Assignment 5

This exercise is part of the course assignment. **Deadline for the assignment 30.11.2022 at 23:59**

The topic of this assignment is explainable AI. For this assignment you should return

- The file main5.m, the four test images you have picked and a file mdl5.mat containing the trained network you have used. Each file should contain your name and student number (of both students if you work in pairs).
- Your answers to the questions in the analysis part. At the end of the course, you should return a single pdf containing the answers to all questions of the assignments. The report should contain also your name and student number (of both students if you work in pairs).

Coding part (3pt)

Take the trained neural network from assignment 3, for each class (smiling/non smiling) find one correctly classified and one misclassified test image (i.e. 4 images in total). **Note!** If you had saved your trained model, make sure that the training-test splitting is consistent. The easiest way to do so is to fix the random seed at the beginning of the script before training the model.. This way whenever you re-run the script the training-test splitting will be the same. In Matlab, this can be done using the rng function. For each image you should plot the heat maps using

- LIME algorithm
- GRAD-CAM algorithm

Hint: look at exercises 2 and 3.

Analysis part (7pt)

Answer the following questions in your report

- Explain how the GRAD-CAM algorithm works. What are its advantages and limitations? (4pt) The original paper about grad-cam can be found here and the pdf is also attached in the folder.

- Analyse the results of your code. What areas in the image affect the decision for correctly classified images? Does about misclassified images? Does it make sense? (3pt)