J. Electrical Systems 20-2s (2024): 1325-1333

[1]Latika Pinjarkar

[2]Jaspal Bagga

[3]Poorva Agrawal

[4]Gagandeep Kaur

[5]Vedant Pinjarkar

[6]Rutuja Rajendra

Deep Learning Model for Retrieving Color Logo Images in Content Based Image Retrieval

Abstract: - Content-Based Image retrieval (CBIR) has gained a magnificent deal of consideration because of the digital image data's epidemic growth. The advancement of deep learning has enabled Convolutional Neural Networks to become an influential technique for extraction of discriminative image features. In recent years, convolutional neural networks (CNNs) have proven extremely effective at extracting unique information from images. In contrast to text-based image retrieval, CBIR gathers comparable images based primarily on their visual content. The use of deep learning, especially CNNs, for feature extraction and image processing has been shown to perform better than other techniques. In the proposed study, we investigate CNNs for CBIR focusing on how well they extract discriminative visual features and facilitate accurate image retrieval. Also Principal Component Analysis and Linear Discriminant Analysis are combined for optimization of features resulting in boosting the retrieval results. Using hierarchical representations learned by CNNs, we aim to improve retrieval accuracy and efficiency. In comparison with conventional retrieval techniques, our proposed CBIR system shows superior performance on a benchmark dataset.

Keywords: CBIR, Deep Learning, CNNs, Feature Extraction, Image Retrieval

I. INTRODUCTION

Effective methods for image retrieval are required due to the rapid proliferation of digital images throughout a variety of industries, including networking sites, e-commerce, and healthcare. This problem is addressed by Content-Based Image Retrieval (CBIR) methods, which allow users to retrieve and search for images based on their visual features. Traditional CBIR methods relied on handcrafted features, like, texture descriptors color histograms (e.g., Local Binary Patterns, Gabor filters,), and shape descriptors (e.g., Scale-Invariant Feature Transform –SIFT. Color-based methods capture the distribution of color information in images, enabling retrieval based on color similarity. Texture-based methods focus on capturing local patterns and variations in image textures, allowing retrieval based on similar textural properties. Shape-based methods extract and compare shape descriptors to retrieve images with similar shapes and objects.

With the advent of deep learning, CNNs have revolutionized CBIR systems by automatically learning discriminative features from images. Transfer learning is commonly employed in CBIR using pre-trained CNN models such as ResNet, VGGNet, or inception which are trained on large-scale datasets like Imagenet [1]. These models serve as powerful feature extractors. The extracted features from CNN can be used in various ways, such as encoding them into fixed-length feature vectors using techniques like Vector of locally Aggregated Descriptors (VLAD) or Fisher Vectors. Salient points, object and shape features, signs, accumulative and global features, and structural combinations of these are considered features for retrieval sorting. The degree of similarity between images and objects in images is examined for every feature category, closely related to the kinds and ways in which an

[1,3,4]Symbiosis Institute of Technology Nagpur Campus, Symbiosis International (Deemed University), Pune, India

[2]Shri Shankaracharya Technical campus Bhilai India

[5]Shri Ramdeobaba College of Engineering and Management Nagpur, India

[6]Vishwakarma Institute of Information Technology, Computer Science Engineering-Artificial Intelligence and machine Learning Department,Pune, India

1latika.pinjarkar@sitnagpur.siu.edu.in, 2baggajaspal@gmail.com, 3poorva.agrawal@sitnagpur.siu.edu.in, 4gagandeep.kaur@sitnagpur.siu.edu.in, 5pinjarkarvs@rknec.edu, 6rutujapat@gmail.com

interaction with the system can provide feedback. databases, system architecture, and assessment—the three facets of system engineering are covered [2].

Another approach is to use the activations of intermediate layers in the CNN as feature representations, capturing both local and global visual information. Siamese networks and triplet networks have been proposed for CBIR where the models are trained to learn similarity metrics directly from image pairs or triplets [3]. Some recent studies explore hybrid features that combine traditional handcrafted features with deep learning-based features. This allows leveraging the strength of both approaches to improve retrieval performance [4].

Recent advancements in deep learning have revolutionized the field. Deep learning models like Convolutional Neural Networks (CNNs), have shown remarkable success in various computer vision tasks, including image classification, object detection, and segmentation. CNNs capture both low-level and high-level visual information by learning hierarchical models for images, which can be used for image retrieval.

To measure the performance of CBIR systems evaluation Metrics commonly used include Mean Average Precision (MAP), Precision, Recall, and Normalized Discounted Cumulative Gain (NDCG). Benchmark Datasets, such as ImageCLEF, PascalVOC, or COCO are widely used to evaluate CBIR methods providing diverse and challenging image collections for retrieval tasks. The evaluation of CBIR systems relies on established metrics and benchmark datasets, enabling fair comparisons and performing assessments.[5].

The remaining sections of the paper are structured as follows. thorough analysis of related work in CBIR utilizing deep learning models is provided in Section 2. The proposed methodology is described in Section 3, along with the deep learning model's design and training procedure. The experimental setup and findings are presented in Section 4, and a discussion is included in Section 5. The work is concluded in Section 6, which also provides a roadmap for future research.

II. LITERATURE REVIEW

This section reviews the existing literature on CBIR and application of Deep Learning models for image retrieval. Various methods and Techniques including traditional feature extraction methods and CNN approaches are discussed in terms of their strengths, limitations, and performance.

The influential AlexNet was introduced which achieved breakthrough results in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). By demonstrating the potential of deep learning for image categorization issues, this work paved the way for later developments in CBIR [6]

Early on in the development of CBIR techniques, numerous methods were discussed for retrieving images based on their visual content. Improvements to the bag-of-features model, a prevalent CBIR method, were suggested by the authors to increase the performance of large-scale picture searches [7]. They proposed the Vector of Locally Aggregated Descriptors (VLAD). By capturing both local and global image information, the image categorization method was introduced to increase the accuracy of image retrieval [8]. Local characteristics, spatial verification, and query expansion were identified as three crucial elements for enhancing object retrieval in CBIR systems [9]. The merits and disadvantages of various deep learning models and strategies for image retrieval were discussed in a thorough study of deep learning-based CBIR [10].

Top-down neural attention through excitation backprop was introduced to identify the most discriminative regions in an image for improved image retrieval performance [11]. Integral max-pooling of CNN activations was presented for retrieving specific objects, improving retrieval performance by taking into account the spatial data captured by CNN features [12]. The Vector of Locally Aggregated Descriptors (VLAD) encoding technique was introduced [13]. It enhanced image retrieval accuracy by capturing the connections between local features.

To facilitate scalable CBIR applications, the authors presented a deep learning-based hashing technique for image retrieval. This method produced compact binary codes for effective image retrieval [14]. The work concentrated on using CNNs for image retrieval and demonstrated that deep features outperform conventional handcrafted features in terms of retrieval performance. The authors used pre-trained CNN models, like VGGNet, as the basis for their feature representations, which were based on the activations of intermediate layers. This method captured both global and local visual information, leading to more accurate retrievals [15]. Expanding on this research, it was focused on CNN model optimization specifically for image retrieval, removing the requirement for human annotation. Through the application of metric learning techniques, the researchers created an architecture that allows the network to automatically extract similarity metrics from image pairs. As a result, the learned features were enhanced in terms of discriminative power, further enhancing retrieval accuracy [16]. In the study global representations of images were proposed to improve picture retrieval. They used Siamese networks to extract similarity metrics from triples of images, enabling efficient retrieval even in massive image repositories. By combining cutting-edge CBIR system performance with customary handcrafted elements, they attained the highest level of performance [17].

This hybrid approach created a more comprehensive representation of images, capturing both high-level semantics and fine-grained visual details. This paper focuses on contributing to the existing body of research on CBIR using deep learning models. We suggest a method that utilizes the use of CNN's power to extract distinct features for image retrieval. The experimental evaluation compares the performance of our proposed approach with state-of-the-art methods on benchmark datasets.

The authors retrieved shape-invariant and shape-specific features using contour-based shape feature extraction and image moment extraction techniques. The integration of color, grayscale, texture, and shape characteristics with the features extracted is achieved through the utilization of Particle Swarm Optimization (PSO) to generate informative features. Subsequently, the query image has been furnished, and the random forest classifier has been educated to discern the intended image. Within this process, the Particle Swarm Optimization technique is employed to arrange grey, color, and texture attributes, to identify the most informative ones [18].

The work employed a Convolutional Neural Network (CNN) that had been trained on a substantial dataset containing 2D images and 3D facial models or scans. This approach effectively overcomes several limitations. The model does not necessitate meticulous alignment or extensive correspondence between images to reconstruct the complete 3D facial structure, including the facial areas that are not visible. A sole 2D facial image suffices, and it demonstrates flexibility in handling various facial poses and expressions [19].

An extensive examination of recent advancements in the domains of CBIR and picture representation was conducted. The key characteristics of various models for image representation and retrieval—from straightforward low-level feature extraction to sophisticated semantic deep-learning approaches—are examined. In-depth assessments of studies based on CBIR and picture representation are provided to promote more studies in this area [20].

III. PREAMBLES

Recent research has demonstrated that CBIR algorithms based on CNN outperform low-level feature extraction methods and yield improved outcomes. Every CNN layer is composed of a large number of 3-dimensional height, breadth, and depth neural connections.

I) DarkNet-19

The Darknet-19 prototype [21], which comprises of 19 conv2D layers, 5 max-pooling layers and a soft-max layer, uses YOLOv2 to categorize objects. The Darknet-19, a C-based CUDA-based neural network framework, is employed. Fast object recognition is essential for real-time prediction applications. To categorize images, this network employs one thousand diverse class images. The prototype has acquired knowledge of particular image attributes through image set classification. The largest image size that the network can process is 256x256.

II) Darknet-53

The objective of Darknet-53 model is features extraction of an input image. The YOLOv2 basic feature extraction and residual module can be used to represent the core idea of this network. There are five repeating blocks in this model, each covered by convolution layers(two) (size as one and thirty-three) and a residual layer. There are fifty-three convolutional layers in all, and after each one there is ReLU activation layer and a batch normalization layer.

Darknet-53 serves as the key feature extraction method for YOLOv3 (You Only Look Once), the real-time object detection network. The core concept of this network combines the residual module with YOLOv2's fundamental feature extraction.

The convolution layer is used to convolve a number of image filters, resulting in a range of feature maps. Stride 2 convolutional layers down sample the feature maps without pooling. Pooling is typically associated with a loss of low-level properties, which is lessened by this.

Selecting relevant images from a set of logo images is the main task of the proposed work.

III) Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)

When class labels for logos are available, combining PCA with LDA might be helpful for retrieving images of logos. PCA minimizes dimensionality while LDA enhances class separation.

A powerful method for improving logo image retrieval is to combine Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). This combo functions as follows:

1. PCA for Dimension Reduction: The first application of PCA is to decrease the number of dimensions in the data.

The first step involves transforming the original feature space into a new space that is optimized to maximize the variability within the dataset. This process aims to decrease the dimensionality of the features while preserving the highest degree of variability present in the data. As PCA is categorized as an unsupervised technique, it does not take into account class labels. Nevertheless, it is capable of filtering out irrelevant information and redundant patterns from the dataset, thereby enhancing the efficacy of subsequent supervised techniques.

2. Utilizing LDA for Enhancing Class Separation: LDA serves as a method for supervised reduction of dimensionality. The primary objective of this approach is to optimize the capacity of the data to differentiate between different classes. The incorporation of class labels aids LDA in discovering a fresh feature space in which the classes are distinctly segregated, thereby rendering it more suitable for tasks related to classification and retrieval. This can ensure that the representation of features effectively discerns between various logos within the domain of logo image retrieval.

There are various benefits to using PCA and LDA together for logo image retrieval.

• Dimension Reduction: PCA greatly reduces the dimensionality of the dataset, which makes it easier to manage and more effective to process.

• Class Separation: LDA maximizes the separation between logo classes, which is important for retrieval tasks because it is a supervised technique.

• Interpretability: The resulting feature space can reveal the most discriminative features for logo retrieval and is frequently easier to understand.

• Noise Reduction: When working with intricate and noisy logo images, PCA can be especially helpful in reducing noise in the feature space.

The following is a detailed procedure for merging PCA and LDA in logo image retrieval:

1. To reduce the dimensionality of the logo image dataset while keeping a subset of the most important principal components, PCA is applied.

2. Taking into consideration the class labels of the logos when using the reduced PCA feature space is given as input to LDA.

3. A feature space with maximum class separation is what LDA produces as its output, which makes it perfect for retrieving logo images.

4. The LDA-transformed features are used to represent the logo images, and in this transformed space, the logo image retrieval algorithm is implemented.

The logo image retrieval performance is improved by reducing the dimensionality of logo images while improving class separation by combining PCA and LDA.

Certainly, using PCA and LDA together is a very effective method for retrieving logo images. This is a more thorough mathematical explanation of how to use the combination of PCA and LDA for this purpose:

PCA(Principal Component Analysis):

Principal Component Analysis, or PCA for short, is a dimensionality reduction technique that puts the data into a new coordinate system where the principal components maximize the variance of the data. The steps involved are as follows:

1. A matrix X containing a dataset of logo images is given, with each row denoting an image and each column a feature.

2. Determine the data's mean, or μ.

3. To centre the data, subtract the mean from each data point (X_centered = X - μ).

4. With n being the number of samples, compute the covariance matrix C = (1/n) * (X_centered^T) * X_centered.

5. Determine the covariance matrix's eigenvalues and eigenvectors.

6. To find the principal components, sort the eigenvectors in descending order of their matching eigenvalues.

7. To reduce the dimensionality, choose the top k principal components; the value of k is usually determined by how much variance you wish to keep (e.g., 95% of the variance)

LDA (Linear Discriminant Analysis):

The goal of LDA, or linear discriminant analysis, is to minimize variance within each class and maximize separation between them. It is a supervised dimensionality reduction technique. The steps involved are as follows:

1. Given a dataset of logo images, each with a class label (e.g., various brands of logos).

2. Determine the scatter matrices between and within classes (Sb and Sw, respectively).

3. The covariance matrices of each class are added up to get Sw, and the separation between the class means is measured to get Sb.

4. Determine the matrix Sw^(-1) * Sb's eigenvalues and eigenvectors.

5. Arrange the eigenvectors in descending order of their matching eigenvalues.

6. To create the new feature space, choose the top k eigenvectors, where k is the

Combining PCA and LDA: The following procedures can be used to combine PCA and LDA for logo image retrieval:

1. Use PCA to lower the dimensionality of your dataset of logo images. The feature space is changed as a result.

2. Enter the LDA with the condensed feature space that was acquired through PCA.

3. Taking the logos' class labels into consideration, apply LDA to the PCA-transformed feature space. LDA maximizes class separation while further reducing dimensionality.

4. The result of LDA is a feature space with optimized class separability, which is ideal for retrieving logo images.

5. You can perform retrieval tasks like classification and similarity search using this combined feature space.

When PCA and LDA are combined, dimensionality is reduced.

IV. PROPOSED FRAMEWORK FOR CBIR SPEECH ACQUISITION:

The suggested framework builds a CBIR model for colored logo images via feature blend. The conventional method extracts the texture, color, or other significant elements of the image and then manually picks a large number of components for fusion. The problem with this approach is that the same characteristic has varied effects on image retrieval depending on the dataset. Finding the features that are best suited for the specific datasets and the ideal feature combination prior to image retrieval procedures may be difficult. To enhance the results, different combinations over the full datasets might be tested, however the procedure is incredibly time-consuming, ineffective, and unreliable.

The pre-trained DarkNet-53 and DarkNet-19 models are used in this investigation to extract image attributes. A vector containing new features is produced by combining the retrieved features. DarkNet-53's last convolutional layer, which is connected to each input neuron and maintains spatial information, is used by us to extract features. The feature extraction layer in DarkNet-19 uses the average pooling layer, often known as "Avg 1," to create a vector with additional features.

V. EXPERIMENTATION FRAMEWORK

By categorising the images according to similarity, a Logo image retrieval framework must be able identify images that correspond with a certain query. The logo examiners might assess if it includes images that are sufficiently similar to the query by choosing a small portion at the top of the list. Determining a logo image retrieval system's capacity to successfully and effectively accomplish this objective is the simple evaluation criterion.

5.1 Selection of Database

One of the most crucial questions to answer while developing any image retrieval system is how to get reliable datasets for conducting experiments, as the dataset is the foundation for all findings. To provide reliable results, a fair experimental dataset must be obtained. This information can be accessed in two different ways. The 1st method is gathering data from a small group of experts such as Logo examiners in this field, while the 2nd entails gathering info from a larger variety of human subjects. The experiments in the proposed framework were conducted using the FlickrLogos-47 collection of logos, as shown in Table 1.1 below.

Table 1. Flickr Logos datasets details

Dataset	**Total No. Images**	**Brand**	**Class**
FlickrLogos-47 Dataset	9200	47	47

5.2 FlickrLogos-47 Dataset:

The FlickrLogos-47 dataset contains images of logos from well-known companies to test logo recognition and detection methods on images. It consists of FlickrLogos-32 dataset iamges and that had their mission annotations corrected and new classes added. For logos with names like Apple, Adidas, Aldi, Becks, BMW, Carlsberg, Coca-Cola, DHL, Erdinger, Esso, Fedex, Ferrari, and Ford, among others, there are 47 different classes.

Some descriptive trademark images from the FlickrLogos-47 Dataset that were utilized in the structure are shown in Figure 1.

Figure 1: FlickrLogos-47 Dataset sample Trademark images

The following algorithmic form illustrates the experimental steps:

Algorithm 1.1: A method for retrieving images

Input: The query Image is I. DB identifies an image database. NO signifies the overall images number.

Output: S, related images retrieved against I.

i) NO: total image count in DB

ii) For I in (1 to NO)

query image I∈ Q^(N X RX 3)

Image resizing to 224 X 224 pixels.

Extract high-level characteristics from the two models of NO images that have been activated in the database. Use the created features for each image to create a database index, or DB Index..

iii) Feature Extraction for the given query I

iv) Reduction of Features by means of PCA and LDA

v) For each image from the DB Index

Comparison of the query image's feature with all the DB Index images through the similarity metric.

vi) Arrangement of the suitable images conferring to Q features

vii) Top S images retrieval against the given query I

VI. EXPERIMENTAL RESULTS IN TERMS OF RETRIEVAL

The evaluation criteria for precision and recall are used to assess the experimentation. For example, Table 2 shows the average precision of the retrieval results for 10 query images using the DarkNet-19 and DarkNet-53 models and the suggested framework. Table 3 presents the average recall of the retrieval results.

Table 2. Retrieval results for sample ten query images (average precision)

Query Image	DarkNet-19	DarkNet-53	Proposed Framework
	80.93	82.4	93.9
	80.78	88.9	95.5
	92.2	93.9	92.7
	91.5	91.2	91.5
	91.8	92.7	90.7
	91.4	97.8	93.4
	92.9	89.9	91.8
	92.9	91.7	87.9
	92.8	95.8	95.7
	84.8	95.2	96.1
Average	89.20	91.95	92.92

Table 3. Retrieval results for sample ten query images (average recall)

Query Image	DarkNet-19	DarkNet-53	Proposed Framework
	16.34	17.12	18.53
	16.78	15.83	18.54
	18. 21	19.79	19.3
	18	18.9	19.8
	17.54	18.7	18.9
	18.84	19.81	18.8

PUMA	17.21	18.76	18.8
NBC	18.12	17.91	18.08
vodafone	18.11	18.56	18.9
BMW	18.9	19.24	18.43
Average	17.76	18.46	18.81

When compared to the current deep learning-based models, the proposed framework which uses the fusion of features method through the DarkNet-19 and DarkNet-53 models performed better in terms of average precision, as demonstrated in Table 4.

Table 4. Comparative analysis with other models based upon deep learning

Results/ Existing Model (Deep Learning)	*VGG16[10]*	*AlexNet[17]*	*ResNet[22]*	*GoogleNet[23]*	*Modified AlexNet [24]*	*Proposed Framework*
Average Precision	89.08	76.22	93.27	95.61	92.37	92.92
Average Recall	58.7	41.58	71.99	79.43	70.92	18.81

VII. CONCLUSION

In order to improve CNN feature illustration and ensure accurate image retrieval for Logo pictures in the CBIR system, this framework aims to design a structure using multiple feature fusion mechanisms. Because it is so closely linked to how humans perceive things, feature extraction and representation in CBIR is a significant problem. The dataset of Logo images is searched for relevant images using the Euclidean distance. The darknet-19 and darknet-53 feature vectors are combined in the proposed method to yield a novel feature vector for extracting logo images. The combination of Principal Component Analysis and Linear Discriminant Analysis for optimization of features have proved the improved results. Compared to other models, the FlickrLogos-47 dataset performs better based on average precision and average recall for retrieval results.

REFERENCES

[1] Adhikari, Tapomoy, Designing a Convolutional Neural Network for Image Recognition: A Comparative Study of Different Architectures and Training Techniques (February 22, 2023). Available at SSRN: https://ssrn.com/abstract=4366645 or http://dx.doi.org/10.2139/ssrn.4366645

[2] W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content-based image retrieval at the end of the early years," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, Dec. 2000, doi: 10.1109/34.895972.

[3] Hajamohideen, F., Shaffi, N., Mahmud, M. *et al.* Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function. *Brain Inf.* **10**, 5 (2023). https://doi.org/10.1186/s40708-023-00184-w

[4] Simran Arora, Patrick Lewis, Angela Fan, Jacob Kahn, and Christopher Ré. 2023. Reasoning over Public and Private Data in Retrieval-Based Systems. *Transactions of the Association for Computational Linguistics*, 11:902–921.

[5] Sreerajeswari, M., and M. Rajalakshmi. "FEATURE EXTRACTION METHODS AND RE-RANKING WITH CLICK-BASED SIMILARITY FOR WEB IMAGES." *Dynamic Systems and Applications*, vol. 30, no. 6, June 2021, pp. 944

[6] Krizhevsky A, Sutskever I and Hinton G E 2012 *Advances in neural information processing systems* **25** 1097–1105

[7] H. Jegou et al., “Aggregatinglocal imagedescriptors intocompact codes,”IEEE Trans.PatternAnal.Mach.Intell.,vol.34,pp.1704–1716,Sept.2012.

[8] F. Perronnin and C. Dance, "Fisher Kernels on Visual Vocabularies for Image Categorization," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 2007, pp. 1-8, doi: 10.1109/CVPR.2007.383266. R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012, pp. 2911-2918, doi: 10.1109/CVPR.2012.6248018.

[9] Chen, X., Yu, G., Cheng, G., & Hao, T. (2019). Research topics, author profiles, and collaboration networks in the top-ranked journal on educational technology over the past 40 years: a bibliometric analysis. Journal of Computers in Education, 6(4), 563-585. https://doi.org/10.1007/s40692-019-00149-1

[10] Zhang, Z., F. Werner, H.-M. Cho, G. Wind, S.E. Platnick, A.S. Ackerman, L. Di Girolamo, A. Marshak, and K.G. Meyer, 2016: A framework based on 2-D Taylor expansion for quantifying the impacts of sub-pixel reflectance variance and covariance on cloud optical thickness and effective radius retrievals based on the bispectral method. *J. Geophys. Res. Atmos.*, **121**, no. 12, 7007-7025, doi:10.1002/2016JD024837.

[11] Tolias G, Sicre R and Jégou H 2016 *arXiv preprint arXiv:1511.05879*

[12] Arandjelovic, R., & Zisserman, A. (2013). All about VLAD. In Proceedings of theIEEE conference on Computer Vision and Pattern Recognition (pp. 1578-1585).

[13] C. Zhou *et al*., "Angular Deep Supervised Hashing for Image Retrieval," in *IEEE Access*, vol. 7, pp. 127521-127532, 2019, doi: 10.1109/ACCESS.2019.2939650.

[14] Babenko, Artem & Slesarev, Anton & Chigorin, Alexandr & Lempitsky, Victor. (2014). Neural Codes for Image Retrieval. LNCS. 8689. 10.1007/978-3-319-10590-1_38.

[15] Radenović, F., Tolias, G., & Chum, O. (2017). Fine-tuning CNN Image Retrieval with No Human Annotation. *ArXiv*. /abs/1711.02512

[16] Gordo, A., Almazan, J., Revaud, J., & Larlus, D. (2016). End-to-end Learning of Deep Visual Representations for Image Retrieval. *ArXiv*. /abs/1610.07940

[17] Subramanian M, Lingamuthu V, Venkatesan C, Perumal S. Content-Based Image Retrieval Using Colour, Gray, Advanced Texture, Shape Features, and Random Forest Classifier with Optimized Particle Swarm Optimization. Int J Biomed Imaging. 2022 Apr 21;2022:3211793. doi: 10.1155/2022/3211793. PMID: 35496641; PMCID: PMC9050319.

[18] Jackson, Aaron & Valstar, Michel & Tzimiropoulos, Georgios. (2016). A CNN cascade for landmark guided semantic part segmentation.

[19] Latif A., Rasheed A., Sajid U., et al. Content-based image retrieval and feature extraction: a comprehensive review. *Hindawi-Mathematical Problems in Engineering* . 2019;**2019**:1–21. doi: 10.1155/2019/9658350

[20] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 21-26 July 2017, 7263-7271. https://doi.org/10.1109/CVPR.2017.690

[21] Maji Sudhadip and Bose Smarajit. 2021. “CBIR using features derived by deep learning “. *ACM/IMS Transactions on Data Science (TDS)*, 2, 3 (Aug. 2021), 1–24. DOI: 021), 24 pages. https://doi.org/10.1145/3470568

[22] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[23] Ashkan Shakarami and Hadis Tarrah “An efficient image descriptor for image classification and CBIR” Optik, Volume 214,2020,pages: 164833, ISSN 0030-4026, https://doi.org/10.1016/j.ijleo.2020.164833.