



# UNIVERSIDAD AUTÓNOMA CHAPINGO

## POSGRADO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA

DESARROLLO DE UN SISTEMA DE VISIÓN ARTIFICIAL  
PARA EL MONITOREO DE PRODUCCIÓN DE JITOMATE  
EN INVERNADERO

### TESIS

COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE:

**MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL  
DEL AGUA**

PRESENTA:

**CHRISTIAN ISRAEL ZUÑIGA AVILA**

BAJO LA SUPERVISIÓN DE:

**DR. JOSÉ ARMANDO RAMÍREZ ÁRIAS**



DIRECCION GENERAL ACADEMICA  
DIRECCION DE SERVICIOS ESCOLARES  
OFICINA DE EXAMENES PROFESIONALES

Chapingo, Estado de México, diciembre de 2019



**DESARROLLO DE UN SISTEMA DE VISIÓN ARTIFICIAL PARA EL  
MONITOREO DE PRODUCCIÓN DE JITOMATE EN INVERNADERO**

Tesis realizada por el **ING. CHRISTIAN ISRAEL ZÚÑIGA ÁVILA** bajo la supervisión del Comité Asesor indicado, aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

**MAESTRO EN INGENIERÍA AGRÍCOLA Y USO INTEGRAL DEL AGUA**

DIRECTOR: \_\_\_\_\_

  
DR. JOSÉ ARMANDO RAMÍREZ ÁRIAS

ASESOR: \_\_\_\_\_

  
DR. EFRÉN FITZ RODRÍGUEZ

ASESOR: \_\_\_\_\_

  
DRA. RAQUEL SALAZAR MORENO

# CONTENIDO

CONTENIDO .....	iii
LISTA DE CUADROS .....	vi
LISTA DE FIGURAS .....	vii
DEDICATORIA .....	ix
AGRADECIMIENTOS .....	x
DATOS BIOGRÁFICOS.....	xi
RESUMEN GENERAL.....	xii
ABSTRACT.....	xiii
1 INTRODUCCIÓN GENERAL.....	14
1.1 Objetivos.....	17
1.1.1 Objetivo general .....	17
1.1.2 Objetivos específicos .....	17
1.2 Hipótesis.....	17
1.3 Estructura de la tesis .....	17
1.3.1 Revisión de literatura.....	18
1.3.2 Detección y clasificación de frutos de jitomate de invernadero con visión artificial y redes neuronales de convolución .....	18
1.3.3 Estimación de peso de frutos de jitomate con visión artificial y redes neuronales artificiales .....	19
2 REVISIÓN DE LITERATURA .....	20
2.1 El cultivo de jitomate ( <i>Solanum lycopersicum</i> L.).....	20
2.2 Sistemas de producción en invernadero.....	20
2.3 Visión artificial.....	21
2.3.1 Dispositivos de visión .....	21
2.3.2 Procesamiento de imágenes.....	23
2.3.3 Algoritmos para la clasificación de frutos de jitomates .....	23
2.3.4 Red neuronal de convolución.....	24
2.4 Estimación del peso de frutos de jitomate .....	29
2.5 Conceptos clave .....	30
2.6 Referencias.....	32

3	DETECCIÓN Y CLASIFICACIÓN DE FRUTOS DE JITOMATE DE INVERNADERO CON VISIÓN ARTIFICIAL Y REDES NEURONALES DE CONVOLUCIÓN.....	38
	RESUMEN.....	39
3.1	Introducción .....	40
3.2	Materiales y Métodos.....	43
3.2.1	Localización .....	43
3.2.2	Sistema de visión artificial .....	43
3.2.3	Selección de la muestra .....	49
3.2.4	Captura de imágenes .....	50
3.2.5	Preprocesamiento .....	51
3.2.6	Evaluación del algoritmo .....	53
3.3	Resultados y Discusión.....	54
3.3.1	Detección de frutos .....	54
3.3.2	Clasificación de jitomates por estado de madurez. ....	57
3.3.3	Segmentación de frutos .....	59
3.4	Conclusiones .....	60
3.5	Referencias.....	61
4	ESTIMACIÓN DE LA MASA EN FRUTOS DE JITOMATE CON VISIÓN ARTIFICIAL Y REDES NEURONALES ARTIFICIALES. ....	65
	RESUMEN.....	66
4.1	Introducción .....	67
4.2	Materiales y Métodos.....	69
4.2.1	Sistema de visión artificial .....	69
4.2.2	Funcionamiento general del sistema.....	70
4.2.3	Medición de características geométricas de los frutos .....	70
4.2.4	Captura de imágenes .....	72
4.2.5	Extracción de características geométricas .....	73
4.2.6	Estimación de la masa de los frutos.....	77
4.3	Resultados y Discusión.....	83
4.3.1	Método 1: Regresión polinomial utilizando una variable de entrada 83	
4.3.2	Método 2: ANN para la estimación de la masa de los frutos con mediciones físicas como parámetros de entrada.....	85

4.3.3	Método 3: ANN para la estimación de la masa de frutos a partir de características obtenidas del procesamiento de imágenes digitales como parámetros de entrada .....	87
4.3.4	Estimación de diámetros .....	93
4.3.5	Método 4: Modelo ANN para la estimación del diámetro ecuatorial y la obtención de la masa de frutos con la regresión del Método 1 .....	94
4.3.6	Método 5: Modelo ANN para la estimación de los diámetros polar y ecuatorial del fruto y la obtención de la masa del fruto con la red del punto 2. 94	
4.3.7	Comparación enfoques para la predicción de masa .....	95
4.4	Conclusiones .....	96
4.5	Referencias.....	97

## LISTA DE CUADROS

Cuadro 1. Grado de madurez de acuerdo a la coloración de jitomate. Norma Mexicana NMX-FF-031-1997.....	49
Cuadro 2. Distribución de frutos por grado de madurez .....	50
Cuadro 3. Número de imágenes etiquetadas por clase .....	52
Cuadro 4. Métricas de desempeño en los diferentes escenarios.....	56
Cuadro 5. Matriz de confusión. ....	58
Cuadro 6. Distribución de frutos por estado de maduración .....	71
Cuadro 7. Desempeño de la ANN durante cada una de las etapas de diseño. ....	85
Cuadro 8. Desempeño ANN en el método 3, con imágenes tomadas en condiciones controladas .....	88
Cuadro 9. Desempeño ANN para estimación de la masa con imágenes de invernadero.....	89
Cuadro 10. Desempeño ANN que utilizó imágenes durante la creación de la red .....	92
Cuadro 11. Desempeño de las ANNs para estimar diámetros .....	94
Cuadro 12. ANN para la estimación del peso en función diámetro estimado y la regresión polinomial.....	94
Cuadro 13. Estimación de peso con red de mediciones físicas e imágenes ....	95
Cuadro 14. Medidas de ajuste con un nuevo set de datos .....	95

## LISTA DE FIGURAS

Figura 1. Estimación de profundidad .....	22
Figura 2. Red neuronal de convolución aplicada en la clasificación de frutos de jitomate .....	25
Figura 5. Diagrama de funcionamiento del sistema de visión .....	45
Figura 6. Arquitectura de una red neuronal de convolución.....	46
Figura 7 Principales operaciones en las CNNs.....	46
Figura 8. Estructura de la red Faster R-CNN implementada. ....	47
Figura 9. Estructura de una red Mask R-CNN .....	48
Figura 10. Grados de madurez del fruto de jitomate.....	49
Figura 11. Detección de frutos individuales en condiciones de laboratorio.....	54
Figura 12. Detección de frutos individuales, con iluminación natural y fondo no uniforme .....	55
Figura 13. Detección de frutos en racimo con iluminación natural y fondo no uniforme .....	55
Figura 14. Detección de frutos en racimos .....	57
Figura 15. Curvas ROC bajo condiciones de invernadero .....	58
Figura 16. Segmentación de frutos: correcta (a, b) e incorrecta (c) .....	59
Figura 17. Esquema de funcionamiento del sistema de visión .....	70
Figura 18. Diámetros del fruto.....	71
Figura 19. Estados de madurez del fruto de jitomate.....	71
Figura 20. Estimación de profundidad .....	74
Figura 21. Estimación de la distancia promedio ( $D_p$ ).....	75
Figura 22. Proceso de obtención de la máscara.....	76
Figura 23. Diagrama de los cinco enfoques propuestos para la estimación de la masa de frutos de jitomate.....	78

Figura 24. Diagrama de flujo del procedimiento para selección de la arquitectura óptima de la ANN.....	82
Figura 25. Matriz de correlación de mediciones físicas: diámetro polar ( $D_1$ ), diámetro ecuatorial ( $D_2$ y $D_3$ ) y estado de maduración (S).....	84
Figura 26. Regresión polinomial de la masa del fruto en función del diámetro ecuatorial ( $D_2$ ) .....	84
Figura 27. Eficiencia de las ANNs propuestas, en función del número de nodos en la capa oculta.....	85
Figura 28. Importancia relativa de las variables medidas de acuerdo con el análisis de Garson. ....	86
Figura 29. Evaluación de una segunda aproximación: Relación masa (g) estimada con la red neural vs datos de medición de frutos .....	87
Figura 30. Evaluación de la aproximación 3 en condiciones de laboratorio: Correlación de masa (g) estimada con la red neural vs datos de medición de frutos.....	88
Figura 31. Eficiencia en cada una de las ANN de acuerdo con el número de nodos en la capa oculta.....	89
Figura 32. Importancia relativa de las características extraídas de las imágenes .....	90
Figura 33. Eficiencia en diversos conjuntos de entradas .....	91
Figura 34. Eficiencia de la ANN en función del número de neuronas en la capa oculta. ....	91
Figura 35. Eficiencia con el incremento de nodos en la segunda capa oculta. .	92
Figura 36. Evaluación de la ANN para la aproximación 3, en condiciones de invernadero: Relación masa (g) estimada con la red neural vs datos de medición de frutos.....	93
Figura 37. Masa de frutos, real vs estimado. Red ANN Imágenes .....	96

## **DEDICATORIA**

A mi madre

A ti, por tejarme alas y enseñarme a volar.

## AGRADECIMIENTOS

A la **Universidad Autónoma Chapingo** y al posgrado en **Ingeniería Agrícola y Uso Integral del Agua** por la formación académica recibida.

Al **Consejo Nacional de Ciencia y Tecnología (CONACyT)** por la beca otorgada para seguir mi formación en la maestría.

A mí comité asesor el **Dr. José Armando Ramírez Árias**, el **Dr. Efrén Fitz Rodríguez** y la **Dr. Raquel Salazar Moreno** que con su sabiduría, conocimiento y apoyo me motivaron a desarrollarme como persona y profesionalista.

A la **Dirección General de Investigación y Posgrado (DGIP-UACH)** por el financiamiento otorgado en la modalidad de Proyectos Convencionales (18395-C-65 y 19293-C-65) y Estratégicos Institucional (18013-EI y 19010-EI) con el cual esta investigación fue posible.

Al **Dr. Efrén Fitz Rodríguez**, por incluirme en su equipo de trabajo en el **Laboratorio de Modelación y Automatización de Biosistemas**, y por permitirme contribuir en su proyecto de desarrollo de Invernaderos Sustentables para México.

A **Harley** y **Mily** por su compañía, el cariño y el apoyo que me han brindado.

A mis amigos y en especial a **Mario** por la bonita amistad que ha surgido.

A mis compañeros y amigos del **Laboratorio de Automatización y Modelación de Biosistemas** por la amistad y sus consejos prestados.

A mi familia por el apoyo que siempre me han brindado, en especial a mi madre que con su ejemplo me ha enseñado a no rendirme jamás.

## **DATOS BIOGRÁFICOS**

### **Datos Personales**

Nombre: Christian Israel Zúñiga Ávila

Fecha de nacimiento: 08 de agosto de 1992

Lugar de nacimiento: Tecomán, Colima

No. Cartilla militar: D-1409285

CURP: ZUAC920808HCMXVH02

Profesión: Ingeniero Mecatrónico

Cédula profesional: 10062083



### **Desarrollo Académico**

Inició sus estudios de preparatoria en el “Bachillerato Técnico N. 20 de la Universidad de Colima” en el área de Técnico Agropecuario. Ingresó en el año 2010 al Instituto Tecnológico de Colima en la capital del estado, donde cursa la carrera de Ing. Mecatrónica egresando en el año 2015, para posteriormente titularse por medio de la modalidad de informe técnico de residencias con el nombre de “Integración de los sistemas de una línea de envasado de miel de agave” bajo la dirección del M.C. Marco Aurelio Vázquez Olvera. A partir del año 2016 trabajó en diferentes empresas privadas, enfocado principalmente en el área de mantenimiento preventivo. En agosto del año 2017, ingresó al Posgrado en Ingeniería Agrícola y Uso Integral del Agua de la Universidad Autónoma Chapingo, obteniendo el grado en diciembre de 2019.

## RESUMEN GENERAL

### DESARROLLO DE UN SISTEMA DE VISIÓN ARTIFICIAL PARA EL MONITOREO DE PRODUCCIÓN DE JITOMATE EN INVERNADERO.

El monitoreo continuo de cultivos en invernadero, en conjunto con la automatización de sus procesos, permite hacer un uso eficiente de los recursos disponibles, tales como agua, energía, fertilizantes, suelo, mano de obra, entre otros; así como también mejorar la calidad, uniformidad y rendimientos en la producción. Con el avance tecnológico en los sistemas de visión artificial, actualmente se puede obtener información útil del estado de los cultivos, de manera no invasiva. En esta investigación se propone implementar un sistema de visión artificial para cuantificar el número, tamaño y estado de madurez de los frutos en un cultivo de jitomate bajo condiciones de invernadero. Esta información resulta útil para el pronóstico de rendimientos semanales y poder realizar una mejor planeación, tanto en el manejo de las cosechas, como en mejores estrategias de mercadotecnia. Se implementó un sistema de visión con una cámara estéreo RGB-D y algoritmos con redes neuronales de convolución para la detección y clasificación por grado de madurez de los frutos de jitomate. Con el sistema desarrollado se obtuvo una sensibilidad del 70 % (capacidad para identificar los frutos) y una precisión de 96 % (la probabilidad de que un objeto identificado sea un fruto de jitomate), en la detección de frutos bajo condiciones de invernadero. Por otro lado, se obtuvo una precisión del 94 % en la clasificación correcta de los frutos por estado de madurez. Para la estimación del peso de los frutos se implementaron varios algoritmos de procesamiento de imágenes para obtener características geométricas, las cuales se utilizaron como entradas en una red neuronal multicapa. En este último sistema se obtuvo una raíz del error cuadrático medio de 11.85 g, con un coeficiente de determinación  $R^2 = 0.88$ .

**Palabras clave:** Detección y clasificación de frutos, Estimación de peso de frutos, Procesamiento de imágenes digitales, Redes neuronales de convolución, Aprendizaje de máquina, Aprendizaje profundo

---

Tesis de Maestría en Ingeniería, Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo

Autor: Christian Israel Zúñiga Ávila

Director de Tesis: Dr. José Armando Ramírez Arias

## ABSTRACT

### DEVELOPMENT OF AN ARTIFICIAL VISION SYSTEM FOR MONITORING GREENHOUSE TOMATO PRODUCTION.

The continuous monitoring of greenhouse crops, along with the automation of their processes, allows for an efficient usage of resources, such as water, energy, fertilizers, soil and workforce, among others; as well as to improve the quality, uniformity and long term yield. With the current technological advances on machine vision systems, useful information on the state of crops can be obtained in a non-invasive way. This research aims to implement an artificial vision system to quantify the number, size and ripeness stage of tomato fruits grown under greenhouse conditions. This information is useful for predicting weekly harvest rates, and then be able to make better decisions on crop management and improved marketing strategies. For the detection and classification of fruits by its ripeness stage, an RGB-D stereo vision camera as well as algorithms with convolutional neural networks were implemented. With the system developed, a 70 % sensibility (capacity of identifying fruits), and a 96 % precision (probability that the identified fruit was in fact a tomato) in the detection of tomato fruits under greenhouse conditions were obtained. On the other hand, on the proper classification of fruits by maturity stage, a 94 % of precision was obtained. For fruit weight estimation, several image processing algorithms were implemented to obtain geometric features. This information was used as input on a multilayer neural network. With this model a root mean square error of 11.85 g was obtained, with a coefficient of determination  $R^2 = 0.88$ .

**Keywords:** *Fruit identification and classification, Fruit weight estimation, Digital image processing, Convolutional neural networks, Machine learning, Deep learning.*

---

Master's degree thesis on engineering, Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo

Author: Christian Israel Zúñiga Ávila

Advisor: PhD José Armando Ramírez Arias

# 1 INTRODUCCIÓN GENERAL

El cultivo de jitomate (*Solanum lycopersicum L.*) ha tenido una amplia difusión mundial, ha modificado las dietas básicas en varios continentes y se ha colocado entre las hortalizas de mayor demanda en el mercado internacional (Cortés, 2015). Hoy en día, el jitomate no solo se consume fresco, también se utiliza de manera industrial en la elaboración de pastas, sopas, jugos, salsas, polvos, etc. La producción mundial superó los 182 millones de toneladas en 2017, donde los diez principales países productores que concentran más del 75 % de la producción total mundial son: China, India, Turquía, Estados Unidos, Egipto, Irán, Italia, España, México y Brasil (FAO, 2017). Durante el 2017 se estima que México tenía alrededor de 50,373 ha de jitomate, una cifra menor comparada con diez años antes donde se registraron 66,635 ha (SIAP, 2017). En ese período, la superficie cultivada en campo abierto se redujo a una tasa promedio anual de 5.9 por ciento, al pasar de 64,663 a 35,175 hectáreas, mientras que la superficie establecida con agricultura protegida pasó de 1,973 a 15,198 hectáreas, es decir, creció a una tasa promedio anual de 22.7 por ciento (FIRA, 2019). La producción obtenida en agricultura protegida pasó de 0.9 por ciento del total en 2003 a 32.2 por ciento en 2010, y hasta 63.3 por ciento del volumen total en 2017 (FIRA, 2019).

Para mejorar la productividad de los cultivos de manera eficiente, es necesario el monitoreo continuo de éstos, y de las condiciones ambientales (Naveen et al., 2018). El monitoreo en la agricultura no es un concepto nuevo, de hecho, éste ya era utilizado en el antiguo Egipto a las orillas del río Nilo (Atzberger, 2013). Los sistemas automatizados de producción en invernadero a menudo requieren del monitoreo del crecimiento de los cultivos, que involucra la cuantificación precisa de las propiedades fisiológicas de la planta (Ryu et al., 2000). El monitoreo se considera parte de la agricultura de precisión donde ha sido ampliamente estudiado en riego, producción, detección de enfermedades, detección de malezas, nutrición, entre otros (De Baerdemaeker, 2000). El desarrollo de la automatización en la agricultura conduce a un monitoreo efectivo de cultivos,

reduciendo la intervención humana en campo (Naveen et al., 2018). Con el monitoreo automatizado se permite realizar labores repetitivas, cuidando la calidad y velocidad de las tareas a realizar.

El avance de la tecnología, como lo es la visión por computadora y el procesamiento de imágenes digitales han demostrado ser herramientas efectivas y no destructivas para el análisis de objetos en el sector agrícola (Pandurng et al., 2015). La implementación de este tipo de tecnologías se debe principalmente a la reducción del costo de los equipos, el aumento de la capacidad computacional y el creciente interés en métodos no destructivos para la evaluación productos y procesos (Mahajana et al., 2015). De acuerdo a Barbedo (2016) estas técnicas presentan ventajas comparadas con los métodos tradicionales basados en trabajo manual, sin embargo, todavía tienen una seria de desafíos que superar. La variación en la intensidad y posición del sol es uno de los retos para el uso de sistemas de visión artificial como dispositivos de monitoreo bajo condiciones reales (Ehret et al., 2001).

La información que se genera en la agricultura moderna, por medio de diferentes sensores o con el procesamiento de imágenes, durante el monitoreo de cultivos o procesos, permite comprender mejor el ambiente de trabajo, lo que conduce a una toma de decisiones más precisa y rápida (Liakos et al., 2018). Por otro lado, con algoritmos de aprendizaje de máquina (machine learning) se pueden analizar grandes volúmenes de información, independientemente de la complejidad, de forma más rápida y precisa (Patrício et al., 2018). Esta técnica está definida como un campo de la ciencia, en el que se otorga a las computadoras la “habilidad de aprender”, sin ser estrictamente programadas (Samuel, 2000). De acuerdo con Patrício et al. (2018) la unión de la visión por computadora, el aprendizaje de máquina y las nuevas tecnologías para el procesamiento de la información, han demostrado ser técnicas prometedoras para resolver problemas en la agricultura. Una clase de redes neuronales artificiales (ANNs), son los algoritmos de machine learning de tipo supervisados, y comúnmente se utilizan en problemas de regresión, clasificación e identificación. Además de las redes neuronales tradicionales, incluyendo las estáticas (perceptrón multicapa) y las dinámicas,

actualmente están tomando auge las redes de aprendizaje profundo “Deep ANNs”, como las redes neuronales de convolución (CNNs) desarrolladas para trabajar con imágenes (Liakos et al., 2018). Una ANN está compuesta por nodos conectados mediante enlaces que están organizados en capas. Estos nodos tienen factores de peso asociados que determinan la “fuerza” de la señal de la conexión. Los pesos de los nodos se actualizan durante el proceso de entrenamiento de la red. Las ANNs reciben una serie de datos en la capa de entrada y éstos son procesados por las capas ocultas, para finalmente enviar las señales a la capa de salida (Russell et al., 2016).

Las Deep ANNs son conocidas comúnmente como redes neuronales profundas (DNNs) (LeCun et al., 2015). Las DNNs difieren de las ANNs convencionales en el sentido de que tienen más nodos, son más complejas y requieren más capacidad computacional. Existen cuatro principales arquitecturas de redes DNNs: 1) Redes pre-entrenadas no supervisadas, 2) redes neuronales recurrentes, 3) redes neuronales recursivas y 4) redes neuronales de convolución (Patrício et al., 2018). Para la clasificación y reconocimiento de patrones en imágenes, las redes neuronales de convolución (CNN) son las más utilizadas. El propósito de este tipo de redes es “aprender” características de los datos a través de las operaciones de convolución (Patrício et al., 2018). Ma et al. (2018) desarrollaron una CNN para la identificación de enfermedades en pepino con imágenes capturadas bajo condiciones de campo. En el estudio reportan que, con este tipo de redes, obtuvieron resultados superiores a los métodos convencionales para la identificación de objetos, en contraste con otros algoritmos como árboles de decisión (decision trees) y las máquinas de soporte vectorial (support vector machine). Por otro lado, Partel et al. (2019) utilizaron las CNNs para la detección de plagas en cítricos, obteniendo una precisión de 80 % y una sensibilidad de 95 %.

El monitoreo en la producción de jitomates de invernadero con sistemas de visión artificial, que implemente CNNs para detectar y clasificar los frutos jitomates y posteriormente estimar el peso con una ANN, resulta de gran interés. De acuerdo con Yamamoto et al. (2014) la información obtenida tiene diversas aplicaciones

como: 1) beneficios económicos a los productores al permitirles una mejor planificación del manejo del cultivo y de las cosechas, 2) permite conocer la capacidad de almacenaje requerida para los frutos cosechados y 3) permite desarrollar mejores estrategias para la colocación de los productos en el mercado. Liu et al. (2019) comentan que el avance en la detección automática de frutos, mejora la eficiencia de los robots cosechadores.

## **1.1 Objetivos**

### **1.1.1 Objetivo general**

Desarrollar un sistema de visión artificial para el monitoreo en la producción de jitomate bajo condiciones de invernadero.

### **1.1.2 Objetivos específicos**

- Detectar los racimos de frutos en un cultivo de jitomates en sistemas de producción bajo invernadero.
- Clasificar los frutos de jitomate de acuerdo con el estado de maduración.
- Estimar el peso de cada uno de los frutos jitomate.

## **1.2 Hipótesis**

Con un sistema de visión artificial se pueden detectar, bajo condiciones ambientales de producción en invernaderos, la ubicación de los racimos de jitomates, cuantificar los frutos en estos, clasificarlos de acuerdo con su grado de madurez y estimar su peso, mediante el procesamiento de imágenes digitales y la implementación de redes neuronales.

## **1.3 Estructura de la tesis**

Este documento está dividido en 4 capítulos, los cuales incluyen: 1) Una introducción general, 2) una revisión del estado del arte y las contribuciones principales que se incluyen en dos manuscritos sometidos para su publicación. Las contribuciones están incluidas en los capítulos 3) Detección y clasificación de frutos de jitomate con visión artificial y redes neuronales de convolución y 4)

Estimación de peso de frutos de jitomate con visión artificial y redes neuronales artificiales.

### **1.3.1 Revisión de literatura**

En el capítulo dos se describen los fundamentos más importantes relacionados con el tema de investigación. Estos incluyen, la descripción física de los frutos de jitomate, las características más importantes, la importancia económica y la clasificación del estado de maduración por color, con referencia a la Norma Mexicana “NMX-FF-031-1997”. Posteriormente se describe la agricultura protegida, la importancia y los objetivos de estos sistemas de producción. Además, en el capítulo se presenta la descripción de la visión artificial, la importancia, aplicaciones en sistemas agrícolas, los principales retos que presenta la aplicación en condiciones no controladas y el uso de procesamiento de imágenes digitales, como método para la descripción geométrica de objetos. Finalmente se tratan los temas relacionados con el procesamiento de la información con aprendizaje de máquina (machine learning) y la aplicación de las redes neuronales artificiales como métodos de predicción de la masa de los frutos, comparados con métodos tradicionales como las regresiones.

### **1.3.2 Detección y clasificación de frutos de jitomate de invernadero con visión artificial y redes neuronales de convolución**

El capítulo tres incluye un manuscrito sometido para su publicación, en el cual se describe el proceso utilizado para la detección y posteriormente la clasificación de acuerdo con el estado de madurez de los frutos de jitomate; comenzando con la creación de una base de datos de frutos de jitomates, la captura de las imágenes en condiciones controladas de luz y fondo uniforme, el etiquetado de imágenes, el uso de redes neuronales de convolución, la detección de los frutos, la segmentación de éstos utilizando una red del tipo Mask R-CNN y finalmente la clasificación de los mismos por su estado de madurez, en diferentes escenarios.

### **1.3.3 Estimación de peso de frutos de jitomate con visión artificial y redes neuronales artificiales**

En el capítulo cuatro se incluye el manuscrito de la investigación realizada para estimar la masa de los frutos de jitomate, partiendo de la selección de un muestreo de frutos de jitomates, la captura de imágenes, el procesamiento de las imágenes para la extracción de características geométricas, la selección de las características más importantes para la estimación del peso. Finalmente se describen cinco métodos propuestos para predecir el peso, utilizando regresiones simples y redes neuronales artificiales.

## 2 REVISIÓN DE LITERATURA

### 2.1 El cultivo de jitomate (*Solanum lycopersicum L.*)

En México, el cultivo de jitomate es la principal hortaliza que se produce, dado por la superficie que ocupa, por el volumen en el mercado nacional y es la que más divisas genera por su exportación. El jitomate, conocido también como tomate, es un producto muy consumido, siendo una importante materia prima para la industria de la transformación. Además, el cultivo de jitomate tiene importancia mundial por razones económicas, gastronómicas y nutricionales.

El jitomate es una planta perenne, de porte arbustivo que se cultiva de forma anual en campo abierto y durante todo el año en invernaderos, con variedades de crecimiento indeterminado, la cual puede desarrollarse de forma rastrera, semirrecta o erecta. El tallo tiene un diámetro de 2 a 4 cm en la base, sobre el tallo se desarrollan hojas, tallos secundarios e inflorescencias. La flor es perfecta o hermafrodita, regular e hipógina, consta de cinco o más sépalos y de seis o más pétalos; tiene un pistilo con cinco estambres unidos en las anteras y formando un tubo que encierra el pistilo. Las hojas se disponen de forma alternativa sobre el tallo (Cortés, 2015).

El fruto es una baya que presenta diferentes tamaños, forma, color, consistencia, y composición, según la variedad. Está constituido por la epidermis o la piel, la pulpa, el tejido placentario y las semillas. Internamente los frutos están divididos en lóculos, que pueden ser bi, tri, tetra, o pluriloculares. Los frutos maduros pueden ser rojos, rosados o amarillos. La maduración del fruto puede presentarse de manera no uniforme (Cortés, 2015).

### 2.2 Sistemas de producción en invernadero

El invernadero es una estructura cerrada con una cubierta y paredes de forma plana y curva, transparente, en el que es posible controlar las condiciones ambientales como: temperatura del aire, concentración de humedad, fotoperiodos, intensidad luminosa, concentración de CO<sub>2</sub> atmosférico, la frecuencia de riegos y la nutrición del cultivo (Dubois, 1980). Los jitomates

cultivados en invernaderos tienen rendimientos superiores a los cultivados a campo abierto, además de que la fructificación se distribuye en periodos más largos y se gana en precocidad (Strijboseh, 1969).

### **2.3 Visión artificial**

La visión artificial, o también denominada visión por computadora, es considerada una rama de la inteligencia artificial que tiene como propósito principal que las maquinas sean capaces de interpretar el entorno que las rodea tal y como lo hacen los seres humanos mediante el sentido de la vista (Sevillano, 2015).

#### **2.3.1 Dispositivos de visión**

Existen una gran cantidad de dispositivos de visión, incluyendo: cámaras RGB, térmicas, espectrales, etc. para la detección de objetos y estimación de sus características. Es necesario que la información que provee el dispositivo permita determinar la profundidad (Distancia del sensor al objeto de interés) y que mediante el procesamiento de imágenes digitales se pueda extraer la forma geométrica de los objetos. Existen diversas tecnologías para la medición de la profundidad en una imagen y se utilizan una amplia variedad de principios, como lo son la visión estéreo, luz infrarroja y rayos láser (Deris et al., 2017). Dentro de los métodos ópticos estos tipos de dispositivos se clasifican en dos categorías: pasivos y activos. En el primer grupo se encuentra la visión estéreo y en el segundo encontramos los métodos por infrarrojo (IR) y tiempo de vuelo (ToF). De los métodos pasivos destaca el de visión estéreo. Con este se estima la profundidad mediante el procesamiento de la disparidad entre dos imágenes tomadas desde diferentes posiciones (Zollhöfer et al., 2018). Los sistemas de visión estéreo dependen en gran medida de las condiciones de iluminación del ambiente. El principal inconveniente se presenta en la estimación de la distancia en los bordes de superficies lisas y curvas (Chen et al., 1995). En la Figura 1 se muestra como es la estimación de la profundidad con cámaras de visión estéreo.

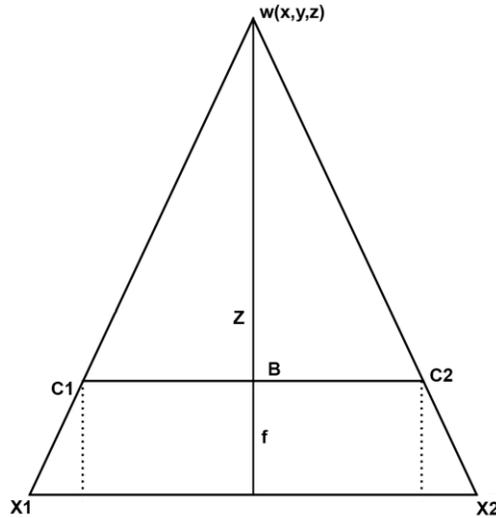


Figura 1. Estimación de profundidad

Donde  $f$  representa la distancia focal,  $B$  la línea base (separación entre centros de las cámaras,  $C1$  y  $C2$  centros de cada una de las cámaras,  $X1$  y  $X2$  corresponde a la localización del pixel en la imagen,  $w$  pixel dentro de la imagen y  $Z$  distancia desde las cámaras hasta el objeto. Para la estimación de  $Z$  se utiliza la Ecuación 1 descrita por Fahmy (2013) donde  $X1-X2$  corresponde a la disparidad  $d$  que se obtiene de la calibración.

$$Z = \frac{B * f}{d} \quad (1)$$

Dentro de los métodos activos existen dos enfoques principales para medir la distancia: mediante triangulación y utilizando el tiempo de vuelo. De acuerdo con Zollhöfer et al. (2018) la tecnología detrás de la medición de profundidad utilizando luz estructurada (Structured Light) o el tiempo de vuelo (time of flight) existe desde hace algunos años. Microsoft comenzó el desarrollo en el 2010 con la primera versión del Kinect utilizando luz estructurada y después de esto surgieron muchos dispositivos RGB-D como lo son Intel RealSense, PrimeSense, Carmine, Google Tango entre otros. De acuerdo con Cheng et al. (2015) para la segunda versión del sensor Kinect utilizaron la tecnología del tiempo de vuelo o Time of Flight (ToF). Este dispositivo fue ampliamente utilizado en la visión por computadora. La diferencia entre el Kinect V2 y la versión anterior se encuentra

en las aplicaciones en exteriores, la luz solar afecta en mayor medida los sensores que proyectan un patrón de luz estructurada (Kinect V1) comparadas con las que utilizan la velocidad de la luz para la medición de la distancia debido a son más robustas (Zollhöfer et al., 2018).

### **2.3.2 Procesamiento de imágenes**

Existe una gran variedad de librerías para el procesamiento de imágenes digitales, ya sean comerciales o versiones de acceso libre. Entre las opciones comerciales se encuentran: 1) MATLAB® (MathWorks, Inc., Natick, MA, EE.UU.) y sus librerías para el procesamiento de imágenes como: “Machine Vision Toolbox” (MVT) e “Image Processing Toolbox” (IPT), los cuales incluyen más de 60 funciones, 2) Matrox Image Library, que es un software para la adquisición y procesamiento de imágenes de uso industrial (Hu et al., 2011). Si bien existen otros paquetes como Khoros, eVision, etc. de acuerdo Arévalo et al. (2002) estos tienen un costo elevado y un ciclo de actualización largo, comparados con las opciones de librerías no comerciales como SimpleCV, Gandalf, ImageLib y OpenCV. Esta última es una librería desarrollada por “Intel”, para visión por computadora que puede ser implementada en C++, Python, Java y Matlab (OpenCV, 2019). OpenCV es utilizada por más de 47 mil usuarios, entre los que destacan compañías y grupos de investigación. De acuerdo con Grange et al. (2003) OpenCV es la librería más ampliamente utilizada en la comunidad de la visión por computadora, debido al extenso repertorio de algoritmos.

### **2.3.3 Algoritmos para la clasificación de frutos de jitomates**

La clasificación de frutos de jitomates se ha realizado desde diferentes puntos de vista. Espita et al. (2017) desarrollaron tres métodos de clasificación de jitomates por tamaño, dos utilizando procesamiento de imágenes digitales (Matlab y OpenCv) y un tercer método basado en el uso de redes neuronales. Los métodos propuestos obtuvieron una eficiencia de 61, 74 y 73 %, respectivamente. Por otro lado, Khoshroo et al. (2014) desarrollaron un algoritmo de segmentación basado en la región de cultivo para guiar a un robot en la cosecha de jitomates en condiciones de luz natural y artificial. El procedimiento consistió en remover el

fondo de la imagen substrayendo las capas R-G de una imagen RGB, y los jitomates que se tocaban entre ellos fueron separados utilizando el método watershed. El algoritmo desarrollado obtuvo una precisión de 82 % en la detección de los tomates maduros. Polder et al. (2002) realizaron un estudio con imágenes espectrales entre los rangos de 396-736 nm para cinco estados de madurez de jitomates y utilizaron una referencia gris para eliminar las variaciones de las fuentes de luz; el estudio demostró un mayor poder de discriminación comparado con el uso estándar de imágenes RGB, logrando una reducción en el error de clasificación de 51 % a 19 %. Wânderson (2007) presentó un modelo para la selección de jitomates basado en el procesamiento de imágenes mediante el tratamiento de color y la segmentación. El sistema propuesto utilizó una cámara digital y una red neuronal perceptrón multicapa para la clasificación de jitomates de uso industrial obteniendo un 100 % de precisión en condiciones controladas.

### **Presentación de resultados**

Existen diversas formas para presentar y visualizar el desempeño de los clasificadores, incluyendo métricas como exactitud, precisión, sensibilidad o el puntaje F1 (combinación de las últimas dos), las cuales describen el desempeño de los algoritmos y permiten hacer comparaciones. Las matrices de confusión describen el desempeño del clasificador en forma de tabla (Sammur et al., 2017) y permiten visualizar de forma fácil los resultados. Por otro lado, gráficos como las curvas ROC (es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario) son una técnica para organizar, visualizar y seleccionar clasificadores de acuerdo con su desempeño. Este tipo de gráficos es ampliamente utilizado en la evaluación de algoritmos de aprendizaje de máquina (Fawcett, 2006).

#### **2.3.4 Red neuronal de convolución**

Lecun et al. (1998) mencionan que las redes neuronales de convolución (CNNs) están específicamente diseñadas para tratar con formas de dos dimensiones y estas han demostrado superar a otras técnicas, en términos de clasificación. De

acuerdo con Yamashita et al. (2018) las CNNs están diseñadas para “aprender” jerarquías de características espaciales, de patrones de bajo a alto nivel. Las CNNs para la clasificación de objetos se componen de tres tipos de capas, principalmente: convolución, agrupación y capas completamente conectadas (Figura 2). En las primeras capas de la red se realiza la extracción de características, obteniendo características más abstractas en las capas intermedias y finalmente se combinan en la última para obtener como salida la clasificación (Yamashita et al., 2018).

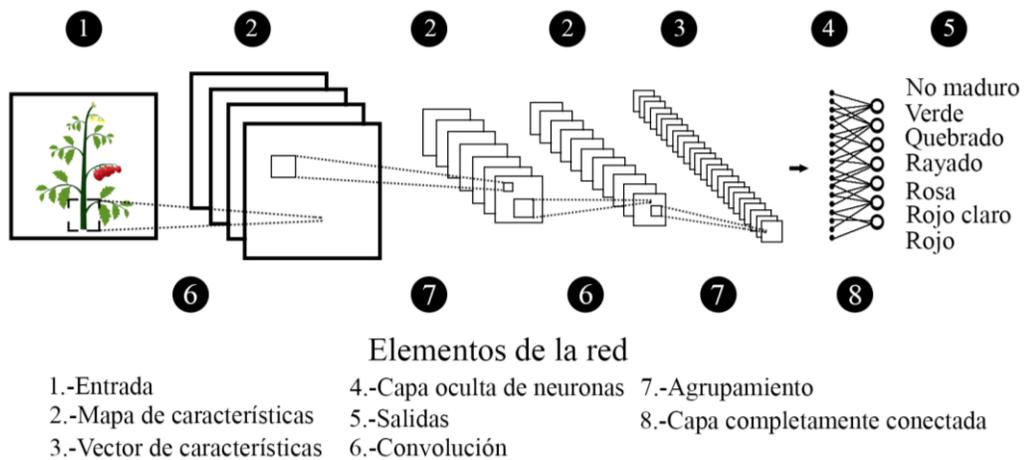


Figura 2. Red neuronal de convolución aplicada en la clasificación de frutos de jitomate

Para la clasificación se utilizó la entropía cruzada (Ecuación 2) como función de pérdida (Loss Function) para el ajuste de los pesos durante la etapa de entrenamiento. La entropía cruzada es un método eficiente para la optimización en problemas determinísticos y estocásticos (Rubinstein et al., 2013).

$$\mathcal{L}_{cls} = - \sum_i^c y_i * \log(\hat{y}_i) \quad (2)$$

Donde  $y_i$  corresponde a la etiqueta o valor real,  $\hat{y}_i$  representa la normalización del valor obtenido mediante la función softmax (Ecuación 3) y  $c$  es el número de clases.

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{i=1}^k e^{z_i}} \quad (3)$$

Donde  $z$  representa el valor obtenido en la salida de la red para la clase  $j$  e  $i$ ,  $k$  es el total de clases y  $j$  la clase a evaluar. Con la normalización se obtiene un valor que está en el rango de 0 a 1.

### **Localización y clasificación de objetos (Faster R-CNN)**

En los algoritmos de localización de objetos se implementan redes CNNs de igual manera, como extractores de características, en conjunto con regresiones para la obtención de coordenadas dentro de las imágenes (Koirala et al., 2019). La red neuronal de convolución conocida como “inception v2” o “GoogLeNet”, es un modelo desarrollado para la localización y clasificación de objetos dentro de una imagen. De acuerdo con Szegedy et al. (2015) la primera versión de esta red (inception V1) utilizó 12 veces menos parámetros, comparado con “AlexNet” para el entrenamiento, lo que permitió una mayor velocidad de entrenamiento y una mayor precisión final. La red fue diseñada para desempeñarse de forma adecuada bajo condiciones estrictas de memoria y presupuesto computacional, teniendo la capacidad de trabajar con mayores cantidades de información o en dispositivos con capacidades limitadas de procesamiento (Szegedy et al., 2016). La principal ventaja del uso de esta red es la detección utilizando el método propuesto por Girshick et al. (2014) conocido como “Regiones con CNN” (R-CNN). Este método combina los enfoques de aplicar la gran capacidad de las redes neuronales de convolución (CNN) en la detección de patrones, con la generación de regiones propuestas para la localización y segmentación, de los objetos de interés dentro una imagen.

Para la localización y clasificación de los objetos se utilizó la Ecuación 4 propuesta por Ren et al. (2017) como función de pérdida (Loss Function) durante la etapa de entrenamiento para el ajuste de los pesos.

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{box} \quad (4)$$

Donde  $L$  es la suma de los errores de clasificación y localización,  $L_{cls}$  corresponde al valor obtenido de la Ecuación 2 para la clasificación y  $L_{box}$  corresponde al error en la localización del fruto obtenido de la Ecuación 5.

$$\mathcal{L}(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i \mathcal{L}_{cls}(p_i, p_i^*) + \frac{\lambda}{N_{box}} \sum_i p_i^* \cdot L_1^{smooth}(t_i - t_i^*) \quad (5)$$

Donde  $i$  corresponde al índice de la caja (coordenadas del objeto) a evaluar,  $p_i$  es la probabilidad de que  $i$  sea un objeto y  $p_i^*$  es la etiqueta real (1 positivo, 0 negativo).  $t_i$  representa un vector de cuatro parámetros con las coordenadas estimadas del cuadro delimitador y  $t_i^*$  corresponde a las coordenadas reales del cuadro. Para la clasificación de los cuadros se utiliza la Ecuación 2 con dos clases (fruto vs no-fruto). Para el error obtenido de la regresión se utiliza la Ecuación 6. Los valores  $N_{cls}$  y  $N_{box}$  son parámetros para normalizar el error del número de clases y el número de cuadros propuestos respectivamente. El símbolo  $\lambda = 10$  para darle el mismo peso a el error de clasificación y de regresión.

$$L_1^{smooth}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{de otra manera} \end{cases} \quad (6)$$

Donde  $x$  corresponde a la suma de la diferencia de los dos vectores (Ecuación 7).

$$x = \sum_{i=1}^N (t_i - t_i^*) \quad (7)$$

Donde  $t_i$  es el vector con las coordenadas estimadas,  $t_i^*$  es el vector con las coordenadas reales y  $N$  es el número de elementos en el vector.

### **Segmentación de objetos (Mask R-CNN)**

Debido a que la “Faster R-CNN” fue creada para la detección, localización y clasificación de objetos se obtienen las coordenadas de los objetos localizados. Estas coordenadas representan una porción de la imagen que contiene píxeles que pertenecen o no al objeto de interés y es necesario realizar la segmentación de éstos para conocer la forma del objeto. La segmentación consiste en hacer una predicción de cada uno de los píxeles y conocer si este pertenece o no a una clase (Long et al., 2018). Un método para realizar la segmentación de los objetos es con la implementación de una red “Mask R-CNN” propuesta por He et al. (2017b). Esta predice una máscara de segmentación dentro de la región de interés, que contiene los píxeles que forman el objeto buscado. Badrinarayanan

et al. (2017) discuten las múltiples aplicaciones que tiene la segmentación en imágenes, que van desde inferir relaciones entre objetos, vehículos autónomos, hasta estimar la geometría de objetos de una escena.

Para la segmentación de los frutos de jitomate se utilizó la Ecuación 8 propuesta por He et al. (2017a) como función de pérdida durante la etapa de entrenamiento para el ajuste de los pesos.

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{box} + \mathcal{L}_{mask} \quad (8)$$

Donde  $L$  es la suma de los errores de clasificación, localización y de la máscara,  $L_{cls}$  corresponde al valor obtenido de la Ecuación 2 para la clasificación,  $L_{box}$  corresponde al error en la localización (Ecuación 5) y  $L_{mask}$  es el error de la predicción de la máscara obtenido de la Ecuación 9.

$$\mathcal{L}_{mask} = -\frac{1}{m^2} \sum_{1 \leq i, j \leq m} [y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log(1 - \hat{y}_{ij}^k)] \quad (9)$$

El algoritmo genera una máscara de  $m \times m$  para cada región de interés y para cada clase;  $k$  es el número de clase. La función de pérdida se obtiene con el promedio de una clasificación binaria de entropía cruzada, solo incluyendo la máscara asociada a la región que corresponde a la clase real. Donde  $y_{(i,j)}$  es el pixel en la máscara real y  $\hat{y}_{(i,j)}$  es el pixel de la máscara obtenida del algoritmo de la clase  $k$ .

### **Transferencia de aprendizaje**

Para la implementación de las redes mencionadas, comúnmente se utiliza la transferencia de aprendizaje. Esta técnica, de acuerdo con Kumaraswamy et al. (2018), es un enfoque para mejorar la eficiencia durante el entrenamiento. Para la implementación de esta técnica se utiliza una red previamente entrenada en una tarea similar y durante la etapa de entrenamiento se ajustan los pesos ya conocidos por la red, con ejemplos de la nueva tarea. La ventaja de este enfoque es utilizar los pesos para la tarea previamente entrenada (en contraste con los iniciados aleatoriamente), lo que resulta más eficiente, en términos de las evaluaciones y ejemplos necesarios durante el entrenamiento. Un ejemplo de

este modo de entrenamiento es la aplicación de la red “Faster R-CNN inception v2”, previamente entrenada con la base de datos “Common Objects in Context” (Lin et al., 2014). Esta base de datos contiene 91 objetos diferentes, con un total de 2.5 millones de etiquetas en 328,000 imágenes. La implementación de esta red proporciona las coordenadas relativas de donde está ubicado el objeto dentro la imagen y la clasificación.

#### **2.4 Estimación del peso de frutos de jitomate**

Los avances en la adquisición y el procesamiento de imágenes digitales han permitido realizar análisis no destructivos, eficientes y en “tiempo real”, utilizando cámaras digitales y computadoras. Las características obtenidas por el procesamiento de imágenes digitales en conjunto con otras técnicas han sido utilizadas para la estimación del peso en frutos. Teoh et al. (2007) utilizaron el área de la imagen de un mango y por medio de una regresión simple lograron predecir el peso de estos. Cömert et al. (2017) obtuvieron el área y el diámetro de manzanas por medio del procesamiento de imágenes digitales, posteriormente estimaron el peso de los frutos utilizando una regresión lineal. Otros autores como Moreno et al. (2013) estimaron el peso de frutos de mango utilizando la densidad de los frutos y la suma de volúmenes transversales, calculando estos como cilindros. Este tipo de predicciones se ven afectadas por la presencia de ruido durante la captura de imágenes, bajo condiciones no controladas como iluminación irregular, diversos elementos presentes en la imagen y la presencia de un fondo no uniforme.

Otros enfoques para la estimación del peso es utilizar redes neuronales artificiales; y autores como Rad et al. (2015) utilizaron las ANN's como modelos de predicción en conjunto con la medición de características de las plantas en etapas tempranas, para predecir el peso final en melones, y posteriormente estimar el rendimiento total antes de la cosecha. Salawu et al. (2014) hicieron uso de las ANN's para predecir el peso de conejos mediante mediciones de características físicas de estos; las ANN's fueron más precisas en capturar, cómo

los cambios en las variables predictoras afectan el peso estimado, comparadas con los modelos de regresión.

## 2.5 Conceptos clave

### Aprendizaje supervisado

El aprendizaje supervisado se refiere al entrenamiento de máquinas en función de entradas y salidas, utilizando ejemplos de los valores de salida (Etiquetas) para el ajuste de parámetros (Sammut et al., 2017).

### Clasificación

La clasificación consiste en determinar la clase a la que pertenece un elemento sin etiqueta de una cantidad finita predefinida de clases (Talia et al., 2016).

### Convolución

La convolución bidimensional es una operación fundamental en el procesamiento de imágenes y videos, donde se transforma una señal de entrada (imagen) en una señal de salida (mapa de características), utilizando una función (Kernel o filtro) (Raschka et al., 2019). La convolución bidimensional está definida matemáticamente como (Ecuación 10).

$$g(x, y) = h(x, y) * f(x, y) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j)h(x - i, y - j) \quad (10)$$

Donde  $g$  representa el resultado de la operación de convolución,  $h$  la imagen de entrada,  $f$  el filtro aplicado,  $M$  es el número de filas de  $g$ ,  $N$  es el número de columnas de  $g$ ,  $x$  e  $y$  son las coordenadas del píxel en  $h$  (imagen) y  $f$  (filtro).

### Detección

La detección de objetos en un problema de visión artificial donde se localizan y clasifican objetos en una imagen (Zhao et al., 2019). La localización de los objetos puede ser mediante cuadros delimitadores o clasificando cada uno de los píxeles en la imagen que contienen el objeto (Segmentación).

### **Gradiente descendiente estocástico**

Es un algoritmo de optimización que busca encontrar los pesos que minimicen la función objetivo, que debe ser derivable y convexa. La actualización de los pesos se hace de forma incremental para cada muestra (conjunto de datos) durante el entrenamiento, a diferencia de utilizar el total de los datos (Raschka et al., 2019).

### **Hiper-parámetros**

Los hiper-parámetros (meta-parámetros) son una serie de especificaciones que se fijan antes de la fase de aprendizaje, con el objetivo de obtener el mejor modelo con la información utilizada, en una cantidad razonable de tiempo (Andonie, 2019).

### **Localización**

La tarea de localización de objetos consiste en describir dónde se encuentra el objeto de interés en la imagen, con un cuadro delimitador indicando la posición, alto y ancho (Zhao et al., 2019).

### **Padding**

Uno de los problemas presentes durante las operaciones de convolución es la pérdida de información en los bordes de la imagen y la reducción de tamaño. El hiper-parámetro *padding* indica si se quiere rellenar los bordes de la imagen con ceros, permitiendo conservar el tamaño de la imagen (Albawi et al., 2017).

### **Pooling**

La operación *pooling* se aplica para reducir la complejidad de las matrices de características para las capas posteriores. El max-pooling es la operación más común de esta clase, donde de una subregión de la matriz solo se toma el valor más alto (Albawi et al., 2017). Se debe destacar que al aplicar esta operación existe pérdida de información.

## ReLU

La función ReLU (Rectificador Linear Unitario) es una función de activación, para introducir la no-linealidad en las redes neuronales (Albawi et al., 2017), donde todos los valores menores a cero, se convierten en cero y los mayores se conservan iguales.

$$ReLU = \max(0, x) \quad (4)$$

## Segmentación

La segmentación consiste en hacer una clasificación de cada uno de los píxeles dentro de una región de interés, para conocer si este pertenece o no a una clase previamente definida (Long et al., 2018).

## Stride

Hiper-parámetro para reducir el número de operaciones de convolución, durante la etapa de entrenamiento (cada cuantos píxeles realizar la operación), eliminando operaciones que presenten traslape con los píxeles vecinos, permitiendo reducir el tamaño de la matriz resultante (Albawi et al., 2017).

## Tensorflow

Es una librería de código abierto desarrollada por Google, utilizada ampliamente en el aprendizaje de máquinas, e implementado en algoritmos como las redes neuronales de convolución y recurrentes. Estas tienen características de simplicidad, flexibilidad y compatibilidad (James, 2020).

## 2.6 Referencias

- Albawi, S., Mohammed, T. A. y Saad, A.-Z. (2017). *Understanding of a convolutional neural network*. Paper presented at the International Conference on Engineering and Technology (ICET), Turkey.
- Andonie, R. (2019). Hyperparameter optimization in learning systems. *Journal of Membrane Computing*, 1, 279-291. doi:<https://doi.org/10.1007/s41965-019-00023-0>

- Arévalo, V. M., González, J. y Ambrosio, G. (2002). La librería de visión artificial OpenCV, Aplicación a la docencia e investigación. *Dpto. De Ingeniería de Sistemas y Automática, Universidad de Málaga. España*, 1-6.
- Atzberger, C. (2013). Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sensing*, 5, 949-981. doi:<https://doi.org/10.3390/rs5020949>
- Badrinarayanan, V., Kendall, A. y Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481-2495. doi:<https://doi.org/10.1109/TPAMI.2016.2644615>
- Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering*, 144, 52-60. doi:<https://doi.org/10.1016/j.tifs.2015.01.001>
- Chen, C. y Zheng, Y. F. (1995). Passive and Active Stereo Vision for Smooth Surface Detection of Deformed Plates. *Institute of Electrical and Electronics Engineers, Transactions on industrial electronics*, 42(3), 6. doi:<https://doi.org/10.1109/41.382141>
- Cheng, H., Yang, L. y Liu, Z. (2015). A Survey on 3D Hand Gesture Recognition. *IEEE: Transactions on Circuits and Systems for Video Technology*, 26(9), 1659-1973. doi:<https://doi.org/10.1109/TCSVT.2015.2469551>
- Cömert, O., Hekim, M. y Adem, K. (2017). Weight and Diameter Estimation Using Image Processing and Machine Learning Techniques on Apple Images. *International Journal of Engineering Research and Development*, 9(3), 147-154 doi:<https://dx.doi.org/10.29137/umagd.350588>
- Corke, P. I. (2015). The Machine Vision Toolbox : A MATLAB Toolbox for vision and vision-based control. *IEEE Robotics & Automation Magazine*, 12(4), 16-25. doi:<https://doi.org/10.1109/MRA.2005.1577021>
- Cortés, A. R. (2015). Diseño de un sistema para control y monitoreo de un invernadero hidropónico. *Instituto Politécnico Nacional*.
- De Baerdemaeker, J. (2000). Process Monitoring and Control for Precision Agriculture. *IF AC Bio-Robotics, Information Technology and Intelligent*, 33(29), 23-30.
- Deris, A., Trigonis, I., Aravanis, A. y Stathopoulou, E. K. (2017). Depth cameras on UAVs: A first approach. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(2), 6. doi:<https://doi.org/doi:10.5194/isprs-archives-XLII-2-W3-231-2017>
- Ehret, D. L., Lau, A., Bittman, S., Lin, W. y Shelford, T. (2001). Automated Monitoring of Greenhouse Crops. *Agronomie*, 21(4), 403 - 414. doi:<https://doi.org/10.1051/agro:2001133>

- Elsayed, A. A. y Yousef, W. A. (2019). Matlab vs. OpenCV: A Comparative Study of Different Machine Learning Algorithms. *arXiv.org*.
- Espita, O. P. y González, I. L. (2017). Clasificador de jitomate redondo aplicando técnicas de vision artificial y redes neuronales artificiales con Matlab y Python. *Universidad Autónoma Chapingo*.
- Fahmy, A. A. (2013). Stereo Vision Based Depth Estimation Algorithm In Uncalibrated Rectification. *International Journal of Video&Image Processing and Network Security*, 13(2), 1-8.
- FAO, Food and Agriculture Organization of the United Nations. (2017). Producción Mundial de tomates. Retrieved 5 de Marzo del 2019 <http://www.fao.org/faostat/es/#data/QC/visualize>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. doi:<https://doi.org/10.1016/j.patrec.2005.10.010>
- FIRA, Fideicomisos Instituidos en Relación con la Agricultura. (2019). Panorama Agroalimentario: Tomate Rojo 2019.
- Girshick, R., Donahue, J., Darrell, T. y Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv.org*.
- Grange, S., Fong, T. y Baur, C. (2003). TLIB: a Real-time Computer Vision Library for HCI. *Digital Image Computing: Techniques and Applications*, 1017-1026.
- He, K., Gkioxari, G., Dollár, P. y Girshick, R. (2017a). Mask R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, 2980-2988. doi:<https://doi.org/10.1109/ICCV.2017.322>
- He, K., Gkioxari, G., Dollár, P. y Girshick, R. (2017b). Mask R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, 2961-2969. doi:<https://doi.org/10.1109/ICCV.2017.322>
- Hu, L. y Wang, C. (2011). Application of core sample image acquisition system to coarse aggregate gradation analysis of cement stabilized aggregates. *Procedia Engineering*, 15, 4511-4515. doi:<https://doi.org/10.1016/j.proeng.2011.08.847>
- James, A. P. (2020). *Deep Learning Classifiers with Memristive Networks*: Springer.
- Khoshroo, a., Arefi, A. y Khodaei, J. (2014). Detection of Red Tomato on Plants using Image Processing Techniques. *Agricultural Communications*, 4(2), 9-15.
- Koirala, A., Walsh, K. B., Wang, Z. y McCarthy, C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, 162, 219-234. doi:<https://doi.org/10.1016/j.compag.2019.04.017>
- Kumaraswamy, S. K., Sastr, P. y Ramakrishnan, K. (2018). Multi-source Subnetwork-level Transfer in CNNs Using Filter-Trees. *IEEE: International*

*Joint Conference on Neural Networks (IJCNN).*  
doi:<https://doi.org/10.1109/IJCNN.2018.8489678>

- LeCun, Y., Bengio, Y. y Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444.  
doi:<https://doi.org/doi:10.1038/nature14539>
- Lecun, Y., Bottou, L., Bengio, Y. y Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11), 2278 - 2324. doi:<https://doi.org/10.1109/5.726791>
- Liakos, K. G., Busato, Patrizia, Moshou, D., Pearson, S. y Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *sensors*, 18(8), 2674-2703. doi:<https://doi.org/doi:10.3390/s18082674>
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. y Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *Springer European Conference on Computer Vision*, 8693, 740-755. doi:[https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- Liu, J., Pi, J. y Xia, L. (2019). A novel and high precision tomato maturity recognition algorithm based on multi-level deep residual network. *Multimedia Tools and Applications*, 1-15. doi:<https://doi.org/10.1007/s11042-019-7648-7>
- Long, J., Shelhamer, E. y Darrell, T. (2018). Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431-3440. doi:<https://doi.org/10.1109/CVPR.2015.7298965>
- Ma, J., Du, K., Zheng, F., Zhang, L., Gong, Z. y Sun, Z. (2018). A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Computers and Electronics in Agriculture*, 154, 18-24. doi:<https://doi.org/10.1016/j.compag.2018.08.048>
- Mahajana, S., Dasa, A. y Sardanaa, H. K. (2015). Image acquisition techniques for assessment of legume quality. *Trends in Food Science & Technology*, 42(2), 116-133. doi:<http://dx.doi.org/10.1016/j.tifs.2015.01.001>
- Moreno, A., Bermúdez, Ballesteros, D., Padilla y Sánchez, G., Torres. (2013). Image analysis for automatic feature estimation of the *Mangifera indica* fruit *Universidad del Magdalena, Santa Marta*.
- Naveen, G. B., Nandhini, V., Mithra, S., Priya, N. y Naveena, R. (2018). IOT Based Smart Crop Monitoring in Farm Land. *Imperial Journal of Interdisciplinary Research (IJIR)*, 4(1), 88-92.
- OpenCV, Developers Team. (2019). OpenCV. Retrieved from <https://opencv.org/about/>
- Pandurng, J. A. y Lomte, S. S. (2015). Digital Image Processing Applications in Agriculture: A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(3), 622-624.

- Partel, V., Nunes, L., Stansly, P. y Ampatzidis, Y. (2019). Automated vision-based system for monitoring Asian citrus psyllid in orchards utilizing artificial intelligence. *Computers and Electronics in Agriculture*, 162, 328-336. doi:<https://doi.org/10.1016/j.compag.2019.04.022>
- Patrício, D. I. y Riederb, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, 153, 69-81. doi:<https://doi.org/10.1016/j.compag.2018.08.001>
- Polder, G., van der Heijden, G. W. A. M. y Young, I. T. (2002). Spectral image analysis for measuring ripeness of tomatoes. *American Society of Agricultural Engineers*, 1155-1161.
- Rad, M. R. N., Koohkan, S., Fanaei, H. R. y Rad, M. R. P. (2015). Application of Artificial Neural Networks to predict the final fruit weight and random forest to select important variables in native population of melon (*Cucumis melo*. Pahlavan). *Scientia Horticulturae*, 181, 108-112. doi:<https://doi.org/10.1016/j.scienta.2014.10.025>
- Raschka, S. y Mirjalili, V. (2019). *Python Machine Learning* (Third ed.). Birmingham B3 2PB, UK.: Packt Publishing Ltd.
- Ren, S., He, K., Girshick, R. y Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137 - 1149. doi:<https://doi.org/10.1109/TPAMI.2016.2577031>
- Rubinstein, R. Y. y Kroese, D. P. (2013). *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning.*: Springer.
- Russell, S. J. y Norvig, P. (2016). *Artificial Intelligence: A Modern Approach* (P. Education Ed.).
- Ryu, K. H., Kim, G. Y. y Chae, H. Y. (2000). Monitoring Greenhouse Plants Using Thermal Imaging. *IF AC Bio-Robotics, Information Technology and Intelligent*, 33(29), 181-186.
- Salawu, E. O., Abdulraheem, M., Shoyombo, A., Adepeju, A., Davies, S., Akinsola, O. y Nwagu, B. (2014). Using Artificial Neural Network to Predict Body Weights of Rabbits. *Open Journal of Animal Sciences*, 4, 182-186. doi:<http://dx.doi.org/10.4236/ojas.2014.44023>
- Sammut, C. y Webb, G. I. (2017). *Encyclopedia of machine learning and data mining*: Springer Publishing Company, Incorporated.
- Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44(1.2), 206 - 226. doi:<https://doi.org/10.1147/rd.441.0206>
- Sevillano, F. J. B. (2015). Equipo de clasificación de semillas basado en visión artificial. *Escuela de Ingenierías Industriales, Universidad de Valladolid*.

- SIAP, Servicio de Información Agroalimentaria y Pesquera (2017). Anuario Estadístico de la Producción Agrícola. Retrieved 5 de Marzo del 2019 <https://nube.siap.gob.mx/cierreagricola/>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. y Rabinovich, A. (2015). Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:<https://doi.org/10.1109/CVPR.2015.7298594>
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. y Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:<https://doi.org/10.1109/CVPR.2016.308>
- Talia, D., Trunfio, P. y Marozzo, F. (2016). *Data Analysis in the Cloud: Models, Techniques and Applications*: Elsevier.
- Teoh, C. C. y Syaifudin, A. R. M. (2007). Image processing and analysis techniques for estimating weight of Chokanan mangoes. *Journal of Tropical Agriculture and Food Science*, 35(1), 183-190.
- Wânderson, O. A. (2007). Selección de tomates para procesamiento industrial por medio de redes neuronales aplicadas a un sistema de vision artificial. *Faculta de de Tecnologia Termomecanica*.
- Yamamoto, K., Guo, W., Yoshioka, Y. y Ninomiya, S. (2014). On Plant Detection of Intact Tomato Fruits Using Image Analysis and Machine Learning Methods. *sensors*, 14(7), 12191-12206. doi:<https://doi.org/10.3390/s140712191>
- Yamashita, R., Nishio, M., Do, R. K. G. y Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, 9, 611-629. doi:<https://doi.org/10.1007/s13244-018-0639-9>
- Zhao, Z.-Q., Zheng, P., Xu, S.-T. y Wu, X. (2019). Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11), 3212 - 3232. doi:<https://doi.org/10.1109/TNNLS.2018.2876865>
- Zollhöfer, M., Stotko, P., Görnitz, A., Theobalt, C., Nießner, M., Klein, R. y Kolb, A. (2018). State of the Art on 3D Reconstruction with RGB-D Cameras. *The Eurographics Association and John Wiley & Sons Ltd*, 37(2), 28. doi:<https://doi.org/10.1111/cgf.13386>

### **3 DETECCIÓN Y CLASIFICACIÓN DE FRUTOS DE JITOMATE DE INVERNADERO CON VISIÓN ARTIFICIAL Y REDES NEURONALES DE CONVOLUCIÓN.**

DETECTION AND CLASSIFICATION OF GREENHOUSE TOMATOES WITH  
ARTIFICIAL VISION AND CONVOLUTIONAL NEURAL NETWORKS

Christian Israel Zuñiga-Avila<sup>1</sup>, Efrén Fitz-Rodríguez<sup>1\*</sup>, José Armando Ramírez-  
Árias<sup>1</sup>, Raquel Salazar-Moreno<sup>1</sup>

<sup>1</sup>Posgrado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma  
Chapingo, carretera México – Texcoco km 38.5, Texcoco, Estado de México, C.P.  
56230, México.

\*Autor de correspondencia:  
efitzr@chapingo.mx (Efrén Fitz-Rodríguez)

Este capítulo fue sometido para su publicación en la revista científica *Computers  
and Electronics in Agriculture*.

## RESUMEN

La detección de frutos en diversos cultivos es de suma importancia por diversas razones, incluyendo el monitoreo de su desarrollo y crecimiento, con lo cual se puede estimar el rendimiento y la automatización de su cosecha. Con el avance tecnológico en los sistemas de visión artificial, actualmente se puede obtener información útil de los cultivos de manera no invasiva. En esta investigación se plantea implementar un sistema de visión artificial para la detección y clasificación por estado de madurez de los frutos en un cultivo de jitomate bajo condiciones de invernadero. Se implementó una red neuronal de convolución pre-entrenada para detectar los frutos en la planta y posteriormente clasificarlos de acuerdo con su estado de madurez, la red fue entrenada mediante transferencia de aprendizaje. El sistema desarrollado obtuvo una sensibilidad (capacidad para identificar los frutos) del 70 % y una precisión (probabilidad de que un objeto identificado sea un fruto de jitomate) del 96 % en la detección de los frutos en la planta, bajo condiciones ambientales de invernadero. Además, se obtuvo una exactitud del 93 % en la clasificación correcta de los frutos por estado de madurez. El algoritmo presentó inconvenientes para la detección de frutos parcialmente ocultos y en la clasificación de los estados de madurez iniciales.

**Palabras Claves:** Tomate, Segmentación, Mask R-CNN, Faster R-CNN, Aprendizaje de máquina.

### **Highlights:**

- Las redes neuronales de convolución son más eficientes que otros algoritmos, en la detección de frutos de jitomate bajo condiciones ambientales de invernaderos.
- Los estados de madurez de los frutos de jitomate se pueden detectar y clasificar adecuadamente con redes neuronales de convolución.
- Las redes Mask R-CNN resultaron más efectivas en la segmentación de los frutos de jitomate.
- El sistema de visión desarrollado puede ser utilizado para la estimación de un mapa de rendimientos.

### 3.1 Introducción

La detección de frutos en “tiempo real” en diversos cultivos es de suma importancia por varias razones: la automatización de su cosecha, el monitoreo de su estado, crecimiento y desarrollo; en la determinación de calidad y en la estimación de rendimientos en cada periodo de cosecha. Con la estimación de rendimientos se pueden planear de mejor manera los procesos que ocurren durante y después de la cosecha, y permite mejores estrategias de mercadotecnia. Con el avance de la tecnología en la capacidad de cómputo, la calidad de las cámaras digitales y algoritmos en el procesamiento de imágenes, los sistemas de visión han demostrado ser una herramienta efectiva y no invasiva en diferentes tareas en la agricultura (Pandurng et al., 2015). Algunas aplicaciones de la visión artificial en la agricultura incluyen: detección de enfermedades (Ma et al., 2018), manejo de riego (Lam et al., 2007), detección de malezas (Wang et al., 2019), detección de deficiencias nutricionales (Xu et al., 2011), inspección de la calidad (Zaborowicz et al., 2017), clasificación del estado de madurez de los frutos (Nandi et al., 2014), estimación del contenido de nutrientes mediante análisis de color (Agarwal et al., 2018), detección de plagas (Partel et al., 2019), estimación del peso de frutos en planta (Cömert et al., 2017) y en robots cosechadores (Zhao et al., 2016a).

La identificación y clasificación de frutos, directamente en la planta, resulta de gran interés, debido a las aplicaciones que tiene el uso de esta información en la estimación de rendimientos. De acuerdo con Yamamoto et al. (2014), la estimación del rendimiento de un cultivo provee información que ayuda a los productores optimizar costos del manejo y cosecha; además permite conocer la capacidad de almacenaje requerida para los frutos cosechados y desarrollar mejores estrategias de mercadotecnia. Por otro lado, una aplicación directa de la detección de frutos de manera autónoma mejora la eficiencia de los robots cosechadores (Liu et al., 2019a).

El proceso de la detección de frutos en las plantas con sistemas de visión, consiste en localizar en qué regiones de las imágenes se encuentran estos y cuáles conforman el fondo de la misma. Koirala et al. (2019) divide la tarea de la

detección de objetos en dos etapas: 1) localización de posibles regiones como frutos y 2) segmentación individual de objetos dentro de esas regiones. Por otro lado, Hamuda et al. (2016) menciona que existen diversos enfoques para la segmentación de imágenes que permiten la detección de los objetos, ya sea por color, un umbral predefinido o mediante algoritmos de aprendizaje.

La segmentación por color es uno de los métodos más comunes para separar los objetos de interés del resto de la imagen. Malik et al. (2018) propusieron un método para detectar frutos de jitomate maduros, bajo condiciones de invernadero, mediante la segmentación por color en el espacio HSV; y separando los frutos sobrepuestos mediante el método de cuencas (watershed). Con este algoritmo removieron el fondo con una alta presencia de ruido bajo diferentes condiciones de iluminación, y lograron separar satisfactoriamente los frutos de jitomates superpuestos mediante el método de cuencas.

La detección de frutos, utilizando una segmentación con umbrales, asume que el problema tiene dos clases, la primera es el objeto de interés y la segunda es el fondo de la imagen. Elegir el umbral apropiado es de suma importancia, ya que de esto depende que se obtenga una sobre o sub, segmentación (Hamuda et al., 2016). Wang et al. (2017) implementaron un detector en cascada utilizando las características de HOG (Histogram of Oriented Gradients) para la identificación de frutos de mango en los árboles. El método propuesto obtuvo una alta precisión para la detección de frutos (la probabilidad de que un objeto identificado sea un mango), pero una tasa pobre en la sensibilidad (capacidad para identificar la totalidad de los frutos) debido principalmente a la presencia de frutos parcialmente ocultos. Liu et al. (2019a) utilizaron este mismo método en conjunto con uno para la eliminación de los falsos positivos (False Color Removal) para la detección de frutos de jitomate; además, implementaron el método Non-Maximum Suppression (NMS) para la eliminación de múltiples detecciones de un mismo fruto. El algoritmo propuesto obtuvo resultados satisfactorios en la detección de jitomates maduros, destacando como principal característica de los frutos no identificados la obstrucción de más de un 50 % por otros frutos o el fondo.

Algunos métodos de aprendizaje no requieren de la extracción manual de características de las imágenes y han tenido gran éxito recientemente. Dentro de estos se encuentran las redes neuronales de convolución (CNN, por sus siglas en inglés). Éste enfoque fue desarrollado para la clasificación de imágenes y se ha desarrollado ampliamente para la detección de objetos dentro de una imagen. Por ejemplo, Partel et al. (2019) implementaron una red neuronal de convolución para la detección de insectos transmisores de enfermedades, logrando una precisión y una sensibilidad del 80 y 95 %, respectivamente. Por otro lado, Liu et al. (2019b) propusieron una Red Densa como arquitectura para detectar frutos de jitomate maduros en imágenes complejas, con el objetivo de mejorar la precisión de un equipo de cosecha. El algoritmo que desarrollaron fue capaz de localizar jitomates maduros con diferentes tamaños bajo condiciones complejas de fondo.

Los principales retos que se presentan en la detección y clasificación de frutos por grado de madurez, son: presencia de frutos con una epidermis no uniforme, obstrucción de frutos dada la orientación de los racimos, lo que se traduce en dificultad para identificarlos individualmente. La iluminación es de suma importancia para la efectiva segmentación, una pobre iluminación resulta en una pobre segmentación.

En este sentido, esta investigación se propone detectar y clasificar por grado de madurez, frutos de jitomate en cultivo de invernadero mediante el uso de visión artificial y redes neuronales de convolución.

## 3.2 Materiales y Métodos

### 3.2.1 Localización

Las imágenes utilizadas en el desarrollo de la investigación se obtuvieron de un cultivo de jitomate (*Solanum lycopersicum L.*) tipo saladette, bajo condiciones de invernadero del *Laboratorio de modelación y automatización de biosistemas* de la Universidad Autónoma Chapingo (19°29'23"N 98°53'37"O) a una altitud de 2,250 msnm.

### 3.2.2 Sistema de visión artificial

En el sistema de visión artificial se utilizó una cámara de visión estéreo RGB-D, modelo Zed-M (Stereolabs, Montrouge, Francia) para la captura de imágenes, conectada a una computadora portátil para la adquisición de imágenes durante monitoreo. Posteriormente se utilizó una computadora de escritorio, para el procesamiento de imágenes (Figura 3).

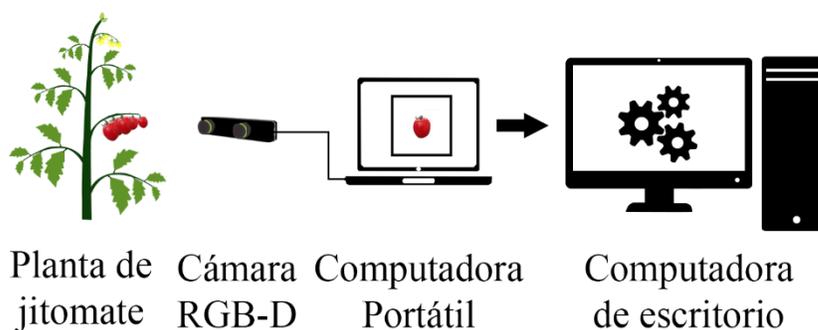


Figura 3. Sistema de visión artificial

Los algoritmos del sistema de visión propuesto (Figura 3), para operar en condiciones ambientales de un invernadero, incluye dos etapas para realizar: 1) la localización y clasificación de los frutos del cultivo y 2) la segmentación de frutos. Las dos etapas en conjunto integran el proceso de detección de frutos de jitomates (Figura 4).

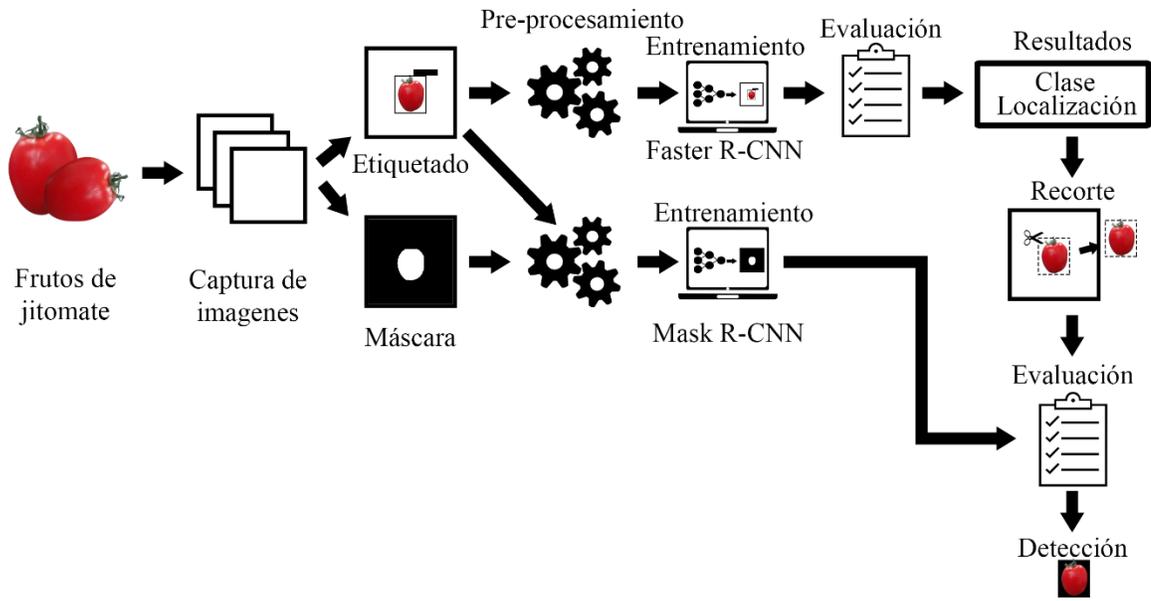


Figura 4. Metodología para la detección y clasificación de frutos.

## Hardware

### *Cámara RGB-D (Zed-M)*

El dispositivo implementado cuenta con dos cámaras de 4 megapíxeles, con una separación de 6.3 cm entre ellas. Se utilizó una resolución de 1280x720 píxeles para la captura de imágenes, con un campo de visión de 54° vertical y 85° horizontal. El rango de operación del dispositivo para la estimación de la profundidad es de 0.15 a 12.0 m y puede ser utilizada en exteriores.

### *Computadora portátil*

La cámara RGB-D se conectó vía un puerto USB 3.0 a una laptop (Intel core I5-4210u con 6 Gb RAM) para la adquisición de imágenes.

### *Computadora de escritorio*

Para realizar el procesamiento de imágenes y los procesos de entrenamiento de las redes Faster R-CNN se utilizó una computadora de escritorio con un procesador Intel core I5-7500, con 16 Gb de memoria Ram y una tarjeta NVIDIA GeForce GTX 1060 de 3 Gb con un total de 1,152 CUDA cores.

## Software: Procedimiento y algoritmos.

### *Funcionamiento del sistema*

Para la detección y clasificación de los frutos el sistema de visión artificial se utilizó el siguiente procedimiento (Figura 5): 1) captura imágenes del cultivo, 2) implementación de una red neuronal de convolución – Rápida, (Faster R-CNN) para la localización y clasificación de jitomates por su grado de madurez, 3) extracción de regiones de la imagen donde se detectaron frutos, y 4) implementación de la red Mask R-NN para la segmentación de los frutos dentro de las regiones de interés. Una vez que se implementa la segmentación los frutos se consideran detectados.

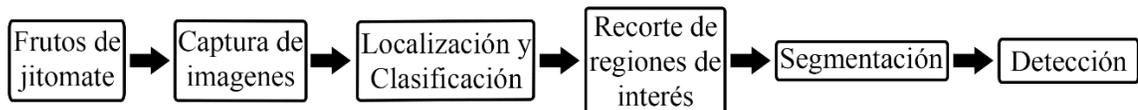


Figura 5. Diagrama de funcionamiento del sistema de visión

### *Localización y clasificación de frutos*

La primera etapa del proceso consiste en la localización de los frutos en las plantas del cultivo. Para ello se implementaron redes neuronales de convolución (CNNs). De acuerdo con Lecun et al. (1998) las CNNs están específicamente diseñadas para tratar con formas de dos dimensiones (2D). Estas redes han demostrado superar otras técnicas en términos de clasificación y estas mismas son utilizadas para la localización. Las CNNs se utilizan como extractores de características, y automáticamente “aprenden” las características de las bases de datos, en lugar de hacerlo de forma manual (Ma et al., 2018). Las CNNs incluyen una capa de entrada y una de salida, con capas intermedias que realizan operaciones de convolución, agrupamiento y activaciones de tipo rectificador lineal unitario (ReLU) (Koirala et al., 2019). Estas operaciones extraen elementos particulares de las imágenes y reducen el tamaño de los mapas de características. En la tarea de clasificación, la red obtiene un vector de características que está conectado con una capa oculta de neuronas. Los valores obtenidos del conjunto de características determinan a que clase pertenece la

imagen. La estructura de una CNN se muestra en la Figura 6. Las principales operaciones utilizadas en las CNNs son la convolución y el agrupamiento máximo Figura 7.

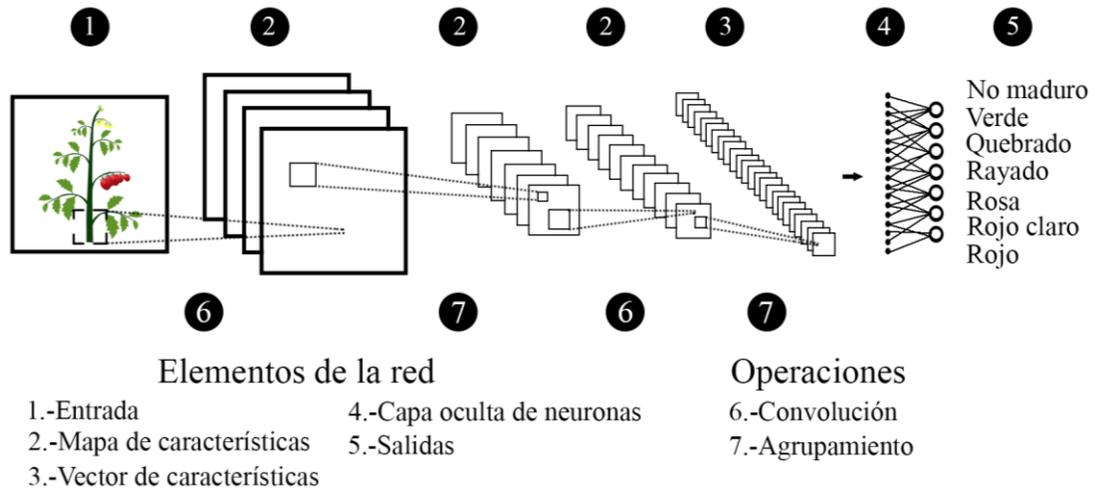


Figura 6. Arquitectura de una red neuronal de convolución

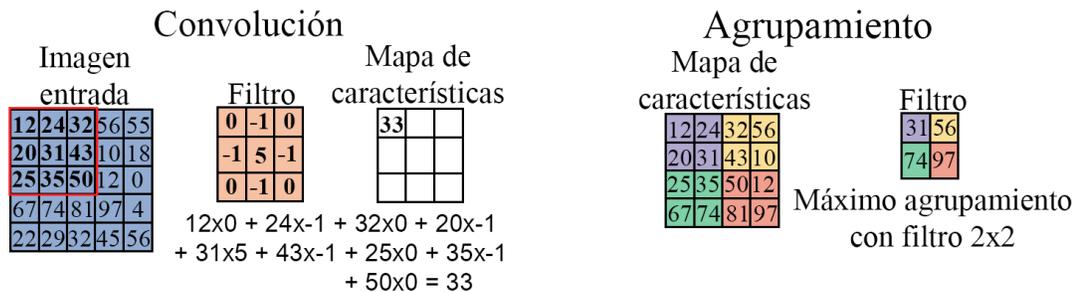


Figura 7. Principales operaciones en las CNNs

Los algoritmos para la localización de objetos utilizan de igual manera las redes CNNs como extractores de características, en conjunto con regresiones para la obtención de las coordenadas dentro de las imágenes (Koirala et al., 2019). En la presente investigación se implementó una red “Faster R-CNN inception v2” como la implementa Ren et al. (2017). Esta red fue previamente entrenada con la base de datos “Common Objects in Context” (Lin et al., 2014). La base de datos contiene 91 objetos diferentes, con un total de 2.5 millones de etiquetas en 328,000 imágenes. La red está compuesta por diversos módulos, incluyendo: a) un extractor de características VGG-16, compuesto por una serie de capas de convolución que produce un mapa de características, b) capas de agrupamientos

(max-pooling) para la reducción del tamaño de los mapas, c) una red para identificar regiones de interés (RPN), conformada por una red de convolución con un kernel de 3x3 que se pasa por el mapa de características y alimenta a una red completamente conectada, para clasificar si existe un objeto de interés en la región o pertenece al fondo; sí la clasificación indica que es una región de interés se realiza una regresión para estimar la posición en la imagen original, d) una capa para la agrupación de regiones de interés (RoI), que transforma el mapa de características de las regiones propuestas en uno de dimensiones fijas, e) una red para la clasificación por medio de vectores de características, que predice la probabilidad de pertenencia a cada una de las clases y estima las coordenadas de la ubicación del objeto en la imagen de entrada (Girshick, 2015). En la Figura 8 se muestra un diagrama simplificado de los elementos de una red Faster R-CNN.

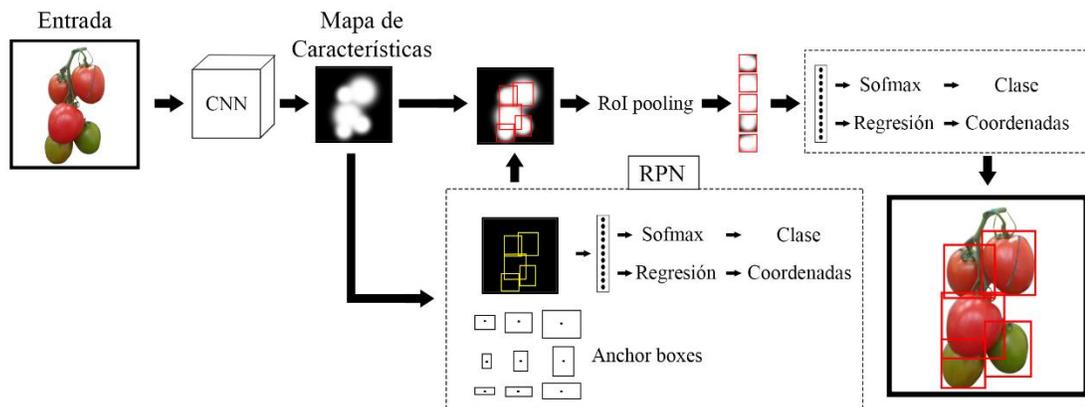


Figura 8. Estructura de la red Faster R-CNN implementada.

Para implementar este tipo de red, se utilizó la transferencia de aprendizaje mediante la librería Tensorflow (Abadi et al., 2016). De acuerdo con Kumaraswamy et al. (2018), la transferencia de aprendizaje es un enfoque para mejorar la eficiencia durante la etapa de entrenamiento de una red. Esta técnica utiliza una red previamente entrenada en una tarea similar, ajustando los pesos ya conocidos durante el entrenamiento con ejemplos de la nueva tarea. La ventaja de este enfoque se encuentra en utilizar los pesos para la tarea previamente entrenada (en contraste con pesos con inicialización aleatoria). Esto

resulta en una forma más eficiente, en términos de las evaluaciones necesarias y el número de ejemplos requeridos para obtener resultados satisfactorios (Kumaraswamy et al., 2018).

### *Segmentación de frutos*

Una vez localizados los frutos de jitomate en las imágenes y clasificados por su estado de madurez, se utilizan sus coordenadas para extraer las partes de la imagen con los objetos de interés (frutos). Dentro de estas porciones de imágenes recortadas, existen píxeles que pertenecen a los frutos de jitomates y otros al fondo de la imagen. Para la clasificación de píxeles se utilizó una red neuronal “Mask R-CNN” (Figura 9) propuesta por He et al. (2017), esta es una modificación de la red Faster R-CNN y se diferencia en que esta hace una estimación de regiones de interés, realiza una predicción de la clase, obtiene sus coordenadas dentro de la imagen y genera una máscara de segmentación binaria para cada una de las regiones. La segmentación consiste en hacer una predicción de cada uno de los píxeles, y conocer si este pertenece o no a una clase (Long et al., 2018). Badrinarayanan et al. (2017) proponen que esta tarea tiene múltiples aplicaciones, que van desde inferir relaciones entre objetos, uso en vehículos autónomos, hasta estimar la geometría de objetos de una escena. Una vez implementada la segmentación se concluye con el proceso de detección de frutos.

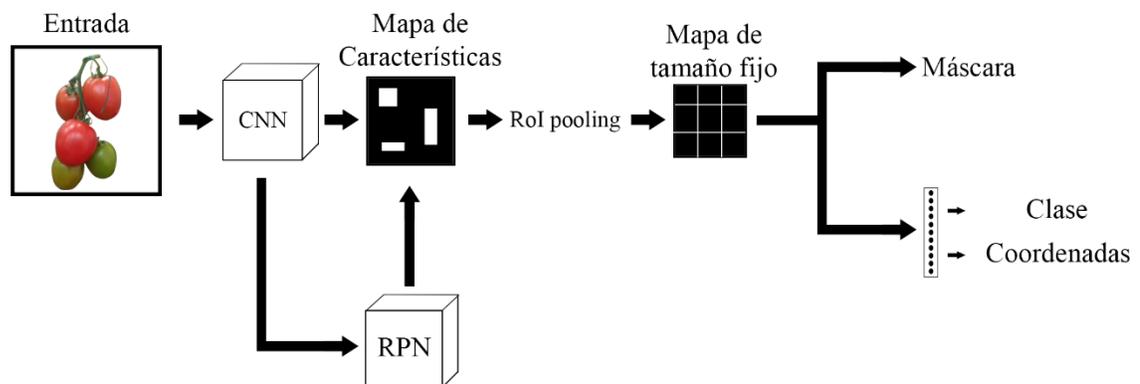


Figura 9. Estructura de una red Mask R-CNN

### 3.2.3 Selección de la muestra

Para entrenar las redes neuronales propuestas en la detección y clasificación de frutos de jitomate, se tomó una muestra aleatoria de 307 frutos de jitomate. Los frutos seleccionados eran una muestra representativa de la población, presentando diferentes características de tamaño, peso y grado de madurez (Figura 10). Se registró el estado de maduración con un número en el rango de 0-6, tomando el 0 para los jitomates no maduros y 6 para los más maduros, siguiendo los requerimientos de color de la Norma Mexicana de productos alimenticios no industrializados (NMX-FF-031-1997-SCFI, 1997) (Cuadro 1) de jitomate fresco, no industrializados para consumo humano.

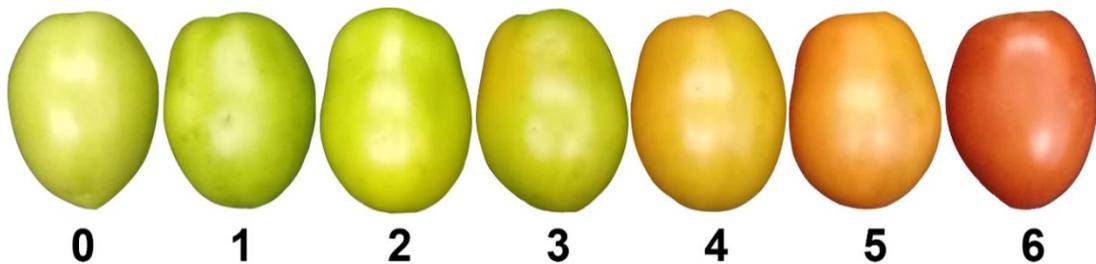


Figura 10. Grados de madurez del fruto de jitomate

Cuadro 1. Grado de madurez de acuerdo con la coloración del fruto.

Representación	Clase	Descripción
0	<b>No maduro</b>	Jitomates cortados antes de alcanzar la madurez fisiológica.
1	<b>Verde</b>	La piel del jitomate está completamente verde. El color verde puede variar de verde claro a oscuro.
2	<b>Quebrado</b>	Hay una interrupción distinta en el color de verde hacia amarillo, rosado o rojo en no más del 10 % de la epidermis.
3	<b>Rayado</b>	La superficie del jitomate muestra un cambio de entre el 10 % y el 30 % del color verde hasta amarillo, rosado o rojo, o una mezcla de éstos.
4	<b>Rosa</b>	Entre el 30 % y el 60 % de la superficie del jitomate presenta un color rosado o rojo.
5	<b>Rojo claro</b>	Entre el 60 % y el 90 % de la superficie tiene color rosado/rojo o rojo.
6	<b>Rojo</b>	Más del 90 % de la superficie del jitomate muestra color rojo.

**Fuente:** Adaptado de la Norma Mexicana (NMX-FF-031-1997-SCFI, 1997) de productos alimenticios no industrializados para consumo humano. Hortalizas frescas. Jitomate (*Solanum lycopersicum* L.).

Si bien la Norma está centrada en la clasificación de jitomates fisiológicamente maduros, se añadió la clase no maduro (0). Este tipo de frutos no se encuentran lo suficientemente desarrollados para madurar a un nivel aceptable (Heuvelink, 2005). La distribución de los jitomates seleccionados en cada una de las clases se presenta en el Cuadro 2.

Cuadro 2. Distribución de frutos por grado de madurez en la muestra

<b>Representación</b>	<b>Clase</b>	<b>N. frutos</b>
0	No Maduro	20
1	Verde	47
2	Quebrando	14
3	Rayado	31
4	Rosa	53
5	Rojo claro	44
6	Rojo	98
<b>Total</b>		<b>307</b>

### 3.2.4 Captura de imágenes

Para construir la base de datos de las imágenes de los frutos de jitomate se usó una cámara RGB-D *Zed-M* y se desarrolló un algoritmo en Python utilizando el Software Development Kit 2.8.3 (SDK 2.8.3). El dispositivo de visión (cámara RGB-D y laptop) almacena la información en un archivo con extensión propia “.svo”, el cual contiene las imágenes sin rectificar de cada una de las cámaras y los datos de los sensores, como el giroscopio y el acelerómetro. Se capturaron alrededor de 10 mil imágenes en diversas condiciones, controladas y no controladas, incluyendo: 1) diferentes tipos de iluminación y fondo, 2) frutos de jitomates individuales o en grupos y 3) presencia de frutos parcialmente ocultos; todas a una resolución de 1280x720 píxeles. Para la captura de imágenes en condiciones controladas se instaló un área con iluminación y fondo uniforme. La cámara fue colocada en un trípode a una distancia de 150 a 1 000 mm del set de imágenes. La distancia mínima se determinó a partir de la distancia mínima de lectura del dispositivo. Se instalaron dos lámparas fluorescentes de 9 watts a 6 500 kelvin para iluminación del área. Este tipo de luz se considera luz blanca fría o “Day light” (Marszalec et al., 1994) y permitió una mejor iluminación de los objetos. Por otro lado, las imágenes en condiciones ambientales no controladas se capturaron

dentro del invernadero, con frutos en las plantas y sobre diferentes superficies para tener una base de datos completa.

Para rectificar las imágenes de cada una de las cámaras del dispositivo, se utilizó la interfaz de programación de aplicaciones (API), la cual incluye un conjunto de métodos dentro del SDK, y que puede ser utilizado por otros programas computacionales. El proceso de rectificación consiste en corregir las imágenes tomadas por dos cámaras para que estén alineadas, y facilitar la estimación de la disparidad, para el cálculo de la profundidad por medio de triangulación (Ortiz et al., 2018). La distancia, que existe del sensor a cada uno de los frutos identificados, se estimó como el promedio de la distancia obtenida del 10 % del área del centro de la imagen.

### **3.2.5 Preprocesamiento**

La etapa de preprocesamiento de la información consta de dos fases: 1) etiquetado de las imágenes, y 2) generación de archivos para entrenamiento.

#### **Etiquetado de imágenes**

Debido a que el entrenamiento de las redes neuronales es de tipo supervisado, se requiere proveer de información correcta para evaluar el error. La red Faster R-CNN requiere la clasificación por grado de madurez y la ubicación de frutos dentro de las imágenes. Para lograr esto se realizó el etiquetado de las imágenes, dibujando un cuadrado alrededor del objeto de interés en cada una de las imágenes y posteriormente se estableció la clase a la que pertenecen. Se utilizó la herramienta *Labelimg* (Tzutalin., 2015) para realizar la tarea de etiquetado. Este software permite guardar la información en un archivo .xml en formato PASCAL VOC. El archivo generado para cada una de las imágenes incluye las coordenadas de cada uno de los frutos dentro de estas y la clase a la que pertenecen. Se obtuvieron un total de 18 294 etiquetas en total con una distribución por clase, como se indica en el Cuadro 3.

La red Mask R-CNN requiere además de la imagen y el archivo xml con la información de los cuadros delimitadores, una imagen .png que contiene la máscara para identificar los píxeles que pertenecen a la clase, donde cada una

de las clases tiene un color diferente y el fondo se considera una clase extra. Para la generación de las máscaras se utilizó la herramienta “PixelAnnotation Tool” (Breheret, 2017). Este programa permite de forma semi-manual la segmentación de la imagen utilizando el algoritmo “watershed marked” de la librería OpenCV (OpenCV Foundation, California, USA). Para realizar la segmentación se debe seleccionar el fondo y los objetos de interés como marcadores, se aplica el algoritmo “watershed”, posteriormente se corrigen los posibles errores en la segmentación.

Cuadro 3. Número de imágenes etiquetadas por clase

<b>Clase</b>	<b>Entrenamiento</b>	<b>Prueba</b>
No Maduro	800	201
Verde	2 619	655
Quebrado	1 159	290
Rayado	1 192	298
Rosa	2 040	510
Rojo claro	2 521	631
Rojo	4 302	1 076
<b>Total</b>	<b>14 633</b>	<b>3 661</b>

### **Generación de archivos de entrenamiento**

Para la transferencia de aprendizaje de las redes neuronales, se utilizó la librería Tensorflow (Abadi et al., 2016) que consiste en un conjunto de algoritmos para el aprendizaje automático. Tensorflow utiliza como formato estándar, para la entrada de datos, los TFRecords (Tensorflow record format) que se crean a partir de las imágenes y los archivos “xmls” provenientes del etiquetado de las imágenes.

### **Entrenamiento de la red**

Para el entrenamiento de la red se utilizó una tasa de aprendizaje de 0.0002 con un número de pasos de 450,000 y un parámetro para detener el entrenamiento de un error de 0.02 en 30 evaluaciones continuas, un tamaño de lote de 5 imágenes, un paso de pixel a pixel de 16 (stride) y una intersección sobre unión de 0.6 (Intersection over Union, Iou).

### 3.2.6 Evaluación del algoritmo

Para evaluar el desempeño del algoritmo implementado, se propusieron cuatro escenarios diferentes: 1) imágenes con frutos individuales con fondo uniforme e iluminación controlada, 2) imágenes con frutos individuales con fondo no uniforme e iluminación natural, 3) imágenes con frutos de jitomates en grupo con iluminación natural y fondo no uniforme, y 4) imágenes con racimos de jitomate en condiciones de invernadero, donde existe la obstrucción parcial de frutos, iluminación natural y fondo no uniforme. Para determinar el desempeño del algoritmo en cada uno de los escenarios y para cada una de las tareas (detección y clasificación), se tomaron como medidas de ajuste las propuestas por Koirala et al. (2019), descritas en las expresiones 11-13 y que incluyen: 1) *precisión*, es la la probabilidad de que un objeto identificado sea un fruto de jitomate, 2) *sensibilidad*, es la capacidad para identificar los frutos y 3) el *índice F1*, el cual relaciona las dos medidas anteriores.

$$Precisión = \frac{TP}{TP+FP} \quad (11)$$

$$Sensibilidad = \frac{TP}{TP+FN} \quad (12)$$

$$F1 = 2 * \frac{Precisión*Sensibilidad}{Precisión+Sensibilidad} \quad (13)$$

Donde **TP** representa los verdaderos positivos (frutos identificados correctamente), **FP** representan los falsos positivos (Fondo identificado como fruto) y **FN** representa los falsos negativos (frutos no identificados).

### 3.3 Resultados y Discusión

#### 3.3.1 Detección de frutos

##### Condiciones controladas y frutos individuales

En la detección de frutos individuales, bajo condiciones de laboratorio (Figura 11) con iluminación controlada y con fondo uniforme, se obtuvo una precisión y sensibilidad del 100 % (Cuadro 4) tomando como frutos detectados aquellos con una probabilidad mayor al 20 %. Los frutos de jitomate fueron detectados y clasificados de forma correcta para todos los estados de maduración, y posicionados a diferentes distancias del dispositivo de visión.

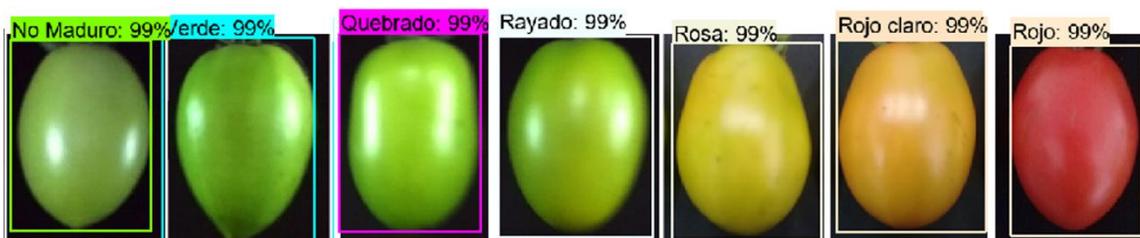


Figura 11. Detección de frutos individuales en condiciones de laboratorio

##### Iluminación natural, fondo no uniforme y frutos individuales

En el caso de la detección de jitomates con iluminación natural y un fondo no uniforme (Figura 12) se presentaron falsos positivos, lo que se traduce en una reducción de la precisión (Cuadro 4). El algoritmo identificó como frutos de jitomate objetos que formaban parte del fondo de la imagen. El algoritmo propuesto por Liu et al. (2019a) para la detección de jitomates mediante histogramas de gradientes orientados (HOG) obtuvo una precisión de 95 % y una sensibilidad de 100 % en condiciones similares. El índice F1 obtenido por Liu et al. (2019a) fue de 0.97, cercano al obtenido en la etapa de experimentación de 0.98. Cabe destacar que Liu et al. (2019a) desarrollaron el algoritmo para la detección de solo una clase (jitomates maduros) en comparación a las seis clases utilizadas en esta investigación.

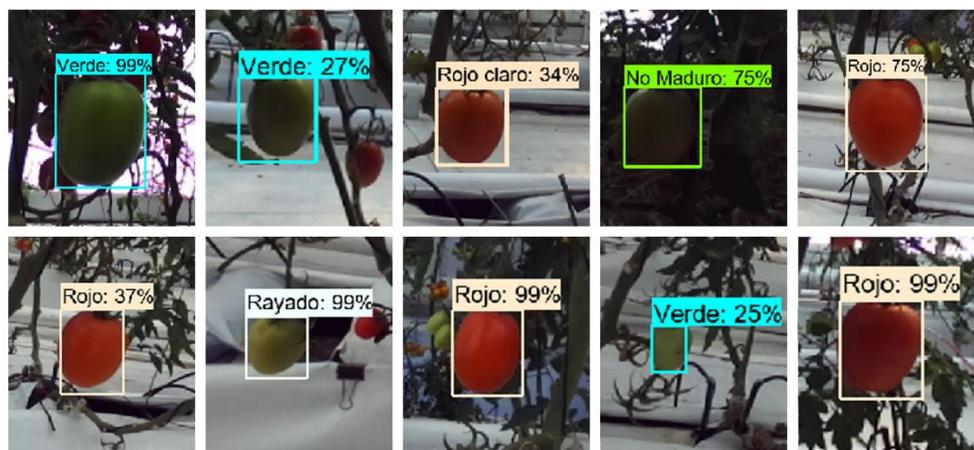


Figura 12. Detección de frutos individuales, con iluminación natural y fondo no uniforme

### Grupo de jitomates con iluminación natural y fondo no uniforme.

La detección de frutos de jitomate en racimos con iluminación natural, fondo no uniforme y sin presencia de frutos parcialmente ocultos, no presentó problemas de falsos positivos, pero algunos frutos no fueron detectados (Figura 13) obteniendo una menor sensibilidad (Cuadro 4). Bajo condiciones similares Liu et al. (2019a) obtuvieron una precisión de 94 % y una sensibilidad de 89 % en la detección de jitomates maduros, y un índice F1 de 0.92 comparado con el 0.96 obtenido en la presente investigación; esto demuestra la sensibilidad del algoritmo implementado es aceptable.

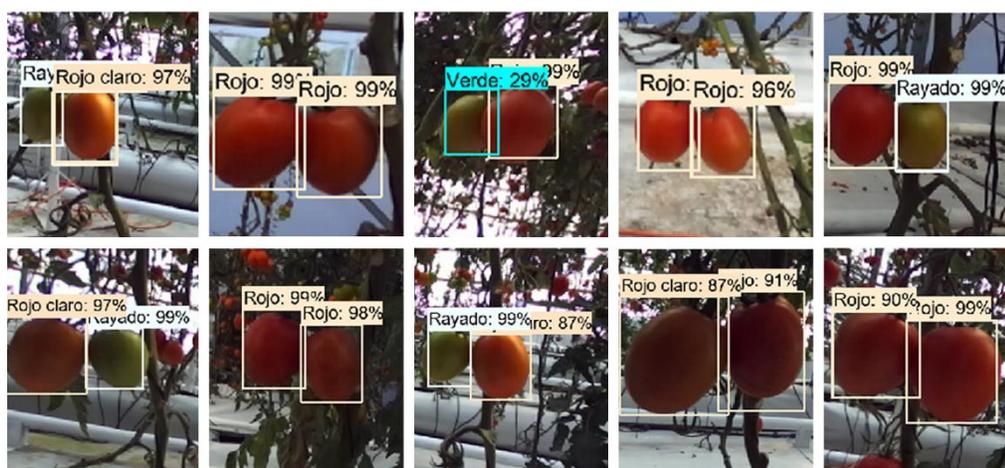


Figura 13. Detección de frutos en racimo con iluminación natural y fondo no uniforme

## Condiciones reales de invernadero

En las condiciones de iluminación reales del invernadero, donde se requiere la implementación del sistema, la detección de frutos en racimos presentó algunos problemas, tanto en la detección, como con la presencia de falsos positivos (Figura 14). Esto se traduce en una reducción en la precisión y la sensibilidad del método utilizado, obteniendo un puntaje F1 de 0.83. Diversos autores como Zhao et al. (2016b), Malik et al. (2018) y Liu et al. (2019a) han desarrollado e implementado diferentes métodos para la detección de frutos de jitomate bajo condiciones de sistemas de producción, obteniendo puntajes F1 de 0.93, 0.88 y 0.95, respectivamente. Si bien todos estos superan el obtenido en la presente investigación, éstos fueron desarrollados para la detección solamente de jitomates maduros (estado 6). Por otro lado, Yamamoto et al. (2014) implementaron un enfoque de aprendizaje de máquina, para la detección de jitomates en tres clases (Maduros, no maduros y jóvenes), los autores reportaron una precisión del 88 % y una sensibilidad del 80 % en la detección de los frutos, con un puntaje F1 de 0.84, similar al obtenido en esta investigación y que resulta bastante aceptable.

*Cuadro 4. Métricas de desempeño en los diferentes escenarios*

<b>Escenario</b>	<b>Condiciones</b>	<b>Precisión</b>	<b>Sensibilidad</b>	<b>F1 Score</b>
Individuales	Controladas	1.00	1.00	1.00
Individuales	No controladas	0.96	1.00	0.98
Grupos	No controladas	1.00	0.93	0.96
Racimos	Invernadero	0.96	0.70	0.83

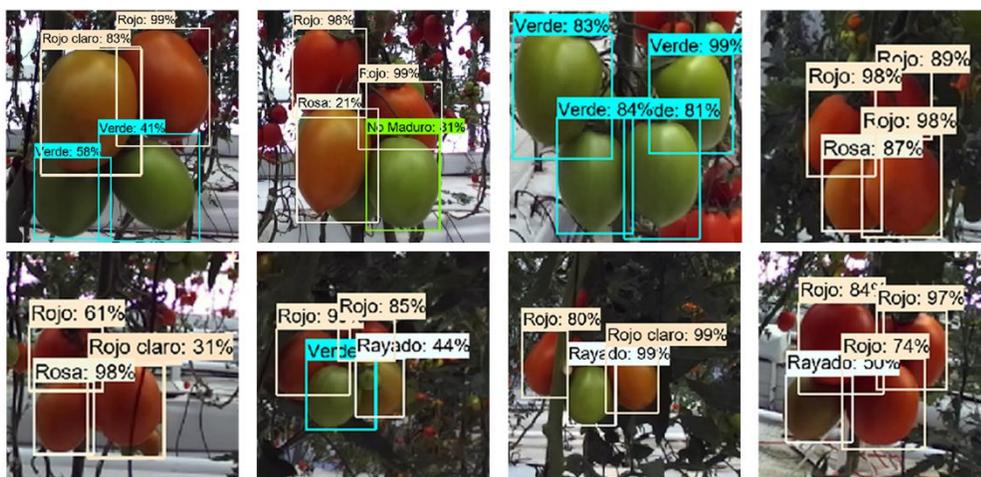


Figura 14. Detección de frutos en racimos

### 3.3.2 Clasificación de jitomates por estado de madurez.

La clasificación de los frutos de jitomate por estado de maduración en condiciones de laboratorio presentó excelentes resultados, sin embargo, en condiciones de invernadero cuando se presentan frutos de jitomate parcialmente ocultos, el desempeño del algoritmo fue menor. Esto se debe a que, en la detección de los frutos en racimo, estos presentan traslape, ocasionando, que en conjunto cumplan las características de una clase diferente. Wan et al. (2018) reportaron una exactitud de 99.3 % en la clasificación de frutos de jitomate con tres clases (rojo, anaranjado, verde), en condiciones de laboratorio. En el Cuadro 5 se muestran los resultados de la clasificación bajo condiciones de invernadero, donde solo fueron evaluados los frutos detectados con una probabilidad mayor al 20 % de pertenecer a una de las clases. Por ejemplo, para la clase “No maduro” se evaluaron 50 frutos de los cuales 47 fueron clasificados correctamente y 3 de manera errónea como pertenecientes a la clase “verde”. Por otro lado, Goel et al. (2015) propusieron un clasificador con lógica difusa, para la estimación de la madurez de jitomates en base a su coloración, siguiendo el estándar del Departamento de Agricultura de los Estados Unidos (USDA), obteniendo una exactitud del 94 %, un valor mayor al obtenido con CNN's en esta investigación. Esto muestra que los clasificadores de lógica difusa son una opción para mejorar el desempeño. En la Figura 15 se muestran las curvas de características

operativas del Receptor (ROC, por sus siglas en inglés) con el desempeño del algoritmo para cada una de las clases. Esta grafica es una representación de la sensibilidad frente a la especificidad para un clasificador binario (clase evaluada vs suma de todas las demás clases). El desempeño es medido con el área bajo la curva (AUC, por sus siglas en inglés) siendo 1 el clasificador perfecto.

Cuadro 5. Matriz de confusión.

Clase	Predicción							Precisión (%)
	0	1	2	3	4	5	6	
No Maduro (0)	47	3	0	0	0	0	0	97.9
Verde (1)	1	44	3	2	0	0	0	83.0
Quebrado (2)	0	5	41	4	0	0	0	91.1
Rayado (3)	0	1	1	48	0	0	0	85.7
Rosa (4)	0	0	0	2	46	2	0	100.0
Rojo Claro (5)	0	0	0	0	0	50	0	96.2
Rojo (6)	0	0	0	0	0	0	50	100
Sensibilidad (%)	94	88	82	96	92	100	100	93.1

**Nota:** Las filas indican el número frutos de jitomate evaluados pertenecientes a cada una de las clases y las columnas la predicción del algoritmo para cada fruto.

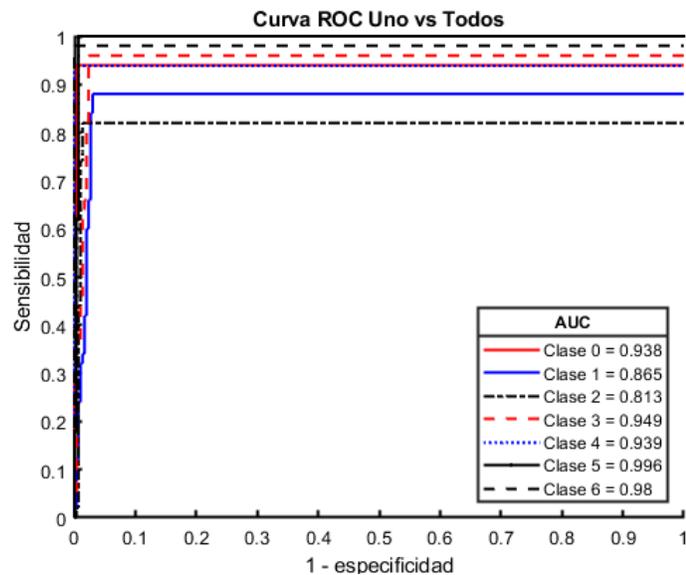


Figura 153. Curvas ROC bajo condiciones de invernadero

### 3.3.3 Segmentación de frutos

La segmentación de los frutos es una parte importante en el proceso de detección adecuada del contorno de éstos. Este paso consiste en la clasificación de los píxeles dentro de una región de interés. Para esta tarea se propusieron dos clases (fruto de jitomate y fondo), la red neuronal Mask R-CNN clasificó los píxeles de acuerdo a éstas clases y se evaluó su precisión y sensibilidad (Ecuaciones 11 y 12, respectivamente), donde se consideran como falsos positivos las detecciones que no son correctas. En la Figura 16 se muestra lo que se consideró como una segmentación correcta e incorrecta. Las segmentaciones correctas cubren casi por completo el fruto y las incorrectas dejan fuera porciones de los frutos de jitomate. El algoritmo obtuvo una precisión de 0.89 % y una sensibilidad de 0.98 % en la segmentación de jitomates, dando como resultado un puntaje  $F1$  de 0.93.

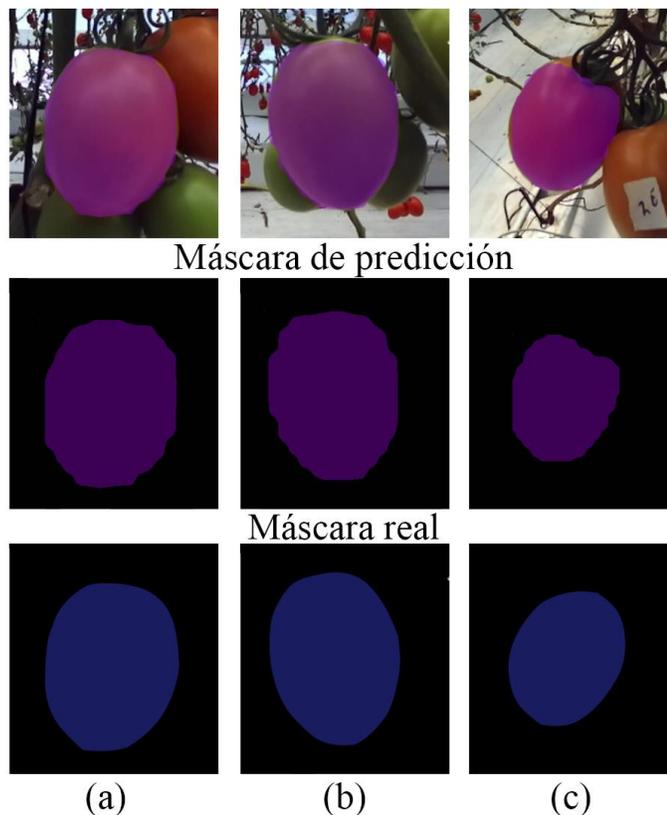


Figura 16. Segmentación de frutos: correcta (a, b) e incorrecta (c)

### 3.4 Conclusiones

Las redes neuronales propuestas en la detección de jitomates en condiciones de laboratorio presentaron buenos resultados, no hubo presencia de errores con ninguna clase. En el caso de condiciones con iluminación natural y fondo no uniforme, algunos objetos del fondo fueron catalogados como jitomates. Esto se debe a que los datos de entrenamiento no fueron incluidos como parte del fondo y cumplían, en cierta medida, con los requisitos necesarios para ser considerados como frutos de jitomates. Los falsos negativos (FN) se presentaron con mayor frecuencia en los escenarios con grupos de frutos. El principal inconveniente en la detección de frutos se presentó cuando existen frutos parcialmente ocultos por otros frutos, hojas y tallos. Esto ocasiona que el puntaje en la identificación sea bajo y sean ignorados. Cabe destacar que, si bien otros autores como Zhao et al. (2016b), Malik et al. (2018) y Liu et al. (2019a) obtuvieron mejores resultados en la detección de jitomates, éstos detectaron jitomates de la clase maduros (6). En el presente estudio el algoritmo utilizado detecta 7 clases diferentes, lo que se traduce en un aumento de complejidad.

La clasificación de los jitomates presentó inconvenientes con las clases verde y quebrado, ya que para cambiar de una clase a la otra hace falta solo un cambio de menos del 10 % del color en la superficie, esto depende en gran medida del ángulo con el cual fue capturada la imagen. Por otro lado, el uso de un clasificador de lógica difusa se muestra como una opción viable como el propuesto por Goel et al. (2015), con el cual han obtenido resultados similares, sin embargo, en esta investigación se utilizó un clasificador con CNNs debido al potencial que ha demostrado en otros campos. Se sugiere implementar un clasificador de lógica difusa como una opción para mejorar el desempeño del clasificador de CNNs.

El entrenamiento de la red Mask R-CNN requirió una menor cantidad de datos para la obtención de resultados similares a los de la red Faster R-CNN de una sensibilidad 70 % bajo condiciones de invernadero, esto como resultado de que el algoritmo “aprende” las características de la región proporcionada con la máscara. Cabe destacar que este tipo de enfoques de redes de convolución requieren de una gran cantidad de información y que esta sea representativa de

la población objetivo. El uso de imágenes, tomadas en condiciones controladas de fondo uniforme e iluminación, en el entrenamiento se traduce en un pobre desempeño en condiciones diferentes, como las encontradas en ambientes de invernadero.

Los resultados obtenidos de esta investigación pueden ser utilizados para la estimación de la masa de los frutos detectados que en conjunto con la geolocalización de los frutos permitiría la estimación de rendimientos en un invernadero.

### 3.5 Referencias

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. y Zheng, X. (2016, November 2–4). *TensorFlow: Large-scale machine learning on heterogeneous systems*. Paper presented at the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), Savannah, GA, USA.
- Agarwal, A. y Dutta, S. G. (2018). Assessment of spinach seedling health status and chlorophyll content by multivariate data analysis and multiple linear regression of leaf image features. *Computers and Electronics in Agriculture*, 152, 281–289. doi:<https://doi.org/10.1016/j.compag.2018.06.048>
- Badrinarayanan, V., Kendall, A. y Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481-2495. doi:<https://doi.org/10.1109/TPAMI.2016.2644615>
- Breheret, A. h. (2017). Pixel Annotation Tool. <https://github.com/abreheret/PixelAnnotationTool>.
- Cömert, O., Hekim, M. y Adem, K. (2017). Weight and Diameter Estimation Using Image Processing and Machine Learning Techniques on Apple Images. *International Journal of Engineering Research and Development*, 9(3), 147-154 doi:<https://dx.doi.org/10.29137/umagd.350588>
- Girshick, R. (2015). Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, 1440-1448. doi:<https://doi.org/10.1109/ICCV.2015.169>
- Goel, N. y Sehgal, P. (2015). Fuzzy classification of pre-harvest tomatoes for ripeness estimation –An approach based on automatic rule learning using

- decision tree. *Applied Soft Computing*, 36, 45-56.  
doi:<http://dx.doi.org/10.1016/j.asoc.2015.07.009>
- Hamuda, E., Glavin, M. y Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, 125, 184–199.  
doi:<http://dx.doi.org/10.1016/j.compag.2016.04.024>
- He, K., Gkioxari, G., Dollár, P. y Girshick, R. (2017). Mask R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, 2980-2988.  
doi:<https://doi.org/10.1109/ICCV.2017.322>
- Heuvelink, E. (2005). *Tomatoes* (First ed.). Wageningen University: The Netherlands.
- Koirala, A., Walsh, K. B., Wang, Z. y McCarthy, C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture*, 162, 219-234.  
doi:<https://doi.org/10.1016/j.compag.2019.04.017>
- Kumaraswamy, S. K., Sastr, P. y Ramakrishnan, K. (2018). Multi-source Subnetwork-level Transfer in CNNs Using Filter-Trees. *IEEE: International Joint Conference on Neural Networks (IJCNN)*.  
doi:<https://doi.org/10.1109/IJCNN.2018.8489678>
- Lam, Y., C. Slaughter, D., W. Wallender, W. y K. Upadhyaya, S. (2007). Machine Vision Monitoring for Control of Water Advance in Furrow Irrigation. *Transactions of the ASABE*, 50(2), 371-378.  
doi:<https://doi.org/10.13031/2013.22628>
- Lecun, Y., Bottou, L., Bengio, Y. y Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11), 2278 - 2324. doi:<https://doi.org/10.1109/5.726791>
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. y Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *Springer European Conference on Computer Vision*, 8693, 740-755.  
doi:[https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- Liu, G., Mao, S. y Kim, J. H. (2019a). A Mature-Tomato Detection Algorithm Using Machine Learning and Color Analysis. *sensors*, 19(9), 2023-2042.  
doi:<https://doi.org/10.3390/s19092023>
- Liu, J., Pi, J. y Xia, L. (2019b). A novel and high precision tomato maturity recognition algorithm based on multi-level deep residual network. *Multimedia Tools and Applications*, 1-15.  
doi:<https://doi.org/10.1007/s11042-019-7648-7>
- Long, J., Shelhamer, E. y Darrell, T. (2018). Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431-3440.  
doi:<https://doi.org/10.1109/CVPR.2015.7298965>

- Ma, J., Du, K., Zheng, F., Zhang, L., Gong, Z. y Sun, Z. (2018). A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Computers and Electronics in Agriculture*, 154, 18-24. doi:<https://doi.org/10.1016/j.compag.2018.08.048>
- Malik, M. H., Zhang, T., Li, H., Zhang, M., Shabbir, S. y Saeed, A. (2018). Mature Tomato Fruit Detection Algorithm Based on improved HSV and Watershed Algorithm. *IFAC-PapersOnLine*, 51(17), 431-436. doi:<https://doi.org/10.1016/j.ifacol.2018.08.183>
- Marszalec, E. y Pietikainen, M. (1994). On-Line Color Camera Calibration. *IEEE Proceedings of 12th International Conference on Pattern Recognition*. doi:<https://doi.org/10.1109/ICPR.1994.576263>
- Nandi, C. S., Tudu, B. y Chiranjib, K. (2014). A Machine Vision-Based Maturity Prediction System for Sorting of Harvested Mangoes. *IEEE Transactions on Instrumentation and Measurement*, 63(7), 1722 - 1730. doi:<https://doi.org/10.1109/TIM.2014.2299527>
- Ortiz, L. E., Cabrera, E. V. y Goncalves, L. M. (2018). Depth Data Error Modeling of the ZED 3D Vision Sensor from Stereolabs. *ELCVIA: Electronic Letters on Computer Vision and Image Analysis*, 17(1), 1-15. doi:<https://doi.org/10.5565/rev/elcvia.1084>
- Pandurng, J. A. y Lomte, S. S. (2015). Digital Image Processing Applications in Agriculture: A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(3), 622-624.
- Partel, V., Nunes, L., Stansly, P. y Ampatzidis, Y. (2019). Automated vision-based system for monitoring Asian citrus psyllid in orchards utilizing artificial intelligence. *Computers and Electronics in Agriculture*, 162, 328-336. doi:<https://doi.org/10.1016/j.compag.2019.04.022>
- Ren, S., He, K., Girshick, R. y Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137 - 1149. doi:<https://doi.org/10.1109/TPAMI.2016.2577031>
- Tzotalin. (2015). Labellmg Git code <https://github.com/tzotalin/labellmg>.
- Wan, P., Toudeshki, A., Tan, H. y Ehsani, R. (2018). A methodology for fresh tomato maturity detection using computer vision. *Computers and Electronics in Agriculture*, 146, 43-50. doi:<https://doi.org/10.1016/j.compag.2018.01.011>
- Wang, A., Zhang, W. y Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Computers and Electronics in Agriculture*, 158, 226-240. doi:<https://doi.org/10.1016/j.compag.2019.02.005>
- Wang, Z., Walsh, K. B. y Verma, B. (2017). On-Tree Mango Fruit Size Estimation Using RGB-D Images. *sensors*, 17(12), 2738-2753. doi:<https://doi.org/10.3390/s17122738>

- Xu, G., Zhang, F., Ghafoor, S. S., Ye, Y. y Mao, H. (2011). Use of leaf color images to identify nitrogen and potassium deficient tomatoes. *Pattern Recognition Letters*, 32, 1584–1590. doi:<https://doi.org/10.1016/j.patrec.2011.04.020>
- Yamamoto, K., Guo, W., Yoshioka, Y. y Ninomiya, S. (2014). On Plant Detection of Intact Tomato Fruits Using Image Analysis and Machine Learning Methods. *sensors*, 14(7), 12191-12206. doi:<https://doi.org/10.3390/s140712191>
- Zaborowicz, M., Boniecki, P., Koszela, K., Przybylak, A. y Przyby, J. (2017). Application of neural image analysis in evaluating the quality of greenhouse tomatoes. *Scientia Horticulturae*, 218, 222-229. doi:<http://dx.doi.org/10.1016/j.scienta.2017.02.001>
- Zhao, Y., Gong, L., Huang, Y. y Chengliang;, L. (2016a). A review of key techniques of vision-based control for harvesting robot. *Computers and Electronics in Agriculture*, 127(311–323). doi:<http://dx.doi.org/10.1016/j.compag.2016.06.022>
- Zhao, Y., Gong, L., Zhou, B., Huang, Y. y Liu, C. (2016b). Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis. *Biosystems Engineering*, 148, 127-137. doi:<http://dx.doi.org/10.1016/j.biosystemseng.2016.05.001>

## **4 ESTIMACIÓN DE LA MASA EN FRUTOS DE JITOMATE CON VISIÓN ARTIFICIAL Y REDES NEURONALES ARTIFICIALES.**

WEIGHT ESTIMATION OF TOMATO FRUITS WITH MACHINE VISION AND ARTIFICIAL NEURAL NETWORKS

Christian Israel Zuñiga-Avila<sup>1</sup>, Efrén Fitz-Rodríguez<sup>1\*</sup>, José Armando Ramírez-Árias<sup>1</sup>, Raquel Salazar-Moreno<sup>1</sup>

<sup>1</sup>Posgrado en Ingeniería Agrícola y Uso Integral del Agua, Universidad Autónoma Chapingo, carretera México – Texcoco km 38.5, Texcoco, Estado de México, C.P. 56230, México.

\*Autor de correspondencia:  
efitzr@chapingo.mx (Efrén Fitz-Rodríguez)

Este capítulo fue sometido para su publicación en la revista científica *Computers and Electronics in Agriculture*.

## RESUMEN

La estimación de la masa de los frutos de un cultivo de jitomate, antes de su cosecha, resulta de interés para el pronóstico de rendimientos, programación de las cosechas, generación de estrategias de mercado, entre otras ventajas. Con los avances tecnológicos en los sistemas de visión artificial, actualmente se puede obtener información útil de los cultivos, de forma no invasiva. De esta manera, se propone estimar la masa de frutos de jitomate, fisiológicamente maduros, bajo condiciones de invernadero directamente en la planta mediante diferentes métodos. En el sistema de visión propuesto se utilizó una cámara de visión estéreo RGB-D para la captura de imágenes de frutos en planta. Para la detección, clasificación y segmentación de los frutos, se implementó una red de convolución pre-entrenada mediante transferencia de aprendizaje. Se utilizó el procesamiento de imágenes digitales, para la extracción de características geométricas de los frutos en un cultivar tipo saladette y se entrenó una red neuronal artificial multicapa para la estimación de la masa de los frutos. El sistema desarrollado con el mejor desempeño obtuvo un error cuadrático medio de 140.47, una RMSE de 11.85 g, un coeficiente de determinación de  $R^2 = 0.88$  y una eficiencia de 0.83. Los resultados obtenidos muestran que el sistema de visión artificial propuesto, en conjunto con las redes neuronales son una opción para una buena estimación de la masa de los frutos de forma no invasiva, en un cultivo de jitomate bajo condiciones de invernadero. Cabe resaltar que los frutos parcialmente ocultos, presentaron retos en la estimación de su masa durante la obtención de la forma geométrica.

**Palabras Claves:** *Procesamiento de imágenes, métodos no invasivos, monitoreo de cultivos, características geométricas.*

### **Highlights:**

- El análisis de Garson resulta adecuado como método de selección de características para la estimación de la masa de jitomates
- La estimación de diámetros ecuatorial y polar, resulta factible por medio del procesamiento de imágenes digitales.

## 4.1 Introducción

Dado que el cultivo de jitomate (*Solanum lycopersicum L.*) es una de las hortalizas más importantes por su consumo a nivel mundial (Heuvelink, 2005), y cuya producción global superó los 182 millones de toneladas en 2017 (SIAP, 2017); resulta de gran interés desarrollar sistemas de monitoreo que provean de información útil para su óptima producción y manejo.

En los sistemas de producción bajo invernadero determinar la masa de los frutos de jitomate, antes de su cosecha, permite conocer el rendimiento durante las diferentes etapas del ciclo de producción, y resulta de interés para los productores; dado que los rendimientos semanales varían durante el ciclo de producción. De acuerdo con Adams et al. (2002), éstas fluctuaciones se deben principalmente al número de frutos cosechados en semana anteriores, más que el tamaño de los frutos. Adams (2002) menciona que existe un gran interés por parte de los productores de jitomate, en sistemas que permitan la estimación de rendimientos semanales, por la necesidad de programar las cosechas de los cultivos con precisión para determinar mejores estrategias de mercado.

Diversos autores han utilizado diferentes estrategias para estimar la masa en diferentes frutos. Por ejemplo, Jorquera et al. (2017) implementaron una regresión lineal, para la estimación del peso de arándanos utilizando la medición del diámetro, en donde obtienen un coeficiente de determinación de 0.98. Por otro lado, Rad et al. (2015) evaluaron el desempeño de las redes neuronales artificiales (ANN, por sus siglas en inglés) como modelos de predicción, en conjunto con mediciones de características de las plantas en etapas tempranas, para predecir el peso final de melones, y posteriormente estimar el rendimiento total antes de la cosecha. La arquitectura de la red que generó el mejor desempeño consistió en 9-5-1, neuronas por capa, con un coeficiente de determinación de 0.87. Salawu et al. (2014) hicieron uso de las ANNs para predecir la masa de conejos, mediante mediciones de características físicas de éstos. Encontraron que las ANNs fueron más precisas en capturar, cómo los cambios en las variables de predicción afectan el peso estimado, en comparación con los modelos de regresión. Si bien para algunos de estos métodos se han

reportado buenos resultados, se debe destacar que la estimación de la masa, utilizando métodos invasivos en plantas, no solo ocasiona daños en los frutos, sino que también, la medición resulta complicada y tediosa; en contraste con los métodos no invasivos, como los sistemas de visión artificial.

Los avances en los sistemas de visión artificial y en el procesamiento de imágenes digitales, ofrecen el potencial de implementar sistemas no invasivos y eficientes utilizando cámaras digitales y computadoras. La extracción de características por medio de procesamiento de imágenes, en conjunto con otras técnicas, como las regresiones lineales y las redes neuronales artificiales, se ha utilizado en la estimación de la masa de frutos. Teoh et al. (2007) capturaron imágenes, en condiciones controladas, con una cámara fija y mediante el procesamiento de imágenes digitales, obtuvieron el área de imágenes de frutos de mangos. La masa de estos la estimaron mediante una regresión simple, obteniendo un coeficiente de determinación de 0.97. Por otro lado, Cömert et al. (2017) determinaron el área y el diámetro de frutos en manzanos, por medio del procesamiento de imágenes digitales. Estimaron la masa de los frutos, utilizando regresiones lineales, obteniendo un coeficiente de determinación de 0.93 para la estimación de los diámetros, y 0.96 en la estimación de la masa. Otros autores como Sa'ad et al. (2015), utilizaron el procesamiento de imágenes para obtener el volumen de frutos de mango, ajustándolo como un cilindro y relacionándolo con su masa por medio de una regresión lineal, cuyo coeficiente de determinación fue de 0.94. Por otro lado, Utai et al. (2018) desarrollaron un algoritmo para el reconocimiento de mangos, utilizando dos cámaras fijas y con el procesamiento de imágenes para la extracción de características. Las características fueron utilizadas como entradas a una ANN para estimar la masa de los frutos. Los autores evaluaron diferentes configuraciones de entradas para las ANN's, utilizando mediciones físicas y procesamiento de imágenes. La red que presentó el mejor desempeño utilizó las características físicas medidas con un vernier como entradas y resultó en una eficiencia de 0.99. Sin embargo, una de las redes que utilizó el procesamiento de imágenes digitales consiguió una eficiencia similar de 0.99, destacando que fue utilizando un método no invasivo. Nyalala et

al. (2019) desarrollaron un método por medio del procesamiento de imágenes digitales para la estimación del volumen de frutos de jitomate cherry, y posteriormente determinaron su masa con una ecuación de regresión. El método que propusieron obtuvo un coeficiente de correlación de 0.96, bajo condiciones de una distancia fija y fondo uniforme.

Los principales inconvenientes que presenta la estimación de la masa de frutos, mediante métodos no invasivos, son las condiciones del ambiente: iluminación irregular, diversos elementos presentes en la imagen y la presencia de un fondo no uniforme. Por esta razón, en de suma importancia implementar nuevos métodos para la descripción geométrica de los objetos. En la presente investigación tiene como objetivo estimar la masa (g) de frutos de jitomate fisiológicamente maduros, directamente en la planta antes de ser cosechados. Se propone utilizar el procesamiento de imágenes digitales y una red neuronal para la segmentación, descripción geométrica y estimación de la masa de los frutos.

## **4.2 Materiales y Métodos**

Para determinar la masa de los frutos de jitomate directamente en las plantas, se obtuvieron imágenes de un cultivo de jitomate (*Solanum lycopersicum L.*) del tipo Saladette, en un invernadero del Laboratorio de Modelación y Automatización de Biosistemas, de la Universidad Autónoma Chapingo. El desarrollo del experimento comprende el periodo de 10 de febrero del 2019 al 20 de febrero del 2020. El sistema de visión propuesto incluye tres etapas: 1) detección y clasificación de frutos bajo condiciones ambientales de invernadero, 2) extracción de características del fruto y 3) estimación de la masa de los frutos de jitomate mediante cinco métodos diferentes.

### **4.2.1 Sistema de visión artificial**

El sistema de visión consiste de una cámara RGB-D de visión estéreo modelo Zed-M (Stereolabs, Montrouge, Francia) para la captura de imágenes y una computadora portátil para el almacenamiento de la información durante el monitoreo. Una computadora de escritorio se utiliza en el post-procesamiento.

Una descripción más detallada del sistema implementado se puede consultar en Zuñiga-Avila et al. (in press).

#### 4.2.2 Funcionamiento general del sistema

El procedimiento (Figura 17) para la estimación de la masa de los frutos con el sistema de visión artificial incluye: 1) captura imágenes de frutos en la planta, 2) implementación de una red Faster R-CNN (Zuñiga-Ávila et al. , in press); esta red fue utilizada para la localización y clasificación de los frutos de jitomate por grado de madurez, 3) implementación de una red Mask R-CNN para la segmentación de los frutos dentro de las regiones de interés, 4) extracción de características geométricas de la máscara, y 5) estimación de la masa del fruto. El procedimiento se diferencia para cada uno de los métodos de estimación de la masa, en el paso 5, donde en cada uno se utilizó un enfoque diferente.



Figura 17. Esquema de funcionamiento del sistema de visión

#### 4.2.3 Medición de características geométricas de los frutos

Para obtener un modelo de regresión y entrenar una red neuronal a partir de mediciones geométricas, que permitieran determinar la masa de los frutos de jitomate, se tomó una muestra aleatoria de 930 frutos de los primeros tres racimos de cada planta en un cultivo, con diferentes características de tamaño, masa y grado de madurez. Se pesaron los jitomates individualmente en una báscula “Cole-Parmer Symmetry” de 30 kg con una precisión de 1g. Se midieron con un vernier tres diámetros, uno polar y dos ecuatoriales ortogonales (Figura 18), cada uno de los frutos. Se registró el estado de maduración asignándole un número en el rango de 0-6 (Figura 19), tomando como referencia la Norma Mexicana - de productos alimenticios no industrializados para consumo humano (NMX-FF-031-1997-SCFI, 1997),.

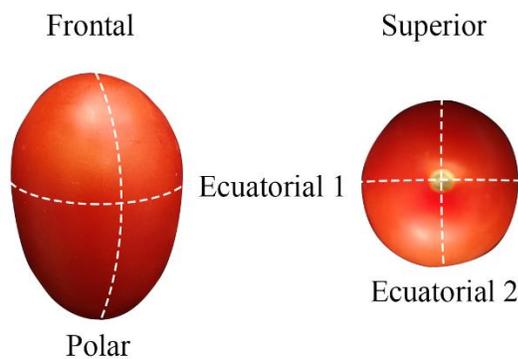


Figura 18. Diámetros del fruto

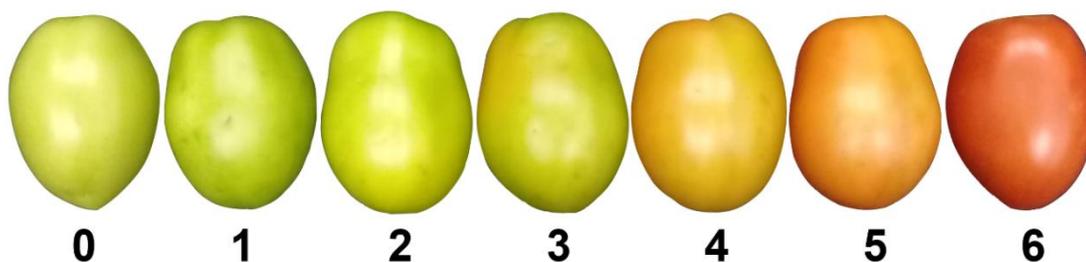


Figura 19. Estados de madurez del fruto de jitomate

De la muestra seleccionada, se eliminaron los frutos que no cumplen las especificaciones de tamaño comercial de acuerdo a la Norma Mexicana (NMX-FF-031-1997-SCFI, 1997), obteniendo una muestra final de 602 frutos de jitomate. El tamaño de los frutos se determinó con base en el diámetro ecuatorial utilizando la Norma Mexicana (NMX-FF-009, 1982). De la muestra seleccionada, se obtuvo una distribución de los frutos en cada una de las clases (Cuadro 6).

Cuadro 6. Distribución de frutos por estado de maduración

<b>Clase</b>	<b>N. frutos</b>
No Maduro	20
Verde	99
Quebrando	33
Rayado	29
Rosa	69
Rojo claro	114
Rojo	238
<b>Total</b>	<b>602</b>

#### **4.2.4 Captura de imágenes**

Para entrenar las redes neuronales propuestas, para la estimación de la masa de los frutos de jitomate, se construyó una base de datos de imágenes de frutos. Las imágenes se obtuvieron con la cámara RGB-D y se desarrolló un algoritmo en Python (Python Software Foundation, Delaware, Estados Unidos) utilizando el “Software Development Kit 2.8.3” (SDK 2.8.3) para la captura, almacenamiento y su posterior rectificación. El dispositivo almacena la información en un archivo con extensión propia (.svo), el cual contiene las imágenes sin rectificar, para cada una de las cámaras y los datos de los sensores como el giroscopio y el acelerómetro. Se capturaron aproximadamente 10 mil imágenes bajo dos condiciones: 1) condiciones controladas en laboratorio, con fondo uniforme y luz blanca y fría (provista con dos lámparas fluorescentes de 9 watts a 6500 Kelvins); las imágenes se tomaron a una distancia de 0.7 m, y 2) condiciones ambientales del invernadero, con frutos en las plantas, sobre diferentes superficies, diferentes tipos de iluminación, con frutos individuales o en grupos y a diferentes distancias de la cámara. Las imágenes se capturaron a una resolución de 1280x720 píxeles. Para rectificar las imágenes de cada una de las cámaras, se desarrolló un algoritmo en Python y utilizando la interfaz de programación de aplicaciones (API), esta interfaz es un conjunto de métodos que ofrece el kit de desarrollo de la cámara Zed-M. El proceso de rectificación consiste en corregir las imágenes tomadas por dos cámaras, para que estén alineadas y facilita la estimación de la disparidad en el cálculo de la profundidad por medio de la triangulación (Ortiz et al., 2018).

#### **4.2.5 Extracción de características geométricas**

Para la estimación de la masa de los frutos, por métodos no invasivos, fue necesario generar una base de datos para entrenar las distintas redes neuronales. De acuerdo con Mortensen et al. (2016), las características geométricas para la estimación de la masa pueden clasificarse en 1D, 2D y 3D dependiendo del espacio dimensional en cual son extraídas. Por ejemplo, el área se considera en un espacio 2D, mientras que el volumen es 3D. En esta investigación solo se aplicaron las dos primeras características.

##### **Características 1D**

###### *Estado de maduración*

Para la obtención del estado de maduración ( $\mathcal{S}$ ) de los frutos, se utilizaron redes neuronales de convolución. Lecun et al. (1998) mencionan que las redes neuronales de convolución están específicamente diseñadas para tratar con formas de dos dimensiones (2D) como las imágenes. Estas redes han demostrado superar a otras técnicas, en términos de clasificación. En esta investigación se implementó la red “Faster R-CNN inception v2”, la cual es una red desarrollada por Ren et al. (2017) y que mediante transferencia de aprendizaje fue entrenada para la detección y clasificación de los frutos en 6 clases de acuerdo con su estado de madurez, como también se propone en Zúñiga-Ávila et al. (in press). Para el entrenamiento de la red se utilizaron 18,000 etiquetas en 10,000 imágenes.

Se debe destacar la importancia de la variable estado de madurez ( $\mathcal{S}$ ), en la estimación de la masa de los frutos. De acuerdo con Priyankara et al. (2017), la masa (g) de los frutos de jitomate incrementa con la madurez del fruto. El incremento en la masa se debe a la acumulación progresiva de los sólidos totales solubles a través de los estados de maduración (Young et al., 1993). Durante el desarrollo del fruto, el contenido de azúcares se incrementa, y estos corresponden a aproximadamente la mitad de sólidos totales, con un 25 % de fructuosa y 22 % de glucosa (Young et al., 1993). Razón por la cual es de suma

importancia el estado de madurez para una mejor estimación de la masa de los frutos de jitomate.

### *Distancia a la cámara*

La distancia que existe desde la cámara a los frutos ( $D_p$ ) es de suma importancia, ya que de esta dependen, en gran medida, las características geométricas obtenidas por el procesamiento de imágenes digitales (a mayor distancia, valores más pequeños). Para la estimación de la distancia se utilizó visión estéreo mediante el procesamiento de la disparidad entre dos imágenes tomadas a diferentes posiciones (Zollhöfer et al., 2018). Los sistemas de visión estéreo dependen en gran medida de las condiciones de iluminación del ambiente. El principal inconveniente se presenta en la estimación de la distancia en los bordes de superficies lisas y curvas (Chen et al., 1995). En la Figura 20 se representa cómo es la estimación de la profundidad con cámaras de visión estéreo.

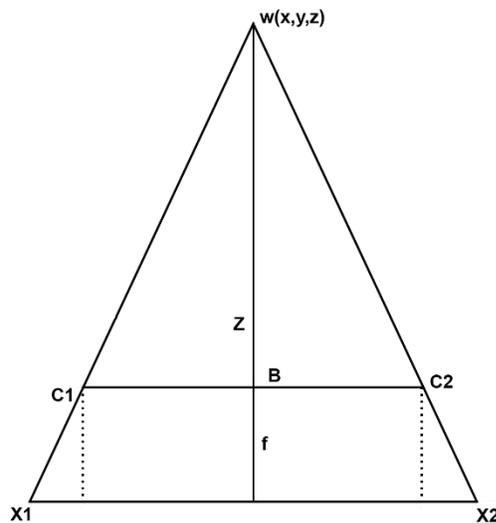


Figura 20. Estimación de profundidad

Donde  $f$  representa la distancia focal,  $B$  la línea base (separación entre centros de las cámaras,  $C1$  y  $C2$  centros de cada una de las cámaras,  $X1$  y  $X2$  corresponde a la localización del píxel en la imagen,  $w$  es un píxel dentro de la imagen y  $Z$  es la distancia desde las cámaras hasta el objeto. Para la estimación

de  $Z$  se utiliza la Ecuación 14 descrita por Fahmy (2013) donde  $X1-X2$  corresponden a la disparidad  $d$  que se obtiene de la calibración

$$Z = \frac{B \cdot f}{d} \quad (14)$$

Debido a que la distancia  $Z$ , obtenida de la ecuación anterior, corresponde a cada uno de los píxeles de la imagen, es necesario agrupar dichos valores para obtener una sola variable y que esta sea representativa. Para obtener este valor se realizó un promedio del 10 % del área central del fruto (área azul, Figura 21). Esta variable se utilizó como entrada en las redes neuronales para la estimación de la masa de los frutos de jitomate.

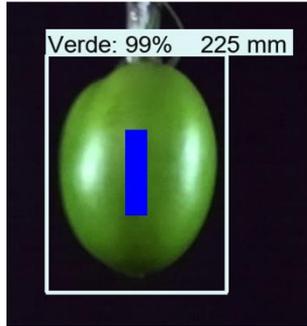


Figura 21. Estimación de la distancia promedio ( $Dp$ )

### Características 2D

Comprenden aquellas características geométricas obtenidas por el procesamiento de las imágenes. Para este caso se procesó la máscara generada por la red Mask R-CNN (Figura 22) de acuerdo a la segmentación de frutos propuesta por Zuñiga-Avila et al. (in press). La extracción de las características geométricas de frutos se implementó con algoritmos de la librería OpenCV (OpenCV Foundation, Palo Alto, Estados Unidos).

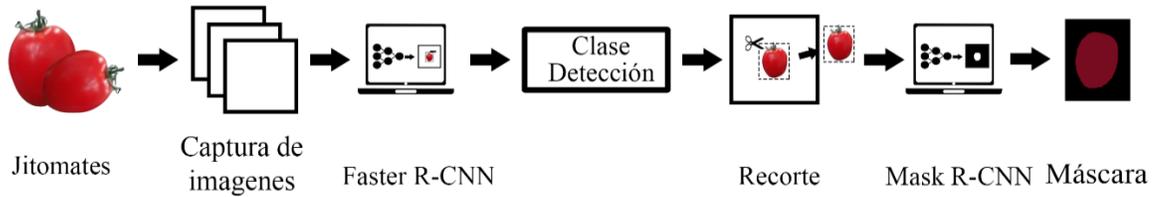


Figura 224. Proceso de obtención de la máscara

Cabe destacar, que si bien se pueden extraer una gran cantidad de características de las imágenes para la estimación de la masa, como las propuestas por Chaithanya et al. (2015) que involucran la textura, color, forma y tamaño de los objetos, algunas de estas ya están relacionadas: 1) el tamaño, está relacionado con la distancia al sensor ( $D_p$ ), y 2) el color, el cual está relacionado con estado de maduración ( $S$ ). Las características que se obtuvieron por medio del procesamiento de imágenes digitales son del tipo geométricas, que nos permiten describir la forma de los frutos. De acuerdo con Mortensen et al. (2016) estas características son las más indicadas para la estimación de la masa. Para la extracción de las características geométricas se implementó un algoritmo en Python. Las imágenes capturadas se almacenaron dentro del espacio de trabajo de OpenCV y posteriormente se les aplicó un preprocesamiento, que consiste en la reducción de ruido mediante un filtro de media, realce de bordes y operaciones morfológicas de apertura y cierre. Se aplicó la función “findContours” que permite generar una lista con todos los objetos encontrados en la región de la imagen (máscara), para posteriormente obtener las características geométricas en cada uno de estas. El número de características obtenidas se debe a la capacidad que tiene la librería OpenCV y se excluyeron aquellas que son combinación de estas. Por lo tanto, se consideraron un total 18 características geométricas por objeto, que son una combinación entre las propuestas por diferentes autores incluyendo: área ( $A$ ), perímetro ( $P$ ), ancho ( $A_n$ ) y alto ( $A_t$ ) del rectángulo circunscrito del área mínima del objeto, tasa de aspecto de los lados del rectángulo ( $TA$ ), radio de un círculo circunscrito de área mínima ( $R_c$ ), ángulo de inclinación de una elipse inscrita dentro del rectángulo rotado ( $A_r$ ), ancho ( $A_{nr}$ ) y alto del rectángulo rotado ( $A_{tr}$ ), eje mayor ( $E_M$ ) y menor de la elipse ( $E_m$ ), la excentricidad ( $E_c$ ), la extensión o rango ( $E_x$ ) que es la relación del área

de la figura entre el área del rectángulo circunscrito, diámetro de un círculo de igual área ( $D_c$ ), cuatro momentos invariantes de  $Hu (m_1, m_2, m_3, m_4)$  los cuales son descriptores de forma y son invariantes a la rotación, translación y escalado de los objetos.

#### **4.2.6 Estimación de la masa de los frutos**

De acuerdo con Jaramillo et al. (2007), los frutos de jitomate tienen diversas formas dependiendo de la variedad, la cual es seleccionada en función del mercado objetivo, ya sea para consumo fresco o procesamiento industrial. La masa de los frutos de jitomate se ha estimado con mediciones directas de los frutos (diámetros ecuatorial y polar). Taheri et al. (2011) utilizaron el diámetro ecuatorial para la estimación de la masa de frutos de jitomate, obteniendo un coeficiente de correlación de 0.91 con una regresión polinomial.

En este trabajo se proponen cinco diferentes enfoques para estimar la masa de los frutos, de acuerdo con las aproximaciones descritas en la literatura para ciertas clases de frutos. Los métodos propuestos se estructuraron en dos grupos, de acuerdo a las características de las variables de entrada: 1) métodos invasivos, aquellos en los cuales se utilizan características físicas medidas en el fruto y 2) métodos no invasivos, aquellos en donde las características geométricas y la masa de los frutos, se extraen del procesamiento de imágenes digitales. Los cinco enfoques que se proponen para la estimación de la masa de los frutos, incluyen : 1) regresión polinomial, utilizando el diámetro ecuatorial del fruto como parámetro de entrada, 2) un modelo de ANN que utiliza mediciones físicas como parámetros de entrada, 3) un modelo ANN con características obtenidas del procesamiento de imágenes como variables de entrada, 4) un modelo ANN para la estimación del diámetro ecuatorial y posteriormente implementar el modelo de regresión obtenido en el primer método y 5) un modelo de ANN para la estimación de los diámetros polar y ecuatorial del fruto, y posteriormente en conjunto con el grado de maduración, implementar la red entrenada en el segundo método. Los diferentes enfoques propuestos se esquematizan en la Figura 23, y se describen detalladamente a continuación.

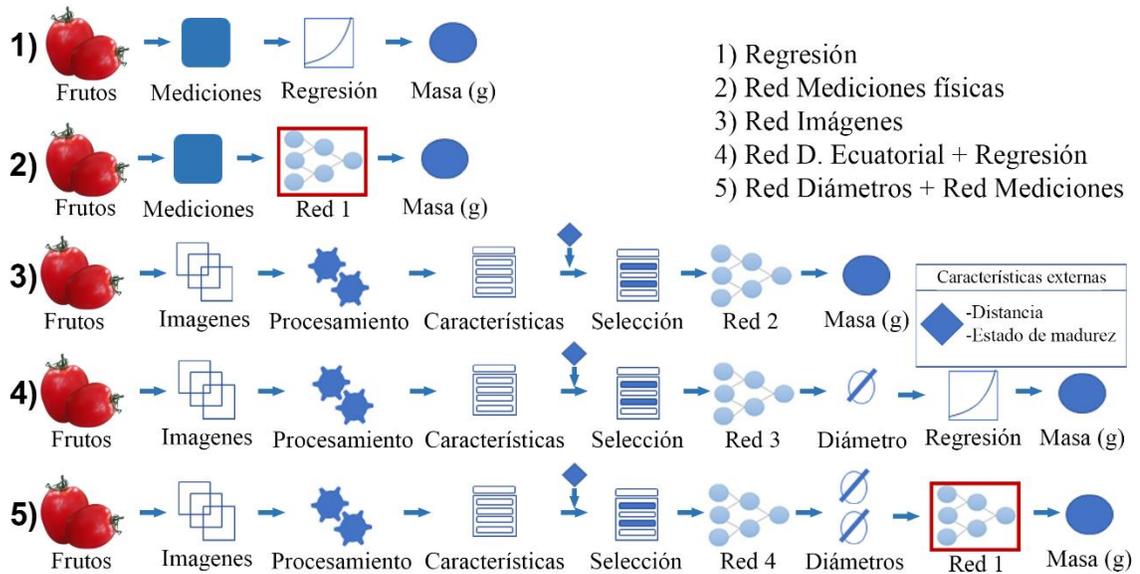


Figura 23. Diagrama de los cinco enfoques propuestos para la estimación de la masa de frutos de jitomate

### Método 1: Regresión polinomial utilizando el diámetro ecuatorial como variable independiente

En este enfoque se utilizó la metodología planteada por Jorquera et al. (2017) para la estimación de la masa, aplicando un modelo de regresión a partir de la medición de características geométricas de los frutos (diámetros ecuatoriales y polar, y estados de maduración) tomadas directamente de los frutos recién cosechados. A partir de una matriz de correlación se seleccionó la variable más relacionada con la masa (g); resultando el diámetro ecuatorial con la correlación más alta, de 0.96. De acuerdo con el análisis realizado por Taheri et al. (2011) para la estimación de la masa por medio de regresiones lineales, polinómicas y exponenciales, la regresión polinomial fue la que obtuvo mejor desempeño utilizando como variable independiente el diámetro ecuatorial (mm) y que corresponde con la aplicada en esta aproximación. Para la implementación del modelo de regresión polinomial se utilizó el método de mínimos cuadrados.

## **Método 2: ANN para la estimación de la masa a partir de mediciones directas como parámetros de entrada**

En esta aproximación para la estimación de la masa, se utilizó la metodología descrita por Utai et al. (2018), en la cual proponen una red neuronal simple, que usa como entradas las mediciones directas de algunas características geométricas de los frutos. En las ecuaciones 15 y 16 se describe el vector de entradas y salidas de la red neuronal.

$$\text{Vector de entrada } X = [D_1, D_2, S] \quad (15)$$

$$\text{Vector de salida } Y = [w] \quad (16)$$

En donde:  $D_1$  y  $D_2$  son los diámetros polar y ecuatorial en milímetros,  $S$  es el estado de maduración y  $w$  es la masa del fruto en gramos.

Para evitar el sobreajuste se eligió el número adecuado de neuronas en la capa oculta siguiendo la metodología recomendada por Haykin (1999), incrementando el número de neuronas en la capa oculta de manera progresiva, y evaluando el desempeño de la red con datos de la etapa de prueba. Finalmente se seleccionó la arquitectura de red con el mejor desempeño.

El entrenamiento de la red se evaluó de 1 a 50 neuronas en la capa oculta con 100 repeticiones para cada una de las ANN y posteriormente se seleccionó la arquitectura con la mayor media recortada de la eficiencia (EF). Esta media es más robusta a los valores extremos, ya que elimina un porcentaje de los datos. Haykin (1999) sugiere dividir la información disponible en dos partes, la primera correspondiente al conjunto de entrenamiento (80 %) y la segunda correspondiente a la validación y prueba (20 %). De acuerdo con Crowther et al. (2005) el conjunto de datos debe ser representativo de la población y lo suficientemente grande para generar resultados significativos desde el punto de vista estadístico. Se utilizó una segunda muestra de 399 frutos para una evaluación, el cual no fue utilizado durante la etapa de entrenamiento de la ANN, y tiene como función evaluar el desempeño de la red neuronal seleccionada.

Para medir el desempeño de las ANNs se utilizaron las siguientes medidas de ajuste: 1) error cuadrado medio, MSE (Ecuación 17), el cual es un estimador que

mide el promedio de los errores al cuadrado; 2) la raíz cuadrada del error cuadrado medio RMSE (Ecuación 18), este es un indicador más intuitivo que el MSE debido a que está en la mismas unidades que los datos; 3) el coeficiente de determinación (Ecuación 19), que indica la proporción de la variación total de los datos observados explicada por los datos pronosticados (Montgomery et al., 2015) y toma valores entre 0 a 1; y 4) la eficiencia, EF (Ecuación 20) que indica el desempeño comparado con un predictor nativo (Media), el cual puede tener valores de 1 para un modelo perfecto, 0 para un modelo que no es mejor que utilizar la media y  $EF < 0$  para un modelo peor que utilizar la media (Wallach et al., 2014).

$$MSE = \left(\frac{1}{n}\right) * \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (17)$$

$$RMSE = \sqrt{\left(\frac{1}{n}\right) * \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (19)$$

$$EF = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (20)$$

Donde  $n$  es el número de valores estimados,  $\hat{y}_i$  es el  $i$ -ésimo valor estimado por el modelo,  $y_i$  es el  $i$ -ésimo valor observado y por último el valor  $\bar{y}$  el cual es la media aritmética de los valores observados.

### **Método 3: ANN para la estimación de la masa con características geométricas obtenidas del procesamiento de imágenes digitales como parámetros de entrada**

Para la estimación de la masa de los frutos, en este enfoque se basó en el método planteado por Utai et al. (2018) donde se utiliza un vector de características (Ecuación 21), obtenidas a partir del procesamiento de imágenes digitales, y que funcionan como variables de entrada a la red neuronal y se obtiene como salida la masa del fruto (Ecuación 22), en gramos. Este método se evaluó en dos situaciones: 1) con imágenes tomadas en condiciones de laboratorio, a una distancia fija a la cámara, y 2) bajo condiciones reales de iluminación ambiental

en un invernadero. Para el entrenamiento de la ANN se dividieron los datos disponibles en dos grupos, 10 % para la evaluación y 90 % restante para el proceso de diseño de la arquitectura de la red, que incluye las etapas de entrenamiento, prueba y validación. El segundo grupo a su vez se dividió en 80/10/10 correspondiente a cada una de las etapas del diseño de una red neuronal. Para determinar la arquitectura de la ANN, que consiste en definir el número de capas y el número de neuronas por capa, se aplicó un proceso iterativo en función de las medidas de ajuste anteriormente mencionadas.

$$\text{Vector de entrada } X = [A, P, T_A, A_n, A_t, A_{nr}, A_{lr}, R_c, A_r, E_M, E_m, D_C, E_x, E_c, D_p, m_1, m_2, m_3, m_4, S] \quad (21)$$

$$\text{Vector de salida } Y = [w] \quad (22)$$

Se evaluó una red neuronal (ANN) con una capa oculta, donde se incrementó el número de neuronas de 1 a 50. Como medida de ajuste se utilizó la media recortada de la eficiencia (Ecuación 20), con 30 repeticiones y como método de selección se tomó como incremento mínimo un 1 % en la medida de ajuste para agregar una neurona a la capa oculta. Con el fin de generar una ANN compacta, se realizó un análisis de sensibilidad para determinar la importancia relativa de cada una de las entradas; aplicando el método de Garson, que utiliza las matrices de pesos de la red para determinar la importancia relativa de cada una de las entradas (Olden et al., 2002). Posteriormente se determinó el número óptimo de nodos en la 1er capa oculta. y se buscó mejorar el ajuste con una 2ª capa oculta variando el número de nodos hasta encontrar el óptimo. El procedimiento aplicado para encontrar la mejor arquitectura de la red neuronal se muestra en la Figura 24.

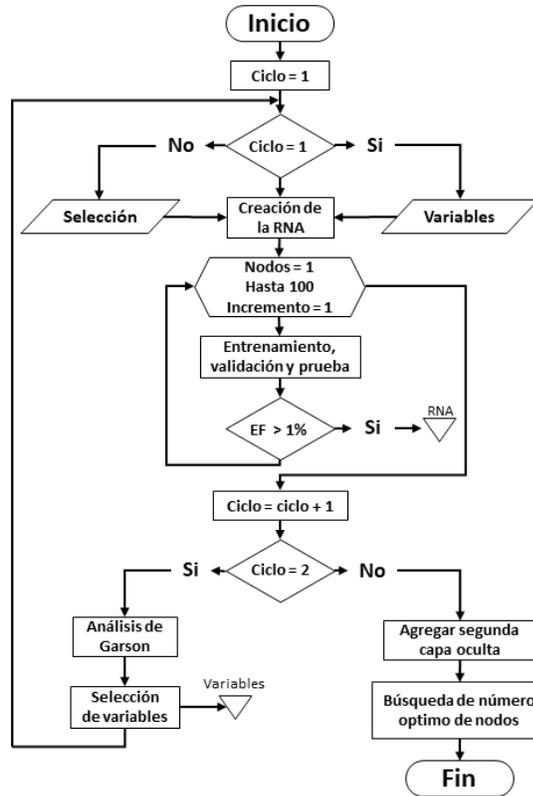


Figura 24. Diagrama de flujo del procedimiento para selección de la arquitectura óptima de la ANN

**Método 4: ANN para la estimación del diámetro ecuatorial y la implementación de la regresión del modelo 1 para obtener la masa del fruto.**

Este enfoque propuesto para la estimación de la masa de los frutos, se dividió en dos etapas: 1) estimación del diámetro ecuatorial ( $D_2$ ) mediante una ANN con un vector de características (Ecuación 23) como entradas, obtenidas a partir del procesamiento de imágenes y 2) aplicación de la regresión lineal (Modelo 1) para la estimación de la masa con el diámetro ecuatorial obtenido con la red. Para determinar el desempeño del algoritmo se utilizaron las medidas de ajuste antes mencionadas (Ecuaciones 17-20).

$$\text{Vector de entrada } X = [A, P, T_A, A_n, A_t, A_{nr}, A_{lr}, R_c, A_r, E_M, E_m, D_C, E_x, E_c, D_p, m_1, m_2, m_3, m_4, S] \quad (23)$$

$$\text{Vector de salida } Y = [D_2] \quad (24)$$

**Método 5: ANN para la estimación de los diámetros polar y ecuatorial del fruto, usados como entradas en la red del método no 2, para determinar la masa del fruto.**

Este enfoque se dividió en dos etapas: 1) implementación de una ANN para la estimación de los diámetros ecuatorial ( $D_2$ ) y polar ( $D_1$ ), utilizando el vector de características obtenidas del procesamiento de imágenes digitales, y en conjunto con el grado de madurez ( $S$ ); este último se obtuvo de la clasificación de los frutos por medio de la red Faster R-CNN, y 2) estimación de la masa de los frutos con la ANN entrenada en el método 2 (Figura 23), usando como parámetros de entrada las variables obtenidas en la etapa 1.

### **4.3 Resultados y Discusión**

#### **4.3.1 Método 1: Regresión polinomial utilizando una variable de entrada**

Con la matriz de correlación entre variables medidas directamente de los frutos, se muestra que existe una alta correlación entre los diámetros y la masa de los frutos (Figura 25); destacando el diámetro ecuatorial ( $D_2$ ) con una correlación positiva de 0.96, con respecto a la masa de los frutos. La variable que presenta menos correlación con la masa es el estado de maduración ( $S$ ) con una correlación muy baja de 0.16, el cual se determinó visualmente.

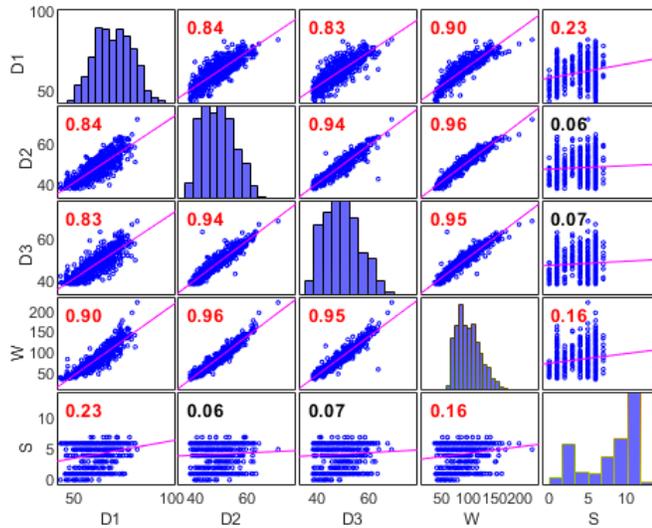


Figura 25. Matriz de correlación de mediciones físicas: diámetro polar ( $D_1$ ), diámetro ecuatorial ( $D_2$  y  $D_3$ ) y estado de maduración ( $S$ ).

El enfoque de una regresión para estimar la masa de los frutos, donde se utiliza como variable independiente el diámetro ecuatorial ( $D_2$ , mm) y como variable dependiente la masa ( $w$ , g), resultó en un MSE = 58, RMSE = 7.61,  $R^2 = 0.93$  y EF = 0.93 (Figura 26). Si bien este método resulta con un desempeño aceptable ( $R^2$  mayor a 0.90), no es recomendable por ser considerado un método invasivo, pero resulta necesario para evaluar los otros enfoques propuestos.

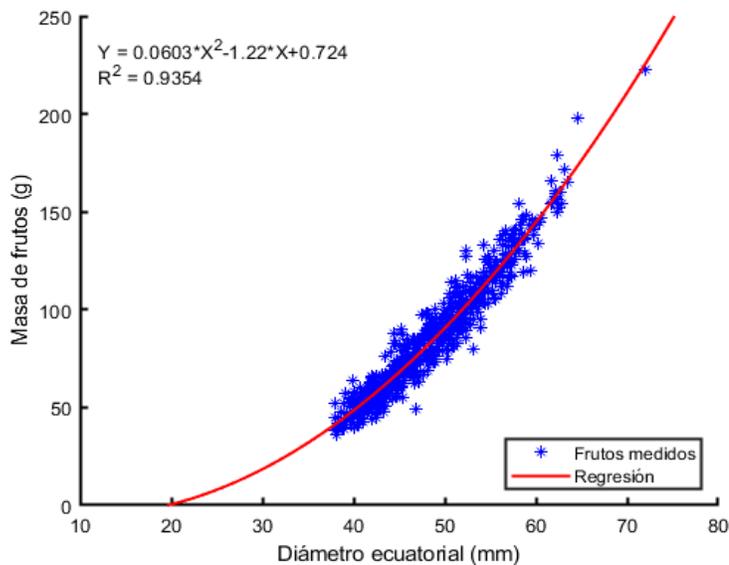


Figura 26. Regresión polinomial de la masa del fruto en función del diámetro ecuatorial ( $D_2$ )

### 4.3.2 Método 2: ANN para la estimación de la masa de los frutos con mediciones físicas como parámetros de entrada

El diseño de una red neuronal con una capa oculta con mediciones físicas ( $D_1$ ,  $D_2$  y  $S$ ) como entradas y la masa de los frutos ( $w$ ) como salida, consiste en determinar el número de neuronas en la capa oculta, a partir de los vectores de entrada y salida. Se probaron de manera heurística el número de neuronas en la capa oculta, desde 1 hasta 50 neuronas. La red neuronal, que arroja los mejores resultados de acuerdo con su eficiencia (Figura 27) y desempeño (Cuadro 7) tiene 5 neuronas en la capa oculta, resultando con una arquitectura de [3 5 1] neuronas en cada una de las capas, entrada, oculta y salida, respectivamente. Esta ANN cumplió el criterio de selección, donde para justificar el aumento de neuronas se tomó como requerimiento mínimo un incremento en la eficiencia del 1 %.

Cuadro 7. Desempeño de la ANN durante cada una de las etapas de diseño.

Entrenamiento				Validación				Prueba			
MSE	RMSE	R2	EF	MSE	RMSE	R2	EF	MSE	RMSE	R2	EF
22.7	4.76	0.97	0.97	34.4	5.87	0.96	0.96	14.37	3.79	0.98	0.98

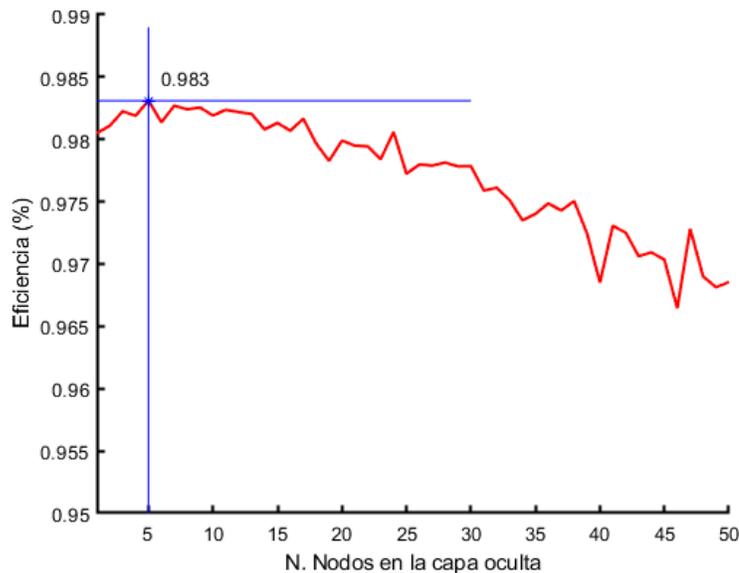


Figura 27. Eficiencia de las ANNs propuestas, en función del número de nodos en la capa oculta

Una vez determinado el número de neuronas óptimas en la capa oculta, se realizó el análisis de Garson para conocer la importancia relativa de cada una de las variables (Figura 28). Este análisis mostró que el diámetro ecuatorial presenta la importancia relativa más alta en la estimación de la masa de los frutos de jitomate. Esto explica por qué la regresión polinomial, que utiliza el diámetro ecuatorial como variable independiente, genera resultados aceptables ( $R^2$  mayor a 0.90) en la predicción de la masa de los frutos.

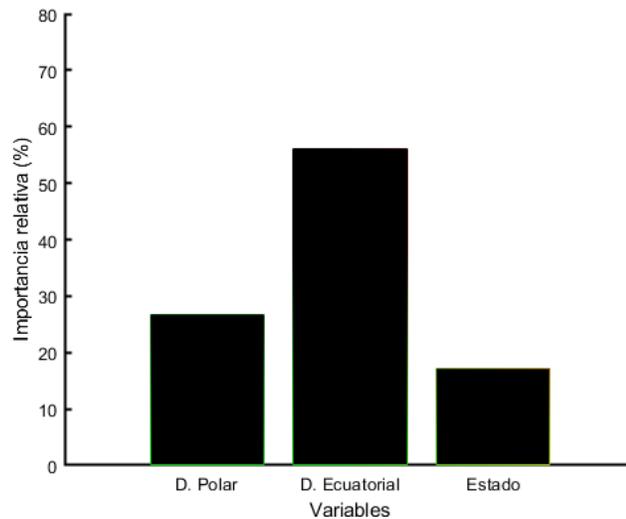


Figura 28. Importancia relativa de las variables medidas de acuerdo con el análisis de Garson.

Una vez establecida la arquitectura de la ANN para la predicción de la masa de los frutos, se realizó una evaluación con una muestra de 120 frutos (Figura 29), en la cual se hace uso de datos que no fueron utilizados durante la etapa diseño y entrenamiento de la red, lo que permitió evaluar de forma real el comportamiento que tiene la ANN ante datos nuevos.

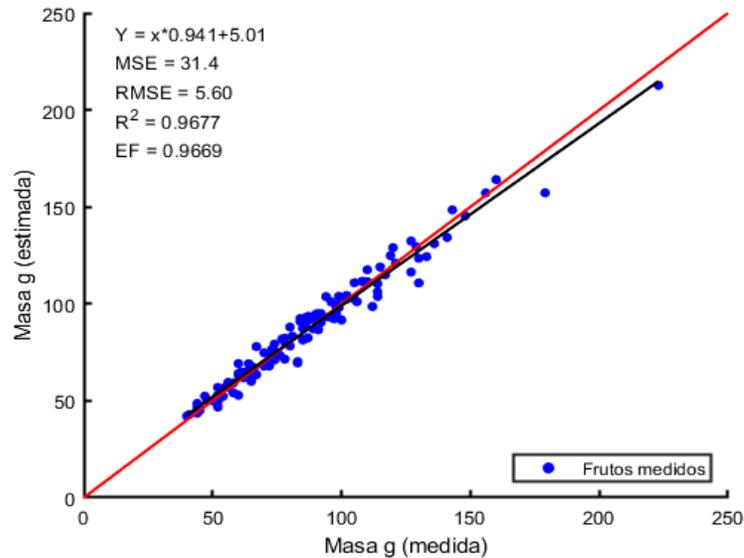


Figura 29. Evaluación de una segunda aproximación: Relación masa (g) estimada con la red neural vs datos de medición de frutos

#### 4.3.3 Método 3: ANN para la estimación de la masa de frutos a partir de características obtenidas del procesamiento de imágenes digitales como parámetros de entrada

En esta aproximación, para determinar la masa de los frutos, la ANN implementada utilizó como entradas las características extraídas (Ecuación 21) de las imágenes por medio del procesamiento digital y la masa de los frutos (g) como variable de salida. La ANN que resultó con un mejor desempeño tiene una arquitectura de 8 neuronas en la capa oculta [20, 8, 1]. Este método se evaluó en dos situaciones: 1) con imágenes tomadas en condiciones de laboratorio, a una distancia fija a la cámara, y 2) con imágenes bajo condiciones ambientales en invernadero.

##### Condiciones de laboratorio

La estimación de la masa de frutos bajo condiciones de laboratorio difiere de la aplicación real en la variable de la distancia a la cámara. Bajo condiciones de laboratorio esta distancia fue constante, por lo tanto, la influencia de la variable en la estimación de masa fue menor. El objetivo de evaluar el algoritmo bajo estas condiciones fue conocer el desempeño en un ambiente más controlado. El desempeño durante la etapa de diseño (entrenamiento y validación) y prueba de

la red se muestra en el Cuadro 8, y la evaluación de la red se muestra en la Figura 30.

Cuadro 8. Desempeño ANN en el método 3, con imágenes tomadas en condiciones controladas

Entrenamiento				Validación				Prueba			
MSE	RMSE	R2	EF	MSE	RMSE	R2	EF	MSE	RMSE	R2	EF
61.5	7.84	0.93	0.93	141.1	11.8	0.83	0.83	93.0	9.64	0.91	0.91

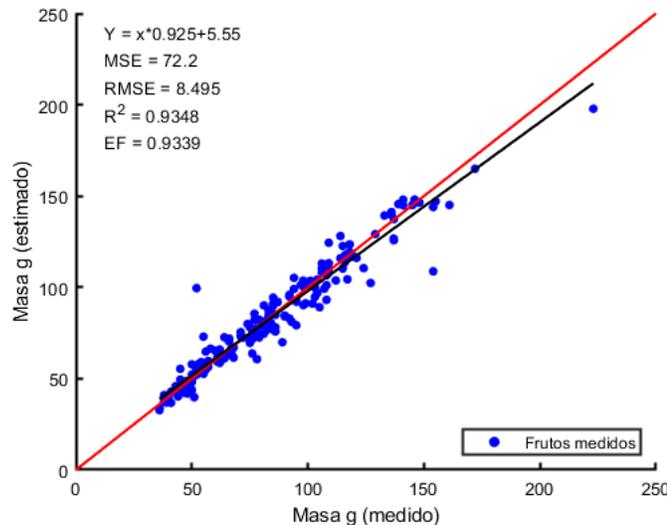


Figura 30. Evaluación de la aproximación 3 en condiciones de laboratorio: Correlación de masa (g) estimada con la red neural vs datos de medición de frutos

### Condiciones de invernadero

Debido a que bajo condiciones de invernadero existe una mayor variación de la *Distancia a la cámara* ( $D_p$ ), la variable tiene una mayor influencia en la estimación de la masa de los frutos. La ANN implementada utilizó como entradas las características extraídas de las imágenes (Ecuación 21) por medio del procesamiento digital y la masa (g) como variable de salida. La ANN que presentó el mejor desempeño resultó con una arquitectura de [20 8 1], donde 8 neuronas se encuentran en la capa oculta. Esta red cumplió el criterio de selección planteado en la Figura 24, donde solo se agrega una neurona si la eficiencia mejora mínimo un 1 %. En la Figura 31 se muestra un comportamiento decreciente en la eficiencia cuando se incrementa el número de neuronas en la

capa oculta. Las medidas de ajuste de la ANN durante la etapa de diseño (entrenamiento, validación y prueba) de la red se presentan en la Cuadro 9.

Cuadro 9. Desempeño ANN para estimación de la masa con imágenes de invernadero.

Entrenamiento				Validación				Prueba			
MSE	RMSE	R2	EF	MSE	RMSE	R2	EF	MSE	RMSE	R2	EF
64.06	8.00	0.85	0.85	54.37	7.37	0.87	0.87	67.18	8.20	0.88	0.87

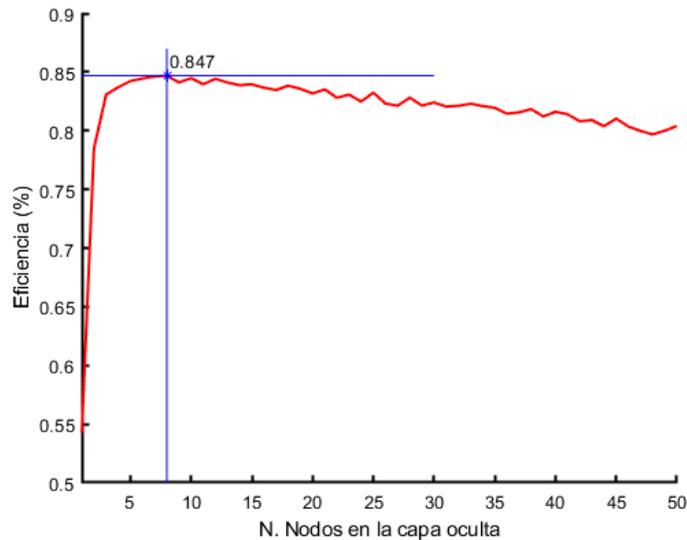


Figura 31. Eficiencia en cada una de las ANN de acuerdo con el número de nodos en la capa oculta.

Una vez diseñada la ANN adecuada, se realizó el análisis de Garson, que permitió conocer la importancia relativa de cada una de las características geométricas, extraídas con algoritmos de procesamiento de imágenes (Figura 32), y que funcionan como entradas de la red neuronal.

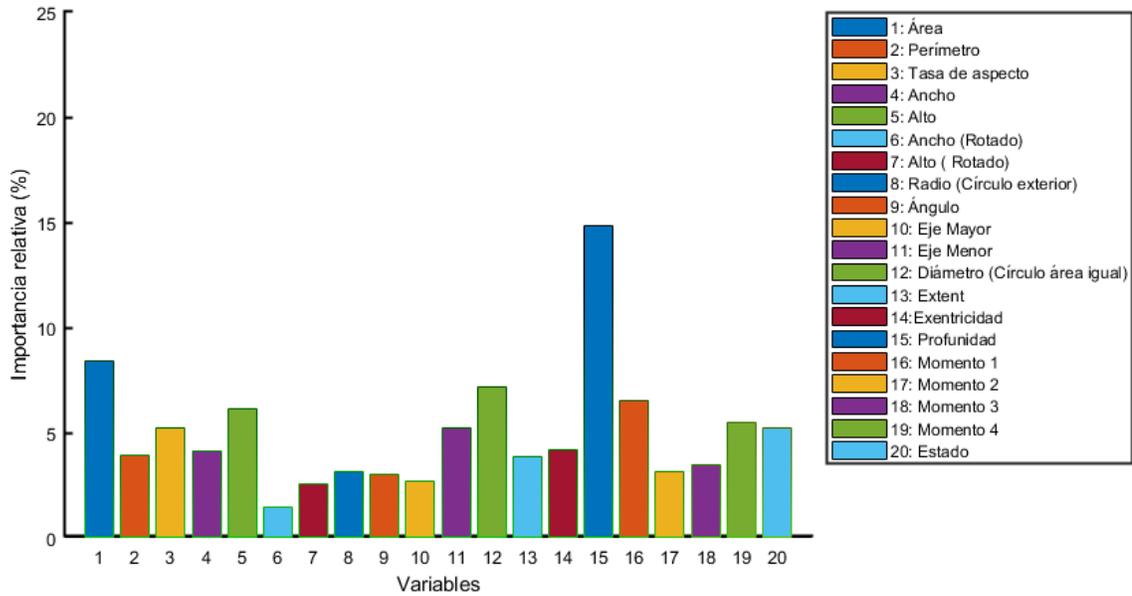


Figura 32. Importancia relativa de las características extraídas de las imágenes

Para la selección de las variables se determinó una reducción máxima del 10 % en la eficiencia del algoritmo. Se dividió el total de variables en 9 grupos de acuerdo a la importancia relativa obtenida del análisis de Garson (1, 2, 3, 4, 5, 6, 7, 8 y 14 %), haciendo referencia a que el grupo 9 incluye todas las variables que obtuvieron una importancia relativa mayor o igual al 14 %, para este caso solamente la variable Distancia a la cámara ( $D_p$ ) y para el grupo 1 incluye las variables mayores o iguales a 1 %, que corresponde a la totalidad de las variables. En la Figura 33 se muestra el desempeño del algoritmo para cada conjunto de entradas, fue seleccionado el grupo 4 que contiene las variables mayores o iguales a 4 % de importancia relativa con un total de 11 variables de entrada (Ecuación 25).

$$\text{Vector de entrada } X = [D_p, A, D_C, m_1, A_t, m_4, S, T_A, E_m, E_x, A_n] \quad (25)$$

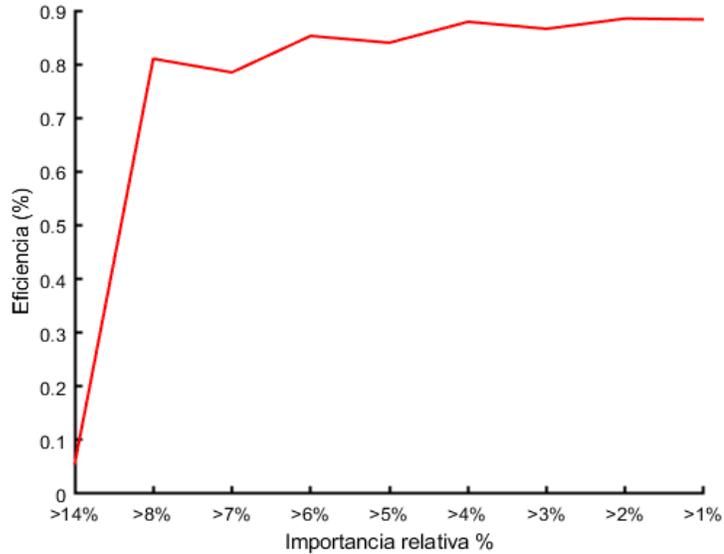


Figura 33. Eficiencia en diversos conjuntos de entradas

Se creó una ANN con un menor número de variables de entradas (Ecuación 25) y se determinó de manera heurística el número óptimo de neuronas en la capa oculta (Figura 34). Se observa una tendencia decreciente en la eficiencia de las redes, al incrementar el número de neuronas en la capa oculta. El número óptimo de neuronas en la capa oculta es de 10 neuronas. En el Cuadro 10 se muestra las medidas de ajuste obtenidas durante la etapa de entrenamiento de la ANN seleccionada.

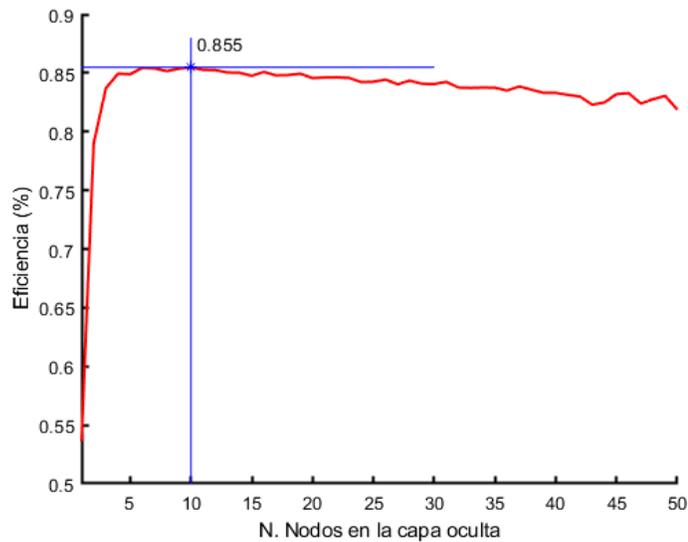


Figura 34. Eficiencia de la ANN en función del número de neuronas en la capa oculta.

Cuadro 10. Desempeño ANN que utilizó imágenes durante la creación de la red

Entrenamiento				Validación				Prueba			
MSE	RMSE	R2	EF	MSE	RMSE	R2	EF	MSE	RMSE	R2	EF
66.83	8.17	0.84	0.84	60.93	7.81	0.85	0.85	65.47	8.09	0.88	0.87

Después de seleccionar el número óptimo de neuronas en la capa oculta de la ANN, con las variables de entrada que presentaron mayor importancia relativa, se realizó el análisis para implementar una segunda capa oculta en la red, los resultados se presentan en la Figura 35.

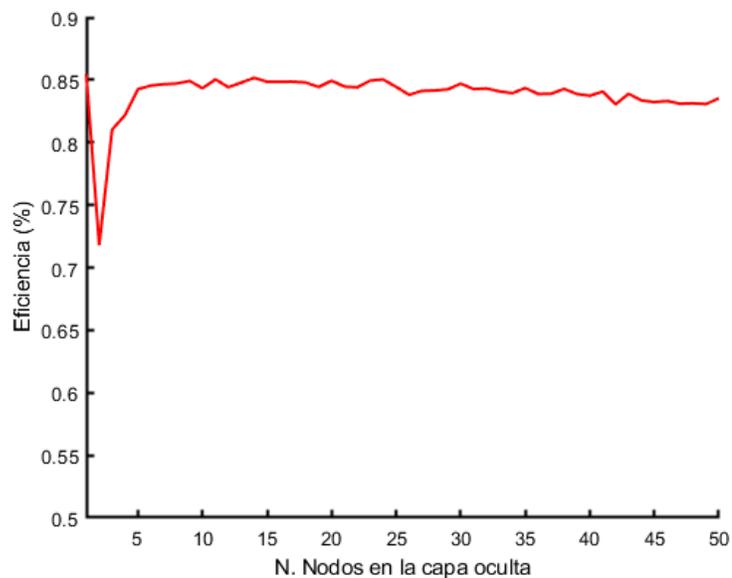


Figura 35. Eficiencia con el incremento de nodos en la segunda capa oculta.

La evaluación realizada con una segunda capa y diferentes números de neuronas (1 a 50) no mostró una mejora significativa en el desempeño de la red, por tal motivo fue seleccionada una arquitectura de ANN con una sola capa oculta.

Finalmente, para conocer el comportamiento de la ANN con información no utilizada durante el proceso de entrenamiento, se realizó la evaluación, la cual resultó con una eficiencia de 0.84 (Figura 36).

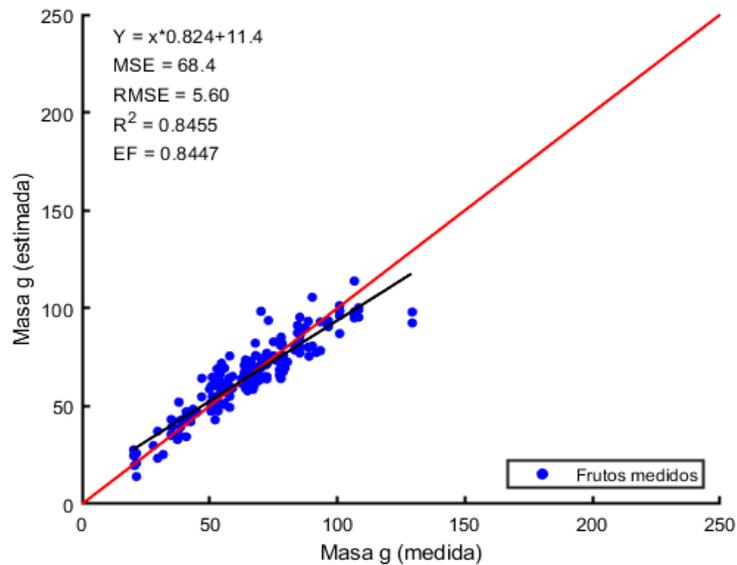


Figura 36. Evaluación de la ANN para la aproximación 3, en condiciones de invernadero: Relación masa (g) estimada con la red neural vs datos de medición de frutos

#### 4.3.4 Estimación de diámetros

La selección del número óptimo de nodos en la capa oculta para cada una de las ANNs se realizó con el procedimiento descrito en la Figura 24. Para la ANN que estimó el diámetro polar ( $D_1$ ) se seleccionaron 6 neuronas en la capa oculta, y para la segunda ANN que estima el diámetro ecuatorial ( $D_2$ ) se seleccionaron 7 neuronas en la capa oculta. Para la ANN de estimación del diámetro polar ( $D_1$ ) no se justificó agregar una nueva capa oculta debido a que no hubo un incremento en la eficiencia. Para la ANN de estimación del diámetro ecuatorial ( $D_2$ ) se obtuvo un incremento en la eficiencia al agregar una segunda capa oculta con 5 neuronas. Del análisis de Garson, se eligieron aquellas variables con una importancia relativa mayor a 4 %.

Una vez diseñadas las ANN, se realizó una evaluación de cada una de las ANNs con datos no utilizados durante la creación y entrenamiento de las redes seleccionadas (Cuadro 11). Se observan desempeños similares para la estimación de los diámetros.

Cuadro 11. Desempeño de las ANNs para estimar diámetros

ANN	Arquitectura	Prueba			
		MSE	RMSE	R2	EF
D. Polar	8-6-1	6.88	2.62	0.89	0.88
D. Ecuatorial	15-7-5-1	3.65	1.91	0.87	0.87

#### 4.3.5 Método 4: Modelo ANN para la estimación del diámetro ecuatorial y la obtención de la masa de frutos con la regresión del Método 1

Este enfoque se realizó con la ANN seleccionada para la estimación del diámetro ecuatorial ( $D_2$ ) en conjunto con la regresión polinomial para la predicción de la masa de los frutos de jitomate. Con una evaluación se muestra un desempeño menor al obtenido por método 2 (ANN Imágenes) (Cuadro 12). Si bien las medidas obtenidas para la estimación del diámetro fueron cercanas a  $R^2=0.90$ , al utilizarlas para predecir el peso con la regresión se incrementó el error en la estimación de la masa del fruto.

Cuadro 12. ANN para la estimación del peso en función diámetro estimado y la regresión polinomial

ANN	Desempeño en la evaluación			
	MSE	RMSE	R2	EF
D. Ecuatorial + Regresión	91.71	9.58	0.81	0.79

#### 4.3.6 Método 5: Modelo ANN para la estimación de los diámetros polar y ecuatorial del fruto y la obtención de la masa del fruto con la red del punto 2.

En esta aproximación para la estimación de la masa de los frutos de jitomate se utilizaron las ANNs seleccionadas para pronosticar el diámetro polar y ecuatorial ( $D_1$  y  $D_2$ ) de los frutos, en conjunto con el estado de maduración ( $S$ ) obtenido de la red de convolución (Faster-CNN). Posteriormente se hizo uso de la red neuronal entrenada con mediciones físicas (4.3.2) para la predicción de la masa. Utilizando los datos para la evaluación se obtuvieron las medidas de ajuste presentadas en el Cuadro 13.

Cuadro 13. Estimación de peso con red de mediciones físicas e imágenes

<b>Evaluación</b>			
MSE	RMSE	R2	EF
73.24	8.43	0.87	0.83

#### 4.3.7 Comparación enfoques para la predicción de masa

Se realizó una evaluación de los diferentes enfoques con datos nuevos de laboratorio e invernadero, los métodos invasivos (1 y 2) fueron evaluados con datos bajo condiciones de laboratorio, los métodos no invasivos (4 y 5) bajo condiciones de invernadero, y el método 3 (ANN imágenes) fue evaluado bajo condiciones de laboratorio e invernadero. Del grupo de métodos no invasivos fue seleccionada la ANN con la más alta eficiencia en la estimación de la masa de los frutos, los resultados se muestran en el Cuadro 14.

Cuadro 14. Medidas de ajuste con un nuevo set de datos

<b>Método</b>	<b>Prueba</b>			
	MSE	RMSE	R2	EF
<b>Regresión polinomial <sup>1</sup></b>	42.58	6.52	0.89	0.89
<b>ANN mediciones físicas <sup>2</sup></b>	41.68	6.45	0.95	0.90
<b>ANN imágenes (laboratorio) <sup>3</sup></b>	72.16	8.49	0.93	0.93
<b>ANN imágenes (invernadero) <sup>3</sup></b>	137.43	11.72	0.88	0.83
<b>ANN ecuatorial + regresión <sup>4</sup></b>	255.63	15.98	0.73	0.69
<b>ANN diámetros + ANN mediciones <sup>5</sup></b>	150.34	12.26	0.85	0.82

<sup>1</sup>Regresión polinomial utilizando el diámetro ecuatorial como variable independiente. <sup>2</sup>ANN para la estimación de la masa con mediciones físicas como parámetros de entrada. <sup>3</sup>ANN para la estimación de la masa con características obtenidas del procesamiento de imágenes digitales como parámetros de entrada. <sup>4</sup>ANN para la estimación del diámetro ecuatorial y la implementación de la regresión del modelo 1, obtener la masa del fruto. <sup>5</sup>ANN para la estimación de los diámetros polar y ecuatorial del fruto, y la obtención de la masa del fruto con la red del método no 2. <sup>6</sup>Error absoluto promedio.

La ANN que presentó la mejor medida de ajuste (Eficiencia) fue la “ANN mediciones físicas”, la ANN se encuentra dentro de la clasificación de métodos invasivos, que utilizó el diámetro polar, diámetro ecuatorial y el estado de maduración de mediciones físicas de los frutos como entradas de una red para la estimación de la masa. La mejor ANN en la clasificación de los métodos no invasivos fue “ANN imágenes”, ésta ANN hizo uso de las características extraídas por el procesamiento de imágenes digitales como entradas del modelo y el peso como salida (Figura 37).

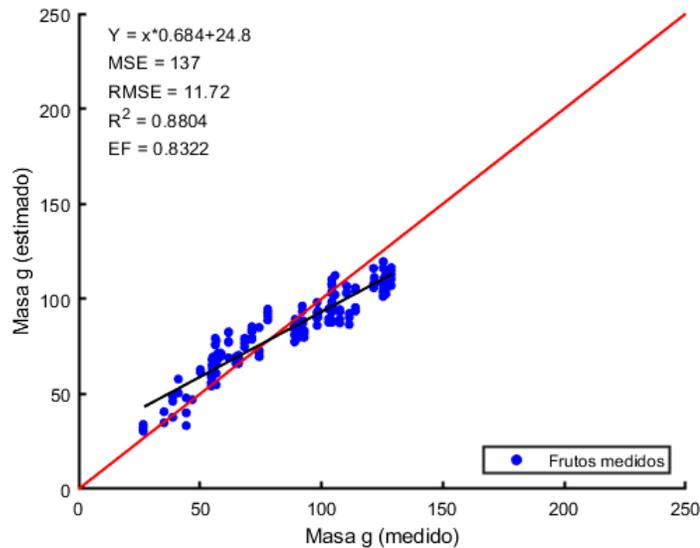


Figura 37. Masa de frutos, real vs estimado. Red ANN Imágenes

Si bien el coeficiente de determinación obtenido fue menor al que presentaron Nyalala et al. (2019) de 0.96 para la estimación de la masa, se tiene que destacar que la condiciones donde se evaluó el modelo de la presente investigación fue bajo condiciones de luz y fondo no uniforme con variaciones de la distancia a los frutos. Estos parámetros incrementan la complejidad del problema y reducen el desempeño del algoritmo propuesto.

#### 4.4 Conclusiones

De acuerdo al coeficiente de determinación el método 2 que utiliza las mediciones físicas como variables de entrada a un ANN fue el que obtuvo el mejor desempeño, sin embargo, este método no es factible de ser aplicado en campo debido a que resulta invasivo y laborioso. El mejor método no invasivo para la estimación de la masa de frutos de un cultivo de jitomate en invernadero fue el método 3. Se demuestra que, con las redes neuronales artificiales en conjunto con sistemas de visión artificial, se logran resultados aceptables en la predicción de la masa de frutos de jitomates directamente en el cultivo.

El procesamiento de imágenes digitales permitió realizar predicciones de forma no invasiva. Esta técnica es aplicable directamente en plantas, antes de la cosecha de los frutos. Sin embargo, aunque esto presenta una ventaja, se debe tener en consideración que esta clase de modelos necesita de una gran cantidad

de datos previos, de buena calidad y representativos de la población objetivo, para obtener buenas predicciones.

La segmentación de los frutos para obtener la forma geométrica, es de suma importancia para generar buenos resultados. Por lo tanto, se sugiere utilizar métodos más avanzados para la segmentación de los frutos en ambientes con mayor presencia de obstáculos y la implementación de algoritmos para la reconstrucción de frutos parcialmente ocultos.

#### 4.5 Referencias

- Adams, S. R. (2002). Predicting the Weekly Fluctuations in Glasshouse Tomato Yields *Acta Horticulturae*, 593, 19-23. doi:<https://doi.org/10.17660/ActaHortic.2002.593.1>
- Adams, S. R. y Valdés, V. M. (2002). The effect of periods of high temperature and manipulating fruit load on the pattern of tomato yields. *Horticultural Science & Biotechnology*, 77(4), 461-466. doi:<https://doi.org/10.1080/14620316.2002.11511522>
- Chaithanya, C. y Priya, S. (2015). Object Weight estimation from 2D images. *ARPN Journal of Engineering and Applied Sciences*, 10(17), 7574-7578.
- Chen, C. y Zheng, Y. F. (1995). Passive and Active Stereo Vision for Smooth Surface Detection of Deformed Plates. *Institute of Electrical and Electronics Engineers, Transactions on industrial electronics*, 42(3), 6. doi:<https://doi.org/10.1109/41.382141>
- Cömert, O., Hekim, M. y Adem, K. (2017). Weight and Diameter Estimation Using Image Processing and Machine Learning Techniques on Apple Images. *International Journal of Engineering Research and Development*, 9(3), 147-154 doi:<https://dx.doi.org/10.29137/umagd.350588>
- Crowther, P. S. y Cox, R. J. (2005). A Method for Optimal Division of Data Sets for Use in Neural Networks *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 3684. doi:[https://doi.org/10.1007/11554028\\_1](https://doi.org/10.1007/11554028_1)
- Fahmy, A. A. (2013). Stereo Vision Based Depth Estimation Algorithm In Uncalibrated Rectification. *International Journal of Video&Image Processing and Network Security*, 13(2), 1-8.
- Grange, S., Fong, T. y Baur, C. (2003). TLIB: a Real-time Computer Vision Library for HCI. *Digital Image Computing: Techniques and Applications*, 1017-1026.

- Haykin, S. (1999). *Neural Networks. A comprehensive Foundation* (A. Dworkin Ed. Second Edition ed.): Tom Robbins.
- Heuvelink, E. (2005). *Tomatoes* (First ed.). Wageningen University: The Netherlands.
- Jaramillo, J. N., Rodríguez, V. P., Guzmán, M. A., Zapata, M., C.; y Rengifo, T. M. (2007). Technical manual: Good Agricultural Practices in the Production of tomato under protected conditions. *Food and Agriculture Organization*, 316.
- Jorquera, E. F., Génard, Michel; Ribera, F. A. y Francke, N. (2017). A simple allometric model for estimating blueberry fruit weight from diameter measurements. *Scientia Horticulturae*, 219, 131–134. doi:<https://doi.org/10.1016/j.scienta.2017.03.009>
- Lecun, Y., Bottou, L., Bengio, Y. y Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11), 2278 - 2324. doi:<https://doi.org/10.1109/5.726791>
- Marszalec, E. y Pietikainen, M. (1994). On-Line Color Camera Calibration. *IEEE Proceedings of 12th International Conference on Pattern Recognition*. doi:<https://doi.org/10.1109/ICPR.1994.576263>
- Montgomery, D. C., Jennings, C. L. y Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting* (Segunda Edición ed.). Hoboken, New Jersey.: John Wiley & Sons, Inc.
- Mortensen, A. K., Lisouski, P. y Ahrendt, P. (2016). Weight prediction of broiler chickens using 3D computer vision. *Computers and Electronics in Agriculture*, 123(319-326). doi:<https://doi.org/10.1016/j.compag.2016.03.011>
- Nyalala, I., Okinda, C., Nyalala, L., Makange, N., Chao, Q., Chao, L., Yousaf, K. y Chen, K. (2019). Tomato volume and mass estimation using computer vision and machine learning algorithms: Cherry tomato model. *Journal of Food Engineering*, 263, 288-298. doi:<https://doi.org/10.1016/j.jfoodeng.2019.07.012>
- Olden, J. D. y Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154(1-2), 135-150. doi:[https://doi.org/10.1016/S0304-3800\(02\)00064-9](https://doi.org/10.1016/S0304-3800(02)00064-9)
- OpenCV, Developers Team. (2019). OpenCV. Retrieved from <https://opencv.org/about/>
- Ortiz, L. E., Cabrera, E. V. y Gonçalves, L. M. (2018). Depth Data Error Modeling of the ZED 3D Vision Sensor from Stereolabs. *ELCVIA: Electronic Letters on Computer Vision and Image Analysis*, 17(1), 1-15. doi:<https://doi.org/10.5565/rev/elcvia.1084>

- Priyankara, G. G. D. S., Karunarathne, C. L. S. M., Sarananda, K. H. y Ariyaratne, M. (2017). Effect of Maturity Stage on Ripening and Quality Characters of Four Tomato (*Solanum lycopersicum* L.) Varieties of Sri Lanka *Tropical Agricultural Research*, 28(4), 496-502.
- Rad, M. R. N., Koohkan, S., Fanaei, H. R. y Rad, M. R. P. (2015). Application of Artificial Neural Networks to predict the final fruit weight and random forest to select important variables in native population of melon (*Cucumis melo*. Pahlavan). *Scientia Horticulturae*, 181, 108-112. doi:<https://doi.org/10.1016/j.scienta.2014.10.025>
- Ren, S., He, K., Girshick, R. y Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137 - 1149. doi:<https://doi.org/10.1109/TPAMI.2016.2577031>
- Sa'ad, F. S. A., Ibrahim, M. F., Shakaff, A. Y. Md., Zakaria, A. y Abdullah, M. Z.; (2015). Shape and weight grading of mangoes using visible imaging. *Computers and Electronics in Agriculture*, 115, 51-56. doi:<http://dx.doi.org/10.1016/j.compag.2015.05.006>
- Salawu, E. O., Abdulraheem, M., Shoyombo, A., Adepeju, A., Davies, S., Akinsola, O. y Nwagu, B. (2014). Using Artificial Neural Network to Predict Body Weights of Rabbits. *Open Journal of Animal Sciences*, 4, 182-186. doi:<http://dx.doi.org/10.4236/ojas.2014.44023>
- SIAP, Servicio de Información Agroalimentaria y Pesquera (2017). Anuario Estadístico de la Producción Agrícola. Retrieved 5 de Marzo del 2019 <https://nube.siap.gob.mx/cierreagricola/>
- Taheri, A. G., Rafiee, S. y Alireza;, K. (2011). Study on some morphological and physical characteristics of tomato used in mass models to characterize best post harvesting options *Australian Journal of Crop Science*, 5(4), 433-438.
- Teoh, C. C. y Syaifudin, A. R. M. (2007). Image processing and analysis techniques for estimating weight of Chokanan mangoes. *Journal of Tropical Agriculture and Food Science*, 35(1), 183-190.
- Utai, K., Nagle, M., Hämmerle, S., Spreer, W., Mahayothee, B. y Müller, J. (2018). Mass estimation of mango fruits (*Mangifera indica* L., cv. 'Nam Dokmai') by linking image processing and artificial neural network. *Engineering in Agriculture, Environment and Food*.
- Wallach, D., Makowski, D., Jones, J. W. y Brun, F. (2014). *Working with Dynamic Crop Models*. (Elsevier Ed. Segunda Edición ed.). United States of America: Academic Press.

- Young, T. E., Juvik, J. A. y Sullivan, J. G. (1993). Accumulation of the Components of Total Solids in Ripening Fruits of Tomato. *Journal of the American Society for Horticultural Science*, 118(2). doi:<https://doi.org/10.21273/JASHS.118.2.286>
- Zollhöfer, M., Stotko, P., Görlitz, A., Theobalt, C., Nießner, M., Klein, R. y Kolb, A. (2018). State of the Art on 3D Reconstruction with RGB-D Cameras. *The Eurographics Association and John Wiley & Sons Ltd*, 37(2), 28. doi:<https://doi.org/10.1111/cgf.13386>
- Zuñiga-Avila, C. I., Fitz-Rodríguez<sup>1</sup>, E., Ramírez-Árias, J. A. y Salazar-Moreno, R. (in press). Detección y clasificación de frutos de jitomate de invernadero con visión artificial y redes neuronales de convolución. *Computers and Electronics in Agriculture*.