Collaborative Multi-Resolution MSER and Faster RCNN (MRMSER-FRCNN) Model for Improved Object Retrieval of Poor Resolution Images

Amitha I C[1], N S Sreekanth[2]
Department of Information Technology
Kannur University, Kannur, Kerala- 670567, India

N K Narayanan[3]
Indian Institute of Information Technology, Kottayam
Valavoor (P.O.) Kottayam-686635 Kerala, India

***Abstract*—Object detection and retrieval is an active area of research. This paper proposes a collaborative approach that is based on multi-resolution maximally stable extreme regions (MRMSER) and faster region-based convolutional neural network (FRCNN) suitable for efficient object detection and retrieval of poor resolution images. The proposed method focuses on improving the retrieval accuracy of object detection and retrieval. The proposed collaborative model overcomes the problems in a faster RCNN model by making use of multi-resolution MSER. Two different datasets were used on the proposed system. A vehicle dataset contains three classes of vehicles and the Oxford building dataset with 11 different landmarks. The proposed MRMSER-FRCNN method gives a retrieval accuracy 84.48% on Oxford 5k building dataset and 92.66% on vehicle dataset. Experimental results show that the proposed collaborative approach outperform the faster RCNN model for poor-resolution conditioned query images.**

Keywords*—Faster RCNN; feature representation; multi-resolution MSER; object detection; object retrieval

I. INTRODUCTION

Object detection and recovery is one of the emerging areas of computer vision. In our daily routine, almost everything is related to object-based and its retrieval. Object retrieval can be applied as a solution to various real-time problems. Retrieving objects present in a given image scene are far more difficult than content-based image retrieval [1]. In content-based image retrieval, the image is recovered as a whole, but in object retrieval, the region of interest (ROI) only retrieved from a database. Various research studies are being conducted in this area through traditional and deep learning-based approaches. On the object retrieval task, the retrieval is mainly based on object-level features. To identify an object location in an image, it is pre-marked with a box called the object's bounding or anchor box. In the initial stage the bounding boxes are drawn for the areas of the essential object [2].

Object recovery is the process of searching for an object in an image from a large image collection or video configuration. Object recovery occurs in two important steps: first, it searches for an image in a large database, and then observes an object with an anchor box. Initially, there were content-based image recovery structures, which were later improved to recover a specific object from a scene utilizing some formal techniques. The fundamental reason for object recognition is to distinguish and find at least one efficient target from a given image or video database. It meticulously incorporates a variety of key technologies, specifically pattern recognition and digital image processing. It has wide application possibilities in different areas such as protection of brand name or logo, detection of defective parts in printed circuit boards (PCB), accident prevention in road traffics [3], alerts about hazardous products in manufacturing plants and military confined region checking [4],[5].

The object recovery process is conventionally settled by physically extricating feature representations, where the normal feature-elements are addressed by histogram-of-oriented gradients (HoG), scale invariant feature transforms (SIFT), Haar-like feature representations and other calculations that depend on grayscale [6]. In addition to the above-mentioned feature extraction procedures, specific object recovery can be performed using support vector machines (SVMs) or AdaBoost algorithms. These conventional feature extraction models are simply ready to recover low level feature components of an image data, such as colour, shape, texture, blobs and edges, and have restrictions in recognizing numerous objects under complex scenes because of their deprived generalization performances. Newer object detection models are mainly depending on deep convolutional neural network (DCNN / Deep conv-net) features. Their results are promising when compared with traditional models [7]. Some of the models based on DCNN are region-based convolutional neural network and its variants, you only look once (YOLO) and single shot multi-box detector (SSD) models. DCNN models are not just concentrating on the detailed surface features from the previous level convolution layer, but on the other hand, can get more significant level data from the next-level convolution layer [8],[9],[10],[11].

In addition to the conventional CNN process, the RCNN variants utilize a counter strategy to assume the target object regions in the feature maps, steadily adjusting the location info and optimize the object's location for categorization and retrieval. Conversely, other object discovery models will concurrently foresee the anchor boxes and categorize straightforwardly in the feature maps by applying diverse convolutional phases. The RCNN model has two activity stages that consider higher location precision, while SSD and YOLO can straightforwardly recognize the arrangement and the position data, which speeds up detection [12]. A faster RCNN model offers better object retrieval accuracy than its

predecessors. The limitation of faster RCNN is that it cannot efficiently recover objects from bad resolution images.

To overcome the limitations of detecting poor / bad resolution images in faster RCNN, the proposed system uses an existing algorithm called multi-resolution MRMSER to formulate a new collaborative MRMSER-FRCNN model that offers better retrieval accuracy, compared to individual faster RCNN or multi-resolution MSER. With this integrated approach, it can achieve better retrieval accuracy on images with poor resolution conditions. The proposed collaborative MRMSER-FRCNN model offers better retrieval accuracy, compared to individual faster RCNN or multi-resolution MSER.

The rest of this paper is arranged as follows. Section 2 gives a detailed study report on various object retrieval techniques, their feasibility and the problems in the existing models. The identification and implementation of collaborative model based on MRMSER and faster RCNN explained in Section 3. Section 4 gives a detailed discussion about the datasets used and the results obtained in the simulation experiment. Finally, Section 5 concludes the proposed model.

II. Related Work

This section provides a detailed review on the object detection and retrieval from images and also explores accessible strategies to summarize in-depth features nearby for creating conservative descriptions for image recovery. Krizhevsky A. et al. [13] achieved better categorization of images in IMAGENET by DCNN. Their study categorizes a large number of high-resolution images. They have designed a neural network consisting of MaxPooling layers and five convolution layers that are fully integrated with SoftMax functionality. An object identification strategy was proposed based on the calculation of a regional proposal by Ren S. et al. in [14]. A region based network (RPN) that expects object boundaries and object simultaneously in each pixel area proposed by Long J. et al. in [15] illustrates a complete CNN, which is remarkably capable of classifying set of images semantically. The primary goal of their study is to create a “fully convolutional network” that receives the variable-sized image inputs and produces equally-sized outcome with viable deductions and information. Here, they configured some recent taxonomic networks, for example, Alex-Net, VGGNet and GoogLeNet for taxonomic purposes. Their fully CNN completes the excellent classification of images. Recovered feature is used as a standard image representation to handle different image object classifications in [16] by Sharif Razavian A.

Hossaine D. et al. suggest a deeper belief NN strategy for object identification. Once the object identification task is done, a live-path is created. The arrangement holds the object and keeps it in a pre-defined place. This deep learning method extracts adequate feature to detect the objects utilizing a computer vision framework [17]. Experimental inferences exhibits that, their proposed model performs better than other strategies. Babenko A and Lempitsky V [18] discussed about different enhancements demonstrated by the image descriptors offered by deep CNN, which greatly enhances image classification and recovery. The convolution layer could be clarified as nearby feature sets that describe image bounds explicitly. Amitha I C and N K Narayanan discuss the state of art about the conventional approaches and their inadequacies for object retrieval in [19]. Amitha I C and N K Narayanan suggested a proficient object recovery method in images utilizing SIFT-RCNN in [20].

Li H. et al. [21] have suggested new recovery method of image objects from an image database. The suggested model utilizes the power of CNNs, they utilized the further developed variant of the faster RCNN. They have tried their technique with various freely accessible databases. Oh I. et al. [22] recommended a strategy for segmenting multi-scale image dependent on maximally stable extremal regions (MSERs). They extended the essential MSERs usefulness (blobs recognition in image) to regular image segmentations [23].

Wang R. et al. [24] have proposed an upgraded, faster RCNN situated on the MSER inference standard for SAR image transport discovery in the harbors. The test result represents excellent recognition, and it distinguishes between their prospective technique and faster RCNN approach. This faster RCNN is utilized for recovering similar objects in various scenes. Faster RCNN is presented to conquer the issues in the fast RCNN model, the previous variant of CNN. Dubey A. K et al. [25] have introduced an efficient way to deal with deformity detecting in a rail route track surfaces utilizing MSERs stamping. Through this strategy the imperfections in the railway track can be assessed without much computation. All the above activities were performed within CNN family to further develop a recovery cycle or MRMSER based calculation to work on bad resolution or improve the lower surface regions in an image. The specific collective method utilizes the remarkable components of faster RCNN and MRMSER computations to recover the queried object with better accuracy from the specified image database.

In this combined model, it inspects the use of the multi-resolution maximally stable extremal regions (MRMSER) estimation to perceive the whole space of an image, regardless of the surface details. Locating blobs in images can be performed effectively through MSER and multi-resolution MSER systems. MSER is generally utilized during edge discovery, which gives better results with the mix of faster RCNN [26],[27],[28],[29],[30]. A layer connection strategy is utilized to perceive objects in low-resolution regions. The suggested method fuses a strategy for layer connection, to distinguish object in lower surface areas.

The main focus of the proposed MRMSER-FRCNN model is to retrieve objects efficiently, even if, the query image is affected due to poor resolution conditions or lower surface areas. Here, improve the poor resolution of the query image by applying specific multi-resolution MSER, and then apply the enhanced image input to the faster RCNN phase, for further object detection and retrieval.

III. Collaborative MRMSER-FRCNN Model

Recovery of objects in images can be done by conventional methods or by deep learning based methods.

With pre-trained CNNs, a large scope image database of high-resolution images can be organized into certain categories [13],[15],[16],[31],[32]. A common problem of not being able to distinguish areas beyond a specific region, were found in both region based CNN (RCNN) [33] and fast region based CNN. As a solution for this issue, the regions proposals networks (RPNs) were presented and joined with the final layer of fast RCNN [34], and the new organization is called as faster RCNN. Faster RCNN model recovery accuracy is lower when the image has lower resolution conditions due to varied features in the image texture [21]. The maximally stable extremal regions (MSERs) method is used to increase the feature extraction capability of faster RCNNs in poor resolution conditions in [35].

Particular object recovery in images is truly challenging in lower resolution conditions. In faster RCNN, the RoI pooling layers uses only the feature-maps of the top (best) convoluted layers, which leads to incorrect movement of feature extractions at lower resolution. Because of this problem, faster RCNN cannot retain the local attributes of an object. The MRMSER algorithm is integrated to solve the problem found in the faster RCNN model for efficient object recovery of the bad resolution state of the query image. The working of this novel collaborative approach is shown in the Fig. 1.

The proposed collaborative model works on the combination of two procedures, a faster RCNN followed by MSER/multi-resolution MSER. The query image is directly applied to the MRMSER or MSER phase. During this phase the MSER algorithm produces such an output image which will help faster RCNN phase detect and retrieve image objects even if, it is from a poor resolution condition or various textured regions. MSER calculations are done through a series of tasks such as grayscale conversion, detecting the MSERs and edge recognition, filtering based on region properties, determination of bounding box region and then produce a segmented image output. The image output from our previous phase is applied as the input to faster RCNN module. The faster RCNN model used in this work consists of 7 convolutional layers marked as conv1, conv2, up to conv7. The first five convolutional layers extract the features from the input image. The last two convolutional layers, conv6 and conv7 are fixed as fully connected layers. The input image is convolved with different filter sizes with varying strides.

The stride is the number of pixels shifts over the input matrix. Apply these convolutional feature map values to region proposal network (RPN) after conv5 layer. After the generation of region proposals, perform RoI pooling and then pass it through the fully connected layer. The output layer contains the bounding box information and the class score values.

Finally, to retrieve the queried objects, perform a ranking by class score generated in the previous stage. The detailed computation of MRMSER and faster RCNN are explained in the following subsections.

A. Multi-Rsolution Maximally Stable Extremal Region (MRMSER) Calculation

MSERs maintain a well-established framework for finding blobs in images. It is utilized to measure the intelligence between image parts of two images with alternate perspectives, providing broader benchmark adjustments. MSER will be utilized for better edge-detection in the proposed object detection framework. The in-depth edge detection ability of MSER overcomes the problems due to poor resolution conditions.

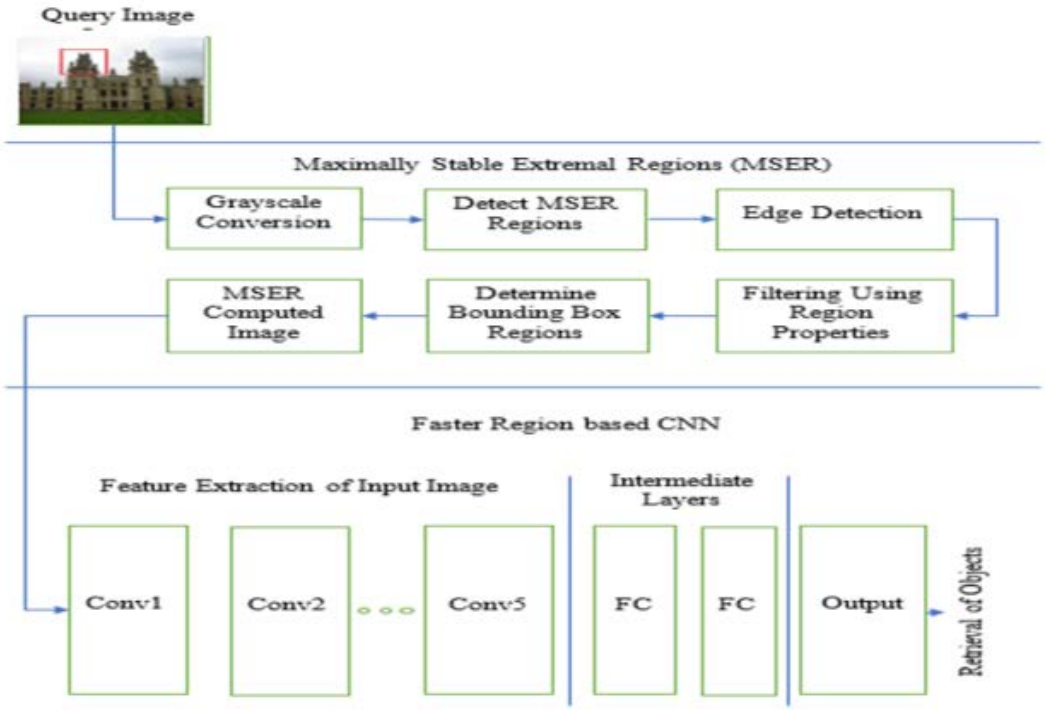

Fig. 1. Collaborative MRMSER-FRCNN Model.

MSER gives fundamental elements of an input image like all other feature detectors. MSER is reliably related fragment of two position sets of input images. This separates the image into several co-variant categories known as MSERs. MSER categories are related with regions describing uniform intensities surrounded by conflicting surroundings. The MSER procedure is described in the subsequent sections. In the initial step perform a grayscale conversion. Two options are there to perform the gray scale transformation in MSER - the average technique and the weighted technique or the luminosity technique. Each technique has its own advantages and disadvantages. Our system uses the weighted technique or the luminosity technique, which can be calculated as in Eq. (1) [24],[36].

$$I_G=0.3R+0.59G+0.11B \quad (1)$$

where I_G is the grayscale transformation and Red, Green and Blue channels are represented by R, G and B.

To guarantee the regions are maximally stable, ought to follow the requirements [35], which shown in Algorithm 1.

Algorithm 1: Algorithm for identification of areas utilizing MSER

Input:

Image I

Delta Parameters: To compute the similarities.

Step 1: For every pixel abbreviated by intensity values.

a. Spot a pixel in an image, when its turn arises.
b. Modify the pattern of the associated parts.
c. Update the region for affected related areas.

Step 2: For every single associated part.

a. Recognize areas with nearby minima concerning the speed of progress of the associated part region with an edge; characterize each such locale as MSE.

Output:

List of nested extremal areas.

Regardless, whether or not an extremely area is maximally consistent, it may be excused if:

- It is excessively large
- It is nearly nothing
- It is extremely shaky.
- It is exorbitantly similar to their parent.

Steps for the execution of MSER extractions are explained in Algorithm 2. Algorithm 3 gives the mathematical formulation of MSERs.

Algorithm 2: MSER extraction steps.

1. Perform the basic brightness of the image and change the intensity range from black to white.
2. Acquire related regions (“Extremal Regions”).
3. Identify a limit when an extremal region is “Maximally Stable”.
4. Estimate the region by an oval (optional).
5. Save those region descriptors as elements.

Algorithm 3: Mathematical Formulation of MSERs

Input:

Image I

I: D⊂Z2 →S

Extremal areas are well defined on image if:

S is completely arranged

A ⊂ D x D

p, q Є D are adjacent(pAq)

iff, $\sum_{i=1}^{d} |p_i - q_i| \leq 1$

∂Q= {q Є D\ Q:]p Є Q}

∂Q is the (Outer) Region Bounds

Extremal Region Q ⊂ D

I (p) > I (q) : Maximum

I (p) < I (q) : Minimum

Q1,….,Qi-1,….,Qi,….be the set of extremal areas/regions.

$Q_i \subset Q_{i+1}$

Q_{i*} is maximally stable

iff $q(i)=|Q_{i+\Delta} \setminus Q_{i-\Delta}| / |Q_i|$

| |: represents the cardinality

Δ Є S is a parameter of the process

To further develop the recovery exactness of the proposed procedure, a variation of MSER calculation is utilized named multi-resolution MSER (MRMSER). An MRMSER can be determined through the accompanying steps, which is shown in Algorithm 4.

Algorithm 4: Algorithm for MRMSER

Step1: Rather finding feature just from the image, generate a scale pyramid with an octave among scales.
Step 2: Identify MSERs exclusively at every resolution.
Step 3: Removes copy MSERs by erasing the best scale MSERs with comparative areas and sizes as MSERs found on the following rough scale.

Scale pyramid can be constructed through obscure and resample by a Gaussian filter.

After computing the MRMSERs of a given query image, generate the output and apply this image as an input to the faster RCNN phase.

B. Faster Region based CNN (FRCNN)

The second phase of the proposed system is based on faster RCNN. A common problem of not being able to distinguish areas beyond a specific region, were found in both region based CNN (RCNN) [33] and fast region based CNN. Faster RCNN model recovery accuracy is low when the image has lower resolution conditions.

FRCNN provides improved detection of analogous objects with the help of MRMSER. The proposed FRCNN consists of seven convolutional layers. The overall implementation task of FRCNN is performed across various convolutional layers. Initially, it receives an input image and pass it to a region proposal network (RPN), which initiates the RPN's task using the anchor boxes. Anchor boxes of different sizes were used depending on the size of the objects. After the generation of anchor boxes, compute intersection over union (IoU) on these bounding blocks / boxes. If IoU$\geq$ 0.5, accept it as the object bounding box and label it as foreground, else reject that box and recognize it as background. At the same time, the first five convolutional layers computes the feature maps and transfer it to RoI Pooling layer, which reduces the size of the feature maps created on the previous layers. A regressor refines the bounding box and classifier categorizes the object.

The real object recovery task is done in the FRCNN stage. The processed and modified image from MRMSER phase is applied as the input to FRCNN phase. The requested query image with poor-resolution condition was corrected by the MSER or MRMSER computation. The different steps associated with the FRCNN stage is described in Algorithm 5.

Algorithm 5: Algorithm for object detection using faster RCNN

1. A queried image is applied as input from MRMSER.
2. Concatenate conv3 and conv5 layers of the prospective system.
3. Finely tune the model.
4. Perform L-2 normalizations, which combines different scales and Norms in conv3 and conv5.
5. Compare the similarity of regional proposals.

a. Confidence score selections is done by setting up a threshold value.

b. According to the similarity score, rank the objects.

IV. Results and Discussion

To evaluate the efficiency of the proposed MRMSER-FRCNN method, the system is being simulated and tested on freely available Oxford buildings and vehicles database.

- Vehicle Dataset: The total number of images in this database is 150, which includes three different types of vehicles: bus, car and motorbike. Each individual vehicle class contains 50 pictures. Out of 150 images, 80% are utilized for training and the remaining 20% for testing.
- Oxford Building Dataset [37]: The total number of images in this dataset is 5062. This dataset provides 11 different landmarks for buildings. Each landmark was given five query images, so they affixed 55 query images to the entire dataset. All query images are marked with bounding boxes of required objects. Here also 80% images are utilized for training and the remaining 20% for testing.

Experimental results shows better object retrieval accuracy than faster RCNN or MSER-FRCNN. To check the retrieval accuracy of poor resolution, such images are tested and results are tabulated. The retrieval accuracy of the proposed method is compared with other similar methods also.

The object retrieval accuracy is calculated as the ratio of correctly retrieved objects and number of total objects in the reference dataset. The proposed MRMSER-FRCNN method gives a retrieval accuracy of 84.48 % in Oxford building dataset and 92.66 % in vehicle dataset. The previous method [35] gives 71.4 % in Oxford buildings dataset and 86 % in vehicle dataset. Faster RCNN shows 63.2 % in Oxford building dataset and 79 % in vehicle dataset. The sample object retrieval from Oxford building dataset is shown in the Fig. 2. The left most image surrounded by red colour box is the query image object and all other pictures surrounded by green colour box are the correctly retrieved image objects.

The detailed result analyses of MRMSER-FRCNN in Oxford building and vehicle dataset are shown in Tables I and II. The result comparison with various methods and proposed method is shown in Table III. Fig. 3 gives the bar chart comparison representation of the newly proposed MRMSER-FRCNN with other methods.

The object retrieval overall accuracy in Oxford building dataset with different methods are listed in Table IV. The proposed model retrieval accuracy is shown in bold values. Their performance comparison is shown in Fig. 4.

Fig. 2. Sample Object Retrieval from Oxford Building Dataset.

TABLE I. OBJECT RETRIEVAL WITH MRMSER-FRCNN IN OXFORD BUILDING DATASET

Land Marks in Oxford dataset	Objects retrieved correctly	Accuracy (%) Oxford Building Dataset -5k
All Souls	125	94.69
Ashmolean	146	81.1
Balliol	140	90.9
Bodleian	174	81.69
Christ Church	450	82.87
Cornmarket	48	81.35
Hertford	55	82.08
Keble	98	80.99
Magdalen	550	80.29
Pitt Rivers	88	81.48
Radcliffe Camera	229	81.20
Overall Accuracy		84.48 %

TABLE II. OBJECT RETRIEVAL WITH MRMSER-FRCNN IN VEHICLE DATASET

Vehicle Class	Vehicles retrieved correctly	Vehicle Dataset Accuracy (%)
Bike	47	94
Bus	46	92
Car	46	92
Overall Accuracy		92.66 %

TABLE III. COMPARISON WITH PROPOSED AND OTHER METHODS

Method	Datasets	
	Vehicle	*Oxford Buildings*
Faster RCNN	79.00	63.20
MSER-FRCNN	86.00	71.40
MRMSER-FRCNN (Proposed Method)	92.66	**84.48**

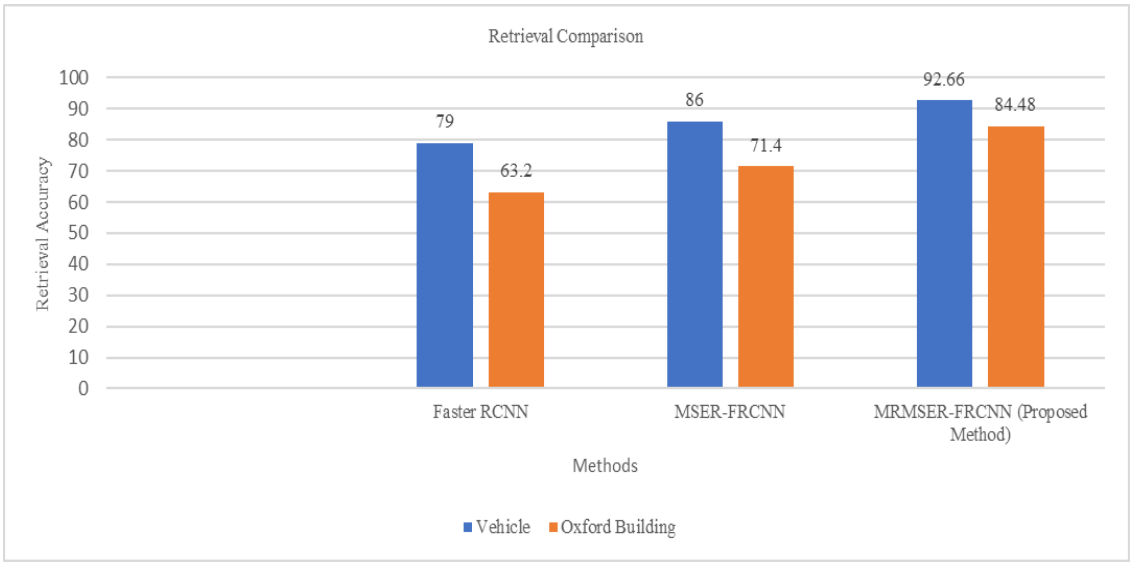

Fig. 3. Comparison with Proposed and other Methods.

TABLE IV. COMPARISON WITH PROPOSED AND OTHER METHODS IN OXFORD BUILDING DATASET

Methods	**Object retrieval Accuracy (%)**
SIFT and CNN [38]	81.60
SIFT and RCNN [20]	82.10
Faster RCNN [21]	63.20
MSER-FRCNN [35]	71.40
MRMSER-FRCNN (Proposed Method)	**84.00**

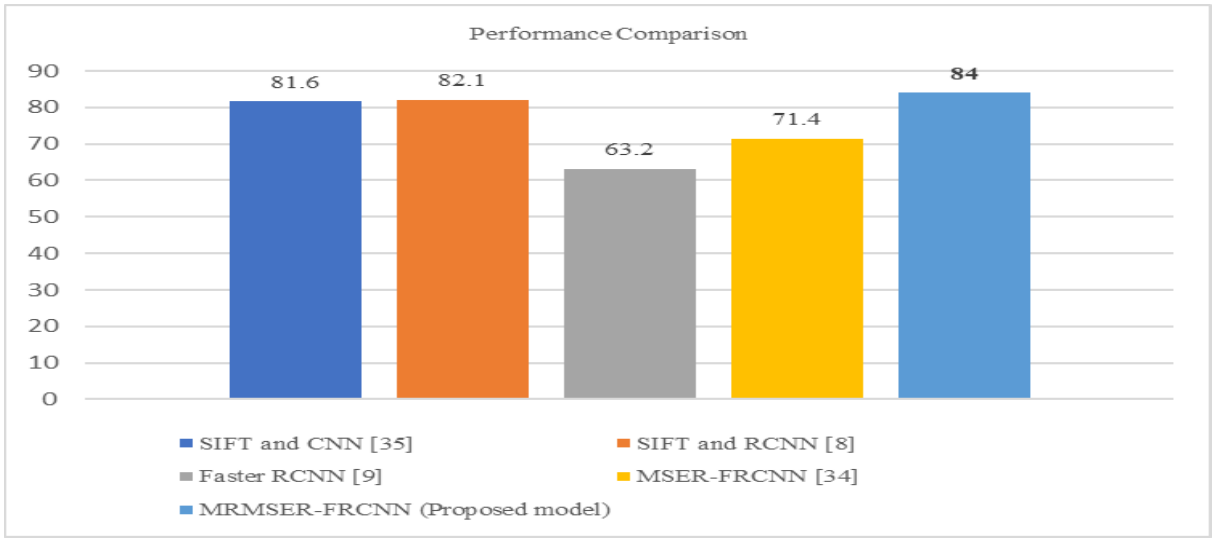

Fig. 4. Comparison with Proposed and other Methods in Oxford Building Dataset.

V. CONCLUSION

In this paper, a new collaborative approach to recovering objects from poor resolution images was proposed. The proposed MRMSER-FRCCN method can overcome the issue with the collaboration of MRMSER strategy. This will take care of smaller objects presented in the image with poor lighting conditions or poor resolution. Experimental results are obtained for collaborative method on two different publicly available datasets. The retrieval accuracies were compared with other individual as well as combined methods. The proposed MRMSER-FRCNN method gives a retrieval accuracy of 84.48% in Oxford building dataset and 92.66% in vehicle dataset. The previous works reported only a maximum

of 71.4% in Oxford building dataset and 86% in vehicle dataset. Hence it is found that the proposed MRMSER-FRCCN method outperforms conventional methods reported in literature.

REFERENCES

[1] Han, Xian-Feng, Hamid Laga, and Mohammed Bennamoun. "Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era." IEEE transactions on pattern analysis and machine intelligence 43, no. 5 (2019): 1578-1604.

[2] Puranik, Vaishali, and A. Sharmila. "Integration of Basic Descriptors for Image Retrieval." In International Conference on Information Management & Machine Intelligence, pp. 629-634. Springer, Singapore, 2019.

[3] Shine, Linu, and C. Victor Jiji. "Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN." Multimedia Tools and Applications (2020): 1-21.

[4] Liu, Jin, Yihe Yang, ShiqiLv, Jin Wang, and Hui Chen. "Attention-based BiGRU-CNN for Chinese question classification." Journal of Ambient Intelligence and Humanized Computing (2019): 1-12.

[5] Cao, Danyang, Menggui Zhu, and Lei Gao. "An image caption method based on object detection." Multimedia Tools and Applications 78, no. 24 (2019): 35329-35350.

[6] Hu, Rui, and John Collomosse. "A performance evaluation of gradient field hog descriptor for sketch-based image retrieval." Computer Vision and Image Understanding 117, no. 7 (2013): 790-806.

[7] He, Xinwei, Yang Zhou, Zhichao Zhou, Song Bai, and Xiang Bai. "Triplet-center loss for multi-view 3d object retrieval." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1945-1954. 2018.

[8] Amitha, I. C., and N. K. Narayanan. "Improved Vehicle Detection and Tracking Using YOLO and CSRT." In Communication and Intelligent Systems, pp. 435-446. Springer, Singapore, 2021.

[9] Amitha, I. C., and N. K. Narayanan. "Object Detection Using YOLO Framework for Intelligent Traffic Monitoring." In Machine Vision and Augmented Intelligence—Theory and Applications, pp. 405-412. Springer, Singapore, 2021.

[10] Tasnim, Zarrin, FM Javed Mehedi Shamrat, Md Saidul Islam, Md Tareq Rahman, Biraj Saha Aronya, Jannatun Naeem Muna, and Md Masum Billah. "Classification of Breast Cancer Cell Images using Multiple Convolution Neural Network Architectures." cancer 12, no. 9 (2021).

[11] Methun, Naimur Rashid, Rumana Yasmin, Nasima Begum, Aditya Rajbongshi, and Md Ezharul Islam. "Carrot Disease Recognition using Deep Learning Approach for Sustainable Agriculture."

[12] Cao, Danyang, Zhixin Chen, and Lei Gao. "An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks." Human-centric Computing and Information Sciences 10, no. 1 (2020): 1-22.

[13] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60, no. 6 (2017): 84-90.

[14] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster RCNN: Towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 39, no. 6 (2016): 1137-1149.

[15] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440. 2015.

[16] Sharif Razavian, Ali, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. "CNN features off-the-shelf: an astounding baseline for recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 806-813. 2014.

[17] Hossain, Delowar, GenciCapi, and Mitsuru Jindai. "Object recognition and robot grasping: A deep learning based approach." In The 34th Annual Conference of the Robotics Society of Japan (RSJ 2016), Yamagata, Japan. 2016.

[18] Babenko, Artem, and Victor Lempitsky. "Aggregating local deep features for image retrieval." In Proceedings of the IEEE international conference on computer vision, pp. 1269-1277. 2015.

[19] Amitha, I. C., and N. K. Narayanan. "Image object retrieval using conventional approaches: a survey." Int J Eng Technol Sci (IJETS) (2018): 1-4.

[20] Amitha, I. C., and N. K. Narayanan. "Object Retrieval in Images using SIFT and RCNN." In 2020 International Conference on Innovative Trends in Information Technology (ICITIIT), pp. 1-5. IEEE, 2020.

[21] Li, Hailiang, Yongqian Huang, and Zhijun Zhang. "An improved faster RCNN for same object retrieval." IEEE Access 5 (2017): 13665-13676.

[22] Oh, Il-Seok, Jinseon Lee, and Aditi Majumder. "Multi-scale image segmentation using MSER." In International Conference on Computer Analysis of Images and Patterns, pp. 201-208. Springer, Berlin, Heidelberg, 2013.

[23] Girshick, Ross. "Fast RCNN." In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448. 2015.

[24] Wang, Rufei, Fanyun Xu, Jifang Pei, Chenwei Wang, Yulin Huang, Jianyu Yang, and Junjie Wu. "An improved faster RCNN based on MSER decision criterion for SAR image ship detection in harbor." In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 1322-1325. IEEE, 2019.

[25] Dubey, Ashwani Kumar, and ZainulAbdinJaffery. "Maximally stable extremal region marking-based railway track surface defect sensing." IEEE Sensors Journal 16, no. 24 (2016): 9047-9052.

[26] LeCun, Yann, YoshuaBengio, and Geoffrey Hinton. "Deep learning." nature 521, no. 7553 (2015): 436-444.

[27] Zhou, Wengang, Houqiang Li, and Qi Tian. "Recent advance in content-based image retrieval: A literature survey." arXiv preprint arXiv:1706.06064 (2017).

[28] Zhao, Zhong-Qiu, Peng Zheng, Shou-tao Xu, and Xindong Wu. "Object detection with deep learning: A review." IEEE transactions on neural networks and learning systems 30, no. 11 (2019): 3212-3232.

[29] Liu, Li, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. "Deep learning for generic object detection: A survey." International journal of computer vision 128, no. 2 (2020): 261-318.

[30] Jo, YoungJu, Hyungjoo Cho, Sang Yun Lee, Gunho Choi, Geon Kim, Hyun-seok Min, and YongKeun Park. "Quantitative phase imaging and artificial intelligence: a review." IEEE Journal of Selected Topics in Quantum Electronics 25, no. 1 (2018): 1-14.

[31] Yim, Junho, Jeongwoo Ju, Heechul Jung, and Junmo Kim. "Image classification using convolutional neural networks with multi-stage feature." In Robot Intelligence Technology and Applications 3, pp. 587-594. Springer, Cham, 2015.

[32] Jaswal, Deepika, S. Vishvanathan, and S. Kp. "Image classification using convolutional neural networks." International Journal of Scientific and Engineering Research 5, no. 6 (2014): 1661-1668.

[33] Fontdevila Bosch, Eduard. "Region-oriented convolutional networks for object retrieval." Bachelor's thesis, UniversitatPolitècnica de Catalunya, 2015.

[34] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.

[35] Amitha, I. C., and N. K. Narayanan. "Collaborative MSER and Faster R-CNN Model for Retrieval of Objects in Images." In Soft Computing for Problem Solving, pp. 673-682. Springer, Singapore, 2021.

[36] Cao, Changqing, Bo Wang, Wenrui Zhang, Xiaodong Zeng, Xu Yan, Zhejun Feng, Yutao Liu, and Zengyan Wu. "An improved faster RCNN for small object detection." IEEE Access 7 (2019): 106838-106846.

[37] Philbin, James, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. "Object retrieval with large vocabularies and fast spatial matching." In 2007 IEEE conference on computer vision and pattern recognition, pp. 1-8. IEEE, 2007.

[38] Zhang, Guixuan, Zhi Zeng, Shuwu Zhang, Yuan Zhang, and Wanchun Wu. "Sift matching with CNN evidences for particular object retrieval." Neurocomputing 238 (2017): 399-409.