Tomato Classification for Quality Assessment Using CNN with Attention Mechanism in Python

Kamalakannan M
M.sc Data Analytics
Bharathiar University

Nivetha C
M.sc Data Analytics
Bharathiar University

Suji D
M.sc Data Analytics
Bharathiar University

Sujitha A
M.sc Data Analytics
Bharathiar University

Sumaiya S
M.sc Data Analytics
Bharathiar University

Abstract

Tomato quality assessment is a critical task in the agricultural industry, impacting both consumer satisfaction and supply chain efficiency. Traditional manual inspection methods are time-consuming , leading to the need for automated approaches. In this work, present a deep learning-based solution for the classification of tomatoes into four quality categories: Ripe, Unripe, Old, and Damaged, using an image dataset. The proposed methodology employs Convolutional Neural Networks (CNN) to automatically extract features from images and classify them into the specified categories. To enhance performance, an attention mechanism is integrated into the CNN architecture, allowing the model to focus on the most informative regions of the images, leading to improved classification accuracy.

The process begins with extensive preprocessing, including image resizing, scaling, and augmentation, to increase dataset variability and robustness. A CNN model is designed with multiple convolutional layers, followed by an attention layer to capture spatial dependencies. The model is then trained and validated on a labeled dataset, with evaluation metrics including accuracy, precision, recall, and F1-score.

Results demonstrate that the attention-enhanced CNN outperforms traditional CNN models in terms of accuracy and classification quality, making it a promising tool for automated tomato quality assessment. This work offers a scalable and efficient solution to improve sorting and grading processes in agricultural production, contributing to reduced waste and improved supply chain management.

Introduction

Assessing tomato quality is essential to ensuring customer satisfaction and reducing waste in the food supply chain. For automated sorting and grading procedures in the agricultural industries, accurate tomato categorization into categories like ripe, unripe, old, and damaged is crucial. Conventional manual inspection techniques require a lot of work and are prone to mistakes by people. Deep learning methods, in particular Convolutional Neural Networks (CNN), have become effective tools for picture classification tasks in order to address these issues.

In this project, an image dataset to identify tomatoes using a deep learning approach and Attention mechanism . Ripe, Unripe, Old, and Damaged are the four different classifications into which the model will be taught to classify tomatoes. CNNs are the foundation of our system because of their capacity to automatically learn and extract information from photos.

An attention mechanism to improve the model's performance, which aids the network in concentrating on the most important areas of the image and raises the accuracy of the classification.

The procedure starts with preprocessing the dataset, which includes scaling, resizing, and augmentation. Next, a CNN model with attention layer integration is built. Accuracy, precision, recall, and F1 score are used to assess the model's performance after it has been trained and verified using the proper metrics. By increasing tomato quality classification's precision and efficiency, this method seeks to enhance agricultural production systems' overall efficacy.

Problem Statement

In the agricultural industry, ensuring the quality of tomatoes is essential for maintaining consumer satisfaction and optimizing supply chain processes. The manual inspection of tomatoes, which classifies them into categories like ripe, unripe, old, and damaged, is both labor-intensive and prone to human error. This can lead to inefficiencies, such as misclassification, product waste, and inconsistent quality in the market. To address these challenges, automated systems using image classification techniques have become increasingly important. However, accurately distinguishing between different quality states of tomatoes presents a unique challenge due to variations in color, texture, and appearance across the different categories.

Methodology

This dataset consists of a total of 7226 images of tomatoes in four different states: Unripe, Ripe, Old, and Damaged. The dataset has been split into a training set and a validation set in a 9:1 ratio, with 90% of the images used for training and the remaining 10% used for validation. This dataset can be used for image classification tasks, particularly for identifying the state of tomatoes. The images were carefully selected to ensure a diverse range of tomato states and were collected from various sources.

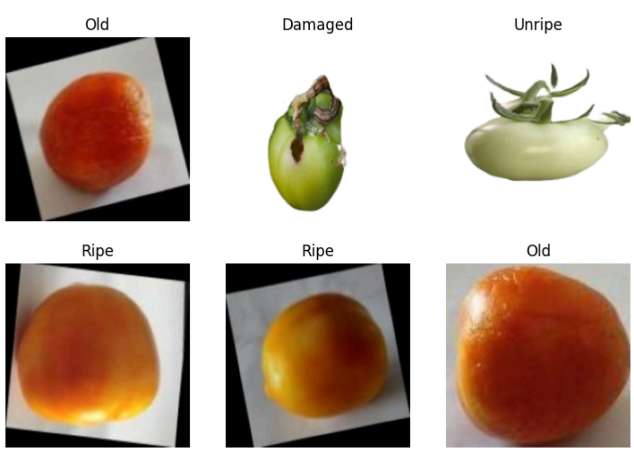

Figure 4.1 Classes of Tomatoes

Some Important Features

- Color,
- Texture,
- Shape

Workflow

The tomato classification system workflow offers a structured approach to building an efficient model, beginning with data acquisition and preprocessing to ensure the dataset is diverse and optimized for training. The design of a CNN with an attention mechanism allows the model to focus on key image features, improving its accuracy in distinguishing between the four quality categories (Ripe, Unripe, Old, and Damaged). By validating and evaluating the model with appropriate metrics like accuracy and F1-score, it ensures reliable performance and generalization. The deployment phase enables practical, automated tomato quality assessment, enhancing efficiency and reducing human error in real-world applications.The tomato classification system workflow offers a structured approach to building an efficient model, beginning with data acquisition and preprocessing to ensure the dataset is diverse and optimized for training. The design of a CNN with an attention mechanism allows the model to focus on key image features, improving its accuracy in distinguishing between the four quality categories (Ripe, Unripe, Old, and Damaged). By validating and evaluating the model with appropriate metrics like accuracy and F1-score, it ensures reliable performance and generalization. The deployment phase enables practical, automated tomato quality assessment, enhancing efficiency and reducing human error in real-world applications.

Proposed Model Architecture Explanation

1. **Input Layer:**

Shape: The input layer expects images of size `(150, 150, 3)`, where `150x150` are the dimensions, and `3` represents the RGB color channels.

Purpose: This layer is the starting point, where images are fed into the network.

2. **Convolutional Layers:**

Layer 1: `Conv2D` with 32 filters, each of size `(3x3)`, followed by `ReLU` activation and `MaxPooling2D`.

■ **Purpose:** Detects low-level features like edges and textures from the input images.

Layer 2: `Conv2D` with 64 filters, each of size `(3x3)`, followed by `ReLU` activation and `MaxPooling2D`.

■ **Purpose:** Captures more complex patterns like shapes and object parts by combining features detected by the previous layer.

Layer 3: `Conv2D` with 128 filters, each of size `(3x3)`, followed by `ReLU` activation and `MaxPooling2D`.

■ **Purpose:** Extracts even more complex and abstract features, such as the overall structure of the tomato.

3 . Attention Mechanism:

Attention Map: A `Conv2D` layer with a single filter of size `(1x1)` and `sigmoid` activation is used to create an attention map. This map highlights important regions of the image that are crucial for the classification task.

Multiplication: The attention map is multiplied with the feature maps from the last convolutional layer to emphasize relevant features and suppress less important ones.

Purpose: Focuses the network on the most relevant parts of the image, improving the accuracy of the classification by helping the model "pay attention" to critical areas like spots, color gradients, or damaged regions of the tomato.

4 . Flatten Layer:

Flattening: The multi-dimensional output of the last convolutional layer is flattened into a one-dimensional vector.

Purpose: Prepares the data for the fully connected layers, which are typically one-dimensional.

Fully Connected (Dense) Layers:

Layer 1: A dense layer with 512 neurons and `ReLU` activation.

■ Purpose: Combines the features extracted by the convolutional layers to learn higher-level patterns and representations.

Output Layer: A dense layer with 4 neurons and `softmax` activation.

■ **Purpose:** Outputs the probabilities of the image belonging to each of the four classes (ripe, unripe, old, damaged). The class with the highest probability is chosen as the predicted label.

RESULTS AND DISCUSSION:

The dataset used is the "Tomatoes Dataset," downloaded from Kaggle. It contains images of tomatoes categorized into four classes: Ripe, Unripe, Old, and Damaged. The dataset has 6,500 training images and 724 validation images distributed across these four classes. The training and validation sets are used for training and evaluating the model.

The data preprocessing and augmentation with resizing images are rescaled by a factor of 1./255, normalizing pixel values to a range of 0 to 1. This helps the model converge faster and avoid issues caused by large pixel value ranges and Data augmentation techniques are applied to the training data to increase the dataset's variability and reduce overfitting. That includes random rotations (up to 20 degrees), width and height shifts (up to 20% of the image size), shear and zoom transformations, horizontal flips, nearest neighbor fill for newly introduced pixels during transformations.

The model consists of three convolutional layers with 32, 64, and 128 filters, respectively. Each convolutional layer uses a 3x3 kernel and `ReLU` activation and maxPooling is applied after each convolution to reduce the spatial dimensions. The attention mechanism has an attention block that is introduced to give the model the ability to focus on important regions of the image. This mechanism helps improve the model's performance by weighting the relevant parts of the image more heavily.

Dense layers after flattening the output of the convolutional layers, a dense layer with 512 units is added with `ReLU` activation. The final output layer has 4 units with a `softmax` activation function, corresponding to the four classes.

- **Optimizer**: The model is compiled using the Adam optimizer.

- **LossFunction**: categorical_crossentropy is used as the loss function,
- **Metrics**: Accuracy is used as the evaluation metric during training and validation. Validation Accuracy = 88.12%, Validation Loss = 0.3184 the model demonstrated consistent improvement in accuracy and reduction in loss over the training epochs, indicating that it was learning effectively from the data.
- **Confusion Matrix Analysis** : The confusion matrix provides insights into how well the model differentiates between the four classes (Damaged, Old, Ripe, Unripe).

Correct Classifications: The diagonal of the confusion matrix shows the number of correct predictions for each class:

- Damaged: 10 correct predictions
- Old: 84 correct predictions
- Ripe: 74 correct predictions
- Unripe: 50 correct predictions

Misclassifications:

- Damaged tomatoes were often misclassified as Old, Ripe, or Unripe.
- Old tomatoes were misclassified as Ripe or Unripe.
- Ripe tomatoes were also misclassified as Old or Unripe.
- Unripe tomatoes were mistaken for other classes at a moderate rate.

This confusion suggests that the model struggles with differentiating between certain visual characteristics of these classes, particularly between Old and Ripe or Unripe tomatoes.

A prediction was made on an individual image from the validation set (Old/o (1122).jpg). The predicted class for this image was Unripe, which indicates that the model was incorrect in this case. This is a specific example of the challenge highlighted in the confusion matrix analysis, where there is some overlap in feature recognition between different classes.

The model achieves a promising validation accuracy of **88.12%**, indicating that it performs well on the tomato classification task. However, the confusion matrix analysis reveals areas where the model struggles, particularly in distinguishing between classes with overlapping features such as Old and Ripe or Unripe tomatoes. By incorporating additional data augmentation, advanced architectures, and more sophisticated attention mechanisms, the model's performance could be further improved.

Reference

[1] Attention mechanisms in computer vision: A survey
Meng-Hao Guo1, Tian-Xing Xu1, Jiang-Jiang Liu2, Zheng-Ning Liu1, Peng-Tao Jiang2, Tai-Jiang Mu1, Song-Hai Zhang1, Ralph R. Martin3, Ming-Ming Cheng2, and Shi-Min Hu1

[2] Deep learning-based approach for tomato classification in
complex scenes Mikael A. Mousse1,* , Bethel C. A. R. K. Atohoun2 , and Cina Motamed3

[3] Type of Tomato Classification Using Deep Learning
Mahmoud A. Alajrami, Samy S. Abu-Naser

[4] Classification of Tomato Fruit Using Yolov5 and Convolutional Neural Network Models by Quoc-Hung Phan 1,*ORCID,Van-Tung Nguyen 1ORCID,Chi-Hsiang Lien 1,*,The-Phong Duong 2,Max Ti-Kuang Hou 1 andNgoc-Bich Le 3,4

[5] Tomato Ripeness Detection and Classification using VGG based CNN Models
Authors
Seetha Ram Nagesh Appe
Research Scholar, Department of Computer Science and Engineering, Annamalai University, Chidambaram