Image Segmentation

Project 3

Deep Learning in Computer Vision

June 2022

In this project, you will be working with data from the ISIC skin lesion segmentation and classification challenge (Fig. 1). The overarching task associated with the dataset is the automatic diagnosis of skin lesions based on images of them – however, the main focus of this project will be on segmenting those same lesions. As you will see in the final part of the project, these segmentations can prove valuable also for the diagnosis part.

The dataset is comprised of images found from `https://challenge.isic-archive.com/data/`, subsampled to create the datasets found on `/dtu/datasets1/02514/isic`.

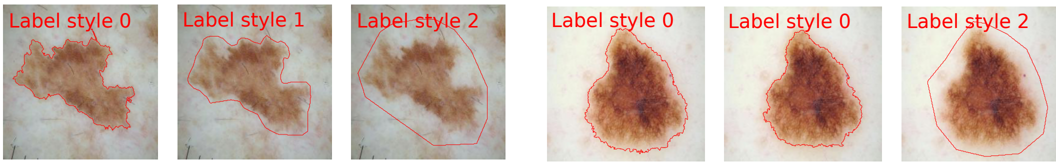

Figure 1: For the ISIC dataset you will find annotations made with three different label styles: Style 0 gives a close, detailed crop; Style 1 gives a relatively close but smooth crop, whereas Style 2 gives a coarse, polygonal crop. In this project, you will study how different label styles can bias your segmentation, and you will also study how the coarse annotations might be used to improve segmentation.

Your tasks are as follows:

1. **A basic segmentation model.** First, you will set up a basic segmentation pipeline using the training set found at `/dtu/datasets1/02514/isic/train_allstyles`. Create a data loader and implement a U-net architecture to segment the images. Train it using the entire training set, selecting a validation set from the training set if you find one necessary. Make sure to describe all your modelling choices on your poster.

   As you will see later, the training set contains several copies of the same image but with different segmentation masks – in this first task, you should treat these as separate training images.

   Test your trained segmentation network using the test set found in `/dtu/datasets1/02514/isic/test_style0`, and compute whichever evaluation metrics you find useful. Please discuss your results, and any weaknesses or strengths that you observe.

2. **Making a better model.** Try to improve the segmentation, e.g. using data augmentation, positive weights to make up for class imbalance, other loss functions, varying architecture, or using various optimization tools. Describe what you did, and what the result was.

3. **Bias from training data.** As you can see in Fig. 1, your training set consists of images with 3 associated annotations each, where these annotations are made using 3 different styles (you will not see every style for every image. Two of the annotation styles (styles 0 and 1) result in a mask that closely follows the boundary of the lesion, whereas the third annotation style is coarser and easier for an annotator to make, as it is comprised of a smaller number of corners that together outline a polygonal border. These segmentations are consistently too big.

   First, compare the sizes of the predictions made by your basic segmentation model on the test data to the sizes of the target masks, which all belong to annotation style 0. Does your basic segmentation model overestimate the masks?

   Next, train your segmentation network using only masks with either annotation style 0, 1, or 2. These can be found in `/dtu/datasets1/02514/isic/train_style0`, `/dtu/datasets1/02514/isic/train_style1` or `/dtu/datasets1/02514/isic/train_style2`. Repeat the experiment to quantify whether or not your segmentation networks are biased towards oversegmenting the lesions. Please describe what you see.

4. **Weak annotations for segmentation.** In class, we discussed the use of weak annotations such as bounding boxes or image level annotations. In this part of the exercise, you will experiment with weak annotations for segmentation.

   First, note that the coarse polygonal masks are essentially just a little finer than bounding boxes. Try to apply the methods discussed in class for bounding box based segmentation to training on these too coarse masks.

   Next, try to compare this to using saliency maps from image classification to obtain object segmentation based on image level labels. For this, you need to train a CNN for image classification. To this end, you will find in `/dtu/datasets1/02514/isic/background` a set of images that do not contain any lesions.

   For both methods, again assess your performance using the same metrics as above, and also checking whether you are overestimating the lesions using the same assessment as above.

   For both of these experiments, please keep your architectures and training procedures as close as possible to those applied earlier, in order to be able to compare results across experiments.

5. **Optional: Using segmentation to alleviate bias for melanoma classification.** As discussed in the second segmentation lecture, CNN-based methods for diagnosing skin lesions have been suspected of being biased by confounding elements in the images: Dermatologists often insert rulers into dermatological images in order to document the size of a lesion, and they are more prone to insert rules when they suspect the lesion to be malignant. On `/dtu/datasets1/02514/isic/ISIC18` you find a dataset, independent of those that you have used earlier in this exercise, which consits of 1) images, 2) lesion diagnosis labels, and 3) a csv file documenting whether or not the image contains the following confounding elements: Rulers, ink, and stickers. For this dataset, rulers are the most commonly found potential confounder.

   Using this dataset, train a classifier to predict skin lesion diagnosis, and try to document that your classifier is confounded by the rulers, namely that it uses the rulers to perform the diagnosis. If you conclude that it is confounded, try to unbias the classifier by segmenting the lesions before training and testing, as this should remove most rulers – they are very rarely touching the lesions. You are also welcome to try other approaches. Please describe what you did, and what you observed.

Your process, performance evaluation and results should be documented and discussed in a PDF poster to be uploaded on DTU Learn.

References

[Kohl 2018] Kohl, Simon, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R. Ledsam, Klaus Maier-Hein, SM Ali Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. "A probabilistic u-net for segmentation of ambiguous images." In Advances in Neural Information Processing Systems, pp. 6965-6975. 2018.