

Рецензия

Video PreTraining (VPT): Learning to Act by Watching Unlabeled Online Videos

Bowen Baker^{*†}
bowen@openai.com

Ilge Akkaya^{*†}
ilge@openai.com

Peter Zhokhov^{*†}
peterz@openai.com

Joost Huizinga^{*†}
joost@openai.com

Jie Tang^{*†}
jietang@openai.com

Adrien Ecoffet^{*†}
adrien@openai.com

Brandon Houghton^{*†}
brandon@openai.com

Raul Sampedro^{*†}
raulsamg@gmail.com

Jeff Clune^{*†‡}
jclune@gmail.com

Основная идея

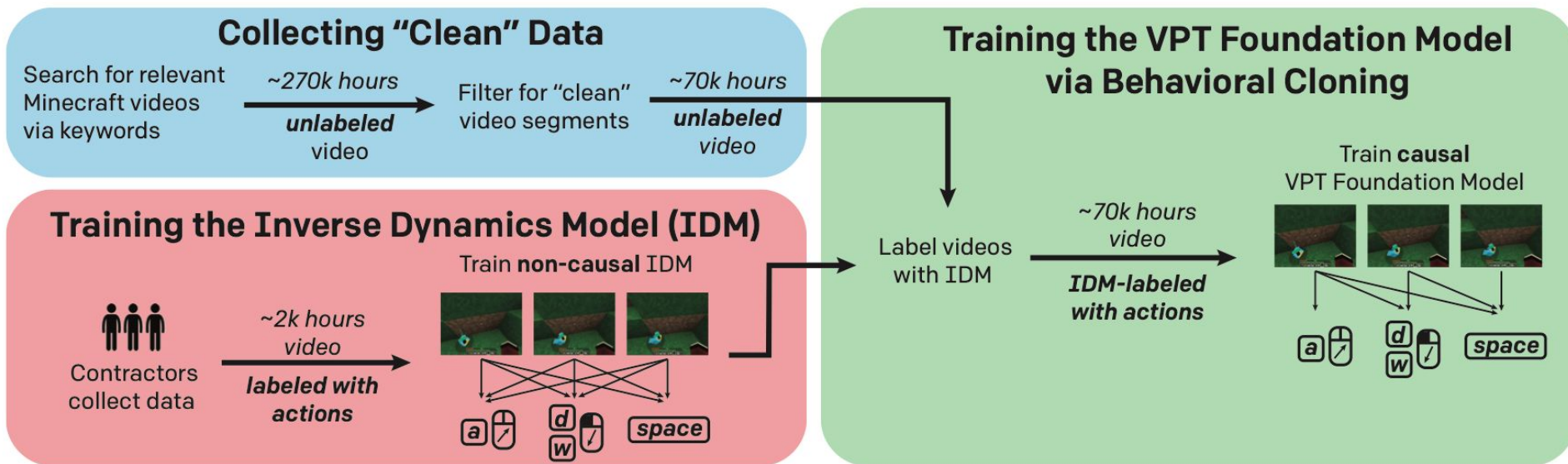


Figure 2: Video Pretraining (VPT) Method Overview.

Авторы



Bowen Baker^{*†}
bowen@openai.com

Ilge Akkaya^{*†}
ilge@openai.com

Peter Zhokhov^{*†}
peterz@openai.com

Joost Huizinga^{*†}
joost@openai.com

Jie Tang^{*†}
jietang@openai.com

Adrien Ecoffet^{*†}
adrien@openai.com

Brandon Houghton^{*†}
brandon@openai.com

Raul Sampedro^{*†}
raulsamg@gmail.com

Jeff Clune^{*†‡}
jclune@gmail.com

Почему Minecraft



- Большое количество данных в интернете
- Большой выбор действий
- Сложного графического интерфейса пользователя.

Работы в области Minecraft

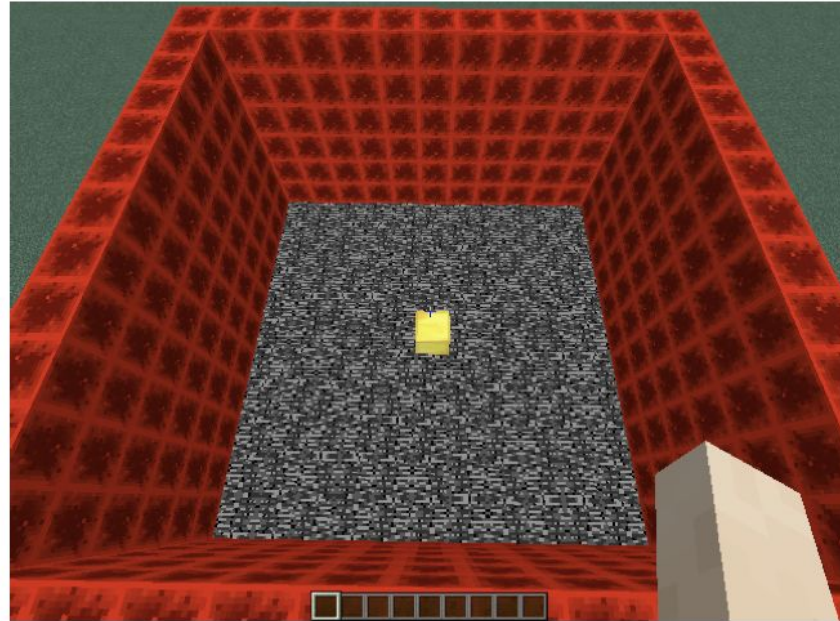
Задачи:



- Навигация
- Работа с блоками
- Следование инструкции
- Восхождение на холмы
- Добыча алмазов

Работы в области Minecraft

Навигация



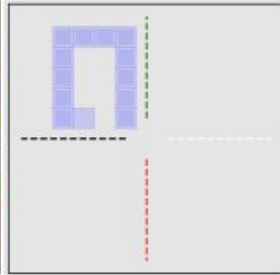
<https://arxiv.org/pdf/1902.04257.pdf>

Работы в области Minecraft

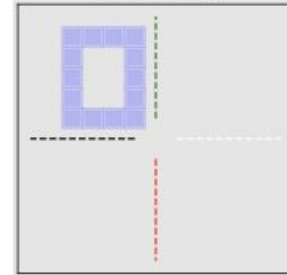
Работа с блоками



Agent State
(visual input) (structure)



Episode
Goal

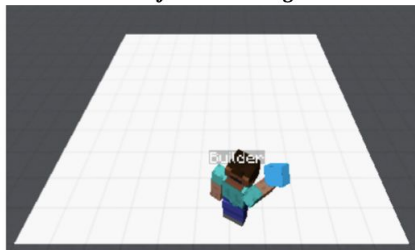


Работы в области Minecraft

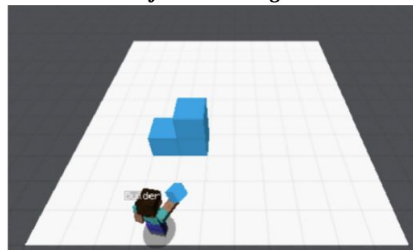
Выполнение инструкций



World State: Before the dialogue



World State: After the dialogue



Dialogue Context

Architect : Apparently, this first structure is a warm up
Architect : It's just 3 blocks arranged in an L shape
Builder : Oh cool
Architect : So two blocks next to each other on the ground
Builder : Any color?
Architect : And then one more block on top of either
Architect : They are blue!

Imitation learning

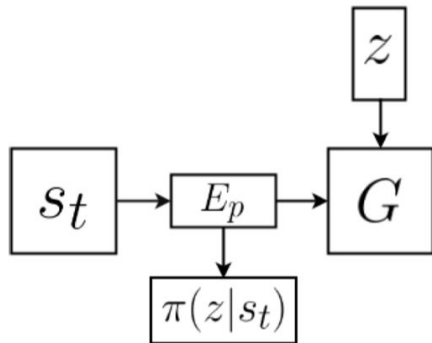


Generative Adversarial Imitation Learning

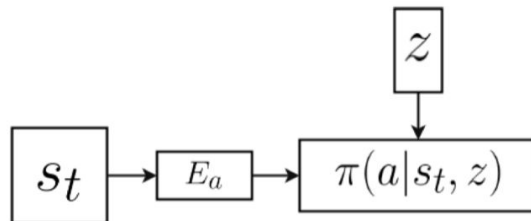
<https://arxiv.org/pdf/1606.03476.pdf>

Imitation learning

Imitating Latent Policies from Observation



(a) Latent Policy Network



(b) Action Remapping Network

Figure 1: The latent policy network learns a latent policy, $\pi(z|s)$, and a forward dynamics model, G . The action remapping network learns $\pi(a|s_t, z)$ to align the latent actions z with ground-truth actions a . We train embeddings, E_a and E_p , concurrently with each network.

Imitation learning



Behavioral Cloning from Observation

<https://arxiv.org/pdf/1805.01954.pdf>

Отличие

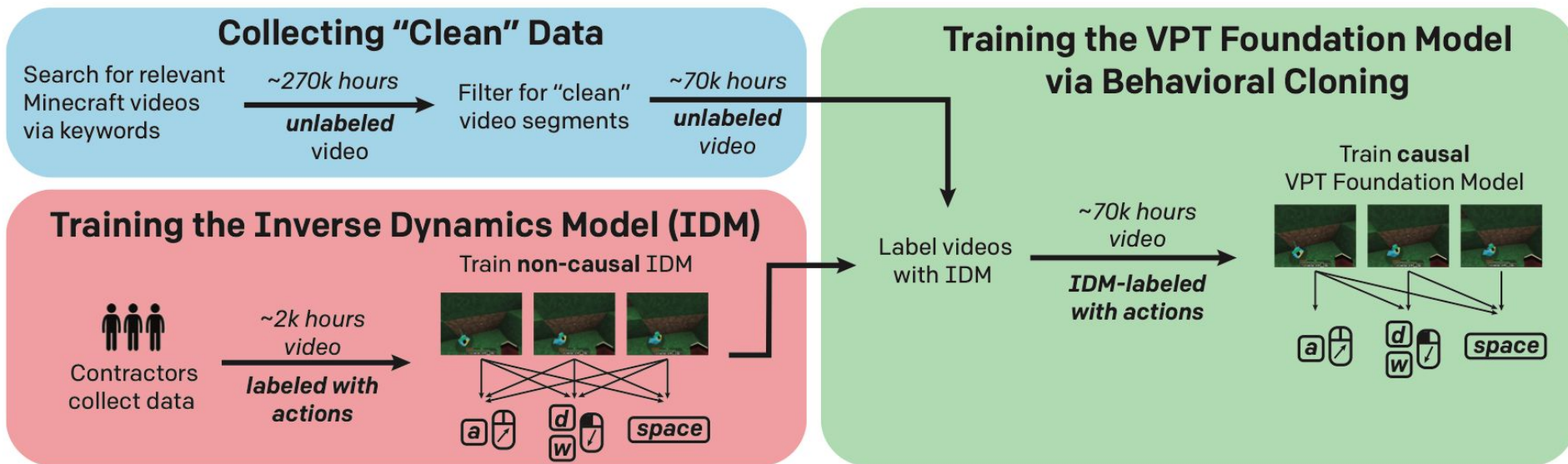


Figure 2: Video Pretraining (VPT) Method Overview.

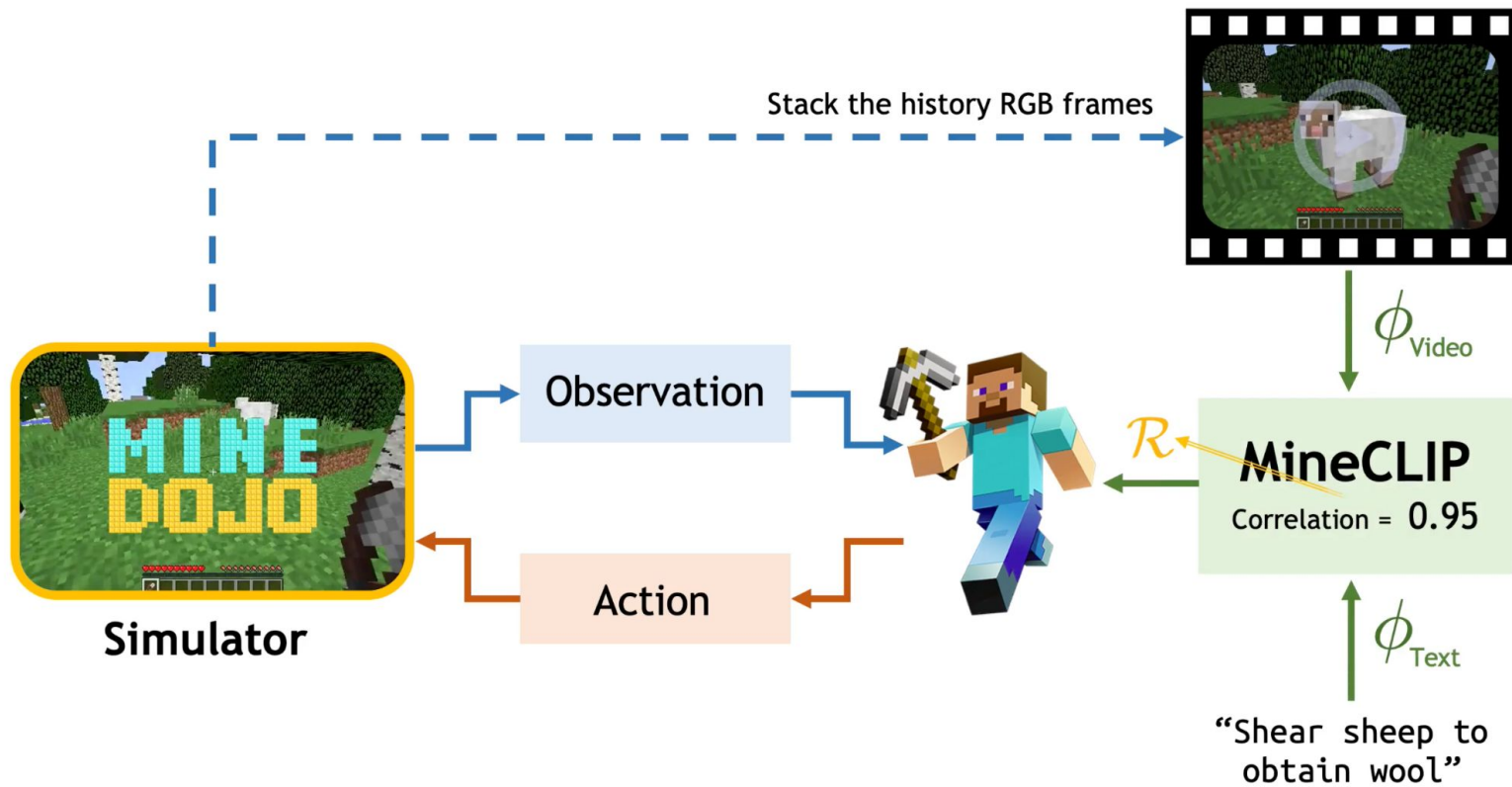
Цитирование



16 цитирований

- Улучшение алгоритма сбора данных
- Использование основной идеи в своих работах

MINEDOJO



Video PreTraining (VPT): Learning to Act by Watching Unlabeled Online Videos

Bowen Baker^{*†}
bowen@openai.com

Ilge Akkaya^{*†}
ilge@openai.com

Peter Zhokhov^{*†}
peterz@openai.com

Joost Huizinga^{*†}
joost@openai.com

Jie Tang^{*†}
jietang@openai.com

Adrien Ecoffet^{*†}
adrien@openai.com

Brandon Houghton^{*†}
brandon@openai.com

Raul Sampedro^{*†}
raulsamg@gmail.com

Jeff Clune^{*†‡}
jclune@gmail.com