Paraguayan Snack Recognition

Jonathan David Gonzalez – F11015113 – UPTP
Juan Jose Minardi – F11015122 – UPTP
Octavio Santacruz – F11015125 – UPTP

ABSTRACT

In this study, we endeavored to bridge a gap in the global food recognition field by developing a machine learning system specifically designed to identify traditional snacks from Paraguay. Due to the lack of pre-existing data, we created our unique dataset, capturing the diverse range of these snacks. Building upon top-tier models from the ImageNet competition, we tailored these architectures to our distinctive collection, enhancing the system's capacity to accurately recognize these specific snacks.

In the pursuit of dynamic interaction and system refinement, we constructed a user-friendly website, inviting individuals to upload images for classification. Crucially, users can correct any misclassifications, enabling our model to learn from its mistakes and improve incrementally. This active participation of users not only bolsters the accuracy of our model, but also underlines the community's role in preserving Paraguay's culinary heritage digitally.

In the project's concluding phase, we incorporated the YOLO object detection model, empowering our system to locate snacks within any image, adding an invaluable layer of practical functionality. This step forward reinforces our system's value in real-world applications, transforming it from a mere food identifier to a versatile tool. Overall, this project provides substantial insights into expanding the horizons of image recognition systems and illustrates the exciting future of machine learning, particularly in its application to preserving global culinary diversity.

1. INTRODUCTION

In the ever-evolving field of machine learning, our project takes a bold step towards enhancing the recognition of traditional Paraguayan snacks. Unlike many datasets readily available for popular food items, the lack of pre-existing data for Paraguayan snacks posed a unique challenge. To tackle this, we curated a distinctive dataset ourselves, capturing the rich diversity of snacks native to Paraguay. With the help of leading architectures from the ImageNet competition, we tailored these models to work with our unique collection, pushing the boundaries of food recognition in less globally recognized cuisines.

We extended our efforts by building an interactive website, enabling users to upload snack images for the system to classify and correct any misclassifications. This collaborative learning approach not only refines our system but also encourages public participation in preserving Paraguay's culinary culture digitally.

In the final phase, we implemented the YOLO model, enabling our system to pinpoint snacks within any given image. This added functionality extends the applicability of our system to real-world scenarios, making it a practical tool for identifying Paraguayan snacks in various contexts.

2. DATA COLLECTION AND PREPARATION

2.1. Folder Creation

Six snacks were selected for detection and a folder was created for each one.

2.2. Dataset Loading

Data generation was used to flow the data from the directory to the dataset.

2.3. Resizing, Normalization and Augmentation of Images

Key preprocessing steps in the domain of deep learning to ensure consistent input and improve model performance. Resizing is done to maintain a uniform input size for the model, since different images can have varying dimensions, yet convolutional neural networks require fixed-size inputs. Normalization scales pixel values to a smaller range, which helps the model converge faster during training by providing numerical stability. Augmentation applies random transformations to the images, effectively increasing the size of the training dataset and helping the model generalize better, by making it robust to such variations in the input data.

2.4. Images and Labels Concatenation

The images and labels are retrieved from the data generator and concatenated to arrays X and y until storing the entire dataset.

3. METHODOLOGY

This project follows a model performance comparison methodology. A subset of the dataset is created in which each class has the same amount of images. The accuracy and loss of each model architecture are compared.

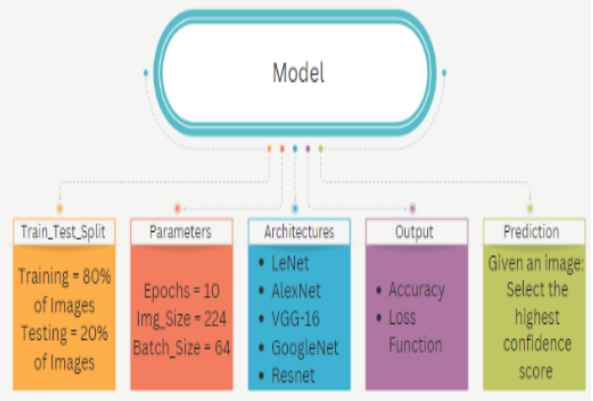

Fig I: Image classification model's structure

3.1. Architectures

3.1.1. LeNet

LeNet is an image recognition convolutional neural network architecture. Convolutional layers for feature extraction, pooling layers for spatial dimension reduction, and fully connected layers for classification make up this system. LeNet was a key component in the creation of CNNs and provided the framework for current deep learning models.

3.1.2. AlexNet

AlexNet introduced deep learning to a wider audience by winning the ImageNet Large Scale Visual Recognition Challenge in 2012. AlexNet features multiple convolutional layers, max-pooling layers, and fully connected layers, demonstrating the effectiveness of deep CNNs in computer vision tasks.

3.1.3. VGG-16

VGG-16 is renowned for being straightforward and efficient. It has 16 layers total, the majority of which are convolutional, followed by fully linked layers. It uses tiny 3x3 convolutional filters throughout the network to provide a deeper and more potent model for feature extraction and image categorization.

3.1.4. GoogleNet

GoogLeNet, also known as Inception v1, is a convolutional neural network architecture that introduced the concept of inception modules. It uses multiple parallel convolutional layers of different filter sizes to capture various scales of information in an efficient manner. GoogLeNet's design reduces computational complexity while maintaining high accuracy, making it suitable for deep learning applications.

3.1.5. ResNet

ResNet addresses the vanishing gradient problem in deep networks. It introduces skip connections or shortcuts that allow the network to learn residual mappings. By propagating the original input along with the learned features, ResNet enables the training of extremely deep networks with improved accuracy and faster convergence.

4. ADDITIONAL FEATURES

4.1. Object Detection

YOLOv8, the most recent iteration of the YOLO object detection algorithm, uses a single neural network to identify and classify objects within images. It accomplishes this by predicting both the boundaries and class probabilities of each detected object. YOLOv8 brings notable improvements like feature fusion, spatial attention, and context aggregation modules that make object detection quicker and more precise. Its architecture splits into two main sections: the 'backbone', which is a tweaked version of the CSPDarknet53 structure, and the 'head', which incorporates a self-attention mechanism and several convolutional layers for making predictions.

One of YOLOv8's unique capabilities is its ability to detect objects of varying sizes using a feature pyramid network. Standout aspects of YOLOv8 include enhanced accuracy and speed, support for different backbones, adaptive learning, advanced data augmentation techniques, a flexible architecture, and the provision of pre-trained models. It has wide-ranging applications in areas like autonomous vehicles, surveillance, and medical imaging due to its high precision and rapid detection speed.

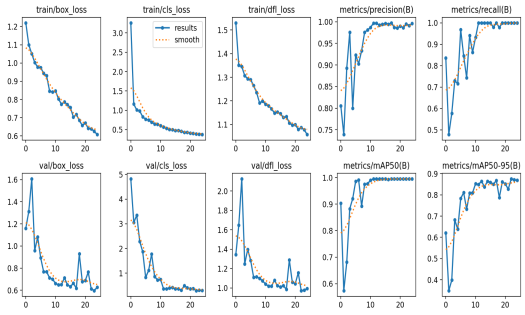

Fig II: Object Detection Model Results

4.2. Website Implementation

This web application plays a crucial role in acquiring and collecting valuable data on Paraguayan food images, which are not widely available in existing datasets. The importance of this website lies in its ability to serve as a platform for data collection, specifically for building a dataset that can be utilized for object detection and image classification tasks.

The Flask-based web application incorporates a pre-trained tensorflow model that specializes in image classification, with a specific focus on Paraguayan cuisine. By allowing users to upload images through the intuitive user interface, the application facilitates the collection of diverse and authentic Paraguayan food images, which are essential for training and improving the accuracy of machine learning models.

The backend logic, managed by the app.py file, efficiently handles the image preprocessing and feeds the uploaded images into the pre-trained machine learning model for classification. The model assigns each image to predefined food categories, enabling the identification and labeling of various Paraguayan dishes accurately. Additionally, the application provides users with the ability to correct misclassifications and store bounding box information, which enhances the dataset's quality and usefulness for object detection tasks.

On the frontend side, the web application consists of three html pages: home.html, index.html, and upload.html. These pages are designed to provide a seamless user experience while supporting both English and Spanish languages. The home page acts as an entry point, featuring an appealing design with a title, description, and an image upload button to encourage users to contribute their Paraguayan food images.

The index page serves as the primary user interface, offering a well-designed form for image upload, a submission button, a convenient return home button, and a canvas for users to draw bounding boxes on the images. By incorporating JavaScript, the page enables dynamic language switching, efficient image upload handling, and interactive canvas manipulation. This enhances the user experience and streamlines the data collection process.

The upload.html page displays the classified results, including the predicted food category and confidence score, along with the uploaded image. In cases where the prediction is incorrect, users are provided with a form to correct misclassifications, contributing to the improvement of the dataset's accuracy. The page further includes navigation buttons and language switch options to ensure user-friendliness. The design of this page emphasizes a structured layout, an appealing background image, and centered text to enhance the overall user experience.

This website plays a vital role in acquiring valuable Paraguayan food image data, which is scarce in existing datasets. By providing an intuitive interface for image upload, preprocessing, and classification, the web application enables the creation of a comprehensive and diverse dataset. This dataset can then be utilized for various applications, including object detection, image classification, and the promotion and preservation of Paraguayan culinary culture.

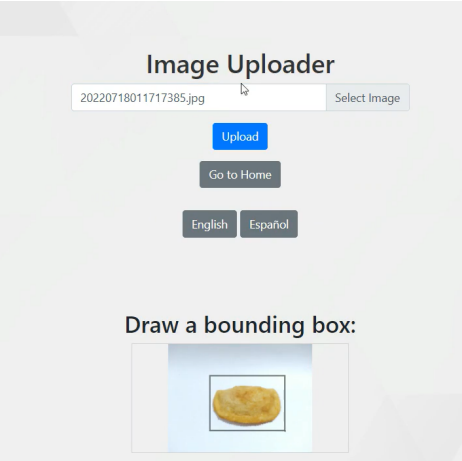

Fig III: Website Upload Page

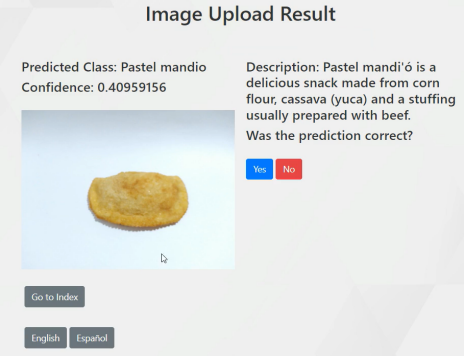

Fig IV: Website Result Page

5. RESULTS

AlexNet and Lenet were the best performing classification models. This might have been attributed to the small size of the dataset, the complex neural networks of VGG-16, GoogleNet and ResNet are not meant to be used to train a small dataset.

Architecture	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
LeNet	0.9714	0.8583	0.1113	0.374
AlexNet	0.9182	0.8583	0.2819	0.442
VGG_16	0.2119	0.1864	1.6072	1.6126
GoogleNet	0.1656	0.1604	2.8722	1.7952
ResNet	0.8406	0.3854	0.5386	42.9204

Fig V: image classification accuracy and loss table comparison between architectures

The object detection model performed on par with AlexNet and LeNet, being able to successfully classify different objects in the same image. The images with multiple objects were created by mixing different images since we could not obtain any more images for the dataset.

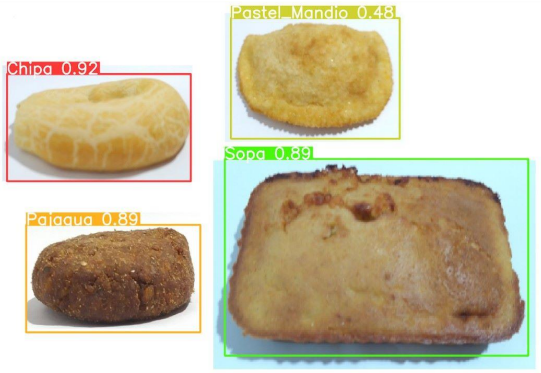

Fig VI: Object Detection sample prediction

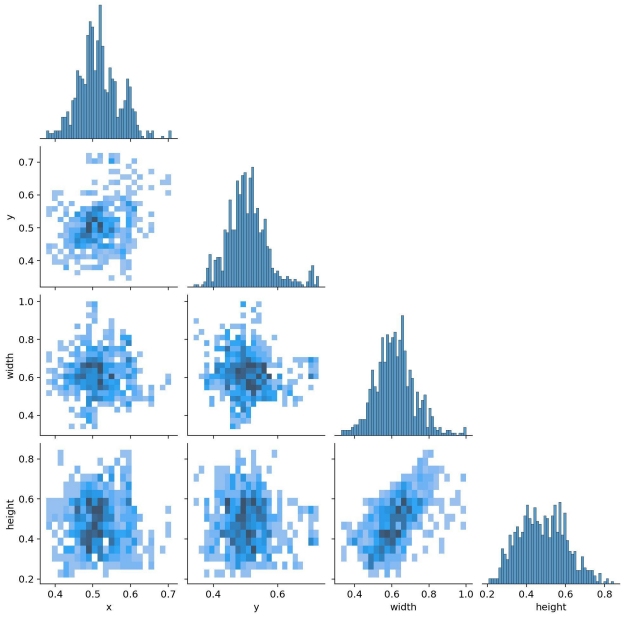

Fig VII: Object Detection Model Labels Correlogram.

6. CONCLUSION

In wrapping up, our project has successfully shown the power of machine learning in a niche domain - recognizing traditional Paraguayan snacks. Through curating our unique dataset, we've filled a gap in the global food recognition field, ensuring Paraguay's rich culinary heritage isn't left behind in the digital era.

The user-interactive nature of our website has fueled continuous system refinement and fostered a communal contribution toward preserving Paraguayan culture.

Finally, by integrating the YOLO object detection model, we've expanded our system's capabilities, cementing its practical value in real-world applications. All in all, this project underlines the importance of inclusivity in the digital recognition of global cuisines and sets a strong foundation for future work in this area.

7. REFERENCES

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems.

[2] Jiang, M. (2019). Food Image Classification with Convolutional Neural Network. Stanford CS230.
[3] Lu, Y. (2016). Food Image Recognition by Using Convolutional Neural Networks. arXiv preprint arXiv:1612.00983.
[4] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[5] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In IEEE Conference on Computer Vision and Pattern Recognition (pp. 248–255). IEEE.
[6] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324.
[7] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting.
[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).
[9] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 779-788).
[10] Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7263-7271).
[11] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Advances in Neural Information Processing Systems.