Deep Learning-based Deconvolution for Astronomical Surveys

Utsav Akhaury

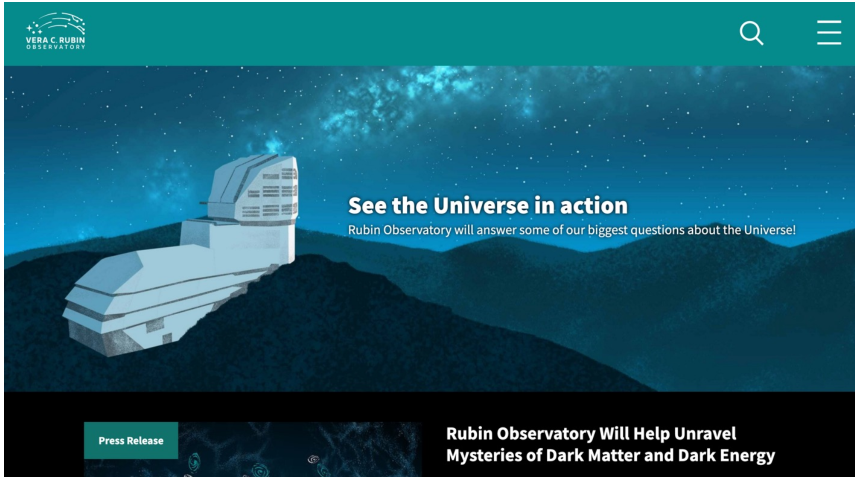
Pascale Jablonka, Jean-Luc Starck, Frédéric Courbin

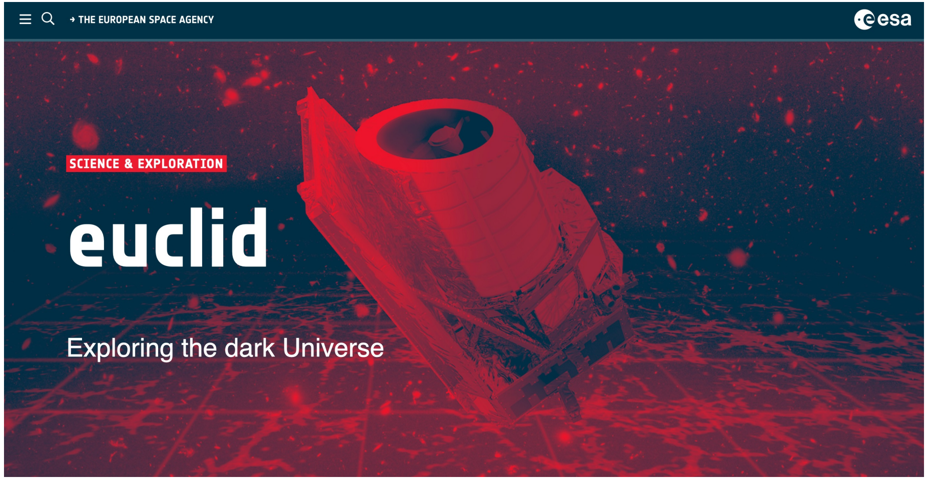
EPFL cea FORTH FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS

VERA C. RUBIN
OBSERVATORY

esa
European Space Agency

esa
euclid





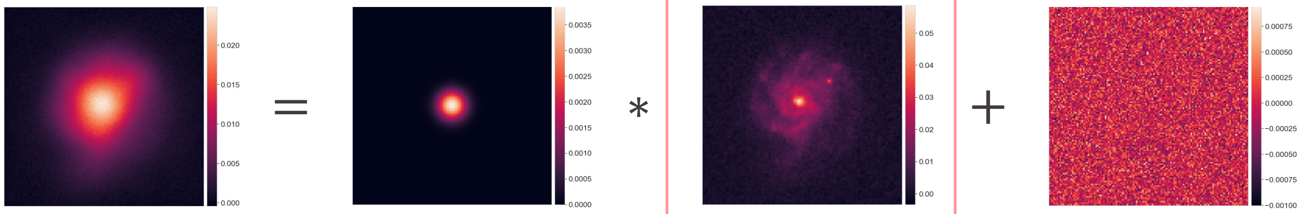
Introduction & Motivation

- Need to be better understand phenomena in galaxy formation & evolution:
  - mechanism of inflow/outflow
  - position of galaxies in the cosmic web and their evolution
  - the importance of mergers
  - star formation and mass maps in high resolution
  - star formation rate
- Multi-wavelength images over a wide field of view only available from **ground-based** telescopes
  - **Degraded** due to atmospheric blurring and instrumental optics
- Access to clear **high-resolution** images
  - Increasing need to develop **fast** and **accurate** deconvolution algorithms that generalize well

The Deconvolution Problem

Model

$$\mathbf{y} = \mathbf{h} * \mathbf{x}_t + \eta$$



Observed Noisy Image | PSF | Ground Truth | Additive Noise

- $\mathbf{y} \in \mathbb{R}^{n \times n}$ — Observed Noisy Image
- $\mathbf{x}_t \in \mathbb{R}^{n \times n}$ — Ground Truth Image
- $\mathbf{h} \in \mathbb{R}^{n \times n}$ — PSF
- $\eta \in \mathbb{R}^{n \times n}$ — Additive Noise

Issues

- The equation is **ill-conditioned** and **ill-posed**
- Problem could be handled by regularization

The Deconvolution Step

Loss Function

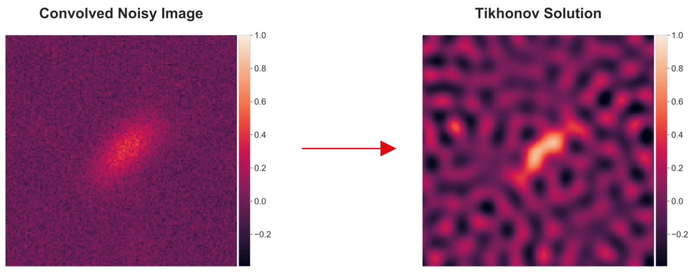
$$L(\mathbf{x}) = \frac{1}{2\sigma^2} \| \mathbf{Hx} - \mathbf{y} \|_2^2 + \lambda \| \mathbf{\Gamma x} \|_2^2$$

Tikhonov Deconvolution

$$\hat{\mathbf{x}} = (\mathbf{H}^\top \mathbf{H} + \lambda\, \mathbf{\Gamma}^\top \mathbf{\Gamma})^{-1}\, \mathbf{H}^\top \mathbf{y}$$

- $\sigma \in \mathbb{R}$ — Noise standard deviation
- $\mathbf{\Gamma} \in \mathbb{R}^{n^2 \times n^2}$ — Linear Tikhonov filter set to a Laplacian high-pass filter (to penalize high frequencies)
- $\lambda \in \mathbb{R}_+$ — Regularization weight
- $\mathbf{H} \in \mathbb{R}^{n^2 \times n^2}$ — Block circulant matrix associated with the convolution operator $\mathbf{h}$

$\mathbf{y}$ $\hat{\mathbf{x}}$

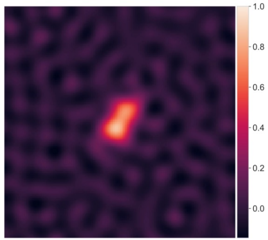


Tikhonov output contains correlated noise

The Denoising Step

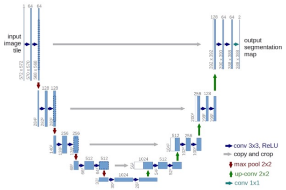
The training is aimed to make the network learn the following mapping while minimizing a suitable loss function:

- Tikhonov output $\hat{\mathbf{x}}$ $\longrightarrow$ ground truth image $\mathbf{x}_t$

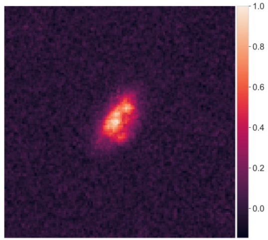


Tikhonov Solution

$\hat{\mathbf{x}}$



- Denoiser

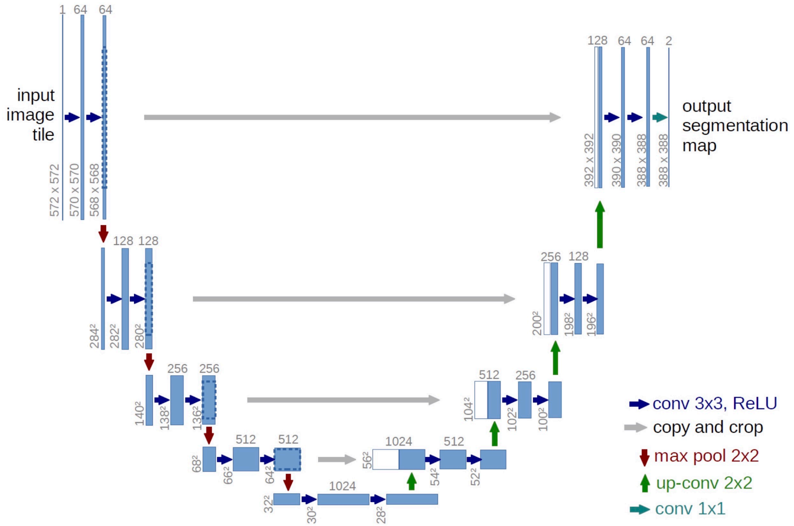


HST Target

$\mathbf{x}_t$

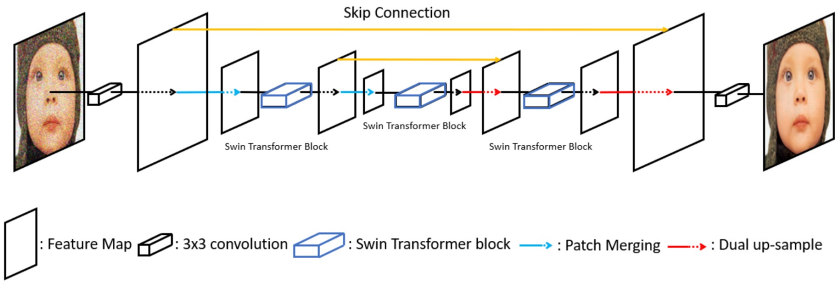
U-net

- Originally developed for biomedical image segmentation
- Relevant to many other imaging problems, like **denoising**
- U-nets consist of a multi-scale approach, allowing the signal to be analyzed at multiple resolutions



U-Net: Convolutional Networks for Biomedical Image Segmentation
Ronneberger et al, 2015

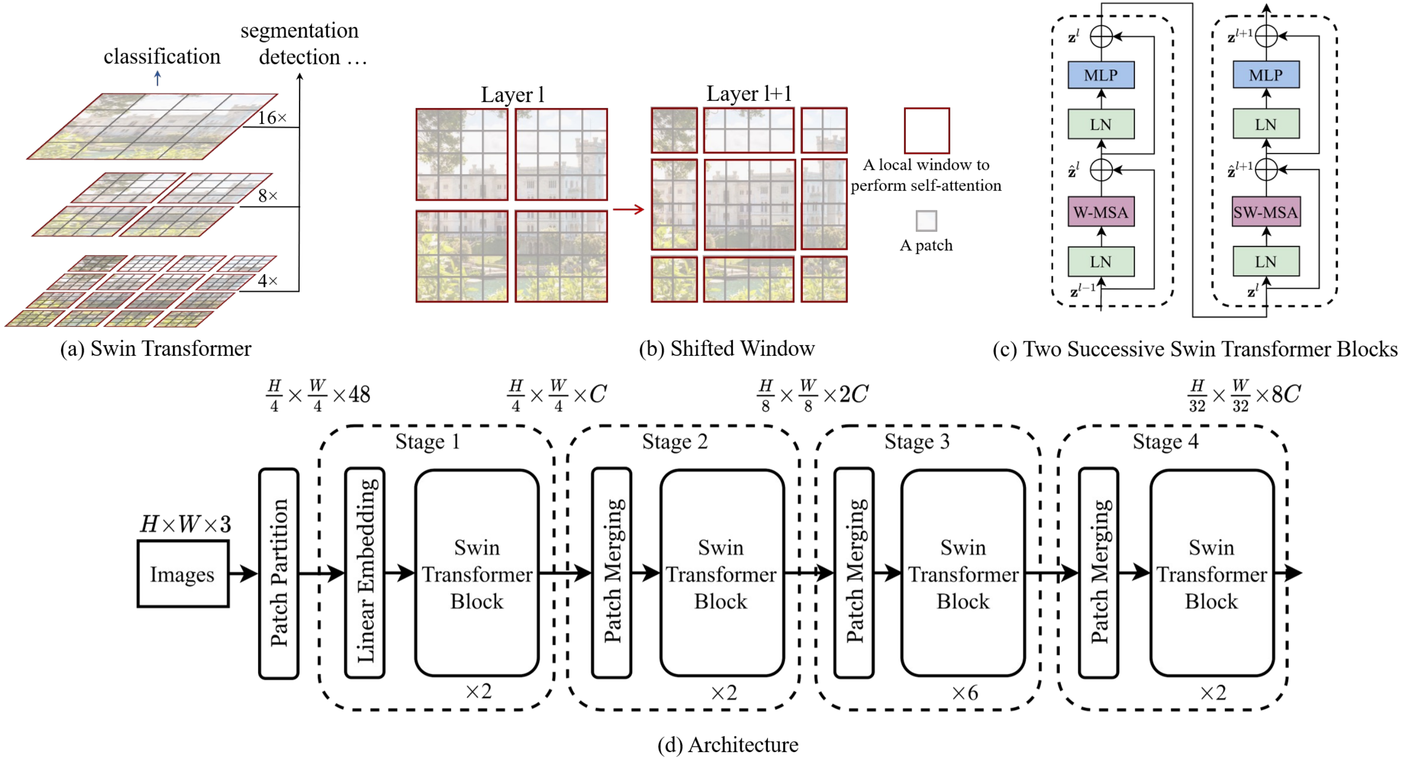
SUNet



- A Unet with **Swin Transformer** blocks incorporated in the architecture

SUNet: Swin Transformer UNet for Image Denoising,
Fan et al, 2022

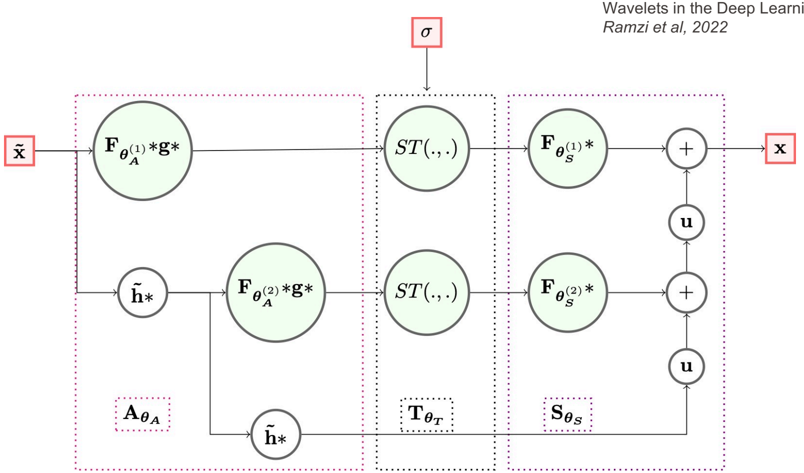
Swin Transformer



Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, *Liu et al, 2021*

Learnlet

Wavelets in the Deep Learning Era
Ramzi et al, 2022



$$\mathbf{f}_\theta(\tilde{\mathbf{x}}, \sigma) = \mathbf{S}_{\theta_s}\left(\mathbf{T}_{\theta_t}\left(\mathbf{A}_{\theta_a}(\tilde{\mathbf{x}}), \sigma\right)\right) \; : \; (\mathbb{R}^{n\times n} \times \Sigma) \; \rightarrow \; \mathbb{R}^{n\times n}$$

- The Learnlet decomposition (Ramzi et al., 2021) aims at learning a filter bank in a denoising setting with backpropagation and gradient descent
- Learnlets exploit the best of both of deep learning and classical algorithms –
  - uses gradient descent to improve the expressive power of wavelets
  - preserves some interesting wavelet properties like exact reconstruction
- m = 5 scales → 44,840 trainable parameters
  128×128

Σ	set of possible values for the noise standard deviation σ
m	Number of scales
$\theta = (\theta_s, \theta_t, \theta_a) \in \Theta_m$	a given set of parameters

EPFL

Dataset Generation & Training

Ground Truth Images

- CANDELS – Five different image mosaics (GOODS-N, GOODS-S, EGS, UDS, COSMOS)
- HST cutouts of 128 × 128 pixels from CANDELS in the *F606W* filter (*V*-band) centred at the object centroid

Barbara A. MIKULSKI ARCHIVE FOR SPACE TELESCOPES

SEARCH MENU

The Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey ("CANDELS")

Primary Investigator: Sandra Faber

Co-Primary Investigator: Harry Ferguson

HLSP Authors: Anton Koekemoer and Norman Grogin (CANDELS HST image mosaics for all fields), Audrey Galametz and Paola Santini (CANDELS-UDS catalogs), Yicheng Guo and Thomas Dahlen (CANDELS-GOODS-S catalogs). Hooshang Nayyeri (CANDELS-COSMOS catalogs), Mauro Stefanon (CANDELS-EGS catalogs), Guillermo Barro (CANDELS multi-band catalogs for all fields), Dritan Kodra and Brett Andrews (CANDELS v2 photo-z catalogs for all fields)

Released: 2011-01-12

Updated: 2022-10-05

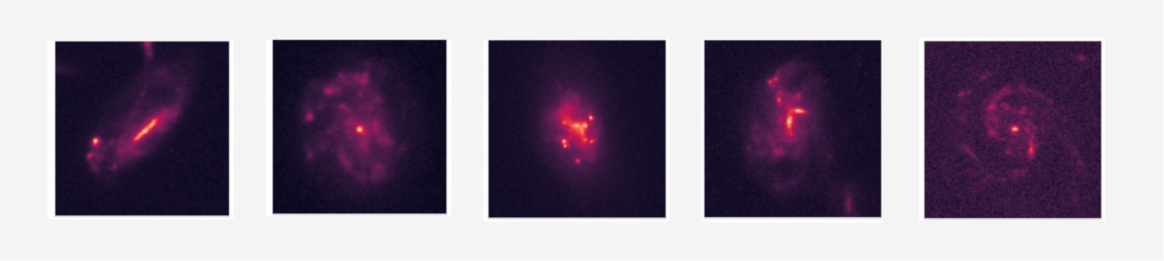
Primary Reference(s): Grogin et al. 2011, Koekemoer et al. 2011

DOI: 10.17909/T94S3X

EPFL

Filtering

- Selected good galaxy candidates and excluded point-sized objects using the following filtering criteria:
  - MAG_AUTO < 26 (AB magnitude in SExtractor "AUTO" aperture)
  - Flux_Radius$_{80}$ > 10 (80% enclosed flux radius in pixels)
  - FWHM > 10 (full width at half maximum in pixels)



Good Candidates

Rejected

Noisy Simulations

- Convolved ~25,000 ground-truth images with a Gaussian PSF having an FWHM of 15 pixels
- Added white Gaussian noise with a standard deviation σ_{noise} having a value such that the faintest object in our dataset has a peak SNR close to 1
- Train-Validation-Test split – 0.8 : 0.1 : 0.1

Normalization

- Normalized each image $\mathbf{x}_{(i)}$ by subtracting its mean $\mu_{(i)}$ and scaling within the $[-1, 1]$ range as follows:

$$\frac{\mathbf{x}_{(i)} - \mu_{(i)}}{\max\left|\mathbf{x}^t_{(i)} - \mu^t_{(i)}\right|}$$

- $\mathbf{x}^t_{(i)}$ - i^{th} target image
- $\mu^t_{(i)}$ - mean of i^{th} target image

Data Augmentation

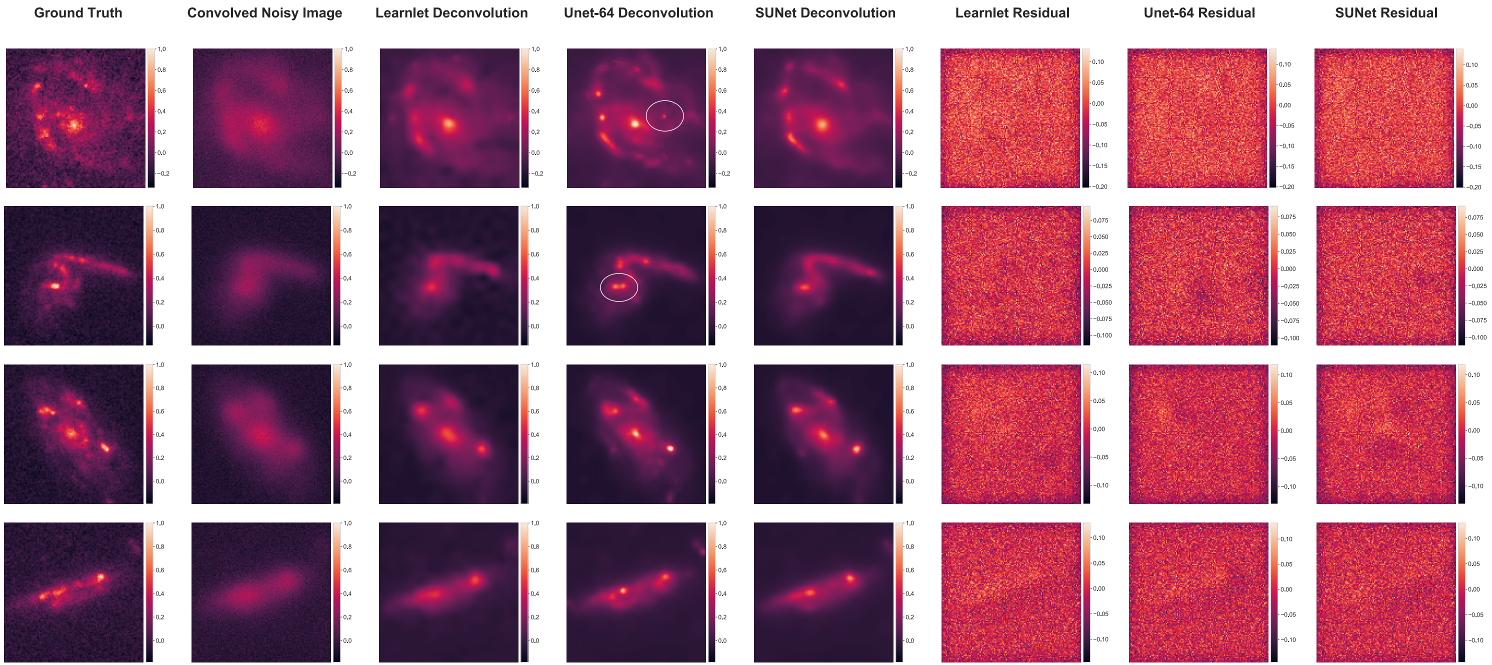
- Random rotations in multiples of 90°, translations and flips along horizontal & vertical axes

Performance Comparison

Method	No. of parameters	Batch Size	Epochs	Training Time (hrs.)	Runtime per image (ms)
Learnlet	44,840	32	150	5.45	30.8
Unet-64	31,023,940	32	500	14.4	26.3
SUNet	38,365,111	16	250	90	15.2

All computations on Titan RTX Turing GPU with 24 GB RAM

Results



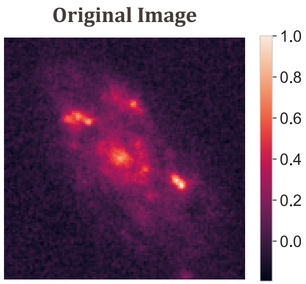
$$\text{Residual} = \mathbf{y} - \mathbf{h} * N_\theta(\hat{\mathbf{x}})$$

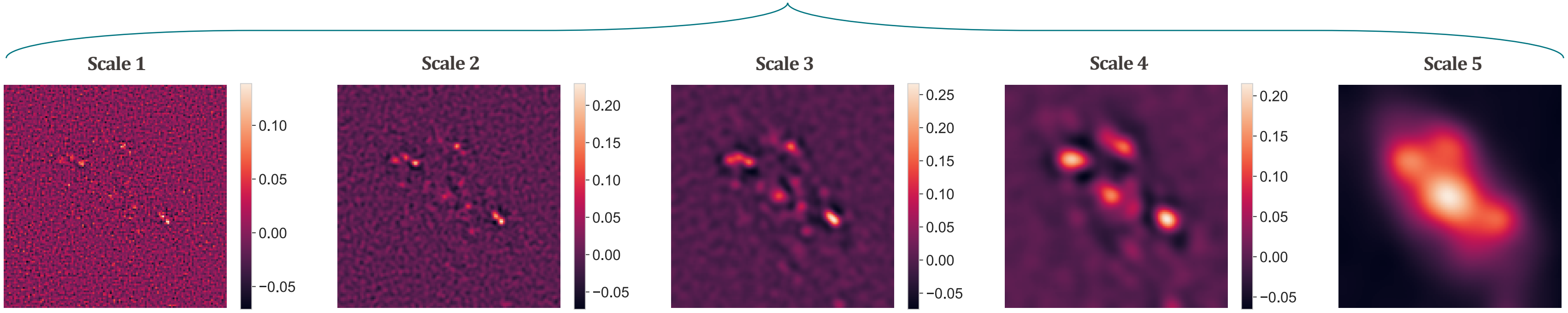
- $\mathbf{y}$ – noisy image
- $\mathbf{h}$ – PSF
- $\hat{\mathbf{x}}$ – noisy tikhonov input
- N_θ – network model

Deep learning-based galaxy image deconvolution
Akhaury et al, 2022

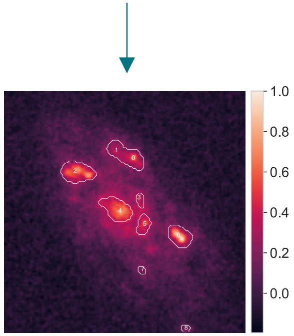
Multi-resolution Analysis

- Wavelet decomposition using SCARLET

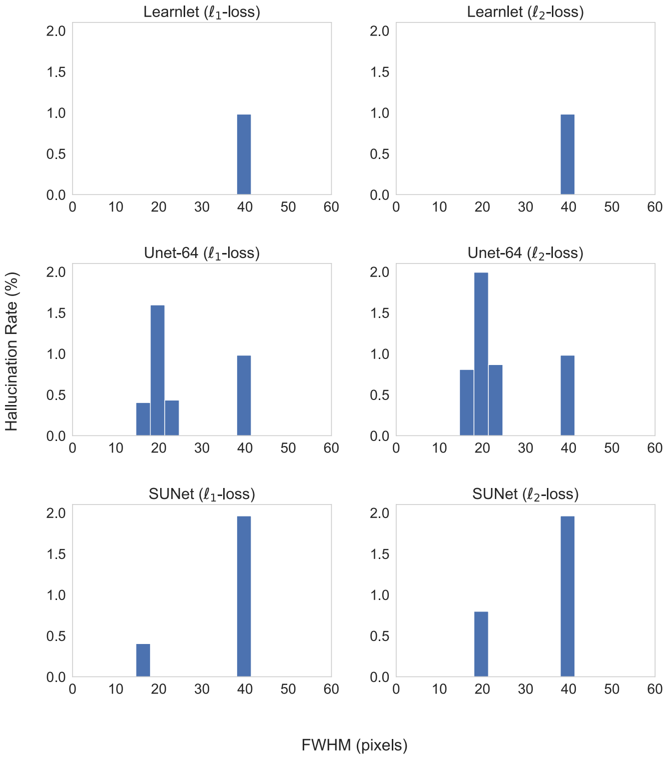


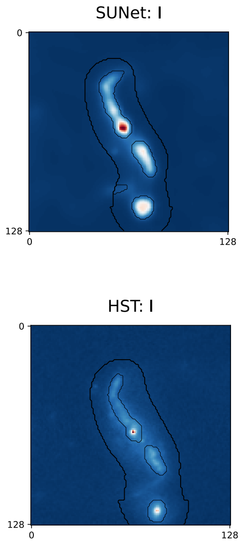


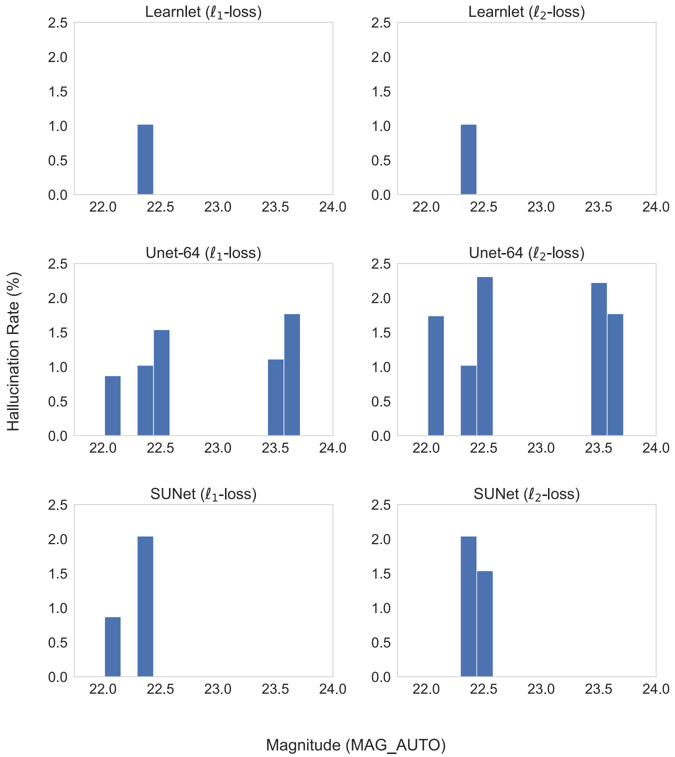
3rd Scale selected to detect clumps



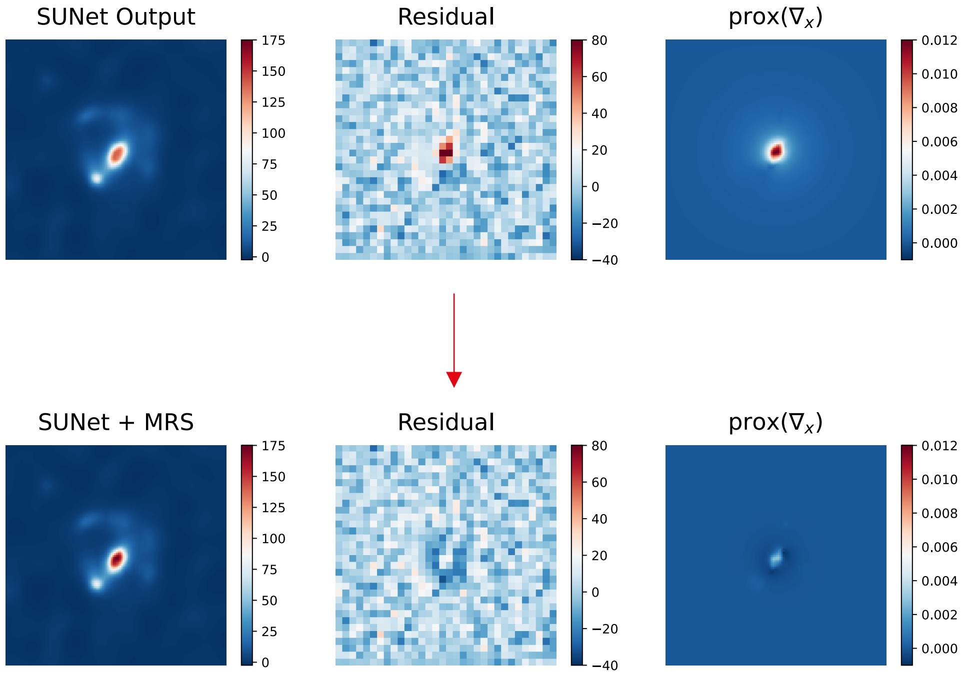
Hallucination Rate

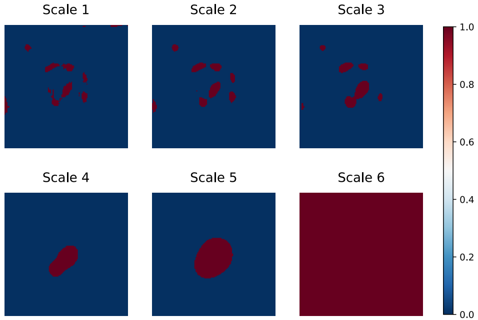






Debiasing with Multi-resolution Support (MRS)

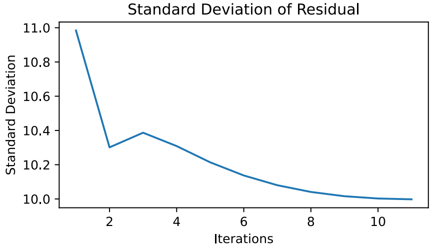




$$\mathbf{x_{j+1}} = \mathbf{x_j} + \mathrm{prox}(\nabla_\mathbf{x})$$

where

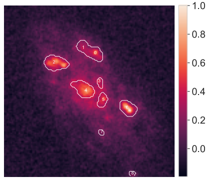
$$\mathbf{r_j} = \mathbf{y} - \mathbf{Hx_j}$$
$$\nabla_\mathbf{x} = \mathbf{H}^\top \mathbf{r_j}$$
$$\mathrm{prox}(\nabla_\mathbf{x}) = (\Phi^\top \mathbf{M} \Phi)\nabla_\mathbf{x}$$
$$\mathbf{M} = \mathrm{MRS}(\Phi(\mathbf{x_0}))$$

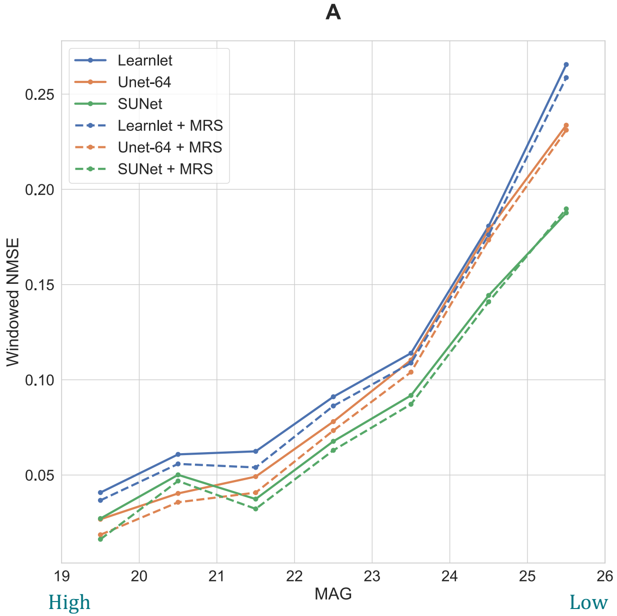


EPFL

Metrics

Extracted Flux in
small-scale
structures

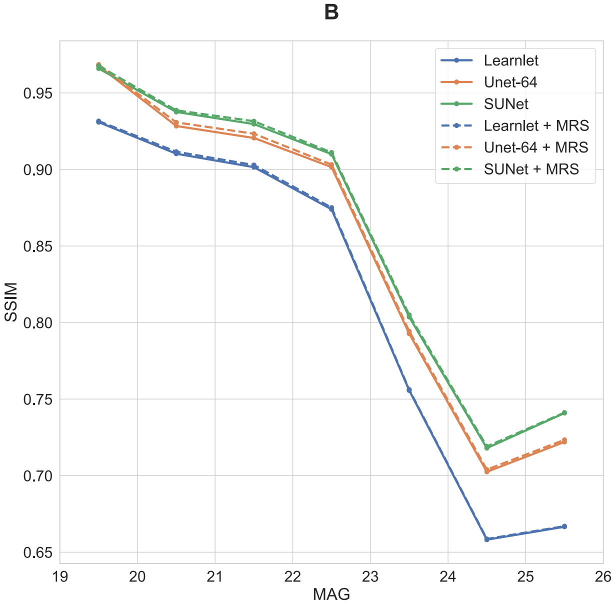




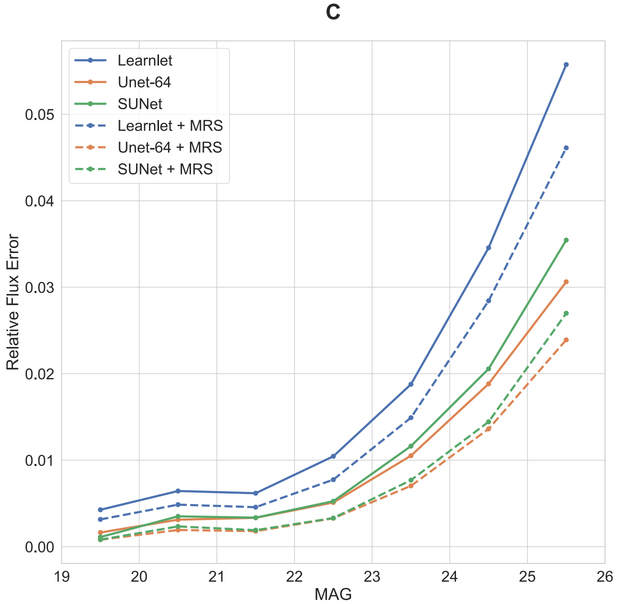
High SNR

Low SNR

NMSE



SSIM



Flux Error

SSIM = 1: Identical
SSIM = 0: Dissimilar

Test on VLT Images

EDisCS – the ESO distant cluster survey★,★★,★★★

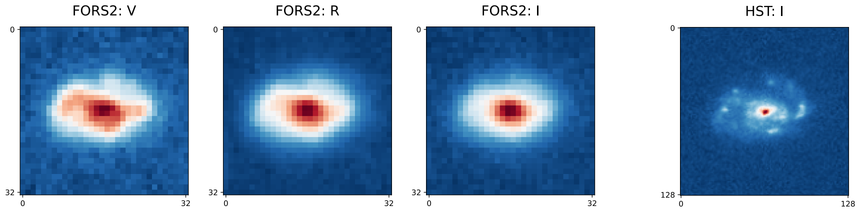
Sample definition and optical photometry

S. D. M. White[1], D. I. Clowe[2], L. Simard[3], G. Rudnick[1], G. De Lucia[1], A. Aragón-Salamanca[4], R. Bender[5], P. Best[6], M. Bremer[7], S. Charlot[1], J. Dalcanton[8], M. Dantel[9], V. Desai[8], B. Fort[10], C. Halliday[11], P. Jablonka[12], G. Kauffmann[1], Y. Mellier[10,9], B. Milvang-Jensen[5], R. Pelló[13], B. Poggianti[14], S. Poirier[12], H. Rottgering[15], R. Saglia[5], P. Schneider[16], and D. Zaritsky[2]

- All 84 cluster members at redshifts: **z ≈ 0.58**, **z ≈ 0.7**, and **z ≈ 0.79**

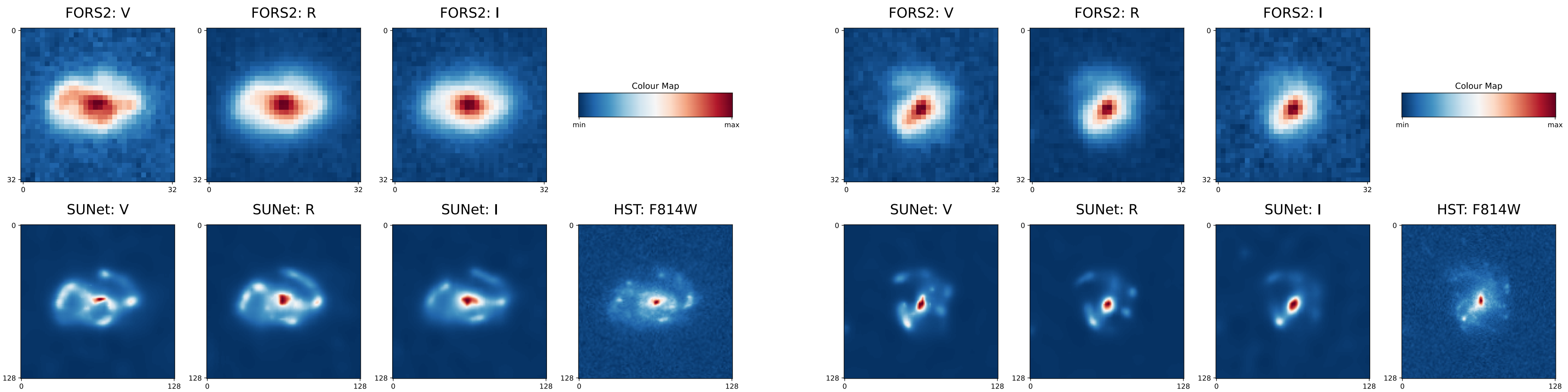
Cluster IDs	Redshift (z_{cl})	Number of samples (N_{cl})
cl1037.9-1243	0.5805	11
cl1040.7-1155	0.7043	19
cl1054.4-1146	0.6972	20
cl1103.7-1245b	0.7031	6
cl1216.8-1201	0.7955	28

Table 1: Summary of the EDisCS clusters considered for analysis.



- **Noisy images:** VLT FORS2 cutouts of 32 × 32 pixels in V (**555nm**), R (**655nm**), and I (**768nm**) bands with resolution = **0.2″**
- **Ground truth:** HST ACS cutouts of 128 × 128 pixels in the ***F814W*** filter (I-band) with resolution = **0.05″**

Outputs



- Able to resolve small-scale structures and recover morphology
- Achieves a resolution close to HST
- Generalizes well to images with completely different noise properties than the training dataset

Reproducible Research

- The ready-to-use version of our SUNet deconvolution method

  https://github.com/utsav-akhaury/SUNet/tree/main/Deconvolution

- The repository fork of the SUNet code used for training the network

  https://github.com/utsav-akhaury/SUNet

- Link to the trained SUNet weights

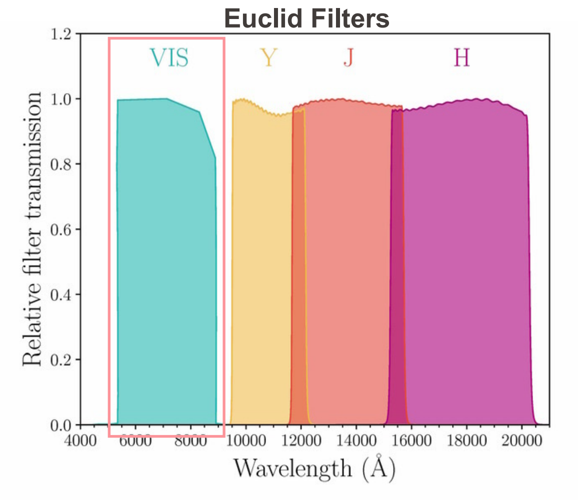
  https://doi.org/10.5281/zenodo.10287213

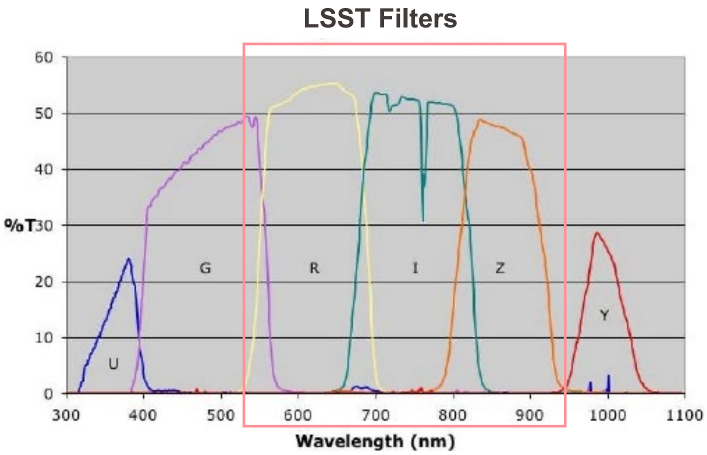
- The Learnlet and Unet-64 codes

  https://github.com/utsav-akhaury/understanding-unets/tree/candels

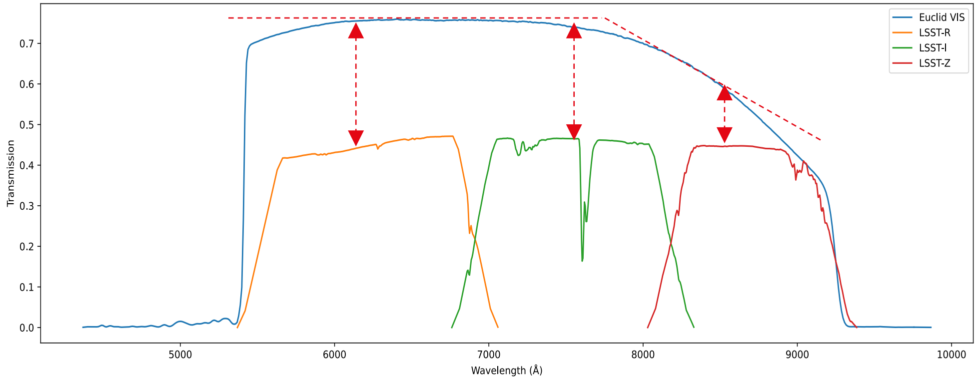
Multi-channel Deconvolution

Motivation





$$\mathbf{x}_{euc} = \alpha_R \mathbf{x}_R + \alpha_I \mathbf{x}_I + \alpha_Z \mathbf{x}_Z$$



Spectral Energy Distributions (SED)

$$\alpha_R,\ \alpha_I,\ \alpha_Z \in \mathbb{R}^n$$

The Multi-channel Deconvolution Problem

Model

$$\mathbf{y}_R = \mathbf{h}_R * \mathbf{x}_R^t + \eta_R$$

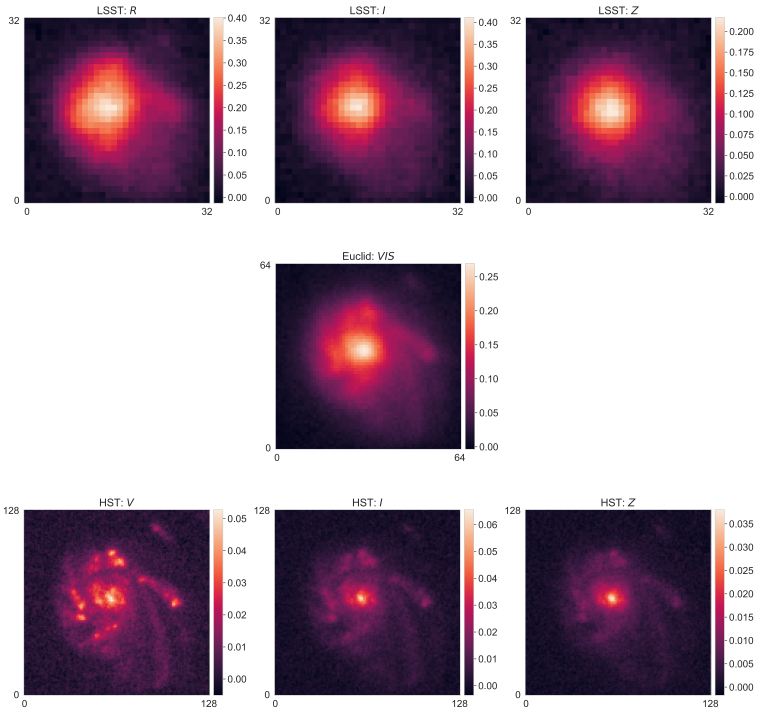
$$\mathbf{y}_I = \mathbf{h}_I * \mathbf{x}_I^t + \eta_I$$

$$\mathbf{y}_Z = \mathbf{h}_Z * \mathbf{x}_Z^t + \eta_Z$$

$$\mathbf{x}_{euc}^t = \alpha_R \mathbf{x}_R^t + \alpha_I \mathbf{x}_I^t + \alpha_Z \mathbf{x}_Z^t$$

$$\mathbf{y}_{euc} = \mathbf{h}_{euc} * \mathbf{x}_{euc}^t + \eta_{euc}$$

- Observed Noisy Images — $\mathbf{y}_R, \mathbf{y}_I, \mathbf{y}_Z \in \mathbb{R}^{n \times n}$
- PSFs — $\mathbf{h}_R, \mathbf{h}_I, \mathbf{h}_Z \in \mathbb{R}^{n \times n}$
- Ground Truth Images — $\mathbf{x}_R^t, \mathbf{x}_I^t, \mathbf{x}_Z^t \in \mathbb{R}^{n \times n}$
- Additive Noise — $\eta_R, \eta_I, \eta_Z \in \mathbb{R}^{n \times n}$
- Spectral Energy Distributions (SED) — $\alpha_R, \alpha_I, \alpha_Z \in \mathbb{R}^{n}$



LSST 0.2”

Euclid 0.1”

HST 0.05”

EPFL

The Loss Functions

$$L_R(\mathbf{x}_R) = \frac{1}{2}\left\|\frac{\mathbf{h}_R * \mathbf{x}_R - \mathbf{y}_R}{\sigma_R}\right\|_F^2 + \lambda_{constr}\left\|\frac{\mathbf{h}_{euc} * \sum_C \alpha_C \mathbf{x}_C - \mathbf{y}_{euc}}{\sigma_{euc}}\right\|_F^2$$

$$L_I(\mathbf{x}_I) = \frac{1}{2}\left\|\frac{\mathbf{h}_I * \mathbf{x}_I - \mathbf{y}_I}{\sigma_I}\right\|_F^2 + \lambda_{constr}\left\|\frac{\mathbf{h}_{euc} * \sum_C \alpha_C \mathbf{x}_C - \mathbf{y}_{euc}}{\sigma_{euc}}\right\|_F^2$$

$$L_Z(\mathbf{x}_Z) = \frac{1}{2}\left\|\frac{\mathbf{h}_Z * \mathbf{x}_Z - \mathbf{y}_Z}{\sigma_Z}\right\|_F^2 + \lambda_{constr}\left\|\frac{\mathbf{h}_{euc} * \sum_C \alpha_C \mathbf{x}_C - \mathbf{y}_{euc}}{\sigma_{euc}}\right\|_F^2$$

where

$C \in \{R, I, Z\}$

$\lambda_{constr} \in \mathbb{R}_+$

- Spectral Energy Distributions (SED) $\alpha_R, \alpha_I, \alpha_Z \in \mathbb{R}^n$
- Noisemaps $\sigma_R, \sigma_I, \sigma_Z \in \mathbb{R}^{n \times n}$

Optimization

$$\hat{\mathbf{x}}_{\{R,I,Z\}} = \underset{\mathbf{x}_{\{R,I,Z\}}}{\operatorname{argmin}} L_{\{R,I,Z\}}\left(\mathbf{x}_{\{R,I,Z\}}\right)$$

Loss Functions iteratively minimized using Gradient Descent

$$\mathbf{x}_{\{R,I,Z\}}^{[k+1]} \longleftarrow \mathbf{x}_{\{R,I,Z\}}^{[k]} - \beta_{\{R,I,Z\}} \nabla L_{\{R,I,Z\}}\left(\mathbf{x}_{\{R,I,Z\}}^{[k]}\right)$$

Step Sizes

$$\beta_R, \beta_I, \beta_Z \in \mathbb{R}^n$$

Gradients of the Loss Functions

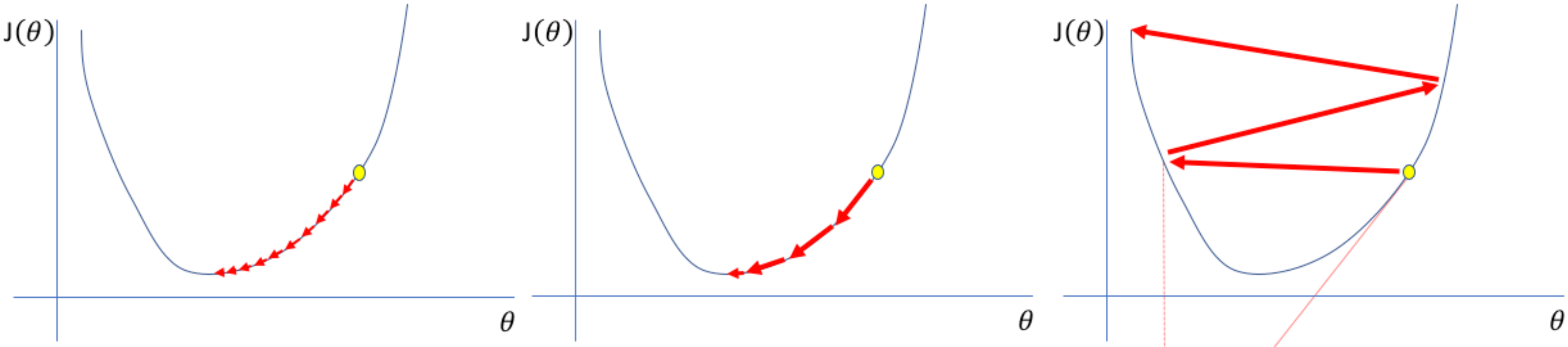
$$\nabla L_R(\mathbf{x}_R) = \frac{\mathbf{h}_R^\top * (\mathbf{h}_R * \mathbf{x}_R - \mathbf{y}_R)}{\|\sigma_R\|_F^2} + 2\lambda_{constr}\alpha_R \mathbf{h}_{euc}^\top * \left[\frac{\mathbf{h}_{euc} * \sum_C \alpha_C \mathbf{x}_C - \mathbf{y}_{euc}}{\|\sigma_{euc}\|_F^2}\right]$$

$$\nabla L_I(\mathbf{x}_I) = \frac{\mathbf{h}_I^\top * (\mathbf{h}_I * \mathbf{x}_I - \mathbf{y}_I)}{\|\sigma_I\|_F^2} + 2\lambda_{constr}\alpha_I \mathbf{h}_{euc}^\top * \left[\frac{\mathbf{h}_{euc} * \sum_C \alpha_C \mathbf{x}_C - \mathbf{y}_{euc}}{\|\sigma_{euc}\|_F^2}\right]$$

$$\nabla L_Z(\mathbf{x}_Z) = \frac{\mathbf{h}_Z^\top * (\mathbf{h}_Z * \mathbf{x}_Z - \mathbf{y}_Z)}{\|\sigma_Z\|_F^2} + 2\lambda_{constr}\alpha_Z \mathbf{h}_{euc}^\top * \left[\frac{\mathbf{h}_{euc} * \sum_C \alpha_C \mathbf{x}_C - \mathbf{y}_{euc}}{\|\sigma_{euc}\|_F^2}\right]$$

Convergence Guarantee & Optimal step size

Too low | **Just right** | **Too high**



A small learning rate requires many updates before reaching the minimum point

The optimal learning rate swiftly reaches the minimum point

Too large of a learning rate causes drastic updates which lead to divergent behaviors

Credits: Jeremy Jordan

Convergence Guarantee & Optimal step size

A function's gradient is Lipschitz continuous if

$$\|\nabla f(\mathbf{x}') - \nabla f(\mathbf{x})\| \leq C\|\mathbf{x}' - \mathbf{x}\|$$

where C is the **Lipschitz constant**

In our case

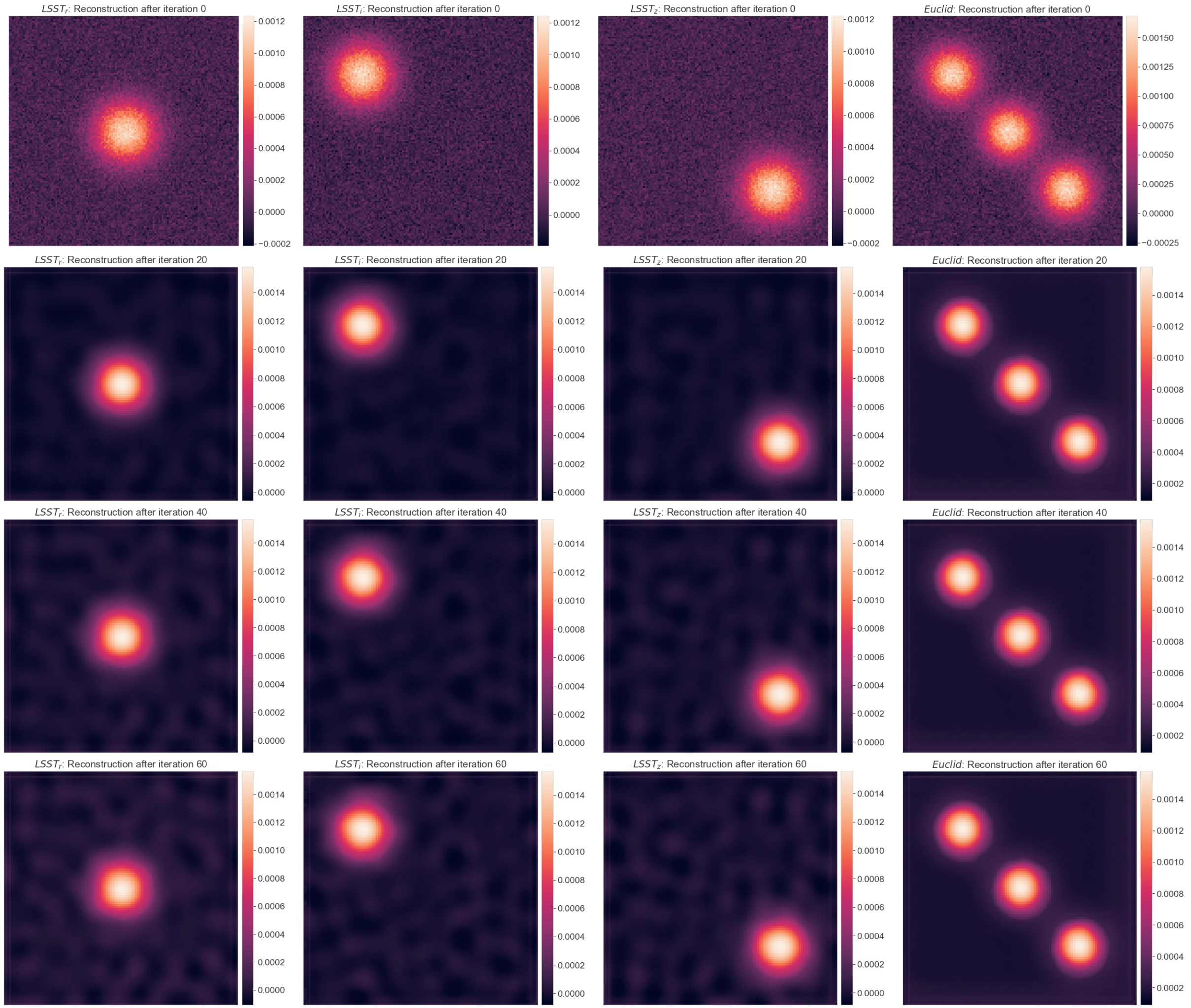
$$\left\|\nabla L_{\{R,I,Z\}}\left(\mathbf{x}'_{\{R,I,Z\}}\right) - \nabla L_{\{R,I,Z\}}\left(\mathbf{x}_{\{R,I,Z\}}\right)\right\| \leq C_{\{R,I,Z\}}\left\|\mathbf{x}'_{\{R,I,Z\}} - \mathbf{x}_{\{R,I,Z\}}\right\|$$

Substituting the individual loss functions, we get

$$C_{\{R,I,Z\}} \geq \frac{\mathbf{h}^{\top}_{\{R,I,Z\}} * \mathbf{h}_{\{R,I,Z\}}}{\left\|\sigma_{\{R,I,Z\}}\right\|_F^2} + \frac{2\lambda_{constr}\alpha^2_{\{R,I,Z\}}\mathbf{h}^{\top}_{euc} * \mathbf{h}_{euc}}{\|\sigma_{euc}\|_F^2}$$

The Optimal Condition for Convergence

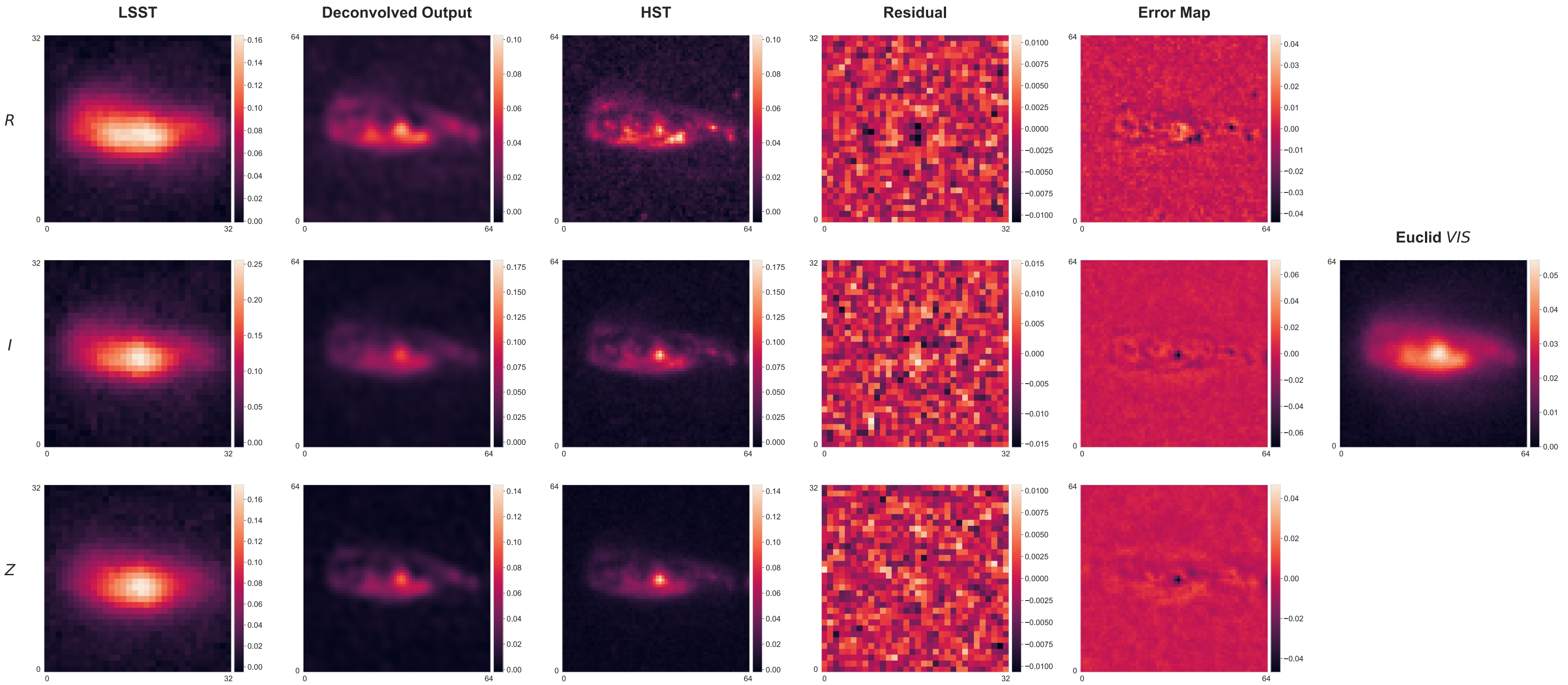
$$\beta_{\{R,I,Z\}} \leq \frac{1}{C_{\{R,I,Z\}}}$$



Flux Leakage Test

- Assume 3 separately placed Gaussians in each channel (corresponding to LSST channels)
- The joint image (Euclid) is a linear sum of these channels
- No Flux Leakage from one channel to another

A Deconvolved Object



Thank You.