CISC3023 Machine Learning

Course Project

Wound area location in animal model images

In this project, you will be given the following data set:

Animal Model Images.

Number of instances (records in your data set): 150
Number of outputs (the values you need to predict from the inputs): 4

Relevant Information:

This data set contains 150 images for animal models, the image for each data is 3264x2448 pixels. One image is shown below for your reference.

Basically, the images are saved as jpg files. In the folder there is a csv file containing the filenames of the images in the folder, the location of the wound area (it is an oval: **x, y, x_width,** and **y_width** are given).

Location of center 🔴 **: [x, y]**

Heigth of oval │ **: y_width**

Width of oval ——— **: x_width**

For image and csv file access, the short demo code included in UMMoodle provides a naive solution.

Project requirement:

In this project, the student should design **two types of machine learning models (KNN is not allowed)** to predict the location of wound area, or the 4 values: **x, y, x_width, y_width** from the animal model image:

- Implement a Python program (**trainer**) that can be used to import historical data from image files, prepare the data and train ML models to predict 4 location-related values. Then the program will save the trained models.
- Implement another Python program (**tester**) that can be used to import the test data from a folder and use the saved ML model to predict wound location from images in the folder. Demonstrate the performance of these models on the test data in this program. (This means we have ground truth of test images in a csv file for your reference)
- Try to find the best predictor (in the two machine learning models with different settings).
- *Input:*
  Your programs should be able to accept (read) historical data files (jpg files and csv files) for the **trainer** and testing data files (jpg files) for **tester**
- *Output:*
  Your **trainer** program should output some performance indexes and the trained prediction models. Your **tester** program should output prediction results for the test data by using the trained models.

What to Submit at UMMoodle before the deadline:

1.   Well annotated Python source code for training and testing.

2.   Stored machine learning models, and the **tester** program can handle test data in a folder.

3.   Project report must include the detail explanation on the design of:

   (1) Two types of machine learning models used. Why did you choose them?
   (2) How do you handle the prediction of 4 outputs?
   (3) How do you prepare your data to get a better result?
   (4) How to choose the parameters of different models?
   (5) How do you make the best model selection?
   (6) What is the performance of the two types of machine learning models, how to improve their performance?
   (7) What are the samples that are difficult to predict the wound location, do you have any explanation about them?
   (8) Can you visualize the prediction results?

4.   You should include a conclusion section for your own additional comments and discussions

5.   A detailed discussion of the comparison of different models and the selection of the best model should be included.

6.   Submit above all your files in one zip file. Provide detailed **instructions** on how to check and run your codes and program.

IMPORTANT NOTE

1. Students may be asked to give a demonstration of their project.
2. There will be at most two students in a group and they can share the code. But the reports should be written separately.
3. Plagiarism is not tolerated.
4. Late submissions will not be accepted. It is your responsibility to submit via UMMoodle.