The Impact of Different Deep Learning Techniques on Food classification

[1] Mrs.K.Priyadharsini
Assistant Professor,
Department of Computer Science
Athoor Cooperative Arts & Science College,Dindigul.

[2] Dr.A.Shanthasheela
Assistant Professor
Department of Computer Science
M.V.Muthiah Arts College,Dindigul

Abstract— The condition known as obesity[1] is brought on by an individual's high muscle to fat ratio. It is a sudden and unusual increase in the ratio of muscle to fat. It can cause hypertension, cholesterol, pulse, heart-related disorders, and other medical issues. Maintaining a food journal is crucial to leading a healthy life and preventing and managing obesity. Computer vision technology were applied to food logging to automate image categorization for nutritional intake tracking. Strong deep learning methods for both generative and descriptive tasks are convolution neural networks. The Food101 dataset is used in this paper's performance comparison of the Convolution Neural Network designs ResNet50, DenseNet-161, and Inception-V3. Keywords: Food Recognition; Food Identification; DenseNet-161, Inception-V3, ReseNet-50; Comparative study; deep learning models;

I.INTRODUCTION

The need of sustaining healthy eating habits has become more widely recognized due to the growth of nutrition-related diseases worldwide. A nutritious diet lowers the risk of food intolerance reactions, weight issues, malnutrition, and certain malignancies. We can manually keep track of what we eat and identify food items prior to consumption using a variety of tools. However, many applications require prior knowledge of the food item for easy identification. Automated food identification methods will come in handy here. One of the main issues with many real-world applications is image classification. Machine learning has received a lot of attention, particularly in relation to neural networks like the Convolutional Neural Network (CNN), which has won image classification competitions [2]. The convolutional layers are the main building blocks of convolutional neural networks. Input vectors, such as an image, filters, such as a feature detector, and output vectors, such as a feature map, are frequently found in this layer. After passing through a convolutional layer, the image is distracted to a feature map, also known as an activation map. Convolutional layers convolve the input before passing the output to the next layer. A neuron's reaction to a single stimulus in the visual cortex is akin to this. Each convolutional neuron processes information only for the receptive field to which it is assigned. The optimization of the Convolutional neural network mostly focused on: the design of the Convolutional layer and pooling layer, the activation function, loss function, regularization, and the application of Convolutional neural networks to real situations [3]. With this knowledge, we are going to compare different architecture models for CNN. The rest of the paper is structured as follows. Section II goes through some previous research material. In section III, Food Dataset and Classification Methods are described in two parts: 1. Dataset, 2. Model architectures. In Section IV, Comparative Evaluation of Results, and in Section V conclusions are discussed.

II. LITERATURE REVIEW

[4] used a CNN with Tensor-flow and Keras. The classifier Back Propagation Multi-Label learning is used which improves the accuracy of the unseen instance and the model. It uses 17 assorted food images such as Asian vegetarian meal, Diabetic Meal, Continental meal, etc. The accuracy of this model is 80%.

[5] proposed a convolution neural network with a variation of multilayer perceptron and require minimal preprocessing for images. It uses a Food-11 dataset for training, validating, and testing the model. The proposed approach produces 92.86% accuracy.

In [6] the Middle Eastern Cuisine dataset with 27 classes was used to train the MobileNetV2 [7] model that aims two-fold that's is predicting the type of food and allowing low latency predictions. It is a shallow architecture and can be easily adapted to mobile applications. It has high accuracy compared to traditional models.

[8] proposed a protocol for an automatic food recognition system that identifies the contents of the meal from the images of the food. We developed a multilayered convolutional neural network (CNN) pipeline that takes advantage of the features of other deep networks and improves efficiency.ETH Food-101 and the newly contributed Indian food image database are used to calculate the efficiency of the system.

[9] detects varieties of food and calculates per serving calories of detected food from an input image. It includes four stages.1. Image acquisition –2.Neural Network Training-Convolution Neural Network (CNN) 3. Image segmentation- Morphological functions and OpenCV is used.4. Calorie estimation - mathematical formulas. It uses the Fruits 360 dataset which contains 90483 images of 131 fruits and vegetables.

III FOOD DATASET AND CLASSIFICATION METHODS

The methodology section is divided into three parts, these are Dataset, deep learning models, evaluation matrices, and implementation procedures. These three processes are described below:

I. Dataset:

The dataset contains several different subsets of the full Food-101[10] data. It consists of 101 food categories with 750 training and 250 test images per category, making a total of 101k images. The idea is to make a more exciting simple training set for image analysis than CIFAR10 or MNIST. For this reason, the data includes massively downscaled versions of the images to enable quick tests. The data has been reformatted as HDF5[11] and specifically Keras HDF5Matrix which allows them to be easily read. The file names indicate the contents of the file.

For example

- foodc101n1000_r384x384x3.h5 means there are 101 categories represented, with n=1000 images, that have a resolution of 384x384x3 (RGB, uint8)
- foodtest c101n1000r32x32x1.h5 means the data is part of the validation set, has 101 categories represented, with n = 1000 images, that have resolution of 32x32x1 (float32 from -

1 to 1).

Figure 1:Sample images from Food-101 Dataset

II Model Architecture:

CNNs are a class of Deep Neural Networks that can recognize and classify particular features from images and are widely used for analyzing visual images

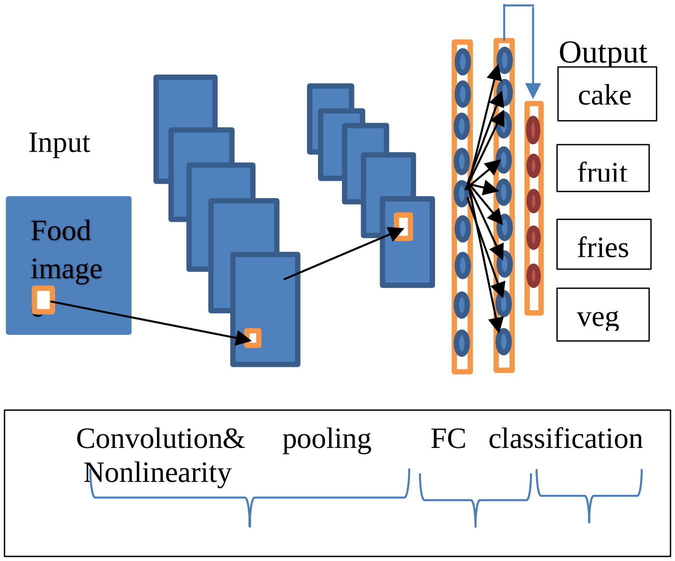

Figure 2:Traditional CNN Architecture

A. DenseNet-161

The variations of CNN architecture is the Dense Convolutional Network (DenseNet-161)[12-14] another state-of-the-art CNN architecture inspired by the cascade-correlation learning[15] architecture proposed in NIPS, 1989. This architecture connects each layer to another layer in a feed-forward method. The normal CNN uses L layers and L connections whereas DenseNet-161 has an L layer and L (L+1) 2 direct connections. All previous layers' feature maps are utilized as inputs into each layer, and their feature maps are used as inputs into all subsequent layers.

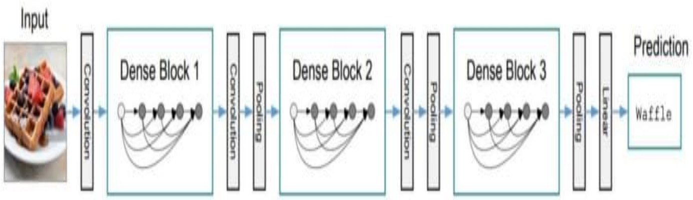

Figure 3: DenseNet-161 Architecture

The assets of Densenet-161 include its ability to address the vanishing gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters.

In DenseNet-161 the training dataset is preprocessed using successive transformations. These transformations are Random rotation, Random resizing crop, Random horizontal flip, Imagenet policy, and finally Normalization in our implementation. These preprocessing transforms are used to reduce image property disparities caused by varied image backgrounds, speed up model learning, and increase output accuracy. The transfer learning method is applied by using the pretrained Densenet-161 model as follows,

- At first, with a pretrained DenseNet-161 model, load a checkpoint. The checkpoints file contains all the tensors after months of training with the ImageNet dataset.

- Secondly, redefined the classifier part of the model (i.e. model. Classifier ()) to fit the number of outputs classes (101) as derived from the input data classes.

This model is evaluated by splitting the dataset into training, test, and validation in the ratio of 8:1:1. That is 80 % of the dataset is used to train and 10% for tests and 10% for validation. The classification accuracy and derived error values are reduced by fine-tuning the network parameters using the Adam optimizer[16] as defined below,

Optimizer=optim.Adam (model.classifier.parameters (), lr=0.001, betas= [0.9, 0.999])

B. ResNet50 model

The ResNet-50[17-19] model is divided into five stages, each with its convolution and identity block. Each convolution block has three convolution levels. There are around 23 million trainable parameters in the ResNet-50.

This model uses the Food-101 dataset. Cropping, padding, and horizontal flipping are standard data augmentation techniques used to train massive neural networks. However, most neural network training methods only employ basic types of augmentation Before training the model, random transforms such as flip, wrap, rotate, zoom, lighting, and contrast randomly are done. Random transformation [20] will increase the variety of image samples and prevents overfitting. some of the random transform codings for the traininareset are given below.

ResNet50 is a derived version of the ResNet[21] version which has 48 Convolution layers in conjunction with 1 MaxPool and 1 Average Pool layer. The architecture has 4 stages. The network can handle images with height and width multiples of 32 and a channel width of 3. Every ResNet design uses 7x7 and 3x3 kernel sizes for initial convolution and max-pooling, respectively. After that, Stage 1 of the network begins, which consists of three Residual blocks, each with three layers. The kernels used to perform the convolution process in all three layers of stage 1's block are 64, 64, and 128 bits in size, respectively.

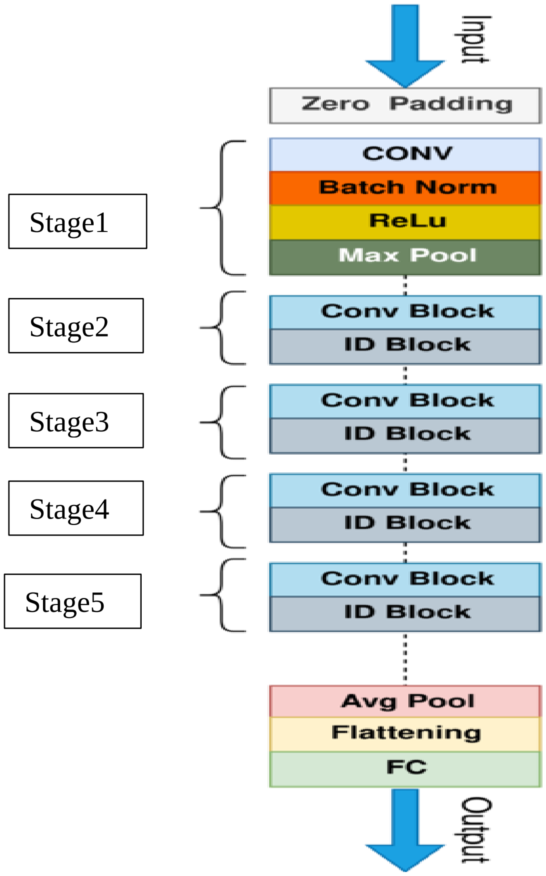

Figure 4: ResNet-50 Architecture

The convolution operation in the residual block will reduce the size of the input in terms of height and width. As moves on from one stage to the next stage the channel width is doubled and size is reduced to half. Three layers are placed one on top of the other for each residual function F. The 3 layers are 1x1,3x3 and 1x1 Convolutions. The 1x1 convolution layers are in charge of shrinking and then expanding the dimensions. Finally, the network has a fully linked layer with 1000 neurons, followed by an Average Pooling layer.

C. Inception-V3:

The Inception-V3[22-24] model consists of 42 layers with less low error rate. The efficiency of this model is relatively the best. The major variations done on the Inception-V3 model are

1. Factorization into Smaller Convolutions
2. Spatial Factorization into Asymmetric Convolutions.
3. Utility of Auxiliary Classifiers.
4. Efficient Grid Size Reduction.

The 5X5 convolution layer is replaced by two 3X3 layers considering its expensiveness. To make the model more efficient Asymmetric convolutions were used with nx1 form. Auxiliary classifier[25] improves the convergence of very deep neural networks. The activation dimension of the network filters is expanded to reduce the grid size efficiently.

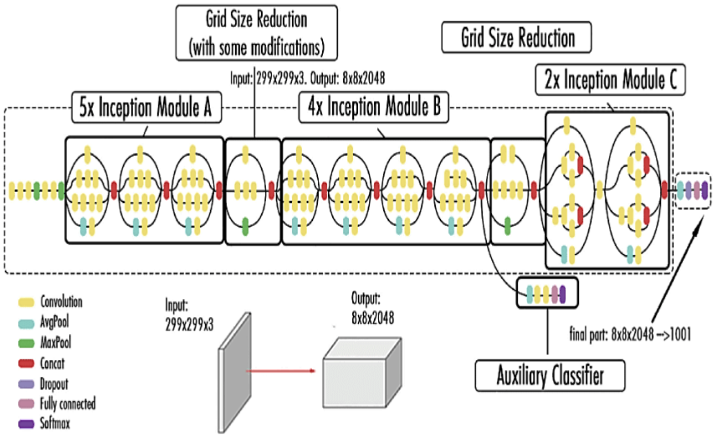

Figure 5:InceptionV3 Architecture

Initially Inception[26] model uses three classes of food from Food-101.Then random images from 11 classes are taken trained. The coding for random selection images to train the model is given below.

```
def pick_n_random_classes(n):
  random.seed(9000)
  food_list = []
    random_food_indices = random.sample(range(len(foods_sorted)),n)
  for i in random_food_indices:
    food_list.append(foods_sorted[i])
  food_list.sort()
  return food_list
```

IV COMPARATIVE EVALUATION OF RESULTS

All the Three models that are given in section III are developed in Python and the results are presented and analyzed in this section. For the classification of ion Food-101, the dataset is used to analyze the efficiency of the models DenseNet-161, ResNet,-50, and Inception-V3.Food-101 is a challenging data collection consisting of 101 food categories and 101'000 photos. 250 manually approved test photos and 750 training images are provided for each class. The dataset was divided into three phases: training, validation, and evaluation. The accuracy of the model is evaluated by true positive (TP), true negative (TN), false positive (FP), and false-negative (FN) after classification[27].

$$Accuracy = \frac{TP + TN}{TP + +FP + TN + FN}$$

Hence before training a model the images should undergo preprocessing. The parameters used by the models are listed in Table 1.

Model	**Parameters**
DenseNet-161	**23M**
ResNet50	**25.8M**
InceptionV3	**26.7M**

Table 1: Parameters for deep learning models.

The Deep learning models DenseNet-161, ResNet-50and Inception-V3 against the dataset Food-101 are compared based on the accuracy.

From the findings, it has been stated that DenseNet-

161 has the highest accuracy of 93.3% for Top-1 and 99.0 % for Top-5. The accuracy of the models is listed in Table 2.

Model	**Top-1 %**	**Top-5 %**
DenseNet-161	**93.3**	**99.0**
ResNet50	**88.4**	**97.8**
InceptionV3	**88.3**	**96.9**

Table2: Accuracy of the models with food-101

V.CONCLUSION

Due to the existence of complex and varied features in the same class, recognizing food images is a difficult task. Multiple aspects, such as the ingredients used, the cooking methods used, the shapes utilized, and others, make it difficult to recognize a food image because there are so many variations of the same item. Deep learning algorithms' recent breakthroughs have attracted interest from a variety of disciplines, including food identification and recognition. Previous research on the food recognition problem yielded encouraging findings, leading to advancements in the field.

On the food recognition task, this research compares the results of the last generation and some untested deep learning systems. In light of past research, the Inception-V3, ResNet-50, and Densenet-161 are evaluated. Experiments have shown that the contributions of deep learning models to the outcomes are apparent in food recognition tasks. According to our results, DenseNet-161 has produced the comparatively best accuracy.

References:

[1] Hruby, Adela, and Frank B Hu. The Epidemiology of Obesity: A Big Picture. *PharmacoEconomics* vol. 33,7 (2015): 673-89. DOI:10.1007/s40273-014-0243-x.

[2] Hussain, M., Bird, J.J., Faria, D.R. (2019). A Study on CNN Transfer Learning for Image Classification. In: Lotfi, A., Bouchachia, H., Gegov, A., Langensiepen, C., McGinnity, M. (eds) *Advances in Computational Intelligence Systems. UKCI 2018. Advances in Intelligent Systems and Computing*, vol 840. Springer, Cham. https://doi.org/10.1007/978-3-319-97982-3_16.

[3] T. Guo, J. Dong, H. Li and Y. Gao, Simple convolutional neural network on image classification, (*2017). IEEE 2nd International Conference on Big Data Analysis (ICBDA), 2017, pp. 721-724,* DOI: 10.1109/ICBDA.2017.8078730.

[4] Andrews Samraj; Sowmiya D.; Deepthisri K.A.; Oviya R.; (2020). Food Genre Classification from Food Images by Deep Neural Network with Tensorflow and Keras. *2020 Seventh International Conference on Information Technology Trends (ITT),*. DOI: 10.1109 /ITT51279.2020.9320870.

[5] M. T. Islam, B. M. N. Karim Siddique, S. Rahman, and T. Jabid, "Image Recognition with Deep Learning," *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2018*, pp. 106-110, doi: 10.1109/ICIIBMS.2018.8550021.

[6] Şeymanur Aktı, Marwa Qaraqe, Hazım Kemal Ekenel,(2022). A Mobile Food Recognition System for Dietary Assessment", *arXiv*.

[7] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.2018: Mobilenetv2: Inverted residuals and linear bottlenecks. In*: Computer Vision and Pattern Recognition.*

[8]P. Pandey, A. Deepthi, B. Mandal, and N. B. Puhan, Foodnet: Recognizing foods using an ensemble of deep networks, *IEEE Signal Processing Letters, vol. 24, no. 12, pp. 1758–1762, Dec. 2017, ISSN: 1558-2361*. DOI: 10.1109/LSP.2017.2758862.

[9] Harshitha S V, Dhanalakshmi S, Mukeshwar Varma D and Mayuri K P. (2020).Food classification and calorie estimation using computer vision techniques.*International*

Journal of Emerging Technologies and Innovative Research, ISSN:2349-5162, Vol.7, Issue pages no.143-14.

[10] Bossard, L., Guillaumin, M., Van Gool, L. (2014). Food-101 – Mining Discriminative Components with Random Forests. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham.*DOI:10.1007/978-3-319-10599-4_29.

[11] Folk, M., Heber, G., Koziol, Q., Pourmal, E., & Robinson, D. (2011, March). An overview of the HDF5 technology suite and its applications. In *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases* (pp. 36-47).

[12] Noh, Kyoung & Choi, Jiho & Hong, Jin & Park, Kang. (2020). Finger-Vein Recognition Based on Densely Connected Convolutional Network Using Score-Level Fusion With Shape and Texture Images, *IEEE Access. PP. 1-1. 10.1109/ ACCESS. 2020. 2996646.*

[13] Challa, Sri Venkata Divya Madhuri & Vaishnav, Hemendra. (2020). Weather Categorization Using Foreground Subtraction and Deep Transfer Learning. 10.1007/978-981-15-2043-3_64.

[14] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[15] Fahlman, S., & Lebiere, C. (1989). The cascade-correlation learning architecture. *Advances in neural information processing systems*, *2*.

[16] Zhang, Z. (2018, June). Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)* (pp. 1-2). IEEE.

[17] Theckedath, D., & Sedamkar, R. R. (2020). Detectingeffectt states using VGG16, ResNet5,0, and SE-ResNet50 networks. *SN Computer Science*, *1*(2), 1-7.

[18] Chu, Y., Yue, X., Yu, L., Sergei, M., & Wang, Z. (2020). Automatic image captioning based on ResNet50 and LSTM with soft attention. *Wireless Communications and Mobile Computing*, *2020*.

[19] Tian, X., & Chen, C. (2019, September). Modulation pattern recognition based on Resnet50 neural network. In *2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP)* (pp. 34-38). IEEE..

[20] Furman, A. (2002). Random walks in groups and random transformations. In *Handbook of dynamical systems* (Vol. 1, pp. 931-1014). Elsevier Science.

[21] Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition*, *90*, 119-133.

[22] Dong, N., Zhao, L., Wu, C. H., & Chang, J. F. (2020). Inception v3-based cervical cell classification combined with artificially extracted features. *Applied Soft Computing*, *93*, 106311.

[23] Tio, A. E. (2019). Face shape classification using inception v3. *arXiv preprint arXiv:1911.07916*.

[24] Barratt, S., & Sharma, R. (2018). A note on the inception score. *arXiv preprint arXiv:1801.01973*.

[25] Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., & Pinheiro, P. R. (2020). Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection. *Ieee Access*, *8*, 91916-91923.

[26] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[27] Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, *161*, 272-279.