

Escuela Superior Politécnica del Litoral

Facultad de Ingeniería en Mecánica y Ciencias de la Producción

Detección inteligente de puntos de poda en viñedos mediante visión por
computador e integración embebida.

INGE-3012

Proyecto Integrador

Previo la obtención del Título de:

Ingenieros en Mecatrónica

Presentado por:

Diego de Jesús Albán Cedeño

Emily Teresa Torres Jara

Guayaquil - Ecuador

Año: 2025

Declaración Expresa

Nosotros, Emily Teresa Torres Jara y Diego de Jesús Albán Cedeño, acordamos y reconocemos que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá a los autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores.

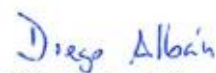
La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por nosotros durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique a los autores que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 20 de octubre del 2025.



Emily Teresa Torres Jara



Diego de Jesús Alban

Cedeño

Evaluadores

Jonathan Roberto León Torres, M.Sc.

Profesor de Materia

Marcelo Rodolfo Fajardo Pruna,

Ph.D.

Tutor de proyecto

Resumen

El presente trabajo desarrolla un sistema embebido de visión por computador para la detección automática del punto de poda en plantas de vid, con el objetivo de mejorar la precisión y eficiencia de los procesos de poda en la viticultura de precisión. La investigación plantea como hipótesis que la integración de técnicas de detección de objetos en 2D y percepción geométrica en 3D permite estimar de forma confiable la ubicación del corte agronómico. Para el desarrollo del proyecto se emplearon cámaras RGB-D Intel *RealSense*, plataformas NVIDIA Jetson Orin y modelos de detección YOLO11 entrenados con estrategias de aumentación de datos. El procesamiento de las imágenes permitió identificar yemas y segmentos de sarmiento, y posteriormente proyectarlos al espacio tridimensional para calcular el punto de corte óptimo. Los resultados obtenidos evidenciaron que el modelo YOLO11m alcanzó el mejor desempeño en precisión y mAP50, validando su viabilidad para la implementación final del sistema. En conclusión, el sistema propuesto constituye una solución tecnológica funcional para la automatización de la poda en vid y sienta las bases para su futura integración en robots agrícolas autónomos.

Palabras Clave: visión artificial, poda inteligente, modelos embebidos, viticultura de precisión.

Abstract

This work develops an embedded computer vision system for the automatic detection of pruning points in grapevine plants, aiming to improve the accuracy and efficiency of pruning processes in precision viticulture. The study hypothesizes that integrating 2D object detection techniques with 3D geometric perception enables reliable estimation of the agronomic cutting location. The project was developed using Intel RealSense RGB-D cameras, NVIDIA Jetson Orin platforms, and YOLO11 detection models trained with optimized data augmentation strategies. Image processing allowed the identification of buds and cane segments, which were then projected into three-dimensional space to compute the optimal pruning point. The obtained results showed that the YOLO11m model achieved the highest performance in terms of precision and mAP50, validating its suitability for the final system implementation. In conclusion, the proposed system represents a functional technological solution for automating grapevine pruning and establishes the foundation for its future integration into autonomous agricultural robots.

Keywords: *computer vision, smart pruning, embedded models, precision viticulture.*

Índice general

Resumen	I
Abstract.....	II
Índice general.....	III
Abreviaturas.....	VII
Simbología.....	VIII
Índice de figuras	IX
Índice de tablas	X
1. Capítulo 1.....	1
1.1 Introducción	2
1.2 Descripción del Problema	3
1.3 Justificación del Problema	4
1.4 Objetivos	5
1.4.1 Objetivo general.....	5
1.4.2 Objetivos específicos	5
1.5 Marco teórico	6
1.5.1 Fundamentos biológicos y morfológicos de la vid	6
1.5.2 Fundamentos de visión por computadora aplicados a la agricultura	12
1.5.3 Procesamiento de imágenes y técnicas clásicas de análisis	14

1.5.4	Fundamentos de sensado y arquitecturas de adquisición de datos	15
1.5.5	Estado del arte.....	17
2.	Capítulo 2.....	19
2.1	Metodología.	20
2.2	Análisis de requerimientos.....	20
2.3	Soluciones evaluadas	23
2.3.1	Alternativas de solución.....	23
2.3.2	Matriz de decisión.....	27
2.4	Metodología de diseño del sistema	29
2.5	Diseño del subsistema de programación	31
2.5.1	Detección de Instancias y Reconstrucción Topológica (Skeletonization)	33
2.5.2	Caracterización Geométrica y Cálculo del Índice de Ravaz.....	34
2.5.3	Identificación del Sistema y Módulos de Poda Específicos	34
2.5.4	Generación de Comandos y Salida	36
2.6	Diseño del subsistema electrónico	36
2.6.1	Derivación de requerimientos electrónicos a partir del pipeline de percepción ..	37
2.6.2	Evaluación de plataformas de procesamiento.....	38
2.6.3	Selección del sensor RGB-D y configuración geométrica	41
2.6.4	Arquitectura electrónica resultante	43

2.6.5	Diferenciación entre validación y despliegue industrial.....	43
3.	Capítulo 3.....	46
3.1	Entorno de experimentación	47
3.2	Implementación de modelos de percepción	49
3.2.1	Implementación del modelo YOLO11n para Jetson Orin Nano.....	49
3.2.2	Implementación del modelo YOLO11m para Jetson AGX Orin	51
3.2.3	Justificación del uso de dos arquitecturas de percepción.....	53
3.3	Evaluación del desempeño del modelo seleccionado	53
3.3.1	Curvas de aprendizaje y métricas de desempeño.....	54
3.3.2	Matriz de confusión	55
3.3.3	Análisis global del desempeño	56
3.4	Generación de percepción espacial (estimación de profundidad).....	56
3.4.1	Comparación visual entre imagen RGB y mapa de profundidad	57
3.4.2	Limitaciones de la estimación de profundidad monocular	58
3.5	Determinación del punto de poda y vector de corte.....	59
3.6	Integración en Sistema Embebido.....	64
3.6.1	Arquitectura electrónica del sistema.....	64
3.6.2	Integración en Plataformas Jetson	66
3.7	Análisis de costos del sistema	67

3.7.1	Costos de Hardware	67
3.7.2	Costos de integración y operación	68
3.7.3	Resumen General de Costos	69
4.	Capítulo 4.....	72
4.1	Conclusiones y recomendaciones	73
4.1.1	Conclusiones	73
4.1.2	Recomendaciones	74
	Referencias	77
	APÉNDICE.....	81

Abreviaturas

3D	Tridimensional
Baseline	Modelo base de referencia para comparación.
Batch Size	Cantidad de imágenes procesadas simultáneamente durante entrenamiento.
DFL Loss	<i>Distribution Focal Loss</i> , error en la regresión de bounding boxes.
Epoch	Iteración completa de entrenamiento sobre el dataset.
ESPOL	Escuela Superior Politécnica del Litoral.
FPS	<i>Frames per second</i> , velocidad de inferencia del modelo.
IoU	<i>Intersection over Union</i> , métrica de solapamiento entre detecciones y etiquetas.
mAP	<i>mean Average Precision</i> , precisión promedio de detección.
mAP50	mAP evaluado a un umbral de IoU = 0.50.
RGB	Rojo, Verde y Azul (Red, Green and Blue); modelo de representación del color.
RGB-D	Imagen RGB con información de profundidad.
UGV	<i>Unmanned Ground Vehicle</i> , vehículo terrestre no tripulado.
YOLO	<i>You Only Look Once</i> , arquitectura de detección de objetos.

Simbología

E	Energía consumida por el sistema durante la jornada de operación.
Hrs	Duración de la jornada laboral.
IoU	Umbral de intersección sobre unión utilizado en la evaluación del modelo.
Pc	Vector de pose del punto de corte.
W	Potencia de operación de la plataforma embebida.
x, y, z	Coordenadas espaciales reales del punto de poda.
α, β, γ	Ángulos de orientación de la herramienta de corte.

Índice de figuras

Figura 1.1 Morfología de la vid (1)	7
Figura 1.2 Morfología de la vid (2)	8
Figura 1.3 Formaciones Guyot	9
Figura 1.4 Formaciones Cordon Royat.....	10
Figura 1.5 Formación en vaso.....	10
Figura 2.1 Diagrama de flujo de Solución 1	24
Figura 2.2 Diagrama de flujo de Solución 2	25
Figura 2.3 Diagrama de flujo de Solución 3	26
Figura 2.4 Metodología de diseño	30
Figura 2.5 Diagrama de Flujo General	32
Figura 2.6 Módulo de Poda Royat (Detalle del Diagrama 2)	35
Figura 2.7 Metodología subsistema electrónico	37
Figura 2.8 Jetson Orin NX.....	40
Figura 2.9 Jetson AGX Orin Industrial.....	41
Figura 2.10 Arquitectura electrónica propuesta para el despliegue final del sistema.....	45
Figura 3.1 Inferencia del modelo YOLO11n sobre el dataset de validación.....	50
Figura 3.2 Inferencia del modelo YOLO11m sobre el dataset de validación.....	52
Figura 3.3 Resultados YOLO11m	54
Figura 3.4 Matriz de confusión normalizada YOLO11m.....	55
Figura 3.5 Comparación imagen RGB (izquierda) y mapa de profundidad estimado (derecha)	57

Figura 3.6 Detección de sarmientos.....	60
Figura 3.7 Visualización de puntos de poda	61
Figura 3.8 Ubicación del punto de poda	62
Figura 3.9 Limitaciones del modelo	63
Figura 3.10 Esquemático del sistema de alimentación	65

Índice de tablas

Tabla 2.1 Requerimientos del sistema	21
Tabla 2.2 Matriz de decisión.....	27
Tabla 2.3 Comparativa de plataformas de procesamiento	39
Tabla 3.1 Configuración de entrenamiento.....	48
Tabla 3.2 Costos de Hardware del sistema	67
Tabla 3.3 Costos de integración y operación	69
Tabla 3.4 Costos totales	70

CAPÍTULO 1

1.1 Introducción

La viticultura constituye una de las principales actividades agrícolas a nivel mundial por su impacto económico, social y cultural. La producción de uva para vino y otros derivados depende en gran medida de un manejo preciso de la planta de vid, donde la poda es la labor más crítica. Esta tarea es determinante para garantizar la calidad de la cosecha, la longevidad de la cepa y la eficiencia global del cultivo. Sin embargo, en la actualidad, la mayoría de las operaciones de poda se realizan de manera manual, un proceso que demanda personal altamente especializado conlleva altos costos laborales y limita significativamente la escalabilidad del cultivo en viñedos de gran extensión [1].

La necesidad de optimizar recursos ha impulsado la agricultura de precisión, que busca la automatización de procesos mediante la inteligencia artificial y el análisis de datos. En este contexto, la visión por computador y el Aprendizaje profundo se han consolidado como herramientas clave, demostrando gran efectividad en tareas agrícolas como la identificación de frutos y la detección de plagas [2]. A pesar de estos avances, la aplicación de estas tecnologías en la viticultura, específicamente en el reconocimiento estructural de la vid y la detección precisa de puntos de poda, sigue siendo un desafío técnico debido a la complejidad morfológica tridimensional de las cepas y a las condiciones variables de iluminación y fondo del entorno agrícola [3].

1.2 Descripción del Problema

El problema central que aborda este proyecto reside en la baja eficiencia y la alta dependencia de mano de obra especializada que caracteriza la poda en la viticultura moderna. Esta actividad es crucial, ya que influye directamente en la calidad y producción de la uva; por ello, su correcta ejecución resulta indispensable para garantizar la salud y longevidad del viñedo. Sin embargo, al ser una labor predominantemente manual, genera altos costos operativos y una marcada variabilidad en los resultados, lo que impacta negativamente la competitividad del sector agrícola [1].

La dificultad para automatizar la poda no radica en la mecánica del proceso, sino en la capacidad de detección e interpretación necesarias para replicar el criterio agronómico experto al seleccionar los puntos de corte. El sistema debe ser capaz de identificar las zonas de poda óptimas siguiendo las directrices de formación de la vid, particularmente en los sistemas de conducción Guyot y Cordon Royat [4]. Esto implica distinguir con precisión entre el cordón permanente, los sarmientos y las yemas productivas, para determinar qué elemento cortar y así mantener el equilibrio entre el vigor y la carga frutal de la planta.

Diversas soluciones en desarrollo han intentado abordar esta tarea mediante algoritmos de análisis estructural y reconstrucción tridimensional de la vid, con el fin de interpretar su morfología y determinar zonas potenciales de corte. La estructura de la planta presenta un entramado tridimensional con múltiples oclusiones y ramificaciones variables entre cepas, mientras que las condiciones ambientales cambiantes —como iluminación irregular, sombras o presencia de residuos— dificultan la adquisición y el procesamiento visual en tiempo real [5].

Por tanto, resulta indispensable desarrollar un sistema capaz de estimar de manera precisa la estructura 3D de la vid para localizar los puntos de corte con la exactitud requerida para una herramienta de poda automatizada. Además, para que la solución sea viable en campo y pueda integrarse en plataformas robóticas, el modelo de procesamiento debe ser computacionalmente eficiente, capaz de operar a alta velocidad y tomar decisiones casi instantáneas. Esto impone una restricción técnica rigurosa sobre el diseño electrónico y la arquitectura del sistema de control.

La problemática afecta directamente al avance de la robótica agrícola, un sector que crece exponencialmente, pero que aún enfrenta un cuello de botella en la automatización de la poda por la ausencia de un “cerebro” preciso y rápido que guíe la acción de los sistemas podadores [6].

1.3 Justificación del Problema

Durante el invierno la vid guarda su energía y reduce su flujo de sabia para esperar la primavera y formar los sarmientos que darán origen a los frutos. La poda invernal constituye una práctica esencial para controlar el vigor, la productividad y la calidad de los frutos en la vid. Actualmente la mano de obra especializada en esta labor ha disminuido [7], esto y la incorporación de nuevas tecnologías en la agricultura, principalmente en la viticultura, han generado un triple impacto: económico-social, científico-tecnológico y académico-formativo.

La poda se ha consolidado como una de las labores agrícolas más costosas y demandantes de tiempo en la viticultura, representando un factor limitante en la escalabilidad y competitividad en la producción [1]. La justificación económica del proyecto radica en la necesidad imperante de reducir los altos costos operativos asociados a la dependencia de mano de obra especializada, la cual es estacional y progresivamente más escasa [7].

Actualmente en el mercado existen múltiples herramientas o sistemas tecnológicos aplicados en la agricultura, que buscan automatizar, reducir costos de producción, no depender completamente de personal especializado, entre otros [19]. Una de las dificultades que se han presentado y que afectan principalmente a estas soluciones, aplicadas en el sector vitícola, está la complejidad de la estructura morfológica de la vid debido a la variabilidad estructural de la planta, las oclusiones entre sarmientos y las condiciones cambiantes de iluminación en campo [5]. Los sistemas existentes son a menudo lentos o inexactos, lo que convierte la detección en el cuello de botella de la automatización robótica de la poda [6].

Estos desafíos requieren un análisis profundo de procesos de percepción visual y modelado tridimensional de la vid, ofreciendo un amplio aprendizaje multidisciplinario de integración electrónica, visión por computadora, sistemas de control [15].

1.4 Objetivos

1.4.1 Objetivo general

Desarrollar un sistema inteligente de detección de puntos de poda en viñedos mediante visión por computador y diseño electrónico embebido, con el fin de optimizar la identificación y toma de decisiones en sistemas de poda automatizada.

1.4.2 Objetivos específicos

1. Analizar la base de datos de imágenes y nubes de puntos de cepas de vid, identificando las características morfológicas relevantes para la detección de puntos de poda.

2. Desarrollar modelos de visión por computador basados en aprendizaje profundo para segmentar, clasificar y determinar los puntos óptimos de corte en cepas formadas en Guyot y Cordón Royat.
3. Diseñar la arquitectura electrónica embebida del sistema, considerando la selección de componentes, sensores, cámara y procesador para el procesamiento local de la información visual.
4. Evaluar el desempeño del modelo de detección y la arquitectura propuesta mediante análisis comparativo y simulaciones conceptuales.

1.5 Marco teórico

En esta sección se presentan los fundamentos conceptuales y técnicos que sustentan el desarrollo del sistema.

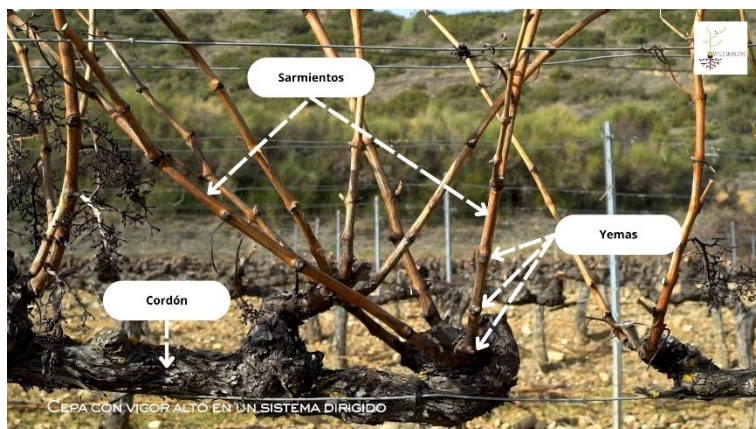
1.5.1 Fundamentos biológicos y morfológicos de la vid

La vid (*Vitis vinifera* L.) es una planta leñosa y trepadora de crecimiento perenne, cuya morfología y fisiología determinan directamente la productividad del viñedo. Comprender la estructura de la planta resulta fundamental para automatizar la poda, dado que el proceso depende de reconocer correctamente los elementos estructurales y las zonas de corte que definen la renovación anual de la vegetación productiva.

1.5.1.1 Partes anatómicas de la cepa y su función

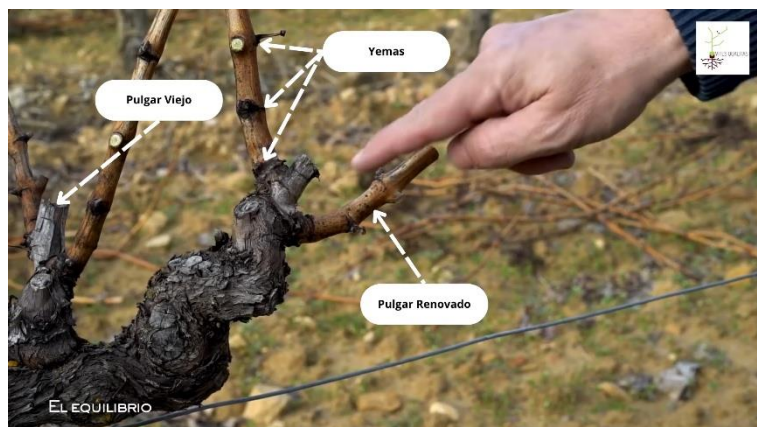
La morfología de la vid, Figura 1.1 y Figura 1.2, condiciona directamente las decisiones de poda; por ello cada estructura debe identificarse y entenderse funcionalmente:

-

Figura 1.1*Morfología de la vid (1)*

Nota. Fotograma extraído [4] y elementos morfológicos etiquetados.

- Tronco. Eje leñoso principal que conecta raíces y parte aérea; tiene como funciones el transporte de agua y savia, y el soporte estructural del sistema de conducción. La integridad del tronco es crítica para la longevidad de la cepa y cualquier daño o infección en esta zona puede comprometer la planta completa [4].
- Cordón. Rama derivada del tronco que actúa como “columna vertebral” en sistemas como el Cordón Royat; distribuye el flujo de savia lateralmente. Su conservación y forma (rectitud, continuidad) influyen en la uniformidad productiva y en la facilidad de mecanización [4], [9].
- Sarmiento. Brote anual que, tras lignificar, porta yemas y potencial productivo. Los sarmientos del año contienen las yemas que producirán los racimos la siguiente temporada; su vigor y longitud condicionan la cantidad y calidad de producción. Visualmente, son elementos temporales que el sistema de visión debe distinguir de las partes permanentes.

Figura 1.2*Morfología de la vid (2)*

Nota. Fotograma extraído [4] y elementos morfológicos etiquetados.

- Yema. Órgano meristemático donde se originan brotes y racimos; existen yemas productivas y yemas menos fértiles. La selección de yemas correctas (localización, orientación, salud) es el criterio central en la decisión de corte.
- Pulgar. Fragmento corto de sarmiento (1 yema ciega y 2 yemas francas) que se deja intencionadamente durante la poda como punto de renovación para la siguiente temporada. Su distribución y número por cordón definen la carga de brotes al año siguiente.
- Chupón. Brote vigoroso, habitualmente indeseado, que nace del tronco o de la base del cordón. Consume recursos y altera el equilibrio; su identificación y eliminación es parte del manejo correcto [4].

(Estos conceptos y definiciones se basan en la Guía de Buenas Prácticas de Poda y en materiales técnicos de formación vitícola, los cuales orientan tanto la práctica como la codificación de reglas para sistemas automáticos) [4], [9].

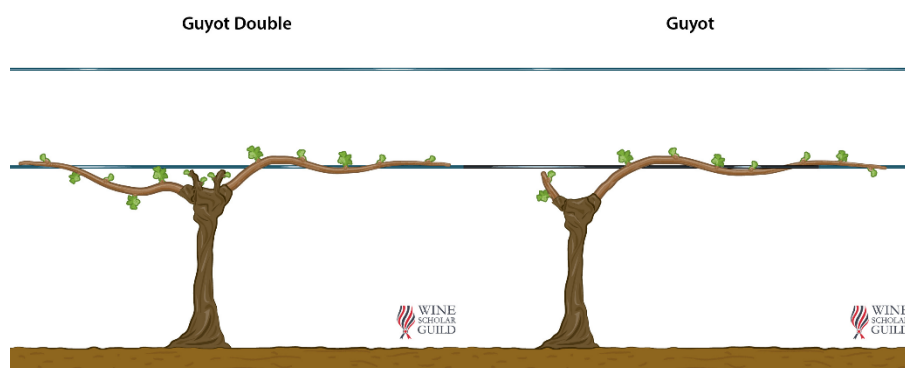
1.5.1.2 Sistemas de formación: diferencias y beneficios

La arquitectura del viñedo varía según el sistema de formación, la geometría condiciona como deberá interpretar el sistema de visión.

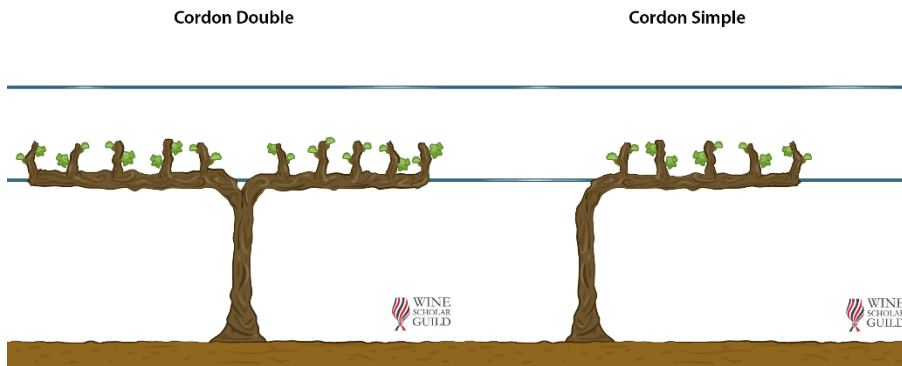
Guyot (simple/doble). Se conserva uno o dos varas largas por planta que se renuevan de sarmientos cada año; es flexible y permite una buena ventilación del follaje como se muestra en Figura 1.3. Beneficios: adaptación a variedades de vigor variable y mayor control manual sobre la renovación. Desventajas: requiere poda más detallada y mano de obra experta. En visión, el Guyot presenta variabilidad en orientación y longitud de sarmientos que complica la segmentación automática [9].

Figura 1.3

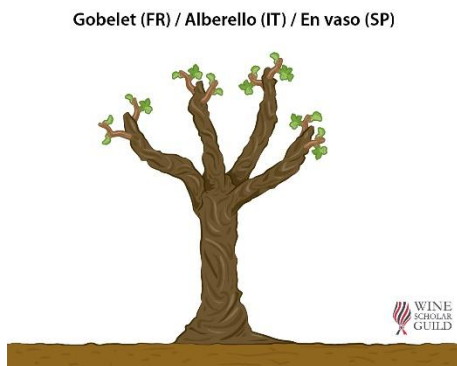
Formaciones Guyot [9]



Cordón Royat. Mantiene uno o dos cordones permanentes con pulgares distribuidos a lo largo, Figura 1.4. Beneficios: geometría regular que facilita la mecanización y la detección sistemática de pulgares/yemas; reduce variabilidad entre plantas y favorece operaciones repetitivas (ideal para soluciones embebidas y robótica). Desventajas: si no se maneja el vigor, puede generar sobrecarga. Desde la perspectiva de visión por computador, el Cordón Royat es ventajoso por su estructura más predecible [10], [11].

Figura 1.4*Formaciones Cordon Royat [9]*

Sistema en vaso. Forma arbustiva, sin tutores, como se ve en Figura 1.5; se usa en zonas áridas o con bajo mecanizado. Beneficio: resistencia a condiciones secas; inconveniente: geometría irregular que dificulta la detección automática [11].

Figura 1.5*Formación en vaso [9]*

La comparación muestra que el Cordón Royat suele ser el más compatible con la automatización (por regularidad geométrica), mientras que el Guyot y el vaso exigen mayor robustez algorítmica por su variabilidad [9], [10], [11].

1.5.1.3 Criterios de poda

La poda de invierno controla el vigor vegetativo y producción de la vid favoreciendo la formación de racimos equilibrados y de calidad de la uva [12]. Esta se basa en la eliminación selectiva de madera no productiva, la conservación de yemas fructífera y el establecimiento óptimo de pulgares para el próximo ciclo vegetativo [13].

Los criterios técnicos de la poda definen qué debe decidir el algoritmo; presento los criterios prácticos que deben traducirse a reglas computacionales:

- Identificación de yemas viables. Conservar yemas sanas, orientadas favorablemente y de tamaño adecuado; descartar yemas dañadas, ciegas o mal orientadas. En datos visuales esto se traduce en detección de forma, color y posición relativa del nudo. [4]
- Número y longitud del pulgar. Mantener de 2 yemas francas por pulgar en sistemas de cordón; en Guyot la longitud del sarmiento y el número de yemas definen la carga. Para el modelo, esto exige conteo y medición relativa de la distancia entre yemas y del diámetro del sarmiento. [4], [10]
- Distancia y orientación del corte. Cortar entre 1.5 a 2 veces el diámetro del sarmiento por encima de la yema, con ángulo y orientación que minimicen el cono de desecación y respeten la dirección del flujo de savia. Esto implica que la detección debe devolver no sólo la coordenada del punto, sino también la orientación local de la rama (vector normal) para definir la alineación del efector. [4]

- Preservación del cordón. Evitar cortes que dañen el cordón permanente: el sistema debe distinguir tejido cordonal (más ancho, continuo) de sarmientos lignificados o chupones. En visión esto requiere segmentación semántica y clasificación de tipo de rama [4], [11].
- Eliminación de chupones y material no deseado. Detectar brotes vigorosos próximos al tronco para marcarlos para eliminación en caso de que el vigor de la vid sobrepase los 12mm. Este criterio implica reglas basadas en proximidad al tronco y vigor relativo (diámetro, color, ángulo) [4].
- Equilibrio vegetativo (regla agronómica). Mantener carga de yemas acorde al vigor estimado: si el sarmiento muestra un diámetro fuera del rango (7mm a 12mm en la base del sarmiento) o hay sobreabundancia de brotes, equilibrar la carga. Para el sistema esto exige estimación del vigor (*proxy* visual: diámetro, densidad de brotes) y reglas de decisión parametrizables [4], [9].

La interacción entre el conocimiento agronómico y las tecnologías de análisis visual constituye la base conceptual sobre la que se sustentan los actuales estudios de segmentación y detección estructural en viticultura.

1.5.2 Fundamentos de visión por computadora aplicados a la agricultura

La visión por computador es una rama de la inteligencia artificial que permite a los sistemas interpretar y analizar información visual proveniente de cámaras o sensores. De acuerdo con González y Woods [14], el flujo básico de un sistema de visión por computador comprende varias etapas: adquisición, preprocesamiento, segmentación, extracción de características y clasificación.

En la adquisición, se capturan imágenes mediante cámaras RGB o sensores de profundidad, cuya resolución y ángulo de visión influyen directamente en la precisión del análisis. El preprocesamiento busca mejorar la calidad de la imagen corrigiendo el contraste, el ruido o los efectos de iluminación, factores que en entornos agrícolas pueden variar significativamente.

Posteriormente, la segmentación divide la imagen en regiones homogéneas con el fin de aislar los objetos de interés —como hojas, frutos o ramas— y facilitar su análisis individual. En esta etapa se aplican técnicas clásicas de umbralización, detección de bordes o modelos más avanzados basados en aprendizaje automático [14] [15]. La extracción de características consiste en representar cada región mediante descriptores geométricos, texturales o de color, los cuales sirven como entrada para algoritmos de clasificación o reconocimiento.

En aplicaciones robóticas agrícolas, esta interpretación visual debe extenderse hacia una representación geométrica del entorno. No basta con clasificar píxeles o regiones, sino que es necesario asociar cada estructura detectada con una ubicación física en el espacio, de modo que las decisiones del sistema puedan convertirse en acciones mecánicas sobre la planta. En el caso de la poda, esto implica que la detección visual de yemas o sarmientos debe estar vinculada a coordenadas métricas que permitan posicionar un actuador de corte con precisión.

En el contexto agrícola, la visión por computador ha demostrado un alto potencial para la automatización de procesos, gracias a su capacidad para analizar grandes volúmenes de datos visuales con precisión y rapidez. Kamilaris y Prenafeta-Boldú [15] destacan que su integración con el aprendizaje profundo ha permitido superar las limitaciones de los métodos tradicionales, al mejorar la tolerancia a la variabilidad de luz, textura y forma de los cultivos.

La literatura reciente muestra que tecnologías de visión por computador y *deep learning* ya se aplican en diversas tareas de agricultura inteligente, incluyendo detección de cultivos, identificación de plagas, conteo de objetos agrícolas y estimación de estados fenológicos, contribuyendo directamente a la agricultura de precisión [19]. En el caso particular de la viticultura, la visión por computador constituye la base tecnológica para el análisis estructural de la planta y la identificación de sus componentes. La capacidad de segmentar la cepa, reconocer sus partes anatómicas y determinar zonas de interés posibilita la automatización de operaciones críticas como la poda. Estas aplicaciones sientan las bases para el desarrollo de sistemas más avanzados de detección y decisión autónoma, que serán abordados en la sección de estado del arte.

1.5.3 Procesamiento de imágenes y técnicas clásicas de análisis

El procesamiento de imágenes constituye la base de los sistemas de visión por computadora, al permitir la extracción, análisis e interpretación de información visual procedente de sensores o cámaras. Su propósito fundamental es transformar los datos visuales en representaciones que faciliten la comprensión automática del entorno por parte de un sistema computacional [14].

En el contexto agrícola, las técnicas de procesamiento digital de imágenes se aplican para detectar, segmentar y clasificar estructuras biológicas como hojas, frutos, tallos o ramas, con el fin de evaluar el estado fisiológico de las plantas y apoyar la toma de decisiones agronómicas [15]. Entre los métodos más empleados se incluyen la conversión de color, filtrado espacial, detección de bordes, segmentación por umbrales y análisis morfológico, los cuales permiten resaltar características relevantes para el análisis posterior.

Estos procedimientos clásicos constituyen la primera capa de interpretación antes del uso de algoritmos más avanzados. Por ejemplo, la detección de bordes mediante operadores como Sobel o Canny se emplea para identificar los límites entre estructuras de interés, mientras que las transformaciones de color en espacios como HSV o Lab facilitan la separación de objetos del fondo bajo diferentes condiciones de iluminación [14].

En plantas leñosas como la vid, la morfología irregular de los sarmientos y la disposición variable de las yemas generan estructuras altamente ramificadas que no pueden describirse adecuadamente mediante modelos geométricos simples. Por esta razón, los procesos de análisis de imagen deben orientarse no solo a detectar regiones, sino también a preservar la continuidad y conectividad de las estructuras vegetales, las cuales son fundamentales para inferir decisiones de poda.

Aunque las técnicas tradicionales presentan limitaciones ante entornos con variabilidad morfológica o condiciones lumínicas cambiantes, siguen siendo indispensables para la preparación y limpieza de datos visuales, así como para reducir el ruido y destacar rasgos estructurales. En los sistemas agrícolas modernos, estas técnicas suelen combinarse con modelos estadísticos o geométricos para mejorar la detección de elementos como los sarmientos, yemas o cordones de la vid, que poseen una morfología irregular y de difícil segmentación [17].

1.5.4 Fundamentos de sensado y arquitecturas de adquisición de datos

La etapa de adquisición constituye el punto de enlace entre el entorno físico y los sistemas de análisis visual, siendo responsable de capturar la información óptica y estructural que alimenta los procesos de interpretación computacional. En la agricultura, los sistemas de sensado se diseñan

para operar en entornos abiertos, donde factores como la variación lumínica, la presencia de polvo o la geometría irregular de las plantas imponen desafíos adicionales.

Entre los principales dispositivos de adquisición en agricultura de precisión se incluyen cámaras RGB y sensores que proporcionan información de profundidad o reconstrucción espacial, utilizados para mapear estructuras vegetales, estimar geometrías y facilitar tareas de percepción automática [20]. Cada uno de estos sistemas proporciona un tipo de información distinto: las cámaras RGB ofrecen datos cromáticos de alta resolución, mientras que los sensores de profundidad permiten obtener la estructura tridimensional del entorno, facilitando la estimación de volúmenes o la localización espacial de los objetos.

Para la integración de sistemas de sensado en entornos agrícolas es necesario definir arquitecturas de adquisición que garanticen calibración geométrica, sincronización temporal y reducción de latencia, de modo que las imágenes captadas puedan procesarse con eficiencia y robustez incluso en tiempo real [16]. La geometría de la disposición de los sensores y la estabilidad de la captura son factores críticos, especialmente en sistemas móviles o robotizados que deben mantener un marco de referencia constante respecto a la planta.

En operaciones como la poda invernal de la vid, realizadas en exteriores y bajo iluminación natural variable, la obtención de información tridimensional resulta esencial para localizar con precisión las estructuras leñosas expuestas. La combinación de datos RGB con información de profundidad permite compensar las limitaciones de la visión bidimensional y establecer una referencia métrica del entorno, requisito indispensable para la posterior ejecución de acciones de corte automatizadas.

El diseño de estas arquitecturas requiere además considerar aspectos como eficiencia energética, robustez ante condiciones ambientales y capacidad de comunicación, parámetros que influyen en la futura implementación electrónica del sistema. En la práctica, las configuraciones híbridas —que combinan cámaras RGB con sensores de profundidad o infrarrojo— han demostrado una mayor capacidad para captar la estructura tridimensional de la vid y mejorar la precisión en la localización de elementos como las yemas o los sarmientos [21].

1.5.5 Estado del arte

El desarrollo de tecnologías aplicadas a la viticultura de precisión ha evolucionado significativamente en la última década, impulsado por los avances en visión por computador, robótica agrícola y sensores inteligentes. Estas herramientas transforman prácticas agrícolas tradicionales mediante técnicas de reconocimiento visual y análisis de datos que automatizan tareas operativas de alto impacto como la poda, la cosecha y el monitoreo fisiológico de los cultivos [19].

Los primeros estudios en automatización de la poda se centraron en el procesamiento bidimensional de imágenes, utilizando técnicas clásicas de segmentación y detección de bordes para identificar las ramas principales y los puntos de corte potenciales [1]. Sin embargo, la complejidad morfológica de la vid y la superposición de sarmientos limitaron la precisión de estos métodos, dando lugar a la incorporación de modelos tridimensionales y enfoques de reconstrucción estructural basados en imágenes RGB-D [3].

Posteriormente, la investigación se orientó hacia la estimación de la estructura 3D de la planta mediante el uso de sensores de profundidad y fotogrametría, permitiendo representar con mayor fidelidad la geometría del cordón, los pulgares y las yemas [5]. En esta línea, la

disponibilidad de conjuntos de datos etiquetados, como el publicado por 3D2Cut SA (2021), ha favorecido el entrenamiento y validación de algoritmos de segmentación estructural de vid bajo diferentes condiciones lumínicas y morfológicas [21].

Estos conjuntos de datos no solo permiten evaluar la capacidad de los modelos para segmentar estructuras de la vid, sino también analizar la coherencia geométrica de las detecciones cuando se combinan con información de profundidad, aspecto clave para la determinación fiable de puntos de corte en sistemas robotizados.

Recientemente, diversos trabajos han explorado la integración de sistemas de percepción avanzados con plataformas robóticas para asistir o automatizar labores específicas en agricultura, demostrando la viabilidad de enfoques basados en visión por computador y redes neuronales para estimar estructuras tridimensionales y localizar regiones de interés en plantas [5] [24]. Asimismo, estudios más recientes, como el de Oliveira et al. (2024), han empleado arquitecturas optimizadas para la detección de nudos y yemas mediante el uso de modelos de detección de objetos en imágenes 2D [18].

Estos desarrollos reflejan el creciente interés en la automatización de la poda de vid, pero también evidencian los desafíos persistentes asociados al procesamiento en tiempo real, la adaptabilidad a distintas estructuras de formación (Guyot, Royat) y la fiabilidad bajo condiciones de campo. En este sentido, la presente investigación se enmarca dentro de esta línea tecnológica, buscando contribuir al desarrollo de un sistema inteligente capaz de detectar y localizar con precisión los puntos de poda en cepas formadas bajo sistemas tradicionales europeos.

CAPÍTULO 2

2.1 Metodología.

Esta sección describe el proceso metodológico seguido para diseñar, implementar y validar un sistema de detección de puntos de poda basado en visión por computadora. A diferencia de enfoques puramente teóricos, la metodología aplicada se centró en la construcción de un pipeline funcional, compatible con los datos disponibles y alineado con los criterios agronómicos empleados en la práctica de poda.

2.2 Análisis de requerimientos

El desarrollo del sistema de detección de puntos de poda requiere identificar los factores técnicos y agronómicos que condicionan la selección de la metodología. Estos requerimientos fueron definidos considerando:

1. Las características estructurales de la vid en reposo invernal.
2. Las limitaciones del dataset disponible.
3. La necesidad de obtener salidas interpretables por especialistas.
4. La viabilidad computacional para su futura integración en sistemas embebidos o prototipos de automatización.

Con el fin de establecer una base objetiva para comparar las alternativas de solución, se formuló un conjunto de criterios ponderados que capturan los aspectos más relevantes del problema: robustez ante oclusiones, nivel de interpretabilidad, compatibilidad con datos RGB-D, capacidad de generalización y exigencias computacionales. Estos criterios fueron organizados en una escala del 1 al 5 según su importancia relativa dentro del proyecto.

La Tabla 2.1 resume los requerimientos establecidos y los pesos asignados, los cuales se emplearon posteriormente en la matriz de decisión para la selección de la solución final.

Tabla 2.1

Requerimientos del sistema

ID	Requisito (Nombre técnico)	Métrica / Variable de control	Descripción cuantitativa (Criterio de aceptación)
RF-01	Robustez ante oclusiones	<i>Recall</i> (Sensibilidad)	El sistema debe mantener un <i>Recall</i> superior al 85 % en la detección de puntos de poda, incluso cuando los sarmientos presenten hasta un 30 % de oclusión visual (hojas u otras ramas).
RF-02	Compatibilidad de entrada RGB-D	de Resolución / Formato	El algoritmo debe procesar nativamente imágenes RGB y mapas de profundidad con una resolución mínima de 1280×720 px, aceptando formatos estándar (JPG, PNG, .bag) sin conversión externa.
RF-03	Precisión de localización (interpretabilidad geométrica)	Error RMSE en coordenadas	El sistema debe entregar las coordenadas del punto de corte ((x,y)) con un RMSE < 5 mm respecto al <i>ground truth</i> validado por expertos agrónomos.

RNF-01	Eficiencia computacional (latencia)	FPS / tiempo de inferencia	El modelo debe ejecutarse en la NVIDIA Jetson Orin Nano a una velocidad mínima de 20 FPS (<i>tiempo de inferencia < 50 ms</i>) para permitir operación en tiempo casi real.
RNF-02	Consumo de recursos	Uso de VRAM / Potencia	El despliegue en contenedor Docker no debe exceder 4 GB de VRAM y debe operar bajo el modo de 15 W (MAXN) de la plataforma embebida.
RF-04	Capacidad de generalización	de mAP@50	El modelo debe alcanzar un mAP@50 > 0.80 al evaluarse sobre un conjunto de prueba que incluya al menos dos sistemas de conducción (por ejemplo, Guyot y Royat).
RF-05	Interoperabilidad robótica (escalabilidad)	Protocolo de comunicación	El sistema debe publicar las detecciones (bounding boxes y puntos de corte) mediante un tópico ROS 2 con una frecuencia estable de 10 Hz.
RF-06	Independencia de profundidad densa	de Tasa de fallo sin canal Depth	El sistema debe mantener una precisión de detección > 70 % en 2D aun cuando el canal de profundidad no esté disponible temporalmente (fallo del sensor 3D).

2.3 Soluciones evaluadas

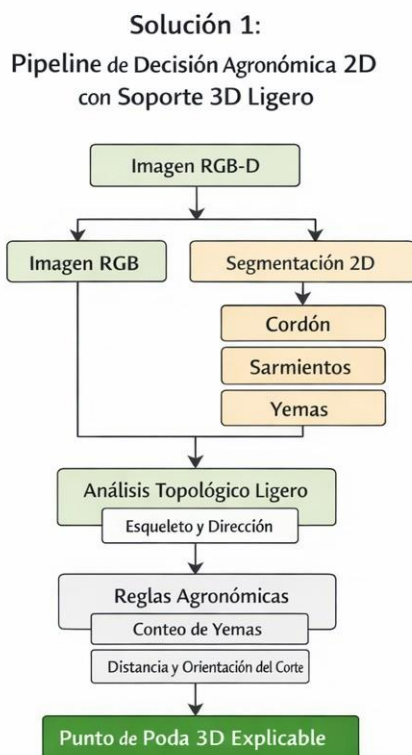
2.3.1 Alternativas de solución

Con base en los requerimientos establecidos en la sección 2.2, se identificaron tres soluciones metodológicas viables para abordar el problema de detección de puntos de poda mediante visión por computadora y datos RGB-D. Estas soluciones representan distintos niveles de complejidad geométrica, dependencia de profundidad e interpretabilidad agronómica, permitiendo un análisis comparativo equilibrado acorde con los objetivos del proyecto.

- **Solución 1: Pipeline híbrido 2D + análisis geométrico 3D ligero (RGB-D asistido)**

La solución se basa en realizar la detección de estructuras de la vid principalmente en imágenes RGB mediante un modelo 2D (por ejemplo, YOLO o un modelo de segmentación), mientras que el canal de profundidad se utiliza únicamente como apoyo geométrico. Las detecciones 2D de sarmientos y yemas se proyectan al espacio 3D usando el mapa de profundidad, permitiendo estimar la orientación y la posición real de cada rama sin reconstruir toda la escena. A partir de estos datos se genera una representación geométrica simplificada que permite calcular distancias y ángulos necesarios para definir el punto de poda mediante reglas agronómicas.

Esta alternativa ofrece un equilibrio entre precisión geométrica, robustez ante oclusiones y eficiencia computacional, siendo compatible con el dataset RGB-D disponible y con plataformas embebidas.

Figura 2.1*Diagrama de flujo de Solución 1*

- **Solución 2: Reconstrucción parcial RGB-D (3D moderado)**

En esta alternativa, la detección 2D se utiliza como punto de partida, pero la información de profundidad se emplea para generar nubes de puntos locales de cada sarmiento. Estas nubes permiten reconstruir con mayor detalle la geometría tridimensional de las ramas.

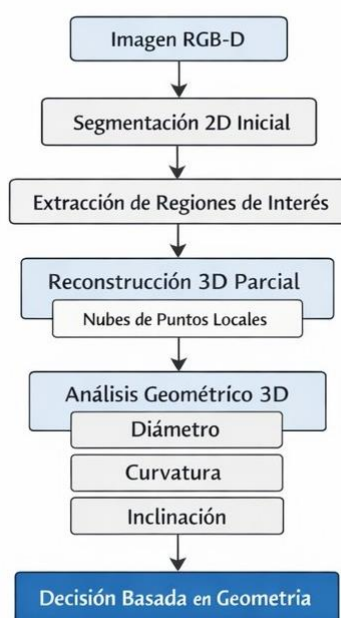
Sobre dichas nubes se aplican métodos de agrupamiento y ajuste geométrico para separar sarmientos cruzados, estimar diámetros y caracterizar la estructura de la planta con mayor precisión. Esta solución proporciona una representación más rica

que la alternativa híbrida ligera, pero incrementa el costo computacional, la complejidad del sistema y la sensibilidad a errores del sensor de profundidad.

Figura 2.2

Diagrama de flujo de Solución 2

Solución 2:
Modelado Geométrico Local de Sarmientos



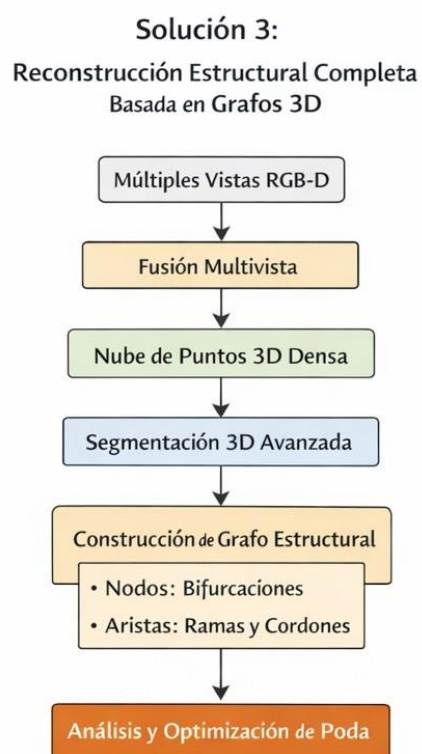
- **Solución 3: Reconstrucción 3D avanzada (*full geometry*)**

Plantea un enfoque de reconstrucción tridimensional avanzada, orientado a representar la arquitectura completa de la vid mediante nubes de puntos densas y modelos basados en grafos estructurales. Este método incorpora fusión de multivista o integración temporal de datos RGB-D para generar representaciones detalladas en 3D, seguidas de algoritmos especializados de segmentación tridimensional como *PointNet++* o *KPCConv*. A partir de ello se construyen grafos

donde los nodos corresponden a bifurcaciones y las aristas representan continuidad morfológica, permitiendo modelar explícitamente longitudes, diámetros, inclinaciones y conectividad estructural. Si bien esta solución ofrece la mayor precisión teórica y un modelado extremadamente completo, también implica una dependencia total del hardware RGB-D, una carga computacional muy elevada y una complejidad que excede el alcance típico de un proyecto de pregrado.

Figura 2.3

Diagrama de flujo de Solución 3



Para determinar la solución más adecuada, estas alternativas fueron comparadas utilizando los criterios y pesos establecidos previamente. La siguiente sección presenta la matriz de decisión y la selección final.

2.3.2 Matriz de decisión

La Tabla 2.2 resume la evaluación comparativa de las tres alternativas, utilizando los criterios definidos en la sección 2.2. Cada criterio se ponderó según su relevancia, y a cada alternativa se le asignó una puntuación del 1 al 5 según su capacidad para cumplir con criterios de precisión agronómica, robustez ante oclusiones, eficiencia computacional, dependencia del sensor de profundidad, interpretabilidad, escalabilidad robótica y compatibilidad con el dataset disponible.

Tabla 2.2

Matriz de decisión

Criterio técnico (derivado de RF/RNF)	Descripción	S1 (2D+3D ligero)	S2 (3D moderado)	S3 (3D completo)
	Capacidad de ubicar			
Precisión agronómica	correctamente el punto de poda (RF-03)	4	5	5
Robustez oclusiones	ante Ramas cruzadas, solapamiento (RF-01)	4	5	4
Eficiencia computacional	FPS y viabilidad en Jetson Orin Nano (RNF-01)	5	3	1
Dependencia del sensor de profundidad	Capacidad de operar si el depth falla (RF-06)	5	3	1

Criterio técnico (derivado de RF/RNF)	Descripción	S1 (2D+3D ligero)	S2 (3D moderado)	S3 (3D completo)
Interpretabilidad agronómica	Facilidad para aplicar reglas de poda	5	3	2
Complejidad de implementación	Riesgo técnico y carga de desarrollo	5	3	1
Escalabilidad robótica	Integración con ROS y sistemas reales (RF-05)	5	4	3
Compatibilidad dataset 3D2Cut	Ajuste a datos reales disponibles	5	4	3
TOTAL		38	30	20

Los resultados muestran que la Solución 1 (*pipeline* híbrido 2D + análisis geométrico 3D ligero) obtiene la puntuación más alta, evidenciando el mejor equilibrio entre desempeño y viabilidad de implementación. Esta alternativa permite estimar puntos de poda con precisión suficiente, mantiene robustez frente a oclusiones mediante el uso selectivo de la profundidad y puede ejecutarse en tiempo real sobre una plataforma embebida. Además, su baja dependencia de una reconstrucción 3D densa facilita la aplicación de reglas agronómicas interpretables y su integración con sistemas robóticos.

La Solución 2 (reconstrucción RGB-D parcial) ofrece mayor fidelidad geométrica, pero incrementa la complejidad computacional y la dependencia del sensor de profundidad. Por su parte, la Solución 3 (reconstrucción 3D avanzada), aunque es la más completa desde el punto de vista teórico, presenta una complejidad y requerimientos de hardware que exceden el alcance de esta investigación.

En consecuencia, la Solución 1 se selecciona como la alternativa más adecuada para el desarrollo del sistema propuesto.

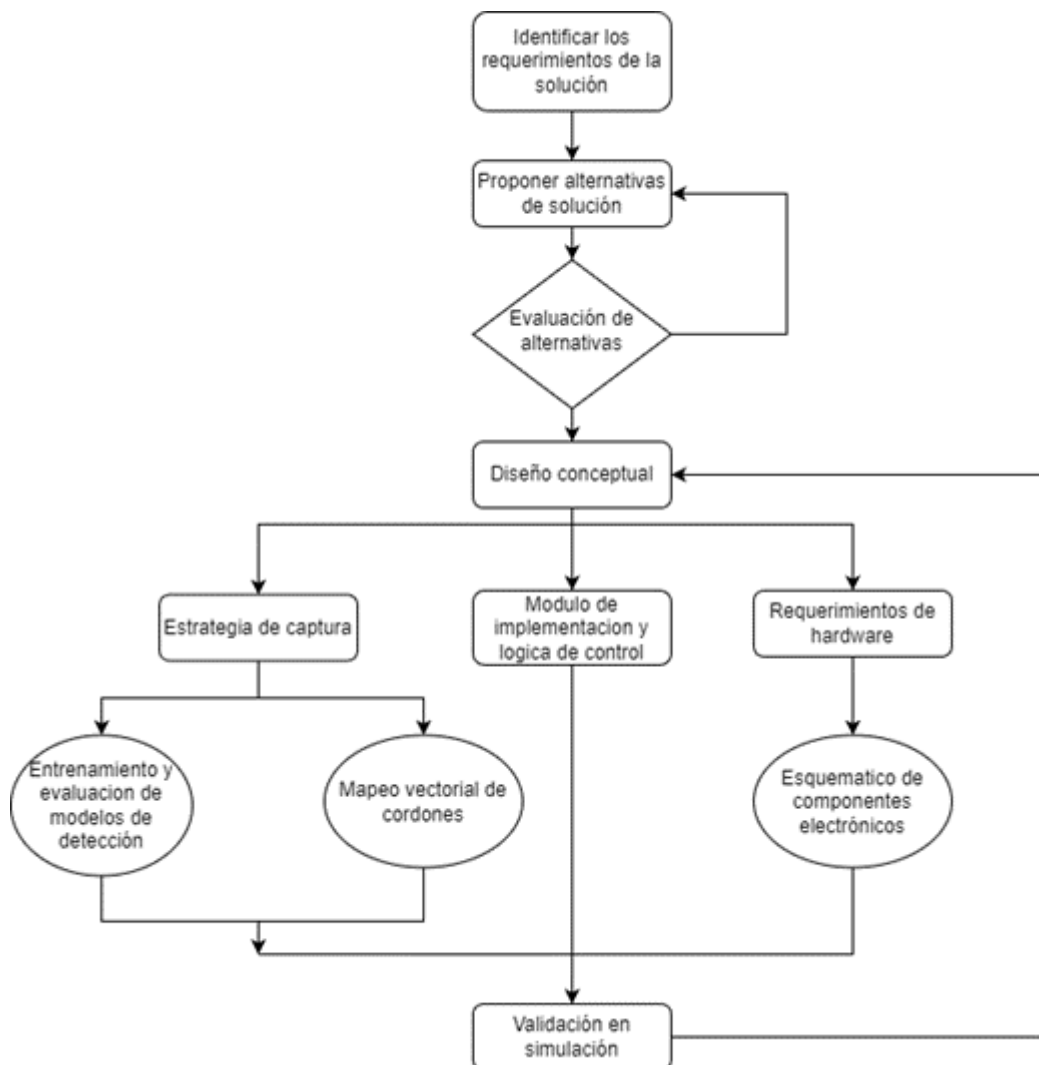
2.4 Metodología de diseño del sistema

La metodología de diseño utilizada para el desarrollo del sistema de detección de puntos de poda se basó en un enfoque secuencial–iterativo, orientado a la ingeniería de sistemas. Este enfoque permitió organizar el proceso desde la identificación de los requerimientos hasta la validación del sistema, integrando de manera coherente los subsistemas de programación y electrónica.

La Figura 2.4 presenta el diagrama general de la metodología empleada, el cual describe el flujo de trabajo seguido y las relaciones de retroalimentación entre sus diferentes etapas.

Figura 2.4

Metodología de diseño



El proceso se inició con la identificación de los requerimientos de la solución, en la que se definieron las necesidades agronómicas y técnicas del sistema. Estos requerimientos sirvieron como base para la formulación de alternativas de solución, en las que se evaluaron distintas estrategias de percepción visual, uso de información de profundidad y esquemas de implementación.

Posteriormente, se realizó la evaluación de las alternativas propuestas, permitiendo seleccionar la opción que ofrecía el mejor equilibrio entre desempeño, viabilidad computacional e integración con sistemas embebidos. Cuando una alternativa no cumplía los criterios establecidos, el proceso regresaba a la etapa de proposición, garantizando un refinamiento iterativo del diseño.

Una vez definida la solución, se desarrolló el diseño conceptual del sistema, donde se establecieron los principales bloques funcionales. A partir de este diseño se ejecutaron en paralelo tres líneas de desarrollo: la estrategia de captura y procesamiento de imágenes, el módulo de implementación y lógica de control, y la definición de los requerimientos de hardware.

En el subsistema de programación se abordó el entrenamiento y evaluación de modelos de detección, así como el mapeo vectorial de las estructuras de la vid, con el fin de obtener representaciones geométricas útiles para la toma de decisiones de poda. En paralelo, el subsistema electrónico se desarrolló a partir del esquemático de componentes necesarios para soportar la adquisición de datos y la ejecución del pipeline de procesamiento.

Finalmente, todos los módulos convergieron en una etapa de validación en simulación, en la que se evaluó el comportamiento del sistema de forma integrada. Los resultados de esta validación permitieron verificar la coherencia del diseño y, cuando fue necesario, retroalimentar las etapas de diseño para realizar ajustes antes de su implementación final.

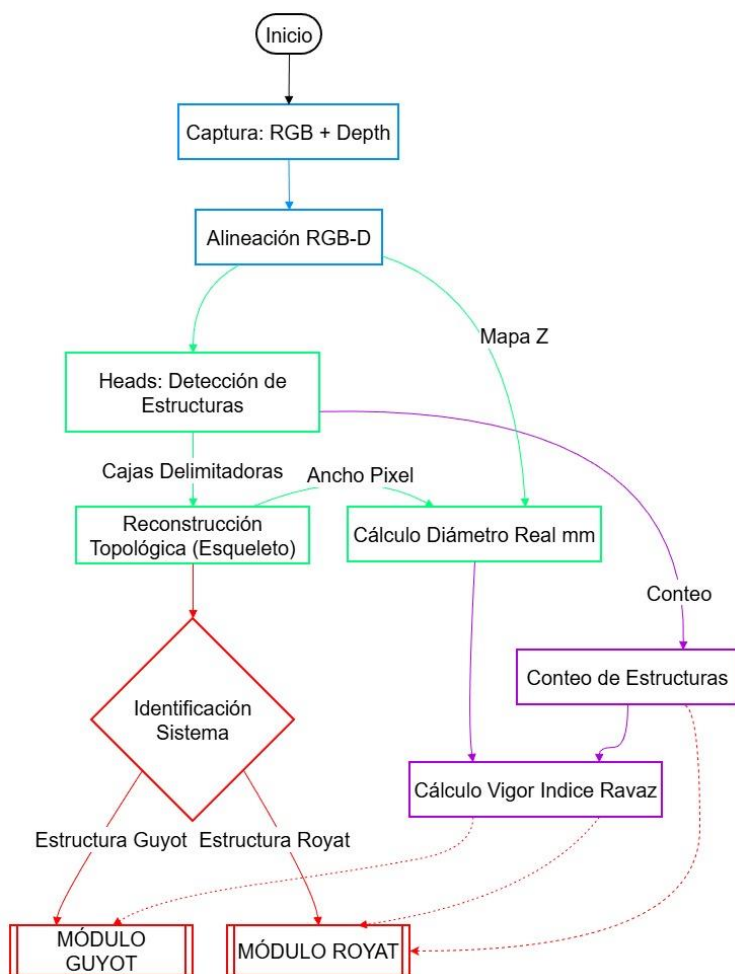
2.5 Diseño del subsistema de programación

El diseño del subsistema de programación implementa un pipeline secuencial de percepción visual orientado a la caracterización agronómica de la vid y la determinación autónoma de puntos de poda. El flujo de procesamiento, detallado en la Figura 2.5, inicia con la adquisición

de datos RGB-D y culmina en la generación de comandos de corte específicos según el sistema de conducción (Guyot o Royat).

Figura 2.5

Diagrama de Flujo General



La arquitectura se divide en cuatro etapas macro: (1) Percepción Multimodal, recibe y comunica la información de la cámara RGB-D; (2) Inferencia y Segmentación, encargada de la extracción de características visuales; (3) Metrología y Estimación de Vigor, donde se calculan

parámetros físicos reales; y (4) Lógica de Decisión Agronómica, donde se aplican las reglas de poda según el estado de la planta.

2.5.1 Detección de Instancias y Reconstrucción Topológica (*Skeletonization*)

La primera etapa del pipeline sustituye la segmentación semántica tradicional por un enfoque de detección de objetos orientado a la reconstrucción topológica de la vid. En lugar de procesar máscaras densas, el sistema emplea una arquitectura de detección (basada en el *backbone* I-JEPA) para inferir Cajas Delimitadoras (*Bounding Boxes*) que encierran las estructuras clave: sarmientos, yemas, pulgares y cordones.

- La generación de la estructura o "esqueleto" de la planta se realiza mediante un algoritmo de post-procesamiento que opera sobre las coordenadas de las cajas detectadas:
- Extracción de Nodos: Se calcula el centroide geométrico (u_c, v_c) de cada bounding box detectada, tratando a cada yema y segmento de sarmiento como un nodo discreto en el plano imagen.
- Conectividad Secuencial: Mediante un análisis de adyacencia espacial y coherencia direccional, el algoritmo enlaza los centroides de las cajas consecutivas. Esto genera un grafo o "esqueleto" lineal que representa la continuidad física del sarmiento.
- Estimación de Espesor: El diámetro de la estructura se infiere directamente a partir de la dimensión menor del *bounding box*, asumiendo que la detección se ajusta a los bordes laterales del sarmiento.

Este enfoque reduce drásticamente la carga computacional respecto a la segmentación por píxeles, permitiendo una inferencia más rápida en la plataforma Jetson Orin Nano, a la vez que proporciona directamente los vectores directores necesarios para el análisis geométrico.

2.5.2 Caracterización Geométrica y Cálculo del Índice de Ravaz

Una vez obtenidas las máscaras, el sistema cruza esta información con el mapa de profundidad (Mapa Z) y el ancho en píxeles de las estructuras para realizar el Cálculo de Diámetro Real en milímetros. Esto permite filtrar estructuras que no cumplan con el grosor mínimo para ser consideradas productivas.

Paralelamente, se ejecuta un módulo de Conteo de Estructuras que, en combinación con los datos biométricos obtenidos, alimenta el bloque de Cálculo de Vigor (Índice de Ravaz). Este índice es determinante para la toma de decisiones, ya que cuantifica el equilibrio entre la producción de fruta y el crecimiento vegetativo, permitiendo al sistema ajustar la intensidad de la poda (número de yemas a dejar) en tiempo real.

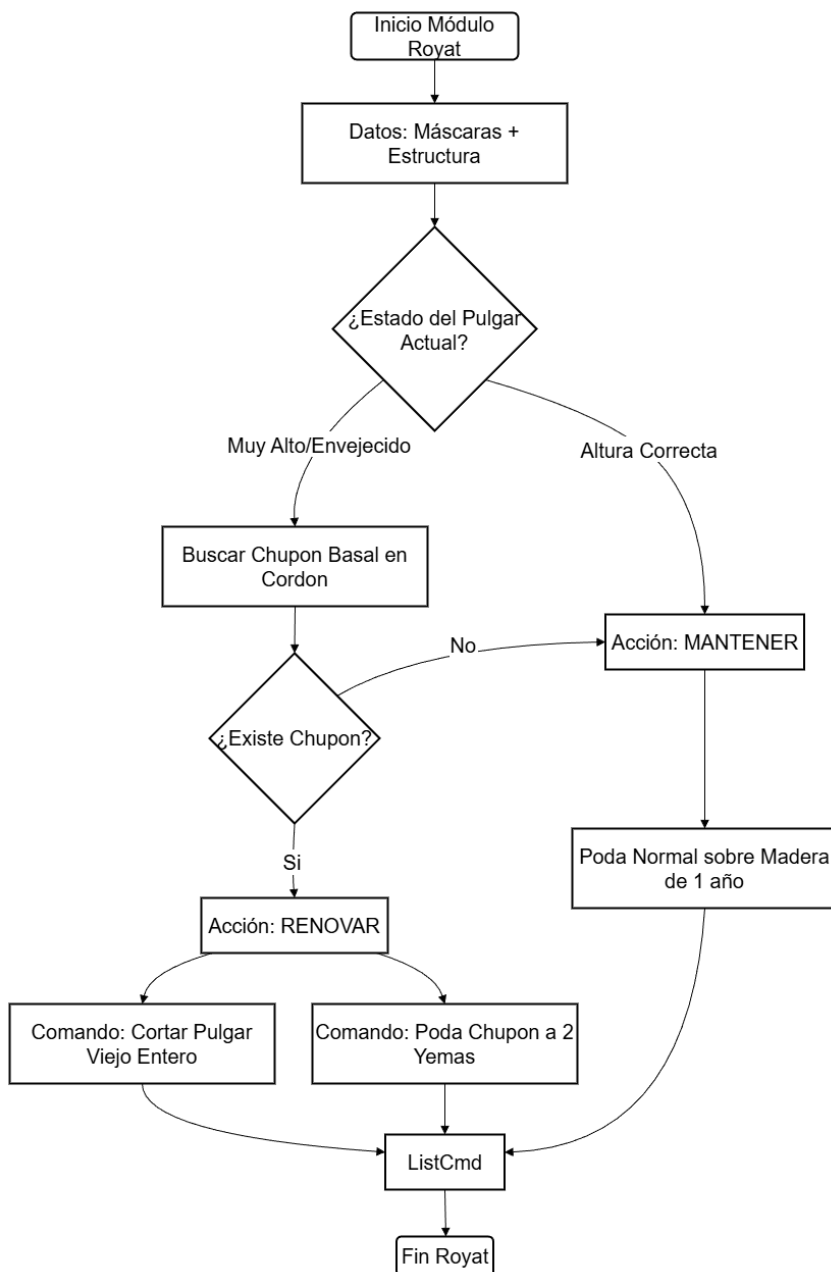
2.5.3 Identificación del Sistema y Módulos de Poda Específicos

El pipeline incluye un bloque de decisión condicional denominado "Identificación de Sistema", el cual clasifica la estructura analizada en uno de los dos sistemas de conducción predominantes: Guyot o Royat. Esta clasificación deriva el flujo de datos hacia módulos de decisión independientes mostrado en la Figura 2.6.

El diagrama de flujo específico para el módulo Royat ilustra una lógica de poda basada en el estado de la madera vieja. El algoritmo sigue la siguiente heurística:

Figura 2.6

Módulo de Poda Royat (Detalle del Diagrama 2)



- Evaluación del Pulgar Actual: Se analiza si el pulgar existente presenta una altura excesiva o signos de envejecimiento ("Muy Alto/Envejecido").

- Búsqueda de Renovación: Si el pulgar está envejecido, el sistema busca activamente un "Chupón Basal" en el cordón.
- Toma de Decisión (Renovar vs. Mantener):
- Acción RENOVAR: Si existe un chupón basal viable, se generan comandos para cortar el pulgar viejo entero y podar el chupón a 2 yemas, rejuveneciendo la estructura.
- Acción MANTENER: Si el pulgar tiene la altura correcta o no existe chupón de renovación, se aplica una poda normal sobre la madera de un año.

2.5.4 Generación de Comandos y Salida

Finalmente, las decisiones lógicas de los módulos Guyot o Royat se traducen en una lista de comandos (*ListCmd*). Estos comandos contienen la información vectorial del punto de corte, expresada matemáticamente como un vector de pose P_c :

$$P_c = [x, y, z, \alpha, \beta, \gamma]^T \quad (2.1)$$

Donde (x, y, z) representan las coordenadas espaciales del punto de incisión y (α, β, γ) la orientación de la herramienta necesaria para ejecutar el corte sin dañar las yemas adyacentes.

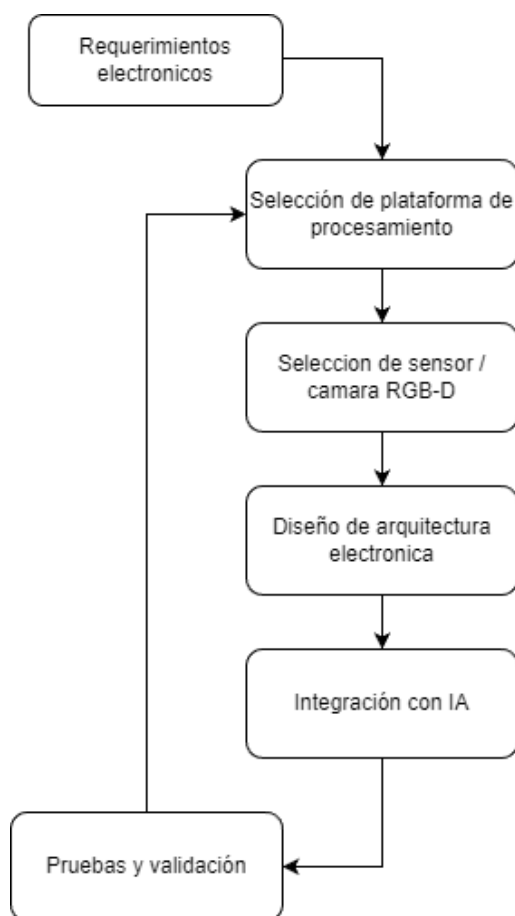
2.6 Diseño del subsistema electrónico

El diseño del subsistema electrónico se estructuró como un proceso secuencial e iterativo, representado mediante el diagrama de bloques de la Figura 2.7. Dicho flujo establece una progresión desde la definición de requerimientos electrónicos hasta las pruebas y validación del sistema integrado, permitiendo que los resultados experimentales retroalimenten decisiones previas de selección de hardware y arquitectura. Esta estructura evita la selección aislada de

componentes y asegura coherencia entre el pipeline algorítmico y la plataforma física que lo ejecuta.

Figura 2.7

Metodología subsistema electrónico



2.6.1 Derivación de requerimientos electrónicos a partir del pipeline de percepción

El diseño del subsistema electrónico se deriva directamente de la arquitectura metodológica definida en las secciones 2.4 y 2.5, donde se estableció un pipeline híbrido basado en detección 2D asistida por información de profundidad (RGB-D) y análisis geométrico tridimensional ligero. En el marco del diagrama metodológico adoptado, esta etapa corresponde al bloque inicial de

“Requerimientos electrónicos”, a partir del cual se condicionan todas las decisiones posteriores de selección de hardware y arquitectura.

Este pipeline impone requerimientos computacionales y de adquisición de datos claramente cuantificables. En particular, el sistema debe:

- Procesar imágenes RGB de alta resolución ($\geq 1280 \times 720$ px) junto con mapas de profundidad sincronizados.
- Ejecutar modelos de detección y segmentación basados en redes neuronales profundas en tiempo cuasi-real.
- Proyectar detecciones 2D al espacio 3D usando mapas de profundidad para calcular posiciones y orientaciones de puntos de poda.
- Mantener latencias compatibles con operación interactiva o robótica futura.

Estos requerimientos excluyen plataformas basadas en microcontroladores o SBC sin aceleración por GPU, y obligan al uso de plataformas embebidas con aceleración paralela para inferencia de IA y procesamiento geométrico.

2.6.2 Evaluación de plataformas de procesamiento

En esta etapa se elige la capacidad de cómputo que hará viable, o no, la ejecución del pipeline de percepción definido en las secciones previas.

A partir de los requerimientos anteriores, se analizaron tres clases de plataformas que representan los niveles de escalabilidad del sistema:

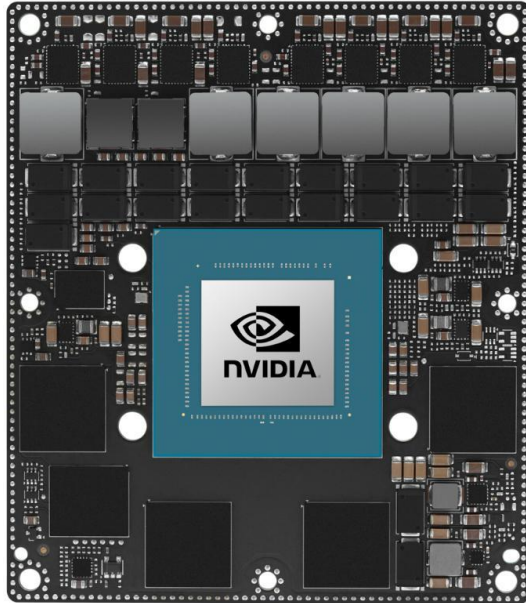
Tabla 2.3*Comparativa de plataformas de procesamiento*

Característica	Jetson Orin NX (16 GB)	Jetson AGX Orin	PC Industrial (Neosys / ADLINK)
Rendimiento IA	100 TOPS	275 TOPS	Variable (RTX 4060 Mobile)
GPU	<i>Ampere – 1024 cores</i>	<i>Ampere – 2048 cores</i>	<i>Ada Lovelace – 3072 cores</i>
CPU	8-core ARM A78AE	12-core ARM A78AE	Intel Core i7/i9
RAM	16 GB LPDDR5	64 GB LPDDR5	Hasta 64 GB DDR5
Consumo (TDP)	10–25 W	15–60 W	>100 W
Factor de forma	SOM compacto	Industrial robusto	Chasis rack

Cabe destacar que, si bien la Jetson AGX Orin presenta un consumo máximo superior al de la Jetson Orin NX, su capacidad de operar en modos intermedios de potencia (≈ 30 W) permite mantener una relación favorable entre desempeño y consumo energético para aplicaciones de percepción en tiempo real.

Figura 2.9

Jetson AGX Orin Industrial [27]



- PC industrial: ofrece máxima capacidad, pero con consumo y tamaño incompatibles con sistemas móviles.

Para la tesis se utiliza Jetson Orin Nano, como plataforma de validación, ya que pertenece a la misma familia Orin y comparte el mismo ecosistema de aceleración por GPU (CUDA, TensorRT, ROS2), lo que permite extrapolar tendencias de desempeño hacia plataformas de mayor capacidad como Orin NX, permitiendo extrapolar resultados de desempeño.

2.6.3 Selección del sensor RGB-D y configuración geométrica

Para la arquitectura de hardware propuesta, se selecciona la cámara Intel RealSense D455 como sensor de percepción espacial del sistema. Este dispositivo resulta adecuado para el despliegue final debido a que proporciona imagen RGB sincronizada con información de

profundidad métrica, permitiendo la reconstrucción tridimensional del entorno de trabajo y la localización espacial de estructuras de interés.

La configuración geométrica adoptada es *eye-to-hand*, es decir, una cámara fija observando la vid desde una posición estable.

Esta decisión se justifica por cuatro razones técnicas directamente ligadas al problema de poda:

1. Baja oclusión en poda invernal

En ausencia de hojas, un punto de vista fijo desde la base permite observar la mayor parte de los sarmientos sin necesidad de mover el sensor.

2. Estabilidad geométrica

Un sensor fijo elimina vibraciones y cambios de pose, mejorando la precisión métrica del mapa de profundidad usado para estimar distancias y ángulos de corte.

3. Menor complejidad computacional

La transformación entre la cámara y el sistema de coordenadas del robot es constante, lo que evita recalibraciones dinámicas y reduce latencia.

4. Ventaja del estéreo activo

La proyección de patrones infrarrojos del sistema *RealSense* mejora la reconstrucción de superficies de madera lisa de los sarmientos, donde el estéreo pasivo pierde calidad.

Esta configuración condiciona directamente la arquitectura electrónica, ya que permite que el procesamiento geométrico se realice en un marco de referencia fijo.

2.6.4 Arquitectura electrónica resultante

A partir de las decisiones anteriores, la arquitectura electrónica queda definida por tres bloques:

1. Sensor RGB-D (*RealSense D455*)

Genera flujos sincronizados de imagen y profundidad.

2. Plataforma embebida Jetson Orin (Nano en la validación)

Ejecuta:

- a. El modelo de detección 2D
- b. La proyección RGB-D
- c. El cálculo geométrico del punto de poda

3. Interfaces de salida

El sistema entrega como resultado un vector de pose de corte, compuesto por:

- a. Posición 3D del punto de poda
- b. Orientación estimada del sarmiento

Este vector constituye la interfaz directa hacia un futuro sistema robótico

2.6.5 Diferenciación entre validación y despliegue industrial

El diseño del sistema prescribe la combinación Jetson Orin Nano más cámara Intel *RealSense D455* como arquitectura objetivo para el despliegue del sistema de poda automatizada. Esta configuración resulta adecuada para evaluar aspectos como la precisión de detección y segmentación de estructuras críticas, robustez ante oclusiones y viabilidad del análisis geométrico tridimensional.

Sin embargo, para la validación metodológica presentada en esta tesis, la evaluación del pipeline algorítmico se realizó empleando *datasets* de imágenes RGB, complementados con estimación de profundidad mediante modelos de inteligencia artificial, lo que permitió simular la entrada de un sensor de profundidad sin depender de adquisiciones en campo en tiempo real.

Para un despliegue industrial a mayor escala, se contempla el uso de plataformas de mayor capacidad, tales como:

- Jetson AGX Orin Industrial o PC industrial
- Cámaras estéreo pasivas de grado industrial (ej. ZED 2i)

Lo cual permitiría una mayor resolución espacial, una operación más robusta en entornos exteriores y la integración con múltiples sensores, sin modificar el pipeline algorítmico desarrollado.

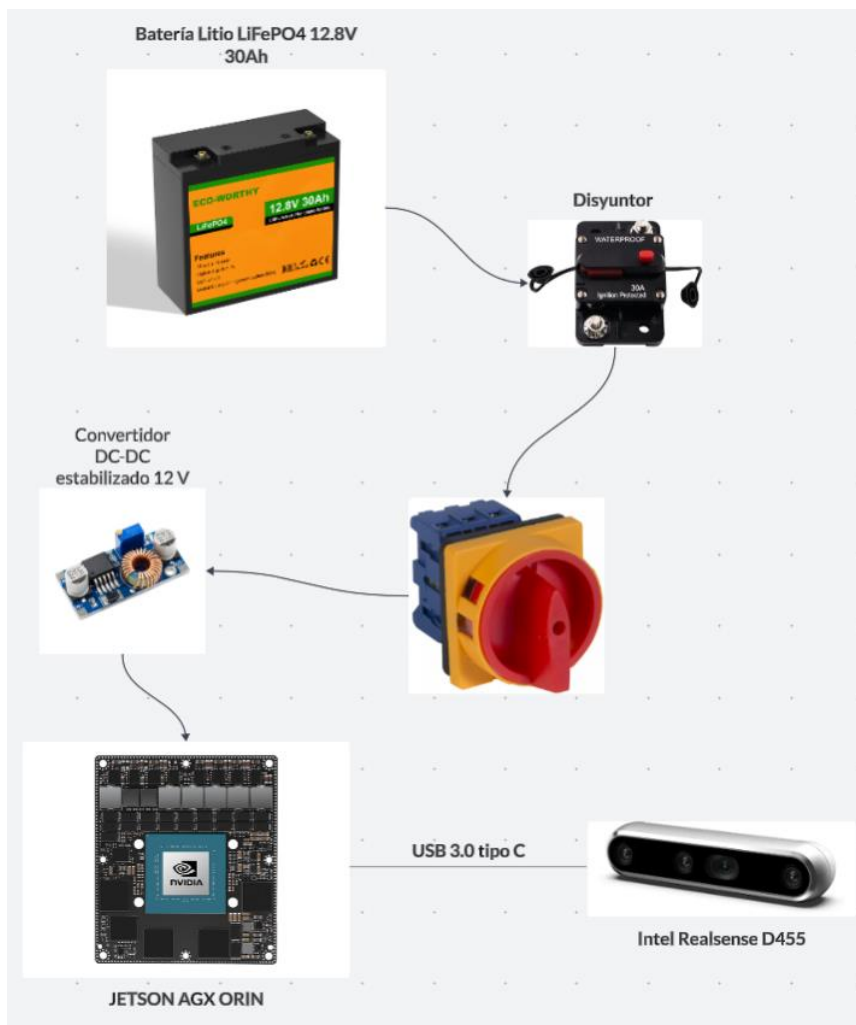
A partir del análisis comparativo de consumo energético de las plataformas evaluadas, se definió que el sistema debía ser compatible con una alimentación autónoma de baja tensión, típica de aplicaciones móviles agrícolas. En consecuencia, el diseño electrónico se orientó a una arquitectura de 12 V DC, adecuada para su futura integración con baterías de tecnología LiFePO₄.

El dimensionamiento específico de la fuente de alimentación y el análisis de autonomía se desarrollan en el Capítulo 3, con base en el consumo efectivo del sistema y el escenario operativo planteado.

La Figura 2.10 presenta la arquitectura electrónica propuesta para el despliegue final del sistema, en la cual se ilustran los principales componentes de hardware y su interconexión dentro del sistema de poda automatizada.

Figura 2.10

Arquitectura electrónica propuesta para el despliegue final del sistema.



CAPÍTULO 3

3.1 Entorno de experimentación

Con el fin de garantizar la reproducibilidad de los resultados y asegurar una correcta convergencia de los modelos de detección, se definió una configuración experimental estandarizada para el entrenamiento de las arquitecturas YOLO11n y YOLO11m. Esta configuración permitió evaluar el desempeño de los modelos bajo condiciones controladas y comparables, evitando que variaciones en los parámetros de entrenamiento influyan en los resultados obtenidos.

El proceso de entrenamiento se realizó utilizando el *framework Ultralytics YOLO*, ampliamente empleado en tareas de detección de objetos en visión por computador, lo que facilitó la gestión del dataset, la definición de hiperparámetros y el monitoreo de métricas durante el entrenamiento. Los experimentos se ejecutaron en una estación de trabajo de alto rendimiento, previo a la migración y validación parcial en plataformas embebidas.

Para adaptar las arquitecturas YOLO a la detección de estructuras finas de la vid, como sarmientos y yemas, se definió un conjunto de hiperparámetros orientados a maximizar la precisión sin comprometer la estabilidad del entrenamiento. La configuración empleada se resume en la Tabla 3.1.

Tabla 3.1*Configuración de entrenamiento*

Parámetro	Valor	Justificación técnica
Épocas (epochs)	100	Número suficiente para asegurar la convergencia de la función de pérdida y la estabilización de las métricas de desempeño, evitando sobreajuste.
Resolución de entrada (imgsz)	640 × 640	Resolución estándar que ofrece un buen equilibrio entre velocidad de inferencia y capacidad para detectar objetos pequeños como yemas.
Tamaño de batch (batch)	16	Ajustado a la capacidad de memoria de la GPU, garantizando estabilidad en la actualización de los pesos durante la retropropagación.
Optimizador	Auto (SGD / AdamW)	Selección automática del optimizador por parte del framework según las características del dataset, permitiendo una minimización eficiente de la función de pérdida.
Aumentación de datos	Estándar	Aplicación de transformaciones geométricas básicas (flip, escalado), desactivando Mosaic en las últimas épocas para refinar la precisión del modelo.

Esta configuración se mantuvo constante para todas las arquitecturas evaluadas (YOLO11n y YOLO11m), lo que permitió realizar una comparación justa del desempeño basada

exclusivamente en la complejidad y capacidad de cada red, y no en variaciones de los parámetros de entrenamiento.

3.2 Implementación de modelos de percepción

La implementación de los modelos de percepción visual constituye una etapa central del sistema propuesto, ya que de ella depende la correcta identificación de las estructuras agronómicas necesarias para la determinación del punto de poda. En esta sección se describen los modelos seleccionados, su implementación experimental y los criterios que motivaron el uso de dos arquitecturas diferenciadas en función de la plataforma de procesamiento.

El enfoque adoptado no busca comparar modelos desde una perspectiva competitiva, sino evaluar su viabilidad práctica dentro de distintos escenarios de despliegue, considerando las limitaciones computacionales y los requerimientos del sistema.

3.2.1 Implementación del modelo YOLO11n para Jetson Orin Nano

El entrenamiento del modelo se realizó utilizando el conjunto de datos previamente preparado, el cual contiene imágenes anotadas de estructuras relevantes de la vid, principalmente sarmientos y yemas. Durante esta fase, el objetivo principal fue verificar que el modelo fuera capaz de detectar dichas estructuras de forma consistente y estable, más que maximizar métricas de precisión absoluta.

Los resultados obtenidos, en la Figura 3.1, muestran que el modelo YOLO11n logra identificar correctamente las estructuras principales en condiciones controladas, generando salidas coherentes que permiten alimentar las etapas posteriores del pipeline, como la recuperación de

información tridimensional y el cálculo geométrico del punto de poda. Si bien su desempeño es inferior al de arquitecturas más grandes, los resultados son funcionales para tareas de validación y pruebas de integración sobre hardware embebido.

Figura 3.1

Inferencia del modelo YOLO11n sobre el dataset de validación



En este contexto, el modelo YOLO11n se utiliza como modelo operativo de validación, permitiendo comprobar la viabilidad del sistema completo sin introducir una carga computacional excesiva sobre la plataforma.

3.2.2 Implementación del modelo YOLO11m para Jetson AGX Orin

Para la versión final y escalable del sistema, se implementó el modelo YOLO11m, orientado a su ejecución sobre una plataforma de mayor capacidad de procesamiento, la Jetson AGX Orin. Esta arquitectura presenta un mayor número de parámetros y una capacidad superior para modelar características visuales complejas, lo que se traduce en una mejora en la precisión y estabilidad de las detecciones.

El modelo YOLO11m fue entrenado utilizando el mismo conjunto de datos base, garantizando consistencia en la evaluación y permitiendo una comparación directa en términos de comportamiento general con lo que se aprecia en la Figura 3.2. Durante el entrenamiento, se observó una convergencia estable de las métricas, así como una mejora significativa en la detección de estructuras pequeñas y en escenas visualmente más complejas.

3.2.3 *Justificación del uso de dos arquitecturas de percepción*

La utilización de dos arquitecturas distintas responde a una decisión metodológica orientada a la validación progresiva y escalable del sistema, más que a una comparación directa de desempeño.

El modelo YOLO11n permite validar el funcionamiento del pipeline completo —detección, recuperación tridimensional y cálculo geométrico— sobre una plataforma embebida de menor capacidad, demostrando que el enfoque propuesto es viable incluso bajo restricciones computacionales. Por su parte, el modelo YOLO11m representa la solución más adecuada para un despliegue final en entornos reales, donde se prioriza la precisión y la robustez del sistema.

De esta manera, ambos modelos cumplen roles complementarios dentro del desarrollo del proyecto:

YOLO11n como herramienta de validación experimental y YOLO11m como arquitectura de referencia para la versión industrial del sistema.

Esta estrategia evita generalizaciones innecesarias y permite al sistema adaptarse a distintos escenarios de implementación sin comprometer los objetivos funcionales definidos al inicio del proyecto.

3.3 **Evaluación del desempeño del modelo seleccionado**

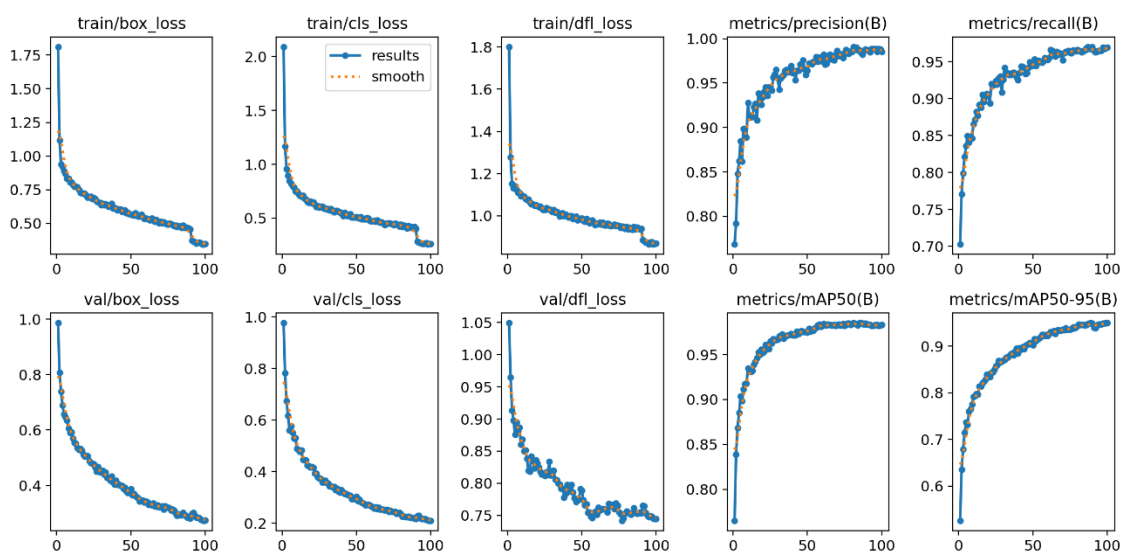
El desempeño del modelo YOLO11m se evaluó a partir de las métricas generadas durante el entrenamiento y validación, considerando las curvas de pérdida, las métricas de precisión, *recall* y mAP, así como la matriz de confusión normalizada.

3.3.1 Curvas de aprendizaje y métricas de desempeño

En la Figura 3.3 se presentan las curvas de pérdida correspondientes al entrenamiento y validación del modelo. La pérdida de localización (*box loss*) disminuye progresivamente desde valores cercanos a 1.8 hasta aproximadamente 0.4 en entrenamiento y 0.28 en validación, lo que indica una mejora sostenida en la precisión espacial de las detecciones. De forma similar, la pérdida de clasificación (*cls loss*) desciende desde valores superiores a 2.0 hasta aproximadamente 0.4 en entrenamiento y 0.22 en validación, evidenciando que el modelo aprende a diferenciar adecuadamente entre sarmientos y fondo.

Figura 3.3

Resultados YOLO11m



Las pérdidas asociadas a la regresión de distancias (*dfl loss*) muestran un comportamiento estable y decreciente en ambas fases, lo que confirma la convergencia del proceso de aprendizaje. En todos los casos, las curvas de validación se mantienen cercanas o por debajo de las curvas de

entrenamiento, indicando ausencia de sobreajuste y una adecuada capacidad de generalización del modelo.

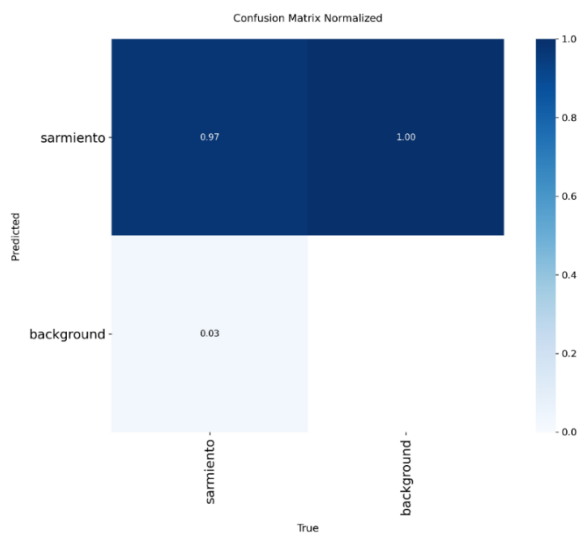
La misma figura muestra la evolución de las métricas de desempeño. El modelo alcanza valores finales de precisión cercanos a 0.98 y *recall* aproximado de 0.96, lo que indica que la mayoría de las detecciones realizadas son correctas y que el sistema logra identificar casi la totalidad de los sarmientos presentes en las imágenes. Asimismo, el valor de *mAP50* alcanza aproximadamente 0.98, mientras que el *mAP50-95* se sitúa alrededor de 0.94, reflejando un desempeño robusto incluso bajo criterios de intersección más estrictos.

3.3.2 Matriz de confusión

La Figura 3.4 corresponde a la matriz de confusión normalizada del modelo YOLO11m, utilizada para analizar el comportamiento de clasificación entre las diferentes clases del sistema.

Figura 3.4

Matriz de confusión normalizada YOLO11m



En esta representación se observa que la clase sarmiento es correctamente clasificada en aproximadamente el 97% de los casos, mientras que solo un 3% de estas estructuras es confundido con el fondo. No se registran errores significativos donde el fondo sea clasificado como sarmiento, lo cual reduce el riesgo de falsas detecciones. Este comportamiento es especialmente relevante para la aplicación de poda, donde es preferible omitir una detección antes que generar un punto de corte incorrecto.

3.3.3 *Análisis global del desempeño*

En conjunto, estos resultados confirman que el modelo YOLO11m presenta un desempeño estable, preciso y confiable para la detección de sarmientos, cumpliendo con los requerimientos necesarios para alimentar el módulo geométrico encargado de la determinación del punto de poda.

3.4 Generación de percepción espacial (estimación de profundidad)

Para la determinación del vector de corte, el sistema requiere coordenadas espaciales (x , y , z). Si bien la arquitectura de hardware propuesta e integrada (ver Sección 2.6) incorpora una cámara *Intel RealSense D455* para la obtención de mapas de profundidad en tiempo real, la validación del algoritmo de poda se realizó sobre el *dataset* de prueba, el cual consta exclusivamente de imágenes RGB bidimensionales sin información de profundidad asociada.

Dado que no fue factible realizar adquisiciones de campo in situ en un viñedo real con el sensor físico debido a restricciones logísticas, se optó por una estrategia de validación basada en la estimación de profundidad monocular. Para ello, se implementó el modelo *Depth Anything*, una arquitectura basada en *transformers* capaz de inferir mapas de profundidad relativa de alta fidelidad a partir de imágenes simples.

Esta aproximación permitió dotar al *dataset* 2D de una "pseudo-profundidad", habilitando la validación matemática del algoritmo de selección de puntos de corte y la orientación del efector final, tal como se visualiza en los resultados siguientes. Es importante notar que, en la fase de despliegue físico, este módulo de estimación por IA es reemplazado directamente por la lectura nativa del sensor *RealSense*, garantizando métricas absolutas en milímetros.

3.4.1 Comparación visual entre imagen RGB y mapa de profundidad

La Figura 3.5 presenta una comparación visual entre la imagen RGB original y el mapa de profundidad estimado mediante *Depth Anything*. A partir de este mapa, se pudo observar que el sistema logra identificar gradientes de profundidad coherentes con la estructura general de la vid, lo que permitió validar el funcionamiento del módulo geométrico en ausencia de datos RGB-D reales.

Figura 3.5

Comparación imagen RGB (izquierda) y mapa de profundidad estimado (derecha)



Aunque la profundidad obtenida no corresponde a valores métricos reales, la distribución relativa de profundidades permite identificar la continuidad del sarmiento y su orientación espacial, información suficiente para validar el cálculo geométrico del punto de poda.

Sin embargo, es importante destacar que *Depth Anything* no forma parte del diseño final del sistema. Las estimaciones obtenidas son relativas y no garantizan una correspondencia métrica precisa con el entorno real, especialmente en escenarios donde los sarmientos se superponen, presentan curvaturas pronunciadas o existen oclusiones parciales. En estos casos, la falta de una profundidad métrica confiable puede conducir a errores en la ubicación espacial del punto de poda.

En consecuencia, el diseño final del sistema contempla el uso de una cámara RGB-D *RealSense*, capaz de proporcionar mediciones de profundidad métricas y consistentes en tiempo real. Esta decisión está alineada tanto con la arquitectura metodológica definida en el Capítulo 2 como con el análisis de costos presentado en la Sección 3.7, donde se consideran explícitamente los componentes de hardware necesarios para una implementación funcional en campo.

En síntesis, el uso de *Depth Anything* en este trabajo debe interpretarse como una herramienta de validación interina y experimental, empleada para comprobar la coherencia del pipeline de percepción espacial, mientras que la obtención de profundidad real y operativa queda reservada al uso de sensores RGB-D dedicados en la implementación final del sistema robótico.

3.4.2 Limitaciones de la estimación de profundidad monocular

A pesar de su utilidad metodológica, la estimación de profundidad monocular presenta limitaciones importantes que deben ser consideradas:

- La profundidad obtenida es relativa, no métrica, por lo que no puede ser utilizada directamente para decisiones de corte en un sistema físico.

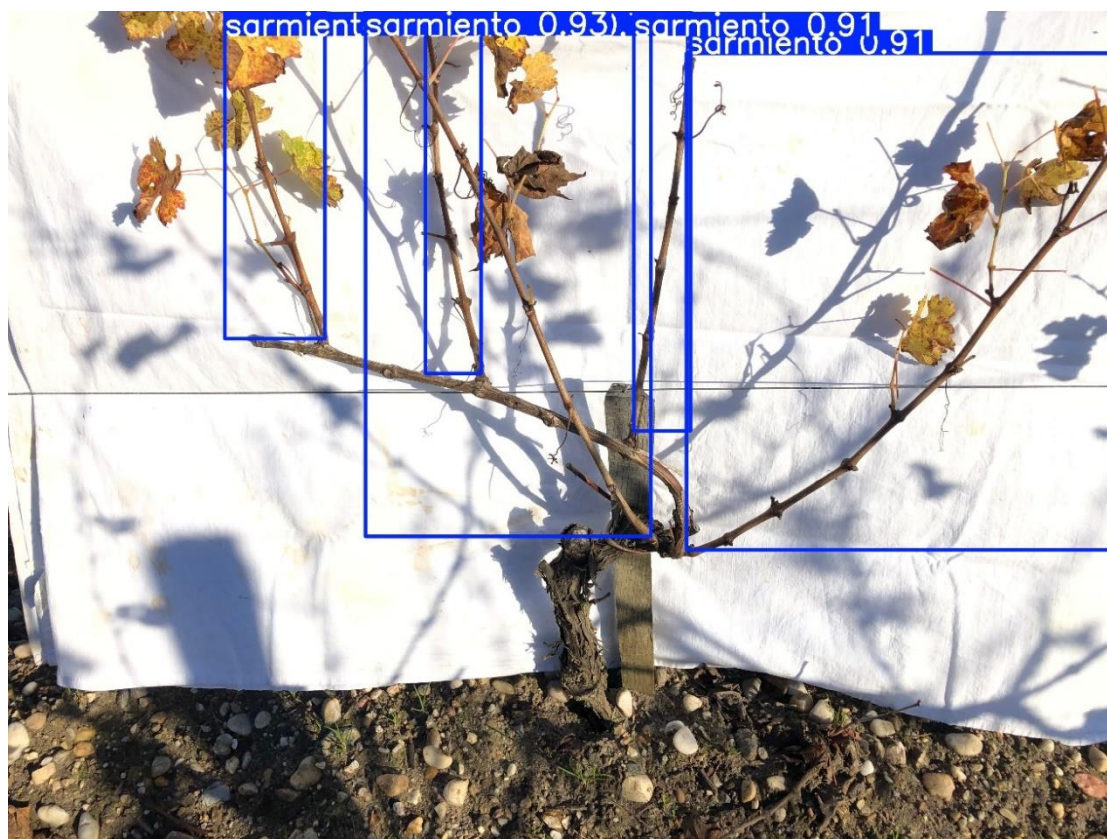
- La precisión de la estimación depende fuertemente de las características visuales de la escena, la iluminación y el contraste.

En una implementación final del sistema, la información de profundidad sería proporcionada por la cámara RGB-D o sistemas estéreo calibrados, permitiendo obtener coordenadas tridimensionales métricas y mejorar la robustez de la determinación del punto de poda.

3.5 Determinación del punto de poda y vector de corte

Esta sección describe el proceso mediante el cual el sistema propuesto utiliza la información obtenida en las etapas de percepción visual y estimación de profundidad para determinar el punto de poda y la orientación del corte sobre el sarmiento. El análisis se centra en la validación funcional del pipeline geométrico a partir de imágenes de validación, sin considerar la ejecución física del corte.

El proceso de determinación del punto de poda inicia con la detección de los sarmientos en imágenes RGB mediante los modelos de percepción seleccionados. A partir de las detecciones generadas por el modelo, se obtienen las regiones de interés correspondientes a cada sarmiento visible en la escena como se muestra en la Figura 3.6; **Error! No se encuentra el origen de la referencia.** a continuación.

Figura 3.6*Detección de sarmientos*

Los resultados obtenidos muestran que el modelo seleccionado identifica de forma consistente los sarmientos principales presentes en las imágenes de validación, proporcionando una base confiable para las etapas posteriores del pipeline.

Una vez detectado el sarmiento, se recuperó información tridimensional asociada a la región de interés. En ausencia de un viñedo real y de un sensor de profundidad físico durante esta etapa del proyecto, la profundidad fue estimada a partir de imágenes RGB mediante un modelo de estimación. Esta información permitió obtener una representación espacial aproximada de la geometría del sarmiento, suficiente para evaluar la coherencia del cálculo del punto de poda y del vector de corte en un entorno de validación controlado.

Como resultado del procesamiento geométrico, el sistema generó un punto de poda tridimensional junto con la orientación estimada del plano de corte mostrados en la Figura 3.7. Esta salida se representa visualmente sobre la imagen original, permitiendo evaluar de forma cualitativa la coherencia espacial de la solución propuesta.

Figura 3.7

Visualización de puntos de poda



Las visualizaciones muestran que el punto de poda se ubica de manera consistente sobre el eje del sarmiento, manteniendo una distancia adecuada respecto a la yema más cercana. Asimismo, el vector de corte presenta una orientación alineada con la geometría local del sarmiento, lo cual concuerda con criterios agronómicos básicos de poda.

Además de la estimación del punto de poda en el plano de la imagen, el sistema proporciona la ubicación espacial tridimensional de dicho punto, expresada en el sistema de coordenadas de la cámara como se aprecia en la Figura 3.8. Esta información se obtiene a partir de la asociación entre las detecciones del modelo de percepción y los valores de profundidad estimados, permitiendo recuperar coordenadas (x, y, z) del punto de corte.

Figura 3.8

Ubicación del punto de poda



Durante la etapa de validación, la información de profundidad fue obtenida mediante el modelo Depth Anything, el cual actúa como un sustituto funcional de un sensor RGB-D real. Si bien esta aproximación no ofrece una medición métrica absoluta, permite evaluar el comportamiento del sistema en términos de coherencia espacial y consistencia geométrica del

punto de poda generado. En un escenario de implementación final, esta referencia espacial estaría directamente asociada a un sensor de profundidad real, permitiendo una localización precisa del punto de corte dentro del entorno físico del viñedo.

Esto lo convierte en una de las limitaciones inherentes, no siempre logra representar correctamente la geometría real en configuraciones complejas como se muestra en la Figura 3.9 donde se observó que en escenarios donde los sarmientos presentan superposición significativa o curvaturas pronunciadas, la estimación del punto de poda pierde precisión.

Figura 3.9

Limitaciones del modelo



En contraste, cuando los sarmientos se presentan de forma aislada y con una geometría claramente definida, el sistema genera puntos de corte coherentes y alineados con los criterios

esperados. Este comportamiento evidencia que las desviaciones observadas no se originan en la lógica de decisión geométrica, sino en la calidad de la información espacial disponible durante la validación.

Estos resultados constituyen una validación funcional del enfoque planteado y sientan las bases para una implementación futura en entornos reales, donde la información de profundidad sería obtenida directamente mediante sensores especializados.

3.6 Integración en Sistema Embebido

La integración del sistema de percepción desarrollado se orienta a su ejecución en plataformas embebidas basadas en NVIDIA Jetson, considerando las restricciones de cómputo, consumo energético y capacidad de procesamiento en tiempo real propias de este tipo de dispositivos. En esta etapa, la validación se realizó principalmente a nivel de procesamiento de imágenes, dejando la implementación completa en hardware físico como una fase posterior del proyecto.

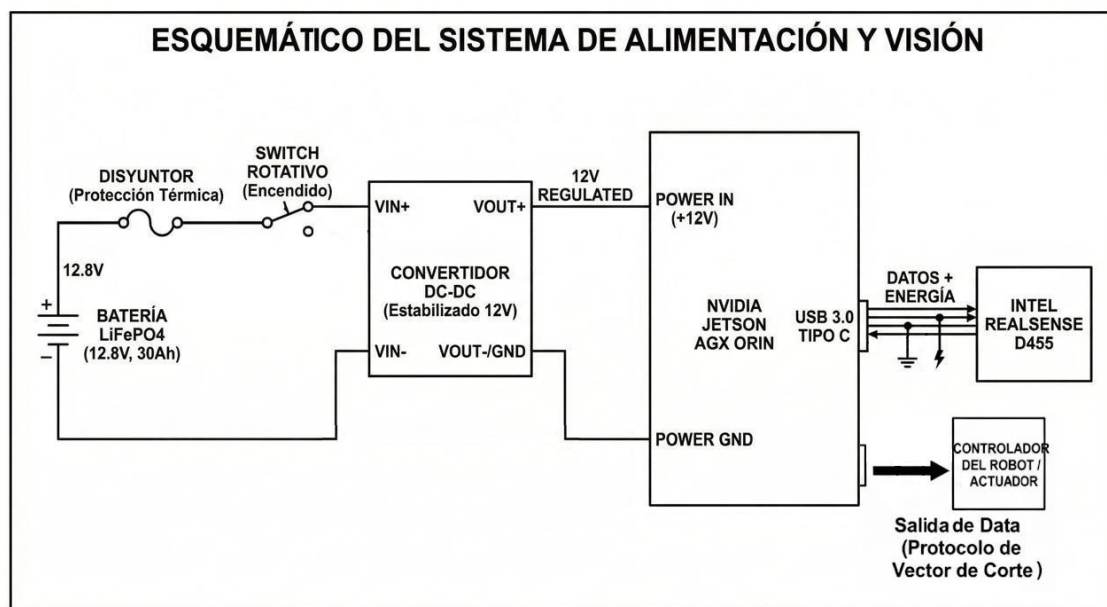
3.6.1 Arquitectura electrónica del sistema

La arquitectura electrónica propuesta contempla como núcleo de procesamiento una plataforma NVIDIA Jetson, encargada de ejecutar los modelos de visión por computador, el cálculo del punto de poda y la generación del vector de corte. Esta unidad se integra con los elementos de adquisición de imagen y los módulos de comunicación necesarios para una futura interacción con sistemas de actuación.

El esquema general de la arquitectura electrónica se presenta como una estructura modular, como muestra la Figura 3.10, permitiendo la sustitución o actualización de componentes sin afectar la lógica principal del sistema. Esta decisión de diseño facilita la adaptación del sistema a distintos entornos operativos y plataformas de cómputo embebido.

Figura 3.10

Esquemático del sistema de alimentación



Para que los resultados de la detección y el cálculo del punto de poda (PCP) tengan un efecto práctico dentro de un sistema robótico, la arquitectura integra un módulo de comunicación encargado de transmitir las coordenadas espaciales generadas al controlador del robot. Una vez procesada la información del sarmiento y calculado el vector de corte en el espacio tridimensional, la plataforma Jetson empaqueta los datos en un mensaje estructurado que incluye la posición [x, y, z] y el ángulo de aproximación θ .

Esta salida de datos ha sido concebida para ser compatible con distintos esquemas de integración, tales como sistemas basados en ROS 2 mediante tópicos de comunicación, o interfaces de comunicación directa (serial o UDP) con microcontroladores, PLCs o controladores de manipuladores. De este modo, el sistema de percepción se desacopla del actuador final, permitiendo su integración en diferentes configuraciones robóticas sin modificar el pipeline algorítmico desarrollado.

3.6.2 Integración en Plataformas Jetson

Como parte de la estrategia de integración, se consideraron distintas variantes de la familia Jetson en función del modelo de percepción empleado. El modelo YOLO11n fue orientado a plataformas de menor capacidad computacional, como Jetson Orin Nano, mientras que el modelo YOLO11m se asoció a plataformas de mayor desempeño, como Jetson AGX Orin.

Durante esta fase, la validación se llevó a cabo mediante la ejecución del pipeline completo de percepción sobre conjuntos de imágenes de validación, verificando la correcta detección de sarmientos, la estimación del punto de poda y la coherencia del vector de corte generado. Los resultados obtenidos confirman que el sistema puede ejecutarse de manera funcional en entornos embebidos, manteniendo consistencia en los resultados visuales respecto a las pruebas realizadas en entorno de desarrollo.

Para la plataforma NVIDIA Jetson Nano, se deja un espacio reservado para la implementación y validación futura, considerando que esta etapa requiere optimización adicional del modelo y ajustes específicos para cumplir con las limitaciones de memoria y procesamiento propias del dispositivo.

3.7 Análisis de costos del sistema

El análisis de costos del sistema propuesto se realizó con el objetivo de estimar la inversión necesaria para el desarrollo e implementación del prototipo funcional. Los costos se dividen en hardware y en costos asociados a integración y operación, considerando únicamente los elementos indispensables para la ejecución del sistema de percepción y estimación del punto de poda.

3.7.1 Costos de Hardware

La Tabla 3.2 presenta los principales componentes de *hardware* requeridos para el sistema, junto con su costo estimado en el mercado al momento del desarrollo del proyecto.

Tabla 3.2

Costos de Hardware del sistema

Elemento	Modelo / Especificación	Cantidad	Costo unitario (USD)	Costo total (USD)
Batería LiFePO₄	12.8 V – 30 Ah	1	180	180
Disyuntor térmico	30 A – <i>waterproof</i>	1	15	15
Switch rotativo industrial	Corte general ON/OFF	1	25	25
Convertidor DC-DC	<i>Step-down</i> 12.8 V → 12 V regulados (≥ 10 A)	1	20	20
Computador embebido	NVIDIA Jetson AGX Orin	1	2000	2000

Cámara RGB-D	Intel <i>RealSense</i> D455	1	350	350
Cable USB 3.0	Alta velocidad	1	15	15
Tipo C				
Cableado y conectores	Grado industrial	—	30	30
Total del				
Hardware				2635

3.7.2 Costos de integración y operación

Además del hardware, se consideraron los costos asociados a la integración del sistema, pruebas y operación básica, los cuales se resumen en la Tabla 3.3. Los costos asociados a estas actividades presentan una variabilidad inherente, ya que dependen de factores como la complejidad de la plataforma robótica, el nivel de experiencia del personal y el alcance del despliegue del sistema

Tabla 3.3*Costos de integración y operación*

Categoría	Actividad / Concepto	Descripción	Horas	Costo unitario (USD/h)	Costo total (USD)
Integración	Integración de Software	Adaptación del sistema a plataformas robóticas	30 h	20	600
Integración	Ajustes de compatibilidad	Formatos de salida, APIs, parámetros y comunicación	10 h	20	200
Capacitación	Capacitación del personal	Uso del sistema y análisis de resultados	—	—	320
Orientación	Mantenimiento anual	Ajuste de modelos, recalibración de cámara y sistema	—	—	700
TOTAL					1820

3.7.3 *Resumen General de Costos*

Para finalizar el análisis de costos, se presenta el Resumen General de Costos del proyecto. La Tabla 3.4 consolida la inversión inicial en componentes de hardware de alto desempeño junto

con los costos operativos y de integración técnica. Este balance permite determinar la inversión total estimada para la puesta en marcha de una unidad funcional de poda automatizada.

Tabla 3.4

Costos totales

Categoría	Costo Total (USD)
Hardware	2635
Integración y operación	1820
Costo total estimado	4455

CAPÍTULO 4

4.1 Conclusiones y recomendaciones

El desarrollo del sistema embebido para la detección de puntos de poda en plantas de vid permitió validar una solución integral que combina visión por computador, metrología 3D y toma de decisiones agronómica en una plataforma de cómputo en el borde.

El prototipo demostró viabilidad técnica para operar en hardware embebido (Jetson Orin Nano) y proyecta su despliegue final en una plataforma industrial (Jetson AGX Orin), manteniendo un mismo *pipeline* de percepción híbrido 2D + 3D.

A partir de los resultados obtenidos en la evaluación experimental del modelo YOLO11 y en la validación de la arquitectura electrónica y energética del sistema, se establecen las siguientes conclusiones y recomendaciones.

4.1.1 Conclusiones

- Se diseñó un sistema integrado de percepción RGB-D y se validó la factibilidad del enfoque híbrido 2D + 3D mediante el uso de estimación de profundidad monocular por aprendizaje profundo. Esto permitió comprobar la lógica de determinación de puntos de poda en coordenadas espaciales sobre *datasets* reales, superando la limitación de acceso físico a viñedos durante la fase de pruebas.
- El modelo YOLO11n demostró ser el más adecuado para la validación experimental en la Jetson Orin Nano, demostró viabilidad técnica para ejecutar el pipeline de procesamiento en hardware embebido, procesando exitosamente las inferencias de visión y la reconstrucción geométrica simulada con latencias compatibles para la operación en tiempo real.

- El modelo YOLO11m, entrenado con técnicas de aumentación de datos optimizadas, obtuvo el mejor desempeño global de detección, alcanzando valores de mAP50 aproximadamente de 0.98, mientras que el *mAP50–95* se sitúa alrededor de 0.94. Por ello, fue seleccionado como el modelo principal para la implementación final del sistema en la plataforma Jetson AGX Orin Industrial.
- El análisis de las curvas de entrenamiento y de la matriz de confusión confirmó que el modelo logra diferenciar adecuadamente las yemas y segmentos de sarmiento en un 97%, lo cual es fundamental para la posterior estimación geométrica y la generación del punto de corte.
- Se desarrolló la arquitectura electrónica embebida del sistema, integrando la selección de la plataforma de procesamiento, el sensor de visión y el subsistema de alimentación autónoma, lo que permite su implementación como un prototipo funcional de bajo consumo.
- El análisis de costos y autonomía energética demostró que la configuración final del sistema es viable para jornadas operativas en campo, proporcionando un balance entre capacidad de procesamiento, precisión visual y estabilidad energética.

4.1.2 Recomendaciones

Tras culminar el desarrollo y validación del sistema propuesto, se plantean las siguientes recomendaciones para su ampliación y mejora en trabajos futuros:

- Realizar pruebas de campo utilizando la cámara Intel *RealSense* D455 en un entorno de viñedo real para calibrar el algoritmo de poda con datos de profundidad

métrica absoluta. Esto permitirá contrastar los resultados de la estimación monocular utilizada en esta tesis con mediciones físicas reales, ajustando las tolerancias del vector de corte.

- Ampliar y diversificar el *dataset* de entrenamiento incorporando imágenes de diferentes sistemas de conducción, épocas de poda y condiciones reales de iluminación, con el fin de incrementar la generalización del modelo.
- Optimizar el proceso de etiquetado y la calidad de las cajas delimitadoras, especialmente en yemas pequeñas, para reducir la confusión con el fondo y mejorar la precisión de detección.
- Explorar el uso de modelos de segmentación semántica ligera o de redes híbridas que combinen detección y segmentación, con el objetivo de mejorar la definición geométrica de las estructuras de poda.
- Evaluar la integración del pipeline de percepción con un manipulador robótico en condiciones de campo, enfocándose en la sincronización entre el mapa de profundidad del sensor físico y el sistema de coordenadas del efector final para validar la precisión del corte en milímetros.
- Investigar el uso de técnicas de cuantización y optimización del modelo para mejorar la velocidad de inferencia en la Jetson Orin Nano sin comprometer la precisión del detector.
- Analizar la incorporación de sensores adicionales (multiespectrales o de mayor resolución) para mejorar la caracterización del vigor de la planta y enriquecer la toma de decisiones agronómicas.

- Desarrollar un sistema de control robótico que permita cerrar el ciclo completo desde la percepción hasta la acción de poda autónoma, consolidando el prototipo como una solución robótica integral para la viticultura de precisión.

Referencias

- [1] X. Wang, Y. Liu, and M. Zhang, “Detection Method for the Buds on Winter Vines Based on Computer Vision,” *Proc. IEEE Int. Conf. Mechatronics and Automation*, pp. 1234–1238, 2015
- [2] M. T. Mallick, S. Banerjee, N. Thakur, H. N. Saha, and A. Chakrabarti, “Evaluation of State-of-the-Art Deep Learning Techniques for Plant Disease and Pest Detection,” *Computers, Materials & Continua*, vol. 85, no. 1, pp. 121-180, 2025, doi: 10.32604/cmc.2025.065250.
- [3]] Z. Chen, Y. Wang, and S. Tong, “Grapevine Branch Recognition and Pruning Point Localization Technology Based on Image Processing,” *Applied Sciences*, vol. 14, no. 8, p. 3327, 2024. [Online]. Available: <https://www.mdpi.com/2076-3417/14/8/3327>
- [4] G. Santesteban et al., GUÍA DE BUENAS PRÁCTICAS DE PODA DE LA VID INTERACTIVA, Proyecto VITES QUALITAS (POCTEFA EFA324/19), 2021.
- [5] T. Gentilhomme, M. Villamizar, J. Corre, and J. Odobez, “Towards smart pruning: ViNet, a deep-learning approach for grapevine structure estimation,” *Comput. Electron. Agric.*, vol. 207, p. 107736, 2023.
- [6] A. Navone, M. Martini, and M. Chiaberge, “Autonomous Robotic Pruning in Orchards and Vineyards: a Review,” *Smart Agric. Technol.*, vol. 101283, 2025.
- [7] OECD, *Labour and Skills Shortages in the Agro-Food Sector*, Paris, France: Organisation for Economic Co-operation and Development, 2023. [Online] https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/01/labour-and-skills-shortages-in-the-agro-food-sector_c9915f4e/ed758aab-en.pdf
- [8] J. Zhang, H. Yu, and X. Song, “Multi-Dimensional Pruned Sparse Convolution for Efficient 3D Object Detection,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1-10, 2023.

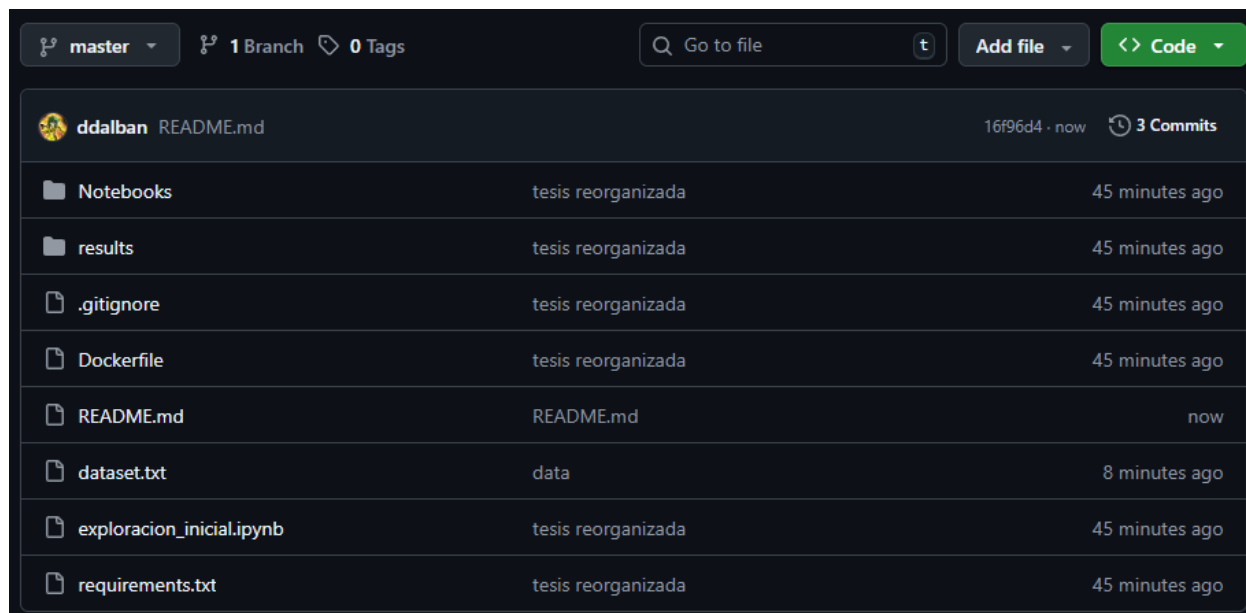
- [9] Wine Scholar Guild, “Vine School Part 1: Common Vine Training Systems,” Wine Scholar Guild Blog, 2021. [Online] <https://www.winescholarguild.com/blog/viticulture-insights/vine-school-part-1-common-vine-training-systems>
- [10] Agroptima, “Guía de poda Royat de la vid,” Blog Agroptima, 2022. [Online] <https://blog.agroptima.com/es/blog/guia-de-poda-royat-de-la-vid/>
- [11] Campo Galego, “El cuidado de la vid III: Tipos de poda,” Campo Galego, 2020. [Online] <https://www.campogalego.es/el-cuidado-de-la-vid-iii-tipos-de-poda/>
- [12] “Winter Pruning of Grapevines – Fact Sheet,” Farming Connect, 2016. [Online] https://businesswales.gov.wales/farmingconnect/sites/default/files/documents/winter_pruning_0.pdf
- [13] “Pruning and Training Principles for Balanced Vines,” UC ANR, Bulletin 118, 2018. [Online] <https://ucanr.edu/sites/default/files/2018-04/281481.pdf>
- [14] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed., Pearson, 2020.
- [15] C. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169917311742>
- [16] S. Casini, P. Ducange, F. Marcelloni, and L. Pollini, “Artificial Intelligence in Agri-Robotics: A Systematic Review of Trends and Emerging Directions,” *Robotics*, vol. 15, 2026. [Online]. Available: <https://www.mdpi.com/2218-6581/15/1/24>
- [17] P. Guadagna et al., “Using deep learning for pruning region detection and plant organ segmentation in dormant spur-pruned grapevines,” *Precision Agriculture*, 2023. [Online]. Available: <https://link.springer.com/article/10.1007/s11119-023-10006-y>

- [18] F. Oliveira et al., “Leveraging State-of-the-Art YOLO Detection Models for 2D Image Analysis in Grapevine Node Detection to Support Pruning Automation,” *Sensors*, vol. 24, no. 4, 2024. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11548613/>
- [19] Z. Cao, S. Sun, and X. Bao, “A Review of Computer Vision and Deep Learning Applications in Crop Growth Management,” *Applied Sciences*, vol. 15, no. 15, 2025. Disponible: <https://www.mdpi.com/2076-3417/15/15/8438>
- [20] Y. Bai and X. Zhang, “Vision-based navigation and guidance for agricultural robots,” *Computers and Electronics in Agriculture*, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0168169922008924>
- [21] J. Corre, «3D2cut Single Guyot Dataset». Zenodo, feb. 27, 2023. doi: 10.34777/azf6-tm83.
- [22] Institut Français de la Vigne et du Vin (IFV), *Fiche technique: Les bonnes pratiques de taille*, 2018. [Online]. Available: https://www.plan-deperissement-vigne.fr/sites/default/files/2018-06/Winetwork-%20les%20bonnes%20pratiques%20de%20taille_0.pdf
- [23] Organisation Internationale de la Vigne et du Vin (OIV), *Viticulture practices for sustainable vineyards*, OIV-VITI 641-2020. [Online]. Available: <https://www.oiv.int/public/medias/7601/oiv-viti-641-2020-en.pdf>
- [24] M. R. Sarker, “Advancing Agriculture Automation Systems: Technological Trends and Applications,” *Advances in Agriculture*, 2025. Available: <https://onlinelibrary.wiley.com/doi/full/10.1155/aia/5518653>
- [25] A. Calvo, “Costes de producción en viñedos: claves para su control,” *Agroptima Blog*, 2025. [Online]. Available: <https://blog.agroptima.com/es/blog/costes-produccion-vinedos/>

- [26] Seed Studio, “reComputer Super J4012 – Advanced Edge AI Computer,” SeedStudio.com.
[Online]. Available: <https://www.seedstudio.com/reComputer-Super-J4012-p-6443.html>
- [27] Arrow Electronics, “NVIDIA Jetson AGX Orin Industrial (Part No. 900-13701-0080-000),”
Arrow.com. [Online]. Available: <https://www.arrow.com/en/products/900-13701-0080-000/nvidia>

APÉNDICE

Repositorio del código: https://github.com/ddalban/Proy_tesis_VPP.git



The screenshot displays the GitHub repository interface for 'ddalban/Proy_tesis_VPP.git'. The repository is on the 'master' branch, with 1 branch and 0 tags. The commit hash is 16f96d4, and it was committed 'now' with 3 commits. The file list includes:

File/Folder	Commit Message	Time
📁 Notebooks	tesis reorganizada	45 minutes ago
📁 results	tesis reorganizada	45 minutes ago
📄 .gitignore	tesis reorganizada	45 minutes ago
📄 Dockerfile	tesis reorganizada	45 minutes ago
📄 README.md	README.md	now
📄 dataset.txt	data	8 minutes ago
📄 exploracion_inicial.ipynb	tesis reorganizada	45 minutes ago
📄 requirements.txt	tesis reorganizada	45 minutes ago