

UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN ELECTRÓNICA Y AUTOMATIZACIÓN, MENCIÓN CONTROL DE PROCESOS

Tema:

"Sistema cuantificador de calidad de cultivo de manzana para monitoreo de la producción utilizando algoritmos de Aprendizaje Profundo con Visión Artificial y Segmentación de Instancias".

Trabajo de Titulación previo a la obtención del Grado Académico de Magister en Electrónica y Automatización, mención Control de Procesos.

Modalidad de titulación: "Proyecto de Desarrollo".

Autor: Ing. Andrés Alejandro Garcés Cadena.

Director: Ing. Álvaro Javier Prado Romo, Ph.D.

Ambato – Ecuador

2022

APROBACIÓN DEL TRABAJO DE TITULACIÓN

A la Unidad de Titulación de Posgrado de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial.

El Tribunal receptor de la Defensa del Trabajo de Titulación presidido por la Ingeniera Elsa Pilar Urrutia Urrutia, Mg., e integrado por los señores: Ingeniero Marlon Antonio Santamaría Villacís, Mg; Ingeniera Ana Pamela Castro Martín, MSc., designados por la Unidad de Titulación de Posgrado de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial de la Universidad Técnica de Ambato, para receptor el Trabajo de Titulación con el tema: SISTEMA CUANTIFICADOR DE CALIDAD DE CULTIVO DE MANZANA PARA MONITOREO DE LA PRODUCCIÓN UTILIZANDO ALGORITMOS DE APRENDIZAJE PROFUNDO CON VISIÓN ARTIFICIAL Y SEGMENTACIÓN DE INSTANCIAS, elaborado y presentado por el señor Ingeniero Andrés Alejandro Garcés Cadena para optar por el Grado Académico de Magíster en Electrónica y Automatización, Mención Control de Procesos una vez escuchada la defensa oral del Trabajo de Titulación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la Universidad Técnica de Ambato.

Ing. Elsa Pilar Urrutia Urrutia Mg.

Presidente y Miembro del Tribunal de Defensa

Ing. Marlon Antonio Santamaría Villacís, Mg.

Miembro de Tribunal de Defensa



Firmado electrónicamente por:
**ANA PAMELA
CASTRO**

Ing. Ana Pamela Castro Martín, MSc.

Miembro de Tribunal de Defensa

AUTORÍA DEL TRABAJO DE TITULACIÓN

La responsabilidad de las opiniones, comentarios, y críticas emitidas en el Trabajo de Titulación presentado con el tema: “SISTEMA CUANTIFICADOR DE CALIDAD DE CULTIVO DE MANZANA PARA MONITOREO DE LA PRODUCCIÓN UTILIZANDO ALGORITMOS DE APRENDIZAJE PROFUNDO CON VISIÓN ARTIFICIAL Y SEGMENTACIÓN DE INSTANCIAS”, le corresponde exclusivamente al Ing. Andrés Alejandro Garcés Cadena, Autor bajo la dirección del Ing. Álvaro Javier Prado Romo, Ph.D., Director del Trabajo de Titulación; y el patrimonio intelectual a la Universidad Técnica de Ambato.



Firmado electrónicamente por:
**ANDRES ALEJANDRO
GARCES CADENA**

Ing. Andrés Alejandro Garcés Cadena.

AUTOR



Firmado electrónicamente por:
**ALVARO
JAVIER PRADO**

Ing. Álvaro Javier Prado Romo, Ph.D.

DIRECTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el Trabajo de Titulación sirva, como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la institución.

Cedo los derechos de mi Trabajo de Titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad Técnica de Ambato.



Firmado electrónicamente por:
**ANDRES ALEJANDRO
GARCES CADENA**

Ing. Andrés Alejandro Garcés Cadena

C.C.: 1804540829

AUTOR

ÍNDICE GENERAL

Contenido

PORTADA	i
APROBACIÓN DEL TRABAJO DE TITULACIÓN	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN	iii
DERECHOS DE AUTOR.....	iv
ÍNDICE DE TABLAS	ix
ÍNDICE DE FIGURAS.....	x
AGRADECIMIENTO	xii
DEDICATORIA	xiii
GLOSARIO.....	xiv
RESUMEN EJECUTIVO	xvi
EXECUTIVE SUMMARY	xviii
1. CAPÍTULO I.....	1
EL PROBLEMA DE INVESTIGACIÓN.....	1
1.1. Introducción.....	1
1.2. Descripción del problema	3
1.3. Preguntas de investigación.....	3
1.4. Solución propuesta.....	4
1.5. Hipótesis.....	4
1.6. Objetivos	4
1.6.1. Objetivo General	4
1.6.2. Objetivos Específicos	4

1.7. Justificación.....	5
2. CAPÍTULO II.....	7
MARCO TEÓRICO Y ESTADO DEL ARTE	7
2.1. Agricultura de Precisión.....	7
2.1.1. Agricultura de Precisión en producción de frutas en el Ecuador.	8
2.2. Inteligencia Artificial enfocada en Visión por Computador aplicada en Agricultura de Precisión	8
2.3. Algoritmos de Visión por Computador.....	10
2.3.1. Redes neuronales profundas (DNN).....	12
2.3.2. Detección de objetos	13
2.3.3. Segmentación semántica	15
3. CAPITULO III	18
MARCO METODOLÓGICO	18
3.1. Enfoque	18
3.2. Ubicación	18
3.3. Equipos y materiales.....	19
3.4. Tipo de investigación.....	23
3.5. Población o muestra.....	23
3.6. Recolección de información.....	25
3.6.1. Adquisición de imágenes	25
3.7. Descripción de métricas	26
4. CAPÍTULO IV	28
RESULTADOS Y DISCUSIONES.....	28
4.1. Sistema de Visión Artificial desarrollado para Agricultura de Precisión en manzanas.....	28

4.1.1 Detección del tipo de manzana	29
4.1.1.1 SSD	30
4.1.1.2 MobileNet	32
4.1.2 Segmentación de la manzana de acuerdo con el nivel de calidad	33
4.1.2.1 FCN ResNet 18.....	34
4.2. Preparación de datos.	36
4.2.1. Preparación de datos para detección.....	36
4.2.1.1. Conjunto de datos para detección – Base de datos 1	38
4.2.1.2. Conjunto de datos para detección – Base de datos 2	39
4.2.1.3. Conjunto de datos para detección – Base de datos 3	39
4.2.2. Preparación de datos para segmentación.	40
4.3. Configuración experimental.....	41
4.3.1. Herramientas de Hardware y Software.....	42
4.3.1.1. Hardware.....	42
4.3.1.2. Software para entrenamiento de la detección.....	44
4.3.1.3. Software para entrenamiento de segmentación.....	45
4.4. Entrenamiento de los sistemas	46
4.4.1. Entrenamiento para detección.....	46
4.4.2. Entrenamiento para segmentación.	47
4.5. Pruebas experimentales.....	50
4.5.1. Pruebas de detección.....	51
4.5.2. Pruebas de segmentación	53
4.6. Resultados	53
4.6.1. Análisis de resultados de la detección del tipo de manzana.....	54

4.6.1.1.	Prueba de Laboratorio 1: Identificación de manzanas en ambiente controlado.....	54
4.6.1.2.	Prueba de Laboratorio 2: Identificación de manzanas en ambiente controlado.....	56
4.6.1.3.	Pruebas de Campo: Identificación de manzanas en ambiente agrícola.....	58
4.6.2	Análisis de resultados de la segmentación semántica de la manzana...	60
4.6.2.1.	Prueba de Campo: Cuantificación de calidad de manzanas en ambiente agrícola.....	60
5.	CAPÍTULO V	63
	CONCLUSIONES Y RECOMENDACIONES.....	63
5.1	Conclusiones	63
5.2	Recomendaciones.....	64
5.3	Trabajo futuro.....	64
5.4	Bibliografía	65

ÍNDICE DE TABLAS

Tabla 1. Comparación de resultados entre modelos para detección.....	14
Tabla 2. Especificaciones de la placa electrónica Jetson Nano.....	21
Tabla 3. Especificaciones del computador.....	22
Tabla 4. Especificaciones de la cámara.....	22
Tabla 5. Variedades de manzanas utilizadas.....	24
Tabla 6. Grados de calidad definidos	25
Tabla 7. Arquitectura MobileNet.....	32
Tabla 8. Variedad de manzanas y cantidad de imágenes en base de datos 1.....	38
Tabla 9. Variedad de manzanas y cantidad de imágenes en base de datos 2.....	39
Tabla 10. Variedad de manzanas y cantidad de imágenes en base de datos 3.....	40
Tabla 11. Cantidad de etiquetas de manzanas por parámetro de calidad.....	40
Tabla 12. Hiperparámetros de la red de entrenamiento	47
Tabla 13. Cantidad de imágenes para entrenamiento y validación.....	48
Tabla 14. Hiperparámetros y características de configuración.....	50
Tabla 15. Matriz de confusión de conjunto de datos de validación 1.....	55
Tabla 16. Matriz de confusión de conjunto de datos de prueba 1.....	56
Tabla 17. Matriz de confusión de conjunto de datos de validación 2.....	57
Tabla 18. Matriz de confusión de conjunto de datos de prueba 2.....	58
Tabla 19. Matriz de confusión de conjunto de datos de validación 3.....	59
Tabla 20. Matriz de confusión de conjunto de datos de prueba 3.....	60
Tabla 21. Análisis de cuantificación de calidad.....	62

ÍNDICE DE FIGURAS

Figura 1. Sectores en donde se aplica la Visión por Computadora	8
Figura 2. Ejemplos de aplicaciones de Visión por Computador.	9
Figura 3. Resultados de detección de varias frutas.	13
Figura 4. Imagen real (a) e imagen segmentada de una flor (b).	15
Figura 5. Segmentación de arándanos de acuerdo con su nivel de madurez	16
Figura 6. Comparación entre AlexNet y ResNet18.	17
Figura 7. Mapa de ubicación del huerto.	18
Figura 8. Imagen panorámica del huerto de manzanas	19
Figura 9. Sistema implementado	20
Figura 10. Equipos utilizados en el sistema: (a) Jetson Nano, (b) Cámara, (c) Monitor	20
Figura 11. Variedades de manzanas.	23
Figura 12. Tipos de manzanas.	24
Figura 13. Manzanas con diferente grado de calidad. (a) Alta calidad, (b) Media calidad y (c) Baja calidad	25
Figura 14. Secciones fundamentales en sistemas de Visión Artificial	28
Figura 15. Proceso de inferencia en tiempo real con Jetson	29
Figura 16. Arquitectura SSD MobileNet	30
Figura 17. Funcionamiento general sistema de detección con SSD MobileNet	30
Figura 18. Arquitectura SSD VGG-16.	31
Figura 19. SSD cuadros delimitadores y mapas de características	32
Figura 20. Ejemplo de resultado de segmentación semántica.	33
Figura 21. Funcionamiento general sistema de segmentación con FCN ResNet18 ...	34
Figura 22. Arquitectura original ResNet18	34
Figura 23. Ejemplo de resultado de segmentación semántica con redes convolucionales	35

Figura 24. Panel de adquisición y etiquetado de imagen	36
Figura 25. Imágenes de los tipos de manzana para cada base de datos	37
Figura 26. Diagrama empleado para adquisición de imágenes de laboratorio.....	38
Figura 27. Panel de etiquetado en el software Labelme	41
Figura 28. Diagrama general del proceso experimental	42
Figura 29. Lugar de ranura para tarjeta microSD	43
Figura 30. Vista frontal y posterior de la carcasa de la Jetson Nano.	43
Figura 31. Vista frontal cámara con el trípode	44
Figura 32 Ejecución de los ensayos de laboratorio	51
Figura 33. Ejecución del sistema en línea para detección de calidad	52
Figura 34. Imágenes de salida con detección de tipo de manzana	52
Figura 35. Imágenes de salida con manzanas segmentadas de acuerdo con grado de calidad.	53
Figura 36. Diagrama de precisión durante el entrenamiento	61
Figura 37. Diagrama de media IoU por cada época del entrenamiento.....	61

DEDICATORIA

A Dios por ser mi fuente inconmensurable de sabiduría.

A mis padres Germán y Anita por siempre estar a mi lado, les entrego el fruto de todo
mi esfuerzo.

A mi esposa Andre por brindarme su amor infinito y apoyo incondicional, este logro
es nuestro.

A mi hermana Pao por estar pendiente de mí en todo momento.

A mi querido Bebé que viene en camino, mi nueva razón para ser mejor.

A mi mascota Scotty por toda su compañía.

A toda mi Familia y Amigos por todo el cariño.

AG7.

GLOSARIO

Agricultura de Precisión (PA) Conjunto de técnicas que mediante el uso de tecnología brindan a los agricultores métodos para monitorear el estado de sus cultivos bajo demanda (Osco, et al., 2020).

Algoritmo Es el conjunto de instrucciones que mediante una secuencia lógica describen la forma para resolver problemas.

Deep Learning (DL) Es el área de la informática que proviene como un subcampo del Machine Learning (ML) y se lo emplea para la resolución de operaciones complejas. Disponen de múltiples capas para el aprendizaje de características a diferentes niveles y cuyo rendimiento va de acuerdo con la cantidad de datos disponibles (Koirala, Walsh, Wang, & McCarthy, 2019).

Entrenamiento Es el proceso de proporcionar información al sistema para la realización de una tarea específica requerida.

Etiquetas Es la clase que se le atribuye a un objeto de una imagen.

Machine Learning (ML) Es el área de informática en la que se otorga a las máquinas la capacidad de aprender sin poseer una programación previa (Koirala, Walsh, Wang, & McCarthy, 2019).

Matriz de confusión Herramienta que permite evaluar el desempeño de un algoritmo de Visión Artificial.

Preprocesamiento Tratamiento digital que se realiza a las imágenes para ajustarlas a una característica requerida.

Redes Neuronales Son una clase de algoritmos de aprendizaje automático que se utilizan para modelar patrones complejos en conjunto de datos.

Segmentación Semántica Es un algoritmo de Deep Learning en el que se relaciona una etiqueta a cada píxel de una imagen.

UNIVERSIDAD TÉCNICA DE AMBATO

**FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E
INDUSTRIAL**

**MAESTRÍA EN ELECTRÓNICA Y AUTOMATIZACIÓN, MENCIÓN
CONTROL DE PROCESOS**

TEMA:

"Sistema cuantificador de calidad de cultivo de manzana para monitoreo de la producción utilizando algoritmos de Aprendizaje Profundo con Visión Artificial y Segmentación de Instancias" .

AUTOR: Andrés Alejandro Garcés Cadena.

DIRECTOR: Ing. Álvaro Javier Prado Romo, Ph.D.

LÍNEA DE INVESTIGACIÓN:

- Tecnología de la información y sistemas de control.

FECHA: Abril 8, 2022

RESUMEN EJECUTIVO

En la actualidad la agricultura es una actividad de marcada influencia en la economía mundial, por lo cual para satisfacer las necesidades progresivas de alimentación el ser humano ha ido introduciendo las herramientas tecnológicas para la optimización de las prácticas agrícolas, esta gestión también es conocida como Agricultura de Precisión (PA).

La Visión Artificial es la tecnología que ha proporcionado un mayor respaldo a la Agricultura de Precisión (PA), otorgando una gama de herramientas que presentan la capacidad de disminuir las dificultades que experimenta el agricultor durante su labor manual.

Con el presente proyecto se busca otorgar una herramienta al agricultor para mejorar el proceso de gestión de cultivos de manzana, mediante el uso de algoritmos de Aprendizaje Profundo (o DL – Deep Learning por sus siglas en inglés) y Visión por Computador.

El desarrollo del sistema presenta dos análisis de estudio: la detección del tipo de manzanas y la cuantificación de la calidad para su inspección y validación empleando un método no invasivo.

Para la detección del tipo de manzana se utilizó el modelo SSD-MobileNet y para la segmentación de acuerdo con el nivel de calidad se usó una red convolucional FCN-ResNet-18. Para los dos estudios se reentrenaron a las redes con bases de datos personalizadas y obtenidas específicamente para el desarrollo de este proyecto.

Finalmente, se presentan los parámetros de evaluación de los sistemas de detección y segmentación con métricas como matrices de confusión, y solapamiento de objetos sobre la IoU respectivamente.

Palabras claves: Visión por computador, Aprendizaje Profundo, Agricultura de Precisión, Redes neuronales convolucionales.

UNIVERSIDAD TÉCNICA DE AMBATO

**FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E
INDUSTRIAL**

**MAESTRÍA EN ELECTRÓNICA Y AUTOMATIZACIÓN, MENCIÓN
CONTROL DE PROCESOS**

THEME:

Apple crop quality quantification system for production monitoring using Deep Learning algorithms with Artificial Vision and Instance Segmentation

AUTHOR: Andrés Alejandro Garcés Cadena.

DIRECTED BY: Ing. Álvaro Javier Prado Romo, Ph.D.

LINE OF RESEARCH:

- Information technologies and control systems.

DATE: April 8th, 2022.

EXECUTIVE SUMMARY

Nowadays, agriculture is an activity of marked influence in the economy world, therefore, in order to satisfy the progressive food needs, human beings have been introducing technological tools for the optimization of agricultural practices, this management is also known as Precision Agriculture (PA).

Artificial Vision is a technology that has given greater support to Precision Agriculture (PA), granting a wide range of tools with the ability to reduce difficulties faced by the farmer during his hand labor.

The aim of this project is to provide farmers a tool to improve the process for apple harvest management, by using Deep Learning (DL) algorithms and a Computer Vision system.

The system development includes two study analyses: apple type detection and quality quantification for its inspection and validation using a non-invasive method.

For apple type detection, SSD-MobileNet model was used and for apple quality segmentation, a fully convolutional network FCN-ResNet-18 was used. For both studies, networks were retrained with customized databases generated specifically for the development of this project.

Lastly, evaluation parameters of the detection and segmentation systems are presented with metrics such as confusion matrices, and overlapping of objects on the IoU, respectively.

Keywords: Computer Vision, Deep Learning, Precision Agriculture, Convolutional Neural Networks.

1. CAPÍTULO I

EL PROBLEMA DE INVESTIGACIÓN

"Sistema cuantificador de calidad de cultivo de manzana para monitoreo de la producción utilizando algoritmos de Aprendizaje Profundo con Visión Artificial y Segmentación de Instancias".

1.1. Introducción

El desarrollo Agricultura de Precisión (PA) en el sector agrario ofrece la oportunidad de manejar de manera eficiente los cultivos (CEPAL, FAO, & IICA, 2020).

La agricultura en la actualidad forma parte del sector primario de la economía mundial teniendo una proyección de incremento del 70% hasta el 2050 debido al crecimiento de la población, por lo cual es necesaria la implementación de prácticas de innovación tecnológica para satisfacer las diferentes y progresivas necesidades de los cultivos.

En el transcurso de los últimos años la PA ha brindado soluciones intrínsecas a la producción de alimentos, fortaleciendo la técnica para detectar, comprender y reaccionar ante inconsistencias que se presentan en los campos de cultivo, adicionalmente las estrategias empleadas permiten la adquisición y el análisis de datos de entornos agrícolas que pueden colaborar en la optimización de las prácticas como el recuento de producto calificado, estimación de rendimiento, detección temprana de enfermedades sobre el cultivo, y clasificación de madurez de la cosecha (Roy, Kislak, Plonski, Luby, & Isler, 2019). La Agricultura de Precisión en Ecuador aún no presenta un desarrollo representativo, debido a que los productores no poseen la información y recursos necesarios para la implementación de estas tecnologías (Villota, 2019). Sin embargo, el hecho que influye directamente en la necesidad de tecnificar la agricultura es el crecimiento de la población acelerado y la influencia del cambio climático que genera incertidumbre al agricultor con respecto a los tiempos de cosecha.

En la actualidad todos los procesos agrícolas deben garantizar una buena calidad para satisfacer los requerimientos de sus clientes, por tal motivo la evaluación del fruto en el árbol es la parte inicial de esta cadena de control y validación. La inspección de calidad del fruto en huertos en los que existe una cantidad representativa de árboles es una tarea compleja para el agricultor y por esto existe un alto riesgo de una evaluación inadecuada y con resultados subjetivos a causa de factores como la repetibilidad de la tarea y al esfuerzo físico que se puede experimentar durante la ejecución constante, tal como lo menciona (Bhargava & Bansal, 2018). Sin embargo, la inspección de productos agrícolas se ha ido fortaleciendo en los últimos años presentando un control eficiente y entregando un producto final competitivo y de buena calidad.

En particular, la estimación del rendimiento de cosecha puede generarse automáticamente con tecnologías de Visión por Computador, Aprendizaje Profundo y de Máquina, debido a que estas tecnologías presentan la capacidad de generar un modelo de pronóstico holístico del ambiente agrícola y su producción. En referencia a lo mencionado se propone realizar un “Sistema cuantificador de calidad de cultivo de manzana para monitoreo de la producción utilizando algoritmos de Aprendizaje Profundo con Visión Artificial y Segmentación de Instancias”. En esta tesis se proyectan 2 contribuciones que emplean técnicas de redes neuronales convolucionales.

1. La primera contribución presenta un método para detectar el tipo de manzana de una imagen de entrada y entrega una imagen en la que se encierra la manzana en un cuadro delimitador junto con su clase y valor de confianza.
2. La segunda contribución presenta un método en la que se recibe una imagen de entrada y se genera una imagen de salida con la superposición de una máscara a nivel píxel que permitirá evidenciar la calidad que presenten las manzanas.

Para el desarrollo del proyecto se dispone de cinco capítulos que se dividen de la siguiente manera:

En el Capítulo I, se describe EL PROBLEMA DE INVESTIGACIÓN y se presenta el análisis de introductorio y justificativo de la investigación, adicionalmente se exponen las hipótesis y objetivos.

En el Capítulo II, se presentan los ANTECEDENTES INVESTIGATIVOS que muestra la información que respalda a la ejecución del proyecto.

En el Capítulo III, se presenta el MARCO METODOLÓGICO en el que se detallan el sitio, los equipos, parámetros para adquisición de información, etc.

En el Capítulo IV, se presentan los RESULTADOS que validan la operación del sistema.

En el Capítulo V, se exponen las CONCLUSIONES Y RECOMENDACIONES obtenidas en base a los objetivos propuestos.

1.2. Descripción del problema

La mayoría de los parámetros para evaluación de calidad de los frutos son visuales y de acuerdo con (Bhargava & Bansal, 2018), al ser una tarea monótona y constante en la cual se deben observar una variada cantidad de frutos por árbol se torna una tarea laboriosa para el agricultor, existiendo el riesgo de que la evaluación presente inconsistencias y con resultados subjetivos.

En Ecuador la mayoría de los pequeños productores no cuentan con las herramientas tecnológicas necesarias que sirvan de apoyo para la tecnificación de los procesos agrícolas (Villota, 2019), lo que genera un alto riesgo de evaluaciones de calidad deficientes, afectando directamente a los ingresos económicos del productor.

1.3. Preguntas de investigación

Luego de haber revisado la descripción del problema se han logrado identificar varios problemas, por lo cual en esta tesis de investigación se direcciona los esfuerzos en las siguientes preguntas de investigación.

¿Es posible detectar e identificar el tipo de manzana analizando su morfología externa y empleando técnicas de Visión Artificial?

¿Es posible cuantificar la calidad de la producción de manzanas mediante identificación, detección, y caracterización del producto haciendo uso de técnicas de Visión Artificial?

1.4. Solución propuesta

La solución se basa en la propuesta de desarrollo de un Sistema cuantificador de calidad de cultivo de manzana para monitoreo de la producción utilizando algoritmos de Aprendizaje Profundo con Visión Artificial en detección y segmentación, que permita la conexión entre la detección del tipo de manzana y la cuantificación de la calidad de dicha producción para generar la conciencia situacional del entorno agrícola, mejorar el proceso productivo, ayudar al agricultor en la toma de decisiones y fomentar el uso particularizado de insumos agrícolas para garantizar la calidad del fruto.

1.5. Hipótesis

Es posible detectar y segmentar imágenes mediante técnicas actuales de Aprendizaje Profundo basado en Visión por Computadora aplicadas sobre diferentes tipos de manzanas con morfología variable que se producen en algunas localidades del Ecuador.

1.6. Objetivos

Los objetivos propuestos que tienen el fin de validar la hipótesis planteada se dividen en general y específicos, los mismos que se detallan a continuación.

1.6.1. Objetivo General

Diseñar un sistema cuantificador de calidad de cultivo de manzana para el pronóstico y diagnóstico de la producción mediante la detección y segmentación de imágenes asociadas al producto en base a técnicas de Aprendizaje Profundo y Visión por Computador.

1.6.2. Objetivos Específicos

- Investigar dentro del estado del arte acerca de las técnicas actuales de Visión Artificial en detección y segmentación de objetos.

- Desarrollar una estrategia de evaluación en tiempo real de producción de manzanas mediante un modelo de estimación y diagnóstico del conteo de fruto de calidad.
- Implementar un marco de estimación automatizado de la cantidad de producto de manzana calificada rojas, verdes y similares bajo condiciones reales de operación.
- Evaluar mediante métricas de desempeño los modelos de Inteligencia Artificial probados para diferentes condiciones ambientales en el que está expuesta la fruta.
- Reportar los resultados de la detección y la segmentación de la manzana en documentos de carácter científico.

1.7. Justificación

El control y la gestión eficientes de las explotaciones agrícolas son esenciales para hacer frente a los requerimientos de la creciente población y a sus múltiples problemas económicos. Las técnicas de Agricultura de Precisión (PA) brindan a los agricultores métodos para monitorear a pie y mano el estado de sus cultivos bajo demanda. Para muchos cultivos elementales (arroz, maíz, trigo), estas técnicas han logrado un desarrollo como para proporcionar mapas bajo demanda e implementar tareas decisivas (Osco, et al., 2020). Sin embargo, para cultivos especiales (como frutas y flores), estas técnicas aún están evolucionando en el mundo e iniciando en los países en vías de desarrollo (Hani, Roy, & Isler, 2020).

La morfología y topología (en términos de variabilidad en el tamaño, color, forma, textura, etc.) del fruto más el ambiente no estructurado dificulta el desarrollo de un método de mapeo de rendimiento general, por lo que constituye un reto para los investigadores que alternativamente desarrollan sistemas de evaluación de rendimiento para cada cultivo (Roy, Kislay, Plonski, Luby, & Isler, 2019). Por lo anterior, el esfuerzo descrito en esta propuesta de tesis se concentra en la obtención de un sistema de evaluación de la calidad en la que se encuentra el producto a ser cosechado.

En Ecuador la aplicación de tecnologías que permitan la optimización y mejora de los procesos agrícolas es limitada, especialmente en el campo de la pequeña producción. Cabe señalar que la mayoría de los productores poseen una superficie de producción mínima y resulta complejo asumir el costo de tecnificar los cultivos (Villota, 2019). Por este motivo es importante asegurar la óptima producción de la fruta considerando todos los detalles de calidad, para incrementar las ganancias y así otorgar estabilidad financiera a los productores. Existe una actividad reciente y significativa para automatizar la estimación del rendimiento en plantaciones de manzanas por su alto impacto en el incremento de la producción (Wang, Nuske, Bergerman, & Singh, 2013).

Algunos de los trabajos se centran en sistemas de estimación de rendimiento de huertos completos, por ejemplo, en (Wang, Nuske, Bergerman, & Singh, 2013) los autores presentaron un sistema para estimación de producción de manzanas para la totalidad de un huerto. En (Hung, Underwood, Nieto, & Sukkarieh, 2015) se presenta un enfoque de aprendizaje basado en características para la identificación solo de manzanas rojas y verde. En (Das, et al., 2015; Das, et al., 2015) desarrollaron un sistema multisensorial para extraer la morfología planar de las manzanas. Sin embargo, el sistema pudiera ser económicamente prohibitivo para las condiciones agrícolas locales. A diferencia de estos sistemas, el trabajo propuesto consiste en desarrollar un sistema de bajo costo para extraer las características de la manzana que permitan determinar su tipo y calidad.

Con la implementación del Sistema cuantificador de calidad de cultivo de manzana para monitoreo de la producción utilizando algoritmos de Aprendizaje Profundo con Visión Artificial en detección y segmentación, se dotará al agricultor de la herramienta tecnológica para apoyo en sus habilidades de gestión y toma de decisiones, contribuyendo paralelamente a la reducción del esfuerzo físico, el costo de la mano obra y optimización de la calidad de inspección (Bhargava & Bansal, 2018).

2. CAPÍTULO II

MARCO TEÓRICO Y ESTADO DEL ARTE

Para el desarrollo adecuado y direccionado del sistema se ha realizado la recopilación y el análisis de información de los trabajos de investigación que presentan una relación directa con el tema planteado. En esta sección se presentará el detalle más relevante de la información que presente relación con el tema propuesto.

2.1. Agricultura de Precisión

La agricultura en los últimos tiempos se ha convertido en un pilar fundamental para la economía de todos los países. Con el crecimiento acelerado de la población también se ha generado el desafío de satisfacer sus necesidades alimentarias. En la actividad agrícola se emplean varias herramientas manuales con el objetivo producir múltiples alimentos necesarios para la nutrición del ser humano. Sin embargo, las prácticas tradicionales presentan riesgos constantes para la producción y rendimiento de los cultivos, para ello las nuevas tendencias tecnológicas aplicadas a la agricultura son fundamentales para el aumento de la producción, reducción del impacto ambiental e incremento de los ingresos económicos generados durante prácticas agrícolas (Roy, Kislay, Plonski, Luby, & Isler, 2019).

El impacto de la tecnología en la agricultura ha impulsado la productividad agrícola y minimizado la degradación del medio ambiente. Los datos obtenidos en las operaciones agrícolas actuales se obtienen a través de una gama de sensores que entregan información precisa de las condiciones ambientales y de los factores que intervienen en el proceso de cultivo, lo que permite agilizar la respuesta ante cualquier evento anormal que se presente durante el proceso agrícola. La Agricultura de Precisión (PA) ha surgido como un conjunto de técnicas de última tecnología para mejorar la productividad y compensar los desafíos alimentarios actuales de la población. Con el uso de esta tecnología el agricultor estará al tanto con precisión de los parámetros que se requieren para mantener un cultivo saludable.

2.1.1. Agricultura de Precisión en producción de frutas en el Ecuador.

En el Ecuador el desarrollo tecnológico en la agricultura se ha visto afectado debido a que las pequeñas y medianas empresas (Pymes) no disponen de la información y las herramientas tecnológicas necesarias, lo que limita el desarrollo de los procesos en las plantaciones agrícolas (Carpio, 2018). Sin embargo, es necesario mantener una buena calidad nutricional empleando técnicas agrícolas sostenibles.

2.2. Inteligencia Artificial enfocada en Visión por Computador aplicada en Agricultura de Precisión

El ser humano en los últimos años ha desarrollado avances en el área de Inteligencia Artificial (IA) dotando a sistemas computacionales de ciertas capacidades como: Visión Artificial, reconocimiento de voz y procesamiento de lenguaje natural. La Visión Artificial destaca de entre las 3 capacidades debido a que ofrece mayor ventaja en el desarrollo de aplicaciones para el uso cotidiano en varios sectores (industria, medicina, agricultura, seguridad, robótica, etc.), como se observa en la Figura 1 (Rouhiainen, 2018).



Figura 1. Sectores en donde se aplica la Visión por Computadora (NVIDIA, 2022).

La Visión por Computadora emplea imágenes para analizar y entender la operación de cualquier proceso físico, en la actualidad esta tecnología se ha utilizado considerablemente para respaldar las tareas de Agricultura de Precisión, entre ellas el

monitoreo y evaluación de cultivos, el mapeo de rendimiento, control de maleza, estado de la cosecha, entre otras. Las imágenes que más se utilizan en un sistema de Visión por Computador son de color (RGB) y simulan un proceso de inspección visual por el ser humano, para esto las imágenes adquiridas por las cámaras de color RGB se centran en el espacio de color. (Bhargava & Bansal, 2018).

Varias características como la textura, la forma, el color, el tamaño y los defectos se los pueden detectar, clasificar y segmentar con el uso de sistemas de Visión por Computadora como se observa en la Figura 2. Adicionalmente, se manejan imágenes multiespectrales (hiperespectrales) de infrarrojo cercano y sensores de rango para mediciones geométricas (Lu, & Young, 2020).

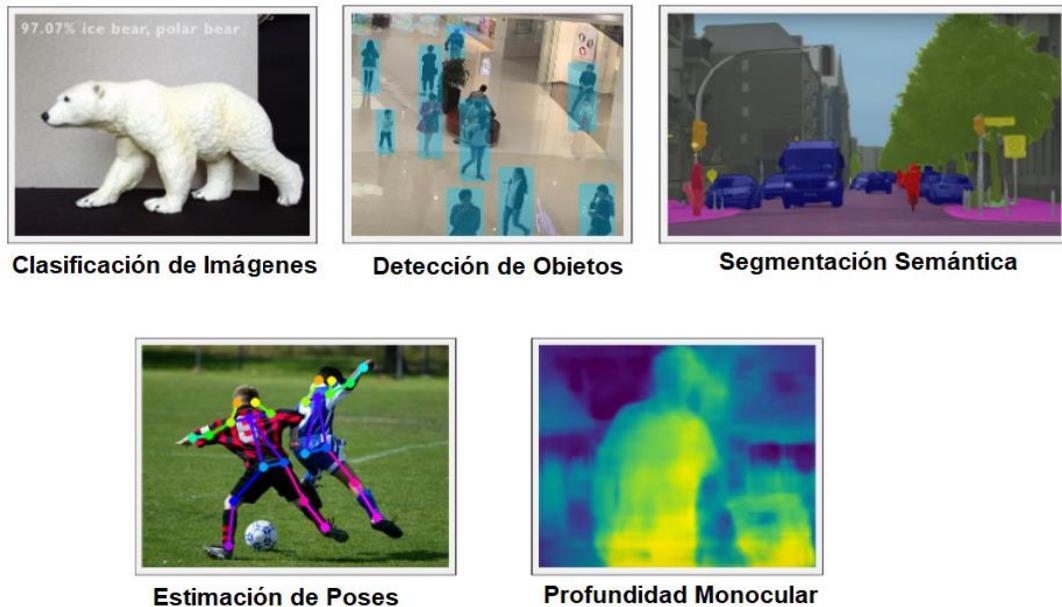


Figura 2. Ejemplos de aplicaciones de Visión por Computador (NVIDIA, 2019).

Según (Mavridou, Vrochidou, Papakostas, Pachidis, & Kaburlasos, 2019) afirma que con la tecnología de Visión por Computador es posible realizar la detección de frutos de manera eficaz, acelerada y no invasiva. También hacen mención de que, debido a la complejidad de separar la planta de fondo y la fruta, el método apropiado para realizar un control de calidad en el fruto es el reconocimiento de imágenes mediante segmentación semántica.

2.3. Algoritmos de Visión por Computador

En la actualidad el análisis de imágenes es un área de investigación de gran importancia en el proceso agrícola y las técnicas inteligentes. Normalmente se están utilizando en identificación, clasificación y detección de frutos que aportan en el análisis de calidad, madurez, anomalías, etc. Las técnicas que se utilizan para analizar imágenes son: el Aprendizaje Automático (ML) K-means, Máquinas de Vectores de Soporte (SVM), Redes Neuronales Artificiales (ANN), polarizaciones lineales, filtrado basado en ondículas, índices de vegetación (NDVI) y análisis de regresión; a estas se adiciona el Aprendizaje Profundo (DL) que se tratan de redes neuronales “con mayor profundidad” y otorga mayores capacidades de aprendizaje al sistema. El autor realiza una comparación de rendimiento entre las técnicas y revela que el Aprendizaje Profundo (DL) ofrece una mejor respuesta con respecto a las técnicas expuestas. (Kamilaris & Prenafeta, 2018).

En (Bhargava & Bansal, 2018) se presenta una comparación de eficiencia de diferentes técnicas de detección de defectos para la evaluación de calidad de frutas y hortalizas, adicionalmente se verifican segmentaciones, características de imagen, descriptores y análisis de calidad de frutas y verduras sobre la base de color, forma, tamaño, textura y el tipo de daño existente. Aquí se indica que las Redes Naturales Artificiales (ANN) presentan un 96.47% de eficiencia en la evaluación de calidad de frutas y hortalizas.

Las ANN son herramientas de modelado computacional muy utilizadas en varias disciplinas para modelar problemas complejos del mundo real. Pueden ser definidas como estructuras compuestas de elementos de procesamiento simples adaptativos densamente interconectados estos elementos son llamados neuronas artificiales o nodos según (Basheer & Hajmeer, 2000). Estas ANN tienen muchas ventajas para implementarse en modelado de sistemas predictivos complejos como se detalla en (Basir, Chowdhury, Islam, & Ashik-E-Rabbani, 2021), estas ventajas son: capacidad de aproximación de funciones universales, resistencia a datos con ruido o faltantes,

ajuste de múltiples variables no lineales con interacciones desconocidas y buena capacidad de generalización.

El modelado utilizando ANN se utilizó con éxito en los problemas de predicción en ingeniería de bioprocesos y Agricultura de Precisión (Basir, Chowdhury, Islam, & Ashik-E-Rabbani, 2021). Además de esto, también se informaron varios estudios sobre modelado utilizando redes neuronales para la predicción de rendimiento de diferentes cultivos basado en parámetros de monitoreo de rendimiento. Por ejemplo, en (Hossain, Uddin, Hossain, & Jang, 2017) desarrollaron un sistema de predicción basado en el clima para el pronóstico del rendimiento del arroz en distintos territorios de Bangladesh. En el territorio local estos sistemas beneficiarían ampliamente debido a que mejoran la productividad y optimizan el recurso de los agricultores.

La agricultura moderna necesita manejar una alta eficiencia en la producción combinada con una alta calidad de los productos obtenidos. Para reunir estos requisitos, se utilizan métodos avanzados de análisis de datos cada vez con más frecuencia, incluidos los derivados de métodos de inteligencia artificial, de estos métodos las redes neuronales artificiales (ANN) son una de las herramientas más populares de este tipo. Son muy utilizados para resolver diversas tareas de clasificación y predicción. Además, pueden formar parte de la Agricultura de Precisión y sistemas de soporte de decisión (Kujawa & Niedbała, 2021). Las ANN están divididas en dos categorías: ANN profundas y ANN tradicionales. El cerebro humano es la base de inspiración para este tipo de redes, emulando funciones cognitivas superiores como el aprendizaje, la generación de patrones y la toma de decisiones. Las ANN profundas también se conocen como redes neuronales profundas (DNN) o Aprendizaje Profundo (DL). (Liakos, Busato, Moshou, Pearson, & Bochtis, 2018).

El Aprendizaje Profundo (DL) emplea modelos más complejos que el Aprendizaje de Máquina (ML), lo que favorece a la resolución de problemas complejos de manera rápida y con mayor precisión, cabe indicar también que posee varios componentes diferentes según la arquitectura de red utilizada. Las aplicaciones de mayor

intervención del Aprendizaje Profundo son (DL) para tipos de datos como imagen y video. Sin embargo, también es posible utilizar datos, audio, etc. (Kamilaris & Prenafeta, 2018). Las técnicas de Aprendizaje Profundo (DL) han presentado resultados alentadores para beneficio de varios dominios, y de acuerdo con lo publicado por (Kamilaris & Prenafeta, 2018) estas técnicas también se las están aplicando al ámbito agrícola.

2.3.1. Redes neuronales profundas (DNN)

Las redes neuronales profundas DNN presentan la capacidad de aprender características de un nivel con mayor complejidad que las redes neuronales menos profundas (Kamilaris & Prenafeta, 2018), las redes neuronales profundas presentan como un tipo específico para la Visión por Computadora a las Redes Neuronales Convolucionales (CNN) y se las utiliza para procesamiento de datos a nivel píxel. Los investigadores recomiendan emplear arquitecturas populares para la construcción de los nuevos modelos, entre ellas se tienen: AlexNet, CaffeNet, VGG, ResNet, etc. (Kamilaris & Prenafeta, 2018).

El desarrollo de algoritmos de Aprendizaje Profundo es más reciente, en comparación con los algoritmos tradicionales basados en aprendizaje automático como se detalla en (Kang & Chen, 2020), los algoritmos basados en Aprendizaje Profundo han demostrado una mayor precisión en la detección y segmentación. Según (Chen, et al., 2017) se indica que mediante las técnicas de Aprendizaje Profundo “Deep Learning” es posible trabajar con las condiciones de los objetos y entornos no estructuradas en la que los métodos tradicionales presentan dificultades.

Según (Hossain, Al-Hammadi, & Muhammad, 2019), los modelos de Aprendizaje Profundo se han utilizado de forma intensiva para la ingeniería en extracción de características de una manera automática. De estos modelos, destacan las redes neuronales convolucionales (CNN), que están conformadas de varias capas neuronales feedforward, las que son capaces de aprender tareas específicas sobre características invariantes de una manera jerárquica. Estos modelos mostraron robustez y se volvieron una solución para hacer frente al rápido crecimiento de los datos a analizar. Estos

modelos lograron una precisión de 96.75% en diferentes enfoques como la detección e identificación de objetos. Por estas razones, (Hossain, Al-Hammadi, & Muhammad, 2019) alienta a emplear el Aprendizaje Profundo en aplicaciones de Agricultura de Precisión como clasificación de frutas.

2.3.2. Detección de objetos

De acuerdo a como lo indica (Hossain, Al-Hammadi, & Muhammad, 2019), la clasificación de frutos es una tarea de apoyo tanto en el plano agrícola o comercial al ser humano, por lo que proponen un sistema clasificador de frutos empleando técnicas de Aprendizaje Profundo y Visión por Computador como se muestra en la Figura 3. Este sistema presenta la posibilidad de detectar y clasificar distintos frutos encontrando como dificultades principales la similitud de forma y color. En este trabajo se realizaron entrenamientos con conjuntos de imágenes propias y obtenidas en la web. Adicionalmente, se emplearon métricas de evaluación como matriz de confusión y precisión del entrenamiento, las cuales emplearemos para la determinación de resultados del proyecto propuesto.



Figura 3. Resultados de detección de varias frutas. (Sa, et al., 2016)

En el artículo de (Sa, et al., 2016) se presenta un sistema para detección de fruta que mediante la utilización de redes convolucionales profundas genera un modelo preciso

y de alta confiabilidad basado en un detector de objetos CNN. Para el entrenamiento del modelo se utilizaron 484 imágenes y para la prueba se utilizaron 118 imágenes distribuidas para 6 clases diferentes de frutas. Para el desarrollo del sistema, se realizó un ajuste en la red VGG 16 en función del modelo pre-entrenado conocido como ImageNet. Como resultados se tiene que la precisión es de 57% y con sensibilidad de 80%. Las condiciones empleadas no se especifican en el documento. Sin embargo, se menciona que los escenarios al aire libre pueden generar inconvenientes en la detección, producto de los cambios continuos en iluminación, oclusión y sombra.

En el artículo (Bargoti & Underwood, 2017) se propone la detección de frutas en huertos, basado en un detector Faster R-CNN. Para el entrenamiento del modelo se utilizaron 2268 imágenes, para la validación y la prueba se utilizaron 482 imágenes para cada una distribuidas en 3 tipos diferentes de frutas. Los resultados finales que presenta este documento son 95.8% de precisión y 86.3% de sensibilidad al emplear la red VGG16 con el modelo pre-entrenado ImageNet. Es muy importante indicar que el autor resalta que en imágenes capturadas en escenas al aire libre se presentan variaciones representativas en los resultados debido a los cambios constantes de iluminación, ángulo de la cámara y otros parámetros. Las imágenes para evaluación se las realizó en horario matutino en ciudades de Australia. (Bargoti & Underwood, 2017).

En la Tabla 1, se muestra la comparación de resultados del modelo MobileNet con los resultados de otros modelos, la evaluación principal se la realizó con la métrica principal del conjunto COCO (Howard, et al., 2017).

Tabla 1. Comparación de resultados entre modelos para detección (Howard, et al., 2017).

Framework Resolution	Model	mAP	Billion Mult-Adds	Million Parameters
SSD 300	deeplab-VGG	21.1%	34.9	33.1
	Inception V2	22.0%	3.8	13.7
	MobileNet	19.3%	1.2	6.8
Faster-RCNN 300	VGG	22.9%	64.3	138.5
	Inception V2	15.4%	118.2	13.3
	MobileNet	16.4%	25.2	6.1
Faster-RCNN 600	VGG	25.7%	149.6	138.5
	Inception V2	21.9%	129.6	13.3
	Mobilenet	19.8%	30.5	6.1

2.3.3. Segmentación semántica

El proceso de segmentación consiste en la evaluación de una descripción semántica de una imagen a nivel de píxel, en lugar de la imagen completa (Bargoti & Underwood, 2017). Por ejemplo, en la Figura 4(a) la imagen de entrada consiste en una flor amarilla con pequeñas hojas y rama verdes sobre un fondo oscuro. En la Figura 4(b), se observa la segmentación ideal que agrupa los píxeles por los objetos semánticos: todos los píxeles que presentan un color amarillo corresponden a los pétalos de la flor, los píxeles de color celeste corresponden a las hojas y ramas, y el color amarillo corresponde al fondo.

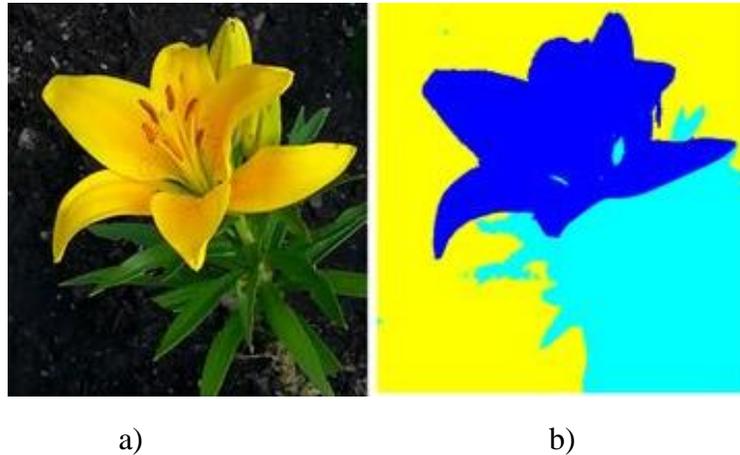


Figura 4. Imagen real (a) e imagen segmentada de una flor (b) (MathWorks, 2022).

En el trabajo desarrollado por (Hung, Underwood, Nieto, & Sukkarieh, 2015) se emplea un algoritmo para segmentación de múltiples imágenes enfocadas a una aplicación de manzanas. El algoritmo que presenta este documento destaca la factibilidad del trabajo en condiciones con iluminación natural y la posibilidad de detectar manzanas de color rojo y verde.

En la investigación realizada por (Ni, Li, Jiang, & Takeda, 2020) se desarrolló un sistema para medir la madurez de arándanos, utilizando el método de segmentación de imágenes con técnicas de Aprendizaje Profundo empleando el modelo Mask R-CNN para la detección y segmentación del fruto como se observa en la Figura 5. La precisión media del modelo fue del 78.3%. Este estudio permite identificar que la técnica de

segmentación de imágenes permite extraer rasgos del fruto para el monitoreo y evaluación de la cosecha.

Como se observa en la Figura 5 los resultados visiblemente son favorables manteniendo condiciones de iluminación controladas, Sin embargo, de acuerdo con lo indicado por (Bargoti & Underwood, 2017), las condiciones al aire libre presentan variaciones en los resultados debido a los factores de iluminación y oclusión que interfieren en el proceso de ejecución del sistema.



Figura 5. Segmentación de arándanos de acuerdo con su nivel de madurez (Ni, Li, Jiang, & Takeda, 2020)

En (Aziz, Bilal, Khan, & Amjad, 2020) se realiza una comparativa del funcionamiento de dos arquitecturas de redes artificiales de las más utilizadas para resolver problemas de clasificación de imágenes a nivel de píxel como son AlexNet y ResNet-18, como se observa en la Figura 6.

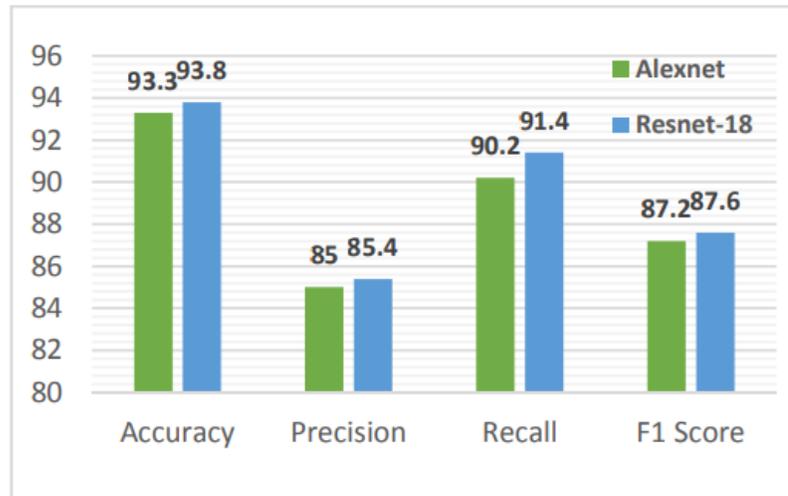


Figura 6. Comparación entre AlexNet y ResNet18. (Aziz, Bilal, Khan, & Amjad, 2020)

Se puede observar un mejor desempeño al utilizar ResNet-18, en este tipo de aplicaciones para estas arquitecturas. Lo cual presenta una referencia para utilizar esta arquitectura para segmentación de imágenes (Aziz, Bilal, Khan, & Amjad, 2020).

3. CAPITULO III

MARCO METODOLÓGICO

En el presente capítulo se describen los parámetros que direccionan la investigación, los equipos y materiales necesarios para la implementación, el método de recolección de datos y las métricas para la evaluación final.

3.1. Enfoque

La modalidad de la presente investigación es de tipo cuantitativa, debido a que con los resultados de los ensayos de laboratorio y de campo evalúan el desempeño del sistema propuesto y la identificación del comportamiento de las variables.

3.2. Ubicación

Los primeros acercamientos se los realizó en un huerto con árboles de manzanas ubicado en la parroquia Santa Rosa de la ciudad de Ambato. La ubicación exacta se muestra en la Figura 7.

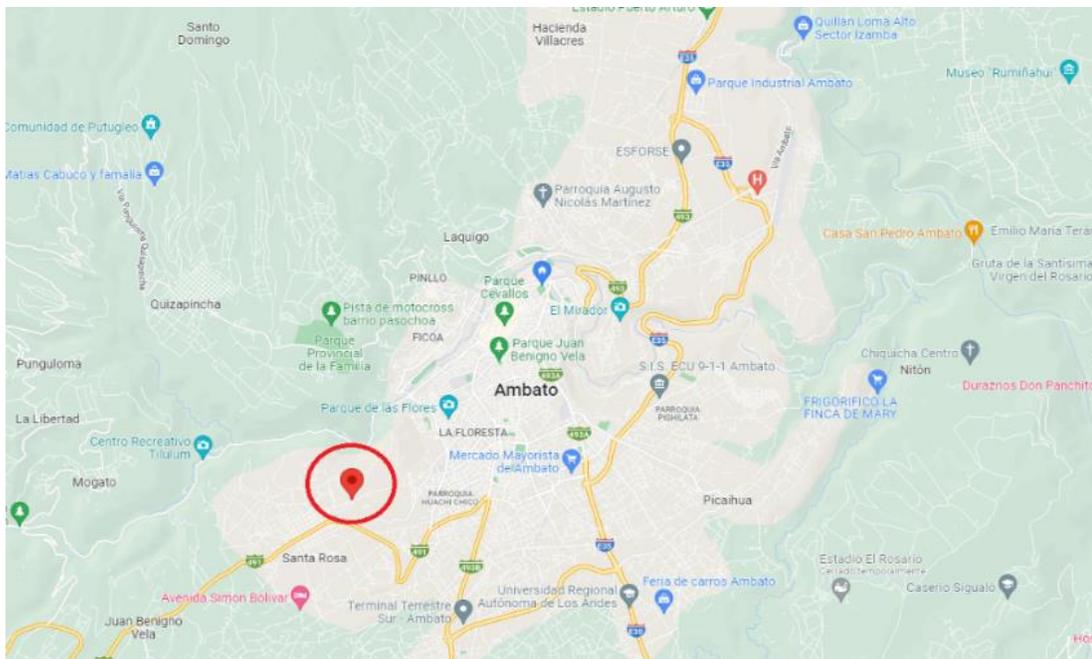


Figura 7. Mapa de ubicación del huerto (GOOGLE, 2022).

En la Figura 8 se muestra la imagen del huerto en donde únicamente se disponía de árboles con un solo tipo de manzana roja denominada Royal Gala y con diferente grado de calidad, por lo que las imágenes posteriormente capturaron en imágenes y se emplearon para el proceso de cuantificación de calidad.



Figura 8. Imagen panorámica del huerto de manzanas (Autor).

3.3. Equipos y materiales

Para la implementación del sistema que maneja imágenes estáticas y video en tiempo real para detectar el tipo de manzana y segmentar el fruto de acuerdo con su grado de calidad, fue necesario utilizar varias herramientas de hardware y software. El sistema propuesto consta principalmente de la placa electrónica Jetson Nano con sus dispositivos de entrada y salida como teclado, monitor y cámara, como se observa en la Figura 9.

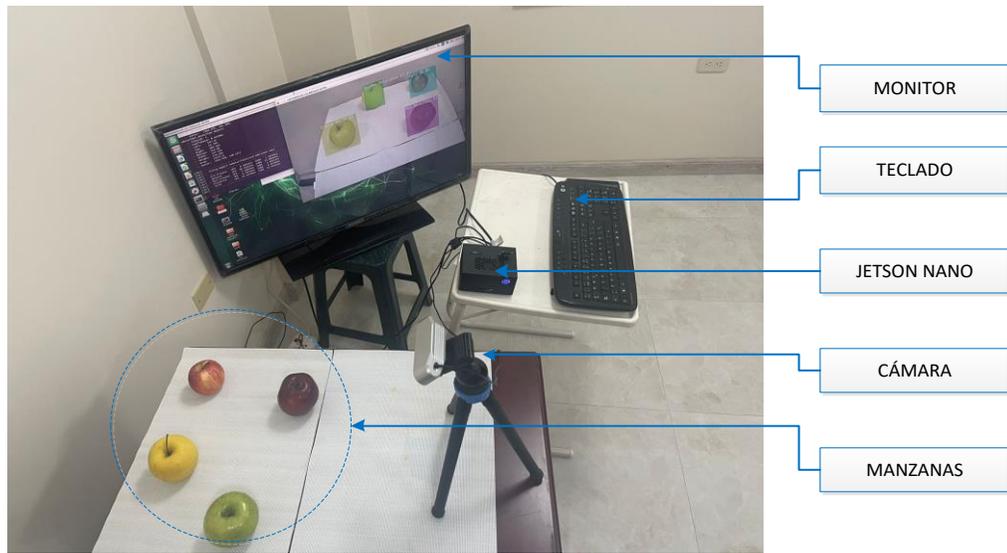


Figura 9. Sistema implementado (Autor).

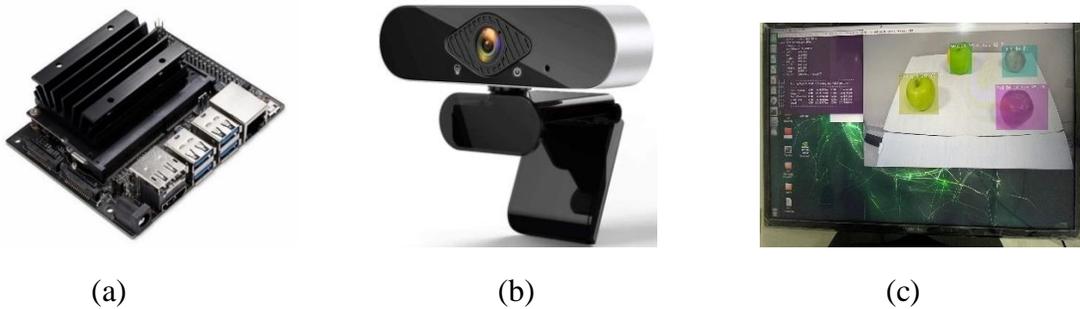


Figura 10. Equipos utilizados en el sistema: (a) Jetson Nano, (b) Cámara, (c) Monitor (NVIDIA, 2022).

El detalle de los equipos utilizados para la ejecución del proyecto se describe a continuación.

- **Placa electrónica Jetson Nano Developer Kit (a)**

Esta placa electrónica está direccionada a la gestión de proyectos de IA, entre las que destacan la Visión por Computador y Redes Neuronales, este dispositivo se muestra en la Figura 10(a).

Una de las ventajas que presta este hardware es la fácil portabilidad debido a su reducido tamaño, esto permite que se la pueda trasladar de manera simple a cualquier

lugar. A continuación en la Tabla 2, se presentan las especificaciones técnicas del ordenador.

Tabla 2. Especificaciones de la placa electrónica Jetson Nano (NVIDIA, 2022).

PARÁMETRO	CARACTERÍSTICA
GPU	128-core NVIDIA Maxwell™
	Quad-core ARM A57 @ 1.43 GHz
Memoria	4GB 64-bit LPDDR4 25.6GB/s
Almacenamiento	microSD (Card not included)
Video encode	4K @ 30 4x 1080 @ 30 9x 720 @ 30 (H.264/H.265)
Video decode	4K @ 60 2x 4K @ 30 8x 1080 @30 18x 720 @ 30 (H.264/H.265)
USB	4x USB 3.0, 1x USB 2.0 Micro-B
Mecanismo	100 mm x 80 mm x 29 mm

Adicionalmente, como accesorios los accesorios que complementaron la operación de la placa electrónica fueron el monitor (c).

- **Computador**

Para agilizar el proceso de entrenamiento para segmentación se utilizó un computador

En la Tabla 3 se muestran las características operativas del computador portátil empleado para entrenamiento del modelo de segmentación semántica.

Tabla 3. Especificaciones del computador (Autor).

PARÁMETRO	CARACTERÍSTICA
Procesador	Intel(R) Core(TM) i7-8750H CPU @ 2.20GHz 2.21 GHz
Memoria RAM	16,0 GB (15,8 GB utilizable)
Tipo de sistema operativo	Sistema operativo de 64 bits, procesador x64.
Tarjeta gráfica	NVIDIA GeForce RTX 2070 with Max-Q Design

- **Cámara (b)**

Para la adquisición de las imágenes se acogió a la recomendación del fabricante y se optó por la selección de una cámara web mostrada en la Figura 10(b), este dispositivo presenta compatibilidad con la placa electrónica Jetson Nano y presenta las características adecuadas para utilización en proyectos de Visión por Computador. (NVIDIA, 2022).

En la Tabla 4 se muestran las principales características de la cámara web empleada para el desarrollo del proyecto.

Tabla 4. Especificaciones de la cámara (Autor).

PARÁMETRO	CARACTERÍSTICA
Video	Resolución: 1.920 x 1.080 píxeles
	FPS: 30
Interfaz	USB 2.0
Campo de visión	Horizontal: 120 grados.
	Vertical: 120 grados.

3.4. Tipo de investigación

El tipo de investigación a emplear es exploratorio, debido a que la información existente es limitada, por lo que fue necesario realizar un reconocimiento inicial del proceso y método empleado de manera directa en el campo y con ello plantear el análisis y propuesta correspondiente.

3.5. Población o muestra

Para la detección del tipo de manzana se adquirieron en un supermercado 5 variedades de entre las cuales se tienen:

- Fuji, Red Delicious, Royal Gala, Grand Smith y Golden Delicious (University of Illinois Board of Trustees, 2022).



Figura 11. Variedades de manzanas.

Para escoger las variedades de manzanas a evaluar en el proyecto se identificó a las que predominan en la producción a nivel mundial publicado con (YARA Knowledge grows, 2022), de las cuales todas se producen en Ecuador.

Tabla 5. Variedades de manzanas utilizadas (YARA Knowledge grows, 2022).

TIPO DE MANZANA	COLOR Y FORMA
Granny Smith (a)	Es de color verde uniforme e intenso, con forma alargada y simétrica
Golden Delicious (b)	Es de color amarillo oro, con tendencia a verde, con forma alargada y redondeada.
Fuji (c)	Es de color rojo y rosa con matices amarillos, con forma redonda.
Royal Gala (d)	Es de color rojo y con matices anaranjados, con forma alargada y simétrica
Red Delicious (e)	Es cónica, alargada y de un tamaño medio con un color rojo oscuro

En la Figura 12 se observan los 5 tipos de manzanas que se utilizaron para el proceso de detección, la información de cada una se la detalla en la Tabla 5.

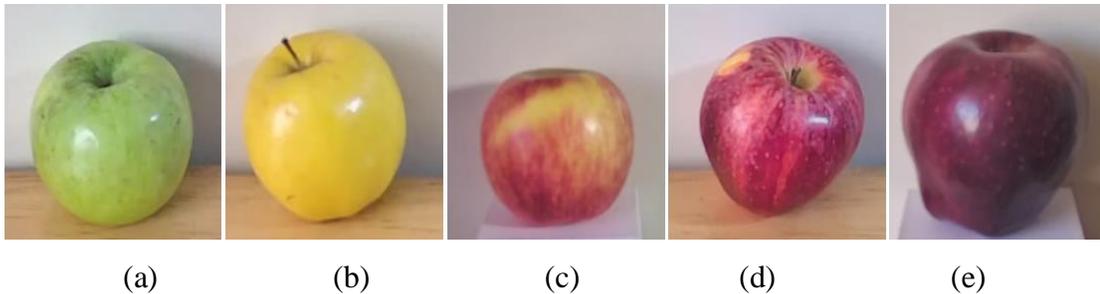


Figura 12. Tipos de manzanas.

Para la segmentación de la cuantificación de calidad de las manzanas la población corresponde a manzanas de tipo “Royal Gala” con diferente grado de calidad ubicadas en un huerto que se indica en la Subsección 3.2.

Tomando como guía el criterio expuesto por (Li, Feng, Liu, & Han, 2021) las categorías para la cuantificación de calidad de las manzanas se las definieron en las categorías expuestas en la Tabla 6.

Tabla 6. Grados de calidad definidos (Autor).

GRADO DE CALIDAD	DETALLE
Alta	No presenta defectos.
Media	Presenta defectos leves y se encuentra apta para el consumo.
Baja	Presenta descomposición que imposibilitan su consumo.

Para el reforzar el criterio del grado de calidad de manzanas en la Figura 13 se muestran tres manzanas con los grados de calidad que se utilizados en el desarrollo del proyecto y descritos en la Tabla 6.



(a) Alta calidad.

(b) Media calidad.

(c) Baja calidad.

Figura 13. Manzanas con diferente grado de calidad. (a) Alta calidad, (b) Media calidad y (c) Baja calidad (Autor).

3.6. Recolección de información

Las bases de datos de los conjuntos de imágenes se los obtuvo durante el mes de marzo del 2021 en horario vespertino (15:00) y específicamente para la ejecución del presente proyecto. El clima que presentó el sector durante el periodo de recopilación de imágenes fue de 19°C y nublado.

3.6.1. Adquisición de imágenes

Las manzanas utilizadas para la detección del tipo fueron adquiridas en un supermercado debido a la ausencia del fruto en la ciudad. Posteriormente se capturaron

5125 imágenes que se distribuyeron en conjuntos para entrenamiento, validación y prueba de los ensayos de laboratorio y campo.

Las imágenes de manzanas que se usaron para el desarrollo de la segmentación para la cuantificación de calidad fueron 338 imágenes, que contienen 2211 manzanas del tipo Royal Gala y con un grado de calidad variable. Las fotografías fueron capturadas durante el mes de marzo del 2021, cuando la fruta estaba madura y a vísperas de su recolección.

Los dispositivos que se utilizaron para la captura de imágenes fueron los siguientes:

- Cámara de teléfono móvil (Iphone 11 Pro Max) la cual dispone de una resolución de 3024 x 4032 píxeles.
- Cámara de fotos Canon PowerShot SX400 IS la cual dispone de una resolución de 3456 x 3456 píxeles.

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, mientras que para las pruebas de campo las condiciones de iluminación fueron naturales en horarios matutinos y vespertinos.

3.7. Descripción de métricas

Existen algunas métricas que sirven para la determinación del rendimiento del modelo en las aplicaciones de detección de frutas en árboles.

Para la detección del tipo de manzana se emplearán matrices de confusión para todos los casos experimentados, mediante la cual se determinará la precisión y sensibilidad del modelo. (Koirala, Walsh, Wang, & McCarthy, 2019).

$$Precisión = \frac{TP}{TP + FP} \quad Eq. 1$$

$$Sensibilidad = \frac{TP}{TP + FN} \quad Eq. 2$$

En donde:

- TP: Número de verdaderos positivos. En este caso el valor real es positivo y la predicción determina que es positivo.
- FP: Número de falsos positivos. En este caso el valor real es negativo y la predicción indica que es positivo.
- FN: Número falsos negativos. En este caso el valor real es positivo y la predicción indica que es negativo (o no fue detectado).
- TN: Número de verdaderos negativos. En este caso el valor real es negativo y la predicción determina que es negativo

La precisión indica la cantidad de detecciones verdaderas del total de detecciones, mientras que la Sensibilidad expone la capacidad del detector para revelar los casos positivos. (Koirala, Walsh, Wang, & McCarthy, 2019).

Para la evaluación de la segmentación se empleará el método de intersección sobre la unión (Rezatofighi, et al., 2019). Se identificó que este método es una buena métrica para medir la superposición entre dos cuadros delimitadores o máscaras.

$$IoU = \frac{\text{Área}_{pred} \cap \text{Área}_{gt}}{\text{Área}_{pred} \cup \text{Área}_{gt}} \quad Eq. 3$$

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

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

4. CAPÍTULO IV

RESULTADOS Y DISCUSIONES

En esta sección se pretende evidenciar el proceso de inferencia y obtener resultados del sistema implementado con su respectivo análisis y validación.

4.1. Sistema de Visión Artificial desarrollado para Agricultura de Precisión en manzanas.

Los sistemas de Visión Artificial se basan en tres secciones fundamentales como se observa en la Figura 14, para desarrollar esta secuencia es necesario conocer los diferentes algoritmos y secciones que se implementan durante el proceso, en este caso el sistema implementado consta de dos secciones independientes, la primera sección resuelve el problema de detección de variedades de manzanas según la Tabla 5, con la secuencia observada en la Figura 17. La segunda sección resuelve el problema de segmentación para mostrar el estado de las manzanas según la Tabla 6, con la secuencia observada en la Figura 21.

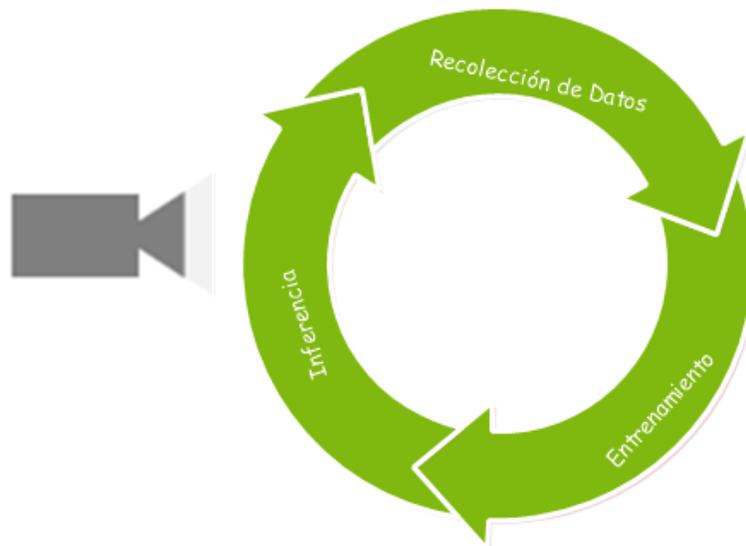


Figura 14. Secciones fundamentales en sistemas de Visión Artificial (NVIDIA, 2019).

Para conseguir un sistema que trabaje bajo la premisa de funcionamiento en tiempo real en dispositivos “on the edge”, como la Jetson Nano, es necesario tener una arquitectura completa que se desarrolle no solo sobre software sino también sobre hardware. Por lo cual se utiliza librerías desarrolladas específicamente para estos dispositivos, basadas en el uso de NVIDIA TensorRT (NVIDIA, 2022), para implementar eficientemente redes neuronales en la plataforma integrada Jetson, mejorando el rendimiento y la eficiencia energética mediante optimizaciones de gráficos, fusión de kernel y precisión FP16/INT8, como se detalla en (NVIDIA, 2022). De este modo se consigue una inferencia en tiempo real para las aplicaciones de detección y segmentación como se observa en la Figura 15.

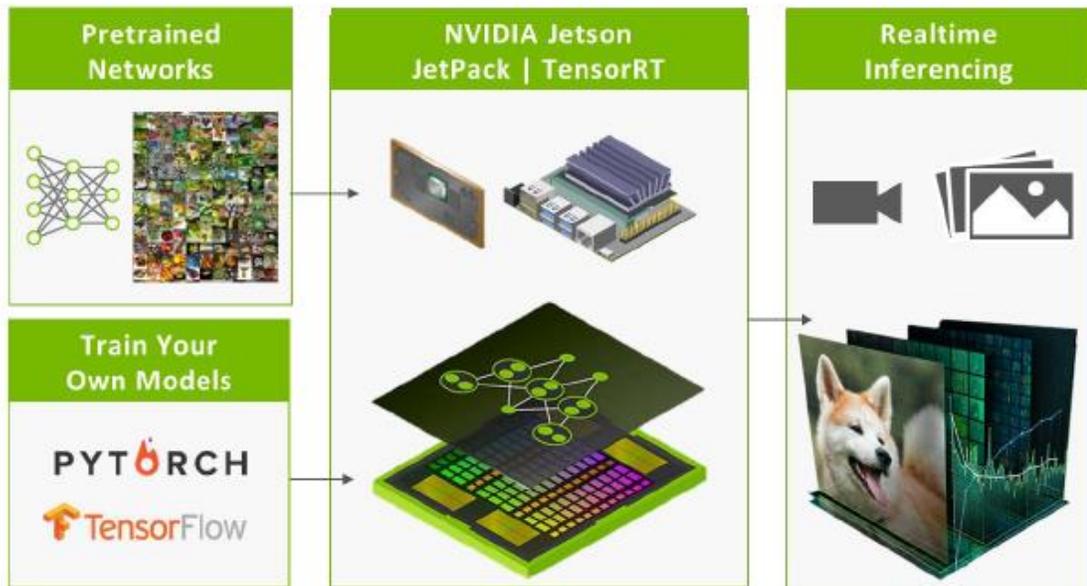


Figura 15. Proceso de inferencia en tiempo real con Jetson (NVIDIA, 2022).

4.1.1 Detección del tipo de manzana

Para la sección de detección se escoge utilizar la arquitectura SSD-MobileNet. La configuración de la red, y programación de los algoritmos fueron implementados utilizando librerías PyTorch, por las características mostradas en Tabla 1, con un mAP

de 19.3% que varía muy alrededor de 3% respecto a modelos mucho más grandes y con mayor costo computacional y con una alta velocidad de inferencia.

SSD-MobileNet (NVIDIA, 2020) es una arquitectura de red para la detección de objetos, se puede observar a detalle sus diferentes secciones en la Figura 16. 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 (Liu, et al., 2016) con una red troncal MobileNet (Howard, et al., 2017), como se observa en la Figura 16.

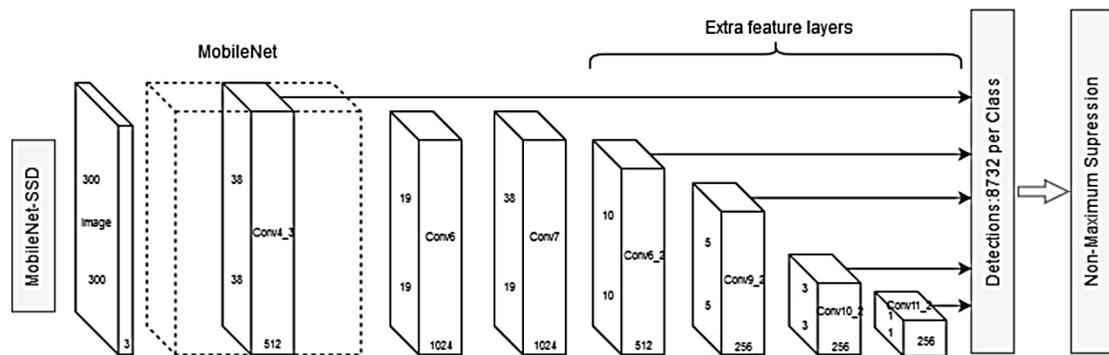


Figura 16. Arquitectura SSD MobileNet (NVIDIA, 2020).

A manera de resumen se identifican los procesos para la detección de la red anteriormente mencionada en base a la arquitectura SSD MobileNet en la Figura 17.



Figura 17. Funcionamiento general sistema de detección con SSD MobileNet (Autor).

4.1.1.1 SSD

La arquitectura SSD se basa en una red convolucional de tipo “feed-forward” que produce una colección de tamaño fijo de cuadros delimitadores y puntuaciones para la

presencia de clases de objetos en esos cuadros, seguido de un paso de supresión no máxima para producir las detecciones finales como se detalla en (Liu, et al., 2016). Las primeras capas de red se basan en una arquitectura estándar utilizada para clasificación de imágenes de alta calidad, denominada red troncal, como se observa en la Figura 17, donde la red troncal es una red MobileNet o como se observa en la Figura 18 donde la red troncal es VGG-16.

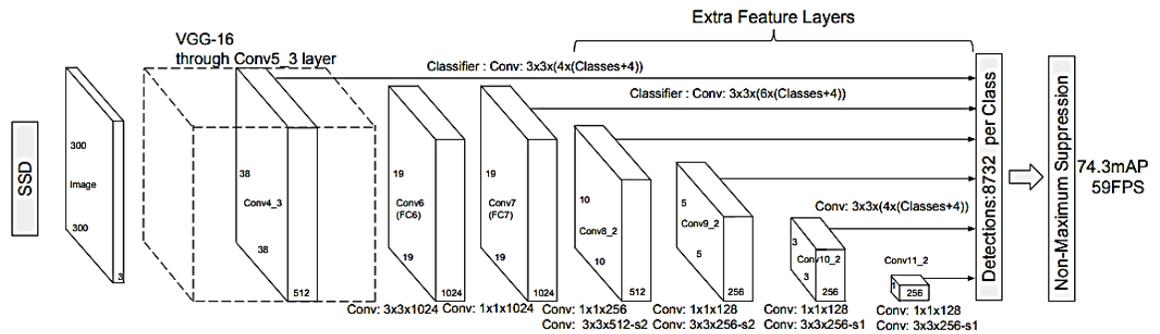


Figura 18. Arquitectura SSD VGG-16 (Liu, et al., 2016).

Esta arquitectura se basa en tener una imagen de entrada a la cual mediante la red troncal se extraen características principales de la misma a diferentes escalas generando un mapa de características como se observa en la Figura 19, los cuales pasan a través de los siguientes bloques de redes convolucionales para generar los cuadros delimitadores y sus diferentes puntuaciones para luego atravesar una capa “Non-Maximum Suppression” como se detalla en (Liu, et al., 2016) para eliminar las superposiciones en la detección y entregar resultados con una precisión promedio (mAP) muy similar a otras arquitecturas pero mucho más pequeña y con velocidades de detección elevadas como se observa en la Tabla 1.

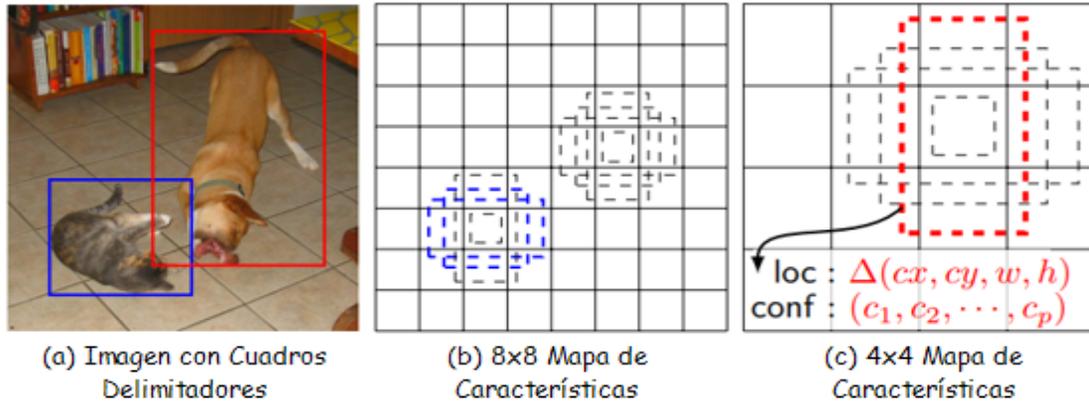


Figura 19. SSD cuadros delimitadores y mapas de características (Liu, et al., 2016).

4.1.1.2 MobileNet

Es un modelo eficiente desarrollado para aplicaciones de visión móviles e integradas se basan en una arquitectura optimizada que utiliza convoluciones separables en profundidad para construir estructuras ligeras de redes neuronales profundas como se detalla en (Howard, et al., 2017).

En la Tabla 7, se enlistan todas las capas que conforman la arquitectura MobileNet.

Tabla 7. Arquitectura MobileNet (Howard, et al., 2017).

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5x Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

La estructura de MobileNet se basa en 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. La arquitectura MobileNet se detalla en la Tabla 7. Todas las capas van seguidas de una “batchnorm” (Wu & Johnson, 2021) y ReLU (Abien, 2018) no lineal con la excepción de la capa final completamente conectada que no tiene no linealidad y se alimenta a una capa softmax (Abien, 2018) para generar su clasificación en porcentaje.

Contando las convoluciones separables en profundidad y en puntos como capas separadas, MobileNet tiene 28 capas.

4.1.2 Segmentación de la manzana de acuerdo con el nivel de calidad

La segmentación semántica se basa en el reconocimiento de imágenes, excepto que las clasificaciones ocurren a nivel de píxel en lugar de la imagen completa. Esto se logra mediante la convolucionización 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 (Shelhamer, Long, & Darrell, 2016). Especialmente útil para la percepción ambiental, la segmentación produce clasificaciones densas por píxel de muchos objetos potenciales diferentes por escena, incluidos los primeros planos y los fondos de la escena (NVIDIA, 2019), como se observa en la Figura 20.

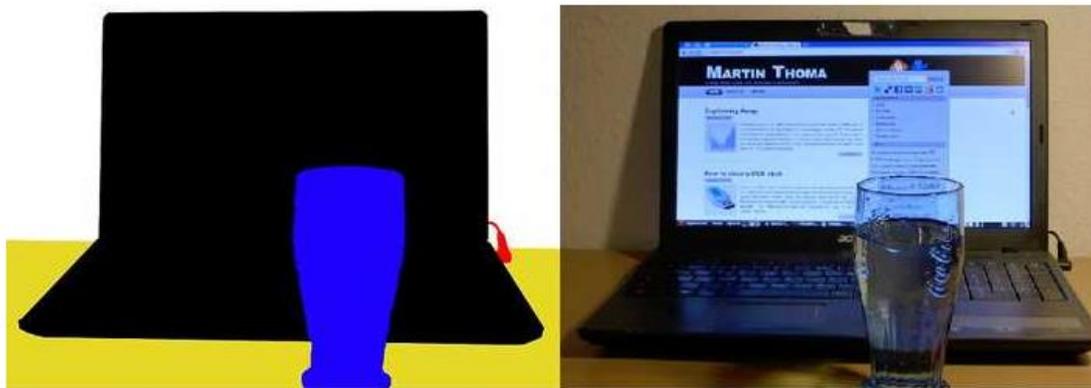


Figura 20. Ejemplo de resultado de segmentación semántica (Polikevicius, 2022).

La librería segNet 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 que se clasificó. El detalle de funcionamiento del sistema implementado en segmentación se observa en la Figura 21, donde se observa la implementación de FCN ResNet18 para la inferencia en segmentación, se escogió esta arquitectura debido a las características que se observan en la Figura 6 y su funcionamiento óptimo para aplicaciones en tiempo real.

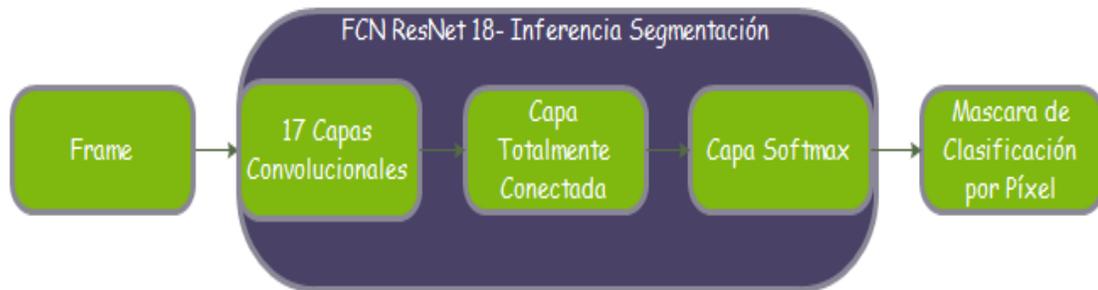


Figura 21. Funcionamiento general sistema de segmentación con FCN ResNet18 (Autor).

4.1.2.1 FCN ResNet 18

ResNet-18 es una red neuronal convolucional que como se describe en (The MathWorks, Inc., 2022) 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 tiene un tamaño de entrada de imagen de 224x224. La arquitectura del ResNet-18 original se muestra en Figura 22.

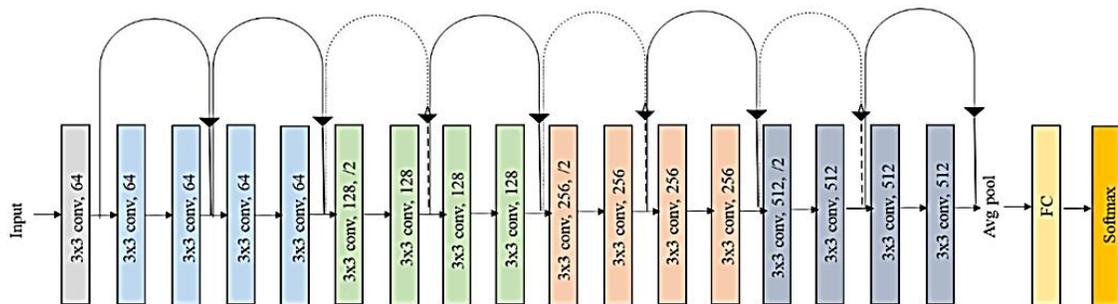


Figura 22. Arquitectura original ResNet18 (Ramzan, et al., 2019).

Las capas convolucionales usan filtros 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 22.

A lo largo de la red, se insertan conexiones de acceso directo residuales entre capas. Hay dos tipos de conexiones como se observa en la Figura 22. El primer tipo de conexiones, indicadas por líneas sólidas, se utilizan cuando al ingreso y la salida tienen las mismas dimensiones. Los segundos tipos de conexiones, indicadas por líneas de puntos, 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 (Shelhamer, Long, & Darrell, 2016). Finalmente se puede observar un ejemplo del funcionamiento de la segmentación con respecto a una arquitectura con redes convolucionales en la Figura 23, donde se puede observar la entrada como una imagen y la salida la predicción píxel por píxel para generar la máscara de segmentación.

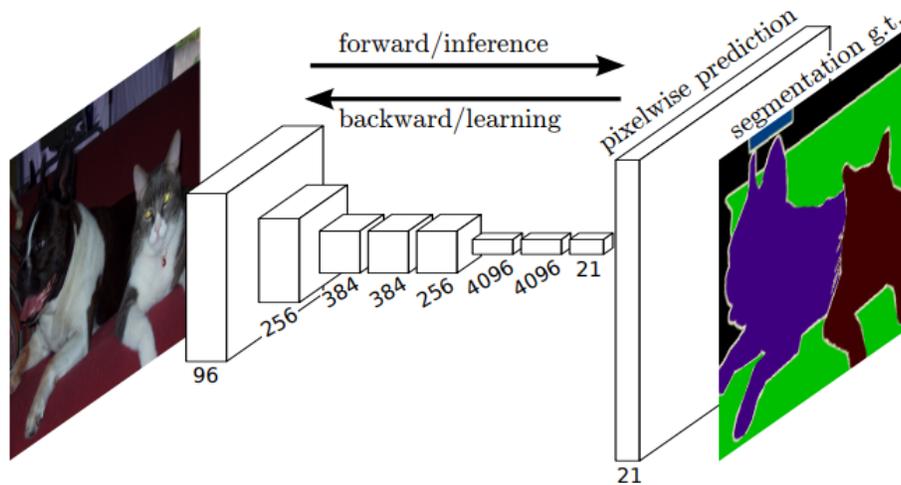


Figura 23. Ejemplo de resultado de segmentación semántica con redes convolucionales (Shelhamer, Long, & Darrell, 2016).

4.2. Preparación de datos.

Culminado el proceso de establecer los modelos de red para la detección y la segmentación se procedió con la elaboración de las bases de datos.

4.2.1. Preparación de datos para detección.

Para recopilación de los conjuntos de imágenes, la creación de archivos con etiquetas se la realizó directamente en el dispositivo Jetson Nano.

Para la adquisición de imágenes se utilizó la herramienta conocida como camera-capture presentada en a Figura 24. Esta aplicación de Jetson Inference (NVIDIA, 2020) permite trabajar paralelamente con la recopilación de datos y el etiquetado para cada clase.

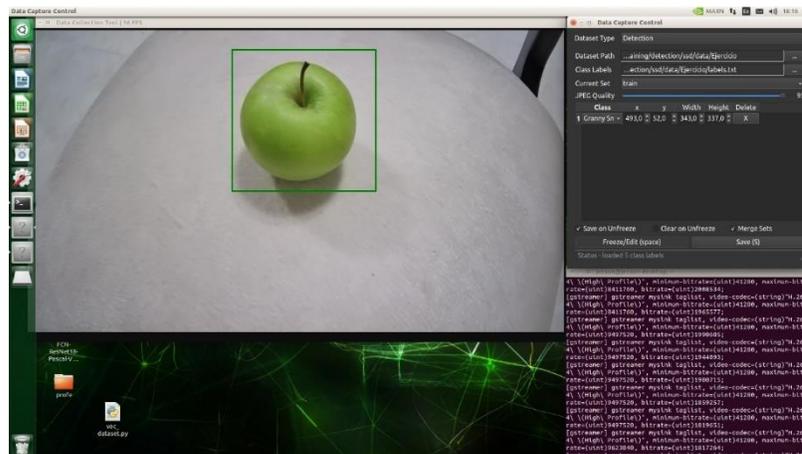


Figura 24. Panel de adquisición y etiquetado de imagen (Autor).

La totalidad de imágenes para la identificación del tipo de manzana fueron obtenidas en formato JPG y con una resolución de 1280 x 720 pixeles.

Para la identificación del tipo de manzana se adquirieron imágenes para pruebas de laboratorio y de campo.

Para los ensayos de laboratorio se generaron 2 bases de datos teniendo como objeto únicamente las manzanas y de fondo con textura uniforme, mientras que para los ensayos de campo se capturaron imágenes de manzanas que se encontraban colgando en los ramales de los árboles, tal como se representa en la Figura 25.



Figura 25. Imágenes de los tipos de manzana para cada base de datos (Autor).

Es importante recalcar que, de acuerdo con la necesidad de una variedad de manzanas rojas y verdes para la ejecución de este análisis, fue preciso obtener los frutos de manera comercial, debido a que durante el periodo de ejecución de este estudio no se generaron este tipo de frutos. Adicionalmente, las manzanas que se colocaron en un campo real para los ensayos de campo fueron capturadas en imágenes para realizar el análisis correspondiente.

La Figura 26 hace referencia al esquema empleado para la adquisición de imágenes requeridas para los ensayos de laboratorio.

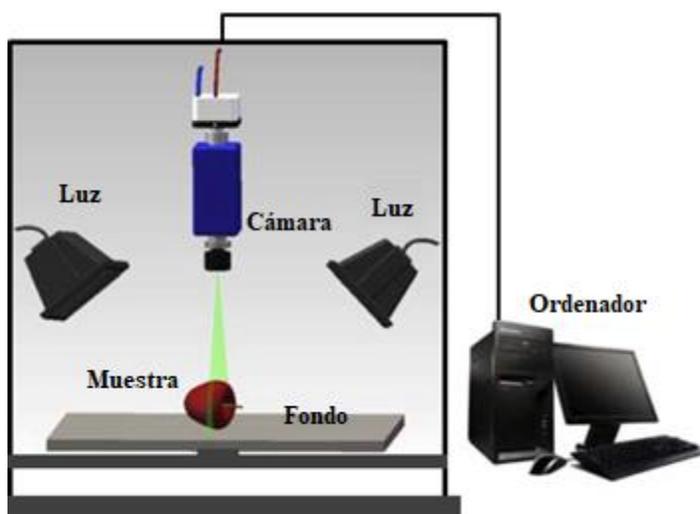


Figura 26. Diagrama empleado para adquisición de imágenes de laboratorio (Zhang, et al., 2014).

4.2.1.1. Conjunto de datos para detección – Base de datos 1

En esta base de datos se presenta por cada imagen una manzana y el fondo con textura de color verde para generar datos artificiales.

Se capturaron 1650 imágenes, distribuidas de la siguiente manera 1500 imágenes de los 5 tipos de manzana distribuidas para el entrenamiento (70%) y la validación (30%) y de manera adicional se capturaron 150 imágenes para las pruebas.

En la Tabla 8 se detalla la cantidad de manzanas para cada variedad, contenida en esta base datos.

Tabla 8. Variedad de manzanas y cantidad de imágenes en base de datos 1 (Autor).

TIPO DE MANZANA	NÚMERO DE IMÁGENES			
	ENTRENAMIENTO	VALIDACIÓN	TOTAL	PRUEBA
Granny Smith	210	90	300	30
Golden Delicious	210	90	300	30
Fuji	210	90	300	30
Royal Gala	210	90	300	30
Red Delicious	210	90	300	30
TOTAL	1050	450	1500	150
PORCENTAJE	70%	30%	100%	10%

4.2.1.2. Conjunto de datos para detección – Base de datos 2

En esta base de datos se presenta por cada imagen una manzana y el fondo con textura de madera.

Se capturaron 1650 imágenes, distribuidas de la siguiente manera 1500 imágenes de los 5 tipos de manzana distribuidas para el entrenamiento (70%) y la validación (30%) y de manera adicional se capturaron 150 imágenes para las pruebas.

En la Tabla 9 se detalla la cantidad de manzanas para cada variedad, contenida en esta base datos.

Tabla 9. Variedad de manzanas y cantidad de imágenes en base de datos 2 (Autor).

TIPO DE MANZANA	NÚMERO DE IMÁGENES			
	ENTRENAMIENTO	VALIDACIÓN	TOTAL	PRUEBA
Granny Smith	210	90	300	30
Golden Delicious	210	90	300	30
Fuji	210	90	300	30
Royal Gala	210	90	300	30
Red Delicious	210	90	300	30
TOTAL	1050	450	1500	150
PORCENTAJE	70%	30%	100%	10%

4.2.1.3. Conjunto de datos para detección – Base de datos 3

En esta base de datos se presentan por cada imagen de 1 a 3 manzanas y como fondo se presentan los ramales de un árbol.

Se capturaron 1825 imágenes, distribuidas de la siguiente manera 1750 imágenes de los 5 tipos de manzana distribuidas para el entrenamiento (70%) y la validación (30%) y de manera adicional se capturaron 175 imágenes para las pruebas.

Para la ejecución de este ensayo se sujetaron las manzanas en un árbol con la finalidad de simular el ambiente agrícola del fruto.

En la Tabla 10 se detalla la cantidad de manzanas para cada variedad contenida en esta base datos.

Tabla 10. Variedad de manzanas y cantidad de imágenes en base de datos 3 (Autor).

TIPO DE MANZANA	NÚMERO DE IMÁGENES			
	ENTRENAMIENTO	VALIDACIÓN	TOTAL	PRUEBA
Granny Smith	245	105	350	35
Golden Delicious	245	105	350	35
Fuji	245	105	350	35
Royal Gala	245	105	350	35
Red Delicious	245	105	350	35
TOTAL	1225	525	1750	175
PORCENTAJE	70%	30%	100%	10%

4.2.2. Preparación de datos para segmentación.

Dado que las imágenes fueron obtenidas desde 2 dispositivos con diferente resolución 3024x4032 píxeles y 3456x3456 píxeles, fue necesario realizar un proceso de redimensionamiento y corte (Anexo 2) para estandarizar la base de datos, mediante 2 códigos elaborados en Pycharm Community que es un IDE para desarrollar programas en Python.

La nueva resolución que se estableció para la totalidad de las imágenes durante el procesamiento fue de 1920x1080 píxeles y con un formato JPG.

Para este caso se disponen 338 imágenes distribuidas de la siguiente manera: 240 entrenamiento (70%), 60 imágenes de validación (30%) y 38 imágenes para pruebas.

Debido a que el conjunto de datos fue obtenido propiamente se utilizó la herramienta Labelme la cual permite etiquetar al objeto a analizar con su respectiva clase.

Tabla 11. Cantidad de etiquetas de manzanas por parámetro de calidad (Autor).

PARÁMETRO DE CALIDAD	NÚMERO
Alta	978
Media	643
Baja	590

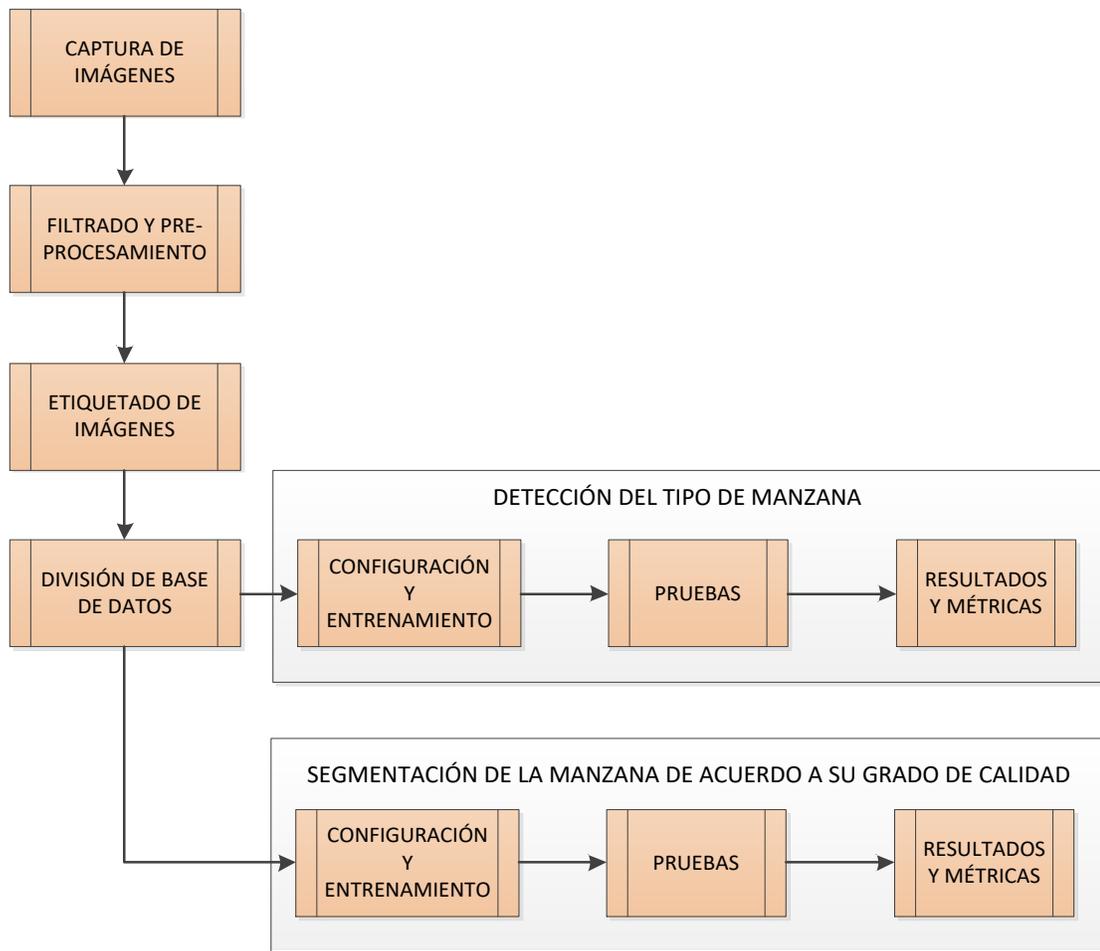


Figura 28. Diagrama general del proceso experimental (Autor).

4.3.1. Herramientas de Hardware y Software.

En esta sección se exponen las herramientas que emplearán en el desarrollo del proyecto, para los procesos de detección y de segmentación de imágenes.

4.3.1.1. Hardware

Como se indica en la Figura 29 se colocó la tarjeta microSD de 64GB en la placa electrónica Jetson Nano, en esta se configurará el software que se detallará en el siguiente capítulo.

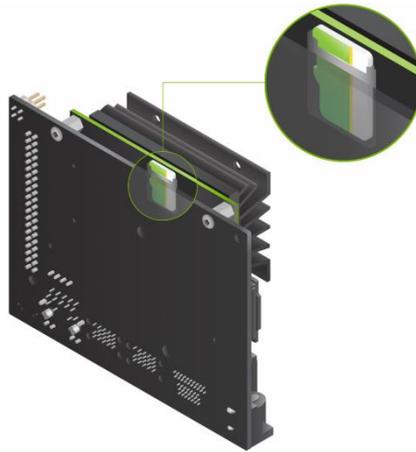


Figura 29. Lugar de ranura para tarjeta microSD (NVIDIA, 2022).

Posteriormente, se ensambló la placa electrónica Jetson Nano en una carcasa metálica para asegurar condiciones de seguridad como se observa en la Figura 30, adicionalmente se colocaron botones para darle mayor robustez para disminuir riesgo de averías en su manipulación.



Figura 30. Vista frontal y posterior de la carcasa de la Jetson Nano (Autor).

En la Figura 31 se observa el ensamblaje de la cámara web en un trípode para evitar movimientos y vibraciones durante la captura de las imágenes.



Figura 31. Vista frontal cámara con el trípode (Autor).

Finalmente, se realizaron las conexiones de los dispositivos de entrada y salida en los puertos de la tarjeta electrónica.

4.3.1.2. Software para entrenamiento de la detección

Dado que el entrenamiento para la detección del tipo de manzana se lo realizó directamente en la placa electrónica Jetson Nano, fue necesario cumplir con el siguiente procedimiento.

- Desde el sitio de NVIDIA DEVELOPER se descargó el controlador para Jetson Nano denominado SDK Jetpack 4.6 que contiene el sistema operativo, controladores y componentes necesarios para desarrollar aplicaciones de IA, lo que simplifica el proceso de instalación individual. Los componentes que incluye este paquete son: Núcleo L4T/BSP Kit de herramientas CUDA, cuDNN, TensorRT, OpenCV, entre otros (NVIDIA, 2022).

- Se realizó la descarga del software SD Card Formatter, se insertó la tarjeta microSD de 64 GB en la ranura de la placa electrónica y se procedió con la ejecución del formato rápido, tal como se muestra en el Anexo 1.
- Se procedió con la descarga del software Balena Etcher que permite montar la imagen en la tarjeta microSD con la información que contiene el SDK Jetpack 4.6. En el Anexo 1 se muestran los pasos de la configuración inicial.
- Con los procesos ejecutados se dispone de la Jetson Nano preparada para la ejecución de las aplicaciones con el sistema operativo Ubuntu 18.04, como se observa en el Anexo 1.
- Para la iniciación del proyecto fue necesario clonar la biblioteca jetson-inference que es compatible con el dispositivo Jetson Nano, adicionalmente instalar el paquete de desarrollo de Python, descargar los modelos de Aprendizaje Profundo que se emplearán durante la ejecución del proyecto, instalar Pytorch y finalmente compilar el proyecto.

4.3.1.3. Software para entrenamiento de segmentación

Para optimizar el tiempo de entrenamiento fue necesario:

- Se empleó la GPU del computador indicado en la Subsección 3.3, en el cual se instaló el CUDA 10.0 que permite potenciar el rendimiento del computador en el procesamiento de imágenes.
- Se instaló el CUDNN 7.6.5 (Red Neuronal Profunda Nvidia CUDA) que es una recopilación de primitivas de NVIDIA para crear redes neuronales profundas que emplean el entrenamiento acelerado por GPU.
- Se instaló el entorno de desarrollo Anaconda, para gestionar adecuadamente el sitio en el que se aislará el proyecto y en este se situó Pytorch que es una de las librerías de Python que permiten el entrenamiento de redes neuronales.
- Se clonó el repositorio Pytorch Segmentation y se procedió con su instalación, al igual que onnx (Open Neural Network Exchange) y pycocotools.
- Se instaló la herramienta de interfaz gráfica para etiquetado de imágenes conocida como Labelme.

4.4. Entrenamiento de los sistemas

En esta sección se comparten los parámetros que se utilizaron para entrenar los modelos preexistentes con las bases de datos generadas específicamente para el desarrollo del proyecto y expuestas en la Subsección 4.1.

4.4.1. Entrenamiento para detección

Dado que el entrenamiento para la detección del tipo de manzana se realizó en la placa electrónica Jetson Nano, se utilizó la guía propuesta por (NVIDIA, 2019) en la que el modelo para detección de objetos es el SSD- Mobilenet.

Las bases de datos que se usaron son las indicadas en el Subsección 4.2.1 para los ensayos de laboratorio y campo.

El lenguaje de programación que se utilizó para el entrenamiento fue Python y en este se configuran los argumentos para el proceso de entrenamiento del modelo de acuerdo con lo indicado en la Tabla 12.

Para la determinación de los hiperparámetros de la red de entrenamiento se tuvo las siguientes consideraciones:

- Para “batch-size” se lo configuró con 2 ya que la memoria disponible en la Jetson Nano no permite realizarlo con un número mayor, debido a su capacidad.
- Para “epochs” se las realizó con 125 ya que se realizaron ensayos previos y se vio la necesidad de mejorar los resultados del modelo.
- Y para “workers” el fabricante recomienda que este parámetro por defecto.

Tabla 12. Hiperparámetros de la red de entrenamiento (Autor).

ARGUMENTO	CARACTERÍSTICAS CONFIGURADAS	DETALLE
--data	data/pcampo	Dirección del conjunto de datos
--model-dir	models/pcampo	Dirección en donde se almacenan los puntos de información de cada época.
--batch-size	2	Es el número de muestras que pasaran a la vez en la red.
--epochs	125	Número de épocas totales ejecutadas durante el entrenamiento.
--workers	2	Número de subprocesos del cargador de datos

En la Anexo 3, se observa el proceso de entrenamiento del modelo en la Jetson nano. El tiempo de entrenamiento fue de 16 horas. Durante este período se ejecutaron las iteraciones por cada época, y de manera progresiva se obtiene una mejor precisión

Finalizado el entrenamiento, convertimos el modelo entrenado a ONNX, para posteriormente cargarlo con el Tensor RT, con esta acción será posible trasladar y ejecutar el archivo directamente en la Jetson Nano Developer Kit.

4.4.2. Entrenamiento para segmentación.

Como se mencionó en la sección de Equipos y Materiales, se utilizó el computador para agilizar el proceso de entrenamiento, debido a que posee mejores características de procesamiento con respecto a la placa electrónica Jetson Nano. El entrenamiento se lo realizó en la suite de código Anaconda.

El proceso inicia con la conversión de los archivos con imágenes etiquetadas en formato JSON a archivos que presenten las imágenes con las máscaras correspondientes a cada nivel de calidad JPG.

El siguiente proceso consiste en ejecutar el directorio de pytorch segmentation el cual convierte las imágenes etiquetadas que poseen formato JSON en máscaras de imagen, la ejecución se lo realiza de acuerdo a lo indicado en el Anexo 4. Las máscaras de imagen presentan la información de las etiquetas en formato JPG, como se puede observar en el Anexo 4, para cada nivel de calidad se presenta una diferente coloración.

En el Anexo 4 se muestra el proceso de ejecución del script split_custom.py el cual permitió generar divisiones de datos para el entrenamiento y la validación. Este código receipta la totalidad de las imágenes y automáticamente las distribuye para cada uno de los conjuntos de datos.

En la Tabla 13 se cuantifica la cantidad de manzanas para el entrenamiento y la validación.

Tabla 13. Cantidad de imágenes para entrenamiento y validación (Autor).

CONJUNTOS	CANTIDAD DE IMÁGENES
Entrenamiento	240
Validación	60

En el Anexo 4 se presenta el proceso de entrenamiento del modelo en la PC.

Con los archivos completos, se inició con el entrenamiento del modelo con la ejecución del código en Python seguido de los siguientes argumentos que influyeron en el adecuado entrenamiento del modelo indicados en la Tabla 14.

Para la determinación de los hiperparámetros de la red de entrenamiento se tuvo las siguientes consideraciones:

- Para “data” se lo configuró con VOC, debido a que el software utilizado emplea este tipo de datos.
- Para “arch” la arquitectura se la configuró con Fcn_ResNet18 debido al análisis previo realizado en el estado del arte.
- Para “width y height” se lo configuró con 1920 x 1080 píxeles debido a que el conjunto de imágenes presenta la resolución indicada.
- Para “batch-size” se lo configuró con 3 ya que la memoria disponible en la computadora no permite realizarlo con un número mayor, debido a su capacidad.
- Para “batch-size” se lo configuró con 3 ya que la memoria disponible en la computadora no permite realizarlo con un número mayor, debido a su capacidad.
- Para “epochs” se las realizó con 500 debido a que para este caso el software de entrenamiento escogerá la mejor época y se observa que en ese rango de épocas la precisión del entrenamiento es aceptable.
- Y para “workers” se configura con 1 debido a que el fabricante recomienda que este parámetro por defecto para el entrenamiento en computadoras.

Tabla 14. Hiperparámetros y características de configuración (Autor).

ARGUMENTO	CARACTERÍSTICAS CONFIGURADAS	DETALLE
--data	VOC	Tipo de datos
--arch	Fcn_resnet18	Arquitectura de la red
--classes	4 (se incluye el fondo)	Número de clases de objetos.
--width	1920 píxeles	Ancho deseado del conjunto de imágenes de entrenamiento.
--height	1080 píxeles	Alto deseado del conjunto de imágenes de entrenamiento.
--batch-size	3	Es el número de muestras que pasaran a la vez en la red.
--epochs	500	Número de épocas totales para ejecutar el entrenamiento.
--workers	1	Número de subprocesos del cargador de datos

Durante la ejecución del entrenamiento, el programa determina el mejor modelo de acuerdo con el que presente una mejor precisión y media IoU. El tiempo total de entrenamiento fue de 7.18 horas sin interrupción.

Finalizado el entrenamiento, convertimos el modelo entrenado a ONNX, para posteriormente cargarlo con el Tensor RT, con esta acción será posible trasladar y ejecutar el archivo directamente en la Jetson Nano Developer Kit.

4.5. Pruebas experimentales.

Finalizados los procesos de entrenamiento de las redes tanto para detección del tipo de manzana y de segmentación de acuerdo con el nivel de calidad se procede a verificar si la red dispone de la capacidad para reconocimiento de los patrones con los cuales fue entrenada.

4.5.1. Pruebas de detección

Las pruebas para la detección del tipo de manzana nos guiamos en el esquema presentado por (Zhang, et al., 2014) y diagramado en la Figura 26. Los ensayos de laboratorio se los realizó sobre una mesa con 2 tipos de fondo. En la Figura 32 se muestra un ejemplo de panorama en el fondo 2.



Figura 32 Ejecución de los ensayos de laboratorio (Autor).

Para la ejecución de los ensayos de campo para la detección de la clase de manzana se ensambló la estructura del hardware en el huerto mencionado en la Subsección 4.3.1 e indicado en la Figura 33, en donde se expone que las manzanas adquiridas en el supermercado fueron colocadas en un árbol para simular la operación dentro del ambiente agrícola.



Figura 33. Ejecución del sistema en línea para detección de calidad (Autor).

En la Figura 34 se muestran imágenes de salida con la detección en las pruebas de laboratorio y campo (Autor).

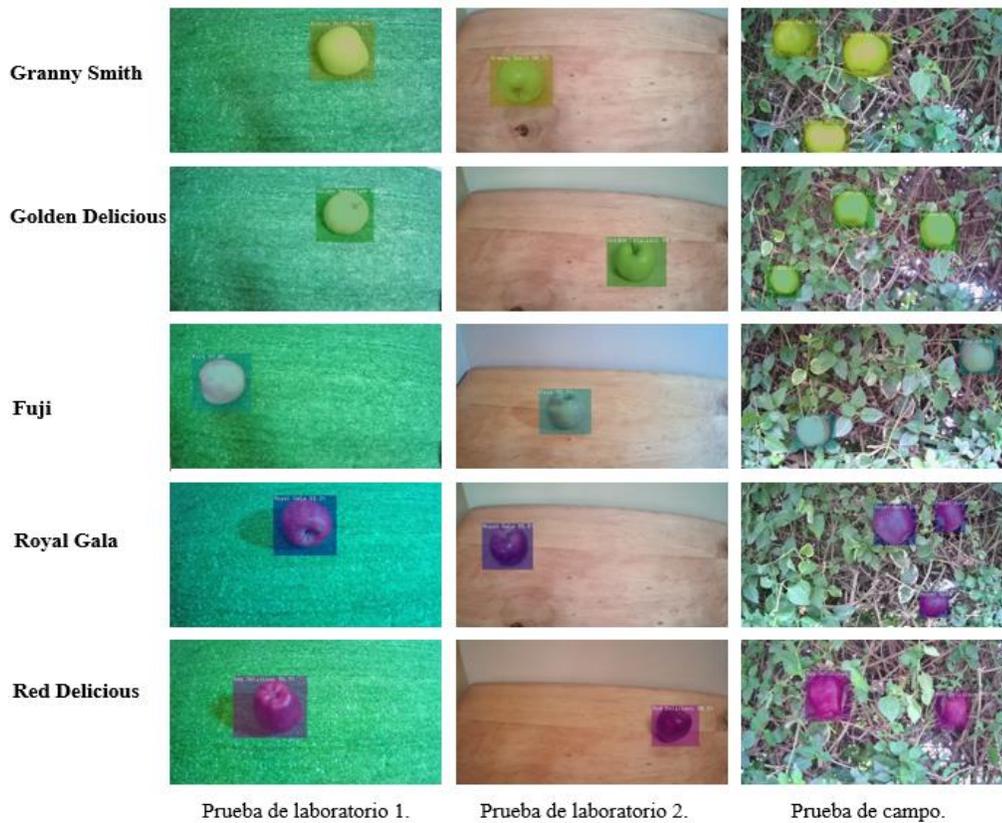


Figura 34. Imágenes de salida con detección de tipo de manzana (Autor).

4.5.2. Pruebas de segmentación

Mediante una memoria USB se movió el archivo generado en formato ONNX desde el computador hasta el dispositivo Jetson Nano, en donde se lo ejecutó con el modelo.

Posteriormente, se ejecutaron los comandos de la librería Segnet (NVIDIA, 2019).

Para este caso se ubicó el sistema en el huerto y se generó la segmentación de las manzanas Royal Gala, como se observa en la Figura 35.



Figura 35. Imágenes de salida con manzanas segmentadas de acuerdo con grado de calidad (Autor).

4.6. Resultados

En esta sección se muestran los resultados obtenidos divididos en 2 secciones:

- En el primer apartado se muestran los resultados de la detección del tipo de manzana, su evaluación se la realizó con matrices de confusión que muestran la precisión y la sensibilidad para cada variedad de manzana en los diferentes ensayos.
- En el segundo apartado se muestran los resultados de la segmentación que muestra el grado de calidad de las manzanas, las métricas que se emplearon para este análisis fueron la IoU (Intersección sobre la Unión por sus siglas en inglés) y precisión durante el entrenamiento.

4.6.1. Análisis de resultados de la detección del tipo de manzana.

A continuación, se presentan las matrices de confusión desarrolladas para cada uno de los casos y conjuntos de datos recopilados. La matriz de confusión muestra en las filas (eje X) las etiquetas reales y en las columnas (eje Y) las etiquetas de predicción para cada tipo de manzana.

También se adicionó una columna en el extremo derecho, la cual muestra la precisión por cada tipo de manzana y una fila en el extremo inferior de la matriz de confusión la cual muestra la sensibilidad por cada clase.

4.6.1.1. Prueba de Laboratorio 1: Identificación de manzanas en ambiente controlado

Analizando los resultados de la matriz de confusión de la Tabla 15 que fue generada con el conjunto de datos de validación 1, de los cinco tipos de manzanas del modelo propuesto. En la columna del extremo derecho se observa que las manzanas Golden Delicious y Fuji presentan la mayor precisión, mientras que las manzanas Royal Gala presentan la menor precisión. Además, la sensibilidad más alta corresponde a la manzana Granny Smith y la más baja corresponde a la manzana Fuji.

La precisión general del modelo propuesto evaluado con el conjunto de datos de validación es del 93.6%.

Tabla 15. Matriz de confusión de conjunto de datos de validación 1 (Autor).

PREDICCIÓN	Granny Smith	90 20,0%	0 0,0%	4 0,9%	0 0,0%	0 0,0%	95,7% 4,3%
	Royal Gala	0 0,0%	88 19,6%	0 0,0%	22 4,9%	1 0,2%	79,3% 20,7%
	Golden Delicious	0 0,0%	0 0,0%	86 19,1%	0 0,0%	0 0,0%	100,0% 0,0%
	Fuji	0 0,0%	0 0,0%	0 0,0%	68 15,1%	0 0,0%	100,0% 0,0%
	Red Delicious	0 0,0%	2 0,4%	0 0,0%	0 0,0%	89 19,8%	97,8% 2,2%
		100,00% 0,0%	97,8% 2,2%	95,6% 4,4%	75,6% 24,4%	98,9% 1,1%	93,6% 6,4%
		Granny Smith	Royal Gala	Golden Delicious	Fuji	Red Delicious	
		REAL					

Analizando los resultados de la matriz de confusión de la Tabla 16 que fue generada con el conjunto de datos de prueba 1, 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 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%.

Tabla 17. Matriz de confusión de conjunto de datos de validación 2 (Autor).

PREDICCIÓN	Granny Smith	90 20,0%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Royal Gala	0 0,0%	90 20,0%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Golden Delicious	0 0,0%	0 0,0%	90 20,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Fuji	0 0,0%	0 0,0%	0 0,0%	90 20,0%	0 0,0%	100,0% 0,0%
	Red Delicious	0 0,0%	0 0,0%	0 0,0%	0 0,0%	90 20,0%	100,0% 0,0%
		100,00% 0,0%	100,0% 0,0%	100,0% 0,0%	100,0% 0,0%	100,0% 0,0%	100,0% 0,0%
		Granny Smith	Royal Gala	Golden Delicious	Fuji	Red Delicious	
	REAL						

Analizando los resultados de la matriz de confusión de la Tabla 18 que fue generada con el conjunto de datos de prueba 2, 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%.

Tabla 18. Matriz de confusión de conjunto de datos de prueba 2 (Autor).

PREDICCIÓN	Granny Smith	30 6,7%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Royal Gala	0 0,0%	24 5,3%	0 0,0%	1 0,2%	4 0,9%	82,8% 17,2%
	Golden Delicious	0 0,0%	0 0,0%	30 6,7%	0 0,0%	0 0,0%	100,0% 0,0%
	Fuji	0 0,0%	2 0,4%	0 0,0%	30 6,7%	2 0,4%	88,2% 11,8%
	Red Delicious	0 0,0%	2 0,4%	0 0,0%	0 0,0%	24 5,3%	92,3% 7,7%
		100,00% 0,0%	85,7% 14,3%	100,0% 0,0%	96,8% 3,2%	80,0% 20,0%	92,6% 7,4%
		Granny Smith	Royal Gala	Golden Delicious	Fuji	Red Delicious	
	REAL						

4.6.1.3. Pruebas de Campo: Identificación de manzanas en ambiente agrícola

Para las pruebas de campo se tomó en cuenta la información expuesta en el estado del arte por (Bargoti & Underwood, 2017) y en la que se indica que las condiciones al aire libre podrían presentar dificultades adicionales debido a la alta variabilidad de iluminación, condiciones climáticas y oclusión, lo que podría generar una inadecuada detección u objetos no detectados. Por tal motivo para este caso se adicionó a la matriz de confusión la clase no detectada para evaluar adecuadamente el sistema con todas las variables existentes.

Analizando los resultados de la matriz de confusión de la Tabla 19 que fue generada con el conjunto de datos de validación 3, de los cinco tipos de manzanas del modelo propuesto, se observa que todas las manzanas presentan una precisión del 100%, mientras que la mayor sensibilidad se presenta en la manzana Royal Gala y la menor sensibilidad en la manzana Fuji.

Para el total de este caso se presentaron 261 manzanas no reconocidas lo que influye en la precisión general del modelo propuesto evaluado con el conjunto de datos de validación es del 82,4%.

Tabla 19. Matriz de confusión de conjunto de datos de validación 3 (Autor).

PREDICCIÓN	Granny Smith	255 17,2%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Royal Gala	0 0,0%	241 16,3%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Golden Delicious	0 0,0%	0 0,0%	244 16,5%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Fuji	0 0,0%	0 0,0%	0 0,0%	231 15,6%	0 0,0%	0 0,0%	100,0% 0,0%
	Red Delicious	0 0,0%	0 0,0%	0 0,0%	0 0,0%	249 16,8%	0 0,0%	100,0% 0,0%
	No detectada	53 3,6%	39 2,6%	46 3,1%	71 4,8%	52 3,5%	0 0,0%	0,0% 100,0%
		82,79% 17,2%	86,1% 13,9%	84,1% 15,9%	76,5% 23,5%	82,7% 17,3%	0,0% 100,0%	82,4% 17,6%
		Granny Smith	Royal Gala	Golden Delicious	Fuji	Red Delicious	No detectada	
		REAL						

Analizando los resultados de la matriz de confusión de la Tabla 20 que fue generada con el conjunto de datos de prueba 3, 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.

Para el total de este caso se presentaron 48 manzanas no reconocidas lo que influye en la precisión general del modelo propuesto evaluado con el conjunto de datos de prueba 3 con un 90.4%.

Tabla 20. Matriz de confusión de conjunto de datos de prueba 3 (Autor).

PREDICCIÓN	Granny Smith	101 19,7%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Royal Gala	0 0,0%	87 17,0%	0 0,0%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Golden Delicious	0 0,0%	0 0,0%	99 19,3%	0 0,0%	0 0,0%	0 0,0%	100,0% 0,0%
	Fuji	0 0,0%	0 0,0%	1 0,2%	90 17,6%	0 0,0%	0 0,0%	98,9% 1,1%
	Red Delicious	0 0,0%	0 0,0%	0 0,0%	0 0,0%	86 16,8%	0 0,0%	100,0% 0,0%
	No detectada	5 1,0%	18 3,5%	5 1,0%	6 1,2%	14 2,7%	0 0,0%	0,0% 100,0%
		95,28% 4,7%	82,9% 17,1%	94,3% 5,7%	93,8% 6,3%	86,0% 14,0%	0,0% 100,0%	90,4% 9,6%
		Granny Smith	Royal Gala	Golden Delicious	Fuji	Red Delicious	No detectada	
		REAL						

En el siguiente enlace se podrá encontrar el video de la ejecución del software para la estimación en tiempo real del proceso de identificación del tipo de manzana:

<https://drive.google.com/file/d/1HYJ6Km0qfOYvIzbsmrQyyODohjJaZIUN/view?usp=sharing>

4.6.2 Análisis de resultados de la segmentación semántica de la manzana

Para los ensayos de cuantificación de la calidad de la manzana empleando la técnica de segmentación semántica se obtuvieron los resultados indicados en la Subsección 4.6.2.1.

4.6.2.1. Prueba de Campo: Cuantificación de calidad de manzanas en ambiente agrícola

La evaluación para esta sección se la realizó de acuerdo con los resultados del entrenamiento en el que se obtuvo que los mejores resultados se presentaron en la época 218 con 94.656 % de precisión y 55.646% de media IOU (Intersection Over Union).

Siendo esta precisión y media IoU, mejores resultados en comparación con los revisados en (Santos, de Souza, dos Santos, & Avila, 2020) donde la precisión y medida IoU fueron 92.3% y 40% respectivamente Aplicados sobre una base de datos similar de 300 imágenes para segmentación de frutas (uvas).

En la Figura 36 se muestra el diagrama de precisión por cada época del entrenamiento.

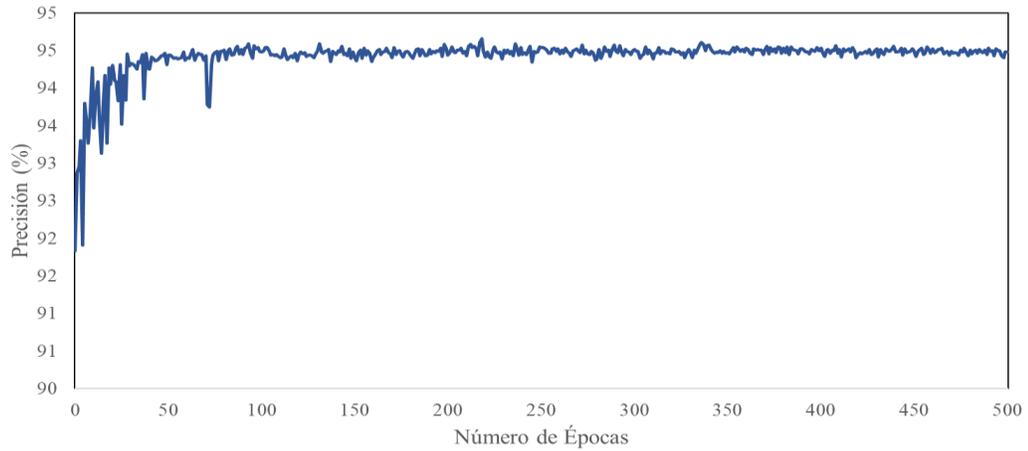


Figura 36. Diagrama de precisión durante el entrenamiento (Autor).

En la Figura 37 se muestra el diagrama de IOU (Intersection Over Union) por cada época del entrenamiento, obteniendo como mejor resultado el de la época 218 con 55.65%.

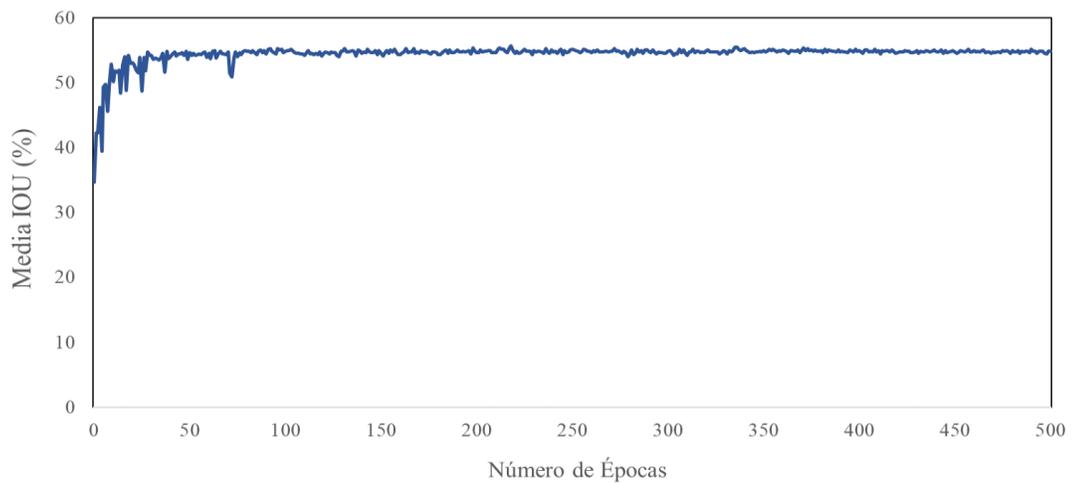


Figura 37. Diagrama de media IoU por cada época del entrenamiento (Autor).

Durante el desarrollo experimental del ensayo de campo empleando las imágenes de los conjuntos de entrenamiento y validación para alimentar al sistema, se obtuvo como resultado que el 96.32% de las imágenes segmentadas de alta calidad coinciden con su etiqueta, el 95.96% de las imágenes segmentadas de media calidad coinciden con su etiqueta y el 88.81% de las imágenes segmentadas de baja calidad coinciden con su etiqueta.

Adicionalmente, se realizó el ensayo de un conjunto de prueba en un huerto en donde se tuvieron árboles con un total de 360 manzanas, obteniendo los resultados indicados en la Tabla 21.

Tabla 21. Análisis de cuantificación de calidad.

GRADO DE CALIDAD	NÚMERO DE MANZANAS	PORCENTAJE
Alta	170	47 %
Media	109	30 %
Baja	81	23 %
TOTAL	360	100%

En el conjunto de datos de la Tabla 21 se observa que el 47% de manzanas del total de la muestra se encuentran aptas para su cosecha inmediata, al 30% se debería realizarlas una nueva inspección visual para su validación y el 23% se encuentran totalmente deterioradas.

Es decir, el resultado de las manzanas de alta calidad y baja calidad es contundente tanto para su aprobación y rechazo, otorgando al agricultor la información suficiente para tomar las decisiones correspondientes. Para las manzanas de nivel medio, se recomienda realizar una inspección en toda su superficie. En el siguiente enlace se podrá encontrar el video de la ejecución del software que cuantifica la calidad del cultivo de manzanas:

https://drive.google.com/file/d/1Sn8F4PxdViBCUbGA3r7JmKZyx6uha8_1/view?usp=sharing

5. CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

Después de haber culminado el proceso de ensayos del sistema y el análisis de los resultados obtenidos, se establecen las siguientes conclusiones y recomendaciones, las cuales aportan información relevante para la realización de proyectos similares.

5.1 Conclusiones

- Se realizó la investigación de los actuales algoritmos de Visión Artificial y se evidenció que las prestaciones de la red SSD-MobileNet presenta una respuesta más rápida en el proceso de detección de objetos con respecto a los modelos deeplab-VGG e Inception V2. Adicionalmente la red FCN ResNet-18 obtiene mejores resultados en precisión y sensibilidad con respecto a la red FCN AlexNet en el proceso de segmentación.
- Se desarrolló el sistema para detección del tipo de manzana y segmentación de acuerdo con su grado de calidad, el sistema opera con imágenes estáticas (fotografías). Sin embargo, también presenta la factibilidad de realizar la evaluación en tiempo real, obteniendo en el video de salida 41 FPS y demostrando que el proceso en línea opera adecuadamente con la superposición de las máscaras definidas que indican el nivel de calidad de las manzanas.
- En el desarrollo experimental de laboratorio para detección del tipo de manzana se obtuvieron mejores resultados con el conjunto de prueba de la base de datos 2 en la que el fondo correspondía a una mesa que presentaba textura de madera, obteniendo un 92.6% de precisión, con respecto a los resultados obtenidos con el conjunto de prueba de la base de datos 1 en la que se obtuvo una precisión del 86.7%, es decir el fondo influye directamente en la precisión del sistema de detección.
- En el desarrollo experimental de campo para la detección del tipo de manzana se obtuvo una precisión del 90.4%, teniendo en cuenta que las frutas se

encuentran ubicadas en el árbol, teniendo directa influencia de variabilidad en la iluminación, sombra y oclusiones.

- En el sistema de segmentación que cuantifica de calidad de la manzana se evaluó la precisión del entrenamiento, obteniendo una media IoU (Intersección sobre la Unión) de 54.65% y una precisión de 94.56%.

5.2 Recomendaciones

- Se recomienda realizar ensayos al sistema incrementando la cantidad de imágenes en la base de datos para analizar el crecimiento de la precisión en los resultados de detección y segmentación.

5.3 Trabajo futuro

- Los resultados obtenidos en la investigación realizada permitirán realizar publicaciones de carácter científico en un futuro, debido a que el campo de análisis está basado procesos tecnológicos actuales que aportan al desarrollo de la mano de obra.

5.4 Bibliografía

- Abien, A. (2018). Deep Learning using Rectified Linear Units (ReLU). *arXiv*, 7.
- Aziz, S., Bilal, M., Khan, M. U., & Amjad, F. (2020). Deep Learning-based Automatic Morphological Classification of Leukocytes using Blood Smears. *IEEE*, 1-5.
- Bargoti, S., & Underwood, J. (2017). Deep Fruit Detection in Orchards. *IEEE*, 3626-3633.
- Basheer, I., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *ELSERVIER*, 3-31.
- Basir, S., Chowdhury, M., Islam, N., & Ashik-E-Rabbani, M. (2021). Artificial neural network model in predicting yield of mechanically transplanted rice from transplanting parameters in Bangladesh. *ELSERVIER*, 8.
- Bhargava, A., & Bansal, A. (2018). Fruits and Vegetables Quality Evaluation Using Computer Vision: A Review. *Journal of King Saud University - Computer and Information Sciences*, 243-257.
- Carpio, L. (2018). El uso de la tecnología en la agricultura. *Pro Sciences*, 25-32.
- CEPAL, FAO, & IICA. (2020). *Perspectivas de la agricultura y del desarrollo rural en las Américas: una mirada hacia América Latina y el Caribe*. San José: Comunicaciones Milenio .
- Chen, S., Shivakumar, S., Dcunha, S., Das, J., Okon, E., Qu, C., & Kumar, V. (2017). *Approach*. Philadelphia: Robotics and Automation Letters IEEE.
- Das, J., Cross, G., Qu, C., Makineni, A., Tokekar, P., Mulgaonkar, Y., & Kumar, V. (2015). *Devices, systems, and methods for automated monitoring enabling precision agriculture*. Gotenburgo: IEEE International Conference on Automation Science and Engineering (CASE).

- GOOGLE. (2022, 04 2022). *GOOGLE MAPS*. Retrieved from <https://www.google.com.ec/maps/@-1.2737011,-78.6472987,15.5z?hl=es>
- Hani, N., Roy, P., & Isler, V. (2018). Apple Counting using Convolutional Neural Networks. *IEEE*, 2559-2565.
- Hani, N., Roy, P., & Isler, V. (2020). A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *Journal of Field Robotics*, 37(2), 263-282.
- Hani, N., Roy, P., & Isler, V. (2020). MinneApple: A Benchmark Dataset for Apple Detection and Segmentation. *IEEE Robotics and Automation Letters*, 852 - 858.
- Hossain, A., Uddin, M. N., Hossain, M. A., & Jang, Y. M. (2017). Predicting rice yield for Bangladesh by exploiting weather conditions. *IEEE*, 589-594.
- Hossain, S., Al-Hammadi, M., & Muhammad, G. (2019). *Automatic Fruit Classification Using Deep Learning for Industrial Applications*. New York: IEEE.
- Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., . . . Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv*.
- Hung, C., Underwood, J., Nieto, J., & Sukkarieh, S. (2015). A Feature Learning Based Approach for Automated Fruit Yield Estimation. *9th Conference on Field and Service Robotics*. Cham.
- Ireri, D., Belal, E., Okinda, C., Makange, N., & Ji, C. (2019). A computer vision system for defect discrimination and grading in tomatoes using machine learning and image processing. *Artificial Intelligence in Agriculture*, 28-37.
- Kamilaris, A., & Prenafeta, F. (2018). Deep learning in agriculture: A survey. *ELSERVIER*, 70-90.

- Kang, H., & Chen, C. (2020). *Fruit Detection, Segmentation and 3D Visualisation of Environments in Apple Orchards*. Melbourne: Computers and Electronics in Agriculture.
- Koirala, A., Walsh, K., Wang, Z., & McCarthy, C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. *ELSEVIER*, 219-234.
- Kujawa, S., & Niedbała, G. (2021). Artificial Neural Networks in Agriculture. *AGRICULTURE*, 497.
- Li, Y., Feng, X., Liu, Y., & Han, X. (2021). *Apple quality identification and classification by image processing based on convolutional neural networks*. Shandong: Scientific Reports.
- Liakos, K., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). *Machine Learning in Agriculture: A Review*. Switzerland: PubMed.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Yang Fu, C., & Berg, A. (2016). SSD: Single Shot MultiBox Detector. *SPRINGER*, 21-37.
- Lu,, Y., & Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *EL SERVIER*, 178.
- MathWorks. (2022). *Image Segmentation*. Retrieved from <https://www.mathworks.com/discovery/image-segmentation.html>
- Mavridou, E., Vrochidou, E., Papakostas, G., Pachidis, T., & Kaburlasos, V. (2019). *Machine Vision Systems in Precision Agriculture for Crop Farming*. Basel: Journal of Imaging.
- Ni, X., Li, C., Jiang, H., & Takeda, F. (2020). *Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield*. Georgia: Horticulture research.

- NVIDIA. (2019). *Deploying Deep Learning*. Retrieved from <https://github.com/dusty-nv/jetson-inference>
- NVIDIA. (2019). *Semantic Segmentation with SegNet*. Retrieved from <https://github.com/dusty-nv/jetson-inference/blob/master/docs/segnet-console-2.md>
- NVIDIA. (2020). *Re-training SSD-Mobilenet*. Retrieved from <https://github.com/dusty-nv/jetson-inference/blob/master/docs/pytorch-ssd.md>
- NVIDIA. (2022). *Getting Started with Jetson Nano Developer Kit*. Retrieved from <https://developer.nvidia.com/embedded/learn/get-started-jetson-nano-devkit#intro>
- NVIDIA. (2022). *Jetson AI courses and certifications* . Retrieved from https://developer.nvidia.com/embedded/learn/jetson-ai-certification-programs#course_outline
- NVIDIA. (2022). *Setting up Jetson with JetPack*. Obtenido de <https://github.com/dusty-nv/jetson-inference/blob/master/docs/jetpack-setup-2.md>
- Osco, L., Dos Santos de Aruda, M., Juniot, J., Da Silva, N., Ramos, A., & Moryia, E. (2020). A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 97-106.
- Polikevicius, J. (2022). *Edge AI: Semantic Segmentation on Nvidia Jetson*. Retrieved from <https://www.highvoltagecode.com/post/edge-ai-semantic-segmentation-on-nvidia-jetson>
- Ramzan, F., Ghani, M., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., & Mehmood, Z. (2019). A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State fMRI and Residual Neural Networks. *Springer*, 1-16.

- Rezatofighi, H., Tsoi, N., Gwak, J. Y., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. *IEEE/CVF*, 658-666.
- Rouhiainen, L. (2018). *Inteligencia Artificial*. Barcelona: Alienta .
- Roy, P., Kislak, A., Plonski, P., Luby, J., & Isler, V. (2019). *Vision-based preharvest yield mapping for apple orchards*. Amsterdam: ELSEVIER.
- Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). *DeepFruits: A Fruit Detection System Using Deep Neuronal Networks*. Brisbane: sensors.
- Santos, T., de Souza, L., dos Santos, A., & Avila, S. (2020). Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. *ELSERVIER*, 17.
- Shafi, U., Mumtaz, R., Ali Hassan, S., García, J., Raza, S., & Iqbal, N. (2019). *Precision Agriculture Techniques and Practices: From Considerations to Applications*. Leeds: sensors.
- Shelhamer, E., Long, J., & Darrell, T. (2016). Fully Convolutional Networks for Semantic Segmentation. *IEEE*, 3431-3440.
- The MathWorks, Inc. (2022). *Resnet18*. Retrieved from <https://www.mathworks.com/help/deeplearning/ref/resnet18.html>
- University of Illinois Board of Trustees. (2022). *Manzanas y Máses*. Retrieved from https://web.extension.illinois.edu/apples_sp/varieties.cfm
- Vasconez, J., Delpriano, J., Vougioukas, S., & Cheein, F. (2020). *Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation*. Elsevier Computers and Electronics in Agriculture.
- Villota, P. E. (2019). *Implementación de una estación prototipo con*. Cuenca: Universidad de Cuenca.

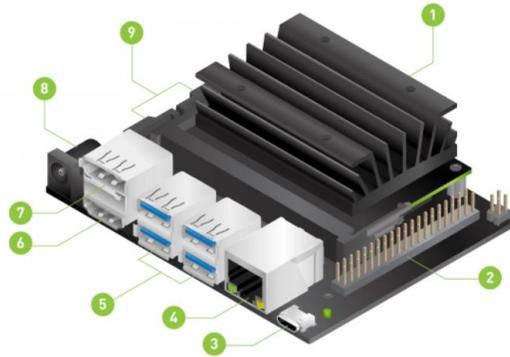
- Wang, Q., Nuske, S., Bergerman, M., & Singh, S. (2013). Automated Crop Yield Estimation for Apple Orchards. *Experimental Robotics*, 745-758.
- Wu, J., Yang, G., Yang, H., Zhu, Y., Li, Z., Lei, L., & Zhao, C. (2020). *Extracting apple tree crown information from remote imagery using deep learning*. Computers and Electronics in Agriculture.
- Wu, Y., & Johnson, J. (2021). Rethinking “Batch” in BatchNorm. *arXiv*, 18.
- YARA Knowledge grows. (2022). *Tipos de frutos de pepita*. Retrieved from <https://www.yara.com.ec/nutricion-vegetal/frutales-de-pepita/pome-fruit-types/>
- Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., & Liu, C. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *ELSERVIER*, 326-343.

ANEXOS

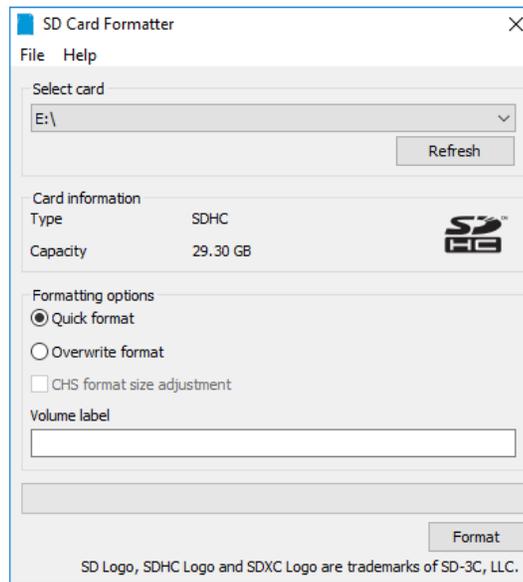
Anexo 1.

Partes de la placa electrónica Jetson Nano

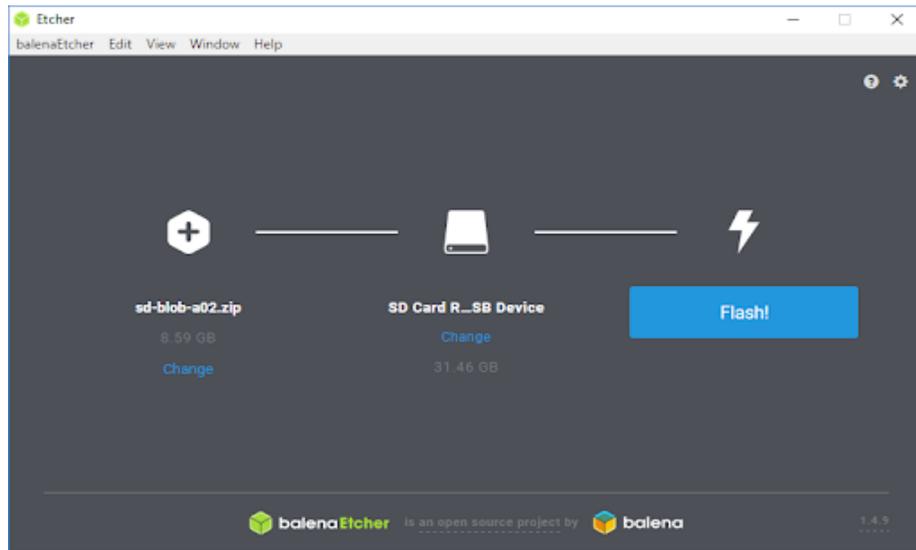
1. Ranura para tarjeta microSD para almacenamiento principal
2. Cabezal de expansión de 40 pines
3. Puerto micro-USB para entrada de alimentación de 5 V o para modo de dispositivo
4. Puerto Ethernet Gigabit
5. Puertos USB 3.0 (x4)
6. Puerto de salida HDMI
7. Conector DisplayPort
8. Conector de barril de CC para entrada de alimentación de 5 V
9. Conectores de cámara MIPI CSI-2



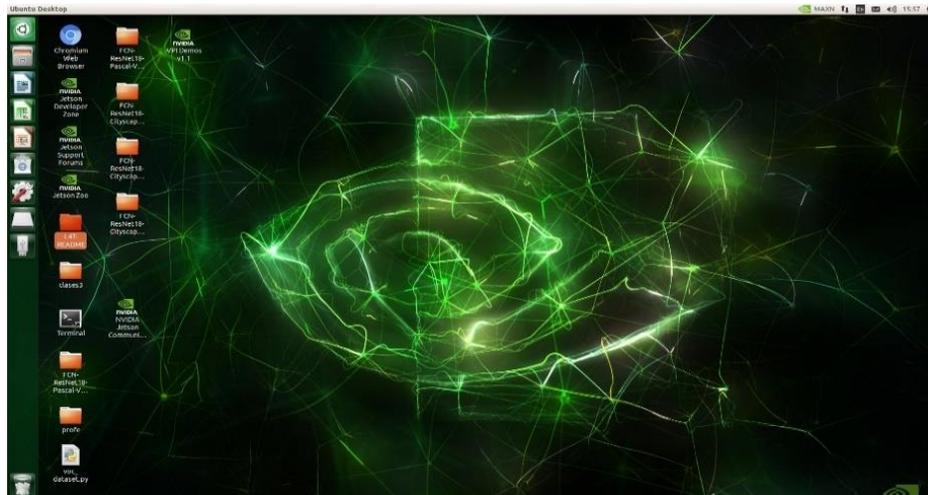
Pantalla para formato de micro SD.



Pantalla de Balena Etcher (Autor).



Pantalla principal del escritorio de la Jetson Nano.



Anexo 2

Código para redimensionamiento y corte de imagen.

```
import cv2
import os
input_images_path = "D:\Manzanas\efinal"
files_names = os.listdir(input_images_path)
print(files_names)
output_images_path = "D:\Manzanas\efinalr"
if not os.path.exists(output_images_path):
    os.makedirs(output_images_path)
    print("Directorio creado: ", output_images_path)
count = 0
for file_name in files_names:
    #print(file_name)
    ""
    if file_name.split(".")[1] not in ["jpeg", "png"]:
        continue
    ""
    image_path = input_images_path + "/" + file_name
    print(image_path)
    image = cv2.imread(image_path)
    if image is None:
        continue
    image = cv2.resize(image, (1344, 1008), interpolation=cv2.INTER_CUBIC)
    cv2.imwrite(output_images_path + "/image" + str(count) + ".jpg", image)
    count += 1
    ""
    cv2.imshow("Image", image)
    cv2.waitKey(0)
cv2.destroyAllWindows()
```

Código para corte de bordes de imagen.

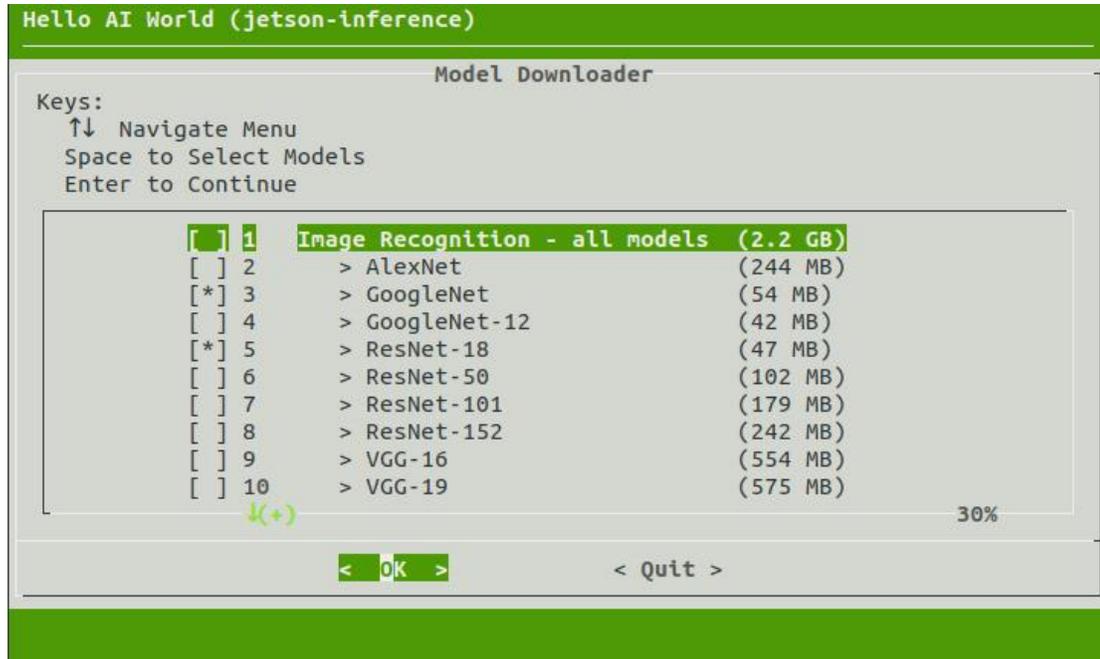
```
import cv2
import os

input_images_path = "D:\Manzanas\Procesar\PorCorte"
files_names = os.listdir(input_images_path)
print(files_names)
output_images_path = "D:\Manzanas\Procesar\Cortadas"
if not os.path.exists(output_images_path):
    os.makedirs(output_images_path)
    print("Directorio creado: ", output_images_path)
count = 0
for file_name in files_names:
    #print(file_name)
    '''
    if file_name.split(".")[1] not in ["jpeg", "png"]:
        continue
    '''
    image_path = input_images_path + "/" + file_name
    print(image_path)
    image = cv2.imread(image_path)
    if image is None:
        continue
    image= image[552:2472, 1476:2586]

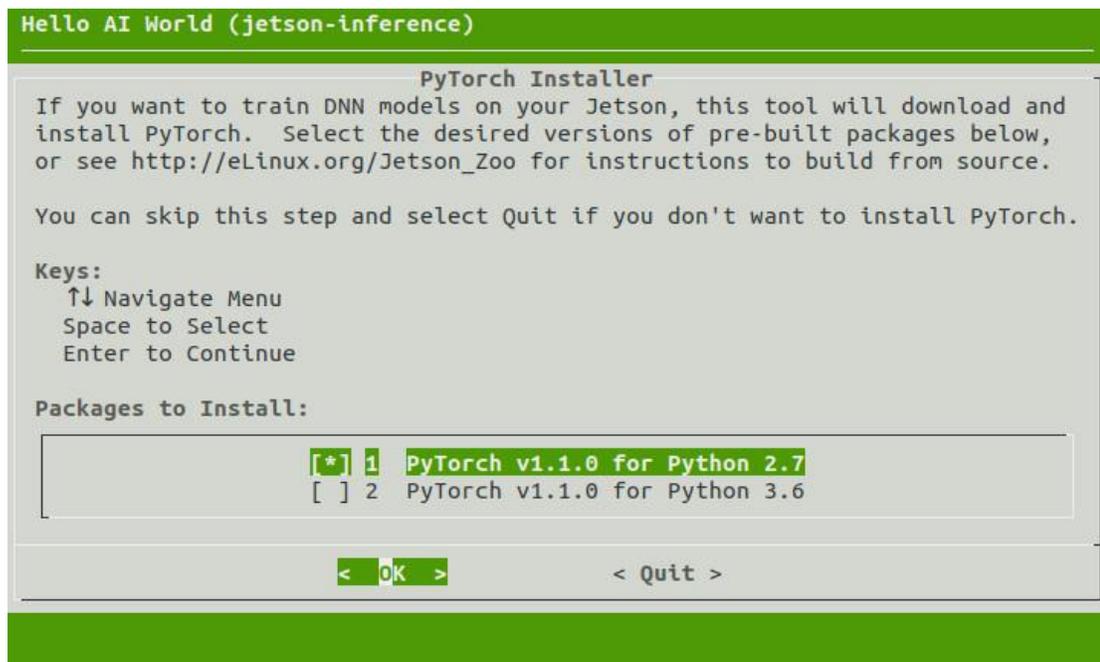
    cv2.imwrite(output_images_path + "/image" + str(count) + ".jpg",
image)
    count += 1
    '''
    cv2.imshow("Image", image)
    cv2.waitKey(0)
cv2.destroyAllWindows()
```

Anexo 3.

Pantalla de proceso de descarga de modelos en Jetson Nano



Pantalla de proceso de instalación de Pytorch para Python en Jetson Nano



Proceso de entrenamiento de modelo en Jetson Nano

```
jetson@jetson-desktop: ~/jetson-inference/python/training/detection/ssd
2022-03-20 16:15:38 - Took 0.51 seconds to load the model.
2022-03-20 16:15:55 - Learning rate: 0.01, Base net learning rate: 0.001, Extra
Layers learning rate: 0.01.
2022-03-20 16:15:55 - Uses CosineAnnealingLR scheduler.
2022-03-20 16:15:55 - Start training from epoch 0.
/home/jetson/.local/lib/python3.6/site-packages/torch/optim/lr_scheduler.py:123:
UserWarning: Detected call of `lr_scheduler.step()` before `optimizer.step()`.
In PyTorch 1.1.0 and later, you should call them in the opposite order: `optimiz
er.step()` before `lr_scheduler.step()`. Failure to do this will result in PyTo
rch skipping the first value of the learning rate schedule. See more details at
https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate
  "https://pytorch.org/docs/stable/optim.html#how-to-adjust-learning-rate", User
Warning)
/home/jetson/.local/lib/python3.6/site-packages/torch/nn/_reduction.py:44: UserW
arning: size_average and reduce args will be deprecated, please use reduction='s
um' instead.
  warnings.warn(warning.format(ret))
2022-03-20 16:16:58 - Epoch: 0, Step: 10/895, Avg Loss: 14.2659, Avg Regression
Loss 5.8286, Avg Classification Loss: 8.4373
2022-03-20 16:17:15 - Epoch: 0, Step: 20/895, Avg Loss: 12.6225, Avg Regression
Loss 6.2365, Avg Classification Loss: 6.3860
2022-03-20 16:17:21 - Epoch: 0, Step: 30/895, Avg Loss: 10.1493, Avg Regression
Loss 4.6815, Avg Classification Loss: 5.4678
```

Anexo 4

Procesamiento de imágenes en PC.

```
Anaconda Prompt (miniconda3) - python labelme2voc.py C:\pytorch-segmentation\total C:\pytorch-segmentation\total\new --labels C:\pytorch-seg...
cls_id = label_name_to_value[cls_name]
KeyError: 'media'

(segmentations) C:\pytorch-segmentation>python labelme2voc.py C:\pytorch-segmentation\total C:\pytorch-segmentation\total\new --labels C:\pytorch-segmentation\classes.txt --noviz
Creating dataset: C:\pytorch-segmentation\total\new
class_names: ('background', 'alta', 'media', 'baja')
Generating dataset from: C:\pytorch-segmentation\total\image1037.json
Generating dataset from: C:\pytorch-segmentation\total\image1041.json
Generating dataset from: C:\pytorch-segmentation\total\image1042.json
Generating dataset from: C:\pytorch-segmentation\total\image1045.json
Generating dataset from: C:\pytorch-segmentation\total\image1047.json
Generating dataset from: C:\pytorch-segmentation\total\image1053.json
Generating dataset from: C:\pytorch-segmentation\total\image1062.json
Generating dataset from: C:\pytorch-segmentation\total\image1063.json
Generating dataset from: C:\pytorch-segmentation\total\image1068.json
Generating dataset from: C:\pytorch-segmentation\total\image107 (2).json
Generating dataset from: C:\pytorch-segmentation\total\image107.json
Generating dataset from: C:\pytorch-segmentation\total\image1073.json
Generating dataset from: C:\pytorch-segmentation\total\image1076.json
Generating dataset from: C:\pytorch-segmentation\total\image1077.json
Generating dataset from: C:\pytorch-segmentation\total\image1078.json
Generating dataset from: C:\pytorch-segmentation\total\image108.json
Generating dataset from: C:\pytorch-segmentation\total\image1106.json
Generating dataset from: C:\pytorch-segmentation\total\image111.json
Generating dataset from: C:\pytorch-segmentation\total\image112.json
Generating dataset from: C:\pytorch-segmentation\total\image113.json
Generating dataset from: C:\pytorch-segmentation\total\image114.json
Generating dataset from: C:\pytorch-segmentation\total\image115.json
```

Ejemplo de las imagen con máscara generadas en formato JPG.

