

FluxFlow.

Conservative Flow-Matching for Astronomical Image Super-Resolution

Generative detail. Conservative flux.

Shuhong Liu^{1,2*†} · Xining Ge^{2*} · Quanfeng Xu^{3*} · Ziteng Cui^{1,2} · Liuzhuozheng Li¹ · Gengjia Chang²
Jun Liu² · Ziying Gu¹ · Dong Li² · Xuangeng Chu^{1,2} · Lin Gu⁴ · Tatsuya Harada^{1,5}

¹The University of Tokyo ²I2WM ³Shanghai Astronomical Observatory ⁴Tohoku University ⁵RIKEN

* equal contribution † corresponding author

PROBLEM

One sky——Two instruments

Task: Reconstruct detail from DESI to HST

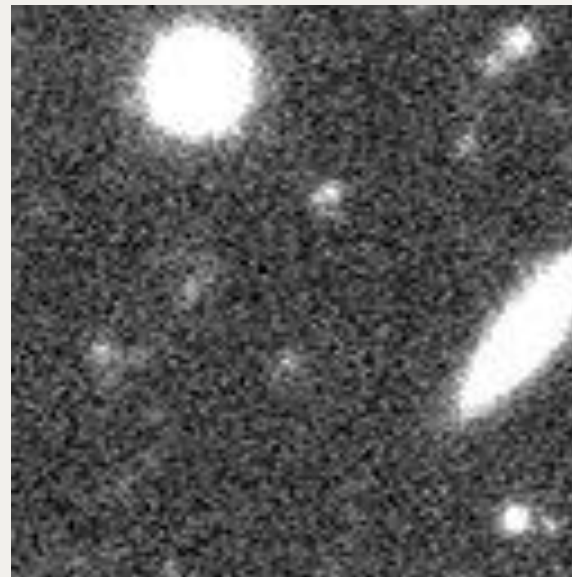
DESI: wide sky

Ground-based images
atmospheric seeing blurs fine detail

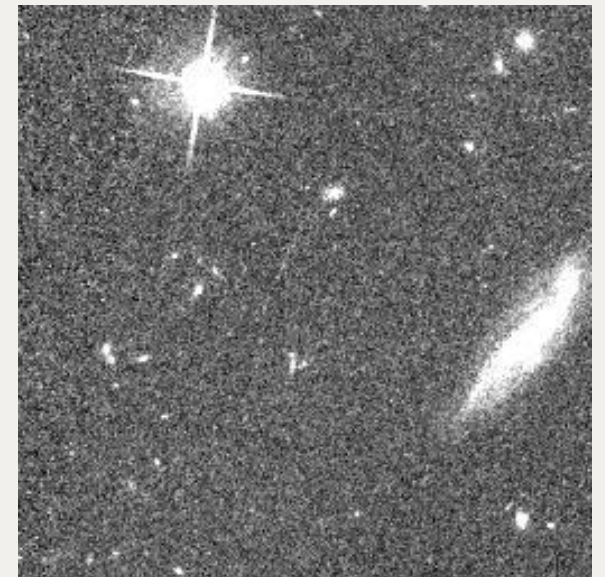
HST: sharp detail

Space-based images;
sharper, but covering smaller fields

Different resolution: X2 / X4

DESI

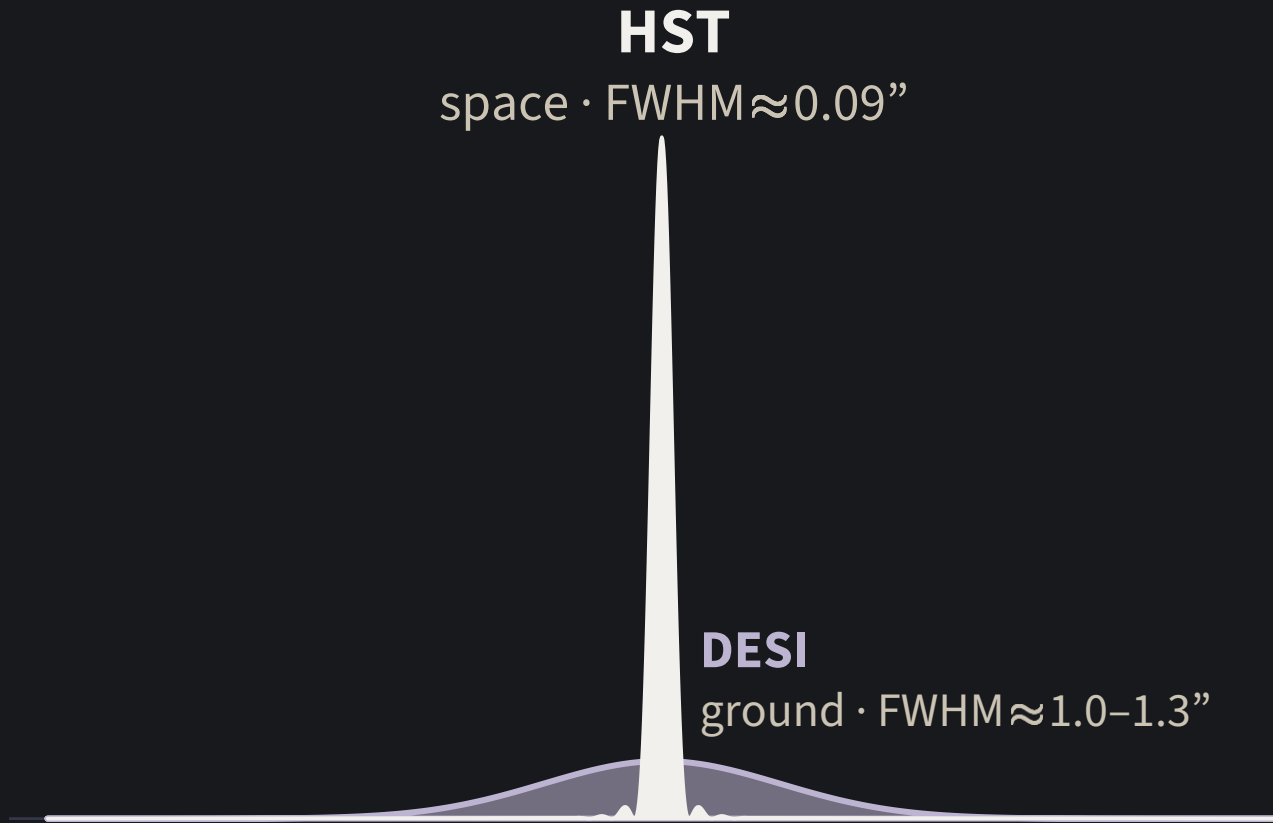
DESI Legacy Imaging Surveys

HST

Hubble Space Telescope

PROBLEM

From the ground, every star is 11–14× wider



11–14×

wider blur from the ground

Upsampling alone cannot solve it!

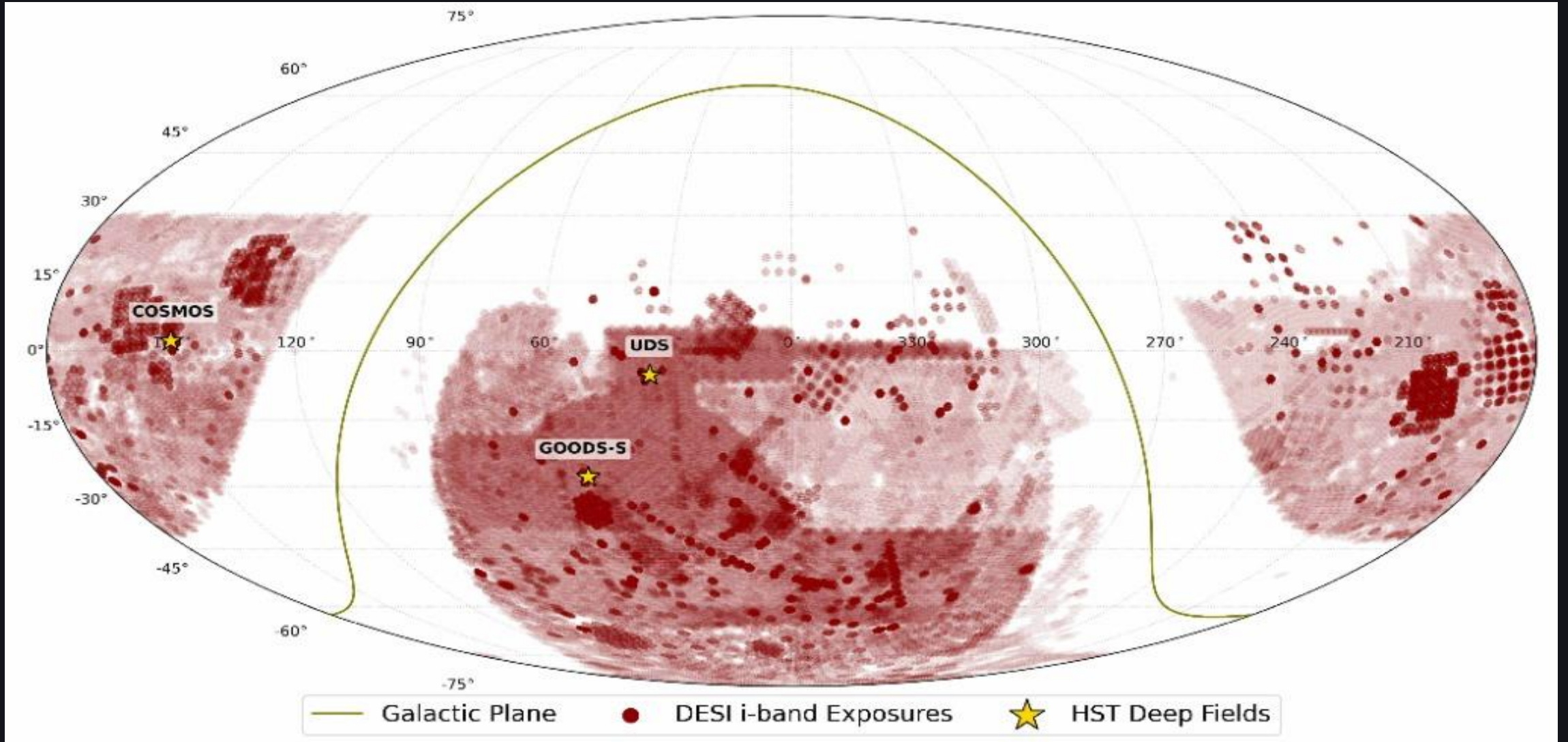
HST sees a tiny patch of sky

DESI surveys thousands of square degrees

Same total light,
drawn to scale: spread $\sim 12\times$ wider $\sim 12\times$ fainter at the peak.

DATA

Same sky, two telescopes——One pair

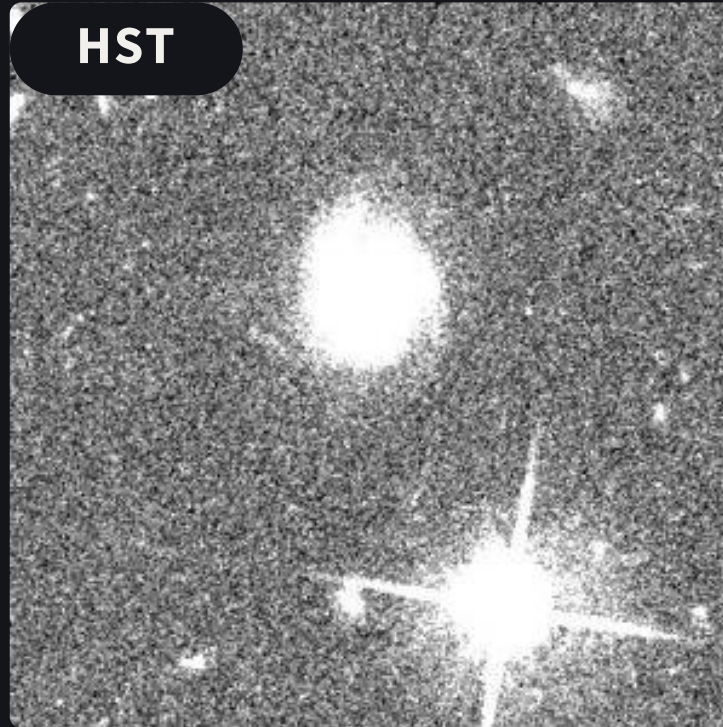


CHALLENGE

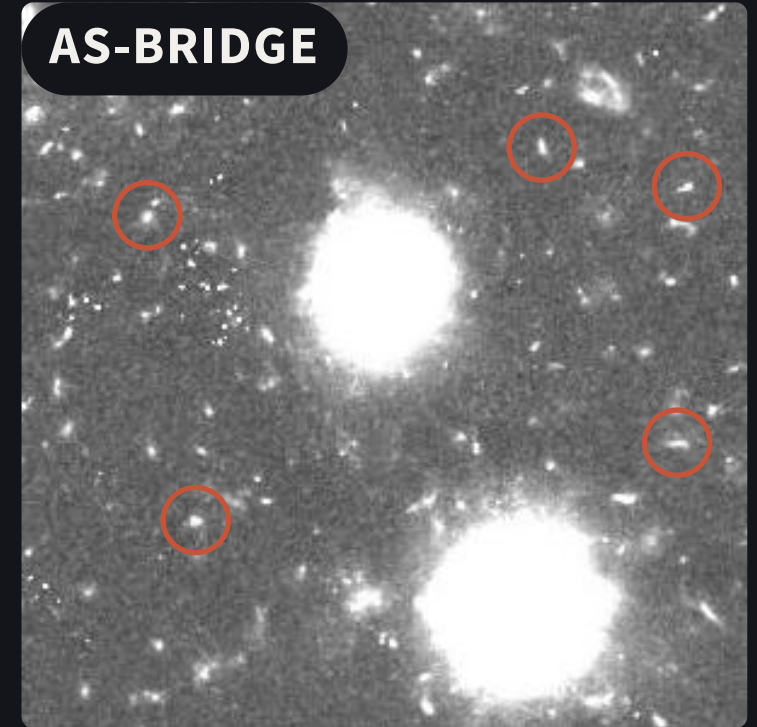
Too smooth, or too creative

**Too smooth**

Fine structure is smoothed away

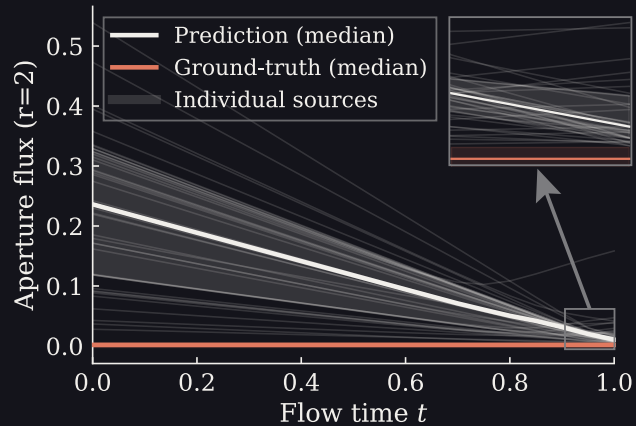
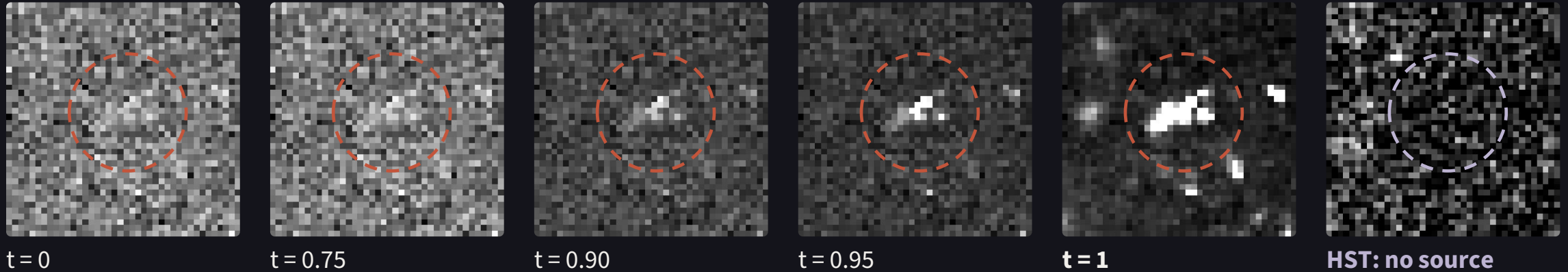
**What is really there**

The real sky, observed from space

**Too creative**

Sharp, but adds sources HST doesn't see

Fake sources grow out of initial noise



A faint noise bump is sharpened late in sampling into a compact source, ending brighter than Hubble at that spot.

Classifier-free guidance doesn't stop it; source masks predicted from $x_{t_{\text{reach}}}$ only 11.3% mIoU.

So FluxFlow checks every step against the observation.

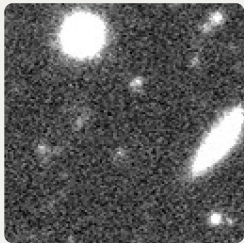
METHOD

FluxFlow in four parts

1 TRAINING



HST target



DESI input

DESI condition c
BicubicUp(y)

Flow state x_t

1 Weighted flow matching loss

Channel
concat $[x_t, c]$

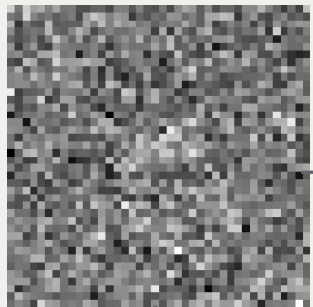
4 Conditional U-Net
velocity field · pixel
space

$t \rightarrow$ FiLM conditioning

v_θ

velocity field
pixel space · no VAE

SAMPLING



Noise

EVERY STEP

Euler step
with v_θ

**2 Measurement
check**

compare with the
DESI image

**3 Wiener
correction**

back-project without
blur

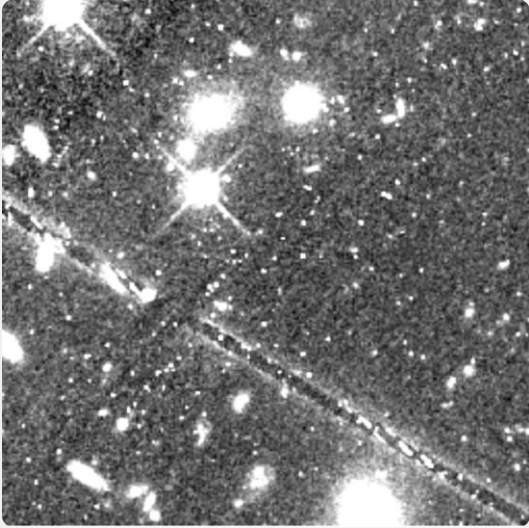
repeat for $N = 10$ steps



FluxFlow

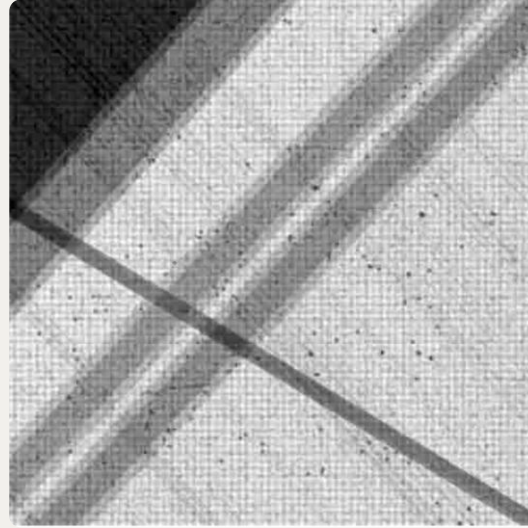
Without **2** and **3**, initial noise can grow into false sources.

Don't learn HST's cosmic rays



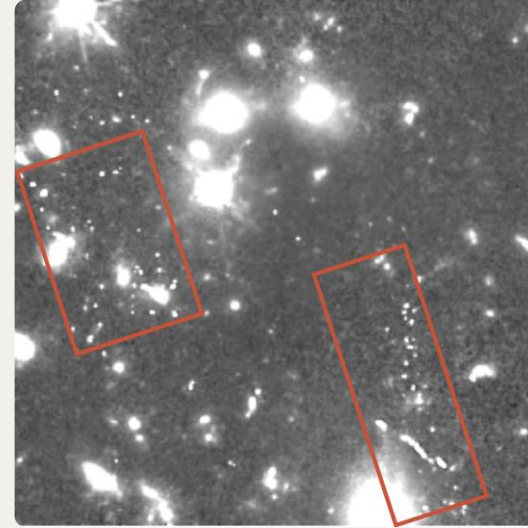
HST target

cosmic-ray trail and chip gap



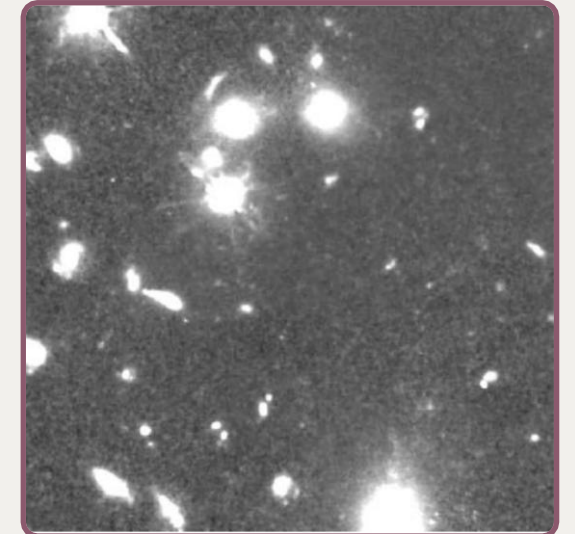
Weight map $\sqrt{W}$

defects get low weight



Trained without $\sqrt{W}$

copies the defects



FluxFlow

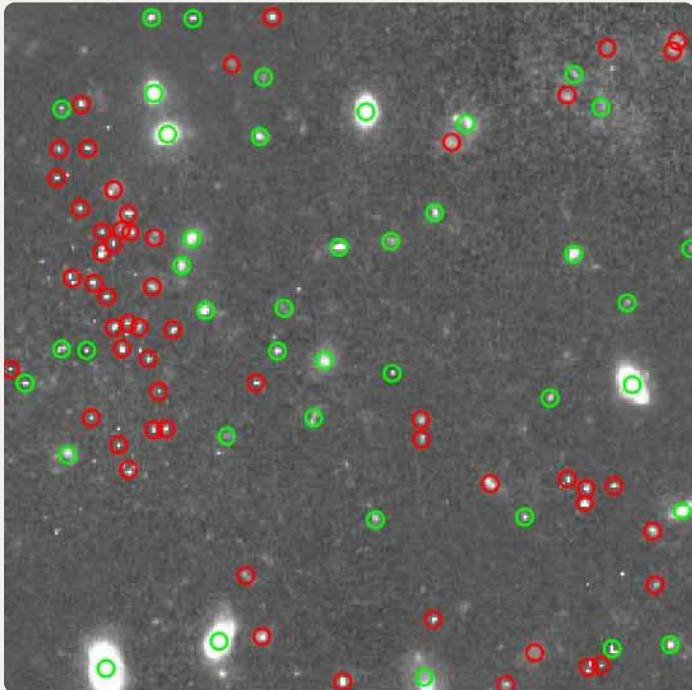
clean reconstruction

Flow-matching loss, weighted per pixel by $\sqrt{W} \cdot (1 + M)$

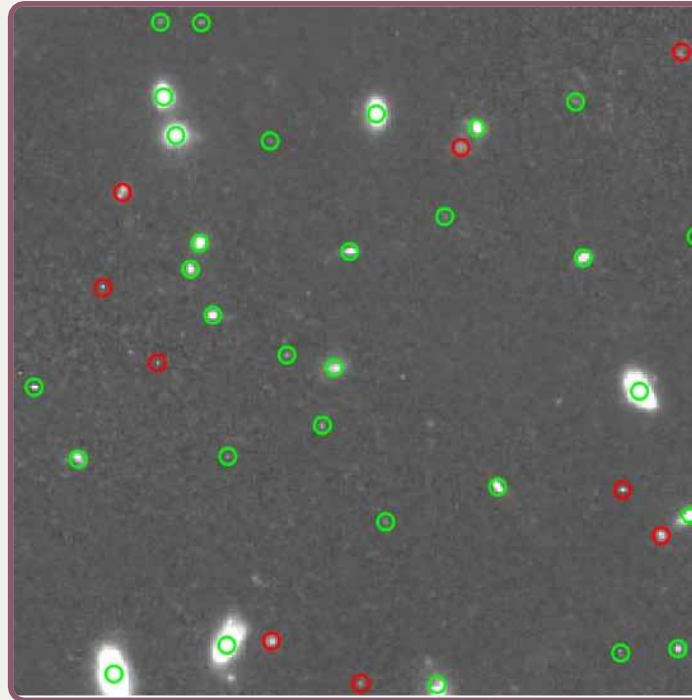
$\sqrt{W}$ HST inverse-variance map: trust

M source mask: source pixels count twice

Check every step against the DESI image



Without the check



FluxFlow



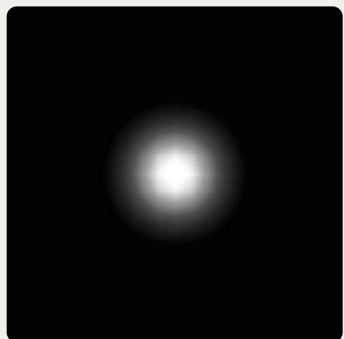
0.602 → 0.738

detection precision

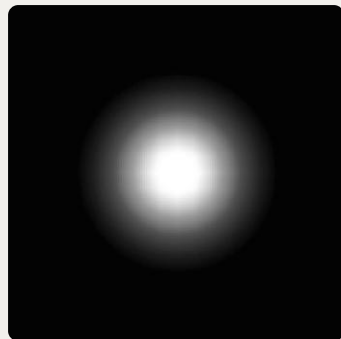
training-free

no gradient through the network

Back-project with Wiener, not the PSF

**HST truth**

FWHM 4.71 px

**Plain correction**

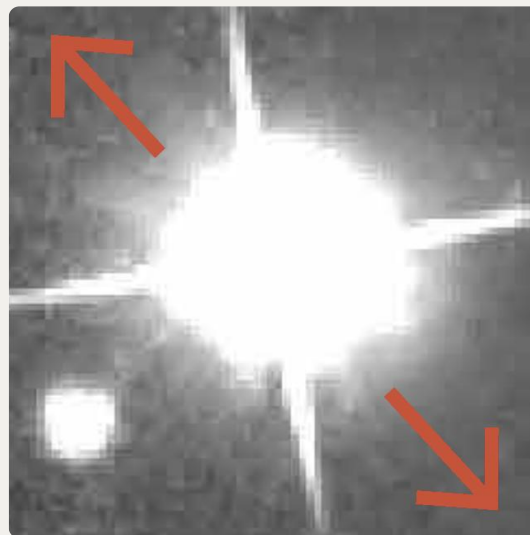
FWHM 6.78 px

**Wiener**

FWHM 4.85 px

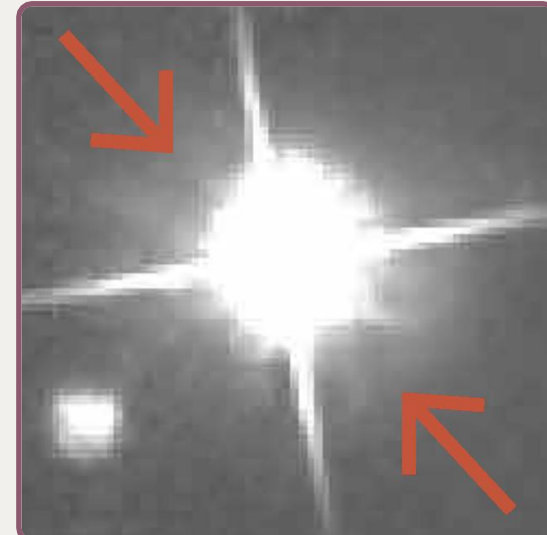
$$\mathcal{W}(k) = \frac{H^*(k)}{|H(k)|^2 + \lambda^{-1}}$$

Replaces the PSF in the back-projection, so repeated corrections stop widening sources.

**Plain correction**

PSNR

Flux-L1

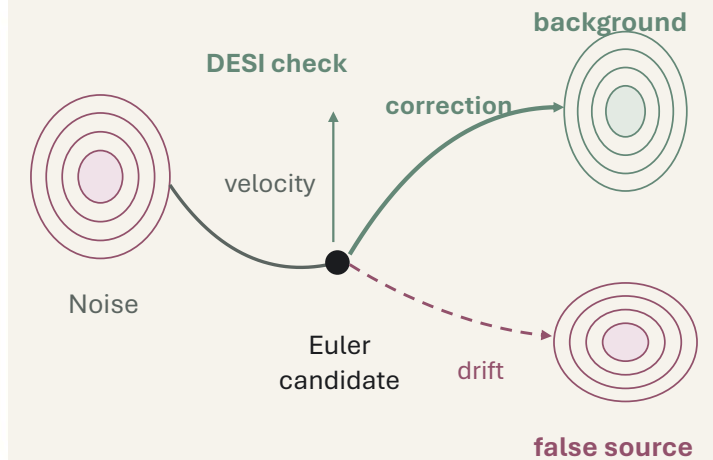
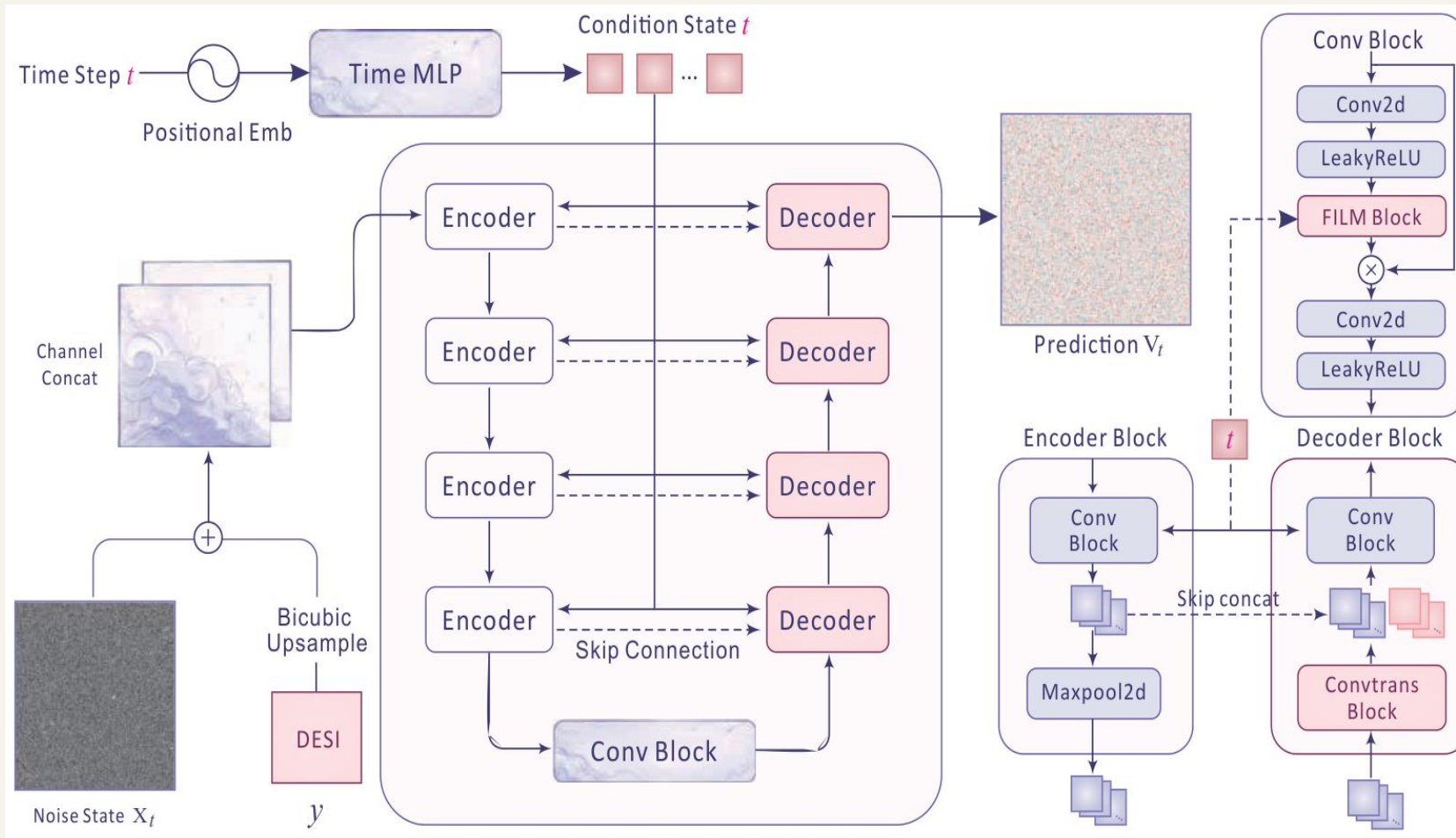
**FluxFlow**

30.79 → 31.23

3.031 → 2.959

The conditional U-Net in FluxFlow

DESI conditions the network at inference; HST provides the training target.



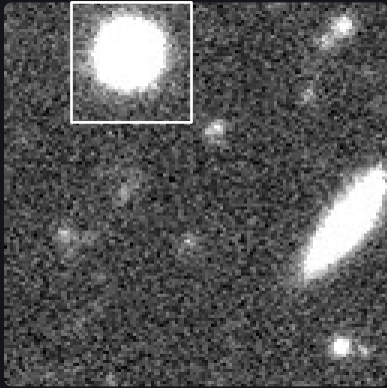
01 Network input and output

Upsampled DESI is channel-concatenated with the flow state. A four-level U-Net with time-conditioned FiLM predicts pixel-space velocity.

02 Ten-step MC-FS

Each Euler candidate is checked in DESI space. Wiener correction maps the residual back to the high-resolution grid.

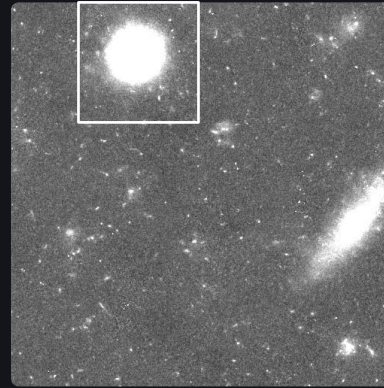
Sharp detail, far fewer invented sources



DESI input ground



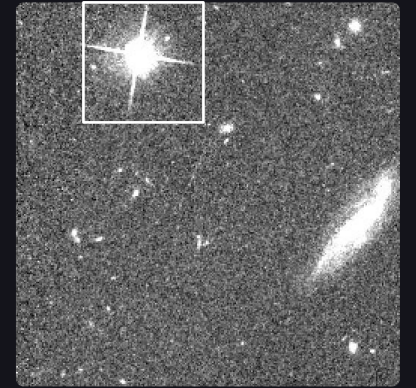
HAT regression



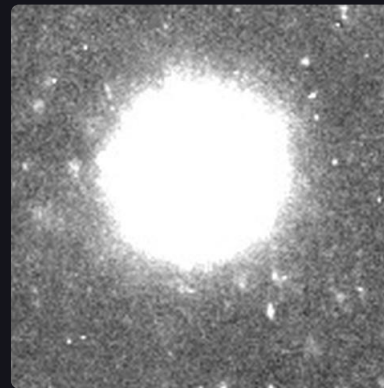
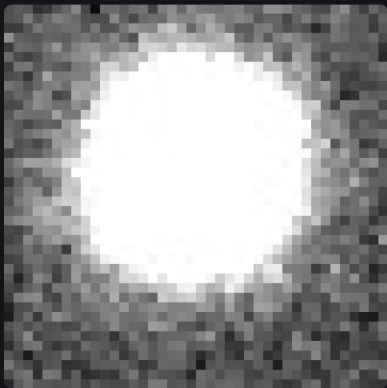
AS-Bridge generative



FluxFlow ours

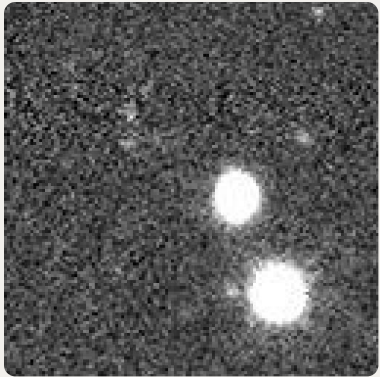


HST reference

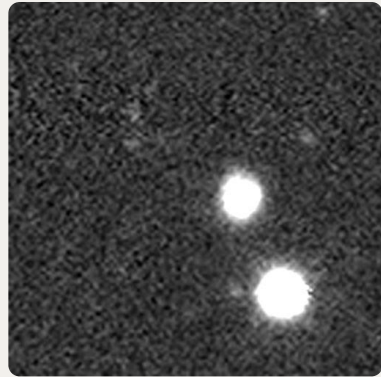


Real model outputs on the same field; lower row zooms on the boxed bright star.

At ×4, against all seven baselines



DESI input input



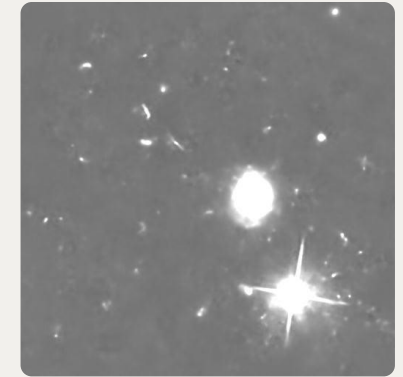
Bicubic interpolation



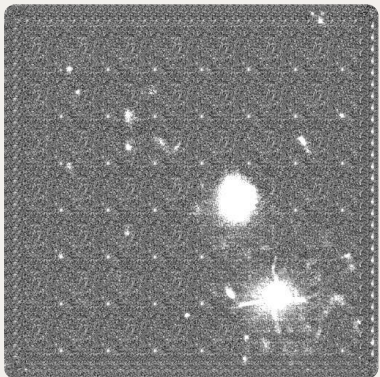
SwinIR regression



HAT regression



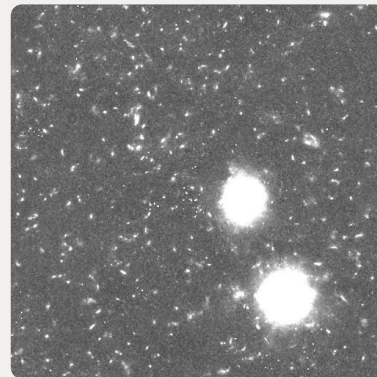
FISR regression



cGAN generative



GD-Net generative



AS-Bridge generative



FluxFlow ours



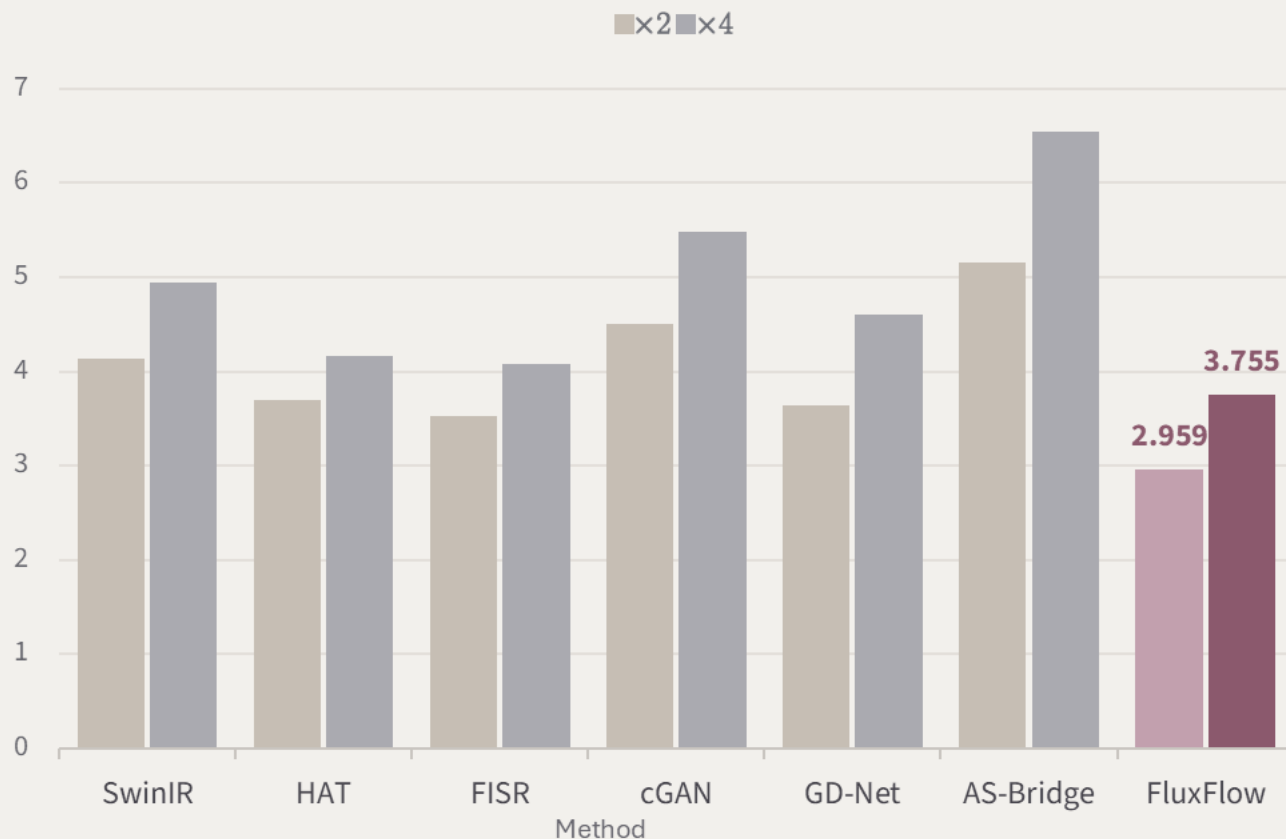
HST reference

Same field and framing for every method (×4: 128 px DESI→512 px).

RESULTS

Best flux error, PSNR within 0.1 dB

Flux-L1 on source regions lower is better



Bicubic omitted for scale (11.25 at $\times 2$, 18.08 at $\times 4$). All numbers from Table 2 of the paper.

-16%

Flux-L1 vs. the best baseline at $\times 2$

2.959 vs. 3.526 (FISR)

-8%

Flux-L1 vs. the best baseline at $\times 4$

3.755 vs. 4.082 (FISR)

0.714

best SSIM at $\times 2$

tied best at $\times 4$ (0.570)

+2.3 dB

PSNR above every generative baseline

31.23 vs. 28.88 (cGAN) at $\times 2$

ABLATION

Removing any part lowers PSNR and F1

Variant	PSNR $\uparrow$	SSIM $\uparrow$	Flux-L1 $\downarrow$	Prec. $\uparrow$	Rec. $\uparrow$	F1 $\uparrow$
Latent VAE instead of pixels	28.90	0.593	5.075	0.859	0.211	0.340
Bridge instead of noise start	30.77	0.663	3.206	0.702	0.393	0.502
without $\sqrt{w}$ weighting	30.17	0.685	2.814	0.688	0.483	0.527
without source weighting	31.16	0.711	3.098	0.729	0.464	0.538
without measurement check	30.78	0.698	3.025	0.602	0.553	0.526
without Wiener	30.79	0.699	3.031	0.708	0.448	0.507
FluxFlow (full)	31.23	0.714	2.959	0.738	0.458	0.541

×2 test set. Detection: SExtractor on each prediction vs. Hubble sources.

Table 3 of the paper.

Pixels beat a latent VAE

F1 0.340 $\rightarrow$ 0.541: compression erases faint sources.

The check buys precision

Precision 0.602 $\rightarrow$ 0.738, recall 0.553 $\rightarrow$ 0.458; F1 rises to 0.541.

$\sqrt{w}$ is a deliberate trade-off

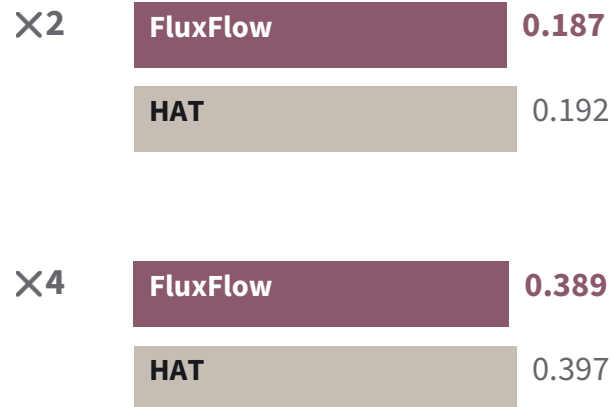
Without it Flux-L1 looks better (2.814) because the model fits detector artefacts. We keep clean structure.

SCIENCE CHECK

Best per-source magnitudes; shapes slightly behind

Magnitude error

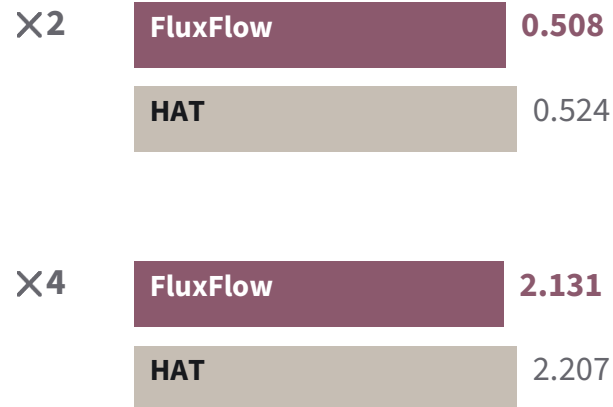
feeds photometric catalogues



Best at X2 and X4

Mass-proxy error

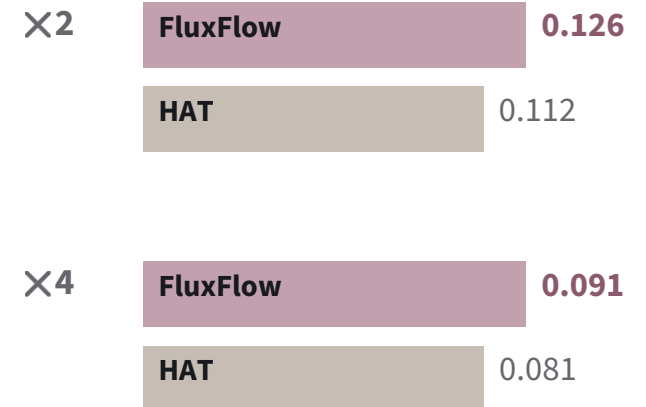
luminosity $\times$ fixed M/L



Best at X2 and X4

Shear error

feeds weak-lensing shapes



Smooth baselines keep shapes slightly better

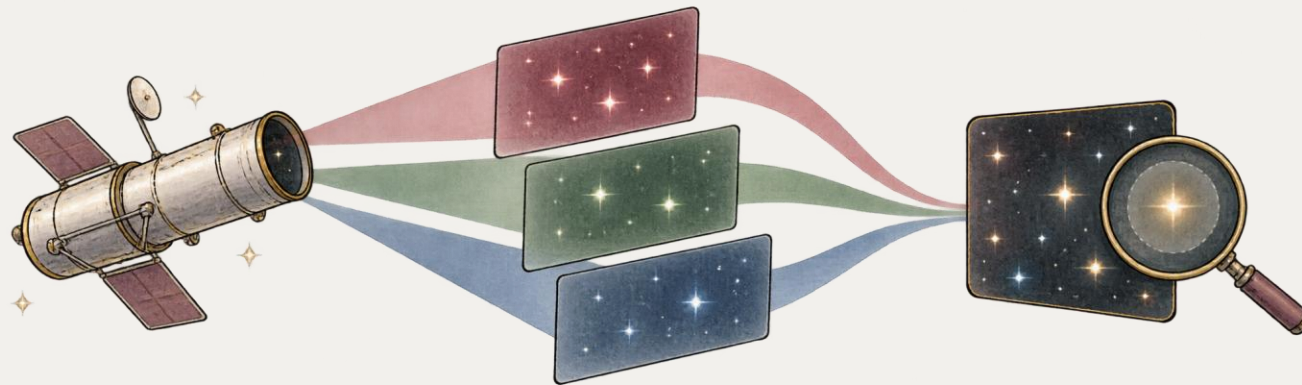
Suppressed, not eliminated

Limitations

- Hallucinations are suppressed, not eliminated.
- Scientific accuracy still falls short of what precision astronomy demands.

Next

- Multi-band joint reconstruction.
- Uncertainty-calibrated sampling to improve the SNR.



FluxFlow

Faithful astronomical super-resolution

Ground-based DESI → space-based HST

Thank you!

github.com/I2WM/FluxFlow

I2WM · University of Tokyo · RIKEN AIP



arxiv.org/abs/2605.03749