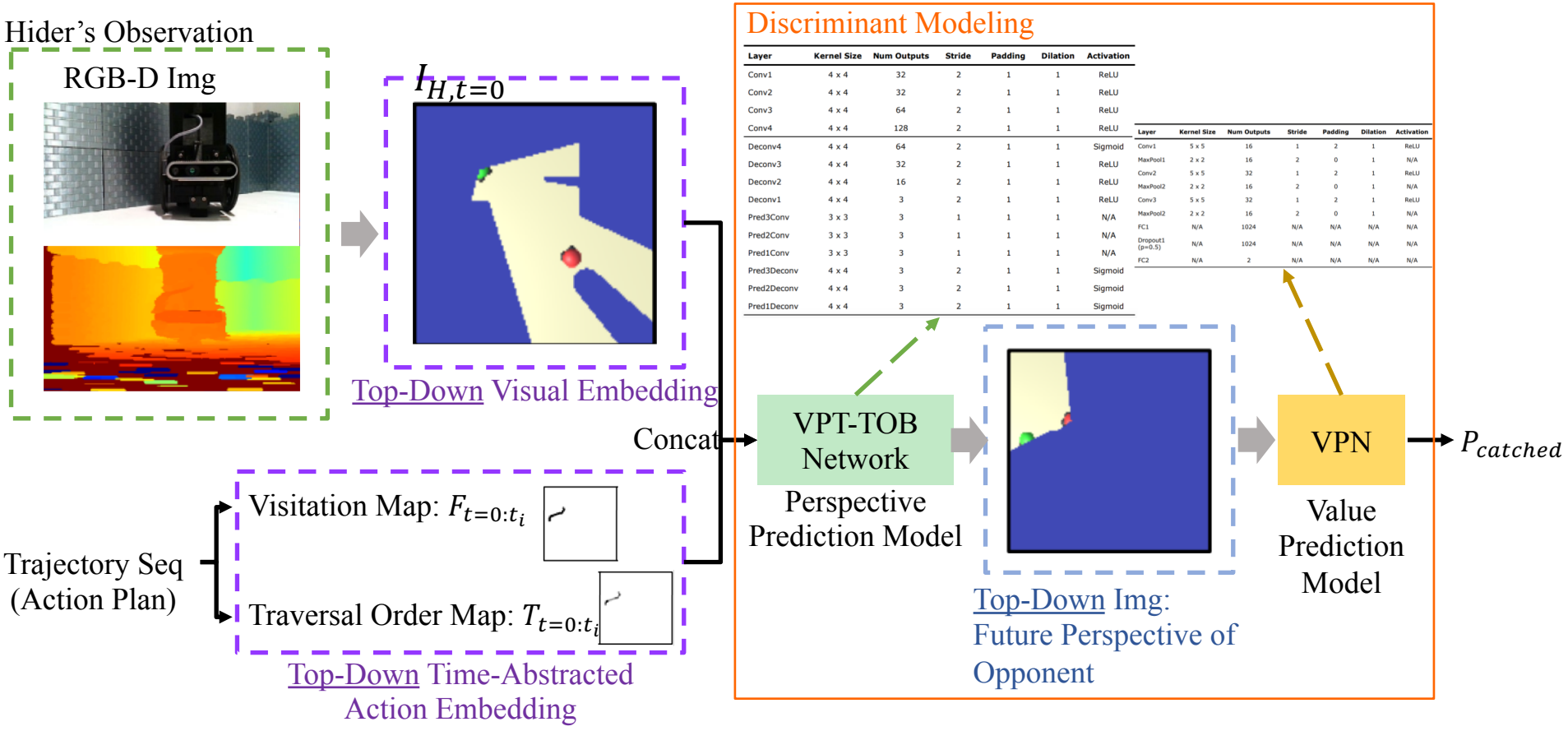
VPT-TOB Learning Poster

Tags: Robot Behavior Modeling, Multi-Agent Learning (MAL)

(Reference: Chen, Boyuan, et al. "Visual perspective taking for opponent behavior modeling." 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021.)

VPT-TOB Framework
(for hider learning)



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

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

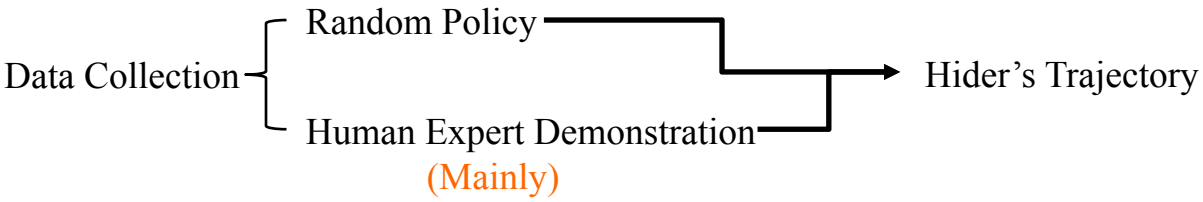
Input

Discriminant Modeling:

$$P(catch|state, action) = f_v(f_p(I_{H,t=0}, F_{t=0:t_i}, T_{t=0:t_i}))$$

Output

Benchmark Dataset



Training

VPT-TOB Network:

1. Loss (pixel-wise MSE loss):

$$\mathcal{L}_{\text{VPT-TOB}} = \text{MSE}\left(f_p\left(I_{H,t=0}, F_{t=0:t_i}, T_{t=0:t_i}\right), I_{S,t=t_i}\right)$$

2. Learning:
   a. Gradually decreasing learning rate.
   b. Mini-batch size: 256
   c. Adam optimizer

VPN:

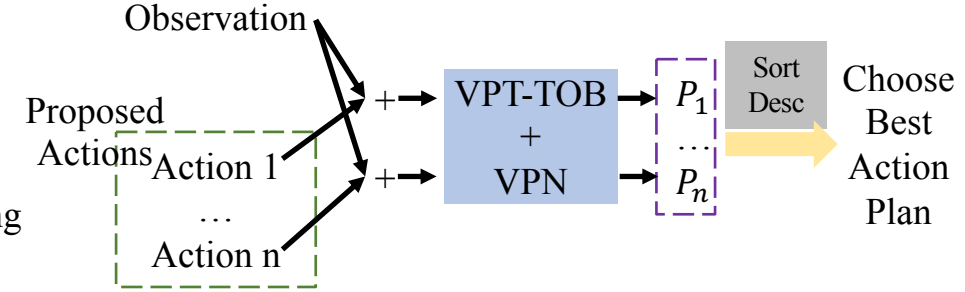
1. Data Augmentation: Randomly rotating the images among 90◦, 180◦ and 270◦.
2. Loss (cross-entropy loss):

$$\mathcal{L}_{\text{VPN}} = -(v\log(f_v(I_{S,t=t_i})) + (1-v)\log(1-f_v(I_{S,t=t_i})))$$

3. Learning:
   a. Gradually decreasing learning rate.
   b. Mini-batch size: 256
   c. Adam optimizer

Inference and Planning

1. Action Plans Proposal:
   Propose a set of action plans with a low-level controller policy (e.g., A*).
2. Determine:
   Observation



Seeker's Policy

A heuristic expert policy:
a. If the hider is visible, the seeker will navigate towards it with A*.
b. If the hider is not visible, the seeker navigate to the last known position of the hider and then continuously explore.

Summarization and Personal Thinking

Summarization:

1. Innovation:
   a. Proposal and Determine Paradigm:
      a.1 Action plans proposal with a low-level controller policy (time efficient).
      a.2 Construct observation-action pairs, modeling as a classification and ranking problem using discriminant VPT-TOB and VPN.
   b. Top-down visual perspective taking for opponent behavior modeling in MAL.
2. To Improve:
   Mainly focus on VPT but don't consider explicitly about TOB and behavior modeling.

Future Work and Personal thinking:

1. Explore the generalization to more robots and study swarm robot (mentioned in paper).
2. Consider scenarios where the seeker policy might change, and thus stochastic optimization with few-shot learning algorithms should be taken into consideration (mentioned in paper).
3. Explicitly consider about behavior modeling and TOB (not purely base on VPT).
4. Online expert intervention and correction when training to make up for the insufficiency of expert demonstration dataset.