Electronics and Computer Science
Faculty of Engineering and Physical Sciences
University of Southampton

Palak Prakash Jain

06/12/2020

Food Detection and Classification for nutrients estimation

Project supervisor: Dr. Jonathan Hare
Second examiner: <2nd Examiner>

A project progress report submitted for the award of
MEng Computer Science

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

ELECTRONICS AND COMPUTER SCIENCE

A progress report submitted for the award of

MEng Computer Science

by Palak Prakash Jain

According to the World Health Organization, more than 1.9 billion adults are overweight and approximately 13% of the adult population is obese[46]. These are alarming figures which prove that there is a need more than ever for us to control and record our eating habits. Mobile applications are also increasingly being used both to share images of food and to record food intake of individuals, resulting in exponential growth in the number of food images on the internet[15]. This has spurred demand for automated and reliable food classification systems.

The purpose of this project is to compare the various robust methods of Object Detection and Classification and apply them to the domain of Food Recognition, by exploring Deep Neural Networks and Traditional approaches. So far, experimentation with various Convolutional Neural Networks, including the Inceptionv3 model has been performed on existing datasets to gain a solid understanding of Deep Learning. A variety of pre-existing datasets have also been selected to train the models.

The plan for the remaining work is to adjust and train various models on the chosen datasets, to obtain highest accuracies on unseen data. The final objective is to create a web application for accurate Food classification.

STATEMENT OF ORIGINALITY

I have read and understood the ECS Academic Integrity information and the University's Academic Integrity Guidance for Students.

I am aware that failure to act in accordance with the Regulations Governing Academic Integrity may lead to the imposition of penalties which, for the most serious cases, may include termination of programme.

I consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to verify whether my work contains plagiarised material, and for quality assurance purposes.

1. I have acknowledged all sources, and identified any content taken from elsewhere.
2. I have not used any resources produced by anyone else.
3. I did all the work myself, or with my allocated group, and have not helped anyone else.
4. The material in the report is genuine, and I have included all my data/code/designs.
5. I have not submitted any part of this work for another assessment.
6. My work did not involve human participants, their cells or data, or animals.

Contents

List of Figures

Nomenclature

CNN	Convolutional Neural Network
DNN	Deep Neural Network
ResNet	Residual Neural Network
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
CV	Computer Vision
SIFT	The scale-invariant feature transform
SURF	The speeded up robust features
KNN	K-Nearest Neighbours
SVM	Support Vector Machine
FAST	Features from Accelerated Segment Test
LBP	Local Binary Patterns
RBF	Radial Basis Function
PFID	Pittsburgh Fast-Food Image Dataset
DL	Deep Learning

Chapter 1

Introduction

1.1 Problem Description

Eating habits directly impact our health and well-being, while ingredients, flavors and recipes shape particular cuisines that build our personal and collective cultural identities[43]. Recent technological advances in CV along with smartphone cameras have powered new uses of technology in the domain of food.

An average physically active person needs about 2,500 calories a day to maintain a healthy weight[36]. Eating large amounts of processed or fast food, however, is a particularly increasing problem other than common undereating/overeating. As a result monitoring food intake is essential to having a healthy lifestyle. Conventional food logging for diet monitoring requires expertise and effort from the user, and is prone to accuracies and forgetting. In contrast an automatic food annotation system could perform automatic analysis, annotation and logging with minimum human intervention.

Food recognition is also perfect for businesses in the travel and hospitality industry. Restaurant review platforms can use Food recognition to categorize user-generated content automatically or manage a library of photos submitted by restaurants[15]. . The model can also make media management much much easier and help surface the right content to users. For customer brands like Pizza Hut etc. accurate Food Recognition models can provide deeper analytical insights into what dishes are most popularly photographed and shared by customers or even the ability to identify and understand eating habits, cuisines and cultures. Moreover, travel recognition models along with food recognition models can categorize which foods are eaten where. The collective knowledge can also be leveraged by recognition models to improve their accuracy.

Reliable food analysis from images is essential for these applications. However, despite remarkable advances in computer vision, Food recognition in the wild still remains a very challenging problem even for humans. We largely rely on contextual and prior information. which makes automatic image-based food recognition a particularly challenging task.

1.2 Project Description

This research paper is a comparison study between the performance of recent Deep Learning and Traditional approach-based models. The DL models are originally trained on the ImageNet dataset and then fine-tuned using a process known as Transfer Learning, to train on food datasets such as Food-11. [42] and Food-101[10]. The study will also delve deeper into the potential for nutrients estimation. All the training and testing is to be conducted using cloud GPUs provided by Google Colab. The best model(s) will be used to create a web application that allows users to input their food pictures to classify them appropriately. The models will be optimised throughout the development and comparison phase to maximise their accuracy rates.

Chapter 2

Background Research and Literature Review

2.1 Traditional approaches

Traditional techniques in Computer Vision have continued to undergo progressive development in recent years despite the ground-breaking results shown by Deep Learning models. Most CV problems are solved by a combination of feature descriptors such as SIFT and SURF and Machine Learning classification algorithms such as SVMs and K-NNs. Problems however, can be solved with much fewer lines of code using traditional CV[35].

The main traditional approaches used for Object Detection and Recognition: SIFT, SURF, FAST, Hough Transforms and Geometric hashing. Most of these algorithms are not class-specific, meaning that their performance is indifferent to the image fed in, whereas in contrast in DL networks, features learnt are specific to the dataset, and will therefore most likely lead to low accuracy on a different training set[33]. For some types of problems, it is obvious to choose the route of traditional approaches, for example to classify two classes of paints in buckets one with yellow paint and the other with red. A DNN will work given that enough data can be collected to train from, however this can be achieved much more easily by simple colour thresholding[35]. There are also instances where DNN performs poorly outside of the training data - the machine may overfit and not be able to generalize for the task at hand, the reason being that DNN has millions of parameters inside of it with complex relationships[28]. Traditional approaches however, are transparent and one can judge whether the solution would work outside of the training environment.

This traditional object detection approach typically constitutes of the following three steps[35]:

1) Informative Region Selection: Objects can be located in very different places in different images, and may have a different ratio and size. Thus, in this part of the algorithm, the object is searched for by scanning the entire image with a multi-scale sliding window.
2) Feature Extraction: Techniques mentioned above(e.g. SIFT) are used to extract the visual features for recognizing the object. All the visual features extracted from the image are known as Bag of Visual Words. These features provide a logical and strong representation of the image.
3) Classification: Target objects from all other categories are classified using classifiers like SVMs to make the representations more logical, informative and hierarchical for visual recognition.

2.2 Deep Learning Approaches

DL is typically used to solve difficult problems (e.g. image colorization, classification, segmentation and detection). Rapid progressions in DL and improvements in device capabilities have improved the performance and cost-effectiveness of further quickened the spread of vision-based applications.

The contents of food dishes are typically deformable objects usually including complex semantics, which makes the task of defining their structure very difficult[26]. DL methods have shown very promising results in such challenges, particularly structures that are built on Transfer Learning. Therefore, it is sensible to present the hypothesis that they are a better technique for food detection and classification in comparison to traditional methods.

2.3 Convolutional Neural Network (CNN)

A convolutional Neural Network (CNN) is a multi-layer deep learning neural network designed to replicate the pattern of neuron connectivity within the human brain by analyzing visual inputs and performing tasks such as image classification, segmentation and object detection.

A CNN is composed of several kinds of layers as depicted in Figure 2.1:[49]

- Convolutional layer: Creates a feature map to extract the high level features by applying a filter that scans the whole image, few pixels at a time.
- Pooling layer: Scales down the amount of information generated by the Convolutional layer for each feature and maintains the most essential information.
- Fully connected input layer: “Flattens” the outputs generated by previous layers to turn them into a single vector that can be used as an input for the next layer.
- Fully connected layer: Applies weights over the result produced by the feature analysis to predict an accurate label.
- Fully connected output layer: Generates the final probabilities to determine a class for the image.

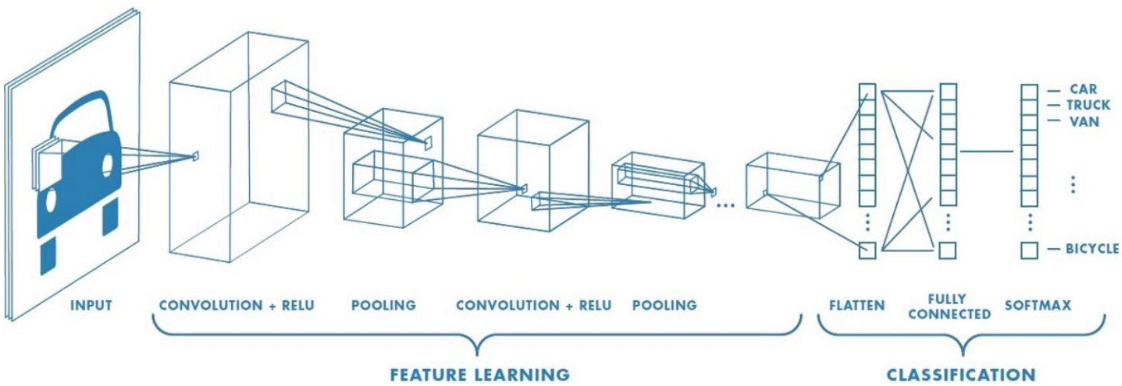


Figure 2.1: A CNN sequence. Sourced from [49].

Transfer Learning is a popular approach in which pre-trained models are used as the starting point on computer vision tasks. The Pre-trained Model Approach of Transfer Learning has been chosen to be deployed in the models of this project, as it is not practical to design Neural Networks in the given time and with current knowledge.

In this approach, a pretrained source model is chosen from pre-existing models which/part of which can be used as a starting point for another task. The model may have to be adapted or refined to input-output pairs for the other task[42].

The architecture of a CNN is a key factor in determining its performance and efficiency. Since 2010, the contenders of the ILSVRC contest have built software programs that attempt to correctly detect and classify objects and scenes within the given 14 million images in the ImageNet dataset[21]. Currently, the challenge uses a cut down list of a thousand separate classes.

When the annual ILSVRC competition began, a good classification rate was 25%, the first major leap in performance was achieved by a network called AlexNet in 2012, which dropped the classification rate by 10%[50]. Over the next years, the error rates dropped to lower percentages and finally exceeded human capabilities. The main DL models released were as follows: LeNet-5(1998), AlexNet(2012) GoogleNet(2014) and VGGNet(2014). Image A depicts the drop in the best classification error rate on the Imagenet dataset from 26% in 2011 (Feature Engineering) to 2.3% (CNNs) in 2017[50].

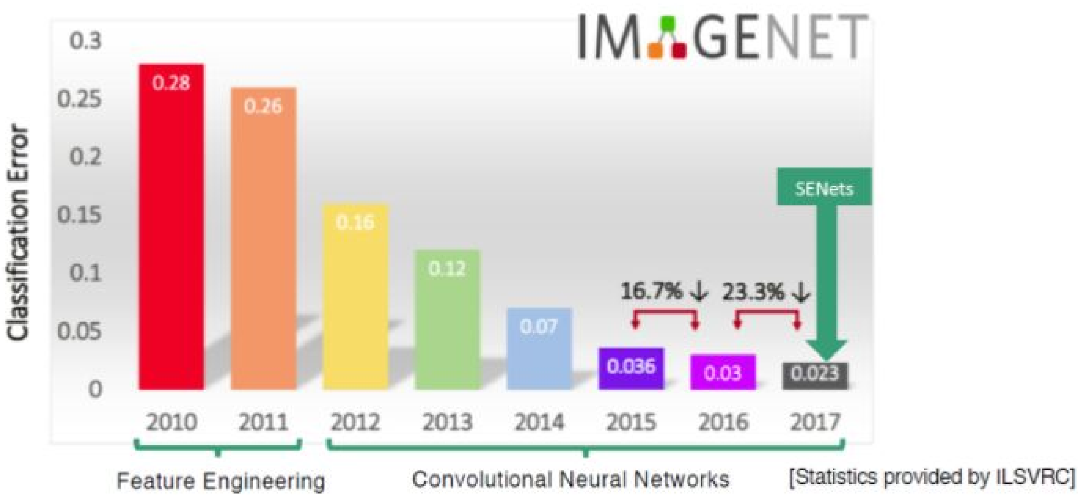


Figure 2.2: Significant decrease in Classification error with CNN in comparison to Feature Engineering. Sourced from [50].

2.4 Existing Implementations

In 2009, a researcher built the PFID composed of 4545 images of 101 different food classes. SVM classifier was applied on this dataset and it achieved a classification accuracy of 11%, with the color histogram method and 24 % with the bag-of-SIFT-features method[56]. Later on, Rashed Mustafa and Prashenjit Dhar proposed a method using Gist and SURF features, which along with the SVM classifier results in 93.3% accuracy, compared to 84.4%, 80.5% and 68.7% using Adaboost, Decision Trees and KNN models respectively, highest recorded using traditional approaches[41].

After reaching almost perfect accuracy results with the binary classification of food, research has progressed into multiclass food classification. Several deep learning-based classifiers were trained using the Food-101 dataset. The most used indicators used are Top-1 and Top-5 classification accuracy. Yanai and Kawano in 2015 used a fine-tuned version of AlexNet and got 70.41% top-1 accuracy[26]. Liu et al. introduced a network named Deep Food that reached 77.40% and 93.70% for Top-1 and Top-5 respectively[28]. Fu, Chen, and Li, in 2017, obtained an even better accuracy of 78.5% and 94.1% for Top-1 and Top-5 using a fine-tuned deep 50-layer ResNet[13]. At the time the classification results of CNN models on the Food-101 dataset were much better than those of traditional methods.

Martinel et al. came to the conclusion that almost all published deep learning models just exploited existing models without specific features of the food being taken into account[31]. As a result, they designed a part convolution unit for extracting common vertical characteristics of food and then added deep residual blocks to make a combination to calculate the classification score, eventually releasing a new CNN structure called WISeR specially for food recognition. Results on the the Food-101, UECFood-256 and UECFood-100 datasets respectively achieved Top-1% and Top-5% rates of 90.27% and 98.71%, 83.15% and 95.45%, and 89.58% and 99.23% respectively[31]. In 2017, Mezgec and Seljak also developed a modified version of the AlexNet architecture, known as ‘NutriNet’ for food and drink classification. They reached an accuracy rate of 86.72% and 94.47% respectively, for the training set and testing set[32] .

The most successful work in Food nutrients estimation was by Parisa Pouladzadeh aet al. which took into account color, size, shape and texture features in the data preparation for SVM, and then applied the RBF kernel, and lastly used SVM for classification[40]. They applied their method to three different categories of food: single, non mixed, and mixed foods, and from the results which shows that the SVMs accuracy is approximately 92.21%, 85%, and 35%–65%, respectively[40].

Chapter 3

Final Design of the System

3.1 Architectures chosen for this project

ResNet (2015)

Microsoft introduced ResNet to improve the training of networks that are substantially deeper than those used previously. Researchers observed that, when training neural networks, there comes a point of saturation where increase in depth causes accuracy to saturate, then degrade rapidly. This is called the degradation problem[51]. In other words, not all neural network architectures are equally easy to optimize. ResNet uses a technique called 'residual mapping' to combat this issue. Instead of hoping that every few stacked layers directly fit a desired underlying mapping, the Residual Network explicitly lets these layers fit a residual mapping. Figure 3.1 shows the building block of a Residual Network[52].

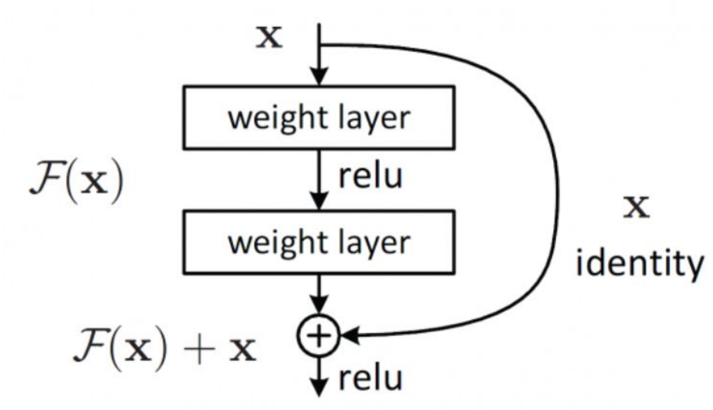


Figure 3.1: "The formulation of F(x)+x can be realized by feedforward neural networks with shortcut connections." Sourced from [54].

Inception

The Inception model, Inceptionv3 in particular, can be distinguished from other models by factorized convolutions, which helps increase the computational efficiency, and network efficiency as it reduces the number of parameters involved in a network. It also replaces bigger convolutions with smaller ones, for faster training. An auxiliary classifier is a small CNN inserted between layers during training, and the loss incurred is added to the main network loss. In GoogLeNet auxiliary classifiers were used for a deeper network, whereas in Inception v3 an auxiliary classifier acts as a regularizer[54]. Moreover, grid size reduction is usually done by pooling operations to combat the bottlenecks of computational costs[53].

Inception-ResNet-v2

The Inception-ResNet-v2 architecture is more accurate than previous state of the art models; it integrates the benefits of both ResNet and Inception models to successfully train deep networks and for enabling significant simplification of the Inception blocks. The table below reports the Top-1 and Top-5 validation accuracies on the ILSVRC 2012 image classification benchmark based on a single crop of the image[55]. Note however, that this new model only requires roughly twice the memory and computation compared to Inception V3[7].

The aim of this project is to reach accuracy level on par with the performance on the ImageNet datasets as shown in the table below:

Model	Architecture	Checkpoint	Top-1 Accuracy	Top-5 Accuracy
Inception-ResNet-v2	Code	inception_resnet_v2_2016_08_30.tar.gz	80.4	95.3
Inception V3	Code	inception_v3_2016_08_28.tar.gz	78.0	93.9
ResNet 152	Code	resnet_v1_152_2016_08_28.tar.gz	76.8	93.2
ResNet V2 200	Code	TBA	79.9*	95.2*

(*): Results quoted in ResNet paper.

Figure 3.2: Comparison of latest Inception and Resnet models performance on Imagenet dataset. Sourced from [55]..

3.2 Datasets

The research project has chosen the following datasets to train the models because of their ease of access, and large quantities and diverse categories of food images:

FOOD-101

This dataset consists of 101 food categories, with 101 000 images. For each class, 250 manually reviewed test images are provided as well as 750 training images which were not cleaned, and thus still contain some amount of noise (intense colors and sometimes wrong labels)[10].

FOOD-11

FOOD-11 dataset contains 16643 food images grouped in 11 major food categories. Similar to the Food-5K dataset, the whole dataset is divided in three parts: training, validation and evaluation [42].

ECUSTFD

"ECUSTFD" is a free public food image dataset which contains 2978 images of 19 types of food specifically for Food nutrient estimation[27]. The dataset is shown in figure 3.3. The number of images and objects for the same type are shown as follows:

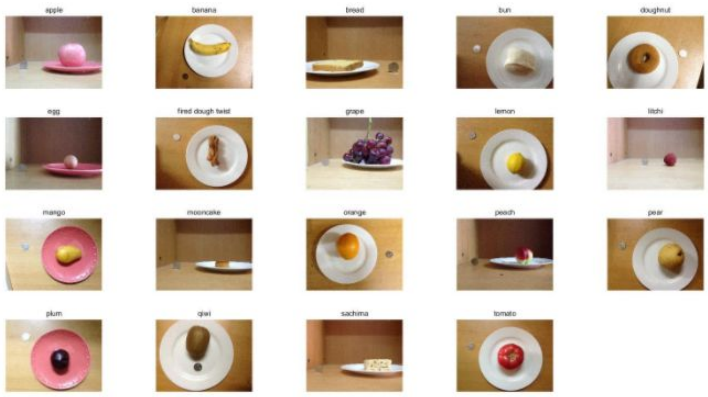


Figure 3.3 : Images in the ECUSTFD dataset. Sourced from [27].

For a single food portion, several groups of images have been taken by smartphones; each group of images contains a top view and a side view of this food. For each image, there is only one coin as a calibration object and no more than two foods in it. If there are two foods in the same image, the type of one food is different from another[27].

Chapter 4

Account of Technical Progress

4.1 Research

The project problems, scope and goals have been discussed earlier, in the project brief. The plan remains almost the same as the beginning of the project, with a priority of optimizing performance of DL and Traditional models on the Food-101 dataset. So far, Python has been chosen to train the DL models. Knowledge and practice with Python DL libraries were built over the summer. The models have particularly been chosen after significant amounts of research and discussion with the supervisor. An overview and the fundamentals of Deep Learning were understood through a Deep Learning workshop held by the supervisor early on in the year. To build on the technical knowledge, Google Codelabs tutorials on Convolutional Neural Networks[6] and Transfer Learning[42] were completed successfully, along with Tutorial 6 of the Deep Learning module[22].

4.2 Technical Experimentation

Open source neural network library Keras, with a Tensorflow backend, has been chosen for the implementation of the project. The process of experimentation on Inceptionv3 is as follows: The images from the Food-101 dataset were extracted from a tar.gz file. All the class titles of images inside the food-101/images folder were listed, following which the images were visualised - an image from each class was displayed and the dimensions were each checked to be 25*25 .The data was split into separate train and test folders with 3:1 ratio respectively. Three classes were picked out of the 101 possible classes, namely samosa, omelette, and pizza and a dataset for solely these three classes was prepared accordingly.

There were several steps taken in training of the model, including data augmentation on all train, test and validation set of images. The weights from the ImageNet dataset were frozen, but the top layer was excluded. Instead a new Pooling layer was defined, followed by Dense and Dropout layers. The predictions were outputted using the Softmax activation function. The model was compiled and the model training function, when called prints the epoch, the accuracy levels at each epoch are printed. The final accuracy rate was 92.12%. The results of training the model can be seen in Figure 4.1:

```
Please use Model.fit, which supports generators.
Epoch 1/5
140/140 [==============================] - ETA: 0s - loss: 1.0347 - accuracy: 0.5040
Epoch 00001: val_loss improved from inf to 0.79321, saving model to bestmodel_3class.hdf5
140/140 [==============================] - 2328s 17s/step - loss: 1.0347 - accuracy: 0.5040 - val_loss: 0.7932 - val_accuracy: 0.7486
Epoch 2/5
140/140 [==============================] - ETA: 0s - loss: 0.7676 - accuracy: 0.7179
Epoch 00002: val_loss improved from 0.79321 to 0.57555, saving model to bestmodel_3class.hdf5
140/140 [==============================] - 2359s 17s/step - loss: 0.7676 - accuracy: 0.7179 - val_loss: 0.5755 - val_accuracy: 0.8533
Epoch 3/5
140/140 [==============================] - ETA: 0s - loss: 0.6024 - accuracy: 0.7936
Epoch 00003: val_loss improved from 0.57555 to 0.44933, saving model to bestmodel_3class.hdf5
140/140 [==============================] - 2350s 17s/step - loss: 0.6024 - accuracy: 0.7936 - val_loss: 0.4493 - val_accuracy: 0.8818
Epoch 4/5
140/140 [==============================] - ETA: 0s - loss: 0.4953 - accuracy: 0.8339
Epoch 00004: val_loss improved from 0.44933 to 0.35738, saving model to bestmodel_3class.hdf5
140/140 [==============================] - 2360s 17s/step - loss: 0.4953 - accuracy: 0.8339 - val_loss: 0.3574 - val_accuracy: 0.9022
Epoch 5/5
140/140 [==============================] - ETA: 0s - loss: 0.4395 - accuracy: 0.8478
Epoch 00005: val_loss improved from 0.35738 to 0.31301, saving model to bestmodel_3class.hdf5
140/140 [==============================] - 2335s 17s/step - loss: 0.4395 - accuracy: 0.8478 - val_loss: 0.3130 - val_accuracy: 0.9212
{'omelette': 0, 'pizza': 1, 'samosa': 2}
```

Figure 4.1: Results from training Inceptionv3 on 3 classes from Food-101 dataset. Screenshot from Google Colab.[55]

4.3 Difficulties Encountered

This project has been a particularly technically challenging project to date, however hands-on practice with Object detection and recognition over the summer (through OpenImaj tutorials[48] and Python Deep Learning online resources) have helped tremendously in building the foundational knowledge. Errors in importing and extracting datasets and correctly implementing the last layer of the Inceptionv3 model have been the most technically challenging aspects of the implementation so far. However, online platforms such as StackOverflow and Google Codelabs have helped solve these problems.

4.4 Remaining Work

As per the Inceptionv3 model, the remaining work includes plotting the accuracy and loss curves against epochs, to deduce the rate of gain in accuracy and drop in loss. The model will further be tested on random images on the internet belonging to these same categories, to explore whether the accuracy remains for images outside of the dataset. The model will then be altered to train and test accuracy with variable number of classes. This entire process will be used again when training other models to compare their performance on the various datasets.

Bibliography

[1] A. Singla, L. Yuan and T. Ebrahimi, "Food/Non-food Image Classification and Food Categorization using Pre-Trained GoogLeNet Model", Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management - MADiMa '16, 2016. Available: 10.1145/2986035.2986039. [Accessed 8 December 2020].

[2] Baxter, J.. “Food Recognition using Ingredient-Level Features.” (2012). Massachusetts Institute of Technology. https://www.semanticscholar.org/paper/Food-Recognition-using-Ingredient-Level-Features-Baxter/6cc1e3fedca5029b47dced0bb6beb26e33f6c819.[Accessed 6 December 2020].

[4] Bedford, Simon. “Simon Bedford.” Simon Bedford Atom, simonb83.github.io/machine-learning-foodclassification.html. [Accessed 6 December 2020].

[5] Ciocca, G., Napoletano, P., & Schettini, R. (2017). Food recognition: A new dataset, experiments, and results. IEEE Journal of Biomedical and Health Informatics, 21(3), 588–598. https://doi.org/10.1109/jbhi.2016.2636441.[Accessed 6 December 2020].

[6] "Convolutional Neural Network (CNN) | TensorFlow Core", TensorFlow, 2020. [Online]. Available: https://www.tensorflow.org/tutorials/images/cnn. [Accessed: 07- Dec- 2020].

[7]C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", arXiv.org, 2016. [Online]. Available: https://arxiv.org/abs/1602.07261v2. [Accessed: 07- Dec- 2020].

[8]C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", arXiv.org, 2015. [Online]. Available: https://arxiv.org/abs/1512.00567v3. [Accessed: 07- Dec- 2020].

[9]Ege, T., & Yanai, K. (2018). Image-based food calorie estimation using recipe information. IEICE Transactions on Information and Systems, E, 101D(5), 1333–1341. https://doi.org/10.1587/transinf.2017MVP0027. [Accessed 6 December 2020].

[10]"Food101 | TensorFlow Datasets", TensorFlow, 2020. [Online]. Available: https://www.tensorflow.org/datasets/catalog/food101. [Accessed: 08- Dec- 2020].

[11]"Food Calorie Measurement and Classification of Food Images", International Journal of Pharmaceutical Research, vol. 12, no. 04, 2020. Available: 10.31838/ijpr/2020.12.04.262. [Accessed 7 December 2020].

[12]F. Ragusa, V. Tomaselli, A. Furnari, S. Battiato and G. Farinella, "Food vs Non-Food Classification", MADiMa '16: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, pp. 77-81, 2016. Available: https://dl.acm.org/doi/10.1145/2986035.2986041. [Accessed 8 December 2020].

[13]Fu, Z. H., Chen, D., & Li, H. Y. (2017). ChinFood1000: A large benchmark dataset for Chinese food recognition. In D. S. Huang, V. Bevilacqua, P. Premaratne, & P. Gupta (Eds.), Intelligent Computing Theories and Application, ICIC 2017, Pt I (Vol. 10361, pp. 273–281). https://doi.org/10.1007/978-3-319-63309-1_25. [Accessed 5 December 2020].

[14]G. Ciocca, P. Napoletano and R. Schettini, "Food Recognition: A New Dataset, Experiments, and Results", IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 3, pp. 588-598, 2017. Available: 10.1109/jbhi.2016.2636441. [Accessed 5 December 2020].

[15]G. Yiğit and B. Özyildirim, "Comparison of convolutional neural network models for food image classification", Journal of Information and Telecommunication, vol. 2, no. 3, pp. 347-357, 2018. Available: https://www.tandfonline.com/doi/full/10.1080/24751839.2018.1446236. [Accessed 8 December 2020].

[16]Heravi, E. J., Aghdam, H. H., & Puig, D. (2015). A deep convolutional neural network for recognizing foods. In Proceedings of 8th International Conference on

Machine Vision (Vol. 9875). https://doi.org/10.1117/12.2228875. [Accessed 8 December 2020].

[17]Heravi, E. J., Aghdam, H. H., & Puig, D. (2017). Classification of foods by transferring knowledge from ImageNet dataset. In Proceedings of 9th International Conference on Machine Vision (Vol. 10341). https://doi.org/10.1117/12.2268737. [Accessed 5 December 2020].

[18]Heravi, E. J., Aghdam, H. H., & Puig, D. (2018). An optimized convolutional neural network with bottleneck and spatial pyramid pooling layers for classification of foods. Pattern Recognition Letters, 105, 50–58. https://doi.org/10.1016/j.patrec.2017.12.007. [Accessed 9 December 2020].

[19]H. Kagaya, K. Aizawa and M. Ogawa, "Food Detection and Recognition Using Convolutional Neural Network", 2020. Available: https://www.researchgate.net/publication/266357771_Food_Detection_and_Recognition_Using_Convolutional_Neural_Network. [Accessed 7 December 2020].

[20]I. Fadelli, "FoodTracker: An AI-powered food detection mobile application", Techxplore.com, 2020. [Online]. Available: https://techxplore.com/news/2019-09-foodtracker-ai-powered-food-mobile-application.html. [Accessed: 08- Dec- 2020].

[21]"ImageNet", Image-net.org, 2020. [Online]. Available: http://image-net.org/. [Accessed: 08- Dec- 2020].

[22]J. Hare, "COMP6248 Differentiable Programming (and Deep Learning)", Comp6248.ecs.soton.ac.uk, 2020. [Online]. Available: http://comp6248.ecs.soton.ac.uk/labs/lab6/. [Accessed: 07- Dec- 2020].

[23]J. Sun, K. Radecka and Z. Zilic, "Exploring Better Food Detection via Transfer Learning", 2019 16th International Conference on Machine Vision Applications (MVA), 2019. Available: 10.23919/mva.2019.8757886 [Accessed 7 December 2020].

[24]Kagaya, Hokuto & Aizawa, Kiyoharu & Ogawa, Makoto. (2014). Food Detection and Recognition Using Convolutional Neural Network. 10.13140/2.1.3082.1120. [Accessed 6 December 2020].

[25]Kiourt, C., Pavlidis, G. and Markantonatou, S., (2020), Deep learning approaches in food recognition, MACHINE LEARNING PARADIGMS - Advances in Theory and Applications of Deep Learning, Springer

[26]K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, 2015, pp. 1-6, doi: 10.1109/ICMEW.2015.7169816. [Accessed 8 December 2020].

[27]Liang-yc. "ECUSTFD_Bifrost Data Search", Datasets.bifrost.ai, 2020. [Online]. Available: https://datasets.bifrost.ai/info/230. [Accessed: 08- Dec- 2020].

[28]Liu C., Cao Y., Luo Y., Chen G., Vokkarane V., Ma Y. (2016) DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment. In: Chang C., Chiari L., Cao Y., Jin H., Mokhtari M., Aloulou H. (eds) Inclusive Smart Cities and Digital Health. ICOST 2016. Lecture Notes in Computer Science, vol 9677. Springer, Cham. https://doi.org/10.1007/978-3-319-39601-9_4. [Accessed 8 December 2020].

[29]L. Zhou, C. Zhang, F. Liu, Z. Qiu and Y. He, "Application of Deep Learning in Food: A Review", Comprehensive Reviews in Food Science and Food Safety, vol. 18, no. 6, pp. 1793-1811, 2019. Available: 10.1111/1541-4337.12492 [Accessed 7 December 2020].

[30]M. A. Subhi and S. Md. Ali, "A Deep Convolutional Neural Network for Food Detection and Recognition," 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia, 2018, pp. 284-287, doi: 10.1109/IECBES.2018.8626720. [Accessed 8 December 2020].

[31]Martinel, N., Foresti, T. L., & Micheloni, C. (2018). Wide-slice residual networks for food recognition. In 2018 IEEE Winter Conference on Applications of Computer Vision (pp. 567–576). https://doi.org/10.1109/WACV.2018.00068. [Accessed 8 December 2020]. [Accessed 5 December 2020].

[32]Mezgec, S., & Seljak, B. K. (2017). NutriNet: A deep learning food and drink image recognition system for dietary assessment. Nutrients, 9(7). https://doi.org/10.3390/nu9070657. [Accessed 2 December 2020].

[33]M. Subhi, S. Ali and M. Abdulameer, "Deep Convolutional Networks for Food Detection and Classification", Journal of Computational and Theoretical Nanoscience, vol. 16, no. 5, pp. 2433-2438, 2019. Available: 10.1166/jctn.2019.7913.[Accessed 2 December 2020].

[34]Myers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., . . . Murphy, K. (2015). Im2Calories: Towards an automated mobile vision food diary. In 2015 IEEE International Conference on Computer Vision (pp. 1233–1241). https://doi.org/10.1109/ICCV.2015.146. [Accessed 3 December 2020].

[35]N. Mahony et al., "Deep Learning vs. Traditional Computer Vision", IMaR Technology Gateway, Institute of Technology Tralee, Tralee, Ireland, 2019. Available: https://www.researchgate.net/publication/331586553_Deep_Learning_vs_Traditional_Computer_Vision. [Accessed 8 December 2020].

[36]"Obesity - Causes", *nhs.uk*, 2020. [Online]. Available: https://www.nhs.uk/conditions/obesity/causes/. [Accessed: 08- Dec- 2020].

[37]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. [Accessed: 03- Dec- 2020].

[38]Parisa Pouladzadeh, Abdulsalam Yassine, Shervin Shirmohammadi, August 1, 2020, "FooDD: Food Detection Dataset for Calorie Measurement Using Food Images", IEEE Dataport, doi: https://dx.doi.org/10.21227/yvk7-qk38. [Accessed: 07- Dec- 2020].

[39]P. McAllister, H. Zheng, R. Bond and A. Moorhead, "Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets", Elsevier Ltd., 2018. Available: https://pubmed.ncbi.nlm.nih.gov/29549733/. [Accessed 8 December 2020].

[40]Pouladzadeh, P., S. Shirmohammadi and R. Almaghrabi. “Measuring Calorie and Nutrition From Food Image.” IEEE Transactions on Instrumentation and Measurement 63 (2014): 1947-1956. [Accessed: 03- Dec- 2020].

[41]R. Mustafa and P. Dhar, "A Method to Recognize Food using Gist and SURF Features," 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, 2018, pp. 127-130, doi: 10.1109/ICIEV.2018.8641072. [Accessed: 07- Dec- 2020].

[42]"Transfer learning and fine-tuning | TensorFlow Core", TensorFlow, 2020. [Online]. Available: https://www.tensorflow.org/tutorials/images/transfer_learning. [Accessed: 07- Dec- 2020].

[42]Verma, Avi. “Food-11.” Kaggle, 5 June 2019. [Online]. Available: https://www.kaggle.com/vermaavi/food11. [Accessed: 08- Dec- 2020].

[43]"What Is Food Culture And How Does It Impact Your Health?", *The Well Essentials*, 2020. [Online]. Available: https://www.thewellessentials.com/blog/what-is-food-culture-and-what-does-it-have-to-do-with-our-health. [Accessed: 09- Dec- 2020].

[44]Yoshiyuki Kawano and Keiji Yanai, Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation, Proc. of ECCV Workshop on Transferring and Adapting Source Knowledge in Computer Vision (TASK-CV), 2014.

[45]Z. Shen, A. Shehzad, S. Chen, H. Sun and J. Hiu, "Machine Learning Based Approach on Food Recognition and Nutrient Estimation", vol. 174, pp. 448-453, 2020.

Available: https://www.sciencedirect.com/science/article/pii/S1877050920316331. [Accessed 8 December 2020].

[46]"Obesity and overweight", *Who.int*, 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight. [Accessed: 09- Dec- 2020].

[47]Konaje, Nayan Kumar. “Food recognition and calorie extraction using Bag-of-SURF and Spatial Pyramid Matching methods.” [Accessed: 09- Dec- 2020].

[48]"OpenIMAJ: Open Intelligent Multimedia Analysis", *Openimaj.org*, 2020. [Online]. Available: http://openimaj.org/tutorial/. [Accessed: 09- Dec- 2020].

[49]S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way", Medium, 2018. [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. [Accessed: 09- Dec- 2020].

[50]J. Brownlee, "A Gentle Introduction to the ImageNet Challenge (ILSVRC)", *Machine Learning Mastery*, 2020. [Online]. Available: https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/. [Accessed: 09- Dec- 2020].

[51]V. Mishra, "CNN Architecture: How ResNet works and why?", *Medium*, 2020. [Online]. Available:https://medium.com/datadriveninvestor/cnn-architecture-how-resnet-works-and-why-1c197b8eba34. [Accessed: 09- Dec- 2020].

[52] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[53]"A Simple Guide to the Versions of the Inception Network", *Medium*, 2020. [Online].

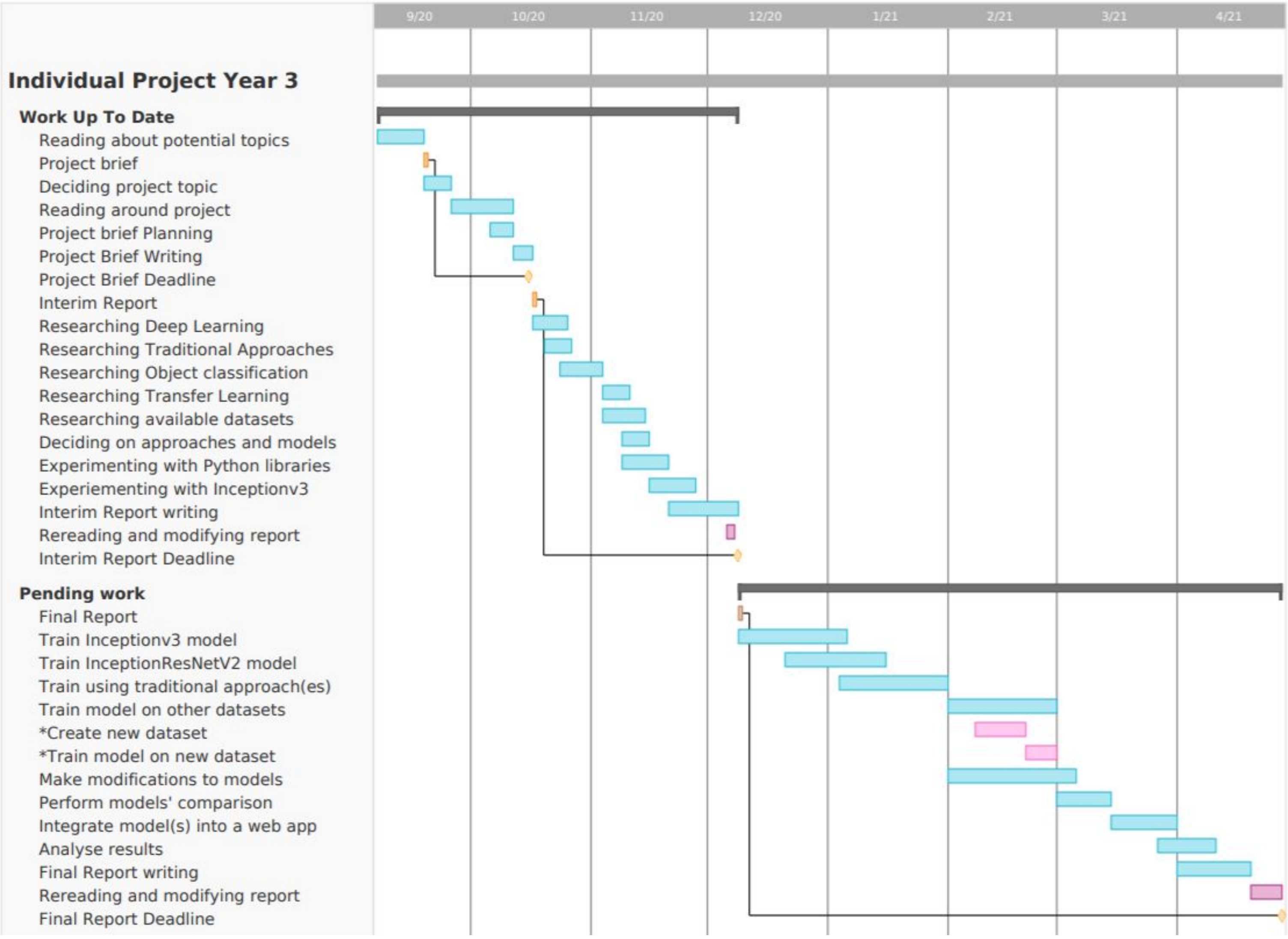
Available:https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202. [Accessed: 09- Dec- 2020].

[54]"A Guide to ResNet, Inception v3, and SqueezeNet | Paperspace Blog", *Paperspace Blog*, 2020. [Online]. Available: https://blog.paperspace.com/popular-deep-learning-architectures-resnet-inceptionv3-squeezenet/. [Accessed: 09- Dec- 2020].

[55]"Improving Inception and Image Classification in TensorFlow", *Medium*, 2020. [Online]. Available: https://medium.com/@penolove15/improving-inception-and-image-classification-in-tensorflow-1e3c2ada4572. [Accessed: 09- Dec- 2020].

[56] Dixon, J. B. (2010). The effect of obesity on health outcomes. Molecular and cellular endocrinology, 316(2), 104-108.

Appendix A: Gantt Chart showing Work up to date and Pending work.



Tasks labelled with * symbols are extensions, and will only be explored if the project is up to date, and it seems feasible to work on these tasks at the time.

Appendix B: Risk analysis and Mitigation

Risk	Probability (1-5)	Severity (1-5)	Risk Exposure (P*S)	Mitigation
Loss of work/dataset	3	5	15	-Keep a copy of written work in the cloud. -For final report, commit changes to Latex file. -Use version control to retrieve recent versions of work. -Keep a hard disk for work conducted locally on the personal computer.
Implemented model is too time-taking and resource intensive	4	3	12	- Use GPUs in cloud (Google Colab). - Connect to university's machines virtually on local machine/book space in B16.
Conflicting deadlines	4	3	12	-Plan and alter Gantt chart accordingly.
Difficulty with implementation of models	3	3	9	- Seek help from supervisor, dedicate more time to learning and practising on less complex problems.
Personal Illness/ Close family member(s) illness	2	4	8	- Inform project supervisor and appropriate members of ECS department for issues related to Covid-19. - For not as serious situation, update Gantt chart accordingly.
Infeasible solutions to problem	2	4	8	- Brainstorm alternative solutions -Discuss a feasible solution with supervisor -Consider alternative solutions
Change in project specification	2	3	6	- Update Gantt chart accordingly. -Contact project coordinator for details if in doubt.

Postponed allocation of second supervisor	3	2	6	- Be prepared for the Viva in advance. -Keep an eye out for second supervisor allocation.
Project is complete too soon	1	1	1	-Work on extensions eg. new dataset/ adding features to web application/ constructing new models