UAV Depth Perception from Visual, Images using a Deep Convolutional Neural Network

Kyle Julian, John Mern, and Rachael Tompa
Department of Aeronautics and Astronautics, Stanford University

SISL Stanford Intelligent Systems Laboratory

Introduction

Unmanned Aerial Vehicle (UAV) technology has led to a proliferation of affordable vehicles for hobbyist and low-end commercial use. Depth maps are critical for guidance and collision avoidance

Common sensors for depth: LIDAR / RADAR

Depth Sensors = $$$$$ → Visual Cameras = $

Task: extract depth maps from single images

Problem Statement

Real time images from a UAV should inexpensively and reliably be translated into depth maps

Datasets

Microsoft Airsim[1]: a sophisticated UAV simulation environment

- Made to generate UAV images for deep learning
- Gathered raw images and depth images from a simulated neighborhood environment
- Collected 1,963 pairs of images

Divided the data as follows:

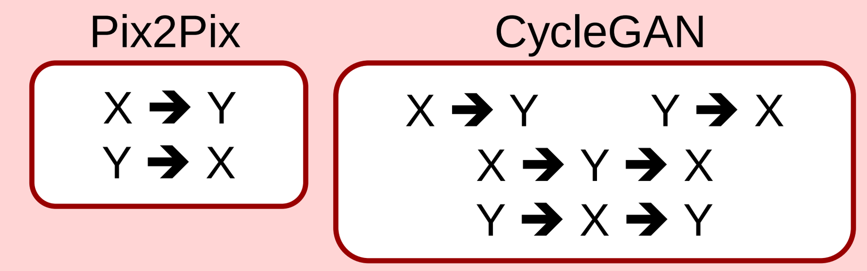
Function	Data %
Train	70
Validate	20
Test	10

Acknowledgments

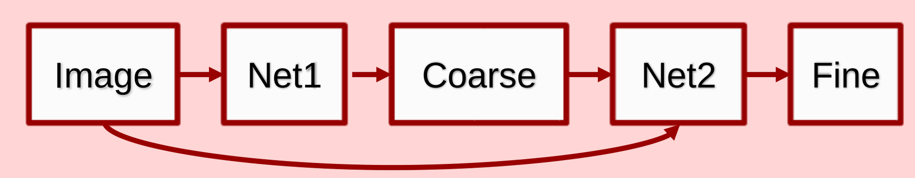
Kyle, John, and Rachael would like to thank Shital Shah for his help with Airsim, Mykel Kochenderfer for his continuous support and enthusiasm about our project, and the CS231N course staff!

Methods

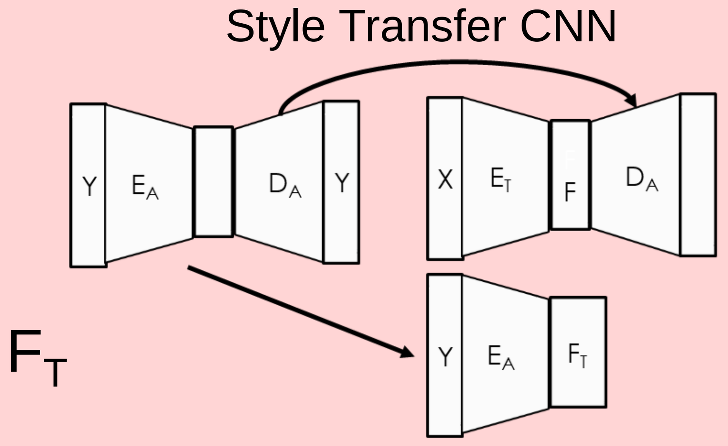
1. Pix2Pix and CycleGAN[2]
   - Image translation approach
   - Unsupervised
   - Multi-objective



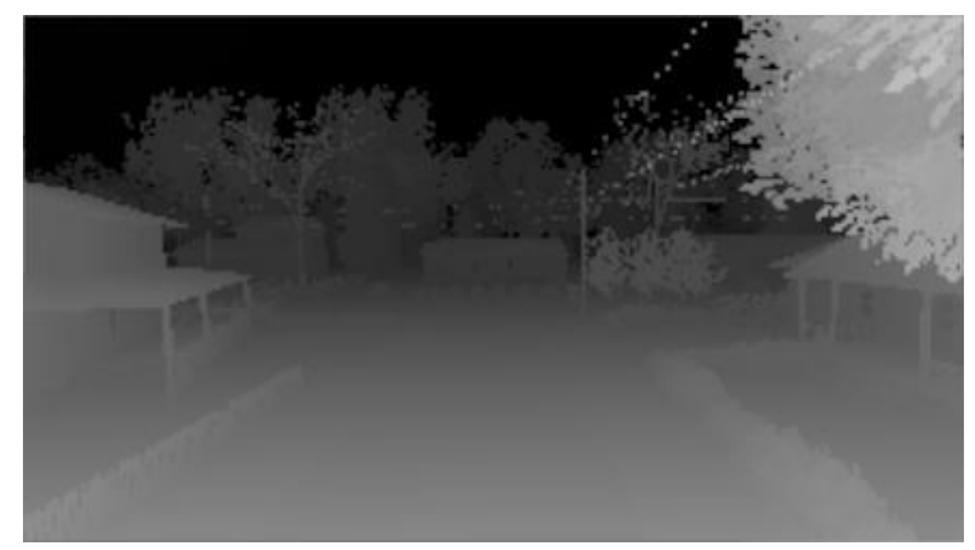
1. Multi-Scale Deep Network[3]
   - Coarse and fine scales
   - Supervised
   - Scale-invariant error



2. Style Transfer CNN[4]
   - Auto-encoder CNN
   - Depths reduced to features
   - Transfer loss on features $L = F - F_T$



Real Scene Image

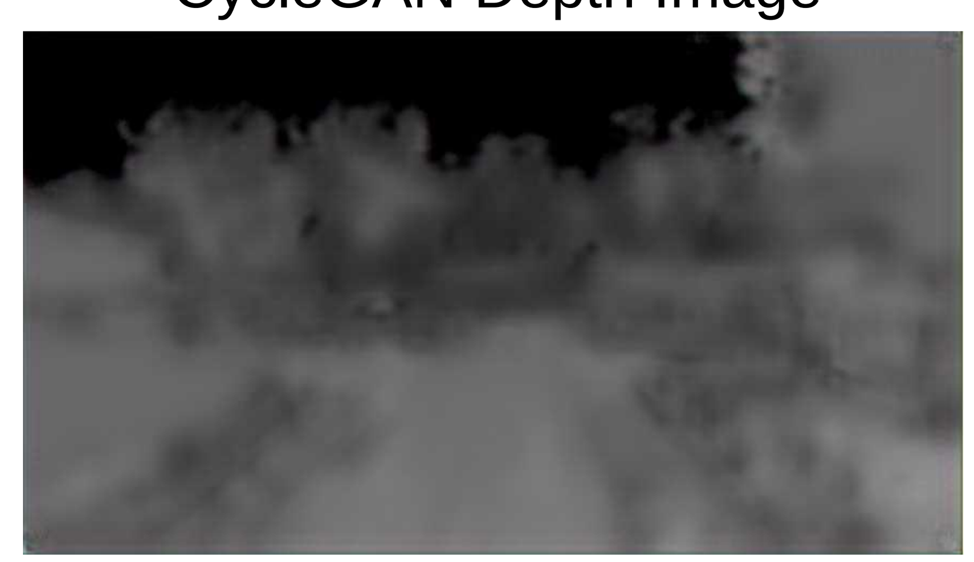


Real Depth Image

Pix2Pix Depth Image

CycleGAN Depth Image

Multi-Scale Depth Image



Style Transfer Depth Image

Experimental Evaluation

The network generated and Airsim generated depth images are compared at a pixel level using the mean squared error (MSE)

The MSEs are averaged for the full test set

Network	Average MSE
Pix2Pix	740
CycleGAN	1443
Multi-Scale Deep Network	660
Style Transfer CNN	3530

Conclusions and Next Steps

Performance Comparison:

- CycleGAN depth appears crisp, but color information must be preserved in the depth map, leading to larger errors in depth values
- Pix2Pix does not have cycle problems, and maintains detail with better depth estimates
- Multi-Scale loses image clarity, but the depth estimates are good on average
- Style Transfer retains features, but spurious details impact depth values

Next Steps:

- Investigate different evaluation metrics
- Sweep through hyper-parameters

Initial claim: based on our results we claim that images can be used to create depth images resulting in affordable 3D scene maps

References

[1] S. Shah, D. Dey, C. Lovett, and A. Kapoor. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. arXiv:1705.05065, 2017.
[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.
[3] D. Eigen, C. Puhrsch, and R. Fergus. Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. arXiv: 1406.2283, 2014.
[4] J. Johnson , A. Alahi, and F. Li. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. http://dblp.uni-trier.de/rec/bib/journals/corr/JohnsonAL16. 2016.