Breast Cancer Detection and Segmentation in Mammograms using Deep Learning Techniques

Adam Jaamour

Masters of Science in Artificial Intelligence
The University of St Andrews
August 2020

Breast Cancer Detection and Segmentation in Mammograms using Deep Learning Techniques

Submitted by: Adam Jaamour

Abstract

TODO.

The code developed for this dissertation can be found online at the following URL: `https://github.com/Adamouization/Breast-Cancer-Detection-and-Segmentation`.

Contents

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Problem Description . . . 1
1.3 Related Systems & Their Applications . . . 1
1.4 Project Aims . . . 1
1.5 Report Structure . . . 1

2 Context Survey **3**
2.1 First part . . . 3
2.2 Second part . . . 3
2.3 Third part . . . 3
2.4 Chapter Summary . . . 3

3 Requirements **5**
3.1 Functional Requirements . . . 5
3.1.1 System Requirements . . . 5
Training Phase . . . 5
Evaluation Phase . . . 5
Data Pre-Processing Phase . . . 5
General Requirements . . . 6
3.1.2 Code Design Requirements . . . 6
3.1.3 Data Requirements . . . 6
3.2 Non-Functional Requirements . . . 6
3.3 Summary . . . 6

4 Ethics **7**

5 Design **9**
5.1 Programming Language . . . 9
5.2 Pipeline Design Analysis . . . 9
5.3 General Project Design . . . 9
5.4 Chosen Solution . . . 10
5.5 Summary . . . 10

6 Implementation 11

7 Evaluation & Critical Appraisal 13

8 Conclusions 15
8.1 Achievements . . . 15
8.2 Future Work . . . 15
8.3 Limitations . . . 15
8.4 Project Summary & Reflections . . . 15

A Potential Programming Languages Comparison 16

B Text and Binary Files Comparison 18

C Raw Console Output Example 19

D Stored Feature Files 24
D.1 Greyscale Histogram Features . . . 24
D.2 RGB Histogram Features (Red channel only) . . . 25
D.3 HSV Histogram Features . . . 25

E 13 Point Ethics Check List 27

F Experiment Survey 29
F.1 Title . . . 29
F.2 What is the project about? . . . 29
F.3 Your role in this experiment . . . 29
F.4 Watch . . . 30
F.5 Rank . . . 31
F.6 Confirmation Message . . . 32

G Raw Experiment Results 33
G.1 Question 1 . . . 33
G.2 Question 2 . . . 33
G.3 Question 3 . . . 34

H Code Listings 36
H.1 Main Module . . . 36
H.1.1 Argument Parsing . . . 36
H.1.2 Initial Off-line Feature Extraction Phase Function . . . 37
H.1.3 Initial On-line Retrieval Phase Function . . . 38
H.1.4 Initial Database Pre-Processing Phase . . . 39
H.2 Histograms Module . . . 40
H.2.1 High Level Module Overview . . . 40
H.2.2 Single Greyscale Histogram Generation . . . 40

H.2.3 Single RGB Histogram Generation 41
H.2.4 Single HSV Histogram Generation 42
H.2.5 Average & Store Greyscale Histogram 43
H.2.6 Average & Store RGB Histogram 43
H.2.7 Average & Store HSV Histogram 44
H.2.8 Histogram Matching 45
2D Histogram matching 45
Matching 3D Histograms 47
H.2.9 Shot Boundary Detection Algorithm 49
H.2.10 Frames Pre-Selection Function 50
H.3 Video Operations Module 51
H.3.1 VideoStabiliser Class 51
H.3.2 ClickAndDrop Class for Manual ROI Selection 51
H.4 Helpers Module 53

List of Figures

F.1 The YouTube video showing the 6 database videos used for the experiment, available online: `https://www.youtube.com/watch?v=BvukbK-sX9A`. 30
F.2 The YouTube video showing the query video used for the experiment, available online: `https://www.youtube.com/watch?v=4JPoO-aSzNE`. 31
F.3 Screenshot of the checkbox grid used to rank the database videos from most likely to match the query video to least likely. 31

G.1 Survey results of the "Which database video does the recorded query match the most?" ranking question. 33

List of Tables

A.1 Table comparing the main pros and cons for using different programming languages to build the system. 16

B.1 Table comparing the pros and cons for storing data in text files and binary files. 18

Acknowledgements

todo.

Chapter 1

Introduction

1.1 Motivation

TODO.

1.2 Problem Description

TODO.

1.3 Related Systems & Their Applications

TODO.

1.4 Project Aims

TODO.

- aim 1
- aim 2
- ...

1.5 Report Structure

- **Introduction**
  Todo: describe.

- **Context Survey**
  Todo: describe.
- **Requirements Specification**
  Todo: describe.
- **Ethics**
  Todo: describe.
- **Design**
  Todo: describe.
- **Implementation**
  Todo: describe.
- **Evaluation & Critical Appraisal**
  Todo: describe.
- **Conclusions**
  Todo: describe.

Chapter 2

Context Survey

Lit review introduction.

2.1 First part

Todo.

2.2 Second part

Todo.

2.3 Third part

Todo.

2.4 Chapter Summary

Todo: review.

Chapter 3

Requirements

Requirements introduction.

3.1 Functional Requirements

3.1.1 System Requirements

TODO

Training Phase

Describe phase

F1 Req 1

F2 Req 1

Evaluation Phase

Describe phase

F1 Req 1

F2 Req 1

Data Pre-Processing Phase

Describe phase

F1 Req 1

F2 Req 1

General Requirements

F1 Req 1

F2 Req 1

3.1.2 Code Design Requirements

F1 Req 1

F2 Req 1

3.1.3 Data Requirements

F1 Req 1

F2 Req 1

3.2 Non-Functional Requirements

NF1 req 1.

NF2 req 2.

3.3 Summary

TODO: review.

Chapter 4

Ethics

Todo

Chapter 5

Design

With the requirements necessary to build the system now formulated, potential solutions to fulfil the list of requirements from Chapter 3 for each aspect of the system can now be analysed before choosing a final solution.

Throughout this chapter, the requirements established from the previous chapter referred as "F" stand for functional requirements, while requirements referred as "NF" stand for non-functional requirements.

TODO: recap section

5.1 Programming Language

TODO

5.2 Pipeline Design Analysis

TODO

5.3 General Project Design

TODO.

5.4 Chosen Solution

1. **Programming Language**:
   (a) Python 3.7[1]
   (b) Third-party libraries:
      i. Keras
      ii. TensorFlow
      iii. NumPy
      iv. Pandas
      v. SciPy
      vi. Matplotlib

5.5 Summary

Review

[1]Version 3.7 is the latest stable release of Python, satisfying NF1.

Chapter 6

Implementation

TODO.

Chapter 7

Evaluation & Critical Appraisal

TODO.

Chapter 8

Conclusions

TODO.

8.1 Achievements

TODO.

8.2 Future Work

TODO.

8.3 Limitations

TODO.

8.4 Project Summary & Reflections

TODO.

The code developed for this dissertation can be found online at the following URL: `https://github.com/Adamouization/Breast-Cancer-Detection-and-Segmentation`.

Bibliography

Appendix A

Ethics Approval Document

Examples of the different compact features stored in plain text files using the butterfly video as an example. These correspond to averaged histograms in greyscale, RGB and HSV colour models.

A.1 Greyscale Histogram Features

The averaged greyscale histogram only has one channel, represented by 255 values for the 255 bins. Only bins 1-5, 124-128 and 250-255 are represented below.

File: *butterfly.mp4/hist-gray*

```
# Greyscale Histogram (255 bins) [normalised]
0.000677
0.000453
0.000588
0.000803
0.001338
...
0.147575
0.149898
0.141548
0.023301
0.108333
...
0.000152
0.000164
0.000087
0.000080
0.000128
```

A.2 RGB Histogram Features (Red channel only)

The averaged RGB histogram has three channels, each represented by 255 values for the 255 bins. Only the red channel is represented in the example below. The same file format is used to represent the green and blue channels. Only bins 1-5, 124-128 and 250-255 are represented below.

File: *butterfly.mp4/hist-r*

```
# 'R' channel of RGB histogram (255 bins) [normalised]
0.001610
0.000436
0.000395
0.000468
0.000833
...
0.150552
0.112711
0.133993
0.113074
0.103523
...
0.002662
0.002867
0.003141
0.003351
0.003299
```

A.3 HSV Histogram Features

The averaged HSV histogram has 8 total slices for the 8 different hue bins. Each hue bin has 12 saturation bins, each with 3 value bins. Only the 1st and 8th hue bins are represented below.

File: *butterfly.mp4/hist-hsv*

```
# HSV Histogram shape: (8, 12, 3) [normalised]
# New slice
1.690998567606915226e-03 9.581707232758741389e-04 1.511106754399158718e-03
2.772998953746123787e-03 4.584945844147692712e-03 3.237609486942264088e-03
4.736755551262335738e-03 6.970707496458833710e-03 4.525759765370326003e-03
5.926002092151479682e-03 7.984556638720359881e-03 9.167456248013133510e-03
5.528808910061012299e-03 1.160728825594891170e-02 8.409987489523535267e-03
8.208146847953850730e-03 1.193360518664121628e-02 6.526554860597984752e-03
```

```
7.324515270407904594e-03 9.202591588043353144e-03 1.104022278873757930e-02
7.212228395722128892e-03 7.106588645414872209e-03 1.543889956718141367e-02
6.635527232323180230e-03 7.380225738002495726e-03 1.792533712630922302e-02
5.477001382545990983e-03 5.517539758743209878e-03 1.022866905920884889e-02
3.716428414918482304e-03 1.813905805730345495e-03 6.794666538057340423e-04
8.167625489560041349e-03 1.257265304130586632e-03 1.215253951985770014e-04
...
# New slice
8.166000294626097689e-04 1.444111869204789400e-03 2.556757863864979764e-03
2.281072126193480019e-03 1.405416374010118492e-03 6.197975771184163719e-04
3.227299312129616737e-03 4.016239108750596642e-04 5.364667876247867172e-05
4.634697921574115753e-03 1.376390735624061679e-04 2.150440096474168000e-04
4.121060758321123084e-03 1.419378115149976904e-04 1.943936631505494006e-04
3.336825863119553554e-03 2.044695773458277588e-04 1.442692149050179185e-04
3.074369096959178940e-03 3.068208203661594103e-04 5.189062966647642284e-04
2.311520511284470558e-03 2.902706883932378913e-04 2.334064346293664594e-03
1.362986608662388561e-03 1.100270561768080737e-04 2.477648943154649262e-03
7.060457216787406187e-04 7.275994308160575492e-05 1.679020986722951645e-03
2.139106257924471488e-04 7.810963734300457872e-05 1.976605589416894021e-04
2.936312653632326707e-04 3.317900460718771412e-05 1.067189321640233325e-05
```