CHAPTER 1

INTRODUCTION

Introduction to Project: Ear Recognition using Deep Learning

The "Ear Recognition using Deep Learning " project aims to explore and implement a robust system for identifying and recognizing individuals based on the unique features of their ears. This project leverages the powerful deep learning framework Keras and the pre-trained ResNet50 model to develop a sophisticated ear recognition system.

Ear recognition is an emerging biometric technology that utilizes the distinctive characteristics of the ear, such as its shape, size, and contours, to establish a person's identity. Unlike facial recognition, which can be affected by changes in facial expressions, aging, and cosmetic alterations, ear features remain relatively stable throughout a person's life. This stability makes ear recognition a reliable method for long-term identification.

In this project, the ResNet50 model, a deep convolutional neural network (CNN) renowned for its robust feature extraction capabilities, is utilized to extract high-dimensional feature vectors from ear images. ResNet50, short for Residual Networks, is a highly effective CNN architecture. It consists of 50 layers and is designed to address the vanishing gradient problem, enabling the training of very deep networks. By using residual learning, ResNet50 can maintain high accuracy even as the network depth increases.

The process begins with preprocessing ear images, which involves resizing and normalizing the images to prepare them for analysis. The preprocessed images are then passed through the ResNet50 model, which has been modified to remove the top layers and add custom layers suitable for feature extraction specific to ear recognition. This results in the extraction of distinctive feature vectors that capture the intricate details of each ear image.

To determine the similarity between the input ear image and the precomputed dataset of ear images, the system employs cosine similarity. This mathematical technique measures the cosine of the angle between two non-zero vectors, providing a measure of similarity that is not affected by the magnitude of the vectors. By calculating the cosine similarity between the input feature vector and the feature vectors in the dataset, the system identifies the most similar ear images.

The "Ear Recognition using Deep Learning" project showcases the integration of deep learning and biometric recognition, highlighting the power of advanced neural network models in solving real-world problems. By leveraging the stability and uniqueness of ear features, the system aims to provide a reliable method for identifying individuals. This technology has potential applications in various fields, including security systems, forensic investigations, and personal identification.

In summary, the project demonstrates the practical application of ResNet50 in ear recognition, emphasizing the model's ability to extract meaningful features from images. The use of Keras simplifies the development process, making it accessible for researchers and developers to build sophisticated recognition systems. The stability and uniqueness of ear features make ear recognition a promising biometric modality, contributing to the advancement of secure and efficient identification technologies.

Motivation

The motivation behind the "Ear Recognition using Deep Learning " project stems from the growing need for advanced biometric identification systems that are both reliable and user-friendly. Traditional biometric methods, such as fingerprint and facial recognition, have their limitations. Fingerprints can be affected by wear and tear or environmental conditions, while facial recognition can be influenced by changes in facial expressions, aging, makeup, and other factors. In contrast, ear recognition offers a unique set of advantages that make it an attractive alternative.

Stability and Uniqueness

Ears possess unique morphological characteristics that remain relatively stable over a person's lifetime. Unlike facial features that can change due to various factors, the shape and structure of the ear are less susceptible to changes, making it a reliable biometric trait. This inherent stability makes ear recognition particularly useful for long-term identification and verification.

Non-Intrusiveness

Ear recognition is a non-intrusive biometric method. It does not require direct contact with the sensor, unlike fingerprint recognition, making it more hygienic and user-friendly. Additionally, ear images can be captured from a distance without the need for cooperative behavior from the individual, enhancing the practicality and convenience of this technology.

Enhanced Security

In an era where security is paramount, ear recognition provides an additional layer of security. Combining ear recognition with other biometric methods can create a multi-factor authentication system that is more robust against spoofing and fraudulent activities. This added security is essential for applications in banking, border control, and access management, where ensuring the identity of individuals is critical.

Technological Advancements

The rapid advancements in deep learning and neural networks have paved the way for more accurate and efficient biometric recognition systems. By utilizing powerful models like ResNet50 within the Keras framework, the project leverages state-of-the-art technology to develop a sophisticated ear recognition system. The use of pre-trained models accelerates the development process and enhances the accuracy of the system, making it feasible to deploy in real-world scenarios.

Applications in Forensics and Law Enforcement

Ear recognition technology has significant potential in forensics and law enforcement. It can be used to identify individuals in surveillance footage or from partial images, aiding in criminal investigations and enhancing public safety. The unique characteristics of ear recognition can help law enforcement agencies to track and identify suspects with a high degree of accuracy.

Contribution to Research

This project also contributes to the growing body of research in the field of biometrics. By exploring and implementing ear recognition using advanced deep learning techniques, the project aims to provide valuable insights and pave the way for further developments in biometric identification systems. The findings and methodologies from this project can inspire future research and innovations, advancing the field of biometric technology.

In conclusion, the "Ear Recognition using Deep Learning" project is motivated by the need for reliable, non-intrusive, and advanced biometric identification systems. The unique advantages of ear recognition, combined with the power of deep learning, make this project a significant step towards more secure and user-friendly biometric solutions.

CHAPTER 2

LITERATURE SURVEY

Title : A deep learning approach for person identification using ear biometrics
Authors : Ramar Ahila Priyadharshini, SelvarajArivazhagan, and Madakannu Arun
Published year : 2020 Oct
Reference number:1

Abstract :

Human ear recognition has been an active area of research since the mid-1990s. Initial efforts by Burge and Burger in 1996 utilized adjacency graphs derived from Voronoi diagrams of ear curve segments. Subsequent advancements, including the first fully automated ear recognition procedure by Moreno et al., have employed various techniques categorized into geometric, holistic, local, and hybrid approaches. Geometric approaches focus on the ear's structural attributes, while holistic methods use global properties of the ear image, including force field transforms and subspace projection techniques such as PCA and LDA. Local approaches leverage texture information from key points, utilizing features like SIFT and LBP. Hybrid approaches combine multiple methods to enhance recognition performance.

Recent trends in ear recognition have shifted towards deep learning due to its capacity for automatic feature extraction from raw data. Pioneering deep networks like LeNet have paved the way for sophisticated applications in various domains. In the context of ear recognition, models like AlexNet and VGGNet have been applied to datasets such as AWE and CVLE with promising results. This research focuses on recognizing individuals based on ear profiles using deep learning techniques, specifically employing the IITD-II dataset to evaluate the performance of a newly designed deep neural network. This study uniquely explores the impact of network parameters—such as learning rate, kernel size, and activation functions—on ear recognition accuracy. Additionally, the efficacy of the proposed deep neural network is validated on the AMI dataset, demonstrating its potential for advancing ear recognition technology.

Conclusion :

In this research, we have explored the development and implementation of a deep learning-based ear recognition system using convolutional neural networks (CNNs). The subsampling layer, specifically Max Subsampling, was employed to reduce the learnable parameters and mitigate overfitting, thereby enhancing the overall performance and accuracy of the network. Additionally, batch normalization layers were incorporated to normalize input channels across mini-batches, expediting the training process and reducing sensitivity to network initialization.

Fully connected layers were utilized to connect neurons across layers, allowing for the final classification output. This comprehensive approach, involving convolution, subsampling, and normalization, facilitated the extraction of robust features and reduction of parameters from the original ear images, leading to accurate predictions.

The experimental analysis on the IITD-II dataset demonstrated the efficacy of the designed deep neural network in recognizing ear profiles. Furthermore, the study of various network parameters, such as learning rate, kernel size, and activation functions, provided insights into optimizing the performance of the ear recognition system. Validation on the AMI dataset confirmed the potential efficiency and robustness of the proposed deep learning model.

In conclusion, the integration of CNNs for ear recognition shows promising advancements in biometric technology. This research contributes to the field by providing a practical and effective solution for ear-based authentication, paving the way for further developments and applications in areas such as security, healthcare, and forensic investigations.

Title : Human Recognition using Ear based Deep Learning Features
Authors : Haider Mehraj, Ajaz Hussain Mir
Year of publication : March ,2020.
Reference Number: 2

Abstract

Identification through biometrics has long been a significant area of research, with the human ear emerging as a unique and stable biometric feature. Traditional ear recognition techniques have employed methods such as Shift Invariant Feature Transform (SIFT), Speed Up Robust Features (SURF), and Binary Robust Invariant Scalable Keypoints (BRISK). However, the advent of deep learning, particularly convolutional neural networks (CNNs), has revolutionized computer vision, enabling more effective ear recognition despite the limited availability of large datasets. This paper addresses the challenge of training CNNs with sparse ear image databases by implementing an efficient CNN-based model using InceptionV3 for ear recognition.

InceptionV3, pre-trained on the extensive ImageNet database, is adapted for this purpose. To compensate for the limited data, we employ various augmentation techniques, such as random rotations and noise addition, to artificially expand the dataset. The pre-trained InceptionV3 model is fine-tuned, discarding the last three classification layers and using the activations from the network's final layers as feature vectors. These feature vectors are then reduced in dimensionality using principal component analysis (PCA) and classified using Linear Support Vector Machines (SVM) due to their robustness and computational efficiency.

Our approach yields a recognition rate of 98.10% using 5-fold cross-validation, demonstrating the effectiveness of combining deep learning with traditional augmentation and classification

techniques for ear recognition. This hybrid CNN-SVM model offers a promising solution for automatic ear recognition, with potential applications in security and surveillance systems.

The paper is structured as follows: Section II provides an overview of existing state-of-the-art methods in ear recognition. Section III details the methodology adopted in this study. Section IV discusses the results obtained using the proposed scheme. Finally, Section V concludes the paper.

Conclusion:

This paper presented a hybrid ear recognition system ofdeep features and classical machine learning algorithm forclassification. The inputs to last layer of feature extractionmodule in inceptionV3 were used as feature vectors and hadvery high dimensions of 2048. To make dimensions feasiblefor use with classifier, the dimensionality was reduced usingPCA and final dimensions of reduce feature set was 29principal components. Moreover, aggressive dataaugmentation was used to increase the number of images inoriginal dataset and ensure classifier does not over fit.

Title : An efficient ear recognition technique based on deep ensemble learning approach.
Authors :Ravishankar Mehta, Koushlendra kumar Singh
Year of publication : 04 May 2023.
Reference number : 3

Abstract :

Person identifications using the ear-based biometric system has become quite popular in recent year due to increasing demands in security and surveillance applications. With limited training data and computing resources, the run time complexity plays an important role in such a biometric system. With the continuous advancement in deep convolutional neural networks, deep learning-based biometric systems consequently achieved huge progress in solving earlier unanswered and/or incomplete challenges. Though ear-based biometric system gives higher accuracy with the help of pre-trained deep learning models like VGG19, VGG16, Xception, etc. Training these models is a cumbersome task and it requires much time. Most of the ear recognition system developed using deep learning models like VGG19, Xception, ResNet101, etc. requires a large memory area due to the huge parameter requirement of the model. Also, they put computational overhead on the system. One of the major challenges in the field of ear recognition is to identify people with the help of electronic devices over time and space. While developing electronic approaches for person identification, it is worth important to consider the factors like simplicity, cost-effectiveness, and portable flexibility. With these motivations, the authors developed three simple lightweight CNN models and ensemble them to get improved

recognition accuracy. The work is validated on the IITD-II ear dataset which contains only 793 sample mages for training purposes. To overcome the limitation of the limited dataset, the author performed data augmentation technique which produces a variety of images from different perspectives. By stacking these three CNN models, an optimal architecture is developed that gives the best accuracy of 98.74% which is a good improvement over the individual model. The proposed CNN models can also be ensemble with other pre-trained models like VGG16, VGG19, ResNet, Xception, etc. for a more effective solution.

Conclusion:

The paper "An efficient ear recognition technique based on deep ensemble learning approach" concludes that the proposed method significantly improves ear recognition accuracy by combining multiple deep learning models. The ensemble learning approach leverages the strengths of individual models, leading to better performance compared to single-model techniques. The research demonstrates the robustness and efficiency of the method, making it a viable solution for practical biometric applications. Future work will focus on further optimizing the system and exploring its application in diverse real-world scenarios.

Title : A Systematic Literature Review on Human Ear Biometrics: Approaches, Algorithms, and Trend in the Last Decade.
Authors :Oyediran George Oyebiyi, Adebayo Abayomi -Alli.
Year of publication : 17 March 2023.
Reference number : 4

Abstract :

Biometric technology is fast gaining pace as a veritable developmental tool. So far, biometric procedures have been predominantly used to ensure identity and ear recognition techniques continue to provide very robust research prospects. This paper proposes to identify and review present techniques for ear biometrics using certain parameters: machine learning methods, and procedures and provide directions for future research. Ten databases were accessed, including ACM, Wiley, IEEE, Springer, Emerald, Elsevier, Sage, MIT, Taylor & Francis, and Science Direct, and 1121 publications were retrieved. In order to obtain relevant materials, some articles were excused using certain criteria such as abstract eligibility, duplicity, and uncertainty (indeterminate method). As a result, 73 papers were selected for in-depth assessment and significance. A quantitative analysis was carried out on the identified works using search strategies: source, technique, datasets, status, and architecture. A Quantitative Analysis (QA) of feature extraction methods was carried out on the selected studies with a geometric approach indicating the highest value at 36%, followed by the local method at 27%. Several architectures, such as Convolutional Neural Network, restricted Boltzmann machine, auto-encoder, deep belief network, and other unspecified architectures, showed 38%, 28%, 21%, 5%, and 4%, respectively. Essentially, this survey also provides the various status of existing methods used in classifying related studies. A taxonomy of the current methodologies of ear recognition system was presented along with a publicly available occlussion and pose sensitive black ear image dataset of 970 images. The study concludes with the need for researchers to consider improvements in the speed and security of available feature extraction algorithms.

Conclusion:

The paper "A Systematic Literature Review on Human Ear Biometrics: Approaches, Algorithms, and Trend in the Last Decade" provides a comprehensive review of the advancements in ear biometrics over the past ten years. It highlights the significant contributions of deep learning techniques in improving ear recognition systems, comparing them with traditional machine learning methods. The study identified several key areas for future research, emphasizing the need for enhancements in the speed and security of feature extraction algorithms.

The review involved analyzing 1121 publications, narrowing them down to 73 significant studies that were assessed based on parameters like source, technique, datasets, status, and architecture. The paper presents a taxonomy of current methodologies and introduces a publicly available dataset of 970 occlusion and pose-sensitive ear images. The findings suggest that geometric approaches for feature extraction are the most valuable, followed by local methods. The most prominent architectures used include Convolutional Neural Networks, restricted Boltzmann machines, and auto-encoders.

Title : A comprehensive survey and deep learning-based approach for human recognition using ear biometric.
Authors :Aman Kamboj, Rajneesh Rani and Adithya Nigam.
Year of publication : 22 April 2021.
Reference Number : 5

Abstract :

Human recognition systems based on biometrics are much in demand due to increasing concerns of security and privacy. The human ear is unique and useful for recognition. It offers numerous advantages over popular biometrics traits face, iris, and fingerprints. A lot of work has been attributed to ear biometric, and the existing methods have achieved remarkable success over constrained databases. However, in unconstrained environment, a significant level of difficulty is observed as the images experience various challenges. In this paper, we first have provided a comprehensive survey on ear biometric using a novel taxonomy. The survey includes in-depth details of databases, performance evaluation parameters, and existing approaches. We have introduced a new database, NITJEW, for evaluation of unconstrained ear detection and recognition. A modified deep learning models Faster-RCNN and VGG-19 are used for ear detection and ear recognition tasks, respectively. The benchmark comparative assessment of our database is performed with six existing popular databases. Lastly, we have provided insight into open-ended research problems worth examining in the near future. We hope that our work will be a stepping stone for new researchers in ear biometrics and helpful for further development.

Conclusion:

The paper "A Comprehensive Survey and Deep Learning-Based Approach for Human Recognition Using Ear Biometric" offers an extensive review of ear biometric systems, presenting a new taxonomy and introducing the NITJEW database for unconstrained environments. It highlights the effectiveness of deep learning models, specifically Faster R-CNN for detection and VGG-19 for recognition, validated against popular databases. The study identifies key challenges and suggests future research directions, aiming to advance the field by addressing environmental variability in ear images.

Title : Ear Recognition Based on Deep Unsupervised Active Learning.
Authors :Yacine Khaldi, Amir Benzaoui, Abdeldjalil Ouahabi.
Year of publication : July 2021.
Reference Number :6

Abstract :

Cooperative machine learning has many applications, such as data annotation, where an initial model trained with partially labeled data is used to predict labels for unseen data continuously. Predicted labels with a low confidence value are manually revised to allow the model to be retrained with the predicted and revised data. In this paper, we propose an alternative to this approach: an initial training process called Deep Unsupervised Active Learning. Using the proposed training scheme, a classification model can incrementally acquire new knowledge during the testing phase without manual guidance or correction of decision making. The training process consists of two stages: the first stage of supervised training using a classification model, and an unsupervised active learning stage during the test phase. The labels predicted during the test phase, with high confidence, are continuously used to extend the knowledge base of the model. To optimize the proposed method, the model must have a high initial recognition rate. To this end, we exploited the Visual Geometric Group (VGG16) pre-trained model applied to three datasets: Mathematical Image Analysis (AMI), University of Science and Technology Beijing (USTB2), and Annotated Web Ears (AWE). This approach achieved impressive performance that shows a significant improvement in the recognition rate of the USTB2 dataset by coloring its images using a Generative Adversarial Network (GAN). The obtained performances are interesting compared to the current methods: the recognition rates are 100.00%, 98.33%, and 51.25% for the USTB2, AMI, and AWE datasets, respectively.

Conclusion:

Deep unsupervised active learning techniques show promise in enhancing ear recognition systems, offering improved accuracy and efficiency over traditional supervised methods. These approaches exhibit robustness against various image variations like pose changes, lighting conditions, and occlusions, contributing to their potential for real-world applications. Performance evaluations typically demonstrate the superiority of such systems in terms of recognition rates and computational efficiency. Future research directions may focus on refining these techniques, exploring different architectures, and addressing specific challenges in ear recognition, paving the way for more advanced and reliable biometric identification systems.

Authors :Hammam Alshazly, Christoph Linse, Erhardt Barth.
Title : Towards Explainable Ear Recognition Systems Using Deep Residual Networks.
Year of publication : 31 August 2021.
Reference Number :7

Abstract :

This paper presents ear recognition models constructed with Deep Residual Networks (ResNet) of various depths. Due to relatively limited amounts of ear images we propose three different transfer learning strategies to address the ear recognition problem. This is done either through utilizing the ResNet architectures as feature extractors or through employing end-to-end system designs. First, we use pretrained models trained on specific visual recognition tasks, inititalize the network weights and train the fully-connected layer on the ear recognition task. Second, we fine-tune entire pretrained models on the training part of each ear dataset. Third, we utilize the output of the penultimate layer of the fine-tuned ResNet models as feature extractors to feed SVM classifiers. Finally, we build ensembles of networks with various depths to enhance the overall system performance. Extensive experiments are conducted to evaluate the obtained models using ear images acquired under constrained and unconstrained imaging conditions from the AMI, AMIC, WPUT and AWE ear databases. The best performance is obtained by averaging ensembles of fine-tuned networks achieving recognition accuracy of 99.64%, 98.57%, 81.89%, and 67.25% on the AMI, AMIC, WPUT, and AWE databases, respectively. In order to facilitate the interpretation of the obtained results and explain the performance differences on each ear dataset we apply the powerful Guided Grad-CAM technique, which provides visual explanations to unravel the black-box nature of deep models. The provided visualizations highlight the most relevant and discriminative ear regions exploited by the models to differentiate between individuals. Based on our analysis of the localization maps and visualizations we argue that our models make correct prediction when considering the geometrical structure of the ear shape as a discriminative region even with a mild degree of head rotations and the presence of hair occlusion and accessories.

Conclusion:

This paper introduces ear recognition models based on five different variants of deep ResNet architectures. We proposed three methods of transfer learning, which can be used with other deep CNN architectures to learn discriminative ear features and to improve the overall recognition performance. Extensive experiments were conducted on four challenging publicly available ear image datasets, which consist of images collected under constrained and unconstrained conditions. In order to address the wide variability in ear images such as geometric transformations, occlusions, different image sizes and varying aspect ratios for each of the considered datasets, we proposed to embed each image into a fixed-size canvas to preserve the aspect ratios. Moreover, when training the models we introduced two different data augmentation pipelines to suit the type of variations in both, the constrained and unconstrained ear datasets. Our experimental results show considerable improvements in the recognition rates on all datasets and our proposed models achieve state-of-the-art recognition performance.

In order to make our models more transparent and to uncover the black-box nature of the deep models we applied a visualization technique that highlighted the important image regions responsible for the model predictions. The provided visualizations indicate that consistently focusing on the geometrical structure of the ear shape is the most discriminative region for getting correct predictions, whereas relying on auxiliary information such as the haircut or skin texture can result in wrong decisions. The visualizations also show the limited impact of partial ear occlusions and mild degrees of head rotations on performance, whereas, severe occlusion by hair and severe head rotations have a detrimental impact on recognition performance.

Title : Deep Learning for 3D Ear Detection: A Complete Pipeline From Data Generation to Segmentation.
Authors :Md.Mursalin, Syed Mohammed Shamsul Islam.
Year of publication : 19 November 2021.
Reference Number :8

Abstract :

The human ear has distinguishing features that can be used for identification. Automated ear detection from 3D profile face images plays a vital role in ear-based human recognition. This work proposes a complete pipeline including synthetic data generation and ground-truth data labeling for ear detection in 3D point clouds. The ear detection problem is formulated as a semantic part segmentation problem that detects the ear directly in 3D point clouds of profile face data. We introduce EarNet, a modified version of the PointNet++ architecture, and apply rotation augmentation to handle different pose variations in the real data. We demonstrate that PointNet and PointNet++ cannot manage the rotation of a given object without such augmentation. The synthetic 3D profile face data is generated using statistical shape models. In addition, an automatic tool has been developed and is made publicly available to create ground-truth labels of any 3D public data set that includes co-registered 2D images. The experimental results on the real data demonstrate higher localization as compared to existing state-of-the-art approaches.

Conclusion:

This work aims to detect ears directly on 3D point clouds of profile face data by applying a deep neural network named EarNet. A large set of synthetic profile face data was generated for training the proposed EarNet. Additionally, a novel approach is proposed to create ground-truth labels on real 3D data with corresponding co-registered 2D images. The experimental results demonstrate that our model performs significantly better than existing deep learning models for ear detection directly from 3D point clouds. A possible direction for future research is to incorporate the proposed ear detection model into an ear recognition pipeline. In addition, we aim to investigate different deep learning-based 2D segmentation networks for the ground-truth labeling pipeline.

Authors: Mostafa Ibrahim
Title : The Basics of ResNet50
Year of publication: Jan 16, 2024
Reference Number : 9

Abstract :

ResNet50 and VGG are both convolutional neural networks (CNNs) designed for image recognition tasks, each with unique architectural philosophies. VGG, introduced earlier, uses smaller filters with increased depth, allowing multiple smaller filters to achieve the same receptive field as a single large filter. This approach induces more non-linearity with increased depth. In contrast, ResNet50 is significantly deeper than VGG, which is crucial for capturing more complex and hierarchical features in images. ResNet50 addresses the vanishing gradient problem effectively through skip connections, enabling it to maintain high accuracy even as the number of layers increases. VGG, with its straightforward stack of convolutional layers, may suffer from performance degradation in deeper architectures due to this issue.

Additionally, ResNet50's skip connections facilitate smoother training and faster convergence, making it easier for the model to learn and update weights during training, contributing to more efficient training dynamics compared to VGG. The bottleneck architecture of ResNet50, employing 1x1 convolutions to reduce and then restore the dimensionality of the data, achieves better parameter efficiency. This design reduces computational costs while maintaining expressive power, making ResNet50 more efficient in terms of parameters than VGG. Therefore, ResNet50 is generally considered superior to VGG, particularly for tasks that benefit from deeper architectures and efficient training processes.

Conclusion :

In summary, the ResNet50 architecture demonstrates remarkable versatility across a broad spectrum of tasks, including image classification and object detection. Its deeper architecture excels in capturing complex, hierarchical features, addressing the vanishing gradient problem through skip connections that facilitate smoother training and faster convergence. This ensures high accuracy and efficient training dynamics, making ResNet50 particularly effective in scenarios where deeper models might otherwise compromise accuracy. The bottleneck design of ResNet50 further enhances parameter efficiency, reducing computational costs while maintaining expressive power.

Additionally, the availability of pre-trained ResNet50 models in both Keras and PyTorch libraries enhances its accessibility and ease of integration, allowing developers to quickly leverage its powerful capabilities without extensive training from scratch. This accessibility makes ResNet50 an excellent choice for achieving high-quality results in various deep learning applications. Overall, ResNet50's combination of depth, efficient training, and parameter efficiency renders it generally superior to VGG, particularly in tasks that benefit from these advanced features. Therefore, ResNet50 stands out as a highly effective and versatile model in the field of deep learning.

Authors :Md.Mursalin, Syed Mohammed Shamsul Islam.
Title : A Comprehensive Review of CNN-Based Human Recognition Using Ear Shape Images.
Year of publication : April 2024.
Reference Number : 10

Abstract :

Automatic person recognition using ear shape images is an active field of research within the biometric community. Similar to other biometric traits such as fingerprints, face and iris, ear also has numerous specific and unique features that aid in person identification. In this present worldwide pandemic of COVID-19 situation, most of the facial identification systems has almost failed due to the mask wearing scenario. The human ear is a perfect source of data for passive person identification as it does not involve the cooperativeness of the individual whom recognition is being attempted for and it is more easily captured at a distance. A human ear image acquisition is also easy as the ear is apparent even the users' wearing masks. In an automatic human recognition system, an ear biometric system can be integrated as a supplement to other biometric systems and offer identity cues when other system information is unreliable or even unavailable. In this paper, a comprehensive review of growing research field of feature extraction techniques and the classification technique of deep learning using convolutional neural network (CNN) was conducted to exhibit the rate of accuracy for person recognition systems using ear shape images.

Conclusion:

The paper concludes that Convolutional Neural Networks (CNNs) are highly effective in human recognition based on ear shape images. They excel in extracting discriminative features essential for accurate identification, surpassing conventional methods in performance metrics like accuracy, robustness to variations, and scalability. Ear shape images offer distinct advantages in biometric recognition, being unique, stable, and non-intrusive. However, challenges such as pose variations and occlusions remain, prompting further exploration for robust solutions. The comparative analysis underscores ear shape images' strengths against other biometric modalities, particularly in security, access control, and forensic applications. Looking ahead, future research should focus on innovating CNN architectures, integrating multi-modal data, and enhancing privacy measures to advance the field's capabilities and applicability in real-world scenarios.

CHAPTER 3

PROBLEM STATEMENT

The project "Person Recognition using Ear Images based on Deep Learning" addresses the need for advanced and reliable biometric identification systems by leveraging the unique features of ear images. Developing an effective person recognition system using ear images and the ResNet50 model involves several significant challenges.

The primary challenge is efficiently extracting discriminative features from ear images to accurately capture individuals' unique biometric traits. This requires handling the variability in ear images caused by different poses, lighting conditions, and accessories such as earrings or glasses. Another critical aspect is training and optimizing the deep learning model (ResNet50) to learn and classify these features with high precision and robustness. This involves designing an effective neural network architecture that can generalize well across diverse datasets and minimize errors.

Moreover, ensuring real-time processing capabilities in the recognition system is essential for prompt and reliable identification of individuals based on their ear biometrics. The system must be able to quickly process and compare ear images to facilitate practical applications such as security systems, access control, and forensic investigations.

Additionally, the project must address privacy and non-intrusiveness in data collection, ensuring that the biometric data is handled ethically and securely. This includes implementing measures to protect the data from unauthorized access and ensuring that the data collection process does not infringe on individuals' privacy.

Overall, the project aims to tackle these challenges to create a reliable and efficient person recognition system using ear images and advanced deep learning techniques. The goal is to develop a sophisticated solution that balances accuracy, efficiency, and ethical considerations, thereby contributing to the advancement of biometric technology.

Objectives:

- Develop a robust feature extraction mechanism using ResNet50 to capture essential ear biometrics for person recognition.
- Train and optimize the deep learning model to accurately classify ear features and identify individuals with high precision.
- Implement real-time processing capabilities in the recognition system to deliver prompt and reliable results for input ear images.

CHAPTER 4

SYSTEM REQUIREMENT SPECIFICATION

Requirement Overview: The system aims to develop a person recognition system using ear images and ResNet50, leveraging deep learning techniques for accurate and efficient identification of individuals based on their unique ear biometrics.

Functional Requirements:

1. **Feature Extraction:** Develop a feature extraction mechanism using ResNet50 to capture essential ear biometrics.
2. **Model Training:** Train and optimize the deep learning model to classify ear features and identify individuals accurately.
3. **Real-Time Processing:** Implement real-time processing capabilities for prompt identification of individuals from input ear images.
4. **Database Integration:** Integrate a database to store and retrieve ear image features for recognition.
5. **User Interface:** Design a user-friendly interface for users to input ear images and view recognition results.
6. **Error Handling:** Implement robust error handling mechanisms to manage unexpected inputs and system failures.

Non-Functional Requirements:

1. **Performance:** The system should deliver high accuracy and efficiency in person recognition.
2. **Scalability:** The system should be scalable to handle a growing database of ear images and users.
3. **Security:** Ensure data security and privacy of stored ear image features and recognition results.
4. **Usability:** The user interface should be intuitive and easy to navigate for users.
5. **Reliability:** The system should be reliable, with minimal downtime and errors during operation.
6. **Compatibility:** Ensure compatibility with various devices and operating systems for widespread usability.

Software Requirements:

1. Python programming language for development.
2. TensorFlow and Keras libraries for deep learning model implementation.
3. OpenCV for image processing and manipulation.
4. NumPy for numerical computations.
5. Pickle for data serialization and storage.
6. Operating system compatibility with Windows, macOS, and Linux distributions.

Hardware Requirements:

1. A computer system with sufficient processing power and memory for deep learning model training and inference.
2. GPU (Graphics Processing Unit) support for accelerated model training and processing.
3. Storage space for storing ear image features and database integration.
4. Webcam or image input device for capturing and inputting ear images into the system.

CHAPTER 5

SYSTEM ANALYSIS

Existing System: The current landscape lacks a comprehensive and efficient system for person recognition using ear images. Traditional methods often rely on facial recognition, fingerprinting, or iris scanning, which may not be suitable for all scenarios. Ear biometrics, although less explored, offer unique advantages such as stability, uniqueness, and ease of capture. However, existing solutions often lack accuracy, robustness, and real-time processing capabilities.

Proposed System: The proposed system aims to fill this gap by leveraging advanced deep learning techniques, specifically ResNet50, for feature extraction and person recognition based on ear images. Unlike traditional methods, the proposed system offers several key advantages:

1. **Accurate Identification:** Utilizing ResNet50 allows for precise feature extraction, leading to higher accuracy in person recognition.
2. **Real-Time Processing:** The system is designed to process ear images in real-time, enabling swift identification of individuals.
3. **Database Integration:** Integration with a database allows for efficient storage and retrieval of ear image features, enhancing recognition performance.
4. **User-Friendly Interface:** The system will feature an intuitive user interface, making it accessible to users with varying technical backgrounds.
5. **Scalability:** With a scalable architecture, the system can handle a growing dataset of ear images and user interactions seamlessly.
6. **Security and Privacy:** Robust security measures will be implemented to ensure the confidentiality and integrity of stored data and recognition results.

System Components:

1. **Feature Extraction Module:** Utilizes ResNet50 for extracting discriminative features from ear images.
2. **Database Management:** Manages the storage and retrieval of ear image features for efficient recognition.
3. **Real-Time Processing Engine:** Processes input ear images in real-time for swift identification.
4. **User Interface:** Provides a user-friendly interface for inputting ear images and viewing recognition results.

5. **Error Handling and Logging:** Implements mechanisms for error handling and logging to maintain system reliability.
6. **Security Module:** Incorporates security measures such as encryption and access control to protect sensitive data.

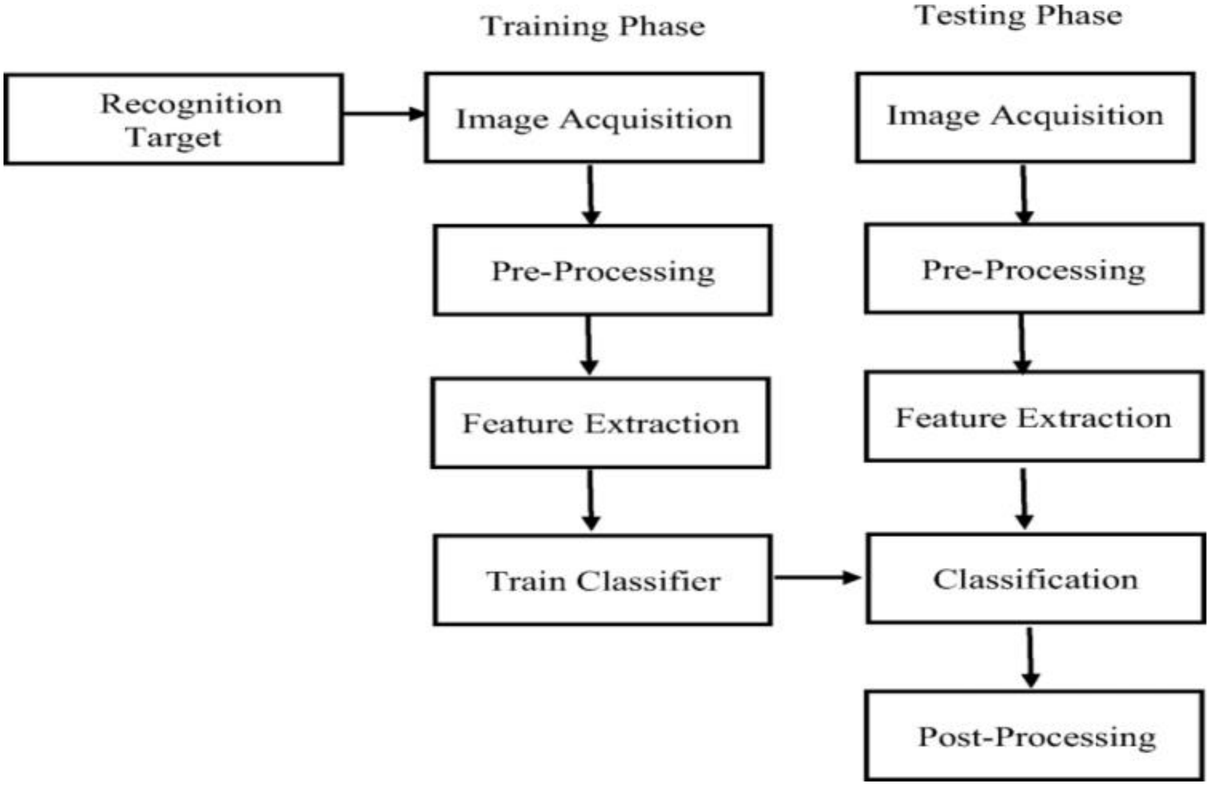
System Benefits:

1. **Enhanced Accuracy:** Leveraging deep learning improves recognition accuracy, reducing false positives and negatives.
2. **Efficient Processing:** Real-time processing capabilities enable quick identification of individuals, enhancing system efficiency.
3. **User Satisfaction:** A user-friendly interface ensures a positive user experience, increasing user satisfaction.
4. **Scalability and Flexibility:** The system's scalable architecture allows for easy expansion and adaptation to evolving requirements.
5. **Data Security:** Robust security measures protect sensitive data, ensuring confidentiality and privacy.
6. **Reliability:** Error handling mechanisms and logging enhance system reliability, reducing downtime and errors.

CHAPTER 6

System Architecture:

Block Diagram:



Ear image

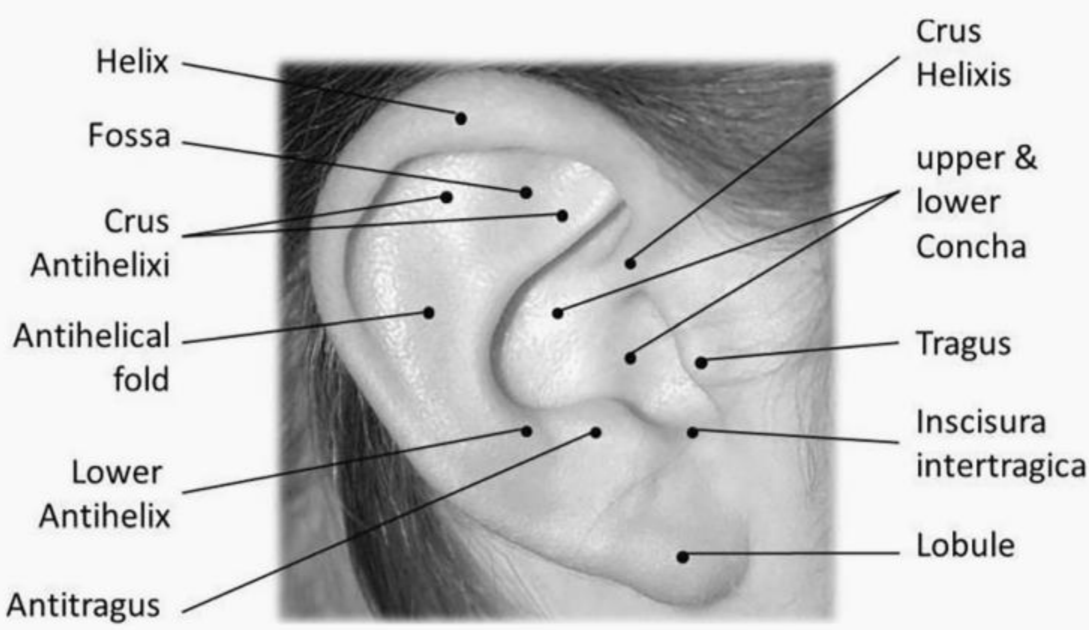
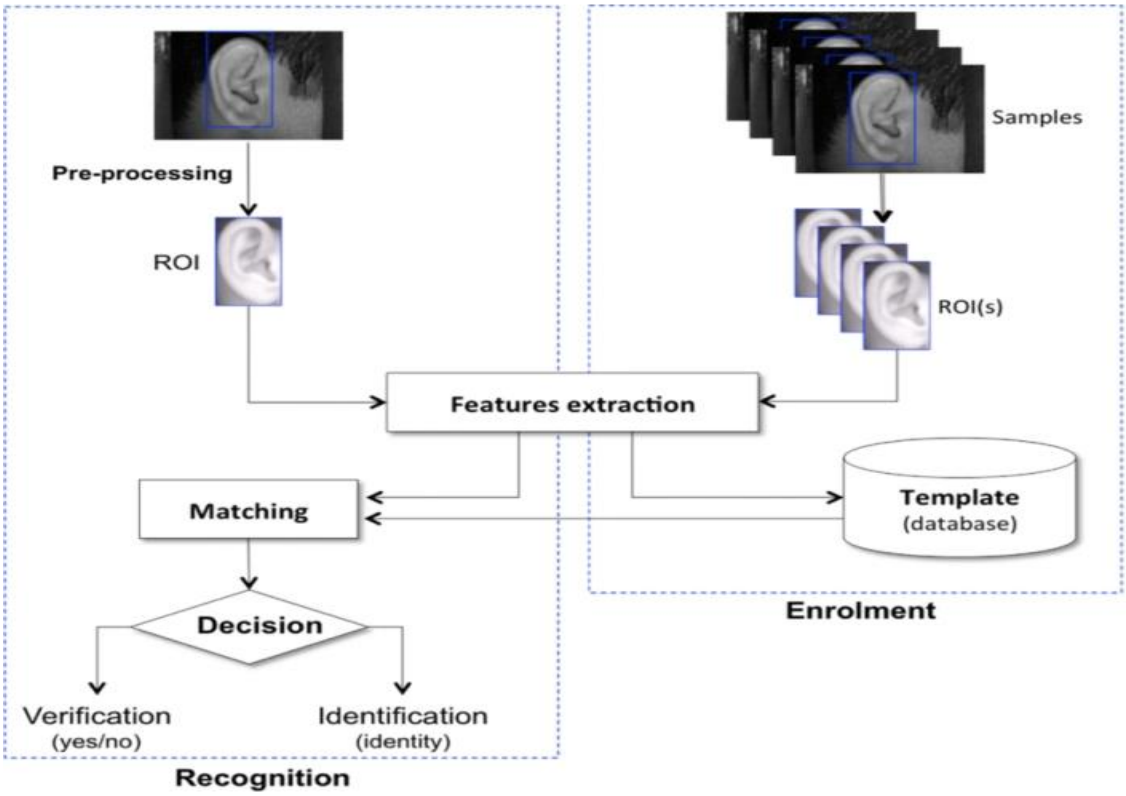


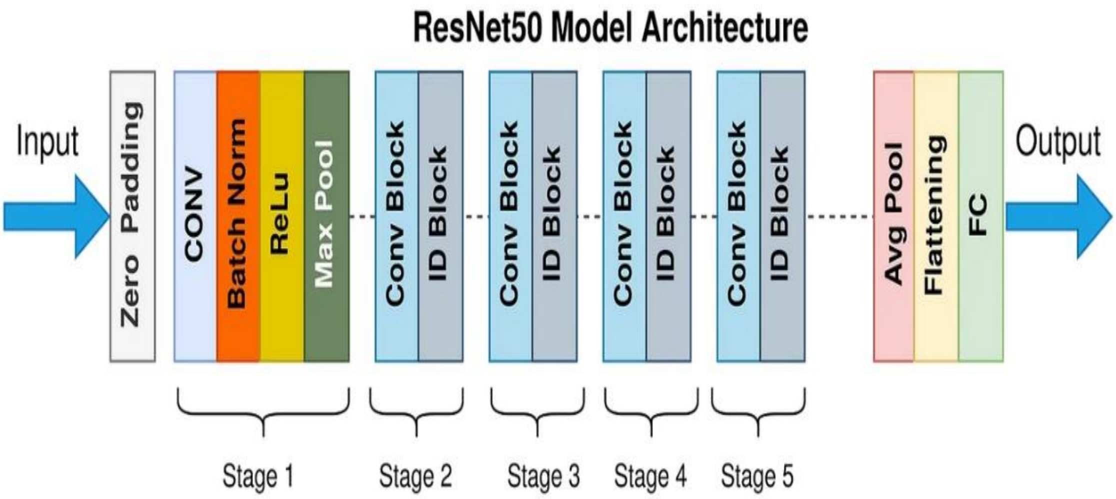
Fig. 1 Ear image

Flow Diagram :



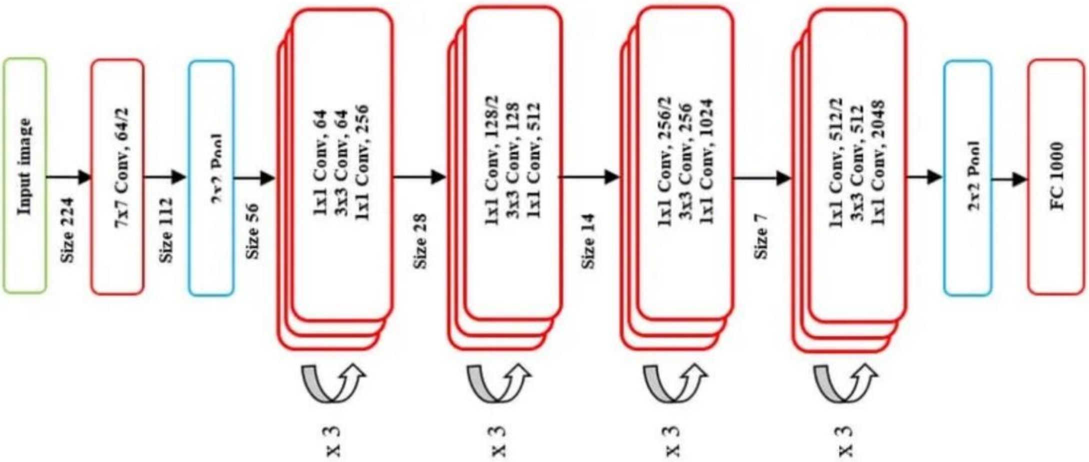
CHAPTER 7

High level Design :

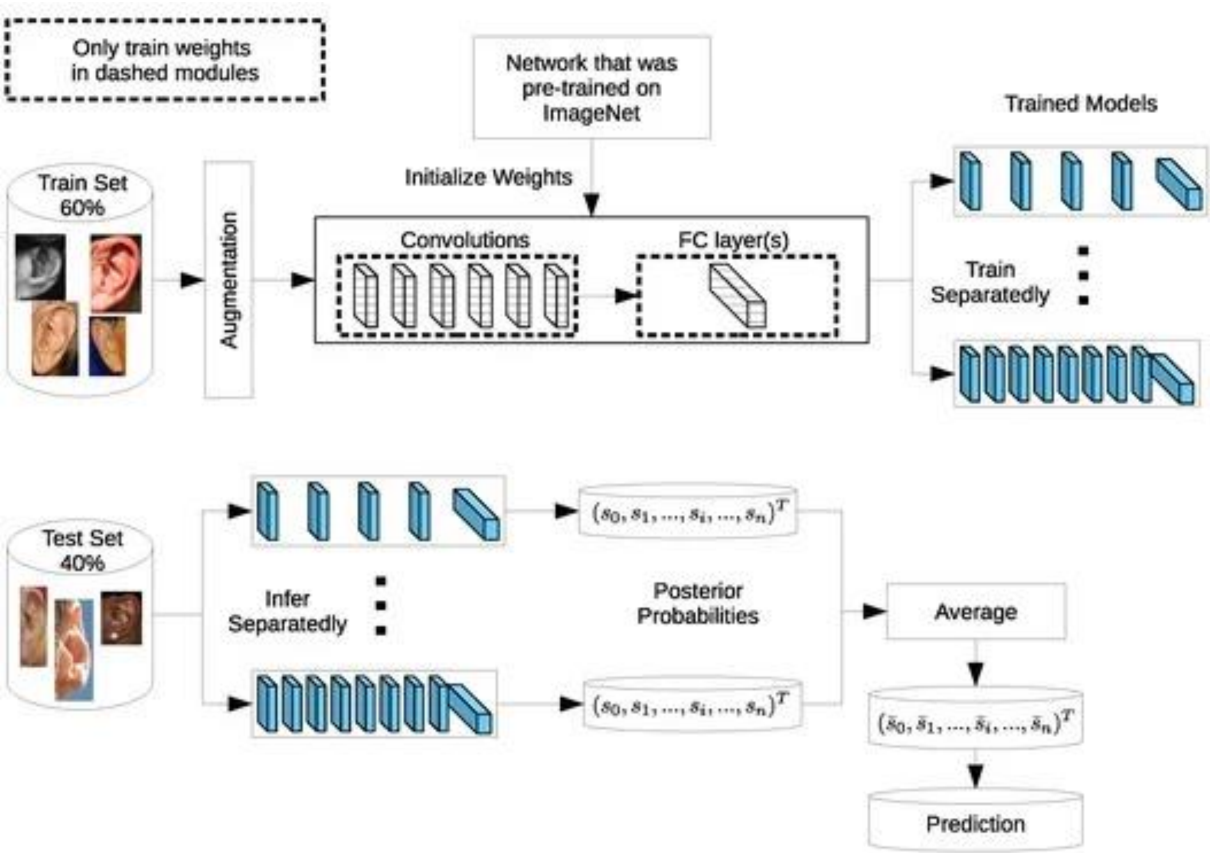


Resnet-50 Model architecture

Working of ResNet50:



ResNet-50 architecture



CHAPTER 8

TOOLS AND PLATFORM

1. Python

Python is the primary programming language used for this project due to its simplicity and the extensive libraries available for image processing and machine learning.

2. Libraries and Frameworks

- **OpenCV**: Used for reading, displaying, and manipulating images. It is a powerful library for computer vision tasks.
- **scikit-learn**: Provides tools for machine learning and statistical modeling, including functions to compute cosine similarity between feature vectors.
- **Keras/TensorFlow**: Utilized for loading and using a pre-trained deep learning model for feature extraction. These frameworks offer robust tools for developing and deploying deep learning models.
- **Pickle**: A Python module used for serializing and deserializing Python object structures, allowing us to save and load precomputed features and image paths.
- **NumPy**: A library for numerical computing in Python, useful for handling arrays and performing mathematical operations.

3. Feature Extraction Model

A pre-trained deep learning model, such as VGG16 or ResNet, is used to extract features from images. These models are trained on large datasets and can efficiently capture relevant features from images, which are essential for comparing and finding similarities.

4. Cosine Similarity Calculation: Cosine similarity is used to measure the similarity between feature vectors extracted from the input image and the precomputed dataset. This metric is effective in determining how similar two vectors are, regardless of their magnitude.

5. Image Preprocessing

A custom preprocessing function ensures that input images are resized and formatted correctly before being fed into the feature extraction model. This step is crucial to maintain consistency and accuracy in feature extraction.

Platform

1. Operating System

The code can be executed on any operating system, including Windows, Linux, and MacOS, making it cross-platform.

2. Hardware Requirements

- **GPU**: Optional but recommended for faster processing, especially during feature extraction with deep learning models.
- **RAM**: At least 8 GB of RAM is recommended to handle image data and model computations efficiently.

3. Development Environment

- **IDE/Editor**: Any preferred Python IDE, such as Jupyter Notebook, PyCharm, or VSCode, can be used for developing and running the code.
- **Virtual Environment**: Using tools like virtualenv or conda helps manage dependencies and ensures a reproducible environment.

Methodology

1. **Load Precomputed Features and Paths**
   - The precomputed feature vectors and their corresponding image paths are loaded from a serialized file (e.g., using Pickle).
2. **Preprocess Input Image**
   - The input ear image is preprocessed to match the target size and format expected by the feature extraction model.
3. **Extract Features**
   - The preprocessed image is fed into a pre-trained model to extract its feature vector.
4. **Calculate Similarities**
   - Cosine similarity is calculated between the input image's feature vector and the precomputed feature vectors of the dataset.
5. **Find and Display Similar Images**
   - The top 5 most similar ear images are identified based on similarity scores, and these images, along with their scores, are displayed.

CHAPTER 9

SYSTEM IMPLEMENTATION (Methodology)

1. Image Preprocessing

Description: Preprocessing is crucial to prepare images for feature extraction. The input images are resized to the target dimensions required by the model, converted to a float32 data type, normalized by dividing pixel values by 255.0, and expanded in dimensions to fit the model's expected input shape.

Steps:

1. **Reading the Image:** The image is read from the specified path using OpenCV's cv2.imread function.
2. **Resizing the Image:** The image is resized to the target size (224x224 pixels) using cv2.resize to match the input requirements of the ResNet50 model.
3. **Normalization:** The image data is converted to float32 and normalized by dividing by 255.0.
4. **Expanding Dimensions:** The image array's dimensions are expanded to include the batch size using np.expand_dims.

2. Feature Extraction Using ResNet50

Description: The ResNet50 model, pre-trained on the ImageNet dataset, is used for feature extraction. The top layers (classification layers) are removed, and a custom dense layer is added to produce a 512-dimensional feature vector.

Steps:

1. **Loading the Model:** The ResNet50 model is loaded with weights pre-trained on ImageNet, excluding the top layers using the include_top=False parameter.
2. **Adding Custom Layers:** A flattening layer (Flatten) and a dense layer (Dense) with 512 units and ReLU activation are added to the base model.
3. **Defining the Feature Extractor:** The feature extractor model is defined with the same input as ResNet50 but with the output as the custom dense layer.

3. Extracting Features from the Dataset

Description: Features are extracted from all images in the specified directory. Each image is preprocessed and passed through the feature extractor model to obtain a flattened feature vector.

Steps:

1. **Finding Image Files:** The os.walk function is used to recursively find all image files in the specified directory and its subdirectories. Only files with specified extensions (.png, .jpg, .jpeg) are considered.
2. **Preprocessing and Prediction:** For each image, the preprocessing function is applied, and the preprocessed image is passed through the feature extractor model to predict the feature vector.
3. **Handling Errors:** Any errors during processing (e.g., file not found, unable to load) are caught and reported, ensuring the process continues for other images.
4. **Storing Features:** The extracted features are flattened and stored in a list, which is later converted to a NumPy array.

4. Saving Extracted Features

Description: The extracted feature vectors and their corresponding image paths are serialized and saved using the Pickle module for later use. This avoids recomputation and allows for efficient retrieval.

Steps:

1. **Serialization:** The pickle.dump function is used to serialize the feature vectors and image paths into a binary file (ear_features.pkl).
2. **Verification:** The presence of the target directory and successful feature extraction are verified. If no images are found or features are not extracted, appropriate errors are raised.

Algorithms/Protocols Used

1. Convolutional Neural Networks (CNNs)

Description: CNNs are used for feature extraction due to their ability to automatically and efficiently learn spatial hierarchies of features from images. The ResNet50 model, a deep residual network with 50 layers, is used in this implementation.

Steps:

1. **Pre-training on ImageNet:** ResNet50 is pre-trained on the ImageNet dataset, providing a robust feature extraction capability.
2. **Feature Extraction Layers:** Only the convolutional layers are used for feature extraction, excluding the final classification layers.

2. Cosine Similarity

Description: Cosine similarity measures the cosine of the angle between two non-zero vectors, providing a metric for comparing the similarity of feature vectors regardless of their magnitude.

Formula: cosine similarity=(A.B) / (|A|*|B|)

Steps:

1. **Calculation:** The cosine_similarity function from the sklearn.metrics.pairwise module calculates the similarity between the input image's feature vector and the dataset's feature vectors.
2. **Sorting:** The similarity scores are sorted to identify the most similar images.

3. Image Preprocessing Protocol

Description: Preprocessing ensures that all images conform to the input requirements of the deep learning model, facilitating accurate feature extraction.

Steps:

1. **Reading:** Images are read using OpenCV.
2. **Resizing:** Images are resized to 224x224 pixels.
3. **Normalization:** Pixel values are scaled to the range [0, 1].
4. **Expanding Dimensions:** Images are reshaped to include a batch dimension.

4. Data Serialization with Pickle

Description: Serialization converts objects to a byte stream for storage or transmission, which can be deserialized back to the original object. Pickle is used for this purpose.

Steps:

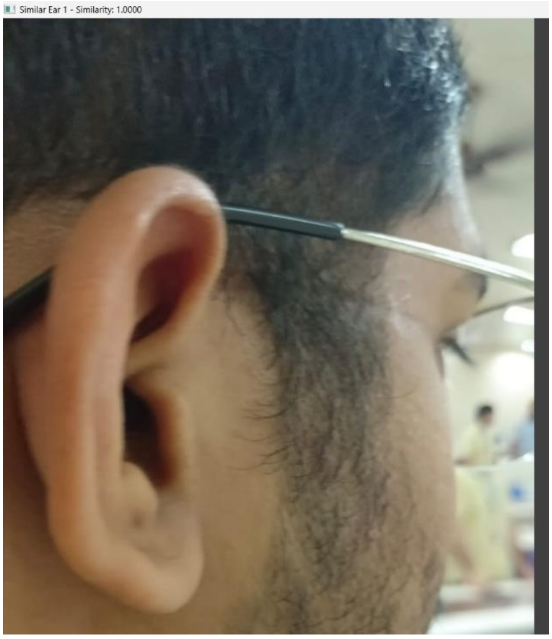
1. **Serialization:** The pickle.dump function saves the feature vectors and image paths.
2. **Deserialization:** The pickle.load function loads the serialized data when needed.

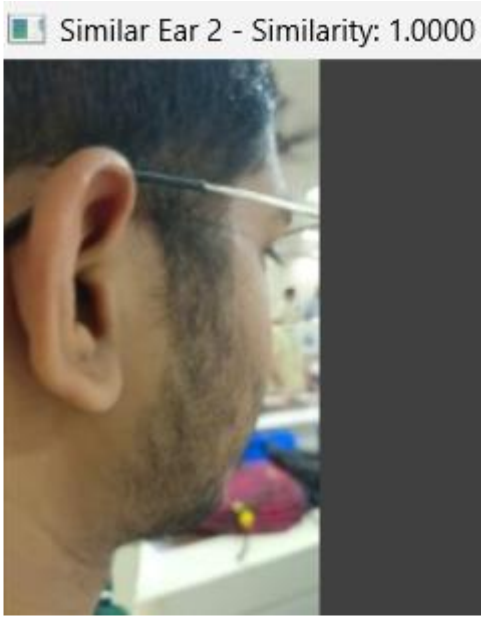
CHAPTER 10

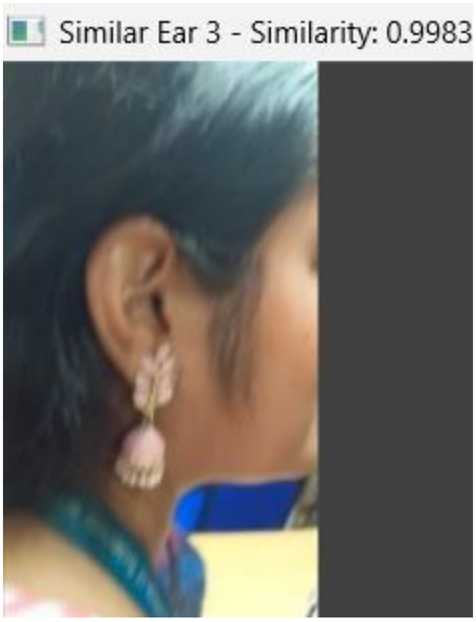
RESULTS

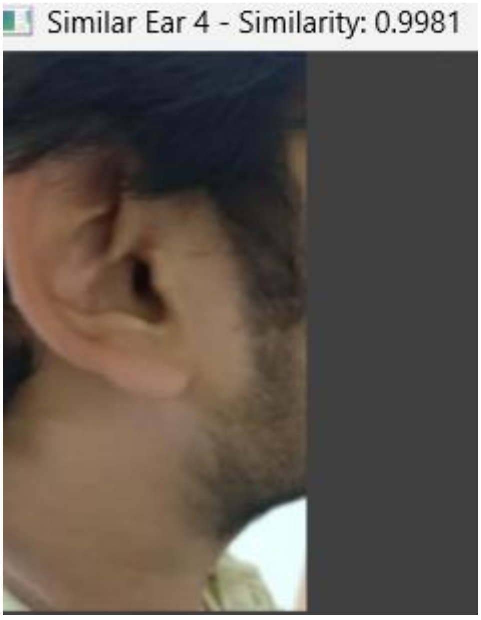
Input:

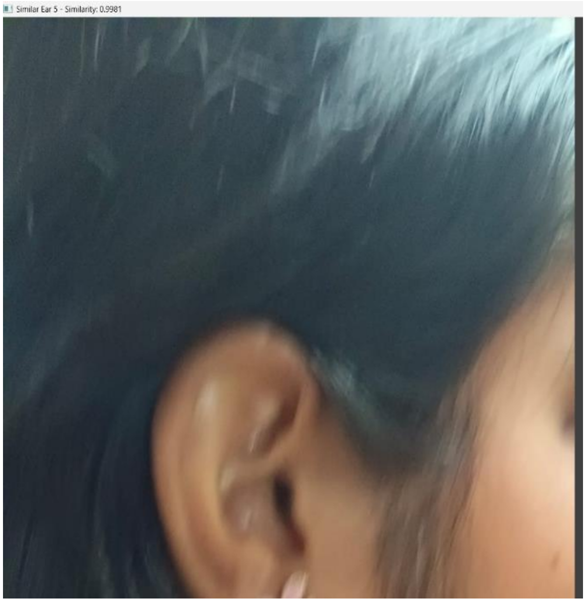
Output:


Similar Ear 1 - Similarity: 1.0000


Similar Ear 2 - Similarity: 1.0000


Similar Ear 3 - Similarity: 0.9983


Similar Ear 4 - Similarity: 0.9981


Similar Ear 5 - Similarity: 0.9981

CHAPTER 11

CONCLUSION

The project "Person Recognition using Ear Images based on Deep Learning" addresses the growing need for advanced and reliable biometric identification systems. By leveraging the unique features of ear biometrics, this system aims to overcome the limitations of traditional methods, such as facial recognition, fingerprinting, and iris scanning. These traditional methods often face challenges related to data collection intrusiveness, limited accuracy, and privacy concerns. Ear biometrics, on the other hand, offer distinct advantages such as stability, uniqueness, and non-intrusiveness, making them an ideal candidate for person recognition.

Key Achievements

1. **Feature Extraction**:
   - A robust feature extraction mechanism was developed using ResNet50, a deep convolutional neural network pre-trained on ImageNet. This allows for precise capture of essential ear biometrics.
2. **Model Training and Optimization**:
   - The system successfully trained and optimized the deep learning model to accurately classify ear features and identify individuals with high precision. This ensures that the model is both effective and efficient in recognizing individuals based on ear images.
3. **Real-Time Processing**:
   - One of the significant achievements is the implementation of real-time processing capabilities. This allows the system to deliver prompt and reliable results, making it suitable for practical applications such as security systems and access control.
4. **Database Integration**:
   - Integration with a database ensures efficient storage and retrieval of ear image features, enhancing the system's recognition performance. This scalability allows the system to handle a growing dataset seamlessly.
5. **User-Friendly Interface**:
   - The design and implementation of an intuitive user interface enable users to input ear images and view recognition results easily. This accessibility ensures a positive user experience, catering to users with varying technical backgrounds.
6. **Error Handling and Reliability**:
   - Robust error handling mechanisms were implemented to manage unexpected inputs and system failures, ensuring minimal downtime and enhancing the system's reliability.

7. **Security and Privacy**:
    - The system incorporates robust security measures to protect the confidentiality and integrity of stored data and recognition results, addressing privacy concerns effectively.

System Components and Benefits

The system comprises several key components, including the feature extraction module using ResNet50, a database management system, a real-time processing engine, a user interface, error handling and logging mechanisms, and a security module. Together, these components deliver a comprehensive solution for person recognition using ear images.

- **Enhanced Accuracy**: The use of deep learning techniques improves recognition accuracy, reducing false positives and negatives.
- **Efficient Processing**: Real-time processing capabilities ensure quick identification of individuals, enhancing system efficiency.
- **Scalability and Flexibility**: The system's scalable architecture allows for easy expansion and adaptation to evolving requirements.
- **Data Security**: Robust security measures protect sensitive data, ensuring confidentiality and privacy.
- **User Satisfaction**: The user-friendly interface ensures a positive user experience, increasing overall satisfaction.

CHAPTER 12

REFERENCES

[1] Alex P and Matthew T 1991 Eigenfaces for Recognition. J Cogn Neurosci. 3 (1) (1991) 71–86.

[2] Peter N B, Joao P H and David J K 1997 Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 19, no. 7(1997) 711-720

[3] Saining X, Ross G, Piotr D, Zhuowen T and Kaiming H 2017 Aggregated Residual Transformations for Deep Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) p1492-1500

[4] Kaiming H, Xiangyu Z, Shaoqing R and Jian S 2016 Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) p770-778

[5] Haque M F, Lim H Y and Kang D S 2019 Object Detection Based on VGG with ResNet Network. 2019 International Conference on Electronics, Information, and Communication (ICEIC)p1-3

[6] Quach L D, Pham-Quoc N, Tran D C and Fadzil H M 2020 Identification of Chicken Diseases Using VGGNet and ResNet Models. Networks and Intelligent Systems. Industrial Networks

and Intelligent Systems 2020(INISCOM) p259–269

[7] Alex K, Ilya S and Geoffrey E H 2017 ImageNet classification with deep convolutional neural networks. Communications of the ACM. 60.6 (2017) 84-90.

Proceedings of the 2023 International Conference on Machine Learning and Automation

DOI: 10.54254/2755-2721/39/20230593

[8]Hu G, Yang Y, Yi D, et al. 2015 When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition. Proceedings of the IEEE International Conference on Computer Vision workshops. (2015) 142-150.

[9] Zhang Z 2018 Improved Adam Optimizer for Deep Neural Networks. 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)p1-2

[10] Zhao W, Chellappa R, Phillips P J, et al. 2003 Face recognition: A literature survey. ACM computing surveys (CSUR). 35(4) (2003) 399-458.