Starduster: A multi-wavelength SED model based on radiative transfer simulations and deep learning

Yisheng Qiu\textsuperscript{1} and Xi Kang\textsuperscript{1,2,3}

\textsuperscript{1} Zhejiang University–Purple Mountain Observatory Joint Research Center for Astronomy, Zhejiang University, Hangzhou 310027, China
\textsuperscript{2} Purple Mountain Observatory, 10 Yuan Hua Road, Nanjing 210034, China
\textsuperscript{3} National Basic Science Data Center, Zhongguancun South 4th Street, Beijing 100190, China

ABSTRACT

We present starduster, a supervised deep learning model that predicts the multi-wavelength SED from galaxy geometry parameters and star formation history by emulating dust radiative transfer simulations. The model is comprised of three specifically designed neural networks, which take into account the features of dust attenuation and emission. We utilise the skirt radiative transfer simulation to produce data for the training data of neural networks. Each neural network can be trained using $\sim 4000 \sim 5000$ samples. Compared with the direct results of the skirt simulation, our deep learning model produces $0.1 \sim 0.2$ mag errors in FUV to FIR wavelengths. At some bands, the uncertainty is only $0.01$ mag. As an application, we fit our model to the observed SEDs of IC4225 and NGC5166. Our model can reproduce the observations, and successfully predicts that both IC4225 and NGC5166 are edge-on galaxies. However, the predicted geometry parameters are different from image-fitting studies. Our analysis implies that the inconsistency is mainly due to the degeneracy in the star formation history of the stellar disk and bulge. In addition, we find that the predicted fluxes at $20 \mu m \sim 80 \mu m$ by our SED model are correlated with bulge radius. Our SED code is public available and can be applied to both SED-fitting and SED-modelling of galaxies from semi-analytic models.

1. INTRODUCTION

Theoretical spectral energy distribution (SED) models play an important role in understanding galaxy formation. They have two main applications. The models can be fit to an observed SED to infer galaxy parameters such as dust mass, stellar mass and star formation rate (e.g. da Cunha et al. 2008; Han & Han 2014; Leja et al. 2017). Secondly, we can also apply the models to producing large samples of mock SEDs using galaxy parameters obtained from simulations, and compare the derived statistical properties such as luminosity functions and colour-magnitude relations with observations (e.g. Qiu et al. 2019; Yung et al. 2019; Vogelsberger et al. 2020). The two applications use the theoretical SED models inversely, and therefore can be put into the same framework, leading to a better connection between observations and simulations. For example, the SED code developed by Robotham et al. (2020) was applied to both SED-fitting (Bellstedt et al. 2020) and SED-modelling of galaxies from a semi-analytic model (Lagos et al. 2019, 2020).

There is a gap in the comparisons between observations and simulations via theoretical SED models. Currently, most SED models use phenomenological parameters to calculate dust attenuation and emission, e.g. the slope of the dust attenuation curve and the spectral index of dust emission, which are not direct predictions of galaxy formation simulations. This problem is typically addressed by employing empirical relations, which link the dust parameters to some relevant galaxy properties (e.g. Cullen et al. 2017; Qiu et al. 2019; Hahn et al. 2021). However, the extra degrees of freedom introduced by the empirical relations weaken the constraining power from observations, leading to difficult interpretation of the results. For instance, by adjusting the free parameters in an empirical dust model, Hahn et al. (2021) found that the UV and optical colour-magnitude relations can be reproduced by three hydrodynamical simulations that predict different intrinsic galaxy properties. An additional problem is that multi-wavelength SED models (e.g. da Cunha et al. 2008; Leja et al. 2017; Boquien et al. 2019; Robotham et al. 2020) assume the same dust attenuation curve to estimate both the UV...
to optical attenuation and the energy absorbed by dust. This is unrealistic, since the dust emission spectrum does not depend on inclination angle while the UV to optical attenuation does.

The above issues can be overcome using dust radiative transfer simulations (e.g. Fioc & Rocca-Volmerange 2019; Camps & Baes 2020; Narayanan et al. 2021, see Steinacker et al. 2013 for a review). These simulations estimate dust attenuation and emission by solving the dust radiative transfer equation given the star-dust geometry and dust extinction curves. The dependence of the dust attenuation on inclination angle is explicitly taken into account by these simulations. However, they are computationally expensive, and therefore are not widely used in SED-fitting or producing large mock galaxy catalogues.

The main purpose of this work is to develop a supervised deep learning model that emulates radiative simulations. Our model is trained using data generated by radiative simulations, and predict the FUV to FIR SED from galaxy geometry parameters and star formation history. As complementary, some studies build the model from the opposite direction, i.e. deriving the galaxy and dust properties from a SED. Such models can be found in Lovell et al. (2019); Dobbels et al. (2020); Gilda et al. (2021). This work also gives an example of using the deep learning model in SED-fitting. An application of our SED model on a semi-analytic model will be presented in a future work.

This paper is organised as follows. Section 2 describes our deep learning model. Section 3 demonstrates an application of the model on SED-fitting. Finally, this work is summarised in Section 4. This paper uses concepts from deep learning directly, e.g. the fully connected layer and the multi-layer perceptron. The reader is referred to Goodfellow et al. (2016) for a nice introductory book of deep learning.

The implementation of the deep learning model includes many technical details and hyperparameters. It is tedious to present all of them. The reader is referred to our public code repository at https://github.com/yqiuu/starduster for the details. Our code is built on PYTORCH 1 (Paszke et al. 2019) and named as STARDUSTER.

2. METHODOLOGY
2.1. Geometry model

We start by describing the geometry model of a galaxy, which serves as the input of radiative transfer simulations. We assume that a galaxy is comprised of a stellar disk, a stellar bulge and a dust disk. Their density profiles are given by the following parametric forms:

\[ \rho_{\text{disk}}(r, z) \propto \exp\left(-\frac{r}{r_{\text{disk}}} - \frac{|z|}{h_{\text{disk}}}\right), \]
\[ \rho_{\text{bulge}}(r, z) \propto S_n\left(\frac{\sqrt{r^2 + z^2}}{r_{\text{bulge}}}\right), \]
\[ \rho_{\text{dust}}(r, z) \propto \exp\left(-\frac{r}{r_{\text{dust}}} - \frac{|z|}{h_{\text{dust}}}\right), \]

where \( S_n \) is the Sérsic function:

\[ S_n(s) = -\frac{1}{\pi} \int_0^\infty \frac{dt}{\sqrt{t^2 - s^2}}, \]
\[ I(t) = \frac{b^{2n}}{\pi \Gamma(2n + 1)} \exp(-bt^{1/n}). \]

This geometry model mainly follows De Geyter et al. (2013); Fioc & Rocca-Volmerange (2019), which are frequently used to study observed spiral galaxies (e.g. De Looze et al. 2014; De Geyter et al. 2014, 2015; Viaene et al. 2017).

2.2. The dust attenuation model
2.2.1. Neural network design

Our dust attenuation model is designed to compute the dust obscured stellar luminosity. With the geometry model described in the previous section and by the linearity of the radiative transfer equation, we can express the dust attenuated luminosity as

\[ L_\lambda^{\text{star}} = 10^{-0.4k_\lambda^{\text{disk}}} L_\lambda^{\text{disk}} + 10^{-0.4k_\lambda^{\text{bulge}}} L_\lambda^{\text{bulge}}, \]

where \( L_\lambda^{\text{disk}}, L_\lambda^{\text{bulge}} \) are the intrinsic luminosities, and \( k_\lambda^{\text{disk}}, k_\lambda^{\text{bulge}} \) are the corresponding dust attenuation curves. We note that the dust attenuation curves can be obtained by simulations that contain only one stellar component, which can be run separately and are independent of the intrinsic luminosity. Accordingly, our approach first carries out two sets of simulations that predict \( k_\lambda^{\text{disk}} \) and \( k_\lambda^{\text{bulge}} \), and then uses the resulting data to train two neural networks.

The design of the proposed neural network is illustrated in Figure 1. To understand the motivation of this design, we show some attenuation curves of the disk component as a heat map in Figure 2. These data are produced in the next section. In order to highlight the shape, these curves are normalised so that the value at the shortest wavelength is equal to one. We suggest that these curves can be decomposed into five components, i.e. a monotonically decreasing baseline, a trough, and tree bumps. Our neural network takes into account

---

1 https://pytorch.org/
Table 1. Summary of the input parameters to run the SKIRT dust radiative transfer simulation.

| Symbol | Description | Value | Unit |
|--------|-------------|-------|------|
| θ      | Inclination angle | Beta distribution with α = β = 0.5 \(^a\) | - |
| r\(_{\text{disk}}\) | Stellar disk radius | Log-uniform distribution from -2.0 to 2.5 | kpc |
| h\(_{\text{disk}}/r_{\text{disk}}\) | Stellar disk height scaled by stellar disk radius | Fixed at 0.1 | |
| r\(_{\text{bulge}}\) | Stellar bulge radius | Log-uniform distribution from -0.5 to 1.5 | kpc |
| r\(_{\text{dust}}/r_{\text{disk}}\) | Dust disk radius scaled by stellar disk radius | Log-uniform distribution from -0.7 to 0.7 | - |
| h\(_{\text{dust}}/r_{\text{disk}}\) | Dust disk height scaled by stellar disk radius | Fixed at 0.1 | - |
| Σ\(_{\text{dust}}\) | Dust column density\(^b\) | Log-uniform distribution from 3.5 to 7.5 | M\(_{\odot}\)/kpc \(^2\) |
| η\(_{\text{tot}}\) | Intrinsic bolometric luminosity | Log-uniform distribution from 6 to 14 | L\(_{\odot}\) |
| α\(_{\text{B/T}}\) | Intrinsic Bulge to total luminosity ratio | Beta distribution with α = β = 0.5 | - |

\(^a\) Inclination angle is scaled to \([0, \frac{π}{2}]\) using linear transformation.

\(^b\) Dust column density is defined by \(Σ_{\text{dust}} = m_{\text{dust}}/(2πr_{\text{dust}}^2)\).

Figure 1. Schematic diagram of the dust attenuation neural network described in Section 2.2.1. The neural network is designed to explicitly take into account the features of the attenuation curves generated by the SKIRT radiative simulation. We suggest that these curves can be decomposed into five components, i.e. a monotonically decreasing baseline, a trough, and tree bumps. These features can be seen from Figure 2.

These features explicitly. We can construct a sequence of data that have one of these features in the following way. Firstly, a monotonically increasing sequence can be obtained by applying a cumulative sum to an array of positive data. Then, by changing the sign, we can create a monotonically decreasing sequence. Finally, we can generate a sequence that has only one peak or trough by multiplying a monotonically increasing sequence and a monotonically decreasing sequence together. We introduce a renormalisation module as a final step in the neural network. This step ensures those modules which represent the features of the curves have similar scales.

2.2.2. Data preparation

We utilise the SKIRT 3D Monte Carlo dust radiative transfer simulation (Baes et al. 2011; Camps & Baes 2015, 2020) to generate training and validation data. The SKIRT code provides a variety of geometry models, and is coupled with DUSTEM (Compiéne et al. 2011) to produce the dust emission spectrum, allowing flexible choices on the dust grain mixture.
Figure 2. Heat map of dust attenuation curves produced by the SKIRT radiative simulation. These curves are generated in Section 2.2.2, which are used to train the dust attenuation neural work of the stellar disk. The curves are such normalised that the value at the shortest wavelength is equal to one.

We run the SKIRT radiative transfer simulation using the geometry model described in Section 2.1. In addition to those defined in equation (1) - (3), the input parameters also include dust mass and inclination angle. We sample these parameters according to Table 1.

In addition, we adopt the Draine & Li (2007) dust grain mixture, and launch $10^7$ photon packages for each run. The dust emission luminosity is not calculated for these simulations, since we are only interested in the dust attenuation curves. The output wavelength grid ranges from $0.09\,\mu m$ to $100\,\mu m$, with 300 bins in logarithmic scale.

Data preprocessing is an essential step in the machine learning context. We scale all input parameters into $[-1, 1]$. Specifically, we apply $f(\theta) = 2 \cos(\theta) - 1$ for the inclination angle, and $f(x) = 2 \left( \frac{\log x - \log x_{\text{min}}}{\log x_{\text{max}} - \log x_{\text{min}}} \right) - 1$ for other parameters. The output curves are noisy due to the Monte Carlo nature of the simulations. To reduce the noise, we smooth the attenuation curves using a Savitzky-Golay filter. Furthermore, we exclude the data for $k(\lambda = 100\,\mu m) < 0.01$ since they are extremely noisy and badly behaved. Table 2 shows the number of samples in each data set.

Table 2. Number of samples in the data set generated by this work.

| Data set                  | Total | Included | Training | Validation |
|---------------------------|-------|----------|----------|------------|
| Dust attenuation (disk)   | 7020  | 5850     | 4094     | 1756       |
| Dust attenuation (bulge)  | 7020  | 6007     | 4204     | 1803       |
| Dust emission             | 8100  | 6772     | 4740     | 2032       |
| Integrated validation     | 2048  | -        | -        | 2048       |

To minimise the loss function, we employ an Adam optimiser with a batch size of 256. We run the optimiser for 3000 epochs and change the learning rate dynamically according to Smith & Topin (2019). We train the neural network for four times with different maximum learning rates at $2.5 \times 10^{-4}$, $5 \times 10^{-4}$, $1 \times 10^{-3}$ and $2 \times 10^{-3}$, and choose the best result from these runs.

The training results are illustrated in Figure 3. The median error on the dust attenuation curves is roughly $3 \times 10^{-3}$ mag for both the stellar disk and bulge models. We will discuss the uncertainties in more detail in Section 2.4, where we combine the dust attenuation and emission neural networks.

2.2.4. Classifiers

As mentioned in Section 2.2.2, we exclude some badly behaved samples from the training data, and consequently the obtained neural networks should not be applied to the excluded parameter space. For this reason, we train two classifiers for the disk and bulge models respectively to identify whether the input parameters are in the effective region. Each classifier is a multi-layer perceptron with one hidden layer. Both classifiers are trained using the same method as in Section 2.2.3.

2.3. The dust emission model

2.3.1. Neural network

Our dust emission neural network aims to predict the dust emission luminosity. A major difficulty is that the dependence of the dust emission luminosity on the intrinsic SED is non-linear, and therefore the training data must include the intrinsic SED. While the intrinsic SED can be derived from star formation history and simple stellar population library, it is almost impossible to fully sample the star formation history. Fortunately, we have recognised that the dust emission luminosity depends on the intrinsic SED only via the dust absorption fraction, which is a linear function with respect to the intrinsic SED. This is the key point behind the design of our dust emission neural network.

The proposed dust emission neural network is illustrated in Figure 4. To understand our approach, we write the dust emission luminosity as

$$L_{\text{dust}}^\lambda = \eta_{\text{tot}}^\lambda \epsilon_{\text{abs}}^\lambda Y_{\text{dust}},$$

where $\eta_{\text{tot}}^\lambda$ and $\epsilon_{\text{abs}}^\lambda$ are the dust absorption fraction and the intrinsic SED, respectively. The dust absorption fraction is a linear function with respect to the intrinsic SED.
A multi-wavelength SED model

**Figure 3.** Error distributions of the trained dust attenuation models. Left and right panels correspond to the models of stellar disk and bulge respectively. The training approach is described in Section 2.2.3.

**Figure 4.** Schematic diagram of the proposed dust emission neural network described in Section 2.3.1. The model assumes a discrete star formation history parametrised by \( c_{i,k} \). While the neural network works for any \( c_{i,k} \), its training by design can be done without sampling \( c_{i,k} \). In addition, the model uses a Planckian mixture to model the FIR dust emission and a transfer module to describe the MIR dust emission.

with

\[ \eta_{\text{tot}} = \int L_{\text{disk}}^\lambda + L_{\text{bulge}}^\lambda \, d\lambda, \]  

(9)

where \( \eta_{\text{tot}} \) is the intrinsic bolometric luminosity, \( s_{\text{abs}} \) is the dust absorption fraction, and \( Y_{\text{dust}} \) is the normalised luminosity. We first focus on the modelling of the absorption fraction. The linearity of the energy balance equation suggests that the absorption fraction can be decomposed into two parts

\[ s_{\text{abs}}^{\text{tot}} = (1 - \alpha_{B/T})s_{\text{abs}}^{\text{disk}} + \alpha_{B/T}s_{\text{abs}}^{\text{bulge}}, \]  

(10)

with

\[ \alpha_{B/T} = \frac{1}{\eta_{\text{tot}}} \int L_{\text{bulge}}^\lambda \, d\lambda, \]  

(11)

where \( s_{\text{abs}}^{\text{disk}} \) and \( s_{\text{abs}}^{\text{bulge}} \) are the fractions of energy contributed from the stellar disk and bulge but absorbed by dust respectively, and \( \alpha_{B/T} \) is the intrinsic bulge to
total luminosity ratio. We suggest that $\xi_{\text{disk}}$ and $\xi_{\text{bulge}}$ can have the following functional forms

$$\xi_{\text{abs}} = \int A_{\lambda}^{\text{disk}} Y_{\lambda}^{\text{disk}} d\lambda,$$

$$\xi_{\text{abs}} = \int A_{\lambda}^{\text{bulge}} Y_{\lambda}^{\text{bulge}} d\lambda,$$

where $A_{\lambda}^{\text{disk}}$ and $A_{\lambda}^{\text{bulge}}$ are the mean dust attenuation functions. They can be modelled using neural networks. Different from the dust attenuation curves defined in Equation (6), these two functions are independent of inclination angle. In principle, they can be derived from $k_{\lambda}^{\text{disk}}$ and $k_{\lambda}^{\text{bulge}}$. However, this approach is not adopted in this work, since it introduces an intermediate training step, resulting in extra errors. Instead, we make the dust absorption fractions linearly depend on the star formation history parameters by using discrete star formation histories, in which case the normalised intrinsic SED can be written as

$$Y_{\lambda} = \sum_{i} \sum_{k} c_{i,k} \langle Y_{\lambda}^{\text{SSP}} \rangle_{i,k}^{\text{abs}},$$

with

$$\langle Y_{\lambda}^{\text{SSP}} \rangle_{i,k}^{\text{abs}} = \int_{t_i}^{t_{i+1}} Y_{\lambda}^{\text{SSP}} (\tau, Z_i) d\tau,$$

where $c_{i,k}$ is the fractional luminosity in $i$-th metallicity and $k$-th stellar age bins, and $Y_{\lambda}^{\text{SSP}}(\tau, Z)$ is the normalised spectrum of simple stellar population (SSP) with stellar age $\tau$ and metallicity $Z$. Substituting equations (14) and (15) into equations (12) and (13) yields

$$\xi_{\text{abs}} = \sum_{i} \sum_{k} c_{i,k} \xi_{\text{abs}}^{\text{disk}},$$

$$\xi_{\text{abs}} = \sum_{i} \sum_{k} c_{i,k} \xi_{\text{abs}}^{\text{bulge}},$$

with

$$\xi_{\text{abs}}^{\text{disk}} = \int A_{\lambda}^{\text{disk}} \langle Y_{\lambda}^{\text{SSP}} \rangle_{i,k}^{\text{abs}} d\lambda,$$

$$\xi_{\text{abs}}^{\text{bulge}} = \int A_{\lambda}^{\text{bulge}} \langle Y_{\lambda}^{\text{SSP}} \rangle_{i,k}^{\text{abs}} d\lambda,$$

The above equations suggest that once the neural network can predict $\xi_{\text{abs}}^{\text{disk}}$ and $\xi_{\text{abs}}^{\text{bulge}}$, the total absorption fraction can be obtained using the linear combination. Hence, when preparing the training data, there is no need to sample the star formation history of the stellar disk and bulge, i.e. $c_{i,k}^{\text{disk}}$ and $c_{i,k}^{\text{bulge}}$. Instead, these parameters can be provided by the one hot encoding. This significantly reduces the dimensions of the problem.

In terms of the neural network to predict $Y_{\lambda}^{\text{dust}}$, we point out that the dependence of the dust luminosity on the intrinsic SED is only via the absorption coefficient in accordance with the energy balance equation. In other words, we can write

$$Y_{\lambda}^{\text{dust}} = \gamma_{\lambda}^{\text{dust}} (\xi_{\text{abs}}^{\text{disk}}, \xi_{\text{abs}}^{\text{bulge}}, \Phi_{\text{geometry}}),$$

where $\Phi_{\text{geometry}}$ is the parameter set describing the geometry model exclusive of the inclination angle. The functional form of $Y_{\lambda}^{\text{dust}}$ is modelled using a neural network.

As illustrated in Figure 5, we introduce a specific module to reproduce the shape of the dust emission luminosity. The FIR emission is modelled by a Planckian mixture, which is a weighted sum of some Planck distributions with different temperatures. The resulting spectrum is then modified to take into account the other emission features at MIR wavelengths using a transfer module. The design of the transfer module is demonstrated in Figure 4. The key idea behind the module is probability conservation.

2.3.2. Data preparation

To prepare the data, we use SKIRT to perform a set of simulations with both stellar disk and bulge components. We apply the same method as in Section 2.2.2 to sample the required parameters for the simulations. Moreover, each run requires two SSP spectra as input for the disk and bulge components respectively. They are randomly chosen from a library. We construct the library using the Flexible Stellar Population Synthesis model (Conroy et al. 2009; Conroy & Gunn 2010) with a Chabrier (2003) initial mass function. Nebular line and continuum emissions are added following Byler et al. (2017). The raw SSP spectra are gridded using Equation (15) with six stellar age bins and six metallicity bins. The simulations require two additional parameters, i.e. the intrinsic bolometric luminosity $\eta_i^{\text{tot}}$ and bulge to total luminosity ratio $\alpha_{B/T}$. We sample them according to Table 1.

For other settings of the simulations, we assume that dust grains are not in local thermal equilibrium, and use the stochastic heating algorithm to calculate the dust emission luminosity. To be consistent with the dust attenuation model, we adopt the Draine & Li (2007) dust grain mixture. The output wavelength grid uses 500 logarithmic bins and ranges from 0.09 $\mu$m and 10,000 $\mu$m. Finally, we launch $10^7$ photon packages for each run.

We adopt the same method as in Section 2.2.2 to preprocess the input parameters. In addition, we use the classifiers obtained in Section 2.2.4 to exclude badly behaved data. We find that the data exclusion can lead
Figure 5. Error distributions of the trained dust emission model. Left and right panels show the errors of the normalised dust emission spectrum and absorption fraction respectively. The errors are defined by the KL divergences in Equations (22) and (23). The training method is described in Section 2.3.3.

For the best result, we calculate the KL divergences for each sample, and demonstrate the corresponding distributions in Figure 5. The median KL divergence is $\sim 7 \times 10^{-4}$ for the normalised emission spectra and $\sim 4 \times 10^{-5}$ for the dust absorption fraction. A more intuitive analysis on the uncertainties will be carried out in the next section, where we combine the dust attenuation and emission models.

2.4. Integrated validation

In this section, we integrate the dust attenuation and emission models to estimate the overall errors. To generate the validation data, we sample the fractional luminosity $c_{\text{disk}}^{i,k}$ and $c_{\text{bulge}}^{i,k}$ using a Dirichlet distribution with a constant consecration parameter of 0.5. The sampling of other parameters follows Table 1. To be consistent, we run the SKIRT simulation using the same configuration as in Section 2.2.2 and Section 2.3.2. Our validation set contains 2048 samples. Figure 6 illustrates nine example samples. The SEDs predicted by SKIRT and our deep learning model are shown as black and red lines respectively. Overall, we find good consistency between them.

In Figure 7, we demonstrate the errors in AB magnitude in twenty filters from FUV to FIR. We only compare the results for $M_{\text{true}}^{AB} < -10$. The errors in FUV to NIR regime are dominated by the dust attenuation model, which are at $\sim 0.01$ mag. An exception is the steep increase of the uncertainties in the $u$-, $g$-, $r$-bands. The reason could be that the adopted wavelength grid is too coarse to capture the line emission features in these bands. This problem may be seen from the third rows of Figure 6. In terms of the dust emission model, the predicted magnitudes in the MIR to FIR bands deviate from the fiducial values by $\sim 0.1$ mag to 0.2 mag.

2 In fact, this step can be done by pre-selecting the input parameters before running the simulations.
Table 3. Best-fitting parameters of IC4225 and NGC5166.

| Symbol | Description                           | Unit       | IC4225 (fixed $r_{\text{bulge}}$) | IC4225 (fixed $r_{\text{bulge}}$) | IC4225 (fixed $r_{\text{bulge}}$) | IC4225 (fixed $r_{\text{bulge}}$) |
|--------|---------------------------------------|------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| $\theta$ | Inclination angle                    | -          | 88.9                              | 86.4                              | 86.1                              | 86.1                              |
| $\Sigma_{\text{dust}}$ | Dust column density                | $M_{\odot}/\text{kpc}^2$ | $9.97 \times 10^4$                  | $6.95 \times 10^4$                  | $7.13 \times 10^4$                  | $8.07 \times 10^4$                  |
| $r_{\text{dust}}/r_{\text{disk}}$ | Dust disk radius scaled by stellar disk radius | -          | 0.83                              | 4.24                              | 4.24                              | 5.01                              |
| $r_{\text{disk}}$ | Stellar disk radius                | kpc        | 7.81                              | 2.05                              | 1.94                              | 3.87                              |
| $r_{\text{bulge}}$ | Stellar bulge radius              | kpc        | 0.32                              | 0.64                              | 1.25                              | 0.32                              |
| $M_{\text{dust}}$ | Dust mass                          | $M_{\odot}$ | $2.64 \times 10^7$                  | $3.30 \times 10^7$                  | $3.02 \times 10^7$                  | $2.10 \times 10^7$                  |
| $m_{\text{disk}}$ | Disk Mass                          | $M_{\odot}$ | $8.68 \times 10^6$                  | $6.78 \times 10^{10}$                 | $6.75 \times 10^{10}$                 | $-$                              |
| $m_{\text{bulge}}$ | Bulge Mass                         | $M_{\odot}$ | $6.72 \times 10^{10}$                 | $3.39 \times 10^8$                  | $8.79 \times 10^6$                  | $-$                              |
| SFR$_{\text{disk}}$ | Disk star formation rate$^b$       | $M_{\odot}/\text{yr}$ | 0.07                              | 0.00                              | 0.00                              | $-$                              |
| SFR$_{\text{bulge}}$ | Bulge star formation rate$^b$      | $M_{\odot}/\text{yr}$ | 0.00                              | 0.47                              | 0.07                              | $-$                              |
| $c$ | Bulge metallicity                   | $Z_{\odot}$ | 0.42                              | 0.002                             | 0.01                              | 0.010                             |

Table 4. Summary of some relevant galaxy properties derived from the best-fitting parameters of IC4225 and NGC5166.

| Symbol | Description                           | Unit       | NGC5166 (fixed $r_{\text{bulge}}$) | NGC5166 (fixed $r_{\text{bulge}}$) | NGC5166 (fixed $r_{\text{bulge}}$) | NGC5166 (fixed $r_{\text{bulge}}$) |
|--------|---------------------------------------|------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| $\theta$ | Inclination angle                    | -          | 86.1                              | 85.5                              | 85.8                              | $-$                              |
| $r_{\text{disk}}$ | Stellar disk radius                | kpc        | 3.87                              | 5.99                              | 3.98                              | 3.92                              |
| $r_{\text{bulge}}$ | Stellar bulge radius              | kpc        | 0.32                              | 0.64                              | 3.17                              | 3.17                              |
| $r_{\text{dust}}$ | dust disk radius                   | kpc        | 19.40                             | 13.55                             | 11.95                             | 5.81                              |
| $m_{\text{dust}}$ | Dust mass                          | $M_{\odot}$ | $1.91 \times 10^8$                  | $1.54 \times 10^8$                  | $1.21 \times 10^8$                  | $4.80 \times 10^7$                  |
| $m_{\text{disk}}$ | Disk Mass                          | $M_{\odot}$ | $1.13 \times 10^{12}$                 | $1.10 \times 10^{11}$                 | $1.06 \times 10^{11}$                 | $-$                              |
| $m_{\text{bulge}}$ | Bulge Mass                         | $M_{\odot}$ | $5.95 \times 10^7$                  | $5.28 \times 10^7$                  | $9.10 \times 10^9$                  | $-$                              |
| SFR$_{\text{disk}}$ | Disk star formation rate$^b$       | $M_{\odot}/\text{yr}$ | 0.00                              | 0.00                              | 0.75                              | $-$                              |
| SFR$_{\text{bulge}}$ | Bulge star formation rate$^b$      | $M_{\odot}/\text{yr}$ | 0.51                              | 0.45                              | 0.00                              | $-$                              |
| $\xi_{\text{abs}}$ | Dust absorption fraction           | -          | 0.21                              | 0.23                              | 0.20                              | $-$                              |

$^a$ Disk mass and bulge mass are obtained by integrating the star formation history without taking into account the mass loss due to the supernova.

$^b$ The timescale of star formation rate is $\sim 100$ Myr.
Figure 6. Example SEDs from the validation data described in Section 2.4. Black and red lines show the SEDs obtained by \textsc{skirt} and our deep learning model respectively.

Figure 7. Uncertainties of our deep learning model in 20 filters. The model combines the dust attenuation and emission networks. The errors are given by the difference in AB magnitude and estimated using the validation data described in Section 2.4.
3. APPLICATIONS

As an application, we fit our deep learning model to the observed SED of IC4225 and NGC5166. We use the data summarised by De Geyter et al. (2015) (hereafter DG15), which are obtained from the GALEX (Martin et al. 2005), the SDSS (Abazajian et al. 2009), the 2MASS (Skrutskie et al. 2006), the WISE (Wright et al. 2010), the IRAS (Moshir et al. 1992), and the Herschel-ATLAS (Eales et al. 2010). The used data are plotted in Figure 8.

For the present work, we do not adopt the uncertainties from the observations. The observed errors are assumed to be Gaussian in flux. However, as discussed in Section 2.4, our deep learning model also introduces some uncertainties, which are Gaussian-like in magnitude. Some of these errors may be even larger than the observations. Hence, for the fittings of IC4225 and NGC5166, we employ a loss function that assumes Gaussian errors in magnitude

\[ \Psi = \sum_{\text{band}} \left( \frac{m_{\text{AB}}^{\text{obs}} - m_{\text{AB}}^{\text{true}}}{\sigma} \right)^2. \]  

(24)

Based on the results illustrated in Figure 7, we assume \( \sigma = 0.1 \) for the SDSS-u SDSS-g, SDSS-r, WISE-12, WISE-22, and IRAS-100 bands, \( \sigma = 0.2 \) for all Herschel bands, and \( \sigma = 0.05 \) for other bands. Since we ignore the observational errors, to avoid double weighting, we do not use the fluxes in IRAS-25, IRAS-100 and Planck bands for NGC5166, which overlap the existing data points. In addition, we assume the distances of IC4225 and NGC5166 are 73.9 Mpc and 62.6 Mpc respectively according to DG15.

The free parameters of the fitting model are the input parameters of our deep learning model. As listed in Table 1, we have six parameters describing the geometry and two parameters specifying the intrinsic bolometric luminosity of the stellar disk and bulge. As introduced in Section 2.3.1, we can have thirty-six parameters for the star formation history. However, such degrees of freedom may be unnecessary for the SED fitting. Instead, we use three parameters to describe the fractional luminosity of three stellar age bins and one parameter to specify the overall metallicity. The reduced stellar age bins are obtained by merging every two bins. The corresponding bin edges can be found in Table 3. The metallicity is applied using linear interpolation. This is a typical discrete star formation history model used by several recent SED-fitting studies (Leja et al. 2017; Boquien et al. 2019; Robotham et al. 2020). We adopt this model for both the stellar disk and bulge. We note that the fractional luminosity parameters \( c_k \) should satisfy

\[ \sum_k c_k = 1 \]  

(25)

To tackle this constrain, we assume that each input parameter satisfies \( 0 < a_k < 1 \), and apply the following transformation

\[ c_k = \frac{\log(1 - a_k)}{\sum_k \log(1 - a_k)} \]  

(26)

If each \( a_k \) is uniformly distributed, the transformed parameters will follow a flat Dirichlet distribution. Overall, our SED-fitting model contains 15 free parameters.

We utilise the particle swarm optimisation (PSO) (Shi & Eberhart 1998) to minimise the loss function defined by Equation (24). We run the optimiser with 32 particles for 2000 iterations. We find that the parameter space is complex and contains multiple local minima. To ensure the results are stable, we perform the optimisation for 10 times for each galaxy, and choose the best-fit from these runs.

The best-fitting SEDs are demonstrated in Figure 8, and the best-fitting parameters are given in Table 3. We find that our SED model can fit the observations well. To test our deep learning model, we run the SKIRT simulation using the best-fitting parameters, and plot the results in Figure 8 as dashed lines. The SEDs predicted by our deep learning model are well consistent with the simulation results.

Table 4 compares some relevant quantities with DG15, who carried out an image fitting analysis on these galaxies using FITSKIRT (De Geyter et al. 2013). Our model also predicts that both IC4225 and NGC5166 are edge-on galaxies, consistent with the galaxy images. However, our predicted geometry parameters are different from DG15. An additional issue is that for NGC5166, our model cannot fit the fluxes in the WISE-22.2 and IRAS-70 bands. This behaviour was also found by DG15. Their results show even poorer consistency with the observations at MIR to FIR wavelengths. The issue is known as the dust energy balance problem, which was reported by several image studies (e.g. De Looze et al. 2012; Mosenkov et al. 2016, 2018).

We find that the bulge radius is related to the discrepancies found in this work. Our predicted bulge radius for IC4225 and NGC5166 is much smaller than DG15, and reaches the minimum value in our model. To examine the dependency of the SED on bulge radius, we perform two additional fittings for each galaxy, doubling the bulge radius and fixing the bulge radius
Figure 8. Best-fitting SEDs of IC4225 and NGC5166. Coloured dots are observational data described in De Geyter et al. (2015). Solid lines with empty circles are the best-fitting results of our deep learning SED model. For the black lines, no parameter is fixed, while for the blue and magenta lines, bulge radius is fixed. In particular, when fitting the magenta lines, we fix the bulge radius at the same value with De Geyter et al. (2015). The dashed lines are obtained by running the SKIRT simulation using the corresponding best-fitting parameters. The model free parameters and the fitting method are described in Section 3. As discussed in Section 3, observational uncertainties are not considered in the fittings, which may be smaller than the intrinsic errors of our deep learning model.
at the same value with DG15\textsuperscript{3}. The additional results can also be found in Figure 8, Tables 3 and 4. Figure 8 shows that increasing the bulge radius lowers the fluxes at 20 \(\mu\)m – 80 \(\mu\)m. This trend is very obvious for NGC5166, in which case the predicted SED has a constraint at \(\sim 70 \mu\)m. In Tables 3 and 4, no apparent trend can be found in the changes of the best-fitting parameters. In particular, the star formation history of the stellar disk and bulge is varied significantly. These findings imply that the parameter space is highly degenerated. In other words, our results find it difficult to constrain galaxy geometry parameters from the spatially integrated SED alone.

4. SUMMARY

In this work, we present a deep learning model which predicts the FUV to FIR SED by emulating the \textsc{skirt} radiative transfer simulation. The SED model is comprised of three specifically designed neural networks. Each neural network can be trained using \(\sim 4000 – 5000\) samples. We estimate that the uncertainties of the deep learning model are \(\sim 0.01 – 0.1\) for the dust attenuation neural network and \(\sim 0.2\) for the dust emission neural network.

As an application, we use our deep learning SED model to fit the observed SEDs of IC4225 and NGC5166, and compare the results with the image-fitting analysis by De Geyter et al. (2015). Our findings can be summarised as follows:

1. Our model can reproduce the observed SEDs, and successfully predict that both IC4225 and NGC5166 are edge-on galaxies.

2. The parameter space of our SED model, particularly the star formation history of the stellar disk and bulge, is highly degenerated. Consequently, our predicted geometry parameters are inconsistent with the image-fitting analysis.

3. The predicted fluxes at 20 \(\mu\)m – 80 \(\mu\)m by our SED model are correlated with bulge radius.

This work demonstrates the feasibility of using the supervised deep learning technique to accelerate the speed of radiative transfer simulations. We expect that our model can be widely used in SED-fitting and SED-modelling of galaxies from semi-analytic models. We have presented an application of our model on SED-fitting in this work, and will apply the model to a semi-analytic model in a future work. Generating theoretical galaxy SEDs is a very complex task, and there are still many aspects that are not considered in this work. We suggest that our SED model can have the following extensions:

1. It is instructive to have the SED model support multiple stellar evolutionary models. While our dust attenuation model is independent of simple stellar population (SSP) libraries, our dust emission model is restricted to the Flexible Stellar Population Synthesis library with the given star formation history bins, initial mass function and nebular emission model. If we want to change the settings, we need to run additional radiative simulations and retrain the model.

2. The wavelength resolution of our SED model can be improved. This is critical for spectroscopic and emission line studies. When the emission lines are included, the attenuation from the birth cloud of stars is significant, which should be taken into account explicitly. Increasing the wavelength resolution could also reduce the model errors in the u-, g-, r-bands.

3. The uncertainty of our dust emission model are not negligible for applications of SED-fitting. This issue might be resolved by improving the design of the neural work. Alternatively, one can also build a model to describe the errors.

This work is partly supported by the NSFC (No.11825303, 11861131006), the science research grants from the China Manned Space project with NO.CMS-CSST-2021-A03, CMS-CSST-2021-B01 and the cosmology simulation database (CSD) in the National Basic Science Data Center (NBSDC-DB-10).

Software: \textsc{astropy}\textsuperscript{4} (Astropy Collaboration et al. 2013, 2018), \textsc{fsps} (Conroy et al. 2009; Conroy & Gunn 2010), \textsc{jupyter-notebook}\textsuperscript{5}, \textsc{matplotlib} (Hunter 2007), \textsc{numpy} (Harris et al. 2020), \textsc{pytorch} \textsuperscript{6} (Paszke et al. 2019), \textsc{python-fsps}\textsuperscript{7}, \textsc{scipy} (Jones et al. 2001), \textsc{sedpy} \textsuperscript{8}, \textsc{skirt} (Camps & Baes 2020).

\textsuperscript{3} We note that the our geometry model of the stellar bulge is slightly different from DG15, whose model includes two additional parameters: the Sérsic index and the intrinsic flattening.

\textsuperscript{4} https://www.astropy.org/

\textsuperscript{5} https://github.com/executablebooks/jupyter-book

\textsuperscript{6} https://pytorch.org/

\textsuperscript{7} https://dfm.io/python-fsps/current/

\textsuperscript{8} https://github.com/bd-j/sedpy
REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543, doi: 10.1088/0067-0049/182/2/543
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabe4f
Baes, M., Verstappen, J., De Looze, I., et al. 2011, ApJS, 196, 22, doi: 10.1088/0067-0049/196/2/22
Bellstedt, S., Robotham, A. S. G., Driver, S. P., et al. 2020, MNRAS, 498, 5581, doi: 10.1093/mnras/staa2620
Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103, doi: 10.1051/0004-6361/201834156
Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44, doi: 10.3847/1538-4357/aa6e66
Camps, P., & Baes, M. 2015, Astronomy and Computing, 9, 20, doi: 10.1016/j.ascom.2014.10.004
—. 2020, Astronomy and Computing, 31, 100381, doi: 10.1016/j.ascom.2020.100381
Chabrier, G. 2003, PASP, 115, 763, doi: 10.1086/376392
Compagnie, M., Verstraete, L., Jones, A., et al. 2011, A&A, 525, A103, doi: 10.1051/0004-6361/201015292
Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833, doi: 10.1088/0004-637X/712/2/833
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486, doi: 10.1088/0004-637X/699/1/486
Cullen, F., McLure, R. J., Khochar, S., Dunlop, J. S., & Dalla Vecchia, C. 2017, MNRAS, 470, 3006, doi: 10.1093/mnras/stx1451
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595, doi: 10.1111/j.1365-2966.2008.13535.x
De Geyter, G., Baes, M., Camps, P., et al. 2014, MNRAS, 441, 869, doi: 10.1093/mnras/stu612
De Geyter, G., Baes, M., Fritz, J., & Camps, P. 2013, A&A, 550, A74, doi: 10.1051/0004-6361/201220126
De Geyter, G., Baes, M., De Looze, I., et al. 2015, MNRAS, 451, 1728, doi: 10.1093/mnras/stv1104
De Looze, I., Baes, M., Bendo, G. J., et al. 2012, MNRAS, 427, 2797, doi: 10.1111/j.1365-2966.2012.22045.x
De Looze, I., Fritz, J., Baes, M., et al. 2014, A&A, 571, A69, doi: 10.1051/0004-6361/201424747
Dobbs, W., Baes, M., Viaene, S., et al. 2020, A&A, 634, A57, doi: 10.1051/0004-6361/201936695
Draine, B. T., & Li, A. 2007, ApJ, 657, 810, doi: 10.1086/511055
Eales, S., Dunne, L., Clements, D., et al. 2010, PASP, 122, 499, doi: 10.1086/653086
Fioc, M., & Rocca-Volmerange, B. 2019, A&A, 623, A143, doi: 10.1051/0004-6361/201833556
Gilda, S., Lower, S., & Narayanan, D. 2021, ApJ, 916, 43, doi: 10.3847/1538-4357/ac0058
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (MIT Press)
Hahn, C., Starkenburg, T. K., Anglés-Alcázar, D., et al. 2021, arXiv e-prints, arXiv:2106.09741.
https://arxiv.org/abs/2106.09741
Han, Y., & Han, Z. 2014, ApJS, 215, 2, doi: 10.1088/0067-0049/215/1/2
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2
Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90, doi: 10.1109/MCSE.2007.55
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python.
http://www.scipy.org/
Lagos, C. d. P., da Cunha, E., Robotham, A. S. G., et al. 2020, MNRAS, 499, 1948, doi: 10.1093/mnras/staa2861
Lagos, C. d. P., Robotham, A. S. G., Trayford, J. W., et al. 2019, MNRAS, 489, 4196, doi: 10.1093/mnras/stz2427
Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, ApJ, 837, 170, doi: 10.3847/1538-4357/aa5fe
Lovell, C. C., Acquaviva, V., Thomas, P. A., et al. 2019, MNRAS, 490, 5503, doi: 10.1093/mnras/stz2851
Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJL, 619, L1, doi: 10.1086/426387
Mosenkov, A. V., Allaert, F., Baes, M., et al. 2016, A&A, 592, A71, doi: 10.1051/0004-6361/201628676
—. 2018, A&A, 616, A120, doi: 10.1051/0004-6361/201832999
Mosher, M., Kopman, G., & Conrow, T. A. O. 1992, IRAS Faint Source Survey, Explanatory supplement version 2 (Pasadena: Infrared Processing and Analysis Center, California Institute of Technology)
Narayanan, D., Turk, M. J., Robitaille, T., et al. 2021, ApJS, 252, 12, doi: 10.3847/1538-4365/abc487
Paszke, A., Gross, S., Massa, F., et al. 2019, arXiv e-prints, arXiv:1912.01703. https://arxiv.org/abs/1912.01703
Qiu, Y., Mutch, S. J., da Cunha, E., Poole, G. B., & Wyithe, J. S. B. 2019, MNRAS, 489, 1357, doi: 10.1093/mnras/stz2233
Robotham, A. S. G., Bellstedt, S., Lagos, C. d. P., et al. 2020, MNRAS, 495, 905, doi: 10.1093/mnras/staa1116
Shi, Y., & Eberhart, R. 1998, in 1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360), IEEE, 69–73
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163, doi: 10.1086/498708
Smith, L. N., & Topin, N. 2019, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11006, Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, 1100612, doi: 10.1117/12.2520589

Steinacker, J., Baes, M., & Gordon, K. D. 2013, ARA&A, 51, 63, doi: 10.1146/annurev-astro-082812-141042
Viaene, S., Baes, M., Tamm, A., et al. 2017, A&A, 599, A64, doi: 10.1051/0004-6361/201629251
Vogelsberger, M., Nelson, D., Pillepich, A., et al. 2020, MNRAS, 492, 5167, doi: 10.1093/mnras/staa137
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868, doi: 10.1088/0004-6256/140/6/1868
Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Popping, G., & Davé, R. 2019, MNRAS, 483, 2983, doi: 10.1093/mnras/sty3241