Quantifying the Use of Domain Randomization

Mohammad Ani[1], Hector Basevi, Aleš Leonardis
Intelligent Robotics Lab, School of Computer Science, University of Birmingham,
[1]mxa563@cs.bham.ac.uk

Abstract

Synthetic image generation provides the ability to efficiently produce large quantities of labeled data, which addresses both the data volume requirements of state-of-the-art vision systems and the expense of manually labeling data. However, systems trained on synthetic data typically under-perform systems trained on realistic data due to mismatch between the synthetic and realistic data distributions. Domain Randomization (DR) is a method of broadening a synthetic data distribution to encompass a realistic data distribution and provides better performance when the exact characteristics of the realistic data distribution are not known or cannot be simulated. However, there is no consensus in the literature on the best method of performing DR. We propose a novel method of ranking DR methods by directly measuring the difference between realistic and DR data distributions. This avoids the need to measure task-specific performance and the associated expense of training and evaluation. We compare different methods for measuring distribution differences, including the Wasserstein and Fréchet Inception distances. We also examine the effect of performing this evaluation directly on images and features generated by an image classification backbone. Finally, we show that the ranking generated by our method is reflected in actual task performance.

Introduction

- Synthetic images provide the ability to generate large amounts of annotated data typically required by state-of-the-art vision systems and the expense of manually annotating data.
- Systems trained on synthetic data typically underperform systems trained on real data due to a mismatch between the synthetic and real data distributions, as seen in Fig. 1. Note the difference in illumination, texture appearance, and background.

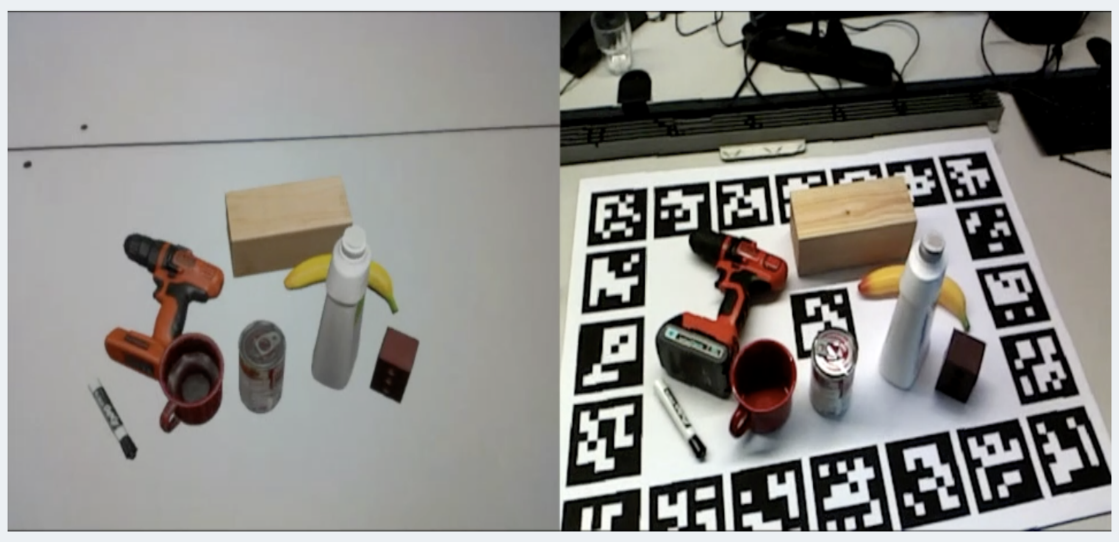

Fig 1. Synthetic images on the left, real-world equivalent on the right [1].

- Domain Randomization (DR) is used to address scenarios where the exact characteristics of the realistic data distribution are unknown. For example, knowing the approximate position for an object of interest, but not the texture as seen in Fig. 2. Samples of DR textures are shown in Fig. 3.

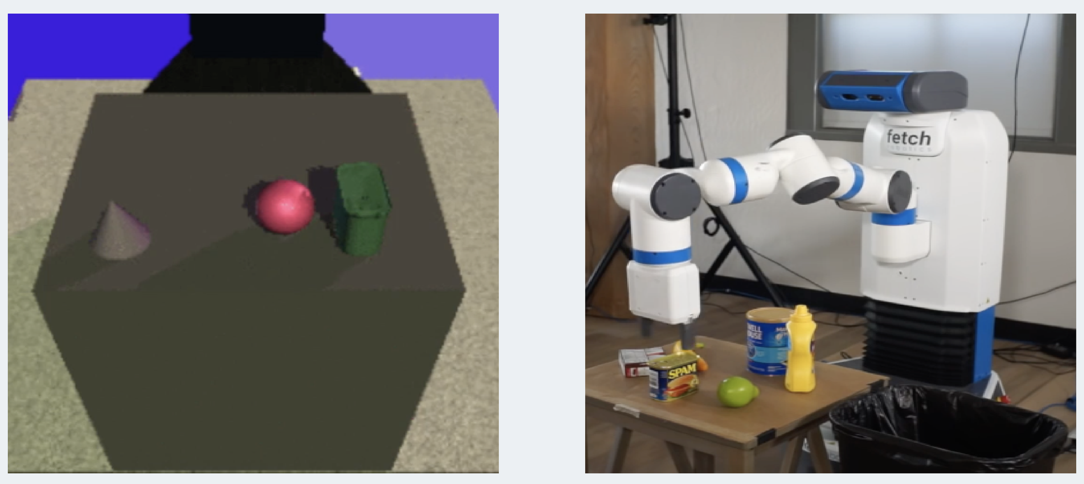

Fig 2. DR image on the left, real-world task on the right [2].

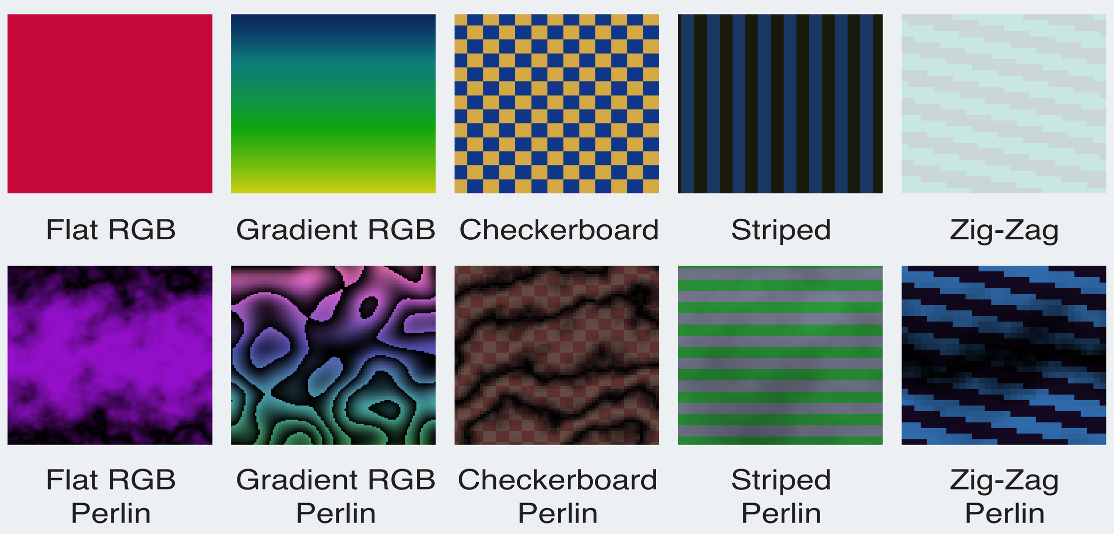

Fig 3. Commonly used textures in existing literature.

Acknowledgements

We acknowledge MoD/Dstl and EPSRC for providing the grant to support the UK academics involvement in a Department of Defense funded MURI project through EPSRC grant EP/N019415/1.

Method

Proposed approach for selecting the most suitable textures for a given task by directly measuring the difference between realistic and DR data distributions to rank the various DR techniques using Wasserstein and Fréchet Inception Distances. Avoids the need to measure task-specific performance and the associated expense of training and evaluation.

- Wasserstein Distance

$$W(P_{data}, P_g) = \inf_{\gamma \in \Pi(P_{data}, P_g)} \mathbb{E}_{(x,y)\sim\gamma} [||x - y||] \tag{1}$$

- Fréchet Inception Distance

$$d^2((m_{aug}, C_{aug}), (m_r, C_r)) = ||m_{aug} - m_r||_2^2 + \mathrm{Tr}(C_{aug} + C_r - 2(C_{aug}C_r)^{\frac{1}{2}}) \tag{2}$$

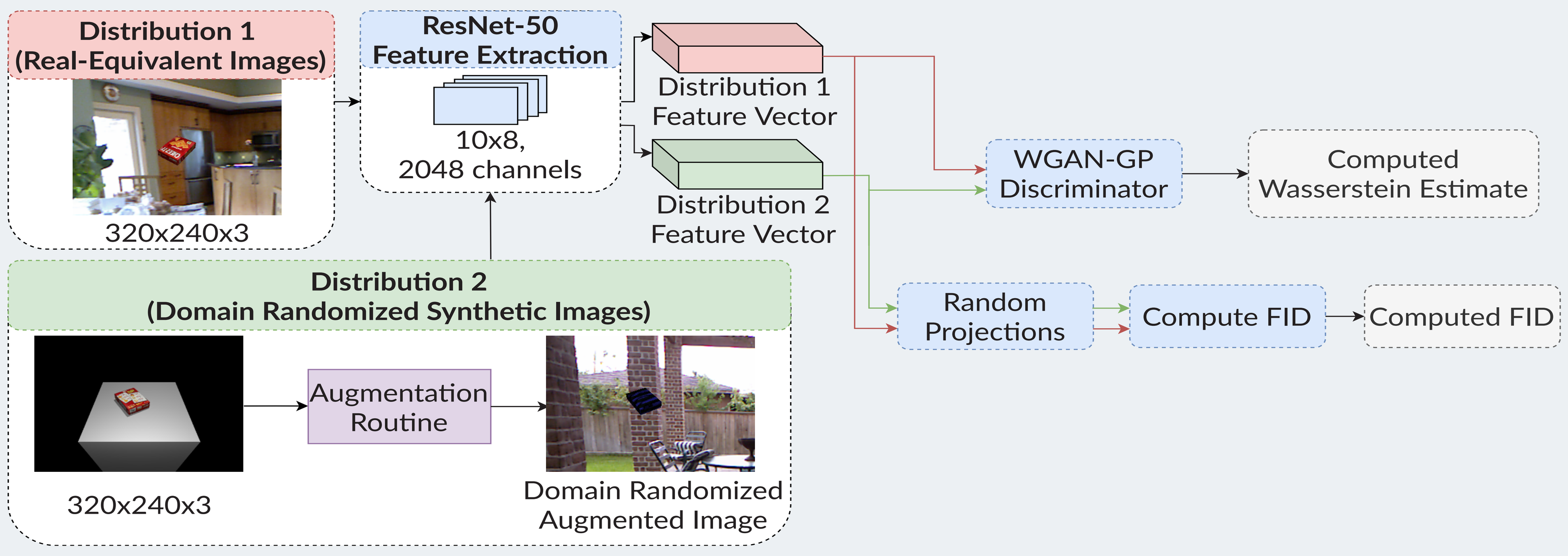

Fig 4. Flow of data through the system.

Results

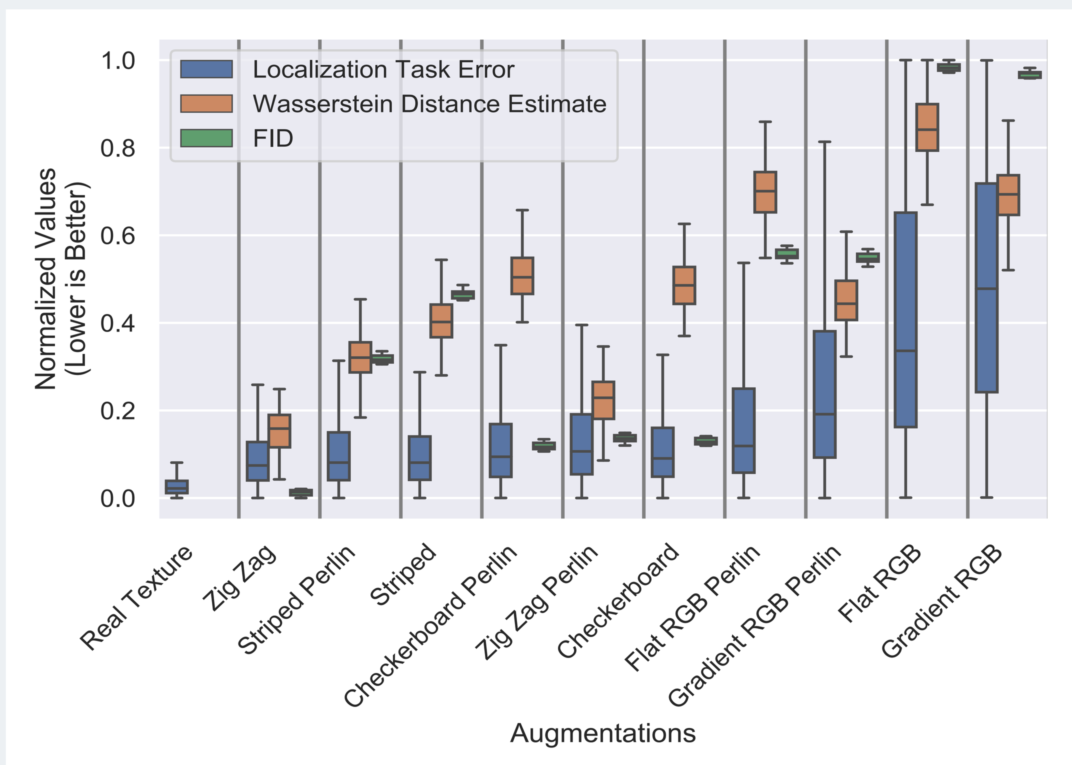

Fig 5. Results on localization task sorted by mean task error. Performance degrades from left to right.

- Upwards trend for both Wasserstein and FID following task-based performance.
- Lower Wasserstein and FID estimates correlates to higher task-based performance.
- Addition of Perlin noise generally aids task-based performance.

Conclusion

- We propose a novel method of quantifying differences between DR data distributions and real-equivalent samples, using neural networks.
- We demonstrate that the method is capable of ranking the different augmentations and is reflected in the performance of an object localization task.
- Based on the produced ranking, generated without task-based training, we recommend using more complex patterned textures when generating DR synthetic data.

References

[1] T. Grenzdörffer, M. Günther, and J. Hertzberg, "YCB-M: A multi-camera RGB-D dataset for object recognition and 6DoF pose estimation," in *2020 IEEE International Conference on Robotics and Automation, ICRA 2020, Paris, France, May 31-June 4, 2020*, IEEE, 2020.

[2] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *IEEE/RSJ International Conference on Intelligent Robots and Systems IROS*, pp. 23–30, 2017.