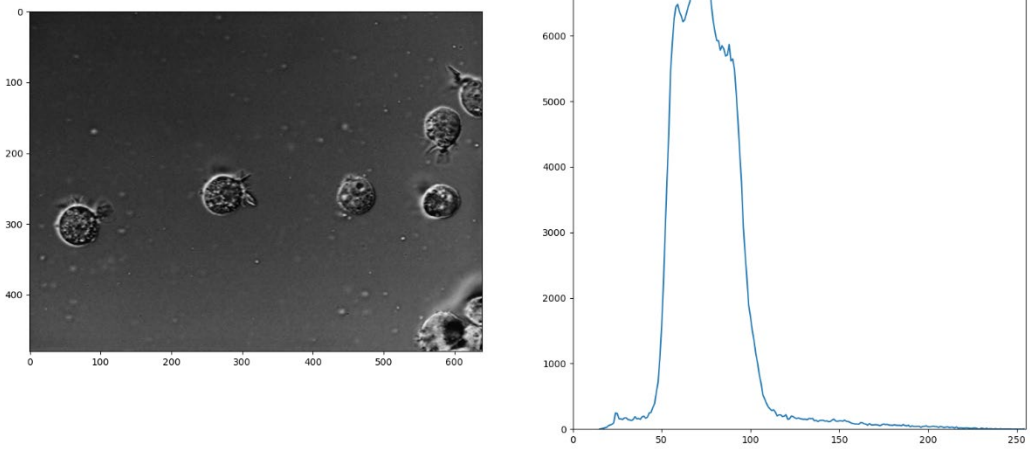
Bachelor's in Data Science and Engineering
2021/22. 4th course

AUDIO PROCESSING, VIDEO PROCESSING AND COMPUTER VISION
ORDINARY FINAL EXAM (19/1/2021)

Student: ……. Grade:

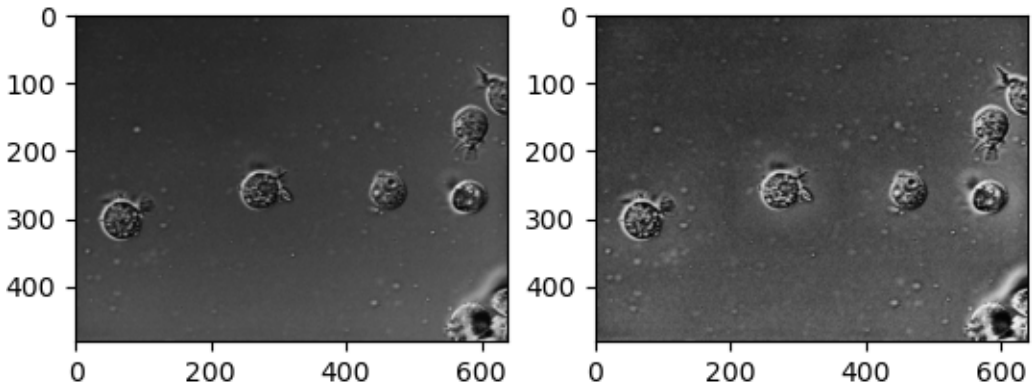
EXERCISE 1 (1 pt)

Consider the following image and its histogram.



a) Draw approximately the pixelwise transformation which implements the histogram equalization. Pay attention to the regions of high slope and label carefully the x-axis.
b) Draw approximately the histogram of image resulting from histogram equalization.

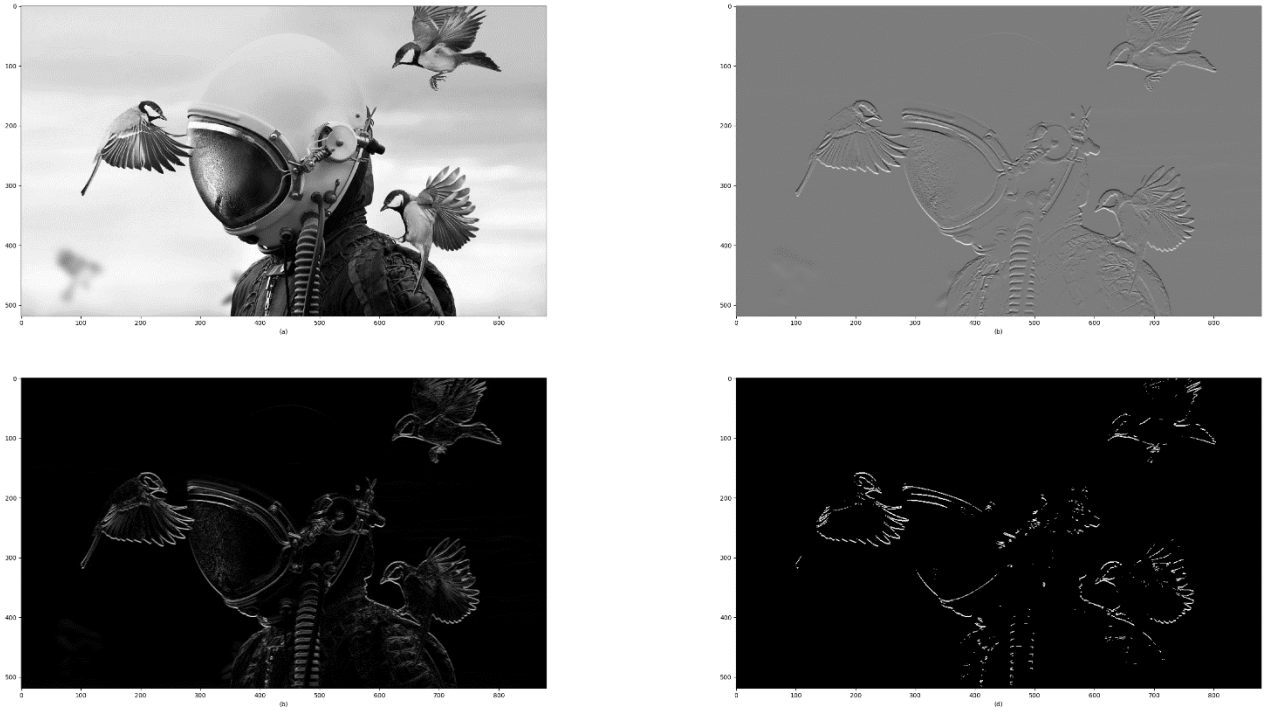
Consider now the image on the right (the image on the left is the original, for comparison purposes)



c) Could it be the result of the histogram equalization? Justify your answer
d) Could it be the result of a contrast-limited histogram equalization? Justify your answer

EXERCISE 2 (1 pt)

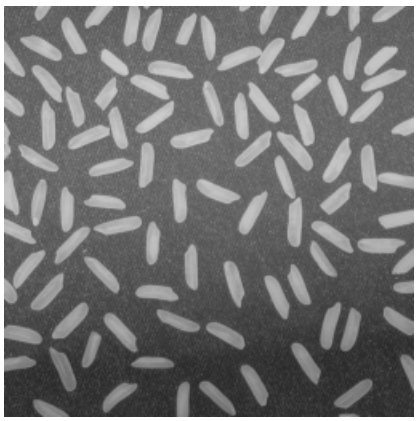
Let us consider the following images, being the one on the upper left of the figure the original image.



a) The image on the upper right has been obtained using a $3x3$ Sobel Filter. Determine the exact coefficients of the (correlation) filter. Justify your answer.
b) How did we get the image on the bottom left?
c) And that on the bottom right?

EXERCISE 3 (1 pt)

Consider the following image showing some rice grains.



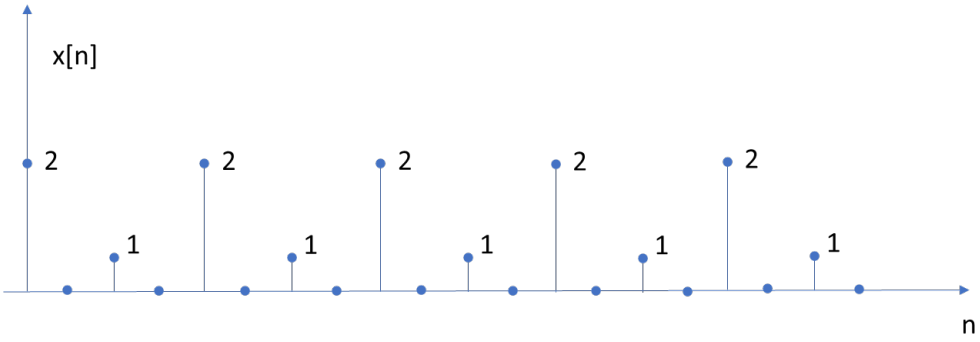
a) Each pixel is replaced with the darkest pixel in a small neighborhood centered on that pixel to produce a new image. Then, the same operation is repeated several times, starting each time with the new image, until the grains disappear. Describe the image that we get as a final result.
b) If we used a global thresholding operation to segment the grains from the background, we wouldn't get a perfect segmentation. Why?
c) How would you use the image resulting from a) to improve the segmentation?
d) Propose an alternate way to produce a good segmentation of the original image.

EXERCISE 4 (1 pt)

The autocorrelation function $R_x[m]$ is useful to estimate the fundamental frequency.

$$R_x[m] = \sum_{n=-\infty}^{\infty} x[n+m]\, x[n]$$

a) How would you use $R_x[m]$ to determine the fundamental frequency of a signal?
b) Provide a numeric example using the following signal:

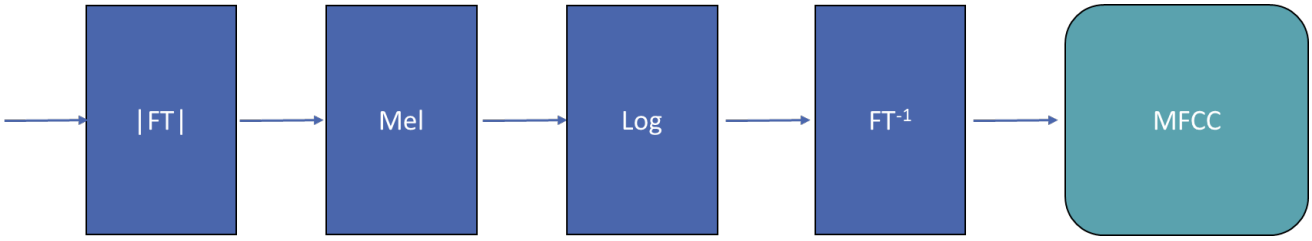


c) Why is it useful the *normalized* autocorrelation function $R'_x[m]$?

$$R'_x[m] = \frac{\sum_n x[n]x[n+m]}{\sqrt{\sum_n x^2[n]}}$$

EXERCISE 5 (1 pt)

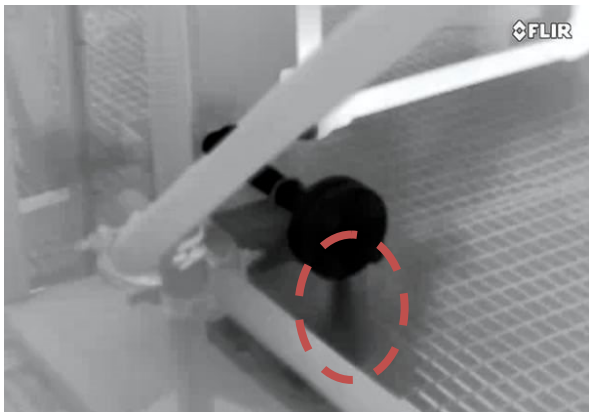
The Mel-Frequency Cepstral Coefficients (MFCCs) are commonly used for speech and audio analysis.

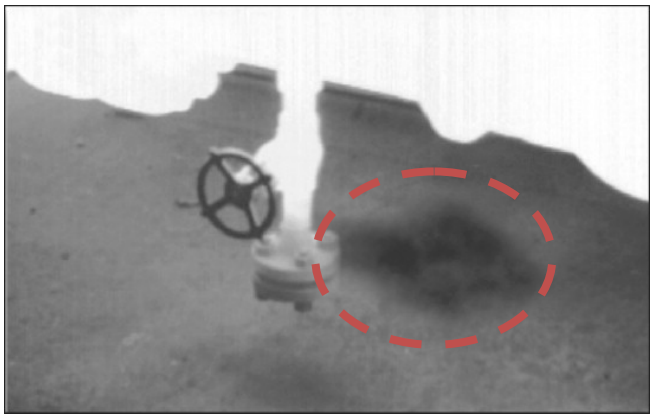


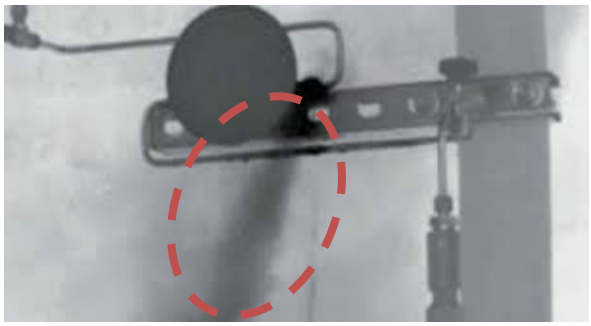
a) What were they designed for?
b) What is the role of the Mel-scale block?
c) What is the role of the logarithm?

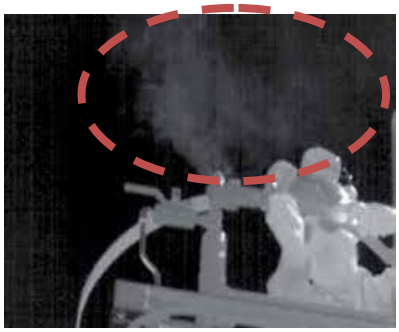
EXERCISE 6 (5 pts)

We aim to design a system to automatically detect gas leaks in videos recorded with infrared cameras. Below are some examples of frames that illustrate our problem:









Inputs to our system are short videoclips (~30 secs video at 25 frames per second) that may contain gas leaks or not. The goal of our task is twofold:

1. First, we need to trigger alarms when a video shows gas. It does not matter if our system shows a short delay of various seconds since the leak starts but it is critical that no video with a gas leak is missed by our system.
2. Second, we would like to estimate the severity of the leak, approximating its flow rate from the number of pixels detected as gas in the video.

We have a training dataset with 1000 short videoclips recorded at different scenarios, in which roughly the 20% correspond to sequences with gas leaks. For testing, the dataset contains 200 videos with similar proportions.

Answer the following questions regarding the system design:

a) (1 pt) Initially, we will tackle the problem of detecting and measuring gas leaks on a per-frame basis, thus considering an individual problem associated with each frame in a video. Compare the following alternatives to address this problem. Discuss factors such as human effort, complexity, accuracy and fulfillment of the required goals:
   a. Label each frame as positive (it includes a gas leak) or negative, and use a classification CNN such as resnet-50 trained on a frame-level binary cross entropy.
   b. Label gas leaks using bounding boxes, and train a faster RCNN with a backbone resnet-50.
   c. Outline gas leaks on each video frame using irregular masks and train a segmentation CNN with a resnet-50 backbone and a pixelwise cross-entropy.

In the following questions, consider that you are using the last approach (c) described in the previous question.

b) (1 pt) Discuss the suitability of the following segmentation CNNs to address our goals:
   a. A fully convolutional CNN in which all layers have a stride=1.
   b. An encoder-decoder architecture with 0<stride<1 in the encoder layers and stride>1 in the

decoder layers.

c. An encoder-decoder architecture with stride>1 in the encoder layers and 0<stride<1 in the decoder layers.
d. A fully convolutional CNN in which all layers have a stride=1, but top layers implement atrous convolutions with factor r>1.

c) (1 pt) Discuss the usefulness of the following data augmentation techniques in your problem: random cropping, rescaling + random cropping, random rotation, mirroring, Color Jitter through brightness, contrast, saturation and hue random changes, Fancy PCA for color augmentation.

d) (1 pt) Explain how you would use the outputs of your frame-level system to produce the intended outcomes (gas detection and severity estimation). Propose metrics to assess the performance of your system in both tasks.

e) (1 pt) Now we would like to enhance the previous system working independently at frame level by considering the sequence of frames in the videos. One simple option would be to average the outputs of our previous system considering a buffer of several previous instants. Can you propose a more advanced method to consider the videos using techniques studied during the course?