

Euclid Quick Data Release (Q1)

XXI. Active galactic nuclei identification using diffusion-based inpainting of Euclid VIS images

Euclid Collaboration: G. Stevens^{1,*}, S. Fotopoulou¹, M. N. Bremer¹, T. Matamoro Zatarain¹, K. Jahnke², B. Margalef-Bentabol³, M. Huertas-Company^{4,5,6,7}, M. J. Smith^{8,9}, M. Walmsley^{10,11}, M. Salvato¹², M. Mezcuca^{13,14}, A. Paulino-Afonso^{15,16}, M. Siudek^{5,13}, M. Talia^{17,18}, F. Ricci^{19,20}, W. Roster¹², N. Aghanim²¹, B. Altieri²², S. Andreon²³, H. Aussel²⁴, C. Baccigalupi^{25,26,27,28}, M. Baldi^{29,18,30}, S. Bardelli¹⁸, P. Battaglia¹⁸, A. Biviano^{26,25}, A. Bonchi³¹, E. Branchini^{32,33,23}, M. Brescia^{34,35}, J. Brinchmann^{16,36}, S. Camera^{37,38,39}, G. Cañas-Herrera^{40,41,42}, V. Capobianco³⁹, C. Carbone⁴³, J. Carretero^{44,45}, M. Castellano²⁰, G. Castignani¹⁸, S. Cavaoti^{35,46}, K. C. Chambers⁴⁷, A. Cimatti⁴⁸, C. Colodro-Conde⁴, G. Congedo⁴⁹, C. J. Conselice¹¹, L. Conversi^{50,22}, Y. Copin⁵¹, A. Costille⁵², F. Courbin^{53,54}, H. M. Courtois⁵⁵, M. Cropper⁵⁶, A. Da Silva^{57,58}, H. Degaudenzi⁵⁹, G. De Lucia²⁶, C. Dolding⁵⁶, H. Dole²¹, M. Douspis²¹, F. Dubath⁵⁹, X. Dupac²², S. Dusini⁶⁰, S. Escoffier⁶¹, M. Farina⁶², S. Ferriol⁵¹, K. George⁶³, C. Giocoli^{18,30}, B. R. Granett²³, A. Grazian⁶⁴, F. Grupp^{12,63}, S. V. H. Haugan⁶⁵, I. M. Hook⁶⁶, F. Hormuth⁶⁷, A. Hornstrup^{68,69}, P. Hudelot⁷⁰, M. Jhabvala⁷¹, E. Keihänen⁷², S. Kermiche⁶¹, A. Kiessling⁷³, M. Kilbinger²⁴, B. Kubik⁵¹, M. Kümmel⁶³, H. Kurki-Suonio^{74,75}, Q. Le Boulc'h⁷⁶, A. M. C. Le Brun⁷⁷, D. Le Mignant⁵², P. B. Lilje⁶⁵, V. Lindholm^{74,75}, I. Lloro⁷⁸, G. Mainetti⁷⁶, D. Maino^{79,43,80}, E. Maiorano¹⁸, O. Marggraf⁸¹, M. Martinelli^{20,82}, N. Martinet⁵², F. Marulli^{17,18,30}, R. Massey⁸³, S. Maurogordato⁸⁴, H. J. McCracken⁷⁰, E. Medinaceli¹⁸, S. Mei^{85,86}, M. Melchior⁸⁷, M. Meneghetti^{18,30}, E. Merlin²⁰, G. Meylan⁸⁸, A. Mora⁸⁹, M. Moresco^{17,18}, L. Moscardini^{17,18,30}, R. Nakajima⁸¹, C. Neissner^{90,45}, S.-M. Niemi⁴⁰, C. Padilla⁹⁰, S. Paltani⁵⁹, F. Pasian²⁶, K. Pedersen⁹¹, W. J. Percival^{92,93,94}, V. Pettorino⁴⁰, G. Polenta³¹, M. Poncet⁹⁵, L. A. Popa⁹⁶, L. Pozzetti¹⁸, F. Raison¹², R. Rebolo^{4,97,98}, A. Renzi^{99,60}, J. Rhodes⁷³, G. Riccio³⁵, E. Romelli²⁶, M. Roncarelli¹⁸, R. Saglia^{63,12}, A. G. Sánchez¹², D. Sapone¹⁰⁰, J. A. Schewtschenko⁴⁹, M. Schirmer², P. Schneider⁸¹, T. Schrabback¹⁰¹, A. Secroun⁶¹, S. Serrano^{14,102,13}, P. Simon⁸¹, C. Sirignano^{99,60}, G. Sirri³⁰, J. Skottfelt¹⁰³, L. Stanco⁶⁰, J. Steinwagner¹², P. Tallada-Crespi^{44,45}, A. N. Taylor⁴⁹, I. Tereno^{57,104}, S. Toft^{105,106}, R. Toledo-Moreo¹⁰⁷, F. Torradeflot^{45,44}, I. Tutusaus¹⁰⁸, L. Valenziano^{18,109}, J. Valiviita^{74,75}, T. Vassallo^{63,26}, G. Verdoes Kleijn¹¹⁰, A. Veropalumbo^{23,33,32}, Y. Wang¹¹¹, J. Weller^{63,12}, A. Zacchei^{26,25}, G. Zamorani¹⁸, F. M. Zerbi²³, I. A. Zinchenko⁶³, E. Zucca¹⁸, V. Allevato³⁵, M. Ballardini^{112,113,18}, M. Bolzonella¹⁸, E. Bozzo⁵⁹, C. Burigana^{114,109}, R. Cabanac¹⁰⁸, A. Cappi^{18,84}, J. A. Escartin Vigo¹², L. Gabarra¹¹⁵, W. G. Hartley⁵⁹, J. Martín-Fleitas⁸⁹, S. Matthew⁴⁹, R. B. Metcalf^{17,18}, A. Pezzotta^{116,12}, M. Pöntinen⁷⁴, I. Risso¹¹⁷, V. Scottez^{118,119}, M. Sereno^{18,30}, M. Tenti³⁰, M. Wiesmann⁶⁵, Y. Akrami^{120,121}, S. Alvi¹¹², I. T. Andika^{122,123}, S. Anselmi^{60,99,124}, M. Archidiacono^{79,80}, F. Atrio-Barandela¹²⁵, D. Bertacca^{99,64,60}, M. Bethermin¹²⁶, L. Bisigello⁶⁴, A. Blanchard¹⁰⁸, L. Blot^{127,77}, S. Borgani^{128,25,26,27,129}, M. L. Brown¹¹, S. Bruton¹³⁰, A. Calabro²⁰, F. Caro²⁰, T. Castro^{26,27,25,129}, F. Cogato^{17,18}, S. Davini³³, G. Desprez¹¹⁰, A. Díaz-Sánchez¹³¹, J. J. Diaz⁴, S. Di Domizio^{32,33}, J. M. Diego¹³², P.-A. Duc¹²⁶, A. Enia^{29,18}, Y. Fang⁶³, A. G. Ferrari³⁰, A. Finoguenov⁷⁴, A. Fontana²⁰, A. Franco^{133,134,135}, J. García-Bellido¹²⁰, T. Gasparetto²⁶, V. Gautard¹³⁶, E. Gaztanaga^{13,14,137}, F. Giacomini³⁰, F. Gianotti¹⁸, M. Guidi^{29,18}, C. M. Gutierrez¹³⁸, A. Hall⁴⁹, S. Hemmati¹³⁹, H. Hildebrandt¹⁴⁰, J. Hjorth⁹¹, J. J. E. Kajava^{141,142}, Y. Kang⁵⁹, V. Kansal^{143,144}, D. Karagiannis^{112,145}, C. C. Kirkpatrick⁷², S. Kruk²², L. Legrand^{146,147}, M. Lembo^{112,113}, F. Lepori¹⁴⁸, G. Leroy^{149,83}, J. Lesgourgues¹⁵⁰, L. Leuzzi^{17,18}, T. I. Liaudat¹⁵¹, J. Macias-Perez¹⁵², M. Magliocchetti⁶², F. Mannucci¹⁵³, R. Maoli^{154,20}, C. J. A. P. Martins^{15,16}, L. Maurin²¹, M. Miluzio^{22,155}, P. Monaco^{128,26,27,25}, G. Morgante¹⁸,

* Corresponding author: grant.stevens@bristol.ac.uk

K. Naidoo¹³⁷, A. Navarro-Alsina⁸¹, F. Passalacqua^{99,60}, K. Paterson², L. Patrizii³⁰, A. Pisani⁶¹,
D. Potter¹⁴⁸, S. Quai^{17,18}, M. Radovich⁶⁴, P.-F. Rocci²¹, G. Rodighiero^{99,64}, S. Sacquegna^{134,133,135},
M. Sahlén¹⁵⁶, D. B. Sanders⁴⁷, E. Sarpa^{28,129,27}, A. Schneider¹⁴⁸, M. Schultheis⁸⁴, D. Sciotti^{20,82},
E. Sellentin^{157,42}, F. Shankar¹⁵⁸, L. C. Smith¹⁵⁹, K. Tanidis¹¹⁵, G. Testera³³, R. Teyssier¹⁶⁰, S. Tosi^{32,33,23},
A. Troja^{99,60}, M. Tucci⁵⁹, C. Valieri³⁰, D. Vergani¹⁸, G. Verza¹⁶¹, and N. A. Walton¹⁵⁹

(Affiliations can be found after the references)

Received 18 March 2025 / Accepted 13 August 2025

ABSTRACT

Light emission from galaxies exhibit diverse brightness profiles, influenced by factors such as galaxy type, structural features, and interactions with other galaxies. Elliptical galaxies feature more uniform light distributions, while spiral and irregular galaxies have complex, varied light profiles due to their structural heterogeneity and star-forming activity. In addition, galaxies with active galactic nuclei (AGN) feature intense, concentrated emission from gas accretion around supermassive black holes, superimposed on regular galactic light, while quasi-stellar objects (QSOs) represent extreme cases in which AGN emissions dominate their host galaxies. The challenge of identifying AGN and QSOs has been discussed many times in the literature, often requiring multi-wavelength observations. This paper introduces a novel approach to identify AGN and QSOs from a single image. Diffusion models have recently been developed in the machine-learning literature to generate realistic-looking images of everyday objects. Utilising the spatial resolving power of the *Euclid* VIS images, we created a diffusion model trained on one million sources, without using any source pre-selection or labels. The model learns to reconstruct light distributions of normal galaxies, since the population is dominated by them. We conditioned the prediction of the central light distribution by masking the central few pixels of each source and reconstructed the light according to the diffusion model. We further used this prediction to identify sources that deviate from this profile by examining the reconstruction error of the few central pixels regenerated in each source's core. Our approach, solely using VIS imaging, features high completeness compared to traditional methods of AGN and QSO selection, including optical, near-infrared, mid-infrared, and X-rays. Our study offers practical insights for refining diffusion models and broadening their applications throughout the *Euclid* survey area, underscoring the utility of this approach in diverse astronomical contexts beyond just AGN identification.

Key words. methods: observational – techniques: image processing – galaxies: active

1. Introduction

Ever since the discovery that the mass or dynamics of a spheroid within a galaxy correlates with the mass of any detectable supermassive black hole (SMBH) at its centre (e.g. Dressler 1989; Kormendy & Richstone 1995; Magorrian et al. 1998; Ferrarese & Merritt 2000, and more recently, Sahu et al. 2019; Davis et al. 2018, 2019 and references therein), it has been believed that most – if not all – galaxies with a spheroidal (or pseudo-spheroidal) component contain a SMBH at their centre. If a galaxy's emission contains evidence of its central SMBH being actively fuelled, it will emit non-thermal radiation, leading it to be classified as an active galactic nucleus (AGN). Depending upon the luminosity of that emission and its contrast with the stellar emission from the surrounding galaxy, it could have a classification ranging from a Seyfert galaxy (for the weakest emission) through to a quasi-stellar object or QSO (for the most luminous).

For imaging in wavebands such as the optical and near-IR, the presence and detectability of emission from the area directly around a SMBH depends on multiple factors. This can include the rate of infall of material into the region around the SMBH, the amount of obscuration (absorption or scattering) towards that region, the contrast between the SMBH-related emission and that of the surrounding galaxy, the redshift, and the intrinsic compactness of the stellar distribution of the host galaxy. The growth and mass of a central SMBH is intimately related to the evolution of its host galaxy, not least through the action of energy, momentum, and radiative feedback from outflows originating from the area around the SMBH on the properties of the stellar populations and ISM (see Harrison & Ramos Almeida 2024 for a recent review). Consequently, identifying and charac-

terising SMBHs, particularly during periods of active feedback, is central to the study of SMBH and galaxy evolution.

A more precise classification of a source as an AGN is possible when a much wider range of multi-wavelength data is available. This can include spectroscopy and considering the much broader spectral energy distribution of the galaxy and/or the candidate AGN. However, it is still worth considering the efficient selection of AGN candidates from single-band imaging as a first step in identifying large samples of AGN for further study. In particular, with the advent of the uniform, deep, and high-spatial-resolution optical imaging dataset from the *Euclid* mission, covering a large fraction of the extragalactic sky, there is the potential to derive unprecedentedly large samples of AGN across a wide range of the (optical) luminosity and redshift parameter space with the use of an appropriate and sufficiently efficient technique. The technique can then subsequently be refined by feeding back the result of using other multi-wavelength data, allowing us to confirm or reject candidates. With sufficient refinement, this further assessment step eventually becomes either unnecessary, or at the very least, an efficient stage in compiling reliable samples. This then opens the door to the production of unprecedentedly large and reliable samples of optically selected AGN.

This work presents a technique that can achieve the above goal. In the following, we describe the use of diffusion models for generating samples of AGN candidates using *Euclid* VIS I_E band (Euclid Collaboration: Mellier et al. 2025; Euclid Collaboration: Cropper et al. 2025) imaging alone. Diffusion-based methods are machine-learning models that generate or reconstruct data by evolving a random distribution of pixels into a structured output over several steps. This technique outperforms traditional methods such as median pixel

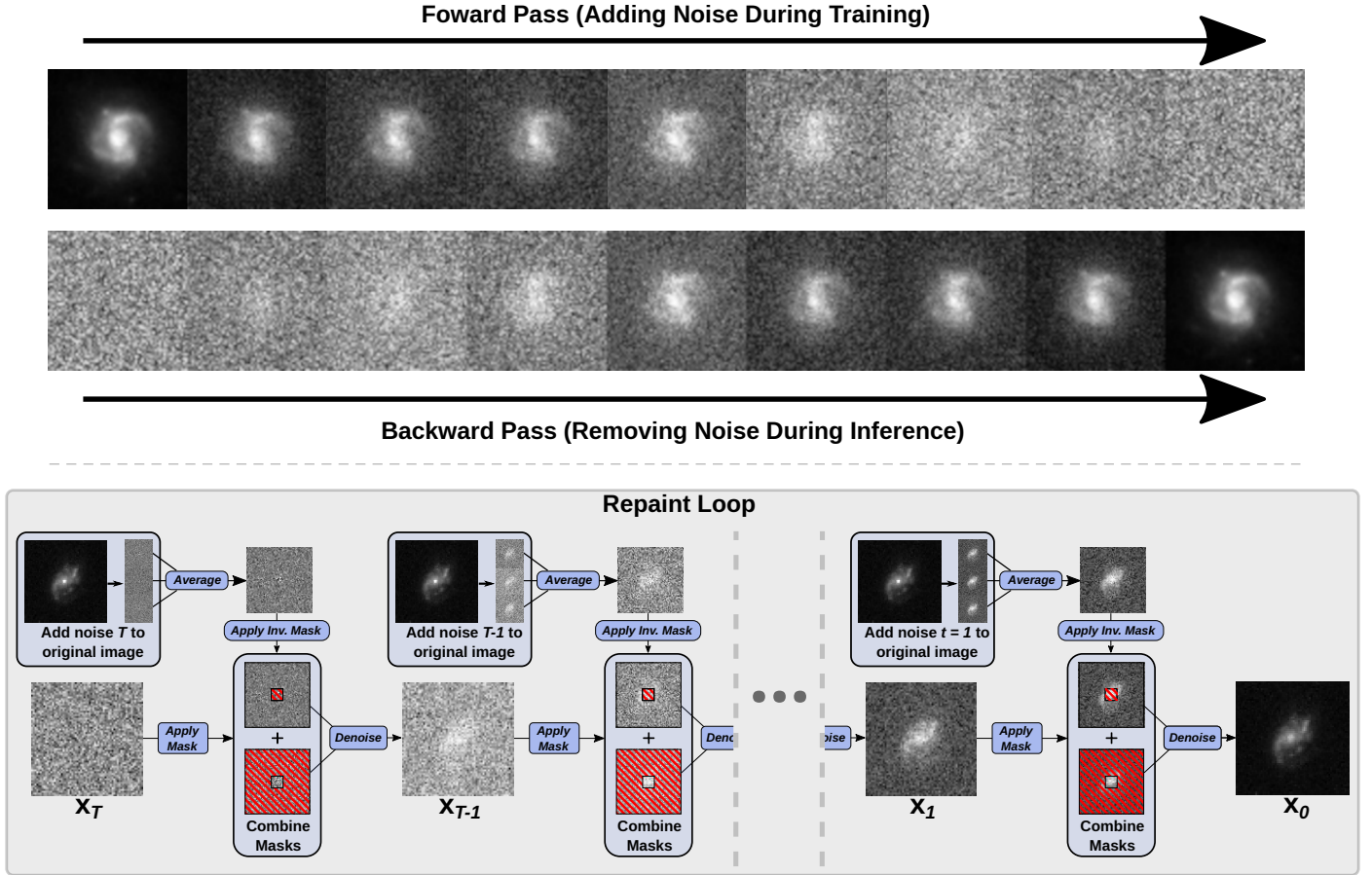


Fig. 1. Diffusion pipeline (top) that progressively adds noise to images, training the model to predict what noise was added from the previous step. Once trained and during inference, the model takes pure Gaussian noise as input and can iteratively remove the noise until a realistic galaxy image remains. Repeat inference runs will provide a different and unique galaxy from those it was trained on. The Repaint pipeline (bottom) takes the trained diffusion model and enables conditioning to allow parts of an existing image to be preserved by masking. At each denoising step, noise levels in the preserved pixels are adjusted to ensure they integrate correctly with the newly generated sections. After T iterations, the output includes the retained pixels and newly generated areas, creating a different yet plausible final image.

substitution or interpolation, which often struggle to recreate complex structures within images. Using these models for ‘inpainting’, whereby we adaptively recreate parts of the image, we can measure the discrepancies between the original and generated pixels. Examining these errors allows us to separate AGN, QSOs, and other sources from the broader galaxy population.

Our method¹ identifies potential AGN by treating the task as an unsupervised anomaly detection problem. It begins with an image of a galaxy, potentially containing AGN emission at its centre, and masks out the central pixels that might be influenced by such emission. A diffusion model, trained on a large, unlabeled dataset of galaxies – most of which are normal, non-active – then predicts or ‘inpaints’ the masked region based on the learned prior distribution of typical galaxy morphologies. Because AGN are rare, there is no requirement to remove them from the training sample; the model is inherently biased towards reconstructing the central region as if it belonged to a typical, non-active galaxy. As a result, any significant discrepancy between the inpainted and actual central region suggests the presence of an AGN. This reconstruction error thus serves as a powerful anomaly score for identifying candidate AGN.

While this paper details the fundamental technique applied to VIS imaging data, as this approach utilises machine learning, future improvements could incorporate findings from a detailed assessment that involves additional datasets. These could include multi-wavelength cross-matching or spectroscopic validation, from which we can ultimately reduce the need for further time-expensive refinement steps. Such enhancements would help to further improve the accuracy and reliability of the AGN candidacy samples generated by the technique. Additionally, the technique presented here uses a standard treatment of diffusion models, which in principle can be significantly further optimised for the specific characteristics of the *Euclid* dataset, thereby improving further the efficiency of the processing of the *Euclid* data. Such work would be the subject of future papers.

2. Diffusion-based inpainting

2.1. Generative models

Generative models are a class of machine learning systems designed to learn an underlying data distribution, $p(x)$, from a finite set of independent and identically distributed (i.i.d.) samples, $\{x_i\}_{i=1}^N \sim p(x)$. The primary goal of these models is to develop a tractable method for generating new, synthetic samples that are statistically indistinguishable from the

¹ The code is publicly available here: <https://github.com/grant-m-s/Astro-Repaint>

training data. Over the years, several distinct families of generative models have emerged, each with unique architectural and training paradigms. A key distinction lies in their optimisation objectives, which often seek to minimise a measure of divergence between the true data distribution, $p(x)$, and the model's learned distribution, $p_\theta(x)$. The choice of divergence has significant implications for model behavior. Prominent examples include variational autoencoders (VAEs, Kingma & Welling 2013), which learn a latent variable model of the data, and generative adversarial networks (GANs, Goodfellow et al. 2014, 2020), which involve a contest between a generator and a discriminator. In practice, GANs implicitly optimise an objective related to the reverse Kullback-Leibler (KL) divergence, $D_{\text{KL}}[p_\theta(x)||p(x)]$, which exhibits a ‘mode-seeking’ property. This encourages the generation of high-fidelity samples but can lead to mode collapse, whereby the model learns only a subset of the data distribution (Thanh-Tung & Tran 2020). In contrast, VAEs and diffusion models optimise objectives related to the forward KL divergence, $D_{\text{KL}}[p(x)||p_\theta(x)]$, which is ‘mass-covering’. This property encourages the model to account for all modes present in the data, leading to greater sample diversity and more comprehensive coverage of the data distribution.

2.2. Diffusion models

Diffusion models (Sohl-Dickstein et al. 2015; Ho et al. 2020), the focus of this work, systematically corrupt data with noise and then learn to reverse the process. They have been shown to achieve superior sample quality and mode coverage compared to other classes of models (Dhariwal & Nichol 2021), making them a compelling choice for modelling the complex and diverse morphologies of galaxies observed by *Euclid*.

For both GANs and VAEs, their significant advantage compared to diffusion models is that they offer faster sample generation and lower computational costs. Conversely, diffusion models trade off speed for superior sample quality, mode coverage, and flexibility in controllable generation, making them our preferred choice in the task of AGN identification.

This paper makes use of one such implementation, denoising diffusion probabilistic models (DDPMs, Ho et al. 2020). Diffusion works on the following premise: by continually adding a small amount of noise to an image, eventually, one is left with an image of complete noise without any remnants of the original input. If the noise is generated in a stochastic but consistent way, such as from a Gaussian distribution, we can use a network to learn the dynamics of the noise and perform a reverse process. By predicting the noise that was added to an image, x , at timestep t , we can remove this noise to produce image x_{t-1} . This forms an iterative process originating from some final timestep, T , an image of pure noise, to a realistic image from the data distribution at timestep 0.

The formal definition of this diffusion process, following Ho et al. (2020) and Nichol & Dhariwal (2021), is expressed as $q(x_1, \dots, x_T)$, which represents the joint distribution of a sequence in which each image, x_t , is progressively noised.

$$q(x_1, \dots, x_T|x_0) := \prod_{t=1}^T q(x_t|x_{t-1}), \quad (1)$$

where $q(x_t|x_{t-1})$ specifies the conditional distribution at each timestep, t , modelling the incremental addition of noise:

$$q(x_t|x_{t-1}) := \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I). \quad (2)$$

The β_t term defines the variance of the Gaussian noise added at each timestep. By ensuring a gradual increase in noise, the model can more easily learn the transition from the data distribution to the noise distribution.

With a large enough T and an adjusted time-dependent variance, β_t , x_T will approach a Gaussian distribution that is isotropic, meaning uniform in all directions. For the denoising process, starting at a $x_T \sim \mathcal{N}(0, I)$, accurately modelling $q(x_{t-1}|x_t)$ is not tractable, requiring a neural network to approximate it. An example of the full diffusion process can be seen in the top panel of Fig. 1.

2.2.1. Background noise versus diffusion noise

It is important to clarify the distinctions between the definition and use of ‘noise’ in diffusion models compared to the traditional understanding of noise in astronomy-based data. In the context of astronomical imaging, noise refers to the random fluctuations inherent in observational data, often caused by, for example, background emission, electronics read-out noise, and atmospheric disturbances. This ‘background noise’ reduces the signal-to-noise ratio (S/N) of the image, affecting the ability to detect subtle details in any analysis.

In contrast, the diffusion process uses an intentional integration of Gaussian noise to facilitate the training of the generative model. The pixels of image x_T follow the distribution $\mathcal{N}(0, I)$, rather than following the distribution of the background noise. The goal of the reverse diffusion process is to remove the artificial, high-variance Gaussian noise that was intentionally added during the forward process, thereby transforming a sample from a simple prior distribution $[\mathcal{N}(0, I)]$ into a realistic galaxy image. A realistic galaxy image, as defined by our training dataset, inherently includes the source’s light profile superimposed on the characteristic background noise of the *Euclid* instrument. Therefore, the model learns to generate images that replicate this entire distribution, including the statistical properties of the background noise. The diffusion pipeline does not aim to produce a ‘denoised’ image in the traditional astronomical sense (i.e. a background-subtracted image).

2.2.2. Use in astronomy applications

Smith et al. (2022) also utilised the DDPM framework to generate a dataset of realistic galaxies. The model was trained using the Photometry and Rotation curve OBservations from Extragalactic Surveys (PROBES) dataset (Stone & Courteau 2019; Stone et al. 2021), a collection of large, well-resolved objects that feature significant internal structure. They provide analysis of the quality of their generated images, showcasing the diffusion model’s ability to create visually realistic images that feature similar physical property distributions such as the half-light radius and flux-space colour values. They also briefly explore the use of inpainting with their model to remove satellite trails.

When researching diffusion models, papers utilising score-based models are often used and spoken of interchangeably, as they are different implementations of the same generative process (Smith et al. 2022). Rather than the fixed sequence of timesteps to denoise data, score-based methods apply stochastic differential equations (SDEs) to estimate the data distribution’s gradient (score). Working with SDEs and gradients allows for a continuous range of possible diffusion paths. It is for this reason that score-based models can be classed as the more general framework for diffusion-based generative models. The use of score-based generative models within

astronomical applications includes galaxy image deconvolution (Adam et al. 2025; Spagnoletti et al. 2024), gravitation lensing analysis (Adam et al. 2022; Remy et al. 2023), and deblending (Sampson et al. 2024).

2.3. Training objective

For training, we used 64×64 pixel VIS cutouts for each of the selected sources that are discussed in Sect. 4.1. We used the training pipeline described in Dhariwal & Nichol (2021)² and adopted a hybrid loss function that combines a simple mean squared error (MSE) term with the full variational lower bound (VLB).

The MSE objective, proposed by Ho et al. (2020), is an MSE loss between the true and predicted Gaussian noise, ϵ ,

$$L_{\text{MSE}} = \mathbb{E}_{t, x_0, \epsilon} \left[\left\| \epsilon - \epsilon_\theta \left(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right\|_2^2 \right], \quad (3)$$

where x_0 is an original image, t is a uniformly sampled timestep, and $\epsilon \sim N(0, I)$ is the true noise. For our implementation $\bar{\alpha}_t$ is described below (the reader is directed to Nichol & Dhariwal (2021) for specifics on its derivation):

$$\bar{\alpha}_t = \cos \left(\frac{t/T + 0.008 \pi}{1.008} \frac{\pi}{2} \right)^2. \quad (4)$$

While this objective is effective for training the mean of the reverse process transition, $\mu_\theta(x_t, t)$, it provides no learning signal for its variance, $\Sigma_\theta(x_t, t)$. To address this, Nichol & Dhariwal (2021) proposed learning this variance by incorporating the full VLB of the log-likelihood into the training objective. The VLB is given by

$$L_{\text{VLB}} = L_0 + L_1 + \dots + L_{T-1} + L_T, \quad (5)$$

$$L_{t-1} = D_{\text{KL}} [q(x_{t-1} | x_t, x_0) \| p(x_{t-1} | x_t)], \quad (6)$$

$$L_T = D_{\text{KL}} [q(x_T | x_0) \| p(x_T)]. \quad (7)$$

Since L_{VLB} explicitly depends on the parameters of Σ_θ , it can be used to train them. We adopted the hybrid objective

$$L_{\text{hybrid}} = L_{\text{MSE}} + \lambda L_{\text{VLB}}. \quad (8)$$

We used a small weighting constant ($\lambda = 0.001$) to prevent the L_{VLB} term from overpowering the L_{MSE} term, which remains the primary driver of sample quality. This hybrid loss allows the model to learn an optimal reverse process variance, which can improve model likelihoods and enable more efficient sampling. A key challenge with astronomical data is its high dynamic range. A standard MSE-based loss is not scale-invariant and will be dominated by the few brightest pixels or sources in a training batch. This can lead to poor performance on the much larger population of fainter objects. To address this, we introduced a normalised version of the MSE loss term. For each image, x_i , we divided its error by its respective maximum pixel value, denoted as $\max(x_i)$. This normalisation ensures that the relative error is penalised, allowing for more equitable training across both bright and faint sources. The final loss used for our primary model is thus

$$L_{\text{normalised}} = \frac{L_{\text{MSE}}}{\max(x_i)} + \lambda L_{\text{VLB}}. \quad (9)$$

Throughout the remainder of the paper, we refer to this variant of the loss as the normalised loss. The training curves of

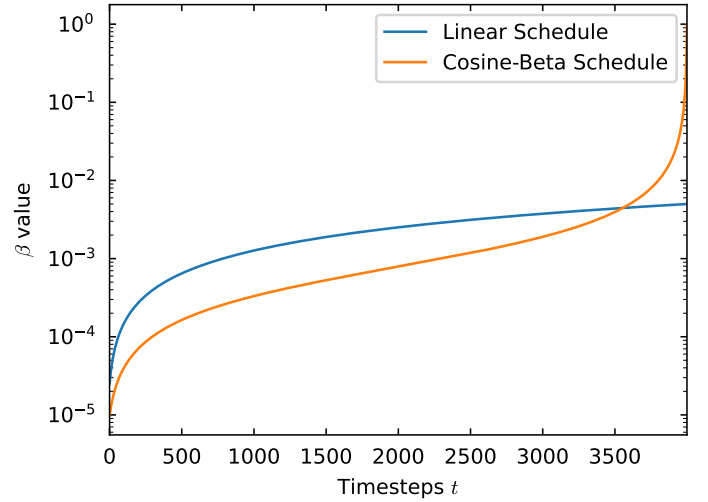


Fig. 2. Linear schedule from the original diffusion implementation, which causes the parameters to converge early in the timesteps, resulting in training images becoming pure noise too soon and leading to suboptimal performance. The switch to the cosine-beta schedule adds noise at a much slower rate, prioritising smaller updates in the early stages, leading to more unique noised images throughout training.

this scale-invariant loss compared with the original loss used in Nichol & Dhariwal (2021) are shown in Fig. B.1. By incorporating this normalised loss, we enable more equitable training across the full range of galaxy brightnesses, rather than being dominated by the brightest objects, which occurs when using the default loss.

2.4. Noise scheduling

The total number of timesteps, T , is a critical hyperparameter in the diffusion process. A larger T corresponds to smaller, more refined steps in both the forward (noising) and reverse (denoising) processes. While this can, in principle, allow the model to learn a more accurate representation of the data distribution, it comes with a significant trade-off. The computational cost of inference (i.e. generating a sample) is directly proportional to T , as the reverse process must be iterated T times. The cost of a single training step, however, is not directly dependent on T , because at each step a random timestep, t , is sampled and the loss is evaluated only for that level of noise.

In the original DDPM paper, Ho et al. (2020) made use of a linear schedule for the amount of noise added to the image at each t . Although the produced images competed with state-of-the-art methods, Nichol & Dhariwal (2021) found that a linear schedule was not well suited to low-resolution images and proposed using a cosine-beta noise schedule as a replacement. By stretching the noise levels to add less noise in early iterations and more in the latter, the process overcame the issue of images becoming too noisy too quickly. Figure 2 shows how the scheduler impacts the β parameter used in Eq. (2). Equation (10) shows how β_t is calculated (the reader is directed to Nichol & Dhariwal 2021 for specifics on its derivation):

$$\beta_t = 1 - \frac{\bar{\alpha}_{t+1}}{\bar{\alpha}_t}, \quad (10)$$

where $\bar{\alpha}_t$ is shown in Eq. (4).

All models in this paper make use of the cosine-beta scheduler. The impact of this scheduler on the high dynamic range nature of astronomy data is explored in detail in Sect. 6.1.

² Available at <https://github.com/openai/guided-diffusion>

2.5. Conditioning the model

Using the pipeline of DDPMs allows us to impose conditioning on our input space. Throughout training, the model is already conditioned on the timestep, t . This conditioning is crucial, as it informs the neural network of the noise level present in the input image, x_t . The network learns a single function, $\epsilon_\theta(x_t, t)$, which is trained to predict the noise component, ϵ , from a noised image, x_t , at any given timestep, t . The characteristics of the noise to be removed differ substantially between a timestep close to T (where the image is almost pure noise) and a timestep close to 0 (where the image is almost fully reconstructed). Providing t as an input allows the single network to adapt its predictions accordingly.

However, we are not limited to only conditioning by timestep; often, multiple parameters are provided alongside each image. Typically, this is to allow ‘class-conditioning’, whereby users can tune the expected output of the diffusion model to a specific class or category (Radford et al. 2021; Zhang et al. 2023). By providing an embedding space during training, the model can learn not only the dynamics of the noise but also the inherent differences between noise and non-noise pixel values for each of the class representations. This conditioning provides small nudges to the random walk, guiding the denoising to the respective population of the class. Providing rich representations to the model has led to the impressive text-to-image capabilities of popular diffusion-based implementations, such as Dall-E (Ramesh et al. 2022) or Imagen (Saharia et al. 2022).

This type of conditioning allows the user to direct the image generation to a particular class or style of image. However, with diffusion’s ability to cover large areas of the search space, a repeat sample of the same prompt will likely produce an image very different from the first, even if the generated image belongs to the requested class. Even with the possibility of detailed prompts and refined classes, which can guide the generation towards a narrower or more specific region of the learned space, it may be difficult to know exactly what the output image will look like. This is due to the one-to-many relationship between classes or prompts and their subsequent generated outputs.

However, for our task, we require a different form of conditioning. Instead of generating an entire image based on a class label, we need to reconstruct a specific, masked region of an existing image, conditioned on the surrounding, unmasked pixels. This task is known as inpainting (Pathak et al. 2016; Yeh et al. 2017; Yu et al. 2018; Wang et al. 2018). To achieve this, we do not train a new, specialised conditional model. Instead, we employ a ‘plug-and-play’ (PnP) approach, whereby conditioning is introduced during the inference phase of an unconditionally trained generative model (Nguyen et al. 2017; Graikos et al. 2022). Our diffusion model, trained on a vast sample of galaxies, serves as a powerful, non-parametric prior for galaxy morphology. The PnP approach allows us to apply this prior to the inpainting task with great flexibility, as the model does not need to be retrained for different images or mask geometries. This is particularly advantageous for large-scale surveys like *Euclid*. Recent works in astrophysics have successfully employed similar PnP or posterior sampling strategies for complex inference tasks (Feng et al. 2023; Dia et al. 2025).

2.6. Repainting

For this work, we used the Repaint algorithm (Lugmayr et al. 2022), a PnP method specifically designed for diffusion models. The core challenge in diffusion-based inpainting is to

ensure semantic and structural consistency between the known (unmasked) and unknown (masked) regions. A naive approach of simply replacing the masked region with noise and then running the reverse process would fail, as the denoising network would alter the known, unmasked pixels, corrupting the conditioning information.

Repaint overcomes this by iteratively enforcing the data constraint at each step, t , of the reverse diffusion process. Given an image, x_0 , and a binary mask, M (where 1 indicates the region to be inpainted and 0 indicates the known region), the algorithm proceeds as follows. At each timestep, t , of the reverse process, starting from a noisy sample, x_t :

1. Sample the known region: The original image, x_0 , is noised to the appropriate level for timestep t using the forward process, $q(x_t|x_0)$, resulting in a sample of the known part, x_t^{known} .
2. Sample the unknown region: The output image from the previous, slightly noisier step, x_{t+1} , is denoised for one step using the unconditional diffusion model to get a sample of the unknown part, x_t^{unknown} .
3. Combine and denoise: The known and unknown parts are combined: $x'_t = (1 - M) \odot x_t^{\text{known}} + M \odot x_t^{\text{unknown}}$. This composite image is then passed to the next iteration of the repaint loop to produce the estimate for step $t - 1$. Once this image has been passed through one step of the learned reverse diffusion process $p_\theta(x_{t-1}|x'_t)$, it will produce the slightly less noisy image x_{t-1} .

This procedure, illustrated in Fig. 1, ensures that the generated region is coherently conditioned on the surrounding galaxy structure at every stage of the generation. The ‘denoise’ operation in this context refers to the application of one step of the learned reverse diffusion process, $p_\theta(x_{t-1}|x_t)$, which involves using the trained neural network, ϵ_θ , to predict the noise in x_t and deriving an estimate for x_{t-1} . By denoising an image composed of both fixed and generated pixels – each matched to the appropriate noise level t – we ensure that the generated region is properly conditioned on the information from the rest of the galaxy, resulting in a final image in which all pixels are consistent and coherent.

While alternative conditioning methods based on guiding the reverse process with a likelihood score function exist (Graikos et al. 2022), we adopted the Repaint algorithm for its straightforward implementation and proven effectiveness in enforcing data consistency in a PnP setting.

Incorporating the Repaint method for AGN identification has the following benefits:

- Diffusion methods have proven their ability in many applications to accurately recreate complex datasets, allowing for the generation of realistic galaxy morphology images.
- A large imbalance of normal galaxies exists compared to AGN, allowing the model to learn a bias. Such a bias can be exploited to create an outlier detection-based classifier without explicitly requiring labels during training.
- There are no constraints on the size and shape of masks used, and since Repainting is only applied at inference, no additional retraining is required.

3. Training and inference pipeline

The diffusion and inpainting methods are two distinct pipelines, one for the initial training of the model and a second for the actual inpainting. Training the diffusion model uses the whole image and learns to predict what noise to remove at each timestep, allowing for novel examples to be generated from the trained distribution.

3.1. Mask creation

Due to the nature of the application, our aim is to mask out any suspected AGN without the need to rely on labels ahead of time. Each mask is positioned so that the brightest pixel in the centre of the source becomes the centre of the mask. Although each image is centred on the source, various interactions and dynamics allow suspected AGN to be offset from the centre. To allow for some deviation, a 9×9 pixel window around the centre of the image is used to detect the brightest pixel. This offset allows for the correct masking of less symmetric sources, whilst limiting the likelihood of masking an adjacent source by mistake. The masks for each image are assigned automatically throughout the data-loading pipeline.

Throughout this paper, we use a 5×5 pixel mask, unless explicitly stated otherwise, centred on the calculated brightest pixel. A 5×5 pixel mask, which covers $0''.5 \times 0''.5$, allows us to effectively cover the AGN contribution, whilst minimising loss of the surrounding galaxy structure. Given that the VIS PSF has a full width half maximum (FWHM) of $0''.13$ (Euclid Collaboration: Mellier et al. 2025), the AGN's light is spread over multiple pixels. Taking into account the redshift of the galaxy, an AGN within nearby sources ($z = 0.01$) will cover 4.8 pixels, ensuring that we can fully encapsulate the core without unnecessary overlap of the host morphology. By retaining as much of the host galaxy information as possible for inpainting, we also reduce the risk of introducing artificial features into the reconstructed image.

3.2. Reconstruction rescaling

Unlike some image processing pipelines that rescale image data to a limited range, such as when working with JPG images that are binned to the standard 0–255 pixel scale range and converted to floats between 0–1 (Walmsley et al. 2020; Euclid Collaboration: Aussel et al. 2024), our approach utilises the raw values direct from the sensor as input to our network. As comparisons of individual pixels are being made on such a small mask, the relative differences become much more impactful and would likely be lost if values were binned.

One potential issue with this set-up is that the libraries used for training are optimised for images within the standard pixel range. Although this did not prevent the model from effectively learning the distribution, it does require a rescaling of the generated output as the pixels are mostly between -1 and $+1$ to improve training stability.

We experimented with simply normalising the input data so that all our images were in the range 0–1; however, due to the large dynamic range of the data, the faintest pixels ended up being 10^{-4} . Given the previous discussions on how MSE does not penalise smaller errors effectively, the model could not accurately recreate sources, outputting a near-empty image of very low pixel values.

To allow the inpainted region to be meaningfully compared (e.g. via MSE) to the original pixel values in that masked region, a post-inference rescaling must be applied. The vast majority of pixels in the image (the unmasked region, which constitutes 99.5% of a 64×64 image when a 5×5 central mask is used) correspond to known pixel values from the original astronomical image. We leveraged these known, unmasked pixels to perform a linear rescaling of the diffusion model's initial output (which is in a normalised range) back to the original image's dynamic range. This procedure ensures consistency in the unmasked regions and appropriately scales the newly inpainted pixels in the

masked region. The following scaling can be applied instantly on output of the repaint procedure using each image's respective values to allow the non-masked pixels of the input and output images to be consistent:

$$\text{output}_{\text{scaled}} = \text{output} \times \text{scale} + \text{offset}, \quad (11)$$

where

$$\text{scale} = \frac{\max(\text{input}_{\text{nonmasked}}) - \min(\text{input}_{\text{nonmasked}})}{\max(\text{output}_{\text{nonmasked}}) - \min(\text{output}_{\text{nonmasked}})}$$

and

$$\text{offset} = \min(\text{input}_{\text{nonmasked}}) - \text{scale} \times \min(\text{output}_{\text{nonmasked}}).$$

By ensuring that the shared pixels are consistent, the inpainted pixels can also be correctly adjusted.

4. Data

4.1. Sample selection

Covering 63.1 deg^2 , the Euclid Q1 Data Release (Euclid Quick Release Q1 2025) contains just under 30 million catalogued sources (Euclid Collaboration: Romelli et al. 2026) in the Euclid Deep Fields North (EDF-N), South (EDF-S), and Fornax (EDF-F). In this work, we use exclusively the VIS images (Euclid I_E band) because we aim to identify AGN candidates on a single image. However, due to the varying distances and sizes of sources, an appropriate subset must be chosen to ensure the model is presented with a sufficiently broad selection of resolved galaxies. The following criteria are applied to the 30 million source catalogue of Q1:

- The source must be detected in the VIS band ($\text{vis_det_euclid} > 0$);
- the source must have a I_E magnitude brighter than 22.5 [$23.5 - 2.5 \log(\text{flux_detection_total_euclid}) < 22.5$];
- the source must not have any defects or abnormalities as signified by the detection flags ($\text{det_quality_flag_euclid} = 0$ and $\text{flag_vis_euclid} = 0$);
- the source must have a low probability of being spurious ($\text{spurious_prob_euclid} < 0.2$);
- the source must have a large segmentation area ($\text{segmentation_area_euclid} > 200$);
- the S/N of the total flux measurement must be sufficiently high ($\frac{\text{flux_detection_total_euclid}}{\text{fluxerr_detection_total_euclid}} > 15$); and
- if the S/N criterion is not met, then the segmentation area of the source must be very large ($\text{segmentation_area_euclid} > 1200$), as defined in Euclid Collaboration: Walmsley et al. (2026).

It is crucial to note that because AGN are intrinsically rare compared to the general galaxy population, the training set does not need to be pre-filtered to remove them. The diffusion model, in learning the data distribution, naturally develops a prior that is heavily biased towards the 'typical' light profiles of non-active galaxies. This inherent bias is the key mechanism that enables the subsequent outlier detection.

After this sample selection, we are left with 1 142 606 sources across the three Q1 fields. VIS image cutouts (64×64 pixels) were created for each source. Problematic images, such as those at the very edge of tiles where a full cutout could not be created, were removed. This provided us with images and catalogue data for 1 060 864 sources. We then applied a 80:20 train and test split. Even though the inpainting

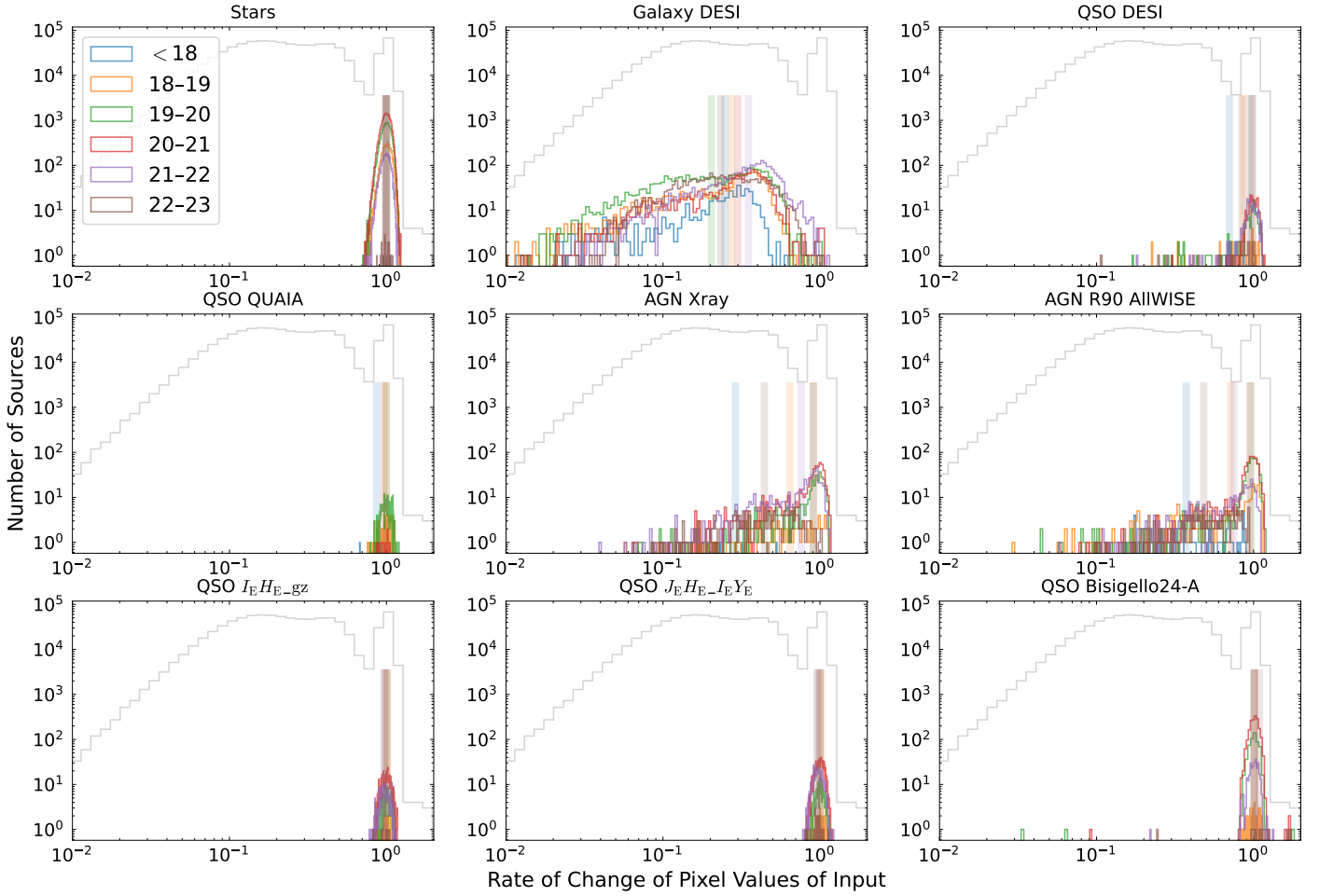


Fig. 3. Initial results for pixel value differences across various selections. Comparing the ratios of the centre’s brightest pixel with the means of the surrounding 1- and 2-pixel-wide regions shows a clear distribution difference between galaxy and non-galaxy classes. The grey histogram shows the distribution of the whole dataset, showing how the images not captured in these selections compare. The median value for each I_E magnitude bin is shown in the respective vertical line.

and its results are a separate pipeline to the diffusion training, all of the decision boundaries created are based solely on the training set to prevent any potential data leakage. When showing classifier performance, all results are from the unseen 20% of the data. This prevents any potential memorisation of the training images from impacting the results. However, in our analysis, no obvious differences have been found in the inpainting results between the two sets, showing that the model is not simply memorising the training set.

To explore the impact point-like sources have on training and the inpainting process, a second smaller selection was created that combined all the selections above with the following:

- The source must have a low probability of being a star. ($\text{phz_star_prob_phz_class} < 0.1$); and
- the source must have a low probability of being point-like. ($\text{point_like_prob_euclid} < 0.2$).

After removing problematic images as before, this smaller dataset contained 794 624 sources.

4.2. Comparisons with other classifications

Although our diffusion model is trained in an unsupervised manner without labels, it is essential to compare the distributions of metrics over different source selection criteria. This allows us to compare the performance of detecting particular sources and

determine what sources or factors may hinder the pipeline’s ability to classify objects sufficiently.

In particular, we use some of the sample definitions discussed in Euclid Collaboration: Matamoro Zatarain et al. (2026). A selection of stars are used from DESI (DESI Collaboration 2024) and *Gaia* (Gaia Collaboration 2023a,b). The normal galaxy sample is also from DESI (DESI Collaboration 2024). Comparisons are made to selections of AGN and quasars using *Euclid* photometry with the $I_E H_{g_z}$ and $J H_{I_E} Y$ selections from Euclid Collaboration: Matamoro Zatarain et al. (2026) and the two-colour selection from Euclid Collaboration: Bisigello et al. (2024, hereafter Bisigello24-A). Additional AGN and QSO selections are used from DESI (DESI Collaboration 2024), *Gaia*/Quaia (Storey-Fisher et al. 2024), and the 90% reliability WISE sample from Assef et al. (2018), which is referred to as R90 for the remainder of the paper. The reader is directed to Euclid Collaboration: Matamoro Zatarain et al. (2026) for a comprehensive overview of each of the selections used throughout our paper.

In this work, when referring to completeness against a specific reference sample (e.g. a catalogue of selected AGN candidates), we mean the fraction of objects in that reference sample that are successfully identified by our method. This is due to the true astrophysical completeness being unknown



Fig. 4. Distribution differences of the ROC value over each class. The narrow peaks from the non-galaxy class are significantly widened after applying the asinh transformation, causing a large overlap between classes.

and extremely difficult to measure from the observational data alone. Our context is equivalent to the standard machine learning metric of recall for the positive class, which is defined in Eq. (13).

When developing a method for initial candidate selection, whereby the goal is to flag objects for follow-up or multi-wavelength analysis, high recall is often prioritised. The aim is to minimise the loss of potentially interesting or rare sources (i.e. reduce false negatives), even if this comes at the cost of lower precision (i.e. including a larger fraction of contaminants or false positives in the selected sample). These contaminants can often be filtered out through subsequent analysis stages or by incorporating additional data. Our primary goal here is to demonstrate the method’s capability to recover a wide variety of known AGN and QSO types, making recall a key indicator of its potential utility in broad searches.

As is particularly true for AGN and QSOs, it is impossible to create a selection criterion that will identify a single population of objects with high completeness and high purity. Instead, we shall use the aforementioned samples, many of which are only candidates themselves, to assess the contamination and purity of our method to the extent that is possible.

4.3. Data expectations

As a baseline for our morphology-based approach, it is first necessary to verify that the *Euclid* VIS images contain sufficient information to resolve the contrast between the compact central emission characteristic of an AGN and the broader light profiles of typical galaxies. We analysed the relationship between the brightest central pixel and its surrounding pixels to quantify a difference between normal galaxies and AGN. By producing ratios comparing the central pixel to 1 and 2-pixels-wide regions around the central pixel, we formed a rate of change (ROC) metric over sources, which is described by

$$\text{ROC} = R_1^2 + R_2^2, \quad (12)$$

where

$$R_1 = \frac{\text{Input}_{\text{max}} - \text{Input}_{1s}}{\text{Input}_{\text{max}}}, \quad R_2 = \frac{\text{Input}_{1s} - \text{Input}_{2s}}{\text{Input}_{1s}},$$

Table 1. Percentage of galaxies that have a ROC value higher than the mean ROC of the other classes.

Galaxy overlap	Without asinh	With asinh
AGN	0.92%	3.98%
QSO	0.13%	0.72%
Stars	0.12%	1.80%

and Input_{1s} and Input_{2s} are the mean values of the surrounding rings of 1 (total 8 pixels) and 2 pixels (total 16 pixels), respectively. $\text{Input}_{\text{max}}$ is the pixel value of the brightest pixel within the centre 9×9 pixels.

Pixels surrounding the brightest pixel, which share values similar to the maximum, will produce a ROC value closer to zero, implying that the source does not feature a prominent active component. Sources featuring a single very bright pixel surrounded by fainter pixels will have a ROC value closer to its maximum of 2. The produced distributions of sources, analysed according to both I_E magnitude and various selection criteria (Sect. 4.2), are shown in Fig. 3. As was expected, sources with a high-flux point spread function (PSF) light profile show the highest ROC values. This effect is seen in the star sample and the QSO-dominated selections (R90 WISE, Quiaia, DESI-QSO, and *Euclid*-QSO). On the other hand, spectroscopically confirmed normal galaxies show very broad ROC distributions.

4.4. Addressing the use of asinh for image scaling

Given its prevalence in galaxy morphology-based research (Lupton et al. 2004), the effect of using an asinh transformation on the data is also investigated. The function is often used to benefit visual inspections by reducing the difference between the brightest and faintest pixels. In our approach, however, when attempting to differentiate galaxy and non-galaxy sources, the asinh function becomes a hindrance, causing the distributions to more significantly overlap, as is shown in Fig. 4. To generate a measure for class overlap, we measured the ROC for each image, taking the mean of each non-galaxy class. We used the percentage of our true galaxy sample with a ROC value over the respective class mean as our overlap metric. The extent of the overlap is showcased in Table 1, where galaxy overlap increases by at least a factor of 4 and in the extreme case, nearly 15 times more galaxies overlap with stars.

In the following sections of the paper, we discuss the results of the diffusion model and analyse the performance of the inpainting pipeline that automatically learns to reconstruct the central emission expected for the normal galaxy population, allowing us to identify stars and AGN as outliers.

5. Results

Although the focus of the paper is on the model’s performance on small areas of a galaxy’s core, Repaint’s adaptability allows the pipeline to handle a variety of mask shapes and sizes. More general examples of the inpainting performance can be seen in Fig. A.1. With larger masks, fewer pixels remain for conditioning, making the output more susceptible to artefacts. However, Lugmayr et al. (2022) demonstrate that this can be mitigated by increasing the number of resamples during the Repaint loop, at the cost of longer inference times.

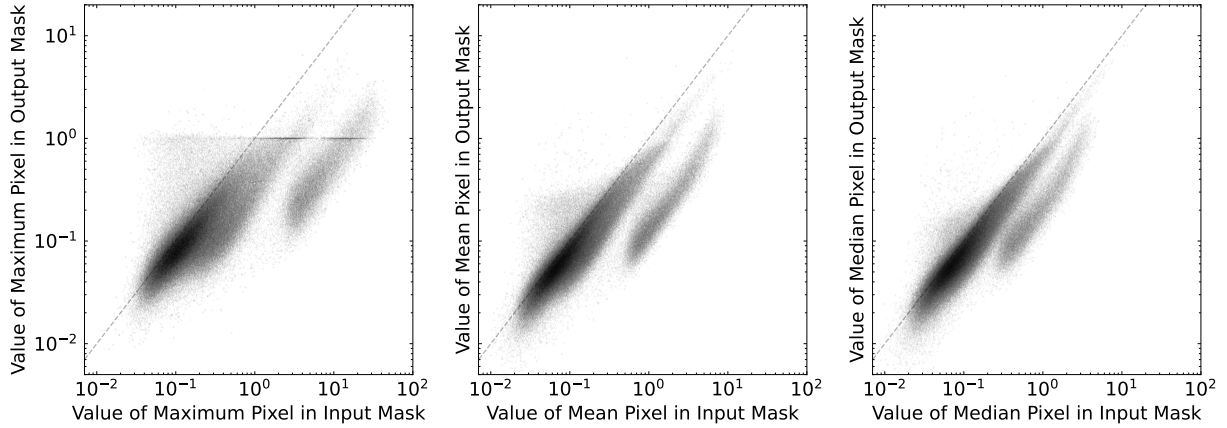


Fig. 5. Comparison between the original masked pixels and the pixels of the generated output. The model’s outputs demonstrated a consistent prediction along the gradient of $y = x$ (dashed line), with a bias for underpredicting the true value. A secondary cluster, also following the same gradient, is shown to reduce the pixel values by a factor of 10. This implies an inherent difference in the input image causing the model to behave differently. This collection of sources could be prime AGN candidates using our model.

5.1. Inpainting of AGN

For our inpainting, we used a 5×5 pixel mask centred on the brightest pixel, following the constraints described in Sect. 3.1. Comparing the expected pixel values with the generated outputs, Fig. 5 shows the model’s performance. The assumption that the model will predict a fainter core than the original does appear to be true, with a near-global bias for underpredicting, as is shown by the dashed $y = x$ line. The majority of sources behave in a consistent way, with the exception of a smaller cluster of sources that predict a core more than an order of magnitude fainter than the original. This distinct cluster could provide sufficient candidates for AGN.

The horizontal bar in the leftmost panel of Fig. 5, where the maximum output value appears to equal one, is a side effect of Eq. (11). The values themselves are not exactly one, with $<1\%$ of the data falling within $\epsilon = 0.01$ of having a maximum output of one. There does exist a convergence point in Eq. (11), where when the maximum and minimum pixel values of an input image equal one, the output of the rescaling always equals one; however, such an image would have all its pixels valued one, and so is not present in the data. Although no specific convergence point explains this trend, the majority of the images that fall on this line occur when the network outputs a maximum pixel value close to one. However, many more images have a network output close to one that is not close to this line, as well as images close to the line without a network output close to one. We have verified that both point-like and extended sources appear within this line; therefore, due to the low number of sources impacted, we leave more in-depth exploration to future work.

To further validate the model’s output and show it is producing realistic outputs, Fig. C.1 shows how the ROC metric compares between input and output images. The significant overlap, especially with respect to the peak of each distribution, shows that the model is accurately recreating the dynamics of the data.

To analyse the reconstruction error for each subset of class candidates, comparisons were made between the original pixels and inpainted reconstruction for each image. The MSE of the pixel values are shown in Fig. 6. Each subset was binned according to its I_E magnitude, providing a distinction between the relative brightness of each image.

There is a clear distribution difference between the respective class candidates. Brighter sources, particularly stars and luminous QSOs, produce higher errors, followed by fainter QSOs

and the lowest-magnitude AGN candidates. Galaxies and the faintest AGN candidates show significantly lower errors, showing the model’s tendency to reconstruct sources without strong AGN features more effectively. This indicates that the model has learned the anticipated bias introduced by the uneven distribution within the dataset. While MSE is scale-dependent and influenced by the source’s VIS I_E magnitude, the separation between the errors from the galaxy class and those of the other classes suggests it remains useful as a selection metric. The presence of low-error AGN candidates may also indicate contamination within the AGN labels.

It is acknowledged that because the model is trained on the full, unlabelled galaxy population, its prior distribution does contain information about AGN, even if they are rare. This could, in principle, lead the model to occasionally reconstruct AGN-like features, resulting in false negatives. However, the strong performance in separating known AGN and QSO populations, as is shown in Fig. 6, suggests that the prior is sufficiently dominated by non-active galaxies for this outlier detection framework to be effective.

To showcase how the performance of the inpainting method impacts the overall image, Fig. 7 presents the input, output, and residuals for a selection of sources, including a close-up of the masked area both before and after inpainting. It is clear how bright central cores can be entirely removed with inpainting and replaced by regions consistent with the surrounding pixels. The model can also accurately recreate images that do not feature any significant core. This shows that the model has a good internal representation of galaxies and how they should behave.

5.2. Max pixel difference ratio

Although the reconstruction error allows stars and a subset of AGN to stand out, a substantial overlap between suspected AGN and DESI-classified galaxies remains, especially at higher magnitudes. Even though measured on a pixel-by-pixel scale, MSE averages away specific pixel differences, leading to less nuanced differences between input and output. For example, if the reconstructed image ended up being the correct values but shifted by one pixel, this would be penalised heavily under MSE. From the perspective of this paper, such a scenario would represent a desirable outcome, as the model would correctly predict the presence of a bright, unresolved component, even if its exact location were

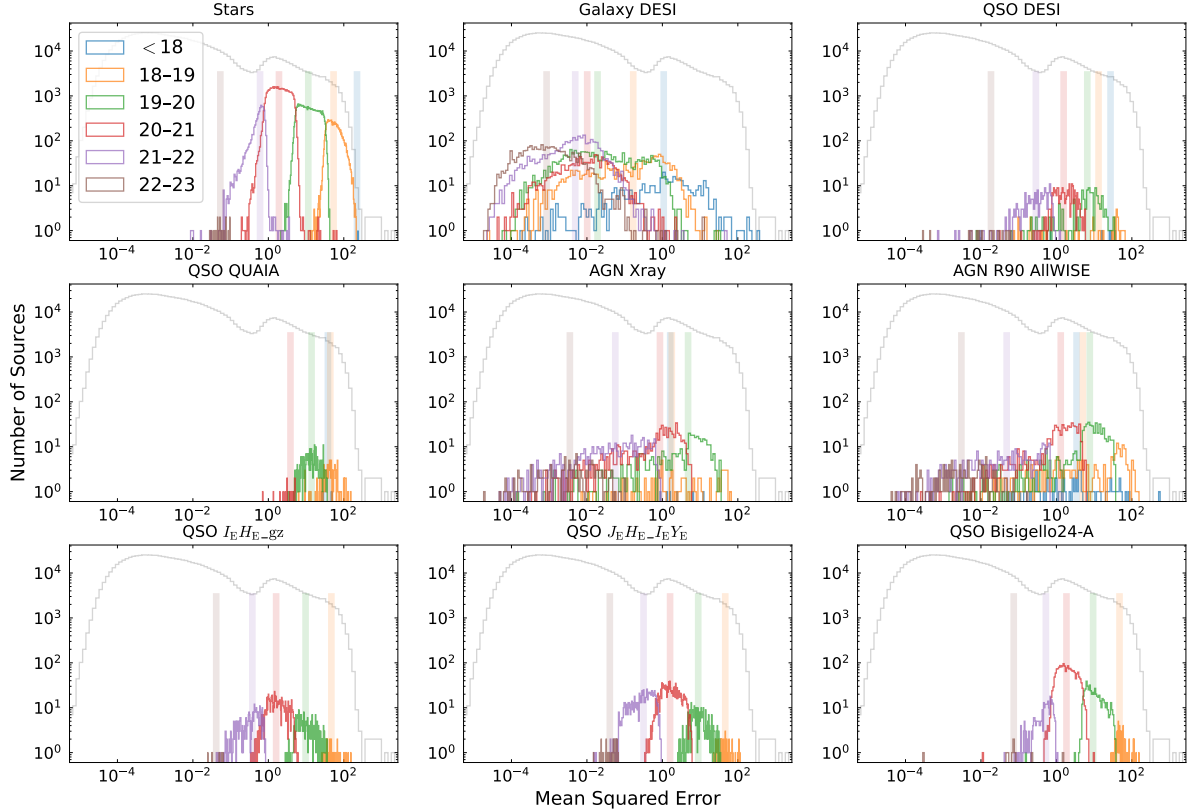


Fig. 6. MSE of the 5×5 pixel mask from the input and generated output. AGN selection show a skew toward higher errors, especially for lower-magnitude sources. The grey histogram shows the distribution of the whole dataset, showing how the errors of images not captured in these selections compare. The median value for each I_E magnitude bin is shown in the respective vertical line.

shifted by a few pixels within the mask. Another example would be if only the brightest pixel were different in the reconstruction. The MSE would significantly reduce the impact of this measurement, potentially scoring similarly to a source that gave minor errors in every pixel within the mask.

To ensure such cases are not overlooked, we measured the pixel value ratio between the brightest pixel from the input mask and the brightest pixel from the reconstruction. As is shown in Fig. 8, the produced distributions show a much tighter spread of values for many of the AGN selections, with a skew towards higher scores. The majority of the galaxy subset produces a ratio of less than 10, whereas a large population of suspected AGN from each selection features a larger ratio.

Another important difference between this metric and MSE is how the results are relative, reducing the impact of magnitude on the recorded values. Many of the high-density peaks remain consistent, even as the magnitude of the sources increases. This implies that this metric is more robust at detecting fainter AGN than the MSE.

The distribution also highlights the scale of the extreme values, with many sources showing pixel values approaching 100 times greater than their counterparts. Exploring these large disparities can provide insights into the types of sources that are significantly sensitive to the model’s capabilities. Sources producing the highest values for the ratio are those for which the reconstruction produces a cutout significantly fainter than the original image. As is seen in Fig. D.1, this space is dominated by point-like objects. Having the sharp fall-off in pixel intensity in such a small area leads to the mask covering the majority of the source, if not the entirety. When the model attempts to reconstruct the missing pixels, it generates values that are coherent

Table 2. Number of sources that produced brighter pixels in the masked area after repainting.

Dataset	Loss function	
	Default	Normalised
All data	55 008 (6.49%)	33 075 (3.90%)
No point-like	52 313 (6.17%)	39 507 (4.66%)

with the surrounding pixels. In these cases, the point-like source contributes very little to the surrounding pixels, leading to the generation of only background noise without any source.

The behaviour of the previous examples matches the model’s expected behaviour and the paper’s initial assumptions. Applying inpainting to this task is expected to produce outputs that are similar to, or produce a flattened light profile of, the original pixels. However, a subset of sources exists that exhibit the opposite behaviour and generate a much brighter collection of pixels. Figure D.2 shows examples of such sources. When analysing the sources that exhibit this behaviour, we found that they have overwhelmingly low-S/N images. Further exploration of the impact of S/N on the diffusion process is shown in Sect. 6.1.

The number of sources whose max pixel difference ratio is less than one for models trained on the two loss functions discussed in Sect. 2.3 is shown in Table 2. The normalised loss significantly reduces the number of sources exhibiting this behaviour. The normalised loss trained on the complete dataset was selected as our applied model because it further reduced the number of erroneous predictions.

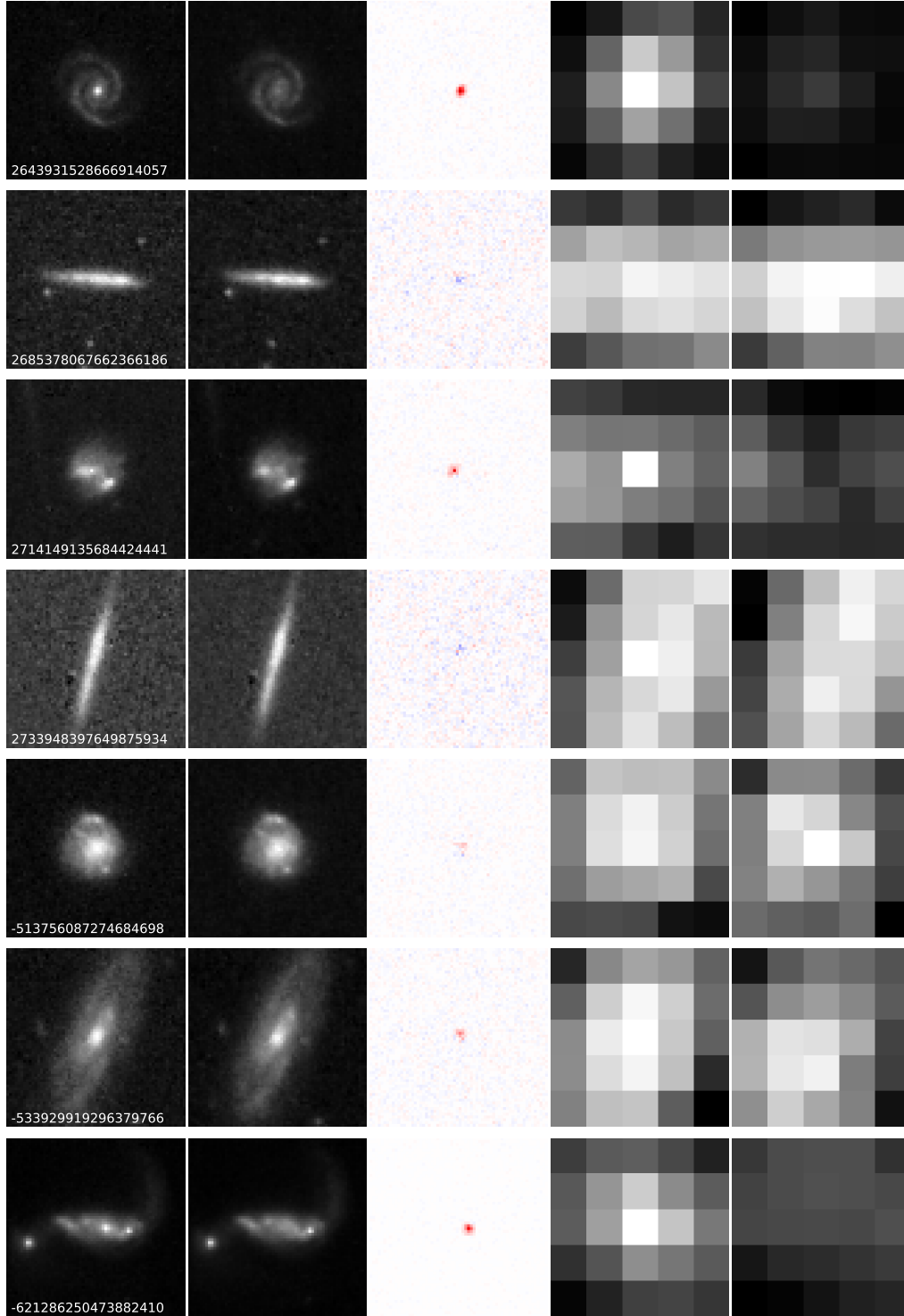


Fig. 7. Example outputs from the repainting model. A 5×5 pixel mask, centred on the brightest pixel within the central region of the input image, is repainted. The unmasked region is used to condition the generation of the new pixels. Each example shows (from left to right): the input image with its corresponding *Euclid* ID; the output image (same scale as input); the residuals of the generated image; the masked pixels from the original image; and the inpainted pixels of the output image. The first three columns are the original 64×64 cutout size, while the last two columns are the 5×5 masked region.

5.3. Morphology-based performance

Given the variety of shapes and sizes of objects that will be witnessed using *Euclid*, it is important to verify that the model can handle the complexities of different galaxy morphologies. Utilising the refined morphology classifications provided from

[Euclid Collaboration: Walmsley et al. \(2026\)](#), we provide examples of edge-on galaxies (`disk-edge-on_yes_fraction > 0.5`), spirals (`has-spiral-arms_yes_fraction > 0.7`), and mergers (`merging_merger_fraction > 0.3`) in Figs. [E.1](#), [E.2](#), and [E.3](#), respectively.

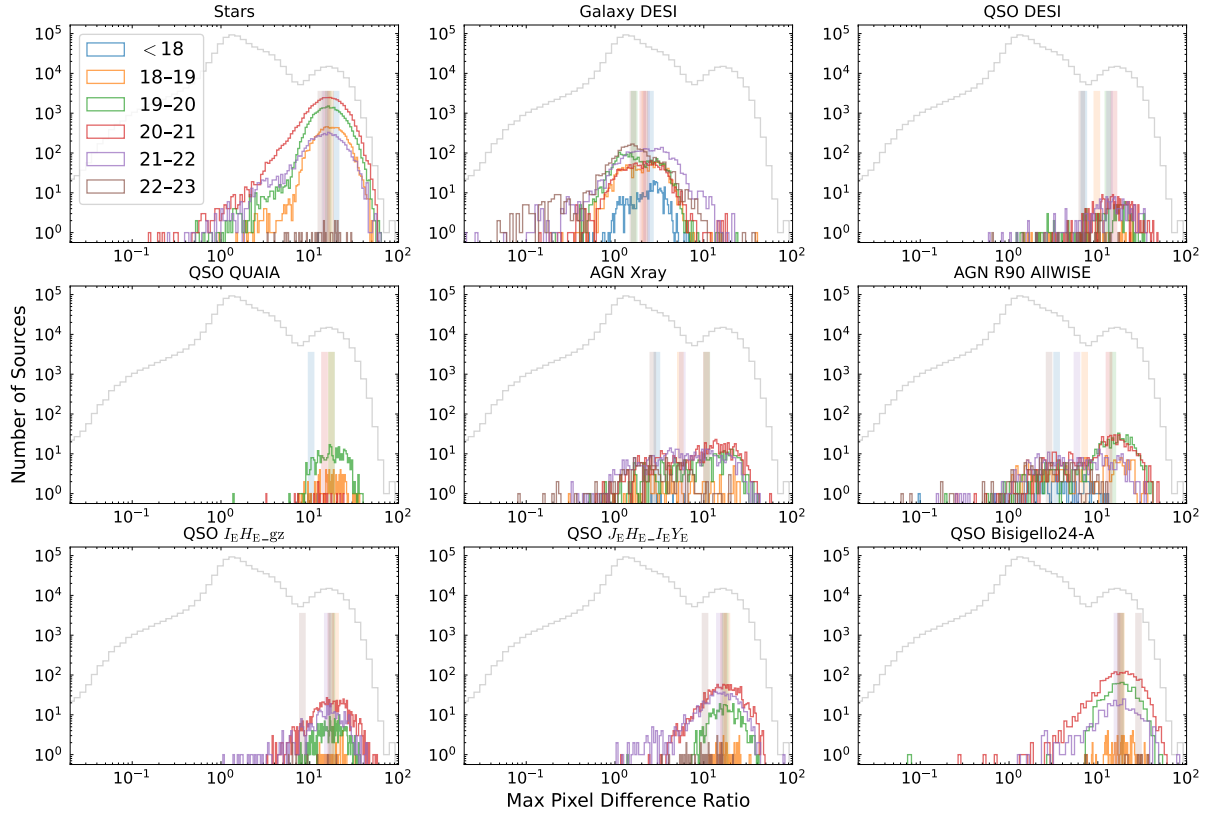


Fig. 8. Measured ratio between the maximum pixel within the masked pixels of the input image and the max of the pixels in the output. The galaxy sample is the only sample that does not feature a right skew, allowing a threshold between this class and the rest. The brightness differences caused by the varying magnitudes do not significantly alter the ratios within each class selection. The grey histogram shows the distribution of the whole dataset, showing how the ratio of images not captured in these selections compares. The median value for each I_E magnitude bin is shown in the respective vertical line.

These images show how our diffusion model and the inpainting pipeline can effectively adapt to different morphologies. For edge-on in particular, the region of bright pixels within the masked area is often narrower than the mask itself. Despite this, the inpainting correctly preserves the orientation of the features rather than uniformly filling the entire masked area with bright pixels.

5.4. Creating a classifier

Due to the difference in distributions between the two metrics, any non-overlapping cases can be found when used in combination. As is seen in Fig. 9, there is a cluster of sources in the top right. The key aspect of this cluster is its clear boundary from the DESI galaxy selection. This is vital for ensuring high reliability in classifications. We created a simple decision boundary between the two clusters of points, as is shown in Fig. 9. Although a more fine-tuned boundary would be possible, especially when incorporating a secondary machine-learning model, it is important to show the benefits of utilising the inpainting method through only its error metrics. The equation of our linear boundary is $y = 5.8x^{-0.06}$.

5.5. Variance of reconstructions

Given the stochastic nature of the diffusion model pipeline, it is important to verify that any produced results can remain consistent if applied to the wider dataset. Although the nonmasked region of the images will remain fixed over repeat samples, the

random walk performed in the inpainting procedure could lead the samples to a different area of the search space. This would result in a different, but possibly equally plausible, output image.

As part of a single application of the pipeline, seven samples are generated. To ensure that any low variance in samples is not caused by the model simply ‘remembering’ the true image from its training examples, all resampling is applied on images the model has never seen before. The mean and variance over a normalised-by-brightness MSE are shown in Fig. 10.

The top figure shows that, irrespective of magnitude, the model achieves a similar relative median error across all images, highlighting the impact of the normalised loss throughout training. However, the difference in the scheduler over images of different quality and brightness is evident in the bottom figure. The consistency of generated outputs varies drastically as fainter (and therefore typically lower S/N) images produce widely different pixels when resampled.

5.6. Comparing to other classifiers

It is important to compare our AGN and QSO predictions to more traditional methods that are also able to utilise only the image data.

5.6.1. Sérsic and PSF flux measurements

To benchmark our non-parametric inpainting method against a more traditional approach, we constructed a classifier based on parametric modelling of galaxy light profiles. We utilised

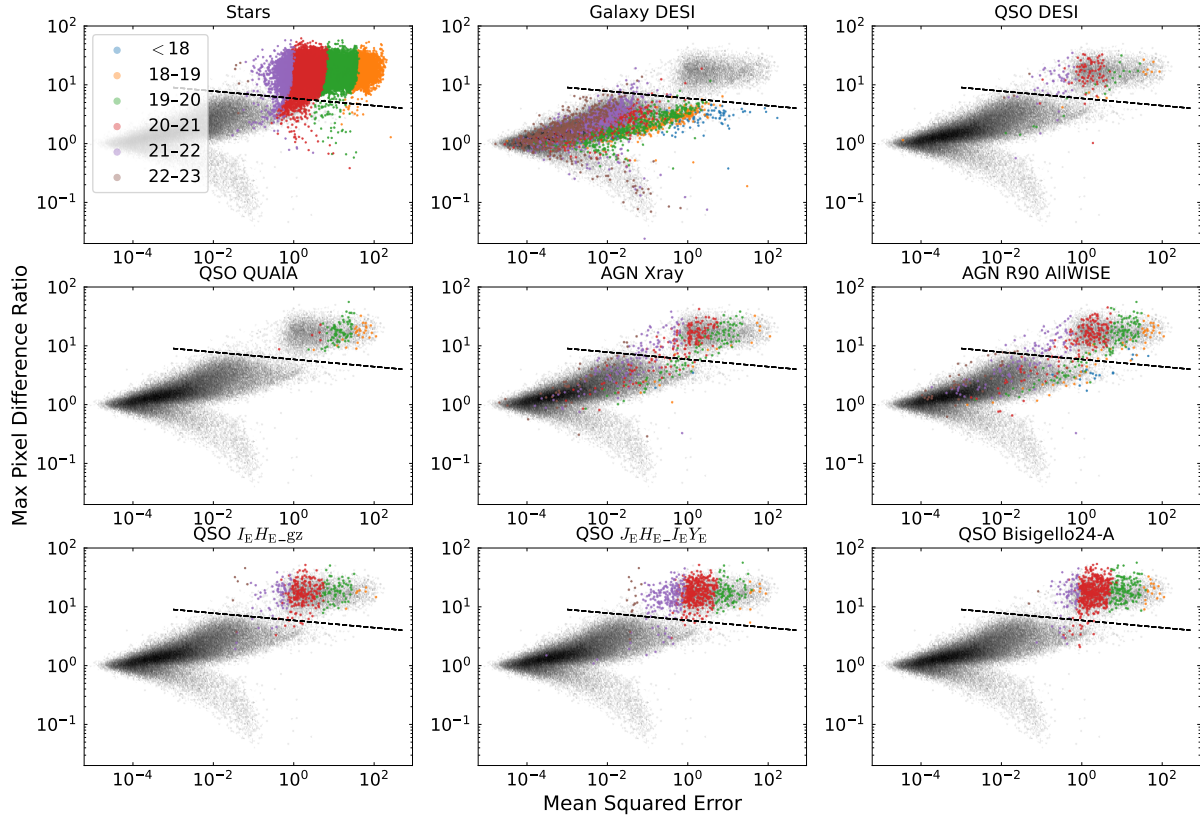


Fig. 9. Threshold produced between the top right cluster of points and the largely galaxy-dominated cluster. The coloured dots represent the VIS I_E magnitude for each positive candidate for the respective selections. The grey dots show the values of all the images in the dataset. The majority of the sources within each non-galaxy class are captured with this simple linear boundary, allowing for high recall across all classes. The sources not captured are typically those of a higher VIS magnitude, and therefore fainter. The tail formed, where sources exhibit a maximum pixel difference ratio of $<10^0$, shows outlier sources with brighter predictions after inpainting.

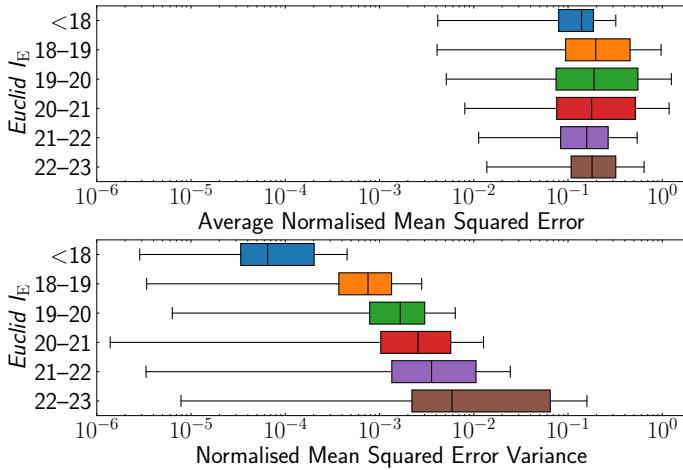


Fig. 10. Variations in the output error from repeat samples of the same fixed pixels. Over seven separate inpainting runs, the model shows a consistent median error when normalised by each respective magnitude (top panel). The variance in the output errors is less consistent with the brighter, lower magnitudes, which provide more robust inpainting than those of fainter images.

morphological parameters sourced directly from the *Euclid* MER Q1 catalogue (Euclid Collaboration: Romelli et al. 2026). Specifically, we used the provided `flux_vis_sersic_euclid` and `flux_vis_psf_euclid` measurements. These parameters were derived by the MER pipeline using the SourceExtractor++

software (Bertin et al. 2020; Kümmel et al. 2020). For each detected source, the pipeline performs two separate and independent model fits. The `flux_vis_psf_euclid` value represents the total flux obtained by fitting a model of the instrumental PSF to the source image. The `flux_vis_sersic_euclid` value represents the total flux from fitting a two-dimensional Sérsic profile, which is designed to model the extended light distribution of a galaxy. The underlying assumption for a classifier based on these parameters is that sources with a dominant, unresolved central component (such as a QSO or a Type-1 AGN) will be described well by a PSF model, resulting in a high `flux_vis_psf_euclid` value. Conversely, normal resolved galaxies will be better described by the extended Sérsic profile, leading to a high `flux_vis_sersic_euclid` value. By comparing the relative strength of these two flux measurements (see Fig. 15), we can construct a simple linear classifier to distinguish between point-like and extended sources. This baseline method is computationally inexpensive as it relies on pre-computed catalogue values, but it is fundamentally limited to this simple morphological distinction and may struggle with identifying AGN in complex host galaxies or at lower relative luminosities.

The distribution of sources with these parameters is shown in Fig. 11. The equation of this linear boundary is $y = x^{1.13}$.

5.6.2. ROC metric

Our second comparison is with the original ROC metric used in Sect. 4.3. As a clear distinction between the galaxy class and

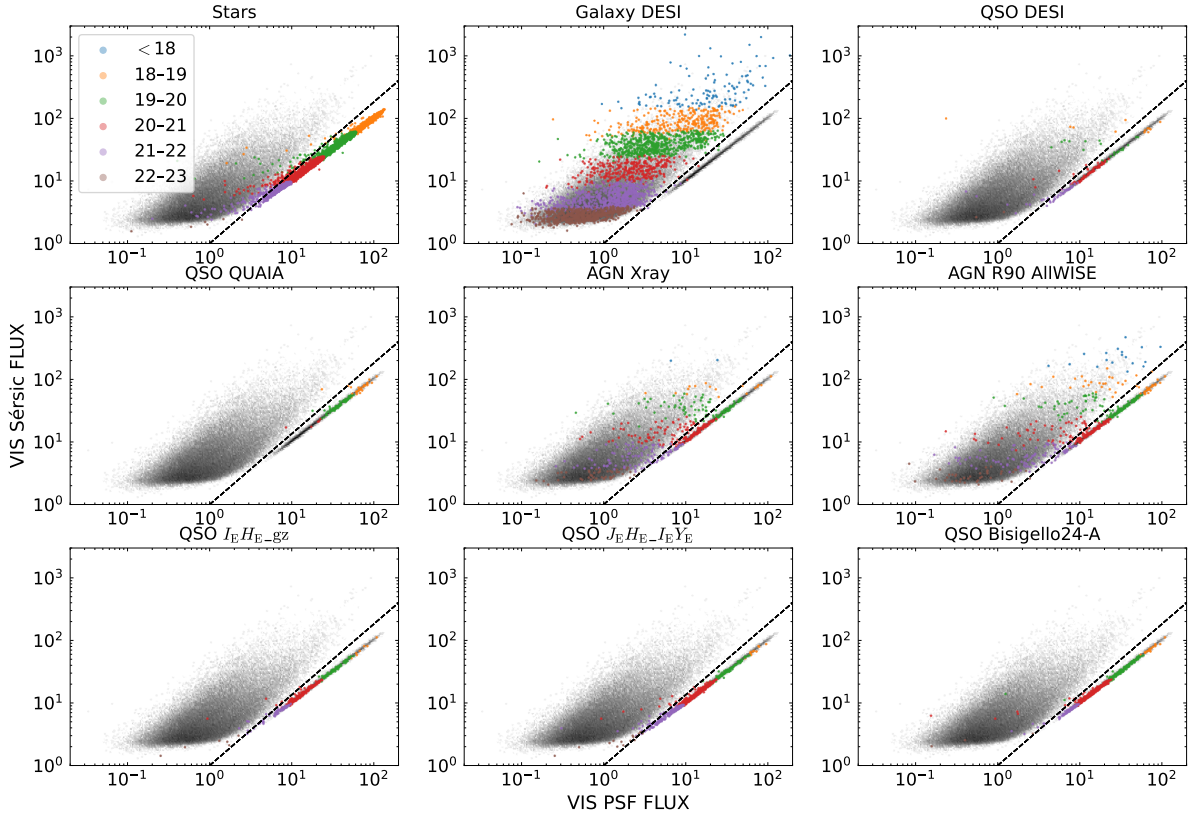


Fig. 11. Comparing the PSF total flux with the Sérsic flux provides a clear separation between the galaxy class and the majority of the non-galaxy selection. This allows for a simple linear boundary for classifying potential AGN, with the majority being point-like objects. However, it is most apparent with selections from AllWISE R90 and X-ray-selected AGN that many sources would not be selected using these features. The coloured dots represent the VIS I_E magnitude for each positive candidate for the respective selections. The grey dots show the scores of all the images in the dataset.

the others is visible, it is important to verify how such a metric performs as a classifier. This is especially true given that such a metric can be generated from the raw image, and no machine-learning model or precomputed values are required. The boundary for this classifier is $x = 0.75$.

5.6.3. Performance against other candidates

Given the variety and differences in the selection methods, there is an inherent incompleteness and potential for contamination. Therefore, our analysis focuses primarily on achieving sufficiently high recall. For an initial candidate selection method, such as the one proposed here, the primary goal is often to compile a large, inclusive sample for potential follow-up observations or more detailed multi-wavelength analysis. In this context, high recall is typically prioritised over high precision, which measures the purity of the selected sample. The scientific cost of missing a potentially interesting source (a false negative) is often considered higher than the operational cost of including a contaminant (a false positive) that can be filtered out in subsequent analysis stages. Therefore, our primary goal is to demonstrate the capability of our model to achieve high recall across a wide variety of AGN and QSO candidates which have been selected via different, more conventional methods and physical mechanisms.

Precision and recall are defined as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (13)$$

For a given classification task (e.g. identifying DESI QSO objects):

- TP (true positives) is the number of objects that belong to the class and are correctly predicted as belonging to it by the model.
- FP (false positives) is the number of objects that do not belong to the class but are incorrectly predicted as belonging to it by the model.
- FN (false negatives) is the number of objects that belong to the class but are not predicted as belonging to it by the model.

The test set scores for each respective selection are shown in Fig. 12, with the collated results for stars, galaxies, AGN, and QSOs shown in Fig. F.1. All predictions are made from images that were neither used during training nor used in the creation of any of the decision boundaries for the classifiers. Due to differing constraints for each selection, we split the analysis to allow for a more representative comparison.

5.6.4. Galaxies

Our galaxy selection was determined by the DESI spectral type classification `SPECTYPE=GALAXY` (DESI Collaboration 2024), with AGN contaminants removed following the procedure detailed in Euclid Collaboration: Matamoro Zatarain et al. (2026), hereafter called MZ26. Negative instances were formed from the remaining spectroscopically determined labels `SPECTYPE=STAR` and `SPECTYPE=QSO` (DESI Collaboration 2024) along with the contaminants removed from the original `SPECTYPE=GALAXY` selections.

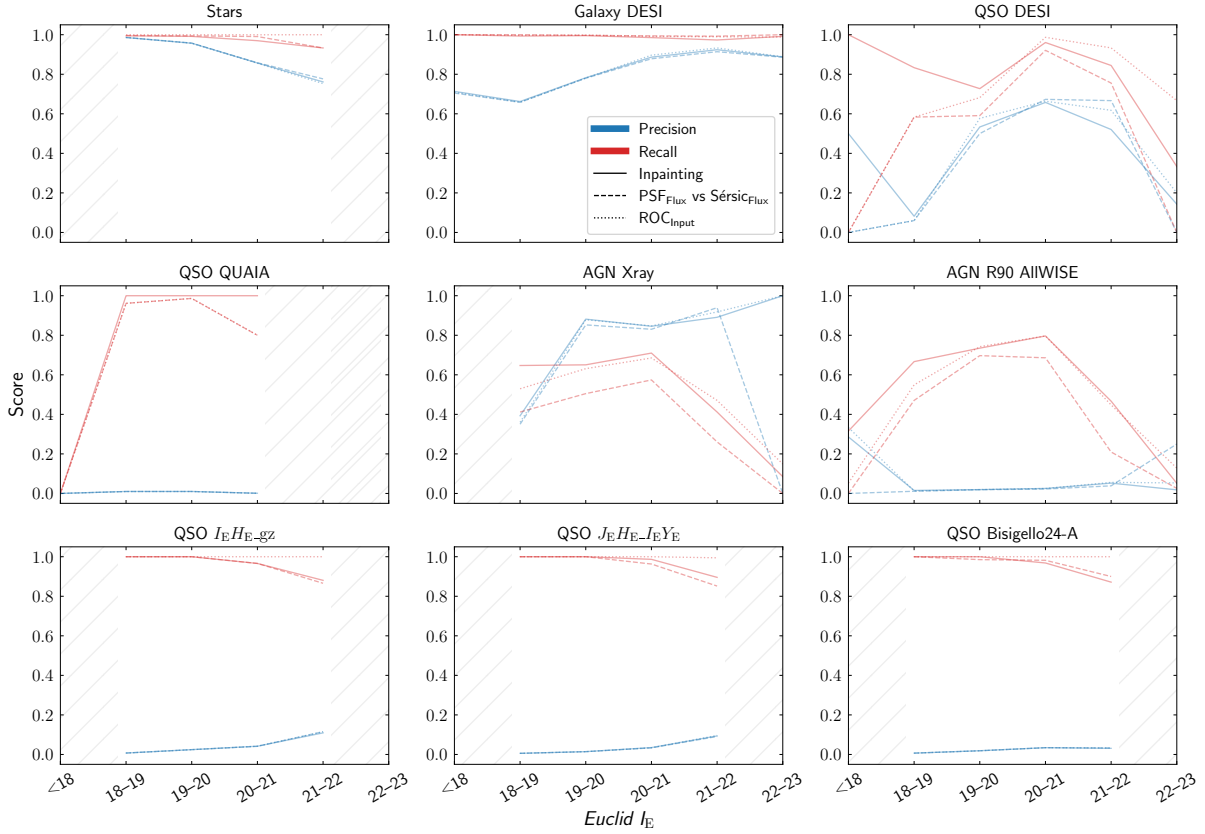


Fig. 12. Precision and recall scores for the diffusion-based AGN predictions (solid line). As the various selections we compare against are not complete, sufficiently high recall allows us to see which traditional methods our selection overlaps with, indicating the types of sources our method is most appropriate for.

The predictions from each classifier for whether a source is a normal galaxy are simply the inverse of the boundary’s original predictions. As Fig. 12 shows, all classifiers capture nearly all galaxies across every magnitude bin. This behaviour is expected, as the selection criteria used by each classifier are explicitly designed to favour AGN-like features, thereby excluding typical galaxies by construction. The precision scores also match extremely closely between classifiers, indicating that similar contaminants prove difficult, especially for the brighter sources. Such contaminants can be seen in Figs. 9 and 11, where stars and DESI-selected QSO fall outside the boundaries.

The ability to separate normal galaxies from other extragalactic or stellar objects benefits both groups of data. Being able to create a pure galaxy sample is vital for the precise measurements needed for cosmology, as well as to further our understanding of dark matter and dark energy, a primary object of the *Euclid* mission (Euclid Collaboration: Mellier et al. 2025).

5.6.5. Stars

The star selection is made up of the positive instances of *Euclid* selected (star_candidate_euclid, MZ26), DESI-selected (SPECTYPE=STAR, DESI Collaboration 2024), and *Gaia*-selected stars. Due to the assumptions made for this method, the incidental capture of stars within our classifications is an expected outcome. The near-perfect recall shows that our classifier is able to capture stars well. This performance is beneficial when attempting to capture a high-purity galaxy sample, as was discussed in the previous section. However, as other works such as Euclid Collaboration: Matamoro Zatarain et al. (2026)

have shown, there are metrics to mitigate the contamination of stars. As a significant number of stars are likely to remain in our sample, for all subsequent selections we applied the cut `phz_star_prob_phz_class < 0.3` to predictions of all classifiers to minimise the contamination of stellar sources.

5.6.6. Euclid-detected QSOs

The *Euclid* photometric-based selections for QSOs include $I_E H_{-g_z}$ and $J_H I_E Y$, as is described in MZ26, and the Bisigello24-A 2-colour selection (Euclid Collaboration: Bisigello et al. 2024). This subset considers only sources with `MUMAX_MINUS_MAG` ≤ -2.6 , a parameter that indicates point-likeness (Euclid Collaboration: Romelli et al. 2026). As these selections have yet to go through a follow-up validation, they remain candidates and as such should not be seen as the complete ground truth.

For the $I_E H_{-g_z}$ and $J_H I_E Y$ selections, we capture around 90% of the candidate QSOs. This coverage remains consistent across the available magnitudes. The performance compared to the other classifiers is similar, with ROC capturing a small percentage more sources, but not different enough to be significant.

The two-colour selection proved harder to match for all classifiers, with recall dropping to 0.6 for the faintest objects. Again, no significant differences are seen between the classifiers.

5.6.7. DESI-selected QSOs

As these sources are assumed to be point-like, applying the above `MUMAX_MINUS_MAG` limit would be reasonable. However,

due to the improvements in spatial resolution from DESI, at about $1''$, to *Euclid*'s now $0''.2$, some sources that appeared point-like through DESI now appear extended. Therefore, to ensure all potential sources were found, we did not apply the restriction to this subset.

This selection provides the first evidence of significant differences in the suggested candidates for each classifier. The inpainting classifier is able to capture a significantly higher selection of lower-magnitude sources compared to the other classifiers. Throughout all magnitude bands, the PSF and Sérsic flux classifier has worse recall. The ROC classifier matched more candidates in the fainter sources.

5.6.8. *Gaia*-detected QSOs

For the *Gaia*-detected QSOs, we used the Quaia selection (Storey-Fisher et al. 2024). The subset consists of positive instances (qso_quaia, MZ26), with negative instances from *Gaia*-detected galaxies (in_galaxy_candidates_gaia, MZ26) and stars (star_candidate_gaia, MZ26). We again applied the $\text{MUMAX_MINUS_MAG} \leq -2.6$ restriction to the sources.

After the initial magnitude bin, our classifier matched with a high proportion of the Quaia-selected QSOs. The other two classifiers achieved equal recall scores, but in the faintest detected sources our classifier was able to match with an additional 20% of sources.

The particularly poor precision for this selection across all the classifiers is due to the considerable imbalance between positive and negative predictions. With only 0.5% of the *Gaia*-detected sources being Quaia candidates, selecting more than just the candidates is very likely, significantly impacting the precision.

5.6.9. AllWISE R90 AGN

Utilising WISE-detected sources allows us to verify our ability to select candidates that may have been obscured by dust. Using the selection criteria derived by Assef et al. (2018), we compared our candidates to the 90% reliability (R90) selection, a high-purity AGN sample.

The Assef et al. (2018) selection provides a more significant test for the classifiers due to no restrictions on point-like sources, as well as fewer contaminants that may have unintentionally improved the overlap of candidates. Similarly to the DESI-selected QSOs, our classifier shows a higher coverage than the other decision boundaries on the brightest sources. In the fainter sources, the diffusion and ROC classifiers are matched. Given the purity of the R90 selection, a large sample of suspected AGN remained unmatched by any of the classifiers, with the most matched magnitude bin achieving 80%.

Much like the *Gaia*-detected QSOs, the small fraction of candidates compared to the full sample of possible sources leads to many of the classifiers selecting sources outside the R90 selection. With a completeness of only 17% (Assef et al. 2018) there will be many mid-infrared AGN in the negative sample, which the classifiers are likely finding.

5.6.10. X-ray-selected AGN

We compared our candidates to the X-ray selection from the work by Euclid Collaboration: Roster et al. (2026). The positive instances are when the calculated probability that the source is a galaxy is sufficiently low ($\text{Gal_proba_roster} < 0.2$, MZ26). The negative instances are sources with $\text{Gal_proba_roster} \geq$

0.2. Only sources with detection in X-rays are included in the results.

The main difference with the results of the X-ray-detected AGN and the other selections is that the precision outperforms the recall across nearly every magnitude bin. This is due to the low number of X-ray sources available. The recall performance between the diffusion model and the ROC classifier is very similar, with the PSF and Sérsic flux classifier again struggling to match as many candidates as the other two.

Each of the selection criteria we compare captures distinct subsets of the AGN and QSO populations, influenced by specific observational biases and methodological constraints. Therefore, it is only when collating these selections and identifying trends and overlaps in the labels that we can be confident that we sufficiently cover the spectrum of possible AGN.

Table 3 summarises the scores shown in Fig. 12. We combine the results into two I_E magnitude bins 18–20 and 20–22.5, separating into brighter and fainter sources, respectively. Due to similar scores in many of the selections, we highlight in bold any scores that outperform the other classifiers by at least 0.05.

The results show that the diffusion method achieves high recall scores across all tested selections, most notably outperforming the other classifiers in the bright QSO DESI, faint QSO QUAIA, and bright R90 selections. In other selections, it performs comparably to the ROC classifier, with both classifiers often significantly outperforming the PSF_{FLUX} versus $\text{Sérsic}_{\text{FLUX}}$ classifier.

5.7. Decomposition of the AGN component

Given our model's ability to classify suspected AGN, it is not unreasonable for the reader to assume that the residuals of the repainting process and the output image can independently represent the AGN and galaxy components. Although utilising inpainting for decomposition is a worthwhile future research direction, the validity of the residuals in accurately representing the AGN component for further scientific analysis has not been tested and, as such, cannot be used. A more significant exploration of which mask to use will also be vital for such a use case. A mask that adapts its size and shape for each input image will likely be necessary. For a deep-learning approach applying AGN decomposition to *Euclid* data, readers are referred to Euclid Collaboration: Margalef-Bentabol et al. (2026), hereafter called MB26, whose method provides AGN fraction (F_{AGN}) measurements that we use to analyse our AGN candidates.

The subset of sources from the Q1 catalogue (Euclid Collaboration: Romelli et al. 2026) used within this work and the subset tested in MB26 feature an overlap of 425 781 sources. The confusion matrix for AGN predictions between the diffusion classifier and sources with an $F_{\text{AGN}} \geq 0.2$ (AGN candidate for MB26) are shown in Table 4. The majority of sources ($\sim 95\%$) are agreed by both methods not to be AGN candidates. For the remaining sources, only $\sim 10\%$ are agreed upon, with over five times more AGN candidates selected by MB26. When comparing all the AGN candidates selected using their respective subsets of data, this number reduces down to 3.5 times more candidates selected by MB26 than the diffusion model (Table 5).

The left panel of Fig. 13 shows the AGN candidates selected by the diffusion classifier but predicted to have a low AGN fraction by the model in MB26. The right panel shows the AGN fraction of sources that are agreed upon between the two methods. Fig. F.2 shows where the MB26 selections are located on the inpainting metrics. As is shown in Table 4, many of the

Table 3. Recall, precision, and area under the curve (AUC) scores for each classifier.

Selection	$18 < I_E < 22.5$	Recall			Precision			AUC		
		Ours	PSF-S	ROC	Ours	PSF-S	ROC	Ours	PSF-S	ROC
Stars	<20	0.99	0.99	1.00	0.96	0.97	0.96	0.50	0.51	0.50
	≥ 20	0.97	0.98	1.00	0.84	0.85	0.84	0.51	0.53	0.50
Galaxy DESI	<20	1.00	1.00	1.00	0.74	0.73	0.73	0.64	0.63	0.64
	≥ 20	0.98	1.00	0.99	0.90	0.90	0.91	0.73	0.72	0.75
QSO DESI	<20	0.77	0.57	0.63	0.18	0.14	0.15	0.84	0.75	0.77
	≥ 20	0.90	0.84	0.96	0.59	0.67	0.62	0.93	0.91	0.96
QSO QUAIA	<20	1.00	0.98	0.98	0.01	0.01	0.01	0.59	0.58	0.58
	≥ 20	1.00	0.80	0.80	<0.01	<0.01	<0.01	0.51	0.41	0.40
AGN X-ray	<20	0.64	0.48	0.61	0.74	0.73	0.74	0.63	0.59	0.62
	≥ 20	0.51	0.38	0.53	0.86	0.86	0.87	0.53	0.51	0.55
AGN R90 AllWISE	<20	0.68	0.57	0.63	0.02	0.02	0.02	0.68	0.63	0.66
	≥ 20	0.59	0.45	0.60	0.03	0.02	0.03	0.75	0.68	0.75
QSO $I_E H_{-gz}$	<20	1.00	1.00	1.00	0.02	0.02	0.02	0.50	0.50	0.50
	≥ 20	0.94	0.92	1.00	0.05	0.05	0.05	0.49	0.47	0.50
QSO $JH_{-I_E Y}$	<20	1.00	1.00	1.00	0.01	0.01	0.01	0.50	0.50	0.50
	≥ 20	0.95	0.91	1.00	0.04	0.04	0.04	0.50	0.46	0.50
QSO Bisigello24-A	<20	1.00	0.99	1.00	0.02	0.02	0.02	0.50	0.50	0.50
	≥ 20	0.96	0.97	1.00	0.03	0.03	0.03	0.50	0.50	0.50

Notes. Each selection is split into brighter and fainter sources, $I_E < 20$ and $I_E \geq 20$, respectively. Bold scores indicate significant improvements over the other classifiers (an increase of at least 0.05). The PSF_{FLUX} versus Sérsic_{FLUX} results have been shortened to PSF-S.

Table 4. AGN candidate comparison between this work and the AGN fractions predicted by MB26.

Matched sources	MB26 AGN	MB26 non-AGN
Diffusion selected	2 263	1 331
Diffusion non-selected	19 758	402 429

Table 5. Total number of AGN candidates from each morphology method and their respective datasets.

Method	Diffusion	MB26
Total AGN candidates	16 053	57 874

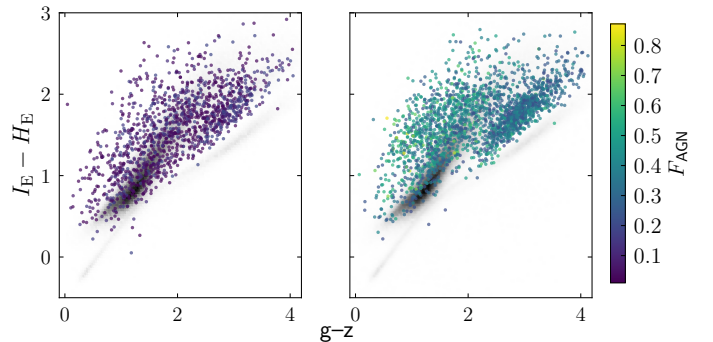
MB26 sources are not located above the diffusion threshold. Of the sources that lie above the threshold, the majority show an $F_{\text{AGN}} > 0.5$. This implies that the F_{AGN} selection is more sensitive to fainter AGN.

Comparing the predictions for each of the *Euclid*-based selections in Fig. 14, we can see that our diffusion model covers areas of the colour space that the photometry-based selections do not. Being able to select both bluer galaxies captured by the photometry selections (MZ26; Euclid Collaboration: Bisigello et al. 2024) and the much redder sources captured by MB26 shows the versatility of the method, highlighting its effectiveness across different galaxy populations.

6. Discussion

6.1. The effect of signal-to-noise

In machine-learning research, the improvements of generative models are often demonstrated with datasets that are of a high resolution and typically noise-free. This provides a level of

**Fig. 13.** Diffusion model AGN candidates, colour-coded according to the AGN fraction of MB26. Left: Not identified as AGN by MB26 ($F_{\text{AGN}} < 0.2$), and right: AGN candidates selected by MB26 ($F_{\text{AGN}} \geq 0.2$).

consistency across the full dataset, ensuring the benchmarks reflect the model's performance under ideal conditions. Whether through loss functions (as is discussed in Sect. 2.3) or in the practical application of the pipelines, the model's response to one part of the data will be similar to another.

However, the difficulty in learning the dynamics of our data is not consistent throughout the images. Higher S/Ns and lower magnitudes provide higher contrast between the source features and the background noise, enabling the model to discern these features more accurately and faster than the low-S/N counterparts. In the context of diffusion models, where the aim throughout training is to learn the dynamics of the introduced noise at different timesteps, the varying S/N of the images could again cause inconsistencies. This can be seen in both the classifier results shown in Fig. 12, where the fainter objects typically have lower recall, as well as the increased variance during inpainting of fainter images shown in Fig. 10.

To explore how the noise schedule impacts the *Euclid* images, Fig. 15 shows the variability of pixel values as they

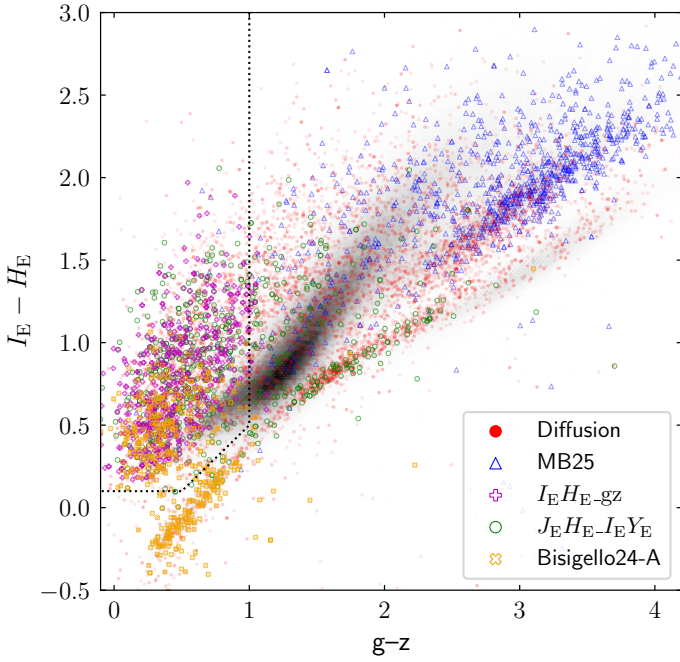


Fig. 14. Comparison of our diffusion model selections with those from photometry-based selections (MZ26; Euclid Collaboration: Bisigello et al. 2024) and the selections from MB26. Although all the diffusion selected sources are plotted, to improve clarity, only a representative subset of the other selections is shown. The photometry-based selections are limited to much bluer sources, highlighted by the dotted line showing the cut introduced in MZ26.

progress through timesteps. We take a row and column of pixels (a slice) through each image, bin them by S/N, and compute the median pixel value at each index within the slice. This allows us to observe how the values change across the image. Since pixels affected by the source have much higher values than those near the edges of the image, Fig. 15 compares slices through the outer regions (top) with slices through the centre of the image (bottom).

In the top row, we see that low-S/N images are more significantly affected by the introduction of noise, and this effect appears earlier in the scheduler. The background noise converges to the Gaussian distribution across all S/N bins between halfway ($t = 2000$) and three quarters ($t = 3000$) of the total noise steps. In contrast, for pixels dominated by the source (bottom row), high-S/N images remain relatively unaffected through much of the schedule. The brightest images do not show noticeable changes until around $t = 2000$, and only converge to the Gaussian distribution within the final 250 steps. It is evident that both the image’s S/N level and the position of the pixels within the image determine the number of timesteps needed for the noise to resemble the target Gaussian distribution.

To visualise the true effect of this differing behaviour, Fig. 16 shows a sample of images throughout different timesteps. Each image shown is a sample from the respective S/N bins below. It is clear that the noise scheduling does not perform uniformly across the images.

To improve the performance of a diffusion model, whether to better recreate a dataset or to model more complex data, the simplest approach is to increase the number of unique denoising steps. The emphasis here is on uniqueness, as simply adding more timesteps can lead to redundancy, especially in the later stages when successive steps become too similar,

add little value, and increase overhead. This is precisely why Nichol & Dhariwal (2021) introduced the cosine-beta scheduler (discussed in Sect. 2.4); the original linear scheduler caused images to degrade into pure noise too early into the process.

The images in Fig. 16 show that for a large fraction of our dataset, the model spends the majority of its time and computational effort learning the trivial task of denoising pure Gaussian noise, which provides no useful learning signal for galaxy morphology. This is a fundamental inefficiency in the implementation of diffusion-based models that not only affects *Euclid* images, but will apply to any scientific dataset that exhibits varying levels of S/N. Although this has implications in regards to the efficiency and optimisation of the pipeline, it highlights the bigger issue of the data quality and variety assumptions implied when using off-the-shelf deep learning methods to the complex, heterogeneous data typical in large astronomical surveys and the wider scientific community. While a detailed implementation is beyond the scope of this work, this finding highlights a critical direction for future research: the development of adaptive noise schedulers that are conditioned on image properties such as S/N. Such an advance would significantly improve both the efficiency and performance of diffusion models in scientific applications.

6.2. Considerations for training and inference costs

Our DDPM inpainting model required approximately 50 hours of training using a single Nvidia A6000 GPU, which aligns with expectations for models of a similar scale and architecture. However, the inference pipeline is heavily constrained by memory rather than computation, as is demonstrated by its near-linear scaling – the inpainting of 1024 and 8192 images took roughly the same amount of time when sufficient GPU memory was available. This memory constraint is due to the need to store and iteratively manipulate large intermediate tensors during inference. Training, although more computationally expensive, has a more structured process whereby gradients are computed and accumulated in a predictable manner.

With expanded computational resources ($4 \times$ Nvidia H100 GPUs), we can inpaint just over 2.5 million images per day. As we have shown in Sects. 2.4 and 3, our model architecture and hyperparameters remain largely unmodified from their standard configurations. Optimisations tailored explicitly for astronomical-based data, especially in regards to the noise scheduler discussed in Sect. 6.1, will significantly reduce the number of iterations and inference time required to process each image. For example, Fig. 16 shows an opportunity for a two- to four-fold speed-up in timesteps when correcting for the high number of pure Gaussian noise images alone.

The results from Sect. 5.6 show that although our predictions have a significant overlap with many of the selections presented, it also highlights that the ROC classifier produces a similar overlap performance. Given that this boundary is formed by a simple calculation from the data, rather than requiring a complex model and therefore near-zero overhead to calculate, it may present itself as the preferred method. However, the ROC metric is static and its performance is fixed by its mathematical definition. It must be reiterated that the current pipeline for the diffusion model has had minimum optimisations for the data and the classification task and so can be expected to improve with larger training sets, more advanced network architectures, or the future incorporation of multi-wavelength data as conditioning information. The ROC classifier is also limited to outputting a single number. The generative capabilities of the diffusion model,

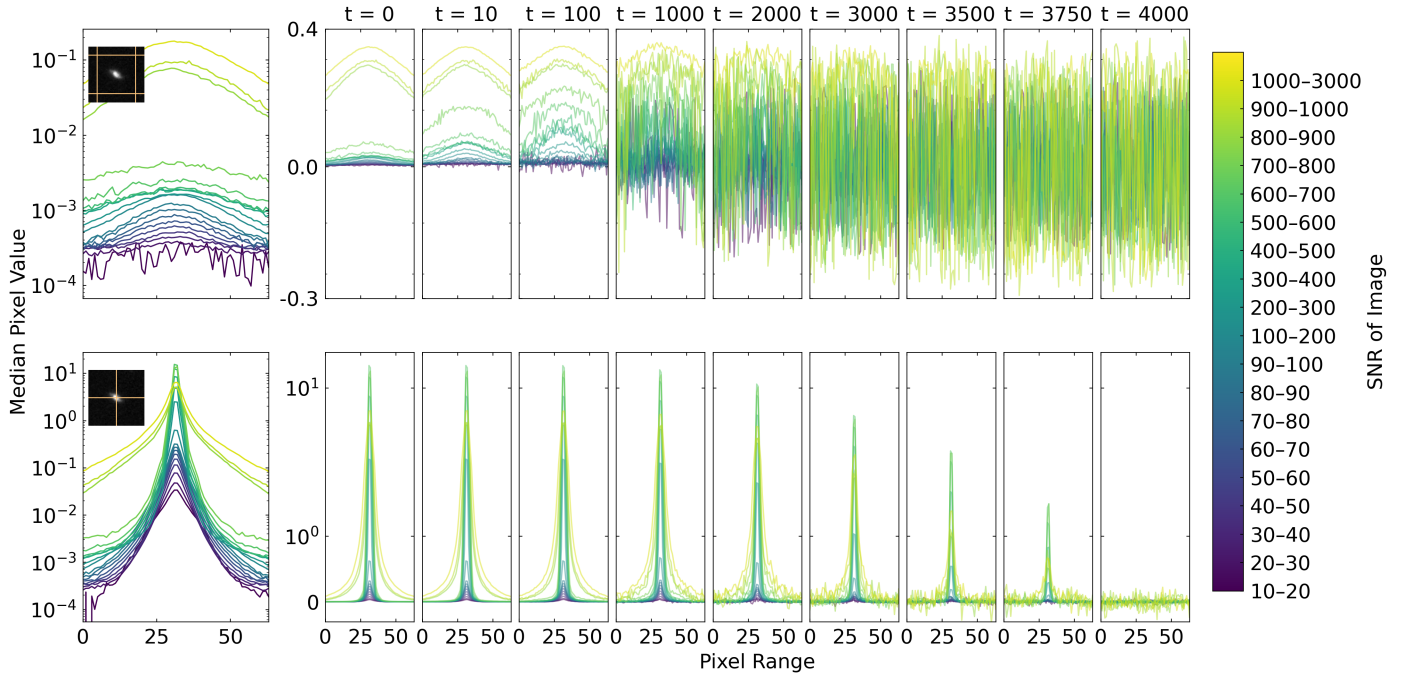


Fig. 15. Impact of the diffusion noise scheduler on the median pixel values of the VIS images. The pixels covered in the median pixel calculation are shown as orange regions overlaid on the example source image in the leftmost plots. The top row shows how the additional noise impacts areas of the image that are more typically background noise, whereas the bottom row focuses on pixels that will be heavily influenced by the source. The binning of images according to their S/N shows how different levels of image clarity are affected by the gradual increase in introduced noise. Images with higher S/N can retain more of the source detail for more timesteps in the scheduler.

where it outputs a reconstructed image, provides the potential for a much richer set of scientific applications beyond simple classification such as decomposition as was discussed in Sect. 5.7. Therefore, although it has shown good performance, its limited approach restricts the ROC classifier to only serve as a fundamental baseline for comparison in our study.

In regards to the PSF and Sérsic flux classifier, although there are no computational costs to the user, traditional parametric fitting is a serial, per-object process that becomes a bottleneck for survey-scale datasets (Tuccillo et al. 2018). Having parallelisable and scalable alternative methods, even if they feature a large one-time training cost, allows us to make a strategic investment as we cover more of the extragalactic sky. For AGN identification specifically, the simple parametric models may not be well suited to handle complex or merged morphologies. This could explain the worse AGN-based recall seen for this classifier in Fig. 12.

Future enhancements to the diffusion model are expected to not only improve the accuracy of identifying AGN candidates but also refine our reconstruction error maps for better analysis of prospective AGN components. The other classifiers do not offer the same potential for improvement, making the diffusion method the most promising moving forward.

7. Conclusion

In this paper, we have used a diffusion-based inpainting model to identify AGN and QSOs using VIS images from the Q1 data (Euclid Quick Release Q1 2025) from the *Euclid* telescope. Our approach focuses on inpainting the brightest pixels within the centre of a galaxy and employs a novel thresholding approach based on the reconstruction errors to differentiate AGN and QSOs from the rest of the galaxy population. Using only VIS images, our method can generate a large, reliable sample of opti-

cally selected AGN without requiring AGN labels ahead of time. Utilising a standard training and inference pipeline with minimal modifications for astronomical data, our method demonstrates high recall and significant overlap with AGN candidates obtained through more traditional methods, across various wavelengths including optical, near-infrared (NIR), mid-infrared, and X-ray.

Our key contributions and conclusions include the following:

- Compared to traditional colour-selection criteria, the approach attains high coverage in AGN candidates while achieving near-perfect separation of galaxies not featuring an AGN component. Reliably separating these different sources allows for the creation of subsets with high purity, which are vital for the cosmological objectives of the *Euclid* mission.
- By demonstrating that VIS images alone can generate reliable and extensive samples of AGN and QSOs, our method showcases the impressive capabilities of *Euclid*'s imaging. Given that the future of the mission will cover $14\,000\text{ deg}^2$, the ability for AGN identification using only VIS allows for the possibility of some science aims to be met without the immediate need to integrate multiple data sources. This can significantly reduce the time and complexity of processing and analysing the data.
- The application of diffusion models with minimal adaptations for astronomical imaging highlights the potential for the use of more advanced machine-learning tools and pipelines for future *Euclid* data, as well as other scientific datasets. The ability to leverage models that can process large volumes of data efficiently, without costly and extensive specialised tuning, will be vital as scientific datasets continue to increase in size and complexity.
- Although our model performs well without significant optimisation for our use case, we provide insights into the behaviour of the model given the challenges presented

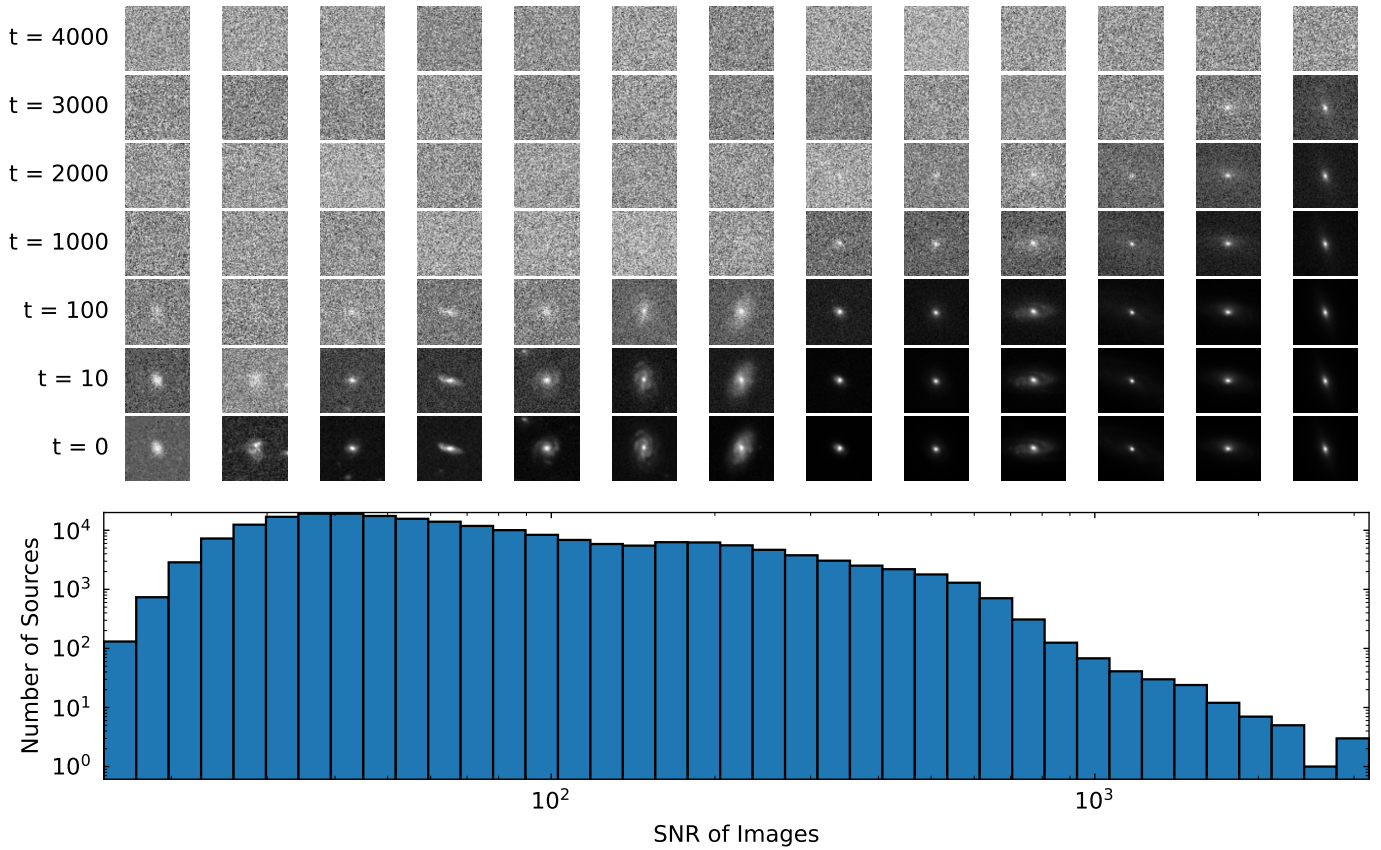


Fig. 16. *Top:* Respective noised images produced by the cosine-beta schedule at different timesteps. Each image is a sample from the respective S/N bin directly below it. Due to the scales of pixel values, the introduced noise has a more significant impact on the typically fainter, low-S/N images, leading to the images converging to Gaussian noise much sooner into the forward process. The relationship between the S/N and the rate of convergence results in the entirety of the top left of the grid of images being pure noise, indicating inefficient training for lower-S/N images. This highlights the difficulty in applying off-the-shelf pipelines to the complexities of real-world astronomical data that feature a high dynamic range and varying quality over images. *Bottom:* Distribution of S/N of galaxy images. Even though the sample is dominated by lower-S/N images, a non-negligible number of sources with S/N > 1000 remains in the training set.

with astronomical imaging. We show how the assumptions of machine-learning literature, particularly regarding fixed noise schedules, can result in non-optimal training on data with a high dynamic range and large variations in S/N. By highlighting these inefficiencies, we identify a clear direction for future progress in both astronomy and machine-learning domains through the development of adaptive, data-aware training strategies.

- Because this is the first use of these pipelines for such an application, we have routinely provided guidance on improvements and next steps to refine this work. Enhancing the model’s precision, improving the efficiency of both training and inference and expanding its capabilities will only increase its impact.

In conclusion, our results demonstrate the feasibility and effectiveness of using diffusion-based inpainting for AGN and QSO identification. The ability to use VIS images exclusively for selecting these candidates provides a potential strategic flexibility to the *Euclid* mission, while reducing reliance on complex multi-spectral data integration. By proving the utility of applying our pipelines to the task of AGN identification, we believe it can be expanded and can significantly contribute to a broader range of astronomical applications, such as identifying gravitational lensing or finding transient objects.

Acknowledgements. The authors would like to thank the referee for their very insightful and constructive feedback during the review process. This work has made use of the *Euclid* Quick Release Q1 data from the *Euclid* mission of the European Space Agency (ESA), 2025, <https://doi.org/10.57780/esa-2853f3b>. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of *Euclid*, in particular the Agenzia Spaziale Italiana, the Austrian Forschungsförderungsgesellschaft funded through BMK, the Belgian Science Policy, the Canadian Euclid Consortium, the Deutsches Zentrum für Luft- und Raumfahrt, the DTU Space and the Niels Bohr Institute in Denmark, the French Centre National d’Etudes Spatiales, the Fundação para a Ciência e a Tecnologia, the Hungarian Academy of Sciences, the Ministerio de Ciencia, Innovación y Universidades, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Nederlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Research Council of Finland, the Romanian Space Agency, the State Secretariat for Education, Research, and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the *Euclid* web site (www.euclid-ec.org). Based on data from UNIONS, a scientific collaboration using three Hawaii-based telescopes: CFHT, Pan-STARRS, and Subaru (www.skysurvey.cc). Based on data from the Dark Energy Camera (DECam) on the Blanco 4-m Telescope at CTIO in Chile (<https://www.darkenergysurvey.org>). This work uses results from the ESA mission *Gaia*, whose data are being processed by the Gaia Data Processing and Analysis Consortium (<https://www.cosmos.esa.int/gaia>). This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. DESI construction and operations is managed by the Lawrence Berkeley National Laboratory. This

research is supported by the U.S. Department of Energy, Office of Science, Office of High-Energy Physics, under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain, and by the DESI Member Institutions. The DESI collaboration is honored to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. Grant Stevens acknowledges financial support from the UKRI for an EPSRC Doctoral Prize Fellowship at the University of Bristol (EP/W524414/1). This work has benefited from the support of Royal Society Research Grant RGS/R1/231450. This research was supported by the International Space Science Institute (ISSI) in Bern, through ISSI International Team project #23-573 "Active Galactic Nuclei in Next Generation Surveys". "ELSA: Euclid Legacy Science Advanced analysis tools" (Grant Agreement no. 101135203) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Innovate UK. Neither the European Union nor the granting authority can be held responsible for them. UK participation is funded through the UK Horizon guarantee scheme under Innovate UK grant 10093177. The authors acknowledge the use of computational resources from the parallel computing cluster of the Open Physics Hub (<https://site.unibo.it/openphysicshub/en>) at the Physics and Astronomy Department in Bologna. The authors acknowledge the use of resources provided by the Isambard-AI National AI Research Resource (AIRR). Isambard-AI is operated by the University of Bristol and is funded by the UK Government's Department for Science, Innovation and Technology (DSIT) via UK Research and Innovation; and the Science and Technology Facilities Council [ST/AIRR/I-A-I/1023].

References

- Adam, A., Coogan, A., Malkin, N., et al. 2022, ArXiv e-prints [arXiv:2211.03812]
- Adam, A., Stone, C., Bottrell, C., et al. 2025, *AJ*, **169**, 254
- Assef, R. J., Stern, D., Noirot, G., et al. 2018, *ApJS*, **234**, 23
- Bertin, E., Schefer, M., Apostolakis, N., et al. 2020, in *Astronomical Data Analysis Software and Systems XXIX*, eds. R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, & H. Verkoeter, *ASP Conf. Ser.*, **527**, 461
- Davis, B. L., Graham, A. W., & Cameron, E. 2018, *ApJ*, **869**, 113
- Davis, B. L., Graham, A. W., & Cameron, E. 2019, *ApJ*, **873**, 85
- DESI Collaboration (Adame, A. G., et al.) 2024, *AJ*, **168**, 58
- Dhariwal, P., & Nichol, A. 2021, *Adv. Neural Inf. Process. Syst.*, **34**, 8780
- Dia, N., Yantovski-Barth, M., Adam, A., et al. 2025, *ApJ*, submitted [arXiv:2501.02473]
- Dressler, A. 1989, in *Active Galactic Nuclei*, eds. D. E. Osterbrock, & J. S. Miller, *IAU Symp.*, **134**, 217
- Euclid Collaboration (Aussel, B., et al.) 2024, *A&A*, **689**, A274
- Euclid Collaboration (Bisigello, L., et al.) 2024, *A&A*, **691**, A1
- Euclid Collaboration (Cropper, M., et al.) 2025, *A&A*, **697**, A2
- Euclid Collaboration (Mellier, Y., et al.) 2025, *A&A*, **697**, A1
- Euclid Collaboration (Margalef-Bentabol, B., et al.) 2026, *A&A*, **711**, A18 (Euclid Q1 SI)
- Euclid Collaboration (Matamoro Zatarain, T., et al.) 2026, *A&A*, **711**, A20 (Euclid Q1 SI)
- Euclid Collaboration (Romelli, E., et al.) 2026, *A&A*, **711**, A4 (Euclid Q1 SI)
- Euclid Collaboration (Roster, W., et al.) 2026, *A&A*, **711**, A16 (Euclid Q1 SI)
- Euclid Collaboration (Walmsley, M., et al.) 2026, *A&A*, **711**, A9 (Euclid Q1 SI)
- Euclid Quick Release Q1. 2025, <https://doi.org/10.57780/esa-2853f3b>
- Feng, B. T., Smith, J., Rubinstein, M., et al. 2023, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10520
- Ferrarese, L., & Merritt, D. 2000, *ApJ*, **539**, L9
- Gaia Collaboration (Vallenari, A., et al.) 2023a, *A&A*, **674**, A1
- Gaia Collaboration (Bailer-Jones, C. A. L., et al.) 2023b, *A&A*, **674**, A41
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, *Adv. Neural Inf. Process. Syst.*, **27**
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2020, *Commun. ACM*, **63**, 139
- Graikos, A., Malkin, N., Jovic, N., & Samaras, D. 2022, *Adv. Neural Inf. Process. Syst.*, **35**, 14715
- Harrison, C. M., & Ramos Almeida, C. 2024, *Galaxies*, **12**, 17
- Ho, J., Jain, A., & Abbeel, P. 2020, *Adv. Neural Inf. Process. Syst.*, **33**, 6840
- Kingma, D. P., & Welling, M. 2013, ArXiv e-prints [arXiv:1312.6114]
- Kormendy, J., & Richstone, D. 1995, *ARA&A*, **33**, 581
- Kümmel, M., Bertin, E., Schefer, M., et al. 2020, in *Astronomical Data Analysis Software and Systems XXIX*, eds. R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, & H. Verkoeter, *ASP Conf. Ser.*, **527**, 29
- Lugmayr, A., Danelljan, M., Romero, A., et al. 2022, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (IEEE)*, 11461
- Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, *PASP*, **116**, 133
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, *AJ*, **115**, 2285
- Nguyen, A., Clune, J., Bengio, Y., Dosovitskiy, A., & Yosinski, J. 2017, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4467
- Nichol, A. Q., & Dhariwal, P. 2021, in *Proceedings of the 38th International Conference on Machine Learning*, eds. M. Meila, & T. Zhang (PMLR), *Proc. Mach. Learn. Res.*, **139**, 8162
- Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. 2016, *IEEE Conference on Computer Vision and Pattern Recognition (IEEE)*, 2536
- Radford, A., Kim, J. W., Hallacy, C., et al. 2021, in *Proceedings of the 38th International Conference on Machine Learning*, eds. M. Meila, & T. Zhang, (PMLR), *Proc. Mach. Learn. Res.*, **139**, 8748
- Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. 2022, arXiv e-prints [arXiv:2204.06125]
- Remy, B., Lanusse, F., Jeffrey, N., et al. 2023, *A&A*, **672**, A51
- Saharia, C., Chan, W., Saxena, S., et al. 2022, in *Advances in Neural Information Processing Systems*, eds. S. Koyejo, S. Mohamed, A. Agarwal, et al. (Curran Associates, Inc.), **35**, 36479
- Sahu, N., Graham, A. W., & Davis, B. L. 2019, *ApJ*, **876**, 155
- Sampson, M. L., Melchior, P., Ward, C., & Birmingham, S. 2024, *Astron. Comput.*, **49**, 100875
- Smith, M. J., Geach, J. E., Jackson, R. A., et al. 2022, *MNRAS*, **511**, 1808
- Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. 2015, *International Conference on Machine Learning (pmlr)*, 2256
- Spagnoletti, A., Boucaud, A., Kabalan, W., Biswas, B., et al. 2024, *38th Conference on Neural Information Processing Systems*
- Stone, C., & Courteau, S. 2019, *ApJ*, **882**, 6
- Stone, C., Courteau, S., & Arora, N. 2021, *ApJ*, **912**, 41
- Storey-Fisher, K., Hogg, D. W., Rix, H.-W., et al. 2024, *ApJ*, **964**, 69
- Thanh-Tung, H., & Tran, T. 2020, *2020 International Joint Conference on Neural Networks (ijcnn)*, 1
- Tuccillo, D., Huertas-Company, M., Decencière, E., et al. 2018, *MNRAS*, **475**, 894
- Walmsley, M., Smith, L., Lintott, C., et al. 2020, *MNRAS*, **491**, 1554
- Wang, Y., Tao, X., Qi, X., Shen, X., & Jia, J. 2018, *Adv. Neural Inf. Process. Syst.*, **31**
- Yeh, R. A., Chen, C., Yian Lim, T., et al. 2017, *IEEE Conference on Computer Vision and Pattern Recognition (IEEE)*, 5485
- Yu, J., Lin, Z., Yang, J., et al. 2018, *IEEE Conference on Computer Vision and Pattern Recognition (IEEE)*, 5505
- Zhang, L., Rao, A., & Agrawala, M. 2023, *IEEE/CVF International Conference on Computer Vision (IEEE)*, 3836

¹ School of Physics, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK

² Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

³ SRON Netherlands Institute for Space Research, Landleven 12, 9747 AD Groningen, The Netherlands

⁴ Instituto de Astrofísica de Canarias, Vía Láctea, 38205 La Laguna, Tenerife, Spain

⁵ Instituto de Astrofísica de Canarias (IAC); Departamento de Astrofísica, Universidad de La Laguna (ULL), 38200 La Laguna, Tenerife, Spain

⁶ Université PSL, Observatoire de Paris, Sorbonne Université, CNRS, LERMA, 75014 Paris, France

⁷ Université Paris-Cité, 5 Rue Thomas Mann, 75013 Paris, France

⁸ School of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

⁹ Aspia Space, Falmouth TR10 9TA, UK

¹⁰ David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario M5S 3H4, Canada

- ¹¹ Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
- ¹² Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
- ¹³ Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans s/n, 08193 Barcelona, Spain
- ¹⁴ Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels, Barcelona, Spain
- ¹⁵ Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal
- ¹⁶ Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal
- ¹⁷ Dipartimento di Fisica e Astronomia “Augusto Righi” – Alma Mater Studiorum Università di Bologna, Via Piero Gobetti 93/2, 40129 Bologna, Italy
- ¹⁸ INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
- ¹⁹ Department of Mathematics and Physics, Roma Tre University, Via della Vasca Navale 84, 00146 Rome, Italy
- ²⁰ INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
- ²¹ Université Paris-Saclay, CNRS, Institut d’astrophysique spatiale, 91405 Orsay, France
- ²² ESAC/ESA, Camino Bajo del Castillo s/n., Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
- ²³ INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
- ²⁴ Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France
- ²⁵ IFPU, Institute for Fundamental Physics of the Universe, Via Beirut 2, 34151 Trieste, Italy
- ²⁶ INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
- ²⁷ INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste, TS, Italy
- ²⁸ SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, TS, Italy
- ²⁹ Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
- ³⁰ INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- ³¹ Space Science Data Center, Italian Space Agency, Via del Politecnico snc, 00133 Roma, Italy
- ³² Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146 Genova, Italy
- ³³ INFN-Sezione di Genova, Via Dodecaneso 33, 16146 Genova, Italy
- ³⁴ Department of Physics “E. Pancini”, University Federico II, Via Cinthia 6, 80126 Napoli, Italy
- ³⁵ INAF-Osservatorio Astronomico di Capodimonte, Via Moiriello 16, 80131 Napoli, Italy
- ³⁶ Faculdade de Ciências da Universidade do Porto, Rua do Campo de Alegre, 4150-007 Porto, Portugal
- ³⁷ Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
- ³⁸ INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
- ³⁹ INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese, (TO), Italy
- ⁴⁰ European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
- ⁴¹ Institute Lorentz, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
- ⁴² Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
- ⁴³ INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
- ⁴⁴ Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
- ⁴⁵ Port d’Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain
- ⁴⁶ INFN section of Naples, Via Cinthia 6, 80126 Napoli, Italy
- ⁴⁷ Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
- ⁴⁸ Dipartimento di Fisica e Astronomia “Augusto Righi” – Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- ⁴⁹ Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
- ⁵⁰ European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
- ⁵¹ Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, Villeurbanne F-69100, France
- ⁵² Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
- ⁵³ Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
- ⁵⁴ Institució Catalana de Recerca i Estudis Avançats (ICREA), Pas-seig de Lluís Companys 23, 08010 Barcelona, Spain
- ⁵⁵ UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 rue Enrico Fermi, 69622 Villeurbanne, France
- ⁵⁶ Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
- ⁵⁷ Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
- ⁵⁸ Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- ⁵⁹ Department of Astronomy, University of Geneva, ch. d’Ecogia 16, 1290 Versoix, Switzerland
- ⁶⁰ INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
- ⁶¹ Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
- ⁶² INAF-Istituto di Astrofisica e Planetologia Spaziali, Via del Fosso del Cavaliere, 100, 00100 Roma, Italy
- ⁶³ Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
- ⁶⁴ INAF-Osservatorio Astronomico di Padova, Via dell’Osservatorio 5, 35122 Padova, Italy
- ⁶⁵ Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029, Blindern 0315, Oslo, Norway
- ⁶⁶ Department of Physics, Lancaster University, Lancaster LA1 4YB, UK
- ⁶⁷ Felix Hormuth Engineering, Goethestr. 17, 69181 Leimen, Germany
- ⁶⁸ Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
- ⁶⁹ Cosmic Dawn Center (DAWN), Denmark
- ⁷⁰ Institut d’Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
- ⁷¹ NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
- ⁷² Department of Physics and Helsinki Institute of Physics, Gustaf Hållströmin katu 2, 00014 University of Helsinki, Helsinki, Finland
- ⁷³ Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
- ⁷⁴ Department of Physics, P.O. Box 64, 00014 University of Helsinki, Helsinki, Finland
- ⁷⁵ Helsinki Institute of Physics, Gustaf Hållströmin katu 2, University of Helsinki, Helsinki, Finland
- ⁷⁶ Centre de Calcul de l’IN2P3/CNRS, 21 avenue Pierre de Coubertin, 69627 Villeurbanne Cedex, France
- ⁷⁷ Laboratoire d’étude de l’Univers et des phénomènes eXtremes, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 92190 Meudon, France
- ⁷⁸ SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK
- ⁷⁹ Dipartimento di Fisica “Aldo Pontremoli”, Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
- ⁸⁰ INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
- ⁸¹ Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
- ⁸² INFN-Sezione di Roma, Piazzale Aldo Moro, 2 – c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy

- ⁸³ Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK
- ⁸⁴ Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
- ⁸⁵ Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
- ⁸⁶ CNRS-UCB International Research Laboratory, Centre Pierre Binetruy, IRL2007, CPB-IN2P3, Berkeley, USA
- ⁸⁷ University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
- ⁸⁸ Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
- ⁸⁹ Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- ⁹⁰ Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- ⁹¹ DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark
- ⁹² Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
- ⁹³ Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
- ⁹⁴ Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
- ⁹⁵ Centre National d'Etudes Spatiales – Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
- ⁹⁶ Institute of Space Science, Str. Atomistilor, nr. 409 Măgurele, Ilfov 077125, Romania
- ⁹⁷ Consejo Superior de Investigaciones Científicas, Calle Serrano 117, 28006 Madrid, Spain
- ⁹⁸ Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain
- ⁹⁹ Dipartimento di Fisica e Astronomia “G. Galilei”, Università di Padova, Via Marzolo 8, 35131 Padova, Italy
- ¹⁰⁰ Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile
- ¹⁰¹ Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
- ¹⁰² Satlantis, University Science Park, Sede Bld 48940, Leioa-Bilbao, Spain
- ¹⁰³ Centre for Electronic Imaging, Open University, Walton Hall, Milton Keynes MK7 6AA, UK
- ¹⁰⁴ Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- ¹⁰⁵ Cosmic Dawn Center (DAWN)
- ¹⁰⁶ Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
- ¹⁰⁷ Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain
- ¹⁰⁸ Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
- ¹⁰⁹ INFN-Bologna, Via Imerio 46, 40126 Bologna, Italy
- ¹¹⁰ Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- ¹¹¹ Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
- ¹¹² Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- ¹¹³ Istituto Nazionale di Fisica Nucleare, Sezione di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- ¹¹⁴ INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
- ¹¹⁵ Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
- ¹¹⁶ INAF – Osservatorio Astronomico di Brera, Via Emilio Bianchi 46, 23807 Merate, Italy
- ¹¹⁷ INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy, and INFN-Sezione di Genova, Via Dodecaneso 33, 16146 Genova, Italy
- ¹¹⁸ Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France
- ¹¹⁹ ICL, Junia, Université Catholique de Lille, LITL, 59000 Lille, France
- ¹²⁰ Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain
- ¹²¹ CERCA/ISO, Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
- ¹²² Technical University of Munich, TUM School of Natural Sciences, Physics Department, James-Frank-Str. 1, 85748 Garching, Germany
- ¹²³ Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
- ¹²⁴ Laboratoire Univers et Théorie, Observatoire de Paris, Université PSL, Université Paris Cité, CNRS, 92190 Meudon, France
- ¹²⁵ Departamento de Física Fundamental. Universidad de Salamanca, Plaza de la Merced s/n, 37008 Salamanca, Spain
- ¹²⁶ Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France
- ¹²⁷ Center for Data-Driven Discovery, Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
- ¹²⁸ Dipartimento di Fisica – Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
- ¹²⁹ ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy
- ¹³⁰ California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA
- ¹³¹ Departamento Física Aplicada, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia, Spain
- ¹³² Instituto de Física de Cantabria, Edificio Juan Jordá, Avenida de los Castros, 39005 Santander, Spain
- ¹³³ INFN, Sezione di Lecce, Via per Arnesano, CP-193, 73100 Lecce, Italy
- ¹³⁴ Department of Mathematics and Physics E. De Giorgi, University of Salento, Via per Arnesano, CP-I93, 73100 Lecce, Italy
- ¹³⁵ INAF-Sezione di Lecce, c/o Dipartimento Matematica e Fisica, Via per Arnesano, 73100 Lecce, Italy
- ¹³⁶ CEA Saclay, DFR/IRFU, Service d'Astrophysique, Bat. 709, 91191 Gif-sur-Yvette, France
- ¹³⁷ Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- ¹³⁸ Instituto de Astrofísica de Canarias, c/ Via Lactea s/n, La Laguna, 38200, Spain. Departamento de Astrofísica de la Universidad de La Laguna, Avda. Francisco Sanchez, La Laguna 38200, Spain
- ¹³⁹ Caltech/IPAC, 1200 E. California Blvd., Pasadena, CA 91125, USA
- ¹⁴⁰ Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
- ¹⁴¹ Department of Physics and Astronomy, Vesilinnantie 5, 20014 University of Turku, Turku, Finland
- ¹⁴² Serco for European Space Agency (ESA), Camino bajo del Castillo s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- ¹⁴³ ARC Centre of Excellence for Dark Matter Particle Physics, Melbourne, Australia
- ¹⁴⁴ Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
- ¹⁴⁵ Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town 7535, South Africa
- ¹⁴⁶ DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

- ¹⁴⁷ Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- ¹⁴⁸ Department of Astrophysics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
- ¹⁴⁹ Department of Physics, Centre for Extragalactic Astronomy, Durham University, South Road, Durham DH1 3LE, UK
- ¹⁵⁰ Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
- ¹⁵¹ IRFU, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette Cedex, France
- ¹⁵² Univ. Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, 53, Avenue des Martyrs, 38000 Grenoble, France
- ¹⁵³ INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy
- ¹⁵⁴ Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy
- ¹⁵⁵ HE Space for European Space Agency (ESA), Camino bajo del Castillo s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- ¹⁵⁶ Theoretical astrophysics, Department of Physics and Astronomy, Uppsala University, Box 515, 751 20 Uppsala, Sweden
- ¹⁵⁷ Mathematical Institute, University of Leiden, Einsteinweg 55, 2333 CA Leiden, The Netherlands
- ¹⁵⁸ School of Physics & Astronomy, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK
- ¹⁵⁹ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- ¹⁶⁰ Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
- ¹⁶¹ Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010 New York, NY, USA

Appendix A: Masks

To explore the ability of the model at recreating galaxy images, we apply a variety of different masks to a selection of unseen images. Figure A.1 shows the original image on the leftmost column, followed by the regenerated images in subsequent columns. With each mask shown on the top row, the white area shows the pixels that are fixed from the original image, whereas the black areas are generated using the diffusion model.

As our model's aim is not to generate large collections of synthetic data, the variety and scientific validity of the produced galaxies are not put under high scrutiny. Instead, we only check that there are no obvious catastrophic mistakes. In the last two columns, we can see that due to the large mask and fewer pixels to guide the reconstruction, some images produce a visible boundary around the image where the combination of fixed and generated pixels is not as seamless as the other images. Given that the masks used for AGN detection are significantly smaller than those presented in Fig. A.1, the fraction of pixels that the inpainting is conditioned on will be much higher, providing tighter constraints on the output compared to these examples. Improvements to the scheduler, as discussed in Sect. 6.1, would also reduce such artefacts.

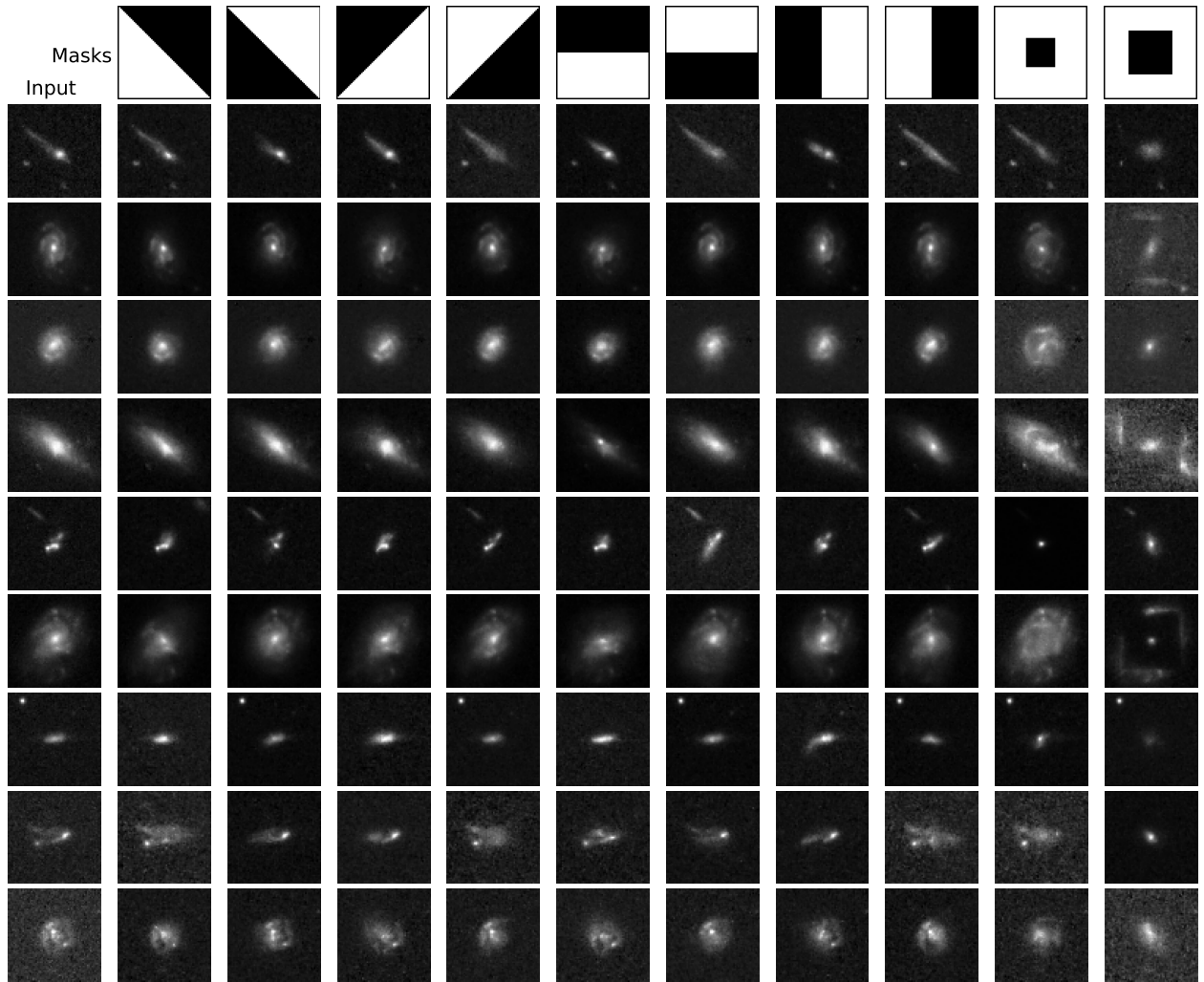


Fig. A.1. The output of applying the inpainting method to a selection of images with example masks (top row). The input image (leftmost column) has the pixels in the white area of each mask fixed, while the diffusion-based inpainting regenerates the black area. Although this model's aim is not explicitly large-scale image generation, applying masks that cover large selections of the image showcases the model's ability to create realistic galaxies. Ensuring that the generated images preserve a shape and brightness consistent with the fixed pixels shows that the model has a good internal representation of the dynamics of galaxy structure.

Appendix B: Alternative loss

Making use of standard metrics whose primary use-case is data on a consistent scale, such as 0–255 pixels, may not function optimally with the high dynamic range of VIS data. Specifically, since MSE is not scale invariant, having images that feature pixels that are an order of magnitude brighter, results in relative errors increasing by a square of that difference. This is typically a benefit as it forces outlier predictions to be severely penalised. However, with our data, this results in brighter images being disproportionately penalised compared to fainter ones, even if the predictions are both the same relative distance away from the true prediction.

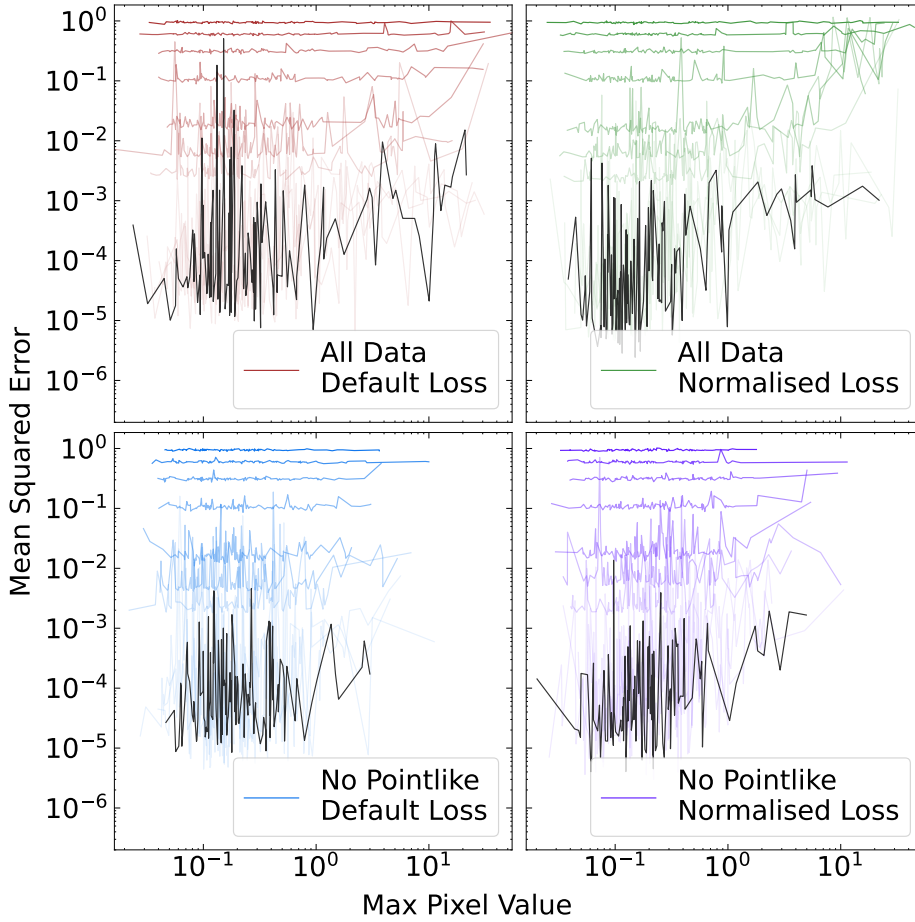


Fig. B.1. MSE of a random batch at different epochs throughout training. The opacity of each line indicates how far into the training epochs the measurement is, with fainter lines at higher epochs. Each training pair exhibits different behaviour as training progresses, indicating the weight each puts on an image's pixel brightness. Because each model was trained on more than solely MSE, these measurements are only to highlight the raw pixel performance of image recreation at each stage of training. The black line indicates the final step during training.

As shown in Sect. 2.3, we applied a normalisation to L_{MSE} by dividing each image's error by its respective maximum pixel value, which could be from a brighter external source. To visualise the impact of amending the loss function, with respect to an images brightness, Fig. B.1 shows how the error from the unmodified MSE changes as the model is trained for longer. Each line represents the error for a collection of images in a random batch, with the increased transparency indicating the passage of time where the model had been trained on more epochs. The final epoch of each model is presented by a black line. The effect of removing point-like sources, as discussed in Sect. 4.1, is also explored.

Although all four models arrive at a similar minimum loss, and begin the training process near-identically, their respective behaviour throughout training differs significantly. For the full dataset with the default loss (top left of Fig. B.1), training prioritises brighter images, leading to some significantly low errors towards the rightmost data. However, this comes at the expense of fainter data, where wild fluctuations in error show that the model's output is not very robust when it comes to precise predictions. This happens even when MSE might suggest low error, since pixel values below 1.0 result in minimal losses despite the predictions being relatively incorrect.

For the normalised loss with the full dataset (top right), since the penalisation is shared more fairly across pixel values, it is the density of images that has the most influence. Early on in training we can see that as the majority of data reduces in error, the brightest images remain at near-starting performance. This continues until the errors of the fainter images become so low that model is forced to prioritise the bright images. The key difference here is this only happens after the rest of the images are at a sufficient standard. The final outcome is fainter images performing well and the entire scale of images having the same maximum error.

The two models which feature the reduced training set (bottom row) share similar dynamics during training, regardless of the loss used. This is largely due to many point-like sources also being some of the brightest images and therefore the scale of the dataset is reduced. As the most problematic data are removed, the impact of disproportionate penalisation is significantly reduced, allowing the fainter sources to be the focus during training.

Appendix C: Reconstruction metrics

Figure C.1 shows the differences in the ROC measurement between input images and the inpainting output.

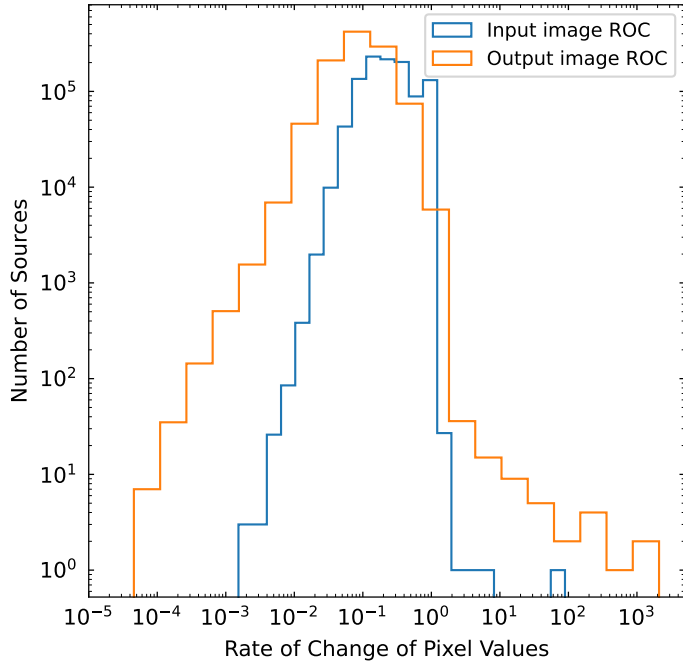


Fig. C.1. Comparison of the ROC between the original images and the inpainted images. The similarity in the peak of the distributions shows that the model is able to accurately recreate the distribution of the data.

Appendix D: Extreme examples

We present output examples from the inpainting pipeline where pixel values deviate significantly from the expected input pixels. Figure D.1 highlights much fainter pixels, while Figure D.2 shows much brighter ones.

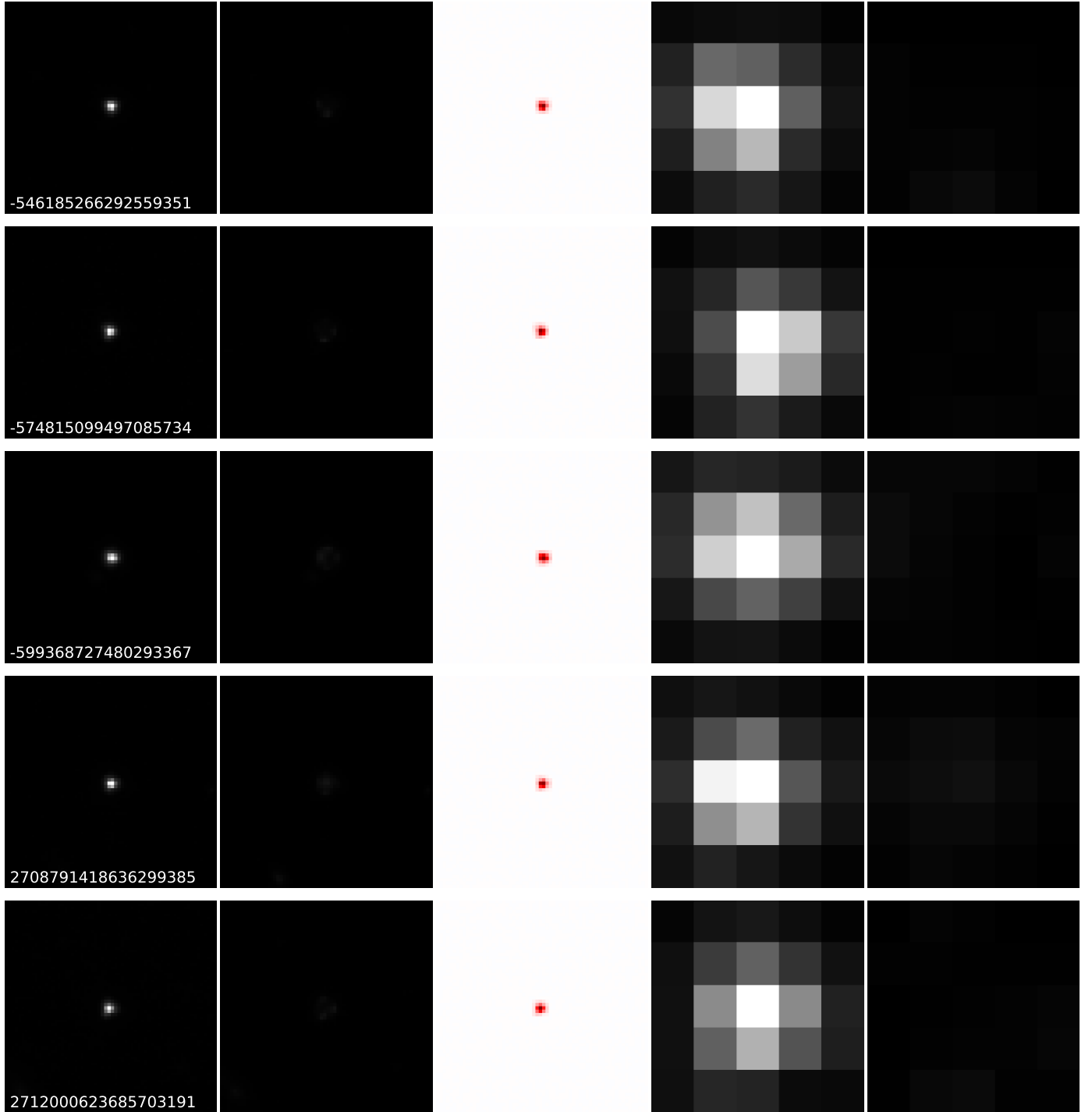


Fig. D.1. Examples of sources that achieved a very high maximum pixel difference ratio.

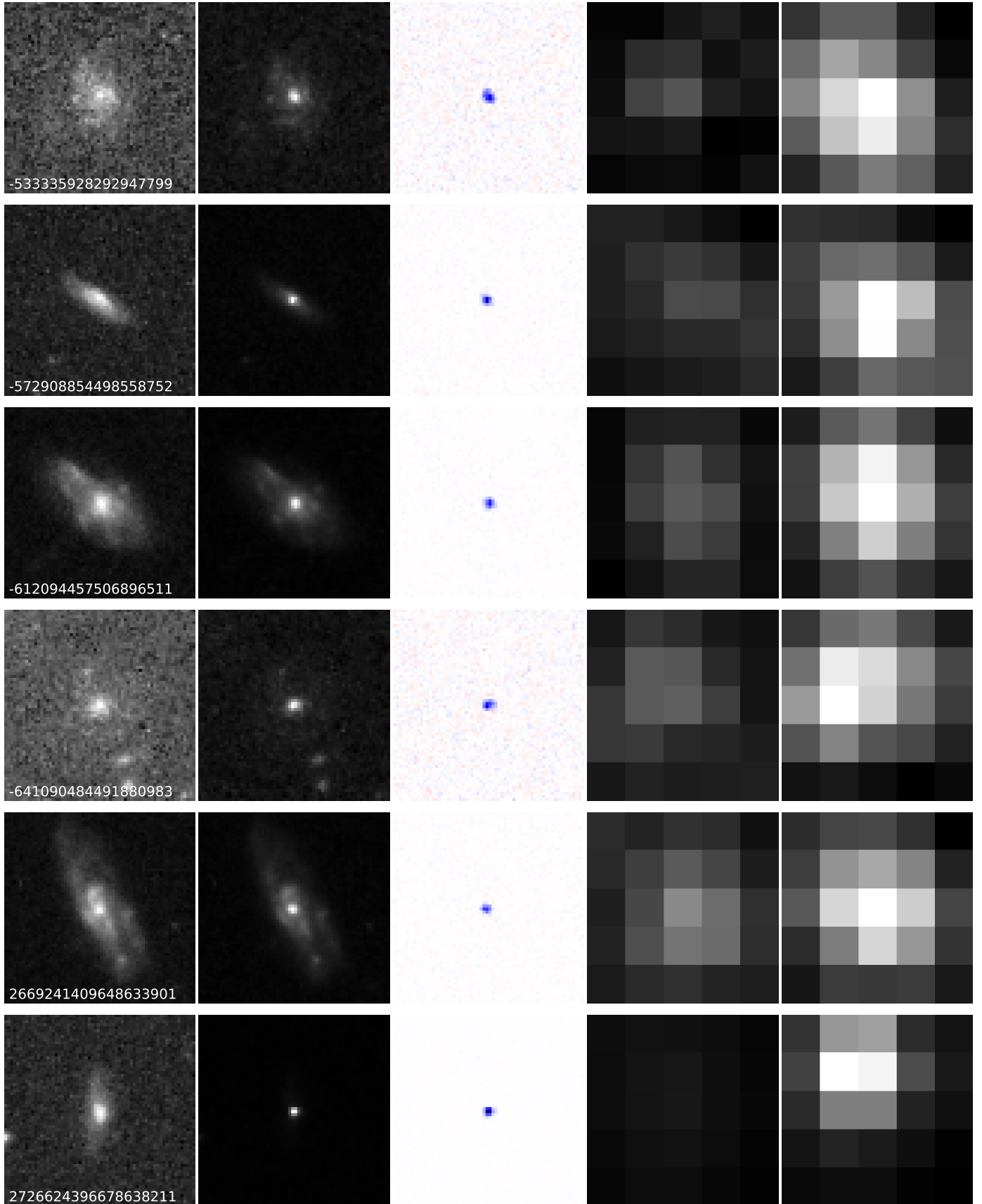


Fig. D.2. Examples of sources that achieved a very low maximum pixel difference ratio.

Appendix E: Morphology examples

We show output examples from various morphology types to showcase how the diffusion model adapts to different galaxy shapes. Edge-on galaxies, spirals and mergers are shown in Figs. E.1, E.2 and E.3, respectively.

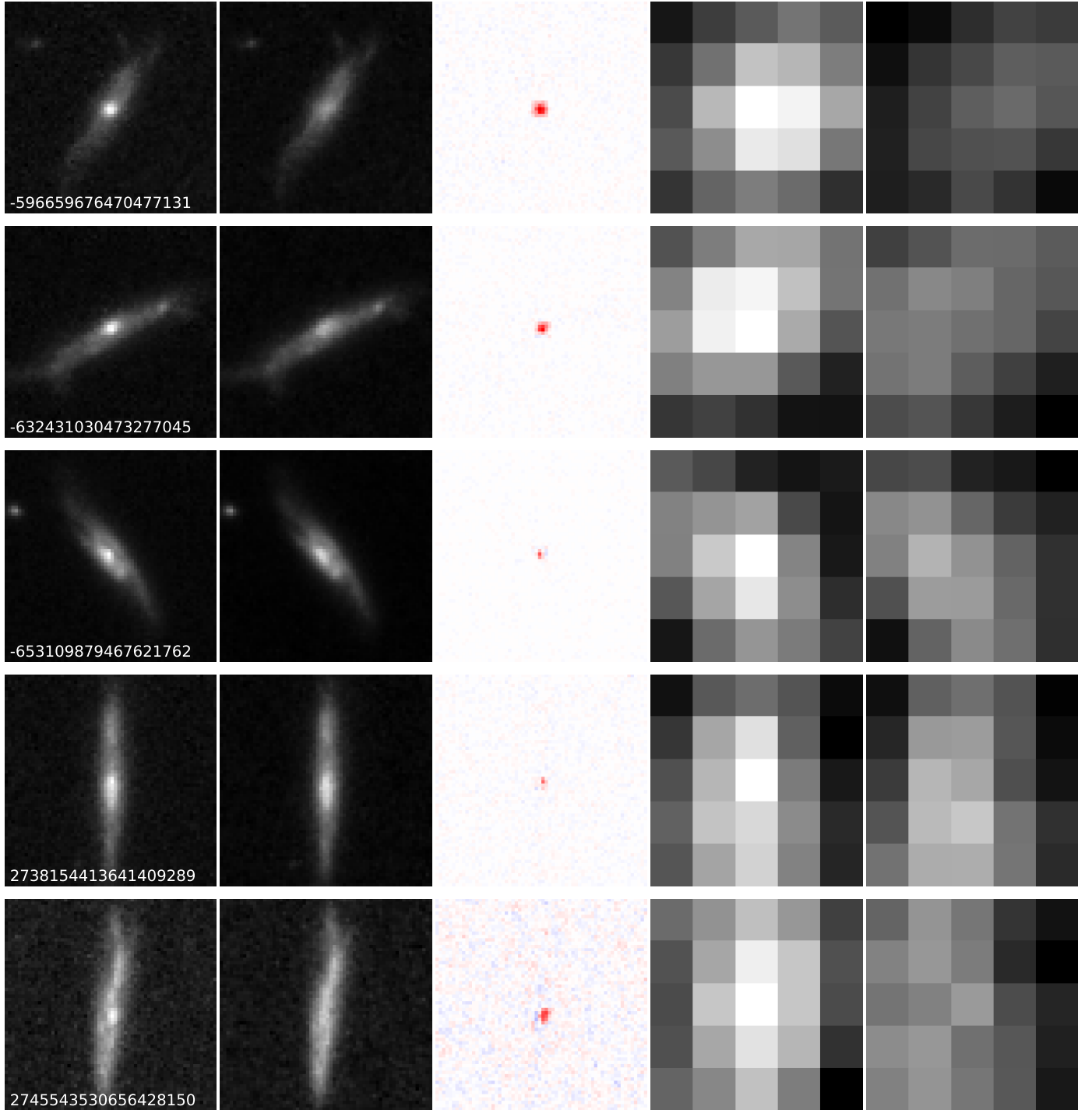


Fig. E.1. Examples of inpainting on edge-on sources.

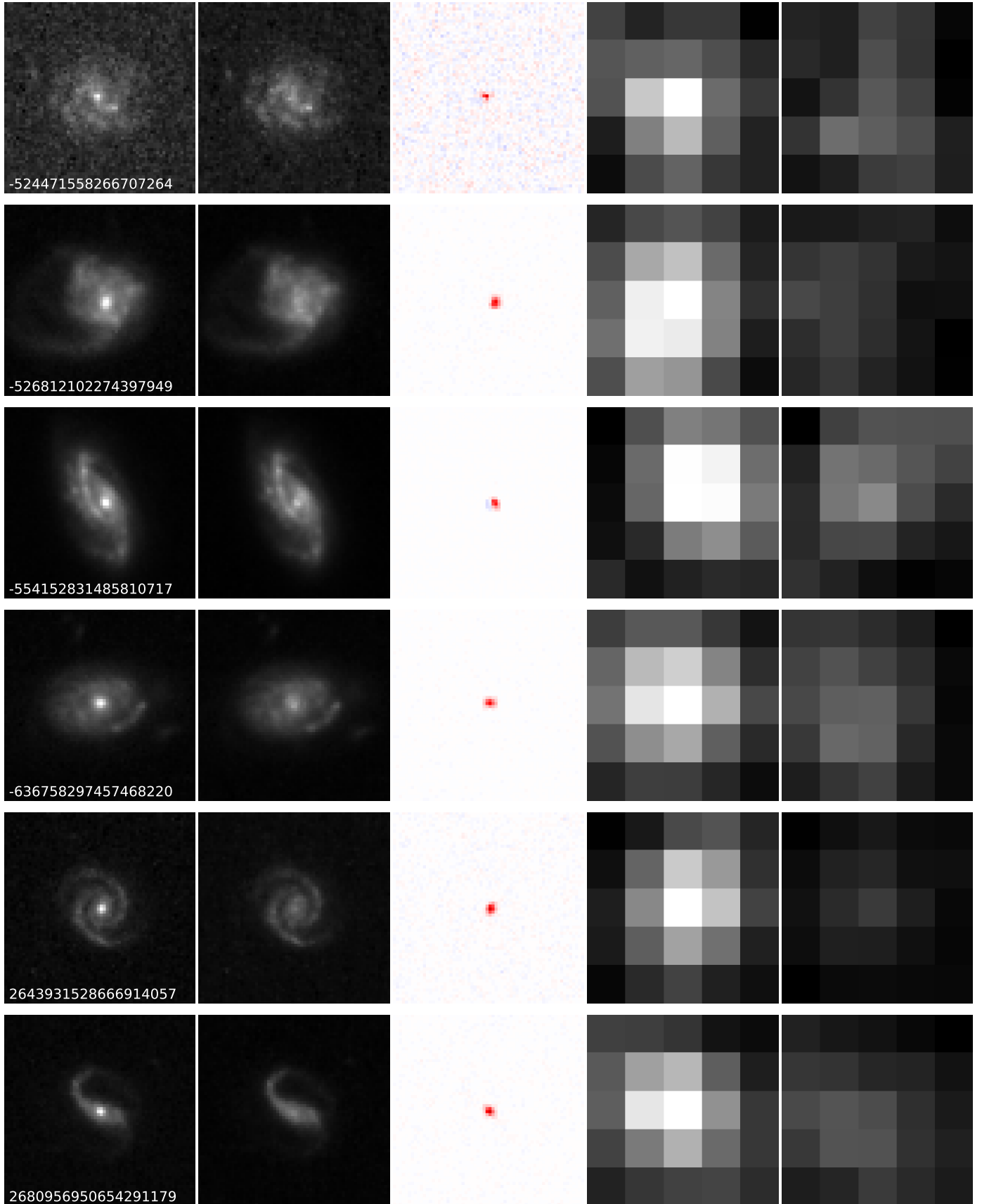


Fig. E.2. Examples of inpainting on spiral sources.

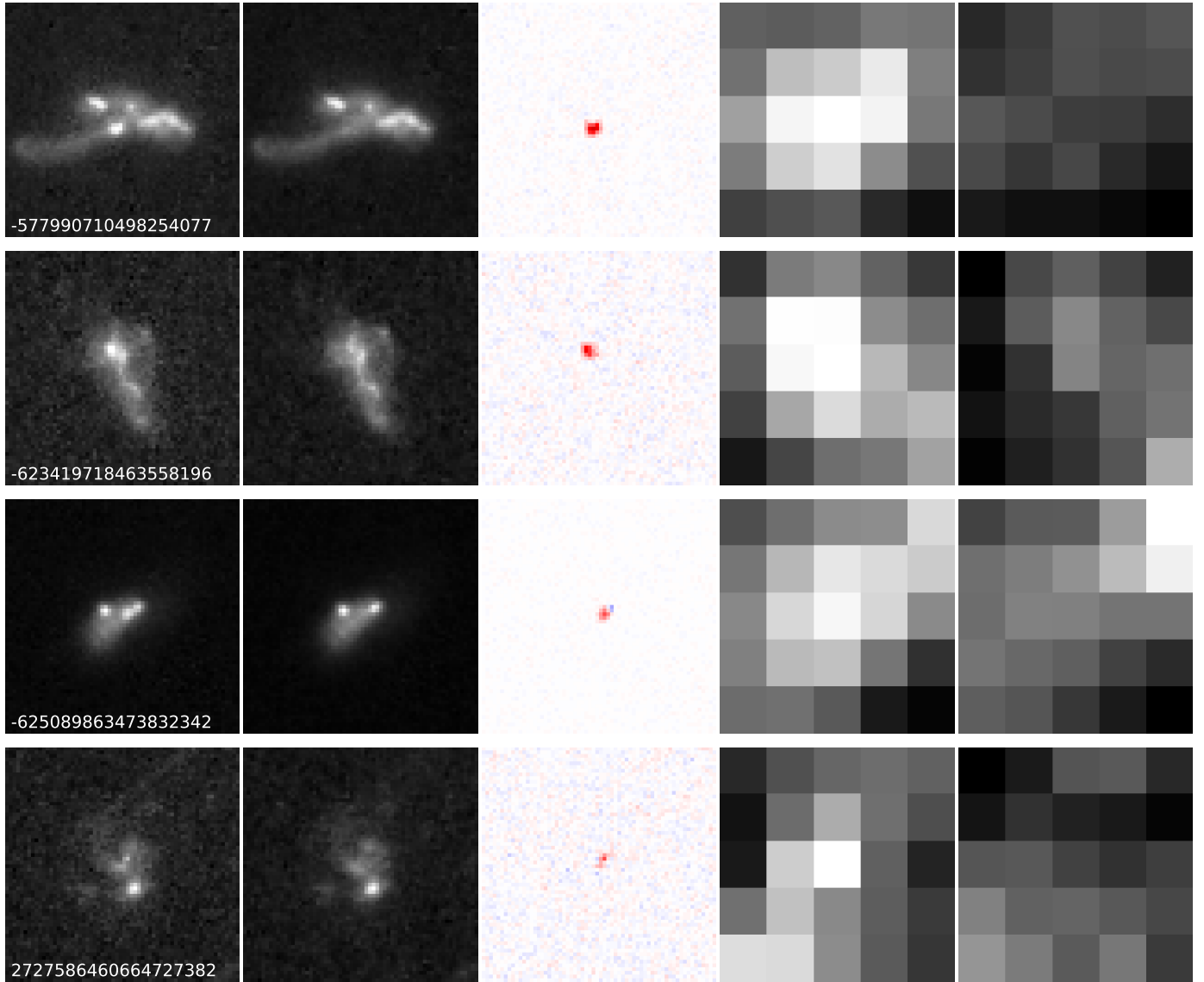


Fig. E.3. Examples of inpainting on mergers.

Appendix F: Additional results

We present collated results for each of the star, galaxy, AGN and QSO selections in Fig. F.1. In Fig. F.2, the MSE and max ratio scores of all the MB26 selections as shown, indicating our diffusion model is more sensitive to selecting brighter F_{AGN} sources.

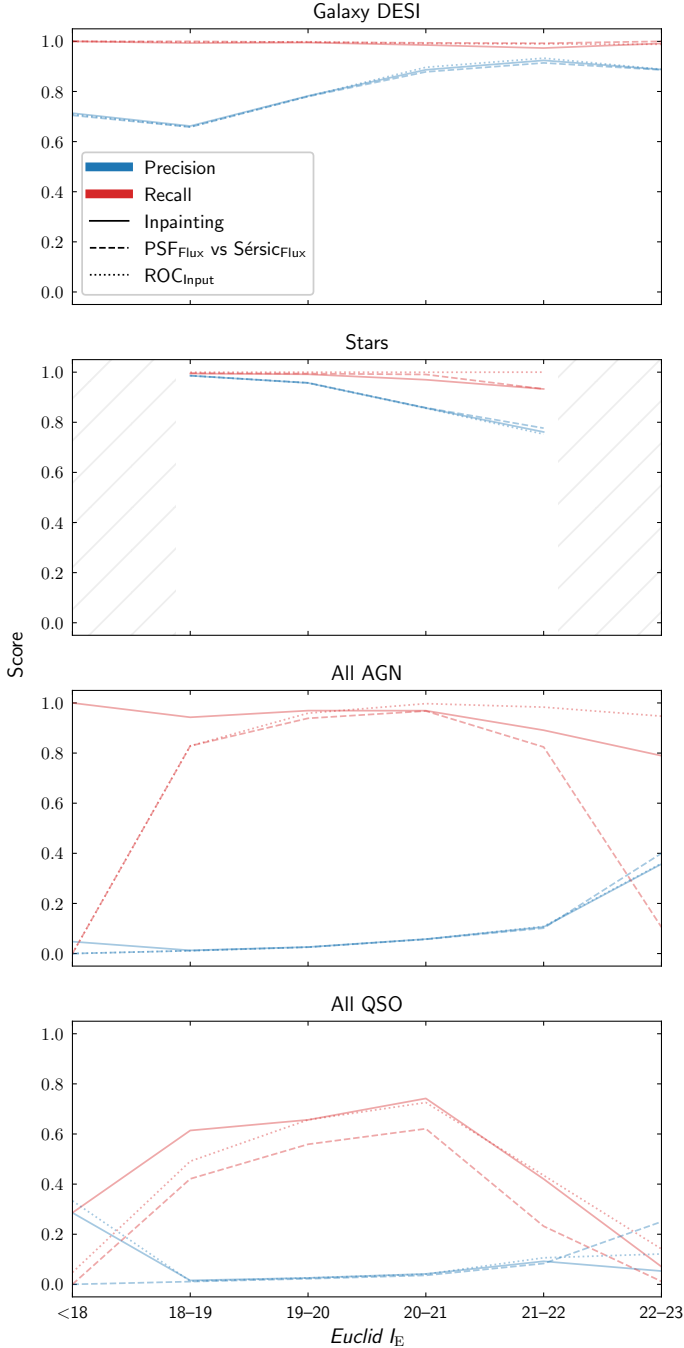


Fig. F.1. Precision and recall scores for the diffusion-based AGN predictions (solid line). All of the respective selection are collated into their common classification types.

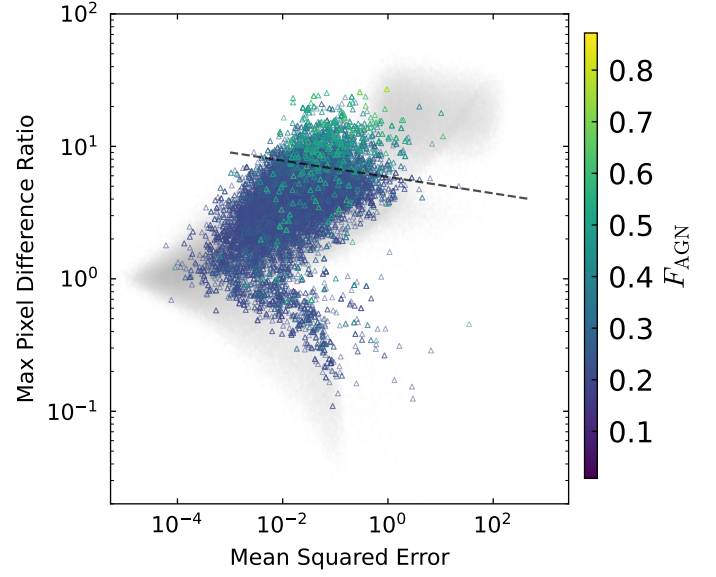


Fig. F.2. MB26 selection on our inpainting metrics. The F_{AGN} values are provided by MB26. All shown sources have an $F_{\text{AGN}} > 0.2$. A large selection of sources within the MB26 selection are not located within our boundary. Of the sources that are within our boundary, many exhibit higher F_{AGN} .