Ensemble learning framework for Biometric Data via deep Hash Ranking

CS4090 Project Final Report

Submitted by

Padala Sawtvik	**B200812CS**
Veerisetty Arun Krishna	**B200799CS**
Kommineni Avinash Krishna	**B200834CS**

Under the Guidance of
T.A Sumesh

Department of Computer Science and Engineering
National Institute of Technology Calicut
Calicut, Kerala, India - 673 601

April 28, 2024

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT, KERALA, INDIA - 673 601

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

2024

CERTIFICATE

Certified that this is a bonafide record of the project work titled

PROJECT TITLE

done by

Padala Sawtvik
Veerisetty Arun Krishna
Komminenei Avinash Krishna

of eighth semester B. Tech in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering of the National Institute of Technology Calicut

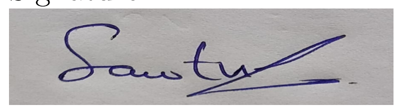
Project Guide
T.A Sumesh
Assistant Professor

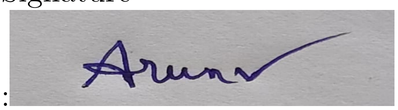
07.05.2024

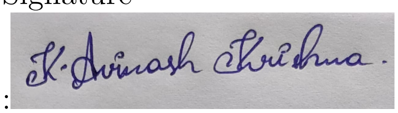
DECLARATION

I hereby declare that the project titled, **Ensemble learning framework for Biometric Data via deep Hash Ranking**, is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or any other institute of higher learning, except where due acknowledgement and reference has been made in the text.

Place : NIT Calicut
Date : 05-05-24

Name : Padala Sawtvik
Reg No : B200812CS
Signature :


Name:Veerisetty Arun Krishna
Reg No : B200799CS
Signature
:

Name: K Avinash Krishna
Reg No : B200834CS
Signature
:

Abstract

Our task is aimed toward improving the retrieval and analysis of biometric facts via the usage of deep hashing methodologies. Biometric statistics encompasses various personal identifiers, which includes fingerprints, facial traits, and iris scans. Efficient retrieval and analysis of biometric data are vital for programs inclusive of identity verification, get right of entry to manipulate, and protection. In the world of biometric facts, traditional strategies have traditionally been employed for retrieval. Nevertheless, this summary underscore the capability of deep hashing techniques to significantly increase the pace and performance of these retrieval procedures. Deep hashing on this context merges feature extraction with hash coding, which involves the conversion of biometric traits into binary hash codes. Furthermore, we introduce an ensemble-based totally deep neural model framework tailor-made for the retrieval and evaluation of biometric facts. This framework has been engineered to collect concise hash codes containing wealthy semantic facts while adhering to hash constraints. Ensemble strategies are carried out to reinforce the performance, which include weighted voting for integrating. Rating data into the retrieval technique. This approach endeavours to improve the performance and precision of biometric statistics retrieval and analysis.

ACKNOWLEDGEMENT

We would like to express our sincere and heartfelt gratitude to our guide Sumesh T A sir who have provided invaluable guidance, support, and encouragement, without sir, we wouldn't have achieved this good progress.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In our increasingly interconnected world, characterised via the speedy evolution of conversation technology, expansive pc networks, and the proliferation of excessive-potential garage solutions, biometric information has emerged as a effective and versatile information source. This facts format distils complicated data right into a concise and intuitive illustration, facilitating a extensive array of applications throughout various domain names. Biometric facts retrieval, corresponding to its counterpart in photo retrieval, performs a pivotal position in addressing a variety of realworld challenges.

The proposed framework encompasses numerous key additives. First, a hidden hash layer is incorporated into the architecture of deep neural networks. This novel layer serves a dual motive: it extracts high-level semantic capabilities from biometric records and generates approximate hash codes. These hash codes are finally quantized to yield particular binary representations, retaining vital records while lowering dimensionality. Second, the take a look at leverages a weighted vote casting strategy for biometric records retrieval, which takes into account the spatial relationships many of the retrieved facts factors. By integrating the region statistics of the retrieved statistics, the technique ensures that the maximum relevant effects are ranked highest, emphasizing the significance of similarity between the topranked records point and the query facts.

In summary, this technique, drawing suggestion from the successes of ensemble gaining knowledge of and deep neural networks in photo retrieval, offers a promising road for enhancing the efficiency and accuracy of biometric records retrieval in the face of ever expanding and complicated datasets.

Chapter 2

Literature Survey

Approximate nearest neighbour

This study paper introduces a singular approach to big-scale face retrieval via addressing the trade-off between pace and accuracy[1]. Approximate Nearest Neighbour Search (ANNS) prioritizes pace over accuracy, probably lacking relevant statistics. To bridge this gap, the paper defines the idea of Neighbourhood-Exact Nearest Neighbour Search (NENNS). NENNS seeks efficient retrieval at the same time as making sure accuracy within a selected question neighbourhood. The proposed technique discretizes continuous capabilities into binary codes and employs binarized Cosine similarity to exclude irrelevant applicants, considerably reducing seek area. The paper offers theoretical evidence and substantial experiments, demonstrating the effectiveness of their method. Overall, this studies presents a valuable contribution to the field of big- scale face retrieval, making sure a balance among search performance and precision, with ability applications in areas including surveillance and image corporation. In current years, ANNS technology has advanced unexpectedly due to its reduced area and time requirements, and its ability to achieve correct retrieval effects. Among ANNS approaches, hashing, as a consultant nearest-neighbour seek technique, is widely used for large-scale image retrieval.

Unsupervised hashing

The unsupervised hashing approach[2] is impartial of the enter data in the course of the manner of designing characteristic statistics mapping schemes. This technique does now not use the label information of the dataset, nor does it do not forget the distribution of the facts; it makes use of the picture itself to learn the hash feature. The fundamental studies instructions regarding conventional unsupervised hashing strategies are hash feature getting to know and hash quantization coding

semisupervised hashing

A semi-supervised hashing method makes use of a small quantity of dataset labels and unlabelled photograph information to analyze a hash feature. The maximum representative semi- supervised hashing approach is semi- supervised hashing (SH [3]), that makes use of labelled statistics to study hash functions to keep semantic similarity and enhance the first-rate of the resulting hash codes. By optimizing the errors between the labelled information point pairs, the generated hash codes are as small as viable while balanced and independent, and best a part of the information contains supervision information. The algorithm has strong robustness and might prevent the occurrence of overfitting to a certain quantity.

supervised hashing

To make full use of label information, scholars have proposed supervised hashing methods. A supervised hashing method targets to apply supervision information (labelled records) for hash function education. Minimal loss hashing (MLH [4]) evaluates the similarities between data primarily based on label facts and obtains powerful hash codes.

Deep hash based retrieval method

With the breakthrough of deep getting to know concept, the capacity of deep learning models to specific high-level semantic records contained in pix has end up increasingly more distinguished. AlexNet [5] has carried out the maximum recognition accuracy fees on ILSVRC2012. Therefore, methods based totally on deep learning, mainly CNNs, have performed awesome outcomes. During 2010 and 2017, some traditional CNN models emerged, inclusive of VGG, ResNet, and DenseNet. Therefore, pupils have tried to integrate deep neural networks with hashing generation; hence, supervised deep hashing methods have emerged and completed correct results. Semantic hashing (SH [6]) turned into the first technique to link deep neural networks with hashing techniques. Based in this new method, academia started to pay attention to the mixture of deep mastering and hashing algorithms.

Ensemble learning

Ensemble learning [7] completes the learning task by constructing and combining multiple individual learners; this method involves using a certain strategy to integrate the results of each individual learner to obtain a better learning effect than those of the individual learners. Ensemble strategies mainly include the averaging, weighted voting, and learning methods. Ensemble learning is based on the premise that even if a certain weak classifier (referring to a classifier with low accuracy) obtains an incorrect prediction, other strong classifiers can correct the error. While training a stable, well-performing model is our ultimate goal, real-world scenarios are rarely perfect, and we occasionally end up with a small number of models that match our preferences. Individual students must be "good but different"—that is, they must possess a certain level of "accuracy"—and diverse (i.e., there must be differences between individual learners) in order to form a good ensemble. [8]. Ensemble learning is also widely used in the field of deep learning.

Chapter 3

Problem Definition

Our aim is to present an ensemble deep neural model tailored to expedite the process of large-scale facial similarity retrieval from datasets. This model is engineered to acquire succinct hash codes, effectively encapsulating extensive semantic information, all the while maintaining compliance with hash constraints. Our approach involves returning a list of images closest to the queried image, and further ranking these images through weighted voting mechanisms.

Chapter 4

Methodology

In this project we are going to used three individual learners which are models based on classic VCG,Resnet and DenseNet.In different tasks,they can be regarded as feature extractors guided by an objective function specifically designed for a single task.The learners perform feature extraction on the same image, their focuses are different,so they are suitable for individual learners.CNNs have strong autonomous learning capabilities and feature expression capabilities.Therefore using the image features extracted by a CNN as image descriptors can be used for biometrics.Ref. [42] showed that the features output by the second fully connected layer fc7 of a neural network can be regarded as image features and used for subsequent biometric classification tasks. However, these features are high-dimensional vectors. For image retrieval with large-scale databases, the high computational cost of this approach leads to low retrieval efficiency. A suitable method of avoiding massive calculations is the conversion of high-dimensional vectors into binary hash codes. Therefore, we introduce a hash layer between the last two fully connected layers, which can reduce the loss of image feature information and generate hash codes with rich semantic information. Thus, the input of the hash layer is divided into two parts: one part comes from fc7 , and the other part comes from fc6 .

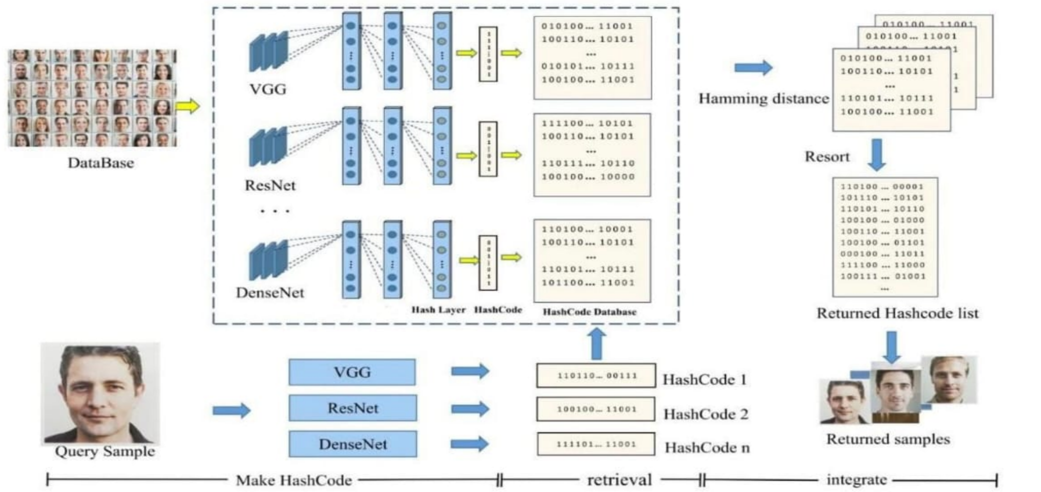


Fig. 1. Each learner in the proposed ERDH method for hashing-based learning adds a hidden layer that generates approximate binary code, and they are all obtained by fine-tuning the classic network.

Fig 1

The hidden hash layer maps the high-dimensional feature space of the image output by the deep CNN to the Hamming space and can be used for efficient retrieval. The hash layer is added after the fully connected layer fc7 . The classification layer fc8 in the model expresses the high-level semantic information of the input image and completes the classification task. The hash layer is not only a generalization of the features of the fully connected layers fc6 and fc7 , but the neurons of the hash layer also receive feedback from the classification layer fc8 during backpropagation. Therefore, the hash layer can be regarded as a bridge between fc7 and fc8 , as. It connects the middle-level features of the deep convolutional network with the high-level semantic features, such that the generated hash codes have rich semantic information. The hash function can be formulated as:

$$h(x:w) = \delta\left(w^T\left[fc_6(x); fc_7(x)\right]\right),$$

where delta of t is a logistic function, w represents the weight of the hash layer, fc6(x) and fc7(x) represent the feature vector outputs of the fully connected layers fc6 and fc7, respectively. To obtain kbit binary codes, the hash layer was set as a k-dimensional fully connected layer. The output range of the sigmoid function is a continuous value between 0 and 1, which is naturally close to 0 or 1, and the function is symmetric at approximately 0.5, which is convenient for the subsequent quantization of hash codes. Hash codes with rich semantic information require that, after the image is encoded, hash codes with similar Hamming distances should be semantically similar, and vice versa. Suppose there are N images [X1 . . . XN], the corresponding labels are [Y1 . . . YN], and the labels predicted by the model are [Yˆ 1 . . . Yˆ N]. After the sigmoid activation function of the hash layer is executed, for each image, the approximate hash codes [B1 . . . BN] can be obtained. To generate compact hash codes for each bit of the hash code, a loss function must be added to the classification layer. Furthermore, in the hash layer, a loss function is required to guide learning of the target task. Two hash functions were used to constrain the hash layer. One is the binary hash loss function because the output value of the hash layer obtained through the sigmoid activation function is continuously distributed between 0 and 1. To obtain binary hash codes that can be used for efficient image retrieval, the loss function must constrain the output value as close to zero or one as possible so that the hash codes lose a minimal amount of semantic information during the quantization process. The binary hash loss function

$$L_1 = \theta - \frac{1}{2N} \sum_{i=1}^{N} (B_i - \zeta)^2 ,$$

is expressed as:

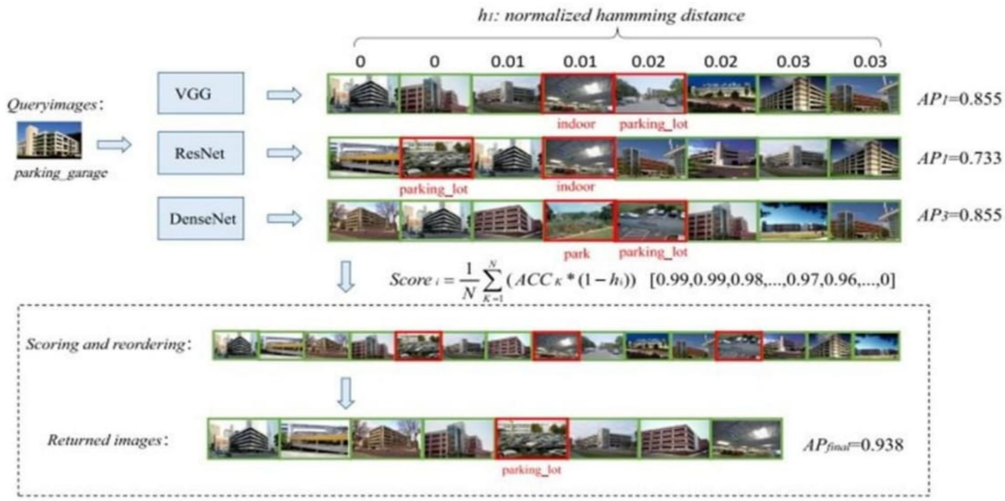


Fig. 2. Examples of the top eight retrieved images for a given query image from the SUN397 dataset. We calculate the score for each image and adjust its position. The images with green borders are the matched images, while those with red borders are incorrect matches.

Fig 2

where theta is 0.25, Bi is the hash layer output, and zeta is 0.5. The second hash loss function makes 0 and 1 each comprise half of the generated hash

$$L_2 = \frac{1}{2N} \sum_{i=1}^{N} \left(\left[\frac{\sum_{j=1}^{K} b_j}{K} \right]_i - \xi \right)^2 ,$$

codes; this is expressed as:

where $\left[\frac{\sum_{j=1}^{K} b_j}{K} \right]_i$ denotes the Bi mentioned above and xi is 0.5. To maintain semantic information during the model training process, we also used a classification loss function. The cross-entropy loss function is used in the last layer of the model and is defined as follows:

$$L_c = -\frac{1}{N} \sum_{i} \sum_{c=1}^{M} y_{ic} \log p_{ic},$$

where i represents the ith sample, N represents the number of images, M represents the number of label categories, and Y ic is the indicator function. To ensure that the hash codes of the images contain rich semantic information while ensuring classification accuracy, the overall optimization goal must include both of these aspects, and the overall loss function is defined

$$L_{all} = \alpha L_c + \beta L_1 + \gamma L_2.$$

as :
where alpha, beta and gamma are hyperparameters

Ensemble learning methods
For each image, the weighted voting method was used to determine the result for the predicted image returned by each of the three individual learners. Assuming that we have N query images, for each query image, the quantized binary hash codes are compared with the image in the database, which not only contains the hash code but also records the name of the image corresponding to the hash code and true label of the image. Generally, after sorting the Hamming matrix in ascending order, the top-k query results can be obtained; however, in this study, the accuracy of each individual learner and the top-k Hamming distances returned to participate in the score calculation process for each image. The score for each returned image can be formulated as follows:

$$Score_i = \left(\frac{1}{N} \sum_{k=1}^{N} ACC_k * \left(1 - \frac{h_i}{\max(H)} \right) \right),$$

$$S_i = descort(Score_i).$$

where ACCk represents the classification accuracy of each individual learner, hi represents the normalized Hamming distance between the ith returned image and the query image, H represents the Hamming distance matrix and N represents thenumber of individual learners. We sort thesescores in descending order and obtain the final score ranking; i.e, the k query results returned at the end are ranked. The overall idea of the reordering method is shown in Fig. 2.

Mean Average Precision (mAP): To assess the performance of our biometric data retrieval system, we will calculate the mean average precision (mAP) across multiple queries. mAP measures the effectiveness of our algorithm in retrieving relevant biometric data accurately, with a particularemphasis on facial features.

Precision for k Samples: This metric evaluates the percentage of true facial features among the top-k retrieved samples, providing insights into the precision of our face detection and biometric data retrieval system.

Datasets

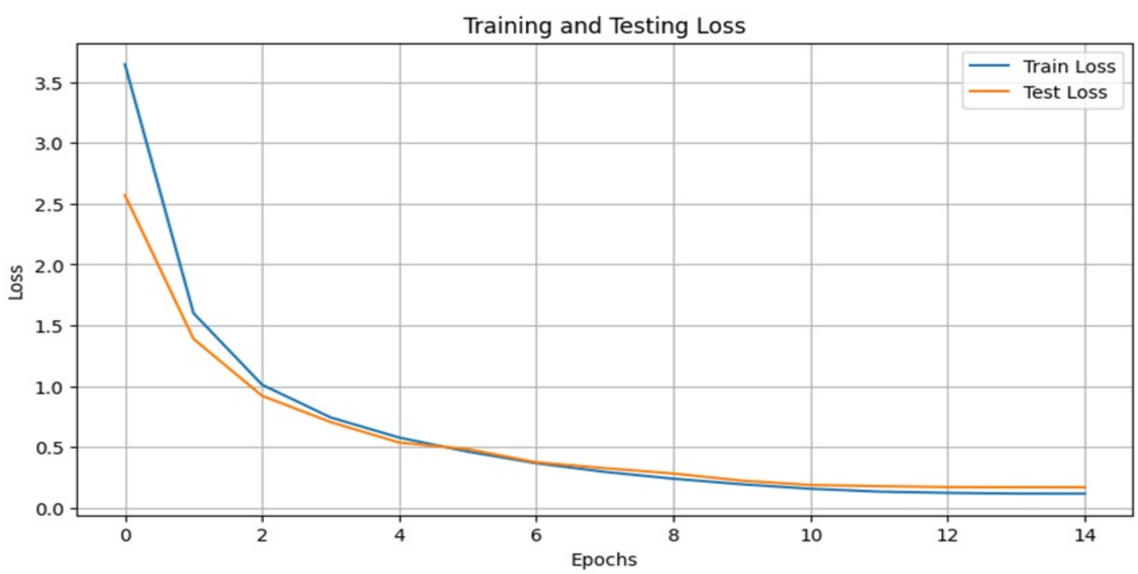
VGGFace2: This dataset is a large-scale face recognition dataset containing over 3 million images of more than 9,000 individuals. It's commonly used for training deep learning models for face recognition tasks. The dataset covers a wide range of variations in pose, age, illumination, and ethnicity, making it valuable for research and development in facial recognition technology.

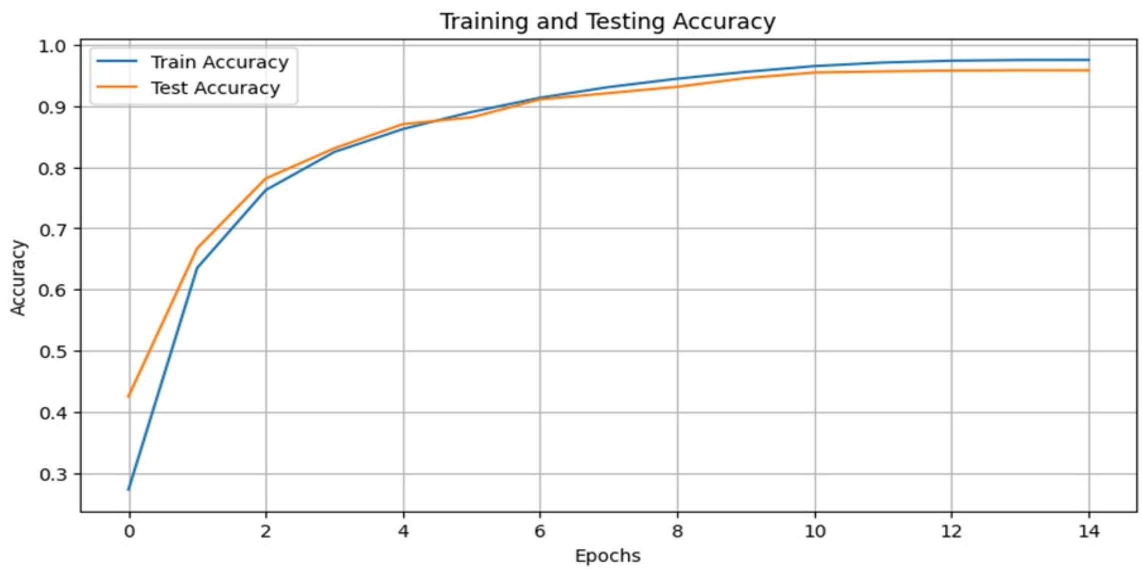
Chapter 5

Results

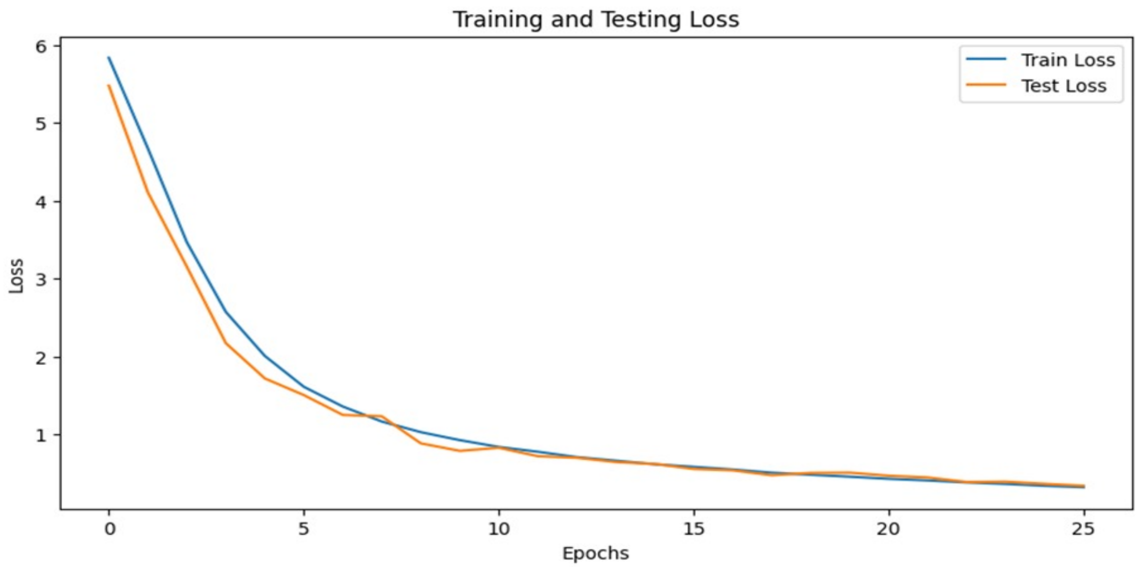
5.1 Loss Accuracy Curves

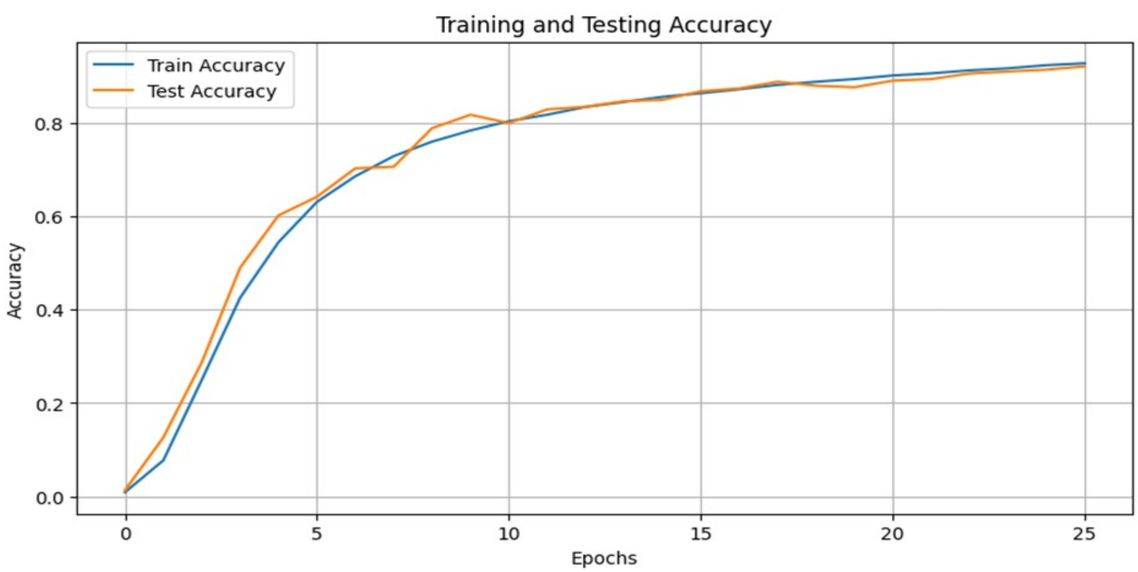
5.1.1 RESNET



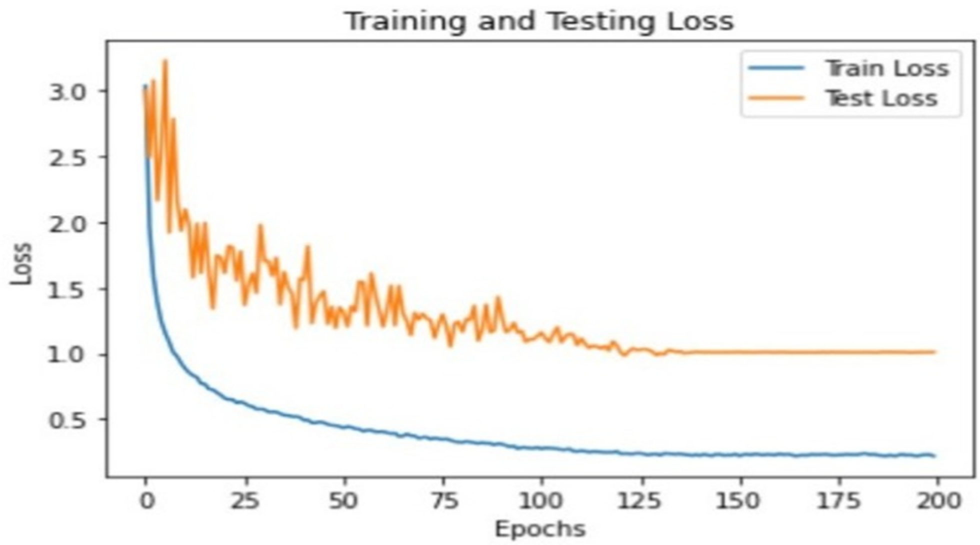

Training and Testing Accuracy
Train Accuracy
Test Accuracy
Accuracy
Epochs
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
2
4
6
8
10
12
14

5.1.2 VGG

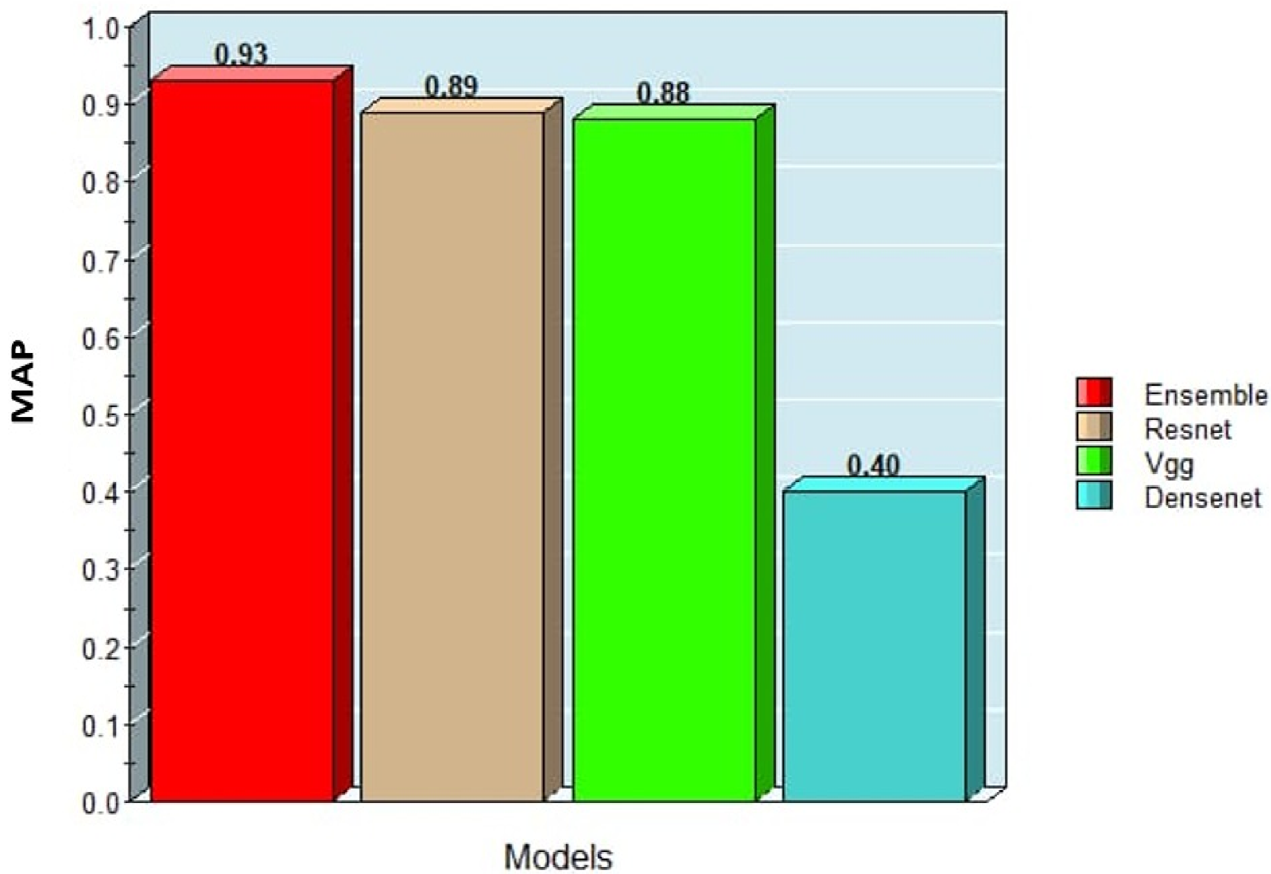

Training and Testing Loss
Train Loss
Test Loss
Loss
Epochs


Training and Testing Accuracy
Train Accuracy
Test Accuracy
Accuracy
Epochs

5.1.3 DENSENET



5.2 Ensemble learning using Vgg, Resnet and Densenet



5.3 Dataset information used in the experiments

1	Dataset	VGGFACE2
2	Train set	187808
3	Query set	39539
4	Size	224x224
5	Labels	480

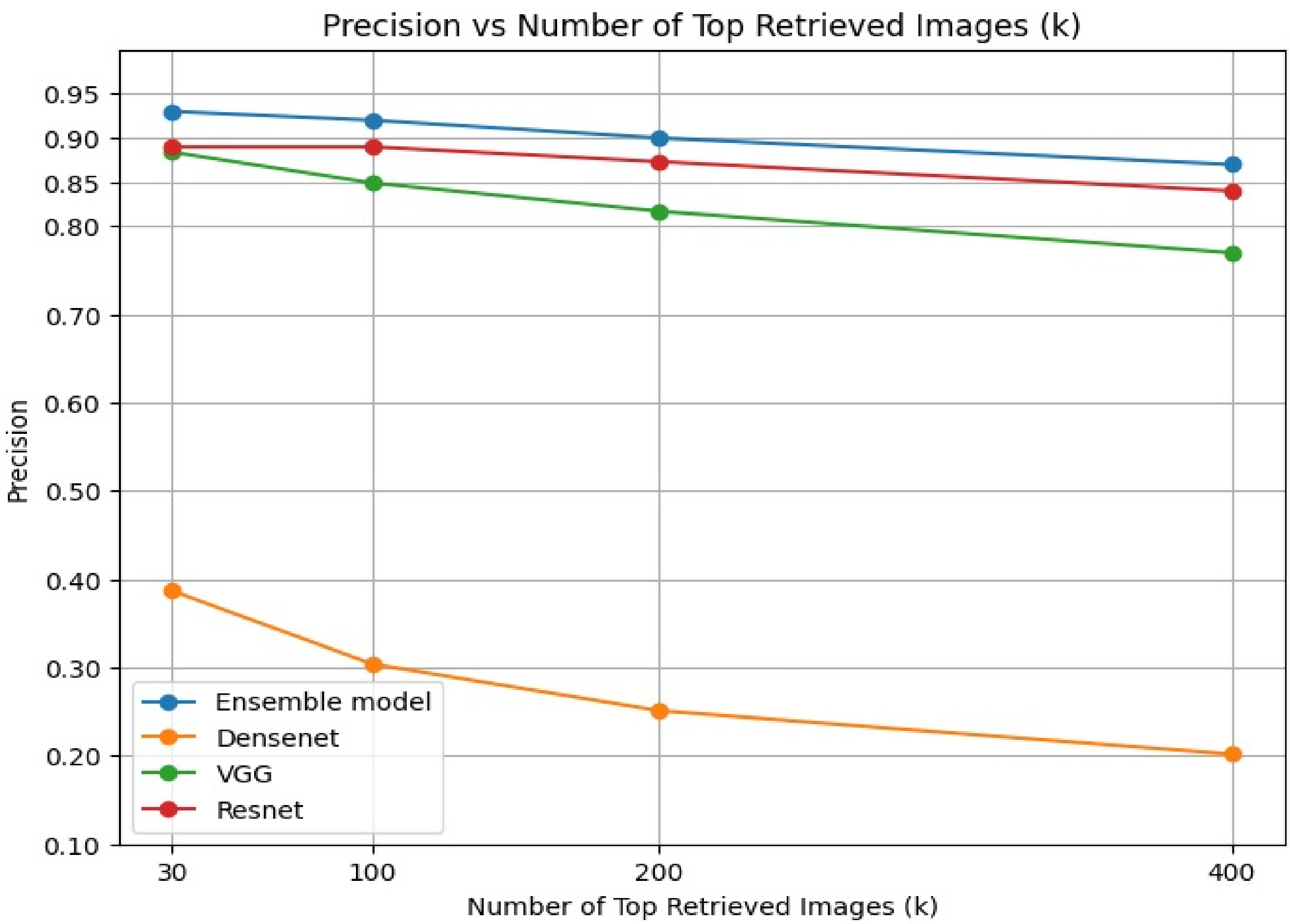
5.4 Precision rates obtained under different numbers of returned samples(k)

Top-k	Top30	Top100	Top200	Top400
Resnet	0.890	0.890	0.8732	0.84
VGG	0.884	0.849	0.8169	0.77
Densenet	0.387	0.304	0.2512	0.202
Ensemble	0.932	0.920	0.9007	0.874

5.5 Performance Metrics

Top-k=400	Precision	Recall	F1 score
Resnet	0.84	0.635	0.723
VGG	0.77	0.495	0.602
Densenet	0.202	0.0296	0.0516
Ensemble	0.874	0.528	0.658

5.6 Precision vs Number of Top Retrieved Images(k) Graph



Chapter 6

Conclusion and Future work

We introduce a novel supervised hashing network, which effectively preserves semantic information between facial images by learning binary hash codes.This is achieved through the joint optimization of a classification loss function and a hash loss function.Our proposed model offers several advantages: It is easy to implement and can be readily achieved by fine-tuning existing deep networks. The ensemble learning approach allows individual learners to compensate for any weaknesses encountered by another learner within the ensemble model, thus enhancing overall performance. Our model not only considers the similarity between the returned image and the query image but also integrates ranking information to ensure consistency between the top-ranked images and the query image. Ensemble model also provides a versatile image retrieval framework based on deep hashing, which can be seamlessly combined with efficient hashing methods to achieve superior performance.

In future work, we envision improvements in the following directions: Exploring methods to enable the network to autonomously learn these parameters could enhance the model's adaptability and performance. Leveraging the feature extraction capabilities of different networks throughout the entire process of supervising hash coding using ensemble methods may lead to further improvements in performance and robustness.

Chapter 7

References

[1] F. Chen, W. Pei, G. Lu, Neighborhood -exact nearest neighbor search for face retrieval, Knowl. -Based Syst. 248 (2022) 108757

[2] Y. Sun, Y. Ye, X. Li, S. Feng, B. Zhang, J. Kang, K. Dai, Unsupervised deep hashing through learning soft pseudo label for remote sensing image retrieval, Knowl. -Based Syst. 239 (2022) 107807.

[3] J. Wang, S. Kumar, S. -F. Chang, Semi - supervised hashing for large -scale search, IEEE Trans. Pattern Anal. Mach. Intell. 34 (12) (2012) 2393 –2406, http://dx.doi.org/10.1109/TPAMI.201 2.48 .

[4] M.E. Norouzi, D.J. Fleet, Minimal loss hashing for compact binary codes, in: Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011, 2011.

[5] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012) 1097 –1105.

[6] . R. Salakhutdinov, G. Hinton, Semantic hashing, Internat.

J. Approx. Reason. 50 (7) (2009) 969 – 978.

[7] T.G. Dietterich, et al., Ensemble learning, Handb. Brain Theory Neural Netw. 2 (2002) 110 –125.

[8] X. Zhou, J. He, C. Yang, An ensemble learning method based on deep neural network and group decision making, Knowl. - Based Syst. (ISSN: 0950 - 7051) 239 (2022) 107801.