

Capstone Proposal

No Man's Sky Explored Space Map

CSCI 490
Computer Science
&
Data Science
Capstone

Derek Borders

Last compiled on
September 27, 2022

No Man's Sky Interactive 3d Galaxy Map

No Man's Sky is a space exploration game that uses extensive procedural generation to create a mindbogglingly massive open-world play space. The game contains 255 galaxies, each of which contains 4.2 billion regions. Each region is a roughly 400 light-year cube of space and home to around 100 to 600 star systems.¹ Finally, each system has 2-6 planets (counting moons).

The number of planets is so vast that the vast majority of them will never be visited by anybody. Even with all users sharing a starting galaxy (Euclid), in normal exploration, an individual player is often the only person who has ever seen a given planet they land on.

Each world in No Man's Sky are addressed and reachable via an in-game portal network using its unique portal address. This address uses a unique 16 glyph system that translates nicely to hexadecimal. In each address, the final 8 characters translate to hexadecimal coordinates in 3d space. Two for the y axis, and three each for the X and Z axis. These coordinates identify the galactic region containing the system.

No Man's Sky includes a 'photo mode' feature which includes an overlay in the lower left corner displaying the current galactic coordinates in glyphs. Using this feature with the screenshot feature, users can take and share images of their travels with coordinates included. An entire community of players engage in this activity on reddit.com/r/NMSCoordinateExchange.

The goal of the proposed project is to create a tool that can take in these screen captures, use machine learning to extract the glyphs, translate them to coordinates, and plot the regions in an intuitive, interactive 3D visualization (or several) in some kind of dashboard UI. Mapping will be from both a local directory of personally captured images representing my own explored space, as well as images scraped from Reddit to stand in as an approximation of globally explored space.

The project would initially start as a local-only Python analysis tool (package), with the possibility of being extended into a small web application.

Reason for Selection

I came up with this idea as a way to get some exposure to basic neural networks and computer vision while doing something I am personally interested in. I also expect to be able to further develop my data visualization skills using some tools I've been aware of but have not had the opportunity to explore.

Data Science

I think the image classification, machine learning, and data visualization elements here are enough to qualify this as a data science project. That said, I will be looking for ways to perform additional analysis on the information I scrape from the web. This project is much more descriptive than predictive analytics. While I understand the DS certificate program's focus on predictive analytics, I think descriptive analytics, particularly for spatial things like mapping, are very interesting and worth exploring further.

¹Each region has address space for 4096 systems but the vast majority of these are inaccessible 'phantom' systems that do not appear on the in-game galaxy map and cannot be reached normally.

Features

Core Features

- mapping personally explored space against space explored the Reddit community as a whole
- options to switch between several different visualizations:
 - 3d scatter plot
 - 3d surface plot
 - 2d heat map
- various quantitative analyses of community explored space and how users, planets, systems, and regions all relate to one another

Additional Features (aspirational)

- A more advanced computer vision model that can additionally determine the subject of a given image:
 - umbrella categorization:
 - * a planet in general
 - * a star ship, freighter, or frigate
 - * fauna
 - * multitool
 - subcategories
 - * lush planet
 - * 'strider' or 'diplo' fauna
 - * fan-wing hauler, x-wing fighter
 - * alien multitool
- the ability to hover over a given region and see some summary statistics about it
 - number of explored systems
 - known individuals to have explored the region
 - closest explored region
- the ability to hover over a given region and the image or gallery of images related to it
- a visualization of regions that appear to be unexplored
- a way to generate possible coordinates for the system farthest from explored space
- support for galaxies beyond the first
- a publicly accessible web application (low priority)

Tools

- Python
 - Pandas, Numpy, PyTorch, Plotly Dash
 - PRAW (Python Redditi API Wrapper)
- TensorFlow, OpenCV
 - possibly other, more appropriate ML/CV tools
 - possibility of adapting an existing OCR engine (like Tesseract)
 - possibly a transformer based on an imagenet model (to categorize images)
- Nvidia RAPIDS suite for GPU acceleration
- Docker
- HTML, JS, CSS
- Default to Django, possibly React or some other framework
 - If I get to a web interface, I'd rather not build it from scratch

Scope Tiers

Stage 1. Proof of Concept - Personal Map

- Scan local directory of personal images
 - Identical format
 - Single galaxy, platform, & version
- Process images
- Extract coordinates
- Visualize with interactive plot
 - base plotly

Stage 2. Dashboard UI

- Upgrade to modular dashboard UI
 - Plotly Dash
- Modular design for efficient addition of later features
- Plotly Dash (or other framework that easily extends to web applications)

Stage 2. Feature: Collective Map

- Scrape reddit for a collection of images
- Start with only initial galaxy, pc platform, & update 3.94
- Include overlaid map and separate personal/collective map

Stage 3. Scale Increase

- Add images from older updates, other platforms, additional galaxies, etc

Stage 4. Upgrade with Miscellaneous Optional Features

- Identify most isolated regions and generate random coordinates to attempt jumping to unexplored space.
- Find an efficient way to make it so you can hover over points and see the image(s) related to that point.
- Scrape with flair searches to add filterable attributes like galaxy, subject, platform
- Use subject specific scraped images to train a transformer to differentiate between base, ship, multi-tool, freighter, and flora/fauna
- Process uploaded images to ‘paint out’ undesirable obstructions (UI/HUD elements) with ML-based technique
- Galactic population density heat map
 - Individual regions have hundreds of systems
 - With sufficient sample size, collisions are likely
 - Some regions or clusters of regions may have images from dozens or hundreds of systems/planets

Stage 5. Publish to Web Application

- Containerization w/ Docker
- Basic web interface