JAISP

A Joint Rubin–*Euclid* Foundation Model for Detection and Astrometry

— and a path to *Roman* —

Shoubaneh Hemmati

with G. Helou, Y.-H. Lin, R. Akeson, C. D. Dowell, A. Faisst, T. Greene, D. Shupe

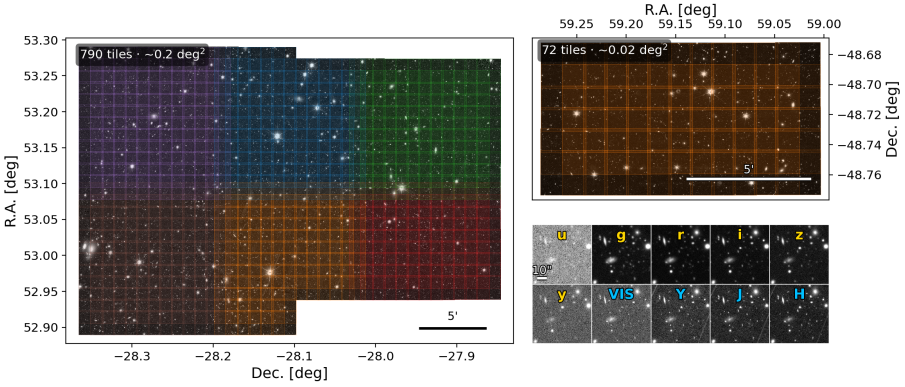
Caltech/IPAC, NASA JPL

Shifting Landscapes in Astrophysics: New Frontiers to Explore with *Roman* · July 2026

Why combine surveys at the pixel level?

- Cosmology is now limited by **systematics, not statistics**: whether the tensions between cosmological probes are real physics or measurement error, settling them requires control of systematics at a level no survey has yet demonstrated.
- Rubin/LSST, *Euclid*, and *Roman* will observe overlapping sky from $\sim$0.3 to 2 μm at sub-arcsecond resolution — combined at the pixel level, they can **cross-check one another**. But that means handling **different resolutions, PSFs, depths, and noise** at once, which no single-survey pipeline was built to do.
- AI will not replace decades of classical pipeline effort — certainly not all at once; reaching that maturity takes **comparable, problem-specific work**, and classical methods **inform the learning** along the way (here: they provide our training labels and every benchmark we test against).

Data: the earliest public release of each survey



- **Rubin DP1**: commissioning camera, *ugrizy*, 0.2″/px, ∼1″ seeing.
- **Euclid Q1**: VIS and near-infrared *YJH*, 0.1″/px.
- **Training field (ECDFS)**: 790 overlapping ten-band tiles over $\sim 0.2\,\mathrm{deg}^2$.
- **Out-of-distribution field (EDF-S)**: **set aside untouched** — the model never sees it — to test at the end whether everything transfers to new sky.

Train one model once — reuse it for every measurement

Self-supervised pretraining (masked-band reconstruction):

- mask one of the ten bands; reconstruct it from the other nine
- **no labels, no catalogs, no simulations**

Freeze the encoder, reuse the features:

- each task is a **lightweight head** on the same cached features
- detection & astrometry today; photometry, shapes, redshifts, … in preparation

Cost structure

Pretraining (once): $\sim$2.5 **GPU-hours**

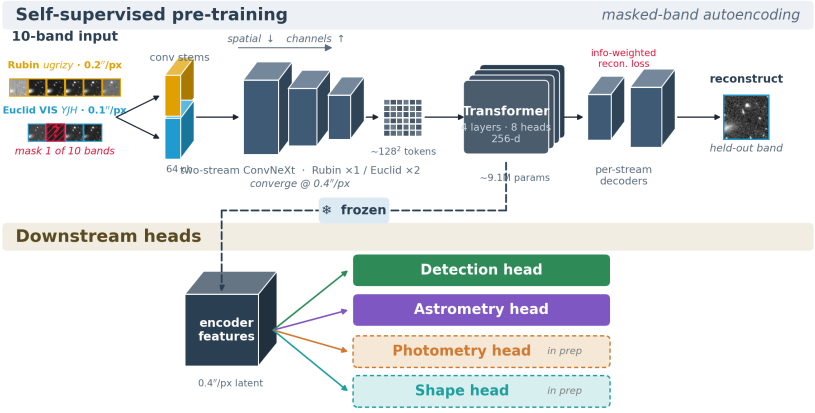
Model size: $\sim$9M parameters

New sky: **one encoding pass** (no training)

Each new task: a lightweight head

$\Rightarrow$ train once · encode each field once · measure many times

Architecture: each instrument at its native resolution



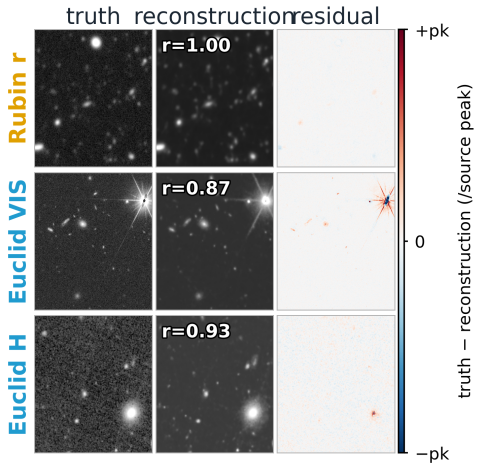
- Rubin stays at 0.2″/px, *Euclid* at 0.1″/px — **no resampling onto a shared grid**, so neither instrument's spatial information is thrown away.
- The branches merge on a common grid, where all ten bands exchange information.

We didn't start here: ten iterations to this design

v1–v2	**Contrastive patch matching** — make matching Rubin/*Euclid* cutouts look similar in feature space	failed: sky-dominated
v3	**Object-centric matching** — discover individual sources, pair them across instruments	abandoned: no precision
v4–v5	**Feature matching on full tiles (JEPA)** — make the two instruments' features agree at corresponding positions	beaten by a simple CNN
v6	**The turn: predict actual pixels** — hide one band, reconstruct its image; no shortcut that discards position	**paradigm shift**
v7–v8	**Mixed resolution** — one branch per instrument at its native pixel scale, merged at a common physical scale	works
v9–v10	**Refinements** — every band gets an equal vote in training; the loss focuses on real sources	**production**

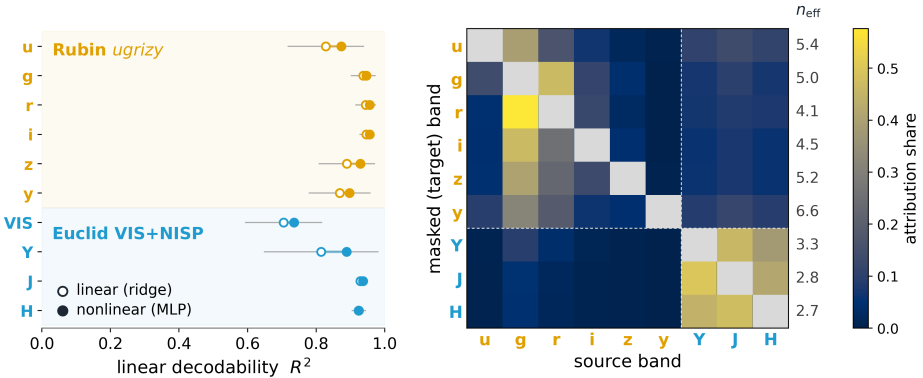
Features that look similar aren't features that know **where** *— contrastive, JEPA, masked autoencoding: we walked the road, and pixels won.*

Did it learn cross-instrument structure?



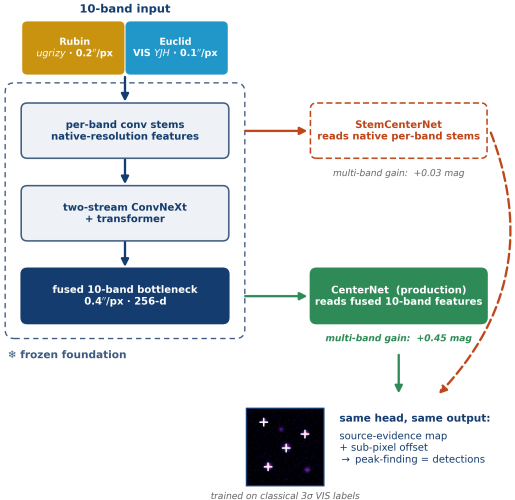
- Held-out tiles: withhold one band, predict it from the other nine.
- The loss is **noise-normalized (per-pixel variance) and weighted toward source pixels** — the model is graded on astronomical structure, not on reproducing blank sky.
- Every band is reconstructed faithfully (three examples shown) — residuals confined to bright cores and faint galaxy wings.
- A band from **one instrument predicted from bands of the other**: genuine cross-instrument correspondence, not within-instrument copying.

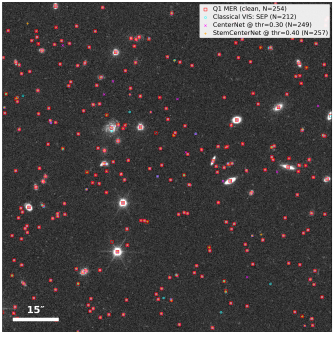
...and is it organized physically?



- **Left:** a simple linear fit reads each band's brightness straight back out of the features ($R^2 \approx 0.7$–0.95) — the information is there and easy to reach.
- **Right:** the model leans on bands the way real galaxy colors would predict: VIS is the sharp scaffold everything is built on; Rubin bands draw on their spectral neighbors; the three near-infrared bands mostly predict one another.

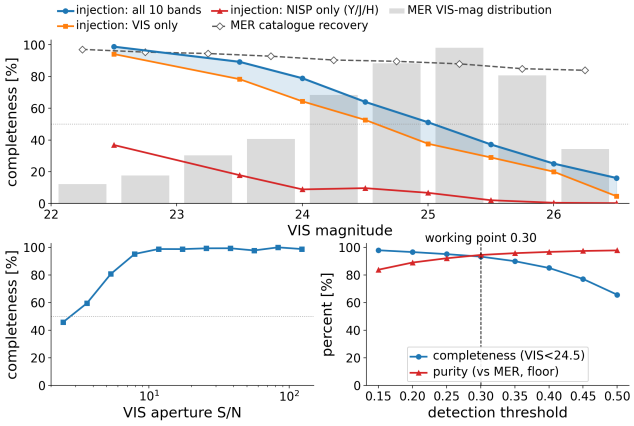
Downstream task 1 — Detection





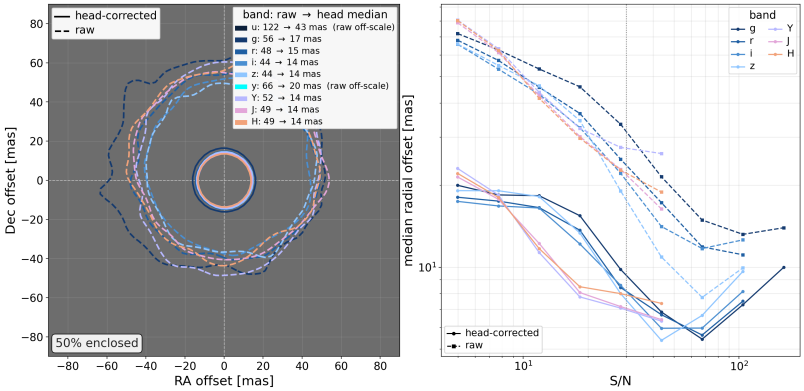
- Same head design, two read-out points: **fused ten-band features** vs. native per-band stems.
- Matches the *Euclid* Q1 MER catalog on held-out sky; **one detection per galaxy** where classical extraction fragments extended systems.

Detection: ten bands buy real depth



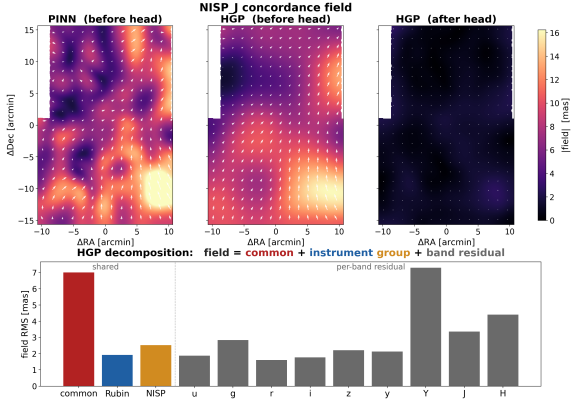
- Against the *Euclid* Q1 MER catalog on held-out sky: **93% completeness, 95% purity**.
- Injection test — real sources, dimmed to a chosen magnitude, hidden in the images: ten-band detection is $+0.45$ **mag deeper** than the same detector on VIS alone.
- On the deeper *Euclid* DR1 imaging, the same injection test reaches **VIS** ≈ 26.0.

Downstream task 2 — Astrometry: carrying VIS precision to every band



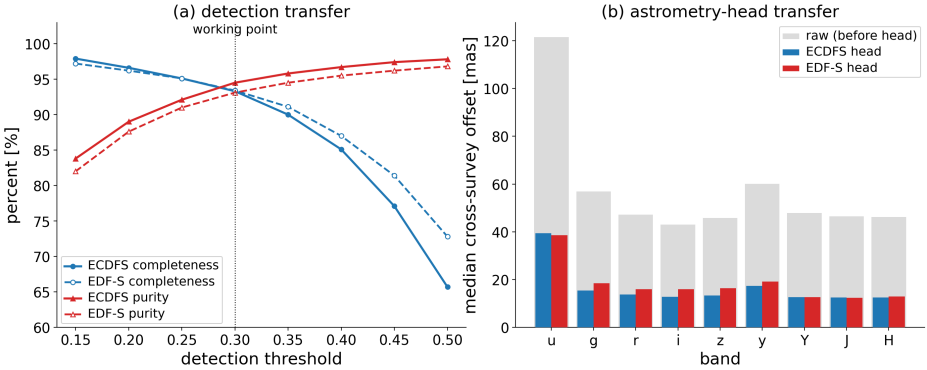
- The same sources measured in both surveys disagree by $\sim$50 mas — mostly photon noise in centroiding faint galaxies. A small correction network reads the model's features at each source and shrinks this to 14–17 mas **in every band**.
- Ground-truth check: inject sources at **known positions** and re-measure. At $\mathrm{S/N} = 5$ the head recovers them to 19 mas; classical centroiding scatters by $\sim$50 mas. That is within $\sim$2 mas of the VIS anchor itself — **VIS-quality positions in every band**.

The concordance field: how well do two Gaia-tied surveys actually agree?



- Beneath the per-source scatter sits a coherent 9–10 mas **pattern** between the two surveys' astrometric solutions — even though both are tied to Gaia; ∼7 mas is shared by **all ten bands**. Two entirely different fitting methods recover the same pattern, and the per-source correction absorbs it (→1.5 mas).
- A few-mas map of **how well two independently calibrated surveys actually agree** — quality assurance no single pipeline can produce.

Does any of this transfer? A second deep field, nothing retrained



- The locked model and its measurement networks, applied to EDF-S: **nothing retrained**.
- Detection: 93.4%/93.1% completeness/purity vs. 93.3%/94.5% on the training field. Positions land on the same $\sim$12–18 mas floor, within 1–3 mas.
- Scaling to more area is a matter of **applying the stack, not rebuilding it**.

Roman: the natural third stream

Why the model is ready:

- one encoder branch per instrument *is* the design — adding *Roman* means **adding a third branch**, not redesigning the model
- training needs only **overlapping public images** — no labels, no simulations — and hours of GPU time, once WFI imaging overlaps Rubin/*Euclid* sky

What Roman adds:

- **sharp, deep near-infrared imaging at $0.11''$/px** — the infrared counterpart of the sharp VIS reference, where today we must rely on resampled NISP
- a wide survey of thousands of deg^2 **inside the LSST footprint**, plus deep time-domain fields

What Roman gains:

- detection that uses all surveys at once — for Rubin+*Euclid* that was worth $+0.45$ mag
- *Roman*-quality positions carried into the ground-based bands
- **three-way agreement maps**: Rubin–*Euclid*–*Roman* astrometry compared at the few-mas level — quality assurance no single pipeline can produce
- cross-instrument alignment solved *inside* the model instead of inherited from preprocessing (the $\sim 3\,\mathrm{mas}$ NISP resampling error we found is exactly that)

Takeaways

- A small (∼9M-parameter) foundation model, trained **without any labels** on public Rubin + *Euclid* images in ∼2.5 GPU-hours, learns a shared ten-band representation precise enough to measure with.
- **Detection**: matches the *Euclid* catalog (93%/95%), and ten-band fusion pushes 0.45 **mag deeper** than VIS alone.
- **Astrometry**: VIS-quality positions in every band — ∼50 → 14–17 mas, verified at 19 mas on hidden test sources — and a few-mas **map of the disagreement** between the two surveys' Gaia-tied solutions.
- The whole stack **transfers unchanged** to a second deep field. All of this on the *earliest* data both surveys have released — a floor, not a ceiling.
- *Roman* slots in as a **third encoder stream**: sharp NIR anchor, three-way cross-survey calibration, and one representation reused across tasks, fields — and instruments.