

# Predicting resolved galaxy properties from photometric images using convolutional neural networks

Tobias Buck<sup>1\*</sup>, Steffen Wolf<sup>2</sup>

<sup>1</sup>Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany

<sup>2</sup>MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge CB2 0QH, United Kingdom

Accepted XXXX . Received XXXX; in original form XXXX

## ABSTRACT

Multi-band images of galaxies reveal a huge amount of information about their morphology and structure. However, inferring properties of the underlying stellar populations such as age, metallicity or kinematics from those images is notoriously difficult. Traditionally such information is best extracted from expensive spectroscopic observations. Here we present the *Painting IntrinsiC Attributes onto SDSS Objects* (PI-CASSO) project and test the information content of photometric multi-band images of galaxies. We train a convolutional neural network on 27 558 galaxy image pairs to establish a connection between broad-band images and the underlying physical stellar and gaseous galaxy property maps. We test our ML algorithm with SDSS *ugriz* mock images for which uncertainties and systematics are exactly known. We show that multi-band galaxy images contain enough information to reconstruct 2d maps of stellar mass, metallicity, age and gas mass, metallicity as well as star formation rate. We recover the true stellar properties on a pixel by pixel basis with only little scatter,  $\lesssim 10\%$  compared to 20 – 30% statistical uncertainty from traditional mass-to-light-ratio based methods. We further test for any systematics of our algorithm with image resolution, training sample size or wavelength coverage. We find that galaxy morphology alone constrains stellar galaxy properties to better than  $\sim 20\%$  thus highlighting the benefits of including galaxy morphology into the parameter estimation. The network architecture and all code is publicly available.

**Key words:** methods: statistical techniques: image processing - galaxies: abundances - galaxies: structure - galaxies: fundamental parameters - galaxies: photometry

## 1 INTRODUCTION

The Sloan Digital Sky Survey (Abazajian et al. 2009) has revolutionised our understanding of the Universe by creating detailed three-dimensional maps of the Universe. Additionally, this survey provides deep multi-colour images of about one third of the sky, supplemented by spectra for more than three million astronomical objects. In particular, our understanding of galaxy formation and evolution was strongly impacted by the characterisations of the galaxy population either via the color bi-modality of galaxies (e.g. Strateva et al. 2001) which divides them into a blue star-forming cloud and a red, passively evolving sequence or via the galaxy mass-metallicity relation (Tremonti et al. 2004).

Large scale surveys such as the Dark Energy Survey

(The Dark Energy Survey Collaboration 2005), EUCLID or the upcoming Large Synoptic Survey Telescope (Ivezić et al. 2019) and the Nancy Grace Roman Telescope will supersede the achievements of SDSS by scanning even larger areas of the sky and including even fainter objects. With these tremendous datasets, astronomy is entering the era of big (imaging) data. Spectroscopic follow-up observations for such huge datasets will become increasingly impractical and new, innovative analysis methods need to be explored in order to exploit the datasets in full depth.

One example of such an innovative analysis method has been explored in the crowd-sourced Galaxy Zoo project (Lintott et al. 2008) for the classical SDSS imaging dataset. In this citizen science project the public was presented with galaxy images and asked to classify them into different groups (edge-on spiral, face-on spiral, barred spiral, ellipticals, etc.). In case enough people participate, each galaxy will be classified by several individuals and a relatively robust classification can be achieved. One drawback of this approach is that it relies on the pattern recognition abilities of

\* E-mail: tbuck@aip.de

individual humans. In this way, no strong conclusions/quantification of the features that lead to the final decision can be made. And more importantly, even these concepts can not cope with the vastness of future datasets.

Fortunately, the field of machine learning (ML) has matured the tasks of image recognition, classification and segmentation (e.g. Ronneberger et al. 2015) and by now, several groups have tested ML techniques on astronomical data with great success. ML techniques have been applied in a large number of astronomical use-cases such as galaxy morphology classification (e.g. Dieleman et al. 2015; Huertas-Company et al. 2015; Beck et al. 2018; Hocking et al. 2018), gravitational lensing analysis (e.g. Hezaveh et al. 2017; Lanusse et al. 2018; Petrillo et al. 2017, 2019; Jeffrey et al. 2020), galaxy cluster mass estimates (e.g. Ntampaka et al. 2015, 2018, 2019; Ho et al. 2019) or cold gas kinematics (Dawson et al. 2020). Furthermore, ML has proven useful for star-galaxy separation (e.g. Kim & Brunner 2017; Bai et al. 2019), galaxy de-blending (Lanusse et al. 2019), super resolution imaging (e.g. Falahkheirkhah et al. 2019) and outlier detection (Margalef-Bentabol et al. 2020). It has been used to estimate galaxy redshifts (Soo et al. 2018; Menou 2019; Wilson et al. 2020; Campagne 2020), to reconstruct star cluster parameters (Pasquato & Chung 2016, 2019) or galaxy merger states (Bottrell et al. 2019), to identify asteroids (e.g. Smirnov & Markov 2017), or to transfer knowledge between different galaxy surveys (Domínguez Sánchez et al. 2019; Pérez-Carrasco et al. 2019). ML has been successfully used to investigate the star formation and quenching processes (Lovell et al. 2019; Bluck et al. 2020), estimate cooling rates in cosmological simulations (Galligan et al. 2019) or interpret the turbulent inter-stellar medium (ISM, Peek & Burkhardt 2019; Van Oort et al. 2019). Other works have explored dark matter halo formation (Lucie-Smith et al. 2019), created galaxy mock catalogs (e.g. Xu et al. 2013; Kamdar et al. 2016), inferred the halo mass distribution function (Charnock et al. 2020) or predicted the galaxy-halo connection (Agarwal et al. 2018; Jo & Kim 2019) using ML techniques. Finally, using ML cosmological parameters could be derived from redshift surveys (Ramanah et al. 2019; Ntampaka et al. 2020) or supernova data (Escamilla-Rivera et al. 2020; Wang et al. 2020).

From observations it is well established that galaxy color is a good predictor for its stellar mass (e.g. Bell et al. 2003; Zibetti et al. 2009) and its stellar mass in turn correlates well with the total stellar or gaseous metallicity (e.g. Tremonti et al. 2004; Gallazzi et al. 2005). Combining this with the color or morphological bi-modality of spirals vs. ellipticals then implies that there might also exist a correlation between morphology, color and fundamental properties such as stellar mass, gas mass, SFR or metallicity. An immediate question arising from this is: *Can we estimate galaxy properties typically derived from spectroscopy using broad band photometry? And how much more accurate can stellar mass estimates based on color get if we include morphological galaxy properties?*

Recently, such questions have been studied by Bonjean et al. (2019) using integrated UV, optical and IR luminosities to derive star formation rate (SFR) estimates via a random forest algorithm and by Dobbels et al. (2019) who investigated a morphology assisted mass-to-light ratio estimate via deep learning methods. Wu & Boada (2019) on

the other hand used RGB 3-color images from SDSS to infer the spectroscopically derived metallicity solely from imaging data using neural networks. And de Diego et al. (2021) combine morphological classification and photometric redshift determination using ML techniques to prevent catastrophic outliers.

In this work we take previous successes of ML techniques in image processing and classification one step further. Whereas previous astronomical studies looked at tasks such as either classifying galaxies (e.g. Domínguez Sánchez et al. 2018) or connecting three-color images and scalar values such as total gas phase metallicity (e.g. Wu & Boada 2019) or redshift. Here we propose to use ML techniques to establish a connection between astronomical broad band imaging data and the underlying resolved physical properties on a pixel-by-pixel basis. This means, we propose to train a CNN to *predict* resolved maps of physical properties such as surface mass density, metallicity or star formation rate (SFR) solely from photometric images. This enables to obtain resolved maps of galaxy properties from relatively cheap photometric images otherwise only possible to derive from expensive integral field unit (IFU) spectroscopic surveys such as SDSS-MANGA (Bundy et al. 2015), SAMI (Bryant et al. 2015) or CALIFA (Sánchez et al. 2012).

The paper is organised as follows: In §2 we present a short summary of the data set used, SDSS mock images created for model galaxies from the Illustris simulation. In §3 we describe the neural network and the training procedure and proceed to present the main results in §4. In this section we further discuss the accuracy and possibilities of our method. Section §5 describes how the methods presented here can be applied to observational datasets and we conclude our work in §6 with a summary.

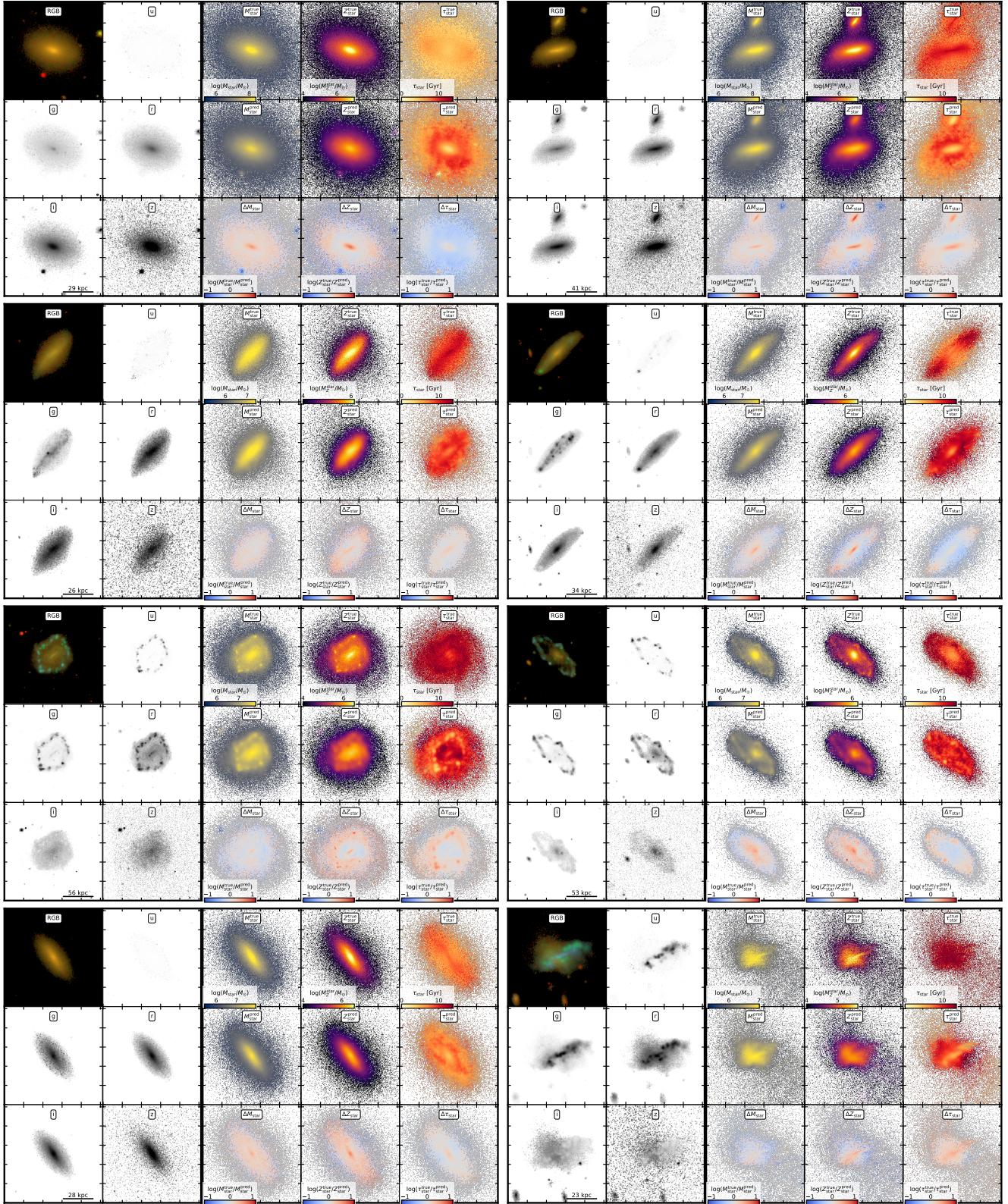
## 2 METHODS

While our proof-of-concept work is based on SDSS mock images to verify our procedure and to quantify its accuracy, the CNN architecture can easily be applied to real photometric surveys or even IFU surveys such as SDSS-MANGA, SAMI or CALIFA leveraging the reliance on hydrodynamical simulations. In this section, we lay out the details of the ML workflow (including a description of the CNN architecture in Section 3) and explain the dataset of synthetic photometric images (Section 2.1) as well as corresponding maps of fundamental galaxy properties (e.g. stellar mass, gas mass or SFR maps). We have combined our training, prediction and analysis procedures into a publicly available python package named *PICASSO* (*Painting Intrinsic Attributes onto SDSS Objects*)<sup>1</sup>.

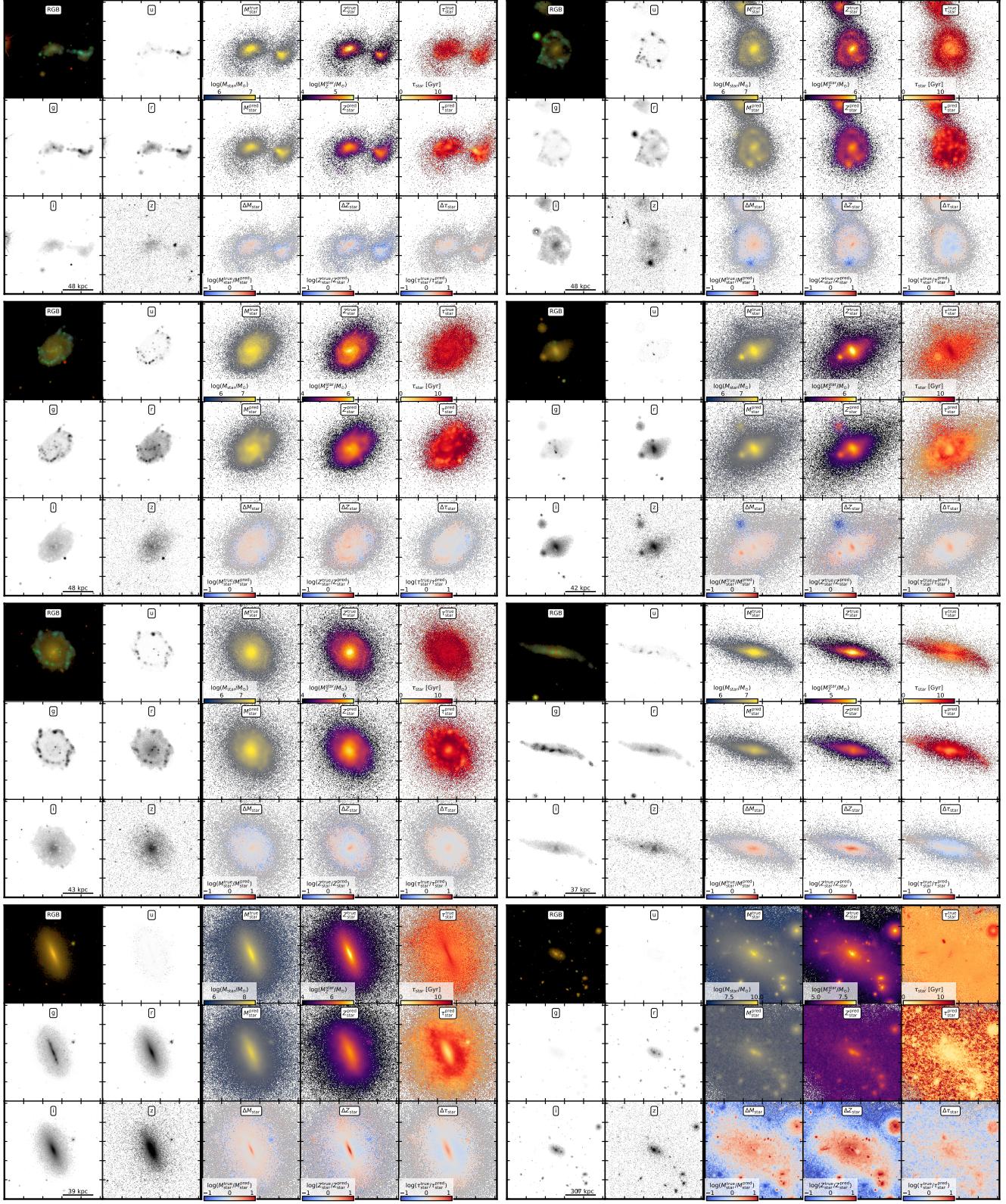
### 2.1 Synthetic observations

Modern cosmological simulations of galaxy formation have mastered the task of calculating model galaxies that resemble statistical properties of the observed galaxy population (e.g. Vogelsberger et al. 2014; Schaye et al. 2015; Wang et al. 2015; Dolag et al. 2016; Hopkins et al. 2018;

<sup>1</sup> <https://github.com/TobiBu/picassso>



**Figure 1.** Randomly selected examples galaxies and their ML reconstruction. In each sub-panel, we show on the left an RGB image plus all five SDSS bands. To the right of that we show in the upper row the true stellar mass, metallicity and age maps. The middle row shows the ML prediction and the bottom row shows the logarithmic residual between true and predicted maps. Note, the SDSS fibre size is 3 arcsecond corresponding to  $\sim 3$  kpc at a redshift of  $z \sim 0.05$  which for most galaxies in SDSS only covers the central parts. In contrast, our ML approach here is able to get information on physical properties for the entire galaxy out to several half-mass radii. This includes even the companion galaxies in merging galaxy pairs shown in the upper right example and the top row of Fig. 2.



**Figure 2.** Same as Fig. 1. Note the excellent reconstruction of merging galaxy pairs in the upper row and the ability of the network to reconstruct morphological features such as edge-on disks or cluster galaxies (bottom right) well.

Pillepich et al. 2018; Buck et al. 2019, 2021). Therefore, such model galaxies are ideally suited to build the training data for machine learning applications. Here, we study the information content of  $\sim 27\,000$  photometric images of galaxies at  $z \sim 0$  taken from the publicly available Illustris Project ([www.illustris-project.org/data](http://www.illustris-project.org/data) Vogelsberger et al. 2014) which is a suite of hydrodynamical simulations of galaxy formation in a periodic volume of size (106.5 Mpc)<sup>3</sup>. The Illustris Project uses the Arepo code for gravity and gas dynamics (Springel 2010) which simultaneously evolves the gas, star, and dark matter components from cosmological initial conditions. The galaxy formation physics models include primordial and metal line gas cooling, star formation, gas recycling, metal enrichment, supermassive black hole (SMBH) growth, and gas heating by feedback from supernovae and SMBHs (Vogelsberger et al. 2013). Here we present results from the highest resolution simulation, which has a baryonic mass resolution of  $1.26 \times 10^6 M_{\odot}$ .

### 2.1.1 Radiative transfer post-processing of model galaxies

For each model galaxy we create images for four viewing directions (face-on, edge-on and two random directions) in the five SDSS filters ( $u, g, r, i, z$ ) following the procedures described in Torrey et al. (2015) and Snyder et al. (2015). These authors generated noiseless high-resolution images with the SUNRISE<sup>2</sup> code (Jonsson 2006).

SUNRISE assigns to each star particle a spectral energy distribution (SED) based on its mass, age, and metallicity using stellar population models by Bruzual & Charlot (2003) with a Chabrier (2003) initial mass function. Each star particle radiates from a region with a size directly proportional to the radius that encloses its 16 nearest neighbours which makes the emission region smaller (larger) in areas with high (low) stellar density. This adaptive smoothing technique leads to much more realistic surface brightness maps in regions of low stellar density (see Torrey et al. 2015, for full details). The publicly available images were created without the dust radiative transfer (RT) option enabled in SUNRISE in order to avoid over-modelling galaxies with ISM resolution insufficient to gain substantial accuracy by performing the RT (see Snyder et al. 2015, for more details). Since the subject of this paper is a proof-of-concept of the ML technique and dust attenuation will not strongly change the morphology of low redshift galaxies (but see e.g. Buck et al. 2017, for the effect of dust on high redshift star forming galaxies). Thus, we are confident that this omission of realism is no major restriction to our study.

### 2.1.2 SDSS mock images

The output of the RT step in SUNRISE is the SED at each position of 512x512 pixels in each of four cameras chosen to view the galaxy face-on, edge-on and from two random directions. From this, Sunrise creates raw mock images by integrating the (optionally redshifted) SED in each pixel over a set of common astronomical filters, from the UV through

IR. In this paper, we aim to create SDSS mock images and thus perform this filter synthesis assuming each galaxy resides at redshift  $z = 0.05$  by adopting the SDSS  $u, g, r, i, z$  filters. The spatial extent of each image is set to ten times the 3D stellar half-mass radius, and therefore the physical pixel scale varies. This results in pixel sizes of roughly 100 – 300 pc which is sufficiently small to simulate SDSS images of sources at  $z \sim 0.05$  regardless of the limited spatial resolution of the simulation.

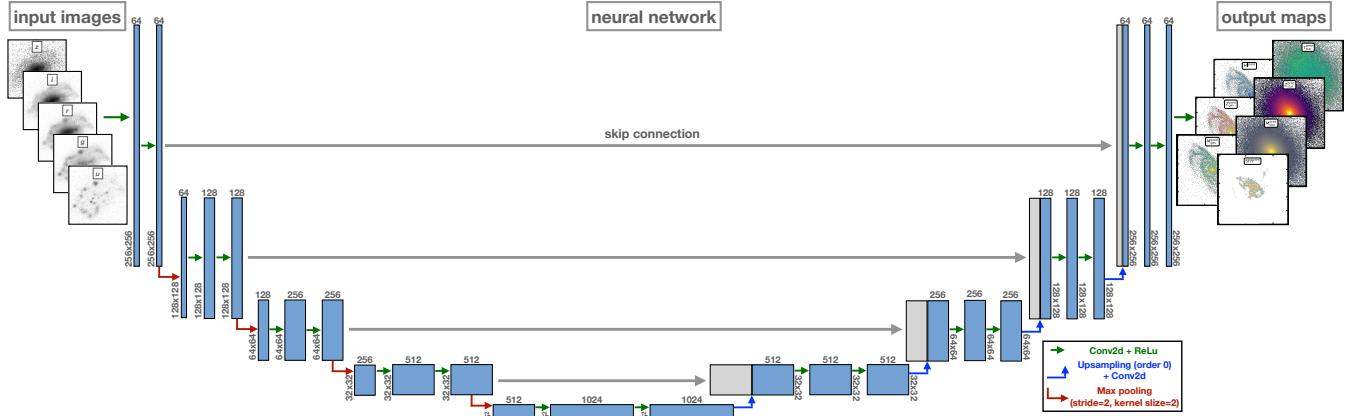
We apply several layers of image realism to the photometric images in order to account for foreground and background sources, point-spread function (PSF) blurring and observational photon noise. Following the procedure explained in Snyder et al. (2015), we convolve each high-resolution image with a Gaussian (PSF) with full-width at half-maximum (FWHM) of 1.0 kpc. Next, the images are rebinned to a constant pixel scale of 0.24 kpc (approximately 1/3 of the Illustris stellar gravitational softening). These parameters are chosen to roughly correspond to 1 arcsecond seeing for observations of a source located at  $z = 0.05$ , where the physical scale is roughly 1 kpc/arcsec. This choice was made such that the image properties roughly correspond to many sources in the Sloan Digital Sky Survey main galaxy sample (Strauss et al. 2002) or to an HST WFC3 survey of sources at  $z \sim 0.5$ .

In a next step, we add sky shot noise assuming Gaussian random noise independently applied to each of the pixels such that the final average signal-to-noise ratio of each pixel is 25. Further we add real SDSS background images in the appropriate photometric band to create fully synthetic  $u, g, r, i, z$  galaxy images (see e.g. 1 and 2). For this, we first download mosaics from the SDSS DR10 (Ahn et al. 2014) Science Archive Server with the mosaic web tool ([data.sdss3.org/mosaics](http://data.sdss3.org/mosaics)) and then randomly select a region of an appropriate size for each synthetic image (assuming the galaxies are at  $z = 0.05$ ) and add it to the simulated galaxy image. More details on this procedure can be found in (Snyder et al. 2015, Section 2.). Finally, for computational reasons we create our target images with a fixed pixel count of (256x256) for our fiducial images and successively lower pixel counts for the resolution study in Section 4.5. Thus, at each resolution level we have a total of  $\sim 2.7 \times 10^5$  synthetic images.

The Figures 1 and 2 show each eight example galaxies from our validation set. In the two left columns of each sub-panel we show a three color mock image created from the SDSS  $g, r, i$  bands as well as all five  $u, g, r, i, z$  single band images used as input to the network. Additionally, we show in the right three columns the resulting reconstructed physical maps (top panels) for stellar mass, stellar metal mass and average stellar age predicted by the ML algorithm from the photometric images. In comparison to this we also show the *true* counterparts (middle panels) extracted from the simulation. The bottom row shows the pixel residual maps between predicted and true maps.

Note, for spectroscopic observations, the SDSS fibre size is 3 arcsecond corresponding to  $\sim 3$  kpc at a redshift of  $z \sim 0.05$  which for most galaxies in SDSS only covers the very central parts of each galaxy. In contrast, our ML approach here is able to get information on physical properties for the entire galaxy out to several half-mass radii and even for merging systems (see lower right panel of Fig. 1).

<sup>2</sup> Sunrise is freely available at <https://bitbucket.org/lutorm/sunrise>



**Figure 3.** Neural network architecture: We use a U-net architecture with skip connections. Each blue box corresponds to a multi-channel feature map with the number of channels denoted on top of each box. The x-y image size is provided at the lower left edge of each box. Grey boxes represent copied feature maps. Red, blue and green arrows denote the different operations of down-sampling by max pooling, up-sampling and 2d-convolutions.

## 2.2 Maps of physical quantities

Additionally to the mock photometric images we create surface density maps of 7 physical galaxy properties with the same pixel sizes as the final photometric images (e.g. 256x256 in the fiducial case). In particular we consider the properties stellar mass,  $M_{\text{star}}$ , stellar metallicity,  $Z_{\text{star}}$ , stellar age,  $t_{\text{star}}$ , star formation rate, SFR, gas mass,  $M_{\text{gas}}$ , gas metallicity,  $Z_{\text{gas}}$ , and neutral hydrogen mass,  $M_{\text{HI}}$ . This amounts to another  $\sim 1.9 \times 10^5$  images in our dataset. No image realism is applied to the true physical images. The reasoning behind this is that we seek to create mock photometric images closely resembling the data from real galaxy surveys while the target physical maps in real surveys might already have accounted for contamination and applied some form of de-noising/de-blurring, modelling the effects of the point spread function (PSF). Thus, we like to study if some kind of automatic de-noising can be learned simultaneously.

### 2.2.1 Analysis pipeline

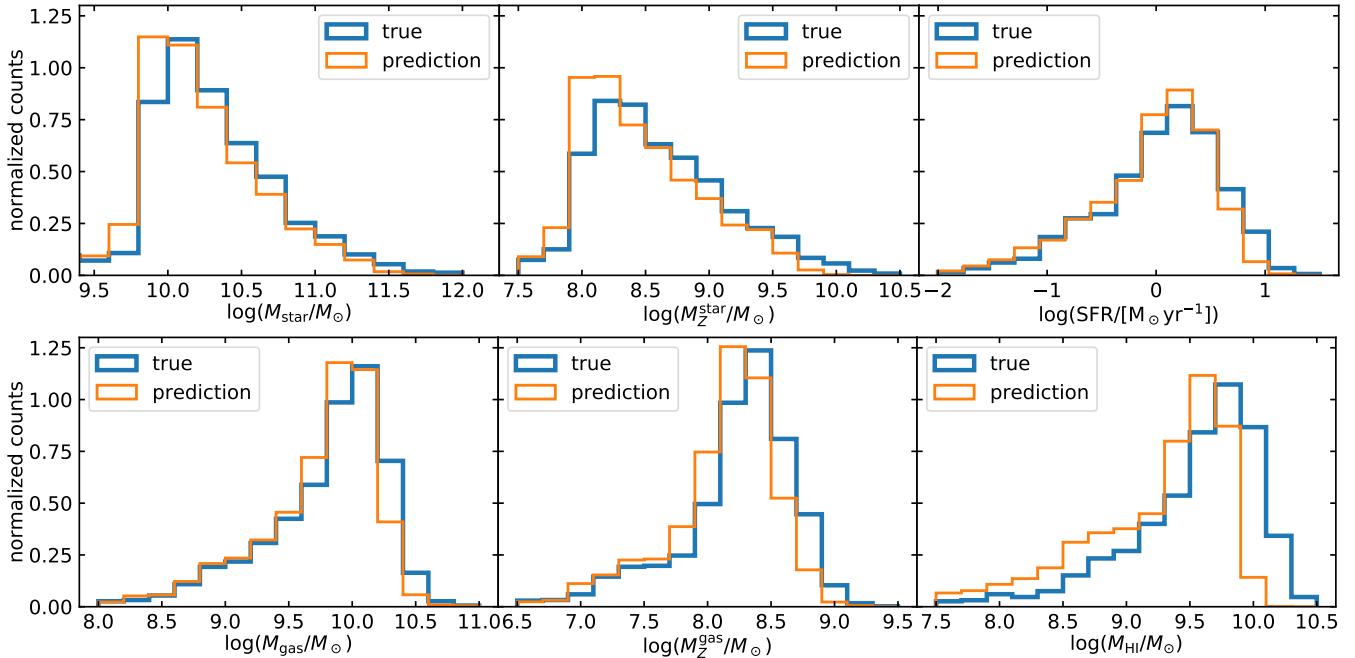
In the course of this paper we will compare galaxy properties derived from the predicted images and the true underlying physical maps. Since galaxy images are scaled to encompass an area of  $\pm 5$  half mass radii around each galaxy and galaxies can be randomly orientated (e.g. not only edge-on or face-on) we apply an ellipse fitting method<sup>3</sup> to the true stellar mass map in order to define which pixels belong to the galaxy. For stability and better comparison, the ellipse found from the true stellar mass image is kept fixed during analysis of all other properties both for true and predicted properties. Fiducial analysis is performed on all pixels inside an ellipse of semi-major axis size of two half mass radii,  $2R_{\text{half}}$ . Using ellipses with different semi-major axis values we are further able to assess any radial dependence of our image reconstruction.

<sup>3</sup> We use the ellipse fitting technique from the publicly available python package photutils (<https://photutils.readthedocs.io/en/stable/>)

Note, the true maps derived from the Illustris galaxies may contain empty pixels, especially in the outskirt of the galaxies where the stellar density or the gas density decreases strongly. Therefore, when we compare predicted vs. true properties, we restrict ourselves to only take into account those pixels, that have non-zero values in the true properties although the network is able to also predict values for originally empty pixels owing to its convolutional nature.

## 3 MACHINE LEARNING SETUP

Thanks to the widespread availability of affordable graphics processing units (GPUs) capable of performing general purpose, highly parallel computing, neural networks have become the tool of choice for image classification and regression problems. Connecting the imaging data to the underlying physical properties derived from spectroscopy is essentially an image regression problem. Deriving spatial relations in images is most readily done using multiple layers of image convolutions and adaptively constraining the filters in use (see e.g. Krizhevsky et al. 2012) – concepts which have been in use in modern astronomy since its beginning. In the field of image recognition and analysis, these concepts have been formalised and combined which resulted in algorithms such as convolutional neural networks (CNN O’Shea & Nash 2015) and/or generative adversarial neural networks (Goodfellow et al. 2014). Such networks efficiently adapt to (learn) relations in the images which are of the order of the convolution filters (or kernels). Like in classical regression the parameters of a model function, here the filters themselves, are the subject of the regression. These filters will be adapted/learned through training the network, i.e. minimising a predefined loss function on a set of images where the property to be learned is known. The more layers of convolution are included in a network the better its capability to adapt to more and more abstract features (Zeiler & Fergus 2014). If the number of layers becomes large enough, the neural network is called *deep*.



**Figure 4.** Distributions of the true (blue) and predicted (orange) galaxy properties. Upper panels show stellar properties, from left to right we display stellar mass, metallicity and SFR. Lower panels show gaseous properties displaying from left to right the gas mass, metallicity and neutral hydrogen abundance.

### 3.1 Network Architecture

In this work, we use a convolutional neural network (CNN) to learn the connection between the 5 band photometry from SDSS *ugriz* and the underlying physical galaxy properties (on a pixel-by-pixel basis). Both input and output of this network are images and we therefore utilize an encoder-decoder style architecture with skip connections, commonly used in these tasks. This architecture type was first developed for biological-image processing (Ronneberger et al. 2015) and is referred to as a U-net due to its use of down-sampling and subsequent upsampling layers. Similar architectures are also referred to as Hourglass networks Newell et al. (2016) or pix2pix networks Isola et al. (2017) for image-to-image translation.

This enables us to not only connect the imaging data to scalar values of total mass, metallicity, stellar age or SFR but to further perform full 2D analysis on the image data itself such as calculation of mass, metal or age gradients or spatial variations of metallicity, SFR or stellar age.

The specific network architecture used for this work is shown in Fig. 3. We employ a U-Net with 4 down-/upsampling layers each of which has 2 convolutional layers and skip connections. We follow Ronneberger et al. (2015) and employ max-pooling to decreases the image resolution in every down-sampling layer and use a ReLU activation function.

#### 3.1.1 Dataset Splits and Augmentation

Our analysis is based on 27,558 pairs  $(x, y)$  of SDSS mock images ( $x \in \mathbb{R}^{5 \times W \times H}$ ) with the corresponding physical properties ( $y \in \mathbb{R}^{7 \times W \times H}$ ). We use 20% of all images for our analysis (test set) and the rest for network training. From

the training set, we again reserve 20% of all image pairs for validation and selecting the best network during the training run (lowest loss on the validation data set).

The training set size can be artificially increased through data augmentation. We utilize the symmetry of our mock images and augment the images with random 90-degree rotations, flips and image transpositions. Random noise is added to the images during the rendering of the mock images but fixed during training.

### 3.2 Network Training

For the training purposes we consider the network as a function  $f_\theta$  with adjustable parameters  $\theta$ . It maps any given mock image  $x$  to a prediction of physical properties  $\hat{y} = f_\theta(x)$ . During the training we aim to find network parameters that minimize the regression loss for all training image pairs  $(x, y)$ :

$$\mathcal{L}(x, y) = \sum_{i=0}^W \sum_{j=0}^H \sum_{c=0}^7 w_{ij} \|\log_{10}(y_{cij}) - f_\theta(x)_{cij}\|_2 \quad (1)$$

The loss decomposes over all pixels (index  $i$  and  $j$ ) and the 7 physical properties. Since the physical properties span multiple orders of magnitude, we chose to train the network to predict the physical properties in log-space. After training, we remap to the true physical properties in our analysis using  $10^{\hat{y}}$ . We use the weight factor  $w_{ij}$  to mask pixels that correspond to bins without any simulated particles. For all pixels with known physical properties we set  $w_{ij} = 1$  and for all other pixels we use  $w_{ij} = 10^{-6}$  and  $y_{ijc} = 10^{-100}$  as a regularization factor.

We train the network for 13 epochs with a batch size of 16 using an Adam optimizer Kingma & Ba (2014). We start

with a learning rate of  $10^{-5}$  and reduce it by a factor of 10 after epoch 8 and 10. One training run takes 10 hours on a single GeForce RTX 2080 Ti.

## 4 RESULTS: THE INFORMATION CONTENT OF PHOTOMETRIC IMAGES

Properly evaluating the quality of the reconstruction of physical properties from photometric data is a difficult problem. A traditional metric measuring the per-pixel mean-squared error does not assess the joint statistics of the result. Therefore, this approach does not measure the very structure encoded in the image that we aim to reproduce with structured losses. Furthermore, we are dealing with galaxy images for which specific quantities are of interest, such as e.g. the total mass within a given radius or the radial gradients of the properties of interest. To more holistically evaluate the quality of our results, we employ two strategies: We evaluate how well we recover the total sum of pixel values in a given area of the image (e.g. total mass within  $R_{\text{half}}$ ) and how well we recover those properties as a function of radius.

Another practical aspect to test our algorithm for is the sample size used for training and the image resolution requirements. While gathering (low resolution) photometric data is relatively cheap and therefore datasets are usually huge, gathering the initial training datasets consisting of photometric input images plus reconstructed physical maps (e.g. stellar mass or metallicity maps) might become a challenge. Therefore, in Section 4.4, 4.5 and 4.6 we independently vary the number of input wavelength bands, image resolution and training sample size in order to evaluate how many different wavelength observations, image pairs or which image resolution is required to meet a given target accuracy. This effectively constrains the survey which can be used as input training set by the number of its available image pairs (e.g. SDSS MANGA with  $\sim 10,000$  images) or vice-versa for a given survey which redshift range might be accessible by constraining the minimum number of pixels per galaxy needed. At the same time, reducing the number of wavelength bands to a single image at fixed resolution tests how important color information is over morphological information. Similarly, reducing the image resolution at fixed color information (e.g. all 5 SDSS bands) studies the importance of morphological features over color.

### 4.1 Galaxy population properties

Looking at Figs. 1 and 2, we can already identify several successes but also failure modes of the model. In general, mass and metals are distributed following the light profile of the galaxy and residuals are small. Both, face-on and edge-on but also inclined galaxies are well reconstructed and internal structures such as rings and clumps or disk warps are recovered. Similarly, ongoing galaxy mergers are well captured by the model and both galaxies, the central one and the merging galaxy, are simultaneously well modelled. However, from the lower right panel of Fig. 2 we see that although the internal structure of cluster galaxies can be well recovered, their magnitude is under-predicted. The reason for this is, that our training set, the Illustris simulation, has

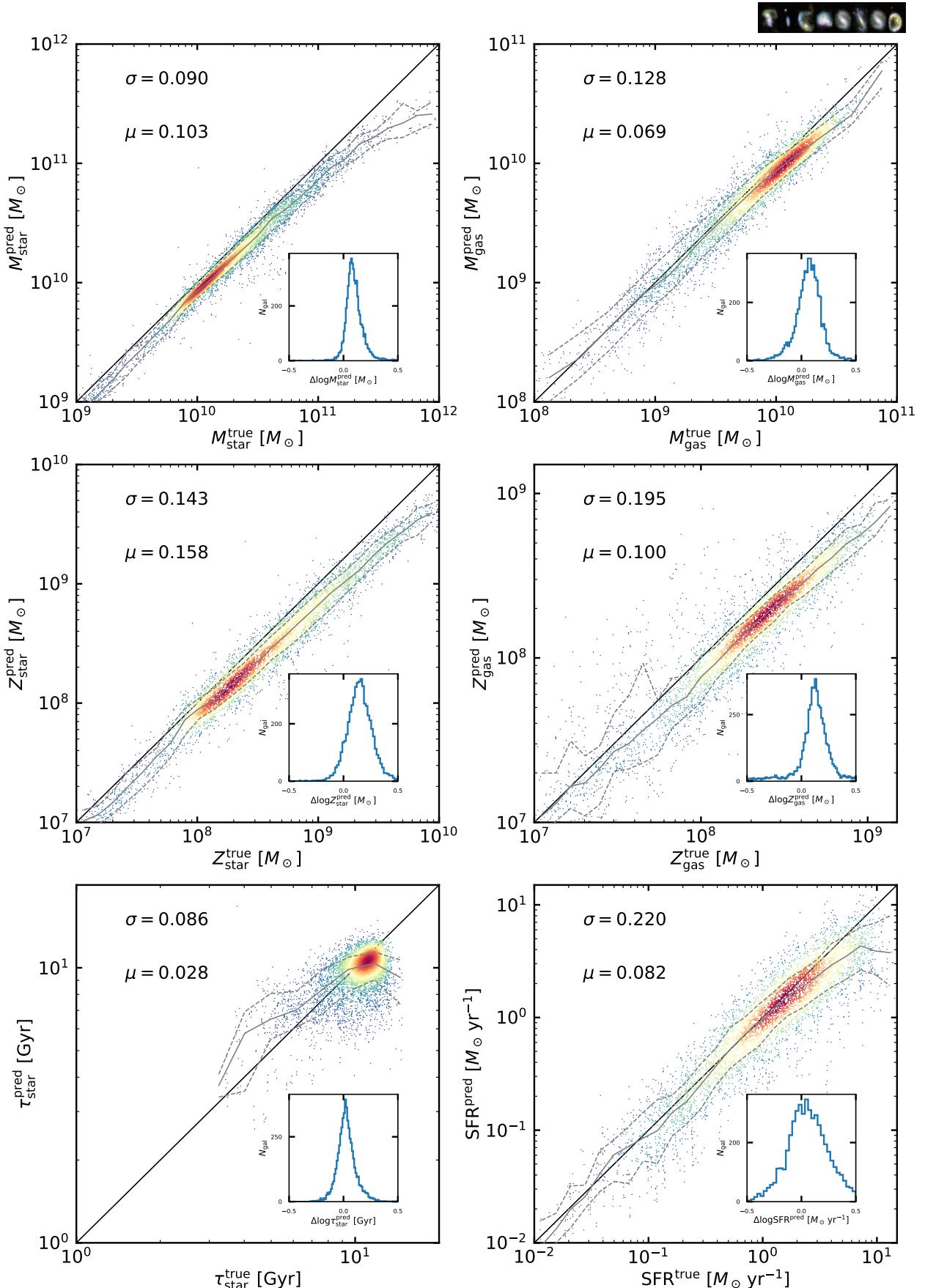
only a small number of such massive clusters and thus, during training the network only encounters very few examples of these clusters. This makes learning their statistical properties difficult.

We perform our final assessment of the quality and accuracy of our method on the validation sample randomly drawn from the whole SDSS mock Illustris data set. The distribution of stellar (top panels) and gaseous properties (bottom panels) of the galaxies from the validation set is shown in Fig. 4 by the blue histograms. Here we have defined the galaxy properties as the sum over all pixels within an ellipse of semi-major axis length of two projected half mass radii. Galaxies in the validation set span a stellar mass range from  $M_{\text{star}} \sim 10^{9.5} - 10^{12} M_{\odot}$  and a metal mass range of  $M_Z^{\text{star}} \sim 10^{7.5} - 10^{10.5} M_{\odot}$ , both peaked at the low mass end with a tail towards larger masses. The overall SFR of those galaxies spans a range between  $\sim 10^{-2} - 10^1 M_{\odot} \text{yr}^{-1}$  with the peak of the distribution around  $\text{SFR} \sim 1 - 2 M_{\odot} \text{yr}^{-1}$ . Corresponding gas masses range from  $M_{\text{gas}} \sim 10^8 - 10^{11} M_{\odot}$  and gas metal masses from  $M_Z^{\text{gas}} \sim 10^7 - 10^9 M_{\odot}$ . Both are peaked at the higher mass end with a tail towards lower masses. The galaxy neutral hydrogen fractions show values between  $M_{\text{HI}}^{\text{gas}} \sim 10^{7.5} - 10^{10.5} M_{\odot}$  with the peak of the distribution again at the high mass end. With orange histograms we show the predicted galaxy properties from the ML algorithm whose statistical properties are in good agreement with the true galaxy properties. We find slight offsets towards lower values of gas metallicity and neutral hydrogen fraction for our predictions.

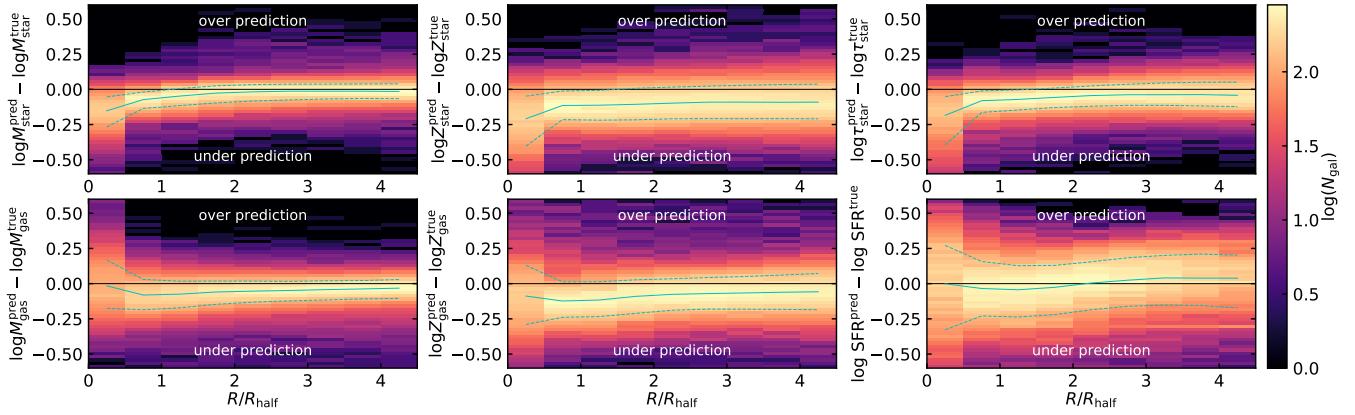
### 4.2 Individual galaxy properties

A more detailed comparison between predicted galaxy properties and their true values is shown in Fig. 5 where we plot the true value on the x-axis against the prediction on the y-axis for each galaxy in our sample. The left panels show stellar properties and the right hand side shows gaseous properties. From top to bottom we show mass, metal mass and stellar age or SFR. In this plot, each point represents one galaxy and compares the total sum of all pixels within two half mass radii,  $2 R_{\text{half}}$ , between the predicted and the true image. Color coding of the points highlights a kernel density estimate of the galaxy number density. In each panel the inset shows the distribution of the residuals between predicted and true value in logarithmic scale defined as  $\Delta\phi = \log(\phi_{\text{true}}) - \log(\phi_{\text{pred}})$ . The corresponding mean residual,  $\mu$ , and the  $1\sigma$  scatter are printed in the upper left corner of each panel. The mean residual can be interpreted as a measure of how well we can predict a spectroscopic property based on photometric images while the corresponding scatter is a measure of how accurate those predictions are on average.

Figure 5 shows that our image reconstruction is unbiased for most galaxy properties and does not show any strong dependence on galaxy property. For each property our predictions scatter along the 1:1 line with relatively small intrinsic scatter of  $\sim 10\%$ . We note some systematic deviations from the 1:1 relation at the highest masses, metal masses as well as at the highest SFRs/stellar ages. Stellar and gas masses are best reproduced by the ML algorithm with systematic residual scatter of  $\sim 0.09$  dex and  $\sim 0.13$  dex, respectively. Correspondingly, metal masses are pre-



**Figure 5.** True vs. predicted galaxy properties as measured from all pixels inside two half mass radii,  $2 R_{\text{half}}$ . The top row shows stellar and gas mass, the middle row shows stellar and gas metallicities and the bottom row shows the average stellar age within  $2 R_{\text{half}}$  and the SFR. The diagonal black line shows the 1:1 relation and the color-coding highlights a kernel-density estimation of the galaxy density. The inset panel shows the logarithmic difference between true and predicted galaxy property. The values for mean,  $\mu$ , and scatter,  $\sigma$ , of this distribution are printed in the upper left corner of each panel. In general, galaxy properties are reconstructed with high precision although we find that all properties are slightly under-predicted.



**Figure 6.** Accuracy of the predicted properties as a function of radius. For each radial bin we plot the distribution of logarithmic errors. Cyan solid lines show the median and dashed lines the 16<sup>th</sup> and 84<sup>th</sup> percentiles. Upper panels show stellar properties, from left to right we display stellar mass, metallicity and age. Lower panels show gaseous properties, from left to right we display gas mass, metallicity and SFR.

dicted to better than  $\sim 0.14$  dex for stellar metal mass and  $\sim 0.19$  dex for gaseous metal mass. Average stellar ages can be predicted to better than  $\sim 0.09$  dex while the SFR has intrinsic residual scatter of  $\sim 0.22$  dex. We further note that all properties are slightly under-predicted, at most for the stellar metal mass with  $\sim 0.16$  dex (see also next section for more details).

### 4.3 Testing for systematics with galaxy radius

In order to exploit the full information content of the two dimensional images it is important to quantify how well we can predict galaxy properties at different galactic radii. For example, we have seen in Fig. 5 that galaxy properties within twice the projected stellar half mass radius are reasonably well reconstructed. However, a priori it is not clear if the ML algorithm achieves this by simply predicting an average pixel value or indeed by reproducing the actual two dimensional structure of the galaxy image. However, Fig. 1 shows the potential of the CNN to actually reconstruct small morphological features of the galaxies as well as merging galaxy pairs.

Thus, in Fig. 6 we investigate how well different properties are reproduced as a function of galactic radius. To this extent, we compare the residual between true and predicted properties in different elliptical annuli ranging from  $0.5 - 5.0 R_{\text{half}}$  in semi-major axis length in bins of  $0.5 R_{\text{half}}$ . The upper panels in Fig. 6 show stellar properties and lower panels the gaseous properties. From left to right we show mass, metal mass and stellar age or SFR. Overall we do not find a strong radial dependence of the reconstructed galaxy properties but we note some interesting features.

While stellar mass is equally well reproduced at each radius with a residual scatter of less than 0.1 dex, the gas mass reconstruction shows increased scatter inside one half mass radius. Looking at the metal mass we find that the stellar metal mass is under-predicted by  $\sim 0.15$  dex at all radii and the scatter is  $\sim 0.2$  dex. Only in the very galaxy center we find a slight increase of under-prediction. Gas metal mass (and similarly average stellar age) does not show much radial dependence and is under-predicted by  $\sim 0.1$  dex at each radial bin. The SFR is surprisingly well predicted with no

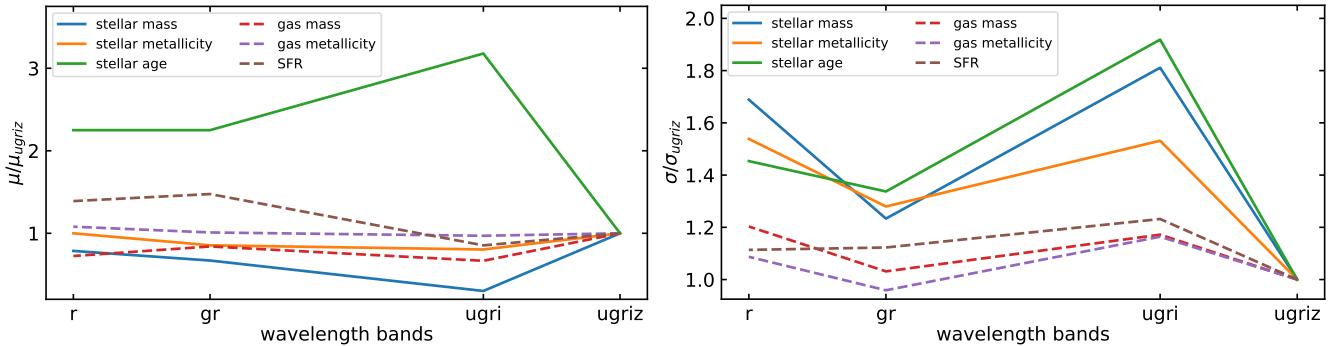
systematic bias but shows the largest scatter of  $\sim 0.25$  dex as we have already seen from Fig. 5.

### 4.4 Evaluating the ML prediction as a function of available wavelength bands

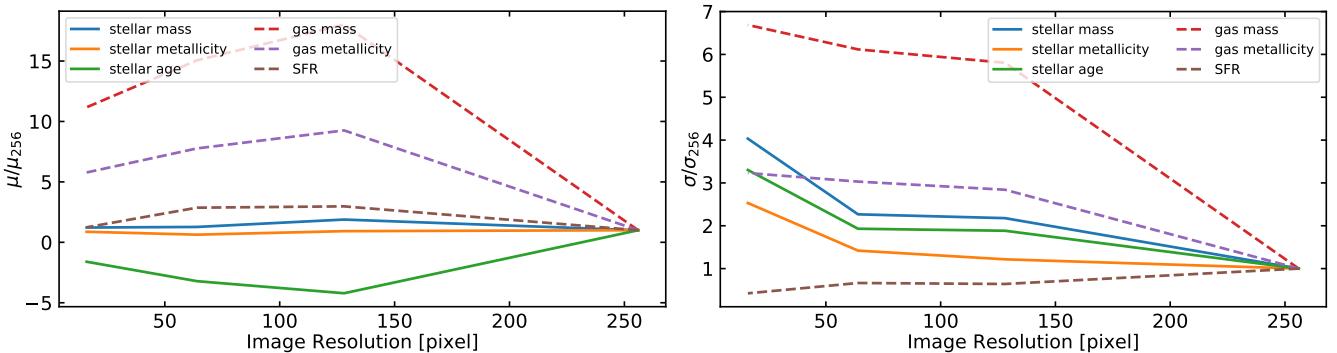
The accuracy of the galaxy reconstruction might depend on the number of available wavelength bands. This simply highlights the fact that stellar populations of different metallicity and age obey a different spectral energy distribution and thus the more bands are available, the better the wavelength coverage and with that a better representation of the entire spectrum can be achieved. Therefore, leaving out specific bands during training reduces the amount of information passed to the network. We have explored a few representative combination of bands to study this effect. We have repeated the analysis of the previous section using only the *ugri*, *gr* and *r* band images during training and prediction. This aims at probing the effect of leaving out the infrared part of the spectrum (*z* and/or *i*-band), the blue part of the spectrum (*u*-band) or only using a single band without any color information.

In Fig. 7 we show how the accuracy and precision of the galaxy reconstruction changes with available wavelength bands. To this end, we compare how the values of  $\mu$  and  $\sigma$  as calculated for Fig. 5 change. For completeness, for each combination of bands we show a similar figure to Fig. 5 in the appendix A (Figs. A1, A2, A3). The left panel of Fig. 7 shows the change of  $\mu$  normalized to  $\mu_{ugriz}$  for the *ugriz* combination of bands. The right panel shows  $\sigma$  normalized to  $\sigma_{ugriz}$ , respectively. This figure shows that, except for stellar age,  $\mu$  does not change much as a function of available bands. Stellar/gas mass and metallicity for the whole population are roughly equally well reconstructed even when using only *r*-band information. This might suggest that total luminosity in combination with morphology seems to be enough to predict these properties. SFR on the other hand is less accurately reconstructed as  $\mu$  increases with decreasing number of bands.

The scatter,  $\sigma$ , for stellar properties generally increases with decreasing number of bands while the one for gas properties only increase mildly by  $\lesssim 20\%$ . Interestingly, the in-



**Figure 7.** Dependence of the reconstruction accuracy and precision as a function of wavelength bands used. The left panel shows the accuracy to which galaxy properties within  $2 R_{\text{half}}$  can be reconstructed and the right panel shows how the precision of the reconstruction changes with wavelength bands used.



**Figure 8.** Dependence of the reconstruction accuracy and precision as a function of image resolution. The left panel shows the accuracy to which galaxy properties within  $2 R_{\text{half}}$  can be reconstructed and the right panel shows how the precision of the reconstruction changes with image resolution.

crease in stellar scatter is largest ( $\sim 60\%$ ) when leaving out only the  $z$ -band and stays roughly constant when reducing the number of bands further. A similar result is further obtained for the radial dependence of the reconstruction. It is yet unclear if this means that the  $z$ -band carries a lot information or if this would also happen for a combination of  $griz$  bands due to the reduced color information. Pinpointing this is currently outside the scope of this study. **Or should we do another test???**

In summary, the analysis done for Fig. 7 shows that on average the accuracy of the prediction is only little affected by color information but the precision decreases with decreasing number of wavelength bands.

#### 4.5 Evaluating the ML prediction as a function of image size

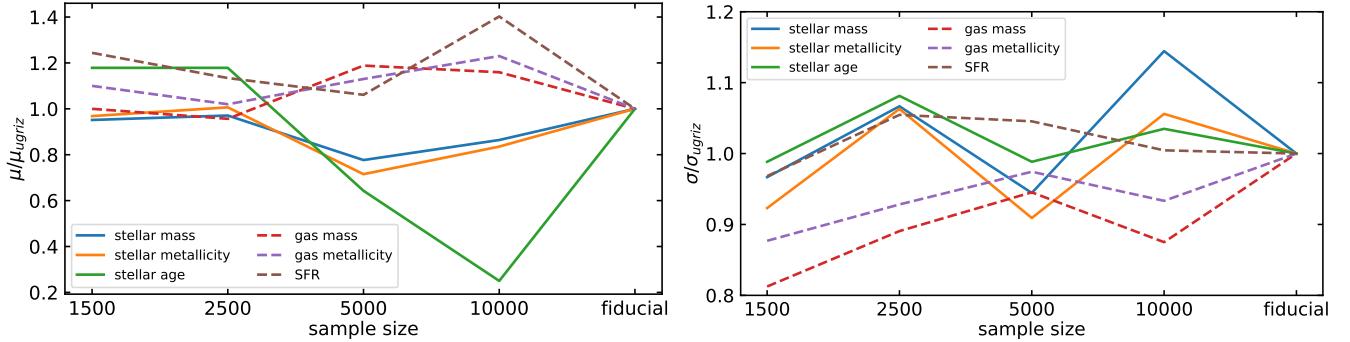
Similar to color information from using different wavelength bands, an important factor for the reconstruction of galaxy properties might come from the morphological information in the image. Here we test, how well spatially resolved a galaxy image must be to achieve a good accuracy in the reconstruction of galaxy properties. Therefore we reduce the spatial resolution of our fiducial images (256x256 pixel) by factors of two and retrain the network. Thus we study how well galaxy properties can be reconstructed when images of 128, 64 and 16 pixels are used.

Since image resolution is affected by either the resolving

power of the telescope or by the distance of the object, this either tests fixed telescope resolution how far away a galaxy sample can be to achieve a good reconstruction or at fixed distance it tests for the minimum resolution of a telescope. We summarise our results with Fig. 8 which in the left panel again shows the change of  $\mu$  normalized to  $\mu_{256}$  for the fiducial 256x256  $ugriz$  combination of bands and the right panel shows  $\sigma$  normalized to  $\sigma_{256}$ , respectively.

We find that stellar properties such as stellar mass, metallicity or age are less affected by image resolution than gas properties. We find that stellar mass and metallicity as well as SFR are equally accurately predicted across all image resolutions. Stellar age on the other hand is strongly overpredicted. Only the precision for stellar mass and metallicity decreases constantly up to a factor of 3 increased scatter at the lowest image resolution compared to the fiducial resolution of 256x256 pixels. The precision for SFR on the other hand stays roughly constant across all resolutions. Gas mass and metallicity are heavily under-predicted (up to a factor of  $\gtrsim 10$  compared to our fiducial resolution) when image resolution is changed and similarly their precision decreases with decreasing image resolution. The precision of gas metallicity decreases similarly to the stellar metallicity and gas mass precision degrades strongly with decreasing image resolution, up to a factor of 7 compared to our fiducial resolution.

We attribute this to the fact that the gas morphology shows more complexity than the stellar body. This can readily be seen from the example galaxy shown in Figs. B1 and



**Figure 9.** Dependence of the reconstruction accuracy and precision as a function of training sample size. The left panel shows the accuracy to which galaxy properties within  $2 R_{\text{half}}$  can be reconstructed and the right panel shows how the precision of the reconstruction changes with the number of galaxy images in the training sample.

**B2** in the appendix B. Furthermore, we train on stellar light from  $ugriz$  SDSS bands and naturally, the main contribution to these images comes from the stellar body of the galaxy. The gas of the galaxy will only marginally affect the SED via absorption and emission lines. Thus, it is not unexpected that gas properties are less well reproduced when the image resolution is changing. In fact, it is actually surprising that gas properties can be reproduced at all (even in our fiducial setup) given that galaxy images are mainly tracing stellar light.

The tests performed in this sub-section make us conclude that galaxy morphology contains a significant amount of information about galactic properties. Foremost gaseous properties such as gas mass or metallicity are mainly learned from morphological features. This becomes obvious when comparing to the results from the previous sub-section where we tested the information content of galaxy colors at fixed image resolution and thus at a fixed amount of morphological features resolved. Note, the total luminosity of the galaxies stays the same when resolution is changed. Thus, the only variable for this test is the resolved morphology. This makes us conclude that galaxy morphology carries much information about galaxy properties, not only in a global sense for total stellar mass but also as a resolved map-like feature.

These findings complement the textbook knowledge of galaxy luminosity and color being a good predictor for its total (stellar) mass (e.g. Kauffmann et al. 2003) and similarly galaxy color being a good predictor for its metallicity (modulo stellar age effects) (e.g. Tremonti et al. 2004; Gallazzi et al. 2005). Furthermore, in contrast to this established knowledge, adding and employing galaxy morphology through modern image recognition tools does not only enable to reconstruct galaxy properties from images alone, it further allows an estimate of the 2d maps of those properties. This presents a major advantage with respect to previous analysis where only a global measurement could be obtained.

#### 4.6 Evaluating the ML prediction as a function of training sample size

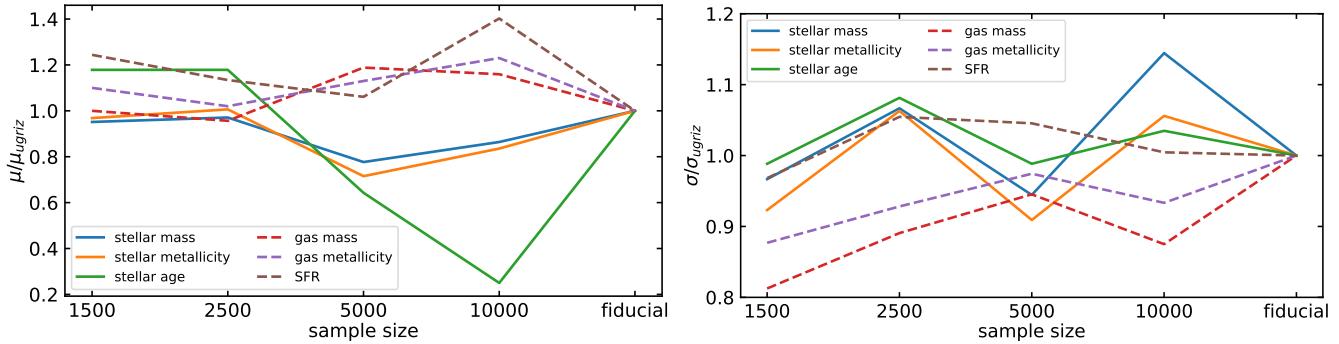
Finally, in order to apply the concept presented in this work to real galaxy images we need to retrain the network on a sample of real galaxy images and their reconstructed physical properties, e.g. from IFU observations. To this extent

is necessary to know the sensitivity of the algorithm to the number of galaxy images in the training sample, i.e. we want to know if a sample of galaxies of the size of SAMI (3000 galaxies) or merely of the size of MANGA (10 000 galaxies) is needed to achieve a good reconstruction accuracy.

We have retrained our neural network varying the training sample size from 1 500 up to the fiducial sample size of 17 637 galaxy image pairs and present the results in Fig. 10. While the accuracy and precision vary by about 10 – 20% when changing the sample size we do not find any clear trend with decreasing sample size. We attribute the change in reconstruction accuracy and precision to the fact that re-training and determining the optimal network weights is a stochastic process. Thus, testing the statistical significance of the  $\sim 10\%$  changes would require retraining a statistical sample of a few ten networks which is outside the scope of this work. We conclude from Fig. 10 that within the reach of current IFU sample sizes the accuracy and precision of the neural network reconstruction does not change significantly. However, when evaluating the accuracy and precision in a real application it will be necessary to train an ensemble of networks to carefully pin-down the reconstruction uncertainties.

## 5 DISCUSSION: APPLICATION TO OBSERVATIONS

Before we conclude, we will discuss in this section how the reconstruction of galaxy properties via ML applications in a real observational setup could work. In this work, we have presented a proof-of-concept how modern CNNs are able to reconstruct physical galaxy properties from multi-band galaxy images by making use of the combined spatial color, luminosity and morphological information contained in the images. There are several advantages of employing ML techniques for this task: (i) morphological features can be taken into account via the CNN, (ii) a generally faster analysis of image data but also that (iii) image analysis can be **automated** easier and **generalizes** better as no manual, human biased parameter choices have to be made (except for hyper-parameter choices of the network). Furthermore, **transferability** to other datasets is easier as a simple retraining on new data is needed. Sometimes, only the higher layers of the network which take care of the survey specific features need



**Figure 10.** Dependence of the reconstruction accuracy and precision as a function of training sample size. The left panel shows the accuracy to which galaxy properties within  $2 R_{\text{half}}$  can be reconstructed and the right panel shows how the precision of the reconstruction changes with the number of galaxy images in the training sample.

to be retrained and deeper layers that cope with the more general features can simply be ported (transfer learning).

In terms of **uncertainty**, the approach presented here relies on the fact, that enough training data to learn the statistical relations of galaxies exists and that the quality of this data is good enough in terms of sampling a fair representation of the objects. As a proof-of-concept we have used mock SDSS *ugriz* images from the Illustris dataset which has enough objects sampling a range of galactic morphologies. additionally, the true galaxy properties can easily be created from the raw simulation data. This in turn means, the network effectively learns the statistical properties of the Illustris Universe which obviously differs from the real Universe. Therefore, it is not recommended to train on mock images of any simulated model and then apply the network to real galaxy images.

There are two options to proceed from here: (i) use a training set, where *ground truth* data is self-consistently derived from observational data, i.e. IFU observations where resolved galaxy properties are derived e.g. via spectral energy distribution fitting. (ii) Modify the supervised ML network to not blindly learn statistical relations of the data but be able to either include prior knowledge about galaxies or to move to unsupervised ML techniques to alleviate the pure reliance on *ground truth* data.

Point (i) is the most straightforward way of applying the proposed reconstruction network to real observational data such SAMI, MANGA or CALIFA. Those IFU studies have enabled the classical reconstruction of galaxy properties such as stellar mass and metallicity but also gas mass and gas metallicity from the spectral information included in the IFU data of several thousands of galaxies. Combining these datasets with images of photometric surveys of the same objects will then results in image-physical property pairs that can be fed to the CNN to train it. Thus, one would be able to transfer the knowledge gained from the "small" sample of IFU observed galaxies onto the larger set of galaxies from photometric images (millions of galaxies already in SDSS). Especially with existing large photometric surveys such as e.g. the Dark Energy Survey (DES Abbott et al. 2018; The Dark Energy Survey Collaboration 2005), the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys (Dey et al. 2019) or upcoming large photometric surveys from the next-generation of 30m class ground-based telescopes or space based observations

like EUCLID (Laureijs et al. 2011), JWST (Gardner et al. 2006), the Rubin Observatory Legacy Survey of Space and Time (LSST LSST Science Collaboration et al. 2009; Ivezić et al. 2019) or the Nancy Grace Roman Telescope in combination with photometric redshift estimates (e.g. Henges et al. 2021) this approach might enable to use their imaging data in an unforeseen way.

## 6 SUMMARY & CONCLUSION

In this work we set out to study the information content of galaxy images. We are specifically interested to evaluate how well physical properties of galaxies such as stellar mass or metallicity maps, stellar age maps or gas mass and metallicity maps can be predicted from multi-band photometric images. To tackle this question we build upon modern ML techniques for image recognition such as CNNs. As a proof-of-concept we have studied more than 27.000 SDSS *ugriz* mock images from the Illustris simulation suite. We have extracted maps of physical quantities (stellar/gas mass and metallicity, stellar age and SFR and neutral HI gas mass) on the same scale and resolution as the corresponding SDSS mock images. The network architecture and all code is publicly available on github (<https://github.com/TobiBu/picassso>).

In our fiducial setup we have used galaxy images in all 5 SDSS bands (*ugriz*) with a resolution of 256x256 pixels as input to the network in order to predict maps of all seven physical quantities (stellar and gas mass/metallicity, HI mass, stellar age and SFR) at the same time. Additionally, we test for the importance of color information in the prediction by reducing the number of input bands from 5 to 1 as well as the importance of image resolution by reducing the number of pixels from 256 to 16 per side. Further, we test how many galaxy images in the training sample are needed to achieve a decent prediction accuracy by varying the number of training images from the fiducial 17 637 down to only 1 500.

Our results are summarized as follows:

- The ML network tested in this work (see Fig. 3 for a sketch of the architecture) is able to predict global stellar and gaseous galaxy properties from broad band photometric images for galaxies spanning a range of more than 3 orders of magnitude in mass and thus luminosity (see Fig. 4). Remarkably, this includes not only the reconstruction of the

central galaxy in an image but also merging or overlapping galaxies (see e.g. Fig. 1).

- In our fiducial setup, using all five *ugriz* SDSS bands, we recover the true stellar properties on a pixel by pixel basis with only little scatter,  $\lesssim 10\%$  compared to 20–30% statistical uncertainty from traditional mass-to-light-ratio based methods. Gaseous properties such as gas mass and metallicity are slightly less well predicted with a scatter of around  $\sim 15\text{--}20\%$ . Similarly, stellar age and SFR are reconstructed with a population wide scatter of  $\sim 9\text{--}22\%$  (c.f. Fig. 5).

- Since our algorithm reconstructs full 2d maps of galaxies from the corresponding images, we are able to evaluate how well each property is reconstructed at any given radius (see Fig. 6). We find no systematic dependence of the reconstruction accuracy with radius except for the very central parts ( $R \sim 0.5 R_{\text{half}}$ ) of galaxies where stellar properties are slightly under-predicted.

- In order to examine how important color information is in comparison to galaxy morphology, we vary the number of input bands at fixed image resolution (see Fig. 7) as well as the image resolution at fixed number of input bands (Fig. 8). We find that reducing the number of input bands causes a  $\sim 40\text{--}60\%$  increase in scatter for stellar properties and  $\sim 20\%$  for gaseous properties. On the other hand, downgrading the resolution blows up the scatter by up to a factor of 3–4 for stellar properties and  $\sim 7$  for gas mass compared to the fiducial resolution. This makes us conclude that galaxy morphology in addition to color carries significant amount of information for the reconstruction of galaxy properties.

- Finally, we explore how large the training sample size needs to be to enable a decent reconstruction accuracy and find no significant variations of accuracy with training sample size. Within the numbers of current IFU surveys ( $\sim 1500\text{--}10\,000$  galaxies) the reconstruction accuracy does not vary much.

## DATA AVAILABILITY

The network architecture and all code is publicly available on github (<https://github.com/TobiBu/picassso>) including Jupyter notebooks containing all analysis and plotting routines.

## ACKNOWLEDGMENTS

We like to thank Paul Torrey for helping us to access the Illustris SDSS mock images. Analysis has been performed on the ISAAC cluster of the Max-Planck-Institut für Astronomie and the HYDRA and DRACO clusters at the Rechenzentrum in Garching. This research made use of the MATPLOTLIB (Hunter 2007), SCI PY (Jones et al. 01 ), ASTROPY (Price-Whelan et al. 2018) and NUMPY, IPYTHON AND JUPYTER (Walt et al. 2011; Pérez & Granger 2007; Kluyver et al. 2016) PYTHON packages. This research made use of Photutils, an Astropy package for detection and photometry of astronomical sources (Bradley et al. 2020). Hyperlink figures to code access are inspired by Sven Buder and Rodrigo Luger.

## References

- Abazajian K. N., et al., 2009, *ApJS*, **182**, 543  
 Abbott T. M. C., et al., 2018, *ApJS*, **239**, 18  
 Agarwal S., Davé R., Bassett B. A., 2018, *MNRAS*, **478**, 3410  
 Ahn C. P., et al., 2014, *ApJS*, **211**, 17  
 Bai Y., Liu J., Wang S., Yang F., 2019, *AJ*, **157**, 9  
 Beck M. R., et al., 2018, *MNRAS*, **476**, 5516  
 Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, *ApJS*, **149**, 289  
 Bluck A. F. L., Maiolino R., Sánchez S. F., Ellison S. L., Thorpe M. D., Piotrowska J. M., Teimoorinia H., Bundy K. A., 2020, *MNRAS*, **492**, 96  
 Bonjean V., Aghanim N., Salomé P., Beelen A., Douspis M., Soubrié E., 2019, *A&A*, **622**, A137  
 Bottrell C., et al., 2019, *MNRAS*, **490**, 5390  
 Bradley L., et al., 2020, astropy/photutils: 1.0.0, doi:10.5281/zenodo.4044744, <https://doi.org/10.5281/zenodo.4044744>  
 Bruzual G., Charlot S., 2003, *MNRAS*, **344**, 1000  
 Bryant J. J., et al., 2015, *MNRAS*, **447**, 2857  
 Buck T., Macciò A. V., Obreja A., Dutton A. A., Domínguez-Tenreiro R., Granato G. L., 2017, *MNRAS*, **468**, 3628  
 Buck T., Obreja A., Macciò A. V., Minchev I., Dutton A. A., Ostriker J. P., 2019, arXiv e-prints, p. [arXiv:1909.05864](https://arxiv.org/abs/1909.05864)  
 Buck T., Rybizki J., Buder S., Obreja A., Macciò A. V., Pfrommer C., Steinmetz M., Ness M., 2021, arXiv e-prints, Bundy K., et al., 2015, *ApJ*, **798**, 7  
 Campagne J.-E., 2020, arXiv e-prints, p. [arXiv:2002.10154](https://arxiv.org/abs/2002.10154)  
 Chabrier G., 2003, *PASP*, **115**, 763  
 Charnock T., Lavaux G., Wandelt B. D., Boruah S. S., Jasche J., Hudson M. J., 2020, *MNRAS*,  
 Dawson J. M., Davis T. A., Gomez E. L., Schock J., Zabel N., Williams T. G., 2020, *MNRAS*, **491**, 2506  
 Dey A., et al., 2019, *AJ*, **157**, 168  
 Dieleman S., Willett K. W., Dambre J., 2015, *MNRAS*, **450**, 1441  
 Dobbels W., Krier S., Pirson S., Viaene S., De Geyter G., Salim S., Baes M., 2019, *A&A*, **624**, A102  
 Dolag K., Komatsu E., Sunyaev R., 2016, *MNRAS*, **463**, 1797  
 Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, *MNRAS*, **476**, 3661  
 Domínguez Sánchez H., et al., 2019, *MNRAS*, **484**, 93  
 Escamilla-Rivera C., Carvajal Quintero M. A., Capozziello S., 2020, *J. Cosmology Astropart. Phys.*, **2020**, 008  
 Falahkheirkhah K., Yeh K., Mittal S., Pfister L., Bhargava R., 2019, arXiv e-prints, p. [arXiv:1911.04410](https://arxiv.org/abs/1911.04410)  
 Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, *MNRAS*, **362**, 41  
 Galligan T. P., Katz H., Kimm T., Rosdahl J., Blaizot J., Devriendt J., Slyz A., 2019, arXiv e-prints, p. [arXiv:1901.01264](https://arxiv.org/abs/1901.01264)  
 Gardner J. P., et al., 2006, *Space Sci. Rev.*, **123**, 485  
 Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, arXiv e-prints, p. [arXiv:1406.2661](https://arxiv.org/abs/1406.2661)  
 Henghes B., Pettitt C., Thiyagalingam J., Hey T., Lahav O., 2021, arXiv e-prints, p. [arXiv:2109.02503](https://arxiv.org/abs/2109.02503)  
 Hezaveh Y. D., Perreault Levasseur L., Marshall P. J., 2017, *Nature*, **548**, 555  
 Ho M., Rau M. M., Ntampaka M., Farahi A., Trac H., Póczos B., 2019, *ApJ*, **887**, 25  
 Hocking A., Geach J. E., Sun Y., Davey N., 2018, *MNRAS*, **473**, 1108  
 Hopkins P. F., et al., 2018, *MNRAS*, **480**, 800  
 Huertas-Company M., et al., 2015, *ApJS*, **221**, 8  
 Hunter J. D., 2007, *Computing In Science & Engineering*, **9**, 90  
 Isola P., Zhu J.-Y., Zhou T., Efros A. A., 2017, in Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1125–1134

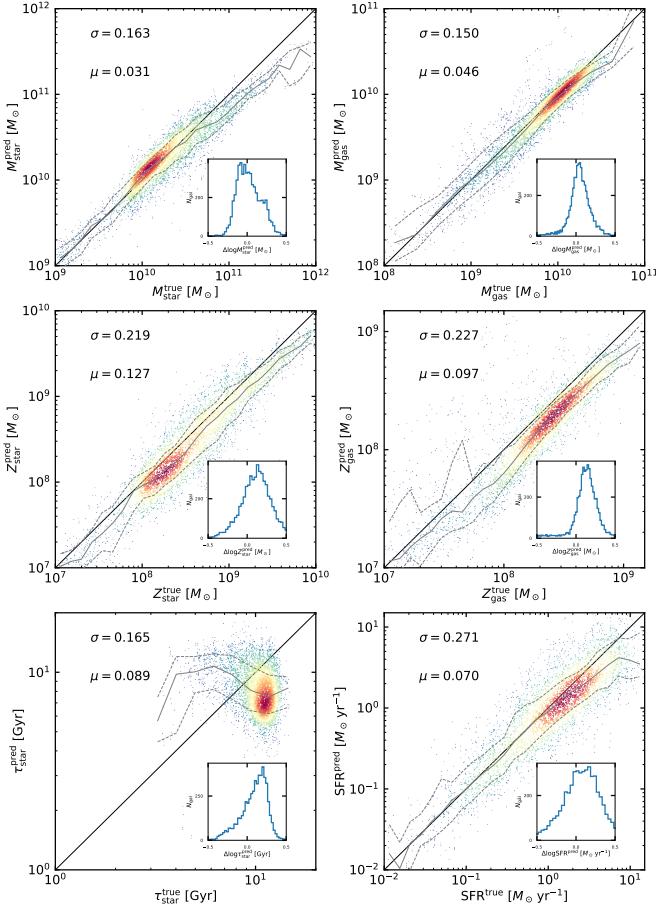
- Ivezić Ž., et al., 2019, *ApJ*, **873**, 111  
 Jeffrey N., Lanusse F., Lahav O., Starck J.-L., 2020, *MNRAS*, **492**, 5023  
 Jo Y., Kim J.-h., 2019, *MNRAS*, **489**, 3565  
 Jones E., Oliphant T., Peterson P., et al., 2001–, SciPy: Open source scientific tools for Python, <http://www.scipy.org/>  
 Jonsson P., 2006, *MNRAS*, **372**, 2  
 Kamdar H. M., Turk M. J., Brunner R. J., 2016, *MNRAS*, **457**, 1162  
 Kauffmann G., et al., 2003, *MNRAS*, **341**, 33  
 Kim E. J., Brunner R. J., 2017, *MNRAS*, **464**, 4463  
 Kingma D. P., Ba J., 2014, arXiv preprint arXiv:1412.6980  
 Kluyver T., et al., 2016, in Loizides F., Schmidt B., eds, Positioning and Power in Academic Publishing: Players, Agents and Agendas. pp 87 – 90  
 Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Pereira F., Burges C. J. C., Bottou L., Weinberger K. Q., eds, , Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp 1097–1105, <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-pdf>  
 LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201  
 Lanusse F., Ma Q., Li N., Collett T. E., Li C.-L., Ravankhsh S., Mandelbaum R., Póczos B., 2018, *MNRAS*, **473**, 3895  
 Lanusse F., Melchior P., Moolekamp F., 2019, arXiv e-prints, p. arXiv:1912.03980  
 Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193  
 Lintott C. J., et al., 2008, *MNRAS*, **389**, 1179  
 Lovell C. C., Acquaviva V., Thomas P. A., Iyer K. G., Gawiser E., Wilkins S. M., 2019, *MNRAS*, **490**, 5503  
 Lucie-Smith L., Peiris H. V., Pontzen A., 2019, *MNRAS*, **490**, 331  
 Margalef-Bentabol B., Huertas-Company M., Charnock T., Margalef-Bentabol C., Bernardi M., Dubois Y., Storey-Fisher K., Zanis L., 2020, arXiv e-prints, p. arXiv:2003.08263  
 Menou K., 2019, *MNRAS*, **489**, 4802  
 Newell A., Yang K., Deng J., 2016, in European conference on computer vision. pp 483–499  
 Ntampaka M., Trac H., Sutherland D. J., Battaglia N., Póczos B., Schneider J., 2015, *ApJ*, **803**, 50  
 Ntampaka M., Trac H., Sutherland D., Fromenteau S., Póczos B., Schneider J., 2018, in American Astronomical Society Meeting Abstracts #231. p. 225.04  
 Ntampaka M., et al., 2019, *ApJ*, **876**, 82  
 Ntampaka M., Eisenstein D. J., Yuan S., Garrison L. H., 2020, *ApJ*, **889**, 151  
 O’Shea K., Nash R., 2015, arXiv e-prints, p. arXiv:1511.08458  
 Pasquato M., Chung C., 2016, *A&A*, **589**, A95  
 Pasquato M., Chung C., 2019, *MNRAS*, **490**, 3392  
 Peek J. E. G., Burkhardt B., 2019, *ApJ*, **882**, L12  
 Pérez F., Granger B. E., 2007, *Computing in Science and Engineering*, **9**, 21  
 Pérez-Carrasco M., Cabrera-Vives G., Martínez-Marin M., Cerulo P., Demarco R., Protopapas P., Godoy J., Huertas-Company M., 2019, *PASP*, **131**, 108002  
 Petrallo C. E., et al., 2017, *MNRAS*, **472**, 1129  
 Petrallo C. E., et al., 2019, *MNRAS*, **484**, 3879  
 Pillepich A., et al., 2018, *MNRAS*, **473**, 4077  
 Price-Whelan A. M., et al., 2018, *AJ*, **156**, 123  
 Ramanah D. K., Lavaux G., Jasche J., Wandelt B. D., 2019, *A&A*, **621**, A69  
 Ronneberger O., Fischer P., Brox T., 2015, arXiv e-prints, p. arXiv:1505.04597  
 Sánchez S. F., et al., 2012, *A&A*, **538**, A8  
 Schaye J., et al., 2015, *MNRAS*, **446**, 521  
 Smirnov E. A., Markov A. B., 2017, *MNRAS*, **469**, 2024  
 Snyder G. F., et al., 2015, *MNRAS*, **454**, 1886  
 Soo J. Y. H., et al., 2018, *MNRAS*, **475**, 3613  
 Springel V., 2010, *MNRAS*, **401**, 791  
 Strateva I., et al., 2001, *AJ*, **122**, 1861  
 Strauss M. A., et al., 2002, *AJ*, **124**, 1810  
 The Dark Energy Survey Collaboration 2005, arXiv e-prints, pp astro-ph/0510346  
 Torrey P., et al., 2015, *MNRAS*, **447**, 2753  
 Tremonti C. A., et al., 2004, *ApJ*, **613**, 898  
 Van Oort C. M., Xu D., Offner S. S. R., Gutermuth R. A., 2019, *ApJ*, **880**, 83  
 Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *MNRAS*, **436**, 3031  
 Vogelsberger M., et al., 2014, *MNRAS*, **444**, 1518  
 Walt S. v. d., Colbert S. C., Varoquaux G., 2011, *Computing in Science and Engg.*, **13**, 22  
 Wang L., Dutton A. A., Stinson G. S., Macciò A. V., Penzo C., Kang X., Keller B. W., Wadsley J., 2015, *MNRAS*, **454**, 83  
 Wang G.-J., Ma X.-J., Li S.-Y., Xia J.-Q., 2020, *ApJS*, **246**, 13  
 Wilson D., Nayyeri H., Cooray A., Häußler B., 2020, *ApJ*, **888**, 83  
 Boada S., 2019, *MNRAS*, **484**, 4683  
 Xu X., Ho S., Trac H., Schneider J., Póczos B., Ntampaka M., 2013, *ApJ*, **772**, 147  
 Zeiler M., Fergus R., 2014, in Computer Vision, ECCV 2014 - 13th European Conference, Proceedings. Springer Verlag, pp 818–833, doi:10.1007/978-3-319-10590-1\_53  
 Zibetti S., Charlot S., Rix H.-W., 2009, *MNRAS*, **400**, 1181  
 de Diego J. A., et al., 2021, arXiv e-prints, p. arXiv:2108.09415

## APPENDIX A: RECONSTRUCTION WITH FEWER WAVELENGTH BANDS

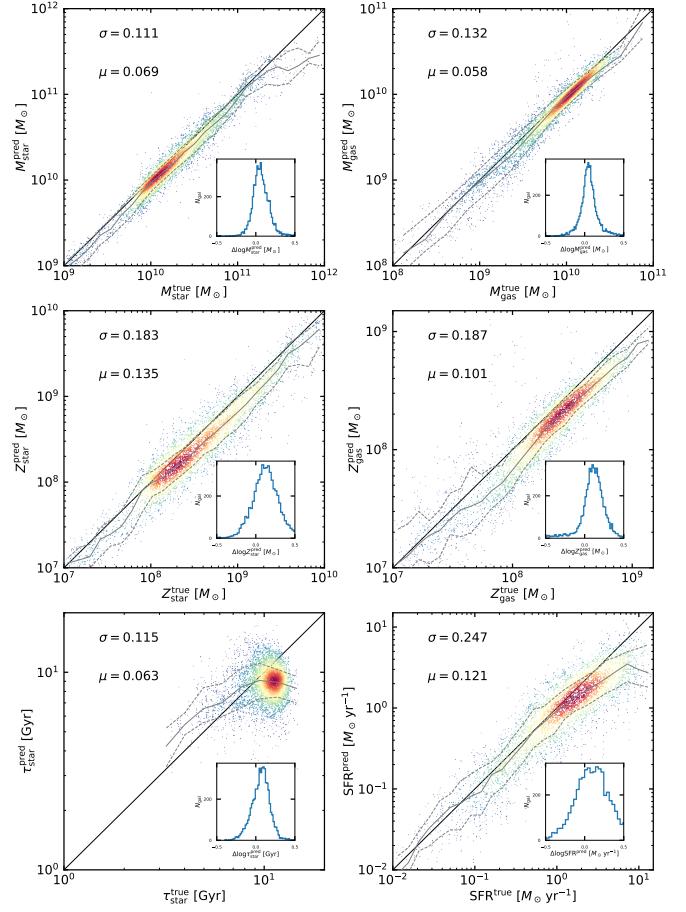
The accuracy of the galaxy reconstruction depends on the number of available wavelength bands.

## APPENDIX B: RECONSTRUCTION WITH LOWER IMAGE RESOLUTION

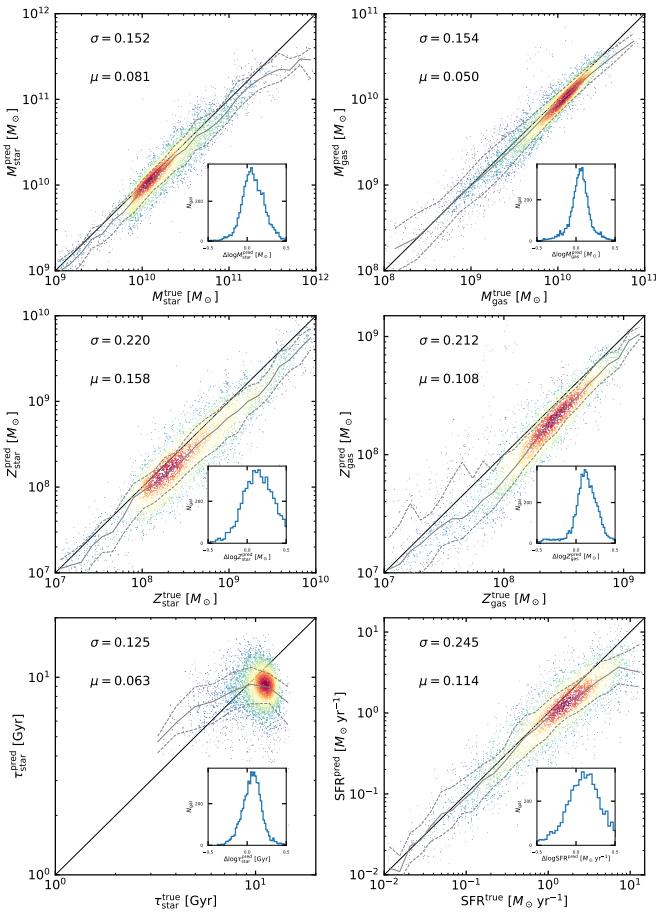
The neural network is able to identify morphological features of galaxies and might be able to use those for the reconstruction of galaxy properties. Thus, the accuracy of the galaxy reconstruction can depend on the image resolution. In Fig. B1 we show a visual comparison between the four different resolutions tested in this work. The upper left panels show the fiducial resolution of 256x256 pixels, the upper right panels have 128x128 pixels and the lower left and right panels have 64x64 and 16x16 pixels, respectively. Additionally, in Fig. B2 we show for the same example galaxy gas properties for the different images resolutions.



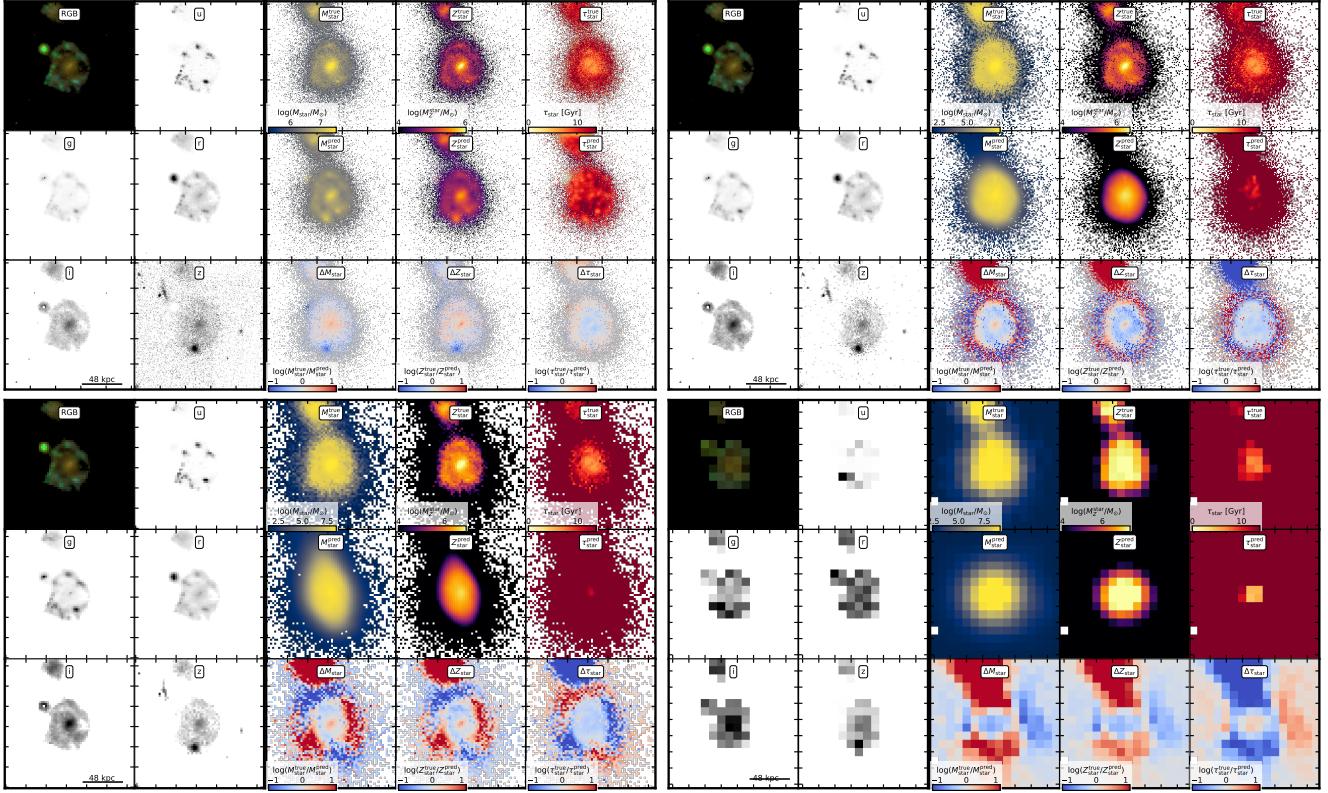
**Figure A1.** Same as Fig. 5 but for training only on the 4 SDSS *ugri* bands.



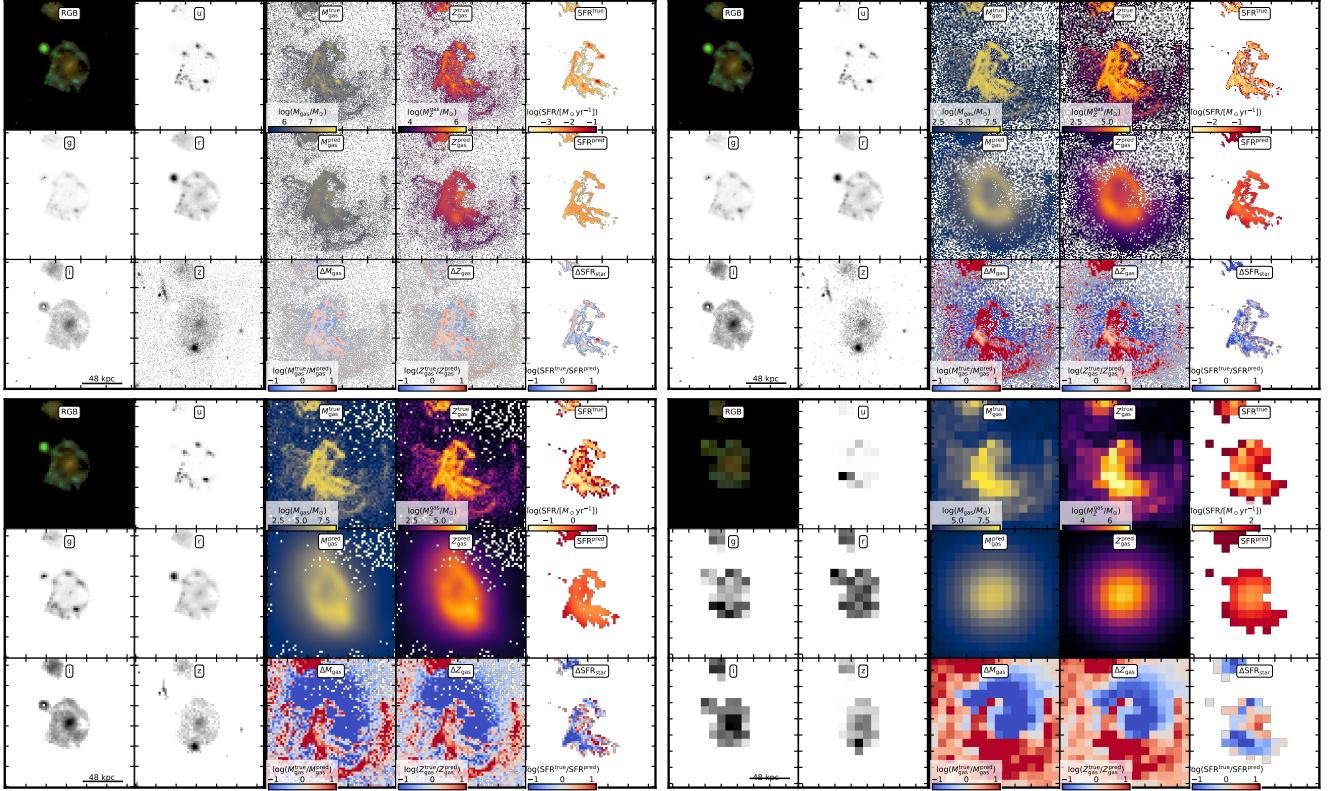
**Figure A2.** Same as Fig. 5 but for training only on the 2 SDSS *gr* bands.



**Figure A3.** Same as Fig. 5 but for training only on the 1 SDSS *r* bands.



**Figure B1.** Same as Fig. 1 but comparing different resolutions. The upper left panels show the fiducial resolution of 256x256 pixels, the upper right panels have 128x128 pixels and the lower left and right panels have 64x64 and 16x16 pixels, respectively.



**Figure B2.** Similar to Fig. 1 but this time showing gas properties instead of stellar properties as indicated by the labels. This figure exemplifies that the more complex gas morphology in comparison to the stellar body is less well predicted if image resolution decreases.