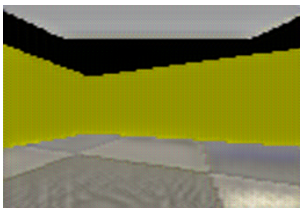
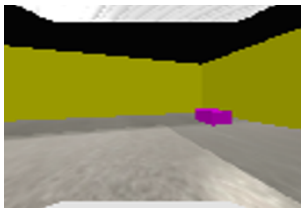
(a) Real image in Source



(b) Generated image in target

Fig. 1: Example of source → target image translation for Experiment 1.



(a) Real image in target

(b) Generated image in source

Fig. 2: Example of target → source image translation for Experiment 1.

APPENDIX

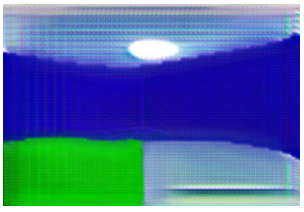
A. CycleGAN translation

CycleGANs are an extension of GANs that learn a mapping between two image domains without requiring paired datapoints in the two domains. This is beneficial for Sim2Real Transfer as obtaining paired observation matches is difficult when there are physical and semantic differences in the simulation and reality.

Here, we illustrate some examples of image translation generated by CycleGAN between the source and the target domains. Fig 1 and Fig 2 highlight CycleGAN image translations for experiment 1. Similarly, Fig 3 and Fig 4 highlight CycleGAN image translations for experiment 3.

From the images, it can be seen that CycleGAN's image translation works well for sim2sim translation. In case of sim2real translation for experiment 3, we see that the differences in the environment make it harder to obtain accurate translation. However, the experiments for goal reaching task obtained a 90% accuracy.

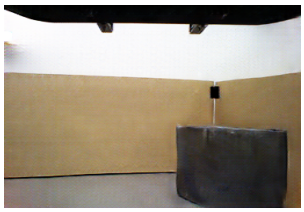
(a) Real image in target



(b) Generated image in source

Fig. 3: Example of target → source image translation for Experiment 3.

(a) Real image in source



(b) Generated image in target

Fig. 4: Example of source → target image translation for Experiment 3.

B. PPO Hyperparameters

Parameter	Value
discount factor γ	0.99
learning rate α	1×10^{-3}
optimizer	Adam
PPO clipping parameter	0.2
Generalized Advantage Estimation λ	0.95
Entropy regularization coefficient	0.001
action distribution	categorical with 4 bins

TABLE I: Parameters used for training the Proximal policy optimization

For our experiments, we used Proximal Policy Optimization to learn the source task. The actor and the critic network consisted of 3 convolutional layers of size 64 with kernel size 3 and stride 1. They were followed by 3 linear layers. The output of the actor network was a probability distribution over the actions, where as the output of the critic network is the expected value. The input to the network were images cropped to size $100 \times 100 \times 3$.