Name: Adrian Flores Camacho
FINAL EXAM 1T
IMT-344 Computer Vision
I - 2021

1. INSTRUCTIONS

Similar to the midterm exam, this test also includes two types of questions: problems and inquiries. This time, the presentation of your results will be made in two ways: on the Neo platform and through a Zoom call, which you will have to schedule at https://calendly.com/edwinsalcedo/20min on a first-come-first-served basis. You will receive a Zoom link some minutes before your appointment. Although your work must be written completely in English, the presentation can be in Spanish.

2. GRADING

- 40% goes to your work (notebooks, scripts, and report) where you will be evaluated as follows:
- **Problem: 50%**
  - Correct or appropriate results.
- **Inquiries: 50%**
  - Answer justification and correct way of presenting your ideas.

- 60% goes to the presentation and questions on Monday 21st, where you will be graded as follows:
  - **Presentation (10 min): 50%** ● **Questions (10 min): 50%**
    - Knowledge, contribution, and correct way of explaining your ideas.

Again, It is highly recommended to read the whole exam first and then start to solve it.

3. PROBLEM

Specifications: If a given solution is proved to have undeclared similarities with a solution of another student or from the Internet, you will be assumed to have incurred plagiarism. This time, you are not allowed to work with anybody so you would lose marks if your answers are found in another test. This time, **you must avoid including comments and justifications in your code when submitting your work.**

Tumor Detection

In the fifth and sixth tutorial, you developed short algorithms to show if a patient with a benign tumour improved after a month of having a surgery. This time, you will have to analyse and implement a more complete strategy to help the Bolivian hospitals identify benign and malign tumours from MRI images. Specifically, you will have to apply Transfer Learning with Pytorch and evaluate which of the following pretrained models are the best for this task: resnet, alexnet, vgg,

squeezenet, densenet, and inception. This decision must be based on metrics and must be well described. Important, you must use the following dataset: https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection

To submit your answer to this problem, you must use an independent ipynb file and a python file capable of accepting an input image and showing a result. Also, a github repository where you commit frequently.

4. INQUIRIES

Specifications: This inquiries' list contains all the questions for the final exam. They are categorized according to two sections of what was covered during the entire course.

A) Image and Video Processing

For this section: choose 5 questions randomly, and answer them thoroughly with examples.

1. Why would you need image feature extractors, such as SIFT or SURF?
SIFT solves image rotation, transformation, intensity and point of view shift in matching features, it has 4 steps to follow which are:

- Estimate an extreme scale space with the Gaussian difference
- Locates key points whereby key point candidates are refined by eliminating low contrast points.
- Performs keypoint orientation based on image gradient and generates a descriptor for each keypoint based on image gradient magnitude and orientation.

Accordingly, the SIFT functionality is a chosen image area and is an algorithm used to identify and explain local properties in a digital image. It locates certain key points of view and then gives them quantitative information that they can use for object recognition. Assuming that the descriptors are invariant under some transformations, even if they represent the same object, they have the possibility of making the image look different.

The SURF method comes from the English Speeded-Up Robust Features, which is an algorithm that allows the extraction, description and comparison of image characteristics.

- Compared to SIFT, this algorithm is faster and more robust for conversions such as zoom, direction, rotation, blur, lighting changes, and noise. Which this algorithm consists of three parts
- The detector finds the characteristic points of interest in an image using the Gaussian difference approximation (DoG) and the Hessian matrix.
- The descriptor constructs a vector of 64 characteristics that describe the position and orientation of the points of interest using the sums of the responses.
- The comparison identifies the same points in two images, comparing the Euclidean distance between the descriptor vectors of the points of interest.

Bibliography:

Websites:

https://41jaiio.sadio.org.ar/sites/default/files/6_EST_2012.pdf

https://repositorio.unibague.edu.co/jspui/bitstream/20.500.12313/841/1/Trabajo%20de%20grado.%20pdf

https://arxiv.org/ftp/arxiv/papers/1710/1710.02726.pdf

https://www.i2tutorials.com/what-are-sift-and-surf/

https://medium.com/@shehan.a.perera/a-comparison-of-sift-surf-and-orb-333d64bcaaea

https://ieeexplore.ieee.org/document/7975187

2. What are common areas for video analysis? That's, what sorts of needs commonly occur in industry that leads to video analysis?

Video analysis is a modern and developed mechanism that can verify and analyze the images captured by surveillance cameras. Its main function is to make the program that detects intruders more accurate. It contains a series of specifications and patterns, and the system checks if the moving object conforms to any of them. If they do, it will immediately issue an intrusion alert. It provides the intelligence to assess what is happening in real time and provides alerts to immediately activate related protocols. The system can divide the video into three categories:

- Violent or criminal activity
- Potentially suspicious
- Sure

It is an architecture based on convolutional and recurrent neural networks, which is described in the following points.

 Can be scaled to thousands of cameras without overwhelming operators with trivial alerts.

- Self-learns to recognize unexpected behavior in any camera's field of view without the need for custom programming or human intervention.
- Interpret video from H.264 and MPEG4 sources in both distributed and centralized environments.
- Reduces the time and expense associated with implementing and operating intelligent video surveillance through an open systems approach.
- Provides real-time insights based on data as events unfold, rather than forensic information applied after the fact.
- Provides direct visibility into industrial safety issues, as well as physical protection, such as automatic smoke, fire and flood recognition.
- Improvement of situational awareness in organizations that seek to know the possible threats in real time.

Bibliography:

Websites:

https://business.adobe.com/es/products/analytics/video-analytics.html

http://www.grupoalava.com/seguridad/productos/analisis-de-video/analisis-de-video-inteligente-basado-en-redes-neuronales/analisis-de-video-inteligente-vs-analisis-de-video-basado-en-reglas-o-patrones/

https://seguridadrgr.com/video-analisis/

https://www.imsel.com/sistemas-de-videoanalisis/

https://grupocartronic.com/deep-learning-hanwha/

https://everislatam.medium.com/video-analysis-to-detect-suspicious-activity-based-on-deep-learning-2218e754cdae

3. Describe the challenges and benefits of processing video in embedded devices, as opposed to transmitting video to a cloud server?

4. What is periodic noise and in what sorts of contexts does it occur?

A common source of periodic noise in images is electrical or electromechanical interference during image capture. The image affected by periodic noise appears as if a repeating pattern has been added on top of the original image. By adding a sine wave of the original frequency, periodic noise can be artificially generated to produce a diagonal pattern in the image.

Bibliography:

Websites:

https://en.wikipedia.org/wiki/Image_noise

https://dergipark.org.tr/tr/download/article-file/433724

Removal of Periodic and Salt & Pepper Noise from an Image.pdf

5. If you were able to apply any technique learned in the course, how would you reduce periodic noise optimally?

Periodic sound can be significantly minimized by filtering in the frequency domain - this kind of sound can be seen as discrete peaks. A significant reduction in this sound can be achieved by implementing notch filters in the frequency domain.

The notch filter rejects frequencies in predefined neighborhoods near a center frequency. The number of notch filters is arbitrary. The manner of the notch areas could furthermore be arbitrary such as rectangular or circular.

Periodic noise reduction depends if fine image details are not considered important, low-pass filtering might be a suitable option.

Bibliography:

Websites:

https://docs.opencv.org/3.4/d2/d0b/tutorial_periodic_noise_removing_filter.html

https://academic.mu.edu/phys/matthysd/web226/Lab03.htm

https://publicaciones.unirioja.es/catalogo/online/VisionArtificial.pdf

6. Imagine that the government is about to share with you a dataset of satellite images collected during the last forest fires in Santa Cruz, nevertheless, the dataset is massive and you are in another country. What sort of compression method or strategy would you use?

7. Given the image on the right, what sort of operations/transformations would you use to get a binary mask with the veins of the leaf?

An adjustment would be made to the brightness and contrast through the addWeighted () method to vary the contrast and brightness through the alpha and beta variables after applying this change, bring the image to grayscale, make a mask that best suits possible to the veins of the leaf, then noise will be observed in the image, which can be applied a medium filter or ersion and expansion filter to eliminate granulated noise, which would have to be applied with the Canny with adjustments in the parameters to achieve leaf veins.

Bibliography:
Websites:
https://qastack.mx/signals/1714/best-way-of-segmenting-veins-in-leaves

B) Introduction to Object Detection

For this section: choose 5 questions randomly, and answer them thoroughly with examples.

1. Explain how a Convolutional Neural Network mimics our visual system.
Understanding how the visual cortex works, especially the cells that cause orientation selectivity and edge detection in visual stimuli in the primary visual cortex V1. These are called simple cells and complex cells.

Simple cells have excitatory and inhibitory zones, both forming necessary elongated patterns in a direction, posture and size especially in each cell. If a visual stimulus reaches the cell with the same orientation and posture, in such a way that it aligns perfectly with the patterns created by the excitatory zones and simultaneously it prevents activating the inhibitory zones, the cell is activated and emits a signal.

Complex cells operate in a similar way. Like simple cells, they have a special orientation to which they are susceptible. However, they do not have posture sensitivity. For this reason, a visual stimulus requires arriving only in the ideal orientation for this cell to be activated.

The visual cortex its structure is the V1 region of the visual cortex, then following the V2, V4 and IT regions, the ideal stimuli increase more and more.

The Neural network consists of several layers of convolutional filters of one or more dimensions. Like any network used for categorization, at the beginning these networks have a characteristic subtraction stage, composed of convolutional neurons, then there is a reduction by sampling and finally we will have simpler perceptron neurons to make the final categorization on the extracted characteristics.

The characteristic subtraction stage resembles the stimulating process in the cells of the visual cortex. As the data progresses during this stage, its dimensionality is reduced, the neurons in distant layers being much less susceptible to disturbances in the input data, however, in parallel, these are activated by increasingly complicated characteristics.

Bibliography:

Websites:

https://es.wikipedia.org/wiki/Red_neuronal_convolucional#:~:text=Una%20red%20neuronal%20convolucional%20es,V1)%20de%20un%20cerebro%20biol%C3%B3gico.

https://www.aprendemachinelearning.com/como-funcionan-las-convolutional-neural-networks-vision-por-ordenador/

http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2225-87872016000100005

https://www.juanbarrios.com/redes-neurales-convolucionales/

2. Given a ResNext deployed in a Raspberry Pi, you are called because the video processing is taking longer and sometimes the device crashes, however, both the device and model are important for the business . What would be your conclusions? Why is the performance suffering?

3. Explain what would you do if a model based on mobilenet overfits after transfer learning?

How would you improve the model's ability to generalize?

4. Explain what kind of metrics you would use for object detection.

The accuracy of the machine learning model is the size used to establish which model is best to detect interrelationships and patterns between changes in a data group in functionality of the input or training data. The better a model is able to generalize to "invisible" data, the better predictions and insights it can produce. Which this measure is measured in percentage which your result to be closer to 100% better will be your prediction of the objects

Loss

The lower the loss, the better the model is going to be. The loss is calculated based on training and validation and its interpretation is how well the model is working for those 2 sets. Unlike accuracy, the loss is not a percentage. It is a sum of the errors made for each example in the training or validation sets. Therefore, its measurement is measured by values close to zero as long as the result is the closest to zero, less loss was obtained in the object detection model.

Bibliography:

Websites:

https://stackoverflow.com/questions/34518656/how-to-interpret-loss-and-accuracy-for-a-machine-learning-model

https://www.datarobot.com/wiki/accuracy/#:~:text=Machine%20learning%20model%20accuracy%20is,input%2C%20or%20training%2C%20data.
https://developers.google.com/machine-learning/crash-course/descending-into-ml/training-and-loss#:~:text=Descending%20into%20ML%3A%20Training%20and%20Loss%20bookmark_border&text=Loss%20is%20the%20penalty%20for,was%20on%20a%20single%20example.&text=The%20goal%20of%20training%20a,on%20average%2C%20across%20all%20examples.

5. Given a supervised learning task, what would be your strategy to segment a dataset that contains 15000 unlabelled images images and 6150 labelled images. What would be your strategy to take advantage of the majority of images?

Make a model with the 6150 tag images in one of the supervised Machine Learning techniques, which will help us through a script that helps us predict the classes of the 15000 so that we will be able to obtain a categorization of the images in each class, there could be an error because the model does not have a large database, which would have lost and confused data in the classes, however, most of the images would be used

Another alternative that is possible is to use unsupervised Machine Learning procedures unlike supervised learning, in unsupervised only the properties are given, without giving the algorithm or a label. Its functionality is grouping, so the algorithm must classify by similarity and be able to generate sets, without having the function of conceptualizing how each individuality of all the members of the set is what we have these 2 methods:

Clustering: classify the output data into groups.

Association - Discover rules within the dataset.

Bibliography:

Websites:

https://blog.bismart.com/es/diferencias-machine-learning-supervisado-no-supervisado
https://medium.com/@juanzambrano/aprendizaje-supervisado-o-no-supervisado-39ccf1fd6e7b

6. What are the differences and examples of discriminative and generative models?

Discriminatory algorithms model the dependence of an output variable Y as a function of an input variable X, which allows us to predict the variable Y from X, we see if our model is probabilistic or not

the logistic regression where the probability of an event is modeled as a function of the independent variables that learn the conditional probability distribution P (Y / X)

Non-probabilistic algorithms where a mapping or direct function is learned from the input variables to the target classes or variables.

Generative algorithms provide a model for how data is generated. They learn the joint probability distribution P (X, Y). An observation is made of how the model was generated.

If the variable Y indicates classes of an object, then the distribution P (X / Y) models how the characteristics that are the input data of the object are distributed, which models the distribution P (X / Y) and P (Y)

Generative algorithms address a more general problem and actually learn how input data is structured and distributed in what discriminatives simply learn to categorize a target variable into input functionality.

Since the generative algorithm makes more assumptions about the composition of the model, it is common that once you have little training data, the generative algorithm behaves better, however, as the group of training data grows, this behavior is reversed.

Bibliography:

Websites:

https://mlearninglab.com/2017/08/23/algoritmos-discriminativos-y-generativos-de-aprendizaje/
https://qastack.mx/programming/879432/what-is-the-difference-between-a-generative-and-a-discriminative-algorithm
https://www.it-swarm-es.com/es/algorithm/cual-es-la-diferencia-entre-un-algoritmo-generativo-y-uno-discriminativo/957513124/

7. Consider the image on the right, where an edge detector has been applied, what would be your method to find the area of the buildings and land.

It would be done using OpenCV with the cv2.cornerHarris () function, which we will use what we learned in tutorial 10 of this function, which allows us to obtain the corners of the found edges, which allows us to know exactly the coordinates of each pixel of the corners of each building and land, which by applying the distance equation between two points, the sides of each area are obtained, which would apply the equation to obtain the area, which depending on the distance from which the image was taken, a scale of the area found would have to be made with true value.

Another method that I managed to investigate is with functions that OpenCV which they use firstly apply the canny to the grayscale image then they use the dilation and the erode in the image then they apply the findContours function which we pass to the contourArea function which would obtain the area of the object in the following script is an example of how they get the area of a box.

```
import cv2 as cv
import numpy as np
import matplotlib.pyplot as plt

image = cv.imread("caja2.jpg")
image_copy = np.array(image)
gray = cv.cvtColor(image, cv.COLOR_BGR2GRAY)

canny = cv.Canny(gray, 10,150)
canny = cv.dilate(canny, None, iterations=1)
canny = cv.erode(canny, None, iterations=1)
cnts,_ = cv.findContours(canny, cv.RETR_EXTERNAL, cv.CHAIN_APPROX_SIMPLE)


#Buscas el contorno más grande
lista_areas = []
for c in cnts:
    area = cv.contourArea(c)
    lista_areas.append(area)


#Te quedas con el area más grande
mas_grande = cnts[lista_areas.index(max(lista_areas))]

#Representas el contorno más grande
area = cv.contourArea(mas_grande)
x,y,w,h = cv.boundingRect(mas_grande)
cv.rectangle(image_copy, (x,y), (x+w, y+h), (0,255,0), 2)
cv.imshow('Calculo de Area y Reconocimiento de imagen', image_copy)
cv.waitKey(0)
```

Bibliography:

Websites:

https://es.stackoverflow.com/questions/419194/c%C3%A1lculo-de-area-con-opencv

https://programmerclick.com/article/11701709304/

https://www.it-swarm-es.com/es/python/obtener-area-dentro-de-los-contornos-de-opencv-python/1050492001/

https://docs.opencv.org/3.4/dd/d49/tutorial_py_contour_features.html

5. SUBMISSION

You will have to submit your solution in a compressed file (.zip) via the Neo platform until Monday 21st at 11:00 a.m.