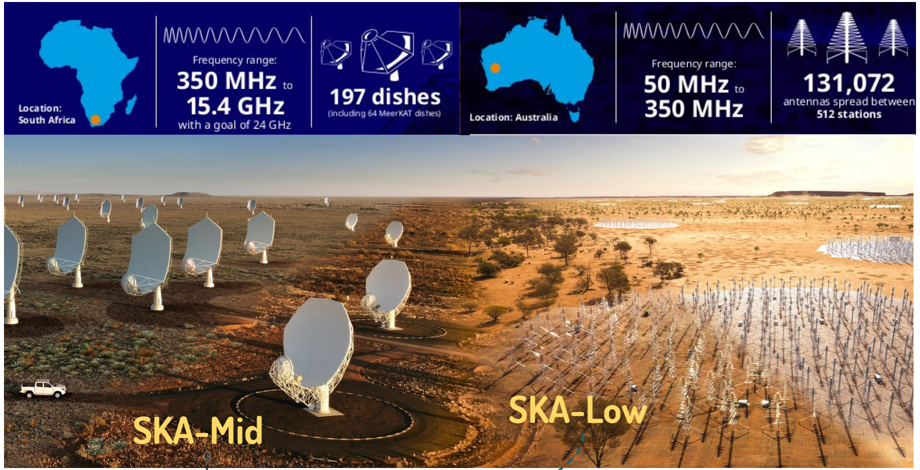
simone.riggi@inaf.it

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

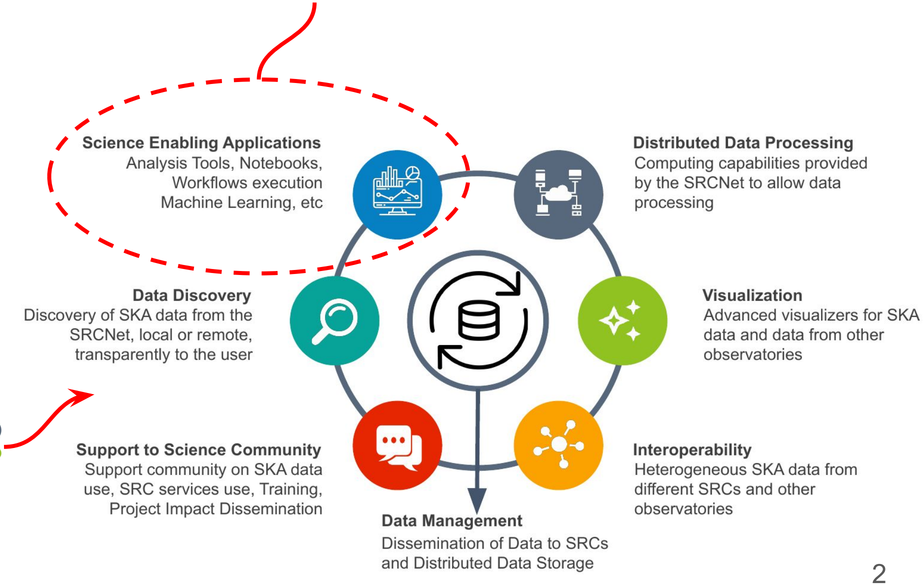
The Square Kilometer Array

SKA will generate massive radio data volumes

✓ *Automated extraction of science-ready data is a major challenge*



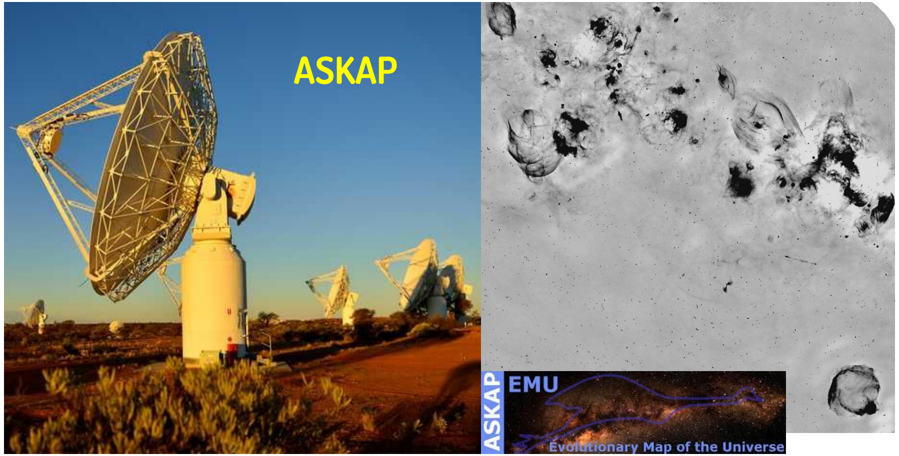
__AI/deep learning__ is a key resource for many SKA science use cases, from source finding & classification to anomaly discovery (--> SKA SRC-0000004)

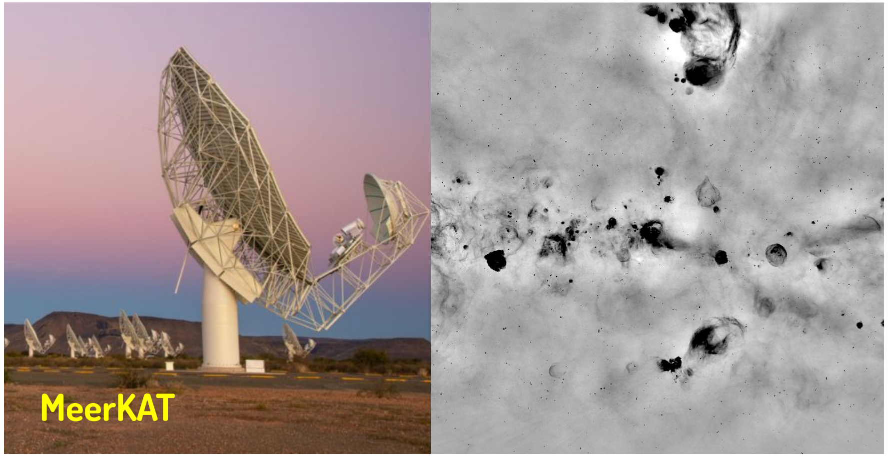


SKA Precursors

SKA precursors surveys and upgrade operations are ongoing ...

✓ *Anticipating SKA challenges, serving as analysis test benches for SKA*
✓ *Boosting new developments in data processing software*





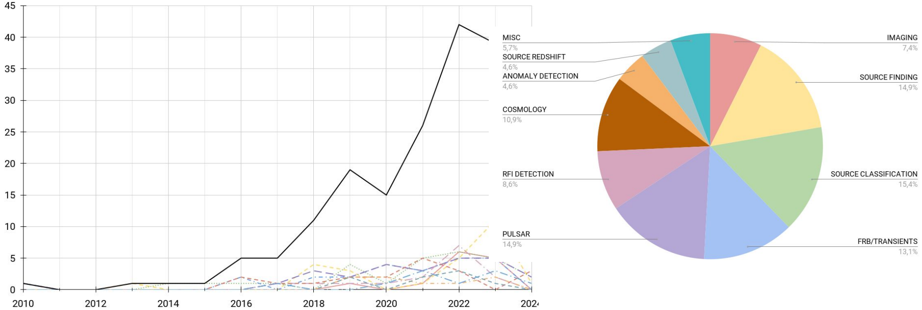
ASKAP EMU Survey

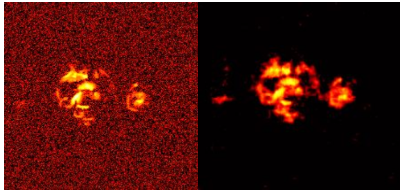
- *Early Science* (SCORPIO field, 912 MHz) (2018-2019)
- *Pilot 1* (SCORPIO field, 1243 MHz) (2019)
- *Pilot 2* (SCORPIO & other GP fields, 943 MHz) (2021)
- *EMU main survey* (943 MHz) started

SARAO MeerKAT Galactic Plane Survey (SMGPS)

- Surveyed area: |b|<1.5°, 2°<l<60°, 252°<l<358°
- Frequency: 886 – 1678 MHz
- Theoretical LAS @ 1.284 GHz: 27 arcmin

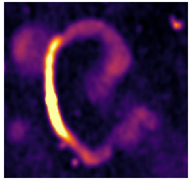
ML in Radio Astronomy





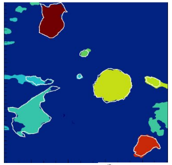
Denoising

e.g. to clean data from background or instrumentation noise patterns



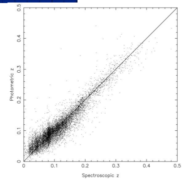
Anomaly Detection

e.g. to search for astrophysical objects with peculiar morphology or to identify outliers/anomalies in the data



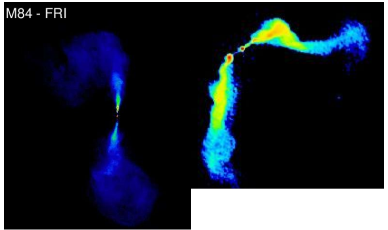
Detection/Segmentation

e.g. to extract sources from maps or to segment source regions



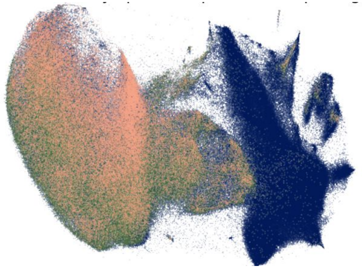
Regression/Inference

e.g. to estimate source physical parameters (redshift, flux, ...) or model parameters



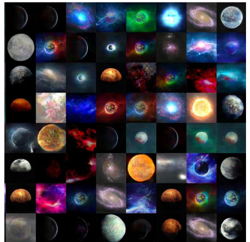
Classification

e.g. to classify source type or morphology, or to find sub-groups



Representation Learning

e.g. to extract features/parameters from the data, view/discover hidden patterns



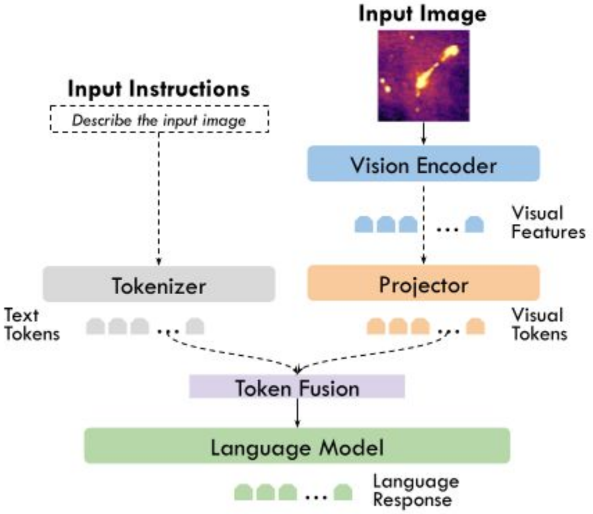
Data Generation

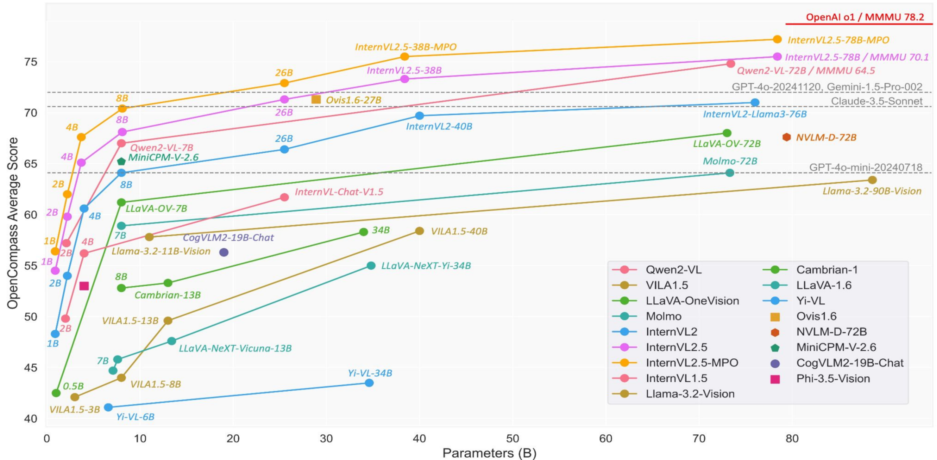
e.g. to generate simulated/synthetic data, map inpainting, etc

Vision Language Models (VLMs)

- **VLMs used for various tasks on image + text inputs**
  - ✓ *Image Captioning*
  - ✓ *Visual Question Answering (VQA)*
  - ✓ *Text-to-Image*
  - ✓ *Image Search/Retrieval*

- **Why exploring VLMs for astronomy?**
  - ✓ *Instruct and perform visual tasks (data quality assessment, source detection/classification, anomaly/similarity search, etc) without coding through a text-based interface and examples*





Leaderboards of open/commercial VLMs (March 2025)

Rank	Method	Param (B)	Avg Score
1	InternVL 2.5-78B-MPO	78	80.3
2	InternVL 2.5-38B-MPO	38	78.3
3	Qwen2.5-VL-72B	73.4	78.1
4	Ovis2-34B	34.9	77.5
5	InternVL 2.5-26B-MPO	26	76.4

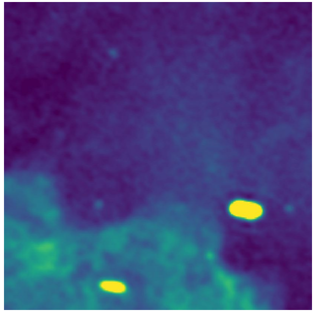
Rank* (UB)	Model	Arena Score
1	ChatGPT-4o-latest (2025-01-29)	1279
1	GPT-4.5-Preview	1269
1	Gemini-2.0-Flash-Thinking-Exp-01-21	1278
3	Gemini-2.0-Pro-Exp-02-05	1241
4	o1-2024-12-17	1232
4	Claude 3.7 Sonnet	1223
4	Gemini-2.0-Flash-001	1227
5	Gemini-1.5-Pro-002	1222

Radio Benchmarks

We have set up 6 benchmarks on radio source detection/classification

B1 – Extended/diffuse radio source detection

- ✓ Multi-label multi-class classification
- ✓ Dataset: ~5,700 ASKAP/MeerKAT labelled images with extended (33%), diffuse (4%) and large-scale diffuse (7%) radio sources



Sample image
{EXTENDED, DIFFUSE-LARGE}

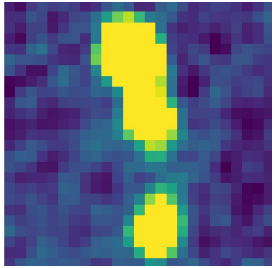
Context: …

Question: **Which of these morphological classes of radio sources do you see in the image?**

EXTENDED
DIFFUSE
DIFFUSE-LARGE
NONE

B2 – Source morphology classification

- ✓ Single-label multi-class classification
- ✓ Dataset: ~3,800 VLA source-zoomed images of 6 morphological classes (~600 per class): 1C-1P, 1C-2P, 1C-3P, 2C-2P, 2C-3P, 3C-3P



Context: …

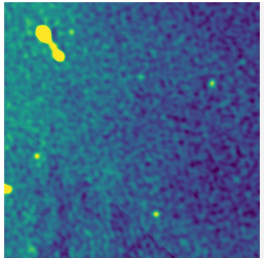
Question: **Which of these morphological classes of radio sources do you see in the image?**

1C-1P
1C-2C
1C-3P
2C-2P
2C-3P
3C-3P

Radio Benchmarks

B3 – Extended Radio Galaxy Detection

- ✓ Binary classification
- ✓ Dataset: ~5,700 ASKAP/MeerKAT images with extended radio galaxies (~18%)



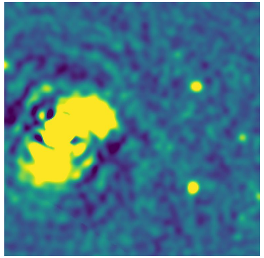
Context: …

Question: **Do you see any likely radio galaxy with an extended morphology in the image?**

YES
NO

B4 – Artefact Detection

- ✓ Binary classification
- ✓ Dataset: ~5,700 ASKAP/MeerKAT images with imaging artefacts (~7%)



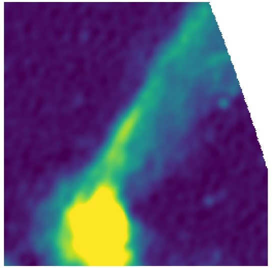
Context: …

Question: **Do you see any imaging artefact with a ring pattern around bright sources in the image?**

YES
NO

B5 – Image peculiarity classification

- ✓ Single-label multi-class classification
- ✓ Dataset: ~5,700 ASKAP/MeerKAT images labelled as ORDINARY (~63%), COMPLEX (~35%), PECULIAR (~3%)



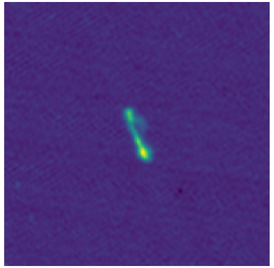
Context: …

Question: **Can you identify which peculiarity class the presented image belongs to?**

ORDINARY
COMPLEX
PECULIAR

B6 – Radio Galaxy Morphology Classification

- ✓ Binary classification
- ✓ Dataset: ~832 VLA images centred on FR-I (~48%) and FR-II (~52%) radio galaxies



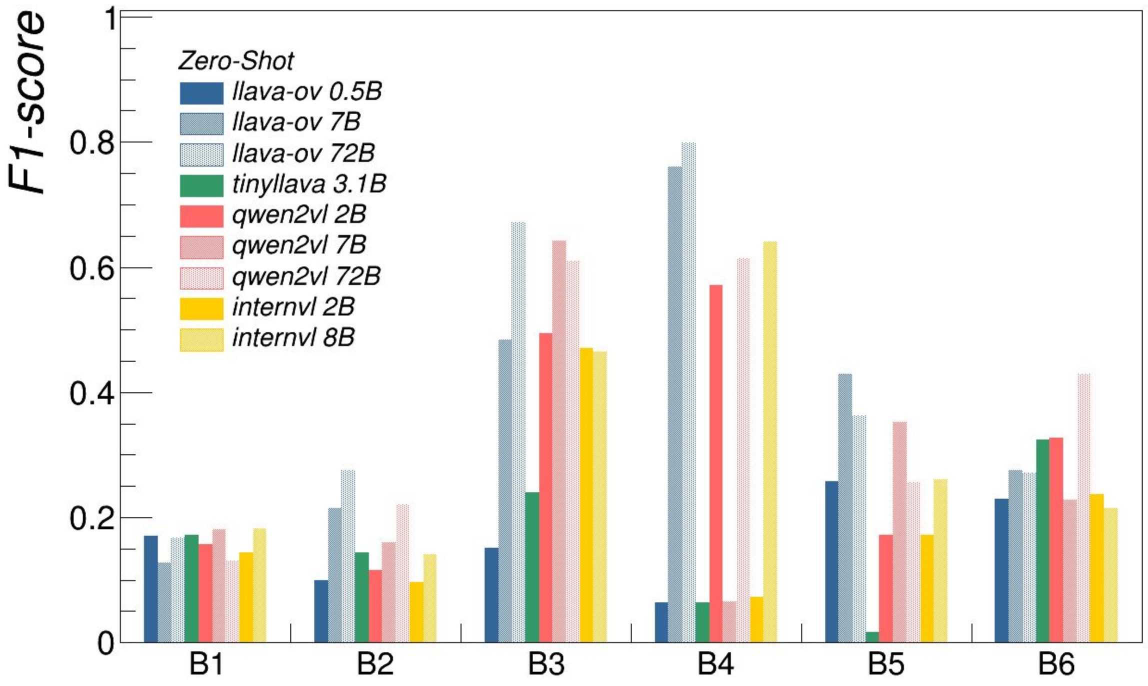
Context: …

Question: **Which of these morphological classes of radio galaxy do you see in the image?**

FR-I
FR-II

Zero-Shot Results

Evaluating open VLMs of various sizes on radio benchmarks



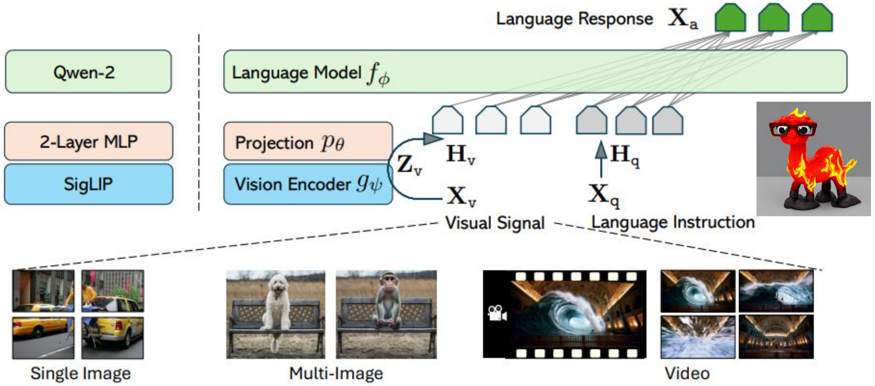
- ✓ *Overall, poor performance across all benchmarks*
- ✓ *Larger models performing better, as expected*
- ✓ *Best results on B3 (galaxy detection) & B4 (artefact detection)*
- ✓ *LLaVA slightly outperforming other models on most benchmarks*

→ **need models specialized on radio tasks**

The *radio-llava* model

Fine-tuning LLaVA-OneVision 7B on radio data

✓ *Vision encoder frozen*
✓ *LLM (qwen2) & projector free*
✓ *Full vs LORA fine-tuning tested*
✓ *Shallow (1 epoch) vs deeper (3 epochs) fine-tuning*
✓ *Standard vs alternative hyperparameter choices*

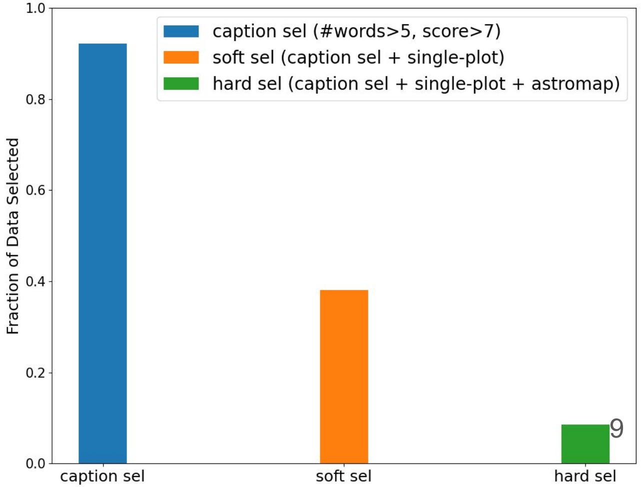


Two training datasets created

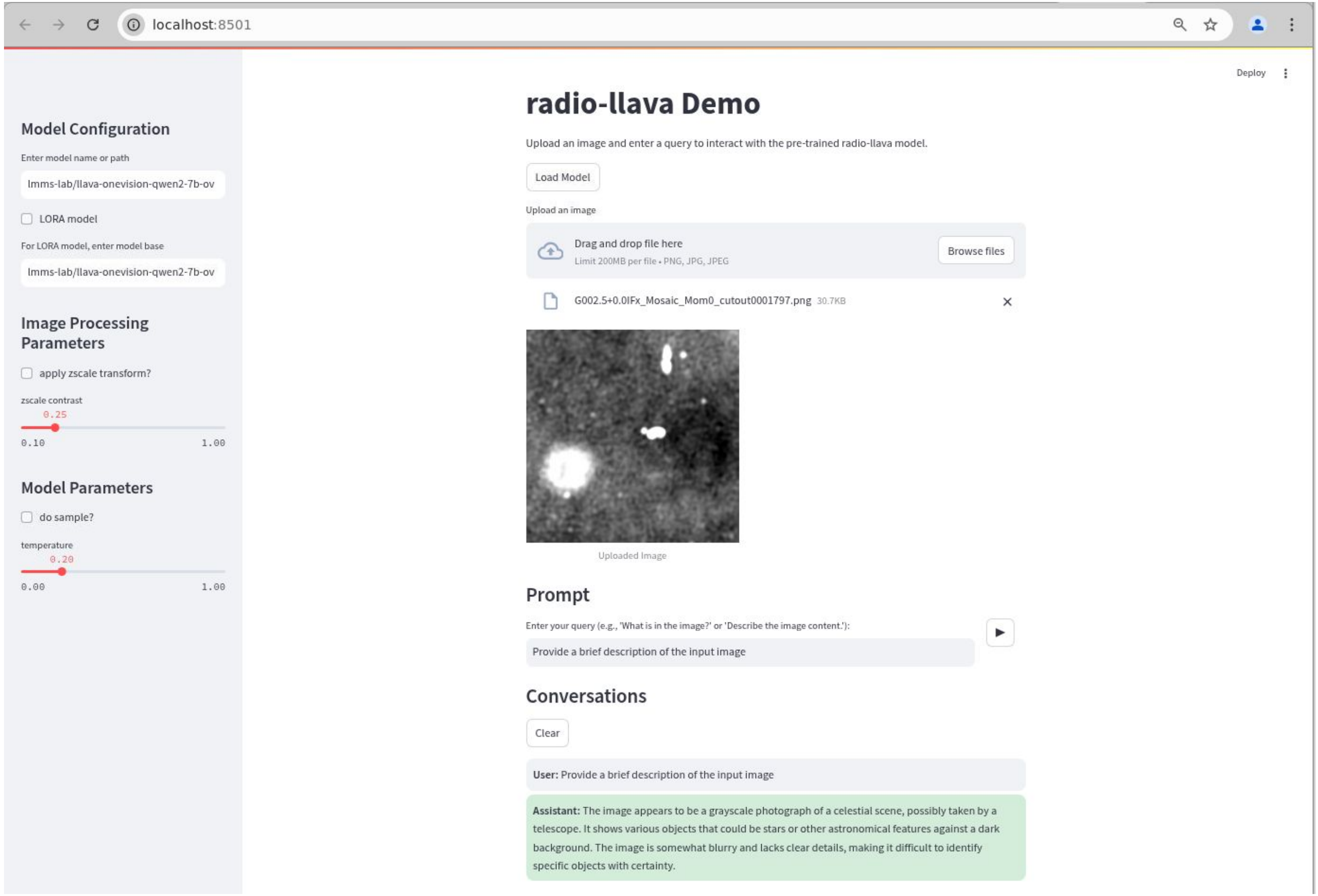
- **Q&A dataset**
  - ✓ assembled from *fine-* & *coarse-grained* annotated radio datasets
  - ✓ 59k radio images (source-zoomed & wide-field, ASKAP/MeerKAT + others)
  - ✓ ~1.5 M user-assistant conversations generated from label information

    Q: Can you describe the image content?
    Q: Can you provide the bounding box of sources with class X in the image?
    Q: Do you see source/pattern X in the image?
    ...
- **Caption dataset**
  - ✓ assembled from arXiv papers (2000-2025) on radioastronomical topics
  - ✓ ~38k figure-caption pairs after quality selection *(soft sel)*

    Q: Can you describe the image



The *radio-llava* model

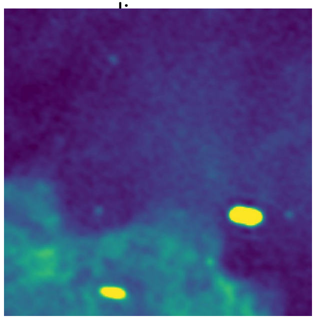


For more details:
Riggi et al, 2025,
https://arxiv.org/abs/2503.23859

Fine-tuning Results (B1)

B1 – Extended/diffuse radio source detection

- ✓ Multi-label multi-class classification
- ✓ Dataset: ~5,700 ASKAP/MeerKAT labelled images with extended (33%), diffuse (4%) and large-scale diffuse (7%)

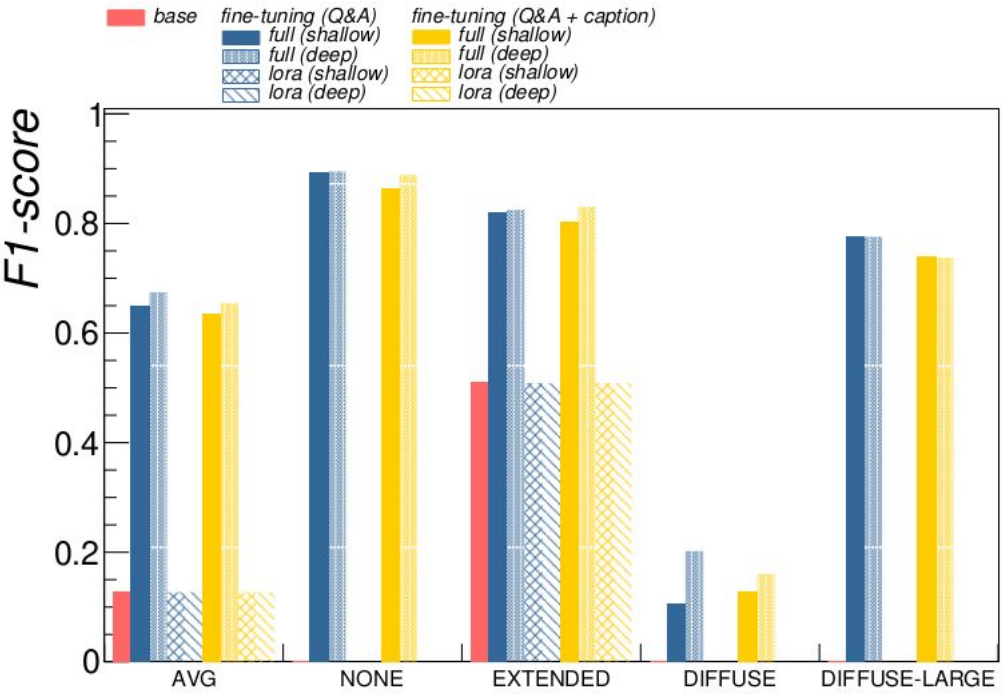


Sample image
{EXTENDED, DIFFUSE-LARGE}

\# Context: ...

\# Question: **Which of these morphological classes of radio sources do you see in the image?**

EXTENDED
DIFFUSE
DIFFUSE-LARGE
NONE

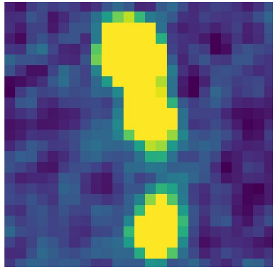


- ✓ *+40% wrt base model*
- ✓ *Caption data slightly degrade*
- ✓ *Marginal improvements with deeper fine-tuning*
- ✓ *Diffuse sources still largely missed*
- ✓ *LORA fine-tuning not effective*
- ✓ *Suboptimal wrt vision-only fine-tuning (F1~80%)*

Fine-tuning Results (B2)

B2 - Source morphology classification

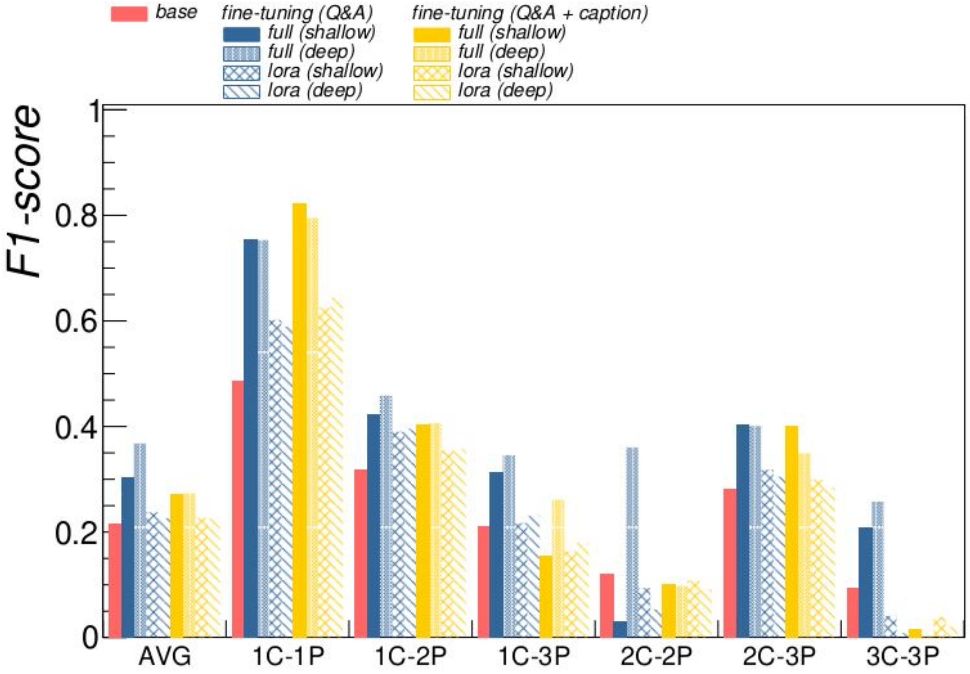
- ✓ Single-label multi-class classification
- ✓ Dataset: ~3,800 VLA source-zoomed images of 6 morphological classes (~600 per class): 1C-1P, 1C-2P, 1C-3P, 2C-2P, 2C-3P, 3C-3P



Context: ...

Question: **Which of these morphological classes of radio sources do you see in the image?**

1C-1P
1C-2C
1C-3P
2C-2P
2C-3P
3C-3P

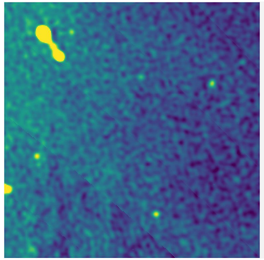


- ✓ *+10% wrt base model*
- ✓ *Caption data slightly degrade*
- ✓ *Marginal improvements with deeper fine-tuning*
- ✓ *Poor classification overall*
- ✓ *LORA fine-tuning not effective*

Fine-tuning Results (B3)

B3 – Extended Radio Galaxy Detection

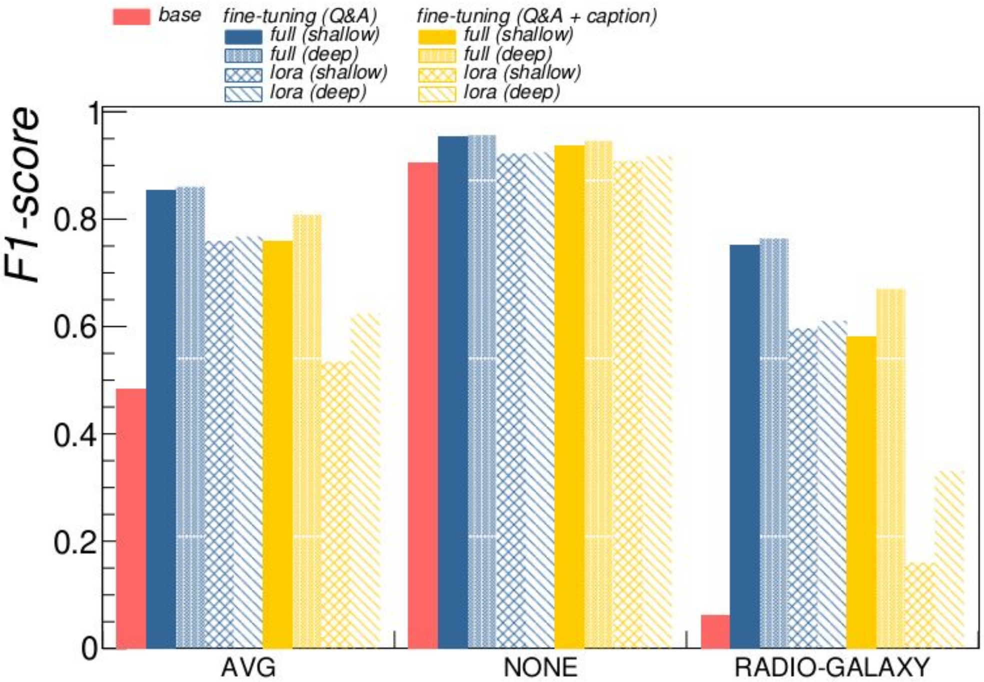
- ✓ Binary classification
- ✓ Dataset: ~5,700 ASKAP/MeerKAT images with extended radio galaxies (~18%)



Context: …

Question: **Do you see any likely radio galaxy with an extended morphology in the image?**

YES
NO

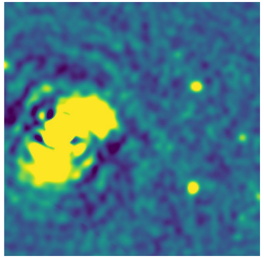


- ✓ *+30% wrt base model*
- ✓ *Caption data slightly degrade*
- ✓ *Negligible improvements with deeper fine-tuning*

Fine-tuning Results (B4)

B4 – Artefact Detection

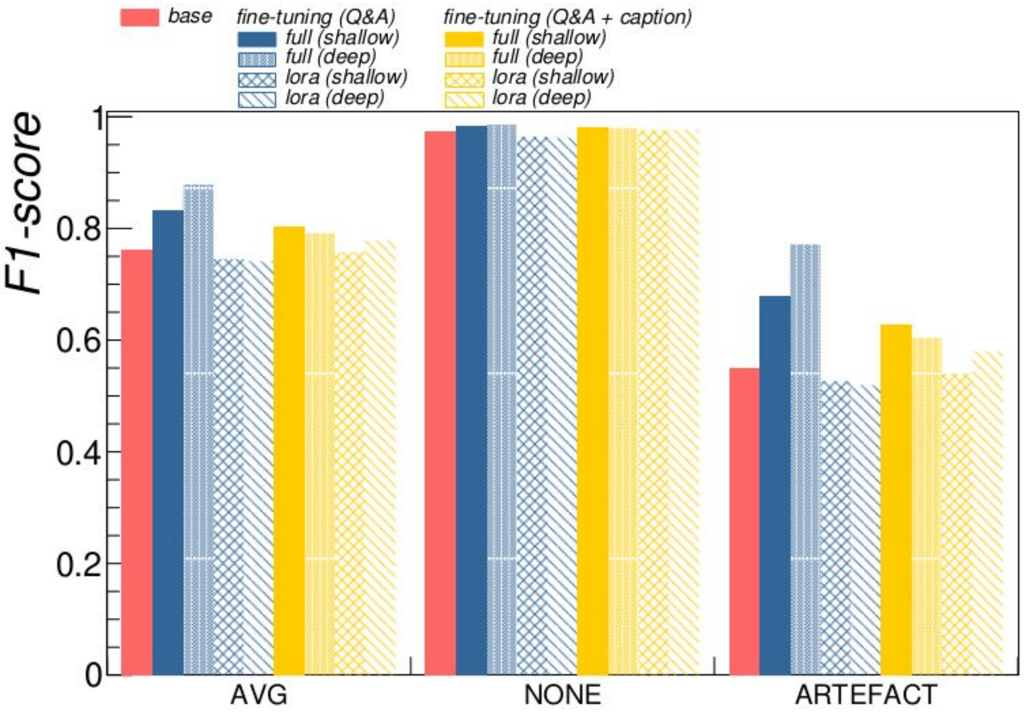
- ✓ Binary classification
- ✓ Dataset: ~5,700 ASKAP/MeerKAT images with imaging artefacts (~7%)



Context: …

Question: **Do you see any imaging artefact with a ring pattern around bright sources in the image?**

YES
NO

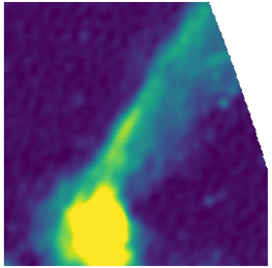


- ✓ *+10% wrt base model*
- ✓ *Caption data slightly degrade*
- ✓ *Marginal improvements with deeper fine-tuning*
- ✓ *LORA fine-tuning not effective*

Fine-tuning Results (B5)

B5 – Image peculiarity classification

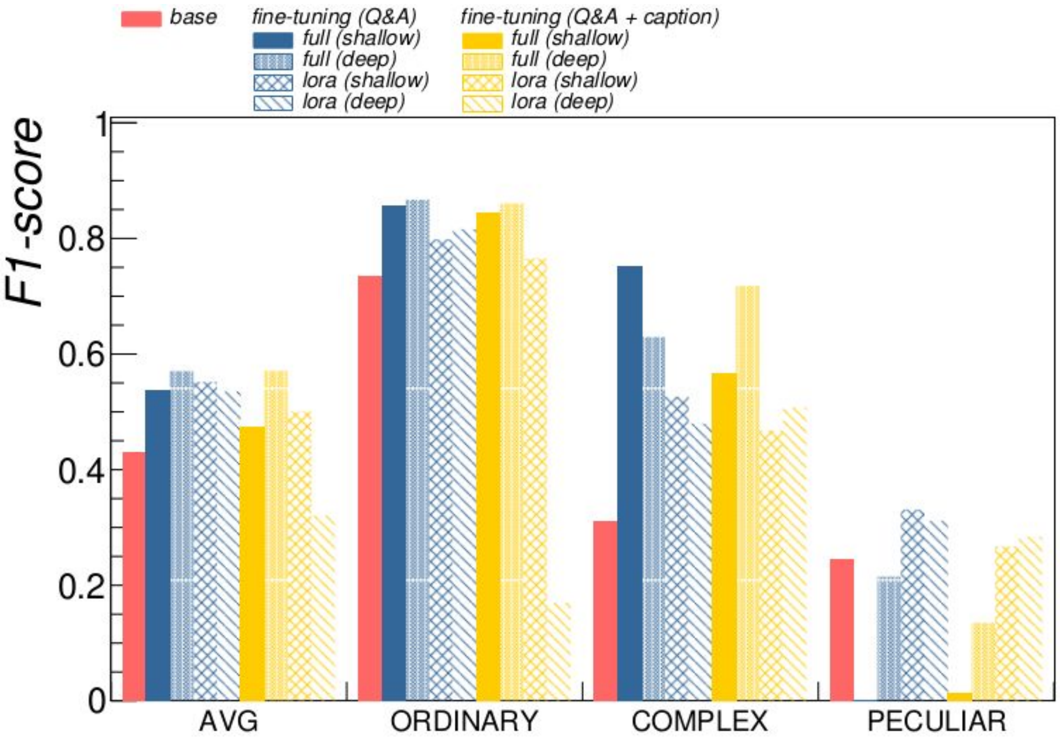
- ✓ Single-label multi-class classification
- ✓ Dataset: ~5,700 ASKAP/MeerKAT images labelled as ORDINARY (~63%), COMPLEX (~35%), PECULIAR (~3%)



Context: …

Question: **Can you identify which peculiarity class the presented image belongs to?**

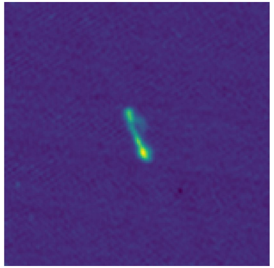
ORDINARY
COMPLEX
PECULIAR



- ✓ *+10% wrt base model*
- ✓ *Caption data slightly degrade*
- ✓ *Peculiar frames largely missed*

B6 – Radio Galaxy Morphology Classification

- ✓ Binary classification
- ✓ Dataset: ~832 VLA images centred on FR-I (~48%) and FR-II (~52%) radio galaxies

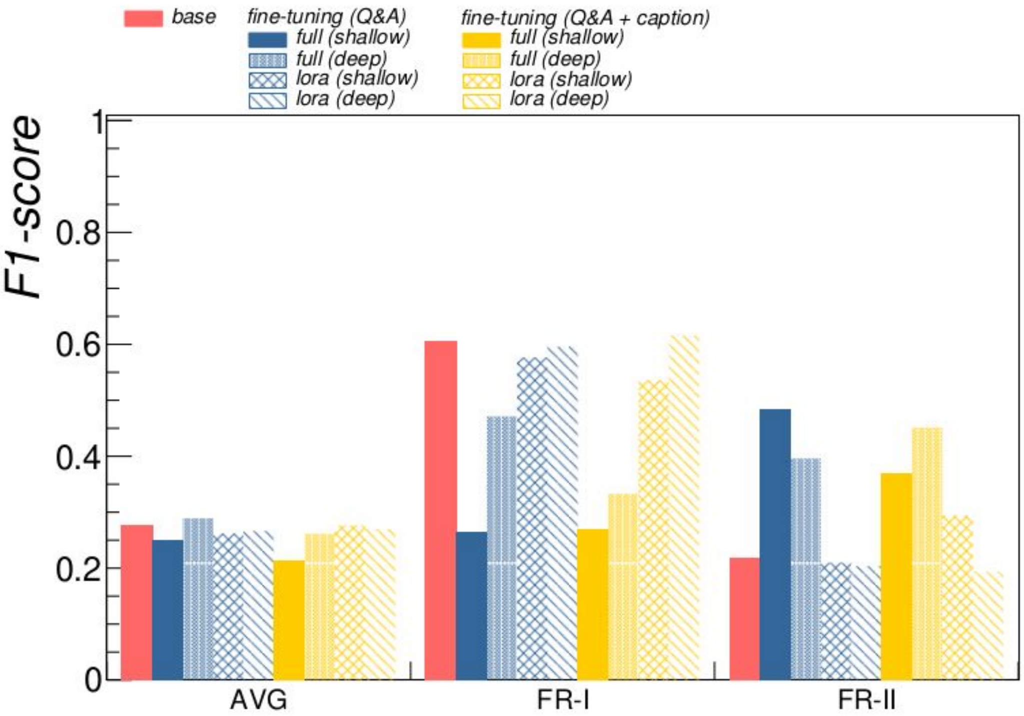


Context: …

Question: **Which of these morphological classes of radio galaxy do you see in the image?**

FR-I
FR-II

- ✓ *No improvements wrt base model*
- ✓ *Poor classification overall*

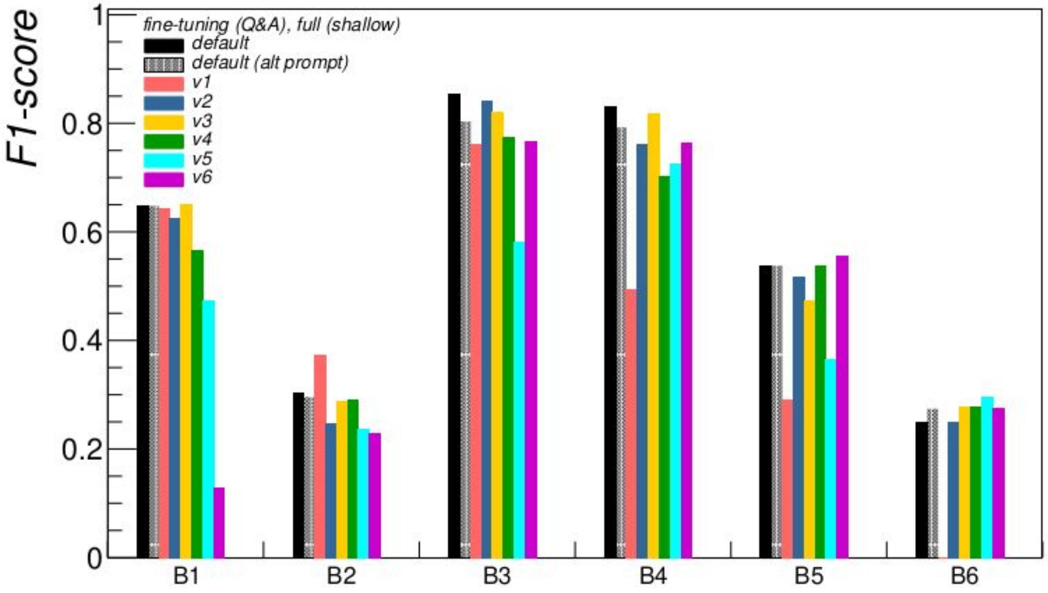


Fine-tuning Results (B1-B6)

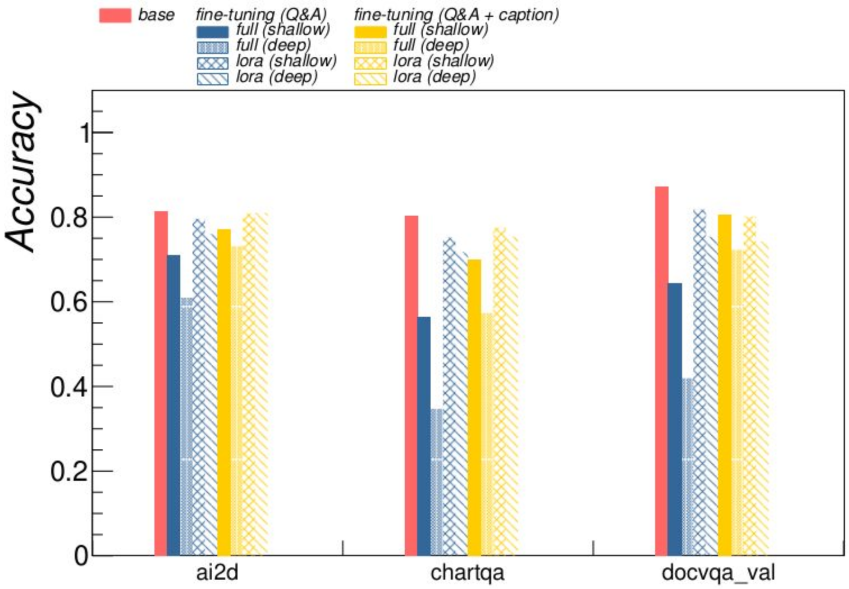
No significant improvement with alternative model hyperparameters

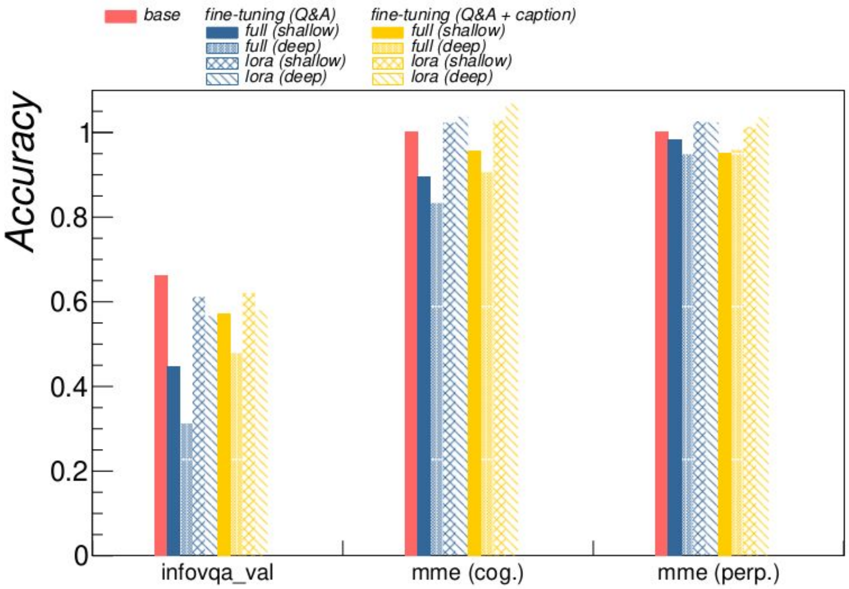
- ✓ *Alternative learning rates or learning schedules*
- ✓ *Alternative batch sizes*
- ✓ *Alternative LORA parameters*

Model variant	Parameters
v1	lr=5×10^{-5}
v2	lr=5×10^{-6}
v3	warmup_ratio=0.01
v4	cosine_with_min_lr (lr=5×10^{-6})
v5	effective_batch_size=256
v6	LORA rank=128, alpha=256



Fine-tuning Results (Multi-Modal Benchmarks)

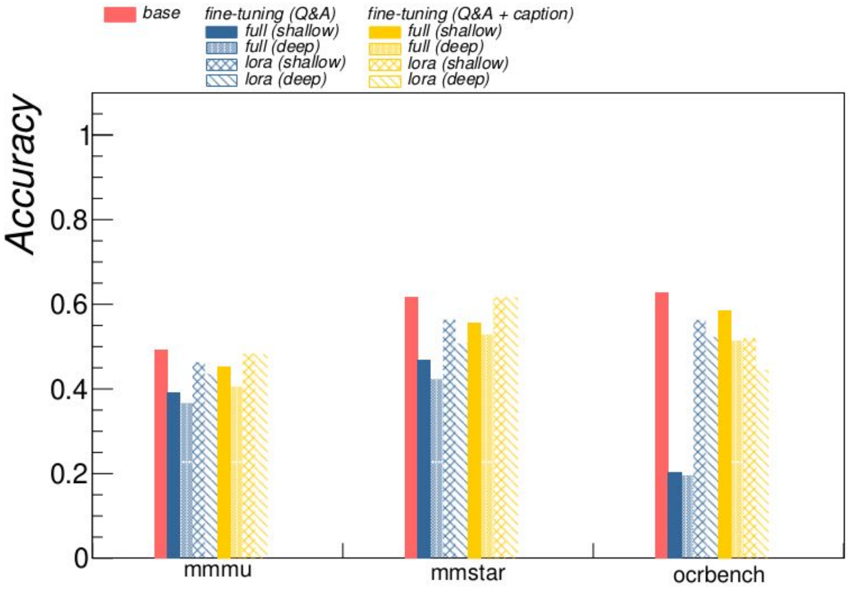


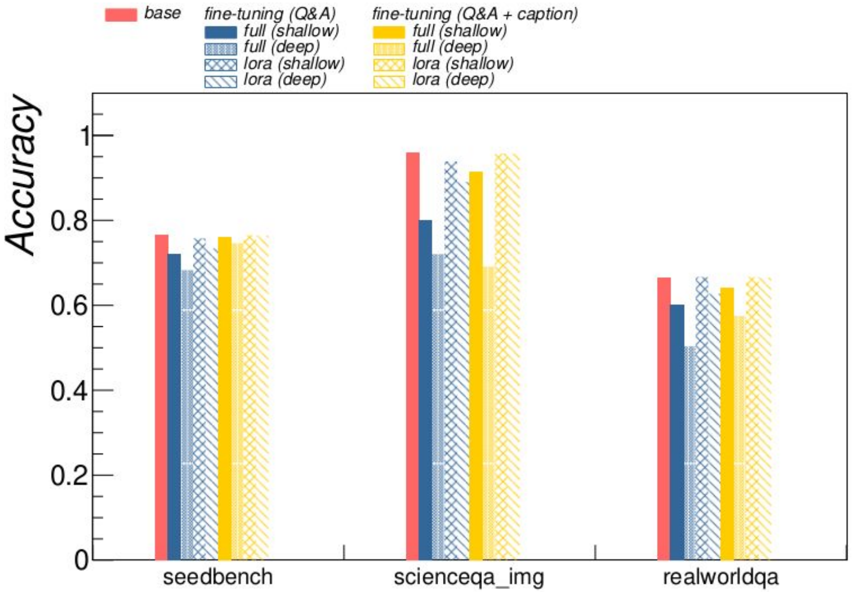


Catastrophic forgetting of previously learned multi-modal tasks

- ✓ *20-30% accuracy drop*
- ✓ *getting worse in deeper training runs*
- ✓ *caption data & LORA fine-tuning mitigating*

Fine-tuning Results (Multi-Modal Benchmarks)





Catastrophic forgetting of previously learned multi-modal tasks

- ✓ *20-30% accuracy drop*
- ✓ *getting worse in deeper training runs*
- ✓ *caption data & LORA fine-tuning mitigating*

Summary

VLMs offering an intuitive text-based interface for running radio source analysis

- **First attempt to evaluate VLMs on radio data**
  - ✓ *Full fine-tuning improves mostly B1 and B3 by >30% F1-score*
  - ✓ *LoRA gains are smaller but avoid catastrophic forgetting*
  - ✓ *Best results still behind vision-only task-customized models*
  - ✓ *Caption data improve instruction following & generalization to non-radio benchmarks*
  - ✓ *Minor gains from prompt or hyperparameter changes*

- **Challenges & limitations**
  - ✓ *Multimodal misalignment & training data quality/diversity are key bottlenecks to performance*
  - ✓ *Strategies to mitigate catastrophic forgetting on non-radio tasks*

Large Language Models (LLMs)

LLMs increasingly used for various tasks

- ✓ *Drafting & analyzing documents*
- ✓ *Generating & modifying software codes*
- ✓ *Inspecting/analyzing/formatting data*

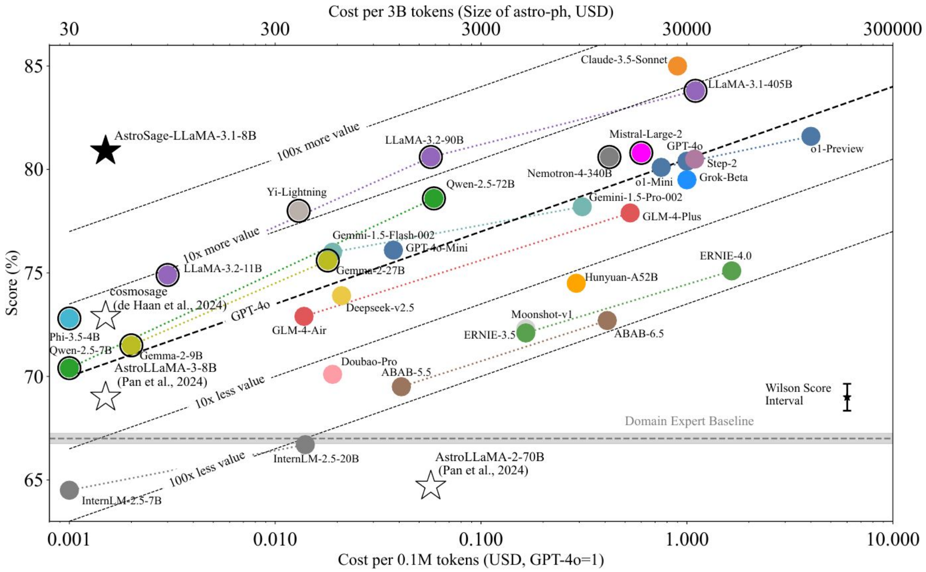
■ Why developing LLMs for astronomy?

- ✓ *Existing LLMs less effective on astro tasks or having prohibitive inference costs for large-scale analysis*
- ✓ *Data privacy reasons*

■ Ongoing projects focusing on small LLMs

Astrollama, cosmosage, astroLLM, astroBERT, AstroSage, ...

- ✓ *Trained on entire ArXiV astroph paper corpus (+ other resources) via instruction fine-tuning or RAG*
- ✓ *Catastrophic forgetting is a major challenge*

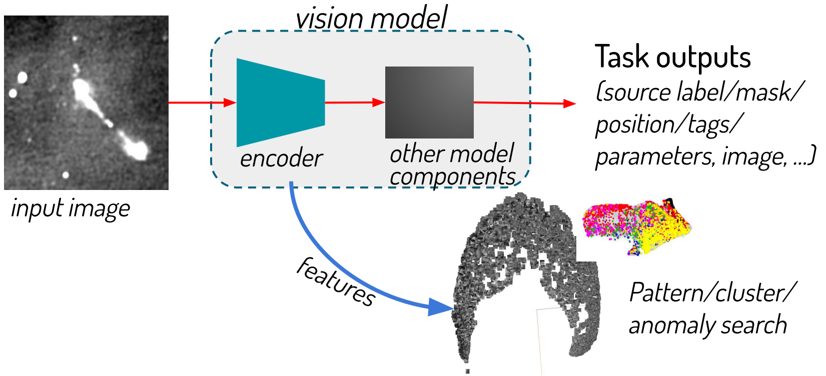


For more details:
AstroMLab papers (I, II, III)

Vision Models (VMs)

VMs used for various tasks on visual data

- ✓ *Source classification*
- ✓ *Source segmentation*
- ✓ *Anomaly search*
- ✓ *Data generation*
- ✓ *Similarity search*



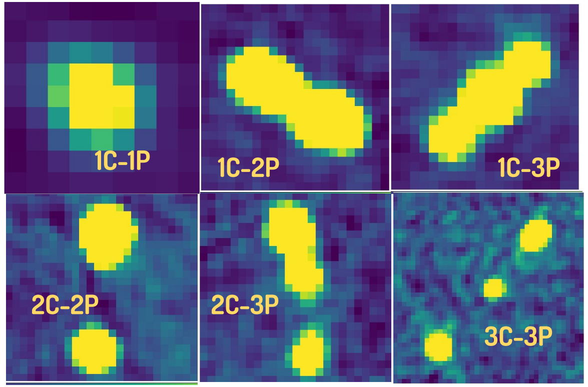
- **VM including vision encoders (CNN/ViT-based)**
  - ✓ *Extracting features from visual inputs*
  - ✓ *Pre-trained on large data samples (often in a self-supervised way)*
  - ✓ *Model & data representation learnt can be used for data inspection and analysis and re-used for different tasks*

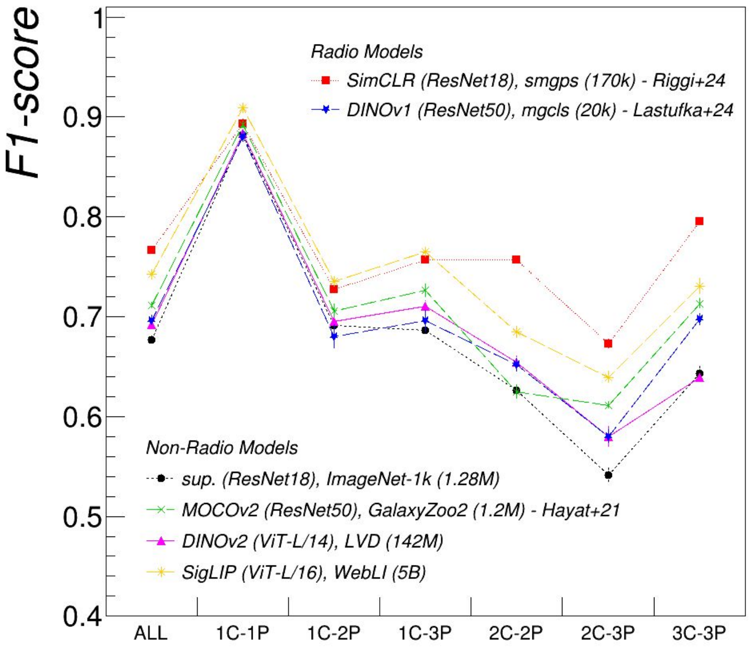
- **Why developing VMs for astronomy?**
  - ✓ *Extract high-quality features from astronomical data*
  - ✓ *Improve supervised analysis in the low-label regime*

Vision Models (VMs)

- **Various foundational models developed**
  - ✓ *Radio SSL models provide best performance on radio data/tasks*
  - ✓ *Marginal improvements in some tasks compared to ViT-pretrained models or fully-supervised models*
  - ✓ *Ongoing efforts: scale-up to millions of images, improve downstream task datasets*

Dataset: Radio Galaxy Zoo (RGZ) DR1
Images/class: 1000 (train), 600 (test)
Classes: 1C-1P, 1C-2P, 1C-3P, 2C-2P, 2C-3P, 3C-3P
Surveys: VLA FIRST
Method: LGBM classifier on SSL features





Evaluating encoders pre-trained on radio vs non-radio (optical, natural images) in a radio source morphology classification problem

For more details:
S. Riggi et al, https://arxiv.org/abs/2411.08519 (2024)
T. Cecconello et al, https://arxiv.org/abs/2411.14078 (2024)
E. Lastufka et al, https://arxiv.org/pdf/2409.11175 (2024)