CS230: Lecture 5

Case Study

Kian Katanforoosh

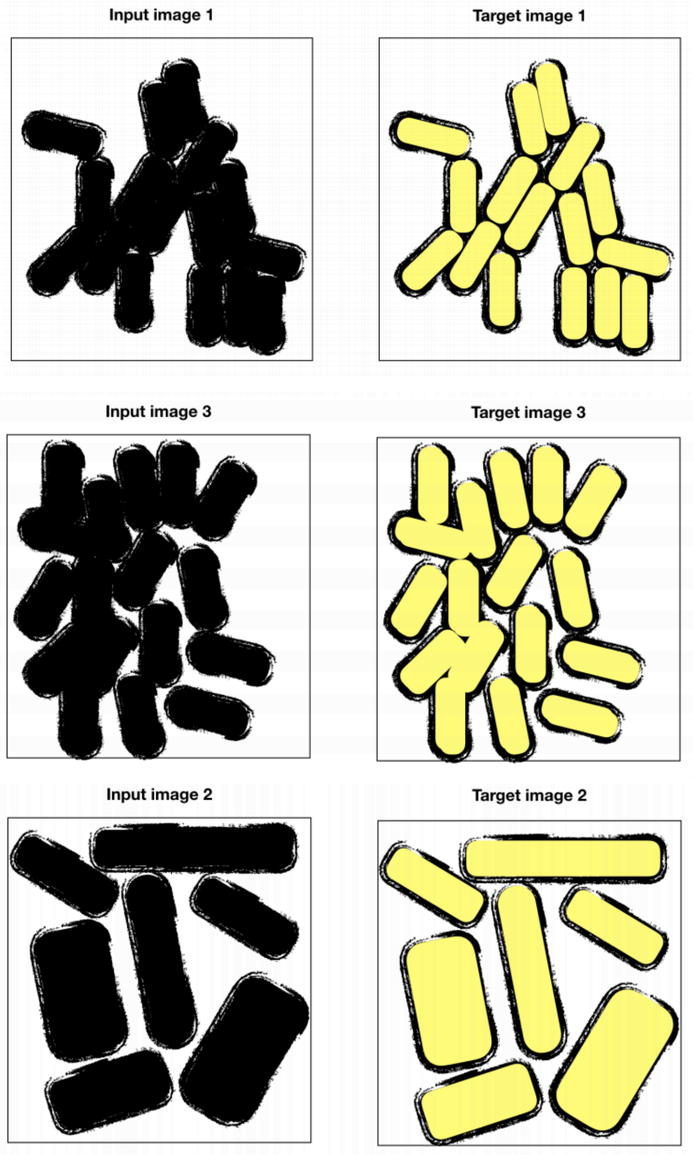
Problem statement: cell segmentation

Goal: Determine which parts of a microscope image corresponds to which individual cells.

Data: Doctors have collected 100,000 images from microscopes and gave them to you. Images have been taken from three types of microscopes:

Type A	50,000 images
Type B	25,000 images
Type C	25,000 images

Question: The doctors who hired you would like to use your algorithm on images from microscope C. How you would split this dataset into train, dev and test sets?



Data

Question: The doctors who hired you would like to use your algorithm on images from microscope C. How you would split this dataset into train, dev and test sets?

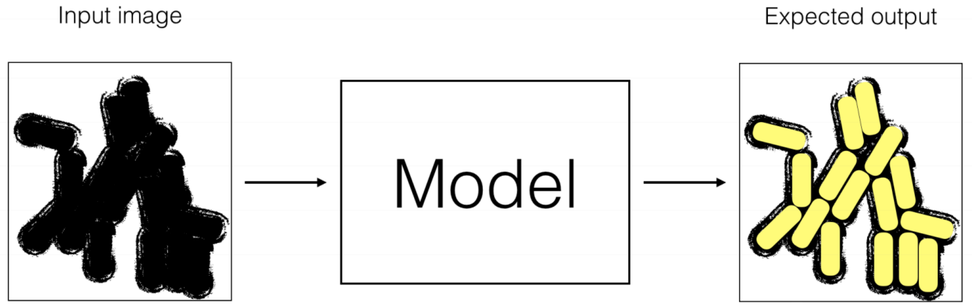
Answer:
i) Split has to be roughly 90,5,5. Not 60,20,20.
ii) Distribution of dev and test set have to be the same (contain images from C).
iii) There should be C images in the training as well, more than in the test/dev set.

Question: Can you augment this dataset? If yes, give only 3 distinct methods you would use. If no, explain why (give only 2 reasons).

Answer: Many augmentation methods would work in this case:
– cropping
– adding random noise
– changing contrast, blurring.
– flip
– rotate

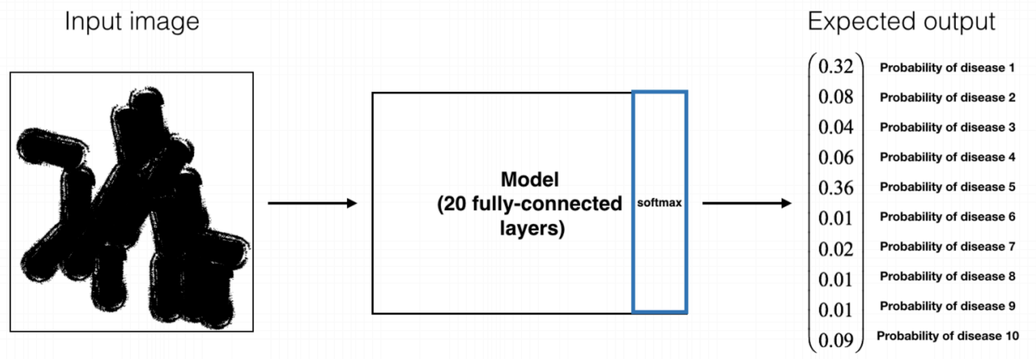
Transfer Learning

First try: You have coded your neural network (model M1) and have trained it for 1000 epochs. It doesn't perform well.



Transfer Learning: One of your friends suggested to use transfer learning using **another labeled dataset** made of 1,000,000 microscope images for skin disease classification (very similar images).

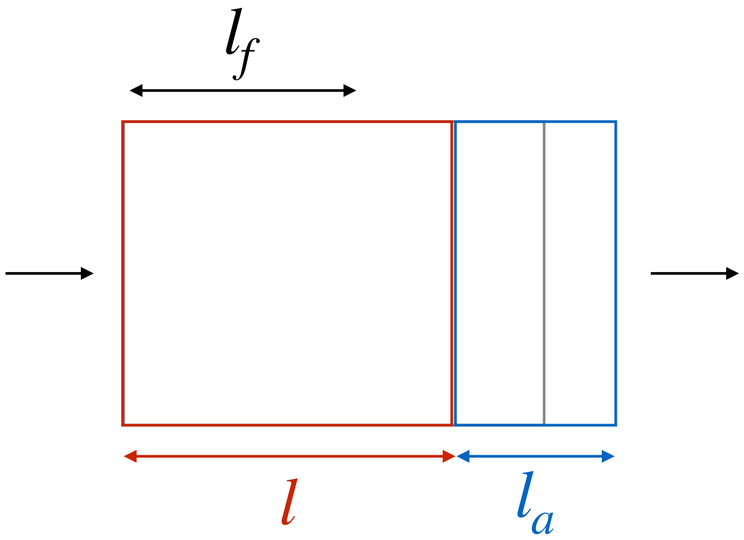
A model (M2) has been trained on this dataset on a 10-class classification. Here is an example of input/output of the model M2.

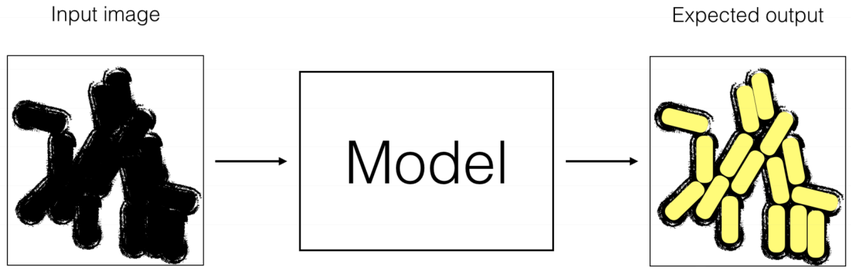


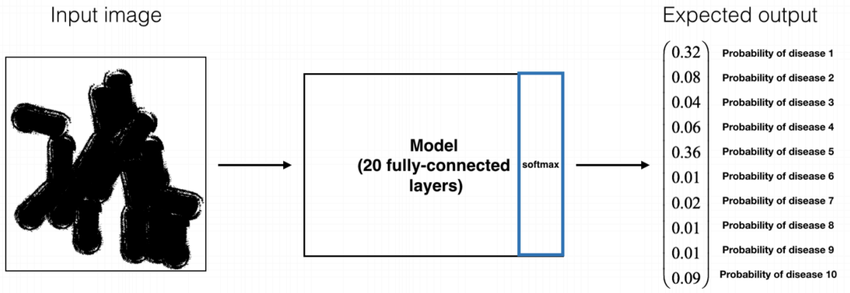
Question: You perform transfer learning from M2 to M1, what are the new hyperparameters that you'll have to tune?

Transfer Learning

Question: You perform transfer learning from M2 to M1, what are the new hyperparameters that you'll have to tune?





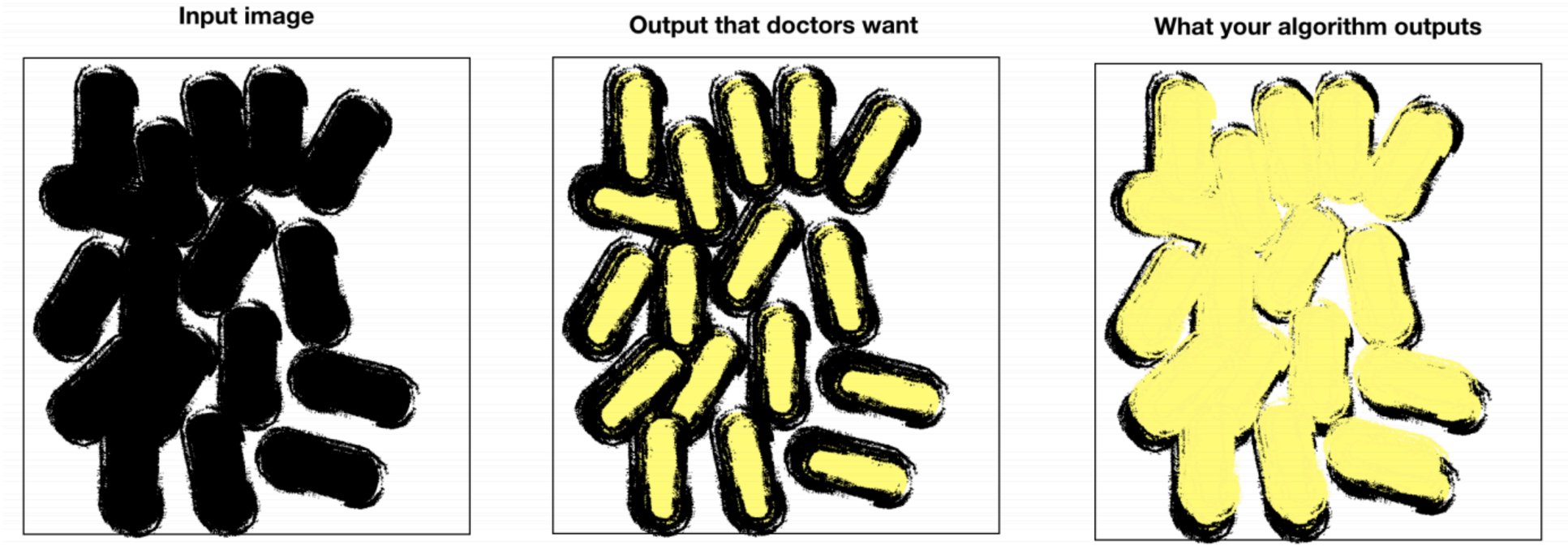


l = number of layers transferred from M2

l_a = number of new layers added to the new model's head

l_f = number of frozen layers

Network modification



Question: How can you correct your model and/or dataset to satisfy the doctors' request?

Answer: Modify the dataset in order to label the boundaries between cells. On top of that, change the loss function to give more weight to boundaries or penalize false positives.

$$L_{binary} = -\sum_{i=1}^{n_y} (3y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i))$$

$$L_{multi-class} = -\sum_{i=1}^{n_y} w_{y_i} y_i \cdot log(\hat{y}_i)$$

$$w_{y_i} = \begin{cases} 1 & if \quad y_i = (1,0,0) \\ 5 & if \quad y_i = (0,1,0) \\ 3 & if \quad y_i = (0,0,1) \end{cases}$$

Kian Katanforoosh

Duties for next week

- TA Section on Friday 05/08
- Project Meeting #2 has to happen before Friday 05/08 at 11:59 PM
- Project Milestone #1 due Friday 05/08 at 11:59 PM
- C4M1 and C4M2 for next Tuesday 05/12 at 9:00 AM