Integración de modelos de inteligencia artificial mediante la herramienta SelSegToYolo para la clasificación y detección de animales en imágenes

J.A. Mestra*, E. Castrillon†, J.C. Amezquita-Rios**
*Correo electrónico: andresmestra11a@gmail.com
†Correo electrónico: cestiven156@gmail.com
**Correo electrónico: arvicami@gmail.com

***Abstract*—Presentamos una solución de inteligencia artificial (IA) diseñada para permitir la clasificación precisa de imágenes, al tiempo que es accesible para aquellos con conocimientos básicos en programación. Nuestra aproximación se apoya en el uso de herramientas establecidas como YOLO (You Only Look Once) y SAM (Selective Attention Mechanism), complementadas por la utilidad de Labelme2YOLO, Esto ha resultado en un integracion que logra una tasa de precisión de aproximadamente el 85% en ejemplos positivos clasificados correctamente.**

La metodología se centra en la adopción de modelos y parámetros previamente optimizados, lo que simplifica significativamente el proceso de aprendizaje. Para aumentar la accesibilidad, hemos desarrollado esta solución de IA de manera que pueda ser ejecutada de manera sencilla en la plataforma de Google Colab, lo que facilita su implementación.

En nuestro estudio, destacamos la importancia de aprovechar las herramientas ya existentes en la creación de IA especializadas. Nuestro análisis detallado respalda cómo la combinación de accesibilidad y eficacia puede abrir nuevas posibilidades en la implementación de soluciones de IA en una variedad de campos.

***Index Terms*—Inteligencia Artificial, Visión de Computadora, Mascara, Segmentación, Clasificación, YOLO, SAM**

I. Introducción

La IA desempeña un papel fundamental en una amplia variedad de campos, y uno de los ámbitos donde ha demostrado su potencial es el análisis de imágenes. Dentro de esta disciplina, la tarea de clasificar imágenes en función de su contenido, ya sea identificando animales, estructuras u objetos, se presenta como un desafío intrigante. En este artículo, exploramos el desarrollo de una integracion destinada a la creación de una IA especializada en la clasificación de imágenes según un conjunto de datos específico.

En el ámbito de la IA, es común aprovechar modelos previamente probados y optimizados para impulsar el desarrollo de soluciones más avanzadas. Esta práctica elimina la necesidad de empezar desde cero, evitando la tediosa tarea de ajustar parámetros y configuraciones óptimas. Siguiendo esta premisa, nuestro estudio presenta la creación de una IA que se enfoca exclusivamente en el objeto de interés. Esta especialización resulta en un proceso de aprendizaje más preciso y eficiente.

A lo largo de este artículo, describiremos las herramientas y métodos que utilizamos, además de presentar los resultados obtenidos. Abordamos este desafío conscientes de su complejidad y de la contribución que nuestro trabajo representa en el campo de la IA.

II. Revisión de la Literatura

Dos herramientas fundamentales en el campo de la visión por computadora y la IA han sido SAM (Segmentation and Annotation of Images) [1] y YOLO (You Only Look Once) [2]. Ambas desempeñaron un papel crucial en la evolución de nuestra investigación.

SAM se destacó como una herramienta esencial en la segmentación de imágenes, permitiendo la identificación y aislamiento de regiones de interés en las imágenes. Una de sus contribuciones más significativas fue la capacidad para reducir el ”ruido” presente en las imágenes, lo que mejoraba la precisión en tareas posteriores de clasificación y deteccion.

YOLO representó un avance revolucionario al introducir un enfoque de detección en tiempo real. A diferencia de las técnicas usualmente usadas, YOLO permite la clasificación y deteccion de objetos en una sola pasada a través de una imagen. Lo que lo convierte en una herramienta poderosa para aplicaciones de detección y clasificación de objetos en tiempo real.

Sinergia entre SAM y YOLO

Una de las contribuciones clave de SAM fue su capacidad para preparar las imágenes al eliminar el ”ruido”. Esta reducción de ruido se tradujo en imágenes más limpias y precisas, lo que resultó beneficioso para YOLO en sus tareas de detección y clasificación.

A medida que avanzamos en este informe, se describirá en detalle las adaptaciones y mejoras realizadas en nuestro metodo de entrenamiento de un modelo de clasificación, destacando cómo hemos aprovechado estas herramientas ya existentes para crear una solución aún más efectiva y accesible para cualquier persona con conocimientos básicos de programación y con la utilización de herramientas de código abierto.

III. Metodología

A. Análisis de datos de las imágenes

En la fase inicial de este proyecto, llevamos a cabo un análisis exhaustivo de las propiedades de todas las imágenes

presentes en el dataset utilizando Python. Comprender estos datos resultó ser un paso crucial para avanzar en la investigación y preparar adecuadamente los datos antes de desarrollar un modelo de IA.

Una de las primeras observaciones destacadas se relacionó con la inconsistencia en los nombres de archivo, careciendo de un patrón coherente que facilitara su identificación y gestión eficiente. Esta variabilidad en los nombres de archivo no solo dificultaba la comprensión de las imágenes, sino que también presentaba desafíos en la manipulación y el procesamiento de datos. Como respuesta a esta problemática, se procedió a estandarizar rigurosamente los nombres de las imágenes en un formato uniforme y fácilmente distinguible. Cada imagen ahora presenta un nombre en el formato 'categoria_N', donde 'N' representa un número de identificación único asignado a la imagen. Esta modificación no solo mejoró la identificación de las imágenes, sino que también simplificó significativamente la gestión y organización de los datos, lo que se tradujo en una mayor eficiencia en el flujo de trabajo del proyecto.

Además, la estandarización de nombres también abordó el desafío de nombres duplicados que se encontraba en las categorías 'birds' y 'fish'. Al asignar identificadores únicos a cada imagen dentro de estas categorías, se garantizó una distinción clara y precisa entre las imágenes, evitando cualquier ambigüedad en la referencia a las mismas. Esta medida resultó esencial para garantizar la integridad y la coherencia de los datos a lo largo del proyecto.

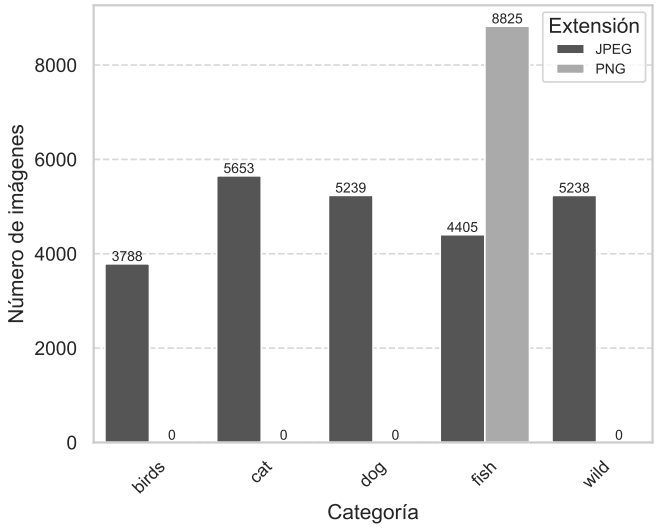

Fig. 1. Número de imágenes por categoría y extensión.

La figura 1 desempeña un papel fundamental en nuestra investigación, ya que proporciona una visión clara y concisa de la distribución de imágenes en nuestro conjunto de datos. Revela patrones significativos y posibles desequilibrios, como la existencia de archivos con diferentes extensiones en la categoría 'fish', lo cual se resuelve transformando las imágenes de png a jpg, logrando así un estándar para todo el dataset.

Además de las transformaciones previamente mencionadas, se implementó una modificación adicional en las imágenes. Esta transformación implicó la introducción de un relleno

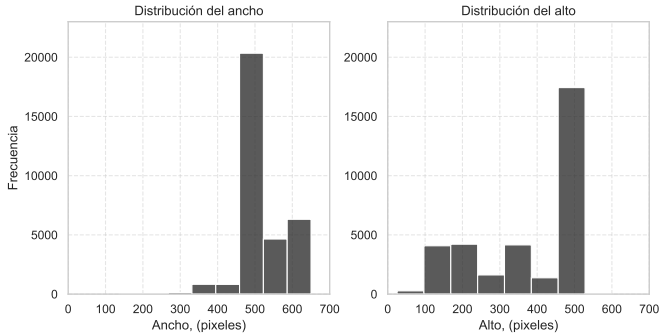

Fig. 2. Distribución del ancho y alto de las imágenes en el dataset.

blanco alrededor de aquellas imágenes que no tenían una relación de aspecto cuadrada original, debido a la diferencia de tamaños como se muestra en la Figura 2. El objetivo de esta modificación fue homogeneizar las dimensiones de todas las imágenes, convirtiéndolas en cuadrados. Posteriormente, se procedió a redimensionar todas las imágenes a un formato estándar de 256 por 256 píxeles. Este enfoque facilita el procesamiento y la manipulación de las imágenes en el contexto del desarrollo y la implementación de modelos de inteligencia artificial, lo que a su vez mejora la eficacia y la velocidad del entrenamiento del modelo.

B. Segmentación de Elementos de Interés

- Se desarrolló una herramienta llamada SelSegToYolo [3] basada en el modelo X de FastSAM [4] y el modelo L de SAM para generar máscaras de segmentación.

C. Definición de Puntos de Interés

- Las máscaras de segmentación generadas por SAM se utilizaron para definir los puntos de interés donde el modelo YOLO, en especifico YOLOv8m se concentra en detectar y clasificar objetos.
- Estos puntos se guardaron en archivos JSON en formato LabelMe, cada uno con su etiqueta de categoría correspondiente.

D. Conversión de Archivos JSON a YOLO

- Se empleó la herramienta "LabelMe2YOLO" [5] para transformar los archivos JSON en formato LabelMe a archivos TXT en formato YOLO. Esta conversión permitió adaptar los datos para su posterior uso con el modelo YOLO.

E. Separación de datos de entrenamiento y validación

- La herramienta "LabelMe2YOLO" también se encargó de dividir las imágenes y etiquetas en conjuntos de datos de entrenamiento y validación que fueron previamente acordados por nosotros, con un 80% para entrenamiento y un 20% para validación. Esta división es esencial para evaluar y mejorar el rendimiento del modelo.

IV. Resultados

Para evaluar el rendimiento del modelo YOLO en nuestro conjunto de datos, llevamos a cabo un proceso de entrenamiento y validación que demandó aproximadamente 30 minutos. En este proceso, empleamos dos métricas fundamentales: precisión (precision) y exhaustividad (recall), las cuales se centran en la evaluación de la calidad de las clasificaciones realizadas por el modelo. Cabe destacar que utilizamos un conjunto de datos reducido en comparación con el conjunto de datos original, compuesto por 1200 imágenes, con 240 imágenes por categoría. Este proceso se llevó a cabo con la GPU gratuita que ofrece el servicio de Google Colaboratory [6], lo que permitió un rendimiento eficiente y efectivo en el entrenamiento y validación del modelo.

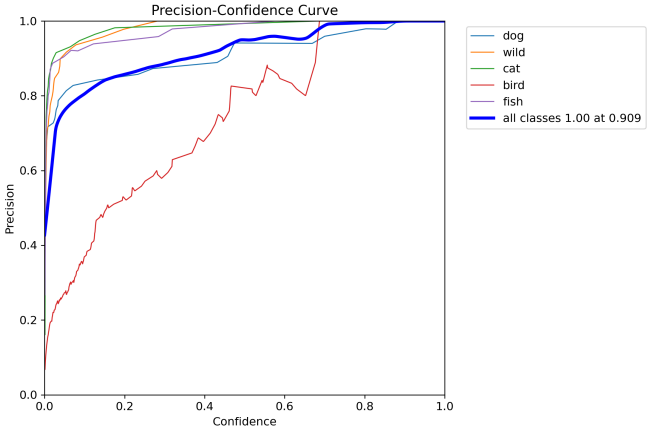

Fig. 3. Gráfica que representa la relación entre la confianza y la métrica de precisión, (precision), en clasificación de imágenes

En cuanto a la métrica de precisión como se observa en la figura 3, la cual se enfoca en la proporción de ejemplos que el modelo clasifica como positivos y cuántos de ellos son verdaderamente positivos., observamos que, al analizar la relación entre la confianza del modelo y la precisión, se obtienen resultados destacados. Con un nivel de confianza cercano al 70%, nuestras curvas de precisión indican que el modelo logra una clasificación efectiva, promediando alrededor del 85% de precisión para cada categoría de imágenes. Este hallazgo sugiere que el modelo YOLO demuestra una capacidad sólida para identificar correctamente objetos en nuestras imágenes de interés.

En relación con la métrica de exhaustividad, presentada en la figura 4 que se centra en la proporción de ejemplos verdaderamente positivos que el modelo ha identificado correctamente como positivos, se observa una distinción sutil en las curvas correspondientes a las categorías 'dog', 'cat', 'wild' y 'fish'. Estas curvas mantienen una consistencia notable a medida que varía la confianza del modelo. Este análisis permite concluir que, con un nivel de confianza del 80%, el modelo logra una clasificación precisa de los casos verdaderamente positivos en estas categorías. Este resultado sugiere un alto nivel de rendimiento y confiabilidad del modelo en la identificación de ejemplos positivos en el conjunto de datos en cuestión, aun así, la exhaustividad de la categoría 'birds'

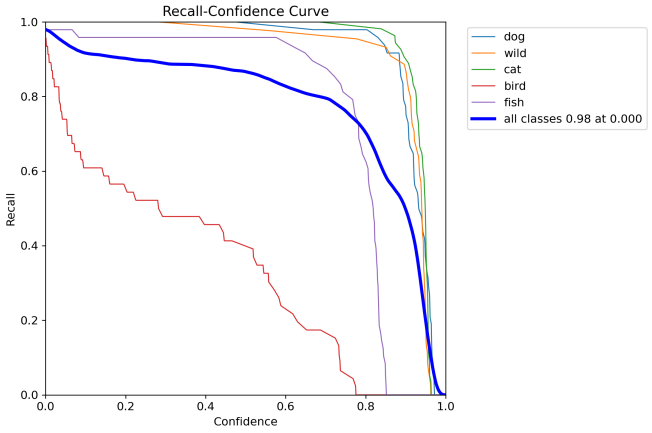

Fig. 4. Gráfica que representa la relación entre la confianza y la métrica de exhaustividad, (recall), en clasificación de imágenes

permanece significativamente por debajo de las demás, similar al caso en el análisis de la precisión.

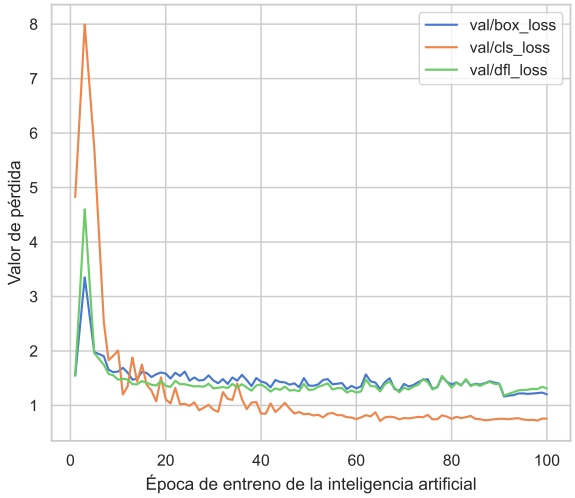

Fig. 5. Pérdidas en el conjunto de validación de las imágenes después de entrenado el modelo

En el análisis de las pérdidas del modelo de acuerdo a la figura 5, se tienen las siguientes definiciones:

- val/box_loss (Pérdida de Caja): se refiere a la pérdida asociada con la precisión de la detección de las ubicaciones de las cajas delimitadoras (bounding boxes) alrededor de los objetos en las imágenes.
- val/cls_loss (Pérdida de Clasificación): esta métrica mide cuán precisamente el modelo puede predecir las clases de los objetos detectados en las imágenes.
- val/dfl_loss (Pérdida de Desplazamiento de Características): se refiere a la pérdida asociada con el desplazamiento de características dentro de las cajas delimitadoras.

Es evidente que las tres métricas de pérdida en el conjunto de validación disminuyen en la mayoría de las épocas posteriores a la época 20. Esta tendencia decreciente indica un

mejor rendimiento del modelo en términos de precisión en la delimitación de las cajas, clasificación de objetos y asignación de categorias en las imágenes, ver figura 6.

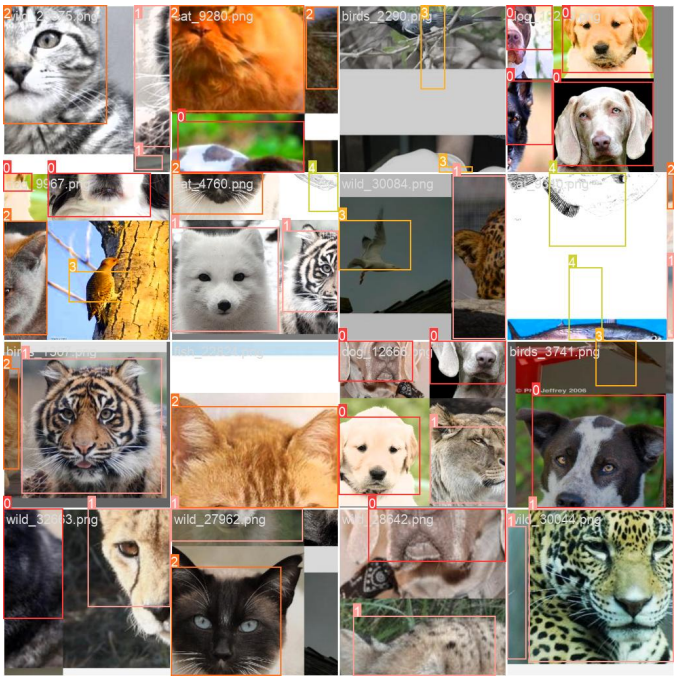

Fig. 6. Representación de cajas de detección

V. Discusión

Las métricas de evaluación utilizadas revelan una tendencia en el modelo a lograr una alta precisión, incluso cuando se entrena con un conjunto de datos de entrenamiento relativamente pequeño, como en este caso, donde se emplearon solamente 1200 imágenes de las 33,000 disponibles en el conjunto de datos original.

No obstante, se observa una tendencia no satisfactoria en la categoría 'birds', que se presenta como un punto crítico del modelo. Las métricas de evaluación muestran un bajo desempeño en esta categoría durante la validación. Esto podría deberse a la notable variabilidad cromática y a las marcadas diferencias morfológicas que existen entre las distintas aves. La diversidad de colores y las variaciones anatómicas en los cuerpos y picos de las aves no siguen patrones predecibles, lo que dificulta la identificación de tendencias distintivas. Además, la presencia de camuflaje natural, común en la naturaleza aviar, añade una capa adicional de complejidad a la tarea de clasificación. Como resultado, la clasificación de aves se ve influenciada por la interacción compleja de estos factores, lo que dificulta la extracción de rasgos significativos o patrones claros en los datos.

Uno de los factores que contribuyeron a la obtención de resultados sobresalientes en la mayoría de las categorías de imágenes se relaciona con el sesgo inherente del dataset original, que se caracteriza por mostrar exclusivamente las caras de los animales. Este sesgo en la recolección de datos implica que las imágenes disponibles se centran principalmente en las regiones faciales de los animales, lo que facilita la tarea de detección y clasificación. La prominencia de las caras en las imágenes proporciona características distintivas y de alto contraste que el modelo puede aprovechar para realizar predicciones precisas. Sin embargo, es importante tener en cuenta que esta focalización en las caras puede limitar la capacidad del modelo para identificar y clasificar adecuadamente otras partes del cuerpo o características menos visibles de los animales, específicamente hablando de las categorías 'cat', 'dog' y 'wild', lo que podría explicar en parte el desafío adicional en la categoría 'birds', donde las diferencias morfológicas y la variabilidad cromática son particularmente notables en áreas no faciales.

VI. Conclusiones

En el curso de esta investigación, se han obtenido conclusiones fundamentales que derivan de la meticulosa metodología aplicada y las discusiones previas. A continuación, se presentan las conclusiones cruciales que se extraen de este informe:

Primeramente, la estandarización de los nombres de archivo y la transformación de imágenes en formato uniforme y dimensiones cuadradas se revelaron como pasos críticos en la gestión eficiente de datos y el aumento de la eficacia del flujo de trabajo del proyecto. Esta estandarización no solo simplificó la administración de datos, sino que también solucionó problemas de duplicación de nombres y garantizó la integridad de los datos, esto unido a la herramienta SelSegToYolo, basada en modelos de segmentación, permitió identificar y segmentar elementos de interés en las imágenes de manera efectiva. Estas máscaras de segmentación desempeñaron un papel esencial en la definición de puntos de interés y en la conversión de archivos JSON a formato YOLO para el entrenamiento del modelo.

Siguiendo con la idea, la definición precisa de puntos de interés mediante las máscaras de segmentación y la posterior conversión de archivos JSON a formato YOLO fueron etapas fundamentales para la preparación de los datos de entrenamiento del modelo YOLOv8m, el cual exhibió un rendimiento sólido en términos de precisión y exhaustividad en la mayoría de las categorías de imágenes. A pesar de haber sido entrenado con un conjunto de datos relativamente pequeño, logró una alta precisión en la detección y clasificación de objetos en las imágenes. Finalmente, la categoría 'birds' presentó un desafío significativo para el modelo, mostrando un rendimiento considerablemente inferior en comparación con otras categorías. Este resultado se atribuyó a la variabilidad cromática, las diferencias morfológicas y la presencia de camuflaje natural en las aves, lo que dificulta la identificación de patrones claros.

En resumen, este estudio proporciona una metodología sólida para la preparación de datos y la implementación de un modelo YOLO en un conjunto de datos de imágenes de animales. A pesar de los desafíos en la categoría 'birds', los resultados generales son prometedores y destacan el potencial de esta aproximación en la detección y clasificación de objetos en imágenes. Para futuras investigaciones, se podría considerar la expansión del conjunto de datos, la exploración de enfoques

de aumento de datos distintos a los que automáticamente genera YOLO durante el entrenamiento, además de explorar el desempeño YOLO en la detección específica de aves.

AGRADECIMIENTOS

Se agradece a todas las personas involucrados en el desarrollo de los modelos preentrenados utilizados en este proyecto, a las personas involucradas en la recolección de los datos, a los profesionales de las diferentes áreas que colaboran en el continuo desarrollo y mejora de la IA.

Adicionalmente, un especial agradecimiento a los divulgadores de los canales de YouTube DotCSV, Aprende e Ingenia y Ringa Tech.

REFERENCES

[1] MetaAI. (2023) Segment anything. Fecha de acceso: 07 de septiembre de 2023. [Online]. Available: https://segment-anything.com/

[2] glenn jocher. (2023) ultralytics. Fecha de acceso: 07 de septiembre de 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[3] a. estivengrx and camiloarios. (2023) Selsegtoyolo. Fecha de acceso: 22 de septiembre de 2023. [Online]. Available: https://github.com/andres-mestra/SelSegToYolo

[4] CASIA-IVA-Lab. (2023) Fastsam. Fecha de acceso: 14 de septiembre de 2023. [Online]. Available: https://segment-anything.com/

[5] rooneysh. (2023) Labelme2yolo. Fecha de acceso: 21 de septiembre de 2023. [Online]. Available: https://github.com/rooneysh/Labelme2YOLO/tree/main

[6] Google. (2023) Colaboratory. Fecha de acceso: 07 de septiembre de 2023. [Online]. Available: https://colab.research.google.com/