Deep Learning - Project description

Wednesday 19-9-2021

Normal	Pneumonia
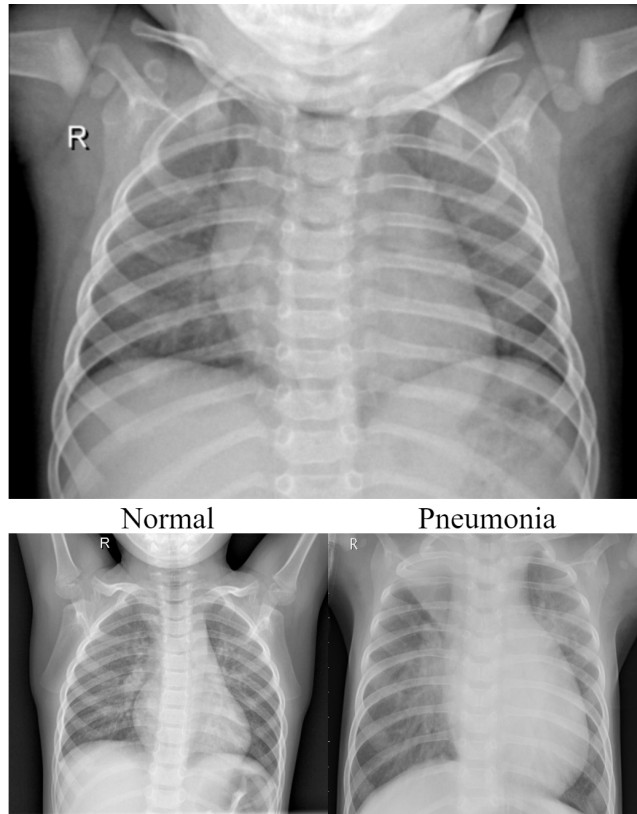
In light of the recent COVID-19 pandemic, we will take a look at one of the severe symptoms of the disease, pneumonia. The dataset we have consists of x-ray images of patients who either has pneumonia or are healthy, and you have to develop a model to distinguish them.

When to submit

You need to submit before Friday 20-8-2021 at 12.00.

What to submit

You must hand-in: your code, your models, and all visualizations saved as images. Your code needs to be heavily commented, as well as specifications as to who worked on each section.

How to submit

You need to submit your material HERE, in the group folder corresponding to the group you were assigned on the first day of the course.

Notes

You must do this assignment in your pre-assigned groups, and only one submission is needed per group. The data is available on ItsLearning under Plans/Day 8/data.zip

Training the networks can be a slow process, keep this in mind. You're NOT allowed to use pretrained networks!. You are allowed to use any networks made during the exercises.

Please read through the entire project description, as there will be a Q&A on Thursday 19th at 12.00. For questions or comments, send and email to *greisager@imada.sdu.dk*. Bear in mind, this should not be used for assistance as in the exercises, as you should try to complete this project on your own.

Task 1 - The Dataset

All images are x-ray images taken in the chest region of children aged 1-5 from the Guangzhou Women and Children's Medical Center. There are 1100 images of healthy humans, as well as 1100 images of people with pneumonia.

Keep in mind, that the image format is jpeg, and there are 3 color channels.

You need to organize the data into directories as shown on Figure 1. You need to determine the training/validation/testing split yourselves, but need to justify your split choice
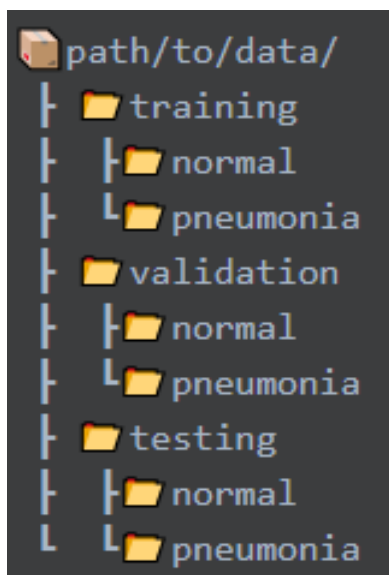

Figure 1: The structure of the dataset.

Task 2 - Create the datagenerators

Due to the high amount of data, you need to write a datagenerator to load the images. This requires you to use the ImageDataGenerator (as shown in Keras Intro - 2) in which you must apply at least one on-the-fly data augmentation. Which type of data augmentation is up to you, but you need to justify your choice in a comment. On top of the ImagaDataGenerator you need to apply the .flow_from_directory function to make it work with the directory tree in Task 1

Task 3 - Constructing the network

After successfully crafting the data generators, you need to craft a neural network, and use the data generators to feed the network. The architecture, complexity and regularizers are all up to you, but you need to justify your choices in comments. You are more than welcome to replicate already known architectures

or architectures we made during the course, but you are NOT allowed to use any pretrained networks. You are also not allowed to use any training data that isn't included on ItsLearning.

Carefully consider which hyperparameters to test and strategically try to find the optimal architecture for the task. In the comments, please describe your method for the optimization and your choice of hyperparameters. Remember that there is an underlying competition, and the highest accuracy wins. The competition will be measured based on the saved model, so make sure to submit only the best one!

Task 4 - Visualizing your results

Finally, you must visualize some aspects of your model. It can be a graph of the training/validation performance, visualization of the filters or feature maps, or anything you can think of. This has to be saved as an image file, and uploaded along with your model and code.