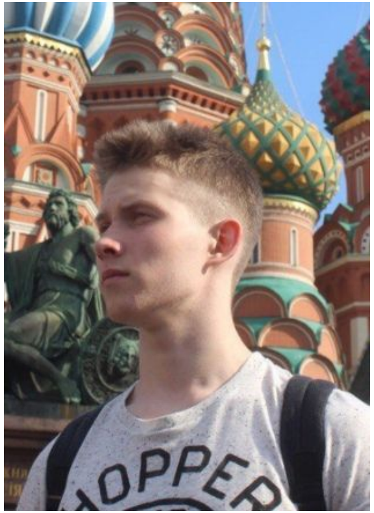
NeuroAI: IGLU in Minecraft, Silent Builder Solution

Linar Abdrazakov



Igor Churin

NEURAL INFORMATION PROCESSING SYSTEMS

MIPT

Agenda

- How it started
- What NeurIPS is
- IGLU in Minecraft competition
- Our solution
- Results

How it started

РАИИ
ЛЕТНЯЯ
ШКОЛА

Сириус
Образовательный центр

Date: 4-17 July 2021
Place: Sochi, Sirius

NeurIPS

- Conference on Neural Information Processing Systems
- One of the most influential conferences gathering the best ML engineers, data scientists, and artificial intelligence researchers from around the world

NEURAL INFORMATION PROCESSING SYSTEMS

https://nips.cc/

NeurIPS 2021 Competition Track

- **Diamond: A MineRL Competition on Training Sample-Efficient Agents**

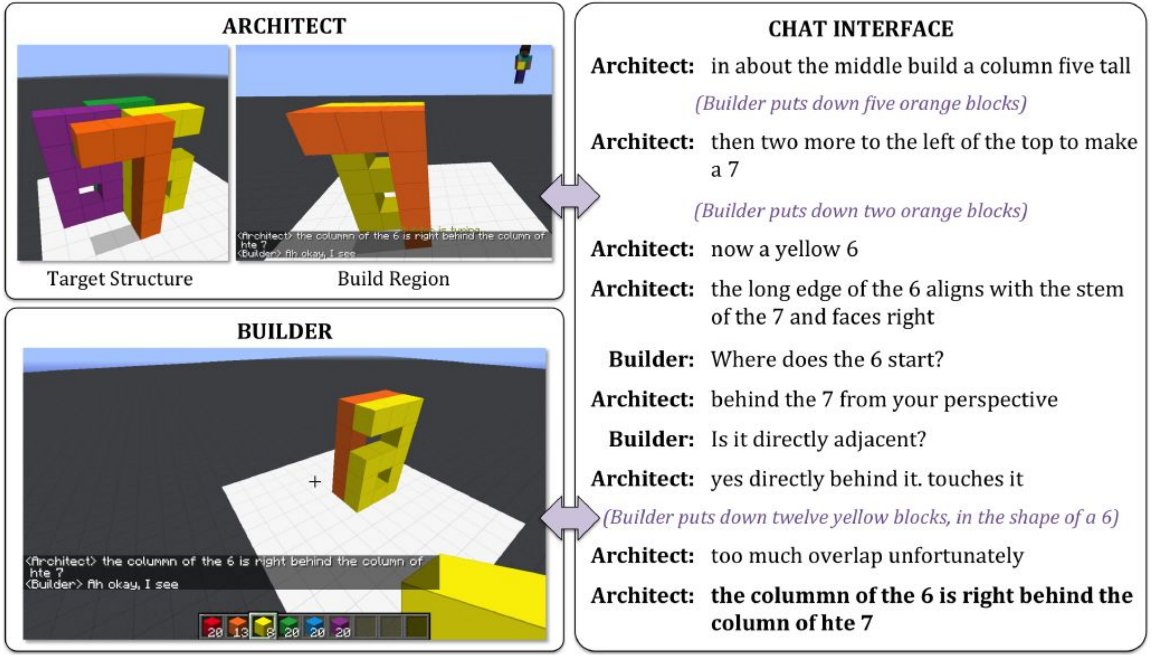
- **Shifts Challenge: Robustness and Uncertainty under Real-World Distributional Shift**

https://neurips.cc/Conferences/2021/CompetitionTrack

IGLU in Minecraft

Interactive Grounded Language Understanding in a Collaborative Environment

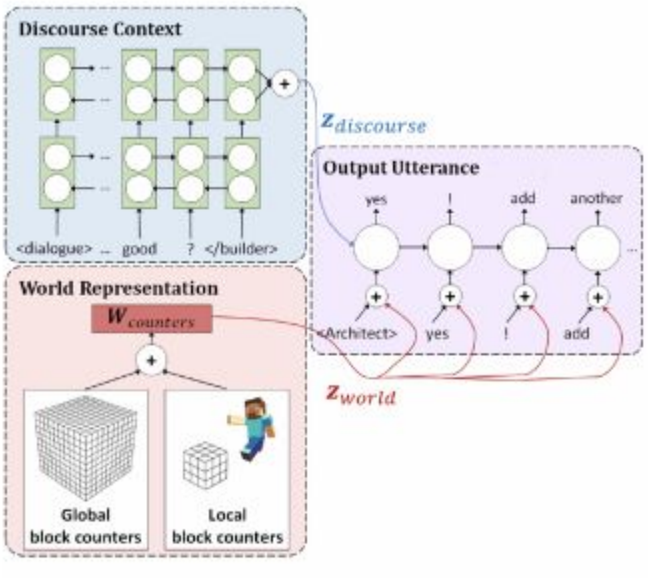
The goal of the competition is to approach the following scientific challenge: *How to build interactive agents that learn to solve a task while provided with grounded natural language instructions in a collaborative environment?*



https://www.iglu-contest.net/

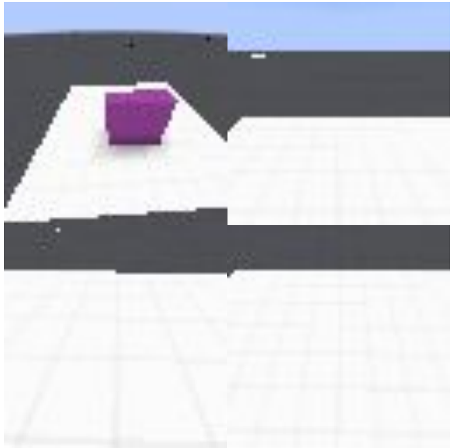
Architect Task

- To generate step instructions for the builder
- Architect is conditioned on half-finished structure and dialog context
- Evaluation using BLEU and keyword Precision/Recall

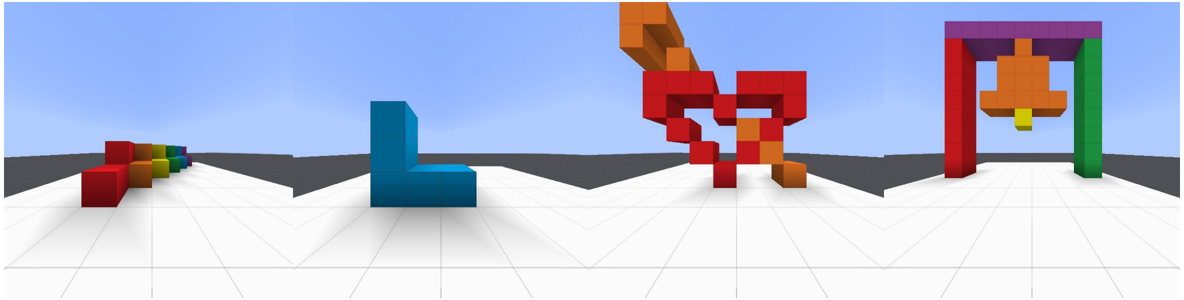


https://www.iglu-contest.net/architect-task

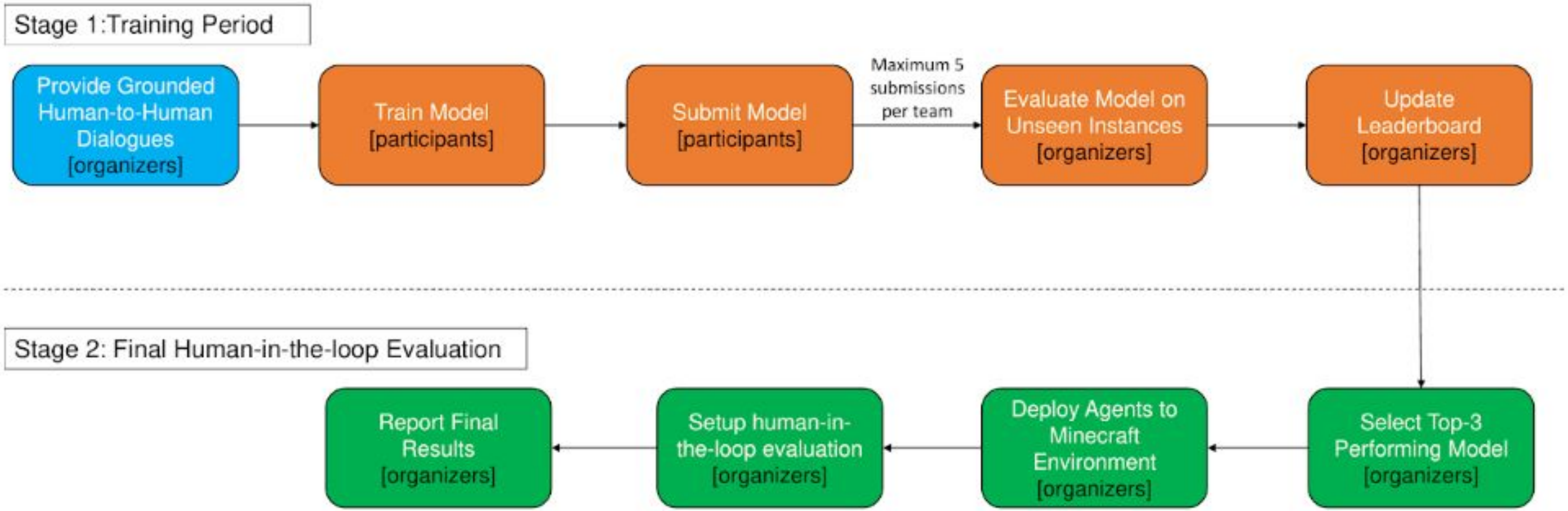
Silent Builder Task



- The goal is to build a target structure given human instructions
- Builder is able to navigate, place and break blocks
- Agent should analyze past conversations to reproduce spatial structures



Competition pipeline



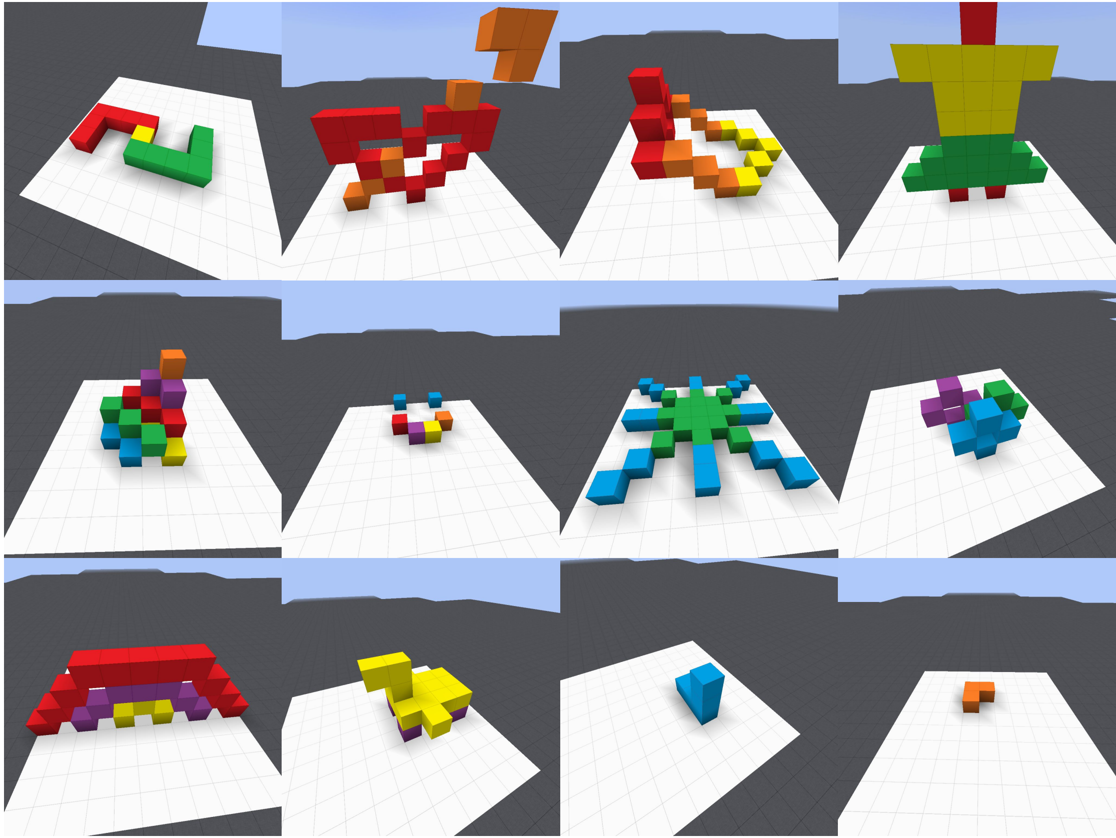
https://www.iglu-contest.net/home

Competition timeline

- July 26 - Stage 1 begins
- (Tentive) October 15 - Stage 1 ends
- October 22 - Stage 2 begins by deploying the top-3 performing agents for human evaluation
- November 26 - The results of Stage 2 are posted, and the list of winning teams per task is released
- December 6 - NeurIPS 2021 begins

Multigoal environment

- Builder task of IGLU features a wide range of different goals
- More than 150 goals
- The difficulty ranges from simple one color 3-6 block goals to complex ones of all six colors



Environment components

Observations:

```
Dict({
    "pov": Box(low=0, high=255, shape=(64, 64, 3)),
    "inventory": Box(low=0, high=20, shape=(6,)),
    "agentPos": Box(low= [-5, 0, -5, 0,  -90],
                    high=[ 5, 8,  5, 360, 90],
                    shape=(5,)),
    "grid": Box(low=-1, high=5, shape=(9, 11, 11)),
    "compass": Dict({"angle": Box(low=-180.0, high=180.0, shape=())}),
    "chat": String()
})
```

Action spaces:

- Human-like movement
- Discrete movement
- “Creative mode” movement

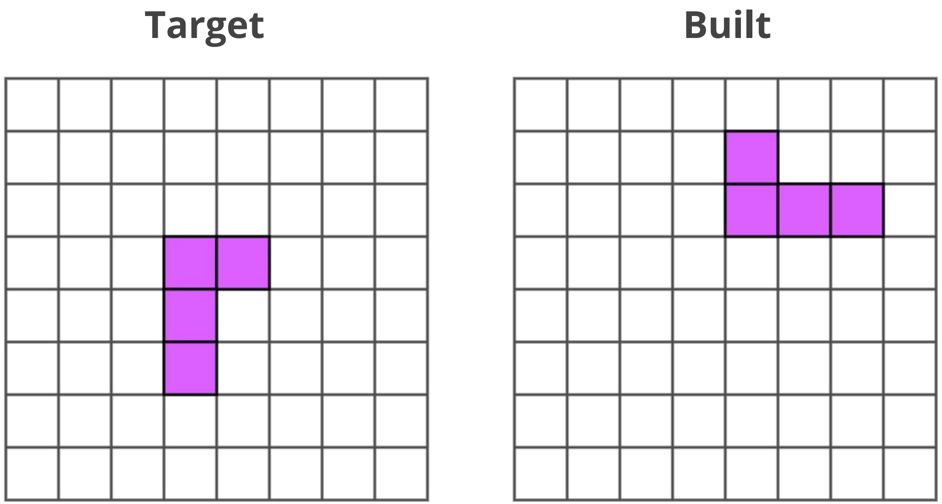
IGLU env repository link

> 2k downloads!

https://github.com/iglu-contest/iglu

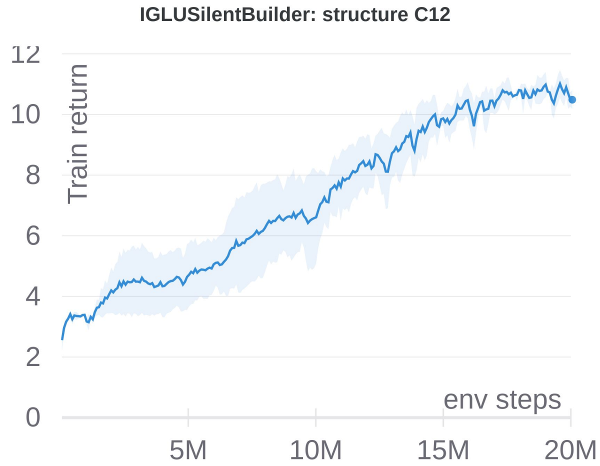
Rewards for building

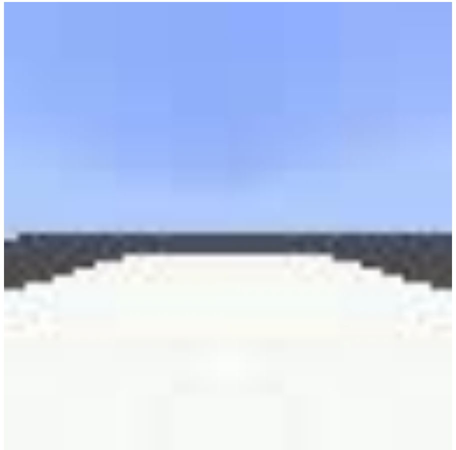
- Is the task solved? - Yes!
- By this, we introduce a bias, yet make the task more accessible for RL agent



Single Task Builder baselines

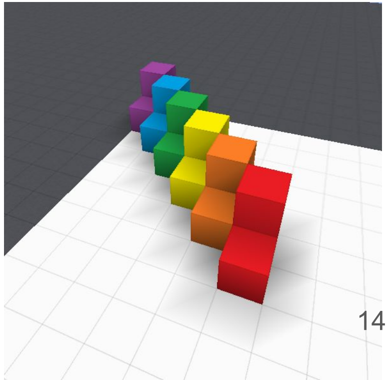
- Model-free RL baseline: IMPALA
- Agent acts given a visual input
- Trains in a day with 20 workers
- Does not use text information (as it's single task)





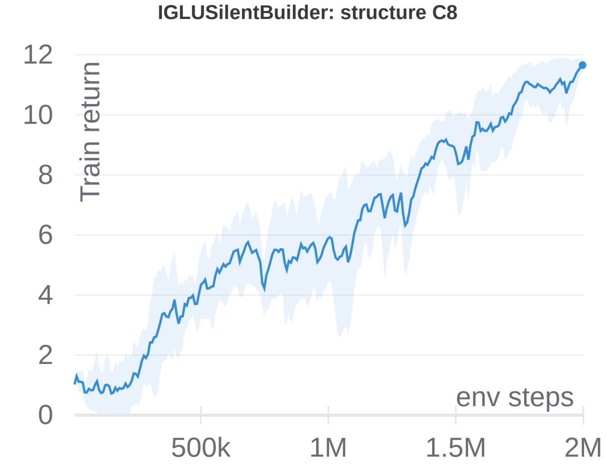
RLlib baselines github repo

Goal: C12 (18 blocks)



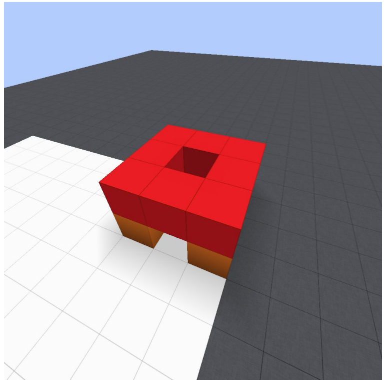
Single Task Builder baselines

- Model-based RL baseline: Dreamer
- Agent acts given a visual input
- Trains in a day with just one GPU and one env worker



Dreamer baseline github repo

Goal: C8 (12 blocks)



DreamerV2

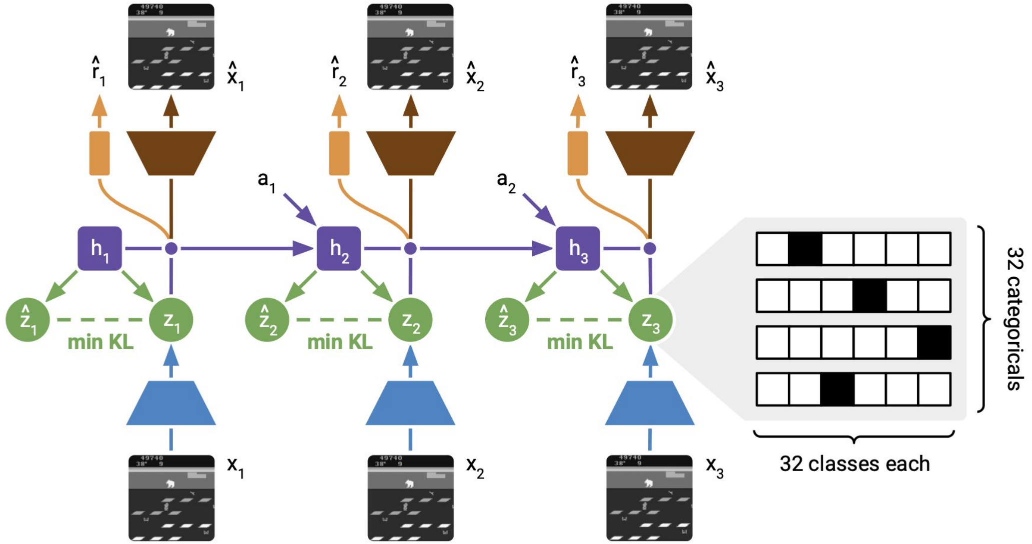


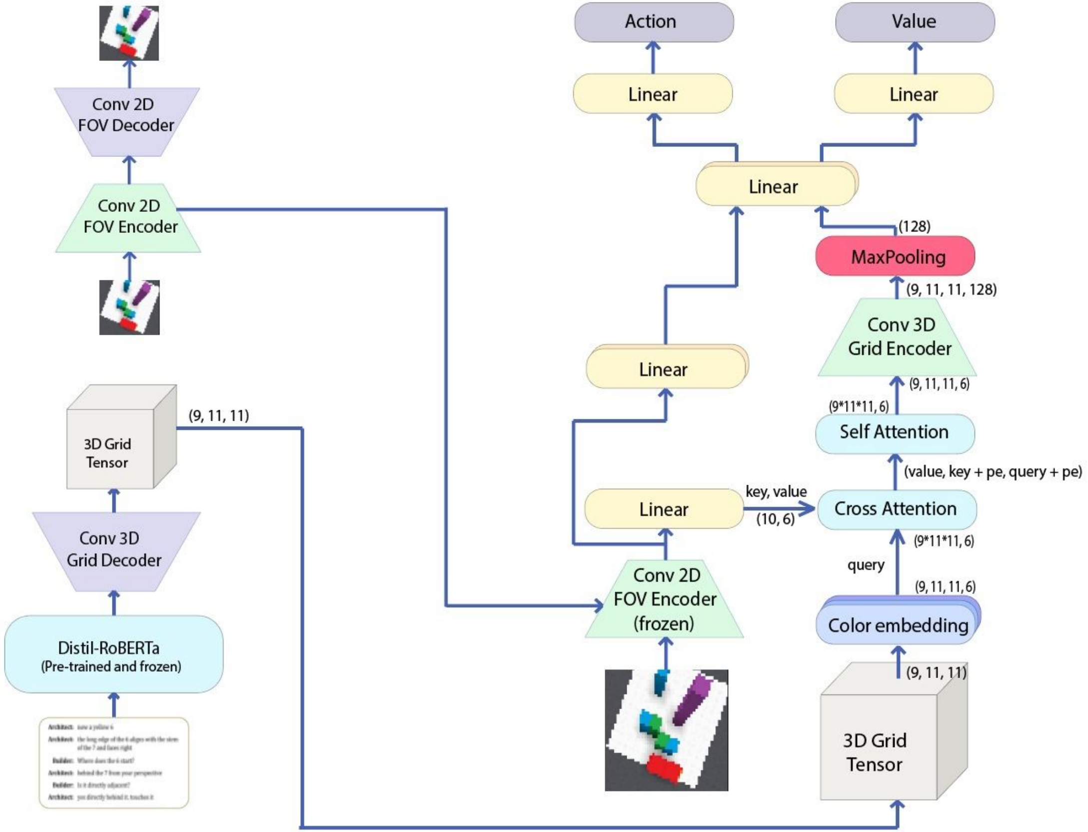
Figure 2: World Model Learning. The training sequence of images x_t is encoded using the CNN. The RSSM uses a sequence of deterministic recurrent states h_t. At each step, it computes a posterior stochastic state z_t that incorporates information about the current image x_t, as well as a prior stochastic state $\hat{z}_t$ that tries to predict the posterior without access to the current image. Unlike in PlaNet and DreamerV1, the stochastic state of DreamerV2 is a vector of multiple categorical variables. The learned prior is used for imagination, as shown in Figure 3. The KL loss both trains the prior and regularizes how much information the posterior incorporates from the image. The regularization increases robustness to novel inputs. It also encourages reusing existing information from past steps to predict rewards and reconstruct images, thus learning long-term dependencies.

https://arxiv.org/pdf/2010.02193.pdf

Our Solution

Visual
Autoencoder
(both)

Text2Target
Network
(Igor Churin)

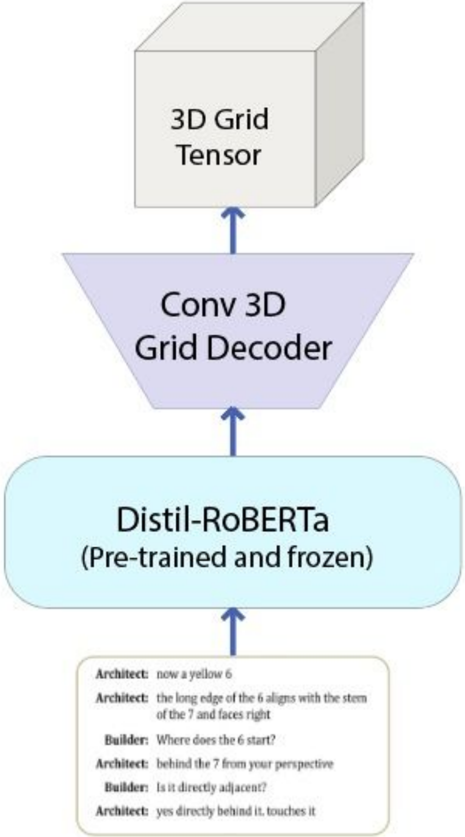


Policy Network
(Linar Abdrazakov)

Positional Encoding: $PE = [sin(x_{pos}/3),\ cos(x_{pos}/3),\ sin(y_{pos}/3),\ cos(y_{pos}/3),\ sin(z_{pos}/3),\ cos(z_{pos}/3)]$

Text2Target Network

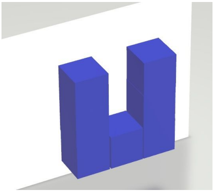
- We use Distil-RoBERTa for higher quality embedding of the whole chat
- We give the chat to the input of the transformer encoder as one sequence



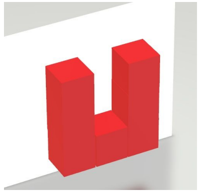
Data Augmentation

Changing colors in chat and target correspondingly (from 154 tasks up to 2850)

Architect: now place **blue** blocks…



Architect: now place **red** blocks…



Randomly removed questions from builder (from 2850 tasks up to 200 000+)

Architect: now a yellow 6

Builder: where does the 6 start?

Architect: behind the 7 from your perspective

Builder: is directly adjacent?

Architect: yes, directly behind it. touches it.

Architect: now a yellow 6

Builder: where does the 6 start?

Architect: behind the 7 from your perspective

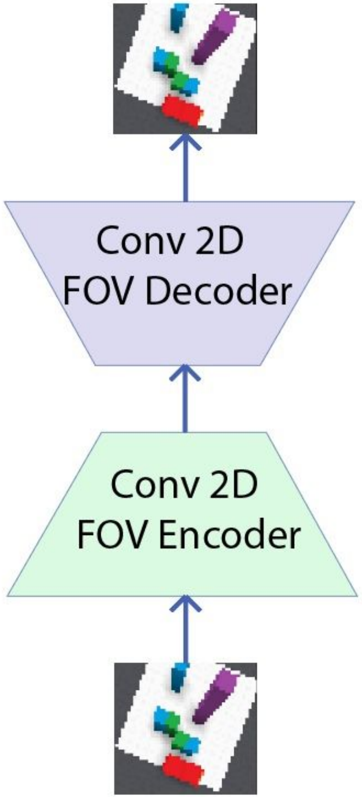
~~Builder: is directly adjacent?~~

Architect: yes, directly behind it. touches it.

Visual Autoencoder

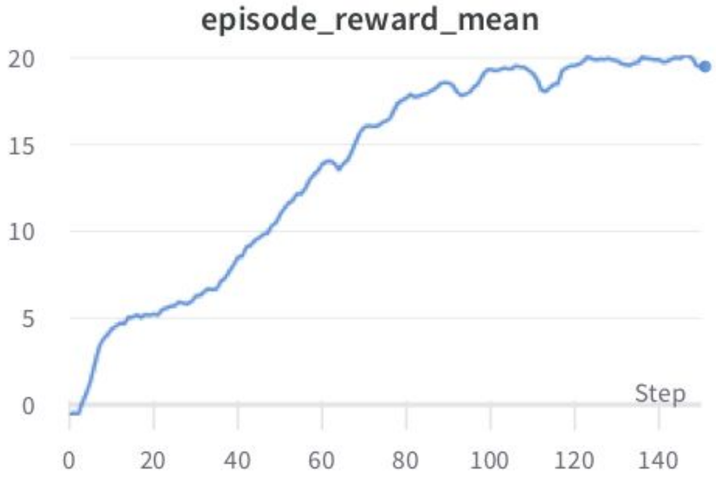
- Dataset was collected with a random agent

- We tuned model and its hyperparameters for better reconstruction

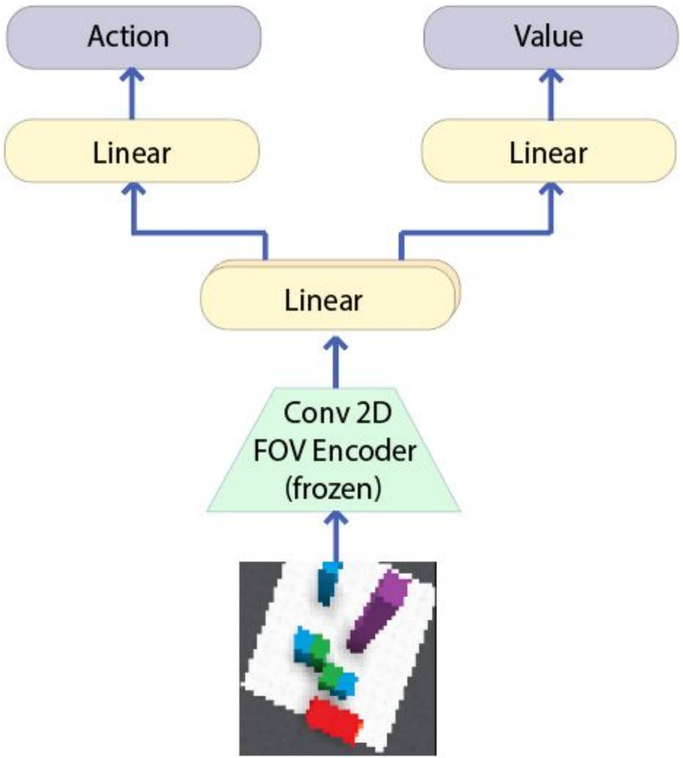


Our Single Task Baseline

- 160 times more sample efficient on task C12 comparing to the IMPALA baseline
- 15 times more sample efficient on task C8 comparing to the DreamerV2 baseline



Learning on C12



Fine-tuning with an unfrozen FOV encoder

- After training policy in one of our experiments, we tried to unfreeze FOV encoder and to fine-tune neural network.
- Surprisingly, the mean reward started to decline.
- After that, all of the training experiments were conducted with a frozen POV encoder without fine-tuning it.

Policy Network

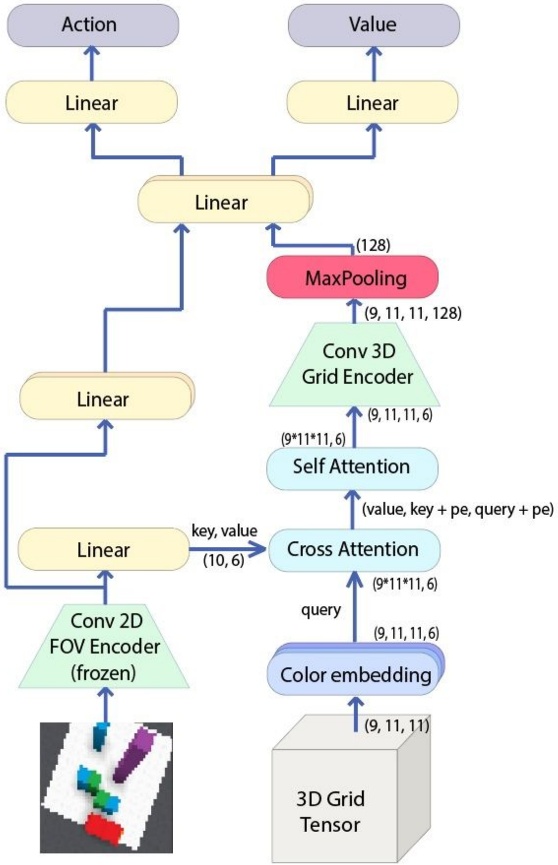
Method: PPO

Framework: RLlib

Parameters:

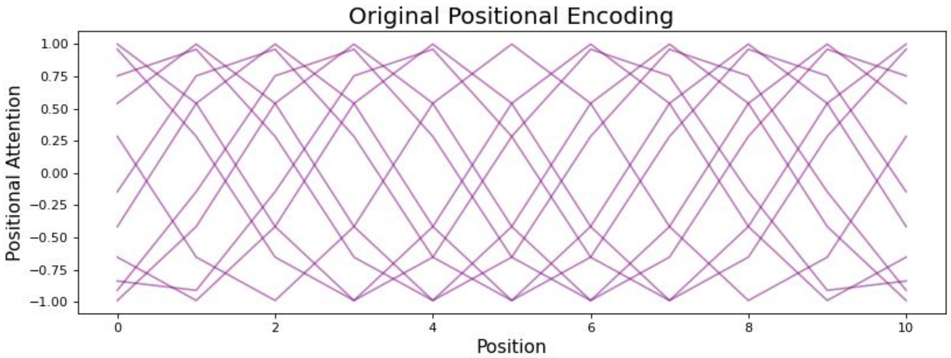
- sgd_minibatch_size: 60
- entropy_coeff: 0.01
- lambda: 0.95
- train_batch_size: 5000

Color embeddings are used for the target tensor. Then, cross-attention is applied to fuze data of different modalities. Self-attention and convolution layers are needed to process target grid features consider local dependencies. Max-Pooling allows to make it invariant to target tensor shifts.

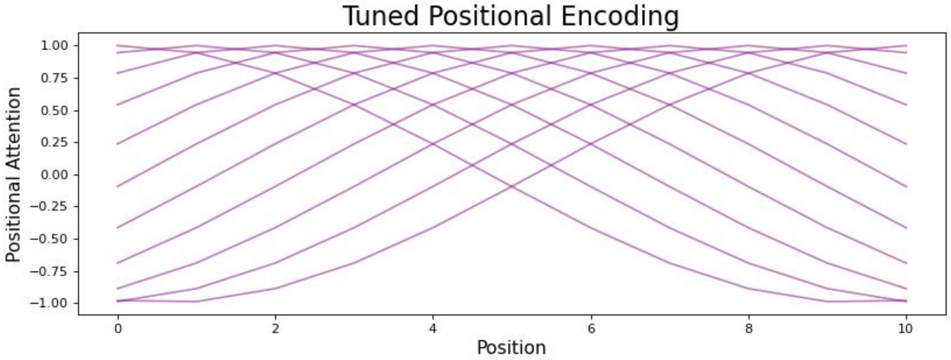


Positional Encoding

Original Positional Encoding: $PE = [sin(x_{pos}),\ cos(x_{pos}),\ sin(y_{pos}),\ cos(y_{pos}),\ sin(z_{pos}),\ cos(z_{pos})]$



Tuned Positional Encoding: $PE = [sin(x_{pos}/3),\ cos(x_{pos}/3),\ sin(y_{pos}/3),\ cos(y_{pos}/3),\ sin(z_{pos}/3),\ cos(z_{pos}/3)]$



Observations, actions and rewards

Observations:

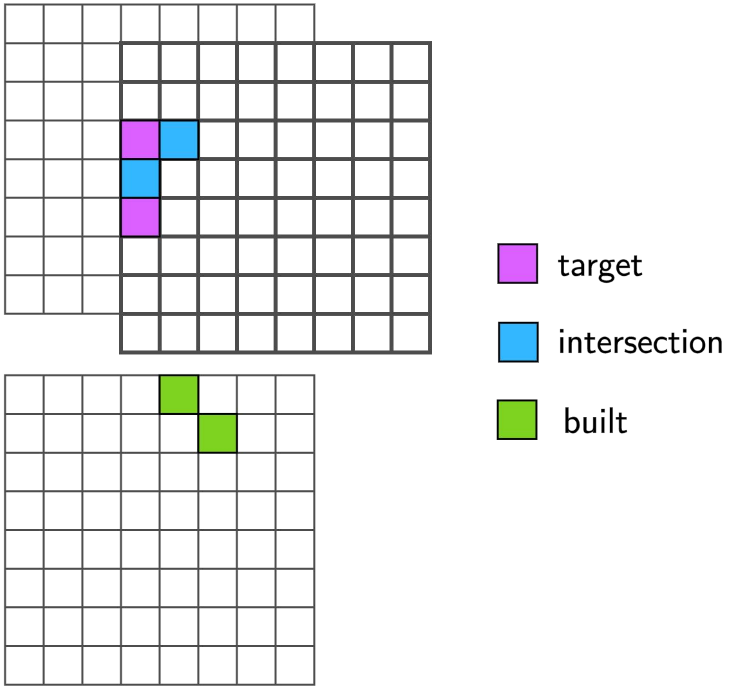
- FOV
- chat

Action space:

- human level

Reward shaping:

- increasing/decreasing the intersection size (between built and target) with 2/-2
- removing/placing a block without a change of the intersection size with 0.1/-0.1
- for any action which is not removing or not placing block -0.01

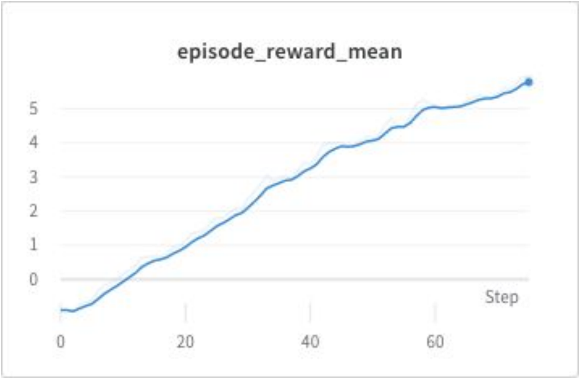


Policy Training

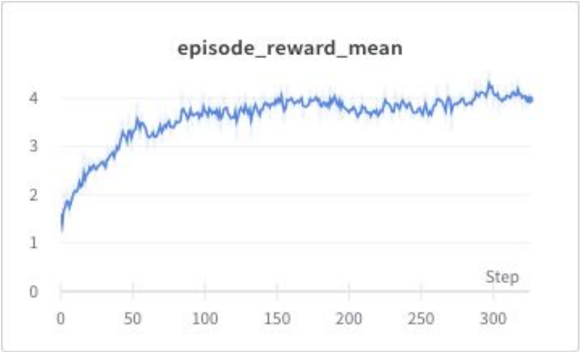
Device:
- Intel core i7 9700 (8 cores)
- Nvidia GeForce RTX 2060

Stage 1:
- Training on tasks C3, C17 and C32 to test the neural network ability of training
- 3 workers
- 150 k environment steps
- 4 hours of training

Stage 2:
- Training on augmented 154 tasks (2850 in result)
- 3 workers
- 3.5 M environment steps
- 50 hours of training



Stage 1: Training on C3, C17, C32



Stage 2: Training on augmented dataset

Policy Training

Stage 3:

- Zeroed rewards for removing/placing a block without a change of the intersection size
- In the result agent was able to build 2.5 blocks correctly on average

Stage 3: Training on augmented dataset with reshaped rewards

Random Agent

Scores:

- One of algorithm: 0.2872
- Random agent: 0.2562
- Our algorithm and random agent: 0.3218

Final Leaderboard

1. Hybrid Intelligence - 0.365 (Putra Manggala, Kata Naszadi, Michiel van der Meer, Taewoon Kim)
2. NeuroAI - 0.34 (Linar Abdrazakov, Igor Churin)

In total, there were 96 submissions and 37 registered participants.

https://www.iglu-contest.net/neurips-2021-winners

Interactive Grounded Language Understanding in a Collaborative Environment: IGLU 2021

Julia Kiseleva[1] Ziming Li[3] Mohammad Aliannejadi[2] Shrestha Mohanty[1] Maartje ter Hoeve[2] Mikhail Burtsev[4,5] Alexey Skrynnik[5] Artem Zholus[4] Aleksandr Panov[4] Kavya Srinet[6] Arthur Szlam[6] Yuxuan Sun[6] Marc-Alexandre Côté[1] Katja Hofmann[1] Ahmed Awadallah[1]

Linar Abdrazakov[7] Igor Churin[8] Putra Manggala[2] Kata Naszadi[2] Michiel van der Meer[10] Taewoon Kim[11]

`julia.kiseleva@microsoft.com`

[1]*Microsoft Research*
[2]*University of Amsterdam*
[3]*Alexa AI*
[4]*MIPT*
[5]*AIRI*
[6]*Facebook AI*
[7]*Moscow Institute of Physics and Technology*
[8]*Kalashnikov Izhevsk State Technical University*
[10]*Universiteit Leiden*
[11]*Vrije Universiteit Amsterdam*

Editors: Douwe Kiela, Marco Ciccone, Barbara Caputo

NeurIPS 2021. URL: https://arxiv.org/pdf/2205.02388.pdf

Some Statistics

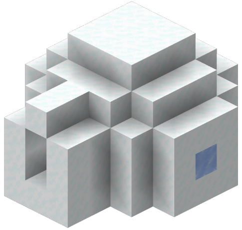
- 250 RL experiments
- 2000 hours of training neural networks
- used up to 5 servers in parallel

Github project

Thank you for your attention!

Linar Abdrazakov: linar200015@gmail.com
Igor Churin: igor.churin19@gmail.com

NEURAL INFORMATION PROCESSING SYSTEMS



MIPT