

# DDSM Software — A User's Guide

## Introduction

This manual describes how to install and use software to:

- download DDSM mammograms and convert them into a usable format;
- obtain the annotations (e.g., tumour segmentations) made by the DDSM radiologists and the associated metadata (i.e., malignancy type etc.)

## Copyright, Licensing, Warranty and Fitness for Use

The accompanying ddsd-smftware.zip file contains software that is in the public domain, is publicly available or has been specially adapted or developed to be used within the Windows and Cygwin environment. You are granted a non-exclusive license to use or modify the software for research purposes only.

The accompanying software must be used for research purposes only and has not been validated for use in clinical or similar environments. There is no warranty for this software, which is supplied “as is”, and no claims are made about fitness for purpose. Portions of the software have, however, been used by many computer-aided mammography researchers for a number of years.

## Software Requirements

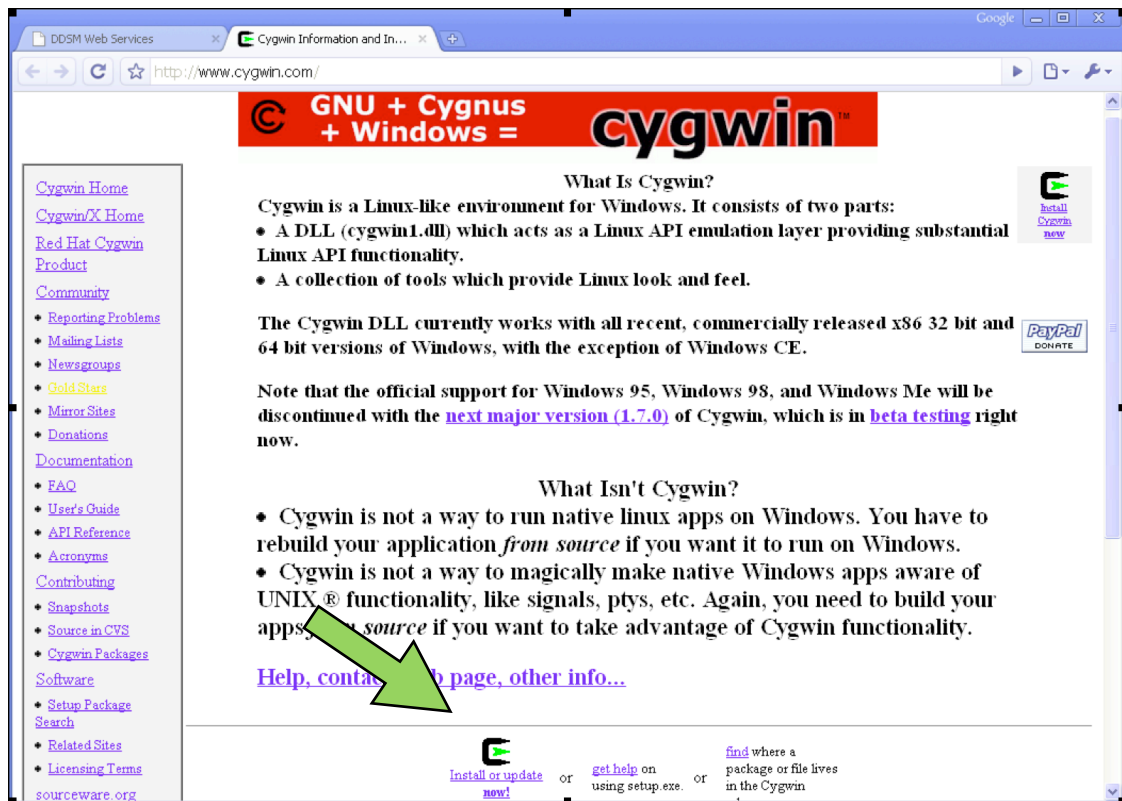
In addition to the provided software, you will need:

- One of:
  - Windows XP Professional with Service Pack 2 (later versions will probably work, but are untested).
  - Linux (x86, 64-bit only). Please note that I have not been able to test the Linux **jpeg** binary.
- If you are using Windows, you will also need:
  - The Cygwin UNIX environment version 1.5.25-15 or later (available as a free download; installation instructions follow later in this manual).
- ImageMagick version 6.4.0.6-1 or later. (Installation instructions are provided later in this document for Windows users.)
- Ruby version 1.8.7-p72-2 or later. (Installation instructions are provided later in this document for Windows users.)
- Matlab version 7.6.0 (R2008a) or later (commercial software, available separately).
- An FTP client (one is built in to Windows Explorer, though other options are available).

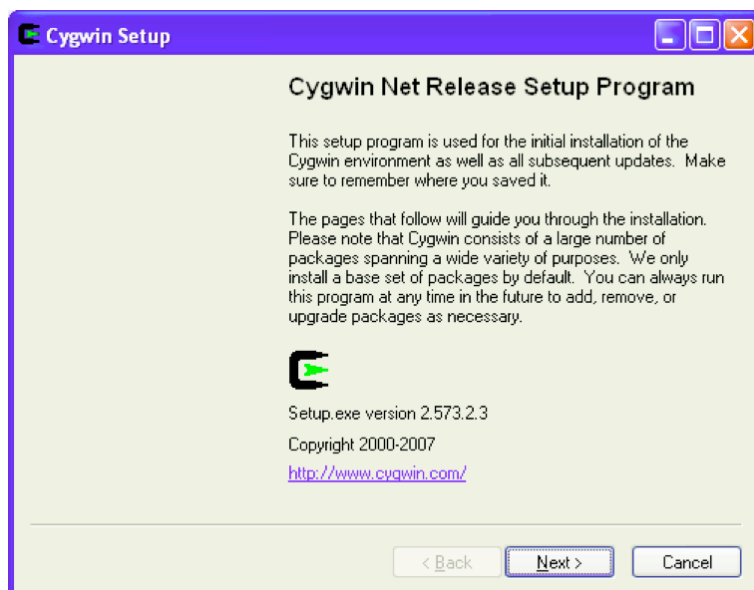
## Windows Installation Instructions

Before you can install the software contained in the **ddsm-software** directory, you must install the Cygwin UNIX-like environment and some supporting software. The following instructions may be a little out of date, but should give you a good idea as to what you need to do:

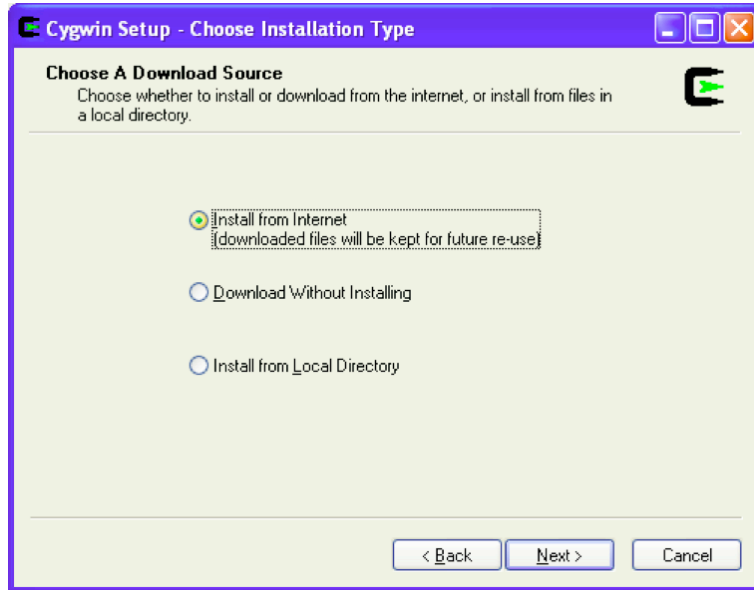
1. Download the Cygwin installer from **<http://www.cygwin.org>** (see green arrow, below).



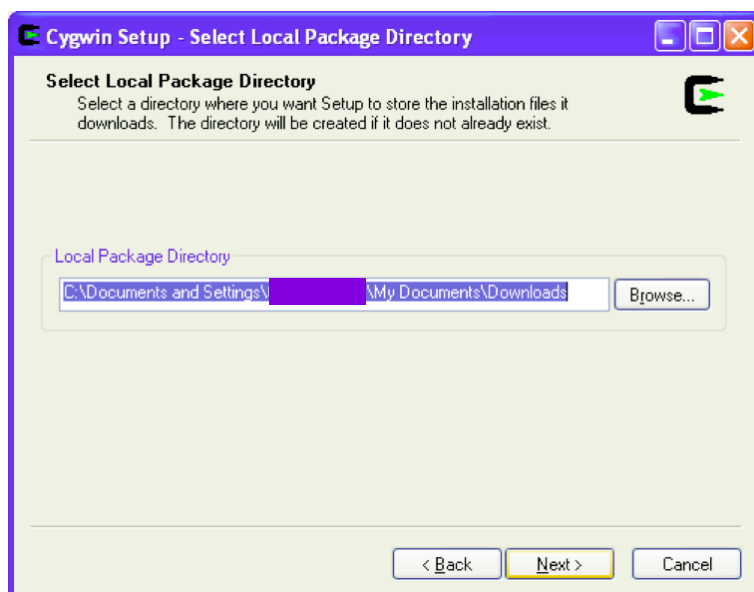
2. Run the Cygwin **setup.exe** file.



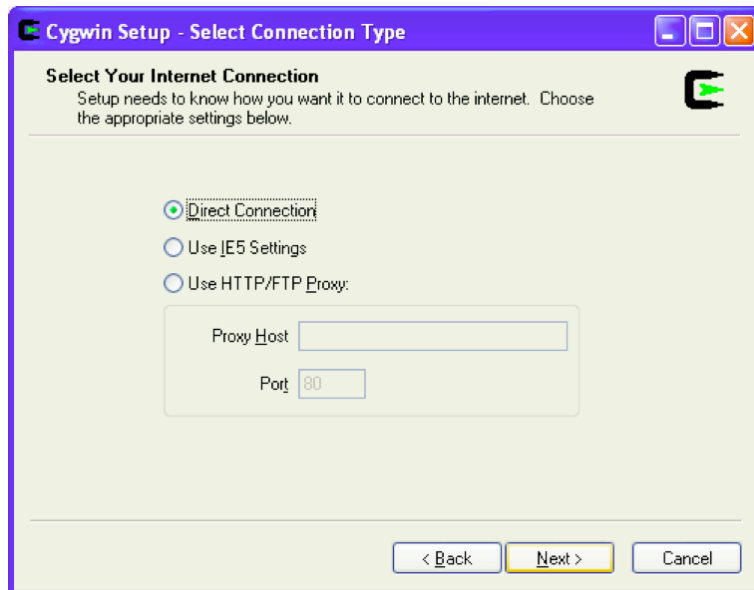
3. Choose Next > Install from Internet > Next.



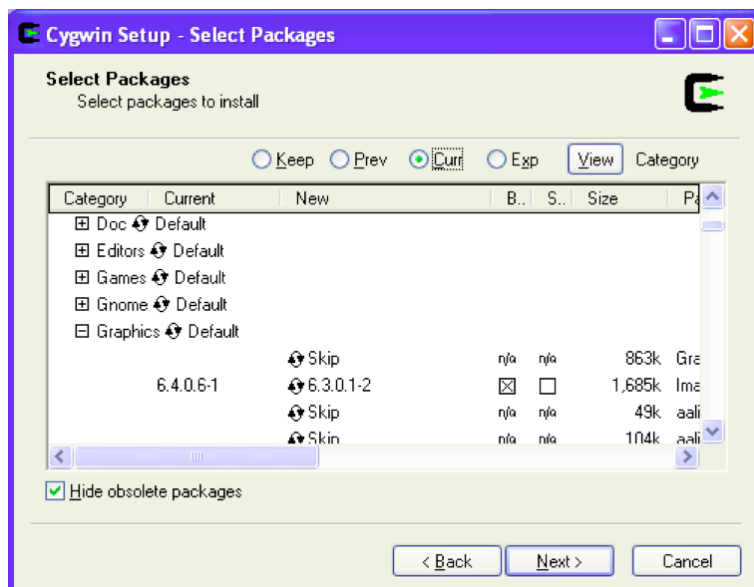
4. Leave Root Directory set to its default: **C:\cygwin**; choose to install for “Just Me”; leave Default Text File Type set to “Unix/binary”. Then click Next.
5. Leave Local Package Directory set to its default and click Next.



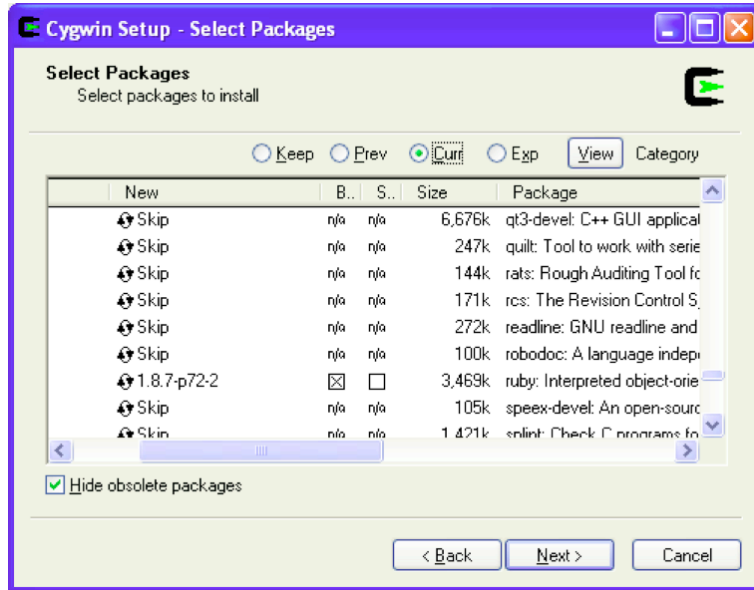
6. Choose “Direct Connection” and click Next.



7. Choose a server that is likely to be close to you and click Next.
8. Under the Graphics category, choose to install ImageMagick. You do this by expanding the Graphics category, resizing the window and/or the column headers so that you can see the package names (which appear to the right-hand side), and then clicking once on the element in the “New” column that says “Skip”, so that it changes to a version number; if you click too many times, you can keep clicking until you get back to “Skip” and then the version number. Next to the “New” column are two columns—which you may need to resize to read their column headers—labelled “Bin?” (for “binary”—i.e., the actual software) and “Src?” (for “source code”); the box for “Bin?” should be checked (as shown), but you don’t need to check the “Src?” box.



9. Under the Devel category, choose to install Ruby, in a similar way to the previous step.



10. Click Next and wait for the installation to complete.
11. When installation has finished, choose to create a Desktop icon and to add an icon to the Start menu; then click Finish.
12. Start a Cygwin session by double-clicking the Cygwin icon on your desktop; this provides a Unix-like command-line interface. Initially, your session will start in you “home” directory, denoted by “~” in Cygwin; this corresponds to the Windows directory **C:\cygwin\home\<Your User Name>** where **<Your User Name>** is your Windows username.
13. Follow the remaining instructions below for both Windows and Linux users.

## Installation Instructions — Windows and Linux Users

14. Copy the **ddsm-software** directory into a convenient directory; Windows users should copy the **ddsm-software** directory into your Cygwin home directory (i.e., **C:\cygwin\home\<Your User Name>**).
15. Change directory to the newly created **ddsm-software** directory by typing the following command into the terminal (Cygwin window for Windows users) and then pressing return:

```
cd ddsm-software
```

16. Verify that the files have been copied correctly: type the following command and then press return:

```
ls
```

You should see a list of the following file names: **ddsmraw2pnm.exe**,  
**get\_ddsm\_groundtruth.m**, **get-ddsm-mammo**, **info-file.txt**, **jpeg.exe**, **jpeg**.

17. Verify that you can run the **jpeg** program (**jpeg.exe** for Windows users): type the applicable command and press return:

**./jpeg.exe** (Don't forget the leading **./** characters; don't confuse **\** for **/**)

**./jpeg** (Don't forget the leading **./** characters; don't confuse **\** for **/**)

You should see some usage instructions printed to the screen; you will never need to run this command directly, however, so you can ignore them.

18. Verify that the Ruby program was installed correctly; type the following and press return:

**ruby -e "p 'Hello world!'"** (Don't miss out the single quotes!)

You should see "**Hello world!**" printed to the screen.

## How to Obtain and Convert a DDSM Mammogram

The software provided makes it very simple to get a DDSM mammogram in a usable file format. The following example shows how we can get the mammogram **A\_1141\_1.LEFT\_MLO** in PNG format. PNG format files can be read by a very large range of software, including Matlab, Photoshop and Windows itself. It is a lossless file format and offers good compression ratios. (Note, however, that the DDSM mammograms were digitised at high resolution and even in PNG format, each mammogram is about 40MB.)

To download and convert the mammogram **A\_1141\_1.LEFT\_MLO** into PNG format:

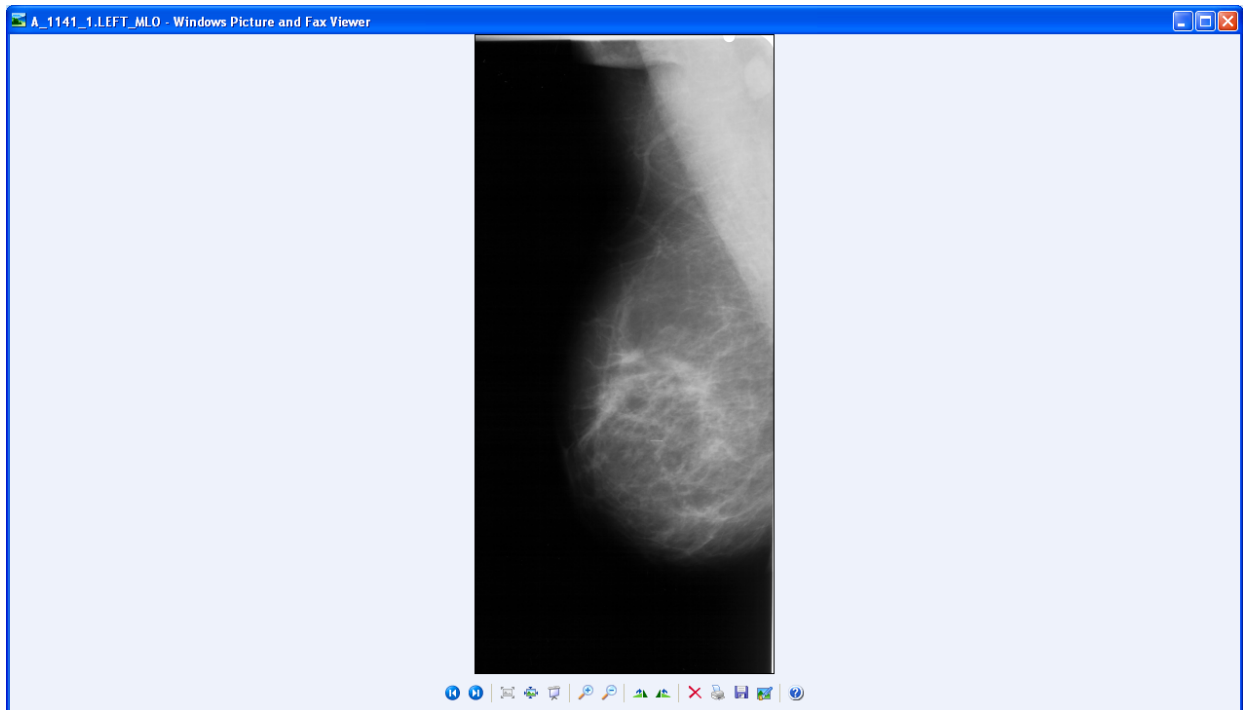
1. If you closed the terminal or Cygwin program, restart it and repeat step 18 above to change to the **ddsm-software** directory.
2. Type the following command and press return:

**./get-ddsm-mammo A\_1141\_1.LEFT\_MLO**

The mammogram will be downloaded and the converted file will be placed in the **ddsm-software** directory within your home directory. The full (Unix) path to the file will be printed to the screen. It is worth moving mammogram files to another directory to avoid clutter.

(Continued overleaf.)

3. View the downloaded and converted file. E.g., in Windows, double-click on the file:



The command in step 2 starts the **get-ddsm-mammo** program, telling it to obtain and convert the mammogram named A\_1141\_1.LEFT\_MLO. The program connects to the DDSM's FTP server, downloads the corresponding “lossless” JPEG file, converts that file to a raw binary format, converts that file to a simple human-readable file format called PNM, converts that file to the desired PNG format, and finally deletes the “lossless” JPEG and all intermediate files. Because the DDSM files are large and conversion is processor intensive, it can take a few minutes for the file to be downloaded and converted. Exactly how long depends on the speed of your Internet connection and how powerful your computer's CPU is; on a domestic 10Mbps cable modem and a 2006-vintage 2GHz Intel Core 2 Duo CPU, **get-ddsm-mammo** takes about 3 minutes to run.

If the **get-ddsm-mammo** program is interrupted, intermediate files may be left in the **ddsm-software** directory; such files will have the suffix “.1” or “.pnm”, and should be deleted as they tend to be quite large (80MB+).

Note that the DDSM's FTP server has a policy of allowing no more than 10 users at a time; if the **get-ddsm-mammo** program fails, the limit on the number of users is a likely reason—simply wait a few minutes and try again. (While testing the software described in this manual, the DDSM's FTP server went offline for several hours, so be aware that this may be a “weak link” in your workflow.)



## How to Obtain DDSM Radiologist Annotations and Metadata

The DDSM radiologist annotations and metadata (which includes all information provided about the annotations in the **.OVERLAY** files, such as the type of abnormality and its subtlety) are made available via the **get\_ddsm\_groundtruth.m** Matlab function. You can install this software as follows:

1. Copy or move the file **get\_ddsm\_groundtruth.m** from the directory **ddsm-software** in your home directory to a directory in which you keep your Matlab software (creating such a directory if one does not exist).
2. Start Matlab.
3. Add the directory into which you placed the **get\_ddsm\_groundtruth.m** file in step 1 to your Matlab path (see <http://bit.ly/set-matlab-path-with-gui>).
4. From the Matlab command prompt, check that Matlab can find the **get\_ddsm\_groundtruth.m** file in its path by entering the following command and pressing return:

```
help get_ddsm_groundtruth
```

Matlab should print the documentation for the **get\_ddsm\_groundtruth** function.

For this example, we will obtain the annotations and metadata for the file **A\_1580\_1.LEFT\_MLO**; this mammogram was chosen because it has multiple abnormalities (boundaries) and one of those has a core annotation, and is therefore an example of a non-trivial mammogram. You can obtain groundtruth (annotations and metadata) for this file as follows.

1. Using your FTP client (such as Windows Explorer), connect to the DDSM FTP server (which is at **figment.csee.usf.edu**; if prompted, use the username "**anonymous**" and an empty password).
2. Navigate to the directory **/pub/DDSM/cases/cancers/cancer\_10/case1580/**
3. Copy the remote file **A\_1580\_1.LEFT\_MLO.OVERLAY** to a convenient directory on your local computer; let us call this directory **<overlay-directory>**.
4. Disconnect from the DDSM's FTP server so that you aren't preventing other users from accessing it.
5. At Matlab's command prompt, enter the following command and press return:

```
cd <overlay-directory>
```

6. At Matlab's command prompt, enter the following command and press return:

```
groundtruth = get_ddsm_groundtruth('A_1580_1.LEFT_MLO.OVERLAY');
```

The rest of this tutorial is provided by the **get\_ddsm\_groundtruth** function's documentation (read from "Example usage"); enter the following command and press return to see the documentation:

```
help get_ddsm_groundtruth
```

Note that to get a binary mask image showing the region annotated by the radiologist, you will need to know the image dimensions. You can obtain this information in one of two ways; you can:

- Use the DDSM's website and view the thumbnails pages; these list the contents of the corresponding case's ICS file, which gives the number of lines (i.e., rows) and pixels per line (i.e., columns) for each mammogram in the case.
- Obtain the corresponding mammogram (using the **get-ddsm-mammo** program) and use Matlab's **size** function to get the number of rows and columns in the mammogram.

Parsing the **.OVERLAY** file is very quick, while creating a binary mask matrix using the anonymous function(s) returned by **get\_ddsm\_groundtruth.m** can take several seconds. The advantage of the approach used by **get\_ddsm\_groundtruth.m** is that the cell array of structs that is returned can easily be saved to disk using Matlab's **save** command, and such files will require very little disk space relative to the full binary mask matrices (which are the same size as their corresponding mammograms).

## What Can Go Wrong

The weakest link in the workflow described in this manual is the DDSM's FTP server, which—to the best of this author's knowledge—permits only 10 users to access the server at any given time; the server also suffered an outage of several hours while testing was being performed on the software described in this manual.

The **get-ddsm-mammo** program should be quite robust, with the exception of its reliance on the DDSM's FTP server. If you experience problems using it, please try connecting to the DDSM's FTP server using your FTP client, so see if this is the cause of the problem.

The **get\_ddsm\_groundtruth.m** function is new and has only been tested with about 10 overlay files. It would be wise to double check the number of boundary and core annotations with the DDSM website's thumbnail preview pages. Please report any bugs that may be found by including a copy of the overlay file that causes the bug.