Reinforcement Learning (CS60077) Term Paper

IMAGE AUGMENTATION AND AUXILIARY LOSS DUO

Team Reinforced Minds:

Faraaz Mallick

Dewang Modi

Amatya Sharma

5th Year DD CSE. IIT Kharagpur

Why learn from Images?

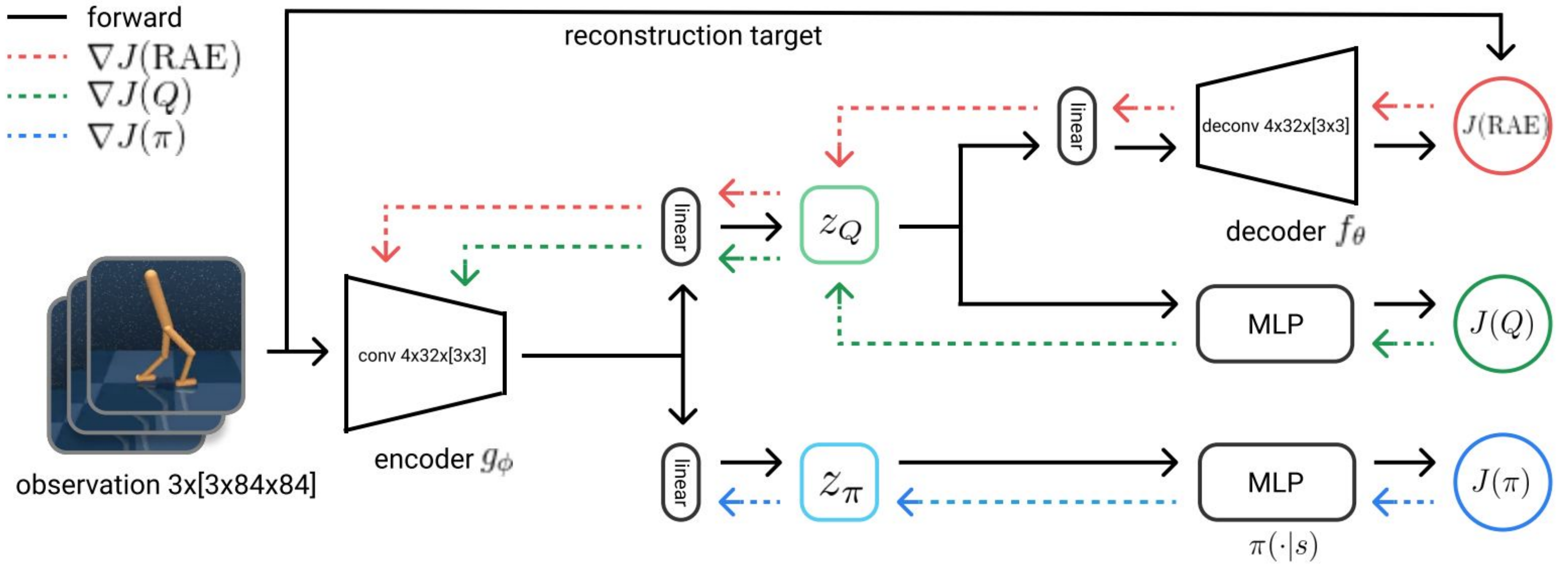
- ➢ Use of images is common (Why? cameras, easy to capture state)

Approach	|	Drawback
➢ Directly learn from Images	← | →	Requires High Dim Data
➢ Learn Latent Representations using AE	← | →	Sample Inefficient
➢ Image Reconstruction Loss (in Off Policy)	← | →	Training Instability

Prior Work

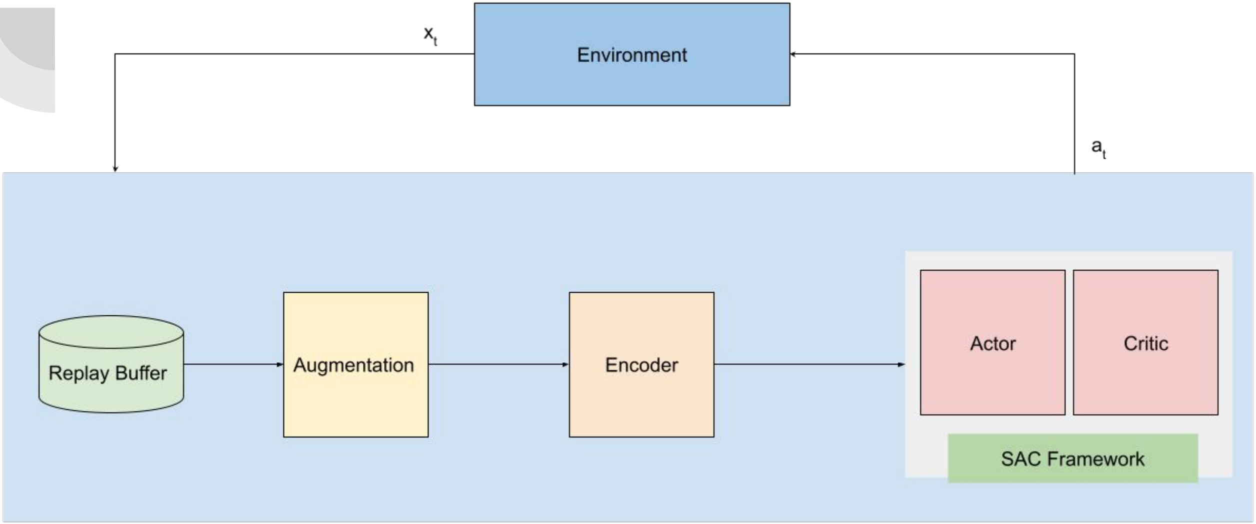
SAC+AE	DRQ	DRQ v2
➢ **Soft-Actor Critic** ➢ Regularized Auto-**Encoder** ➢ Image reconstruction loss	➢ **Soft-Actor Critic** ➢ **Image Augmentation**	➢ **DDPG** instead of SAC ➢ Image Augmentation

SAC+AE



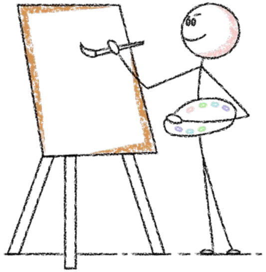
- ➢ VAE => divergence & instability
- ➢ Jointly learns Latent Representations & Policy
- ➢ At-par with model-based algorithms
- ➢ Sample efficient

DRQ

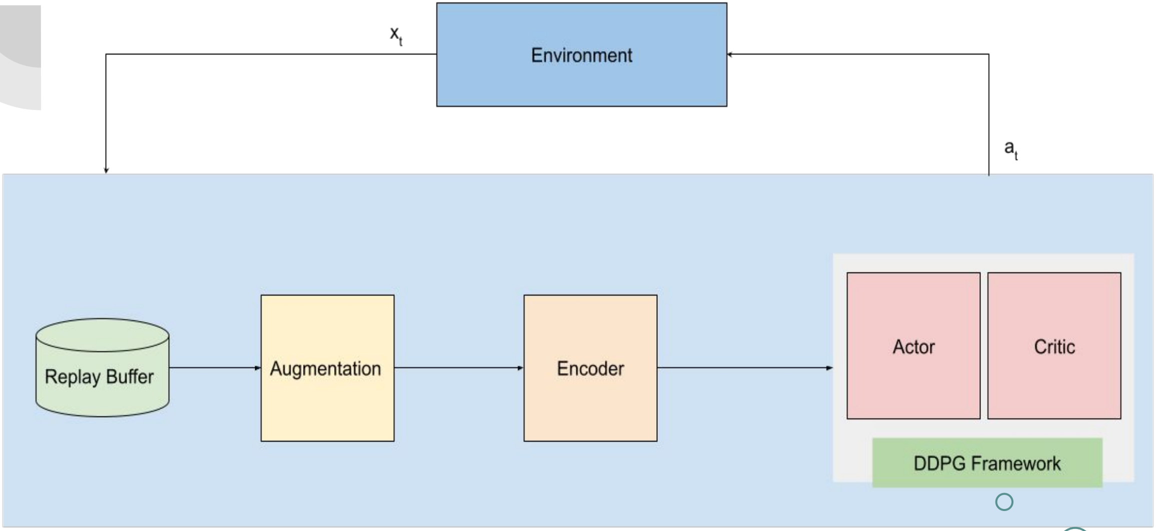


can be combined with any model-free reinforcement learning algorithm

- ➢ Introduced the use of **image augmentation** with SAC
- ➢ **No decoder** or image reconstruction loss
- ➢ Choice of augmentation - **Random shifts**



DRQ:v2



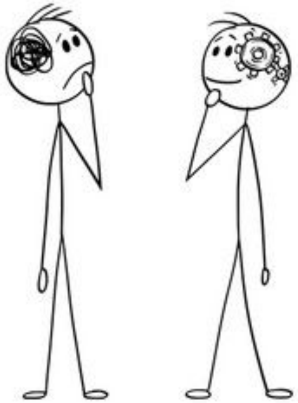
- ➢ Added **bilinear interpolation**
- ➢ Improved the hyperparams
- ➢ Changed the algorithm to **DDPG**

Why DDPG > SAC?

- **N-step return**
- **Automatic Entropy Adjustment**

Our Approaches

Reproduce SAC+AE



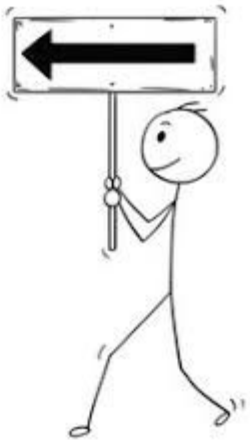
The *AD*-Model

Our Approach 1: Reproduce SAC+AE

- ➢ Test the reproducibility of the SAC+AE model
- ➢ Submit our findings to ML Reproducibility Challenge 2021
- ➢ Setup:
    - ○ Same model structure and value of hyperparameters as considered in the original paper
    - ○ 104GB RAM 1xTesla T4 GPU

Our Approach 2: The *AD*-Model

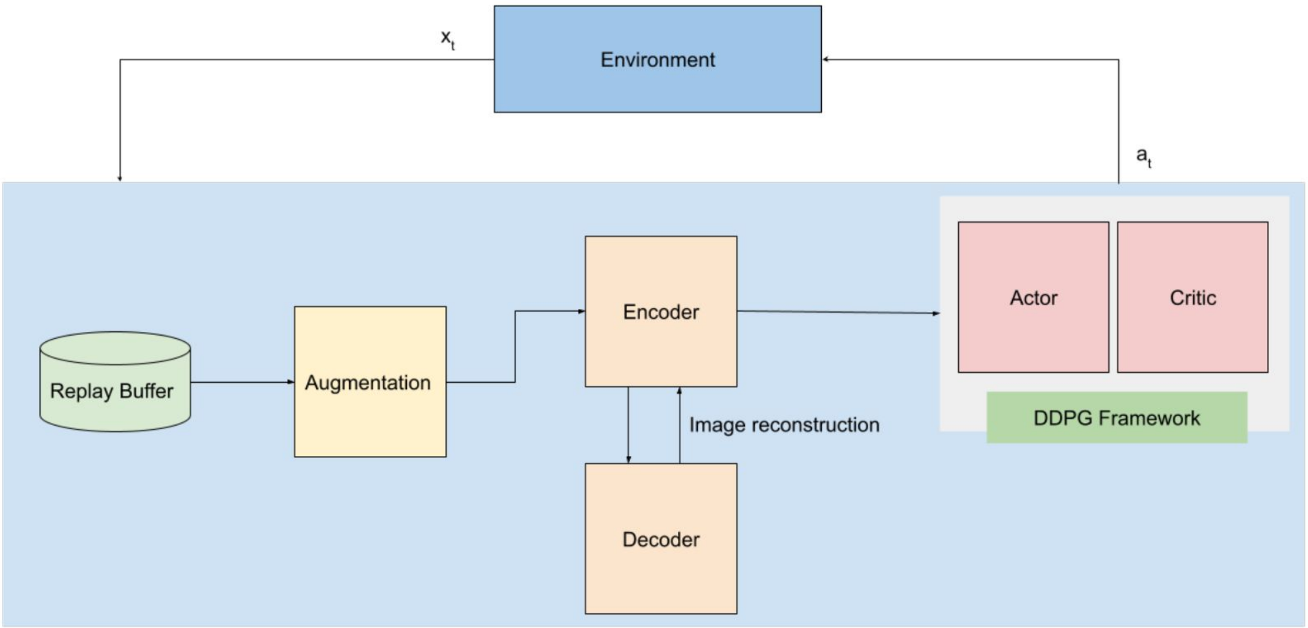
- Combine
  - Image reconstruction loss using decoder (D)
  - Image augmentation (A)



- Test the implementation in Mujoco Env on *Walker-stand* task

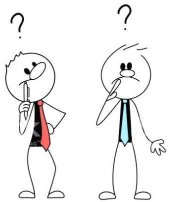
  Deepmind Control suite

AD-Model

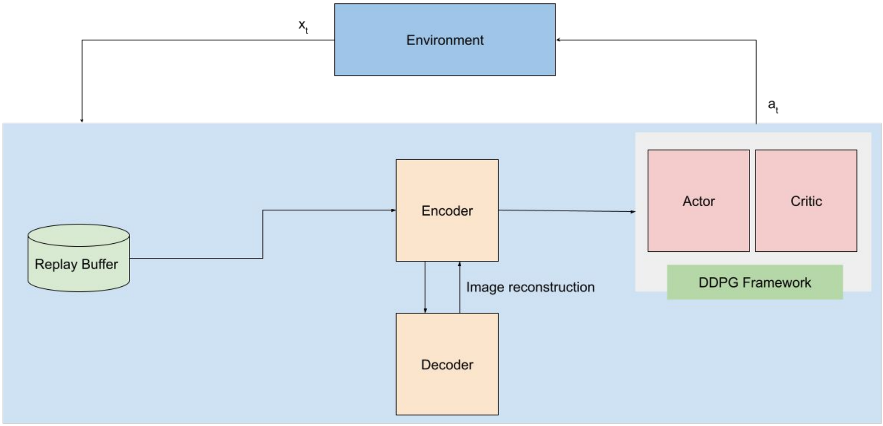


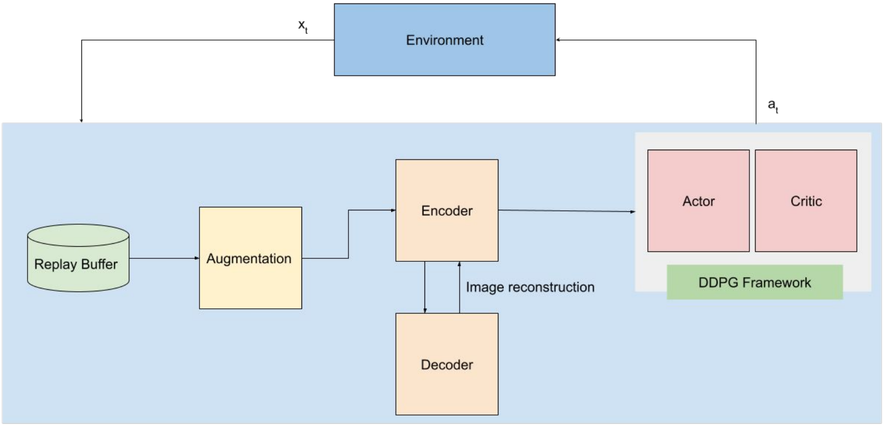
- Replay Buffer
- **Augmentation Unit**
- Encoder
- **Decoder**
- Actor and Critic as per DDPG algorithm

The four variations



- **AD(0,0):** No augmentation, no reconstruction loss
- **AD(0,1):** No augmentation, added reconstruction loss
- **AD(1,0):** Added augmentation, no reconstruction loss
- **AD(1,1):** Added augmentation, added reconstruction loss





Loss functions

- ➢ **Critic** loss

$$\mathcal{L}_{\theta_k,\xi}(\mathcal{D}) = \mathbb{E}_{\tau\sim\mathcal{D}}\big[(Q_{\theta_k}(\boldsymbol{h}_t, \boldsymbol{a}_t) - y)^2\big] \quad \forall k \in \{1, 2\}$$

- ➢ **Actor** loss

$$\mathcal{L}_{\phi}(\mathcal{D}) = -\mathbb{E}_{\boldsymbol{x}_t\sim\mathcal{D}}\big[\min_{k=1,2} Q_{\theta_k}(\boldsymbol{h}_t, \boldsymbol{a}_t)\big]$$

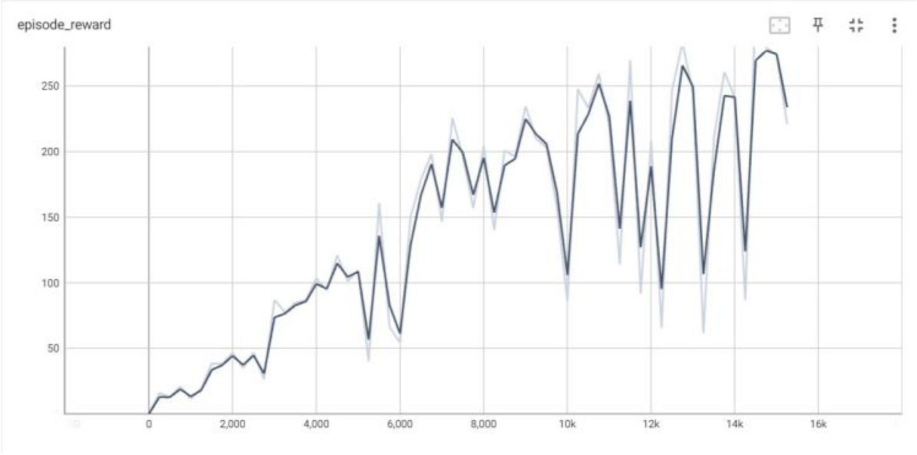
- ➢ **Image Reconstruction** Loss

$$\mathcal{L}_{AE}(\mathcal{D}) = \mathbb{E}_{\rtimes_t\sim\mathcal{D}}\big[\mathrm{MSE}(\mathbf{o}_t, \mathbf{z}_t) + \lambda_{\mathbf{z}}||\mathbf{z}_t||^2\big]$$

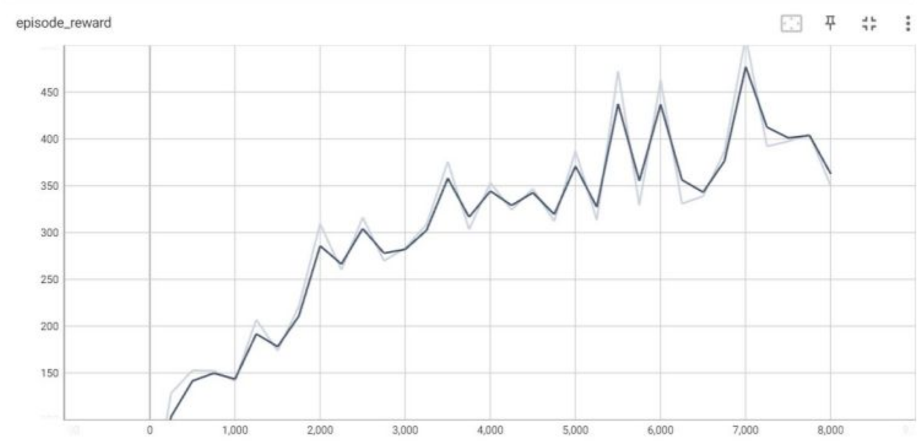
Results

Results1

- ➢ **Buggy environment** creation
- ➢ **Logic** aptly **reproduced**
- ➢ **Experimental setup** on GCP with 16 cores, 104 GB RAM and 1x Tesla T4 GPU
- ➢ **Train Time**: 12 hours for 16,000 training steps.



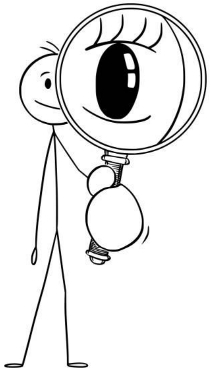
(a) SAC+AE Cheetah Run

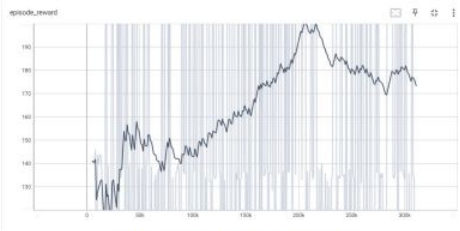


(b) SAC+AE Walker Walk

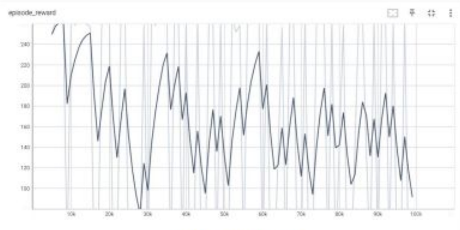
Results2

- ➢ ***AD(1,0)*** model **outperforms** in the Walker Stand task
- ➢ ***AD(0,1)*** results are **negative**, average episode reward tends to decrease
- ➢ ***AD(0,0)*** **increases** but optimal is not attained
- ➢ ***AD(1,1)*** average episode reward **oscillates.** (in the end a small peak is observable)

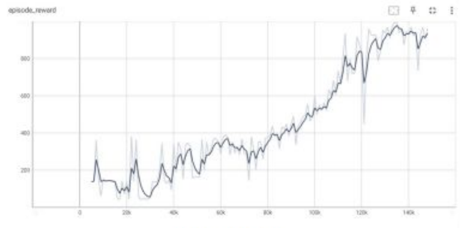




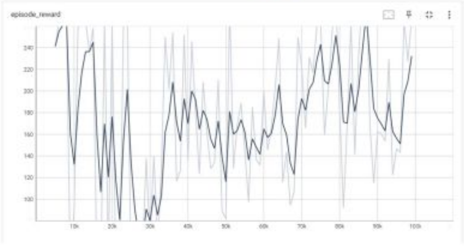
(a) No Image Aug or Recon Loss



(b) Recon Loss



(c) Image Aug



(d) Image Aug + Recon Loss

Explanation

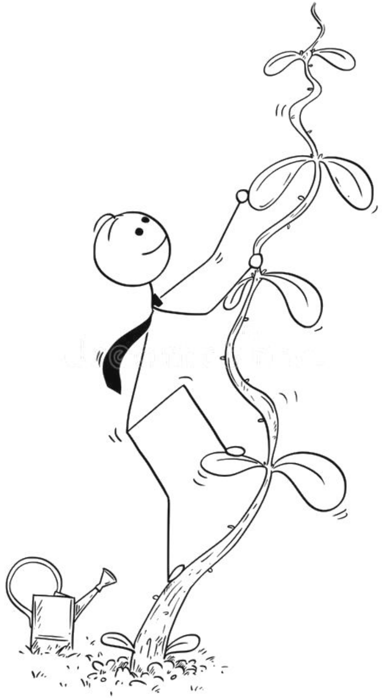
- ➢ **Conflicting effects**

  Image augmentation
  => similar latent vectors for augmented images
  => *effect on Decoder*

- ➢ Walker_Stand is **Easy** ← | → Models is **Complex**

- ➢ Limitations in computing power
  - Limited to 99k frame steps
  - May perform with more training

Future Directions



- ➢ **Contrastive** learning
- ➢ Composite **image augmentations**
- ➢ Robustness under background **noise**

Q&A