

Clasificación de calidad de manzana para monitoreo de cosechabilidad utilizando visión por computador y algoritmos de aprendizaje profundo

Apple quality grading for harvestability monitoring using computer vision and deep learning algorithms

Andrés Alejandro Garcés Cadena¹  <https://orcid.org/0009-0009-2410-0034>
Oswaldo Aníbal Menéndez Granizo²  <https://orcid.org/0000-0001-7101-0319>
Edgar Patricio Córdova¹  <https://orcid.org/0000-0002-2866-287X>
Álvaro Javier Prado Romo^{2*}  <https://orcid.org/0000-0001-7192-1681>

Recibido 06 de abril de 2023, aceptado 06 de junio de 2023

Received: April 06, 2023 Accepted: June 06, 2023

RESUMEN

La industria agrícola comprende una actividad de marcada influencia sobre el crecimiento económico y calidad de vida de las personas. Dada la necesidad de cubrir la demanda de alimentos debido al crecimiento poblacional, actualmente se requieren de sistemas capaces de optimizar el rendimiento del cultivo. Es así como este trabajo contribuye con una herramienta práctica para asistir al agricultor en tareas de reconocimiento de calidad de fruta, la misma que le permite mejorar el proceso de cuantificación de manzana y monitoreo del estado cosechable de la fruta mediante el uso visión por computador y algoritmos de aprendizaje profundo. El sistema propuesto presenta i) la detección del tipo de manzanas para el conteo y ii) la clasificación de su calidad para la inspección y validación de la fruta por categoría. Para la detección del tipo de manzana se utiliza el modelo de red de detección SSD-MobileNet y para la segmentación de instancias de calidad a nivel de píxel se emplea una red neuronal convolucional rápida FCN-ResNet 18. El sistema fue entrenado, validado y puesto a prueba en varios ensayos experimentales de laboratorio y campo, empleando dos bases de datos de imágenes construidas en ambientes controlados y en entornos agrícolas reales. Los resultados muestran que es posible detectar y clasificar el estado de calidad de manzanas durante la cosecha, obteniendo una precisión que varían entre el 86,7% y 92,6% para la detección y de $94,7 \pm 2,5\%$ para la segmentación, superando en ambos casos los resultados presentados en trabajos relacionados.

Palabras clave: Clasificación de manzana, visión por computador, aprendizaje profundo, redes neuronales convolucionales, segmentación de instancias.

ABSTRACT

The agricultural industry comprises an activity that has a marked influence on economic growth and people's quality of life. Given the need to meet the food demand due to population growth, systems capable of optimizing crop yield are currently required. This work contributes with a practical tool that

¹ Universidad Técnica de Ambato. Facultad de Ingeniería en Sistemas, Electrónica e Industrial, Ambato, Ecuador.
E-mail: agarcesc7@gmail.com; edgarpcordovac@uta.edu.ec

² Universidad Católica del Norte. Departamento de Ingeniería de Sistemas y Computación. Antofagasta, Chile.
Email: alvaro.prado@ucn.cl, oswaldo.menendez@ucn.cl

* Autor de correspondencia: alvaro.prado@ucn.cl

aids farmers in recognizing fruit quality, which is aimed at improving the apple quantification process and harvestability monitoring based on object identification employing computer vision techniques and deep learning algorithms. The development of the system presents i) detection of the apple classes for product counting and ii) quality classification for inspection and validation of the fruit by category. For grading of apple types, the SSD-MobileNet detection network model was used. A fast convolutional neural network FCN-ResNet 18 was used to segment quality instances at the pixel level. The proposed system was trained, validated, and tested on several experimental laboratory and field scenarios using two image databases generated in controlled and real agricultural environments. Results show that it is possible to detect and classify the quality status of apples during harvest, obtaining an accuracy ranging between 86.7% and 92.6% for detection and $94.7 \pm 2.5\%$ for segmentation, exceeding the results presented in related works in both cases.

Keywords: *Apple grading, computer vision, deep learning, convolutional neural networks, instance segmentation.*

INTRODUCCIÓN

La agricultura es una de las actividades que forman parte del sector primario de la economía mundial, la misma que proyecta un crecimiento del 70% debido a la expansión poblacional para el 2050 [1]. Por consiguiente, es necesaria la implementación de desarrollos tecnológicos que permitan optimizar los procesos de producción para cubrir las diferentes y progresivas necesidades de los cultivos frente a dicho crecimiento [2]. La Agricultura de Precisión (AP) ha evolucionado hacia un enfoque de información, cuyo objetivo es adquirir la mayor cantidad de datos posibles del cultivo para la toma de decisiones sobre diversas tareas que anteriormente se realizaban de forma manual, como la gestión del entorno agrícola, administración eficiente de las plantaciones, optimización de costos y tiempos de producción, entre otras [2]. Durante la última década, la producción sostenible de alimentos a nivel mundial se ha visto influenciada por el desarrollo de tecnologías asociadas con el paradigma de la AP [3], demostrando su efectividad a lo largo de sus procesos productivos [4]. Entre las principales aplicaciones de tecnologías en AP se encuentra la detección de plaga y maleza, riego distribuido y temporizado, aplicación automatizada de agroquímicos y fertilizantes, fenotipado de plantaciones, mapeo de producción, cosecha y otros [5]. A pesar de varias aplicaciones ya implementadas en la práctica, el concepto de AP aún no presenta un desarrollo distintivo debido a que los productores usualmente no poseen herramientas y recursos necesarios para tecnificar la producción en su totalidad (5). De aquí surge el motivo principal de este trabajo, el cual se

centra en la construcción de un sistema de asistencia al agricultor que le permita identificar frutas en árboles de manzanos y clasificar su calidad.

En la actualidad, la mayoría de los productores agrícolas deben asegurar que las frutas, verduras y vegetales cumplan con estándares de calidad con el fin de satisfacer requerimientos en los exigentes mercados globales. Por lo tanto, el productor debe monitorear continuamente el proceso de producción de ciertos productos en cada una de sus etapas, es decir, desde la preparación y siembra, pasando por el crecimiento, hasta la cosecha y entrega). En particular, la inspección de calidad del fruto durante su crecimiento es fundamental para la identificación temprana de riesgos potenciales; sin embargo, esta tarea es compleja dependiendo del tipo de fruta y entorno de producción. Por ejemplo, en huertos de manzanas, la identificación y clasificación manual de fruta en medio de regiones densas de árboles y en condiciones ambientales hostiles significa una tarea usualmente engorrosa para el operador, creando así condiciones de riesgo de evaluación inadecuada y de resultados equívocos a causa de factores como la repetibilidad de la tarea y el esfuerzo físico [6]. Entonces, métodos automáticos y no-invasivos para la evaluación de la calidad aparente de la fruta mediante técnicas de visión por computador y procesamiento de imágenes combinadas con estrategias de aprendizaje profundo son más que bienvenidos.

Este estudio presenta el diseño y evaluación de un sistema terrestre de base portable para la clasificación de calidad de manzanas, cuyo hardware fue

modificado para fines de este trabajo. El sistema está desarrollado con una estrategia de detección automática de la fruta y un enfoque de segmentación de instancias de forma y borde, las mismas que están dispuestas sobre árboles frutales en huertos de producción agrícola real. Para la detección, una red sujeta a un Detector Multicuadro de Disparo Único se consideró como meta estructura y se juntó con la base de las redes neuronales convolucionales para dar forma a una red SSD-MobileNet, mientras que para la inferencia de la forma de manzana a nivel de píxel se emplea un modelo de red neuronal residual con 18 capas de profundidad FCN-ResNet 18. Cada manzana dentro del conjunto de entrenamiento, validación y prueba es detectada automáticamente mediante un agente que clasifica la fruta bajo estudio en cinco categorías tales como Royal Gala, Red Delicious, Fuji, Grand Smith, y Golden Delicious, las mismas que son las más comunes en la exportación [7]. Además, el clasificador considera tres posibles estados semánticos de calidad que permiten discernir el grado de aceptabilidad de la fruta, incluyendo alto, medio y bajo. Dos conjuntos de entrenamiento para cada escenario son considerados, y una variedad de condiciones de posición de árboles, iluminación, tamaño, fondo y color de fruto también son tomados en cuenta para la generación de la base de datos de imágenes y para las pruebas experimentales realizadas. Para comparar cuantitativamente el desempeño del sistema propuesto frente a diferentes condiciones de operación, varias métricas de evaluación fueron consideradas, incluyendo la precisión, exactitud, exhaustividad, valor-F y matrices de confusión para la detección, mientras que la media de la intersección bajo la unión para la segmentación.

El resto de este escrito está estructurado de la siguiente manera. En la siguiente Sección se describe brevemente los trabajos del estado del arte que están relacionados con esta investigación. La posterior Sección presenta la metodología propuesta en este trabajo junto con las arquitecturas de red SSD-MobileNet y FCN-ResNet 18, las mismas que son empleadas en los sistemas de detección y segmentación de manzanas, respectivamente. La configuración de integración de hardware, la descripción de las pruebas experimentales y el análisis de los resultados en base a métricas de desempeño son presentados en la Sección de Resultados Experimentales. Finalmente, la última

Sección incluye las conclusiones extraídas de esta investigación.

TRABAJO RELACIONADOS

La visión por computador emplea la extracción automática de información proveniente de imágenes para analizar y generar modelos de estimación de procesos físicos, y en la actualidad esta tecnología se ha extendido significativamente hacia la automatización de ciertas tareas de la agricultura. Entre algunas de estas tareas están el monitoreo y evaluación de la siembra, el mapeo de rendimiento de cultivo, identificación de áreas de maleza y suelo de cultivo, estado de salud de la cosecha, entre muchas otras [8]. Recientes estudios sugieren que los dispositivos basados en técnicas de visión por computador pueden ser considerados como herramientas efectivas, rápidas y no invasivas para el fenotipado de plantas en ambientes agrícolas [9]. De la misma forma, los autores de estos trabajos sugieren que las técnicas de identificación de objetos y de segmentación basadas en métodos clásicos de aprendizaje de máquina son adecuadas para la detección, identificación y clasificación de frutos [5]. Existen métodos de aprendizaje de máquina, considerados ya tradicionales [10], que ampliamente han sido utilizados para el análisis de imágenes de frutos como regresiones lineales y no lineales [11], análisis de componentes principales [12], agrupamiento por K-means [13], mezcla de Gaussianas [14], clusterización iterativa [15], histograma de gradiente direccional [4], máquinas de soporte vectorial [16], y otras técnicas que soportan las anteriores para la extracción de características como transformada de Hough [17], transformadas de ondículas [18], espectro de potencia y muchos más [19]. Dichos métodos generalmente se basan en características geométricas, color, forma o textura, lo que resulta ser un desafío para lograr un resultado sólido en un entorno agrícola real [20].

De modelos basados en Redes Neuronales Artificiales, destacan las Redes Neuronales Convolucionales (Convolutional Neural Network - CNN) por el nivel de precisión o exactitud en la estimación que pueden alcanzar [21]. Las CNNs están conformadas de varias capas neuronales convolucionadas con retroalimentación. Estas redes son capaces de aprender en base a características extraídas mediante bancos de filtros que se disponen por etapas de forma

jerárquica. Los modelos de aprendizaje profundo en base a convolución se han utilizado de forma intensiva para la ingeniería en extracción de características de una manera automática [22]. Además, las redes neuronales profundas habitualmente presentan capacidades para aprender características de mayor complejidad que su contraparte (las redes neuronales menos profundas). Trabajos de investigación actual revelan que los métodos de aprendizaje profundo basados en CNN y con capacidades avanzadas de expresión semántica podrían mejorar la robustez en la detección ya que podrían aprender de forma autónoma las características superficiales y profundas de los objetos [23]-[25]. También, mediante las técnicas de aprendizaje profundo es posible trabajar con condiciones de objetos y entornos no estructurados en la que los métodos tradicionales presentan dificultades, tales como invarianza a la rotación y acercamiento al objeto dentro de la escena, entre otras [26]. Varios investigadores sugieren emplear arquitecturas convolucionales preconstruidas (o preentrenadas) para la obtención de modelos de estimación de fruta dada la eficiencia computacional, entre algunas de ellas están AlexNet, CaffeNet, VGG, MobileNet, ResNet y otros [24]. Sin embargo, se indica que la preferencia de CNNs en aplicaciones de portabilidad sean tan profundas como las capacidades de hardware en términos de memoria y capacidad de procesamiento puedan ser eficientes.

Redes RBF (Radial Basis Function) combinadas con CNN han sido entrenadas para detección y conteo de verduras [27], mostrando un alto grado de desempeño en condiciones de iluminación ideales. Aquí se indica la relevancia de la extracción de características en base al tamaño, forma y color como entradas iniciales al sistema de detección. Sin embargo, para alcanzar los resultados presentados se requirió el entrenamiento sobre un elemento de cómputo demandante como tarjetas aceleradoras gráficas. Redes CNN para identificar falencias en la forma de manzanas han sido también presentadas en la literatura [28]. Por ejemplo, una red profunda que mejora el modelo LeNet-5 fue evaluada en manzanas de mesa [6]. No obstante, comparado con máquinas de soporte vectorial, el desempeño del modelo propuesto solo mejoró cuando la cantidad de imágenes de entrenamiento fue lo suficientemente grande. Estos modelos lograron una precisión de alrededor del 96% para diferentes enfoques de postula de la cámara sobre la detección

e identificación manzanas Red Fuji. En [29], se presenta un modelo de estimación para detectar y clasificar distintos tipos de manzanas de similitud apariencia en términos de forma y color. En este trabajo se realizaron entrenamientos utilizando varios conjuntos de imágenes obtenidas en la web, bajo características ideales de fondo, color y posición de la cámara. En [30], se presenta un sistema para detección de pimientos dulces y otras frutas que, mediante la utilización de Faster R-CNNs, se genera un modelo en base a imágenes RGB y NIR del conjunto de datos ImageNet. Para el entrenamiento del modelo se utilizaron 484 imágenes y para la prueba se utilizaron 118 imágenes distribuidas en 6 clases diferentes de la fruta, incluyendo melón, palta, mango, naranja y manzana. Para el desarrollo del sistema, se realizó un ajuste de hiper-parámetros en la red preentrenada VGG 16. Como resultados se obtuvo una precisión del 57% y una sensibilidad del 80%. Este trabajo indica que los escenarios al aire libre frecuentemente generan inconvenientes en la detección, producto de los cambios continuos en iluminación, oclusión y sombra.

Basado en una arquitectura Faster R-CNN [17] compuesta por una red ZF y VGG16, se propone la detección de frutas en huertos (manzana, mango y almendras). La base de datos de las imágenes es ImageNet con tamaño inicial de 3496×2472 pxls, considerando para el entrenamiento del modelo 2268 imágenes y para la validación y prueba 482 imágenes de cada clase. Los resultados presentan un 95,8% de precisión y 86,3% de sensibilidad considerando todo el conjunto de datos; sin embargo, es importante indicar que los resultados son sensitivos a cambios de iluminación constante a pesar de haber realizado las pruebas solo en horarios matutinos. Una mayor mejora se realizó en [31] para arándanos, donde un alto porcentaje de precisión es alcanzado, pero los resultados varían en base a un ajuste de un conjunto considerable de hiperparámetros de la red. En [26], se muestra la comparación de resultados del modelo MobileNet con los resultados de otros modelos, en el cual el modelo escogido presenta la menor cantidad de parámetros de ajuste. En comparación con redes de detección de doble etapa, las redes de detección en una sola etapa generalmente muestran una velocidad de detección superior [32]. Por ejemplo, YOLO (You Only Look Once por sus siglas en inglés) es un algoritmo para CNNs de tiempo-real que se ha

utilizado para obtener una mayor velocidad en la detección de frutas. En [33], se aplicó YOLOv4 para detectar manzanas y se obtuvo una tasa del 92,6%, que fue alrededor de 8,0% más alta que Fast R-CNN y un 7,0% más alta que SSD. De manera similar, se utilizó la red YOLOv4-tiny para la detección de manzana [34], alcanzando una tasa del 94,4% en 18ms para procesar una imagen con resolución de 1280 x 720. Es así como se considera que esta última arquitectura tiene el potencial de lograr una detección de alta precisión a una velocidad relativamente rápida. Sin embargo, las capacidades de hardware son exigentes en computadores de limitadas capacidades de memoria y procesamiento.

Por otra parte, en [35] se ha desarrollado un sistema para cuantificar el estado de madurez de arándanos, utilizando el método de segmentación de imágenes con técnicas de aprendizaje profundo empleando el modelo Mask R-CNN. La precisión media del modelo fue del 78,3%. Este estudio permite extraer rasgos del fruto para el monitoreo y evaluación de la cosecha. Un mayor avance se encuentra en la clasificación del tipo de madurez de manzana utilizando imágenes RGB bajo una arquitectura modificada En-UNet [36], alcanzando una precisión del 97,54% en la validación del modelo de segmentación. Sin embargo, dichos porcentaje se reducen drásticamente a menos del 93,1% cuando el nivel intermedio de madurez de la fruta no está bien definido. En [37], se realiza una comparación de resultados de dos arquitecturas de red utilizadas para resolver problemas de clasificación de imágenes a nivel de píxel como son AlexNet y ResNet-18. En la segunda arquitectura se presenta menor carga computacional con respecto a la primera, seleccionándose ésta última para la segmentación de objetos dado que se requiere implementar en un sistema de alta portabilidad con un procesador de capacidades reducidas. Algunos de los trabajos mencionados anteriormente se resumen en Tabla 1.

METODOLOGÍA

En esta sección se presentan arquitecturas de las redes para detección del tipo de fruta y segmentación de acuerdo con su grado de calidad. Para la detección del tipo de manzana se escoge utilizar la arquitectura SSD-MobileNet debido a que el costo computacional es reducido en aplicaciones de hardware con capacidades reducidas. La configuración de la

red, y programación de los algoritmos fueron implementados utilizando librerías de PyTorch. SSD-MobileNet [36] es una arquitectura de red para la detección de objetos, donde una de sus características fundamentales es su funcionamiento en tiempo real en dispositivos móviles e integrados. Además, esta red combina el detector SSD-300 Single-Shot MultiBox [37] con una red troncal MobileNet [36], como se observa en la Figura 1.

La estructura de red se compone de convoluciones separables en profundidad, a excepción de la primera capa que es una convolución completa, las cuales extraen las características necesarias para la detección de objetos en las imágenes. Brevemente se identifican los procesos para la detección de la red anteriormente mencionada en base a la arquitectura SSD-MobileNet en la Figura 2.

La segmentación semántica, por otro lado, se basa en el reconocimiento de imágenes, excepto que las clasificaciones de objetos ocurren a nivel de píxel en lugar de objetos compactos en la imagen completa. Esto se logra mediante la múltiple convolución de una red troncal de reconocimiento de imágenes previamente entrenada, que transforma el modelo en una red totalmente convolucional (FCN) capaz de etiquetar por píxel [38]. La segmentación produce clasificaciones densas por píxel de muchos objetos potenciales por escena, incluidos los primeros planos y los fondos de la escena [39].

El detalle de funcionamiento del sistema implementado para la segmentación se observa en la Figura 3, donde se observa el procedimiento de implementación de FCN ResNet18 para la inferencia del modelo de predicción. El algoritmo núcleo de segmentación para el entrenamiento está bajo el soporte de segNet, el cual acepta como entrada la imagen 2D y genera una segunda imagen con la superposición de máscara de clasificación por píxel. Cada píxel de la máscara corresponde a la clase de objeto categorizado.

ResNet-18 es una red neuronal convolucional que como se describe en [40], [41] tiene 18 capas de profundidad de las cuales 17 son capas convolucionales, 1 capa totalmente conectada y una capa softmax adicional para realizar la tarea de clasificación. La red permite un tamaño de imagen de entrada de 224 x 224. La arquitectura de la red original ResNet-18 se muestra en la Figura 4.

Tabla 1. Resumen de algunas contribuciones de redes neuronales profundas en aplicaciones de detecciones de defectos en alimentos.

Referencia	Tipo de dato	Tipo de característica	Método	Índices de desempeño
[5]	Cuatro sets de imágenes de manzanas. Cada set consiste de 270, 568, 274, y 414 imágenes.	Imágenes de color, formas geométricas, y características SIFT	Perceptrón multicapa + aprendizaje semi-supervisado	Precisión: 91,98-94,81%
[9]	Survey con diferentes tipos de datos, como: plantas, frutas, vegetales, hortalizas, y semillas	Variadas como: Color (HSV) y morfológicas, texturas de wavelets, información espacial (Profundidad y LiDAR)	Variados como: CNN, SURF+GMM, Deep CNN, basado en encoders, SVM, entre otros	Altos desempeños que varían entre: Precisión: 85%-99,04%, dependiendo del caso.
[11]	200 imágenes RGB de atún y manzanas Gala.	Rugosidad y frescura mediante color	Clasificador Lineal y SVM	Precisión: 91,98% en condiciones controladas
[14]	Imágenes RGB	Color RGB	AlexNet using a CNN + GMM	Precisión: 94% in three of four datasets Exactitud: 96-97%
[15]	Imágenes RGB y HSV	Color RGB, HSV y forma YCbCr	Clusterización iterativa lineal sobre bloque de super-píxeles de segmento. RCNN + HOG usado para descripción de forma.	Precisión: 95,12%
[16]	Tamaño de naranja y puntos de detección de bacterias	Tamaño, color y textura: RGB y CMYK.	SVM	Precisión: 67,74% en el mejor de los casos
[17]	50 imágenes de mango RGB	Bordes, distancias y morfología de imágenes B/N y RGB	Backpropagation Neural Network de 3 capas y Transformada de Hough	Precisión: 96,26%
[18]	Imágenes de manzana en escala de grises	Histogramas de frecuencia y características wavelets	RNN+wavelets con transformación de grises IBT	No especificado
[20]	Set de imágenes públicas de 4035 manzanas defectuosas	Imágenes RGB	UNet	Precisión: 97,54%
[22]	2633 imágenes JPG (1024 × 768), conteniendo palta, naranja, papaya, melón, kiwi, zanahoria	Dos grupos de imágenes RGB	CNN y una red pre-entrenada y reajustada	Precisión: 99,75% para modelo de validación

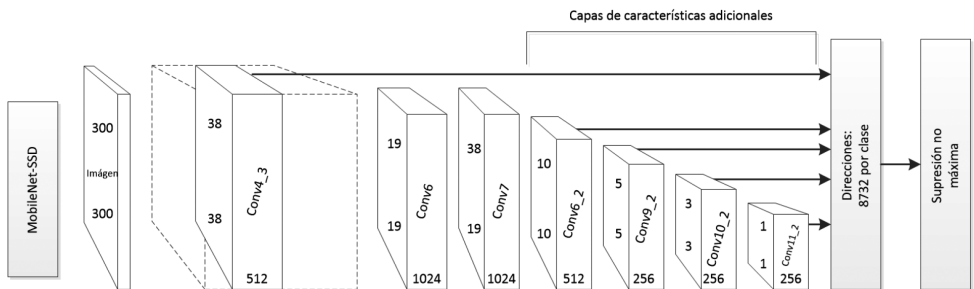


Figura 1. Estructura de la red SSD MobileNet. Las capas de convolución permiten generar mapas de extracción de características mediante filtros de convolución. El modelo de red SSD está diseñado para ser independiente de la red base, de tal forma que puede ser acoplada a otras como VGG, YOLO o MobileNet.

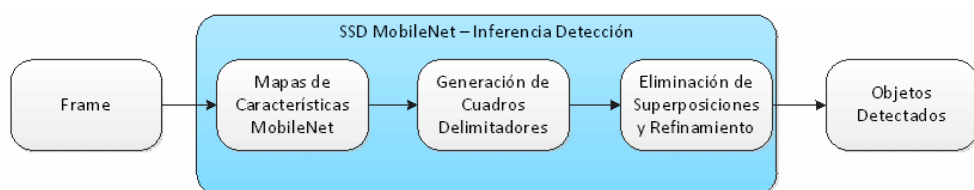


Figura 2. Esquema general del sistema de detección con SSD-MobileNet.

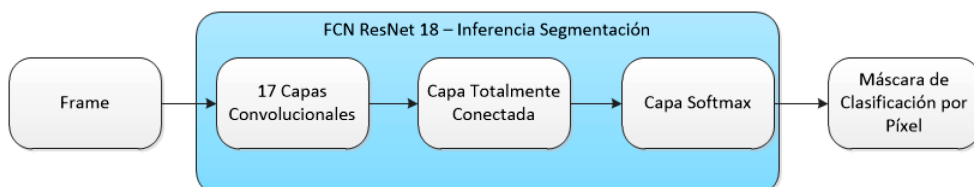


Figura 3. Esquema general del sistema de segmentación con FCN ResNet18.

Las capas convolucionales usan filtros de 3×3 y la red es diseñada de tal manera que, si el mapa de características de salida es del mismo tamaño, entonces las capas tienen el mismo número de filtros. Sin embargo, los filtros se duplican en las capas, si el mapa de características de salida se reduce a la mitad. La reducción de muestreo se realiza mediante capas convolucionales que tienen un paso de 2. Por último, hay una agrupación promedio seguida de una capa totalmente conectada con una capa softmax como se observa en la Figura 4.

A lo largo de la red propuesta, se insertan conexiones de acceso directo residuales entre capas, las mismas que son de dos tipos:

- El primer tipo de conexiones, indicadas por líneas sólidas, se utilizan cuando al ingreso y la salida tienen las mismas dimensiones.

- El otro tipo de conexiones, indicadas por líneas punteadas, se utilizan cuando aumentan las dimensiones, en este tipo de conexión todavía funciona mapeo de identidad, pero con relleno de ceros para aumentar las dimensiones con un paso de 2 como se detalla en [17].

RESULTADOS DE DETECCIÓN DE FRUTA

En esta Sección se presenta la configuración experimental, proceso de inferencia, los resultados obtenidos del sistema implementado, así como la respectiva validación y análisis de resultados.

Configuración Experimental

En la Figura 5 se muestra el diagrama del proceso experimental de cada uno de los procesos que se han llevado a cabo para la obtención del modelo, incluyendo la recolección de datos, etiquetado de

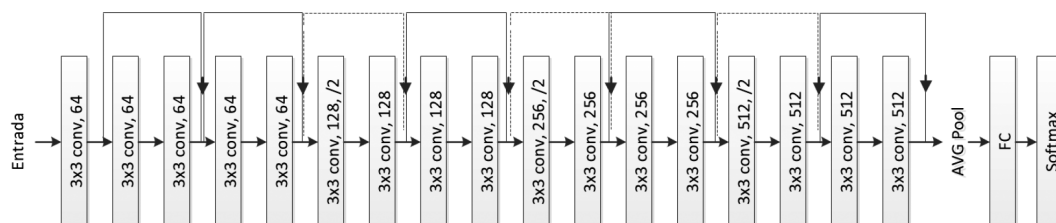


Figura 4. Arquitectura ResNet18 propuesta para extracción de características de los objetos (manzanas). Como se muestra en esta figura, existen dos tipos de conexiones de acuerdo con la dimensionalidad de los datos. Líneas sólidas representan las conexiones de la misma dimensionalidad, mientras que líneas entrecortadas denotan diferente dimensionalidad.

imágenes, entrenamiento, pruebas y análisis de resultados, para los procesos de detección del tipo de manzana y segmentación de acuerdo con su grado de calidad.

Los procesos preliminares detallados en la Figura 5 se detallan brevemente:

- Captura de imágenes: Proceso en el cual se adquirieron las bases de datos de imágenes de manzanas en el laboratorio y campo.
- Filtrado y pre-procesamiento: Proceso en el que se limpian los datos, se excluyen las imágenes que no agregan valor a la base de datos, se estandariza la resolución de las imágenes, y se normalizan los vectores de entrada.
- Etiquetado de imágenes: Proceso que consiste en delinear con cuadrados (detección) y polígonos (segmentación) al objeto asignando su característica correspondiente.
- División de base de datos: En este proceso se distribuyen las imágenes para el entrenamiento, validación y prueba.

A continuación, se detallan los procesos específicos para los procesos de detección del tipo de manzana y segmentación de acuerdo con su grado de calidad.

Configuración del Hardware del Sistema

En esta Sección se muestran los equipos utilizados para la implementación del sistema portable, el método de recolección de datos y los indicadores para la evaluación final.

El sistema propuesto consta de un minicomputador con placa electrónica Jetson Nano cuyo procesador es un ARM Cortex -A57 con máxima frecuencia de operación de 921MHz, y una memoria de 4GB. Tiene una unidad de procesamiento gráfico CUDA de 128 núcleos, incluida ya con Open GL (Figura 6). Aquí es importante indicar que el computador para el entrenamiento es distinto al que se utilizó para la validación y prueba dadas las facilidades de usabilidad y transportabilidad. Entonces, el computador de entrenamiento tiene un procesador x64 Intel(R) Core i7-870H CPU@ 2.2 y 2.21GHz, con memoria RAM de 15.8GB utilizable, más una tarjeta gráfica NVIDIA GeForce RTX 2070 con diseño Max-Q. El sistema operativo para el computador de entrenamiento como prueba y validación es Linux con distribución Ubuntu 18.04. Los dispositivos que se utilizaron para la obtención de datos de imágenes para el entrenamiento, validación y prueba del sistema fueron los siguientes; a) cámara para obtención de imágenes de entrenamiento Canon PowerShot

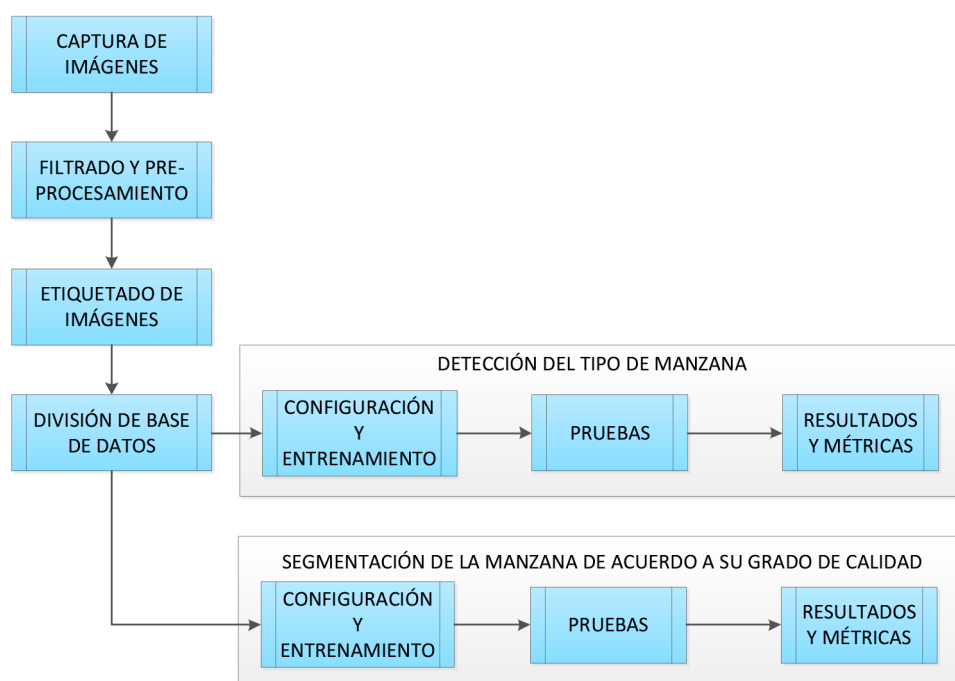


Figura 5. Diagrama general del proceso experimental.

SX400 IS la cual dispone de una resolución de 3456 x 3456 píxeles con FPS de 30 y campo de visión de 120° horizontal y vertical, b) cámara portable para la prueba del modelo con resolución de 3024 x 4032 píxeles.

Imágenes para conjunto de entrenamiento

Para la detección del tipo de manzana se adquirieron las 5 variedades que comúnmente están dentro de la gama exportable de los mercados globales [2] (Figura 7).

El primer tipo de manzana es Granny Smith, esta presenta una coloración verde generalmente uniforme e intenso, con forma alargada y simétrica. El segundo tipo de manzana es Golden Delicious, la misma que es de color amarillo oro, con tendencia a verde y con forma alargada y redondeada. El tercer tipo es la manzana Fuji, siendo de color rojo y rosa con matices amarillos y de forma redonda. El cuarto tipo es la manzana Royal Gala, y es de color rojo con matices anaranjados. Su forma es alargada y

relativamente simétrica. El último tipo es la manzana Red Delicious, la misma que es cónica, alargada y de un tamaño medio con un color rojo oscuro.

Para la cuantificación de calidad de manzana se definen tres categorías: alta, media y baja, las mismas que fueron ya analizadas previamente [23]. Esta selección de niveles de calidad comprende el criterio mínimo de validez de la fruta, el mismo que bordea al menos un elemento medio entre alto y bajo. Debido a que este rango de calidad es aceptable en la práctica de la industria, se adoptaron estas categorías para el presente estudio, las mismas que pueden ser diferenciadas visualmente en la Figura 8. Aquí se muestran tres imágenes de manzanas con los grados de calidad indicados. Para el proceso de segmentación se emplearon solamente manzanas de tipo Royal Gala dada la complejidad de la detección del estado de salud del fruto frente a las otras bajo prueba. En lo siguiente se especifica el grado de calidad descrito de forma semántica:



Figura 6. Configuración del experimento y sistema de adquisición de imágenes.

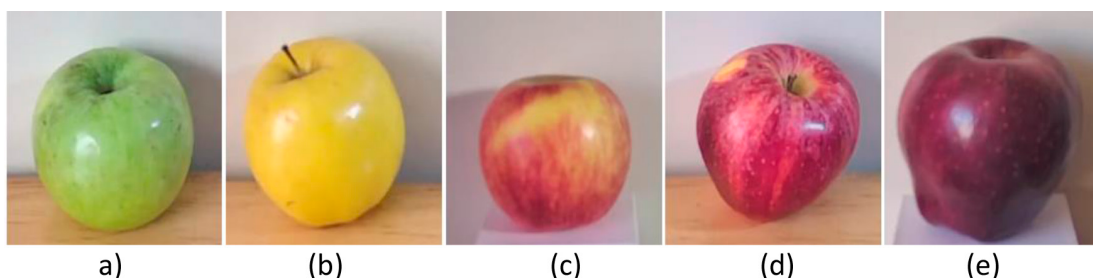


Figura 7. Granny Smith(a), Golden Delicious (b), Fuji (c), Royal Gala (d), Red Delicious (e).

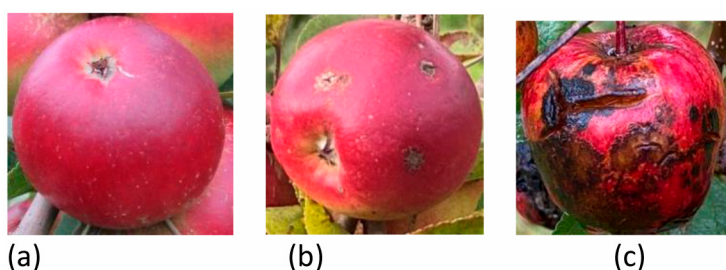


Figura 8. Manzanas con diferente grado de calidad. Alta calidad (a), Media calidad (b) y Baja calidad (c).

- Alta: No presenta defectos.
- Media: Presenta defectos leves y se encuentra apta para el consumo.
- Baja: Presenta descomposición que imposibilitan su consumo.

El conjunto de datos para obtención del modelo de detección fueron 5125 imágenes, mientras que las imágenes de manzanas que se usaron para la segmentación del nivel de calidad fueron 338 imágenes, conteniendo 2211 manzanas del tipo Royal Gala y con un grado de calidad variable, cuando la fruta estaba madura y a vísperas de la cosecha. Estas imágenes se distribuyeron en conjuntos para entrenamiento, validación y prueba de los ensayos de laboratorio y campo. Las condiciones de iluminación fueron controladas para las pruebas de laboratorio, es decir se mantuvo un nivel de

luz artificial constante durante todo el proceso de entrenamiento y ensayos. Por otro lado, para las pruebas de campo, las condiciones de iluminación fueron naturales. Las bases de datos de los conjuntos de imágenes se los obtuvo durante el mes de marzo del 2022 en horario vespertino (alrededor de las 15:00 horas) a una temperatura promedio de 19 °C y las condiciones climáticas generalmente presentaron días nublados. El sistema de detección de calidad de manzana fue implementado en un parque industrial de procesamiento de fruta, cerca de un huerto de árboles de manzana como se puede visualizar en la Figura 9.

Indicadores de Evaluación del Sistema de Detección y Segmentación

En esta Sección se presenta el esquema de evaluación empleado para la medición del rendimiento de los



Figura 9. Ambiente de pruebas experimentales en campo. La imagen de la izquierda muestra un mapa del sitio en donde se encuentra la planta industrial para el procesamiento de fruta (a), mientras que la imagen de la derecha es el huerto en donde se encuentran las frutas para las pruebas (b).

métodos propuestos. Varias métricas se han utilizado para la determinación del rendimiento del modelo entrenado para la detección de frutas en los árboles de manzanos. La precisión (ecuación (1)), exactitud (ecuación (2)), exhaustividad (ecuación (3)), valor-F (ecuación (4)) y matrices de confusión se utilizaron para la detección del tipo de manzana de acuerdo con la clasificación de las muestras, mientras que la intersección bajo la unión (ecuación (5)) para la segmentación [24]. Las ecuaciones que describen cada una de las métricas indicadas se presentan a continuación:

$$\text{Precisión} = \frac{VP}{VP + FP} \quad (1)$$

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2)$$

$$\text{Exhaustividad} = \frac{VP}{VP + FN} \quad (3)$$

$$\text{Valor} - F = 2 \frac{\text{Precisión} \times \text{Exactitud}}{\text{Precisión} + \text{Exactitud}} \quad (4)$$

donde:

- VP: Verdaderos Positivos. En este caso el valor real es positivo y la predicción determina que es positivo.
- FP: Falsos Positivos. En este caso el valor real es negativo y la predicción indica que es positivo.
- FN: Falsos Negativos. En este caso el valor real es positivo y la predicción indica que es negativo (o no fue detectado).
- VN: Verdaderos Negativos. En este caso el valor real es negativo y la predicción determina que es negativo.

La precisión indica el nivel de dispersión del conjunto de valores clasificados frente a todas las detecciones positivas de manzanas, mientras que la exactitud indica la cantidad de detecciones acertadas de manzanas frente al total de manzanas de muestras clasificadas. La exhaustividad expone la capacidad del estimador para revelar los casos positivos de manzanas identificadas, mientras que el valor-F indica la relación de precisión y exhaustividad, denotando la distribución de clases estimadas [24]. Para la evaluación del desempeño del sistema en la segmentación se emplea el método de Intersección sobre la Unión (IsU) [25]. Se identificó que este método es una métrica adecuada para medir la

superposición entre dos cuadros delimitadores o máscaras (ecuación (5)).

$$IsU = \frac{\text{Área}_{pred} \cup \text{Área}_{gt}}{\text{Área}_{pred} \cap \text{Área}_{gt}} \quad (5)$$

- Área_{pred} : Es el área resultante de la predicción de la segmentación.
- Área_{gt} : Es el área etiquetada en base a datos.
- $\text{Área}_{pred} \cup \text{Área}_{gt}$: Es la unión del área resultante de la predicción de la segmentación y el área etiquetada.
- $\text{Área}_{pred} \cap \text{Área}_{gt}$: Es la intersección entre el área resultante de la predicción de la segmentación y el área etiquetada.

Si la predicción es completamente correcta, $IsU = 1$. Cuanto menor sea el IoU, disminuye la calidad del resultado de la predicción.

Pruebas del Sistema de Detección de Fruta

Para la detección de fruta, existen tres tipos de pruebas. El primer conjunto de pruebas se realizó en un laboratorio bajo condiciones controladas de fondo de imagen, el siguiente conjunto considera las imágenes sin fondo, mientras que la tercera prueba se realiza en campo. En la Figura 10 y Figura 11 se muestra un ejemplo del panorama utilizado en los ensayos de la primera prueba y su detección. Para la ejecución de los ensayos de campo para la detección de la clase de manzana se ensambló la estructura del hardware en un huerto en donde las manzanas estaban dispuestas en un árbol sujetas a diferentes condiciones ambientales, de oclusión, sombra, traslape entre frutas.

En las imágenes de la Figura 11 se observan los cuadros delimitadores que indican el tipo y valor de confianza de cada manzana.

Pruebas del Sistema de Segmentación de Fruta

Para los ensayos de segmentación se tomaron muestras de las imágenes de árboles de manzanos del mismo huerto, procurando que sean del mismo tipo como se observa en la Figura 12. También se muestran las máscaras correspondientes de acuerdo con el grado de calidad de cada manzana.

Evaluación de los Resultados Experimentales

En esta Sección se muestran primeramente los resultados obtenidos en las pruebas de detección y seguidamente los resultados de segmentación.



Figura 10. Imágenes de los ambientes de laboratorio (a) y campo (b).

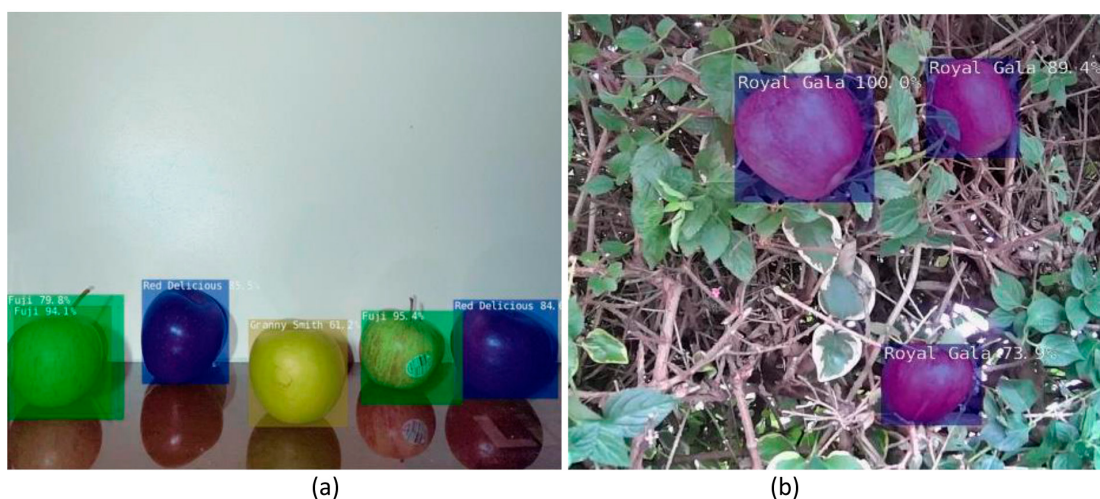


Figura 11. Resultado del sistema de detección en los ambientes de laboratorio (a) y campo (b).

Resultados del sistema de detección

Los resultados de los ensayos del primer conjunto de prueba de laboratorio se muestran en la matriz de confusión de la Tabla 2. De los cinco tipos de manzanas del modelo propuesto se observa que la manzana Golden Delicious presenta la mayor precisión y en el caso opuesto la manzana Royal Gala presenta la menor precisión. Además, la sensibilidad más alta corresponde a las validaciones con las manzanas Granny Smith y Red Delicious, mientras que la manzana Fuji presenta menor sensibilidad. La precisión general del modelo propuesto evaluado con el conjunto de datos de prueba es del 86,7%.

Los resultados de los ensayos del segundo conjunto de prueba de laboratorio se muestran en la matriz de confusión de la Tabla 3. De los cinco tipos de manzanas del modelo propuesto, se observa que las manzanas Granny Smith y Golden Delicious presentan la mayor precisión y en el caso opuesto la manzana Royal Gala presenta la menor precisión. Además, la sensibilidad más alta corresponde a las manzanas Granny Smith y Golden Delicious, mientras que la manzana Red Delicious presenta menor sensibilidad. La precisión general del modelo propuesto evaluado con el conjunto de datos de prueba es del 92,6%.



Figura 12. Imágenes originales y con máscara de segmentación.

Tabla 2. Matriz de confusión de conjunto de datos de prueba 1.

Predicción	Granny Smith	30	0	4	0	0	88,2%
		6,7%	0,0%	0,9%	0,0%	0,0%	11,8%
	Royal Gala	0	24	0	10	0	70,6%
		0,0%	5,3%	0,0%	2,2%	0,0%	29,4%
	Golden Delicious	0	0	26	0	0	100%
		0,0%	0,0%	5,8%	0,0%	0,0%	0,0%
	Fuji	0	1	0	20	0	95,2%
		0,0%	0,2%	0,0%	4,4%	0,0%	4,8%
	Red Delicious	0	5	0	0	30	85,7%
		0,0%	1,1%	0,0%	0,0%	6,7%	14,3%
		100%	80,0%	86,7%	66,7%	100%	86,7%
		0,0%	20,0%	13,3%	33,3%	0,0%	13,3%
		Granny Smith	Royal Gala	Golden Delicious	Fuji	Red Delicious	
		Real					

Los resultados de los ensayos del conjunto de prueba de campo se muestran en la matriz de confusión de la Tabla 4. De los cinco tipos de manzanas del modelo propuesto, se observa que cuatro tipos de manzanas

presentan una precisión del 100% y la manzana Fuji con una menor precisión. Adicionalmente, la manzana que mayor sensibilidad presenta es la Granny Smith y la manzana de menor sensibilidad en la Red Delicious. Es

Tabla 3. Matriz de confusión de conjunto de datos de prueba 2.

Predicción	Granny Smith	30	0	0	0	0	100%
		6,7%	0,0%	0,0%	0,0%	0,0%	0,0%
	Royal Gala	0	24	0	1	4	82,8%
		0,0%	5,3%	0,0%	0,2%	0,9%	17,2%
	Golden Delicious	0	0	30	0	0	100%
		0,0%	0,0%	6,7%	0,0%	0,0%	0,0%
	Fuji	0	2	0	30	2	88,2%
		0,0%	0,4%	0,0%	6,7%	0,4%	11,8%
	Red Delicious	0	2	0	0	24	92,3%
		0,0%	0,4%	0,0%	0,0%	5,3%	7,7%
		100%	85,7%	100%	96,8%	80%	92,6%
		0,0%	14,3%	0,0%	3,2%	20,0%	7,4%

	Granny Smith	Royal Gala	Golden Delicious	Fuji	Red Delicious	
Real						

Tabla 4. Matriz de confusión de conjunto de datos de prueba 3.

Predicción	Granny Smith	101	0	0	0	0	0	100%
		19,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	Royal Gala	0	87	0	0	0	0	100%
		0,0%	17,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	Golden Delicious	0	0	99	0	0	0	100%
		0,0%	0,0%	19,3%	0,0%	0,0%	0,0%	0,0%
	Fuji	0	0	1	90	0	0	98,9%
		0,0%	0,0%	0,2%	17,6%	0,0%	0,0%	1,1%
	Red Delicious	0	0	0	0	86	0	100%
		0,0%	0,0%	0,0%	0,0%	16,8%	0,0%	0,0%
	No detectada	5	18	5	6	14	0	0,0%
		1,0%	3,5%	1,0%	1,2%	2,7%	0,0%	100%

	Granny Smith	Royal Gala	Golden Delicious	Fuji	Red Delicious	No detectada	
Real							

importante indicar que para esta evaluación se adicionó la clase “no detectada”, debido a que las condiciones externas del ambiente agrícola presentaron dificultades en el proceso de detección debido a la variabilidad en las condiciones de iluminación, climáticas y oclusión.

Para el total de este caso (tercera prueba) se presentaron 48 manzanas no reconocidas lo que influye en la precisión general del modelo propuesto evaluado con el conjunto de datos con un 90,4% de exactitud de la red. Sin embargo, los mayores porcentajes de precisión se exhibieron en esta prueba con respecto a las anteriores. Mayor detalle de

los resultados alcanzados en los tres conjuntos de pruebas, se indican las evaluaciones de las métricas propuestas dentro de la Tabla 5.

Resultados del sistema de segmentación

La evaluación para esta sección se realizó de acuerdo con los resultados del entrenamiento en el que se obtuvo que los mejores resultados se presentaron en la época 218 con 94,7% de precisión. En la Figura 13 se observa la evolución de precisión alcanzada durante el entrenamiento, mientras que la media de la Intersección bajo la unión de todas las imágenes entrenadas se indica en la Figura 14.

Tabla 5. Resumen de resultados en los tres conjuntos de prueba.

Tipo de Manzana	Laboratorio						Campo		
	Prueba 1: Fondo controlado			Prueba 2: Fondo no controlado			Prueba 3: Ambiente externo		
	Precisión (%)	Exhaustividad (%)	Valor-F (%)	Precisión (%)	Exhaustividad (%)	Valor-F (%)	Precisión (%)	Exhaustividad (%)	Valor-F (%)
Granny Smith	88	100	94	100	100	100	100	95	98
Royal Gala	71	80	75	83	86	84	100	83	91
Golden Delicious	100	87	93	100	100	100	100	94	97
Fuji	95	67	78	88	97	92	99	94	96
Red Delicious	86	100	92	92	80	86	100	86	92
Exactitud	86,7%			92,6%			90,4%		

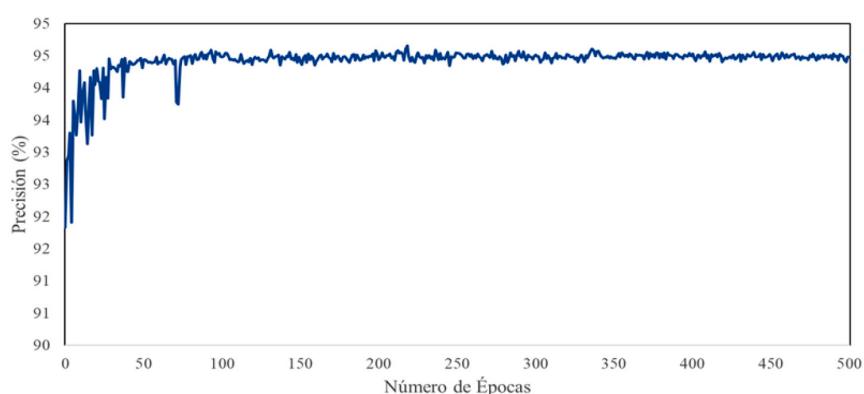


Figura 13. Evolución de la precisión de la red de segmentación FCN-ResNet 18 durante el entrenamiento de imágenes de manzanas.

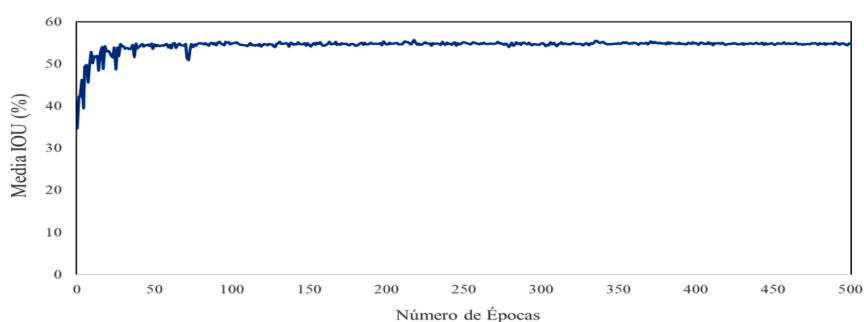


Figura 14. Evolución de la media de la Intersección bajo la unión de la red de segmentación FCN-ResNet 18 durante el entrenamiento de imágenes de manzanas.

Es importante indicar la evolución de la media de la intersección bajo la unión de la red de segmentación entrenada converge, lo que significa que en promedio el nivel de solapamiento es constante.

CONCLUSIONES

En este trabajo se desarrolla, implementa y valida un sistema portable de detección de calidad para la

identificación del estado cosechable de manzana. Para ello, un sistema de visión por computador y aprendizaje de máquina es puesto a prueba para la detección y segmentación de fruta. Se presenta un método de detección del tipo de variedad de manzana que permite el conteo de fruta, y un segundo método de segmentación que permite medir el nivel de calidad para su aceptabilidad de consumo. Para el primer método se emplea un modelo de red SSD-MobileNet, el mismo que es modificado para alcanzar el mejor rendimiento de desempeño con las imágenes de entrada. Para el segundo método, se utiliza una red neuronal convolucional rápida FCN-ResNet 18 con el objetivo de ser implementada sobre hardware de limitadas capacidades de memoria y procesamiento. Con estos dos métodos, el trabajo se direcciona a la provisión de una herramienta que permite al agricultor estimar la calidad de la fruta in-situ. Varias pruebas experimentales fueron llevadas a cabo tanto en entornos de laboratorio como en campo real. Los resultados indican que es posible identificar y clasificar semánticamente la calidad de manzanas previo a la cosecha, alcanzando una precisión entre el 86,7% y 92,6% para la detección y alrededor del 94,7% para la segmentación. En la detección del fruto se realizaron pruebas sobre ambientes controlados, semicontrolados, y naturales de luz, siendo éste último el más complejo por su influencia sobre la calidad de la imagen, distorsión y oclusión de objetos. Como hallazgo se identificó que una fuente de luz adicional permite obtener una mejor de mejor calidad y a su vez reflejar una mayor precisión de la red. En lo que respecta a la segmentación, la identificación del objeto y sus bordes mediante la estimación del tipo de manzana es difuso cuando existe gran cantidad de oclusión. Sin embargo, mediante una red FCN ResNet18, un mejor desempeño del clasificador es posible para hardware portable de limitadas capacidades de memoria. Lo indicado aquí sugiere el uso del sistema en aplicaciones reales donde la identificación de la calidad del fruto es usualmente realizada de forma manual. La fusión de características encontradas en ambos métodos propuestos y la extracción de nuevas características a nivel de píxel para ser utilizadas en combinación con otras arquitecturas similares son parte de la investigación en curso y trabajo futuro de los autores.

AGRADECIMIENTO

Los autores agradecen el soporte de ANID bajo el proyecto Fondecyt Iniciación 2023 con N° 11230962.

Esta investigación fue soportada y financiada por el proyecto 202203010029-VRIDT-UCN de la Universidad Católica del Norte, Anillo de Investigación en Ciencia, y Tecnología ACT210052, y Fondef IDEA I+D 2021 Cod. ID21 | 10181.

REFERENCIAS

- [1] CEPAL, FAO, IICA. Perspectivas de la agricultura y del desarrollo rural en las américas: una mirada hacia américa latina y el caribe”, San José, Costa Rica: Comunicaciones Milenio, 2021. [En línea]. <https://www.cepal.org/es/publicaciones/47208-perspectivas-la-agricultura-desarrollo-rural-america-mirada-america-latina>.
- [2] S. Condran, M. Bewong, M.Z. Islam, L. Maphosa, and L. Zheng, “Machine learning in precision agriculture: a survey on trends, applications and evaluations over two decades,” *IEEE Access*, vol. 10, pp. 73786-73803, 2022, doi: 10.1109/ACCESS.2022.3188649.
- [3] H.R. Seireg, Y.M.K. Omar, F.E.A. El-Samie, A.S. El-Fishawy, and A. Elmahalawy, “Ensemble machine learning techniques using computer simulation data for wild blueberry yield prediction,” *IEEE Access*, vol. 10, pp. 64671-64687, 2022, doi: 10.1109/ACCESS.2022.3181970.
- [4] F. Yandun Narvaez, G. Reina, M. Torres-Torriti, G. Kantor, and F.A. Cheein, “A survey of ranging and imaging techniques for precision agriculture phenotyping,” *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 6, pp. 2428-2439, 2017, doi: 10.1109/TMECH.2017.2760866.
- [5] P. Roy, A. Kislay, P. Plonski, J. Luby, and V. Isler, “Vision-based preharvest yield mapping for apple orchards,” *Computers and Electronics in Agriculture*, vol. 164, pp. 104897, 2019, doi: 10.1016/j.compag.2019.104897.
- [6] A. Bhargava and A. Bansal, “Fruits and vegetables quality evaluation using computer vision: a review,” *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 3, pp. 243-25, 2021, doi: 10.1016/j.jksuci.2018.06.002.
- [7] University of Illinois Board of Trustees, “Manzanas y másas”, 2022. [Online].

- Available: https://web.extension.illinois.edu/apples_sp/varieties.cfm
- [8] F.A. Cheein *et al.*, "Human-robot interaction in precision agriculture: Sharing the workspace with service units," *2015 IEEE International Conference on Industrial Technology (ICIT)*, pp. 289-295, 2015, doi: 10.1109/ICIT.2015.7125113.
- [9] E. Mavridou, E. Vrochidou, G. Papakostas, T. Pachidis, and V. Kaburlasos, "Machine vision systems in precision agriculture for crop farming," *Journal of Imaging*, vol. 5, no. 12, pp. 89, 2019, doi: 10.3390/jimaging5120089.
- [10] A. Kelliher, "Machines learning culture," *IEEE MultiMedia*, vol. 22, no. 2, pp. 18-22, Apr-June 2015, doi: 10.1109/MMUL.2015.43.
- [11] A. Savchenko, A. Galuza, N. Protsay, I. Kolenov, O. Tevyasheva, and O. Tonitsa, "Construction of a linear food quality classifier based on their color analysis", in *2021 11th International Conference on Advanced Computer Information Technologies (ACIT)*, 2021, pp. 5-8, doi: 10.1109/ACIT52158.2021.9548333.
- [12] M. Andruta, E. Sestras, F. Ariana, and M. Mădălina, "Chemometric comparison and classification of 22 apple genotypes based on texture analysis and physico-chemical quality attributes," *Horticulturae*, vol. 8, 2022, doi: <https://doi.org/10.3390/horticulturae8010064>.
- [13] P. Gao *et al.*, "Canopy volume measurement of fruit trees using robotic platform loaded lidar data," *IEEE Access*, vol. 9, pp. 156246-156259, 2021, doi: 10.1109/ACCESS.2021.3127566.
- [14] N. Häni, P. Roy, and V. Isler, "Apple counting using convolutional neural networks", *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2559-2565, doi: 10.1109/IROS.2018.8594304.
- [15] X. Liu, D. Zhao, W. Jia, W. Ji, and Y. Sun, "A detection method for apple fruits based on color and shape features," *IEEE Access*, vol. 7, pp. 67923-67933, 2019, doi: 10.1109/ACCESS.2019.2918313.
- [16] H. Patel, R. Prajapati, and M. Patel, "Detection of quality in orange fruit image using svm classifier," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 74-78, doi: 10.1109/ICOEI.2019.8862758.
- [17] K. Nanaa, M. Rizon, M.N. Abd Rahman, Y. Ibrahim, and A. Z. Abd Aziz, "Detecting mango fruits by using randomized hough transform and backpropagation neural network", in *2014 18th International Conference on Information Visualization*, 2014, pp. 388-391, doi: 10.1109/IV.2014.54.
- [18] C. Zhang, X. Wang, and H. Zhang, "Contrast enhancement for fruit image by gray transform and wavelet neural network," in *2006 IEEE International Conference on Networking, Sensing and Control*, 2006, pp. 1064-1069, doi: 10.1109/ICNSC.2006.1673299.
- [19] S. Kumar *et al.*, "A back-sampling chain technique for accelerated detection, characterization, and reconstruction of radiation-induced transient pulses," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 29, no. 12, pp. 2086-2097, Dic. 2021, doi: 10.1109/TVLSI.2021.3119462.
- [20] K. Roy, S. Chaudhuri, and S. Pramanik, "Deep learning based real-time Industrial framework for rotten and fresh fruit detection using semantic segmentation," *Microsystem Technologies*, vol. 27, no. 9, pp. 3365-3375, 2021, doi: 10.1007/s00542-020-05123-x.
- [21] Q. Liang, J. Long, W. Zhu, Y. Wang, and W. Sun, "Apple recognition based on convolutional neural network framework," in *2018 13th World Congress on Intelligent Control and Automation (WCICA)*, 2018, pp. 1751-1756, doi: 10.1109/WCICA.2018.8630705.
- [22] M. Hossain, M. Al-Hammadi, and G. Muhammad, "Automatic fruit classification using deep learning for industrial applications," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1027-1034-04861, 2018, doi: 10.1109/TII.2018.2875149.
- [23] Y. Li, X. Feng, Y. Liu, and X. Han, "Apple quality identification and classification by image processing based on convolutional neural networks," *Scientific Reports*, vol. 11, no. 1, pp. 1-15, 2021, doi: 10.1038/s41598-021-96103-2.
- [24] A. Kamilaris and F. Prenafeta-Boldú, "Deep learning in agriculture: a survey," *Computers and electronics in agriculture*, vol. 147, pp. 70-90, 2018, doi: 10.1016/j.compag.2018.02.016.
- [25] H. Kang and C. Chen, "Fruit detection, segmentation and 3d visualization of

- environments in apple orchards,” *Computers and Electronics in Agriculture*, vol. 171, pp. 105302, 2020, doi: 10.1016/j.compag.2020.105302.
- [26] A. Howard *et al.*, “Mobilenets: efficient convolutional neural networks for mobile vision applications,” arXiv, 2017. doi: 10.48550/arXiv.1704.04861.
- [27] Y. Yu, S. A. Velastin, and Fei Yin, “Automatic grading of apples based on multi-features and weighted K-means clustering algorithm,” *Information Processing in Agriculture*, vol. 7, pp. 556-565, 2020, doi: 10.1016/j.inpa.2019.11.003.
- [28] J. Lv, F. Wang, Z. Ma, and H. Rong, “Yellow apple recognition method under natural environment,” in *7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2015, pp. 46-49, doi: 10.1109/IHMSC.2015.91.
- [29] S. Bargoti and J. Underwood, “Deep fruit detection in orchards,” in *2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017*, pp. 3626-3633, doi: 10.1109/ICRA. 2017.7989417.
- [30] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, “Deepfruits: a fruit detection system using deep neural networks,” *Sensors*, vol 16. no. 8, pp. 1222, 2016, doi: 10.3390/s16081222.
- [31] X. Ni, C. Li, H. Jiang, and F. Takeda, “Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield,” *Horticulture Research*, vol. 7, 2020, doi: 10.1038/s41438-020-0323-3.
- [32] A. Kuznetsova, T. Maleva, and V. Soloviev, “Using yolov3 algorithm with pre- and post-processing for apple detection in fruit-harvesting robot,” *Agronomy*, vol. 10, no. 7, 2020, Art. No. 1016, doi: 10.3390/agronomy10071016.
- [33] S. Lu, W. Chen, X. Zhang, and M. Karkee, “Canopy-attention-yolov4-based immature/mature apple fruit detection on dense- foliage tree architectures for early crop load estimation,” *Computers and Electronics Agriculture*, vol. 193, pp. 106696. 2022, doi: 10.1016/j.compag.2022.106696.
- [34] F. Gao *et al.*, “Apple detection and counting using real-time video based on deep learning and object tracking,” *Transactions of the Chinese Society of Agricultural Engineering*, 2021, vol. 37, no. 21, pp. 217-224.
- [35] S. Aziz, M. Bilal, M. Khan, and F. Amjad, “Deep learning-based automatic morphological classification of leukocytes using blood smears”, in *2020 International Conference on Electrical, Communication and Computer Engineering*, 2020, pp. 1-5, doi: 10.1109/ICECCE49384.2020.9179246.
- [36] X. Zuo, H. Lin, D. Wang, and Z. Cui, “A method of crop seedling plant segmentation on edge information fusion model,” *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3187825.
- [37] NVIDIA, “Re-training ssd-mobilenet,” github.com, 2019. [Online]. Available: <https://github.com/dusty-nv/jetson-inference/blob/master/docs/pytorch-ssd.md>
- [38] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, “Ssd: single shot multibox detector”, in *European conference on computer vision*, Springer, pp. 21-37, 2016, doi: 10.1007/978-3-319-46448-0_2.
- [39] NVIDIA, “Semantic segmentation with segnet,” github.com, 2019. [Online]. Available: <https://github.com/dusty-nv/jetson-inference/blob/master/docs/segnet-console-2.md>
- [40] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation” in *IEEE conference on computer vision and pattern recognition*, pp. 3431-3440, 2015.
- [41] The MathWorks, “Resnet18”, mathworks.com, 2022. [Online]. Available: <https://www.mathworks.com/help/deeplearning/ref/resnet18.html>