

AstrID – Week 2 Plan - Meeting 2

Project Summary

Focus this week: agree on which surveys to use for a small pilot, define simple preprocessing defaults, and confirm plain-language success measures before scaling up.

Data Sources (Shortlist)

- MAST (Pan-STARRS1 references; Hubble for deep context)
- SkyView (DSS2, SDSS, GALEX, 2MASS, WISE) with a standard cutout size and pixel scale; set a fallback order
- VizieR catalogs (Gaia DR3, SDSS DR16, Pan-STARRS DR2, AllWISE) for cross-matching
- ZTF public data (small image stamps) if accessible for time-variable examples

Near-Term Milestones

1. Ingestion pilot: fetch a small set of cutouts from SkyView and Pan-STARRS via MAST
2. Preprocessing defaults: alignment, background subtraction, consistent scaling
3. Image differencing on the pilot set; note workable settings
4. Baselines: a U-Net style model plus a simple traditional method for comparison
5. Experiment tracking: name runs clearly and save parameters/artifacts
6. Orchestration: a minimal flow (ingest → preprocess → subtract → detect)

Evaluation (Metrics)

- Precision: of the flagged candidates, how many are real?
- Recall: of the real events present, how many did we find?
- Localization match: how closely the highlighted region lines up with the true spot (percent overlap)
- False alarms per image: average spurious detections per image
- Processing time per image: rough speed on current hardware

Risks & Open Questions

- Limited ground truth; balance synthetic injections with a few known events
- Cross-survey differences (filters, depth, image sharpness)
- Access and rate limits; may need caching/staging
- Path to larger scale (rough compute/storage estimates)

Requests for Advisor

- Confirm survey and catalog priorities for the pilot
- Feedback on the metrics above and acceptable target ranges
- Science focus guidance (e.g., supernovae vs. cataclysmic variables; preferred sky regions)
- Any help with data access or broker introductions