Using Simulation & Domain Adaptation to Improve Deep Robotic Vision

Sreya Francis

Authors: Konstantinos Bousmalis, Sergey Levine

Overview

- Collecting Datasets – Very time consuming and expensive

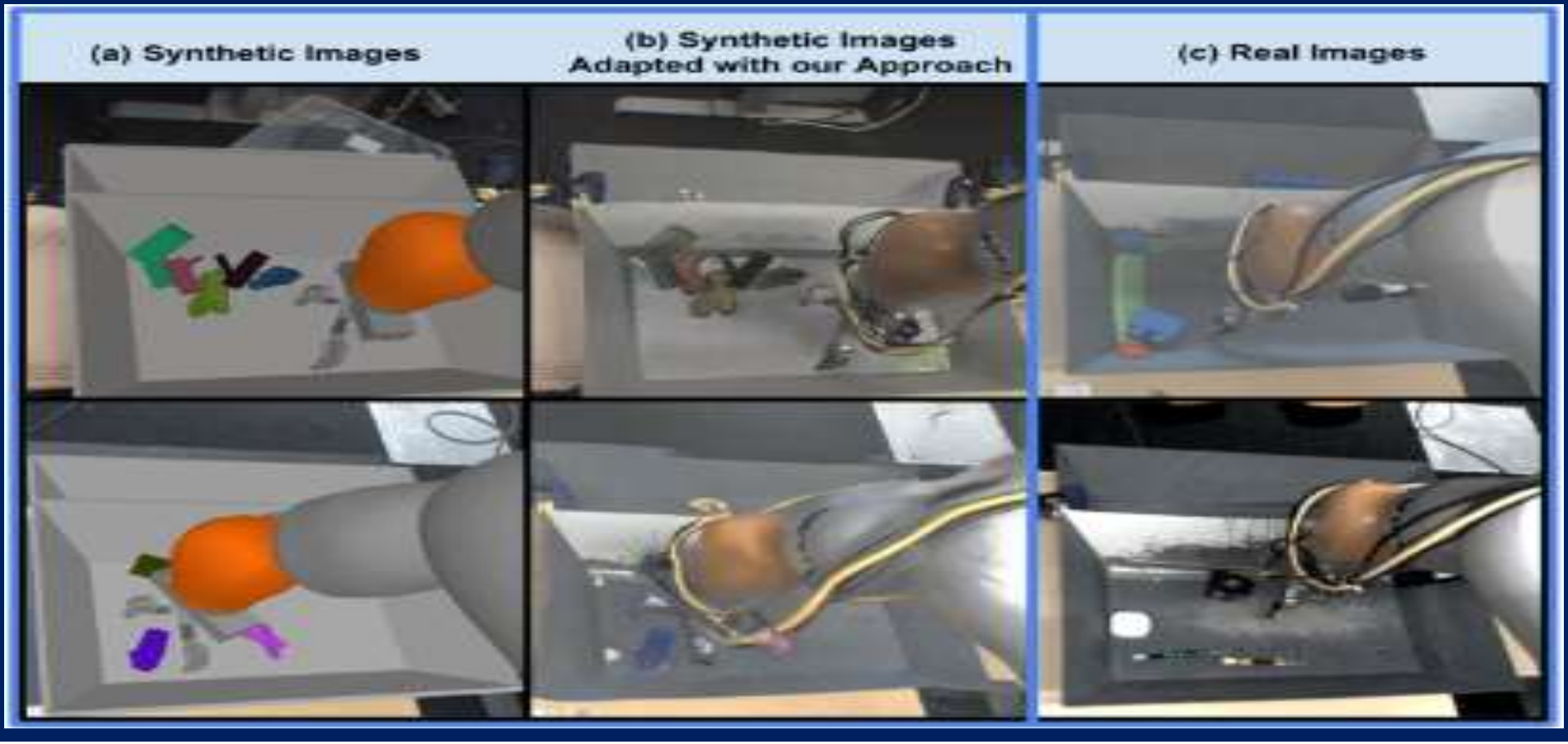
- Use off-the-shelf simulators to generate synthetic data

  - ❖ Ground truth annotations are automatically generated.
  - ❖ Models trained purely on synthetic data fail to generalize in real world.

- Use simulated environments and domain adaptation methods.

  - ❖ Employs pixel level domain adaptation called GraspGAN.

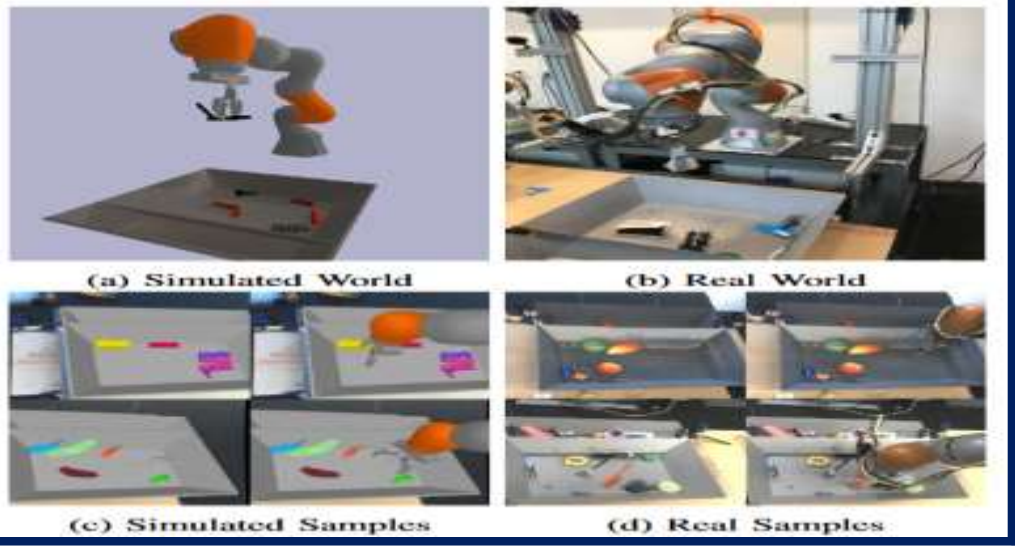
Bridging the Reality Gap

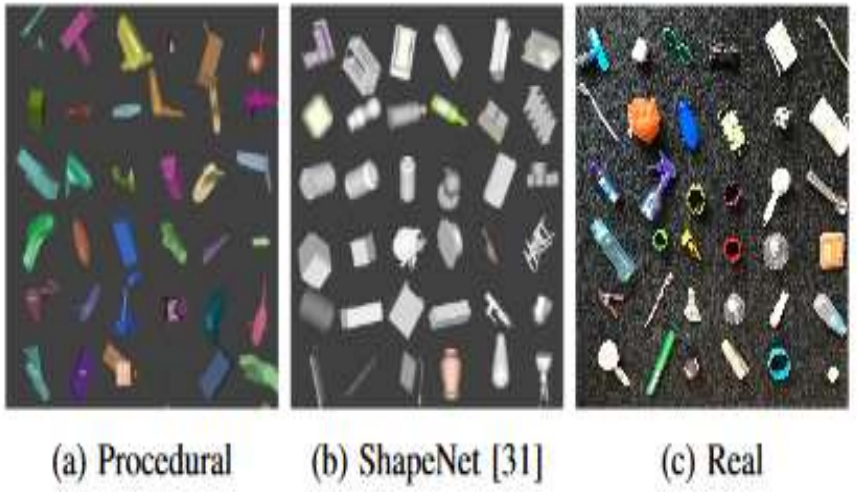


a) Synthetic images produced by simulator
b) Adapted images that look similar to real world ones
c) Real world images produced by camera

- Train the network with real and adapted images
- Refine with feature level adaptation

Simulation





•Construct simulators which ensure diversity sufficient for effective generalization to real world settings.

•To evaluate simulation , collect one real world dataset and multiple such datasets in simulation.

•Built a basic virtual environment(Bullet physics simulator and simple renderer).

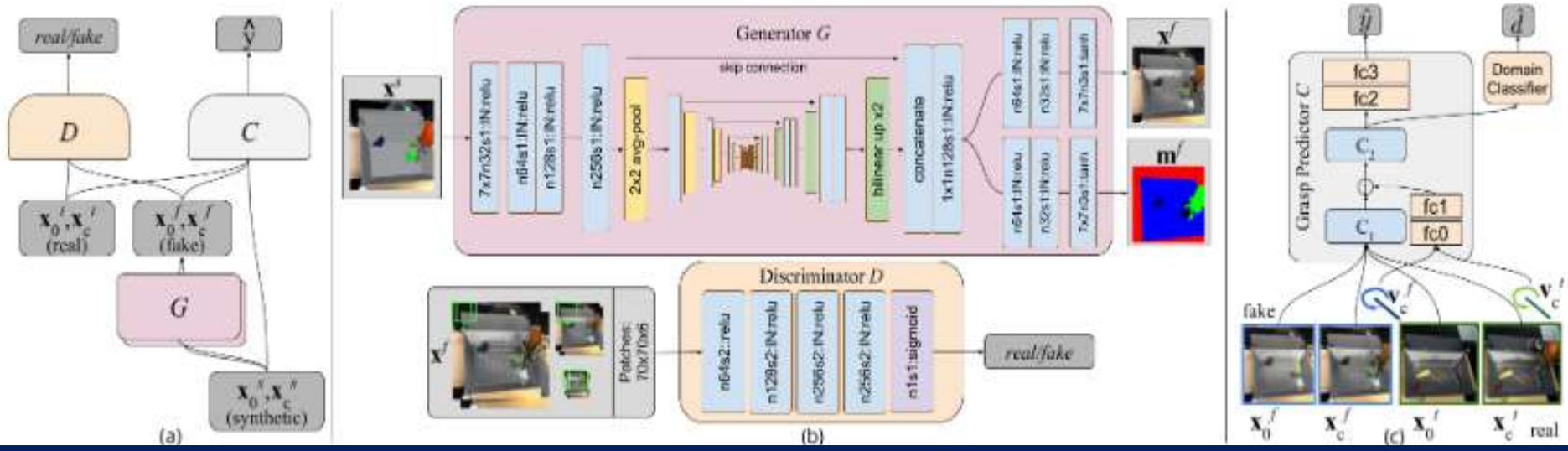
•Apply Virtual Scene Randomization

- ❖ Visual Randomization :Vary background texture, object texture and color, lighting direction and brightness.
- ❖Dynamic Randomization :Vary object mass, object lateral/rolling/spinning friction coefficients.

Domain Adaptation

•Pixel level adaptation using GraspGAN.

❖ Given original synthetic images, GraspGAN produces adapted images that look more realistic.



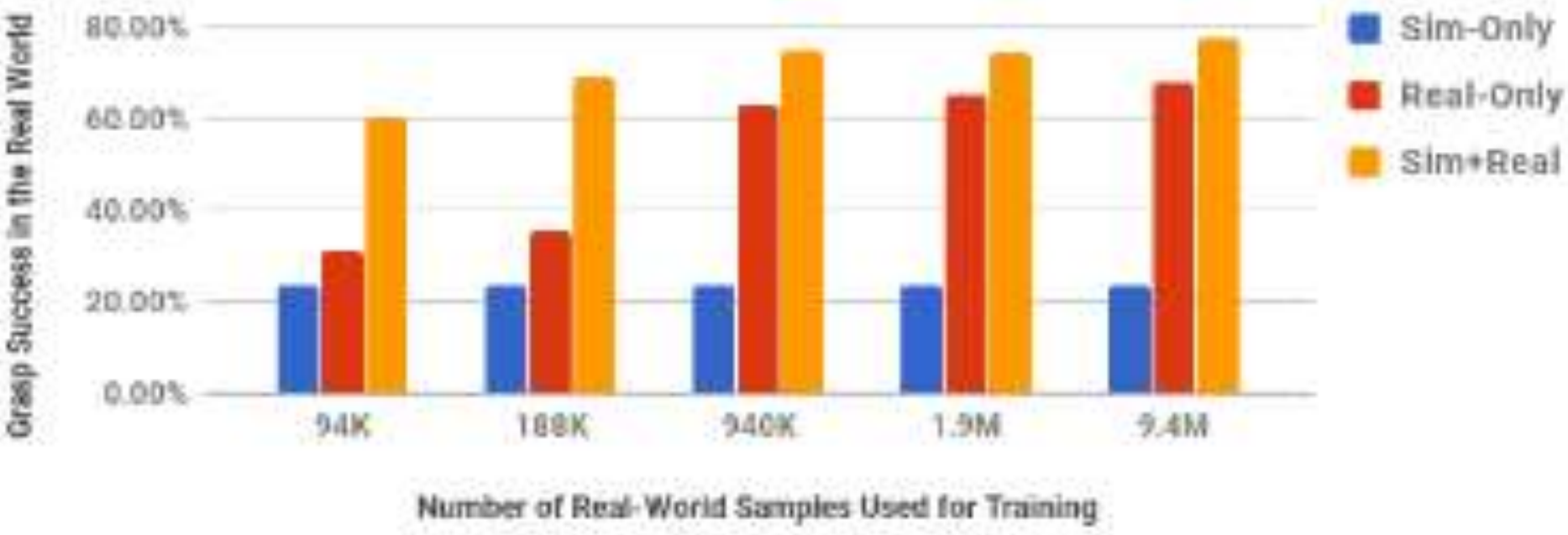
•D (Discriminator)gets unlabeled real world images xt and xf .D is trained to distinguish them.

•C is the grasp success prediction network trained in parallel.

•G (Generator)gets feedback from D and C.

•Simulated images are fed into G to produce realistic versions(xf).

Evaluation/Results



- Simulated data from simulator aids in improving vision based performance.
- We do not need realistic 3d models for simulated training to obtain these gains.
- Using random procedurally generated data/shapes in combination with 10 percent of real world data gives better results.
- Achieves better performance with 50 times fewer real world samples.