Molegazer

Astronomy meets dermatology

Chris Frohmaier - PyData Southampton - 17 Oct 2023

Dr Chris Frohmaier

- University of Southampton

- Senior Research Fellow in Physics and Astronomy

- Research Software Engineer: Legacy Survey of Space and Time

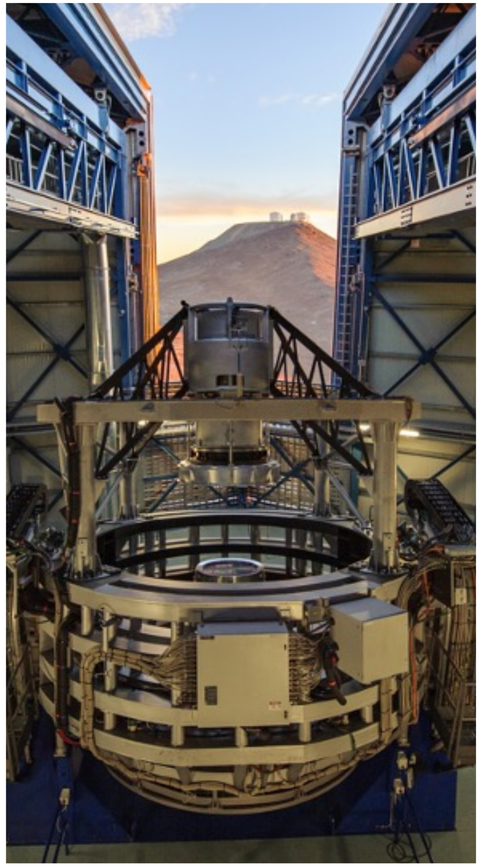
✉️ cfro.astro@gmail.com

chrisfrohmaier

Context

Astronomy and Dermatology?

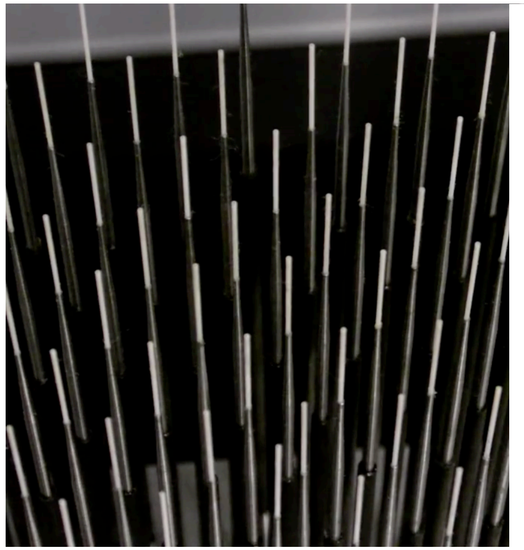
Get more data!



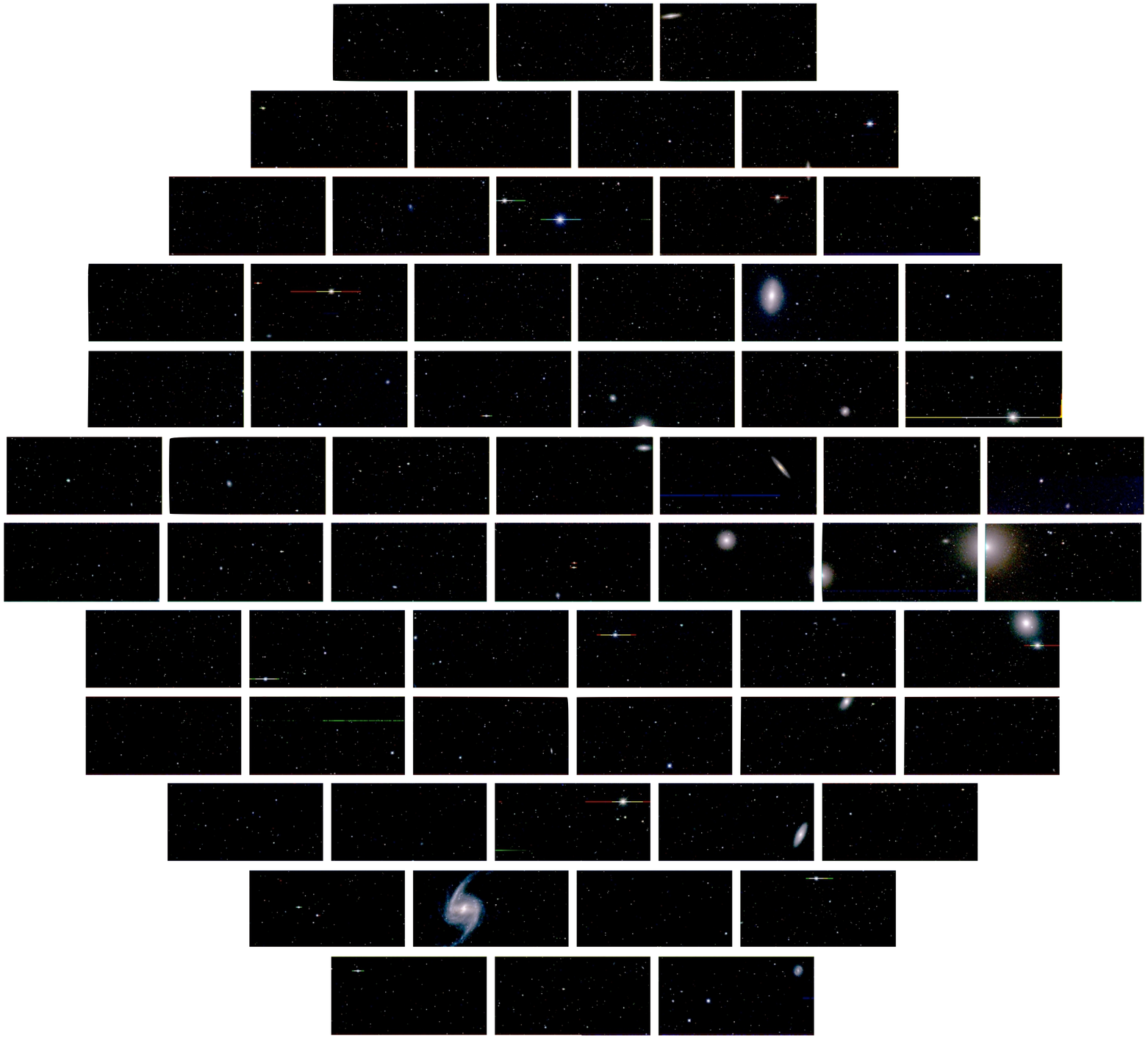
Legacy Survey Space Time

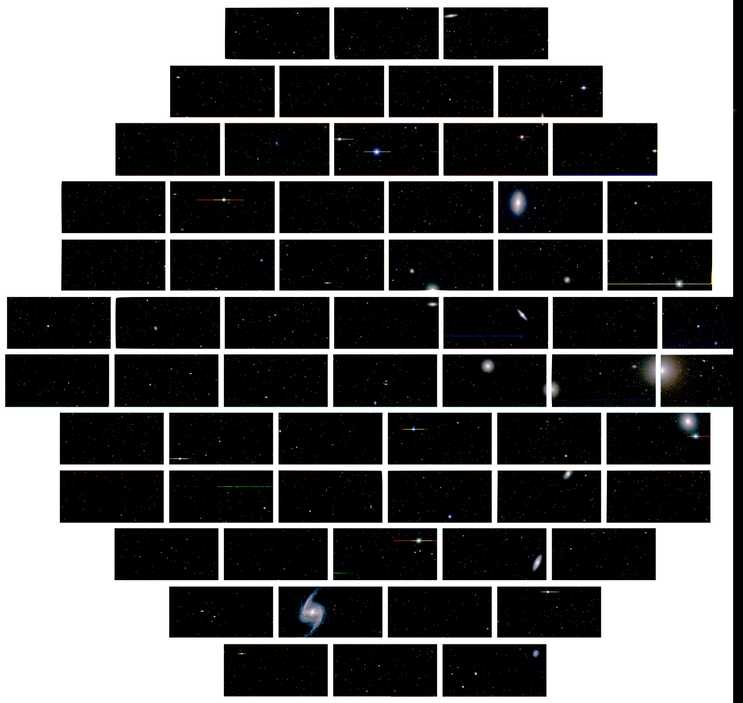
Discover exploding stars (supernova)

20 TB/night - I orchestrate with prefect.io



15 PB after 10 year experiment



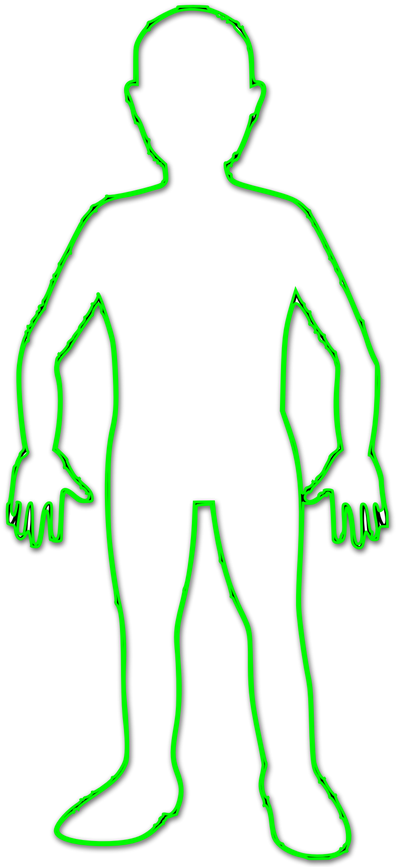


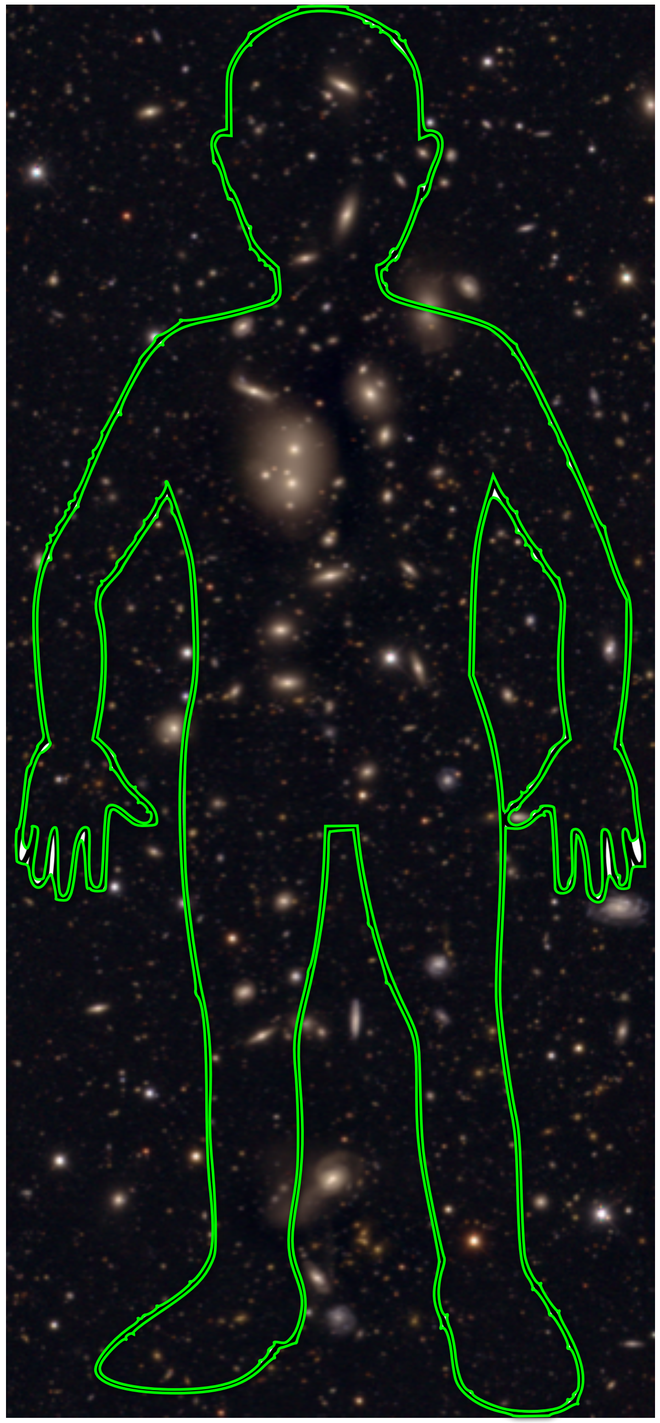
Dermatology

CANCER RESEARCH UK

Science and Technology Facilities Council

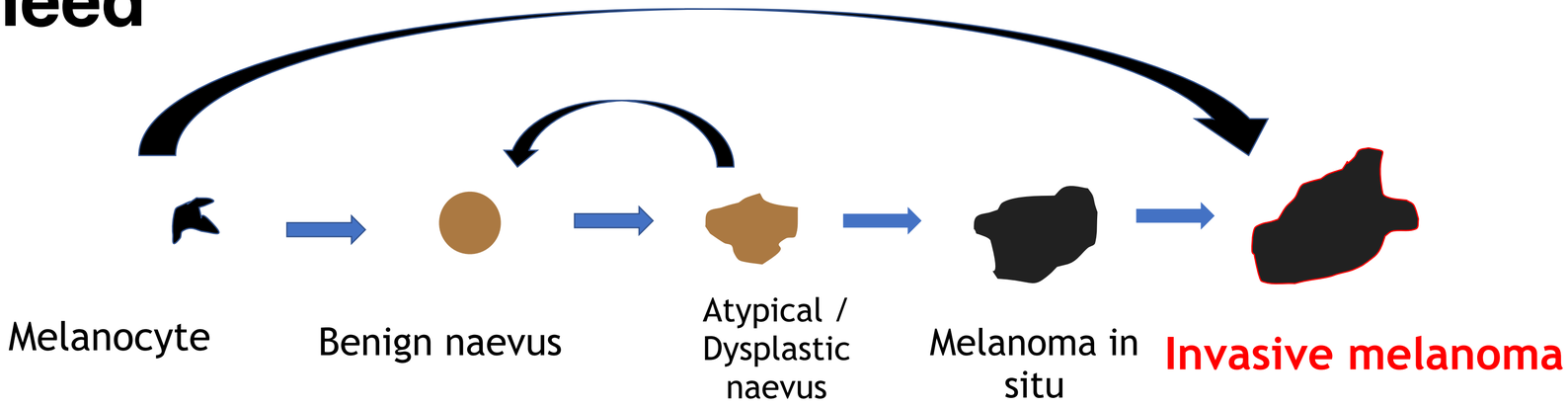
Dr Rubeta Matin PhD FRCP
Consultant dermatologist
Senior Lecturer
Oxford University Hospital Trust





Context 2

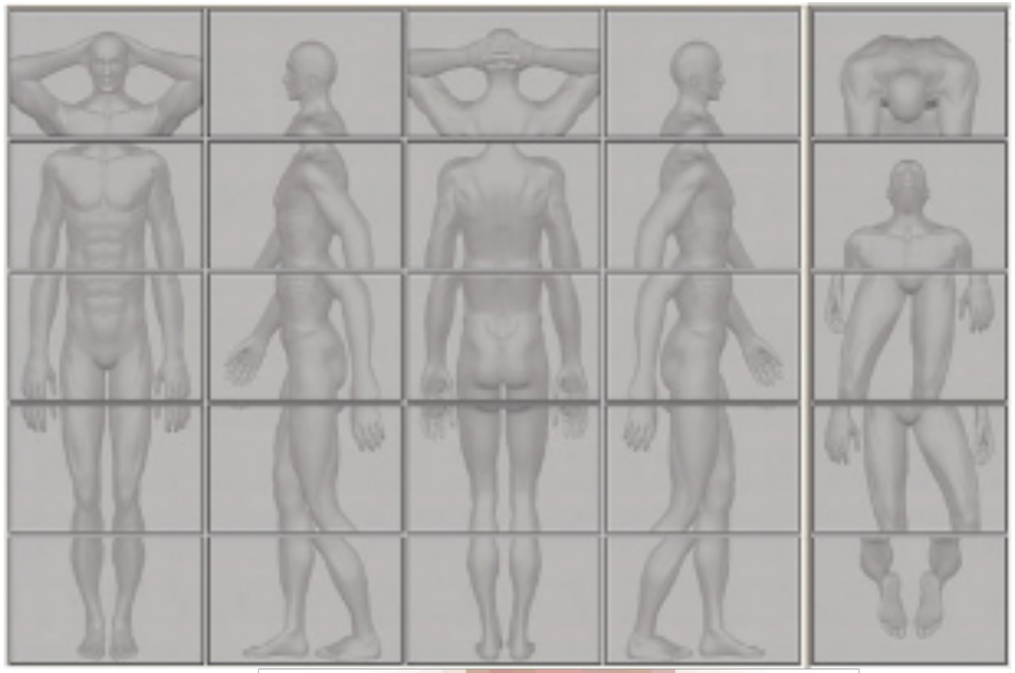
Medical need

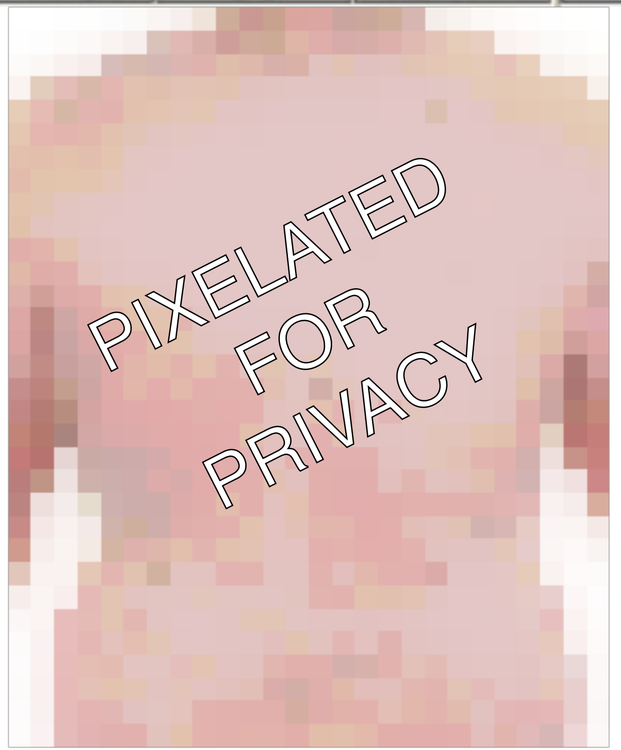


- 60% of melanoma comes from pre-exisiting naevus (mole)
- >6/day die of melanoma in UK
- Sequential monitoring of patients is labour intensive and not routinely performed
  - Moles are checked by eye and compared to a PDF printout by a consultant!

The Data

- 50 NHS patients (Oxford cohort)
  - High risk of developing melanoma
- Clinical setting
- Total Body Photography ~3 months collected over 3 years
- 500 patients with single epoch snapshot photography
- >8,000 individual images taken to-date
- All identifying images are removed before sending to Southampton
- Nikkon NEF format

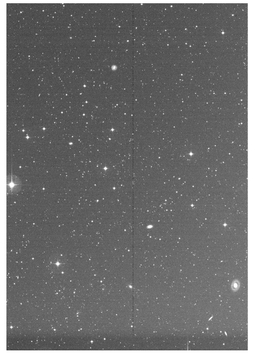




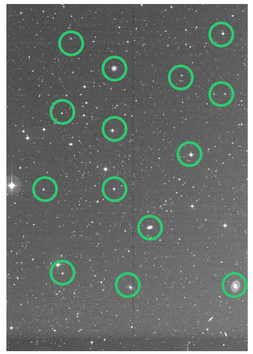
Project Aim:
Predict which naevi will evolve into melanoma

(work-in-progress)

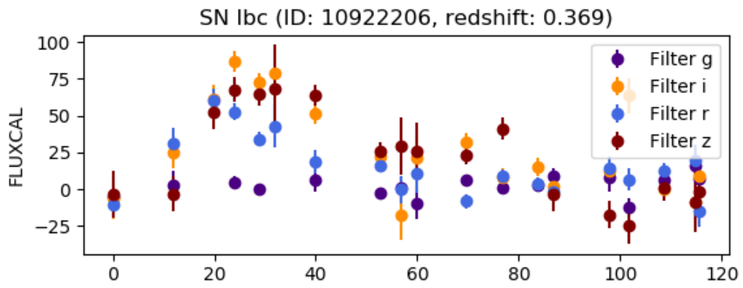
Aquire Image



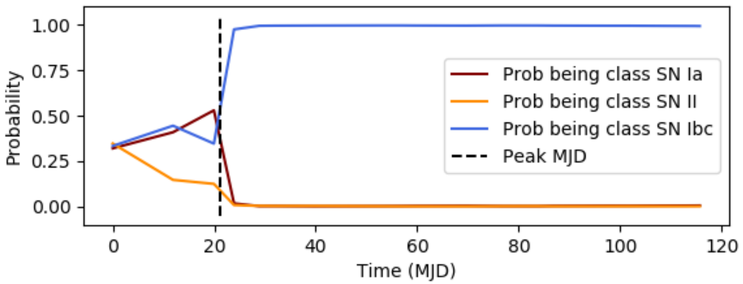
Source Detection

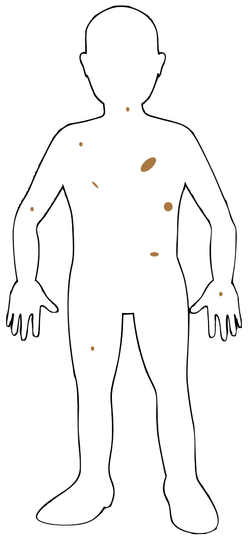


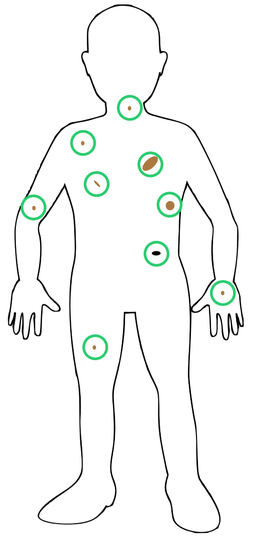
Compare History

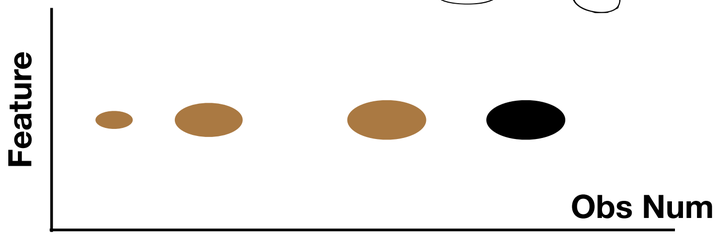


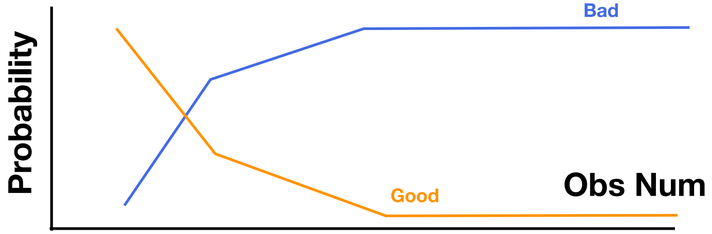
Predict classification/ evolution





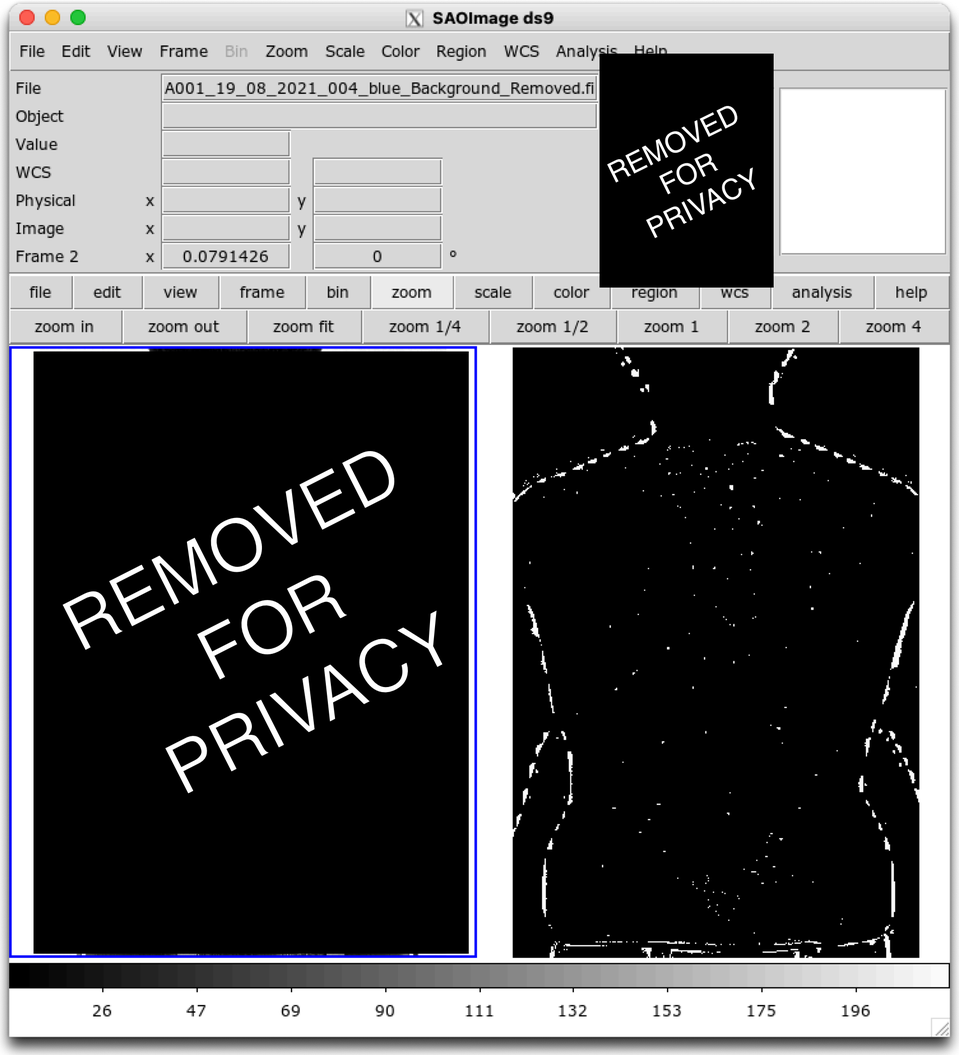


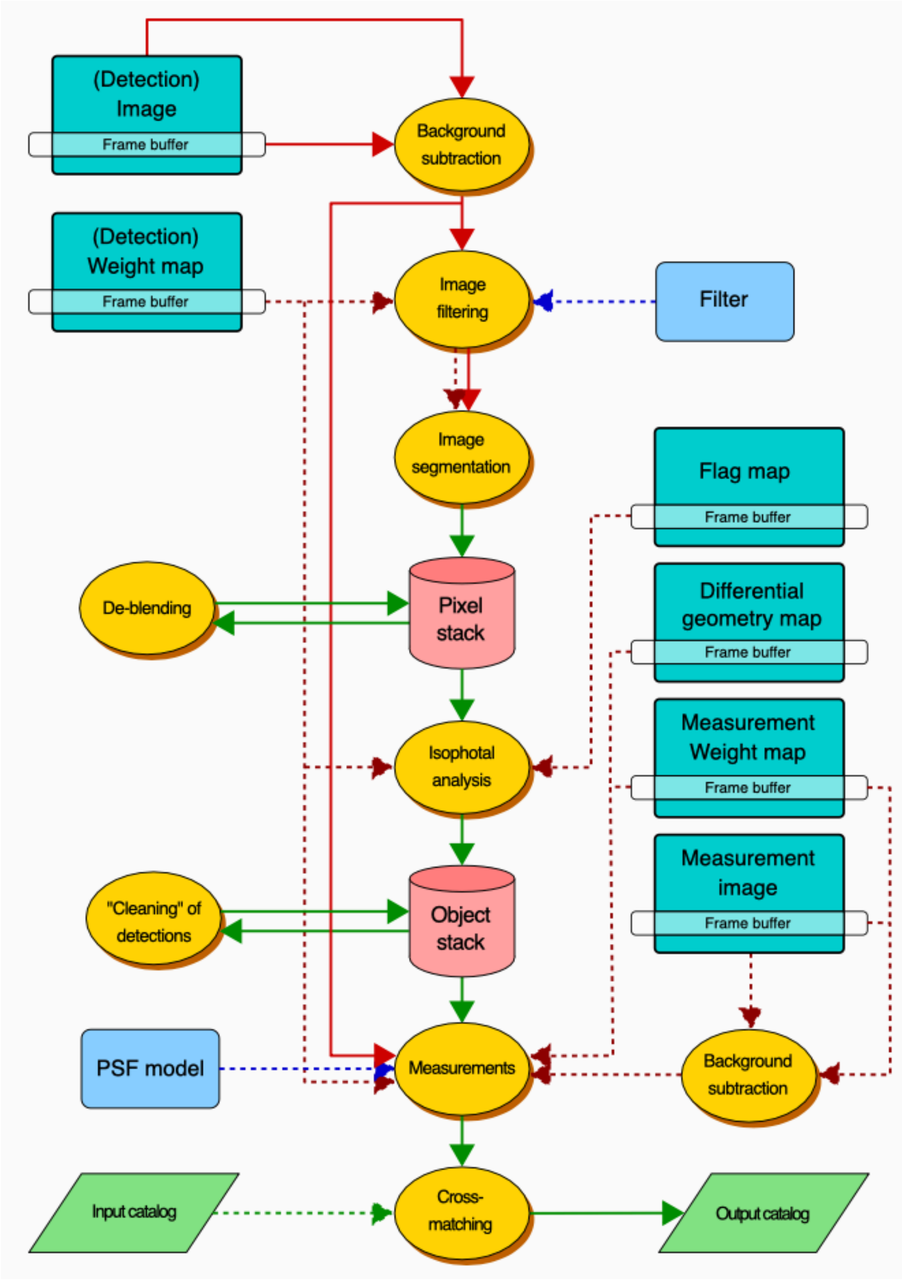




The astronomy solution

Source Extractor





We could detect moles… but it wasn’t a scalable solution

Research Software Group

University of Southampton

Dr Sam Mangam

Senior Research Software Engineer

University of Southampton

Dr Ed Parkinson

Research Software Engineer

University of Southampton

Dr Sam Mangam

Dr Ed Parkinson

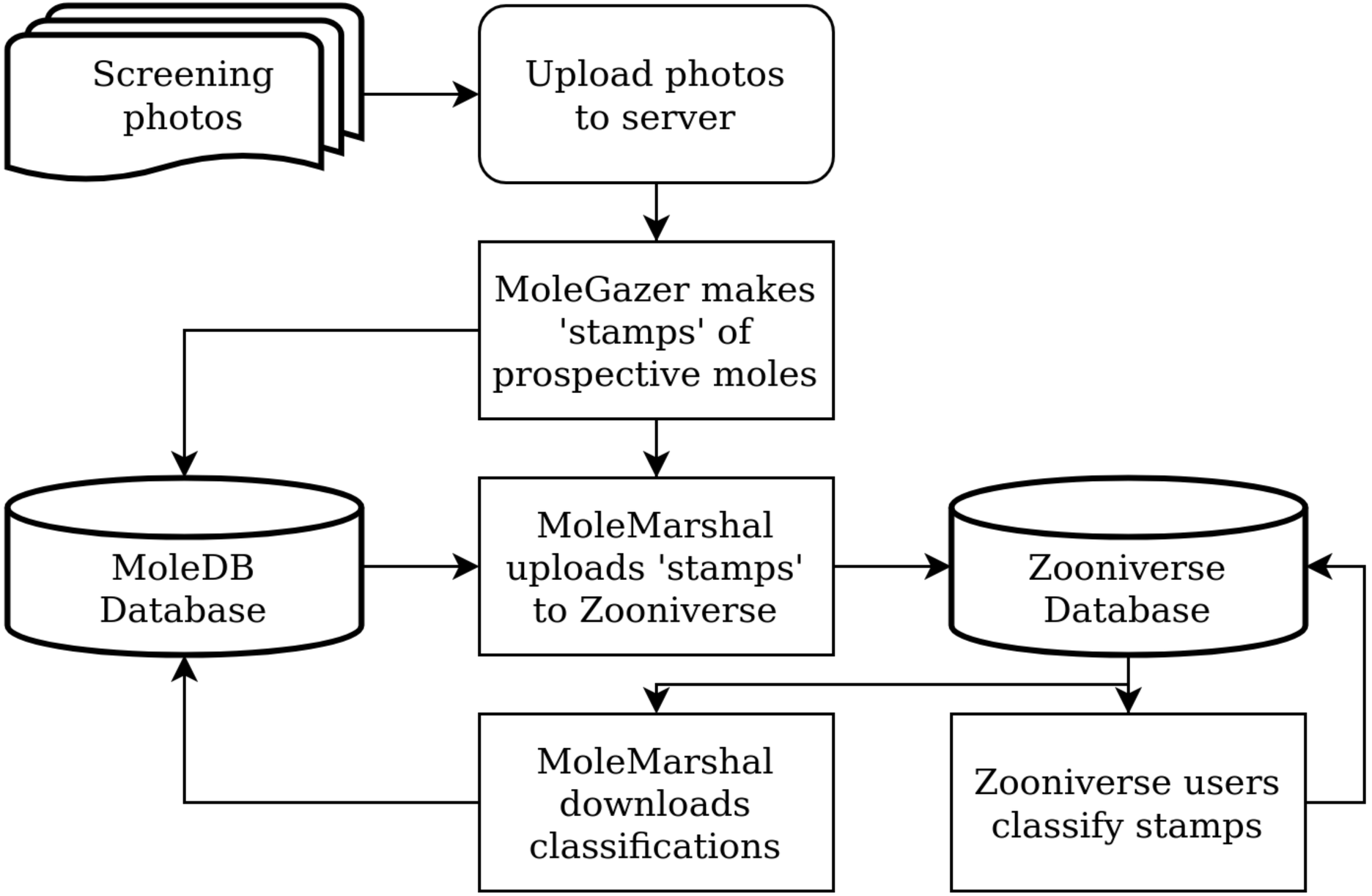
Improve our mole detection algorithms within the Python framework

Create a system for dermatologists to label our data to train future models

MoleGazer

Mole Marshal

The Moleverse

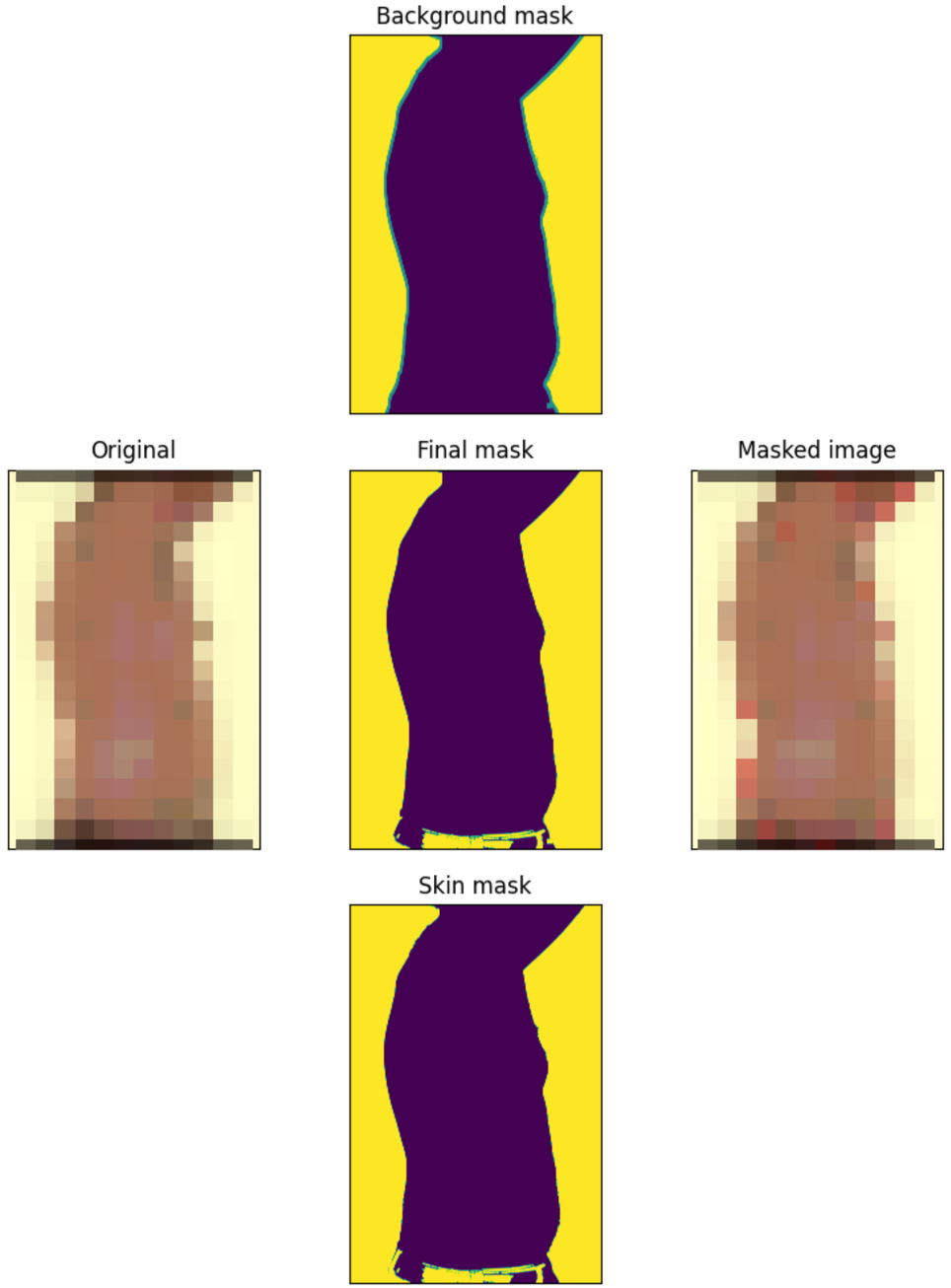

Screening photos
Upload photos to server
MoleGazer makes 'stamps' of prospective moles
MoleDB Database
MoleMarshal uploads 'stamps' to Zooniverse
Zooniverse Database
MoleMarshal downloads classifications
Zooniverse users classify stamps

Molegazer

Detection skin via hue and saturation, then detect background colour.

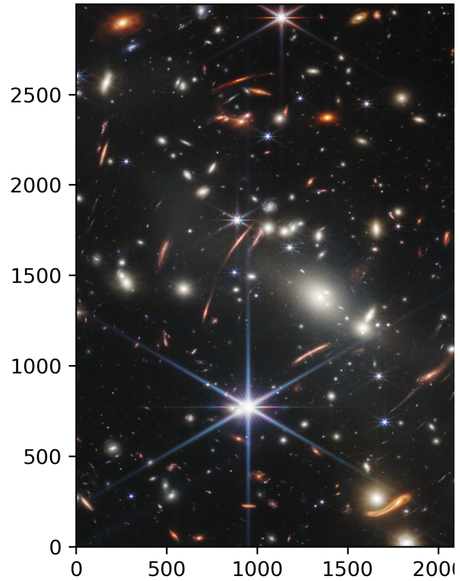
Use the Laplacian of Gaussian method

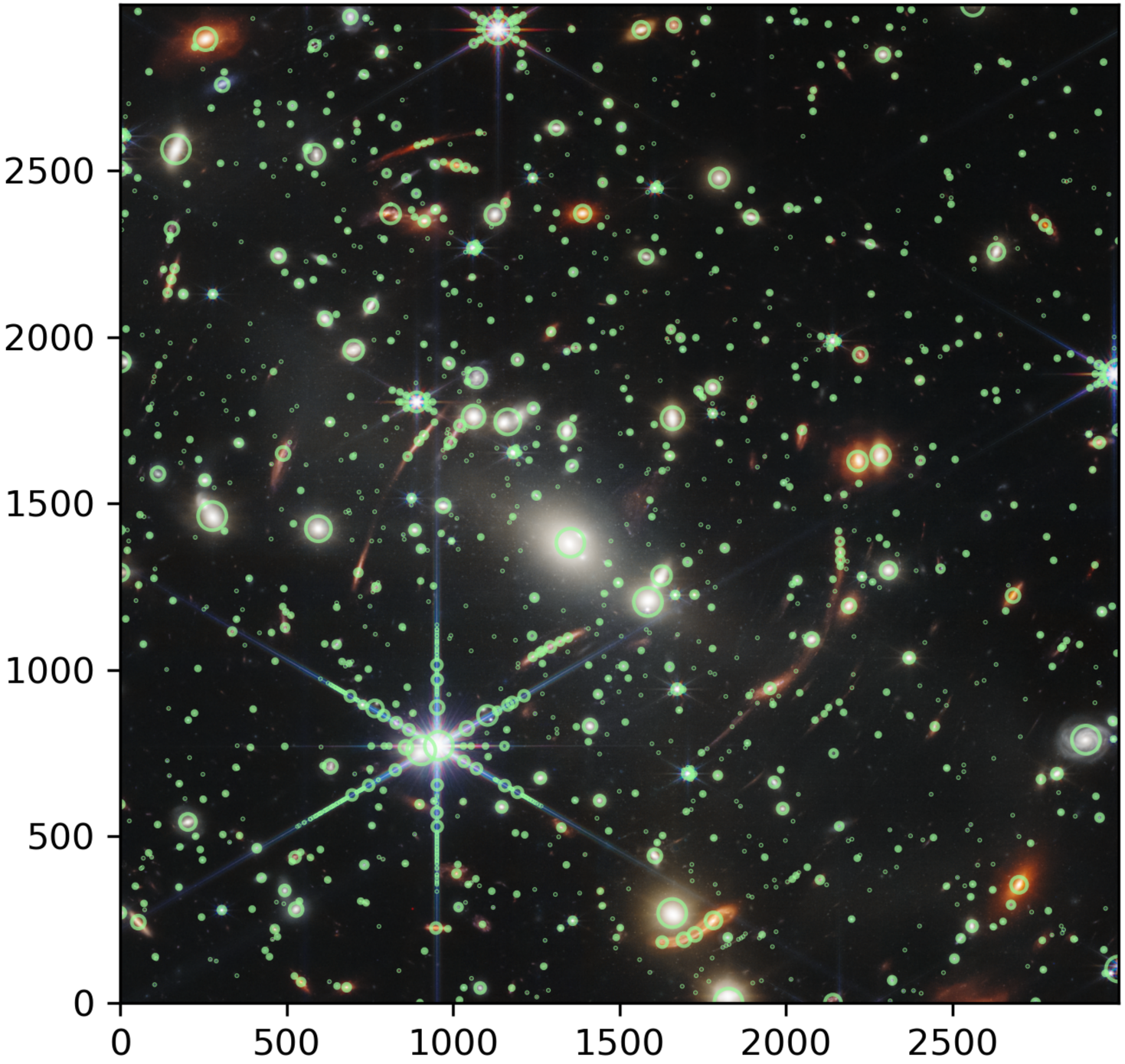
- Good for blob detection and edge detection
- Greyscale the image
- Apply a Gaussian blur, then take the second derivative of the pixel intensity map.
  - When crosses 0 consider it an edge.

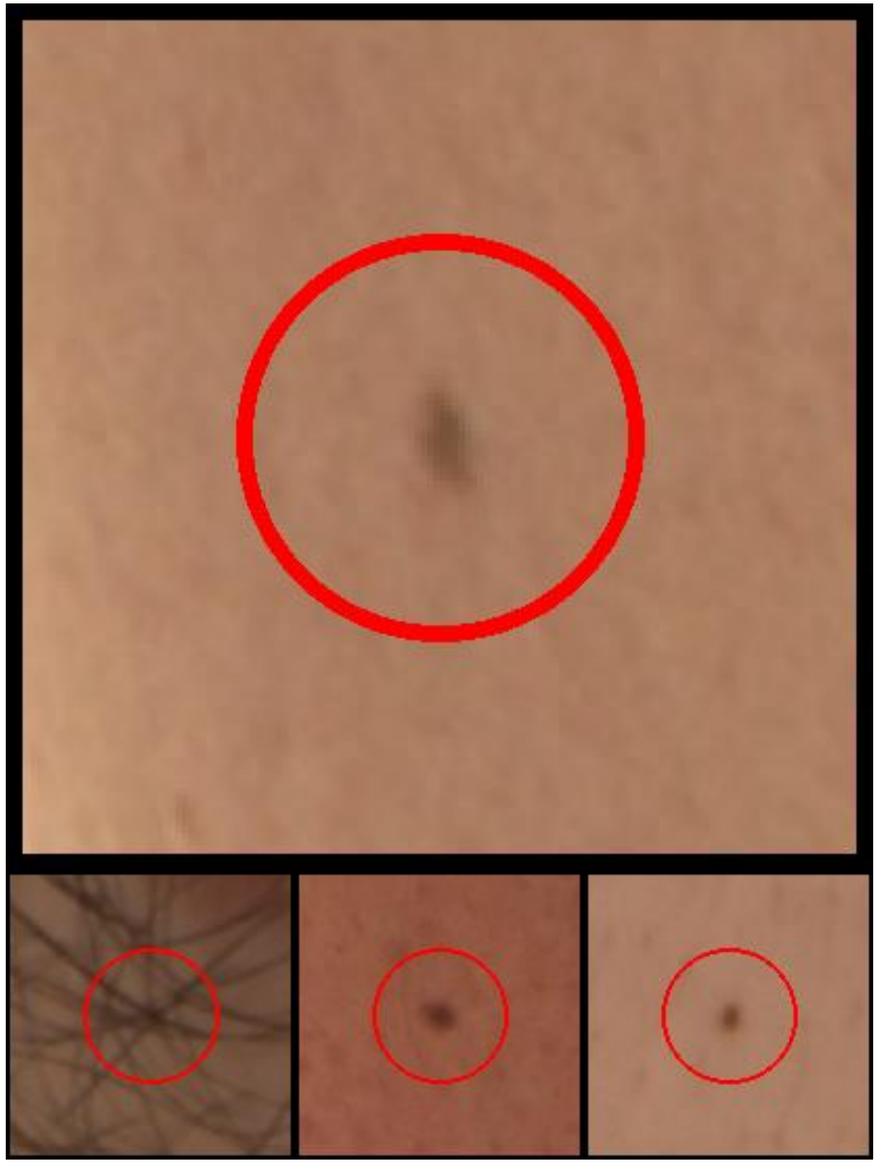


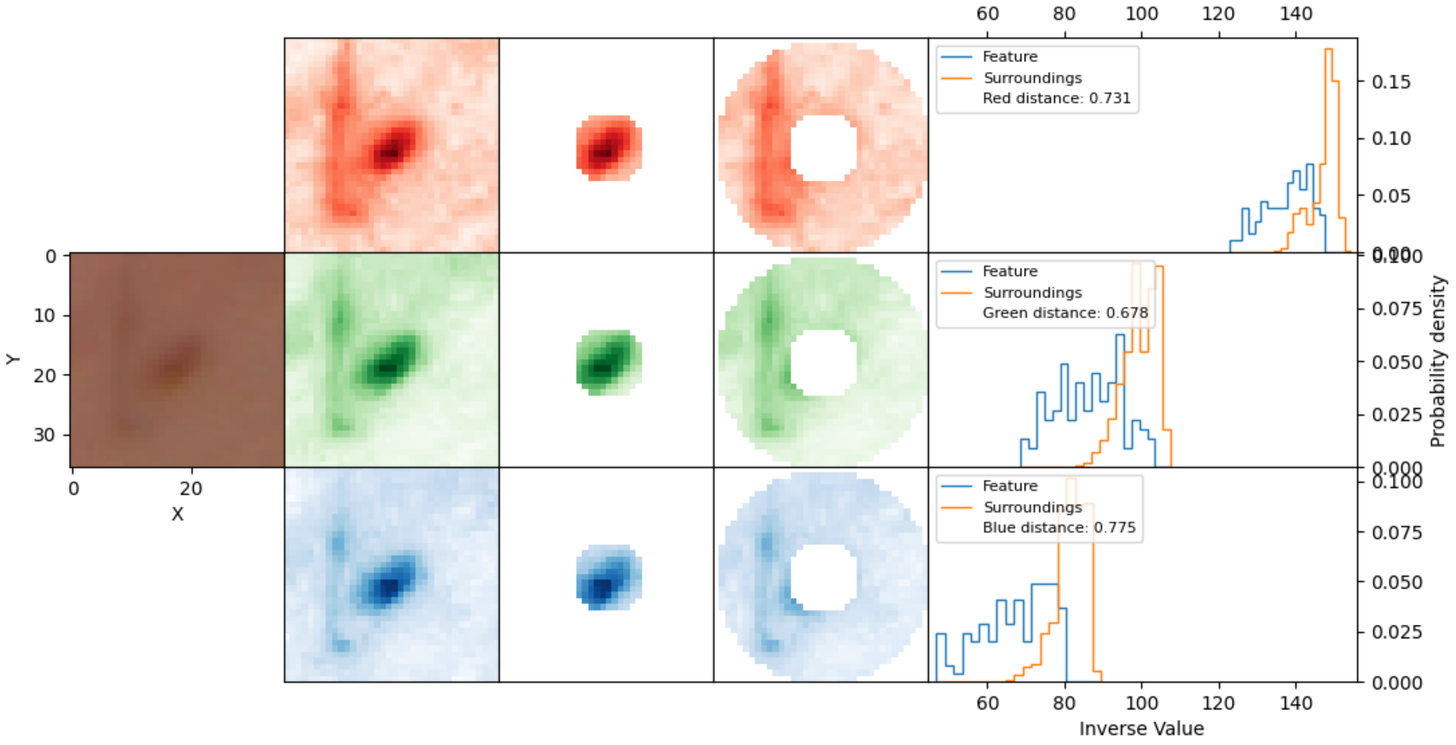
```python
from skimage.feature import blob_log
from skimage.color import rgb2gray
from skimage.io import imread
import numpy as np
import matplotlib.pyplot as plt

webb_deep = imread('./Webb_s_first_deep_field.png') #Open JWST deep field
webb = webb_deep[1000:4000,1000:4000] #Small area for demo

webb_gray = rgb2gray(webb) #Convert to Grayscale

#Compute Laplacian of Gaussian
blobs_log = blob_log(webb_gray, max_sigma=30, num_sigma=10, threshold=.1)

#Size is blob is sqrt(2)*sigma for 2D image
blobs_log[:, 2] = blobs_log[:, 2] * np.sqrt(2)

fig, axes = plt.subplots(1, 1, figsize=(5, 5))
axes.imshow(webb, origin='lower')
for blob in blobs_log:
        y, x, r = blob
        c = plt.Circle((x, y), r, color='palegreen', linewidth=1, fill=False, alpha=0.7)
        axes.add_patch(c)
plt.show()
```

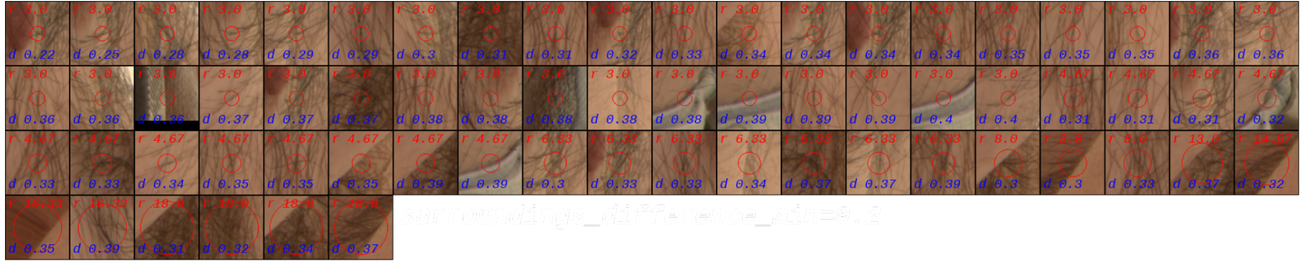
```python
from skimage.feature import blob_log
from skimage.color import rgb2gray
from skimage.io import imread
import numpy as np
import matplotlib.pyplot as plt

webb_deep = imread('./Webb_s_first_deep_field.png') #Open JWST deep field
webb = webb_deep[1000:4000,1000:4000] #Small area for demo

webb_gray = rgb2gray(webb) #Convert to Grayscale

#Compute Laplacian of Gaussian
blobs_log = blob_log(webb_gray, max_sigma=30, num_sigma=10, threshold=.1)

#Size is blob is sqrt(2)*sigma for 2D image
blobs_log[:, 2] = blobs_log[:, 2] * np.sqrt(2)

fig, axes = plt.subplots(1, 1, figsize=(5, 5))
axes.imshow(webb, origin='lower')
for blob in blobs_log:
        y, x, r = blob
        c = plt.Circle((x, y), r, color='palegreen', linewidth=1, fill=False, alpha=0.7)
        axes.add_patch(c)
plt.show()
```

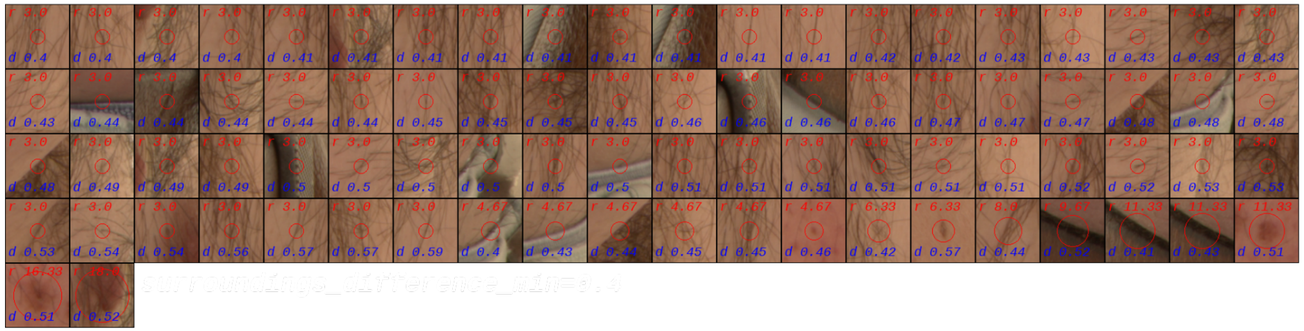




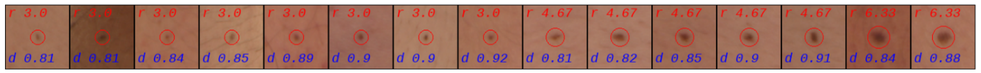



Feature
Surroundings
Red distance: 0.731
Feature
Surroundings
Green distance: 0.678
Feature
Surroundings
Blue distance: 0.775
Y
X
Inverse Value
Probability density


surroundings_difference_min=0.2


surroundings_difference_min=0.4


surroundings_difference_min=0.6


surroundings_difference_min=0.8

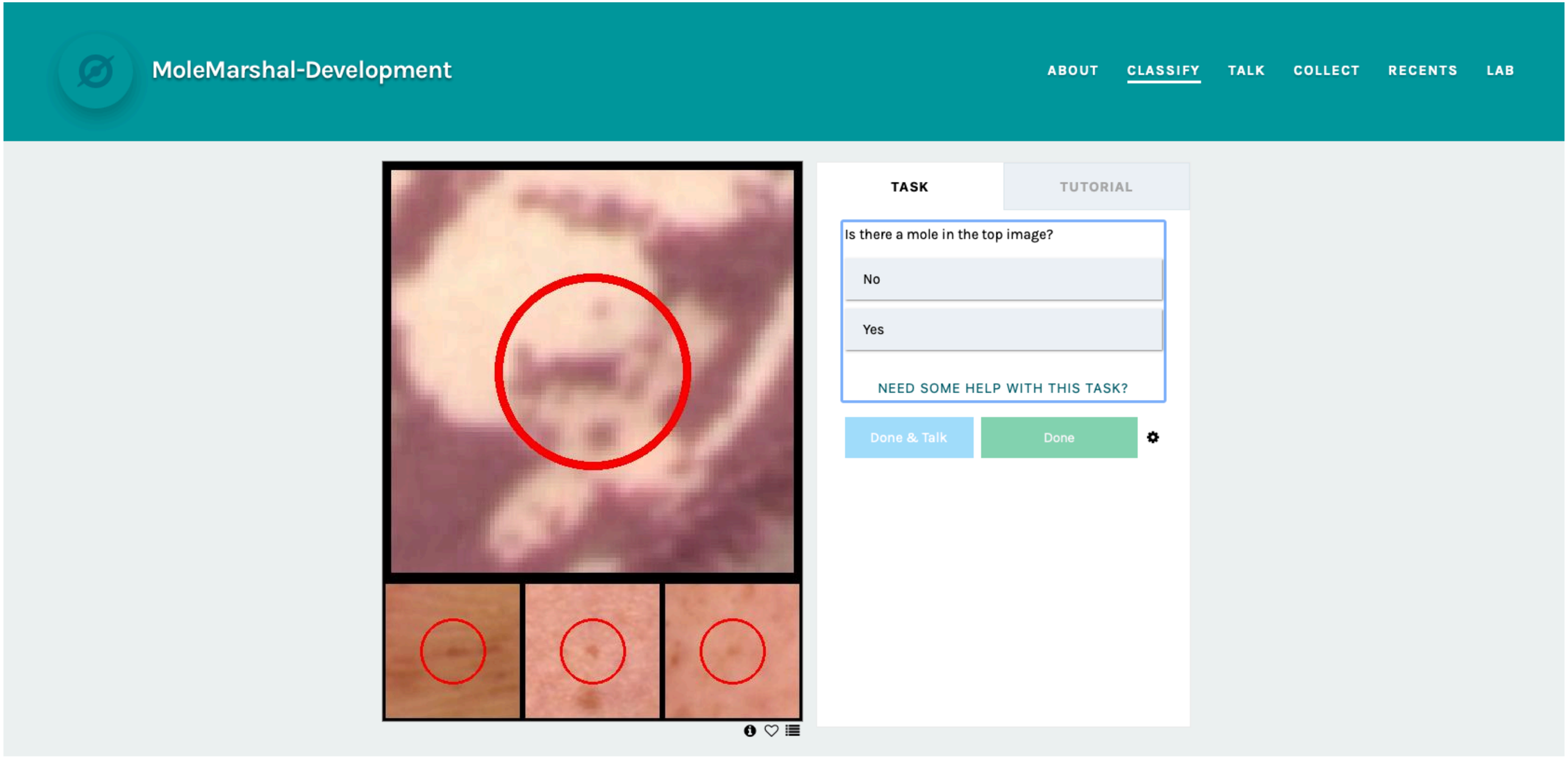

PIXELATED FOR PRIVACY

MoleMarshal

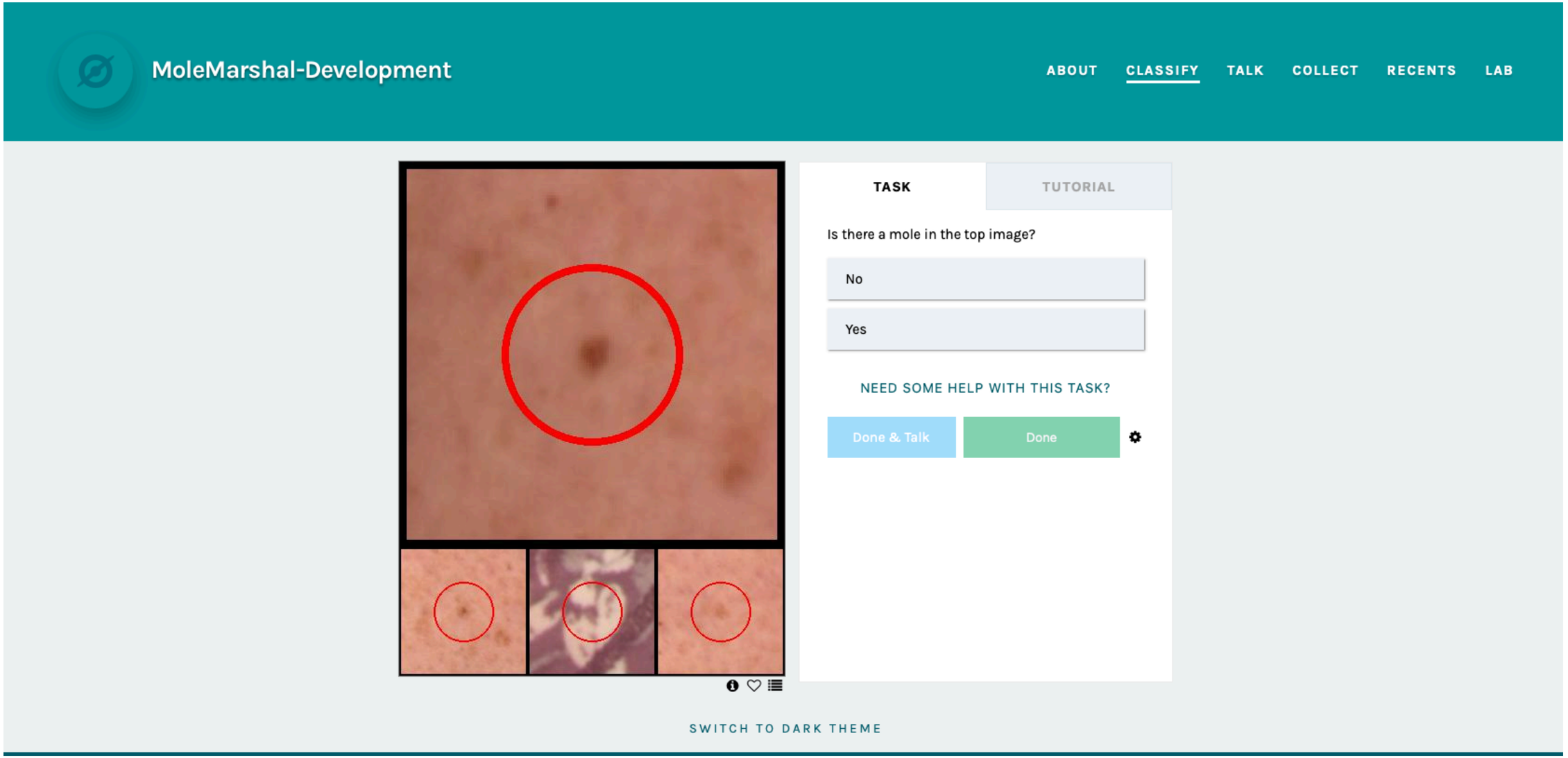
Labelling the data

- Citizen Science platform engaging the public with research
- Made famous for asking public to label images of galaxies
- MoleMarshal is a private project
  - Patient sensitive data
  - Only accessed from behind UoS firewall
- Consultant dermatologists can look through the data and label it
- (I am not allowed to demo this - patient confidentiality)

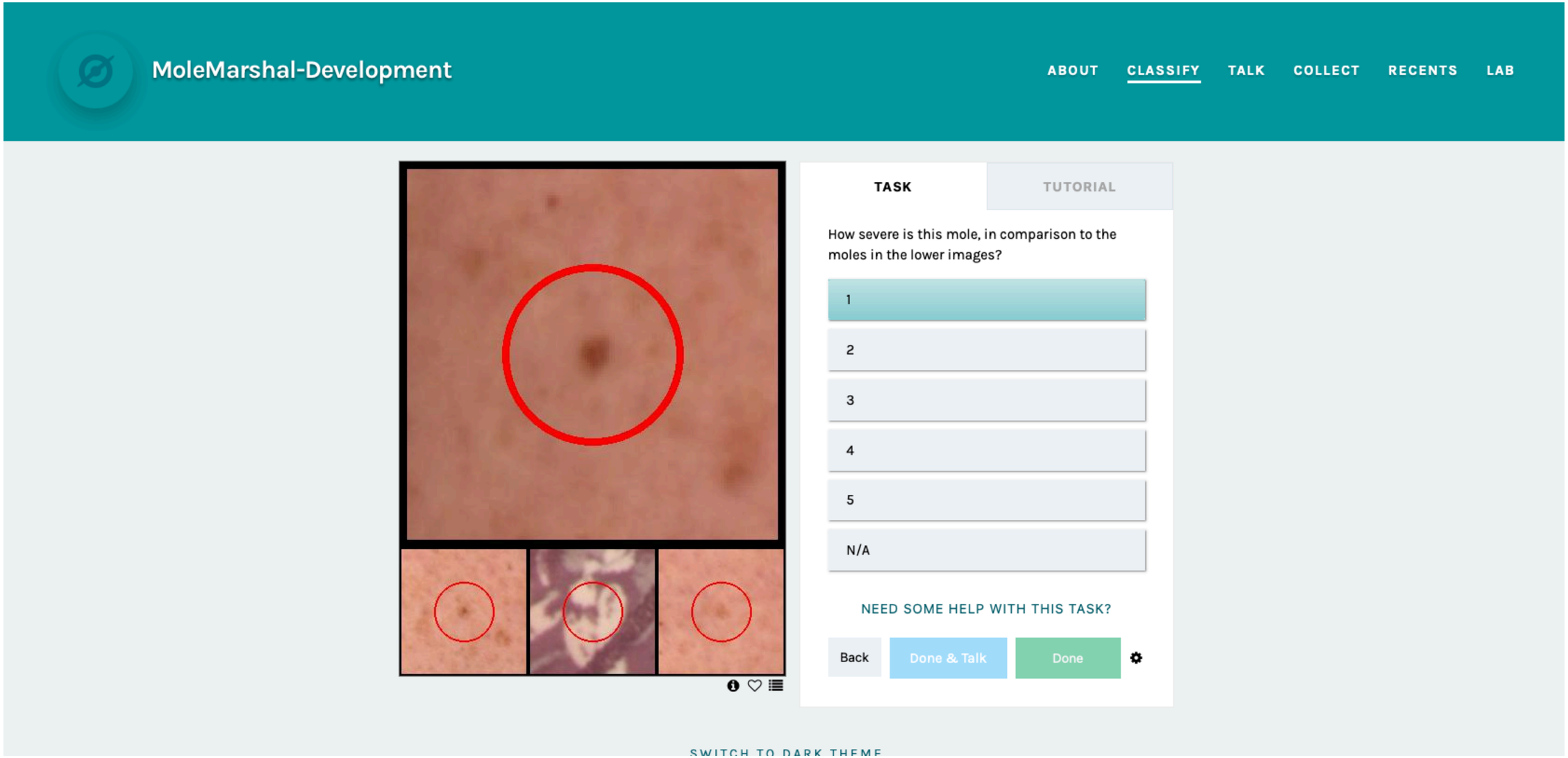
Interface

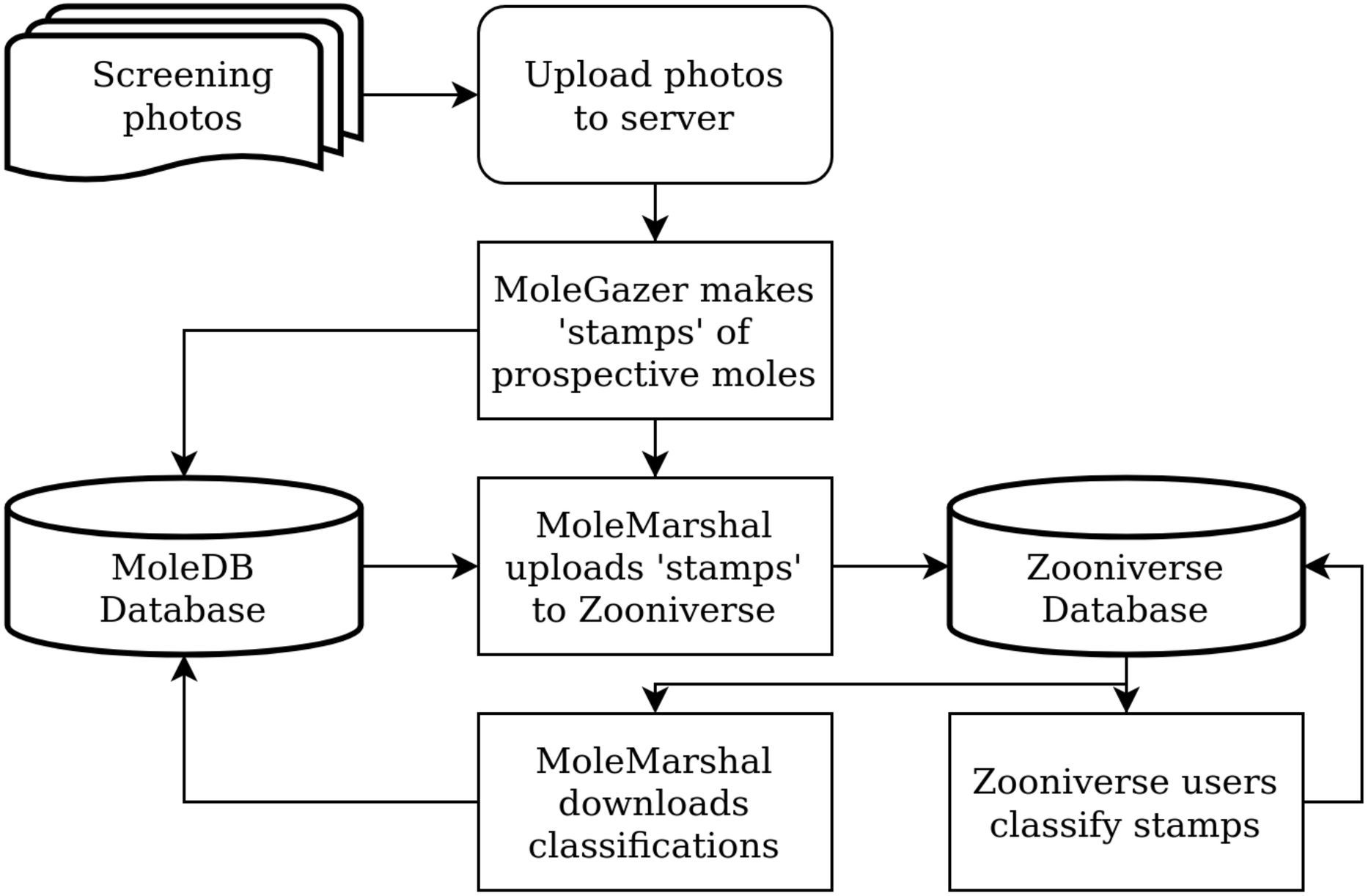

MoleMarshal-Development
ABOUT
CLASSIFY
TALK
COLLECT
RECENTS
LAB
TASK
TUTORIAL
Is there a mole in the top image?
No
Yes
NEED SOME HELP WITH THIS TASK?
Done & Talk
Done

Interface

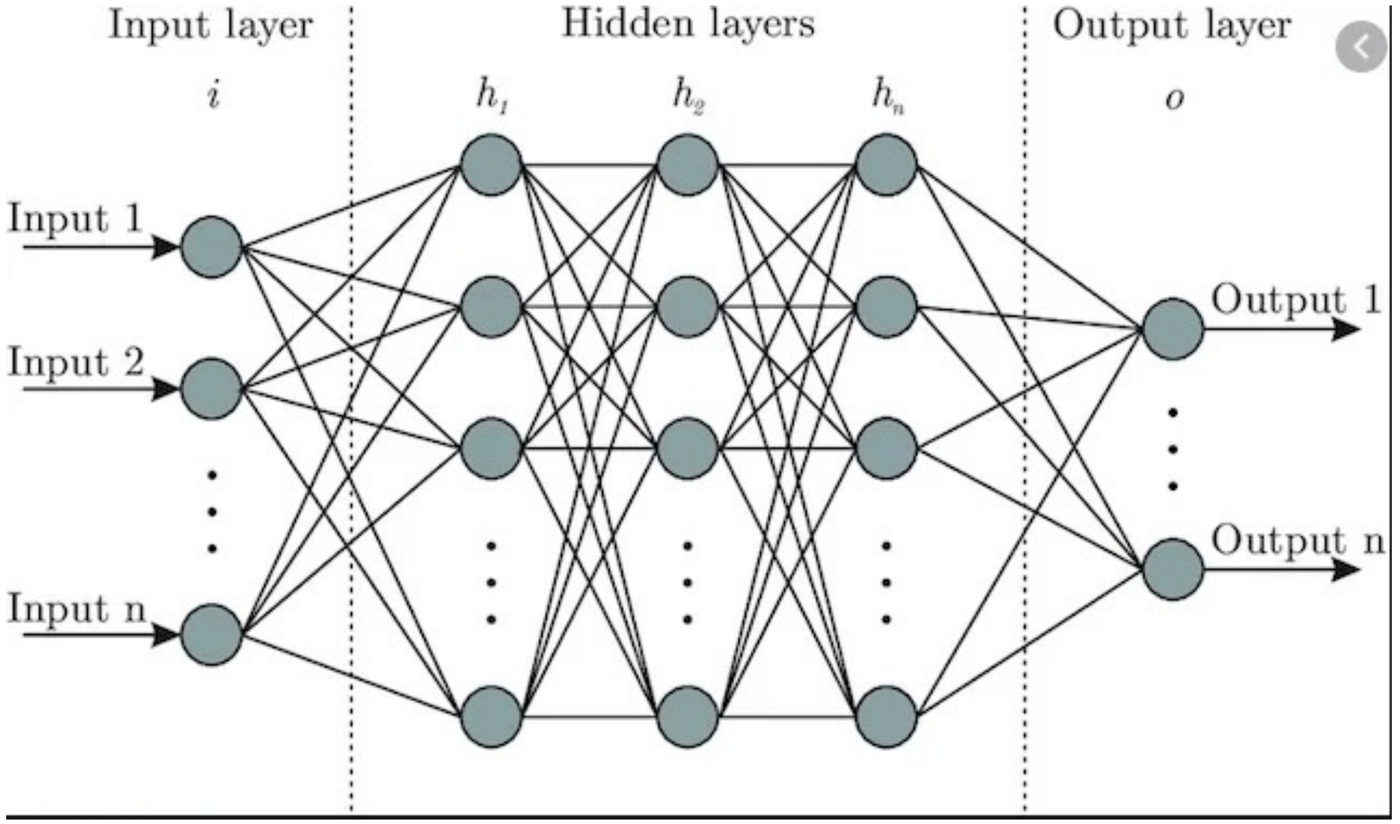

MoleMarshal-Development
ABOUT
CLASSIFY
TALK
COLLECT
RECENTS
LAB
TASK
TUTORIAL
Is there a mole in the top image?
No
Yes
NEED SOME HELP WITH THIS TASK?
Done & Talk
Done
SWITCH TO DARK THEME

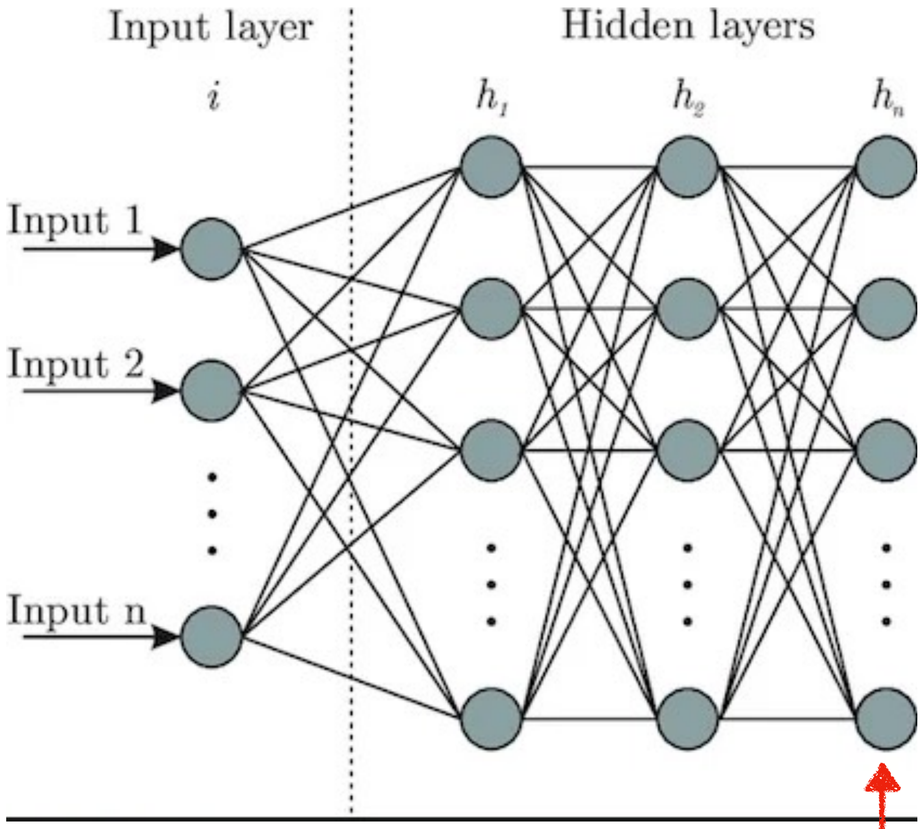
Interface


MoleMarshal-Development
ABOUT
CLASSIFY
TALK
COLLECT
RECENTS
LAB
TASK
TUTORIAL
How severe is this mole, in comparison to the moles in the lower images?
1
2
3
4
5
N/A
NEED SOME HELP WITH THIS TASK?
Back
Done & Talk
Done
SWITCH TO DARK THEME

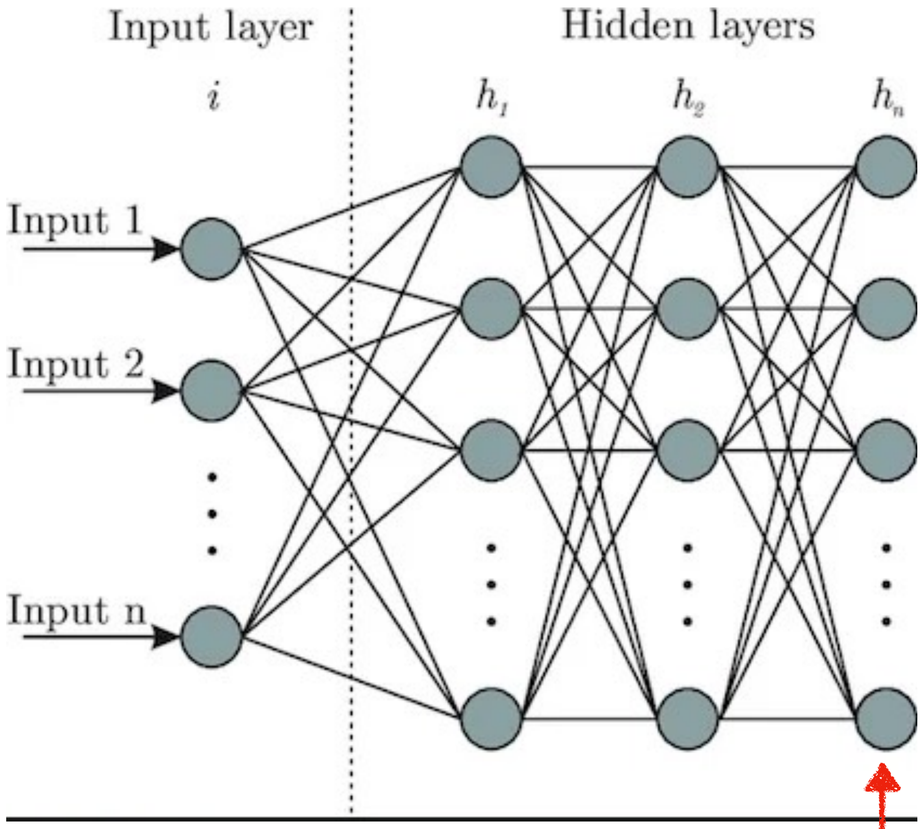

Screening photos
Upload photos to server
MoleGazer makes 'stamps' of prospective moles
MoleDB Database
MoleMarshal uploads 'stamps' to Zooniverse
Zooniverse Database
MoleMarshal downloads classifications
Zooniverse users classify stamps

Finding similar moles… efficiently label data


Input layer
Hidden layers
Output layer
i
h_1
h_2
h_n
o
Input 1
Input 2
Input n
Output 1
Output n



Only use the values at these nodes



Only use the values at these nodes

Similar regions in this nD-space would be occupied by similar looking moles?

```
# for loading/processing the images
from keras.preprocessing.image import load_img
from keras.preprocessing.image import img_to_array
from keras.applications.inception_resnet_v2 import preprocess_input

# models
from keras.applications.inception_resnet_v2 import InceptionResNetV2
from keras.models import Model

# clustering and dimension reduction
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA

# for everything else
import os
import numpy as np
import matplotlib.pyplot as plt
from random import randint
import pandas as pd
import pickle
import glob
```

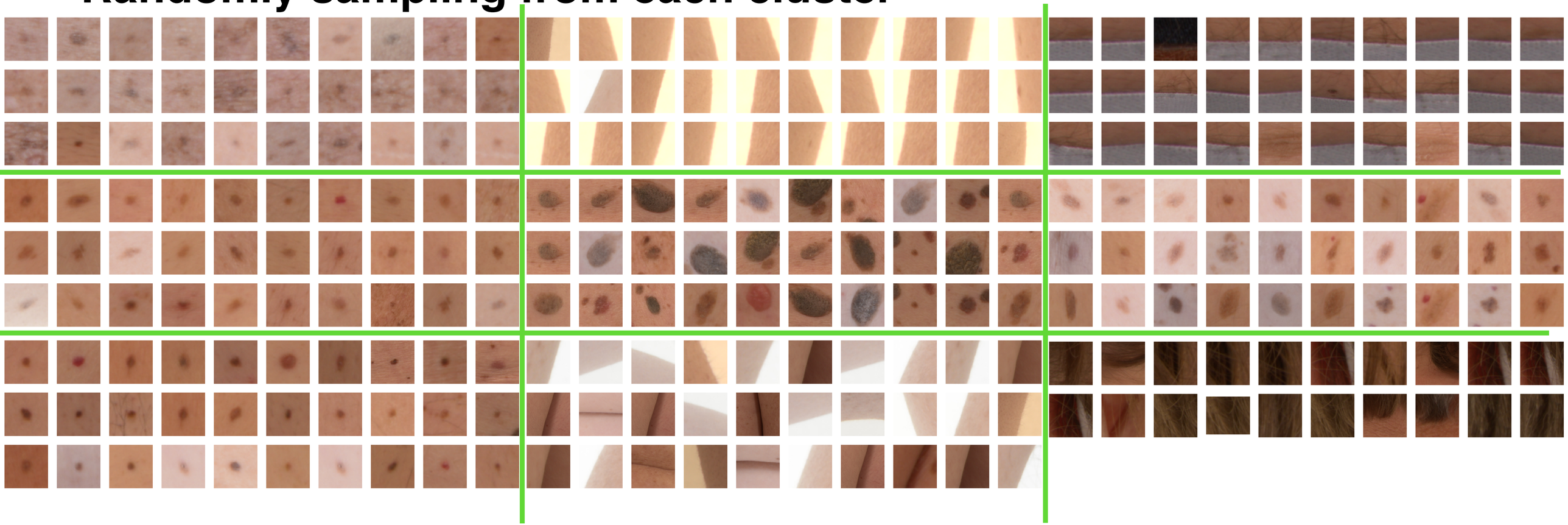

# load model
model = InceptionResNetV2(
weights='imagenet', # Load weights pre-trained on ImageNet.
input_shape=(299, 299, 3),
)
model = Model(inputs=model.inputs, outputs=model.layers[-2].output)

```
imageList = glob.glob('/my/very/private/medical/data/stamps/*.png')

def extract_features(file, model):

    img = load_img(file, target_size=(299,299))
    img = np.array(img)
    reshaped_img = img.reshape(1,299,299,3)
    imgx = preprocess_input(reshaped_img)

    features = model.predict(imgx, use_multiprocessing=True)

    return features

data = {}
for mole in imageList:
    feat = extract_features(mole,model)
    data[mole] = feat
```

```
pca = PCA(n_components=100, random_state=42)
pca.fit(feat)
x = pca.transform(feat)

numberOfClasses = 10 #Guess how many classes

kmeans = KMeans(n_clusters=numberOfClasses, random_state=42)
kmeans.fit(x)
```

Plot the clusters

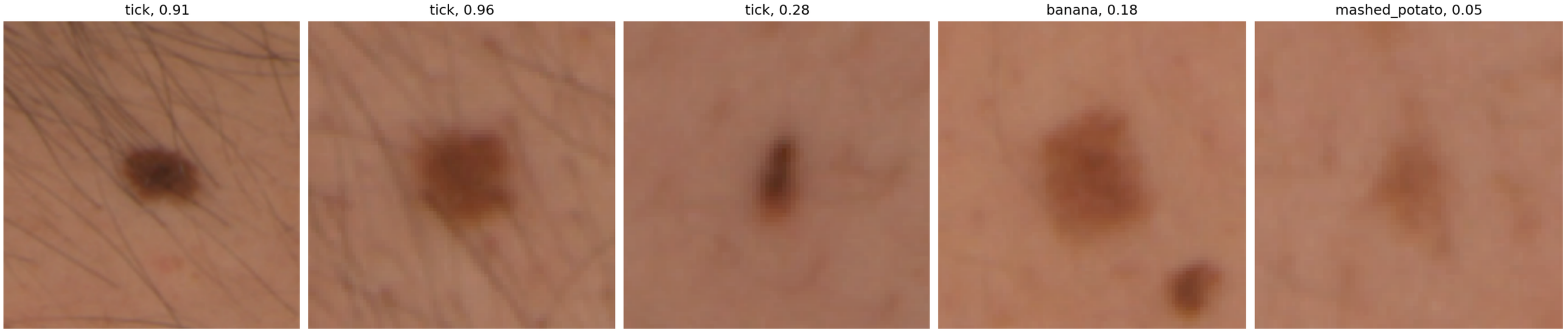
Randomly sampling from each cluster



A more efficient labelling system

(For fun)

What did the classifier predict?



Future work

- Things seem to work on single snapshots!
  - We can detect moles on patients
- More diverse cohort!
- Need to build in time series
  - Tracking moles over multiple years
  - How to align patients so we know we are looking at the same moles
  - Astronomy techniques don’t work reliably for this problem.

Questions?

Dr Chris Frohmaier

chrisfrohmaier ✉️ cfro.astro@gmail.com cfrohmaier