DeepShape: Radio Weak Lensing Measurements using Deep Learning

Cosmology in the Alps 2026, Les Diablerets

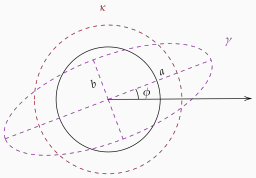
Priyam Tripathi

Supervisors: Prof. André Ferrari and Prof. Simon Prunet

20 March, 2026

Observatoire de la Côte d'Azur

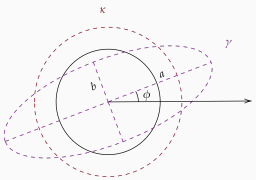
Cosmic shear: Weak lensing induces small coherent distortions in galaxy shapes.



$$\boldsymbol{\epsilon} \approx \boldsymbol{\epsilon}^I + \boldsymbol{\gamma}$$

$\boldsymbol{\epsilon}, \boldsymbol{\epsilon}^I, \boldsymbol{\gamma} \in \mathbb{R}^2$

Cosmic shear: Weak lensing induces small coherent distortions in galaxy shapes.



$$\boldsymbol{\epsilon} \approx \boldsymbol{\epsilon}^I + \boldsymbol{\gamma}$$

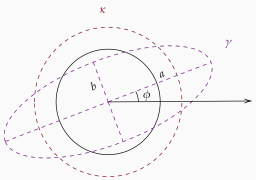
$\boldsymbol{\epsilon}, \boldsymbol{\epsilon}^I, \boldsymbol{\gamma} \in \mathbb{R}^2$

Radio-specific challenges

- Shear requires high precision shapes for large samples: $\hat{\boldsymbol{\gamma}} = \langle \boldsymbol{\epsilon} \rangle$

Introduction

Cosmic shear: Weak lensing induces small coherent distortions in galaxy shapes.



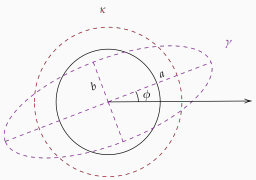
$$\boldsymbol{\epsilon} \approx \boldsymbol{\epsilon}^I + \boldsymbol{\gamma}$$

$\boldsymbol{\epsilon}, \boldsymbol{\epsilon}^I, \boldsymbol{\gamma} \in \mathbb{R}^2$

Radio-specific challenges

- Shear requires high precision shapes for large samples: $\hat{\boldsymbol{\gamma}} = \langle \boldsymbol{\epsilon} \rangle$
- Visibility domain: accurate but slow and model dependent

Cosmic shear: Weak lensing induces small coherent distortions in galaxy shapes.



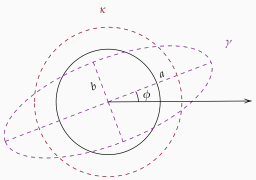
$$\boldsymbol{\epsilon} \approx \boldsymbol{\epsilon}^I + \boldsymbol{\gamma}$$

$\boldsymbol{\epsilon}, \boldsymbol{\epsilon}^I, \boldsymbol{\gamma} \in \mathbb{R}^2$

Radio-specific challenges

- Shear requires high precision shapes for large samples: $\hat{\boldsymbol{\gamma}} = \langle \boldsymbol{\epsilon} \rangle$
- Visibility domain: accurate but slow and model dependent
- Image domain: reconstruction unreliable for faint sources

Cosmic shear: Weak lensing induces small coherent distortions in galaxy shapes.



$$\boldsymbol{\epsilon} \approx \boldsymbol{\epsilon}^I + \boldsymbol{\gamma}$$

$\boldsymbol{\epsilon}, \boldsymbol{\epsilon}^I, \boldsymbol{\gamma} \in \mathbb{R}^2$

Radio-specific challenges

- Shear requires high precision shapes for large samples: $\hat{\boldsymbol{\gamma}} = \langle \boldsymbol{\epsilon} \rangle$
- Visibility domain: accurate but slow and model dependent
- Image domain: reconstruction unreliable for faint sources

Goal: scalable and accurate shape measurement with deep learning.

DeepShape
isolated galaxies
[Tripathi et al., 2025]

Plug-and-Play Image Reconstruction

⇓

Shape Measurement Network

DeepShape
isolated galaxies
[Tripathi et al., 2025]

Faceting for Source Isolation

⇓

Plug-and-Play Image Reconstruction

⇓

Deblending Network (VAE + Attention)

⇓

Shape Measurement Network

DeepShape II
wide-field extension
[Tripathi et al., 2026]
(submitted)

Wide Field Measurements

- SKA-MID primary beam: $\sim 1\ \mathrm{deg}^2$, up to 10 000 galaxies.
- In the visibility domain, galaxies overlap — separating sources is difficult.

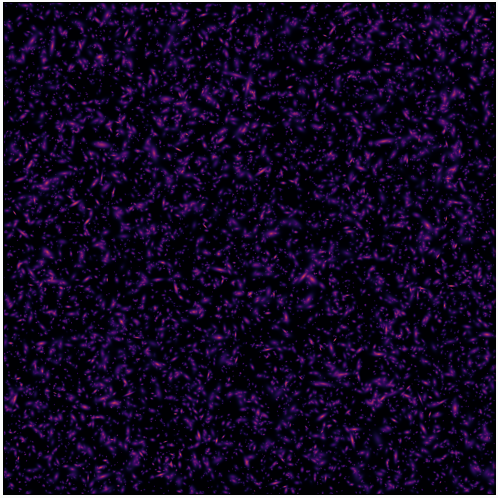


Figure 1: Simulated true sky. 8192×8192 px / $\sim$400 arcmin^2.

Wide Field Measurements

- SKA-MID primary beam: $\sim 1\ \mathrm{deg}^2$, up to 10 000 galaxies.
- In the visibility domain, galaxies overlap — separating sources is difficult.
- Standard approach: model-based method
  [Chang et al., 2004, Rivi et al., 2016] — jointly fit all sources in the FoV, but can be slow.

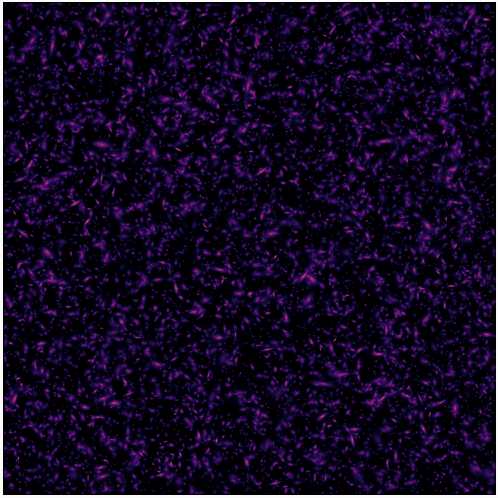


Figure 1: Simulated true sky. 8192×8192 px / $\sim$400 arcmin^2.

Wide Field Measurements

- SKA-MID primary beam: $\sim 1\ \mathrm{deg}^2$, up to 10 000 galaxies.
- In the visibility domain, galaxies overlap — separating sources is difficult.
- Standard approach: model-based method
  [Chang et al., 2004, Rivi et al., 2016] — jointly fit all sources in the FoV, but can be slow.
- Our approach: small facets per source $\rightarrow$ reconstruct $\rightarrow$ deblend $\rightarrow$ measure shapes.

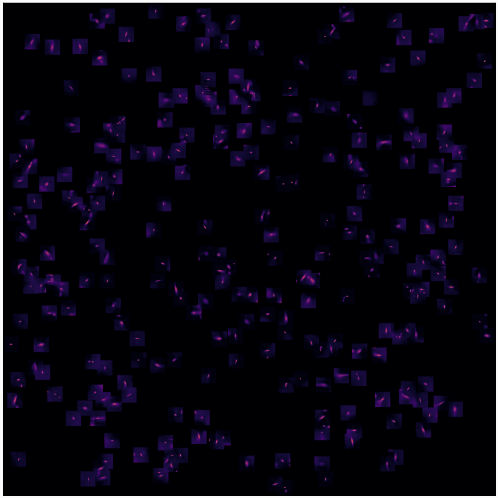


Figure 1: Reconstructed sky using facets

Simulation Setup

1. Galaxy Population

- T-RECS SFG catalog [Bonaldi et al., 2018]
- Flux range: 1–200 μJy
- Surface density: $n_g \approx 32$ gal/arcmin^2
- Sérsic light profile

$$I(r) = I_e \, e^{-b_{n_S}[(r/r_e)^{1/n_S}-1]}$$

2. Dataset Construction

- Wide simulation (train/val)
- 150 disjoint $1° \times 1°$ patches
- $\approx 25\,200 \times 25\,200$ pixels / patch
- Deep simulation (test)
- Realistic clustering from P-Millennium simulation [Baugh et al., 2019]
- 128×128 pixel stamps at $0.143''$ pixel scale

3. Visibility Simulation

- SKA-MID AA4 configuration
- $\nu_0 = 1.4$ GHz, 8 hr observation
- Simulated with RASCIL
- Briggs weighting ($R = -0.5$)
- Image RMS $\sigma \approx 0.71$ μJy/beam

- Measured visibilities (relative to pointing centre):

$$V(u,v,w) = \iint I(l,m)\,\frac{e^{-2\pi j[ul+vm+w(n-1)]}}{n}\,\mathrm{d}l\,\mathrm{d}m, \quad n = \sqrt{1-l^2-m^2}$$

Faceting

- Measured visibilities (relative to pointing centre):

$$V(u,v,w) = \iint I(l,m)\,\frac{e^{-2\pi j[ul+vm+w(n-1)]}}{n}\,\mathrm{d}l\,\mathrm{d}m, \quad n = \sqrt{1-l^2-m^2}$$

- To form a facet centred at (l_g, m_g), apply a phase shift:

$$V'(u,v,w) = V(u,v,w)\,e^{2\pi j(ul_g+vm_g+wn_g)}$$

Faceting

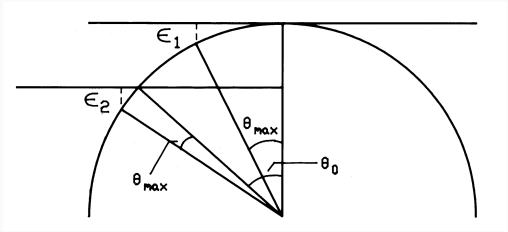
- Measured visibilities (relative to pointing centre):

$$V(u,v,w) = \iint I(l,m)\,\frac{e^{-2\pi j[ul+vm+w(n-1)]}}{n}\,\mathrm{d}l\,\mathrm{d}m, \quad n = \sqrt{1-l^2-m^2}$$

- To form a facet centred at (l_g, m_g), apply a phase shift:

$$V'(u,v,w) = V(u,v,w)\,e^{2\pi j(ul_g+vm_g+wn_g)}$$

- Phase shifting limits the usable FoV within each facet.



Faceting

- Measured visibilities (relative to pointing centre):

$$V(u,v,w) = \iint I(l,m)\,\frac{e^{-2\pi j[ul+vm+w(n-1)]}}{n}\,\mathrm{d}l\,\mathrm{d}m, \quad n = \sqrt{1-l^2-m^2}$$

- To form a facet centred at (l_g, m_g), apply a phase shift:

$$V'(u,v,w) = V(u,v,w)\,e^{2\pi j(ul_g+vm_g+wn_g)}$$

- Phase shifting limits the usable FoV within each facet.
- Coplanar faceting [Kogan and Greisen, 2009] corrects (u,v) coordinates:

$$u' = u - w\,\frac{l_g}{\sqrt{l_g^2+m_g^2}}, \qquad v' = v - w\,\frac{m_g}{\sqrt{l_g^2+m_g^2}}$$

Faceting

- Measured visibilities (relative to pointing centre):

$$V(u,v,w) = \iint I(l,m)\,\frac{e^{-2\pi j[ul+vm+w(n-1)]}}{n}\,\mathrm{d}l\,\mathrm{d}m, \quad n = \sqrt{1-l^2-m^2}$$

- To form a facet centred at (l_g, m_g), apply a phase shift:

$$V'(u,v,w) = V(u,v,w)\,e^{2\pi j(ul_g+vm_g+wn_g)}$$

- Phase shifting limits the usable FoV within each facet.
- Coplanar faceting [Kogan and Greisen, 2009] corrects (u,v) coordinates:

$$u' = u - w\,\frac{l_g}{\sqrt{l_g^2+m_g^2}}, \qquad v' = v - w\,\frac{m_g}{\sqrt{l_g^2+m_g^2}}$$

- Coarse UV grid ($\Delta u \uparrow$) $\Rightarrow$ smaller facets (FoV $\sim 1/\Delta u$), faster to process.

Faceting

- Measured visibilities (relative to pointing centre):

$$V(u,v,w) = \iint I(l,m)\,\frac{e^{-2\pi j[ul+vm+w(n-1)]}}{n}\,\mathrm{d}l\,\mathrm{d}m, \quad n = \sqrt{1-l^2-m^2}$$

- To form a facet centred at (l_g, m_g), apply a phase shift:

$$V'(u,v,w) = V(u,v,w)\,e^{2\pi j(ul_g+vm_g+wn_g)}$$

- Phase shifting limits the usable FoV within each facet.
- Coplanar faceting [Kogan and Greisen, 2009] corrects (u,v) coordinates:

$$u' = u - w\,\frac{l_g}{\sqrt{l_g^2+m_g^2}}, \qquad v' = v - w\,\frac{m_g}{\sqrt{l_g^2+m_g^2}}$$

- Coarse UV grid ($\Delta u \uparrow$) $\Rightarrow$ smaller facets (FoV $\sim 1/\Delta u$), faster to process.
- Trade-off: some small-scale structure is smoothed and aliasing increases

Measurement model

$$\mathsf{i}_D = H\mathsf{i}_T + \mathsf{N}$$

Plug-and-Play Reconstruction

Measurement model

$$\mathsf{i}_D = H\mathsf{i}_T + \mathsf{N}$$

HQS objective

$$\mathcal{L}(\mathsf{i}, \mathsf{z}) = \frac{1}{2\sigma^2}\|\mathsf{i}_D - H\mathsf{i}\|^2 + \lambda\,\Phi(\mathsf{z}) + \frac{\mu}{2}\|\mathsf{i} - \mathsf{z}\|^2$$

Measurement model

$$\mathsf{i}_D = H\mathsf{i}_T + \mathsf{N}$$

HQS objective

$$\mathcal{L}(\mathsf{i}, \mathsf{z}) = \frac{1}{2\sigma^2}\|\mathsf{i}_D - H\mathsf{i}\|^2 + \lambda\,\Phi(\mathsf{z}) + \frac{\mu}{2}\|\mathsf{i} - \mathsf{z}\|^2$$

Data step

$$\mathsf{i}_{k+1} = (H^\dagger H + \mu_k\sigma^2 I)^{-1}(H^\dagger \mathsf{i}_D + \mu_k\sigma^2 \mathsf{z}_k)$$

Measurement model

$$\mathsf{i}_D = H\mathsf{i}_T + \mathsf{N}$$

HQS objective

$$\mathcal{L}(\mathsf{i}, \mathsf{z}) = \frac{1}{2\sigma^2}\|\mathsf{i}_D - H\mathsf{i}\|^2 + \lambda\,\Phi(\mathsf{z}) + \frac{\mu}{2}\|\mathsf{i} - \mathsf{z}\|^2$$

Data step

$$\mathsf{i}_{k+1} = (H^\dagger H + \mu_k\sigma^2 I)^{-1}(H^\dagger \mathsf{i}_D + \mu_k\sigma^2 \mathsf{z}_k)$$

Regularisation step

$$\mathsf{z}_{k+1} = DN_{\hat{\theta}}(\mathsf{i}_{k+1}; \tau_k), \quad \tau_k = \sqrt{\lambda/\mu_k}$$

Measurement model

$$\mathsf{i}_D = H\mathsf{i}_T + \mathsf{N}$$

HQS objective

$$\mathcal{L}(\mathsf{i}, \mathsf{z}) = \frac{1}{2\sigma^2}\|\mathsf{i}_D - H\mathsf{i}\|^2 + \lambda\,\Phi(\mathsf{z}) + \frac{\mu}{2}\|\mathsf{i} - \mathsf{z}\|^2$$

Data step

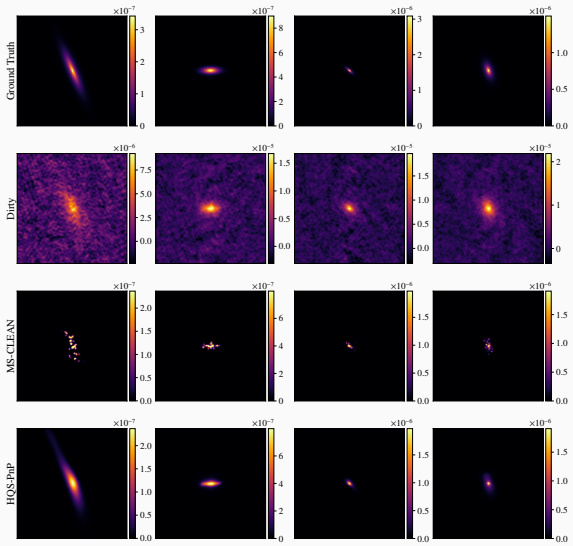
$$\mathsf{i}_{k+1} = (H^\dagger H + \mu_k\sigma^2 I)^{-1}(H^\dagger \mathsf{i}_D + \mu_k\sigma^2 \mathsf{z}_k)$$

Regularisation step

$$\mathsf{z}_{k+1} = DN_{\hat{\theta}}(\mathsf{i}_{k+1}; \tau_k), \quad \tau_k = \sqrt{\lambda/\mu_k}$$

- Learned prior implemented via DRUNet denoiser [Zhang et al., 2021]
- Naturally supports variable PSFs

Comparison of HQS-PnP and MS-CLEAN



- MS-CLEAN images are **model-only** (no CLEAN beam convolution, residuals not added) → improves shape measurement
- Only 1 major cycle: roughly equal computation time

Comparison of HQS-PnP and MS-CLEAN

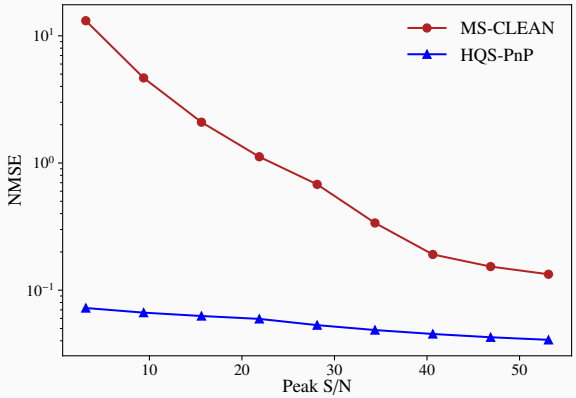
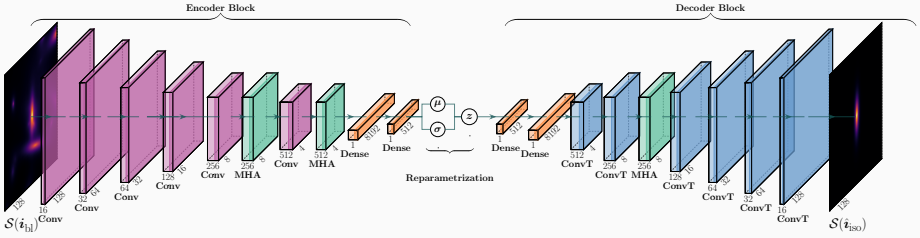


Figure 2: Variation of NMSE $\equiv \frac{\sum_p \|\hat{x}^p - i_T^p\|_2^2}{\sum_p \|i_T^p\|_2^2}$ as a function of peak S/N. The statistics were calculated on 20,000 test images split into 9 peak S/N bins.

Deblender Network

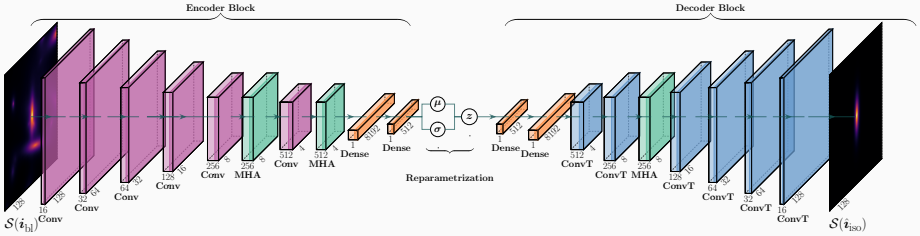


- Variational Autoencoder (VAE) inspired network learns a latent representation $\boldsymbol{z}$ by maximizing the ELBO:

$$\log p_{\boldsymbol{\theta}}(i_{\mathrm{iso}}) \geq \mathbb{E}_{\boldsymbol{z}\sim q_{\boldsymbol{\phi}}(\boldsymbol{z}|i_{\mathrm{bl}})}\left[\log p_{\boldsymbol{\theta}}(i_{\mathrm{iso}} \mid \boldsymbol{z})\right] - \mathbb{D}_{\mathrm{KL}}\left[q_{\boldsymbol{\phi}}(\boldsymbol{z} \mid i_{\mathrm{bl}}) \parallel p(\boldsymbol{z})\right]$$

- $p_{\boldsymbol{\theta}}(i_{\mathrm{iso}})$ evidence
- $p_{\boldsymbol{\theta}}(i_{\mathrm{iso}}|\boldsymbol{z})$ likelihood (decoder)
- $q_{\boldsymbol{\phi}}(\boldsymbol{z}|i_{\mathrm{bl}})$ posterior (encoder)
- $p(\boldsymbol{z})$ latent prior

Deblender Network



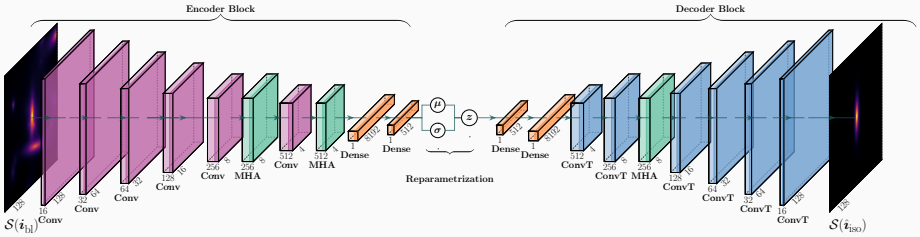
- Variational Autoencoder (VAE) inspired network learns a latent representation $\boldsymbol{z}$ by maximizing the ELBO:

$$\log p_{\boldsymbol{\theta}}(i_{\mathrm{iso}}) \geq \mathbb{E}_{\boldsymbol{z}\sim q_{\boldsymbol{\phi}}(\boldsymbol{z}|i_{\mathrm{bl}})}\left[\log p_{\boldsymbol{\theta}}(i_{\mathrm{iso}} \mid \boldsymbol{z})\right] - \mathbb{D}_{\mathrm{KL}}\left[q_{\boldsymbol{\phi}}(\boldsymbol{z} \mid i_{\mathrm{bl}}) \parallel p(\boldsymbol{z})\right]$$

  - $p_{\boldsymbol{\theta}}(i_{\mathrm{iso}})$ evidence
  - $p_{\boldsymbol{\theta}}(i_{\mathrm{iso}}|\boldsymbol{z})$ likelihood (decoder)
  - $q_{\boldsymbol{\phi}}(\boldsymbol{z}|i_{\mathrm{bl}})$ posterior (encoder)
  - $p(\boldsymbol{z})$ latent prior

- First term enforces reconstruction fidelity; KL term regularizes the latent space.

Deblender Network



- Variational Autoencoder (VAE) inspired network learns a latent representation $\boldsymbol{z}$ by maximizing the ELBO:

$$\log p_{\boldsymbol{\theta}}(i_{\mathrm{iso}}) \geq \mathbb{E}_{\boldsymbol{z}\sim q_{\boldsymbol{\phi}}(\boldsymbol{z}|i_{\mathrm{bl}})}\left[\log p_{\boldsymbol{\theta}}(i_{\mathrm{iso}} \mid \boldsymbol{z})\right] - \mathbb{D}_{\mathrm{KL}}\left[q_{\boldsymbol{\phi}}(\boldsymbol{z} \mid i_{\mathrm{bl}}) \parallel p(\boldsymbol{z})\right]$$

  - $p_{\boldsymbol{\theta}}(i_{\mathrm{iso}})$ evidence
  - $p_{\boldsymbol{\theta}}(i_{\mathrm{iso}}|\boldsymbol{z})$ likelihood (decoder)
  - $q_{\boldsymbol{\phi}}(\boldsymbol{z}|i_{\mathrm{bl}})$ posterior (encoder)
  - $p(\boldsymbol{z})$ latent prior

- First term enforces reconstruction fidelity; KL term regularizes the latent space.
- Multi-Head Attention (MHA): captures long-range dependencies; improves separation of overlapping sources.

Deblending Examples

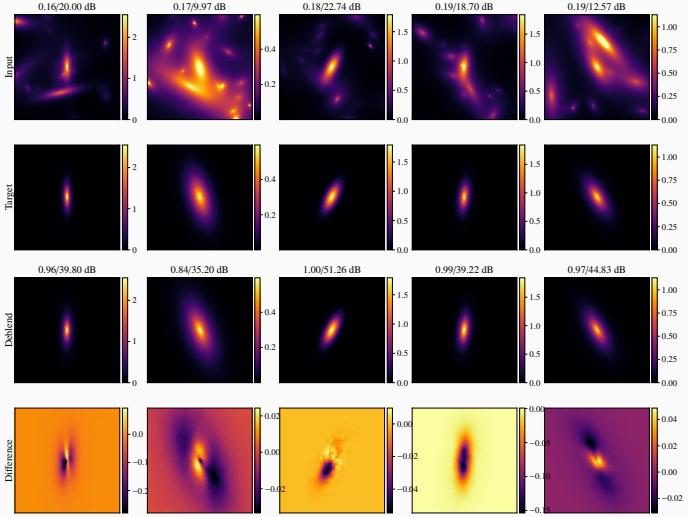
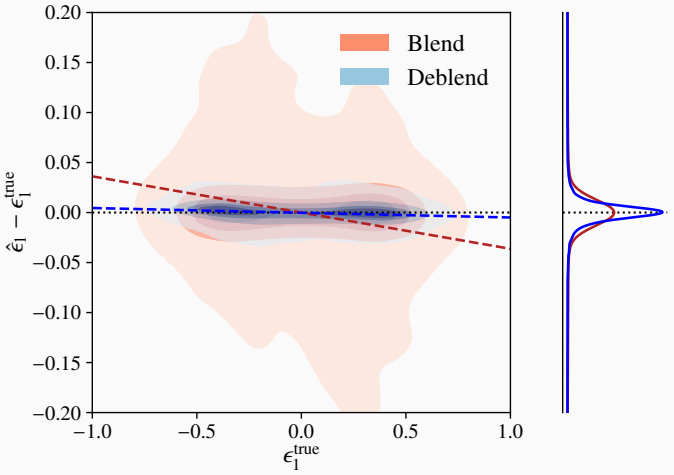


Figure 3: Trained/tested on true images: no PSF effect or noise

Shape Measurement: Blended vs Deblended



Blended:

$m_1 = (-36.2 \pm 2.2) \times 10^{-3}$

$c_1 = (-7.1 \pm 7.7) \times 10^{-4}$

Deblended:

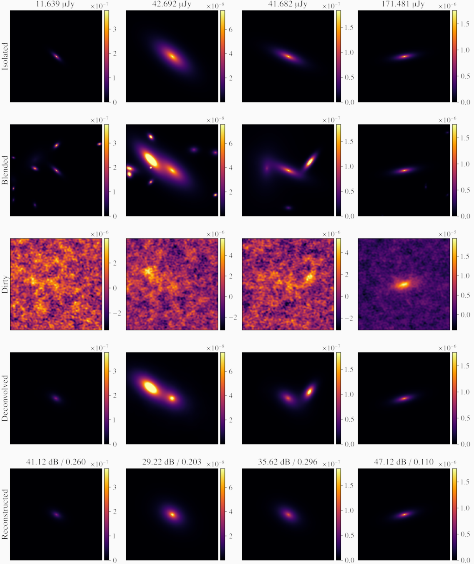
$m_1 = (-4.7 \pm 0.4) \times 10^{-3}$

$c_1 = (6.9 \pm 1.3) \times 10^{-4}$

Deblending significantly reduces multiplicative bias.

Figure 4: Shape measurement comparison for blended (red) and deblended (blue) stamps using GalSim.

Reconstruction Examples (Wide Field)



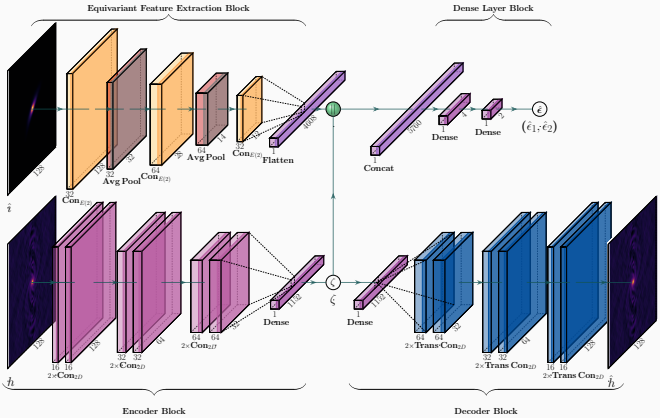
$$\mathrm{PSNR} = 10 \log_{10} \frac{\max(i_{\mathrm{iso}})^2}{\langle (\hat{i}_{\mathrm{iso}} - i_{\mathrm{iso}})^2 \rangle}$$

Ellipticity error: $\Delta\epsilon = \|\hat{\boldsymbol{\epsilon}} - \boldsymbol{\epsilon}^{\mathrm{true}}\|_2$

Feature Extraction

$E(2)$-equivariant CNN using Steerable kernels:

$$k(x|w) = \sum_{\ell=1}^{8} w_\ell(r)\, Y_\ell(\alpha), \quad Y_\ell(\alpha) = e^{i\ell\alpha}$$



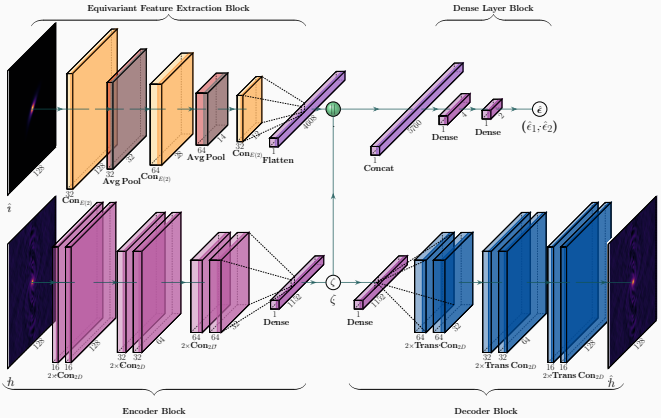
Feature Extraction

$E(2)$-equivariant CNN using Steerable kernels:

$$k(x|w) = \sum_{\ell=1}^{8} w_\ell(r)\, Y_\ell(\alpha), \quad Y_\ell(\alpha) = e^{i\ell\alpha}$$

PSF Encoding

PSF compressed using a pre-trained CNN autoencoder



Feature Extraction

$E(2)$-equivariant CNN using Steerable kernels:

$$k(x|w) = \sum_{\ell=1}^{8} w_\ell(r)\, Y_\ell(\alpha), \quad Y_\ell(\alpha) = e^{i\ell\alpha}$$

PSF Encoding

PSF compressed using a pre-trained CNN autoencoder

Shape Prediction

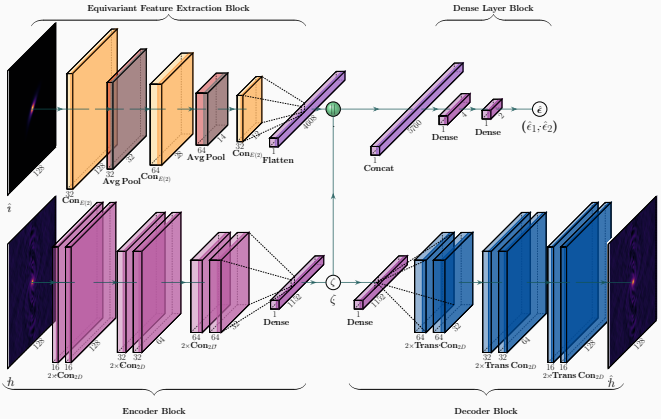
Features + Encoded PSF

⇓

Dense layers

⇓

$\hat{\epsilon}$



Validation on Isolated Galaxies

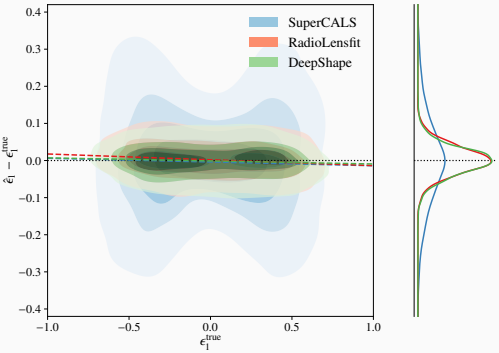
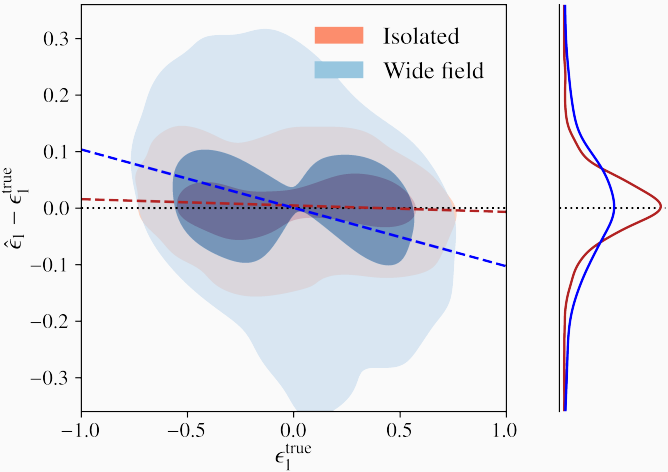


Figure 5: Shape residuals from three methods: DeepShape, SuperCALS, and RadioLensfit. Contours via KDE; dashed lines show best-fit trends.

- **SuperCALS** [Harrison et al., 2020]: An image-domain approach that uses MS-CLEAN for deconvolution. Biases are estimated by injecting known model sources and then used to calibrate the recovered shapes.
- **RadioLensfit** [Rivi et al., 2016]: A visibility-domain parametric model-fitting method based on an exponential brightness profile. Ellipticities are obtained by marginalising the posterior over nuisance parameters.

DeepShape: Wide-Field vs Isolated Performance



Bias increases from isolated → wide-field, but remains comparable to literature values.

Key Points:

Bias & Timing

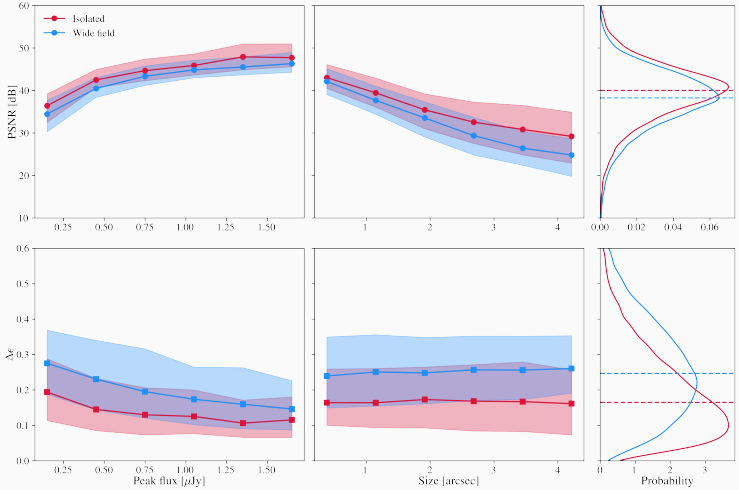
	Bias	Timing [s]
DeepShape II	∼10%	0.85
RadioLensfit	∼9%	10
Requirement	∼1%	–

Trade-off: Accuracy slightly worse, but massive gain in speed.

To reach required bias levels with DeepShape, stringent quality cuts are needed.

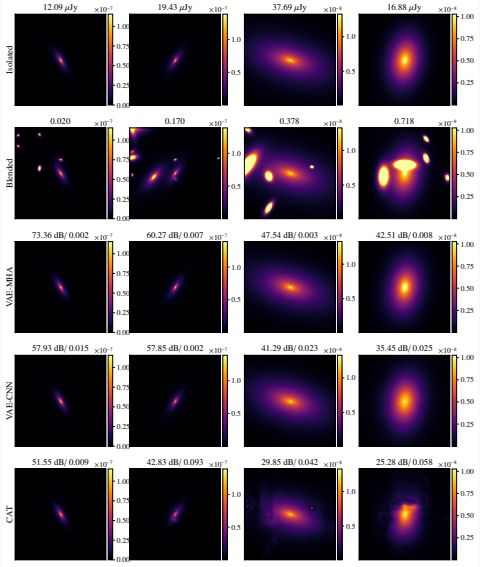
Thank you for your time

Reconstruction Fidelity vs Galaxy Properties

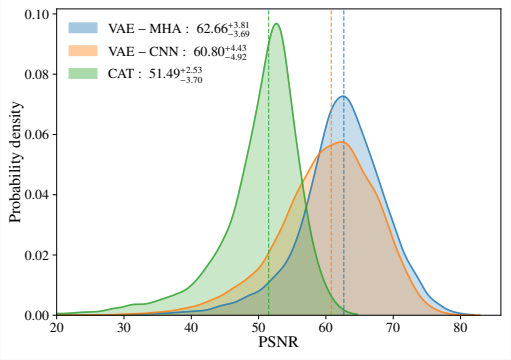


Top row: PSNR, **Bottom row:** ellipticity error $\Delta\epsilon$. **Columns:** median in bins of peak flux, median in bins of galaxy size (shaded = IQR), KDE with median line. Shapes measured using GalSim.

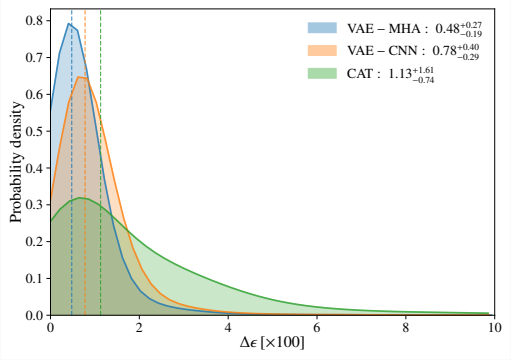
Comparison with other deblender networks



Comparison II



(a) *PSNR*



(b) *Ellipticity error* $\Delta\boldsymbol{\epsilon} = \|\hat{\boldsymbol{\epsilon}} - \boldsymbol{\epsilon}^{\mathrm{true}}\|_2$

Figure 6: Comparison of reconstruction metrics for the three models.

SuperCLASS Data

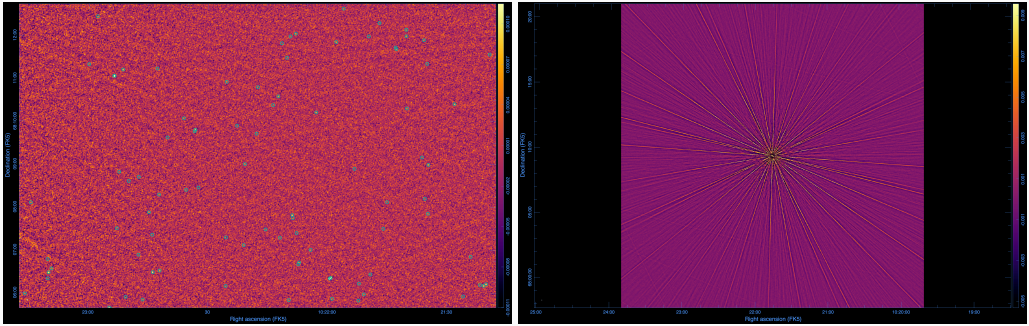
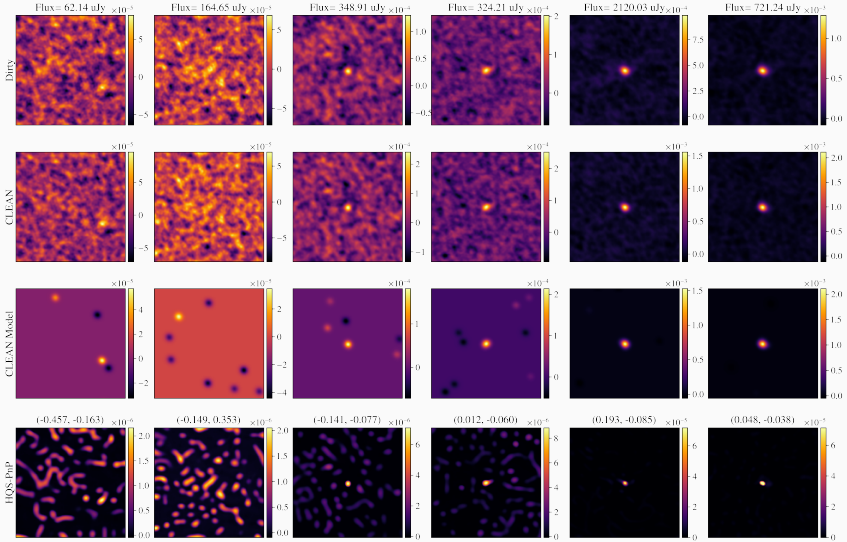
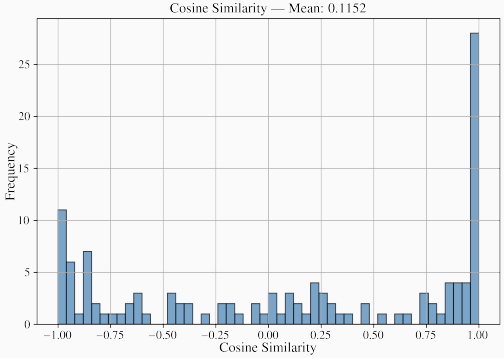


Figure 7: Example P28 pointing from SuperCLASS (VLA+e-MERLIN). Left: Zoomed-in CLEAN image with identified sources. Right: Corresponding PSF

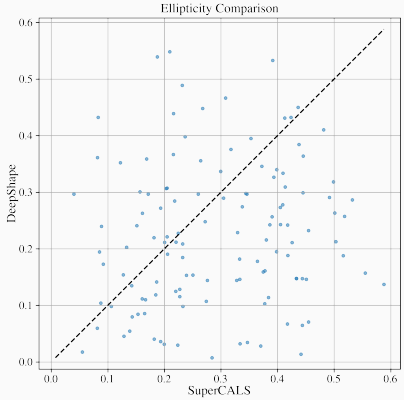
Reconstruction



Shape Measurement



Cosine similarity between DeepShape and SuperCALS measurements



Compaison of ellipticty measurements: $|\epsilon|$

References I

Baugh, C. M., Gonzalez-Perez, V., Lagos, C. D. P., Lacey, C. G., Helly, J. C., Jenkins, A., Frenk, C. S., Benson, A. J., Bower, R. G., and Cole, S. (2019).
Galaxy formation in the Planck Millennium: the atomic hydrogen content of dark matter haloes.
, 483(4):4922–4937.

Bonaldi, A., Bonato, M., Galluzzi, V., Harrison, I., Massardi, M., Kay, S., De Zotti, G., and Brown, M. L. (2018).
The tiered radio extragalactic continuum simulation (t-recs).
MNRAS, 482(1):2–19.

Chang, T.-C., Refregier, A., and Helfand, D. J. (2004).
Weak lensing by large-scale structure with the FIRST radio survey.
The Astrophysical Journal, 617(2):794–810.

References II

Harrison, I., Brown, M. L., Tunbridge, B., Thomas, D. B., Hillier, T., Thomson, A. P., Whittaker, L., Abdalla, F. B., Battye, R. A., Bonaldi, A., Camera, S., Casey, C. M., Demetroullas, C., Hales, C. A., Jackson, N. J., Kay, S. T., Manning, S. M., Peters, A., Riseley, C. J., and Watson, R. A. (2020).
Superclass – iii. weak lensing from radio and optical observations in data release 1.
MNRAS, 495(2):1737–1759.

Kogan, L. and Greisen, E. W. (2009).
Faceted imaging in aips.
AIPS memo, 840.

Rivi, M., Miller, L., Makhathini, S., and Abdalla, F. B. (2016).
Radio weak lensing shear measurement in the visibility domain – i. methodology.
MNRAS, 463(2):1881–1890.

References III

Tripathi, P., Ferrari, A., and Prunet, S. (2026).
Deepshape ii: Source separation for radio weak-lensing measurements using deep learning.
submitted.

Tripathi, P., Wang, S., Prunet, S., and Ferrari, A. (2025).
Deepshape: Radio weak-lensing shear measurements using deep learning.
A&A, 696:A216.

Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., and Timofte, R. (2021).
Plug-and-Play Image Restoration with Deep Denoiser Prior.
arXiv:2008.13751 [cs, eess].