sciendo INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS

Research Article | DOI: 10.2478/ijssis-2025-0031 Issue 1 | Vol. 18 (2025)

An improved similarity matching model for the content-based image retrieval model

Manimegalai Asokaraj[1,2,*], Josephine Prem Kumar[3,*] and Nanda Ashwin[4,*]

[1]Research Scholar, Visvesvaraya Technological University, Belagavi, India

[2]Department of CSE, East Point College of Engineering and Technology, Bengaluru, India

[3]Department of CSE, Cambridge Institute of Technology, Bengaluru, India

[4]Department of CSE, East Point College of Engineering and Technology, Bengaluru

*E-mail: lathuramesh@gmail.com, d_prem_k@yahoo.com, nandaashwin@eastpoint.ac.in

Received for publication August 30, 2024.

Abstract

Content-based retrieval (CBR) is an essential process to retrieve images from databases based on metadata from the image. Metadata in an image refers to image colors, textures, and shapes, or any other important information that can be derived from the image itself. The goal of CBR is to search for the relevant image and retrieve it from databases. Many CBR models achieved reliable results in analyzing the image. However, the computation cost of image retrieval is a challenging task due to the growth of image traits. This paper presents an Optimized Hybrid Ensemble Model (OHEM) R-2,C-1. The proposed model aims to improve the process of similarity matching for the efficient analysis of the query image while minimizing computation time. The purpose of OHEM is twofold. First, OHEM analyzes the features within the query image and performs similarity matching within the database, achieving this with reduced computational complexity. Subsequently, in accordance with the established objective function, it identifies and evaluates the pertinent features. Two distinct datasets, ROxford and RParis, are utilized to assess the model's performance. Several assessment criteria, including F1-score, recall, precision, and computation time, have been used to assess the model. The computation and evaluated outcomes are compared to six distinct algorithms, such as CSM, equilibrium propagation (EP), DCM, GA-based IR, and IRT. The comparison findings suggested that the proposed method performs better than the other models. R-2,C-1.

Keywords

Deep learning, Image retrieval, Content-Based Image Retrieval, Evaluation metrics, Optimization

I. Introduction

Recent technological advancements have led to a rise in the use of digital cameras, smartphones, and the Internet. The amount of shared and stored multimedia data is increasing, which makes it a difficult research problem to search for or retrieve pertinent images from an archive. Any image retrieval model's primary requirement is to find and sort images that have a visual semantic relevance to the user's query. Most Internet search engines rely on text-based methods to obtain images, which necessitate the inclusion of captions. Digital images are used in various industries, including health care, design, education, and spatial imaging. Images can be stored and retrieved using a wide range of strategies and approaches. Still, most of these search engines rely on metadata (keywords, descriptions, and tags) [1].

Open Access. Published by Sciendo. BY-NC-ND
© 2025 Asokaraj et al., This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License https://creativecommons.org/licenses/by-nc-nd/4.0/

To process digital image data, the visual characteristics of the image must be extracted and represented in an organized manner. Content-based retrieval (CBR) is one of the widely used models for retrieving information from digital images. CBR expands on conventional information retrieval methods for digital image search by including information that can only be found by directly examining the image's pixel values [2–3].

Content-based image retrieval (CBR) method is used to extract image features automatically by deriving the attributes, including color, texture, and shape. Extracting the features from the image is a tedious task and a multi-step process that includes various phases such as feature extraction, pattern matching, extracting queries, and matching queries. The general CBR system's process is visualized in Figure 1. The CBR process can be thought of as a simple collection of modules that work together to obtain images and process the comparison in the databases, R-2,C-4. Multi-dimensional feature vectors are used in standard content-based image retrieval systems to extract and describe the visual content of the images stored in the database [4]. A feature database is created from the feature vectors of the images stored in the database. The number of accumulated images is growing rapidly due to the advent of technology, web users, and the availability of recent mobile technologies such as digital cameras and image scanners [5]. Users in a variety of fields, such as publishing, architecture, crime prevention, fashion, remote sensing, and health, need reliable software or applications to handle digital images that are generated from the Internet and smartphones. Numerous general-purpose image retrieval systems have been designed with this goal in mind. Some of the CBR systems have achieved considerable results in retrieving the image from the database, and the learning model uses a similarity matching model to match features from query images [6]. Various models have contributed effectively to achieving reliable results in retrieving the image features. This research study validates various facets of the images, such as image shape, image colors, texture of the images, and image contents, to comprehend the statistical reasons [7]. However, there is a considerable research gap that currently exists to retrieve features between low-level and high-level characteristics, pertinent pattern analysis, problems with data modification, and retrieving images in the least computation time.

Many of the CBR models are developed based on a similarity-matching approach to retrieve images from the database. To create content-based image retrieval systems, contemporary research incorporates various content-based image retrieval algorithms. To improve the accuracy of the image retrieval process, a novel multiple-feature extraction from the query image is proposed [5]. A content-based image retrieval system was constructed using the color auto-correlogram feature, Gabor wavelet feature, and wavelet transform feature together. The improved precision is evident from the results, which show that

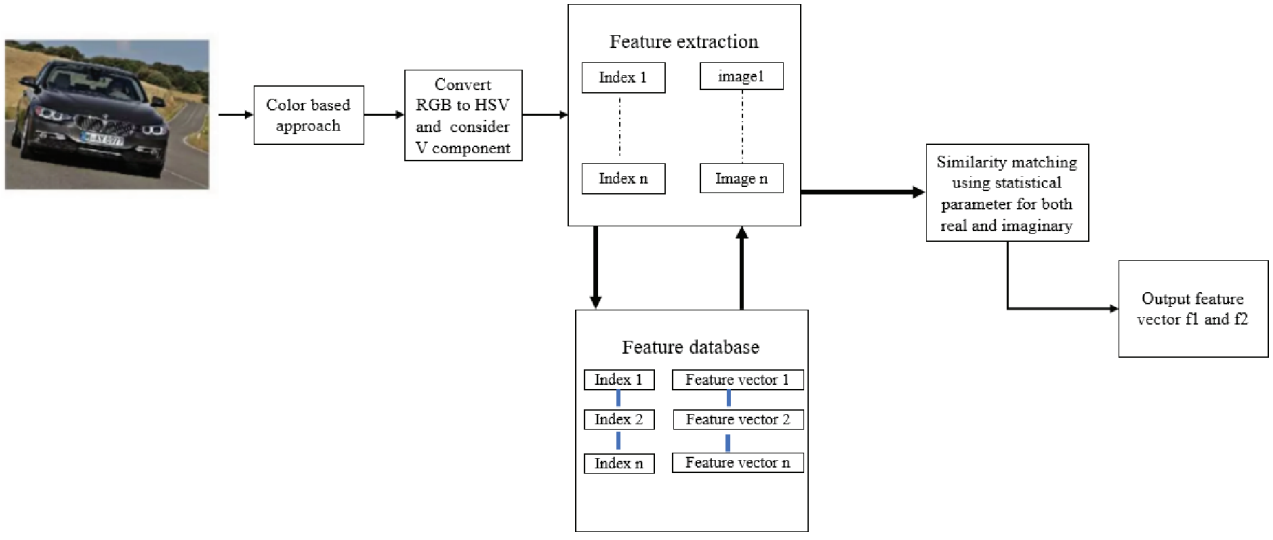

Figure 1: General workflow of CBIR.

multiple feature extraction creates a path to retrieve optimal images from the database effectively. Due to computational overhead, systems utilizing various feature extraction techniques may see a drop in retrieval speed as accuracy increases. Enhancing the system's performance is crucial because it boosts the system's accuracy. Computation time and semantic gap are two categories of CBR problems. The semantic gap refers to the difference that exists between low-level image pixels and high-level semantics that are understood by humans [8–10]. The other problem is computation time, or the amount of time needed to analyze, index, and search through images [11]. The color feature is extracted based on color string coding. Then, we compare one string with another string, and finally, their matching weights are returned. As the process is automated using RPA technology, this improves the perfection of images, which has significantly increased the accuracy while retrieving images that are appropriate to the given input image [29]. The recent advancements in deep learning-based image retrieval, starting with an introduction to the CBIR problem, key datasets, and content-based deep image retrieval methods, with a focus on network models, deep feature extraction, and retrieval types [39]. Finding appropriate methods for picture categorization, prediction, and retrieval systems is a challenging task for researchers. The research aims to develop a dynamic CBR model for retrieving optimal images from databases based on query image features with low computational time. Objectives of the research study can be achieved as follows:

1. To investigate appropriate methodologies for query image techniques and ascertain those exhibiting minimal computational duration.
2. To examine pertinent predictive and retrieval methodologies to recognize those characterized by superior performance metrics.
3. To amalgamate the recognized classification, predictive, and retrieval methodologies to establish a highly efficient case-based reasoning (CBR) system.R-2,C-8

a. Related work

In a vast database, it may be difficult to locate a specific image object due to the wide variety of image formats available on the Internet. A technique used in many domains is the retrieval of comparable images based on differences in the content of query images. Digitally knowledgeable libraries, criminal prevention, fingerprint identification, biodiversity information systems, health care, historical site research, and more are some of these domains.

CBR is a unique method of image retrieval that deviates from keyword-based techniques by emphasizing the examination of visual characteristics present in images rather than depending only on predefined keywords. A method called CBR makes use of visual cues, including color, form, and texture, to help with the difficult task of identifying visual objects [12].

CBR is one of the many applications that fall under the broad category of computer vision. The retrieving images from a database with many images is known as CBR [17]. This paper aims to investigate the real-world implementation of a two-phase method for content-based image retrieval. In the first stage, convolutional neural networks (CNNs) are used for image detection. CBR is one of the many applications that fall under the broad category of computer vision. The method of retrieving images from a database with a large number of images is known as CBR. This paper aims to investigate the real-world implementation of a two-phase method for content-based image retrieval. In the first stage, CNNs are used for image detection [14,30]. Three different levels of clouds are distinguished by the cloud classification system: high, intermediate, and low. K-means clustering methods and content-CBR are used in the cloud categorization process. Clouds are divided into three different categories by the developed method: low, medium, and high. The kind of cloud has a significant impact on how much precipitation falls. There is a lack of established research and inconsistency about the impact of high resolutions on search precision and result arrangement. This [15] study's main goal is to investigate how picture resolution affects search accuracy and result sorting. Resizing images is highly advised before uploading them to the image database, particularly if the images are affected in any way by resizing.

To improve image retrieval and identification through content-based image recognition, Das et al. presented a technique for feature extraction through image binarization. A total of 3,688 images from two public datasets were used by the authors to test their technique. This procedure, irrespective of the image dimensions, decreased the size of features to 12. For the aim of evaluation, statistical measures based on recall and precision findings were used. Misclassification of query images is one drawback of this methodology, which could impact retrieval performance compared to other available techniques [16].

Jianbo Ouyang et al. [31] proposed a re-ranking method that refines top-K retrieval results by

encoding each image into an affinity feature vector based on comparisons with anchor images. These features are refined using a transformer encoder to incorporate contextual information and then used to recalculate similarity scores for re-ranking. Our approach, enhanced with a novel data augmentation scheme, is robust and compatible with results from various retrieval algorithms. Ashraf et al. [13] proposed a unique feature extraction model for CBR, which reliably retrieves object information contained in an image by the bandelet transform technique for image representation and feature extraction. Three public datasets, such as Coil, Corel, and Caltech 101, were utilized to evaluate the system's performance and accomplishments in image retrieval using artificial neural networks. The retrieval efficiency was evaluated using precision and recall values.

This CSM algorithm applies this concept to a contrastive function that is created by gently pushing the multilayer similarity-matching goal function's output neurons. Consequently, between their earlier and later levels, the hidden layers pick up intermediate representations.

This model was inspired by deep learning techniques, such as single feedforward networks and networks with feedback connections. The objective function in CSM is determined with the help of a hyperparameter of anti-Hebbian and Hebbian plasticity to manage neuron weight to retrieve features from the query images [18].

In equilibrium propagation (EP), weight updates in selecting the features are defined with a gradient proportion value of the error signal, which aids in comparing the parameter tuning of a local approximation phase. The approximation phase of propagation is primarily characterized by the nudged equilibrium phase (NP) and clamped phase. One way to conceptualize the learning process is minimization in the approximation phase by the contrastive objective function. Another phase of EP is the reconfiguring phase, which updates the landscape matrix score to an energy score, which helps in removing false rigid features to increase the stability of the fixed points of weights. However, the model performed well in selecting features but failed in matching the image content due to the presence of large local features [19].

DCM's extremely easy method of obtaining a representation appropriate for geometric verification from CNN activations for re-ranking. To align the activation tensors and compare, it should ideally estimate the geometric transformation. However, as mentioned earlier, this is not feasible. DCM uses two characteristics of the activations—that high values are more significant and that the activations are sparse—to simulate this process. Therefore, a limited number of extremal regions can accurately mimic each channel [20,37].

Girgis et al. [21] suggested a GA-based IR technique that modifies a query's keyword weights to produce an optimal or nearly optimal query vector. Every query described in this method is represented by a chromosome. The method generates a new population by performing the genetic operator processes of selection, crossover, and mutation on the present population. Each member of the new population is then subjected to a local search approach to obtain optimal solutions. Until an optimal query chromosome for document retrieval is achieved, this process is repeated. Fitness functions, which evaluate the quality of a potential answer, and direct the evolution of those possibilities.

A technique for image retrieval utilizing statistical tests, such as Welch's *t*-tests and F-ratio, was proposed by Seetharaman and Selvaraj [15]. We examined both input query images, which were either textured or structured. In the experiment, the textured image considers the full image, but the structured image divides the shape into several parts according to its characteristics [23]. Applying the F-ratio test is the first stage of the exam, and images that pass this stage proceed to the energy spectrum testing. The determination that the images are comparable was made if they passed both tests. If not, they are distinct. The performance was validated and verified using the mean average precision score [22]. A component of the case-based reasoning scenario is similarity-based image retrieval, which has grown in importance within CBR. Similar images from a database are retrieved using similarity-based retrieval, often in order of similarity. The two main problems with CBR systems are the semantic gap, or the difference between semantic-based and content-based image retrieval. Similarity matching can be crucial to the task of identifying the important features from the query image. The primary problem has been the semantic gap that exists between low-level image pixels and high-level semantics that people understand [24]. An efficient CBIR system using pre-trained CNN models, VGG16 and ResNet-50, to extract deep features and improve retrieval performance. It uses transfer learning from ImageNet, enhancing both accuracy and efficiency, with results evaluated based on precision [38]. Existing deep learning methods often struggle with large scene data due to insufficient consideration of spatial dependencies, both inter-image and

intra-image. Spatial-driven network (SDN) captures discriminative features by addressing these dependencies [36]. The other problem is computation time, or the amount of time needed to analyze, index, and search through images. The HFFR-SR net integrates hierarchical feature representations. It uses a novel transformer block to expand the receptive field and a lightweight CBAM attention module to enhance high-frequency details, resulting in excellent results in both quantitative and subjective evaluations [40]. AutoLoss-GMS is a method for automatically discovering an optimal loss function within the space of generalized margin-based softmax loss functions for person re-identification. It uses a forward method to generate richer loss function forms compared to existing backward methods and introduces cross-graph mutation to enhance diversity, along with a loss-rejection protocol, equivalence-check strategy, and a predictor-based promising-loss chooser to boost search efficiency [34]. Wu et al. [33] proposed a flexible contextual similarity distillation framework that trains the small query model with a novel similarity consistency constraint, ensuring compatibility with the large model's output without requiring labels. This approach preserves both first-order feature representations and second-order ranking relationships, achieving state-of-the-art results on the Revisited Oxford and Paris datasets.

b. Similarity matching

Similarity between two or more images is defined using similarity metrics. The degree to which the content of the retrieved images resembles the user query determines their ranking. The image features are converted into vectors and represented by Raj and Mohanasundaram [25], using the vector space model, and the weights of the terms in the vector were calculated using the TFIDF measure. To determine how similar the image features are, the authors used cosine and Jaccard similarity measures. Various similarity measures were tested by Sutojo et al. [26] for text categorization and grouping. For image classification, they tested three similarity metrics, including cosine, Euclidean distance, and similarity measures. When compared to other similarity measures, they found that the performance of the similarity measure was comparable to that of the conventional model. Similarity matching in content-based image retrieval (CBR) refers to the process of finding images in a database that are like a query image based on their content features. CBR systems analyze the visual content of images to perform retrieval tasks, rather than relying on metadata or annotations. Similarity mapping in CBR typically involves calculating the similarity between feature vectors representing images, which is expressed in Eq. (1) R2,C-7.

$$SM = \frac{1}{N^2}\sum_{N=1}^{N}\sum_{n'=1}^{n}\left[\left(X_N{}^n X_{n'}{}^i - Y_N{}^i Y_{n'}{}^i\right)^2 + (Y_N{}^n Y_{n'}{}^i - Z_N{}^n Y_{n'}{}^i)^2\right] \quad (1)$$

where $X_N \in F^N$, $F^N \rightarrow$ is the corresponding output, and training data inputs $N = 1, 2, 3, \ldots, n$, be the total number of data points generated, corresponding to the output F. The data input Y_N, Z_N is the input vector within a network layer. Whereas $X_{n'}{}^i, Y_{n'}{}^i, Z_{n'}{}^i$ are the Bernoulli trial derivatives for the corresponding inputs. Training data X_N generate the similarity values of each feature with a non-trivial solution to the output Y_N.

c. Proposed method

Similarity matching is an important step in query processing because the feature generator vector consumes much time in processing feature comparisons, which degrades the computation time of the query processing. The query of the image and the database's collection of images are compared once the features are extracted from the image. In the suggested paper, the Euclidean distance is used to determine how similar the two images are to one another. Eq. (2) defines the Euclidean distance formula. R-2,C-7. This research study develops the optimized hybrid ensemble model (OHEM) by considering the limitations of standard deep learning of heavy architecture described in Algorithm 1. It combines heterogeneous retrieval techniques with the ensemble architecture of many deep learning models, resulting in increased efficiency. OHEM reduces the computation time in image retrieving by using the objective function $F(\Phi)$, which can be redefined in Eq. (3). R-2,C-7.

$$ED = \sqrt{\sum_{i=1}^{n}(x_1 - x_2)^2} \quad (2)$$

$$F(\Phi) = \min \frac{1}{N^2}\sum_{N=1}^{N}\sum_{n'=1}^{n}\left[\frac{\left(X_N{}^n X_{n'} - Y_N{}^n Y_{n'}\right)^2}{R^I\left[X_{n'} - Y_{n'}\right]} + \frac{(Y_N{}^n Y_{n'} - Z_N{}^n Z_{n'})^2}{R^I\left[Y_{n'} - Z_{n'}\right]}\right] \quad (3)$$

$$R^I = \max_{0 \le x \le 1}\left\{xe^{-x^2}\right\},\ \beta$$

where, $X_N \in F^N$, $F^N \rightarrow$ is the corresponding output $N = 1, 2, 3, \ldots, n$, be the total number of data points. The proposed function represents the learning difference of R^I to the product of input $X_{n'}$ and desired output $Y_{n'}$. The objective of the cost function is to achieve non-trivial solution, which leads to the

interaction of different data points in images. R^{l} of the square value of the output feature, which represents the similarity score of the input vector and the output vector scores. The quantization function's centroids are data points that describe the space's structure. The image set is kept in a frozen state while the query set is trained. Every training sample is mapped into two distinct embeddings by the query and image set. The structural similarities are then ascertained by calculating the similarities between the two embeddings and contrasting them with the centroids. In conclusion, by limiting the degree of consistency between two structural similarities, this strategy seeks to improve the query set. The proposed model reveals a thorough definition of the embedding space and chooses the appropriate data points in the retrieved model. These data points are embedded space references that translate gallery and query features into comparable architectural forms. To create a collection of data points, OHEM uses a similar methodology. According to Eq. (4), a vast number of data points p_x (x = 1, 2, 3, …, n) and py (y = 1, 2, 3, …, n) are needed for this method to accurately characterize the space structure $X_1|$, $X_2|$, $X_3|$, …, $X_n|$ R-2,C-7. Due to their similar nature, the query set and the image set data points show a high degree of alignment in their embedding spaces after training. Input query image convert into matrix, and the value of the image features is calculated based on the segmentation score as defined:

$$R^{l} \equiv (p_x \text{ and } p_y)\, px$$

$$\begin{matrix} X_{1'} & X_{2'} & X_{3'} & & X_{n'} \end{matrix}$$

$$P_x = \{1, 2, 3 \ldots P_n\} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \ldots\ldots \rightarrow \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix} py \quad (4)$$

$$P_y = \{1, 2, 3 \ldots P_n\} \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} \ldots\ldots \rightarrow \begin{bmatrix} 2 & 2 \\ 1 & 4 \end{bmatrix} \quad (5)$$

Evolved from Eqs (4) and (5) contain pixel values p_x and p_y with parameter values. Let us assume, suppose input image $X_{n'}$ contain pixels px and py with respect to the directions vertical v_i and horizontal h_i. Assume P_x = {1, 2, 3, …, P_n} is a horizontal domain space whereas vertical space domain is P_y = {1, 2, 3, …, P_n} are the product function $F(\Phi)$ in given β R – 2,C – 7. Parameter β is an optimization minimax function used to fix bias in the image query match. Considering all defined parameters, this research formulates the proposed function, which is defined as:

$$F(\Phi)\,(p_x, p_y/\beta)$$

$$= \#\,\{(X_N^{n}, P_x), (X_{n'}, Px), (X_N^{n}, Y_N^{n}, Z_N^{n},) \in (P_x, P_y)\}$$

$$F(\Phi)\,(p_x, p_y/\beta)$$

$$= \#\,\{(X_N^{n}, Y_N^{n}, Z_N^{n}) \in P_x, P_y)\ \&\ X_{n'}, Y_{n'}, Z_{n'} \in \beta$$

$$= X_1^{'}, \beta - |P_x, P_y)| * X_2^{'}, \beta - |P_x, P_y), X_3^{'}, \beta - |(P_x, P_y)|$$

$$= \beta[X_1^{'}, \beta - |(P_x, P_y)| * X_2^{'} - |(P_x, P_y)| * X_3^{'} - |(P_x, P_y)$$

$$F(\Phi)\,(p_x, p_y/\beta)$$

$$= \beta\left[X_1^{'}, -\left\| \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \right\| * X_2^{'} - \left\| \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \right\| * X_3^{'} - \left\| \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \right\| \right] \quad (6)$$

Let $(X_N^{n}, Y_N^{n}, Z_N^{n},) \in ((P_x, P_y)$, i.e., the correspondin feature vector $Y_1^n \ldots Y_n^n$ is proportional to the feature component e^{2x+1} as defined in Eq. (6) $R-2, C-7$. The first stage involves using the features collected by the image set to train a feature space compressor (FSC). The eigenvalue decomposition of R^{l} yields the optimal $F(\Phi) < \{0 > (P_x, P_y), < 1\}$ when the top m modes are retained and the other modes are set to zero. Attaining a representational similarity matrix that is the average of the input and output layers, optimal P_x, precisely equals P_y) if m rank count + 1. The feature component P_x, P_y is noted as a contrastive minimization problem with optimal weights. To efficiently retrieve images from the database, this study proposed OHEM to integrate the concepts of contrastive learning and supervised similarity matching to create a biologically plausible supervised learning system. R-1 C-2. The training samples are needed for the adaptive clustering within the computational complexity with multiple centroids. The high clustering cost and the huge set of centroids are included. Here, the quantization function is used to increase the number of data points efficiently and economically. Here, the training data is indicated by (X_N^{n}, Y_N^{n}, Z_N^{n},) to facilitate the creation of data points. The objective function $F(\Phi)$ is used to determine the OHEM simultaneously. Compared to the minimax objective, Z_N from Eq. (1) is the proof of the problem statement that any identity component in the feature vector of the query image is always proportional to the objective function $F(\Phi)$.. R-1 C-2.

d. Proposed Algorithm-1

Input: Query images-N, sample inputs (X_N^{n}, Y_N^{n}, Z_N^{n},), training input: P_x, P_y

Output: processed images F(Φ)

for each query image are train in X_N^n, β **do**
for all Let $(X_N^n, Y_N^n, Z_N^n, \in (P_x, P_y))$ **do**
draw 4 * 4 matrix

$$X_1^{'}, \beta - |(P_x, P_y)|, X_2^{'}, \beta - |(P_x, P_y)|, X_3^{'}, \beta - |(P_x, P_y)|$$
$$= \beta \left[X_1^{'}, -\left\| \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \right\| * X_2^{'} - \left\| \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \right\| * X_3^{'} - \left\| \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \right\| \right]$$

the first feature vectors

$$X_1^{'}, \beta = (P_{x1}, P_{x2}, P_{x3} \ldots\ldots P_{xn})$$
$$Y_1^{'}, \beta = (P_{y1}, P_{y2}, P_{y3} \ldots\ldots P_{yn})$$
$$Z_1^{'}, \beta = (P_{z1}, P_{z2}, P_{z3} \ldots\ldots P_{zn})$$

for all $F(\Phi) \in (P_x, P_y)$ **then do**
#pairwise similarity between query images
for each X_1 which present (P_x, P_y)
count = count+1
end for
return $\{(P_x, P_y)/\text{count}\}$
else return $F(\Phi) < \{0 > (P_x, P_y) < 1\}$
end for
end for
return

$$F(\Phi) = min \frac{1}{N^2} \sum_{N=1}^{N} \sum_{n'=1}^{n} \left[\frac{\left(X_N^n X_{n'} - Y_N^n Y_{n'}\right)^2}{R^l \left[X_{n'} - Y_{n'}\right]} + \frac{(Y_N^n Y_{n'} - Z_N^n Z_{n'})^2}{R^l \left[Y_{n'} - Z_{n'}\right]} \right]$$

end for

II. Result

a. Dataset description

The studies were conducted using the widely recognized ROxford and RParis datasets in the field of image retrieval, and data can be freely accessible at https://paperswithcode.com/dataset/roxford [12] R-2 C-6. The Oxford Buildings Dataset comprises 5062 images collected from Flickr through a specific Oxford landmark search. Toto provides a thorough ground truth for 11 distinct landmarks, each represented by 5 potential queries, and the collection has been painstakingly annotated. This provides a set of 55 queries that can be used to assess image retrieval using content to ease the process. We categorize 5,062 into 10 categories. The RParis dataset consists of 501,356 geotagged images gathered from Panoramio and Flickr. The dataset was not gathered using keyword queries, but rather from a geographic bounding box that encompassed the whole dataset. The graphic on the right illustrates how the images have a "natural" distribution as a result. Due to the high percentage of unrelated images, including pictures of parties, pets, and other events, as well as the existence of near-duplicates and duplicates, the dataset is extremely difficult to work with. The details of the images are shown in Figure 2. The collection comprises 70 query images, divided into three categories: easy, medium, and hard. A mix of easy and difficult questions is included in the medium split, while the hard split is intended to tackle only the most difficult issues.

The precision, recall, and *F*-measure values are used to assess the suggested methodology. Precision is the metric used to quantify retrieval accuracy [11, 27]. Precision measures the ratio of all the images obtained to the number of relevant images retrieved, expressed as:

$$\text{Precison} \rightarrow Pr = \frac{\text{Total number of relevant image retrieved}}{\text{Total number of image retrieved}}$$

Recall serves as a measurement of the system's robustness in retrieval. The term recall refers to the proportion of relevant images that were successfully retrieved from all the relevant images stored in the database. Recall can be expressed as:

$$\text{Recall} \rightarrow Re = \frac{\text{Total number of relevant image retrieved}}{\text{Total number relevant image in Data base}}$$

The calculation of the F-measure indicates the unified performance measure. The *F*-measure is used to indicate how much recall is less important than precision and is expressed as:

$$F\text{measure} \rightarrow F = \frac{(1+\alpha) \times Pr \times Re}{(\alpha \times Pr) + Re}$$

where $\alpha = 0.25$ indicates that precision is more important than recall. Table 1 presents a tabulation of the resulting precision and recall values. Figure 2 displays the graph of the precision, recall, and *F*-measure values for these methods when used separately and in combination. The precision over the ROxford dataset achieved was% higher than CSM [18], EP [9], and nearly 7% higher than DCM [20] than the other conventional model shown in Table 1 and Figure 2. R-1,C-1 and recall are also greater than other methods [28]. The query image was processed and gained a reliable result to extract features that are fair enough to prove the proposed model outperforms

Figure 2: Sample query image.

Table 1: Comparison of precision and recall on the datasets

	ROxford			RParis		
Method	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
CSM [18]	83.29	71	73.58	84.28	71.93	78.83
EP [19]	83.12	70.1	71.45	84.56	72.09	81.23
DCM [20]	79.12	65.1	61.45	80.09	69.01	75.41
GA-based IR [21]	84.16	66.6	78.56	85.56	72	74.51
IRT[22]	80.14	69.5	71.26	78.41	68.47	73.25
OHEM	85.56	73.3	81.45	86.25	75.59	87.89

EP, equilibrium propagation; OHEM, optimized hybrid ensemble model.

other approaches. However, extracting the features from the query image is slightly slow when applying it to the database. The RPairs dataset also proves that the precision and recall values are higher than those of the other models. The significance value "α" is slightly adjusted with parameter L_1 R-1,C-1. First, it struggled to obtain the result; later, its training improved compared to that of other models. Recall, precision, and F1-score levels are visualized in Figures 3 and 4, which makes them very helpful for assessing systems when retrieving all pertinent images is crucial. The recall value is comparatively good than the other models by utilizing parameter adjustment with EP [19]. The F1-score of the proposed model achieved 81.45% and 87.89% on ROxford and RParis datasets, respectively, which is slightly better than the other five models. The overall performance results are tabulated in Table 1 and visualized in Figures 3 and 4. R-1,C-1.

The computation time to retrieve the image from the database is measured in terms of seconds. Each dataset contains a large number of images with unique features; OHEM worked well in retrieving optimal images from the database. The computational complexity measure through our local system, which has a configuration of 2.5 GHz speed, 8 GB of RAM, a 64-bit OS, and Windows 12. To streamline the process, we divided the images into five categories, and each category had an equal number of images. Each group of images is retrieved, and the time taken to retrieve is measured and counted as average time and visualized. The computational analysis study reveals that OHEM takes less time to retrieve the images from the database, plot, and visualize in Figure 5.

III. Conclusion

This research used an enhanced hybrid ensemble model referred to as OHEM to demonstrate considerable advancements in the domain of content-based image retrieval from databases, predicated on similarity matching principles. OHEM addresses the essential challenges associated with the management of large-scale images within the database, as well as the computational time required to retrieve analogous images R-1,C-3. A considerable component of OHEM pertains to the extraction of features through the integration of a hybrid ensemble model for the retrieval of images in a coherent fashion. Utilizing two distinct databases (ROxford and RParis) encompassing a variety of image types,

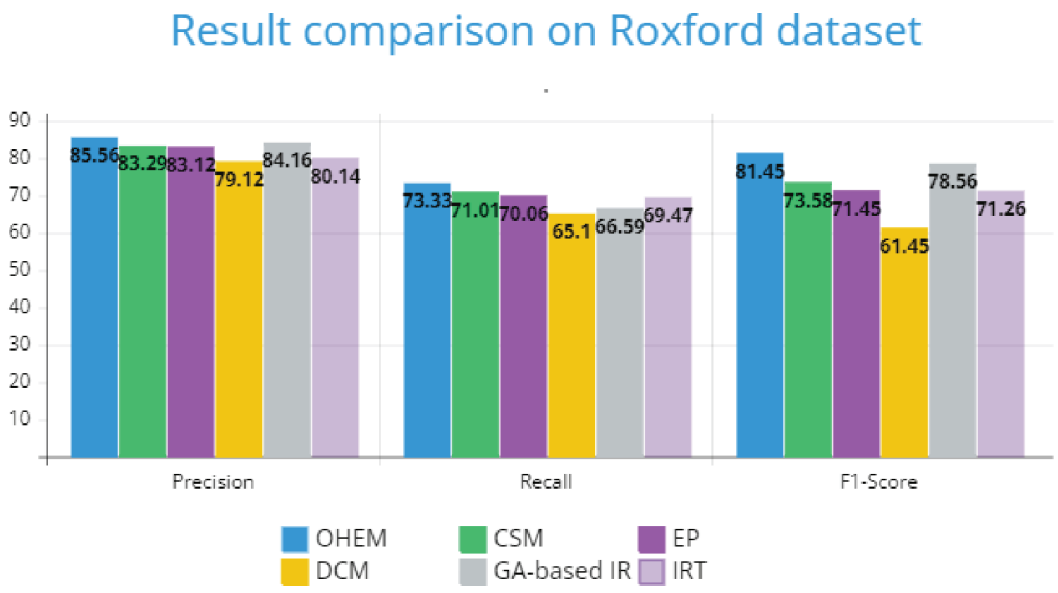

Figure 3: Result comparisons with different algorithms on the ROxford dataset. EP, equilibrium propagation; OHEM, optimized hybrid ensemble model.

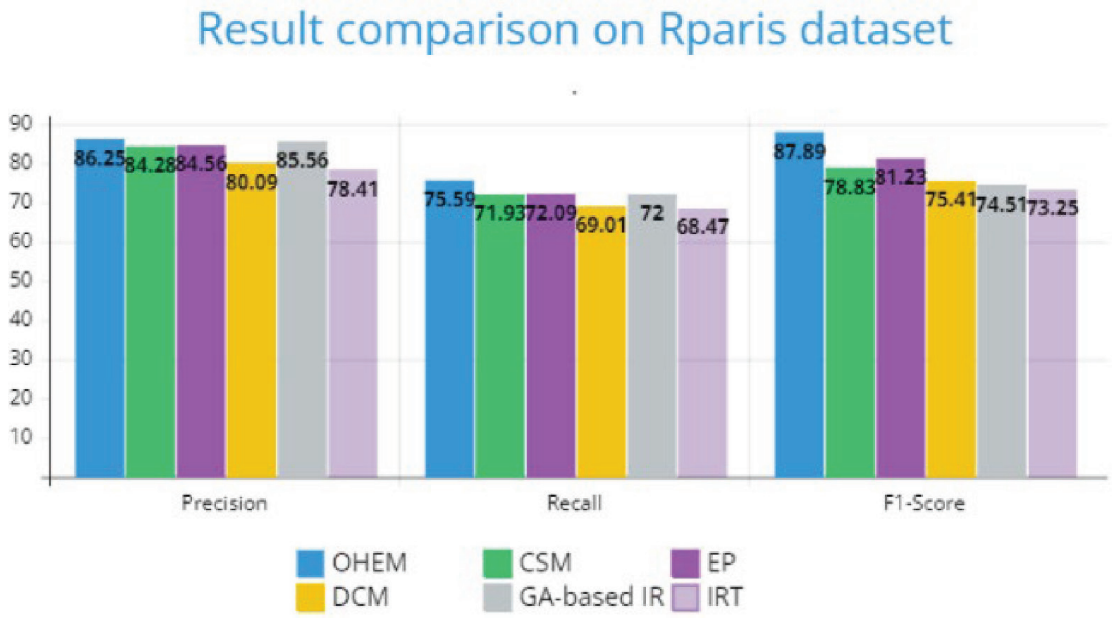

Figure 4: Result comparisons with different algorithms on the RParis dataset. EP, equilibrium propagation; OHEM, optimized hybrid ensemble model.

the proposed research evaluated five disparate algorithms (CSM, EP, IRT, GA-based IR, and DCM) R-1,C-2. Similar images may be accurately and quickly retrieved by entering a query image. Additionally, the proposed paper analyzes the precision and recall of each algorithm using both databases. It concludes that the OHEM algorithm has an accuracy of 85%, which is 3% greater than that of the CSM, EP, IRT, GA-based IR, and DCM R-1,C-1. Future papers on the system will incorporate more low-level feature images, such as spatial position and shape features, to fortify it. Semantically based image retrieval and image feature matching are the other two essential parts of the OHEM system.

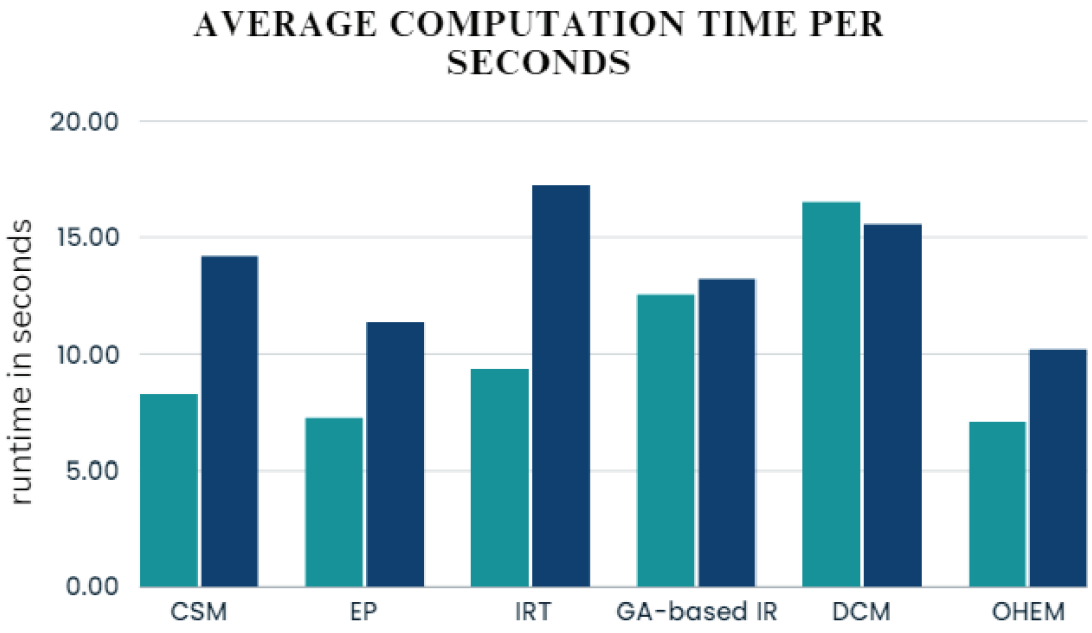

Figure 5: Comparison of computation time of different models on ROxford and RParis datasets. EP, equilibrium propagation; OHEM, optimized hybrid ensemble model.

References

[1] D. Srivastava, S. S. Singh, B. Rajitha, M. Verma, M. Kaur and H. -N. Lee, "Content-Based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation, and Evaluation Parameters," in IEEE Access, vol. 11, pp. 95410-95431, 2023, doi: 10.1109/ACCESS.2023.3308911.

[2] G. Sumbul, J. Xiang and B. Demir, "Towards Simultaneous Image Compression and Indexing for Scalable Content-Based Retrieval in Remote Sensing," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-12, 2022, Art no. 5630912, doi: 10.1109/TGRS.2022.3204914.

[3] Cengiz Pehlevan, Anirvan M Sengupta, and Dmitri B Chklovskii. Why do similarity matching objectives lead to hebbian/anti-hebbian networks? Neural computation, 30(1):84–124, 2018.

[4] Liu, G.-H.; Yang, J.-Y. Content-based image retrieval using color difference histogram. Pattern Recognit. 2013, 46, 188–198.

[5] Tian, D. Support Vector Machine for Content-based Image Retrieval: A Comprehensive Overview. J. Inf. Hiding Multim. Signal Process. 2018, 9, 1464–1478.

[6] Mehmood, Z.; Mahmood, T.; Javid, M.A. Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine. Appl. Intell. 2018, 48, 166–181.

[7] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "SURF: Speeded Up Robust Features." Computer Vision – ECCV 2006 Lecture Notes in Computer Science (2006): 404-17. Web.

[8] Lindeberg, T.: Feature detection with automatic scale selection. IJCV 30(2) (1998) 79–116.

[9] Mikolajczyk, K., Schmid, C.: Indexing based on scale invariant interest points. In: ICCV. Volume 1. (2001) 525–531.

[10] H. Tamura, S. Mori, and Y. Yamawaki, "Textural Features Corresponding to Visual Perception", IEEE Transactions on Systems, Man, and Cybernetics, SMC-8, (1978), pp. 460-473.

[11] P. Manipoonchelvi and K. Muneeswaran, Multi region-based image retrieval system, Sadhana Indian Acad. Sci. 39 (2014), 333–344.

[12] Radenović F., Iscen A., Tolias G., Avrithis Y., Chum O. Revisiting Oxford, and Paris: Large-Scale Image Retrieval Benchmarking, CVPR, 2018.

[13] Ashraf, R.; Bashir, K.; Irtaza, A.; Mahmood, M.T. Content Based Image Retrieval Using Embedded Neural Networks with Bandletized Regions. *Entropy* 2015, *17*, 3552-3580.

[14] Lin Feng, Jun Wu, Shenglan Liu, Hongwei Zhang, Global Correlation Descriptor: A novel image representation for image retrieval, Journal of Visual Communication, and Image Representation, Volume 33, 2015, Pages 104-114.

[15] Seetharaman, K. and Selvaraj, S., 2016. Statistical tests of hypothesis based color image retrieval.

Journal of Data Analysis and Information Processing, *4*(2), pp.90-99.

[16] Abhishek Jain, Aman Jain, Nihal Chauhan, Vikrant Singh, Narina Thakur, "Information Retrieval using Cosine and Jaccard Similarity Measures in Vector Space Model", International Journal of Computer Applications, Volume 164, No 6, PP.28-30, 2017.

[17] Komal Maher, Madhuri S. Joshi, "Effectiveness of Different Similarity Measures for Text Classification and Clustering", International Journal of Computer Science and Information Technologies, Vol. 7, No.4, pp.1715-1720, 2016.

[18] Obeid, D., Ramambason, H., and Pehlevan, C. (2019). Structured and deep similarity matching via structured and deep hebbian networks. In Advances in Neural Information Processing Systems, pages 15377-15386.

[19] D.M Raj, R. Mohanasundaram (2020), "A New Improved Filter-based Feature Selection Model for High-Dimensional Data" Journal of Supercomputing, Springer, volume No. 76, Issue No. 3, pp 5745-5762.

[20] O. Siméoni, Y. Avrithis and O. Chum, "Local Features and Visual Words Emerge in Activations," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 11643-11652, doi: 10.1109/CVPR.2019.01192.

[21] Moheb Ramzy Girgis, Abdelmgeid Amin Aly & Fatima Mohy Eldin Azzam "The Effect of Similarity Measures on Genetic Algorithm-Based Information Retrieval", International Journal of Computer Science Engineering and Information Technology Research. Vol. 4, Issue 5, Oct 2014, pp. 91-100.

[22] A. El-Nouby, N. Neverova, I. Laptev, and H. Jegou, "Training vision ´ transformers for image retrieval," arXiv preprint arXiv:2102.05644, 2021, doi. org/10.48550/arXiv.2102.05644.

[23] Y. Zhang, Q. Qian, H. Wang, C. Liu, W. Chen and F. Wang, "Graph Convolution Based Efficient Re-Ranking for Visual Retrieval," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2023.3276167.

[24] X. Zhu, H. Wang, P. Liu, Z. Yang, and J. Qian, "Graph-based reasoning attention pooling with curriculum design for content-based image retrieval," Image and Vision Computing, vol. 115, p. 104289, 2021, doi.org/10.1016/j.imavis.2021.104289.

[25] Raj, R. Mohanasundaram (2020), "An Efficient Filter-Based Feature Selection Model to Identify Significant Features from High-Dimensional Microarray Data" Arabian Journal for Science and Engineering", Springer, volume-45, pp. 2619–2630.

[26] T. Sutojo, P. S. Tirajani, D. R. Ignatius Moses Setiadi, C. A. Sari and E. H. Rachmawanto, "CBR for classification of cow types using GLCM and color features extraction," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 2017, pp. 182-187, doi: 10.1109/ICITISEE.2017.8285491.

[27] Chauhan, S., Prasad, R., Saurabh, P., Mewada, P. (2018). Dominant and LBP-Based Content Image Retrieval Using Combination of Color, Shape and Texture Features. In: Pattnaik, P., Rautaray, S., Das, H., Nayak, J. (eds) Progress in Computing, Analytics and Networking. Advances in Intelligent Systems and Computing, vol 710. Springer, Singapore. DOI: 10.1007/978-981-10-7871-2_23.

[28] Sengupta, A., Pehlevan, C., Tepper, M., Genkin, A., and Chklovskii, D. (2018). Manifold-tiling localized receptive _elds are optimal in similarity-preserving neural networks. In Advances in Neural Information Processing Systems, pp.7080-7090.

[29] Manimegalai A, Dr. Josephine Prem Kumar, "Automating Image Retrieval using Ulpath (RPA) by Extricating Color Feature with String comparison in CBIR", International Journal of New Innovations in Engineering and Technology (IJNIET), Volume14 Issue 2-July2020, pp.14-20.

[30] Manimegalai A, Sonika, Sunil KK, Sunil BA, Vishwas V, "Enhancing Signature Forgery Detection System using CNN-SVM", International Journal of Innovative Research in information Security, Volume 10, Issue 04, May 2024, DOI: 10.26562/ijiris.2024.v1004.33.

[31] H. Wu, M. Wang, W. Zhou, H. Li, and Q. Tian, "Contextual similarity distillation for asymmetric image retrieval," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9489-9498, 2022, doi: 10.1109/cvpr52688.2022.00927.

[32] Y. Song, R. Zhu, M. Yang, and D. He, "Dalg: Deep attentive local and global modeling for image retrieval," arXiv preprint arXiv:2207.00287, 2022.

[33] H. Wu, M. Wang, W. Zhou, H. Li, and Q. Tian, "Contextual similarity distillation for asymmetric image retrieval," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 9489–9498.

[34] H. Gu, J. Li, G. Fu, C. Wong, X. Chen, and J. Zhu, "Autoloss GMS: Searching generalized margin-based softmax loss function for person re-identification," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 4744–4753.

[35] G. Wu, X. Zhu, and S. Gong, "Learning hybrid ranking representation for person re-identification," Pattern Recognition, vol. 121, p. 108239, 2022.

[36] T. Si, F. He, H. Wu, and Y. Duan, "Spatial-driven features based on image dependencies for person re-identification," Pattern Recognition, vol. 124, p. 108462, 2022.

[37] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Learning generalizable omni-scale representations for person re-identification," IEEE Transactions on Pattern

Analysis and Machine Intelligence, vol. 44, no. 9, pp. 5056–5069, 2022.

[38] Panel Giriraj Gautam, Anita Khanna,"Content Based Image Retrieval System Using CNN based Deep Learning Models", Elsevier, Vol 235, Pages 3131-3141, 2024, DOI: 10.1016/j.procs.2024.04.296

[39] Chi Zhang, Jie Liu, "Content Based Deep Learning Image Retrieval: A Survey", ICCIP 2023: 2023 the 9th International Conference on Communication and Information Processing (ICCIP), DOI: 10.1145/3638884.3638908

[40] Yiwei Jia, Yiwei Jia, Shiyong Huang, Xueming Li, "HFFR-SR: Hierarchical Fusion Feature Representations for Super Resolution of Old Images", ICCIP 2023: 2023 the 9th International Conference on Communication and Information Processing (ICCIP), pages 1–5, DOI: 10.1145/3638884.3638885.