Summer school

Robotics Multimodal Deep Learning for Object Recognition

Yupei Wu, He Liu

Department of Computer Science
Tsinghua University

Outline

Background

Tsinghua University

Robots





Background

Tsinghua University

KINECT sensor to obtain data

RGB-D dataset

RGB-D sensors are ubiquitous in many robotic systems. They are inexpensive, widely supported by open source software, do not require complicated hardware and provide unique sensing capabilities.

Background

Target:

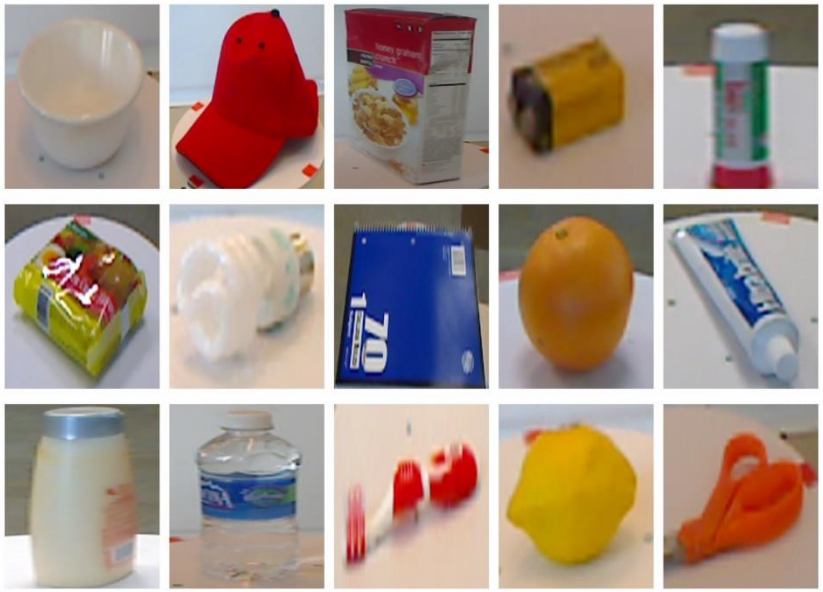
Design a multimodal (combine RGB and depth) network architecture recognition.

Outline

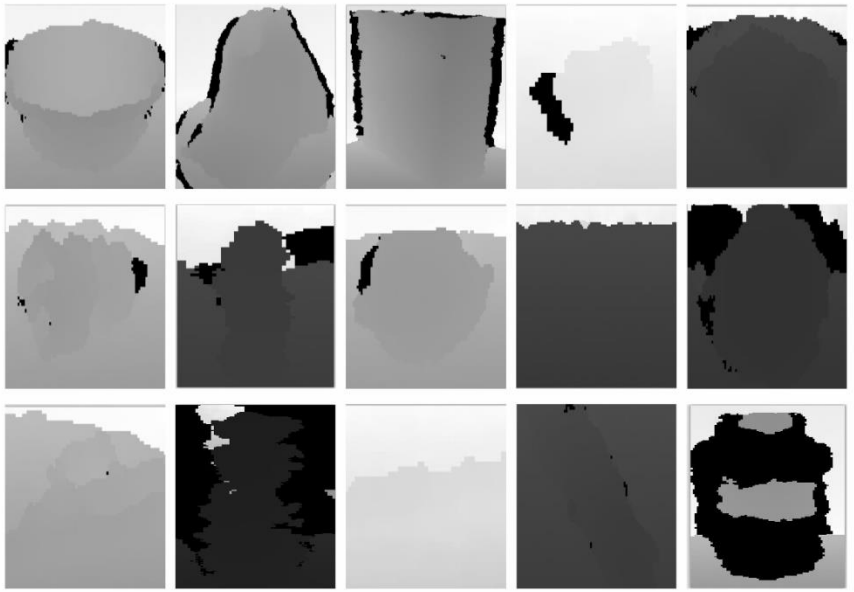
1. Background
2. **RGB-D Dataset**
3. Related technology
4. Network Architecture
5. Points and bonus

RGB-D dataset

RGB images



Depth images



15 classes totally

Outline

Related technology

Tensorlayer is a Deep Learning and Reinforcement Learning library extended from Google TensorFlow.

Advantages:

- **Easily customized**
- **Assembled for tackling real-world machine learning problems**

Outline

1. Background
2. RGB-D dataset
3. Related technology
4. **Experiments**
5. Points and bonus

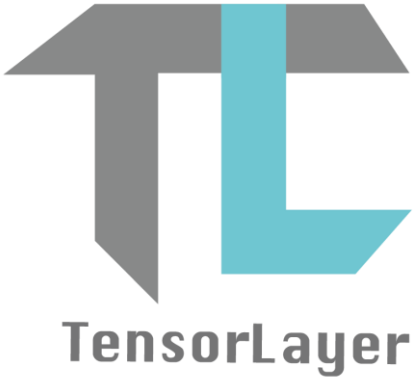
Experiment

- **Target:**
  - multimodal(combine RGB and depth) classification

- **Task:**
  - Design your own net for single model(RGB and depth apart) classification
  - Analyze the single model result(give your loss, acc…)
  - Design your own network for multimodal classification

- **Data:**
  - Dataset: Washington RGB-D
  - Classes: 15 classes of different objects, both have RGB and depth images

Experiment tools

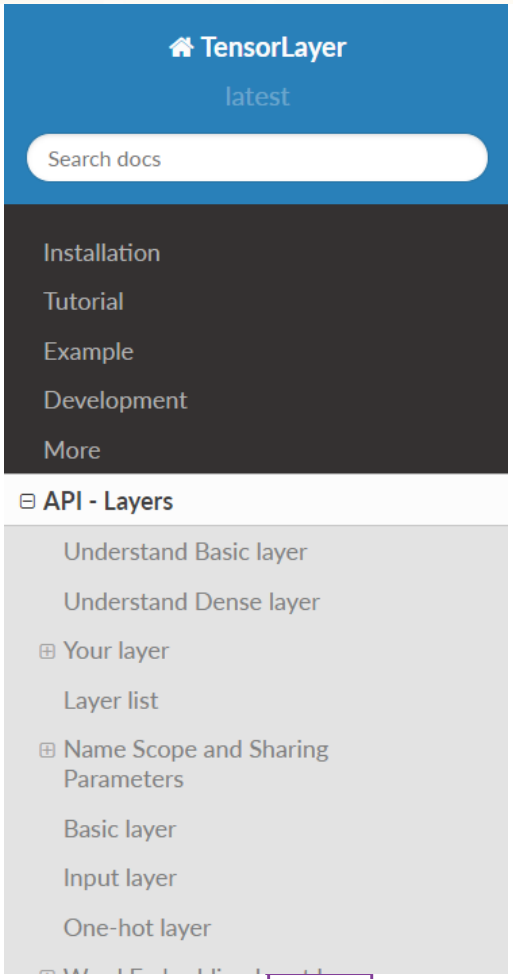


Install:
pip install tensorlayer

Reference:
http://tensorlayer.readthedocs.io/en/latest/index.html#

Experiment details

- **How to design your own net?**
  - Open tensorlayer website
  - Choose API-Layers， all layers supported by tensorlayer are listed
  - If you don’t know how to use them, see the example of mnist

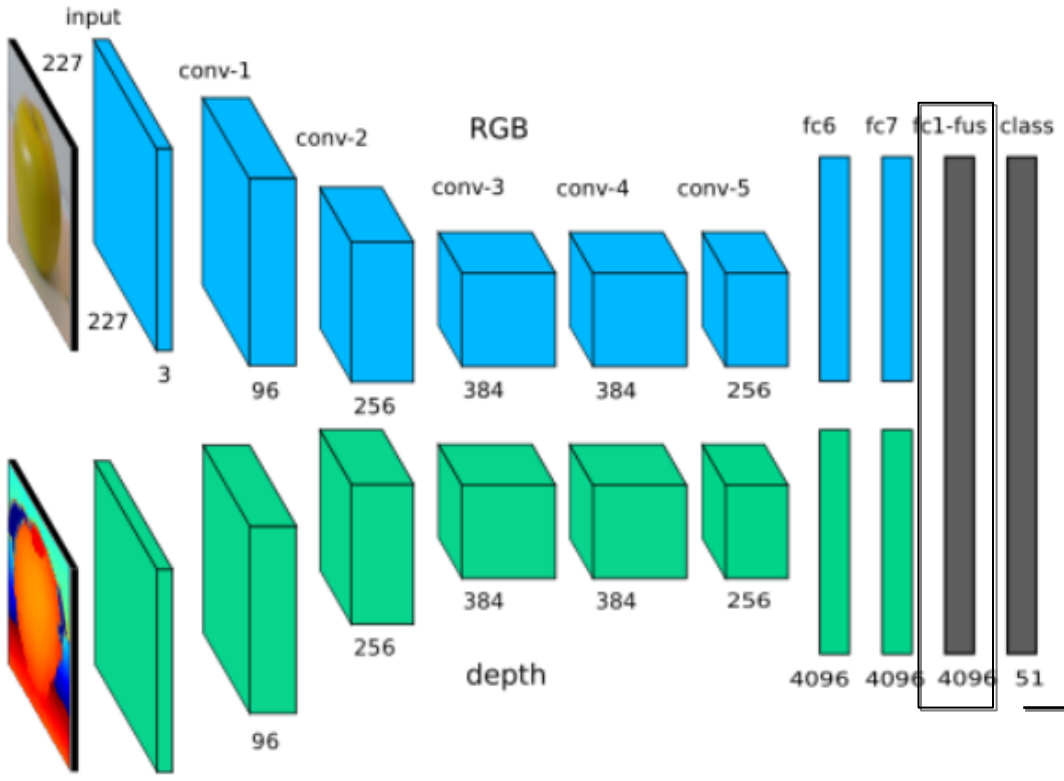


Experiment details

- **How to combine different model together?**

RGB net

Depth net



Combine (concat layer)

Just a example of network architecture

Outline

Points and Bonus

Points:

1.Train the two individual stream networks using RGB image and depth image and give the accuracy.

50 points

2.Train the fusion network and give the accuracy

45 points

3.Analyze the result and give the confusion matrix if time is enough.

5 points

4.Give the brief report to show your work.

Points and Bonus

Bonus:

The more tricks you use to arise your accuracy of the network, the higher points you will get.

Tricks:

1. Various ways of data augmentation
2. More reasonable Network

.....

Points and Bonus

- **Where to get the presentation slides?**
  - **https://github.com/liuhe1305/Summer-school-of-THU**
- **How to download dataset, loadData.py, main.py?**
  - **https://github.com/liuhe1305/Summer-school-of-THU**
- **Send your report to the Email address:**
  - **liuhe_work@126.com before 5:00 p.m today(20th July).**

THANK YOU