Hand Gesture Recognition

Miniaturized for Raspberry Pi

Overview

A program to identify hand gestures in real time, lightened to the point it can be run on a raspberrry pi unit connected to a camera, ultimatly focusing on rock paper sissors gestures. This could be used as groundwork for a number of uses including hand gesture comands, digital sign language translation in real time, and creating a program that will finally prove that that one friend cheats at rock paper sissors.

Application

- Electronics can be remoted by hand gesture.
- Example:
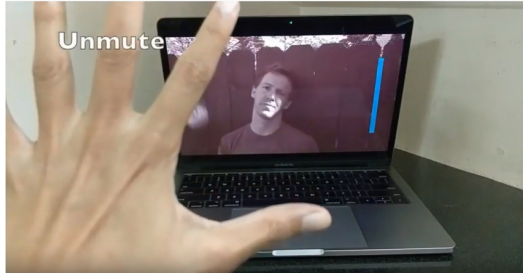
https://www.youtube.com/watch?v=-_9WFzgI7ak&feature=emb_rel_pause

Application

- Example:

Why Raspberry Pi?

- Portable
- Low cost
- Good tools for studying computer vision on IoT

	IoT	Cell Phone	Laptop	Desktop	Server
Model	Rasp. 3B	iPhone 11	Apple MacBook Air	Dell Inspiron Tower	Dell PowerEdge T340
CPU Freq.	1.2 GHz	2.65 GHz	1.6 GHz	3.0 GHz	3.3 GHz
Memory Size	1 GB	4 GB	8 GB	16 GB	32 GB
Size	85.60 × 56.5 mm	150.9 x 75.7 x 8.3 mm	304 x 212 x 15.6 mm	14.7 x 6.3 x 11.6 inches	16.9 x 8.6 x 23.7 inchs
Price	43 USD	700 USD	1100 USD	890 USD	2900 USD

Challenges

- The trade-off of Speed and Accuracy for inference.

    - How many layers/neurons for the CNN?

    - How to simply the model as much as possible?
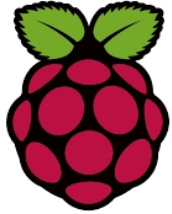
    - How many images for training?

Hardware Specification

Raspberry Pi 3 Model B

- Quad Core 1.2GHz Broadcom BCM2837 64bit CPU

- 1GB RAM

- Micro USB power supply (2.1 A)

- Equivalent to Roku streaming devices, cell phones

Software Reuqirement

Raspberry Pi 3 Model B

● OS: NOOBs 3.2.1

● Tensorflow 2.0

● Keras

● OpenCV4

● Python 3.7

Please check attacment for installation Instruction.

Foxit Reader PDF Document

Dataset Descritpion

- Three different hand gestures: paper, scissors, rock
- Image size: 80x60
- Color: gray scale
- Each gesture has 2000 images.
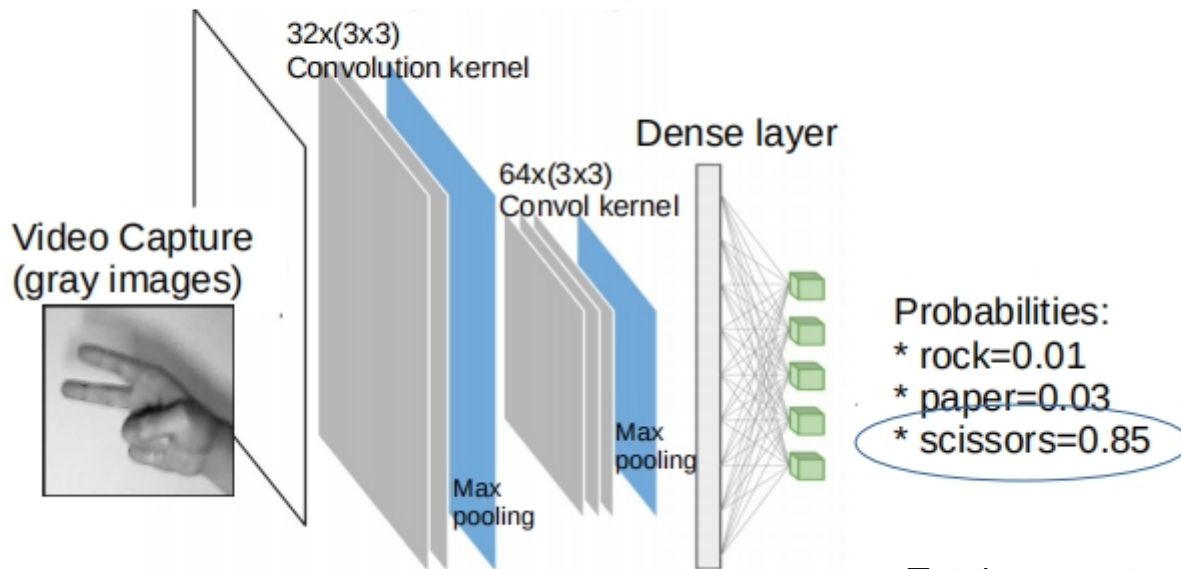

paper

scissors

rock

Question

- Will these hand gesture confuse the model?

paper, scissors, rock?

Model Description

Simple CNN: 2 convolutional, max pooling layers

Video Capture (gray images)

32x(3x3) Convolution kernel

64x(3x3) Convol kernel

Dense layer

Max pooling

Max pooling

Probabilities:
* rock=0.01
* paper=0.03
* scissors=0.85

Total parameters: 1,936,259

Memory: 399 MB

Model Training

- Model is trained on Google Colab with GPU.
- Total images of Training/Testing: 4800/1200
- Training time: < 5min
- Accuracy: 97%

```
Epoch 1/5
100/100 - 23s - loss: 0.8732 - accuracy: 0.6040 - val_loss: 0.5373 - val_accuracy: 0.8263
Epoch 2/5
100/100 - 22s - loss: 0.3731 - accuracy: 0.8600 - val_loss: 0.3205 - val_accuracy: 0.8706
Epoch 3/5
100/100 - 22s - loss: 0.2343 - accuracy: 0.9155 - val_loss: 0.1310 - val_accuracy: 0.9606
Epoch 4/5
100/100 - 22s - loss: 0.1116 - accuracy: 0.9645 - val_loss: 0.0958 - val_accuracy: 0.9719
Epoch 5/5
100/100 - 22s - loss: 0.0927 - accuracy: 0.9720 - val_loss: 0.0888 - val_accuracy: 0.9719
```

Hand Detection

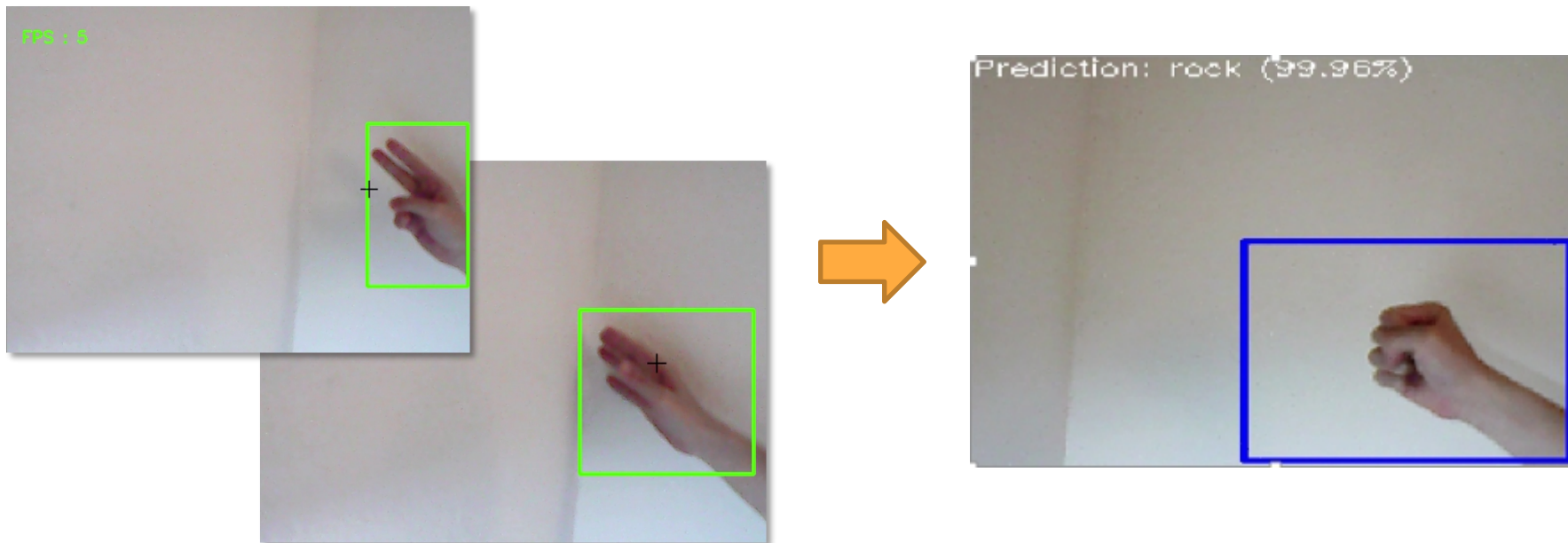- SSD with Mobilenet _v1_coco

- Latency 5 sec. on Raspberry Pi

COCO-trained models

Model name	Speed (ms)	COCO mAP[^1]	Outputs
ssd_mobilenet_v1_coco	30	21	Boxes
ssd_mobilenet_v1_0.75_depth_coco ☆	26	18	Boxes
ssd_mobilenet_v1_quantized_coco ☆	29	18	Boxes
ssd_mobilenet_v1_0.75_depth_quantized_coco ☆	29	16	Boxes
ssd_mobilenet_v1_ppn_coco ☆	26	20	Boxes
ssd_mobilenet_v1_fpn_coco ☆	56	32	Boxes
ssd_resnet_50_fpn_coco ☆	76	35	Boxes
ssd_mobilenet_v2_coco	31	22	Boxes
ssd_mobilenet_v2_quantized_coco	29	22	Boxes
ssdlite_mobilenet_v2_coco	27	22	Boxes
ssd_inception_v2_coco	42	24	Boxes
faster_rcnn_inception_v2_coco	58	28	Boxes
faster_rcnn_resnet50_coco	89	30	Boxes
faster_rcnn_resnet50_lowproposals_coco	64		Boxes
rfcn_resnet101_coco	92	30	Boxes
faster_rcnn_resnet101_coco	106	32	Boxes
faster_rcnn_resnet101_lowproposals_coco	82		Boxes
faster_rcnn_inception_resnet_v2_atrous_coco	620	37	Boxes
faster_rcnn_inception_resnet_v2_atrous_lowproposals_coco	241		Boxes
faster_rcnn_nas	1833	43	Boxes
faster_rcnn_nas_lowproposals_coco	540		Boxes
mask_rcnn_inception_resnet_v2_atrous_coco	771	36	Masks
mask_rcnn_inception_v2_coco	79	25	Masks
mask_rcnn_resnet101_atrous_coco	470	33	Masks
mask_rcnn_resnet50_atrous_coco	343	29	Masks

https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md

Solutions

● In order to solve latency issue, we use fixed boundary box instead of SSD
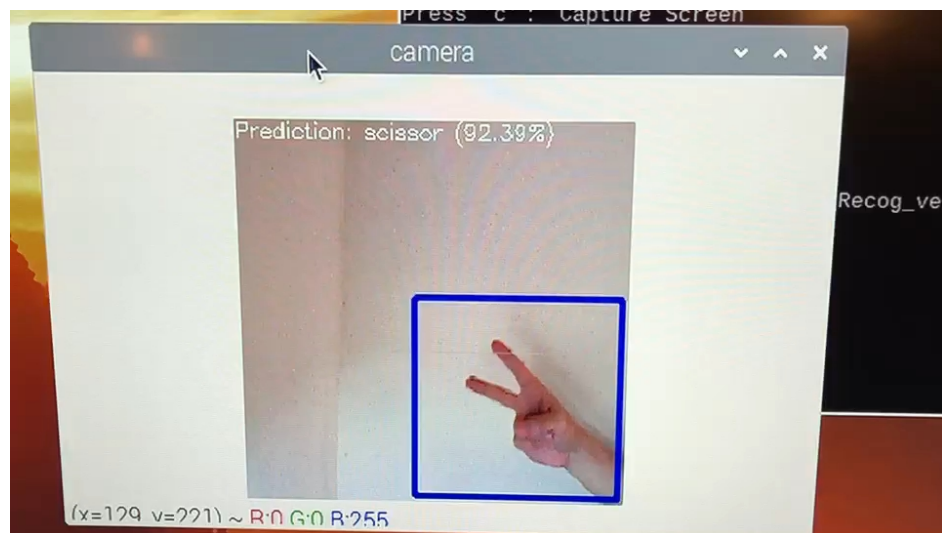
Demo

On laptop

Demo

On Raspberry Pi

Summary

- The accuracy of training/validation can reach 97%
- Very simple CNN (2 Conv. Layers only)
- Model is trained by 1600 images for each hand gesture.
- Image size: 80 x 60; Image color: gray scale
- Following action: solving background issue.

Resources

- https://towardsdatascience.com/training-a-neural-network-to-detect-gestures-with-opencv-in-python-e09b0a12bdf1
- https://colab.research.google.com/github/filipefborba/HandRecognition/blob/master/project3/project3.ipynb
- https://colab.research.google.com/github/zaidalyafeai/Notebooks/blob/master/tf_handBbox_esitmation.ipynb