

Research Background

My research interests are machine learning and computer vision. I mostly focus on **computer vision** and **machine learning**. I also devote myself on other potential combinations with computer vision, including **robotics** and **natural language processing**.

Breakfast Savior

Language: Java

Topic: Android App development

We develop an android app, which is an ordering system. Users can order from home, from the bus or while stuck in traffic and get the breakfast right away! The breakfast supplier can receive the order from the clients and broadcast advertisement to all the clients, too.



Video Title Generation

Framework: Tensorflow; Language: Python

Our goal is to teach a machine to automatically generate a title in natural language from a user generated video, average about 45 seconds. We first create a highlight detector to find where the highlight clip is, and use the feature of the chosen clip to initialize the hidden state in LSTM cell. Finally, generate the title from the output of the LSTM cell. We also try many methods to increase our performance, including spatial attention, sentence augmentation, using C3D feature and so on. My main contribution is at the **spatial attention** and **sentence augmentation**.

Combine Faster-RCNN with Tracker:

Our idea is to use object detection (faster-rcnn) to detect where the objects are, and we extract 5 boxes per frame, which forms a feature pool. Since we focus on objects in the video, we have to use a tracker to produce tracklet of the objects. Finally, we use the weighted sum of the feature as the LSTM input.

(I have write a **tutorial about training faster rcnn on github**, please refer to

<https://github.com/andrewliao11/py-faster-rcnn>)



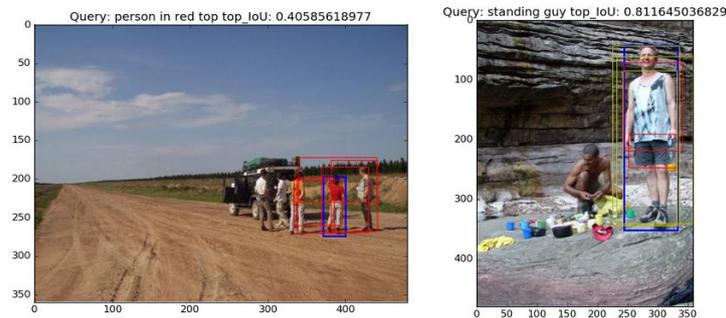
Figure 1 The visualization of the faster rcnn. I have a github tutorial on this topic

Grounding via natural language

Framework: Torch; Language: Lua

In real AI, you may refer something with its location, color, or other characteristics. The AI

robot should know where the corresponding object is. Inspired by the Ronghang Hu's work, I re-implemented the Hu's work and furthermore, **utilizing multi-task concept and region proposals network to boost its performance**. This work is done during my UmboCV internship.



Deep Reinforcement Learning

I'm right now conducting the research on **deep reinforcement learning** at Computational Intelligence Technology Center in Industrial Technology Research Institute. I like to study the algorithms in reinforcement learning and I have shared some of **my survey on github** (<https://github.com/andrewliao11/Deep-Reinforcement-Learning-Survey>). Besides studying, I'll share my survey to the members (more than 30 people) in my department.



Video Question Answering

Framework: Tensorflow; Language: Python

We extend the visual question answering to video question answering. Since the current visual question answering dataset is still not large enough, we propose **an automatically question generator** from video description crawled from the internet. This method is able to scale up to a very large size. However, the auto-generated data isn't perfect. We propose a novel way called **self-pace learning** to handle the noise from the auto-generated QA dataset. **This paper won the runner-up of Digital Drift Best Paper on Deep Learning for Visual Analysis, CVGIP 2016.**

Side Projects

Besides the vision-language research area, I also devote to open source development. I love to implement something cool and novel.

- DNI-tensorflow on github
- Natural-Language-Object-Retrieval-tensorflow on github