International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 739-750
https://doi.org/10.70917/ijcisim-2026-2731
Article

Intra-User Image Ranking Using Computation-Efficient Quantum Convolutional Neural Networks for Selecting Highly Relevant and Non-Redundant Images

Murlidhar Mouriya [1], P.sammulal[2]

[1]Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad (JNTUH), Kukatpally, Hyderabad, India
[2]Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad (JNTUH), Kukatpally, Hyderabad, India
Corresponding Author: Murlidhar Mouriya
Email: murli.cool927@gmail.com.

Abstract: Intra-user image ranking refers to the process of organizing and prioritizing images within a single user's collection based on their relevance and quality.The rapid growth of user-generated images has created a major challenge of organizing and identifying the most meaningful images within a single user's collection as low-quality and lack popularity information such as likes and views. To overcome these issues,Intra-User Image Ranking Using Computation-Efficient Quantum Convolutional Neural Networks for Selecting Highly Relevant and Non-Redundant Images (IUIR-CEQCNN-HRNI) is proposed.Computation-Efficient Quantum Convolutional Neural Networks (CEQCNN) captures complex image patterns using fewer parameters by handling high-dimensional image effectively.The method is used to select most relevant and non-relevant images from personal image collections.These features are then encoded into quantum states using angle encoding allowing multiple data values through quantum superposition. The QCNN applies parameterized quantum convolution layers composed of small entangled qubit circuits like convolution layers to capture local patterns, followed by quantum pooling operations that reduce number of qubits and retain most informative quantum states.Dynamic Binary Swordfish Movement Optimization Algorithm (DBSMOA) is used to solve optimization problems based on swarms where the hunting process of swordfish is simulated. It uses adaptive techniques of exploration and exploitation to dynamically evolve the candidate solutions and maps the movement to a binary decision process to the best possible subset. The model is tested using the Flickr image data set and the results show the model with 99.41% accuracy, 98.41% recall and 98.72% f1-score outperforming existing techniques by a significant margin. The proposed IUIR-CEQCNN-HRNI method provides reduced redundancy, improved image quality selection, higher ranking accuracy and efficient management of large personal image collections.

Keywords: Computation-efficient Quantum Convolutional Neural Networks, Dynamic Binary Swordfish Movement Optimization Algorithm, Deep feature embedding, Redundancy reduction, Pattern recognition, Visual similarity learning.

1. Introduction:

The rapid growth of digital platforms and social media has led to an exponential increase in user-generated visual content, resulting in large collections of images associated with individual users. With the proliferation of the internet and the emergence of various online communities, social networks and platforms witnessed an explosion in the number of images uploaded by users [1]. The management and organization of such large amounts of images are often a large amount of redundant, near-duplicate and low-quality images present in the image repository. Under such conditions, intra-user image ranking becomes a critical process that helps identify and sort out the most important images belonging to that particular user [2].Due to the fast development to the typical image retrieval systems which mainly emphasize the inter-user ranking, the intra-user ranking aims to improve the representation of an individual user's image group [3]. It plays a crucial role in some specific application scenarios; including social media feed presentation, personal image grouping and recommendation systems [4]. Neither the textual nor the visual image feature-based search deliver a perfect result. Regarding the

development of existing digital image re-ranking technologies, most existing re-ranking techniques do not deal with the semantic gap between the background visual feature and semantic level of the images.It created the similarity graph of the images based on their backgrounds [5].

With the rapid growth of user-generated images, managing large personal image collections has become challenging due to the presence of redundant and less informative images. The management of large image collections is now becoming quite challenging because of the huge increase in the creation of user-created images. Existing image ranking methods often focus only on visual features and fail to effectively combine with semantic and contextual information. Therefore, there is a need for intelligent systems that automatically rank and select relevant, high-quality and non-redundant images from a user's image collection. The proposed IUIR-CEQCNN-HRNI method is motivated by the need to effectively manage large collections of user-generated images by selecting only the most relevant and non-redundant content. It aims to develop systems that automatically select the best, unique and meaningful images in an efficient way.

The proposed CEQCNN introduce a quantum-inspired architecture that captures complex image patterns using fewer parameters. It integrates visual, semantic and artificially generated popularity features to achieve robust and meaningful image representation. It enables efficient and scalable processing of large image collections.The DBSMOA automatically tunes the network parameters to improve ranking accuracy and reduce redundancy. Overall, the method provides an efficient and scalable solution for selecting highly relevant and non-duplicate images.

The major contribution are given below,

- The paper proposes the IUIR-CEQCNN-HRNI framework that integrates perform accurate intra-user image ranking for selecting highly relevant and non-redundant images.
- The ANUKF improves image quality by effectively removing noise and stabilizing pixel variations. It enhances normalization and image clarity, leading to more reliable and better model performance.
- AQT enables fast and efficient extraction of complex visual and semantic features from images. It improves feature quality which helps enhance overall image ranking.
- The CEQCNN capture complex image patterns with fewer parameters and improve ranking accuracy and generalization making the system efficient and scalable for large image collections.
- The DBSMOA is utilized to optimize CEQCNN parameters, ranking accuracy and diversity.

Remaining manuscript is structured as follows: Section 2 outlines literature survey, Section 3 displays proposed method, Section 4 presents results and discussions, and Section 5 concludes this manuscript.

2. Literature Review

Several recent studies have presented on image re-ranking using deep learning. Some of the most relevant works are summarized below,

In 2025, Wei, et.al., [6] presented a dynamic visual semantic sub-embeddings (DVSE) and fast re-ranking for image-text retrieval.It acquires a series of heterogeneous visual sub-embeddings by means of a dynamic orthogonal constraint loss. To promote diversity in visual embeddings by assigning varying levels of importance to the learning process, a mixed distribution and variance-sensitive weighting loss function was used. In addition, a Fast Re-ranking approach (FR) was suggested. It provides high precision and low f1-score.

In 2024, Pivoňka, T. and Přeučil, L., [7] presented an on model-free re-ranking for visual place recognition along deep learned local features. It concerned with the issue of model-free re-ranking using conventional local visual features and how these can be applied to long-term autonomy systems. It takes a similar approach to model-free re-ranking, which has mostly been proposed for use with deep-learned local visual characteristics. These local characteristics highly resilient to various appearance modifications, which regarded as a crucial component of long-term autonomy systems. It provided low MAPE and ROC.

In 2024, Wang, K., et.al, [8] suggested a neighborhood re-ranking-base method for pedestrian text-image retrieval.It introduced Implicit Neighbourhood Reranking Network (INRNet) that adopts bilateral feature extraction scheme to learn global text-image alignment and utilizes nearby neighbors as prior information to select positive instances. The adopt bilateral feature extraction scheme to extract text and pedestrian image features and text-image alignment algorithm to get initial global alignment between texts and pedestrian images. Next it suggests Neighborhood Data Construction Mechanism (NDCM). It provided low RMSE and precision.

In 2024, Sabahi, F.,et.al, [9] presented a RefinerHash: a new hashing-basis re-ranking technique for image retrieval.A computational efficiency approach for re-ranking was introduced using the quick and efficient use of image hashing technique as a method for image representation. The proposed approach is introduced by applying an image's hash value based on its Discrete Cosine Transform (DCT) and Discrete Wavelet Transform

(DWT) values. Three balanced binary search trees are built: two based on the most and least significant bits of the hash codes from images using DWT values, respectively, and one using hash codes from images based on DCT values. It provided higher ROC and lower accuracy.

In 2025, Yang, X., et.al, [10] developed a Re-ranking through image description for object re-identification.It involves a framework for ReID (Re-Ranking through Image Description), which leverages deep semantic data for the purposes of improving re-ranking by using context-rich features. As such, this results in a significant improvement in the accuracy of object re-identification. The key contributions involved in generating a richly annotated dataset made up of real-world images used to train an object semantic extraction model. The development of framework for ReID that analyzes deeply the appearance similarity of object within an end-to-end match algorithm. It provides high recall and high MSE (Mean Squared Error).

Several works in recent literature have attempted various images ranking for relevant and non-relevant images. Wei et al. (2025) is characterized by increased diversity in visual embeddings through the use of dynamic orthogonal constraints and variance-aware weighting, re-ranking algorithms based on deep-learned local visual features were suggested by Pivoňka and Přeučil (2024); exhibit robustness to appearance changes yet pay attention only to local features. The INRNet method for pedestrian text-image retrieval, offered by Wang et al. (2024), achieves enhanced global alignment using neighborhood information yet lacks precision. Hashing-based re-ranking algorithm RefinerHash, developed by Sabahi et al. (2024), achieves increased computational efficiency at the cost of lower accuracy. The ReID model with deep semantic image descriptions by Yang et al. (2025), enhances re-ranking accuracy and recall yet demonstrates higher error values.The proposed IUIR-CEQCNN-HRNI method overcomes these limitations by combining visual, semantic, and popularity features for better representation. The integration of CEQCNN with optimization improves ranking accuracy and reduces redundancy. It shows more efficient and reliable intra-user image ranking system.

3. Proposed Methodology

The proposed IUIR-CEQCNN-HRNI method starts with input images acquired from the Flickr image data set and artificially popularity score generation. The obtained images are preprocessed using ANUKF in order to normalize, enhance the image, and eliminate noise. The preprocessed images are fed into the feature extraction process, where AQT is utilized to extract valuable information from the image. Further, the extracted features such as visual, semantic, and popularity are fused together through IRDGF method. The resulting fused features are used as input to the CEQCNN to obtain the ranked images. Ranking scores are assigned in relation to relevance and uniqueness. The parameters of the model are optimized through DBSMOA is used for tuning of the network to optimize the rankings.

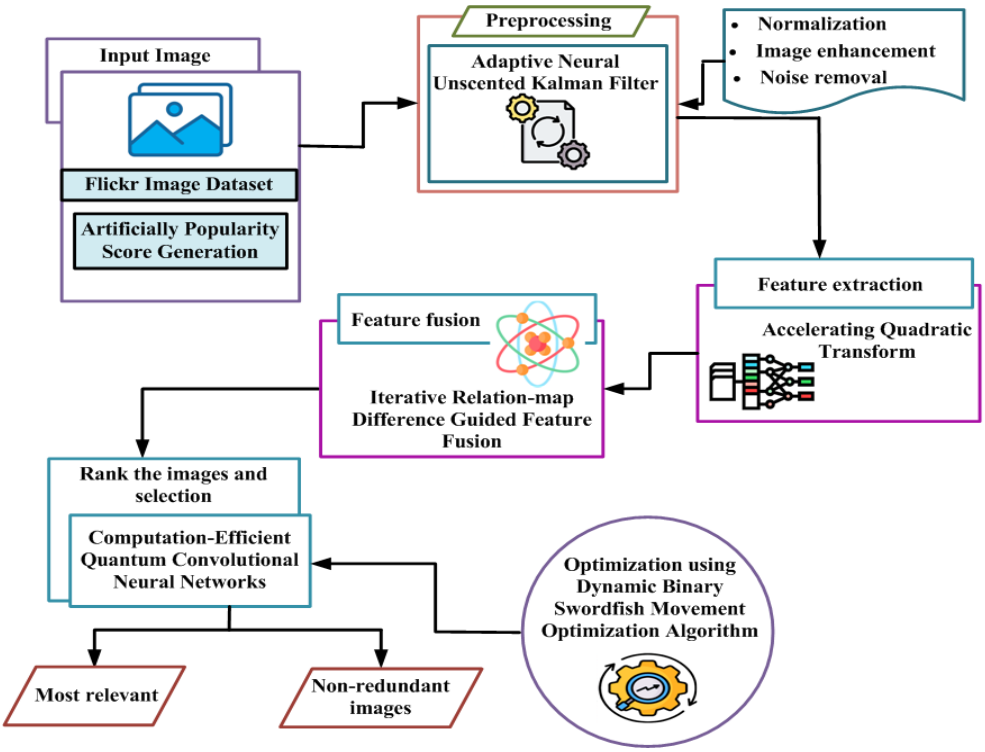


Figure 1: Block diagram of proposed IUIR-CEQCNN-HRNI method

3.1 Image Acquisition

The input images are collected via Flickr Image Dataset [11]. Flickr Image Dataset is one of the extensively employed multimodal datasets that comes from the Flickr30k dataset. It has about 31,000 real-life images along with roughly 158,000 human-generated image descriptions, with each image having many sentence descriptions. It also includes many other annotations, such as around 244,000 co-referenced chains and 276,000 bounding boxes that connect the text description to certain parts of images. It allows the researchers to gain more insight into the correlation between the images and languages. It is used for various purposes, like image captioning, image and text matching, object localization, and multimodal learning. The input images are divided as 70% for training, 20% for testing and 10% for validation.

Additionally, the dataset does not include popularity features such as likes or views, generate a popularity score is artificially generated using features extracted from the images.These features are used to compute a relevance or importance score for each image, which serves as a proxy for popularity.

3.2 Preprocessing using Adaptive neural unscented Kalman filter

ANUKF [12] is discussed to image enhancement, noise removal and normalization.It applied to achieve better image quality through effective noise and uncertainty handling. This method effectively uses the capabilities of networks along with the unscented kalman filter to achieve an adaptive state estimation and correction of nonlinear distortions in image. It achieves more accurate state estimation without linearization, which is required in other state estimation filters. It effectively preserves structural information removing noise during preprocessing resulting in refined and normalized images.The parameters of ANUKF, including the initial state vector, covariance matrix and noise parameters are initialized based on the input image characteristics. Then the image acquisition and sigma point generation is calculated in equation (1),

$$S_{ib}^{b} = S_{ib,t}^{b} + C_a + x_a \text{----------(1)}$$

Where,S_{ib}^{b} denotes sigma point generation at ibof image state atb; $S_{ib,t}^{b}$indicates initial or previous sigma point derived from the current state estimate; $C_a$signifies covariance adjustment factor at aand $x_a$indicates process noise factor. In this step, sigma points are computed based on the state estimate of the image. The computed sigma points are used to better capture the non-linear nature of the system compared to linear estimation methods. The generated sigma points help in predicting the possible variations in pixel intensity values. Then the prediction of image state is calculated in equation (2),

$$\partial y = K\,\partial y + H\,\partial w \text{---------- (2)}$$

Where, ∂yindicates prediction of image state; Kdenotes pixel size; Hsignifies the observed image and ∂wspecifies image resizing. At this stage, sigma points are propagated through a nonlinear transformation, resulting in an estimate of the predicted image state. This prediction involves assessing the pixel values is change, taking into consideration noise and system dynamics. Then the measurement update and noise reduction is calculated in equation (3),

$$W_{c+1} = G_{c+1} \cdot S_{c+1}^{*} \cdot G_{c+1}^{T} \text{ ----------(3)}$$

Where, S_{c+1}^{*}denotes measurement update and noise reduction;G_{c+1} represents the image variance of components; W_{c+1}signifies intensity of the image of components; G_{c+1}^{T}denotes distribution of the features and $c+1$ represents the pattern strength.Finally, the images is enhanced, denoised and normalized using ANUKF. Then the preprocessed images are fed into feature extraction phase.

3.3 Feature Extraction using Accelerating Quadratic Transform

In this section, AQT [13] is discussed to extract visual, semantic and popularity features from the images. The visual features describe the appearance of the images to extracts colour, Inverse and Correlation. Then the semantic feature denotes the meaning of the images. It identifies objects and scenes. Then the popularity features estimate the importance of the images used to generate image quality factors like brightness, sharpness and contrast.In addition, feature discriminability is enhanced through the use of nonlinear variations of intensity and spatial dependency of the region of interest. This way, more robust feature representations are obtained leading to enhanced classification and faster convergence of the learning models. This transformation enhances the texture and structural patterns present in the features. Then the feature space transformation using AQT is calculated in equation (4).

$$y_i^{*} = \left(\sum_{j=1}^{n} G_{ij} x_j x_j^{H} B_{ij}^{H}\right)^{-1} (R_i v_i) \text{ ----------(4)}$$

Where, y_i^*represents the image generated from the featuresi,G_{ij}denotes corner intensity of images ,x_j indicates images in some dimensionj, x_j^Hindicates to detected image jand edge informationH, B_{ij}^Hspecifies fused feature set obtained by aggregating different descriptors,$R_i$denotes strength of corners of i and v_i denotes binary mask logic. In this step, AQT then enables the identification of key features between pixels. This is done using higher-order statistical features that identify intensity dependencies and structural uniformity. Furthermore, inverse difference features are also identified to determine homogeneity in the segmented images. Then the extraction of visual features is exhibited in equation (5).

$$C_i^{k-1} = x_i^{k-1} + \rho_{k-1}\left(x_i^{k-1} - x_i^{k-2}\right) \text{ ----------(5)}$$

Where, C_i^{k-1}specifies textual features, x_i^{k-1}indicates processing of total image$k-1$, ρ_{k-1}denotes modulation of bias vector and x_i^{k-2} signifies total number of rounds. At this stage, visual features describe the appearance of the images to extracts colour, Inverse and Correlation. Then the semantic feature denotes the meaning of the images. It identifies objects and scenes. Then the popularity features estimate the importance of the images is generated using image quality factors like brightness, sharpness and contrast. Table 1shows extracted features using AQT

Table 1: Extracted features using AQT

Features	Equations	Description
Color distribution	$H(c) = \frac{1}{N}\sum_{j=1}^{N}\delta(I_j = c)$	$H(c)$ represents the normalized histogram for color c, I_j represents the color intensity at pixel j, N signifies number of leafs, δ implies indicator function
Inverse	$X = \emptyset^{-1}(F) = W^{+}F$	Where, X denotes reconstructed original image, $\emptyset^{-1}(F)$ specifies indicator function ; W^{+} indicates stability and F denotes feature vector
Contrast	$\text{Contrast} = \sum (p,m)^2\, h(p,m)$	Here, $h(p,m)$ represents input image
Correlation	$\text{correlation} = \frac{\sum_{h-0}^{F-1}\sum_{l=0}^{F-1}(l-n_h)(i-j_n)p(l,n)}{\delta_h\delta_h}$	Here, F represents the functions scale, l represents immediate estimate of frequency and δ represents total image.

Brightness	$\text{Brightness} = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} I(x,y)$	Where, M denotes number of rows in the image, N indicates number of columns in the image, $I(x,y)$ specifies intensity value of pixel at position at x,y
Sharpness	$\text{Sharpness} = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N} (\nabla^2 I(x,y) - \mu_L)^2$	Where,∇^2 indicates Laplacian value at pixel,μ_L denotes mean of laplacian values.

Finally, the visual features,semantic feature and popularity features are extracted throughAQT. Then the extracted features are given to the feature fusion.

3.4Feature Fusion using Iterative Relation-map Difference Guided Feature Fusion

In this section,IRDGF [14] is discussed toextract fusionfor accurate image ranking. The feature fusion technique enhances the image representation through fusion of many different kinds of features. The result is a more powerful feature set that considers visual cues, semantic knowledge, and popularity all at once. This is increases the accuracy of the ranking and reduces the redundancy by allowing differentiation between very similar images. Then the efficient feature learning is calculated in equation (6),

$$Q_i, F_i, V_i = F_i.\left[C_i^q, C_i^k, C_i^v\right] \text{ ----------(6)}$$

Where,Q_i denotesfeature dimensions at i,F_i specifies spatial regions,V_i indicates fused feature, C_i^q signifies captures both modality specific patters at q,C_i^k denotes cross-modal interactions at kand C_i^v represents short-term and long-term patterns across modalities at v. In this step,In CEQCNN, the process of learning highly dimensional patterns of images is accomplished via quantum-inspired convolutional layers. The network uses fewer parameters than conventional CNNs, hence saving computation costs. Then the accurate and scalable ranking is calculated in equation (7),

$$V_{p,b} = V_b + \tau_b \cdot V_t \text{ ----------(7)}$$

Where,$V_{p,b}$ denotes robust multimodal analysis at p,V_b indicates cross-modal interactions at b,τ_b specifies scaling factor and$V_t$represents fusion of image.CEQCNN efficiently identifies relevant images and eliminates any redundant ones. Theextracted features are fused byIRDGF. Finally, thefused features are fed into CEQCNN.

3.5 Image ranking and selection using Computation-Efficient Quantum Convolutional Neural Networks

In this section, CEQCNN [15]is discussed to ranking image and selecting most relevant and non-redundant images. It is capable to providing efficient feature in terms of reduced computation costs through fewer parameters that capture complicated image features. The model is able to handle high-dimensional images, thus providing accurate ranking results. It is also capable of reducing redundancies within similar images. This attributes from the three domains is then used as an input into the CEQCNN. This feature vector is comprehensive since it contains information about the quality and content of the image. Then the feature learning using quantum convolution layers is calculated in equation (8),

$$|\varphi\rangle_i = T_Y(y_i^r)R_X\left(x_i^g\right)T_z\left(y_i^b\right)|0\rangle \text{ --------(8)}$$

Where, $|\varphi\rangle_i$ denotesmodulation feature map at i,T_Y specifiesquantum feature encoding at Y,y_i^r signifiesspatio-temporal feature learning at j,R_X denotesprogressive multi-Stage learning at x,x_i^g indicatesdenotes input feature map extracted from convolutional layers at g,T_z denotesfunction linear at zand y_i^b specifiesnormalized output feature at i. In this step,the CEQCNN analyzes the fused features using quantum-

inspired convolutional layers. This stage enables learning and facilitates of dimensional representation learning within the system. Then the image ranking score generation is calculated in equation (9),

$$|\beta_\theta\rangle_i = U_T(\theta)|\beta\rangle_i \text{ ----------(9)}$$

Where, $|\beta_\theta\rangle_i$ denotesrepresents batch variance and batch mean,U_T indicates quantum convolutional operations at T,θ specifies attenuation function and$|\beta\rangle_i$ signifiesmodulation feature map i. At this stage,once all images are analyzed and modelled, the CEQCNN ranks the images depending on their relevance and uniqueness. Images that have high levels of significance and high quality are ranked highly. Then the selection of non-redundant images is calculated in equation (10),

$$\langle H\rangle = \{\langle\langle H(|\beta_\theta\rangle_i)\rangle \; i\epsilon N(0, F_0 \times W_0)\} \text{ ----------(10)}$$

Where,$\langle H\rangle$ specifiesconvolutional operation,β_θ denotesoutput feature map at the current stage,iϵN specifiesinput feature map, F_0 denotesmapping function andW_0 indicatessubstitution values.In the last stage of the ranking process, all images are arrangedas descending order according to their rankings. This enables removal of any duplicates to increase diversity.Finally, the CEQCNN identified ranking image and selected most relevant and non-redundant images. DBSMOA is utilized for improving CEQCNN parameters β and $W_0$by effectively tuning its weights and biases.

3.6 Optimization using Dynamic Binary Swordfish Movement Optimization Algorithm

DBSMOA [16] is used to enhance the parameterβ and $W_0$of proposed CEQCNN.The DBSMOA ensures the global search process is carried out effectively by striking the balance among exploration and exploitation during the process of optimization. It automatically adjusts parameters of the network to enhance its performance. Optimization makes it possible for the model to overcome the problem of getting stuck in local minima and enhances its convergence rate. Therefore, the convergence speeds, detection accuracy, stability of the optimized CEQCNN model are improved. Then the initialization is exhibited in equation (11).

$$\vec{S}_i = [d_{i1}, d_{i2}, \ldots\ldots, d_{id}] d_{ij} \in \{0,1\} \text{----------(11)}$$

where, $\vec{S}_i$specifies number of initial processes of i and $d_{i1}, d_{i2}, \ldots\ldots, d_{id}$indicates optimal applicant at iteration. At first, the DBSMOA initializes a population of candidate solutions for the parameters of CEQCNN. Each position of the swordfish is converted to a binary format to facilitate the search process in a high-dimensional search space. The weight β and $W_0$is generated randomly using method based upon obvious hyper parameter circumstances. Random solution is created from initialization. Fitness function is assessed for optimizing weight parameter with the help of parameter optimization value expressed in equation (12).

$$\text{Fitness function} = \text{optimizing } (\beta \text{ and } W_0) \text{ ---------- (12)}$$

Where, βdenotes increase accuracy, $\omega_0$denotes decrease MSE. In this step, each weight vector for the candidates is calculated by a fitness function related to predictive accuracy of tumor properties. At this stage, candidate solutions are assessed and evaluated against the fitness function. This fitness function is dependent on the accuracy of detection, loss value, and rate of convergence. This stage of DBSMOA identifies highly performing positions of the swordfish. Such positions provide better object detection outcomes. Less desirable solutions are discarded over time, and better solutions are kept for further processing. Then the dynamic movement and binary update mechanism is calculated in equation (13).

$$\vec{b}_i(t+1) = \vec{T}_i(t+1) + \beta.\vec{M}_i(t+1) + (Z.K).\vec{J}_i(t+1) \text{----------(13)}$$

Here, $\vec{b}_i$indicates dynamic movement and binary update mechanism, $\vec{T}_i$specifies the value of the step control, Z indicates updated position,$\vec{M}_i$denotes factor of adaptive control , Z. K specifies random influence ,$\vec{J}_i$denotes iteration progress and $t+1$indicates control value. In this step, DBSMOA imitates the attack and hunting process of the swordfish. This process involves dynamic movement of the positions of the swordfish. This movement strategy adjusts the parameters of the network. This process avoids local optima and enhances the rate of convergence to global optimum.Then the optimal parameter integration is expressed in equation (14).

$$w(\vec{S}) = w_0 \cdot E(\vec{S}) + (1-\delta) \cdot \frac{|\vec{S}|}{X} \text{ ----------(14)}$$

Where, $w(\vec{S})$denotes optimal parameter integration, $E(\vec{S})$signifies personalized weights parameter, $1-\delta$specifies convergence matrix, $|\vec{S}|$specifies update operation and Xdenotes computational time. Finally, the best-optimized parameter set is integrated into the CEQCNNmodel. The refined network achieves improved precision and stability in detecting objects from outdoor natural scenes.The weight parameter is used to optimize the weight parameter βdenotes higher accuracy, $\omega_0$represents low MSE the generator values from DBSMOA. The repeated

repeatedly until the stopping requirement $s = s + 1$is satisfied. Finally, DBSMOA optimized the weight parameters of CEQCNNwith better accuracy by reducing error rate.

4. Result and Discussion

This section discusses the experimental results ofIUIR-CEQCNN-HRNI approach. The implementation is carried out in Python utilizing 2.80 GHz Intel(R) Core(TM) i7 CPU M60. The IUIR-CEQCNN-HRNI method is evaluated using the mentioned metrics. The efficacy of IUIR-CEQCNN-HRNI method is compared with existing techniques: visual semantic sub-embeddings and fast re-ranking for image-text retrieval (DVS-FRR-ITR) [6], Model-Free Re-ranking for Visual Place Recognition including Deep Learned Local Features (MFRR-CPR-DL) [7] and Neighborhood Re-Ranking-Based Model for Pedestrian Text-Image Retrieval (NRR-PTIR) [8].

4.1 Performance Measures

The performance of the IUIR-CEQCNN-HRNI approachis determined by the following metrics.

4.1.1 Accuracy

The rate of appropriately predicted instances between allprocesses. It reflects effectively a model or system accurately classifies ranking. This is given in equation (15),

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{----------(15)}$$

Where, TP implies true positive, TN represents true negative, FP signifies the false positive and FN denotes false negative.

4.1.2 Precision

It computes the rate of true positive among all predicted positives using equation (16),

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{---------(16)}$$

4.1.3 F1-score

This metrics consolidates precision and sensitivity into a single value by determining their harmonic mean. It reflects a balance among appropriately predicted positive instances and the model's capacity to find out each actual positives making it especially useful for imbalanced datasets. This is measured by equation (17),

$$f1 - \text{score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad \text{---------(17)}$$

4.1.4 Recall

It computes the ability of classification method to appropriately detect each real positive samples. It determines the proportion of actual threats appropriately detected by the system using equation (18),

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad \text{----------(18)}$$

4.1.5 Mean Squared Error

It calculates the average of the square of the differences between the estimated value and the actual value using equation (19),

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \text{----------(19)}$$

Here, N denotes number of samples, y_i indicates actual value and $\hat{y}_i$ specifies predicted value.

4.2 Performance Analysis

The proposed IUIR-CEQCNN-HRNI methodology demonstrates the significant improvements in metrics. Figures 2-5 show the performance analyze of the proposed approach.

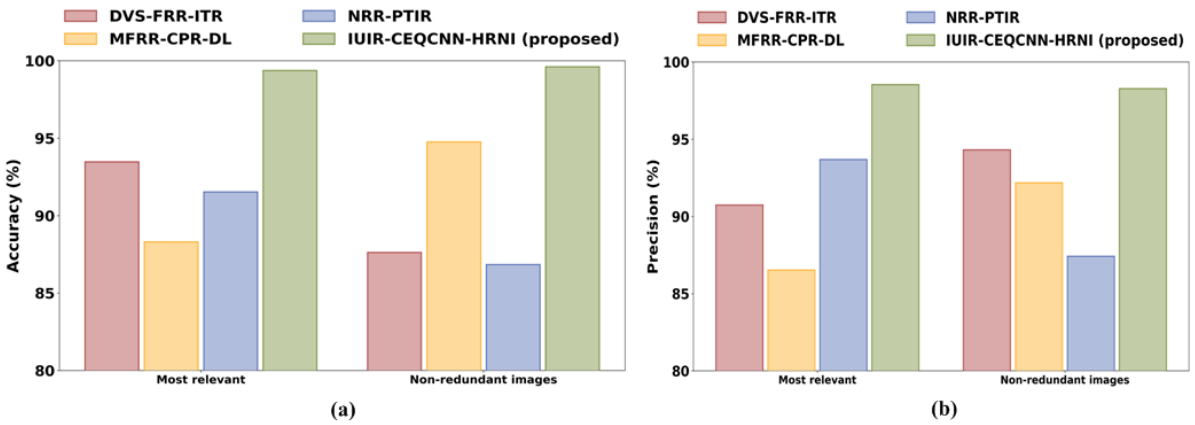


Figure 2:Comparative Performance of (a) Accuracy and (b) Precision

Figure 2 (a) shows accuracy analysis. The high level of accuracy offered by the IUIR-CEQCNN-HRNI approach is because of the application of CEQCNN. It utilizes relatively few numbers of parameters to extract highly dimensional image feature sets, avoiding overfitting of the training process and resulting in good generalization properties. This helps in recognizing important and non-redundant images. Here the IUIR-CEQCNN-HRNI method attains 6.16%, 7.39% and 8.36% higher accuracy for most relevant; 5.29%, 7.96% and 8.23% higher accuracy for non-redundant images compared with existing DVS-FRR-ITR, MFRR-CPR-DL and NRR-PTIR respectively.

Figure 2 (b) shows comparative performance assessment with respect to precision. The precision is quite high because of the application of the DBSMOA. This algorithm helps to optimize the parameters of the model, which allows for an improved boundary for making decisions regarding the relevant and irrelevant images. As a result, the precision is achieved because of the reduction in the number of false positives. Here the IUIR-CEQCNN-HRNI method attains 5.73%, 7.07% and 9.15% higher precision for most relevant; 6.82%, 7.96% and 9.63% higher precision for non-redundant images compared with existing DVS-FRR-ITR, MFRR-CPR-DL and NRR-PTIR respectively.

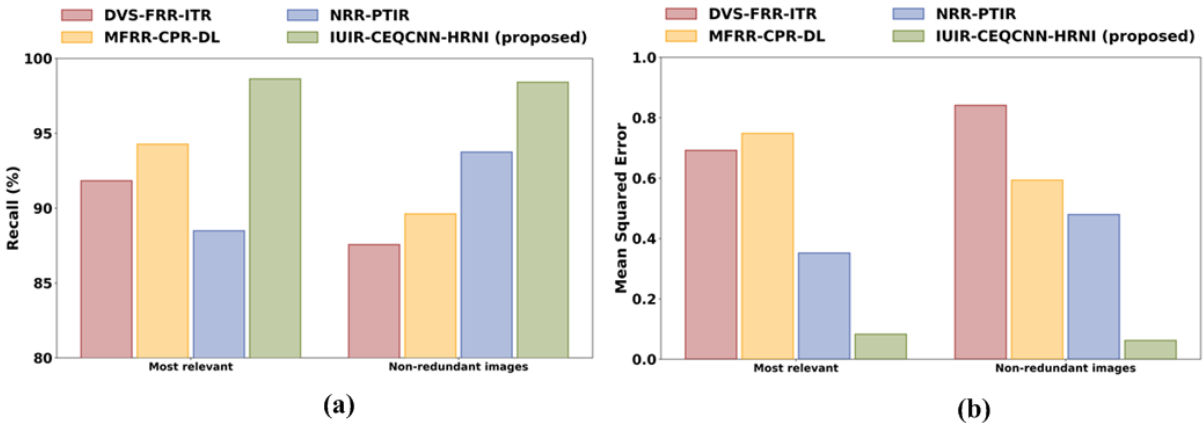
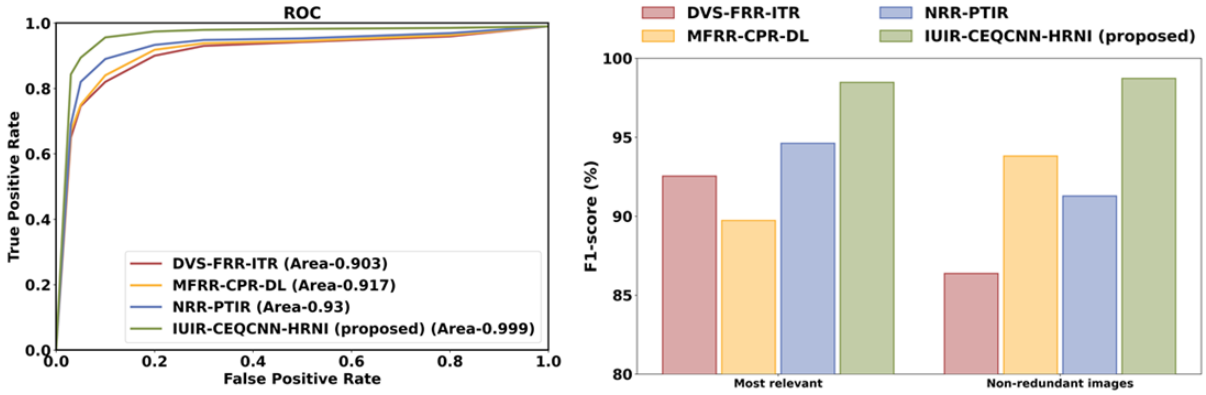


Figure 3:Comparative Performance Assessment of (a) Recall and (b) mean squared error

Figure 3 (a) shows recall analysis. The high recall is attributed to the ANUKF algorithm which improves the quality of images by efficiently removing noise and normalizing images, enabling the model to capture more relevant images and minimize false negatives. Here the IUIR-CEQCNN-HRNI method attains 5.82%, 6.17% and 8.95% higher recall for most relevant; 5.76%, 7.36% and 8.56% higher recall for non-redundant images compared with existing DVS-FRR-ITR, MFRR-CPR-DL and NRR-PTIR respectively.

Figure 3 (b) shows comparative analysis of mean squared error. Hence, the discrepancy between predictions and actual results becomes minimal. The CEQCNN uses quantum-inspired feature mapping to enable an image to be mapped to a higher dimensional space. The CEQCNN is better able to learn the correlation and patterns of features associated with images. Here the IUIR-CEQCNN-HRNI method attains 9.04%, 6.82% and 5.45% lower mean squared error for most relevant; 7.58%, 5.16% and 8.47% lower mean squared error for non-redundant images compared with existing DVS-FRR-ITR, MFRR-CPR-DL and NRR-PTIR respectively.

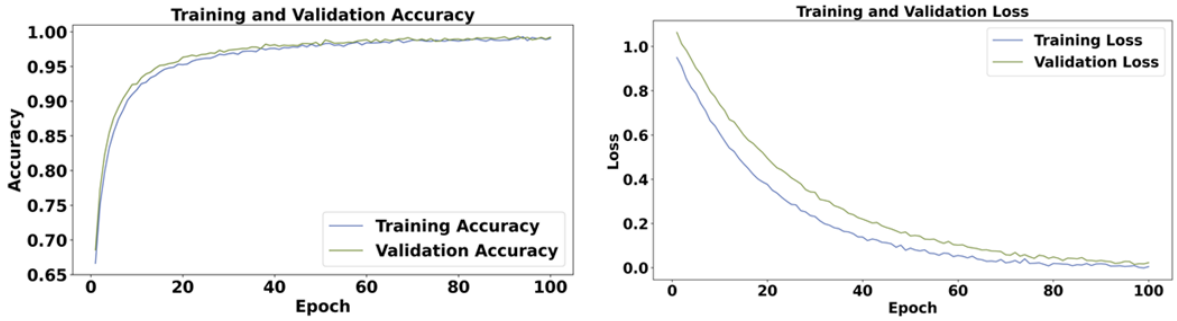
Figure 4 (a) shows ROC analysis. The CEQCNN has the ability to capture the complexity and nonlinearity of the relations in the fusion feature space. The CEQCNN maps the input features to the higher dimensional space, making it easier to separate the important images from the unimportant ones. Thus, it produces better classifications for various thresholds. Here the IUIR-CEQCNN-HRNI method attains 5.32%, 9.57% and 9.78% higher ROC compared with existing DVS-FRR-ITR, MFRR-CPR-DL and NRR-PTIR models.



(a) (b)

Figure 4:Comparative Performance Assessment with Respect to (a) ROC and (b) F1-Score

Figure 4 (b) shows F1-Score analysis. The IRDGF algorithm has effectiveness in integrating visual, semantic, and popularity features into a feature set without redundancy. The IRDGF algorithm improves the discriminating ability of the integrated features to differentiate between relevant and irrelevant images. Here, the IUIR-CEQCNN-HRNI method attains 6.03%, 7.21%, 9.55% high F1 score for most relevant; 5.22%, 7.76% and 9.26% high F1 score for non-redundant images compared with existing DVS-FRR-ITR, MFRR-CPR-DL and NRR-PTIR respectively.



(a) (b)

Figure 5:Analysis of (a) Accuracy vs Epoch and (b) Loss vs Epoch

Figure 5 (a) represents analysis of accuracy vs epoch. The plot shows the changes in the accuracy of the model for both training and validation purposes. At the beginning of the epoch, the training and validation accuracy of the model is relatively lower (around 0.68-0.70), but after some number of epochs, it increases very sharply, meaning that the model has learned important features from the training dataset. However, after some point in time, the improvements made become much smaller and more stable, converging to 0.98-0.99. One very interesting feature about the accuracy of the training process is the similarity between the two curves, meaning that there is no sign of overfitting.

Figure 5 (b) indicates analysis of loss vs epoch. The above graph show the relationship between loss for the purpose of training and validation over the epochs. From this graph, one can observe that the loss gradually goes down as the model continues to learn. Initially, the value for both training and validation losses are fairly high, meaning that the model has not yet learned the features embedded in the data set. However, with an increase in the number of epochs, there is a sharp reduction in the loss values. This shows an indication of effective learning. Furthermore, the loss curves tend to converge at very low level. Most importantly, there is a close proximity of the loss values, signifying that the model does not over-fit.

Table 2: Benchmark analysis using literature review

Author	Performance Metrics			
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Wei, et.al., [6]	87.62	94.31	87.56	86.37
Pivoňka, T. and Přeučil, L., [7]	94.76	92.17	89.61	93.81
Wang, K., et.al, [8]	86.84	87.42	93.74	91.28
Sabahi, F.,et.al, [9]	92.54	89.32	90.43	90.23
Yang, X., et.al, [10]	92.87	91.19	93.85	94.62
IUIR-CEQCNN-HRNI(proposed)	99.41	98.27	98.41	98.72

Table 2 shows the benchmark analysis using literature review.It is evident from Table 5 that although the existing approaches such as Wei etal. [6], Pivoňka and Přeučil [7], Wang etal. [8], Sabahi etal. [9], and Yang etal. [10] have performed well in some metrics and some have performed exceedingly well in other metrics like precision or recall, they lack the required level of effectiveness due to their lack of balanced approach towards each metric. For instance, Wang et al. [8] have performed excellently in recall whereas Yang et al. [10] perform extremely well with an F1-score. In contrast, the proposed algorithm has far exceeded the existing approaches with a significantly high value of all four parameters with 99.41% accuracy, 98.27% precision, 98.41% recall, 98.72% F1-score compared with existing methods.

4.3 Discussion

This proposed IUIR-CEQCNN-HRNI method incorporates well-preprocessing, feature extraction, feature fusion, deep learning, and optimization techniques in order to enhance image ranking within users. Through the incorporation of the visual, semantic, and popularity features, the system acquires better insights about the quality of the images. The CEQCNN model increases the effectiveness of ranking, while the optimization technique makes it easy to tune parameters. Consequently, the method is able to eliminate unnecessary images and ensure that only relevant images remain.In contrast, the proposed algorithm has far exceeded the existing approaches with a significantly high value of all four parameters with 99.41% accuracy, 98.27% precision, 98.41% recall, 98.72% F1-score comparedto the existing methods.The proposed approach performs excellently by utilizing visual, semantic, and popularity attributes, hence increasing ranking precision while minimizing repetitive pictures. Utilization of the CEQCNN algorithm along with optimization helps enhance efficiency and scalability to handle large picture collections. In addition, this model performs effectively even in the absence of actual popularity values by creating artificial ones. Nevertheless, the effectiveness of this model heavily relies on the quality of artificially generated popularity attributes. This approach requires significant computational power to perform feature fusion and optimization.

5.Conclusion

The proposed method offers a computationally efficient approach to ranking images within the same user by combining pre-processing, feature fusion, deep learning, and optimization. This approach enables the selection of the most relevant and non-repetitive images from extensive personal image data sets. The employment of CEQCNN contributes to improving ranking results while minimizing computational effort. The optimization process additionally helps optimize performance by fine-tuning model parameters.Here the IUIR-CEQCNN-HRNI method attains 6.03%, 7.21% and 9.55% higher F1- score and 9.04%, 6.82% and 5.45% lower mean squared error compared with existing DVS-FRR-ITR, MFRR-CPR-DL and NRR-PTIR respectively.In future work, blockchain technology will be incorporated to securely store image metadata, ownership and popularity information such as likes and usage history. This will improve trust, transparency and data authenticity in the ranking process. Overall, combining blockchain will enhance security, scalability and real-world applicability.

References

1. Qiao, S., Wang, R. and Chen, X., 2025. Learning interpretable binary codes via semantic alignment for customized image retrieval. Pattern Recognition, p.112380.
2. Song, C., Liu, X., Hui, C., Zhu, H., Mi, Y., Geng, K., Wu, J., Zhou, Z. and Jiang, F., 2026. Affective-aware fine-grained image quality assessment via multi-modal large language models. Pattern Recognition, p.113420.
3. Nguyen-Le, H.H., Tran, L., Nguyen, D.S.A., Le-Khac, N.A. and Nguyen, T., 2025. Privacy-preserving speaker verification system using Ranking-of-Element hashing. Pattern Recognition, 159, p.111107.
4. Wang, D., Tian, J., Liang, X., Tian, Y. and He, L., 2025. Global-aware Fragment Representation Aggregation Network for image–text retrieval. Pattern Recognition, 159, p.111085.
5. Liu, Z., Xu, J., Gao, S. and Chen, Z., 2025. CSA: Cross-scale alignment with adaptive semantic aggregation and filter for image–text retrieval. Pattern Recognition, 165, p.111647.
6. Wei, W., Gui, Z., Wu, C., Zhao, A., Peng, D. and Wu, H., 2025. Dynamic visual semantic sub-embeddings and fast re-ranking for image-text retrieval. IEEE Transactions on Multimedia, 27, pp.3781-3796.
7. Pivoňka, T. and Přeučil, L., 2024. On Model-Free Re-ranking for Visual Place Recognition with Deep Learned Local Features. IEEE Transactions on Intelligent Vehicles.
8. Wang, K., Wang, Y., Xue, L. and Li, Q., 2024. INRNet: Neighborhood Re-Ranking-Based Method for Pedestrian Text-Image Retrieval. IEEE Access, 13, pp.1470-1480.
9. Sabahi, F., Ahmad, M.O. and Swamy, M.N.S., 2024. RefinerHash: a new hashing-based re-ranking technique for image retrieval. Multimedia Systems, 30(3), p.119.
10. Yang, X., Ge, J., Li, H., Li, L. and Wu, B., 2025. ReID: Re-ranking through image description for object re-identification. Pattern Recognition, p.112552.
11. https://www.kaggle.com/datasets/hsankesara/flickr-image-dataset
12. Levy, A. and Klein, I., 2026. Adaptive neural unscented Kalman filter. IEEE Transactions on Intelligent Vehicles.
13. Shen, K., Zhao, Z., Chen, Y., Zhang, Z. and Cheng, H.V., 2024. Accelerating quadratic transform and WMMSE. IEEE Journal on Selected Areas in Communications, 42(11), pp.3110-3124.
14. Shen, J., Zhan, H., Zuo, X., Fan, H., Yuan, X., Li, J. and Yang, W., 2026. IRDFusion: Iterative relation-map difference guided feature fusion for multispectral object detection. Pattern Recognition, p.113189.
15. Roh, E.J., Im, C., Jeong, W. and Park, S., 2025. Computation-efficient quantum convolutional neural networks for autonomous driving applications: EJ Roh et al. The Journal of Supercomputing, 81(8), p.1033.
16. Rizk, F.H., Gaber, K.S., Eid, M.M., Khafaga, D.S., Alhussan, A.A. and El-kenawy, E.S.M., 2026. Dynamic binary swordfish movement optimization algorithm for feature selection. Journal of Big Data, 13(1), p.8.