

Supplementary Materials

Visual Perspective Taking for Opponent Behavior Modeling

Boyuan Chen*, Yuhang Hu, Robert Kwiatkowski, Shuran Song, Hod Lipson

The PDF file includes:

- Robot hardware parameters.
- Details of the neural network architectures

Supplementary Materials

Robot Hardware Parameters

Our real world robotic platform is a 3D printed two wheeled robot. We show the parameters of the body of the robot excluding the camera in Fig. S1. The camera is an Intel Realsense Depth camera (D435) attached to the outside body of the robot (Fig. S2). We control our robot with a Raspberry Pi 3B+ module with Wi-Fi communication to our remote server which performs the main inference while the on-board controller is mainly responsible for observation processing and communication with the low-level control commands on the actuators. We use two LOBOT LX-16A motors. Both the hider robot and the seeker robot share the exact same hardware parameters.

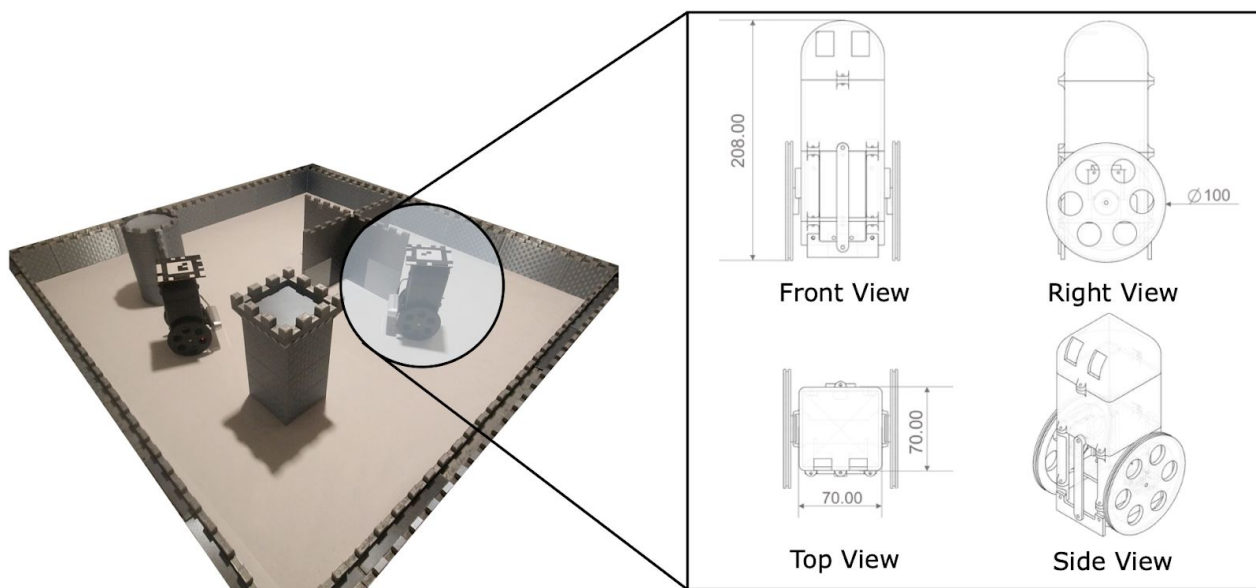


Fig. S1. CAD model of our physical robot. We show the arena of the environment with obstacles on the left, and different views of the CAD model to present the parameters of the robot.



Fig. S2. Physical Robot Demonstration. Our physical robots are two wheeled robots with a first-person RGB-D camera.

VPT-TOB Network Architecture

Layer	Kernel Size	Num Outputs	Stride	Padding	Dilation	Activation
Conv1	4 x 4	32	2	1	1	ReLU
Conv2	4 x 4	32	2	1	1	ReLU
Conv3	4 x 4	64	2	1	1	ReLU
Conv4	4 x 4	128	2	1	1	ReLU
Deconv4	4 x 4	64	2	1	1	Sigmoid
Deconv3	4 x 4	32	2	1	1	ReLU
Deconv2	4 x 4	16	2	1	1	ReLU
Deconv1	4 x 4	3	2	1	1	ReLU
Pred3Conv	3 x 3	3	1	1	1	N/A
Pred2Conv	3 x 3	3	1	1	1	N/A
Pred1Conv	3 x 3	3	1	1	1	N/A
Pred3Deconv	4 x 4	3	2	1	1	Sigmoid
Pred2Deconv	4 x 4	3	2	1	1	Sigmoid
Pred1Deconv	4 x 4	3	2	1	1	Sigmoid

Table S3. VPT-TOB Network Architectures. The upper portion of the table shows the detailed parameters of the Encoder network, while the lower portion of the table shows the detailed parameters of the Predictor network.

Value Prediction Network Architecture

Layer	Kernel Size	Num Outputs	Stride	Padding	Dilation	Activation
Conv1	5 x 5	16	1	2	1	ReLU
MaxPool1	2 x 2	16	2	0	1	N/A
Conv2	5 x 5	32	1	2	1	ReLU
MaxPool2	2 x 2	16	2	0	1	N/A
Conv3	5 x 5	32	1	2	1	ReLU
MaxPool2	2 x 2	16	2	0	1	N/A
FC1	N/A	1024	N/A	N/A	N/A	N/A
Dropout1 (p=0.5)	N/A	1024	N/A	N/A	N/A	N/A
FC2	N/A	2	N/A	N/A	N/A	N/A

Table S4. Value Prediction Network Architecture. Our value prediction network has three convolutional layers to process the future envisioned view image of the seeker robot. It then uses two fully connected layers to produce the probability of whether the hider robot will be safe at the corresponding future timestamp.