International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 3878 – 3884

Face and Food Recognition Using Convolutional Neural Networks

Nidhi Shah[1], Vanshika Singh[2], Samiksha Shirbhate[3],Pranjal Kumavat[4],Dr.Meenakshi.A. Thalor[5]

Information Technology,Savitribai Phule Pune University [1]nidhish77988@gmail.com
[2]vanshikalksh@gmail.com [3]samikshashirbhate2016@gmail.com [4]pranjal.kumavat25@gmail.com
[5]meenakshi.thalor@gmail.com

Abstract

This paper proposes a system with models which will recognize face and food from the given input images using CNN. The result will be saved and displayed. Shallow neural network is built over pre trained models for better results. The system can be used by schools which provide lunch to their students for keeping a proper record. The paper focuses on the models used for face and food recognition. The dataset for faces is built with around 400 images divided into 31 classes. Food dataset is built with around 500 images also with 31 classes. The validation accuracies of the models is 91% and 88% for face and food models respectively

Keywords— *Face recognition, Food recognition, CNN, EigenFaces, Haar*

I. INTRODUCTION

The motivation behind developing these models was for the mid-day meal program. It is a program where school children are fed and the food is provided by the government. It is one of the ways to attract children from poor families to come to school. Our system is designed to check the meals they provide and that students receive their fair share of food. Face recognition is used in some high-tech security systems. The use of videos or images and real-time face detection helps in maintaining proper reports and also helps in security management. Maintaining a good diet starts with eating healthy food. So, recognizing the food items we are consuming on daily basis is also important. These processes are made automatically using different algorithms on various datasets.

There are various techniques for face and food detection and recognition. These techniques have evolved over a significant timeline. We have discussed different techniques like Eigen Faces [1], PCA [2][3], CNN[4] for face recognition. Similarly, in food recognition, different techniques have been discussed. Each paper has used different datasets and has proposed a model with the highest achievable accuracy.

However, by studying all the papers and we decided to use CNN technique because it is a proven technique and its high accuracy. In this paper, we have presented face and food models which are trained in 31 classes each. Each dataset has around 400-500 images on which the models are trained. Section 2 gives a summary of the literature survey and different techniques for face and recognition. In section 3 our models are explained in depth.

II. LITERATURE SURVEY

A. Face Recognition Techniques

In [1], face detection and recognition are done on a real-time basis. The Haar algorithm is used for face detection. This algorithm performs well for illumination changes. Recognition of faces is done using EigenFaces and Gabor Feature algorithms. The accuracy of the system using EigenFaces is approximately 65% and using the Gabor feature is approximately 60%. In [2] real-time attendance system is proposed which detects the faces of students using different modes like snap and video. The Haar method is used for face detection in this system as well. It is built on Raspberry Pi and the

ISSN: 2233-7853 IJFGCN

software module uses open-source computer vision libraries. The accuracy of this system is 74%. Another attendance system is proposed in [3], which uses Haar for face detection and PCA and EigenFaces for recognition of faces. The proposed system stores the records of the students in the database. The captured image undergoes the detection process. The process is carried out by checking the landmarks of the face using AdaBoost Algorithm. The recognition is done by Principle Component Analysis with EigenFaces.

In [4], Convolutional Neural Network (CNN) is used for face detection. CNN is used for image classification and object recognition, as it works on raw pixel strength of image which is given as input, flat vector. The accuracy of the model is increased from 68.85% to 79.41% after the test set. In [5], the CNN technique on FPGA (field-programmable gate array) is used for face recognition. This paper proposes real-time face recognition by using FPGA. The system claims to be 99.25% better than other systems built on CPU, GPU, and even other algorithms. The only drawback of this is that there isn't much research done on face recognition using FPGA. It is also difficult to code a CNN algorithm on an FPGA. The hardware of this system can be a limiting factor.

In [6], the system specifically focuses on face recognition when there are some obstacles in capturing the face. The recognition of the face is done by calculating the Euclidean distance. In this, the dataset had 30% occlusion over the faces. The CNN model proposed in this paper has a recognition rate of 98%. Multiple layer model for real-time face detection is discussed in [7], it works on the CPU. Multiple Images are detected using CNN architecture which uses shallow layers that produce light variations so the system can work as close to real-time. They have achieved this using Average Precision for measuring the accuracy of the system which is 90.84%. This can be further improved by using pre-processing.

B. Food Recognition Techniques

In [8] deep convolutional neural networks (DCNN) are proposed to classify food images. Food image analysis has three steps: 1. image segmentation 2.food recognition 3.quantity analysis. In food analysis color, intensity, texture, etc are important features. The model proposed consists of 54 layers. Feature extraction is done using Inception v3. The model shows 92.3% accuracy for food detection. The only drawback of this model is that even after using a pre-trained model the computational resource requirement is quite high. In [9] ensemble network containing GoogleNet, ResNet, and AlexNet is used for food image detection. It is trained and tested on ETH-101 and the Indian food dataset. The feature extraction is done with the help of an external pre-trained model. AlexNet and GoogleNet are used in the max-pooling layer of the CNN module and ResNet are used in the last output layer. The accuracy of this model is 73% for the Indian food database which in comparison to AlexNet, ResNet and GoogleNet individually are quite high. The hardware used in this experiment includes an NVIDEA processor with a system requirement of 128GB of ram. The architecture of each sub-network is different which poses a challenge in coding, reshaping, and debugging.

In [10] a method for food image classification and furthermore how to determine its calories is discussed. Food recognition is done by 2 techniques namely, Graph cut segmentation and deep learning neural network. From these two methods, CNN has proven to be more efficient. The dataset used for training and testing purposes contains images of single foods. The CNN approach proposed in the paper [11] is built on 5-layer network architecture. The first 4 layers are convolutional-pooling layers and the last layer is fully connected. UEC-FOOD100 and an open-source database are used in

ISSN: 2233-7853 IJFGCN

the training and testing of the model. The accuracy on a single food item is 80.8 and 60% on multi-food item datasets.

In [11] CNN method is proposed to recognize the image of foods. In this paper the food Activation maps i.e food probability were generated in a global average pooling layer. Fine Tuning was done for FAM generation in which the softmax layer was settled and convolutional layers were added with stride. For classification of the test image into its appropriate class Bag Of Feature is used. The accuracy is 86.5% on datasets.

III. SYSTEM ARCHITECTURE

CNN has performed well under all the different scenarios as seen in the above literature survey. The system will detect and recognize faces and food from the given input images. The major two steps in image classification are feature extraction and classification, which are done by CNN models. The models are trained in a supervised manner.

Images of faces and food will be uploaded separately on the website. In the face module, face detection is done using the MT- CNN algorithm which detects the faces in the given input image and the array of detected faces is its output. It is in 224x224x3shape. The pre-trained model VGGFace which acts as feature extractor has input as 3 feature maps of the size 224 x 224. The output layer is excluded while using VGGFace. After feature extraction, the output is given as input to our model. This model consists of 3 dense layers which are fully connected layers of the CNN. This model generates a list of probabilities concerning each class. The maximum of this is chosen as the final output. It corresponds that the input image is of the person with maximum probability among the labeled list.

Similarly, in the food module, VGG16 is used as a feature extractor. The input shape for the images is 224x224x3. Here 224x224 is the size of the input image. VGG16 is used as a feature extractor by excluding the output layers of the model. In VGG16 input is given as an image and it is passed from the convolutional layer in which filters are present. The convolution stride is fixed to 1 pixel. The spatial padding of convolution layer input is done in a way that the spatial resolution is preserved after convolution. Spatial pooling is done by a maximum of five pooling layers followed by some convolutional layers. The three fully-Connected layers follow a stack of convolutional layers, the first two contain 4096 channels each and the third layer contains 1000 channels. These layers have been replaced with a new output layer with an activation layer of 31 categories. The output layer is a shallow neural network that has dense and dropout layers with a softmax activation layer. All the layers are linearly stacked using Keras Sequential model. VGG16 outperforms the previous generation of models because it achieved the best result. The model consists of 1 global average pooling layer and dense layers.

The output of VGG16 is given as input to our model. It gives us a labeled list of probabilities for each class in the training dataset, the maximum of which is the output of the module. After recognition of faces and the food items, the result is stored in the database and a report is also made available. Fig 1 shows the above process.

ISSN: 2233-7853 IJFGCN

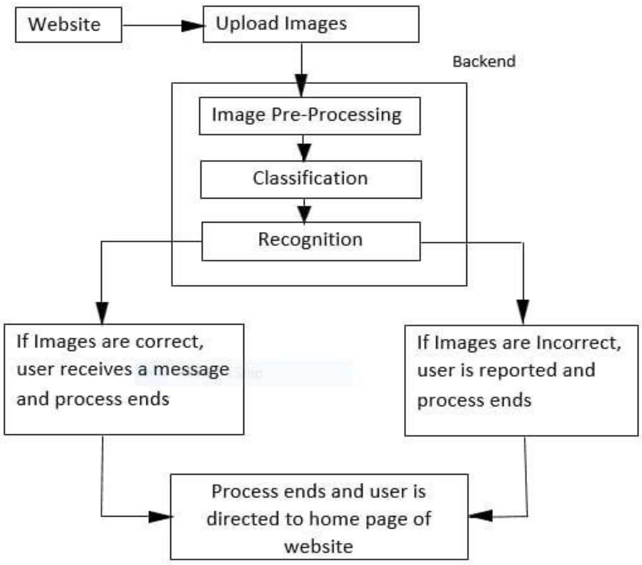

Fig 1.System Architecture

IV. EXPERIMENTAL RESULTS

The process of using the pre-trained model as a feature extractor for a new model is called transfer learning. The face model is trained in 31 classes. It is trained on around 400 images. The dataset is split into 7:3 for training and testing purposes. In this we have created a database of around 700 images of a face as shown in Fig 2. The database was created using agumentation technique. The face model shows 91% accuracy. Fig 3. shows the accuracy graph of the face model.

Fig 2. Data set of images of students

ISSN: 2233-7853 IJFGCN

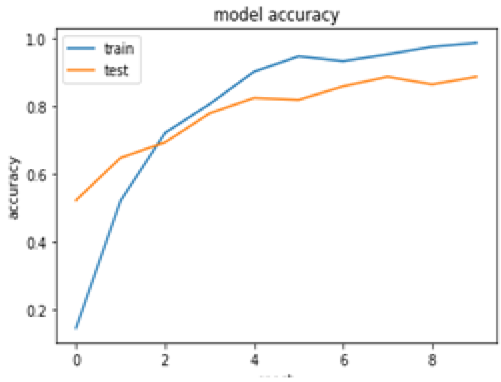

Fig 3. Accuracy graph of Face model

The food model is also trained on 31 classes with 500 images. It contains images of food like roti, chole, rice, and even fruits like grapes, apples, etc. paper. We have created a database of around 700 images of food as shown in Fig 4. The food model shows 88% accuracy. Fig 5 shows the accuracy graph of the food model.

Fig 4. Data set of images of food

‘

Fig 5.Accuracy Graph of Food model

ISSN: 2233-7853 IJFGCN

V. CONCLUSIONS

Many studies have proven that CNN is among the best algorithms for accurate face and food recognition. The proposed system recognizes the face and food with a great accuracy. Pre-trained models are used for feature extraction as it is one of the crucial steps in recognition process. Our neural networks have performed very well and shown accuracy of 91% for the face model and 88% for the food model.

REFERENCES

[1] S. V. Tathe ,A. S. Narote,S. P. Narote, " Face Detection and Recognition in Videos", IEEE. Bangalore Section, Institute of Electrical and Electronics Engineers. India Council, & Institute of Electrical and Electronics Engineers. (n.d.). *2016 IEEE Annual India Conference (INDICON) : 16-18 Dec. 2016.*

[2] V, T., Vellore Institute of Technology, Institute of Electrical and Electronics Engineers. Madras Section, IEEE Communications Society., & Institute of Electrical and Electronics Engineers. (n.d.). *Conference proceedings, International Conference on Vision Towards Emerging Trends In Communication and Networking (ViTECoN 2019) : 30-31, March 2019, Vellore, Tamilnadu, India*.

[3] Shreyak Sawhney, Karan Kacker,Samyak Jain, Shailendra Narayan Singh, Rakesh Garg,"Real-Time Smart Attendance System using Face Recognition Techniques" 9th International Conference On Cloud Computing, Data Science and Engineering : Confluence 2019 : 10-11 January 2019, Uttar Pradesh, India.(Shah et al., 2018)

[4] Di Wang , Ding Wang , Hongzhi Yu , Guanyu Li " Face Recognition System Based On CNN" 978-1-7281-9837-8/20/$31.00 ©2020 IEEE

[5] Qu, X., Wei, T., Peng, C., & Du, P. (2018). A Fast Face Recognition System Based on Deep Learning. Proceedings - 2018 11th International Symposium on Computational Intelligence and Design, ISCID 2018, 1, 289–292. https://doi.org/10.1109/ISCID.2018.00072.

[6] Gui Wu1, Jun Tao, Xun Xu " Occluded Face Recognition Based on the Deep Learning "IEEE *31st Chinese Control and Decision Conference (2019 CCDC) : 3-5 June, 2019, Nanchang, China*.

[7] M. D. Putro, Wahyono and K. -H. Jo, "Multiple Layered Deep Learning Based Real-time Face Detection," 2019 5th International Conference on Science and Technology (ICST), Yogyakarta, Indonesia, 2019, pp. 1-5.doi: 10.1109/ICST47872.2019.9166172

[8] Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M., & Cagnoni, S. (2016). Food image recognition using very deep convolutional networks. *MADiMa 2016 - Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, Co-Located with ACM Multimedia 2016*, 41–49. https://doi.org/10.1145/2986035.2986042

[9] Pandey, P., Deepthi, A., Mandal, B., & Puhan, N. B. (2017). FoodNet: Recognizing Foods Using Ensemble of Deep Networks. *IEEE Signal Processing Letters*, *24*(12), 1758–1762. https://doi.org/10.1109/LSP.2017.2758862

[10] Pouladzadeh, P., Kuhad, P., Peddi, V. B., Yassine, A., & Shirmohammadi, S. (n.d.). *Food Calorie Measurement Using Deep Learning Neural Network*.

[11] Zhang, W., Zhao, D., Gong, W., Li, Z., Lu, Q., & Yang, S. (2015). Food Image Recognition

ISSN: 2233-7853 IJFGCN

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)*, 690–693. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.139

[12] *SRIGURULEKHA. K, Mr.V.RAMACHANDRAN, Food image recognition using CNN, 978-1-7281-4514-3/20/$31.00© 2020 IEEE.*