International Research Journal of Education and Technology
ISSN 2581-7795

AUTOMATED FOOD IMAGE CLASSIFICATION USING DEEP LEARNING APPROACH

USHA SREE JAKKA[1],THUMMAALA BHAGAVAN REDDY[2], PANDITI AMULYA[3],YERRAMIYA SHAIK YUNUS AHMED[4]

[1](BTech, ECE), Dept of Electronics and Communications Engineering, Sree Vidyanikethan Engineering College, Tirupathi
[2](BTech, ECE), Dept of Electronics and Communications Engineering, Sree Vidyanikethan Engineering College, Tirupathi
[3](BTech, ECE), Dept of Electronics and Communications Engineering, Sree Vidyanikethan Engineering College, Tirupathi
[4](BTech, ECE), Dept of Electronics and Communications Engineering, Sree Vidyanikethan Engineering College, Tirupathi

---

Abstract -*Growing importance in the health and medical fields, food image categorization is a new research subject. Automated food recognition techniques will undoubtedly aid in the development of diet monitoring systems, calorie estimation, and other similar applications in the future. Automated food classification methods based on deep learning algorithms are discussed in this research. For food image classification, SqueezeNet and VGG-16 Convolutional Neural Networks are utilized. It was shown that applying data augmentation and fine-tuning the hyper parameters improved the performance of these networks, making them appropriate for practical applications in the health and medical domains. Because SqueezeNet is a lightweight network, it is easy to set up and maintain. SqueezeNet can attain a high level of accuracy even with fewer parameters. Extraction of complex features from food photographs improves the accuracy of food image classification. The suggested VGG-16 network improves the performance of automatic food image classification. The planned VGG-16 has improved significantly as a result of increased network depth*

***Key Words*:Food Classification, Image processing, Squeeze Net, VGG-16 Network, Transfer learning.**

1.Introduction

It's no secret that the amount of calories people eat and drink has a direct impact on their weight: Consume the same number of calories that the body burns over time, and weight stays stable. Consume more than the body burns, weight goes up. Less, weight goes down.

1.1 Deep Learning

Deep learning was first theorized in the 1980s, there are two main reasons it has only recently become useful:

1. Deep learning requires large amounts of **labelled data**. For example, driverless car development requires millions of images and thousands of hours of video.
2. Deep learning requires substantial **computing power**. High-performance GPUs have a parallel architecture that is efficient for deep learning. When combined with clusters or cloud computing, this enables development teams to reduce training time for a deep learning network from weeks to hours or less.

Most deep learning methods use **neural network** architectures, which is why deep learning models are often referred to as **deep neural networks**. One of the most popular types of deep neural networks is known as convolutional neural networks **(CNN** or **ConvNet)**. A CNN convolves learned features with input data, and uses 2D convolutional layers, making this architecture well suited to processing 2D data, such as images. CNNs eliminate the need for manual feature extraction, so you do not need to identify features used to classify images. The CNN works by extracting features directly from images[1].

1.2 CNN Layers

1.Convolution layers: Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus. Each convolutional neuron processes data only for its receptive field. Although fully connected

feedforward neural networks can be used to learn features and classify data, this architecture is generally impractical for larger inputs such as high-resolution images[2-3].

2.Pooling layers:

Convolutional networks may include local and/or global pooling layers along with traditional convolutional layers. Pooling layers reduce the dimensions of data by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Local pooling combines small clusters, tiling sizes such as 2 x 2 are commonly used. Global pooling acts on all the neurons of the feature map. There are two common types of pooling in popular use: max and average.

3.Fully connected layers

Fully connected layers connect every neuron in one layer to every neuron in another layer. It is the same as a traditional multi-layer perceptron neural network (MLP). The flattened matrix goes through a fully connected layer to classify the images.

2.Literature Survey

Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y: In this, Deep learning has been proved to be an advanced technology for big data analysis with a large number of successful cases in image processing, speech recognition, object detection, and so on. In this, we provided a brief introduction of deep learning and detailed described the structure of some popular architectures of deep neural networks and the approaches for training a model. The result of our survey on various aspects indicates that deep learning outperforms other methods such as manual feature extractors, conventional machine learning algorithms, and deep learning as a promising tool in food quality and safety inspection. The encouraging results in classification and regression problems achieved by deep learning will attract more research efforts to apply deep learning into the field of food in the future.[8]
Summary: Brief introduction to deep learning algorithms.
Farinella, G. M., Moltisanti, M., &Battiato, S: In this ,the classification of food images is an interesting and challenging problem since the high variability of the image content which makes the task difficult for current state-of the-art classification methods. Images are processed with a bank of rotation and scale invariant filters and then a small codebook of Textons is built for each food class. The learned class-based Textons are hence collected in a single visual dictionary. The food images are represented as visual words distributions (Bag of Textons) and a Support Vector Machine is used for the classification stage.[9]
Summary: About Texture features, SVM classifier, Food classification.
Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A: In this, Scene recognition is one of the hallmark tasks of computer vision, allowing definition of a context for object recognition. Current deep features trained from ImageNet are not competitive enough. Here, we introduce a new scene-centric database called Places with over 7 million labeled pictures of scenes. We propose new methods to compare the density and diversity of image datasets and show that Places is as dense as other scene datasets and has more diversity. Using CNN, we learn deep features for scene recognition tasks, and establish new state-of-the-art results on several scene-centric datasets.[10]
Summary: About Convolutional Neural Network, High level features.
Rahmani, G. A: In this paper, for each segment obtained from the MS, a color feature vector and several texture representatives are obtained in previous section. The pairwise affinities have been computed with a different approach and a combination of texture and color features have been used as similarity graph in spectral clustering.[11]
Summary: About Segmentation, color space, Gabor filter

3. Proposed Method

Below figure shows the model diagram of our proposed method.

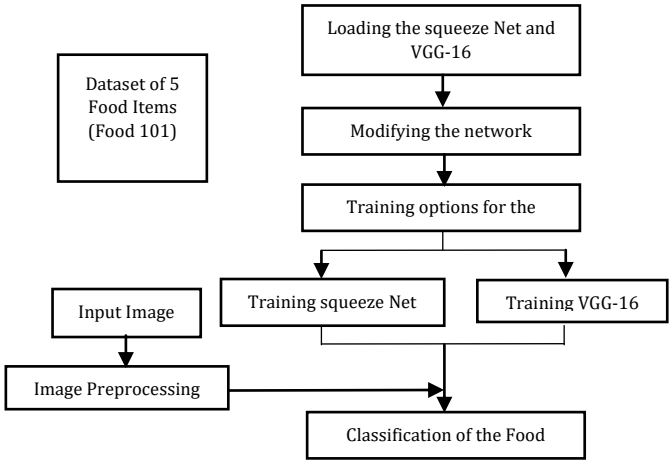

Fig-1 : The block diagram of the proposed method
Dataset: We obtained the dataset Food 101 from the Kaggle public dataset. The dataset consists of 101 classes of food.

Squeeze Net: The main ideas of SqueezeNet are:

- Using 1x1(point-wise) filters to replace 3x3 filters, as the former only 1/9 of computation.
- Using 1x1 filters as a bottleneck layer to reduce depth to reduce computation of the following 3X3 filters.
- Down sample late to keep a big feature map.

International Research Journal of Education and Technology
ISSN 2581-7795

VGG16 Net: The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3 (which is the smallest size to capture the notion of left/right, up/down, centre). In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e., the padding is 1-pixel for 3×3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2-pixel window, with stride 2[4-7].

CNN Layers: In a regular Neural Network, there are three types of layers:

1.Input Layer: It's the layer in which we give input to our model. The number of neurons in this layer is equal to total number of features in our data.

2.Hidden Layer: The input from Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layer can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by addition of learnable biases followed by activation function which makes the network nonlinear.

3.Output Layer: The output from the hidden layer is then fed into a logistic function like sigmoid or softmax which converts the output of each class into probability score of each class.

Padding:

To handle the edge pixels there are several approaches:

- Losing the edge pixels
- Padding with zero value pixels
- Reflection padding.

Stride:

It is common to use a stride two convolution rather than a stride one convolution, where the convolutional kernel strides over 2 pixels at a time, for example our 3x3 kernel would start at position (1, 1), then stride to (1, 3), then to (1, 5) and so on, halving the size of the output channel/feature map, compared to the convolutional kernel taking strides of one.

4.Results

Fig-2 : Input image

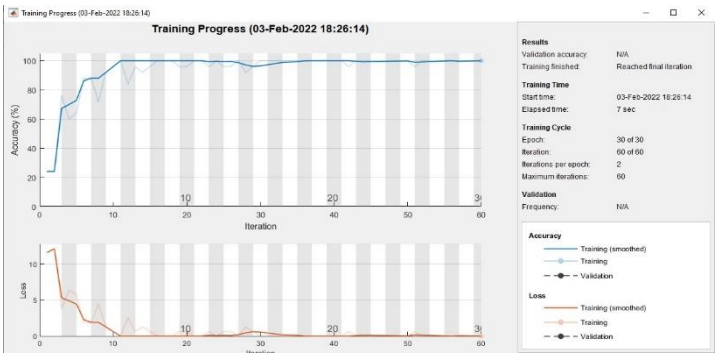

Fig-3 : Training progress of SqueezeNet

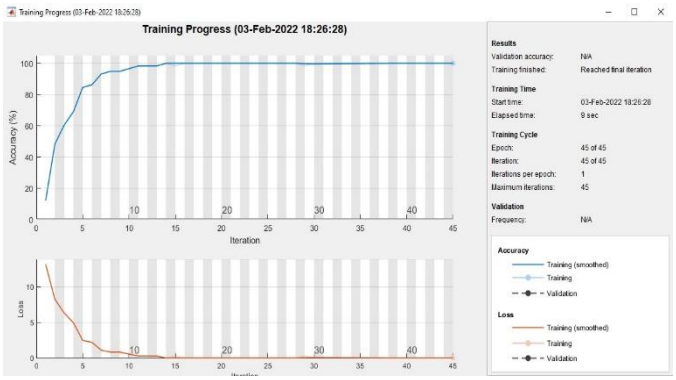

Fig-4 : Training progress of VGG 16 Net

Fig-5 : Accuracy details

5. CONCLUSION

In this study, we proposed deep learning algorithms which are SqueezeNet and VGG 16 Net which are the neural networks for the task of food classification was successfully completed with the better accuracy. The system is able to classify the food contained image from the provided 5 classes dataset.

6. Future Scope

The future scope for the project is to apply the system with larger datasets (Thousands of Images) of food and develop an application for calorie calculation of the input food image, the vitamins that are presented in the food etc.

7. References

[1] Tallapragada, V.S., Kumar, G.P., Reddy, D.V. and Narasihimhaprasad, K.L., 2021. Image Denoising Using Low Rank Matrix Approximation in Singular Value Decomposition. *REVISTA GEINTEC-GESTAO INOVACAO E TECNOLOGIAS*, *11*(2), pp.1430-1446.

[2] D.Venkat Reddy, V. V. Satyanarayana Tallapragada*, K. Raghu, M. Venkat Naresh. (2020). Hybrid Tone Mapping with Structural and Edge-preserving Priors. International Journal of Advanced Science and Technology, 29(7), 5135-5143. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/23592

[3] Tallapragada, V.V., Manga, N.A., Kumar, G.V. and Naresh, M.V., 2020. Mixed image denoising using weighted coding and non-local similarity. SN Applied Sciences, 2(6), pp.1-11.

[4] Satyanarayana Tallapragada VV, Potlabathini R, Narmada A (2019)Rain streak removal using sparse coding. Int J Sci Technol Res8:1828–1833

[5] Tallapragada, V.V., Alivelu Manga, N., Nagabhushanam, M.V. and Venkatanaresh, M., 2022. Greek Handwritten Character Recognition Using Inception V3. In *Smart Systems: Innovations in Computing* (pp. 247-257). Springer, Singapore.

[6] Satyanarayana Tallapragada, V.V., Bhaskar Reddy, B., Ramamurthy, V. and Sunkara, J.K., 2020. Effective Compression of Digital Images Using SPIHT Coding with Selective Decomposition Bands. In *Advances in Electrical and Computer Technologies* (pp. 955-961). Springer, Singapore.

[7] Tallapragada, V.V., Kullayamma, I., Kumar, G.V. and Venkatanaresh, M., 2022. Significance of Internet of Things (IoT) in Health Care with Trending Smart Application. In *Smart Systems: Innovations in Computing* (pp. 237-245). Springer, Singapore.

[8] Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y, "Application of Deep Learning in Food: A Review," Comprehensive Reviews in Food Science and Food Safety, vol. 18, pp. 1793-1811, 2019.

[9] Farinella, G. M., Moltisanti, M., &Battiato, S., "Classifying food images represented as Bag of Textons," IEEE International Conference on Image Processing (ICIP), Paris, pp. 5212-5216, doi: 10.1109/ICIP.2014.7026055, 2014.

[10] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A., "Learning deep features for scene recognition using places database," Proceedings of the 27th International Conference on Neural Information Processing Systems, vol. 1, pp. 487-495, ACM, 2014.

[11] Rahmani, G. A.,"Efficient Combination of Texture and Color Features in a New Spectral Clustering Method for PolSARImage Segmentation" National Academy Science Letters, vol. 40, pp. 117-120, 2017, https://doi.org/10.1007/s40009-016-0513-6.