A Graphical Approach for Image Retrieval Based on Five Layered CNNs Model

Mohammad Khalid Imam Rahmani

College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia

Corresponding author: Mohammad Khalid Imam Rahmani, Email: m.rahmani@seu.edu.sa

Image processing is an important field in the computer vision domain. A lot of work has been done for the processing of image data in various fields like science and technology, defense, medical, space science for satellite imagery analysis, seismology, traffic control, crime control,publishing, and other emerging research areas.There are different levels of complexities for the accurate retrieval of images as most of the images are affected by different kinds of noise and other factors. In this proposed work, I have performed the work of image retrieval using two methods: firstly, processing for denoising and filtering of the data set of images taking density parameter 0.7 and adaptive gamma parameter constant value 0.5. The obtained images are then processed by Convolutional neural networks (CNNs). The 5-layer convolutional neural network has been used for the best features extraction and then the algorithm is finally optimized using GA (Genetic Algorithm). In my work I have used 5*5 fold convolutional layers and compared the results with the previous approach Deep Convolutional neural network (DCNN). Finally, the Genetic Algorithm is implemented to obtain the best-optimized value. The proposed work is validated with a graphical-based approach using the mathematical results in terms of peak signal-to-noise ratio (PSNR), mean-squared error (MSE), and the processing time of the algorithm. The result parameters of the proposed algorithm clearly show better performance as compared to the previous approach.

Keywords: PSNR, MSE, GA, CNNs, Image Processing, Image Retrieval.

2021. In Raju Pal & Praveen K. Shukla (eds.), *SCRS Conference Proceedings on Intelligent Systems*, 261–270. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-93-91842-08-6-26

1 Introduction

The multiplication of shrewd cell phones has prompted a blast in the measure of pictures that are caught, put away, and dissected. Then again, the accessibility of expanded register force and web network has empowered the utilization of complex PC vision calculations for visual examination. For sure, the products of such advances have brought about applications, for example, Google Lens that works on the nature of lives of people by giving an abundance of data. In any case, the exhibition of PC vision calculations on camera caught pictures can debase because of an assortment of bends like clamor, goal, pressure, and light [1-3].

To give a solid base for the extraction of visual investigation, a requirement to betterthe heartiness of the PC vision calculations within the sight of such twists is paramount. Ifocuson a particular example of this heartiness question by thinking about issues of picture recovery. I consider the plan of picture improvement calculations to guarantee the vigorous exhibition of recovery calculations within the sight of mutilations because of clamor and low goal [4-5].

We note that the picture recovery calculation we allude to here is the old-style recovery issue where the objective is to recover pictures from an information base with comparative substance or semantic likeness. Picture recovery dependent on the pack of words model has been concentrated widely. A few upgrades have likewise been tried to defeat the impediments of highlight identifiers/descriptors, descriptor examination measurements, and descriptors quantization [6-7]. By and by, the exhibition of picture recovery within the sight of mutilations and how to further develop execution in such situations has been significantly less contemplated. Different procedures created so far have advanced to upgrade the perception levels of the enhanced pictures. Worked with measurable priors on normal pictures and misusing the comparability of fixed content across the picture has prompted picture denoising calculations with incredible execution [8-11].

The hypothesis of meager sign portrayals has been utilized to foster cutting-edge single picture super goal calculations [12]. The best-in-class execution can be accomplished for both these issues utilizing basic designs of CNNs. From the outset sight, the streamlining of traditional denoising, as well as single picture super goal calculations for picture recovery errands, gives off an impression of being tested. This is part of the way because the denoising or super goal calculations themselves are perplexing including non-direct activities of different boundaries that should be advanced. While the utilization of profound CNNs works on the improvement activity partially, networks are normally advanced for cost capacities, for example, regularized mean squared mistake or perceptual quality files, for example, the primary closeness file [13-15]. While these expense capacities might be applicable for perceptual quality, their importance in working on the presentation of picture recovery isn't clear. The estimation of picture recovery execution includes two segments, the recovery calculation itself and the presentation assessment of the yield of the recovery calculation as far as measurements like normal accuracy by contrasting the yield and a commented-on data set. These include an unpredictable succession of tasks that can't be composed as a shut structure articulation. In this manner, it's anything but clear how a differentiable expense capacity can be acquired that can be utilized to streamline the picture improvement calculations [16-20].

There is a gap in the literature work as most of the work is focused on image retrieval based on direct CNN or processed after denoising [21-26]. The image retrieval may be retained compromising the image quality [27-29]. In the present work, I have focusedmy study on image retrieval [30, 31] provided the image quality should be retained. The parameters for deciding image quality are PSNR, MSE and the proposed algorithm should be light in processing so that the processing time taken is less as compared to previous approaches [32]. The remainder of the paper is outlinedin the manner: I have discussed the proposed methodology in section 2, the results and discussion in section 3, and finally, section 4 stated the conclusion and future work.

2 Proposed Methodology

In the proposed work the dataset of images is first processed with the binarization process known as pre-processing. The images are first denoised and then processed to enhance the overall quality of the image. The raw dataset images may contain various noise which affects the performance of the algorithm. So, we first remove the Gaussian noise using density function (0.7) and then use the gamma correction function to enhance the image quality. In the proposed work we have used 0.5 gamma parameter as compared to the previous work which utilizes variable parameter which results in over enhancement of the image quality results in pixel distortion. The deep features of the images are then extracted using a Convolutional neural network using 5 layered structures.The proposed model of the work is shown in Fig. 1.

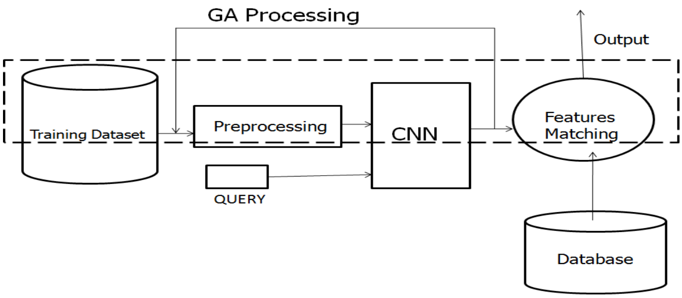

Fig. 1. Proposed model

CNN has had all the earmarks of being the best in the classification system for picture retrieval and examination. CNN is a finished bundle for picture examination and doesn't need a specific segment for planning. This methodology is profitable since we don't have to investigate different component extractors. Profound learning models tentatively animated by data preparing, and correspondence diagrams in tangible frameworks yet have extraordinary separation from the associate and utilitarian properties of regular minds. The basic Architecture of the CNN is discussed in the next subsection.

2.1 The architecture of the VGG-F Model

This investigation is restricted to the models VGG-F and ResNet. The Visual Geometry Group (VGG) model uses five convolutional layers of which three are completely associated with a large number of learning boundaries. Normally, the last layer is the 'Softmax' work arrangement layer. This model is prepared by forecast misfortune minimization. The normal misfortune capacity is describedin equation (1) below:

$$J(w) = \frac{1}{N}\sum_{i=1}^{N} L\big(f\big(w, x(i), y(i)\big)\big) + \lambda R(w) \tag{1}$$

where N = No. of iterations/data instances
L= Loss function
f = Predicted output
w = Weights
λR = Weight decay with the Lagrange multiplier λ.

The visual geometry group (VGG) uses a 3*3 filter for the classification of images. The VGG Model performs well for the exact localization of image features. In this work, I have used this model to avoid any pixel loss and then a ResNet model is applied which understands the layering topology for the advanced categorization, localization, and detection within the same model. The model is intelligent within its layers such that the next layer learns from the input layer and if distortion occurs corrects at its level for further processing of the next layer. As stated in equation(1), I have corrected the weight and lambda parameterto minimize the loss function which would be helpful in the retrieval of images with more accuracy. The Softmax is the last layer utilized in the work. This is the lightest model for the processing and is given by equation (2) [8].

$$\phi_i = \frac{\exp(\theta_i^T z)}{\sum_{j=1}^{k} \exp(\theta_i^T z)} \tag{2}$$

here, j varies from 1 to k where k is the total number of set images of the database. T is the threshold value, and z is the input function that varies exponentially for the training purpose and gives output in terms of Φ.

The proposed model as shown in Fig. 1 indicates the flow of the algorithm. First of all, the training dataset containing images is being processed for the pre-processing which involves the conversion of RGB images into gray form and then binarization of the images like extracting the pixel. The dataset has been passed through a Gaussian filter to remove the noise present and then the distorted pixel is restored using the image enhancement technique. Further, the different models of the deep convolutional neural network have been implemented for deep features extraction and retrieval of the final image.

The proposed algorithm is further optimized using the Genetic algorithm which repeats the iteration until the best fitness value is achieved. In the proposed work the genetic algorithm takes the input from the last CNN layer then processes the same from the beginning and matches the previous threshold value which must be well optimized otherwise takes the previous value. This way the algorithm iterates for n number of times to get the optimized performance.Here I have defined the initial population and after processing, it obtained the final value which further moves to the initial level for further processing and the event keeps continuing until obtains the final value better than the threshold value.

3 Results and Discussion

Here I present the results and discuss them for accurate image retrieval based on a convolutional neural network. The algorithm has been processed through different techniques to obtain the best iterative value. The algorithm involves pre-processing of an image, going through the denoising phenomenon, and then enhancement of the image. The mathematical results are obtained in terms of PSNR with equation (3), MSE with equation (4), and processing time. The parameters have shown that the proposed algorithm has outperformed the existing results. The mathematical equations for PSNR and MSE are defined below:

$$PSNR = 10\, Log_{10} \frac{(L-1)^2}{MSE} \tag{3}$$

where, L is the number of pixels.

$$MSE = \frac{\sum\sum |x(i,j) - y(i,j)|^2}{N} \tag{4}$$

where, x(i, j) is the input image and y(i, j) is the output image.

Training has been done using discrete cosine transforms (DCT), Zernike moments (Z Moment) and testing has been done using DCT and Genetic algorithm (GA). Firstly, the image folder was uploaded to train with the dataset, then the image was processed using DCT for image division then the Z Moment was appliedto get the enhanced image.In the proposed work the CNN model uses convolutional layers with filters with a small receptive filed: 5*5. It has a convolutional stride of 1 pixel and spatial padding of 0, 1, or 2. Five max-pooling layers are used to do spatial pooling after convolutional operations. Max-pooling is done with a 2*2 pixel window, with stride 2. In the proposed work I have used 3 Fully-Connected layers. The soft-max layer is the last one that usesthe soft-max activation function. Further, I have useda 10-foldcross-validation technique isto ensure that outcomes are not biased to any specific set of data. 9-foldswere used for training purposes, and 1-foldwas used for testing purposes. Around 2656 images were taken in the positive set which was taken froma camera. For the negative image set,1000 imageswere collected from the internet.

The mathematical results obtained have been summarized and compared with the previous work. Here, I have considered Ref [22] as the base for the comparison. In the previous approach, it has been implemented using a simple Deep Convolution Neural Network (DCNN). I have again implemented the previous algorithm to obtain the mathematical value and implemented the enhanced approach for comparing the two approaches. The processing time is calculated using the in-built function of MATLAB tic-toc which is used to calculate the processing time of the algorithms. It is equivalent to the CPU time utilization during the execution of codes.

In Table 1, a comparison of mathematical results of PSNR, MSE, and Processing time obtained by the proposed work and the previous work has been summarized.

Table 1. Comparison of PSNR, MSE, and processing timebetween the proposed work and the previous work

Image No.	PSNR		MSE		Processing time	
	Proposed work	Previous work	Proposed work	Previous work	Proposed work	Previous work
1	49.59	42.28	0.07195	0.43276	0.719	1.23
2	51.37	41.23	0.03488	0.72375	0.453	2.86
3	46.83	37.97	0.09854	0.43268	0.875	0.94
4	53.32	43.34	0.03784	0.37426	0.546	1.83
5	48.82	38.42	0.088754	0.34267	0.676	1.37

The result clearly shows that the parameter values of the proposed algorithm are outperforming the previous work.In Fig. 2 the values of the PSNRobtained by the proposed work have been shown.

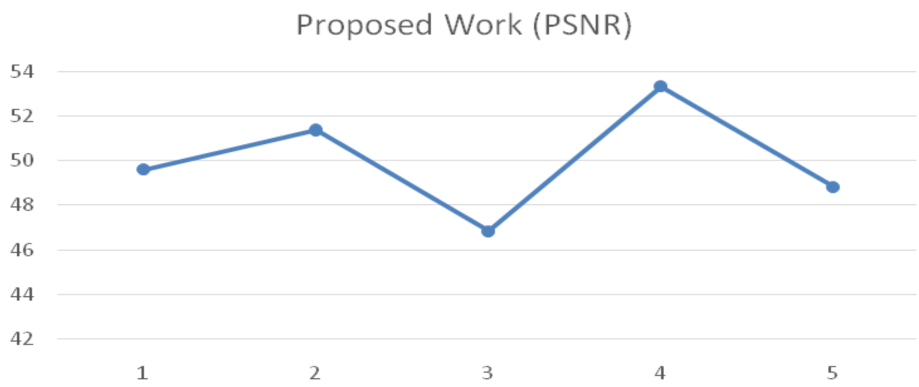

Fig. 2. The PSNRs of the proposed work

In Fig. 3 the values of the PSNR obtained by the previous work have been shown.

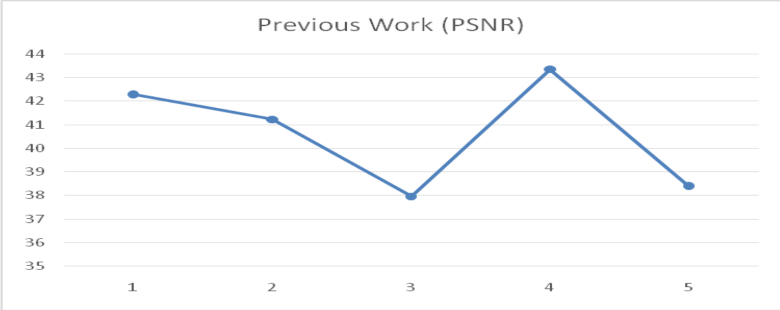

Fig. 3. The PSNRs of the previous work

In Fig. 4 the values of the PSNR obtained by both the proposed and previous works for comparison have been shown.

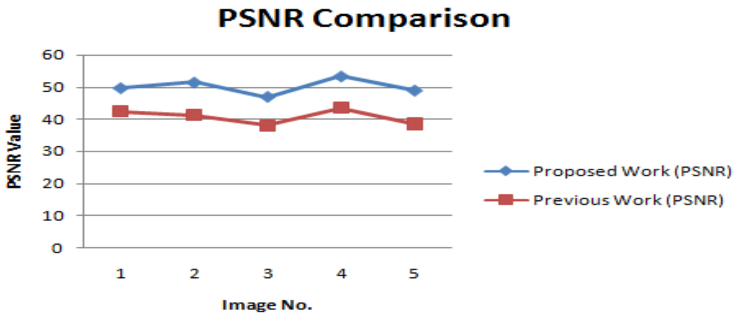

Fig. 4. Comparison of PSNRs obtained by the proposed work and the previous work

It is clerly shown that proposed work is outperforming the previous work in terms of the PSNR.

In Fig. 5 the MSE values obtained by the proposed work have been shown.

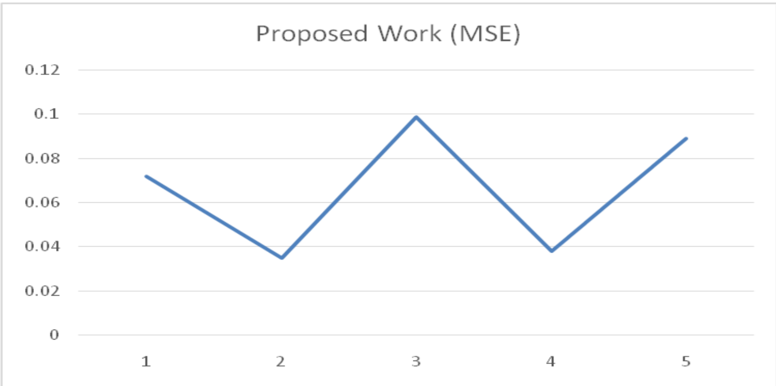

Fig. 5. The MSE of the proposed work

In Fig. 6 the MSE values of the previous work have been shown.

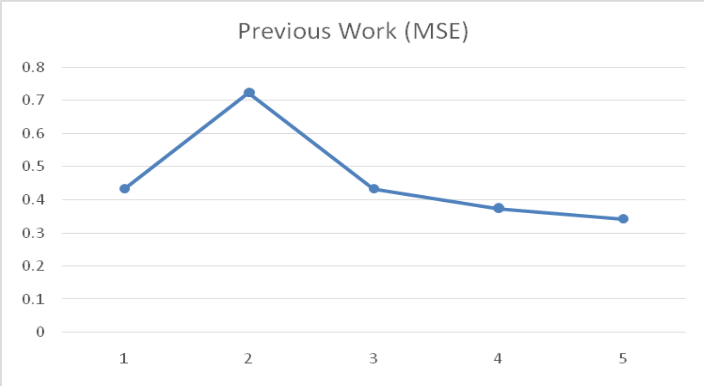

Fig. 6. The MSE of the previous work

In Fig. 7 the MSE values of both the proposed and previous works for comparison have been shown.

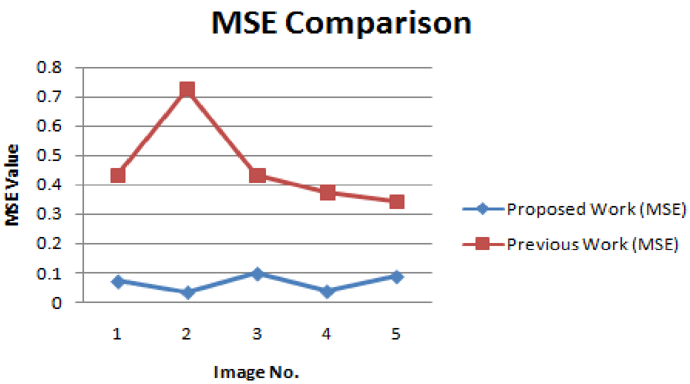

Fig. 7. Comparison of MSE obtained by the proposed work and the previous work

It is clerly shown that proposed work is outperforming the previous work in terms of the MSE values.

In Fig. 8 the processing timestaken by the proposed work have been shown.

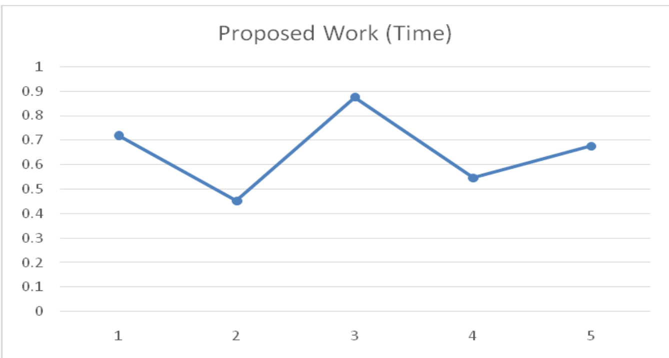

Fig. 8. Processing time forthe proposed work

In Fig. 9 the processing timestaken by the previous work have been shown.

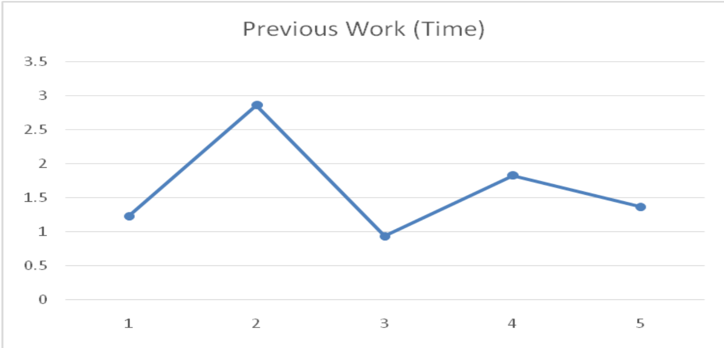

Fig. 9. Processing time,taken bythe previous work

In Fig. 10 the comparison of processing timestaken byboth the proposed and previous works have been shown.

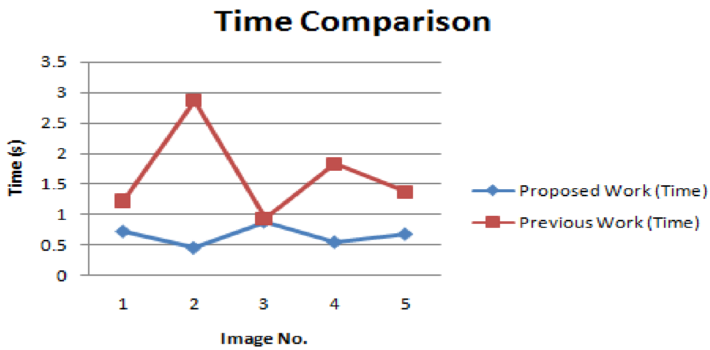

Fig. 10. Comparison of processing times taken by the proposed work and the previous work

It is clerly shown that proposed work is outperforming the previous work in terms of the Processing time.

4 Conclusion and Future Work

The proposed work is implemented in a well-defined manner based on CNN. First of all, the dataset is processed for the removal of noise to improve the image quality, and then the images are enhanced using the gamma correction method. The Convolutional neural network is implemented for the deep features extraction which may help for the image matching during validation. The work is further extended to process the algorithm using a genetic algorithm for optimizing the parameter which enhances the results. Finally, the results are compared with the previous approach with the parameters PSNR, MSE, and the processing time. The proposed algorithm outperforms the existing work. The work may further be extended in the future considering aerial images and video signals extracting its frames.

References

[1] Philbin, J. et al. (2007). Object retrieval with large vocabularies and fast spatial matching. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.
[2] Philbin, J. et al.(2008). Lost in quantization: Improving particular object retrieval in large scale image databases. *IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.
[3] Dabov, K. et al. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080-2095.
[4] Zoran, D. and Weiss, Y. (2011). From learning models of natural image patches to whole image restoration. *International Conference on Computer Vision*, 479-486.
[5] Yang, J. et al. (2010). Image super-resolution via sparse representation. *IEEE Transactions on Image Processing,* 19(11):2861–2873.
[6] Zhang, K. et al. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142-3155.
[7] Wang, Z. et al. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing,* 13(4):600-612.
[8] Gupta, V., Gupta, M. and Singla, P. (2021). Ship Detection from Highly Cluttered Images Using Convolutional Neural Network. *Wireless Personal Communications*, 121:287-305.
[9] Singhal, N. et al. (2018). Comparing CNN and RNN for prediction of judgment in video interview based on facial gestures. *5th International Conference on Signal Processing and Integrated Networks,* 175-180.
[10] Gupta, V. and Gupta, M. (2020). Ships Classification using Neural Network based on Radar Scattering. *International Journal of Advanced Science and Technology*, 29:1349-1354.
[11] Szegedy, C. et al. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 4278–4284.
[12] Szegedy, C. et al. (2015). Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.
[13] Tan, M. and Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*, 6105-6114.
[14] Gupta, V. and Gupta, M. (2020). Automated Object Detection System in Marine Environment. *Mobile Radio Communications and 5G Networks, Lecture Notes in Networks and Systems*, 140:225-235.
[15] Gordo, A. (2016). Deep image retrieval: Learning global representations for image search. *European conference on computer vision*, 241-257.
[16] Hu, J., Shen, L. and Sun, G. (2018). Squeeze-and-excitation networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7132-7141.
[17] Huang, G., Liu, Z. and Weinberger, K. Q. (2016). Densely connected convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 2261-2269.
[18] Dawood, H. et al. (2019). Correlated microstructure descriptor for image retrieval. *IEEE Access,* 7:55206-55228.
[19] Raza, A., Nawaz, T. and Dawood, H. (2019). Square texton histogram features for image retrieval. *Multimedia Tools and Applications*, 78(3):2719-2746.
[20] Alzu'bi, A., Amira, A. and Ramzan, N. (2017). Content-based image retrieval with compact deep convolutional features. *Neurocomputing*, 249:95-105.
[21] Qayyum, A. et al. (2017). Medical image retrieval using deep convolutional neural network. *Neurcomputing*, 266:8-20.
[22] Biju, S., et al. (2020). Robust image retrieval by cascading a deep quality assessment network. *Signal Processing: Image Communication*, 80:115652.
[23] Radenovic, F., Tolias, G. and Chum, O. (2019). Fine-tuning CNN image retrieval with no human annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 41(7):1655-1668.
[24] Liu, Z. et al. (2020). Structured optimal graph based sparse feature extraction for semi-supervised learning. *Signal Processing,* 170:107456.
[25] Tahir, M. and Anwar, S. (2021). Transformers in Pedestrian Image Retrieval and Person Re-Identification in a Multi-Camera Surveillance System. *Applied Sciences,* 11(19):9197.
[26] Tahir, M. et al. (2021). Deep localization of protein structures in fluorescence microscopy images. *arXiv preprint arXiv.*
[27] Liu, Z. et al. (2019). Discriminative low-rank preserving projection for dimensionality reduction. *Soft Computing*, 85:105768.
[28] Pradhan, J. et al. (2018). A hierarchical CBIR framework using adaptive tetrolet transform and novel histograms from color and shape features. *Signal Processing,* 82(1):258-281.
[29] Pradhan, J., Pal, A. K. and Banka, H. (2019). Principal texture direction based block level image reordering and use of color edge features for application of object based image retrieval. *Tools and Applications,* 78(2):1685-1717.

[30] Rahmani, M. K. I. and Ansari, M. A. (2013). Fuzzy image retrieval: Recenttrends. *Indian Journal of Industrial and Applied Mathematics*, 4(2):131–137.
[31] Rahmani. M. K. I. and Ansari, M. A. (2013). A color based fuzzy algorithm for CBIR. *The Next Generation Information Technology Summit (4th International Conference)*, 363-370.
[32] Chu, K. and Liu, G. H. (2020). Image Retrieval Based on a Multi-Integration Features Model. *Mathematical Problems in Engineering,* 2020:1461459.