DEEPCRAFT™ Ready Model for Gesture Detection

Introduction

In this document, we describe the DEEPCRAFT™ Ready Model for Gesture Detection, a radar-based AI model developed by Imagimob, an Infineon Technologies company. This model detects when a person is performing the following five hand gestures in front of a radar sensor:

- Push
    - Open palm moving horizontally in parallel to the ground from the user's body towards the radar sensor

- Orthogonal Swipes
    - Swipe Left - open palm moving horizontally in parallel to the ground from right side to left side of the radar sensor
    - Swipe Right - open palm moving horizontally in parallel to the ground from left side to right side of the radar sensor
    - Swipe Up - open palm moving vertically, perpendicular to the ground, from the bottom side to the top side of the radar sensor
    - Swipe Down - open palm moving vertically, perpendicular to the ground from the top side to the bottom side of the radar sensor

We provide details about the technical specifications of this machine learning model, its performance in common scenarios, and various test results for the model including the real-time testing on an Infineon PSOC™ 6 board connected to an Infineon XENSIV™ 60GHz radar sensor.

Contents

Introduction.....1
Model Specification.....2
Model Overview.....2
Expected Performance.....2
Operations.....3
Model Tech Specs and Deployment.....4
Data Properties.....4

Testing.......... 4
Validation Set Results.......... 4
On-device testing..........6
Appendix I - Radar Configuration..........9
Appendix II - Results of On-Device Model Testing..........9

Model Specification

Model Overview

The DEEPCRAFT™ Ready Model for Gesture Detection is designed to detect in real-time the above-mentioned five gestures with the purpose of enhancing and/or assisting human-machine interaction and experience. Such a model is developed with the intent to run, for instance, on a monitor or on any device connected to a monitor in order to provide contactless control. The user is expected to stand or sit in front of the radar sensor without any obstacle in between. The model will not distinguish between fast or slow gestures, diagonal swipes, right-handed versus left-handed gestures. The model's predictions will be the input of an application used to interact with any device.

Expected Performance

The aim of this model is to detect the user's hand gestures, defined as push and swipe movements, performed with one hand. The hand movement is expected to:

- last between one and two seconds
- occur within 10 cm and 70 cm from the radar sensor
- be within an angle of 10 degrees in the field of view of the radar sensor

The model is expected to detect 92% of gesture events for a standing or sitting user. A gesture is expected to be misclassified or to trigger one of the other gestures fewer than five times out of every 100 gesture events. This performance can significantly increase with the user's practice and experience in using the model.

Besides different distances, angles and velocities, the model is expected to be robust against different hand shapes, arm lengths and user heights, as well as hand and body movements different from the actual gestures.

Operations

The DEEPCRAFT™ Ready Model for Gesture Detection is designed to detect the above mentioned five gestures to empower the user with touchless control of any device. Missed detections or unwanted triggers may occur when the five gestures are performed in the following ways:

- faster or slower than 1-2 seconds
- outside the 10-70 cm range from the radar sensor
- more than 10 degrees from the radar sensor's field of view
- with obstacles between the user's hand and the radar sensor
- when holding an object, especially one with moving parts
- when wearing clothes with long, hanging sleeves
- when wearing a hanging bracelet or a gadget close to the hand

In addition to the relative positioning of the radar sensor and the user, the accuracy of the gesture model strongly depends on the way the user moves their hand and arm. Any movement which resembles a push or a swipe will be detected as such. Diagonal hand movements, as well as stopping the gesture halfway, may result in unexpected model output.

This model detects any of the five gestures in its vicinity without distinguishing the person who performed it; this means that the model may detect the gestures of someone else close by. Also, if one or more people are moving in between the radar sensor and the user, the model's output is expected to be less accurate. On the other hand, people moving or walking behind the user might have no impact on model performance.

Performance degradation may occur because of complementary movements that may resemble one of the five gestures. For instance, a swipe down can be triggered after a swipe up or a push, because the user will naturally bring their hand down. Similarly, a swipe left or right can be preceded by a push because of the hand movement towards the radar sensor that the user has to make to engage with the radar sensor.

The model is in general robust against body movements as long as they are not the same speed as a gesture, although one of the gestures might be detected when a user is sitting, standing, bending down quickly, jumping, or running.

It is recommended that this model should be used to enhance the user's overall experience in controlling a device/system rather than focusing on each single gesture detection.

Model Tech Specs and Deployment

The DEEPCRAFT™ Ready Model for Gesture Detection is able to detect a gesture from radar sensor data, parsed on device, with the following characteristics:

- Features: range, velocity, azimuth, elevation, magnitude
- Features sample rate: 33 Hz

For more details about the radar configuration see Appendix I.

The model's C library has the following memory footprint:

- Flash: 45 kB
- RAM: 3 kB

Its inference time, excluding the radar data parsing time, is about 8 ms when running on a XENSIV™ KIT CSK BGT60TR13C with a PSOC™ 6 and a 60GHz radar sensor. The model outputs a prediction every 91 ms, namely 11 times per second.

Data Properties

This Ready Model has been built using positive and negative radar sensor data. The positive data consists of gestures from several individuals with different heights performing gestures at different distances and angles. Most of the participants were right-handed. The negative data consists of different random movements in front of the radar as well as sessions without any movement.

For the collection of all the data, the radar sensor was mounted about 1 m from the ground. The person performing the gestures was standing mostly in front of the radar sensor (azimuth angle zero degrees) and within the range 0.5-1 m from the radar sensor.

Testing

Validation Set Results

The overall accuracy of the model on the Validation set is 96.6% when the output of the model with the highest confidence level or probability is counted as a trigger. The table below is the so-called confusion matrix of the model which shows its performance on the Validation set in more detail. The meaning of the shown percentages is as follows:

- Top Left Value (True Negatives): actual negative/non-gesture data predicted as negative/non-gesture data
- Other Values in “Unlabelled” Column (False Positives): actual negative/non-gesture data predicted as positive/gesture data
- Other Values in “Unlabelled” Row (False Negatives): actual positive/gesture data predicted as negative/non-gesture data
- Other Values in Diagonal (True Positives): actual positive/gesture data predicted as positive/gesture data
- All Other Off-Diagonal Values (False Positives): actual positive/gesture data predicted as another positive/gesture data

The sample-based confusion matrix below is obtained by splitting the data into short windows of about 485 ms. Even if this is a more accurate representation of the model’s performance since it looks at samples short in time, it can overestimate False Negatives. This means that the True Positives percentages in the table below set a roughly lower limit of the real True Positives of this model.

Samples-Based Confusion Matrix	Actual non-gesture	Actual push	Actual swipe down	Actual swipe left	Actual swipe right	Actual swipe up
Predicted non-gesture	97.24%	6.07%	6.59%	4.96%	8.41%	6.13%
Predicted push	0.76%	91.44%	0.39%	0.25%	0.35%	0.10%
Predicted swipe down	0.78%	1.41%	91.22%	0.35%	0.56%	1.72%
Predicted swipe left	0.57%	0.65%	0.36%	93.59%	1.77%	1.32%
Predicted swipe right	0.30%	0.22%	0.47%	0.50%	88.61%	0.96%
Predicted swipe up	0.35%	0.22%	0.97%	0.35%	0.30%	89.77%

In this case, another way to interpret the high value for the False Negatives is that the model may miss gestures which are performed further from and/or not aligned with the radar sensor as well as gestures which are executed with incorrect movements. Typically, this happens more when the user interacts with the radar sensor for the first time.

When looking at the True Positives (diagonal values), we see that the model has the highest detection recall with Swipe Left (93.6%) and lowest with Swipe Right (88.6%). Overall, this means that it is easier to trigger a swipe when moving the hand from right to left than the opposite. This 5% difference or asymmetry can be a consequence of the fact that most of the people who contributed to model building with their data are right-handed. Using the right hand to perform a Swipe Left can indeed be easier, more consistent and more accurate than a Swipe Right, which, has more variation and requires more training data.

On the other hand, we notice that most of the False Positives are below 1% except for four cases: Push detected as Swipe Down (1.41%), Swipe Up detected as Swipe Down (1.72%) or Swipe Left (1.32%), and Swipe Right detected as Swipe Left (1.77%). In all these cases, the False Positives may be triggered by complementary movements that the users do in order to perform a gesture. For instance, the user will put his hand down after a Push or a Swipe Up, or execute a Swipe Right after moving the hand to the left side of the radar sensor.

Overall, by summing up the False Positives per gesture, Swipe Up and Swipe Right have the highest false positives rates with a total of 3.0% and 4.1%, respectively. On the other hand, Swipe Left is expected to have the highest precision since it has in total a false positives rate equal to 1.5%.

On-device testing

To test this model, the following steps need to be done:

1. Obtain the ready model library from Imagimob
2. Obtain a PSOC™ 6 board XENSIV Connected Sensor Kit with a BGT60TR13C radar sensor, such as XENSIV™ KIT CSK BGT60TR13C
3. Use the example project from Imagimob
4. Use the provided API calls and example code in the library header
5. Create a UI for displaying the library outputs. E.g. a printf statement to a terminal

Test Results

We performed on-device testing with 20 people that were not part of the model training. For this test, we deployed the model on a XENSIV™ Connected Sensor Kit board with PSOC™ 6 mounting a BGT60TR13C radar sensor and used a post-processing algorithm that combines the usage of a confidence threshold and the counting of multiple triggers. We summarise the results in the table below. The detailed results of this testing are reported in the table in Appendix II.

Gesture	Recall	Precision
Push	88.1 %	93.3 %
Swipe Down	91.0 %	96.1 %
Swipe Left	93.8 %	97.5 %
Swipe Right	91.7 %	95.0 %
Swipe Up	94.4 %	94.4 %
TOTAL	91.8%	95.3%

The participants had different heights ranging from 160 cm till 190 cm, and they were all right-handed. Most importantly, for the majority of the participants it was the first time they interacted with a radar gesture model.

To test the model on-device, we equally divided the participants into groups depending on the positioning of the radar sensor and the positioning of the person's body and height. In particular, the test has been run with the following conditions:

- Person standing or sitting
- Radar mounted at about 1 m from the ground
- For sitting, the radar sensor placed on a table
- Two radar sensor-to-person's body distances: about 1 m and 0.5 m
- Gestures performed in front of the radar (zero degrees azimuth angle)

Each participant performed each gesture several times in a row, with and without a person moving close by (see details in tables in Appendix II).

The total recall value for this test turned out to be 91.8%, the total precision was 95.3%, namely 5% of misclassified events. By looking at the performance per gesture, this test shows that the model is more likely to detect Swipe Up and Swipe Left events than Swipe Down and Push. The overall detection increases if the person is sitting close to the radar (see detailed results in Appendix II). In general, the further the user stands or sits, the higher the number of missed gestures.

The performance of the model running on-device also shows that Swipe Left (93.8%) is more accurate than Swipe Right (91.7%), in agreement with the results on the validation set. Once again, the 2% difference is most likely due to the fact that all participants were right-handed.

When analysing the misclassifications, only a few gestures were actually detected as another gesture, most of the recorded False Positives in the test occurred because of complementary movements. Similarly to what is discussed in the Validation Set Results section, this happened most often in Swipe Up events which were followed by a Swipe Down gesture because the participants were naturally lowering their hand and arm down. Also, we observed that in some cases a Swipe Left or Right could be preceded by a Push, as some participants were moving their hand towards the radar to engage with it in a way similar to a Push gesture. Using additional post-processing that filters out unwanted triggers may result in a more precise detection and a lower learning effort for the user. An example is applying a long enough delay after each detected gesture to block the additional triggers.

Most importantly, during this test, we clearly observed that the more each participant performed the gestures, the fewer False Negatives and False Positives occurred. In particular, the Push gesture was the most difficult to trigger and to learn. This is one of the main reasons why the recall for Push is only 88.1%.

Besides performing the five gestures, we also tested the robustness of the model against non-gestures and other body movements. No False Positives were triggered for instance for the following movements:

- Hand waving
- Hand/finger rotation
- Hands moving while talking
- Clapping
- Moving around the radar
- Walking to/away from the radar

In some cases, however, misclassifications could occur when a hand or body movement was fast enough to resemble a gesture. Here are some examples:

- Sitting on a chair or bending to the ground in a relatively fast way in front of the radar sensor could trigger Push
- Jumping from the right to left (or vice versa) side of the radar sensor could be detected as Swipe Left (or Swipe Right)
- Quickly putting both hands on the table, where the radar sensor was placed, when standing up from a chair could trigger one of the gestures

Appendix I - Radar Configuration

The radar needs to be configured with the following parameters:

- Start Frequency: 58.5 GHz
- End Frequency: 62.5 GHz
- Samples per chirp: 64
- Chirps per frame: 32
- Number of Receivers: 3
- Number of Transmitters: 1
- Sample rate: 2 MHz
- Chirp repetition time: 0.299787 ms
- Frame repetition time: 30.0446 ms

Appendix II - Results of On-Device Model Testing

In the tables below, we report true positives (TP) and false positives (FP) for the on-device testing with 20 people as well as values for the Recall and Precision of the model in different scenarios. For the TPs we also write the total number of expected gestures per each gesture, which is 20 for most of the participants.

Person-Radar Distance: about 0.5 m (close case), about 1 m (far case).

Standing Close

Pers	**Swipe Right**		**Swipe Left**		**Swipe Up**		**Swipe Down**		**Push**	
	TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
1	20/20	0	20/20	0	19/20	4	20/20	0	20/20	0
2	10/20	0	17/20	2	19/20	1	16/20	0	14/20	5
3	20/20	0	20/20	0	18/20	3	20/20	0	16/20	1
4	17/20	3	19/20	0	19/20	5	20/20	3	18/20	6

Standing Far

Pers	**Swipe Right**		**Swipe Left**		**Swipe Up**		**Swipe Down**		**Push**	
	TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
1	20/20	3	20/20	3	19/20	0	20/20	0	18/20	0

2	19/20	0	15/20	0	18/20	4	16/20	0	20/20	0
3	20/20	3	20/20	0	20/20	1	20/20	0	17/20	0
4	14/20	0	14/20	0	14/20	0	11/20	2	15/20	3

Sitting Close

Pers	**Swipe Right**		**Swipe Left**		**Swipe Up**		**Swipe Down**		**Push**	
	TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
1	19/20	0	20/20	0	20/20	0	19/20	0	20/20	0
2	18/20	1	20/20	0	19/20	0	19/20	0	19/20	2
3	17/20	2	19/20	0	20/20	0	20/20	0	20/20	0
4	20/20	0	20/20	0	20/20	0	18/20	0	16/20	0
5	15/15	0	13/13	0	16/16	0	16/17	1	17/17	0
6	10/10	0	6/10	1	10/10	2	8/10	0	8/10	0

Sitting Far

Pers	**Swipe Right**		**Swipe Left**		**Swipe Up**		**Swipe Down**		**Push**	
	TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
1	19/20	0	19/20	1	19/20	0	19/20	5	18/20	4
2	19/20	3	20/20	1	18/20	1	20/20	2	16/20	1
3	18/20	0	19/20	1	19/20	0	16/20	0	15/20	0
4	19/20	0	19/20	0	18/20	0	19/20	0	16/20	1
5	20/20	2	20/20	0	20/20	0	18/20	1	19/20	0
6	10/10	1	10/10	0	10/10	0	8/10	0	10/10	1

Performance

Overall	Swipe Right	Swipe Left	Swipe Up	Swipe Down	Push	TOTAL
Recall	91.7 %	93.8 %	94.4 %	91.0 %	88.1 %	91.8 %
Precision	95.0 %	97.5 %	94.4 %	96.1 %	93.3 %	95.3 %

Standing	Swipe Right	Swipe Left	Swipe Up	Swipe Down	Push	TOTAL
Recall	87.5 %	90.6 %	91.3 %	89.4 %	86.3 %	89.0 %
Precision	94.0 %	96.7 %	89.0 %	96.6 %	90.2 %	93.2 %

Sitting	Swipe Right	Swipe Left	Swipe Up	Swipe Down	Push	TOTAL
Recall	94.9 %	96.2 %	96.8 %	92.2 %	89.4 %	93.9 %
Precision	95.8 %	98.1 %	98.6 %	95.7 %	95.6 %	96.8 %

Close	Swipe Right	Swipe Left	Swipe Up	Swipe Down	Push	TOTAL
Recall	89.7 %	95.1 %	96.8 %	94.1 %	89.8 %	93.1 %
Precision	96.5 %	98.3 %	92.3 %	97.8 %	92.3 %	95.4 %

Far	Swipe Right	Swipe Left	Swipe Up	Swipe Down	Push	TOTAL
Recall	93.7 %	92.6 %	92.1 %	87.9 %	86.3 %	90.5 %
Precision	93.7 %	96.7 %	96.7 %	94.4 %	94.3 %	95.1 %