# Visual Transfer between Atari Games using Competitive Reinforcement Learning

Akshita Mittel [*] Purna Sowmya Munukutla [*]
Robotics Institute, Carnegie Mellon University
{amittel, spmunuku}@andrew.cmu.edu

## Abstract

*Modern deep Reinforcement Learning (RL) methods are highly effective at selecting optimal policies to maximize rewards. The combination of these methods with Deep Learning approaches shows promise for challenging tasks by leveraging rich visual information for policy selection [5]. In this paper, we explore the use of visual representations to transfer the knowledge of an RL agent from one domain to another. More specifically, we propose a method that can generalize for a target game using an RL agent trained for a source game in Atari 2600 environment. Instead of fine-tuning a pre-trained model for the target game, we propose a learning approach to update the model using multiple RL agents trained in parallel with different representations of the target game. The visual representations of the target game are generated by learning a visual mapping between the source game and the target game in an unsupervised manner. The visual mapping between sequences of transfer pairs has been shown to derive new representations of the target game; training on which improves the RL agent updates in terms of performance, data efficiency and stability. In order to demonstrate the effectiveness of this approach, the transfer learning procedure is evaluated on two pairs of Atari games taken in contrasting settings.*

## 1. Introduction

Convolutional neural networks (CNN) have shown remarkable performance for training RL agents to learn optimal policy from high dimensional visual inputs [5]. RL agents trained on [5] reach human-level game playing performance in real-time for Atari games. We aim to extend these efforts by developing a generalizable framework that can apply previously gained knowledge to new domains within Atari games. In our paper, a frame taken from the source game is mapped to the target game and a trained policy learned from the source game knowledge is used to play the target game.

It has been shown that human-expert performances can be achieved on several games with the model complexity of a single agent using transfer learning methods [1, 2]. Significant improvements in performance, stability and learning speeds in target games are also seen by pre-training with source game agent [1]. Building on these ideas, we propose a transfer learning approach that leverages underlying similarities between different games to represent common knowledge with Unsupervised Image-to-image Translation (UNIT) Generative adversarial networks (GANs) [3].

In this paper, we propose a competitive training strategy where the RL agent is simultaneously trained with images from the UNIT GAN and images that are queried directly from the environment. The UNIT GAN acts as a visual mapper between source, target pairs and captures underlying latent space shared across games. Thus, the two representations of target game compete for model capacity aiding transfer learning process. The RL agent is trained using Asynchronous Advantage Actor-Critic (A3C) network [4].

In this work, our contribution is two fold:

1) We leverage the multi-threaded nature of the A3C network by utilizing UNIT GAN as a visual mapper, between source and target pairs, to simultaneously train the network using multiple visual representations of a particular game.

2) We further demonstrate that training a network competitively amongst multiple representations instead of a single representation yields better performance, which has been evaluated on different transfer learning metrics.

## 2. Method

The goal of this paper is to use an RL agent to generalize between two related but vastly different pairs of Atari games like Pong, Breakout and Assault, Demon-Attack. This is done by learning visual mappers across games: given a frame from the source game, we should be able to generate the analogous frame in the target game. Building on the existence of these mappers, the training method we propose is to simultaneously learn two representations of the target game, effectively making them compete for representational space within the neural network.

[*] equal contribution

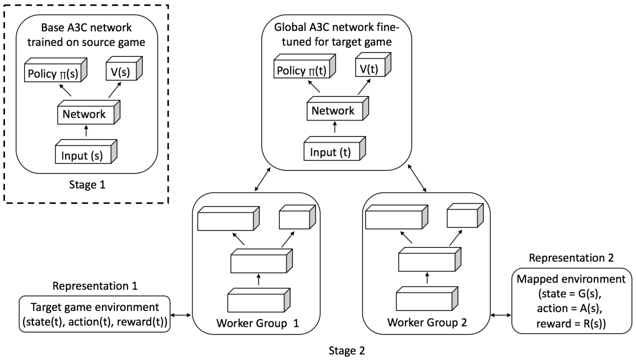
Base A3C network trained on source game
Policy π(s)
V(s)
Network
Input (s)
Stage 1
Global A3C network fine-tuned for target game
Policy π(t)
V(t)
Network
Input (t)
Representation 1
Target game environment (state(t), action(t), reward(t))
Worker Group 1
Worker Group 2
Representation 2
Mapped environment (state = G(s), action = A(s), reward = R(s))
Stage 2

Figure 1. Transfer learning process with training for the source game in stage 1 and training for the target game using two competing representations in stage 2. Representation 1 is for the target game which is queried directly from the environment and representation 2 for the target game is extracted using visual transfer of states and static mapping of actions, rewards from source game.

### 2.1. Visual mappers across Atari games

To create visual analogies between games, we rely on the core ideas explored in [2] to learn the mapping $G : s \rightarrow t$ between source game ($s$) and target game ($t$) in an unsupervised manner assuming a shared latent space between the two games. The visual mapper is trained with pre-processed frames obtained by binarizing the input after subtracting the median pixel and applying dilation operator to enlarge relevant object sizes. The frames are also rotated to ensure that main axis of motion in source and target game is horizontal. The pre-processed frames from source game and target game are mapped to the same latent representation in a shared latent space across both games. We assume that a pair of corresponding images in two different domains of source and target game can be inferred by learning two encoding functions, to map images to latent codes and two generation functions that map latent codes to images. The visual mapper is implemented with the network architecture of UNIT GAN [3] with unit consistency loss. The encoding and generating functions are implemented using CNNs and the shared-latent space assumption is enforced with a weight sharing constraint across these functions. Adversarial discriminators for the respective domains are also trained to evaluate whether the translated images are realistic.

### 2.2. Transfer Learning Method

One of the challenges addressed in this paper is to prove or disprove that visual analogies across games are necessary and sufficient to transfer the knowledge of playing one particular game to another.

In the recent times, policy gradient methods like A3C [4] are shown to be extremely effective in learning the world of Atari games. They are currently set as the baseline for Atari games like Pong, Breakout, Seaquest in terms of training time vs rewards. The idea of A3C networks is to use multiple workers in parallel that can each interact with the environment and update the shared model simultaneously. Thus, A3C networks asynchronously exchange multiple agents in parallel instead of using experience replay.

We use the baseline A3C network trained for source game in the first stage of our training process and transfer the knowledge from this model to learning to play target game. We measure the efficiency of transfer learning method in terms of training time and data efficiency across parallel actor-learners. In the second stage of training process, we use two representations of the target game amongst the workers in parallel. The first representation of transfer process uses the target game frames taken directly from the environment. The second representation uses the frames learned from the visual mapper i.e., $G(S)$. This paper further leverages the multi-threaded nature of A3C architecture, by providing different representations of a game to different subsets of workers and letting them train competitively. We show that by providing a subset of A3C workers with visual representations generated by the visual mapper, the network is able to learn the target domain significantly faster. The ratio of number of workers that train directly on frames queried from the target game and frames mapped from the source game is a hyper-parameter that is determined through experimentation.

In this way, we transfer the knowledge from source game to target game by competitively and simultaneously fine-tuning the model using two different visual representations of the target game. We have experimented with pairs of games like Pong, Breakout and Assualt, Demon-Attack in this paper. Since Pong and Breakout have similar game strategies of controlling a paddle to hit the ball to obtain a certain objective, it is intuitive to determine a meaningful static mapping of actions. The six discrete actions of Pong $\{$No Operation, Fire, Right, Left, Right Fire, Left Fire$\}$ is mapped to four actions of Breakout as $\{$Fire, Fire, Right, Left, Right, Left$\}$ respectively. The rewards are mapped directly from source game to target game without any scaling. It can be seen that multiple Atari games like Pong, Breakout, Tennis share similar objectives and the transfer learning process is shown on a subset of these games. Since the strategies of these games are very similar, it can be intuitively reasoned that this transfer learning method can be extended to other sets of games.

## 3. Results

From Figure 2, we can conclude that not only is it possible to transfer the learning across two domains, it is also possible to speed up the learning process using our method. In Figure 2(b), we observe that the worker configuration 3:1 has a greater speed-up than all other configurations. This particular configuration also has better performance for other metrics like total rewards and transfer ratio as seen in

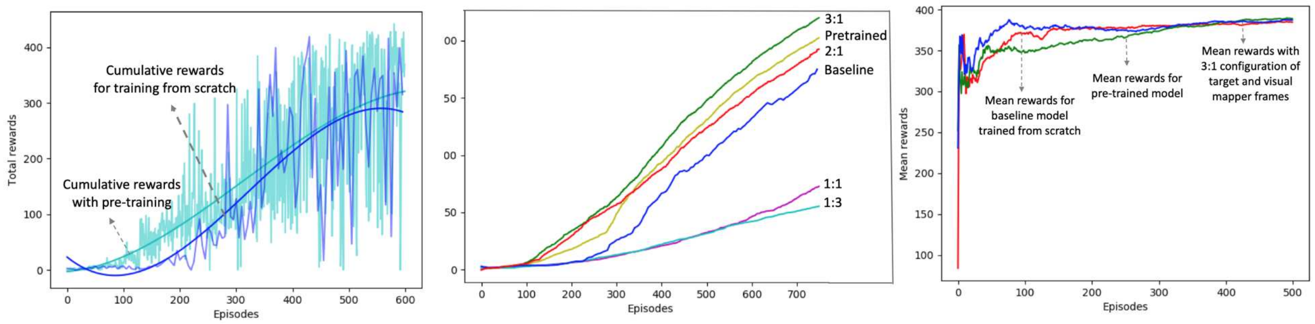

Figure 2. (a) The figure depicts the total reward per episode versus number of epochs for transfer learning between Pong to Breakout. Here the blue curve indicates the behaviour without pre-training and the cyan curve indicates the behaviour after pre-training. The smoothed out curves show the relative behaviour between the two models, indicating that on an average the cumulative rewards obtained after pre-training are better than training from scratch. (b) The figure depicts the mean reward per episode versus number of epochs for Pong to Breakout. The graph depicts the performance of the baseline (blue), pretrained model (yellow), and the various visual-mapper model configurations {3:1, 2:1, 1:1, 1:3}. Here we see that the model with the worker-configuration 3:1 (green) outperforms baseline and pre-trained model indicated in blue and yellow. (c) The figure depicts the mean reward per episode versus number of epochs for Demon-Attack to Assault. The graph depicts the performance of the baseline (red), pre-trained model (green), and the visual-mapper model with configurations 3:1 (blue). Here also, we see that the model with the worker-configuration 3:1 (blue) outperforms both the baselines indicated in green and red.

| Worker Configurations | Jumpstart | Epoch to Threshold | Total Rewards | Transfer Ratio | Jumpstart | Epoch to Threshold | Total Rewards | Transfer Ratio |
|---|---|---|---|---|---|---|---|---|
| Target to Source Game | Breakout to Pong | | | | Assault to Demon-Attack | | | |
| Baseline Network | - | 435 | 47960 | - | - | - | 185745 | - |
| Pre-trained Network | - | 437 | 63955 | 1.334 | 231 | - | 183035 | 0.985 |
| 3:1 (Mappers vs Target) | - | 357 | **74932** | **1.562** | **252** | - | **188380** | **1.014** |
| 2:1 (Mappers vs Target) | - | **319** | 65376 | 1.363 | - | - | - | - |
| 1:1 (Mappers vs Target) | - | 746 | 18400 | 0.384 | - | - | - | - |
| 1:3 (Mappers vs Target) | - | 872 | 17403 | 0.355 | - | - | - | - |

Table 1. Evaluation metrics [6] across different experiment settings with worker configurations (frames taken from native Atari target game vs frames generated from source game using visual mappers). The metrics evaluated are as follows: (a) *Jumpstart*: This compares the initial performance of an agent in the transfer learning task. (b) *Epoch to threshold*: It measures the time taken to reach a particular level of performance. Though not a standard threshold, the threshold for learning to play Breakout for this particular experiment is kept at a cumulative reward of 400. We show that introducing a competitive setting amongst the workers obtains the best epoch to threshold. (c) *Total Rewards*: The total rewards are essentially the area under the graph of mean reward per episode vs total number of episodes for each model. It can be inferred that the agents trained with a higher proportion of native workers have higher total rewards. (d) *Transfer Ratio*: The transfer ratio is the ratio of the total rewards obtained from the transfer learning experiment when compared to the baseline. Its magnitude specifies the extent of efficient knowledge transfer, where a value of more than one signifies transfer of knowledge.

Table 1 showing the effectiveness of our approach. We have also trained a model using only single form of representation that has been derived from visual mapper of source game. The model has been trained for over 300 epochs and gained only little improvement in learning the expert policy of target game. Though it did not have significant performance on its own, it acts as stabilizer to the transfer process.

## References

[1] C. E. Chaitanya Asawa and D. Pan. Using transfer learning between games to improve deep reinforcement learning performance and stability, 2017. 1

[2] L. W. Doron Sobol and Y. Taigman. Visual analogies between atari games for studying transfer learning in rl, 2018. 1, 2

[3] M. Liu, T. Breuel, and J. Kautz. Unsupervised image-to-image translation networks. *CoRR*, abs/1703.00848, 2017. 1, 2

[4] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu. Asynchronous methods for deep reinforcement learning. *CoRR*, abs/1602.01783, 2016. 1, 2

[5] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013. 1

[6] M. E. Taylor and P. Stone. Transfer learning for reinforcement learning domains: A survey. *J. Mach. Learn. Res.*, 10:1633–1685, Dec. 2009. 3