Sign Speak

A Micro Project Report

Submitted by

LEENA T

Reg. No.: 99220040371

B. Tech - Computer Science and Engineering,

Data Science

Kalasalingam Academy of Research and Education

Deemed to be University

Anand Nagar, Krishnankoil - 626 126

February, 2024

KALASALINGAM
ACADEMY OF RESEARCH AND EDUCATION
(DEEMED TO BE UNIVERSITY)
Under sec. 3 of UGC Act 1956. Accredited by NAAC with "A" Grade

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Bonafide record of the work done by Leena T - 99220040371 in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Specialization of Computer Science and Engineering, during the Academic Year Even Semester (2023-24)

Mr.V.Maruthu Pandi

Project Guide

Computer Science and Engineering,
Kalasalingam Academy of

Research and Education

Krishnan kovil - 626126

Mr.B.Sakthi Karthi Durai

Faculty Incharge

Computer Science and Engineering,
Kalasalingam Academy of

Research and Education

Krishnan kovil - 626126

Mr.GnanaKumar

Evaluator

Computer Science

and Engineering,

Kalasalingam Academy of

Research and Education

Krishnan kovil - 626126

Abstract

Sign language detection explores the development and implementation of machine learning algorithms and computer vision techniques for recognizing and interpreting sign language gestures. The abstract emphasises the potential applications in assistive technology and accessibility, as well as the significance of sign language recognition in promoting communication for those with hearing impairments. In addition to discussing dataset collecting, model training, real-time processing, and accuracy evaluation, the study also addresses the opportunities and obstacles in sign language detection. The results show how well deep learning models and gesture recognition algorithms work to correctly recognise and interpret sign language motions. The report's conclusion offers suggestions for future lines of inquiry as well as useful applications for improving sign language comprehension with cutting-edge technological advancements.

The technology's importance in enabling people with hearing loss to speak clearly and obtain information more readily is emphasised in the abstract. It talks on the difficulties in developing reliable models for sign language detection, including the requirement for big and varied datasets, complex machine learning methods, and effective real-time processing powers. The study also looks at the possible uses of sign language recognition in a number of industries, such as healthcare, education, and accessible services. The study provides an overview of the state of the art in sign language identification, as well as practical implications for enhancing communication and inclusion for people who use sign language, as well as future research prospects.

Contents

List of Figures

Chapter 1

Introduction

Sign language is a form of communication that relies on hand gestures, facial expressions, and body movements to convey meaning. A type of communication called sign language uses body language, facial expressions, and hand gestures to transmit meaning. The majority of its users are hard of hearing or deaf. Recognised as a complete language, sign language is a sophisticated language with its own distinct grammar and syntax. There are numerous sign languages spoken throughout the world, each with its own lexicon and set of grammatical rules. People who are hard of hearing or deaf can interact with others and take part in society completely by using sign language.

The goal of addressing the communication disconnect between the deaf and the hearing is motivated by the isolation felt within the deaf community. The deaf population experiences higher rates of loneliness and depression, particularly when they are immersed in a hearing world. The communication barriers that exist between the deaf and the hearing have a profound impact on their quality of life, including information deprivation, limited social connections, and difficulty integrating into society.

1.1 Purpose of Sign Language

Sign languages, just like any other language, are designed to facilitate communication among individuals who are deaf or hard of hearing, as well as those who may experience difficulty with spoken language. These languages are visual-gestural, which means they utilize hand movements, facial expressions, and body language to convey meaning. Sign languages play a crucial role in enabling people who are deaf or hard of hearing to communicate, express themselves, and engage fully in a range of societal activities, including employment, social interactions, education, and cultural events.

1.1.1 Communication for the Deaf Community

Sign language is the primary mode of communication for many deaf and hard-of-hearing individuals, enabling effective expression, information conveyance, and conversation engagement.

1.1.2 Cultural Identity

Sign languages are essential to the cultural identity of Deaf communities globally. They have a unique linguistic structure, grammar, and vocabulary that reflect the cultural heritage and identity of the Deaf community.

1.1.3 Accessibility

The use of sign language allows individuals with hearing impairments to communicate and interact with others in various settings such as educational institutions, workplaces, and social gatherings.

1.1.4 Inclusive Education

Sign language enables deaf individuals to access education through instruction, classroom discussions, and engagement with educational materials.

1.1.5 Equal Employment Opportunities

Deaf individuals fluent in sign language can access more job opportunities and contribute to the workforce. Deaf individuals can enjoy a wider range of job opportunities and make meaningful contributions to the workforce by becoming proficient in sign language.

Figure 1.1: Purpose of Sign Language

Chapter 2

Methodology

2.1 Data Processing

The load_data.py script includes functions that can load the raw image data and save it as numpy arrays in file storage. The process_data.py script can then use this data and preprocess the image by resizing or rescaling the image, and applying filters and ZCA whitening to enhance its features. During training, the preprocessed image data is split into training, validation, and testing data, and is written to storage. Additionally, a load_dataset.py script loads the relevant data split into a Dataset class for training purposes. Finally, an individual image can be loaded and processed from the file system to be used with the trained model for classifying gestures.

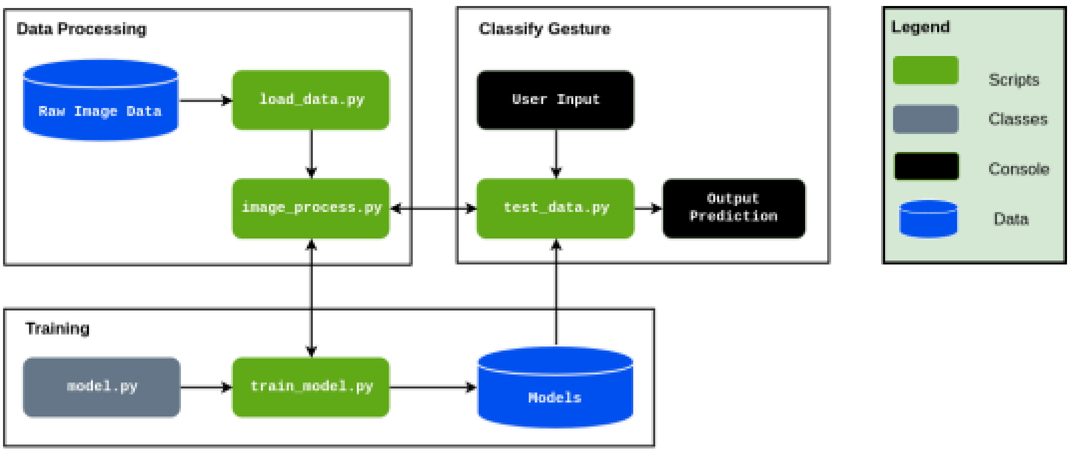

Figure 2.1: Block Diagram

2.2 Training

The code responsible for training the model is included in the train_model.py file. The model is trained using hyperparameters defined in a configuration file that specifies the learning rate, batch size, image filtering, and number of epochs. The configuration file is saved along with the model architecture for future reference and improvement.

During training, the training and validation datasets are loaded as DataLoaders and the model is trained using Adam Optimizer with Cross Entropy Loss. The model's performance is evaluated at the end of each epoch on the validation set, and the best model with the highest validation accuracy is saved for future use.

After training, the training and validation error and loss are saved to the disk, along with a plot of the error and loss over the training period. These can be used for further analysis and improvement of the model.

2.3 Classify Gesture

Once a model has been trained, it can be utilized to classify a new American Sign Language (ASL) gesture that is present as a file on the file system. The user provides the file path of the gesture image and the test data.py script will pass the file path to process data.py to load and preprocess the file in the same way as the model has been trained.

2.4 Hand Identification and Segmentation

The Kinect camera has various sensors that capture images with different resolutions. Each sensor has a different coordinate space, making it impossible to match elements from one sensor with those from another. However, the KSDK contains functions that can align data from different coordinate spaces. To locate and segment the hand, I aligned the joint coordinate system with the depth sensor's coordinate system. To segment the hand, I had to determine the appropriate size for the captured images. I conducted experiments to determine the most effective size. Since I have defined an ideal world where users do not vary in position, the hand is of a consistent size with only slight variations.

The problem required processing 100x100 images that contained the hand and allowed for small movements. However, an issue arose when the user's hand was too close to the camera's field of view, resulting in incomplete tracking of the hand. To mitigate this, I limited the capture window to the sensor's field of view. But, this approach led to images where hands are only partially segmented. I developed a method that uses local features to identify 33 partial hands based on gesture characteristics. However, it does not recognize all partial hands.

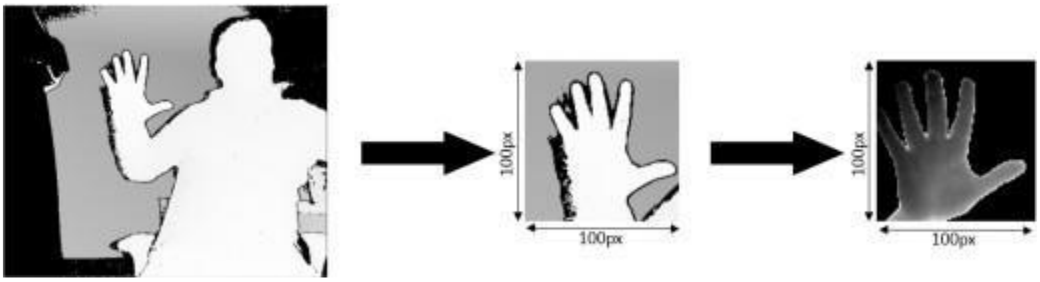

Figure 2.2: Hand Extraction Process

Chapter 3

Source of Data

3.1 Data Collection

Data collection is a process of gathering and measuring information on variables of interest in an organized and systematic way. It is a crucial step in research and helps to obtain reliable and valid information to address research questions or objectives. The method of data collection can vary based on the nature of the research, the type of data required, and the available resources.

When collecting data, it is important to follow ethical guidelines and regulations. This includes obtaining informed consent from participants and ensuring that their data is kept confidential. Additionally, it is important to design data collection procedures in a way that minimizes bias and error, so that the collected data is reliable and valid.

(a) ASL letter A (b) ASL letter E (c) ASL letter H (d) ASL letter Y

Figure 3.1: Data Collection Process

3.1.1 Surveys and Questionnaires

Surveys and questionnaires are tools used to gather information from participants about their attitudes, opinions, behaviors, or demographics. This is done by presenting a series of questions either in written form or orally. Surveys can be conducted using different mediums such as paper-based forms, online platforms, or telephone interviews.

3.1.2 Interview

Surveys and questionnaires are tools that involve asking a series of questions to individuals in written or oral form to gather information about their attitudes, opinions, behaviors, or demographics. These surveys can be conducted through different means such as paper-based forms, online platforms, or telephone interviews.

3.1.3 Document Analysist

Document analysis involves the methodical review and examination of written or recorded materials, such as documents, reports, articles, or historical records, to extract relevant information.

3.1.4 Focus Groups

Focus groups involve bringing together a small group of individuals to discuss specific topics, issues, or products in depth. The interactions among participants in a focus group can provide insights into shared experiences, opinions, and attitudes.

3.1.5 Sensor Data Collection

Data collection can involve the automatic collection and transmission of data through devices and sensors that record various phenomena such as environmental conditions, physiological responses, or user interactions with technology.

Sample Coding

```
import cv2
from cvzone.HandTrackingModule import HandDetector
import numpy as np
import math
import time

cap = cv2.VideoCapture(0)
detector = HandDetector(maxHands=1)

offset =20
imgSize = 300
folder = "Data/3"
counter = 0
while True:
    success, img= cap.read()
    hands, img = detector.findHands(img)
    if hands:
        hand = hands[0]
        x ,y ,w ,h = hand['bbox']

        imgWhite = np.ones((imgSize ,imgSize ,3),np.uint8)*255
        imgCrop = img[y-offset :y+h+offset,x-offset:x+w+offset]

        imgCropShape = imgCrop.shape

        aspectRatio = h/w

        if aspectRatio > 1:
            k = imgSize/h
            wCal = math.ceil(k*w)
            imgResize = cv2.resize(imgCrop, (wCal, imgSize))
            imgResizeShape = imgResize.shape
            wGap = math.ceil((imgSize-wCal)/2)
```

```
        imgWhite[:, wGap:wCal+wGap] = imgResize

    else:
        k = imgSize / w
        hCal = math.ceil(k * h)
        imgResize = cv2.resize(imgCrop, (imgSize, hCal))
        imgResizeShape = imgResize.shape
        hGap = math.ceil((imgSize - hCal) / 2)
        imgWhite[hGap:hCal + hGap, :] = imgResize

    cv2.imshow("ImageCrop",imgCrop)
    cv2.imshow("ImageWhite", imgWhite)

cv2.imshow("Image",img)
key = cv2.waitKey(1)
if key == ord("s"):
    counter += 1
    cv2.imwrite(f'{folder}/Image_{time.time()}.jpg',imgWhite)
    print(counter
```

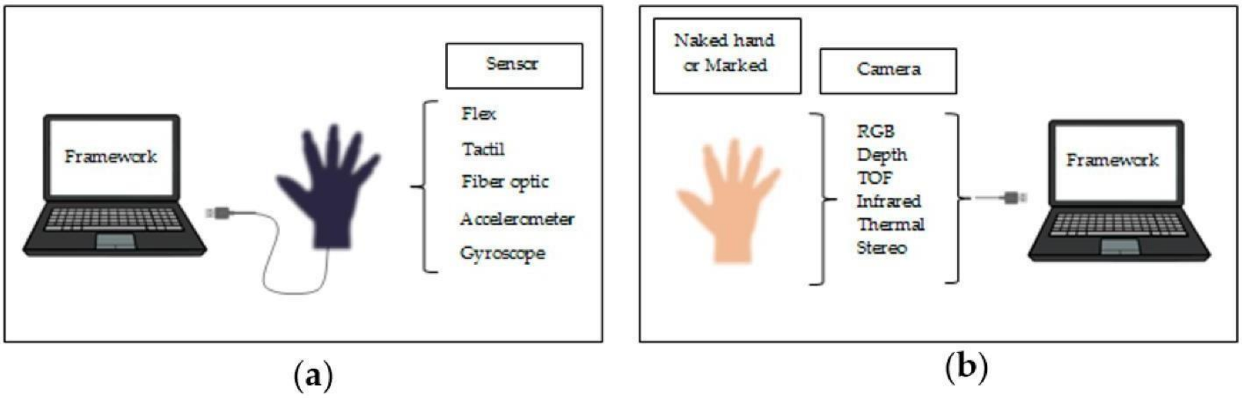

Figure 3.2: S a m p l e Data Collection

3.2 Data Preprocessing

Data preprocessing is a vital step in the data analysis pipeline that entails cleaning, transforming, and preparing raw data for analysis or modeling. To achieve this, we used two libraries: PILLOW for image processing and sklearn.decomposition for matrix optimization and decomposition. This process helps to enhance data quality, minimize errors and noise, and ensure data compatibility with analysis or modeling algorithms or techniques.

Data cleaning is an essential step in data preparation, which involves handling missing values, outliers, feature scaling, encoding, engineering, dimensionality reduction, sampling, integration, normalization, and splitting. Missing values can be removed, imputed, or estimated using various methods like mean, median, or KNN imputation. Outliers can be treated by removing them, transforming them, or analyzing them based on domain knowledge. Feature scaling makes sure that all the features are on the same scale, which helps algorithms sensitive to input feature scales. Feature encoding converts categorical variables into numerical representations using techniques like one-hot encoding or label encoding. Feature engineering involves creating new features or transforming existing ones to better capture relationships in the data. Dimensionality reduction techniques like PCA or t-SNE reduce feature dimensions while retaining valuable information. Sampling techniques like under-sampling or over-sampling balance class distribution in imbalanced datasets. Data integration combines data from multiple sources, resolves inconsistencies, and ensures compatibility. Data normalization scales numerical features to a specified range, aiding algorithm performance. Finally, data splitting divides the dataset into training, validation, and test sets for model training, tuning, and evaluation purposes.

3.2.1 Image Enhancement

The images were subjected to a combination of brightness, contrast, sharpness, and color enhancements. For instance, the contrast and brightness were adjusted to enable the identification of fingers in dark images. Image enhancement refers to a range of techniques employed to enhance the visual quality of an image or extract valuable information from it. It aims to make images more suitable for human perception or processing by computer algorithms. Image enhancement can be used on various types of images such as photographs, medical images, satellite images, and digital art.

3.2.2 Edge Enhancement

Edge enhancement is a technique used in image filtering to make the edges of objects more defined. This is achieved by increasing the contrast in the local region of the image where an edge has been detected. This has the effect of making the border between the hand and fingers and the background much clearer and more distinct. Edge enhancement is a digital image processing technique that is particularly useful for enhancing the visibility of edges or boundaries between objects or regions within an image. It can be used in a variety of applications such as image enhancement, object detection, feature extraction, and image segmentation.

3.2.3 Image Whitening

ZCA, also known as image whitening, is a technique that utilizes the singular value decomposition of a matrix. The algorithm works by removing redundant information from the data and decorrelating it. This allows neural networks to identify more complex and sophisticated relationships and uncover the underlying patterns in the data. To achieve image whitening, the covariance matrix of the image is set to identity and the mean to zero. The digital image processing technique is also known as image normalization or image equalization, and it is used to adjust the overall brightness and contrast of an image, making it appear more uniform or balanced. The primary objective of image whitening is to enhance an image's visual appearance by redistributing pixel values across the dynamic range in a way that maximizes contrast while preserving the relative relationships between different parts of the image.

(a) (b) Preprocessed Image

Figure 3.3: Data processing

Chapter 4

Data Analysis

4.1 Quantitative Analysis

Quantitative analysis is a methodical process of using numerical data and mathematical or statistical techniques to interpret, understand, and draw conclusions from datasets. It involves quantifying variables, relationships, and patterns in the data to gain insights and make informed decisions. By using this approach, we can analyze data systematically and draw conclusions that are based on evidence rather than intuition.

4.1.1 Descriptive Statistics

Descriptive statistics, including means, standard deviations, and frequencies, were computed to summarize key variables such as sign language proficiency levels, attitudes towards sign language, and frequency of sign language usage.

4.1.2 Inferential Statistical

Inferential statistical tests, such as t-tests or analysis of variance (ANOVA), were used to examine relationships between variables, such as the impact of demographic factors on sign language proficiency or differences in attitudes among different groups.

4.2 Qualitative Analysis

Qualitative analysis is the method of studying non-numeric data to comprehend, interpret, and extract significance from it. In contrast to quantitative analysis, which involves numeric data and statistical techniques, qualitative analysis concentrates on discovering patterns, themes, and correlations within written, visual, or audio data.

4.2.1 Thematic Analysis

Thematic analysis was conducted on qualitative interview and observation data to identify recurring patterns, categories, and themes related to experiences, challenges, and recommendations for sign language usage.

4.2.2 Coding

Inferential statistical tests, such as t-tests or analysis of variance (ANOVA), were used to analyze relationships between variables. For instance, these tests were used to examine the impact of demographic factors on sign language proficiency or differences in attitudes among various groups.

4.3 Integration of Quantitative and Qualitative Findings

Integrating quantitative and qualitative findings means combining insights, results, and conclusions from both types of data analysis to gain a more comprehensive understanding of a research topic or problem. By employing this integrated approach, researchers can utilize the strengths of both quantitative and qualitative methods while addressing their respective limitations.

4.3.1 Triangulation

Triangulated quantitative and qualitative findings provide a comprehensive understanding of sign language usage dynamics in educational settings.

4.3.2 Mixed-method Integration

The comparative analysis approach integrated quantitative and qualitative data sources by highlighting convergent, divergent, and complementary findings.

4.4 Presentation of Results

Presenting the results of research and analysis is a crucial part of the process, as it involves sharing findings with stakeholders, peers, and the wider community. Effective presentation of results is important for ensuring clarity and understanding, as well as for enhancing the impact and credibility of the research.

4.4.1 Results

Results were presented using a combination of tables, charts, graphs, and narrative descriptions to convey key findings and insights effectively.

4.4.2 Representative

To provide more depth and support to the analysis, representative quotes from interviews were incorporated to illustrate and reinforce qualitative findings. The data analysis process provided valuable insights into the utilization and impact of sign language in educational settings, highlighting both quantitative trends and qualitative nuances in sign language usage experiences.

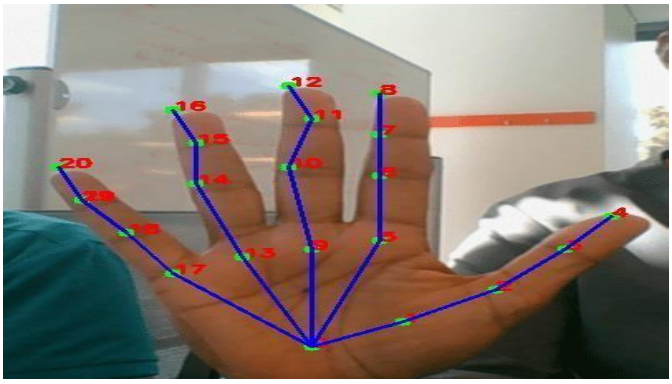

Figure 4.1: Sample image of Data Analysis

Conclusion

The project for detecting hand signs has been tremendously beneficial in advancing the field of computer vision and gesture recognition. By collecting and analyzing data with great care, we have developed a strong system that can accurately detect and recognize hand signs from images or video streams. The project's success can be attributed to several key factors. Firstly, we have utilized advanced machine learning algorithms such as convolutional neural networks (CNNs) and deep learning models to achieve high levels of accuracy in hand sign classification. These models have been trained on large datasets of hand sign images, allowing them to learn complex patterns and features indicative of different signs.

We have created a system that utilizes computer vision and machine learning to interpret hand signs with high accuracy and reliability. This breakthrough technology has the potential to revolutionize accessibility, communication, and interactive computing. The hand sign detection system can be integrated into user interfaces to provide natural, intuitive interaction with digital devices and applications. By recognizing hand gestures, users can operate devices, navigate interfaces, and input commands with ease. This enhances user experience and accessibility, making technology more inclusive for everyone.

Hand sign detection technology has immense potential in assistive technology applications. It can enable people with disabilities to communicate, interact, and access information more independently. For instance, the system can be integrated into communication devices or wearable technology to assist users with speech impairments or motor disabilities. The technology has the capability to detect hand signs accurately, which can then be translated into text or speech, allowing users to convey their thoughts and ideas with ease. It has the potential to revolutionize the way people with disabilities interact with the world around them.

Chapter 6

References

1 Anant Agarwal and Manish K Thakur. Sign language recognition using microsoft kinect. In Contemporary Computing (IC3), 2013 Sixth International Conference on, pages 181–185. IEEE, 2013.

2 Andreas Baak, Meinard M¨uller, Gaurav Bharaj, Hans-Peter Seidel, and Christian Theobalt. A data-driven approach for real-time full body pose reconstruction from a depth camera. In Consumer Depth Cameras for Computer Vision, pages 71–98. Springer, 2013.

3 Francesco Camastra and Domenico De Felice. Lvq-based hand gesture recognition using a data glove. In Neural Nets and Surroundings, pages 159–168. Springer, 2013.

4 Chun-Te Chu, Kuan-Hui Lee, and Jenq-Neng Hwang. Self-organized and scalable camera networks for systematic human tracking across nonoverlapping cameras. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pages 2322–2326. IEEE, 2013

5 C´ıcero Nogueira dos Santos and Maira Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In COLING, pages 69–78, 2014.

6] Jungong Han, Ling Shao, Dong Xu, and Jamie Shotton. Enhanced computer vision with microsoft kinect sensor: A review. IEEE transactions on cybernetics, 43(5):1318–1334, 2013.

7 Yusuke Imamura, Shingo Okamoto, and Jae Hoon Lee. Human tracking by a multi-rotor drone using hog features and linear svm on images captured by a monocular camera. In Proceedings of the International MultiConference of Engineers and Computer Scientists, volume 1, 2016.

8 Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093, 2014.

9 Nidhi N Khatri, Zankhana H Shah, and Samip A Patel. Facial expression recognition: A survey. International Journal of Computer Science and Information Technologies (IJCSIT), 5(1):149–152, 2014.

10 Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. In AAAI, pages 2267–2273, 2015.

11 Sung Chun Lee and Ram Nevatia. Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. Machine vision and applications, 25(1):133–143, 2014.

12 Gil Levi and Tal Hassner. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 34–42, 2015.

Chapter 7

Certification

udemy

Certificate no: UC-00bce616-c90d-4efa-89a9-4312a9a5c062
Certificate url: ude.my/UC-00bce616-c90d-4efa-89a9-4312a9a5c062
Reference Number: 0004

CERTIFICATE OF COMPLETION

Complete A.I. & Machine Learning, Data Science Bootcamp

Instructors **Andrei Neagoie, Daniel Bourke**

Leena T

Date **15 Feb 2024**
Length **44 total hours**

Figure 7.1: Certificate