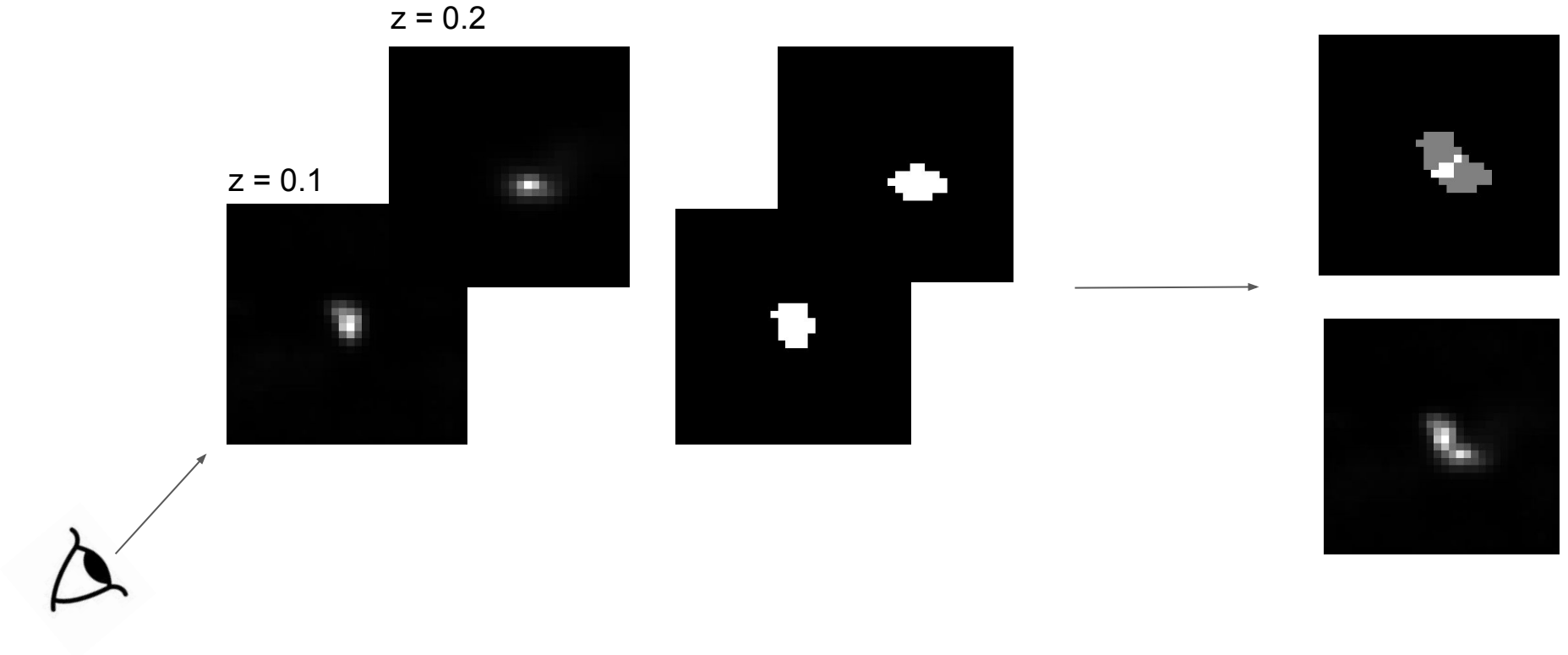
# Reconstructing blended galaxies with Machine Learning

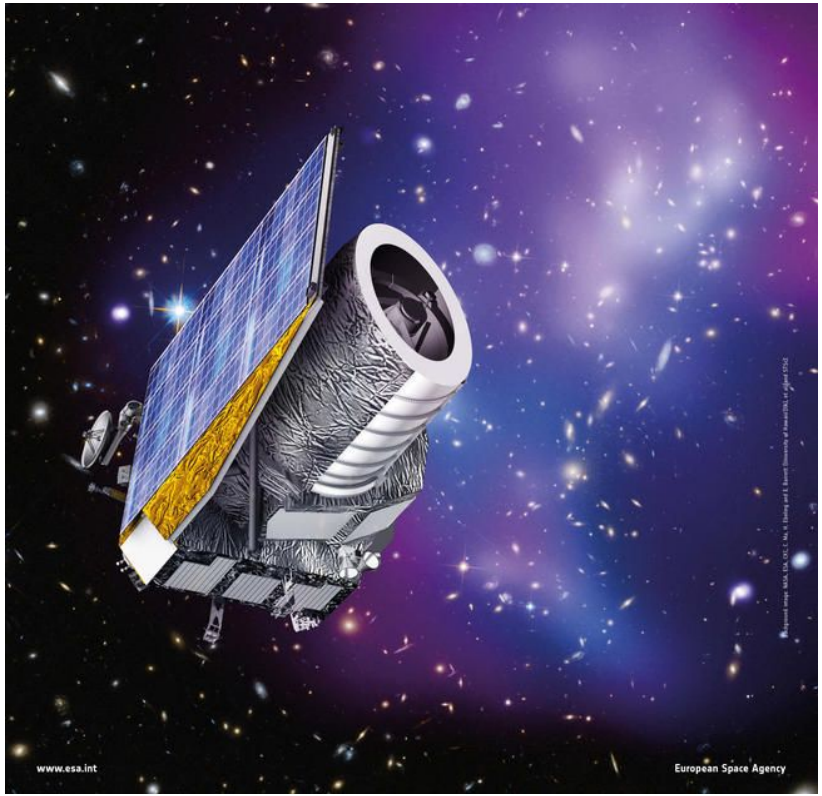
Lavanya Nemani (INAF OAR)
Advisors: Prof. Adriano Fontana, Dr. Emiliano Merlin, Dr. Fernando Caro

# Problem: Blending

Confusion effect[1] → Projection of photons to 2D plane with the same line of sight.



# Motivation




Credits: ESA

**Why is *accurate* flux measurement important?**

f → SED fitting →Distance → z → Mass

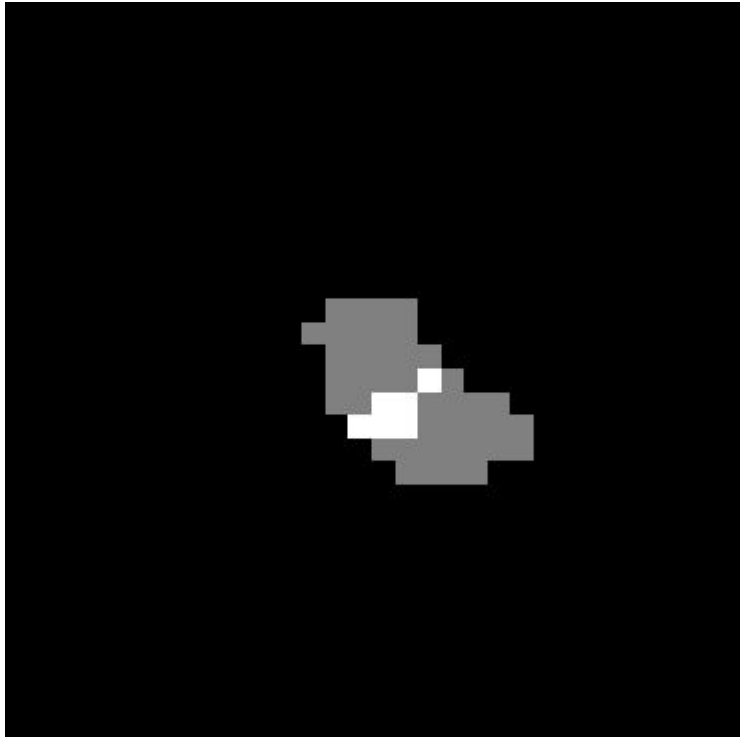
**Why the need for light profile?**

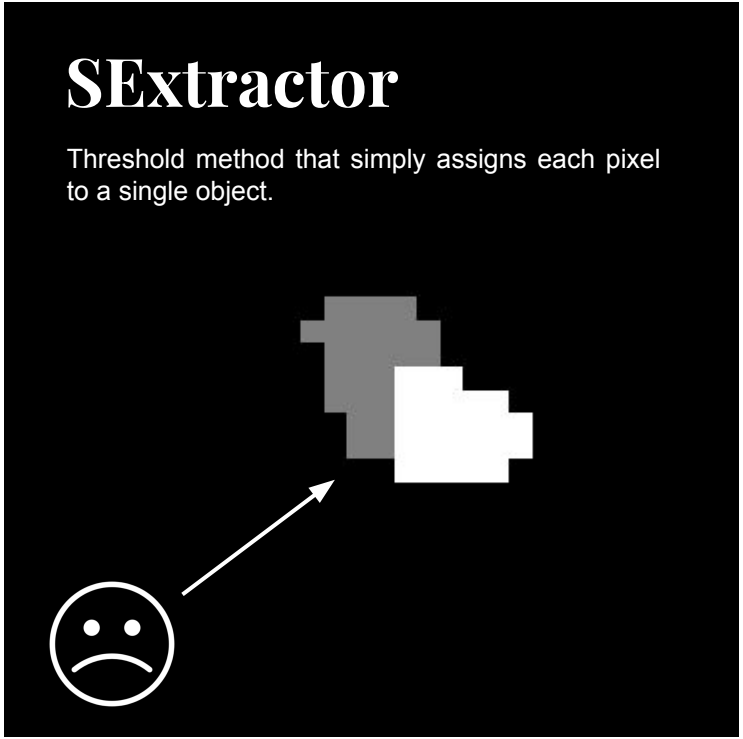
Priors for template fitting codes (TPHOT[3])

**Why we cares about de-blending?**

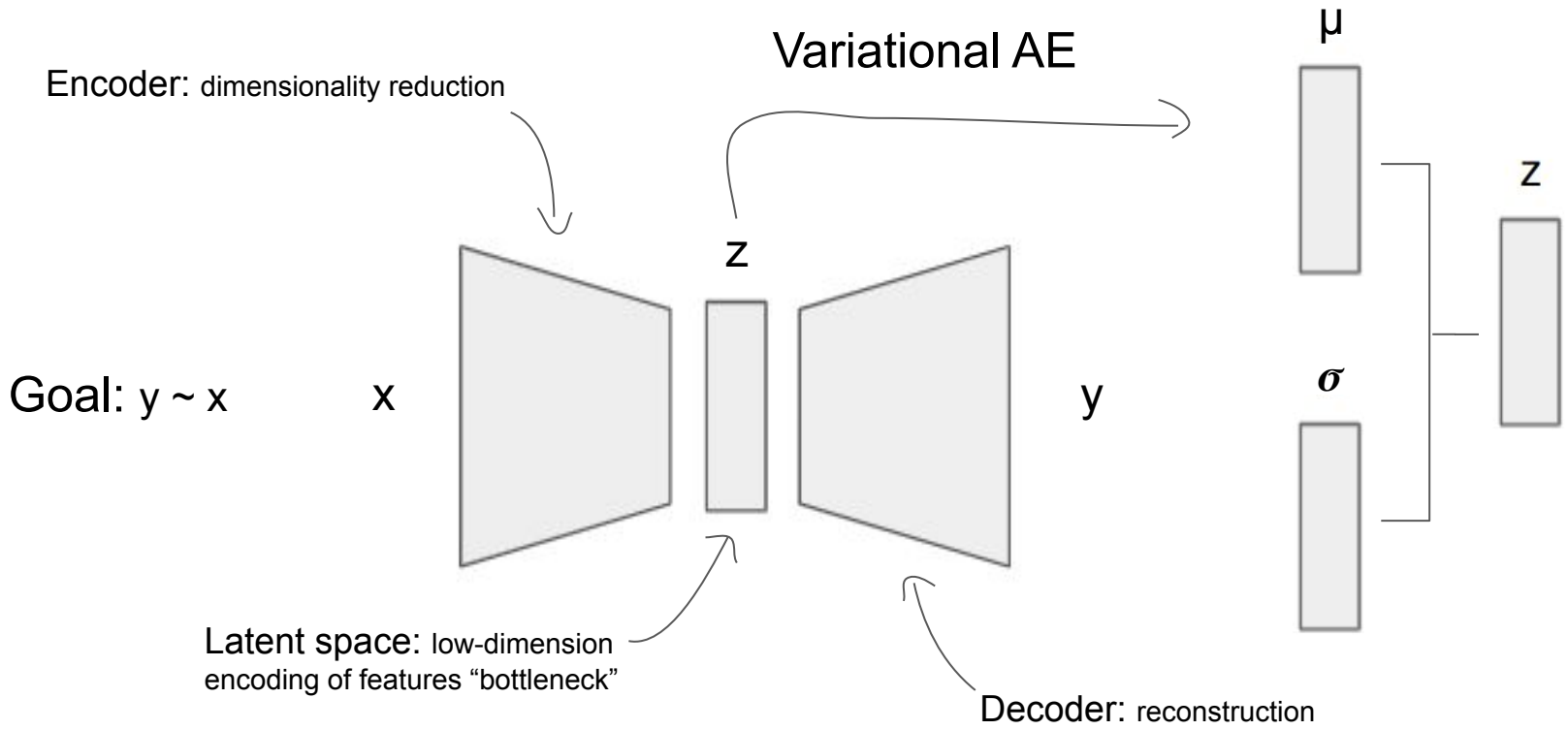
The upcoming deep extragalactic surveys like LSST and **EUCLID** expect to see a blending fraction of up to ***50%*** in the densest regions[2].

# Standard method: SExtractor[4]



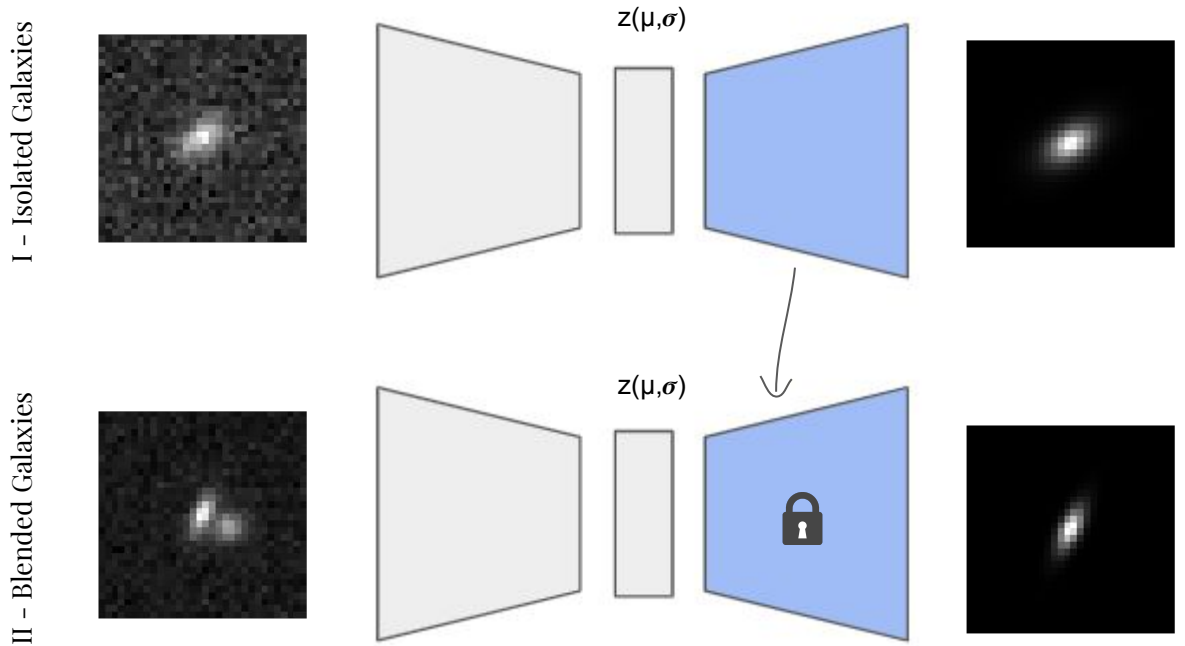


# Auto-Encoders



# VAEs for De-blending

Single galaxies; Learns to reconstruct galaxies




Blended galaxies; Learns to map central galaxy to z(μ,σ) s.t. reconstruction ~ original isolated

# Analytic Simulations - Dataset


*EGG is a code that can generate mock galaxy catalogs with realistic positions, morphologies and fluxes from the far-ultraviolet to the far-infrared.*

Catalog Generation

- EUCLID **VIS Band**

Masking

- 23.0 < Magnitude < 25.0
- 0.01 < Bulge Radius < 0.03
- 0.05 < Disk Radius < 0.35

**~250,000 Galaxies!**

*Stamps are created using **GalSim**, an image simulation toolkit.*

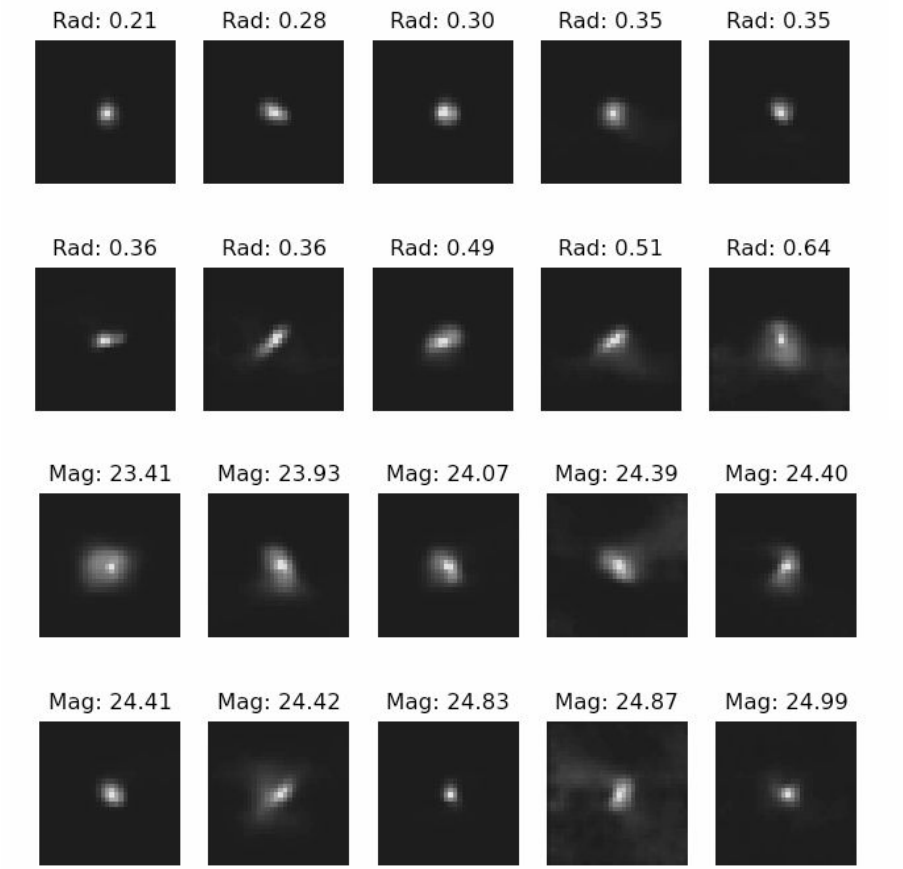
# Realistic Simulations - Dataset

## EUCLID MORPHOLOGY CHALLENGE

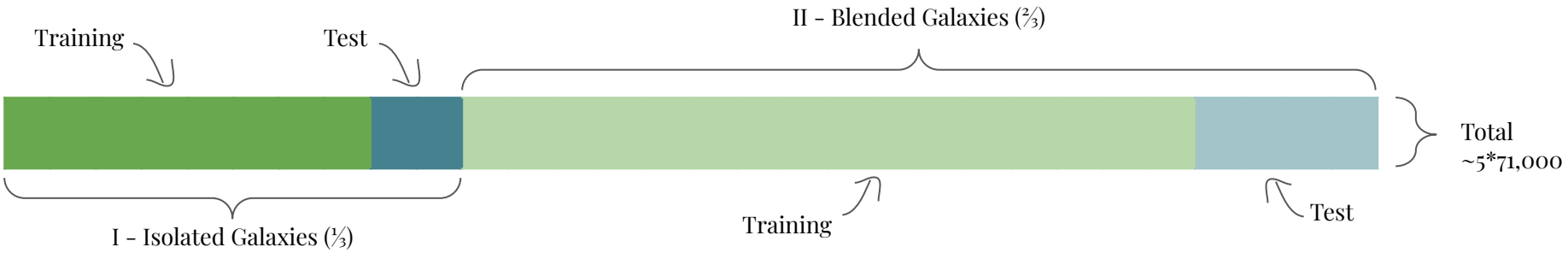
M. Castellano, M. Huertas-Company, E. Merlin, D. Tuccillo, H. Bretonniere

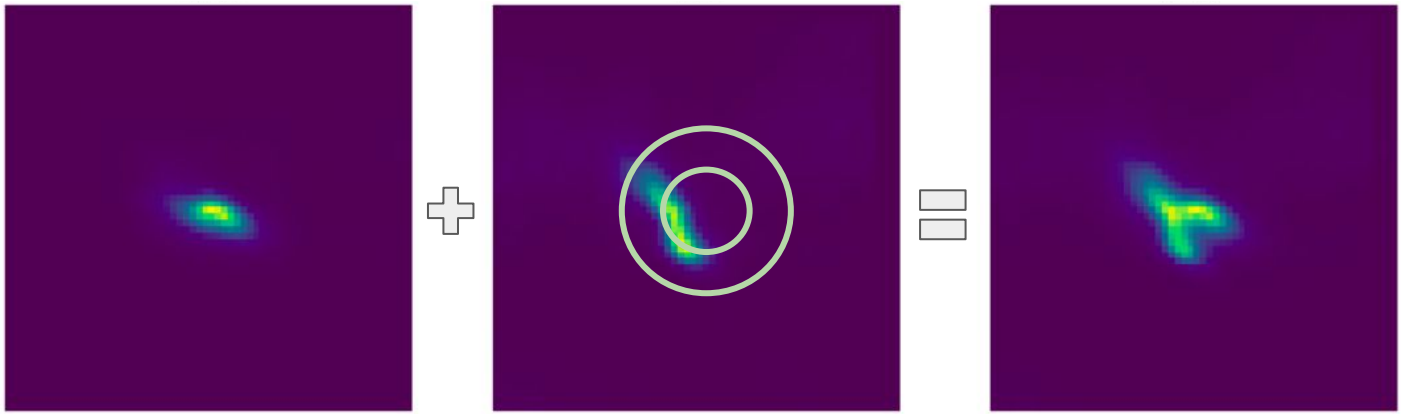
- 5 fields at 0.1 arcsec/pixel scale, corresponding to a FoV of ~0.5 sq. degrees containing ~71000 objects per field (after mag, radius masks)

- **Realistic morphologies**
  - Galaxies simulated using a Variational Auto-Encoder (VAE) trained on the COSMOS field.
  - + Normalizing Flows trained on Latent space of VAE (Input; 1-comp Sersic model)
  - The galaxies are convolved with the Euclid PSF (Point spread function)

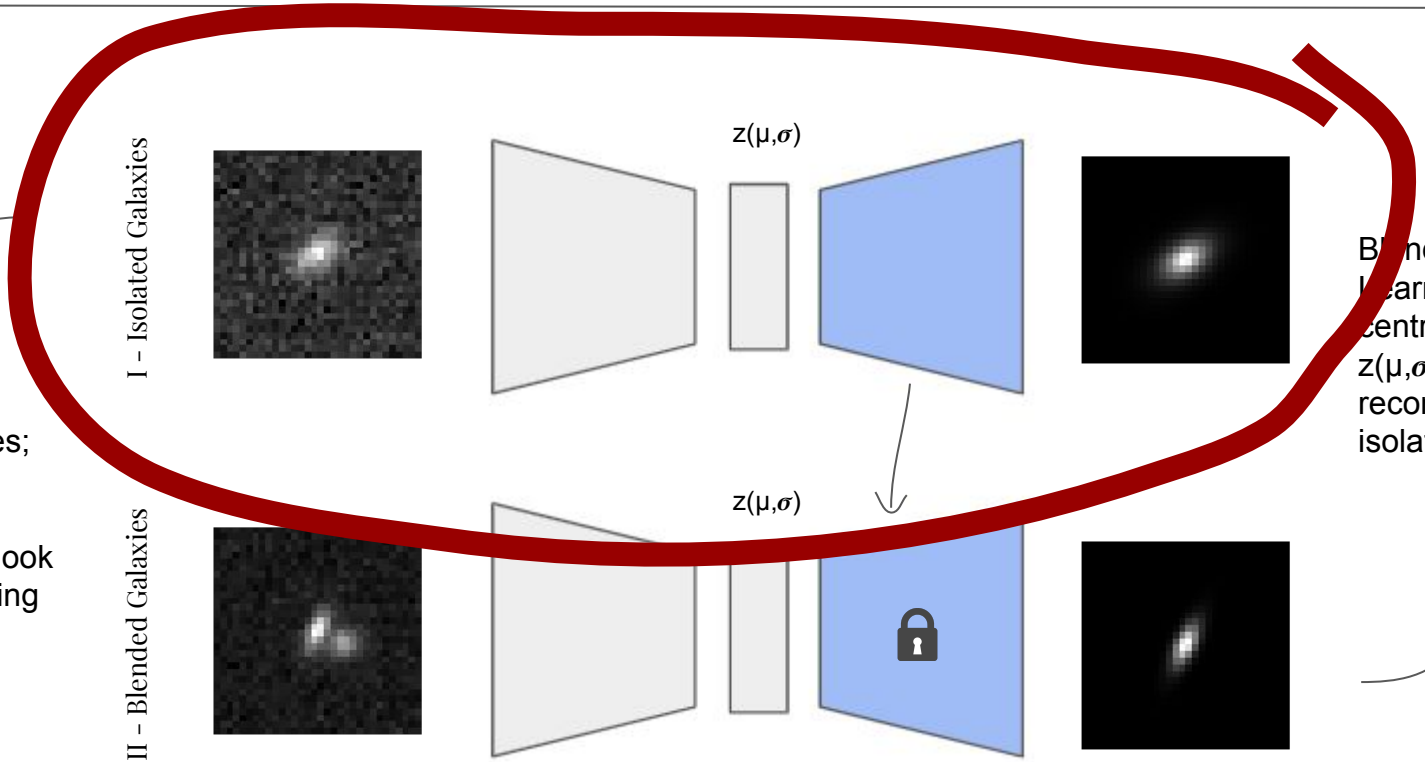


# Isolated & Blended



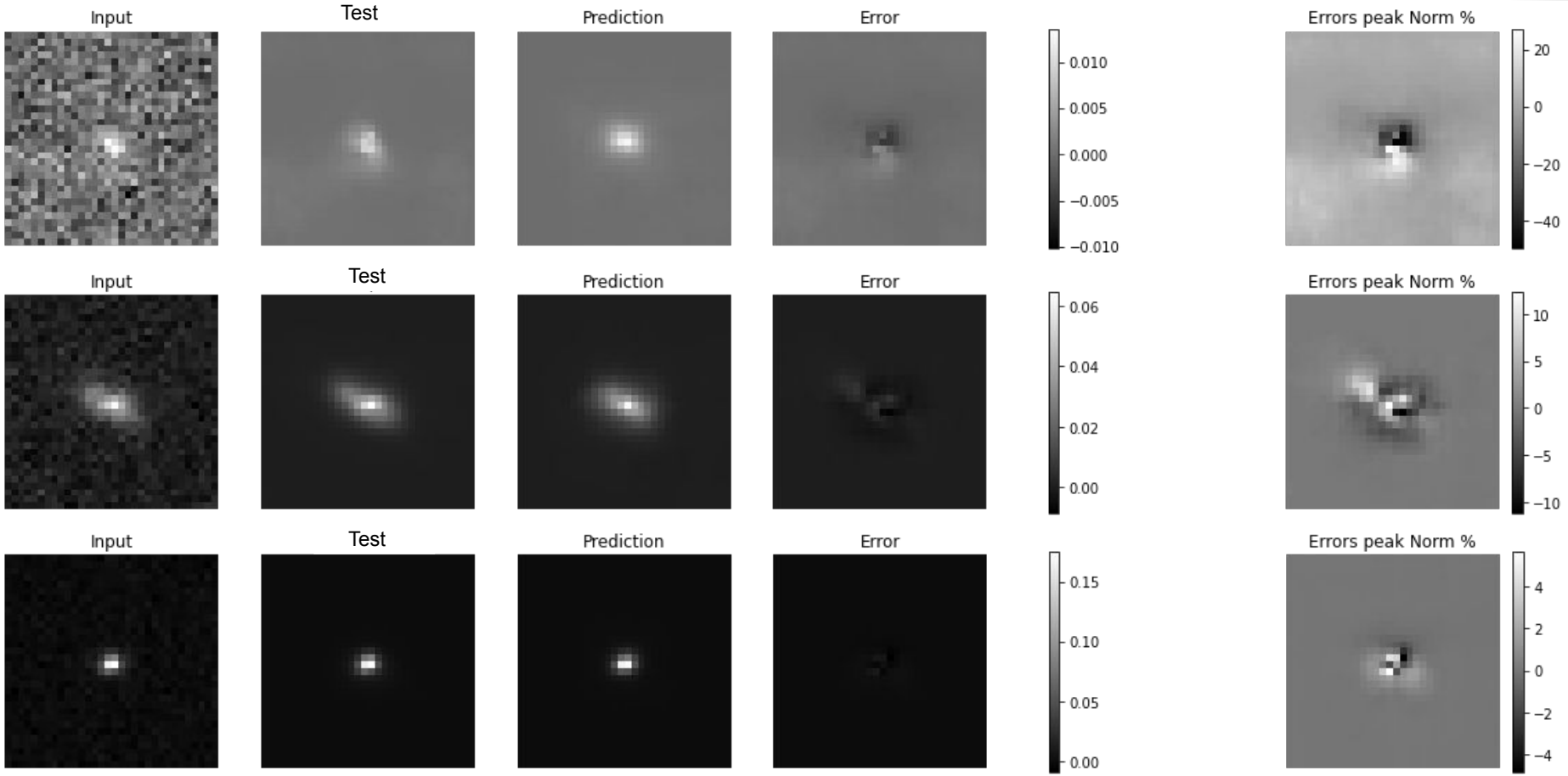
# VAEs for De-blending




Single galaxies; Learns to reconstruct galaxies that look like your training data

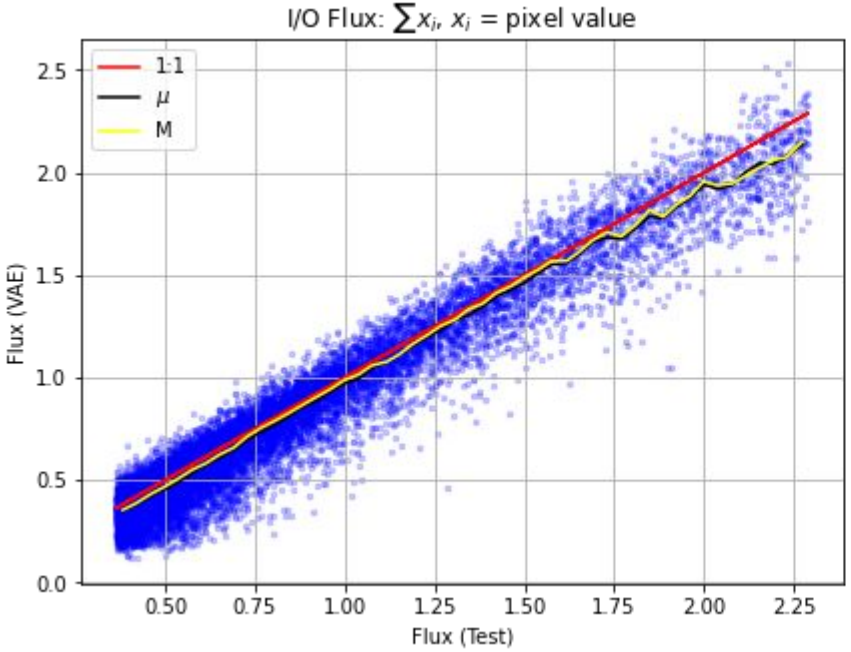
Blended galaxies; Learns to map central galaxy to $z(\mu,\sigma)$ s.t. reconstruction ~ isolated

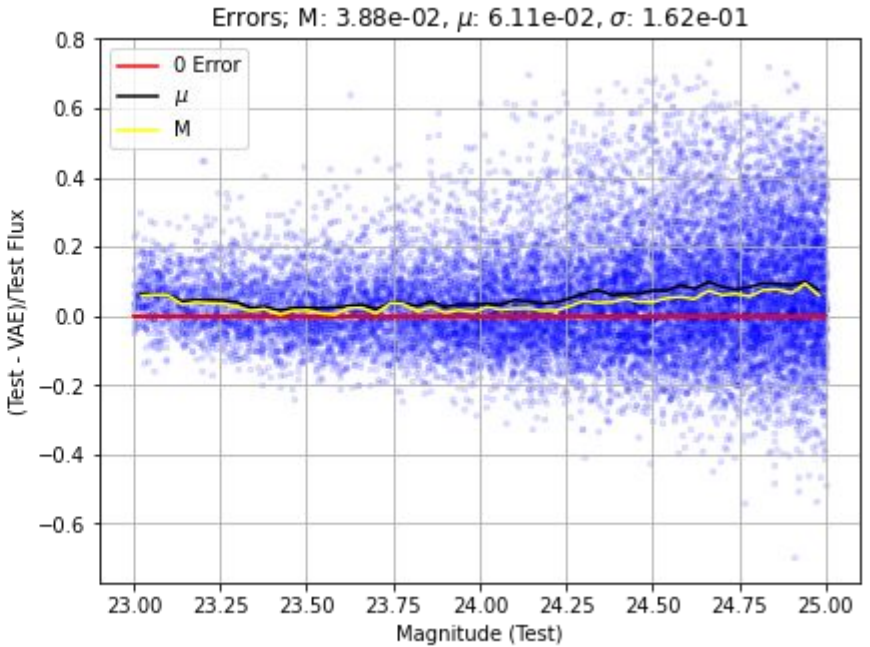
# VAE Results: I – Isolated Galaxies (Light Profile Reconstruction)



# VAE Results: I – Isolated Galaxies (Flux)

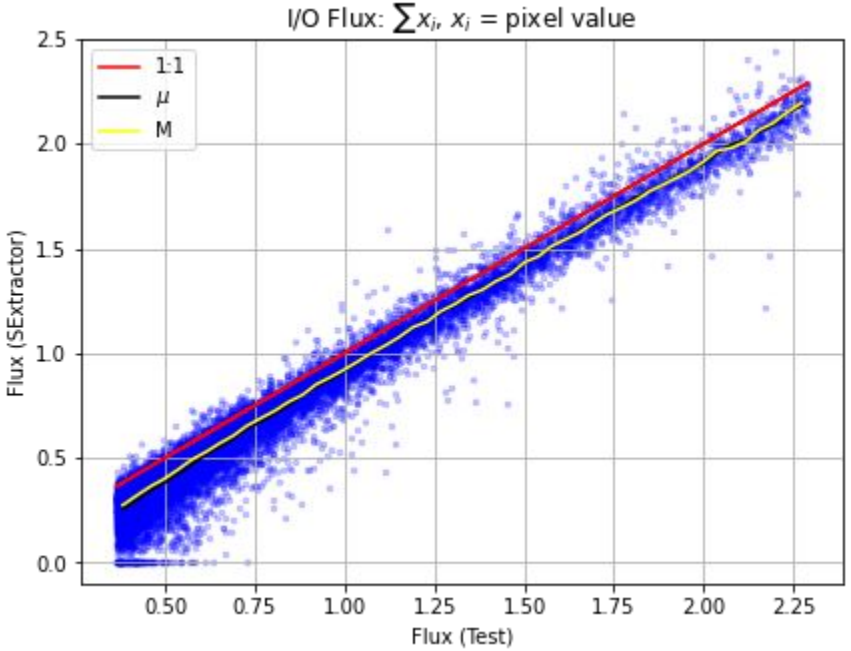
**Flux estimates**; The $\sigma$ of predicted fluxes on the **test images** is ~**16%** from the true value.

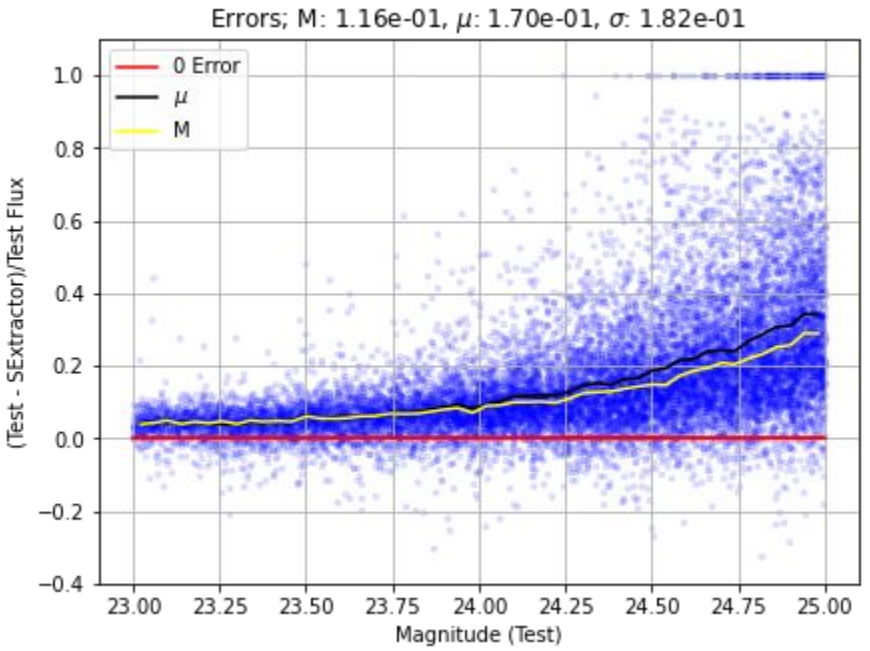


# SExtractor Results: Isolated Galaxies (Flux)

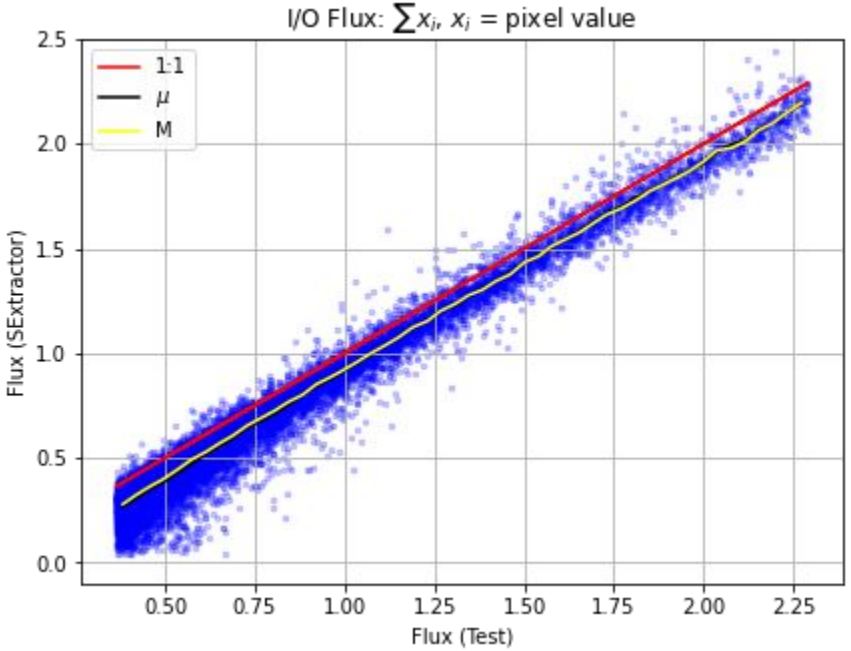
**Flux estimates**; The $\sigma$ of predicted fluxes on the **test images** is ~**18%** from the true value.

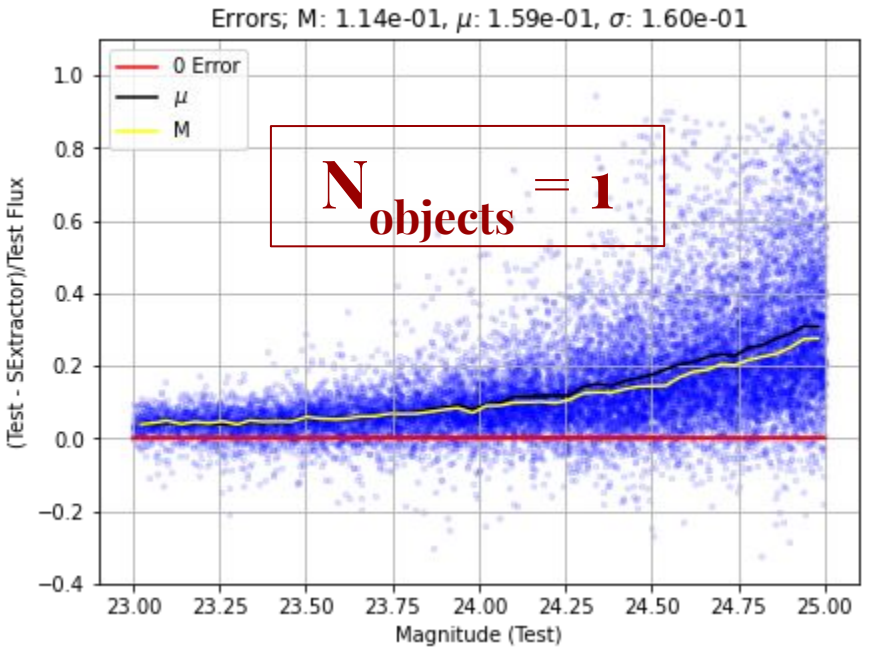


# SExtractor Results: Isolated Galaxies (Flux)

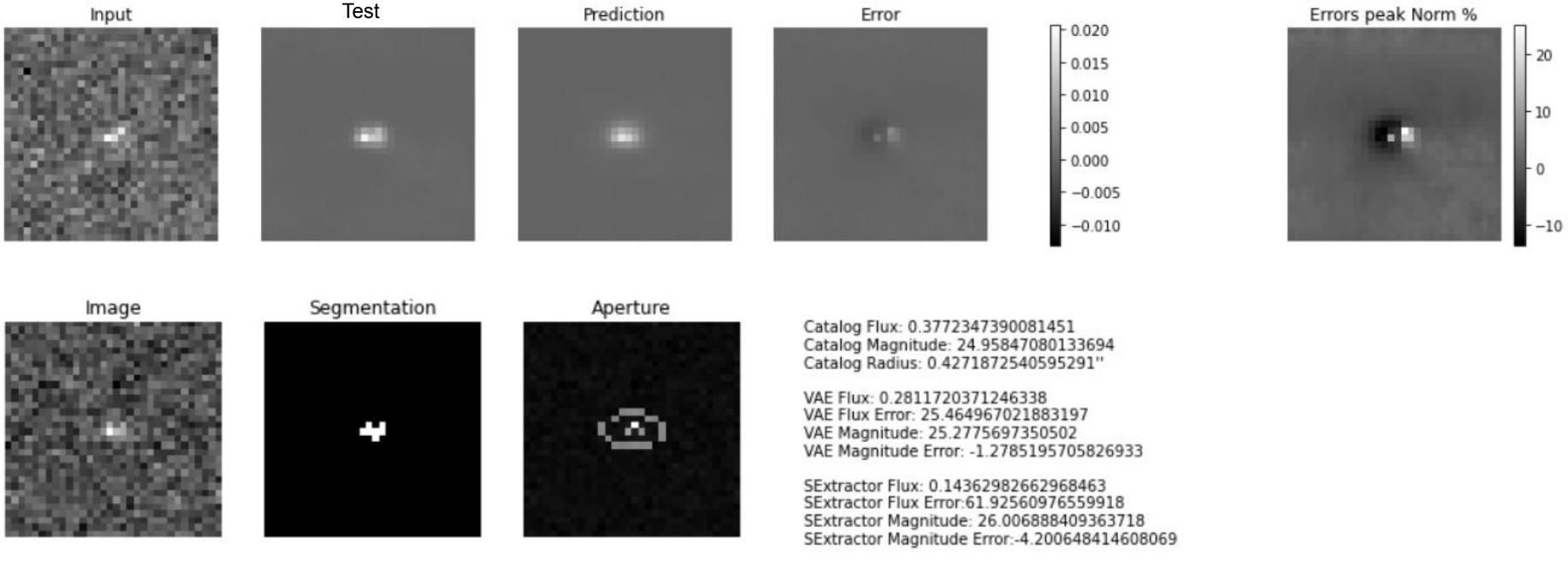
**Flux estimates**; The $\sigma$ of predicted fluxes on the **test images** is ~**16%** from the true value.



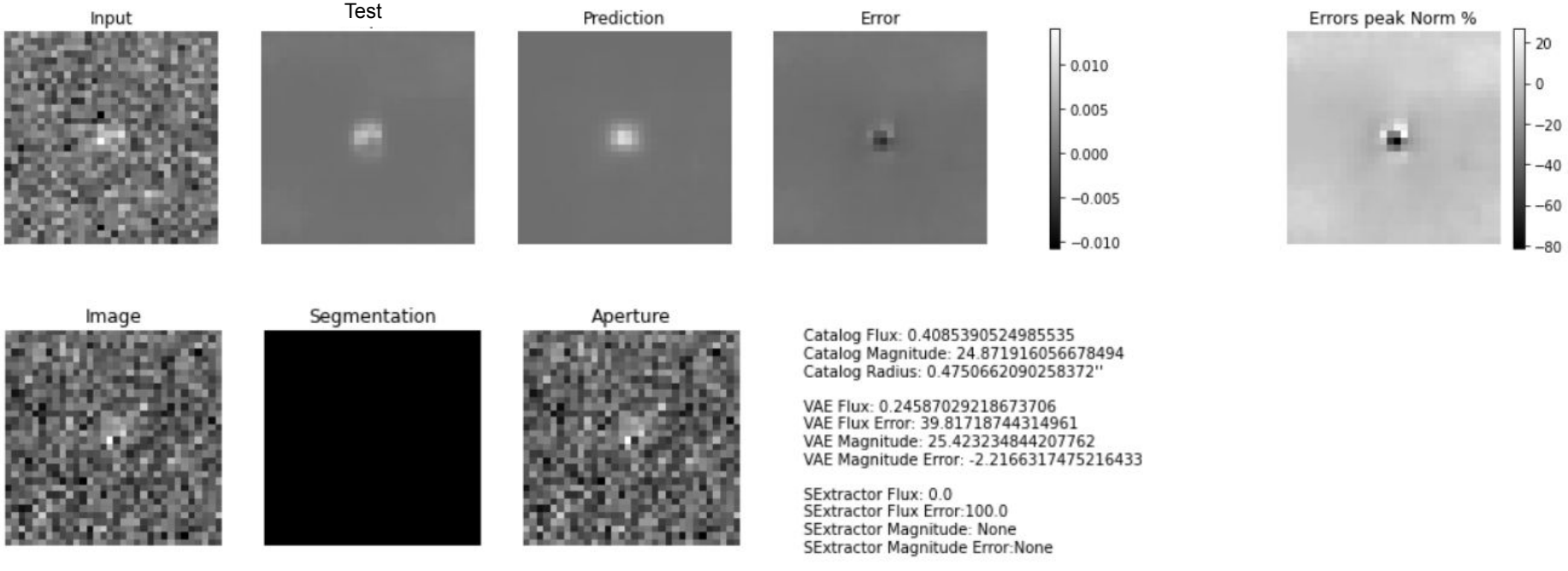
# SExtractor Results: Isolated Galaxies (Whats going wrong?)

**$N_{objects}$ = 1, 50% < SExtractor error < 75%**



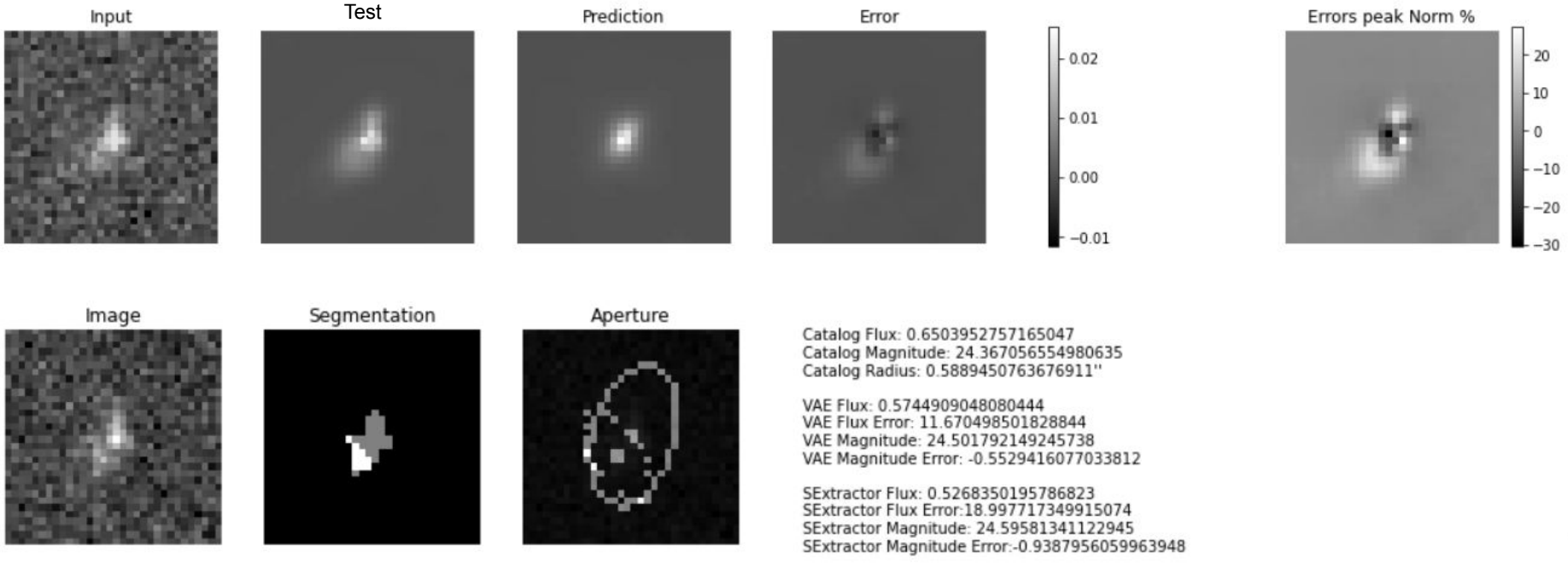
# SExtractor Results: Isolated Galaxies (Whats going wrong?)

**$N_{objects}$ = 0, SExtractor error = 100 %**



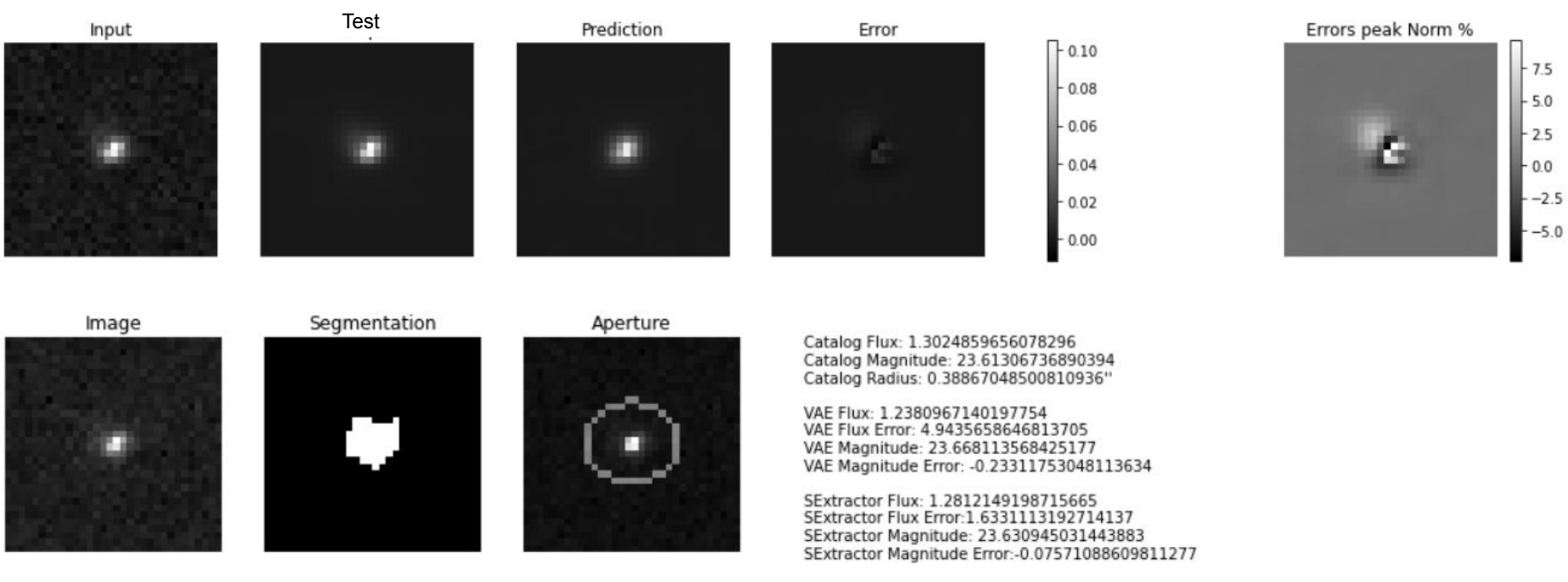
# SExtractor Results: Isolated Galaxies (Whats going wrong?)

**$N_{objects}$ = 2 or more**



# SExtractor Results: Isolated Galaxies

**Better than VAE**

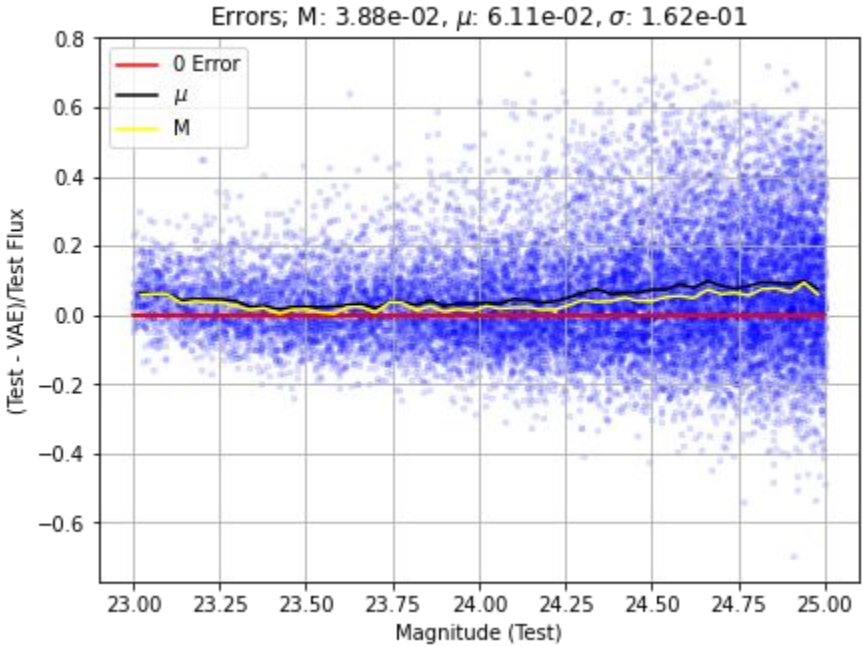


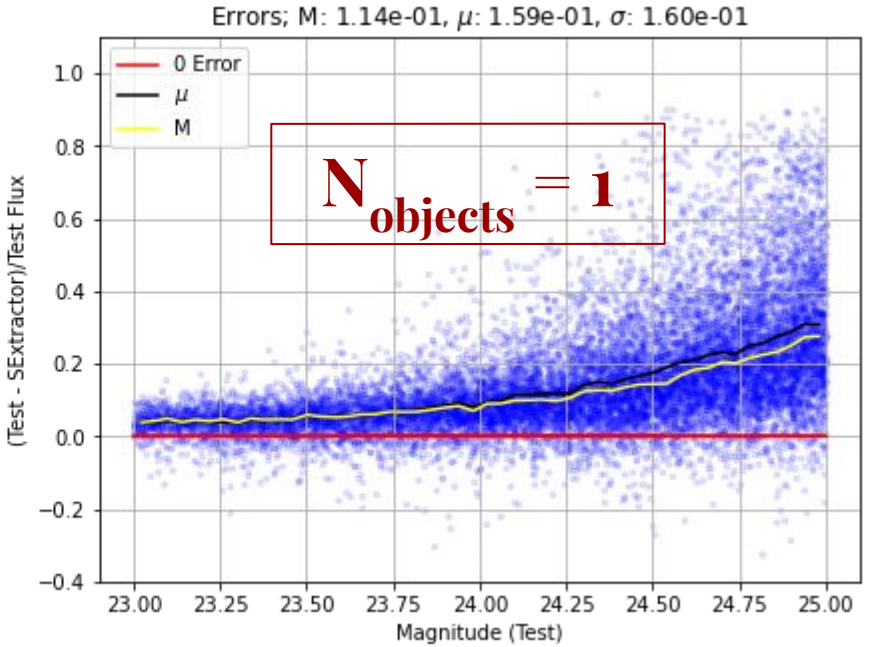
# VAE v/s SExtractor

Test

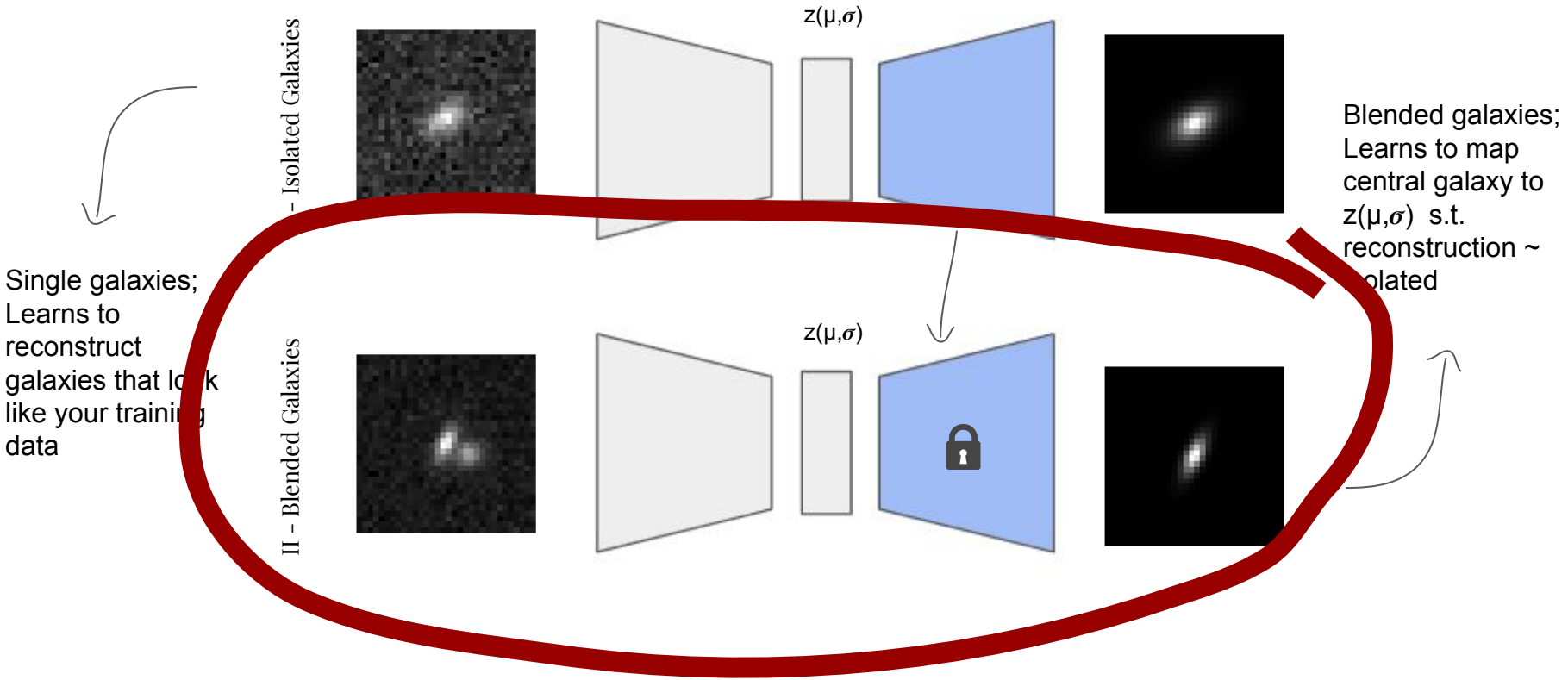
Total
~15,000

VAE Outperforms SExtractor in **~80%** of the cases;
For the rest, $\mu$ error on Flux (VAE) is **~17%**

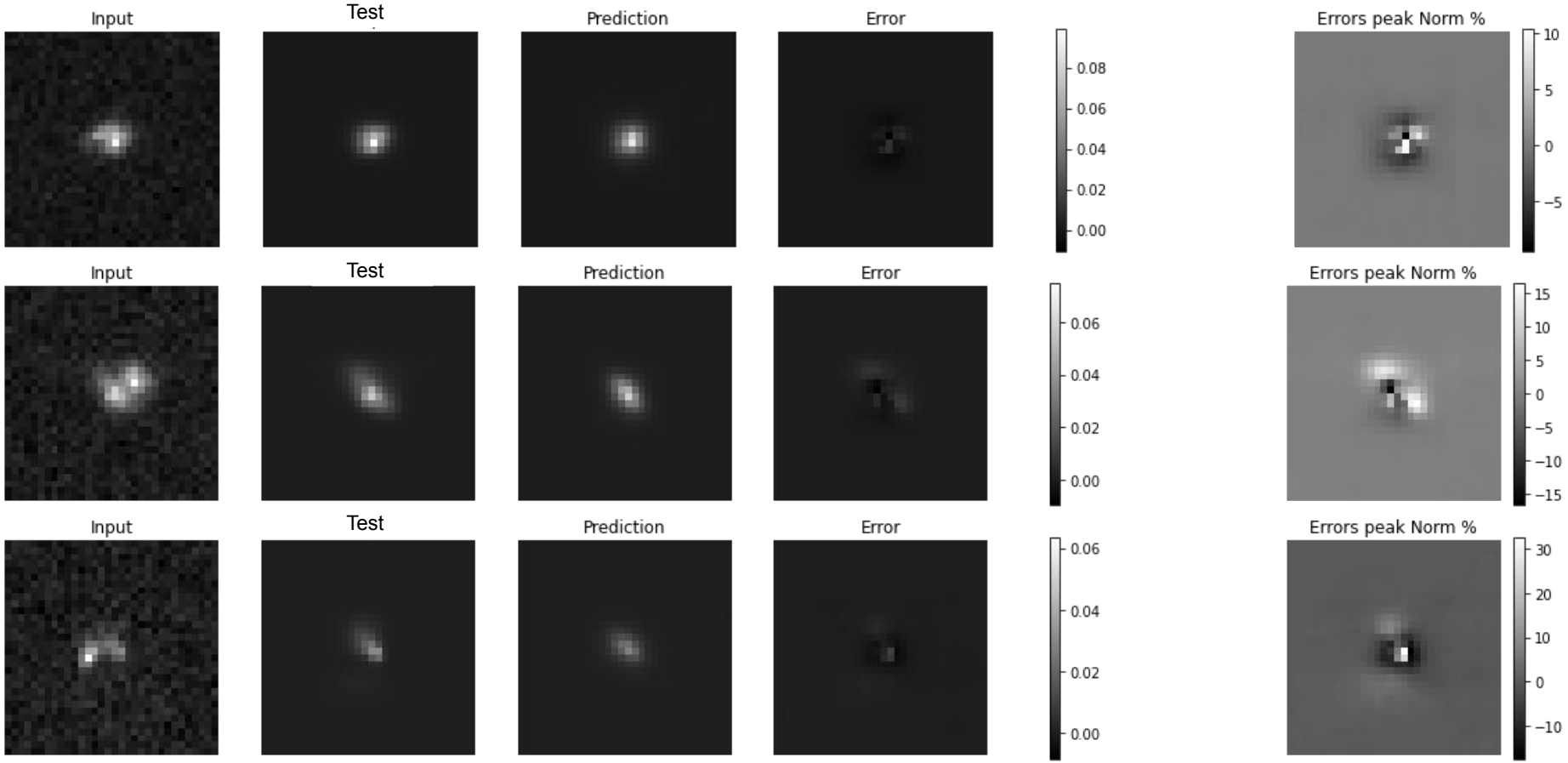


# VAEs for De-blending

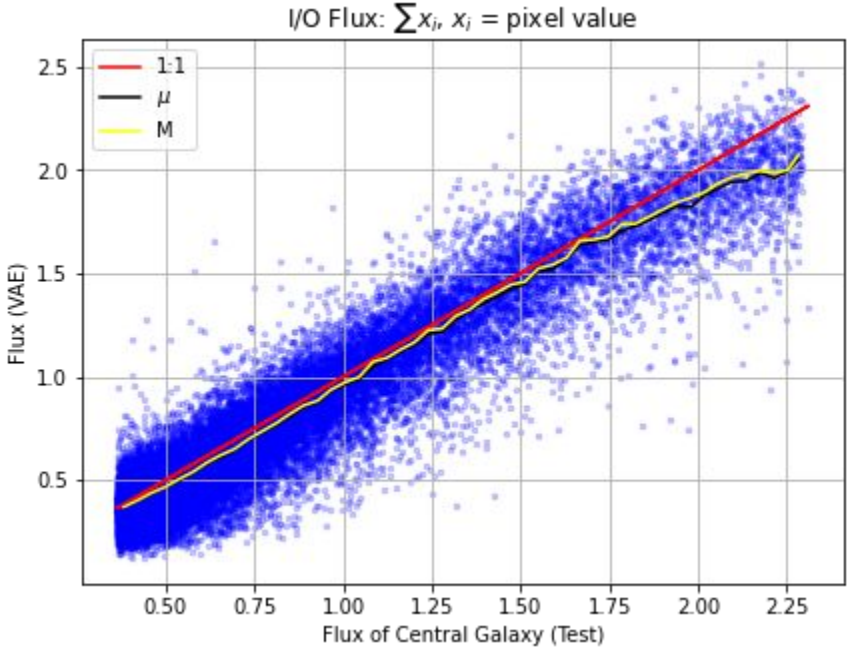


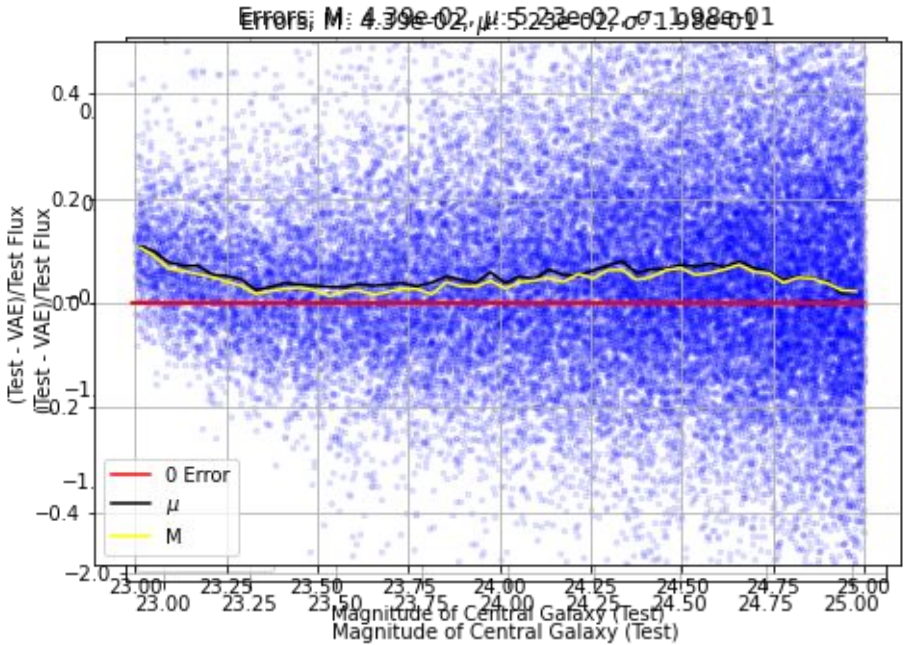
# VAE Results: II - Blended Galaxies (Light Profile Reconstruction)



# VAE Results: II – Blended Galaxies (Flux)

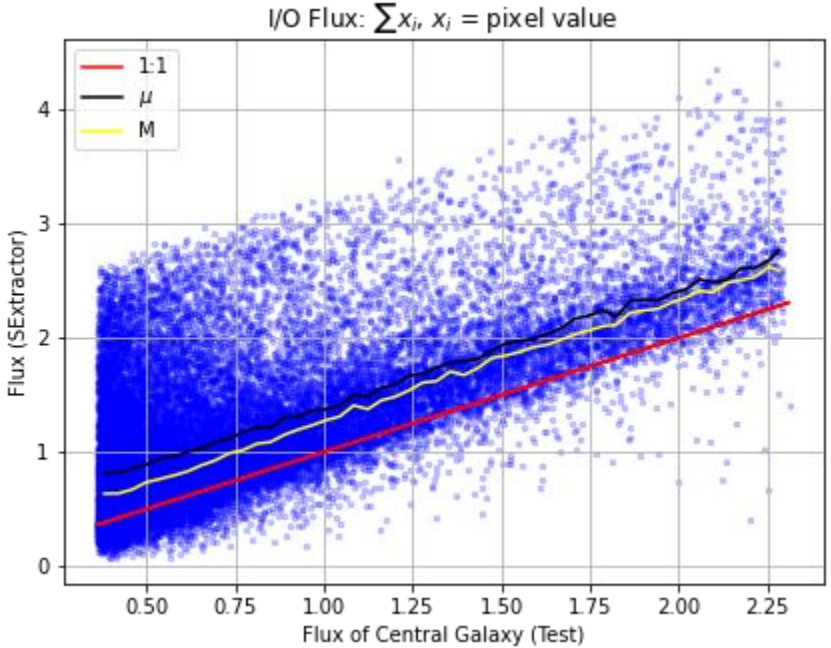
**Flux estimates**; The $\sigma$ of predicted fluxes of **central galaxies** on the **test images** is ~**20%** from the true value.

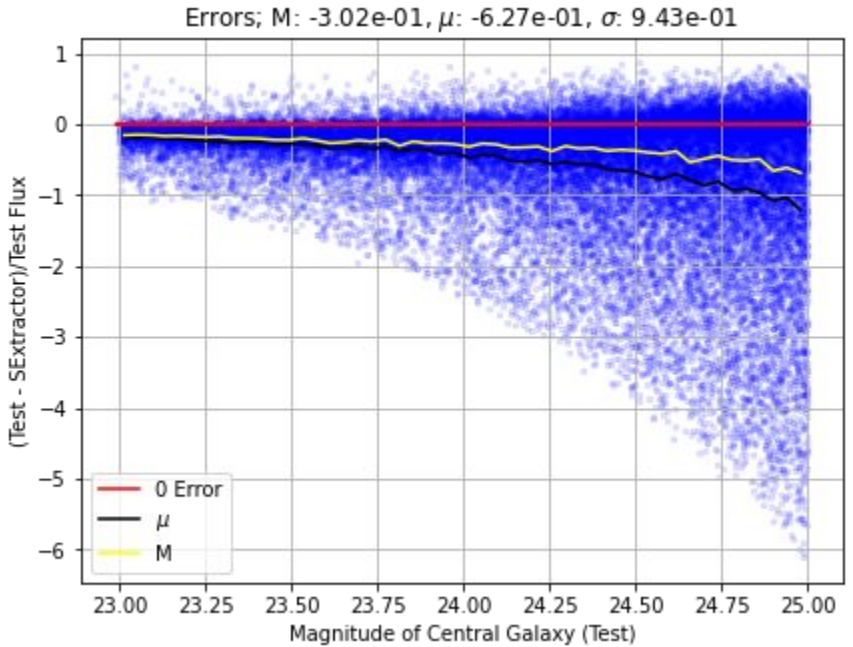


# SExtractor Results: Blended Galaxies (Flux)

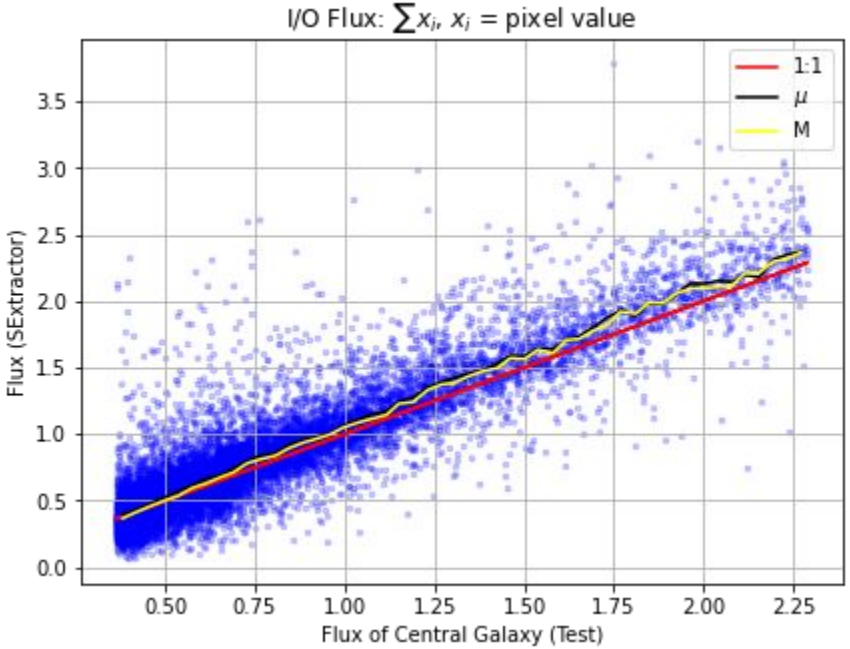
**Flux estimates**; The $\sigma$ of predicted fluxes of **central galaxies** on the **test images** is **~94%** from the true value.

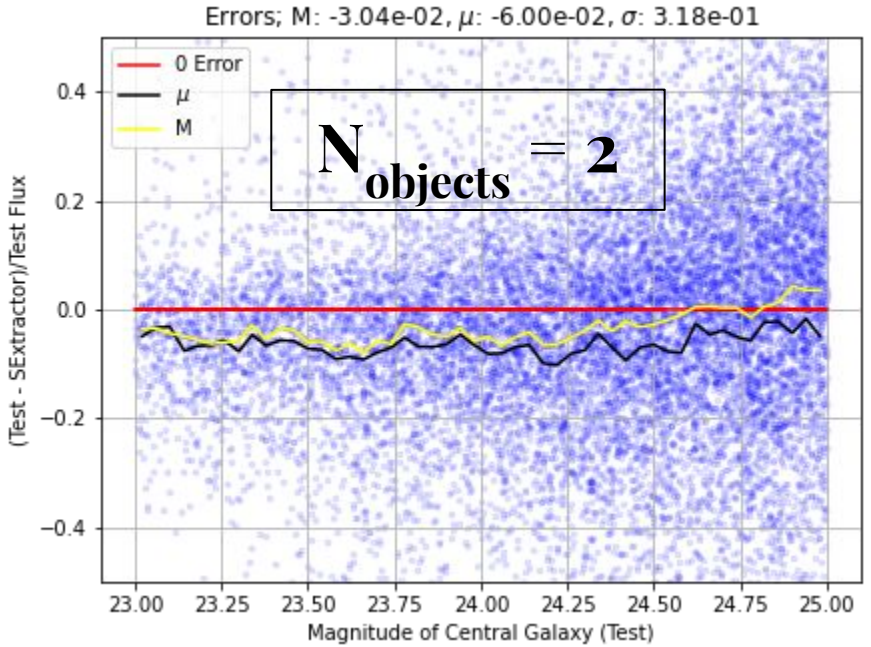


# SExtractor Results: Blended Galaxies (Flux)

**Flux estimates**; The $\sigma$ of predicted fluxes of **central galaxies** on the **test images** is ~**31%** from the true value.



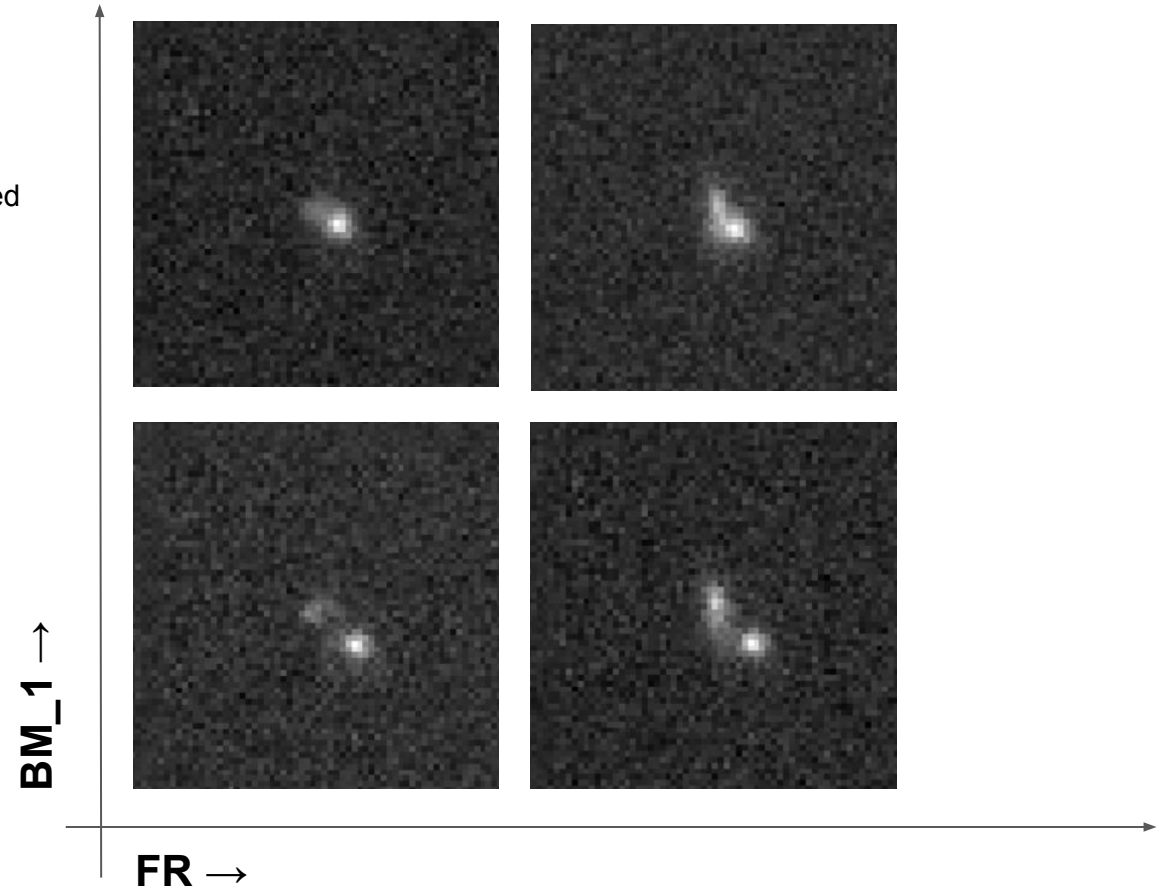
# Complexity of Blending

**BM_1** → Blended/Centre galaxy area

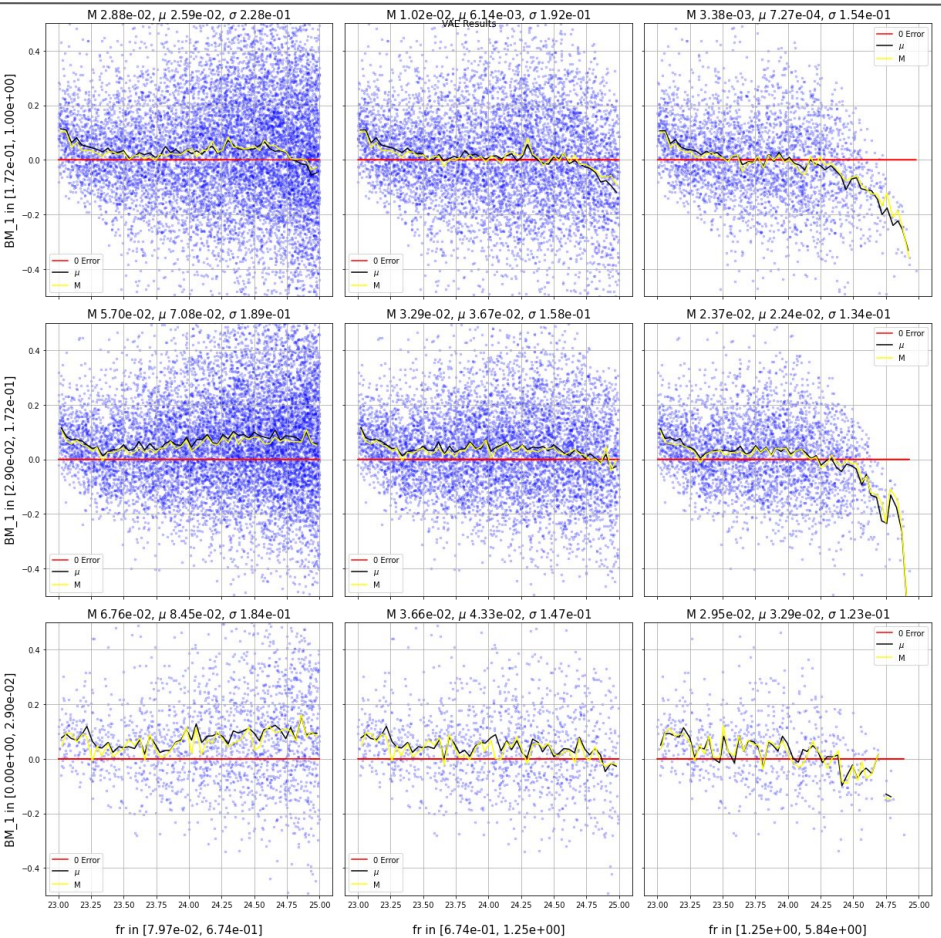
The smaller BM_1 is the less blending has occured

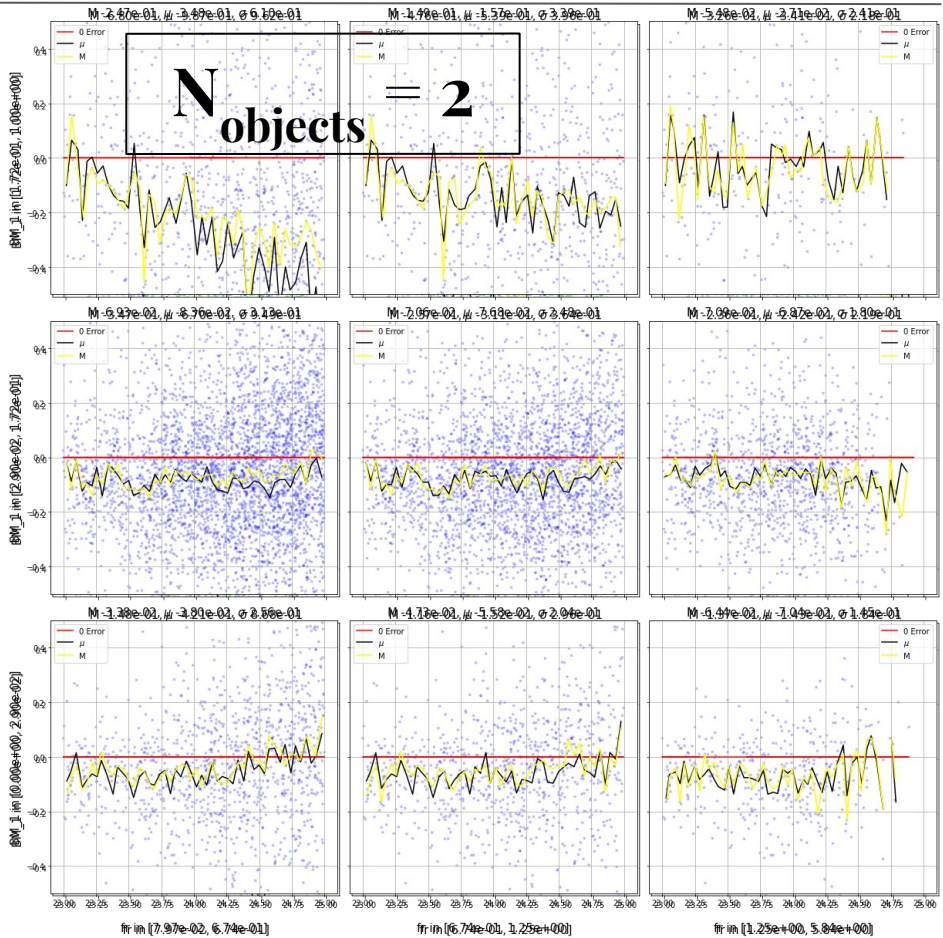
**FR** → Flux ratio of Centre/Companion

The closer FR is to 1, the more important it becomes to de-blend..



# Complexity of Blending : VAE v/s SExtractor



# Other things:

- Hyper-parameter tuning; Eg. Optimizing size of latent space, no of layers, loss ratios (reconstruction loss, kl divergence loss)
- Learning Rate Schedulers, Augmentation (Flips,Rotations, etc.)
- Custom losses for higher reconstruction accuracy; Eg. reconstruction loss with higher weight in the centre of the stamps
- Pipeline for De-blending EUCLID images - Train on isolated galaxies, deblend others

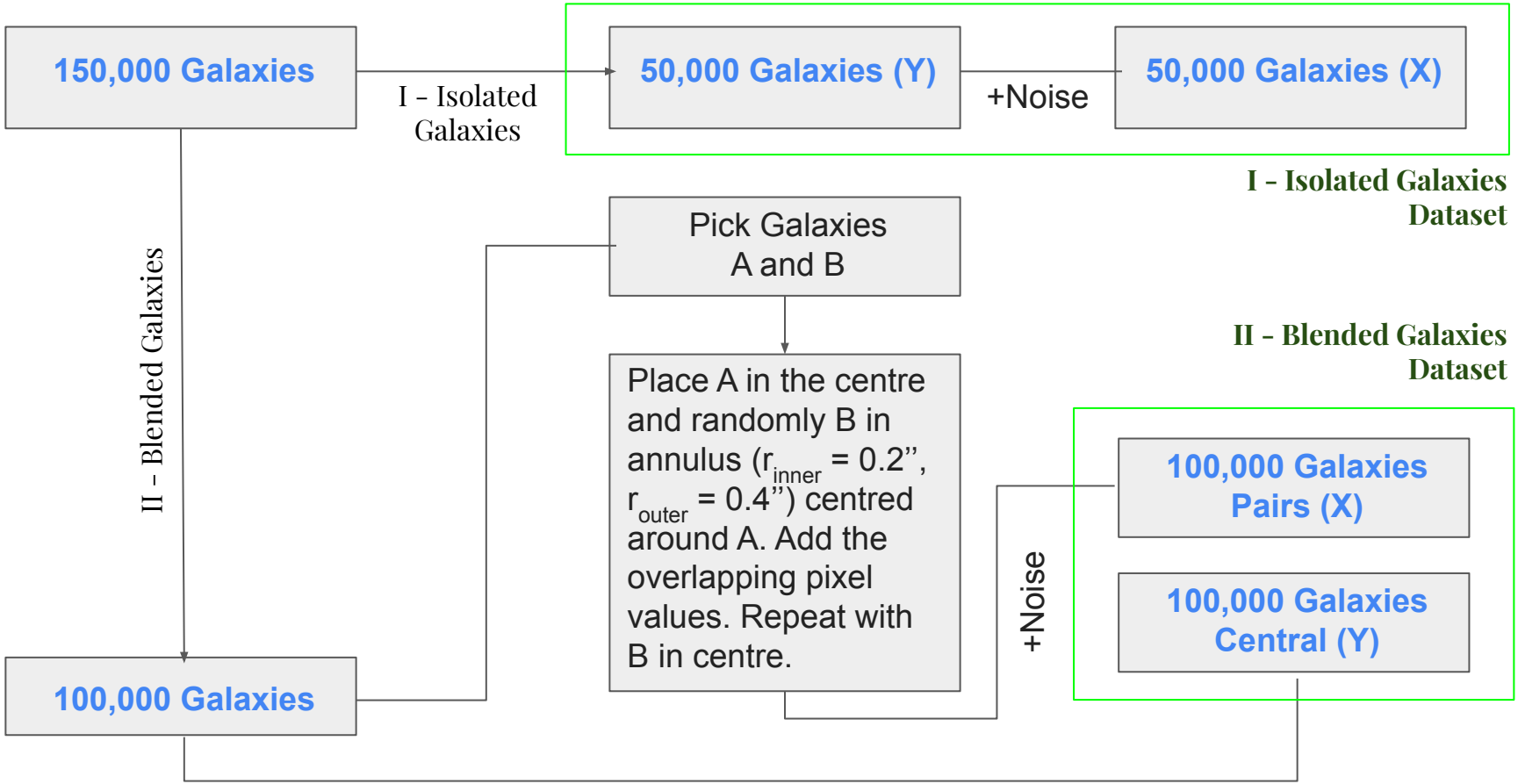
***Thanks for listening!***

# References


1. Dawson, W.A. and Schneider, M.D. (2014). Complementarity of lsst and wfirst: Regarding object blending U.S. Department of Energy Office of Scientific and Technical Information, doi: 10.2172/1122197
2. Reiman, D.M. and Göhre, B.E. (2019). Deblending galaxy superpositions with branched generative adversarial networks Monthly Notices of the Royal Astronomical Society, doi: 10.1093/mnras/stz575
3. Merlin, E. et al. (2015). T-PHOT: A new code for PSF-matched, prior-based, multiwavelength extragalactic deconfusion photometry Astronomy &amp Astrophysics, doi: 10.1051/0004-6361/201526471
4. Bertin, E. and Arnouts, S. (1996). SExtractor: Software for source extraction Astronomy &amp Astrophysics supplement series, doi: 10.1051/aas:1996164
5. Arcelin, B., Doux, C. et al. (2020). Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach Monthly Notices of the Royal Astronomical Society, doi: 10.1093/mnras/staa3062.
6. Schreiber, C. et al. (2017). EGG: hatching a mock Universe from empirical prescriptions Astronomy &amp Astrophysics, doi: 10.1051/0004-6361/201629123
7. Rowe, B., Jarvis, M. et al. (2014). GalSim: The modular galaxy image simulation toolkit arXiv, doi: 10.48550/arxiv.1407.7676

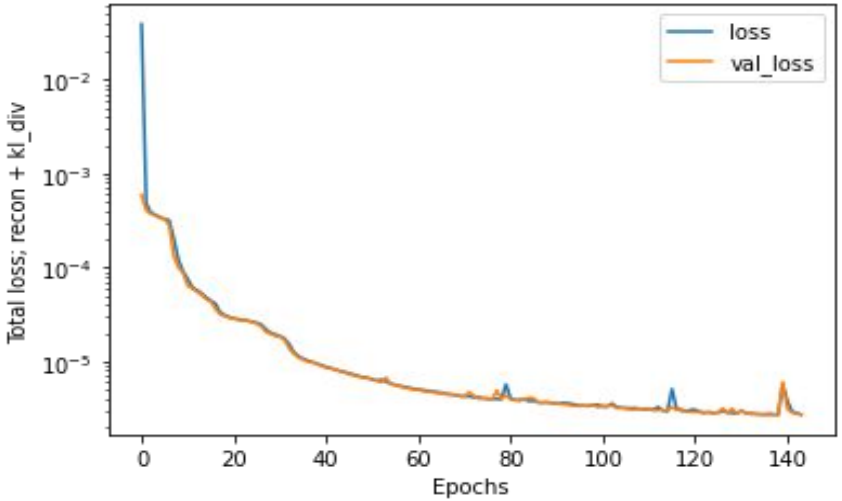
# *Extras*

# Analytic Simulations - Dataset

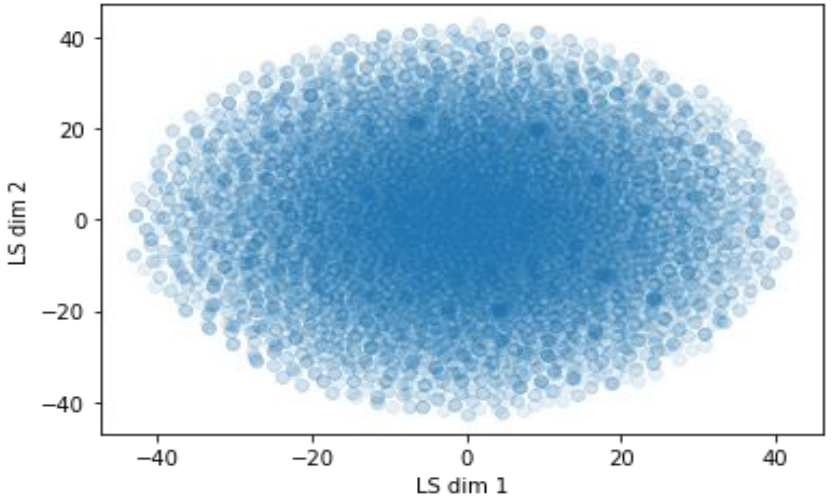


# Isolated Galaxies: Training

**Loss curve**; The total loss is the sum of the reconstruction loss (for reproducing images accurately) and kl divergence loss (for ensuring latent space distribution has good properties).




**Visualization of Latent Space**; TSNE is used to visualize the 100-dimensional latent space in 2-dimensions. These dimensions can be thought of as hidden features that the network has learned.

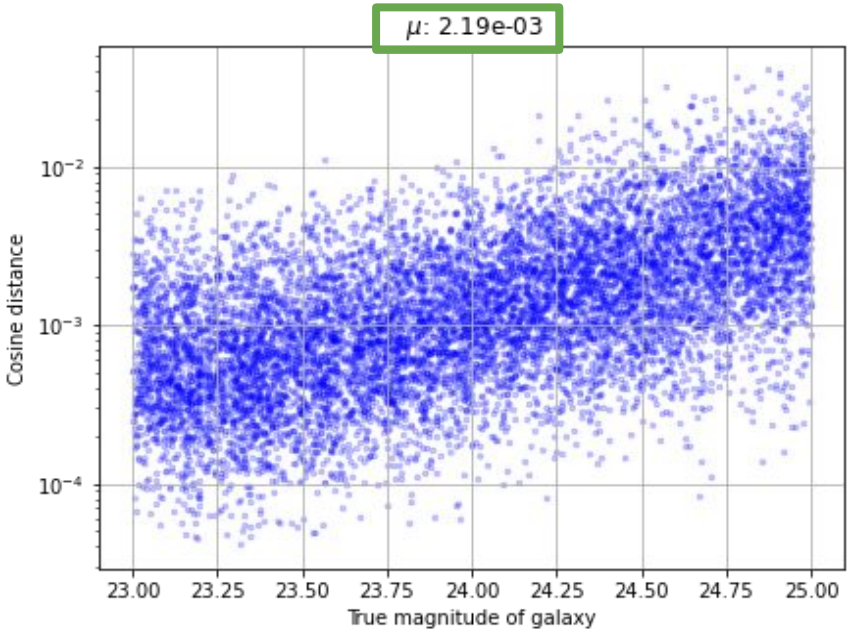


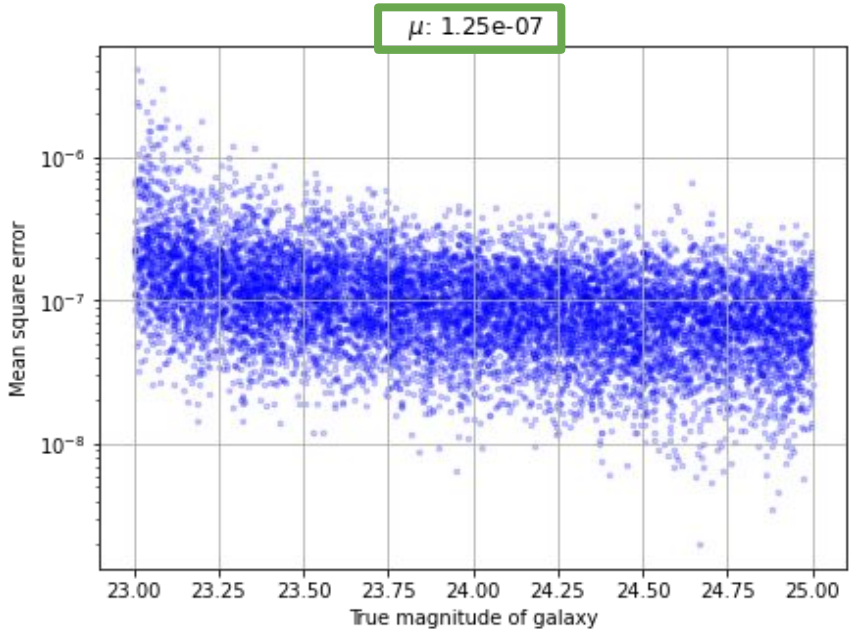

L.J.P. van der Maaten and G.E. Hinton. **Visualizing High-Dimensional Data Using t-SNE**.

# Isolated Galaxies: Results

$$\text{Cosine distance} = 1 - \frac{\sum_i^{\text{pixel}} x_i y_i}{\sqrt{\sum_i^{\text{pixel}} x_i^2}\sqrt{\sum_i^{\text{pixel}} y_i^2}}$$

$$\text{MSE} = \frac{1}{\text{no. pixels}} \sum_i^{\text{pixel}} (x_i - y_i)^2$$

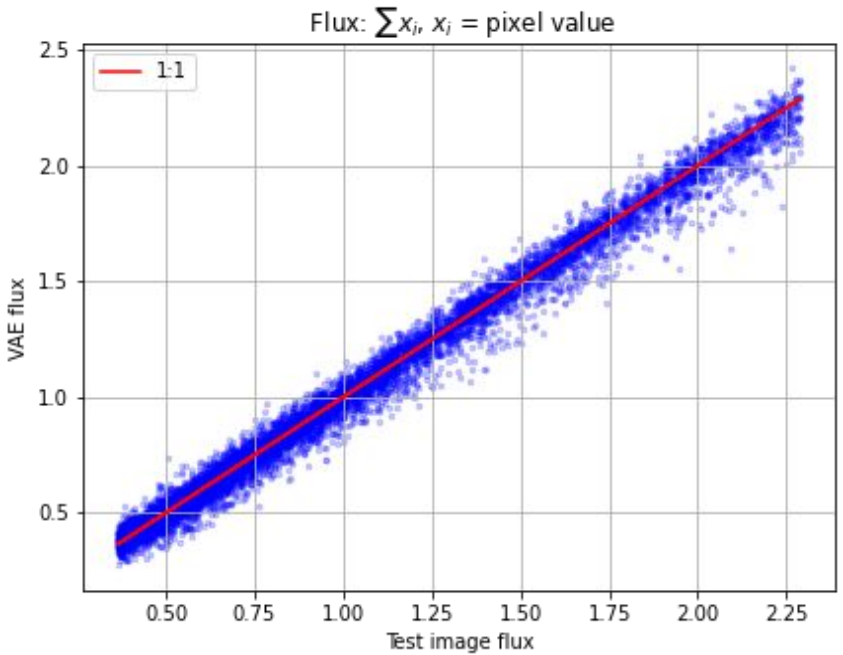


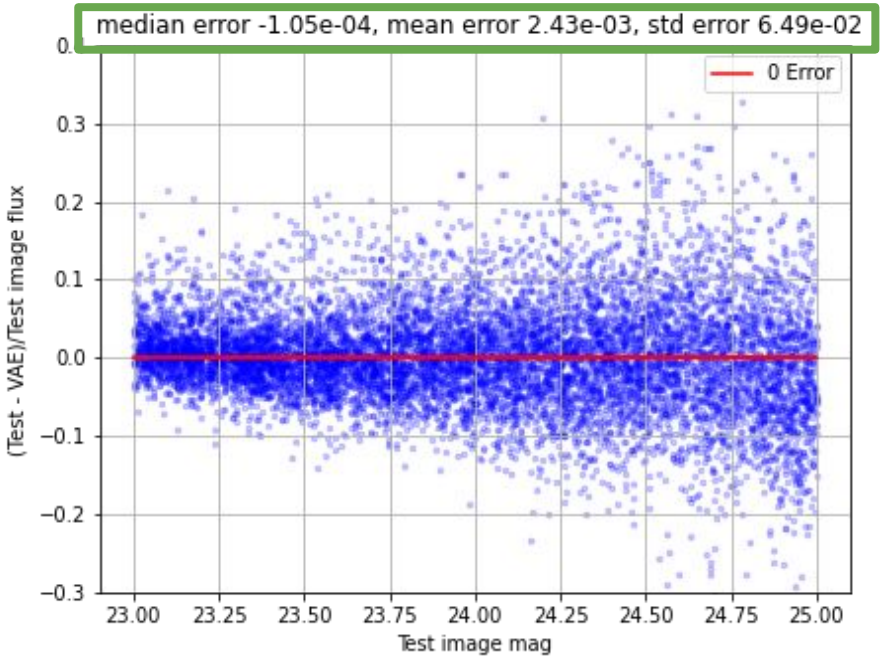




*NOTE:* CD is a relative error; relative noise in fainter galaxies is higher; hence, increasing error with mag

# Isolated Galaxies: Results

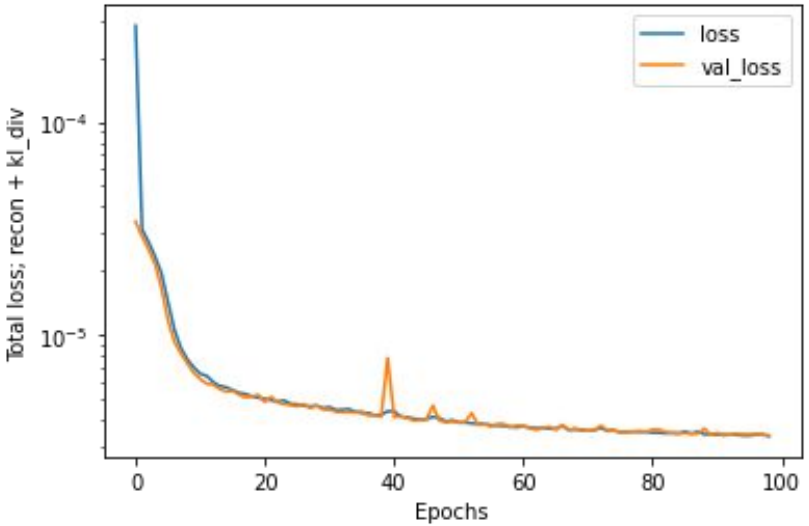
**Flux estimates**; The predicted fluxes on the **test images** is within **10%** for most galaxies.



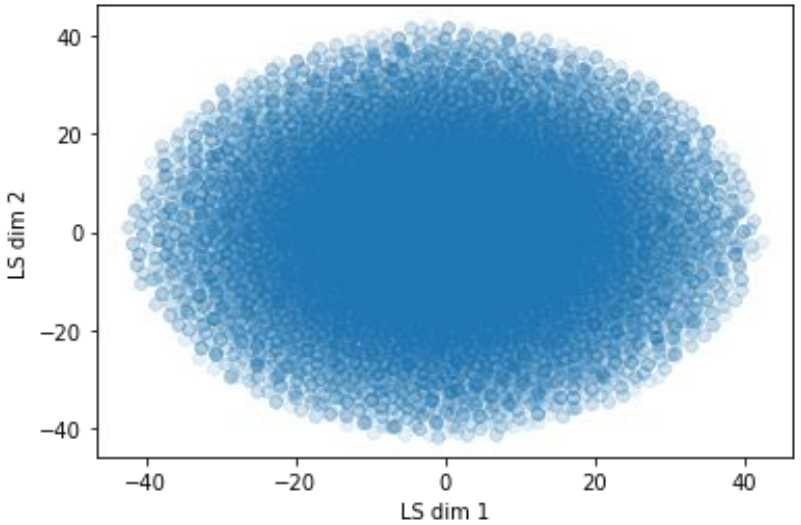



# Blended Galaxies: Training

**Loss curve**; The total loss is the sum of the reconstruction loss (for reproducing images accurately) and kl divergence loss (for ensuring latent space distribution has good properties).



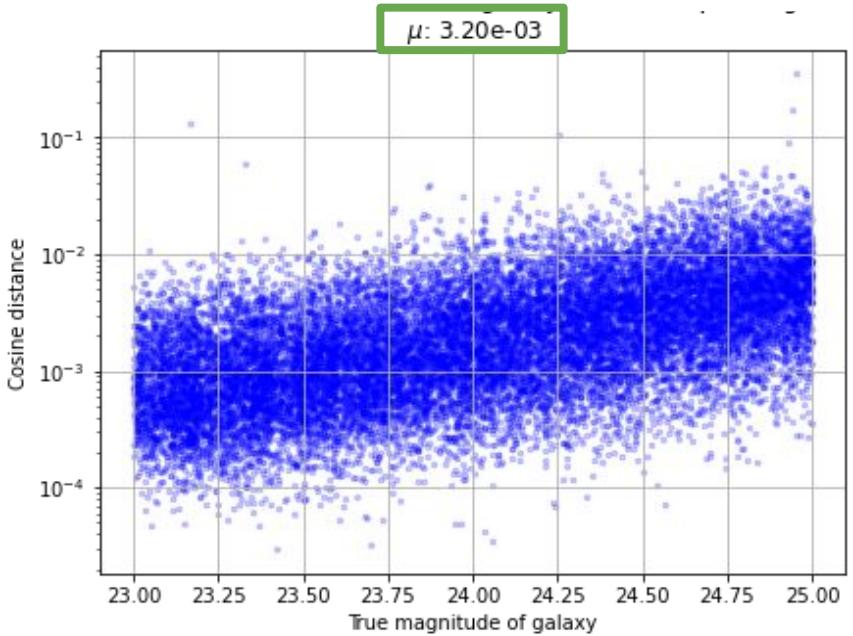

**Visualization of Latent Space**; TSNE is used to visualize the 100-dimensional latent space in 2-dimensions. These dimensions can be thought of as hidden features that the network has learned.

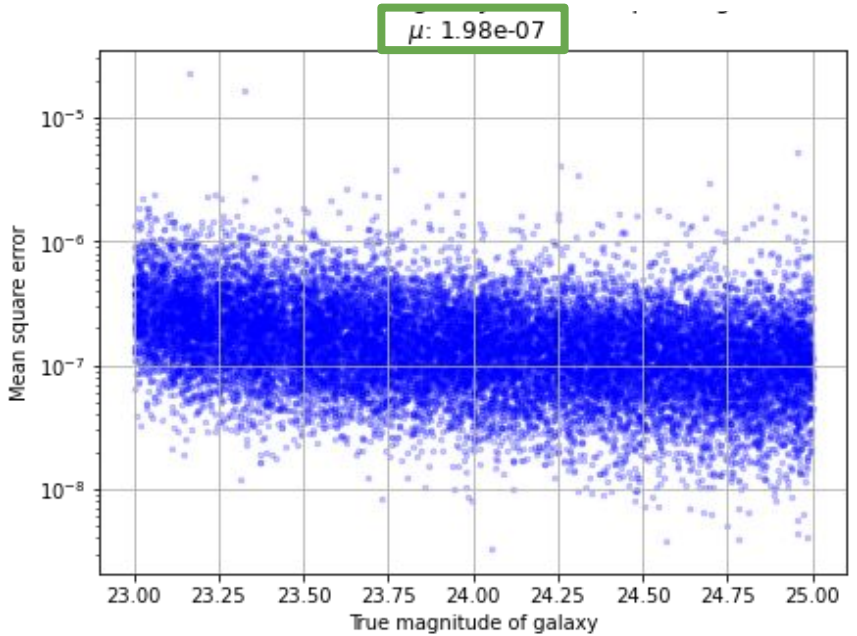


# Blended Galaxies: Results

$$\text{Cosine distance} = 1 - \frac{\sum_i^{\text{pixel}} x_i y_i}{\sqrt{\sum_i^{\text{pixel}} x_i^2}\sqrt{\sum_i^{\text{pixel}} y_i^2}}$$

$$\text{MSE} = \frac{1}{\text{no. pixels}} \sum_i^{\text{pixel}} (x_i - y_i)^2$$

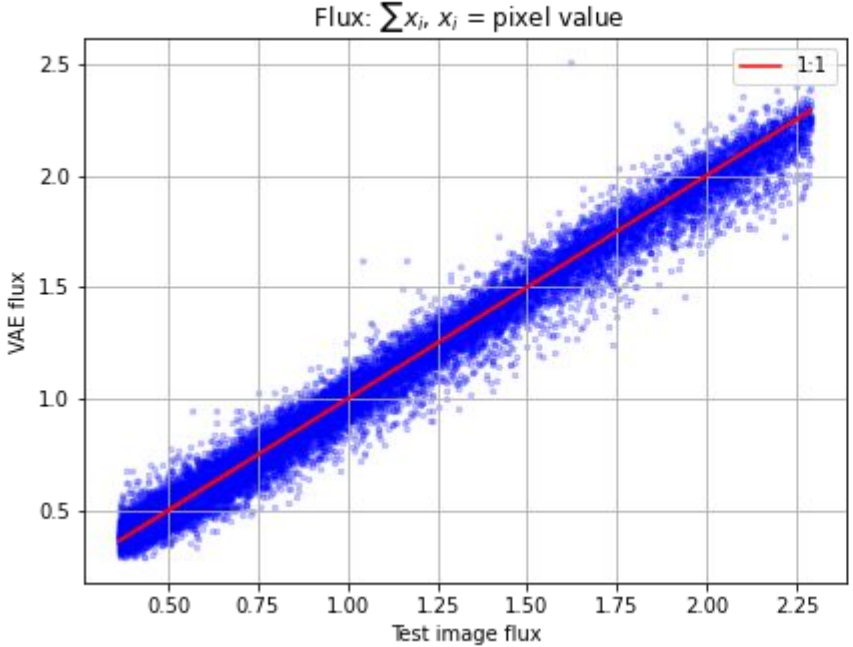


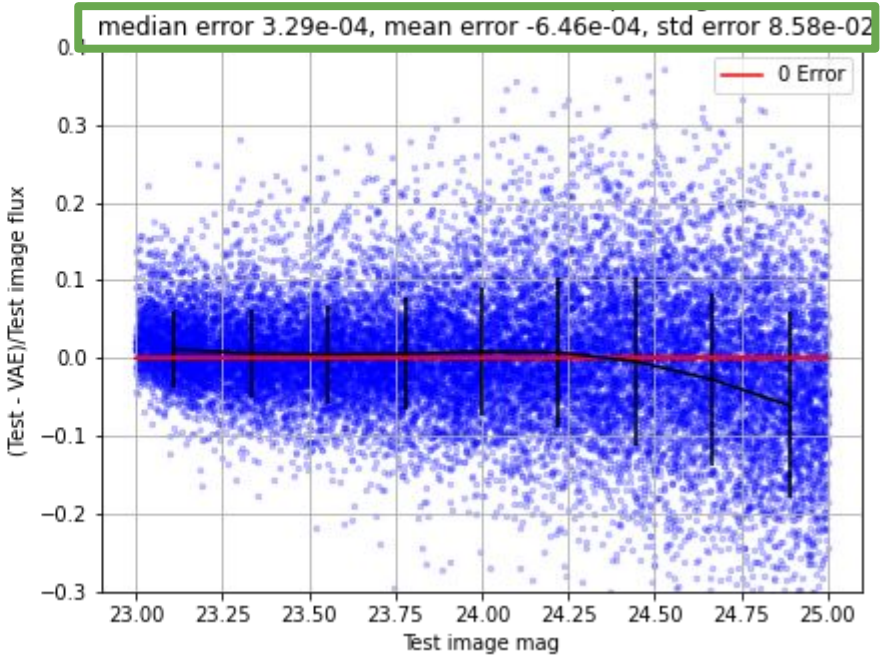




*NOTE:* CD is a relative error; relative noise in fainter galaxies is higher; hence, increasing error with mag

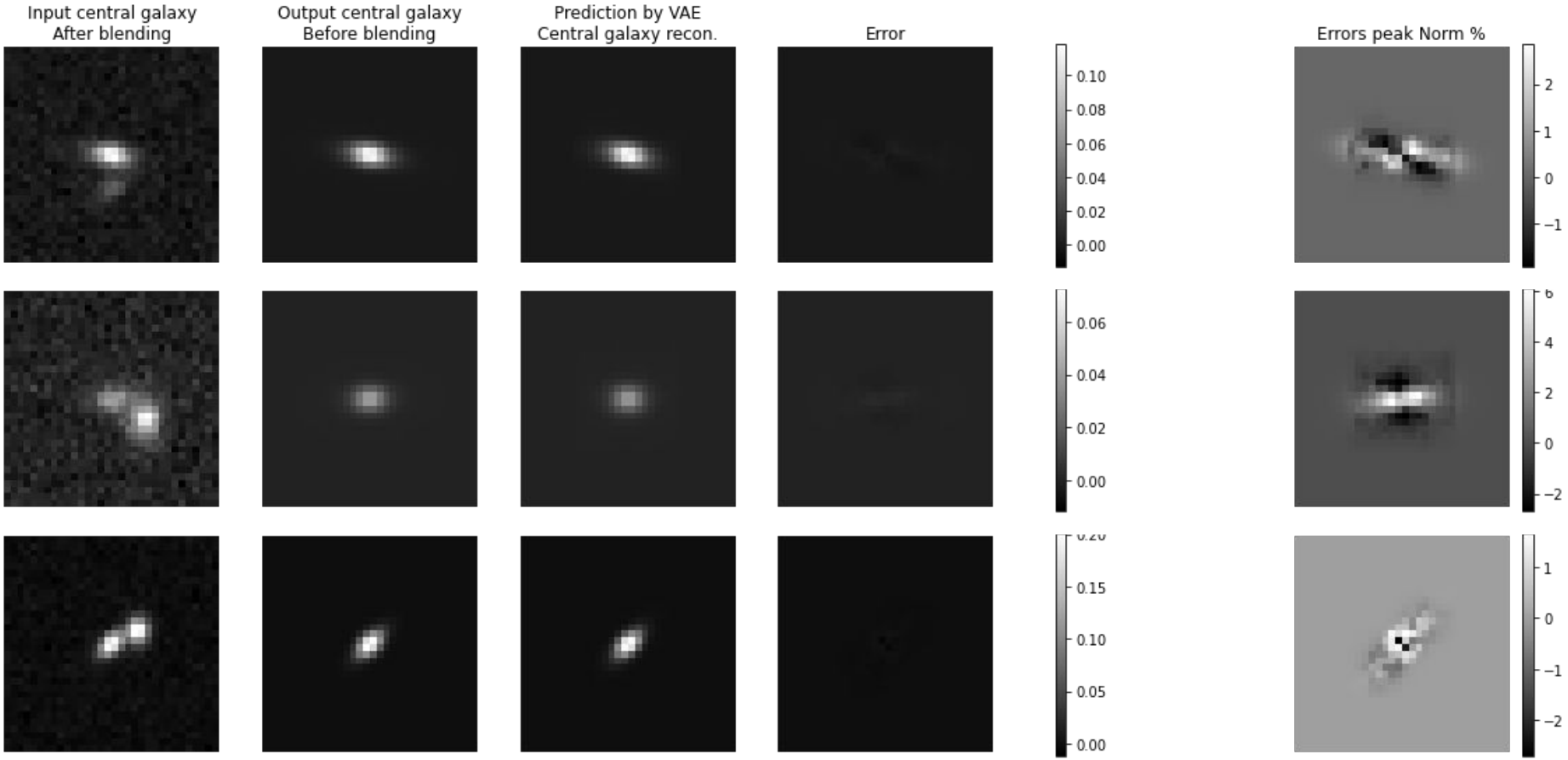
# Blended Galaxies: Results

**Flux estimates**; The predicted fluxes  on the **test images** is within **15%** for most galaxies.



# Examples of Deblending



# Realistic images - Dataset (Augmentation)

Rotation, Horizontal & Vertical Flips (Each with p = 0.5). Rate = 100%