International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Modify Resnet 101: Quality Evaluation of Fruit Recognition and Classification based on Deep Learning

Nareen O. M. Salim[1], Ahmed Khorsheed Mohammed [2]

Submitted: 28/01/2024 **Revised:** 06/03/2024 **Accepted:** 14/03/2024

Abstract: Recognizing and categorizing fruits is a unique challenge in the field of computer vision due to the wide variation in their appearance, shape, and texture. In this research, the effectiveness of an enhanced ResNet-101 deep convolutional neural network (CNN) architecture to achieve higher accuracy in fruit recognition and classification is explored compared to other deep learning (DL) models. The improved ResNet-101 model leverages its depth, skip connections, and pre-trained weights to capture intricate and distinctive features, making it particularly adept at handling complex visual tasks. In our experiments, it was observed that this enhanced ResNet-101 model outperforms several other cutting-edge DL models when applied to a diverse and demanding fruit classification dataset. The model's robustness in dealing with a wide range of fruits, including those with subtle variations and occlusions, is highlighted in our study. Furthermore, the significance of tuning hyperparameters, applying data preprocessing techniques, and utilizing data augmentation strategies is delved into. These aspects are deemed crucial for optimizing the performance of the enhanced ResNet-101 model. By striking the right balance between model complexity and dataset size, the issue of overfitting is addressed while still achieving exceptional accuracy. In essence, the presented model showcases enhanced accuracy performance within 15 epochs, achieving an impressive 99.97% accuracy across a dataset comprising 40 different types of fruits. The findings underscore the superior performance of the proposed model, surpassing the efficacy of several commonly employed methods in current use. This model's effectiveness and robustness for fruit recognition and classification are demonstrated. By reducing the number of trainable parameters by 81%, the overall complexity of the model is decreased. Faster training times, fewer computational resources required, potentially improved generalization, and a reduced likelihood of overfitting can be achieved through this reduction in complexity.

Keywords: *Deep Learning (DL), Convolutional neural network (CNN), Recognition, Classification, ResNet101.*

1. Introduction

Fruits are a crucial part of our daily diet, providing us with essential nutrients, vitamins, and minerals that are vital for our health and overall well-being. The ability to recognize and categorize fruits plays a significant role in various fields, spanning agriculture, food processing, nutrition tracking, and even computer vision applications [1,2]. In recent years, the domain of fruit recognition and classification has seen remarkable progress, primarily due to the remarkable capabilities of ML and DL techniques, with a particular focus on Convolutional Neural Networks (CNNs), including the powerful ResNet-101 model [3, 4].

The significance of fruit recognition and classification cannot be overstated. In agriculture, the accurate identification of fruits holds immense importance for estimating crop yields, managing pests and diseases, and assessing the overall quality of harvests. Within the food industry, fruit classification ensures consistent product quality and aids in automating sorting and packaging procedures [5]. Furthermore, the capacity to identify fruits in images or videos is indispensable for applications like smart agriculture, autonomous robotics, and even personalized dietary monitoring. ML, and more specifically, DL, have brought about a revolution in the realm of fruit recognition and classification [6]. CNNs, which draw inspiration from the human visual system, have proven themselves exceptionally adept at learning intricate patterns and features from image data. Among these models, ResNet-101, a variant of CNN, stands out for its exceptional depth and performance, rendering it an indispensable tool for various image classification tasks. In this era of technological advancement, the integration of CNNs like ResNet-101 into fruit recognition and classification tasks has yielded highly accurate and efficient solutions [5, 7]. These DL models have the ability to handle extensive datasets, discern complex attributes of fruits, and generalize effectively across diverse fruit varieties and environmental conditions [8]. As a result, they empower industries to achieve greater productivity, maintain quality control, and foster innovation in the domains of fruit processing, agriculture, and beyond [9]. This research [10] paper presents a thorough investigation into the utilization of DL models for categorizing various apple ieties. The investigation utilizes five distinct convolutional neural

1 Department of Information Technology, Faculty of Computers & Information, Duhok Polytechnic University, Duhok, 42001 Iraq nareen.mohameed@dpu.edu.krd;
2 Department of electrical and computer engineering, College of Engineering, University of Duhok, Duhok 42001, Iraq
ahmad.khorshed@uod.ac

networks (CNNs), derived from two distinct architectural structures: series networks (namely, AlexNet and VGG-19) and DAG networks (consisting of ResNet-18, ResNet-50, and ResNet-101). The primary aim is to recognize and categorize 13 different apple classes. Notably, the study highlights a substantial impact of dataset configuration on classification outcomes [10]. All models achieve an overall classification accuracy surpassing 96.1% on dataset A (with a training-to-testing split of 4:1) and 94.5% on dataset B (with a training-to-testing split of 3:2). Moreover, the authors delve into enhancing the interpretability of their models using techniques such as feature visualization images, scrutinizing the most robust activations, and incorporating LIME images. These methodologies illuminate the processes and reasons behind the classification decisions made by diverse trained models. [10]. However, it's important to note that certain behaviors exhibited by these models remain unexplained, owing to the inherent "black box" nature of DL technology, which still lacks a full understanding of its internal mechanisms. In the exploration detailed in research paper [11], DL algorithms take center stage in the classification of fruit maturity and quality detection. The authors meticulously reviewed 35 papers that investigated ML and DL techniques for fruit maturity classification. Their comprehensive analysis led them to the firm conclusion that Convolutional Neural Networks (CNN) emerge as the most suitable algorithm for this specific task. Within their study, the authors introduced two distinctive methodologies utilizing DL techniques—specifically CNN and AlexNet—for the thorough examination of banana fruit images. The assessment covered a spectrum of datasets, including their original dataset, an augmented dataset, and a well-established dataset (Fruit 360). To enhance the richness of the dataset, they implemented augmentation techniques and devised an advanced CNN model tailored to accurately discern and categorize the maturity stages of bananas. The paper provides a detailed exposition of their proposed methodology, covering aspects such as data collection, model development, and training and validation procedures [11]. The outcomes and discussions underscore that the proposed models effectively achieve exceptional classification and prediction accuracy for fruits across various ripening stages. It is noteworthy that the CNN model demonstrated an accuracy rate of 98.25% on the original dataset, while the AlexNet model achieved 81.75%. On the augmented dataset, the CNN model reached an impressive accuracy of 99.36%, and the AlexNet model outperformed with an even higher accuracy of 99.44%.

2. Methods and Materials

In general, the process of recognizing and classifying images involves a combination of various techniques, including structural analysis, statistical measurements, and spectral analysis. When it comes to statistical measurements, common metrics like mean, variance, and entropy are typically employed. Structural analysis, on the other hand, is used for object recognition and identification. Spectral approaches come into play when identifying images based on characteristics like intensity, texture, and color features, although it's not always standard to employ all these methods. To address the challenges outlined above, this article employs DL methods, specifically CNN models such as AlexNet, ResNet50, and ResNet101, for the classification and identification of fruit images [8, 11]. The proposed simulation and experimental work for fruit recognition using DL techniques are divided into several distinct phases.

2.1 Image Acquisition

In the course of performing experiments and simulations, the initial phase entails acquiring images. Subsequent to obtaining the images, the subsequent step involves carefully selecting specific images and subsequently resizing them to conform to a standardized dimension. In the context of experimental work, it's crucial to organize these images based on their sizes. As for simulations, it becomes necessary to define the parameters used for classification. Given the diverse and distinct characteristics of different fruits, a critical requirement is to categorize them based on specific features. Among these features, the classification of intensity and texture presents a significant challenge, and this complexity is effectively addressed through the use of CNNs [13].

2.2 Pre-processing

Pre-processing operates at an abstract level on the fruit images and aims to enhance the intensity levels. The primary objective is to mitigate any undesired distortions in the images and to optimize the specific portions of the acquired image that are subsequently employed in the processes of feature selection and extraction [14].

2.3 Feature extraction

The most effective features are obtained following the pre-processing stage. Features are essentially numerical values derived from image data through mathematical operations [15]. When it comes to recognition tasks, the selection of these features is of the utmost significance. Optimal feature subset selection, determined by the performance of classifiers, enhances the classification accuracy and success rate of the recognition system [16].

2.4 Image Classification

A classification task involves identifying a specific input using one of several distinct classifiers. In the study presented, DL techniques are employed to categorize fruit images [17]. The process of fruit image classification comprises several steps, such as image enhancement, object detection, recognition, and classification.

2.5 Evaluation

The evaluation phase marks the concluding step in the process of image classification. The proposed work tests a multi-model approach for classifying fruit images [18]. The phases of this multi-model DL model are illustrated in Figure 1 below.

Fig. 1. Phases of the proposed work.

2.6 Proposed Fruit Classification Multi-Mode

A CNN structure model (AlexNet, ResNet50, and ResNet101) was proposed for the recognition of fruits using DL algorithms. Recently, Deep Learning has garnered popularity as a method for recognizing and classifying images [18]. It is within the unsupervised learning branch of Artificial Neural Networks that DL has emerged as a promising technique for image recognition and classification. The primary purpose of utilizing CNN architectures like AlexNet, ResNet50, and ResNet101 for fruit recognition and classification is to harness the capabilities of DL for the precise and efficient categorization of fruit images [19]. These CNN models are particularly well-suited for tasks involving image recognition due to their capacity to autonomously acquire and extract pertinent features from the input images. Regarding their individual contributions:

AlexNet: AlexNet is renowned for its pioneering work in the realm of DL and image classification. Its design is specifically tailored to detect and distinguish objects within images, making it valuable for identifying various types of fruits based on their visual attributes [20].

ResNet50 and ResNet101: ResNet, or Residual Networks, represents a deep neural network architecture that excels in the extraction of features and the classification of objects. ResNet50 and ResNet101, as variants of ResNet with 50 and 101 layers, respectively, exhibit the capability to capture intricate details and characteristics within fruit images, leading to more precise and detailed classification outcomes [23, 25].

The adoption of these CNN structures serves the objective of enhancing the accuracy and robustness of fruit recognition and classification, particularly in scenarios where fruits exhibit variations in shape, color, and size. These DL models automate the process and elevate the system's competence in classifying a wide array of fruits with a high degree of accuracy.

2.6.1 Conventional neural networks (CNN)

Traditional neural networks, alternatively known as artificial neural networks, are machine learning algorithms inspired by the neural networks observed in the human brain [13]. These networks are comprised of interconnected nodes, referred to as neurons, possessing the ability to process and convey information in the form of electrical signals [9]. Neurons are organized into layers, with each layer being linked to the next. The weights assigned to connections between neurons play a crucial role in determining the strength of these connections. In the neural network, the input layer is where initial data is received, while the output layer produces the final results. Situated between these layers are hidden layers, responsible for learning patterns in the input data and generating the desired output. Throughout the training process, optimization techniques such as backpropagation are employed to adjust connection weights until the desired output is achieved. This layered structure empowers traditional neural networks to effectively learn and adapt to new data, establishing them as a valuable tool in the realm of machine learning applications [13]. Figure 2 visually represents the fundamental structure of the CNN network.

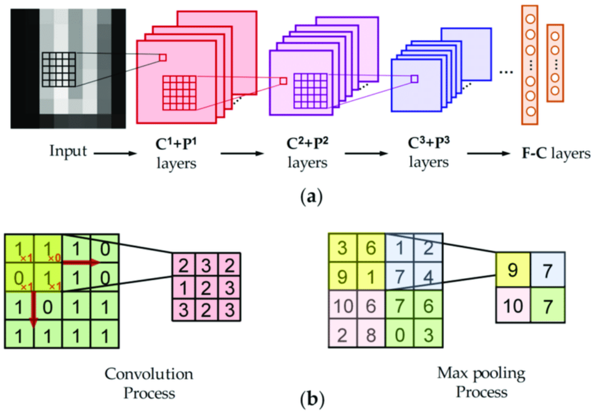

Fig. 2. CNN network's structure.

2.6.1.1 AlexNet

AlexNet, created by Alex Krizhevsky, is a deep CNN that achieved a groundbreaking victory in the 2012 ImageNet Large Scale Visual Recognition Challenge [21]. It marked a significant advancement in computer vision and serves as the foundational framework for most contemporary deep neural networks. AlexNet comprises eight layers, including five convolutional layers, three fully-connected layers, and a concluding output layer. The initial layer in AlexNet functions as the input layer, receiving and passing input images to subsequent convolutional layers [20]. These convolutional layers are instrumental in extracting image features and identifying objects within them. The first two convolutional layers employ 96 and 256 filters, with filter sizes of 11×11 and 5×5, respectively. After every convolutional layer, a subsequent max-pooling layer is implemented to reduce the spatial dimensions of the convolutional layer's output. The third convolutional layer utilizes 384 filters and is succeeded by another max-pooling layer. Similarly, the fourth convolutional layer employs 384 filters and is followed by an additional max-pooling layer. The fifth and final convolutional layer consists of 256 filters. Responsible for object categorization in the images, the three fully-connected layers are each equipped with 4096 neurons, and the last layer contains 1000 neurons [22]. The ultimate SoftMax layer in the output is instrumental in predicting the classes of objects depicted in the images. The structural layout of AlexNet is presented in Figure 3.

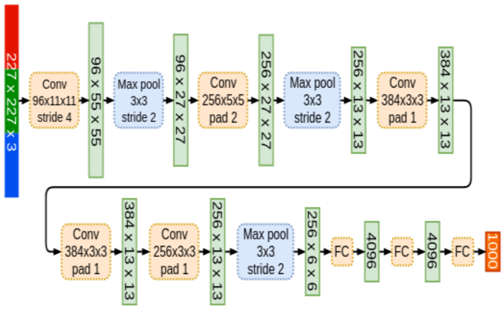

Fig.3. The AlexNet architecture

2.6.1.2 ResNet50

ResNet50, a deep convolutional neural network (CNN) comprising 50 layers, is structured into five phases, each housing a convolution and identity block. Whether convolutional or identity, each block consists of three layers. The ResNet50 architecture boasts over 23 million trainable parameters. Notably, the top layers in ResNet50 are left unfrozen, enabling them to learn through backpropagation, while the remaining layers remain frozen [23]. This process involves crafting a ResNet50 model with pre-trained weights derived from the ImageNet dataset, allowing the model to classify images across 1,000 different categories. Figure 4 illustrates the architecture of the employed ResNet50 model, showcasing a deep residual learning framework with 50 layers and five stages [24]. This framework is specifically tailored for the large-scale classification of fruits in their natural environment. The cross-entropy loss function for the ResNet50 model is expressed as follows:

$$Loss = -\left(\frac{1}{n}\right)\sum_{i=0}^{n}\log P(N) \quad (1)$$

Where n is the output size of the classification network.

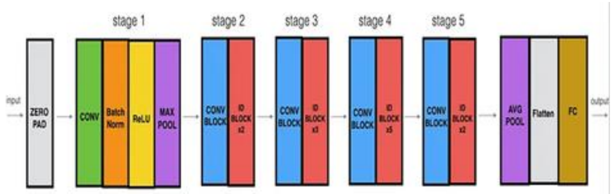

Fig.4. ResNet50 Architecture.

2.6.1.3 ResNet101

As mentioned before, ResNet101 is a deep CNN developed by Microsoft Research for image classification. It is based on the ResNet architecture, which consists of a series of convolutional layers, pooling layers, and fully connected layers. ResNet101 has 101 layers, including a total of 101 convolutional layers, and the number of parameters is about 44 million. The network has a large number of learnable parameters, making it more accurate than other networks [25]. ResNet101 was designed to improve the generalization of object recognition and scene understanding from images. In addition to object recognition, the network is also used for semantic segmentation, image captioning, and image generation. ResNet101 is widely used in a variety of computer vision tasks and applications, such as autonomous driving, medical imaging, and facial recognition [26]. It has been used in many research projects and applications, and it continues to be an important model for computer vision. This neural network is included in the MATLAB DL toolbox and pre-trained on the ImageNet dataset. The network is called by the command resnet101 and displayed as a two tables network [27]. First table consists of 347 × 1 layer which are conventional layers and batch normalization layers with the ReLU activation functions layer and max pooling layers. The second table consists of the connections between these different layers as a source connection layer and destination one [28]. Figure 5 shows this structure in MATLAB, and Figure 6 shows the layers of the ResNet101 network in MATLAB.

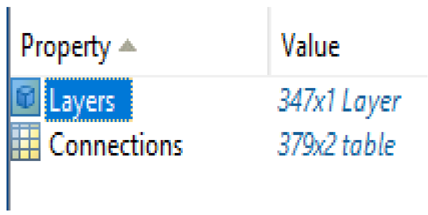

Fig.5. structure of ResNet in MATLAB, consists of layers and connections

	1
1	1x1 ImageInputLayer
2	1x1 Convolution2DLayer
3	1x1 BatchNormalizationLayer
4	1x1 ReLULayer
5	1x1 MaxPooling2DLayer
6	1x1 Convolution2DLayer
7	1x1 BatchNormalizationLayer
8	1x1 ReLULayer
9	1x1 Convolution2DLayer
10	1x1 BatchNormalizationLayer
11	1x1 ReLULayer
12	1x1 Convolution2DLayer
13	1x1 BatchNormalizationLayer
339	1x1 ReLULayer
340	1x1 Convolution2DLayer
341	1x1 BatchNormalizationLayer
342	1x1 AdditionLayer
343	1x1 ReLULayer
344	1x1 AveragePooling2DLayer
345	1x1 FullyConnectedLayer
346	1x1 SoftmaxLayer
347	1x1 ClassificationOutputLayer

Fig.6.example of ResNet101 layers in MATLAB Toolbox

The input layer is of size $224 \times 224 \times 3$, so all the images need to be resized to this size to work with this neural network. where the output of the network is one of the 1000 classes the network was trained on. This network was trained on several objects in the ImageNet dataset; some of them were the main types of fruits. It was not trained on all types of fruits. So, when it was used for fruit prediction, the accuracy was 27% because a large number of fruits ware not in the ImageNet dataset. Figure 7 shows the output layer of ResNet101 in MATLAB software.

Property ▲	Value
Name	'ClassificationLayer_p...
Classes	1000x1 categorical
OutputSize	1000
LossFunction	'crossentropyex'

Fig. 7. output layer of ResNet101

2.6.1.4 ResNet101 modified.

ResNet, short for residual network, plays a pivotal role in addressing computer vision challenges. ResNet101 [28] is comprised of 104 convolutional layers organized into 33 blocks, with 29 of these blocks directly incorporated into previous ones. Initially trained on the ImageNet dataset featuring 1000 object classes, the original architecture is depicted in Figure 8. This illustration highlights the processing of input images through residual blocks, each consisting of multiple layers. In our study, we made modifications by eliminating the fully connected (FC) layer associated with the original 1000 object classes. Instead, we introduced a new FC layer tailored to our specific number of classes. In our selected dataset, the number of classes is forty types of fruit, such as (Apple golden 3,Apple Red 1, Apple Red Yellow 1,Apricot, Avocado, Banana Lady Finger, Bana Red, Beetroot, Blueberry, Cherry1, Cherry Rainier, Cocos, Corn, Cucumber Ripe, Cucumber Ripe 2, Dates, Eggplants, Fig, Grape Pink, Grape White, Kaki, Kiwi, Lemon, Lychee, Mango, Mulberry, Onion Red, Onion White, Orange, Papaya, Peach, Pear, Pepper Green, Plum, Potato White, Raspberry, Strawberry, Tomato 1, Watermelon). The adjusted model maintains a uniform input size of 224 × 224 × 3 and yields an output of N × 3. Figure 8 visually represents this modified model, highlighting its configuration with a reduced number of layers while preserving the same sequence.

Building this network in MATLAB begins with the definition of the layers, conventional layer sizes. Padding was performed on corners and edges for accurate filter passing. Batch normalization layers and the ReLU activation function layer. The Pooling layers were chosen of the type “Average” because the colors of the near pixels of the fruit was much similar. figure 8 shows the structure of the network was used in proposed method.

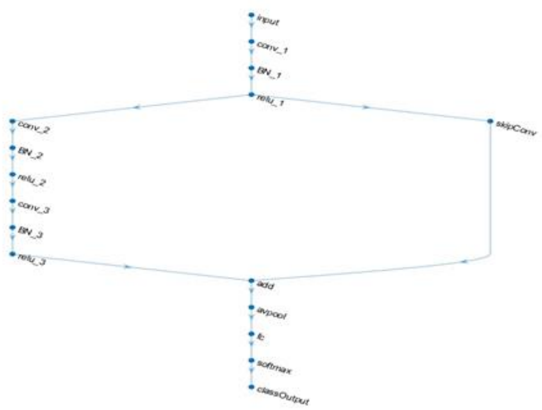

Fig.8. smaller structure of ResNet101 for fruit recognition

2.7 Dataset

The dataset employed for training and testing, known as Fruits 360 and publicly available on Kaggle, contributed all the images used in this study. Comprising approximately 90,380 images, the dataset encompasses

131 diverse fruits and vegetables [29]. In the image acquisition process, a motor rotated the fruits, generating frames for photographing. A white piece of paper served as the backdrop, and to mitigate uneven lighting, a flood-fill technique was applied to isolate the fruits from the background. Following background elimination, all fruits were standardized to 100x100 pixels in standard RGB format (thus, three values for each pixel). As depicted in figure 9, the training dataset comprises 67,692 photos, while the test dataset includes 22,688 photos [29]. These images represent a variety of fruits sourced from the 360 datasets.

Fig. 9. fruits classes in 360 datasets.

2.8 Confusion Matrix

The dataset employed for training and testing, known as Fruits 360 and publicly available on Kaggle, contributed all the images used in this study. Comprising approximately 90,380 images, the dataset encompasses 131 diverse fruits and vegetables [29]. In the image acquisition process, a motor rotated the fruits, generating frames for photographing. A white piece of paper served as the backdrop, and to mitigate uneven lighting, a flood-fill technique was applied to isolate the fruits from the background. Following background elimination, all fruits were standardized to 100x100 pixels in standard RGB format (thus, three values for each pixel). As depicted in figure 9, the training dataset comprises 67,692 photos, while the test dataset includes 22,688 photos [29]. These images represent a variety of fruits sourced from the 360 datasets.

1. True Positives (TP): Instances where the model accurately predicted the positive class.

2. True Negatives (TN): Situations where the model correctly predicted the negative class.

3. False Positives (FP): Cases where the model predicted the positive class incorrectly, given that the actual class was negative.

4. False Negatives (FN): Instances where the model inaccurately predicted the negative class, whereas the true class was positive.

The Confusion Matrix is frequently depicted as a figure, roughly resembling this:

	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Fig.10. 2*2 Confusion Matrix

Calculating the proportion of samples that were correctly classified to all samples, as demonstrated in the example below, can be used to evaluate the classification models' accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \qquad (2)$$

Calculating accuracy involves dividing True Positives (TP) by the total sum of items with positive labels (the sum of TP and False Positives, FP), as indicated in formula 3. A high precision value signifies that the model and categorization process are generating a greater proportion of meaningful and accurate results.

$$Precision = \frac{TP}{TP+FP} \qquad (3)$$

Determining recall involves dividing the total count of components truly belonging to the positive class (True Positives, TP). This computation enables the assessment of the model's sensitivity, as outlined in formula 4.

$$Sensitivity = \frac{TP}{TP+FN} \qquad (4)$$

The F1-Score is derived by computing the harmonic mean of sensitivity and precision. This metric provides a balanced and comprehensive evaluation of the model's performance, considering both the ability to correctly identify positive instances (sensitivity) and the precision in categorizing them accurately.

$$F1 = 2 * \frac{(Precision * \text{sensitivity})}{(Precision + \text{sensitivity})} \quad (5)$$

Data scientists and ML practitioners can assess how well their models are working and make educated decisions regarding model modification and deployment using the Confusion Matrix and related metrics [41].

3. Results and Discussions

This section provides an in-depth exploration of the experimental procedures, encompassing details on the experimental setup, dataset characteristics, chosen evaluation metrics, and the ensuing results. The dataset utilized in this study encompasses a diverse array of 40 fruit classes. Post-preprocessing, the experimental dataset comprises a total of 24,636 images, representing 40 unique fruit categories. The allocation of the preprocessed dataset into training and testing subsets follows an 80:20 ratio, with 80% designated for training purposes and the remaining 20% reserved for rigorous testing, as outlined in Table 1.

The DL models employed in this study automatically extracted fruit features through a sequence of convolutional operations, eliminating the need for

Table 1. Name and number of each Fruit.

Number	Fruit Label	Training number	Testing number
1	Apple Golden 3	481	161
2	Apple Red 1	492	164
3	Apple Red Yellow 1	492	164
4	Apricot	492	164
5	Avocado	427	143
6	Banana	490	166
7	Banana Lady Finger	450	152
8	Banana Red	490	166
9	Beetroot	450	150
10	Blueberry	462	154
11	Cherry 1	492	164
12	Cherry Rainier	738	246
13	Cocos	490	166
14	Corn	450	150
15	Cucumber Ripe	392	130
16	Cucumber Ripe 2	468	156
17	Dates	490	166
18	Eggplant	468	156
19	Fig	702	234
20	Grape Pink	492	164
21	Grape White	490	166
21	Kaki	490	166
23	Kiwi	466	156
24	Lemon	492	164
25	Lychee	490	166
26	Mango	490	166
27	Mulberry	492	164
28	Onion Red	450	150
29	Onion White	438	146

30	Orange	479	160
31	Papaya	492	164
32	Peach	492	164
33	Pear	492	164
34	Pepper Green	444	148
35	Plum	447	151
36	Potato White	450	150
37	Raspberry	490	166
38	Strawberry	492	164
39	Tomato 1	738	246
40	Watermelon	475	157

manual feature extraction from the dataset containing 24,636 images of diverse fruits. The training process involved initializing epochs at 7, 10, and 15, with 3696, 5280, and 7920 iterations, a mini-batch size of 32, and a learning rate set to 0.001. The Adam optimizer was utilized for the learning process. A ten-fold cross-validation was conducted, employing multiple classifiers, and each classifier was assessed based on recall rate, precision, accuracy, and f1-score.

All simulations for this study were carried out using MATLAB 2022b. The computational system utilized was equipped with a Core i7 processor, 16GB of RAM, and an 8GB graphics card.

3.1 Results of Tested CNN pre-training Models

The proposed models were assessed for their performance in various fruit recognition scenarios depicted in images, employing diverse classification assessment tools such as accuracy, precision, sensitivity, specificity, and F1-score. The evaluation involved the utilization of a confusion matrix to scrutinize the proposed model, along with plotting the training accuracy. The suggested methodology underwent training for both 5 epochs and 15 epochs, utilizing a dataset initially partitioned into three segments: 80% for training, 10% for testing, and 10% for validation. The training and validation accuracy results for 5 epochs are illustrated in Table 2. Display the outcome classification of several types of CNN pre-training models, such as (AlexNet, ResNet50, and a compressed version of ResNet101), and modify ResNet101 achieving a maximum training accuracy of (100%), and the ResNet50 get the less accuracy (96.66%) during these models

Table 2. Performance evaluation of CNN pre-training models of 15 types of fruits

Model	Accuracy	Error-rate	Precision	Sensitivity	F1_ score
AlexNet	99.67%	0.33%	99.56%	99.62%	99.59%
ResNet50	96.66%	3.33%	97.77%	96.66%	97.21%
Modify ResNet101	100%	0.00%	100%	100%	100%

In the subsequent phase, we expanded the dataset by incorporating an additional 40 types of fruits, as outlined in Table 3. The training process involved retraining our models with the augmented dataset, maintaining the same number of epochs. This process yielded a set of distinct rates. Notably, during the pretraining of models, including AlexNet, ResNet50, and a modified version of ResNet101, it was observed that the compressed variant of ResNet101 achieved the highest accuracy among them.

Table 3. Evaluation of CNN pre-training models of 40 fruits

Model	Accuracy	Error-rate	Precision	Sensitivity	F1_score
AlexNet	99.67%	0.33%	99.56%	99.62%	99.59%
ResNet50	99.62%	0.37%	99.66%	99.63%	99.64%
Modify ResNet101	100%	0.00%	100%	100%	100%

In the third phase, following the success of the modified ResNet101 in outperforming other CNN pre-training models, we aimed to further explore and showcase the robustness and capabilities of the model. To achieve this, we experimented with altering the number of epochs and adjusting the image sizes within the dataset. This step was undertaken to assess the model's adaptability and to highlight its efficacy under varying training conditions. Both Table 4 and Table 5 encompass a range of values corresponding to the modified model, each obtained through the utilization of distinct numbers of epochs and varied image sizes. These tables collectively provide a comprehensive overview of the model's performance under diverse training configurations, offering a nuanced understanding of its adaptability and effectiveness across different scenarios.

Table 4. shown the Performance evaluation modify ResNet101 model in different type of epochs with various size of images(156×156)

No. of epochs	Accuracy	Error-rate	Precision	Sensitivity	F1_ score
7	99.57%	0.42%	99.62%	99.55%	99.58%
10	99.90%	0,09%	99.91%	99.94%	99.93%
15	99.95%	0.04%	99.95%	99.96%	99.95%

Table 5. shown the Performance evaluation modify ResNet101 model in different type of epochs with various size of images (224×224)

No. of epochs	Accuracy	Error-rate	Precision	Sensitivity	F1_score
7	99.73%	0.26%	99.70%	99.60%	99.70%
10	99.85%	0.14%	99.86%	99.83%	99.84%
15	99.97%	0.02%	99.97%	99.98%	99.97%

3.2 Discussion

Tables 2 and 3 showcase the remarkable performance of the proposed model, surpassing its counterparts in accuracy. Notably, when comparing models, ResNet50 exhibited lower accuracy than the proposed model. Despite the fact that the training images for the AlexNet model were of larger dimensions (227×227) compared to the modified ResNet101, the proposed model outperformed AlexNet in accuracy. This underscores the efficacy of the proposed model, which achieved superior results even with a smaller image size, showcasing its robustness and efficiency in fruit image classification tasks.

In this study, the utilization of confusion matrix plots was integral, seamlessly integrated with both training and validation plots. Furthermore, classification result plots were generated to vividly illustrate the diverse outcomes associated with the categorization of fruits. These results were meticulously analyzed across various DL (CNN pre-training) models, providing a comprehensive insight into the nuanced performance of each model in fruit classification. Figures 11, 12, and 13 provide a comprehensive overview of the outcomes derived from the training of three distinct models—AlexNet, ResNet50, and modified ResNet101. These models underwent 5 epochs of training for the classification of 15 different types of fruits. The figures vividly illustrate the accuracy rates and loss rates, encapsulating the performance nuances of each model in the fruit classification task. In the visual representation of a confusion matrix, the vertical axis denotes the predicted labels, while the horizontal axis corresponds to the true labels. The entries above and below the main diagonal in the confusion matrix depict instances of misclassification, with the diagonal elements capturing accurately identified occurrences. Figures 14 and 15 vividly illustrate the confusion matrix plots for the evaluated proposed model. These visuals provide a detailed breakdown of classification outcomes when the model is applied to datasets comprising 15 and 40 types of fruits, respectively. The graphical depictions offer insights into the model's performance by highlighting both correct and erroneous predictions

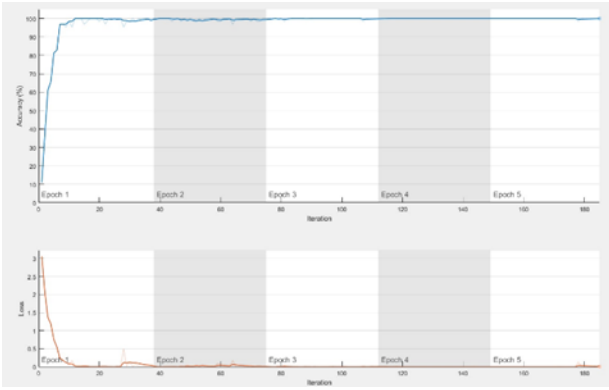

Fig.11: AlexNet training and validation plots for 5 epochs to recognize of 15 fruits dataset.

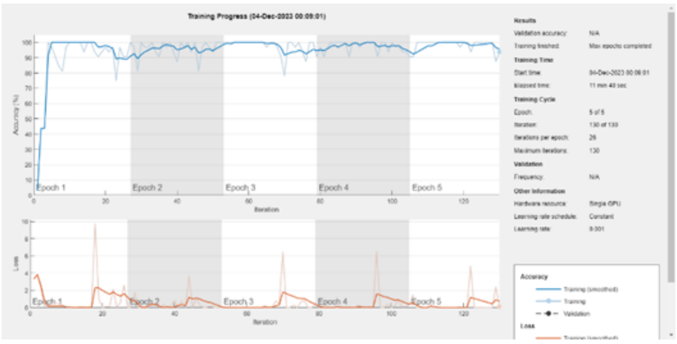

Fig.12: Evaluation of ResNet50's performance through the analysis of accuracy and loss metrics during the training and validation phases on a dataset containing 15 different fruits.

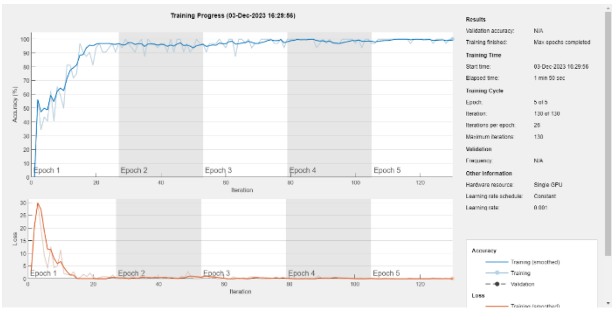

Fig.13: Suggest an evaluation of ResNet101's performance by conducting a thorough analysis of accuracy and loss metrics during both the training and validation stages, utilizing a dataset comprising 15 distinct fruit classes.

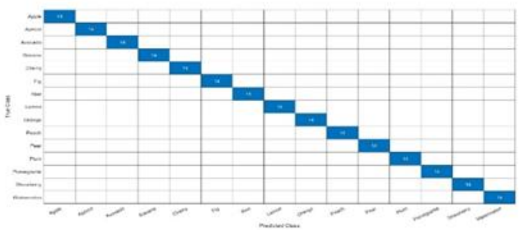

Fig.14: Present a confusion matrix illustration for the testing dataset containing 15 varieties of fruits, showcasing a detailed representation of classification results.

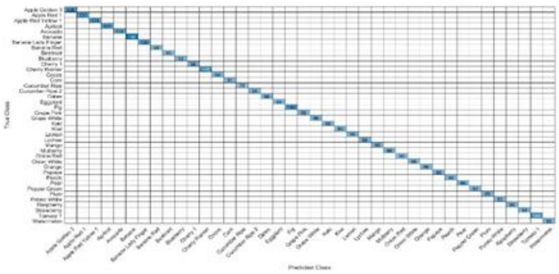

Fig.15: Provide a visual depiction of a confusion matrix for the testing dataset, encompassing 40 diverse types of fruits, to offer a comprehensive overview of the classification outcomes.

Furthermore, by augmenting the number of epochs and exploring diverse image sizes, the proposed model demonstrated its capacity to attain superior accuracy. This noteworthy achievement is visually depicted in the ensuing plots and the accompanying confusion matrix, illustrating the model's robust performance across various training configurations. The Figures 16 and 17 of plots showcase the discernible impact of increased epochs and varied image sizes on enhancing the model's accuracy, providing a comprehensive insight into the versatility and adaptability of the proposed approach.

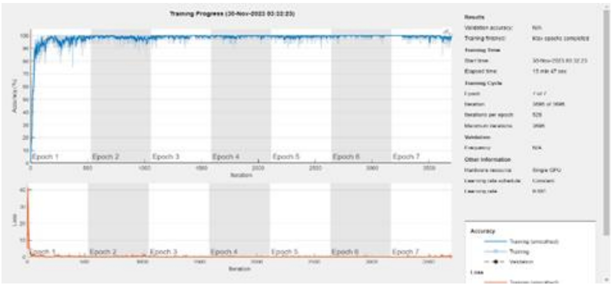

Fig.16: Representation encapsulates a thorough performance analysis of the modified ResNet101 model with 7 epochs and image size (156× 156).

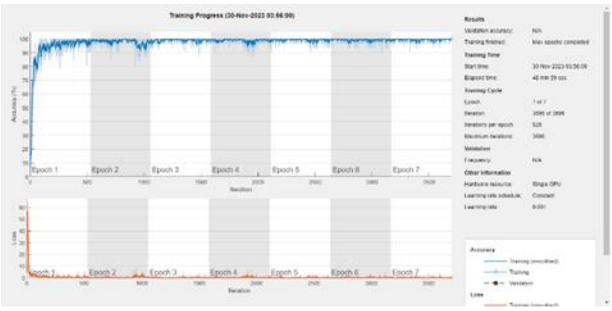

Fig. 17: Representation encapsulates a thorough performance analysis of the modified ResNet101 model with 7 epochs and image size (224× 224).

4. Conclusion

Commencing with a concise overview in the results section, this addition serves to delve into the analysis of the proposed framework. The findings prominently highlight the commendable performance of the proposed model on the selected dataset. Significantly, the proposed framework achieves heightened accuracy while utilizing a substantially reduced number of trainable parameters (weights) compared to the ResNet101 model, boasting an impressive 81% reduction. This reduction in parameters translates to a notable decrease in both training and testing times.

The proposed model continues to exhibit remarkable results across a spectrum of fruit types, ranging from 15 to 40, drawn from a database of 360. Notably outperforming the ResNet50 and AlexNet pre-trained CNN models, the proposed ResNet101 stands as a robust choice for recognition tasks. This success underscores its efficacy in enhancing accuracy while concurrently streamlining computational processes.

While the proposed ResNet101 algorithm represents a significant stride in improving recognition accuracy, it is acknowledged that the accuracy of advanced deep models is still an area for improvement and falls short of the latest techniques. Recognizing this, the algorithm was introduced in light of its potential advantages. A key limitation lies in the choice of employing both ResNet50 and AlexNet models for comparison, presenting an avenue for future exploration by focusing on a singular model to deepen the analysis. Future studies can pursue further investigation by utilizing databases featuring diverse fruit images characterized by the presence of multiple instances within the same fruit category or instances of various fruits captured in a single image and incorporating fruit images set against intricate backgrounds, such as those where fruits naturally coexist, to simulate real-world scenarios and enhance the robustness of the dataset.

References

[1] Pham, T. T., Nguyen, L. L. P., Dam, M. S., & Baranyai, L. (2023). Application of Edible Coating in Extension of Fruit Shelf Life. AgriEngineering, 5(1), 520-536.

[2] Bùi, V. P. (2023). Fruit type and weight recognition for payment purposes using computer vision (Doctoral dissertation, FPTU Hà Nội).

[3] Sugandi, B., Wahyu, R., Apriliana, S., & Putri, F. R. (2023). Fruit Recognition and Weight Scale Estimation Based on Visual Sensing. International ABEC, 234-241.

[4] Indira, D. N. V. S. L. S., Goddu, J., Indraja, B., Challa, V. M. L., & Manasa, B. (2023). A review on fruit recognition and feature evaluation using CNN. Materials Today: Proceedings, 80, 3438-3443.

[5] Lin, M. (2023). Fruit Classification based on ResNet and Attention Mechanism. Highlights in Science, Engineering and Technology, 34, 163-167.

[6] Wangchuk, K., Dorji, T., Dhungyel, P. R., Galey, P., & Chhetri, M. Fruits and Vegetables Recognition System in Dzongkha Using Visual Geometry Group Network.

[7] Kumari, R. S. S., & Gomathy, V. (2023). Fruit shop tool: Fruit classification and recognition using deep learning. Agricultural Engineering International: CIGR Journal, 25(2).

[8] Hussain, D., Hussain, I., Ismail, M., Alabrah, A., Ullah, S. S., & Alaghbari, H. M. (2022). A simple and efficient deep learning-based framework for automatic fruit recognition. Computational Intelligence and Neuroscience, 2022.

[9] Wang, J. J. (2022). Recognition system for fruit classification based on 8-layer convolutional neural network. EAI Endorsed Transactions on e-Learning, 7(23).

[10] Yu, F., Lu, T., & Xue, C. (2023). Deep Learning-Based Intelligent Apple Variety Classification System and Model Interpretability Analysis. Foods, 12(4), 885.

[11] Aherwadi, N., Mittal, U., Singla, J., Jhanjhi, N. Z., Yassine, A., & Hossain, M. S. (2022). Prediction of fruit maturity, quality, and its life using deep learning algorithms. Electronics, 11(24), 4100.

[12] Hassan, N. M. H., Elshoky, B., Hassan, N. M. H., Elshoky, B. R. G., & Mabrouk, A. M. (2023). Quality of performance evaluation of ten machine learning algorithms in classifying thirteen types of apple fruits. Indones. J. Electr. Eng. Comput. Sci, 30, 102-109.

[13] Gill, H. S., Murugesan, G., Mehbodniya, A., Sajja, G. S., Gupta, G., & Bhatt, A. (2023). Fruit type classification using deep learning and feature fusion. Computers and Electronics in Agriculture, 211, 107990.

[14] Tu, Bing, et al. "Feature extraction via joint adaptive structure density for hyperspectral imagery classification." IEEE Transactions on Instrumentation and Measurement 70 (2021): 1-16.

[15] Mutlag, W. K., Ali, S. K., Aydam, Z. M., & Taher, B. H. (2020, July). Feature extraction methods: a review. In Journal of Physics: Conference Series (Vol. 1591, No. 1, p. 012028). IOP Publishing.

[16] Nixon, M., & Aguado, A. (2019). Feature extraction and image processing for computer vision. Academic press.

[17] Dhiman, G., Kumar, A. V., Nirmalan, R., Sujitha, S., Srihari, K., Yuvaraj, N., ... & Raja, R. A. (2023). Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications. Multimedia Tools and Applications, 82(4), 5343-5367.

[18] Ekramirad, N., Khaled, A. Y., Donohue, K. D., Villanueva, R. T., & Adedeji, A. A. (2023). Classification of Codling Moth-Infested Apples Using Sensor Data Fusion of Acoustic and Hyperspectral Features Coupled with Machine Learning. Agriculture, 13(4), 839.

[19] Gao, G., Xu, Z., Li, J., Yang, J., Zeng, T., & Qi, G. J. (2023). Ctcnet: A cnn-transformer cooperation network for face image super-resolution. IEEE Transactions on Image Processing, 32, 1978-1991.

[20] Hamid, N. N. A. A., Razali, R. A., & Ibrahim, Z. (2019). Comparing bags of features, conventional convolutional neural network and AlexNet for fruit recognition. Indonesian Journal of Electrical Engineering and Computer Science, 14(1), 333.

[21] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360.

[22] Zhu, L., Li, Z., Li, C., Wu, J., & Yue, J. (2018). High performance vegetable classification from images based on alexnet deep learning model. International Journal of Agricultural and Biological Engineering, 11(4), 217-223.

[23] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable cities and society, 65, 102600.

[24] Alzamily, J. Y. I., Ariffin, S. B., & Abu-Naser, S. S. (2022). Classification of Encrypted Images Using Deep Learning–Resnet50. Journal of Theoretical and Applied Information Technology, 100(21), 6610-6620.

[25] Reynaldi, D., Negara, B. S., Sanjaya, S., & Satria, E. (2021, July). Covid-19 classification for chest x-ray images using deep learning and resnet-101. In 2021 International Congress of Advanced Technology and Engineering (ICOTEN) (pp. 1-4). IEEE.

[26] Sarwinda, D., Paradisa, R. H., Bustamam, A., & Anggia, P. (2021). Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. Procedia Computer Science, 179, 423-431.

[27] Lin, S. L. (2021). Application combining VMD and ResNet101 in intelligent diagnosis of motor faults. Sensors, 21(18), 6065.

[28] Zhang, Q. (2022). A novel ResNet101 model based on dense dilated convolution for image classification. SN Applied Sciences, 4, 1-13.

[29] Rathnayake, N., Rathnayake, U., Dang, T. L., & Hoshino, Y. (2022). An efficient automatic fruit-360 image identification and recognition using a novel modified cascaded-ANFIS algorithm. Sensors, 22(12), 4401.

[30] Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. IEEE Access, 8, 90847-90861.

[31] Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-label confusion.