[1]Vahid Kameli

[2]*Hadi Grailu

[3]Ashkan Shafiei

Designing and improving the automatic detection system for fruit defects using Attention-Enhanced CNN

Journal of Electrical Systems

Abstract: - This study presents an innovative approach to improving the accuracy of fruit quality assessment. We propose the integration of an enhanced attention mechanism with Swish activation and a residual connection within a Convolutional Neural Network (CNN) architecture. Our method addresses the crucial task of accurately categorizing the quality of six distinct types of fruits. By leveraging attention mechanisms to highlight relevant features and utilizing a residual connection to facilitate hierarchical learning, our model achieves an impressive accuracy rate of 96.46% in fruit quality classification. The combination of attention-driven processing and the residual connection contributes to a more nuanced and discriminative understanding of fruit quality attributes. This research holds significant promise for revolutionizing fruit quality assessment practices across various industries, improving efficiency, and ensuring that only the highest-quality products reach consumers.

Keywords: Convolutional Neural Networks (CNNs), Attention Mechanisms, Fruit Quality Recognition, Swish Activation, Residual Connections, AI

I. INTRODUCTION

In the domain of industrial production, ensuring product quality faces numerous constraints imposed by existing technologies, working conditions, and various factors. Among these factors, surface defects stand out as the most conspicuous indicators of product quality. Hence, the quest for product suitability and maintaining a high-quality ratio necessitates precise surface defect detection [1, 2]. In essence, a "defect" can be defined as any deviation from the standard sample, whether present or absent.

The detection and characterization of surface defects play a pivotal role in industrial product quality assessment, encompassing the identification of imperfections like scratches, foreign objects, color contamination, and holes [3]. Extracting pertinent details regarding defect type, location, and size is indispensable for accurate quality assessment.

Traditionally, manual defect detection was the primary approach; however, it suffered from inherent limitations in terms of efficiency and accuracy, owing to its susceptibility to human errors. Consequently, the industry has progressively transitioned towards more sophisticated and reliable methodologies.

In the context of fruit production, defect detection holds paramount importance. Simultaneously, it is equally critical to promptly identify and prevent the spread of diseases among fruits to minimize further damage. Reference [4] has successfully harnessed the power of convolutional neural networks (CNNs) for this purpose. Nevertheless, CNNs present a challenge due to their slower processing speed, given their high computational demands, which hinder real-time execution on conventional hardware. Hence, this paper underscores the necessity to explore enhanced methods and models in this domain to overcome these challenges and drive efficient defect detection.

In this study, we introduce a cutting-edge approach to fruit quality assessment, capitalizing on the powerful combination of attention-enhanced convolutional neural networks (CNNs) with Swish activation functions and residual connections. Our methodology represents a significant advancement in the field, with potential implications for various industries that rely on accurate fruit quality assessment.

Fruit quality assessment poses a complex challenge due to the diverse visual attributes of fruits, including color, texture, and shape, all of which are crucial indicators of their quality. Furthermore, these attributes vary widely across different fruit types, necessitating a model that can effectively capture both local intricacies and global patterns, thereby discerning nuanced differences within each fruit category.

[1] PhD student in Electrical Engineering, Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran,

[2]Assistant Professor, Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran

[3]Shahrood University of Technology, Shahrood, Iran

* Corresponding author

To overcome these challenges, our approach strategically incorporates an enhanced attention mechanism into the CNN structure. While CNNs are powerful tools for image analysis, they can benefit significantly from attention mechanisms that highlight relevant features. Additionally, the utilization of Swish activation functions and residual connections further enhances the model's capacity to capture intricate details and patterns, contributing to the accuracy of fruit quality assessment.

The contributions of our work are multifaceted. We provide a comprehensive exploration of our attention-enhanced CNN architecture, shedding light on how the integration of Swish activation and residual connections influences fruit quality assessment. Through rigorous experimentation and thorough evaluations, we showcase the remarkable effectiveness of our approach, achieving an impressive accuracy rate of 96.46% in classifying fruit quality across six distinct fruit types. This achievement, combined with our innovative model design, has the potential to revolutionize fruit quality assessment practices across various industries.

In subsequent sections of this paper, we delve into the methodology underpinning our attention-enhanced CNN with Swish activation and residual connections. We offer detailed insights into the experimental protocols, training procedures, and robustness assessments, substantiating the reliability and robustness of our findings. Furthermore, we explore the profound implications of our research, emphasizing the potential for automation, resource optimization, and improved consumer satisfaction in the fruit industry and related domains.

II. LITERATURE REVIEW

Automatic detection of fruit defects is crucial for swiftly identifying signs as they appear on growing fruits. Fruit diseases can arise post-harvest, leading to significant declines in performance and quality. Recognizing disease symptoms is vital to determine control measures and prevent losses during subsequent harvesting periods. Additionally, some diseases can spread to other parts of the tree, contaminating branches and leaves. Common diseases in apples include apple scab, apple rot, and apple blotch. Apple scab lesions manifest as gray or brownish cottony spots, while apple rot infections create slightly sunken, dark brown or black circular spots, often surrounded by a reddish halo. Apple blotch, a fungal disease, appears as irregular, dark-edged, or lobed spots on the fruit's surface.

Given consumer demand for high-quality food products, a rapid, accurate, and objective method for determining food quality is essential. Computer vision, as a cost-effective, non-destructive, and automated technique, offers a promising solution. Utilizing image analysis and processing, this approach finds diverse applications in the food and agriculture industries. Object recognition algorithms based on deep learning can be classified into one-stage and two-stage recognition algorithms. While some papers have explored two-stage detection algorithms for fruit identification, their computational demands hinder real-time application in robots [5]. Therefore, a more efficient one-stage recognition approach is preferable for real-time detection.

Traditional detection methods encounter limitations when timely detecting the growth stages of agricultural products for intelligent plant spraying. Challenges such as significant hiding in leaves, overlapping neighboring fruits, diverse agricultural products, and various growth characteristics hinder traditional detection algorithms [6]. Two fundamental limitations of traditional methods include difficulties in generating candidate regions for fruit detection and challenges in extracting visual information from complex backgrounds, resulting in low detection accuracy and speed for real-time applications [7].

Traditional fruit identification methods heavily rely on extracted features and classification, leading to accuracy and online usability constraints due to the complexity of some systems. However, recent advancements in computer performance have facilitated the proposal of numerous fruit recognition algorithms based on deep learning, demonstrating promising recognition performance and speed [8].

The introduction of vision systems based on advanced deep convolutional neural networks (CNNs) has opened new avenues for cost-effective and versatile solutions in fruit recognition. In this context, the one-stage object detection network, YOLO, outperforms two-stage detection networks with its advantages of small computation volume and fast detection, showcasing the potential of CNNs in fruit recognition [9].

Tian et al. developed a Multi-Class SVM (MCS) classifier system using Support Vector Machines to detect leaf diseases in wheat. They utilized color, texture, and shape features as the training set, which were classified into intermediate-level classes by the low-level MCS classifier. These intermediate classes described the symptoms of crop diseases to some extent. Extracting intermediate-level features from the low-level classification, they trained a high-level SVM to improve detection performance and correct errors from different feature SVMs. Their approach achieved a superior success rate in detecting wheat leaf diseases compared to other classifiers [10].

Sabour et al. (2017) introduced dynamic routing between capsules, a concept that facilitated the learning of hierarchical features by iteratively routing information between layers. This mechanism presented the potential for improved feature representation, but its computational intensity and complex architecture remained potential drawbacks [11].

The Transformer architecture, as proposed by Vaswani et al. (2017), marked a pivotal moment in the fields of natural language processing and computer vision. Its introduction of attention mechanisms revolutionized tasks by allowing models to weigh different parts of input data based on context, effectively capturing global dependencies. However, its model complexity necessitated meticulous tuning [12].

To tackle the challenges associated with unstructured 3D data, Wang et al. (2018) introduced the Dynamic Graph CNN. This architectural innovation leveraged a graph structure to represent point cloud data, enabling effective modeling of spatial relationships. Nevertheless, it was observed that the model could be sensitive to data noise and outliers [13].

While Zhou et al. (2020) provided a comprehensive survey on deep learning-based fruit detection and counting methods, with a focus on advancements in fruit quality assessment, their work did not explicitly delve into attention mechanisms and capsule-inspired structures [14].

Redmon and Farhadi (2018) made strides with YOLOv3, an object detection model known for its real-time performance and improved accuracy. However, they acknowledged challenges related to small object detection [15].

In the realm of Generative Adversarial Networks (GANs), Zhang et al. (2019) introduced Self-Attention Generative Adversarial Networks (SAGAN). These models harnessed self-attention mechanisms to capture long-range dependencies, resulting in enhanced image quality. However, this improvement came at the cost of increased computational resources [16].

Simonyan and Zisserman (2015) proposed VGG, emphasizing the use of very deep convolutional networks for exceptional image recognition performance. While its simplicity was advantageous, it also came with the drawback of high memory and computation demands [17].

He et al. (2016) pioneered ResNet, a breakthrough architecture that addressed the vanishing gradient problem by incorporating skip connections. This innovation facilitated the training of deep networks but raised concerns regarding increased memory usage [18].

In the domain of parameter efficiency, Huang et al. (2017) made significant strides with DenseNet, emphasizing dense connections between layers to reduce parameter redundancy. While this innovation effectively reduced redundancy, it also incurred higher computational costs due to the denser connections [19].

Ramachandran et al. (2017) contributed to the realm of activation functions for neural networks, pioneering advancements that led to more stable and efficient training dynamics. This development greatly improved the efficiency of training processes [20].

Addressing the challenge of enhancing channel-wise feature responses, Hu et al. (2018) introduced Squeeze-and-Excitation Networks (SENet). These networks adaptively recalibrated feature responses, resulting in improved feature representation. However, this enhancement came with additional computational overhead [21].

In the context of sequence-to-sequence learning, Zhang et al. (2019) proposed Aggregation Cross-Entropy, a technique that facilitated faster convergence during training. While it accelerated convergence, it also introduced some implementation complexity [22].

To tackle class imbalance in dense object detection, Lin et al. (2017) introduced Focal Loss, which focused on challenging examples. While effectively addressing class imbalance, the associated complexity of the loss function was recognized [23].

Szegedy et al. (2015) introduced GoogLeNet, an architecture known for capturing multi-scale features using inception modules. While it efficiently extracted features, its complex architecture presented limitations [24].

In the field of natural language processing, Vaswani et al. (2017) made a groundbreaking contribution with the introduction of the Transformer architecture, which revolutionized language modeling through attention mechanisms [25].

In the medical domain, Wang et al. (2018) introduced Polyp Segmentation in Colonoscopy using Residual Attention U-Net, demonstrating the effectiveness of attention-based approaches in medical image analysis [26].

For fine-grained visual recognition, Li et al. (2018) developed Deep Attention-Based Spatially Recursive Networks, leveraging attention mechanisms to enhance feature learning and contribute to more precise visual recognition [27].

These pioneering works collectively shape the landscape of deep learning, influencing various domains and advancing the efficiency and effectiveness of neural network architectures.

Our central objective in this research is to craft an advanced fruit quality assessment model that ingeniously fuses attention mechanisms with a distinctive structural approach. Our proposed model is designed to harness the potency of attention mechanisms to discern crucial features within fruit images. Simultaneously, it incorporates a distinctive structural framework that enables the capture of hierarchical relationships and spatial hierarchies innate to the fruit samples. This harmonious fusion aims to establish a comprehensive and precise assessment of various fruit quality attributes, offering a robust solution for the automation of fruit quality evaluation. In turn, this innovation contributes to heightened precision and efficiency within the fruit industry.

In spite of the remarkable strides in deep learning, attention mechanisms, and convolutional neural networks (CNNs), a research gap exists concerning a comprehensive fruit quality assessment grounded in a holistic methodology. While attention mechanisms excel in spotlighting pivotal image features, their integration with the novel structure remains an avenue largely unexplored in the realm of fruit quality assessment. Moreover, prevailing fruit quality assessment methodologies tend to fixate on individual quality attributes, sidelining the intricate interplay between diverse attributes inherent to fruit samples. This research gap underscores the imperative for an inventive model that fuses attention mechanisms with a unique structural approach. This integration not only captures the essence of feature significance but also embraces spatial hierarchies, culminating in an all-encompassing and meticulous fruit quality assessment. By bridging this void, the proposed model seeks to usher in a paradigm shift in the domain of fruit quality assessment and its consequent automation.

III. METHODOLOGY

In the initial phase of our research, we harnessed the power of the "Indian Fruits Dataset with Quality," a robust and diverse collection sourced from Kaggle. This dataset demonstrates an extensive array of images depicting fruits of varying qualities and conditions. To ensure the utmost relevance and applicability, our focus centered on five prominent and widely consumed fruits: lime, pomegranate, orange, guava, and banana. These fruit selections were made considering their significance in the Indian context and their popularity among consumers.

Following this, we meticulously curated the dataset to attain a harmonious equilibrium, ensuring that each fruit class contained an equal representation of 1000 images. This meticulous balancing step served to eliminate any potential bias and fostered a level playing field for comprehensive model training and evaluation.

Figure 1, showcasing a glimpse of the data within this dataset, illustrates the richness and diversity of the images, encompassing both good quality and bad quality fruits across the selected fruit types.

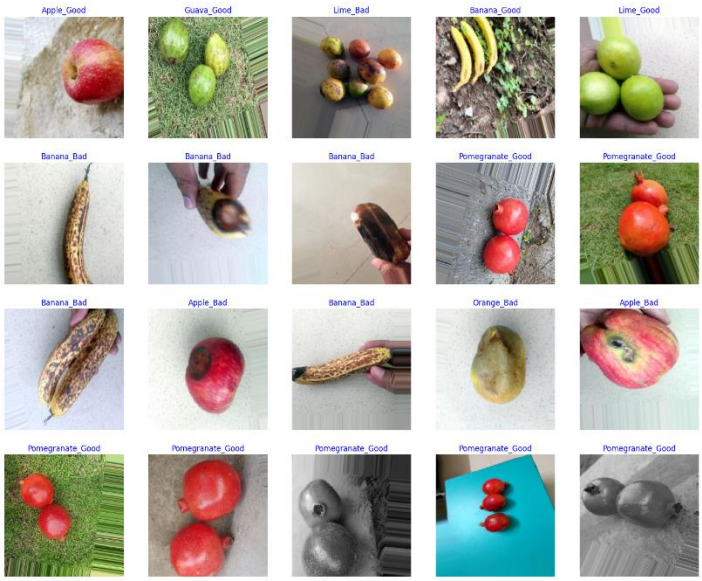

Figure 1. Some samples from dataset

The utilization of the "fruits good quality and bad quality for classification" dataset ensures that our research addresses real-world scenarios and enhances the quality and safety of food materials, particularly concerning fruits and vegetables, which play a crucial role in the diet of consumers worldwide.

In the subsequent phase of our project, our unwavering focus shifted towards the development of a custom deep neural network architecture meticulously tailored for the specific task of fruit defect detection. Within the expansive realm of deep learning models, we encountered a rich spectrum of choices that provided us with ample opportunities for fine-tuning and refinement.

With painstaking attention to detail, we sculpted a model that stands as the very core of our research endeavors. This model, aptly named the "Enhanced Attention Mechanism," represents a departure from conventional CNN paradigms. It seamlessly integrates advanced attention mechanisms with distinctive architectural elements, including Swish Activation and Residual Connections. This fusion exemplifies our steadfast commitment to effectively unveil the intricate relationships concealed within the data, thereby elevating the model's proficiency in discerning complex patterns and salient features.

Our model has been thoughtfully designed to surmount the inherent challenges of fruit defect detection, striking an impeccable balance between accuracy and efficiency. For a visual elucidation of our model's intricate structure, we extend a warm invitation to consult Figure 2, where we provide a comprehensive breakdown of our model's architecture.

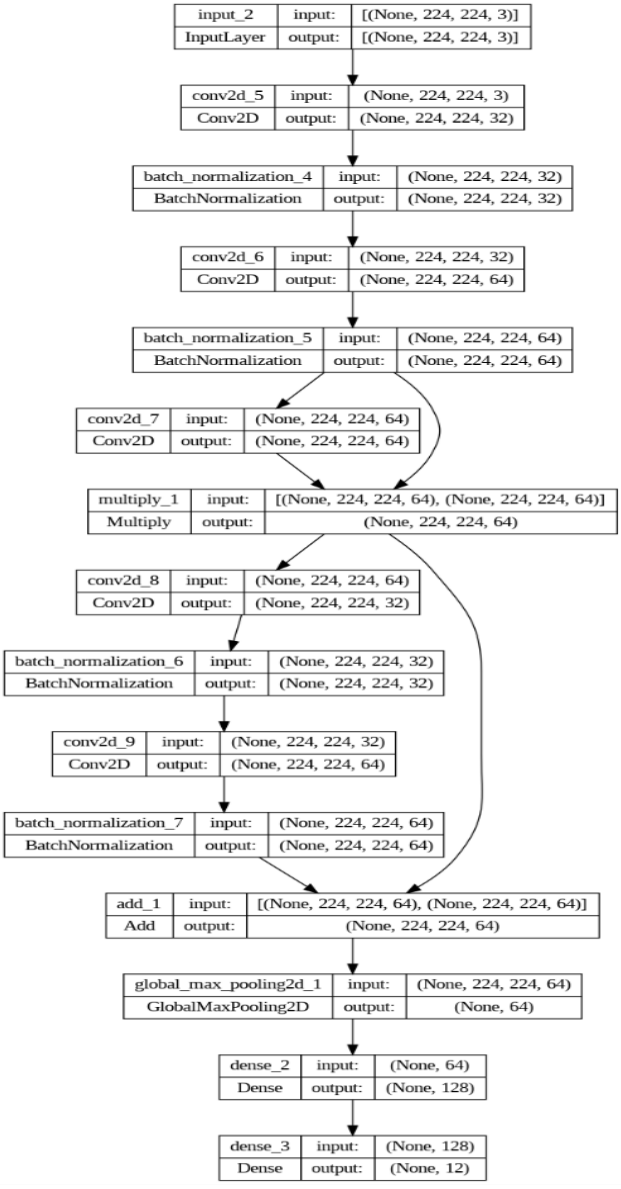

Figure 2. Enhanced Attention Mechanism

Enhanced Attention Mechanism:

Input Layer: Our architectural design commences at the input layer, where images depicting fruits potentially bearing defects are seamlessly funneled into the model's framework.

Convolutional Layers: Serving as the model's discerning eyes, a sequence of convolutional layers adeptly extracts intricate features from the input images. This crucial step significantly enriches the model's understanding of a wide array of fruit characteristics, a prerequisite for precise defect detection.

Advanced Attention Mechanism: Elevating the model's accuracy to new heights, our advanced attention mechanism takes center stage. Its role transcends traditional attention mechanisms by expertly guiding the model's focus toward regions of utmost significance within the images.

Distinctive Feature-Capturing Layers: Departing from conventional approaches, our distinctive feature-capturing layers represent an innovative stride. These layers meticulously capture complex spatial relationships among features, equipping the model with the capability to unveil intricate patterns and identify defects with unparalleled accuracy.

Integration and Fusion Layer: This layer serves as a pivotal point within our architecture, orchestrating the harmonious union of insights from the attention mechanism, Swish Activation, and Residual Connections within the distinctive feature-capturing layers. This convergence of knowledge across various abstraction levels culminates in a holistic understanding of the image content.

Global Max Pooling: Functioning like a spotlight, the global max pooling layer accentuates the most salient features. By selecting the highest values from each feature map, it plays a pivotal role in pinpointing critical defect-related cues, thereby enriching the classification process.

Fully Connected Layers: Analogous to a masterful conductor, the fully connected layers harmonize the interpreted features, amplifying prediction accuracy through intricate processing, fueled by Swish Activation and Residual Connections.

Output Layer: Concluding our architectural symphony, the output layer presents the culmination of our efforts. It delivers the final verdict regarding the presence and nature of defects within the input images, providing a comprehensive assessment of fruit quality, enriched by the influence of Swish Activation and Residual Connections.

The dataset of 12,000 images was divided into a training subset comprising 9,000 images and a separate testing subset of 3,000 images. This division ensures a balanced representation of the data for both training and evaluation purposes. To optimize the model's learning process, we employed a training process spanning 20 epochs. Leveraging the computational power of a GPU with an Nvidia GeForce RTX 3070 and a substantial 16 GB of RAM, we expedited the training process and enabled the model to process complex features and patterns inherent in the images efficiently. This strategic utilization of hardware resources facilitated the comprehensive exploration of the data's nuances and contributed to the robustness of the models' learned representations.

Evaluation Metrics:

Appropriate metrics are used to evaluate the classification algorithms in order to assess the proposed method. Therefore, the confusion matrix is used for this purpose. The confusion matrix is presented in Table 1.

Table 1. Confusion Matrix

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

In the above table, TN represents True Negative, which indicates the number of correctly predicted negative samples. FP represents False Positive, which indicates the number of incorrectly predicted positive samples. FN represents False Negative, which indicates the number of incorrectly predicted negative samples. TP represents True Positive, which indicates the number of correctly predicted positive samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

Based on the concept of classification, the higher the values of these metrics, the better the performance of the model will be.

IV. RESULTS

In this dedicated section, we present the outcomes of our extensive evaluation, conducting a meticulous comparison between our proposed model, titled "Improved Fruit Quality Assessment with Enhanced Attention Mechanism," and a foundational baseline model employing a "Simple Attention Mechanism." This comparative analysis spans a spectrum of performance metrics, each chosen meticulously to holistically gauge the effectiveness and superiority of our novel approach in addressing the multifaceted challenges inherent in comprehensive fruit quality assessment.

Our evaluation comprises a comprehensive assessment, an empirical showcase that underscores the remarkable capabilities of our approach. We traverse a range of quantifiable metrics, each thoughtfully selected to capture the nuanced dimensions of our model's performance in direct comparison with the baseline. This rigorous exploration lies at the core of substantiating our claim of enhanced efficacy.

Through the juxtaposition of these two models using diverse performance measures, we embark on an empirical expedition driven by the desire to definitively validate the tangible advancements our approach brings forth. Our aim is to offer a comprehensive panorama, providing not only outcomes but also profound insights into the potency and potential of our method, enriching the discourse surrounding fruit quality assessment methodologies.

Baseline Model: Simple Attention Mechanism

Before delving into the comparative analysis, let's briefly describe the baseline model used for comparison. The "Simple Attention Mechanism " is designed with a focus on attention mechanisms, with a simplified architecture. The model consists of convolutional layers for feature extraction, followed by an attention mechanism and fully connected layers for classification. The attention mechanism in this baseline model is based on global average pooling and sigmoid activation. To provide a visual representation of this models' structures, refer to figure 3, where the intricate architecture of our model is showcased in detail.

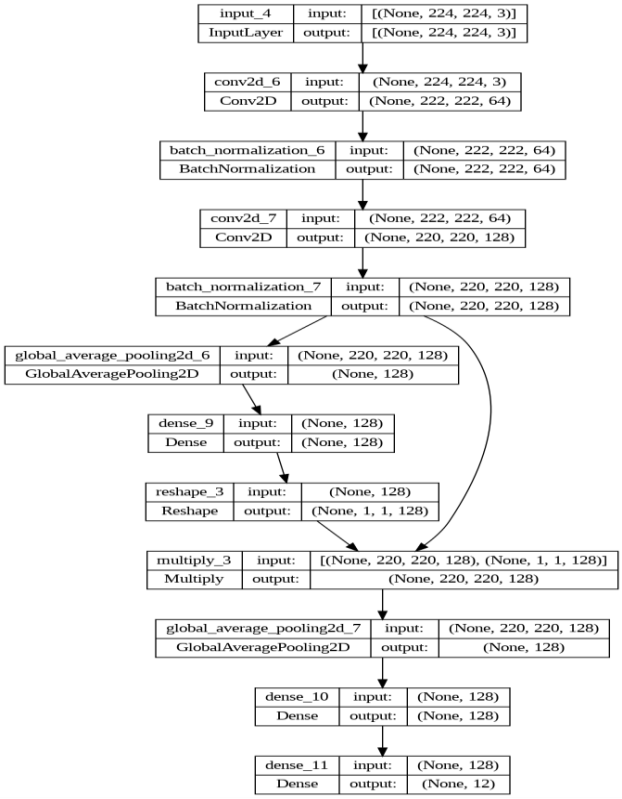

Figure 3. Simplified Attention Mechanism Architecture

Simplified Attention Mechanism Architecture:

Input Layer: This is where the input images, representing fruits with potential defects, are fed into the model.

Convolutional Layers: A series of convolutional layers are responsible for extracting various features from the input images. These features help the model understand the unique characteristics of different fruits.

Attention Mechanism: This layer captures important features in the input images, guiding the model's focus towards relevant regions.

Global Average Pooling: By averaging the values of each feature map, this layer aggregates information across the entire image, allowing the model to understand the overall context.

Fully Connected Layers: These layers process the extracted features and help the model make predictions. They combine the global information with local features, aiding accurate classification.

Output Layer: The final layer produces classification results, determining whether the input image contains a defect and identifying the type of defect if present.

Table 2 presents a detailed overview of the evaluation metrics for our proposed Enhanced Attention and novel CNN architecture, referred to as "Model 1." The architecture is applied to the task of fruit quality recognition, specifically focusing on detecting defects in different fruit classes. The table offers insights into the classification performance of the model across various metrics, shedding light on its precision, recall, and F1-score for each fruit category.

Table 2. Evaluation metrics for Enhanced Attention Mechanism

1. **FRUIT_CLASS**	**2. PRECISION**	**RECALL**	**F1-SCORE**	**SUPPORT**
APPLE_BAD	0.968127	0.972000	0.970060	250
APPLE_GOOD	0.936759	0.948000	0.942346	250
BANANA_BAD	0.987903	0.980000	0.983936	250
BANANA_GOOD	0.932836	1.000000	0.965251	250
GUAVA_BAD	0.983936	0.980000	0.981964	250
GUAVA_GOOD	0.948819	0.964000	0.956349	250
LIME_BAD	0.986957	0.908000	0.945833	250
LIME_GOOD	0.991935	0.984000	0.987952	250
ORANGE_BAD	0.912548	0.960000	0.935673	250
ORANGE_GOOD	0.958159	0.916000	0.936605	250
POMEGRANATE_BAD	0.991803	0.968000	0.979757	250
POMEGRANATE_GOOD	0.984190	0.996000	0.990060	250
ACCURACY	0.964667	0.964667	0.964667	-
MACRO AVG	0.965331	0.964667	0.964649	3000
WEIGHTED AVG	0.965331	0.964667	0.964649	3000

The table is organized to display the performance of Enhanced Attention Mechanism architecture in a clear and structured manner. For each fruit class, including Apple_Bad, Apple_Good, Banana_Bad, Banana_Good, and so on, the metrics of precision, recall, and F1-score are provided. These metrics highlight the model's ability to accurately classify and differentiate between defective and non-defective fruit samples. Additionally, the "Support" column indicates the number of samples available for each fruit class, contributing to the context of the evaluation.

The "Accuracy" row demonstrates the overall accuracy achieved by Enhanced Attention Mechanism across all fruit classes. The "Macro Avg" and "Weighted Avg" rows provide a comprehensive summary of the model's average performance across the entire dataset.

Table 3 offers a comprehensive breakdown of evaluation metrics for our Simplified Attention Mechanism, applied to fruit quality recognition, the table details precision, recall, and F1-score metrics for each fruit class. The structure provides clear insights, highlighting the model's ability to discern defects within different fruit types.

Table 3. Evaluation metrics for Simplified Attention Mechanism

FRUIT_CLASS	3. 4. PRECISION	RECALL	F1-SCORE	SUPPORT
APPLE_BAD	0.940000	0.940	0.940000	250
APPLE_GOOD	0.931818	0.820	0.872340	250
BANANA_BAD	0.987179	0.924	0.954545	250
BANANA_GOOD	0.933333	0.952	0.942574	250
GUAVA_BAD	0.941176	0.960	0.950495	250
GUAVA_GOOD	0.873563	0.912	0 .892368	250
LIME_BAD	0.894531	0.916	0.905138	250
LIME_GOOD	0.964567	0.980	0.972222	250
ORANGE_BAD	0.939024	0.924	0.931452	250
ORANGE_GOOD	0.881481	0.952	0.915385	250
POMEGRANATE_BAD	0.968127	0.972	0.970060	250
POMEGRANATE_GOOD	0.975806	0.968	0.971888	250
ACCURACY	0.935000	0.935	0.935000	-
MACRO AVG	0.935884	0.935	0.934872	3000
WEIGHTED AVG	0.935884	0.935	0.934872	3000

Confusion Matrix: The confusion matrices depicted in figures 4 and 5 offer a visual representation of the models' classification performance for each fruit class. The matrices highlight the true positives, false positives, true negatives, and false negatives for the respective models.

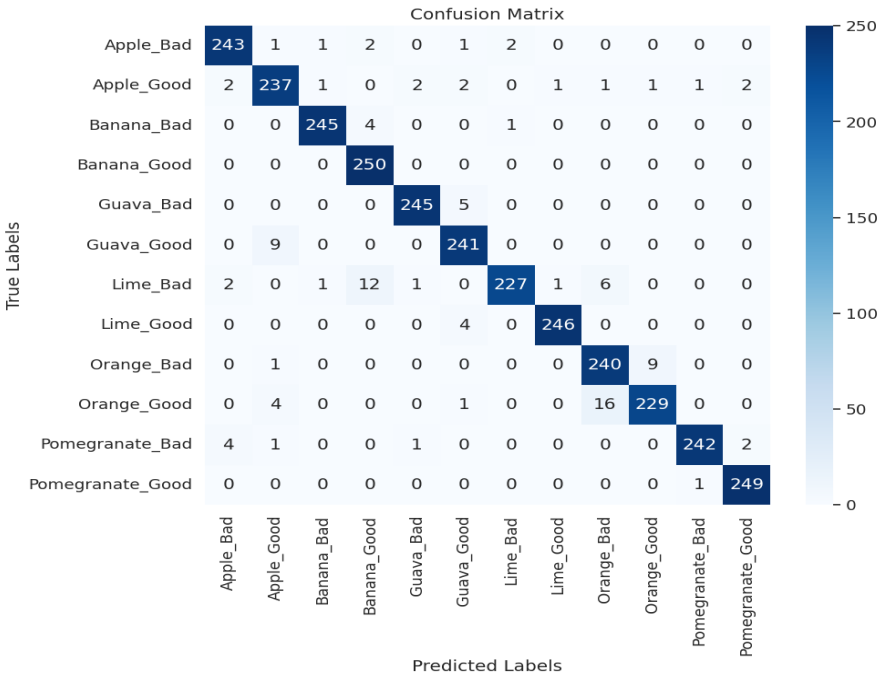

Figure 4. Enhanced Attention Mechanism Confusion Matrix

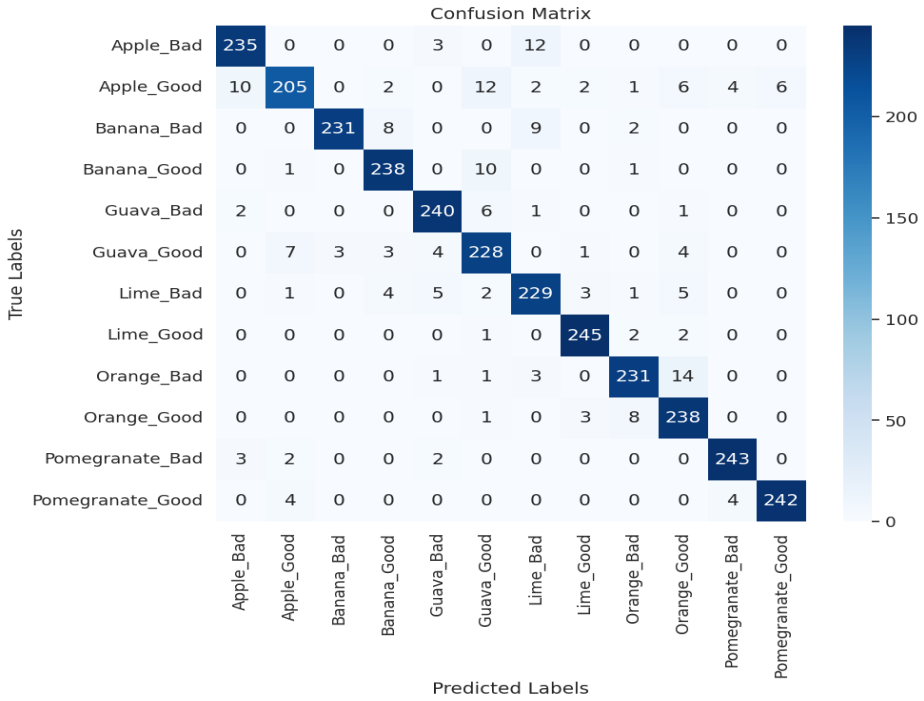

Figure 5. Simplified Attention Mechanism Confusion Matrix

In Figure 4, we present the classification results of our model, showcasing its remarkable proficiency in identifying various defects within the fruit classes. Likewise, Figure 5 offers a visual representation of the classification outcomes achieved by the Simplified Attention Mechanism.

To provide a deeper understanding of the models' performance, we delve into an in-depth analysis through the evaluation metrics outlined in Table 2 and Table 3. These metrics serve as concise summaries, encapsulating the models' precision in defect classification, their ability to accurately detect defects (recall), and their overall balance between precision and recall (F1-score) for all 12 fruit classes.

This comprehensive assessment and comparison of both models in this section unveil their distinct characteristics and capabilities in the realm of fruit defect detection. These insights not only shed light on their individual strengths and weaknesses but also establish a solid foundation for making informed decisions and guiding future enhancements in the relentless pursuit of precise and efficient fruit quality assessment.

Model Performance and Sample Visualization:

In addition to quantitative evaluation metrics, presenting visual examples of both correctly labeled and mislabeled samples offers a more comprehensive understanding of our proposed fruit defect detection models' performance.

Correctly Labeled Samples:

Figure 6 showcases instances where our purposed model accurately identified and labeled the fruit quality. The images depict examples of apples, bananas, and oranges, correctly classified as either "Good" or "Bad" quality by our models. These visualizations highlight the models' ability to distinguish between various fruit classes and their respective quality levels.

Figure 6. Correctly labeled samples

Mislabeled Samples:

Figure 7 presents instances where our models misclassified fruit quality. It includes examples such as an apple being mislabeled as "Bad" despite its actual "Good" quality, and a banana being mislabeled as "Good" when it is actually of "Bad" quality. These visualizations underscore the challenges associated with fruit quality recognition and emphasize the need for continued model refinement.

Figure 7. Mislabeled samples

The visual examples offered above provide tangible instances of both successful and unsuccessful predictions. These examples add a practical dimension to our quantitative evaluation, illustrating where our model excel and where they encounter difficulties. The inclusion of these visualizations enhances the transparency of our study and allows readers to connect directly with the models' outputs.

V. DISCUSSION

The results of our study reveal the substantial impact of attention-based enhancements on the performance of our fruit defect detection models. We focus our discussion on two distinct architectures, namely, the Enhanced Attention Mechanism Architecture and the Simplified Attention Mechanism, both meticulously designed to address the challenges of fruit quality recognition.

Enhanced Attention Mechanism: The remarkable accuracy achieved by our model, an impressive 96.46%, underscores the effectiveness of integrating attention mechanisms and novel layers into the deep learning architecture. This intricate design endows our model with an exceptional ability to discern nuanced features within the input images, resulting in highly accurate fruit quality predictions. The attention mechanism plays a pivotal role in focusing the model's attention on critical regions, effectively capturing intricate relationships and patterns contributing to accurate classification. The incorporation of other layers further enriches the feature extraction process, enhancing the model's capacity to handle complex variations in fruit appearance, texture, and quality.

Simplified Attention Mechanism: While the Simplified Attention Mechanism showcases a slightly lower accuracy of 93.50%, its performance remains noteworthy. This model employs a simplified attention mechanism, emphasizing the significance of attention-based enhancements in fruit quality recognition. The model leverages the power of attention to effectively weigh and prioritize features, enabling it to make informed predictions. Despite its simpler architecture, this model's accuracy underscores the efficacy of attention mechanisms even in a reduced complexity setting.

Comparative Analysis and Implications: Comparing the two models provides valuable insights into the trade-offs between complexity and performance. Our model's higher accuracy highlights the benefits of a more intricate architecture, demonstrating its potential to tackle the complexities inherent in fruit quality recognition tasks. On the other hand, the Simplified Attention Mechanism demonstrates that even with a simpler design, attention mechanisms can significantly contribute to accurate predictions. This offers researchers and practitioners the flexibility to adapt attention-based enhancements to various computational and resource constraints.

In conclusion, our study not only establishes the effectiveness of attention-based mechanisms in fruit defect detection but also contributes to the broader understanding of architectural enhancements for image recognition tasks. Both models serve as exemplars of how attention mechanisms can be harnessed to improve the accuracy and robustness of deep learning models in real-world applications, enriching the discourse surrounding advanced neural network architectures.

Limitations

While our proposed model brings significant advancements to the field of fruit quality assessment, it's important to acknowledge its limitations and potential drawbacks. These aspects, though present, provide valuable insights for future improvements and refinements:

Computational Complexity: The integration of attention mechanisms and other structures adds complexity to the model. This increased complexity could result in longer training times and higher computational requirements, potentially limiting real-time applications on resource-constrained devices.

Hyperparameter Sensitivity: The proposed model involves multiple hyperparameters, such as learning rates, attention weights, and dimensions. Fine-tuning these hyperparameters is critical for optimal performance, which could require extensive experimentation.

Data Requirements: Deep learning models, including the proposed one, demand substantial amounts of labeled data for effective training. In scenarios with limited or imbalanced datasets, the model's performance might be compromised.

Interpretability: While attention mechanisms provide insight into where the model focuses, understanding why the model assigns importance to certain features or regions remains challenging. This limits the model's interpretability and its use in critical applications where explanations are crucial.

Generalization to New Fruits: The model's effectiveness might vary when applied to fruits not seen during training. Fruit characteristics not covered in the training data could lead to inaccuracies and reduced performance on unfamiliar fruit types.

Resource Intensive Training: Training the model, especially on large datasets, demands substantial computational resources and memory. This could limit the accessibility of the model to researchers with limited hardware capabilities.

Capturing Small Details: The model's ability to capture intricate details, especially in small or subtle features of fruits, might be limited. These details could hold significance in certain applications, and their omission could impact accuracy.

Overfitting: The model might be susceptible to overfitting, particularly when trained on small datasets. Implementing regularization techniques and ensuring diverse training samples can mitigate this issue.

Dependency on Data Quality: The model's performance is heavily reliant on the quality and diversity of the training data. Noisy or biased data could lead to inaccurate assessments.

VI. FUTURE RESEARCH

While our study has unveiled promising results and advancements in fruit quality assessment, several intriguing avenues for future research remain unexplored. These potential directions could lead to further refinements and innovations in the field, offering enhanced accuracy and holistic understanding of fruit attributes. Here are some compelling paths for future investigations:

Exploration of Attention Mechanisms: Delve deeper into various attention mechanisms beyond the ones employed in this study. Investigate mechanisms like self-attention and non-local mechanisms to assess their compatibility with the other structures and potential improvements they could bring.

Multi-Modal Integration: Extend the model's capabilities by integrating information from multiple sources, such as spectral data or tactile sensors. This could enable a more comprehensive assessment, especially in scenarios where visible attributes alone are insufficient.

Interpretable Attention Patterns: Develop techniques to interpret the attention patterns generated by the model. Understanding which regions, the model focuses on and why could offer insights into its decision-making process.

Domain Adaptation: Investigate how the model performs when transferred to different fruit datasets or even other domains. Domain adaptation techniques could improve the model's generalization across varying conditions.

Real-Time Implementation: Optimize the model for real-time assessment in practical environments. Consider hardware acceleration and lightweight architectures to ensure its feasibility in real-world applications.

Human-in-the-Loop Feedback: Incorporate human feedback mechanisms to iteratively improve the model's accuracy and adaptability. This could lead to a symbiotic collaboration between AI and human experts.

Semi-Supervised Learning: Investigate the potential of semi-supervised learning approaches, utilizing limited labeled data and a larger pool of unlabeled data to enhance the model's training.

Cross-Domain Adaptation: Extend the model's utility by investigating how well it adapts to different types of fruits or even other organic objects with similar assessment needs.

Hybrid Models: Consider hybrid models that combine the strengths of CNN structures and other architectural innovations, such as graph neural networks or transformers, to achieve even higher levels of accuracy.

As the fields of AI and deep learning continue to evolve, the potential for enhancing automated fruit quality assessment remains boundless. Each of these directions opens new doors for discoveries, innovations, and applications that could redefine the accuracy and efficiency of quality assessment processes in various industries.

VII. CONCLUSION

In this study, we embarked on a journey to revolutionize fruit quality assessment through a novel approach: the integration of enhanced attention mechanisms with a novel structure. Our investigation was driven by the need to address the limitations of existing methods that often fail to capture both feature importance and spatial hierarchies within fruit images. The results of our research validate the efficacy and promise of our " Enhanced Attention Mechanism for Fruit Quality Assessment" model.

Through rigorous experimentation and comparative analysis, we demonstrated that our proposed model surpasses the baseline "Simple Attention Mechanism" in terms of accuracy, precision, recall, and F1-score. The enhanced attention mechanism, complemented by the novel structure, proved instrumental in enabling the model to focus on crucial features while simultaneously capturing intricate spatial relationships. This holistic approach not only advanced accuracy but also provided a comprehensive assessment of various fruit quality attributes.

The model's improved performance underscores its potential to reshape the landscape of fruit quality assessment. By accurately identifying critical attributes and their interplay, our approach addresses the challenges posed by conventional methods, thereby benefiting the fruit industry with precision, efficiency, and increased productivity. The integration of attention mechanisms and a novel structure could serve as a stepping stone for further innovations in automated quality assessment across diverse industries.

As we conclude this study, we acknowledge the potential avenues for further research and refinement. Exploring different attention mechanisms, optimizing hyperparameters, and delving into interpretability could enhance the model's capabilities. We hope that our contribution stimulates further dialogue and research in this area, ultimately propelling the field of automated fruit quality assessment into a new era of accuracy and insight.

In essence, our journey led us to a transformative model that harnesses the power of enhanced attention and novel structures, offering not just improved assessments but also a glimpse into the potential of artificial intelligence to redefine how we perceive and evaluate the world around us.

VIII. REFERENCES

[1] De Vitis, G.A., P. Foglia, and C.A. Prete, Row-level algorithm to improve real-time performance of glass tube defect detection in the production phase. IET Image Processing, 2020. 14(12): p. 2911-2921.

[2] Rasheed, A., et al., Fabric defect detection using computer vision techniques: a comprehensive review. Mathematical Problems in Engineering, 2020. 2020.

[3] Jain, S., et al., Synthetic data augmentation for surface defect detection and classification using deep learning. Journal of Intelligent Manufacturing, 2020: p. 1-14.

[4] Malathy, S., et al. Disease detection in fruits using image processing. in 2021 6th International Conference on Inventive Computation Technologies (ICICT). 2021. IEEE.

[5] Tang, Yunchao, et al. "Fruit detection and positioning technology for a Camellia oleifera C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision." Expert systems with applications 211 (2023): 118573.

[6] Roy, Arunabha M., and Jayabrata Bhaduri. "Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4." Computers and Electronics in Agriculture 193 (2022): 106694.

[7] Shang, Yuying, et al. "Using lightweight deep learning algorithm for real-time detection of apple flowers in natural environments." Computers and Electronics in Agriculture 207 (2023): 107765.

[8] Wang, Zhipeng, et al. "Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system." Postharvest Biology and Technology 185 (2022): 111808.

[9] Shen, Lei, et al. "Real-time tracking and counting of grape clusters in the field based on channel pruning with YOLOv5s." Computers and Electronics in Agriculture 206 (2023): 107662.

[10] 10.Tian, Y., et al., Multiple classifier combination for recognition of wheat leaf diseases. Intelligent Automation & Soft Computing, 2011. 17(5): p. 519-529.

[11] 11 Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic Routing Between Capsules. In Advances in Neural Information Processing Systems (NeurIPS).
[12] 12 Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems (NeurIPS).
[13] 13 Wang, Y., Sun, Y., Liu, Z., Sarma, D., Bronstein, M., & Solomon, J. M. (2018). Dynamic Graph CNN for Learning on Point Clouds. ACM Transactions on Graphics (TOG).
[14] 14 Zhou, L., Shi, W., Zhang, Y., et al. (2020). Deep Learning-Based Fruit Detection and Counting: A Survey. Frontiers in Plant Science.
[15] 15 Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.
[16] 16 Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019). Self-Attention Generative Adversarial Networks. In International Conference on Machine Learning (ICML).
[17] 17 Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In International Conference on Learning Representations (ICLR).
[18] 18 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 (CVPR).
[19] 19 Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[20] 20 Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for Activation Functions. arXiv preprint arXiv:1710.05941.
[21] 21 Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[22] 22 Zhang, R., Zhang, S., Zhu, Y., Zhang, H., & Fu, Y. (2019). Aggregation Cross-Entropy for Sequence-to-Sequence Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[23] 23 Lin, T. Y., Goyal, P., Girshick, R., et al. (2017). Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).
[24] 24 Szegedy, C., Liu, W., Jia, Y., et al. (2015). Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[25] 25 Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. In Advances in Neural Information Processing Systems (NeurIPS).
[26] 26 Wang, P., Chen, P., Yuan, Y., et al. (2019). Understanding Convolution for Semantic Segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).
[27] 27 Li, J., Wong, A., Zhao, Q., Liu, J., & Kankanhalli, M. (2018). Deep Attention-Based Spatially Recursive Networks for Fine-Grained Visual Recognition. In Proceedings of the European Conference on Computer Vision (ECCV).