

Desarrollo de un framework para la segmentación auto-supervisada de series temporales de imágenes mediante flujo óptico

Fernando Frassia¹, Enzo Ferrante¹, Nicolás Gaggion¹

¹ FCEyN, Universidad de Buenos Aires

Resumen En el campo de la biología vegetal, comprender la dinámica de crecimiento radicular es crucial para diversas aplicaciones, que abarcan desde la mejora genética de cultivos hasta el estudio de respuestas fisiológicas ante diferentes condiciones ambientales. Para analizar dicha dinámica mediante series temporales de imágenes o videos, es indispensable segmentar las raíces para poder finalmente fenotiparlas, extrayendo así parámetros como la longitud, el ángulo de crecimiento, entre otros. Actualmente, los métodos de segmentación más utilizados se basan en redes neuronales convolucionales (CNNs), las cuales requieren un gran volumen de datos etiquetados manualmente para su entrenamiento. El presente trabajo busca abordar este desafío mediante el desarrollo de un método auto-supervisado para la generación de etiquetas, aprovechando la redundancia temporal inherente en videos de raíces en crecimiento. Este enfoque permite reducir significativamente la necesidad de anotaciones manuales, haciendo el proceso de segmentación más eficiente.

Palabras clave: Flujo óptico - segmentación radicular - fenotipado

Development of a Framework for Self-Supervised Segmentation of Image Time Series Using Optical Flow

Abstract In the field of plant biology, understanding root growth dynamics is crucial for various applications, ranging from crop genetic improvement to the study of physiological responses under different environmental conditions. To analyze this dynamic behavior through time-lapse image sequences or videos, it is essential to segment the roots in order to ultimately phenotype them, thereby extracting parameters such as length, growth angle, among others. Currently, the most commonly used segmentation methods are based on convolutional neural networks (CNNs), which require a large amount of manually labeled data for training. This work aims to address this challenge through the development of a self-supervised method for label generation, leveraging the temporal redundancy inherent in videos of growing roots. This approach significantly reduces the need for manual annotations, making the segmentation process more efficient.

Keywords: Optic flow - root segmentation - phenotyping

1. Introducción

El sistema radicular de las plantas es fundamental para comprender su desarrollo, función y adaptación a diferentes condiciones ambientales. Las raíces desempeñan roles cruciales en el anclaje, absorción de nutrientes, interacción con microorganismos del suelo y respuesta a estrés ambiental, convirtiéndolas en elementos determinantes para el rendimiento de los cultivos y la sostenibilidad agrícola. Sin embargo, la naturaleza subterránea de las raíces presenta desafíos significativos para su estudio, limitando nuestra capacidad para caracterizar de manera eficiente sus fenotipos.

Los avances en técnicas de imagen y aprendizaje automático han revolucionado el fenotipado vegetal, permitiendo extraer automáticamente parámetros arquitectónicos radiculares. El sistema ChronoRoot (Gaggion et al., 2021) representa un avance significativo al combinar hardware de bajo costo con modelos de aprendizaje profundo para el fenotipado de alta resolución temporal, revelando parámetros y patrones de desarrollo antes inaccesibles. No obstante, los métodos actuales de segmentación basados en redes neuronales convolucionales (CNNs) (Ronneberger et al., 2015), como U-Net, dependen de grandes volúmenes de anotaciones manuales, un proceso laborioso que constituye un importante cuello de botella. Además, los conjuntos de datos disponibles se concentran principalmente en *Arabidopsis thaliana*, y los modelos entrenados con estas anotaciones muestran rendimiento reducido cuando se aplican a otras especies debido a variaciones morfológicas, limitando su transferibilidad.

El objetivo principal de este trabajo es desarrollar un método auto-supervisado para segmentación de raíces en series temporales que: (1) reduzca significativamente la necesidad de anotaciones manuales, (2) aproveche la redundancia temporal en videos de crecimiento radicular, y (3) genere pseudo-etiquetas de calidad suficiente para pre-entrenar modelos supervisados adaptables a múltiples especies vegetales. Nuestro enfoque se basa en las propiedades del flujo óptico para detectar movimiento en secuencias de imágenes, distinguiendo entre el crecimiento dinámico de las raíces y el fondo estático.

2. Métodos

2.1. Flujo Óptico

El flujo óptico es una técnica que permite detectar y cuantificar el movimiento entre imágenes consecutivas. En términos simples, estima cómo cada pixel se desplaza de una imagen a la siguiente, generando vectores de movimiento (u, v) para cada posición. Esta técnica resulta particularmente adecuada para nuestro problema de segmentación de raíces, ya que durante el crecimiento, las raíces se desplazan gradualmente mientras el fondo permanece estático. La estimación del flujo óptico se basa en dos principios intuitivos: la intensidad de un objeto permanece similar aunque se mueva (conservación de brillo), y los movimientos entre frames consecutivos son pequeños. Expresado matemáticamente, esto con-

duce a la ecuación de restricción del flujo óptico: $I_x u + I_y v + I_t = 0$, donde I_x , I_y e I_t representan cómo cambia la intensidad de la imagen en las direcciones horizontal, vertical y temporal.

2.1.1. Método de Farneback

Para nuestro trabajo elegimos específicamente el algoritmo de Farneback (Farneback, 2003), que ofrece dos ventajas para la segmentación de raíces: proporciona un campo de flujo denso (calcula el movimiento para todos los píxeles, no solo para puntos característicos) y es robusto ante cambios sutiles de iluminación. El método funciona aproximando vecindarios de píxeles mediante polinomios y estimando cómo estos vecindarios se transforman entre frames consecutivos. Esta densidad del flujo resulta fundamental para nuestra aplicación, ya que necesitamos detectar el movimiento en toda la imagen.

2.2. Descripción del framework

Nuestro pipeline consta de dos etapas principales: preprocesamiento y generación de segmentaciones mediante flujo óptico.

2.2.1. Preprocesamiento

Dado que las imágenes originales contienen múltiples raíces de *Arabidopsis thaliana*, implementamos un procedimiento para aislar raíces individuales:

1. El usuario dibuja bounding boxes alrededor de cada raíz de interés en la imagen original.
2. Para cada bounding box, se selecciona un punto semilla en la parte superior de la raíz como referencia.
3. Se extraen sub-imágenes correspondientes a cada bounding box, generando secuencias individuales.

Esta intervención manual inicial es significativamente menos costosa que el etiquetado completo de cada imagen y solo se requiere una vez por secuencia, como se puede observar en la Figura 1.

2.2.2. Procesamiento mediante flujo óptico

Una vez obtenidas las secuencias individuales, aplicamos un proceso iterativo que aprovecha la naturaleza acumulativa del crecimiento radicular, aplicando el pipeline presente en la Figura 2. Primero, calculamos el flujo óptico entre imágenes consecutivas mediante el método de Farneback. La clave de nuestro enfoque está en la acumulación temporal: sumamos los vectores de flujo a lo largo de toda la secuencia, lo que permite resaltar el movimiento consistente de las raíces y atenuar el ruido aleatorio. Esta acumulación se realiza sumando los vectores completos (u, v) o sus magnitudes $\sqrt{u^2 + v^2}$.

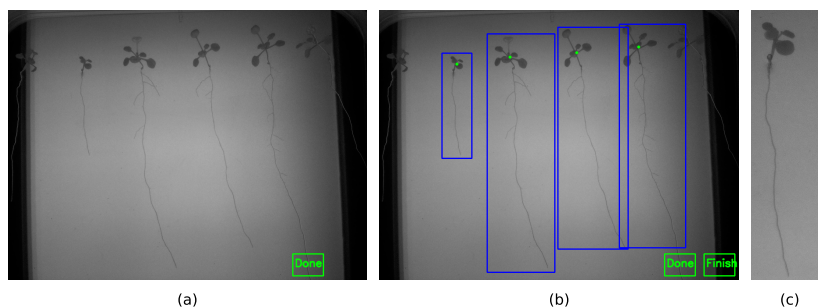


Figura 1: Preprocesamiento. (a) imagen original, (b) imagen con bounding boxes y punto semilla, (c) imagen de raíz seleccionada.

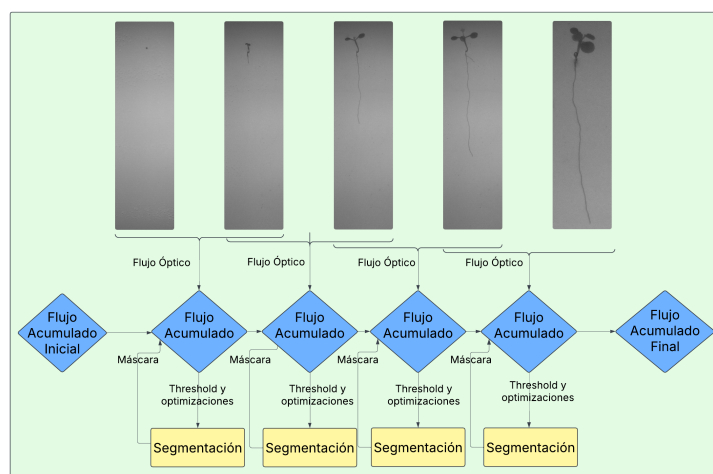


Figura 2: Pipeline de procesamiento. Acumulación de flujo óptico, generación de segmentaciones y máscaras.

Tras aplicar un umbral a esta matriz acumulada, obtenemos una segmentación inicial que refinamos mediante el uso de operaciones morfológicas y esqueletonización (Zhang y Suen, 1984). Un paso crucial es utilizar el punto semilla seleccionado inicialmente para identificar la componente conectada correspondiente a la raíz principal, eliminando falsos positivos. Finalmente, utilizamos la segmentación obtenida como máscara para filtrar el flujo en iteraciones posteriores, creando un proceso que se retroalimenta y mejora progresivamente.

2.3. Dataset

Para la validación de nuestro modelo, utilizamos dos conjuntos de datos complementarios adquiridos mediante el sistema ChronoRoot (Gaggion et al., 2021), el cual captura imágenes a intervalos regulares (cada 15 minutos) mediante cámaras fijas e iluminación infrarroja.

El primero (Dataset1) consta de 19 secuencias temporales que suman más de 12000 imágenes de raíces de plantas, donde cada secuencia incluye entre 5 y 6 imágenes con anotaciones manuales realizadas por expertos biólogos, garantizando así un estándar de calidad apropiado para la evaluación.

El segundo conjunto (Dataset2) comprende 24 secuencias temporales con más de 30000 imágenes en total, de mucha mayor calidad y mucho menor ruido en comparación con el Dataset1, debido a mejoras en el protocolo de captura. Para este conjunto, disponemos de segmentaciones generadas mediante un modelo nnUNet previamente entrenado con una base de datos de anotaciones manuales, lo que nos permite tener un proxy de evaluación de nuestro método comparándolo con uno del estado del arte.

3. Resultados

Evaluamos las segmentaciones mediante dos métricas estándar en el campo de segmentación de imágenes: el coeficiente Dice, que cuantifica la superposición espacial entre la predicción y el ground truth, y el Recall, que determina la proporción de la estructura radicular correctamente identificada. En la Figura 3 podemos observar que en ambos conjuntos de datos se obtuvo un Recall promedio cercano a 0.83 (línea punteada vertical), demostrando una consistente capacidad de detección. La media del coeficiente Dice (línea punteada vertical) presentó variaciones significativas: 0.21 en el primer dataset y 0.35 en el segundo, diferencia atribuible a las mejoras implementadas en el protocolo de captura y la consecuente reducción de ruido en las imágenes.

Un análisis más detallado revela que, en el primer conjunto, predominan secuencias con valores muy bajos de Dice, generando un pico en los valores cercanos a 0.1, mientras que en el segundo, la mayoría se concentra en las zonas más altas de Dice, alrededor de 0.45. Esta diferencia sugiere que la calidad del hardware y las condiciones de captura influyen directamente en la calidad de la segmentación generada mediante nuestro método. Asimismo, se observó una mayor proporción de secuencias con altos valores de Recall (entre 0.8 y 1.0) en Dataset 2 (75 %) en comparación con Dataset 1 (59 %). Esto indica que nuestro enfoque logra recuperar una fracción considerablemente mayor de la raíz en contextos con imágenes de mejor calidad.

Estos resultados preliminares demuestran que el método propuesto, si bien presenta sensibilidad a condiciones de captura y niveles de ruido, logra un rendimiento que nos presenta buenas perspectivas a futuro.

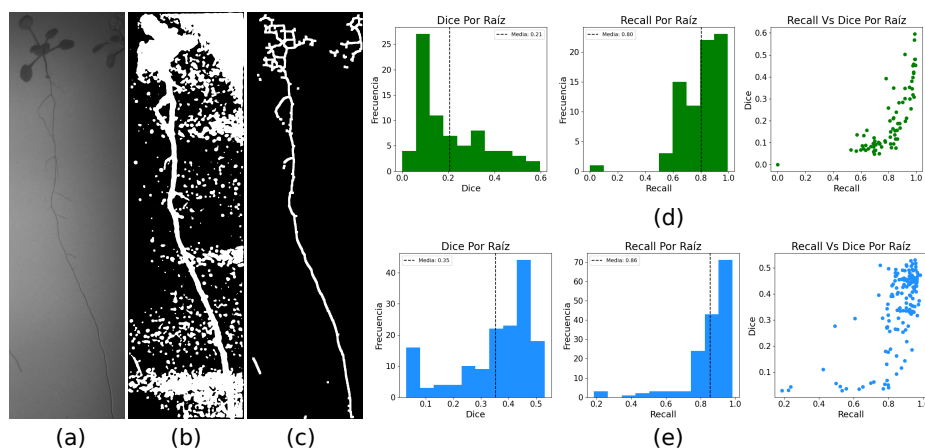


Figura 3: Resultados. (a) Imagen original, (b) Segmentación obtenida mediante flujo óptico, (c) Segmentación optimizada mediante esqueletización, (d) Resultados cuantitativos de Dice y Recall en Dataset 1, (e) Resultados cuantitativos de Dice y Recall en Dataset 2.

4. Conclusiones y Trabajo Futuro

Presentamos un método auto-supervisado de segmentación de raíces basado en flujo óptico. Este enfoque aminora la necesidad de anotaciones extensas y muestra resultados prometedores. Un aspecto destacable es que el enfoque es completamente reutilizable para otras especies o escenarios donde el movimiento diferencial entre la estructura de interés y el fondo sea significativo. Esto abre la puerta a aplicaciones más allá del fenotipado vegetal, como el seguimiento de estructuras móviles en estudios biomédicos. Como parte de los trabajos futuros, planeamos evaluar el método en raíces de plantas de tomate, con el objetivo de validar su capacidad de generalización a especies con morfologías más complejas. Por otro lado, planeamos también usar las etiquetas generadas para preentrenar un modelo nnUNet y evaluar su desempeño cuando se tienen pocos datos etiquetados. Finalmente exploraremos comparaciones con frameworks recientes como SAM2 (Ravi et al., 2024), que podría complementar nuestro enfoque para mejorar la robustez ante variaciones en las condiciones de captura.

Referencias

Farneback, G. (2003). Two-Frame Motion Estimation Based on Polynomial Expansion. In: Bigun, J., Gustavsson, T. (eds) Image Analysis. SCIA 2003. Lecture Notes in Computer Science, vol 2749. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45103-X_50

Gaggion, N., Ariel, F., Daric, V., Lambert, É., Legendre, S., Roulé, T., Camoirano, A., Milone, D., Crespi, M., y Ferrante, E. (2021). ChronoRoot: High-throughput phenotyping by deep segmentation networks reveals novel temporal parameters of plant root system architecture. *GigaScience*, 10(7), giab052. <https://doi.org/10.1093/gigascience/giab052>

Ravi, N., Gabeur, V., Hu, Y., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C., Girshick, R., Dollár, P., y Feichtenhofer, C. (2024). SAM 2: Segment Anything in Images and Videos. arXiv. <https://arxiv.org/abs/2408.00714>

Ronneberger, O., Fischer, P., y Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Cham: Springer international publishing. <https://arxiv.org/abs/1505.04597>

Zhang, T. Y., y Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Commun of the ACM*, 27(3), 236-239.