Is Self-Supervision Helpful? Exploring Self-Supervision in Reinforcement Learning and Classification

Prabhat Agarwal, Abhishek Sinha and Kumar Ayush

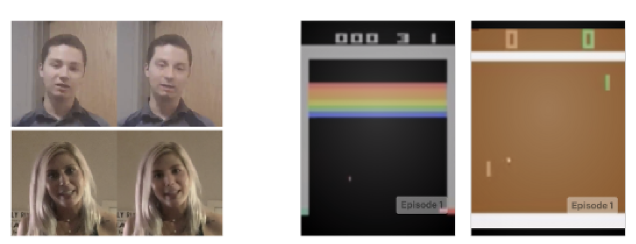
Department of Computer Science, Stanford University

Introduction

- **Motivation**: Human labels are expensive to collect and hard to scale up. Thus, there has been an increasing research interest to investigate learning representations from unlabeled data. In particular, the image itself already contains structural information which can be utilized. Self-supervised learning approaches attempt to capture this.
- **Problem Definition**: Self-supervised tasks such as rotation task loss as an auxiliary loss function have shown to improve classification performance across a range of classification datasets [4]. Self-supervised representation learning to extract useful information when building classifiers or other predictors have also shown promise. In this work, we explore the benefits of such self-supervision in two classes of tasks: (a) **reinforcement learning**, (b) **classification** (e.g. detection of forgery in manipulated facial images).

Dataset

- **Reinforcement Learning**: We perform our experiments on two Atari Games - Pong and Breakout. We use the open AI gym simulator for the game.
- **Classification**: We sample frames from real (368) and manipulated videos (400) from Deep Fake Detection Challenge (DFDC) [1] dataset . We consider two train/test splits: (a) Random Split: Real or manipulated frames of a video may be in both train and test, (b) Hard Split: All frames of a video are either in train or test.



Modelling

- In addition to supervised losses, we consider self-supervised losses based on labeled data $x \to (\hat{x}, \hat{y})$ that can be systematically derived from inputs x alone.
- We augment standard supervised losses with those derived from self-supervised tasks to improve representation learning.
- We also experiment with building classifiers over data representations obtained using a Bidirectional GAN (trained with ImageNet) for DeepFake Classification.

Classification: We use two kinds of self-supervision, (a) Rotation Task (See Fig. 3) as an auxiliary task in feed-forward CNNs (input image x is rotated by an angle $\theta \in \{0°, 90°, 180°, 270°\}$ to obtain $\hat{x}$ and the target label $\hat{y}$ is the index of the angle), (b) Self-Supervised representation learning using Bidirectional GAN [2] (trained with ImageNet) and training a classifier over the learned features (See Fig. 2).

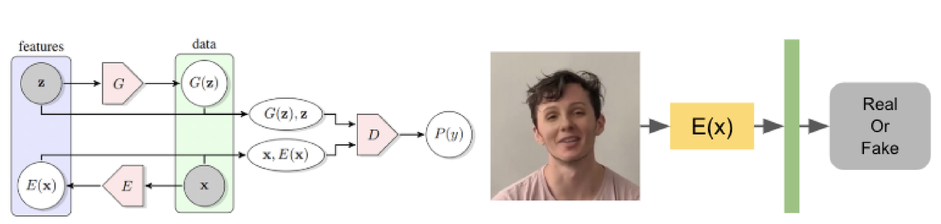


Fig. 2: Self-Supervised representation learning using Bidirectional GAN (pre-trained with Imagenet)

Reinforcement Learning: We use two kinds of self-supervision, (a) Rotation Task, and (b) Sequence Prediction (an auxiliary task where sequence of the four input frames, $x = [f_1, f_2, f_3, f_4]$, is permuted to obtain $\hat{x}$ and the network is tasked to identify the class of the permutation).
We incorporate self-supervision in Dueling Deep Q-Learning [5] algorithm to estimate $Q(s, a)$ value for a (state, action) pair.

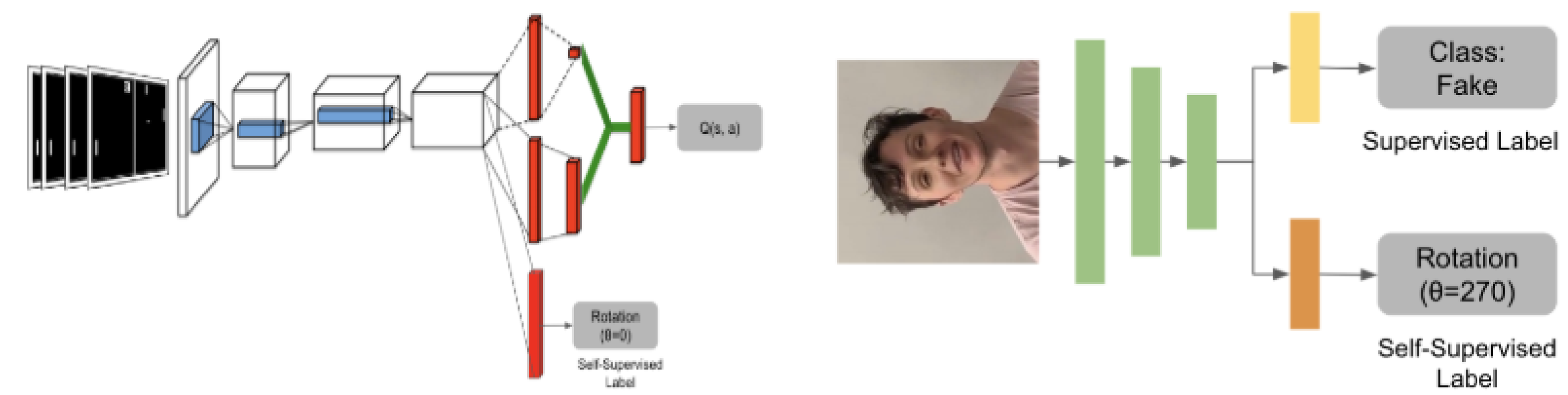


Fig. 3: Self-supervision in Pong and DeepFake Classification with angle of rotation prediction

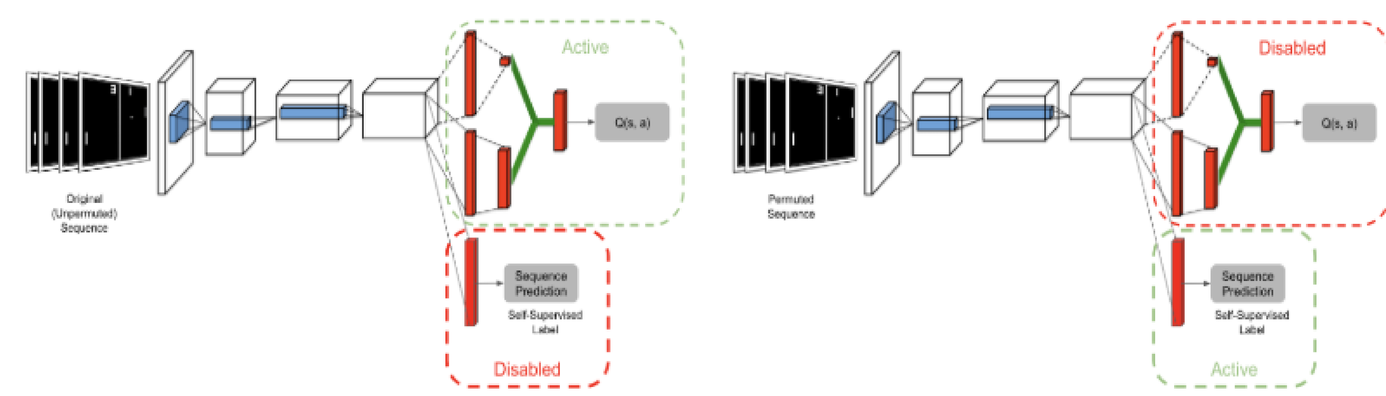


Fig. 4: Self-supervision in Pong using sequence prediction. We alternate between $Q(s, a)$ prediction and sequence prediction during training.

Results

We train classifiers for DeepFake Classification using different backbones and weight initializations. We also generate the attribution maps, using SmoothGrad [3], which computes map by performing squared average of the gradients with respect to the input for various noisy inputs. It can be seen that the model has learnt to identify relevant regions in the input image.

Model	Random Split	Hard Split
ResNet-50 (Random Initialization)	56.7	56.2
ResNet-50 (ImageNet pre-trained)	96.1	91.1
AlexNet (ImageNet pre-trained)	81.5	73.5
AlexNet (Backbone trained with ImageNet & Rotation)	75.8	71.1
Classifier over BigBiGAN (ResNet-50 Enc.)	92.3	70.8
Classifier over BigBiGAN (RevNet-50X4 Enc.)	96.5	92.7

TABLE 1: Binary Classification Accuracy for DeepFake Detection using different methods.

Agent	Pong	Breakout
Random	-21	1.25
DQN	20.0	9.9
Human	-3	30.5
World Record	21.0	864
Rotation	-20.1	7.6
Sequence	18.2	14.9

TABLE 2: Mean Score in 100 episodes for Pong and Breakout for various agents.

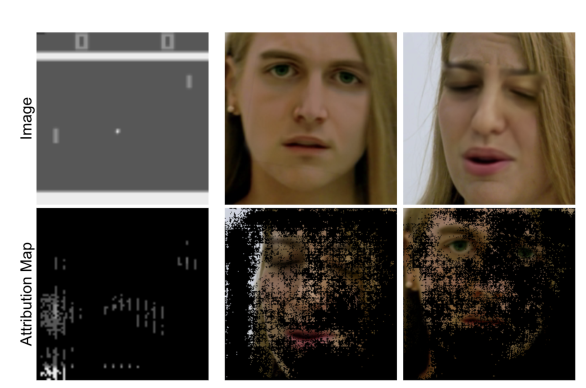


Fig. 5: Attribution map generated using SmoothGrad [3]

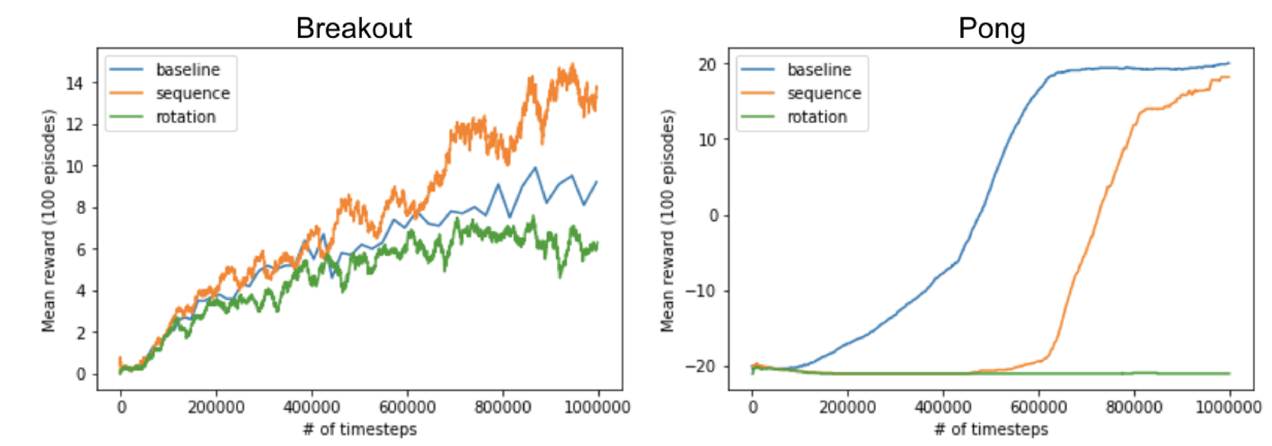


Fig. 6: Mean Reward with number of episodes

Key Observations and Discussion

- **Classification**: Even though the BigBiGAN is trained using ImageNet, a classifier trained over features from BigBiGAN encoder performs with high accuracy. This shows that self-supervised representation learning extracts useful features which can withstand domain shifts in datasets. Contrary to our expectations, rotation task does not seem to perform well (See TABLE I row 3 and 4).
- **Reinforcement Learning**: Results show that agents trained with sequence prediction self-supervision leads to learning of good features, and augmenting it with the Q-learning loss can lead to faster learning. We also see that rotation task does not provide any benefit, rather it deteriorated the performance. We believe this is because the rotation task is very easy for the model as it can be predicted by looking only at the pedal or score bar.

Self-supervised losses act as a data-dependent regularizer for representation learning.

Challenges and Future Work

- The classifier trained over BigBiGAN features shows erratic behaviour during training. Some runs resulted in divergent training and the classifier getting stuck at 50% accuracy.
- We aim to explore other techniques such as jigsaw puzzle task, image colorization, for image-based self-supervision and predicting the arrow of time for sequence-based self-supervision.

References

[1] Brian Dolhansky et al. “The Deepfake Detection Challenge (DFDC) Preview Dataset”. In: *arXiv preprint arXiv:1910.08854* (2019).

[2] Jeff Donahue and Karen Simonyan. “Large scale adversarial representation learning”. In: *Advances in Neural Information Processing Systems*. 2019, pp. 10541–10551.

[3] Daniel Smilkov et al. “Smoothgrad: removing noise by adding noise”. In: *arXiv preprint arXiv:1706.03825* (2017).

[4] Jong-Chyi Su, Subhransu Maji, and Bharath Hariharan. “Boosting Supervision with Self-Supervision for Few-shot Learning”. In: *arXiv preprint arXiv:1906.07079* (2019).

[5] Ziyu Wang et al. “Dueling network architectures for deep reinforcement learning”. In: *arXiv preprint arXiv:1511.06581* (2015).