Requirement 2 Report - Research Datasets

What are the most frequently used public Minecraft datasets in machine learning? How are these datasets structured and labeled?

MineDojo supplies several huge datasets for Minecraft machine learning [1]. They have 730,000+ narrated Minecraft YouTube videos divided into two datasets: one consisting of tutorials, illustrating step by step guides to tasks, and one of general gameplay. Each video includes time-aligned transcripts and accompanying metadata with a video id, duration, etc. They also have a dataset from scraping Minecraft wiki pages. This dataset organizes entries as Python dictionaries with fields metadata, screenshot, texts, images, sprites, and tables. The entries preserve the layout of the page within these fields. Lastly, MineDojo has a reddit dataset of 340,000 posts and 6.6 million comments. Each entry is again a Python dictionary with fields for id, title, type, content, comments, etc. All data is stored in JSON files with metadata and urls to content except the Reddit data, which requires obtaining a Reddit API key and running a script.

Another large dataset is the MineRL-v0 dataset from MineRL with up to 130-734 GB depending on the resolution [2]. This dataset was built from 500+ hours of recordings of human demonstrations of tasks for the MineRL competitions and is organized as state-action pairs. Each state is an image frame and metrics like inventory, health, hunger, etc., and the action consists of keyboard presses, camera movements, and crafting. Tasks comprising the dataset include navigating, chopping wood, obtaining a bed, obtaining meat, obtaining an iron pickaxe, obtaining diamond, finding a cave, creating a waterfall, creating an animal pen, and building a house near a village.

Lastly, OpenAI has a dataset that was created by recording video and action data (keyboard presses and mouse movements) from contractors playing Minecraft and using it to train an inverse dynamics model to label a larger unlabeled dataset [3]. The result is a 70,000 hours of labeled gameplay pairing state and actions. The majority of this dataset is general gameplay–similar to MineDojo's.

What kinds of gameplay do these datasets feature? Are they missing any key components of gameplay for our goals?

The MineDojo datasets and OpenAI dataset feature virtually all Minecraft gameplay imaginable. Referring to our requirements stack, the MineRL-v0 dataset may be insufficient for farming, illuminating regions, building shelter, and fighting hostile mobs.

How feasible is it to consolidate disparate datasets? Which datasets can and cannot be combined?

Videos with similar formatting and annotations are ideal to consolidate. It was demonstrated in OpenAI's Minecraft Machine Learning model that undocumented videos can be documented with another ML model “MineCLIP” using custom-created annotations with gameplay recordings. Similarly, NVIDIA trained their MineDojo model with Natural Language Processing input, including captions and external websites. This demonstrates how multiple mediums can be utilized in tandem, and any type of data capable of being integrated into the model can be combined as input. However, this scale of computing power is more feasible for tech giants, and it would be more practical for our project to choose one type. It would be best to use video in our use case, since it is an information-rich medium that closely resembles gameplay. If the command annotations are provided, that would be useful; but if not, we could consider adding our own or creating a system to process video captions.

What role will these datasets play in our AI's training?

These datasets are integrated into the model during the Imitation Learning phase of the Model Training process, if utilized. The choice of using this stage is discussed in Report 5. If it is decided to solely use Reinforcement Learning without Imitation Learning, the video input would not be necessary, and the agent would learn from scratch by interacting with the environment. However, incorporating the pre-training stage may significantly accelerate the process while having a higher degree of computational efficiency.

References

[1] “Building generally capable AI agents with Minedojo,” NVIDIA Technical Blog, https://developer.nvidia.com/blog/building-generally-capable-ai-agents-with-minedojo/ (accessed Sep. 24, 2023).

[2] “Learning to play Minecraft with video pretraining,” Learning to play Minecraft with Video PreTraining, https://openai.com/research/vpt (accessed Sep. 24, 2023).

[3] W. H. Guss et al., “Minerl: A large-scale dataset of Minecraft demonstrations,” arXiv.org, https://arxiv.org/abs/1907.13440 (accessed Sep. 23, 2023).

[4] “Use the knowledge base,” Use the Knowledge Base - MineDojo 0.1.0 documentation, https://docs.minedojo.org/sections/getting_started/data.html (accessed Sep. 23, 2023).

[5] B. Baker et al., “Video pretraining (VPT): Learning to act by watching unlabeled online videos,” arXiv.org, https://doi.org/10.48550/arXiv.2206.11795 (accessed Sep. 23, 2023).