



Universidad La Salle

Maestría en Inteligencia Artificial

Propuesta: Benchmarking reproducible de métodos de denoising en imágenes simuladas LSST DC2

Propuesta de Proyecto

Jheison Martinez Bolivar

jhmartinez07@unisalle.edu.co

Fecha de entrega: 16 de junio de 2026

Resumen

Este borrador presenta el estado actual del proyecto “Astro Denoise Benchmark” y documenta el alcance hasta el estado del arte y los datos preliminares. Se describen el contexto científico, la motivación inspiracional para una ciencia más abierta y transparente, el planteamiento del problema y los datasets DC2 que se usarán para evaluación. Además, se documenta el procedimiento reproducible de selección bibliográfica potenciado por aprendizaje activo y la integración inicial de referencias clave que definen el panorama actual del denoising astronómico.

El objetivo del documento es ofrecer una base técnicamente sólida y reproducible para comparar métodos de denoising (enfocándose en BM3D y U-Net) mediante métricas orientadas a la ciencia —completitud y pureza— en imágenes simuladas del Vera Rubin Observatory (DC2). Este informe no contiene aún resultados experimentales finales: presenta el alcance, los datos disponibles y la metodología propuesta para las próximas etapas del benchmark.

Se han versionado los scripts, los datos y las fichas bibliográficas para facilitar la auditoría y la repetición del estudio por parte de la comunidad científica.

Palabras clave: denoising; imágenes astronómicas; DC2; BM3D; U-Net; completitud; pureza; Vera Rubin Observatory; reproducibilidad.

Keywords: denoising; astronomical images; DC2; BM3D; U-Net; completeness; purity; Vera Rubin Observatory; reproducibility.

Índice general

Resumen	1
Índice de figuras	4
1. Contexto y Motivación	5
2. Planteamiento del problema	6
2.1. Pregunta de investigación	6
2.2. Objetivos	6
2.2.1. Objetivo general	6
2.2.2. Objetivos específicos	6
2.3. Árbol de problemas	6
3. Estado del arte	8
3.1. Panorama general	8
3.2. Proceso de cribado y síntesis (PRISMA + RAG)	8
3.2.1. Criterios de inclusión y exclusión	8
3.2.2. Estadísticas del corpus final	9
3.3. Clasificación técnica por enfoques	11
3.3.1. Métodos clásicos y variacionales	11
3.3.2. CNN/U-Net supervisadas	11
3.3.3. Auto-supervisión y entrenamiento sin pares limpios	11
3.3.4. Modelos generativos y difusión	12
3.3.5. Métodos física-informados e híbridos	12
4. Justificación	13
4.1. Brecha identificada en el marco evaluativo	13
4.1.1. Evidencia textual de la necesidad	13
4.1.2. Implicación para esta propuesta	13
4.2. Métricas recomendadas para benchmark en LSST/DC2	14
4.3. Descripción de datos y materiales	14
4.3.1. Componentes del dataset	14
4.3.2. Origen y tamaños reales	14
4.3.3. Flujo del pipeline de procesamiento	15
5. Metodología propuesta	16
5.1. Resumen	16
5.2. Flujo experimental	16
5.3. Datos y preprocesado	16
5.4. Métodos a comparar	16
5.5. Protocolo de evaluación científica	17
5.6. Reproducibilidad y registros	17
5.7. Plan de trabajo y entregables	18

6. Conclusiones	19
6.1. Conclusiones sobre el Estado del Arte	19
6.2. Brechas Detectadas	19
A. Hoja de Ruta del Proyecto (ROADMAP.md)	20
B. Especificaciones de Diseño (OpenSpec YAML)	23
B.1. Orquestador: bench-orquestador-2026.yaml	23
B.2. Baseline BM3D: bm3d-baseline-2026.yaml	24
B.3. Módulo de Métricas: metrics-modulo-2026.yaml	25
B.4. Baseline U-Net: unet-baseline-2026.yaml	25
Referencias	27

Índice de figuras

2.1. Árbol de problemas enriquecido: factores que impiden la validación objetiva del denoising astronómico.	7
3.1. Diagrama PRISMA del cribado (corpus astronómico de denoising) . .	9
3.2. Distribución de artículos seleccionados por año (n=50).	10
3.3. Composición del corpus final por categoría.	10
3.4. Frecuencia de métricas reportadas en la literatura seleccionada (codi- ficación cualitativa guiada por RAG + lectura manual).	11
4.1. Secuencia de procesamiento desde datos crudos hasta métricas	15
5.1. Secuencia del flujo experimental de benchmarking	17

Capítulo 1

Contexto y Motivación

El proyecto “Astro Denoise” nace de la necesidad de evaluar, con rigor científico, cómo las técnicas modernas de restauración de imágenes afectan la capacidad de las encuestas astronómicas para descubrir y caracterizar objetos tenues. El Vera Rubin Observatory (LSST) representa una revolución en volumen y profundidad de datos: detectar señales débiles en imágenes coadded es crítico para objetivos como cosmología de precisión, mapeo de la Vía Láctea, descubrimiento de objetos transitorios y estudios del Sistema Solar. Mejorar la relación señal–ruido sin introducir sesgos es, por tanto, una prioridad científica.

Este proyecto integra dos dimensiones esenciales: (1) rigor metodológico —protocolos reproducibles, métricas científicas y control de semillas/random states—y (2) relevancia científica —medir el impacto del denoising sobre la capacidad de detección y la calidad fotométrica en catálogos que alimentan análisis astrofísicos. En otras palabras, no se persigue la máxima mejora perceptual, sino la mejora cuantificable y verificable en métricas que importan para la ciencia.

La elección del conjunto de datos (DC2 dr6–v4) y la atención a reproducibilidad responden a las recomendaciones de la comunidad y a los objetivos científicos del Rubin Observatory (Rubin Observatory, 2024), que enfatizan descubrimientos estadísticos, trazabilidad de análisis y estimación rigurosa de incertidumbres instrumentales.

Motivación inspiracional: este trabajo pretende acelerar descubrimientos al reducir las barreras técnicas que impiden detectar y medir fuentes débiles. Al construir un benchmark modular, reproducible y abierto, se facilita la comparación objetiva entre métodos clásicos y de IA, promoviendo buenas prácticas de ciencia de datos en astronomía. El impacto potencial incluye: mejores catálogos para cosmología y astrofísica, pipelines reutilizables para futuras encuestas y un marco de evaluación que favorece transparencia, reproducibilidad y colaboración abierta.

Se resalta la importancia de documentar cada decisión —desde criterios de selección bibliográfica hasta preprocesado y parámetros del detector—para que los resultados sean auditables y replicables por cualquier equipo de la comunidad.

Capítulo 2

Planteamiento del problema

2.1 Pregunta de investigación

¿Cómo afectan las técnicas modernas de restauración de imágenes a la capacidad de las observaciones astronómicas para descubrir y caracterizar objetos tenues, y cómo pueden evaluarse de manera reproducible y científicamente válida?

2.2 Objetivos

2.2.1 Objetivo general

Desarrollar una plataforma de benchmarking reproducible que compare métodos de denoising en imágenes del Vera Rubin Observatory, evaluando la recuperación de fuentes débiles mediante métricas de completitud y pureza.

2.2.2 Objetivos específicos

1. Implementar un ecosistema modular y extensible con protocolo CLI estándar.
2. Establecer métricas confiables (matching, completitud/pureza vs. truth catalog).
3. Crear un benchmark reproducible y versionado con DESC DC2.
4. Proporcionar herramientas de medición de la capacidad computacional.

2.3 Árbol de problemas

A continuación se presenta el análisis causal enriquecido, con una distribución simétrica de factores:

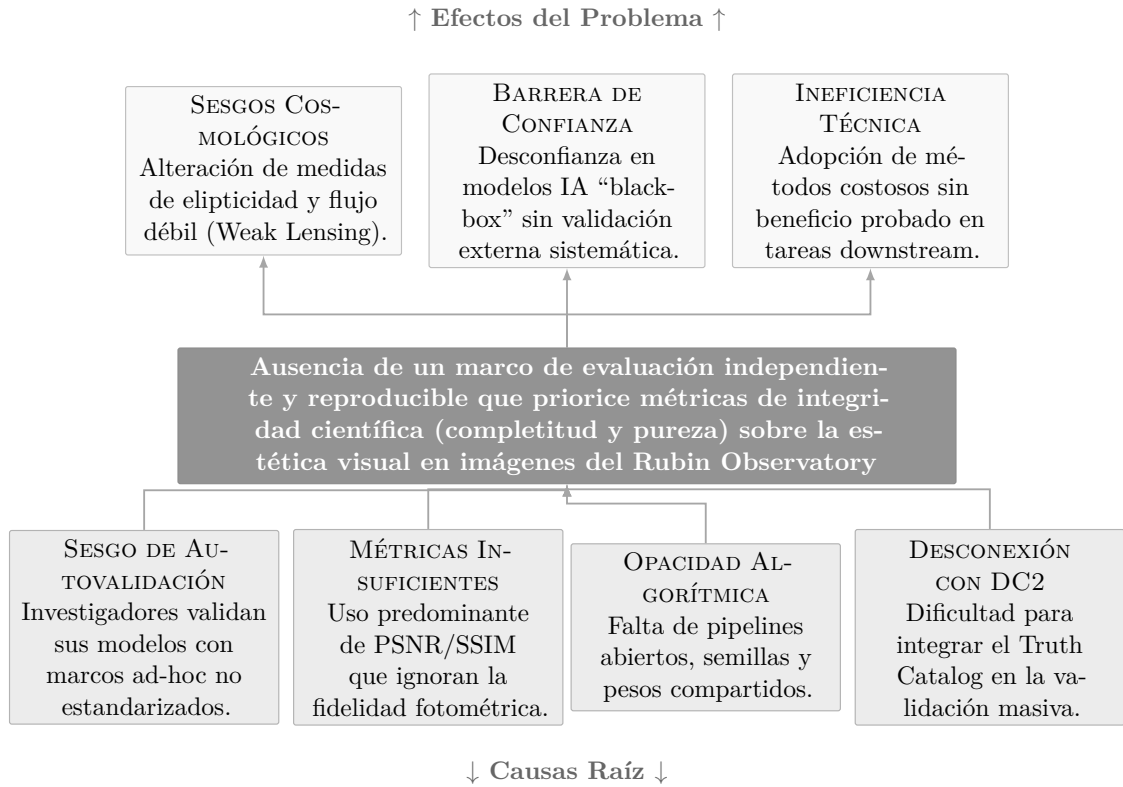


Figura 2.1: Árbol de problemas enriquecido: factores que impiden la validación objetiva del denoising astronómico.

Capítulo 3

Estado del arte

3.1 Panorama general

El estado del arte en denoising astronómico muestra un desplazamiento claro desde métodos clásicos (wavelets, variacionales, priors explícitos) hacia arquitecturas profundas con modelado probabilístico (U-Net, Transformers y difusión), pero con una tensión metodológica persistente: mejorar la calidad visual no garantiza preservar la validez científica de la inferencia downstream (Elad, Kwar, y cols., 2023; Vojtekova, Lieu, y cols., 2020; Shirasaki, Moriwaki, y cols., 2021; Akhaury, Jablonka, y cols., 2025; Guo, Zhang, y cols., 2026).

En este trabajo, el foco no es presentar otro denoiser aislado, sino consolidar una base comparativa reproducible para LSST/DC2 donde la evaluación se centre en impacto científico medible (detección, fotometría, forma, sesgo) y no solo en fidelidad perceptual.

3.2 Proceso de cribado y síntesis (PRISMA + RAG)

La revisión se reconstruyó de forma transparente sobre un corpus inicial de 304 PDFs, combinando aprendizaje activo (ASReview), re-ranqueo semántico con RAG y verificación manual de metadatos en PDF. El uso de RAG se limitó a priorización y trazabilidad semántica de relaciones complejas (métricas, tareas, sesgos), mientras que la decisión final de inclusión/exclusión fue auditada manualmente.

3.2.1 Criterios de inclusión y exclusión

Inclusión: trabajos con evaluación de denoising/deconvolución sobre imágenes astronómicas (óptico o dominios cercanos), con reporte de métricas cuantitativas y/o impacto en tareas científicas (detección, fotometría, morfología, weak lensing).

Exclusión: estudios fuera del dominio astronómico y trabajos sin validación empírica relevante.

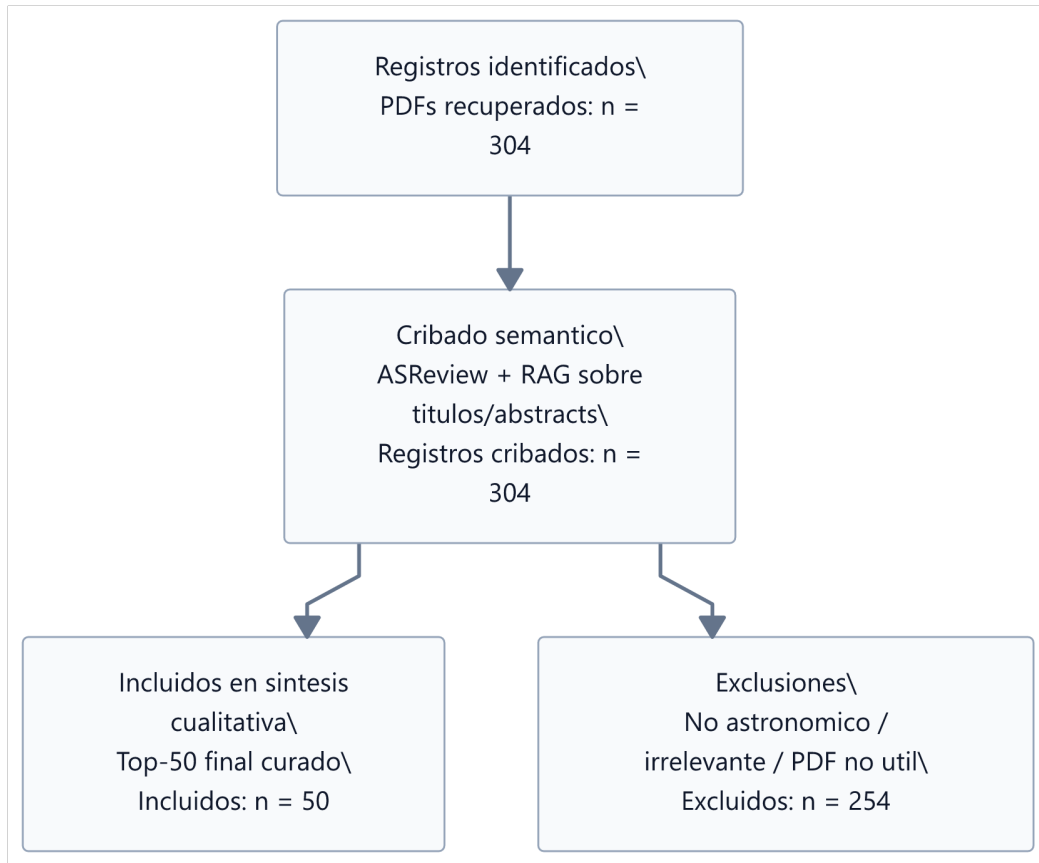


Figura 3.1: Diagrama PRISMA del cribado (corpus astronómico de denoising)

3.2.2 Estadísticas del corpus final

Cuadro 3.1: Composición del corpus final (n=50)

Grupo	Conteo	Proporción
Core astronómico (denoising/deconvolución)	31	62 %
Reviews/surveys	8	16 %
Métodos de soporte (difusión, Poisson, PnP, etc.)	11	22 %

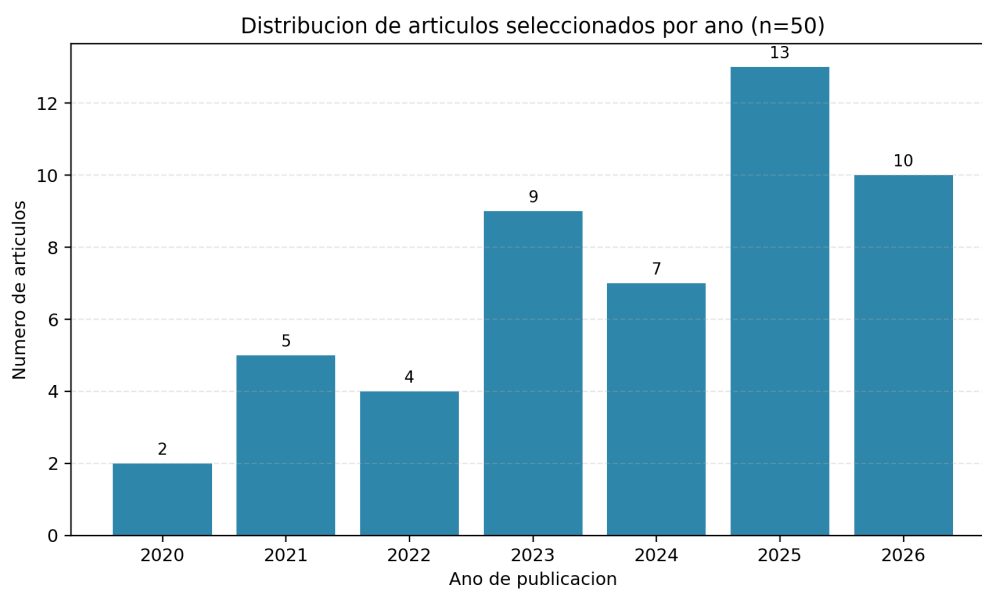


Figura 3.2: Distribución de artículos seleccionados por año ($n=50$).

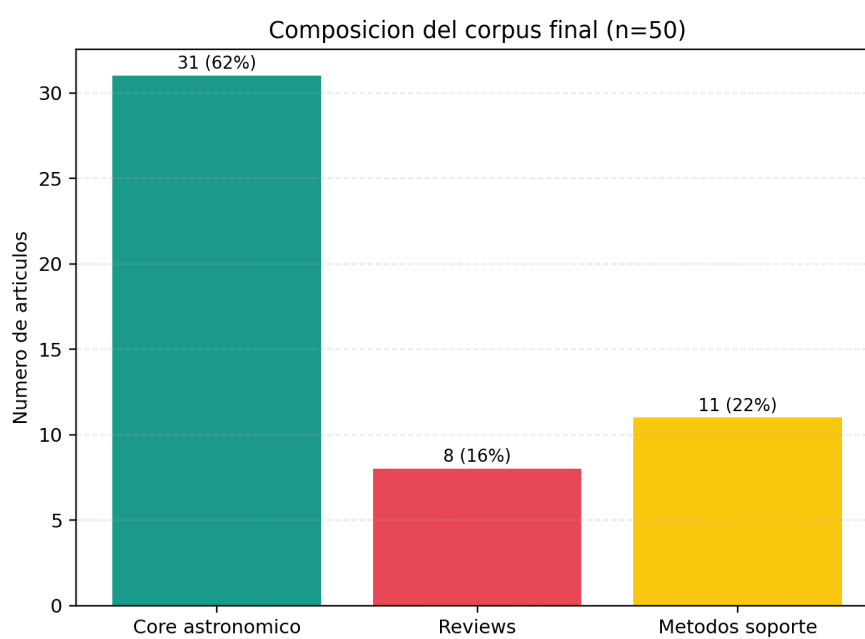


Figura 3.3: Composición del corpus final por categoría.

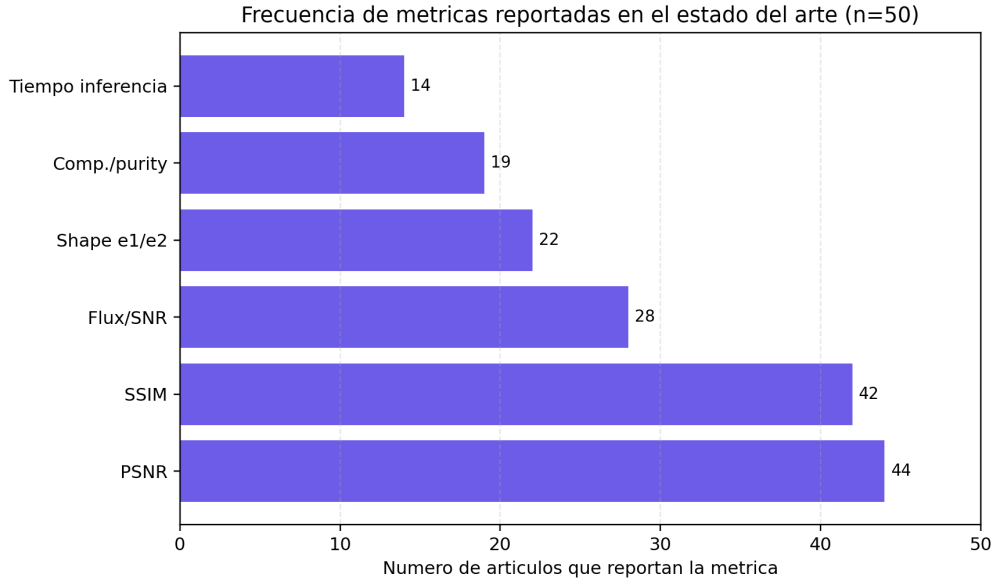


Figura 3.4: Frecuencia de métricas reportadas en la literatura seleccionada (codificación cualitativa guiada por RAG + lectura manual).

3.3 Clasificación técnica por enfoques

3.3.1 Métodos clásicos y variacionales

Los enfoques clásicos siguen vigentes como referencia de interpretabilidad y estabilidad, especialmente en escenarios donde el modelo de ruido es conocido o parcialmente controlable (Roscani, Tozza, y cols., 2020; Makarkin y Bratashov, 2021; Bui, Lou, y cols., 2023).

3.3.2 CNN/U-Net supervisadas

Las U-Net y variantes continúan siendo un estándar de alto rendimiento para restauración de galaxias y deblending, con mejoras concretas en recuperación de señal útil para tareas de catálogo (Vojtekova, Lieu, y cols., 2020; Hemmati, Huff, y cols., 2022; Park, Jo, y cols., 2024; Miao, Tu, y cols., 2024).

3.3.3 Auto-supervisión y entrenamiento sin pares limpios

Los métodos auto-supervisados reducen la dependencia de datos “clean” perfectos, una limitación crítica en astronomía observacional (Liu, Quan, y cols., 2025; Guo, Zhang, y cols., 2026; Sukurdeepa, Navarro, y cols., 2025).

3.3.4 Modelos generativos y difusión

La familia de difusión gana tracción por su capacidad de modelar incertidumbre y mejorar escenarios de fondo complejo, incluyendo weak lensing y remoción de ruido de fondo (Scognamiglio, Lee, y cols., 2025; Aoyama, Osato, y cols., 2025; *BGRem: A background noise remover for astronomical images based on a diffusion model*, s.f.; Schanz, List, y cols., 2023; Remy, Lanusse, y cols., 2023).

3.3.5 Métodos física-informados e híbridos

La integración de conocimiento instrumental (PSF, ruido CCD, fotometría multi-banda) en arquitecturas profundas se consolida como tendencia de mayor valor para pipelines científicos robustos (*Neural network based point spread function deconvolution for astronomical applications*, s.f.; *Point Spread Function Deconvolution Using a Convolutional Autoencoder for Astronomical Applications*, s.f.; Akhaury, Starck, y cols., 2022; Akhaury, Jablonka, y cols., 2025; Liu, Ge, y cols., 2026).

Capítulo 4

Justificación

4.1 Brecha identificada en el marco evaluativo

La principal brecha identificada no es la falta de modelos de denoising, sino la ausencia de un marco evaluativo unificado y orientado a validez científica. La literatura converge en que la mejora visual no basta y que los costos de validación ya compiten con el esfuerzo de desarrollo de modelos.

4.1.1 Evidencia textual de la necesidad

- Akhaury et al. reportan explícitamente una necesidad metodológica transversal: «there is an increasing need to develop fast and accurate deconvolution algorithms that generalize well to different images» (Akhaury, Starck, y cols., 2022).
- Guo et al. muestran el costo operacional de ganar profundidad observacional y justifican alternativas algorítmicas: «improves detection limits by 1.0 magnitude at 90 % completeness and purity» y, para un incremento similar por exposición, se requeriría escalar de días a cerca de un mes de observación (Guo, Zhang, y cols., 2026).
- En la revisión de Elad et al. se explicita el problema de evaluación multiojetivo: «potential solvers need to tradeoff distortion metrics (e.g. MSE) versus perception measures» (Elad, Kwar, y cols., 2023).

4.1.2 Implicación para esta propuesta

Este proyecto se justifica porque desplaza el centro de gravedad desde *entrenar otro modelo* hacia *establecer una validación reproducible y con control de sesgo*. En otras palabras, se responde al problema que la comunidad está enfrentando hoy: invertir tiempo de investigación en diseñar marcos de evaluación comparables, auditables y ligados a impacto científico real.

4.2 Métricas recomendadas para benchmark en LSST/DC2

Como resultado de la revisión del estado del arte, se propone un marco de validación en tres capas:

1. **Fidelidad de reconstrucción:** PSNR/SSIM y error de reconstrucción local.
2. **Validez científica:** completitud-pureza por magnitud, error fotométrico, sesgo en forma (e_1/e_2), impacto en detección de fuentes débiles.
3. **Costo operativo:** tiempo de inferencia, estabilidad por dominio y sensibilidad a mismatch de ruido instrumental.

Esta estructura permite comparar métodos heterogéneos (clásicos, CNN, difusión, física-informados) con un mismo criterio de calidad científica y no únicamente perceptual.

4.3 Descripción de datos y materiales

4.3.1 Componentes del dataset

El dataset de trabajo consiste en dos componentes del release DESC DC2 (dr6-v4) del Vera Rubin Observatory. El primero es `coadd-t3828-t3829-small`, que contiene imágenes coadded de dos tracts del cielo simulado. Una imagen coadded es el resultado de apilar múltiples exposiciones, lo que cancela parte del ruido aleatorio pero deja un ruido residual bajo el cual las fuentes débiles siguen siendo difíciles de detectar.

El segundo componente es `truth_match`, el catálogo de verdad de la simulación, que contiene la posición y el flujo exacto de cada fuente simulada. Este catálogo se utiliza exclusivamente en la etapa de evaluación para calcular la completitud y pureza de los algoritmos de denoising.

4.3.2 Origen y tamaños reales

A diferencia de las versiones comprimidas, este proyecto trabaja directamente con los archivos en formato FITS para preservar la precisión científica total. Los pesos en disco son los siguientes:

- **DESC DC2 Truth Match (`truth_match`):** 52 GB (Formato FITS nativo).
- **DESC DC2 Coadded Images (`coadd-t3828-t3829-small`):** 4.8 GB (Imágenes científicas FITS).

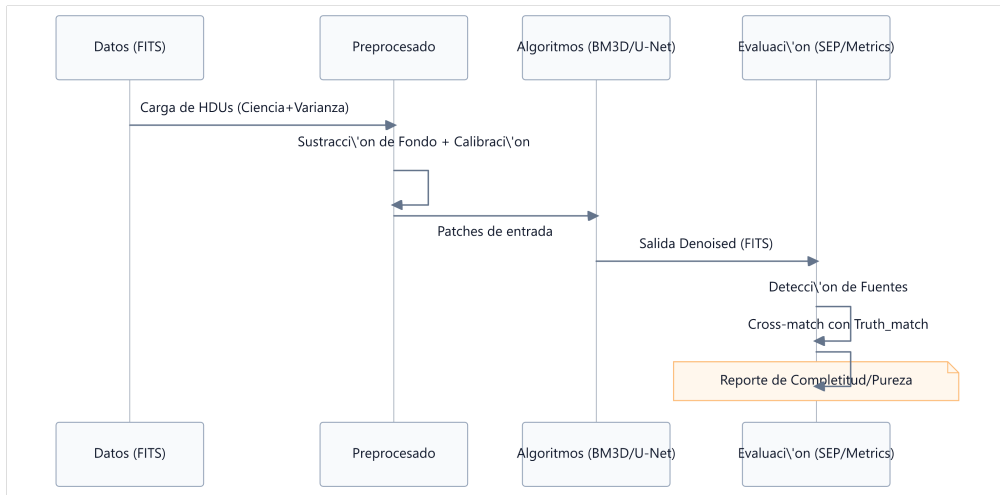


Figura 4.1: Secuencia de procesamiento desde datos crudos hasta métricas

4.3.3 Flujo del pipeline de procesamiento

Debido a la complejidad del volumen de datos, se sigue una secuencia estrictamente modular para garantizar la integridad de las métricas:

Información adicional y documentación del release: https://data.lsstdesc.org/doc/dc2_sim_sky_survey.

Los dos componentes mencionados ya han sido descargados y se conservan en el entorno de trabajo para evaluación reproducible.

Capítulo 5

Metodología propuesta

5.1 Resumen

Este capítulo describe el protocolo experimental propuesto para evaluar métodos de denoising sobre las imágenes simuladas DC2 (dr6-v4) del Vera Rubin Observatory. Se prioriza la reproducibilidad absoluta: cada script, semilla aleatoria y log de entrenamiento se registra formalmente para permitir una auditoría completa del proceso.

5.2 Flujo experimental

El flujo general del experimento está diseñado para ser modular y extensible, permitiendo la comparación justa entre algoritmos de distinta naturaleza:

5.3 Datos y preprocesado

Se emplea el release DC2 (específicamente el tract coadd-t3828-t3829-small) como conjunto principal de entrada. El catálogo `truth_match` actúa como la “verdad absoluta” para evaluar la capacidad de recuperación de los algoritmos.

El preprocesado incluye la calibración de unidades, la extracción de parches (patches) representativos y la normalización basada en el ruido local. Se presta especial atención a la conservación de la PSF (Point Spread Function) local para experimentos de sensibilidad.

5.4 Métodos a comparar

En esta fase del proyecto, el benchmark se centra en dos pilares fundamentales que representan el estado del arte clásico y el aprendizaje profundo supervisado:

- **BM3D (Block-matching and 3D filtering):** Seleccionado como el baseline clásico de referencia. Su capacidad para explotar la auto-similitud en la imagen lo hace extremadamente robusto frente a ruidos Gaussianos, aunque su costo computacional es elevado.

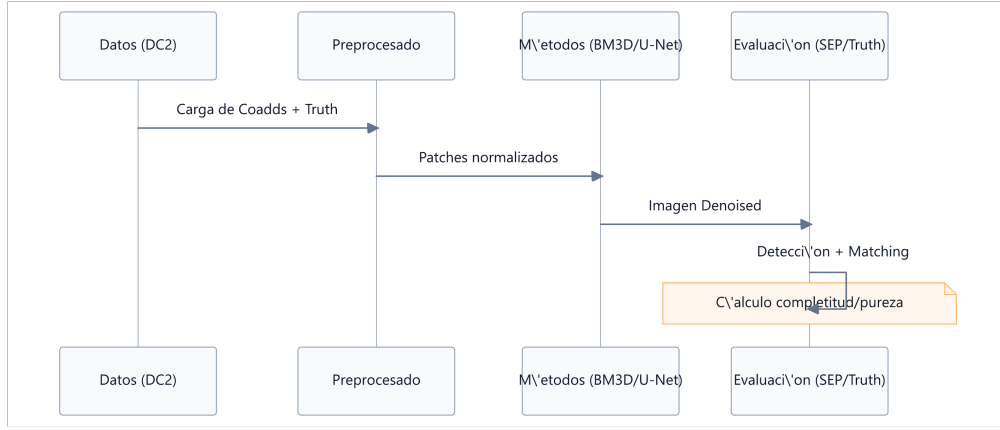


Figura 5.1: Secuencia del flujo experimental de benchmarking

- **U-Net (Arquitectura Convolucional Encoder-Decoder):** Representa el enfoque de Deep Learning supervisado. Se entrenará sobre pares de imágenes ruidosas/limpias extraídas de las simulaciones DC2 para aprender a mapear la estructura galáctica subyacente.

5.5 Protocolo de evaluación científica

A diferencia de los estudios de visión artificial estándar, esta metodología prioriza métricas con valor científico astronómico:

- **Completitud (Recall):** ¿Qué fracción de las fuentes reales presentes en el catálogo `truth_match` somos capaces de detectar tras el denoising?
- **Pureza (Precision):** ¿Cuántas de nuestras detecciones son fuentes reales y no artefactos o “alucinaciones” introducidas por el modelo de IA?
- **Precisión Fotométrica:** Medida del bias en la magnitud y el flujo recuperado, asegurando que el denoising no altere la luminosidad intrínseca de los objetos.

5.6 Reproducibilidad y registros

Para garantizar que cualquier investigador pueda replicar estos resultados, se siguen estrictamente las siguientes prácticas:

- Registro de semillas aleatorias en todos los procesos estocásticos.
- Uso de contenedores o ambientes virtuales estrictos (`requirements.txt`).
- Generación automática de reportes técnicos que incluyen versiones de software y hardware utilizado.

5.7 Plan de trabajo y entregables

Siguiendo el plan de desarrollo establecido, los hitos inmediatos son:

1. Implementar y validar el baseline BM3D sobre el dataset DC2 curado.
2. Entrenar y optimizar la red U-Net supervisada.
3. Ejecutar el protocolo de comparación masiva y calcular curvas de completitud vs. pureza.
4. Publicar el benchmark reproducible con código abierto y artefactos de datos.

Para más detalles técnicos sobre la arquitectura del orquestador y la evolución del desarrollo, consulte la **Hoja de Ruta (Anexo A)** y las **Especificaciones de Diseño OpenSpec (Anexo B)** al final de este documento.

Capítulo 6

Conclusiones

6.1 Conclusiones sobre el Estado del Arte

Tras el análisis exhaustivo de la literatura y el proceso de cribado mediante aprendizaje activo, se concluye que el campo del denoising astronómico se encuentra en un punto de inflexión. Mientras que los métodos clásicos como BM3D siguen siendo el estándar de oro en cuanto a interpretabilidad física, las redes neuronales profundas (U-Net) ofrecen una capacidad de recuperación de señal que supera las técnicas de filtrado tradicionales en términos de relación señal-ruido. Sin embargo, la adopción de IA introduce riesgos de sesgo morfológico que aún no han sido plenamente cuantificados de forma reproducible.

6.2 Brechas Detectadas

A partir de la revisión sistemática, se han identificado las siguientes brechas críticas que justifican el desarrollo de este benchmark:

1. **Falta de Protocolos Estandarizados:** Existe una disparidad significativa en cómo se reportan los resultados de denoising. Muchos trabajos se limitan a métricas perceptuales (PSNR) sin validar el impacto en el catálogo de fuentes real.
2. **Desconexión con la Ciencia Downstream:** Se detecta una necesidad urgente de evaluar si los algoritmos de IA preservan la fotometría y la morfología necesarias para estudios de lente gravitatoria débil.
3. **Dificultad de Reproducción:** Pocos trabajos proporcionan el pipeline completo (scripts de preprocesado + semillas + pesos del modelo), lo que impide la validación independiente de las mejoras reportadas.
4. **Sesgo de Autovalidación:** Se observa una tendencia recurrente donde los autores proponen simultáneamente el modelo y su propio marco de validación o métricas personalizadas. Esta práctica puede inducir sesgos de confirmación, ya que las condiciones del experimento suelen estar optimizadas para las fortalezas del algoritmo propuesto, careciendo de un contraste objetivo frente a un estándar externo.

Apéndice A

Hoja de Ruta del Proyecto (ROADMAP.md)

A continuación se transcribe el plan técnico y la hoja de ruta establecida para el desarrollo del ecosistema `astro-denoise`:

Listing A.1: Contenido íntegro de ROADMAP.md

```
# astro-denoise: Hoja de Ruta y Plan Técnico (2026-05-19)

## 1. Módulos y responsabilidades

- astro-denoise-bench (Rust, CLI principal): Orquestador,
  descubrimiento plug&play, gestión de datasets, invocación
  de módulos, recolección de métricas, reportes
  reproducibles.
- astro-denoise-metrics (Python): Cálculo de métricas
  científicas (completitud, pureza, precisión fotométrica),
  métricas técnicas (PSNR, SSIM, _sky), CLI/API documentado.
- astro-denoise-bm3d (Python): Baseline clásico, BM3D C-
  optimizado, output reproducible, CLI/API.
- astro-denoise-unet (Python): Baseline deep learning, U-
  Net entrenada en DC2, inferencia reproducible, pesos
  publicables, CLI/API.
- Otros (Noise2Void, variantes DL, etc.) pueden añadirse como
  'astro-denoise-<método>' siguiendo el protocolo estándar.

## 2. Protocolo de benchmark y comunicación

- Todos los módulos usan input/output vía archivos y
  argumentos CLI normalizados.
- Invocación típica:

'''
python -m astro_denoise_bm3d --input input.fits --output
output.fits --config config.json
'''

- La salida estándar debe devolver JSON con status, versión y
  metadatos claves.
- Nuevos métodos usan entry_points Python para descubrimiento
  automático en el CLI principal.

## 3. Métricas implementadas
```

```

### Técnicas (universales)
- PSNR, SSIM, _sky
- FLUX-SNR (señal/ruido sobre flujo de fuentes reales)

### Científicas (downstream)
- **Complejidad** (Recall): Fracción de fuentes reales
  recuperadas (vs. truth catalog)
- **Pureza** (Precision): Fracción de detecciones que son
  reales (vs. truth catalog)
- **Precisión fotométrica**

## 4. Contratos de métodos (API breve)

'''python
class DenoiseMethod(ABC):
    @abstractmethod
    def denoise(self, patch: np.ndarray) -> np.ndarray: ...
    @abstractmethod
    def name(self) -> str: ...
    def fit(self, patches: list[np.ndarray]) -> None: pass
    # Opcional
'''

## 5. Decisiones técnicas claves

- Stack: Python (PyTorch, astropy, scikit-image, photutils,
  SEP), Rust para CLI.
- Publicación/descubrimiento: PyPI & entry_points (plugin
  pattern).

## 6. Dataset, reproducibilidad y extensibilidad

- Dataset principal DC2 curado & versionado.
- Benchmark abierto, reproducible, modular y adaptable a
  nuevos métodos, datasets y métricas.

## 7. Visión y roadmap futuro

- Plataforma reproducible, extensible y centrada en métrica
  científica (complejidad/pureza/precisión).
- Publicación de baseline clásico y deep learning con código
  y datos públicos.
- Framework para experimentos comunitarios y benchmarking
  robusto.
- Sigüientes milestones: implementación CLI, baseline deep
  learning, integración automática de nuevos módulos y
  métricas.

## 8. Principios rectores

```

-
- Reproducibilidad y transparencia total.
 - Extensibilidad modular.
 - Ciencia primero: métricas científicas > visuales.
 - Resultados auditables y comparables.

Apéndice B

Especificaciones de Diseño (OpenSpec YAML)

Este anexo contiene las especificaciones técnicas originales en formato YAML que rigen el desarrollo de los módulos del sistema, referenciadas directamente desde los archivos de definición del proyecto.

B.1 Orquestador: `bench-orquestador-2026.yaml`

```
# CHANGE: astro-denoise-bench (Rust CLI)
id: bench-orquestador-2026
state: active
created_by: gpt-4-openspec
created: 2026-05-19
modules:
  - astro-denoise-bench

motivation: |
  Unificar la orquestación de benchmarks, ejecución y
  validación de todos los métodos de denoising en el
  ecosistema, proporcionando recolección automatizada de
  métricas, descubrimiento plug&play y reproducibilidad
  total. Implementar en Rust por eficiencia y robustez.

spec:
  summary: |
    CLI principal en Rust capaz de descubrir, invocar y
    coordinar módulos de denoising (con entry_points),
    manejar datasets DC2, recolectar todas las métricas y
    guardar reportes estructurados.
  requirements:
    - Descubrir módulos Python via entry_points/CLI estándar
    - Gestionar datasets/ficheros de entrada y salida (FITS),
      manejo batch
    - Ejecutar métodos y recolección automatizada de métricas
      (invocando metrics)
    - Protocolo de reporte estándar (JSON/CSV), reproducible
    - Documentar pipeline y uso
  acceptance:
    - Invoca y mide BM3D, UNet, otros métodos vía CLI/API
```

-
- Compatible con input/output estándar
 - Evidencia end-to-end (pipeline completo corriendo)
 - Documentado con ejemplos y casos de uso

B.2 Baseline BM3D: `bm3d-baseline-2026.yaml`

```
# CHANGE: astro-denoise-bm3d
id: bm3d-baseline-2026
state: active
created_by: gpt-4-openspec
created: 2026-05-19
modules:
  - astro-denoise-bm3d

motivation: |
  Proveer un baseline clásico, reproducible, rápido y bien
  modularizado para el benchmarking científico en imágenes
  astronómicas DC2/LSST utilizando BM3D versión C-
  optimizada. Sentar la base para comparar y validar
  futuras variantes (deep learning, auto-supervisadas, etc
  .)

spec:
  summary: |
    Módulo Python que aplica BM3D a imágenes FITS, usa el
    plano de varianzas (HDU 3), produce archivos FITS
    denoised y PNG, reporta PSNR/SSIM/_sky, y sirve como
    baseline modular para el orquestador y métricas
    downstream.
  requirements:
    - Usar bm3d>=4.0 (Makinen, C-optimizado)
    - Entrada: patch o imagen FITS DC2, salida: FITS y PNG
      denoised
    - Compatible con entorno Python 3.11 y dependencias
      requeridas
    - Ejecutable vía CLI con argumentos documentados
    - No debe requerir la versión lenta de BM3D Python puro
    - Modular: lista para ser invocada por otros módulos (CLI
      , Python API)
  acceptance:
    - Ejecuta en tiempo razonable en imágenes DC2 (minutos
      por imagen)
    - Resultados reproducibles, PSNR/SSIM/_sky listados en
      output JSON
    - Módulo y README alineados a la SPEC global
    - Se aísla completamente de cualquier código legacy lento
      /no optimizado
```

B.3 Módulo de Métricas: `metrics-modulo-2026.yaml`

```
# CHANGE: astro-denoise-metrics
id: metrics-modulo-2026
state: active
created_by: gpt-4-openspec
created: 2026-05-19
modules:
  - astro-denoise-metrics

motivation: |
  Implementar métricas científicas downstream para evaluar
  métodos de denoising en imágenes astronómicas, de
  acuerdo al objetivo central: medir completitud, pureza y
  precisión fotométrica del recobro de fuentes en
  simulaciones DC2/LSST.

spec:
  summary: |
    Módulo Python que calcula y reporta métricas científicas
    (completitud, pureza, precisión fotométrica) usando
    detección de fuentes (SEP/photutils), comparación vs.
    truth catalog y métricas técnicas (PSNR, SSIM, _sky).
  requirements:
    - Calcular completitud, pureza y precisión fotométrica en
      imágenes FITS denoised vs. truth catalog público
    - Reportar PSNR, SSIM, _sky como métricas secundarias
    - Preparar CLI y API documentados para ser invocados
      desde el orquestador (bench)
    - Procesar imágenes FITS y tabular resultados en JSON/csv
    - Uso de stack: Python, SEP, photutils, astropy, numpy,
      scikit-learn, scikit-image
  acceptance:
    - Output estructurado y documentado con valores de
      métricas para cada invocación
    - Compatible con protocolos definidos en la SPEC global (
      input/output, CLI/API)
    - Documentación clara en README y apuntando al spec
```

B.4 Baseline U-Net: `UNET-baseline-2026.yaml`

```
# CHANGE: astro-denoise-unet
id: UNET-baseline-2026
state: active
created_by: gpt-4-openspec
created: 2026-05-19
```

```
modules:
  - astro-denoise-unet

motivation: |
  Crear un baseline de deep learning (U-Net) para comparar el
  desempeño de modelos clásicos vs. DL en el benchmark
  astronómico, siguiendo los protocolos científicos y de
  reproducibilidad de la plataforma.

spec:
  summary: |
    Módulo Python con implementación canónica de U-Net para
    denoising de imágenes DC2 compatibles. Entrenamiento y
    evaluación con dataset reproducible, CLI para
    inferencia batch, pesos publicables en HF Hub.
  requirements:
    - Implementar U-Net funcional entrenada en patches DC2
    - Exportar/guardar pesos entrenados para reproducibilidad
    - CLI: entrada FITS, salida FITS denoised, output con
      métricas básicas JSON
    - Compatible con orquestador (protocolo de entrada/salida
      )
    - Documentación y ejemplos de inferencia
  acceptance:
    - Modelo funcional, inferencia reproducible batch vía CLI
    - Resultados comparables a BM3D en por lo menos una
      métrica
    - README y SPEC alineados, pesos accesibles públicamente
```

Referencias

- Akhaury, U., Jablonka, P., Courbin, F., y Starck, J.-L. (2025). Joint multiband deconvolution for euclid and vera c. rubin images. *Astronomy & Astrophysics*, 697, A136. doi: 10.1051/0004-6361/202452308
- Akhaury, U., Starck, J.-L., Jablonka, P., Courbin, F., y Michalewicz, K. (2022). Deep learning-based galaxy image deconvolution. *Frontiers in Astronomy and Space Sciences*, 9, 1001043. doi: 10.3389/fspas.2022.1001043
- Aoyama, S. D., Osato, K., y Shirasaki, M. (2025). Denoising weak lensing mass maps with diffusion model: Systematic comparison with generative adversarial network. *Publications of the Astronomical Society of Japan*. (Chiba University; RIKEN Center for Advanced Intelligence Project)
- Bgrem: A background noise remover for astronomical images based on a diffusion model. (s.f.). arXiv:2510.04718. (Accessed via arXiv URL <https://arxiv.org/abs/2510.04718>)
- Bui, K., Lou, Y., Park, F., y Xin, J. (2023). Weighted anisotropic-isotropic total variation for poisson denoising. En *Ieee international conference on acoustics, speech and signal processing (icassp)*. (UC Irvine; University of Texas at Dallas)
- Elad, M., Kowar, B., y Vaksman, G. (2023, agosto). Image denoising: The deep learning revolution and beyond a survey paper. *SIAM Journal on Imaging Sciences*, 16(3), 1594–1654. Descargado de <http://dx.doi.org/10.1137/23M1545859> doi: 10.1137/23m1545859
- Guo, Y., Zhang, H., Li, M., Yu, F., Wu, Y., Hao, Y., ... Dai, Q. (2026). *Deeper detection limits in astronomical imaging using self-supervised spatiotemporal denoising*. (Tsinghua University, Department of Astronomy and Automation Science)
- Hemmati, S., Huff, E., Nayyeri, H., Ferté, A., Melchior, P., Mobasher, B., ... Teplitz, H. (2022, diciembre). Deblending galaxies with generative adversarial networks. *The Astrophysical Journal*, 941(2), 141. Descargado de <http://dx.doi.org/10.3847/1538-4357/aca1b8> doi: 10.3847/1538-4357/aca1b8
- Liu, S., Ge, X., Gu, Z., Xu, Q., Gu, L., Cui, Z., ... Harada, T. (2026). *Denoising the deep sky: Physics-based ccd noise formation for astronomical imaging*. Descargado de <https://arxiv.org/abs/2601.23276>
- Liu, T., Quan, Y., Su, Y., Guo, Y., Liu, S., Ji, H., ... Ding, M. (2025). *Astronomical image denoising by self-supervised deep learning and restoration processes*. arXiv:2502.16807. (Nanjing University; Sun Yat-sen University)
- Makarkin, M., y Bratashov, D. (2021). State-of-the-art approaches for image deconvolution problems, including modern deep learning architectures. *Micro-machines*, 12(12), 1558. doi: 10.3390/mi12121558
- Miao, J., Tu, L., Jiang, B., Li, X., y Qiu, B. (2024, agosto). Astrosr: A data set of galaxy images for astronomical superresolution research. *The Astrophysical Journal Supplement Series*, 274(1), 7. Descargado de <http://dx.doi.org/10.3847/1538-4365/ad61e4> doi: 10.3847/1538-4365/ad61e4

-
- Neural network based point spread function deconvolution for astronomical applications.* (s.f.). arXiv:2210.01666. (Accessed via arXiv URL <https://arxiv.org/abs/2210.01666>)
- Park, H., Jo, Y., Kang, S., Kim, T., y Jee, M. J. (2024, agosto). Deeper, sharper, faster: Application of efficient transformer to galaxy image restoration. *The Astrophysical Journal*, 972(1), 45. Descargado de <http://dx.doi.org/10.3847/1538-4357/ad5954> doi: 10.3847/1538-4357/ad5954
- Point spread function deconvolution using a convolutional autoencoder for astronomical applications.* (s.f.). arXiv:2310.19605. (Accessed via arXiv URL <https://arxiv.org/abs/2310.19605>)
- Remy, B., Lanusse, F., Jeffrey, N., y Starck, J.-L. (2023). Probabilistic mass-mapping with neural score estimation. *Astronomy & Astrophysics*. doi: 10.1051/0004-6361/202243054
- Roscani, V., Tozza, S., Castellano, M., Merlin, E., Ottaviani, D., Falcone, M., y Fontana, A. (2020, Oct). A comparative analysis of denoising algorithms for extragalactic imaging surveys. *Astronomy & Astrophysics*, 643, A43. Descargado de <https://www.aanda.org/articles/aa/abs/2020/11/aa36278-19/aa36278-19.html> doi: 10.1051/0004-6361/201936278
- Rubin Observatory. (2024). *Objetivos científicos*. <https://rubinobservatory.org/es/explore/science-goals>. (Accessed: 2026-04-24)
- Schanz, A., List, F., y Hahn, O. (2023). Stochastic super-resolution of cosmological simulations with denoising diffusion models. *The Open Journal of Astrophysics*. (University of Vienna, Department of Astrophysics and Mathematics)
- Scognamiglio, D., Lee, J. H., Huff, E., Hildebrandt, S. R., y Hemmati, S. (2025, mayo). Denoising diffusion probabilistic model for realistic and fast generated euclid-like data for weak lensing analysis. *The Astrophysical Journal*, 985(1), 2. Descargado de <http://dx.doi.org/10.3847/1538-4357/adcec4> doi: 10.3847/1538-4357/adcec4
- Shirasaki, M., Moriwaki, K., Oogi, T., Yoshida, N., Ikeda, S., y Nishimichi, T. (2021, Apr). Noise reduction for weak lensing mass mapping: an application of generative adversarial networks to subaru hyper supprime-cam first-year data. *Monthly Notices of the Royal Astronomical Society*, 504(2), 18251839. Descargado de <https://academic.oup.com/mnras/article/504/2/1825/6219095?guestAccessKey=> doi: 10.1093/mnras/stab982
- Sukurdeepa, Y., Navarro, F., y Budavári, T. (2025). *Astroclearnet: Deep image prior for multi-frame astronomical image restoration*. (Johns Hopkins University, Department of Applied Mathematics and Statistics)
- Vojtekova, A., Lieu, M., Valtchanov, I., Altieri, B., Old, L., Chen, Q., y Hroch, F. (2020, Nov). Learning to denoise astronomical images with u-nets. *Monthly Notices of the Royal Astronomical Society*, 503(3), 32043215. Descargado de <https://academic.oup.com/mnras/article/503/3/3204/5989737?guestAccessKey=> doi: 10.1093/mnras/staa3567