



# CVGesture

## Performance Report

2018-02-07

**OPEN AI LAB**

## Revision Record

| Date       | Rev   | Change Description | Author       |
|------------|-------|--------------------|--------------|
| 2017-10-13 | 0.1.0 | QA Test            | Luo Hao      |
| 2017-12-25 | 0.2.0 | QA Test            | Luo Hao      |
| 2018-01-25 | 0.3.0 | QA Test            | Luo Hao      |
| 2018-02-07 | 0.3.1 | QA Test            | Shulian Shen |

# catalog

|  |           |
|--|-----------|
| <b>1 INTRODUCTION .....</b>                  | <b>3</b>  |
| <b>2 TEST ENVIRONMENT .....</b>              | <b>3</b>  |
| <b>3 DEFINITION OF GESTURES .....</b>        | <b>3</b>  |
| <b>4 PERFORMANCE ON DIFFERENT CORES.....</b> | <b>5</b>  |
| 4.1 SIMPLE BACKGROUND .....                  | 5         |
| 4.2 COMPLEX BACKGROUND .....                 | 6         |
| <b>5 ACCURACY TEST .....</b>                 | <b>7</b>  |
| <b>6 STRESS TEST .....</b>                   | <b>10</b> |
| <b>7 CONCLUSION.....</b>                     | <b>10</b> |
| <b>8 TESTING ISSUES .....</b>                | <b>10</b> |

# 1 Introduction

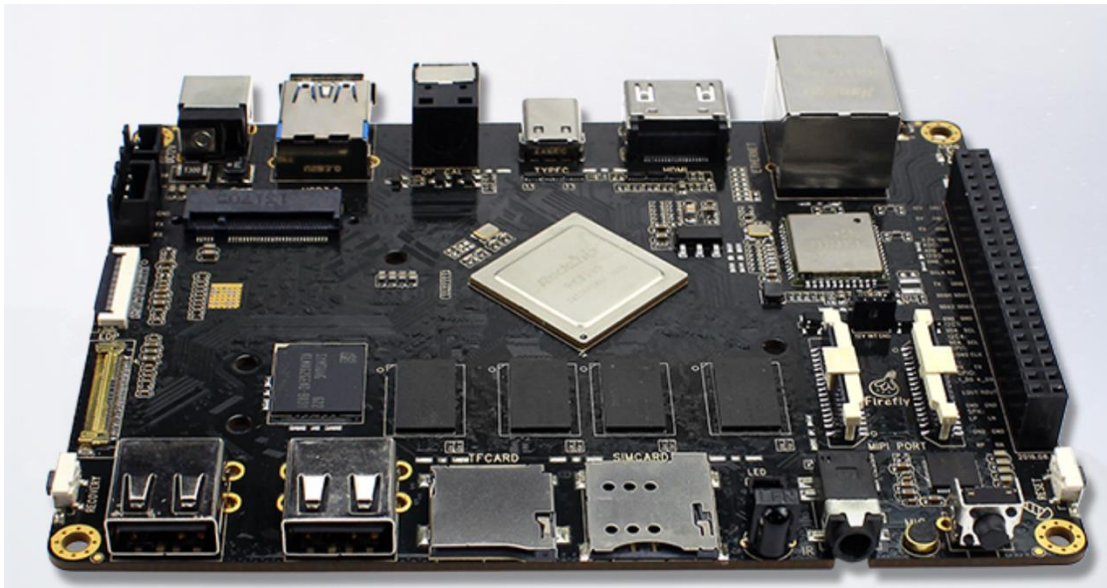
This report is tested on RK3399 platform. The report includes performance and accuracy information.

## 2 Test Environment

## Hardware SoC: firefly

<http://www.t-firefly.com/product/rk3399.html>

- GPU: Mali T864 (800MHz)
- RAM: 2G
- CPU: Dual-core Cortex-A72 up to 2.0GHz (real frequency is 1.8GHz); Quad-core Cortex-A53 up to 1.5GHz (real frequency is 1.4GHz)
- Operating System: Ubuntu 16.04
- Software: OpenCV 3.3.0



### 3 Definition of Gestures

In order to test the performance of the application, there should be some definitions and limitations to different gestures:

- **Palm:** with five fingers open (do not close five fingers together, as fig.3), facing directly to the camera (as fig.1). The angle of rotation of hand in the direction of **front and back should not go beyond 30 degrees**. The angle of rotation of hand in the direction of **left and right should not go beyond 45 degrees**. The standard is the same for both left and

right hand.

- **Fist:** with fist clenched, facing directly to the camera (as fig.2). The angle of rotation of hand in the direction of **front, back, left and right should not go beyond 30 degrees**. The standard is the same for both left and right hand.



Figure 1. Palm facing directly to camera



Figure 2. Fist facing directly to camera



Figure 3. Wrong palm gesture

## 4 Performance on Different Cores

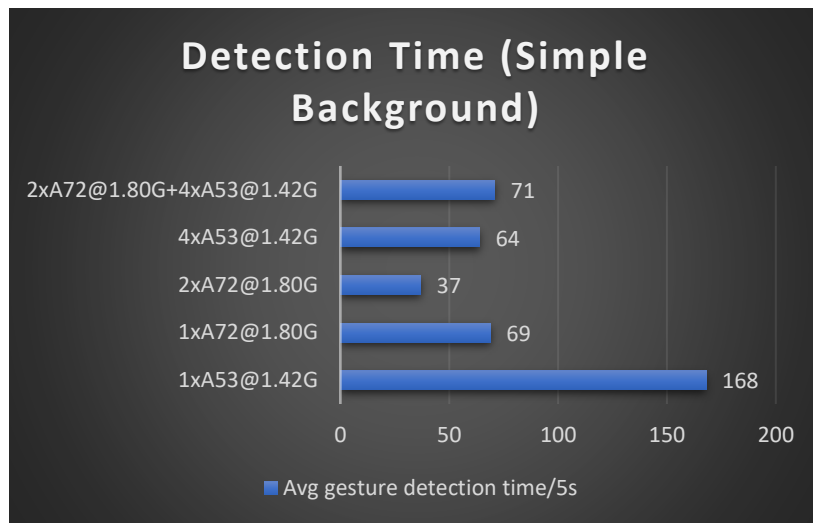
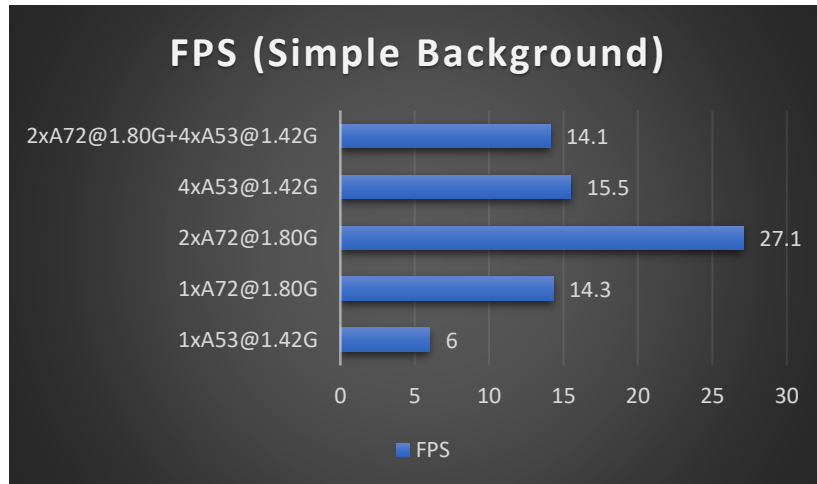
Calculate the FPS (Frames Per Second) in five seconds, and print the results in terminal. Skip the results of first five seconds. Recognize two gestures: palm and fist. The palm and fist detection time are also averaged in five seconds, skip the results of first five seconds. As the complexity of background will significantly affect the performance, the performance under different backgrounds will be given.

### 4.1 Simple background

Simple background means that the background does not contain object of irregular shape like a plain wall. Performance under simple background will be stronger.

FPS on different cores, 640x480 resolution

|                             | FPS  | Avg. gesture detection time (ms) in 5 seconds |
|-----------------------------|------|---|
| <a href="#">1xA53@1.42G</a> | 6    | 168   |
| <a href="#">1xA72@1.80G</a> | 14.3 | 69.   |
| 2xA72@1.80G                 | 27.1 | 37  |
| 4xA53@1.42G                 | 15.5 | 64  |
| 2xA72@1.80G+4xA53@1.42G     | 14.1 | 71  |

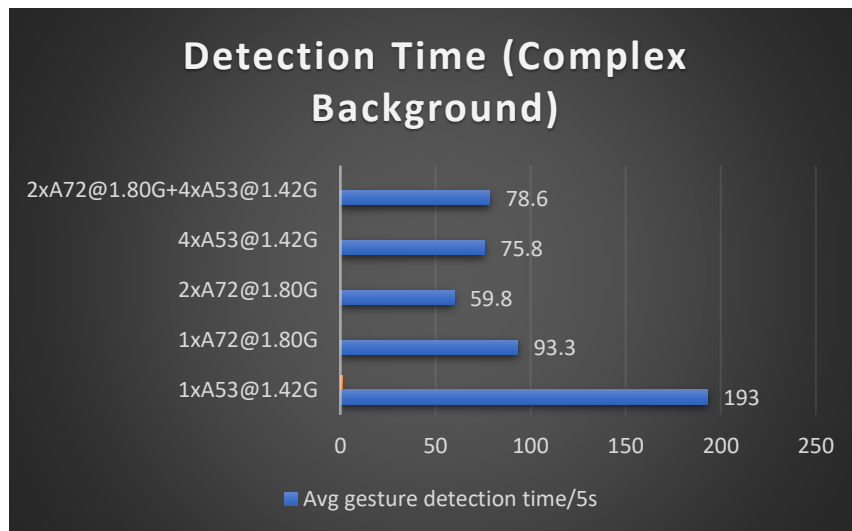
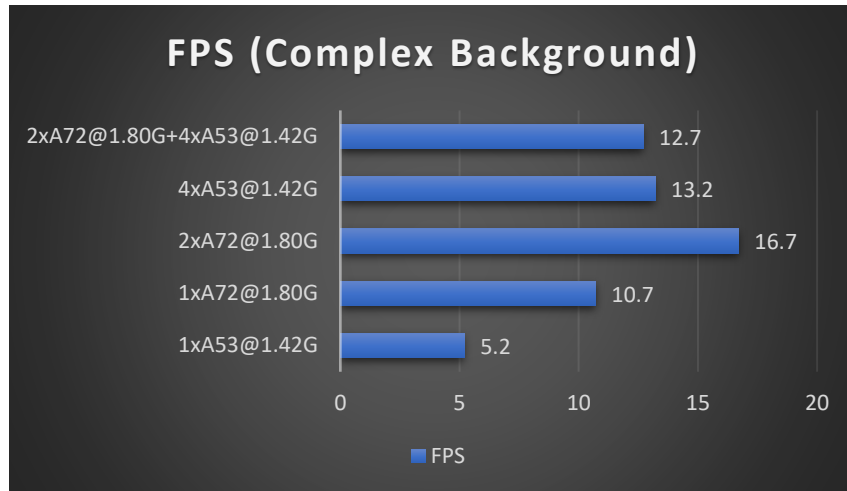


## 4.2 Complex background

Complex background means that the background contains many objects of irregular shape. Performance under complex background will be weaker.

FPS on different cores, 640x480 resolution

|                         | FPS  | Avg. gesture detection time (ms) in 5 seconds |
|-------------------------|------|---|
| 1xA53@1.42G             | 5.2  | 193   |
| 1xA72@1.80G             | 10.7 | 93.3  |
| 2xA72@1.80G             | 16.7 | 59.8  |
| 4xA53@1.42G             | 13.2 | 75.8  |
| 2xA72@1.80G+4xA53@1.42G | 12.7 | 78.6  |



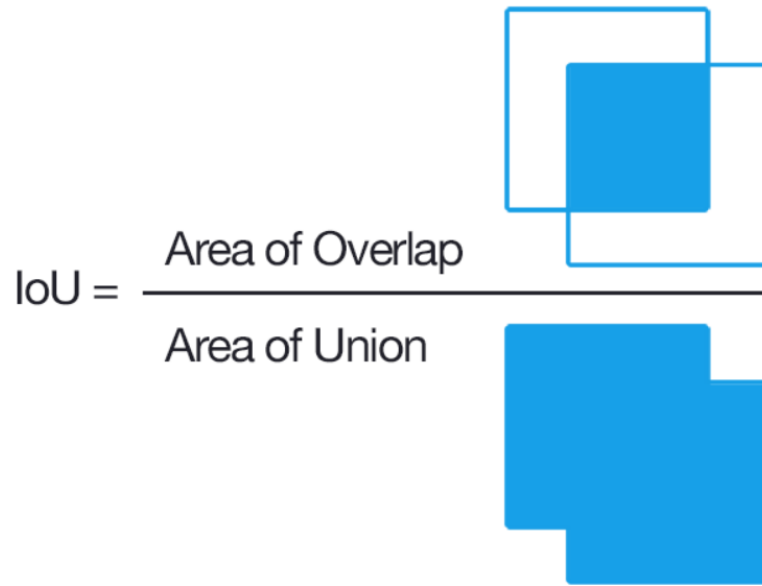
## 5 Accuracy test

To test accuracy of the classifiers, a small number of images of palms and fists are collected from the internet and labeled with ground truth. Following indices are used to evaluate detection and recognition accuracy:

For detection:

- IoU: Intersection over Union, intersection means the overlap area of detection and ground truth and union mean the union area of detection and ground truth (as shown below)





For recognition (as shown below):

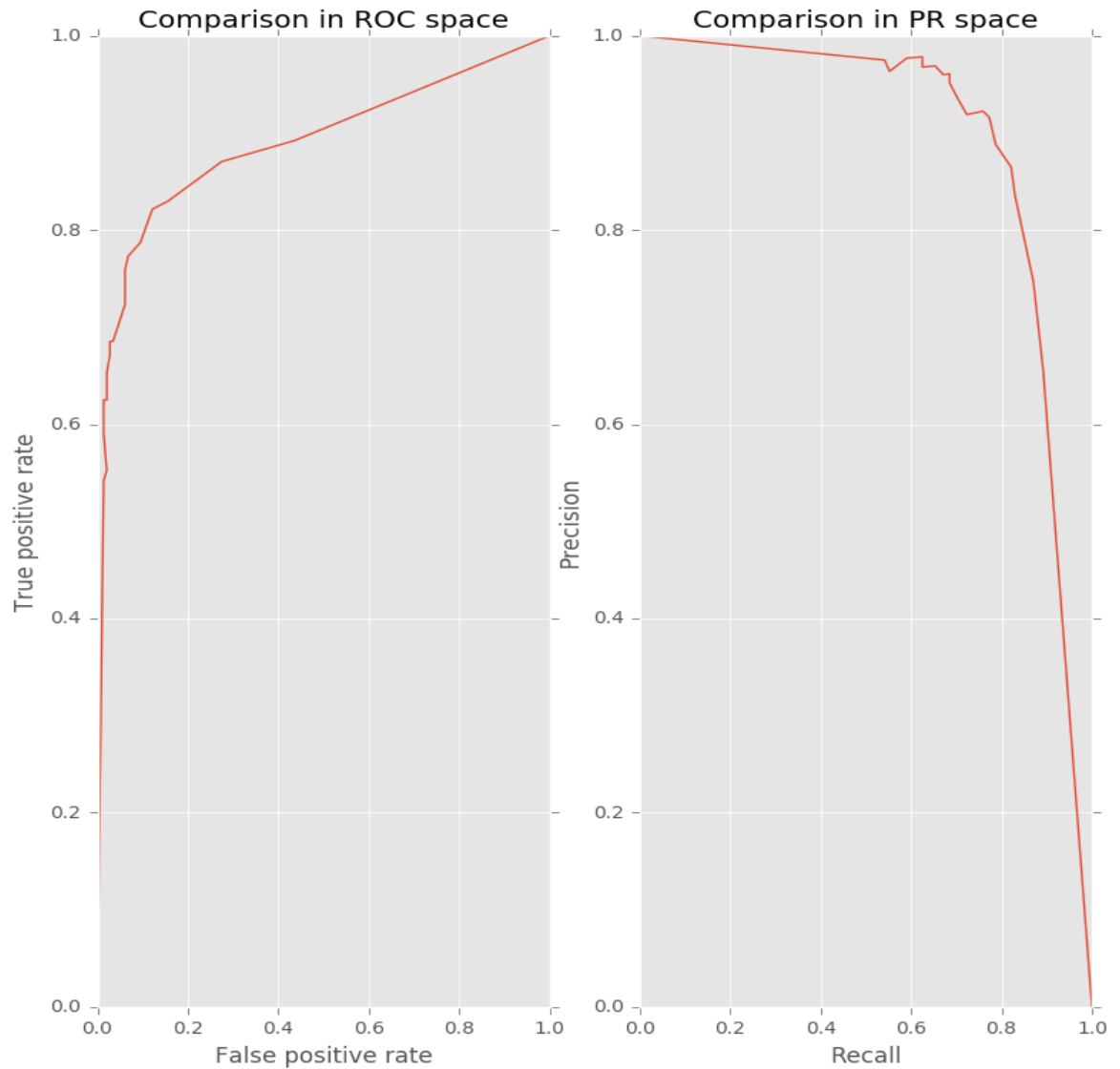
- Precision: Ture positives / (Ture positives + False positives)
- Recall: Ture positives / (Ture positives + False negatives)
- Accuracy: (Ture positives + Ture negatives) / Total
- Ture positives: IoU of recognition  $\geq 0.5$  and the label of class is correct

|                           |          | <u>True class</u> |                 |  |                          |
|---------------------------|----------|-------------------|-----------------|--|--------------------------|
|                           |          | <b>p</b>          | <b>n</b>        |  |                          |
| <u>Hypothesized class</u> | <b>Y</b> | True Positives    | False Positives | fp rate = $\frac{FP}{N}$                                     | tp rate = $\frac{TP}{P}$ |
|                           | <b>N</b> | False Negatives   | True Negatives  | precision = $\frac{TP}{TP+FP}$                               | recall = $\frac{TP}{P}$  |
| Column totals:            |          | <b>P</b>          | <b>N</b>        | accuracy = $\frac{TP+TN}{P+N}$                               |                          |
|                           |          |                   |                 | F-measure = $\frac{2}{1/\text{precision} + 1/\text{recall}}$ |                          |

The accuracy results of the classifiers are given below, the results are averaged from 20 different confidence level (setting *minNeighbors* parameter in *cv.detectMultiScale* from 1 to 20) and the min:

| Avg. IoU | Avg. Precision | Avg. Recall | Avg. Accuracy |
|----------|----------------|-------------|---------------|
| 0.75     | 0.90           | 0.70        | 0.81          |

The ROC curve and PR curve



## 6 Stress test

The gesture detection passed over night (about 16 hours) test with different gestures in complex background.

## 7 Conclusion

From the above test cases, we can deduce that:

- the performance on 2xA72 is the best
- the performance on 4xA53+2xA72 is similar versus 4xA53

## 8 Testing Issues

There are many known factors that will significantly affect the performance of the application:

- The **version of OpenCV should be 3.3.0**, using OpenCV2 will slow down the detection
- The performance of the application depends on the complexity of background, running under **simple background like a plain wall** will be significantly faster