TEXAS
The University of Texas at Austin

Efficiently Guiding Imitation Learning Agents with Human Gaze

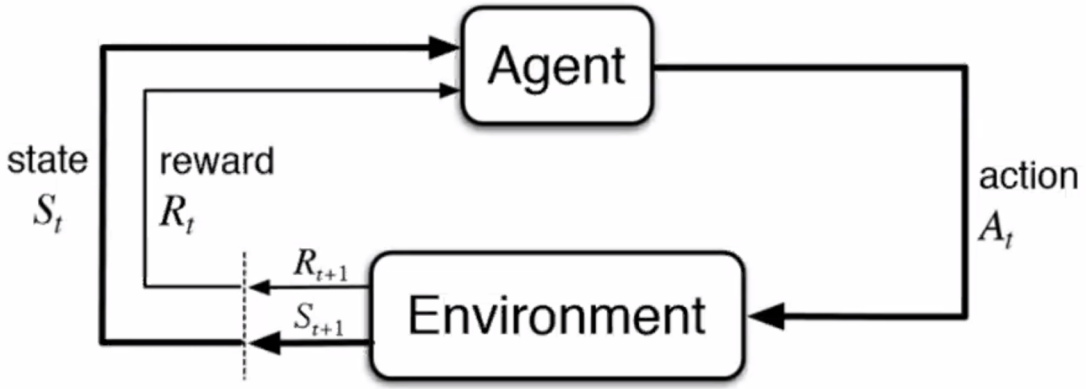
Akanksha Saran, Ruohan Zhang, Elaine Short, Scott Niekum

AAMAS 2021




Reinforcement Learning with predefined Reward Functions

$$\langle S, A, T, R, \gamma \rangle$$

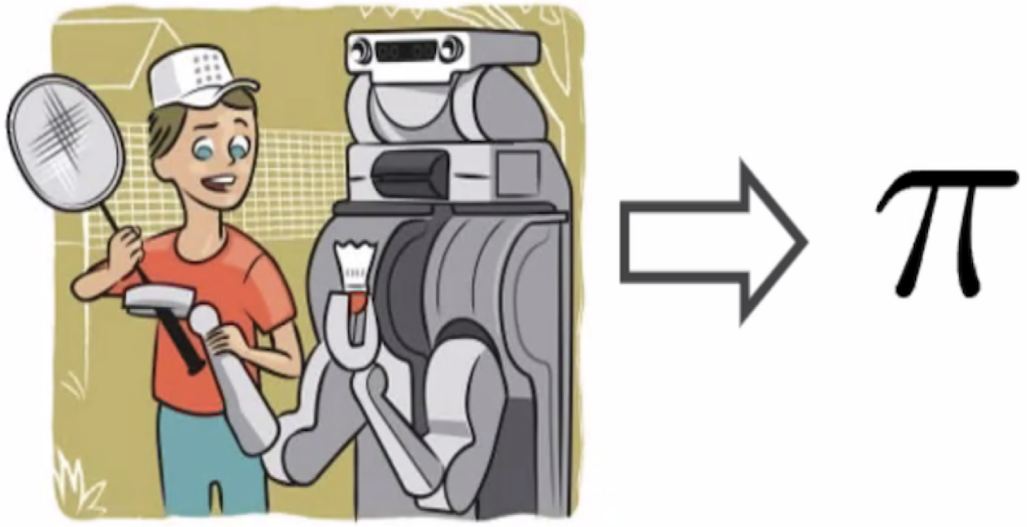


$$\pi : \mathbb{S} \rightarrow A$$

$$\pi^* = \arg\max_{\pi} J(\pi)$$

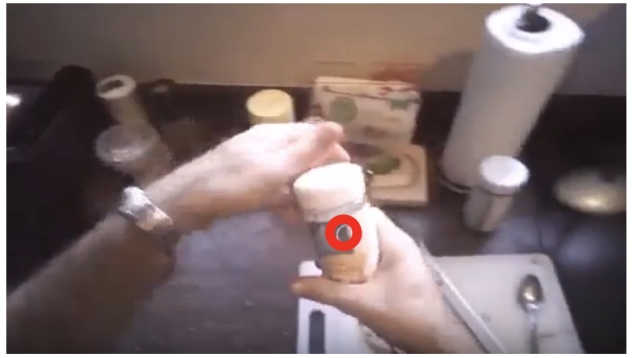
$$J(\pi) = \mathbb{E}_{\pi}\left[\sum_{t=1}^{\infty} \gamma^t r_t \middle| \pi\right]$$

Imitation Learning does **not** require predefined Reward Functions



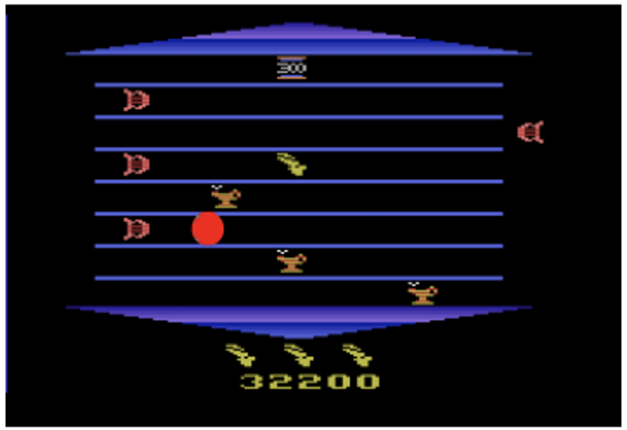
π

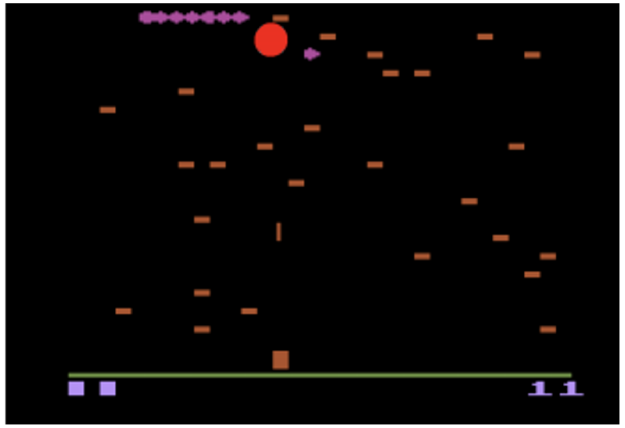
Gaze is a rich signal of Human Intent



Argyle, M. Non-verbal communication in human social interaction. 1972.
Hayhoe, M & Ballard, D. Eye movements in natural behavior. Trends in cognitive sciences, 9(4), 2005.

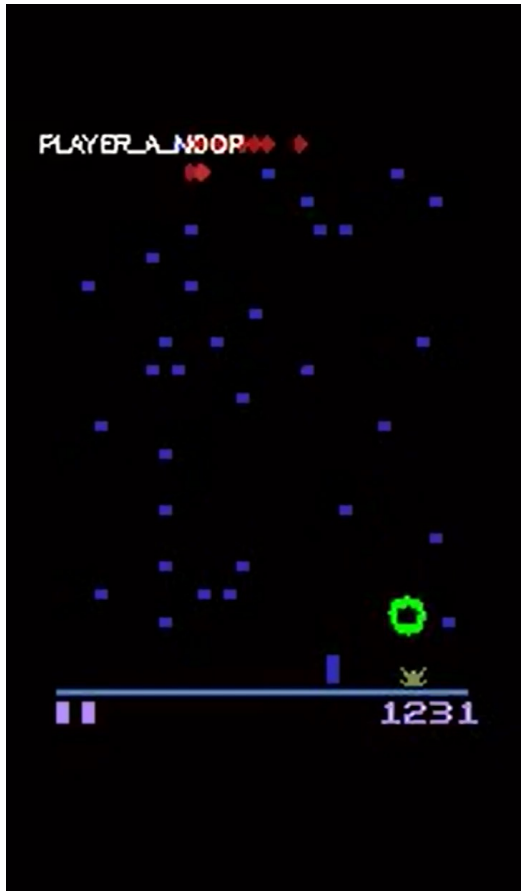
Domain for this work: Atari Game Playing Agents





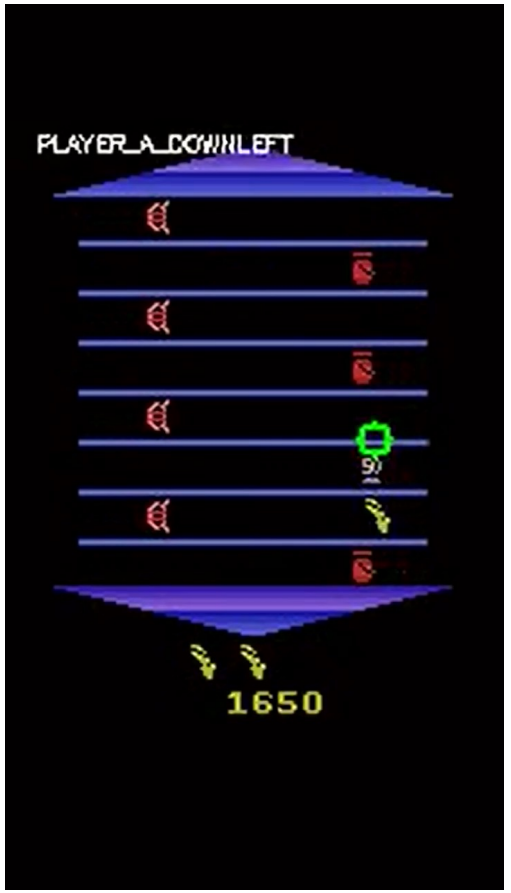
Gaze is a rich signal of Human Intent

Centipede



Gaze indicates where the human might shoot next

Asterix

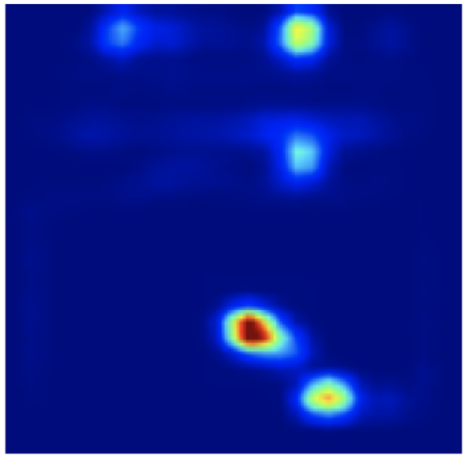


Gaze on hamburgers that should be eaten and dynamites which should be avoided

Can we draw inspiration from high-scoring RL agents to close the gap between the performances of IL and RL agents?

Attention of RL Agents

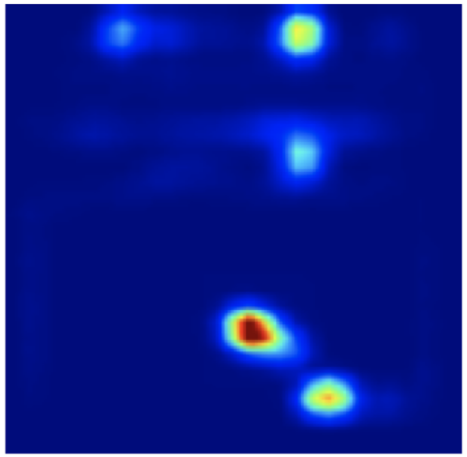
(a) Game State



(b) RL Attention

Attention of RL Agents

(a) Game State



(b) RL Attention

Perturbation based method to compute RL attention

$$S_\pi(i,j) = \frac{1}{2}||\pi(I) - \pi(\phi(I,i,j))||^2$$

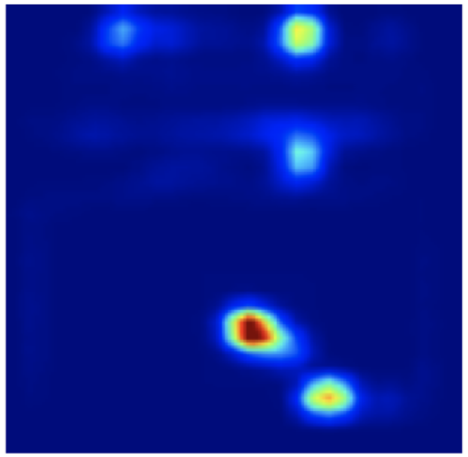
Change in policy by perturbing the image at a pixel

Greydanus, S., Koul, A., Dodge, J., & Fern, A. Visualizing and understanding Atari agents. ICML, 2018.

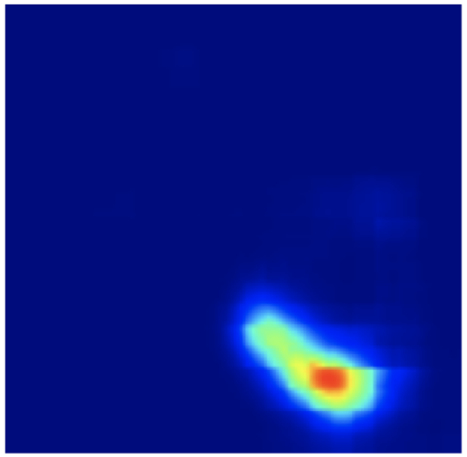
RL agent Attention “covers” regions attended by Human Gaze

… while also attending to other regions

(a) Game State



(b) RL Attention



(c) Human Gaze

Covert and Overt Attention

- Selective Attention
  Overt Attention: directing sensory organs towards specific stimuli
  Covert Attention: using information from working memory

- Being attended by the human gaze model is a sufficient, but not necessary condition for the features to be important.

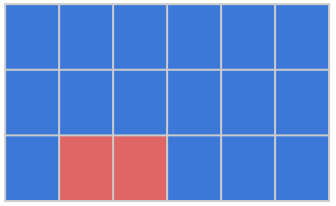
- High-performing RL agents pay attention to at least the features attended by humans.

- Quantifying comparisons with RL agent attention: use a metric that is sensitive to false negatives if we treat human attention as the ground truth.

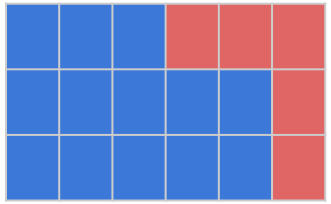
Petrosino, G., D. Parisi, and S. Nolfi. Selective attention enables action selection: evidence from evolutionary robotics experiments. Adaptive Behavior, 2013.

Coverage Metric

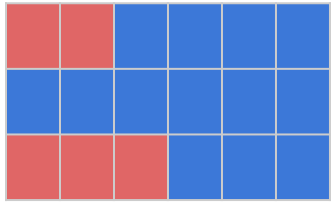
$$KL(P||Q) = \sum_i \sum_j P(i,j) \log \left(\frac{P(i,j) + \epsilon}{Q(i,j) + \epsilon} \right)$$



P: Human Gaze map



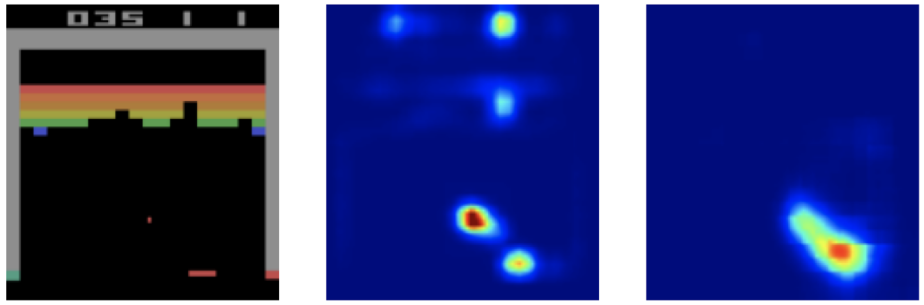
Q: RL Attention Map (**No Coverage**)
KL(P || Q) = **8.5**



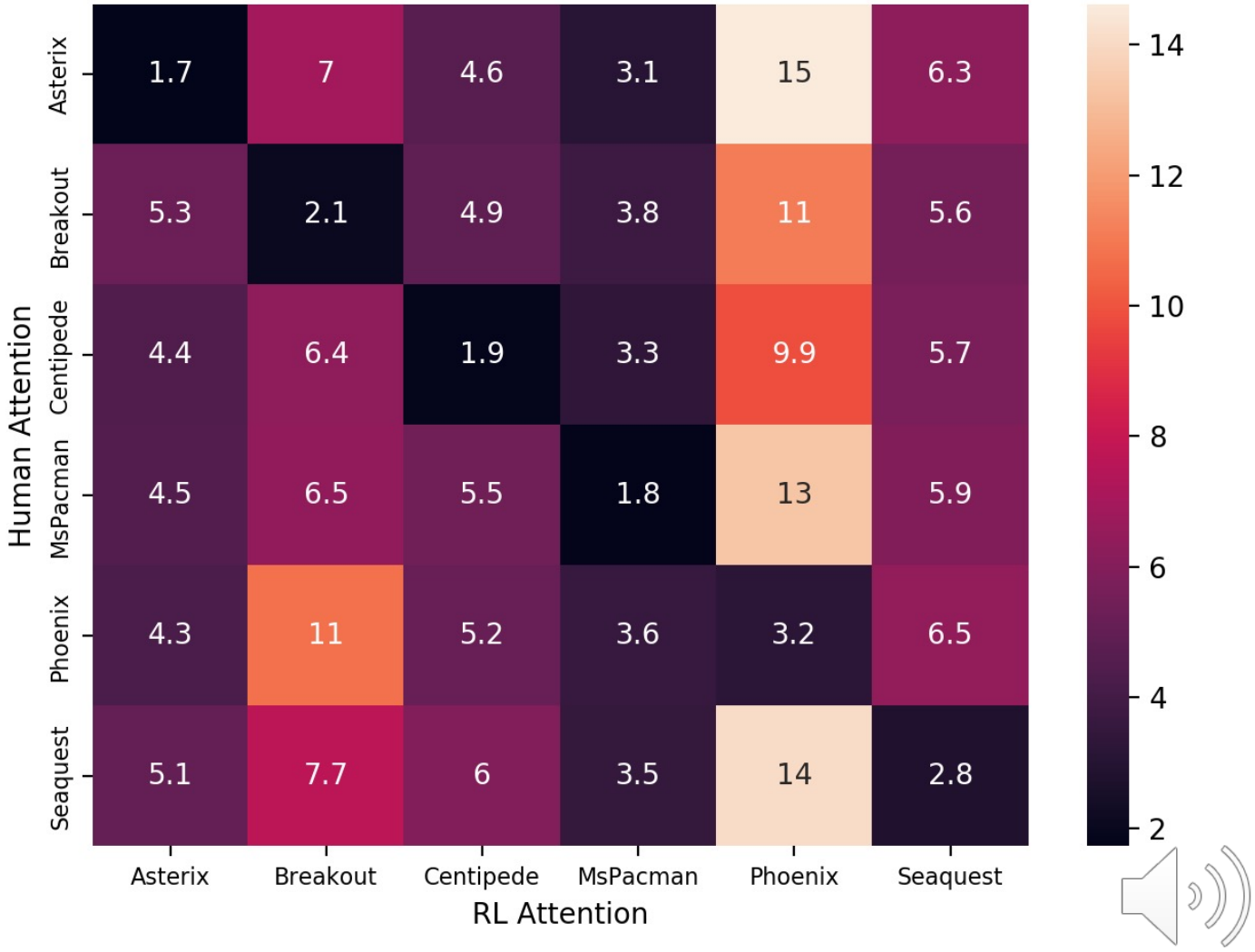
Q: RL Attention Map (**Has Coverage**)
KL(P || Q) = **0.9**

Bylinskii, Z., T. Judd, A. Oliva, A. Torralba, and F. Durand. “What do different evaluation metrics tell us about saliency models?”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

Comparison of Human Attention and RL agent Attention



(a) Game State (b) RL Attention (c) Human Gaze

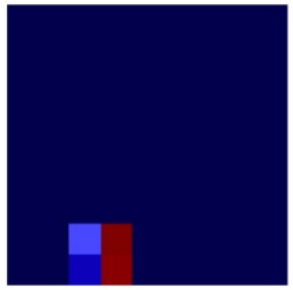


CGL – Our Approach to Leverage Human Gaze

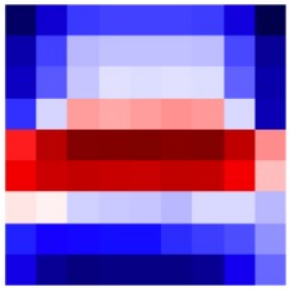
- We propose an auxiliary **coverage-based gaze loss** (CGL) based on KL divergence between human attention and agent's activations

- Guide the attention of existing Imitation Learning methods towards features that humans consider important for decision making

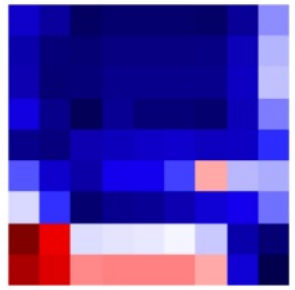
(a) Input image stack



(b) Gaze heatmap



(c) Network activation without gaze loss



(d) Network activation with gaze loss

Prior Approaches Leveraging Human Gaze for Imitation Learning

- We consider two baseline methods which also leverage gaze for Imitation Learning: (1) AGIL and (2) GMD
- AGIL increases the learnable parameters of models by using gaze heatmaps as input to networks
- AGIL and GMD require gaze information at both train and test time

Zhang et al. , “AGIL: Learning attention from human for visuomotor tasks, in Proceedings of the European Conference on Computer Vision (ECCV), 2018
Yuying Chen, Congcong Liu, Lei Tai, Ming Liu, and Bertram E Shi. 2019. Gaze Training by Modulated Dropout Improves Imitation Learning. arXiv preprintarXiv:1904.08377(2019).

Advantages of CGL over prior Gaze utilization methods

- Augment any existing Imitation Learning algorithm to utilize human gaze data
- No additional learnable parameters added to existing models
- Gaze data only required at train time along with demonstrations
- No need for human gaze prediction models for test time gaze

Experimental Methodology

- Three Imitation Learning methods:
  - Behavioral Cloning (BC)
  - Behavioral Cloning from Observation (BCO)
  - Trajectory-ranked Reward Extrapolation (T-REX)

- Three Baseline methods:
  - Attention Guided Imitation Learning (AGIL)
  - Gaze modulated Dropout (GMD)
  - Motion instead of Gaze in the CGL formulation

- Measure average performance improvement for 20 Atari Games

Atari-HEAD: Atari demonstration Dataset with Gaze

- Human gaze and demonstration data for 20 Atari Games
- EyeLink 1000 eye tracker at 1000Hz
- Total data of 117.07 hours collected with 4 users

Zhang, R., C. Walshe, Z. Liu, L. Guan, K. S. Muller, J. A. Whritner, L. Zhang, M. M. Hayhoe, and D. H. Ballard. “Atari-head: Atari human eye-tracking and demonstration dataset.” AAAI, 2020.

Algorithm #1: Behavioral Cloning (BC)

Zhang et al. , “AGIL: Learning attention from human for visuomotor tasks, in Proceedings of the European Conference on Computer Vision (ECCV), 2018

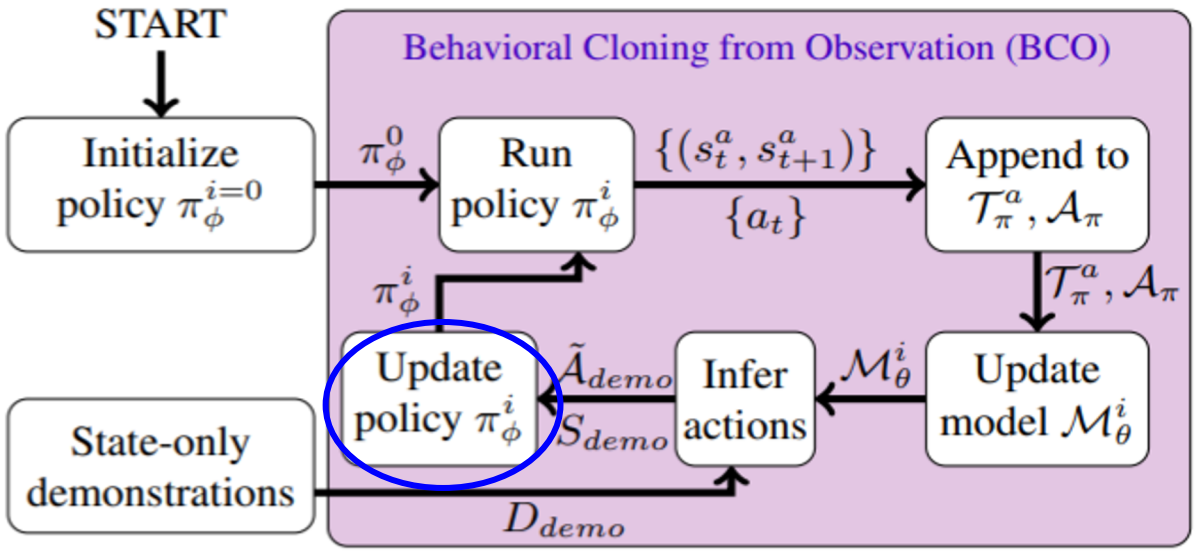
Algorithm #1: Behavioral Cloning (BC)

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \left[-\log \pi_\theta(a_i|s_i) + \alpha\, CGL(g(s_i), c_3(s_i)) \right]$$

Cross-entropy loss for action prediction

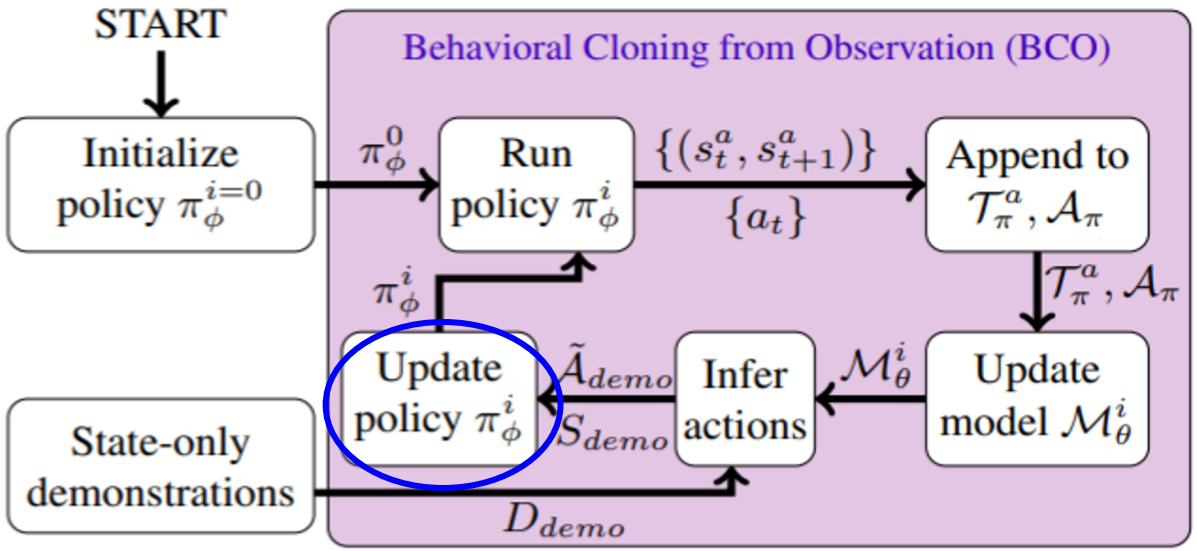
Coverage Loss

Algorithm #2: Behavioral Cloning from Observation (BCO)



Torabi, F., G. Warnell, and P. Stone. “Behavioral cloning from observation.” IJCAI, 2018.

Algorithm #2: Behavioral Cloning from Observation (BCO)

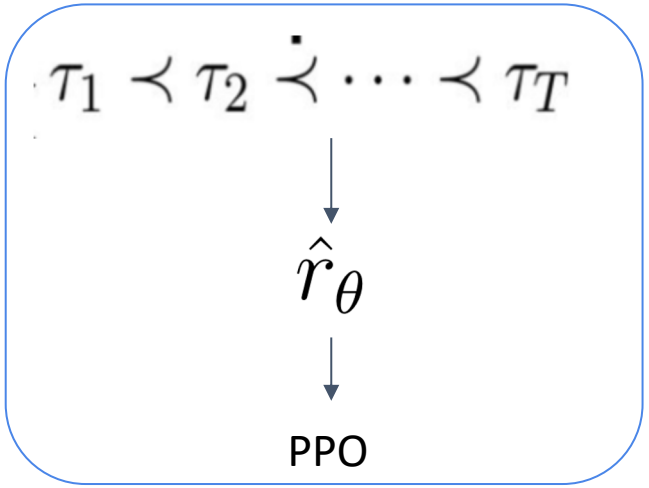


$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \left[-\log \pi_\theta(\tilde{a}_i | s_i) + \alpha \, CGL(g(s_i), c_2(s_i)) \right]$$

Cross-entropy loss for action prediction

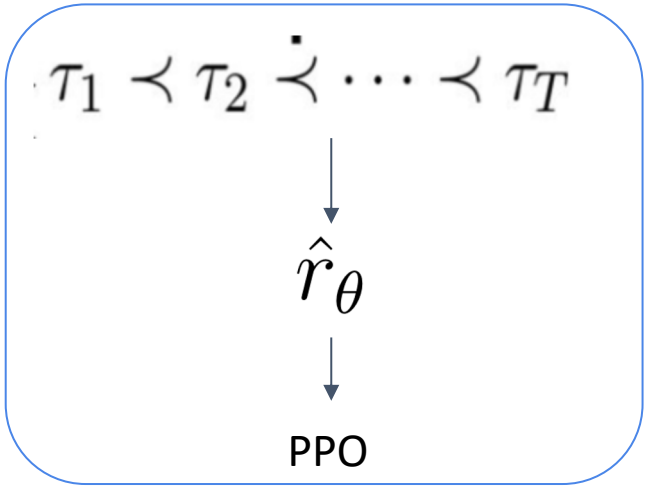
Coverage Loss

Algorithm #3: Trajectory-ranked Reward Extrapolation (T-REX)



Brown, D. S., W. Goo, P. Nagarajan, and S. Niekum. “Extrapolating beyond suboptimal demonstrations via inverse reinforcement learning from observations.”, ICML, 2019.

Algorithm #3: Trajectory-ranked Reward Extrapolation (T-REX)



$$\mathcal{L}(\theta) = -\sum_{\tau_i \prec \tau_j} \log \frac{\exp \sum_{s \in \tau_j} \hat{r}_\theta(s)}{\exp \sum_{s \in \tau_i} \hat{r}_\theta(s) + \exp \sum_{s \in \tau_j} \hat{r}_\theta(s)} + \alpha \left| \sum_{s \in \tau_i} CGL(\tau_i^g(s), c_1(s)) + \sum_{s \in \tau_j} CGL(\tau_j^g(s), c_1(s)) \right|$$

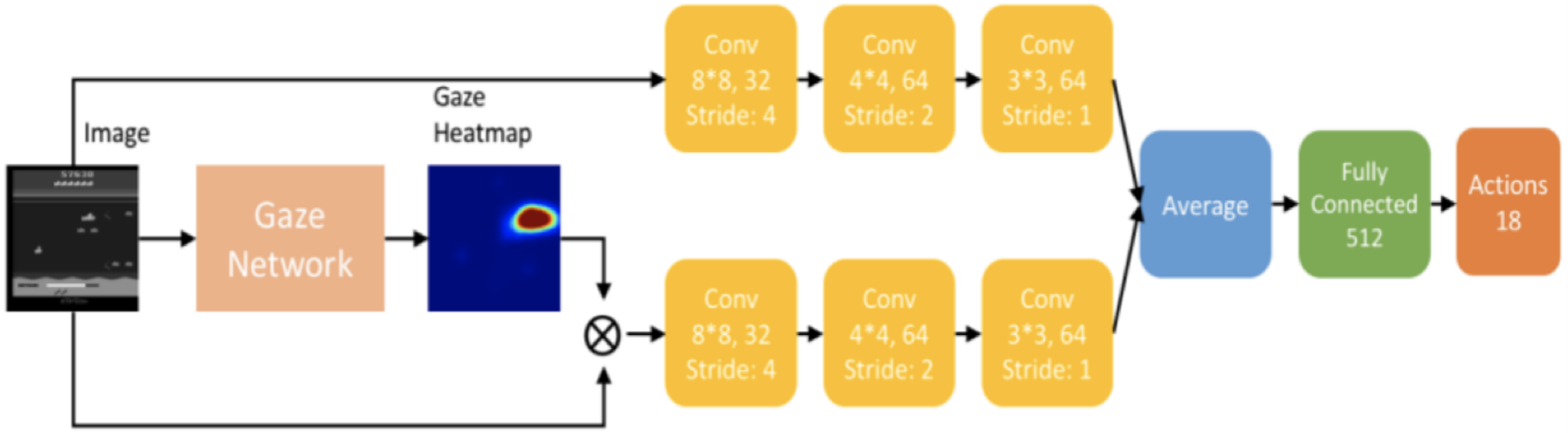
Pairwise Trajectory Ranking Loss

Coverage Loss

CGL improves performance for 3 Imitation Learning Agents

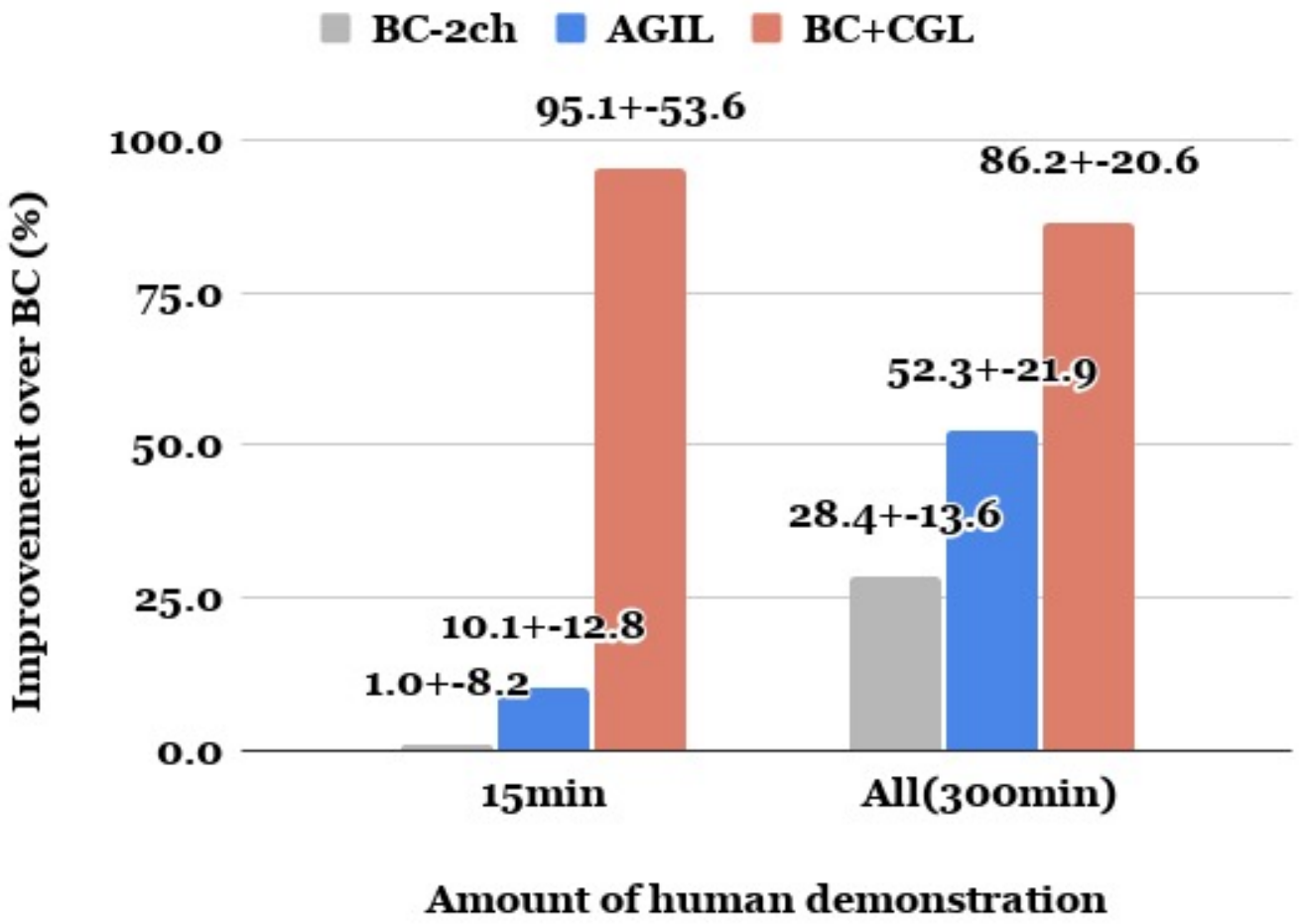
IL Algorithm	% Improvement with CGL
BC	95%
BCO	343%
TREX	390%

CGL outperforms AGIL

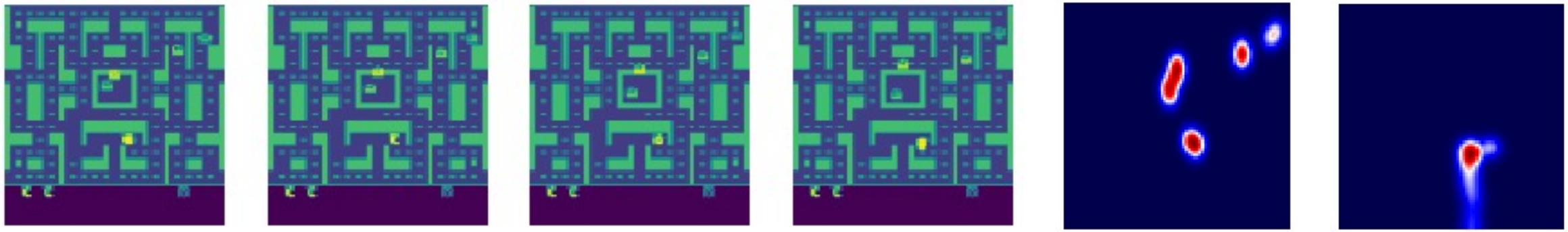


Zhang et al. , “AGIL: Learning attention from human for visuomotor tasks, in Proceedings of the European Conference on Computer Vision (ECCV), 2018

CGL outperforms AGIL



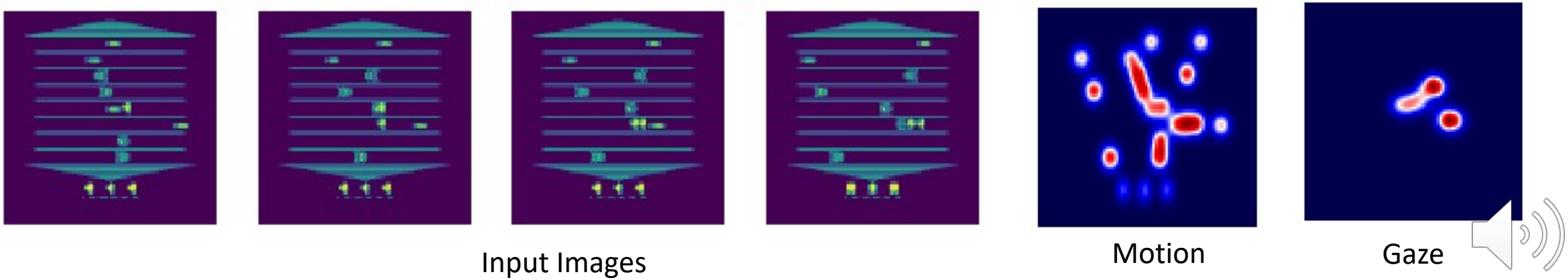
Motion information instead of Gaze as a baseline for CGL



Input Images

Motion

Gaze

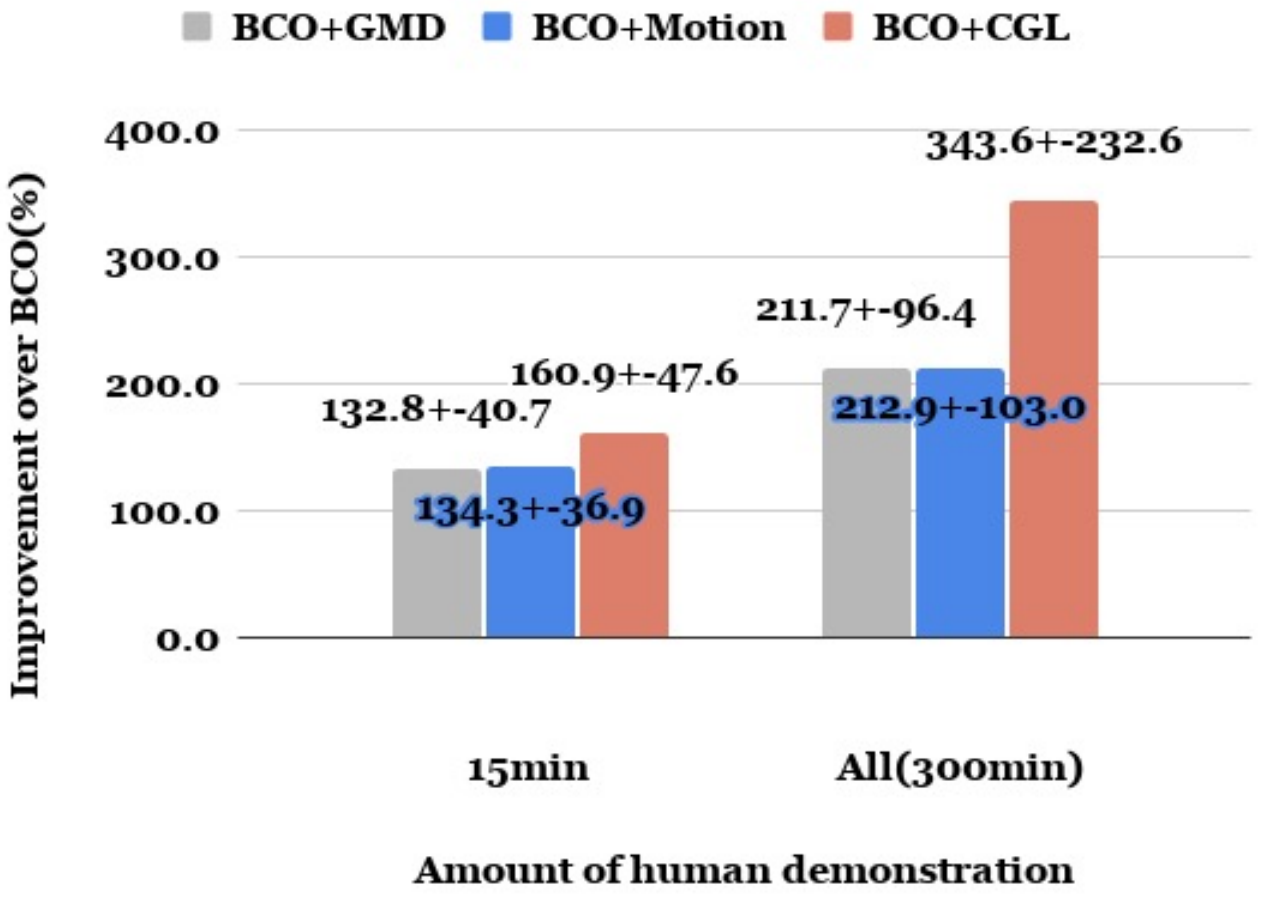


Input Images

Motion

Gaze

CGL outperforms Motion Baseline and Gaze Modulated Dropout (GMD) Baseline



Understanding the performance gains of CGL

CGL reduces causal confusion compared to baseline BC algorithm

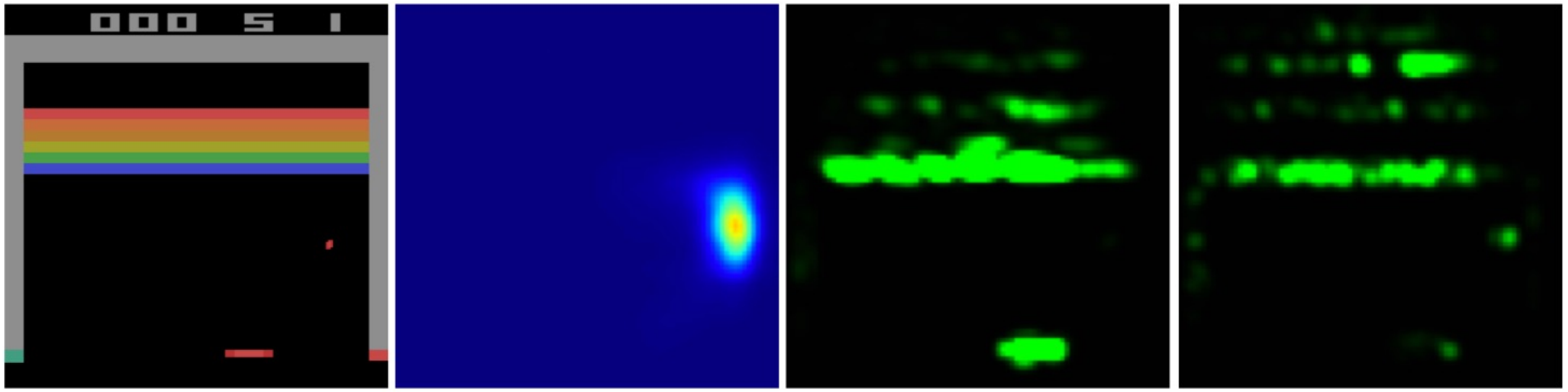
Confounded images with correlated past actions as part of the state space

(a) Breakout (b) Asterix (c) Demon Attack (d) Freeway

CGL suffers less with confounded data and hence reduces causal confusion compared to BC

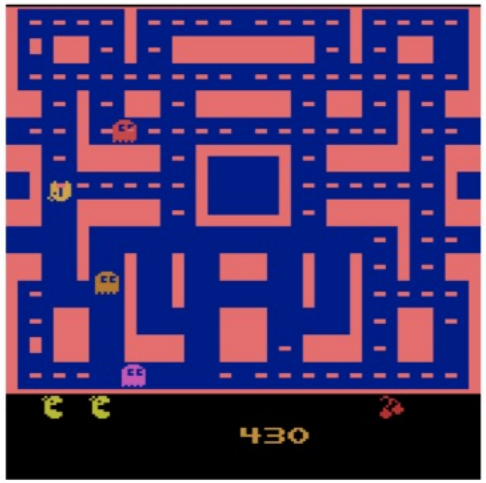
Algorithm tested with confounded images	Performance reduction with confounded images (lower is better)
BC confounded	-47.8 %
BC+ CGL confounded	**-34.0 %**

Visualizing Attention of CGL Agents

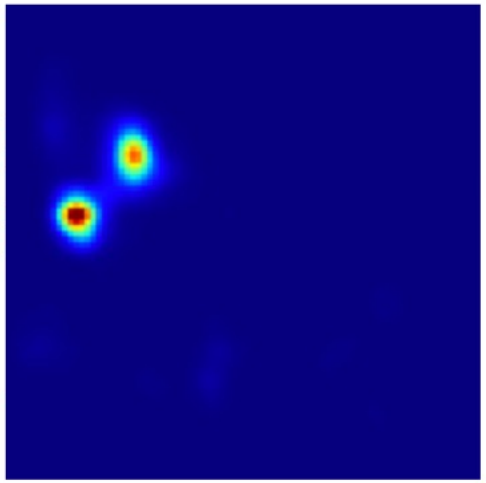


(a) Input image (b) Human (c) T-REX (d) T-REX+CGL

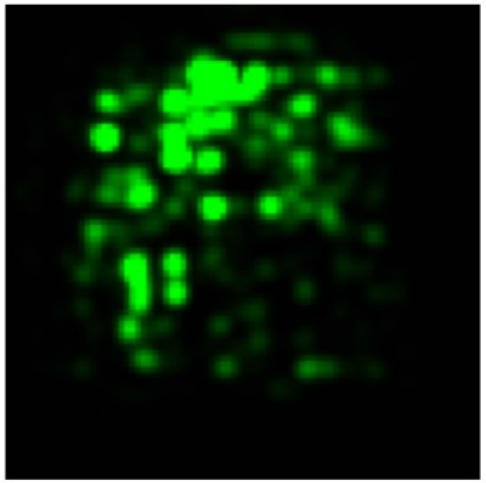
Failure modes of CGL



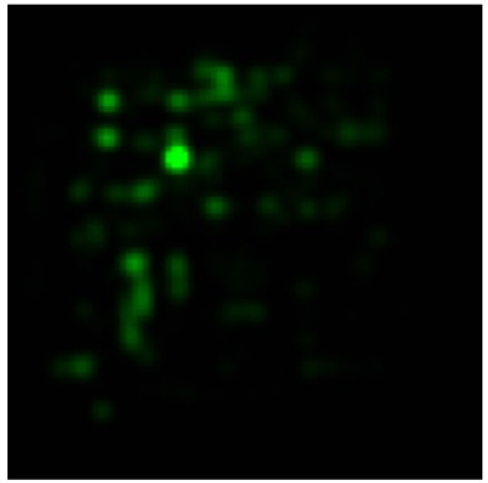
(e) Input image



(f) Human



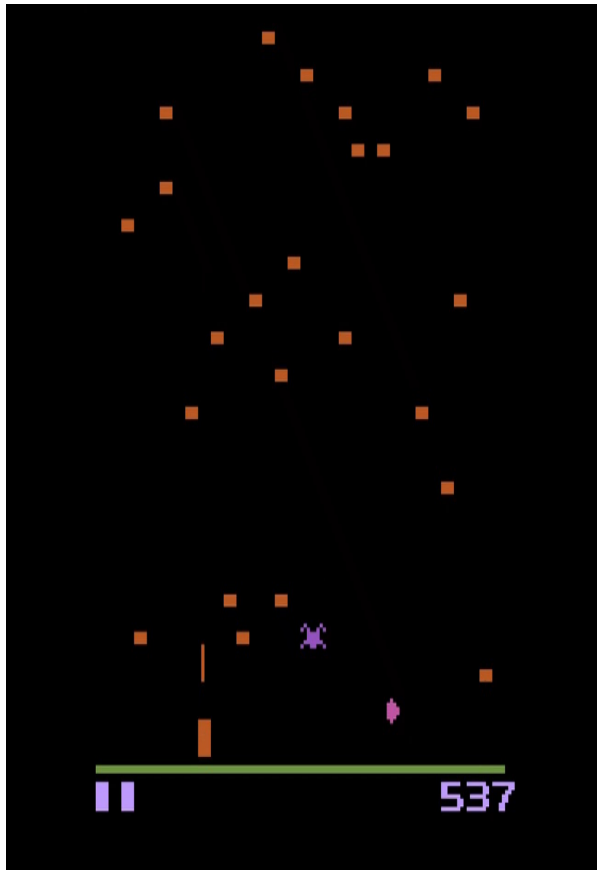
(g) T-REX



(h) T-REX+CGL

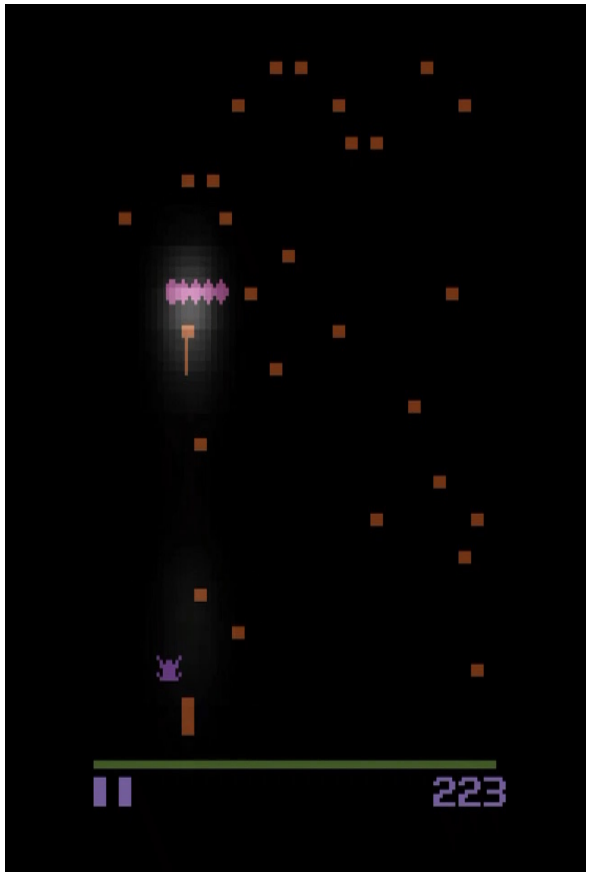
Visualizing CGL agent policies for BC

BC



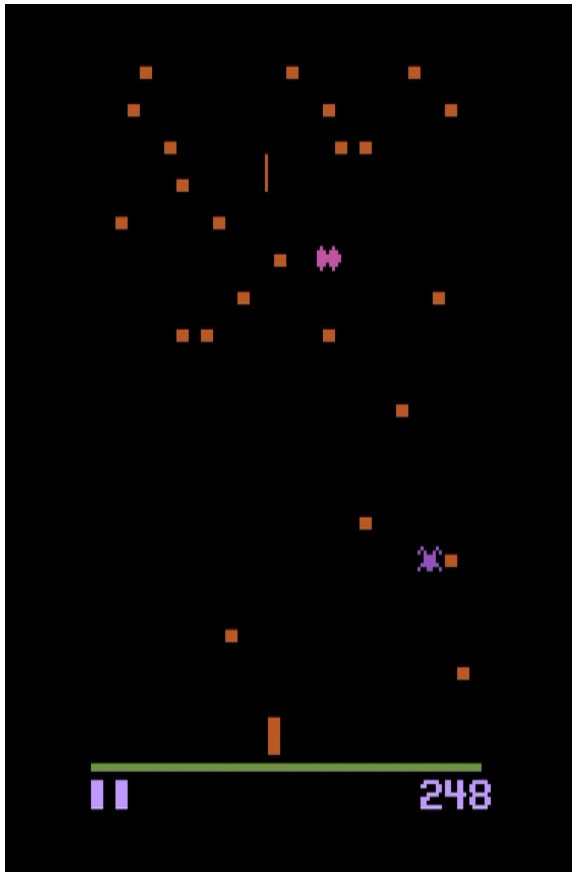
Does not learn to actively shoot the spider

AGIL



Shoots the spider when it comes directly above the agent

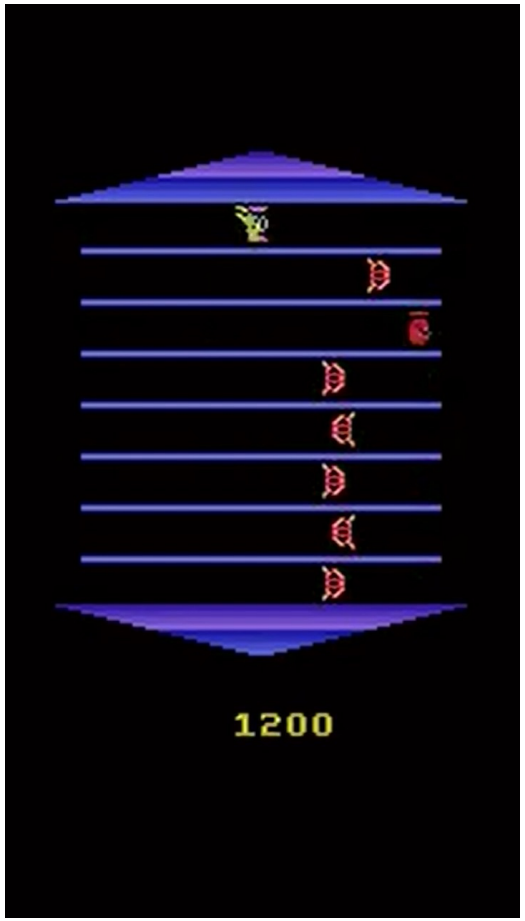
BC +CGL



Actively goes and shoots the spider

Visualizing CGL agent policies for T-REX

T-REX



Agent learns to move slowly between levels and does not eat many hamburgers

T-REX + CGL

Agent learns to move fast and advance to many levels by eating more hamburgers

Closing the gap between RL and IL

Game	Algorithm (#demo)	Score	DQN Score
alien	AGIL (300min)	**2104.7**	1620.0
asterix	T-REX+CGL (30min)	**66445.0**	4359.0
bank_heist	BC-2ch (300min)	174.3	**455.0**
berzerk	BCO+CGL (15min)	**687.67**	585.6
breakout	T-REX+CGL (300min)	**438.4**	385.5
centipede	T-REX+CGL (30min)	**20762.5**	4657.7
demon_attack	T-REX+CGL (300min)	**17589.0**	12149.4
enduro	BC+CGL (300min)	445.1	**729.0**
freeway	BC-2ch (300min)	**31.4**	30.8
frostbite	BC+CGL (30min)	**5897.7**	797.4
hero	BC+CGL (15min)	19023.2	**20437.8**
montezuma	BC+CGL (300min)	**1720.0**	0.0
ms_pacman	BC+CGL (30min)	2739.7	**3085.6**
name_this_game	AGIL (300min)	5817.0	**8207.8**
phoenix	AGIL (300min)	5140.0	**8485.2**
riverraid	T-REX+CGL (300min)	7370.0	**8316.0**
road_runner	BC+CGL (300min)	33510.0	**39544.0**
seaquest	T-REX+CGL (30min)	759.3	**5860.6**
space_invaders	T-REX+CGL (300min)	1563.7	**1692.3**
venture	BC+CGL (15min)	**376.7**	163.0

Summary

- Coverage Based Gaze Loss (CGL) - A novel approach to incorporate gaze information of the human demonstrator to any existing Imitation Learning Algorithm with convolutional layers
- Encourages a network to focus on parts of the state space where the humans fixated, while also being able to focus on other regions
- CGL improves performance of 3 Imitation Learning algorithms for Atari Game Playing
- CGL does not increase learnable parameters of the existing Imitation Learning approaches and does not require gaze data at test time.

Thank you!

Personal Autonomous Robotics Lab