MNRAS 000, 1–8 (2020) Preprint 5 October 2020 Compiled using MNRAS LATEX style file v3.0

# Pix2Prof: fast extraction of sequential information from galaxy imagery via deep learning

Michael J. Smith,[1,2]★ Nikhil Arora,[3] Connor Stone,[3] Stéphane Courteau,[3] and James E. Geach[1,2]
[1]*Centre for Astrophysics Research, Department of Physics, Astronomy & Mathematics, University of Hertfordshire, Hatfield, AL10 9AB, UK*
[2]*Centre of Data Innovation Research, Department of Physics, Astronomy & Mathematics, University of Hertfordshire, Hatfield, AL10 9AB, UK*
[3]*Department of Physics, Engineering Physics, and Astronomy, Queen's University, Kingston, ON, K7L 3N6, Canada*

5 October 2020

**ABSTRACT**

We present "Pix2Prof", a deep learning model that eliminates manual steps in the measurement of galaxy surface brightness (SB) profiles. We argue that a galaxy "profile" of any sort is conceptually similar to an image caption. This idea allows us to leverage image captioning methods from the field of natural language processing, and so we design Pix2Prof as a float sequence "captioning" model suitable for SB profile inferral. We demonstrate the technique by approximating the galaxy SB fitting method described by Courteau (1996), an algorithm with several manual steps. We use $g$, $r$ and $i$-band images from the Sloan Digital Sky Survey (SDSS) Data Release 10 (DR10) to train Pix2Prof on 5367 image–SB profile pairs. We test Pix2Prof on 300 SDSS DR10 galaxy image–SB profile pairs in each of the $g,r$, and $i$ bands to calibrate the mean SB deviation between interactive manual measurements and automated extractions, and demonstrate the effectiveness of Pix2Prof in mirroring the manual method. Pix2Prof processes ∼1 image per second on an Intel Xeon E5-2650 v3 CPU and ∼2 images per second on a NVIDIA TESLA V100 GPU, improving on the speed of the manual interactive method by more than two orders of magnitude. Crucially, Pix2Prof requires no manual interaction, and since galaxy profile estimation is an embarrassingly parallel problem, we can further increase the throughput by running many Pix2Prof instances simultaneously. In perspective, Pix2Prof would take under an hour to infer profiles for $10^5$ galaxies on a single NVIDIA DGX-2 system. A single human expert would take approximately two years to complete the same task. Automated methodology such as this will accelerate the analysis of the next generation of large area sky surveys expected to yield hundreds of millions of targets. In such instances, all manual approaches – even those involving a large number of experts – would be impractical.

**Key words:** methods: data analysis – methods: statistical – galaxies: photometry

## 1 INTRODUCTION

Large astrophysical surveys such as the Sloan Digital Sky Survey (York et al. 2000, hereafter SDSS), the Panoramic Survey Telescope and Rapid Response System (Chambers et al. 2016, hereafter Pan-STARRS), the Hyper Suprime Cam (Aihara et al. 2017, hereafter HSC) Subaru Strategic Program Survey, or the upcoming Vera C. Rubin Observatory Legacy Survey of Space and Time (Ivezić et al. 2019, hereafter LSST), carry an inherent scaling problem. SDSS has observed over 35% of the sky, cataloguing over 1 billion astronomical objects (Ahumada et al. 2019) with a data rate of raw multi-band imagery approaching 200 GB per night. In January 2019, Pan-STARRS second Data Release totalled 1.6 PB of imaging data. A precursor to LSST, HSC's 990 megapixel camera has already produced over 1 PB of imaging data (Aihara et al. 2019). These surveys will be dwarfed by the upcoming LSST project. LSST's 3.2 gigapixel camera will be the largest ever made, and will survey the entire visible sky twice per week, generating ∼500 PB of imaging data over its decade-long mission. The "firehose" of data from surveys such as LSST will require correspondingly efficient and fully automated procedures to curate and analyse the data, enabling new astrophysical findings and making unanticipated discoveries.

In this study, we are concerned with the automated direct analysis of galaxy imagery towards estimating galaxy properties such as size, luminosity, colour and stellar mass. To calculate these properties, one typically applies a photometric analysis that involves extracting and characterising the spatial distribution of a galaxy's light, described by a surface brightness (SB) profile. The galaxy structural parameters as reflected by the SB profile can be used to infer a suite of other important characteristics such as light concentration, age, star formation rate, and assembly history (e.g. Strom et al. 1976; Bell et al. 2003; Shen et al. 2003; Bernardi et al. 2005; Fernández Lorenzo et al. 2013; Trujillo et al. 2020).

Numerous catalogues of galaxy structural properties already exist (Jedrzejewski 1987; Courteau 1996; Brinchmann et al. 2004; Blanton et al. 2005; Hall et al. 2012; Gilhuly & Courteau 2018). Unfortunately, the methods used in these compilations are either time consuming, requiring human supervision, or fast but unreliable since they often make flawed *a priori* assumptions about a galaxy's disc profile shape and other features. Even with semi-automated methods, the accurate extraction of *all* the useful information from existing surveys would still take years. With the data volume expected to grow

★ E-mail: m.smith28@herts.ac.uk

significantly in the coming years, this becomes an intractable problem. Of great concern is the possibility that important discoveries and insight could be missed or delayed significantly due to the technical challenges imposed by the unprecedented data volume. Clearly, there is a pressing need for entirely new and efficient automated methods that significantly reduce, and ideally circumvent, human interactions. Machine learning is ideally suited for this task, and we apply it in this paper towards the specific problem of extracting SB profiles from multi-band imaging data. Our approach takes advantage of a set of SB profiles already determined via classical, interactive methods (Courteau 1996; Gilhuly & Courteau 2018). We describe the classical method used to produce the training dataset in the next section. The remainder of the paper is organised as follows: Section 2 introduces our approach; our results and validation are presented in Section 3; Section 4 addresses our global findings, and concludes with suggestions for broader application of the algorithm.

## 2 METHOD

### 2.1 The classical surface brightness profile extraction algorithm

In the surface photometry of galaxies (e.g. Courteau 1996, and references therein), the spatially-resolved light profile of a galaxy is extracted by fitting progressively larger isophotes about a common centre. The fitting technique assumes that projected isophotes are well-represented by ellipses. A galaxy's centre is found by identifying the brightest pixel in a manually selected region. Given a manually defined galaxy centre, the classical algorithm determines the parameters needed to define each ellipse via a least squares optimisation. The algorithm then generates isophotal solutions at each radius well into the faint outskirts of the galaxy. In these regions of lower signal-to-noise, where fitting algorithms are challenged, the algorithm radially extends the last fitted isophote in the previous operation with a set of concentric isophotes out to an arbitrarily large radius, usually taken to be the edge of the image.

The isophotes may vary in ellipticity and position angle as a function of galactocentric radius. This can become problematic when fitting to non-axisymmetric structures in galaxies, such as bars and spiral arms that can cause large twists in the fitted isophotal map. This issue can be corrected by manually applying a smoothing function to some portions of the image. The latter consists of manually smoothing the contour fits (i.e. uncrossing twisted isophotes), and replacing poorly fitted data with a polynomial smoothing function. Note that, prior to applying isophotal fitting to galaxy images, some pre-processing is also required: the galaxy centre must be identified as described above; the "sky" background must be estimated and removed from the image; nuisance foreground objects (such as unassociated galaxies or foreground stars) must be identified and masked. These steps add to the manual supervision of the task.

Besides the assumption that galaxies are circular when viewed face-on, and thus generally of elliptical appearance when projected onto the plane of the sky, the algorithm purposefully avoids using *a priori* knowledge of galactic disc profile shapes and other features such as bars, rings, and spiral arms. This avoids biasing the isophotal solution to any pre-determined, and possibly incorrect, shape – a problem especially acute in the faint outer edges of a galaxy.

While the semi-automated steps outlined above yield high quality SB profiles, the process of obtaining a single profile is slow and systematic variations may exist between different profile extraction methods. The interactive nature of certain steps may indeed give rise to marked profile differences, especially in low SB regimes where the isophotal solutions are less robust. The low SB regimes will always remain the bane of galaxy image analysis, whether automated or interactive, but the elimination of subjective steps goes a long way towards reducing systematic differences between profiles. Therefore, it becomes desirable to eliminate all interactive steps while retaining all the benefits of classic algorithms such as Courteau (1996) described above. In this work, we present a fully automated solution that incorporates the extant knowledge base of SB profile fitting methodology, but avoids human interaction.

**Table 1.** A summary of the Courteau (1996) surface brightness profile fitting algorithm's processes. An approximate wall time per galaxy is given for the manual sections. The automated sections' time contributions are negligible.

| Process | Automated? | Wall time (s/gal) |
|---|---|---|
| Identify galaxy centre | No | 5 |
| Estimate & remove sky background | Yes | – |
| Remove foreground objects | No | 120 |
| Fit contours | Yes | – |
| Extend contours to galaxy extent | No | 30 |
| Smooth isophotes | Yes | – |
| Interpolate poorly fitted data | No | 120 |

### 2.2 Borrowing from automated image captioning

In recent years, the field of automated image captioning has benefited greatly from advances in deep learning. We were strongly influenced by these developments when designing the architecture of our "Pix2Prof" profile estimator. In this section, we briefly review pure recurrent neural network (hereafter RNN) based encoder–decoder architectures, or models that only use a single encoder and decoder to generate captions. A comprehensive review on deep learning methods applied to image captioning can be found in Hossain et al. (2018).

We primarily draw inspiration from gated RNN based encoder–decoder architectures, as seen in Sutskever et al. (2014) for sequence-to-sequence translation, and in Vinyals et al. (2014), Jia et al. (2015), and Wang et al. (2016) for image-to-sequence translation. Sutskever et al. (2014) uses a Long Short Term Memory (Hochreiter & Schmidhuber 1997, hereafter LSTM) network to encode a given sentence to a latent descriptive vector, and a second LSTM network to decode the descriptive vector into a different feature space. One can use this technique to translate text between two different languages, for example.

Vinyals et al. (2014), Jia et al. (2015), and Wang et al. (2016) all use a convolutional neural network (Fukushima 1980; LeCun et al. 1989, hereafter CNN) to first encode an image to a latent descriptive vector, and then use an LSTM network to decode this vector into a text description (caption) of a given image. Xu et al. (2015) use a CNN encoder, and a LSTM that attends over the CNN output. Attention allows this approach to link each word in the caption with an associated part of the image. This approach works well for images that are crowded with multiple objects, but a simpler approach is preferred for our case where each image is dominated by a single, central galaxy.

A galaxy profile can be thought of as analogous to a text caption describing that galaxy. Both a text caption and galaxy profile can be encoded as a list of floats or integers, and both have a length and content dependent on the context of the conditioning image. Both also need to terminate once a complete sentence or profile is generated. Again, this is a subjective task well suited to a machine learning solution. Additionally, galaxy profiles and text captions can

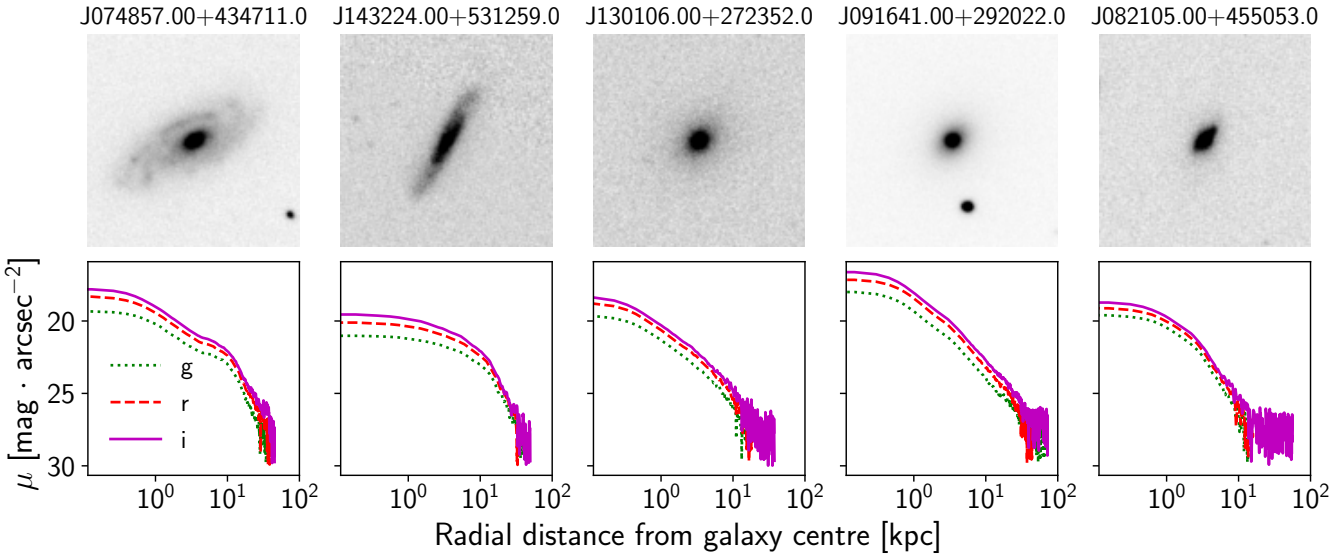


**Figure 1.** SDSS images of sample galaxies in the $g$ band (top row), and corresponding surface brightness "ground truth" profiles (bottom row). $\mu$ is the surface brightness. The galaxy names above each image refer to their J2000 celestial coordinate.

both be approximated with an appropriate RNN. With this in mind, it is natural to consider an encoder–decoder network for the specific task of estimating challenging galaxy profiles. Importantly, since the proposed model directly learns the transformation between an unprocessed galaxy image, and the galaxy's corresponding SB profile, it eliminates all of the manual steps described in §2.1 and Table 1.

While we develop Pix2Prof within the context of galaxy profile extraction, the model is equally applicable to any array → float sequence translation task.

### 2.3 Training set

We exploit galaxy imaging from the SDSS DR10 (York et al. 2000; Eisenstein et al. 2011; Ahn et al. 2014) in the $g$, $r$, and $i$-bands. We pair these images with manually-extracted SB profiles calculated using the Courteau (1996) method described in §2.1. Figure 1 presents a random sample of training set galaxies, and their corresponding, manually extracted SB profiles. The 1953 galaxy image–SB profile pairs in each of the $g$, $r$, and $i$ bands yield a total of 5859 pairs. This full dataset is then divided into training, validation, and testing sets. There are 5367 galaxy image–profile pairs in the training set, 192 galaxy image–profile pairs in the validation set, and 300 galaxy image–profile pairs in the test set. The sets are randomly assigned, with the condition that a given galaxy's three photometric bands are kept within the same set. The subset sizes are chosen to maximise the training set efficacy while retaining most of the training set variance in the test set.

The only destructive pre-processing performed on the galaxy imagery is a 99.9th percentile clipping. This clipping mitigates the issue of single bright (i.e. "hot") pixels reducing image contrast when the galaxy images are normalised, which would reduce training efficacy. To this end, we apply a fixed min-max normalisation, defined as

$$\bar{x} = \frac{x - A}{B - A}, \tag{1}$$

where $A = 2.0$ nanomaggies is the floor of the minimum value in the training set, and $B = 30.0$ nanomaggies is the ceiling of the 99.9th percentile value in the training set.

We crop the galaxy images to a shape of `[256, 256]` pixels and train using the full 32-bit depth of the original data as measured. Good quality data is paramount when training a neural network, and we therefore cut the profiles when the signal-to-noise ratio reaches a quality threshold. We terminate the profile when the signal-to-noise ratio of a 1D convolution with length 40 reaches a threshold of 4. We define signal-to-noise so that it is equivalent to the ratio of the power of a signal to the power of background noise: $(\mu/\sigma)^2$, where $\mu$ is the mean and $\sigma$ is the standard deviation of the convolutional window.

### 2.4 Network architecture

We write our model in PyTorch (Paszke et al. 2019), using a ResNet-18 (He et al. 2015; Srivastava et al. 2015) encoder, and a Gated Recurrent Unit (Cho et al. 2014, hereafter GRU) decoder architecture. This architecture takes an arbitrarily sized single channel image input, and outputs a sequence of floats of arbitrary length. The same network is trained on images in the $g$, $r$, and $i$ bands, and therefore can produce a SB profile in any one of these bands. Figure 2 shows a representation of the architecture used.

We use the standard ResNet-18 architecture as described in He et al. (2015). The GRU is stacked to three layers. We apply a rectified linear unit (Glorot et al. 2011, hereafter ReLU) activation and a dense neural layer after the three layer stack to reduce the number of output values to one. As a regularising measure, we apply dropout at a 20% rate (Srivastava et al. 2014). The ResNet first encodes the incoming galaxy image to a latent space vector $z$ of length 512. This vector is then used as the initial hidden state $h_0$ of a GRU. In this way Pix2Prof encodes and passes relevant information from the image to the GRU. The GRU then unrolls to estimate properties of the galaxy from $z$. In this paper's case, we demonstrate this process by using $z$ to estimate a galaxy's SB profile.

To start estimation, the GRU is fed a start of sequence token. This token is set as an array of zeros. In place of an end of sequence token, the GRU is programmed to halt after 100 predictions are output that

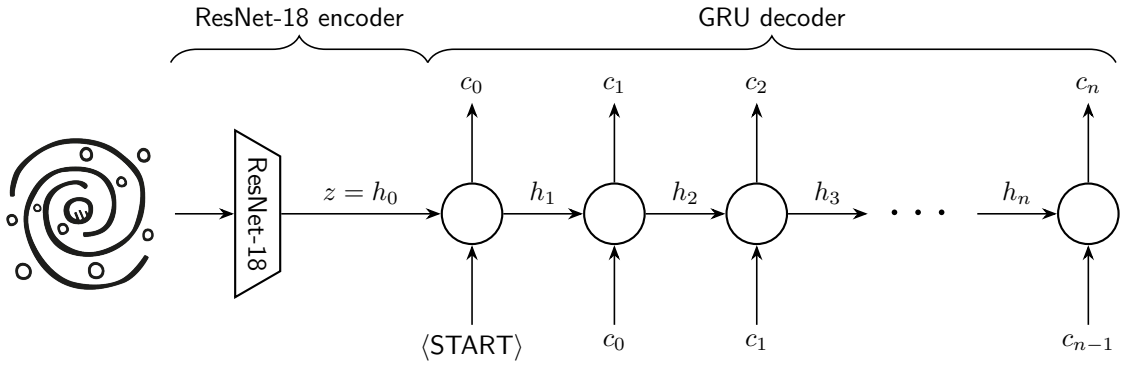


**Figure 2.** The ResNet → GRU encoder-decoder architecture used in this work. The hidden state $h_i$ is the internal state of the GRU, and is dependent on both the galaxy latent encoding $z$, and the previous profile predictions $c_i$.

have a standard deviation of 0.01 or less. This ensures that the GRU halts estimation once it encounters the background sky.

We use the Adam optimiser (Kingma & Ba 2014) to train Pix2Prof via gradient descent (Robbins & Monro 1951). Using manual search, we set the learning rate as $2 \times 10^{-4}$. Due to the logarithmic nature of magnitude, we want to penalise large deviations from our ground truth SB profiles at a higher rate compared to small deviations, and so we use the mean squared error loss:

$$\text{MSE} = \frac{1}{b}\sum_{i=1}^{b}(y_i - p_i)^2, \tag{2}$$

where $b$ is the batch size, $y$ is the ground truth, and $p$ is a prediction.

### 2.5 Training the model

We augment the galaxy images by applying a "wobble". This wobble is a random small shift in the centre of the image. Each band is treated independently. We do this to encourage the network to work with the slightly off-centre galaxies that will be encountered in real data. This is required to make Pix2Prof robust to poorly centred galaxy images. We also exploit the rotational axisymmetry of galaxies and further augment the data by randomly rotating an input image through 90, 180, and 270 degrees.

We train the model for 100,000 global steps on a single NVIDIA TESLA V100 GPU. Training takes approximately 20 minutes per epoch of 500 galaxy images, a rate of 0.4 galaxies per second.

## 3 RESULTS & VALIDATION

We validate the model during training once per epoch using the validation set. We test the trained model on 100 randomly sampled observed galaxies in the $g$, $r$, and $i$ bands (for 300 total image–profile pairs) drawn from the dataset and which are set aside entirely during training. We use the model with the lowest validation loss; epoch 160. We run an entirely automated inference on an Intel Xeon CPU E5-2650 v3 at a rate of 0.9 galaxies per second.

Figure 3 shows a random selection of 25 Pix2Prof inferred test set SB profiles superimposed onto the Courteau (1996) SB profiles. Figure 4 shows the error distribution of the test set as well as the test set error per distance in physical units from the galaxy centre. We define error as the absolute of Figure 3's residual, the absolute deviation:

$$\eta = |y - p|, \tag{3}$$

where $p$ is a prediction, and $y$ is measured via Courteau (1996). The units of SB call for additional care in defining our errors. Since SB values are defined on a logarithmic scale, Equation 3 is really a form of fractional error:

$$\begin{aligned} \eta &= \left| A \log_{10}\frac{I_y}{B} - A\log_{10}\frac{I_p}{B} \right| = \left| A \log_{10}\frac{I_y}{I_p} \right|, \\ \therefore \frac{I_y}{I_p} &= 10^{\eta/|A|}, \end{aligned} \tag{4}$$

where $A$ is a constant ($-\sqrt[5]{100}$) and $B$ is a constant reference brightness. $\{I_p, I_y\}$ are brightnesses in linear units.

We take the median of this error per galaxy profile to produce Figure 4a and Figure 4b, and we take the median of this error across profiles to produce Figure 4c. Figure 4c shows that the error increases with radius away from the galaxy centre towards regions containing less signal, as expected. We find that the median test set absolute deviation is 0.41 mag arcsec$^{-2}$ with an interquartile range of 0.21 mag arcsec$^{-2}$. We also find that the median test set absolute deviation for $y$ values brighter than the SDSS limiting SB (26.5 mag arcsec$^{-2}$) is 0.34 mag arcsec$^{-2}$, with an interquartile range of 0.22 mag arcsec$^{-2}$. Errors of this scale mean that profiles generated via Pix2Prof will be immediately useful for rough searches; it would be possible to categorise galaxies roughly by brightness, isophotal radius, scale length, and other structural parameters. Further refinement of the model may reduce error, enabling more sophisticated processing and analysis of generated SB profiles. Possible refinements are described in §4.

In Figure 5, we separate the three bands' median errors as a function of galactocentric radius. Close to the galaxy, there is little difference in the three bands' median predictions. However, as we proceed outwards, the $r$-band's error is higher than the $g$-band's, and the $i$-band's error is higher still. This is likely due to a difference in the instrumental noise between the three bands, as evidenced in the difference in the spectral bands' median galaxy image signal-to-noise ratios: $\text{SNR}_g = 41.6$; $\text{SNR}_r = 35.8$; $\text{SNR}_i = 28.4$.

Figure 3 and Figure 4 show that Pix2Prof can successfully approximate a complicated astrophysical image processing pipeline with low deviation (0.34 mag arcsec$^{-2}$ averaged over the test set). Processing 0.9 galaxies per second on an Intel Xeon E5-2650 v3 CPU, Pix2Prof improves on the speed of the classical image analysis method of Courteau (1996) by more than two orders of magnitude. For comparison, an astronomer trained to use the Courteau (1996) method can typically process ~150 galaxies in a full eight hour working day (or ~0.005 galaxies per second). Regrettably however, even astronomers

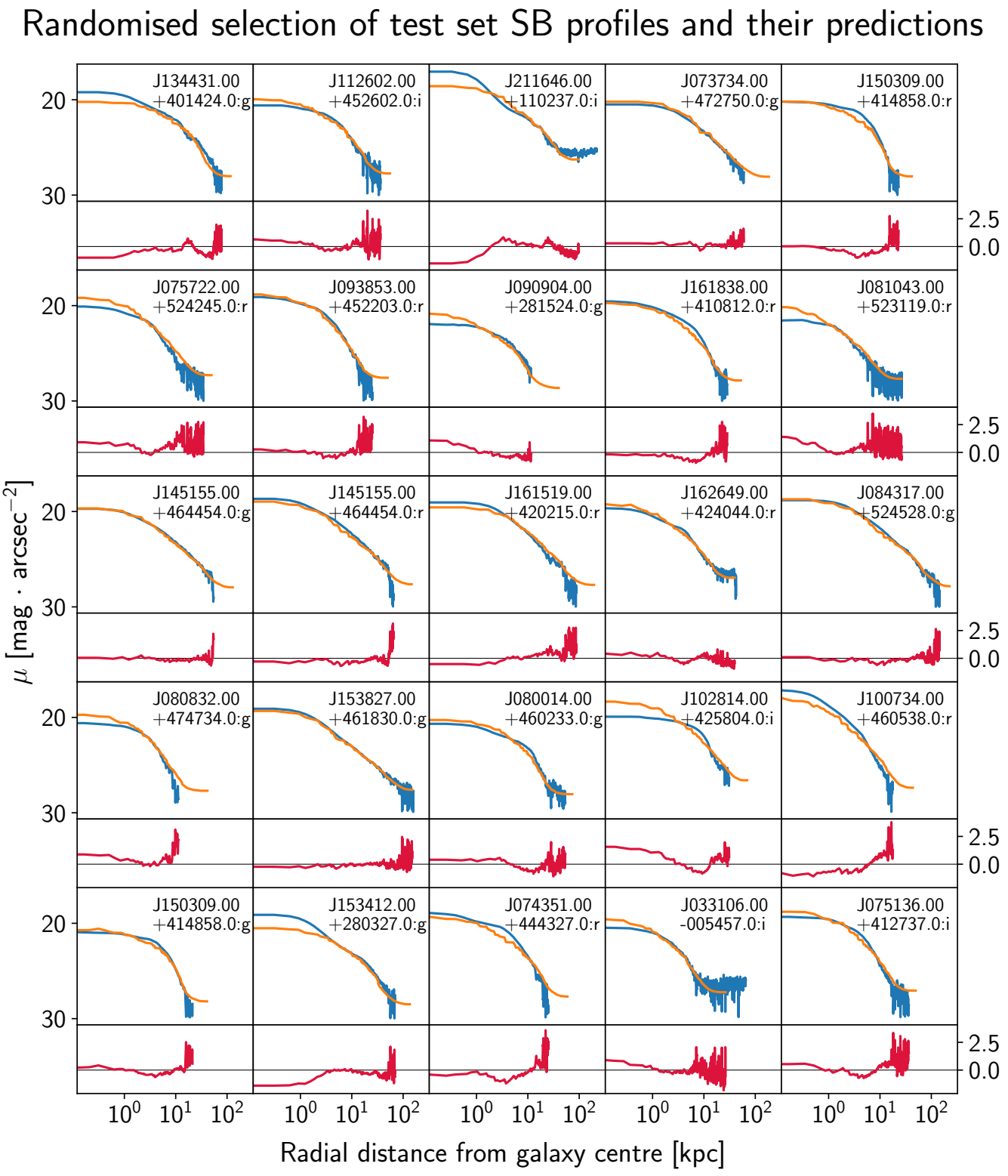


**Figure 3.** Randomly sampled test set predicted SB galaxy profiles (orange) superimposed onto SB profiles measured via the Courteau (1996) method (blue). $\mu$ is the surface brightness. Distances from centre are in log scale to emphasise divergences in the high signal-to-noise ratio region closer to the galaxies' centres. Below each SB profile plot is the residual defined as res = $y - p$, where $y$ is the profile as measured according to §2.1, and $p$ is the prediction. The SDSS name and spectral band are indicated at the top right of each graph.

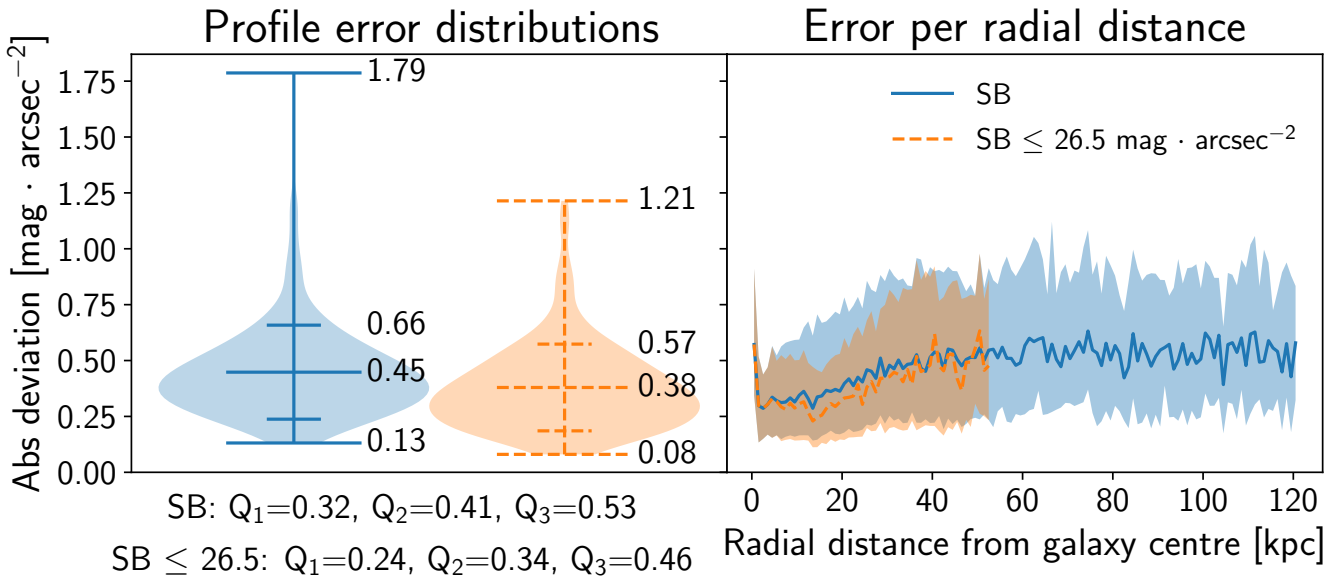


**Figure 4.** Approximation errors as defined in Equation 3. The leftmost violin plot (*4a*) is the distribution of median test set errors. The rightmost violin plot (*4b*) shows the same distribution for only SB values below the SDSS limiting SB of 26.5 mag arcsec$^{-2}$. The maximum, minimum, mean, and (mean + standard deviation) are labelled. Below the violin plots are their distribution quartiles. The right panel (*4c*) shows the median error per kpc from the galaxy centre, with the interquartile range shaded. To reduce the effect of small sample size variability, the profiles in Fig. 4c are terminated once 90% of the SB profiles reach their galaxy's extent.

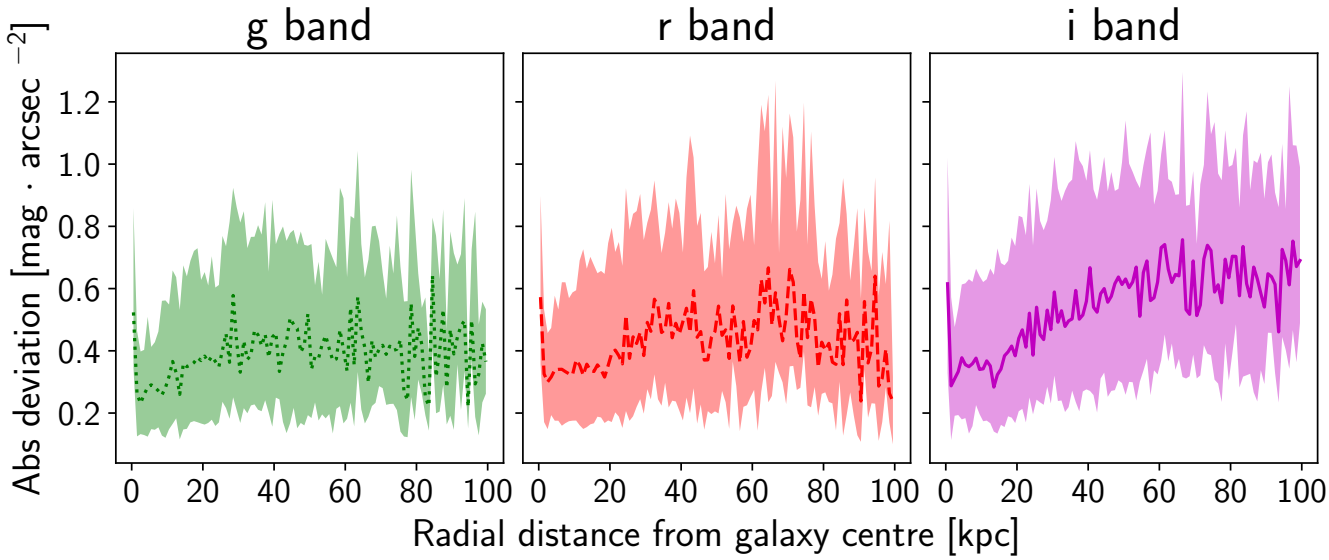


**Figure 5.** Median test set error per kpc from the galaxy centre, with the interquartile range shaded, split into the three bands present in the test set.

must rest and so the true working rate for a human would be ~150 galaxies per 24 hours, or ~0.002 galaxies per second.

As Table 2 shows, Pix2Prof eliminates any manual interaction from SB profile inference, alleviating the issue of subjectivity in the different methods developed for such tasks; Pix2Prof will infer the same profile every time for a given galaxy image, whereas a human may not. The full automation of Pix2Prof enables a complete parallelisation, and thus significant gain in parallel throughput of galaxy profile estimation.

## 4 DISCUSSION AND CONCLUSIONS

While Pix2Prof can rapidly and accurately produce profiles of arbitrary length, there are some limitations to this technique. Principally, any profile produced will be biased to the training set. For instance, if Pix2Prof is trained on primarily nearby galaxies, it may not yield accurate profiles for more distant systems whose images will be poorly resolved. Similarly, if the model is trained on galaxy image–profile pairs as produced by numerical simulations the model will encode any flaws, incompleteness, or bias inherent to each simulation

**Table 2.** Pix2Prof eliminates all interactive steps in the Courteau (1996) algorithm, alleviating subjectivity and speeding up inference significantly.

| Process | Automated in: Courteau (1996)? | Pix2Prof? |
|---|---|---|
| Identify galaxy centre | No | Yes |
| Estimate & remove sky background | Yes | Yes |
| Remove foreground objects | No | Yes |
| Fit contours | Yes | Yes |
| Extend contours to galaxy outskirts | No | Yes |
| Smooth isophotes | Yes | Yes |
| Interpolate poorly fitted data | No | Yes |

and will not encode instrumental effects (e.g. read-out noise) unless properly included. The same issue will occur if we train on galaxy image–profile pairs sampled from one survey and deploy the trained model on a dissimilar survey, for example SDSS, and LSST (York et al. 2000; Ivezić et al. 2019). It may be possible to mitigate this problem with an image domain translator (i.e. Zhu et al. 2017; Isola et al. 2016; Choi et al. 2017) that could transform observations so that they match a given survey. Of course, the Courteau (1996) measured profiles may also not entirely reflect the "true" SB profile, due to modelling assumptions, human bias, and inherent noisiness in measurement. As neural networks typically require very large datasets, our relatively small dataset is likely not reflecting the true potential of the model. Therefore, a larger set of training data could improve the results presented here. Generating a larger dataset from simulated galaxies for training Pix2Prof will be a future project.

As described in Jia et al. (2015), due to the vanishing gradient problem a LSTM or GRU may "forget" an image encoding as it unrolls. For Pix2Prof, this will manifest in a loss of accuracy at larger galactocentric radius. We see this effect in Figures 4c and 5, but we cannot disentangle the individual contributions from image noise and the model architecture. However, assuming that the noise is significantly caused by GRU "forgetfulness", future Pix2Prof models could imitate Jia et al. (2015) and counteract the noise by reinjecting the image encoding into the GRU's hidden state periodically as it unrolls. Another solution could involve adopting an architecture that suffers less from the vanishing gradient problem, such as the Transformer (Vaswani et al. 2017). The non-sequential nature of a Transformer would also allow us to parallelise output at inference time, reducing processing time even further.

In §1 we stressed the need for efficient and fully automated methods for timely analysis of ultra-large scale astrophysical imaging survey data. We believe that Pix2Prof addresses this challenge. Pix2Prof can predict any galaxy profile, given the right simulated or observed dataset. Training Pix2Prof on simulated galaxy images offers additional benefits; the model could be trained on information that is only inferred indirectly in observations. For instance, Pix2Prof could train on sets of galaxy image–mass profile pairs directly in order to predict dark matter halo profiles, as mass profiles cannot be recovered classically by direct imaging observations. Furthermore, Pix2Prof has the potential to automate any galaxy profile fitting routine and be ported to other forms of galaxy image analysis that may not rely on isophotal analysis, but still produce a float sequence given a multi-dimensional array. These analyses could include galaxy component decompositions, the characterisation of galaxy interactions and distortions, pixelised stellar population synthesis, inference of galaxy mass distributions, and more (e.g. Eneev et al. 1973; Vazdekis 2001; Peng et al. 2002). In a future paper, we will demonstrate how Pix2Prof can be used to recover simultaneously the galaxy surface brightness profile as well as the ellipticity profile and curve of growth of a galaxy.

An exciting future investigation involves building a system that can predict properties of unseen classes of objects. This could be achieved by building up a "prior" that encodes known objects into a latent space and interpolates between their latent spatial representations at inference time. A generative model like the Generative Adversarial Network (Goodfellow et al. 2014, GAN) or Variational Autoencoder (Kingma & Welling 2013, VAE) could achieve this (i.e. Spindler et al. 2020). Such a model could quickly identify astrophysically "interesting" objects in a large field survey. The ability to search for rare objects in large unstructured datasets will become increasingly more important as new large scale astronomical surveys come online (Chambers et al. 2016; Aihara et al. 2017; Ivezić et al. 2019).

The training of machine learning models requires considerable energy, contributing to carbon emission and therefore climate change (Strubell et al. 2019; Lacoste et al. 2019). The energy used while training Pix2Prof on a single NVIDIA V100 GPU is estimated to be ~20 kWh (5.54 kg $CO_2$ eq) according to the Machine Learning Impact calculator described in Lacoste et al. (2019). For comparison, an economy class round trip flight from London to New York City would emit 670 kg $CO_2$ eq[1]. To counteract further emission from redundant retraining, we follow the recommendations of Strubell et al. (2019) and make available the fully trained model, as well as the code to run it. Also, we will make available trained models for any improvements that we make to Pix2Prof in the future.

In summary, we have introduced a fully automated deep learning model for the extraction of sequential data from galaxy imagery. We have tested this model by applying it to the specific problem of estimating galaxy surface brightness profiles, a process that previously required manual, time-consuming human intervention. We have tested our model on unseen galaxy images and found that our model has an average absolute deviation of 0.34 mag arcsec$^{-2}$ with an interquartile range of 0.22 mag arcsec$^{-2}$, while inferring surface brightness profiles over two orders of magnitude faster than the classic (interactive) algorithm it automates. This is quite remarkable.

## ACKNOWLEDGEMENTS

This research made use of the University of Hertfordshire's High Performance Computing facility (http://uhhpc.herts.ac.uk/). We are also grateful to the Natural Sciences and Engineering Research Council of Canada, the Ontario Government, and Queen's University for critical support through various scholarships and grants. J.E.G. thanks the Royal Society for support. We also thank reviewers at the NeurIPS 2020 conference for helpful comments and suggestions.

## DATA AND CODE AVAILABILITY

The code and trained model used in this paper is available at https://github.com/Smith42/pix2prof. The profile dataset used to train the network will be released separately (Arora et al. 2021, *in prep*).

## REFERENCES

Ahn C. P., et al., 2014, Astrophysical Journal Supplement Series, 211, 17

[1] According to the International Civil Aviation Organization (ICAO).

Ahumada R., et al., 2019, arXiv e-prints, p. arXiv:1912.02905
Aihara H., et al., 2017, Publications of the Astronomical Society of Japan, 70
Aihara H., et al., 2019, Publications of the Astronomical Society of Japan, 71
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Bernardi M., Sheth R. K., Nichol R. C., Schneider D. P., Brinkmann J., 2005, AJ, 129, 61
Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005, The Astrophysical Journal, 631, 208
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, Monthly Notices of the Royal Astronomical Society, 351, 1151
Chambers K. C., et al., 2016, arXiv e-prints, p. arXiv:1612.05560
Cho K., van Merrienboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y., 2014, preprint (arXiv:1406.1078)
Choi Y., Choi M., Kim M., Ha J.-W., Kim S., Choo J., 2017, arXiv e-prints, p. arXiv:1711.09020
Courteau S., 1996, Astrophysical Journal Supplement Series, 103, 363
Eisenstein D. J., et al., 2011, The Astronomical Journal, 142, 72
Eneev T. M., Kozlov N. N., Sunyaev R. A., 1973, A&A, 22, 41
Fernández Lorenzo M., Sulentic J., Verdes-Montenegro L., Argudo-Fernández M., 2013, MNRAS, 434, 325
Fukushima K., 1980, Biological Cybernetics, 36, 193
Gilhuly C., Courteau S., 2018, Monthly Notices of the Royal Astronomical Society, 477, 845
Glorot X., Bordes A., Bengio Y., 2011, in Gordon G., Dunson D., Dudík M., eds, Proceedings of Machine Learning Research Vol. 15, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, Fort Lauderdale, FL, USA, pp 315–323, http://proceedings.mlr.press/v15/glorot11a.html
Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, in Ghahramani Z., Welling M., Cortes C., Lawrence N. D., Weinberger K. Q., eds, , Advances in Neural Information Processing Systems 27. Curran Associates, Inc., pp 2672–2680, http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf
Hall M., Courteau S., Dutton A. A., McDonald M., Zhu Y., 2012, Monthly Notices of the Royal Astronomical Society, 425, 2741
He K., Zhang X., Ren S., Sun J., 2015, arXiv e-prints, p. arXiv:1512.03385
Hochreiter S., Schmidhuber J., 1997, Neural Computation, 9, 1735
Hossain Z., Sohel F., Shiratuddin M. F., Laga H., 2018, arXiv e-prints, p. arXiv:1810.04020
Isola P., Zhu J.-Y., Zhou T., Efros A. A., 2016, arXiv e-prints, p. arXiv:1611.07004
Ivezić Ž., et al., 2019, The Astrophysical Journal, 873, 111
Jedrzejewski R. I., 1987, MNRAS, 226, 747
Jia X., Gavves E., Fernando B., Tuytelaars T., 2015, arXiv e-prints, p. arXiv:1509.04942
Kingma D. P., Ba J., 2014, arXiv e-prints, p. arXiv:1412.6980
Kingma D. P., Welling M., 2013, arXiv e-prints, p. arXiv:1312.6114
Lacoste A., Luccioni A., Schmidt V., Dandres T., 2019, arXiv e-prints, p. arXiv:1910.09700
LeCun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D., 1989, Neural Computation, 1, 541
Paszke A., et al., 2019, in Wallach H., Larochelle H., Beygelzimer A., d Alché-Buc F., Fox E., Garnett R., eds, , Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp 8024–8035, http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266
Robbins H., Monro S., 1951, The Annals of Mathematical Statistics, 22, 400
Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978
Spindler A., Geach J. E., Smith M. J., 2020, arXiv e-prints, p. arXiv:2009.08470
Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R., 2014, Journal of Machine Learning Research, 15, 1929
Srivastava R. K., Greff K., Schmidhuber J., 2015, arXiv e-prints, p. arXiv:1505.00387
Strom S. E., Strom K. M., Goad J. W., Vrba F. J., Rice W., 1976, ApJ, 204, 684
Strubell E., Ganesh A., McCallum A., 2019, arXiv e-prints, p. arXiv:1906.02243
Sutskever I., Vinyals O., Le Q. V., 2014, arXiv e-prints, p. arXiv:1409.3215
Trujillo I., Chamba N., Knapen J. H., 2020, MNRAS, 493, 87
Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I., 2017, arXiv e-prints, p. arXiv:1706.03762
Vazdekis A., 2001, Ap&SS, 276, 921
Vinyals O., Toshev A., Bengio S., Erhan D., 2014, arXiv e-prints, p. arXiv:1411.4555
Wang C., Yang H., Bartz C., Meinel C., 2016, arXiv e-prints, p. arXiv:1604.00790
Xu K., Ba J., Kiros R., Cho K., Courville A., Salakhutdinov R., Zemel R., Bengio Y., 2015, arXiv e-prints, p. arXiv:1502.03044
York D. G., et al., 2000, Astronomical Journal, 120, 1579
Zhu J.-Y., Park T., Isola P., Efros A. A., 2017, arXiv e-prints, p. arXiv:1703.10593

This paper has been typeset from a TEX/LATEX file prepared by the author.