Efficiently Guiding Imitation Learning Algorithms with Human Gaze

Akanksha Saran
Computer Science Department
Univeristy of Texas at Austin
asaran@cs.utexas.edu,

Ruohan Zhang
Computer Science Department
Univeristy of Texas at Austin
zharu@utexas.edu,

Elaine Schaertl Short
Computer Science Department
Tufts University
eshort@tufts.edu

Scott Niekum
Computer Science Department
Univeristy of Texas at Austin
sniekum@cs.utexas.edu

Abstract

Human gaze is known to be an intention-revealing signal in human demonstrations of tasks. In this work, we use gaze cues from human demonstrators to enhance the performance of state-of-the-art inverse reinforcement learning (IRL) and behavior cloning (BC) algorithms, without adding any additional learnable parameters to those models. We show how to augment any existing convolutional architecture with our auxiliary gaze loss (coverage-based gaze loss or CGL) that can guide learning toward a better reward function or policy. We compare our approach against two baseline methods which utilize gaze, highlighting trade-offs for each of them. Our auxiliary gaze loss function improves performance of both BC and IRL methods on a variety of Atari games. It outperforms a baseline approach for incorporating gaze with IL methods, called gaze-modulated dropout, and is comparable to another approach (AGIL) which uses gaze as input to the network.

Introduction

Learning agents can outperform humans at tasks such as Atari game playing when provided with well-defined goals or rewards using reinforcement learning (RL) (Sutton and Barto 2018). However, designing reward functions by hand can be difficult for complex tasks, even for experts. Imitation learning (IL) (Schaal 1997),(Argall et al. 2009) is an alternative methodology to infer an optimal policy from demonstrations. A challenge in training and utilizing IL agents in the real world is learning from few demonstrations to minimize the burden on end-users, while also sufficiently resolving ambiguity in user intentions and avoiding overfitting. Gaze, an additional informative modality from the demonstrator apart from state-action pairs, can help extract more information out of the same number of demonstrations (Zhang et al. 2018).

Human attention in the form of eye gaze has been known to encode top-down attention versus bottom-up salience when performing goal directed tasks (Land 2009; Hayhoe and Ballard 2005; Rothkopf, Ballard, and Hayhoe 2007; Tatler et al. 2011). Gaze has been shown to improve performance of imitation learning algorithms (Zhang et al. 2019),

---

particularly for autonomous driving (Chen et al. 2019; Xia et al. 2019) and Atari game playing (Zhang et al. 2018). However, most prior approaches utilizing gaze for IL algorithms either use gaze heat maps as input to the agent's networks in addition to the game state (Zhang et al. 2018; Liu et al. 2019), or predict gaze heatmaps along a network pathway in conjunction with learning the policy (Xia et al. 2019). By contrast, we propose using an auxiliary gaze loss during training of imitation learning algorithms to improve performance of existing methods without increasing model complexity, data requirements or requiring test-time gaze.

Our methodology utilizes a demonstrator's gaze fixations on the image as part of a surrogate loss function (coverage-based gaze loss or CGL) during the training phase. Encoding priors in loss functions for label-free supervision of neural networks has been suggested by Stewart el al. (Stewart and Ermon 2017). Similar to their approach, we show that a loss term guiding a network to attend to the demonstrator's gaze locations can help improve performance on different Atari Games when using IL. Such an auxiliary gaze loss can guide the learning of any agent using image based state representations and convolutional layers as part of it's model architecture. A critical advantage of our approach is that gaze is not required at test time and instead used as a weak supervisory signal, in contrast to several prior approaches utilizing gaze.

We evaluate our auxiliary gaze loss function on 11 unique Atari games and 3 different IL approaches. Our experiments show that CGL can improve performance for both IRL and BC frameworks: T-REX (Brown et al. 2019), BCO (Torabi, Warnell, and Stone 2018) and BC (Zhang et al. 2018) networks, compared to not using any gaze information at all. Moreover, we show that to improve performance, human gaze is more informative than information already encoded in the visual state space in the form of motion of the visual scene. Our auxiliary loss does not show performance gains when using motion instead of gaze. We also show improved results compared to another recent method for incorporating gaze called gaze-modulated dropout (GMD) (Chen et al. 2019). We compare against this baseline as it is the closest work in the literature in terms of not increasing the complexity of existing models. CGL outperforms GMD for 5 out of 6 games when tested with BCO. We also compare against a recent work utilizing gaze as an input to a behavior cloning algorithm (AGIL) (Zhang et al. 2018). We show that our

(a) Input image stack

(b) Gaze heatmap

(c) Network activation without gaze loss
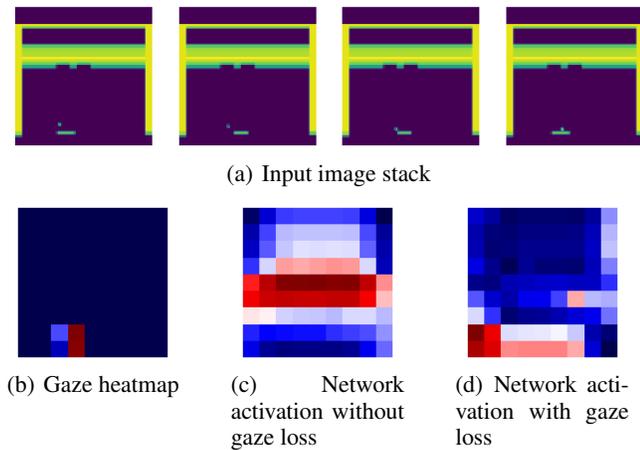
(d) Network activation with gaze loss

Figure 1: Our auxiliary gaze loss (CGL) guides a convolutional network to focus on parts of the state space which the human attends to. An example of network activations for the BCO network are shown in (c) without utilizing gaze and (d) when CGL is incorporated as part of training. These activation heatmaps represent the collapsed and normalized outputs from the second convolutional layer as described in Equation (2). We find that when CGL is utilized, network activations are more heavily focused (in red) on the bottom of the input image stack (a), where the human also paid attention during demonstrations (b). Without CGL, the network chooses to focus more on the middle rows of the input image stack.

auxiliary gaze loss improves performance over most games compared to AGIL. We provide an analysis that a significant performance improvement in AGIL comes from increased number of model parameters and test-time gaze data, while our approach does not suffer from these issues.

Related Work

Imitation Learning

When learning from demonstrations, Atari game playing has been attempted with various imitation learning approaches. Behavior Cloning (Bain and Sommut 1999; Ross, Gordon, and Bagnell 2011; Daftry, Bagnell, and Hebert 2016) is a standard approach used for imitation learning, where an agent trains a classifier or regressor to replicate an expert's policy by using the demonstrated states and actions as input data for supervised learning. Behavioral Cloning from Observation (BCO) (Torabi, Warnell, and Stone 2018) is a two-phase, iterative imitation learning technique – first allowing the agent to acquire experience in a self-supervised fashion in a task-independent pre-demonstration phase, which is then used to learn a model for a specific task policy only from state observations of expert demonstrations (without access to actions). The self supervision produces an inverse dynamics model to infer actions, given state observations. This model is then used to infer expert actions from state-only demonstrations. The inferred actions along with state information is then used to perform imitation learning for

the agent's policy. GAIL (Ho and Ermon 2016) is an adversarial imitation learning approach trained by alternating the learning updates between a generator policy network and a discriminator network distinguishing between the demonstrated and generated trajectories. It achieved state-of-the-art performance for low-dimensional domains. BCO shows comparable performance to GAIL on low-dimensional MuJoCo benchamrks (Todorov, Erez, and Tassa 2012) with increased learning speed.

Behavior Cloning does not explicitly model the goals or intentions of the demonstrations which a succinct reward function attempts to capture in IRL. Typically, such a succinct inferred reward function makes IRL have better generalization properties compared to behavior cloning (Ross, Gordon, and Bagnell 2011). Most deep learning-based IRL methods either require access to demonstrated actions (Ibarz et al. 2018) or do not scale to high-dimensional tasks such as video games (Finn, Levine, and Abbeel 2016; Fu, Luo, and Levine 2017; Qureshi, Boots, and Yip 2018). Tucker et al. (2018) showed that their adversarial IRL method is difficult to train and fails at high-dimensional tasks of Atari game playing, even with extensive parameter tuning. Aytar et al. (2018) learn a reward function from observations for three Atari games. They guide the agent to exactly imitate the checkpoints from provided demonstrations, assuming access to high-quality demonstrations. T-REX (Brown et al. 2019) is a reward learning from observation algorithm, that extrapolates beyond a set of ranked and potentially suboptimal demonstrations. T-REX outperforms other imitation learning methods such as BCO and GAIL, on Atari and MuJuCo benchmarks (Todorov, Erez, and Tassa 2012) and also demonstrates the ability to extrapolate intentions of a suboptimal demonstrator.

Utilizing Gaze for Learning

Prior studies have shown that human fovea move to the correct place at the right time to extract task-relevant information, making visual attention a feature selection mechanism for humans (Rothkopf, Ballard, and Hayhoe 2007). Novice human learners can benefit from observing experts' gaze (Vine et al. 2012) for learning complex surgical skills. Yamani et al. (2017) showed that viewing the expert gaze videos can improve the hazard anticipation ability of novice drivers. Saran et al. (2019) showed the advantage of incorporating a human demonstrator's gaze for learning robotics manipulation tasks. Penkov et al. (2017) learn the mapping between abstract plan symbols and their physical instances in the environment using eye gaze. Gaze has been exploited in prior imitation learning approaches for autonomous driving (Chen et al. 2019; Xia et al. 2019) and Atari Games (Zhang et al. 2018), but to the best of our knowledge, our work is the first attempt to incorporate gaze in a deep IRL algorithm.

A common method of incorporating human attention is to simply use the gaze map as an additional image-like input (Liu et al. 2019) or predict the gaze heatmap and further use high-resolution parts of the image to improve learning (Zhang et al. 2018; Xia et al. 2019). Zhang et al. (2018) show improved learning on Atari games for imitation learn-

ing (AGIL) but use predicted heatmaps corresponding to demonstration states as part of the input. Gaze-modulated dropout (GMD) was proposed by Chen et al. (2019) to implicitly incorporate gaze into an IL framework for autonomous driving, instead of using gaze as an additional input. An estimated gaze distribution is used to modulate the dropout probability of units at different spatial locations in the first two convolutional layers. GMD is tested with an autonomous driving dataset on an imitation learning network called PilotNet (Liu et al. 2019). Both GMD (Chen et al. 2019) and our auxiliary gaze loss CGL do not increase the complexity of an agent's network, and hence we use it as a baseline for comparison. However, an auxiliary gaze loss only requires gaze data at train time, whereas GMD requires gaze both at train and test time.

Approach

To enable existing IL algorithms to take advantage of human gaze signals accompanying demonstrations, we propose an auxiliary coverage-based gaze loss term. This loss term guides the network to attend to features that human demonstrators attend to. Our approach does not increase the model complexity of existing algorithms in terms of the number of learnable parameters, and can be easily applied to the training of any neural network with convolutional layers.

Data Collection for Atari Game Playing

We use demonstrations collected in the Arcade Learning Environment (ALE) (Bellemare et al. 2013) for a total of 11 unique games. Two types of users provide these demonstrations—experts and novices. Expert users are experienced at playing the games and provide very high-scoring demonstrations. Novice users only get one practice session before playing a game and have no prior experience with any of them. The demonstrations provided by novice users score much lower than those of experts. Expert demonstrations are from the publicly available Atari-HEAD dataset (Zhang et al. 2020). Demonstrations from 2 novice users are collected using the same procedure as followed by Zhang et al. (2020). The subjects were only allowed to play for 15 minutes and were required to rest for at least 5 minutes before the next trial. Users had the option to start the next trial from the point they left off or begin a new trial.

Gaze data was recorded using an EyeLink 1000 eye tracker at 1000 Hz. The game screen was 64.6×40.0 cm (or 1280×840 in pixels), and the distance to the subjects' eyes was 78.7 cm. The visual angle of the screen was 44.6×28.5 visual degrees, where the visual angle of an object is a measure of the size of the object's image on the retina. In the default ALE setting, the game runs continuously at 60 Hz, a speed that is very challenging even for expert human players. An innovative feature of the Atari-HEAD (Zhang et al. 2020) setup is that the game pauses at every frame, until a keyboard action is taken by the human player. This allows users to fixate at all critical locations of the state space before taking an action, producing a richer gaze signal at every time step. If desired, the subjects can hold down a key and the game will run continuously at 20Hz, a speed that

is reported to be comfortable for most players. While providing demonstrations, we observe that users only pause the game at rare instances involving a critical action, but proceed with game play at 20 Hz in most other instances. To generate gaze heatmaps, the discrete gaze positions are converted into a continuous distribution (Bylinskii et al. 2018) by blurring each fixation location using a Gaussian with a standard deviation equal to one visual degree (Le Meur and Baccino 2013).

Coverage-based Gaze Loss

We propose adding an auxiliary loss term based on KL divergence, to the existing loss function for a network, modifying the training procedure of any IL algorithm. Our loss term will penalize the network if it does not focus on parts of the image that the demonstrator focused on, but will have no penalty for activations where the demonstrator did not pay attention. We refer to the proposed loss function as a coverage-based gaze loss (CGL).

CGL operates on the gaze heatmap from the demonstrator and the output of a convolutional layer. Given a 3D feature map f of size $h \times w \times c$ from a convolutional layer, Equation 2 shows the collapse of such a feature map to 2D by summing over the dimension of feature channels, and a further normalization of this 2D feature map to values between 0 and 1. Given a normalized 2D gaze heatmap g of size $h \times w$, CGL is computed as:

$$CGL(g, f'') = \sum_{i \in (1,h), j \in (1,w)} g_{i,j} \left[g'_{i,j} \log \frac{g'_{i,j}}{f''_{i,j}} \right] \quad (1)$$

$$f'_{i,j} = \frac{\sum_{k \in (1,c)} f_{i,j,k} - min(\sum_{k \in (1,c)} f_{i,j,k})}{max(\sum_{k \in (1,c)} f_{i,j,k}) - min(\sum_{k \in (1,c)} f_{i,j,k})} \quad (2)$$

$$g'_{i,j} = \begin{cases} \epsilon, & \text{if } g_{i,j} = 0 \\ g_{i,j}, & \text{otherwise} \end{cases} \quad (3)$$

$$f''_{i,j} = \begin{cases} \epsilon, & \text{if } f'_{i,j} = 0 \\ f'_{i,j}, & \text{otherwise} \end{cases} \quad (4)$$

The loss function adds a penalty if activations from none of the convolutional filters are non-zero on areas where the demonstrator's gaze fixates during game play. If activations are non-zero in other areas where the demonstrator does not focus, there is no penalty. This is because only regions of the gaze map which have a non-zero value contribute to the auxiliary loss, and other regions of the convolutional output which are not fixated on by the demonstrator do not effect the loss term. Hence, our loss term encourages *coverage* of the demonstrator's attention space. The magnitude of penalty is computed using a smoothed (Equations 3, 4 with $\epsilon = 1e^{-10}$) KL divergence term between the normalized gaze map and the collapsed and normalized convolutional map, and is then weighted by the amount of gaze fixation an image region gets (Equation 1). Instead of forcing the filter weights to exactly match the demonstrator's gaze, CGL guides the network to focus on aspects of the state space which might be missed by the network, for example, areas

of the image which are not feature-rich but are critical for decision-making, eventually leading to better performance. A loss function which encourages a network to attend proportional to the human's gaze frequency instead, will be more restrictive.

Auxiliary Gaze Loss for BC For behavior cloning (BC), the gaze coverage loss is added as an auxiliary loss term in addition to the log likelihood action classification loss:

$$\mathcal{L}(\theta) = -\sum_{i=1}^{N} \left[\log \pi_\theta(a_i|s_i) + \alpha \, CGL(g(s_i), c_3(s_i)) \right] \tag{5}$$

The network architecture is similar to the one used in Zhang et al.(2018) – comprised of three convolutional layers and one fully-connected layer. It takes in a single game image as input, and outputs a vector that gives the probability of each action. The gaze coverage loss is applied to the feature maps at the third convolutional layer. $g(s_i)$ is the gaze map of size 16×16, $c_3(s_i)$ is the collapsed and normalized feature map (Equation (2)) from the third convolutional layer, and $N = 50$ is the batch size. We found $\alpha = 0.01$ or $\alpha = 0.1$ worked well in our setting.

Auxiliary Gaze Loss for BCO For BCO (Torabi, Warnell, and Stone 2018), we incorporate CGL as part of learning the imitation policy after the agent learns an inverse-dynamics model of the environment. Similar to Torabi et al. (2018), we use a neural network with three convolutional layers and one fully-connected layer using a stack of four consecutive frames as input. The output is the probability distribution over the discrete action space of the Atari domain. The network is learned using maximum likelihood estimation (MLE), finding the network parameters that best match the provided state-action pairs – states s_i obtained from a demonstrated trajectory τ_i and actions $\widetilde{a}_i$ recovered from the inverse dynamics model. The new loss function is a weighted combination of the standard cross-entropy loss for MLE and CGL applied to the intermediate output of the second convolutional layer (a spatial map of size $32 \times 9 \times 9$) as shown below.

$$\mathcal{L}(\theta) = -\sum_{i=1}^{N} \left[\log \pi_\theta(\widetilde{a}_i|s_i) + \alpha \, CGL(g(s_i), c_2(s_i)) \right] \tag{6}$$

Here, π_θ is the imitation policy network, $g(s_i)$ is the gaze map of size 9×9, $c_2(s_i)$ is the collapsed and normalized feature map (Equation (2)) from the second convolutional layer (we tested CGL on all convolutional layers), and $N = 32$ is the batch size. The Adam (Kingma and Ba 2014) optimizer is used to solve for the network parameters. We found $\alpha = 0.01$ or $\alpha = 0.1$ worked well for our experiments.

Auxiliary Gaze Loss for T-REX T-REX (Brown et al. 2019) is concerned with the problem of reward learning from observation, using rankings of demonstrations to efficiently infer a reward function. To the best of our knowledge, gaze has not been incorporated as part of inverse reinforcement learning for Atari game playing. Given a sequence of m demonstrations ranked from worst to best, $\tau_1, \ldots, \tau_m$,

a parameterized reward network $\hat{r}_\theta$ is trained with a cross-entropy loss over a pair of trajectories ($\tau_i \prec \tau_j$), where τ_j is ranked higher than τ_i. We add CGL to the reward network's loss, so the new loss function becomes:

$$\mathcal{L}(\theta) = -\sum_{\tau_i \prec \tau_j} \log \frac{\exp \sum_{s \in \tau_j} \hat{r}_\theta(s)}{\exp \sum_{s \in \tau_i} \hat{r}_\theta(s) + \exp \sum_{s \in \tau_j} \hat{r}_\theta(s)} +$$
$$\alpha \left[\sum_{s \in \tau_i} CGL(\tau_i^g(s), c_1(s)) + \sum_{s \in \tau_j} CGL(\tau_j^g(s), c_1(s)) \right] \tag{7}$$

$\tau_i^g(s)$ represents the gaze map corresponding to the state s from the trajectory snippet τ_i and $c_1(s)$ represents the convolutional output of the first layer for the same state s. The loss function accumulates gaze over the entire trajectory snippet for both trajectories used as input to the network.

We found $\alpha = 0.01$ worked well for our experiments. We use the same hyper-parameters and network architectures from Brown et al. (2019). The reward network has four convolutional layers with sizes 7×7, 5×5, 3×3 and 3×3 with strides 3, 2 and 1. Each convolutional layer used 16 filters and LeakyReLU non-linearities. The gaze loss is computed over the first convolutional layer output – a spatial map of size $16 \times 26 \times 26$. At the end, a fully connected layer with 64 hidden units with a single scalar output is used to determine the ranking between a pair of demonstrations.

Similar to the implementation of Brown et al. (2019), the trajectories are first subsampled by maximizing over every 3rd and 4th frame, from which a stack of 4 consecutive frames with pixel values normalized between 0 and 1 is passed as input to the reward network. We subsampled 6,000 trajectory pairs with an observation length of 50 time steps. The Adam optimizer (Kingma and Ba 2014) is used with a learning rate of $5e^{-5}$ for $30,000$ steps. The snippets are ranked based on the ground truth rewards or cumulative game scores of the trajectories they are sampled from.

Other Techniques to Incorporate Gaze

Gaze-modulated Dropout (GMD) As a baseline for learning from human gaze, we implement GMD (Chen et al. 2019) after the first convolutional layer of the BCO policy network (Torabi, Warnell, and Stone 2018). The BCO policy network does not originally use dropout layers. Gaze maps are generated using a convolution-deconvolution network (Zhang et al. 2018), trained separately for each game on the Atari-HEAD dataset (Zhang et al. 2020). The gaze prediction network uses as input a stack of 4 consecutive game frames, each of size 84×84, along with their optical flow (Farnebäck 2003) and saliency maps (Itti, Koch, and Niebur 1998). Details of the network architecture are similar to Zhang et al. (2018). We employ this network for gaze prediction, as it has been shown to work well for the Atari domain, instead of the Pix2Pix network (Isola et al. 2017) used by (Chen et al. 2019) for the autonomous driving domain. The generated gaze map is then used as a mask for the additional dropout layer added after first the convolutional layer. Units of the convolutional layer near the estimated

Table 1: BCO performance with and without the usage of expert human demonstrators' gaze

Game	Human	BCO	+GMD	+Motion	+CGL
Asterix	88000-537500	203.3	305.0	276.6	**355.0**
Breakout	344-554	0.9	0.2	0.0	**2.17**
Centipede	39737-251961	2765.0	**3172.4**	2287.1	2195.1
MsPacman	27731-36061	70.0	60.0	70.0	**470.0**
Phoenix	22410-27570	317.3	52.6	619.0	**755.6**
Seaquest	35870 - 445860	0.0	80.0	80.0	**184.0**

Table 2: T-REX performance with and without the usage of expert human demonstrators' gaze

Game	Human	T-REX	T-REX+CGL
Asterix	88000-537500	3746.7	**76383.3**
Breakout	344-554	50.9	**68.9**
Centipede	39737-251961	5742.1	**7586.3**
MsPacman	27731-36061	484.7	**522.7**
Phoenix	22410-27570	166.0	**534.7**
Seaquest	35870 - 445860	0.0	**10.0**

Table 3: Behavior cloning's (BC) and AGIL's performance with and without the usage of expert human demonstrators' gaze

Game	Human	BC	BC-2ch	AGIL	BC+CGL
Asterix	88000-537500	256.7	268.3	368.3	**415.0**
Breakout	425-554	1.2	1.5	**7.0**	3.3
Centipede	18853 - 204294	6360.1	7185.2	9270.4	**9794.1**
MsPacman	27731-36061	1258.7	1121.0	1047.0	**1804.7**
Phoenix	22410 - 57580	1641.0	2925.3	3355.3	**3909.0**
Seaquest	35870 - 445860	198.0	164.7	**326.0**	242.3

Table 4: Behavior cloning's (BC) and AGIL's performance with and without the usage of novice human demonstrators' gaze

Game	Human	BC	BC-2ch	AGIL	BC+CGL
Beamrider	1692 - 8024	368.1	422.4	524.0	**616.8**
Breakout	66 - 427	2.9	2.2	**16.5**	6.6
Enduro	278 - 742	27.2	36.2	107.6	**123.6**
Pong	0-9	-18.1	-17.9	**-11.5**	-15.9
Q*bert	7700 - 11950	623.3	745.8	1111.7	**1213.3**
Seaquest	1600 - 67710	118.0	135.3	161.3	**165.3**
Space Invaders	845 - 2035	207.5	189.3	175.5	**278.3**

gaze location are assigned a lower dropout probability than units far from the estimated gaze location. This is similar to conventional dropout (Srivastava et al. 2014), but with non-uniform dropout probability for spatial units corresponding to different parts of the image space, as described by Chen et al. (2019).

Attention Guided Imitation Learning (AGIL) AGIL adds more parameters to a BC network to utilize gaze. Zhang et al. (2018) train a gaze-prediction network using a stack of 4 image frames denoting the game state, optical flow and saliency (Itti, Koch, and Niebur 1998) as input. The output of the gaze network (gaze saliency map) is then used as an additional input to a modified version of standard behavior cloning. AGIL consists of two channels of 3 convolutional layers. One channel takes as input a single image frame (game state) and another uses a masked image which is an element-wise product of the original image and predicted gaze saliency map. Finally the outputs of the two channels are averaged to predict one of the 18 actions within ALE.

Experiments and Results

We work with demonstrations from 11 unique games in ALE (Bellemare et al. 2013). We evaluate CGL with expert demonstrations for BC, BC, TREX and evaluate CGL with novice demonstrations for BC and its variants. These algorithms are implemented in the OpenAI Gym platform (Brockman et al. 2016), which contains Atari 2600 video games with high-dimensional observation space (raw pixels). All reported results are game scores averaged over 30 different rollouts of the learnt policy. We use the default settings from OpenAI baselines (Dhariwal et al. 2017) for parameters of ALE. Experiments are conducted on server clusters with NVIDIA Titan V or DGX GPUs.

CGL for Imitation Learning Algorithms

We augment BC, BCO and T-REX with CGL and find that CGL improves performance over all algorithms for most games. For BCO (Table 1), CGL outperforms basic BCO for 5 out of 6 games with expert demonstrators. For T-REX (Table 2), CGL outperforms on all games with expert demonstrations. For BC (Tables 3, 4), CGL outperforms basic BC on all games for both experts and novice demonstrations. These results confirm that the gaze information is beneficial for imitation learning algorithms.

CGL versus GMD

We test GMD and CGL with BCO and find that for 5 out of 6 games, CGL outperforms GMD (columns 4, 6 in Table 1). One may expect that convolutional dropout helps generalization by reducing over-fitting. However, it is far less advantageous, since the shared-filter and local-connectivity architecture in convolutional layers is a drastic reduction in the number of parameters and this already reduces the possibility to overfit (Hinton et al. 2012). Empirical results by Wu et al. (2015) confirm that the improvement in generalization to test data from convolutional dropout is often inferior to max-pooling or fully-connected dropout.

CGL versus AGIL

For this experiment, we investigate how well the coverage-based gaze loss compares against another common method of incorporating gaze in imitation learning networks, i.e. using gaze as an input. To tease apart whether improvement in prior IL approaches (such as AGIL (Zhang et al. 2018)) comes from increased parameters to standard behavior cloning or from the gaze information itself, we perform

(a) A masked frame stack for Ms. Pacman
(b) Motion heatmap
(c) Gaze heatmap

(d) A masked frame stack for Asterix
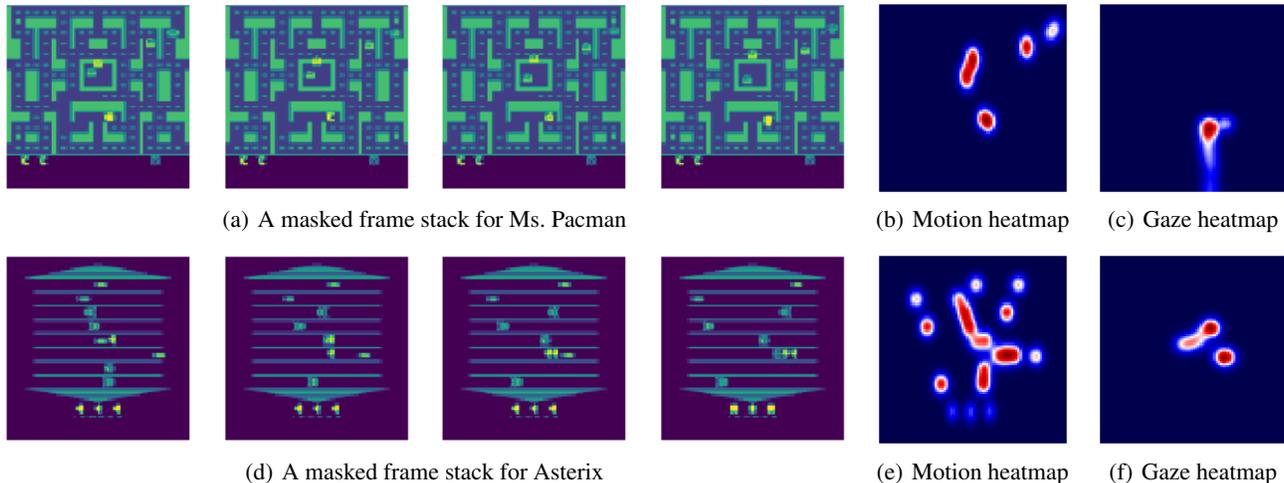(e) Motion heatmap
(f) Gaze heatmap

Figure 2: Motion in the visual game state, i.e. the difference between the last and first frame in an input image stack, cannot alone explain human attention. This is further highlighted with minimal performance gains when CGL utilizes the motion heatmap instead of human gaze.

the following experiment. We re-train the AGIL network, but instead of using gaze heatmaps, we pass the original image as input to the gaze pathway, referred as the BC-2ch model. This helps us disambiguate if more parameters in the model help more versus the gaze information itself. As shown in Tables 3 and 4, we find that the BC-2ch model improves performance over BC for 4 out of 6 games with expert demonstrators (Table 3) and 5 out of 7 games with novice demonstrators (Table 4). This hints at the fact that increasing model complexity alone without using any additional information as input proves beneficial.

Our work is focused on utilizing the gaze information itself without increasing number of model parameters. Our method (BC+CGL) utilizes gaze without changing model complexity and does not require gaze prediction for the input state at test time. BC+CGL consistently outperforms basic BC and BC-2ch models in all games by a significant margin, as shown in Tables 3 and Table 4. Compared to the previous state-of-the-art model AGIL, CGL performs better in 4 out of 6 expert game evaluations and 5 out of 7 novice game evaluations, even though AGIL adds learnable parameters to BC and requires gaze prediction at test time.

CGL for different convolutional layers

To choose a specific layer for computing and reporting results on the CGL loss for each IL algorithm, we test different combinations of α values $(0.01, 0.1, 1.0, 10.0)$ along with applying CGL on the output of different convolutional layers. Table 5 shows the results of applying CGL to each convolutional layer of the BCO network. As we know from Table 1, except Centipede CGL outperforms BCO and GMD for all games. However in Table 5, we also see that for two values of $\alpha = 0.01, 0.1$, CGL applied to the second convolutional layer outperoms CGL applied to other layers for most games. Based on such an analysis, we choose the layer which performs the best for each IL algorithm individually.

Table 5: Analysis of CGL applied to different layers of the BCO network for expert users. BCO+CGL$_1$ refers to CGL being computed with the outputs of the first convolutional layer and so on.

Game	BCO+CGL$_1$	BCO+CGL$_2$	BCO+CGL$_3$	BCO+CGL$_1$	BCO+CGL$_2$	BCO+CGL$_3$
	$\alpha = 0.01$			$\alpha = 0.1$		
Asterix	131.7	**293.3**	240.0	220	**355.0**	318.3
Breakout	0.4	**1.8**	0.2	1.9	**2.2**	0.5
Centipede	1762.8	2195.1	**2207.9**	**2630.6**	1916.6	703.9
MsPacman	60.0	**470.0**	70.0	60.0	60.0	**70.0**
Phoenix	532.7	**755.7**	370.3	**546.7**	443.3	459.3
Seaquest	**210.0**	184.0	124.0	160.0	**170.0**	104.0

For each game, we report results for the best α value for the chosen layer.

CGL performs the best on different convolutional layers for different algorithms, however, the architecture for each algorithm is different as well. Spatial resolution of the gaze heatmaps and convolutional feature maps can vary the magnitude of the loss penalty. Initial layers will apply the loss to specific regions of human attention using a higher resolution heat map, hence penalizing a lack of attention coverage on the low-level visual features. Whereas deeper in the network, the loss will be computed over a broader region of the input image via a low resolution heat map, penalizing a lack of attention coverage on high-level visual features. T-REX gains the most with gaze using its first convolutional layer. There might be subtle differences in terms of low-level features that helps distinguish better between a pair of trajectory snippets. BC on the other hand directly predicts actions given the state of the game. It benefits the most with gaze from the last convolutional layer which depicts high-level features representing the state and loses low-level visual details.

Motion versus human gaze attention

Prior work has established that human gaze encodes attention which is different from salient regions in a scene (Land 2009; Hayhoe and Ballard 2005; Rothkopf, Ballard, and Hayhoe 2007; Tatler et al. 2011). We test whether our proposed approach extracts the additional information from human gaze data, in comparison to what might already be encoded in the visual game state, such as motion. We replace the gaze heat maps used by CGL with heatmaps representing the normalized motion in an input image frame stack (difference between the last and first frame in a stack). We test this motion-based CGL loss for BCO (Table 1) which didn't show as much of an improvement in comparison to human attention heat maps.

Conclusions and Future Work

In this work, we introduce an auxiliary coverage-based gaze loss (CGL) term which can guide the training of any imitation learning network with convolutional layers. Our results show improved performance on several Atari games over state-of-the-art imitation learning algorithms. Our approach provides these gains without requiring gaze prediction at test time or increasing the model complexity of existing algorithms. We outperform a baseline method (GMD), which also does not increase model complexity. On most games, CGL outperforms another method (AGIL) which utilizes gaze in the form of an additional input to a BC algorithm. AGIL requires gaze prediction at test time and is shown to gain performance by increasing model complexity alone. Our approach improves performance by utilizing gaze without these shortcomings.

We highlight that utilizing human gaze provides additional information to what is encoded implicitly in the game state (such as motion). Our work confirms prior research showing gaze can help extract more information from a demonstrator than traditional state-action pairs. Moreover, we present a new methodology to incorporate gaze effectively by augmenting existing algorithms, and only requiring access to human gaze data at training time. We hope our work encourages the research community to innovate on other novel ideas for efficiently incorporating human attention. Human gaze and actions from demonstrations may be correlated in time. Our approach only utilizes gaze per game state, and so do all other approaches we compare against. Utilizing temporal connections in the gaze signal is a direction for future work.

References

Argall, B. D.; Chernova, S.; Veloso, M.; and Browning, B. 2009. A survey of robot learning from demonstration. *Robotics and autonomous systems* 57(5):469–483.

Aytar, Y.; Pfaff, T.; Budden, D.; Paine, T.; Wang, Z.; and de Freitas, N. 2018. Playing hard exploration games by watching youtube. In *Advances in Neural Information Processing Systems*, 2930–2941.

Bain, M., and Sommut, C. 1999. A framework for behavioural claning. *Machine intelligence* 15(15):103.

Bellemare, M. G.; Naddaf, Y.; Veness, J.; and Bowling, M. 2013. The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research* 47:253–279.

Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.; Schulman, J.; Tang, J.; and Zaremba, W. 2016. Openai gym. *arXiv preprint arXiv:1606.01540*.

Brown, D. S.; Goo, W.; Nagarajan, P.; and Niekum, S. 2019. Extrapolating beyond suboptimal demonstrations via inverse reinforcement learning from observations. *arXiv preprint arXiv:1904.06387*.

Bylinskii, Z.; Judd, T.; Oliva, A.; Torralba, A.; and Durand, F. 2018. What do different evaluation metrics tell us about saliency models? *IEEE transactions on pattern analysis and machine intelligence* 41(3):740–757.

Chen, Y.; Liu, C.; Tai, L.; Liu, M.; and Shi, B. E. 2019. Gaze training by modulated dropout improves imitation learning. *arXiv preprint arXiv:1904.08377*.

Daftry, S.; Bagnell, J. A.; and Hebert, M. 2016. Learning transferable policies for monocular reactive mav control. In *International Symposium on Experimental Robotics*, 3–11. Springer.

Dhariwal, P.; Hesse, C.; Klimov, O.; Nichol, A.; Plappert, M.; Radford, A.; Schulman, J.; Sidor, S.; Wu, Y.; and Zhokhov, P. 2017. Openai baselines. https://github.com/openai/baselines.

Farnebäck, G. 2003. Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, 363–370. Springer.

Finn, C.; Levine, S.; and Abbeel, P. 2016. Guided cost learning: Deep inverse optimal control via policy optimization. In *ICML*, 49–58.

Fu, J.; Luo, K.; and Levine, S. 2017. Learning robust rewards with adversarial inverse reinforcement learning. *arXiv preprint arXiv:1710.11248*.

Hayhoe, M., and Ballard, D. 2005. Eye movements in natural behavior. *Trends in cognitive sciences* 9(4):188–194.

Hinton, G. E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. R. 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Ho, J., and Ermon, S. 2016. Generative adversarial imitation learning. 4565–4573.

Ibarz, B.; Leike, J.; Pohlen, T.; Irving, G.; Legg, S.; and Amodei, D. 2018. Reward learning from human preferences and demonstrations in atari. In *Advances in Neural Information Processing Systems*, 8011–8023.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *CVPR*, 1125–1134.

Itti, L.; Koch, C.; and Niebur, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (11):1254–1259.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Land, M. F. 2009. Vision, eye movements, and natural behavior. *Visual neuroscience* 26(1):51–62.

Le Meur, O., and Baccino, T. 2013. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior research methods* 45(1):251–266.

Liu, C.; Chen, Y.; Tai, L.; Ye, H.; Liu, M.; and Shi, B. E. 2019. A gaze model improves autonomous driving. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 33. ACM.

Penkov, S.; Bordallo, A.; and Ramamoorthy, S. 2017. Physical symbol grounding and instance learning through demonstration and eye tracking. *International Conference on Automation and Robotics (ICRA)*.

Qureshi, A. H.; Boots, B.; and Yip, M. C. 2018. Adversarial imitation via variational inverse reinforcement learning. *arXiv preprint arXiv:1809.06404*.

Ross, S.; Gordon, G.; and Bagnell, D. 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 627–635.

Rothkopf, C. A.; Ballard, D. H.; and Hayhoe, M. M. 2007. Task and context determine where you look. *Journal of vision* 7(14):16–16.

Saran, A.; Short, E. S.; Thomaz, A.; and Niekum, S. 2019. Understanding teacher gaze patterns for robot learning. *Conference on Robot Learning (CoRL)*.

Schaal, S. 1997. Learning from demonstration. In *Advances in neural information processing systems*, 1040–1046.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15(1):1929–1958.

Stewart, R., and Ermon, S. 2017. Label-free supervision of neural networks with physics and domain knowledge. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Sutton, R. S., and Barto, A. G. 2018. *Reinforcement learning: An introduction*. MIT press.

Tatler, B. W.; Hayhoe, M. M.; Land, M. F.; and Ballard, D. H. 2011. Eye guidance in natural vision: Reinterpreting salience. *Journal of vision* 11(5):5–5.

Todorov, E.; Erez, T.; and Tassa, Y. 2012. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5026–5033. IEEE.

Torabi, F.; Warnell, G.; and Stone, P. 2018. Behavioral cloning from observation. In *IJCAI*, 4950–4957. AAAI Press.

Tucker, A.; Gleave, A.; and Russell, S. 2018. Inverse reinforcement learning for video games. *arXiv preprint arXiv:1810.10593*.

Vine, S. J.; Masters, R. S.; McGrath, J. S.; Bright, E.; and Wilson, M. R. 2012. Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills. *Surgery* 152(1):32–40.

Wu, H., and Gu, X. 2015. Towards dropout training for convolutional neural networks. *Neural Networks* 71:1–10.

Xia, Y.; Kim, J.; Canny, J.; Zipser, K.; and Whitney, D. 2019. Periphery-fovea multi-resolution driving model guided by human attention. *arXiv preprint arXiv:1903.09950*.

Yamani, Y.; Bıçaksız, P.; Palmer, D. B.; Cronauer, J. M.; and Samuel, S. 2017. Following expert's eyes: Evaluation of the effectiveness of a gaze-based training intervention on young drivers' latent hazard anticipation skills.

Zhang, R.; Liu, Z.; Zhang, L.; Whritner, J. A.; Muller, K. S.; Hayhoe, M. M.; and Ballard, D. H. 2018. Agil: Learning attention from human for visuomotor tasks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 663–679.

Zhang, R.; Torabi, F.; Guan, L.; Ballard, D. H.; and Stone, P. 2019. Leveraging human guidance for deep reinforcement learning tasks. *IJCAI*.

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