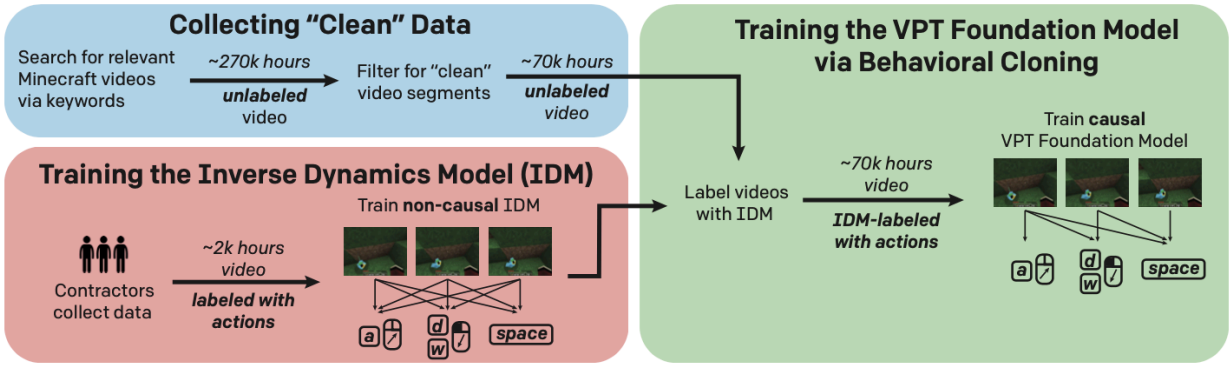
English Summary of Article 2206.11795 by Ramtin Kosari

Video PreTraining

Learning to Act by Watching Unlabeled Online Videos



Abstract

Pretraining on large, noisy internet datasets is a well-known method for training models in various domains like text and images. However, in sequential decision environments such as robotics and video games, labeled data may not always be readily available. To tackle this issue, this article introduces a new approach called semi-supervised imitation learning, where agents learn by observing unlabeled online videos. By utilizing a small amount of labeled data, we can train a model to accurately label a vast source of unlabeled data, such as videos of people playing Minecraft. This trained model can perform complex tasks with minimal supervision and can even outperform humans in certain scenarios. Additionally, we demonstrate that this model can be fine-tuned with reinforcement learning to tackle challenging tasks, showcasing its versatility and efficiency. This work represents a significant advancement in the field of artificial intelligence, with these agents showcasing exceptional abilities such as crafting diamond tools in record time.

OpenAI VPT Project

OpenAI trained a neural network to play Minecraft by Video PreTraining (VPT) on a massive unlabeled video dataset of human Minecraft play, while using only a small amount of

labeled contractor data. With fine-tuning, our model can learn to craft diamond tools, a task that usually takes proficient humans over 20 minutes (24,000 actions). Our model uses the native human interface of keypresses and mouse movements, making it quite general, and represents a step towards general computer-using agents.

Why Minecraft ?

OpenAI choses to validate its method in Minecraft because it (1) is one of the most actively played video games in the world and thus has a wealth of freely available video data and (2) is open-ended with a wide variety of things to do, similar to real-world applications such as computer usage.

Foundation Model

Model has been trained on 70,000 hours of IDM-labeled online video, our behavioral cloning model (the “VPT foundation model”) accomplishes tasks in Minecraft that are nearly impossible to achieve with reinforcement learning from scratch. It learns to chop down trees to collect logs, craft those logs into planks, and then craft those planks into a crafting table; this sequence takes a human proficient in Minecraft approximately 50 seconds or 1,000 consecutive game actions.

Fine-tuning with Behavioral Cloning

Foundation models are designed to have a broad behavior profile and be generally capable across a wide variety of tasks. To incorporate new knowledge or allow them to specialize on a narrower task distribution, it is common practice to fine-tune these models to smaller, more specific datasets. As a case study into how well the VPT foundation model can be fine-tuned to downstream datasets, we asked our contractors to play for 10 minutes in brand new Minecraft worlds and build a house from basic Minecraft materials. We hoped that this would amplify the foundation model’s ability to reliably perform “early game” skills such as building crafting tables. When fine-tuning to this dataset, not only do we see a massive improvement in reliably performing the early game skills already present in the foundation model, but the fine-tuned model also learns to go even deeper into the technology tree by crafting both wooden and stone tools. Sometimes we even see some

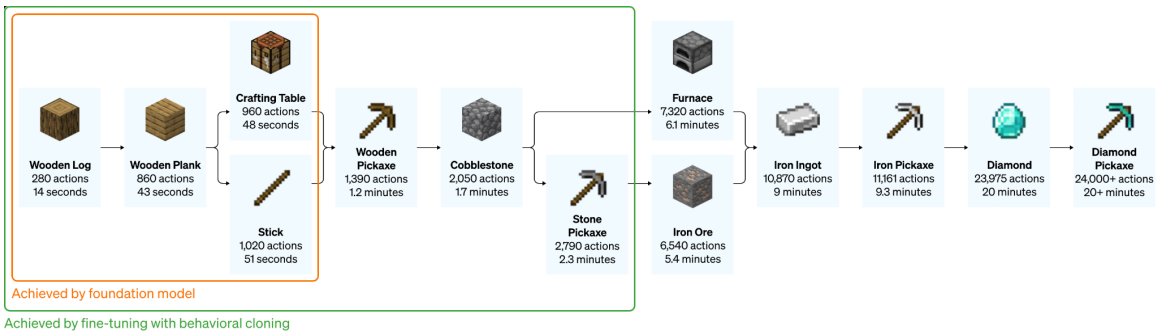
rudimentary shelter construction and the agent searching through villages, including raiding chests.

Fine-tuning with Reinforcement Learning

When it is possible to specify a reward function, reinforcement learning (RL) can be a powerful method for eliciting high, potentially even super-human, performance. However, many tasks require overcoming hard exploration challenges, and most RL methods tackle these with random exploration priors, e.g. models are often incentivized to act randomly via entropy bonuses. The VPT model should be a much better prior for RL because emulating human behavior is likely much more helpful than taking random actions. We set our model the challenging task of collecting a diamond pickaxe, an unprecedented capability in Minecraft made all the more difficult when using the native human interface.

Crafting a diamond pickaxe requires a long and complicated sequence of subtasks. To make this task tractable, we reward agents for each item in the sequence.

We found that an RL policy trained from a random initialization (the standard RL method) barely achieves any reward, never learning to collect logs and only rarely collecting sticks. In stark contrast, fine-tuning from a VPT model not only learns to craft diamond pickaxes (which it does in 2.5% of 10-minute Minecraft episodes), but it even has a human-level success rate at collecting all items leading up to the diamond pickaxe. This is the first time anyone has shown a computer agent capable of crafting diamond tools in Minecraft, which takes humans over 20 minutes (24,000 actions) on average.



Conclusion

VPT paves the path toward allowing agents to learn to act by watching the vast numbers of videos on the internet. Compared to generative video modeling or contrastive methods that would only yield representational priors, VPT offers the exciting possibility of directly learning large scale behavioral priors in more domains than just language. While we only experiment in Minecraft, the game is very open-ended and the native human interface (mouse and keyboard) is very generic, so we believe our results bode well for other similar domains, e.g. computer usage.