

UNIVERSIDAD NACIONAL SAN ANTONIO ABAD DEL CUSCO
FACULTAD DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA,
INFORMÁTICA Y MECÁNICA
ESCUELA PROFESIONAL DE INGENIERÍA INFORMÁTICA Y DE
SISTEMAS



TESIS

**CONTABILIZACIÓN DE CAMÉLIDOS SUDAMERICANOS
DOMÉSTICOS MEDIANTE VISION COMPUTACIONAL APLICADA
A IMÁGENES OBTENIDAS POR VEHÍCULOS AÉREOS NO
TRIPULADOS**

PRESENTADO POR:

Br. ALEX HELDER HUANCARA
CCOLQQUE
Br. DAVID SONCCO CACHURA

**PARA OPTAR AL TÍTULO PROFESIONAL
DE INGENIERO INFORMÁTICO Y DE
SISTEMAS**

ASESOR:

Dr. RONY VILLAFUERTE SERNA

CO-ASESOR:

Ph.D. CARLOS FERNANDO MONTOYA
CUBAS

**FINANCIADO POR YACHAYNINCHIS
WIÑARINANPAQ – UNSAAC**

CUSCO – PERÚ

2026



Universidad Nacional de San Antonio Abad del Cusco

INFORME DE SIMILITUD

(Aprobado por Resolución Nro.CU-321-2025-UNSAAC)

El que suscribe, el Asesor RONY VILLAFUERTE SERNA
..... quien aplica el software de detección de similitud al
trabajo de investigación/tesis titulada: CONTABILIZACIÓN DE CAMELIDOS SUDAMERICANOS
DOMÉSTICOS MEDIANTE VISION COMPUTACIONAL APLICADA
A IMÁGENES OBTENIDAS POR VEHÍCULOS AÉREOS NO TRIPULADOS

Presentado por: ALEX HELDER HUANCARA COLQUE DNIN° 73473647 ;
presentado por: DAVID SÓNECO CACHURA DNIN°: 74140956
Para optar el título Profesional/Grado Académico de INGENIERO INFORMÁTICO
Y DE SISTEMAS

Informo que el trabajo de investigación ha sido sometido a revisión por 03 veces, mediante el
Software de Similitud, conforme al Art. 6° del **Reglamento para Uso del Sistema Detección de**
Similitud en la UNSAAC y de la evaluación de originalidad se tiene un porcentaje de 1 %.

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No sobrepasa el porcentaje aceptado de similitud.	☒
Del 11 al 30 %	Devolver al usuario para las subsanaciones.	☐
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, conforme al reglamento, quien a su vez eleva el informe al Vicerrectorado de Investigación para que tome las acciones correspondientes; Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	☐

Por tanto, en mi condición de Asesor, firmo el presente informe en señal de conformidad y **adjunto**
las primeras páginas del reporte del Sistema de Detección de Similitud.

Cusco, 17 de agosto de 2026


Firma

Post firma RONY VILLAFUERTE SERNA

Nro. de DNI 23457778

ORCID del Asesor 0000-0003-4607-522X

Se adjunta:

- Reporte generado por el Sistema Antiplagio.
- Enlace del Reporte Generado por el Sistema de Detección de Similitud: oid: 27259:614411024

Alex_David Huancara_Soncco contadorCamelidosV3.pdf

 Universidad Nacional San Antonio Abad del Cusco

Detalles del documento

Identificador de la entrega

trn:oid:::27259:614411024

Fecha de entrega

14 ago 2026, 4:29 p.m. GMT-5

Fecha de descarga

14 ago 2026, 4:37 p.m. GMT-5

Nombre del archivo

contadorCamelidosV3.pdf

Tamaño del archivo

14.4 MB

158 páginas

30.560 palabras

168.158 caracteres

1% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado
- Coincidencias menores (menos de 20 palabras)

Exclusiones

- N.º de coincidencias excluidas

Fuentes principales

- 1%  Fuentes de Internet
- 0%  Publicaciones
- 1%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alerta de integridad para revisión

-  **Caracteres reemplazados**
53 caracteres sospechosos en N.º de páginas
Las letras son intercambiadas por caracteres similares de otro alfabeto.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Dedicatoria

A mis padres y a mi familia por su apoyo incondicional, confianza y acompañamiento constante a lo largo de este camino, brindándome fortaleza en cada etapa; ellos hicieron más llevadero este proceso y contribuyeron a mantenerme enfocado en el logro de mis objetivos.

Alex H.

Con gratitud a mis padres, por su constante apoyo y por impulsarme a seguir adelante; a nuestros amigos productores de camélidos, quienes hicieron posible este proceso y contribuyeron en el logro de nuestros objetivos.

David S.

Agradecimientos

Nuestro más sincero agradecimiento a los profesores que formaron parte de esta investigación, a quienes contribuyeron en mi formación académica y brindaron las bases necesarias para el desarrollo de este proyecto.

También extendemos nuestros agradecimientos a los fundos ganaderos, entre familias y empresas, que facilitaron el acceso a sus instalaciones y permitieron la recolección de datos en condiciones reales, hecho que fue importantísimo para el desarrollo y validación de la investigación: Empresa Comunal de Llamas de Hanccollahua, fundo ganadero “Fundo Pallallani” (Hanocca), productor de alpacas cabaña “k’ori Paccocha” (Phinaya), familia Maccarcco Hilachoque (Hanccollahua), familia Choque Gonzales (Hanccollahua), familia Huillca (Phinaya), Sr. Alejandro Huillca (Phinaya) y al Sr. Mario Panocca (Chisicata).

De igual manera, se agradece al Vicerrectorado de Investigación que, a través del Programa Yachayninchis-Wiñarinpapaq, financió parte de esta investigación en su modalidad de Subvención de Tesis de Pregrado.

Resumen

La contabilización de camélidos sudamericanos domésticos en zonas altoandinas, como la llama y la alpaca, presenta limitaciones debido a las condiciones geográficas y depende enteramente de los métodos manuales; además, representan un área poco estudiada. En la presente investigación se propone la implementación de una solución tecnológica basada en modelos de visión computacional para la contabilización automática de dos especies morfológicamente semejantes (llamas y alpacas) utilizando vehículos aéreos no tripulados. Para ello, se implementó una solución tecnológica que integra modelos de detección de objetos, técnicas de seguimiento multiobjeto y una aplicación móvil capaz de procesar video en tiempo real de forma local sin conexión a internet, con persistencia de datos, siguiendo la metodología de desarrollo Cross-Industry Standard Process for Data Mining (CRISP-DM). Los resultados obtenidos evidencian una exactitud promedio superior al 95 % en condiciones reales de conteo validados por un especialista en camélidos sudamericanos, así como una reducción en el tiempo de conteo con respecto a los métodos tradicionales. Estos resultados demuestran la viabilidad de aplicación del sistema propuesto, que puede ser una alternativa para la gestión ganadera en zonas remotas.

Palabras clave: Camélidos sudamericanos, Visión computacional, UAV, Conteo de camélidos, Procesamiento en tiempo real.

Abstract

The counting of domestic South American camelids in high Andean regions, such as llamas and alpacas, presents significant limitations due to geographic conditions and relies entirely on manual methods. Additionally, this area remains relatively underexplored. This research proposes the implementation of a technological solution based on computer vision models for the automatic counting of two morphologically similar species (llamas and alpacas) using unmanned aerial vehicles (UAVs). To achieve this, a system was developed that integrates object detection models, multi-object tracking techniques, and a mobile application capable of processing real-time video locally without requiring an internet connection, including data persistence, following the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. The results obtained demonstrate an average accuracy exceeding 95 % under real counting conditions, validated by a specialist in South American camelids, as well as a reduction in counting time compared to traditional methods. These findings confirm the feasibility of the proposed system as an effective alternative for livestock management in remote areas.

Keywords: South American Camelids, Computer Vision, UAV, Camelid Counting, Real-Time Processing.

Introducción

El conteo automático de camélidos sudamericanos domésticos, como la llama y la alpaca, representa una actividad esencial para el sustento económico en las zonas altoandinas. Tradicionalmente, el conteo ha sido realizado de forma manual, que conlleva un alto consumo de tiempo, personal encargado y probabilidad de error, generalmente en escenarios con gran cantidad de animales.

Actualmente, la visión computacional junto con los UAVs han demostrado ser herramientas valiosas para procesos de automatización. En ese contexto, esta investigación propone la implementación de una solución tecnológica basada en modelos de detección de objetos, implementación de algoritmos de tracking, y una aplicación móvil capaz de procesar videos provenientes de UAVs, para identificar y contar de forma automática llamas y alpacas, mejorando y optimizando el monitoreo animal.

La presente investigación se encuentra estructurada en varios capítulos que permiten desarrollar de manera progresiva el problema planteado. En el primer capítulo se abordan los aspectos generales de la investigación, incluyendo la descripción y formulación del problema, los objetivos, la justificación, la delimitación del estudio, así como la metodología de investigación utilizada. En el segundo capítulo se presenta el marco teórico, donde se revisan los fundamentos conceptuales relacionados con la visión computacional, la detección de objetos, el uso de vehículos aéreos no tripulados y las técnicas de seguimiento aplicadas al conteo de animales. En el tercer capítulo se describe el desarrollo de la investigación, detallando

la construcción del dataset, el proceso de entrenamiento de los modelos, la comparación de arquitecturas y la implementación del sistema propuesto. En el cuarto capítulo se presentan los resultados obtenidos, incluyendo el análisis del desempeño de los modelos y la validación en condiciones reales de campo. Finalmente, se exponen las conclusiones y recomendaciones derivadas del estudio, así como las posibles líneas de trabajo futuro.

Listado de Abreviaturas

- **CNN:** Convolutional Neural Network (Redes Neuronales Convolucionales).
- **YOLO:** You Only Look Once.
- **IoU:** Intersection over Union (Intersección sobre Unión).
- **SGD:** Stochastic Gradient Descent (Descenso de Gradiente Estocástico).
- **MuSGD:** nuevo optimizador híbrido que combina SGD con Muon. Inspirado en Kimi K2 de Moonshot AI.
- **TP:** True Positive (Verdader positivo).
- **FP:** False Positive (Falso positivo).
- **FN:** False Negative (Falso negativo).
- **mAP:** Mean Average Precision
- **SVM:** Support Vector Machine (Máquina de Vectores de Soporte).
- **SSD:** Single Shot Multibox Detector.
- **R-CNN:** Region-based Convolutional Neural Network (Red Neuronal Convolucional basada en Regiones).
- **Faster R-CNN:** Versión optimizada de R-CNN que utiliza una red de propuesta de regiones (RPN).

- **RPN:** Region Proposal Network (Red de Propuesta de Regiones).
- **UAV:** Unmanned Aerial Vehicle (Vehículo Aéreo No Tripulado).
- **FPS:** Frames per Second (Cuadros por segundo)
- **MIDAGRI:** Ministerio de Desarrollo Agrario y Riego
- **ONU:** Organización de las Naciones Unidas
- **ROI:** Region of Interest (Región de Interés).
- **CVAT:** Computer Vision Annotation Tool
- **CRISP-DM:** Cross-Industry Standard Process for Data Mining (Proceso Estándar de la Industria para la Minería de Datos).
- **MSDK:** Mobile Software Development Kit (Kit de Desarrollo de Software Móvil).
- **Precisión:** Métrica que mide la exactitud de las detecciones positivas.

Índice general

Dedicatoria	II
Agradecimientos	III
Resumen	IV
Abstract	V
Introducción	VI
Listado de Abreviaturas	IX
Índice General	X
Índice de figuras	XV
Índice de tablas	XVIII
1. Aspectos Generales	1
1.1. Planteamiento del Problema	1
1.1.1. Descripción del problema	1
1.1.2. Identificación del problema	3
1.2. Formulación del Problema	4
1.2.1. Problema General	4

1.2.2.	Problemas Específicos	5
1.3.	Objetivos	5
1.3.1.	Objetivo General	5
1.3.2.	Objetivos Específicos	5
1.4.	Justificación	6
1.4.1.	Conveniencia	6
1.4.2.	Relevancia	6
1.4.3.	Implicancias Prácticas	7
1.4.4.	Valor Teórico	7
1.4.5.	Utilidad Metodológica	8
1.5.	Delimitación de estudio	9
1.5.1.	Delimitación Espacial	9
1.5.2.	Delimitación Temporal	9
1.5.3.	Delimitación Operativa y Ambiental	10
1.6.	Limitaciones	10
1.7.	Método de investigación	12
1.7.1.	Enfoque	12
1.7.2.	Tipo	13
1.7.3.	Alcance	13
1.7.4.	Diseño	13
1.7.5.	Metodología de desarrollo	14
1.7.6.	Cronograma de Actividades	17
2.	Marco Teórico	18
2.1.	Antecedentes	18
2.1.1.	Antecedentes Internacionales	18

2.1.2.	Antecedentes Nacionales	26
2.2.	Bases Teóricas	27
2.2.1.	Camélidos sudamericanos	27
2.2.2.	Visión Computacional	34
2.2.3.	Procesamiento de Imágenes	35
2.2.4.	Deep Learning	42
2.2.5.	Detección de Objetos	45
2.2.6.	Modelos de detección de objetos	46
2.2.7.	Métricas de Evaluación	48
2.2.8.	Tracking	52
2.2.9.	Aplicaciones de drones en la Ganadería	52
2.2.10.	Sistemas autonomos	52
2.2.11.	Herramientas	53
3.	Desarrollo del proyecto	56
3.1.	Fase 1: Entendimiento del proyecto	56
3.1.1.	Análisis del problema	56
3.1.2.	Zona de estudio	57
3.1.3.	Requerimientos de hardware	59
3.1.4.	Requerimientos funcionales	60
3.1.5.	Requerimiento no funcionales	61
3.2.	Fase 2: Adquisición de datos	61
3.2.1.	Captura de datos con UAV	62
3.2.2.	Condiciones de captura de datos	65
3.3.	Fase 3: Preparación de datos	66
3.3.1.	Procesamiento de datos	66

3.3.2.	Etiquetado y anotación	71
3.3.3.	División del dataset	75
3.3.4.	Aumento de datos	77
3.4.	Fase 4: Entrenamiento del modelo	78
3.4.1.	Modelos implementados y evaluados	78
3.4.2.	Configuración de hiperparámetros	80
3.4.3.	Proceso de entrenamiento	82
3.5.	Fase 5: Evaluación del modelo	84
3.6.	Fase 6: Implementación de la solución tecnológica	91
3.6.1.	Arquitectura del sistema CamelidCounter	91
3.6.2.	Desarrollo de la Aplicación	93
3.6.3.	Tecnologías utilizadas	93
3.6.4.	Recepción de video MSDK	94
3.6.5.	Procesamiento y conteo en tiempo real	94
3.6.6.	Visualización de resultado	99
3.7.	Fase 7: Despliegue y validación en campo	104
3.7.1.	Condiciones de vuelo	105
3.7.2.	Validación en conteo de alpacas (Comunidad de Phinaya)	105
3.7.3.	Validación en conteo de alpacas (Comunidad de Hanccollahua)	106
3.7.4.	Validación en conteo de llamas (Comunidad de Hanccollahua)	107
3.7.5.	Evaluación del tiempo de conteo	108
4.	Análisis y discusión de resultados	110
4.1.	Análisis de resultados respecto a los objetivos	110
4.2.	Discusión de resultados respecto a los antecedentes	112

4.3. Análisis referencial de inversión y viabilidad económica solución tecnológica de	
conteo automatizado	114
4.3.1. Estimación del costo del sistema	114
4.3.2. Potencial de retorno	115
4.3.3. Aceptación por parte de los productores	115
Conclusiones	116
Recomendaciones	118
Bibliografía	120
Anexos	126
1. Dataset Público en Mendely Data	126
2. Aplicación Android de solución tecnológica	126
3. Registro fotográfico de las zonas de validación en campo	128
4. Documento de validación técnica emitido por el especialista en camélidos sud-	
americanos domésticos.	129
5. Documento de recolección de datos: Cuestionario aplicado a productores . .	137
6. Video de captura de datos y validación en campo	139

Índice de figuras

1.1. Metodología CRISP-DM modificada para el conteo automático de camélidos.	15
2.1. Camélidos sudamericanos	28
2.2. Fenotipo de llamas, Expo Feria Internacional Reyes Espinar 2025	29
2.3. Fenotipo de Alpacas, Expo Feria Internacional Reyes Espinar 2025	30
2.4. Representación gráfica de una imagen digital	37
2.5. Componentes RGB de una imagen a color	38
2.6. Representación de un modelo de Machine Learning	43
2.7. Representación de un modelo de Deep Learning	43
2.8. Hitos en la detección de objetos.	46
3.1. Lugar de estudio: Provincia de Espinar y Canchis - Perú	58
3.2. UAVs utilizados de DJI para la la investigación	60
3.3. Captura de datos a 30 y 70 metros de altura, en la comunidad de Hanccollahua	63
3.4. Ángulo de visión de la cámara a -90 grados	64
3.5. Ángulo de visión de la cámara con inclinación	65
3.6. Extracción de fotogramas de los videos	68
3.7. Redimensionamiento de imágenes a 1280x720	71
3.8. Carga de imágenes para el etiquetado en CVAT	72
3.9. Etiquetado de instancias de llamas en CVAT	73

3.10. Etiquetado de instancias de alpacas en CVAT	74
3.11. División de datos para entrenamiento: 70 % train, 20 %val y 10 % test	76
3.12. Métricas de rendimiento (YOLO11)	86
3.13. Métricas de rendimiento (YOLO26n)	86
3.14. Métricas de rendimiento (Faster R-CNN)	87
3.15. Métricas de rendimiento (SSD)	87
3.16. Matriz de confusión para YOLO11 y YOLO26	88
3.17. Arquitectura de la solución tecnológica	92
3.18. Pipeline de procesamiento y seguimiento en tiempo real	95
3.19. Interfaz principal de la aplicación	100
3.20. Interfaz durante el vuelo	101
3.21. Interfaz de reporte un vuelo	101
3.22. Interfaz de historial de vuelos	102
3.23. Interfaz de modificación de parámetros de conteo	102
3.24. Parámetros de configuración de visualización	103
A.1.1Dataset CAMEL-UAV publicado	126
A.2.1Interfaz de inicio de la Aplicación	127
A.2.2Interfaz durante el vuelo para el conteo de camélidos sudamericanos domésticos	127
A.2.3Interfaz reportes (resultados de conteo)	128
A.3.1Alpacas: Ganadería de la familia Huillca Cochama, Pitumarca, Phinaya, Sec- tor Canlline a 4,800 m s.n.m	128
A.3.2Alpacas: Ganadería de la familia Pedro Maccarcco, Espinar, Hanccollahua, Sector Paccollani a 4,700 m s.n.m.	129
A.3.3Llamas: Ganadería de la familia Choque Gonzales, Espinar, Hanccollahua, Sector Paccollani a 4,700 m s.n.m.	129

A.5.1 Encuesta técnica a productores (Pag. 1 de 2)	137
--	-----

Índice de tablas

3.1. Resultados de la encuesta a productores de camélidos	57
3.2. Zonas de estudio	58
3.3. Cantidad de imágenes obtenidas de -90 grados	68
3.4. Cantidad de imágenes obtenidas de -45 hasta -20 grados	69
3.5. Cantidad de imágenes obtenidas de -45° hasta -20° y -90° grados	69
3.6. Cantidad de imagenes obtenidas en 45° y 90° grados	70
3.7. Dataset 4: Cantidad de anotaciones por clase	75
3.8. Dataset de entrenamiento (70 %)	76
3.9. Dataset de validación (20 %)	76
3.10. Dataset de prueba (10 %)	77
3.11. Hiperparámetros de entrenamiento, optimización y arquitectura.	81
3.12. Hiperparámetros de data augmentation.	82
3.13. Resultados de modelos evaluados con datos de validación	85
3.14. Resultados de modelos evaluados con datos de prueba	85
3.15. Resultados de modelos evaluados con datos de validación	90
3.16. Resultados de conteo de alpacas en Phinaya, pastoreo extensivo	106
3.17. Resultados de conteo de alpacas en Hanccollahua, cerco perimétrico	107
3.18. Resultados de conteo de llamas en Hanccollahua, pastoreo extensivo	108
3.19. Comparación de tiempo de conteo manual y digital	108

4.1. Costo estimado de implementación de la solución tecnológica propuesta . . .	114
--	-----

Capítulo 1

Aspectos Generales

1.1. Planteamiento del Problema

1.1.1. Descripción del problema

El Perú es el mayor productor de camélidos sudamericanos en el mundo. La mayor parte de estas especies habita por encima de los 3,800 metros sobre el nivel del mar, y el país concentra aproximadamente el 87 % de la población mundial de alpacas y el 35 % de la población mundial de llamas (MIDAGRI, 2024c,b,a). Estas especies desempeñan un papel fundamental en los ecosistemas altoandinos, así como en la economía y cultura de las poblaciones locales. Sin embargo, su manejo presenta particularidades importantes, ya que las llamas y alpacas poseen hábitos de pastoreo distintos: las llamas prefieren gramíneas de mayor altura, mientras que las alpacas muestran un comportamiento más selectivo y consumen una mayor proporción de herbáceas (Pezo et al., 2014).

Los camélidos sudamericanos domésticos, como la llama y la alpaca, desempeñan un papel fundamental en los ecosistemas andinos, así como en la economía y la cultura local. Estas especies, aunque comparten características similares en cuanto a tamaño, cuerpo y pelaje, presentan diferencias clave en su comportamiento, adaptación y uso por parte de los cria-

dores. Las alpacas son valoradas principalmente por su lana de alta calidad, utilizada en la producción de artesanías y prendas, mientras que las llamas son empleadas como animales de carga y para la producción de carne. Un ejemplo destacado es el caso de Awana Kancha, ubicado en el Valle Sagrado de los Incas en Cusco, Perú, donde se crían estos camélidos y se muestra el proceso completo de producción de artesanías, además de ofrecer una experiencia turística familiar. Para llevar un control efectivo de estos hatos separados, el manual recomienda un sistema de identificación individual mediante aretes numerados y collares con discos, que permite el registro de peso vivo y producción de fibra por animal (Pezo et al., 2014, p. 116). Sin embargo, el monitoreo y conteo de estas especies se realiza principalmente mediante métodos manuales, los cuales son poco eficientes, costosos en términos de tiempo y personal de apoyo, y propensos a errores por el constante movimiento de estos.

Aunque los documentos específicos sobre llamas y alpacas no siempre describen de manera explícita el procedimiento de conteo manual, la literatura peruana sobre camélidos evidencia que su manejo exige registro, control poblacional e identificación individual. En el caso de las vicuñas, los censos oficiales se realizan mediante conteo visual en campo, lo que constituye una referencia metodológica para entender que el conteo de camélidos en zonas altoandinas es una actividad territorial, manual y apoyada por brigadas de observación (Servicio Nacional de Áreas Naturales Protegidas por el Estado, 2025). En esa misma línea, Ocaña (Ocaña Wilcapi et al., 2026) señala que el censo poblacional de alpacas se desarrolló en campo mediante registros productivos y reproductivos, obteniéndose información individual por animal sobre caravana, sexo, edad y estado fisiológico reproductivo; el estudio tuvo una duración de 90 días.

La problemática se hace más evidente en las zonas de estudio. De acuerdo con la encuesta técnica (Anexo 5) aplicada a algunos productores de camélidos sudamericanos domésticos (llamas y/o alpacas), esta situación se ve reflejada en la experiencia de los propios ganaderos.

Por ejemplo, un productor local de Pitumarca, del fundo Oqemarca, en la provincia de Canchis, señala: “El conteo se realiza de forma manual, con ayuda de 3 a 5 personas para una cantidad de 800 alpacas. Hay margen de error y los animales se mueven constantemente, lo que nos hace perder tiempo y no obtener datos confiables”. Asimismo, los productores manifiestan interés en la incorporación de equipos o soluciones tecnológicas que contribuyan a optimizar el proceso de contabilización.

Como referencia, Peru Flying Labs en colaboración con el Centro Experimental Pacamarca-Fundo Mejoramiento Genético Alpaca (Labs, 2023). Desarrollaron un sistema de visión computacional para el conteo automatizado de alpacas en áreas cercadas, utilizando imágenes aéreas capturadas por un vehículo aéreo no tripulado. Para realizar el conteo, se diseñó un protocolo de verificación, donde los ejemplares fueron liberados individualmente del cerco mientras el vehículo aéreo no tripulado realizaba simultáneamente el conteo automático de la población total, permitiendo así contrastarlo con el número real de especímenes. Aunque esta investigación aún no cuenta con documentación formal publicada, esto representa un avance muy importante para poder automatizar el proceso en las zonas altoandinas.

En consecuencia, el proceso de contabilización demanda una cantidad considerable de tiempo y esfuerzo humano, incrementa la posibilidad de errores por omisión o duplicidad en el conteo y dificulta la disponibilidad de información confiable para apoyar la gestión del hato. Esta situación evidencia la necesidad de comprender las características del proceso actual de contabilización y las dificultades que enfrentan los productores durante su ejecución.

1.1.2. Identificación del problema

La zona de estudio comprende el distrito de Pitumarca (comunidad de Phinaya), el distrito de Marangani (comunidad de Hanocca) y el distrito de Espinar (comunidades de Huarca, Chisicata y Hanccollahua, esta última perteneciente a la Empresa Comunal de Hanccollahua

de Llamas). En estas localidades, la crianza de camélidos constituye una actividad económica importante; sin embargo, la contabilización de los animales se realiza principalmente mediante procedimientos manuales, la ubicación de las áreas de estudio se muestra en la Fig. 3.1. De acuerdo con la encuesta técnica aplicada a los productores de llamas y alpacas, el 100 % realiza el conteo en cerco perimétrico; además, el 40 % indicó que el proceso demanda entre 30 minutos y 1 hora para contabilizar entre 150 y 250 ejemplares, mientras que el 60 % señaló que requiere 2 horas o más cuando se trata de cantidades comprendidas entre 500 y 800 ejemplares, como se muestra en la Tabla. 3.1. Asimismo, todos los encuestados manifestaron en que este procedimiento demanda demasiado tiempo, presentan errores durante su ejecución y que estarían dispuestos a adquirir una solución tecnológica que automatice el proceso.

En las unidades de producción de camélidos sudamericanos domésticos, estas condiciones generan limitaciones que afectan la eficiencia y precisión del proceso de contabilización. La movilidad constante de los animales, las grandes extensiones de pastoreo y la posibilidad de errores por omisión o duplicidad durante el registro dificultan la obtención de información confiable y oportuna para el monitoreo del hato y la toma de decisiones relacionadas con su manejo productivo.

1.2. Formulación del Problema

1.2.1. Problema General

¿Cómo mejorar la contabilización de camélidos sudamericanos domésticos en las comunidades altoandinas, considerando las limitaciones del método manual de conteo?

1.2.2. Problemas Específicos

- ¿Qué factores afectan la eficiencia del proceso manual de contabilización de camélidos sudamericanos domésticos en zonas altoandinas?
- ¿Qué factores afectan la precisión y confiabilidad del conteo manual de camélidos sudamericanos domésticos?
- ¿Cómo puede automatizarse el proceso de contabilización de camélidos sudamericanos domésticos utilizando información obtenida durante el monitoreo del hato?
- ¿En qué medida una solución automatizada puede reducir el tiempo requerido para la contabilización de camélidos sudamericanos domésticos respecto al método manual?

1.3. Objetivos

1.3.1. Objetivo General

Implementar una solución tecnológica basada en visión computacional, aplicada a imágenes de vehículos aéreos no tripulados, para la contabilización automática de camélidos sudamericanos domésticos.

1.3.2. Objetivos Específicos

- Desarrollar un conjunto de datos representativo de camélidos sudamericanos domésticos obtenido mediante imágenes capturadas por vehículos aéreos no tripulados, para el entrenamiento y evaluación de modelos de visión computacional.
- Determinar el modelo de visión computacional que proporcione el mejor desempeño para la detección y contabilización automática de camélidos sudamericanos domésticos.

- Desarrollar una solución tecnológica que integre el modelo de visión computacional con una aplicación móvil conectada a un vehículo aéreo no tripulado para la contabilización automática de camélidos sudamericanos domésticos.
- Evaluar el desempeño de la solución tecnológica mediante la comparación del conteo automático con el método manual en condiciones reales de campo.

1.4. Justificación

1.4.1. Conveniencia

La investigación es conveniente porque optimiza el conteo ganadero a través de la automatización. Al alinear la solución tecnológica con una metodología basada en CRISP-DM, se asegura que esta solución no sea solo un experimento aislado, sino la integración de un UAV, modelos de visión computacional y una aplicación móvil que siguen un proceso estandarizado, desde el entendimiento del problema en el sector auquénido hasta el despliegue funcional en un aplicativo móvil. Esto permite una transición eficiente de los métodos de conteo tradicionales a un sistema de precisión basado en visión computacional. Además los resultados de la encuesta técnica evidencian que el conteo manual continúa siendo la práctica predominante en las unidades productivas, lo que demanda tiempo, esfuerzo humano y presenta errores durante el registro. Esta situación refuerza la conveniencia de desarrollar una solución automatizada que responda a una necesidad real identificada en campo.

1.4.2. Relevancia

El proyecto tiene una alta relevancia social para las comunidades altoandinas cuya economía depende de la crianza de llamas y alpacas. Al implementar una solución que trabaja con drones y procesamiento en tiempo real sin conexión a internet, adecuada para zonas altoandinas

donde la señal es nula o deficiente, se dota a los fundos ganaderos de alta tecnología que reduce el tiempo de conteo y monitoreo de sus especímenes, aplicando visión computacional para resolver problemas reales de sostenibilidad económica de la región. Adicionalmente, la encuesta aplicada a los productores muestra una alta aceptación hacia la incorporación de una herramienta tecnológica para automatizar el conteo, lo que confirma la pertinencia social y productiva de la investigación.

1.4.3. Implicancias Prácticas

Desde una perspectiva práctica, esta tesis resuelve la dificultad de identificar llamas y alpacas debido a su similitud morfológica en tomas aéreas. La implementación de una aplicación móvil conectada a un UAV para la transmisión de video en tiempo real permite una intervención inmediata en el campo. Siguiendo la fase de despliegue de CRISP-DM, la solución se traslada del laboratorio al entorno real, superando las limitaciones de altura, resolución y errores de identificación manual que afectan la precisión del conteo actual.

1.4.4. Valor Teórico

El valor teórico reside en la selección y optimización de los modelos de detección de objetos más recurrentes (como las diversas versiones de YOLO, SSD y Faster R-CNN) adaptados específicamente para la detección de llamas y alpacas. Investigaciones previas como la de (Shao et al., 2020) demostraron que YOLOv2 alcanza un 0.95 en conteo de ganado, pero su rendimiento decae en datasets distintos. Asimismo, (de Lima Weber et al., 2023) logró un mAP de 0.97 con YOLOv5 en entornos controlados, mientras que (Seijas et al., 2019) obtuvo precisiones de 0.86 con YOLOv3 y YOLOv4 en el conteo de ciervos. Esta tesis aporta al adaptar y optimizar modelos más recientes para la implementación de un conteo automatizado en tiempo real, así que no solo se trata de probar un modelo, sino de determinar

cuáles tienen mejor rendimiento en una implementación real para superar la similitud visual de los camélidos en entornos no controlados o zonas de pastoreo de fundos ganaderos. Al basarse en la fase de evaluación de CRISP-DM, la tesis además aporta conocimiento científico mediante la validación en campo entre los resultados por la solución y los métodos manuales ejecutadas por un especialista zootecnista. Adicionalmente los resultados de la encuesta sirven como base para contrastar el desempeño de la solución propuesta frente al método manual, permitiendo evaluar si la automatización realmente mejora la precisión y eficiencia del conteo en condiciones reales de campo.

1.4.5. Utilidad Metodológica

La utilidad metodológica es doble: primero, la construcción de un dataset especializado de imágenes aéreas, lo cual cierra la brecha de datos existente para estas especies; segundo, la aplicación rigurosa de las fases de CRISP-DM para un problema de visión computacional con UAVs. Este enfoque metodológico sirve como una hoja de ruta técnica para futuros investigadores que deseen replicar el proceso de entrenamiento, validación y puesta en marcha de sistemas que busquen integrar drones y modelos de visión computacional en el monitoreo de fauna silvestre o doméstica, garantizando la calidad y la mejora continua del modelo. Asimismo, la información recogida en la encuesta contribuye a definir los requerimientos funcionales del sistema y a orientar la construcción de un proceso metodológico ajustado a las condiciones productivas de las comunidades estudiadas.

1.5. Delimitación de estudio

1.5.1. Delimitación Espacial

La presente investigación se desarrolla en diversas zonas de producción pecuaria altoandina dentro de la región Cusco, seleccionadas por su alta densidad poblacional de camélidos y su representatividad geográfica. El estudio se despliega en las siguientes locaciones:

- **Sector alpacas:** Se trabajó en la comunidad de Phinaya (distrito de Pitumarca, provincia de Canchis) a una altitud de 4800 m s. n. m., específicamente al sector Canlline cabaña K'ori Paccocha. Asimismo, se incluyó la comunidad de Hanocca (distrito de Marangani, provincia de Canchis) con altitudes superiores a los 3800 m s. n. m. Estas zonas presentan un entorno de puna con presencia de bofedales, pastos naturales dominados por el ichu, zonas rocosas y montañas pronunciadas.
- **Sector llamas:** La recolección de datos se centró en la provincia de Espinar, abarcando la comunidad de Hanccollahua (Empresa Comunal de Hanccollahua - Llamas), las comunidades de Huarca y Chisicata (background images), y la Expo Feria Internacional Reyes 2026. Estas locaciones se sitúan en una meseta altoandina a una altitud promedio de 3930 m s. n. m., caracterizada por un relieve predominantemente plano y clima seco.

La diversidad de estas ubicaciones garantiza que el sistema de visión computacional sea evaluado en distintos niveles altitudinales y tipos de terreno, asegurando su aplicabilidad en la accidentada geografía de los Andes.

1.5.2. Delimitación Temporal

El proyecto se desarrolló en un periodo de doce meses, comprendido entre marzo de 2025 y febrero de 2026. Este tiempo abarcó desde la planificación inicial y la recolección de datos

en campo en distintas estaciones del año, procesamiento de datos, hasta el entrenamiento de los modelos de visión computacional, desarrollo del aplicativo móvil, integración del dron con el aplicativo y la validación final del sistema.

1.5.3. Delimitación Operativa y Ambiental

El estudio se limitó a capturas de imagen y video realizadas en un intervalo horario de 10:00 h a 14:00 h para garantizar una iluminación óptima. Asimismo, las pruebas y la formación del dataset consideraron condiciones climáticas variadas, incluyendo días soleados y nublados, con el propósito de mejorar la robustez del modelo frente a variaciones naturales de iluminación.

Desde el punto de vista tecnológico, la solución se delimitó al uso de UAVs para la captura y transmisión de video en tiempo real, así como al procesamiento mediante modelos de detección de objetos basados en la familia YOLO como modelo final para la implementación de la solución tecnológica, siendo estos optimizados para su ejecución en dispositivos móviles.

1.6. Limitaciones

- El sistema se limitó exclusivamente a la detección de las especies llama y alpaca. Si bien se recolectaron imágenes de fondo (background images) que incluyen otras especies con el fin de reducir falsos positivos durante el entrenamiento, el modelo no fue diseñado para la detección de camélidos silvestres, como vicuñas o guanacos, ni de otras especies ganaderas. Adicionalmente, debe considerarse que la producción de llamas y alpacas se realiza de manera separada para evitar el contacto entre poblaciones y preservar sus características productivas y genéticas. Asimismo, la producción de estos camélidos se desarrolla en distintas ubicaciones geográficas descritas en la Sec. 1.5.

- La autonomía de vuelo del dron y la estabilidad de la transmisión de video se vieron condicionadas por factores climáticos presentes en las zonas de Phinaya y Espinar. Condiciones como vientos superiores a los parámetros de seguridad del dron, lluvias, nevadas y variaciones bruscas de temperatura limitaron la realización continua de vuelos y la captura de datos.
- Debido a las limitaciones de hardware de los dispositivos móviles utilizados para la inferencia, se priorizaron modelos de detección con arquitecturas livianas que permitieran equilibrar la precisión con la velocidad de procesamiento (FPS). Esta decisión pudo implicar una reducción potencial en la precisión que podrían alcanzar arquitecturas más complejas ejecutadas en servidores de alto rendimiento.
- La ubicación remota de las comunidades y la topografía accidentada de la zona dificultaron la logística de transporte de equipos y el acceso a una mayor cantidad de rebaños en condiciones de pastoreo extensivo. En consecuencia, la variabilidad del fondo (background) presente en el dataset quedó delimitada principalmente a las zonas geográficas consideradas en el estudio.
- La dificultad para incorporar más imágenes negativas al dataset debido al difícil acceso a las zonas de estudio, lo que impidió realizar viajes recurrentes para recolectar más datos adicionales. Además, no se encontraron estudios actuales que definan una proporción matemática óptima de imágenes negativas, por lo que su selección se realizó de forma referencial según el entorno y el comportamiento del modelo.
- En escenarios donde la dispersión de los camélidos sea excesiva, el campo de visión de la cámara podría verse restringido, requiriendo maniobras de reencuadre del UAV. Esta condición podría afectar la continuidad del seguimiento y generar duplicidad u omisión en el conteo. En consecuencia, el vuelo del UAV se delimitó a terrenos donde

los camélidos no se encuentren excesivamente dispersos, condición que generalmente se presenta entre las 10:00 y las 12:00 horas.

- Otra de las limitaciones del presente estudio fue la ausencia de datos estadísticos confiables sobre el gasto económico real asociado al conteo y manejo de camélidos sudamericanos en las comunidades altoandinas. Esta condición impidió elaborar un análisis de rentabilidad sustentado en información financiera histórica de los productores, por lo que la evaluación económica se restringió a una estimación referencial del costo de inversión del sistema propuesto (ver Tabla 4.1). En consecuencia, los resultados obtenidos deben interpretarse como una aproximación inicial de viabilidad económica y no como una evaluación financiera definitiva.

1.7. Método de investigación

1.7.1. Enfoque

El estudio se desarrolló bajo un enfoque cuantitativo con base en la metodología de Hernández-Sampieri and Mendoza Torres (2018), debido a que se basó en la recopilación, procesamiento y análisis de datos numéricos obtenidos a partir del desempeño de modelos de visión computacional y del sistema implementado. La evaluación del modelo se realizó utilizando métricas cuantificables propias del campo de la detección de objetos, tales como precisión, recall y mAP durante las etapas de entrenamiento y validación.

Asimismo, en las pruebas de campo se efectuó una comparación entre el conteo automático generado por la solución tecnológica y el conteo manual realizado por un especialista zootecnista (validación en campo), obteniendo así indicadores de error de conteo y la diferencia porcentual entre ambos métodos.

1.7.2. Tipo

La presente es una investigación aplicada, ya que se orientó al desarrollo e implementación de una solución tecnológica basada en visión computacional para abordar un problema práctico. Se desarrolló un sistema capaz de realizar la contabilización automática de llamas y alpacas mediante el análisis de imágenes capturadas por UAVs. Este tipo de investigación buscó generar conocimiento con el objetivo de mejorar los métodos tradicionales de conteo utilizados en comunidades altoandinas.

1.7.3. Alcance

De acuerdo con la clasificación metodológica de Hernández-Sampieri and Mendoza Torres (2018), la investigación presenta un alcance descriptivo y explicativo. El estudio inicialmente adopta un alcance descriptivo al recolectar, categorizar y comparar características morfológicas de los camélidos sudamericanos y parámetros de vuelo de UAV para la captura de datos. Posteriormente, la investigación se consolida como un estudio de alcance explicativo al evaluar la relación causa-efecto entre la selección de modelos y el rendimiento de la solución tecnológica. Se demuestra que la elección de la arquitectura afecta directamente la precisión, el tiempo de inferencia y los FPS, para finalmente validar la solución tecnológica en un entorno operativo real comparando el conteo automatizado con el conteo manual.

1.7.4. Diseño

El diseño de la presente investigación fue no experimental, ya que no se manipularon deliberadamente variables independientes ni se establecieron condiciones controladas de experimentación. En su lugar, el estudio se centró en la recolección de datos en entornos reales y en el desarrollo e implementación de una solución tecnológica basada en visión computacional para la detección y conteo de camélidos sudamericanos domésticos.

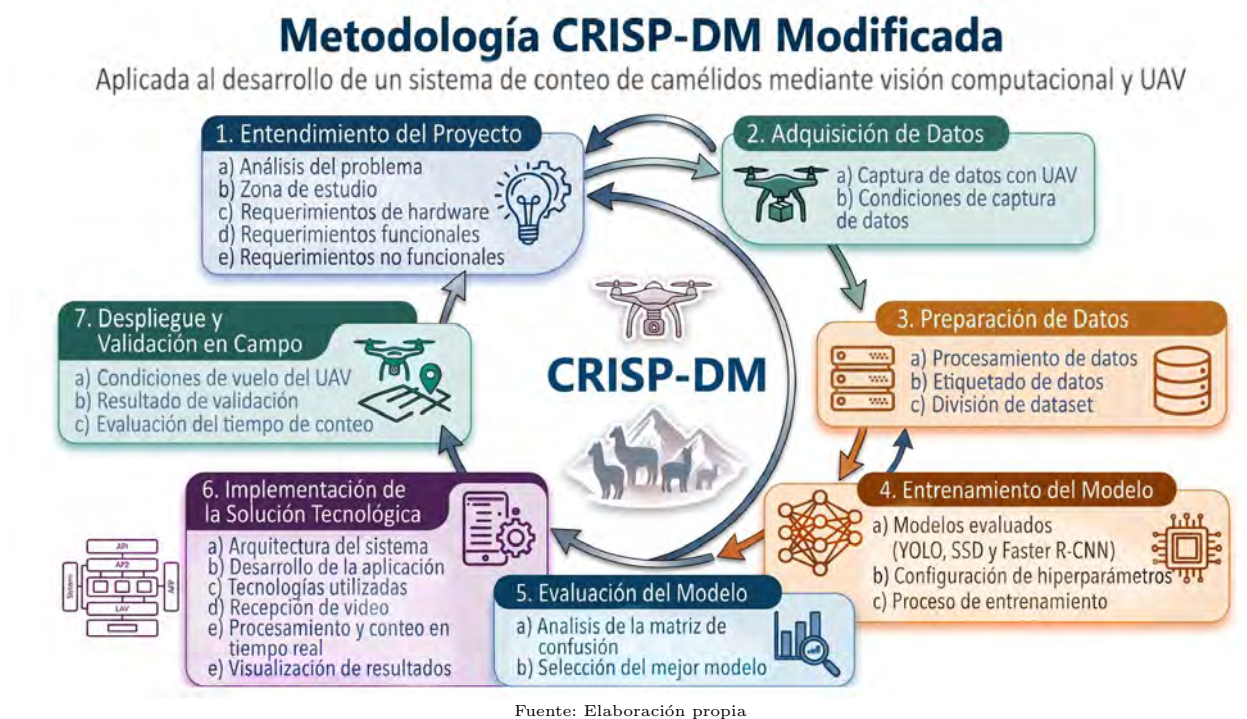
Asimismo, la investigación presentó un carácter longitudinal en la fase de recolección de datos, debido a que las imágenes utilizadas para la construcción del dataset fueron capturadas en diferentes periodos del año y bajo diversas condiciones ambientales. Esta estrategia se incorporó para variar la iluminación, el clima y las características del entorno altoandino, con la finalidad de mejorar la capacidad de generalización de los modelos de detección entrenados. Posteriormente, el sistema desarrollado fue evaluado mediante pruebas de campo utilizando un UAV conectado a una aplicación móvil capaz de procesar video en tiempo real, comparando los resultados del conteo automático con los métodos manuales tradicionales.

1.7.5. Metodología de desarrollo

Para el desarrollo de la solución tecnológica de la presente investigación, se utilizó una modificación de la metodología Cross Industry Standard Process for Data Mining (CRISP-DM), adaptada al contexto de un sistema de visión computacional aplicado a imágenes y videos capturados mediante UAVs. Esta modificación permitió incorporar etapas específicas relacionadas con la implementación de la solución tecnológica y su validación en condiciones reales de campo. La Fig. 1.1 detalla la adaptación de esta metodología.

Figura 1.1

Metodología CRISP-DM modificada para el conteo automático de camélidos.



El proceso en la investigación se estructuró en las siguientes fases:

1) Entendimiento del proyecto

En esta fase se analizó el problema de investigación relacionado con la contabilización manual de camélidos sudamericanos domésticos en zonas altoandinas. Asimismo, se definieron los objetivos del sistema, las coordinaciones con fundos ganaderos, los requerimientos tecnológicos y los aspectos viables en la solución tecnológica.

2) Adquisición de datos

Se realizó la captura de imágenes y videos mediante UAVs en distintas comunidades altoandinas. La recolección de datos se llevó a cabo en diferentes periodos del año y bajo diversas condiciones ambientales, con el objetivo de obtener un conjunto de datos lo más parecido posible a la realidad.

3) Preparación de datos

En esta etapa se realizó el preprocesamiento, la organización, la limpieza y el etiquetado de las imágenes recolectadas. Se generó un dataset estructurado mediante la anotación de los camélidos presentes en las imágenes utilizando *bounding boxes*.

4) Entrenamiento del modelo

Se entrenaron diferentes modelos para la detección de objetos, incluyendo arquitecturas de la familia YOLO, SSD y Faster R-CNN, con el objetivo de identificar la arquitectura más adecuada para el conteo automatizado en tiempo real.

5) Evaluación del modelo

Los modelos entrenados fueron evaluados utilizando métricas de desempeño propias de la detección de objetos, tales como precisión, recall y mean Average Precision (mAP), con el fin de determinar su rendimiento y seleccionar el modelo óptimo en términos de precisión y velocidad de inferencia.

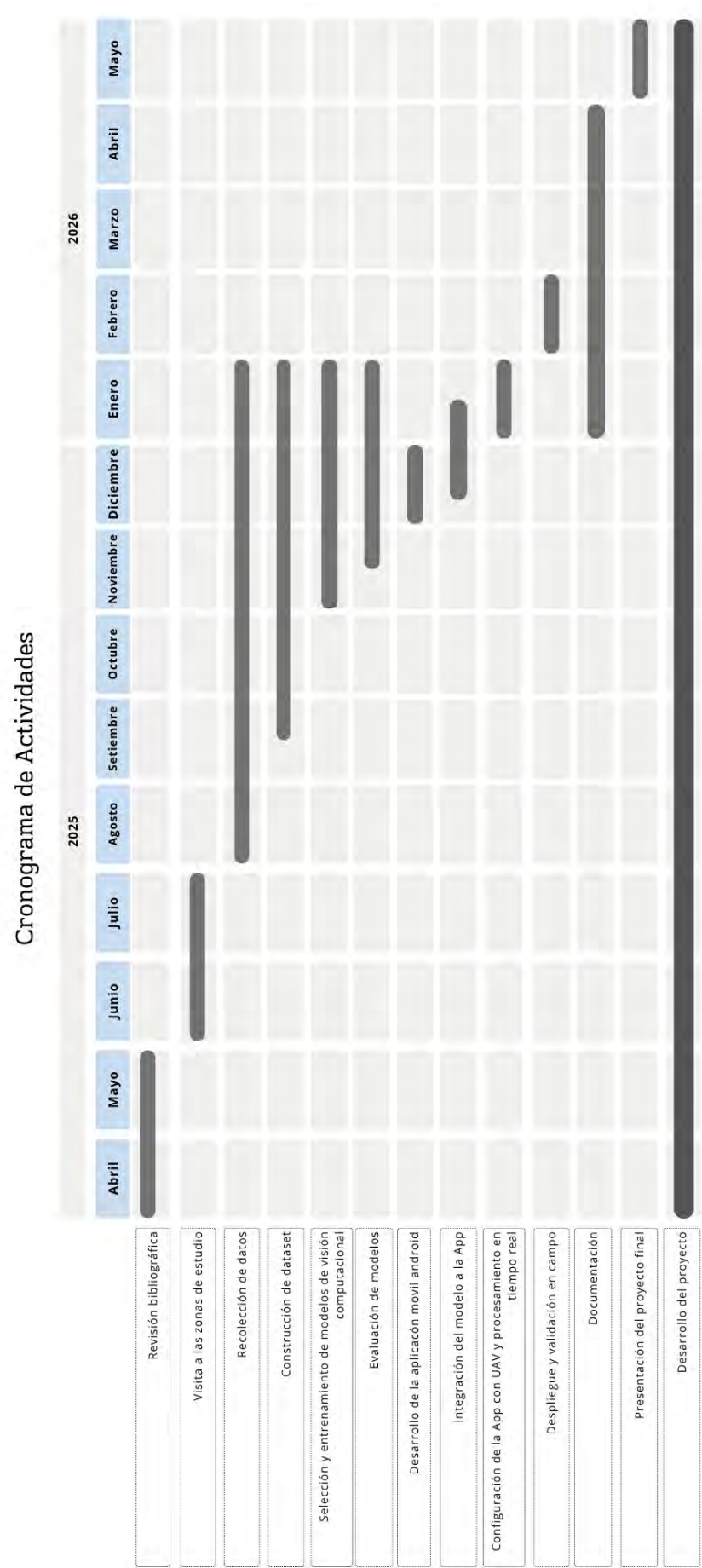
6) Implementación de la solución tecnológica

El modelo seleccionado fue integrado en una aplicación móvil capaz de procesar video en tiempo real proveniente de un UAV, permitiendo la contabilización automática de camélidos durante el vuelo del dron.

7) Despliegue y validación en campo

Finalmente, el sistema desarrollado fue validado mediante pruebas de campo, comparando los resultados del conteo automático con los métodos manuales tradicionales. Esta etapa permitió evaluar el desempeño del sistema en condiciones reales de operación.

1.7.6. Cronograma de Actividades



Capítulo 2

Marco Teórico

2.1. Antecedentes

2.1.1. Antecedentes Internacionales

(Patel, 2023), *A Comprehensive Study on Object Detection Techniques in Unconstrained Environments*, The University of Texas at San Antonio, Estados Unidos.

Conclusiones:

- Se realizó un estudio completo y un análisis profundo sobre los retos asociados a la detección de objetos en entornos sin restricciones.
- Se ofreció una visión general de las técnicas más avanzadas de detección de objetos.
- Se compararon los rendimientos de varios métodos, destacando sus puntos fuertes y débiles, con conjuntos de datos populares.

Comentario: El siguiente trabajo presenta avances recientes en la detección de objetos utilizando redes neuronales convolucionales (CNN), aspecto que es fundamental para este proyecto, ya que se enfoca en los desafíos de entornos no controlados utilizando el data-

set COCO. Además, realiza un análisis comparativo de métodos modernos, destacando sus ventajas y desventajas. Para los métodos de dos etapas, se incluyen R-CNN (Region-based Convolutional Neural Networks) con un 53.3 % de precisión promedio (mAP), Fast R-CNN con un 70.0 % y Faster R-CNN con un 73.2 %. Por otro lado, entre los modelos de una sola etapa, se analizan YOLOv3 (You Only Look Once) con un 57.9 %, SSD (Single Shot Multi-box Detector) con un 72.1 % y RetinaNet con un 74.8 % de mAP. Los resultados muestran que los modelos de dos etapas alcanzan una mayor precisión, pero son más lentos en términos de cuadros por segundo (fps), mientras que los métodos de una sola etapa son más rápidos. Este análisis resulta clave para seleccionar el modelo más adecuado para nuestra tarea de detección de camélidos sudamericanos domésticos en tiempo real.

(Hu et al., 2021), *Influence of Image Quality and Light Consistency on the Performance of Convolutional Neural Networks for Weed Mapping*, Department of Soil and Crop Sciences, Texas A&M University, Estados Unidos.

Conclusiones:

- El estudio demostró que las degradaciones comunes en imágenes, como la reducción de resolución, sobreexposición, desenfoque gaussiano, desenfoque de movimiento y ruido, afectan significativamente el rendimiento de los modelos CNN (Faster R-CNN, Mask R-CNN y Deeplab-v3) en tareas de detección de objetos, segmentación de instancias y segmentación semántica.
- Aunque se intentó minimizar las degradaciones en las imágenes sin procesar, no fue posible eliminarlas por completo. Esto plantea la incógnita de cuál sería el rendimiento máximo de las CNN si se utilizaran imágenes perfectas, sin degradaciones.
- El estudio se centró en el mapeo de malezas y cultivos en etapas tempranas de crecimiento. Queda pendiente evaluar el rendimiento de las CNN en la detección y segmen-

tación de plantas maduras, una tarea que podría presentar desafíos adicionales.

Comentario: En este estudio, se investigó cómo la calidad de las imágenes afecta el rendimiento de los modelos de redes neuronales convolucionales (CNN). Para ello, se utilizaron imágenes obtenidas en la Granja de Investigación AgriLife de la Universidad de Texas A&M en junio de 2020. Un dron equipado con una cámara FUJIFILM GFX 100 RGB capturó imágenes automáticamente durante 24 horas, con una resolución de 100 megapíxeles. A partir de este conjunto de datos, se simuló una variedad de degradaciones comunes, como cambios en la resolución, exposición, desenfoque gaussiano, desenfoque de movimiento y ruido, aplicando diferentes umbrales de alteración. Como resultado, se generaron 1458 imágenes tomadas al mediodía, 500 durante el atardecer y 500 en condiciones nubladas, todas ellas modificadas para evaluar el impacto de la calidad de imagen.

El estudio se centró en la detección de objetos de cinco categorías: algodón, soja, campanilla, pastos y otros (esta última categoría incluye diversas especies de maleza). Para ello, se utilizaron tres modelos CNN ampliamente reconocidos: Faster R-CNN (detección de objetos), Mask R-CNN (segmentación de instancias) y Deeplab-v3 (segmentación semántica). Cada modelo fue sometido a múltiples pruebas bajo las condiciones de degradación simuladas.

Los resultados mostraron que los modelos de segmentación semántica (Deeplab-v3) presentaron un mejor rendimiento frente a las alteraciones de imagen, seguidos por los modelos de detección de objetos (Faster R-CNN). Por otro lado, los modelos de segmentación de instancias (Mask R-CNN) fueron los más afectados, mostrando mayores dificultades en la clasificación de malezas. Estos hallazgos confirman que la calidad de las imágenes tiene un impacto significativo en el rendimiento de los modelos CNN, especialmente en tareas de detección y segmentación en entornos agrícolas.

(Seijas et al., 2019), *Identification of Rodent Species Using Deep Learning*,

Conclusiones:

- El artículo presenta la efectividad de la aplicación de un clasificador construido con la conexión de una CNN pre-entrenada como extractor de características y una SVM como clasificador multiclase, en el desafiante problema de identificación de especies animales.
- La base de datos es de roedores, morfológicamente parecidos, lo cual hace que el problema de discernir entre especies sea más difícil. Los elevados registros de exactitud (error de clasificación mínimo) indican que el desempeño del clasificador desarrollado supera los niveles alcanzados por otros clasificadores con diferentes arquitecturas y metodologías; en ese sentido, lo anterior debe justificarse en el uso de la combinación de CNNs pre-entrenadas de excelente desempeño, junto con las ventajas de usar SVM, que se corresponden con las manifestadas cotidianamente.
- Es la conocida controversia, de comparar redes neuronales multicapa vs SVM; y se habla de ventajas como: ausencia de estimación de capas intermedias y neuronas por capa, ausencia de mínimos locales, etc. El trabajo futuro a desarrollar es construir localizadores de las especies en las imágenes mediante métodos de búsqueda por regiones propuestas u otras estrategias de punta.

Comentario: Se implementó un modelo de clasificación para cuatro especies de roedores utilizando la arquitectura VGG-16 preentrenada, a la cual se le modificaron las tres últimas capas y se incorporó una Máquina de Vectores de Soporte (SVM) para la clasificación de las especies, las cuales presentan similitudes morfológicas. La recolección de datos se realizó mediante cámaras trampa de activación automática, capturando imágenes de los animales en su hábitat natural. El conjunto de datos constó de 1411 imágenes correspondientes a las

cuatro especies, distribuidas en un 80 % para entrenamiento y un 20 % para validación y prueba. Los resultados obtenidos se analizaron por clase, alcanzando precisiones del 96.8 % para la clase *Coiba agouti*, 99.0 % para *Paca*, 96.7 % para *Spinirat* y 99.7 % para *Woodmouse*. Estos resultados son altamente satisfactorios; sin embargo, es importante mencionar que las imágenes utilizadas en el modelo fueron preprocesadas manualmente mediante recortes con bounding box para aislar completamente al roedor como ente principal. Este preprocesamiento podría haber influido significativamente en el desempeño del modelo y es un factor a considerar en futuras implementaciones y mejorar e nuestro proyecto de investigación con modelos más modernos de CNN.

(Rančić et al., 2023), *Animal Detection and Counting from UAV Images Using Convolutional Neural Networks*, University of Novi Sad, Serbia.

Conclusiones:

- Se presentaron modelos de visión por computadora basados en imágenes de UAV que pueden asistir a los humanos en la detección y conteo de animales, una tarea crucial tanto para la conservación ambiental como para aspectos económicos.
- Se entrenaron cuatro modelos de redes neuronales convolucionales para detectar y contar ciervos, realizando un transfer learning y ajustando a datos propios. Las CNN lograron resultados sobresalientes incluso con imágenes de gran tamaño, con una precisión comparable a la detección humana.
- La detección de animales con YOLO alcanzó una precisión del 86 % y un mAP de hasta 70.45 %. La precisión media de los modelos en la configuración óptima fue entre 65-70 %, excepto YOLOv3, que solo obtuvo un 38 %. La mayor precisión, de 70.45 %, se logró con YOLOv4 en la segunda configuración, cuando había un número igual de muestras positivas y negativas. La versión comprimida de YOLO tuvo un mAP

ligeramente menor, de 68.87 %, pero permitió procesos más rápidos de entrenamiento, prueba y conteo. Este modelo detectó 157 de 169 ciervos, superando a los otros modelos en el proceso de conteo.

- La combinación de imágenes de UAV con redes neuronales convolucionales es una técnica prometedora para el monitoreo de fauna silvestre. Se demostró que es posible generalizar la detección a poblaciones fuera del sitio de entrenamiento, y al recolectar más datos de imágenes UAV diversas, el rendimiento de los modelos podría mejorar aún más.

Comentario: El trabajo de investigación tiene un gran valor, ya que se enfoca en el conteo de ciervos de una sola especie, es de suma importancia para nuestro proyecto de tesis. Este estudio sirve como base para desarrollar métodos de detección y conteo de llamas y alpacas, permitiendo la selección y mejora con modelos más actuales. Además, contribuye a justificar el uso de drones, puesto que ellos trabajaron con 3 tipos de drones en distintos escenarios. Además, nos ayuda a fundamentar la parte teórica relacionada con las redes neuronales convolucionales y justifica la implementación.

(de Lima Weber et al., 2023), *Counting cattle in UAV images using convolutional neural network*, Federal University of Mato Grosso do Sul, Brazil.

Conclusiones:

- El modelo YOLOv5-m permite contar ganado Nelore en imágenes de drones de forma rápida y precisa, con resultados de hasta 0.972 en mAP@.5, ayudando a los ganaderos a llevar un control fácil del inventario. Al ser un método automático y barato, los ganaderos pueden revisar cuántos animales hay en un lote con más frecuencia y detectar si faltan o se mezclaron entre grupos. Como no hay que mover o tocar al ganado para contarlos, este sistema también mejora el bienestar de los animales.

- Las imágenes se tomaron solo dos veces, por lo que no incluyen todas las posibles condiciones de luz o clima que podrían afectar el conteo. Solo se usaron imágenes de ganado Nelore, dejando fuera otras razas con diferentes tamaños y colores que podrían cambiar los resultados.
- Las fotos se capturaron a pocas alturas (20, 40, 80 y 100 metros), lo cual limita la variedad de perspectivas para el modelo.

Comentario: El trabajo de investigación se centra en el conteo de vacas de una especie específica y aporta un valor significativo a nuestro proyecto, ya que también emplearemos UAVs para realizar un conteo automático. Sin embargo, este estudio se limita al conteo en imágenes estáticas, es decir, determina la cantidad de vacas presentes en una fotografía. Además, sirve como base para desarrollar e implementar métodos de detección y conteo de llamas y alpacas, facilitando la selección y optimización de modelos más avanzados y actuales. Asimismo, contribuye a justificar el uso de UAVs, proporcionándonos información clave sobre las altitudes de vuelo que debemos considerar para la captura óptima de imágenes y videos.

(Shao et al., 2020), *Cattle detection and counting in UAV images based on convolutional neural networks*, The University of Tokyo, Japan.

Conclusiones:

- Se desarrolló un sistema para detectar y contar ganado usando imágenes de vehículos aéreos no tripulados (UAV). Optimizando la resolución de las imágenes de entrenamiento según la altura fija de vuelo, se logró que el tamaño del ganado coincidiera con una celda del mapa de características, obteniendo una precisión de 0.957, un recall de 0.946 y una medida F1-score de 0.952. El conteo fue efectivo en áreas de 1.5 a 5 hectáreas cuando el movimiento del ganado era mínimo (menos de su longitud

corporal).

- El rendimiento del sistema disminuye al aplicarlo a nuevos conjuntos de datos (Dataset 2), lo que indica la necesidad de hacerlo más robusto. Se propone aumentar los datos de entrenamiento con diferentes fondos y estaciones, y explorar técnicas de adaptación de dominio para mejorar su desempeño en áreas distintas. Además, el sistema enfrenta dificultades al contar animales que se mueven más rápido o al fusionar resultados de diferentes secciones de vuelo debido a desplazamientos entre posiciones.
- El sistema es aplicable a otros animales de movimiento lento, pero requiere mejoras para animales rápidos y para la reidentificación de individuos entre secciones de vuelo. Se sugiere incorporar características específicas de los animales, como patrones de piel o formas corporales, para superar las limitaciones actuales y mejorar la precisión en escenarios más complejos.

Comentario: El trabajo de investigación en esta tesis se usará como un referente técnico para la detección y conteo de llamas y alpacas mediante vehículos aéreos no tripulados (UAV). Su enfoque de optimización de resolución resulta útil para ajustar modelos como YOLOv2 que fue evaluada según el tamaño de las especies objetivo a alturas de vuelo específicas. Asimismo, la fusión de resultados utilizando un modelo tridimensional ofrece una estrategia para gestionar solapamientos en imágenes, mientras que las limitaciones identificadas ante el movimiento de los animales subrayan la necesidad de incorporar un módulo de seguimiento, como DeepSORT, para animales en desplazamiento. Las métricas reportadas (F1-score de 0.952) pueden citarse para respaldar la viabilidad técnica del sistema propuesto. Finalmente, sus recomendaciones, como diversificar los datos de entrenamiento y emplear resoluciones adaptativas, se adaptarán para garantizar la robustez de nuestra solución tecnológica en los entornos variables de las regiones altoandinas.

2.1.2. Antecedentes Nacionales

(Carrasco-Sauñe and Ibarra-Cabrera, 2023), *Determination of the best image recognition algorithm for South American camelids using machine learning*,

Universidad Nacional de Micaela Bastidas de Apurimac, Perú.

Conclusiones:

- Se demuestra que el algoritmo mas eficiente para el reconocimiento de camélidos sudamericanos, es Mask R-CNN con ResNet101 como extractor de características, seguido de YOLOv8 y finalmente SSMD.
- Los algoritmos fueron probados con 146 imágenes de vicuñas tomadas en Apurimac, Perú y 142 imágenes de alpacas extraidas de Kaggle.

Comentario: En la investigación se implementó tres modelos de Deep Learning, YOLOv8, Mask R-CNN con ResNet101 y SSMD con MobileNet 2.0. Se realizó con el propósito de detectar camélidos sudamericanos, específicamente vicuñas y alpacas. Para las vicuñas, se obtuvieron métricas de precisión de 1.00, 1.00 y 0.90, respectivamente. En el caso de las alpacas, la precisión fue de 1.00, 1.00 y 1.00, respectivamente.

La recolección de datos para el entrenamiento se llevó a cabo en Apurimac (Perú), donde se capturaron imágenes exclusivamente de vicuñas. Para las alpacas, se utilizó un dataset público de Kaggle, aunque es importante destacar que este dataset no es exclusivo de alpacas, sino que incluye tanto llamas, alpacas, vicuñas y guanacos, lo cual podría haber afectado el rendimiento del modelo debido a la falta de distinción clara entre estas especies.

Los resultados indican que Mask R-CNN fue el modelo más óptimo para esta tarea, superando a YOLOv8 y SSMD en términos de precisión. Sin embargo, la limitación del dataset de alpacas sugiere que el modelo podría no haber sido entrenado de manera óptima debido a la confusión entre especies similares.

- La investigación puede ser un punto de partida para continuar con la detección de camélidos sudamericanos.
- Destaca la importancia de mejorar y crear un dataset adecuadamente etiquetado de camélidos sudamericanos para entrenar modelos más precisos.
- La investigación se abre a nuevos desafíos como crear un datast de camelidos con imágenes obtenidas con vehículos aéreos no tripulados y entrenar modelos de vision con estas características.

2.2. Bases Teóricas

2.2.1. Camélidos sudamericanos

En la actualidad existen cuatro especies de camélidos sudamericanos; dos de ellas son silvestres: guanaco (*Lama guanicoe*) y la vicuña (*Vicugna vicugna*), y dos domésticos: llama (*Lama glama*) y alpaca (*Vicugna pacos*). Las cuatro especies de camélidos sudamericanos que se muestran en la Fig. 2.1 poseen características comunes, como la presencia de glándulas metatarsianas, labio leporino, utilización de estercoleros o bosteaderos comunes, organización social polígama, ausencia de marcado dimorfismo sexual, ovulación inducida con la producción de una sola cría por parto o estación reproductiva (Wheeler, 1991).

La investigación genética de Barreta et al. (2013) indica la procedencia de los camélidos sudamericanos domésticos, donde se confirma que el origen de las llamas son los guanacos, y que las alpacas tienen un origen mixto de guanacos y vicuñas, así como también un mestizaje de llamas y vicuñas. Las llamas fueron domesticadas mucho antes que las alpacas.

Figura 2.1

Camélidos sudamericanos



(a) *Guanaco, Parque Nacional Los Cardones, Argentina*

Fuente: (Milne, 2025).



(b) *Vicuña, Reserva Nacional de Salinas y Aguada Blanca, Perú*

Fuente: (Arenas, 2021).



(c) *Llamas, Expo Feria Internacional Reyes Espinar 2025*



(d) *Alpacas, Expo Feria Internacional Reyes Espinar 2025*

Características biológicas de las llamas y alpacas

I) Llama: Es el más grande de los camélidos sudamericanos domesticados y comparte numerosas características morfológicas y de comportamiento con su ancestro salvaje, el guanaco. Al igual que esta especie silvestre, cuenta con una amplia distribución geográfica, aunque en la actualidad esta es más reducida en comparación con la que se estimaba antes de la llegada de los europeos. El pelaje de la llama es muy variable, desde el blanco hasta el negro y el marrón, pasando por toda la gama de colores intermedios, con tendencia a presentar manchas de varios colores en un mismo animal. El tamaño y las dimensiones corporales son muy similares a los del guanaco, pudiendo en muchos casos superar los 130 kg de peso vivo,

como producto de la selección realizada para obtener un animal de carga, para la utilización de su carne y de su fibra, según menciona de Lamo (2011).

Existen dos variedades de fenotipos de llamas, las cuales se muestran en la Fig. 2.2:

- **Llama Ch'aku:** También conocida como “llama lanuda”, esta variedad se caracteriza por poseer un vellón denso que cubre la mayor parte de su cuerpo. Su fibra es más abundante en la cabeza y las orejas, aunque las piernas suelen estar menos cubiertas. Ver Fig. 2.2a.
- **Llama K'ara:** Conocida como “llama pelada”, presenta un pelaje más corto y menos denso, con ausencia de fibra en la cara y las extremidades. Esta variedad ha sido tradicionalmente utilizada como animal de carga, más que para la producción de lana. Ver Fig. 2.2b.

Figura 2.2

Fenotipo de llamas, Expo Feria Internacional Reyes Espinar 2025



(a) *Llama K'ara*



(b) *Llama Ch'aku*

II) Alpaca: Son uno de los camélidos más pequeños y comparten muchas características con su ancestro silvestre, la vicuña. Con un peso promedio de 55 a 65 kg y una altura de 80-90 cm, han sido domesticadas principalmente para la producción de fibra, lo que ha impulsado una selección genética hacia pelajes uniformes, especialmente de color blanco, debido a su alta demanda en la industria textil. Biológicamente, están adaptadas a los ecosistemas

altoandinos, sobreviviendo a altitudes de hasta 5000 msnm, donde enfrentan temperaturas extremas y recursos alimenticios limitados. Su metabolismo eficiente les permite alimentarse de gramíneas y hierbas, siendo herbívoras selectivas. Además, son resistentes a condiciones climáticas adversas y pueden tolerar largos períodos sin agua, aunque normalmente beben diariamente. Su reproducción incluye ovulación inducida y un período de gestación de 342 a 345 días, con una cría por parto de Lamo (2011).

Existen dos variedades de fenotipos de alpacas, las cuales se muestran en la Fig. 2.3:

- **Alpaca Huacaya:** Es la variedad más abundante y se distingue por su fibra densa, rizada y esponjosa, que cubre todo su cuerpo, incluidas las piernas, la frente y las mejillas. En algunos casos, el pelo de la cabeza forma un copete que puede cubrir los ojos. Ver Fig. 2.3a.
- **Alpaca Suri:** Se caracteriza por una fibra larga, sedosa y lisa, que crece en mechones y cae desde la línea media del lomo hacia los lados. Esta estructura hace que su pelaje luzca más brillante y con mayor movimiento en comparación con la variedad Huacaya. Ver Fig. 2.3b.

Figura 2.3

Fenotipo de Alpacas, Expo Feria Internacional Reyes Espinar 2025



(a) Alpaca Huacaya



(b) Alpaca Suri

Hábitat natural y distribución

Hábitat: Los camélidos sudamericanos se distribuyen principalmente a lo largo de la Cordillera de los Andes, desde Ecuador hasta la Tierra del Fuego, con mayor concentración en el altiplano peruano-boliviano y en el norte de Chile y Argentina, habitando altitudes entre 3600 y 5000 metros MIDAGRI (2023).

Distribución: Perú es el principal productor mundial, con más de 5 millones de cabezas, de las cuales 3.6 millones son alpacas, lo que representa más del 85 % de la población global. Sin embargo, la crianza de alpacas ya no se limita a Sudamérica, ya que desde la década de 1980 se ha expandido a países como Estados Unidos, Australia, Nueva Zelanda y Canadá, donde se mantiene un creciente interés por su producción. A nivel global, la población de alpacas supera los 4 millones, con Perú liderando la cifra, seguido por Bolivia y otros países con menor presencia MIDAGRI (2023).

Impacto cultural, económico y ambiental de los camélidos sudamericanos

Los camélidos sudamericanos, entre los que se encuentran las llamas, alpacas, vicuñas y guanacos, desempeñan un papel crucial en la vida de millones de familias que habitan ecosistemas extremos, contribuyendo significativamente a la seguridad alimentaria, la reducción de la pobreza y el fortalecimiento de las comunidades rurales mediante la creación de empleos sostenibles. Estas especies no solo son un pilar económico, sino que también forman parte integral de la identidad cultural de los pueblos indígenas, preservando conocimientos ancestrales y una valiosa biodiversidad genética. Su aprovechamiento proporciona recursos esenciales como proteínas, fibras textiles de alta calidad y fertilizantes, además de servir como medio de transporte en las regiones andinas. Asimismo, los camélidos juegan un papel vital en la mitigación del cambio climático, particularmente en zonas áridas y semiáridas. En reconocimiento a su importancia multidimensional, la Organización de las Naciones Unidas

(ONU) declaró el 2024 como el “Año Internacional de los Camélidos”, con el objetivo de promover la concienciación global sobre su valor económico, social y ambiental (ONU, 2024).

Conservación y amenazas

Para su conservación se han estado implementando programas como el desarrollo de rebaños selectos (mejoramiento genético), la mejora de la producción y la asistencia técnica a los criadores. En el caso de la vicuña, las Comunidades Campesinas han logrado beneficiarse de la comercialización de su fibra bajo la supervisión del Estado, promoviendo un uso sostenible. Además, existen proyectos de repoblamiento para expandir su distribución. La crianza de camélidos sudamericanos en las zonas altoandinas es clave para su economía rural, representando del 70 al 80 % del ingreso anual de más de 150 mil familias en condiciones de pobreza, lo que resalta su importancia social y económica (MIDAGRI, 2021).

Las principales amenazas o riesgos de los camélidos sudamericanos incluyen la erosión genética, especialmente en las variedades de alpaca Suri y llama K'ara y Chaku, que requieren mayor atención para evitar su extinción. La falta de programas adecuados de mejora genética y conservación podría reducir la diversidad y la calidad de la fibra. Por otro lado, en cuanto a las poblaciones silvestres, la vicuña fue declarada en peligro de extinción en 1969, aunque hoy en día, gracias a los proyectos de conservación, ha mejorado su situación a “vulnerable” (MIDAGRI, 2021).

Conservación de fauna y tecnología

Según BioScience (2021), los UAVs se han consolidado como una de las herramientas fundamentales para la conservación de hábitats naturales, ya que permiten acceder a áreas remotas o de difícil acceso, facilitando la recolección de datos sin exponer al personal de campo a riesgos innecesarios. Además, estos dispositivos posibilitan el monitoreo de extensas áreas en períodos reducidos de tiempo, proporcionando una cobertura eficiente y una perspecti-

va aérea única que enriquece el análisis de la distribución y el estado de los ecosistemas. Su implementación también reduce la dependencia de vuelos tripulados, incrementando la seguridad y disminuyendo los costos operativos en comparación con tecnologías alternativas, como los satélites. Estas ventajas convierten a los drones en una solución versátil y económicamente viable para la vigilancia y protección de especies y hábitats naturales.

Aplicaciones de detección de objetos en diferentes áreas

- **Videovigilancia y Seguridad:** La videovigilancia es una herramienta con múltiples aplicaciones en ámbitos públicos y privados, como la seguridad nacional, la prevención de delitos, el control de tráfico, la detección de accidentes, el cuidado de pacientes, ancianos y niños en el hogar. Esta tecnología se ha convertido en un área clave de investigación dentro de la visión por computadora, con el objetivo de extraer información útil de manera eficiente a partir de grandes volúmenes de video capturados por cámaras de seguridad (Sánchez et al., 2020).
- **Transporte:** En el sector del transporte, la detección de objetos es crucial para los sistemas de conducción autónoma, donde identifica peatones, vehículos y señales de tráfico para garantizar la seguridad vial. También se utiliza en la gestión del tráfico para monitorear el flujo de vehículos y detectar incidentes como accidentes o congestiones. En logística, ayuda a clasificar y rastrear paquetes en centros de distribución, optimizando la eficiencia operativa.
- **Medicina:** En el área de la medicina, la detección de objetos se utiliza para analizar imágenes diagnósticas, como radiografías, resonancias magnéticas y tomografías, con el fin de identificar anomalías, tumores u otras estructuras relevantes. También es fundamental en cirugías asistidas por robots, donde ayuda a localizar órganos o tejidos específicos durante procedimientos quirúrgicos. Además, se emplea en el monitoreo de

pacientes para detectar movimientos o comportamientos anómalos en entornos clínicos.

- **Agricultura y Ganadería:** En esta área, la detección de objetos se utiliza para el conteo y monitoreo de animales, asegurando su bienestar y optimizando su manejo. Además, permite identificar y contar especies consideradas plagas, como roedores o insectos, para implementar medidas de control más efectivas.
- **Tecnología de drones y satélites:** Los drones equipados con cámaras y algoritmos de detección de objetos pueden cubrir grandes áreas para identificar y contar animales, especialmente en terrenos extensos o de difícil acceso. El análisis de imágenes satelitales también permite detectar manadas o grupos de animales en regiones remotas, como desiertos o tundras.

2.2.2. Visión Computacional

Según Szeliski (2022), la visión computacional es una disciplina que desarrolla técnicas matemáticas para recuperar la forma tridimensional y la apariencia de los objetos en las imágenes. Básicamente, busca realizar el proceso inverso de los gráficos por computadora, describir el mundo que vemos en una o más imágenes y reconstruir sus propiedades como la forma, la iluminación y las distribuciones de color para tener una comprensión completa de la escena. El autor destaca que la visión es difícil porque se trata de un problema inverso, en el cual se busca recuperar incógnitas a partir de información insuficiente para especificar completamente la solución. Por esta razón, la tarea principal consiste en intentar invertir el proceso de adquisición de imágenes para obtener la mejor descripción posible de la escena, recurriendo a modelos probabilísticos, físicos o al aprendizaje automático para encontrar soluciones.

Desafíos de Visión Computacional en Entornos Reales

Se pueden identificar las siguientes dificultades según el texto leído del autor Szeliski (2022):

- **Naturaleza de problema inverso:** Su objetivo es recuperar incógnitas (como la forma tridimensional, la iluminación y el color) a partir de imágenes que normalmente no contienen la información suficiente para especificar una solución única.
- **Alta complejidad del mundo real:** Modelar el mundo visual con toda su complejidad es extremadamente difícil. Mientras que nosotros los seres humanos y los animales interpretamos estas variaciones sin esfuerzo, los algoritmos informáticos son muy propensos a cometer errores.
- **Desfase entre modelos matemáticos y la realidad:** Es muy común formular descripciones matemáticas que, en la práctica, no coinciden con las condiciones realistas del mundo real o que no permiten una estimación estable de las incógnitas que se buscan resolver.
- **Sensibilidad al ruido y recursos limitados:** Los algoritmos deben ser robustos frente al ruido y tolerar las desviaciones que ocurren respecto a los modelos teóricos. Además, deben lograr esto siendo razonablemente eficientes en el uso de recursos espaciales y tiempos de ejecución.
- **Dificultad de validación:** Es muy fácil crear un algoritmo que funcione de forma aceptable en unas pocas imágenes de prueba, pero que falle en la práctica. No basta con usar datos limpios, el algoritmo debe someterse a ruido y probarse en una amplia variedad de datos del mundo real para garantizar que puede manejar la complejidad de imágenes que no se ajustan a suposiciones simplificadas.

2.2.3. Procesamiento de Imágenes

Es un conjunto de operaciones matemáticas que transforman los valores de los píxeles de una o más imágenes de entrada para generar una nueva imagen de salida. Podría entenderse como

la primera etapa fundamental en los algoritmos de visión computacional, cuyo objetivo es preparar, limpiar o adecuar la imagen capturada para facilitar su posterior análisis de manera más eficiente.

Representación de imágenes digitales

Matemáticamente se representa como una función bidimensional $f(x, y)$, donde x e y son coordenadas espaciales y la amplitud de f en cualquier par de coordenadas representa la intensidad (o nivel de gris) en ese punto. Cuando las coordenadas espaciales y los valores de intensidad son cantidades finitas y discretas, a esta función se le llama imagen digital, ver Fig. 2.4. Para crear una imagen digital a partir de datos continuos obtenidos por sensores, se requieren dos procesos fundamentales: el muestreo (digitalizar los valores de las coordenadas) y la cuantificación (digitalización de los valores de amplitud o intensidad). Computacionalmente, el formato preferido para procesar una imagen digital es representarla como una matriz (Gonzalez and Woods, 2018).

Imagen digital: $f(x, y)$ (representación matricial)

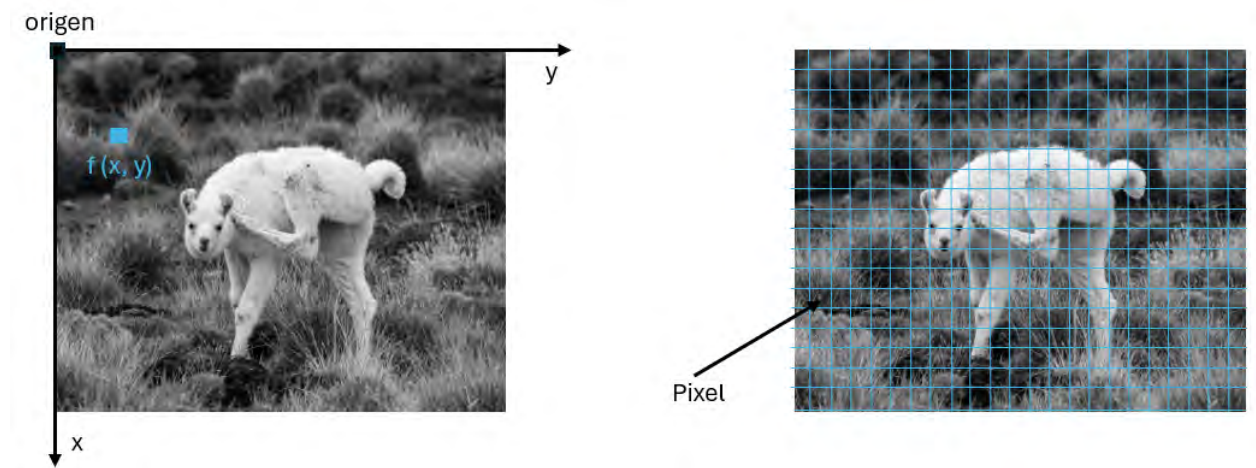
x y y : Coordenadas espaciales

$f(x, y)$: Nivel de gris en la imagen en el punto (x, y)

$$f(x, y) \approx \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, N-1) \\ f(1, 0) & f(1, 1) & \dots & f(1, N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1, 0) & f(M-1, 1) & \dots & f(M-1, N-1) \end{bmatrix} \quad (2.1)$$

Figura 2.4

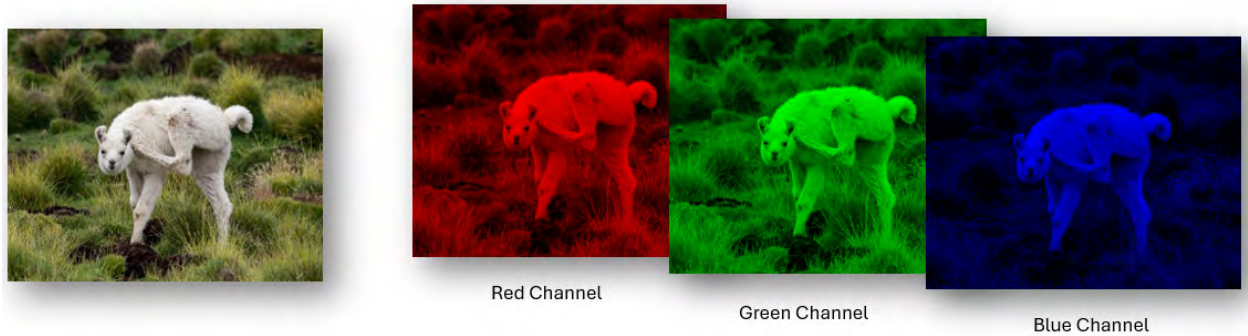
Representación gráfica de una imagen digital



- **Píxeles:** Una imagen digital está compuesta por un número finito de elementos discretos, cada uno ocupando una ubicación particular y poseyendo un valor numérico. A los elementos individuales que conforman el arreglo o matriz matemática de la imagen se les conoce como elementos de imagen o píxeles.
- **Canales (RGB):** En el procesamiento de imágenes en color, uno de los sistemas más utilizados es el modelo RGB, en el que cada color se representa mediante sus componentes espectrales primarios: rojo, verde y azul. Una imagen de este tipo está conformada por tres imágenes de componentes o canales del mismo tamaño, una para cada color primario. Dado que la imagen tiene tres canales, cada píxel de una imagen RGB es en realidad un triplete de valores espaciales (R,G,B) que funciona como un vector tridimensional, ver Fig. 2.5. En las pantallas, estos tres canales se combinan de forma aditiva para producir la ilusión de una sola imagen a todo color.

Figura 2.5

Componentes RGB de una imagen a color



- **Resolución:** Esta se divide principalmente en dos métricas fundamentales que afectan cómo se visualiza el contenido:

- Resolución espacial: Es una medida cuantitativa del detalle más pequeño que se puede discernir en una imagen. Comúnmente se mide a través de los puntos (o píxeles) por unidad de distancia, como por ejemplo la medida de puntos por pulgada (dpi) utilizada en imprenta. Cabe destacar que las dimensiones de la imagen (por ejemplo, 1024×1024 píxeles) no tienen significado completo como resolución espacial a menos que se especifique el tamaño físico real que abarca la imagen.

$$M = 2^m$$

$$N = 2^n$$

m: Número de bits para representar a las filas.

n: Número de bits para representar a las columnas.

M: Número de filas de la imagen.

N: Número de columnas de la imagen.

- Resolución de intensidad: Se refiere al cambio discernible más pequeño posible en el nivel de intensidad. En la práctica, depende de cuántos bits se utilizan para la cuantificación. Por ejemplo, es habitual referirse a una imagen donde la intensi-

dad se cuantifica en 256 niveles uniformes (gracias a 8 bits de almacenamiento) como una imagen que tiene “8 bits de resolución de intensidad”. La falta de una buena resolución de intensidad puede generar artefactos visuales perjudiciales en la imagen, como el falso contorno.

$$L = 2^k$$

L: Cuantización

k: Número de bits

Preprocesamiento de imágenes

Redimensionamiento

El redimensionamiento es el proceso de cambiar la resolución espacial de una imagen digital. Según Gonzalez and Woods (2018), esta operación es esencial para estandarizar las dimensiones de entrada de las Redes Neuronales Convolucionales, que operan con tensores de tamaño fijo.

La interpolación es el método para asignar valores a los nuevos píxeles tras el escalado. La interpolación bilineal utiliza los cuatro vecinos más cercanos para calcular un promedio ponderado, mientras que la bicúbica emplea 16 vecinos, ofreciendo una mayor suavidad y preservación de bordes en objetos pequeños, como camélidos capturados a gran distancia por un UAV (Szeliski, 2022).

Normalización

Es una transformación que ajusta el rango de valores de los píxeles, que de forma típica son de $[0, 255]$ en imágenes de 8 bits, a un nuevo dominio, usualmente $[0, 1]$ o $[-1, 1]$. De acuerdo con Szeliski (2022), la normalización ayuda a que el algoritmo de optimización (descenso de gradiente) converja más rápido. Al estandarizar la escala de los datos, se evita que las funciones de activación de la red neuronal se saturen y se asegura que todas las características

de la imagen contribuyan equitativamente al aprendizaje del modelo.

Conversión de formatos

Las imágenes digitales se representan en diversos modelos matemáticos de color. El estándar en cámaras de drones es el modelo RGB (Red, Green, Blue). En Gonzalez and Woods (2018) definen los espacios de color como sistemas para especificar colores de manera estándar. En visión computacional, la conversión de formatos como de BGR a RGB es vital, ya que la interpretación incorrecta de los canales cromáticos altera las características de textura y color que el modelo utiliza para diferenciar especies como el color de la fibra de las llamas y alpacas.

Mejora de calidad

La mejora de calidad comprende técnicas para acentuar detalles específicos de la imagen o reducir el ruido visual. Una técnica fundamental es la ecualización del histograma, que busca distribuir uniformemente las intensidades de los píxeles para mejorar el contraste global. Según Gonzalez and Woods (2018), esto es particularmente útil para entornos reales; por ejemplo, en entornos altoandinos donde la alta reflectancia solar o las sombras densas pueden ocultar detalles morfológicos de los animales, dificultando su detección.

Aumento de datos

El aumento de datos se define como un conjunto de técnicas empleadas para incrementar de forma artificial la diversidad de un conjunto de datos de entrenamiento mediante la aplicación de transformaciones aleatorias a las imágenes originales. Según Szeliski (2022), estas operaciones permiten generar nuevas variantes de una imagen manteniendo la integridad de las etiquetas de clase, esto permite que el modelo aprenda representaciones más robustas de los objetos de interés.

Importancia de aumento de datos en Deep Learning

En el contexto del aprendizaje profundo, el aumento de datos actúa como una técnica de regularización, especialmente cuando no se tienen suficientes datos para realizar un determinado trabajo de investigación o los datos son difíciles de conseguir. Su importancia radica en mitigar el sobreajuste (overfitting), un fenómeno donde el modelo memoriza el ruido de los datos de entrenamiento pero falla al generalizar en datos nuevos. De acuerdo con la documentación de Jocher et al. (2023) y Rančić et al. (2023), el uso de aumentos permite que el modelo sea capaz de realizar inferencias de mejor forma en condiciones no vistas durante el entrenamiento, mejorando la capacidad de generalización ante variaciones de iluminación, escala y oclusión. Este aumento actúa sobre cuatro aspectos importantes:

- **Conjunto de datos ampliado:** Al crear variaciones de las imágenes existentes, se puede aumentar eficazmente el tamaño del conjunto de datos de entrenamiento sin recopilar nuevos datos.
- **Generalización mejorada:** Los modelos aprenden a reconocer objetos en diversas condiciones, lo que los hace más robustos en aplicaciones del mundo real.
- **Reducción del sobreajuste:** Al introducir variabilidad en los datos de entrenamiento, es menos probable que los modelos memoricen características específicas de las imágenes.
- **Rendimiento mejorado:** Los modelos entrenados con un aumento adecuado suelen lograr una mejor precisión en los conjuntos de validación y prueba.

Tipos de técnicas de aumento de datos

Transformaciones geométricas: Estas técnicas modifican la geometría de la imagen para que el sistema sea invariante a la posición y perspectiva del objeto: como el *reflejo*

(invierte la imagen horizontal o verticalmente), *rotación* (aplica giros aleatorios en un rango de grados determinado) y *escalado* (ajusta el tamaño del objeto en la imagen).

Transformaciones fotométricas: Modificaciones de tono, la saturación, valor de brillo, desenfoque y ruido que se introducen distorsiones visuales a la imagen.

Técnicas avanzadas de aumentación: Los modelos modernos como YOLO introducen métodos de composición que elevan el rendimiento en escenarios de alta densidad: *Mosaic Augmentation* es la técnica distintiva de YOLO que combina cuatro imágenes de entrenamiento en una sola, y *Mixup* superpone dos imágenes distintas con diferentes niveles de opacidad (Jocher et al., 2023).

2.2.4. Deep Learning

El aprendizaje profundo es un subárea del aprendizaje automático que se centra en algoritmos basados en la estructura y funciones del cerebro, conocidos como redes neuronales artificiales que se adaptan a partir de grandes cantidades de datos. A diferencia del Machine Learning esta herramienta extrae las características relevantes de forma automática directamente de la base datos (Ortega Candel, 2025). En la Fig. 2.6 y Fig. 2.7 se puede observar la diferencia entre ambos enfoques. Este enfoque DL se fundamenta en la evolución de los modelos de redes neuronales, desde el perceptrón inicial hasta los perceptrones multicapa, junto con la incorporación de algoritmos de retropropagación que permiten a las redes ajustar sus parámetros mediante el aprendizaje basado en el error (Science Direct, 2022).

Figura 2.6

Representación de un modelo de Machine Learning

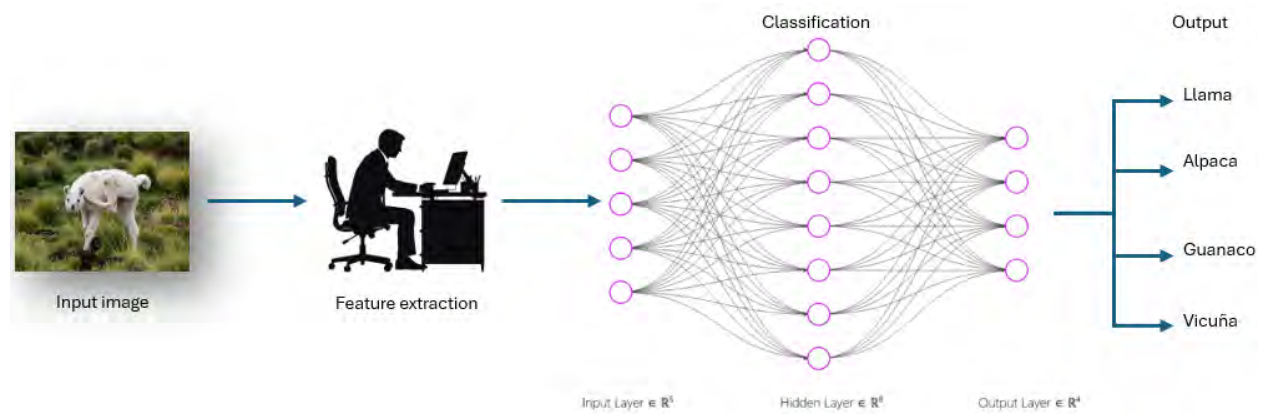
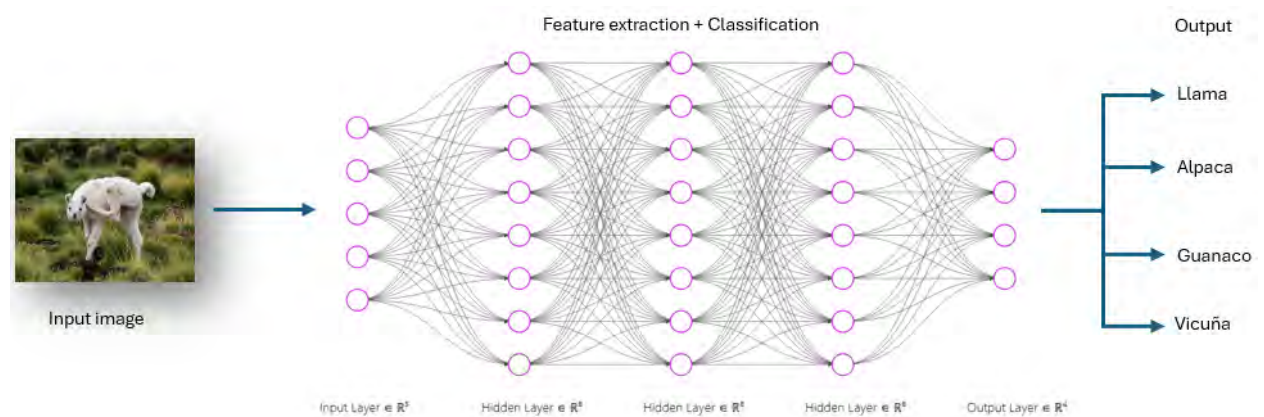


Figura 2.7

Representación de un modelo de Deep Learning



Redes Neuronales Aritificiales

Las Redes Neuronales Artificiales (ANN) son modelos computacionales inspirados en la estructura biológica del cerebro. De acuerdo con Szeliski (2022), una ANN está compuesta por capas de unidades de procesamiento llamadas neuronas, interconectadas mediante pesos sinápticos. Donde cada capa realiza una combinación lineal de sus entradas seguida de una función de activación no lineal. Este proceso permite al modelo aproximar funciones matemáticas complejas, mapeando los datos de entrada hacia una salida específica, como la clasificación de un objeto.

Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales (CNN) son una variante especializada de las ANN. Estas presentan más de una capa oculta y están diseñadas específicamente para procesar datos con estructura de rejilla (ver Fig. 2.4), como las imágenes. Gonzalez and Woods (2018) señala que estas redes utilizan una operación matemática denominada convolución en lugar de multiplicaciones matriciales generales en al menos una de sus capas.

- **Capas Convolucionales:** Utilizan filtros o kernels que se desplazan sobre la imagen para detectar patrones locales como bordes, texturas y formas.
- **Capas de Agrupamiento (Pooling):** Reducen la dimensionalidad espacial de los datos, permitiendo que el modelo sea invariante a pequeñas traslaciones del objeto en la imagen (Szeliski, 2022). Esta arquitectura es la base fundamental de modelos de detección como YOLO.
- **Capas totalmente conectadas:** Interpretan las características extraídas para generar una salida, como la clasificación de una imagen. Las CNN han demostrado un desempeño excepcional en tareas de visión por computadora, como clasificación, detección de objetos y segmentación de imágenes (Wu, 2017).

Transferencia de Aprendizaje

El aprendizaje por transferencia es una técnica donde un modelo entrenado para una tarea inicial se reutiliza como punto de partida para una segunda tarea relacionada. Según Ortega Candel (2025), en lugar de entrenar una red desde cero, proceso que requiere millones de imágenes y gran capacidad de cómputo, se utilizan pesos preentrenados en grandes datasets. La ventaja de esta metodología es que permite que el modelo aproveche el conocimiento previo sobre formas y colores básicos, ajustando únicamente las capas finales (fine-tuning)

para reconocer objetos específicos (Szeliski, 2022). En nuestro caso, esos objetos son los camélidos sudamericanos domésticos en entornos altoandinos. Esto optimiza drásticamente el tiempo de entrenamiento y mejora la precisión en datasets limitados.

2.2.5. Detección de Objetos

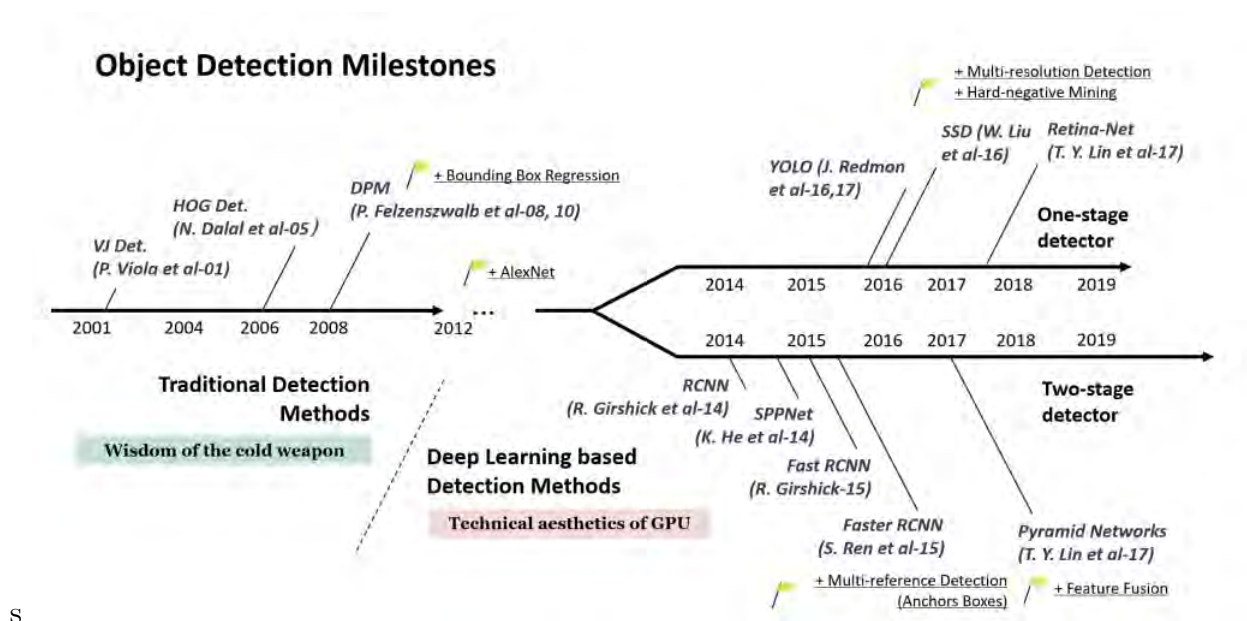
La detección de objetos es una tarea fundamental en el campo de la visión por computadora, está enfocada en identificar y localizar instancias de objetos pertenecientes a categorías específicas (como personas, animales o vehículos) dentro de imágenes digitales. Su objetivo principal es crear modelos y métodos computacionales capaces de responder a una de las preguntas esenciales en este ámbito: ¿Qué objetos están presentes y en qué ubicación? Como uno de los desafíos centrales de la visión por computadora, la detección de objetos sirve como base para otras tareas relacionadas, como la segmentación de instancias, la generación de descripciones de imágenes y el seguimiento de objetos, entre otras. Además en los últimos años, el avance de las técnicas de aprendizaje profundo ha revolucionado este campo, logrando progresos significativos y posicionándolo como un área de investigación de gran relevancia. (Zou et al., 2023). Ver Fig. 2.8 sobre el avance de modelos de detección de objetos.

Cuadros delimitadores (Bounding boxes)

Los bounding boxes son imprescindible en tareas de detección de objetos, estas se definen como una región rectangular delimitada por un conjunto de coordenadas que encierra un objeto específico dentro de una imagen o fotograma de video para indicar su localización. Según la documentación oficial de Ultralytics (2026), estos cuadros constituyen las anotaciones fundamentales para el entrenamiento de sistemas de Inteligencia Artificial. En el contexto de esta investigación, el bounding box no solo identifica al animal, sino que proporciona los datos de coordenadas necesarios para evitar conteos duplicados mediante algoritmos de seguimiento

(tracking).

Figura 2.8
Hitos en la detección de objetos.



Nota. Tomado de Zou et al. (2023).

2.2.6. Modelos de detección de objetos

YOLO (You Only Look Once)

El modelo YOLO es un enfoque innovador en la detección de objetos que, a diferencia de métodos que requieren múltiples pasos, realiza la identificación y localización en una sola pasada. YOLO aborda la detección como un problema de regresión unificado, dividiendo la imagen en una cuadrícula de celdas y prediciendo, para cada una, múltiples cajas delimitadoras (bounding boxes) junto con sus probabilidades de clase (Class Probabilities). Este enfoque multidimensional permite una detección precisa y eficiente, eliminando la necesidad de redes adicionales para proponer regiones candidatas, lo que lo hace significativamente más rápido y adecuado para aplicaciones en tiempo real. (Lee and Kim, 2020)

SSD (Single Shot MultiBox Detector)

El modelo emplea un enfoque basado en cajas de anclaje y regresión, diseñando previamente cajas de diferentes tamaños y utilizando mapas de características multinivel para predecir objetos de diversas escalas. Utiliza VGG16 como red troncal para la extracción de características, introduciendo modificaciones como la convolución ampliada para aumentar el campo de visión y la convolución con $\text{stride} = 2$ para reducir el tamaño de los mapas de características, obteniendo así seis mapas de diferentes niveles. Durante el entrenamiento, SSD combina la confianza de categoría y el ratio de intersección (IoU) para eliminar cuadros redundantes mediante la supresión de valores no máximos, generando predicciones finales. Además, SSD aborda el desequilibrio entre muestras positivas y negativas seleccionando solo las muestras negativas con mayores errores para el entrenamiento, asegurando una proporción equilibrada. Aunque SSD mejora la precisión y velocidad en la detección de objetos de una sola etapa, su representación semántica y precisión aún tienen margen de mejora (Bai et al., 2022).

Faster R-CNN

Faster R-CNN se organiza en tres componentes principales: la extracción de características mediante una red convolucional, la generación de regiones candidatas a través de la Region Proposal Network (RPN), y la clasificación junto con la regresión para corregir la localización de las cajas delimitadoras. Las características extraídas se utilizan como entrada de la RPN, que emplea capas de clasificación y regresión para distinguir objetos del fondo y ajustar las propuestas. Las regiones generadas se transforman en Regiones de Interés (ROI), las cuales son normalizadas mediante ROI Pooling y posteriormente procesadas por capas totalmente conectadas que producen la clase del objeto y la posición final de la caja delimitadora (Wang and Xiao, 2023).

Este es un algoritmo que destaca por su eficiencia y precisión en comparación con sus predecesores, gracias a la incorporación de una red de propuesta de regiones (Region Proposal Network, RPN). Esta red, similar a una red completamente convolucional, tiene como función principal generar sugerencias de detección y localización de objetos en una imagen. El RPN toma una imagen como entrada y produce puntuaciones de clasificación para las propuestas de objetos, junto con rectángulos delimitadores ajustados. Mediante una ventana deslizante aplicada al mapa de características, se generan regiones de interés, evaluando múltiples propuestas en cada posición. La precisión de estas propuestas se verifica utilizando la métrica de intersección sobre unión. Gracias al mecanismo de la RPN, el costo computacional para estimar las regiones donde se ubican los objetos se reduce significativamente, lo que permite que Faster R-CNN sea más rápido y eficiente que otros métodos. Esta ventaja ha posicionado al algoritmo como una de las soluciones más competitivas en el campo de la detección de objetos (Sahin et al., 2023).

2.2.7. Métricas de Evaluación

Las métricas de evaluación son herramientas clave para medir el rendimiento de los modelos de computación visual. Se utilizan para comparar qué tan precisas son las predicciones del modelo respecto a los datos reales, conocidos como “verdad fundamental”. Una métrica común es la Intersección sobre Unión (IOU), que evalúa la precisión de los cuadros delimitadores (Rančić et al., 2023).

Intersección sobre unión (IoU)

Es una métrica de evaluación fundamental en la detección de objetos que mide el grado de superposición entre dos cuadros delimitadores: el cuadro predicho por el modelo (B_p) y el cuadro real de referencia o ground truth (B_{gt}). Según Szeliski (2022), esta métrica es

indispensable para determinar la exactitud de la localización espacial de un algoritmo.

El IoU se calcula como el cociente entre el área de la intersección de ambos cuadros y el área de la unión de los mismos. La fórmula matemática se expresa de la siguiente manera:

$$IoU = \frac{\text{Área de la Intersección}}{\text{Área de la Unión}} = \frac{Area(B_p \cap B_{gt})}{Area(B_p \cup B_{gt})}$$

$IoU = 1,0$: Coincidencia perfecta.

$IoU > 0,5$: Generalmente se considera una detección buena, esto depende del umbral elegido.

$IoU = 0,0$: No hay ningún solapamiento.

Matriz de Confusión

La matriz de confusión en la detección de objetos es distinta de la clasificación estadística convencional (Hoang, 2024), ya que integra la precisión de la localización espacial mediante la métrica IoU. Según Szeliski (2022), para determinar si una predicción es correcta, no basta con la coincidencia de etiquetas; es obligatorio que el cuadro delimitador predicho se superponga significativamente con el cuadro real (Ground Truth).

Verdadero Positivo (TP): Se produce cuando el modelo predice la clase correcta y la superposición espacial con el objeto real es mayor o igual al umbral establecido ($IoU \geq 0,5$).

Esto confirma una detección exitosa en identidad y posición.

Falso Positivo (FP): Se genera bajo dos condiciones:

- *Error de clasificación*, cuando el modelo detecta un objeto donde no existe (detección de "fantasmas." ruido visual).
- *Error de localización*, cuando el modelo identifica la clase correcta, pero el cuadro delimitador tiene un $IoU < 0,5$ respecto al objeto real, indicando una imprecisión espacial inaceptable.

Falso Negativo (FN): Ocurre cuando un objeto no es detectado por el modelo, o la predicción asociada posee un IoU insuficiente o demasiado bajo para ser considerada válida.

Verdadero Negativo (TN): No considerada para problemas de detección de objetos, como se señala la documentación de Jocher et al. (2023) de Ultralytics, una imagen normalmente tienen infinidad de cuadros que no contienen un objeto.

Precisión

Mide el porcentaje de predicciones positivas que el modelo acertó en relación con el total de predicciones positivas que realizó (Rančić et al., 2023), (Szeliski, 2022). Expresado de otra forma mide la exactitud de los objetos detectados, esto indica cuántas detecciones fueron correctas (Jocher et al., 2023). La precisión se calcula de la siguiente forma:

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (2.2)$$

Recall

Se define como la proporción entre las instancias positivas detectadas correctamente por el modelo y todas las instancias positivas etiquetadas (Szeliski, 2022; Rančić et al., 2023). También se define como la capacidad del modelo para identificar todas las instancias de objetos en las imágenes (Jocher et al., 2023), el recall se calcula de la siguiente forma:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.3)$$

F1-Score

Es una métrica que combina la precisión y el recall en un solo valor, equilibrando ambos aspectos para evaluar el rendimiento de un modelo. Ésta métrica busca un balance entre estos dos conceptos. Su valor oscila entre 0 y 1, donde 0 indica un rendimiento pobre (ya

sea por baja precisión, bajo recall o ambos) y 1 representa un rendimiento perfecto, es decir, cuando el modelo tiene tanto alta precisión como alto recall (Rančić et al., 2023) F1-Score e define como:

$$F_1 \text{ score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (2.4)$$

Mean Average Precision (mAP)

Existe un compromiso intrínseco entre la precisión y el recall; al incrementar el umbral de confianza del modelo, la precisión aumenta y el recall disminuye, y viceversa. Ambas métricas por sí sola no determinan el rendimiento del modelo de detección de objetos, por lo que se empleó la Precisión Media Promedio como métrica global. El mAP se define como la media aritmética de la Precisión Promedio (AP) calculada para cada una de las clases del conjunto de datos, de acuerdo con la siguiente expresión:

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i$$

Donde AP_i representa la precisión promedio para la clase i , y C corresponde al número total de clases presentes en el dataset (en este estudio, llamas y alpacas). El procedimiento detallado para el cálculo del área bajo la curva que define el AP se basa en la metodología mencionada en Rančić et al. (2023).

Frames per Second (FPS)

Los FPS son la métrica que determina la fluidez del procesamiento de una secuencia de video. Representan la cantidad de cuadros individuales que el modelo de visión computacional puede procesar y anotar en un segundo. Matemáticamente, los FPS están inversamente relacionados con la latencia total de procesamiento por cuadro.

Latencia de inferencia

La latencia de inferencia es el tiempo transcurrido desde que se ingresan los datos al modelo hasta que se genera una salida, como un cuadro delimitador o una etiqueta de clase. Según la documentación de Ultralytics (2026), este proceso se descompone en tres etapas medidas en milisegundos (ms).

- *Preprocesamiento*: Incluye el redimensionamiento de la imagen al tamaño de entrada y la normalización de píxeles.
- *Inferencia*: Es el tiempo que tarda la imagen en realizar el paso hacia adelante (forward pass) que pasa por el modelo hasta calcular una predicción.
- *Postprocesamiento*: Incluye operaciones adicionales como la Supresión No Máxima (NMS) para refinar los cuadros delimitadores y eliminar duplicados (más adelante se verá como modelos nuevos de un solo paso ya no aplican NMS).

2.2.8. Tracking

El seguimiento de objetos es el proceso de estimar el estado de un objeto (como una alpaca o llama) a medida que se desplaza a través de una secuencia de video. Según Szeliski (2022), el objetivo es establecer una correspondencia consistente entre las detecciones de cuadros consecutivos, permitiendo que el sistema reconozca que una instancia en el cuadro t es la misma que apareció en el cuadro $t - 1$.

2.2.9. Aplicaciones de drones en la Ganaderia

2.2.10. Sistemas autonomos

Los sistemas autonomos requieren una minimia intervencio humana, son capaces de operar en entornos peligrosos, desconocidos y no estructurados, adaptándose a eventos ines-

perados mientras realizan tareas específicas. Nuestra investigación abarca diversos aspectos de estos sistemas, incluyendo vehículos aéreos no tripulados (UAV), utilizados en vigilancia y exploración; vehículos submarinos no tripulados (UUV), empleados en exploración submarina, detección de contaminación y cartografía industrial; y vehículos terrestres no tripulados (VANT), que operan en tierra, interactúan con el entorno y manipulan objetos, como en el caso de excavadoras autónomas para investigaciones del subsuelo. Estos sistemas son clave para aplicaciones en entornos remotos y desafiantes (Department of Electrical and Electronic Engineering, University of Manchester, 2024)

2.2.11. Herramientas

Python

Es un lenguaje de programación de código abierto, orientado a objetos y de propósito general. Cientos de miles de desarrolladores lo utilizan tanto para programas independientes como para aplicaciones de scripting, desarrollo de aplicaciones web, en una amplia variedad de ámbitos. Python está diseñado para optimizar la productividad de los desarrolladores, la calidad del software, la portabilidad de los programas y la integración de componentes.

Python es un lenguaje de programación multiplataforma, lo que significa que funciona en los principales sistemas operativos. Cualquier programa Python que escribas debería ejecutarse en cualquier ordenador moderno que tenga Python instalado. Python evoluciona a medida que surgen nuevas ideas y tecnologías siendo la versión más actual Python 3, además, los desarrolladores de python han hecho que el lenguaje sea cada vez más versátil y potente incluso para aplicaciones de inteligencia artificial.

PyTorch

Es una biblioteca optimizada para el aprendizaje profundo que utiliza tanto GPUs como CPUs para procesar tensores de manera eficiente. Su documentación clasifica las características en tres categorías según su estado de desarrollo: Estable, que garantiza mantenimiento a largo plazo, compatibilidad hacia atrás y documentación completa; Beta, que incluye funciones en mejora continua, sujetas a cambios basados en retroalimentación y sin garantía de compatibilidad futura; y Prototipo, que son características en fase inicial de prueba, generalmente no disponibles en distribuciones binarias y destinadas a recibir comentarios (PyTorch Developers, 2024)

OpenCV

Es una biblioteca de código abierto diseñada principalmente para aplicaciones de visión por computadora. Ofrece una amplia variedad de funciones y algoritmos que permiten realizar tareas como seguimiento de movimiento, reconocimiento facial, detección y segmentación de objetos, entre otras. Esta herramienta es muy versátil, ya que permite manipular tanto imágenes como flujos de video en tiempo real, adaptándolos a diversas necesidades (Adusumalli et al., 2021)

Anotación de Datos

La anotación de datos es un proceso esencial en el desarrollo de modelos de visión por computadora, ya que permite a estos sistemas interpretar y responder con precisión al mundo visual. Consiste en etiquetar datos visuales, como imágenes, videos o texto, con información descriptiva o identificativa, proporcionando el contexto necesario para que los modelos reconozcan patrones, objetos y escenarios (OpenCV Team, 2024). Existen muchas herramientas que facilitan a realizar esta tarea de anotación de imágenes, entre ellas tenemos como:

- **Roboflow:** Es una plataforma que facilita la gestión de datos de visión por computadora, permitiendo a los usuarios cargar, organizar, anotar y exportar imágenes para entrenar modelos de aprendizaje automático. El autor destaca las características clave de Roboflow, como su interfaz intuitiva para la anotación de imágenes, herramientas de preprocesamiento y aumentación de datos, y la capacidad de exportar conjuntos de datos en formatos compatibles con diversos frameworks de aprendizaje profundo (Choudhary, 2024)
- **CVAT (Computer Vision Annotation Tool):** Es una herramienta de anotación de código abierto creada por Intel, diseñada para tareas como detección de objetos, segmentación de imágenes y anotación de videos. Es compatible con múltiples formatos de anotación y ofrece APIs para una fácil integración con otros sistemas. Al ser una plataforma basada en la web, puede instalarse en servidores locales o en la nube, lo que la hace altamente personalizable y escalable para diversos proyectos de visión por computadora (Musleh et al., 2023)

Entornos de Desarrollo Google Colab

Google Colab es una plataforma gratuita basada en Jupyter Notebook que ofrece acceso a GPUs y TPUs para la ejecución de modelos de inteligencia artificial. Su integración con Google Drive y la posibilidad de ejecutar código en la nube sin necesidad de configuración avanzada lo hacen una opción popular para el desarrollo de proyectos de inteligencia artificial.

Capítulo 3

Desarrollo del proyecto

Para el desarrollo del siguiente proyecto, se utilizó una modificación del Cross-Industry Standard Process for Data Mining, al cual llamamos CRISP-DM Modified, que se encuentra en la Sec. 1.7.5, donde se menciona con más detalle sobre las etapas y funcionamiento.

3.1. Fase 1: Entendimiento del proyecto

3.1.1. Análisis del problema

Tal como se planteó en los aspectos generales de la presente investigación (ver Capítulo 1), la dificultad del conteo manual requiere una solución automatizada. Dicho problema se traduce en la necesidad de crear una solución tecnológica para realizar el conteo automático de camélidos sudamericanos en tiempo real sin conexión a internet mientras un UAV sobrevuela el hato de animales, adaptada para entornos altoandinos con computación de borde (End to End). Aquí también se presentan los objetivos generales de la investigación, los cuales son las especificaciones técnicas y los requerimientos del sistema.

Para comprender la magnitud del problema y las necesidades reales de los productores, se diseñó y aplicó una encuesta técnica (Anexo 5) dirigida a cuatro productores de camélidos

sudamericanos domésticos en la región altoandina. El instrumento recopiló información sobre: cantidad de animales por hato, método de conteo utilizado, tiempo estimado por conteo, principales dificultades encontradas, número de personal involucrado, disposición a adquirir una solución tecnológica y preferencias sobre el modelo de adquisición (compra directa o individual). La Tabla 3.1 resume los hallazgos obtenidos.

Tabla 3.1

Resultados de la encuesta a productores de camélidos

P	Especie	Conteo	Uni.	T. m/h	Dificultad	Dispo.	Forma
1	llama	manual	250	1/2h	tiempo/errores	si	colectiva
2	alpaca	manual	540	+2h	tiempo/errores	si	colectiva
3	alpaca	manual	800	+2h	tiempo/errores	si	individual
4	alpaca/llama	manual	800	1/2h	tiempo/errores	si	individual

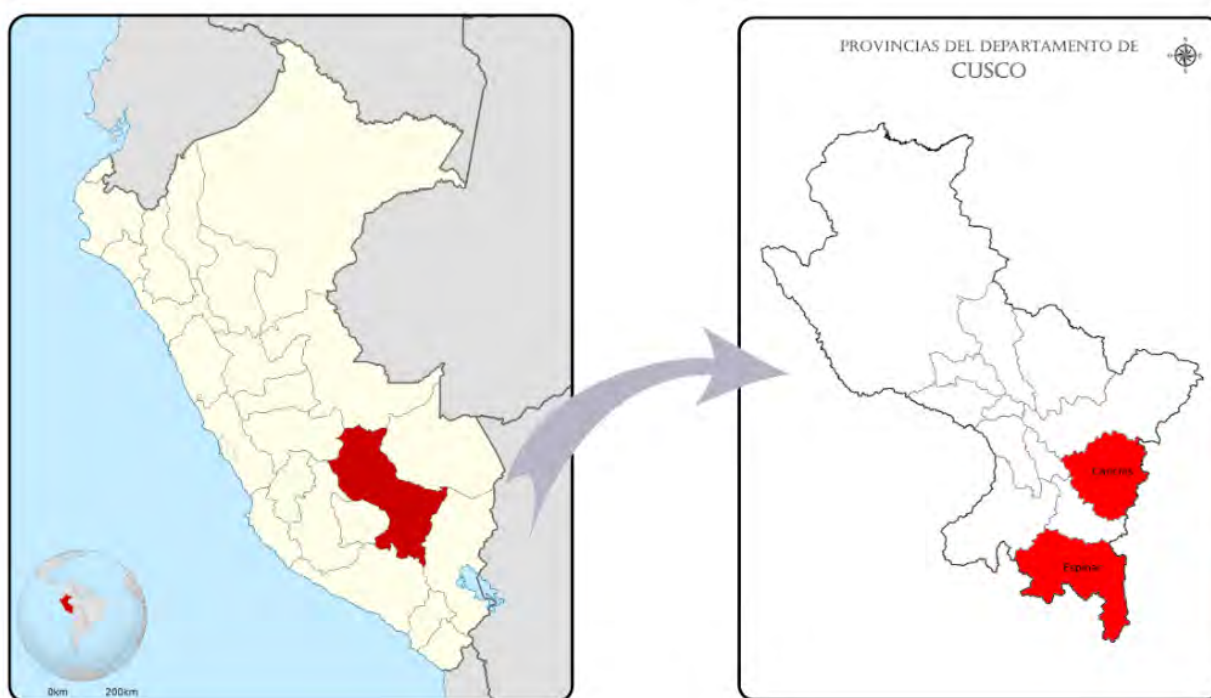
Los resultados de la encuesta evidencian que el conteo manual es un proceso lento y propenso a errores, lo que refuerza la necesidad de implementar una solución automatizada.

3.1.2. Zona de estudio

El estudio se realizó en dos provincias estratégicas que se dedican a la crianza de camélidos sudamericanos (Espinar y Canchis), como se puede ver en la Fig. 3.1. Las zonas en estudio se detallan en la Tabla 3.2. Además, es importante mencionar que las comunidades altoandinas se ubican a más de 3900 msnm y se caracterizan por un relieve de montañas pronunciadas con terrenos irregulares, donde predominan extensos pajonales (ichu) y bofedales que conforman el ecosistema de pastoreo. Las condiciones climáticas son adversas y variables, registrándose vientos fuertes de manera frecuente, así como la presencia recurrente de truenos y tormentas eléctricas, particularmente durante la temporada de lluvias.

Figura 3.1

Lugar de estudio: Provincia de Espinar y Canchis - Perú



En cuanto a la accesibilidad, las zonas de estudio se ubican en áreas rurales de difícil acceso, siendo necesario el desplazamiento exclusivamente con vehículo propio o privado para llegar a los fundos ganaderos, debido a la ausencia de transporte público y a las condiciones precarias de las vías de acceso.

Finalmente, en cuanto a la cantidad de animales de los fundos ganaderos o productores de camélidos sudamericanos, las llamas presentan poblaciones reducidas de hasta 200 ejemplares, mientras que las alpacas mantienen poblaciones significativas que pueden llegar hasta los 2000 ejemplares, lo que influye directamente en el volumen de datos recolectados por clase.

Tabla 3.2

Zonas de estudio

Provincia	Comunidad	Especie	Latitud — Longitud	Altitud
Espinar	Hanccollahua	llamas,alpacas	15°1'48.79" S : 71°16'28.42" O	4361
Espinar	Chisicata	otros	14°47'3.03" S : 71°26'15.29" O	3900
Espinar	Huarca	llamas,alpacas,otros	14°47'47.74" S : 71°25'54.20" O	3915
Canchis	Phinaya	alpacas	13°53'32.95" S : 70°58'6.05" O	4693
Canchis	Marangani	alpacas	14°31'42.93" S : 71° 1'41.47" O	4500

3.1.3. Requerimientos de hardware

Para la viabilidad del proyecto se requirieron los siguientes componentes físicos:

- **Dron:** Para el desarrollo del proyecto completo se utilizaron 3 drones, y para la prueba final solo un dron. Con la finalidad de facilitar la transmisión de video en tiempo real, se utilizó un Dron DJI Mini 3 con control RC-N1 (ver Fig. 3.2a), seleccionado por su sensor CMOS de 1/1.3 μ resistencia a ráfagas de viento en altitud. Tiene calidad de video 4K a 30 FPS y MSCK para que los desarrolladores puedan crear aplicaciones propias. Además, es importante que el dron tenga baterías de repuesto para la recolección de datos y las pruebas de vuelo.
- **Smartphone:** Los dispositivos compatibles con el sistema solo pueden ser Android con arquitectura ARM, capaces de ejecutar modelos optimizados mediante aceleración por GPU o NPU. También es importante que el celular sea de gama media a alta para tener una transmisión fluida de video entre el celular y el dron, y para no tener pérdida de FPS. En la aplicación de campo se optó por utilizar un Samsung S25 Ultra, con las siguientes características:
 - Procesador (SoC): Qualcomm Snapdragon 8 Elite for Galaxy
 - RAM: 12GB
 - GPU: Adreno 830

Figura 3.2

UAVs utilizados de DJI para la la investigación



(a) Dron DJI Mini 3 con RC-N1



(b) Dron DJI Mavic 3 Enterprise

3.1.4. Requerimientos funcionales

- **RF-01 Detección:** Detectar y clasificar los cuadros delimitadores (bounding boxes) de las clases llama y alpaca.
- **RF-02 Streaming:** Interceptar y decodificar el flujo de video del dron en tiempo real dentro del aplicativo.
- **RF-03 Tracking:** Asignar un identificador a cada animal detectado para mantener el seguimiento entre fotogramas del video.
- **RF-04 Conteo acumulativo:** Realizar la sumatoria total de individuos detectados, evitando el doble conteo de un mismo animal, teniendo en cuenta si pertenece a la clase llama o alpaca.
- **RF-05 Despliegue del aplicativo:** Mostrar en pantalla, en tiempo real, el contador de cada clase, así como las cajas delimitadoras (*bounding boxes*) y la línea de conteo.
- **RF-06 Configuración:** Permitir al usuario modificar parámetros de operación y visualización, como el umbral de confianza, la selección del modelo, mover las líneas de conteo y ajustar los parámetros de tracking.

- **RF-07 Reporte de vuelo:** Generar un reporte al finalizar cada vuelo, que contenga al menos: fecha, hora, duración del vuelo, total de alpacas y total de llamas detectadas.
- **RF-08 Historial de vuelos:** Almacenar y permitir consultar un historial de vuelos previos con sus respectivos reportes.

3.1.5. Requerimiento no funcionales

- **RNF-01 Inferencia offline:** El sistema debe ejecutar el modelo de detección de objetos de manera local, sin depender de tener conexión a internet.
- **RNF-02 FPS:** El sistema debe mantener una tasa de procesamiento de al menos 15 fotogramas por segundo en el smartphone, para que garantice un conteo fluido.
- **RNF-03 Optimización de modelo:** El sistema empleará un modelo optimizado para dispositivos móviles, utilizando TensorFlow Lite con técnicas de cuantización, a fin de reducir el tamaño del modelo y el consumo de recursos computacionales.
- **RNF-04 Latencia:** El tiempo de respuesta entre la captura de cada fotograma y la visualización de la detección correspondiente debe ser rápido para que los bounding boxes aparezcan sobre los animales.
- **RNF-05 Precisión del modelo:** El modelo de detección debe alcanzar un mAP@50 superior al 85 % en el conjunto de validación y prueba para tener un buen desempeño en una aplicación en tiempo real de campo.

3.2. Fase 2: Adquisición de datos

La cantidad de datos recolectados es un factor determinante para el desempeño de los modelos de detección de objetos, puesto que influye directamente en la capacidad de generalización

de estos. No solo debe considerarse el número de imágenes por clase, sino también la cantidad total de anotaciones, dado que una misma imagen puede contener múltiples instancias de interés. Estudios como el de Wang and Xiao (2023) demuestran que es posible obtener resultados satisfactorios con conjuntos de datos moderados, siempre que exista una cantidad suficiente de anotaciones y una adecuada captura de datos en distintos escenarios.

Es así que la adquisición de datos se realizó de forma iterativa durante el año 2025 en sus distintas estaciones, con la finalidad de adquirir la mayor cantidad de datos posible en distintos escenarios para que nuestro sistema generalice mejor. Para este propósito se realizaron coordinaciones constantes con los fundos ganaderos para tener acceso cuando se requiriera a través de un UAV.

3.2.1. Captura de datos con UAV

Debido a la falta de datasets de imágenes aéreas sobre camélidos sudamericanos domésticos, según los requisitos técnicos de la investigación, se optó por la creación de un dataset propio. La recolección de datos se realizó en zonas de pastoreo mediante el uso de UAVs, lo que permitió capturar videos e imágenes aéreas bajo ciertos criterios que se verán a continuación:

Especificaciones técnicas

Durante el proceso de recolección de datos se utilizaron 3 drones, estos son: DJI Mini 3 (ver Fig. 3.2a), DJI Mavic 3 Enterprise (ver Fig. 3.2b) y DJI Phantom 4, cada uno con 2 baterías de repuesto, teniendo cada batería aproximadamente 20 minutos de autonomía de vuelo. Estos equipos superan los 1.5 km de distancia sin interferencia para transmisión en tiempo real y cuentan con cámaras Ultra High Definition 4K a 30 FPS.

Configuración de cámaras de vuelo

Las tomas de video se realizaron en resolución 4K y 2k a 30 FPS. Se fijó el balance de blancos y la velocidad de obturación para evitar el desenfoque por movimiento, garantizando que cada fotograma extraído fuera apto para el entrenamiento.

Altura de vuelo

Se determinó una altura operativa desde 25 hasta 70 metros. Esta altitud proporciona detalles que aún permiten identificar rasgos distintivos entre la llama y la alpaca (cuello, orejas o tamaño), sin que el ruido de los motores disperse al ganado. No es posible definir una altura específica para cada vuelo debido a la irregularidad del terreno, pero dentro de los metadatos de las imágenes se colocó una altura aproximada.

Figura 3.3

Captura de datos a 30 y 70 metros de altura, en la comunidad de Hanccollahua



(a) *Captura de datos a 30 metros de altura en ángulo cenital (Hanccollahua)*



(b) *Captura de datos a 70 metros de altura en ángulo cenital (Hanccollahua)*

Ángulo de inclinación de la cámara

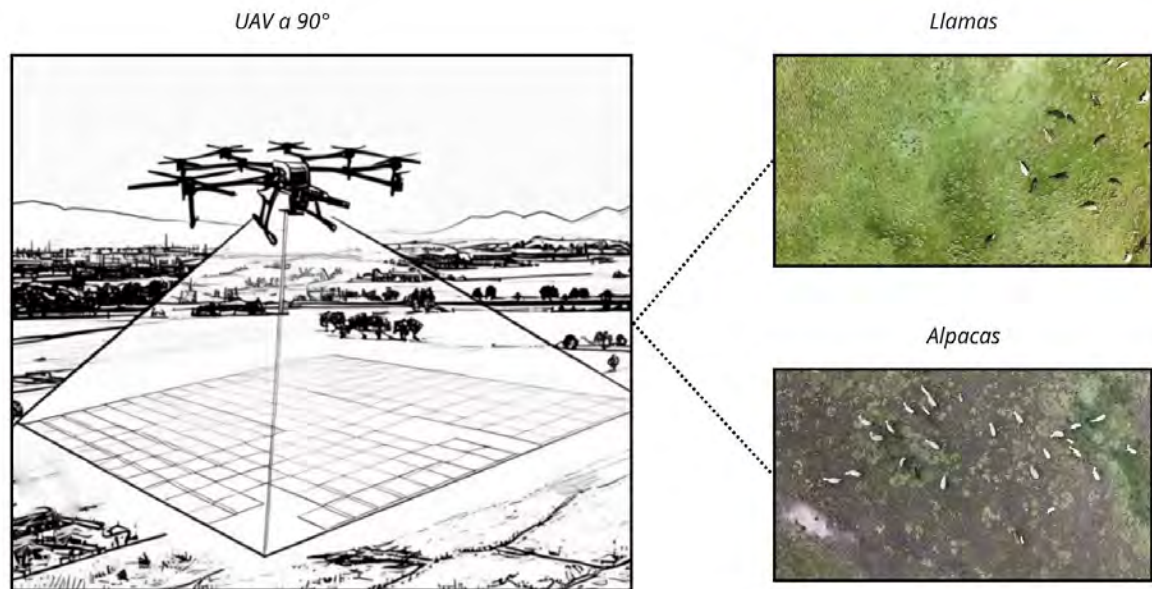
Se emplearon ángulos de -90° y ángulos de inclinación desde -45° hasta -20° . Estas perspectivas son indispensables para el conteo automático, puesto que se pretende cubrir escenarios de conteo en cerco perimétrico y campo abierto (pastoreo extensivo).

Ángulo -90 grados: El ángulo de -90 grados respecto a la horizontal representa la posición cenital de la cámara. En este ángulo, el campo de visión del dron es muy limitado; sin

embargo, es una posición perfecta para realizar conteos en cercos perimétricos porque no existe oclusión entre animales. Ver Fig. 3.4.

Figura 3.4

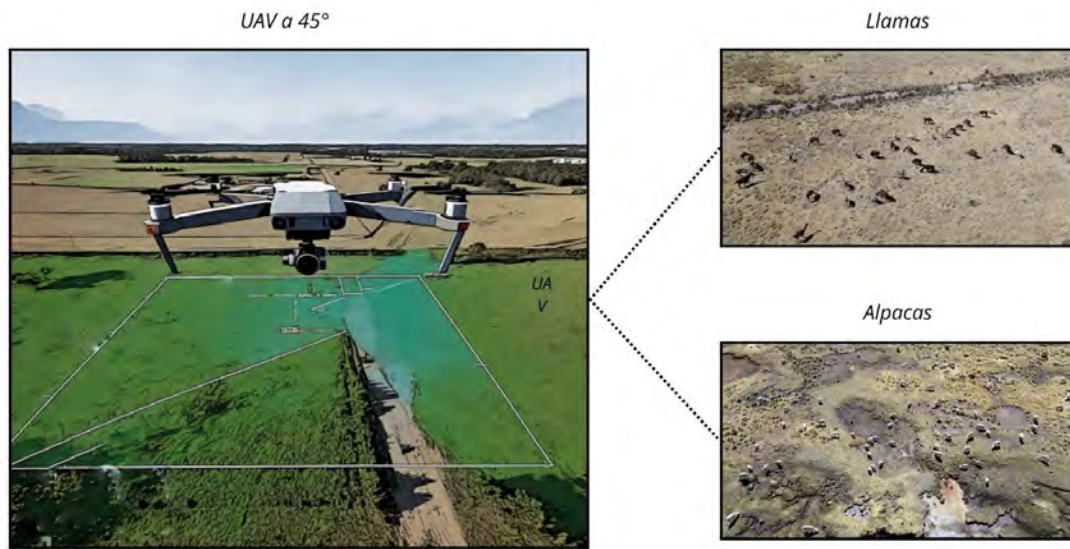
Ángulo de visión de la cámara a -90 grados



Ángulo entre -45 y -20 grados: Este tipo de inclinación tiene una vista panorámica. El dron tiene un mejor campo de visión, pero existe el problema de oclusión cuando los animales se juntan demasiado. Esta inclinación de cámara es perfecta para el conteo y monitoreo en campo abierto. Ver Fig. 3.5.

Figura 3.5

Ángulo de visión de la cámara con inclinación



3.2.2. Condiciones de captura de datos

Horario

Para asegurar que el dataset sea robusto y contenga un conjunto diverso de datos, la captura de videos se realizó en bloques horarios entre las 10:00 h y las 14:00 h. Durante este intervalo se abarcaron diferentes condiciones de iluminación, incluyendo luz matutina con sombras alargadas, luz cenital de mediodía y condiciones de luz difusa en tardes nubladas.

Clima

La limitación del horario estuvo basada en las condiciones climáticas de la zona en campo abierto como fuertes vientos y lluvias. Además, la captura se realizó a lo largo del año 2025 en sus diferentes estaciones, permitiendo registrar variaciones ambientales y de iluminación propias de cada periodo y contribuyendo a un conjunto diverso y representativo de datos.

Entorno geográfico

La captura de datos se desarrolló en una zona de estudio correspondiente al paisaje andino en campo abierto. Se buscó cubrir la variabilidad cromática del entorno según las características presentes en cada sector de la zona estudiada, incluyendo pasto seco, bofedales, zonas de tierra, lugares pedregosos, pasto verde y pajonales.

Conjunto diverso

Finalmente, se tomó en consideración la variabilidad de escenarios. Dentro de ellos se encuentran: capturas de videos de solo llamas, capturas de videos de solo alpacas, capturas de videos de llamas y alpacas juntas, capturas de video de imágenes negativas (background e imágenes de otros animales). También se obtuvieron videos de todo tipo de animales, objetos y personas en conjunto, incluyendo llamas y alpacas de la Expo Feria Internacional Reyes 2026, esto con el objetivo de que nuestro modelo detecte a la llama y a la alpaca en todos los escenarios posibles y aprenda a diferenciarlos de otros animales de la zona.

3.3. Fase 3: Preparación de datos

En esta fase de la metodología CRISP-DM Modificada, los datos crudos obtenidos (videos) en campo se transforman en un dataset estructurado. Los videos se convertirán en imágenes con anotaciones y una división de dataset listo para entrenamiento.

3.3.1. Procesamiento de datos

Los datos obtenidos durante la fase previa se encuentran en formato MP4 con resolución de Ultra High Definition (4K) y HD (2k) a 30 FPS. Estos son convertidos en imágenes a través de la *extracción de frames*. Para este proceso se utilizó un script en Python (ver pseudocódigo 1), que extrae imágenes mediante la selección de frames en intervalos definidos,

con el propósito de evitar redundancia de imágenes consecutivas y garantizar la diversidad visual en el conjunto de datos.

Algorithm 1 Algoritmo de extracción de fotogramas

```

1: Entrada: Video en formato MP4 (4K, 30 FPS)
2: Salida: Conjunto de imágenes JPG
3: Parámetro: Intervalo de extracción  $k \leftarrow 30$  frames

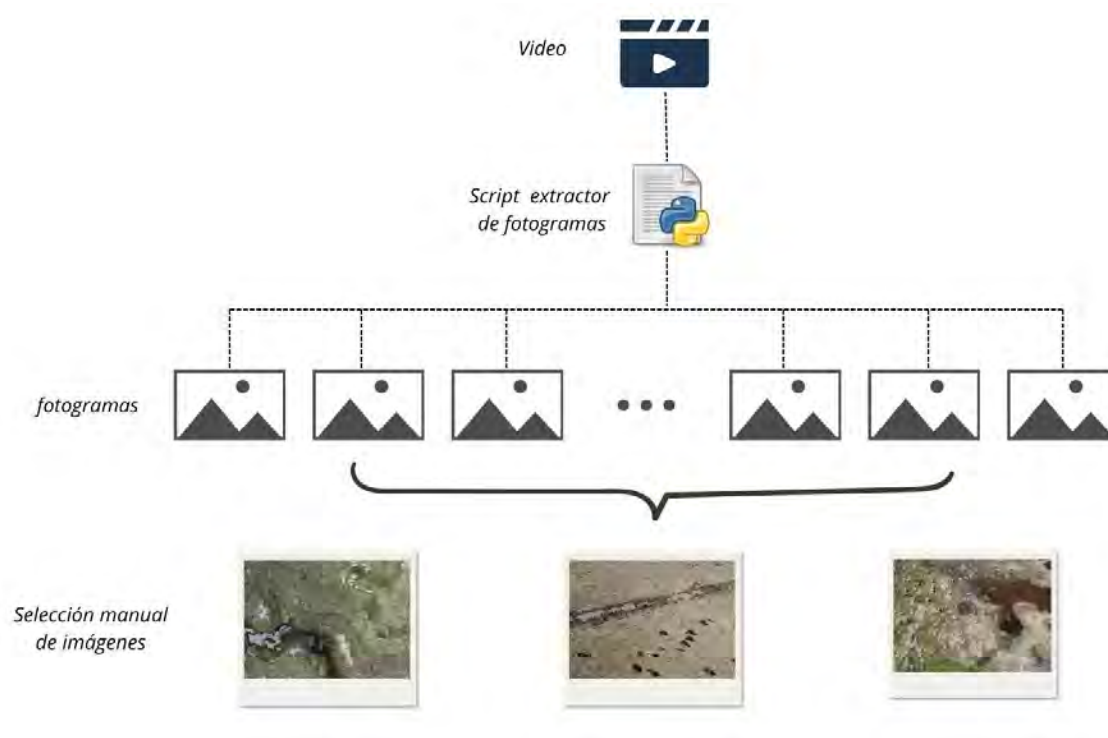
4: Crear carpeta de salida frames_video
5: Inicializar captura de video  $cap \leftarrow VideoCapture(video)$ 
6: Obtener total_frames y fps del video
7: Inicializar  $frame\_counter \leftarrow 0$ ,  $saved\_images \leftarrow 0$ 
8: while existen frames en video do
9:    $ret, frame \leftarrow cap.read()$ 
10:  if no ret then
11:    break
12:  end if
13:  if  $frame\_counter \% k == 0$  then
14:     $saved\_images \leftarrow saved\_images + 1$ 
15:     $filename \leftarrow DJI\_animal\_ubicacion\_anguloInclinacion\_alturaVuelo$ 
16:     $imwrite(filename, frame)$ 
17:  end if
18:   $frame\_counter \leftarrow frame\_counter + 1$ 
19: end while
20:  $cap.release()$ 
21: return  $saved\_images$ 

```

Seguidamente, una vez obtenidos los fotogramas, se realiza un filtrado manual de las imágenes. Los cuadros afectados por vibraciones extremas de la cámara del dron o por cambios bruscos de exposición son eliminados, obteniendo así un dataset más limpio. Este proceso se puede observar en la Fig. 3.6.

Figura 3.6

Extracción de fotogramas de los videos



Construcción de datasets

En el proceso de extracción de frames, se obtuvieron 4 datasets, los cuales se detallan a continuación.

Dataset 1: La Tabla 3.3 muestra la cantidad de imágenes obtenidas para llamas y alpacas considerando un ángulo de inclinación de la cámara de -90 grados y a una altura de vuelo del dron entre 25 y 70 metros respecto al nivel del punto de partida del UAV.

Tabla 3.3

Cantidad de imágenes obtenidas de -90 grados

Especie	Cantidad de imágenes	Angulo	Altura (m)
Llamas	685	-90°	25 hasta 70 m.
Alpacas	375	-90°	25 hasta 70 m.
Total	1060	—	—

Dataset 2: La Tabla 3.4 muestra la cantidad de imágenes obtenidas para llamas y alpacas considerando un ángulo de inclinación de la cámara de -45 a -20 grados y una altura de vuelo del dron entre 25 y 70 metros respecto al nivel del punto de partida del UAV.

Tabla 3.4*Cantidad de imágenes obtenidas de -45 hasta -20 grados*

Especie	Cantidad de imágenes	Angulo	Altura (m)
Llamas	259	-45° hasta -20°	25 hasta 70 m.
Alpacas	220	-45° hasta -20°	25 hasta 70 m.
Total	479	—	—

Dataset 3: La Tabla 3.5 muestra la unión del conjunto de datos de la Tabla 3.3 y la Tabla 3.4. Esto supone la cantidad de imágenes obtenidas para llamas y alpacas con ángulos de inclinación de la cámara de -45 a -20 grados y de -90 grados con respecto al nivel del punto de partida del UAV.

Tabla 3.5*Cantidad de imágenes obtenidas de -45° hasta -20° y -90° grados*

Especie	Cantidad de imágenes	Angulo	Altura (m)
Llamas	944	-45° hasta -20° ; -90°	25 hasta 70 m.
Alpacas	595	-45° hasta -20° ; -90°	25 hasta 70 m.
Total	1539	—	—

Dataset 4: En esta última iteración de obtención de datos se vio la necesidad de incluir imágenes negativas (imágenes que contienen objetos que no son de nuestro interés) en tareas de detección de objetos, además de contar con imágenes que contienen anotaciones de las clases objetivo (llama y alpaca). Una forma de enseñar al modelo a no detectar otras especies como si fueran clases objetivo es mediante la inclusión de imágenes negativas, es decir, ejemplos que no contienen ningún objeto de interés, como fondos, otros animales o piedras. Esta es una práctica recomendada para mejorar la capacidad del modelo de diferenciar entre el objeto de interés y el fondo. Estudios recientes han demostrado que la introducción de estas muestras permite reducir la tasa de falsos positivos, ya que el modelo aprende a no generar predicciones incorrectas cuando se enfrenta a contenido sin objetos relevantes (Nogueira-Rodríguez et al., 2023). No obstante, la proporción óptima de imágenes negativas no sigue una regla matemática fija, sino que depende del contexto de la imagen. (Gao et al., 2019) Demostraron experimentalmente que, en imágenes con una gran área de fondo o tierra, se

requieren imágenes negativas para suprimir las falsas alarmas, las que ofrecieron un mejor rendimiento.

La Tabla 3.6 muestra la unión del dataset de la Tabla 3.5 agregando imágenes negativas.

Tabla 3.6

Cantidad de imagenes obtenidas en 45° y 90° grados

Especie	Cantidad de imágenes	Angulo	Altura (m)
Llamas	944	-45° hasta -20° ; -90°	25 hasta 70 m.
Alpacas	595	-45° hasta -20° ; -90°	25 hasta 70 m.
Alpacas y llamas	80	-45° hasta -20° ; -90°	25 hasta 70 m.
Negativos	230	-45° hasta -20°	25 hasta 70 m.
Total	1841	—	—

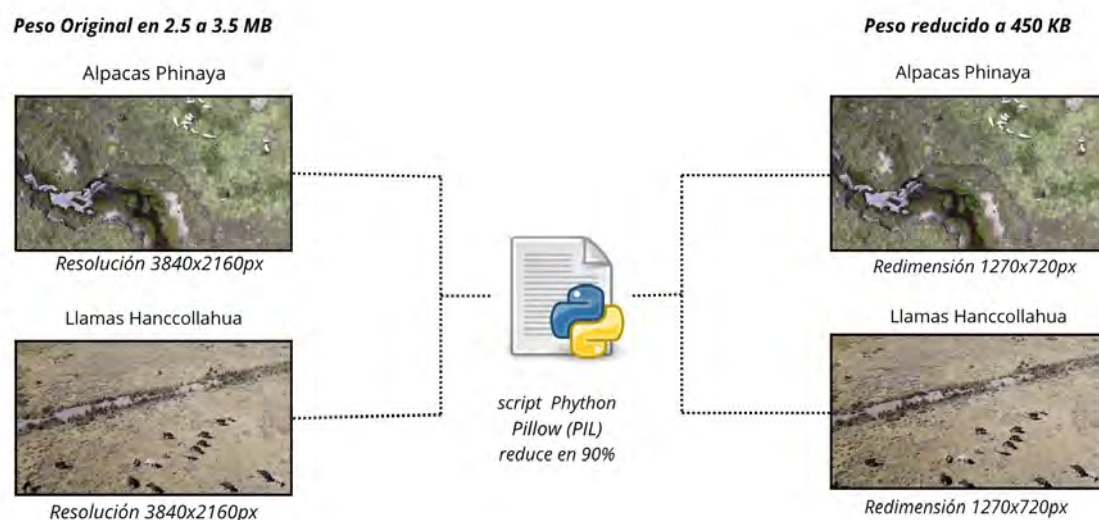
Reducción de dimensión de imágenes

Para reducir costos computacionales, se aplicó una reducción de resolución a las imágenes de entrada, disminuyendo su tamaño de 3840×2180 píxeles a 1280×720 píxeles, como se ve en la Fig. 3.7. Esta práctica juega un papel importante debido a que disminuye la demanda de costo computacional. En el estudio de Morlier et al. (2025) se ha identificado la reducción de la resolución de entrada como una técnica de compresión efectiva para sistemas de deep learning, que permite mantener un rendimiento competitivo en tareas como clasificación y segmentación, mientras reduce los requerimientos computacionales y de memoria.

Nota: Sólo se realizó la reducción de la dimensión espacial de las imágenes.

Figura 3.7

Redimensionamiento de imágenes a 1280x720



3.3.2. Etiquetado y anotación

Los modelos de detección de objetos de la familia YOLO, Faster R-CNN y SSD necesitan datos etiquetados mediante cuadros delimitadores (bounding boxes) que rodeen cada instancia del objeto de interés, indicando su posición dentro de la imagen y la clase a la que pertenece (ver Fig. 3.9 y Fig. 3.10). Este tipo de aprendizaje se conoce como *aprendizaje supervisado*. Kaur and Singh (2022) señala que la calidad y precisión del etiquetado influyen directamente en el desempeño de los modelos de detección entrenados con esas imágenes, y que la disponibilidad de anotaciones detalladas y bien estructuradas es esencial para obtener resultados precisos en tareas de detección de objetos.

Computer Vision Annotation Tool (CVAT)

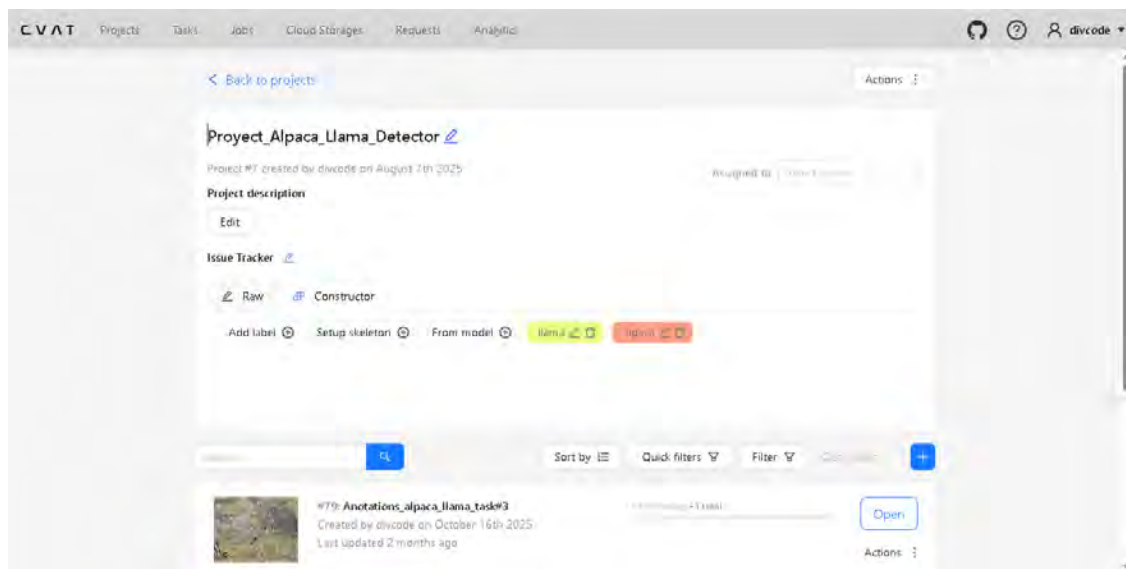
Para la realización de anotaciones de bounding boxes, se utilizó CVAT, una plataforma open-source que sirve específicamente para etiquetar imágenes y videos de forma local para

aplicaciones de visión computacional (ver Fig. 3.8). CVAT soporta diversos tipos de anotaciones para tareas de clasificación, detección, segmentación y estimación de poses. Asimismo, permite exportar las etiquetas en formatos como YOLO, COCO o Pascal VOC. Su interfaz intuitiva facilita la creación de anotaciones de calidad, reduciendo errores y permitiendo controlar de forma precisa la información de cada objeto etiquetado.

El despliegue y la ejecución se realizaron en un entorno local utilizando contenedores Docker, lo que permitió un mayor control del proceso de etiquetado manual, garantizando la seguridad y disponibilidad de las imágenes recolectadas.

Figura 3.8

Carga de imágenes para el etiquetado en CVAT

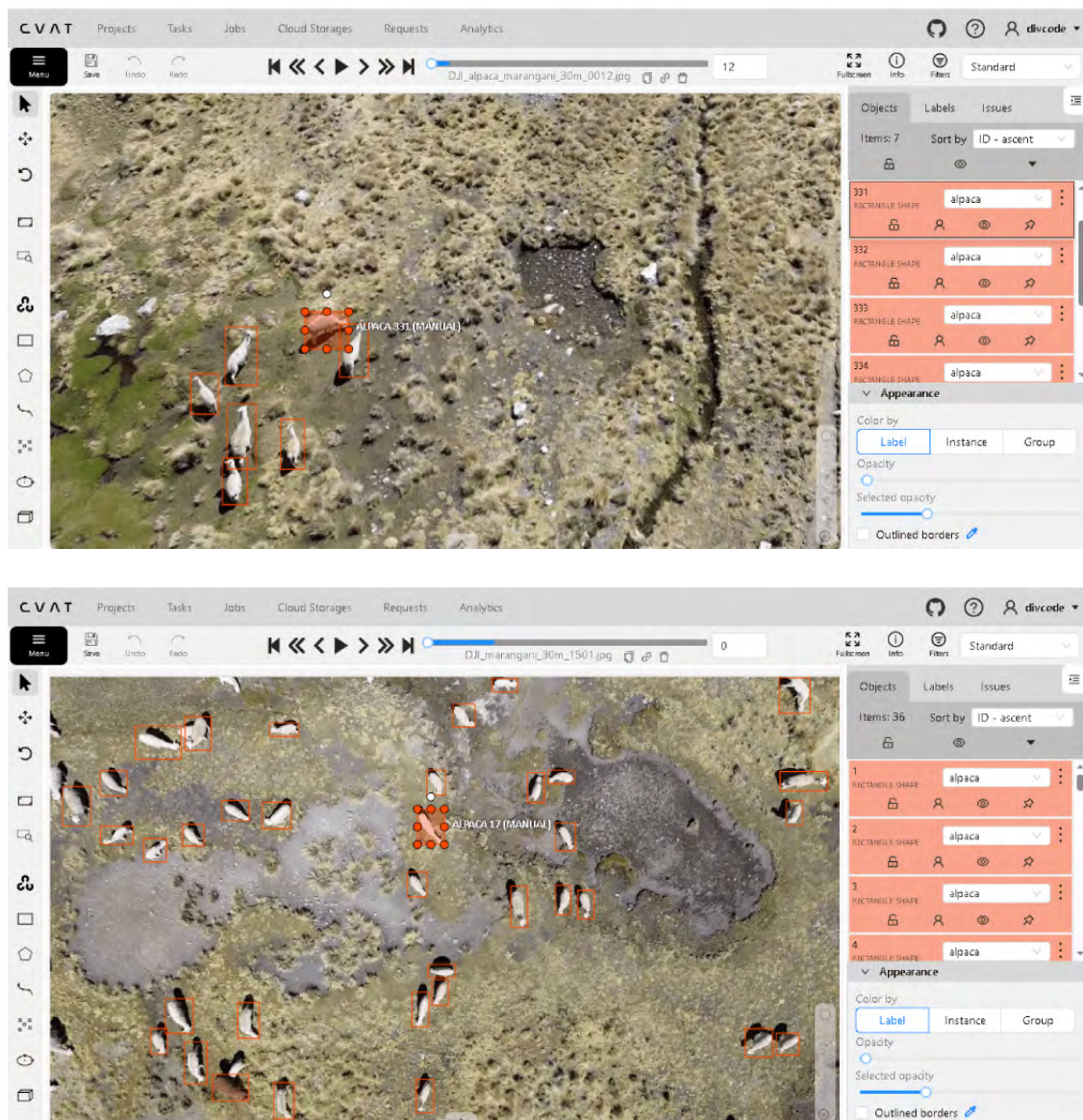


En las Fig. 3.9 y Fig. 3.10 se muestra el proceso de etiquetado de llamas y alpacas utilizando la herramienta CVAT. Se traza un cuadro delimitador alrededor de cada instancia de interés y se asigna la etiqueta correspondiente.

Figura 3.9
Etiquetado de instancias de llamas en CVAT



Figura 3.10
Etiquetado de instancias de alpacas en CVAT



Seguidamente, se realiza el balanceo de anotaciones. Como indica Oksuz et al. (2021), el desbalance de anotaciones es el principal problema de un modelo de detección de objetos, puesto que las anotaciones desbalanceadas dan mayor importancia a la clase que tiene mayor cantidad de anotaciones y menor importancia a la clase con menor cantidad de anotaciones. Una distribución equilibrada contribuye a mejorar la capacidad de generalización de los modelos. Se verificó este balanceo en los 4 datasets antes de la exportación de los mismos. Finalmente, se realiza la exportación del dataset final en formato YOLO, que tiene la siguiente estructura: `<clase><x.centro><y.centro><ancho><alto>`. Se puede observar en

la Tabla 3.7 la cantidad de anotaciones de las clases llama y alpaca. El proceso se realizó para cada dataset en cuestión; sin embargo, para fines prácticos solo se mostrará el dataset 4 (final), que une todas las anotaciones de los datasets previos.

Tabla 3.7

Dataset 4: Cantidad de anotaciones por clase

Especie	Cantidad de imágenes	Anotaciones
Llamas	1025	17008
Alpacas	672	17122
Negativos	230	0
Total	1927	34130

3.3.3. División del dataset

En tareas de detección de objetos, una sola imagen puede contener múltiples instancias de diferentes clases, lo que hace que la simple contabilización del número de imágenes por clase no refleje adecuadamente la verdadera distribución de datos. Esto se observa en la Tabla 3.7, donde la clase *alpaca*, con pocas imágenes, contiene la misma cantidad de anotaciones que la clase *llama*. Esto se debe a que cada imagen contiene una mayor densidad de instancias de esa clase.

Por ello, durante la división del dataset con un script de Python, la cantidad de anotaciones se consideró como la principal estrategia, dado que el dataset fue dividido en 3 subconjuntos: 70 % para entrenamiento, 20 % para validación y 10 % para prueba. Cada uno de ellos tiene distribuida de forma uniforme la cantidad de anotaciones de llama y alpaca para disminuir el sesgo hacia una clase específica. Este proceso se realizó para el dataset exportado en formato COCO y YOLO.

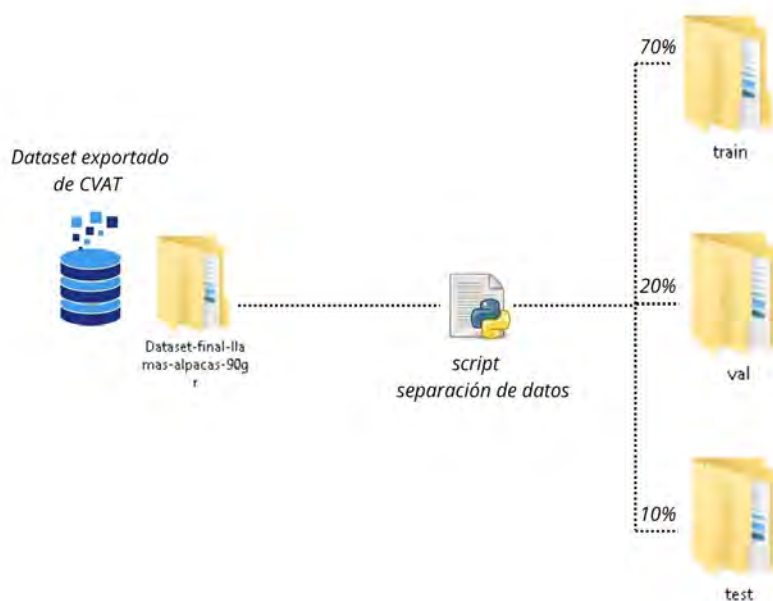
- **Train:** Subconjunto del dataset utilizado exclusivamente para el entrenamiento de los modelos YOLO, Faster R-CNN y SSD, donde se ajustan los pesos de los modelos.
- **Validation:** Subconjunto del dataset empleado para la evaluación y ajuste de hiper-

parámetros durante el proceso de entrenamiento; nos permite verificar el sobreajuste.

- **Test:** Subconjunto del dataset reservado para la evaluación final e imparcial de los modelos, midiendo su desempeño real en datos no vistos durante el entrenamiento.

Figura 3.11

División de datos para entrenamiento: 70 % train, 20 %val y 10 % test



A continuación en las Tabla 3.8, 3.9 y 3.10, se observa la distribución de anotaciones.

Tabla 3.8

Dataset de entrenamiento (70 %)

Especie	Cantidad de imágenes	Anotaciones
Llamas	699	11949
Alpacas	467	11957
Negativos	160	0
Total	1326	23906

Tabla 3.9

Dataset de validación (20 %)

Especie	Cantidad de imágenes	Anotaciones
Llamas	221	3427
Alpacas	134	3472
Negativos	46	0
Total	401	6899

Tabla 3.10
Dataset de prueba (10 %)

Especie	Cantidad de imágenes	Anotaciones
Llamas	105	1632
Alpacas	71	1693
Negativos	24	0
Total	200	3325

3.3.4. Aumento de datos

Para mejorar la capacidad de generalización del modelo en condiciones reales de prueba, se aplicaron técnicas de aumento de datos durante el entrenamiento que adoptan tecnologías más modernas. Estas transformaciones permiten simular variaciones propias del entorno de captura con dron en la zona de estudio, sin necesidad de recolectar nuevas imágenes.

Se aplicaron transformaciones orientadas a simular las variaciones geométricas generadas por el movimiento del dron durante el vuelo.

- **Escalamiento:** Permite al modelo reconocer animales a diferentes alturas de vuelo, variando su tamaño en la imagen.
- **Perspectiva:** Introduce ligeras deformaciones que simulan cambios en el ángulo de captura.
- **Rotación:** Representa rotaciones asociadas al balanceo del dron por efecto del viento.
- **Distorsión:** Simula distorsiones laterales causadas por movimientos bruscos o aceleraciones.
- **Inversión horizontal:** Aplica una inversión horizontal aleatoria, aumentando la variabilidad espacial de las detecciones.

Además, agregamos transformaciones que buscan replicar condiciones reales de agrupamiento y oclusión entre animales:

- **Borrado aleatorio:** Utilizado para simular oclusiones parciales al eliminar aleatoriamente partes de la imagen, lo que fácilmente podría representar animales parcialmente cubiertos dentro de la manada.
- **Traslación:** Desplaza los animales dentro del encuadre, lo que permite al modelo detectar individuos en los bordes de la imagen.

También se incorporaron ajustes en el espacio de color para representar la variabilidad lumínica característica del entorno:

- **Tono:** Variamos los tonos manteniendo la naturalidad de los colores.
- **Saturación:** Modificamos la saturación para representar diferencias entre pastos secos y vegetación verde (simulación de captura de datos en diferentes estaciones del año).
- **Brillo:** Simulamos condiciones de días soleados, sombras intensas y cielos nublados.

3.4. Fase 4: Entrenamiento del modelo

Durante esta fase se describirán los modelos evaluados de una y dos etapas, mostraremos los valores utilizados en aumento de datos, la configuración de hiperparámetros y mostrar el proceso de entrenamiento.

3.4.1. Modelos implementados y evaluados

Los modelos evaluados están basados de acuerdo al estado del arte de la presente investigación y, además, deben ser capaces de diferenciar objetos pequeños y tener la capacidad de generalizar datos nuevos en distintos escenarios. Finalmente, deben tener un buen equilibrio entre precisión y velocidad. Entre ellos se encuentran Faster R-CNN, SSD, YOLO11 y YOLO26 (arquitectura nueva con aparición a inicios del año 2026), siendo estas tres últimas de una sola etapa.

Faster R-CNN

En la presente investigación se implementó Faster R-CNN usando la librería Detectron2, desarrollada por Facebook AI Research. Este es un modelo de deep learning de dos etapas, y se evaluó su capacidad de precisión. Aunque ofrece una alta tasa de acierto en la localización de objetos pequeños, su arquitectura basada en una Red de Propuesta de Regiones (RPN) genera una latencia incompatible con el procesamiento en el borde (Edge Computing). Este aspecto es muy necesario para aplicaciones en tiempo real, y más si se trata de aplicaciones en dispositivos con recursos limitados, como los dispositivos móviles.

Single Shot Multibox Detector

Considerado un modelo de deep learning de una sola etapa, fue evaluado por su rapidez y equilibrio entre precisión y velocidad. Aunque presentó dificultades en la detección de ejemplares pequeños en este proyecto, en general fue el modelo que obtuvo el menor rendimiento en la etapa de entrenamiento. Para la implementación se utilizó la arquitectura SSD300 con backbone VGG16, disponible en la librería Torchvision.

YOLO11

Modelo de una sola etapa, se utilizó la versión nano de esta arquitectura. YOLO11 permite una extracción de características más profunda con un menor número de parámetros, ideal para la integración en dispositivos de borde. La ventaja es su optimización en la extracción de características para una detección más precisa. Fue el modelo que tuvo mejor desempeño en la etapa de entrenamiento, pero presentó problemas en el tiempo de inferencia para aplicaciones en tiempo real. Su implementación es de uso simplificado utilizando la librería Ultralytics.

YOLO26

Este modelo es el más óptimo teóricamente para aplicaciones en tiempo real, puesto que tiene una nueva arquitectura End to End, lanzada en enero de 2026, con una velocidad de procesamiento 43 % más rápida en la CPU. Estas mejoras se logran con la eliminación de NMS en la etapa de post-procesamiento, haciendo que el tiempo de inferencia sea más rápido. Este nuevo modelo fue introducido junto con el optimizador MuSGD, que es una combinación de Muon con SGD, inspirado en Kimi K2 de Moonshot AI para el entrenamiento de LLMs. Las pruebas de este modelo se verán más adelante. Durante el entrenamiento, fue el modelo que tuvo un desempeño sobresaliente, casi similar a YOLO11n. Al igual que su predecesor, este modelo se encuentra dentro de la librería Ultralytics.

3.4.2. Configuración de hiperparámetros

Para garantizar que nuestros modelos de detección de objetos generalicen mejor, se configuraron hiperparámetros específicos para cada uno. Se debe tener en cuenta que estos hiperparámetros son las configuraciones finales.

- **Epochs:** Indica el límite máximo de iteraciones sobre todo el dataset.
- **Batch Size:** Tamaño del lote de imágenes que procesa la GPU a la vez.
- **Image size:** La resolución a la que se redimensionan todas las imágenes aéreas antes de entrar a la red.
- **Learning Rate:** Indica la tasa de aprendizaje inicial del entrenamiento.
- **Optimizer:** Es el algoritmo de optimización que ayuda en la convergencia del modelo.
- **Patience:** Número de épocas que el modelo espera sin mejora antes de detener el entrenamiento (early stopping). Le dice al modelo: “si no mejoras, hasta aquí llegaste;

no podemos caer en overfitting”.

- **Momentum:** Ayuda a mantener la dirección de la optimización y a saltar mínimos locales.
- **Weight Decay:** Penaliza pesos demasiado grandes para que el modelo sea más simple y generalice mejor.
- **Data augmentation:** Son transformaciones que se incluyen en tiempo de ejecución durante el entrenamiento. Estas se mencionan en la Sección 3.3.4 y se incluyen para agregar diversidad de datos al modelo, ayudar con la generalización y evitar el overfitting.

En la Tabla 3.11 se muestra la configuración de hiperparámetros estructurales y de optimización, mientras que en la Tabla 3.12 se muestran la configuración de valores para data augmentation.

Tabla 3.11

Hiperparámetros de entrenamiento, optimización y arquitectura.

Hiperparameter	YOLO11	YOLO26	SSD	Faster R-CNN
Epochs	139/1000	174/1000	90/1000	80/100
Batch	32	34	16	32
Image size	640	640	640	(1280,720)
Learning rate	0.001	0.001	0.002	0.00025
Patiance	20	20	10	5
Optimizer	MuSGD	MuSGD	SGD	SGD
Weight decay	0.0005	0.0005	0.0005	0.0001
Momentum	0.937	0.937	0.9	0.937

Tabla 3.12*Hiperparámetros de data augmentation.*

Hiperparameter	YOLO11	YOLO26	SSD	Faster R-CNN
Scaling	0.6	0.6	-	0.2
Perspertive	0.001	0.001	0.001	-
Rotation (Degrees)	10.0	10.0	10.0	10.0
Distortion	2.5	2.5	-	-
Horizontal flip	0.5	0.5	0.5	0.5
Random erasing	0.4	0.4	-	0.8
Traslacion	0.1	0.1	0.1	0.1
Hue	0.015	0.015	0.01	-
Saturation	0.7	0.7	0.3	0.5
Brightness	0.4	0.4	0.2	0.4
Mixup	0.0	0.0	0.0	0.0
Mosaic	0.0	0.0	0.0	0.0
Vertical flip	0.0	0.0	0.0	0.0
Shear	2.5	2.5	2.5	-

3.4.3. Proceso de entrenamiento

Hardware utilizado

El proceso de entrenamiento de los modelos de detección de objetos se llevó a cabo en la plataforma Google Colab Pro con GPU NVIDIA Tesla T4, GPU NVIDIA L4 y GPU NVIDIA A100, que proporcionó recursos computacionales adecuados para el entrenamiento intensivo de modelos de detección de objetos. Los detalles del hardware se muestran a continuación:

GPU NVIDIA Tesla T4

- Memoria RAM de sistema : 51GB
- Memoria de GPU : 15GB
- Capacidad de almacenamiento : 235.7GB

GPU NVIDIA L4

- Memoria RAM de sistema : 53GB
- Memoria de GPU : 22.5GB
- Capacidad de almacenamiento : 235.7GB

GPU NVIDIA A100

- Memoria RAM de sistema : 167.1GB
- Memoria de GPU : 80GB
- Capacidad de almacenamiento : 235.7GB

El uso de la GPU permitió acelerar significativamente las operaciones propias de las redes neuronales convolucionales, disminuyendo el tiempo de entrenamiento y facilitando la ejecución de múltiples experimentos a lo largo del proceso de realización del presente proyecto. Este entorno fue empleado para el entrenamiento de los cuatro modelos considerados en la investigación: YOLO11, YOLO26, SSD y Faster R-CNN.

Tiempo de entrenamiento

El tiempo de entrenamiento se condicionó por diversos factores, tales como la complejidad de las arquitecturas de cada modelo utilizado, el tamaño del conjunto de datos, la resolución de entrada de las imágenes, la configuración de los hiperparámetros y, finalmente, el hardware utilizado influye en el tiempo de entrenamiento, tener en cuenta que los tiempos indicados serán de los de mayor rendimiento de cada tipo.

Se observó que Faster R-CNN no requirió muchas épocas para converger. Esto se debe a que su arquitectura incluye una Region Proposal Network que dibuja gran cantidad de bounding boxes y elige aquellos que tienen mayor confianza (entrenamiento con A100). Este modelo requiere un alto costo computacional; el entrenamiento de Faster R-CNN convergió en la época 80 con un tiempo de entrenamiento de 121 minutos. Por el contrario, los modelos de una sola etapa, como YOLO11, convergieron en la época 139 con un tiempo de 32.1 minutos (entrenamiento con GPU NVIDIA A100). También YOLO26 convergió en la época 174 con un tiempo de entrenamiento de 61.8 minutos (entrenamiento con GPU NVIDIA L4). Finalmente, el modelo SSD convergió en la época 90 con un tiempo de entrenamiento de 4h 30m (entrenamiento con NVIDIA Tesla T4), siendo el modelo con menor rendimiento.

Durante el entrenamiento se aplicó Early Stopping para controlar el entrenamiento basado

en la evaluación continua del conjunto de validación, esto a través de un umbral de paciencia, que indica que si este no mejora durante las X épocas siguientes, el modelo deja de entrenar.

Estrategias de entrenamiento

La estrategia de entrenamiento se basó en el uso de modelos preentrenados aplicando técnicas de transfer learning con el fin de adaptar las arquitecturas al problema específico de detección de camélidos sudamericanos domésticos.

Durante el proceso se consideraron los siguientes aspectos: primero, la inicialización de los modelos con pesos preentrenados; segundo, el ajuste de hiperparámetros; tercero, la incorporación de técnicas de aumento de datos para mejorar la capacidad de generalización del modelo; y finalmente, la aplicación de early stopping para evitar overfitting.

3.5. Fase 5: Evaluación del modelo

En esta fase se observará la comparación de las métricas de rendimiento definidas en la Sec. 2.2.7 de todos los modelos, así como la comparación con el estado del arte y el análisis de resultados a priori antes de la implementación final de la solución tecnológica en campo. El objetivo es determinar qué modelo es el adecuado para la implementación en un entorno de procesamiento en tiempo real, priorizando un equilibrio entre precisión y velocidad (FPS y tiempo de inferencia). Estas evaluaciones fueron realizadas con el conjunto de datos de *validación* y *test*.

Para esta evaluación se emplearon métricas usadas en tareas de detección de objetos, como precisión, recall, mAP@50 y mAP@50-95. Estas métricas son un estándar para analizar tareas de detección. Recordemos que la precisión mide la proporción de detecciones correctas respecto a todas las detecciones realizadas; el recall evalúa la capacidad del modelo para detectar todos los objetos presentes en un fotograma; el mAP@50 considera la precisión

media promedio con IoU al 50 %, y mAP@50-95 considera el IoU dentro del rango de 50 a 95 %.

Los resultados obtenidos para los modelos YOLO11, YOLO26, SSD y Faster R-CNN se presentan en la Tabla 3.13, donde se muestran las métricas de desempeño con datos de validación, y en la Tabla 3.14, donde se muestran las métricas de desempeño con datos de test.

Tabla 3.13

Resultados de modelos evaluados con datos de validación

Modelo	Precisión	Recall	mAP@50	mAP@50-95
YOLO11n	0.850	0.875	0.922	0.556
YOLO26n	0.858	0.864	0.919	0.555
Faster R-CNN	0.827	0.870	0.868	0.496
SSD	0.715	0.413	0.528	0.191

Tabla 3.14

Resultados de modelos evaluados con datos de prueba

Modelo	Precisión	Recall	mAP@50	mAP@50-95
YOLO11n	0.886	0.904	0.932	0.600
YOLO26n	0.877	0.886	0.925	0.596
Faster R-CNN	0.853	0.875	0.884	0.513
SSD	0.713	0.469	0.562	0.208

Se observa claramente que el modelo YOLO11n presenta el mejor desempeño en las 4 métricas principales para los datos de test, logrando un mAP50 y mAP50-95 de 93.2 % y 60.0 %, respectivamente; seguido del modelo YOLO26n, que presenta un desempeño muy cercano con mAP50 y mAP50-95 de 92.5 % y 59.6 %, respectivamente; como tercer modelo está Faster R-CNN con mAP50 de 88.4 % y mAP50-95 de 51.3 %. Finalmente, el modelo SSD presenta un desempeño menor en todas las métricas evaluadas, quedando con un mAP50 y mAP50-95 de 56.2 % y 20.8 %, respectivamente.

Para complementar, se analizó la evolución de las métricas durante el entrenamiento, donde se observa una convergencia estable en la familia YOLO, SSD y Faster R-CNN.

Figura 3.12

Métricas de rendimiento (YOLO11)

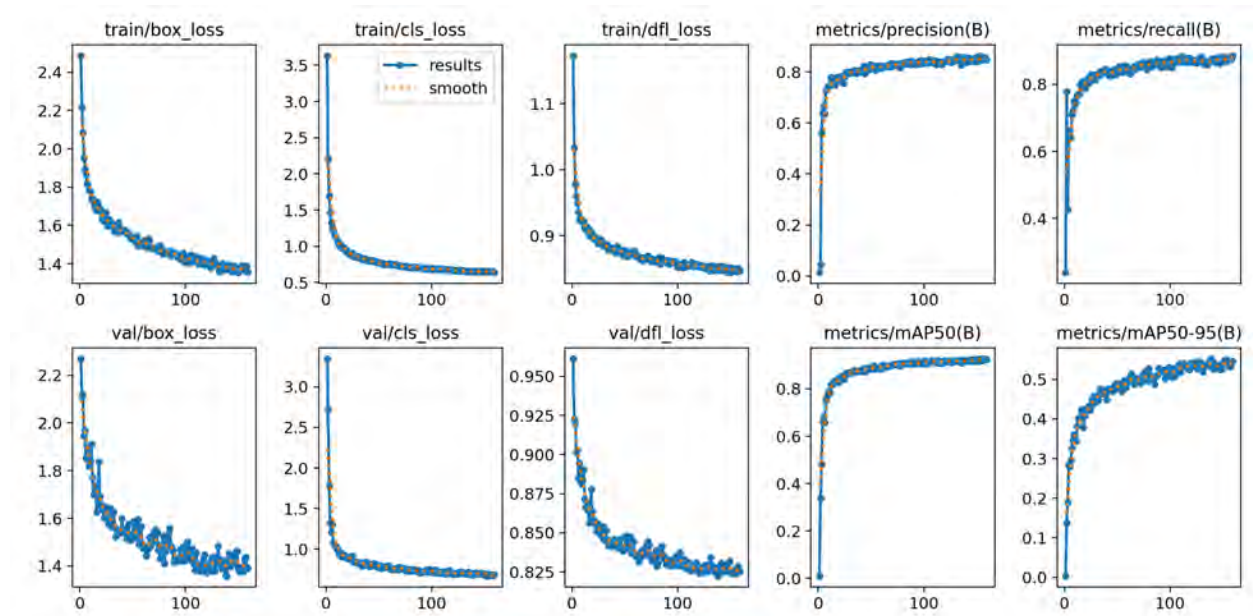


Figura 3.13

Métricas de rendimiento (YOLO26n)

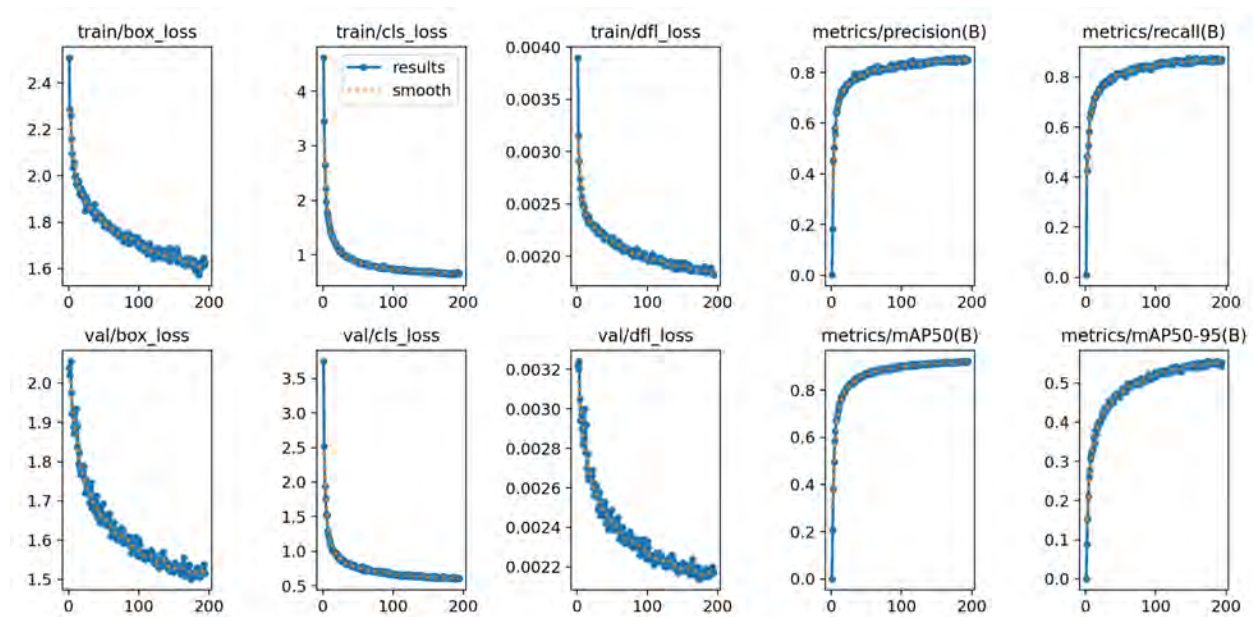


Figura 3.14
Métricas de rendimiento (Faster R-CNN)

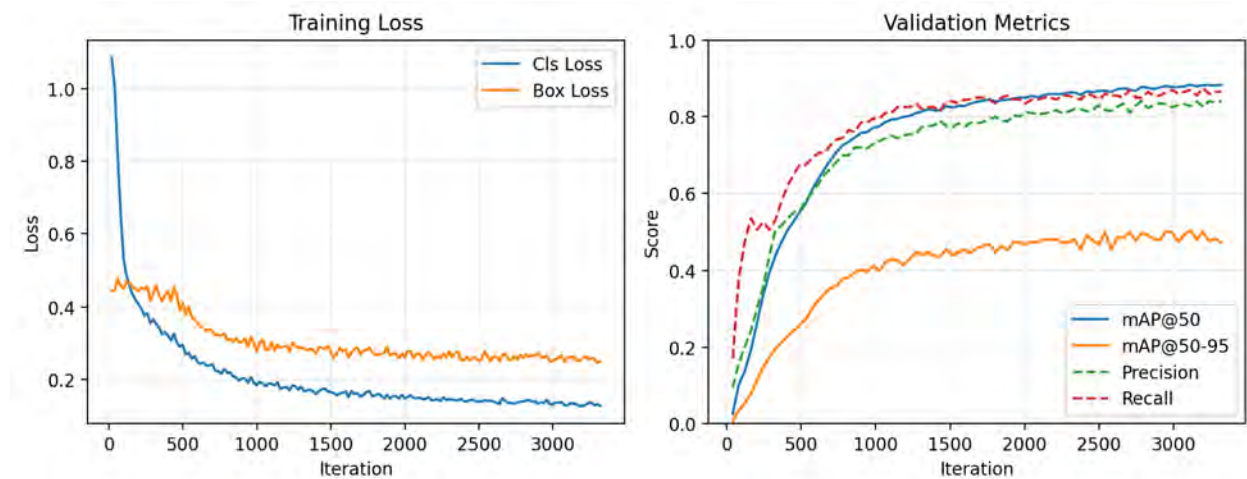
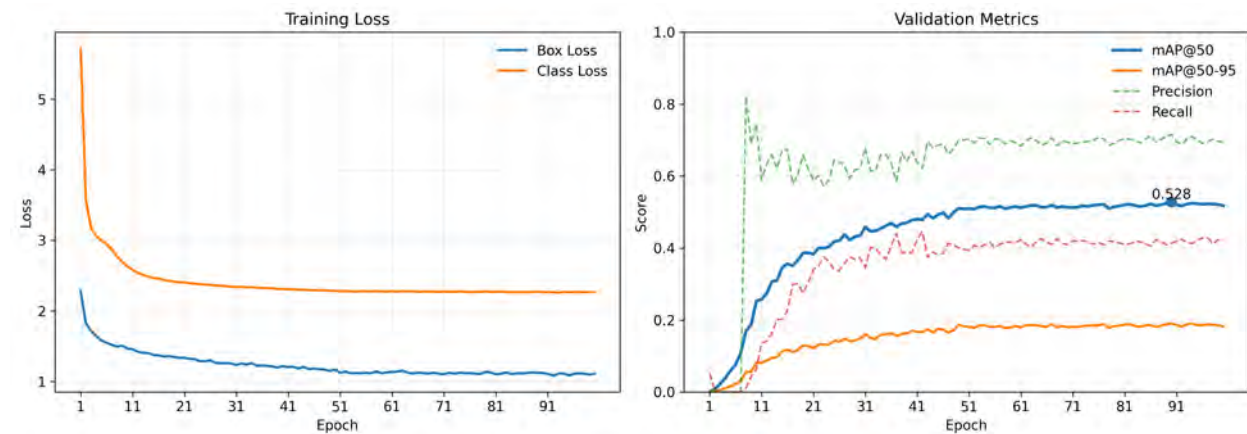


Figura 3.15
Métricas de rendimiento (SSD)



De las Figuras 3.12, 3.13, 3.15 y 3.14, se puede observar que las pérdidas de box y pérdida de clasificación tiene una distribución similar para los datos de validación y test, un indicativo claro de que no hay una aparente presencia de overtiffing, esto se puede corroborar con las métricas de las Tablas 3.13 y 3.14 donde se observa los valores de las métricas de validación y prueba son muy cercanas, además, que cada modelo fue entrenado con early stopping.

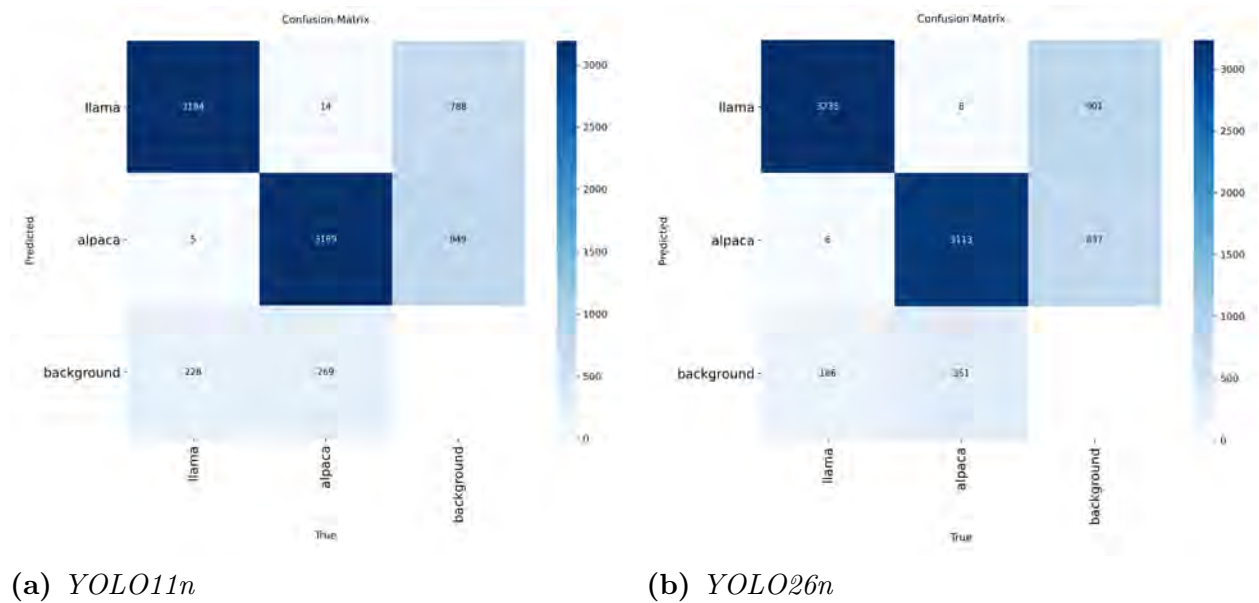
Análisis de matriz de confusión

Adicionalmente se realizó el análisis cuantitativo, evaluando el comportamiento de los modelos a través de las matrices de confusión. En la Fig. 3.16a y Fig. 3.16b se muestra a matriz

de confusión del modelo de YOLO11n y YOLO26n, respectivamente, correspondiente a los dos modelos con mejor desempeño.

Figura 3.16

Matriz de confusión para YOLO11 y YOLO26



Los resultados de las matrices de confusión vistas líneas arriba nos proporcionan un comportamiento bastante notable para los datos de validación que se evaluaron durante el entrenamiento.

En general, tanto YOLO11 como YOLO26 distinguen bien a la llama y a la alpaca. Esto se puede observar en la diagonal principal y en la diagonal secundaria de la matriz de confusión. El principal desafío de los modelos no radica en confundir la llama y la alpaca entre sí, sino en confundir elementos del entorno como otros animales.

Los modelos generan falsos positivos al predecir la existencia de llamas y alpacas en zonas donde corresponde exclusivamente fondo (background). Este problema será abordado con más detalle en la etapa de validación durante las pruebas de campo, pero podemos saber a priori que el modelo confunde el paisaje pedregoso, las variaciones de formas de bofedales, rocas y aglomeraciones de ichus que pueden parecerse a algunos de nuestros objetos de interés. Sin embargo, es importante mencionar que también se debe a la inclusión de imágenes

negativas en el proceso de entrenamiento (abordado en la Sec. 3.3.1). Esto se realizó con el objetivo de enseñar a los modelos a no detectar cualquier tipo de animal o fondo que se encuentre en el entorno como si fueran las clases de interés, para dar solución a un problema común en modelos basados en redes neuronales convolucionales. Finalmente, se observa que el modelo genera falsos negativos donde omite detectar cierta cantidad de llamas y alpacas, clasificándolos como fondo. Esto se debe a la redimensión de las imágenes, donde ciertos animales pueden padecer de mimetismo con el terreno por la baja saturación, generalmente en capturas cenitales o de -90 grados.

Selección del mejor modelo

Como último punto a tratar es la elección del modelo con mejor desempeño para la etapa de construcción de la solución tecnológica. Dado el contexto, y considerando que los modelos YOLO11n y YOLO26n presentan el mejor desempeño, se realizó un análisis más detallado para la selección final del modelo más idóneo que cumpla con los requerimientos de procesamiento en tiempo real en un dispositivo móvil. Para cumplir con este propósito se realizó la conversión de los modelos a TensorFlow Lite, con redimensión de entrada de imagen a 320×320 y cuantización de punto flotante de 32 bits a 16 bits. Este proceso sacrifica un poco de precisión, de forma imperceptible para nuestro proyecto.

Las evaluaciones se aplicaron para calcular el *tiempo de inferencia* y los *FPS*. Las pruebas se desarrollaron en una GPU NVIDIA Tesla T4 de Google. Como entrada se utilizó un video capturado con el UAV DJI Mini 3 a 4K con 30 FPS, con un umbral de 200 fotogramas para la evaluación. El pseudocódigo 2 muestra el proceso de implementación para la evaluación de tiempo de inferencia y FPS con un modelo previamente convertido en TFLite y los resultados se muestran en la Tabla 3.15.

Algorithm 2 Benchmark de latencia y FPS del modelo TFLite

```
1: Entrada: Ruta del modelo TFLite model_path, ruta del video video_path, número máxi-  
mo de frames max_frames  $\leftarrow 200$   
2: Salida: Tupla (latencia_promedio_ms, fps)  
  
3: Cargar intérprete de TFLite desde model_path  
4: Asignar tensores de entrada y salida  
5: Obtener dimensiones de entrada input_shape (alto, ancho)  
  
6: Inicializar captura de video cap  $\leftarrow \text{VideoCapture}(\text{video\_path})$   
7: Inicializar lista vacía times  $\leftarrow []$   
8: frame_count  $\leftarrow 0$   
  
9: while cap está abierto y frame_count < max_frames do  
10:   ret, frame  $\leftarrow \text{cap.read}()$   
11:   if no ret then  
12:     break  
13:   end if  
  
14:   frame_resized  $\leftarrow \text{resize}(\text{frame}, (\text{input\_shape.ancho}, \text{input\_shape.alto}))$   
15:   Convertir frame_resized de BGR a RGB  
16:   Normalizar: frame_resized  $\leftarrow \text{frame\_resized}/255,0$  (float32)  
17:   Añadir dimensión de lote: frame_resized  $\leftarrow \text{expand\_dims}(\text{frame\_resized}, \text{axis} = 0)$   
  
18:   start  $\leftarrow \text{tiempo\_actual}()$   
19:   Asignar frame_resized al tensor de entrada del intérprete  
20:   Ejecutar inferencia: interpreter.invoke()  
21:   end  $\leftarrow \text{tiempo\_actual}()$   
  
22:   latency_ms  $\leftarrow (\text{end} - \text{start}) \times 1000$   
23:   Agregar latency_ms a times  
24:   frame_count  $\leftarrow \text{frame\_count} + 1$   
25: end while  
  
26: cap.release()  
27: avg_latency  $\leftarrow \text{media}(\text{times})$   
28: fps  $\leftarrow 1000/\text{avg\_latency}$   
29: return (avg_latency, fps)
```

Tabla 3.15*Resultados de modelos evaluados con datos de validación*

Model	format	Frames	Time inference (latencia)	FPS
YOLO11n	.tflite	200	55.36 ms.	18.06
YOLO26n	.tflite	200	38.93 ms.	25.69

A partir de los resultados obtenidos en las pruebas de inferencia y FPS, se observa que el modelo más adecuado para la implementación en aplicaciones en entornos de edge computing

es YOLO26n. Este modelo tiene un mejor equilibrio entre precisión, velocidad y bajo costo computacional, por lo que será el modelo que estaremos evaluando en las pruebas de campo.

3.6. Fase 6: Implementación de la solución tecnológica

En esta fase se integra el modelo de detección YOLO26n previamente entrenado, donde realizaremos la materialización de la solución tecnológica propuesta para su funcionamiento en tiempo real, considerando las restricciones de hardware. La ejecución del sistema funciona de forma local en el dispositivo móvil sin conexión a internet, reflejando la realidad de las zonas altoandinas.

3.6.1. Arquitectura del sistema CamelidCounter

Se implementó una arquitectura de procesamiento en el borde (conocida como Edge Computing). Este proceso se puede observar en la siguiente Fig. 3.17.

Figura 3.17

Arquitectura de la solución tecnológica



Donde el smartphone actúa como la unidad central de inteligencia (procesamiento). El flujo de datos inicia en el UAV (DJI Mini 3), el cual transmite el video en tiempo real mediante el protocolo DJI OcuSync 2.0 hacia el control remoto. Este, a su vez, entrega el flujo crudo al aplicativo móvil, donde se ejecutan la inferencia neuronal y el tracking antes de mostrar los resultados en la interfaz de usuario.

Para este punto, el modelo entrenado en PyTorch (.pt) fue convertido a un modelo TensorFlow Lite (.tflite) con cuantización de punto flotante de 16 bits (Float16). Esta técnica permite reducir el peso del modelo aproximadamente en un 50 % y ayuda a acelerar la inferencia en dispositivos móviles. Esta conversión sacrifica cierta precisión, que se considera de menos del 1 %.

3.6.2. Desarrollo de la Aplicación

El desarrollo de la aplicación móvil se realizó de forma nativa para maximizar el rendimiento del hardware. Este fue implementado con el objetivo de tener un sistema funcional, minimalista y accesible para el usuario, donde se pueda observar desde la recepción del video hasta la visualización de los resultados, cumpliendo los requisitos funcionales y no funcionales planteados en la Sec. 3.1.4 y en la Sec. 3.1.5, respectivamente.

3.6.3. Tecnologías utilizadas

La implementación del sistema se realizó utilizando diversas tecnologías. Se optó por la construcción de un aplicativo móvil para sistemas operativos Android. Entre las principales herramientas utilizadas se encuentran:

- Kotlin: lenguaje de programación utilizado para la implementación del aplicativo de forma nativa en Android con XML, puesto que tiene una excelente interoperabilidad con librerías de bajo nivel de Android (ir a Kotlin).
- Android Studio: entorno de desarrollo integrado para Android, con acceso a diversos dispositivos virtuales para la visualización de las interfaces (ir a Android Studio).
- TensorFlow Lite: encargado de cargar el modelo YOLO26n y de ejecutar modelos de detección de objetos en dispositivos móviles a través de tensores de entrada y salida (ir a TensorFlow Lite).
- DJI Mobile SDK: utilizado para la recepción de video en tiempo real desde el dron. Esta herramienta conlleva la utilización de Kotlin y Java en el desarrollo de la aplicación debido a la documentación del MSDK (ir a DJI MSDK).

3.6.4. Recepción de video MSDK

La recepción del video en tiempo real se implementó mediante el uso del DJI Mobile SDK (MSDK), disponible en la documentación de DJI, el cual permite establecer comunicación directa entre el dron y el dispositivo móvil a través del control remoto RC-N1. Mediante este SDK se obtiene el flujo de video proveniente de la cámara del dron, el cual es procesado frame a frame dentro de la aplicación.

Este mecanismo permite trabajar en condiciones offline, ya que la transmisión de video se realiza de manera local entre el dron y el dispositivo, garantizando la disponibilidad del sistema en zonas sin conectividad. Por lo tanto, el tiempo de inferencia es mucho más rápido debido a la inexistencia de un puente o almacenamiento en la nube. Sin embargo, es importante mencionar que el primer vuelo requiere conexión a internet para consumir de forma efectiva la API de desarrolladores que se crea.

3.6.5. Procesamiento y conteo en tiempo real

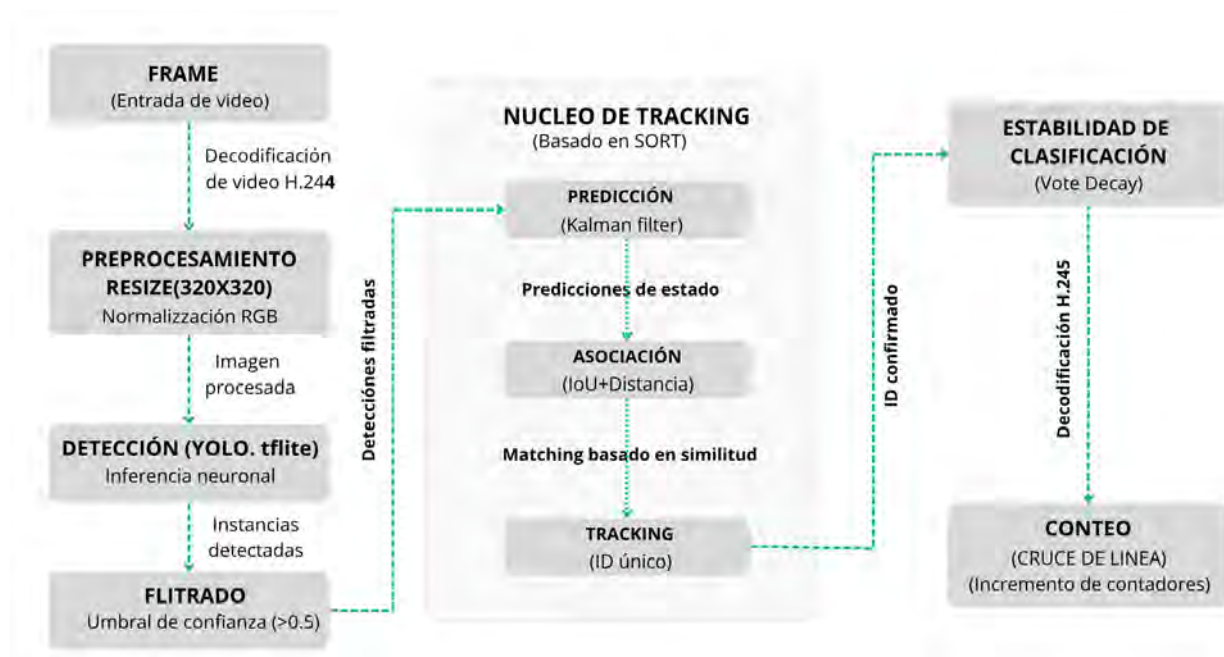
Este paso es necesario porque constituye la base de la solución tecnológica para el conteo y evitar duplicidad.

Pipeline de procesamiento

El flujo de procesamiento en tiempo real se estructura como una secuencia de etapas encadenadas que transforman el video recibido en información cuantitativa útil. Este pipeline inicia con la adquisición del frame y finaliza con el conteo acumulado de animales detectados. En la Fig. 3.18 se muestra el proceso completo del pipeline.

Figura 3.18

Pipeline de procesamiento y seguimiento en tiempo real



Detección de objetos

Como se mencionó en la Sec. 3.5, el modelo utilizado en pruebas de campo será YOLO26n, previamente entrenado y posteriormente convertido al formato TFLite. Cada frame capturado es redimensionado al tamaño de entrada del modelo (320×320) y normalizado a RGB antes de ser procesado. Como resultado, el modelo genera un conjunto de tensores de la forma $[1, 300, 6]$, donde 1 representa el batch, 300 indica que el modelo siempre genera 300 filas, que representan los 300 objetos que cree haber visto en sus detecciones, y el número 6 está representado por:

- 👉 Bounding boxes: $[x1, y1, x2, y2]$, donde $x1, y1, x2, y2$ son flotantes (float).
- 👉 Nivel de confianza: punto flotante.
- 👉 ID de la clase del objeto detectado: “llama” ó “alpaca”.

Posteriormente, se aplica un proceso de filtrado basado en un umbral de confianza configurado en el aplicativo, descartando aquellas detecciones que no cumplen con el valor seleccio-

nado. Esto nos permite reducir los falsos positivos y mejorar la calidad de las detecciones que serán utilizadas en el proceso de seguimiento.

Algoritmo de tracking

Para el proceso de conteo en tiempo real, es imprescindible aplicar seguimiento de objetos. La razón es perceptible: el problema no se simplifica en volar el dron a una altura lo suficientemente alta donde se observe todo el hato de camélidos y contarlo. Esto sería sencillo si se tratase de cantidades pequeñas de hatos, porque a mayor altura la visibilidad de cada objeto de interés decae significativamente. Más aún cuando la imagen se tiene que redimensionar para ingresar a un modelo de detección de objetos, lo que afecta directamente la precisión de la detección y deriva en un conteo deficiente. Por lo tanto, una estrategia real es sobrevolar el UAV de un extremo a otro de forma horizontal a una altura razonable donde se pueda observar con claridad los objetos de interés. De esta manera, podemos seguir a cada animal con un identificador único (ID) a lo largo del tiempo, siempre que se encuentre dentro del campo de visión del UAV. Seguidamente, se dibuja una línea digital para evitar el conteo duplicado de animales, de modo que si un animal cruza varias veces, sea contado solo una vez. Dadas las casuísticas anteriores, se implementó un algoritmo de seguimiento multiobjeto en tiempo real basado en técnicas de asociación de datos, inspirado en el enfoque SORT (Simple Online and Realtime Tracking), pero adaptado al contexto específico del conteo aéreo de camélidos. En esta sección se utilizará el Algoritmo de Seguimiento por Asociación de Datos (Data Association Tracking).

Modelado del estado: Cada objeto rastreado es representado mediante un vector de estado de ocho dimensiones:

$$X = (cx, cy, w, h, v_x, v_y, v_w, v_h) \quad (3.1)$$

Donde (cx, cy) representan el centro del objeto, (w, h) su tamaño y (v_x, v_y, v_w, v_h) las velo-

ciudades.

Predicción con filtro de Kalman: Para cada frame, el sistema utiliza un filtro de Kalman que permite predecir la posición futura de cada objeto. Esta predicción es fundamental para mantener el seguimiento en presencia de oclusiones temporales o fallos en la detección.

Asociación de detecciones: La asociación entre las detecciones actuales y las instancias de objeto previamente rastreadas se realiza utilizando una función de similitud híbrida que combina dos métricas: IoU (Intersection over Union), con un peso del 70 %, y la distancia euclidiana entre centroides, con un peso del 30 %. Estos pesos pueden cambiar de acuerdo con nuestras necesidades, como se explica en la investigación de Yuan et al. (2022).

La función utilizada es:

$$Score = 0,7 \times IoU + 0,3 \times \left(\frac{1}{1 + Distancia/100} \right) \quad (3.2)$$

Cada detección se asigna al objeto con mayor puntaje, siempre que este supere un umbral mínimo.

Estabilidad de clasificación: Para reducir la variabilidad en la clasificación, se implementa un sistema de votación con decaimiento temporal. Cada objeto mantiene un historial de probabilidades por clase, donde las observaciones más recientes tienen mayor influencia. Dicho de otras palabras, cuando una instancia se solapa con otra o se pierde por temas de mimetismo, el sistema mantiene un voto para esa instancia o una probabilidad de que ese objeto siga ahí. El decay del voto reduce ese valor gradualmente. Si el objeto no reaparece tras cierto tiempo, el sistema olvida el objeto para no malgastar recursos.

Pseudocódigo del algoritmo de tracking

A continuación, se presenta el pseudocódigo del algoritmo implementado:

Algorithm 3 Seguimiento multi-objeto para conteo de camélidos

```
1: Entrada: Lista de detecciones  $D_t$  (cada una con rectángulo, clase, confianza), línea de
   conteo  $L$ , alto del frame  $H$ 
2: Salida: Lista de objetos a visualizar, conteos actualizados  $C_{llama}, C_{alpaca}$ 

3: Inicializar
4:  $Tracks \leftarrow \emptyset$ 
5:  $C_{llama} \leftarrow 0, C_{alpaca} \leftarrow 0$ 
6:  $Contados \leftarrow \emptyset$ 
7: for cada frame  $t$  do
8:   for cada  $track$  en  $Tracks$  do
9:      $track.guardarPosicionAnterior()$ 
10:     $track.estado \leftarrow \text{KalmanPredict}(track.estado)$   $\triangleright (cx, cy, w, h, vx, vy, wx, wh)$ 
11:     $track.votos \leftarrow track.votos \times \delta$   $\triangleright$  Decaimiento de votos ( $\delta < 1$ )
12:   end for
13:    $Asoc, D_{libres}, T_{libres} \leftarrow \text{Asignar}(Tracks, D_t)$   $\triangleright \text{score} = 0.7 \cdot \text{IoU} + 0.3 / (1 + \text{dist} / 100)$ 
14:   for cada  $(track, det)$  en  $Asoc$  do
15:      $track.actualizar(det)$   $\triangleright$  Kalman update, incrementa hits, actualiza votos
16:   end for
17:   for cada  $det$  en  $D_{libres}$  do
18:     if  $det.centroY > 0,8 \times H$  then  $\triangleright$  Zona de inhibición inferior
19:       skip  $\triangleright$  No registrar objetos nuevos muy abajo
20:     else
21:        $Tracks \leftarrow Tracks \cup \{\text{CrearNuevoTrack}(det)\}$ 
22:     end if
23:   end for
24:   for cada  $track$  en  $Tracks$  do
25:     if  $track$  en  $T_{libres}$  then
26:        $track.inactivos \leftarrow track.inactivos + 1$ 
27:     end if
28:   end for
29:    $Tracks \leftarrow Tracks - \{track \mid track.inactivos > umbral\_perdido\}$ 
30:    $Tracks \leftarrow Tracks - \{track \mid track.centroY > 0,95 \times H\}$   $\triangleright$  Delete x salida de view
31:   for cada  $track$  en  $Tracks$  do
32:     if  $track.id \notin Contados$  y  $track.inactivos \leq 2$  then  $\triangleright$  Solo contar si activo recién
33:        $cruce\_normal \leftarrow (track.centroY_{prev} < L \text{ y } track.centroY_{act} \geq L)$ 
34:        $cruce\_rapido \leftarrow (track.centroY_{inicial} < L \text{ y } track.centroY_{act} \geq L)$ 
35:       if  $cruce\_normal$  o  $cruce\_rapido$  then
36:         if  $track.clase = llama$  then
37:            $C_{llama} \leftarrow C_{llama} + 1$ 
38:         else
39:            $C_{alpaca} \leftarrow C_{alpaca} + 1$ 
40:         end if
41:          $Contados \leftarrow Contados \cup \{track.id\}$ 
42:       end if
43:     end if
44:   end for
45:   return  $\{track \mid track.confirmado \text{ y } track.inactivos \leq 1 \text{ o } track.inactivos = 0\}$ 
46: end for
```

Lógica de conteo

El conteo de llamas y alpacas se realiza mediante el análisis del movimiento de los objetos rastreados respecto a una línea digital dibujada en la pantalla del aplicativo. La ubicación de esta línea puede ser configurada a voluntad.

Se considera que una llama o alpaca ha sido contabilizada cuando su centro geométrico cruza dicha línea en dirección vertical, como se puede observar en la Fig. 3.20. Es decir, se considera un conteo válido si el centro del animal en el instante anterior y_{t-1} estaba por encima de la línea, y en el instante actual y_t está por debajo, cumpliendo la siguiente condición:

$$y_{t-1} < y_{\text{linea}} < y_t \quad (3.3)$$

Adicionalmente, se implementa una zona de inhibición en la parte inferior de la pantalla, limitada por la línea de conteo. Esta zona evita que los animales que aparecen debajo de la línea sean contados, porque se considera que ya fueron contabilizados. Como el UAV siempre avanza hacia adelante para cubrir todo el hato de camélidos, los animales abandonan el campo de visión. En síntesis, el conteo se basa en un evento de cruce de línea y no en la presencia continua de un objeto.

Finalmente, el conteo total se obtiene a partir del número de identificadores únicos que han cumplido la condición de cruce, diferenciando además entre las clases detectadas (llamas y alpacas).

3.6.6. Visualización de resultado

Los resultados del procesamiento son presentados al usuario en tiempo real a través de la interfaz del aplicativo CamelidCounter. Las detecciones se visualizan mediante bounding boxes superpuestas sobre el video en cada instancia de objeto, acompañadas de etiquetas

que indican la clase del objeto (llama o alpaca), las cuales se actualizan dinámicamente conforme avanza el procesamiento del video.

Además, para cada clase, los colores de los bounding boxes son diferenciados. Asimismo, si algún animal fue contado, el color de su cuadro cambia. En la parte superior izquierda de la pantalla se muestra un contador acumulativo en tiempo real que desglosa el total por especie. El aplicativo es bastante intuitivo para que cualquier persona pueda hacer uso de él.

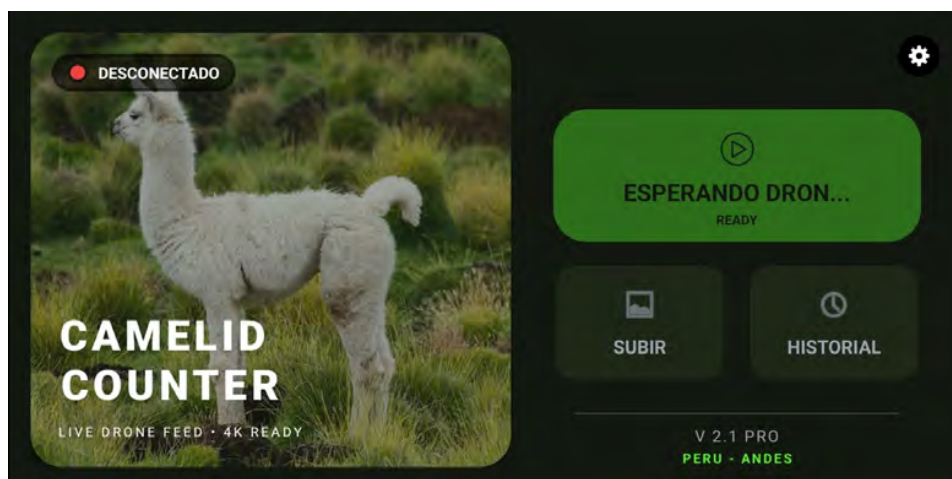
Resultados de los requisitos funcionales

Las funcionalidades principales del sistema se presentan a continuación, evidenciando el cumplimiento de los requisitos funcionales establecidos en la Sec. 3.1.4.

- **Interfaz principal:** permite acceder a las distintas funcionalidades del sistema. Aquí se realiza la conexión al dron, la configuración de vuelo, la carga de un video para realizar el conteo de forma asíncrona y el acceso al historial de vuelos (ver Fig. 3.19).

Figura 3.19

Interfaz principal de la aplicación

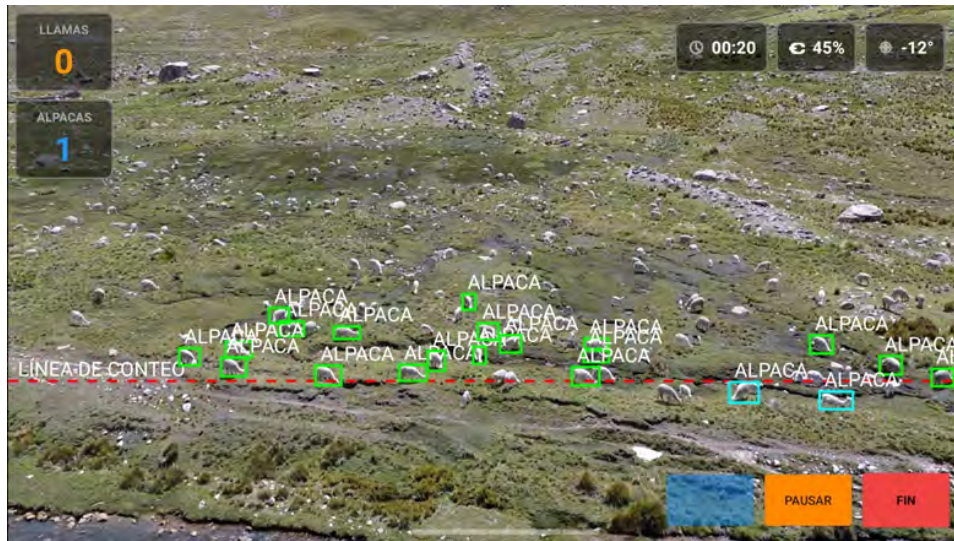


- **Interfaz de visualización:** permite la visualización en tiempo real de la cantidad de llamas, la cantidad de alpacas, el tiempo de vuelo, el porcentaje de batería del dron, el ángulo de inclinación del gimbal (cámara), la altura de vuelo y la opción de

modificación de brillo (ver Fig. 3.20).

Figura 3.20

Interfaz durante el vuelo



- **Interfaz de reporte de vuelo:** se visualiza el detalle del último vuelo realizado. Se muestra después de culminar un vuelo o al abrir un vuelo dentro del historial de vuelos. Aquí se muestra la cantidad total de llamas y alpacas contabilizadas durante el vuelo, la fecha, la hora y la duración del vuelo (ver Fig. 3.21).

Figura 3.21

Interfaz de reporte un vuelo



- **Interfaz de historial de vuelos:** en este apartado se observa el almacenamiento de vuelos desde el más reciente hasta el más antiguo (persistencia de datos). Cada vuelo cuenta con un resumen de su contenido, desde donde también se puede acceder al reporte de cualquier vuelo seleccionado. Aclaramos que no se almacena el video, solo

los datos del conteo realizado. (ver Fig. 3.22).

Figura 3.22

Interfaz de historial de vuelos



- **Interfaz de configuración del sistema:** aquí se puede configurar parámetros del modelo, de visualización y de tracking. Algunas opciones son: confianza del modelo de detección, salto entre frames, modelo a utilizar, línea de conteo, búsqueda del siguiente píxel, etc. (ver Fig. 3.23 y 3.24).

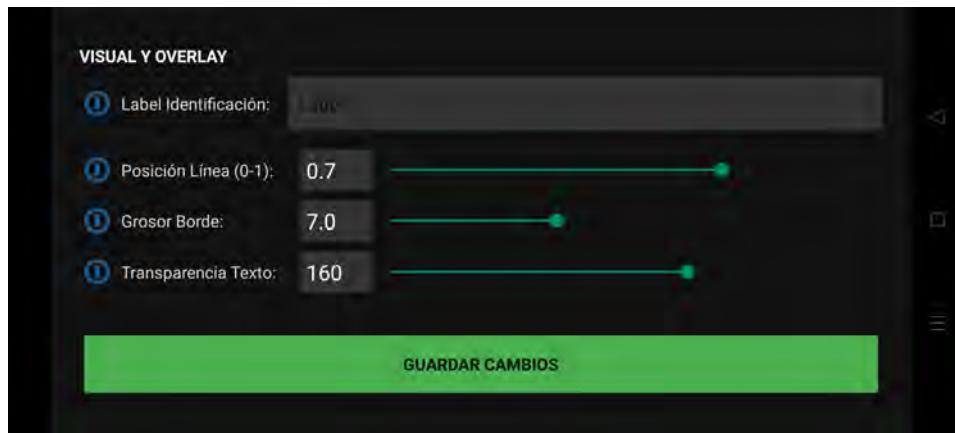
Figura 3.23

Interfaz de modificación de parámetros de conteo



Figura 3.24

Parámetros de configuración de visualización



Resultados de los requisitos no funcionales

En cumplimiento a los requisitos no funcionales establecidos en la Sec. 3.1.5, se obtuvieron los siguientes resultados:

- La ejecución del modelo de detección se realiza completamente en el dispositivo móvil, sin requerir conexión a internet durante la operación.
- Con el objetivo de garantizar un procesamiento fluido en dispositivos móviles, se incorporaron mecanismos de optimización, como el uso de TensorFlow Lite (ver Tabla 3.15). Estas medidas permitieron obtener un desempeño adecuado para la ejecución en tiempo real del sistema.
- El modelo fue convertido y optimizado utilizando TensorFlow Lite, reduciendo el tamaño del archivo y el consumo de recursos computacionales, este proceso está descrito en la Sec. 3.5.
- La latencia observada entre la captura del fotograma y la visualización de las detecciones es suficientemente baja para permitir una interacción en tiempo real (ver Tabla 3.15).

- El modelo de detección alcanzó una de precisión de $\text{mAP@50} = 95\%$, como se describe en la Sec. 3.5, validando su buen desempeño en la aplicación en tiempo real.

3.7. Fase 7: Despliegue y validación en campo

La validación de la solución tecnológica se llevó a cabo mediante pruebas en condiciones reales de pastoreo, con el objetivo de poner a prueba el funcionamiento de la solución tecnológica CamelidCounter. Para ello, días antes de la validación de campo, se realizaron pruebas de vuelo en el Complejo Arqueológico de Chincheros para la detección de llamas. Una vez validadas las pruebas de vuelo integradas con la aplicación, se elaboró el cronograma de ejecución de la validación en campo en condiciones reales.

Las zonas geográficas consideradas para las pruebas de campo presentan características distintas, mencionadas en la Sec. 1.5 y Sec. 3.1.2: la comunidad de Phinaya (distrito de Pitumarca), correspondiente a un entorno de pastoreo extensivo de alpacas en terreno abierto, y la comunidad de Hanccollahua (distrito de Espinar), donde se realizó el conteo automatizado de alpacas en cerco perimétrico y el conteo de llamas en campo abierto durante el pastoreo extensivo.

Las pruebas de vuelo in-situ se realizaron con el apoyo del ingeniero zootecnista Abel Ccama Mendigure, con CIP N° 232458, quien ayudó con las coordinaciones ganaderas y efectuó el conteo manual de los animales en simultáneo con la solución tecnológica propuesta CamelidCounter, permitiendo así una validación de campo entre ambos métodos. Este procedimiento constituye el referente de “verdad de campo” para la evaluación del sistema. Información más detallada sobre el desarrollo de las pruebas, los procedimientos aplicados y los resultados obtenidos se presenta en el documento adjunto incluido en el Anexo 4.

3.7.1. Condiciones de vuelo

Las evaluaciones se realizaron bajo condiciones reales de operación, que representan un escenario complejo. A continuación, los factores a considerar in-situ:

- Altitud de vuelo: 4300 a 4800 msnm.
- Iluminación variable, radiación solar intensa con sombras marcadas.
- Temporada de verano con vientos fuertes y lluvias.
- Escenario con presencia de rocas blancas, vegetación irregular con presencia de ichu, bofedales y dispersión de animales en campo abierto.
- Altura de vuelo: entre 20 y 40 metros.
- Procesamiento en tiempo real (offline).

3.7.2. Validación en conteo de alpacas (Comunidad de Phinaya)

Condiciones de evaluación

- Las evaluaciones se realizaron en zona de pastoreo ubicada a 4,800 m s.n.m.
- Los vuelos se efectuaron entre las 12:00 y las 14:00 horas, a alturas comprendidas entre 20 y 40 m sobre el terreno, bajo condiciones de iluminación variable, radiación solar intensa y presencia de vientos característicos de la temporada de verano, como se muestra en la Fig. A.3.1. del Anexo 3.
- El entorno de evaluación presentó pendientes pronunciadas, la presencia de rocas blancas que dieron un contraste excesivo a la cámara del dron, bofedales y vegetación altoandina.

- Las alpacas se encontraban bajo un sistema de pastoreo extensivo, es decir, pastoreaban libremente en áreas abiertas de pastizales naturales.
- Un aspecto no contemplado inicialmente fue el nacimiento de crías, ya que entre los meses de enero y abril las alpacas se encuentran en temporada reproductiva.
- Debido a la influencia del ángulo de visión del dron, para el conteo en pastoreo extensivo, fue necesario juntar un poco a los animales de manera que el UAV pudiera cubrir la totalidad del hato y así obtener mejores resultados.

Resultado de validación

En la siguiente Tabla 3.16 se puede observar una ligera subestimación de la solución tecnológica. Los resultados son aceptables, teniendo como mejor resultado una precisión del 90 %, como consecuencia de las casuísticas mencionadas anteriormente.

Tabla 3.16

Resultados de conteo de alpacas en Phinaya, pastoreo extensivo

Sesión	Especie	Conteo Manual	Conteo Digital	Precisión
Vuelo 01	Alpaca	188	165	87.77 %
Vuelo 02	Alpaca	188	171	90.96 %

3.7.3. Validación en conteo de alpacas (Comunidad de Hanccollahua)

Condiciones de evaluación

- El conteo de alpacas en cerco perimétrico fue realizado en el sector Paccollani de la comunidad de Hanccollahua, a 4,700 m s.n.m., entre las 12:00 y 13:00 horas, como se muestra en la Fig. A.3.2. del Anexo 3.
- La familia ganadera encerró las alpacas con fines de venta y matanza, situación que fue aprovechada para realizar el proceso de conteo.

- El cerco estaba conformado por piedras blancas y contaba con un árbol en el centro, lo que limitaba parcialmente la visibilidad.
- Las alpacas se encontraban agrupadas, a una distancia aproximada de entre 20 cm y 1 m entre sí; algunas estaban alimentándose, mientras que otras permanecían observando su entorno.
- Se utilizó un ángulo de inclinación del gimbal de -90° y una altura de vuelo entre 20 metros respecto al nivel de partida del dron.

Resultado de validación

En la siguiente Tabla 3.17 Se obtuvo una precisión del 98.55 %, con un error de una alpaca y dos alucinaciones del entorno como llama.

Tabla 3.17

Resultados de conteo de alpacas en Hanccollahua, cerco perimétrico

Sesión	Especie	Conteo Manual	Conteo Digital	Precisión
Vuelo 01	Alpaca	69	68	98.55 %
Vuelo 01	Llama	0	2	-

3.7.4. Validación en conteo de llamas (Comunidad de Hanccollahua)

Condiciones de evaluación

- El conteo de llamas fue realizado en el sector Paccollani de la comunidad de Hanccollahua, a 4,700 m s.n.m., entre las 13:00 y 14:00 horas, como se muestra en la Fig. A.3.3. del Anexo 3.
- Las llamas se encontraban libremente en un área abierta de pastizales naturales y paja andina (pastoreo extensivo). La mayoría de los individuos estaba concentrada en su alimentación, mientras que algunos permanecían observando su entorno.

- Se utilizó un ángulo de inclinación del gimbal de -90° y una altura de vuelo entre 20 y 40 metros respecto al nivel de partida del dron.

Resultado de validación

En la siguiente Tabla 3.18, se obtuvo a una precisión de conteo del 96.84 %. Una de las observaciones durante el conteo con CamelidCounter fue el campo de visión: algunos animales salen del campo de visión del UAV antes de cruzar la línea de conteo, por lo que no son contabilizados. Otro inconveniente es el mimetismo animal: hubo una ínfima cantidad de llamas que no fueron detectadas.

Tabla 3.18

Resultados de conteo de llamas en Hancollahua, pastoreo extensivo

Sesión	Especie	Conteo Manual	Conteo Digital	Precisión
Vuelo 01	Llama	95	90	94.74 %
Vuelo 02	Llama	95	92	96.84 %

3.7.5. Evaluación del tiempo de conteo

Uno de los aspectos importantes de la implementación de la solución tecnológica es cómo se puede mejorar el conteo automatizado dadas las limitaciones de la visión computacional, optimizando el tiempo, como se muestra en la Tabla 3.19.

Tabla 3.19

Comparación de tiempo de conteo manual y digital

Nro de animales	Manual	Automático
95	125 seg.	35 seg.
200	263 seg.	45 seg.
300	395 seg.	50 seg.
400	526 seg.	55 seg.

Es importante mencionar que la primera fila de la Tabla 3.19 corresponde a un conteo válido realizado durante las pruebas de campo (es importante mencionar que se trata del mejor caso y optimas condiciones, dado que para hatos grandes se realizan recuentos y mas de una persona para realizar el conteo).

Para la estimación del tiempo de conteo manual de 200, 300 y 400 animales se utilizó la siguiente ecuación:

$$T_{manual} \approx \frac{125}{95} * NroAnimales \quad (3.4)$$

Mientras que para el conteo automático realizamos un razonamiento simple, el número de frames se puede expresar con la ecuación.

$$NroFrames = FPS * T_{vuelo} \quad (3.5)$$

Entonces

$$T_{automatic} \approx NroFrames * T_f \quad (3.6)$$

Donde FPS son los fotogramas por segundo del dron (30 FPS para el dron utilizado), T_{vuelo} es el tiempo de vuelo del dron en segundos y T_f es el tiempo de inferencia por fotograma calculado en la Tabla 3.15, lo que nos da una buena aproximación al tiempo de vuelo durante el conteo con drones.

A diferencia del conteo manual, que requiere una evaluación secuencial de cada instancia de objeto, el modelo de detección procesa la escena de manera paralela, permitiendo identificar múltiples objetos en un mismo paso computacional, lo que explica su comportamiento escalable en los modelos de detección de objetos. Para entenderlo mejor, la cantidad de animales no interviene en la ecuación, porque el tiempo de conteo es directamente proporcional al tiempo de vuelo, dado que el FPS es constante y T_f también es constante, con ligeras variaciones imperceptibles. Además, a mayor cantidad de animales habrá una mayor densidad de instancias dentro del campo de visión del dron, por lo que no afecta significativamente su consideración.

Capítulo 4

Análisis y discusión de resultados

El análisis de resultados se realizó en función de cada objetivo planteado, mientras que las discusiones están basadas en los antecedentes de la investigación.

4.1. Análisis de resultados respecto a los objetivos

En relación con el objetivo general, la implementación de una solución tecnológica basada en visión computacional para la contabilización automática de camélidos domésticos demuestra cumplir los requerimientos establecidos. La integración de un modelo de detección (YOLO26n) en una aplicación móvil capaz de procesar video en tiempo real desde un UAV permite realizar conteos en condiciones de pastoreo o confinamiento; el desempeño alcanzó una exactitud promedio superior al 95 %, lo que valida su aplicabilidad en entornos ganaderos.

- Respecto al primer objetivo específico, la construcción del dataset demostró que la diversidad y calidad de los datos son factores determinantes para el rendimiento final de la solución tecnológica. La incorporación de imágenes capturadas en distintas localidades altoandinas, bajo diversas condiciones de iluminación, topografía y características del terreno, así como el uso de imágenes negativas, contribuyó a mejorar la capacidad

de generalización del modelo. Como aporte adicional de la investigación, el dataset fue enviado a un repositorio de acceso público, que permitirá su reutilización y validación por parte de otros investigadores (Huancara, 2026). No obstante, los errores asociados al mimetismo entre los animales y el entorno evidencian la necesidad de incorporar escenarios más complejos.

- Para la selección del modelo con mejor desempeño, se analizaron distintas arquitecturas, donde los modelos de la familia YOLO tuvieron un rendimiento sobresaliente, siendo YOLO26n con un mAP@50 de 92.5 %, tiempo de inferencia de 38.93 ms y FPS de 25.69, la mejor opción en comparación con YOLO11, que obtuvo un mAP@50 de 93.2 %, tiempo de inferencia de 55.36 ms y FPS de 18.06. Mientras tanto, el modelo de dos etapas Faster R-CNN alcanzó un mAP@50 de 88.4 %, pero estuvo limitado en tiempo de inferencia por la complejidad de su arquitectura. Finalmente, se analizó el modelo SSD, cuyo desempeño estuvo muy por debajo del de los demás. Estos resultados justifican la selección del modelo con mejor desempeño para la implementación final.
- Con respecto a la implementación del modelo en una aplicación móvil conectada a un UAV, los resultados confirman la viabilidad técnica de la integración propuesta. El sistema fue capaz de recibir, procesar y visualizar en tiempo real el flujo de video de manera offline, incorporando además un módulo de seguimiento multiobjeto para la contabilización basada en trayectorias. La estabilidad observada durante las pruebas de campo confirma su capacidad de funcionamiento en condiciones reales de operación.
- Finalmente, respecto al objetivo de validación en campo, los resultados muestran una concordancia aceptable entre el conteo manual realizado por el especialista zootecnista y el conteo automático generado por CamelidCounter. Las diferencias identificadas

estuvieron asociadas principalmente a factores externos, como la visibilidad parcial de los animales y las limitaciones del campo de visión del UAV, más que a deficiencias propias del modelo de detección. Asimismo, se observó una reducción significativa en el tiempo requerido para realizar los conteos, lo que constituye una ventaja importante frente a los métodos tradicionales empleados en la actualidad. Este hallazgo se alinea con los resultados de la encuesta técnica aplicada a los productores, quienes señalaron que el conteo manual demanda tiempo.

4.2. Discusión de resultados respecto a los antecedentes

Aunque no se identificaron suficientes estudios específicos sobre camélidos sudamericanos, se toman como referencia trabajos relacionados con especies ganaderas como la alpaca y la vicuña para analizar aspectos metodológicos y técnicos de la solución.

- Como antecedente nacional, Perú Flying Labs (Labs, 2023) reportó una experiencia de contabilización de camélidos mediante drones e inteligencia artificial en el centro experimental Pacamarca, donde se menciona la captura de imágenes aéreas y la ejecución de vuelos para apoyar el conteo automatizado de alpacas. Sin embargo, al tratarse de una publicación divulgativa, no se detallan métricas formales de desempeño ni una validación técnica exhaustiva del sistema. En contraste, nuestro trabajo desarrolla un modelo de visión computacional para el conteo automático de camélidos sudamericanos domésticos, con entrenamiento sobre un dataset propio, pruebas en campo y evaluación de desempeño bajo métricas cuantitativas.
- En el trabajo presentado por Ocaña Wilcapi et al. (2026), desarrollaron el censo poblacional de alpacas de la Cooperativa de Productores Alpaqueros de Chimborazo me-

dian­te trabajo de campo y estadística descriptiva, registrando variables como caravana, estado fisiológico reproductivo y categorías productivas, con un total de 979 alpacas en 90 días. Por su parte, el V Censo de vicuñas de Servicio Nacional de Áreas Naturales Protegidas por el Estado (2025) constituye un antecedente metodológico relevante, ya que su conteo en zonas altoandinas se realiza mediante observación visual en campo y brigadas de monitoreo. Ambos antecedentes evidencian la importancia del conteo y la necesidad de información confiable para la gestión de camélidos; sin embargo, también muestran la persistencia de métodos manuales, mientras que la presente investigación propone una solución automatizada mediante UAV y visión computacional para rea­lizar el conteo de camélidos sudamericanos domésticos de forma más eficiente y con menor intervención humana.

- Finalmente, la investigación de Carrasco-Sauñe and Ibarra-Cabrera (2023) presenta resultados de alta precisión utilizando modelos como Mask R-CNN; sin embargo, estos resultados deben interpretarse con cautela debido a limitaciones en el dataset, como la mezcla de especies (falta de conocimiento de los camélidos que se refleja bastante al considerar al guanaco, llama y alpacas bajo el apelativo “ALPACA”) y no ser validadas por un especialista. En contraste, la presente investigación desarrolla un dataset específico de camélidos sudamericanos domésticos (llama y alpaca) capturado mediante UAV en condiciones reales, lo cual proporciona una base más representativa para el entrenamiento del modelo. Asimismo, aunque Mask R-CNN mostró alto rendimiento en dicho estudio, su aplicación en este trabajo se ve limitada por requerimientos computacionales que dificultan su uso en tiempo real en dispositivos móviles, lo cual refuerza la elección de modelos más ligeros de una sola etapa como YOLO.

En conjunto, estos resultados confirman el potencial de la visión computacional y los UAV como herramientas de apoyo para la gestión sostenible de poblaciones de camélidos sudame­

ricanos domésticos, al ofrecer una solución reproducible, eficiente y técnicamente validada en condiciones de uso reales. Además, automatiza el proceso de conteo manual, reduciendo la intervención humana y optimizando significativamente el tiempo requerido para la obtención de resultados.

4.3. Análisis referencial de inversión y viabilidad económica solución tecnológica de conteo automatizado

4.3.1. Estimación del costo del sistema

Debido a la ausencia de datos estadísticos confiables sobre el gasto económico real en las comunidades altoandinas, no fue posible desarrollar un análisis de rentabilidad basado en información financiera histórica de los productores. Por esta razón, se presenta únicamente una estimación del costo de inversión del sistema propuesto en la siguiente Tabla 4.1, con el fin de proporcionar un valor referencial que permita a los productores de camélidos sudamericanos domésticos evaluar su viabilidad de adquisición.

Tabla 4.1

Costo estimado de implementación de la solución tecnológica propuesta

Ítem	Descripción	Cant.	Costo unitario (USD)	Costo total (USD)
1	UAV con cámara HD y transmisión en tiempo real	1	1,200	1,200
2	Dispositivo móvil de media/alta gama para visualización y control	1	450	450
3	Desarrollo de aplicativo móvil conectado al UAV	1	1,000	1,000
4	Desarrollo e integración del modelo de detección de objetos	1	1,000	1,000
5	Implementación de tracking y conteo automático	1	1,000	1,000
6	Capacitación de uso	1	300	300
7	Mantenimiento y soporte técnico inicial	1	700	700
Total estimado				5,650

4.3.2. Potencial de retorno

El sistema propuesto presenta un potencial de retorno favorable, principalmente por el ahorro de tiempo que genera en el proceso de conteo de camélidos sudamericanos. Además, al automatizar la detección y el registro de los animales mediante visión computacional, se reduce la probabilidad de errores asociados al conteo manual, lo que contribuye a una mayor precisión en los resultados. De esta manera, se optimiza el proceso de conteo y se mejora la eficiencia operativa general en las actividades de monitoreo y control de las alpacas y llamas.

4.3.3. Aceptación por parte de los productores

De acuerdo con los resultados obtenidos en las encuestas realizadas. Ver (Anexo 5). Se evidencia un interés positivo por parte de los productores en adquirir e incorporar soluciones tecnológicas aplicadas a sus actividades productivas. Este resultado refleja una predisposición favorable hacia la adopción de herramientas innovadoras que permitan mejorar la gestión y el conteo de camélidos sudamericanos domésticos, lo cual fortalece la viabilidad de implementación del sistema propuesto.

Conclusiones

1. En relación con la construcción del dataset, se logró construir un conjunto de datos aéreos representativo de llamas y alpacas en condiciones reales de pastoreo ganadero, incluyendo la variabilidad de altitud de captura de datos, iluminación, horas de pastoreo, temporadas del año 2025, variabilidad de las zonas altoandinas de la región del Cusco y distribución en la densidad de animales, lo que permitió alcanzar una mejor generalización durante las pruebas, aunque se tuvieron complicaciones con el mimetismo ambiental.
2. Respecto a la elección de modelos de visión computacional, se realizó un análisis comparativo con distintos modelos de una y dos etapas, donde las arquitecturas de la familia YOLO tuvieron mejor desempeño entre precisión y velocidad (tiempo de inferencia y FPS), como se mencionó en la Sec. 3.5. Particularmente, YOLO26n mostró un desempeño superior frente a YOLO11n, Faster R-CNN y SSD, con valores de mAP@50, tiempo de inferencia y FPS de 92.5 %, 38 ms y 25.69 respectivamente, siendo seleccionado para la implementación final debido a su tiempo de inferencia y FPS, que lo hacen adecuado para aplicaciones en tiempo real en dispositivos móviles.
3. En cuanto a la implementación de un aplicativo móvil, se logró integrar el modelo YOLO26n en formato TFLite dentro de una aplicación móvil de forma local, y este fue conectado a un dron DJI Mini 3 a través de su MSDK, permitiendo la recepción, el procesamiento y la visualización de video en tiempo real. Adicionalmente, la integra-

ción de un algoritmo de tracking basado en predicción mediante el filtro de Kalman, asociación de datos, vote decay y la incorporación de una línea de conteo permitieron mantener la consistencia en las detecciones, evitando conteos duplicados y mejorando la estabilidad del conteo.

4. Finalmente, en relación con la validación en campo de la solución tecnológica, se compararon los conteos automatizados realizados por el sistema CamelidCounter con los conteos manuales in-situ realizados por el especialista en camélidos sudamericanos, evidenciándose una exactitud promedio superior al 95 %, dato que confirma la confiabilidad del sistema en condiciones reales de operación. Las diferencias observadas se atribuyen principalmente al mimetismo ambiental y a las limitaciones del campo de visión del UAV. Asimismo, se evidenció una mejora en la eficiencia operativa durante el proceso de conteo, reduciendo el tiempo requerido respecto al método manual tradicional, que crece de forma lineal a medida que aumenta la densidad de llamas y alpacas. En consecuencia, los resultados demuestran que CamelidCounter constituye una solución técnicamente viable, operativamente eficiente y aplicable en contextos reales de operación.

Recomendaciones

- Se recomienda estandarizar y ampliar el dataset, agregando mayor diversidad de datos espaciales, temporales, variaciones estacionales, ecosistemas diferentes y configuraciones de densidad poblacional. Además, se pueden agregar las demás especies de camélidos sudamericanos que tienen una morfología similar.
- Se recomienda incluir más imágenes negativas en el dataset para reducir falsos positivos y mejorar la capacidad del modelo para diferenciar camélidos de fondos, rocas u otros animales. Además, esta estrategia podría favorecer su adaptación a otras especies, como la vicuña. La cantidad de estas imágenes negativas debe definirse según las características del entorno y el comportamiento del modelo durante el entrenamiento.
- Durante el pastoreo extensivo se cuenta con un pastor o personal encargado del cuidado de las alpacas y llamas para evitar la invasión de terrenos y proteger los pastizales. Por ello, se recomienda realizar el conteo cuando el grupo de camélidos se encuentre más compactado antes de iniciar el vuelo del UAV.
- Utilizar un pipeline con Digital Image Processing en el dominio Espacial y de Frecuencia en el preprocesamiento para mejorar el rendimiento de los modelos de detección de objetos.
- Utilizar modelos más avanzados como instance segmentation o arquitecturas basadas en vision transformers, que han demostrado ser herramientas con mejor desempeño en

detección para escenarios complejos, aplicando optimizaciones para entornos de edge computing.

- Puesto que la actual investigación usa una versión modificada de SORT para el tracking, se recomienda evaluar algoritmos avanzados de tracking como DeepSORT o enfoques recientes de tracking end to end, que pueden ayudar a mejorar el seguimiento ante oclusiones prolongadas. Se podrían obtener mejores resultados en comparación con métodos basados solo en geometría.
- Optimizar el modelo para su ejecución en dispositivos móviles mediante cuantizaciones avanzadas, tiling, pruning o el uso de aceleradores como GPU móvil o NPU.
- Integrar funcionalidades adicionales en el aplicativo móvil, como una persistencia de datos más robusta, georreferenciación de datos o sistemas de información geográfica (SIG) que permitan georreferenciar los conteos y obtener mapas de distribución de animales en tiempo real, así como la generación de reportes automatizados, lo cual sería de gran valor para los usuarios finales de los fundos ganaderos.
- Finalmente, se propone extender el sistema a ámbitos agropecuarios, como el monitoreo de salud animal, la detección de comportamientos anómalos o la estimación de biomasa, lo cual requeriría nuevos sensores y modelos especializados.

Bibliografía

- Adusumalli, H., Kalyani, D., Sri, R., Pratapteja, M., and Rao, P. V. R. D. P. (2021). Face mask detection using opencv. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, pages 1304–1309.
- Arenas, B. O. (2021). Observación en inaturalist: Vicuña (vicugna vicugna). <https://www.inaturalist.org/observations/258002832>. Acceso: 26 de enero de 2025.
- Bai, D., Sun, Y., Tao, B., Tong, X., Xu, M., Jiang, G., Chen, B., Cao, Y., Sun, N., and Li, Z. (2022). Improved single shot multibox detector target detection method based on deep feature fusion. *Concurrency and Computation: Practice and Experience*, 34(4):e6614.
- Barreta, J., Gutiérrez-Gil, B., Iñiguez, V., Saavedra, V., Chiri, R., Latorre, E., and Arranz, J. J. (2013). Analysis of mitochondrial dna in bolivian llama, alpaca and vicuna populations: a contribution to the phylogeny of the south american camelids. *Animal Genetics*, 44(2):158–168.
- BioScience (2021). A comprehensive overview of technologies for species and habitat monitoring and conservation. *BioScience*, 71(10):1038–1050.
- Carrasco-Sauñe, F. and Ibarra-Cabrera, M. (2023). Determinación del mejor algoritmo de reconocimiento de imágenes de camélidos sudamericanos mediante machine learning, apurímac, 2023. *C&T Riqchary Revista de investigación en ciencia y tecnología*, 5:57–62.

- Choudhary, N. (2024). An image annotation guide using roboflow for object detection. <https://medium.com/@nikhilc2209/>. Accedido el 7 de febrero del 2025.
- de Lamo, D. A. (2011). Camélidos sudamericanos: Historia, usos y sanidad nimal. *Ministerio de Agricultura, Ganaderia y Pesca - SENASA, Argentina*, pages 12–13.
- de Lima Weber, F., de Moraes Weber, V. A., de Moraes, P. H., Matsubara, E. T., Paiva, D. M. B., de Nadai Bonin Gomes, M., de Oliveira, L. O. F., de Medeiros, S. R., and Cagnin, M. I. (2023). Counting cattle in uav images using convolutional neural network. *Remote Sensing Applications: Society and Environment*, 29:100900.
- Department of Electrical and Electronic Engineering, University of Manchester (2024). Autonomous Systems Research at the University of Manchester. <https://www.eee.manchester.ac.uk/research/expertise/autonomous-systems>. Accedido el 7 de febrero de 2025.
- Gao, L., He, Y., Sun, X., Jia, X., and Zhang, B. (2019). Incorporating negative sample training for ship detection based on deep learning. *Sensors*, 19(3).
- Gonzalez, R. C. and Woods, R. E. (2018). *Digital Image Processing*. Pearson Education Limited, 4th (global edition) edition.
- Hernández-Sampieri, R. and Mendoza Torres, C. P. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Interamericana Editores, Ciudad de México, 1 edition.
- Hoang, J. (2024). How to evaluate computer vision models with confusion matrix. <https://www.datature.io/blog/how-to-evaluate-computer-vision-models-with-confusion-matrix>. Accedido el 6 de febrero de 2025.

- Hu, C., Sapkota, B. B., Thomasson, J. A., and Bagavathiannan, M. V. (2021). Influence of image quality and light consistency on the performance of convolutional neural networks for weed mapping. *Remote Sensing*, 13(11).
- Huancara, A. H. (2026). CAMEL-UAV.
- Jocher, G., Qiu, J., and Chaurasia, A. (2023). Ultralytics YOLO.
- Kaur, J. and Singh, W. (2022). Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, 81(27):38297–38351.
- Labs, P. F. (2023). Contabilización de camélidos utilizando imágenes con drones e inteligencia artificial. <https://bit.ly/flyinglabs-contabilizacion-de-camelidos>. Accedido el 28 de enero de 2025.
- Lee, Y.-H. and Kim, Y. (2020). Comparison of cnn and yolo for object detection. *Journal of the Semiconductor & Display Technology*, 19(1):85–92.
- MIDAGRI (2021). Camélidos sudamericanos: Situación de las actividades de crianza y producción. <https://www.midagri.gob.pe/portal/40-sector-agrario>. Accedido el 3 de febrero de 2025.
- MIDAGRI (2023). Camélidos sudamericanos. <https://www.midagri.gob.pe/portal/40-sector-agrario/>. Accedido el 7 de febrero de 2025.
- MIDAGRI (2024a). Gobierno reconoce a productores de ocho regiones en el día nacional de la llama. <https://www.gob.pe/es/n/1062556>. Accedido el 30 de enero de 2025.
- MIDAGRI (2024b). Midagri: El 87 % de la población mundial de alpacas se encuentra en el Perú. <https://www.gob.pe/es/n/636546>. Accedido el 30 de enero de 2025.

- MIDAGRI (2024c). Midagri: Más de 37 mil familias productoras de puno serán beneficiadas mediante actividades por el año internacional de los camélidos 2024. <https://www.gob.pe/es/n/911794>. Accedido el 30 de enero de 2025.
- Milne, H. (2025). Observación en inaturalist: Guanaco (lama guanicoe). <https://www.inaturalist.org/observations/258943432>. Acceso: 26 de enero de 2025.
- Morlier, J., Leonardon, M., and Gripon, V. (2025). Input resolution downsizing as a compression technique for vision deep learning systems. *arXiv preprint arXiv:2504.03749*.
- Musleh, A., Qasem, A., and AlRyalat, S. A. (2023). Image annotation software for artificial intelligence applications. *High Yield Medical Reviews*, 1(2):14–45.
- Nogueira-Rodríguez, A., Glez-Peña, D., Reboiro-Jato, M., and López-Fernández, H. (2023). Negative samples for improving object detection—a case study in ai-assisted colonoscopy for polyp detection. *Diagnostics*, 13(5).
- Ocaña Wilcapi, J. J., Ponce Cepeda, E. V., Oleas López, J. M., and Vaca Cárdenas, M. L. (2026). Censo poblacional de alpacas de la cooperativa de los productores alpaqueros de chimborazo. *Dominio de las Ciencias*, 12(1):573–583.
- Oksuz, K., Cam, B. C., Kalkan, S., and Akbas, E. (2021). Imbalance problems in object detection: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3388–3415.
- ONU (2024). Informe de la Segunda Comisión sobre Desarrollo Sostenible. Technical Report A/72/420, Asamblea General de las Naciones Unidas, Nueva York.
- OpenCV Team (2024). Data Annotation for AI: Overview and Tools. <https://opencv.org/blog/data-annotation/>. Accedido el 7 de febrero de 2025.

- Ortega Candel, J. M. (2025). *Inteligencia Artificial: Investigaciones, aplicaciones y avances*. Títulos Especiales. Anaya Multimedia, Madrid, España.
- Patel, H. (2023). A comprehensive study on object detection techniques in unconstrained environments.
- Pezo, D., Franco, E., García, W., Franco, F., Bravo, W., Alarcón, V., and San Martín, F. (2014). *Manual del técnico alpaquero*. Soluciones Prácticas, Lima, 2da edition.
- PyTorch Developers (2024). PyTorch: An Open Source Machine Learning Framework. <https://pytorch.org/docs/stable/index.html>. Accedido el 7 de febrero del 2025.
- Rančić, K., Blagojević, B., Bezdan, A., Ivošević, B., Tubić, B., Vranešević, M., Pejak, B., Crnojević, V., and Marko, O. (2023). Animal detection and counting from uav images using convolutional neural networks. *Drones*, 7(3).
- Sahin, M. E., Ulutas, H., Yuce, E., and Erkoc, M. F. (2023). Detection and classification of covid-19 by using faster r-cnn and mask r-cnn on ct images. *Neural Computing and Applications*, 35(18):13597–13611.
- Science Direct (2022). Introduction to deep learning in computer science. <https://www.sciencedirect.com/topics/computer-science/deep-learning-technique>. Accessed: 2025-12-19.
- Seijas, C., Montilla, G., and Frassato, L. (2019). Identificación de especies de roedores usando aprendizaje profundo. *Computación y Sistemas*, 23(1):257–266.
- Servicio Nacional de Áreas Naturales Protegidas por el Estado (2025). El sernanp inicia conteo de vicuñas en áreas naturales protegidas en el marco del v censo nacional. Nota de prensa institucional.

- Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., and Naemura, T. (2020). Cattle detection and counting in uav images based on convolutional neural networks. *International Journal of Remote Sensing*, 41(1):31–52.
- Szeliski, R. (2022). *Computer Vision: Algorithms and Applications*. Springer, 2nd edition.
- Sánchez, D. A., Diez, H. R. G., and Heredia, Y. H. (2020). Revisión de algoritmos de detección y seguimiento de objetos con redes profundas para videovigilancia inteligente. *Revista Cubana de Ciencias Informáticas*, 14(3):165–195.
- Ultralytics (2026). What is a bounding box? object detection guide. Fundamentos de localización y anotación en visión artificial.
- Wang, H. and Xiao, N. (2023). Underwater object detection method based on improved faster rcnn. *Applied Sciences*, 13(4):2746.
- Wu, J. (2017). Introduction to convolutional neural networks. Technical Report Lecture Notes, Nanjing University, China.
- Yuan, D., Shu, X., Fan, N., Chang, X., Liu, Q., and He, Z. (2022). Accurate bounding-box regression with distance-iou loss for visual tracking. *Journal of Visual Communication and Image Representation*, 83:103428.
- Zou, Z., Chen, K., Shi, Z., Guo, Y., and Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276.

Anexos

1. Dataset Público en Mendely Data

Figura A.1.1.
Dataset CAMEL-UAV publicado

Acceso al dataset (registrese con cualquier correo personal para visulizar):

https:

//data.mendeley.com/preview/86g56zw9ts?a=90400dfb-cd44-4708-a7fb-32718da62e7b

CAMEL-UAV

Published: 31 July 2026 | Version 1 | Reserved DOI: 10.17632/86g56zw9ts.1
Contributor: Alex Helder Huancaira

Description

a UAV high-resolution imagery and corresponding bounding box annotations for computer vision applications of South American domestic camelids (llama and Alpaca). The data was collected throughout the year 2025 across the provinces of Espinar and Canchis, located in the Department of Cusco, Peru. To ensure environmental and operational diversity, the imagery captures livestock in distinct seasonal conditions (dry and wet seasons) and was acquired at various flight altitudes.

Download All 1.91 GB

Files

CAMEL-UAV dataset

Categories

Computer Vision, Object Detection, Deep Learning, Database

Draft version

Version 1 (Draft)
Published: Draft
Reserved 10.17632/86g56zw9ts.1
DOI:

Cite this dataset:

Huancaira, Alex Helder (2026), "CAMEL-UAV", Mendeley Data, V1, doi: 10.17632/86g56zw9ts.1

*DOI is reserved but not active.

Copy to clipboard

2. Aplicación Android de solución tecnológica

Figura A.2.1.
Interfaz de inicio de la Aplicación

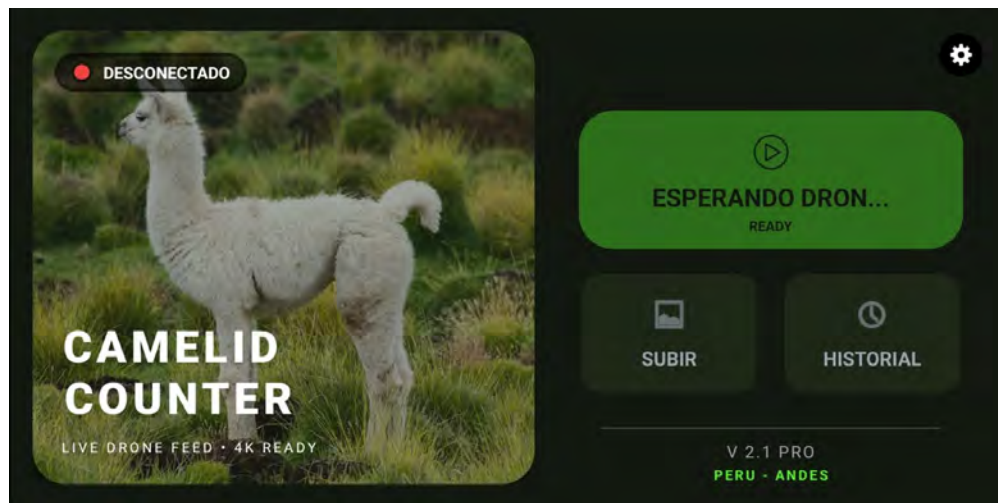


Figura A.2.2.
Interfaz durante el vuelo para el conteo de camélidos sudamericanos domésticos



Figura A.2.3.
Interfaz reportes (resultados de conteo)



3. Registro fotográfico de las zonas de validación en campo

Figura A.3.1.
Alpacas: Ganadería de la familia Huillca Cochama, Pitumarca, Phinaya, Sector Canlline a 4,800 m s.n.m



Figura A.3.2.

Alpacas: Ganadería de la familia Pedro Maccarcco, Espinar, Hanccollahua, Sector Paccollani a 4,700 m s.n.m.



Figura A.3.3.

Llamas: Ganadería de la familia Choque Gonzales, Espinar, Hanccollahua, Sector Paccollani a 4,700 m s.n.m.



4. Documento de validación técnica emitido por el especialista en camélidos sudamericanos domésticos.

INFORME DE VALIDACIÓN TÉCNICA N° 001-2026-ACCM-ECS

A : Dr. Abel Franklin Canal Céspedes
Director (e) de Gestión de la Investigación

De : Ing. Abel Ccama Mendigure
Especialista en Camélidos Sudamericanos.

Asunto : Validación de conteo automatizado de camélidos sudamericanos domésticos mediante la Solución Tecnológica CamelidCounter.

Fecha : 15 de marzo de 2026

Tengo el agrado de dirigirme a usted para saludarlo cordialmente y, al mismo tiempo, remitir el Informe de Validación del conteo automatizado de camélidos sudamericanos domésticos, realizado mediante la solución tecnológica *CamelidCounter*. Dicho proceso se deriva del trabajo de tesis titulado: **"CONTABILIZACIÓN DE CAMÉLIDOS SUDAMERICANOS DOMÉSTICOS MEDIANTE VISIÓN COMPUTACIONAL APLICADA A IMÁGENES OBTENIDAS POR VEHÍCULOS AÉREOS NO TRIPULADOS"**, cuyos detalles técnicos se exponen a continuación:

1. ANTECEDENTES Y METODOLOGÍA DE VALIDACIÓN

En mi calidad de Ingeniero Zootecnista colegiado con CIP N°232458, se procedió a realizar la validación técnica de campo del sistema CamelidCounter desarrollado por el Bach. Alex Helder Huancara Ccolque y Bach. David Soncco Cachura, de la tesis titulada: **"CONTABILIZACIÓN DE CAMÉLIDOS SUDAMERICANOS DOMÉSTICOS MEDIANTE VISION COMPUTACIONAL APLICADA A IMÁGENES OBTENIDAS POR VEHÍCULOS AÉREOS NO TRIPULADOS"**

La validación consistió en la comparación directa entre:

- Conteo manual in situ realizado por el suscrito.
- Conteo automatizado generado por el sistema CamelidCounter, el cual procesa video en un dispositivo móvil mientras el dron realiza el sobrevuelo.

Las pruebas fueron ejecutadas en dos zonas altoandinas juntamente con los tesisistas mencionados arriba, bajo condiciones reales de manejo ganadero:

- **Comunidad Phinaya - Distrito de Píumarca:** 4,800 m s.n.m. (pastoreo extensivo).
- **Comunidad Hanccollahua (Sector Paccollani) - Distrito de Espinar:** 4,700 m s.n.m. (cerco y pastoreo extensivo).



2. REGISTRO DE PRUEBAS Y VALIDACIÓN CRUZADA

2.1 Evaluación en Phinaya (Pastoreo)

Rebaño evaluado : 188 alpacas.
Altura de vuelo : 40 m.
Ángulo de cámara : -40°.
Terreno : Campo abierto con presencia de formaciones rocosas.

Vuelo	Especie	Conteo Manual	Conteo Digital	Observación Técnica	Precisión
01	Alpaca	188	165	01 piedra detectada como alpaca, y resto de animales no detectados por mimetismo con las rocas.	87.77%
02	Alpaca	188	171	03 piedras detectadas como alpaca. 02 alpacas tuvieron detecciones como llama, y casos no detectados por mimetismo.	90.96%

Se identificó sensibilidad del sistema ante la presencia de rocas blancas, generando detecciones erróneas o no detectar todas las alpacas, debido a similitud cromática con el vellón de alpacas blancas en su mayoría.

2.2 Evaluación en Hanccollahua (Cerco)

Rebaño evaluado : 69 alpacas.
Altura de vuelo : 20 m.
Ángulo de cámara : 90° grados (en posición cenital).
Terreno : Cerco de piedras blancas con presencia de un árbol en el centro.

Vuelo	Especie	Conteo Manual	Conteo Digital	Observación Técnica	Precisión
01	Alpaca	69	68	Omisión de 01 ejemplar no visible.	98.55%
	Llama	0	2	Error de clasificación de especie	--

2.3 Evaluación en Hanccollahua (Pastoreo)

Rebaño evaluado : 95 llamas.
Altura de vuelo : 40 m.
Ángulo de cámara : -40°.
Terreno : Zona de pastoreo plano con paja andina.

Vuelo	Especie	Conteo Manual	Conteo Digital	Observación Técnica	Precisión
01	Llama	95	90	03 ejemplar no detectados a causa de la velocidad de dron. 02 ejemplares no detectados a causa de estar fuera del campo de visión del dron.	94.74%

02	Llama	95	92	• 03 ejemplares no detectados por la velocidad del vuelo del dron.	96.84%
----	-------	----	----	--	--------

3. CONSIDERACIONES TÉCNICAS POST-EVALUACIÓN

Tras el análisis de los resultados obtenidos, se establecen las siguientes conclusiones técnicas para la optimización del sistema de monitoreo:

Interferencia por Reflectancia y Camuflaje Natural: En ecosistemas altoandinos con afloramientos rocosos de tonalidades claras, se identificó una tasa marginal de falsos positivos debido a la similitud cromática entre el sustrato y el pelaje de los ejemplares. Asimismo, se registraron omisiones en la detección (falsos negativos) vinculadas a la mimetización morfológica y a las variaciones en la firma visual del ganado frente al entorno rocoso.

Desafíos en la Clasificación Taxonómica: Se observó una disminución en la precisión de la clasificación entre llamas y alpacas en escenarios de alta densidad poblacional (agrupamientos) o cuando la captura de imagen se realiza bajo un ángulo cenital estricto (nadir), donde las diferencias morfológicas distintivas se vuelven menos perceptibles.

Atribución de Variabilidad en el Conteo: Las discrepancias cuantitativas en el inventario de llamas no son atribuibles a deficiencias en la arquitectura del algoritmo de detección, sino que se asocian a limitaciones operativas de encuadre y al campo de visión (Field of View) del sensor durante el desplazamiento del dron.

Evaluación del Impacto Etológico: Los protocolos de vuelo ejecutados a altitudes de 20, 30 y 40 metros demostraron neutralidad acústica y visual, al no inducir alteraciones conductuales ni respuestas de estrés en el ganado. Esto valida la metodología para levantamientos de inventarios ganaderos sin afectación al bienestar animal.

Escalabilidad y Valor Estratégico: Para hatos que superan las 150 cabezas, la implementación de esta solución tecnológica adquiere un carácter crítico. La automatización mitiga el sesgo del conteo manual y la tendencia al abandono de registros censales en unidades productivas de gran escala, donde la complejidad logística excede las capacidades de la supervisión tradicional.

4. RECOMENDACIONES.

1. Optimización de la Captura de Datos (Dron y Sensor). - Ajuste del Ángulo de Inclinación: Se sugiere configurar el gimbal de la cámara con una inclinación de 15° a 20° respecto al nadir (vertical estricto). Esto permitirá capturar rasgos morfológicos laterales (cuello, perfil facial y altura de la cruz) que facilitan la distinción taxonómica entre llamas y alpacas.

Planificación de Vuelo Según la Pendiente: Para evitar discrepancias en el encuadre, los planes de vuelo deben ser terreno-seguídos (terrain-following), manteniendo una altura relativa constante respecto al suelo, especialmente en la accidentada topografía altoandina.

Configuración del Sensor: Utilizar filtros de polarización o ajustar el balance de blancos de forma manual para reducir el resplandor de las rocas claras, mejorando el contraste entre el pelaje del animal y el sustrato geológico.

2. Refinamiento del Algoritmo de Detección

Aumento del Dataset en Entornos Rocosos: Se recomienda reentrenar los modelos de visión artificial integrando un mayor volumen de imágenes que contengan "negativos difíciles" (rocas con formas y colores similares al ganado) para reducir la tasa de falsos positivos.

Segmentación por Talla: Implementar un umbral de detección basado en el área de píxeles ocupada por el ejemplar, permitiendo que el sistema ignore objetos demasiado pequeños para ser ganado o demasiado grandes para ser un individuo aislado.

3. Protocolos Operativos y de Bienestar Animal

Ventanas Horarias de Vuelo: Priorizar los sobrevuelos durante las primeras horas de la mañana o media tarde. Esto no solo mejora la calidad visual por la suavidad de las sombras, sino que coincide con los periodos de pastoreo donde los animales están más dispersos y son más fáciles de contabilizar.

Validación de Altitud: Aunque se comprobó que a 20 metros no hay estrés, se recomienda estandarizar la altura de trabajo a 30 metros como margen de seguridad para evitar cualquier impacto acústico imprevisto ante cambios de clima (viento) que exijan más potencia a los motores del dron.

4. Gestión de la Información y Escalabilidad

Digitalización Obligatoria para Hatos Mayores: Establecer la fotogrametría y el conteo automatizado como el método estándar de auditoría para beneficiarios con más de 150 cabezas, eliminando el conteo manual que es propenso al error humano y al subregistro.

Georreferenciación de Hallazgos: Vincular cada conteo automático a coordenadas GPS para generar mapas de calor de ocupación de pastizales, permitiendo una gestión más eficiente de la carga animal en los fundos ganaderos.

5. Conteo en Recintos de Confinamiento (Cercos): Para unidades poblacionales de alta densidad, se recomienda ejecutar la captura de imagen con el sensor en **posición cenital (90° respecto al plano horizontal)**. Esta perspectiva minimiza el error por **superposición de ejemplares** y permite una segmentación más precisa de cada individuo mediante el análisis de su proyección aérea, facilitando el procesamiento por algoritmos de visión artificial.

6. Conteo en Flujo de Tránsito (Control de Acceso): Como alternativa de alta precisión, se propone el monitoreo durante el desplazamiento de los ejemplares en los puntos de ingreso o salida de los cercos. Para este escenario, se recomienda una configuración de cámara con un **ángulo de inclinación de entre -30° y -45°**, o en su defecto una vista cenital. La inclinación oblicua mejora la visibilidad de los rasgos morfológicos laterales, lo que optimiza la tasa de éxito en la identificación y conteo dinámico frente a ejemplares en movimiento.

Estas recomendaciones buscan transformar el sistema de una herramienta de conteo a una plataforma de gestión ganadera de precisión.

5. CONCLUSIONES

En base a las pruebas de validación cruzada realizadas en condiciones reales de campo altoandino, se concluye que el sistema digital evaluado presenta:

- ✓ **Alta Fiabilidad Operativa:** El sistema demuestra una robustez técnica consistente bajo condiciones orográficas y climáticas exigentes, asegurando la continuidad en la captura de datos en entornos altoandinos sin degradación del rendimiento del hardware ni del software.
- ✓ **Métricas de Exactitud:** Se determinó una **precisión promedio general superior al 95.00%**, lo que posiciona a esta solución tecnológica dentro de los estándares de alta confianza para la toma de decisiones en activos biológicos.
- ✓ **Autonomía Tecnológica en Zonas Remotas:** Se confirma la viabilidad técnica para ejecutar inventarios en áreas de difícil acceso y **nula conectividad (offline)**. La capacidad de procesamiento local garantiza la operatividad del sistema en campo sin dependencia de redes externas o servicios de nube.
- ✓ **Mitigación del Error Humano y Optimización de Censos:** La automatización del proceso es determinante para hatos que superan los **150 ejemplares**, logrando neutralizar los factores de error críticos presentes en los métodos tradicionales:
 - **Fatiga Cognitiva y Visual:** Elimina el sesgo por agotamiento del observador en conteos masivos.
 - **Dinamismo y Oclusión:** El algoritmo procesa de forma simultánea ejemplares en movimiento y aquellos parcialmente ocultos, superando la limitación del seguimiento ocular humano en terrenos irregulares.

- **Integridad de la Trazabilidad:** El registro digital evita la pérdida de datos o distracciones comunes en el conteo manual en cerco, asegurando que el inventario final sea auditable y preciso.
- **Eficacia de la Solución Tecnológica:** Se confirma que el uso de aeronaves no tripuladas (RPA) y algoritmos de detección automática constituye una herramienta de alta precisión para el inventario ganadero en zonas altoandinas, superando las limitaciones logísticas y los sesgos del conteo manual tradicional en terrenos accidentados.
- **Determinación de Parámetros de Vuelo Óptimos:** Los resultados validan que un rango de altitud de entre **20 y 40 metros** asegura un equilibrio operativo crítico: permite una resolución de imagen suficiente para la discriminación de ejemplares sin inducir respuestas etológicas negativas (estrés) en las especies evaluadas, garantizando el cumplimiento de estándares de bienestar animal.
- **Identificación de Variables Críticas de Precisión:** La precisión del sistema está condicionada por factores exógenos, principalmente la **reflectancia del entorno rocoso** y la densidad de agrupamiento del ganado. La transición de una perspectiva cenital estricta a una **oblicua (30° a 45°)** en escenarios de flujo dinámico mitiga significativamente los errores de clasificación taxonómica entre llamas y alpacas.
- **Escalabilidad y Mitigación de Subregistros:** La implementación de esta metodología se vuelve estrictamente necesaria en unidades productivas que superan las **150 cabezas**. La automatización del conteo no solo optimiza los tiempos de supervisión, sino que garantiza la integridad de la base de datos, eliminando la tendencia histórica al abandono de registros en hatos de gran magnitud.
- **Viabilidad de Implementación en el Sector Público:** El sistema demuestra ser una solución técnica robusta para la supervisión de proyectos de transferencia tecnológica, permitiendo una valorización precisa de los activos biológicos y un seguimiento eficiente del impacto de los proyectos de inversión en comunidades altoandinas.

La solución tecnológica representa un recurso de alta fiabilidad para la optimización de los procesos de inventariado. Su implementación permite **mitigar el sesgo operativo** derivado de la fatiga visual y **maximizar la eficiencia operativa** en el manejo ganadero de los camélidos sudamericanos.

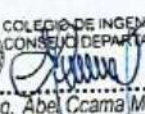
En virtud de los resultados obtenidos, se dictamina que el sistema cumple con los estándares técnicos de **interoperabilidad y precisión** necesarios, validando su aplicabilidad práctica en las unidades de producción de camélidos sudamericanos."

Es cuanto informo a su despacho para los fines del trámite administrativo y las acciones que su jefatura considere pertinentes.

Adjunto:

- ✓ Anexos: tomas fotográficas

Atentamente.

 COLEGIO DE INGENIEROS DEL PERU
CONSEJO DEPARTAMENTAL CUSCO

Ing. Abel Ccama Mendigure
INGENIERO ZOOTECNISTA
CIP 232468

5. Documento de recolección de datos: Cuestionario aplicado a productores

Figura A.5.1.

Encuesta técnica a productores (Pag. 1 de 2)

 UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
FACULTAD DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA, INFORMÁTICA Y MECÁNICA
ESCUELA PROFESIONAL DE INGENIERIA INFORMATICA Y DE SISTEMA
"Año de la Esperanza y el Fortalecimiento de la Democracia" 

ENCUESTA TÉCNICA

TRABAJO DE INVESTIGACIÓN:

"CONTABILIZACIÓN DE CAMÉLIDOS SUDAMERICANOS DOMÉSTICOS MEDIANTE VISIÓN COMPUTACIONAL APLICADA A IMÁGENES OBTENIDAS POR VEHÍCULOS AÉREOS NO TRIPULADOS"

Investigadores:

- Br. Alex Helder Huancara Ccolque
- Br. David Soncco Cachura

Docente Asesor:

- Dr. Rony Villafuerte Serna

PRESENTACIÓN

La presente encuesta tiene como finalidad recopilar información sobre los métodos utilizados para la contabilización de camélidos sudamericanos domésticos (alpacas / llamas) y las dificultades asociadas a este proceso. Los resultados serán utilizados únicamente con fines académicos y de investigación.

DATOS DEL ENCUESTADO

Nombre y Apellidos: Benedicta Jacinta Ccarita

DNI: 24695859

Comunidad/Asociación/Ganadería: Fondo Oquemarca - Pitumarca

Productor de (alpacas y/o llamas): Alpaca y llama

Fecha: 16 / 07 / 2026


Br. ALEX H. HUANCARA
CCOLQUE
TESISTA 1


Br. DAVID SONCCO CACHURA
TESISTA 2

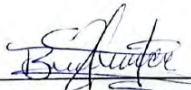

FIRMA DEL ENCUESTADO
24695859.

Figura A.5.1. Encuesta técnica a productores (Pag. 2 de 2)



UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
FACULTAD DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA, INFORMÁTICA Y MECÁNICA
ESCUELA PROFESIONAL DE INGENIERÍA INFORMÁTICA Y DE SISTEMA
"Año de la Esperanza y el Fortalecimiento de la Democracia"



ENCUESTA TÉCNICA

CUESTIONARIO

1. ¿Cuántas ejemplares tiene actualmente bajo su manejo? () Entre 101 y 300 () Más de 300 Cantidad aproximada: <u>800</u> 2. ¿Qué tipo de método utiliza para contabilizar? ☒ Conteo manual directo () Registros escritos () Otro: _____ 3. ¿En qué condiciones realiza el conteo? () Pastoreo extensivo (campo abierto) ☒ Cerco perimétrico (enjaulados) () Otro: _____ 4. ¿Con qué frecuencia realiza la contabilización? ☒ Mensualmente () Trimestralmente () Otra: _____ 5. ¿Cuánto tiempo le toma realizar un conteo completo? () Menos de 30 minutos () Entre 30 minutos y 1 hora ☒ Entre 1 y 2 horas () Más de 2 horas 6. ¿Qué dificultades encuentra durante el proceso de conteo? (Respuesta múltiple) ☒ Demanda mucho tiempo	☒ Los animales están en constante movimiento ☒ Posibilidad de errores en el conteo ☒ Requiere apoyo de otras personas () Condiciones climáticas desfavorables () Otra: _____ 7. ¿Qué cantidad de personal requiere para el proceso de conteo? () Solo yo Cantidad aproximada: <u>3 a 5</u> 8. ¿Estaría dispuesto(a) a utilizar una solución tecnológica que automatice el proceso de conteo? ☒ Sí () No () Tal vez ¿Por qué?: _____ 9. Considerando que la herramienta tecnológica tiene un costo ¿Cómo le gustaría adquirir?: ☒ Adquirirla individualmente () Adquirirla mediante financiamiento () Compartir la inversión con otros productores/asociaciones () Otro: _____ <i>¡Gracias! por su participación ...</i>
--	--

6. Video de captura de datos y validación en campo

[https://drive.google.com/drive/folders/1yQ0HVdXy152Awinr9kTpwT4_0PWQfRLt?
usp=drive_link](https://drive.google.com/drive/folders/1yQ0HVdXy152Awinr9kTpwT4_0PWQfRLt?usp=drive_link)