

Herramienta interactiva para segmentación de imágenes con SAM2

A. Yubero^{1*}, C. Borja¹, M. Mulero-Pázmány², E. Montijano¹, A. C. Murillo¹

¹ Grupo de Robótica, Visión por Computador e Inteligencia Artificial (Ropert)
Instituto de Investigación en Ingeniería de Aragón (I3A)

Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.

*e-mail: 842236@unizar.es

²Afiliación: Dpto. Biología Animal, Universidad de Málaga. Spain

Resumen

Este trabajo presenta una herramienta para la segmentación interactiva de información semántica relevante en imágenes de entornos naturales basado en modelos fundacionales. La nueva herramienta facilita el uso del modelo SAM2, por parte de usuarios no expertos en el manejo de modelos fundacionales de visión. A modo de ejemplo, se muestra el uso de la herramienta en la segmentación de caminos en ortofotos aéreas.

Introducción

El etiquetado manual de imágenes de alta resolución constituye un cuello de botella habitual en los flujos de trabajo de visión por computador. Los grandes avances en el campo de la segmentación automática basada en *deep learning* han motivado el desarrollo de algoritmos que permiten segmentar objetos de manera genérica, a partir de un punto dado por el usuario, entre los que destaca la familia de modelos Segment Anything Model [1]. Encontramos trabajos que aprovechan estos modelos como ayuda a tareas de segmentación complejas. Por ejemplo, SSeg [2] propone aumentar anotaciones muy dispersas en imágenes complejas para obtener una anotación densa de la imagen gracias a dichos modelos. SAM también se ha utilizado por ejemplo como ayuda en el etiquetado de residuos en plantas de reciclaje [3].

Un problema habitual es la dificultad de investigadores de otras disciplinas para utilizar estos modelos. Aunque herramientas como SAM son muy avanzadas, su aplicación real requiere conocimientos técnicos en visión por computador e informática (gestión de dependencias, formatos de datos y parámetros). Esto crea una barrera para usuarios que no buscan desarrollar IA, sino aplicarla como herramienta en sus áreas de estudio.

Para mitigar esta limitación, presentamos una herramienta que simplifica el uso de SAM2, permitiendo a usuarios sin experiencia en visión por computador segmentar imágenes de forma intuitiva.

El usuario solo debe cargar la imagen y hacer *clicks* con el ratón para marcar las regiones de interés, cuyos resultados se pueden exportar en formatos estándar. Como ejemplo de uso, el trabajo muestra la segmentación de caminos en fotografías aéreas del PNOA (Plan Nacional de Ortofotografía Aérea).

Metodología

Segment Anything Model 2 (SAM2) es un modelo de inteligencia artificial diseñado para segmentar objetos en imágenes y vídeos de forma general, es decir, sin necesidad de haber sido entrenado específicamente para cada categoría de objeto. Su objetivo es poder separar automáticamente cualquier elemento visible de una escena, tales como personas, coches, o vegetación, a partir de indicaciones simples del usuario, como un píxel en la región que se quiere extraer. SAM2 utiliza una arquitectura basada en transformers que analiza la imagen para extraer características relevantes de la imagen, o *embeddings*, y, a partir de ellas, generar las máscaras correspondientes de segmentación.

Herramienta de interacción con SAM2.

El **funcionamiento** de la herramienta propuesta se resume en la Fig. 1, y consta de dos etapas.

En la primera, cuando el usuario sitúa el cursor sobre la imagen, la herramienta extrae un parche de 1024x1024 píxeles centrado en dicha posición y lo codifica mediante ese el *encoder* del modelo para obtener un *embedding*. El tamaño del parche se ha elegido para que coincida con la dimensión de entrada del *encoder* de SAM2. El *embedding* se almacena en caché, lo que permite que este paso solo se repita cuando el cursor abandona la región ya codificada, evitando cálculos innecesarios.

En la segunda parte, el *decoder* recibe el *embedding* en caché y las coordenadas del cursor en el espacio del parche, y genera tres máscara binarias y otorga una puntuación a cada una de ellas, como se puede

ver en la Fig. 2. La aplicación permite obtener la más adecuada a la aplicación final en lugar de la que SAM2 selecciona por defecto.

Todas las máscaras generadas pueden exportarse en formato GeoTIFF como mapas binarios, asociados a un identificador de su clase correspondiente. Asimismo, con el objetivo de facilitar la anotación de elementos lineales (como caminos), se implementó una opción de exportación en formato tipo línea basada en el algoritmo de adelgazamiento de Zhang-Suen [4]. Mediante este procedimiento, las máscaras se reducen al grosor de un solo píxel para, posteriormente, trazar segmentos rectos entre ellos y vectorizar la información a exportar.

Resultados

Como evaluaciones de esta versión de la herramienta, la Fig. 2 muestra un ejemplo de las tres máscaras devueltas por SAM2, de las cuáles la primera sería marcada como correcta por defecto, al tener la puntuación más alta. En todas ellas se observa una agrupación de píxeles basada en la similitud: por semántica en la máscara 1, que delimita con precisión el camino, o por apariencia en la máscara 3, donde el camino y el campo se confunden por su color parecido.

Conclusiones

Se ha presentado una herramienta de anotación interactiva que emplea SAM2 y facilita de manera

satisfactoria la tarea de anotación al usuario, llegando a encontrar máscaras que se ajustan a las formas de los objetos presentes en la imagen. Como siguientes pasos se plantea incluir una mayor automatización que permite sugerir máscaras antes de que el usuario mueva siquiera el cursor, así como una evaluación más exhaustiva y cuantitativa de la calidad de la segmentación, en colaboración con expertos en temas de conservación de entornos naturales.

REFERENCIAS

- [1]. RAVI, N., GABEUR, V., HU, Y.-T., HU, R., CARION, N., PANG, T., HAZIZA, B., SINGH, G., GIRSHICK, R., DOLLÁR, P. y FEICHTENHOFER, C. SAM 2: Segment Anything in Images and Videos. Int. Conf. on Learning Representations (ICLR). 2025.
- [2]. BORJA, C., PLOU, C., MARTINEZ-CANTIN, R. y MURILLO, A.C. SSeg: Active Sparse Point-Label Augmentation for Semantic Segmentation. Proc. of IEEE/CVF Wint. Conf. on Applications of Computer Vision. 2026, pp. 1424-1434.
- [3]. CASAO, S., PEÑA, F., SABATER, A., CASTILLÓN, R., SUÁREZ, D., MONTIJANO, E., Y MURILLO, A. C. Spectralwaste dataset: Multimodal data for waste sorting automation. In 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 5852-5858..
- [4]. ZHANG, T.Y. y SUEN, C.Y. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*. 1984, vol. 27, núm. 3, pp. 23

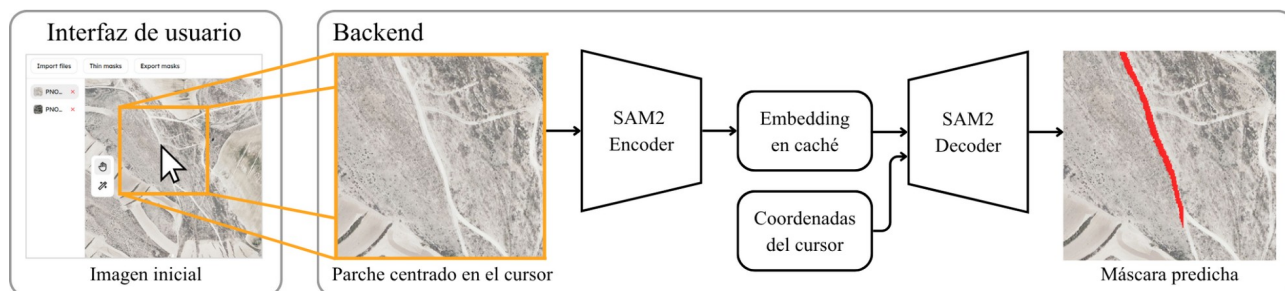


Figura 1. Diagrama de funcionamiento del sistema.

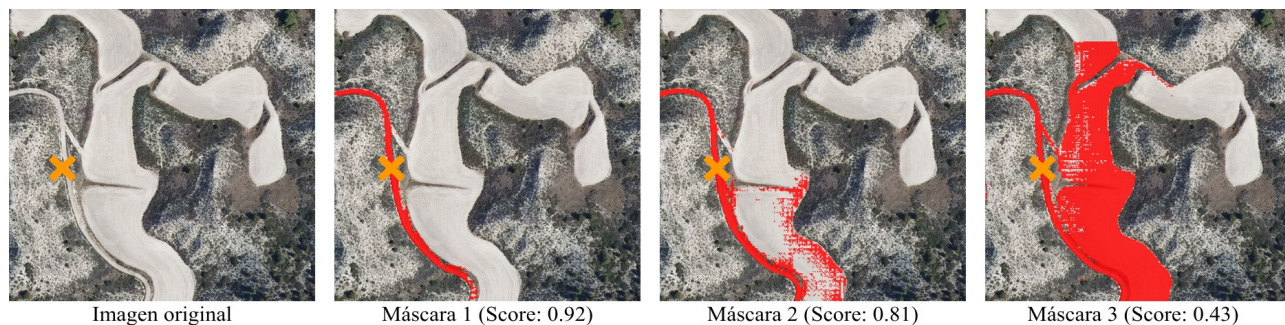


Figura 2. Las tres máscaras y sus respectivas puntuaciones devueltas por SAM2, al introducir en su *decoder* el embedding en caché y las coordenadas del punto marcado en amarillo.