

Fine Grained Food Image Segmentation through EA-DCNNs

Vishwanath.C.Burkapalli, Priyadarshini.C.Patil

Abstract:-The recognition of Indian food can be considered as a fine-grained visual recognition due to the same class photos may provide considerable amount of variability. Thus, an effective segmentation and classification method is needed to provide refined analysis. While only consideration of CNN may cause limitation through the absence of constraints such as shape and edge that causes output of segmentation to be rough on their edges. In order overcome this difficulty, a post-processing step is required; in this paper we proposed an EA based DCNNs model for effective segmentation. The EA is directly formulated with the DCNNs approach, which allows training step to get beneficial from both the approaches for spatial data relationship. The EA will help to get better-refined output after receiving the features from powerful DCNNs. The EA-DCNN training model contains convolution, rectified linear unit and pooling that is much relevant and practical to get optimize segmentation of food image. In order to evaluate the performance of our proposed model we will compare with the ground-truth data at several validation parameters.

Keywords: Deep Convolutional Neural Networks (DCNNs); Rectified Linear Unit (ReLU); Edge Adaptive (EA); Pooling; Convolution.

I. INTRODUCTION

A food is very essential part of our day today life, food has a directly effect on the human body. Whereas the ingredient, food shape, cooking recipes and flavours are specific for each cuisines. There has been a remarkable progress in the field of computer technologies that allows the food recognition, but still there are lots of difficulty present at that even for the human to identify the food. The cuisine contains contextual and prior information, which can be used to build the automatic model for food analysis. In this study, we are going to review some recent study regarding of food analysis model and afterwards we will proposed an optimized model for food segmentation. The user comment on a particular food can be used to provide an interesting insight to understand the eating habits, cuisines and cultures of food [1]. However, the collective information can be used by the automated model to provide better-classified and segmented accuracy [2].

Revised Manuscript Received on November 05, 2019.

Vishwanath.C.Burkapalli, Dept. Of Information Science and Engineering, PDA College of Engineering, Kalaburgi, Karnataka, India
Email-vishwa_bc@rediffmail.com

Priyadarshini.C.Patil, Dept. Of Computer Science and Engineering, PDA College of Engineering, Kalaburgi, Karnataka, India
Email-priyadarshinicpatil@gmail.com

The recognition of Indian food seems to be the fine-grained category of visual recognition, whereas the similar category can provide the significant amount of variability at recognition stage [3]. So the efficient method of classification and segmentation is needed to predict the particular and enhanced analysis. The restaurant recipe

contains the various numbers of recipes that can be only exploded due to dishes name present in menu card. The similar dish with different appearance style (i.e., restaurants, cooking style, and presentation) can create variability for recognition of food items. This confirms the long tail due to appearance style and rare dishes with limited number of training data causes difficulty at visual appearance and to build the recognition system.

In general, the segmentation approach consist only the spectral information which causes unfortunate information variability for the same target class [4], which also results an unsatisfactory performance of model. Considering this kind of difficult problem, a segmentation of food images faces the key challenges in practice, therefore, some method can be considered to tackle this type of issues. To develop a model using both spectral and spatial information will be an efficient way to design, which provides additional information regarding ingredient, size, shape and several other kind of important features [5]. Advancement in the field multi-layer neural network (NN) will provide huge variety of machine learning jobs like as regression and classification, also provide independency during training and testing phase. Deep network type model are able to extract more abstract data features, which provide more ability to get enhanced classification accuracy with respect to the other traditional classifiers [6]. A convolutional neural networks (CNN) has adopted widely for the image processing task and majorly used for labeling problems at the pixel-level among some deep learning approaches. Therefore, considering a CNN model allows the good representation of features and provides end-to-end labeling [7].

CNN provide an effective way to determine the spatial features under the surrounding input data patches, that is why the outcome maps of classification become smoother, which is not possible for other approach during indirectly modeling. The local minima reaching possibility during training phase of CNN with presence of noise at the input side may cause for development of isolated area in classification map. While comparing with other models of machine learning, CNN is limited by the absence of constraints such as shape and edge that causes output of segmentation to be rough on their edges. Due to this limitation of CNN, a post processing stage is needed after the segmentation through CNN to get optimal output. In this study, we proposed an EA approach which will model the required features such as; connectivity, shape, region and contextual information. The EA is directly formulated with the DCNNs approach, which allows training step to get beneficial from both the approaches for spatial data relationship. Therefore, the EA will help to get better-refined output after receiving the features from powerful

DCNNs. For the computer applications, it is required to train the model with huge samples values, or else it face the over fitting that is representation capability may not able to work at test data. It is required to increase the training sample size in order to reduce the issue of over fitting. In our approach, each of the input images are sub-divided that called as patches and this patches are able to get more accurate and effective structural data description. The proposed EA based DCNNs model start developing the feature coarse map, which get through DCNN model and afterwards the EA is performed to build the final segmented image. In order to get the significant extent, the EA-DCNN training model contains convolution, rectified linear unit (ReLU) and pooling that is much relevant and practical to get optimize segmentation of food image. To evaluate the performance of our proposed model we will compare with the ground-truth data at several evaluation parameters in the result analysis section.

II. LITERATURE SURVEY

This considered approach is linked with the earlier work in the area of image segmentation, so in this section, we briefly review before introducing our proposed approach. The object segmentation has been very interesting field for the researchers in computer vision analysis from a long time, while the traditional methods have been mostly based upon the feature-handcrafted model. Therefore, we can sort the mostly of traditional method into the object segmentation that based upon the texture, color and low-level appearances [8]. Also to the higher-level object information such as; global image information gain and object detectors [9] [10], whereas the methods are also present which contains both the low-level and higher-level information that we called it as hybrid methodologies [11]. The hybrid type of approaches works well in the several scenarios for what they have designed, but in some complex cases it generally break down [12] [13]. In some cases, the color of image may be affected through the conditions of lightning and local features of contrast may scuffle with the homogeneous areas. CNNs have gained a significant amount of attention due to its advances in the image segmentation area or classification process, CNN has ability to acquire high-level complex information without need of any handcrafted features design [14].

In order to perform more precise and robust image object segmentation, some methods have been recently applied for job of salient object segmentation through rephrasing the segmentation task as the pixel-wise binary classification. Several methods has utilized the patch based technique [15] [16], where the image patches is presented with CNN and this trained in order to classify central super-pixel or the pixel of a considered patch. These type of approaches generally have inefficient in terms of memory requirement and computation, so the approaches have to accomplish a forward authorization for each and every pixel in an image. The performance of traditional non-deep learning approaches illustrating the CNNs potential for the segmentation of object, the fully end-to-end CNN has performed segmentation of pixel-to-pixel in a sole forward authorization, in ordered to replace the patch based methods.

This type of network can be considered as the design of encoder and decoder, whereas the raw image is generally mapped under a minor resolution representation, afterwards it mapped into a raw architecture through using fractionally dependent convolutions [14]. Where the networks count enhance to design an end-to-end trainable model design, which can perform segmentation and extract features for complete color image. Because of the advantages at patch based methods both computational and performance efficiency enhancement is performed in some resent image segmentation study [12], [17].

Afterwards, in order to optimize the performance related to image segmentation, some of the extra components have added in to the model design; study in [17] has added the recurrent type neural network at the design implementation to fine tune the object segmentation at some necessary time steps. In paper [12], considered the two-stream design, where one-stream design have a fully CNN, which generates pixel wise mask. On other side, the second stream design performs the segment level segmentation at the original image, which has segmented towards the super-pixels. Considering both the masks, the fully connection of random fields can be merge to the streams in order to enhance spatial coherence. The method of adversarial training has been considered in [18], the training of discriminator is done to differentiate between the ground truth and predicted outcome. The loss of discriminator and segmentation can be overcome through minimizing the high order uncertainties and inconsistencies at the ground truth and prediction.

In paper [19], proposed a DSS approach that performs a skip-layer design based upon the nested edge detector prototype [20] through implementing the short linkage. Moreover, in paper [21] introduced a multi-tasking design to forecast the salient objects number and to map object using adaptive layer of weight which encodes the mathematical features. In paper [22], uses a low level based handcrafted features in order to complement higher-level CNN features to enhance the performance and in [23] studied the block wise recurrent network to enhance object localization. Moreover, the used refinement system in order to improve the boundary estimation, furthermore a contextual based pixel-wise system can be considered to learn more about the local and global contextual maps that tends to integrate more context areas information for the final segmentation. In [19], proposed MSR network that consist the information at several scales, which can be useful for classification and segmentation task and it also make use of a multiple scale fine-tuned network that get importance through integrating different scales. Afterwards, this can be enhanced by the fine-tuning of MSR network for contour detection; multiple scale alliance is then performed to convert these types of contours to salient segmentation proposals.

III. PROPOSED METHODOLOGY

Here, we will provide an end-to-end Deep-CNN network, where input color images consists of three channels (i.e., individual channel for each color) and the obtained output images will be at the form of binary segmented maps.

In order to reduce the dimension, here we use conventional layer and at last, Softmax function is used to modify the neural network results to class problem. Which allow unremitting segmentation of large images by the overlap-tile strategy.

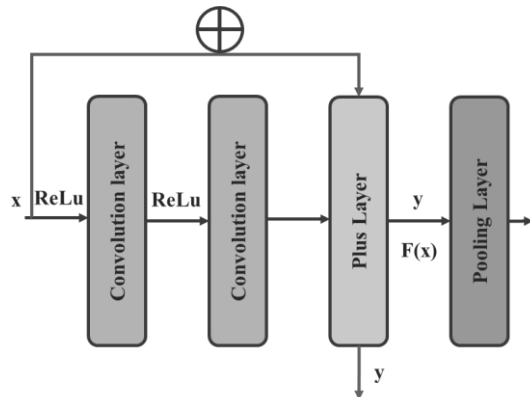


Figure III.1: Design of Down-Block

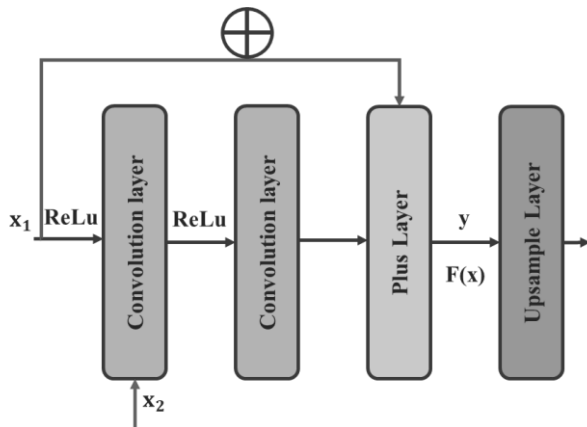


Figure III.2: Design of Up-Block

Here, the proposed outlines structure contains the two blocks known as down and up blocks. Figure III.1 shows the architecture of Down-Block and, fig III.2 shows the architecture of Up-Block, where the model structure is symmetrical. At each of the up and down blocks, two convolutional layers are considered and concatenated through the ReLu layer. Also the plus layer is taken after convolutional layers that shown in grey box, which are considered to optimize the structure and allows to solve error loss problem whenever the network goes deep. This process allows to avoid the training step convergence for the solution of local optimum, so it guarantee the deep learning system to accomplish the good performance at a complex image segmentation process. The output of plus layer is considered to be y , x considered to be the input of Down-Block, thus the plus layer authorize x and provide the y optimized outcome to max pooling layer. However, the up-block model structure is similar to the model of Down-Block, but this uses to accumulate much precise output. The model of deep CNN allows to get information of hierarchical features from raw input data, whereas the training samples consists of image patches, which further turned into features vectors in order to fit input layer at considered system. The convolutional task is been performed by using kernels on the spectral cubes, the common exercise is to convolve the spatial neighbourhood along with kernels under all spectral channels. Considering

the input image that have the A number of bands can be sub-partition into many images with the neighbourhood bands B (whereas $B \ll A$). Convolutional filters is provided to individual exclusive band group of image, so these bands groups can provide much relevant spectral and spatial information that tends to unique wavelength ranges to deliver distinct result at the final segmented image. Therefore, the repeated convolution operation produces multiple number of feature maps with spectral cubes, so here (x, y) provides a location at the spatial dimension, z indicates a band index at the spectral dimension, therefore the (x, y, z) position value on the D^{th} feature mapis given as;

$$PV_{cn(d)}^{xyz} = f \left(\sum_{h=1}^g \sum_{l=0}^{L_c-1} \sum_{k=0}^{K_c-1} \sum_{m=0}^{M_c-1} P_{chn}^{lkm} PV_{(c-1)hn}^{(x+l)(y+k)(z+m)} + a_{cn} \right) \quad (1)$$

The current layer is given by c , the feature map number is represented by g in $(c - 1)$ th layer (formerly considered layer), n denotes the current kernel, h shows the currently feature map at $(c - 1)$ th layer which connected through a feature map layer of $c - th$. P_{chn}^{lkm} is $(l, k, m) - th$ kernel associated value to the $h - th$ feature map at preceding layer. K_c and L_c symbolizes the width and height of kernel, M_c indicates the kernel size with spectral dimension.

An each of the feature map are considered to be independent, so the PV_{chn}^{xyz} is calculated by convolving feature map of preceding layer along with the current layer of kernel. Whereas in this procedure, the number of feature maps at preceding layer will multiply by the number of kernel in current layer. This generates several feature maps at the result of $c - th$ convolutional layer, so the convolutional method can realm the spectral information of input data.

The transitional situated feature maps going over the pooling layers and activation function, so that the feature maps comprise of data cubes are then converted to the feature vectors. Moreover, this feed forward towards a fully end-to-end network layer, which extracts the preceding learned features of the deep neural network. Afterwards, the output are given to a linear classifier such as Softmax function [24] to predict the required classification result, therefore all network is now trained through using the back propagation algorithm [25]. The DCNN trains the Q_R learned parameter that contains distinct information of the all band with $R \in A$ spectral channel. So on the subsequent feature map is used to formulate the proposed deep EA prototype.

3.1 EA-DCNN

DCNN is applied to our raw data, initially we got segmentation outcome from the feature map, which generated through the DCNN, but it is limited by the absence of shape and edge constraints. That results the rough segmented outcome on the edges. Though considering these conditions, a post-processing phase should be done after the DCNN segmentation to obtain a better-refined result. Therefore, we proposed a complete DCNN and EA model, which will be used to get the important characteristics like as, shape, contextual information between the regions, region, connectivity, and etc.

The proposed EA based model is suitable for the neighborhood voxel, so it can be appropriate for food image data. However, DCNN has used several bands, so EA-DCNN will consider the complete spectral channels at the input side of network. The initial feature map comprises a powerful depiction of local features (i.e., spectral and, the spatial features) with the different values of wavelength.

The nodes, which present at EA prototype corresponds to individual voxel in the feature map with B number of bands, the $c \in Y$ represents voxels labels. The edges are created in between the nodes, so it can provide pair-wise association at the neighborhood voxels in EA prototype by connecting one node to the other existing neighborhood nodes. So the EA model can be written as;

$$L(c|s_{(w,R)}; Q_R) = \frac{1}{Z(s_{(w,R)})} e^{-E(c|s_{(w,R)}; Q_R)} \quad (2)$$

The network parameter Q with the different wavelengths R is to be learned, the $E(c|s_{(w,R)}; Q_R)$ represents the energy function which used to verify the compatibility with s input voxel. The s is defined through spatial coordinates $w = \{x, y\}$ with R spectral area and specific predicted label denoted by c . The partition function represented by $Z(s_{(w,R)})$;

$$Z(s_{(w,R)}) = \sum e^{-E(c|s_{(w,R)}; Q_R)} \quad (3)$$

In order to get the essential contextual information, it is important to build the voxels relationships at EA model. So that the energy function can be given as;

$$E(c, s_{(w,R)}) = \sum_{\substack{l \in G \\ R \in A}} T(c_l, s_l; Q_R) + \sum_{\substack{(k,l) \in V \\ R \in A}} U(c_l, c_k, s_l, s_k; Q_R) \quad (4)$$

Here, unary potential function T is computed for each voxels and pair wise function U is determined with respect of compatibility at surrounded voxels.

In order to compute the function of unary potential, we generate a feature maps under a fully linked layer. Where the final outcome is produced for each potential function with the voxel and R , and to compute unary potential value for individual voxel that shows nodes in EA graph can be given as;

$$T(c_l, s_l; Q_R) = -\log L(c_l|s_l; Q_R) \quad (5)$$

Where, it is important to identify the Q_R network parameters that are accustomed at first CNN in accordance to wavelength R for several group of bands, which may be deliver much important information regarding the data.

The pairwise functions can be computed through considering the adjustment between voxels pairs and all possible matches, therefore the function of pairwise potential can be given as;

$$U(c_l, c_k, s_l, s_k; Q_R) = \beta(s_l, s_k) \eta_{l,k,c_l,c_k}(f_l, f_k; Q_R) \quad (6)$$

Where, $\beta(\cdot)$ shows the function label compatibility that encodes the estimated probability of voxel pairs (s_l, s_k) , which labeled as (c_l, c_k) through taking account of possible pairs combinations. The η_{l,k,c_l,c_k} is the CNNs output value that applied towards the node pairs and described through their corresponding feature vectors such as f_l, f_k that previously computed through initial CNN. Whereas, the Q_R function consist of CNN parameters, which required to be acknowledge for pairwise function with respect of complete R spectral channels. Initially, the CNN is applied to input image in order to provide feature vector for the each voxels at feature map, so that feature edge can created through concatenating two neighboring voxels feature vectors.

Table III.1: Deep Edge Adaptive Algorithm

1:	Input: feature map vectors are generated from the DCNN approach, then G voxels at $\{s_1, s_2, \dots, s_G\}$
2:	While s in G do
3:	Add s in EA graph
4:	While at each (s_h, s_n) do
5:	If considered s_h is contiguous to the s_n then
6:	Associate the edge in between s_h and s_n in EA
7:	End if
8:	End while
9:	End while
10:	While s in G do
11:	Estimate unary potential function T by (5)
12:	End while
13:	While at each (s_l, s_k) in V do
14:	Estimate pairwise potential function U by (6)
15:	End while
16:	Estimate the two-kernel potential function
17:	Train EA model
18:	Perform mean-field-inference process
19:	Output: Segmented Regions

Initially, the first step in soft-max function initialized to be the unary potential at all labels for the individual performed voxels, the next step is used to pass the data which applies convolutional through defined Gaussian kernels at the present predicted voxels.

Which show the robust correlation between voxels s_i and s_k , that are related to other voxels and V is number of sub-cubes at R . Back propagation is considered to calculate the derivative errors at filter replies; afterwards it's used to calculate the weighted summation of filtered output for individual voxels label. The label is considered after the typical filter convolution and it is reformulated with output labels and, input voxels.

Table II.1 shows the proposed deep EA algorithm, which shows all the important stages that present in deep EA model, where the input is taken from the DCNN approach, and convolutional network is function ate during pairwise computation of EA, also it produce the optimized outcome at the last segmentation stage. The error computation can be accomplished easily because of outputs and input that are knows at the back-propagation operation, which also allows the automated filter weights gain. Subsequently, the step of compatibility is performed that followed through assimilating real unary potential for the each voxel acquired by the primary DCNN. Afterwards, the step of iterated normalization can be considered as the diverse soft-max function that provides the final output of segmented image.

IV. RESULTS AND ANALYSIS

Initially, our segmented algorithm EA-DCNNs is used to train the model, which consist of 80 dishes that taken from yumly_66K/Indian dataset [26]. Afterwards, when the model is completely build using 80% of training data, then we tested with remaining 20% of data and the segmentation of code is executed in visual studio 2017 with the configuration of Windows 10 OS, Intel i5 processor and 8GB DDR4 RAM. Here, we used Python scripting language, which provide rich number of libraries. We manually mark the boundary of ground-truth of each dish image. Figure IV.1 shows the segmented outcome of recipes, where 1st row the selected input data for testing, 2nd row shows the intermediate result in that red color is superimposed and lastly the 3rd row shows the segmented output from our EA-DCNNs approach. The ground-truth images $X = \{x_1, x_2, \dots, x_\omega\}$ is compared with the obtained segmented food region $Y = \{y_1, y_2, \dots, y_\omega\}$, where ω is the number of testing set and can be used to evaluate the result performance by our proposed approach. In order to validate the performance, several evaluation parameters has considered such as F-score value, precision, specificity, accuracy, Jaccard score and Dice Coefficient [27]. In the experiment, initially we resized the image size to 512x512. The Dice Coefficient can be given as;

$$DC = 2(|X \cap Y|) / (|X| + |Y|) \quad (6)$$

Also, the Jaccard score can be given as;

$$JS = (|X \cap Y|) / (|X \cup Y|) \quad (7)$$

Figure IV.2 shows the F-Score value at various recipes. Figure IV.3 shows for the precision value at various recipes. Similarly, bar chat analysis is done by using specificity and accuracy performance evaluation parameters in fig IV.4 and fig IV.5. Furthermore, we computed Jaccard score and dice coefficient that shows in fig IV.6 and fig IV.7, where blue box shows the standard deviation and red line shows the

mean value. Table4.1 shows the evaluation parameters values at different recipes.



Figure IV.1: Segmented outcome of Recipes

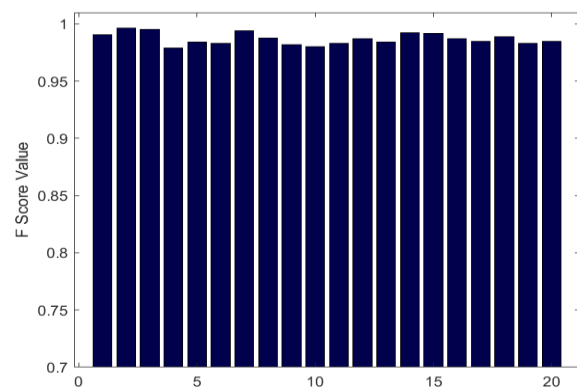


Figure IV.2: Estimated F-Score Value

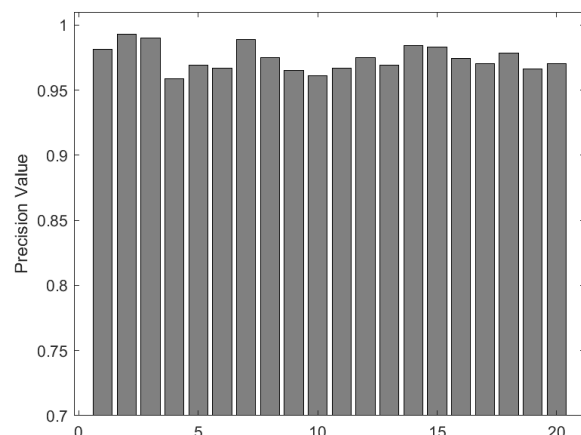


Figure IV.3: Estimated Precision value

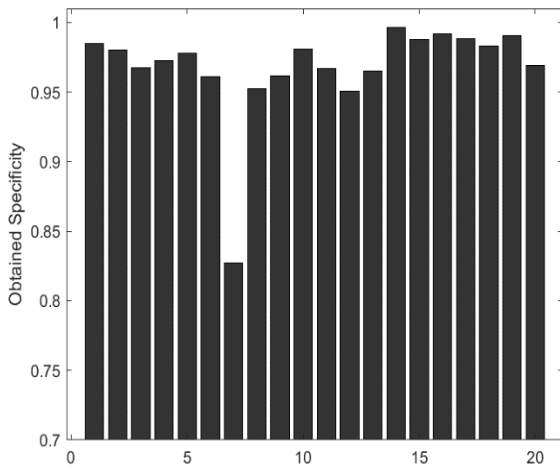


Figure IV.4: Obtained Specificity value

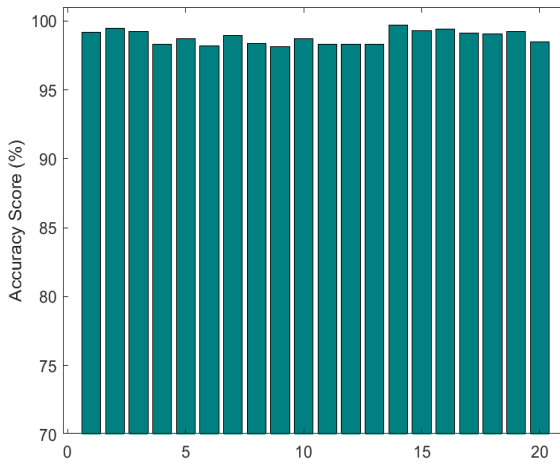


Figure IV.5: Accuracy score (%)

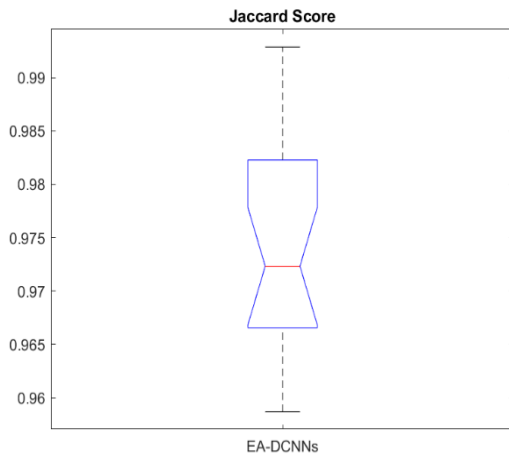


Figure IV.6: Estimated Jaccard Score

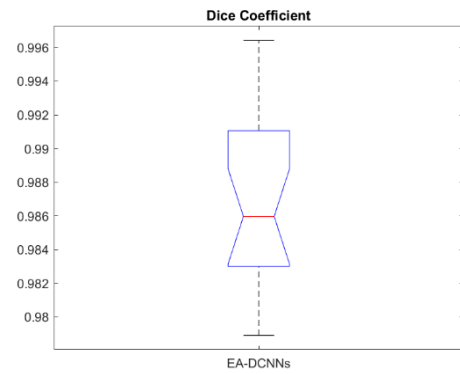


Figure IV.7: Estimated Dice Coefficient Value

TableIV.1: Evaluation Parameters Values at different Recipes(R)

Sl.No.	Accur acy	Specifi city	Precis ion	F Score	DI	JA
R-1	0.9916 08	0.9849 58	0.9813 7	0.990 597	0.990 597	0.981 37
R-2	0.9947 2	0.9801 54	0.9928 58	0.996 416	0.996 416	0.992 858
R-3	0.9924 39	0.9674 7	0.9902 46	0.995 099	0.995 099	0.990 246
R-4	0.9833 37	0.9728 31	0.9586 91	0.978 91	0.978 91	0.958 691
R-5	0.9870 19	0.9780 09	0.9692 88	0.984 405	0.984 405	0.969 288
R-6	0.9816 25	0.9608 11	0.9665 59	0.982 995	0.982 995	0.966 559
R-7	0.9892 65	0.8269 69	0.9886 85	0.994 31	0.994 31	0.988 685
R-8	0.9833 87	0.9520 74	0.9752 04	0.987 446	0.987 446	0.975 204
R-9	0.9812 74	0.9613 33	0.9649 6	0.982 168	0.982 168	0.964 96
R-10	0.9868 55	0.9805 47	0.9610 4	0.980 133	0.980 133	0.961 04
R-11	0.9830 86	0.9668 99	0.9665 68	0.983	0.983	0.966 568
R-12	0.9830 4	0.9508 02	0.9747 7	0.987 224	0.987 224	0.974 77
R-13	0.9832 69	0.9650 24	0.9689 22	0.984 216	0.984 216	0.968 922
R-14	0.9969 52	0.9962 33	0.9842 77	0.992 076	0.992 076	0.984 277
R-15	0.9927 75	0.9874 92	0.9831 82	0.991 52	0.991 52	0.983 182
R-16	0.9938 09	0.9919 07	0.9743 28	0.986 997	0.986 997	0.974 328
R-17	0.9913 83	0.9880 01	0.9703 28	0.984 941	0.984 941	0.970 328
R-18	0.9904 71	0.9831 89	0.9784 74	0.989 12	0.989 12	0.978 474
R-19	0.9925 88	0.9905 79	0.9664 13	0.982 92	0.982 92	0.966 413
R-20	0.9846 92	0.9694 18	0.9702 6	0.984 905	0.984 905	0.970 26
Aver age Value	0.9881 8	0.9677 35	0.9743 21	0.986 97	0.986 97	0.974 321

V. CONCLUSION

In this paper, proposed a complete EA-DCNN approach that will provide an important characteristic to optimized segmented result. Initially, the feature vectors are obtained from the DCNN approach and it given as input to EA for final process. The convolutional network is function ate during pairwise computation of EA and it also produces the optimized outcome at the last stage of segmentation. The error computation can be accomplish easily because of outputs and input that are acknowledge through the back-propagation operation, which also allows the automated filter weights gain. Here, we manually mark the boundary of ground-truth of each dish image, the ground-truth images is compared with the obtained segmented food region. In order to evaluate the result performance by our proposed approach, several validation parameters has considered that shows in the result analysis section. Though our segmentation approach has performed considerably well at each considered recipe and, further we can do identification and classification analysis based upon the obtained segmented outcome.

REFERENCES

1. W. Min, B. K. Bao, S. Mei, Y. Zhu, Y. Rui, and S. Jiang, "You are what you eat: Exploring rich recipe information for cross-region food analysis," *IEEE Transactions on Multimedia*, vol. PP, no. 99, p. 1, 2017.
2. W. Min, S. Jiang, J. Sang, H. Wang, X. Liu, and L. Herranz, "Being a supercook: Joint food attributes and multimodal content modeling for recipe retrieval and exploration," *IEEE Transactions on Multimedia*, vol. 19, no. 5, pp. 1100–1113, 2017.
3. Dr. Vishwanath.C.Burkapalli*,Priyadarshini.C.Patil**, "Segmentation and Identification of Indian food items from Images", *Int. Journal of Engineering Research and Application* www.ijera.com ISSN: 2248-9622, Vol. 8, Issue3,(Part -1) march2018, pp.49-52.
4. G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.
5. M. Fauvel, Y. Tarabalka, J. Benediktsson, J. Chanussot, and J. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
6. N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. Rodriguez-Sanchez, and L. Wiskott, "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1847–1871, 2013.
7. E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
8. F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 733–740.
9. X. Shen and Y. Wu, "A unified approach to object detection via low rank matrix recovery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 853–860.
10. S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware object detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.
11. Y. Jia and M. Han, "Category-independent object-level detection," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1761–1768.
12. G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 478–487.
13. Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of object segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 280–287.
14. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2015, pp. 3431–3440.
15. R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multicontext deep learning," in *Computer Vision and Pattern Recognition*. IEEE, 2015, pp. 1265–1274.
16. G. Li and Y. Yu, "Visual saliency based on multi-scale deep features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2015, pp. 5455–5463.
17. L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Saliency detection with recurrent fully convolutional networks," in *Proceedings of the European Conference on Computer Vision*, 2016, pp. 825–841.
18. J. Pan, C. Canton, K. McGuinness, N. E. O'Connor, J. Torres, E. Sayrol, and X. Giro-i Nieto, "Salgan: Visual prediction with generative adversarial networks," *Conference on Computer Vision and Pattern Recognition Scene Understanding Workshop*, 2017.
19. Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. Torr, "Deeply supervised salient object detection with short connections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
20. S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1395–1403.
21. S. He, J. Jiao, X. Zhang, G. Han, and R. W. Lau, "Delving into object subitizing and detection," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
22. G. Lee, Y.-W. Tai, and J. Kim, "Eld-net: An efficient deep learning architecture for accurate detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 7, pp. 1599–1610, 2018.
23. T. Wang, L. Zhang, S. Wang, H. Lu, G. Yang, X. Ruan, and A. Borji, "Detect globally, refine locally: A novel approach to object detection," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
24. S. Horiguchi, S. Amano, M. Ogawa and K. Aizawa, "Personalized Classifier for Food Image Recognition," in *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2836-2848, Oct. 2018.
25. W. Zhang, D. Zhao, W. Gong, Z. Li, Q. Lu and S. Yang, "Food Image Recognition with Convolutional Neural Networks," 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, 2015, pp. 690-693.
26. Weiqing Min, Bing-Kun Bao, Shuhuan "Supplementary Material You Are What You Eat: Exploring Rich Recipe Information for Cross-Region Food Analysis" 2018.
27. Hossin, M and Sulaiman, M.N., "A review on evaluation metrics for data classification evaluations", *International Journal of Data Mining & Knowledge Management Process (IJDMP)* Vol.5, No.2, March 2015.

AUTHORS PROFILE



Dr. Vishwanath C Burkapalli is working as a Professor in the department of Information science & Engg., PDA College of Engg., Kalaburgi, Karnataka, India. He has received Ph.D in computer science from VTU Belagum, Karnataka. He has published many research papers in international journals and presented several research papers in international conferences. His research area of interest is Image Processing and Pattern Recognition.

E-mail: vishwa_bc@rediffmail.com.



Mrs. Priyadarshini C Patil is working as an Assistant professor in the Department of computer science AND Engg., and Research Scholar at Department of Information Science and Engg, PDA College of Engineering., Kalaburgi, Karnataka, India. She did her M.Tech from Vishveswaraiah Technological University, Belgaum, Karnataka, India and pursuing Ph.D in same university. E-mail: priyadarshinicpatil@gmail.com.