Gesture Symphony: Playing music using hand gestures

Wednesday 17th May, 2023 - 11:50

Hugo BARTHELEMY
University of Luxembourg
Email: hugo.barthelemy.001@student.uni.lu

This report has been produced under the supervision of:
Kayhan LATIFZADEH
University of Luxembourg
Email: kayhan.latifzadeh@uni.lu

Abstract

Artificial Intelligence could revolutionize the way people are currently interacting with machines. This paper is the result of the work of a first year student pursuing a Bachelor of Computer Science at the University of Luxembourg. It is divided into two distinct parts, the scientific deliverable which answer to "How can you play simplified musical instruments using hand gestures recognition?" and a technical deliverable which presents the development of a demo application.

1. Introduction

The skyrocket expansion of Artificial Intelligence (AI) is currently happening right before our eyes. With the development of the extremely effective chat bot ChatGPT, the awesome image generator Midjourney and the tons of AI tools that are being developed while you are reading this paper, people can't deny anymore that Artificial Intelligence is a fascinating topic. Which is why this paper can be seen as a first step inside the AI world. It will provide solid foundations on specific topic related to Artificial Intelligence as well as, includes materials about hand gestures recognition and app development. The knowledge that this paper is willing to convey is supported by the development of a demo application called Gesture Symphony, in which you can play musical instrument by using hand gestures.

2. Project description

2.1. Domains

This section introduces the different scientific and technical domains covered in this project.

2.1.1. Scientific. The scientific work of this project refers to concepts from human-computer interaction, artificial intelligence and data science be it: user input, usability, deep learning as well as notions related to data-set such as data acquisition, data cleaning and data labelling.

2.1.2. Technical. The technical deliverable domains are at the intersection of human-computer interaction and artificial intelligence. It includes the development of an application, a deep learning model and the implementation of functionalities using hand gestures recognition.

2.2. Targeted Deliverables

This section describes both deliverables.

2.2.1. Scientific deliverables. The scientific deliverable of this project focuses on answering to the following question: **How can you play simplified musical instruments using hand gestures recognition?**
The answer consists of 4 main parts, an introduction to the subject, then two main axes, and finally the intersection of those two axes.

- Introduction to the subject: Definitions related to hand gestures
- First axe: Artificial Intelligence
- Second axe: Human-Computer Interaction
- The intersection of AI and HCI

Introduction: It presents the most important concepts related to hand gestures and hand gestures recognition (e.g. static gestures, hand segmentation).
First axe, Artificial Intelligence: In the first axe, deep learning, convolutional neural network (CNN) and computer vision are emphasized to understand how can we use those techniques [5] to achieve hand gestures recognition. Each concepts presented in this part is defined beforehand using mostly mathematical expressions (e.g. activation and loss functions)[4], natural languages and figures (e.g. neural networks layers).
Second axe, Human-Computer Interaction: The second axe aims to add Human-Computer Interaction notions such as usability [7] and user experience [6]. Defining those notions helps to understand how to enhance the user experience which

is a step forward in order to simplify the way the user could play a musical instrument.
The intersection of AI and HCI: This part uses concept that are defined on both axes to demonstrate how to simplify interactions between the human and the machine.

2.2.2. Technical deliverables. The technical deliverable of this project is an application developed in Python that allows its users to play simplified musical instrument using hand gestures. This application aims to attract a range of users that don't know how to play any musical instrument, but sill want to play music in an easy way. The development of this software is composed of 3 phases.

- Development of the hand detection tool
- Development of the deep learning model
- Development of the final application

Development of the hand detection tool: The first phase aims to develop a python programs that detect the hand of the user using the built-in camera of the computer. This programs should track hand gestures and segment the hand in real time [1] [2]. To develop this program the libraries OpenCV and Mediapipe are used.
Development of a deep learning model: The second phase is dedicated to the development of a deep learning model that recognizes static hand gestures. This model should be composed of a convolutional neural network [4] that takes as input a set of images of size 28 by 28 pixels. Those images represent static hand gestures. In order to build the convolutional neural network and handle the data the libraries Pandas, Numpy and Keras [3] are used.
Development of an application: The final phase is oriented towards the development of the final application that allows a user to play simplified musical instruments using hand gestures. The application should uses the deep learning model in order to associates a static hand gesture to a musical note. After this step is done, the application should uses the hand detection tool to detect when the user makes a recognized hand gestures so that it play the musical note associated to the hand gesture.

3. Pre-requisites

This section denotes both the scientific and technical pre-requisites that are required to understand this paper and its related project.

3.1. Scientific pre-requisites

From a scientific point of view, it is required to have a certain foundation in mathematics, notably in linear algebra and analysis. For linear algebra, it is enough to understand how matrix calculations work, and in analysis you need to know how to use functions and derivatives.

3.2. Technical pre-requisites

From a technical point of view, it is required to have a strong knowledge in the programming language Python (e.g. Pandas, Numpy) and inobject oriented programming.

4. How can you play simplified musical instruments using hand gestures recognition?

4.1. Requirements

This section defines what are the requirements regarding the scientific deliverable.

4.1.1. General Description. The scientific work of this project must answer to the following scientific question: **How can you play simplified musical instrument using hand gestures recognition ?** it should explain from beginning to end, how can we use hand gestures recognition as a new medium of interaction, here using the context of music. Therefore, this paper should gives well structured definitions of different concepts related to the question (e.g. convolutional neural network, dynamic gestures) and use those concepts afterwards to give a proper answer of good scientific quality.

4.1.2. First axe, Artificial Intelligence. The first axe of this paper must explain the backstage of hand gestures recognition using convolutional neural network.

- Hand gesture vs Hand posture
- Process of hand recognition using fingers and palm segmentation

In the final stage, the concept of convolutional neural network is to be defined so this stage will give information on:

- Convolutional Layers
- Pooling Layers
- Flatten Layer
- Dense Layers

4.1.3. Second axe, Human-Computer Interaction. The second axe must shows that the interaction between human and computer is currently limited by the use of common input devices (e.g. mouse and keyboard). Hence, the following definition are required:

- Usability
- User Experience

Usability and User Experience are two notions that belongs to the field of HCI, their definitions will helps to estimate weither or not hand gestures recognition technology as input device can produce application that are both usable and pleasant to use.

4.1.4. The intersection of AI and HCI. This final part must use the content of both axes to create a link between two field of Computer Science, HCI and AI. The requirement of this part is to use the intersection AI and HCI to answer to the

scientific question and so, to demonstrate that the interaction between human and machine can be bring to a new stage using the demo application **Gesture Symphony** as an example.

4.2. Design

4.3. Production

4.3.1. Hand gestures vs Hand postures. Hand gestures and hand postures are two concepts of computer vision and their are mainly used as a new way of interaction between a user and a machine. However, those concepts are distinct and that is why we define them.

Definition of hand gestures: Hand gestures involves movements of the hand that often has a meaning. When hand gestures recognition happens the computer is supposed to recognize multiple frames of a video that capture a movement from the hand, which is called a hand gesture. It could be clapping hands or waving the right hand.

Definition of hand postures: Hand postures as opposed to hand gestures are characterize by the lack of motion. When hand postures recognition happens, the computer recognize a single frame, which is basically a picture of a hand posture. It could be the posture thumps up or thumps down.

So with those two definitions, the main difference that exists between hand gestures and hand postures is the presence of motion. However, in a project like Gesture Symphony, both hand gestures and hand postures are interesting to use in order to detect, for example hand postures between the multiple frame of hand gestures.

4.3.2. Process of hand recognition using fingers and palm segmentation. Hand recognition is a computer vision concept that entails the detection and the analyze of the features of a human hand in an image or video. In order, to recognize the hand and extract information from it, the first thing is to detect the hand in an image.

Detection of the hand: The detection of the hand in an image or video can be done using multiple computer vision techniques such as:

- **HSV color space**
- **Machine learning** (see subsection 4.3.3)
- **Background subtraction**

HSV color space: The HSV color space helps to divide a pixel color into three different part, the hue which contains the color information, the saturation denotes the saturation and the value which characterize the brightness. By dividing the pixel information, this color space makes it easier to differentiate specific color range, especially the hand skin color which has a specific color range. However, the main drawback about the HSV color space is that the brightness value may vary a lot based on lighting condition and skin color vary too between each individual.

Background subtraction: The background subtraction method detect the hand in an image by applying a simple process. It starts by taking the initial image without the hand and store it, then it subtracts the initial image with the stored image with no hand on it. By doing that the background of the image will be removed and the only pixels that will still have a non zero value are the hand pixel. This method is not affected by the lighting exposures or the hand skin color variation and so it is more efficient.

Once the hand is detected, the segmentation of the palm and fingers can be performed.

Palm segmentation The palm of the hand is segmented into different areas and points. The **palm point** which is the center of the palm, **the palm mask** which is the segmented hand palm, **the wrist points** which are the two end points of the wrist and **the wrist line** which is the line that go through the wrist points.[**2**]

The palm point is obtained by using the technique of distance transform.

Distance Transform Method: The distance transform of an image is another image where each pixel value is now the equal to the distance of itself and the nearest boundary pixel, where a boundary pixel is a pixel which is at the edge of the detected hand.[2]

0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	0
0	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0

(a)

0	0	0	0	0	0	0	0	0
0	1	1	1	1	1	1	1	0
0	1	2	2	2	2	2	1	0
0	1	2	3	3	3	2	1	0
0	1	2	3	4	3	2	1	0
0	1	2	3	3	3	2	1	0
0	1	2	2	2	2	2	1	0
0	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0

(b)

Fig. 1. Distance Transform Example with Binary Matrix
a) Initial matrix b) Matrix after the distance
0: Background pixel, 1: Hand pixel
Source: [2]

Using this method, the palm point is easily defined since it is supposed to be at the center of the detected hand, it is the pixel with the greatest value (e.g. the pixel with value 4 in Figure 1 b).

Palm Mask Algorithm: The palm mask algorithm uses several steps to get the palm mask using the x and y coordinates of the palm point. Therefore, this algorithm can be explained as follows:

Step 1: Acquire a pixel(x,y) near the palm point(X, Y)

$$x = cos(\frac{angle*\pi}{180}) * radius + X$$

$$y = sin(\frac{angle*\pi}{180}) * radius + Y$$

where *angle* is between 0 and 180 and *radius* is between 1 and the image size.
Step 2: Check the value of the pixel(x,y), if it is 0 then go to step 3, else go to step 4
Step 3: Check if the following property holds:

$$\forall P(x + dx, y + dy) == 0, P(x + dx, y + dy) \in N$$

where N is the set of the 8 neighbors pixels of the pixel(x,y).
If it holds, insert $P(x + dx, y + dy)$ in the palm mask array, else go to step 4
Step 4: Increase the radius and angle by 1 and go to step 1

Fig. 2. Palm Mask Algorithm
Source: Made by the student, information coming from[2]

To get the palm mask use this algorithm up until all the pixel are scanned and connect all the points that are in the palm mask array.

Wrist Points & Wrist line The wrist points are obtain by checking the distance between two successive points from the palm mask, if the distance is large then those two points are supposed to be the wrist points. The formula used to check the wrist points distance is the following:

$$\arg\max \text{ dist } (P_i, P_{i+1}),\ P_i, P_{i+1} \in S$$

where S is the set of the point in the palm mask circle and dist(_ , _) is the distance between two points. [2]

Afterwards, the wrist line is obtained by linking the two wrist points.

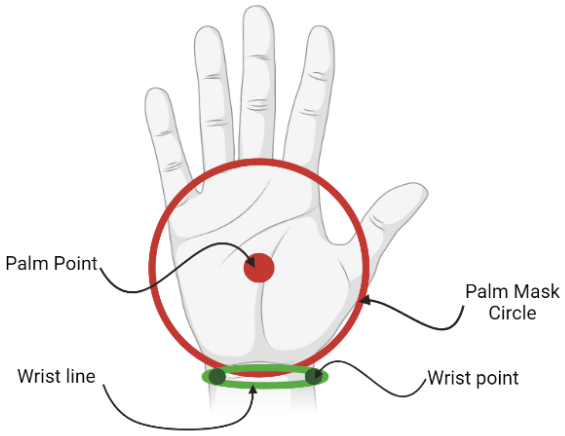

Fig. 3. Palm Segmentation
Source: Made by the student, information coming from [2]

Fingers Segmentation: While palm segmentation involves a lot of mathematics, finger segmentation is much simpler. Here is a simple algorithm to segment fingers:

Step 1: Perform palm segmentation
Step 2: Remove the palm mask from the hand pixel
Step 3: Store the remaining hand pixels in a list
This list denotes the segmented fingers pixels.

Fig. 4. Algorithm for Fingers Segmentation
Source: Made by the student

4.3.3. Convolutional Neural Network. A convolutional neural network(CNN) is a specific type of neural network that performs well in solving image processing problems. CNN have the particularity of including convolutional layer(CN) and pooling layer(PL) inside their model which make them more efficient at detecting patterns or features such as circles, corners or edges in an image. However, CNN do not only have CL and PL they also include fully connected layers called Dense Layer(DL).

4.3.4. The Different Layers of a CNN. Convolutional layers are locally connected layers, that means that each of their neurons is not connected to all neurons from the previous layers but only a subset of them. This fact helps CL to extract patterns and features from an image while using a reduced amount of parameters due to the reduced amount of connected neurons. The operations performed inside convolutional layers are mathematical convolution operations over tensors(multidimensional matrices). They involves moving a tensor, called a kernel, over an image. The result of this operation is a new tensor, called a patterns map or features map, where each value is the dot product of the kernel and the pixel of the image covered by the kernel. The size of the new tensor takes many parameters in account: The image size, the kernel size, the stride and the padding. The stride is the distance that is covered by the kernel when moving and the padding is the number of 0s that should be added to the image to preserve its size.

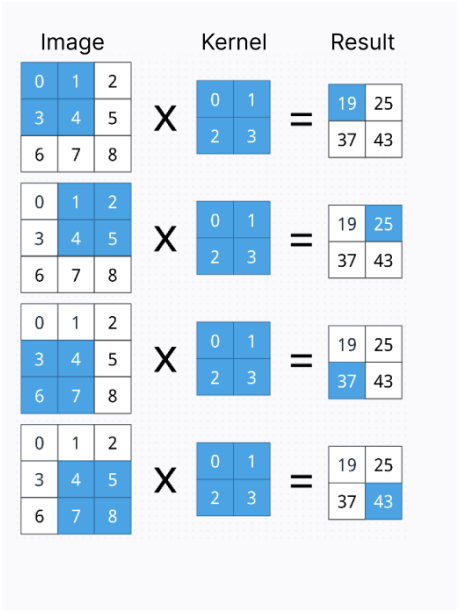

Fig. 5. Convolution Operation
Source: Made by the student

This figure shows a convolution operation with an image of size 3x3, a kernel of size 2x2, a stride of 0 and a padding of 0.
The purpose of those layers is to detect patterns or features from the input images, each kernel is used to extract a specific patterns be it edges, circles or corners.
Moreover, CL are usually followed by a PL.

Pooling Layers:
Pooling layers performs the exact inverse operation. They take as input a patterns map coming from a convolutional layer output, they divide it into non-overlapping tensors and compute a summary statistic such as the average, the minimum or the maximum. By doing this, pooling layers are used to reduce the amount of parameters in the CNN since they simplify the patterns map. It also helps to prevent the network from over fitting, which happens when the CNN is too used to the same data. Furthermore, it introduces the network to translation invariance, which means that it will be able to recognize patterns or features even if they don't appears in the exact same location. Finally, they can increase slightly the robustness of the network by reducing the impact of some parameters in the image like lighting exposure. Even if it is called a convolutional layers, it also contains normal layers which are called dense layers. However, dense layers cannot take as input a features map from CL that is why CNN needs a specific layer called the flatten layer.

Flatten Layer:
The flatten layer is usually unique in a CNN model and it is used to make the transition between CL and DL. CL works with tensors but DL works with single dimensional vectors. That is why, to transit from multidimensional to single dimensional the output must me flatten. Once, it is flatten it can be feed directly to the DL.

Dense Layer:
Dl are fully connected layers meaning that each of their neurons are connected to all the neurons of the previous layer. This fact means that those layers take much more parameters but are more efficient for classification tasks, which is why those layers are often placed at the end of a CNN to classify the flattened output of the CL into a label.

4.3.5. Hand gestures recognition as an input device.

Hand gestures recognition is an interesting input device that could revoke the usual way of interaction between people and machine. From the start, hand gestures interaction has an assets that makes it a really good input device, it is the fact that hand gestures is a natural way of interaction between humans. Therefore, implementing an already existing way of interaction is a lot more logical and natural than using a mouse and a keyboard.
Moreover, this new kind of input device could remove some boundaries in our way of interaction with machine regarding certain applications. The different applications of hand gestures recognition can be categorised into three main fields:

- Health
- Human Machine Interaction
- Entertainment

Health:
Hand gesture recognition can be an application in the health domain in a number of ways.

- **Performing hands free operation**
  Allowing medics to perform hands free operation is a huge assets in the medical field since it helps them to interact with medical robot and controlling surgery tools with more ease and accuracy.
- **Improving the communication between patients and medical staff**
  This application allows patients with temporary or permanents disabilities, such as motor or voice disabilities, to communicate more easily with the medical staff in case of emergency.
- **Developing medical tools**
  Improving the medical fields by developing more efficient treatment or tools is always a high priority research's topic and hand gestures recognition could bring something in this field. New medical robot or surgery tools controllable with hand gestures that would be commanded by a surgeon or remote monitoring devices that recognize hand gestures to gives a more deeper analysis of hand injuries.

Human Computer Interaction:
In the domain of HCI, such a technology is definitely something you want to use. HCI looks for the best way of interaction between humans and machines. Therefore, hand gestures recognition has many applications in this field such as:

- **Redefine interface controller**
  This new input device can be used to redefine the current

way interfaces are being developed. Developing new interfaces that are controllable by using hand gestures is one of the current goals of HCI. It can be done by using gloves based techniques or vision based techniques.

- **Improve efficiency**
  Hand gestures recognition, if used with a efficient interface controller system, can improve the efficiency of certain tasks by achieving them with less times and ease of use.
- **Improve accessability**
  People with disabilities that prevent them from using the usual keyboard and mouse can be helps by hand gestures recognition technology.

Entertainment:
Entertainment is also one of the field where hand gestures recognition technology as numerous possible applications. Here is some of the possible application in the entertainment's field:

- **Video games**
  Video games is the most known application of hand gestures recognition since it is presents in two new fields inside the video games development industry which are Augmented Reality and Virtual Reality. In both of those fields, hand gestures recognition has a great impact.
- **TV Controller**
  Hand gestures recognition can also have applications such as controlling the television, adjusting the sound volume or pause the video.

However, hand gestures recognition technology requires high performance to be efficient and usable. Therefore, the definitions of usability and user experience are going to be given in order to estimate if hand gestures recognition could really be a better input device than usual ones.

4.3.6. Definition of usability. Usability is a notion in HCI that is use to assess, from an objective point of view, the ease of use of an application. Therefore, by passing different usability test, an application can be qualified either as highly usable or poorly usable. Moreover, usability can be evaluated by using five different components [7] which are:

- **Learnability**
  Learnability evaluate how easy it is for a new user to learn how to use the application.
- **Efficiency**
  Efficiency defines how fast a user can complete tasks on the application.
- **Memorability**
  Memorability assesses how easy it is for a user remember how to use the application after a long period of time where he didn't use it.
- **Errors**
  Errors computes the number of errors done by the user when using the different functionality of the application.
- **Satisfaction**
  Satisfaction evaluate the subjective satisfaction of a user regarding the application.

Each of these components can be evaluated using different metrics and based on the result we can objectively estimate if an application is qualified as highly usable or poorly usable.

4.3.7. Definition of user experience. User experience(UX) is a concept that is part of Human-Computer Interaction. This concept denotes the experience and satisfaction of a user when using a specific application. However as each user is different, this concept includes subjectivity and needs to perform what is called user research. User research is the action of gathering user feedback and then develop the application based on the user insight. To estimate the quality of the UX it is possible to use four different aspects[6]:

- **Non Instrumental Qualities**
  Non instrumental qualities include the design of the user interface and the aesthetics aspects of the application such as the beauty of the visuals or the quality of the acoustic.
- **Instrumental Qualities**
  Instrumental qualities, as opposed to non instrumental qualities, are technical qualities which can be associated with usability and the usefulness of the application.
- **Emotional User Reaction** Emotional user reaction is another aspect that is important in user experience. It underlines the subjective feeling of the user, his motor expression, especially the facial expression that a user makes when using the application. It includes also, physiological reaction such as change in the heart beating rate or the eye pupil dilatation.

4.3.8. Conclusion: Hand gestures recognition as a medium of interaction. This paper shows that hand gestures recognition technology is an amazing tools to be used to improve the interaction between humans and machines and make it more natural. It has a numerous way of implementation using multiple hand gestures recognition technique and it has also a numerous way of application in domains such as health, HCI and entertainment. Furthermore, the interaction between humans and machines can be assessed using different criteria such as usability and user experience. However, for hand gestures technology to be good enough it requires to highly efficient and robust programs with a high performance hardware. The main challenges about hand gestures technology using camera based technique is the difficulty of building a robust and efficient systems.

4.4. Assessment

5. Gesture Symphony

5.1. Requirements

5.1.1. Description of the application. The technical deliverable of this project is based on the development of a demo application called **Gesture Symphony**. It is developed using the programming language python and the libraries Keras,

OpenCV and Mediapipe. Gesture Symphony allows its users to play simplified musical instrument using the user's hand gestures. It targets a large audience that wants to have fun playing music without having to learn a particular instrument. Therefore, in order to develop the application and make it work correctly, there are general requirements that must be met:

5.1.2. General requirements. This section denotes the general requirements of that are needed to develop **Gesture Symphony**.
Creation of a dataset: The student must build a large enought data set on its own to train the deep learning model of the application.
Creation of a deep learning model A deep learning model that uses convolutional neural network to recognizes hand gestures must be build by the student using the library Keras.
Implementation of the model inside the application The model should be saved beforehand and loaded inside the application so that it can recognize hand gestures and associate musical notes to each of them.
Creation of a user interface The last step is to build a good looking user interface for **Gesture Symphony**.

5.1.3. Functional requirements. This subsection denotes the requirements that are part of the application functionality

- **Recognize_Hand_Gestures** (application function)
  This is the most important feature of the application as it recognises the hand gestures made by a user. This function makes it possible to develop a large number of new functions such as `Play_Music`.
- **Play_music** (user function)
  This function allows the user to play simplified musical instrument using hand gesture along with the function `Recognize_Hand_Gestures`.

5.1.4. Non-functional requirements. This subsection denotes the requirements that are not part of the application functionality.

- The user must have a webcam on their computer.

5.2. Design

This section describes and explains the key choices that were made by the student to develop the application.

5.2.1. Tools List. It list each tools that the student used (e.g. programming languages, libraries).
Programming Language: Python is the programming language used to develop this application. It was chosen since it is a versatile programming language that will do the work both for the development a deep learning model and for the development of the application.
Libraries:

- Keras - A Python library for AI used to build the deep learning model
- OpenCv - A Python library for real time computer vision chose to perform video capture using the built-in camera.
- Mediapipe - A Python library for computer vision problem used in this project to get hand landmarks.
- Pandas - A Python library for data analysis chose for the creation and management of datasets.

Google Colab: A coding environment that is like a notebook in which GPU are allocated for your work. This tool was used due to incompatibility with libraries and the deep learning model was built, trained and saved on Google Colab.

5.2.2. Project Directory Structure. This section shows the directory structure of the project.

5.2.3. Project Components. This section denotes each components of this project (e.g. convolutional neural network, user interface).
Datasets:
Convolutional Neural Network:
Hand Gestures and Musical Note:
User Interface:

5.3. Production

5.4. Assessment

Acknowledgment

The authors would like to thank the BiCS management and education team for the amazing work done.

6. Conclusion

7. Plagiarism statement

This 350 words section without this first paragraph must be included in the submitted report and placed after the conclusion. This section is not counting in the total words quantity.

I declare that I am aware of the following facts:

- As a student at the University of Luxembourg I must respect the rules of intellectual honesty, in particular not to resort to plagiarism, fraud or any other method that is illegal or contrary to scientific integrity.
- My report will be checked for plagiarism and if the plagiarism check is positive, an internal procedure will be started by my tutor. I am advised to request a precheck by my tutor to avoid any issue.
- As declared in the assessment procedure of the University of Luxembourg, plagiarism is committed whenever the source of information used in an assignment, research report, paper or otherwise published/circulated piece of work is not properly acknowledged. In other words, plagiarism is the passing off as one's own the words, ideas or

work of another person, without attribution to the author. The omission of such proper acknowledgement amounts to claiming authorship for the work of another person. Plagiarism is committed regardless of the language of the original work used. Plagiarism can be deliberate or accidental. Instances of plagiarism include, but are not limited to:

1) Not putting quotation marks around a quote from another person's work
2) Pretending to paraphrase while in fact quoting
3) Citing incorrectly or incompletely
4) Failing to cite the source of a quoted or paraphrased work
5) Copying/reproducing sections of another person's work without acknowledging the source
6) Paraphrasing another person's work without acknowledging the source
7) Having another person write/author a work for oneself and submitting/publishing it (with permission, with or without compensation) in one's own name ('ghost-writing')
8) Using another person's unpublished work without attribution and permission ('stealing')
9) Presenting a piece of work as one's own that contains a high proportion of quoted/copied or paraphrased text (images, graphs, etc.), even if adequately referenced

Auto- or self-plagiarism, that is the reproduction of (portions of a) text previously written by the author without citing that text, i.e. passing previously authored text as new, may be regarded as fraud if deemed sufficiently severe.

References

[1] Feng-Sheng Chen, Chih-Ming Fu, and Chung-Lin Huang. "Hand gesture recognition using a real-time tracking method and hidden Markov models". In: *Image and vision computing* 21.8 (2003), pp. 745–758. URL: https://www.sciencedirect.com/science/article/pii/S0262885603000702.

[2] Zhi-hua Chen et al. "Real-time hand gesture recognition using finger segmentation". In: *The Scientific World Journal* 2014 (2014). URL: https://downloads.hindawi.com/journals/tswj/2014/267872.pdf.

[3] F. Chollet. *Deep Learning with Python, Second Edition*. Manning, 2021. ISBN: 9781638350095. URL: https://books.google.fr/books?id=mjVKEAAAQBAJ.

[4] Thippa Reddy Gadekallu et al. "Hand gesture recognition based on a Harris hawks optimized convolution neural network". In: *Computers and Electrical Engineering* 100 (2022), p. 107836. URL: https://www.sciencedirect.com/science/article/pii/S004579062200129X.

[5] Farah Farhana Mod Ma'asum, Suhana Sulaiman, and Azilah Saparon. "An overview of hand gestures recognition system techniques". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 99. 1. IOP Publishing. 2015, p. 012012. URL: https://iopscience.iop.org/article/10.1088/1757-899X/99/1/012012/pdf.

[6] Sascha Mahlke. "User experience: usability, aesthetics and emotions in human-technology interaction". In: *Towards a UX Manifesto* 26 (2007). URL: https://www.irit.fr/recherches/ICS/projects/cost294/upload/506.pdf#page=32.

[7] Jakob Nielsen. *Usability engineering*. Morgan Kaufmann, 1994. URL: https://books.google.fr/books?hl=en&lr=&id=95As2OF67f0C&oi=fnd&pg=PR9&dq=J.+Nielsen,+Usability+engineering+.+Morgan+Kaufmann,+1994.&ots=3cGAFlftVq&sig=Kj4nuR8_XIlbT4cXqyZFix-qKro#v=onepage&q=J.%20Nielsen%2C%20Usability%20engineering%20.%20Morgan%20Kaufmann%2C%201994.&f=false.

8. Appendix

8.1. State of the art

[1] This reference talks about a real time hand tracking method used to do hand gestures recognition. I chose it because it explains the process of a real time tracking method.
[2] This reference explains how the palm and fingers can be segmented in a real time hand tracking model. I think this reference can be useful since it includes the palm and fingers segmentation notions.
[3] This reference is a book which explains how deep learning work using theory and application (i.e. using Keras library). I bought this book based on my tutor's advice and it helps me to understand deep learning and to program my first convolutional neural network.
[4] This reference talks about using an optimized convolutional neural network to perform hand gestures recognition. I chose it since I thought it could be a great assets to complement the book.
[5]This reference is an article that presents an overview of different hand gestures technique. I think this article can be a great first reference to read before starting this project, since it talks about many different hand gestures recognition techniques, it could give me a global view of what is hand gestures recognition.
[6] This document is an article that explains a framework to assess the user experience of an application. It includes an introduction that present the different keywords of the paper such as user experience, instrumental and non instrumental quality and emotional user reaction. It also includes the presentation of the framework used to assess UX and finally it presents different study cases. This article helped me to understand the definition of user experience and how can we evaluate it.
[7] This book relates one definition of the usability. It explains

how to assess it by using different criteria such as learnability, efficiency, memorability, errors and satisfaction. This book was really useful to understand one definition of usability and how to estimate the usability of an application.