🏆 Featured Prediction Competition

2018 Data Science Bowl

Find the nuclei in divergent images to advance medical discovery

$100,000
Prize Money

Booz Allen · Booz Allen Hamilton · 738 teams · 6 months ago

Overview   Data   Kernels   **Discussion**   Leaderboard   Rules   New Topic

---

Annotation strategy 1: Hand-outlining using GIMP   🔶 ▲ 15

posted in 2018 Data Science Bowl 9 months ago

Anne Carpenter

Hi everyone!

It took us a while to work out a reasonably efficient scheme for annotation so we thought we would share. Strategy 2 was semi-automated; Allen Goodman is going to share the tool for that shortly.

I do not know what is typical for these competitions, but if you are doing annotations, please consider tracking the images and their permissions so that others can make use of them after the challenge is over (for example, all the images we annotated for this challenge are licensed CC0 - public domain, per our thread).

Note that these are very rough instructions we wrote for internal use so please don't expect them to be exhaustively nicely documented. The thing missing from below is the last step: how to turn outlines into masks of the format used in the challenge. More info on that soon.

P.s. as far as I know we are intentionally withholding information about the split between strategies 1 & 2 for Stage 1 & 2 test sets. So please do not ask!

SETUP of GIMP

- Open GIMP
- If the Toolbox is not shown, go to Windows > Toolbox
- If Tool Options are not shown, go to Windows > Dockable Dialogs > Tool Options
- If Layers are not shown, go to Windows > Dockable Dialogs > Layers
- If Undo History is not shown, go to Windows > Dockable Dialogs > Undo History Choose the pencil tool in the Toolbox
- Set its color to red in the Toolbox (See the two largish rectangles with a curvy arrow

- Set its color to red in the Toolbox (See the two largish rectangles with a curvy arrow pointing to them both? Click the topmost rectangle (foreground) and type in HTML notation = ff0000 to get red, then click OK)

- Set the background color to blue by selecting the bottommost rectangle (background) and type in HTML notation = 0000ff to get blue, then click OK)

- Set the pencil size to 1 pixel in the Tool Options

- Choose the Eraser tool and set it to "Dynamics off", and "Hard edge"

- To save the current GIMP settings, go to "Edit" -> "Preferences" menu. In the "Preferences" window, click on "Tool Options" in the left hand sidebar. Then click on "Save Tool Options Now" button. At this point, the current GIMP settings will have been changed permanently, and the next time you run GIMP, you will see the same current settings.

FOR EACH IMAGE

- Open a new image (File > Open): be sure to choose one that does NOT already have a .xcf companion file.

- Go to Image > Mode > RGB to change it to colors (unless it's already RGB!)

- Create a new layer for outlines using menu bar Layer > New Layer. Call it "outlines" and leave settings at default (e.g., "Transparency" for fill type)

- That layer will now be selected: it's highlighted in the Layers window and must remain so to do your work.

- Making sure to remain in the "outlines" layer, and making sure the pencil tool is selected, draw red outlines around nuclei! Well, actually just inside nuclei: you want the outline you draw to be included in the final object, not outside the final object. -- If you are too far inside the boundary, just continue pushing the pixels closer to the edge: it's ok if the outlines are a few pixels thick (eventually each outline will be filled in anyway). EXCEPTION: if you are near the junction between two cells you should erase the extra pixels because the system won't know where to put the real boundary. The pipeline seems to do a pretty decent job of putting the dividing line towards the middle, but I'm not 100% sure about this. -- If you go outside the boundaries, just hit Undo (yellow arrow in the Undo History panel) or Ctrl-Z if a keyboard is handy. If you only notice it later, you can use the Eraser tool which is found somewhat nearby the pencil tool.

- Click at least one blue pixel somewhere in the background of the image (where no nuclei exist), especially around the edges of the image if some background is "trapped" at the edge between two actual nuclei. To choose blue instead of red, click the arrow in between the largish rectangles in the Toolbox window so blue becomes foreground. Don't forget to switch back! In cases where there is a small chunk of background in between the junctions of several nuclei, you also need to give it at least one blue pixel so post-processing does not think it's a small nucleus. EXCEPTION: if there is just a single pixel of background in between two nuclei, don't worry about it; the pipeline will exclude these based on being one pixel in size plus having most of their borders touching other nuclei.

- Save your work frequently! The first time you do File > Save it will offer to save it with the same name, but using a .xcf extension. This is what you want. Save it to your desktop folder of images. You can stop and start anytime using that file.

EXPORTING FINAL RESULT

EXPORTING FINAL RESULT

- Save the image first using File > Save as described above. This is critical.
- Hide the "background" layer (by clicking the "eye" icon), so that only the "outline" layer is visible.
- Then File > Export As and choose PNG. The default values are OK. Add "outlines" to the filename so it doesn't overwrite the original image.
- Close the .xcf image. Important: it will ask if you want to save changes. You want to DISCARD changes because you have the background layer hidden!

CHECKING FOR ERRORS

- Run the "outlines" version of each image through the CellProfiler pipeline we made (Make_objects_from_outlinesNN.cppipe. It will save some images to your output folder that you can examine. Look for places in the image that might be: -- missing blue dots such that background is detected as a nucleus, especially at image edges and inside clumps of nuclei -- not-completely-closed outlines, which will appear as rings instead of whole nuclei -- Too-thick border between or even right nearby adjacent nuclei that cause leakage of the border
- Correct errors as needed in the .xcf file and when finished, re-export the outlines layer as a PNG, as before.

SETUP of the tablet (if needed)

- If the icons/menus on a Surface tablet are too small, go to Windows menu, settings, resolution, change the size of text apps to ~200%
- On a Surface tablet, go to Windows menu, settings, pen: change "Ignore touch input when I'm using my pen" to On

Options

Comments (7)

Sort by   Hotness ▾

Please sign in to leave a comment.

Anne Carpenter   [Competition Host]   •   6 months ago   •   Options   ∧ 1

As asked in another thread, we didn't keep track of how many images/nuclei can be done by this method per hour. Besides, it might make me cry to know the total amount of my life dedicated to this task.

If I had to guess, I would say 1 nucleus took 2 seconds but the downstream work is not insignificant so probably double that for the "overhead" of file handling and checking the annotations via the CellProfiler pipeline, etc.

CPMP   •   6 months ago   •   Options   ∧ 1

We cannot thank you enough for having labelled that data. This was not only fun for us to work on the result, but it will probably help advance science quite a bit. Thanks a lot for that.

Abner Chou · (341st in this Competition) · 9 months ago · Options

∧ 0

Thanks for sharing. This should pin to the top of forum.

devin · 9 months ago · Options

∧ 0

Hi Anne,

I have a question about where to place the blue background pixels. Should they be inside the cell nuclei? And if so how many are required, you mentioned that they should be particularly near the edges, is this the edge of the image, or of the division between the nuclei, and non nuclei regions.

Thanks for the annotation strategy outline, it has spead up the process of getting a pipeline running greatly.

Anne Carpent... [Competition Host] · 9 months ago · Options

∧ 1

The blue background pixels should be in the background of the image (where no nuclei exist). I've updated the strategy to be a bit more clear about this.

Amir · 22 days ago · Options

∧ 0

Thanks for your sharing!

I am wondering if the pipeline(Make_objects_from_outlinesNN.cppipe) you mentioned in STEP 'CHECKING FOR ERRORS' is available ? my email address is yuewensong@163.com

I am working on a similar pipeline , but I met a configuration error

"*cannot run pipeline: The pipeline cannot be started because of a configuration problem. The pipeline did not identify any image sets. Please correct any problems in your input module settings and try again.*"

would you give me a hint? Looking forward to your reply! Thanks!

HELENE LECOQ · a day ago · Options

∧ 0

Hello, I want to use your solution but when I use the Hide the "background" layer (by clicking the "eye" icon), so that only the "outline" layer is visible. But all hide. Why. Thanks a lot

Our Team   Terms   Privacy   Contact/Support