

## Euclid Quick Data Release (Q1)

### Active galactic nuclei identification using diffusion-based inpainting of *Euclid* VIS images

Euclid Collaboration: G. Stevens<sup>\*1</sup>, S. Fotopoulou<sup>1</sup>, M. N. Bremer<sup>1</sup>, T. Matamoro Zatarain<sup>1</sup>, K. Jahnke<sup>2</sup>, B. Margalef-Bentabol<sup>3</sup>, M. Huertas-Company<sup>4,5,6,7</sup>, M. J. Smith<sup>8,9</sup>, M. Walmsley<sup>10,11</sup>, M. Salvato<sup>12</sup>, M. Mezcua<sup>13,14</sup>, A. Paulino-Afonso<sup>15,16</sup>, M. Siudek<sup>5,13</sup>, M. Talia<sup>17,18</sup>, F. Ricci<sup>19,20</sup>, W. Roster<sup>12</sup>, N. Aghanim<sup>21</sup>, B. Altieri<sup>22</sup>, S. Andreon<sup>23</sup>, H. Aussel<sup>24</sup>, C. Baccigalupi<sup>25,26,27,28</sup>, M. Baldi<sup>29,18,30</sup>, S. Bardelli<sup>18</sup>, P. Battaglia<sup>18</sup>, A. Biviano<sup>26,25</sup>, A. Bonchi<sup>31</sup>, E. Branchini<sup>32,33,23</sup>, M. Brescia<sup>34,35</sup>, J. Brinchmann<sup>16,36</sup>, S. Camera<sup>37,38,39</sup>, G. Cañas-Herrera<sup>40,41,42</sup>, V. Capobianco<sup>39</sup>, C. Carbone<sup>43</sup>, J. Carretero<sup>44,45</sup>, M. Castellano<sup>20</sup>, G. Castignani<sup>18</sup>, S. Cavuoti<sup>35,46</sup>, K. C. Chambers<sup>47</sup>, A. Cimatti<sup>48</sup>, C. Colodro-Conde<sup>4</sup>, G. Congedo<sup>49</sup>, C. J. Conselice<sup>11</sup>, L. Conversi<sup>50,22</sup>, Y. Copin<sup>51</sup>, A. Costille<sup>52</sup>, F. Courbin<sup>53,54</sup>, H. M. Courtois<sup>55</sup>, M. Cropper<sup>56</sup>, A. Da Silva<sup>57,58</sup>, H. Degaudenzi<sup>59</sup>, G. De Lucia<sup>26</sup>, C. Dolding<sup>56</sup>, H. Dole<sup>21</sup>, M. Douspis<sup>21</sup>, F. Dubath<sup>59</sup>, X. Dupac<sup>22</sup>, S. Dusini<sup>60</sup>, S. Escoffier<sup>61</sup>, M. Farina<sup>62</sup>, S. Ferriol<sup>51</sup>, K. George<sup>63</sup>, C. Giocoli<sup>18,30</sup>, B. R. Granett<sup>23</sup>, A. Grazian<sup>64</sup>, F. Grupp<sup>12,63</sup>, S. V. H. Haugan<sup>65</sup>, I. M. Hook<sup>66</sup>, F. Hormuth<sup>67</sup>, A. Hornstrup<sup>68,69</sup>, P. Hudelot<sup>70</sup>, M. Jhabvala<sup>71</sup>, E. Keihänen<sup>72</sup>, S. Kermiche<sup>61</sup>, A. Kiessling<sup>73</sup>, M. Kilbinger<sup>24</sup>, B. Kubik<sup>51</sup>, M. Kümmel<sup>63</sup>, H. Kurki-Suonio<sup>74,75</sup>, Q. Le Boulc’h<sup>76</sup>, A. M. C. Le Brun<sup>77</sup>, D. Le Mignant<sup>52</sup>, P. B. Lilje<sup>65</sup>, V. Lindholm<sup>74,75</sup>, I. Lloro<sup>78</sup>, G. Mainetti<sup>76</sup>, D. Maino<sup>79,43,80</sup>, E. Maiorano<sup>18</sup>, O. Marggraf<sup>81</sup>, M. Martinelli<sup>20,82</sup>, N. Martinet<sup>52</sup>, F. Marulli<sup>17,18,30</sup>, R. Massey<sup>83</sup>, S. Maurogordato<sup>84</sup>, H. J. McCracken<sup>70</sup>, E. Medinaceli<sup>18</sup>, S. Mei<sup>85,86</sup>, M. Melchior<sup>87</sup>, M. Meneghetti<sup>18,30</sup>, E. Merlin<sup>20</sup>, G. Meylan<sup>88</sup>, A. Mora<sup>89</sup>, M. Moresco<sup>17,18</sup>, L. Moscardini<sup>17,18,30</sup>, R. Nakajima<sup>81</sup>, C. Neissner<sup>90,45</sup>, S.-M. Niemi<sup>40</sup>, C. Padilla<sup>90</sup>, S. Paltani<sup>59</sup>, F. Pasian<sup>26</sup>, K. Pedersen<sup>91</sup>, W. J. Percival<sup>92,93,94</sup>, V. Pettorino<sup>40</sup>, G. Polenta<sup>31</sup>, M. Poncet<sup>95</sup>, L. A. Popa<sup>96</sup>, L. Pozzetti<sup>18</sup>, F. Raison<sup>12</sup>, R. Rebolo<sup>4,97,98</sup>, A. Renzi<sup>99,60</sup>, J. Rhodes<sup>73</sup>, G. Riccio<sup>35</sup>, E. Romelli<sup>26</sup>, M. Roncarelli<sup>18</sup>, R. Saglia<sup>63,12</sup>, A. G. Sánchez<sup>12</sup>, D. Sapone<sup>100</sup>, J. A. Schewtschenko<sup>49</sup>, M. Schirmer<sup>2</sup>, P. Schneider<sup>81</sup>, T. Schrabback<sup>101</sup>, A. Secroun<sup>61</sup>, S. Serrano<sup>14,102,13</sup>, P. Simon<sup>81</sup>, C. Sirignano<sup>99,60</sup>, G. Sirri<sup>30</sup>, J. Skottfelt<sup>103</sup>, L. Stanco<sup>60</sup>, J. Steinwagner<sup>12</sup>, P. Tallada-Crespí<sup>44,45</sup>, A. N. Taylor<sup>49</sup>, I. Tereno<sup>57,104</sup>, S. Toft<sup>105,106</sup>, R. Toledo-Moreo<sup>107</sup>, F. Torradeflot<sup>45,44</sup>, I. Tutusaus<sup>108</sup>, L. Valenziano<sup>18,109</sup>, J. Valiviita<sup>74,75</sup>, T. Vassallo<sup>63,26</sup>, G. Verdoes Kleijn<sup>110</sup>, A. Veropalumbo<sup>23,33,32</sup>, Y. Wang<sup>111</sup>, J. Weller<sup>63,12</sup>, A. Zacchei<sup>26,25</sup>, G. Zamorani<sup>18</sup>, F. M. Zerbi<sup>23</sup>, I. A. Zinchenko<sup>63</sup>, E. Zucca<sup>18</sup>, V. Allevato<sup>35</sup>, M. Ballardini<sup>112,113,18</sup>, M. Bolzonella<sup>18</sup>, E. Bozzo<sup>59</sup>, C. Burigana<sup>114,109</sup>, R. Cabanac<sup>108</sup>, A. Cappi<sup>18,84</sup>, J. A. Escartin Vigo<sup>12</sup>, L. Gabarra<sup>115</sup>, W. G. Hartley<sup>59</sup>, J. Martín-Fleitas<sup>89</sup>, S. Matthew<sup>49</sup>, R. B. Metcalf<sup>17,18</sup>, A. Pezzotta<sup>116,12</sup>, M. Pöntinen<sup>74</sup>, I. Risso<sup>117</sup>, V. Scottez<sup>118,119</sup>, M. Sereno<sup>18,30</sup>, M. Tenti<sup>30</sup>, M. Wiesmann<sup>65</sup>, Y. Akrami<sup>120,121</sup>, S. Alvi<sup>112</sup>, I. T. Andika<sup>122,123</sup>, S. Anselmi<sup>60,99,124</sup>, M. Archidiacono<sup>79,80</sup>, F. Atrio-Barandela<sup>125</sup>, D. Bertacca<sup>99,64,60</sup>, M. Bethermin<sup>126</sup>, L. Bisigello<sup>64</sup>, A. Blanchard<sup>108</sup>, L. Blot<sup>127,77</sup>, S. Borgani<sup>128,25,26,27,129</sup>, M. L. Brown<sup>11</sup>, S. Bruton<sup>130</sup>, A. Calabro<sup>20</sup>, F. Caro<sup>20</sup>, T. Castro<sup>26,27,25,129</sup>, F. Cogato<sup>17,18</sup>, S. Davini<sup>33</sup>, G. Desprez<sup>110</sup>, A. Díaz-Sánchez<sup>131</sup>, J. J. Diaz<sup>4</sup>, S. Di Domizio<sup>32,33</sup>, J. M. Diego<sup>132</sup>, P.-A. Duc<sup>126</sup>, A. Enia<sup>29,18</sup>, Y. Fang<sup>63</sup>, A. G. Ferrari<sup>30</sup>, A. Finoguenov<sup>74</sup>, A. Fontana<sup>20</sup>, A. Franco<sup>133,134,135</sup>, J. García-Bellido<sup>120</sup>, T. Gasparetto<sup>26</sup>, V. Gautard<sup>136</sup>, E. Gaztanaga<sup>13,14,137</sup>, F. Giacomini<sup>30</sup>, F. Gianotti<sup>18</sup>, M. Guidi<sup>29,18</sup>, C. M. Gutierrez<sup>138</sup>, A. Hall<sup>49</sup>, S. Hemmati<sup>139</sup>, H. Hildebrandt<sup>140</sup>, J. Hjorth<sup>91</sup>, J. J. E. Kajava<sup>141,142</sup>, Y. Kang<sup>59</sup>, V. Kansal<sup>143,144</sup>, D. Karagiannis<sup>112,145</sup>, C. C. Kirkpatrick<sup>72</sup>, S. Kruk<sup>22</sup>, L. Legrand<sup>146,147</sup>, M. Lembo<sup>112,113</sup>, F. Lepori<sup>148</sup>, G. Leroy<sup>149,83</sup>, J. Lesgourgues<sup>150</sup>, L. Leuzzi<sup>17,18</sup>, T. I. Liaudat<sup>151</sup>, J. Macias-Perez<sup>152</sup>, M. Magliocchetti<sup>62</sup>, F. Mannucci<sup>153</sup>, R. Maoli<sup>154,20</sup>, C. J. A. P. Martins<sup>15,16</sup>, L. Maurin<sup>21</sup>, M. Miluzio<sup>22,155</sup>, P. Monaco<sup>128,26,27,25</sup>, G. Morgante<sup>18</sup>, K. Naidoo<sup>137</sup>, A. Navarro-Alsina<sup>81</sup>, F. Passalacqua<sup>99,60</sup>, K. Paterson<sup>2</sup>, L. Patrizii<sup>30</sup>, A. Pisani<sup>61</sup>, D. Potter<sup>148</sup>, S. Quai<sup>17,18</sup>, M. Radovich<sup>64</sup>, P.-F. Rocci<sup>21</sup>, G. Rodighiero<sup>99,64</sup>, S. Sacquegna<sup>134,133,135</sup>, M. Sahlén<sup>156</sup>, D. B. Sanders<sup>47</sup>, E. Sarpa<sup>28,129,27</sup>, A. Schneider<sup>148</sup>, M. Schultheis<sup>84</sup>, D. Sciotti<sup>20,82</sup>, E. Sellentin<sup>157,42</sup>, F. Shankar<sup>158</sup>, L. C. Smith<sup>159</sup>, K. Tanidis<sup>115</sup>, G. Testera<sup>33</sup>, R. Teyssier<sup>160</sup>, S. Tosi<sup>32,33,23</sup>, A. Troja<sup>99,60</sup>, M. Tucci<sup>59</sup>, C. Valieri<sup>30</sup>, D. Vergani<sup>18</sup>, G. Verza<sup>161</sup>, and N. A. Walton<sup>159</sup>

(Affiliations can be found after the references)

March 17, 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 an active galactic nucleus (AGN) feature intense, concentrated emission from gas accretion around supermassive black holes, superimposed on regular galactic light, while quasi-stellar objects (QSO) are the extreme case of the AGN emission dominating the galaxy. The challenge of identifying AGN and QSO has been discussed many times in the literature, often requiring multi-wavelength observations. This paper introduces a novel approach to identify AGN and QSO from a single image. Diffusion models have been recently 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 condition the prediction of the central light distribution by masking the central few pixels of each source and reconstruct the light according to the diffusion model. We further use 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.** Galaxies: active – techniques: image processing – methods: data analysis – methods: observational

## 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 galaxy’s emission shows evidence for 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 characterising 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 by using a much wider range of multi-wavelength data. 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 data set from the *Euclid* mission, covering a

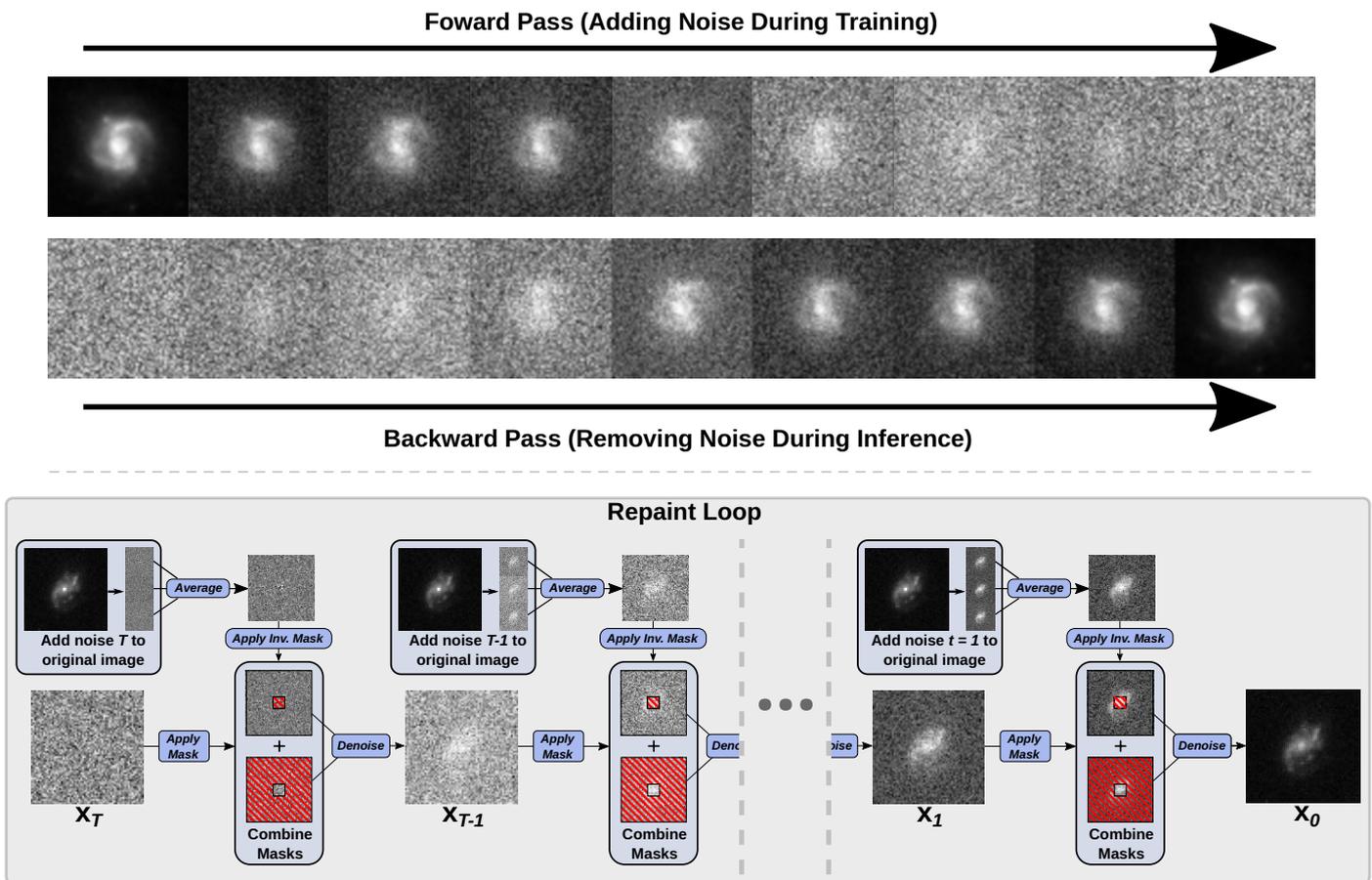
large fraction of the extragalactic sky, there is the potential to derive unprecedentedly large samples of AGN across a wide range of (optical) luminosity and redshift parameter space with the use of an appropriate and sufficiently efficient technique. The technique can then be subsequently refined by feeding back the result of using other multi-wavelength data, allowing us to confirm or reject candidates. With sufficient refinement, this further assessment step subsequently becomes either unnecessary or at the very least an efficient stage in compiling reliable samples. This then opens the door to production of unprecedentedly large and reliable samples of optically-selected AGN.

In this work, we present 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. 2024; *Euclid* Collaboration: Cropper et al. 2024) imaging alone. Diffusion-based methods are machine-learning models used to generate or reconstruct data by evolving a random distribution of pixels into a structured output over several steps. This technique outperforms traditional methods like median pixel substitution or interpolation, which often struggle to recreate complex structures within images. By using these models for ‘inpainting’, where 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.

Briefly, the technique takes an image of a galaxy that may or may not have AGN emission at its centre and excludes those central pixels that may be influenced by AGN emission. The model then predicts (‘inpaints’) the underlying stellar light profile in those pixels from a comparison of the rest of the galaxy image to a model derived from a very large sample of similar galaxy images. Any significant difference between the prediction and true image in those central regions indicates the potential presence of AGN emission. Because AGN are comparatively rare, the sample of galaxy images used in the generation of the model does not need to be filtered to reject AGN, the method will be inherently biased to the ‘typical’ behaviour of galaxy light distributions and so naturally down-weights or excludes predicting central regions with AGN emission.

While this paper details the basic technique applied to VIS imaging data, as this is a machine-learning technique, future enhancements would allow incorporation of the results of a detailed assessment step, involving other data sets, in order to further improve the accuracy and reliability of the AGN candidate

\* e-mail: grant.stevens@bristol.ac.uk



**Fig. 1.** The diffusion pipeline (*top*) 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 as input pure Gaussian noise and is able to 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 both the retained pixels and newly generated areas, creating a different yet plausible final image.

samples produced directly by the technique. Such steps could include multi-wavelength cross-matching or spectroscopic validation, from which we can ultimately reduce the need for further time-expensive refinement steps. Additionally, the technique as presented here uses a standard treatment of diffusion models which in principle can be significantly further optimised for the specific characteristics of the *Euclid* data set, 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 such as generative adversarial networks (GANs, Goodfellow et al. 2014, 2020) or variational autoencoders (VAEs, Kingma 2013) are trained with the aim to learn the underlying distribution of a data set to generate novel examples.

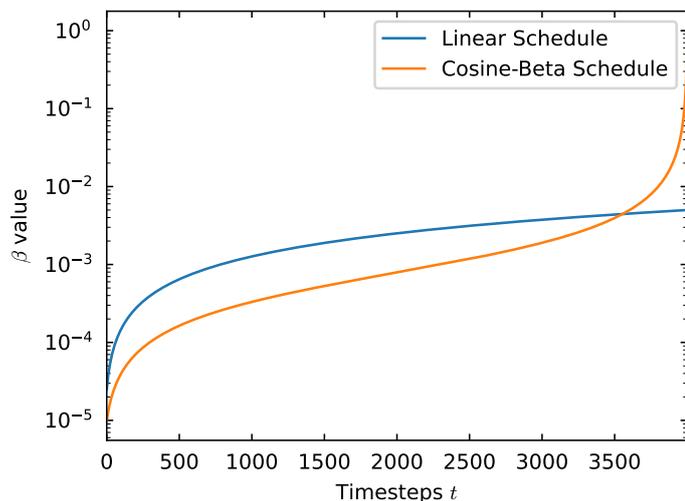
Using a pair of competing networks, GANs ensure that once trained the generated images are of high quality, leading to the creation of many artificially generated data sets that are indistinguishable from the original data set (Karras et al. 2020). However, the adversarial training loop is often difficult to train, often leading to mode collapse, meaning the generator focuses on

only a small subset of the total search space leading to generating samples of similar instances (Thanh-Tung & Tran 2020).

On the other hand, VAEs use an encoder network to compress the data into a latent space of fewer dimensions. The network then uses the latent space to recreate the input through an often symmetric decoder network. VAEs provide better coverage over the search space, improving stability during training as well as increased variety in generated outputs. However, the drawback for VAEs becomes apparent when handling more complex, high dimensional data. The use of latent spaces does not guarantee a lossless compression of information, often leading to accurate but blurry outputs.

### 2.2. Diffusion models

Diffusion-based models provide an alternative process for image generation that allows for both good coverage, as well as the ability to represent complex data distributions (Dhariwal & Nichol 2021). For both GANs and VAEs, their significant advantage with respect to diffusion models is that they offer faster sample generation and lower computational costs. Whereas 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.



**Fig. 2.** Linear schedule from the original diffusion implementation, showing the parameter converging early in the timesteps. This results in training images becoming pure noise too early into the timesteps 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 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 where 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  $\beta_t$  term defines the variance of the Gaussian noise added at each timestep. By ensuring a gradual increase in noise, the model is more easily able to learn the transition from the data distribution to the noise distribution.

With a large enough  $T$  and an adjusted time-dependent variance  $\beta_t$ ,  $x_T$  will approach a Gaussian distribution that is isotropic, meaning uniform in all directions. For the denoising process, starting at a  $x_T \in \mathcal{N}(0, I)$ , 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.

Smith et al. (2022) also utilised the DDPM framework to generate a data set of realistic galaxies. The model is trained us-

ing the Photometry and Rotation curve OBServations from Extragalactic Surveys (PROBES) data set (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 not only visually realistic images, but also how they feature similar physical property distributions such as 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 include galaxy image deconvolution (Adam et al. 2023; Spagnoletti et al. 2024), gravitation lensing analysis (Adam et al. 2022; Remy et al. 2023), and deblending (Sampson et al. 2024).

### 2.3. Noise scheduling

The larger the number of total timesteps  $T$ , the more refined the model can become at accurately generating realistic images, since less complex noise dynamics are required to be learnt for each timestep  $t$ . This does, however, require the model to learn the dynamics of more timesteps, leading to extra overhead in both training and inference, since more steps must be performed to traverse from pure noise to realistic images.

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 were competing 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 so that less noise is added in early iterations and more in the later, the process overcame the issue of images became too noisy too quickly. Figure 2 shows how the scheduler impacts the  $\beta$  parameter used in Eq. (2). Equation (3) shows how  $\beta_t$  is calculated, where the reader is directed to Nichol & Dhariwal (2021) for specifics on its derivation:

$$\beta_t = 1 - \frac{\bar{\alpha}_{t+1}}{\bar{\alpha}_t}, \quad (3)$$

where

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

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. 5.

### 2.4. 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  $t$ . Due to the cosine scheduler discussed previously, the model will alter an image very differently at a timestep

close to  $T$ , where the image is still very noisy, than when the image is close to timestep 0 and near completion. However, we are not limited to only conditioning by timestep, often multiple parameters are provided alongside each image. Typically, this is in respect to 'class-conditioning', where 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 learns not only the dynamics of the noise but the inherent differences of noise and non-noise pixel values for each of the 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 the possibility of complex prompts and classes, it may be difficult to know exactly what the output image will look like. Furthermore, with diffusion's ability to cover large areas of the search space, a repeat sample of the same prompt is likely to produce an image very different from the first, even if the generated image belongs to the requested class. This is due to the one-to-many relationship between classes/prompts and their subsequent generated outputs.

In a situation where the user requires consistency in part of the image, a more direct manipulation of the produced pixels is required. A process known as 'inpainting' (Pathak et al. 2016; Yeh et al. 2017; Yu et al. 2018; Wang et al. 2018), forces conditioning of specific pixels within the image, allowing the model to fix certain parts of the image whilst generating others – all whilst remaining coherent and consistent across the image as a whole. Allowing the user to selectively remove and replace certain aspects of images has gained a lot of popularity in the last few years due to its mainstream adoption by nearly all smartphone brands (Sargsyan et al. 2023).

## 2.5. Repainting

This paper uses a specific implementation of inpainting using diffusion known as Repaint (Lugmayr et al. 2022). Due to the model being conditioned on timestep  $t$ , where it is expecting the whole image to be at a certain noise level, we can not simply take an image and a corresponding mask and replace the masked pixels with noise. The model itself does not know that a mask has been applied nor does it know that the inverse mask should remain unaffected. Instead, it will continue to remove noise across the whole image as usual, leading to the clean parts of the image becoming corrupted. It is also not possible to generate an entirely new image and copy over the masked pixels to the original image as there will be no continuity between the existing and new elements.

To combine the parts of the image we want unchanged, and the parts we want to generate, both the forward pass and the backward pass of the diffusion process must be utilised. This pipeline needs to generate an image  $x_t$  that incorporates the mask and the fixed pixels for each timestep  $t$ , as shown in Fig. 1.

### Input

The main input of each timestep  $t$  of the repaint pipeline is the denoised image,  $x_t$ , from the previous timestep. To aid in the

understanding of the pipeline, this image is referred to as  $x_t^{\text{denoised}}$  for this discussion. At the first iteration when  $t = T$ , the input is simply an image of pure Gaussian noise.

### Fixed pixels

To ensure that the pixels we want unchanged are still present at the end of the diffusion process, we must interject these fixed pixels during each timestep. However, as the model is expecting all pixels to be at a specific noise level, as described in Eq. (2), we must utilise the same scheduler used in training to add the correct amount of noise for the current timestep  $t$ . Due to the stochasticity present in adding Gaussian noise to the image, the pipeline allows us to create  $N$  different versions of the input with the required noise. By averaging these samples, the important details of the image are more consistently preserved, ensuring that once all  $T$  iterations are complete, the fixed pixels look identical to the original input. We refer to the image formed from the average of the  $N$  samples as  $x_t^{\text{noised}}$ .

### Utilising the mask

To combine the parts of the image we want to keep and the parts we want to generate, we utilise a mask  $M$ , which is an array of equal size to our image, with pixel values of either zero or one. Pixels valued as one will be generated from scratch by the diffusion process, with pixels valued zero being replaced at each noise step with the fixed pixels in  $x_t^{\text{noised}}$  from the process described above. The two parts can then be combined into a coherent image and then passed back into the diffusion model to be further denoised and progress onto timestep  $t - 1$ ,

$$x_{t-1} = \text{denoise}[M \times x_t^{\text{denoised}} + (1 - M) \times x_t^{\text{noised}}]. \quad (4)$$

By denoising the combination of pixels we want to keep with the generated pixels, we condition the diffusion process to ensure that in the final image, all pixels are coherent, being conditioned on the details found within the fixed pixels.

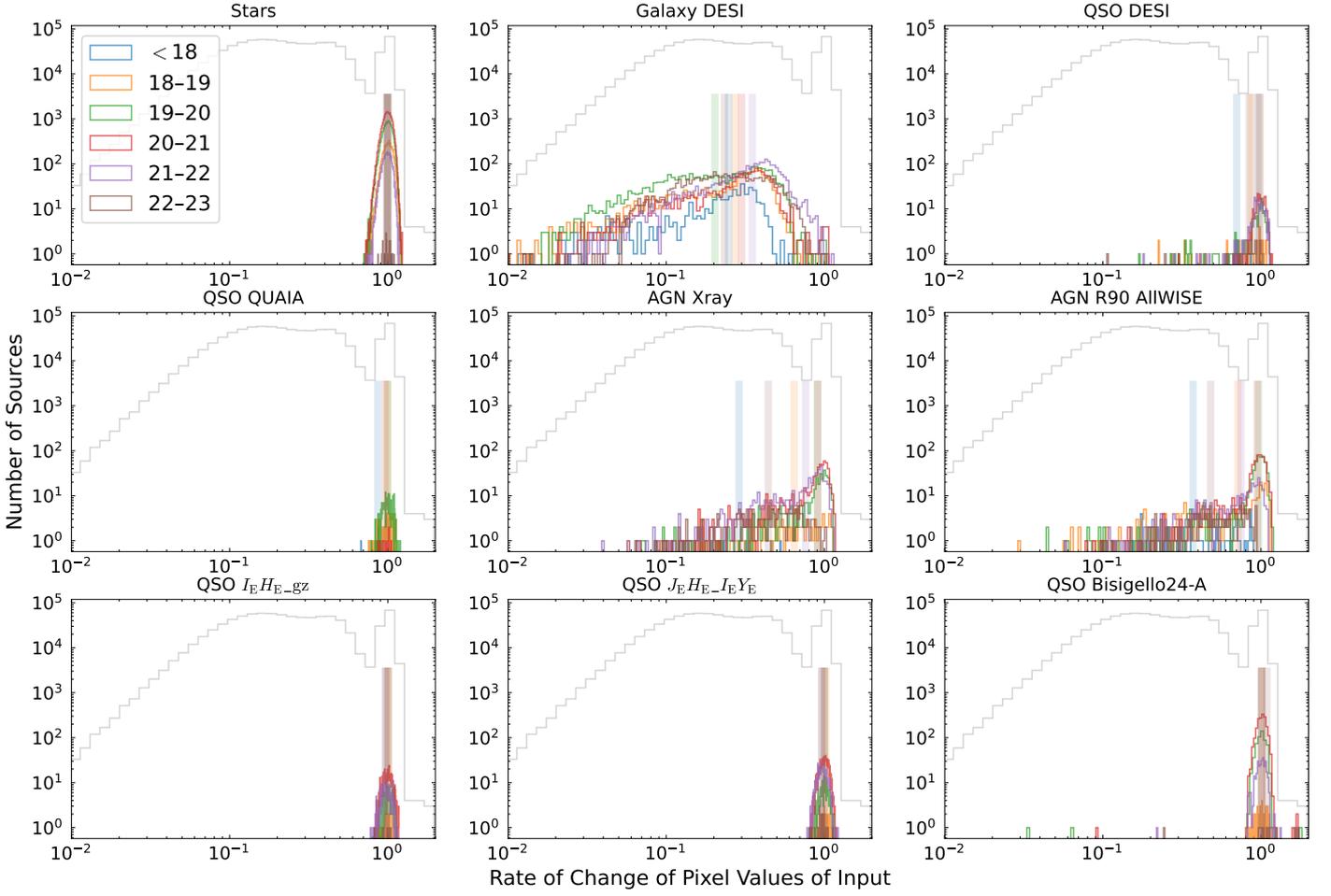
Incorporating the Repaint method for AGN identification has the following benefits

- Diffusion methods have been proven their ability in many applications to accurately recreate complex data sets, allowing for realistic generation of galaxy morphology images.
- 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.
- There exists a large imbalance of normal galaxies compared to AGN, allowing for the model to learn a bias. This can be exploited to create an outlier detection-based classifier without explicitly requiring labels during training.

## 3. Data

### 3.1. Sample selection

Covering  $63.1 \text{ deg}^2$ , the Euclid Q1 Data Release (Euclid Quick Release Q1 2025) contains just under 30 million catalogued sources (Euclid Collaboration: Romelli et al. 2025) 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



**Fig. 3.** Initial results for pixel value differences across various selections. Comparing the ratios of the centre brightest pixel with the means of the surrounding 1 and 2 pixel wide regions, show a clear distribution difference between galaxy and non-galaxy classes. The grey histogram shows the distribution of the whole data set, 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.

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 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 signal-to-noise ratio of the total flux measurement must be sufficiently high ( $\frac{\text{flux\_detection\_total\_euclid}}{\text{fluxerr\_detection\_total\_euclid}} > 15$ ); and
- if the signal-to-noise 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. \(2025\)](#).

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 of the respective sources, where 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 apply a 80:20 train and test split. Even though the inpainting and its results are a separate pipeline to the diffusion training, to prevent any potential data leakage, all of the decision boundaries created are based solely on the training set. When showing classifier performance, all results are from the unseen 20% of the data. This is to prevent any potential memorisation of the images from training 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 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 data set contained 794 624 sources.

### 3.2. Comparisons with other classifications

Although the training of the diffusion model does not make use of any labels, it is important to compare the distributions of metrics over different source selection criteria. This allows us to compare both the performance of detecting particular sources but also to 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. \(2025\)](#). A selection of stars are used from DESI ([DESI Collaboration et al. 2024](#)) and *Gaia* ([Gaia Collaboration: Vallenari et al. 2023; Gaia Collaboration: Bailer-Jones et al. 2023](#)). The normal galaxies sample is also from DESI ([DESI Collaboration et al. 2024](#)). Comparisons are made to selections of AGN and quasars using *Euclid* photometry with the  $I_E H_{gz}$  and  $JH_{I_E Y}$  selections from [Euclid Collaboration: Matamoro Zatarain et al. \(2025\)](#) and the 2-colour selection from [Euclid Collaboration: Bisigello et al. \(2024, hereafter Bisigello24-A\)](#). Additional AGN and QSO selections are used from DESI ([DESI Collaboration et al. 2024](#)), *Gaia*/Quaia ([Storey-Fisher et al. 2024](#)), as well as the 90% reliability WISE sample from ([Assef et al. 2018](#)), which is referred to as R90 for the remainder of the paper.

As is particularly true for AGN and QSO, it is impossible to create a selection criterion that will identify both with high completeness and high purity a single population of objects. Instead, we will use the aforementioned samples to assess contamination and purity of our method to the extent that is possible.

### 3.3. Data expectations

Although, by definition, a difference between a galaxy with and without a bright central component should be found, it is important to validate that this behaviour can be measured with the *Euclid* VIS images. To quantify this difference we analyse the relationship between the brightest central pixel and its surrounding pixels. By producing ratios comparing the central pixel to 1 and 2 pixels wide regions around the central pixel, we form a rate of change (ROC) metric over sources, as described by

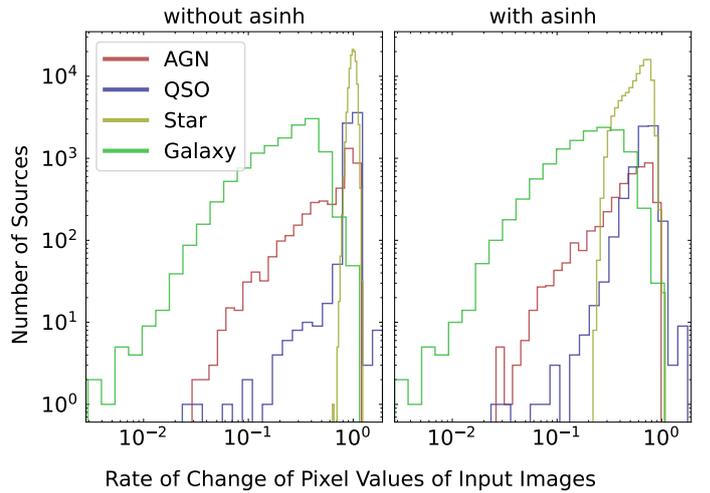
$$\text{ROC} = R_1^2 + R_2^2, \quad (5)$$

where

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

and  $\text{Input}_{1s}$  and  $\text{Input}_{2s}$  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  $9 \times 9$  pixels.

Pixels surrounding the brightest pixel, which share similar values 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. 3.2](#)), are shown in [Fig. 3](#). It is evident that sources with a high flux point-spread function (PSF) light profile show the highest ROC values, as expected. This effect is seen in the star sample and the QSO dominated selections (R90 WISE, Quaia, DESI-QSO, and Euclid-QSO). On the other hand, spectroscopically confirmed normal galaxies show very broad ROC distributions.



**Fig. 4.** Distribution differences of the rate of change 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.

**Table 1.** Percentage of galaxies that have rate of change value higher than the mean rate of change of the other classes. Due to its ability to compress the dynamic range of the image, the separability of the classes becomes more difficult.

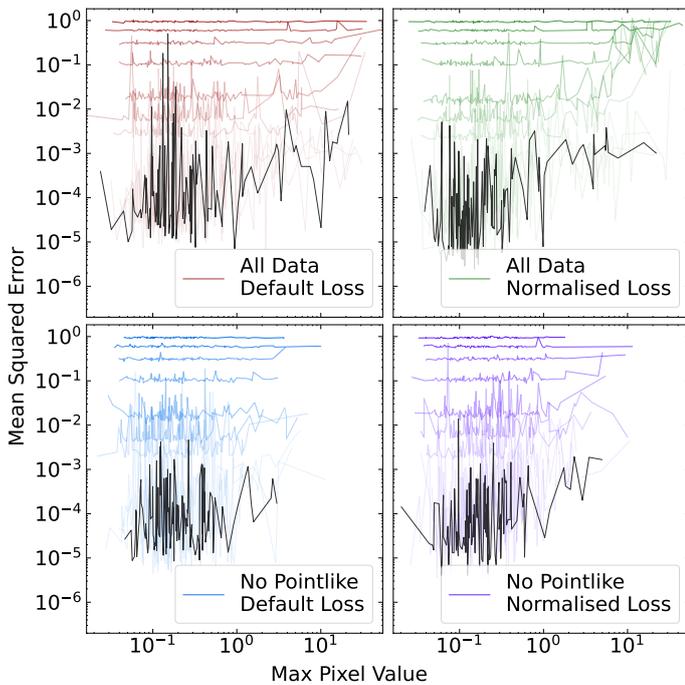
| Galaxy Overlap | Without asinh | With asinh |
|----------------|---------------|------------|
| AGN            | 0.75%         | 1.15%      |
| QSO            | 0.20%         | 7.82%      |
| Stars          | 0.17%         | 1.86%      |

### 3.4. 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, which is often to the benefit of visual inspections, reduces the differences between the brightest and faintest pixels. However, when attempting to differentiate galaxy and non-galaxy sources, the asinh function becomes a hindrance, causing the distributions to more significantly overlap, as shown in [Fig. 4](#). To generate a measure for class overlap we measure the ROC for each image, taking the mean of each non-galaxy class. We use the percentage of our true galaxy sample that have 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 2.5 and in the extreme case, nearly 50 times more galaxies overlap with stars. In the following sections of the paper we will discuss the set up, training and performance of the diffusion model that automatically learn to reconstruct the central emission expected for the normal galaxy population, and thus by contrast allow us to identify stars and AGN.

## 4. 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 makes use of the whole image and learns to predict what noise to remove at each timestep. This allows for the generation of novel examples from the trained distribution.



**Fig. 5.** 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 exhibit different behaviour as training progresses, indicating the respective weighting each put 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.

#### 4.1. Training the diffusion model

For training, we use  $64 \times 64$  pixel VIS cutouts for each of the selected sources discussed in Sect. 3.1. We use the training pipeline described in Dhariwal & Nichol (2021)<sup>1</sup> with the loss function

$$L_{\text{hybrid}} = L_{\text{MSE}} + \lambda L_{\text{VLB}}. \quad (6)$$

The hybrid loss is a combination of the mean squared error (MSE) and the variational lower bound (VLB). Using MSE ensures the output pixels of the model are numerically similar to the original image, maintaining structural and visual consistency within the generated images.

The first term in 6 is

$$L_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (\epsilon_i - \hat{\epsilon}_i)^2, \quad (7)$$

where  $\epsilon$  and  $\hat{\epsilon}$  are the true and predicted noise added for a timestep, respectively.

The VLB loss, shown in Eqs. (8) to (10), measures how well the diffusion process has modelled the underlying distribution of the image data, ensuring that the image as a whole is coherent and plausible for the data. The VLB loss has a weighting  $\lambda$  applied so as to not overpower the MSE loss:

$$L_{\text{VLB}} = L_0 + L_1 + \dots + L_{T-1} + L_T, \quad (8)$$

$$L_{t-1} = D_{\text{KL}} [q(x_{t-1} | x_t, x_0) \| p(x_{t-1} | x_t)], \quad (9)$$

<sup>1</sup> Available at <https://github.com/openai/guided-diffusion>

$$L_T = D_{\text{KL}} [q(x_T | x_0) \| p(x_T)]. \quad (10)$$

By utilising the Kullback-Leibler (KL) divergence (Kullback & Leibler 1951), a measure of the difference of two distributions, the loss acts as a regulariser to prevent simple memorisation and instead optimising for learning general features and patterns present in real data.

By combining both losses, the model becomes accurate at generating realistic data at both the micro (pixel-wise) and macro (image-wise) levels. This is vital in ensuring the model learns the dynamics of morphology images through the diffusion process, allowing it to effectively recreate essential features through re-painting. Throughout the rest of the paper the loss presented in Eq. (6) is referred to as the default loss.

#### Alternative loss

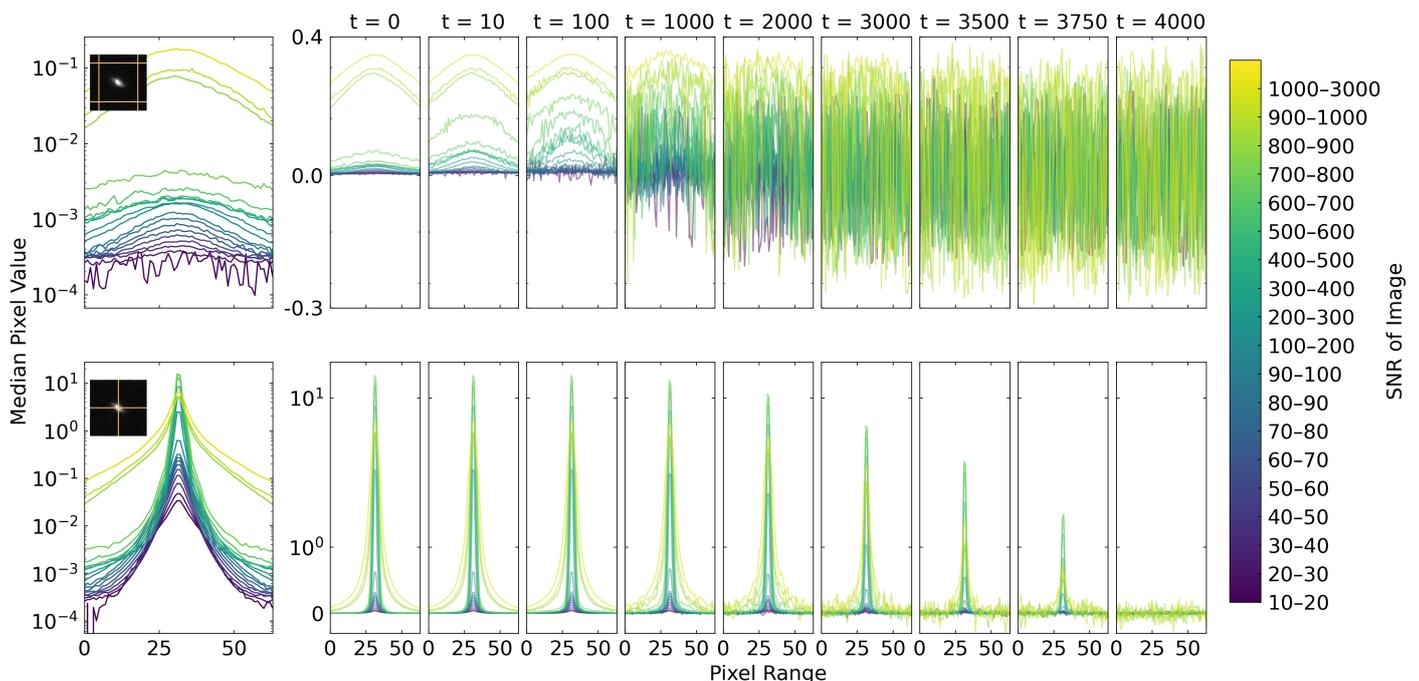
Making use of standard metrics whose primary use-case is data on a consistent scale, such as 0–255 pixels, may not function optimally with the high dynamic range of VIS data. Specifically, since MSE is not scale invariant, having images that feature pixels that are an order of magnitude brighter, results 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.

To address this, we applied a normalisation to  $L_{\text{MSE}}$  by dividing each image’s error by its respective maximum pixel value, which could be from a brighter external source. This allows for the errors of both bright and faint images to be scaled more equitably. The output of this loss is still combined with errors calculated in  $L_{\text{VLB}}$ . Throughout the remainder of the paper, we refer to this variant of the loss as the normalised loss.

To visualise the impact of amending the loss function, with respect to an images brightness, Fig. 5 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. 3.1, is also explored.

An initial observation that can be made with this representation is the distribution of the number of images of a particular brightness. By focusing solely on the black lines, where the impact is most noticeable, each vertex on the line represents an individual image. The more frequent updates the line has within a range of maximum pixel values, the more images in our data set exist within that region. As the line has more updates when the max pixel value is less than one, there are significantly more images with fainter pixels. This further highlights the disproportionate penalisation of MSE as it assigns the majority of the error to a minority of the data.

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 data set with the default loss (top left of Fig. 5), 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



**Fig. 6.** 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 left most plots. The top row shows how the additional noise impact 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 of introduced noise. Images with higher S/N are able to retain more of the original source detail for more timesteps in the scheduler.

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 data set (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 data set 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. After analysing the results from each of the four models, which will be discussed in Sect. 6.2, we chose to use the model trained on the full data set using the normalised loss.

#### 4.2. 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 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 for suspected AGN to be offset from the centre. To allow for

some deviation, a  $9 \times 9$  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 dynamically throughout the data loading pipeline.

Throughout this paper we use a  $5 \times 5$  pixel mask, unless explicitly stated otherwise, centred on the calculated brightest pixel. A  $5 \times 5$  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. 2024), the AGN’s light is spread over multiple pixels. Taking into account the redshift of the galaxy, 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 as the host galaxy information as possible for inpainting, we also reduce the risk of introducing artificial features into the reconstructed image.

#### 4.3. Reconstruction rescaling

An important difference with the processing of data in this work compared to other morphology-based papers is how the data input into the network are the raw values direct from the sensor, rather than images binned to the standard 0–255 pixel scale range. 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 setup is that the libraries used for training are optimised for images within the standard pixel range. Although this did not prevent the model from effective learning the distribution, it does require a rescaling of the gen-

erated output as the range is restricted to 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.

Instead, to allow the comparison of the original and outputted pixel values, we make use of the fact that 99.5% of the image (4071 of 4096 pixels when using a  $5 \times 5$  pixel mask) are shared between the input and output. The following scaling can be dynamically applied using each images respective values to allow the non-masked pixels 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.

## 5. Background noise versus diffusion noise

It is important for clarification to address 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 e.g. 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 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. It is unclear whether using the background noise as the pure noise sample would cause issues for the model. Given the aim of the diffusion process is to denoise a noisy image so that a realistic one emerges, the original noise should be largely, if not entirely, removed. Given the background noise constitutes any area that is not significantly affected by a source, the model may attempt to remove it which would create an unrealistic image. On the other hand, it may result in less pixels needing to be denoised, optimising the models ability at learning and processing the source features. Such an experiment is left for future exploration.

### 5.1. The effect of signal-to-noise

In machine-learning research, the improvements of generative models are often demonstrated with data sets that are of high resolution and typically noise free. This provides a level of consistency across the full data set, ensuring the benchmarks reflect the model’s performance under ideal conditions. Whether through loss functions (as discussed in Sect. 4) or in the practical application of the pipelines, how the model behaves on one part of the data will be similar to another.

However, the difficulty in learning the dynamics of our data are not consistent throughout the images. Higher signal-to-noise ratios 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 noise at different timesteps, the varying S/N of the images could again cause inconsistencies during training.

Figure 6 explores how this variability in data impacts the scheduled noise added to the data. By measuring the specific pixels that are dominantly background noise (top) and those that are significantly affect by the source (bottom), and further splitting images into S/N bins, we show the transformation from raw pixels to Gaussian noise across the entire scale of our data.

Analysing the top row, we can see that low S/N images are more significantly affected by the introduction of noise. This also has an effect much earlier on in the scheduler. The background noise converges to the diffusion’s pure Gaussian distribution, across all S/N bins, at between  $1/2$  ( $t = 2000$ ) and  $3/4$  ( $t = 3000$ ) of the total noise steps. When focusing on the pixels dominated by the source (bottom row), high S/N images are fairly unaffected for large parts of the schedule, with the brightest images not changing significantly until half way ( $t = 2000$ ). It is not until the scheduler is within the final 250 steps that the brightest images converge to the Gaussian distribution.

To visualise the true effect of this differing behaviour across S/N, Fig. 7 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.

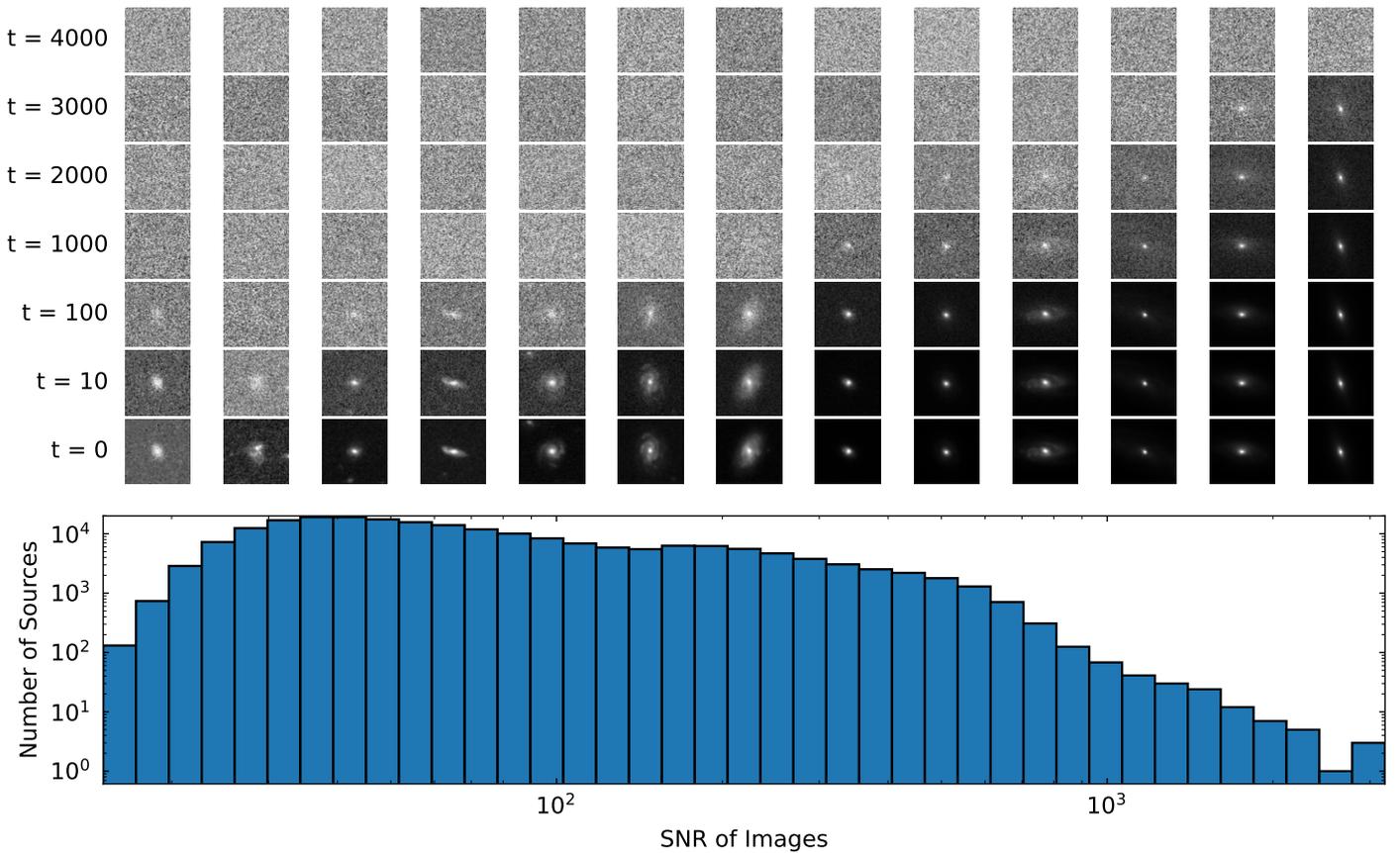
To increase performance of a diffusion model, whether to recreate a data set more accurately or increase the complexity of the data you are modelling, the simplest way is to increase the number of unique denoising steps performed. The important part here is the requirement to be unique, as simply adding more timesteps often results in the additional timesteps (especially in the latter stages) looking too similar and therefore become just additional overhead. This is the exact reason Nichol & Dhariwal (2021) implemented the cosine-beta scheduler, discussed Sect. 2.3, the original linear scheduler produced images that were pure noise too early into the process.

The images in Fig. 7 show that there are further optimisations that could be made to improve the scheduler for our data set. For optimal performance, one should expect the entire grid of images to behave like those in final 2 or 3 columns. A scheduler that adapts to the specific S/N of the image would be required and is left for future work. As our work is specifically using the reconstruction loss of inpainting as a classifier, the current scheduler is sufficient for our science goals.

## 6. Results

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 8 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



**Fig. 7.** *Top:* Respective noised images produced by the cosine-beta schedule at different timesteps. Each image is a sample from the respective signal-to-noise 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 forward process. The relationship of S/N and 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 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.

columns, we can see that due to the large mask and fewer pixels to guide the reconstruction, some of the images produce a visible boundary around the image where the combination of fixed pixels and generated pixels are not as seamless as the other images. Given the masks used for AGN detection are significantly smaller than those presented in Fig. 8, the fraction of pixels that the inpainting is conditioned on will be much higher, providing tighter constraints on the output compared to these examples.

### 6.1. Inpainting of AGN

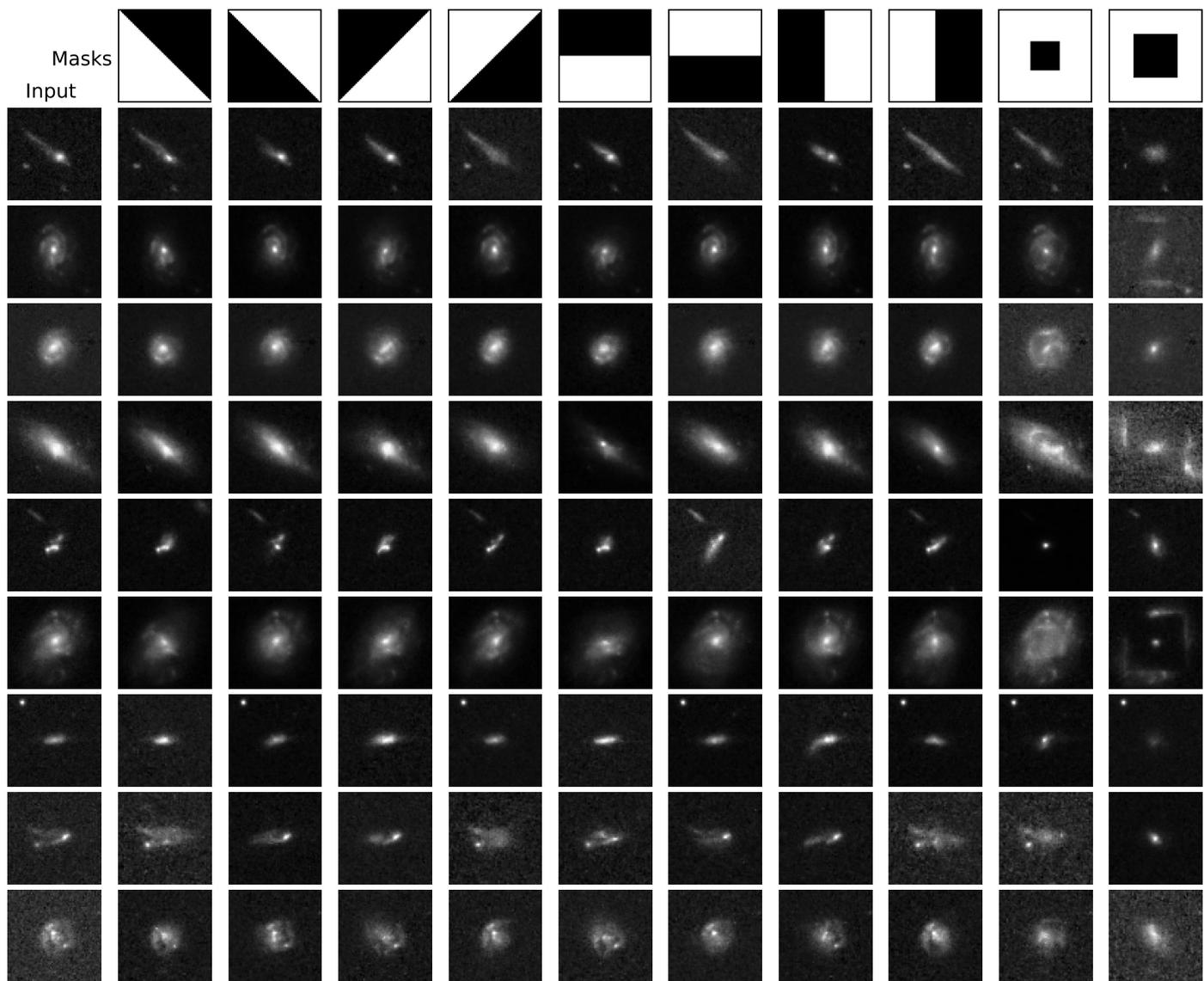
For our inpainting, we use a  $5 \times 5$  pixel mask centred on the brightest pixel, following the constraints described in Sect. 4.2. Comparing the expected pixel values with the generated outputs, Fig. 9 shows the model's performance. The assumption that the model will predict a fainter core than the original does appear to be true, with a near global bias for underpredicting as 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 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. 9, 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 fall 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 there is not a specific convergence point that 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, there are many more images that have a network output close to one that are 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 source 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 output of the model and show it is producing realistic outputs, Fig. A.1 shows how the ROC metric compares between input and output images. The significant overlap, especially with respect to peak of each distribution, shows that the model is accurately recreating the dynamics of the data.

To analyse how the error of each subset of class candidates have performed, comparisons are made between the original pixels and inpainted reconstruction for each image. The MSE of the pixel values are shown in Fig. 10. Each subset is binned ac-



**Fig. 8.** Applying the inpainting method to a selection of images with example masks (top row). The input image (leftmost column) has the pixels in 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 showcase the model’s ability at creating realistic galaxies. Ensuring that the generated images preserve a shape and brightness that is consistent with the fixed pixels shows that the model has a good internal representation of the dynamics of galaxy structure.

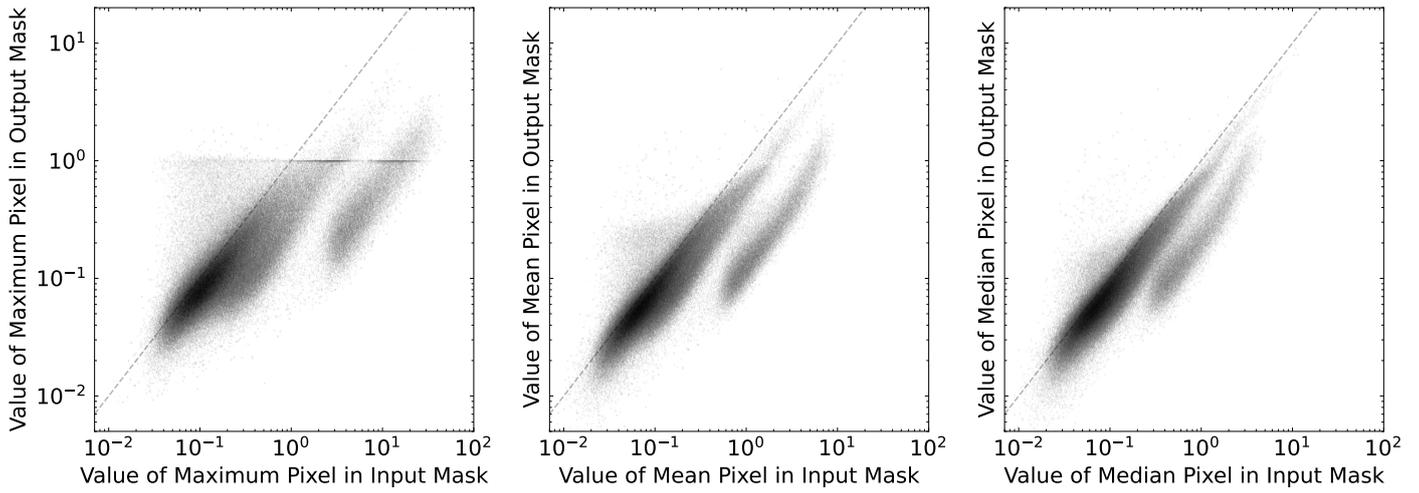
according to its  $I_E$  magnitude, providing a distinction between the relative brightness of each image.

There is a clear distribution difference between the respective class candidates. Stars perform the most consistently across magnitudes, producing the highest MSEs. There is an inherent trend for how high the error of the predictions are with the relative brightness of the classes, with stars and the brightest QSOs featuring the highest errors across all magnitudes. This is then followed by QSOs at higher magnitudes, and the lowest magnitude AGN candidates. Finally, the galaxy sample, along with the faintest AGN candidates produce much lower errors. Since these classes are candidates, it could also highlight the presence of contamination within the AGN labels. However, the large number of DESI-detected galaxies showing vastly reduced errors implies that the model does develop a bias for recreating galaxies with little or no AGN component.

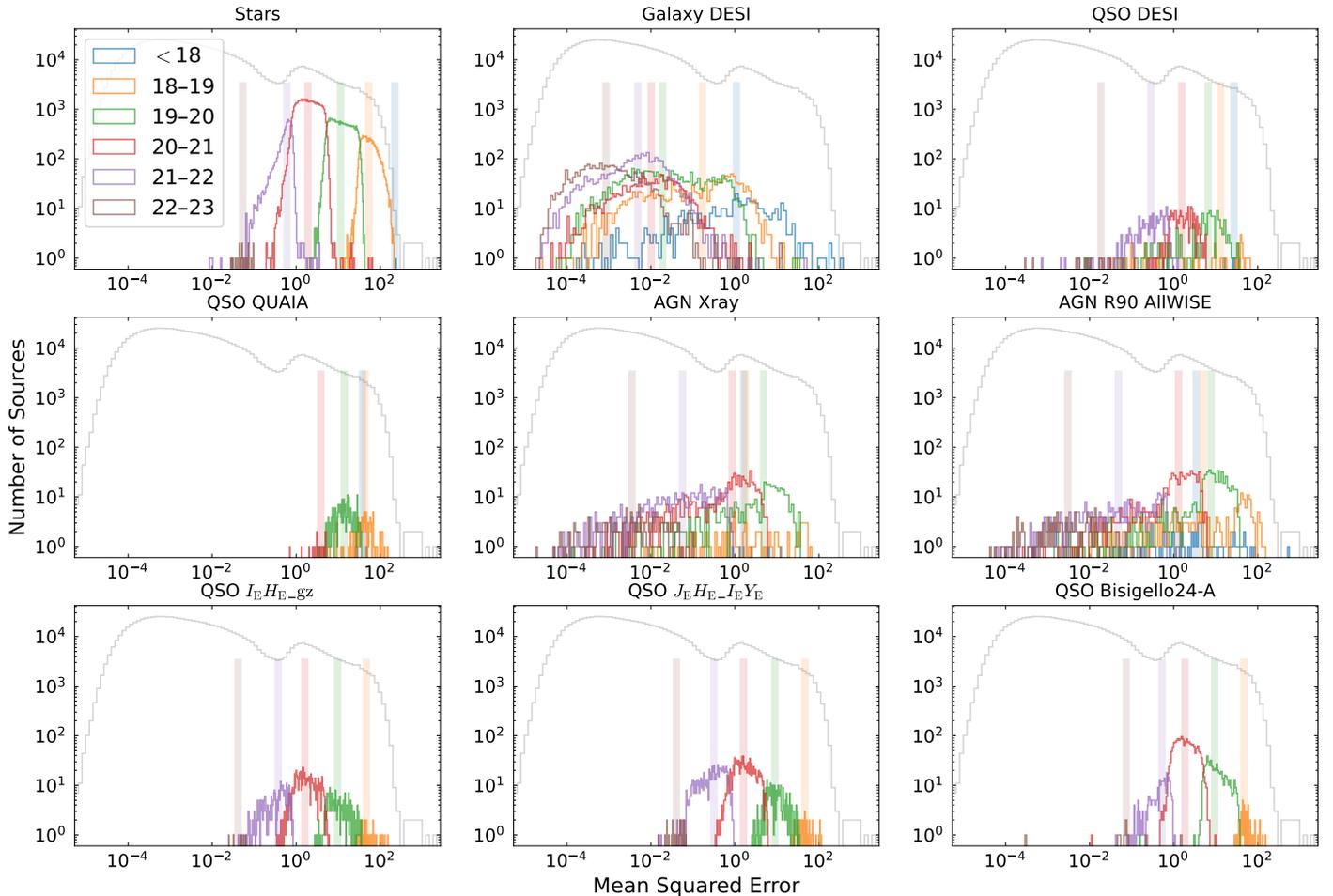
The second important finding is the trend between VIS  $I_E$  magnitude and error rate across all classes. The brighter a source

is, the higher its recorded MSE. This matches the initial assumptions of the method, where, as the sources increase in magnitude, differences in brightness values across galaxies become harder to differentiate. This leads to fainter potential AGN candidates looking more similar to a normal galaxy, causing the inpainting to reconstruct more accurately and therefore less likely to be included in our selection. However, the relationship between brightness and error is once again more influenced by the scale-variant nature of MSE. However, the lack of significant overlap for many of the selections with the galaxy class still make this metric useful.

To showcase how the performance of the inpainting method impacts the overall image, Fig. 11 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 to see how bright central cores can be entirely removed with inpainting, being replaced by regions that are consistent with the surrounding pixels. The model is also able to recreate images that do not



**Fig. 9.** Comparison between the original masked pixels and the pixels of the generated output. The model 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, implying an inherent difference in the input image causing the model to behave different. This collection of sources could be prime AGN candidates using our model.

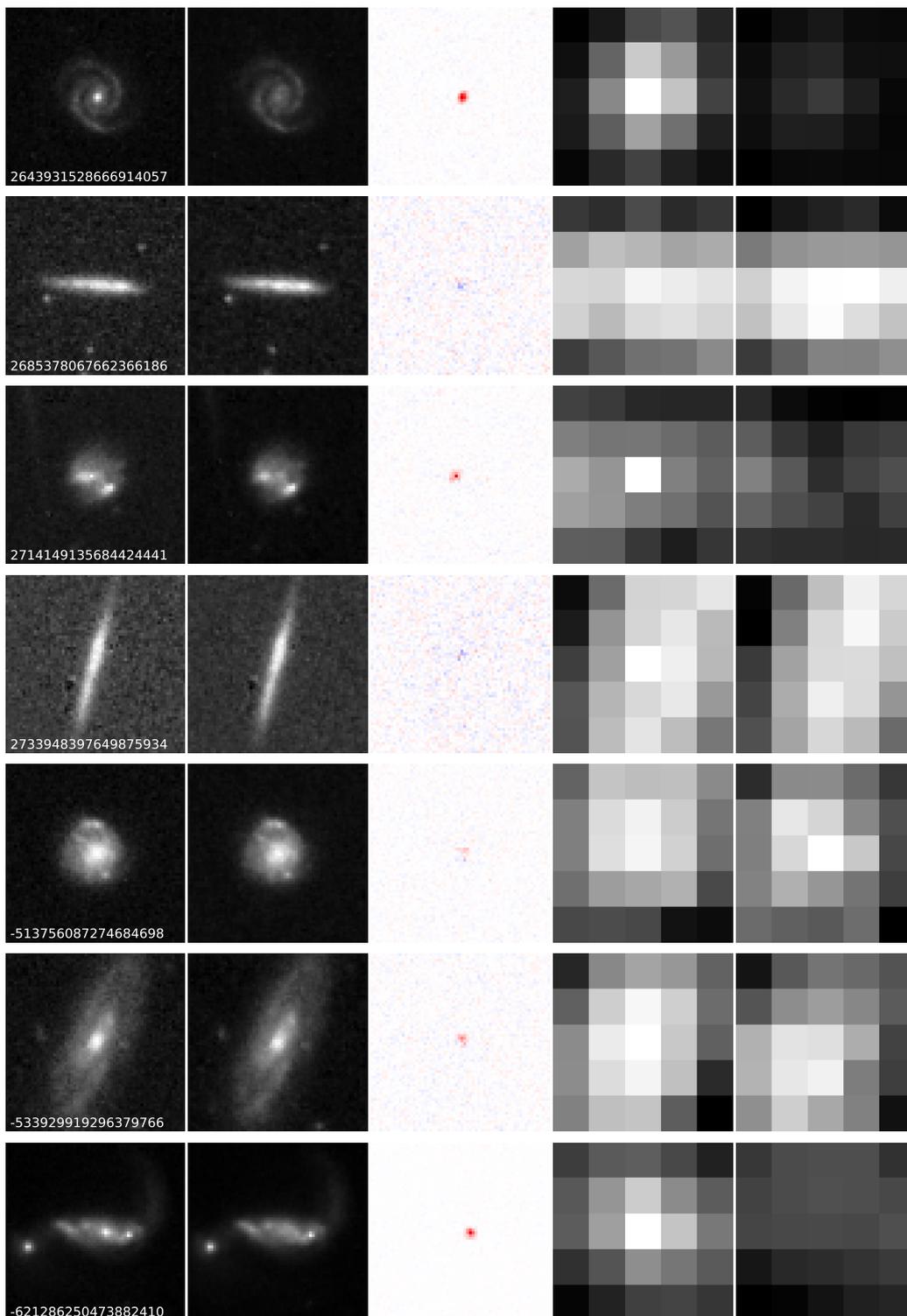


**Fig. 10.** MSE of the  $5 \times 5$  pixel mask from the input and generated output. AGN selection show a skew toward higher errors, especially for lower magnitude sources. The grey histogram shows the distribution of the whole data set, showing how the errors of images not captured in these selections compare. The median value for each  $I_E$  magnitude bin is shown in the respective vertical line.

feature any significant core accurately. This shows the model has a good internal representation of galaxies and how they should behave.

## 6.2. Max pixel difference ratio

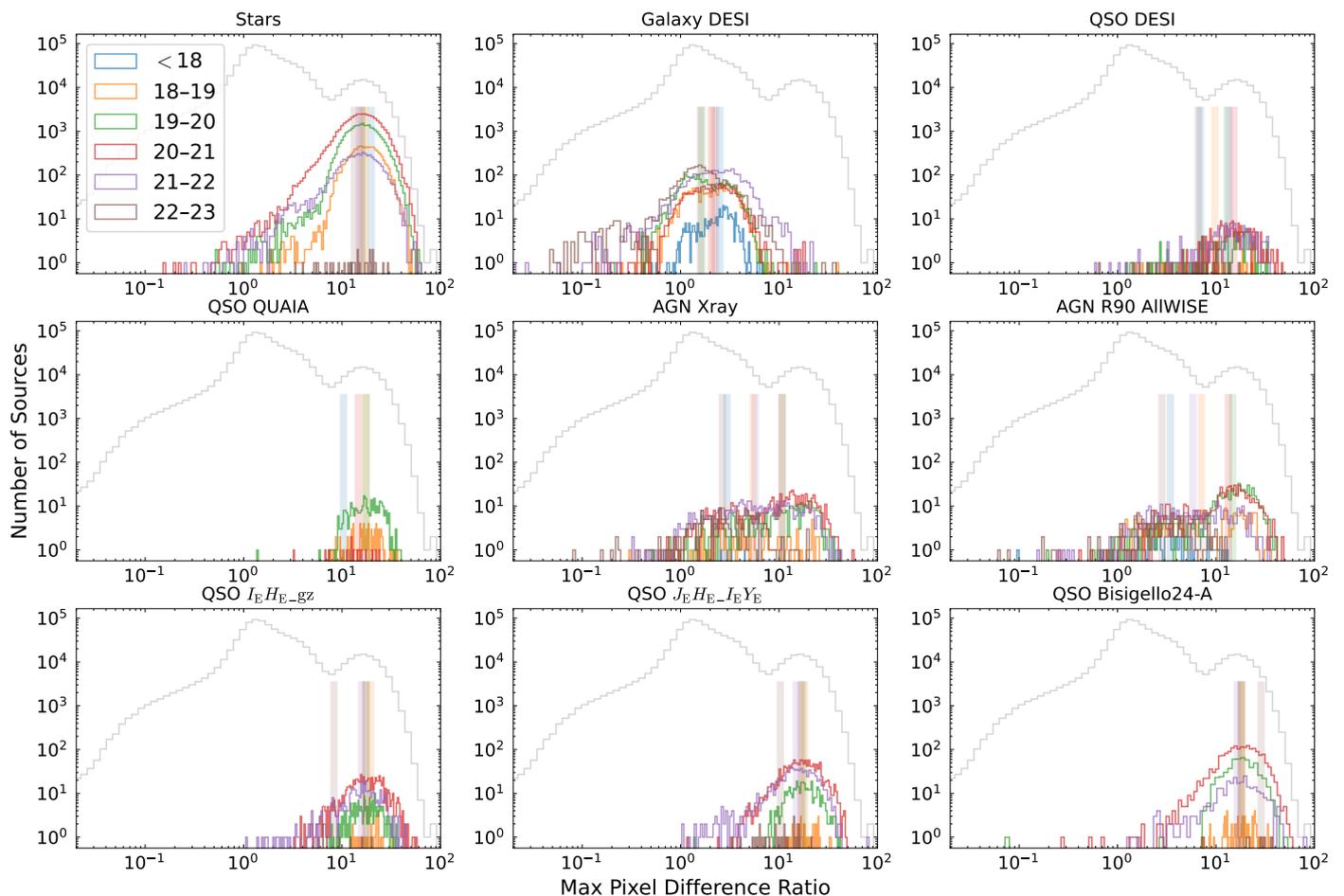
Although reconstruction error allows stars and a subset of AGN to stand out from the rest, a substantial overlap between sus-



**Fig. 11.** Example outputs from the repainting model. A 5x5 pixel mask, centred on the brightest pixel within the central region of the input image is masked and 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 64x64 cutout size, while the last two columns are the 5x5 masked region.

pected 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. As an example, if the reconstructed image ended up being the correct values

but just shifted by one pixel, this would be penalised heavily under MSE. From the perspective of this paper, such a scenario would imply that the model knew a bright pixel was valid and therefore, should not be penalised. Another example would be if only the brightest pixel is different in the reconstruction. MSE



**Fig. 12.** Measured ratio between the maximum pixel within the masked pixels of the input image, with the max of the pixels in the output. The galaxy sample is the only sample not to feature a right skew and therefore allows a threshold between this class and the rest. The brightness differences caused by the varying magnitudes do not significantly differ the ratios within each class selection. The grey histogram shows the distribution of the whole data set, showing how the ratio of images not captured in these selections compare. The median value for each  $I_E$  magnitude bin is shown in the respective vertical line.

would significantly reduce the impact of this measurement, potentially scoring similar to a source that gave small errors in every pixel within the mask.

To ensure such cases are not overlooked, we measure the pixel value ratio between the brightest pixel from the input mask and the brightest pixel from the reconstruction. As shown in Fig. 12, 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 produce a ratio less than 10, whereas a large population of suspected AGN from each selection feature a larger ratio.

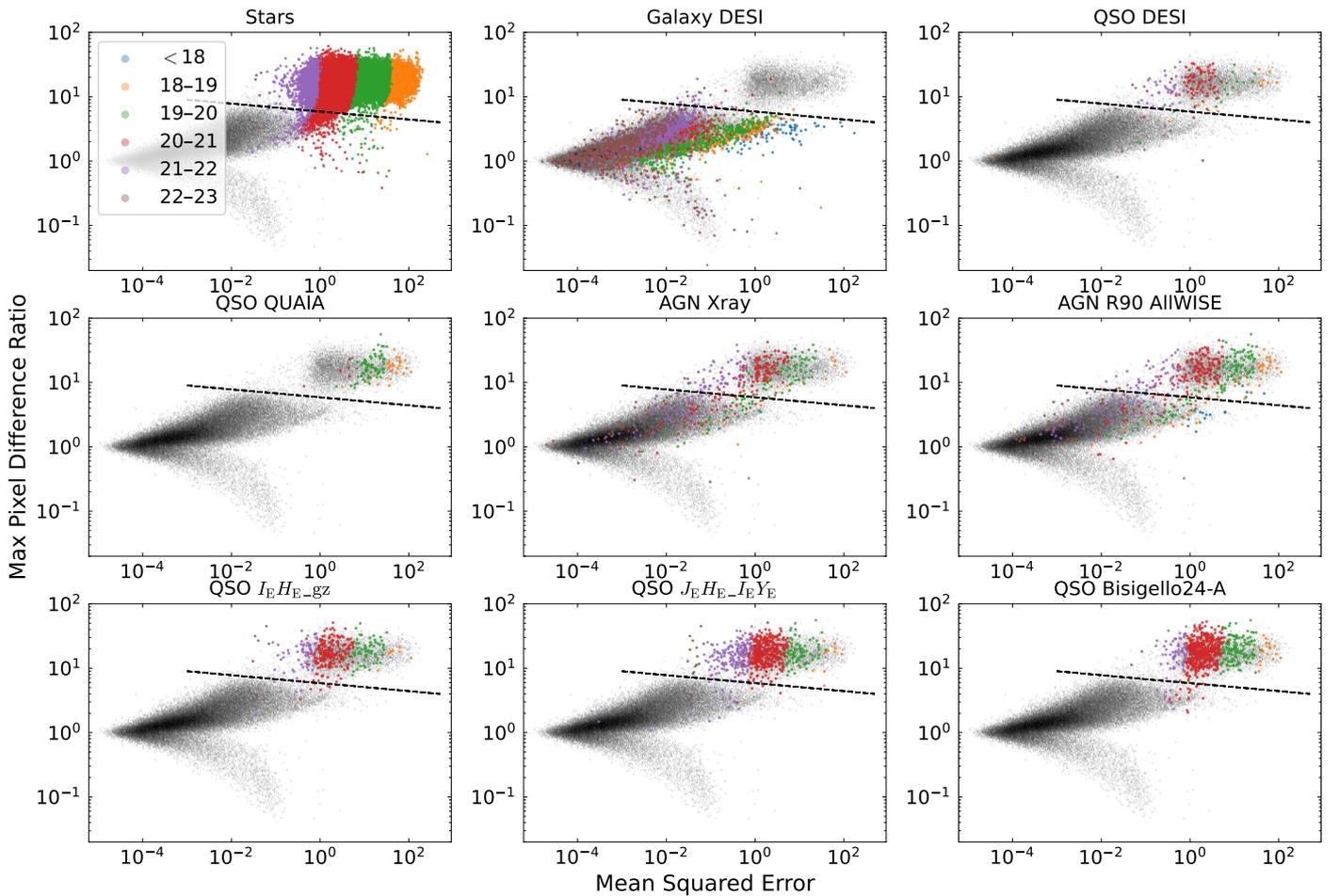
Another important difference between this metric and MSE is how the results are relative, reducing the impact of magnitude on the recorded values. Many of the high density peaks remain consistent, even as the magnitude of the sources increase. This implies that this metric is more robust at detecting fainter AGN than 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 where the reconstruction produces a cutout that is significantly fainter than the original image. As can be seen in Fig. B.1, this space is domi-

nated by point-like objects. Having the sharp falloff of pixel intensity in such a small area leads to the mask covering the majority, if not the entirety of the source. When the model attempts to reconstruct the missing pixels, it generates values that are coherent with the surrounding pixels. In these cases, the best match for empty space being more empty space.

The behaviour of the previous examples match with the expected behaviour of the model and the initial assumptions of the paper. 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, there exists a subset of sources that exhibit the opposite behaviour and generate a much brighter collection of pixels. Figure B.2 shows examples of such sources. When analysing the sources that exhibit this behaviour we find that they are overwhelmingly low S/N images. Given the difference in behaviour of these images during training, as shown in Fig. 7, it is clear this has had an impact on the model’s behaviour with these types of images.

Because we would like to limit this behaviour, we analyse how many of the unseen sources produce this type of output. The number of sources whose max pixel difference ratio is less than one for each of the four models discussed in Sect. 4, are shown in Table 2. The normalised loss helps significantly in reducing the number of sources exhibiting this behaviour. The normalised



**Fig. 13.** 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 data set. The majority of the sources within each of the non-galaxy classes are captured with this simple linear boundary, allowing for high recall across all classes. The sources not captured are typically those of higher VIS magnitude and therefore fainter. The tail formed, where sources exhibit a maximum pixel difference ratio  $< 10^0$ , show outlier sources which have predictions that are brighter after inpainting.

**Table 2.** The number of sources that produced brighter pixels in the masked area after repainting. The normalised loss produces a clear advantage in reducing the number of erroneous results.

| Data set      | 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%) |

loss trained on the full data set was selected as our applied model due to it further reducing the number of outlier predictions.

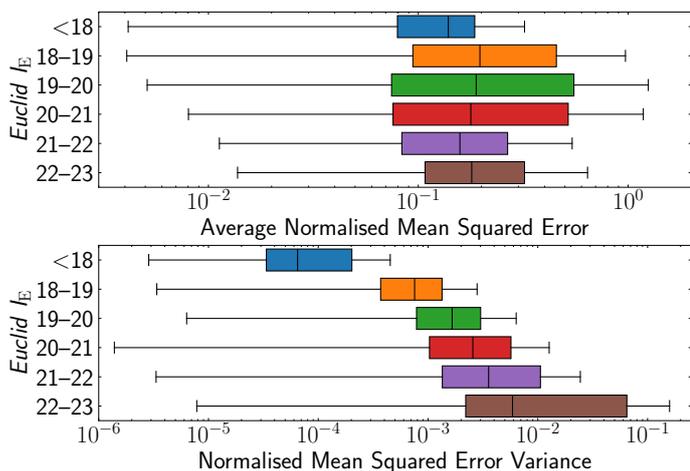
### 6.3. Morphology-based performance

Given the variety of shapes and sizes of objects that will be witnessed using *Euclid*, it is important to verify that the model is able to handle the complexities of different galaxy morphologies. Utilising the refined morphology classifications provided from *Euclid* Collaboration: Walmsley et al. (2025), we provide examples of edge-on (`disk-edge-on_yes_fraction`  $> 0.5$ ), spiral (`has-spiral-arms_yes_fraction`  $> 0.7$ ), and mergers (`merging_merger_fraction`  $> 0.3$ ) in Figs. C.1, C.2, and C.3, respectively.

These images show how our diffusion model and the inpainting pipeline 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.

### 6.4. 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 seen in Fig. 13, 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 create a simple decision boundary between the two clusters of points, as shown in Fig. 13. Although a more fine-tuned boundary would certainly be possible, especially when incorporating a further 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  $y = -0.01x + 9$ .



**Fig. 14.** Variations in the output error from repeat samples of the same fixed pixels. Over seven separate inpainting runs, the model shows 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 providing more robust inpainting than those of fainter images.

### 6.5. 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 data set. 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.

We perform the inpainting pipeline seven separate times on the same subset of the data. 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. 14.

The top figure shows that irrespective of magnitude, the model achieves 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 made evident in the bottom figure. The consistency of generated outputs vary drastically as fainter (and therefore lower S/N) images produce widely different pixels when resampled.

### 6.6. Comparing to other criteria

It is important to compare our AGN and QSO predictions to more traditional methods that are also able to utilise only the image data. Given the assumption of our model that sources containing an AGN will feature a bright central component with a sharp falloff in the galaxy’s light profile, comparing to Sérsic flux and point spread function (PSF) flux measurements is crucial because they capture the key concepts of our assumptions. The Sérsic flux helps quantify galaxy light distribution and differentiate an AGN’s central brightness from the galaxy’s structural brightness. PSF flux isolates the galaxy’s core, capturing point-like behaviour. The distribution of sources with these parameters is shown in Fig. 15. The equation of this linear boundary is  $y = 2x - 1$ .

Our second comparison is with the original rate of change metric used in Sect. 3.3. As we were able to see a clear distinction between the galaxy class and the others, it is important to verify how such a metric performs as a classifier. This is especially true given such a metric can be generated from the raw image, with no machine-learning model required. The boundary for this classifier is  $x = 0.75$ .

The test set scores for each respective selection is shown in Fig. 16, with the collated results for stars, galaxies, AGN and QSOs shown in Fig. D.1. All predictions are on images that were not used during training, nor used in the creation of any of the decision boundaries for the classifiers.

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. High recall ensures that our method is capable of identifying the broad range of AGN and QSO sources that are detectable through existing techniques. A high recall metric indicates that our classifier effectively captures the variety of sources each respective selection criterion is designed to detect.

While precision is also a consideration, our primary metric of success in this context is the diffusion model’s recall. Having the ability to recover a large fraction of known AGN and QSO types, across various selection methods, validates the robustness and comprehensiveness of our classifier. Due to differing constraints for each of the selections, we split the analysis to allow for a more representative comparison.

### Galaxies

Our galaxy selection is determined by the DESI spectral type classification SPECTYPE=GALAXY (DESI Collaboration et al. 2024), with AGN contaminants removed following the procedure detailed in Euclid Collaboration: Matamoro Zatarain et al. (2025), hereafter called MZ25. Negative instances are formed from the remaining spectroscopically determined labels SPECTYPE=STAR and SPECTYPE=QSO (DESI Collaboration et al. 2024) along with the contaminants removed from the original SPECTYPE=GALAXY selections.

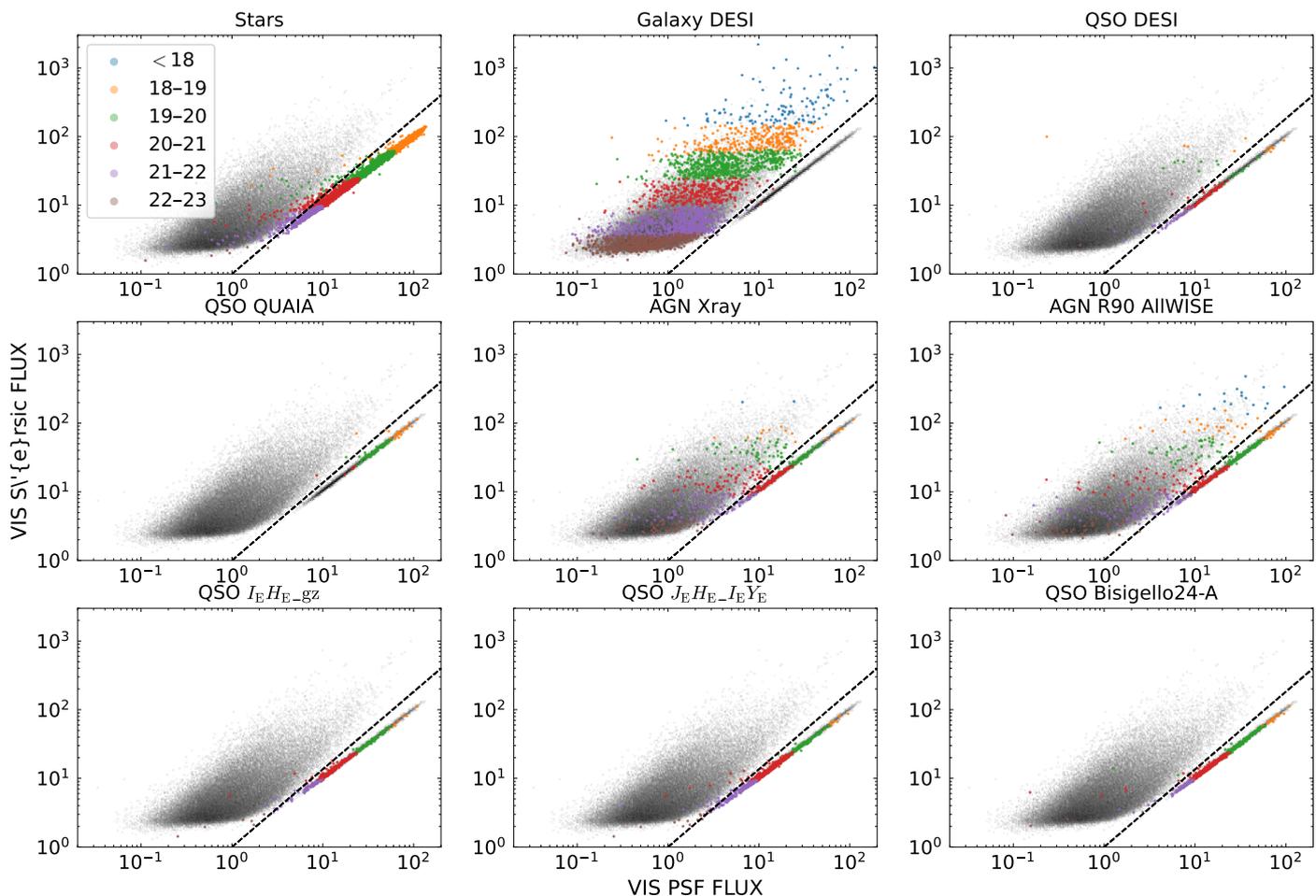
The predictions from each classifier for whether a source is a normal galaxy is simply the inverse of the boundary’s original predictions. As Fig. 16 shows, all classifiers capture nearly all galaxies across every magnitude bin. The precision scores also match extremely closely between classifiers, indicating similar contaminants proving difficult, especially for the brighter sources. Such contaminants can be seen in Figs. 13 and 15, where stars and DESI-selected QSO fall outside the boundaries.

The ability to separate normal galaxies from other extragalactic or stellar objects has benefits for 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 understand of dark matter and dark energy, a primary object of the *Euclid* mission (Euclid Collaboration: Mellier et al. 2024).

### Stars

The star selection is made up of the positive instances of *Euclid* selected (star\_candidate\_euclid, MZ25), DESI-selected (SPECTYPE=STAR, DESI Collaboration et al. 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



**Fig. 15.** 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 AGNs, with the majority being point-like objects. However, it is most apparent with selections from AllWISE R90 and X-ray-selected AGN that there are many sources that 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 data set.

capture stars well. This performance is beneficial when attempting to capture a high purity galaxy sample as discussed in the previous section. However, as other works such as [Euclid Collaboration: Matamoro Zatarain et al. \(2025\)](#) have shown, there are metrics to mitigate the contamination of stars. This is shown in practise by the high precision, highlighting how we are able to isolate this star selection from other point-like sources.

However, a significant number of stars are likely to remain in our sample. Therefore, for all subsequent selections, we apply the cut `phz_star_prob_phz_class < 0.3`, to predictions of all classifiers to minimise the contamination of stellar sources.

#### Euclid-detected QSOs

The *Euclid* photometric-based selections for QSOs include  $I_E H_{-g_z}$  and  $JH_{-I_E Y}$ , as described in [MZ25](#), and the Bisigello24-A 2-colour selection ([Euclid Collaboration: Bisigello et al. 2024](#)). This subset consider only sources with `MUMAX_MINUS_MAG`  $\leq -2.6$ , a parameter that indicates point-likeness ([Euclid Collaboration: Romelli et al. 2025](#)).

For the  $I_E H_{-g_z}$  and  $JH_{-I_E Y}$  selections, we capture around 90% of the candidate QSOs. This coverage remains consistent across the available magnitudes. The performance compared to

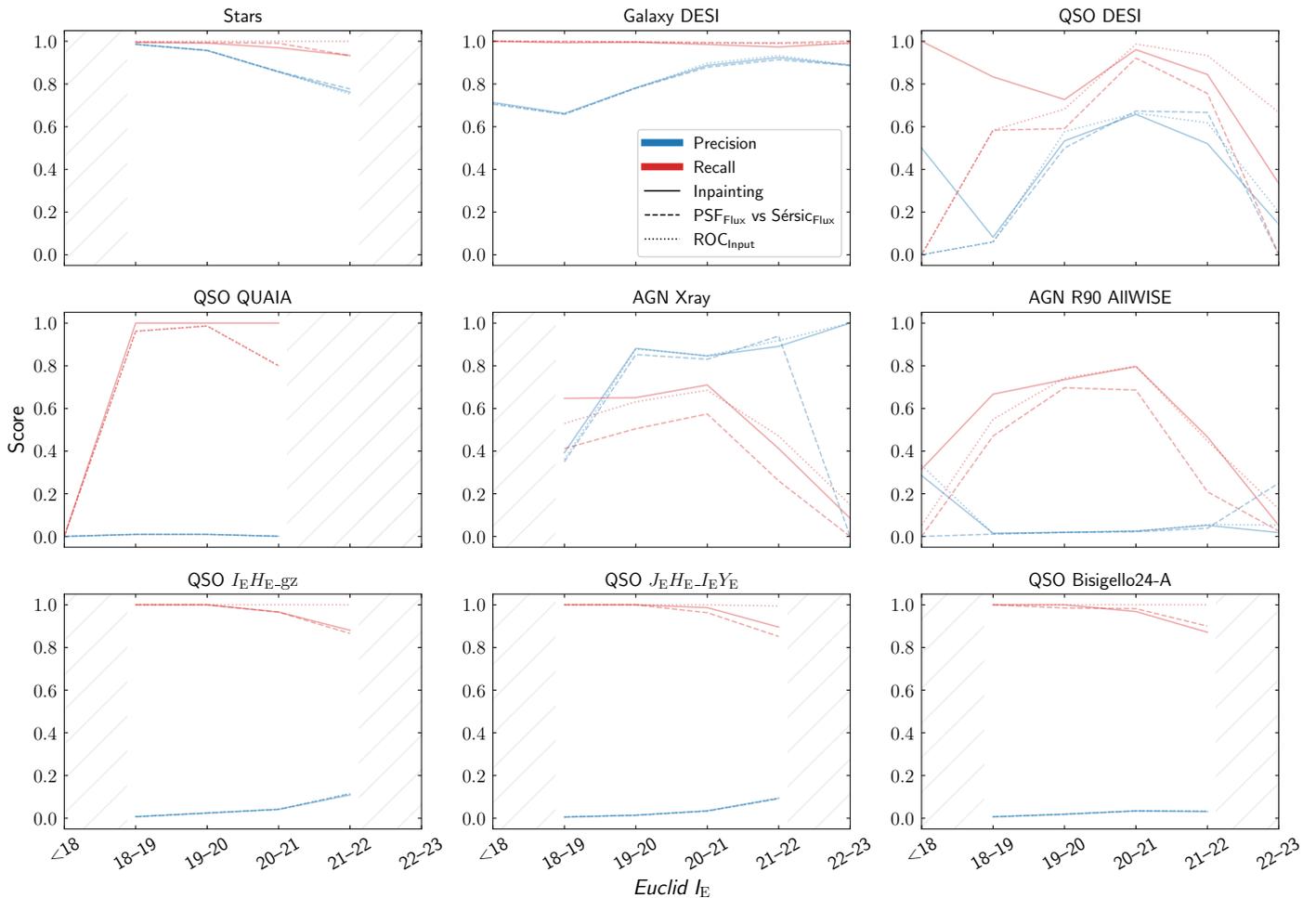
the other classifiers is similar, with ROC capturing a small percentage more sources, but not different enough to be significant.

The 2-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.

#### DESI-selected QSOs

As these sources are assumed 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 are found, we do not apply the restriction to this subset.

This selection provides the first evidence of significant differences within the candidates for each classifier. The inpainting classifier is able to capture a significant amount of the selection compared to the other classifiers. Throughout all magnitude bands, the PSF and Sérsic flux classifier has worse recall. Whereas the diffusion model has higher coverage over the brightest sources (lowest magnitudes), the ROC classifier matched more candidates in the fainter sources.



**Fig. 16.** 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.

### *Gaia*-detected QSOs

For the *Gaia*-detected QSOs, we use the Quiaia selection (Storey-Fisher et al. 2024). The subset consists of positive instances (qso\_quaia, MZ25), with negative instances from *Gaia*-detected galaxies (in\_galaxy\_candidates\_gaia, MZ25) and stars (star\_candidate\_gaia, MZ25). We once again apply the MUMAX\_MINUS\_MAG restriction to the sources.

After the initial magnitude bin, our classifier was able to match with a high proportion of the Quiaia-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 Quiaia candidates, selecting more than just the candidates is likely, significantly impacting the precision.

### 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 compare

our candidates to the 90% reliability (R90) selection which is 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.

### Xray-selected AGN

We compare our candidates to the X-ray selection from the work by Euclid Collaboration: Roster et al. (2025). The positive instances are when the calculated probability that the source is a galaxy is sufficiently low ( $\text{Gal\_proba\_roster} < 0.2$ , MZ25). The negative instances are sources with  $\text{Gal\_proba\_roster}$

**Table 3.** Recall, precision and area under the curve (AUC) scores for each of the classifiers. Each selection we compare to 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 (increase of at least 0.05). The PSF<sub>FLUX</sub> versus Sérsic<sub>FLUX</sub> results have been shortened to PSF–S.

| 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 |
|                         | $\geq 20$         | 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 |
|                         | $\geq 20$         | 0.98        | 1.00  | 0.99        | 0.90      | 0.90        | 0.91  | 0.73        | 0.72  | 0.75 |
| QSO DESI                | $< 20$            | <b>0.77</b> | 0.57  | 0.63        | 0.18      | 0.14        | 0.15  | <b>0.84</b> | 0.75  | 0.77 |
|                         | $\geq 20$         | 0.90        | 0.84  | <b>0.96</b> | 0.59      | <b>0.67</b> | 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 |
|                         | $\geq 20$         | <b>1.00</b> | 0.80  | 0.80        | <0.01     | <0.01       | <0.01 | <b>0.51</b> | 0.41  | 0.40 |
| AGN Xray                | $< 20$            | 0.64        | 0.48  | 0.61        | 0.74      | 0.73        | 0.74  | 0.63        | 0.59  | 0.62 |
|                         | $\geq 20$         | 0.51        | 0.38  | 0.53        | 0.86      | 0.86        | 0.87  | 0.53        | 0.51  | 0.55 |
| AGN R90 AllWISE         | $< 20$            | <b>0.68</b> | 0.57  | 0.63        | 0.02      | 0.02        | 0.02  | 0.68        | 0.63  | 0.66 |
|                         | $\geq 20$         | 0.59        | 0.45  | 0.60        | 0.03      | 0.02        | 0.03  | 0.75        | 0.68  | 0.75 |
| QSO $I_E H_{\text{gz}}$ | $< 20$            | 1.00        | 1.00  | 1.00        | 0.02      | 0.02        | 0.02  | 0.50        | 0.50  | 0.50 |
|                         | $\geq 20$         | 0.94        | 0.92  | <b>1.00</b> | 0.05      | 0.05        | 0.05  | 0.49        | 0.47  | 0.50 |
| QSO $JH_{\text{I}} Y$   | $< 20$            | 1.00        | 1.00  | 1.00        | 0.01      | 0.01        | 0.01  | 0.50        | 0.50  | 0.50 |
|                         | $\geq 20$         | 0.95        | 0.91  | <b>1.00</b> | 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 |
|                         | $\geq 20$         | 0.96        | 0.97  | 1.00        | 0.03      | 0.03        | 0.03  | 0.50        | 0.50  | 0.50 |

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

The main difference with the results of the X-ray-detected AGN is how the precision outperforms the recall across nearly every magnitude bin. This is due to low number of X-ray sources compared to the other selections. 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 with 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 by 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. 16. 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<sub>FLUX</sub> versus Sérsic<sub>FLUX</sub> classifier.

### 6.7. 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 rep-

resent the AGN and galaxy components. Although utilising inpainting for decomposition is a worthwhile future research direction, the validity of the residuals to accurately represent the AGN component for further scientific analysis have not been tested and as such cannot be used. 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 on Euclid data, readers are referred to [Euclid Collaboration: Margalef-Bentabol et al. \(2025\)](#), hereafter called [MB25](#), whose method provides AGN fraction ( $F_{\text{AGN}}$ ) measurements that we use to analyse our AGN candidates.

The subset of Q1 used within this work and the subset tested in [MB25](#), 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 [MB25](#)) are shown in [Table 4](#). The majority of sources ( $\sim 95\%$ ) are in agreement by both methods not to be AGN candidates. With the remaining sources, only  $\sim 10\%$  are in agreement, with over five times more AGN candidates selected by [MB25](#). 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 [MB25](#) than the diffusion model ([Table 5](#)).

The left panel of [Fig. 17](#), shows the AGN candidates selected by the diffusion, but predicted to have low AGN fraction by the model in [MB25](#). Whereas the right panel shows the AGN fraction of agreed sources between the two methods. [Fig. D.2](#) shows where the [MB25](#) selections are located on the inpainting metrics. As shown in [Table 4](#), many of the [MB25](#) 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  $F_{\text{AGN}}$  selection is more sensitive to fainter AGN.

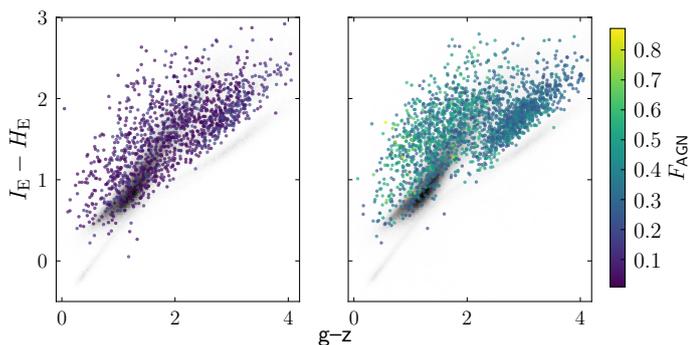
Comparing the predictions for each of the *Euclid*-based selections in [Fig. 18](#), we can see that our diffusion model covers

**Table 4.** AGN candidate comparison between this work and the AGN fractions predicted by MB25. There is an overlap of 425 781 sources between the subset of *Euclid* Q1 sources selected by both methods.

| Matched sources        | MB25 AGN | MB25 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 data sets. The majority of AGN candidates from the two respective data sets are sources not used within the other’s data set.

| Method               | Diffusion | MB25   |
|----------------------|-----------|--------|
| Total AGN Candidates | 16 053    | 57 874 |



**Fig. 17.** Diffusion model AGN candidates, colour-coded according to the AGN fraction of MB25. *Left:* not identified as AGN by MB25 ( $F_{\text{AGN}} < 0.2$ ), and *right:* AGN candidates selected by MB25 ( $F_{\text{AGN}} \geq 0.2$ ).

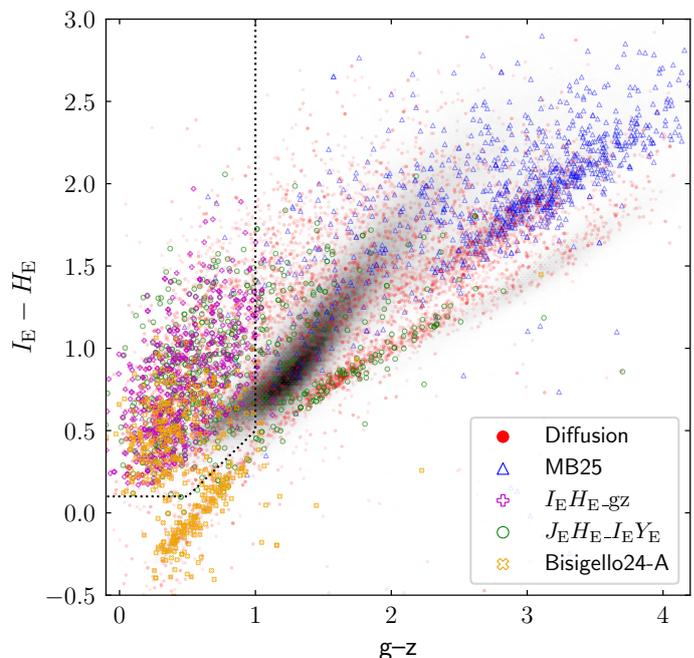
areas of the colour space that the photometry-based selections do not. Being able to select both bluer galaxies, as captured by the photometry selections (MZ25; Euclid Collaboration: Bisigello et al. 2024), and the much redder sources captured by MB25, shows the versatility of the method, highlighting its effectiveness across different galaxy populations.

## 7. Summary and conclusions

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 optically-selected AGN, without the requirement of 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 selections obtained through more traditional methods, across various wavelengths including optical, near-infrared (NIR), mid-infrared, and X-ray.

### 7.1. Considerations on 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 similar scale and architecture. How-



**Fig. 18.** Comparison of our diffusion model selections with those from photometry-based selections (MZ25; Euclid Collaboration: Bisigello et al. 2024) and the selections from MB25. Although all the diffusion selected sources are plotted, to improve clarity, only a representative subset of the other selections are shown. The photometry-based selections are limited to much bluer sources, highlighted by the dotted line showing the cut introduced in MZ25.

ever, the inference pipeline is heavily constrained by memory rather than computation, as 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. Whilst training, although more computationally expensive, has a more structured process where gradients are computed and accumulated in a predictable manner.

With an expanded compute resource (4× Nvidia H100 GPUs), we are able to inpaint just over 2.5 million images per day. As we have shown in Sects. 2.3 and 4, 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, will significantly reduce the number of iterations and inference time required to process each image. For example, Fig. 7 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. 6.6 show that although our predictions have a significant overlap with many of the selections presented, it also highlights that the ROC classifier produces similar overlap performance. Given this boundary is formed by a simple calculation on the data, rather than requiring a complex model, it may present itself as the preferred method. However, it must be reiterated that the current pipeline has had minimum optimisations for the data and the classification task. Therefore, the ROC classifier serves as a fundamental baseline for comparison in our study. Future enhancements to the 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 po-

tential for improvement, making the diffusion method the most promising moving forward.

Our key contributions and conclusions include the following

- Compared to traditional colour-selection criteria, the approach is attaining high coverage in AGN candidates while achieving near-perfect separation of galaxies not featuring an AGN component. Reliably separating these different sources allows for 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 the future of the mission will cover 14 000 deg<sup>2</sup>, the ability for AGN identification using only VIS allows for the possibility of some science aims to be met without the immediate need for integration of multiple data sources. This can significantly reduce time and complexity for 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 data sets. The ability to leverage models that can process large volumes of data efficiently, without costly and extensive specialised tuning, will be vital as scientific data sets 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 with astronomical imaging. We show how the assumptions of machine-learning literature can result in non-optimal training, especially when it comes to the noise scheduler and training losses. By highlighting these inefficiencies, we can see where progress can be made in both astronomy and machine-learning domains.
- 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, as well as 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 whilst aiding in reducing the reliance on complex multispectral 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.* 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](http://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](http://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/openphysics/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].

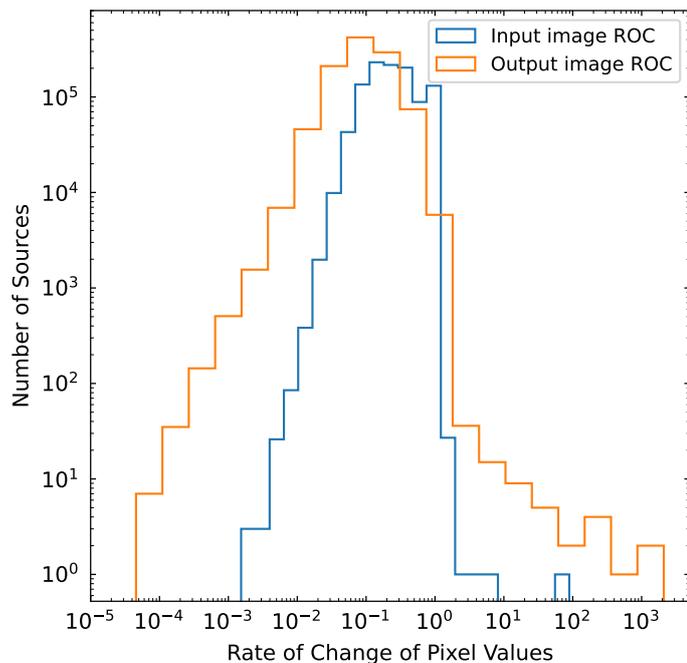
## References

- Adam, A., Coogan, A., Malkin, N., et al. 2022, arXiv preprint arXiv:2211.03812
- Adam, A., Stone, C., Bottrell, C., et al. 2023, arXiv preprint arXiv:2311.18002
- Assef, R. J., Stern, D., Noirot, G., et al. 2018, *ApJS*, 234, 23
- 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., Aguilar, J., et al. 2024, *AJ*, 168, 58
- Dhariwal, P. & Nichol, A. 2021, *Advances in neural information processing systems*, 34, 8780
- Dressler, A. 1989, in *IAU Symposium*, Vol. 134, *Active Galactic Nuclei*, ed. D. E. Osterbrock & J. S. Miller, 217
- Euclid* Collaboration: Bisigello, L., Massimo, M., Tortora, C., et al. 2024, *A&A*, 691, A1
- Euclid* Collaboration: Cropper, M., Al Bahllawan, A., Amiaux, J., et al. 2024, *A&A*, accepted, arXiv:2405.13492
- Euclid* Collaboration: Margalef-Bentabol, B., Wang, L., La Marca, A., et al. 2025, *A&A*, submitted
- Euclid* Collaboration: Matamoros Zatarain, T., Fotopoulou, S., Ricci, F., et al. 2025, *A&A*, submitted
- Euclid* Collaboration: Mellier, Y., Abdurro'uf, Acevedo Barroso, J., et al. 2024, *A&A*, accepted, arXiv:2405.13491
- Euclid* Collaboration: Romelli, E., Kümmel, M., Dole, H., et al. 2025, *A&A*, submitted
- Euclid* Collaboration: Roster, W., Salvato, M., Buchner, J., et al. 2025, *A&A*, submitted
- Euclid* Collaboration: Walmsley, M., Huertas-Company, M., Quilley, L., et al. 2025, *A&A*, submitted

- Euclid Quick Release Q1. 2025, <https://doi.org/10.57780/esa-2853f3b>
- Ferrarese, L. & Merritt, D. 2000, *ApJ*, 539, L9
- Gaia Collaboration: Bailer-Jones, C. A. L., Teyssier, D., Delchambre, L., et al. 2023, *A&A*, 674, A41
- Gaia Collaboration: Vallenari, A., Brown, A. G. A., Prusti, T., et al. 2023, *A&A*, 674, A1
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, *Advances in neural information processing systems*, 27
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2020, *Communications of the ACM*, 63, 139
- Harrison, C. M. & Ramos Almeida, C. 2024, *Galaxies*, 12, 17
- Ho, J., Jain, A., & Abbeel, P. 2020, *Advances in neural information processing systems*, 33, 6840
- Karras, T., Laine, S., Aittala, M., et al. 2020, in *IEEE/CVF conference on computer vision and pattern recognition (IEEE)*, 8110–8119
- Kingma, D. P. 2013, arXiv preprint arXiv:1312.6114
- Kormendy, J. & Richstone, D. 1995, *ARA&A*, 33, 581
- Kullback, S. & Leibler, R. A. 1951, *The annals of mathematical statistics*, 22, 79
- Lugmayr, A., Danelljan, M., Romero, A., et al. 2022, in *IEEE/CVF conference on computer vision and pattern recognition (IEEE)*, 11461–11471
- 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
- Nichol, A. Q. & Dhariwal, P. 2021, in *Proceedings of Machine Learning Research*, Vol. 139, *Proceedings of the 38th International Conference on Machine Learning*, ed. M. Meila & T. Zhang (PMLR), 8162–8171
- Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. 2016, in *IEEE conference on computer vision and pattern recognition (IEEE)*, 2536–2544
- Radford, A., Kim, J. W., Hallacy, C., et al. 2021, in *Proceedings of Machine Learning Research*, Vol. 139, *Proceedings of the 38th International Conference on Machine Learning*, ed. M. Meila & T. Zhang (PMLR), 8748–8763
- 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*, ed. S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh, Vol. 35 (Curran Associates, Inc.), 36479–36494
- Sahu, N., Graham, A. W., & Davis, B. L. 2019, *ApJ*, 876, 155
- Sampson, M. L., Melchior, P., Ward, C., & Birmingham, S. 2024, *Astronomy and Computing*, 49, 100875
- Sargsyan, A., Navasardyan, S., Xu, X., & Shi, H. 2023, in *IEEE/CVF International Conference on Computer Vision (IEEE)*, 7335–7345
- Smith, M. J., Geach, J. E., Jackson, R. A., et al. 2022, *MNRAS*, 511, 1808
- Spagnoletti, A., Boucaud, A., Huertas-Company, M., Kabalan, W., & Biswas, B. 2024, arXiv preprint arXiv:2411.19158
- 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, in *2020 international joint conference on neural networks (ijcnn)*, 1–10
- Wang, Y., Tao, X., Qi, X., Shen, X., & Jia, J. 2018, *Advances in neural information processing systems*, 31
- Yeh, R. A., Chen, C., Yian Lim, T., et al. 2017, in *IEEE conference on computer vision and pattern recognition (IEEE)*, 5485–5493
- Yu, J., Lin, Z., Yang, J., et al. 2018, in *IEEE conference on computer vision and pattern recognition (IEEE)*, 5505–5514
- Zhang, L., Rao, A., & Agrawala, M. 2023, in *IEEE/CVF International Conference on Computer Vision (IEEE)*, 3836–3847
- 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 Moiriarlo 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

- <sup>44</sup> Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
- <sup>45</sup> Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain
- <sup>46</sup> INFN section of Naples, Via Cinthia 6, 80126, Napoli, Italy
- <sup>47</sup> Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
- <sup>48</sup> Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- <sup>49</sup> Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
- <sup>50</sup> European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
- <sup>51</sup> Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, Villeurbanne, F-69100, France
- <sup>52</sup> Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
- <sup>53</sup> Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
- <sup>54</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Luísa Companys 23, 08010 Barcelona, Spain
- <sup>55</sup> UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 rue Enrico Fermi, 69622 Villeurbanne, France
- <sup>56</sup> Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
- <sup>57</sup> Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
- <sup>58</sup> Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- <sup>59</sup> Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290 Versoix, Switzerland
- <sup>60</sup> INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
- <sup>61</sup> Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
- <sup>62</sup> INAF-Istituto di Astrofisica e Planetologia Spaziali, via del Fosso del Cavaliere, 100, 00100 Roma, Italy
- <sup>63</sup> Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
- <sup>64</sup> INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
- <sup>65</sup> Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, 0315 Oslo, Norway
- <sup>66</sup> Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK
- <sup>67</sup> Felix Hormuth Engineering, Goethestr. 17, 69181 Leimen, Germany
- <sup>68</sup> Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
- <sup>69</sup> Cosmic Dawn Center (DAWN), Denmark
- <sup>70</sup> Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
- <sup>71</sup> NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
- <sup>72</sup> Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
- <sup>73</sup> Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109, USA
- <sup>74</sup> Department of Physics, P.O. Box 64, 00014 University of Helsinki, Finland
- <sup>75</sup> Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
- <sup>76</sup> Centre de Calcul de l'IN2P3/CNRS, 21 avenue Pierre de Coubertin 69627 Villeurbanne Cedex, France
- <sup>77</sup> Laboratoire d'étude de l'Univers et des phénomènes eXtremes, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 92190 Meudon, France
- <sup>78</sup> SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK
- <sup>79</sup> Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
- <sup>80</sup> INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
- <sup>81</sup> Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
- <sup>82</sup> INFN-Sezione di Roma, Piazzale Aldo Moro, 2 - c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
- <sup>83</sup> Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham, DH1 3LE, UK
- <sup>84</sup> Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
- <sup>85</sup> Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
- <sup>86</sup> CNRS-UCB International Research Laboratory, Centre Pierre Binetruy, IRL2007, CPB-IN2P3, Berkeley, USA
- <sup>87</sup> University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
- <sup>88</sup> Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
- <sup>89</sup> Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- <sup>90</sup> Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
- <sup>91</sup> DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark
- <sup>92</sup> Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
- <sup>93</sup> Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
- <sup>94</sup> Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
- <sup>95</sup> Centre National d'Etudes Spatiales – Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
- <sup>96</sup> Institute of Space Science, Str. Atomistilor, nr. 409 Măgurele, Ilfov, 077125, Romania
- <sup>97</sup> Consejo Superior de Investigaciones Científicas, Calle Serrano 117, 28006 Madrid, Spain
- <sup>98</sup> Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain
- <sup>99</sup> Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, Via Marzolo 8, 35131 Padova, Italy
- <sup>100</sup> Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile
- <sup>101</sup> Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
- <sup>102</sup> Satlantis, University Science Park, Sede Bld 48940, Leioa-Bilbao, Spain
- <sup>103</sup> Centre for Electronic Imaging, Open University, Walton Hall, Milton Keynes, MK7 6AA, UK
- <sup>104</sup> Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
- <sup>105</sup> Cosmic Dawn Center (DAWN)
- <sup>106</sup> Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
- <sup>107</sup> Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain
- <sup>108</sup> Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
- <sup>109</sup> INFN-Bologna, Via Irnerio 46, 40126 Bologna, Italy
- <sup>110</sup> Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- <sup>111</sup> Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA

- <sup>112</sup> Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- <sup>113</sup> Istituto Nazionale di Fisica Nucleare, Sezione di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
- <sup>114</sup> INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
- <sup>115</sup> Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK
- <sup>116</sup> INAF - Osservatorio Astronomico di Brera, via Emilio Bianchi 46, 23807 Merate, Italy
- <sup>117</sup> INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy, and INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- <sup>118</sup> Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014, Paris, France
- <sup>119</sup> ICL, Junia, Université Catholique de Lille, LITL, 59000 Lille, France
- <sup>120</sup> Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain
- <sup>121</sup> CERCA/ISO, Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
- <sup>122</sup> Technical University of Munich, TUM School of Natural Sciences, Physics Department, James-Frank-Str. 1, 85748 Garching, Germany
- <sup>123</sup> Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
- <sup>124</sup> Laboratoire Univers et Théorie, Observatoire de Paris, Université PSL, Université Paris Cité, CNRS, 92190 Meudon, France
- <sup>125</sup> Departamento de Física Fundamental. Universidad de Salamanca. Plaza de la Merced s/n. 37008 Salamanca, Spain
- <sup>126</sup> Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France
- <sup>127</sup> Center for Data-Driven Discovery, Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
- <sup>128</sup> Dipartimento di Fisica - Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
- <sup>129</sup> ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy
- <sup>130</sup> California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA
- <sup>131</sup> Departamento Física Aplicada, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia, Spain
- <sup>132</sup> Instituto de Física de Cantabria, Edificio Juan Jordá, Avenida de los Castros, 39005 Santander, Spain
- <sup>133</sup> INFN, Sezione di Lecce, Via per Arnesano, CP-193, 73100, Lecce, Italy
- <sup>134</sup> Department of Mathematics and Physics E. De Giorgi, University of Salento, Via per Arnesano, CP-193, 73100, Lecce, Italy
- <sup>135</sup> INAF-Sezione di Lecce, c/o Dipartimento Matematica e Fisica, Via per Arnesano, 73100, Lecce, Italy
- <sup>136</sup> CEA Saclay, DFR/IRFU, Service d'Astrophysique, Bat. 709, 91191 Gif-sur-Yvette, France
- <sup>137</sup> Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- <sup>138</sup> 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
- <sup>139</sup> Caltech/IPAC, 1200 E. California Blvd., Pasadena, CA 91125, USA
- <sup>140</sup> Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
- <sup>141</sup> Department of Physics and Astronomy, Vesilinnantie 5, 20014 University of Turku, Finland
- <sup>142</sup> Serco for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafraña del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- <sup>143</sup> ARC Centre of Excellence for Dark Matter Particle Physics, Melbourne, Australia
- <sup>144</sup> Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
- <sup>145</sup> Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town, 7535, South Africa
- <sup>146</sup> DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK
- <sup>147</sup> Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
- <sup>148</sup> Department of Astrophysics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
- <sup>149</sup> Department of Physics, Centre for Extragalactic Astronomy, Durham University, South Road, Durham, DH1 3LE, UK
- <sup>150</sup> Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
- <sup>151</sup> IRFU, CEA, Université Paris-Saclay 91191 Gif-sur-Yvette Cedex, France
- <sup>152</sup> Univ. Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, 53, Avenue des Martyrs, 38000, Grenoble, France
- <sup>153</sup> INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125, Firenze, Italy
- <sup>154</sup> Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy
- <sup>155</sup> HE Space for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafraña del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- <sup>156</sup> Theoretical astrophysics, Department of Physics and Astronomy, Uppsala University, Box 515, 751 20 Uppsala, Sweden
- <sup>157</sup> Mathematical Institute, University of Leiden, Einsteinweg 55, 2333 CA Leiden, The Netherlands
- <sup>158</sup> School of Physics & Astronomy, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK
- <sup>159</sup> Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- <sup>160</sup> Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
- <sup>161</sup> Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010, New York, NY, USA



**Fig. A.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 A: Reconstruction metrics

Figures Fig. A.1 show the differences in the rate of change measurement between input images and the inpainting output.

## Appendix B: Extreme examples

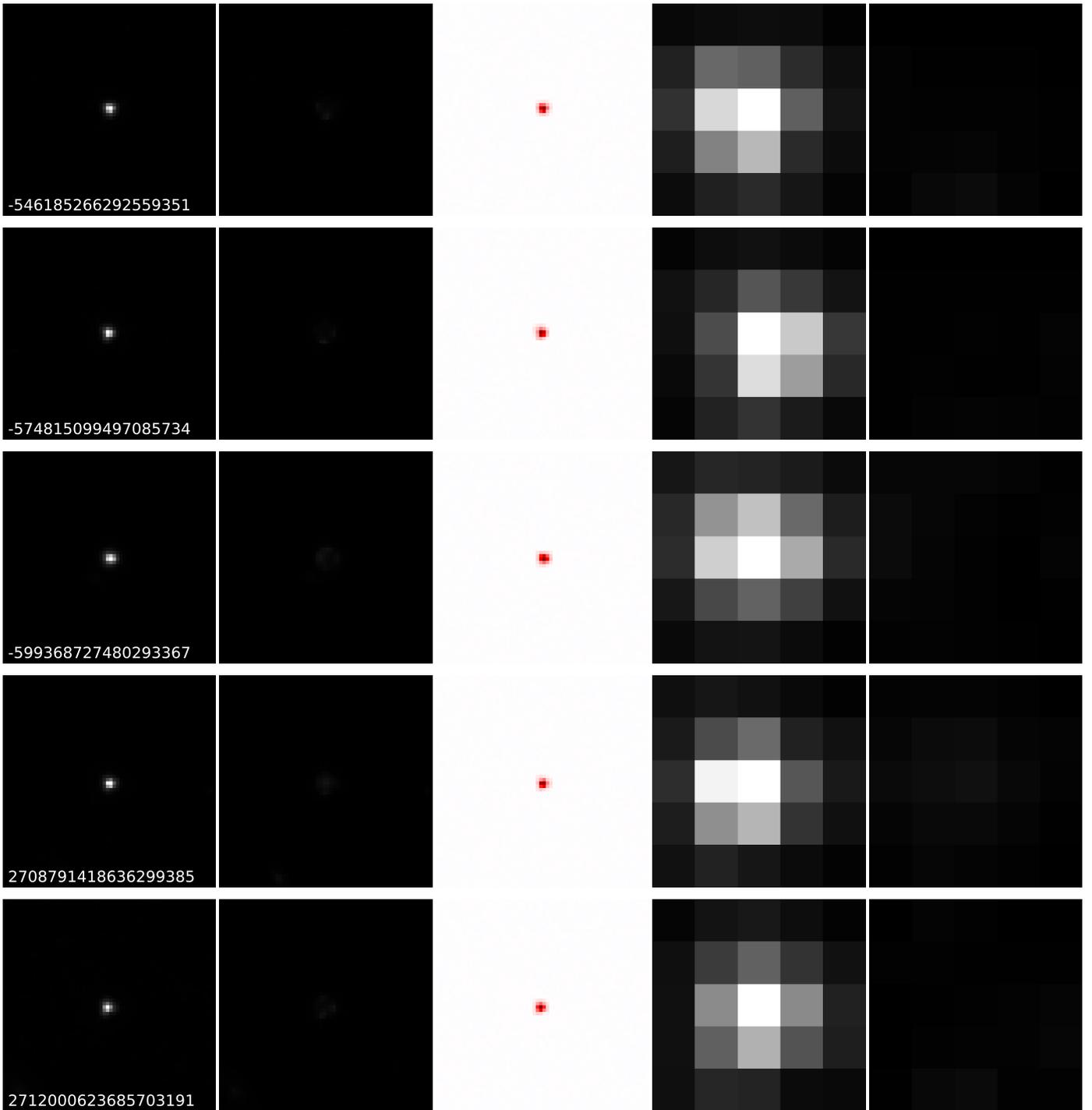
We present output examples from the inpainting pipeline where pixel values deviate significantly from the expected input pixels. Figure B.1 highlights much fainter pixels, while Figure B.2 shows much brighter ones.

## Appendix C: Morphology examples

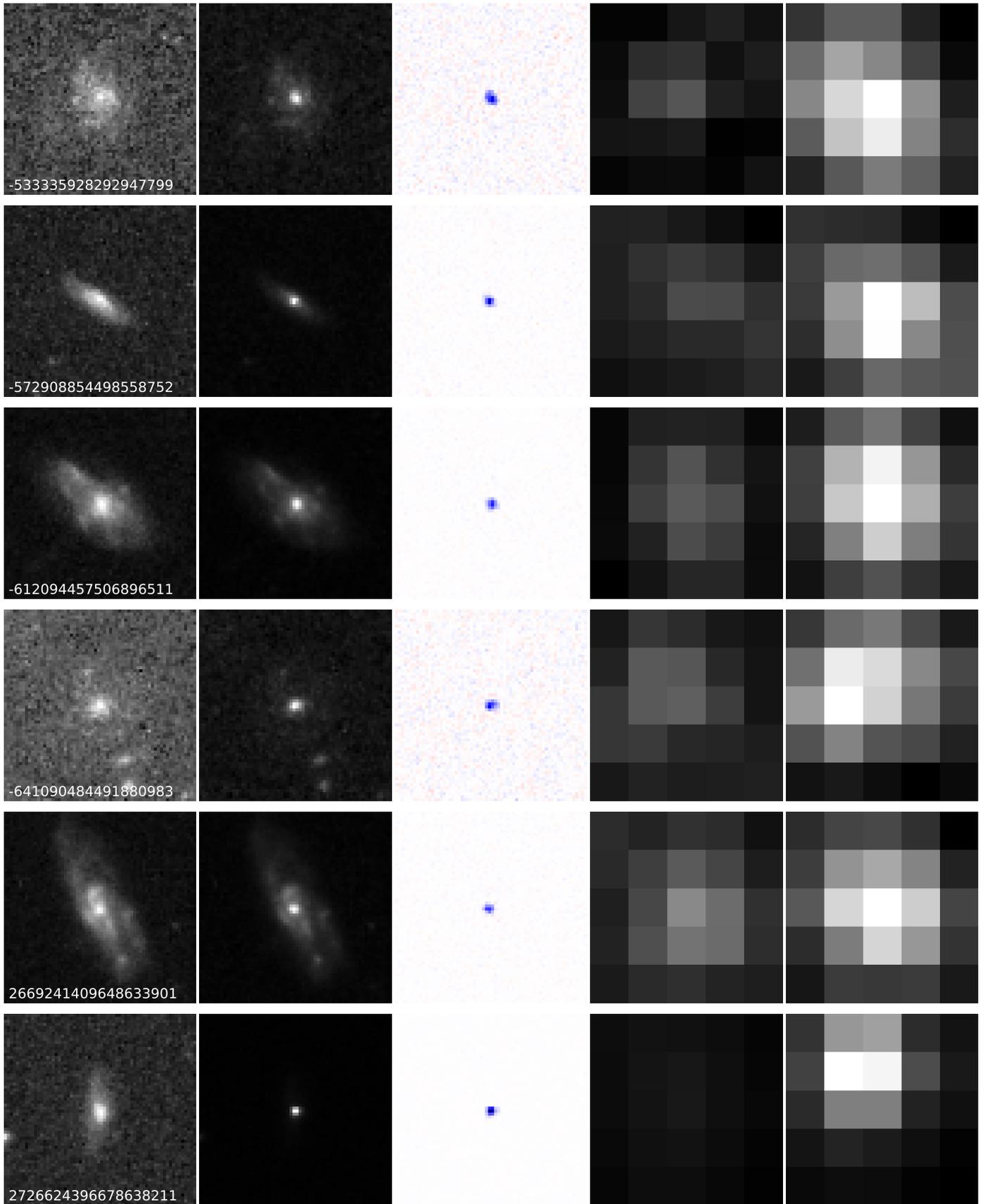
We show output examples from various morphology types to showcase how the diffusion model adapts to different galaxy shapes. Edge-on galaxies, spirals and mergers are shown in Fig. C.1, Fig. C.2 and Fig. C.3, respectively.

## Appendix D: Additional results

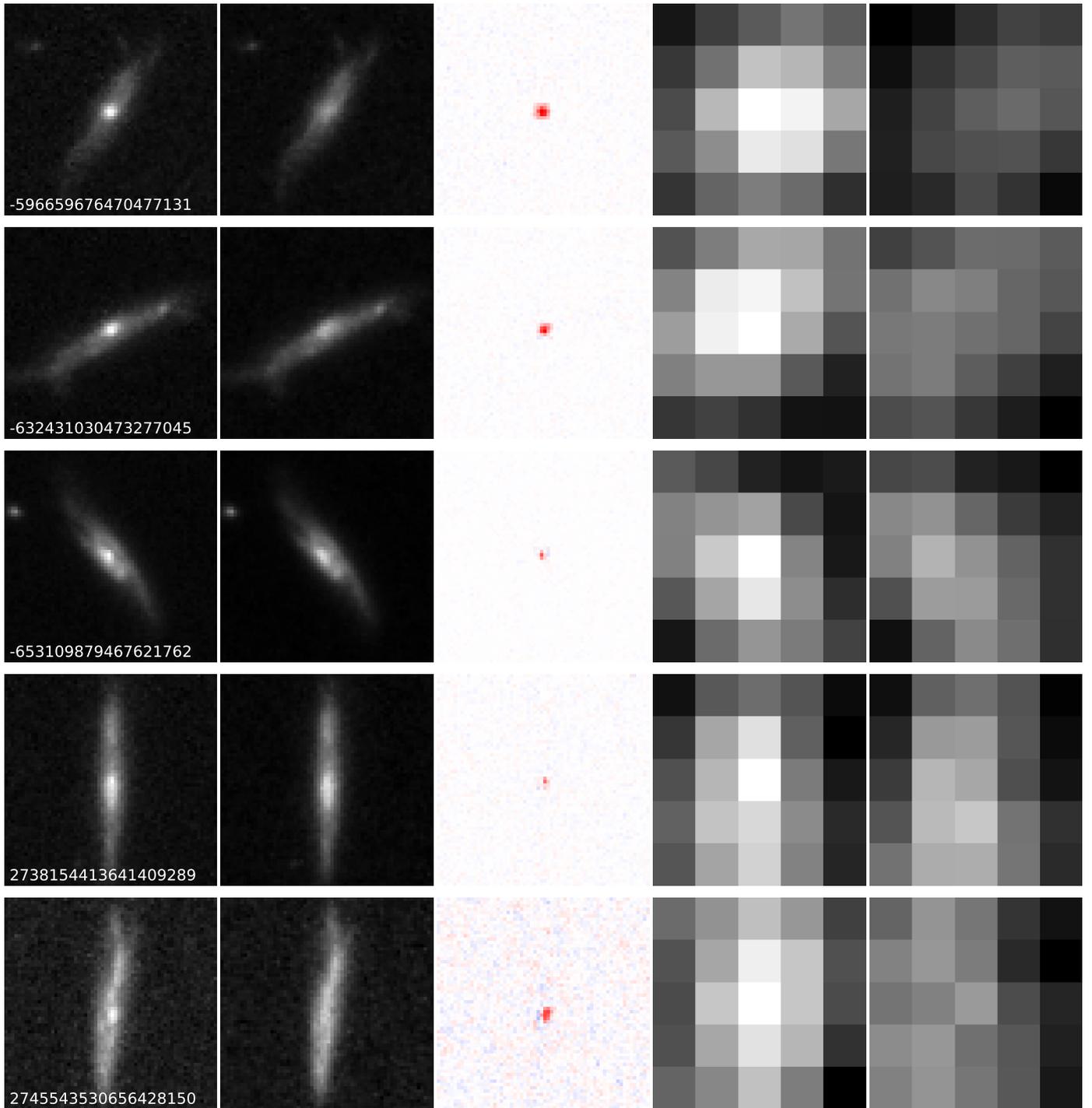
We present collated results for each of the star, galaxy, AGN and QSO selections in Fig. D.1. In Fig. D.2, the MSE and max ratio scores of all the MB25 selections as shown, indicating our diffusion model is more sensitive to selecting brighter  $F_{\text{AGN}}$  sources.



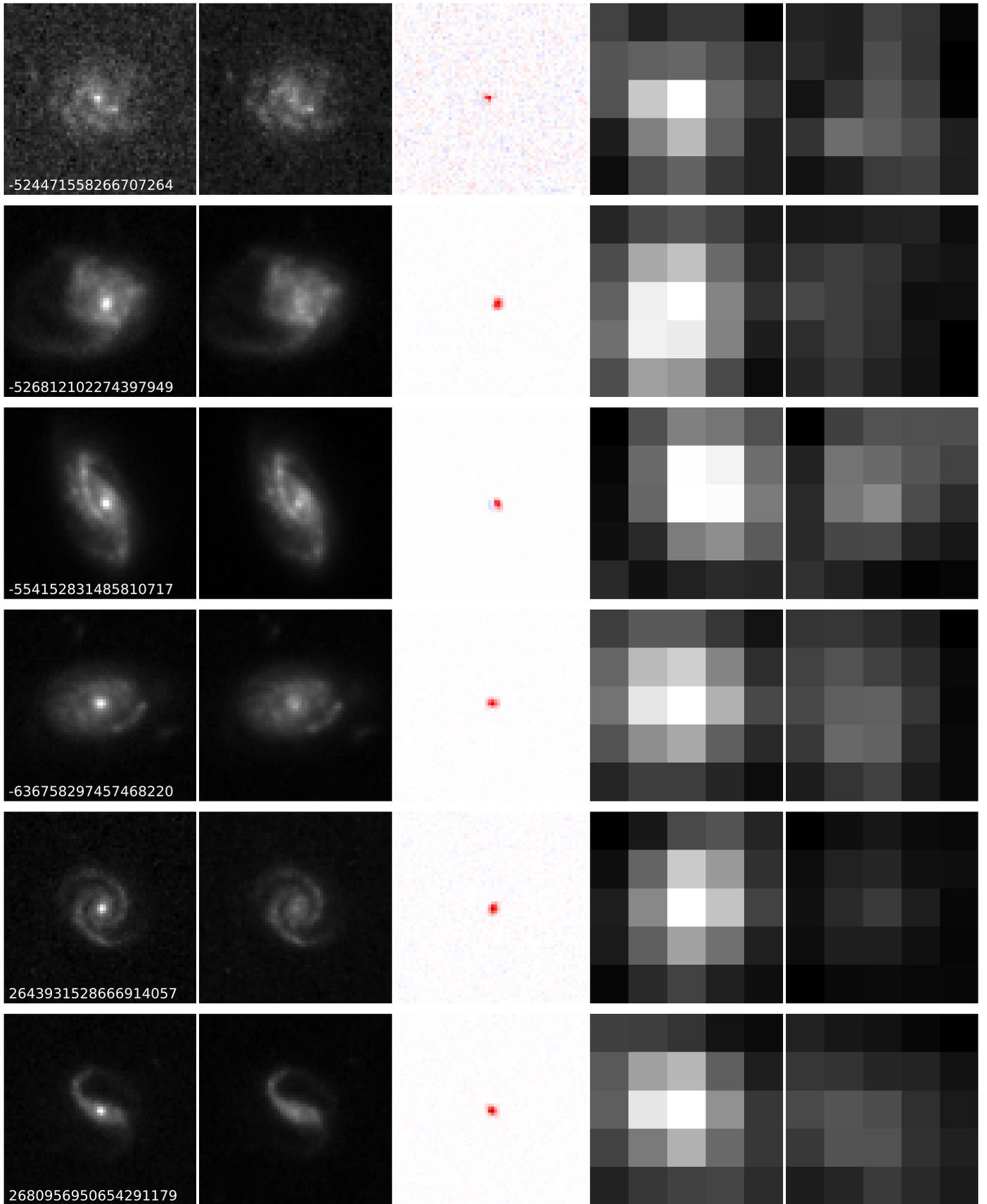
**Fig. B.1.** Examples of sources that achieved a very high maximum pixel difference ratio.



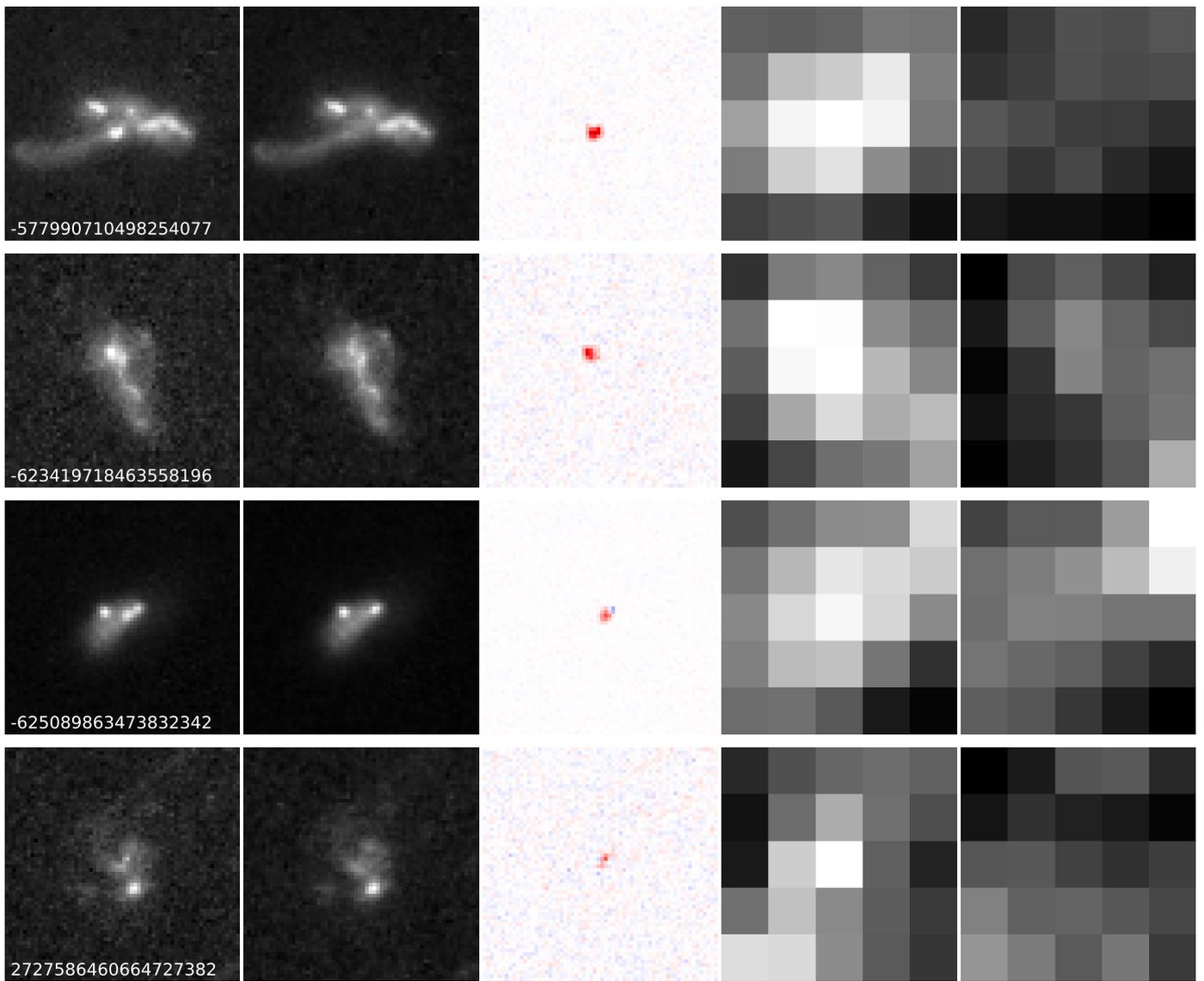
**Fig. B.2.** Examples of sources that achieved a very low maximum pixel difference ratio.



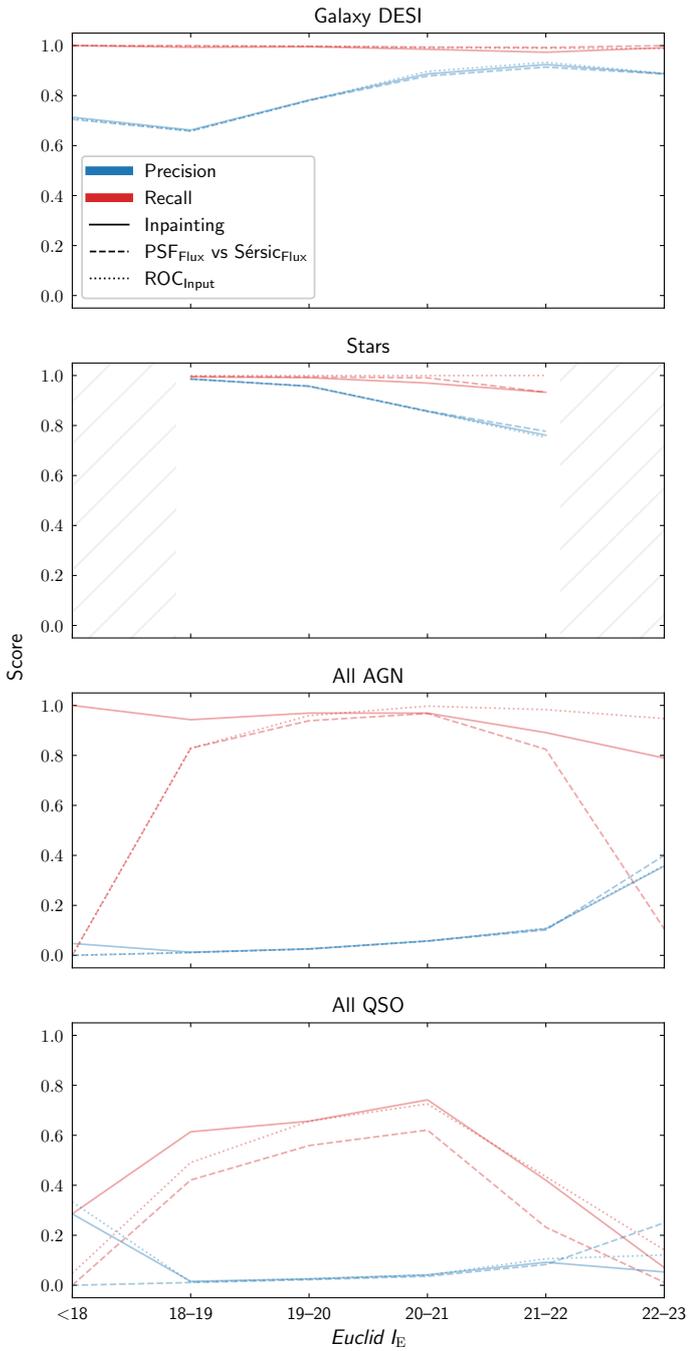
**Fig. C.1.** Examples of inpainting on edge-on sources.



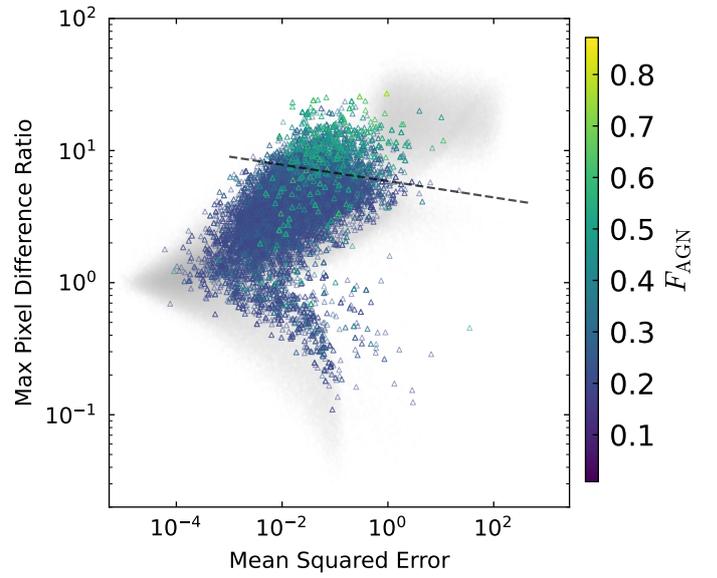
**Fig. C.2.** Examples of inpainting on spiral sources.



**Fig. C.3.** Examples of inpainting on mergers.



**Fig. D.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. D.2.** MB25 selection on our inpainting metrics. The  $F_{AGN}$  values are provided by MB25. All shown sources have an  $F_{AGN} > 0.2$ . A large selection of sources within the MB25 selection are not located within our boundary. Of the sources that are within our boundary, many exhibit higher  $F_{AGN}$ .