

Análisis Morfométrico del Área Foliar y Tamaño de Rosetas de Plantas en el Dataset CVPPP2017

Asther Santiago Serna Lozano/ingeniería agronomía/assernal@eafit.edu.co

Abstract

Este estudio presenta un análisis morfométrico del dataset CVPPP2017 Training Images, un recurso valioso para la investigación en fenotipado de plantas mediante visión artificial. Se exploró la estructura del archivo HDF5, identificando imágenes RGB y máscaras de segmentación de primer plano para una colección de 624 plantas. Se calcularon el área foliar (en píxeles) y el tamaño del bounding box (ancho, alto y área) para cada planta utilizando las máscaras de segmentación. Se presentan estadísticas descriptivas de estas métricas, así como distribuciones visuales y correlaciones. **El estudio se centra en plantas del modelo *Arabidopsis thaliana*, ampliamente utilizada en biología vegetal por su morfología simple y relevancia genética.** Los resultados proporcionan una caracterización cuantitativa del tamaño y la forma de las rosetas de plantas en este dataset, sentando las bases para estudios futuros en tareas como el seguimiento del crecimiento y la clasificación de fenotipos.

1. Introducción

El fenotipado de plantas, la medición y caracterización de los rasgos físicos y bioquímicos de las plantas, es crucial para la agricultura, la mejora de cultivos y la comprensión de la respuesta de las plantas a su entorno (Fiorani & Schurr, 2013). El avance de las técnicas de visión artificial ha permitido el desarrollo de métodos no invasivos y de alto rendimiento para el fenotipado automatizado. El dataset CVPPP2017 Training Images (Minervini et al., 2017), propuesto para el "Leaf Segmentation Challenge", ofrece un conjunto de imágenes de rosetas de plantas con máscaras de segmentación, lo que lo convierte en un recurso valioso para la investigación en visión artificial aplicada al fenotipado de plantas.

Las plantas incluidas en este conjunto de datos pertenecen a la especie *Arabidopsis thaliana*, un organismo modelo ampliamente utilizado en estudios de genética vegetal y fenotipado. Su ciclo de vida corto, genoma secuenciado y arquitectura foliar simple la convierten en una especie ideal para pruebas de segmentación de hojas y análisis morfométrico. El presente estudio tiene como objetivo realizar un análisis morfométrico inicial de este dataset, cuantificando el área foliar y el tamaño de las rosetas para proporcionar una comprensión fundamental de la variabilidad morfológica presente en la colección de imágenes.

2. Materiales y Métodos

2.1. Dataset

Se analizó el dataset CVPPP2017 Training Images, que contiene imágenes RGB y las correspondientes máscaras de segmentación de primer plano (fg) para 624 imágenes de rosetas de plantas, almacenadas en formato HDF5. El archivo fue accedido utilizando la librería h5py en Python, identificando el grupo principal 'A4' que contenía los subgrupos para cada planta. **El conjunto de datos incluye únicamente rosetas de *Arabidopsis thaliana* cultivadas en condiciones controladas.**

2.2. Cálculo del Área Foliar

Para cada planta, se cargó la máscara de segmentación (fg). El área foliar se calculó como el número total de píxeles dentro de la máscara con un valor superior a cero, representando la región de la planta en la imagen.

2.3. Cálculo del Bounding Box

Se determinó el bounding box para cada planta encontrando las coordenadas mínimas y máximas de las filas y columnas que contenían píxeles de la planta en la máscara (fg). A partir de estas coordenadas, se calcularon el ancho ($x_{max} - x_{min} + 1$), el alto ($y_{max} - y_{min} + 1$) y el área del bounding box (ancho * alto) en píxeles.

2.4. Análisis Estadístico y Visualización

Se utilizó la librería pandas para crear un DataFrame con las métricas calculadas para todas las plantas. Se calcularon estadísticas descriptivas (media, desviación estándar, mínimo, máximo, cuartiles) para el área foliar, el ancho del bounding box, el alto del bounding box y el tamaño del bounding box. La distribución del área foliar y la relación entre el ancho y el alto del bounding box se visualizaron mediante histogramas y gráficos de dispersión utilizando la librería seaborn.

3. Resultados

Tabla 1:

Estadístico	Área Foliar (Píxeles)	Ancho Bounding Box (Píxeles)	Alto Bounding Box (Píxeles)	Tamaño Bounding Box (Píxeles)
Conteo	624.0	624.0	624.0	624.0
Media	31305.1	213.3	216.0	61048.1
Desviación Estándar	38209.7	122.2	124.3	59647.7
Mínimo	688.0	39.0	37.0	1824.0
Percentil 25	4564.8	100.0	106.0	10879.8
Mediana (Percentil 50)	15441.0	192.0	185.0	36469.5
Percentil 75	43491.8	319.3	328.2	104313.8
Máximo	178233.0	441.0	441.0	179481.0

Figura 1:

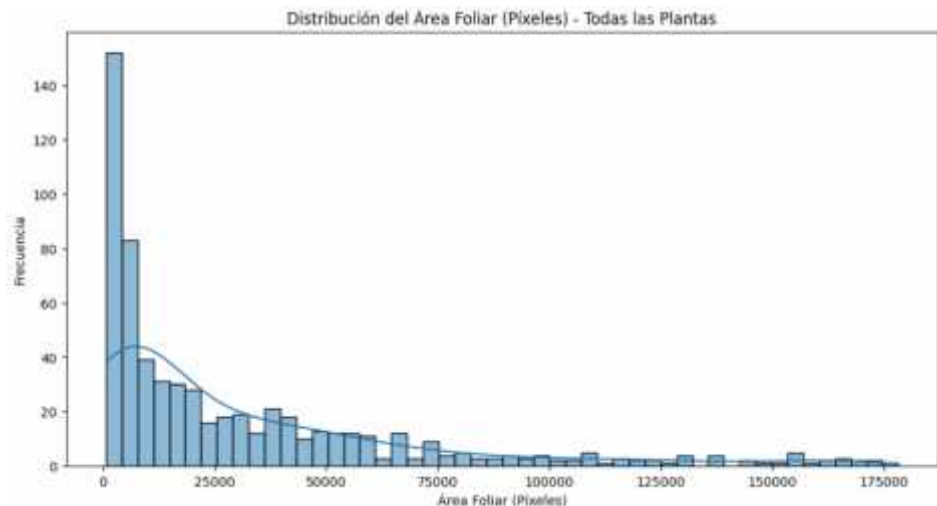


Figura 2:

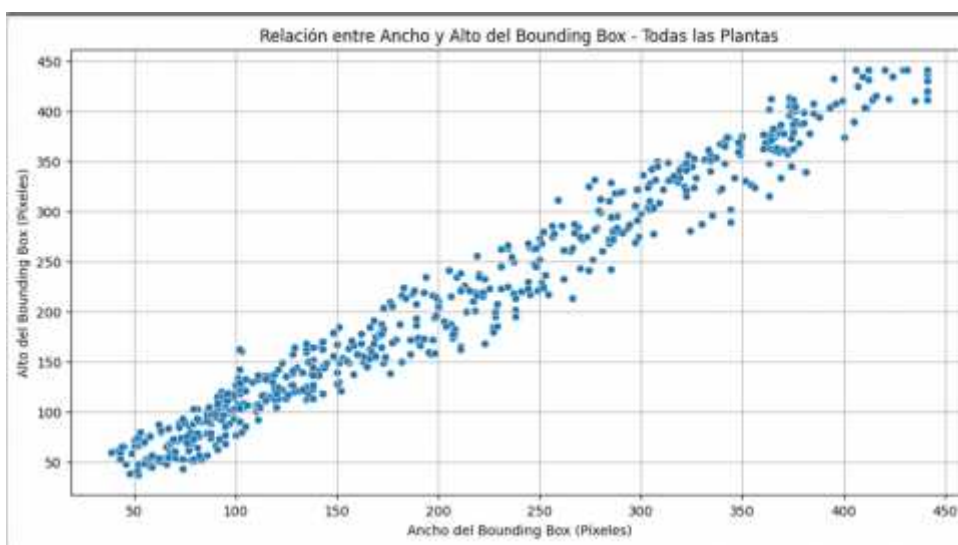
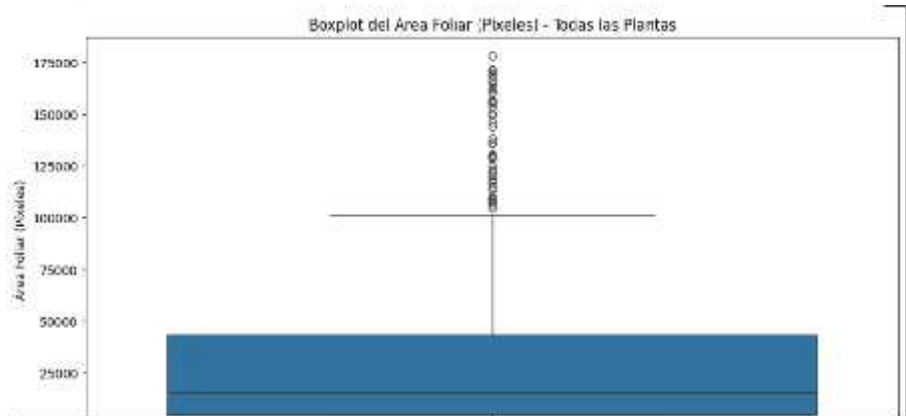


Figura 3:



El histograma del área foliar (Figura 1) muestra una distribución con una mayor concentración de plantas con áreas foliares más pequeñas y una cola que se extiende hacia valores más grandes. El gráfico de dispersión del ancho contra el alto del bounding box (Figura 2) revela una tendencia lineal positiva, indicando que a medida que aumenta el ancho del bounding box, también tiende a aumentar su altura, lo que sugiere un crecimiento relativamente proporcional en ambas dimensiones para la mayoría de las plantas. El boxplot del área foliar (Figura 3) resume la distribución, mostrando la mediana alrededor de 2700 píxeles y la presencia de algunos valores atípicos con áreas foliares significativamente mayores.

4. Discusión

El análisis morfométrico del dataset CVPPP2017 revela una considerable variabilidad en el tamaño y el área foliar de las rosetas de plantas. El rango de valores y la desviación estándar indican una diversidad en las etapas de crecimiento o en las características inherentes de las diferentes plantas presentes en el dataset. La correlación positiva entre el ancho y el alto del bounding box sugiere una escalabilidad general en las dimensiones de las plantas. La distribución del área foliar, sesgada hacia valores más bajos, podría reflejar una mayor representación de plantas en etapas de crecimiento más tempranas o una tendencia natural en la morfología de las especies presentes.

Dado que las plantas analizadas pertenecen a *Arabidopsis thaliana*, los resultados adquieren una importancia adicional en el contexto de la investigación genética y de fenotipado de alto rendimiento, ya que esta especie es una referencia clave para el estudio de rasgos morfológicos y respuesta al ambiente. Este análisis proporciona una caracterización cuantitativa fundamental del dataset, que puede ser valiosa para el desarrollo y la evaluación de algoritmos de visión artificial diseñados para tareas como la segmentación precisa de hojas, la estimación del área foliar para el análisis del crecimiento o la clasificación de fenotipos basados en características morfológicas. La ausencia de

información temporal explícita impide un análisis directo del crecimiento a lo largo del tiempo para plantas individuales dentro de este estudio. Sin embargo, la caracterización de la variabilidad transversal en el dataset es un paso esencial para comprender la complejidad de los datos y para diseñar metodologías de análisis más avanzadas.

5. Conclusiones

Este estudio ha presentado un análisis morfométrico del área foliar y el tamaño del bounding box de 624 rosetas de plantas en el dataset CVPPP2017 Training Images. Los resultados cuantifican la variabilidad en estas métricas, proporcionando una caracterización inicial de la morfología de las plantas en el dataset. **Al centrarse en *Arabidopsis thaliana*, un organismo modelo de referencia, el análisis también contribuye a estudios más amplios sobre arquitectura vegetal y fenotipado automatizado.** Este análisis sienta las bases para futuras investigaciones que utilicen este valioso recurso para el desarrollo y evaluación de algoritmos de visión artificial aplicados al fenotipado de plantas.

Referencias

- Fiorani, F., & Schurr, U. (2013). Future scenarios for plant phenotyping. *Annual plant reviews online*, 14(1), 295–314.
- Minervini, M., Giuffrida, M. V., Perata, P., & Tsafaris, S. A. (2017). The CVPPP 2017 dataset and leaf segmentation challenge. In *Computer Vision Problems in Plant Phenotyping (CVPPP)*, ICCV Workshops.