International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Multi-Modal Region Based Convolution Neural Network (MM-RCNN) for Ethnicity Identification and Classification

C. Christy, S. Arivalagan, P. Sudhakar,

Abstract*: Human facial images help to acquire the demographic information of the person like ethnicity and gender. At the same time, the ethnicity and gender acts as a significant part in the face-related applications. In this study, image-based ethnicity identification problem is considered as a classification problem and is solved by deep learning techniques. In this paper, a new multi-modal region based convolutional neural network (MM-RCNN) is proposed for the detection and classification of Ethnicity to determine the age, gender, emotion, ethnicity and so on. The presented model involves two stages namely feature extraction and classification. In the first stage, an efficient feature extraction model called ImageAnnot is developed for extracting the useful features from an image. In the second stage, MM-RCNN is employed to identify and then classify ethnicity. To validate the effective performance of the applied MM-RCNN model, various evaluation parameters has been presented and the simulation outcome verified the superior nature of the presented model compared to existing models.

***Index Terms*: Classification, Deep learning, Ethnicity, Faster R-CNN, Feature extraction**

I. INTRODUCTION

An essential portion of identification of human is ethnicity data and it is a helpful identifier for different application that varies between targeted advertisement and video surveillance. In recognition of biometric, ethnicity acts as an essential part. Classifying the people in order to the nationality, race and ethnicity comprise huge implicational effect over advertisement, social media profiling and surveillance. Redefining of the group identification of nation, race and regions are modifying constantly [1]. In social science, the procedure of categorization is a major subject. In educational, socioeconomic and health care studies, it is proven to be highly helpful. In market study, it comprises commercial rates mainly in multi-ethnic nation. In numerous circumstances, classification of gender acts as a significant role. Information of gender depends on few biometrics as similar to one among the parameters of demographic classification that offers ancillary data of a distinct identity data. And also, it enhances the face recognition performance. In numerous applications, it is employed extensively to offer smart services like smart interface, smart advertising and visual surveillance.

Manuscript published on 30 September 2019.

***Correspondence Author(s)**
C. Christy, Research Scholar, Department of CSE, Annamalai University
Dr. S. Arivalagan, Assistant Professor, Department of CSE, Annamalai University
Dr. P. Sudhakar, Assistant Professor, Department of CSE, Annamalai University

In gender classification, different modalities are employed that involves iris [3], human face, [2] and hand shape. Most of the traditional works over classification of gender employed the human face modality. Over various datasets, to carry out ethnicity classification, various researchers attempt to employ traditional recognition techniques [4-11] in the past years. Many techniques depend on face images that offers quick and highly direct manner to examine the demographic data or person's ethnicity. A whole review is done by Siyao Fu et al. [4] over little standard advancement in applications and algorithms, feature representation models and face-race perception. Through employing linear discriminant analysis (LDA), classification among non-Asian and Asian is researched by Xiaoguang Lu et al. [5] over a 2,630 face images containing database out of 263 domains. Multi-regularized learning (MRL) is the novel method employed [6] that is extracted from multi-task features learning (MTFL) and multi-stage learning (MSL) to use the problem of race recognition. For the Support Vector Machine (SVM) classifier training, retina sampling and Gabor wavelets transformation are used [7] to build a machine interfaces that are human-friendly in nature for classification of ethnicity. The operator local binary pattern (LBP) is used by Zhiguang Yang et al. [8] in order to categorize the gender, ethnicity and age. But, low-level feature representation is employed by these traditional methods to attain effective performance. In some years, drastic success and development have been seen in computer vision domain by deep learning. In numerous tasks that are vision-related involving classification of images [10], semantic segmentation or object detection, CNNs had been proven to attain better performance depending on deep framework. Autonomous image representation is offered by the learned features of the deep networks and are demonstrated that it attains superior performance where the human being might feel it as complex. For ethnicity classification, few recent studies make use of the frameworks and methods of deep learning. Recognition of ethnicity through Artificial Intelligence (AI) functions is a highly creative technique. The model merges the techniques of deep learning with AI that enables the model to gather the data constantly and to enhance its service depending on that. Each novel data piece aids the model to evolve and create it to bring about highly accurate outcomes while it falls on the diversity recognition techniques. Many kinds of advantages are provided by the deep learning towards the AI technique.

It is the procedure to feed constantly the novel data towards an AI model and enhances the data sum in the databases employed for numerous tasks that involves history mapping and prediction guiding of an AI model. To derive the AI technique that aid to decide the precise facial points, this novel data is employed for facial recognition model. This novel data aids in deciding the highly precise detection score of facial ethnicity diversion when given with the diversion detection Test and Ancestry Facial ethnicity. In deep learning, the novel data is fed constantly that employed new and existing data for the facial feature identification and precisely to decide the diversion detection score and facial ethnicity is a significant portion in the superior scoring and accuracy development. To enhance the facial recognition procedure accuracy constantly, deep learning is employed through comparison of new person's face photos with the growing photos database that is examined before for diversion detection and facial ethnicity. To enhance the confidence score of ethnicity detection, deep learning is employed through the comparison of prior face features and its facial diversion as well as ethnicity detection scores with novel photos to create a scoring curve of higher precision in facial diversion & ethnicity scores. By employing deep learning, [12] projected to identify the semantic facial features. Many techniques are executed by Haoxuan Chen et al. [13] that involve SVM classifier, K-nearest neighbor technique, CNN and two-layered NN to classifier training to the prediction of Japanese, Korean and Chinese. By the classification of 3-class, it attains 89.2% as prediction rate. By employing CIFAR10 network, classification of ethnicity is done by Wei Wang et al. [14] that aim at white and black people classification, Uyghurs and Han Chinese people and non-Chinese and Chinese people. The implementation includes self-collected and public datasets together. In order to predict three ethnicities like Caucasian, Negro and Mongolian that comprise of 447 gathered images, [15] uses CNN and the images are derived from database called FERET. This technique makes use of the benefit of color data and geometric features derived out of the NN. While comparing with MLP network, this outcome demonstrates some enhancement. Even though, its types are rather unique like variation among East Asian and European people, white and black people, the outcomes or less significant as every kind might be varied simply through the machine or human with huge complexity. It is highly significant commonly when to categorize people with near relations geographically. Classification carried out by CNN-based method is depending on the associated layer with softmax classifier [16]. These technique does not exploit the richer intermediate features and therefore the available data might not be employed. The classifier cannot adequately learn and this might offer improvement in poor fitting for circumstances wherever small dataset are there for training. The outcomes of the derived classifier are not autonomous. An effective Race Recognition Framework (RRF) is projected which involve preprocessing and face detection (FD&P), information collector (IC), and RR modules. Two independent models are projected by this research for RR module. By employing deep convolutional neural network (CNN), the primary model is RR. The next is a fine-tuning model for RR that depends on popular trained model and VGG for object recognition [17]. Though several techniques have been developed for ethnicity recognition, there is still a room to improve the efficiency of the classification model. At the same time, the application of multi-model in ethnicity recognition shows the novelty of the work. Keeping this view, a new feature extraction technique is presented in prior to classification process to improve the overall efficiency. In this paper, a new multi-modal region based convolutional neural network (MM-RCNN) is introduced for the detection and classification of Ethnicity to determine the age, gender, emotion, ethnicity and so on. To validate the effective performance of the applied MM-RCNN model, various evaluation parameters has been presented and the simulation outcome verified the superior nature of the presented model compared to existing models. The upcoming parts are planned as follows. A detailed explanation of MM-RCNN model is discussed in next section. An extensive simulation takes place in Section 3 and concluding remarks are given in Section 4.

II. PROPOSED METHOD

The overall process of the MM-RCNN is shown in Fig. 1 and the steps are clearly provided in Algorithm 1. The presented model involves two stages namely feature extraction and classification. In the first stage, an efficient feature extraction model called ImageAnnot is developed for extracting the useful features from an image. In the second stage, MM-RCNN is employed to identify and then classify ethnicity. Submit your manuscript electronically for review.

A. ImageAnnot based Feature Extraction

Extraction of feature is defined as relative shape data comprised within a pattern in order to that the pattern classifying task is performed easily through the formal process. A specific format of reduction in dimensionality is feature extraction in image processing and pattern recognition. The major target of extracting feature is to derive the highly relative data out of the actual data and demonstrating the data in low dimensional space. While the algorithm input data to be processed is so large and finds to be repeating, then the input data might be modified towards a decrease feature set representation. The input data transforming into the feature set is known as feature extraction. When the derived feature are selected carefully, it is desired that the relative data will be derived from features set from input data to carry the expected job by employing this minimized representation in spite of whole input size. For every label, through independent binary classifier learning, the issue is resolved and uses binary classifiers for the label prediction for unlabeled images. To enhance the AI accuracy through the technique of graph-based learning, the MLMC model focus at using label correlations to enhance the consistency of label assignment on the full image similarity. Node is responsible for un- labeled or labeled data points in the MLMC model and the similarity among data points are reflected by the edges. In order to measure the two data points' proximity, both the label and pairwise data similarity are employed.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

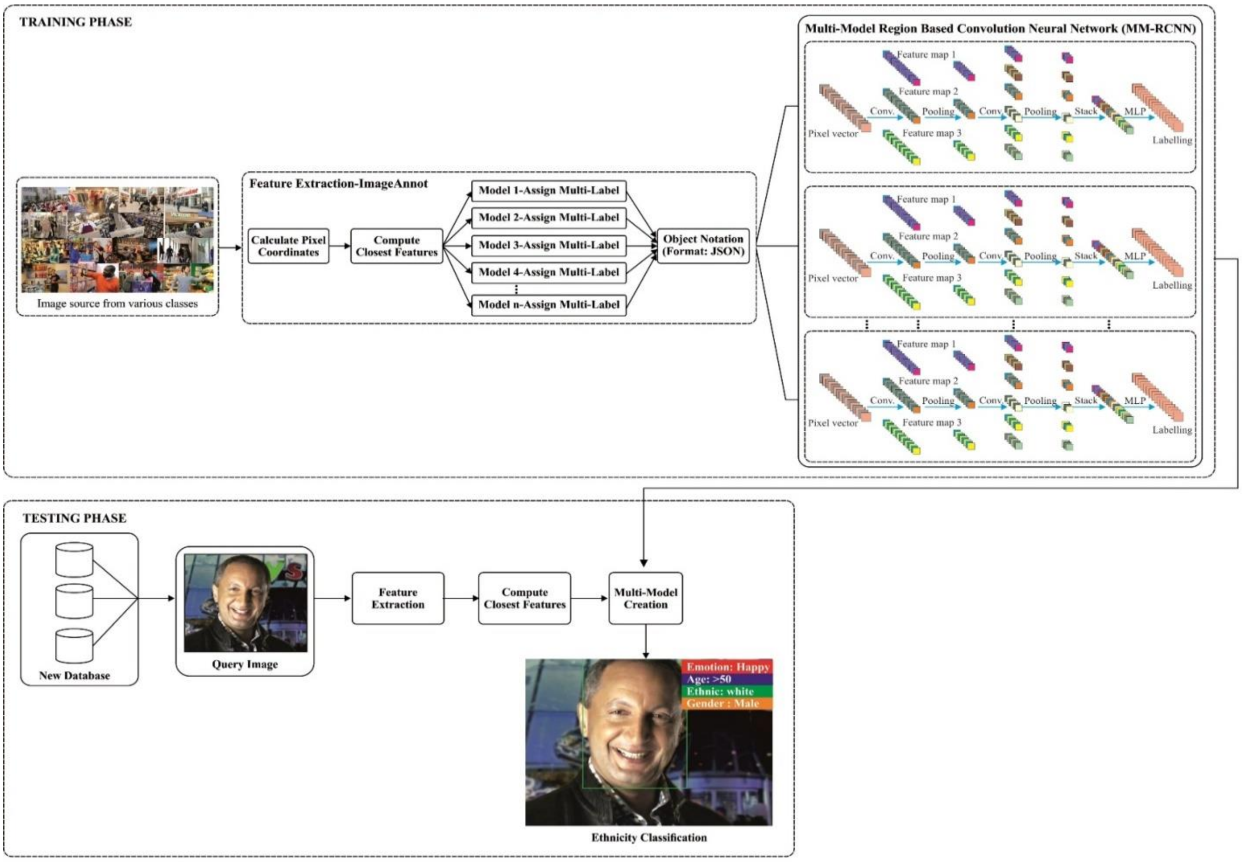

Fig.1. Overall process of proposed method

The weight w is employed to denote the similarity among node over a weighted graph:

$$W = W_X + \gamma W_L \qquad (1)$$

The similarity among pairwise image data are denoted by W_X and the similarity among pairwise labels are denoted through W_L. To manage the two similarity matrices influences, parameter γ is employed.

To estimate the similarity among two images, Gaussian function is employed. For a data point x_i, $\{z_1, z_2, \ldots z_n\}$ is a k -dimensional vector that denotes the indication vector of label assignment. x is the image that is annotated when $z_i(k) = 1$with label k . The label similarity among two images is estimated as below:

$$W_L(i,j) = \frac{z_i^T c z_j}{\|z_i\| \|z_j\|} \qquad (2)$$

Where square matrix is denoted by c to denote the label correlations, $r_l$and $r_k$denotes the l th and k th rows of Z that denote labeled images through a respective keyword. Through a cosine similarity, label correlation among two keywords is measured.

$$c_{kl} = \cos(r_k, r_l) = \frac{\langle r_k, r_l \rangle}{\|r_k, r_l\|} \qquad (3)$$

The MLMC is changed to a multimodal perspective in the study work.

When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables.

B. Faster R-CNN based detection and classification

Region based Convolution Neural Network (R-CNN) is the object detection model that is made of two modules as shown in Fig. 2. Deep fully convolutional network is the primary module that projects areas and the next is the Fast R-CNN detector which employs projected areas. For object detection, the whole model is a unified and single network. Fast R-CNN module conveys the information about RPN module by employing well known neural networks terminology.

Algorithm I: MM-RCNN

Stage 1: Feature Extraction

Input: Image(i) (where i=1,2,3…n)
Output: Extracted Features
For ∀i Є Image Set do
  Determine CNN Features for extracting correlation from the image
End For

Stage 2: Class Labelling

Input: Predicted class labels like {Age, Emotion, Gender, Ethnicity}
Output: Multi-Model Creation {where model_1, model_2, model_3…model_n}
For i=1 to N
  For∀i Є Image Set do
    allocate_class_label(i)
  End For
 Multi-Model Generated for every class label
End For

Stage 3: Ethnicity Classification

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Input: New Input Images
Output: Ethnicity Classified Image Set
For$\forall_i$ € Image Set do
    Determine CNN Features for extracting correlation from the image
    For$\forall_i$ € Region in Image do
        Labelling (i)
    End For
End For

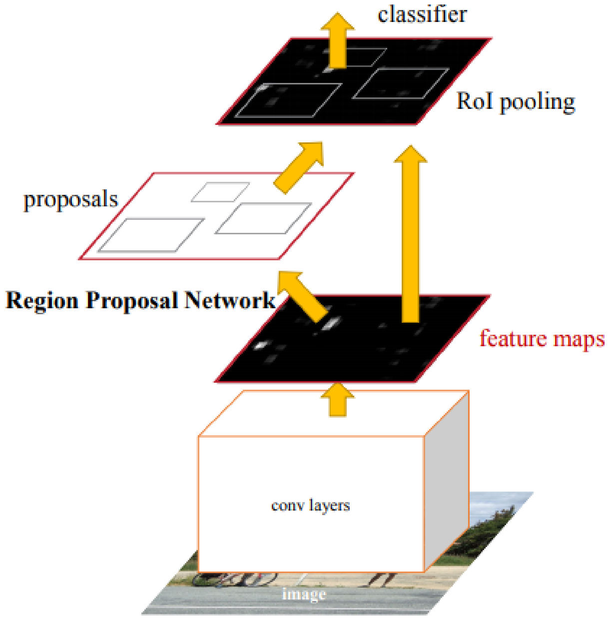

Fig. 2. Region based Convolution Neural Network

Region Proposal Networks (RPN)

An RPN derives the different sized images as input, and it offers the output as rectangular object set proposals with a distinct objectness score. This procedure is designed as CNN. Therefore, the major focus is to computation sharing with Fast R-CNN, it considers that nets distribute the similar convolution layer sets. Over convolution feature map, a small network is slided for the region generation proposals through the end distributed convolutional layer. n×n spatial window is taken as input towards the small network of input convolutional feature map. Towards low dimensional feature, each sliding window is mapped and is offered to the two entire associated layers: box-classification layer and box-regression layer.

Anchors

To region boxes ranking, ranking procedure is performed by RPN and it describes highly possible one that holds the objects. In Faster R-CNN execution, anchors acts as a significant role. A sum of 9 anchors will be contained through the anchor which is a box simply at an image position. The 9 anchors placement with size (600, 800) and (320, 320) position is demonstrated in Fig. 3.

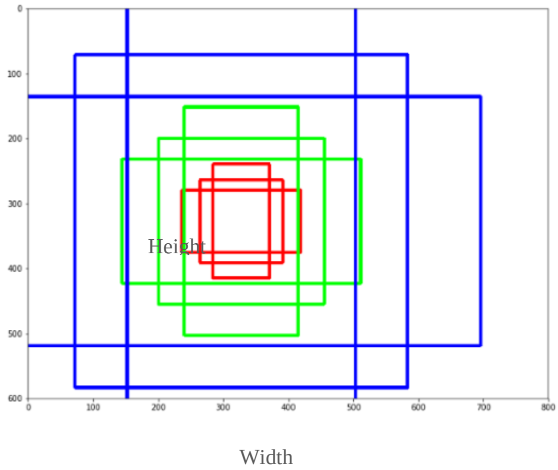

Fig. 3. Anchors

Loss Function

Towards each anchor, it is required to assign binary class label in RPN training. Towards two anchor types, positive label is allocated: (i) with real box, anchors with higher Intersection-over-Union (IoU) merge (ii) anchor that comprise higher than 0.7 IoU. It is denoted that the distinct real box may be assigned positive labels towards numerous anchors. The anchors that cannot fall in negative or positive are not included while in the training procedure.Region based CNN focus to reduce the goal function follows the loss of multi-task within Fast R-CNN. Eq. (4) demonstrates the image loss function

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i,p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i,t_i^*) \quad (4)$$

Let in a mini-batch, anchor index is denoted through i and the anchor i predicted probability is denoted by p_i for an object. While with positive anchor, real label is p_i^* and zero if it is negative anchor. Vectors that denote the four parameterized coordinate points is t_i of predicted bounding box, and the real box associated with positive anchor is t_i^*. L_{cls} is the classification loss. $L_{reg}(t_i,t_i^*) = R(t_i - t_i^*)$ is used for regression loss, wherever robust loss function is denoted through R.

Regression loss is denoted by $p_i^* L_{reg}$ that is the $p_i^* = 1$ predicted probability and for $p_i^* = 0$, it will be in inactive state. $\{t_i\}$ and $\{p_i\}$ are held by the output of reg and cls layers. The attributes that is used for bounding box regression is expressed as Eq. (5):

$$\begin{gathered} t_x = (x - x_a)/w_a, \quad t_y = (y - y_a)/h_a \\ t_w = log\,(w)/w_a, \\ t_h = log(h)/h_a \\ t_x^* = (x^* - x_a)/w_a, \quad t_x^* = (y^* - y_a)/h_a \\ t_w^* = log(w^*)/w_a, \quad t_h^* = log\,(h)/h_a \end{gathered} \quad (5)$$

Let box's center coordinates are denoted as x, y and the height and width are denoted through h and w. The anchor box, ground truth box and predicted box are denoted by the parameters x_a, x^* and$x$.

Retrieval Number: J91020881019/19©BEIESP
DOI: 10.35940/ijitee.J9102.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

ROI Pooling

The projected areas of different sizes are offered from the RPN output. It is composite to create an efficient model to operate over the different feature sizes. The problem is simplified by the region of Interest (ROI) pooling through reducing the feature maps to similar sizes. ROI pooling segments the feature map of input towards the preset count when compared to Max-Pooling of similar areas approximately and towards every area, Max-Pooling would be used. The ROI Pooling output is commonly represented as k that is irresponsible for size of the input.

Sharing Convolutional Features for Region Proposal and Object Detection

The training procedure is performed by Fast R-CNN and RPN independently. In different manners, the convolution layers would be changed. There exists a requirement to model a method which allows distributing the convolution layers among networks in spite of learning in two distinct networks. It is difficult to describe the unique network that comprises Fast R-CNN and RPN together and by employing back propagation, it would be optimized. This is due to the Fast R-CNN training that depends on the predetermined proposals of object.

For shared features learning, four phase training technique is employed by other optimization. RPN should be trained primarily. By Fast R-CNN, distinct detection network might be trained subsequently by employing the generated proposals out of RPN in prior phase. The convolution layers cannot be shared by Fast R-CNN and RPN at that instant. In the next phase of RPN training initialization, detector network is employed. Layers of shared convolution are fixed and the distinct RPN layer might be adjusted however. Convolution layers might be shared at that instant. The shared convolution layers managing at last might be adjusted. Therefore, the Fast R-CNN and RPN distribute the similar convolution layers and unified network is developed.

In ethnicity classification, numerous classes are included, to extract individually the model, there exist a concept, for every class and then to combine the entire models to work out single multi-model for classification. With a certain class, every model is individually trained. In the dataset, when the entire included classes are trained through the implication of distinct multiple models, they are combined to single multi-model to recognize efficiently and classifying the image ethnicity.

III. PERFORMANCE VALIDATION

A. Dataset Used

For ensuring the superior characteristics of the MM-RCNN, a set of experimentations are carried out on benchmark dataset [17] and the details are provided in Table 1. Since the MM-RCNN model identifies multiple classes, the sub-classes under every entity are also given in the table. For instance, the sub-classes under the entity 'Emotions' are Emotion_Happy, Emotion_Sad, Emotion_Neutral and Emotion_Angry. The dataset contains a set of 119 instances under five entities namely No_face, Emotions, Age, Ethnicity and gender. Some of the sample test images are given in Fig. 4.

Table 1 Dataset Description

Entities	Number of Classes	Number of Instances
No_Face	Not_Face	
Emotions	Emotion_Happy Emotion_Sad Emotion_Neutral Emotion_Angry	
Age	Age_below_20 Age_20_30 Age_30_40 Age_40_50 Age_above_50	119
Ethnicity	E_Asian E_White E_Black E_Hispanic E_Indian E_Arab	
Gender	G_Male G_Female G_Other	

B. Accuracy

Next, the results are validated by the use of an important measure called accuracy. It indicates the total percentage of properly classified instances and is indicated by means of %. To available superior classifier results, the classification accuracy value should be nearer to 100 as defined in Eq. (6):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Where TP-true positive, TN- true negative, FP-false positive and FN-false negative.

Fig. 4. Sample Test Images

Table 2 Test image 001 and its annotated Points

Image 001	**Annotation Details**

	"image_name": "image_001",
	"content":
	"http://com.dataturks.a96-i23.open.s3.amazonaws.com/2c9fafb06477f4cb0164895548a600a3/66127d05-93eb-498f-bac3-85a19bcbbbc7___2538464.main_image.jpg.jpeg",
	"annotation": [{"label": ["Emotion_Happy", "Age_below20", "E_White", "G_Male"],
	"notes": "",
	"points": [{
	"x": 0.19876868953386104,
	"y": 0.23148148148148148
	}, {
	"x": 0.3245382585751979,
	"y": 0.45987654320987653
	}],
	"imageWidth": 1300,
	"imageHeight": 741
	}, {
	"label": ["Emotion_Happy", "Age_20_30", "E_White", "G_Female"],
	"notes": "",
	"points": [{
	"x": 0.6024626209322779,
	"y": 0.16049382716049382
	}, {
	"x": 0.7361477572559367,
	"y": 0.36882716049382713
	}],
	"imageWidth": 1300,
	"imageHeight": 741
	}],
	"extras": null,
	"metadata": {
	"first_done_at": 1531657992000,
	"last_updated_at": 1531657992000,
	"sec_taken": 23,
	"last_updated_by": "HWbICv9u4uSnWrU830DjuF7FfMK2"
	}

C. Results analysis

Once the test input image is provided, the presented MM-RCNN model will initially annotate the image by the use of ImageAnnot technique. The annotation details of the sample test image 'Image 001' is tabulated in the Table 2. Since the MM-RCNN follows multi-modal classification, a conversion takes place from uni-modal classification to multi-modal classification. Initially, all the classes are identified separately and then merged together to define multi-class labels. As shown in Table 3, for the applied test image 001, individual classes are defined for the applied individual image.

Retrieval Number: J91020881019/19©BEIESP
DOI: 10.35940/ijitee.J9102.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

As five types of classes are present, al the five class labels are given individually. Then, the presented MM-RCNN model merges all the classes together and the classes are represented in the input image itself. As shown in Table, every individual class is indicated in second row and the classes are merged together in third row.

Similarly, the ethnicity classified final image 001 by the presented MM-RCNN model is shown in Table 4. Here, it can be seen that all the classes for every individual present in the test image is clearly identified and classified.

To further highlight the betterment of MM-RCNN model, a detailed comparative study is made under varying cross validation (CV) values. A range between 1 to 10 CV values is taken for performance analysis. In addition, a comparison with RR-CNN, RR-VGG0, RR-VGG1 and RR-VGG2 [16] is made with respect to accuracy as shown in Table 5. Fig. 5 also shows the comparative analysis among various methods interms of accuracy. As shown in the table 5, under the CV-1, a maximum of 91.9% is achieved by the presented MM-RCNN model and a minimum of 82.15% is attained by the existing RR-VGG0 model. At the same time, the RR-VGG1 and RR-VGG2 shows better accuracy over RR-CNN, but not than the presented model.

Table 3 Uni-Modal to Multi-Model Ethnicity Classification

Original Image 001

Age

Emotion

Ethnicity

Gender

Multi-Model Ethnicity Classification

Similarly, under the CV-2, highest accuracy of 90.67% is achieved by the presented MM-RCNN model and a lowest accuracy of 51.15% is attained by the existing RR-VGG1 model. On the other hand, the RR-CNN tries to exhibit better classification with the accuracy of 89.34. However, it fails to excel better results over the MM-RCNN model. Under the CV-2, a maximum of 92.37% is achieved by the presented MM-RCNN model and a minimum of 75.21% is attained by the existing RR-VGG0 model. At the same time, RR-VGG2 shows competitive performance with the accuracy of 90.66%. But, the MM-RCNN model exhibits better results with an accuracy of 92.37%. Similarly, under the CV-5, a maximum accuracy is achieved by the presented MM-RCNN model and a minimum of 40.64% is attained by the existing RR-VGG0 model. At the same time, RR-VGG2 shows competitive performance with the accuracy of 91.48%. But, the MM-RCNN model exhibits better results with an accuracy of 91.98%.

Retrieval Number: J91020881019/19©BEIESP
DOI: 10.35940/ijitee.J9102.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Likewise, under the CV-6, a maximum of 92.90% is achieved by the presented MM-RCNN model and a minimum of 73.24% is attained by the existing RR-VGG0 model. At the same time, the RR-VGG1 and RR-VGG2 shows better accuracy of 80.98% and 88.85% over RR-CNN, but not than the presented model. Identical to the above CV values, the presented model shows excellent performance over the compared models under all values of CV. It is shown that a highest accuracy of 96.59 is achieved by represented MM-RCNN model on the applied test images under the CV of 10. At the same time, a maximum accuracy of only 91.64% is achieved by the existing RR-CNN model. In addition, a clear average accuracy analysis is made between the presented MM-RCNN and existing models on the applied test images. As shown in Fig. 6, a maximum average accuracy of 92.98% is obtained by the MM-RCNN model. Next to that, RR-VGG2 shows slightly lower average accuracy of 88.88% whereas the RR-CNN, RR-VGG0 and RR-VGG1 exhibits moderate performance with the average accuracy values of 88.64%, 72.37% and 78.7% respectively. In overall, the presented MM-RCNN model is the superior model for ethnicity recognition and classification over the compared methods in a significant way.

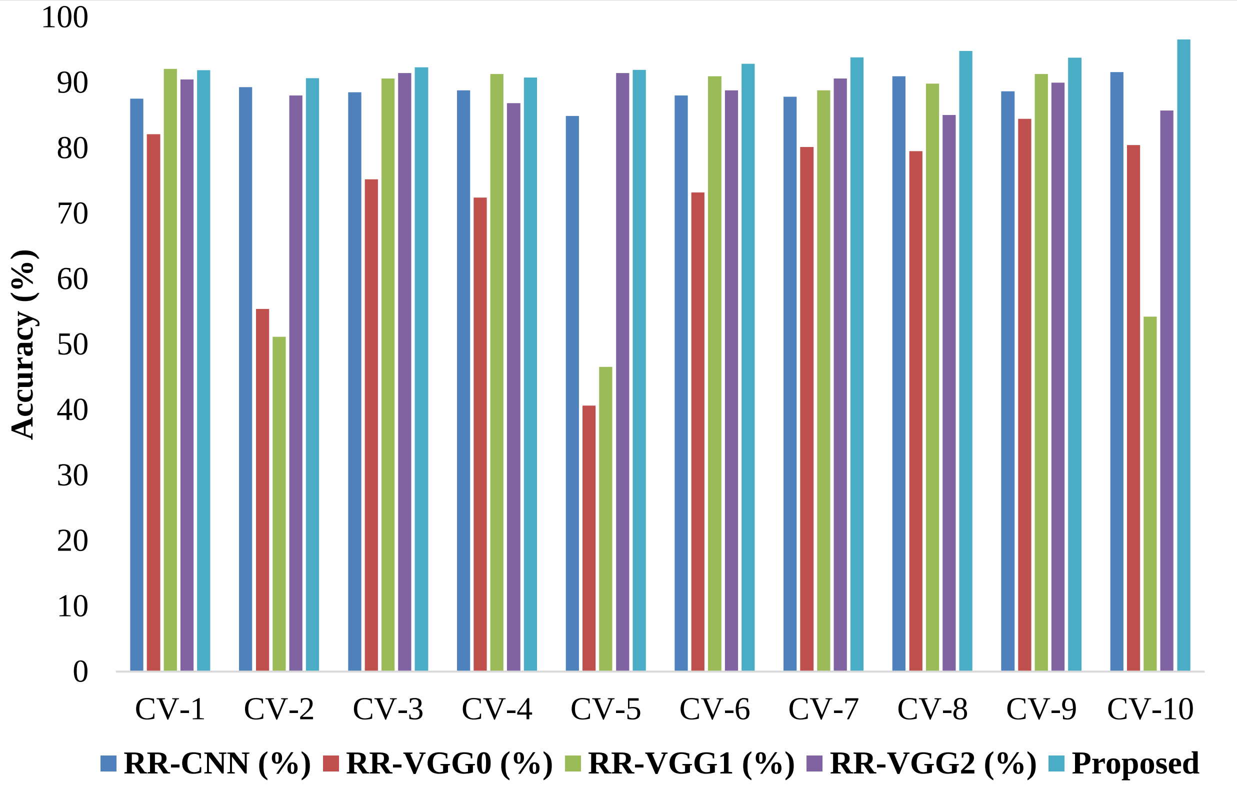

Fig. 5. Accuracy analysis of different models for applied dataset

Table 5 Accuracy of different models for applied dataset

Cross Validation	Methods				
	RR-CNN (%)	RR-VGG0 (%)	RR-VGG1 (%)	RR-VGG2 (%)	Proposed
CV-1	87.54	82.15	92.13	90.49	91.90
CV-2	89.34	55.42	51.15	88.03	90.67
CV-3	88.52	75.21	90.66	91.48	92.37
CV-4	88.85	72.46	91.31	86.89	90.78
CV-5	84.92	40.64	46.56	91.48	91.98
CV-6	88.03	73.24	90.98	88.85	92.90
CV-7	87.87	80.15	88.85	90.66	93.89
CV-8	90.98	79.56	89.84	85.08	94.85
CV-9	88.69	84.46	91.31	90.00	93.84
CV-10	91.64	80.46	54.26	85.74	96.59

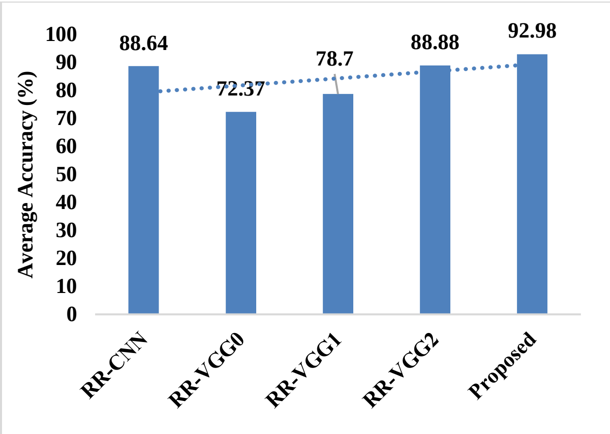

Fig. 6. Average accuracy analysis of different methods

Retrieval Number: J91020881019/19©BEIESP
DOI: 10.35940/ijitee.J9102.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

IV. CONCLUSION

In this paper, a new multi-modal based ethnicity recognition and classification model called MM-RCNN is introduced to determine the age, gender, emotion, ethnicity and so on. The presented model involves two stages namely feature extraction and classification. In the first stage, an efficient feature extraction model called ImageAnnot is developed for extracting the useful features from an image. In the second stage, MM-RCNN is employed to identify and then classify ethnicity. To validate the effective performance of the applied MM-RCNN model, several experiments have been carried out. From the experimental outcome, it is evident that a maximum average accuracy of 92.98% is obtained by the MM-RCNN model whereas the RR-CNN, RR-VGG0 and RR-VGG1 exhibits moderate performance with the average accuracy values of 88.64%, 72.37% and 78.7% respectively. In overall, the presented MM-RCNN model is the superior model for ethnicity recognition and classification over the compared methods in a significant way.

REFERENCES

1. D. R. Williams, N. Priest, and N. B. Anderson, "Understanding associations among race, socioeconomic status, and health: Patterns and prospects.," Health Psychology, vol. 35, no. 4, p. 407, 2016.
2. V. Tomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, "Learning to predict gender from iris images," in Proceedings of the 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–5, Crystal City, VA, USA, September 2007.
3. G. Amayeh, G. Bebis, and M. Nicolescu, "Gender classifcation from hand shape," in Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, USA, June 2008.
4. S. Fu, H. He, and Z.-G. Hou, "Learning race from face: A survey," IEEE transactions on pattern analysis and machine intelligence, vol. 36, no. 12, pp. 2483–2509, 2014.
5. X. Lu, A. K. Jain, et al., "Ethnicity identification from face images," in Proceedings of SPIE, vol. 5404, pp. 114–123, 2004
6. M. A. Borgi, M. El'Arbi, D. Labate, and C. B. Amar, "Face, gender and race classification using multi-regularized features learning," in Image Processing (ICIP), 2014 IEEE International Conference on, pp. 5277–5281, IEEE, 2014.
7. S. Hosoi, E. Takikawa, and M. Kawade, "Ethnicity estimation with facial images," in Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, pp. 195–200, IEEE, 2004.
8. Z. Yang and H. Ai, "Demographic classification with local binary patterns," Advances in Biometrics, pp. 464–473, 2007.
9. R. Girshick, J. Donahue, T. Darrell, and J. 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.
10. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
11. S.B. Sasi, N. Sivanandam, A survey on cryptography using optimization algorithms in WSNs, Indian J. Sci. Technol. 8 (2015) 216–221.
12. A. Gudi, "Recognizing semantic features in faces using deep learning," arXiv preprint arXiv:1512.00743, 2015.
13. H. Chen, Y. Deng, and S. Zhang, "Where am i from? -east asian ethnicity classification from facial recognition," Project study in Stanford University, 2016.
14. W. Wang, F. He, and Q. Zhao, "Facial ethnicity classification with deep convolutional neural networks," in Chinese Conference on Biometric Recognition, pp. 176–185, Springer, 2016.
15. S. Masood, S. Gupta, A. Wajid, S. Gupta, and M. Ahmed, "Prediction of human ethnicity from facial images using neural networks," in Data Engineering and Intelligent Computing, pp. 217–226, Springer, 2018.
16. A. Gudi, "Recognizing semantic features in faces using deep learning," arXiv preprint arXiv:1512.00743, 2015
17. Vo, T., Nguyen, T. and Le, C., 2018. Race recognition using deep convolutional neural networks. Symmetry, 10(11), p.564.
18. https://dataturks.com/projects/Mohan/Face Dataset With Emotion_Age_Ethnicity

AUTHORS PROFILE

C.Christy, MCA., M.Phil., SET, is working as an Assistant Professor in PG and Research Department of Computer Science at St. Joseph's College of Arts and Science (Autonomous), Cuddalore-1.She has 17 Years of Teaching Experience. She is currently pursuing PhD in Department of Computer Science and Engineering, Annamalai University of Tamilnadu, Chidambaram, since 2017.Her Research interest mainly focus on Machine Learning, Deep Learning, Image Processing, and Data Mining. She has published 15 papers in International, National Conferences and one Paper in IEEE Conference. She has Published 13 Research Papers in International and National Journals and one Research Paper in SCOPUS Journal.

Dr. S. Arivalagan obtained his Bachelor of Engineering Degree in Computer Science and Engineering during 2003 from Annamalai University and Master of Engineering Degree in Computer Science and Engineering in 2005 from the same University. He started his carrier as a Lecturer in September 2005 in the Department of Computer Science and Engineering in the Faculty of Engineering and Technology, Annamalai University, Annamalainagar, Tamilnadu, India and currently serving as an Assistant Professor in the same Department. He is a member of Computer Society of India, Indian Society for Technical Education. His areas of interest include Image Processing, Soft Computing, Data Mining and Big Data Analytics. Currently, he is guiding six research scholars at Annamalai University.

Dr. P. Sudhakar obtained his Bachelor of Engineering Degree in Electronics and Communication Engineering in 2000 from Bharathidasan University, Tiruchirappalli through A.V.C College of Engineering, Mayiladuthurai and his Master of Engineering Degree in Computer Science and Engineering in 2005 from Annamalai University, Annamalai nagar. He started his carrier as a Lecturer in August 2000 in the Department of Computer Science and Engineering in the Faculty of Engineering and Technology, Annamalai University, Annamalai nagar, Tamilnadu, India and currently serving as an Associate Professor in the same Department. He is a member of Computer Society of India, Institution of Electronics and Telecommunication Engineering, Institution of Engineers (India), Indian Science Congress Association, Indian Society for Technical Education, VLSI Society of India etc., His areas of interest include Software Engineering, Data Mining, and Big Data Analytics. He has already produced two Doctorates and other six research scholars are in the pipeline.

Retrieval Number: J91020881019/19©BEIESP
DOI: 10.35940/ijitee.J9102.0981119
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)