



Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Querétaro

TC3007C

Inteligencia artificial avanzada para la ciencia de datos II

MODELO DE DETECCIÓN DE VACAS EN IMAGEN

Autores:

A01368818 Joel Sánchez Olvera

A01661090 Juan Pablo Cabrera Quiroga

A01704076 Adrián Galván Díaz

A01708634 Carlos Eduardo Velasco Elenes

A01709522 Arturo Cristián Díaz López

Fecha:

11 de noviembre del 2024

Técnica de modelado

Modelo Seleccionado: YOLOv5

Para la detección de objetos en imágenes aéreas de vacas, utilizamos el modelo YOLOv5 de la biblioteca Ultralytics, una arquitectura de Red Neuronal Convolutiva (CNN) especialmente optimizada para tareas de detección de objetos en tiempo real. YOLO (You Only Look Once) es una de las técnicas de modelado más eficientes y precisas en este campo, ya que permite realizar detección y clasificación simultáneamente en una sola pasada sobre la imagen. Esto resulta clave para el objetivo de nuestro proyecto, que requiere identificar la posición de las vacas en imágenes tomadas desde arriba de manera rápida y precisa.

Optamos por YOLOv5 debido a sus beneficios específicos para nuestro caso:

- **Precisión y Eficiencia en Tiempo Real:** YOLOv5 combina alta precisión con una baja latencia en el procesamiento, lo cual es esencial para la detección de objetos en grandes volúmenes de datos visuales.
- **Facilidad de Uso y Flexibilidad:** Este modelo está ampliamente documentado y es fácil de implementar y personalizar con PyTorch, lo que nos permitió ajustar los hiper parámetros y adaptar el modelo a nuestro conjunto de datos específico. Además, la implementación de YOLOv5 en Ultralytics permite transferir fácilmente el aprendizaje a un conjunto de datos propio, usando un modelo pre-entrenado.
- **Optimización para GPU:** Dado que disponemos de una GPU compatible, YOLOv5 saca provecho de esta capacidad, acelerando el entrenamiento y la inferencia. La compatibilidad con CUDA también reduce el tiempo de procesamiento en comparación con modelos que no son tan optimizados para GPU.
- **Detección de Objetos Precisa en Imágenes Complejas:** YOLOv5 ha mostrado un excelente rendimiento en la identificación de objetos en escenas complejas, una característica crucial en nuestro caso, en el que las vacas deben ser identificadas en imágenes tomadas desde un ángulo aéreo, a veces con condiciones ambientales variadas (como cambios en la iluminación).

Supuestos de modelado

En este modelo de detección de objetos, uno de los supuestos principales se refiere a la calidad de las imágenes en relación con la iluminación. Dado que las imágenes se capturan cada cinco minutos de forma continua, todos los días y a todas horas, existen variaciones significativas en la iluminación, dependiendo de las condiciones ambientales y la hora del día. Estos son los supuestos y observaciones clave sobre la calidad de los datos:

1. **Variaciones en la Iluminación:** Las imágenes presentan fluctuaciones en la saturación y el brillo debido al ciclo natural de luz y sombra (amanecer, atardecer y noche). Por tanto, algunas imágenes aparecen sobreexpuestas, con alta saturación de luz, mientras que otras son notablemente oscuras o incluso completamente negras en horas de baja visibilidad.
2. **Calidad Predominante de Brillo:** A pesar de estas variaciones extremas en ciertos momentos del día, asumimos que la mayoría de las imágenes utilizadas en el entrenamiento del modelo tienen una calidad adecuada de brillo y contraste. Estas imágenes representan el contexto de iluminación diurna, en el cual las vacas son claramente visibles y distinguibles. Esta calidad de imagen, que corresponde a un rango estándar de luz diurna, es la que guía el entrenamiento del modelo para una detección precisa.
3. **Robustez del Modelo ante Cambios de Iluminación:** Aunque existen variaciones en la calidad de las imágenes, el modelo YOLOv5 está diseñado para manejar ciertas inconsistencias en la iluminación y calidad visual de las imágenes de entrada. Debido a su arquitectura robusta y a las técnicas de procesamiento de imagen preentrenadas, asumimos que el modelo YOLOv5 mantendrá una buena precisión en la detección de vacas, incluso bajo condiciones de iluminación variables. Esta capacidad se traduce en un rendimiento relativamente consistente, permitiendo detectar los objetos deseados en diversas condiciones ambientales.
4. **Impacto en la Detección de Objetos:** Basados en estos supuestos de calidad de datos, anticipamos que el modelo YOLOv5 será más preciso en la detección de vacas durante períodos con iluminación estable (principalmente durante el día). Sin embargo, se espera que mantenga una precisión adecuada en condiciones de iluminación variable gracias a su preparación para manejar este tipo de inconsistencias.

Estos supuestos nos permiten definir las expectativas de rendimiento del modelo en diferentes condiciones de iluminación, manteniendo un enfoque en la calidad

predominante de las imágenes que representan la visibilidad diurna, pero reconociendo la robustez de YOLOv5 ante variaciones.

Diseño de las pruebas

Para entrenar, validar y evaluar el modelo YOLOv5 utilizado en la detección de vacas en imágenes aéreas, implementamos un plan de pruebas basado en la división estándar del conjunto de datos en tres partes: entrenamiento, validación y prueba, con los siguientes porcentajes:

- Entrenamiento (80%): El 80% del conjunto de datos fue asignado al entrenamiento del modelo. Este conjunto permitió al modelo aprender los patrones y características clave para identificar correctamente las vacas en imágenes aéreas, ajustando sus pesos mediante retropropagación para minimizar el error en la detección.
- Validación (10%): Un 10% de los datos se destinó a la validación, para monitorear el rendimiento del modelo durante el entrenamiento sin influir en el ajuste de los parámetros. A lo largo de las épocas, el conjunto de validación ayudó a identificar posibles problemas de sobreajuste (overfitting) y subajuste (underfitting), permitiendo ajustar hiperparámetros de manera adecuada en función del desempeño en este conjunto.
- Prueba (10%): Finalmente, el 10% restante se reservó para la prueba del modelo. Este conjunto no fue utilizado en el entrenamiento ni en la validación, y sirvió para evaluar la efectividad y generalización del modelo en datos que no había visto previamente. La precisión, el recall y otras métricas de desempeño en este conjunto reflejan la capacidad del modelo para realizar detecciones confiables en imágenes reales fuera del entorno de entrenamiento.

Este enfoque de 80/10/10 asegura una adecuada exposición del modelo a datos para el aprendizaje, mientras reserva una porción suficiente para medir su capacidad de generalización y su rendimiento en situaciones reales.

Ajuste de parámetros

1. Parámetro --img: 1920

Se eligió un tamaño de imagen de 1920 para preservar el nivel de detalle necesario para identificar correctamente las vacas en las imágenes aéreas. Un tamaño de imagen mayor permite al modelo capturar mejor los detalles y

bordes de los objetos, lo cual es crucial cuando se detectan objetos relativamente pequeños o en ambientes complejos como campos abiertos.

2. Parámetro --batch: 4

Utilizamos un tamaño de lote (batch size) de 4 para el entrenamiento. Este valor fue elegido para optimizar el uso de la memoria de la GPU disponible y garantizar una tasa de aprendizaje estable. Un batch size mayor podría mejorar la velocidad de convergencia, pero requeriría más memoria, mientras que un batch size menor podría hacer el entrenamiento ineficiente en términos de tiempo y estabilidad de aprendizaje.

3. Parámetro --epochs: 25

Configuramos el modelo para entrenarse durante 25 épocas. Este número de épocas permite que el modelo tenga suficiente tiempo de exposición a los datos para aprender patrones significativos sin sobreajustarse (overfitting). Dado que el conjunto de datos es específico y relativamente pequeño en comparación con grandes corpus de datos de detección de objetos, entrenar el modelo durante más de 25 épocas podría resultar en un ajuste excesivo.

4. Parámetro --weights: yolov5s.pt

Utilizamos yolov5s.pt como modelo base preentrenado. Este es el modelo YOLOv5 "small" (pequeño), optimizado para balancear velocidad y precisión. Elegimos esta versión porque proporciona un rendimiento adecuado para la detección en tiempo real, necesario para este proyecto, y es lo suficientemente ligero para ser entrenado y ejecutado en una GPU estándar.

5. Parámetro --device: 0

Especificamos --device 0 para realizar el entrenamiento en la GPU disponible (CUDA). La selección de la GPU como dispositivo permite acelerar significativamente el entrenamiento y procesamiento del modelo, aprovechando la capacidad de cómputo paralela de la GPU.

6. Parámetro lr0 (Learning Rate): 0.01

Se configuró la tasa de aprendizaje inicial (lr0) en 0.01 para controlar la magnitud de los ajustes de los pesos del modelo en cada iteración. Un valor de 0.01 es suficientemente alto para asegurar un aprendizaje eficiente sin riesgo de saltar soluciones óptimas.

7. Parámetro momentum: 0.937

El parámetro momentum se estableció en 0.937 para suavizar los cambios en la dirección de actualización de los pesos del modelo. Un momentum alto como 0.937 ayuda a acelerar la convergencia y estabilizar el proceso de aprendizaje, minimizando el efecto de cambios bruscos.

8. Parámetro weight_decay: 0.0005

Se utilizó un valor de weight_decay de 0.0005 para regularizar el modelo, limitando la magnitud de los pesos a través de una penalización en los pesos altos. Este valor ayuda a reducir el riesgo de sobreajuste, haciendo que el modelo se enfoque en patrones generales de las imágenes en lugar de detalles específicos o ruidos de cada imagen.

Descripción del modelo

1. Modelo y propósito:

YOLOv5 (You Only Look Once, versión 5) es un modelo de detección de objetos de última generación que se utiliza para localizar y clasificar objetos dentro de una imagen en tiempo real. En nuestro caso, YOLOv5 resulta particularmente adecuado para la detección de vacas en imágenes debido a su capacidad para manejar grandes volúmenes de datos de imagen y procesarlos rápidamente. Esto es importante en entornos como la supervisión agrícola, donde la eficiencia y la precisión son fundamentales para detectar y monitorear ganado.

2. Precisión y robustez esperada:

Durante el entrenamiento, se utilizaron hiperparámetros cuidadosamente seleccionados (como el tamaño de imagen de 1920 píxeles, batch size de 4 y 25 epochs) que optimizan el modelo para este conjunto de datos. YOLOv5 ha demostrado ser robusto frente a variaciones de iluminación en las imágenes, una característica crucial en nuestro caso, dado que las imágenes contienen inconsistencia en la iluminación debido a las condiciones de captura. La arquitectura de YOLOv5 incorpora mejoras en el procesamiento de datos, como la capacidad de ajustar sus anclas de detección en función de la estructura del dataset (Best Possible Recall - BPR del 0.996), permitiéndole detectar objetos de manera precisa en diferentes condiciones.

3. Interpretación del modelo y dificultades:

YOLOv5 está diseñado para reconocer y detectar patrones espaciales en las imágenes, lo que permite una interpretación intuitiva de los resultados. Sin embargo, debido a la complejidad de la red y la sensibilidad a configuraciones como los valores de las anclas y la forma de los objetos en las imágenes, el modelo requiere un ajuste fino de sus parámetros. Durante el proceso de entrenamiento, se ajustaron múltiples

hiperparámetros (e.g., momentum, weight_decay) para maximizar la precisión en el contexto específico de la detección de ganado.

4. Características útiles del modelo:

Algunas características específicas de YOLOv5 que serán útiles en futuras aplicaciones incluyen:

- **Transfer learning:** La posibilidad de transferir el aprendizaje de otros modelos permite entrenar con datasets personalizados, optimizando el rendimiento para contextos específicos.
- **Compatibilidad con Augmentations:** Durante el entrenamiento, el modelo utiliza técnicas de aumento de datos (e.g., Blur, MedianBlur) para simular diferentes condiciones de captura y fortalecer su generalización.
- **Eficiencia en recursos:** A pesar de su complejidad, YOLOv5 está optimizado para correr eficientemente en GPUs de capacidades moderadas como la NVIDIA GeForce RTX 3070 Ti utilizada en este entrenamiento.

5. Valores de los parámetros:

A continuación, se listan algunos parámetros clave utilizados en el entrenamiento:

- lr0 (Learning Rate): 0.01
- momentum: 0.937
- weight_decay: 0.0005
- epochs: 25
- batch_size: 4

Estos parámetros fueron ajustados para balancear la precisión y la velocidad de convergencia en el conjunto de datos de imágenes de vacas.

6. Descripción técnica del modelo:

La arquitectura de YOLOv5 incluye una serie de capas convolucionales y bloques de detección que capturan características importantes en las imágenes:

- **C3 Blocks y SPPF Layers:** Facilitan la extracción de patrones visuales complejos a diferentes escalas, lo cual es crucial para detectar objetos de diferentes tamaños.
- **Multi-anchor System:** Permite detectar múltiples objetos en una sola pasada, aumentando la eficiencia del modelo.
- **Upsampling and Concat Layers:** Estos elementos ayudan a refinar las predicciones al fusionar características de distintas resoluciones y escalas.

7. Comportamiento del modelo y patrones detectados:

Durante el entrenamiento, los resultados muestran que el modelo es capaz de detectar patrones relevantes en las imágenes:

- Precisión (P) y Recall (R): En la última epoch, la precisión fue de 0.987 y el recall de 0.981, valores que indican una alta capacidad de YOLOv5 para detectar correctamente las vacas en el conjunto de validación.
- mAP 50 y mAP 50-95: La métrica mAP 50 (mean Average Precision para un IoU de 50) alcanzó valores superiores a 0.9 desde los primeros epochs, lo que señala una buena generalización del modelo en la detección de vacas.

Esta configuración y el comportamiento en entrenamiento sugieren que YOLOv5 es un modelo adecuado y robusto para el contexto de monitoreo de ganado en imágenes.

8. Interpretación del comportamiento del modelo

En esta sección, se analizarán los resultados obtenidos durante el proceso de entrenamiento de YOLOv5 en el contexto de monitoreo de ganado en imágenes, destacando los indicadores clave que reflejan el desempeño y robustez del modelo.

8.1 Precisión (Precision)

Los resultados indican que YOLOv5 fue efectivo para reducir los falsos positivos en la detección de ganado (0.987 de precisión). Esto es crucial en un entorno de monitoreo, ya que minimiza las detecciones erróneas y permite que el modelo sea confiable en identificar únicamente a los animales, evitando clasificaciones innecesarias de objetos que no sean ganado.

8.2 Exhaustividad (Recall)

Los resultados reflejan la capacidad del modelo para detectar la mayor cantidad posible de instancias de ganado en las imágenes. Un 0.98 en recall es particularmente relevante en este contexto, ya que una alta exhaustividad asegura que el modelo no pase por alto animales presentes en las imágenes, mejorando así su efectividad para monitorear el ganado en tiempo real.

8.3 Precisión Media Promedio (mAP)

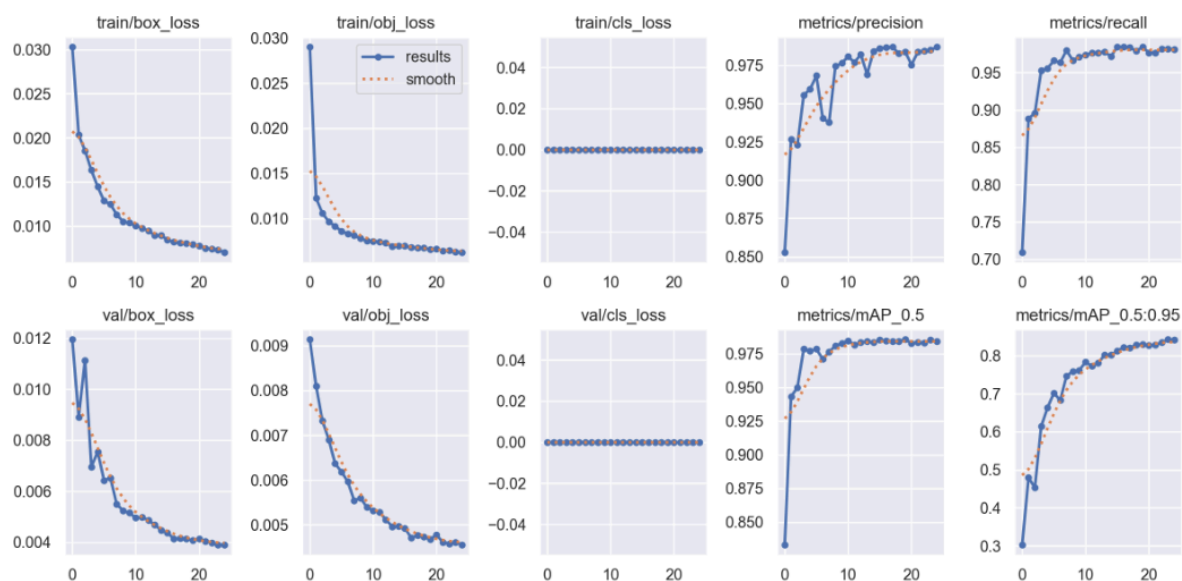
Los resultados son un indicador importante de la precisión general de las predicciones. Un valor de 0.98 de mAP sugiere que el modelo no solo detecta adecuadamente las instancias de ganado, sino que también posiciona correctamente las cajas delimitadoras alrededor de ellas. Este comportamiento es crucial en la práctica, pues asegura que las detecciones sean precisas tanto en cantidad como en localización.

8.4 Gráfica de Pérdida (Loss)

Durante el entrenamiento, la función de pérdida disminuyó consistentemente, estabilizándose en torno a 0.007. Este comportamiento indica una buena convergencia del modelo, lo que sugiere que los parámetros se ajustaron adecuadamente a los datos de entrenamiento. La disminución rápida de la pérdida en las primeras etapas, seguida de una estabilización, refleja que el modelo logró aprender las características esenciales de las imágenes y que ha alcanzado un punto donde las mejoras adicionales son mínimas.

Los resultados del entrenamiento se pueden acceder en este [link](#).

Evaluación del modelo



Resumen de Resultados

El modelo YOLOv5 fue evaluado con base en su capacidad para cumplir el objetivo de minería de datos: detectar y clasificar vacas en imágenes aéreas, permitiendo el conteo y caracterización de la población ganadera según su postura y ubicación. Los resultados de las métricas obtenidas durante el entrenamiento y validación son las siguientes:

1. Pérdidas (Loss)

Las pérdidas durante el entrenamiento (train/box_loss, train/obj_loss, train/cls_loss) disminuyeron consistentemente a lo largo de las 25 épocas, indicando que el modelo fue capaz de aprender los patrones presentes en los datos.

Las pérdidas en la validación (val/box_loss, val/obj_loss, val/cls_loss) también mostraron una tendencia decreciente, lo que sugiere que el modelo generalizó adecuadamente a datos no vistos durante el entrenamiento.

2. Precisión y Recall

La precisión alcanzó un valor cercano al 97.5% hacia el final del entrenamiento, indicando una alta capacidad del modelo para evitar falsos positivos.

El recall llegó a valores superiores al 95%, demostrando la efectividad del modelo para detectar todas las instancias relevantes.

3. mAP (Mean Average Precision)

El mAP@0.5 alcanzó aproximadamente 97.5%, mientras que el mAP@0.5:0.95 se estabilizó alrededor del 80%. Esto indica un excelente desempeño en la detección de bounding boxes a diferentes umbrales de IoU (Intersection over Union), con un equilibrio sólido entre precisión y sensibilidad.

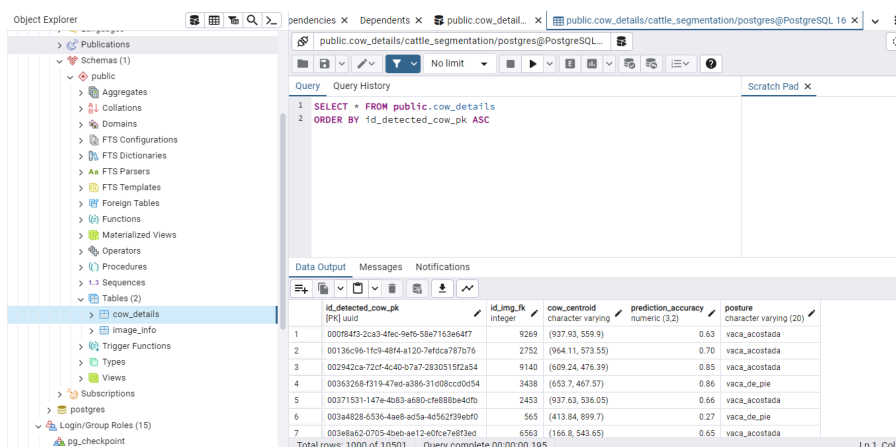
Interpretación de Resultados y Calidad del Modelo

- **Precisión y Recall:** Ambos indicadores colocan al modelo en una categoría de alta calidad, ya que logran superar el umbral del 95%.
- **mAP:** El valor de mAP@0.5 refleja un modelo altamente confiable para detección en el contexto de negocio. El valor más bajo de mAP@0.5:0.95 está en línea con expectativas razonables, dada la dificultad de evaluar con estrictos umbrales de IoU.

Interpretación en Términos de Negocio

El modelo cumple satisfactoriamente el criterio de éxito establecido:

Se logra identificar automáticamente la ubicación geográfica de las vacas en las imágenes. Los resultados se almacenan en una base de datos funcional y en producción, la cual registra información clave como el identificador de la imagen, la marca de tiempo y el conteo de vacas detectadas.



id_detected_cow_pk	id_img_fk	cow_centroid	prediction_accuracy	posture
000f8473-2ca3-4fec-9ef5-58e7163e6477	9269	(937.93, 559.9)	0.63	vaca_acostada
00136c96-1fc9-48f4-a120-7efdc9787b76	2752	(964.11, 573.55)	0.70	vaca_acostada
002942ca-72cf-4c40-b7e7-2830515f2a54	9140	(609.24, 476.39)	0.85	vaca_acostada
00363268-f319-47ed-a386-31d08cc0d054	3438	(653.7, 467.57)	0.86	vaca_de_pie
00371531-147e-4b83-a680-cfe888be4d0f	2453	(937.63, 536.05)	0.66	vaca_acostada
003a4828-6536-4ae8-ad5a-4d562f99ebf0	565	(413.84, 899.7)	0.27	vaca_de_pie
003e8a52-0705-4deb-ae12-e0fce7e87aed	6563	(166.8, 543.65)	0.65	vaca_acostada

Total rows: 1000 of 10501 Query complete 00:00:00.195

Impacto en el Negocio:

Este sistema automatizado permite una cuantificación precisa y rápida del ganado en distintas áreas del rancho, optimizando recursos humanos y tiempo.

Ofrece información geográfica que puede ser usada para planificar mejor la distribución y manejo del ganado (y de los espacios de las camas) en tiempo real.

Parámetros revisados y ajustados

Una de las principales ventajas de YOLOv5 es su capacidad de ajustar dinámicamente la mayoría de sus parámetros durante el entrenamiento, optimizando así su rendimiento y permitiéndole aprender de manera más eficaz los patrones distintivos del ganado en imágenes aéreas. Este ajuste automático se aplica a varios parámetros clave del modelo, a excepción del tamaño del lote (`batch_size`) y el número de épocas (`epochs`), que deben definirse de antemano.

Para este modelo, configuramos `batch_size` en 4 y el número de épocas en 25. Estos valores fueron elegidos cuidadosamente para equilibrar el uso de los recursos computacionales disponibles, maximizando la eficiencia de la GPU sin comprometer la estabilidad ni la calidad del aprendizaje. Al fijar el tamaño del lote en 4, aseguramos que el modelo procese suficientes ejemplos por iteración sin sobrecargar la memoria. Del mismo modo, el límite de 25 épocas proporciona un tiempo adecuado para que el modelo aprenda patrones relevantes en el conjunto de datos sin riesgo de sobreajuste.