Yuan-Hong Liao

National Tsing Hua University

Hsinchu City 30010, Taiwan

andrewliao11@gmail.com

Research Background

Most of my work is related to computer vision and machine learning (especially deep learning).

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.

Handcam Recognition

Framework: Caffe; Language: Python

Topic: Image Classification

Our goal is to fine-tune on the pre-trained Imagenet network. We want to apply the techniques to the IoT. We teach a machine that can recognize the objects around our daily life, like kettle, phone, cup and so on. Our data is from the camera on the user's hands, which is why we call it "Handcam Recognition"

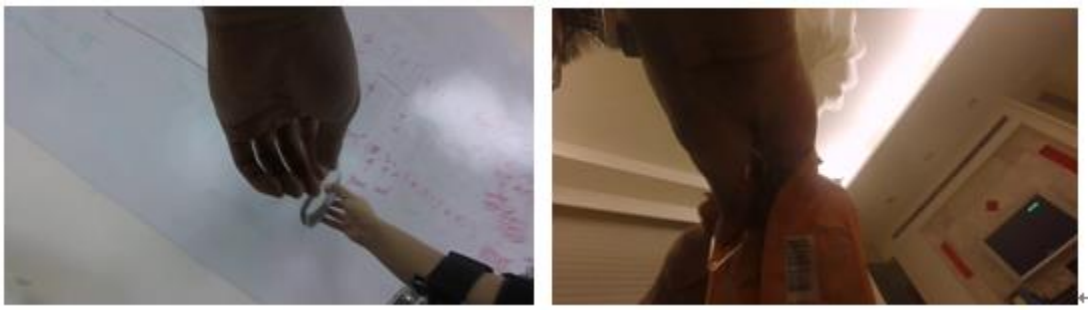


Figure 1 The left picture is taken in our lab, while the right is taken in a house.

Method: data augmentation, and tune the hyper parameters in the network

Video Title Generation

Framework: Tensorflow; Language: Python

Our goal is to teach a machine to automatically generate a title in natural language from a personal video, about 45 seconds. We first create a highlight detector to find which the highlight clip is, and use the feature of the chosen clip to initialize the hyper parameter in LSTM model. Finally, generate the title from the output of the LSTM model. 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 part and sentence augmentation.

Combine Faster-RCNN with Tracker:

Our idea is to use object detection to extract where the objects are, and we extract 5 boxes per frame, which forms a feature pool. We train the faster rcnn on Imagenet 200 categories and use the tracker to visualize the video object detection. Then we use the weighted sum of the feature as the input of the LSTM.

(I have write a tutorial about training faster rcnn on github, please refer to https://github.com/andrewliao11/py-faster-rcnn)

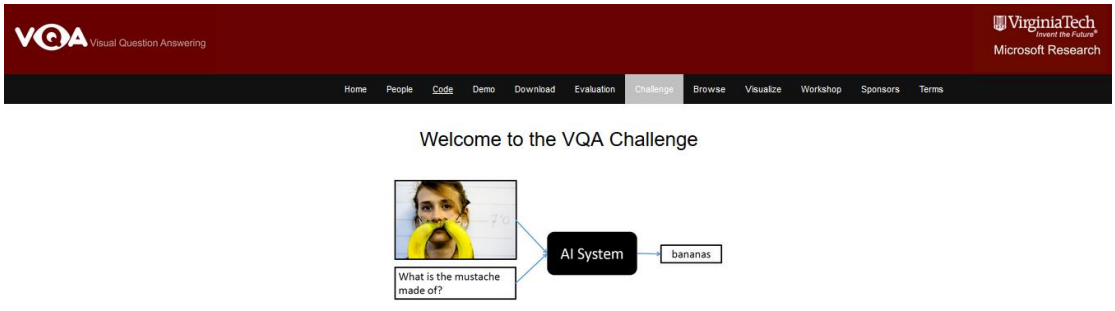
Figure 2 The visualization of the faster rcnn. I have a github tutorial on this topic

Visual Question Answering **[ongoing]**

Framework: Tensorflow; Language: Python

We attempt to extend the image/video title generation to image/video question answering. Being able to generate the caption or title implicitly means the machine can understand the image/video. So, we use the knowledge of this to develop a question answering machine.

We are going to participate the Visual Question Answering Challenge held by Virginia Tech on May 27, 2016.



Deep Reinforcement Learning

I’m right now an intern in Industrial Technology Research Institute (ITRI) at Computational Intelligence Technology Center. We mainly focus on deep reinforcement learning and apply it to the real product.