

Modelado III

Adrián Matute, Pablo Martínez, Osvaldo Del Valle, Andres Callealta, Jorge Martínez
Tecnológico de Monterrey Campus Querétaro, México

hiperparámetros estándar.

1. Técnica de modelado

En esta iteración se busca entrenar y evaluar un modelo basado en la arquitectura Faster R-CNN como alternativa al modelo híbrido, diurno y nocturno de YOLOv8. La elección de Faster R-CNN se realizó para explorar sus capacidades de detección de objetos en escenarios controlados y con condiciones de iluminación variables, considerando su reputación por la precisión en bounding boxes.

- Fortalezas:
 - Alto nivel de precisión en bounding boxes debido a su enfoque basado en propuestas de región (RPN).
 - Capacidad de trabajar con imágenes con objetos solapados o en movimiento.
 - Respaldo de investigaciones académicas y aplicaciones prácticas en tareas de visión computacional.
- Limitaciones:
 - Velocidad: Inferencia más lenta que YOLOv8, especialmente en tiempo real.
 - Requerimientos computacionales: Alta demanda de recursos, lo que puede limitar su uso en hardware con capacidades restringidas.
 - Adaptabilidad: Menos eficiente para condiciones mixtas, especialmente en iluminación nocturna.

2. Supuestos de modelado

- Dataset:
 - Imágenes etiquetadas con precisión y alineadas con las condiciones del proyecto (día y noche).
 - Incluye variaciones de densidad y situaciones de vacas en movimiento o solapadas.
- Configuración del Modelo:
 - Entrenamiento inicial con un modelo pre entrenado en COCO, lo que permite aprovechar características generales preaprendidas.
 - Evaluación inicial con

3. Diseño de pruebas

La detección de objetos por medio de múltiples bounding boxes presenta 3 resultados que utilizaremos para medir la precisión de la detección.

- Las cajas predichas, coordenadas del centro de la caja, ancho y alto.
- Las clases a las que pertenece cada objeto para dichas cajas, en nuestro caso es únicamente *vaca o fondo*.
- La confianza que tiene el modelo de que la clase predicha.

3.1 Métricas de evaluación

Para evaluar la calidad y validez del modelo en la tarea de detección y conteo de vacas en la fila de espera, se utilizaron las siguientes métricas clave:

- **Precisión:**
 - a. La proporción de verdaderos positivos entre todas las predicciones de clases, nos muestra que tan propenso es el modelo a generar falsos positivos.
 - b. Resultado esperado: Superior a 85% para garantizar un conteo confiable en imágenes diurnas y nocturnas.
- **Recall (Sensibilidad):**
 - a. La proporción de verdaderos positivos sobre los positivos esperados, mostrando la habilidad del modelo de predecir todos los objetos presentados.
 - b. Resultado esperado: Superior al 90%, para minimizar vacas no detectadas.
- **mAP@50 y mAP@50-95:**
 - a. mAP@50: Precisión media para un umbral de intersección sobre unión (IoU) del 50%.
 - b. mAP@50-95: Precisión media en múltiples umbrales de IoU (50%-95%).
 - c. Resultado esperado:
 - i. mAP@50: Superior al 95% en condiciones diurnas.
 - ii. mAP@50-95: Superior al 80% para condiciones mixtas.

Estas métricas permiten calcular un porcentaje de error y observar cómo el modelo se ajusta a los objetivos planteados, tanto en términos de detección como de conteo.

3.2 División del Dataset

El dataset total de 8,087 imágenes fue dividido cuidadosamente para asegurar una representación balanceada de condiciones diurnas y nocturnas:

1. Porcentaje de división:
 - 80% para entrenamiento: El modelo ajusta sus parámetros para mejorar la detección de vacas.
 - 20% para validación: Evalúa el rendimiento del modelo en imágenes no vistas previamente.
2. Subconjuntos por iluminación:
 - Imágenes diurnas: Capturadas entre las 6:00 am y las 8:00 pm.
 - Imágenes nocturnas: Capturadas entre las 8:01 pm y las 5:59 am.
 - Cada modelo fue entrenado y validado en su respectivo subconjunto, mejorando la precisión en cada condición de iluminación.

3.3 Plan de entrenamiento, pruebas y evaluación

Para la fase de pruebas y entrenamiento del modelo, se tomó la decisión de hacer dos iteraciones, esto para conocer cuales son los resultados generados en la primera iteración, seguido de hacer diversos cambios en el modelo para identificar de qué manera el modelo puede mejorar o reducir la calidad de los resultados.

- **Entrenamiento:**
 - Modelo base: Se utilizó Transfer Learning con un modelo pre entrenado en el dataset COCO, para aprovechar las características generales aprendidas.
 - Épocas: Se realizaron hasta 150 iteraciones
- **Validación y pruebas:**
 - El conjunto de validación fue utilizado para calcular las métricas de precisión, recall y mAP en imágenes no vistas previamente.
 - Se generaron gráficos de pérdida y métricas para observar la estabilidad y el rendimiento del modelo durante las épocas de entrenamiento.
- **Optimización y ajustes:**
 - Tasa de aprendizaje: Ajustada iterativamente para garantizar una convergencia eficiente.
 - Tamaño de lote: Modificado según los recursos de hardware disponibles.
 - Parada temprana: Implementada para detener el entrenamiento si no se observaron mejoras significativas en las métricas de validación.

3.4 Resultados Esperados y Actividades de Prueba

1. Conteo de vacas en la fila de espera:
 - Prueba: Evaluar la precisión del modelo para identificar la cantidad exacta de vacas en imágenes.
 - Condiciones: Considerar situaciones con vacas solapadas o en movimiento.
 - Métricas clave: Precisión y recall superiores al 85%.
2. Análisis de aglomeraciones para identificar cuellos de botella:
 - Prueba: Detectar patrones de aglomeración en secuencias de imágenes y evaluar cambios en la densidad de vacas en el área de espera.
 - Evaluación: Se utilizaron métricas de recall en series temporales para determinar momentos de alta congestión.

3.5 Justificación del Diseño de Test

El diseño de test asegura que las métricas elegidas sean representativas de los objetivos planteados:

- Precisión y Recall: Directamente alineadas con la capacidad del modelo para cumplir con el conteo de vacas.
- mAP: Garantiza que las detecciones sean consistentes a través de múltiples umbrales, lo que es crucial para minimizar errores en el análisis de aglomeraciones.

Estas pruebas validan que el modelo cumple con los requisitos técnicos y de negocio, maximizando su capacidad para detectar patrones útiles para la toma de decisiones en el sistema de ordeño.

4. Ajuste de parámetros

Para el modelo de Faster R-CNN, se consideraron los siguientes parámetros a modificar para las dos iteraciones realizadas:

Optimizador

- Su función es ajustar los pesos y sesgos de la red neuronal de manera que el error en las

predicciones se minimice

Threshold de puntuación

- Define el nivel mínimo de confianza que una predicción debe tener para ser considerada válida.

5. Descripción del modelo Faster-RCNN

5.1 Primera iteración

El primer modelo se entrenó con el 80% del dataset (6,470 imágenes), contemplando las dos condiciones (noche y día) en la que trabajará este modelo, el 20% restante se usó para validación (1,617 imágenes).

- **Épocas: 5**
 - Se eligió este número de épocas como punto inicial para observar la convergencia de la pérdida. Además, en la próxima iteración se mantuvo el mismo valor para no generar un tiempo de entrenamiento extremadamente largo, ya que esta red llega a usar muchas variables, por cada época llega a tardar unos 40 minutos.
- **Tamaño de lote: 4**
 - Seleccionado debido a las limitaciones de memoria del hardware disponible (GPU). Intentar tamaños mayores resultaba en un aumento excesivo del tiempo por época o errores de memoria.
- **Weight decay: 0.0005**
 - Este valor se mantuvo porque valores mayores (por ejemplo, cercanos a 1) generaban un incremento en la pérdida, afectando la estabilidad del modelo. En cambio, valores más pequeños no ofrecieron mejoras notables en el rendimiento.
- **Taza de aprendizaje: 0.0001**
 - Se utilizó un valor bajo para evitar que el modelo hiciera saltos grandes durante el ajuste de los pesos, lo que podría haber evitado la convergencia. Tasas de aprendizaje más altas provocan inestabilidad.
- **Threshold de puntuación: 0.5**
 - Este threshold filtra las predicciones para conservar sólo aquellas con alta confianza. Se optó por este valor para tener una buena predicción de las

detecciones, al igual que buscamos un buen valor de recall.

- Optimizador: SGD
 - Se utilizó el optimizador SGD con un momentum de 0.9. Este optimizador fue elegido en la primera iteración para comparar su rendimiento en términos de estabilidad y velocidad de convergencia.

5.2 Segunda iteración

Al igual que la primera iteración, se entrenó con el 80% del dataset y el 20% restante se usó para validación.

- **Épocas: 5**
 - Se mantuvo el mismo número de épocas para no generar un aumento excesivo de tiempo de entrenamiento.
- **Tamaño de lote: 4**
 - Se mantuvo el tamaño de lote en 4 debido a las limitaciones del hardware disponible.
- **Weight decay: 0.0005**
 - Este valor se mantuvo porque valores mayores generaban un incremento en la pérdida, afectando la estabilidad del modelo.
- **Taza de aprendizaje: 0.0001**
 - Se mantuvo un valor bajo de tasa de aprendizaje para evitar saltos grandes durante el ajuste de los pesos.
- **Threshold de puntuación: 0.75**
 - Este valor fue seleccionado para filtrar las predicciones de baja calidad, priorizando precisión sobre recall. En la segunda iteración con Adam, se mantuvo este umbral para asegurar una evaluación de las predicciones más confiable.
- Optimizador: Adam
 - Se cambió a Adam debido a su capacidad para ajustar dinámicamente los parámetros de aprendizaje, lo que resultó en un entrenamiento más rápido y eficiente, especialmente con datasets más complejas. Adam ayudó a reducir la oscilación de la pérdida, mostrando mejoras notables en el tiempo de entrenamiento y precisión.

6. Evaluación del modelo

6.1 Primera iteración

Métrica	Valor
Precisión	20.50%
Recall	17.10%
mAP@50	20.47%
mAP@50-95	16.19%

6.1.1 Robustez

El modelo mostró un desempeño limitado en las condiciones actuales de iluminación. Aunque se obtuvo una precisión general decente, la capacidad de detección no fue consistente debido a la baja mAP@50 y mAP@50-95, lo que indica que el modelo tiene dificultades en identificar correctamente los objetos en la mayoría de las pruebas. El modelo no logra distinguir con claridad entre objetos pequeños y medianos, lo que se refleja en los resultados nulos para el área "small" y "medium". Además, el recall fue relativamente bajo, especialmente en IoU, lo que sugiere que muchas vacas no fueron detectadas en el conjunto de pruebas.

6.1.2 Limitaciones

- **Detección de vacas:** El modelo presenta errores frecuentes en la detección de las vacas, sobre todo cuando las condiciones de iluminación son cambiantes. La tasa de recall baja demuestra que muchas vacas no se detectan adecuadamente.
- **Objetos superpuestos:** El modelo presenta problemas al intentar identificar vacas parcialmente cubiertas o en posiciones complejas, lo que provoca la detección de "súper vacas" o combinaciones incorrectas de diferentes partes de las vacas.
- **Confusión con seres humanos:** En ciertas ocasiones, el modelo confundió personas con vacas, lo que indica que aún existe un solapamiento significativo en las características visuales.

6.1.3 Dificultades

- **Tiempo de entrenamiento:** El modelo tardó aproximadamente 3 horas en entrenar durante 5 épocas en una GPU 4050. Esto representa un tiempo considerable debido a la cantidad de datos procesados, lo que puede ser optimizado en iteraciones futuras.
- **Generalización:** El modelo presentó problemas de generalización debido a las condiciones de iluminación volátiles. Aunque el entrenamiento se realizó bajo condiciones diurnas, los cambios de luz durante el día o la transición a la noche afectaron la precisión. Además, el modelo muestra un rendimiento pobre en la detección de vacas pequeñas y medianas, lo cual debe abordarse para mejorar la detección en diferentes condiciones de tamaño y luz.

6.2 Segunda iteración

Métrica	Valor
Precisión	45.50%
Recall	25.79%
mAP@50	40.5%
mAP@50-95	35.2%

6.2.1 Robustez

El modelo mostró un desempeño ligeramente mejor en términos de mAP@50 con respecto a la primera iteración, aunque la mAP@50-95 y la recall siguen siendo bajas. Aunque se ha observado una mejora promedio en la precisión, sigue siendo un desafío identificar objetos en áreas pequeñas y medianas. A pesar de algunos avances, la capacidad de generalización en condiciones diversas de iluminación y tamaños de vacas sigue siendo limitada.

6.2.2 Limitaciones

- **Detección de vacas:** La tasa de recall sigue siendo baja, lo que indica que el modelo no detecta correctamente muchas vacas en el conjunto de pruebas. Los resultados nulos para las áreas pequeñas y medianas son una limitación importante, pues este tipo de vacas parece ser difícil de detectar.
- **Confusión de objetos:** El modelo sigue generando predicciones incorrectas, como la detección de seres humanos como vacas, lo que refleja la necesidad de mejorar la capacidad del

modelo para diferenciar entre vacas y otros objetos similares.

6.2.3 Dificultades

- **Tiempo de entrenamiento:** El modelo tardó aproximadamente 3 horas en entrenar 5 épocas en una GPU 4050, lo cual es un tiempo relativamente largo para el rendimiento obtenido. Sin embargo, en comparación con la iteración anterior, el tiempo de entrenamiento no varió considerablemente.
- **Generalización:** Aunque la mAP@50 no mostró cambios significativos, la generalización en condiciones cambiantes sigue siendo una dificultad importante. El modelo sigue enfrentando problemas con la variabilidad de la iluminación, especialmente con vacas pequeñas y medianas que no se detectan adecuadamente en varias situaciones.

Comparación de Resultados

Análisis de Precisión y Recall:

- El modelo tiene una precisión consistente de alrededor del 45.50%, lo cual indica que, aunque el modelo identifica vacas, sigue siendo relativamente poco preciso en cuanto a la cantidad correcta de vacas que se detectan por imagen.
- Recall es baja, con un máximo de 17.10% en la primera iteración, lo que implica que el modelo no detecta una gran proporción de las vacas presentes en las imágenes.

mAP (Mean Average Precision):

- El **mAP@50** y **mAP@50-95** no muestran una gran diferencia entre las iteraciones, indicando que el modelo ha tenido dificultades en detectar correctamente en varias situaciones de IoU (Intersección sobre la Unión) y áreas pequeñas o medianas. Este comportamiento es relevante, ya que la **mAP@50-95** refleja la capacidad del modelo para generalizar mejor en diferentes casos de intersección, lo que es crucial cuando se enfrentan objetos con distintos tamaños y condiciones.

Evaluación con respecto a los criterios de éxito en los objetivos de minería de datos

Detección de Aglomeraciones:

- **Impacto en el objetivo de minería de datos:** El modelo muestra dificultades para detectar vacas en áreas pequeñas y medianas, lo que impide realizar un análisis detallado sobre las aglomeraciones. A pesar de obtener un mAP@50 decente, el modelo aún no es suficientemente robusto para detectar con precisión aglomeraciones de vacas en diferentes condiciones de iluminación, lo que afecta su capacidad para identificar patrones significativos en la fila.

Identificación de Cuellos de Botella:

- **Impacto en el objetivo de minería de datos:** La falta de una alta tasa de recall limita la capacidad del modelo para identificar de manera confiable los cuellos de botella. Si las vacas no son detectadas correctamente, se pierde información clave sobre las aglomeraciones, lo que dificulta la identificación de los cuellos de botella en el proceso de ordeño.

Conclusión de resultados de los modelos

Con los resultados proporcionados por parte de las tres soluciones (YOLOv8 híbrido, diurno y nocturno, y Faster R-CNN híbrido), se puede realizar una evaluación comparativa basada en criterios clave como precisión, recall, mAP y la efectividad en tareas de minería de datos.

1. Modelo Faster R-CNN:

- Precisión: 45.50%
- Recall: 17.10%
- mAP@50: 20.47%
- mAP@50-95: 16.19%

Este modelo, a pesar de haber sido entrenado con técnicas avanzadas como Faster R-CNN, se presenta un desempeño menor a YOLOv8, con resultados más bajos en precisión y recall.

2. Modelo YOLOv8 híbrido:

- Precisión: 95.14%
- Recall: 95%
- mAP@50: 96%
- mAP@50-95: 84%

El modelo YOLO v8 se destaca por su capacidad para generalizar en condiciones mixtas, mostrando un alto desempeño en tareas como el conteo de vacas y la detección de aglomeraciones. Su precisión y recall elevados lo hacen una solución sólida para el monitoreo en condiciones variables de luz.

3. Modelo YOLOv8 diurno:

- Precisión: 96%
- Recall: 96%
- mAP@50: 98%
- mAP@50-95: 87%

Este modelo ha sido entrenado exclusivamente con imágenes diurnas, lo que le permite aprovechar las condiciones de luz natural para obtener una mayor precisión y robustez en la detección de vacas. Es ideal para supervisión durante el día, mostrando una alta capacidad para detectar detalles específicos y patrones de flujo.

4. Modelo YOLOv8 nocturno

- Precisión: 92%
- Recall: 92%
- mAP@50: 95%
- mAP@50-95: 71%

Aunque el modelo nocturno ha sido entrenado con imágenes tomadas en condiciones de poca luz, su precisión es ligeramente inferior en comparación con el modelo diurno. Esto es esperado debido a la limitación de datos y la dificultad inherente de trabajar con imágenes nocturnas. Sin embargo, sigue siendo útil para monitorear aglomeraciones y realizar tareas de supervisión en ambientes de baja iluminación.

Modelos y los Criterios de éxito de minería de datos

Precisión y Recall:

El modelo diurno sobresale en precisión con un 96.4%, seguido del modelo híbrido con un 95.14%, el modelo nocturno con un 92.9% y, finalmente, el modelo Faster R-CNN con resultados significativamente más bajos.

En cuanto al recall, el modelo híbrido y el modelo diurno obtienen valores elevados (95% y 96.4%, respectivamente), lo que implica que ambos son muy buenos para identificar patrones de densidad y flujo de vacas, lo que es clave para la optimización de la fila de ordeño.

mAP@50 y mAP@50-95:

El modelo híbrido obtiene el mejor desempeño en mAP@50 con un 96%, mientras que el modelo diurno tiene un valor cercano del 96.4%. Ambos son notablemente más altos que el modelo nocturno (92.9%) y el Faster R-CNN (con mAP en torno al 16%).

El modelo híbrido también tiene un buen desempeño en mAP@50-95 (84%), mientras que el modelo nocturno y el modelo diurno tienen un mAP@50-95 considerablemente más bajo, alrededor del 70% en el diurno y 63% en el nocturno.

Evaluación del Desempeño en Términos de Negocio

1. Conteo de Vacas y Aglomeraciones:

- El modelo híbrido con un 95.14% de precisión y 95% de recall es el más adecuado para contar vacas en condiciones de luz mixtas (día/noche), asegurando que el proceso de monitoreo sea fiable en cualquier momento del día.
- Los modelos diurno y nocturno son más especializados y pueden ser útiles cuando se requiere un alto nivel de precisión en condiciones controladas de luz. El modelo diurno tiene la ventaja de un rendimiento superior en el monitoreo diurno, mientras que el modelo nocturno es útil en condiciones de poca luz pero presenta algunas limitaciones.
- El modelo de Faster R-CNN no es lo suficientemente efectivo para detectar precisamente vacas en diferentes condiciones de iluminación y cantidades, lo que no nos permite hacer un análisis de patrones significativos en la fila.

2. Optimización de la Programación y Cuellos de Botella:

- La capacidad de los modelos YOLOv8 (especialmente el híbrido) para detectar aglomeraciones es crucial para identificar cuellos de botella en la fila de ordeño, lo que puede mejorar la eficiencia operativa. El modelo híbrido tiene la ventaja de funcionar bien en entornos cambiantes, lo que lo convierte en la

opción más flexible para operaciones en tiempo real.

[performance-metrics/#object-detection-metrics](#)

Impacto en el Objetivo de Minería de Datos

- El modelo híbrido ha demostrado ser la opción más robusta y precisa para el conteo y monitoreo de vacas en la fila de espera, permitiendo un análisis detallado de las aglomeraciones y el comportamiento de las vacas en diferentes condiciones.
- Aunque el modelo diurno sobresale en condiciones de luz natural, su aplicabilidad en condiciones nocturnas es limitada.
- El modelo nocturno es útil para supervisión básica, pero su precisión más baja hace que no sea tan eficaz en situaciones de alta densidad.

Conclusiones y Selección de los Mejores Modelos

1. **Mejores Modelos:**
 - **Modelo YOLOv8 Híbrido:**
Seleccionado como el modelo más eficiente y robusto debido a su capacidad de generalizar en condiciones mixtas, su precisión y recall elevados, y su eficiencia en tiempo de entrenamiento.
 - **Modelo Diurno YOLOv8:**
Seleccionado para tareas específicas durante el día, especialmente cuando se necesita una alta precisión en un entorno bien iluminado.
2. **Impacto en los Objetivos de Negocio:**
 - El modelo híbrido cumple con los objetivos de negocio al permitir un monitoreo preciso y eficiente tanto durante el día como en la noche, lo que optimiza la programación y mejora el bienestar animal al reducir las aglomeraciones.

7. Referencias

Ultralytics. (2024, 1 octubre). YOLO Performance Metrics. Ultralytics YOLO Docs.
<https://docs.ultralytics.com/guides/yolo->