



# Deeper, Sharper, Faster: Application of Efficient Transformer to Galaxy Image Restoration

Hyosun Park<sup>1</sup> , Yongsik Jo<sup>2</sup> , Seokun Kang<sup>2</sup> , Taehwan Kim<sup>2</sup> , and M. James Jee<sup>1,3</sup> 

<sup>1</sup> Department of Astronomy, Yonsei University, Seoul, Republic of Korea; [hyosun.park@yonsei.ac.kr](mailto:hyosun.park@yonsei.ac.kr), [mkjee@yonsei.ac.kr](mailto:mkjee@yonsei.ac.kr)

<sup>2</sup> Artificial Intelligence Graduate School, UNIST, Ulsan, Republic of Korea

<sup>3</sup> Department of Physics and Astronomy, University of California, Davis, CA, USA

Received 2024 March 29; revised 2024 May 29; accepted 2024 May 31; published 2024 August 23

## Abstract

The Transformer architecture has revolutionized the field of deep learning over the past several years in diverse areas, including natural language processing, code generation, image recognition, and time-series forecasting. We propose to apply Zamir et al.'s efficient transformer to perform deconvolution and denoising to enhance astronomical images. We conducted experiments using pairs of high-quality images and their degraded versions, and our deep learning model demonstrates exceptional restoration of photometric, structural, and morphological information. When compared with the ground-truth James Webb Space Telescope images, the enhanced versions of our Hubble Space Telescope-quality images reduce the scatter of isophotal photometry, Sérsic index, and half-light radius by factors of 4.4, 3.6, and 4.7, respectively, with Pearson correlation coefficients approaching unity. The performance is observed to degrade when input images exhibit correlated noise, point-like sources, and artifacts. We anticipate that this deep learning model will prove valuable for a number of scientific applications, including precision photometry, morphological analysis, and shear calibration.

*Unified Astronomy Thesaurus concepts:* [Galaxy physics \(612\)](#); [Galaxy photometry \(611\)](#); [Galaxy structure \(622\)](#); [Astronomy image processing \(2306\)](#); [Neural networks \(1933\)](#)

## 1. Introduction

A deeper and sharper image of an astronomical object offers an opportunity for gaining new insights and understanding of it. One of the most notable examples is the advent of the James Webb Space Telescope (JWST), launched in 2021, which has since continuously expanded our knowledge of the universe through its discoveries, thanks to its unprecedented resolution and depth. This trend is reminiscent of a similar era three decades ago when the Hubble Space Telescope (HST) became the most powerful instrument available to us at that time.

Astronomers' ongoing efforts to enhance the depth and clarity of astronomical imaging extend beyond advancements in instrumentation to include significant developments in the algorithmic domain. Early techniques relied on straightforward Fourier deconvolution techniques (e.g., Simkin 1974; Jones & Wykes 1989). A primary challenge with this approach is noise amplification, and linear filtering was suggested to remedy the frequency-dependent signal-to-noise ratio issue (e.g., Tikhonov & Goncharov 1987). However, the filtering method provided only band-limited results. With the development of computer technologies, more computationally demanding, sophisticated approaches based on the Bayesian principle with various regularization schemes were introduced (e.g., Richardson 1972; Lucy 1974; Shepp & Vardi 1982). Some implementations of these Bayesian approaches were shown to outperform the early Fourier deconvolution methods. Nevertheless, the regularization could not recover both compact and extended features simultaneously well. Multiresolution or spatially adaptive approaches were suggested to overcome this limitation (e.g., Wakker & Schwarz 1988; Yan et al. 2012).

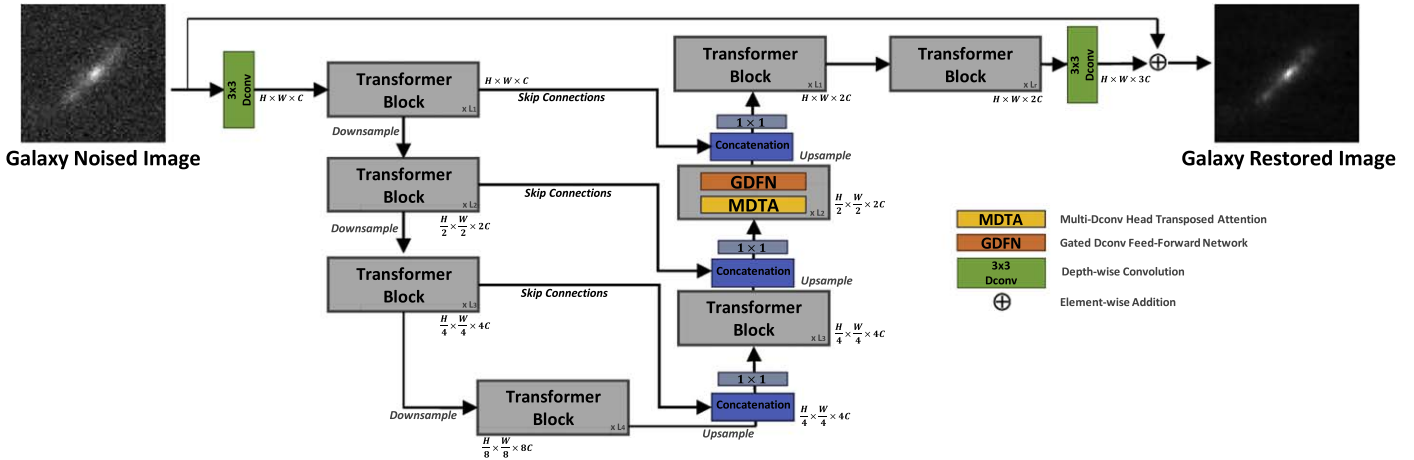
In recent years, the emergence of deep learning has significantly impacted the general field of image restoration and enhancement. Also, a growing number of studies are reporting notable results in astronomical contexts as well (e.g., Schawinski et al. 2017; Sureau et al. 2020; Lanusse et al. 2021; Akhaury et al. 2022; Sweere et al. 2022). Deep learning algorithms, particularly convolutional neural networks (CNNs; Krizhevsky et al. 2012), have shown promising results in automatically learning complex features from images and effectively enhancing their depth and resolution (e.g., Zhang et al. 2017, 2022; Díaz Baso et al. 2019; Elhakiem et al. 2021). By training on large data sets of both raw and enhanced images, deep learning models can learn intricate patterns and relationships within the data, allowing for more precise and tailored image enhancement.

Deep learning techniques offer advantages over traditional methods, such as Fourier deconvolution and Bayesian approaches, by providing greater flexibility and adaptability to diverse data sets. The integration of CNNs with other deep learning architectures, such as generative adversarial networks (Goodfellow et al. 2014) and recurrent neural networks (Williams & Zipser 1989), has further expanded the capabilities of image enhancement (e.g., Ledig et al. 2017; Schawinski et al. 2017; Tripathi et al. 2018; Alsaiani et al. 2019; Rajeev et al. 2019; Wang et al. 2020; Liu et al. 2021; Tran et al. 2021; Kalele 2023). These hybrid approaches enable the generation of highly realistic and detailed images, pushing the boundaries of what is achievable with traditional methods alone. One of the outstanding limitations of the CNN-based model is its restricted receptive field. That is, long-range pixel correlations may not effectively be captured by the model. Another critical drawback is its limited adaptability to input content during inference.

The Transformer architecture (Vaswani et al. 2017), which has revolutionized the field of deep learning over the past



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



**Figure 1.** Restormer architecture employed in this paper for galaxy image restoration. The architecture uses a multiscale hierarchical design incorporating efficient Transformer blocks. The two core modules of the Transformer block are MDTA and GDFN (Zamir et al. 2022).

several years in diverse areas including the well-known large language models, can be considered a potent alternative to overcome these limitations of the CNN-based models. However, with its original implementation structure comprised of so-called self-attention layers, it is infeasible to apply the Transformer model to large images because the computing complexity increases quadratically with the number of pixels.

Zamir et al. (2022) devised an innovative scheme to substitute the original self-attention block with the multi-Dconv transposed attention (MDTA) block. The MDTA block, implementing self-attention in the feature domain rather than the pixel domain, ensures that the complexity increases only linearly with the number of pixels, making its application to large images feasible. Zamir et al. (2022) demonstrated that their model, named Restormer, attained state-of-the-art results for image deraining, single-image motion deblurring, defocus deblurring, and image denoising. However, its performance has not been evaluated in the astronomical context.

In this paper, we propose to apply Zamir et al.’s efficient transformer to perform deconvolution and denoising to enhance astronomical images. In particular, we investigate the feasibility of enhancing HST images to achieve the quality of JWST ones in both resolution and depth. We build our model based on Zamir et al. (2022)’s implementation and employ the transfer learning approach, initially pretraining the model using simplified galaxy images and then fine-tuning it using realistic galaxy images.

Our paper is structured as follows. In Section 2, we describe the overall architecture of Restormer and the implementation details employed in the current galaxy restoration model. Section 3 explains how we prepare training data sets. Results are presented In Section 4. We show the results when the model is applied to real HST images In Section 5 and discuss the limitations In Section 6 before we conclude In Section 7.

## 2. Method

### 2.1. Overview

Throughout the paper, we use the term restoration to refer to the process that simultaneously improves resolution and reduces noise. Our goal is to restore the HST-quality image to the JWST-quality image based on the Restormer

implementation (Zamir et al. 2022) of the Transformer architecture (Vaswani et al. 2017) shown in Figure 1. We first briefly review the encoder–decoder architecture in Section 2.2. Section 2.3 describes the Transformer architecture including the Zamir et al. (2022)’s implementation. Implementation details are presented in Section 2.4.

### 2.2. Encoder–Decoder Architecture

The encoder–decoder architecture allows neural networks to learn to map input data to output data in a structured and hierarchical manner. The encoder captures characterizing features from the input data and encodes them into a compressed representation, while the decoder reconstructs or generates the desired output based on this encoded representation. This architecture has been widely used in various applications, including image-to-image translation, image segmentation, language translation, and more.

U-net (Ronneberger et al. 2015) is a classic example of an encoder–decoder architecture based on CNN. It consists of a contracting path, which serves as the encoder, an expanding path, which functions as the decoder, and skip connections that link corresponding layers between the contracting and expanding paths. In the convolutional layers of the contracting path, the dimension is reduced by increasing the number of channels to capture essential image features. The expanding path employs only low-dimensional encoded information to decrease channel numbers and increase dimensions, aiming to restore high-dimensional images. To mitigate information loss in the contracting path, the skip connection is employed to concatenate features obtained from each layer of the encoding stage with each layer of the decoding stage.

### 2.3. Transformers for Image Restoration

In Transformer, the encoder consists of multiple layers of self-attention mechanisms followed by position-wise feed-forward neural networks. “Attention” refers to a mechanism that allows models to focus on specific parts of input data while processing it. It enables the model to selectively weigh different parts of the input, giving more importance to relevant information and ignoring irrelevant or less important parts. The key idea behind attention is to dynamically compute weights for different parts of the input data, such as words in a

sentence or pixels in an image, based on their relevance to the current task. In self-attention, each element (e.g., word or pixel) in the input sequence is compared to every other element to compute attention weights, which represent the importance of each element with respect to others. These attention weights are then used to compute a weighted sum of the input elements, resulting in an attention-based representation that highlights relevant information.

The Transformer decoder also consists of multiple layers of self-attention mechanisms, along with additional attention mechanisms over the encoder’s output. The decoder predicts one element of the output sequence at a time, conditioned on the previously generated elements and the encoded representation of the input sequence.

The Transformer architecture was initially proposed and applied to the task of machine translation, which involves translating text from one language to another. The success of the Transformer in machine translation tasks demonstrated its effectiveness in capturing long-range dependencies in sequences and handling sequential data more efficiently than traditional architectures. This breakthrough sparked widespread interest in the Transformer architecture, leading to its adoption and adaptation for various image-processing tasks. Transformers show promising results in tasks such as image classification, object detection, semantic segmentation, and image generation, traditionally dominated by CNNs. Transformer models capture long-range pixel correlations more effectively than CNN-based models.

However, using the Transformer model on large images becomes challenging with its original implementation, which applies self-attention layers on pixels. This is because the computational complexity escalates quadratically with the pixel count. Zamir et al. (2022) overcame this obstacle by substituting the original self-attention block with the MDTA block, which implements self-attention in the feature domain and makes the complexity increase only linearly with the number of pixels. We propose to use Zamir et al. (2022)’s efficient Transformer `Restormer` to apply deconvolution and denoising to astronomical images. We briefly describe the two core components of `Restormer` in Sections 2.3.1 and 2.3.2. Readers are referred to Zamir et al. (2022) for more technical details.

### 2.3.1. MDTA Block

MDTA stands as a crucial module within `Restormer`. By performing self-attention in the channel dimension, MDTA calculates interactions between channels in the input feature map, creating query-key interactions. Through this process, MDTA effectively models interactions between channels in the input feature map, facilitating the learning of the global context necessary for image restoration tasks.

MDTA also employs depth-wise convolution to accentuate local context. This enables MDTA to emphasize the local context of the input image, ultimately allowing for the modeling of both global and local contexts.

### 2.3.2. GDFN Block

GDFN, short for gated-Dconv feed-forward network, stands as another crucial module within `Restormer`. Utilizing a gating mechanism to enhance the feed-forward network,

GDFN offers improved information flow, resulting in high-quality outcomes for image restoration tasks.

GDFN controls the information flow through gating layers, composed of element-wise multiplication of two linear projection layers, one of which is activated by the Gaussian error linear unit nonlinearity. This allows GDFN to suppress less informative features and hierarchically transmit only valuable information. Similar to the MDTA module, GDFN employs local content mixing. Through this, GDFN emphasizes the local context of the input image, providing a more robust information flow for enhanced results in image restoration tasks.

## 2.4. Implementation Details

In all experiments, we use the following training parameters unless mentioned otherwise. Our `Restormer` employs a four-level encoder–decoder. From level 1 to level 4, the number of Transformer blocks is [4, 6, 6, 8], attention heads in MDTA are [1, 2, 4, 8], and the number of channels is [48, 96, 192, 384]. The refinement stage contains four blocks. The channel expansion factor in GDFN is  $r = 2.66$ . We train models with the AdamW (Loshchilov & Hutter 2019) optimizer ( $B1 = 0.9$ ,  $B2 = 0.999$ , and a weight decay of  $10^{-4}$ ) and L1 loss for 300,000 iterations with the initial learning rate  $3 \times 10^{-4}$  gradually reduced to  $10^{-6}$  with the cosine annealing.

Every input image is min–max normalized. This normalization scheme is needed because astronomical images span a wide dynamic range. Without the proper normalization, the training will be biased toward restoring the brightest galaxy images. Also, we find that this standardization helps to train our model more efficiently. Finally, the min–max normalization minimizes the effects of field-dependent calibration systematics.

We use a transfer learning approach shown in Figure 2, where a model trained on one dataset is reused for another related dataset. First, we train on the pretraining dataset (simplified galaxy images) for 150,000 iterations, followed by an additional 150,000 iterations on the fine-tuning dataset (realistic galaxy images). The batch size remains fixed at 64. Additionally, to compare and analyze the performance of training on individual data sets, we also conduct training separately solely based on either the pretraining or fine-tuning data sets. Our inference model is publicly available.<sup>4,5</sup>

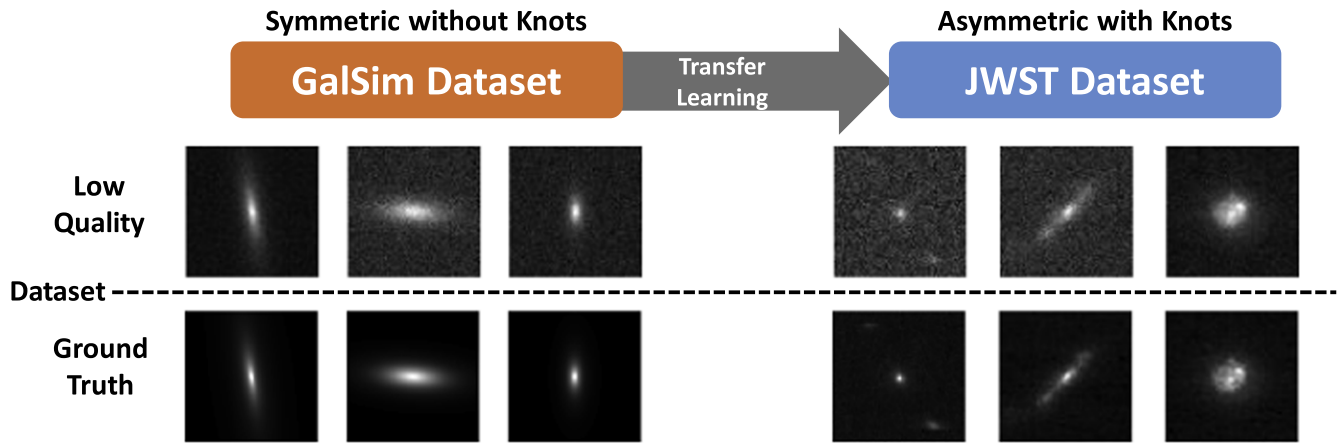
## 3. Data

The goal of the current study is to construct a Transformer model that enhances the quality of HST images to that of the JWST. Thus, our training dataset consists of pairs of JWST-quality ground-truth (GT) images and their degraded, low-quality (LQ) versions. We employ a transfer learning approach by first pretraining our model with GT–LQ pairs of simplified galaxy images and fine-tuning the pretrained model with GT–LQ pairs of realistic galaxy images. Therefore, we prepare two sets of training data sets.

The GT image of the pretraining dataset is generated with `GalSim` (Rowe et al. 2015) while its LQ counterpart is obtained by simply degrading the resolution of the GT image with the HST point-spread function (PSF) and adding Gaussian

<sup>4</sup> On GitHub: <https://github.com/JOYONGSIK/GalaxyRestoration>.

<sup>5</sup> On Zenodo: doi:10.5281/zenodo.11378660; Park et al. (2024).



**Figure 2.** Galaxy image restoration through transfer learning. Initially, we pretrain our model with LQ-GT pairs of simplified galaxy images created with GalSim. Then, this model is finetuned with LQ-GT pairs of realistic galaxy images sampled from deep JWST images. We find that this transfer learning scheme shows improved performance over the models trained with a single dataset.

noise. The input parameters of GalSim should not be arbitrary but must be chosen carefully to represent the structural features of real galaxies. This is achieved by harvesting real galaxy images from the HST archive and measuring their properties.

The GT image of the fine-tuning dataset is obtained by collecting galaxy images from the JWST archive. Following the same procedure in the pretraining dataset generation, we create their LQ counterparts.

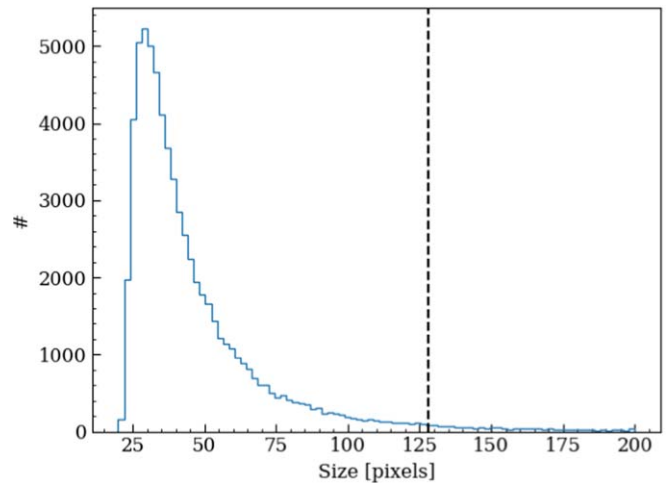
We provide detailed descriptions of the HST, GalSim, and JWST data sets in Sections 3.1, 3.2, and 3.3, respectively. Below we briefly outline the purpose of each dataset.

1. HST dataset: To define the properties of galaxies for creating synthetic images and determine the appropriate noise level for generating low-quality images. This dataset is not used in training.
2. GalSim dataset: To pretrain the model on noise-free synthetic images of galaxies to learn core features.
3. JWST dataset: To fine-tune the model using real JWST galaxy images to learn diverse forms and characteristics of galaxies not covered by the GalSim dataset.

### 3.1. HST Dataset

We collect multiband galaxy images from the following HST Advanced Camera for Surveys (ACS) deep fields: the Hubble Ultra Deep Field 12 (HUDF12), HUDF-Parallel 2 (HUDFP2<sup>6</sup>; Ellis et al. 2013; Koekemoer et al. 2013), and the Great Observatories Origins Deep Survey North and South (GOODS-N&S<sup>7</sup>; Illingworth et al. 2016; Whitaker et al. 2019) fields. The F606W, F775W, and F814W filters are used from HUDF12. For HUDFP2, only F814W is selected. We employ F606W and F814W from GOODS-North&South.

SExtractor (Bertin & Arnouts 1996) was used to detect sources by searching for five connected pixels 1.5 times above the rms noise level. We cropped a  $w \times w$  pixel postage-stamp image centered on each source. The image size  $w$  is set to  $2[4a + 5]$ , where  $a$  is the semimajor axis output by SExtractor and the operation  $[x]$  represents the greatest integer not exceeding  $x$ . We discard the source if



**Figure 3.** Size distribution of the HST postage-stamp images. The postage-stamp image size  $w$  is determined by the following equation:  $w = 2 \times [4a + 5]$ , where  $a$  is the semimajor axis output by SExtractor and the operation  $[x]$  represents the largest integer that does not exceed  $x$ . We used images smaller than  $128 \times 128$  pixels.

1. its image size  $w$  is greater than 128 pixels;
2. either its semiminor axis or half-light radius is smaller than the FWHM of HST/ACS PSF;
3. its S/N value is less than 20;
4. the maximum value of its neighboring sources is larger than 0.8 times its peak value; or
5. the gradient within the central  $10 \times 10$  pixels is sharper than the HST PSF.

The first and second criteria eliminate excessively large or small sources. The third condition removes excessively faint sources. The fourth condition excludes the cases where similarly bright galaxies are present within the postage-stamp image. The last condition is necessary because the HST automatic cosmic ray detection/removal pipeline often fails near the bright peaks of stars or galaxies. In addition to these criteria, we explicitly identified and removed stars using their distinct locus in the size–magnitude diagram. These conditions define our domain of interest and ensure the purity of our dataset. On the other hand, the omission of these galaxies may restrict the domain in which the model can perform well. We obtain a total of 66,092 postage-stamp images. Figure 3 shows

<sup>6</sup> <https://archive.stsci.edu/prepds/hudf12/>

<sup>7</sup> <https://archive.stsci.edu/hlsp/hlf/>

the distribution of the HST postage-stamp images prior to the  $w = 128$  cut.

As mentioned above, our pretraining dataset is created by `GalSim` using the characteristics derived from our HST dataset. We choose to characterize HST galaxies by their best-fit elliptical Sérsic parameters and fluxes. The circular Sérsic profile is defined as

$$I(R) = I(R_{50}) \exp \left( -b_n \left[ \left( \frac{R}{R_{50}} \right)^{1/n} - 1 \right] \right), \quad (1)$$

where  $n$  represents the Sérsic index,  $R$  is the radius,  $R_{50}$  is the half-light radius, and  $I(R_{50})$  is the surface brightness at  $R_{50}$ . The elliptical Sérsic profile is implemented by replacing  $R$  with the following:

$$R = (ab)^{1/2} \times \left[ \left( \frac{x}{a} \cos \theta + \frac{y}{a} \sin \theta \right)^2 + \left( -\frac{x}{b} \sin \theta + \frac{y}{b} \cos \theta \right)^2 \right]^{1/2}, \quad (2)$$

where  $a$  and  $b$  denote the semimajor and semiminor axes, respectively.

Because of the rapid change in  $I(R)$  near the center ( $R = 0$ ), evaluating  $I(R)$  only at discrete positions (i.e., at the center of each pixel) leads to a significant discrepancy between the rendered and real galaxy images. To mitigate this, we first oversampled  $I(R)$  with ten times higher resolution in the central region ( $R < 5$  pixel) and then applied  $10 \times 10$  nearest neighbor-averaging to render the galaxy. This Sérsic galaxy image is convolved with the HST PSF and then is fit to the HST galaxy image. Fluxes are measured by summing up the background-subtracted pixel values from the HST galaxy images.

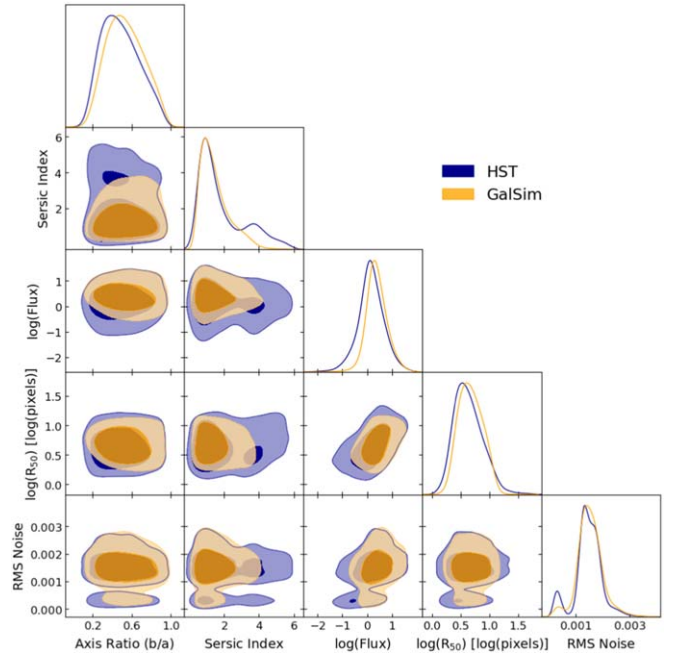
### 3.2. GalSim Dataset

For the generation of the pretraining dataset, we use `GalSim` based on the distributions of the Sérsic parameters derived from the HST dataset to reproduce not only the marginalized distribution of each parameter but also the correlations between the parameters. We constructed the probability density functions (PDFs) in the five-dimensional parameter space (axis ratio, Sérsic index, flux, size, and noise) with the `scipy.stats` Gaussian kernel density estimator module. Then, 100,000 new input parameter sets for `GalSim` were generated by resampling based on the derived PDFs.

The image size  $w$  varies in the original HST postage-stamp dataset. However, the galaxy images in the training dataset need to have an equal size. We chose a size of  $64 \times 64$  pixels since the choice provides a good balance between efficiency and image fidelity. A smaller image would truncate the profile too prematurely while a larger image would be dominated by background values.

The 100,000 noiseless galaxy images output by `GalSim` constitute the GT images in the pretraining dataset. Subsequently, these GT images were convolved with a  $7 \times 7$  pixel Gaussian kernel whose width matches the HST PSF. Finally, Gaussian noise was added to align the rms noise level with that of the HST dataset. These degraded images comprise the LQ images.

Following the same procedure used for determining the best-fit Sérsic parameters from the HST galaxy images, we measured the Sérsic parameters from the LQ images.



**Figure 4.** Galaxy property comparison between the HST and `GalSim` data sets. After measuring the HST galaxy properties, we construct a five-dimensional PDF. Input parameters for `GalSim` are generated through resampling from the pdf while preserving correlations between parameters. Then, we determine the galaxy properties from the resulting `GalSim` images with the same procedure used in the HST case. Overall, the `GalSim` galaxy images follow the HST properties reasonably well. However, the `GalSim` distribution is inevitably smoother because the input parameters as well as the `GalSim` images contain noise.

Figure 4 compares the Sérsic parameter distributions between the HST and `GalSim` LQ images. One must understand that in principle the two distributions cannot align exactly because the input Sérsic parameters used for the `GalSim` image generation are derived from the noisy HST images. Nevertheless, we find that overall the covariances and marginalized probability distributions are similar between the two data sets. Since the `GalSim` galaxy images are used to pretrain the model, which is subsequently finetuned by the JWST-based images, we do not think that the differences in detail matter.

### 3.3. JWST Dataset

We applied nearly the same procedures used for creating the HST dataset to sample the JWST image. The only difference is the criteria for the noise level. Since the JWST images prior to degrading serve as GT, the noise level must be considerably lower. We accepted an image only if its rms noise after the min-max normalization became less than 0.02.

One nontrivial issue is how to handle the difference in pixel scale between JWST and HST. NIRCAM’s native pixel scale is  $0''.03$  ( $0''.06$ ) for the short (long) wavelength channel, which is different from the HST/ACS pixel scale of  $0''.05$ . We considered resampling the JWST pixel scale to  $0''.05$ . However, the interpolation noise degraded the original quality nonnegligibly. Therefore, we maintained the original scale. Therefore, our JWST galaxies are 40% larger or 20% smaller than the HST ones on average. However, as we demonstrate in Section 4, our HST image restoration is not significantly affected by this pixel scale inconsistency.

Another tricky issue is the JWST’s intrinsic PSF. Although JWST’s PSF is on average much smaller than that of HST, its

long wavelength channel PSF width can be similar to that of the HST/ACS PSF. One may consider deconvolving the JWST image with its own PSF as an attempt to further enhance its resolution and thus better represent the truth. Perhaps this is achievable with an approach similar to the current method explored here. However, this is beyond the scope of this study. In the current investigation, we use JWST images without deconvolution. Thus, the effective PSF of the LQ images that we describe below includes the convolutions by both JWST's and HST's PSFs.

We standardize the postage-stamp image size to  $64 \times 64$  pixels. A smaller image is expanded by padding pixels to its four edges in such a way that both the noise and background levels match the original image. A larger image is trimmed to  $64 \times 64$  pixels. We discarded the galaxy images whose profiles were prematurely truncated by the trimming procedure. We collected 113,485 postage-stamp galaxy images from the JWST NIRCам F115W, F200W, F277W, and F444W images from various fields publicly available in 2023 August the Mikulski Archive for Space Telescopes (MAST)<sup>8</sup> doi:10.17909/crbp-3069. The dataset serves as the GT component in our fine-tuning dataset.

The LQ counterpart is generated by convolving the JWST GT galaxy images with the HST PSF and adding Gaussian noises. Although the pixel scale is different from the HST image, we scaled the HST PSF size, assuming the  $0''.05$  pixel scale. This is because, eventually, we desire to restore the HST images whose pixel scale is  $0''.05$ . The applied noise distribution follows that of the HST dataset. The ratios among the training, validation, and test data sets are 8:1:1. The GalSim and JWST train set subsamples are publicly available.<sup>9</sup>

#### 4. JWST Test Dataset Results

In this section, we present the results when our deep learning model is applied to the JWST test dataset. In this case, for each LQ image, there exists a corresponding GT image, which makes the comparison straightforward. However, readers are reminded that when we apply our model to real images (Section 5), where there are no matching GT images, we can only use GT proxies obtained also from real images for comparisons.

Below, we define the two evaluation metrics: peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM; Section 4.1), describe our visual inspection (Section 4.2), and compare morphological (Section 4.3) and photometric properties (Section 4.4).

##### 4.1. PSNR and SSIM

The PSNR and SSIM metrics are employed to evaluate the similarity between LQ and GT images and guide the selection of our best models. The mean squared error (MSE) is defined as:

$$\text{MSE} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [x(i, j) - y(i, j)]^2, \quad (3)$$

where  $m$  and  $n$  represent the size of the image,  $x(i, j)$  [ $y(i, j)$ ] is the pixel value at the position  $(i, j)$  in the GT (RS) image. MSE

**Table 1**

Comparison of PSNR and SSIM Performance between the Model Trained through Transfer Learning and the Models Trained with Either the GalSim or JWST Dataset

Training Dataset (1)	PSNR (2)	SSIM (3)
GalSim	32.2997	0.4711
JWST	38.7868	0.8534
<b>GalSim → JWST</b> no training	<b>38.8416</b> 16.4488	<b>0.8622</b> 0.1883

**Note.** The model obtained with transfer learning (bold) shows improvements in both PSNR and SSIM. The last row presents the PSNR and SSIM measured directly between the GT and LQ data sets, displaying the initial quality of the data sets.

measures the average squared difference between corresponding pixels in two images. With the MSE defined above, PSNR is given by:

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}^2}{\text{MSE}} \right), \quad (4)$$

where MAX is the peak value of the GT image. Compared to MSE, PSNR takes the image-by-image difference in peak values into account.

The SSIM metric is defined as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (5)$$

where  $\mu_x$  ( $\mu_y$ ) and  $\sigma_x^2$  ( $\sigma_y^2$ ) are the mean and variance, respectively, of the  $x$  ( $y$ ) image, and  $\sigma_{xy}^2$  is the covariance between the  $x$  and  $y$  images.  $C_1$  and  $C_2$  are introduced to prevent the equation from diverging when the denominator is small, and the exact values depend on the dynamic range of the target images. SSIM is a perceptual metric that evaluates the structural similarity between two images and considers luminance, contrast, and structure. The value ranges from  $-1$  to 1 with 1 being the perfect similarity.

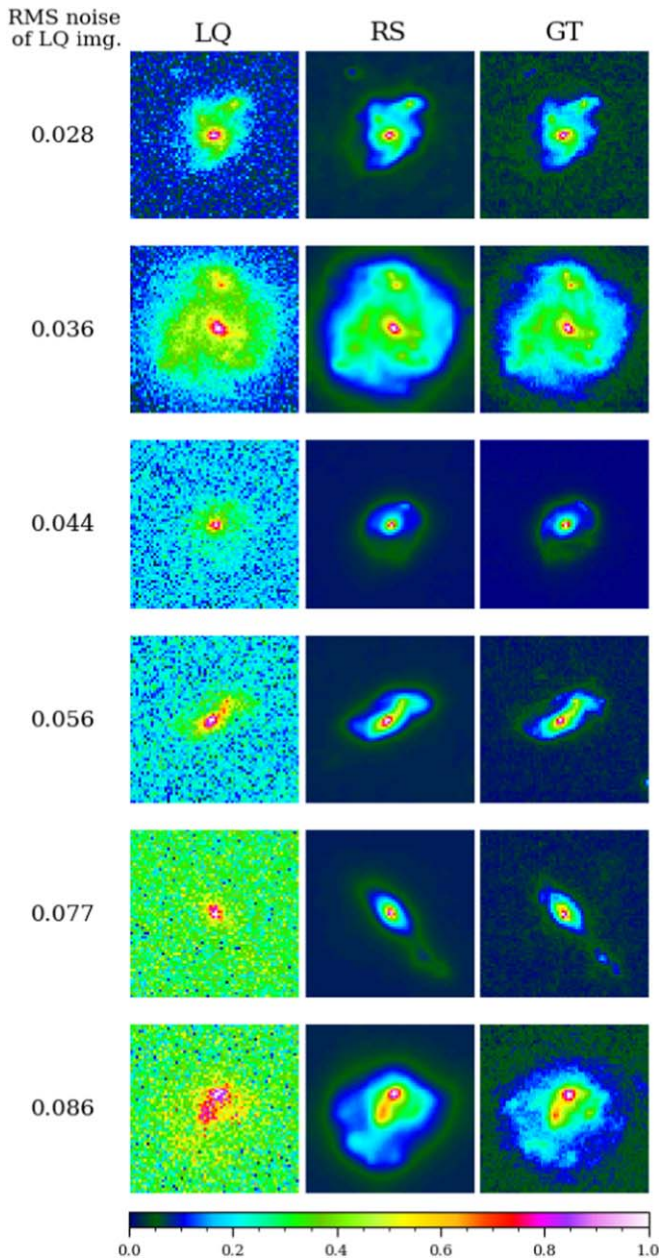
In Table 1, we compare the PSNR and SSIM values between our fiducial model (transfer learning) and experimental (nontransfer learning) models, including the values measured directly between GT and LQ data sets without model training. The transfer learning model provides the highest values in both PSNR and SSIM. Although in terms of PSNR the transfer learning shows only a marginal ( $\sim 0.33\%$ ) improvement over the JWST-only training case, the improvement in SSIM is 4.67% when normalized by the SSIM value for the no-training case.

##### 4.2. Visual Inspection

Figure 5 shows several image restoration cases in ascending order of the rms noise level. We select examples, which have rich substructures, low-surface-brightness features, and extended morphologies. The visual inspection indicates remarkable improvements in both resolution and noise level. Detailed substructures such as star-forming clumps and spiral arms have been restored with high fidelity. Also, the restored overall morphological features such as the ellipticity and core size are consistent with those of the GT images. In addition, the

<sup>8</sup> <https://archive.stsci.edu/>

<sup>9</sup> <https://drive.google.com/file/d/1dOemrQXFr2UxHgvUTUfLqLXIHeEkRbZ/view?usp=sharing>



**Figure 5.** Examples of restored images. We compare the low-quality (left), restored (middle), and ground-truth (right) images. Here, we select examples that have rich substructures, low-surface-brightness features, and extended morphologies. From top to bottom, the rms noise level increases. Each galaxy’s overall shape, low-surface brightness features, and substructures are remarkably well restored regardless of the noise level in the low-quality image.

low-surface-brightness edges are restored remarkably well. Finally, we note that the performance is not very sensitive to the input noise level, which in this example varies by a factor of 3.

### 4.3. Restoration of Morphological Parameters

Although one can characterize galaxy morphologies in various ways, we employ best-fit Sérsic parameters to quantify the comparison. In particular, we compare the two ellipticity components:  $e_1$  and  $e_2$ , Sérsic index  $n$ , half-light radius  $R_{50}$ , and the intensity at  $R_{50}$ .

The two components of the ellipticity are motivated by the weak lensing convention, which utilizes not only the absolute ellipticity  $e$  of a galaxy but also its position angle  $\theta$  as follows:

$$e_1 = e \cos(2\theta), \quad (6)$$

$$e_2 = e \sin(2\theta). \quad (7)$$

The ellipticity  $e$  is defined as  $(a - b)/(a + b)$ , where  $a$  and  $b$  are the semimajor and semiminor axes, respectively.

Figure 6 compares the five morphological parameters. Across all parameters, the RS images exhibit stronger correlations with the GT images in terms of the scatter and slope of the correlations. The ellipticity comparison shows that the LQ galaxies are systematically rounder than the GT galaxies, which is not surprising because they are generated by convolving the GT galaxies with the circular HST PSF. This bias nearly disappears in the RS images. The improvement in the Sérsic index  $n$  is remarkable. The scatter is reduced by more than a factor of 3. Also, the Pearson correlation coefficient improves from 0.61 to 0.90. Given that the Sérsic index  $n$  is one of the most difficult parameters to restore, its recovery showcases the stability of the restoration performance. The  $R_{50}$  recovery is also noteworthy with the reduction in the scatter by more than a factor of 4 and the increase in the Pearson correlation coefficient of  $\sim 48\%$ . Finally, the  $I(R_{50})$  intensity comparison shows a scatter reduction of  $\sim 46\%$  and an increase in the correlation of  $\sim 9\%$ .

### 4.4. Restoration of Photometric Parameters

One of the immediate scientific utilities of image restoration is enhancing photometry. Here we compare aperture flux, isophotal flux, and individual pixel values measured from the RS images with the GT images to evaluate the performance in the photometric context.

Since we use the min–max normalization consistently across our training and input data sets, the dynamic range of the RS images is also restricted. To extract photometry from the LQ and GT images, we opt to use the original (prenormalization) images. Consequently, it is necessary to rescale the RS images. To ensure a fair comparison, this rescaling process must be executed independently, without relying on the information available from the corresponding GT images.

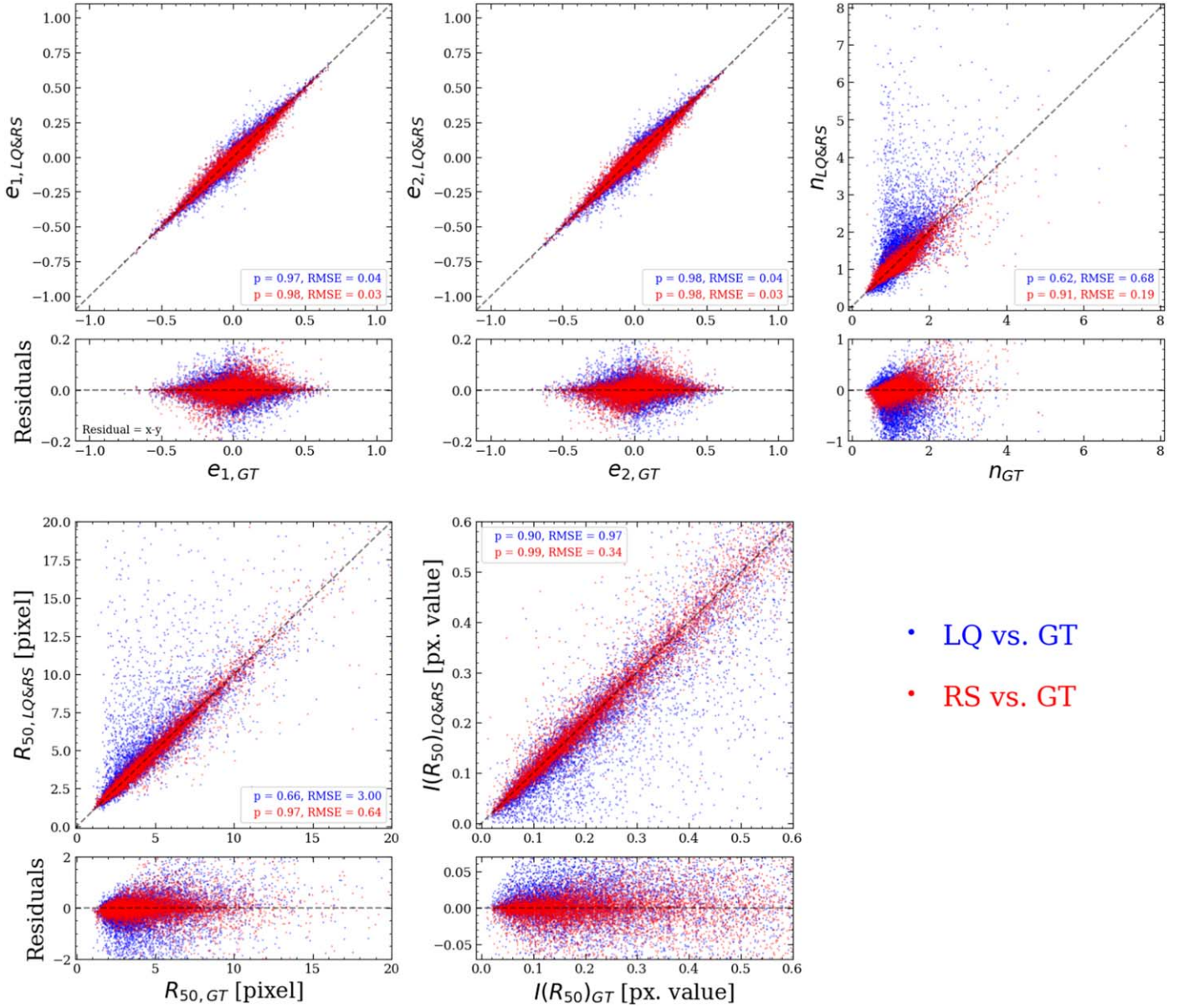
We aligned the dynamic range of the RS images to the LQ images as follows. First, we measured the lower (background) level  $l_{LQ}$  from the LQ image using the combination of  $\sigma$ -clipping and mode estimation implemented in `SEXTRACTOR`. The upper level of the LQ image  $u_{LQ}$  is determined by averaging the central  $4 \times 4$  pixel values. Then, we convolved the RS image with the HST PSF to match the LQ image resolution. With the same procedure, we measured the lower and upper levels ( $l_{RS}$  and  $u_{RS}$ , respectively) from this PSF-convolved RS image. Finally, we rescaled the RS image by first multiplying  $s$  and then adding  $c$ , where

$$s = \frac{u_{LQ} - l_{LQ}}{u_{RS} - l_{RS}} \quad (8)$$

and

$$c = -sl_{RS} + l_{LQ}. \quad (9)$$

For aperture flux measurement, We defined a Kron ellipse using the LQ image using `SEXTRACTOR` and applied it consistently across the LQ, RS, and GT images. Similarly,



**Figure 6.** Comparison of morphological parameters between GT and LQ images (blue) and between GT and RS images (red). We investigate  $e_1$ ,  $e_2$ ,  $n$  (Sérsic index),  $R_{50}$  (half-light radius), and  $I_{R50}$  (intensity at  $R_{50}$ ) determined from the Sérsic fitting. Black dashed lines denote one-to-one correlations. The RS images have a stronger correlation ( $p$ ) with a smaller scatter (RMSE).

isophotal areas were defined from the LQ image and used consistently across the LQ, RS, and GT images.

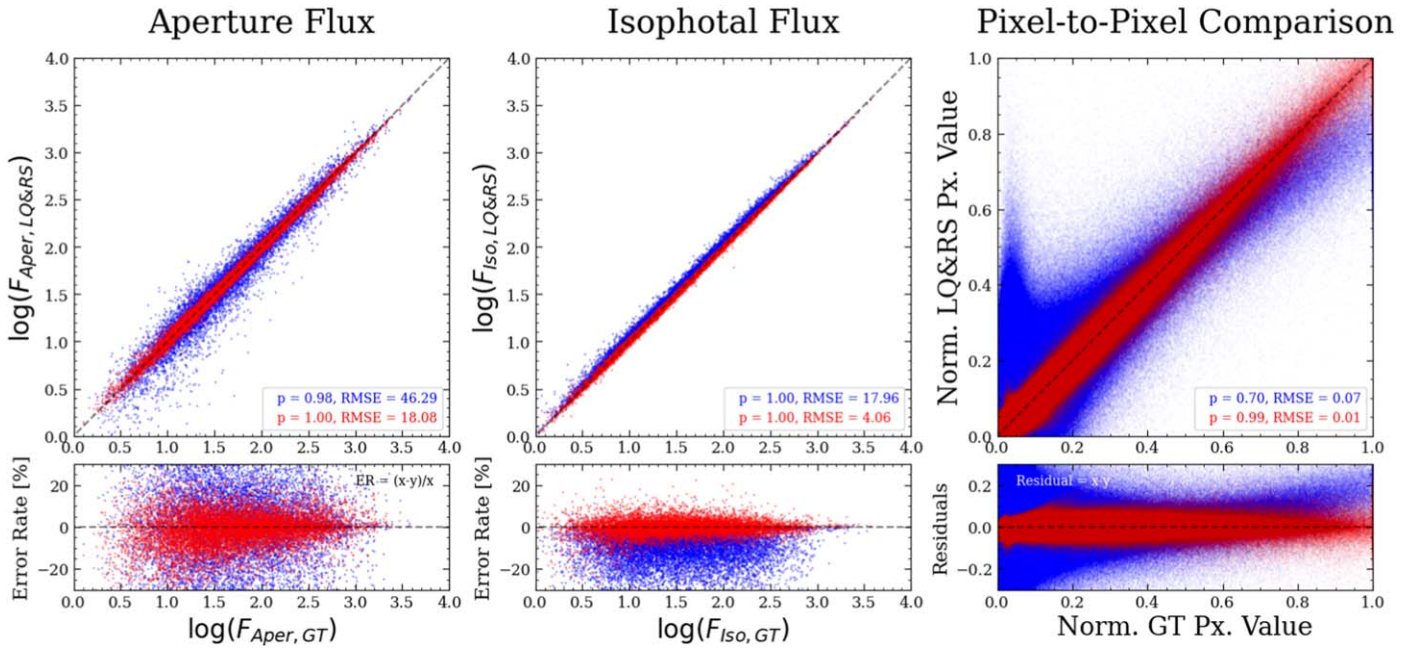
Figure 7 displays the resulting flux comparisons. The aperture fluxes from the RS images are in good agreement with those from the GT images. Compared to the LQ images, the scatter is reduced by  $\sim 60\%$ . The scatter reduction is similar ( $\sim 55\%$ ) in isophotal flux. It is worth noting that the isophotal fluxes in the LQ images are systematically overestimated because the isophotal area is defined from the LQ image.<sup>10</sup> This bias is significantly reduced in the RS images. Finally, the pixel-to-pixel comparison illustrates a tight 1:1 correlation between the RS and GT images across the entire dynamic range, while the LQ images show a slope significantly less than

unity because of their larger PSF. The pixel-to-pixel scatter reduction is by a factor of 7.

## 5. Application to Real HST Images

In Section 4, we have demonstrated that our deep learning model successfully restores the simulated HST images to the JWST-quality level in both resolution and depth. Here we present our results when the same model is applied to real HST images. It is important to note that we cannot perform identical comparisons as presented in Section 4 because there are no equivalent GT images for the real HST images. In Section 5.1, we restore single-epoch (short-exposure) images and compare their RS images to multiepoch (long-exposure) images, using the latter as proxies for the GT dataset. Since our deep learning model not only reduces the noise level but also the effective PSF, the comparison cannot be fair because the multiepoch image suffers from the same HST PSF. In addition, the

<sup>10</sup> That is, noise can make some pixel values near the edge of the isophotal area in the LQ image higher than the GT values.



**Figure 7.** Comparison of photometric information between GT and LQ images (blue) and between GT and RS images (red). We investigate correlations of aperture flux, isophotal flux, and individual pixel values. For flux comparison, since the RS images were scaled to the range [0, 1], we rescaled the RS images to enable quantitative comparisons (see text for details). We stress that we *do not* use the information from GT for rescaling. We use an elliptical aperture defined with SExtractor’s semimajor, semiminor axes, and orientation angle from the LQ image. The isophotal area is also determined from the LQ image. For comparison of individual pixels, we use the pixels only within the elliptical aperture. Flattened pixel values in LQ images due to convolution are restored to their original values, reproducing a one-to-one slope. The photometric information is recovered remarkably well, with a significant reduction in scatter compared to the LQ-GT comparison.

multiepoch image is not sufficiently deep and often noisier than the RS images. Section 5.1, we restore HST images and compare their RS images to deep JWST images. In this case, the resolutions between the two data sets are aligned. However, since their filters are different, we cannot expect the pairs to have identical photometric and morphological properties even if the deep learning model provides perfect restoration.

### 5.1. Enhancement of Single-epoch Images and Comparison with Multiepoch Images

We utilize the data sets from the HUDF and GOODS-N & S observed by the HST ACS. The single-epoch low-quality (SG-LQ) images are retrieved from the MAST archive. We aim to enhance the fully calibrated and flat-fielded (FLT) images not their “drizzled” (DRZ) images because “drizzling” induces inter-pixel noise correlations. When a postage-stamp image is contaminated by cosmic rays, we remove them with the `lacosmic script`<sup>11</sup> (van Dokkum 2001; Bradley 2023). Since the multiepoch high-quality (MT-HQ) mosaic images are corrected for geometric distortions unlike the FLT images, we transformed them back to the distorted, CCD coordinate system to enable a fair comparison with the FLT images.

In Figure 8(a), we display four enhancement examples. Similar to the results observed in the experiment with the JWST test dataset, the restoration qualities are similarly excellent here. The low-surface-brightness features, which cannot be identified in the LQ images, are well restored in the RS images. One notable difference from the results in the JWST test dataset is the PSF size. Since our deep learning model is trained to restore the JWST resolution, the PSF sizes of the RS images are smaller than those of the HQ images.

Another important difference is on average the higher noise level in our single-epoch LQ images.

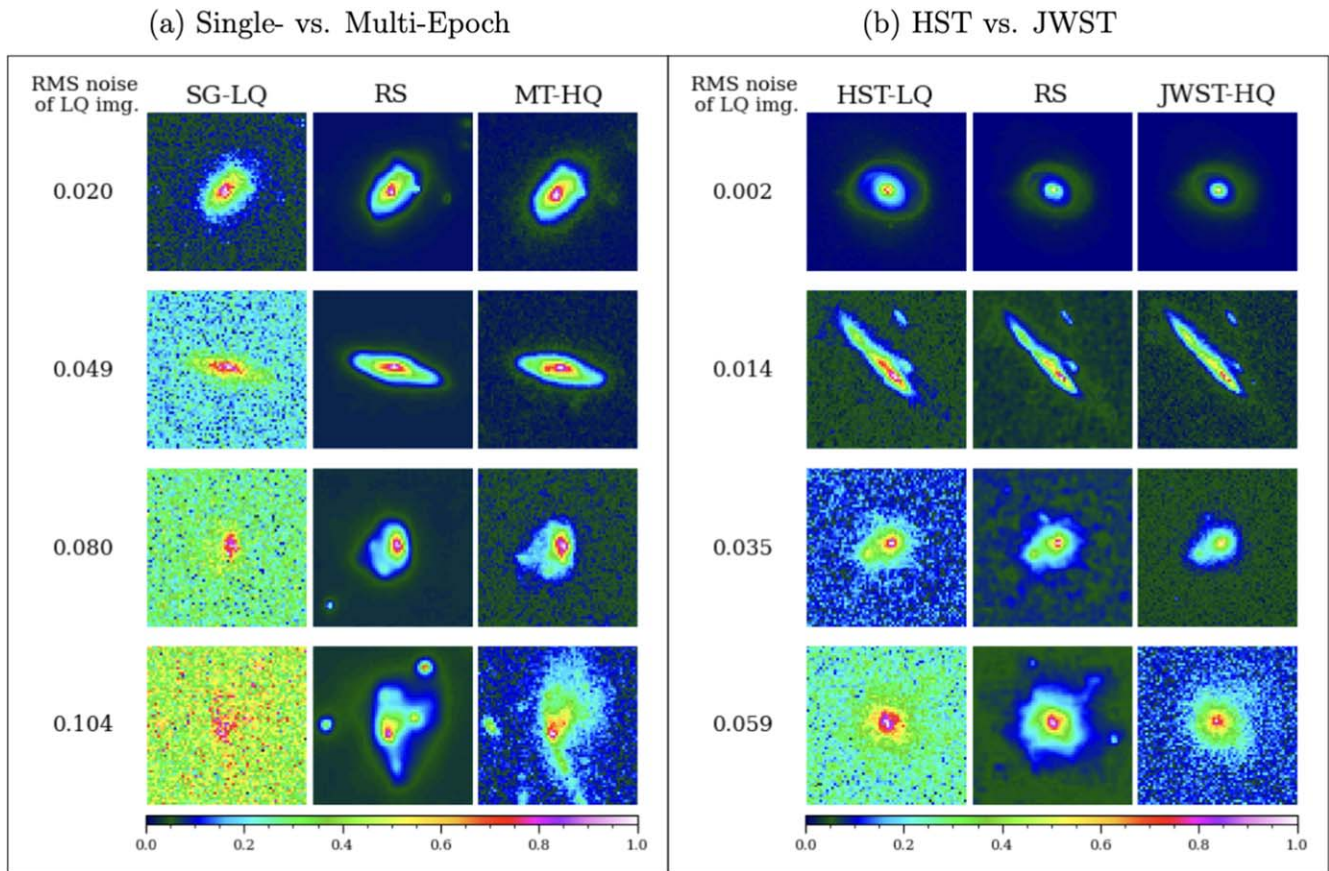
Figure 9 shows a comparison of the five morphological parameters. Significant improvements over the LQ images are clear in terms of the correlation strength and reduction in the scatter. Readers are reminded that we used the multiepoch image as a proxy for the GT image after applying an inverse geometric distortion correction to it. Therefore, compared to the JWST dataset experiment, the measurement of the five morphological parameters from the GT surrogate is substantially noisier here. Also, as mentioned above, the PSF size of the RS image is smaller than that of the HQ image, which can contribute to a systematic difference in shapes between the two data sets.

We display a comparison of photometric parameters in Figure 10. Again, significant improvements over the single-epoch LQ images are apparent. In particular, the pixel-to-pixel comparison shows a remarkable difference in both correlation strength and scatter reduction. Since the caveats mentioned in the morphological parameter comparison above apply here too, the improvements are less pronounced than those in the experiment with the JWST test dataset.

### 5.2. Enhancement of Multiepoch HST Images and Comparison with Multiepoch JWST Images

We utilize the publicly available HST and JWST multiepoch mosaic images of Abell 2744. In particular, we apply our Transformer model to the HST ACS/Wide Field Camera F606W images and compare their RS images with the JWST NIRCcam F115W images. We resampled the JWST images to match the HST pixel scale ( $0''.05 \text{ pix}^{-1}$ ). Since galaxies must have different morphological and photometric parameters

<sup>11</sup> <https://lacosmic.readthedocs.io/en/stable>



**Figure 8.** Application to HST images. We apply our restoration model discussed in Section 2 without any modification to HST images. (a) Comparison of single-epoch RS images with multi-epoch high-quality images. Here, we use deep multi-epoch images as a proxy for GT. The single-epoch LQ images were enhanced using our model. Note that since the model is designed to perform deconvolution to match the JWST PSF, the resolution of the RS images is higher than that of the deep multi-epoch HST images. Also, the noise levels in the RS images are lower than in the multi-epoch data. (b) Comparison of multi-epoch RS images with deep JWST images. Here, we use deep JWST images as a proxy for GT. Since the JWST/NIRCam images are near-IR (F115W) images while the HST/ACS images are optical (F606W) images, only limited comparison is possible. Nevertheless, both the reduction in noise and improvement in resolution make the RS images look more similar to the JWST images than the corresponding LQ images.

between optical and near-IR, we provide only a visual comparison here.

Figure 8(b) shows four examples. Despite the difference in filter, the RS images look remarkably similar to the JWST images. In particular, when the noise level is relatively low (i.e., top two panels), the enhancement in both resolution and low-surface brightness feature is excellent. When the noise level is relatively high, the RS images still resemble the JWST images more closely than their input HST images. We suspect that the correlated noise in the input multi-epoch HST image might have a nonnegligible impact on the result in this noise regime.

## 6. Limitations

### 6.1. Degradation in Restoration Quality Due to High Noise Level

Although our Transformer-based deep learning model provides state-of-the-art performance in both resolution enhancement and noise reduction for moderate noise levels, inevitably, proper restoration becomes impossible when the noise level exceeds a certain threshold.

Figure 11 shows four such examples. The top row illustrates a case, where the distinct spiral arm in the GT image is not restored. Upon scrutiny of the corresponding LQ image, we

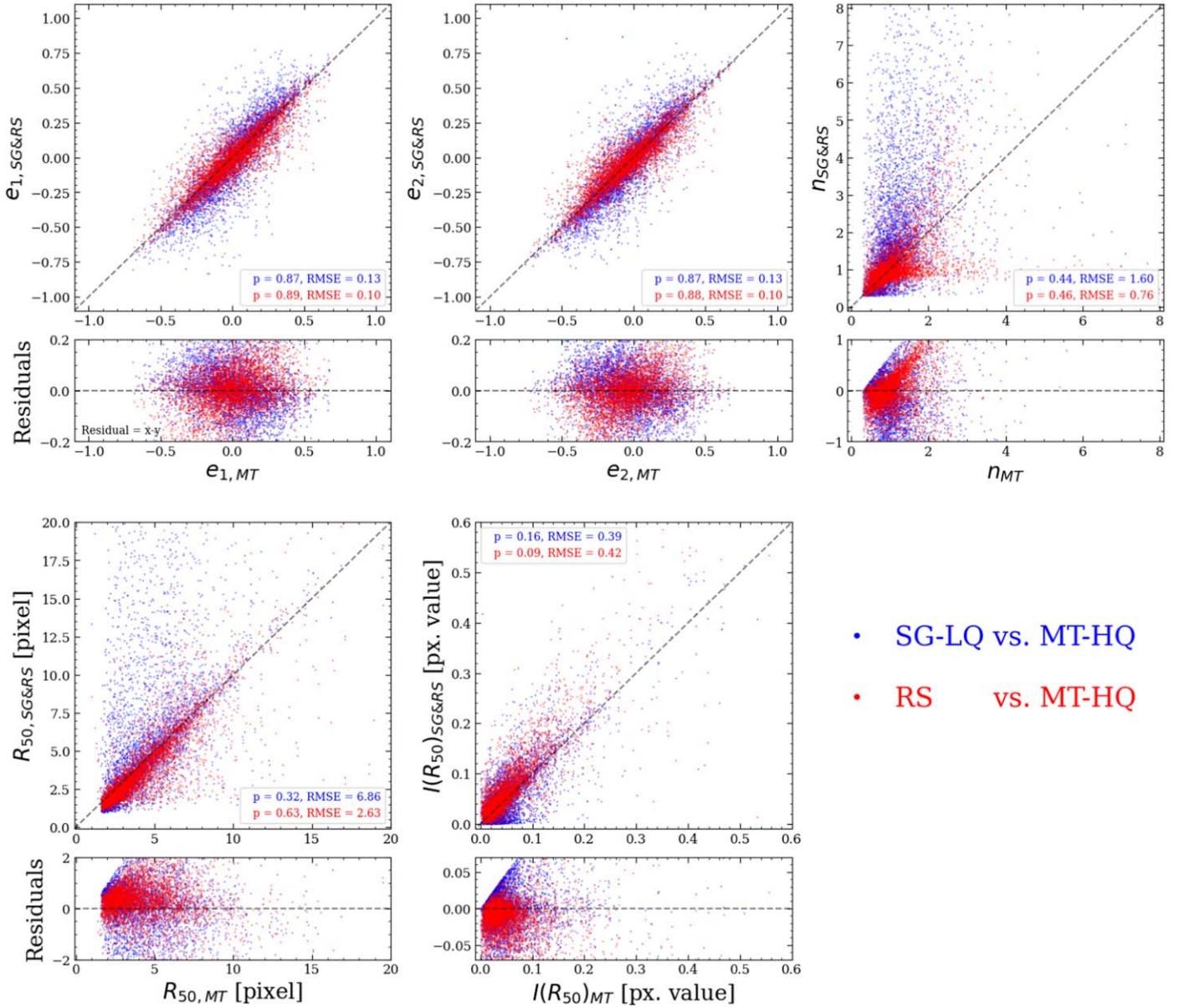
believe that the information loss due to noise is too significant to hint at the presence of the spiral arm. The second row shows a case, where the position angle of the GT image is not properly restored. Again, we suspect that the LQ image is too noisy to enable a proper inference of the position angle of the GT image. The third row is an example, where the ellipticity of the RS image is much rounder than that of the GT image. In the last row, we illustrate a case, where the deep learning model fails to unveil the two distinct peaks of the GT image.

Since the exact definition of restoration failure is a subjective matter, it is difficult to quote an exact threshold. Nevertheless, from visual inspections, we suggest that the failure frequency increases noticeably when the rms value (after min-max normalization) of the LQ image is larger than  $\sim 0.1$ .

### 6.2. Point Source Recovery Test

In principle, the perfect deconvolution algorithm should restore a point source to a delta function, which is infinitely smaller than a pixel. In traditional deconvolution in the Fourier domain, it is a challenging task because the operation is numerically unstable. The resulting images often exhibit many ringing effects around a bright central peak.

Since we excluded stars from the training dataset, our deep learning model did not explicitly learn to deconvolve point source images. Thus, it is interesting to examine how well our



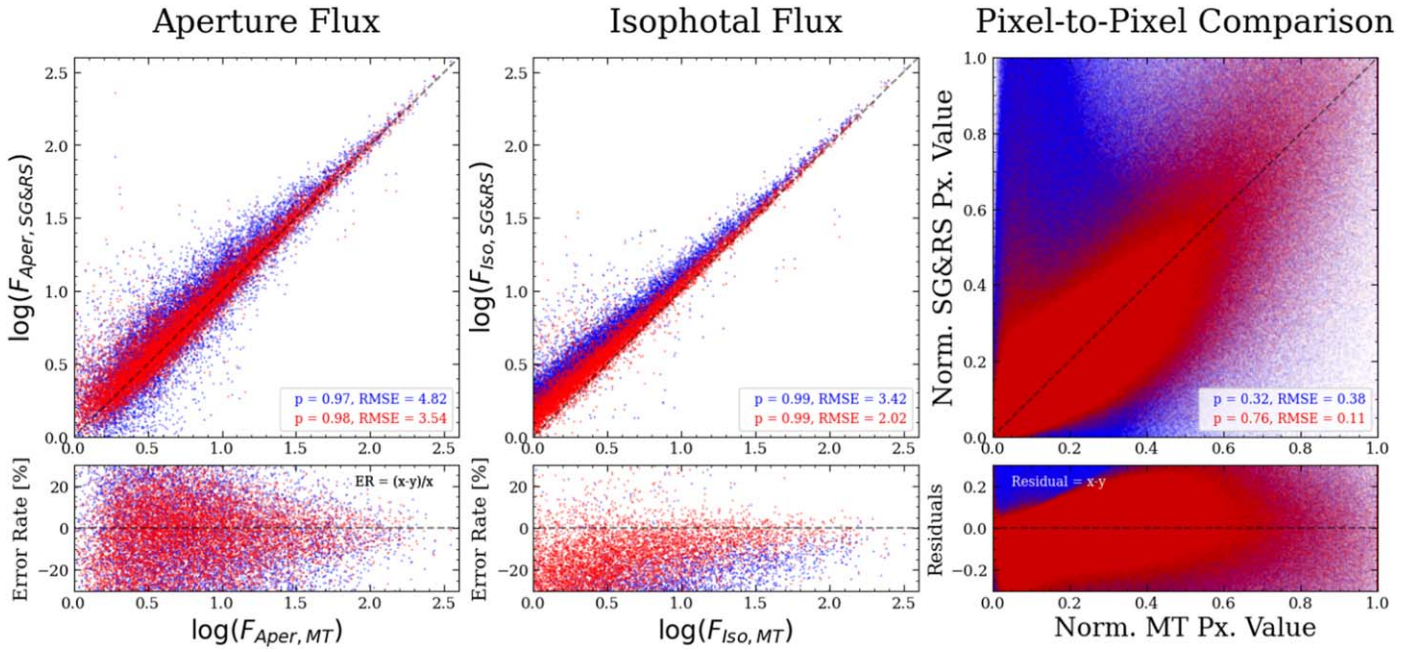
**Figure 9.** Comparison of morphological parameters between MT-HQ and SG-LQ images (blue) and between MT-HQ and RS images (red). We follow the same format as in Figure 6. Note that the MT-HQ images have higher noise than the GT images created from the JWST images. Also, the SG-LQ images are noisier than the LQ images in the JWST dataset. Since SG images are in the geometrically distorted CCD coordinate, we applied the due transformation to the MT-HQ images in the rectified coordinate to align them with the SG images in the CCD coordinate. The properties measured in RS images are more closely correlated with those in MT-HQ images than those in the SG-LQ images.

deep learning model, trained with only galaxy images, restores point sources. We perform a point source recovery test as follows. First, we created 1000 JWST-quality star GT images. Because we did not remove the JWST PSF from the training dataset, the star GT images should not resemble delta functions but the JWST PSF. We implemented this by convolving a single pixel with a Gaussian whose kernel size matches the JWST PSF. Note that we randomize the positions of the stars within the central  $24 \times 24$  pixels of the  $64 \times 64$  postage-stamp images. Then, we created their LQ versions by further convolving the GT images with the HST PSF and adding noise. Finally, these LQ images are restored by our deep learning model. To investigate the systematic effect, we stack the 1000 GT and RS images separately after aligning their centers.

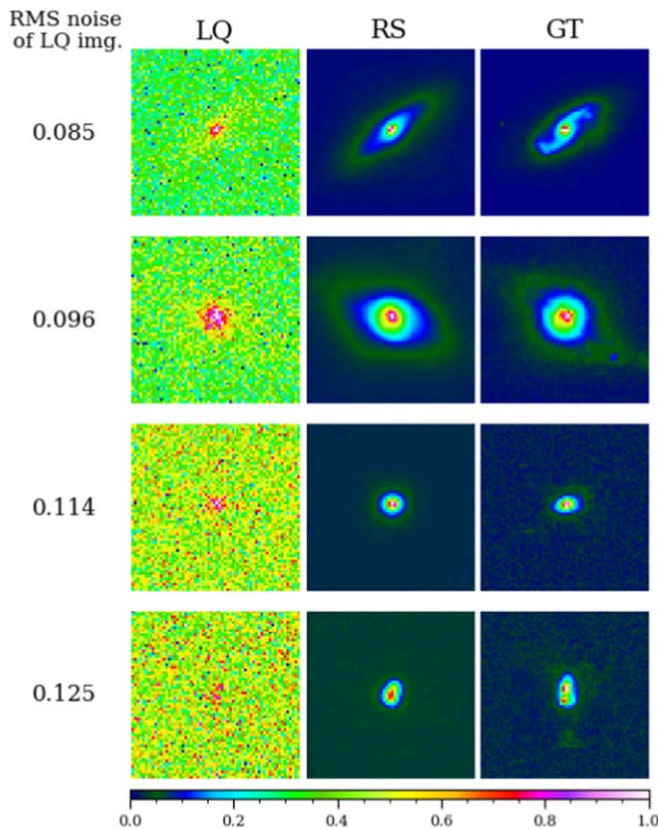
Figure 12 shows that the difference is subtle when we compare the GT (left) and RS (middle) stacks visually. However, the residual image (right) illustrates that the PSF of the RS stack is systematically larger. Thus, we conclude that our deep learning model, trained solely with galaxy images, performs less than ideal for point sources.

### 6.3. Artifacts Due to Pixel Correlation

In our generation of LQ images, we assume that the noise is Gaussian. However, in real astronomical images, especially when we create deep images by stacking many dithered exposures, there exist significant inter-pixel noise correlations. We find that these inter-pixel noise correlations create nonnegligible artifacts.

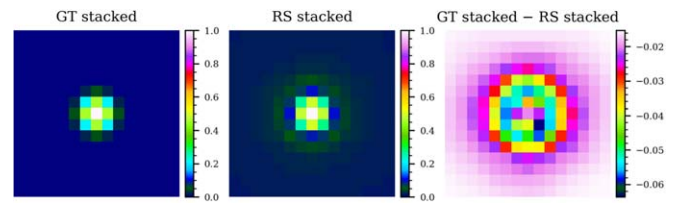


**Figure 10.** Comparison of photometric information between MT-HQ and SG-LQ images (blue) and between MT-HQ and RS images (red). Significant improvements over the SG-LQ images are clear. In particular, the pixel-to-pixel comparison shows remarkable enhancement in both correlation strength and scatter reduction.



**Figure 11.** Performance degradation cases. As the noise level increases, the restoration depends on fewer and fewer pixels. Thus, more frequent degradation cases occur with higher noise levels. Shown here are examples of loss of spiral arms (first row), incorrect ellipticity (second and third row), and failed de-blending of multiple peaks (fourth row).

Figure 13 displays some examples of these artifacts. The LQ images here are sampled from multiepoch images drizzled with square and Lanczos3 kernels using AstroDrizzle



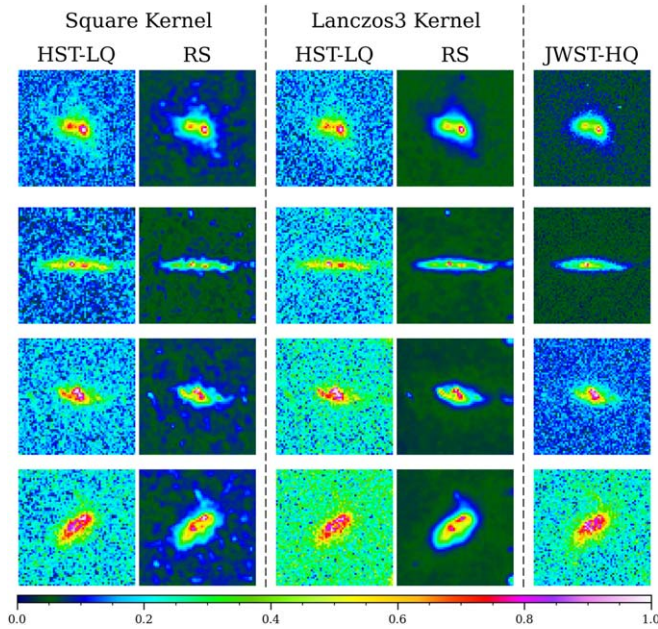
**Figure 12.** Point source restoration test. We created 1000 LQ-GT pairs of point source images and compared the mean RS image with the mean GT image (see text for details). The profile in the RS image is slightly more extended than that in the GT image.

(Fruchter et al. 2010) in DrizzlePac (Hoffmann et al. 2021). With square-kernel drizzling, correlated noise is apparent even from visual inspection. The RS images show that the correlated noise creates some low-surface-brightness artifacts in the galaxy outskirts, which are absent in the JWST images. These artifacts are mitigated when we use the Lanczos3 kernel, which helps reduce the pixel correlation in the mosaic-drizzled image. The background and edge of the galaxy in the RS images are clean and smooth. Addressing this issue could involve strategies such as employing a different drizzling kernel for image stacking or leveraging more advanced deep learning algorithms. Exploring these solutions will be a key focus of our future work.

#### 6.4. Discussion on Potential Hallucination Effects

A hallucination effect in generative models refers to the phenomenon where the model generates outputs that are not grounded in reality or the input data, but instead are fabricated or “hallucinated” by the model, leading to false interpretations. Since virtually all generative models are subject to hallucination effects in varying degrees, we discuss potential hallucination effects in our restoration model.

A detailed examination of Figure 8(a) reveals that the restored images of the single-epoch data in the third and fourth rows show some compact, point-like sources. Because these



**Figure 13.** Application to HST galaxies sampled from multiepoch drizzled images. The first and third columns are the HST-LQ multiepoch images drizzled with square and Lanczos3 kernels, respectively. Pixel values are significantly correlated for the LQ image produced with the square kernel. In their corresponding RS images, these correlations manifest as numerous artifacts, which are absent in the JWST-HQ images. The Lanczos3 kernel reduces pixel correlations in the LQ images and also minimizes artifacts in the corresponding RS images.

features do not exist in the corresponding much deeper multiepoch data, they may be regarded as arising from hallucination effects. However, scrutiny of the corresponding LQ images indicates that the pixel levels of the regions where the hallucination effect is suspected are elevated compared to the neighboring regions. They may be outliers in noise fluctuations, uncorrected bad pixels, or residual cosmic rays that were not identified. We verified that these compact, point-like features do not appear in the RS images without their counterparts in the GT images when we restored images from the LQ images generated by degrading the JWST GT images (Figure 5). Therefore, we conclude that the artifacts in Figure 8(a) are grounded in the real features of the original LQ images rather than caused by the hallucination effect. Nevertheless, the discussion provides an important lesson: when we apply our model to real LQ images, we need to ensure that the input images do not possess glitches (e.g., bad pixels, cosmic rays) distinct from general noise.

## 7. Conclusions

We have showcased astronomical image restoration from HST quality to JWST quality using the efficient Transformer model via transfer learning. The pretraining dataset was created by rendering GT galaxy images based on analytic profiles and generating corresponding LQ versions by reducing their resolution and introducing noise. The fine-tuning dataset was produced by sampling GT galaxy images from deep JWST images, which were then degraded to the LQ images in a similar fashion.

With the test dataset, we find that the restored images show significantly enhanced correlations with the GT images than their original LQ versions, reducing the scatters of isophotal

photometry, Sérsic index, and half-light radius by factors of 4.4, 3.6, and 4.7, respectively, with Pearson correlation coefficients approaching unity. We also visually confirm that the restored images are superior in terms of resolution and noise level. When we applied our model to real low-exposure HST images, the enhanced images also showed significantly improved correlations with their multiepoch versions, although the absence of their real GT images limits our interpretations.

We discuss a few limitations of our model. First, the performance degrades in high noise regimes, where the background rms approaches  $\sim 10\%$  of the object peak values. Second, highly correlated noise can be misinterpreted as astronomical features, leading to the manifestation of low-surface-brightness features. Third, the restoration of point sources is less than optimal. Additionally, the model can create false object images from locally brightened regions caused by noise.

Although it is possible to further improve the model with larger training data sets and enhanced training strategies, we anticipate that our current Transformer-based deep learning model will prove useful for a number of scientific applications, including precision photometry, morphological analysis, and shear calibration.

## Acknowledgments

This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT; No. 2021-0-02068, Artificial Intelligence Innovation Hub). M.J.J. acknowledges support for the current research from the National Research Foundation (NRF) of Korea under the programs 2022R1A2C1003130 and RS-2023-00219959. This work is based on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST.

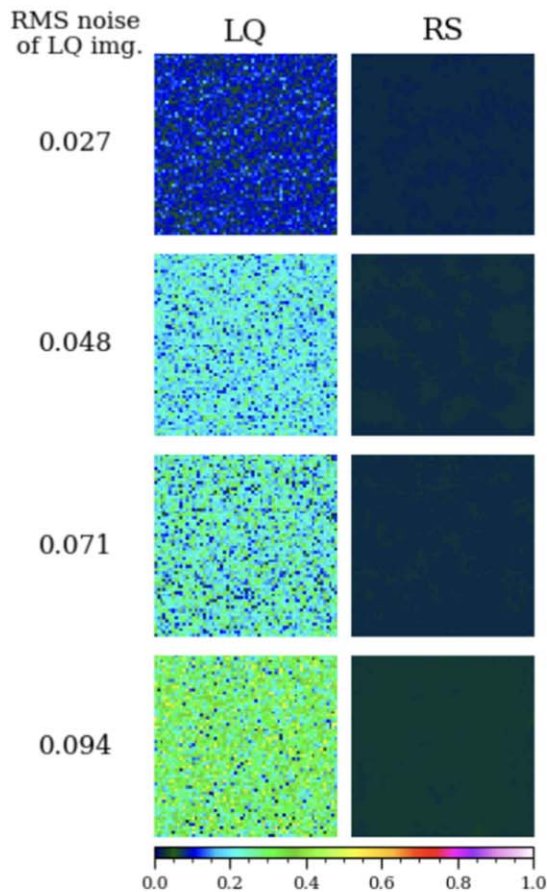
*Facilities:* JWST (NIRCam), HST (ACS).

*Software:* numpy (Harris et al. 2020), scipy (Virtanen et al. 2020), matplotlib (Hunter 2007), astropy (Astropy Collaboration et al. 2013, 2018), photutils (Bradley et al. 2023), SExtractor (Bertin & Arnouts 1996), GalSim (Rowe et al. 2015), lacosmic (Bradley 2023), Drizzlepac (Hoffmann et al. 2021).

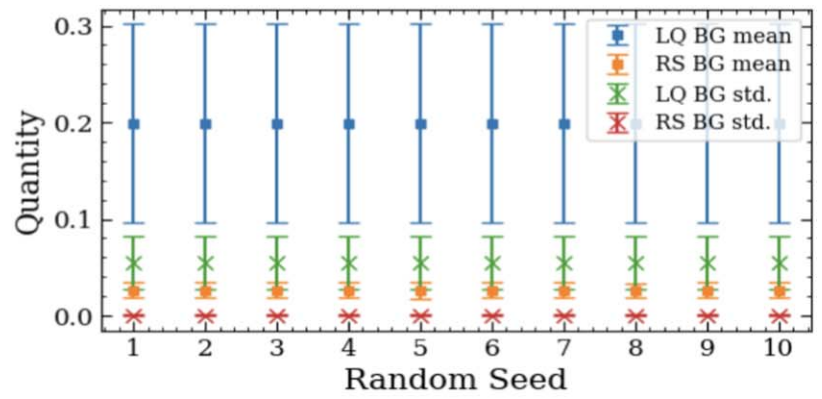
## Appendix

### Image Restoration Test with Blank Noise-only Images

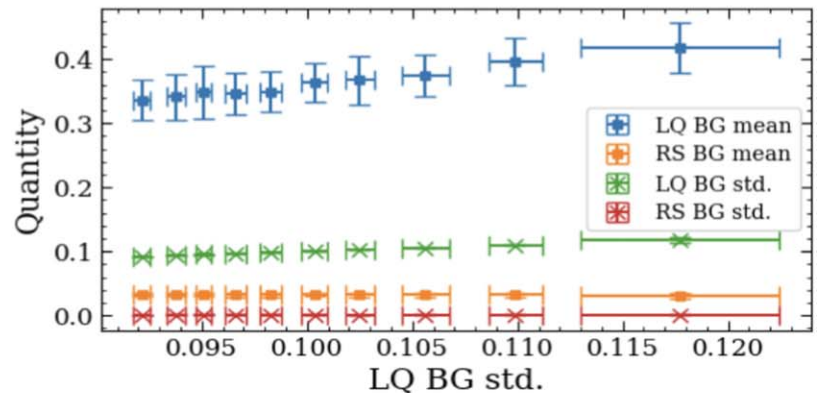
One of the key requirements of our deep-learning-based restoration model is that the model should not generate any false object images by overinterpreting the noise when the image contains no real astronomical source. To test this, we created blank noise-only images by varying the random seed and the noise level. We used 10 different random seeds, and for each random seed, we generated 1000 images, where the mean and standard deviation of the noise were set to mimic those of a randomly selected galaxy image from the JWST train dataset. The RS images created from these blank images were carefully inspected. No image was found to contain pseudo-sources. Figure 14 presents the test results.



(a) Blank image restoration examples



(b) Random Seed Dependency



(c) Background Properties

**Figure 14.** Noise-only image restoration test. (a) Examples of blank image restoration. The rms noise increases from top to bottom. No pseudo-object is produced while the noise is significantly reduced. (b) Test with different random seeds. Each data point represents statistics from 1000 images. The mean and standard deviation of the RS images are consistent across different random seeds. (c) Same as (b) but for different noise levels. Regardless of the input (LQ) image's noise level, the output (RS) image's background mean and noise level stay consistently low.

### ORCID iDs

Hyosun Park <https://orcid.org/0009-0007-7093-1758>  
 Yongsik Jo <https://orcid.org/0009-0005-9082-4875>  
 Seokun Kang <https://orcid.org/0000-0003-4243-8214>  
 Taehwan Kim <https://orcid.org/0000-0002-6571-4632>  
 M. James Jee <https://orcid.org/0000-0002-5751-3697>

### References

- Akhaury, U., Starck, J.-L., Jablonka, P., Courbin, F., & Michalewicz, K. 2022, *FrASS*, **9**, 357
- Alsaiani, A., Rustagi, R., Alhakamy, A., Thomas, M. M., & Forbes, A. 2019, in *IEEE 2nd Int. Conf. Information and Computer Technologies (ICICT)* (Piscataway, NJ: IEEE), 126
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bradley, L. 2023, larrybradley/lacosmic: v1.1.0, Zenodo, doi:10.5281/zenodo.10145563
- Bradley, L., Sipőcz, B., Robitaille, T., et al. 2023, astropy/photutils: v1.10.0, Zenodo, doi:10.5281/zenodo.1035865
- Díaz Baso, C. J., de la Cruz Rodríguez, J., & Danilovic, S. 2019, *A&A*, **629**, A99
- Elhakiem, A. A., Elsaid Ghoniemy, T., & Salama, G. I. 2021, in 2021 Tenth Int. Conf. Intelligent Computing and Information Systems (ICICIS) (Piscataway, NJ: IEEE), 51
- Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, *ApJL*, **763**, L7
- Fruchter, A. S., Hack, W., Dencheva, N., Droettboom, M., & Greenfield, P. 2010, in 2010 Space Telescope Science Institute Calibration Workshop (Baltimore, MD: STScI), 382
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, *Advances in Neural Information Processing Systems* (New York: ACM), 27
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hoffmann, S. L., Mack, J., Avila, R., et al. 2021, *AAS Meeting*, **53**, 216.02
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Illingworth, G., Magee, D., Bouwens, R., et al. 2016, arXiv:1606.00841
- Jones, R., & Wykes, C. 1989, *Holographic and Speckle Interferometry* (Cambridge: Cambridge Univ. Press)
- Kalele, G. 2023, in 1st Int. Conf. Innovations in High Speed Communication and Signal Processing (IHCSPP) (Piscataway, NJ: IEEE), 212
- Koekemoer, A. M., Ellis, R. S., McLure, R. J., et al. 2013, *ApJS*, **209**, 3
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, *Advances in Neural Information Processing Systems* (Red Hook, NY: Curran Associates Inc.), 25, [https://proceedings.neurips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)
- Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., et al. 2021, *MNRAS*, **504**, 5543
- Ledig, C., Theis, L., Huszar, F., et al. 2017, in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (Piscataway, NJ: IEEE), 105
- Liu, M.-Y., Huang, X., Yu, J., Wang, T.-C., & Mallya, A. 2021, in *Proc. of the IEEE, 109, Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications* (Piscataway, NJ: IEEE), 839

- Loshchilov, I., & Hutter, F. 2019, in 7th Int. Conf. on Learning Representations, ICLR 2019, <https://openreview.net/forum?id=Bkg6RiCqY7>
- Lucy, L. B. 1974, *AJ*, **79**, 745
- Park, H., Jo, Y., Kang, S., Kim, T., & Jee, M. J. 2024, Galaxy Restormer: Galaxy Denoising and Deconvolution Deep Learning Model, v0.0.1, Zenodo, doi:[10.5281/zenodo.11378660](https://doi.org/10.5281/zenodo.11378660)
- Rajeev, R., Samath, J., & Karthikeyan, N. 2019, *J. Med. Syst.*, **43**, 234
- Richardson, W. H. 1972, *JOSA*, **62**, 55
- Ronneberger, O., Fischer, P., & Brox, T. 2015, in 16th Europlanet Science Congress 2022 (Berlin: Springer), 234
- Rowe, B., Jarvis, M., Mandelbaum, R., et al. 2015, *A&C*, **10**, 121
- Schawinski, K., Zhang, C., Zhang, H., Fowler, L., & Santhanam, G. K. 2017, *MNRAS*, **467**, L110
- Shepp, L. A., & Vardi, Y. 1982, *IEEE Trans. Med. Imaging*, **1**, 113
- Simkin, S. M. 1974, *A&A*, **31**, 129
- Sureau, F., Lechat, A., & Starck, J. L. 2020, *A&A*, **641**, A67
- Sweere, S. F., Valtchanov, I., Lieu, M., et al. 2022, *MNRAS*, **517**, 4054
- Tikhonov, A. N., & Goncharky, A. V. 1987, *Ill-posed Problems in the Natural Sciences* (Moscow: Mir)
- Tran, L., Nguyen, S., & Arai, M. 2021, *GAN-Based Noise Model for Denoising Real Images* (Berlin: Springer), 560
- Tripathi, S., Lipton, Z., & Nguyen, T. 2018, arXiv:[1803.04477](https://arxiv.org/abs/1803.04477)
- van Dokkum, P. G. 2001, *PASP*, **113**, 1420
- Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, *Advances in Neural Information Processing Systems* (Red Hook, NY: Curran Associates, Inc.), 5998
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wakker, B. P., & Schwarz, U. J. 1988, *A&A*, **200**, 312
- Wang, Z., Wang, L., Duan, S., & Li, Y. 2020, *JPhCS*, **1550**, 032127
- Whitaker, K. E., Ashas, M., Illingworth, G., et al. 2019, *ApJS*, **244**, 16
- Williams, R. J., & Zipser, D. 1989, *Neural Comput.*, **1**, 270
- Yan, L., Fang, H., & Zhong, S. 2012, *OptL*, **37**, 2778
- Zamir, S. W., Arora, A., Khan, S., et al. 2022, in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 5728
- Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. 2017, *ITIP*, **26**, 3142
- Zhang, Q., Xiao, J., Tian, C., Lin, J., & Zhang, S. 2022, *CAAI Trans. Intell. Technol.*, **8**, 331