A Style Transfer Approach to Appearance Agnostic Agents

Sreeganesh Valathara Rajendran

CS/ECE 523 Deep Learning Project
Boston University

Task

Sim2Real is the task of training an agent in simulation and deploying in a real-environment. One of the main methods of implementing Sim2Real is by **domain randomization**. In this project, we're investigating Neural Style Transfer as a domain randomization approach to implementing Sim2Real for autonomous navigation.

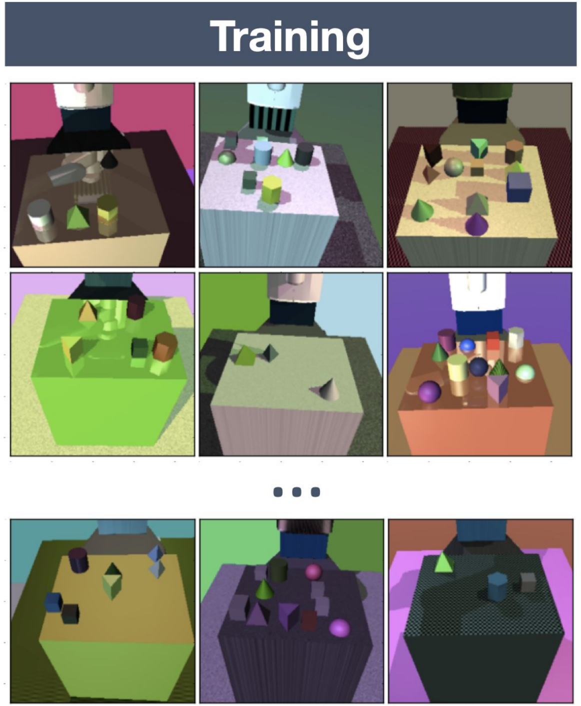
Why Sim2Real is important

1. Abundance of Data. We can simulate a variety of outlier scenarios.
2. Mistakes in simulation will not affect the real world. Important in the context of autonomous vehicles for very obvious reasons.

Related work

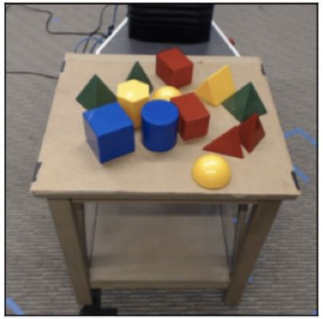
1. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P. (2017). Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. arXiv:1703.06907. [cs.RO]
2. So, J., Xie, A., Jung, S., Edlund, J., Thakker, R., Agha-mohammadi, A., Abbeel, P., & James, S. (2022). Sim-to-Real via Sim-to-Seg: End-to-end Off-road Autonomous Driving Without Real Data (No. 2210.14721). arXiv.
3. Loquercio, A., Kaufmann, E., Ranftl, R., Dosovitskiy, A., Koltun, V., & Scaramuzza, D. (2020). Deep Drone Racing: From Simulation to Reality With Domain Randomization. IEEE Transactions on Robotics, 36(1), 1–14. http://dx.doi.org/10.1109/TRO.2019.2942989

Paper 1

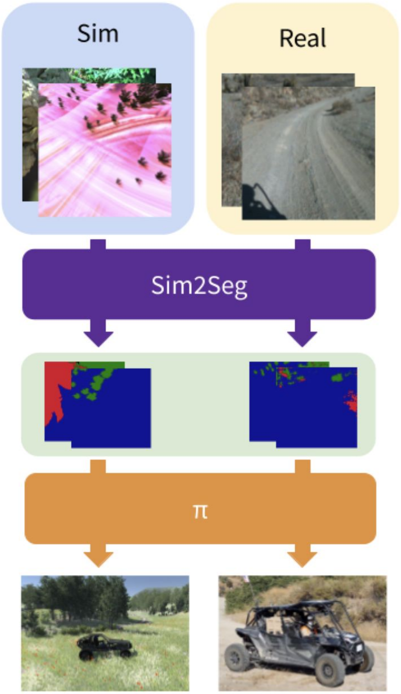


1. Aim: Enhance object localization using domain randomization.
2. Context: Enable grasping by a robotic manipulation.
3. Characterics randomized
   a. Position
   b. Texture
   c. Lighting
   d. Colour
   e. Random Noise
4. Disadvantage
   a. Significant programming in simulation environment to create dataset.

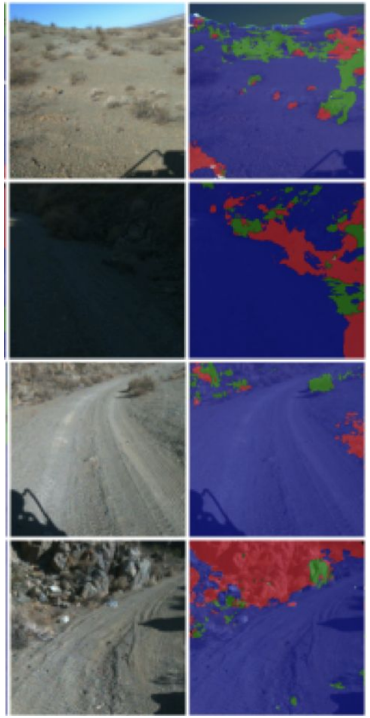
Test



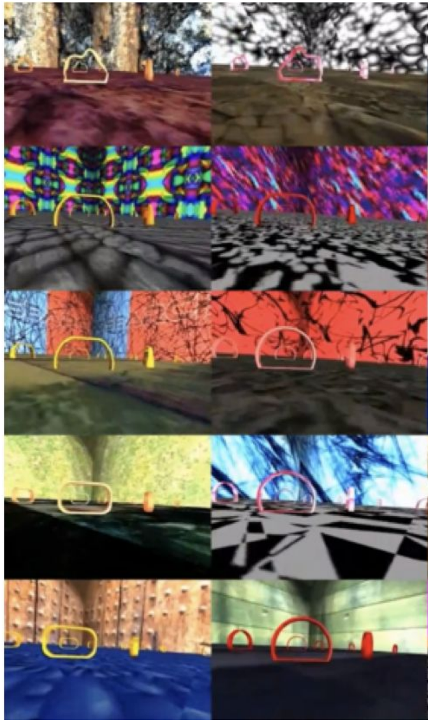
Paper 2



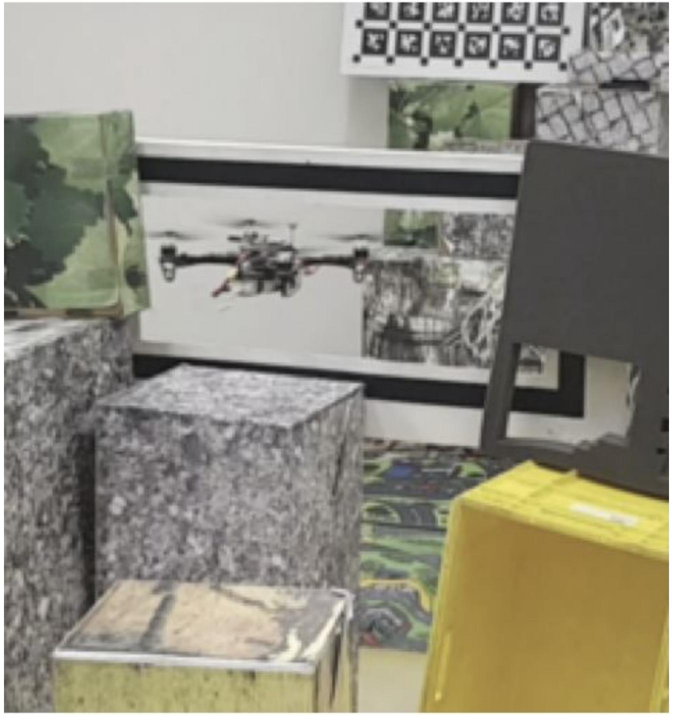
1. Aim: Enable Sim2Real for an autonomous off-road vehicle using segmentation
2. Method: Added a segmentation step in their perception-to-action pipeline
3. This allowed for the decision pipeline to perform zero-shot performance since all inputs are received as segment masks.
4. Disadvantage: Increased pipeline length. Increases latency.



Paper 3

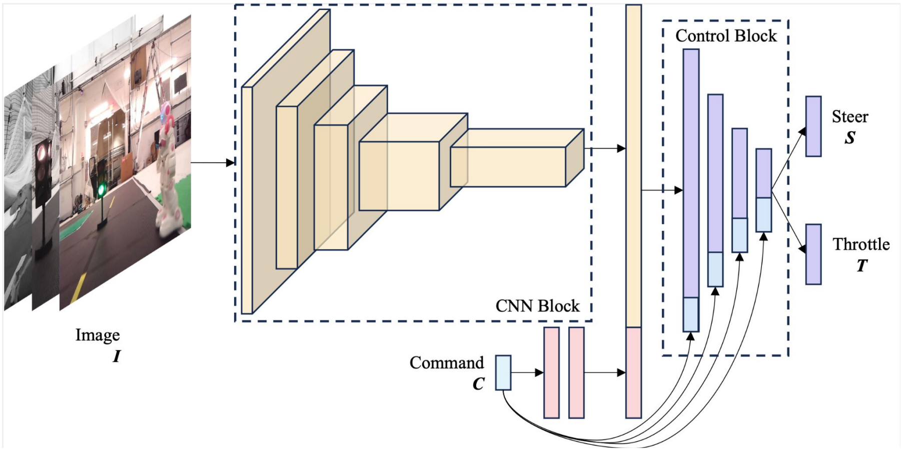


1. Aim: Enable Sim2Real Drone Flying/Racing with only Simulation Training
2. Method: Train agent with inputs where the followed are randomized
    a. Colour
    b. Texture
    c. Background
    d. Brightness
3. Disadvantage: Significant Programming in simulation to create dataset.



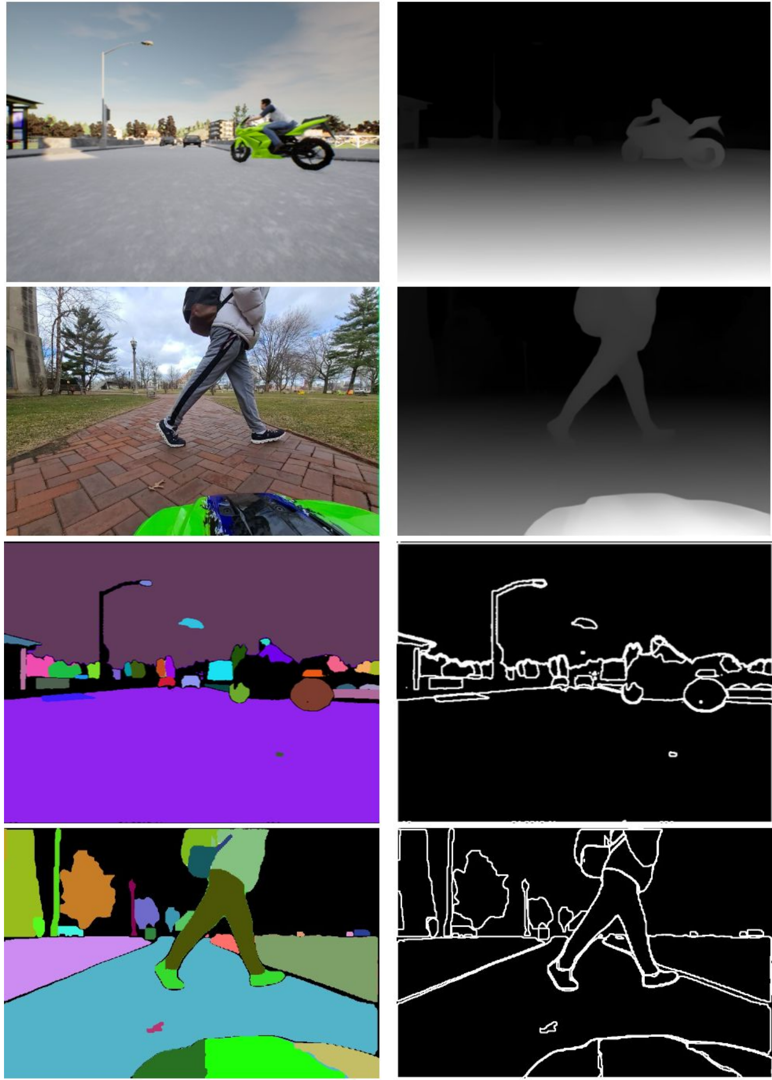
Approach

1. Obtain expert driving dataset. Obtained from Chitta, “Transfuser”.
2. Style Transfer Dataset
3. Train conditional imitation learning agent.
   a. RegNetY-002 Model + custom five-layer convolution neural network
   b. RegNetY-004 Model + custom five-layer convolution neural network
4. Offline evaluate agent on test expert data, from Aich, “Towards Closing The Generalization Gap In Autonomous Driving”.
   a. Indoor Dataset
   b. Outdoor Dataset

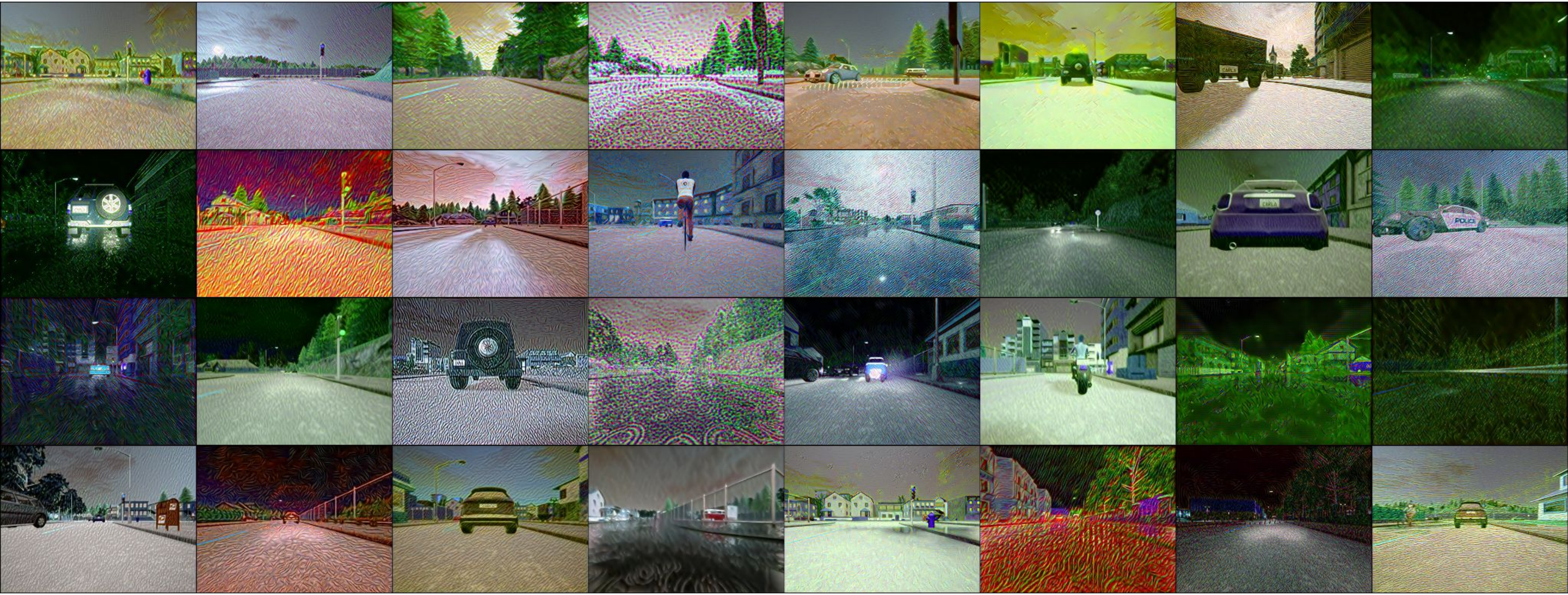


Dataset (s)

1. Train Data: From expert agent in Transfuser by Chitta et al. 2023
2. Test Data: 200,000 frames of expert data. Stores RGB frames in addition to expert control instructions. Obtained from Aich, “Towards Closing The Generalization Gap In Autonomous Driving”, Boston University 2024.
    a. Indoor Dataset: A model city using a black-coloured mat with taped lanes to represent roads. Various obstacles, such as toys, robotic figures are strategically placed to mimic urban obstacles.
    b. Outdoor Dataset: Obtained by an expert controlling an RC car equipped with camera, driven through narrow paths. (Marsh Plaza)



Style Transferred Inputs



Evaluation metric(s)

1. Steer Mean Absolute Error
    a. MAE between expert steer-input vs model's decision.
2. Speed Weighted Mean Absolute Error
3. Action Mean Absolute Error
    a. MAE between expert steer-input and throttle-input vs model's steer-input and throttle-input.
4. Speed Weighted Mean Absolute Error

$$\mathbf{L}_{MAE} = ||a_i - \hat{a}_i||_1$$

$$\mathbf{L}_{SW-MAE} = \frac{1}{N}\sum_{i=0}^{N-1} ||v(a_i - \hat{a}_i)||_1$$

$$\mathbf{L}_{MAE} = ||[a_i, s_i] - [\hat{a}_i, \hat{s}_i]||_1$$

$$\mathbf{L}_{SW-MAE} = \frac{1}{N}\sum_{i=0}^{N-1} ||v([a_i, s_i] - [\hat{a}_i, \hat{s}_i])||_1$$

Results: Indoor, RegnetY-002

RegnetY-002	Steer MAE	Speed-Weighted Steer MAE	Action MAE	Speed-Weighed Action MAE
RGB (Baseline)	0.294	0.391	0.319	0.85
RGB+Depth	0.276	0.383	0.331	**0.812**
SAM-Mask	0.295	0.397	0.304	0.955
SAM-Mask + Depth	0.308	0.405	0.332	0.861
Contour + Depth	0.324	0.414	0.347	0.876
SAM-Mask + Contour + Depth	0.284	0.377	0.312	0.853
RGB-NST04	0.27923	0.38324	0.28695	0.93556
RGB-NST05	0.27185	0.36734	0.28558	0.93269
RGB-NST06	0.27627	0.38895	0.29564	0.89651
RGB-NST07	0.26733	0.37641	0.28837	0.9218
RGB-NST08	**0.23931**	**0.34121**	**0.2647**	0.91443
RGB-NST09	0.59785	0.60309	0.60806	0.92599
RGB-NST10	0.30979	0.42027	0.31789	0.9683

Results: Indoor, RegnetY-004

RegnetY-004	Steer MAE	Speed-Weighted Steer MAE	Action MAE	Speed-Weighed Action MAE
RGB (Baseline)	0.278	0.376	0.296	0.935
RGB+Depth	0.304	0.407	0.321	0.945
SAM-Mask	0.253	0.355	0.278	0.888
SAM-Mask + Depth	0.321	0.418	0.331	0.892
Contour + Depth	0.296	0.397	0.33	0.838
SAM-Mask + Contour + Depth	0.3	0.392	0.317	0.909
RGB-NST04	0.26959	0.37282	0.28365	0.94059
RGB-NST05	**0.23466**	**0.33212**	0.28339	0.87438
RGB-NST06	0.26265	0.35814	0.29365	0.82078
RGB-NST07	0.35429	0.45833	0.39736	**0.76461**
RGB-NST08	0.25557	0.35246	**0.27438**	0.94198
RGB-NST09	1.2225	1.1184	0.8471	1.6286
RGB-NST10	0.29287	0.39248	0.33297	0.81092

Results: Outdoor, RegnetY-002

RegnetY-002	Steer MAE	Speed-Weighted Steer MAE	Action MAE	Speed-Weighed Action MAE
RGB (Baseline)	0.08	0.178	0.357	1.586
RGB+Depth	0.079	0.179	**0.195**	**0.782**
SAM-Mask	0.1	0.22	0.439	1.986
SAM-Mask + Depth	0.086	0.191	0.335	1.475
Contour + Depth	0.122	0.264	0.394	1.753
SAM-Mask + Contour + Depth	0.097	0.214	0.316	1.37
RGB-NST04	**0.070168**	**0.16123**	0.42705	1.9455
RGB-NST05	0.18072	0.39289	0.47923	2.1633
RGB-NST06	0.072733	0.16189	0.40775	1.8347
RGB-NST07	0.074743	0.17182	0.40871	1.8513
RGB-NST08	0.11071	0.24416	0.36804	1.6446
RGB-NST09	0.43561	0.93371	0.25707	0.9663
RGB-NST10	0.16737	0.37257	0.43696	1.9682

Results: Outdoor, RegnetY-004

RegnetY-004	Steer MAE	Speed-Weighted Steer MAE	Action MAE	Speed-Weighed Action MAE
RGB (Baseline)	0.093	0.205	0.481	2.199
RGB+Depth	0.082	0.19	**0.242**	**1.015**
SAM-Mask	0.082	0.194	0.435	1.977
SAM-Mask + Depth	0.116	0.251	0.37	1.641
Contour + Depth	0.097	0.214	0.338	1.486
SAM-Mask + Contour + Depth	0.102	0.223	0.365	1.611
RGB-NST04	0.077964	0.17797	0.41577	1.8774
RGB-NST05	0.071103	0.16338	0.3812	1.7169
RGB-NST06	0.082481	0.18936	0.34499	1.5261
RGB-NST07	**0.068497**	**0.157**	0.31122	1.3787
RGB-NST08	0.07926	0.18112	0.40725	1.8437
RGB-NST09	0.98152	2.1218	0.75437	3.2444
RGB-NST10	0.093697	0.21122	0.38808	1.7472

Results Summary

For Indoor,

- RegNetY-002 with NST produced an decrease of **20%** in Steer-MAE, **12%** in Steer-SW-MAE, **16%** in Action-MAE.
- RegNetY-004 with NST produced an decrease of **14%** in Steer-MAE, **11%** in Steer-SW-MAE, **7%** in Action-MAE and **18%** in Action-SW-MAE.

For Outdoor,

- RegNetY-002 with NST produced an decrease of **12%** in Steer-MAE, **9.5%** in Steer-SW-MAE.
- RegNetY-004 with NST produced an decrease of **26.8%** in Steer-MAE, **23.4%** in Steer-SW-MAE.

Conclusion

So this project shows the promise of Neural Style Transfer as an approach to Sim2Real, specifically in the context of autonomous driving.

Advantage:

1. Empirically proven to improve performance
2. Ease of expanding existing dataset and thus avoid overfitting

Disadvantage:

1. Performance depends on the style images chosen.
2. Performance depends on parameters such as style-weight which implies manual tuning.

A future direction of this project is to test my hypothesis with a number of different models a number of different style transfer parameters to see how the parameters influence the performance of the models.

Thanks!