Evaluating Accuracy and Efficiency of Fruit Image Generation Using Generative AI Diffusion Models for Agricultural Robotics

Kun Zhao, Minh Nguyen, Weiqi Yan
Department of Computer Science & Software Engineering
Auckland University of Technology
Auckland, New Zealand
kun.zhao@autuni.ac.nz, minh.nguyen@aut.ac.nz, weiqi.yan@aut.ac.nz

2024 39th International Conference on Image and Vision Computing New Zealand (IVCNZ) | 979-8-3315-1877-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/IVCNZ64857.2024.10794462

***Abstract*—The automation of fruit picking using robotic systems is rapidly advancing within precision agriculture, relying heavily on the accurate detection and segmentation of fruit objects through computer vision. This study explores the use of generative artificial intelligence (AI) to create extensive fruit image datasets, concentrating on cherries, bananas, oranges, apples, and pineapples. Five generative models—Stable Diffusion v1.4, Stable Diffusion v1.5, Stable Diffusion v2, Stable Diffusion XL Refiner 1.0, and SDXL Turbo—were employed to generate 1000 images per fruit type, each set against a white background. Thresholding techniques were used to automatically extract fruit boundaries, which were then utilized to train segmentation models. The performance of these models was evaluated based on their accuracy and processing speed. SDXL Turbo consistently delivered the highest accuracy across all fruit types, though it required more processing time per image. Stable Diffusion XL Refiner 1.0 also exhibited strong accuracy but balanced performance differently. In contrast, Stable Diffusion v2 showed significant shortcomings, particularly in producing accurate images for cherries and oranges. This comparative analysis highlights the potential of advanced generative models in enhancing synthetic dataset creation for fruit object detection and segmentation in agricultural robotics. Future research will focus on refining these models to improve accuracy, broaden their applicability to a wider variety of fruits and environments, and optimize the trade-off between image quality and generation speed to meet the practical needs of agricultural automation.**

***Index Terms*—Precision agriculture, fruit detection, object segmentation, generative AI, synthetic dataset, Stable Diffusion, agricultural robotics, image processing, computer vision, automation**

I. INTRODUCTION

A. Transforming Agriculture: The Role of AI and Robotics in Fruit Harvesting

The advent of AI and robotics in agriculture is set to revolutionize fruit harvesting, particularly in automating the generation of fruit images for training machine learning models. The growing demand for efficient and sustainable agricultural practices is driven by labor shortages, high costs, and the physically demanding nature of agricultural work. As the global population continues to rise, there is an urgent need to enhance agricultural productivity and efficiency through automation. This paper presents a comparative study of various AI models used to generate fruit images, specifically cherries. The comparison focuses on quality, generation speed, and adaptability to different fruit types.

Fig. 1: Robots may soon have a place in fruit orchards [1].

Automating fruit harvesting has been a topic of extensive research. As shown in Figure 1, robots are being integrated into fruit orchards to automate various tasks [1]. Traditional methods of fruit harvesting involve manual labor, which is not only time-consuming but also prone to inconsistencies and damage to the produce. According to Li et al. , the development of automatic fruit harvesting systems has the potential to address these challenges by providing a reliable and efficient alternative to manual labor [2]. These systems leverage advanced technologies such as computer vision and robotics to accurately identify and harvest fruits with minimal damage.

In recent years, significant advancements have been made in the field of robotic fruit harvesting. Liu et al. highlight the importance of developing damage-free robotic harvesting systems to ensure the quality of the produce [3]. The integration of robotic arms with vision systems, as discussed by Reddy and Nagaraja, has enabled precise and efficient fruit picking, reducing the reliance on manual labor and increasing overall productivity [4].

In fruit harvesting, there is a definite trend towards increased automation, which is evident now and will continue into the future [5], [6]. As agricultural automation becomes increasingly important to address labor shortages and increase efficiency, the need for large, high-quality datasets to train AI systems grows. Generating these datasets manually is time-consuming and costly. Therefore, leveraging AI to generate realistic images of fruits such as cherries can significantly streamline the development of these systems. The generated images can be used to train computer vision models for tasks such as fruit detection, classification, and quality assessment. By improving the accuracy and efficiency of these models, the agricultural industry can benefit from reduced labor costs and minimized fruit damage.

B. AI-Driven Object Detection in Agriculture: Challenges and Innovations

AI's ability to detect objects, such as fruits, has significantly advanced through the use of Convolutional Neural Networks (CNNs) [7] and transformer-based [8]. These technologies have shown remarkable success in accurately identifying and classifying various objects in agricultural settings. Models like Faster R-CNN [9], YOLO (You Only Look Once) [10], and the Vision Transformer (ViT) [8] have set new benchmarks in object detection tasks.

However, the cornerstone of training these sophisticated models is the availability of annotated datasets. Manual annotation, where each fruit in an image is labeled by human annotators, is a labor-intensive and time-consuming process. For instance, to develop a model that can detect and classify different types of fruits, extensive datasets containing thousands or even millions of labeled images are required. This process not only demands significant labor but also incurs substantial costs.

The agricultural industry, characterized by a vast diversity of fruits, faces unique challenges in this context. Each new fruit variety or harvesting condition necessitates the creation of a new dataset, which must be meticulously annotated to ensure the accuracy of the resulting models. This dependency on manual annotation creates a bottleneck, hindering the rapid deployment of AI solutions.

To address this challenge, generative AI techniques are emerging as a promising solution. By leveraging models such as Generative Adversarial Networks (GANs) [11] and Variational Autoencoders (VAEs) [12], it is possible to automatically generate annotated datasets. These synthetic datasets can be used to train object detection models, reducing the need for manual labor and significantly accelerating the development process.

For example, a GAN-based system can be designed to create realistic images of various fruits, complete with annotations. These generated datasets can then be used to train CNNs or transformer models for fruit detection. This approach not only saves time and effort but also ensures that the models are trained on diverse and comprehensive datasets, enhancing their accuracy and robustness.

The potential of generative AI to automate the dataset creation process represents a significant leap forward in agricultural automation. By streamlining the development of object detection models, the agricultural industry can more swiftly adopt AI-driven solutions, leading to increased efficiency and productivity.

C. Primary Goal of This Paper

The primary objective of this study is to rigorously compare various AI models to identify the most efficient in generating realistic and usable fruit images for training computer vision algorithms. With the rapid advancements in image generation technologies, numerous models have emerged, each offering unique strengths and weaknesses. Some models excel in producing high-quality images, while others are optimized for speed or adaptability to diverse fruit types.

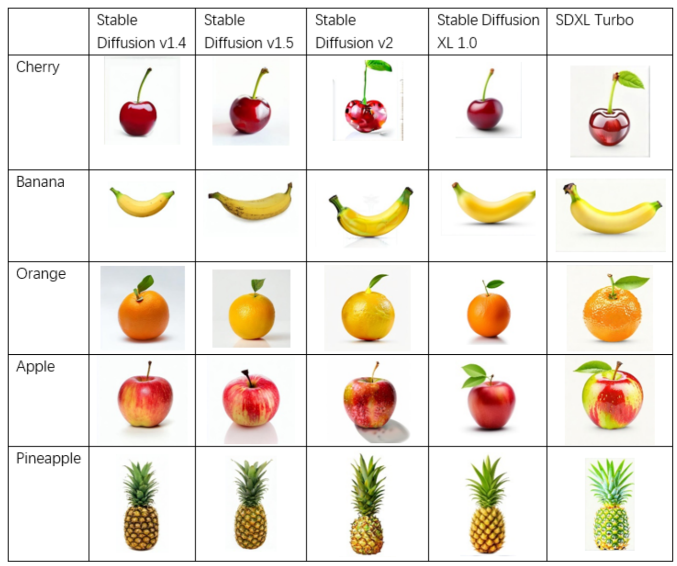


Fig. 2: Sample outputs of the generative models for various fruit types.

In this paper, we conduct a thorough evaluation of these models based on three critical criteria:

1) **Image Quality**: The realism and visual accuracy of the generated fruit images are paramount for effective training of computer vision algorithms. We critically assess how well each model replicates the appearance of various fruits.
2) **Generation Speed**: The efficiency of image generation, particularly in producing large datasets, is a significant factor. Models that generate images faster can considerably reduce the time required for dataset creation.
3) **Applicability to Different Fruit Types**: The versatility of the models in generating images across different fruit types is essential for their practical application in diverse agricultural settings. We evaluate the adaptability of each model to ensure their utility across various scenarios.

To ensure a comprehensive and fair evaluation, we integrate human expert assessments to evaluate the quality of the generated images. Experts in the field review these images to provide insights into their realism and usability for training purposes. This qualitative assessment, combined with quantitative metrics, offers a balanced and in-depth analysis of each model's performance.

The findings from this study are intended to guide the selection of the most suitable models for generating large datasets necessary for training AI systems in agricultural automation. By identifying the top-performing models, we aim to streamline the creation of training datasets, thereby accelerating the adoption of AI-driven solutions in agriculture. This can lead to enhanced productivity, reduced labor costs, and minimized fruit damage, ultimately contributing to more efficient and sustainable agricultural practices.

II. Methodology

This study evaluates five different AI models for generating fruit images, focusing on their performance in terms of image quality, generation speed, and versatility in handling various fruit types. The models selected for this evaluation are Stable Diffusion v1.4, Stable Diffusion v1.5, Stable Diffusion v2, Stable Diffusion XL Refiner 1.0, and SDXL Turbo [13], [14]. The methodology involves several key steps, from image generation to the final performance evaluation of segmentation models.

A. Model Selection Justification

The chosen models represent a range of advancements in the field of image generation, each offering distinct capabilities that are crucial for our study:

- **Stable Diffusion v1.4 and v1.5**: These models are well-regarded for their balance of image quality and computational efficiency, making them suitable for generating a wide variety of images with reasonable resource usage [13], [14].
- **Stable Diffusion v2**: This model incorporates improvements in text-to-image generation, enhancing both the quality and diversity of generated images. It is particularly effective in producing more detailed and visually appealing outputs [13].
- **Stable Diffusion XL Refiner 1.0**: Known for its ability to produce high-resolution images with refined details, this model is ideal for tasks requiring greater image fidelity and fine-grained visual features [14].
- **SDXL Turbo**: This model leverages advanced distillation techniques and NVIDIA's TensorRT acceleration to achieve real-time image generation speeds while maintaining high image quality. It is particularly advantageous for applications requiring rapid generation of large datasets [14].

B. Fruit Image Generation

For each fruit type—cherries, bananas, oranges, apples, and pineapples—1000 images were generated using each AI model. This resulted in a total of 5000 images per model, encompassing a diverse range of visual characteristics for each fruit. The process for image generation was as follows:

1) **Model Setup**: Each of the five AI models was set up and configured according to their respective guidelines and specifications.
2) **Image Generation**: Using the predefined prompts for each fruit type, 1000 images were generated by each model. The generation process was automated to ensure consistency and efficiency.
3) **Data Storage**: The generated images were stored in a structured manner, categorized by fruit type and the model used for generation.

C. Image Quality Analysis

Once the images were generated, each of them was carefully evaluated by human experts to classify whether the generated fruit images appeared realistic. The images were assessed based on their adherence to several quality criteria. If an image failed to meet these criteria, it was rejected. The rejection criteria included:

1) **Quantity**: If there was more than one fruit present in the image.
2) **Texture and Color**: If the texture and color of the fruit did not look correct.
3) **Structure and Shape**: If the structure and shape of the fruit were not accurate.
4) **Overall Appearance**: If the output did not look correct overall.

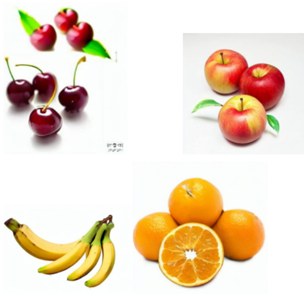
Examples of images that were rejected based on these criteria are illustrated in Figure 3.

This rigorous evaluation ensures that only high-quality, realistic fruit images are used for further processing and model training, thereby enhancing the effectiveness of the trained models in practical applications.

D. Performance Evaluation

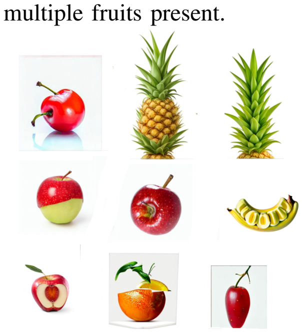
The performance of the generative AI models was evaluated based on two key metrics: accuracy and processing speed. The evaluation process included:

1) **Metric Calculation**: For each fruit type and model, accuracy and processing speed were calculated using standard formulas. The accuracy of the generated images was determined by the ratio of correctly identified fruit images to the total number of images evaluated. The processing speed was calculated as the time taken to generate a specified number of images, divided by the number of images. The formulas used are as follows:
   - **Accuracy**:
   $$\text{Accuracy} = \frac{\text{Correct Images}}{\text{Total Images}} \times 100$$
   - **Processing Speed**:
   $$\text{Speed (sec/image)} = \frac{\text{Total Time (sec)}}{\text{Total Images}}$$



(a) Example of rejection due to multiple fruits present.

(b) Example of rejection due to incorrect texture and color.



(c) Example of rejection due to incorrect structure and shape.

(d) Example of rejection due to overall incorrect appearance.

Fig. 3: Examples of generated fruit images that were rejected based on various quality criteria.

2) **Human Expert Evaluation**: Human experts reviewed a subset of the generated images to provide qualitative assessments of the segmentation accuracy.
   a) **Evaluation Criteria**: Experts assessed each image based on predefined criteria, including:
      - **Realism**: The extent to which the generated image resembles a real fruit.
      - **Texture and Color Accuracy**: The accuracy of the fruit's texture and color compared to natural fruit.
      - **Structural Integrity**: The correctness of the fruit's shape and structure.
      - **Overall Appearance**: The general appearance and any noticeable artifacts or errors.
   b) **Scoring System**: Each criterion was scored as either "Good" or "Bad," with "Good" indicating the image meets the quality standard and "Bad" indicating it does not.
   c) **Compilation of Scores**: The scores for each criterion were compiled to provide an overall quality score for each image.
   d) **Analysis of Results**: The scores were analyzed to identify trends and determine the overall performance of each model in generating realistic fruit images.
3) **Result Compilation**: The performance metrics were compiled into a comprehensive table for comparison across different models and fruit types.

By following this methodology, we aim to provide a thorough comparison of the five AI models, highlighting their strengths and weaknesses in generating realistic and usable fruit images. The results from this study will inform the selection of the most suitable model for creating large datasets required for training AI systems in agricultural automation, ultimately contributing to more efficient and sustainable agricultural practices.

III. Results and Discussion

The performance evaluation of the five AI models—Stable Diffusion v1.4, Stable Diffusion v1.5, Stable Diffusion v2, Stable Diffusion XL Refiner 1.0, and SDXL Turbo—is presented in detail in Table I. Each model was assessed based on image quality, generation speed, and versatility in handling different fruit types. The following subsections provide an in-depth analysis of these evaluations.

Overall, many models achieved very good results, with accuracy rates close to 100%. For instance, Stable Diffusion v1.4 and v1.5 performed exceptionally well in generating images of pineapples, achieving accuracy rates of 0.99. On the other hand, some models performed poorly for certain fruit types. Notably, Stable Diffusion v2 almost failed completely in generating accurate images of cherries, with an accuracy of just 0.01, and also showed poor performance in generating orange samples with an accuracy of 0.11.

The generation speed across all models ranged between 0.5 seconds to just under 1.5 seconds per image. Despite the high accuracy in some cases, the speed varied significantly, with SDXL Turbo being the fastest but less consistent in accuracy for some fruits.

A. Average Performance of Generative Models on Fruit Image Generation

Table II provides a comprehensive overview of the average accuracy and speed of five different Stable Diffusion models: v1.4, v1.5, v2, XL 1.0, and SDXL Turbo. This table is instrumental in understanding the overall performance trends and efficiency of these models in generating fruit images.

According to the table, the SDXL Turbo model exhibits the highest average accuracy at 0.891, indicating its superior performance in accurately generating fruit images. This is closely followed by Stable Diffusion XL 1.0, which has an average accuracy of 0.852, and Stable Diffusion v1.4 with an average accuracy of 0.840. These results demonstrate that the XL and Turbo variants of the Stable Diffusion models are particularly adept at producing high-quality fruit images.

However, when it comes to speed, the models show significant variation. The Stable Diffusion v2 model, despite its relatively low average accuracy of 0.500, is the fastest, with an average generation speed of 0.728 seconds per image. This makes it an efficient choice for scenarios where speed is a critical factor, despite its lower accuracy. On the other end of the spectrum, SDXL Turbo, while being the most accurate, is the slowest with an average speed of 1.140 seconds per image. Similarly, Stable Diffusion XL 1.0 also shows slower

TABLE I: Performance Comparison of Models on Fruit Image Generation (speed in seconds)

Fruits	Model	Accuracy	Speed (sec/image)
Cherry	SD v1.4	0.85	0.466
Cherry	SD v1.5	0.89	0.494
Cherry	SD v2	0.01	0.434
Cherry	SD XL 1.0	0.91	0.762
Cherry	SDXL Turbo	0.93	1.012
Banana	SD v1.4	0.56	0.778
Banana	SD v1.5	0.47	0.794
Banana	SD v2	0.60	0.688
Banana	SD XL 1.0	0.74	1.138
Banana	SDXL Turbo	0.94	1.092
Orange	SD v1.4	0.98	0.788
Orange	SD v1.5	0.99	0.804
Orange	SD v2	0.11	0.676
Orange	SD XL 1.0	0.95	1.134
Orange	SDXL Turbo	0.68	1.094
Apple	SD v1.4	0.83	0.798
Apple	SD v1.5	0.83	0.796
Apple	SD v2	0.81	0.690
Apple	SD XL 1.0	0.98	1.152
Apple	SDXL Turbo	0.93	1.108
Pineapple	SD v1.4	0.99	1.008
Pineapple	SD v1.5	0.99	1.030
Pineapple	SD v2	0.97	0.854
Pineapple	SD XL 1.0	0.70	1.356
Pineapple	SDXL Turbo	0.97	1.394

performance with an average speed of 1.108 seconds per image.

In summary, SDXL Turbo produces the best results in terms of accuracy, making it ideal for applications where image quality is paramount. On the contrary, if speed is more crucial, Stable Diffusion v2 would be the preferred choice due to its faster generation times, albeit at the cost of accuracy. The trade-offs between speed and accuracy in these models highlight the importance of selecting the appropriate model based on the specific requirements of the task at hand.

TABLE II: Average Performance of Generative Models on Fruit Image Generation

Model	Avg. Accuracy	Avg. Speed (sec/image)
SD v1.4	0.84	0.768
SD v1.5	0.83	0.784
SD v2	0.50	0.728
SD XL 1.0	0.85	1.108
SDXL Turbo	0.89	1.140

B. Average Performance of Generative Models by Fruit Type

Table III provides valuable insights into the performance of generative models when tasked with creating images of various fruits. The metrics used to evaluate performance include average accuracy and average speed (in seconds per image) for cherries, bananas, oranges, apples, and pineapples.

From this table, it is evident that the models perform best when generating images of pineapples, achieving the highest average accuracy of 0.924. This high accuracy suggests that the models are particularly effective at capturing the distinct features and textures of pineapples, resulting in high-quality images. Apples also have a relatively high average accuracy of 0.872, indicating that the models are similarly proficient in generating accurate images of this fruit type.

Conversely, bananas have the lowest average accuracy at 0.662, suggesting that the models struggle more with generating images of bananas compared to other fruits. This could be due to the simpler and more uniform appearance of bananas, which might be more challenging for the models to capture distinctively.

In terms of speed, there is considerable variation among the fruit types. Cherries are the fastest to generate, with an average speed of 0.553 seconds per image, making them the most efficient in terms of generation time. On the other hand, pineapples take the longest to generate, with an average speed of 0.87 seconds per image. This longer generation time for pineapples could be attributed to their complex texture and intricate details, which require more processing time to accurately render.

Overall, the table illustrates a trade-off between the accuracy and speed of generating different fruit images. While models perform exceptionally well in generating high-quality images of pineapples and apples, they do so at the cost of longer generation times. Conversely, images of cherries can be generated more quickly, though with slightly lower accuracy. Understanding these performance dynamics can help in selecting the appropriate model and fruit type based on specific requirements for accuracy and efficiency.

TABLE III: Average Performance of Generative Models by Fruit Type

Fruit	Average Accuracy	Average Speed (sec/image)
Cherry	0.717	0.553
Banana	0.662	0.776
Orange	0.742	0.741
Apple	0.872	0.748
Pineapple	0.924	0.870

IV. CONCLUSION AND FUTURE WORK

This comparative analysis of AI models for fruit image generation has provided valuable insights into their performance, highlighting both their strengths and limitations. SDXL Turbo emerged as the top performer in terms of accuracy, consistently producing high-quality images across various fruit types, albeit at the cost of slower generation times. Stable Diffusion XL Refiner 1.0 also showed strong performance, particularly excelling in accuracy, positioning it as a viable option for applications requiring high-fidelity images.

Our findings emphasize the potential of advanced generative models, particularly those optimized for efficiency, in generating synthetic datasets that are essential for fruit object detection and segmentation in agricultural robotics. The high accuracy of models like SDXL Turbo and Stable Diffusion XL

1.0 can significantly enhance the precision of automated systems in identifying and processing different fruits, contributing to advancements in agricultural technology.

However, notable discrepancies were observed, particularly with Stable Diffusion v2, which struggled to generate accurate images for certain fruits, such as cherries and oranges. This highlights the need for continued refinement of these models to improve their versatility and consistency.

Future research will focus on refining AI algorithms to enhance image generation accuracy and expand the scope to include a wider range of fruit types. Additionally, efforts will aim to optimize the balance between image quality and generation speed to increase the practical usability of these models in real-world agricultural applications. Such improvements will not only streamline fruit image generation but also broaden the applicability of generative models across diverse agricultural environments.

By addressing these challenges and building on the strengths of advanced AI models, agricultural robotics can achieve greater accuracy and automation in fruit detection and processing, ultimately contributing to more efficient and sustainable agricultural practices.

References

[1] D. Nield, "Robots may soon have a place in fruit orchards," 2023. Accessed: 2023-07-23.

[2] P. Li, S.-h. Lee, and H.-Y. Hsu, "Review on fruit harvesting method for potential use of automatic fruit harvesting systems," *Procedia Engineering*, vol. 23, pp. 351–366, 2011.

[3] J. Liu, Z. Li, and P. Li, *Rapid damage-free robotic harvesting of tomatoes*. Springer, 2021.

[4] R. Reddy and S. Nagaraja, "Integration of robotic arm with vision system," in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1–5, IEEE, 2014.

[5] T. Yoshida, Y. Onishi, T. Kawahara, and T. Fukao, "Automated harvesting by a dual-arm fruit harvesting robot," *ROBOMECH journal*, vol. 9, no. 1, p. 19, 2022.

[6] H. Zhou, X. Wang, W. Au, H. Kang, and C. Chen, "Intelligent robots for fruit harvesting: Recent developments and future challenges," *Precision Agriculture*, vol. 23, no. 5, pp. 1856–1907, 2022.

[7] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 12, pp. 6999–7019, 2021.

[8] K. Islam, "Recent advances in vision transformer: A survey and outlook of recent work," *arXiv preprint arXiv:2203.01536*, 2022.

[9] W. Li, "Analysis of object detection performance based on faster r-cnn," in *Journal of Physics: Conference Series*, vol. 1827, p. 012085, IOP Publishing, 2021.

[10] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia computer science*, vol. 199, pp. 1066–1073, 2022.

[11] A. Aggarwal, M. Mittal, and G. Battineni, "Generative adversarial network: An overview of theory and applications," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100004, 2021.

[12] A. Singh and T. Ogunfunmi, "An overview of variational autoencoders for source separation, finance, and bio-signal applications," *Entropy*, vol. 24, no. 1, p. 55, 2021.

[13] OpenAI Journey, "15 best stable diffusion models you should try out [2024]," 2024.

[14] N. D. Blog, "New stable diffusion models accelerated with nvidia tensorrt," 2024.