**Self-Supervised Object Recognition Using Synthetic Data for E-Waste Classification**

***Matt Travers & Howie Choset***

***Carnegie Mellon***

Object recognition from raw camera/video data is effectively a commodity capability given contemporary deep neural network models. Deep networks now make it possible to recognize a large variety of different objects with near human-level accuracy. However, the main issue facing existing recognition models is that they rely on there existing a large collection of labeled data that is necessary to tune their many internal parameters (i.e., to train the models). The standard approach to address this data reliance is through brute force, otherwise known as “hand labeling,” where human operators go through individual images and manually draw boxes around portions of the pictures that show a particular object class. Unfortunately, hand labeling data is expensive because it requires humans to spend a great deal of time performing a relatively tedious task. Even worse, a large collection of images that picture a particular class of object may not exist to be labeled in the first place; in such cases, “scrubbing” the internet is not an option.

This work proposes that there are a large number of potential applications where these specific issues related to “data existing” and “data labeling” are currently prohibitive bottlenecks. We thus develop a framework that automatically generates and then labels a novel data set with a specific focus on industrial-scale sorting (e.g., sorting products at a recycling plant). As a motivational example, consider sorting products at an electronics recycling (e-recycling) plant. Depending how products are shipped, the first step at an e-recycling plant divides the mixed waste into distinct piles, and thus for this process to be automated, we must have a system that is capable of recognizing different classes of products (e.g., iPhone 8s, Samsung Galaxy S9, *etc.*).

Existing robotic systems that are currently deployed in recycling plants across the world to sort mixed products use computer vision systems to recognize different object classes. However, these systems are almost exclusively pre-trained using hand-labeled data sets. The primary issue with this approach is that when a new product needs to be recognized, the system needs to be re-trained. This in turn means more labeling. However, the sheer scale and number of “new” products that may only exist in small quantities and may not be documented on the web means that this model does not scale well. This work thus proposes to develop a solution to this problem, wherein the process for collecting new data, labeling it, and retraining an object recognition model is fully automated.

In particular, we draw inspiration from recent efforts in the computer vision and robotics literature that automatically generate pre-labeled, *synthetic* data. The proposed approach to be developed will execute a specific sequence of steps:

* A new object class (e.g., a new type of cell phone) that is not currently recognized by a pre-trained classifier (e.g., a phone-type classifier) is identified;
* An RGB-D camera attached to the end effector of a 6DOF manipulator is then manipulated around the object to create a full 3D scan (thus needing the physically “flip” the object over);
* A 3D mesh for the object is generated from the point cloud data, removing most of the effects due to lighting conditions and texture;
* Given the 3D mesh, a physics simulator is used to simulate a number of different scenes where the number of objects, relative clutter amongst objects, object textures, camera viewpoints, and lighting conditions are varied and used to produce a large number of synthetic images;
* The physics simulator’s collision checking module is used to create “realistic” physical scenes.
* A viewpoint projection method is used to automatically label the simulated data, effectively determining if an object from the “unknown” class is viewable in a given synthetic image; if so a bounding box is automatically drawn around the viewable area;
* The synthetic data and real images generated while building the initial point clouds are used to train a CNN that recognizes the new object class when given real images at run-time.

The approach outlined above is directly motivated by recent results in the computer vision literature that relate to successful demonstrations of something referred to as “domain randomization (DR)” [1,2]. The idea behind DR is simple yet eloquent: if enough variation, in terms of viewpoint, texture, lighting, *etc.*, can be created in a batch of synthetic training data, during deployment in the “real world,” the natural lighting condition, textures, *etc.* will appear as just another variation in the data. The primary innovation to be developed in this work is in determining what type of variations work best for the specific target application of industrial-scale sorting, with a specific focus on sorting e-waste. Of particular interest is determining how to best work with texture variations. For example, damage, such as a cracked screen, can potentially be added or subtracted to object models to simulate a large variety of different potential conditions that have never “actually” been seen in the real data.

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

[2] Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S. and Birchfield, S., 2018. Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 969-977).

Milestones:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tasks | Month | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Learn to identify “new” object class using pre-trained classifier |  |  |  |  |  |  |  |  |  |  |  |  |
| Manipulate objects and move RGB-D camera to create point clouds of “entire” object |  |  |  |  |  |  |  |  |  |  |  |  |
| Generate “smoothed” meshes from point cloud data to remove textures/defects/etc. |  |  |  |  |  |  |  |  |  |  |  |  |
| Simulate different physically realistic scenarios while using domain randomization |  |  |  |  |  |  |  |  |  |  |  |  |
| Explore ray tracing and other geometric methods to automatically label simulated data |  |  |  |  |  |  |  |  |  |  |  |  |
| Use combination of simulated and real data to train a CNN to identify new object class |  |  |  |  |  |  |  |  |  |  |  |  |

**Budget:**

Faculty Salaries 33,863

Engineer Salary 50,000

Student Tuition 44,577

Student Stipend 34,455

Benefits 39,298

**Total Personnel Costs 185,455**

Computing 2,281

Equipment 15,000

**Total Operating Expense 17,281**

**Total Travel Expense 7,578**

**Total Direct Cost 210,314**

**Facilities & Administration 104,414**

**Total Project Cost $314,728**