University of Sheffield

Multi Sensor Fusion Face Identification in Secure AI Powered Robot

University of Sheffield

Mohammad Aadil Minhaz

Supervisor: Dr. Prosanta Gope

A report submitted in fulfilment of the requirements
for the degree of MSc in Cybersecurity and Artificial Intelligence

in the

Department of Computer Science

September 13, 2023

Declaration

All sentences or passages quoted in this report from other people's work have been specifically acknowledged by clear cross-referencing to author, work and page(s). Any illustrations that are not the work of the author of this report have been used with the explicit permission of the originator and are specifically acknowledged. I understand that failure to do this amounts to plagiarism and will be considered grounds for failure in this project and the degree examination as a whole.

Name: Mohammad Aadil Minhaz

Signature: Mohammad Aadil Minhaz

Date: September 06, 2023

I would like to dedicate this thesis to my beloved Parents and my Cat.

Abstract

In the rising AI era, the application of AI in robotics has become a primary research focus in various areas, including healthcare, delivery, manufacturing, and more. For instance, the increasing demand for package and food delivery necessitates a substantial workforce and expenditure. In healthcare, the surge in demand for health and support workers during the pandemic necessitates innovative solutions. AI-powered Smart Robotics offer potential remedies for these real-world challenges. Helper Robots can assist patients and perform basic tasks, minimizing the risk to workers from contagious diseases. Intelligent Drones and ground robots can significantly reduce package delivery costs. While several companies have already deployed such robots for Last-Mile Delivery and care, the use of AI Robots presents numerous challenges to overcome. In this dissertation project, we focus on the challenges related to the secure interaction of robots with users. Our proposal addresses three major problems related to User Identification by an AI-enabled Robot: Face Identification in real-world unfavorable scenarios, Multi-Sensor Fusion Face Identification, Identification of newly registered users with an unbalanced dataset (few pictures of new users), and Security against a physical adversary such as the Flat-Face Adversary. We have addressed these issues with our proposed solution on a Turtle-Bot3 Robot platform.

Acknowledgement

I would like to start by expressing my gratitude towards my supervisor and mentor Dr. Prosanta Gope for his guidance, expertise and unwavering support which has been invaluable throughout this research. I also want to thank my family and friends, they have been a constant source of encouragement and emotional support. Their belief in me has been a driving force behind my academic achievements. I want to show my appreciation to my fellow students and colleagues who engaged in discussions, shared insights and helped me navigate the challenges of academic life.

Contents

1 Introduction **1**
- 1.1 Related Work . . . 2
  - 1.1.1 AI in Robotics for Delivery . . . 2
  - 1.1.2 Robots in healthcare . . . 2
  - 1.1.3 Multi-Sensor Fusion . . . 2
  - 1.1.4 Research in Academia . . . 3
- 1.2 Aims and Contributions . . . 3
- 1.3 Structure of the Thesis . . . 4

2 Preliminaries **5**
- 2.1 Machine Learning for Robots and IOT . . . 5
  - 2.1.1 MobileNetV3 . . . 5
  - 2.1.2 ShuffleNetV2 . . . 5
  - 2.1.3 Vision Transformer . . . 6
- 2.2 Face Identification . . . 6
  - 2.2.1 Face Embedding . . . 6
  - 2.2.2 AAMSoftmax . . . 7
  - 2.2.3 ArcFace . . . 7
  - 2.2.4 ROC curve Thresold . . . 8
  - 2.2.5 Cosine Similarity . . . 8
- 2.3 Digital and Physical Adversaries . . . 9
  - 2.3.1 Clever Hans . . . 9
  - 2.3.2 FGSM - Fast gradient sign method . . . 9
  - 2.3.3 Flat-Image Input . . . 10
- 2.4 RGB-IR Combo Dataset . . . 10

3 Methodology **11**
- 3.1 Proposed System Model and Adversary . . . 12
  - 3.1.1 Overall System Design . . . 12
  - 3.1.2 Proposed Adversary . . . 12
- 3.2 Authentication by Face Identification . . . 12
  - 3.2.1 Multi-Sensor Fusion . . . 13

3.2.2 Training Process . . . 13
3.2.3 The Authentication Process . . . 14
3.2.4 Attack Process . . . 14
3.2.5 Flat-Image Input . . . 15
3.2.6 FGSM - Fast Gradient Sign Method attack . . . 15
3.3 Summary . . . 15

4 Implementation and Evaluation **16**
4.1 Training - on Multi-Sensor Fusion Input Embedding . . . 16
4.2 Model Performance and Evaluation . . . 17
4.3 Testing . . . 19
4.4 Adversarial Experiment - FGSM Attack . . . 20
4.5 Adversarial Experiment - Flat-Image Input Attack . . . 21
4.6 Summary . . . 21

5 Multi-Sensor Fusion Face Recognition on Robot **22**
5.1 AI-powered secure Robot . . . 22
5.2 Multi-Sensor Fusion Module . . . 23
5.3 Registration and Identification Process . . . 24
5.4 Summary . . . 24

6 Discussion **25**
6.1 Comparison with Previous Work on the Robot . . . 25
6.1.1 User privacy . . . 25
6.1.2 Multi Image Sensor Authentication . . . 25
6.1.3 Model Efficiency in Unfavourable Conditions . . . 26
6.1.4 Secured against Flat-Image Input Attack . . . 26
6.1.5 Seamless Authentication . . . 26

7 Conclusion and Future Research **27**
7.1 Conclusion . . . 27
7.2 Future Research . . . 27

Appendices **33**

A Code **34**

List of Figures

2.1 Architecture of MobileNetV3 6
2.2 Architecture of Vision Transformer 7
2.3 Architecture of ArcFace 8
2.4 9
2.5 10

3.1 Process 11
3.2 Fusion Patch Embedding Layer 13
3.3 Overall Architecture 14

4.1 ROC curve to calculate the best threshold 17
4.2 18
4.3 Authentication of a person wearing glasses 19
4.4 Authentication of a person in darkness 19
4.5 Authentication of a person in darkness 20
4.6 Flat-Image Input attack on the robot 21

5.1 Overall System Components 22
5.2 Multi-Sensor Fusion User Interaction Module 23
5.3 Registration and Identification Process on the Robot 24

List of Tables

4.1 Training Hyper-Parameters 16
4.2 Evaluation Results with different Optimisers and K-Folds Cross Validation . 18
4.3 Performance of Model with FGSM attack 20

5.1 Details of overall system components 23

6.1 Comparison With Previous Works 25

Chapter 1

Introduction

The rise in AI and recent research on AI-based systems opens up many possibilities for real-life problem solutions using AI-powered devices such as Robots. Simple commercial robots are being used in various sectors such as Last-mile delivery, Healthcare and many other places. Much research and implementations have been going on in the area of Last Mile Delivery using Automated Smart Robot [25] [7].

According to the McKinsey Global Institute's report [18] study, e-commerce grew in the United States, China, United Kingdom, Spain, Germany, India, France, and Japan three to five times faster than before in 2020. Also during the past five years, the global food delivery market has tripled to a value of $150 billion. [1].It is projected that the global market for Last Mile Delivery will be worth US $269.2 billion by 2030, growing at a compound annual growth rate of 6.7% between 2022 and 2030 [11]. The last mile is responsible for approximately 41% to 53% of total supply chain costs [30].

As an alternative to traditional delivery methods, robotic unmanned delivery systems are able to mitigate the challenges involved in traditional delivery methods, primarily by reducing human labour and other logistic costs. Additionally, the unmanned mode of deliveries can also reduce the risk of infections of COVID-19 due to their efficiency and ability to carry out deliveries rapidly and efficiently in open-air environments where minimal interaction with humans is required.

Healthcare has several growing challenges such as the surge in the demand for healthcare workers, the need for care for the growing elderly population and the safety of healthcare workers in critical times of Pandemics (COVID-19). Globally, the number of older people is expected to double over the next three decades, reaching over 1.5 billion by 2050[32]. A major concern in providing healthcare is the insufficient availability of healthcare workers as well as the declining productivity caused by burnout [23] [5]

A variety of robots with varying specifications and capabilities have been studied in recent AI and robotics research in healthcare[14] [33]. Many of the applications and surveys of Care Robots are geared towards providing mental and emotional support to older patients. Helping elder adults with basic needs includes tasks such as providing medicines as per the schedule and delivering essential items. Smart Robot Pepper [21] from Softbank Robotics

and PHAROS [20] are examples of social robots with a potential application in care.

The use of AI in robotics can solve several real-life problems in Health, Delivery and other areas. It can provide fast, cost-effective and safe alternatives to human labour. Nevertheless, AI in Robotics faces several problems in identifying and authenticating the user. Different modes of authentication are present to interact with the user in order to authenticate such as Facial, One-Time-Passwords, Finger Prints and Voice Recognition but in real-world scenarios, these methods face many challenges like identification in low-light and identification with minimal cooperation from the user. In this thesis, the challenges associated with the identification of users using facial identification by the AI robot are discussed. The thesis focuses on the use of AI robots to identify users and the challenges associated with this process. It evaluates the potential of AI robots to identify users accurately and discusses the potential risks associated with the use of this technology.

1.1 Related Work

1.1.1 AI in Robotics for Delivery

Much research has been done related to Delivery using AI-powered Robots. [28]. Several companies are already employing ADRs to perform their delivery operations such as Amazon's delivery robot [2], Titan AI robots by Terminus and Starship's food delivery Robots[29].

1.1.2 Robots in healthcare

In healthcare, various nonsurgical Robotics solutions have been implemented to help users.

Pepper robot by Softbank Robotics uses facial recognition models and identifies facial expressions and voice tones to interact with the user. It provides assistance to people needing support [21].

Another proposal, PHAROS [6], an assistive robot that monitors and evaluates users' daily physical activity at home. For this purpose, machine learning techniques (such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs)) are used to identify and evaluate exercise performance. Additionally, it incorporates a recommendation system that generates the workout every day so that the individual is working on what is necessary to remain healthy [20].

1.1.3 Multi-Sensor Fusion

Research related to the fusion of multi-sensors to fuse an Infrared and a Visible Image has been carried out by Dongry Rao et al. [24]. The author proposed an algorithm for fusing infrared and visible images by using a lightweight transformer module and adversarial learning. A unique approach is employed in the study to learn global fusion relations. In order to refine the fusion relationship within the spatial scope and across channels, shallow features are extracted by CNN and combined with transformer fusion modules.

1.1.4 Research in Academia

In academia, Yang et al. [36] proposed a secure shipping robot that used cooperative and non-cooperative modes to interact and identify the user to complete the delivery of ordered items. The author conducted several experiments to benchmark the framework's safety against adversarial attacks. Another approach related to secure delivery using an AI-enabled Robot was carried out by Prosanta Gope et al.[35] that implements a facial and voice recognition framework to identify the customer.

The related works presented here address several challenges in real-life scenarios and provide cost-effective, safe, and efficient solutions. Nevertheless, further research is necessary to address further challenges related to the interaction of the Robot with the user so that the user can be identified in a safe and seamless manner without requiring much input on the user's part.

1.2 Aims and Contributions

This thesis highlights the challenges related to the Identification of users using Facial Authentication by the AI Robot. Facial authentication is a reliable and secure way to identify users, but it can be difficult for a robot to accurately identify a user's face due to factors such as lighting, angle, and expression. Additionally, robots may be more likely to make mistakes in authentication if the user is wearing glasses or other items that interfere with facial recognition. In addition, a solution is offered to address attacks against simple adversaries, such as flat-image input, against the robot's facial recognition.

Our proposed system aims to address major challenges associated with facial identification authentication.

- Identification of the User in Unfavourable Conditions, such as low light or darkness, subject with a different emotion, subject wearing glasses and many more.
- Model should work with unbalanced, low data of the new registered user
- The Robot's User Identification system should withstand Adversarial attacks such as Flat-Image input and FGSM attacks.

The major contributions of this thesis:

- A new Fusion Patch Embedding module is introduced that takes multi-sensor inputs, including RGB and IR-Thermal cameras, performs patch size splitting and then projects them into a single embedding as the input for the Vision Transformer.
- In this paper, we propose an enhanced Vision Transformer Model with ArcFace loss in order to achieve higher similarity for intraclass samples and increased diversity for interclass samples (users). Additionally, a customized ROC curve is used to determine the best threshold for identifying the cosine similarity between the input and the user's original image.

- The proposed model is implemented on a real-life robot with dual cameras (RGB and IR-Thermal) to assess its security against adversaries, such as flat-image input. In this scenario, the robot would fail an authentication attempt when presented with an image (printed or on some screen) of the real user by an adversary.

1.3 Structure of the Thesis

The dissertation is structured as follows:

- **Chapter 2: "Preliminaries"** We analyse and provide brief highlights about Literature in the relevant fields.
- **Chapter 3: "Methodology"** We present our proposed methodology, User identification by facial authentication of the user by the robot.
- **Chapter 4: "Implementation and Evaluation"** We developed the proposed Vision Transformer with a unique Patch Embedding layer to fuse facial data from two different sensors, RGB and an IR camera. We carried out several experiments to measure our model performance and its robustness.
- **Chapter 5: "Robot Implementation"** We implement our proposed framework on the real-world robot platform.
- **Chapter 6: "Discussion"** we compare our proposed framework with respect to the existing related works on robot delivery systems.
- **Chapter 7: "Conclusion and Future Research "** we conclude our work and discuss future research direction.

Chapter 2

Preliminaries

The chapter provides a brief overview of all the related technologies and techniques used throughout the project. This chapter is divided into 4 sections: Machine Learning for Robots, Face Identification, Digital and Physical Adversaries, and RGB-IR-Thermal Combo Dataset used in this project.

2.1 Machine Learning for Robots and IOT

Limited computational resources makes it important to consider resource-efficient models to perform on devices such as Smart Robots to perform tasks like recognise the user. We have explained popular Deep Models for computational intensive devices [13] [27] [15] [26]

2.1.1 MobileNetV3

As Convolutional Neural Networks (CNN) have extensive use in image classification [16], image segmentation, and other computer vision problems, it has received a lot of attention. CNNs usually consist of two parts, namely, the Features Extraction Part and the Classifying part. The former is made of convolutional layers and pooling layers while the latter is made of many stacked fully connected layers. Due to this, there are many architecture variants based on CNN used for various tasks. For classification tasks in mobile devices or computationally constrained devices, we require low memory cost as well as efficiency in computation all while having high accuracy. This brings the requirement for models for mobile terminals. MobileNetV3 [16] is one such lightweight and efficient model which also provides good accuracy.

Figure 2.1 [16] shows architecture of MobileNetV3.

2.1.2 ShuffleNetV2

The ShuffleNet V2, developed by Ma et al. [19] and their team, shows us that we cannot simply measure the number of calculations of a neural network to determine its performance. In addition, we should consider where we intend to use it. According to ShuffleNet, we should

2.1

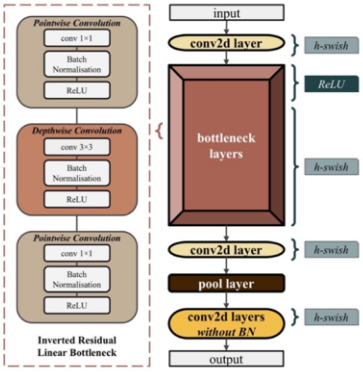


Figure 2.1: Architecture of MobileNetV3

test different models on the actual device where they will be used in order to determine which works best. As an improvement on ShuffleNet V1, ShuffleNet V2 adds a feature called "channel split" to make the model more accurate and efficient.

2.1.3 Vision Transformer

For the purpose of processing images, ViT uses a transformer architecture, which was originally developed for natural language processing tasks (explained in [34].

ViT explained in paper [38] divides images into patches, treats them as sequences of data, and processes them using self-attention mechanisms rather than convolutions like CNNs. ViT is able to capture long-range dependencies in images through this method, and it has demonstrated promising results for a variety of computer vision tasks, including image classification, object detection, and segmentation.

Vision Transformers combine transformer-based language models with computer vision to improve performance and understanding of visual information. Figure 2.2 [9] shows the architecture of Vision Transformer.

2.2 Face Identification

2.2.1 Face Embedding

Face embeddings are numerical representations of a person's facial features that capture essential characteristics of their face in a compact and standardised format. These embeddings serve as a condensed and discriminative way to represent faces, making it easier to compare and recognize them [3]. Once face embeddings are obtained, they can be used for various tasks such as face recognition, face verification, or face clustering.

2.2

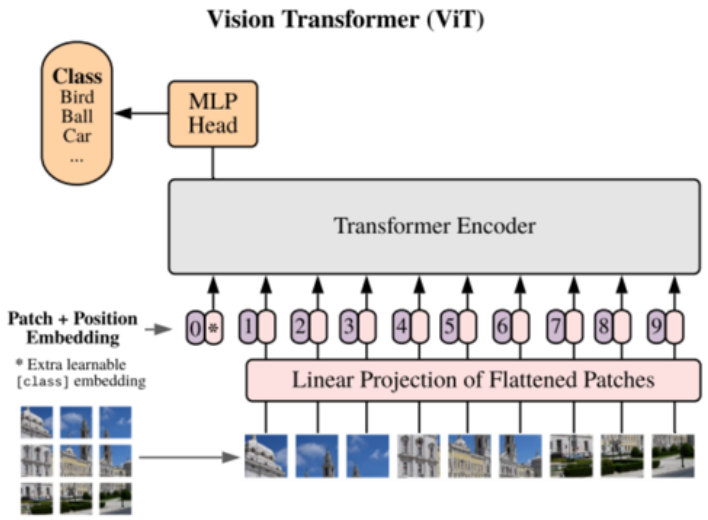


Figure 2.2: Architecture of Vision Transformer

2.2.2 AAMSoftmax

Softmax classifiers have been used widely in deep learning Frameworks to address various classification tasks. A softmax classifier can discriminate identities in training dataset based on large-scale training data and CNN architectures. The Additive Angular Margin softmax (AAM-Softmax) loss [17] was introduced to improve this discriminative power used for identification in a model further by incorporating an additive angular margin constraint to push features to the angular space with a fixed radius.

$$\text{AAMSoftmax: } \mathrm{L}_{\mathrm{AAMSoftmax}} = -\sum_{i=1}^{N} \log \frac{e^{s(\cos(\theta_{y_i} - m) - \cos(\theta_i))}}{e^{s(\cos(\theta_{y_i} - m) - \cos(\theta_i))} + \sum_{j \neq y_i} e^{s\cos(\theta_j)}} \tag{2.1}$$

2.2.3 ArcFace

ArcFace [8], or Additive Angular Margin Loss, is a loss function that is used in the recognition of faces. These tasks are traditionally performed using the softmax. As a result, under large intra-class appearance variations, the softmax loss function does not explicitly optimize the feature embedding in order to ensure higher similarity for intra-class samples and diversity for inter-class samples. Thus, deep face recognition suffers from a performance gap. ArcFace is similar to AAMSoftmax 2.2.2 with some additive advantage to dynamic marginal distribution.

Figure 2.3 [8] shows the architecture of ArcFace.

2.3

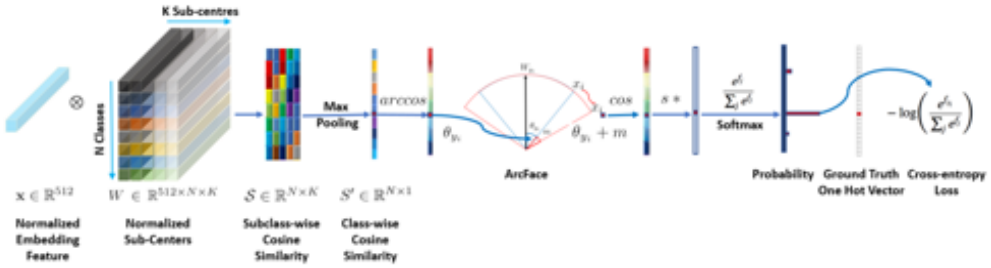


Figure 2.3: Architecture of ArcFace

2.2.4 ROC curve Thresold

In binary classification, a Receiver Operating Characteristic (ROC) [4] curve is a graphical plot of the false positive rate (FPR) against the true positive rate (TPR) at different classification thresholds. The FPR is the ratio of negative instances that are incorrectly classified as positive, while the TPR is the ratio of positive instances that are correctly classified as positive.The ROC curve threshold is the point on the curve that strikes a balance between the FPR and the TPR. A higher threshold will result in a lower FPR, but it will also result in a lower TPR. A lower threshold will result in a higher TPR, but it will also result in a higher FPR [10].

The optimal ROC curve threshold is the one that best meets the specific needs of the application. For example, an application that requires a high TPR may choose a lower threshold, while an application that requires a low FPR may choose a higher threshold [10].

To evaluate the performance of the model or test across varying threshold values, an ROC curve is constructed. The ROC curve graphs the True Positive Rate (Sensitivity) against the False Positive Rate at different threshold values. As the threshold is adjusted, the True Positive Rate and False Positive Rate undergo alterations, yielding distinct points on the ROC curve. The ROC curve's overall performance is summarised through the calculation of the Area Under the Curve (AUC)[10]. A higher AUC signifies superior discriminative capability of the model or test across an array of threshold values.

2.2.5 Cosine Similarity

Cosine similarity is a mathematical metric used to quantify the similarity between two vectors by calculating the cosine of the angle formed between them. It is a valuable tool in various domains, including information retrieval, natural language processing, and machine learning, where assessing data or feature similarity is crucial. Cosine similarity operates on vectorized data representations, with each vector encapsulating a set of features. It calculates the cosine of the angle between two such vectors, designated as vector A and vector B, using the formula:

$$CosineSimilarity(A, B) = (AB)/(||A|| * ||B||)$$

$Here, (AB) represents the dot product of vectors A and B, while ||A|| and ||B|| signify the magnitudes or lengths of th$

In summary, cosine similarity provides a mathematical means to assess the similarity between vectors based on their relative orientation. Its versatility and applicability make it a fundamental metric for similarity assessment across various research and practical domains.

2.3 Digital and Physical Adversaries

2.3.1 Clever Hans

[22] As mentioned in [22]. The Clever Hans is a metaphor for machine learning models. Sometimes, these Machine Learning models can do really well on a specific test, but they don't really understand the underlying details of the task. So, when it is tested on different kinds of tests, the result might not be good. The Clever Han GitHub library provides a Python library that can be used to benchmark the vulnerability of machine learning systems to adversarial examples.

2.3.2 FGSM - Fast gradient sign method

Author Good Fellow et al. [12] explains A Fast Gradient Sign Method (FGSM) as a type of adversarial attack in machine learning and deep neural networks. This technique involves making small, intentional changes to input data in order to impact the decision of a model. FGSM generates perturbed input data by calculating the gradient of the loss function with respect to the input and then adjusting the input in a manner that maximizes the loss. As a result of this perturbation, the model is likely to misclassify the input, even though it may not be noticeable to humans. By using FGSM attacks, machine learning models are tested for robustness against vulnerabilities. Figure 2.4 [12] shows the perturbed input impact.

2.4

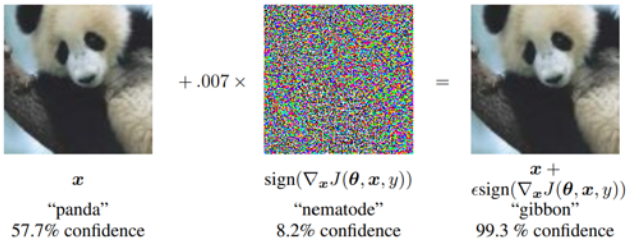


Figure 2.4:

2.3.3 Flat-Image Input

An adversary uses the "flat-image input attack" (see section 3.2.1) to bypass authentication in face recognition systems by presenting a static image of the user (printed or displayed on a screen). The limitation of having only one RGB input sensor prevents the system from distinguishing between a real, live human face and a two-dimensional representation, such as an image or picture. Therefore, the lack of depth or thermal perception and the inherent limitations of one sensor make it difficult to confirm the authenticity of a presented face, making the system vulnerable to such attacks.

2.4 RGB-IR Combo Dataset

We have used a special RGB-IR combo face dataset for the training and evaluation of our proposed face identification model. The dataset is an open-source project by Tufts University, Boston, USA [31].

There are 1532 images in the dataset, consisting of images of 113 individuals. Each person has five to nine RGB photographs as well as IR-thermal photographs. The size of each image is 128x128 pixels. There are images of users wearing glasses as well as images of users in various emotional states, angles, and lighting conditions.

Figure 2.5 shows some data samples.

2.5

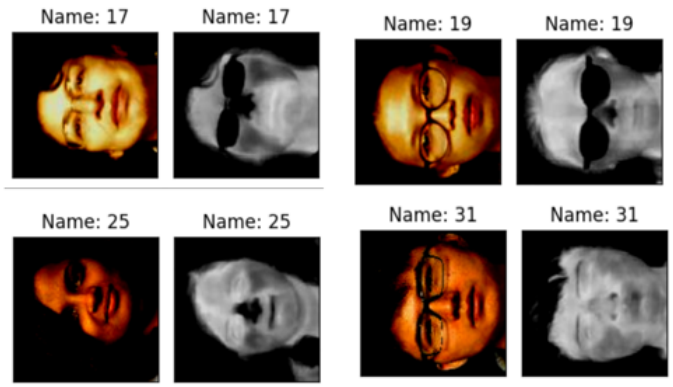


Figure 2.5:

Chapter 3

Methodology

This chapter explains our proposed system. To explain the overall system, we provide an overview of Face Identification by the Robot after the registration of a new user, explained by Figure 3.1. Followed by the architecture of the newly designed "Fusion Patch Embedding" along with ArcFace implementation using a custom Vision Transformer and finally evaluation of the Robot's Face Recognition module for accuracy and robustness against adversaries. This project solely focuses on the Authentication of the User by the Robot using the Multi-Sensor Face Identification module. Potential Applications by the robot after the authentication such as giving a package to the user, navigation, and providing essential items to the user are not covered in the scope of this thesis.

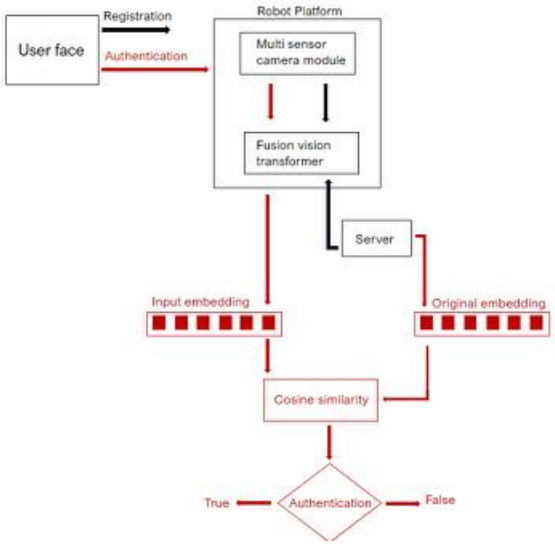


Figure 3.1: Process

3.1 Proposed System Model and Adversary

3.1.1 Overall System Design

The complete system is divided into two sections: Registration of the User by the Robot's User Interaction Module and Authentication of the User by Face Identification Module.

The registration process involves Capturing the user's face via the Multi-Sensor Fusion Module, putting a label (we used an anonymous label for the anonymity of the user) and storing the captured embedding in the database. The Registration module ensures the privacy of the user by storing only the face embedding and not the visible image.

The Robot's User Identification system works on the concept of going to the user as per the instructions and initiating the authentication by face identification. Robot has trained Fusion-Based Face Recognition Vision Transformer loaded already before performing the task. The face identification module captures the user's face embedding using a fusion mechanism and compares them with already registered reference face embedding to verify if the user is authenticated or not.

3.1.2 Proposed Adversary

To secure our robot's user identification module against adversaries, we must consider all potential threats. Challenges related to the vulnerability of the system include communication between the client and the robot. Additionally, the robot communicates with the server to register the user via the Internet, which can be susceptible to attackers. Ensuring the security of the code and machine learning models used in the system is also crucial to protect it from adversaries. In the proposed system, we are addressing two major adversaries, Clever Han' FGSM Attack [22] and Flat-Image Input.

3.2 Authentication by Face Identification

The section explains the working of the proposed Multi-Sensor Fusion-based Face Identification module to authenticate the user. To solve the user identification problem we have used a Deep Neural Network Vision Transformer that extracts the Face Embedding of the user. The model takes the input of the user's face from two camera sensors, RGB Camera and IR-Thermal Camera. The two inputs are then fused together to project as one embedding as the input to the transformer, section 3.2.1.

The proposed deep-learning model doesn't use the classifier to identify the face of a person to avoid the problem of retraining. Since the system's goal is to identify the registered person, when a new user gets registered in the system, the model should be able to identify the new user. In the usual classification technique, the model has to be trained again with new data in order to classify the newly registered users. Retraining could be significantly computationally expensive. To solve this, the face embedding of the User is captured and compared with the already stored Registered User's face embedding. At last, cosine-similarity between input face embedding and stored face embedding of the user is used for the final identification.

3.2.1 Multi-Sensor Fusion

This thesis focuses on the Major challenges related to face identification such as Identification in unfavourable conditions like low light or Darkness and protection against Flat-Image Input adversary explained in section 2.3.3.

To solve the problem of identification in real-world unfavourable scenarios, along with an RGB camera, We have introduced an additional IR-Thermal camera as an extra input sensor in the interaction module, to capture the user's face in low light or in darkness. The IR-Thermal camera enables the face identification module to capture the user's face embedding even in such scenarios, when RGB image would be noisy or completely dark.

Moreover, using an IR-Thermal camera, the model has extra information in the embedding and can determine whether the input was taken directly from the face or whether an adversary used a printed photo or a screenshot on a digital screen such as a phone or tablet. In order to achieve this, the thermal features of the face are compared to those of the original face. The model can detect forgeries if the thermal features are different. For example, the thermal camera can spot a difference in temperature between the skin of the face and a printed photo, which indicates that the photo is not the real face.

We have developed a new Fusion Patch Embedding layer explained in Fig.3.2, that takes input from an RGB camera as well as an IR-Thermal camera and performs the fusion of inputs from these two different camera sensors by first splitting the images into patches based on the patch size configuration and flattening them by projection into one single embedding. The embedding output of this layer goes as the input to the Vision Transformer.

Therefore, the new Fusion Patch Embedding layer plays an important role in solving real-world challenges related to face identification and providing security against common but highly effective physical attacks.

Figure 3.2 shows the architecture of Fusion Patch Embedding Layer

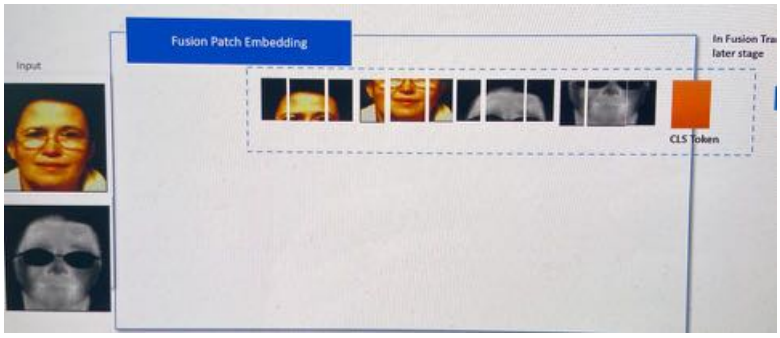


Figure 3.2: Fusion Patch Embedding Layer

3.2.2 Training Process

As with the proposed Fusion Patch Embedding layer discussed in /ref[Multi-Sensor Fusion], the rest of the Vision Transformer implementation is similar to that described in [9]. The

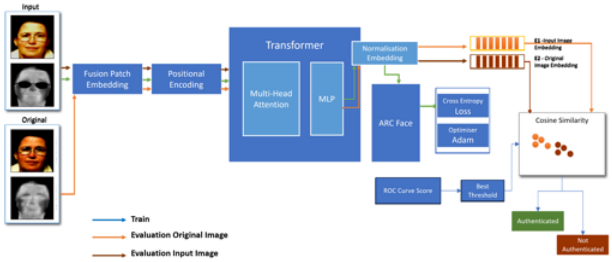


Figure 3.3: Overall Architecture

user face embedding captured and processed by the Fusion Patch Embedding module goes as input to the Vision Transformer for processing through all the sub-components of the Encoder such as Positional Encoding, Multi-Head Attention Layer, followed by the Last MLP(Multi-Layer Perceptron) layer of the Encoder. Unlike a classifier Vision Transformer, the output of the Encoder doesn't map to the number of output classes but uses output embedding of the Encoder to calculate ArcFace loss (explained in **??**) and an optimiser is applied. With ArcFace, the model separates inter-class embeddings and makes intra-class embeddings closer together. Even in scenarios with large variations in pose, lighting, and facial expressions, the ArcFace approach improves face recognition accuracy.

3.2.3 The Authentication Process

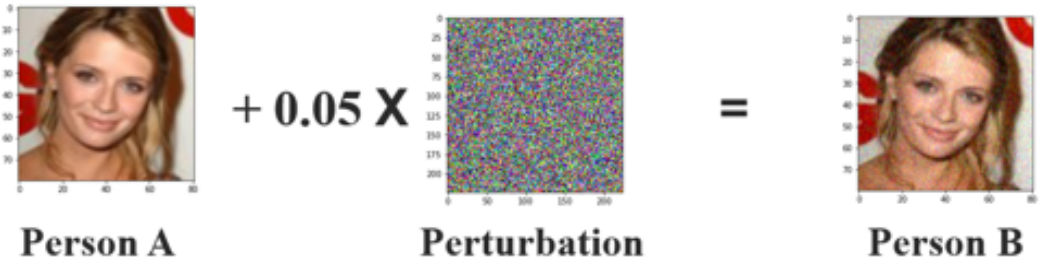
The authentication process uses the already stored embedding of the registered user to validate with the input of the user Fig **??**. The first step of authentication is to capture the image of the user through both RGB and IR-thermal cameras. The captured images then go through the Fusion Patch Embedding Layer, as explained in the training section 3.2.2. As with the training process, the embedding from the Fusion Patch Embedding Layer is then processed through the Vision Transformer. A cosine-similarity check is performed on the output embedding produced by the encoder as opposed to the training phase of the process. In this study, we use a custom calculation module that calculates the best threshold for identifying the person as similar or different while performing a cosine similarity analysis between the two embeddings.

Figure 3.3 shows the authentication flow.

3.2.4 Attack Process

If an adversary is able to influence the captured user information during the Face Identification process, they may launch an attack. As an example, perturbation generation using an attack algorithm can be used in such an attack.

Figure 3.2.4 shows the use of perturbation. With the objective of influencing the face recognition results of the system, a relatively small perturbation is introduced to the captured user input.



Adversarial Samples

3.2.5 Flat-Image Input

Other than the attack scenario explained in Section 3.2.4, there is a possibility of an adversarial attack on the physical level. In flat-image input, a user image (printed or displayed on a screen) is presented by an adversary to the Robot's Face Identification Module in order to pass the authentication process. This attack is simple to conduct and can break advanced face recognition systems which rely on common single-camera input backed by CNN models. Our proposed system mitigates this issue as explained in detail in section 3.2.1

3.2.6 FGSM - Fast Gradient Sign Method attack

For adversarial training, we use the Fast Gradient Sign Method (FGSM) [12], incorporating adversarial samples into the model during training in order to enhance its adaptability. Good Fellow et al. [12] to improve the robustness of the model against adversarial samples. In section **??** We have highlighted the performed experiments and outcomes in section 4.4

3.3 Summary

This chapter 3 introduced the proposed work, its different modules and phases. Section 3.1 presents the high-level design and working of the User Face Identification Module along with different adversaries considered in the proposed system. Section 3.2 explains details about methodologies used such as Fusion Embedding 3.2.1 and proposed Vision Transformer architecture to carry out the Face Identification. In the next chapter, we will explain the implementation of our proposed system and present the results of the experiments.

Chapter 4

Implementation and Evaluation

In this chapter, we discuss the training 4.3, evaluation4.2, and testing of the model 4.3, including experiments 4.4 conducted to validate the hypothesis of our proposed system. The chapter is divided into four sections: Training, Performance, Testing, and Adversarial Experiments.

4.1 Training - on Multi-Sensor Fusion Input Embedding

The data used in the thesis is discussed in section 2.4. Out of the dataset of 113 people, we used 90 people to train our model. We use an RGB-IR combo dataset, where every person has RGB images as well as an IR-thermal image. The dataset also contains images of users with different emotions, lights and with and without glasses. The aim of training with such a dataset is to train the recognition model, to work in real-world scenarios with large variations in pose, lighting, and facial expressions. The training using newly designed Fusion Patch Embedding is explained in the Multi-Sensor Fusion Module section 3.2.1. Since the dataset is not general, where one image will be mapped with one label, we have customised several stages in the whole workflow of the system such as Patch Embedding 3.2.1, calculation of Validation Loss and ROC curve calculation (explained in section 4.2) to find the best threshold. We use the following parameters to train our model:

Table 4.1: Training Hyper-Parameters

Epochs	Batch Size	Learning Rate	ArcFace s	ArcFace m	Normalization
25	16	0.001	0.3	64.0	(0.5,0.5,0.5)

Depending on the quality of data in the batch size, the validation loss varies, for example in a few samples of users the RGB and IR-Thermal images are not exactly the same. We used minimum validation loss across the epochs iteration to save the model with the best evaluation (discussed in 4.2).

4.2 Model Performance and Evaluation

We have used several different parameters to validate the performance of the model.

- **ROC Curve** calculation to find the best threshold, we have implemented a custom ROC Curve calculation Since we are not using classification, instead of using "y_prediction", we have calculated the distances of predictions with original labels, 0 for not similar and 1 for similar. The best threshold is calculated from the thresholds obtained from the ROC Curve. The selected best threshold is used to calculate the cosine similarity of embeddings to decide if the person is the same or not.

  ROC curve to calculate the best threshold 4.1

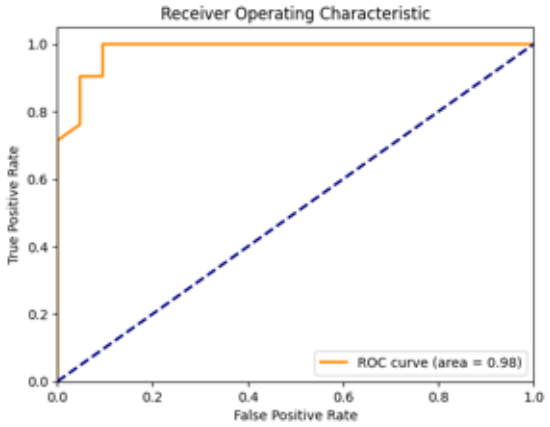


Figure 4.1: ROC curve to calculate the best threshold

- **True Positive and True Negative** We calculated the positive for an input if the cosine-similarity is similar. To calculate TP and TN, we use the following formula:

$$\text{TP} = \frac{\textbf{same personś samples identified as same user}}{\textbf{Total Samples}}$$

$$\text{TN} = \frac{\textbf{different personś samples identified as different user}}{\textbf{Total Samples}}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Model Evaluation Based on Validation Loss

In order to evaluate the performance of the model, Validation loss is used. The best-trained model is selected based on the lowest validation loss value. Our formula for calculating the Validation Loss is as follows.

4.2

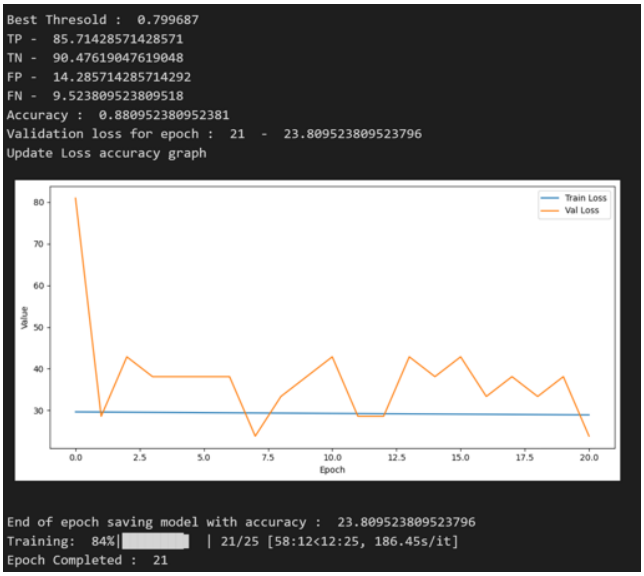


Figure 4.2:

$$\text{Validation Loss} = TP + TN - \text{TP} + \text{TN}$$

- **Accuracy Metrics using different optimisers and K-Fold Validation**. To avoid over-fitting the model, we used K-Fold Cross Validation; since our dataset is small, we performed cross-validation with three different folds of Train-Test data. Using the ADAM and ADAMW optimisers, we trained models and evaluated their performance.

Table 4.2: Evaluation Results with different Optimisers and K-Folds Cross Validation

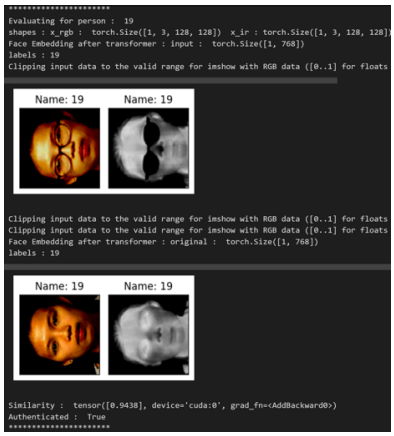
Optimiser	K-Fold	TP	TN m	Accuracy
ADAM	1	90.47	90.47	90.47.0
ADAM	2	76.19	95.23	85.71
ADAM	3	61.90	80.95	71.42
ADAMW	1	85.71	90.47	92.85
ADAMW	2	90.47	95.23	92.85
ADAMW	3	71.42	71.42	71.42

In general, results with Optimiser ADAM and ADAMW were similar, but ADAMW generally performed better due to its characteristic of decoupling weight decay from moving averages, resulting in better training stability and performance. The accuracy varied in different K-Folds because of the quality of the data samples.

4.3 Testing

We performed our testing with a test dataset of 21 people out of 113 samples. In our testing, we used data samples in different real-world scenarios like low light, darkness, and User wearing glasses. The goal of our proposed model is authentication and we considered True Negative as an important metric for the evaluation of our model. We have highlighted a few important results to showcase models working in different scenarios:

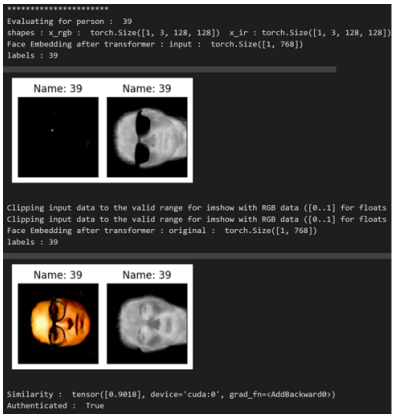
- **Authenticating Positive for the same user, with glasses in input image**



4.3

Figure 4.3: Authentication of a person wearing glasses

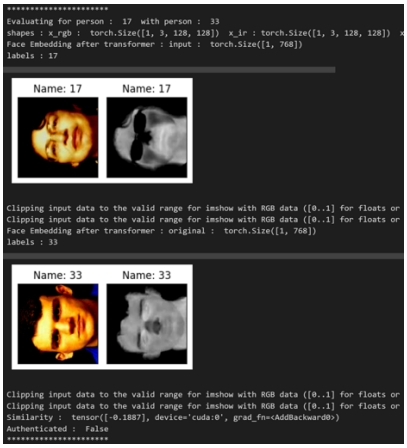
- **Authenticating Positive for the same user, with input in a dark environment**



4.4

Figure 4.4: Authentication of a person in darkness

- **Authenticating Negative for different users**



4.5

Figure 4.5: Authentication of a person in darkness

Our proposed model has been tested in a variety of scenarios, resulting in some interesting authentication results. A person wearing glasses can be identified by the model even if they were not wearing them at the time of registration. With a cosine similarity value of above 0.9, the model is able to identify the user in darkness when the RGB input is almost dark. If a different or unauthorised individual attempts to authenticate, the model produces a negative cosine similarity value.

4.4 Adversarial Experiment - FGSM Attack

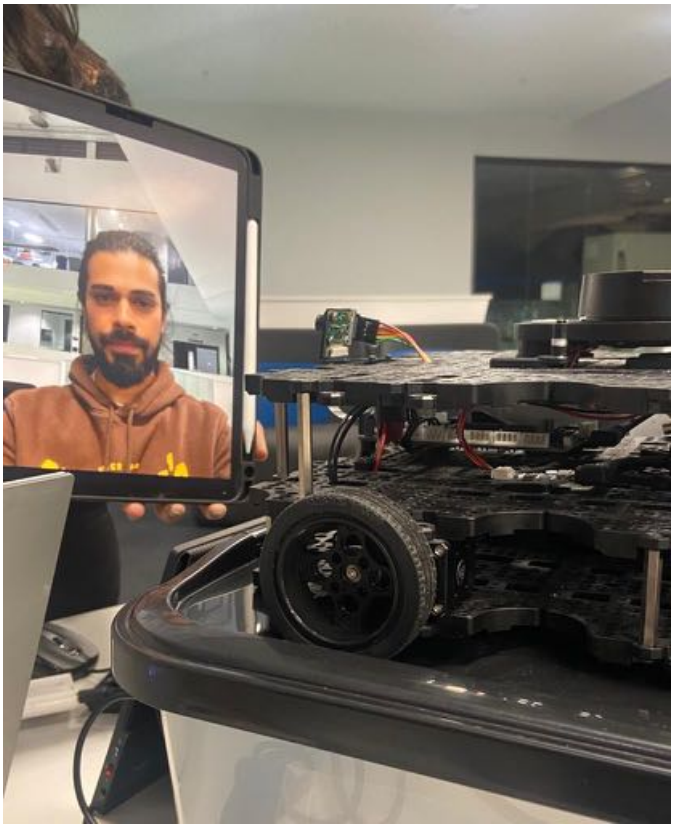
We conducted this experiment in order to evaluate the robustness of the Face Recognition Vision Transformer model against the adversary sample discussed in 3.2.4. We used Clever Han's library [22] to evaluate the performance of the model with input modified with added perturbation using FGSM attack. We also trained the model to identify the malicious input and evaluated the performance. Table 3.3 shows the results of this experiment.

Table 4.3: Performance of Model with FGSM attack

Model	FSGM Input	Adversary Trained	Accuracy
Original Model	False	False	90.0
Original Model	True	False	08.67
Original Model	True	True	75.20

4.5 Adversarial Experiment - Flat-Image Input Attack

This experiment evaluates the proposed system against an effective physical adversary discussed in 2.3.3. In section 3.2.1, we have highlighted how our model should mitigate the issue of Flat-Image Input attacks. We evaluated our model on the implemented Robot (discussed in **??**) by placing a picture of the user on an iPad screen. The model successfully restricted the authentication due to a lack of IR-Thermal facial features on the picture on the iPad screen. Fig 4.6 shows the experiment.



??

Figure 4.6: Flat-Image Input attack on the robot

4.6 Summary

In this chapter, we discussed the implementation details of the training of the proposed model (see Section 4.1), performance, and evaluation (see Section 4.2), where we highlighted all the results of the model's performance using various parameters. Section 4.3 showcased the results to validate all the aims and contributions of this thesis. Finally, we discussed the experiments (see Sections 4.4 and 4.5) involving different adversaries and their impact on our proposed system. In the next chapter, we will showcase the implementation of our approach on a real-world Robot platform.

Chapter 5

Multi-Sensor Fusion Face Recognition on Robot

This chapter shows the implementation details of the proposed model on a real-world Robot Platform 5.1. Multi-Sensor Fusion User Interaction Module is explained in section 5.2. Section 5.3 registration process of the user.Authentication of a user is discussed in section **??**

5.1 AI-powered secure Robot

The complete implementation of the proposed system on the Robot platform consists of the following modules: TurtleBot3 Robot, Multi-Sensor Fusion User Interaction Module, Server, Machine Learning Model - Fusion Vision Transformer.

5.1

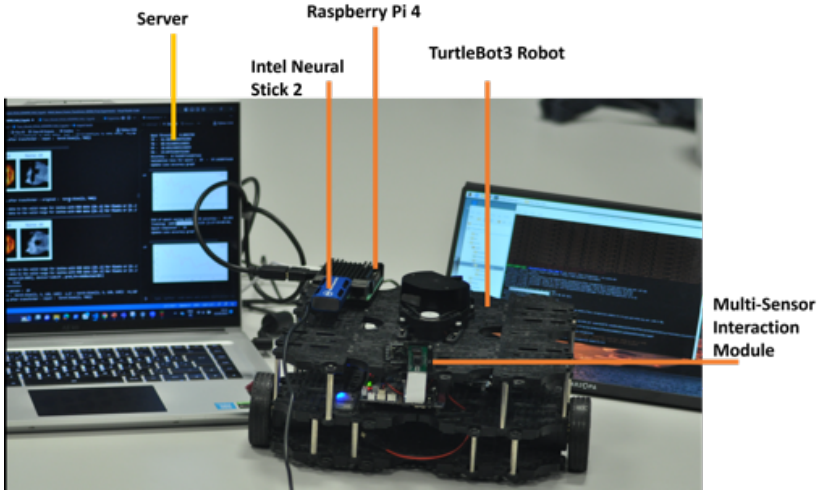


Figure 5.1: Overall System Components

Figure 5.1 shows all the components of the proposed system. The following table provides

details about each component.

Table 5.1: Details of overall system components

Component	***Device***	***Version***
Robot Plarform	TurtleBot	3.0
Server	Gigabyte Aero 16	Windows 11 Pro
Multi-Sensor Interaction Module	Rasberry Pi Camera Module	2.0
Multi-Sensor Interaction Module	IR-Thermal Camera	MLX90640
ML Model	Raspberry Pi	4
AI Inference Dev Kit	Intel Neural Stick Pi	3
Python	Server	3.9.6
PyTorch	Server	1.12.0
Python	Raspberry Pi	3.9.6
PyTorch	Raspberry Pi	1.12.0

5.2 Multi-Sensor Fusion Module

Figure 5.2 represents components of the Multi-Sensor User Interaction Module. This module is the interface between Robot and the user.

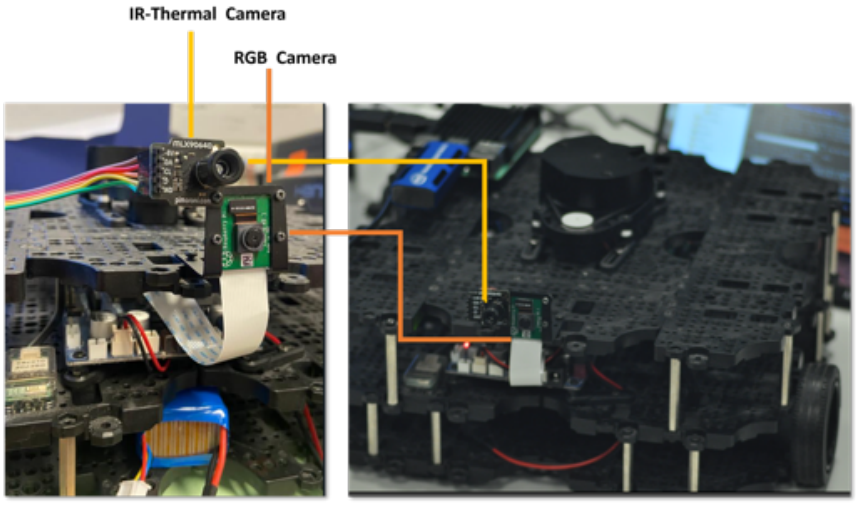


Figure 5.2: Multi-Sensor Fusion User Interaction Module

The functionality of the Multi-Sensor module is explained in section 3.2.1

5.3 Registration and Identification Process

Figure 5.3 shows the registration process of the new User. The robot takes the input and stores the embedding after processing it through the Fusion Vision Transformer model in the system. This stored embedding is used to identify and authenticate the user in the authentication process (see section 3.2.3).

Figure 5.3: Registration and Identification Process on the Robot

5.4 Summary

Our implementation on a real-world platform is presented in this chapter. We provided a brief description of the overall system components, including the version and the device details, in Section 5.1. Compared to last year's research, we have made several hardware improvements. A thermal camera has been added to the User Interaction Module as part of these updates. The proposed system now utilizes a Raspberry Pi 4, replacing the Raspberry Pi 3 used in the previous study.

An overview of this thesis' discussions will be presented in the next chapter.

Chapter 6

Discussion

This chapter compares the proposed approach with previous works. We evaluate the proposed system with 6 major parameters.

6.1 Comparison with Previous Work on the Robot

Comparisons with previous works are illustrated in Table 5.1.

Table 6.1: Comparison With Previous Works

Schemes	User privacy	Multi Image Sensor Authentication	Model Efficiency in Unfavourable Conditions	Secured against Flat-Image Input Attack	Seamless Authentication
Yang [37]	Yes	No	Low	No	No
Wang [35]	Yes	No	Low	No	No
Proposed	Yes	No	High	Yes	Yes

6.1.1 User privacy

As explained in the workflow of the proposed system, in Section 3.1.1, the registration process stores only the user's embedding as a reference for calculating cosine similarity during authentication. Additionally, during the authentication process, the Multi-Sensor Fusion layer converts input images into embeddings. As a result, the robot does not store an actual image of the user, ensuring user privacy.

6.1.2 Multi Image Sensor Authentication

Unlike previous related works, our proposed approach utilizes a Dual Image sensor to capture both an RGB and an IR-Thermal image (as explained in section 3.2.1) of the user's face.

The Multi-Sensor module provides dual-layer facial authentication, enhancing the robot's resilience against various adversaries.

6.1.3 Model Efficiency in Unfavourable Conditions

As discussed in section 3.2.1, this approach addresses real-world challenges associated with face identification, such as identification in low light or darkness.

6.1.4 Secured against Flat-Image Input Attack

This simple physical attack on the single-sensor face-recognition models may identify an adversary as the real user. This vulnerability poses a significantly bigger threat to face recognition systems. As explained in section 3.2.1 and experiment 4.5 shows the resilience of our proposed system against Flat-Image Input attack.

6.1.5 Seamless Authentication

Unlike previous works our proposed system relies on minimal input from the user to provide a more robust face identification solution in various difficult real-world scenarios.

Chapter 7

Conclusion and Future Research

7.1 Conclusion

We have proposed a new model based on the Vision Transformer, featuring an innovative multi-sensor fusion embedding module (section 3.2) to authenticate users. We introduced a new Patch Embedding module that combines user images from two sensors (RGB and IR) to create a single embedding as input for the Vision Transformer (section 3.2.1. This approach significantly enhances the robot's user identification capabilities in real-world unfavourable scenarios. Furthermore, it adds an additional layer of security to guard against attacks such as FGSM-based data corruption methods and Flat-Image input physical spoofing. Additionally, we conducted K-Folds cross-validation to prevent over-fitting of the proposed Transformer Model (see experiment 4.2). We evaluated performance using various optimization techniques and customized ROC (Receiver Operating Characteristic) curves to determine the optimal threshold. 4 presents a performance analysis, demonstrating the enhanced security compared to other available systems. To conclude, our proposed approach performs well with less training data set, solves face identification challenges in real-world unfavourable scenarios such as low light or darkness, and provides robust security against physical adversarial attacks. Hence satisfies all the aims and contributions, mentioned in section 1.2.

7.2 Future Research

Future work on our proposed system can be divided into various categories:

- **Dataset** We have used a specialized RGB-IR combo dataset for training and evaluating our model. Although the dataset was smaller, the model performed well. The robot's multi-sensor module will be able to generate more real-world data in the future. We can expect a significant improvement in the accuracy and efficiency of the face recognition mechanism if the model is trained with more real data.
- **Depth Sensor** To provide more robustness to our proposed face identification system, a Depth Camera Sensor can be introduced to the Multi-Sensor Fusion module. With

more research and enhancement to the Fusion Vision Transformer, the authentication module will be able to identify more essential features of the user's face and the system will be resilient to evolving AI-enabled attacks.

- **Application of the Robot** This thesis focuses on the authentication of the user by face identification. Applications such as giving food packets to the user after successful authentication can be researched and implemented. More interaction modules can be added, for example robot can be equipped with a Thermometer and a Pulse Analysis device to perform the basic task of taking the vitals of the user and performing registration of the new user in the system with relevant information.

- **Test on more types of Robots** The proposed system requires certain 5.1 hardware to run our model. The complete system can be deployed on various robots and drones to perform identification of the user in various applications.

Bibliography

[1] Ahuja, K., Chandra, V., Lord, V., and Peens, C. Ordering in: The rapid evolution of food delivery. *McKinsey & Company 22* (2021).

[2] Amazon. meet-scout, 2022. https://www.aboutamazon.com/news/transportation/meet-scout, Last accessed on 2022-09-11.

[3] Barsoum, E., Zhang, C., Ferrer, C. C., and Zhang, Z. Training deep networks for facial expression recognition with crowd-sourced label distribution. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (New York, NY, USA, 2016), ICMI '16, Association for Computing Machinery, p. 279–283.

[4] Bradley, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition 30*, 7 (1997), 1145–1159.

[5] Chew, C. M. Caregiver shortage reaches critical stage. *Provider (Washington, DC) 43*, 5 (2017), 14–28.

[6] Costa, A., Martinez-Martin, E., Cazorla, M., and Julian, V. Pharos—physical assistant robot system. *Sensors 18*, 8 (2018), 2633.

[7] Das, S., Wei, Z., and Ravuri, V. Safety and operations of automated delivery vehicles: A scoping review.

[8] Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., and Zafeiriou, S. ArcFace: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 44*, 10 (oct 2022), 5962–5979.

[9] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

[10] Fawcett, T. Introduction to roc analysis. *Pattern Recognition Letters 27* (06 2006), 861–874.

[11] Global Industry Analysts, I. Last mile delivery - global strategic business report, 2023. https://www.researchandmarkets.com/reports/4845817/last-mile-delivery-global-strategic-business/, Last accessed on 2023-09-06.

[12] Goodfellow, I. J., Shlens, J., and Szegedy, C. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

[13] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., and Xu, C. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 1580–1589.

[14] Hayley Robinson, B. M. . E. B. The role of healthcare robots for older people at home: A review. *International Journal of Social Robotics 6* (January 2014), 575–591.

[15] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

[16] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision* (2019), pp. 1314–1324.

[17] Li, Z., Liu, Y., Li, L., and Hong, Q. Additive phoneme-aware margin softmax loss for language recognition, 2021.

[18] Lund, S., Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., Meaney, M., and Robinson, O. The future of work after covid-19. *McKinsey global institute 18* (2021).

[19] Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)* (2018), pp. 116–131.

[20] Martinez-Martin, E., Costa, A., and Cazorla, M. Pharos 2.0—a physical assistant robot system improved. *Sensors 19*, 20 (2019), 4531.

[21] Pandey, A. K., and Gelin, R. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics Automation Magazine 25*, 3 (2018), 40–48.

[22] Papernot, N., Faghri, F., Carlini, N., Goodfellow, I., Feinman, R., Kurakin, A., Xie, C., Sharma, Y., Brown, T., Roy, A., Matyasko, A., Behzadan, V., Hambardzumyan, K., Zhang, Z., Juang, Y.-L., Li, Z., Sheatsley, R., Garg, A., Uesato, J., Gierke, W., Dong, Y., Berthelot, D., Hendricks, P., Rauber, J., and Long, R. Technical report on the cleverhans v2.1.0 adversarial examples library. *arXiv preprint arXiv:1610.00768* (2018).

[23] Poghosyan, L., Clarke, S. P., Finlayson, M., and Aiken, L. H. Nurse burnout and quality of care: Cross-national investigation in six countries. *Research in nursing & health 33*, 4 (2010), 288–298.

[24] Rao, D., Wu, X.-J., and Xu, T. Tgfuse: An infrared and visible image fusion approach based on transformer and generative adversarial network, 2022.

[25] Rojas Viloria, D., Solano-Charris, E. L., Muñoz-Villamizar, A., and Montoya-Torres, J. R. Unmanned aerial vehicles/drones in vehicle routing problems: a literature review. *International Transactions in Operational Research 28*, 4 (2021), 1626–1657.

[26] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 4510–4520.

[27] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[28] Srinivas, S., Ramachandiran, S., and Rajendran, S. Autonomous robot-driven deliveries: A review of recent developments and future directions. *Transportation Research Part E: Logistics and Transportation Review 165* (2022), 102834.

[29] Starship. Starship robot, 2022. https://www.starship.xyz/, Last accessed on 2022-09-11.

[30] Sykes, P. Last mile delivery costs — challenges and solutions, 2023. https://blog.routific.com/blog/last-mile-delivery-costs/, Last accessed on 2023-09-02.

[31] TUFTUniversity. $\mathrm{Td}_i r_r gbfaceimages.$, Last accessed on 2022-09-11.

United Nations, Department of Economic and Social Affairs, Population Division. *World Population Ageing 2019*. ST/ESA/SER.A/444. United Nations, 2020. https://digitallibrary.un.org/record/3898412/files/undesa_pd-2020_world_population_ageing_highlights.pdf, Last accessed on 2023-09-02.

Vandemeulebroucke, T., Dierckx de Casterlé, B., and Gastmans, C. The use of care robots in aged care: A systematic review of argument-based ethics literature. *Arch Gerontol Geriatr 74* (January 2018), 15–25.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need, 2023.

Wang, W., Gope, P., and Cheng, Y. An ai-driven secure and intelligent robotic delivery system. *IEEE Transactions on Engineering Management* (2022), 1–16.

Yang, J., Gope, P., Cheng, Y., and Sun, L. Design, analysis and implementation of a smart next generation secure shipping infrastructure using autonomous robot. *Computer Networks 187* (2021), 107779.

Yang, J., Gope, P., Cheng, Y., and Sun, L. Design, analysis and implementation of a smart next generation secure shipping infrastructure using autonomous robot. *Computer Networks 187* (2021), 107779.

Zhong, Y., and Deng, W. Face transformer for recognition, 2021.

Appendices

Appendix A

Code

The code can be found on under the link - https://drive.google.com/file/d/1mrYdk8EueNKDSVpWVLDAcL0nj

The project is divided into modular .py files. The directory Experiment_Result/Evals has testing and evaluation results in .ipynb files.