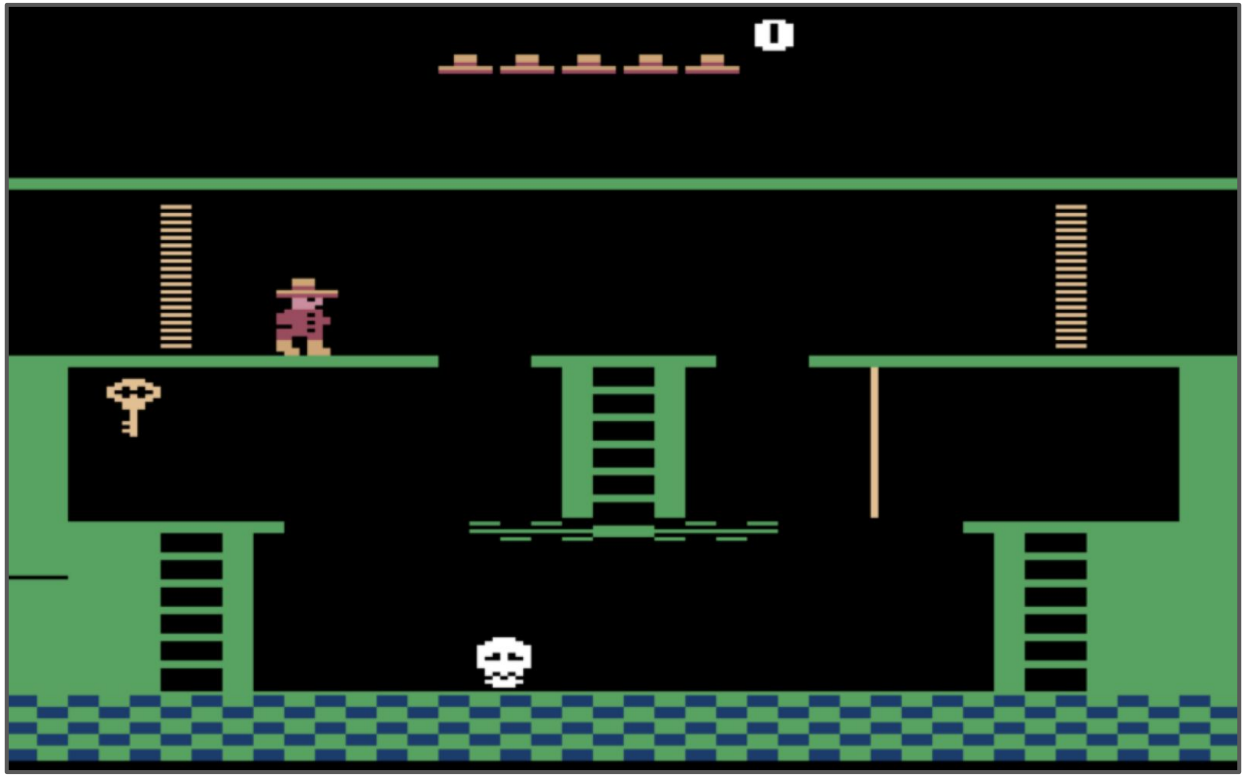
Deep Reinforcement Learning Doesn't Work Yet

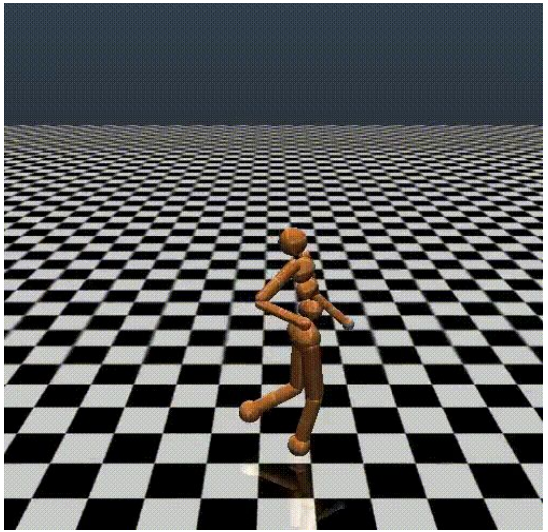
Alex Irpan

Обучение с подкреплением

Глубинное обучение с подкреплением

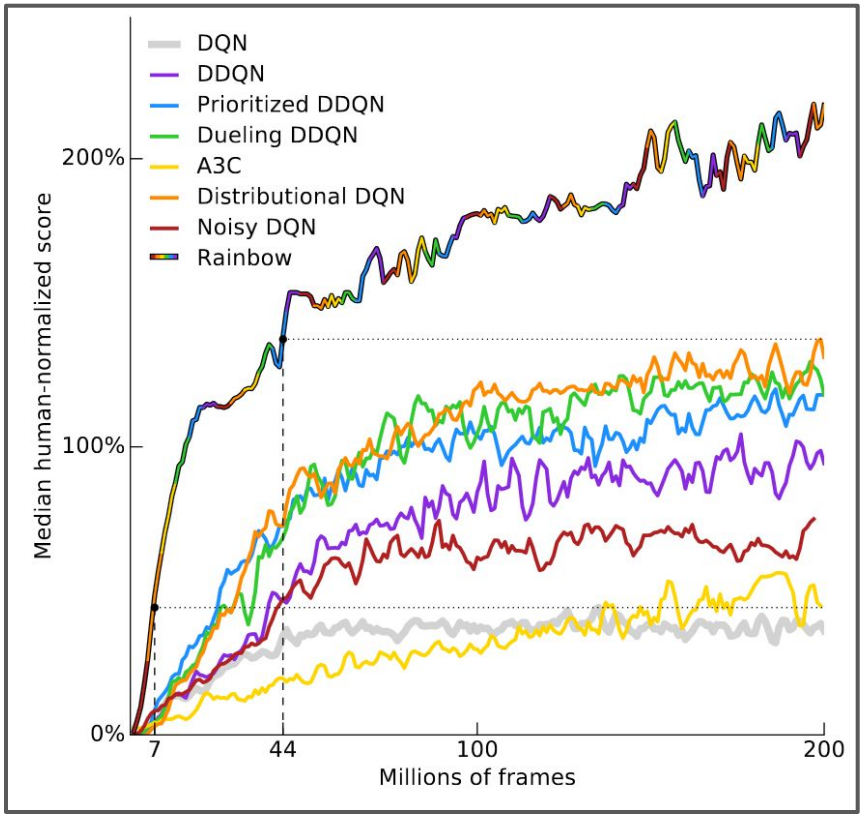


Успешные применения Deep RL



НО

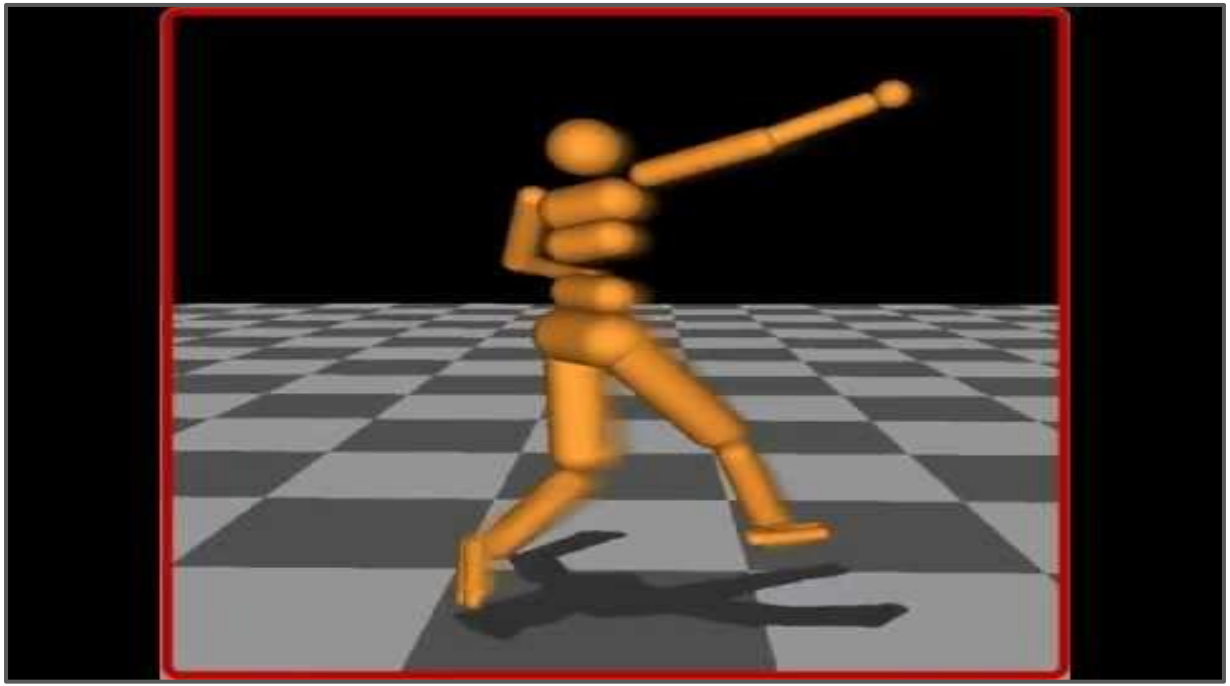
Неэффективность



- 100% достигаются за 18 миллионов фреймов (~ 83 часа игры)
- Можно ли игнорировать эффективность?

Другие методы часто справляются лучше

- Нужен только результат - можно использовать знания об окружении



- Atari

Agent	*B.Rider*	*Breakout*	*Enduro*	*Pong*	*Q*bert*	*Seaquest*	*S.Invaders*
DQN	4092	168	470	20	1952	1705	581
-best	5184	225	661	21	4500	1740	1075

Agent	*B.Rider*	*Breakout*	*Enduro*	*Pong*	*Q*bert*	*Seaquest*	*S.Invaders*
UCT	7233	406	788	21	18850	3257	2354

- Boston Dynamics не используют RL?

Обычно нужна функция награды

Сложность разработки функции вознаграждения

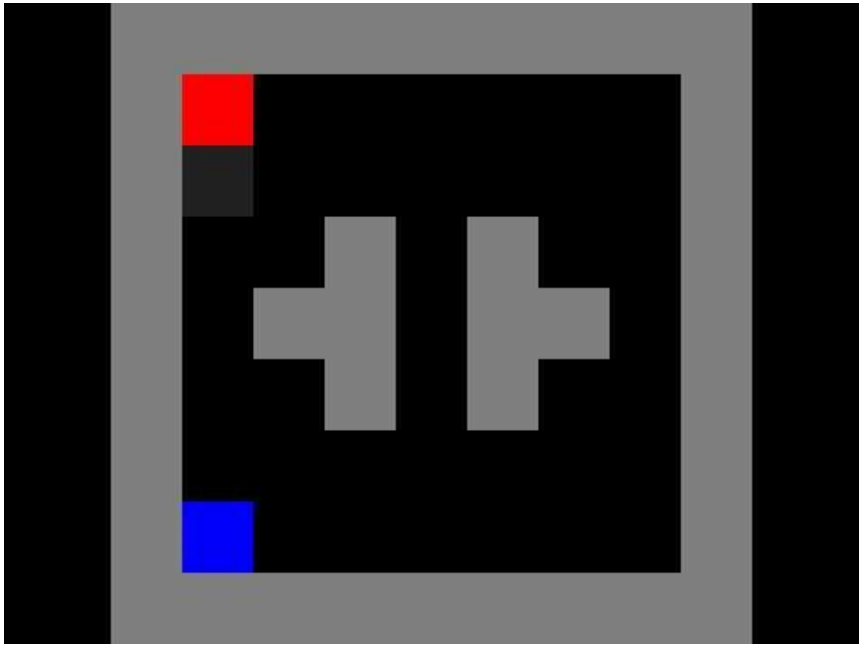
Model	ROUGE-1	ROUGE-L
Nallapati et al. 2016 (abstractive)	35.46	32.65
Nallapati et al. 2017 (extractive baseline)	39.2	35.5
Nallapati et al. 2017 (extractive)	39.6	35.3
See et al. 2017 (abstractive)	39.53*	36.38*
Our model (RL only)	**41.16**	**39.08**
Our model (supervised+RL)	39.87	36.90

Баттона лишили его 100-й гонки за McLaren после того, как ERS не пустила его на старт. Так завершились неудачные выходные для британца. Баттона опередили в квалификации. Финишировал впереди Нико Росберга в Бахрейне. У Льюиса Хэмилтона. В 11 гонках.. Гонка. Чтобы лидировать 2000 кругов.. В... И. — Paulus et al, 2017

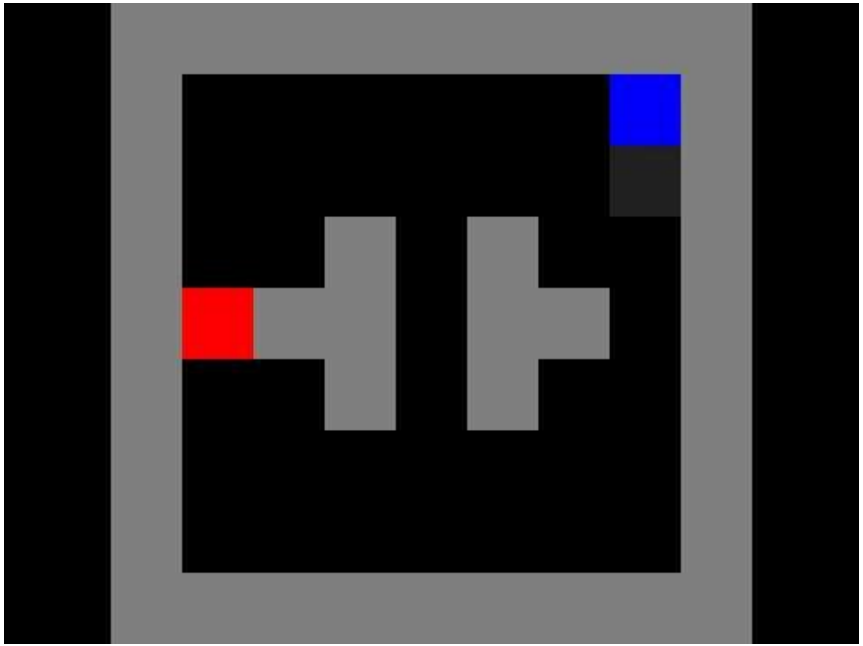
Трудно избежать локального оптимума

Возможно переобучение

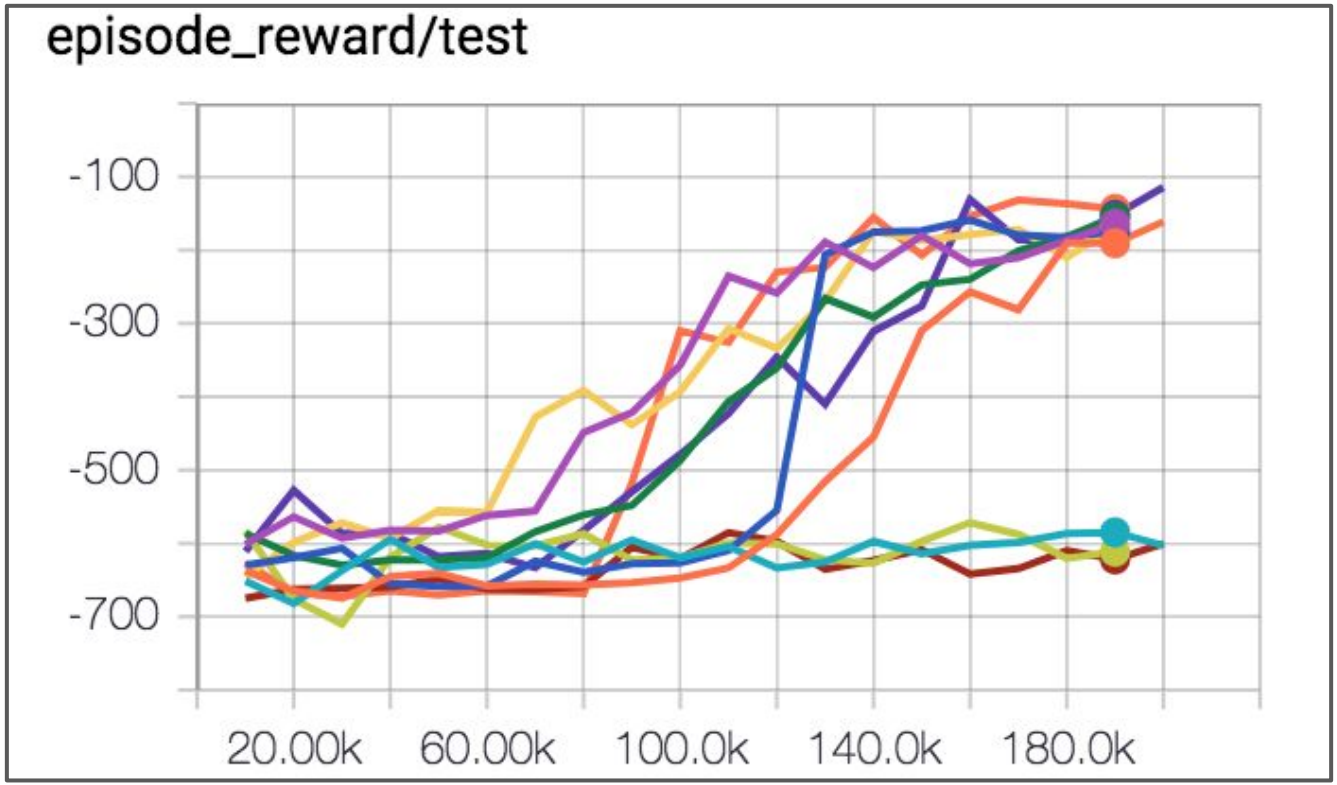
Обученные вместе



Обученные раздельно



Нестабильность результатов



Когда применять все же стоит

- Возможна генерация практически неограниченного опыта
- Возможно упрощение задачи
- Есть возможность обучаться в самостоятельной игре
- Есть способ дать максимально ясное вознаграждение
- Если нужно дать непрерывное вознаграждение, то оно должно быть максимально подробным

Взгляд в будущее

- Локальные оптимумы - не так плохо
- Железо поможет
- Добавление сигналов обучения
- Обучаемые функции награды
- Transfer learning
- Сложные окружения могут оказаться проще