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

AUDIO PROCESSING, VIDEO PROCESSING AND COMPUTER VISION
ORDINARY FINAL EXAM (20/12/2021)

Student: …….. Grade:

EXERCISE 1 (1 pt)

A camera with a focal length of $f = 50\ mm$ is used to take a photo pointing to a vertical column that is $12\ m$ high and is $95\ m$ away from the camera in the direction of the camera axis. For simplicity, consider that the camera axis is also parallel to the ground.

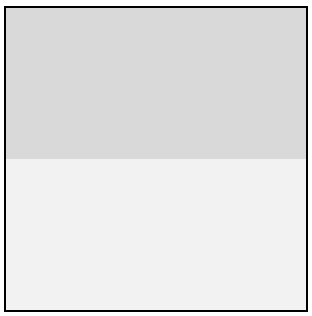
a) Determine the height of the column in the image.
b) Determine the number of pixels corresponding to that height assuming that the camera sensor has a resolution of 4000 dots per inch (1 inch = 2.54 cm).

EXERCISE 2 (1 pt)

Draw approximately the pixelwise transformation implementing the histogram equalization of this image (original image on the left; resulting image on the right).

EXERCISE 3 (1 pt)

The 128x128 image in the figure has 4-bit intensity resolution, with gray levels in the range [0, 15], and the two gray levels actually appearing in the image are 8 and 12.



Assume that we filter the image with a low-pass 3x3 filter mask (all coefficients equal to $1/9$), padding the image beyond its boundaries when necessary by repeating the last pixel. Draw the histograms before and after the filtering. Label the histograms with as much detail as possible.

EXERCISE 4 (1 pt)

The following formula represents the process of extracting a time-domain parameter from a speech or audio signal:

$$Q[n] = \sum_{m=-\infty}^{\infty} T\{s[m]\}\, w[n-m]$$

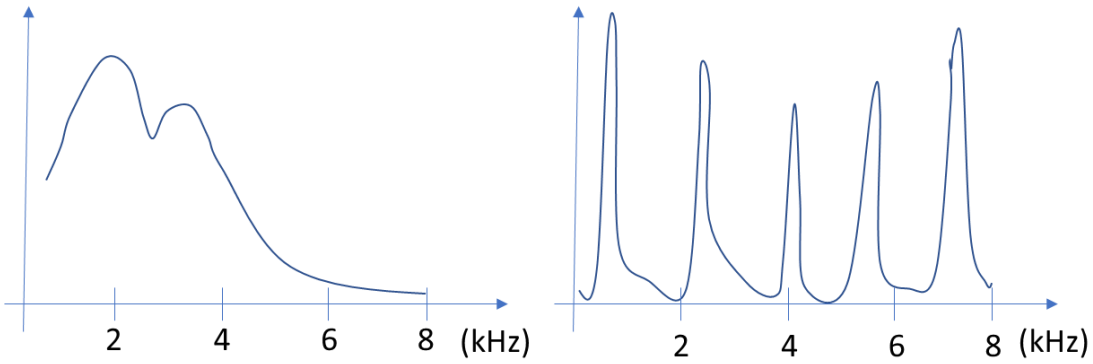
where $s[n]$ and $w[n]$ are the signal and analysis window, respectively, and T denotes a transformation.

a) Illustrate the process of computing $Q[n]$ by drawing the signal, the transformed signal and the window.

b) Why can we interpret this process as a low-pass filtering of the transformed signal?

EXERCISE 5 (1 pt)

The figure shows the average spectrum of two different classes of sounds.



Propose two features that allows us to discriminate between them.

EXERCISE 6 (5 pts)

We aim to design a system to automatically predict the price of a house based on a set of outdoor and indoor images. Next, we provide a description of the main aspects of our problem:

- To develop our system, we have a training dataset that contains 5000 houses distributed along the capital of the Spanish provinces. Data are distributed following the same proportion of city populations (e.g. Madrid is the city with more images in the training dataset, followed by Barcelona and Valencia).
- For each house, we have 5-10 images including: building façade or exteriors, living room, kitchen, bedrooms, bathroom, terrace / garden. Each photo is in turn labeled with the room it depicts.
- Images may have different sizes.
- Each case is labeled with the market price, the city where it is located, and a textual description with information such as: the number of rooms, square meters, date of building, facilities, services, etc.

In order to illustrate the problem, we next include an example of two houses with available images

Façade House 1

Kitchen House 1

Room House 1

Room House 1

Room House 1

Bathroom House 1

Façade House 2

Kitchen House 2

Room House 2

Room House 2

Room House 2

Bathroom House 2

Answer the following questions regarding the system design:

a) (1 pt) Choose the most appropriate architecture to implement your solution among the following options, detailing the reason for your choice:
   a. Alexnet CNN
   b. Resnet-50 CNN
   c. Faster RCNN with a Resnet-50 backbone
   d. Encoder-Decoder architecture with a fractional stride decoder.
   e. A modified Resnet-50, in which the stride in the last layers is removed (they are set to stride=1) and incorporate atrous convolutions.
   f. Convolutional GAN

b) (1 pt) Discuss the usefulness of the following data augmentation techniques in your problem: random cropping, rescaling + random cropping, random rotation, horizontal mirroring, vertical mirroring, Fancy PCA for color augmentation.

c) (0.5 pts) Chose the loss function that is more appropriate for your problem: image level cross-entropy, pixel-level cross-entropy, MSE loss.

d) (1 pt) Compare the following two alternatives discussing their advantages and disadvantages:

    a. Training a general system that receives one image at the input and estimates the price. In test, for each house, the system produces as many estimations as images are available for the house, and the final price is computed averaging the different estimations.

    b. Training expert-estimation systems for each room-type (e.g. one for exteriors, one for rooms, one for kitchens, etc.). In test, use the system that correspond with each image type, and compute the final price averaging the estimations.

e) (1 pt) Based on the alternative (b) in the previous question, propose an improvement to integrate the fusion of the different available images.

f) (0.5 pts) Discuss the kind of deep learning techniques that you will use to enhance the system by analyzing the textual descriptions that come with the images.