Deep Image Retrieval Using Optimizer-aware Convolutional Neural Network with Adaptive Adam-AMSGrad Fusion

Krishna Murthy Sankar [1,*], Woothukadu Thirumaran Chembian [2], Mahamuni Suganthy [3], Marimuthu Venkatesan [4], and Periyannan Raman [5]

[1] Department of Artificial Intelligence and Data Science,
Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College (Autonomous), Chennai, India
[2] Department of Computer Science and Engineering,
Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College (Autonomous), Chennai, India
[3] Department of Electronics and Communication Engineering,
Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India
[4] Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai, India
[5] Department of Business Administration, Panimalar Engineering College, Chennai, India
Email: ksankar@velhightech.com (K.M.S.); drchembianwt@gmail.com (W.T.C.);
msuganthy@veltechmultitech.org (M.S.) venkatesan5488@gmail.com (M.V.); ramanp@panimalar.ac.in (P.R.)
*Corresponding author

***Abstract*—Over the past decade, image retrieval has become the most important research subject in computer vision, attracting significant attention from researchers. The main challenge in image retrieval is significant image depiction and the extraction of large features. Hence, this research proposes an innovative method of cooperative learning between the actual Adam norm approach and Adaptive Moment Estimation with Gradient Correction (AMSGrad) within a Convolutional Neural Network (Adaptive Adam-CNN) for image retrieval. The proposed approach improves the discriminability of learned features and maintains high precision even for larger data. This research primarily employs the three standard datasets, namely Corel 1K, Corel 5K, and Corel 10K, to evaluate the effectiveness of the introduced approach. Next, a preprocessing technique, such as min-max normalization, is employed to scale the input data uniformly for improved image retrieval. The experimental findings demonstrate that proposed Adaptive Adam-CNN approach achieves higher accuracy of 99.63%, 98.37%, and 95.28% on Corel 1K, Corel 5K, and Corel 10K datasets, respectively, compared to the Content-Based Image Retrieval with Accuracy Noise Reduction (CBIR-ANR) and the Extended version of Local Neighborhood Difference Patterns (ELNDP).**

***Keywords*—Adam norm, adaptive moment estimation with gradient correction, convolutional neural network, cooperative method, image retrieval, min-max normalization**

I. INTRODUCTION

Recent advancements in the Internet and the increased use of images have made image retrieval a key research focus [1]. High-quality images are employed in healthcare, the military, and academia [2]. Thus, this has led to the development of larger image databases [3]. Large-scale data retrieval is a critical issue in conventional database development, as traditional text-based databases fail to meet the specific requirements of image databases [4]. The conventional approaches rely on textual descriptions for image retrieval, where images are interpreted through associated text. As a result, retrieval is performed using databases that primarily focus on textual data rather than the visual features necessary for accurate image identification [5, 6]. Conventional image retrieval approaches are inefficient in terms of space and time, which motivates the introduction of an innovative model [7]. Various image retrieval approaches are used, including semantic, text-based, and content-related approaches. In Content-Based Image Retrieval (CBIR), both the features extracted from images and the distance metrics used to compare them are important. Image retrieval systems that rely solely on distance measurements often experience significant delays in query processing [8, 9]. This CBIR system types enhance a retrieval response time through performing a sequential search across greater image collection. Furthermore, it does not effectively retrieve results based on high-level semantic human understanding [10].

CBIR models typically utilize two kinds of image features, such as local and global features. Local features determine the data, such as gradience, edges, and corners, locally [11], while global features are focused on an entire image, and local features such as histograms, textures, and colors [12]. Mostly, local features are recommended for precise image retrieval, although they tend to be more

Manuscript received February 23, 2026; revised March 18, 2026; accepted June 2, 2026; published August 12, 2026.

time-consuming, whereas global features are more suitable for faster processing. Therefore, various advanced CBIR methods focus on local features rather than global features [13]. To identify visually similar images in a dataset, visual features such as low-level features are extracted in CBIR [14, 15]. Although CBIR has seen significant advancements, current approaches still face challenges that limit their effectiveness. A semantic gap occurs among low-level visual features extracted through CBIR and high-level semantic understandings perceived by the users [16, 17]. The curse of dimensionality is another major challenge, as high-dimensional feature spaces often result in inefficient and less accurate retrieval outcomes [18]. To overcome these limitations various learning-based techniques are introduced [19]. These learning approaches provide rapid responses to image queries and are similar to human perception. A learning model's complexity, that utilizes the actual image as input, is greater than that of Machine Learning (ML) approaches [20].

One of the most common problems in existing approaches is the noticeable decline in precision as the number of nearest images retrieved increases for the provided query. The previous methods have converted the minimized dimensional feature vectors into binary. This transition maps various features to values between 0 and 1, leading to data loss and inaccurate outcomes. Thus, this research proposes a novel cooperative approach combining the Adam approach and Adaptive Moment Estimation with Gradient Correction (AMSGrad) in a Convolutional Neural Network (Adaptive Adam-CNN) for image retrieval. The proposed Adaptive Adam introduces an adaptive p-norm gradient scaling mechanism, where the norm value is computed from the current and historical gradient magnitudes. This enables dynamic step-size modulation personalized to the curvature of the loss landscape, thereby avoiding local minima and saddle points. The CNN backbone is synergistically tuned with the AMSGrad optimizer, where optimization feedback actively controls the training dynamics at convolutional and dense layers. The proposed method demonstrates competitive performance against existing CNN-based retrieval approaches.

The key innovations of this research are listed below:

- This research proposes a novel cooperative method of adaptive Adam-CNN for image retrieval, which aims to achieve more accurate and semantically relevant results.
- The min-max normalization technique is employed in the preprocessing phase to equally scale all the input data to enhance the approach's effectiveness.
- The implementation of the work is evaluated through multiple datasets based on the different performance metrics.

This research paper is structured as follows: Section II presents a comprehensive review of existing content-based image retrieval approaches, highlighting the strengths and limitations of handcrafted descriptors, deep learning-based retrieval frameworks, and optimization strategies employed in prior studies. Section III describes the proposed Adaptive Adam-CNN framework in detail, including the network architecture, adaptive optimization mechanism, normalization strategy, and the overall retrieval pipeline. Section IV outlines the experimental setup and discusses the obtained results through quantitative comparisons, ablation studies, computational complexity analysis, cross-dataset validation, and statistical significance evaluation across multiple benchmark datasets. Finally, Section V summarizes the major findings of this study, emphasizes the key contributions of the proposed approach, and provides directions for future research.

II. Literature Survey

Rahbar and Taheri [21] developed a triplet loss function in terms of binary cross-entropy for training the Siamese approach in image retrieval. The implemented approach enabled large metric learning and designed a discriminatory feature space through greater discrimination among classes and lower intra-class distance. Primarily, image features were extracted by a CNN after training the Siamese model to develop a discriminatory feature space through large-scale model learning. The feature space was visualized using a statistical approach. However, the approach was dependent on pre-trained CNNs and did not incorporate adaptive optimization approaches to escalate generalization across datasets.

Narayanaswamy *et al.* [22] introduced the 2-Dimensional CNN (2DCNN) assisted CBIR. Initially, the authors collected diverse datasets to evaluate the proposed 2DCNN approach; later, a normalization technique was employed to remove the noise during preprocessing. Then, important features were extracted using the Principal Component Analysis (PCA) approach. Finally, the relevant images were retrieved using the proposed 2DCNN approach. However, 2D-CNN-based CBIR systems that utilized PCA for handcrafted features and dimensionality reduction suffered from the loss of important nonlinear characteristics.

Vieira *et al.* [23] presented the CBIR, which pursued an Accuracy Noise Reduction named CBIR-ANR mechanism that modified the query responses and enhanced the confidence in image retrieval. Moreover, the presented mechanism integrated the multiple low-level features to design a 187-dimensional feature vector, which was size-effective for large-scale data groups and a competitive model. However, the CBIR-ANR approach lacked model integration for hierarchical feature extraction, which constrained its ability to learn high-level abstractions from images.

Taheri *et al.* [24] developed the CBIR approach through the fusion of handcrafted features from the semantic pyramid of images. The semantic pyramid was formed by fusing feature maps from different stages of the Deep Neural Network (DNN). The developed feature vector involved image data and semantic features. The t-SNE approach was utilized for an interpretation of the feature vector's discriminability among classes in a dataset. However, the model did not adaptively learn optimal feature combinations and relied on t-SNE for visualization, without improving retrieval effectiveness.

Vieira *et al.* [25] implemented the novel image description approach for the representation of visual features of the images. Moreover, the pattern among diverse CBIR systems was noticed in that the initial images retrieved were more confident than those in the center and last locations. Subsequently, the interaction among the primary images retrieved from CBIR models was investigated to ignore the new approach for the minimization of accuracy in the noise data. However, the model's capability to handle noisy data in a scalable way was unclear.

Taheri *et al.* [26] introduced a CBIR approach, that focused on extraction of important features. The developed feature vector was an integration of low and mid-level image features. Extraction of low-level features involved shape, texture, and color, which were employed through Gabor wavelet transform, multi-level fractal dimension, and auto-correlogram estimation. Mid-level features extraction was employed through Deep Boltzmann Machines (DBMs), which leveraged both low-level features and relationships among them. The output feature vector, combining low-level features and DBM outputs, was refined using the collected database. However, the DBM suffered from high computational complexity and reliance on manual feature engineering, which limited scalability.

Ranjith *et al.* [27] implemented the novel CBIR approach through Deep Learning (DL)-based Inception v3, named DL-based Image Mining (DLIM), for significant image retrieval from the collected datasets. The DLIM approach extracted the features of all the images that existed in the database and stored them as feature vector. After providing a Query Image (QI), the DLIM approach extracted features from the QI and identified the similarity measurement through features placed in the datasets. The images with the greatest similarity were retrieved from the database. However, the DLIM approach lacked the optimizer variations or hyperparameter tuning for the improvement of the approach's effectiveness further on diversified datasets.

Kelishadrokhi *et al.* [28] developed an image retrieval method integrating color and texture features. They introduced the Extended version of Local Neighborhood Difference Patterns (ELNDP) to attain the discriminative features. ELNDP exploited the benefits of Local Binary Pattern (LBP) as well as texture descriptors. Moreover, for global feature extraction, the optimized color histogram features in Hue, Saturation, and Value (HSV) color space were utilized for the extraction of color features. Eventually, a prolonged Canberra distance metric was utilized for the retrieval of more important images, which lacked sensitivity to the minimum values, such as traditional Canberra. However, the ELNDP model did not generalize well over moderately sized benchmark datasets.

Liu and Yang [29] presented the image retrieval framework for the identification of efficient, constrained, or invalid feature maps through ranking-assisted Haralick's mechanism. This approach understood the image data and identified the objects, as these mechanisms contained effective physical importance. A Gabor filter was utilized to mimic human-aligned strategies, enabling the deep features to effectively depict alignment and thus enhance discriminative power. The benefits of conventional texture as well as deep features were considered to provide the compact depiction.

Hu *et al.* [30] developed the rapid dual-feature extraction approach in terms of a firmly joined lightweight network. This approach extracted both local and global features in a unified model that involved a lightweight backbone. Subsequently, learned step-size quantization was used to minimize computational overhead during the inference phase. Moreover, an effective channel attention model ensured the capability of feature depiction.

Tyagi *et al.* [31] introduced Deep Feature Fusion (DFF), which integrated ConvNeXt and EfficientNet to enhance large-scale model understanding. The introduced approach fused the layer embeddings from different networks and allowed more detailed modeling of local texture descriptors and global semantic abstractions. This method contributed to a scalable, architecture-agnostic paradigm for deep feature fusion in CBIR and highlighted the potential of large hybrid embeddings for enhanced visual search in moderately sized benchmark datasets.

Humadi *et al.* [32] presented a multi-tier DL approach for image retrieval. Initially, a Supervised Neural Network (SNN) reduced dimensionality and enhanced interpretability through exchanging a HSV color space into semantic 2D color labels using pixel-level classification. Then, a CNN improved computational effectiveness through processing these labels instead of actual samples. Adaptive fusion resolved the constraints of fixed-weight similarity measures in maintaining heterogeneous query kinds.

From the overall analysis, existing CBIR methods either depend on handcrafted features or conventional classification approaches or use fixed optimization mechanisms that constrain adaptability and generalization. Moreover, the lack of cooperative optimization and inadequate performance on large or diverse datasets present common limitations. To overcome these challenges, the proposed method integrates a CNN with Adaptive Adam and employs automated feature learning, adaptive regularization, and improved training. The proposed Adaptive Adam-CNN framework introduces a synergistic cooperative optimization strategy that combines the strengths of Adam and AMSGrad with an adaptive normalization mechanism to enhance model learning dynamics. This contributes to faster convergence, improved generalization, and stable training across varying dataset complexities. The optimizer adaptively tunes the learning rate based on dynamic gradient norms, leading to more discriminative feature extraction and reduced semantic ambiguity during retrieval. It also demonstrates enhanced accuracy, computational efficiency, and robustness over different benchmark datasets, solving the core issues identified in existing approaches.

III. Proposed Methodology

An innovative framework is proposed in this research by performing diverse methods for image retrieval. The proposed method integrates CNN with Adaptive Adam and employs automated feature learning, adaptive regularization, and improved training. Unlike AdaBound, which only bounds the learning rate, RAdam, which rectifies variance, and AdaBelief, which estimates gradient belief, the proposed Adaptive Adam jointly integrates adaptive norm estimation and optimizer switching. The proposed method dynamically alternates between Adam and AMSGrad according to the relationship between current and historical gradients. This provides both rapid convergence and stable generalization, which are not simultaneously achieved by the existing optimizers. This research involves different steps: data collection, preprocessing, and image retrieval. Fig. 1 illustrates an outline of working procedure of introduced approach for image retrieval.

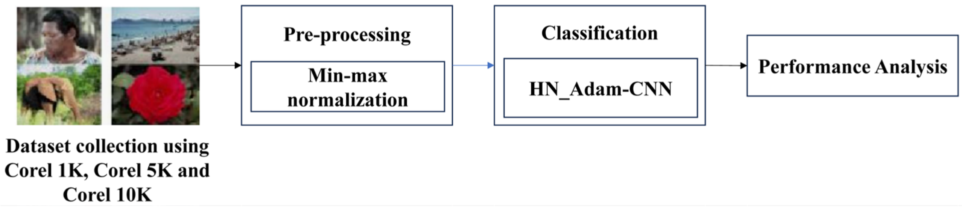


Fig. 1. Outline of working principle of proposed approach for image retrieval.

A. Dataset Collection

This research tests the existing and the proposed approach using different labeled datasets, such as Corel 1K [33], Corel 5K [34], Corel 10K [35], Oxford 5K [36], Paris 6K [36], and Caltech-101 [37]. The images in the collected datasets involve large classes, orientation of the object, spatial data, color, texture, and shape, which are used to evaluate the model's effectiveness. These datasets are utilized for estimating the proposed approach in terms of identifying the comparisons between various image genres. The details of the datasets are described below.

1) Corel 1K

This dataset consists of 1000 color images of various sizes 256×384 or 384×256 pixels. This dataset is arranged into 10 distinct classes, each containing 100 images. It includes diverse categories such as buildings, animals, landscapes, and people.

2) Corel 5K

This dataset contains 5000 colored images in the size of 187×126 or 126×187, that are kept in JPEG format. All the images in this dataset are classified into 50 classes.

3) Corel 10K

Corel-10k is the most utilized benchmark dataset for image classification as well as retrieval. It involves diverse classes ranging from humble contextual images to complex artifacts. This dataset comprises 10,000 images, which are partitioned as 80 groups. An image count in each class differs between 100 and 550 in every instance involving a resolution of $85 \times 128 \times 3$ or $128 \times 85 \times 3$ pixels. The dataset includes classes such as nations, buses in various colors, buildings, dinosaurs through minimal contextual backgrounds, elephants, more flowers, horses in different settings, mountains resembling beach scenes, and food.

4) Oxford 5K and Paris 6K

The Oxford buildings dataset involves 5062 samples acquired through searching for specific Oxford landmarks. An acquisition is physically annotated for producing the complete ground truth for 11 various benchmarks; each demonstrated through five possible queries. This offers the group of 55 queries across which an object retrieval system is estimated.

5) Caltech-101

The Caltech-101 dataset involves around 9364 images grouped into 101 object classes. This dataset represents the non-uniform class distribution, with the count of images per class differing significantly, ranging from over 40 to 800 images per category. These datasets collectively provide a robust and complete estimation environment for the CBIR models, encompassing both balanced and imbalanced data distributions.

B. Preprocessing Using Min-Max Normalization

Data normalization is an approach that modifies the values estimated on various scales to a theoretically basic scale. The min-max normalization is an easy approach, which rescales data values to a range from 0 to 1 based on the minimum and maximum values of an actual input. A collected input image contains various scales, which leads to inaccurate results; thus, it is important to rescale the data. The procedure of normalization converts the numerical values into another range by assigning the mathematical function [38]. Normalization maps the values in the data to a specific range, as described in Eq. (1).

$$v' = \frac{v - min_A}{max_A - min_A}\left(new\_max_A - new\_min_A\right) + new\_min_A \quad (1)$$

where A denotes the attribute data; min_A and max_A illustrates the minimum and maximum absolute value of A individually. v denotes an actual value, and v' denotes the normalized value.

C. Image Retrieval Using CNN

CNNs are most appropriate for image retrieval tasks, as they automatically learn hierarchical and spatial features directly from actual image data, unlike conventional classification approaches like K-Nearest Neighbor (KNN), Random Forest (RF), and Support Vector Machine (SVM), which depend on manual feature extraction. CNNs have the capability to model deep semantic representations, which are more significant in determining the visual similarities, and design them more precisely for large datasets. Their ability to generalize across different image patterns makes them an excellent choice for image retrieval applications. Moreover, there are different motives to employ the CNN rather than benchmark Multi-Layer Perceptron (MLP) approach for image retrieval. A significant motive is a weight-sharing feature that minimizes the count of trainable network parameters whereas enhancing generalization effectiveness and ignoring the model overfitting. Moreover, understanding the feature extraction and classification layers simultaneously leads to an approach's result that is greatly improved and depends on the extracted features.

The CNN framework consists of various layers: input, feature-extraction layer through convolutional and pooling layers, dense layers, and an output layer. Every CNN layer is characterized with particular network hyperparameters that are crucial for model training and development, as provided in the following sections.

1) Convolutional layer

CNN uses different kinds of convolutional procedures, such as Same Padding (SAME) and Valid Padding (VALID). The SAME convolution generates the feature maps through similar dimensions as the input data through zero padding. On the other hand, VALID convolution produces smaller feature maps through the lack of padding. The convolutional block applies the filters through a preidentified height and width for the generation of feature maps from input data. The elements of feature maps are estimated by adding the product of filter elements, as well as respective input elements wrapped with a filter.

2) Pooling layer

Pooling is complex in CNNs to achieve local conversion invariance and to produce outputs on down sampled feature maps, thereby enhancing sensitivity to differences in feature positions in input data. This research utilizes max pooling, which selects the greatest value between the modeled elements. The pooling comprises applying the kernel through predefined height, width, and type to the input data. The down-sampled feature maps are produced through sliding a kernel between top-left and bottom-right, directed by pre-identified strides.

3) Dense layer

To develop the output feature maps in a dense layer, feature maps are flattened and reshaped into a vector. A classification model comprises one or different dense layers, involving neurons which process the input to produce an output. These layers attain the inputs from previous layers through connections by applying the weights. The dense layer uses backpropagation to understand the network weights. The training of the CNN is intended to minimize the errors between predictions and the output of the input dataset through minimization.

While the proposed CNN architecture pursues a conventional layer-wise structure, its distinctiveness lies in the integration with the adaptive Adam optimizer, which introduces optimizer-aware gradient modulation. During backpropagation, the adaptive norm-based updates dynamically adjust the learning rate across layers, particularly enhancing the learning behavior in the final convolutional blocks. This facilitates the capture of more refined and discriminative spatial hierarchies. Additionally, the use of adaptive dropout, influenced by gradient norms, strengthens generalization in the dense layers by preventing co-adaptation of neurons. This synergy between network architecture and adaptive optimization results in improved convergence and retrieval performance across diverse datasets.

The details of the proposed optimizer are provided in the subsection.

4) Adaptive adam optimizer

This research proposes an improved approach based on the benchmark Adam optimizer. The core innovation of the proposed adaptive normalization mechanism lies in its ability to intelligently regulate the gradient magnitude during backpropagation by computing a norm value based on both instantaneous and historical gradients. This optimizer is widely used and gaining popularity; however, it faces some issues, such as limited generalization and slow convergence. Thus, this research aims to overcome those issues by integrating the benchmark Adam for attaining better results. The proposed approach is described as adaptive Adam, which incorporates an adaptive norm and a cooperative method combining the Adam approach and AMSGrad optimizers. The terms "H" and "N" represent the cooperative method and adaptive norm, individually. To enhance a generalization capability of conventional Adam optimizer, this research employs a cooperative mechanism that combines enhancements from both the Adam and AMSGrad approaches.

The primary concept of the proposed approach is to dynamically modify a step size of learning rate according to an adaptive norm based on both current and previous gradients. The norm function computes the Euclidean distance between pairs of binary units to quantify their similarity. An adaptive norm represents a norm value that is adaptively modified according to gradient values acquired in each epoch. Initially, adaptive Adam trains an approach through Adam, and adaptive norm enhances the learning rate's step size, eliminating the local minima issue. As the Adaptive Adam approach utilizes the dynamic norm value, the absolute value of the gradient is estimated.

A threshold value of a norm (A_{t0}) is selected arbitrarily in a range among 2 and 4. A norm value $A(t)$ is then significantly estimated based on value of an absolute gradient $|g_t|$ and an exponential moving average of a preceding gradient m_{t-1}, as formulated in Eq. (2).

$$A(t) \leftarrow A_{t0} - \frac{m_{t-1}}{m_{max}} \tag{2}$$

where m_{max} demonstrates the maximum value between $|g_t|$ and m_{t-1}. Adaptive Adam introduces an adaptive norm adjustment to the benchmark Adam method by modifying its energy value at each update. This is employed through data of prior gradient updates. In Eq. (2), the ratio $\frac{m_{t-1}}{m_{max}}$ is less than or equal to 1; this denotes a norm value will be in an interval between 1 and 4. Hence, a sequence is changed to an AMSGrad approach in a constraint that $A(t) < 2$. This means that Adaptive Adam adopts a modified Adam approach with greater exploration capability, using a maximum norm value in an interval of 2 to 4. This observes that Adaptive Adam utilizes a small step size of learning rate when an absolute gradient $|g_t|$ is close to m_{max}. The mathematical expression for step size formulations of Adam and Adaptive Adam are expressed in Eqs. (3) and (4).

$$\Delta\theta_t^{Adam} = \eta \times \frac{m_t^{\Delta}}{\sqrt{v_t^{\Delta}}} \tag{3}$$

$$\Delta\theta_t^{HN\_Adam} = \eta \times \frac{m_t^{\Delta}}{v_t^{1/A}} \tag{4}$$

where $|\Delta\theta_t|$ denotes a step size for parameter update at time t. A Stochastic Gradient Descent (SGD) algorithm, as compared to an ideal optimizer, takes fewer steps because it is proportional to the Exponential Moving Average (EMA) of gradient m_t. This norm modulates the learning rate dynamically by enhancing it for small gradient magnitudes and reducing it in unstable regions, thus promoting robust convergence.

In the proposed Adaptive Adam framework, gradient flow is not treated uniformly across layers and epochs but is instead adaptively modulated related to both short-term and long-term gradient behaviors. Specifically, the optimizer maintains first-order (mean) and second-order (variance) moment estimates, m_t and v_t, which guide the parameter updates. Unlike the conventional Adam optimizer, Adaptive Adam enhances this process by introducing an adaptive norm p_t that influences the second moment calculation. This norm is computed dynamically from the absolute value of current gradient $|g_t|$ and exponential moving average of previous gradients $\hat{g}_{t-1}$. As a result, the curvature sensitivity is adaptively tuned: parameters in flatter regions of the loss surface receive larger updates to encourage exploration, while parameters in steep regions are updated more conservatively to maintain stability. At each iteration, the optimizer estimates m_t and v_t, computes p_t, and uses this value to modulate the learning rate for the parameter update. This integration of historical gradient information enables the optimizer to make smoother, context-aware updates, improving convergence and generalization during training.

Adaptive Adam ensures bounded and stable updates through combining norm-aware scaling. Compared to AdaBelief, which shifts second-moment estimation toward gradient similarity, Adaptive Adam provides optimal step-size control under gradient fluctuations. Instead of merely correcting for variance (as in RAdam) or reinterpreting second-moment beliefs (as in AdaBelief), an effective p-norm is computed to modulate the direction and scale of parameter updates. This norm is bounded dynamically between values 2 and 4 depending on the relative magnitude of the current and past gradients. If gradients are smooth and consistent, a higher norm enhances exploration by reducing over-correction. When gradients are volatile or noisy, the norm lowers the learning rate, thereby ensuring stability. The norm dynamically switches between Adam and AMSGrad regimes depending on gradient convergence behavior.

The proposed method guarantees convergence through sustaining the non-increasing learning rate condition of AMSGrad while enhancing adaptivity via gradient-norm feedback, improving generalization, and reducing overfitting. The proposed optimizer enables the preservation of momentum while dynamically scaling learning updates based on gradient volatility. Unlike RAdam, which rectifies variance by warm-starting the variance term, or AdaBelief, which shifts the second moment estimate toward belief in gradient consistency, the proposed method maintains convergence stability via AMSGrad logic while enhancing responsiveness through adaptive norm modulation.

In this research, the 80:20 train-test split is considered for training and testing. The hyperparameter settings, such as a batch size of 32 and 100 epochs with an initial learning rate of 0.0001, are considered for tuning the proposed method. Moreover, the dropout rate of 0.2 to 0.5 is considered to be a dense layer for solving the overfitting issue. The proposed Adaptive Adam optimizer is used for parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. These hyperparameter values are effectively chosen for reducing the model complexity, improving learning stability, and enhancing the convergence speed. The chosen settings ensure an approach that effectively optimizes the performance on the validation set. All hyperparameters are selected through manual tuning to minimize validation loss. To analyze robustness, an ablation study is performed by varying the adaptive norm threshold (A_0) from 1.5 to 2.5 and the dropout rate from 0.2 to 0.5. The best performance is achieved at $A_0 = 2.0$ and dropout = 0.5. Performance variation remains within ±0.8%, indicating that the proposed method is relatively insensitive to moderate changes in these parameters. Fig. 2 demonstrates the layer-wise architecture of the proposed method. Algorithm 1 outlines the pseudocode of Adaptive Adam for better reproducibility. The convergence of the proposed Adaptive Adam optimizer follows from the convergence guarantees of AMSGrad under bounded gradients and diminishing effective learning rates. Since the proposed method activates AMSGrad whenever the adaptive norm exceeds the threshold, the second-order moment estimate remains non-decreasing. Therefore, the sufficient convergence conditions established for AMSGrad remain valid. Furthermore, the adaptive norm only scales the effective learning rate within the bounded interval [2, 4], ensuring that the update magnitude remains finite and stable.

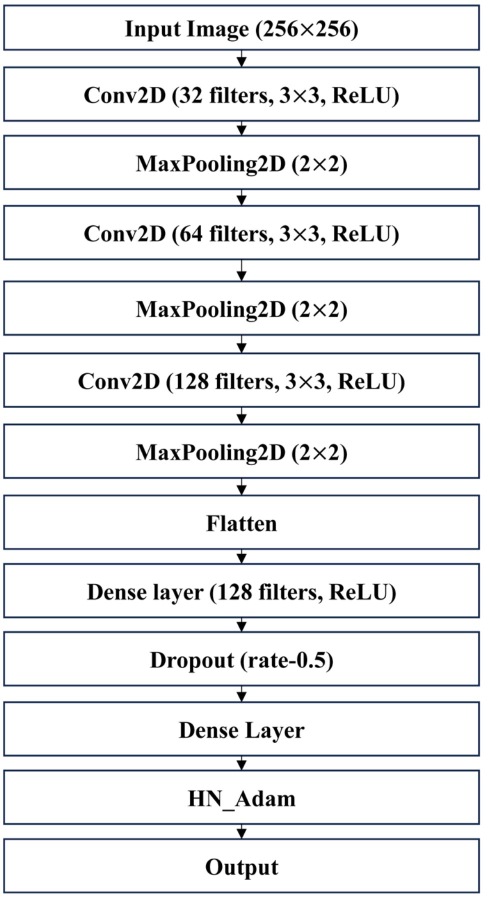


Fig. 2. Layer-wise architecture of the proposed method.

During the training process, each image from the Corel datasets (1K, 5K, and 10K) is passed through the proposed CNN model in a forward propagation step, where feature extraction is performed, and predictions are generated. These predictions are compared against the ground-truth labels through cross-entropy loss function. In the subsequent backpropagation phase, the optimizer calculates the gradient g_t of loss for each trainable parameter. The Adaptive Adam optimizer then uses both the current gradient g_t and exponentially weighted average of past gradients $\hat{g}_{t-1}$ to compute an adaptive norm p_t, which modulates the update behavior. This norm is applied to adjust the first and second moment estimates m_t and v_t, allowing the optimizer to dynamically scale learning rates based on gradient behavior. Parameters θ are updated using a norm-aware step size, ensuring stable and curvature-sensitive updates. This entire process is repeated iteratively across all mini-batches (batch size = 32) over 100 epochs, with training conducted on 80% of the dataset and validation performed on the remaining 20%. This integration allows the Adaptive Adam optimizer to directly influence every step of the learning process, enabling efficient convergence, improved generalization, and better handling of gradient variability in moderately sized benchmark datasets.

Algorithm 1 illustrates how the optimizer dynamically modifies a learning rate based on gradient fluctuations, preventing issues such as local minima entrapment and unstable convergence. This tracking mechanism ensures a more stable and efficient training process, especially in high-dimensional image retrieval tasks. Steps 1 to 10 illustrate the standard initialization and update rules of an Adam optimizer. Steps 11 to 15 model the core contribution of the proposed method. These steps develop an adaptive norm-based adjustment mechanism, where the update of the learning rate and gradient scaling is dynamically overseen through gradient norms. Specifically, Step 11 estimates the gradient norm, and Steps 12 to 14 apply norm-based correction factors to modify first and second moment estimates of an optimizer. This mechanism directly enhances stability at optimization by eliminating immediate spikes or vanishing updates in noisy gradients. Step 15 finalizes a parameter update with this modified mechanism. This part differentiates the proposed Adaptive Adam from standard optimizers by ensuring enhanced convergence, better handling of complex landscapes, and enhanced retrieval performance, especially for high-dimensional or sparse datasets.

Algorithm 1. Pseudocode of Cooperative and Adaptive Norming of Adam through AMSGrad (Adaptive Adam)

1. Reset the parameter θ_0, step size η, and batch size N_B.
2. Exponential decay rates β_1, β_2 and ε dataset $\{x_i, y_i\}_{i=1}^{N}$.
3. Reset: $m_0 = 0$, $v_0 = 0$, AMSGrad =False, $v_{hat(0)} = 0$.
4. For all $t = 1, \ldots, T$, do
5. Draw an arbitrary batch $\{x_{ik}, y_{ik}\}_{i=1}^{NB}$ from the dataset.
6. $g_t \leftarrow \sum_{k=1}^{N} \nabla l \{(x_{ik}, y_{ik}, \theta_{t-1})\} // f' = (\theta_{t-1})$
7. $m_t \leftarrow \beta_1 \times m_{t-1} + (1 - \beta_1) \times g_t$// moving average
8. $m_{max} \leftarrow Max(m_{t-1}, |g_t|)$
9. $A(t) \leftarrow A_{t0} - \frac{m_{t-1}}{m_{max}}$
10. $v_t \leftarrow \beta_2 \times v_{t-1} + (1 - \beta_2) \times (|g_t|)^{A(t)}$
11. $A_t = A_\theta + |g_t| / max(|g_t|, EMA|g_t|)$
12. If $A_t < 2$, then
13. $v_t = \beta_2 \times v_{(t-1)} + (1 - \beta_2) \times {g_t}^2$
14. Else:
15. $\hat{v}_t = max\left(\hat{v}_{(t-1)}, \beta_2 . v_{t-1} + (1 - \beta_2) \times {g_t}^2\right)$
16. End if
17. End for
18. Return the final parameter θ_T.

IV. Experimental Results

The proposed approach for image retrieval is experimented with in Python 3.11.12 by the configurations of Windows 10 OS, NVIDIA RTX 3060 Gated Processing Unit (GPU) with 16GB RAM, Intel i5 processor, and TensorFlow/Keras framework. The datasets used are split into an 80:20 ratio for training and testing. The adaptive norm threshold (A_0) is initialized at 1.5 after grid-search analysis within the range [2, 4]. Hyperparameter tuning is conducted manually in the preliminary stage and subsequently verified through the Adam Optimizer. In this research, the various performance metrics, accuracy, specificity, precision, sensitivity or recall, F1-Score, and mean Average Precision (mAP) are used to validate the proposed method's effectiveness. The mathematical formulations for these indices are illustrated in Eqs. (5)–(9).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (7)$$

$$F1-score = \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (8)$$

$$mAP = \frac{\sum_{q=1}^{Q} AP(Q)}{Q} \quad (9)$$

Here, TP, TN, FP, and FN denote True Positive, True Negative, False Positive, and False Negative. AP denotes the average precision, and Q specifies the count of query images. Table I specifies the hyperparameter details of the proposed method for image retrieval.

TABLE I. HYPERPARAMETER DETAILS OF THE PROPOSED METHOD FOR IMAGE RETRIEVAL

Parameters	Ranges	Values
Learning rate	0.0001–0.01	0.001
Batch size	8–64	32
Number of epochs	50–200	100
Dropout rate	0.2–0.5	0.5
Weight decay	0.00001–0.001	0.0001
Adam $\boldsymbol{\beta_1}$	0.80–0.99	0.90
Adam $\boldsymbol{\beta_2}$	0.90–0.999	0.999
AMSGrad coefficient	0.1–1.0	0.5
Adaptive scaling factor	0.5–2.0	1.2
Norm threshold (A_0)	1.5–2.5	2.0
Convolution kernel size	$\mathbf{3 \times 3}$–$\mathbf{7 \times 7}$	$\mathbf{3 \times 3}$
Number of filters	32–256	64, 128, 256
Early stopping patience	5–20	10

A. *Performance Estimation*

The proposed method's performance in comparison with other methods is demonstrated here using different datasets. The proposed CNN method is validated through the existing methods, namely CNN, Residual Network (ResNet), Visual Geometry Group Network (VGGNet), and Inception v3. A proposed CNN method has better capability to automatically learn and recognize the important features in the images, because other existing methods are incapable of modeling the local patterns as well as the spatial hierarchy. All the experiments are repeated five times using random seeds of 7, 21, 42, 84, and 100. The reported values correspond to mean ± standard deviation. Tables II–IV, VII and IX evaluate the classification performance of the feature extraction model using classification-oriented metrics, whereas Tables V and VI assess the image retrieval effectiveness using retrieval metrics. Therefore, classification-oriented measures such as accuracy, Average Precision (ARP), Average Recall (ARR), and F1-Score are reported. In contrast, Tables V–VII evaluate the actual image retrieval performance and generalization ability of the CBIR system. Hence, retrieval-oriented metrics such as P@10, mAP, R@10, and NDCG@10 are used, as they better measure ranking quality and the relevance of the retrieved images. P@10, mAP, R@10, and NDCG@10 are commonly adopted CBIR evaluation metrics that measure the precision and recall performance within the top-10 retrieved images for each query image. Therefore, the variation in metrics is intentional and corresponds to different objectives of classification and retrieval evaluation.

Table II specifies the performance estimation of various image retrieval methods using multiple datasets. The proposed methods reach the higher accuracy of 99.63%, 98.37%, and 95.28% on Corel 1K, Corel 5K, and Corel 10K datasets individually. Table III specifies the K-fold analysis of proposed method with previous approach in terms of multiple datasets. To reduce optimistic bias, nested 5-fold cross-validation is performed, where hyperparameter tuning is conducted within the inner folds and model evaluation is carried out in the outer folds. The multiple K-values, such as 2, 3, 4, and 5, are used for estimating an approach's effectiveness. The K-value equal to 5 attains improved results as it provides an optimal balance among variance and bias, ensuring reliable and precise predictions. The lower K-values lead to overfitting, whereas the higher K-values lead to over-smoothed significant patterns. These values demonstrate the effectiveness of the utilized cross-validation folds, where K = 3 and 5 offer the minimum bias and greater variance. Between these, K = 5 reliably attains the improved performance in every dataset, representing that it provides an optimal balance between training robustness and testing variability. Hence, K = 5 is chosen to ensure effective and generalizable estimation without overfitting.

Table IV demonstrates the computational complexity of the proposed approach using multiple datasets. The computational indices include inference time, training time, Floating Point Operations (FLOPs), model size, and number of parameters. Although the proposed Adaptive Adam-CNN incurs slightly higher inference time due to adaptive optimization and deeper feature extraction, it achieves improved retrieval performance compared to existing methods. If the model size is large, the proposed method's FLOP and latency analysis demonstrate enhanced model efficiency. Particularly, the proposed approach requires training time of 49.11 s per batch and inference time 19.47 ms per image, compared with training time 39.37 s and inference time 12.39 ms for Inception v3. Nevertheless, this moderate enhancement in computational cost is justified through the significant enhancement in retrieval accuracy reported in this analysis. The confidence interval values further demonstrate that the training time remains consistent over repeated runs, with the proposed method illustrating 49.11 ± 1.34 s, which confirms the reliability and reproducibility of the computational performance. The 95% confidence interval values are included to demonstrate the consistency of the measured training time across repeated runs. Although the proposed adaptive Adam-CNN requires slightly higher computational cost, the narrow confidence interval confirms that its execution time remains stable and reliable.

TABLE II. PERFORMANCE ESTIMATION OF DIFFERENT IMAGE RETRIEVAL METHODS USING MULTIPLE DATASETS

Dataset	Methods	Accuracy (%)	ARP (%)	ARR	F1-Measure (%)
Corel 1K	CNN	90.37	89.37	11.59	13.36
	ResNet	92.38	91.38	13.48	15.48
	VGGNet	94.27	93.38	15.48	17.48
	Inception v3	97.27	96.38	16.74	19.37
	Proposed Adaptive Adam-CNN	99.63	99.51	18.82	27.28
Corel 5K	CNN	91.32	90.82	13.48	28.46
	ResNet	93.13	91.47	15.48	30.38
	VGGNet	94.28	93.21	16.47	33.48
	Inception v3	95.28	95.38	19.74	36.53
	Proposed Adaptive Adam-CNN	98.37	97.13	21.28	38.38
Corel 10K	CNN	85.38	85.31	19.38	34.38
	ResNet	87.56	86.38	21.47	36.32
	VGGNet	92.30	88.42	23.91	39.47
	Inception v3	94.28	91.37	25.38	40.98
	Proposed Adaptive Adam-CNN	95.28	94.38	28.42	42.38
Oxford 5K	CNN	90.12	89.67	26.42	33.12
	ResNet	91.26	90.81	28.38	35.42
	VGGNet	93.72	92.26	30.19	37.84
	Inception v3	95.42	94.21	32.67	39.45
	Proposed Adaptive Adam-CNN	96.73	95.44	33.33	41.26
Paris 6K	CNN	89.45	87.12	14.33	31.78
	ResNet	90.86	89.84	16.41	34.12
	VGGNet	92.11	90.67	17.03	36.52
	Inception v3	94.17	92.38	18.29	38.41
	Proposed Adaptive Adam-CNN	95.44	94.12	19.82	40.37

TABLE III. CROSS-FOLD VALIDATION OF THE PROPOSED ADAPTIVE ADAM-CNN METHOD WITH EXISTING METHODS BASED ON MULTIPLE DATASETS

Dataset	K-values	Accuracy (%)	ARP (%)	ARR	F1-Measure (%)
Corel 1K	K = 2	92.47	93.38	12.53	15.45
	K = 3	94.38	95.56	14.24	17.43
	K = 4	96.13	97.27	16.53	19.42
	K = 5	99.63	99.51	18.82	27.28
Corel 5K	K = 2	91.35	91.32	15.48	31.22
	K = 3	94.28	93.50	17.47	33.12
	K = 4	96.12	95.39	18.32	36.48
	K = 5	98.37	97.13	21.28	38.31
Corel 10K	K = 2	90.38	84.38	22.21	35.34
	K = 3	91.32	88.37	23.12	38.38
	K = 4	93.48	91.37	25.28	40.47
	K = 5	95.28	94.38	28.42	42.38
Oxford 5K	K = 2	92.27	91.42	28.26	35.32
	K = 3	94.18	93.44	30.52	37.41
	K = 4	95.62	94.62	32.03	39.12
	K = 5	96.73	95.44	33.33	41.26
Paris 6K	K = 2	91.42	89.87	12.18	34.16
	K = 3	93.11	91.76	14.12	36.22
	K = 4	94.76	92.84	17.16	38.23
	K = 5	95.44	94.12	19.82	40.37

TABLE IV. COMPUTATIONAL COMPLEXITY OF THE PROPOSED METHODS USING THE OVERALL DATASETS

Methods	Training Time (per batch) (s)	Inference Time (per image) (ms)	GPU Memory Usage (MB)	FLOP (G)	Params (M)	Confidence Interval for Training Time (s)
CNN	47.38	18.32	580	8.3	33.28	47.38 ± 1.21
ResNet	42.48	16.38	610	7.9	35.45	42.48 ± 1.08
VGGNet	41.38	15.44	570	7.5	37.37	41.38 ± 0.96
Inception v3	39.37	12.39	590	7.1	39.38	39.37 ± 0.89
Proposed Adaptive Adam-CNN	49.11	19.47	630	8.4	41.38	49.11 ± 1.34

Table V illustrates the performance analysis of the proposed approach with previous approaches using CBIR metrics. The values are reported as mean ± standard deviation over multiple runs to illustrate the stability of the proposed approach. The proposed adaptive Adam-CNN achieves the highest mean values with lower variation compared to all existing methods across all datasets. These results confirm that the adaptive optimizer not only improves CBIR performance but also provides stable and reliable retrieval results.

TABLE V. PERFORMANCE ANALYSIS OF THE PROPOSED ADAPTIVE ADAM-CNN METHOD WITH EXISTING METHODS USING CBIR METRICS

Dataset	Methods	P@10	mAP	R@10	NDCG@10
Corel 1K	CNN	0.9037 ± 0.0124	0.8732 ± 0.0151	0.8842 ± 0.0068	0.9118 ± 0.0106
	ResNet	0.9248 ± 0.0116	0.8941 ± 0.0137	0.9017 ± 0.0062	0.9314 ± 0.0098
	VGGNet	0.9456 ± 0.0102	0.9164 ± 0.0124	0.9243 ± 0.0058	0.9527 ± 0.0086
	InceptionV3	0.9728 ± 0.0087	0.9483 ± 0.0106	0.9518 ± 0.0051	0.9788 ± 0.0071
	Contrastive learning	0.9842 ± 0.0068	0.9562 ± 0.0094	0.9635 ± 0.0048	0.9871 ± 0.0059
	Vision Transformer	0.9913 ± 0.0054	0.9728 ± 0.0082	0.9786 ± 0.0041	0.9938 ± 0.0047
	Proposed Adaptive Adam-CNN	0.9987 ± 0.0042	0.9816 ± 0.0068	0.9989 ± 0.0031	0.9942 ± 0.0051
Corel 5K	CNN	0.8912 ± 0.0143	0.8126 ± 0.0172	0.8427 ± 0.0074	0.9044 ± 0.0128
	ResNet	0.9146 ± 0.0131	0.8365 ± 0.0161	0.8614 ± 0.0071	0.9268 ± 0.0114
	VGGNet	0.9385 ± 0.0124	0.8728 ± 0.0153	0.8892 ± 0.0067	0.9491 ± 0.0107
	InceptionV3	0.9563 ± 0.0116	0.8947 ± 0.0142	0.9146 ± 0.0062	0.9637 ± 0.0099
	Contrastive learning	0.9628 ± 0.0108	0.9054 ± 0.0136	0.9278 ± 0.0058	0.9712 ± 0.0091
	Vision Transformer	0.9684 ± 0.0101	0.9132 ± 0.0129	0.9384 ± 0.0051	0.9753 ± 0.0087
	Proposed Adaptive Adam-CNN	0.9738 ± 0.0106	0.9213 ± 0.0142	0.9467 ± 0.0049	0.9784 ± 0.0091
Corel 10K	CNN	0.8536 ± 0.0168	0.7142 ± 0.0214	0.8218 ± 0.0082	0.8721 ± 0.0157
	ResNet	0.8761 ± 0.0152	0.7426 ± 0.0198	0.8442 ± 0.0078	0.8938 ± 0.0141
	VGGNet	0.9124 ± 0.0146	0.7815 ± 0.0192	0.8735 ± 0.0071	0.9267 ± 0.0133
	InceptionV3	0.9317 ± 0.0142	0.8064 ± 0.0187	0.8961 ± 0.0067	0.9425 ± 0.0126
	Contrastive learning	0.9368 ± 0.0139	0.8121 ± 0.0185	0.9047 ± 0.0061	0.9478 ± 0.0121
	Vision Transformer	0.9395 ± 0.0136	0.8183 ± 0.0181	0.9132 ± 0.0057	0.9496 ± 0.0118
	Proposed Adaptive Adam-CNN	0.9426 ± 0.0138	0.8214 ± 0.0172	0.9218 ± 0.0053	0.9518 ± 0.0116
Oxford 5K	CNN	0.9014 ± 0.0136	0.8265 ± 0.0164	0.8517 ± 0.0069	0.9128 ± 0.0118
	ResNet	0.9187 ± 0.0128	0.8517 ± 0.0152	0.8764 ± 0.0064	0.9314 ± 0.0109
	VGGNet	0.9426 ± 0.0117	0.8768 ± 0.0141	0.9038 ± 0.0058	0.9556 ± 0.0098
	InceptionV3	0.9568 ± 0.0108	0.8879 ± 0.0135	0.9214 ± 0.0053	0.9667 ± 0.0091
	Contrastive learning	0.9617 ± 0.0103	0.8914 ± 0.0132	0.9328 ± 0.0051	0.9692 ± 0.0088
	Vision Transformer	0.9642 ± 0.0099	0.8936 ± 0.0130	0.9417 ± 0.0048	0.9711 ± 0.0085
	Proposed Adaptive Adam-CNN	0.9673 ± 0.0098	0.8945 ± 0.0131	0.9486 ± 0.0045	0.9728 ± 0.0084
Paris 6K	CNN	0.8945 ± 0.0141	0.8018 ± 0.0173	0.8362 ± 0.0072	0.9041 ± 0.0124
	ResNet	0.9132 ± 0.0133	0.8367 ± 0.0161	0.8627 ± 0.0068	0.9246 ± 0.0113
	VGGNet	0.9354 ± 0.0125	0.8624 ± 0.0152	0.8915 ± 0.0061	0.9472 ± 0.0106
	InceptionV3	0.9473 ± 0.0118	0.8731 ± 0.0144	0.9074 ± 0.0057	0.9568 ± 0.0098
	Contrastive learning	0.9508 ± 0.0115	0.8772 ± 0.0141	0.9183 ± 0.0054	0.9594 ± 0.0093
	Vision Transformer	0.9526 ± 0.0113	0.8798 ± 0.0139	0.9276 ± 0.0051	0.9608 ± 0.0091
	Proposed Adaptive Adam-CNN	0.9544 ± 0.0112	0.8816 ± 0.0153	0.9364 ± 0.0049	0.9619 ± 0.0097
Caltech-101	CNN	0.9128 ± 0.0132	0.8341 ± 0.0163	0.8724 ± 0.0067	0.9238 ± 0.0115
	ResNet	0.9286 ± 0.0124	0.8523 ± 0.0151	0.8918 ± 0.0063	0.9396 ± 0.0107
	VGGNet	0.9412 ± 0.0116	0.8715 ± 0.0143	0.9087 ± 0.0059	0.9527 ± 0.0098
	InceptionV3	0.9498 ± 0.0111	0.8812 ± 0.0138	0.9216 ± 0.0054	0.9601 ± 0.0093
	Contrastive learning	0.9521 ± 0.0109	0.8838 ± 0.0135	0.9314 ± 0.0051	0.9624 ± 0.0091
	Vision Transformer	0.9547 ± 0.0108	0.8856 ± 0.0132	0.9398 ± 0.0048	0.9638 ± 0.0089
	Proposed Adaptive Adam-CNN	0.9563 ± 0.0107	0.8894 ± 0.0128	0.9482 ± 0.0045	0.9751 ± 0.0093

Table VI specifies the cross-dataset analysis of the proposed with previous approaches using CBIR metrics for better generalization. The cross-dataset analysis demonstrates the generalization capability of the proposed Adaptive Adam-CNN under unseen data distributions. Since Corel 1K and Corel 5K are smaller subsets of the same dataset family as Corel 10K, only Corel 10K is considered for cross-dataset validation. Corel 10K provides greater diversity and a larger number of images, making it more suitable for evaluating the generalization capability of the proposed model. When trained on Corel datasets and evaluated on Oxford 5K, Paris 6K, and Caltech-101, the proposed method achieves the highest P@10 of 0.9724 and mAP of 0.9138, demonstrating the effective transferability of learned features. Likewise, training on Oxford 5K and Paris 6K still enables the proposed model to outperform all baseline methods on the Corel datasets with a Normalized Discounted Cumulative Gain at rank 10 (NDCG@10) of 0.9716. The best overall cross-dataset performance is obtained when the model is trained on Corel 10K, where it reached 0.9756 P@10 and 0.9187 mAP. These results confirm that the adaptive switching between Adam and AMSGrad improves robustness and prevents overfitting to a single dataset distribution.

Although the proposed Adaptive Adam-CNN exhibits higher FLOPs due to adaptive optimization and enhanced feature extraction, it achieves improved retrieval effectiveness and feature representation capability. The proposed Adaptive Adam-CNN maintains competitive computational complexity while achieving improved retrieval accuracy and feature representation capability.

Table VII depicts the ablation study of proposed method using multiple datasets. The proposed adaptive Adam-CNN approach performs significantly better compared to others in all performance metrics. The enhanced inclination in accuracy through adaptive Adam suggests that this improvement enhances the generalization capability of the model. The AdaBound and RAdam achieve the lowest retrieval performance because their fixed learning rate leads to slower and less stable convergence. AdaBelief-based CNN (AdaBelief-CNN) improves the performance by adapting the learning rate according to the gradient history; however, it still produces lower accuracy than the proposed method. These improvements prove that optimizer selection crucially impacts performance and validate the proposed cooperative optimizer integration. This analysis demonstrates that combining adaptive norming with the robustness of AMSGrad and the exploration of Adam significantly enhances the optimizer's ability to converge faster and generalize better.

TABLE VI. CROSS-DATASET ANALYSIS OF THE PROPOSED METHOD WITH EXISTING METHODS USING CBIR METRICS FOR BETTER GENERALIZATION

Dataset	Methods	P@10	mAP	R@10	NDCG@10
Trained on Corel 10K; Tested on Oxford 5K	CNN	0.8738	0.7816	0.8424	0.8862
	ResNet	0.8944	0.8063	0.8618	0.9018
	VGGNet	0.9172	0.8327	0.8837	0.9261
	Inception v3	0.9368	0.8614	0.9046	0.9448
	Contrastive Learning	0.9481	0.8783	0.9214	0.9562
	Vision Transformer	0.9567	0.8892	0.9368	0.9645
	Proposed Adaptive Adam-CNN	0.9724	0.9138	0.9527	0.9781
Trained on Corel 10K; Tested on Paris 6K	CNN	0.8617	0.7544	0.8316	0.8723
	ResNet	0.8823	0.7815	0.8524	0.8917
	VGGNet	0.9042	0.8148	0.8743	0.9153
	Inception v3	0.9271	0.8436	0.8965	0.9368
	Contrastive Learning	0.9388	0.8615	0.9138	0.9486
	Vision Transformer	0.9464	0.8743	0.9284	0.9562
	Proposed Adaptive Adam-CNN	0.9637	0.9024	0.9462	0.9716
Trained on Corel 10K; Tested on Caltech-101	CNN	0.8895	0.7924	0.8518	0.9012
	ResNet	0.9086	0.8147	0.8726	0.9183
	VGGNet	0.9264	0.8389	0.8943	0.9364
	Inception v3	0.9448	0.8673	0.9157	0.9521
	Contrastive Learning	0.9537	0.8814	0.9316	0.9618
	Vision Transformer	0.9612	0.8938	0.9448	0.9694
	Proposed Adaptive Adam-CNN	0.9756	0.9187	0.9617	0.9813
Trained on Oxford 5K; Tested on Corel 10K	CNN	0.8426	0.7368	0.8247	0.8542
	ResNet	0.8643	0.7581	0.8462	0.8726
	VGGNet	0.8879	0.7844	0.8694	0.8941
	Inception v3	0.9085	0.8117	0.8918	0.9168
	Contrastive Learning	0.9216	0.8293	0.9085	0.9297
	Vision Transformer	0.9324	0.8465	0.9246	0.9414
	Proposed Adaptive Adam-CNN	0.9498	0.8731	0.9413	0.9587
Trained on Oxford 5K; Tested on Paris 6K	CNN	0.8741	0.7684	0.8468	0.8836
	ResNet	0.8957	0.7923	0.8675	0.9021
	VGGNet	0.9174	0.8198	0.8894	0.9237
	Inception v3	0.9382	0.8476	0.9117	0.9445
	Contrastive Learning	0.9496	0.8651	0.9276	0.9568
	Vision Transformer	0.9571	0.8784	0.9418	0.9647
	Proposed Adaptive Adam-CNN	0.9713	0.9042	0.9947	0.9775
Trained on Oxford 5K; Tested on Caltech-101	CNN	0.8568	0.7492	0.8364	0.8675
	ResNet	0.8784	0.7741	0.8572	0.8864
	VGGNet	0.9012	0.8018	0.8795	0.9096
	Inception v3	0.9235	0.8493	0.9028	0.9442
	Contrastive Learning	0.9357	0.8493	0.9184	0.9442
	Vision Transformer	0.9448	0.8635	0.9326	0.9537
	Proposed Adaptive Adam-CNN	0.9617	0.8898	0.9498	0.9696
Trained on Paris 6K; Tested on Corel 10K	CNN	0.8367	0.7284	0.8216	0.8469
	ResNet	0.8579	0.7526	0.8438	0.8665
	VGGNet	0.8815	0.7798	0.8667	0.8893
	Inception v3	0.9032	0.8076	0.8893	0.9118
	Contrastive Learning	0.9167	0.8423	0.9061	0.9365
	Vision Transformer	0.9278	0.8423	0.9217	0.9365
	Proposed Adaptive Adam-CNN	0.9456	0.8697	0.9384	0.9548
Trained on Paris 6K; Tested on Oxford 5K	CNN	0.8684	0.7597	0.8442	0.8782
	ResNet	0.8896	0.7835	0.8651	0.8974
	VGGNet	0.9118	0.8124	0.8874	0.9206
	Inception v3	0.9346	0.8413	0.9098	0.9417
	Contrastive Learning	0.9462	0.8594	0.9257	0.9539
	Vision Transformer	0.9541	0.8728	0.9394	0.9624
	Proposed Adaptive Adam-CNN	0.9694	0.8996	0.9568	0.9735
Trained on Paris 6K; Tested on Caltech-101	CNN	0.8512	0.7425	0.8337	0.8624
	ResNet	0.8727	0.7688	0.8546	0.8823
	VGGNet	0.8954	0.7962	0.8763	0.9051
	Inception v3	0.9188	0.8265	0.8984	0.9274
	Contrastive Learning	0.9306	0.8447	0.9148	0.9401
	Vision Transformer	0.9397	0.8586	0.9297	0.9495
	Proposed Adaptive Adam-CNN	0.9578	0.8854	0.9472	0.9668

Trained on Caltech-101; Tested on Corel 10K	CNN	0.8474	0.7398	0.8282	0.8581
	ResNet	0.8695	0.7641	0.8496	0.8779
	VGGNet	0.8923	0.7916	0.8725	0.9004
	Inception v3	0.9148	0.8193	0.8951	0.9227
	Contrastive Learning	0.9276	0.8378	0.9118	0.9356
	Vision Transformer	0.9385	0.8547	0.9274	0.9473
	Proposed Adaptive Adam-CNN	0.9554	0.8816	0.9446	0.9649
Trained on Caltech-101; Tested on Oxford 5K	CNN	0.8635	0.7538	0.8415	0.8738
	ResNet	0.8847	0.7786	0.8627	0.8934
	VGGNet	0.9076	0.8064	0.8846	0.9167
	Inception v3	0.9295	0.8352	0.9027	0.9386
	Contrastive Learning	0.9423	0.8537	0.9235	0.9518
	Vision Transformer	0.9508	0.8675	0.9378	0.9606
	Proposed Adaptive Adam-CNN	0.9665	0.8943	0.9544	0.9737
Trained on Caltech-101; Tested on Paris 6K	CNN	0.8558	0.7467	0.8358	0.8662
	ResNet	0.8764	0.7714	0.8564	0.8857
	VGGNet	0.8992	0.7988	0.8788	0.9089
	Inception v3	0.9217	0.8284	0.9014	0.9308
	Contrastive Learning	0.9344	0.8466	0.9718	0.9437
	Vision Transformer	0.9435	0.8602	0.9325	0.9524
	Proposed Adaptive Adam-CNN	0.9606	0.8875	0.9496	0.9687

TABLE VII. ABLATION STUDY OF THE PROPOSED METHOD BASED ON MULTIPLE DATASETS

Dataset	K-values	Accuracy (%)	ARP (%)	ARR	F1-measure	t-test	*p*-value	ANOVA
Corel 1K	CNN-SGD	91.38	93.28	08.66	16.49	0.039	0.036	0.036
	Adam-CNN	93.22	96.22	09.12	19.78	0.032	0.034	0.031
	AMSGrad-CNN	95.42	97.38	11.66	20.67	0.023	0.031	0.027
	Adam-norm-CNN	97.38	97.74	13.22	20.13	0.027	0.030	0.029
	AMSGrad-norm-CNN	98.42	98.66	15.58	20.88	0.019	0.029	0.032
	AdaBound-CNN	96.48	96.53	16.32	23.58	0.017	0.028	0.028
	RAdam-CNN	97.57	97.87	17.62	25.68	0.015	0.031	0.033
	AdaBelief-CNN	98.41	98.59	17.97	26.68	0.016	0.029	0.030
	Adaptive Adam-CNN	99.63	99.51	18.82	27.28	0.014	0.028	0.026
Corel 5K	CNN-SGD	88.37	92.22	09.67	31.38	0.023	0.038	0.037
	Adam-CNN	93.38	94.82	10.56	34.27	0.021	0.033	0.030
	AMSGrad-CNN	97.38	95.38	11.12	36.37	0.018	0.030	0.028
	Adam-norm-CNN	92.48	93.12	12.48	37.06	0.056	0.038	0.027
	AMSGrad-norm-CNN	93.48	94.47	14.58	35.48	0.024	0.034	0.045
	AdaBound-CNN	95.87	95.44	15.53	36.68	0.022	0.031	0.037
	RAdam-CNN	96.12	96.24	17.78	37.64	0.018	0.029	0.035
	AdaBelief-CNN	97.43	96.89	18.91	37.93	0.017	0.028	0.025
	Adaptive Adam-CNN	98.37	97.13	21.28	38.31	0.016	0.027	0.023
Corel 10K	CNN-SGD	88.47	89.38	19.38	35.12	0.078	0.048	0.026
	Adam-CNN	91.37	90.37	20.28	37.47	0.067	0.041	0.029
	AMSGrad-CNN	94.29	92.47	22.37	40.37	0.056	0.034	0.024
	Adam-norm-CNN	93.18	87.37	24.28	39.42	0.053	0.037	0.039
	AMSGrad-norm-CNN	94.02	90.38	26.30	40.57	0.049	0.032	0.035
	AdaBound-CNN	93.53	91.43	26.91	40.65	0.032	0.030	0.042
	RAdam-CNN	94.68	92.43	27.54	41.49	0.025	0.028	0.045
	AdaBelief-CNN	94.87	93.13	27.89	41.70	0.020	0.027	0.026
	Adaptive Adam-CNN	95.28	94.38	28.42	42.38	0.015	0.026	0.024
Oxford 5K	CNN-SGD	95.39	95.29	23.32	18.58	0.025	0.033	0.030
	Adam-CNN	96.37	96.23	24.48	19.48	0.023	0.031	0.029
	AMSGrad-CNN	97.37	97.43	26.39	20.48	0.020	0.028	0.036
	Adam-norm-CNN	91.12	89.38	27.38	15.49	0.042	0.039	0.037
	AMSGrad-norm-CNN	96.38	91.38	29.58	16.73	0.030	0.038	0.035
	AdaBound-CNN	97.42	94.44	30.76	19.43	0.026	0.034	0.038
	RAdam-CNN	98.53	96.24	31.95	23.48	0.021	0.027	0.032
	AdaBelief-CNN	98.78	97.42	32.97	25.49	0.020	0.026	0.027
	Adaptive Adam-CNN	99.63	99.51	33.33	27.28	0.019	0.024	0.030
Paris 6K	CNN-SGD	93.24	91.39	08.56	33.58	0.028	0.032	0.031
	Adam-CNN	94.49	93.38	10.68	34.47	0.026	0.031	0.040
	AMSGrad-CNN	96.38	94.83	12.57	35.48	0.025	0.028	0.049
	Adam-norm-CNN	89.58	88.40	14.47	28.46	0.035	0.038	0.043
	AMSGrad-norm-CNN	90.48	89.67	15.48	30.57	0.033	0.036	0.041
	AdaBound-CNN	93.53	91.43	16.83	33.48	0.028	0.031	0.037
	RAdam-CNN	95.57	93.58	17.53	34.58	0.024	0.028	0.023
	AdaBelief-CNN	96.47	94.24	18.48	35.68	0.025	0.026	0.045
	Adaptive Adam-CNN	97.38	95.38	19.82	36.37	0.023	0.025	0.027

Each experimental finding is statistically estimated through t-tests, p-value, and ANOVA. The statistical significance tests were conducted by comparing each optimization-based CNN model against the baseline approach using paired t-test and ANOVA analysis. To thoroughly assess the effectiveness of the proposed Adaptive Adam-CNN approach, statistical hypothesis tests are performed on important effectiveness indices. To thoroughly evaluate the performance of proposed adaptive Adam-CNN approach, a set of statistical hypothesis tests was conducted. The null hypothesis (H_0) assumes that there is no significant difference between the proposed approach and the comparative models, while the alternative hypothesis (H_1), postulated that Adaptive Adam-CNN significantly outperforms existing methods.

The t-test is employed to compare mean performance index, such as accuracy and F1-measure, among different approaches over multiple folds. The p-value is utilized for identifying the likelihood that observed variations are present through chance, with significance described at $p < 0.05$. The one-way ANOVA test is employed further to estimate a difference between different optimization variants. Effect size analysis indicates moderate to large improvements across datasets. These tests are performed to estimate whether the experimental enhancements in retrieval effectiveness are statistically significant and not because of chance. This statistical evidence firmly supports the consistency and robustness of proposed approach in image retrieval across diverse data distributions.

Additionally, confidence intervals are computed to ensure the robustness and reliability of the reported metrics. These tests are used for estimating whether the observed enhancements in retrieval performance are statistically crucial and not because of chance. All these values are below standard threshold of 0.05, demonstrating strong statistical importance. This statistical evidence firmly supports the reliability and robustness of the proposed approach in image retrieval over different data distributions.

Fig. 3 represents the sample retrieval performance acquired by the adaptive Adam-CNN approach. It demonstrates that a group of appropriate sample images is retrieved on the application of similar images as a query image. It illustrates that the adaptive Adam-CNN approach has properly retrieved the images from the applied database.

Fig. 4 illustrates the graphical representation of confusion matrix of the adaptive Adam-CNN approach. In the confusion matrix, the most query images are correctly matched to their corresponding categories with minimal inter-class misclassification. The confusion matrix shows that correctly classified samples are dominant compared to misclassified samples.

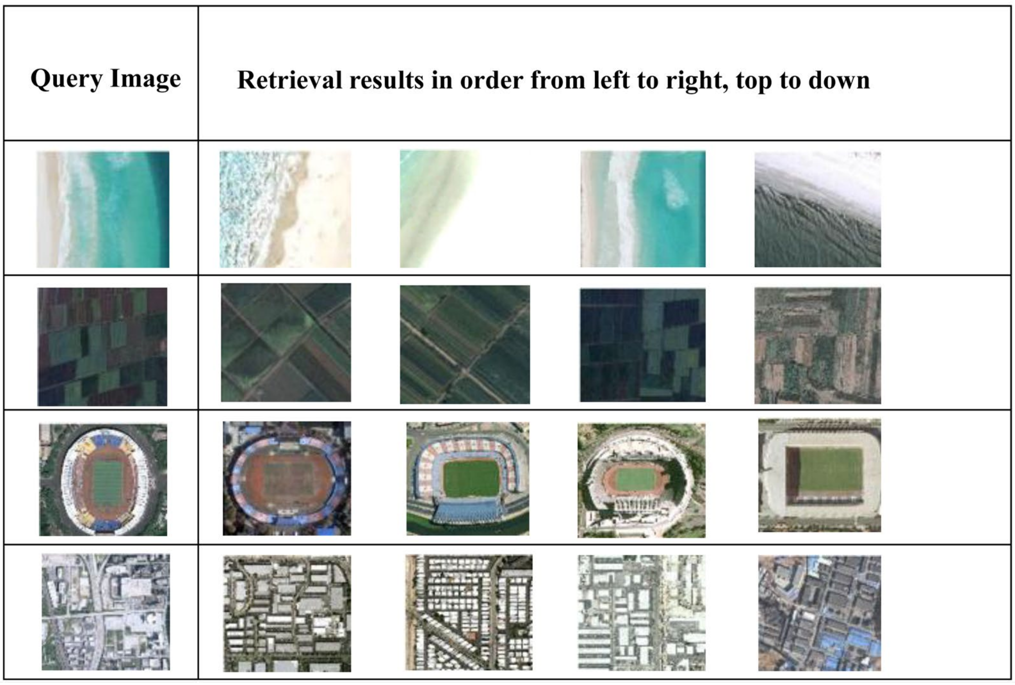
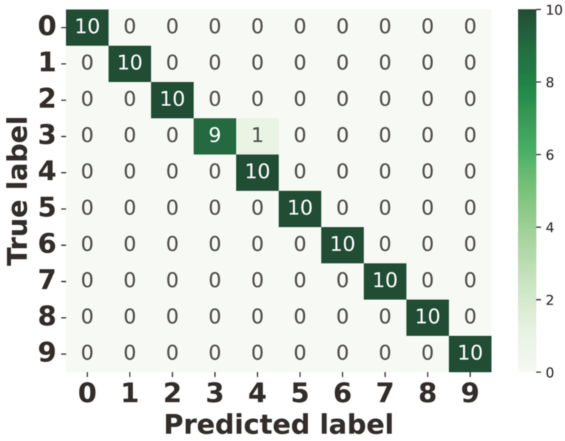
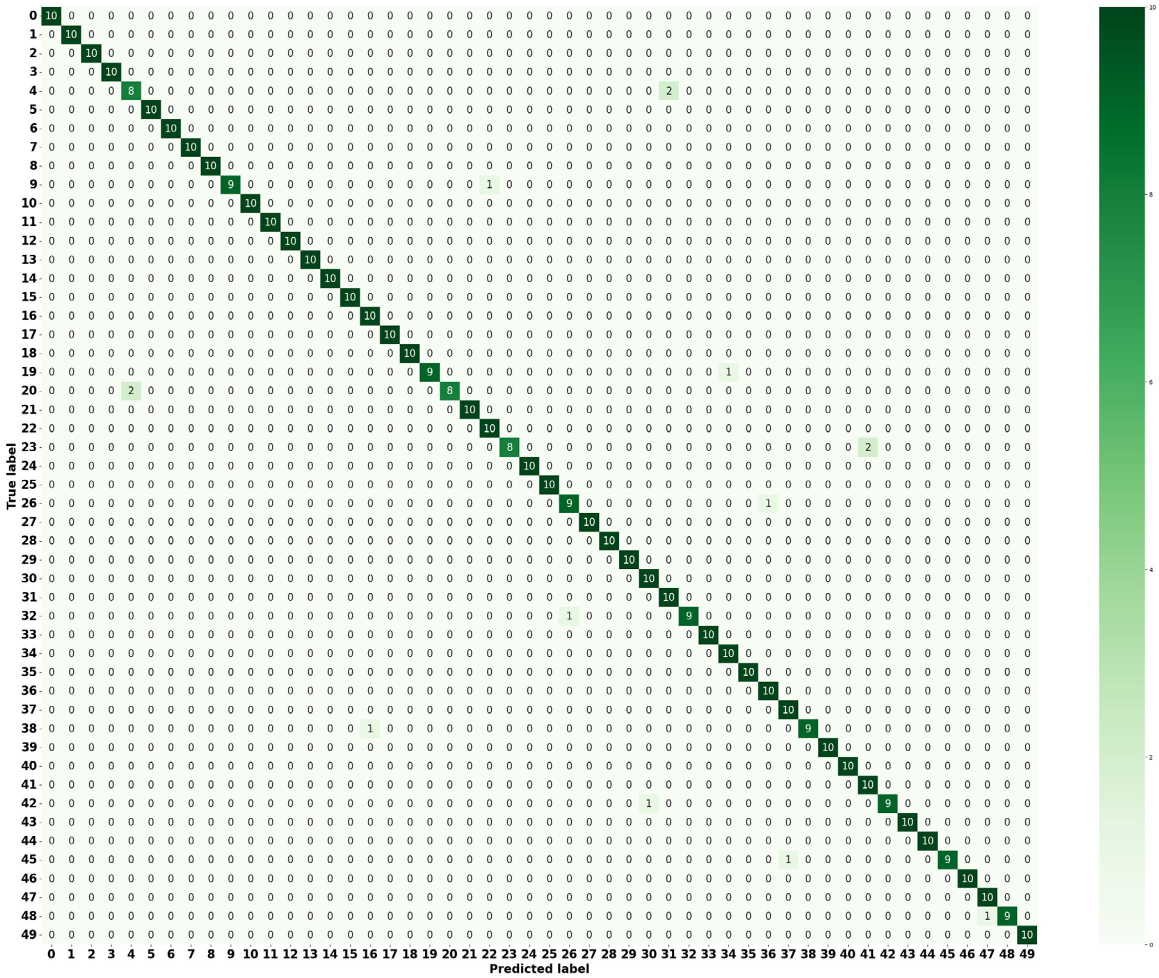


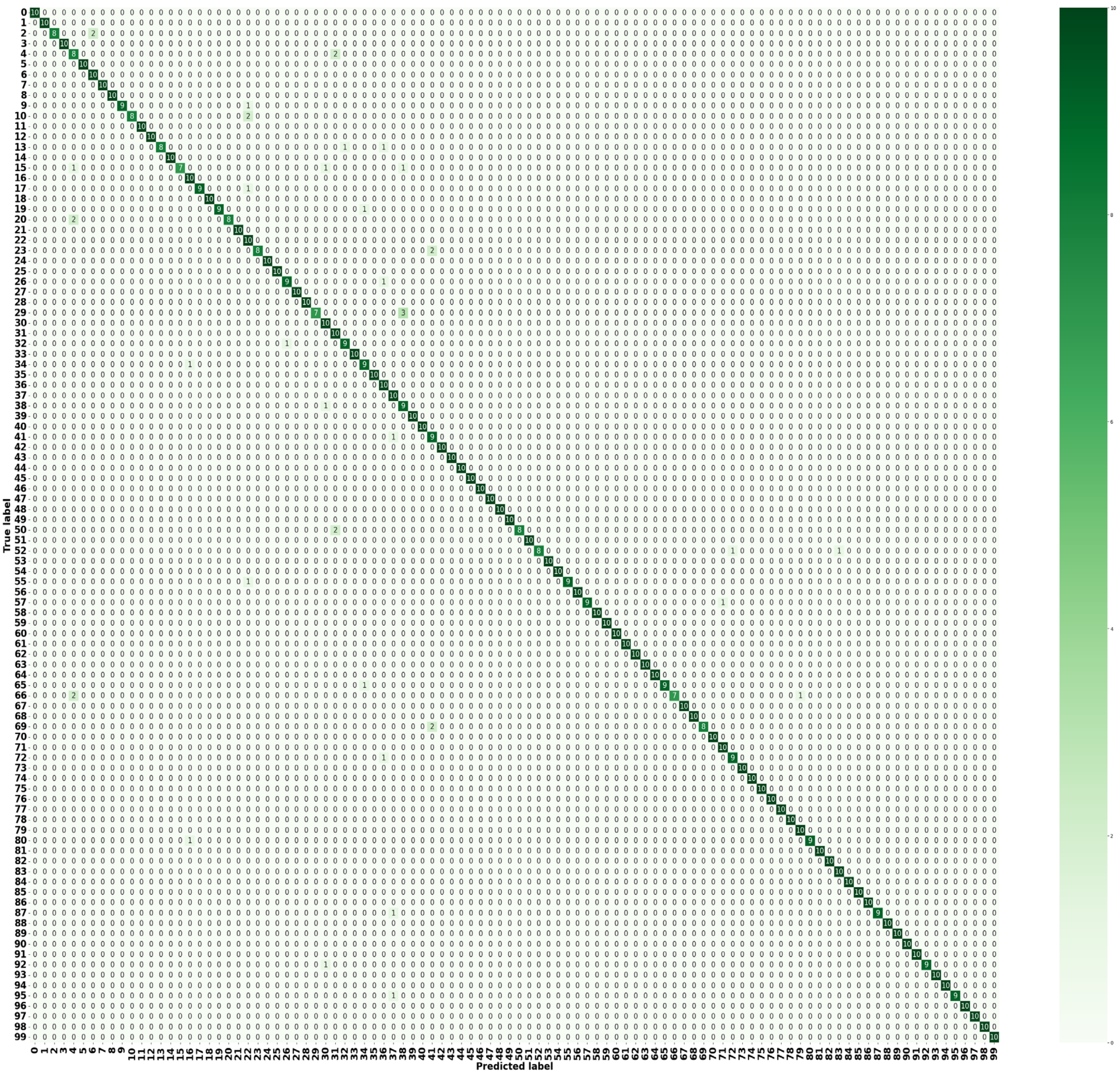
Fig. 3. Sample retrieval performance acquired by the adaptive Adam-CNN approach.



(a)



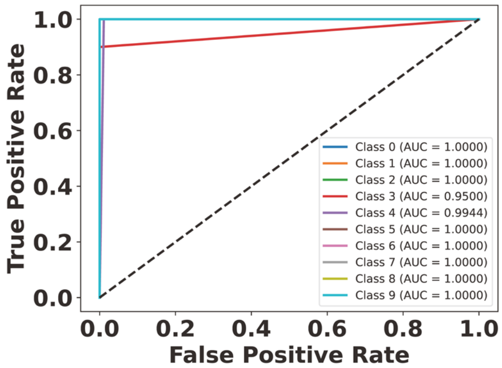
(b)



(c)

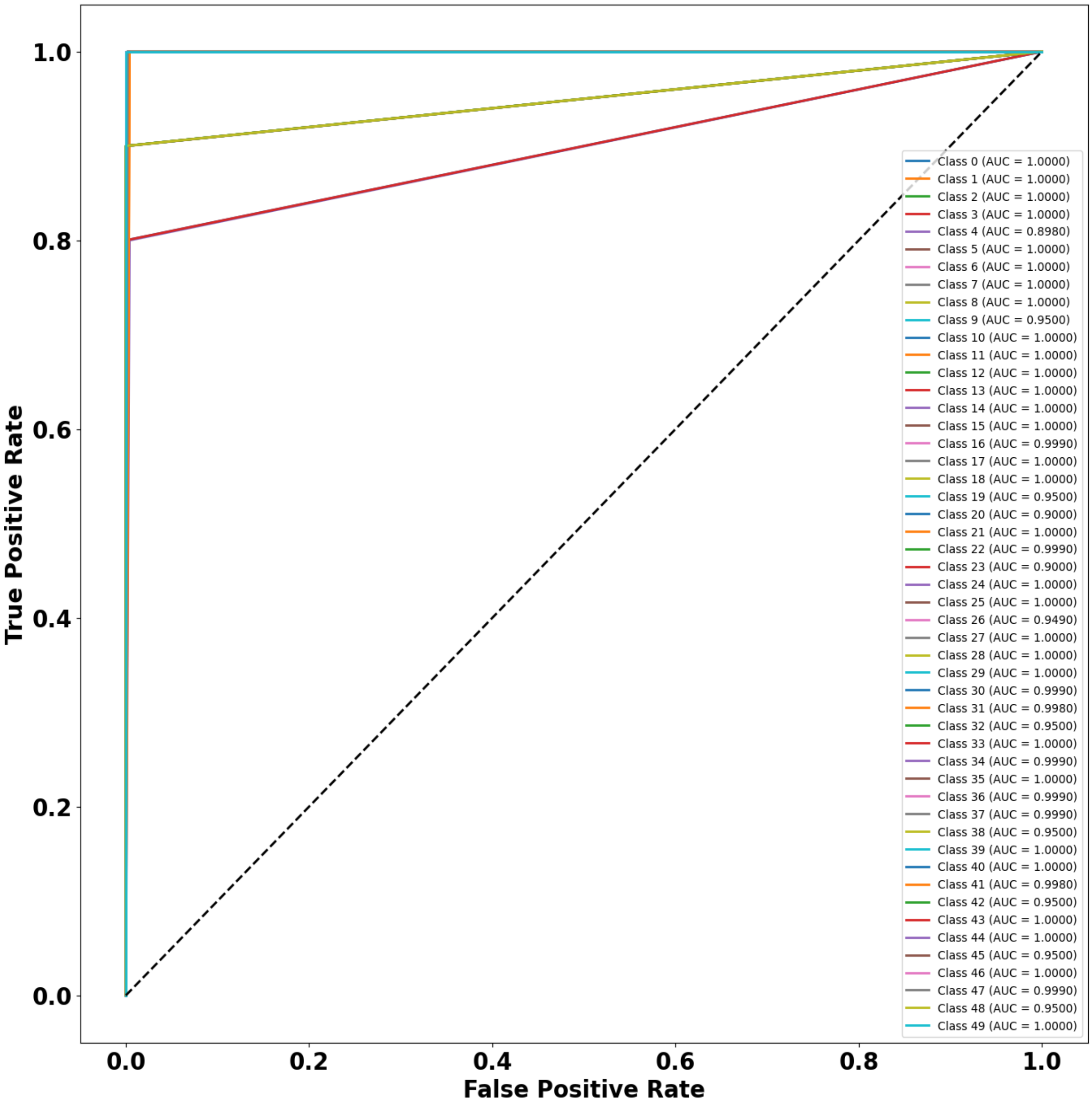
Fig. 4. Confusion matrix of the proposed method for (a) Corel 1K dataset; (b) Corel 5K dataset; and (c) Corel 10K dataset.

Fig. 5 illustrates the graphical depiction of Receiver Operating Characteristic (ROC) curve of the adaptive Adam-CNN approach. The Area Under the Curve (AUC) values of the datasets used are consistently close to 1.0, demonstrating effective class-wise discrimination with reduced misclassification rates. An improved AUC indicates that the classifier handles greater specificity and sensitivity even if the decision threshold differs. A curve remainders in the upper left quadrant for all datasets, reflecting strong performance over multiple operating points.



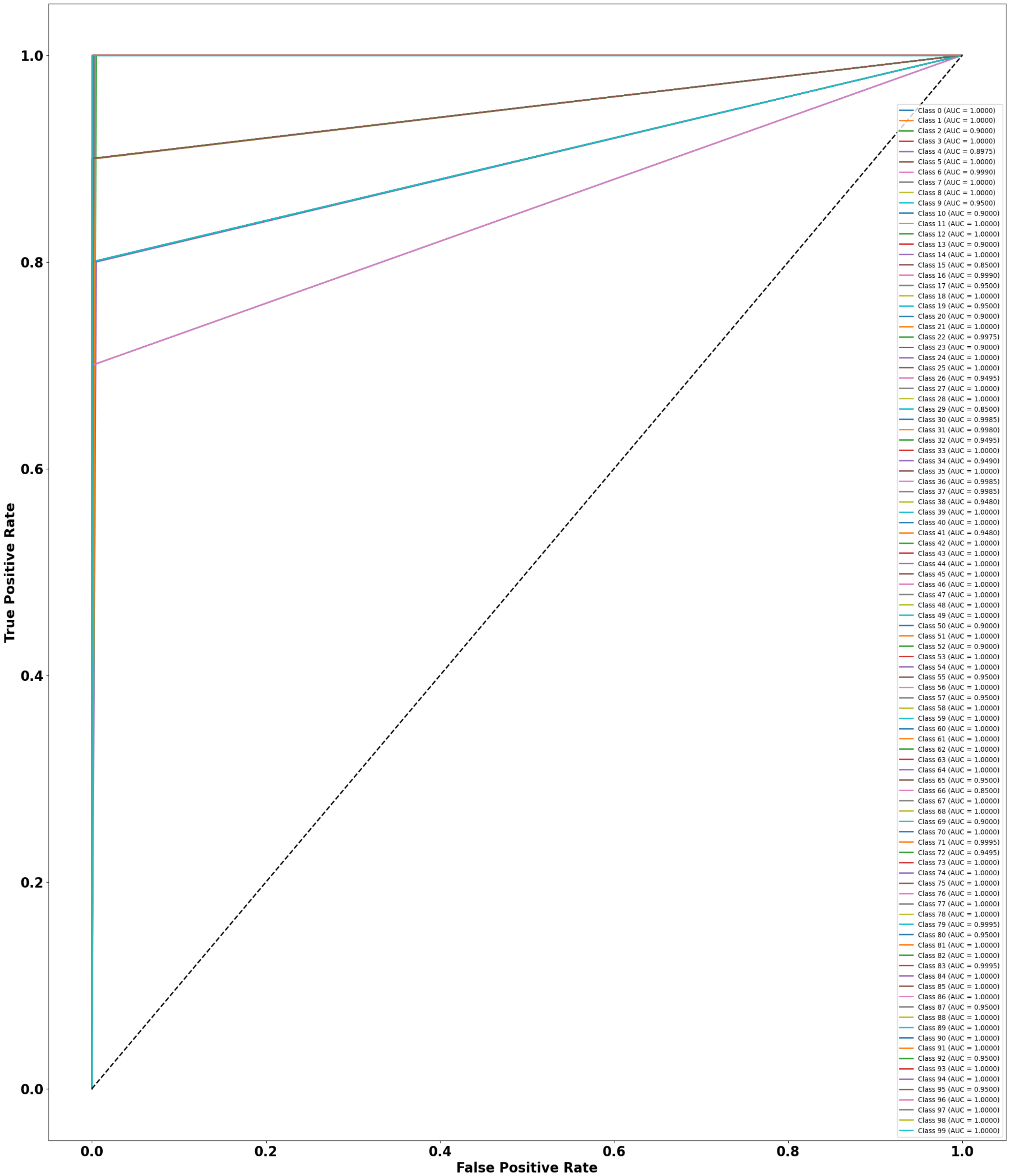
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class 0 (AUC = 1.0000)
Class 1 (AUC = 1.0000)
Class 2 (AUC = 1.0000)
Class 3 (AUC = 0.9500)
Class 4 (AUC = 0.9944)
Class 5 (AUC = 1.0000)
Class 6 (AUC = 1.0000)
Class 7 (AUC = 1.0000)
Class 8 (AUC = 1.0000)
Class 9 (AUC = 1.0000)

(a)



True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class 0 (AUC = 1.0000)
Class 1 (AUC = 1.0000)
Class 2 (AUC = 1.0000)
Class 3 (AUC = 1.0000)
Class 4 (AUC = 0.8980)
Class 5 (AUC = 1.0000)
Class 6 (AUC = 1.0000)
Class 7 (AUC = 1.0000)
Class 8 (AUC = 1.0000)
Class 9 (AUC = 0.9500)
Class 10 (AUC = 1.0000)
Class 11 (AUC = 1.0000)
Class 12 (AUC = 1.0000)
Class 13 (AUC = 1.0000)
Class 14 (AUC = 1.0000)
Class 15 (AUC = 1.0000)
Class 16 (AUC = 0.9990)
Class 17 (AUC = 1.0000)
Class 18 (AUC = 1.0000)
Class 19 (AUC = 0.9500)
Class 20 (AUC = 0.9000)
Class 21 (AUC = 1.0000)
Class 22 (AUC = 0.9990)
Class 23 (AUC = 0.9000)
Class 24 (AUC = 1.0000)
Class 25 (AUC = 1.0000)
Class 26 (AUC = 0.9490)
Class 27 (AUC = 1.0000)
Class 28 (AUC = 1.0000)
Class 29 (AUC = 1.0000)
Class 30 (AUC = 0.9990)
Class 31 (AUC = 0.9980)
Class 32 (AUC = 0.9500)
Class 33 (AUC = 1.0000)
Class 34 (AUC = 0.9990)
Class 35 (AUC = 1.0000)
Class 36 (AUC = 0.9990)
Class 37 (AUC = 0.9990)
Class 38 (AUC = 0.9500)
Class 39 (AUC = 1.0000)
Class 40 (AUC = 1.0000)
Class 41 (AUC = 0.9980)
Class 42 (AUC = 0.9500)
Class 43 (AUC = 1.0000)
Class 44 (AUC = 1.0000)
Class 45 (AUC = 0.9500)
Class 46 (AUC = 1.0000)
Class 47 (AUC = 0.9990)
Class 48 (AUC = 0.9500)
Class 49 (AUC = 1.0000)

(b)



(c)

Fig. 5. ROC Curve of the proposed method for (a) Corel 1K dataset; (b) Corel 5K dataset; and (c) Corel 10K dataset.

The validation curves demonstrate that the proposed adaptive Adam-CNN converges steadily without exhibiting significant overfitting. Over all datasets, the gap between training and validation accuracy remained within 0.8–1.2%, while the validation loss decreased consistently and stabilized after approximately 100 epochs. No sudden increase in validation loss is observed during later epochs, indicating that the model did not memorize the training samples. This behavior significantly contributes to the combined effect of dropout regularization, adaptive norm-based optimizer switching, batch normalization, and early stopping. Specifically, the adaptive transition between Adam and AMSGrad stabilizes the gradient updates and prevents more parameter oscillation, thereby improving generalization even though the network involves a comparatively large number of trainable parameters.

Table VIII demonstrates the sensitivity analysis of the proposed adaptive Adam-CNN. Table IX demonstrates the robustness analysis of the proposed adaptive Adam-CNN using mAP. In Table VIII, the highest mAP of 0.9868 is achieved when the dropout rate is 0.5, and the adaptive norm threshold (A_0) is 2.0, confirming that these settings provide the best balance between regularization and optimization stability. Values lower or higher than these settings reduce performance. In Table IX, the robustness analysis further depicts that the proposed method maintains mAP values above 0.94 even under Gaussian noise, salt-and-pepper noise, and occlusion. Therefore, the performance improvement is not limited to a specific dataset and remains stable under parameter variations and image perturbations.

Fig. 6 specifies the convergence behavior of the proposed optimizer. The proposed Adaptive Adam-CNN reaches stable convergence within nearly 35 epochs, whereas Adam and AMSGrad require more than 50 epochs. In addition, the proposed method achieves lower final loss and higher validation accuracy, confirming that the adaptive switching mechanism improves optimization stability and convergence speed.

TABLE VIII. SENSITIVITY AND ROBUSTNESS OF THE PROPOSED METHOD USING MAP

Dropout	Norm Threshold (A_0)	mAP
0.2	1.5	0.9603
0.3	1.5	0.9682
0.4	1.5	0.9727
0.5	1.5	0.9729
0.2	2.0	0.9641
0.3	2.0	0.9735
0.4	2.0	0.9814
0.5	2.0	0.9868
0.2	2.5	0.9688
0.3	2.5	0.9724
0.4	2.5	0.9751
0.5	2.5	0.9773

TABLE IX. ROBUSTNESS OF THE PROPOSED METHOD USING MAP

Noise type	Noise Level (%)	mAP
Gaussian noise	5	0.9724
Gaussian noise	10	0.9588
Salt-and-pepper noise	5	0.9647
Salt-and-pepper noise	10	0.9513
Occlusion	10	0.9468

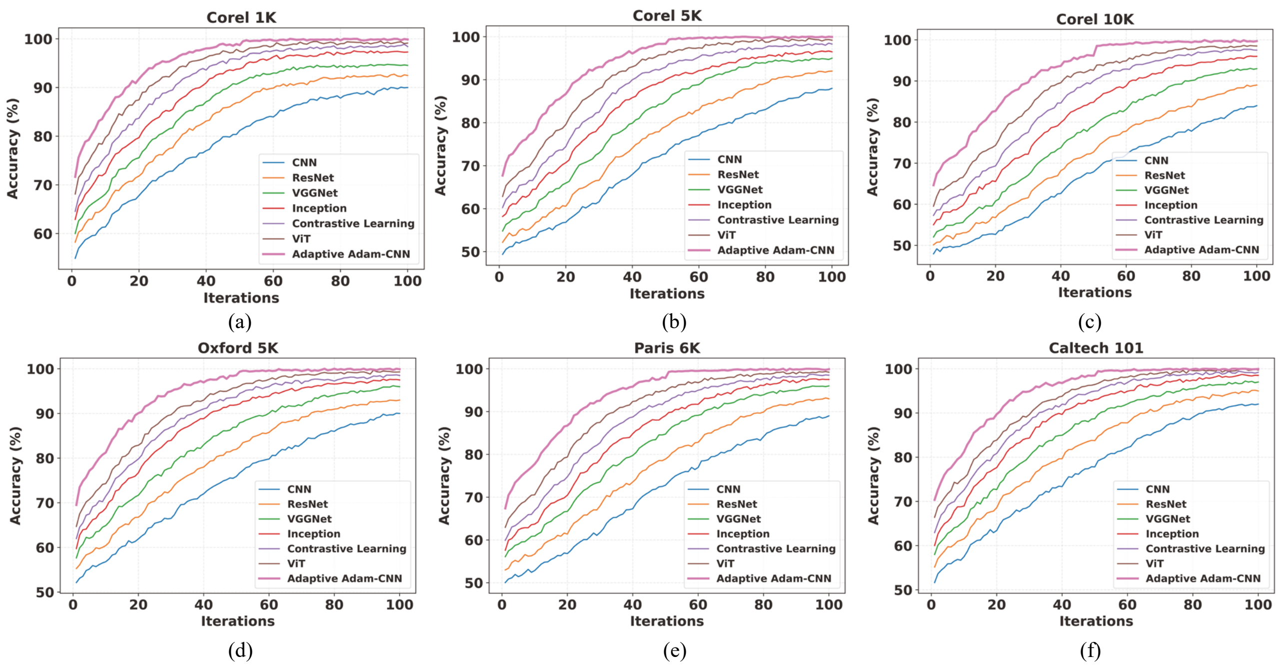


Fig. 6. Convergence behavior of the proposed optimizer (a) Corel 1K; (b) Corel 5K; (c) Corel 10K; (d) Oxford 5K; (e) Paris 6K; and (f) Caltech-101.

B. Comparative Analysis

Table X presents the comparative analysis of proposed methods and existing approaches using multiple datasets. Through estimating the data in the subsequent table, it is clear that the proposed approach delivers enhanced results for image retrieval over each dataset. Table XI demonstrates the comparison of proposed approach through Deep Texture Feature Histogram (DTFH) [29] and Model Quantization (M-Q) [30]. In this analysis, totally 12 number of retrieved images are considered for the proposed method. In this analysis, the exisiting DFF method [31] are reimplemented and FedAvg-RF [30] are locally implemented and reproduced the exact data splits, specific missing-value imputation and identical random seed considered for the proposed method. Table XII illustrates the comparative analysis of the proposed Adaptive Adam-CNN approach with the earlier Deep Feature Fusion (DFF) approaches using CBIR metrics.

TABLE X. COMPARATIVE ANALYSIS OF THE PROPOSED METHOD WITH EXISTING APPROACHES

Dataset	Methods	ARP (%)	ARR	F1-Measure
Corel 1K	CBIR-ANR [23]	84.38	10.13	-
	Color + Textures from semantic pyramid [24]	99.42	15.8	24.23
	cl-MSD-LBP-LDiPv [25]	84.38	10.13	-
	LB-CBIR [26]	99.2	9.2	16.83
	Proposed Adaptive Adam-CNN	99.51	18.82	27.28
Corel 5K	CBIR-ANR [23]	72.86	8.74	-
	Color + Textures from semantic pyramid [24]	95.33	19.32	31.83
	cl-MSD-LBP-LDiPv [25]	72.86	8.74	-
	LB-CBIR [26]	98.2	18.04	30.06
	Proposed Adaptive Adam-CNN	97.13	21.28	38.38
Corel 10k	CBIR-ANR [23]	63.31	7.60	-
	Color + Textures from semantic pyramid [24]	91.53	23.63	37.56
	cl-MSD-LBP-LDiPv [25]	63.31	7.60	-
	LB-CBIR [26]	90.2	22.1	36.08
	ELNDP [28]	59.65	26.85	-
	Proposed Adaptive Adam-CNN	94.38	28.42	42.38

TABLE XI. mAP PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH DTFH [29] AND M-Q [30]

Methods	Oxford 5k	Paris 6K
DTFH-VGG16 [29]	0.545	0.695
M-Q [30]	0.686	0.706
Proposed Adaptive Adam-CNN	0.894	0.881

TABLE XII. COMPARATIVE ANALYSIS OF THE PROPOSED ADAPTIVE ADAM-CNN METHOD WITH THE EXISTING DFF METHOD USING CBIR METRICS

Dataset	Methods	P@10	mAP	R@10	NDCG@10
Corel 1K	DFF [31]	0.9903 ± 0.0716	0.9628 ± 0.0909	0.9985 ± 0.0072	0.9910 ± 0.9910
	Proposed Adaptive Adam-CNN	0.9987 ± 0.0042	0.9816 ± 0.0068	0.9989 ± 0.0031	0.9942 ± 0.0051
Corel 10K	DFF [31]	0.8996 ± 0.2174	0.7424 ± 0.2576	0.0909 ± 0.0220	0.9091 ± 0.2064
	Proposed Adaptive Adam-CNN	0.9426 ± 0.0138	0.8214 ± 0.0172	0.9218 ± 0.0053	0.9518 ± 0.0116
Caltech-101	DFF [31]	0.9282 ± 0.1870	0.8549 ± 0.2258	0.1038 ± 0.0816	0.9307 ± 0.1825
	Proposed Adaptive Adam-CNN	0.9563 ± 0.0107	0.8894 ± 0.0128	0.9482 ± 0.0045	0.9751 ± 0.0093

C. Discussion

In this section, the limitations of preceding works and the findings of the proposed method are described. The limitations of existing works are 2DCNN-based CBIR with PCA [22] based on handcrafted features and dimensionality reduction, which may lead to the loss of important nonlinear characteristics. The CBIR-ANR approach [23] lacks model integration for hierarchical feature extraction, which constrains its capability to learn high-level abstractions from images. The model [24] does not adaptively learn optimal feature combinations and depends on t-SNE only for visualization without an enhancement of retrieval effectiveness. The Deep Boltzmann machine [26] suffered from high computational complexity and manual feature engineering, minimizing scalability. The DLIM approach [27] lacks consideration of optimizer variations and hyperparameter tuning, which could improve further model's performance on different datasets. The ELNDP model does not generalize well over moderately sized benchmark datasets. Thus, the proposed method addresses this challenge by employing a robust feature extraction and optimization strategy named the adaptive Adam optimizer with CNN, which improves the discriminability of learned features and maintains high precision even for larger values of k. This ensures more accurate and semantically relevant retrieval results, thereby illustrating the scalability and robustness of the proposed approach for diverse and moderately sized benchmark datasets.

D. Ethical and Practical Implications

The benchmark datasets used in this research may contain class imbalance and geographic bias, which cause unfairness in retrieval results. Moreover, large-scale image retrieval systems may raise privacy and surveillance concerns if deployed without user consent or appropriate governance. Although the proposed approach employs effectively on datasets up to 10K images, additional indexing and distributed retrieval strategies would be needed for industrial-scale repositories containing millions of images. Real-time deployment may also be constrained by GPU memory and latency requirements.

V. CONCLUSION

In the period of rapidly growing visual data, CBIR played a critical role in enabling effective access to relevant information across diverse fields. Nevertheless, existing CBIR methods often suffer from degraded precision when retrieving a large number of images, constrained feature discriminability, and enhanced computational burden. To overcome this challenge, this research proposed the adaptive Adam-CNN approach for effective image retrieval. A combination of the adaptive Adam optimizer improved convergence speed and generalization capability, while the ablation study helped fine-tune hyperparameters for optimal performance. This resulted in better feature discriminability, higher retrieval precision, and robust performance for image retrieval. The experimental results demonstrated that the proposed adaptive Adam-CNN method attained better results as compared to existing methods. In future work, attention functions and Vision Transformer (ViT) models will be used to further enhance the semantic feature learning. The framework will also be tested in dynamic, real-world environments.

NOTATION LIST

Notations	Descriptions
$\boldsymbol{A}$	attribute data
$\boldsymbol{min_A}$ and $\boldsymbol{max_A}$	minimum and maximum absolute value of $\boldsymbol{A}$, respectively
$\boldsymbol{v}$	actual value
$\boldsymbol{v'}$	normalized value
$(\boldsymbol{A_{t0}})$	threshold value of a norm
$\boldsymbol{A(t)}$	norm value
$\boldsymbol{\lvert g_t \rvert}$	absolute gradient
$\boldsymbol{m_{t-1}}$	exponential moving average of the preceding gradient
$\boldsymbol{m_{max}}$	maximum value between $\boldsymbol{\lvert g_t \rvert}$ and $\boldsymbol{m_{t-1}}$
$\boldsymbol{\theta_0}$	model parameter at iteration $\boldsymbol{t}$
$\boldsymbol{\eta}$	step size
$\boldsymbol{N_B}$	batch size
$\boldsymbol{\varepsilon}$	constant
$\boldsymbol{m_t}$	first moment (mean) estimate at time $\boldsymbol{t}$
$\boldsymbol{m_t}$	second moment (uncentered variance) estimate at time $\boldsymbol{t}$
$\boldsymbol{g_t}$	gradient of the loss function at time $\boldsymbol{t}$
$\boldsymbol{T}$	total number of training iterations
$\boldsymbol{\nabla l\left\{\begin{pmatrix} x_{ik}, y_{ik}, \\ \theta_{t-1} \end{pmatrix}\right\}}$	gradient of the loss function w.r.t. parameters $\boldsymbol{\theta}$ on input $\boldsymbol{x_i}, \boldsymbol{y_i}$

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

KMS conceptualized the research work, designed the methodology, and supervised the overall implementation; WTC contributed to system design, algorithm development, and technical validation; MS carried out data collection, experimentation, and result analysis and contributed to manuscript drafting; MV assisted in data preprocessing, model implementation, and performance evaluation; PR contributed to statistical analysis, interpretation of results, and manuscript review and editing. All authors discussed the results, contributed to refining the manuscript, and approved the final version of the paper.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the management and administration of Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College (Autonomous), Avadi, Chennai, and Panimalar Engineering College, Varadharajapuranm, Chennai, for providing the necessary infrastructure and support to carry out this research work.

REFERENCES

[1] A. Orman, "Content-based image retrieval using composite feature vectors with edge features based on color and pixel similarity," *Traitement du Signal*, vol. 41, no. 4, pp. 1945–1952, 2024.

[2] T. Rajasenbagam, S. Jeyanthi, and J. A. Pandian, "Detection of pneumonia infection in lungs from chest X-ray images using deep convolutional neural network and content-based image retrieval techniques," *J. Ambient Intell. Hum. Comput.,* pp. 1–8, 2021.

[3] D. Singh, J. Mathew, M. Agarwal, and M. Govind, "DLIRIR: Deep learning-based improved reverse image retrieval," *Eng. Appl. Artif. Intell.,* vol. 126, 106833, 2023.

[4] M. S. Sayed, A. A. Gad-Elrab, K. A. Fathy, and K. R. Raslan, "Unsupervised content-based image retrieval using pre-trained CNN and PCNN features extractors," *International Journal of Intelligent Engineering & Systems.,* vol. 16, no. 1, pp. 584–596, 2023.

[5] A. Khan and V. S. Kumar, "Optimizing image retrieval through fusion of GLCM and local binary features," *Iconic Research and Engineering Journals.,* vol. 8, no. 9, pp. 1567–1578, 2025.

[6] M. Alrahhal and K. P. Supreethi, "Integrating machine learning algorithms for robust content-based image retrieval," *Int. J. Inf. Technol.,* vol. 16, no. 8, pp. 5005–5021, 2024.

[7] Y. Tang, J. Yu, K. Gai, J. Zhuang, G. Xiong, Y. Hu, and Q. Wu, "Context-i2w: Mapping images to context-dependent words for accurate zero-shot composed image retrieval," in *Proc. of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, no. 6, pp. 5180–5188.

[8] M. Sotoodeh, A. Kohan, M. Roshanzamir, S. K. Jagatheesaperumal, M. R. C. Qazani, J. H. Joloudari, R. Alizadehsani, and P. Pławiak, "Comparison between state-of-the-art color local binary pattern-based descriptors for image retrieval," *IEEE Access,* vol. 12, pp. 162432–162449, 2024.

[9] S. Sikandar, R. Mahum, and A. Alsalman, "A novel cooperative approach for a content-based image retrieval using feature fusion," *Appl. Sci.,* vol. 13, no. 7, 4581, 2023.

[10] L. K. Pavithra, P. Subbulakshmi, N. Paramanandham, S. Vimal, N. S. Alghamdi, and G. Dhiman, "Enhanced semantic natural scenery retrieval system through novel dominant colour and multi-resolution texture feature learning model," *Expert Syst.,* vol. 42, no. 2, e13805, 2025.

[11] F. A. A. Salih and A. A. Abdulla, "Two-layer content-based image retrieval technique for improving effectiveness," *Multimedia Tools Appl.,* vol. 82, no. 20, pp. 31423–31444, 2023.

[12] F. A. Alghamdi, "An effective cooperative framework based on combination of color and texture features for content-based image retrieval," *Arabian J. Sci. Eng.,* vol. 49, no. 3, pp. 3575–3591, 2023.

[13] G. V. E. Rao and B. Rajitha, "HQF-CC: Cooperative framework for automated respiratory disease detection based on quantum feature extractor and custom classifier model using chest X-rays," *Int. J. Inf. Technol.,* vol. 16, no. 2, pp. 1145–1153, 2024.

[14] V. H. Vu, "Content-based image retrieval with fuzzy clustering for feature vector normalization," *Multimedia Tools Appl.,* vol. 83, no. 2, pp. 4309–4329, 2023.

[15] S. Kumar, Ruchilekha, M. K. Singh, and M. K. Mishra, "DFFnet: Delay feature fusion network for efficient content-based image retrieval," *Pattern Anal. Appl.,* vol. 28, no. 2, 82, 2025.

[16] B. Zhang, Y. Xu, Y. Yan, and Z. Wang, "Privacy-preserving image retrieval based on additive secret sharing in cloud environment," *Cluster Comput.,* vol. 27, no. 4, pp. 5021–5045, 2024.

[17] R. Hidayat, A. Harjoko, and A. Musdholifah, "A robust image retrieval method using multi-hierarchical agglomerative clustering and Davis–Bouldin index," *International Journal of Intelligent Engineering & Systems.,* vol. 15, no. 2, pp. 441–453, 2022.

[18] M. Alrahhal and K. P. Supreethi, "Enhancing image retrieval accuracy through multi-resolution HSV-LNP feature fusion and modified K-NN relevance feedback," *International Journal of Information Technology,* pp. 1–15, 2024.

[19] M. Singh and M. K. Singh, "Content-based medical image retrieval using deep learning and handcrafted features in dimensionality reduction framework," *Computers and Electrical Engineering,* vol. 127, 110581, 2025.

[20] K. Suneetha and N. Gupta, "Employing tensor decomposition and contextual inference for reducing the semantic discrepancy in image processing," *in Proc. 2023 International Conference on*

Communication, Security and Artificial Intelligence (ICCSAI), 2023, pp. 210–214.

[21] K. Rahbar and F. Taheri, "Enhancing image retrieval through entropy-based deep metric learning," *Multimedia Tools Appl.,* vol. 84, pp. 9065–9091, 2024.

[22] R. A. Narayanaswamy, V. K. Gowda, and G. N. K. Murthy, "Content-based image retrieval using the 2-dimensional convolutional neural network," *International Journal of Intelligent Engineering & Systems.,* vol. 17, no. 4, pp. 67–76, 2024.

[23] G. S. Vieira, A. U. Fonseca, and F. Soares, "CBIR-ANR: A content-based image retrieval with accuracy noise reduction," *Software Impacts,* vol. 15, 100486, 2023.

[24] F. Taheri, K. Rahbar, and Z. Beheshtifard, "Content-based image retrieval using handcrafted feature fusion in semantic pyramid," *International Journal of Multimedia Information Retrieval,* vol. 12, no. 2, 21, 2023.

[25] G. S. Vieira, A. U. Fonseca, N. M. Sousa, J. P. Felix, and F. Soares, "A novel content-based image retrieval system with feature descriptor integration and accuracy noise reduction," *Expert Syst. Appl.,* vol. 232, 120774, 2023.

[26] F. Taheri, K. Rahbar, and P. Salimi, "Effective features in content-based image retrieval from a combination of low-level features and deep Boltzmann machine," *Multimedia Tools Appl.,* vol. 82, no. 24, pp. 37959–37982, 2022.

[27] E. Ranjith, L. Parthiban, T. P. Latchoumi, S. A. Kumar, D. G. Perera, and S. Ramaswamy, "An effective content-based image retrieval system using deep learning-based inception model," *Wireless Pers. Commun.,* vol. 133, no. 2, pp. 811–829, 2024.

[28] M. K. Kelishadrokhi, M. Ghattaei, and S. Fekri-Ershad, "Innovative local texture descriptor in joint of human-based color features for content-based image retrieval," *Signal, Image Video Process.,* vol. 17, no. 8, pp. 4009–4017, 2023.

[29] G. H. Liu and J. Y. Yang, "Exploiting deep textures for image retrieval," *Int. J. Mach. Learn. Cybern.,* vol. 14, no. 2, pp. 483–494, 2022.

[30] X. Hu, Zhou, L. Lyu, and C. Lan, "Fast dual-feature extraction based on tightly coupled lightweight network for visual place recognition," *IEEE Access,* vol. 11, pp. 127855–127865, 2023.

[31] S. Tyagi, P. Shukla, and P. Singh, "DFF-CBIR: Deep feature fusion framework for enhanced content-based image retrieval," *IEEE Access*, vol. 14, pp. 22678–22698, 2026.

[32] A. M. Humadi, M. Sadeghzadeh, H. A. Younis, and M. Mosleh, "Adaptive two-tier deep learning for content-based image retrieval and classification with dynamic similarity fusion," *IET Image Proc.,* vol. 19, no. 1, e70192, 2025.

[33] Akil Elkamel. (2020). Corel1k dataset. [Online]. Available: https://www.kaggle.com/datasets/elkamel/corel-images

[34] Ali Salar. (2022). Corel5k dataset. [Online]. Available: https://www.kaggle.com/datasets/parhamsalar/corel5k

[35] Michel Wilson. (2022). Corel10k dataset. [Online]. Available: http://kaggle.com/datasets/michelwilson/corel10k

[36] SkyLord. (2020). Oxford 5K and Paris 6K dataset. [Online]. Available: https://www.kaggle.com/datasets/skylord/oxbuildings

[37] Bikram Saha. (2022). Caltech101 dataset. [Online]. Available: https://www.kaggle.com/datasets/imbikramsaha/caltech-101

[38] V. Yuvaraj and D. Maheswari, "Lung cancer classification based on enhanced deep learning using gene expression data," *Meas.: Sens.,* vol. 30, 100902, 2023.