International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

A Novel Multi Hidden Layer Convolutional Neural Network for Content Based Image Retrieval

K. Ramanjaneyulu, K. Veera Swamy, Ch. Srinivasa Rao

Abstract: ***The applications of a content-based image retrieval system in fields such as multimedia, security, medicine, and entertainment, have been implemented on a huge real-time database by using a convolutional neural network architecture. In general, thus far, content-based image retrieval systems have been implemented with machine learning algorithms. A machine learning algorithm is applicable to a limited database because of the few feature extraction hidden layers between the input and the output layers. The proposed convolutional neural network architecture was successfully implemented using 128 convolutional layers, pooling layers, rectifier linear unit (ReLu), and fully connected layers. A convolutional neural network architecture yields better results of its ability to extract features from an image. The Euclidean distance metric is used for calculating the similarity between the query image and the database images. It is implemented using the COREL database. The proposed system is successfully evaluated using precision, recall, and F-score. The performance of the proposed method is evaluated using the precision and recall.***

Keywords: Convolutional neural network, Euclidean distance and performance measures

I. INTRODUCTION

Over the last few years, the worldwide web (WWW) has become the best information source available today. It needs an effective method to acquire a considerable amount of information from the Internet. Image data are larger than text data, hence, the research community has focused more on the content-based image retrieval (CBIR), very popular method for test image retrieval from a large dataset. The most popular method is convolutional neural network for a huge content-based database. The retrieval system is not sufficient to detect a test image from a large dataset. Therefore, prediction is required for untrained test images. Machine learning is for the good localization of training images in the fields of bio informatics (Bastanlar&Ozuysal,2014), medicine, entrainment, and security.

A content-based image retrieval system is mainly dependent on the data mining of the different patterns of machine learning techniques. The data mining ("Data Min. Pract. Mach. Learn. Tools Tech.," 2016) process has been segregated into different formats of feature extraction techniques such as supervised learning and unsupervised learning algorithms. A support vector network represents (Cortes & Vapnik, 1995) the process of the radial basis function to extract the features of a neural network. It is focused on classification and regression using supervised machine learning techniques. The current trends of the machine learning technique have been studied (Jordan & Mitchell, 2015), certain perspectives and prospects have been used to improve the recognition efficiency of the system. A content-based image retrieval system is the most powerful technique in the field of artificial intelligence. Machine learning (ML) is widely implemented on real-time applications. Thus far, content-based image retrieval has been applied to the health conditions of the human body, military, e-commerce, and many financial models (Harrington, 2012). It has been used to represent innovation thoughts to implement the real-time applications in the areas of ML and deep learning. The advantages of the various actions of machine learning are reliability, space missions, and real-time applications. Machine learning, a probabilistic perspective (Robert, 2015), is represented using different feature extraction models and implemented in different classification methods. The foundation and trends of machine learning (Bengio, 2009) are focused on the concept of a deep convolutional neural network and the design of different layers and an artificial neural network for the system. It has been implemented with a Boltzmann machine, and the deep models of content-based image retrieval (Neapolitan & Neapolitan, 2018) are focused on image recognition, speech recognition, and pattern recognition.

DWT (Demirel & Anbarjafari, 2011) is a multiresolution technique. It can be decomposed into different coefficients such DC coefficients (Dremin, Ivanov, & Nechitailo, 2001) and AC coefficients. The DC coefficients are represented as the LL band of the discrete wavelet decomposition. The AC coefficients are denoted as the LH, HL, and HH bands (Layer & Tomczyk, 2015).

Revised Manuscript Received on February 05, 2020.
* Correspondence Author
K. Ramanjaneyulu*, Research scholar, JNTUK, Associate Professor, Department of ECE, QISIT, Andhra Pradesh, India.
E-mail: ramu36nba@gmail.com
K. Veera Swamy, Professor, Department of ECE, Vasavi College of Engineering, Hyderabad, India.
E-mail: k.veeraswamy@staff.vce.ac.in
Ch. Srinivasa Rao, Professor, Department of ECE, JNTUKUCEV, Vizianagaram.,Andhra Pradesh, India. E-mail:chsrao.ece@jntukucev.ac.in.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

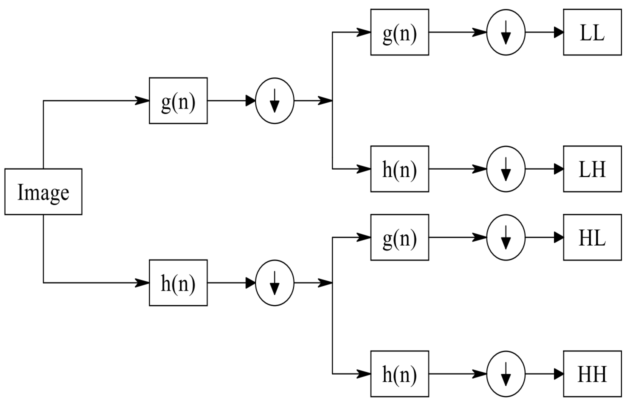

Figure 1 Decomposition of discrete wavelet transform

The discrete cosine transform (DCT) (Ahmed, Natarajan, & Rao, 1974) is widely used in compression and recognition tasks. When applying 2D DCT (Matsui *et al.*, 1994) to an image, in general, information can be represented as a 2D function and is referred to as the low frequency. This low frequency contains the major features of the original image. The 2D DCT (Simoncelli & Adelson, 1991) spectrum C(u, v) of an N × N image Im(x, y) is defined as follows

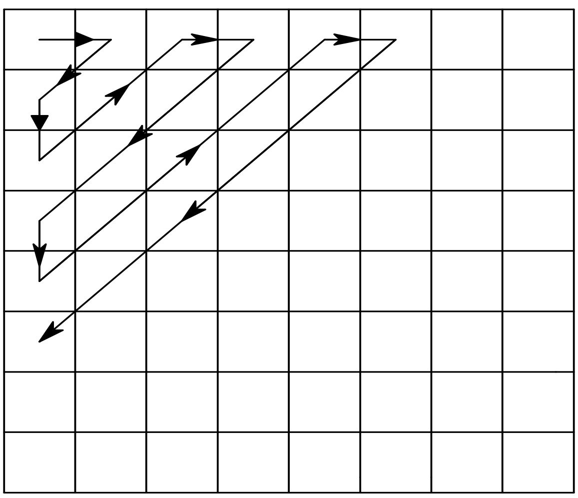

Figure 2 Extraction of DCT coefficients in a zigzag manner

$$C(U,V) = \frac{2}{N}\alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} Im(x,y)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

$$where \;\; \alpha(a) = \begin{cases} \sqrt{\frac{1}{N}}, & For \;\; a = 0 \\ \sqrt{\frac{2}{N}}, & otherwise \end{cases}$$

To construct the feature vector, we took the coefficients of the transformed image in the zigzag manner from the low to the mid frequency. Figure 3 illustrates this technique for an 8 × 8 image.

II. CNN ARCHITECTURE

CNN (Lavin & Gray, 2016) is one of the deep learning technique for image retrieval and detection. The architecture creates a model for feature extraction of the input images. The feature extaction is focused on the different kernels of the input images. Therefore, CNN (Simard, Steinkraus, & Platt, 2003) yields good results in the case of a retrieval system. A CNN (Pang, Sun, Jiang, & Li, 2018) is used to extract the features of the images. To overcome this, it attempts a possible position for extracting the features of an image and makes it a filter/mask.

2.1 Convolution Layers

Convolution (Pang et al., 2018) is basically a combined of input image and kernel of the feature. A feature detector (Harris & Stephens, 2013) is used to extract the features from a large database by using a convolutional operation. The purpose of a convolution layer is to extract the features of the input image by the convolution of the feature detector (Martin, 1994). A feature detector (Mita, Kaneko, & Hori, 2005) of any size mostly depends on the input image size.

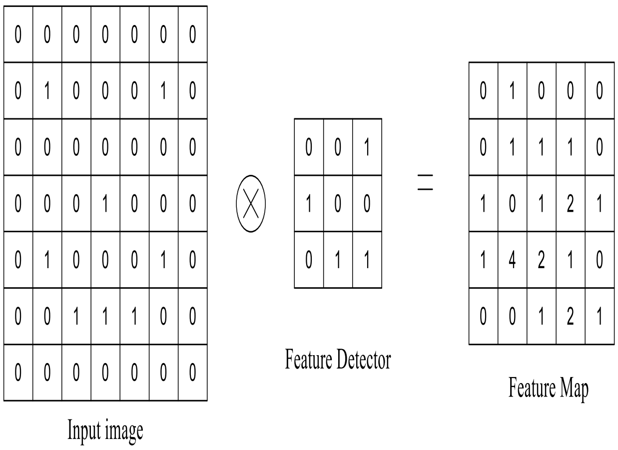

Figure 3 Generation of a featured map using the convolutional process

The ReLu layer (Glorot, Bordes, & Bengio, 2011)(Shang, Sohn, Almeida, & Lee, 2016) is the most powerful tool to extract features. The rectifier linear unit is used to reduce the noisy data of the convolutional layer while building models. It has the output 0 if the input is less than 0; otherwise, the output is raw. However, when it is invented a classification problems, ReLu cannot help much. To overcome this, it can use a SoftMax function (Memisevic, Zach, Hinton, & Pollefeys, 2010) (Larochelle & Lauly, 2012). The SoftMax function is used to compress the data between 0 and 1.

2.2 Pooling Layers

Pooling layer is the most important part in deep convolutional neural network to extract features of preprocessing method of the convolutional operation. (Boureau, Ponce, & Lecun, 2010), it minimizes the process of featured map data. It can be implemented of the process of stride operation of the pooling technique. Pooling is also called down-sampling (Haris, Shakhnarovich, & Ukita, 2018).

The technique consists of two pooling operations: subsampling and max pooling. We used the max pooling technique.

The pooling function using the maximum can be represented as follows:

$$a_j = tanh\left(\beta \sum_{NxN} a_i^{nxn} + b\right)$$

the window function and the scaling factor β

$$a_j = \max_{NxN}(a_i^{nxn}\, u(n,n))$$

Flattering is a technique to express the features are rearranged as the vector form. It is used for to design the input layer model of the system.

III. PROPOSED ALGORITHM

Here, we present three algorithms: 1. CBIR using DCT, 2. CBIR using DWT, and 3. CBIR using the CNN architecture.

3.1 CBIR using DCT

The presentation of the CBIR using DCT has been represented in the following steps:

1. The database can be segregated as the training and testing sets in the ratio of 70:30.
2. Each color image can be converted into a gray image of size 128 × 128.
3. Then, each image is portioned into 8 × 8 images by non-overlapping.
4. DCT is applied to each block of size 8 × 8.
5. The four coefficients (DC, AC1, AC2, and AC3) of each block in an image are considered, and a feature vector is created.

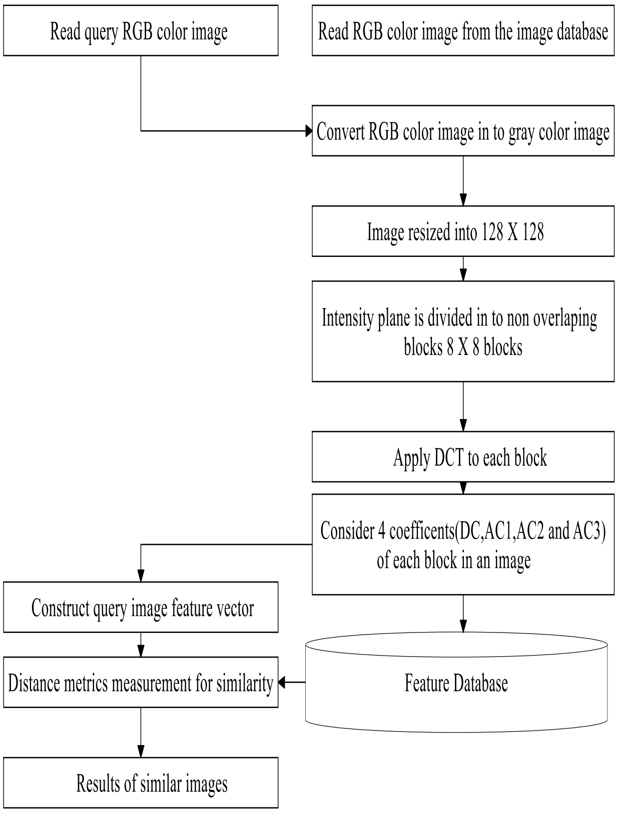

Figure 4 Proposed algorithm using sub-block DCT methodology

6. The feature vector is computed for the testing image.
7. The features are extracted for the testing image to find the similarities from the database.
8. The distance measure technique is applied to retrieve the relevant images.
9. The performance measure is calculated using precision, recall, and F-score.

3.2 CBIR for DWT

The CBIR using DCT can be represented by the following steps:

1. The database can be segregated as the training and testing sets in the ratio of 70:30.
2. Each color image can be converted into a gray image of size 128 × 128.
3. The image can be divided into non-overlapping blocks of size 8 × 8.
4. The two-level DWT is applied to each block as the LL, LH, HL, and HH coefficients.
5. The features are extracted from the LL sub-band for further processing.
6. The same procedure is applied to all the sub-bands of the input image.

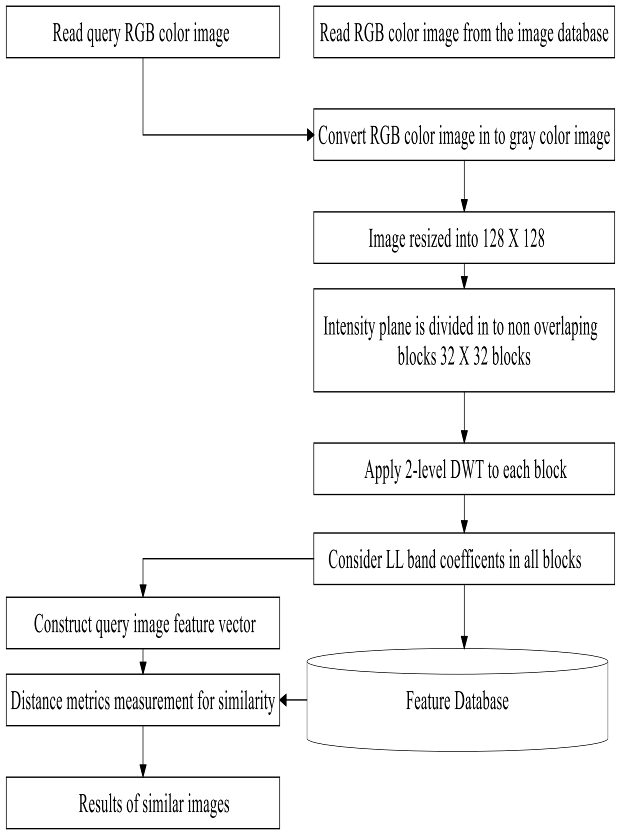

Figure 5 Proposed algorithm using sub-block DWT methodology

7. The features are constructed on the basis of the largest coefficients from the sub-band.
8. A similar image for the test image is obtained from the large database by using the distance measure parameters.
9. The performance measure is calculated using the precision, recall, and F-score.

3.3 CBIR using CNN architecture

1. The experiment is conducted on the CoReL-1k database.
2. The neural network is trained and tested on 1000 **images,** which contain 10 similar people, buildings, etc.
3. Initially, the input image is resized to 128 × 128, and the input is batch processed as 32 images at a time.
4. The neural network architecture for training and testing is as shown.
5. The output from the upper layers is fed to the fully connected layer, and classification is done.

Table 1 Feature extraction process for different layers of CNN

Layer	Dimension of the image	Featured map	Filter size
Original image size	128 X 128	1	No filter
Convolutional layer	128 X 128	16	3 X 3
Pooling layer (Max)	64 X 64	16	2 X 2
Convolutional layer	64 X 64	32	3 X 3
Pooling layer (Max)	32 X 32	32	2 X 2
Convolutional layer	32 X 32	64	3 X 3
Pooling layer (Max)	16 X 16	64	4 X 2

IV. EXPERIMENTAL RESULTS

The experimental results were obtained for the COREL-1K database (Veit, 2015) and used for testing the CNN architectures. The Corel (Di Benigno, Cross, & de Bessonet, 1986) database consists of 1000 images with the RGB color images divided into 10 different categories. Precision is represented as the ratio of the number of relevant retrieval images to the total number of retrieval images.

Precision = True positives/(True positive + False positive).

The range of the precision value is 0 to 100 in percentage.

The precision results for CBIR using CNN architecture is given in Table 3.

Figure 6 Sample images from the Corel database

Figure 6 Retrieval images for testing image "ROSE".

Table 2 Confusion matrix for CBIR using CNN architecture

	Pe	B	B	B	Di	El	Fl	H	M	F
People	77	3	12	0	0	0	0	2	0	7
Beach	2	86	2	0	0	4	0	3	4	0
Building	4	9	77	2	0	4	0	2	3	0
Buses	2	2	2	91	0	0	0	0	0	4
Dinosaur	0	0	0	0	98	0	0	0	2	0
Elephant	3	2	0	0	0	93	0	0	3	0
Flowers	6	0	0	0	0	0	91	2	0	2
Horses	2	0	0	2	0	6	0	90	2	0
Mountain	2	18	6	0	0	4	0	0	72	0
Food	6	0	0	0	0	2	6	0	0	87

Figure 7 Retrieval images for testing image "HORSE"

Table 3 Precision results for CBIR using CNN architecture

Category	CNN (Proposed)	DWT	DCT
People	76.05	65.24	62.01
Beach	73.08	70.00	69.19
Buildings	79.41	74.21	70.11
Buses	96.97	79.16	74.11
Elephants	82.27	84.05	80.28
Flowers	94.12	96.00	91.18
Horses	92.64	92.48	90.15
Mountains	84.71	83.10	80.15
Food	87.14	60.2	63.2
Dinosaurs	100	100	100
Overall	86.64	80.44	78.09

The comparison graph for precision results has been represented in Figure 9

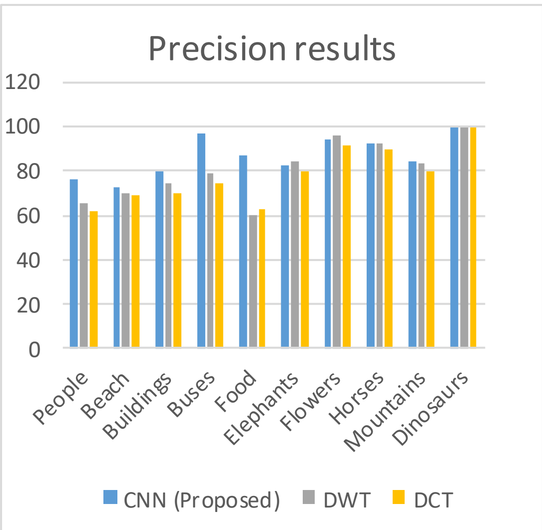

Figure 8 Comparison graph for precision results

Recall is represented as the ratio of number of relevant retrieval images to the total number of images in the database.

Recall = True positive / (True positive + False negative)

The range of the recall value is 0 to 100 in percentage.

The recall results for CBIR using CNN architecture over DWT and DCT is given in Table 4

Table 4 Recall results for CBIR using CNN architecture

Category	CNN (Proposed)	DWT	DCT
People	71.17	53.11	61.45
Beach	85.71	59.06	67.19
Buildings	77.14	59.06	69.18
Buses	91.43	61.42	74.21
Elephants	92.86	77.10	79.29
Flowers	91.43	81.14	89.54
Horses	90.00	74.57	80.00
Mountains	71.29	71.23	76.14
Food	87.14	54.09	57.26
Dinosaurs	98.57	90.37	92.17
Overall	86.64	80.44	78.09

Table 5 F-Score for CBIR using CNN architecture

Category	CNN (Proposed)	DWT	DCT
People	76.61	57.22	63.29
Beach	78.89	63.72	68.57
Buildings	78.26	64.11	71.61
Buses	94.12	67.17	76.61
Elephants	87.24	78.63	81.60
Flowers	92.76	85.87	92.66
Horses	91.30	81.62	85.79
Mountains	77.42	75.43	79.47
Food	87.14	58.29	58.69
Dinosaurs	99.28	94.94	95.93
Overall	86.64	80.44	78.09

The comparison graph for precision results has been represented in Figure 10.

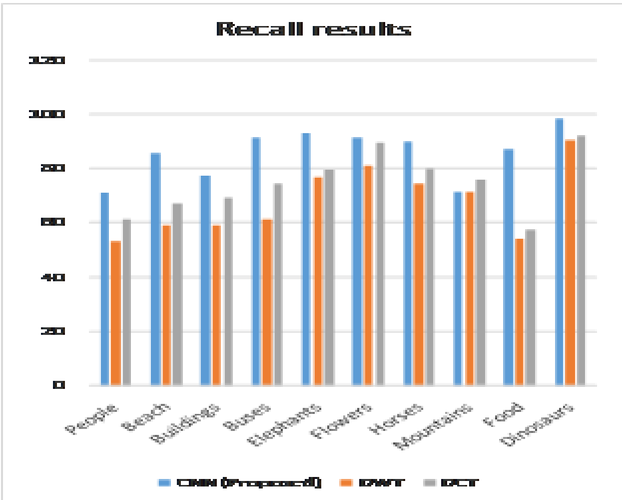

Figure 9 Comparison of recall values with different methods

V. CONCLUSION

A CNN was successfully implemented in this study. The discrete cosine transform for content-based image retrieval for different categories such as people, beach, building, buses, elephants, flowers, horses, mountains, food, and dinosaurs was 63.29%, 68.57%, 71.61%, 76.61%, 81.60%, 92.66%, 85.79%, 79.47%, 58.69%, 95.93%, and 78.9%, respectively. The discrete wavelet transform for content-based image retrieval for different categories such as people, beach, building, buses, elephants, flowers, horses, mountains, food, and dinosaurs was 57.22%, 63.72%, 64.11%, 67.17%, 78.63%, 85.87%, 81.62%, 75.43%, 58.29%, 94.94%, and 80.44%, respectively. The discrete wavelet transform for content-based image retrieval for different categories such as people, beach, building, buses, elephants, flowers, horses, mountains, food, and dinosaurs was 76.61%, 78.89%, 78.26%, 94.12%, 87.24%, 92.76%, 91.30%, 77.42%, 87.14%, 99.28%, and 86.64%, respectively. Therefore, we concluded that the proposed method gave better results than the existing methods.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Dr. S. Srinivasa Rao, Vice Chancellor, JNTUA, Ananthapuram, for his moral support and valuable suggestions for the completion of this work.

REFERENCES

1. Ahmed, K. R. Discrete Cosine Transform. *IEEE Transactions on Computers*. 1974,https://doi.org/10.1109/T-C.1974. 223784.
2. Baştanlar, M. Introduction to machine learning. *Methods in Molecular Biology*. 2014. https://doi.org/10.1007/978 -1-62703-748-8_7.
3. Bengio, Y. Learning Deep Architectures for AI.FTML. 2009 https://doi.org/10.1561/2200000006.
4. Boureau, A theoretical analysis of feature pooling in visual recognition. *ICML 2010 - Proceedings, 27th ICML* 2010--*Proceedings, 27th I C on Machine Learning*.
5. Cortes,. Support-Vector Networks. *Machine Learning*. 1995. https://doi.org/10.1023/A:1022627411411.
6. Data Mining:Practical Machine Learning Tools&Techniques. In *Data Mining: PMLTT*. 2016. https://doi.org/10.1016/c2009-0-19715-5.
7. Demirel, G. IMAGE resolution enhancement by using discrete and stationary wavelet decomposition. *IEEE Transactions on Image Processing*.2011.https://doi.org/10.1109/TIP.2010.2087767.
8. Di C. G. *COREL*. https://doi.org/10.1145/253168.253199, 1983.https://doi.org/10.1145/253168.253199.
9. Dremin,V. A.Wavelets and their applications. *Uspekhi Fizicheskikh Nauk*.2001.
10. Glorot, Y. ReLU. *AISTATS '11: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*. 2011. https://doi.org/10.1.1.208.6449.
11. Haris, N.Deep Back-Projection Networks for Super-Resolution. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2018. https://doi.org/10.1109/CVPR. 2018.00179.
12. Harrington, Machine Learning in Action. In *Machine Learning*.2012. https://doi.org/10.1007/s10994-011-5249-4.
13. Harris,M. *A Combined Corner and Edge Detector*.2013. https://doi.org/10.5244/c.2.23.
14. Jordan,T. M. Machine learning: Trends, perspectives, and prospects. *Science*. 2015. https://doi.org/10.1126/science.aaa8415.
15. Larochelle, A neural autoregressive topic model. *Advances in Neural Information Processing Systems*.2012.
16. Lavin, Fast Algorithms for Convolutional Neural Networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2016. https://doi.org/10.1109/CVPR.2016.435.
17. Layer, Wavelet transform. In *Studies in Systems, Decision and Control*. 2015. https://doi.org/10.1007/978-3-319-13209-9_5.
18. Martin, A brief history of the "feature detector." *Cerebral Cortex*.1994. https://doi.org/10.1093/cercor/4.1.1
19. Matsui,S. A 200 MHz 13 mm2 2-D DCT Macrocell Using Sense-Amplifying Pipeline Flip-Flop Scheme. *IEEE Journal of Solid-State Circuits*. 1994. https://doi.org/10.1109/4.34042.
20. Memisevic, M.Gated softmax classification. *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*.
21. Mita,. Joint Haar-like features for face detection. *Proceedings of the IEEE International Conference on Computer Vision*. 2005..https://doi.org/10.1109/ICCV.2005.129.
22. Neapolitan, Neural Networks and Deep Learning. In *Artificial Intelligence*. 2018. https://doi.org/10.1201/b22400-15..
23. Pang, Convolution in convolution for network in network. *IEEE Transactions on Neural Networks and Learning Systems*. 2018. https://doi.org/10.1109/TNNLS.2017.2676130.
24. Robert, C.Machine Learning, a Probabilistic Perspective .*CHANCE*. 2015. https://doi.org/10.1080/ 09332480. 2014.914768.
25. Shang, H. Understanding and improving convolutional neural networks via concatenated rectified linear units. *33rd ICML 2016*.
26. Simard, Best practices for convolutional neural networks applied to visual document analysis. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. 2013. https://doi.org/10.1109/ ICDAR.2003.1227801.
27. Simoncelli, Subband Transforms. In *Subband Image Coding*.199. https://doi.org/10.1007/978-1-4757-2119-5_4.
28. Veit, H. M. Processing techniques. In *Electronic Waste: Recycling Techniques*. 2015. https://doi.org/10.1007/978-3-319-15714-6_3.

AUTHORS PROFILE

K. Ramanjaneyulu, is a research scholar of ECE department in JNTUCEK, Kakinada, A.P, India. He received his M.Tech degree from JNTUCEA, Hyderabad in 2008. He has thirteen years teaching experience for undergraduate and post graduates students. His research interests in the area of pattern recognition, machine learning and deep learning.

Dr. Veera Swamy, is a Professor at Vasavi College of Engineering, Hyderabad, Telengana, India. He received M.Tech and Ph.D. degrees from JNTUK, Kakinada in 1999 and 2009 respectively. He worked 10 Years at Bapatla Engineering College, Bapatla. He served as a Principal at QIS College of Engineering and Technology, Ongole from 2010 to 2018. He is having 20 years of teaching experience and 9 years of research experience. He received a grant worth of 12.5 lakhs from AICTE under RPS Scheme during 2013-16 as PI. He executed one MODROBS and one consultancy project. He published 86 research papers in various reputed international journals and conferences. He published one paper in the patent journal. His interesting research areas are Digital Image Processing, Image fusion, Image Compression, image Watermarking, and Networking Protocols.

Ch. Srinivasa Rao, is working as professor of ECE department in JNTUCEV, Vijayanagaram, JNTUK, A.P, India. He has received his Doctor of Philosophy in 2009 from JNTUCEK, Hyderabad. He has nineteen years teaching experience for undergraduate and post graduates students. His research interests in the area of content based image retrieval, Image and video processing, communication systems and deep learning .