Initial Plan

Mirror detection from RGBD images

CM3203 One Semester Individual Project

40 credits

Author: Damjan Dimovski

Supervisor: Dr Jing Wu

Contents

Project Description

Mirrors appear quite often in image datasets meant for computer vision tasks such as object detectors. They can easily cause ambiguities in computer vision systems that do not consider their existence by considering the objects in the reflection of the mirror to be real. One application where this might be very problematic is in the field of robotics. Imagine a robot navigating through a scene where there is a large mirror. The robot might not be able to differentiate the real scene from the scene reflected in the mirror, resulting in it trying to navigate in the reflection. A way to alleviate this problem is creating a system that will differentiate reflections from real scenes, i.e. a system that can detect mirrors. Currently, there exist some methods that preform mirror segmentation in images, and one example is MirrorNet [1]. This model uses semantical and low-level colour/texture discontinuities between the contents inside and outside of the mirrors to perform mirror detection. MirrorNet has been created by utilising only RGB image data. The aim of this project is to propose a novel model that would integrate depth data, captured using an RGBD camera, alongside RGB data to perform mirror detection, and to see if this would improve the performance over MirrorNet. Depth data of an RGB image is an image where each pixel relates to a distance between the camera and the corresponding object(s) in the RGB image. An example of such data can be seen in Figure 1.

Figure 1. Example of RGB image and its corresponding depth image[1]

My plan is to implement a Multi-Task Learning model to achieve this. This model would have one encoder and two decoders that would each perform a different task. The first decoder would perform mirror segmentation and the second would perform depth reconstruction.

I will start with constructing a model that will perform mirror segmentation, similar to MirrorNet, the difference being I will use a different model architecture for the mirror segmentation. I will train this model using the MSD[2] dataset which was also used to train MirrorNet. An example of the MSD dataset can be seen in Figure 2.

[1] These two example images were taken from the RGB-D Object Dataset which is free to be downloaded from: https://rgbd-dataset.cs.washington.edu/

[2] The MSD (Mirror Segmentation Dataset) used for MirrorNet is free to be downloaded from: https://mhaiyang.github.io/ICCV2019_MirrorNet/index

Figure 2. Example RGB image from MSD dataset and its corresponding mirror mask

After implementing the mirror segmentation, I will move on to make a Multi-Task Learning model by adding another decoder to the previous model that only performs mirror segmentation. This decoder will perform depth reconstruction. One advantage of having a Multi-Task Learning model is that we could alternatively train both branches (decoders) of the model using different datasets. For the depth reconstruction branch I will create my own dataset that will be composed of at least one hundred RGB images and their corresponding depth masks. This dataset will be collected with an Intel RealSense SR300 camera.

Given an RGB image the Multi-Task Learning model will be able to output the image's estimated depth and segmentation map. I will then perform tests to compare the accuracy of the mirror segmentation results between my model and MirrorNet.

A desirable goal, that I also have in plan, is to replace the mirror segmentation branch with MirrorNet, while keeping the depth reconstruction branch the same. The accuracy of the 3 models will then be analysed.

Project Aims and Objectives

This project aims to test the following hypothesis: “Integration of depth information alongside RGB information in a mirror detection model can improve performance over current mirror detection models that use RGB information only”. I aim to produce a Multi-Task Learning model that would perform depth reconstruction and mirror segmentation. One model that could help me achieve this is a Multi-task learning model with Real-Time Joint Semantic Segmentation and Depth Estimation Using Asymmetric Annotations [2], but this idea could change depending on the difficulty of the implementation of the model.

I will use two datasets to train my model, one of them is the MSD dataset I mentioned before and the other is my dataset I construct using an Intel RealSense SR300 camera. My implementation will be using a ML framework called PyTorch which is already very popular for implementing a Multi-task learning neural-network that consists of mirror segmentation and depth estimation.

Aims

Initial Goal

To construct a Single-Task Learning model with one decoder that would perform mirror segmentation, which given an RGB image as an input would be able to output the image’s segmentation map.

Target Goal

To construct a Multi-Task Learning model with 2 decoders, one would perform mirror segmentation and the other depth reconstruction, which given an RGB image as an input would be able to output the image’s estimated depth and segmentation map.

Objectives

1. Construct a dataset of at least one hundred RGB images and their corresponding depth masks using an Intel RealSense SR300 camera.
2. Implement a Single-Task Learning model that performs mirror detection using a semantic segmentation algorithm.
3. Implement a Multi-task learning model by adding another decoder to the model of objective (2) that would perform depth reconstruction.
4. Preform testing between my model and MirrorNet.
5. **Desirable Objective:** Construct a model based on the one from objective (3) and replace the mirror segmentation branch with MirrorNet. Preform testing between this model and MirrorNet.

Work Plan

I have outlined a plan on a weekly basis. This may alter slightly as the project develops. I have agreed to have a half-hour meeting with my supervisor once every week, where we will go over the previous week's deliverable(s).

Week 1: I will create the dataset needed for the training of the models. This completes objective (1). I will also make sure with my supervisor that the vision for the project goal is clear.

Week 2 and Week 3: I will work on trying to understand an example code[3] of an implementation of a Multi-task learning model with Real-Time Joint Semantic Segmentation and Depth Estimation Using Asymmetric Annotations [2]. I will then work on implementing the Single-Task Learning model (one decoder that does mirror detection). I will be checking with my supervisor that the implementation is going correctly to make sure I do not make any mistakes especially during the start of the implementation. By the end of week 3 I should complete objective (2). I will also make sure that I am writing what I have achieved, until now, in the final report.

Week 4, Week 5 and Week 6: This week will focus on completing objective (3), adding a second decoder to the model that will perform depth reconstruction. If needed I will use some time to finish off any left work I have on objective (2). I will also make sure that I am writing what I have achieved, until now, in the final report.

Week 7: Evaluating the Multi-Task Learning model I have created against MirrorNet. This completes objective (4).

Week 8: This week would be an extra week in case I am behind on some deliverable(s) from the previous weeks. I will also use this week to make sure I have added content I have achieved, until now, to the final report.

Week 9 and 10: If I have time I will focus on the **Desirable Objective (5).** This will include replacing the mirror segmentation branch from my model with MirrorNet. Then I would perform evaluation between my new model, my previous model and MirrorNet to see if performance were improved.

Week 11 and 12: Focus on finishing the writing of the final report, combining all the work achieved in the previous weeks.

[3] The example code for implementing the model can be found here: https://densetorch.readthedocs.io/en/latest/MULTITASKREADME.html

References

[1] X. Yang, H. Mei, K. Xu, X. Wei, B. Yin and R. W. H. Lau, “Where Is My Mirror?,” *CoRR,* vol. abs/1908.09101, 2019.

[2] V. Nekrasov, T. Dharmasiri, A. Spek, T. Drummond, C. Shen and I. D. Reid, “Real-Time Joint Semantic Segmentation and Depth Estimation Using Asymmetric Annotations,” *CoRR,* vol. abs/1809.04766, 2018.