HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 1 (2022)

AOS: Anti-Obesity System with Deep Learning-Based Classification Model Using a Novel Data Augmentation Technique

Sungmin Kim[1] and Kyoung-Hyoun Kim[#]

[1]Bergen County Academics, South Korea
[#]Advisor

ABSTRACT

Obesity has been a worldwide leading cause of associated health risks for decades, yet not much is being done to solve this issue. Previously, minimal efforts have been made to address this issue in regards to mandating nutrition labels on the back of packaged foods. However, the mundane and growingly elusive packing labels have diminished the proper effects of this constitution as people rarely ever take the time nor effort to thoroughly read through the labels. Moreover, most dishes served from restaurants do not serve with any nutritional information, which places many of the people dining out in complete darkness of what they are truly consuming. To solve this problem, we propose the deep learning-based AOS (Anti-Obesity System), which analyzes images of common junk foods and unhealthy meals to classify the category of food. The proposed system consists of a deep learning-based classifier and a post-processing module that outputs relevant nutritional information regarding each category of food. In addition, we propose a novel data augmentation technique in order to make the trained model produce better results. We also conduct the ablation study to experimentally prove that the proposed method enforces the model to work better in real-world situations. We achieve an AP of 59.30% on the proposed food dataset. Code will be available at https://github.com/smkim0508/ObesityNutritionClassifier.git.

Introduction

The obesity epidemic has rapidly spread across much of the world over the past decade. Since 1975, the worldwide prevalence of obesity has nearly tripled, with 39% of adults being overweight and 13% being obese in 2016 (World Health Organization, 2021). As shown by these statistics, the world has long suffered from obesity and its trend does not seem to be fading in the current stage. Although there are numerous contributors to this worldwide phenomenon, the increase in calorie consumption is considered to be one of the most crucial factors involved (PublicHealth, 2021).

In the past, there have been many studies to address this problem–a common method being an application that tracks the dietary consumption of users through what the user inputs as text and outputting the corresponding caloric levels. However, these systems are largely limited by the capability of the user to accurately input the text information of the products they have been consuming. This becomes an issue since the user may not always be able to precisely recount the name of the food. Moreover, the food in question may not even be enlisted if it is an authentic dish rather than a manufactured product. In addition, it is labor-intensive and tiresome for the users to have to constantly input the name of their food.

To solve this issue, we propose AOS (Anti-Obesity-System), a novel deep learning-based system made to raise caution when consuming meals and delicious treats on a day-to-day basis as well as suggest healthier options for one to resort to, ultimately aiming to resolve the crisis of never-ending obesity in the world. The

proposed system takes RGB images of common junk foods as input and outputs relevant information regarding the specific food. We trained the system to look out for various food types as a classification task. The proposed system exploits ResNet18 (Kaiming He, et al., 2016), a model that has consistently shown outstanding performances in various computer vision tasks. We also proposed a domain-specific data augmentation technique in order to boost the proposed model accuracy. We collected a diverse dataset of images from the food domain, through which we achieved a final test accuracy of 59.30%.
Contributions of this paper are summarized as:

1. To the best of our knowledge, we first attempt to address the classification task targeted specifically at promoting healthier food and addressing the issue regarding the worldwide consumption of junk foods.
2. We develop a novel data augmentation method called Random Conjunction. This particular technique allows us to achieve a 0.23% higher accuracy, and it is promising to be applied in more practical real-world scenarios.
3. To train the proposed system, we collect a diverse dataset, which we expect to be exploited in future studies.

Methods

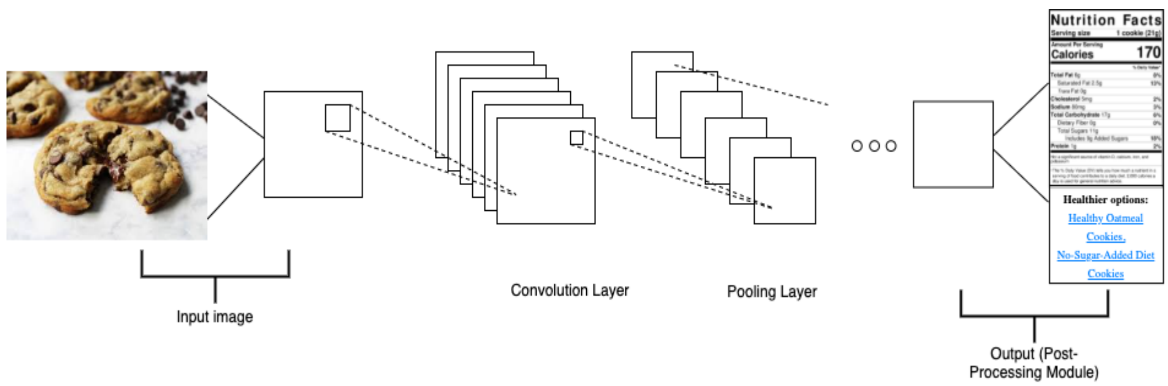


Figure 1. The Architecture of the Proposed System

Figure 1 displays a simplified diagram of the architecture used in the proposed system. The input image of food filters through a series of convolution layers and pooling layers to output a class label and relevant information such as a nutritional label and external links to recipes of similar food.

Architecture Design

Although deep learning networks such as AlexNet (Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, 2012), VGG (Simonyan, Karen and Andrew Zisserman, 2014), ZFNet (Zeiler, M. D. and Fergus, R., 2014), and GoogleNet (Szegedy, C., et al., 2015) have been presenting promising results in classification, there are numerous problems that deter the success of networks. One main problem is the gradient vanishing, in which the gradient value becomes excessively small after processing backpropagation when the network depth becomes too large.

This is a well-known issue with deep learning networks; however, Resnet18 (Kaiming He, et al., 2016) implements the residual block to add a copy of the original signal prior to taking the gradient in order to solve this issue as shown in Figure 1. Due to these advantageous properties, we exploit Resnet18 in our proposed system.

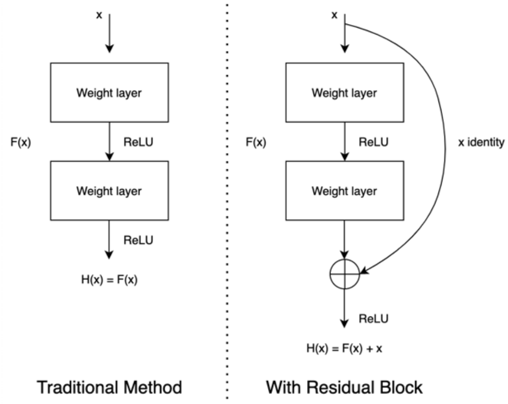


Figure 2. Example of Residual Block

Implementation Details

To train the proposed system, we use Adam optimizer (D. Kingma and J. Ba. Adam, 2014) with beta1 set to 0.9 and beta2 to 0.99.

We set the mini-batch size to 48 samples. We train the proposed system for 200-epoch. The learning rate is initialized to be 0.001 and decreases by a factor of 10 every 80 epochs using MultiLRStep method implemented in PyTorch (Adam Paszke, et al., 2019).

(a) (b) (c) (d) (e) (f)

Figure 3. Data Augmentation used in the Proposed Method

(a) Original Image, (b) Padded Image, (c) Perspective Image, (d) Rotated Image, (e) Horizontally Flipped Image, and (f) Vertically Flipped Image

Figure 3 illustrates some examples of the data augmentation methods we applied to the proposed system. Figure 3 (a) represents the original image without any data augmentation applied. Figure 3 (b) demonstrates the image padding, which inserts a border on the outer edge of the image. We found that a border of 1 pixel worked the best. Figure 3 (c) shows the random perspective augmentation, which enforces the trained model to yield better results on real-world images taken from various perspectives. We also utilize the random rotation shown in Figure 3 (d) to rotate the image by a random degree about the horizontal axis, helping to identify images taken from different angles.

Lastly, the random horizontal flip and random vertical flip, illustrated in Figure 3 (e) and Figure 3 (f), respectively, are two similar functions that either invert the x or y coordinates of the pixels about the center of the image to "flip" the image horizontally or vertically. This augmentation was essential to helping the trained model to better operate on unseen data, real-world samples, and for various orientations of the same class of images.

Results and Discussion

HIGH SCHOOL EDITION
Journal of Student Research

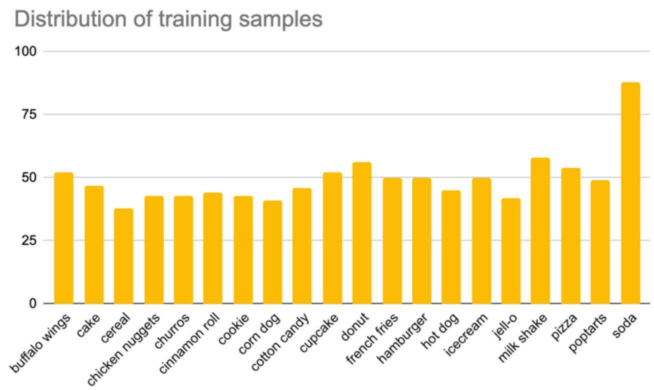


Figure 4. Histogram of Distribution of Each Category Amongst Training Samples

Figure 5. Snippet of the proposed dataset

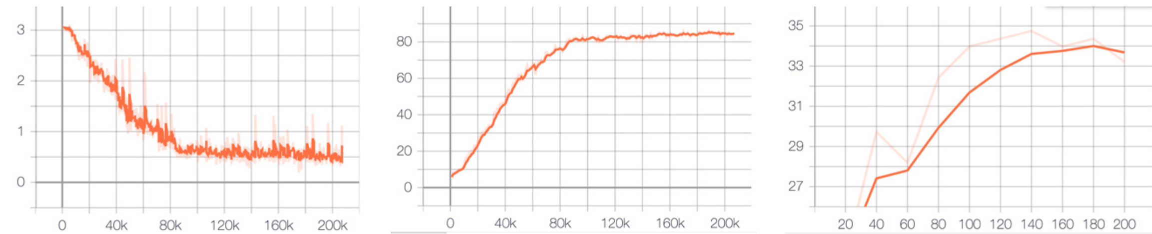
Figure 5 shows the snapshot of the dataset we used to train the proposed system. We collected the dataset from popular google image searches of each class of images with a high resolution. Altogether, there are 21 different classes which each contain approximately 60 samples. In total, there are 1322 sample images. Of these, the dataset is divided into 1059 training samples and 262 test samples.

Evaluation

Table 1. Comparison of accuracy and training loss in the state-of-the-art architectures.

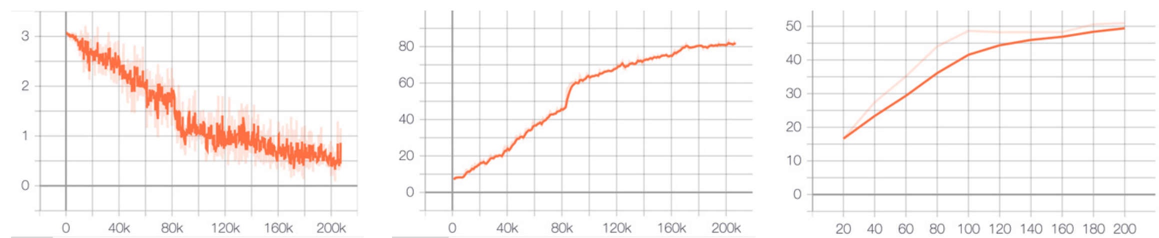
Architecture	Top-1 Test Accuracy (%)
AlexNet (Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, 2012)	34.74
VGG16 (Simonyan, Karen and Andrew Zisserman, 2014)	50.96
Resnet18 (Kaiming He, et al., 2016)	**59.30**

Volume 11 Issue 1 (2022)



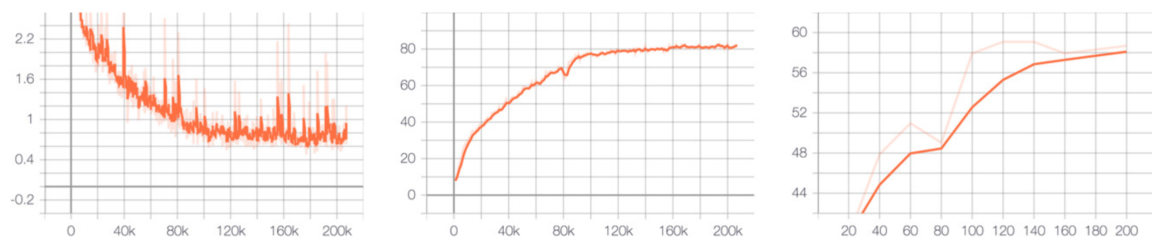
(a) (b) (c)

Figure 6. AlexNet experimental results
(a) Training loss, (b) Training accuracy, and (c) Validation accuracy



(a) (b) (c)

Figure 7. VGG16 experimental results
(a) Training loss, (b) Training accuracy, and (c) Validation accuracy



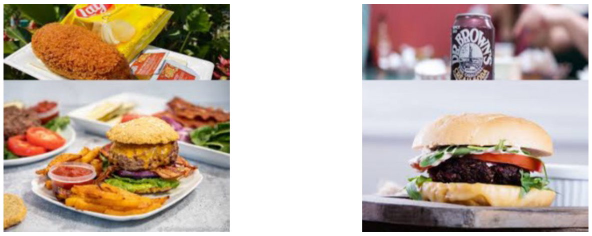
(a) (b) (c)

Figure 8. Resnet18 experimental results
(a) Training loss, (b) Training accuracy, and (c) Validation accuracy

Table 1 displays the comparison in the test accuracies of the AlexNet, VGG16, and Resnet18 architectures. Resnet18 outperformed the AlexNet by 24.56% and the Vgg16 model by 8.34%. At first, AlexNet was chosen as a base architecture for the model, but upon experimental trials, we found that the validation accuracy quickly saturated around only 34.74% as shown in Figure 6.

To address this issue, we exploit Resnet18 and VGG16 architectures which have deeper architecture depth in order to achieve a higher non-linearity power. The validation accuracy of ResNet18 saturates around 59.07% and displays little to no improvements further on as illustrated in (c) in Figure 8. VGG16, on the other hand, saturates at 50.96% test accuracy shown on (c) in Figure 7. Lastly, although AlexNet displays the highest train accuracy described on (b) in Figure 6., it performs the worst in terms of test accuracy amongst the three architectures tested. A likely hypothesis that is supported by this evidence is that the model had become overfit to the training samples and could not perform well when encountered with a new test case, which could be caused as a result of the architecture depth in AlexNet being too shallow.

Ablation Study



(a) (b)

Figure 9. Sample Images of the Proposed Data Augmentation

In real-life circumstances, food is oftentimes served with other accommodating dishes. Due to this reason, common images of food frequently experience partial or whole interference of additional dishes on the side, which can cause the trained model to produce poor inference results. To solve this issue, we propose a novel data augmentation technique called Random Conjunction, which combines two random images into one sample, with the "correct" image being larger in proportion to the "additional" image. With the application of this technique alone, the validation accuracy has risen by 0.23% as shown in Table 2.

Table 2. Comparative results of ablation study of the proposed data augmentation

Method	**Top-1 Test Accuracy (%)**
Ours (w/o RC[1])	59.07
Ours (w/ RC[1])	**59.30**

[1]RC denotes Random Conjunction Data Augmentation Technique.

Application

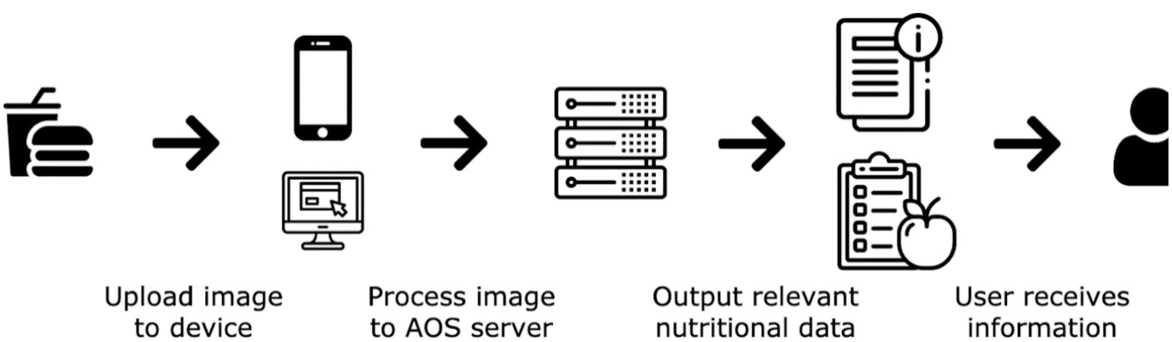


Figure 10. Flowchart of the potential applications of the system

Figure 10 describes a potential application of the proposed system in the real world. Users can upload an image of common foods, so long as it is within the system's accepting range of categories, which will be processed into the AOS (Anti-Obesity System) server. Then, the server will classify which category of image has been uploaded in order to output relevant nutritional data for the user to receive. Such application of this system could largely benefit people who are looking to avoid unhealthy diets and want to learn about what exactly they are consuming without worries.

Conclusion

In this paper, we proposed the deep learning-based AOS (Anti-Obesity System) that performs classification of unhealthy food and outputs vital nutritional information subjected to each category through a post-processing module. For the classification network, we exploited Resnet18 (Kaiming He, et al., 2016) to attain a higher non-linearity performance. We also proposed a novel data augmentation method to enforce the trained network to work better in real-world situations. We also conducted the ablation study for the proposed data augmentation method. Ultimately, applying the proposed data augmentation improved the test accuracy of the training model by 0.23.

Limitations

This new application of the classification task does face some challenges in training due to the intense variety of sub-categories of food, however, with future additions of new categories and more dataset variants, we hope to show promising improvements in both the accuracy and flexibility of the system. As an addition, the proposed system can be further utilized to offer helpful information such as nutrition labels and healthier alternative dishes, altogether hoping to raise awareness of how unhealthy common junk foods are as well as to provide people with an option to make healthier dietary choices while maintaining the same amount of sensory satisfaction.

Acknowledgments

I would like to thank Dr. Kyoung-Hyoun Kim in the Division of Healthcare and Artificial Intelligence at the Korea National Institute of Health for assisting me through the development of our proposed system.

References

"Obesity and Overweight." *World Health Organization*, World Health Organization, 9 June 2021, www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

Writers, Staff. "Why Are Americans Obese?" *Public Health*, PublicHealth.org, 15 Sept. 2021, www.publichealth.org/public-awareness/obesity/.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012): 1097-1105.

He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. https://doi.org/10.1109/cvpr.2016.90

Simonyan, Karen and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

D. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
Paszke, Adam, et al. "Pytorch: An imperative style, high-performance deep learning library." *Advances in neural information processing systems* 32 (2019): 8026-8037.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Zeiler, M. D. and Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer, Cham.