Transfer Learning for Related Reinforcement Learning Tasks via Image-to-Image Translation

Shani Gamrian
Yoav Goldberg
Bar-Ilan University

ICML 2019, Long Beach
Deep Reinforcement Learning
Transfer Learning

- Deep Reinforcement Learning is effective but fails to generalize.

Can we **TRANSFER** knowledge between related RL tasks?
Generalization Failures of Deep-RL Breakout

Figure 1: Various variations of the Breakout game: (a) Standard version, (b) A Constant Rectangle, (c) A Moving Square, (d) Green Lines, (e) Diagonals.
Generalization Failures of Deep-RL Transfer Learning via Finetuning

The results show that fine-tuning takes as long or longer than training from scratch!

Figure 2: A comparison between the different baselines on Breakout. The y-axis on each one of the plots shows the average reward per episode of Breakout during training. The x-axis shows the total number of training epochs where an epoch corresponds to 1 million frames. The plots are averaged on 3 runs with different random seeds.
Problem: finetuning fails to transfer between related tasks.

Our Solution: Transfer by visual mapping.

How?: map the input images from the target task to the source task.
UNsupervised Image-to-Image Translation (UNIT)

Generative Adversarial Networks (GANs)

https://deeplearning4j.org/generative-adversarial-network
We initialize the layers with the values of the trained network.

We run the game and translate each image from the target task to source task.

Our model accuracy is the score of the game.
Analogy-based Zero-Shot Transfer with GANs

Breakout

<table>
<thead>
<tr>
<th>Source</th>
<th>Frames</th>
<th>Score</th>
<th>Target</th>
<th>Frames</th>
<th>Score</th>
<th>Target with GANs</th>
<th>Frames</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Constant Rectangle</td>
<td>43M</td>
<td>302</td>
<td></td>
<td>122</td>
<td>0</td>
<td>260K</td>
<td>362</td>
<td></td>
</tr>
<tr>
<td>A Moving Square</td>
<td>43M</td>
<td>302</td>
<td></td>
<td>100</td>
<td>0</td>
<td>384K</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Green Lines</td>
<td>43M</td>
<td>302</td>
<td></td>
<td>186</td>
<td>2</td>
<td>288K</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Diagonals</td>
<td>43M</td>
<td>302</td>
<td></td>
<td>100</td>
<td>0</td>
<td>383K</td>
<td>330</td>
<td></td>
</tr>
</tbody>
</table>

Our method is 100x more data efficient than training from scratch!
Road Fighter
Analogy-based Zero-Shot Transfer with GANs

Road Fighter
# Analogy-based Zero-Shot Transfer with GANs

## Road Fighter

<table>
<thead>
<tr>
<th>Level</th>
<th>Score (no transfer)</th>
<th>Score (analogy transfer)</th>
<th># Frames (analogy)</th>
<th># Frames (from scratch)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2</td>
<td>0</td>
<td>5350</td>
<td>250K</td>
<td>12.4M</td>
</tr>
<tr>
<td>Level 3</td>
<td>0</td>
<td>5350</td>
<td>250K</td>
<td>31M</td>
</tr>
<tr>
<td>Level 4</td>
<td>0</td>
<td>2050</td>
<td>250K</td>
<td>13.6M</td>
</tr>
</tbody>
</table>
Accelerating RL with Imitation Learning

- Our transfer method is limited by the imperfect GAN generation and generalization abilities.

- We propose to use the visual-transfer based policy as imperfect demonstrations.

- We combine off-policy supervised updates and on-policy RL updates to accelerate the training process.

- We apply this method on Road Fighter.
Road Fighter
Accelerating RL with Imitation Learning
Road Fighter Results

With transfer + imitation learning, agent manages to complete the levels with just 20% of the needed frames.