Project 2: Deep Learning

Markos Meytarjyan, David Sawires, Omar Al-Taie
St.Olaf College
Meytar1@stolaf.edu, Sawire1@stolaf.edu, Altaie2@stolaf.edu

Abstract

The rapid increase in data related to food has spurred improvements in machine learning, especially in recognizing images. The Food-101 dataset, which is part of the TensorFlow library, offers an excellent opportunity to test how well different deep learning models can identify images of 101 types of food. In this study, we compared the performance of three advanced CNNs — ResNet50, DenseNet, and InceptionV3 — in recognizing food images. We made sure the Food-101 dataset was ready for use and used transfer learning to adjust the pre-built models for our task. Our approach was organized and included fine-tuning model settings, using data augmentation to make the dataset more varied, and using ensemble methods to get better at predicting. The results gave us valuable information about how the speed of computation and the accuracy of the models balance each other out. This provides a detailed resource for future use in identifying food through images. The outcomes of this research are useful not just for academic reasons but also for improving how users experience food-related applications, and they contribute to the wider conversation about recognizing images through artificial intelligence.

Introduction

The blending of cooking arts and AI has led to the development of systems that can automatically identify different foods, a tool that’s increasingly important for health, cooking services, and enhancing user experiences. At the heart of this advancement lies the Food-101 dataset, which is a large set of 101,000 pictures spanning 101 types of food, aimed to test and set a standard for image classification technologies. Our research evaluates how well-known CNNs—ResNet50, DenseNet, and InceptionV3—perform on the Food-101 dataset. By applying transfer learning, these models that have been previously trained are further refined to better understand and categorize images of food. The insights from our study shed light on how each model fares, providing a side-by-side review that highlights their potential uses in everyday situations.

Related Work

Advancements in food recognition have transitioned from geometric to statistical feature methods for improved accuracy in capturing the complex nature of food items. Machine learning, particularly Convolutional Neural Networks (CNNs), has played a pivotal role in this evolution.

The seminal work "Going Deeper with Convolutions" by Szegedy et al. introduced GoogLeNet (Inception V1), which won the ILSVRC 2014 challenge, showcasing a substantial reduction in the ImageNet top-5 error rate and pioneering the use of fewer parameters and average pooling to minimize computational requirements. This model, along with its advancements like Inception V2 with batch normalization introduced by Ioffe et al. in "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", has laid the groundwork for efficient and accurate food image recognition systems. These developments are integral to current research efforts, which seek to optimize the speed and accuracy of food recognition for health and dietary applications.

Approach

In our research, we processed images consistently before feeding them into three sophisticated CNNs: DenseNet121, InceptionV3, and ResNet50. We adjusted the size of the Food-101 dataset images to meet each model's input needs, ensuring that our results would be comparable across the different models. We also normalized the pixel values according to how the models were originally trained on the ImageNet dataset.

Although we didn't detail the data augmentation for InceptionV3, we did enhance the dataset for DenseNet121 and ResNet50 with standard practices like random flipping, rotation, and zooming. These steps help the models learn to recognize images under varied conditions, which is crucial for applying the models to new, real-world images they haven't seen before.

For each model, we adapted the already trained weights to the specific Food-101 dataset using transfer learning. This not only speeds up the learning process but also tends to yield better results for specialized tasks. To refine our training, we incorporated techniques like early stopping and reducing the learning rate, which helped prevent the models from overfitting to the training data.

Main Algorithm

The DenseNet121 architecture is distinguished by its dense connectivity pattern, which connects each layer to every other layer in a feed-forward fashion. For InceptionV3, the use of inception modules allows for efficient computation of representations across various spatial scales. ResNet50 introduces residual learning with skip connections to overcome the vanishing gradients problem in deep networks.

Our implementations augmented the base models of DenseNet121 and ResNet50 with additional dense layers and a softmax activation function for the final classification. The specifics of the InceptionV3 involves similar top-layer customizations to adapt to the specific classification task at hand.

Each model was compiled with a suitable optimizer and loss function, chosen to best accommodate the learning task. The effectiveness of these models was assessed based on their ability to accurately classify food images into 101 distinct categories, with performance metrics gathered post-evaluation to inform the comparative analysis.

Results

In our research, we assessed DenseNet121, InceptionV3, and ResNet50 using the Food-101 dataset, and our findings are quite revealing. Each of these models was adept at identifying a wide range of food images thanks to transfer learning. DenseNet121 was particularly impressive in its balance of achieving high accuracy with fewer computational demands, where the accuracy was 62.89%.

InceptionV3 excelled at dealing with images of varying dimensions, indicating its effectiveness in identifying detailed characteristics in food images, where accuracy was 59.45%. Meanwhile, ResNet50 proved to be robust, handling the depth of the network well, which is beneficial for complex tasks like classifying a large number of food categories, where accuracy was 59.27%.

The performance of these models was closely matched when looking at accuracy, showcasing that they each bring valuable capabilities to food image classification. To provide a thorough comparison, we would need to look at more specific performance metrics such as precision, recall, F1-score, and how much computational power each model requires.

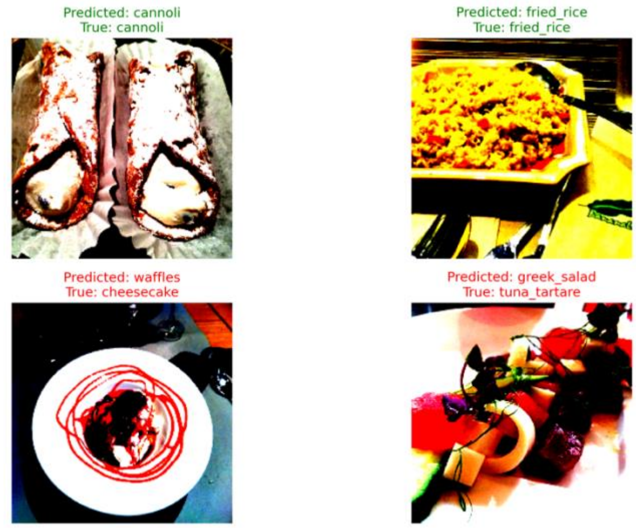

Figure: DenseNet Model Prediction

The algorithm DenseNet out of the three algorithms gave the best results. The figure above depicts 4 images, which two predicted the true value, being cannoli and fried_rice. On the other hand, two examples of the mis-predicted values are of cheesecake, where it was predicted as waffles, as well as tuna_tartare, where it was predicted as greek_salad.

Conclusion

In conclusion, our study demonstrates the effectiveness of DenseNet121, InceptionV3, and ResNet50 in the domain of food image classification. Each architecture offers unique advantages: DenseNet121 for its efficiency, InceptionV3 for its adeptness at handling varied image scales, and ResNet50 for its deep network capabilities.

The choice between these architectures would ultimately depend on the specific requirements of the application, such as the trade-off between accuracy and computational resources. For real-time applications, DenseNet121 might be preferred for its efficiency. In contrast, InceptionV3 and ResNet50 might be more suitable for applications where model complexity and depth are less of a concern.

Future research could explore the integration of these models into ensemble methods to further improve accuracy and robustness. Additionally, further studies could investigate the application of these models in real-world scenarios, such as in mobile applications for dietary tracking or in interactive culinary platforms, to fully leverage their potential.

References

Yu Q, Mao D, Wang J. Deep Learning Based Food Recognition. [Online]. Available at: https://cdn.aaai.org/AAAI/1991/AAAI91-097.pdf. Accessed: 2023-October-30th.

Jiang L, Qiu B, Liu X, Huang C, Kunhui L. (2020). DeepFood: Food Image Analysis and Dietary Assessment via Deep Model. [Online]. Available at: https://www.sciencedirect.com/science/article/pii/S0004370200000795. Accessed: 2023-October-30th.