Implementation on Privacy-Preserving Content-Based Image Retrieval in Cloud Image Repositories

Mr. Atish Gopichand Jadhav, Mr. Namdev Sawant, Mr. Subhash Pingale.
Department of Computer Science and Engineering, SKN Sinhgad College of Engineering Pandharpur, Punyashlok Ahilyadevi Holkar, Solapur University, Solapur.

ABSTRACT Without knowing the name of the picture, searching through a collection of images that resemble the input images using a pursuing framework that uses the CBIR concept is essential. Overall, CBIR systems compare visual elements including colour, picture edge, surface, and the consistency of names between input images and images in the database. CNN is the characterisation method, while cosine comparability is used for recovery. This essay addresses the problem of large-scale image recovery, focusing on enhancing its accuracy and robustness. We focus on elements that might affect search vigour, such as different levels of illumination, object size and shape, fractional obstacles, and disordered foundations. These characteristics are particularly important when a hunt is conducted across extraordinarily huge datasets with high changeability. We suggest a brand-new CNN-based global descriptor termed REMAP, which is prepared from beginning to end with a triplet misfortune and learns and totals a progressive system of deep highlights from various CNN layers. REMAP categorically acquires discriminative cues that are typically constant and correlated at various semantic levels of visual reflection.

Keywords: CNN, REMAP, CBIR, Cloud Dataset

I. INTRODUCTION

CBIR is a system that extracts images from a photo database using visual information. This framework is now essential since it can successfully address the earlier-composed problems. A few techniques, including the histogram and division, are used in CBIR to remove visual content. Likewise, multidimensional element vectors depict. The element vectors and similitude metrics primarily influence how the substance-based picture recovery (CBIR) architecture recovers images. There is consistently a semantic difference between the high level semantics that humans perceive and the low level semantics that machines capture in image pixels. We are motivated to address this issue in order to enhance the presentation of CBIR as a result of the continuous successes of deep learning techniques, notably CNN.

CBIR is a system that extracts images from a photo database using visual information. This framework is now necessary since it can successfully address the problems previously outlined. Histogram and division are two techniques used in CBIR to reduce visual material. Likewise, multidimensional element vectors are used to represent. The component vectors and similitude measurements primarily affect how the substance-based picture recovery (CBIR) architecture recovers images. Between the important level semantics that people perceive and the low-level image pixels that computers capture, there is constantly a semantic difference. We decided to address this issue in order to enhance the presentation of CBIR because of the continuous successes of deep learning techniques, notably Convolutional Neural Networks (CNN), in handling the problem of PC vision applications CBIR.

One of the most well-known subfields of research in the area of example inquiry and machine insight is visual pursuit. With the emotional advancement of the audiovisual sector, the need for a potent and computationally effective visual online search engine has become increasingly important. The goal is to retrieve unique images that depict events of a client-determined article, scenario, or location from a vast corpus of images. The management of media material, portable commerce, observation, therapeutic imaging, augmented reality, apply autonomy, association of individual photos, and other significant applications are among the applications. Due to factors including changing article look, viewpoints, and scale, midway obstructions, moving foundations, and imaging circumstances, it can be difficult to provide powerful and precise visuals.

Due of the enormous amounts of media content available, the current frameworks also need to be adaptable to billions of photographs. A focused and discriminating visual depiction is needed to overcome these challenges. Convolutional Neural Networks (CNNs) provided effective solutions for various PC vision tasks, such as picture grouping. However, they currently don't seem to be able to address the picture recovery problem with anticipated performance enhancements, especially for really large scopes. The main reason is that, despite everything, there are still two major questions that are largely unanswered: (1) how to best combine profound highlights that have been split up by a CNN arrangement into conservative and discriminative picture level portrayals, and (2) how to set up the resulting CNN aggregator engineering for picture recovery tasks. This study addresses the problem of very large-scale picture recovery, focusing on enhancing its robustness and accuracy.

II. LITERATURE REVIEW

By putting forth a revolutionary location-based overall approach that uses multi-layered profound features and builds up the associated design that is trainable in a start to finish fashion, Husain et al. [1] overcome the challenges already discussed. The name of our descriptor, Region-Entropy1 based Multi-layer Abstraction Pooling, mirrors the fundamental innovations.

A widely used method for recovering images from large, unlabeled photo libraries is content-based image recovery (CBIR). However, customers are not satisfied with the usual data recovery techniques. Additionally, the quantity of images that are available to customers as well as the growth of online development and transmission systems continue to grow. In this way, a consistent and extensive production of improved pictures takes place throughout several nations. Fast access to these enormous collections of images and the recovery of a comparable image of a particular image (Query) from them therefore provide substantial challenges and need for efficient approaches. A substance-based picture recovery framework's presentation urgently depends on similitude estimate and component representation. Thus, for speedy picture recovery based on highlight extraction and grouping, Ouhda Mohamed et al. [2] presented a straightforward yet effective profound learning structure relying on Convolutional Neural Networks (CNN) and Support Vector Machine (SVM).

Without knowing the name of the picture, searching through a collection of images that resemble the input images necessitates the use of a hunt framework that implements the concept of content-based image recovery (CBIR). When everything is said and done, CBIR frameworks match input images with images in the database using visual elements including colour, picture edge, surface, and reasonableness of names. Convolutional neural networks (CNN) are used for grouping, while cosine likeness is used for recovery. Each of the five masterclasses in the data set has five subclasses. The class that is used for recovery is a masterclass, where the images of each enormous class are composed images of the enormous class's subclasses. Rian, Z., from the investigations According to et al. [3], the CNN technique has been successful in assisting the recovery effort by sorting picture classes.

Interactive media content investigation is applied in various true PC vision applications, and advanced pictures establish a significant piece of sight and sound information. In most recent couple of years, the multifaceted nature of mixed media substance, particularly the pictures, has developed exponentially, and on consistent schedule, more than a huge number of pictures are transferred at various chronicles, for example, Twitter, Facebook, and Instagram. To scan for a pertinent picture from a chronicle is a difficult research issue for PC vision examine network. The vast majority of the web indexes recover pictures based on customary content put together methodologies that depend with respect to subtitles and metadata. Over the most recent two decades, broad research is accounted for content-based picture recovery (CBIR), picture grouping, and investigation. In CBIR and picture characterization based models, significant level picture visuals are spoken to as highlight vectors that comprises of numerical qualities. The exploration shows that there is a huge hole between picture highlight portrayal and human visual comprehension. Because of this explanation, the examination introduced right now engaged to diminish the semantic hole between the picture include portrayal and human visual comprehension. Right now, expect to introduce a thorough survey of the ongoing advancement in the zone of CBIR and picture portrayal. Latif, A. et al [4] investigated the fundamental parts of different picture recovery and picture portrayal models from low-level component extraction to late semantic profound learning draws near.

As of late, picture portrayal based upon Convolutional Neural Network (CNN) has been appeared to give powerful descriptors to picture search, outflanking pre-CNN includes as short-vector portrayals. However such models are not perfect with geometry-mindful re-positioning techniques and still outflanked, on some specific article recovery benchmarks, by customary picture search frameworks depending on exact descriptor coordinating, geometric re-positioning, or inquiry development. Paper [5] returns to both recovery stages, to be specific starting hunt and re-positioning, by utilizing a similar crude data got from the CNN. G. Tolias et al fabricate reduced element vectors that encode a few picture districts without the need to take care of various contributions to the system. Moreover, they stretch out vital pictures to deal with max-pooling on convolutional layer actuations, permitting us to proficiently limit coordinating items. The subsequent bouncing box is at long last utilized for picture reranking.

R. Arandjelovi'c et al [6] handled the issue of huge scope visual spot acknowledgment, where the assignment is to rapidly and precisely perceive the area of a given inquiry photo. Creators present the accompanying three head commitments. In the first place, we build up a convolutional neural system (CNN) design that is trainable in a start to finish way straightforwardly for the spot acknowledgment task. The principle segment of this design, NetVLAD, is another summed up VLAD layer, enlivened by the "Vector of Locally Aggregated Descriptors" picture portrayal ordinarily utilized in picture recovery. The layer is promptly pluggable into any CNN engineering and agreeable to preparing by means of back engendering. Second, we build up a preparation strategy, in light of another pitifully regulated positioning misfortune, to learn parameters of the engineering in a start to finish way from pictures delineating similar places after some time downloaded from Google Street View Time Machine. At long last, we show that the proposed engineering essentially beats non-learnt picture portrayals and off-the-rack CNN descriptors on two testing place acknowledgment benchmarks, and improves over current best in class conservative picture portrayals on standard picture recovery benchmarks.

In [7], A. Gordo et al contend that purposes behind the disappointing aftereffects of profound strategies on picture recovery are triple: I) boisterous preparing information, ii) improper profound design, and iii) imperfect preparing system. We address every one of the three issues. To start with, we influence a huge scope however loud milestone dataset and build up a programmed cleaning technique that creates an appropriate preparing set for profound recovery. Second, we expand on the ongoing RMAC descriptor; show that it tends to be deciphered as a profound and differentiable design, and present upgrades to improve it. Last, we train this system with a siamese engineering that consolidates three streams with a triplet misfortune. Toward the finish of the preparation procedure, the proposed design creates a worldwide picture portrayal in a solitary forward pass that is appropriate for picture recovery. Broad analyses show that our methodology essentially outflanks past recovery draws near, including best in class techniques dependent on exorbitant neighborhood descriptor ordering and spatial check. On Oxford 5k, Paris 6k and Holidays, we individually report 94.7, 96.6, and 94.8 mean normal accuracy.

More profound neural systems are progressively hard to prepare. K. He et al [8] present a leftover learning system to facilitate the preparation of systems that are generously more profound than those utilized already. They unequivocally reformulate the layers as learning lingering capacities regarding the layer contributions, rather than learning unreferenced capacities. This paper gives far reaching exact proof demonstrating that these remaining systems are simpler to improve, and can pick up precision from impressively expanded profundity. On the ImageNet dataset we assess leftover nets with a profundity of up to 152 layers—8× more profound than VGG nets yet at the same time having lower multifaceted nature. A troupe of these leftover nets accomplishes 3.57% mistake on the ImageNet test set. This outcome won the first spot on the ILSVRC 2015 order task. We additionally present investigation on CIFAR-10 with 100 and 1000 layers. The profundity of portrayals is of focal significance for some, visual acknowledgment assignments. Exclusively because of our incredibly profound portrayals, we acquire a 28% relative enhancement for the COCO object location dataset. Profound lingering nets are establishments of our entries to ILSVRC and COCO 2015 competitions1, where we additionally won the first places on the errands of ImageNet identification, ImageNet restriction, COCO recognition, and COCO division.

S. Xie et al [9] gives a simple, deeply modularized picture grouping organise engineering. Our system is created by rehashing a structural barrier with a topology that has undergone several alterations. Our fundamental strategy results in a homogenous, multi-branch architecture that only requires a few hyper-parameters to be configured. In addition to the dimensions of profundity and width, this method revealed another measurement, which we refer to as "cardinality" (the size of the arrangement of changes), as a crucial feature. On the ImageNet-1K dataset, we precisely demonstrate how increasing cardinality might enhance arrangement precision in the constrained condition of addressing unpredictability. Also, expanding cardinality is more successful than going further or more extensive when we increment the limit. Our models, named ResNeXt, are the establishments of our entrance to the ILSVRC 2016 grouping task in which we made sure about second spot. We further examine ResNeXt on an ImageNet-5K set and the COCO identification set, additionally indicating preferable outcomes over its ResNet partner. The code and models are freely accessible on the web.

In [10], K. Simonyan and A. In the context of large-scale image recognition, Zisserman investigates the impact of the convolutional network depth on its accuracy. Their primary contribution is a detailed analysis of networks with increasing depth utilising an architecture with extremely tiny (3 3) convolution filters, which demonstrates that extending the depth to 16–19 weight layers may significantly outperform existing setups. These results served as the foundation for our entry to the 2014 ImageNet Challenge, which helped our team win first and second place in the localization and classification tracks, respectively. The authors also demonstrate how effectively these representations generalise to different datasets, where they produce cutting-edge outcomes. We have made our two best-performing ConvNet models publicly available to facilitate further research on the use of deep visual representations in computer vision

III. PROPOSED WORK

Our REMAP descriptor's structure identifies two key problems for dealing with content-based picture recovery: A accelerated gathering of multi-area and multi-layer portrayals using end-to-end preparation, as well as (I) a revolutionary collecting component for multi-layer profound convolutional highlights separated by a CNN system. The key peculiarity of our technology is to combine a series of significant deep highlights from multiple CNN layers, which are specifically designed to speak to varied and reciprocal degrees of visual component contemplation, essentially increasing recognition. Importantly, our multi-layer design is expressly and fully ready for recognition from beginning to end. This suggests that many CNN layers are prepared in concert to be:

- Discriminative exclusively (under the particular collection plans utilized inside layers),
- Complementary to one another in acknowledgment errands, and
- Supportive to the extraction of the highlights required at consequent layers.

ISSN 2515-8260 Volume 10, Issue 01, 2023

These differ from the MS-RMAC structure in that fixed loads of the pre-processed CNN are used as an element extractor instead of the CNN being fully prepared. The important and innovative component of our REMAP approach is multi-layer end-to-end finetuning, which simultaneously improves CNN channel loads, relative entropy loads, and PCA+Whitening loads via stochastic gradient descent (SGD) and triplet misfortune work. The CNN must be properly prepared from beginning to end since it unambiguously approves intra-layer include complementarity, which significantly improves performance. The highlights from the additional layers, while occasionally useful, are not prepared to be discriminative or reciprocal without such combined multi-layer learning. The REMAP multi-layer handling can be found in Figure 1, where various equal preparing strands begin from the convolutional CNN layers, each including the ROI-pooling, L2-standardization, relative entropy weighting and Sum-pooling, before being connected into a solitary descriptor.

The district entropy weighting is another important advancement in our technique that is suggested. The objective is to determine how disparaging specific individual characteristics are in each neighbourhood area and then use this knowledge to, hopefully, manage the subsequent whole pooling activity. The relative entropy between the circulations of separations for coordinating and non-coordinating image descriptor sets, calculated using the KL-dissimilarity method, is what is referred to as the location entropy. Areas that provide strong differentiation (high KL-dissimilarity) between coordinating and non-coordinating disseminations are given heavier burdens as a result of their growing enlightenment. Our entropy-controlled pooling allows us to combine a denser set of location-based features without running the risk of the greatest providers being overwhelmed by less knowledgeable areas. The KL-dissimilarity Weighting (KLW) block in the REMAP design is essentially implemented using a convolutional layer with loads introduced by the KL-difference esteems and enhanced using stochastic gradient descent (SGD) on the triplet loss task. A global picture descriptor is created via connecting, PCA brightening, and L2-standardizing the gathered vectors.

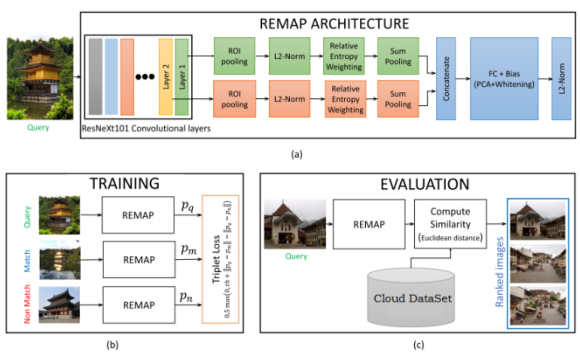


Fig.1. (a) Proposed REMAP architecture with KL-divergence based weighting (KLW) and Multi-layer aggregation (MLA) (b) training of REMAP CNN using triplet loss on Landmarks dataset, (c) Evaluation of REMAP on state-of-the-art datasets.

The REMAP arranges allows for the preparation of the entire engineering project from beginning to end since each square corresponds to a differentiable activity. To fully understand the Cloud datasets and the preparation process, it is recommended that you read the Experimental Section. We prepare the Landmarks-recovery Cloud dataset using triplet misfortune. Additionally, the Product Quantization (PQ) method is used to encode the REMAP marks for the test datasets in order to reduce the memory need and unpredictability of the recovery framework.

Neural networks help us cluster and classify. You can think of them as a clustering and classification layer on top of the data you store and manage. They help to group unlabeled data according to similarities among the example inputs, and they classify data when they have a labelled dataset to train on. (Neural networks can also

ISSN 2515-8260 Volume 10, Issue 01, 2023

extract features that are fed to other algorithms for clustering and classification; so you can think of deep neural networks as components of larger machine-learning applications involving algorithms for reinforcement learning, classification and regression.)

Deep learning maps inputs to outputs. It finds correlations. It is known as a “universal approximate”, because it can learn to approximate an unknown function $f(x) = y$ between any input x and any output y, assuming they are related at all (by correlation or causation).

1) Classification

All classification tasks depend upon labeled datasets; that is, humans must transfer their knowledge to the dataset in order for a neural network to learn the correlation between labels and data. This is known as supervised learning.

- Detect faces, identify people in images, recognize facial expressions (angry, joyful)
- Identify objects in images (stop signs, pedestrians, lane markers…)
- Recognize gestures in video
- Detect voices, identify speakers, transcribe speech to text, recognize sentiment in voices
- Classify text as spam (in emails), or fraudulent (in insurance claims); recognize sentiment in text (customer feedback)

Any labels that humans can generate, any outcomes that you care about and which correlate to data, can be used to train a neural network.

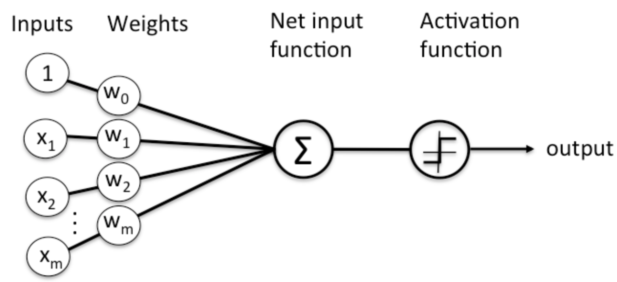
2) Clustering

Clustering or grouping is the detection of similarities. Deep learning does not require labels to detect similarities. Learning without labels is called unsupervised learning. Unlabeled data is the majority of data in the world. One law of machine learning is: the more data an algorithm can train on, the more accurate it will be. Therefore, unsupervised learning has the potential to produce highly accurate models.

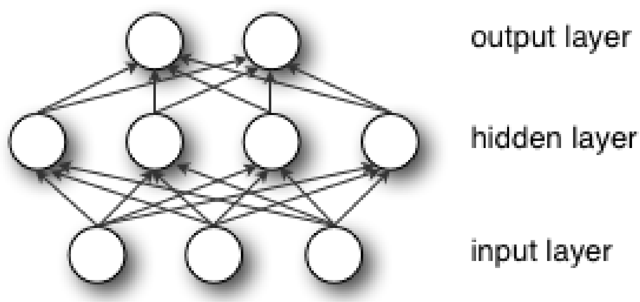
- Search: Comparing documents, images or sounds to surface similar items.
- Anomaly detection: The flipside of detecting similarities is detecting anomalies, or unusual behavior. In many cases, unusual behavior correlates highly with things you want to detect and prevent, such as fraud.

3) Neural Network Elements

Deep learning is the name we use for “stacked neural networks”; that is, networks composed of several layers. The layers are made of *nodes*. A node is just a place where computation happens, loosely patterned on a neuron in the human brain, which fires when it encounters sufficient stimuli. A node combines input from the data with a set of coefficients, or weights, that either amplify or dampen that input, thereby assigning significance to inputs with regard to the task the algorithm is trying to learn; e.g. which input is most helpful is classifying data without error? These input-weight products are summed and then the sum is passed through a node’s so-called activation function, to determine whether and to what extent that signal should progress further through the network to affect the ultimate outcome, say, an act of classification. If the signals pass through, the neuron has been “activated.”



A node layer is a row of those neuron-like switches that turn on or off as the input is fed through the net. Each layer's output is simultaneously the subsequent layer's input, starting from an initial input layer receiving your data.



Pairing the model's adjustable weights with input features is how we assign significance to those features with regard to how the neural network classifies and clusters input.

Deep-learning networks are distinguished from the more commonplace single-hidden-layer neural networks by their depth; that is, the number of node layers through which data must pass in a multistep process of pattern recognition. Deep-learning networks perform automatic feature extraction without human intervention, unlike most traditional machine-learning algorithms. Given that feature extraction is a task that can take teams of data scientists years to accomplish, deep learning is a way to circumvent the chokepoint of limited experts. It augments the powers of small data science teams, which by their nature do not scale.

When training on unlabeled data, each node layer in a deep network learns features automatically by repeatedly trying to reconstruct the input from which it draws its samples, attempting to minimize the difference between the network's guesses and the probability distribution of the input data itself. Restricted Boltzmann machines, for examples, create so-called reconstructions in this manner.

Deep-learning networks end in an output layer: a logistic, or softmax, classifier that assigns a likelihood to a particular outcome or label. We call that predictive, but it is predictive in a broad sense. Given raw data in the form of an image, a deep-learning network may decide, for example, that the input data is 90 percent likely to represent a person.

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision. The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep

Learning has been constructed and perfected with time, primarily over one particular algorithm — a Convolutional Neural Network.

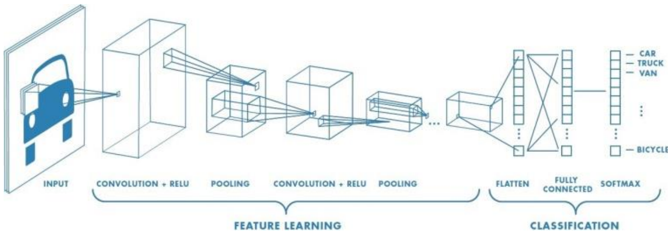


Fig 1 *architect CNN*

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.

The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

IV. RESULTS

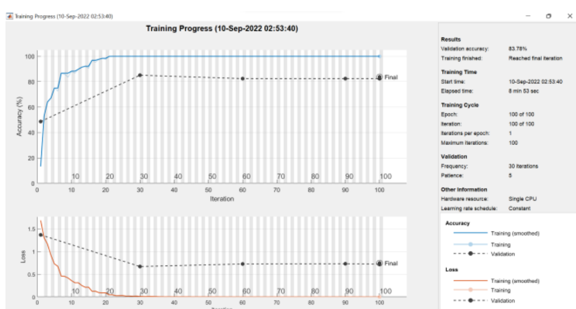
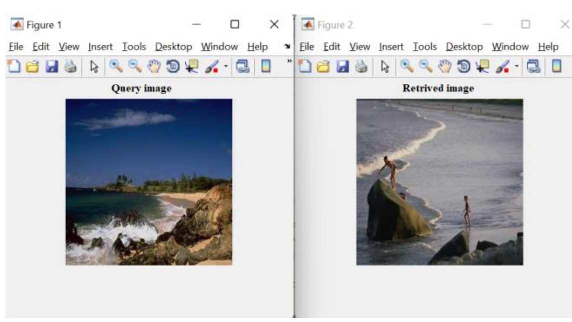


Fig.2. Training Process



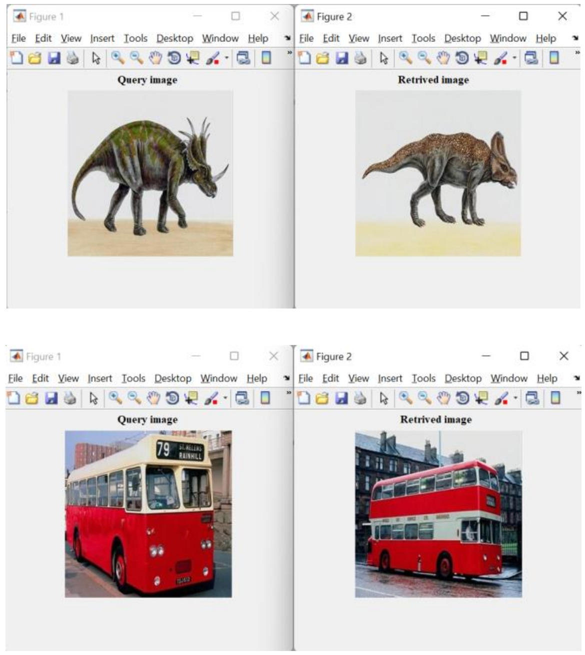


Fig.3. Input and Out put

TABLE 1 performance parameters of existing system

TP	FP	PERFORMANCE PARAMETERS of previous system	
FN	TN		
227	23	sensitivity, recall, hit rate, or true positive rate (TPR)	0.828467153
47	203	specificity, selectivity or true negative rate (TNR)	0.898230088
		precision or positive predictive value (PPV)	0.908
		negative predictive value (NPV)	0.812
		miss rate or false negative rate (FNR)	0.171532847
		fall-out or false positive rate (FPR)	0.101769912
		false discovery rate (FDR)	0.092
		false omission rate (FOR)	0.188
		accuracy (ACC)	0.86
		F1 score	0.866412214
		MCC	0.72334087

Accuracy of face detection in existing system is 86% whereas by using proposed system it increases to 98%, similarly for F1 score, for existing it is 0.866 and for proposed system it is 0.9798. In both accuracy and F1 score, proposed system is found to be better than existing one. False discovery and false recovery rate are negligible in proposed system, only 0.028 and 0.012 respectively which can be neglected while performing detection.

TABLE 2 performance parameters of proposed system

TP	FP	PERFORMANCE PARAMETERS of proposed system	
FN	TN		
243	7	sensitivity, recall, hit rate, or true positive rate (TPR)	0.987804878
3	247	specificity, selectivity or true negative rate (TNR)	0.972440945
		precision or positive predictive value (PPV)	0.972
		negative predictive value (NPV)	0.988
		miss rate or false negative rate (FNR)	0.012195122
		fall-out or false positive rate (FPR)	0.027559055
		false discovery rate (FDR)	0.028
		false omission rate (FOR)	0.012
		accuracy (ACC)	0.98
		F1 score	0.97983871
		MCC	0.960122904

V. CONCLUSION

Using a grandiose CNN-based design known as REMAP, we may learn a chain of command of deep highlights that speak to diverse and correlated degrees of visual reflection. We combine a substantial collection of these dispersed CNN highlights, pooled over several geographic regions, and pack them with weights that reflect their distinctness. The loads are established using KL-difference values for each spatial location and streamlined from beginning to end using SGD, as well as the CNN features. Triplet misfortune is used throughout the whole structure's preparation, and extensive testing show that REMAP virtually outperforms the most advanced technology.

REFERENCES

[1] Husain, S. S., & Bober, M. (2019). REMAP: Multi-layer entropy-guided pooling of dense CNN features for image retrieval. IEEE Transactions on Image Processing, 1–1.

[2] Content-Based Image Retrieval Using Convolutional Neural Networks Ouhda Mohamed, El Asnaoui Khalid, Ouanan Mohammed, and Aksasse Brahim Springer International Publishing AG, part of Springer Nature 2019 J. Mizera-Pietraszko et al. (Eds.): RTIS 2017, AISC 756, pp. 463–476, 2019.

[3] Rian, Z., Christanti, V., & Hendryli, J. (2019). Content-Based Image Retrieval using Convolutional Neural Networks. 2019 IEEE International Conference on Signals and Systems (ICSigSys). doi:10.1109/icsigsys.2019.8811089

[4] Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., Khalil, T. (2019). Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review. Mathematical Problems in Engineering, 2019, 1–21. doi:10.1155/2019/9658350

[5] G. Tolias, R. Sicre, and H. Jégou, “Particular object retrieval with integral max-pooling of CNN activations,” in Proc. Int. Conf. Learn Representations, 2016.

[6] R. Arandjelovi´c, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “NetVLAD: CNN architecture for weakly supervised place recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 6, pp. 1437–1451, Jun. 2018.

[7] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, “End-to-end learning of deep visual representations for image retrieval,” Int. J. Comput. Vis., vol. 124, no. 2, pp. 237–254, Sep. 2017.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[9] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 5987–5995

[10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Represent. 2015.

[11] Mulani, Altaf O., and P. B. Mane. "Watermarking and cryptography based image authentication on reconfigurable platform." Bulletin of Electrical Engineering and Informatics 6.2 (2017): 181-187.

[12] Jadhav, Makrand M. "Machine Learning based Autonomous Fire Combat Turret." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.2 (2021): 2372-2381.

[13] Swami, Shweta S., and Altaf O. Mulani. "An efficient FPGA implementation of discrete wavelet transform for image compression." 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS). IEEE, 2017.

[14] Shinde, Ganesh, and Altaaf Mulani. "A robust digital image watermarking using DWT-PCA." International Journal of Innovations in Engineering Research and Technology 6.4 (2019): 1-7.

[15] Kulkarni, Priyanka R., Altaaf O. Mulani, and P. B. Mane. "Robust invisible watermarking for image authentication." Emerging Trends in Electrical, Communications and Information Technologies. Springer,

Singapore, 2017. 193-200.

[16] Bhanudas Gadade and Altaf Mulani, “Automatic System for Car Health Monitoring”, International Journal of Innovations in Engineering Research and Technology, 57–62, 2022

[17] Pratima Amol Kalyankar, Altaf O. Mulani, Sampada P. Thigale, Pranali Gajanan Chavhan and Makarand M. Jadhav, “Scalable face image retrieval using AESC technique”, Journal Of Algebraic Statistics Volume 13, No. 3, p. 173 – 176, 2022

[18] A. O. Mulani and G. N. Shinde, “An approach for robust digital image watermarking using DWT- PCA”, Journal of Science and Technology, Vol.6, Special Issue 1, 2021 DOI: https://doi.org/10.46243/jst.2021.v6.i04.pp59- 62

[19] U. P. Nagane and A. O. Mulani, “Moving Object Detection and Tracking Using Matlab”, Journal of Science and Technology, Vol.6, Special Issue 1, 2021 DOI: https://doi.org/10.46243/jst.2021.v6.i04.pp63- 66

[20] Priyanka Kulkarni and A. O. Mulani, “Robust Invisible Digital Image Watermarking using Discrete Wavelet Transform”, International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue01, pp.139- 141, Jan.2015

[21] Mulani, Altaf O., and Pradeep B. Mane. "High-Speed Area-Efficient Implementation of AES Algorithm on Reconfigurable Platform." Computer and Network Security (2019): 119.

[22] Deshpande, Hrushikesh S., Kailash J. Karande, and Altaaf O. Mulani. "Area optimized implementation of AES algorithm on FPGA." 2015 International Conference on Communications and Signal Processing (ICCSP). IEEE, 2015.

[23] Godse, A. P., and A. O. Mulani. Embedded systems. Technical Publications, 2009.

[24] Mulani, Altaf O., and P. Mane. "Secure and area efficient implementation of digital image watermarking on reconfigurable platform." Int. J. Innov. Technol. Explor. Eng.(IJITEE) 8.2 (2018): 1.

[25] Rahul G. Ghodake and A. O. Mulani, “Microcontroller Based Drip Irrigation System”, Techno-societal 2016, International conference on advanced technologies for societal applications, pp. 109–115.

[26] Amruta Mandwale and A. O. Mulani, “Different Approaches For Implementation of Viterbi decoder”, IEEE International Conference on Pervasive Computing (ICPC), Jan. 2015.

[27] Amruta Mandwale and A. O. Mulani, “Implementation of Convolutional Encoder & Different Approaches for Viterbi Decoder”, IEEE International Conference on Communications, Signal Processing Computing and Information technologies, Dec. 2014.

[28] Amruta Mandwale and A. O. Mulani, “Implementation of High Speed Viterbi Decoder using FPGA”, International Journal of Engineering Research & Technology (IJERT), Feb. 2016

[29] D. M. Korake and A. O. Mulani, “Design of Computer/Laptop Independent Data transfer system from one USB flash drive to another using ARM11 processor”, International Journal of Science, Engineering and Technology Research, 2016.

[30] Rahul G. Ghodake and A. O. Mulani, “Sensor Based Automatic Drip Irrigation System”, Journal for Research, 53-56, 2016.

[31] Rahul Shinde and A. O. Mulani, “Analysis of Biomedical Image”, International Journal on Recent & Innovative trend in technology (IJRITT), July 2015

[32] Rahul Shinde and A. O. Mulani, “Analysis of Biomedical Image using Wavelet Transform”, International Journal of Innovations in Engineering Research and Technology (IJIERT), July 2015

[33] A. O. Mulani and P. B. Mane, "Area optimization of cryptographic algorithm on less dense reconfigurable platform,"2014 International Conference on Smart Structures and Systems (ICSSS), Chennai, 2014, pp. 86- 89

[34] A.O.Mulani, M. M. Jadhav and Mahesh Seth, “Painless Non- invasive blood glucose concentration level estimation using PCA and machine learning” in the CRC Book entitled Artificial Intelligence, Internet of Things (IoT) and Smart Materials for Energy Applications, 2022.

[35] Kamble, Akshata, and A. O. Mulani. "Google Assistant based Device Control." Int. J. of Aquatic Science 13.1 (2022): 550-555.

[36] Pathan, Atik N., et al. "Hand Gesture Controlled Robotic System." Int. J. of Aquatic Science 13.1 (2022): 487-493.

[37] Kolekar, Supriya D., et al. "Password Based Door Lock System." Int. J. of Aquatic Science 13.1 (2022): 494-501.

[38] Swapnil Takale, Dr. Altaaf Mulani, "Video Watermarking System", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 10, Issue III, Mar-2022.

[39] J. P. Patale et al. "Python Algorithm to Estimate Range of Electrical Vehicle", Telematique, Volume 21, No. 1, 2022.

[40] Jayshri Prakash Patale, A. B. Jagadale, A. O. Mulani, and Anjali Pise. "A Systematic Survey on Estimation of Electrical Vehicle". Journal of Electronics,Computer Networking and Applied Mathematics(JECNAM) ISSN : 2799-1156, vol. 3, no. 01, Dec. 2022, pp. 1-6, doi:10.55529/jecnam.31.1.6.

[41] Kashid, M.M., Karande, K.J., Mulani, A.O. (2022). IoT-Based Environmental Parameter Monitoring Using Machine Learning Approach. In: Kumar, A., Ghinea, G., Merugu, S., Hashimoto, T. (eds) Proceedings of the International Conference on Cognitive and Intelligent Computing. Cognitive Science and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-19-2350-0_5

[42] Swapnil Takale, and Dr. Altaaf Mulani. "DWT-PCA Based Video Watermarking". Journal of Electronics, Computer Networking and Applied Mathematics(JECNAM) ISSN : 2799-1156, vol. 2, no. 06, Nov. 2022, pp. 1-7, doi:10.55529/jecnam.26.1.7.

[43] A. O. Mulani and Dr. P. B. Mane, "High throughput area efficient FPGA implementation of AES Algorithm", in the Intech Open Access Book entitled Computer and Network Security, Feb. 2019.

[44] Jadhav, H.M., Mulani, A. and Jadhav, M.M. (2022). Design and Development of Chatbot Based on Reinforcement Learning. In Machine Learning Algorithms for Signal and Image Processing (eds D. Ghai, S.L. Tripathi, S. Saxena, M. Chanda and M. Alazab). https://doi.org/10.1002/9781119861850.ch12

[45] V. B. Utpat, Dr. K. J. Karande, Dr. A. O. Mulani, "Grading of Pomegranate Using Quality Analysis", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 10 Issue II Feb 2022.

[46] Aasawari Boxey, Anushri Jadhav, Pradnya Gade, Priyanka Ghanti, & Dr.A.O. Mulani. (2022). Face Recognition using Raspberry Pi. Journal of Image Processing and Intelligent Remote Sensing(JIPIRS) ISSN 2815-0953, 2(04), 15–23. https://doi.org/10.55529/jipirs.24.15.23