Exploring Coarse-grained Pre-guided Attention to Assist Fine-grained Attention Reinforcement Learning Agents*

Haoyu Liu
School of Statistics
Renmin University of China
Beijing, China
2017201665@ruc.edu.cn

Yang Liu
Language Understanding Lab
Samsung Research China, Beijing
Beijing, China
yang9004.liu@samsung.com

Xingrui Wang
School of Statistics
Renmin University of China
Beijing, China
XingruiWang@ruc.edu.cn

Hanfang Yang†
School of Statistics
Renmin University of China
Beijing, China
hyang@ruc.edu.cn

***Abstract*—Recently, people have applied the attention mechanism to deep reinforcement learning (DRL), which commits to helping agents focus on crucial factors to learn the task more effectively. However, there is still some margin between the current attention methods and natural human attention since evidence suggests that human attention can be pre-guided before they perform a task, allowing humans to quickly catch areas of important factors at the beginning of the task and then gradually refine fine-grained attention to learn the details during training. This allows humans to use their attention more efficiently. In this paper, we propose an attention method that mimics human attention for DRL in the Atari Games. The proposed method contains a fusion attention module, for which we build a simulated human coarse-grained pre-guided (SHCP) attention module to assist the original fine-grained attention of RL agents. The proposed SHCP attention module contains information about key objects for game tasks and is implemented as a coarse-grained attention region. The experimental results demonstrate that our method can quickly boost performance in the early stages and then outperform the current state-of-the-art fine-grained attention methods significantly in sample efficiency, just like human attention. Further analysis shows that, with fusion attention, agents can not only capture rich features of pre-guided attention but also extend to more improved features after training, which suggests the pre-guided attention signal acts as a good initializer. Therefore, we consider our work reveals a potential and promising direction that combines human attention signals to affect agents' behavior via attention mechanisms.**

***Index Terms*—deep reinforcement learning, attention mechanism.**

I. Introduction

When playing Atria games, humans can quickly understand what the important objects are in the games and reach high scores after a few attempts. A main reason is the human attention mechanism, which allows us to ignore massive and irrelevant information, and focus on the key factors related to tasks [1], [2]. Inspired by this, the attention mechanism has been widely applied in prevailing RL methods recently. They focus on speed up the training process by mimicking human attention ability [3]–[6] and showing comparable performance and great interpretability in RL attention agents [7]–[9].

These methods can indeed capture important factors during training, while holding the limitation that the attention modules are learned from scratch at pixel-level (aka fine-grained attention), which takes a large number of steps to first learn how to attend at the beginning of training.

While for humans, we can receive pre-guided information for attention training before performing the task, such as observing the task, reading the relevant manual instructions, and so on. In detail, when we play Atria games, we can benefit a lot from reading the game manual to identify the important characters or observe the inner logic to achieve a higher score [10]. Besides, in the Atria human eye-tracking dataset [11], people notice that humans are constantly tracking important entities and objects during tasks. In [12], it suggests that humans can also master games faster through the guide of the game manual, which reminds us of the key information and makes our attention more selective.

Inspired by these works, we wonder if the pre-guided process can be applied to DRL. In fact, the learning process between agents and humans holds differences in that humans tend to pay attention to key information in a game at the beginning, and this attention can be prompted in advance by some game-related experience such as the game manual, which differs from the scratch learning as agents. Thus, we tend to explore a more human-like attention mechanism, which we call simulated human coarse-grained pre-guided (SHCP) attention, to strengthen the learning of RL agents, achieving higher sample efficiency and intuitive interpretability.

In this paper, we propose a human-like attention model – simulated human coarse-grained pre-guided attention procedure. We first generate the Simulated Human Coarse-grained Pre-guided (which we call SHCP below) attention information through selective entities in Atria and matching modules. Second, we use a popular top-down style fine-grained attention as our backbone system, and it will constantly generate fine-grained attention signals based on features at pixel-level. Finally, we embed the SHCP attention information in the

* This work is supported by Research Center for Metaverse of Renmin University of China and Hanfang Yang was supported by the National Key R&D Program of China (Grant No. 2018YFC2000302)

† Corresponding author

fine-grained attention process through a fusion module we designed and assist in the training of the RL attention agents. We tested our method on several public Atari environments available on OpenAI Gym. The experimental results suggest that our method can outperform the current SOTA systems significantly with better sample efficiency and stability while costing meagre human effort and being extremely easy to use. Moreover, the interpretability analysis suggests that our method can lead the agent to not only focus on critical objects where human attention is pointed to, but also capture more factors vital to tasks. We believe that our work reveals a potential and promising direction that combines human pre-guided attention signals to affect agents' behavior via attention mechanisms.

II. Related Work

Deep learning enables RL to scale to problems with high dimensional state and action spaces such as video games [13]–[15]. In particular, deep RL can now tackle tasks directly from screen pixels, using variants of the DQN algorithm [16] and actor-critic algorithms [17]–[21]. These successful models are known as model-free algorithms [21], [22]. As the game environment becomes more complicated, the time cost and sample efficiency for these algorithms become heavy. Although the sample complexity has substantially improved in recent years, deep RL methods still require far more experience than human players to learn in each game environment [10]. To improve the training process of RL agents, people have tried to apply attention mechanism in RL agents [23]–[26]. Generally, the attention mechanism has achieved huge success in NLP [27], [28], computed vision [29]–[32] and virtual application systems [33]–[35]. And for DRL agents, the attention mechanism enables them to focus on key features of observations, thus improving the performance and sample efficiency [3], [5], [7]. However, there is still some margin between the attention in the current DRL and natural human attention. Evidence suggests that human attention can be pre-guided at very coarse-grained scales [10], which allows humans to quickly catch areas of important targets and gradually refine the fine-grained attention to learn the details during training [36]. While for DRL agents, their attention modules are all learnt from scratch and require enormous training steps. In fact, this human pre-guided attention mechanism can be applied to DRL agents greatly, where we generate a simulated human coarse-grained pre-guided (SHCP) attention to assist the original fine-grained attention mechanism of DRL agents in this paper.

III. Method

Our simulated human coarse-grained pre-guided (SHCP) attention procedure mainly contains two steps. (1) Generate the simulated human coarse-grained pre-guided attention information consulting reliable resource such as humans playing experience and the game manual. (2) Embed the simulated human coarse-grained pre-guided attention information into the self-attention process to affect the agents' decision making in the controller loop.

A. SHCP Attention Maps Generation

In this section, we will detailedly introduce the process of generating simulated human coarse-grained pre-guided (SHCP) attention information. In detail, our SHCP attention information is attention maps (a matrix consisting of 0 and 1) which reflect key entities in a game. Since real human attention data is hard to obtain and we've confirmed that humans tend to focus on moveable and changeable objects in games, we tend to mimic this attention behavior and generate our SHCP attention information as a similuated human attention signal, which focuses on the key entities' region. According to current studies [10]–[12], we tend to focus on (1) key roles mentioned in the game manual and (2) movable and changeable entities. In other words, we only focus on movable and changeable entities mentioned in the game manual. Besides, benefiting from the relatively simple Atria game environment and highly distinguishable entities in color, We design a simple and efficient object tracking system suitable for the Atria environment to generate the SHCP attention information.

1) Template Matching: Template matching is the technique of finding the location of a given template image in a target (larger) image. We use it to locate the potential existence of moveable objects (template image) in the input observation (target image). This module will return a corresponding binary matrix (SHCP attention map) to represent the location of entities and affect the attention behavior after being fused with self-attention maps.

2) Collecting Templates and Generating SHCP attention Maps: The template matching algorithm requires a template image to locate the area in the target image. So, we first collect templates for each entity in a given Atria environment. As shown in Figure 2, the target image is an observation frame of the MsPacman Atria environment. In this environment, there are two types of moveable entities: the yellow pacman and the four monsters (The pacman needs to eat as more beans as possible and avoid touching the monsters to achieve higher scores). Thus, we first cut out all the possible templates for each type of entity: 4 template images for Pacman and 4 template images for the monsters. Next, we need to locate the entities in the target image, and the same type of entity will generate a single SHCP attention map which contains all the areas of this type of entity.

Formally, given the template images, template matching can recognize the related objects in the task scenario. With a source observation image $O \in \mathbb{R}^{H\times W\times C}$ and a template image $I^n \in \mathbb{R}^{h\times w\times C}$ for one specific type of entity, template matching attention module will render a SHCP attention map matrix $A^n_{\mathrm{H}} \in \mathbb{R}^{H\times W}$ to represent the areas of the template image in the observation image.

Specifically, we generate M attention map matrices for M types of entities. Each single SHCP attention map A^n_{H} is calculated as:

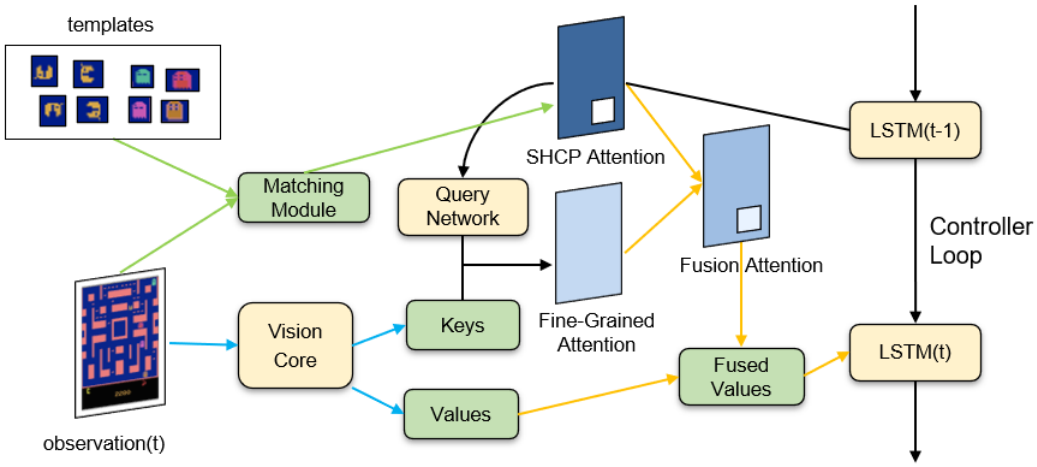

Fig. 1. The architecture of the attention fusion procedure 1) It starts with accessing the simulated human coarse-grained pre-guided attention (SHCP) attention information via entity templates. 2) With the matching module, we can locate pre-guided template images within the whole observation and generate SHCP attention maps. 3) With a top-down self-attention mechanism, we can also generate self-attention maps with queries from query network calculated based on LSTM controller state and keys from the vision core, which can encode observations from the environment. 4) We fuse the simulated human coarse-grained pre-guided attention maps with the self-attention maps to create fusion attention maps. 5) Fusion attention maps combine with values to generate final representation of the observation and then are sent to LSTM controller to make action decisions.

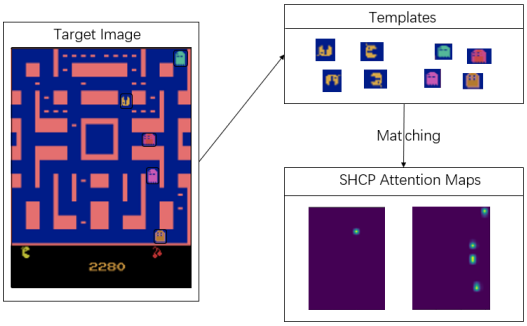

Fig. 2. An example of generating SHCP attention maps in a specific Atria environment (MsPacman here). By default, entities of the same type will be contained in one single map (The SHCP attention signal of the four monsters in the target image is generated in one map here).

$$\tilde{A}_{\mathrm{H},i,j} = O_{i,j} \cdot I^n, \tag{1}$$

$$A_{\mathrm{H},i,j} = \begin{cases} 1, & \tilde{A}_{\mathrm{H},i,j} \geq \tau \\ 0, & \tilde{A}_{\mathrm{H},i,j} < \tau \end{cases}, \tag{2}$$

where $\tau > 0$ is the threshold of similarity to transform the tensor $\tilde{A}_{\mathrm{H}}$ into the binary matrix attention map A_{H}. We concat all types of entities' attention maps together as our SHCP attention maps A_H^M.

B. Fine-grained attention in Self-Attention Mechanism

To combine the SHCP attention information (the SHCP attention maps generated) with RL agents' attention, we preserve a complete top-down self-attention system that contains queries, keys, and values. A query consists of N vectors, which are generated by the LSTM controller. It can actively search for important information in observation.

For each step, features $F \in \mathbb{R}^{h_1 \times w_1 \times c}$ is extracted from the observation $O \in \mathbb{R}^{H \times W \times C}$ by a "vision core" denoted as $\mathrm{Vis}(\cdot)$, which is a multi-layer CNN and a recurrent layer (see Vision Core in Figure 1).

$$F, s_{\mathrm{vis}}(t) = \mathrm{Vis}(O, s_{\mathrm{vis}}(t-1)), \tag{3}$$

where $s_{\mathrm{vis}}(t)$ is the recurrent layer state in vision core of time step t.

We split feature F along the channel dimension into two tensors, namely keys $K \in \mathbb{R}^{h_1 \times w_1 \times c_k}$ and values $V \in \mathbb{R}^{h_1 \times w_1 \times c_v}$. Keys are utilized to detect the task related features. Once the query observes the vital key-value pattern, it will produce an attention map, which is the weight of corresponding values and the result will be pushed into next step controller. This step is similar to the self-attention mechanism used in transformer.

An LSTM controller in the start will produce the state $s_{\mathrm{LSTM}}(t-1)$ from the previous time step $t-1$, which will be passed through a Query Network Q into N query vectors, names as N attention heads $q^1, \ldots, q^N \in \mathbb{R}^{c_k}$, matching the channel of $K \in \mathbb{R}^{h_1 \times w_1 \times c_k}$:

Meanwhile, attention works with spatial information and we also use a spatial encoding to summarize representation [1].

$$q^1, \ldots, q^N = Q(s_{\mathrm{LSTM}}(t-1)). \tag{4}$$

For a single attention head q^n, the fine-grained attention map $A_{\mathrm{f}} \in \mathbb{R}^{h_1 \times w_1}$ for the current step is calculated by two steps.

[1]We use same spatial encoding method described in [7]. The spatial encoder can preserve 2D location information for representation.

We first calculate $\tilde{A}_{\mathrm{f}}$ by taking the inner product between all locations of key tensor K and the query vector q^k:

$$\tilde{A}^n_{\mathrm{f},i,j} = \sum_c q^n_c \cdot K_{i,j,c}. \tag{5}$$

Then we apply the softmax function along the spatial axis and obtain the fine-grained map $A_{\mathrm{f}} \in \mathbb{R}^{h_1 \times w_1}$ which can encode the spatial information for different locations from observation:

$$A^n_{\mathrm{f},i,j} = \frac{\exp(\tilde{A}^n_{\mathrm{f},i,j})}{\sum_{i',j'} \exp(\tilde{A}^n_{\mathrm{f},i',j'})}. \tag{6}$$

C. Attention Fusion Module

Our goal is to aggregate information of SHCP attention maps with the fine-grained attention maps learned from observation. We fuse the SHCP attention maps and self-attention maps by a fusion module.

1) Fusion Module: First, the fusion module concatenates two groups of attention maps together along channel dimension into a tensor $\tilde{A}^n$:

$$\tilde{A}^N = \mathrm{Concat}(A^N_{\mathrm{f}}, A^M_{\mathrm{H}}), \tag{7}$$

where each map of A_{h} is resized to the same size as A^N_{self}.

In order to blend the information of two sources of attention maps thoroughly and match the dimension of tensor V, we reduce the dimension of $\tilde{A}^N$ by a convolution block and obtain the final attention map $A^N \in \mathbb{R}^{h_1 \times w_1 \times N}$ for N attention heads.

$$A^N = \mathrm{Conv}(\tilde{A}^N). \tag{8}$$

Then, the weighted state representation of observation $R^n \in \mathbb{R}^{c_v}$ for each attention head is the summation of point product between every new map and the values:

$$R^N_l = \sum_{i,j} A^N_{i,j} \cdot V_{i,j,l}, \tag{9}$$

where $l = 1, \ldots, c_v$.

a) Fusion Attention for Controller Loop: The weighted state representation R^N of each attention head q^N is concatenated to R and fed into the LSTM controller:

$$o(t), s_{\mathrm{LSTM}}(t) = \mathrm{LSTM}(R, q, s_{\mathrm{LSTM}}(t-1)) \tag{10}$$

Finally, the output is utilized to fit the policy network and value function in actor-critic RL. The controller LSTM is followed by a policy and a value function network to output actions and state values. The controller takes the query, answer, reward, and a one-hot encoding of action from the previous state as its input. In summary, our method generates and combines the SHCP attention information with the fine-grained attention in the self-attention mechanism for RL agents to affect the decision-making controller.

IV. Experiment

In this section, we describe our experiment setup, including environment, evaluation metrics, baseline systems, and model implementation details.

A. Environment

We conduct our experiments on mainly 4 popular Atari environments available on the OpenAI gym platform, which is a common setting for RL research, including MsPacman, KungfuMaster, Seaquest, and Pong. The size of a single observation frame in these environments is (210,160,3), and the player can control the movable agents to play the games[2]. Each game can output a score to evaluate the performance of RL agents immediately. In the four gym games, a better RL agent is considered to achieve higher scores with fewer training steps.

B. Evaluation Metrics

We evaluate performance using two metrics defined and employed in previous work [37]:

- Average reward, which is the area under the reward curve divided by the total steps.
- Asymptotic performance, which is the rewards over 10 episodes of the same training steps.

The first metric emphasizes the sample efficiency that a method can achieve during learning in the target domain within constant steps. The second one evaluates the ability to achieve optimal performance on the task. In our case, we set 5e7 as our total steps. In general, an ideal method should perform well on both the two metrics above.

C. Baseline Systems

In order to explore the effectiveness of our fusion attention reinforcement learning system, we consider the following different conditions for ablation study:

- LSTM Controller with Non-attention. We use the LSTM version of IMPALA implement as one of our benchmark system. IMPALA is an open-source popular RL library provided by DeepMind [21].
- Top-Down Self-attention. We employ a RL with top-down self-attention method as our another benchmark, which is open-source and competitive with several state-of-the-art baselines [7].

D. Experimental Settings

The vision core consists of a 3-layer CNN followed by a convolutional LSTM. The query network is a 3-layer MLP, and it produces 4 attention queries (See Appendix A.1 for the full architecture of the network). We use an actor-critic setup, a VTRACE loss with an RMSProp optimizer, and the IMPALA to train our agents. To perform our method, we prepare the different SHCP attention maps for specific task environments. An example is shown in Figure 2. The number of elements for each environment is shown in Table I. For experiments, we used four GTX2080 GPUs for about 60 hours.

E. Experimental Results

Figure 3 shows the score curves along with $5e7$ steps and Table II and Table III shows results evaluated by the metrics

[2]We use skip-frame setting here, system applies same action to interact with environment in each 4 frames. Frameskip is the number of frames an action is repeated before a new action is selected.

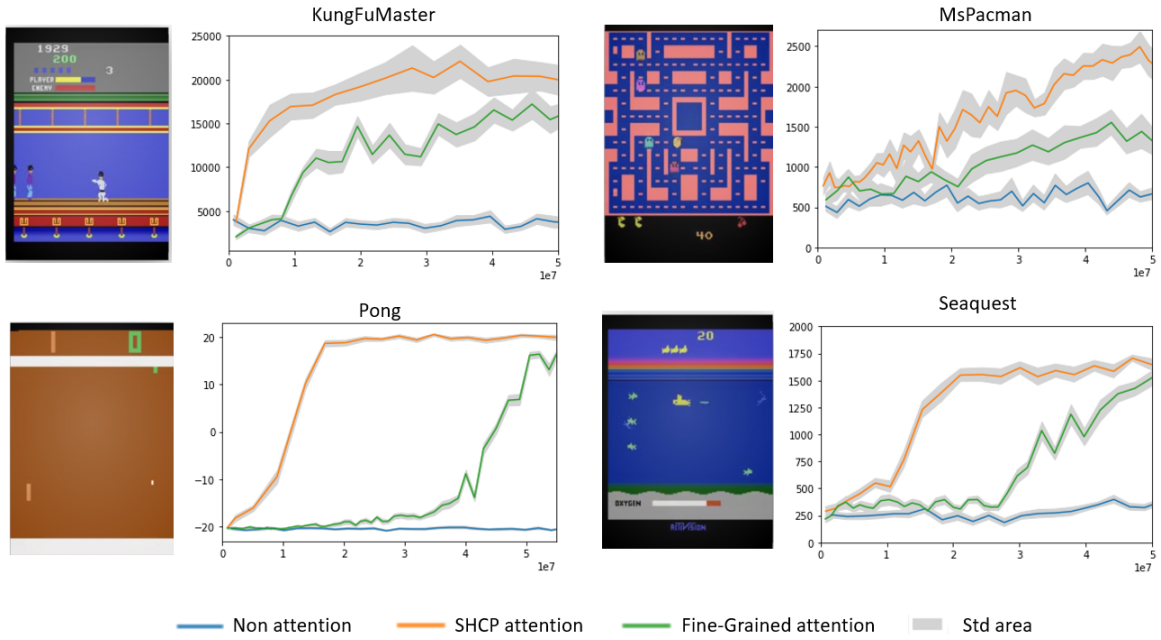

Fig. 3. Reward curves for four different Atari games. The orange line denotes our method, the blue denotes the non-attention LSTM RL and the green denotes the top-down self-attention RL system. The results show that our method can outperform the other two methods in sample efficiency by a great deal. Meanwhile, our method reached the highest scores in all four of the game environments.

TABLE I
THE HYPER-PARAMETERS FOR EACH TASK. **MAPS** INDICATES THE NUMBER OF SHCP ATTENTION MAPS, **IMAGES** INDICATES THE NUMBER OF TEMPLATE IMAGES WE USED.

Env	**Maps**	**Images**
MsPacmans	2	8
Pong	2	2
Seaquest	3	5
Kungfu master	2	6

TABLE II
EXPERIMENTAL RESULTS EVALUATED USING ASYMPTOTIC METRIC.

Env	No-Att	Self-Att	Prior-Att
MsPacman	$663_{\pm 49}$	$1589_{\pm 89}$	$\mathbf{2583}_{\pm 153}$
Kungfu Master	$3872_{\pm 463}$	$19862_{\pm 5242}$	$\mathbf{24910}_{\pm 3371}$
Seaquest	$267_{\pm 43}$	$1668_{\pm 121}$	$\mathbf{1842}_{\pm 12}$
Pong	$-20_{\pm 0.31}$	$19_{\pm 0.78}$	$\mathbf{21}_{\pm 0}$

TABLE III
EXPERIMENTAL RESULTS EVALUATED USING AVERAGED REWARD METRIC.

Env	Non-Att	Self-Att	Prior-Att
MsPacman	644.7	1094.5	**1629.7**
KungfuMaster	3458.2	12232.5	**18444.8**
Seaquest	296.1	747.6	**1304**
Pong	0.57	11.2	**33.5**

above. In general, our method can outperform LSTM non-attention and top-down self-attention methods significantly. The score curve and average reward (see Table III) results suggest our method has much better sample efficiency. In particular, we observe that in most of the tasks, the performance improves quite fast in the early stages. In our opinion, the SHCP attention information can guide the fine-grained attention in an agent to find the key factors for the task at the beginning, instead of searching blindly and learning the bias from scratch. According to the asymptotic results (see Table II), our method can output actions with less variance than the original top-down self-attention method, which contains only fine-grained attention information. The reason behind this is that our model provides more concentrated focus areas than the fine-grained self-attention learning within the RL framework, which is rather unstable due to the nature of RL. Finally, our method achieves extraordinary improvement with negligible human effort cost (see Table I). In the following parts, we will explain why our model works with a series of visualization results.

F. Behavior of Fusion Attention

Figure 4 shows the difference between the fusion attention map and the fine-grained attention map corresponding to the same observation frame in the game MsPacman. The fusion attention map comes from integrating the SHCP attention map with the preserved fine-grained attention map, and its focus area is highly concentrated. A sensible reason is that the fusion attention map is strongly affected by the SHCP attention information and reinforced during the training process as

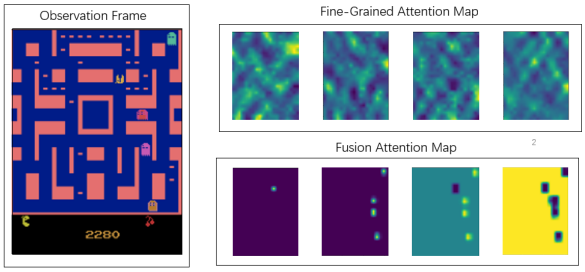

Fig. 4. This figure shows the visualization of 4 attention heads of fine-grained attention map (top) and the fusion attention map (down) respectively. The image left is the input game frame. In the visualization result, we can clearly observe that the fusion attention capture both the coarse-grained areas which containing moveable entities and the background environment information. In fact, it's an excellent combination of the SHCP attention information and the fine-grained attention information. The results also reflects that the SHCP attention information assists the fine-grained attention process to only focus on details, reducing the search space and difficulty of learning.

the areas that contain key entities are relevant to the task. The SHCP attention information gives the agent advantages to quickly improve performance in the early stage since it accesses to the priors of key factors via SHCP attention signal.

G. SHCP Attention Signal Comparison

The observations above naturally raise a doubt about whether the fusion attention makes the agent only focus on the SHCP attention information and whether they can learn more from training. To explain this issue, we show a series of actions of our fusion attention head in the game (see Figure 5). For this task, the "player" and the "fighter" are two types of moveable entities related to images marked in the black box. In this scenario, the fusion attention heads tend to notice the areas where fights are likely to happen instead of only tracking the player and the other fighters. These observations further indicate that the SHCP attention information acts more like a guidance signal than a supervised signal. In other words, agents can learn better from interaction by treating SHCP attention information as a beneficial assistance information.

V. Conclusion

We explore and propose a human-like coarse-grained pre-guided attention to assist fine-grained reinforcement learning attention agents. The results show we can achieve significant improvement in several 2D Atari games, and the further analysis suggests that our SHCP attention information can effectively combine with the top-down self-attention mechanism, fusing the SHCP attention information with the fine-grained attention information to assist training for RL agents. Moreover, we consider our SHCP attention signal to act more like a guidance signal than a supervised signal, which providing a reliable source of information to help agents quickly catch the coarse-grained information and then learn fine-grained attention faster. Besides, our fusion attention map holds reliable interpretability, which reveals that the SHCP attention signal can greatly assist the RL agents to attend to

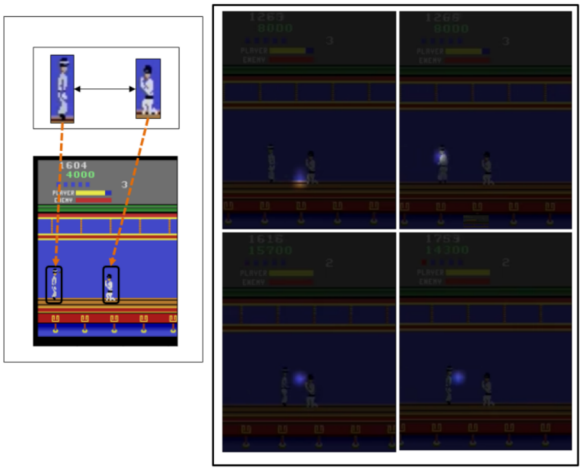

Fig. 5. Fusion attention behaviors for KungFu-Master environment. Agent focuses on the area conflicts may happen rather than "player" or "fighter" where the SHCP attention focus on.

key entities and learn better. We believe that our work reveals a potential and promising direction that combines human pre-guided attention signals to affect agents' behavior via attention mechanisms.

References

[1] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A brief survey of deep reinforcement learning," *IEEE Signal Processing Magazine*, vol. 34, no. 6, 2017.

[2] T. Osa, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, and J. Peters, "An algorithmic perspective on imitation learning," *Foundations and Trends® in Robotics*, 2018.

[3] J. Choi, B.-J. Lee, and B.-T. Zhang, "Multi-focus attention network for efficient deep reinforcement learning," *arXiv preprint arXiv:1712.04603*, 2017.

[4] S. Blakeman and D. Mareschal, "Selective particle attention: Visual feature-based attention in deep reinforcement learning," *arXiv preprint arXiv:2008.11491*, 2020.

[5] I. Sorokin, A. Seleznev, M. Pavlov, A. Fedorov, and A. Ignateva, "Deep attention recurrent q-network," *CoRR, abs/1512.01693*, 2015.

[6] L. Li, "Focus of attention in reinforcement learning," 2007.

[7] A. Mott, D. Zoran, M. Chrzanowski, D. Wierstra, and D. J. Rezende, "Towards interpretable reinforcement learning using attention augmented agents," in *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada* (H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, eds.), pp. 12329–12338, 2019.

[8] A. Manchin, E. Abbasnejad, and A. v. d. Hengel, "Reinforcement learning with attention that works: A self-supervised approach," in *International Conference on Neural Information Processing*, pp. 223–230, Springer, 2019.

[9] S. Mousavi, M. Schukat, E. Howley, A. Borji, and N. Mozayani, "Learning to predict where to look in interactive environments using deep recurrent q-learning," *arXiv preprint arXiv:1612.05753*, 2016.

[10] P. A. Tsividis, T. Pouncy, J. L. Xu, J. B. Tenenbaum, and S. J. Gershman, "Human learning in atari," in *2017 AAAI spring symposium series*, 2017.

[11] R. Zhang, C. Walshe, Z. Liu, L. Guan, K. Muller, J. Whritner, L. Zhang, M. Hayhoe, and D. Ballard, "Atari-head: Atari human eye-tracking and demonstration dataset," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, pp. 6811–6820, 2020.

[12] S. Branavan, D. Silver, and R. Barzilay, "Learning to win by reading manuals in a monte-carlo framework," *Journal of Artificial Intelligence Research*, vol. 43, pp. 661–704, 2012.

[13] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *et al.*, "Mastering the game of go without human knowledge," *nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[14] R. R. Torrado, P. Bontrager, J. Togelius, J. Liu, and D. Perez-Liebana, "Deep reinforcement learning for general video game ai," in *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, pp. 1–8, IEEE, 2018.

[15] G. Tesauro, "Temporal difference learning and td-gammon," *Communications of the ACM*, vol. 38, no. 3, pp. 58–68, 1995.

[16] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, 2016.

[17] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*, pp. 1928–1937, PMLR, 2016.

[18] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[19] M. Babaeizadeh, I. Frosio, S. Tyree, J. Clemons, and J. Kautz, "Reinforcement learning through asynchronous advantage actor-critic on a gpu," *ICLR*, 2016.

[20] Y. Wu, E. Mansimov, S. Liao, R. Grosse, and J. Ba, "Scalable trust-region method for deep reinforcement learning using kronecker-factored approximation," *NIPS*, 2017.

[21] L. Espeholt, H. Soyer, R. Munos, K. Simonyan, V. Mnih, T. Ward, Y. Doron, V. Firoiu, T. Harley, I. Dunning, *et al.*, "Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2018.

[22] M. Hessel, J. Modayil, H. Van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, and D. Silver, "Rainbow: Combining improvements in deep reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[23] X. Zhang, R. Zhang, J. Cao, D. Gong, M. You, and C. Shen, "Part-guided attention learning for vehicle instance retrieval," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[24] R. Iyer, Y. Li, H. Li, M. Lewis, R. Sundar, and K. Sycara, "Transparency and explanation in deep reinforcement learning neural networks," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 144–150, 2018.

[25] Y. Niv, R. Daniel, A. Geana, S. J. Gershman, Y. C. Leong, A. Radulescu, and R. C. Wilson, "Reinforcement learning in multidimensional environments relies on attention mechanisms," *Journal of Neuroscience*, vol. 35, no. 21, pp. 8145–8157, 2015.

[26] L. Zhang, L. Sun, L. Yu, X. Dong, J. Chen, W. Cai, C. Wang, and X. Ning, "Arface: attention-aware and regularization for face recognition with reinforcement learning," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *NIPS*, 2017.

[28] K. M. Hermann, T. Kočiskỳ, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom, "Teaching machines to read and comprehend," *In Advances in Neural Information Processing Systems*, 2015.

[29] Q. You, H. Jin, Z. Wang, C. Fang, and J. Luo, "Image captioning with semantic attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4651–4659, 2016.

[30] M. Shan and N. Atanasov, "A spatiotemporal model with visual attention for video classification," *arXiv preprint arXiv:1707.02069*, 2017.

[31] D. Rudoy, D. B. Goldman, E. Shechtman, and L. Zelnik-Manor, "Learning video saliency from human gaze using candidate selection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1147–1154, 2013.

[32] H. Liu, X. He, Y. Bai, X. Liu, Y. Wu, Y. Zhao, and H. Yang, "Nightlight as a proxy of economic indicators: Fine-grained gdp inference around chinese mainland via attention-augmented cnn from daytime satellite imagery," *Remote Sensing*, vol. 13, no. 11, p. 2067, 2021.

[33] A. Palazzi, D. Abati, F. Solera, R. Cucchiara, *et al.*, "Predicting the driver's focus of attention: the dr (eye) ve project," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1720–1733, 2018.

[34] M. Hossain, M. Hosseinzadeh, O. Chanda, and Y. Wang, "Crowd counting using scale-aware attention networks," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1280–1288, IEEE, 2019.

[35] H. Liu, Y. Liu, H. He, and H. Yang, "Lebp–language expectation & binding policy: A two-stream framework for embodied vision-and-language interaction task learning agents," *arXiv preprint arXiv:2203.04637*, 2022.

[36] L. Rong and C. Li, "Coarse-and fine-grained attention network with background-aware loss for crowd density map estimation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 3675–3684, 2021.

[37] K. R. Narasimhan, R. Barzilay, and T. S. Jaakkola, "Grounding language for transfer in deep reinforcement learning," *Journal of Artificial Intelligence Research*, 2018.