

# mid-term presentation: realistic MVS dataset

Peter Trost

Eberhard Karls Universität Tübingen

*peter.trost@student.uni-tuebingen.de*

May 22, 2019

# Overview

## 1 Related Work

- synthetic datasets
- Generative Adversarial Networks

## 2 Current Progress

- dataset
- GANs

## 3 References

# Overview

thesis goals:

- 1 create MVS dataset including RGB-images, depth-maps and camera parameters of frames from virtual city
- 2 use neural networks (GANs) to make results more realistic

# Overview

## 1 Related Work

- synthetic datasets
- Generative Adversarial Networks

## 2 Current Progress

- dataset
- GANs

## 3 References

# A naturalistic open source movie for optical flow evaluation

[Butler et. al]



Figure: ground truth flow (top), corresponding RGB-images (bottom)

# A naturalistic open source movie for optical flow evaluation

- dataset for optical flow estimation
- derived from open source 3D animated short film *Sintel*
- contains long sequences, large motions, specular reflections, motion blur, defocus blur, atmospheric effects and more
- authors use Blender to get ground truth optical flow maps

# Playing for data: Ground truth from computer games

[Richter et. al]



Figure: RGB images (left), corresponding semantic segmentation (right)

# Playing for data: Ground truth from computer games

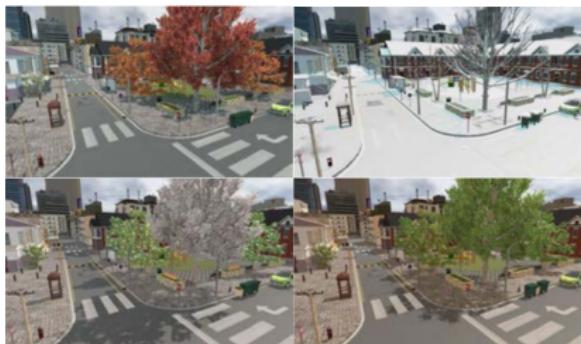
- dataset containing RGB-images, depth-maps and pixel accurate semantic segmentation of frames from Grand Theft Auto V
- accomplished by saving rendering commands for geometry, textures, shaders from the game
- hash geometry, texture, shader to obtain object signatures
- while labeling, labels of objects get propagated to all frames containing the same object

# The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes

[Ros et. al]



**Figure:** RGB images (top),  
corresponding depth-maps (bottom)



**Figure:** from top left to bottom  
right: fall, winter, spring, summer

# The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes

- dataset containing RGB-images, depth-maps and pixel-level semantic segmentation of virtual city
- includes 4 different seasons (different weather conditions), day- and nighttime (varying lighting conditions)
- SYNTHIA-Rand: 13,400 frames of the city (camera randomly moved through city)
- SYNTHIA-Seq: one 50,000 frames video for each season (simulated car driving through city)

# SyB3R: A Realistic Synthetic Benchmark for 3D Reconstruction from Images

[Ley et al.]

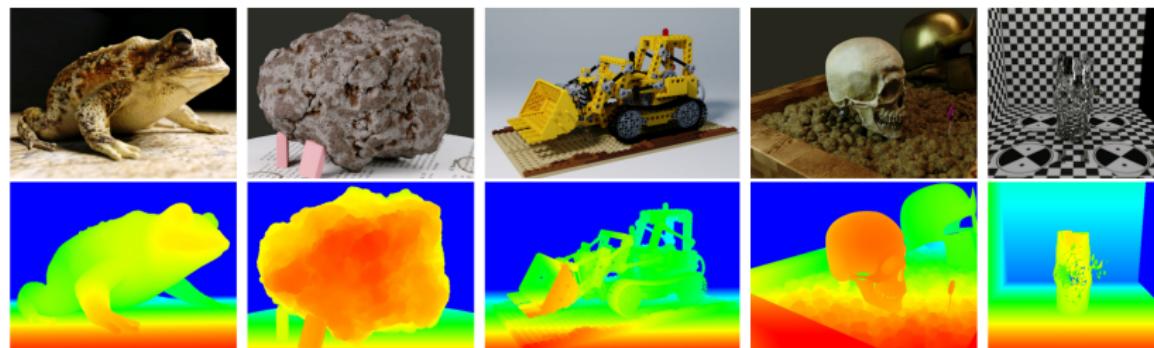


Figure: RGB-images (top), depth-maps (bottom)

# SyB3R: A Realistic Synthetic Benchmark for 3D Reconstruction from Images

- framework to evaluate 3D reconstruction algorithms
- all camera parameters and 3D structure of scene is known
- includes real world effects like motion blur and noise



Figure: SyB3R's modular post processing pipeline

# DeepMVS: Learning Multi-view Stereopsis

# Overview

## 1 Related Work

- synthetic datasets
- Generative Adversarial Networks

## 2 Current Progress

- dataset
- GANs

## 3 References

# Image-to-Image Translation with Conditional Adversarial Networks

[Isola et. al]

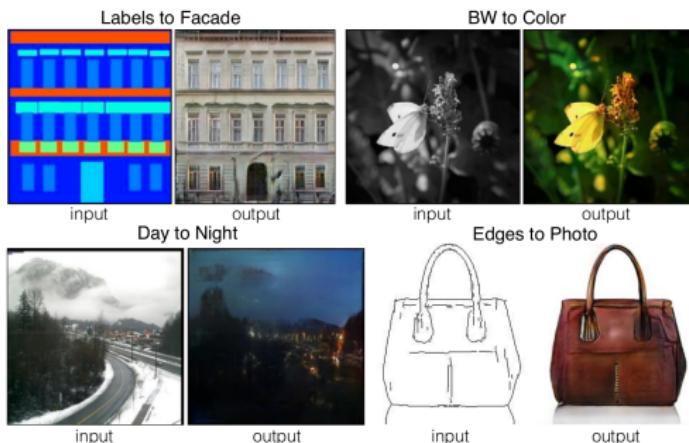


Figure: inputs and corresponding outputs to conditional GAN

# Image-to-Image Translation with Conditional Adversarial Networks

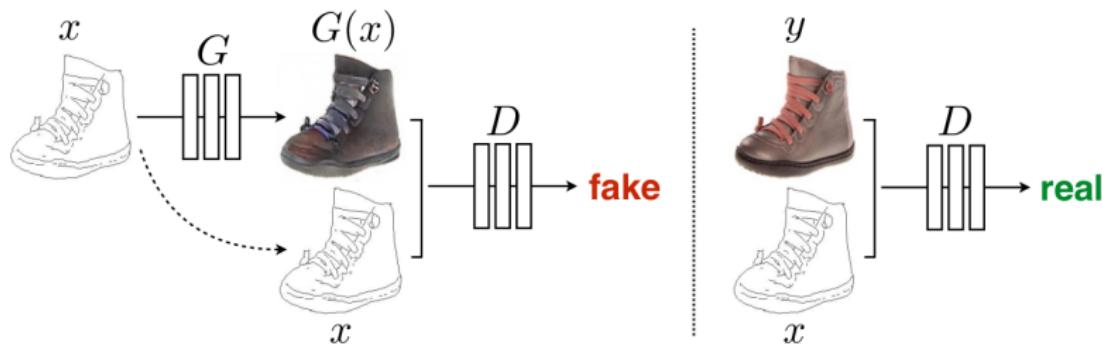


Figure: schematic overview of how GANs work

# Image-to-Image Translation with Conditional Adversarial Networks

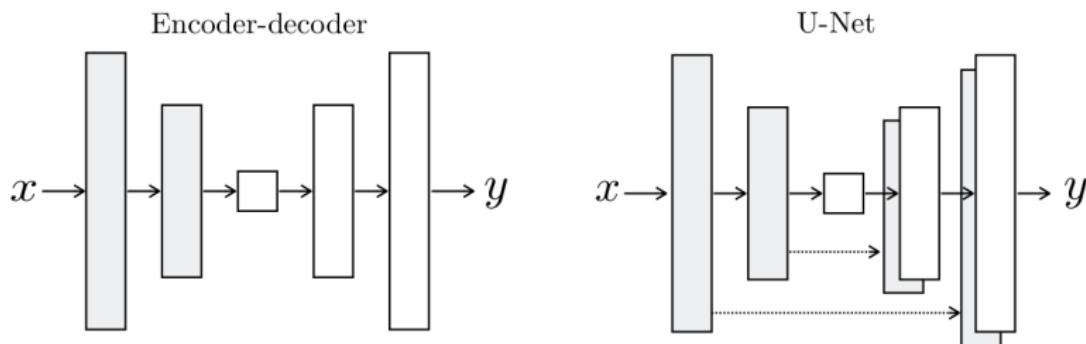


Figure: encoder-decoder and U-Net networks

# Image-to-Image Translation with Conditional Adversarial Networks

- TODO

# Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

[Zhu et. al]

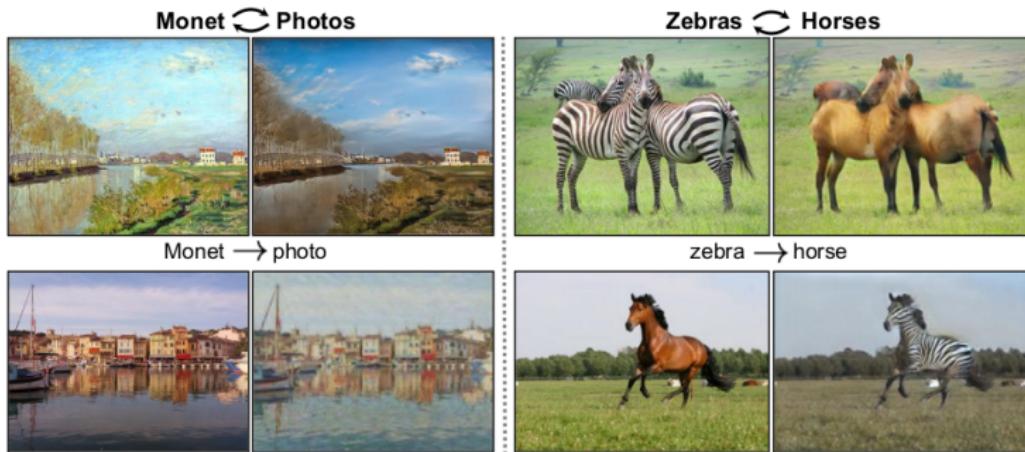


Figure: images translated by cycleGAN

# Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

[Zhu et. al]

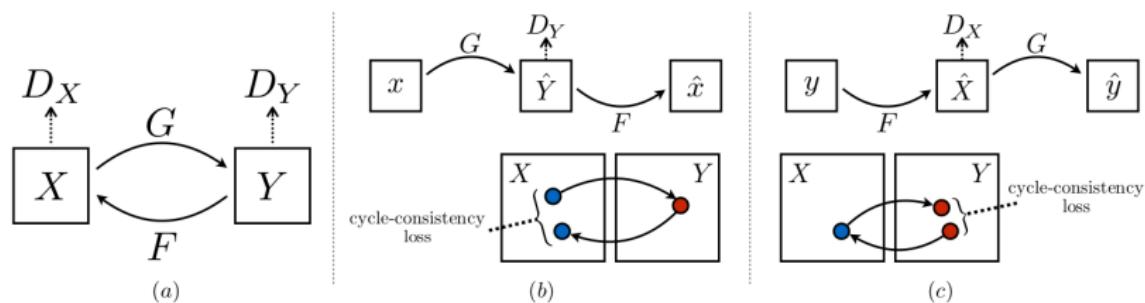


Figure: schematic overview of full cycleGAN with domains  $X$  and  $Y$ , discriminators  $D_X$  and  $D_Y$  and Generators  $G$  and  $F$

# Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

- TODO

# Overview

## 1 Related Work

- synthetic datasets
- Generative Adversarial Networks

## 2 Current Progress

- dataset
- GANs

## 3 References

# RenderDoc

- Website: *renderdoc.org*
- graphics debugger
- allows single-frame capture with detailed introspection of applications using Vulkan, D3D11 and more
- free under *MIT license*
- used in playing-for-data and available at  
<https://bitbucket.org/visinf/projects-2016-playing-for-data/>

# Captures

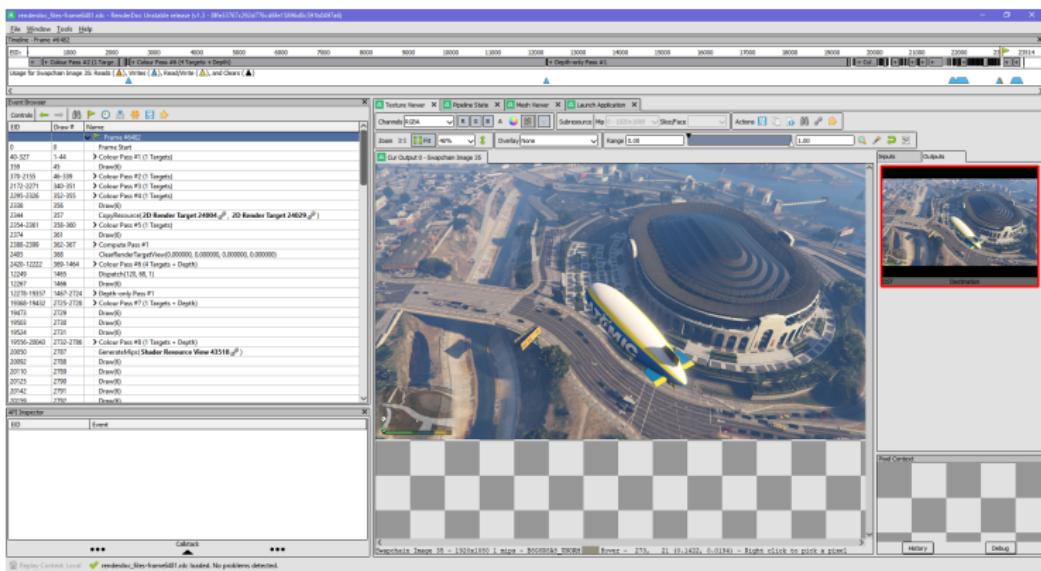


Figure: RenderDoc Capture from GTAV

# extract.py

- python script to extract RGB-images, depth-maps and camera parameters
- original version available at  
*<https://bitbucket.org/visinf/projects-2016-playing-for-data/>*
- takes around 20-30 seconds per frame

# challenges

- playing-for-data from 2016, due to GTAV updates their RenderDoc version can't capture anymore
- current RenderDoc version not compatible with extract.py
- playing-for-data RenderDoc included openEXR to save depth-maps, standard RenderDoc does not → depth-maps accuracy too low
- find the right drawcalls for depth-maps and RGB-images

## current state

- modified RenderDoc to automatically capture every 40th frame ingame
- updated extract.py to work with current RenderDoc API
- not able to save depth-maps in high accuracy due to problems with dependencies while adding openEXR to current RenderDoc build

# looking forward

- goal: finish thesis by end of July
- therefore: skip capturing frames (for now)
- use MVS-Synth dataset which also includes RGB-images, depth-maps and camera parameters from GTAV frames to train GANs

# Overview

## 1 Related Work

- synthetic datasets
- Generative Adversarial Networks

## 2 Current Progress

- dataset
- GANs

## 3 References

# References



Butler et al.

A naturalistic open source movie for optical flow evaluation

*European Conf. on Computer Vision (ECCV)* Part IV, LNCS 7577, 611 – 625. Springer-Verlag, October 2012



Ley et al.

*SyB3R: A Realistic Synthetic Benchmark for 3D Reconstruction from Images*, pages 236 – 251. Springer International Publishing, 2016



Ros et al.

The SYNTHIA Dataset: A large collection of synthetic images for semantic segmentation of urban scenes

2016



Richter et al.

Playing for data: Ground truth from computer games

In Bastian Leibe et al., editors *European Conf. on Computer Vision (ECCV)*, LNCS 9906, 102 – 118. Springer International Publishing, 2016

# References



Huang et al.

DeepMVS: Learning Multi-view Stereopsis

*CoRR*, abs/1804.00650, 2018



Isola et al.

Image-to-image translation with conditional adversarial networks

*CoRR*, abs/1611.07004, 2016



Zhu et al.

Unpaired image-to-image translation using cycle-consistent adversarial networks

*CoRR*, abs/1703.10593, 2017

# mid-term presentation: realistic MVS dataset

Peter Trost

Eberhard Karls Universität Tübingen

*peter.trost@student.uni-tuebingen.de*

May 22, 2019