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

Prabhat Agarwal, Abhishek Sinha, Kumar Ayush
{prabhat8, a7b23, kayush}@cs.stanford.edu

I. Introduction

Human labels are expensive to collect and hard to scale up. To this end, there has been 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. Self-supervised tasks such as rotation task loss, jigsaw puzzle task loss as auxiliary loss functions have shown to improve classification performance across a range of classification datasets. In this work, we aim to explore the benefits of such self-supervision in two classes of tasks: (a) off-policy reinforcement learning, (b) classification task (e.g. detection of forgery in manipulated facial images).

II. Task Definition

In addition to supervised losses, we will consider self-supervised losses based on labeled data $x \rightarrow (\hat{x}, \hat{y})$ that can be systematically derived from inputs x alone. We will consider self-supervised tasks like jigsaw puzzle task, rotation task, or colorization task in our work. The jigsaw task rearranges the input image and uses the index of the permutation as the target label, while the rotation task uses the angle of the rotated image as the target label. A separate function h is used to predict these labels from the shared feature backbone f. Our final loss function will combine the two losses and thus the self-supervised losses act as a data-dependent regularizer for representation learning.

We will incorporate the aforementioned self-supervision in (a) off-policy reinforcement learning, and (b) classification such as detection of forgery in manipulated facial images. An example of self-supervision in each class of task is shown in Fig 1 and 2.

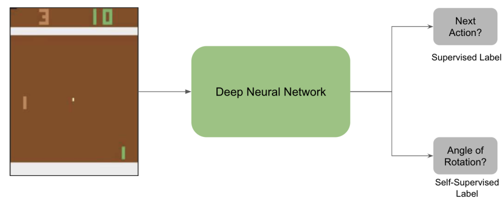

Fig. 1: Self-supervision in Pong with angle of rotation prediction

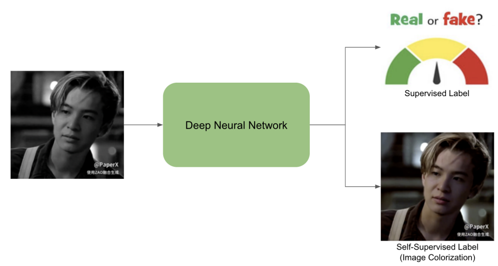

Fig. 2: Self-supervision in classification with image colorization

III. Related Work

There is a rich line of work on self-supervised learning. One class of methods removes part of the visual data (e.g., color information) and tasks the network with predicting what has been removed from the rest (e.g., greyscale images) in a discriminative manner [1], [2]. Su et al. [3] showed that self-supervision improves transferability of representations for few-shot learning tasks on a range of image classification datasets. They introduced self-supervised tasks such as rotation task loss, jigsaw puzzle task loss as auxiliary loss functions.

IV. Dataset

Classification: In order to develop state-of-the-art forgery detection method tailored to facial manipulations, a large-scale dataset of manipulations based on the classical computer graphics-based methods Face2Face [4] and FaceSwap as well as the learning-based approaches DeepFakes and NeuralTextures has been released.

Reinforcement Learning: We will use different Atari Games such as Pong, Pacman, Space Invaders, etc.

V. Evaluation Metric

Classification: We will be using the following metrics for evaluation: Binary Detection Accuracy, Precision, Recall and F-Score.

Reinforcement Learning: We will use Mean Score in 100 episodes for evaluating the performance of the trained RL agent.

VI. Baseline and Oracle

The baselines and oracle depend upon the type of task where we are applying self-supervision. We present baseline and oracle for an example of two types of task below.

A. Reinforcement Learning: Atari Games

1) Baseline: Two baselines have been explored for playing Pong (an Atari game), one using a non deep learning method and one using state-of-the-art Deep Q-Learning. The baselines are described below and the evaluation results shown in Table I.

- **Random Agent**: The simplest baseline is an agent taking random actions. At every game state, the agent samples an action uniformly.
- **Dueling DQN**: We use state-of-the-art dueling deep Q learning algorithm to estimate Q value for a state. A deep convolutional network consisting of 3 CONV layers and 3 FC layers is used to get the Q values of a given input state image.

2) Oracle: In game playing tasks, the oracle is generally a human and we measure the performance with respect to a human. The average human mean score and the world record score are shown in Table I

Agent	Mean score in 100 episodes
Random	-21
DQN	19.3
Human [5]	-3
World Record [6]	21.0

TABLE I: Mean score in 100 episodes for Pong for various agents

B. Classification Tasks: FakeChallenge++

1) Baseline: For the purpose of baselines, we evaluate two approaches known from the literature to solve the classification task of detecting facial forgery. The results are specified in Table II.

2) Oracle: The oracle for this task is the human accuracy on detecting forged video/images in an actual scenario. We utilize the user study accuracy from [7] which is presented in Table II. The participants in the study were ordinary people who were given 2-6 seconds to analyse the image and provide the judgement. This mimics the general setting where users might encounter such fake images on social media. A stronger oracle would be annotating the test set using forensic experts but that is too costly and out of scope of the project.

Method	Accuracy
XceptionNet [8]	82.01
Steganalysis Features + SVM [9]	97.63
Human [7]	68.69

TABLE II: Binary detection accuracy of various methods.

VII. Conclusion

This project aims to augment standard supervised losses with those derived from self-supervised tasks to improve representation learning in reinforcement learning and classification. Our work will be a stepping stone for future research in the domain of using self-supervision in Reinforcement Learning and Classification.

REFERENCES

[1] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. In *European Conference on Computer Vision*, pages 577–593. Springer, 2016.

[2] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016.

[3] Jong-Chyi Su, Subhransu Maji, and Bharath Hariharan. Boosting supervision with self-supervision for few-shot learning. *arXiv preprint arXiv:1906.07079*, 2019.

[4] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2387–2395, 2016.

[5] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[6] Marin Toromanoff, Emilie Wirbel, and Fabien Moutarde. Is deep reinforcement learning really superhuman on atari? *arXiv preprint arXiv:1908.04683*, 2019.

[7] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: Learning to detect manipulated facial images. *arXiv preprint arXiv:1901.08971*, 2019.

[8] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[9] Jessica Fridrich and Jan Kodovsky. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, 2012.