Envisioning Distant Worlds: Training a Latent Diffusion Model with NASA's Exoplanet Data

Marissa A. Beaty
Creative Computing Institute
University of Arts London
London, UK
marissabeaty2@gmail.com

Terence Broad
Creative Computing Institute
University of Arts London
London, UK
t.broad@arts.ac.uk

Abstract

There are some 5,500 confirmed Exoplanets beyond our solar system. Though we know these planets exist, most of them are too far away for us to know what they look like. In this paper, we develop an algorithm and a model to translate any given exoplanet's numeric data into a text prompt that can be input into a trained latent diffusion model to generate a predictive visualization of that exoplanet. This paper describes a novel approach of translating numeric data to textual descriptors formulated from prior accepted astrophysical research. These textual descriptions are paired with photographs and artistic visualisations from NASA's public archives to build a training set for a latent diffusion model, which can produce new visualizations of unseen distant worlds.

1 Introduction

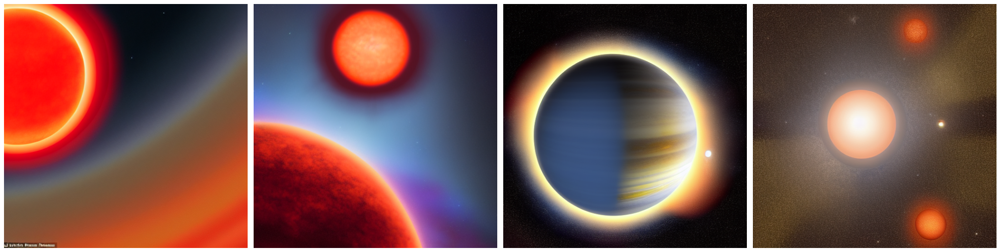
Latent diffusion models offer a powerful and easily controllable form of image generation [12]. When conditioned with text and guided with a contrastive learning algorithm like CLIP [10], models can be trained to learn powerful associations between text and images. Using the numerical data provided in NASA's exoplanet dataset [1], we develop an algorithm that creates descriptive text prompts for each planet, based on established principals in astrophysics. We pair these generated prompts with available photographs and artistic renderings of near and distant planets from NASA's image library [3], to build a dataset of text image prompts. A text-conditioned, latent diffusion model is trained on this dataset, which is capable of generating visualizations of previously unseen exoplanets.

2 Describing Exoplanets with Natural Language

The NASA exoplanet dataset contains 35,086 records of more than 5,500 confirmed exoplanets [1]. A multitude of data points for these planets is recorded, such as the name of the exoplanet and star system, their respective masses, the distance from the planet to the star and the orbital period of the planet (see Table 2 for a full breakdown of variables used in this work). This data is collected, reviewed, and confirmed by orbiting satellites through the radial velocity or the transit method [15]. Using this available data, we create an algorithm to provide textual descriptions that define the stellar size, stellar color, planet category, planet size, planet color, planet spin (which determines weather patterns), and determine whether a planet is tidal-locked.

The stellar size is determined by stellar mass as it is listed in the dataset, as a ratio between stellar mass and planet mass, or as a textual descriptor ranging from tiny to massive (Table 14). The stellar color is defined using one of three methods. If the stellar category is given using the Harvard spectral classification system [6] (the foundation for the Morgan-Keenan classification system [17]) then that

37th Conference on Neural Information Processing Systems (NeurIPS 2023).



(a) (b) (c) (d)

Figure 1: Exoplanet visualisation generated with our latent diffusion model, exoplanets visualised are as follows: (a) 24 Boo b (b) 2MASS J01225093-2439505 b (c) HD 106315 b (d) HD 15906 c.

is used to determine the stellar color (Table 11). If that is not available, then the stellar temperature (Table 12) or the stellar mass is used (Table 13).

We categorized each planet into one of four categories: terrestrial, super-earth, neptune-like, and gas-giant [2] (Table 4). These categories allow us to make assumptions about a planet, such as it's size and temperature. A planet's temperature can define planet color by researching the states of matter and spectral emission and absorption for common chemicals in our Universe. As these planets are too far away to conduct spectroscopy, the relationship between a planets equilibrium temperature and chemical behaviors is a reasonable alternative [9]. For each of the four planet categories, we provide text descriptions of their color based on their predicted chemical composition (Tables 5, 6 & 7).

We define the planet spin to predict potential weather patterns based on the planets orbital period and the planet category [5]. Planet's that spin quickly are more likely to have turbulent weather, and thus, more visible cloud patterns [5]. The range of cloud formations defined based on this data is available in Tables 8, 9, & 10. Finally, we determined whether the exoplanet is tidal-locked using the Roche limit (Figure 3). Tidal locking is when a planet's axial spin is the same as the planet's orbital period causing the side of the planet facing the sun to be extremely hot while the other side is extremely cold [4]. If the calculated Roche limit is less than the planets impact parameter then the exoplanet is likely tidal-locked.

3 Model Training

The images acquired for training the latent diffusion model were all sourced from NASA's image library [3]. The image dataset consists of 127 images split approximately 70% artistic renderings to 30% photographs. As the images are categorised by planet name, the corresponding text descriptions of both the near planets (in our solar system) and exoplanets can be generated using the method described above. A latent diffusion model is then trained on this dataset using the publicly available Stable Diffusion 1.5 [11] as the starting model for training Low Rank-Adaptation matrices (LoRA) [7] to adapt the model to the new data domain. Results from the trained model can be seen in Figure 1.

4 Discussion

Our aim was to translate numeric data within NASA's exoplanet data into predictive images. Table 1 shows the end to end process of our system. By developing a bespoke algorithm for describing exoplanets with natural language, we are able to create a meaningful mapping from numerical exoplanet data to visual imagery. Given the available data, the chemical composition and the visual characteristics cannot be known exactly, the descriptions are best guesses based on existing literature. Since completing the work, predictions from our approach for the exoplanet K2-18 b have already been validated by real-world data released by the James Webb Space Telescope [14]. A comparison of the rendering from our approach and NASA artistic rendering can be seen in Figure 2. While obtaining ground truth images for most of these exoplanets remains unfeasible with current technology, our approach serves as a valuable tool for conceptualizing and comprehending these remote celestial entities.

5 Ethical Concerns

NASA's Exoplanet dataset and Image and Video Library used for this project are in the public domain and do not contain any personal, private, or secure information that could be potentially unethical. The models and training of these models involve ethical concerns that require acknowledgment and consideration for future use in this work. In April 2023, a lawsuit was filed against Stable Diffusion for violating copyright protections of visual artists [16]. A resolution has not been reached, but should Stable Diffusion be found guilty, further use of this research could result in the endorsement of unethical and illegal business practices. This may not be a current concern for those interested in this research but should be taken seriously for future iterations and applications. Furthermore, running these GPUs consumes a significant amount of energy. It is estimated that training an AI model, such as Stable Diffusion, could produce 626,000 pounds of CO2 [13]. Though fine-tuning a model produces less CO2 than training one, it is necessary to consider the environmental effects of reproducing and building on this type of research, particularly should it be expanded upon with larger datasets or trained for a longer duration.

References

[1] National Aeronautics and Space Administration. Planetary systems table. `https://exoplanetarchive.ipac.caltech.edu/`, 2021. DOI:10.26133/NEA12.

[2] National Aeronautics and Space Administration. Exoplanet exploration: Planets beyond our solar system. `https://exoplanets.nasa.gov/what-is-an-exoplanet/planet-types/overview/`, 2022.

[3] National Aeronautics and Space Administration. NASA image and video library. `https://images.nasa.gov/`, no date.

[4] Fani Dosopoulou, Smadar Naoz, and Vassiliki Kalogera. Roche-lobe overflow in eccentric planet–star systems. *The Astrophysical Journal*, 844(1):12, 2017.

[5] Scott D Guzewich, Jacob Lustig-Yaeger, Christopher Evan Davis, Ravi Kumar Kopparapu, Michael J Way, and Victoria S Meadows. The impact of planetary rotation rate on the reflectance and thermal emission spectrum of terrestrial exoplanets around sunlike stars. *The Astrophysical Journal*, 893(2):140, 2020.

[6] Harvard. Spectral classification. `https://lweb.cfa.harvard.edu/~pberlind/atlas/htmls/note.html`, no date.

[7] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[8] ESA J. Olmsted (STScI) Science: N. Madhusudhan (Cambridge University) NASA, CSA. Artist illustration: K2-18b.

[9] National Instititute of Standards and Technology. Chemical name search. `https://webbook.nist.gov/chemistry/name-ser/`, 2023.

[10] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[11] Robin Rombach and Patrick Esser. Stable diffusion 1.5 model card. `https://huggingface.co/runwayml/stable-diffusion-v1-5`, 2022.

[12] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. pages 10684–10695, 2022.

[13] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, 2019.

[14] Webb Team. Webb discovers methane, carbon dioxide in atmosphere of k2-18 b. 2023.

[15] T. Trifonov and J. A. Caballero et. al. A nearby transiting rocky exoplanet that is suitable for atmospheric investigation. *Science*, 371(6533):1038–1041, 2021. doi: 10.1126/science.abd7645. URL `https://www.science.org/doi/abs/10.1126/science.abd7645`.

[16] James Vincent. Getty images sues ai art generator stable diffusion in the us for copyright infringement. 2023.

[17] Morgan William Wilson, Keenan Philip Childs, and Kellman Edith. An atlas of stellar spectra, with an outline of spectral classification. 1943.

Appendix

(a) (b)

Figure 2: Depictions of K2-18 b: (a) Visualization Published by the NASA, CSA & ESA James Webb Telescope, reproduced under creative commons licence (CC BY 4.0 DEED) [8] (b) Visualization by our trained model.

$$\text{Roche Limit} = 2.44 \cdot R \left(\frac{D}{d}\right)^{\frac{1}{3}} \tag{1}$$

Where:

R is the radius of the larger body.
D is the density of the larger body.
d is the density of the smaller body.

Figure 3: Roche Limit formula with explanations.

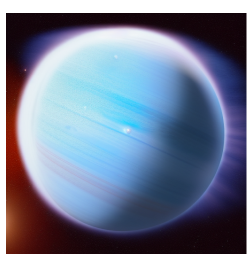
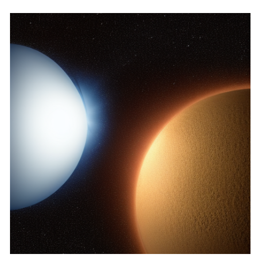
Exoplanet name	HD 56957 b	HIP 41378 c
Exoplanet data	• sy_snum: 1 • sy_pnum: 1 • sy_mnum: 1 • pl_orbper: 29.94992 • pl_rade: 3.71 • pl_bmassse: 13.3 • pl_dens: 1.43 • pl_eqt: 1133 • pl_imppar: 0.34 • pl_orbsmax: 0.229 • st_spectype: N/A • st_teff: 8500 • st_rad: 1.75 • st_mass: 1.89	• sy_snum: 1 • sy_pnum: 5 • sy_mnum: 0 • pl_orbper: 15.572098 • pl_rade: 2.507 • pl_bmassse: 6.83 • pl_dens: 2.38 • pl_eqt: 697 • pl_imppar: 0.53 • pl_orbsmax: N/A • st_spectype: F6 • st_teff: 6226 • st_rad: 1.34 • st_mass: 1.17
Generated prompt	"A white, large star with a Neptune-like planet. The planet is mostly blue mixing with brown and red colors, has clearly defined striped light and dark clouds, and only has one side of the planet facing the sun. The side facing the sun is extremely hot and the side that faces away from the sun is dark and cold."	"A yellow white, medium star with a super-earth planet. The planet is white and pale yellow in color, is hot and rotating quickly with little to no clouds, and only has one side of the planet facing the sun. The side facing the sun is extremely hot and the side that faces away from the sun is dark and cold."
Generated image		

Table 1: Examples of two exoplanets, with the original NASA data, the respective text description and the resulting images generated from our latent diffusion model.

Variable	Description	Unit of Measurement
sy_snum	Number of stars in the system	Star(s)
sy_pnum	Number of planets in the system	Planet(s)
sy_mnum	Number of moons in the system	Moon(s)
pl_orbper	Orbital period	Earth days
pl_rade	Planet radius	Earth radius
pl_bmasse	Planet mass	Earth mass
pl_dens	Planet density	g/cm^3
pl_eqt	Planet equilibrium temperature	Kelvin (K)
pl_imppar	Impact parameter	Distance between stellar and planet disc / stellar radius
pl_orbsmax	Radius of longest elliptic orbit	Astronomical units (AU)
st_spectype	Spectral type	Morgan-Keenan classification
st_teff	Star effective temperature	Kelvin (K)
st_rad	Stellar radius	Radius of the Sun
st_mass	Stellar mass	Mass of the Sun

Table 2: Data variables available from NASAs exoplanet database used for visualisation.

Planet mass	Planet size
$0 < x \leq 0.0553$	Tiny
$0.0553 < x \leq 0.107$	Very Small
$0.107 < x \leq 0.815$	Small
$0.815 < x \leq 1.0$	Medium Small
$1.0 < x \leq 14.5$	Medium
$14.5 < x \leq 17.1$	Large
$17.1 < x \leq 95.2$	Giant
$95.2 < x$	Massive

Table 3: Planet mass descriptions where x is planet mass (given in units of masses of the Earth)

Planet mass	Planet category
$0 < x \leq 2.0$	Terrestrial
$2.0 < x \leq 10.0$	Super Earth
$10 < x \leq 17$	Neptune-Like
$17 < x$	Gas Giant

Table 4: Planet category descriptions where x is planet mass (given in units of masses of the Earth).

Planet temperature	Description
$0 < x \leq 20.0$	Hydrogen and Helium producing a distinct white color
$20.0 < x \leq 200.0$	High quantities of methane known for its rich blue color
$200.0 < x \leq 400.0$	Blue methane and yellow ammonia, dominant color from blue liquid water
$400.0 < x \leq 600.0$	Water vapor producing a true blue color mixing with methane
$600.0 < x \leq 800.0$	Carbon dioxide and hydrocarbons, varying shades of blue and white
$800.0 < x \leq 1200.0$	White carbon dioxide and pale yellow sulfur compounds
$1200.0 < x \leq 1700.0$	Pale yellow sulfur compounds and blue and white water vapor
$1700.0 < x$	Covered in lava due to high temperatures

Table 5: Terrestrial and Super-Earth planet color descriptions where x is equilibrium temperature (given in Kelvin).

Planet temperature	Description
$0 < x \leq 90.0$	Mostly helium and hydrogen, white with light blue from frozen methane
$90.0 < x \leq 110.0$	Azure blue with methane as a liquid
$110.0 < x \leq 275.0$	Deep blue with methane as a gas
$275.0 < x \leq 375.0$	Dark blue methane, lighter blue water vapor, traces of ammonia
$375.0 < x \leq 500.0$	Methane breaking down, pale yellow sulfur
$500.0 < x \leq 800.0$	Methane breaking down, varying shades of blue from hydrocarbons
$800.0 < x \leq 900.0$	Deep blue methane breaking down, silvery white from alkali metals
$900.0 < x \leq 1400.0$	Deep blue methane breaking down, neutral or red from aerosols and thermal emissions
$1400.0 < x$	Purple to red from aerosols, thermal emissions, and high-temperature gases

Table 6: Neptune-like planet color descriptions where x is equilibrium temperature (given in Kelvin).

Planet temperature	Description
$0 < x \leq 70.0$	Dull yellow from frozen ammonia, mixed with blue methane
$70.0 < x \leq 150.0$	Dominant yellow from ammonia clouds
$150.0 < x \leq 250.0$	Darker yellow, ammonia as a liquid
$250.0 < x \leq 350.0$	Mostly white with slight blue tint from water vapor
$350.0 < x \leq 800.0$	Uniform blue
$800.0 < x \leq 900.0$	Transitioning from blue to silvery white from carbon monoxide and alkali metals
$900.0 < x \leq 1400.0$	Silvery white from carbon monoxide and alkali metals
$1400.0 < x$	Red from silicate and iron clouds

Table 7: Gas-giant planet color descriptions where x is equilibrium temperature (given in Kelvin).

Planet temperature	Description
$0 < x \leq 88.0$	Hot and rotating quickly with little to no clouds
$88.0 < x \leq 224.0$	Hot and rotating quickly with swirling clouds in light and dark shades
$224.0 < x \leq 687.0$	Thick clouds of various sizes
$687 < x$	Thin clouds of various sizes

Table 8: Terrestrial and Super-Earth planet spin descriptions where x is orbital period (Earth days).

Planet temperature	Description
$x \leq 30589.0$	Clearly defined stripes of light and dark clouds
$30589.0 < x \leq 59800.0$	Softly defined stripes of light and dark clouds
$59800.0 < x$	Clouds of various colors blending together

Table 9: Neptune-like planet spin descriptions where x is orbital period (Earth days).

Planet temperature	Description
$x \leq 30589.0$	Clear, sharp-edge stripes of thick clouds
$30589.0 < x \leq 59800.0$	Softly defined stripes of light and dark clouds
$59800.0 < x$	Clouds of various colors blending together

Table 10: Gas-Giant planet spin descriptions where x is orbital period (Earth days).

Spectral type	Stellar color
M	orange red
K	light orange
G	yellow
F	yellow white
A	white
B	blue white
O	blue
T	violet
L	magenta
WD or D	white

Table 11: Stellar color coding based on the Harvard spectral classification [6].

Stellar temperature	Stellar color
$x \leq 3500.0$	orange red
$3500.0 < x \leq 5000.0$	light orange
$5000.0 < x \leq 6000.0$	yellow
$6000.0 < x \leq 7500.0$	yellow white
$7500.0 < x \leq 11000.0$	white
$11000.0 < x \leq 25000.0$	blue white
$25000.0 < x \leq 100000.0$	blue
$100000.0 < x$	white

Table 12: Stellar color coding where x is the effective temperature (given in Kelvin).

Stellar mass	Stellar color
≤ 0.45	orange red
$0.45 < x \leq 0.8$	light orange
$0.8 < x \leq 1.04$	yellow
$1.04 < x \leq 1.4$	yellow white
$1.4 < x \leq 2.1$	white
$2.1 < x \leq 16$	blue white
$16 < x$	blue

Table 13: Stellar color coding where x is the stellar mass (given in units of masses of the Sun).

Stellar mass	Stellar size
$0 < x \leq 0.08$	Tiny
$0.08 < x \leq 0.45$	Very Small
$0.45 < x \leq 0.8$	Small
$0.8 < x \leq 1.04$	Medium Small
$1.04 < x \leq 1.4$	Medium
$1.4 < x \leq 2.1$	Large
$2.1 < x \leq 16$	Giant
$16 < x$	Massive

Table 14: Stellar mass descriptions where x is stellar mass (given in units of masses of the Sun)