Seeking particle occupancy info for 2D classes

Mon, 05/28/2018 - 14:45

#1

drichman

Seeking particle occupancy info for 2D classes

I'd like to:

(a) sort the 2D class images (in Results > 2D Classify, after selecting the refinement package and classification round) by particle occupancy (% of particles going into each class)

(b) recover that particle occupancy information for all the classes.

So far I've tried snooping around the GUI (I understand cisTEM is pretty new and might not have these features yet), looking in the metadata of the .mrcs in Assets/ClassAverages, and looking at what kinds of info are stored in class average-related tables in the SQL database but I haven't come across particle occupancy data yet.

Many thanks for this powerful and pleasant program!

timgrant

Hi,

The 2D classification in cisTEM is integrated over angles, shifts and class. The individual likelihoods for all the different angles/shifts/classes are not stored (as it would be a lot of information). The class with the best occupancy is stored, but the actual occupancies are not stored on disk, they are only held in memory.

If you would like to know the best class for each particle you can get that information from the database. First you need to know the classification_id for the classification you are interested in. Running :-

```
sqlite3 name_of_your_database.db "select * from
classification_list
```

will list all the classifications. Find the one you want and read off the classification id (it will be the first number in the column).

You can then get the best class for each particle with the following command :-

```
sqlite3 name_of_your_database.db "select best_class
from classification_result_$your_class_id;"
```

I hope this is of some use, although I realise it is not the full occupancy info.

Cheers,

Tim

Mon, 05/28/2018 - 19:27 <u>(Reply to #2)</u>

drichman

Thanks & some bash

Thanks Tim, that makes sense, looks like I can basically count up the number of times a class is chosen as the best one for the particle and that's another way of looking at which are the most occupied classes. In case it's of any use to anyone, here's what I've done:

To get the looong list of class identities into a file:

```
sqlite3 name_of_your_database.db "select best_class
from classification_result_21;" >
best_class_for_each_particle.txt
```

Then a bash loop to count how many times each class identifier (number from 0 to 80) appears in that list:

```
for i in `seq 0 80`; do grep -o $i
best_class_for_each_particle.txt | wc -l; done
```

Now this output can be used in a histogram or compared with the images of each class. The class numbers seem to go from 1-80 so I'm guessing 0 where some unclassifiable particles wound up?

Tue, 05/29/2018 - 09:31 <u>(Reply to #3)</u>

drichman

Although, I'm a bit puzzled,

Although, I'm a bit puzzled, the counts add up to more than the number of particles in the list. I'll update this if I figure out why...

Tue, 05/29/2018 - 13:22 <u>(Reply to #4)</u>

timgrant

The problem is, is that when

The problem is, is that when you grep 1 you will also include all those in class 11, 12, 13 etc, and you have similar problems with other numbers. I believe you can fix this by adding the -w flag which matches whole words only, so you would run :-

```
for i in `seq 0 80`; do grep -ow $i
best_class_for_each_particle.txt | wc -l; done
```

Also, class 0 is used for those images not included in the classification, this can be because less than 100% of particles were used, or if you have exclude blank edges set to yes, those excluded will come out as best class 0.

Thanks!

Tim

timgrant

Thanks so much for posting

Thanks so much for posting this!

I'm sure other people will find this useful.

Cheers,

Tim

Log in or register to post comments