International Journal for Modern Trends in Science and Technology
Volume 10, Issue 03, pages 195-197.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol10issue03.html
DOI: https://doi.org/10.46501/IJMTST1003031

Comparing VGG-16 and Squeeze Net in Automated Food Image Classification: A Deep Learning Perspective

Dr. Arvind S[1], Dr. Devika SV[2], V Moshe Rani[2], Kondala Rao[2]

[1]Department of CSE, Hyderabad Institute of Technology and Management, Hyderabad, India
[2]Department of ECE, Hyderabad Institute of Technology and Management, Hyderabad, India

To Cite this Article
Dr. Arvind S, Dr. Devika SV, V Moshe Rani, Kondala Rao, Comparing VGG-16 and SqueezeNet in Automated Food Image Classification: A Deep Learning Perspective, International Journal for Modern Trends in Science and Technology, 2024, 10(03), pages. 195-197.https://doi.org/10.46501/IJMTST1003031

Article Info
Received: 02 February 2024; Accepted: 26 February 2024; Published: 03 March 2024.

ABSTRACT

In contemporary times, the significance of food image classification has seen a rapid surge, particularly within the realms of health and medical research. A noteworthy trend in applied research involves the categorization of food images, offering novel perspectives and insights. The automated classification and categorization of food hold considerable promise in monitoring daily dietary habits and caloric intake. This classification is facilitated by leveraging deep learning algorithms and Convolutional Neural Networks (CNN).

In the realm of CNN, two key networks, VGG-16 and SqueezeNet, play pivotal roles in this automated food image classification. While both networks are pre-trained, they differ in the number of layers they encompass. SqueezeNet boasts 18 layers, providing a distinctive architecture, while VGG-16 comprises 16 layers. Employing these networks through MATLAB, we conducted a comprehensive analysis, yielding noteworthy results that exhibit variations between the two networks. This research contributes to the evolving landscape of automated food image classification, showcasing the potential of deep learning algorithms and CNN architectures in advancing our understanding of dietary patterns and nutritional monitoring.

Keywords - Food Classification, Deep learning, Convolution Neural Networks, Squeeze Net, VGG-16 Network

1. INTRODUCTION

The Automatic food recognition is an emerging research topic not only for the social network domain aspect. Indeed, researchers are focusing on this area because of its increasing
benefits for medical point of view. Automatic food recognition tools will help in facilitating the decision-making process of calories estimation, quality detection of food, build diet monitoring systems to combat obesity and so on. On the other hand, food is inherently deformable and shows high divergence in appearance. Since food images have high intra-class variance and low inter-class variance due to which classic approaches do not recognize complex features. This makes food recognition a difficult task for which complex features are not recognized by classic approaches. CNNs can easily identify these features automatically, thus increasing classification accuracy. Therefore, this paper attempts to use CNNs for food image classification

2. Deep Learning

Deep learning is a subset of machine learning techniques which teaches computers in the humanly way of learning. With the usage if Deep learning everything around us can be automated, considering an example where it can be applied in automobiles to scan the things and road signs present around the vehicle and intimating the passengers present inside the automobile. Deep learning is having multiple applications where human mind is involved in the process.

Since the bulk of deep learning approaches use neural network topologies, deep learning models are occasionally referred to as deep neural networks. The word "deep" often denotes how many hidden layers are present in the neural network. Traditional neural networks typically have two or three layers, whereas deep networks can have as many as 150 hidden layers. Deep learning models are trained using massive labelled data sets and automatically learning neural network topologies that extract features from the data.

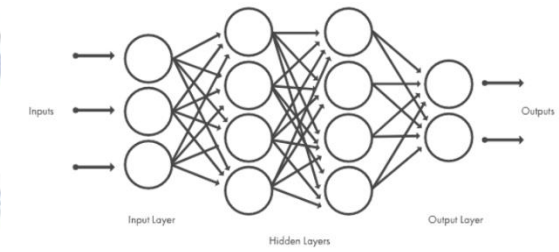

Fig 1: Neural Networks

Convolution Neural Networks

One of the most popular types of deep neural networks is convolutional neural networks (CNN or ConvNet). The architecture of a CNN is appropriate for processing 2D data, such as images, because it uses 2D convolutional layers and convolves learned features with input data. Because CNNs carry out the manual feature extraction for you, you do not need to be familiar with the features that are used to categorize images. CNN uses direct feature extraction from images. When the network is trained on a series of images, the relevant features are not pre-learned; rather, they are found. Because of this automatic feature extraction, deep learning models are incredibly accurate for computer vision applications like object categorization.

CNNs can learn to recognize numerous elements of a picture by using tens or hundreds of hidden layers. The learned picture properties become more complex with each buried layer. For instance, the first hidden layer might learn how to recognize edges, while the final one might learn how to recognize more intricate shapes tailored to the geometry of the object we're trying to identify.

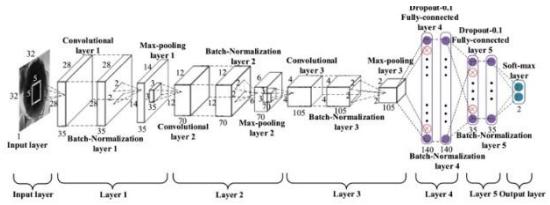

Fig 2: Layers of CNN

Convolutional kernels are a group of filters that make up each convolutional layer. The kernel is the same size as the filter, which is an integer matrix applied to some of the input pixel values. To make things simpler, each pixel is multiplied by the kernel value that corresponds to it, and the result is then summed to form a single value that, much like a pixel, represents a grid cell in the output channel/feature map. A subtype of affine functions, convolutions are all linear transformations. RGB images with three channels are frequently utilised as computer vision input.

3. Proposed Methodology

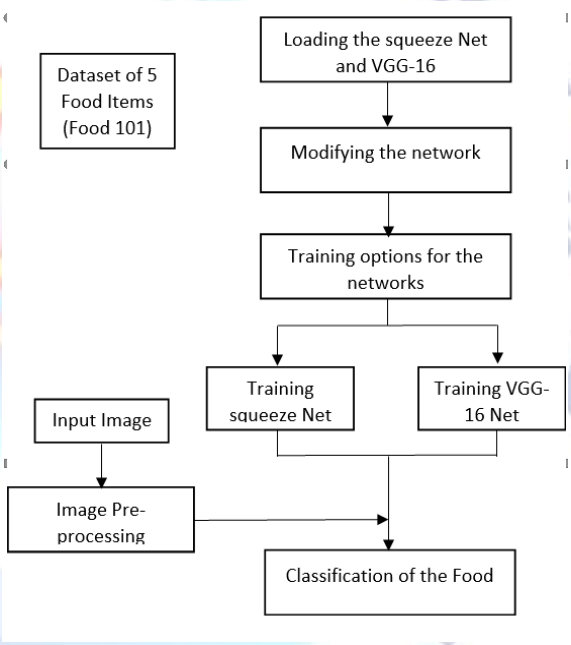

Fig 3: Flow Chart

Training Methods

We are using two networks in the training progress

3.1 Squeeze Net

A convolutional neural network of 18 layers deep is called SqueezeNet. The Image Net database contains a pre-trained version of the network that has been trained on more than a million photos. The pre-trained network can categorise photos into 1000 object categories, including several animals, a keyboard, a mouse, and a pencil. SqueezeNet's fundamental concepts are:

1. Replacing 3x3 filters with 1x1(point-wise) filters since the former require just 1/9 of the calculation.
2. Using a bottleneck layer of 1x1 filters to reduce depth and the subsequent 3x3 filters' computation time.

3. Keep a large feature map by down sampling rate.

Fig 4: Squeeze Net training progress

3.2 VGG 16

VGG_16 is a type of Convolution Neural Network(CNN). Typically VGG-16 has 16 layers of network and VGG refers to Visual Geometry Group. The term "deep" refers to the quantity of layers with VGG-16 or VGG-19 having 16 or 19 layers, respectively. The ImageNet database contains a pretrained version of the network that has been trained on more than a million photos. As VGG-16 is a pretrained network it has the capability to classify and categorize the photos into 700+ object categories which can include multiple character types like Animals, Home accessories , Food items etc.The network accepts RGB images with a resolution of 224 x 224.

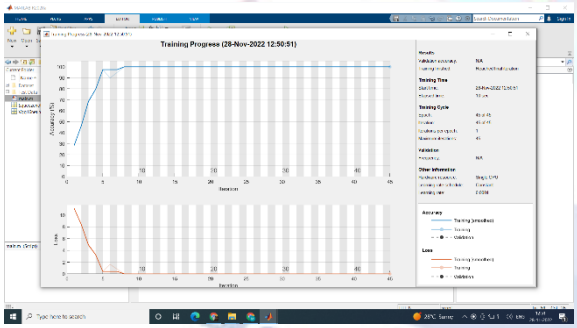

Fig 5: VGG 16 Training Progress

4. Conclusion

This paper presents two deep learning algorithms—Squeeze Net and VGG 16 Net—that use neural networks to successfully perform the task of classifying foods with more accuracy. The system can identify the food-related image from the input dataset of five classes (as per requirement number of classifications). By extracting high-level complicated features, classification of food images performed better. SqueezeNet and VGG-16 deep learning models have been applied for this. To enhance network performance, data augmentation techniques and hyper parameters were used in the design of these networks. SqueezeNet, which had a smaller model and fewer parameters, was found to perform well, with an accuracy of 85% to 89% approx. The proposed VGG-16 is a deeper, more parameterized network than SqueezeNet. As a result, the proposed VGG-16 had substantially better performance and could classify food photos with an accuracy of 93% to 97% approx.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] 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.

[2] 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.

[3] 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.

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

[5] Brownlee, J., "A Gentle Introduction to the RectifiedLinear Unit (ReLU)," retrieved January 24, 2021, from MachineLearningMastery:https://machinelearningmastery.com/rectified-linearactivation-function-for-deep-learning-neural

[6] Chaib, S., Liu, H., Gu, Y., & Yao, H., "Deep Feature Fusionfor VHR Remote Sensing Scene Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, pp. 4775-4784, 2017, doi:10.1109/TGRS.2017.2700322.

[7] Simard, P. Y., Steinkraus, D., & Platt, J. C., "Best Practices for Convolutional Neural Networks," 12th International Conference on Document Analysis and Recognition, vol. 2. IEEE Computer Society, 2003.

[8] Bazargani, Anjos, M. &., Lobo, A. & Mollahosseini, F. &,Shahbazkia, A. &, & Hamid, "Affine Image Registration Transformation Estimation Using a Real Coded," Proceedings of the 14th annual conference companion on Genetic and evolutionary computation, pp. 1459-1460, ACM, 2012, https://doi.org/10.1145/2330784.2330990)