SCAP
SOCIEDADE DE CIÊNCIAS AGRÁRIAS DE PORTUGAL

DeepFake Weeds: integración de redes neuronales y stable diffusion para la detección de malas hierbas en cultivos de tomate

DeepFake Weeds: Integration of neural networks and stable diffusion for weed detection in tomato crops

Hugo Moreno, Adrià Gómez, Angela Ribeiro & Dionisio Andújar*

Centro de Automática y Robótica, Consejo Superior Investigaciones Científicas (CSIC), Madrid, España
*(*E-mail: d.andujar@csic.es)*
https://doi.org/10.19084/rca.34972

Recibido/received: 2024.01.15
Aceptado/accepted: 2024.02.28

RESUMEN

Las malas hierbas afectan negativamente el rendimiento y calidad de las cosechas al competir con el cultivo por los recursos. Detectarlas a tiempo permite optimizar el control mediante la aplicación precisa de herbicidas y reducir su impacto ambiental. Sin embargo, su detección y clasificación es un desafío debido a la gran diversidad de especies y similitudes con los cultivos. Los métodos tradicionales de aprendizaje profundo (Deep Learning) han permitido el desarrollo de redes neuronales convolucionales (CNN) para la clasificación de especies de mala hierba. Sin embargo, las CNN requieren amplios y variados conjuntos de datos para su entrenamiento. Este estudio propone una metodología innovadora utilizando clasificadores de CNN, YOLOv8l y RetinaNet aumentados con datos de Stable Diffusion (SD), implementado imágenes de malas hierbas generadas artificialmente. Para ello, se configuraron tres conjuntos de datos para el entrenamiento de las CNNs (real, artificial y mixto) en cultivos de tomate comercial infestados por *Solanum nigrum* L., *Portulaca oleracea* L. y *Setaria verticillata* L. Los resultados mostraron un alto grado de acuerdo con imágenes artificiales, alcanzando un *mean Average Precision* (mAP) máximo de 0,93 en ambas redes para el conjunto de datos mixto. SD permitió generar grandes conjuntos de datos de alta calidad a partir de un conjunto limitado de imágenes, reduciendo la necesidad de obtener un gran número de imágenes en campo y su posterior etiquetado manual. Además, el método puede adaptarse a otros cultivos y especies de mala hierba, contribuyendo así al avance de los sistemas automatizados de gestión de malas hierbas.

Palabras clave: agricultura de precisión, control de malas hierbas, aprendizaje profundo, imágenes artificiales.

ABSTRACT

Weeds negatively affect crop yield and quality by competing with the crop for resources. Detecting them in time allows us to optimize control through the precise application of herbicides and reduce their environmental impact. However, its detection and classification are challenging due to the great diversity of species and similarities with crops. Deep learning methods have allowed the development of convolutional neural networks (CNN) for the classification of weed species. However, CNNs require large and varied data sets for training. This study proposes an innovative methodology using CNN, YOLOv8l and RetinaNet classifiers augmented with Stable Diffusion data, implemented artificial images of weeds. Three data sets were configured for training CNNs (real, artificial and mixed) in commercial tomato crops infested by *Solanum nigrum* L., *Portulaca oleracea* L. and *Setaria verticillata* L. The results showed a high degree of agreement with artificial images, reaching a maximum mean Average Precision (mAP) of 0.93 in both networks for the mixed data set. Stable Diffusion made it possible to generate large, high-quality data sets from a limited set of images, reducing the need to obtain a large number of images in the field and their subsequent manual labeling. Furthermore, the method can be adapted to other crops and weed species, thus contributing to the advancement of automated weed management systems.

Keywords: precision agriculture, weed control, deep learning, artificial images.

INTRODUCCIÓN

Las malas hierbas causan aproximadamente el 35% de las pérdidas mundiales de rendimiento de los cultivos, y el uso excesivo de herbicidas químicos ha provocado numerosos efectos adversos, como resistencia, contaminación y daños a organismos que no son el objetivo (Pérez-Ortiz *et al*., 2015). Su detección es un paso crucial en el manejo específico localizado de malas hierbas, para controlarlas ya que compiten con los cultivos, e introducen enfermedades o plagas, reduciendo el rendimiento (Lati *et al*., 2022). El aprendizaje profundo, como subcampo del aprendizaje automático, ha logrado avances significativos en el análisis de imágenes extrayendo patrones complejos. Sin embargo, se basa en grandes conjuntos de datos etiquetados, que requieren una cantidad considerable de tiempo y requerir expertos en malherbología. Stable Diffusion (SD), se presenta como un novedoso modelo de aprendizaje profundo para generar imágenes de alta calidad y realistas mediante la aplicación de un proceso de difusión controlada en un espacio latente. Dicho modelo ha sido implementado en este estudio para generar imágenes artificiales de malas hierbas para cultivos de tomate. El estudio tiene como objetivo comparar el rendimiento de los modelos de aprendizaje profundo Yolov8 y RetinaNet entrenados en imágenes de malas hierbas reales y artificiales y evaluar el aumento de precisión al utilizar un conjunto de datos mixto.

MATERIALES Y MÉTODOS

En mayo y junio de 2021 y 2022 se tomaron imágenes de las malas hierbas dicotiledóneas *Solanum nigrum* L. y *Portulaca oleracea* L., y de la mala hierba monocotiledónea *Setaria verticillata* L., en Santa Amalia, Extremadura (España), utilizando una cámara Canon PowerShot SX540 HS. Las imágenes se tomaron en tres campos comerciales de tomate, con un ángulo entre el eje de la cámara y el suelo de -70° a 90° a una altura media de 1 m. Expertos identificaron visualmente cada especie de mala hierba, garantizando la exactitud de las etiquetas. El conjunto de datos de imágenes reales se amplió con imágenes artificiales a través de SD con el objetivo de aumentar el volumen de datos para el entrenamiento de los modelos de detección de objetos, las redes neuronales convolucionales Yolov8l y RetinaNet. La investigación se realizó utilizando una metodología en dos fases: 1) evaluación del rendimiento del entrenamiento de los modelos de detección de malas hierbas utilizando imágenes reales y artificiales por separado; 2) evaluación de la mejora del rendimiento conseguida utilizando un conjunto de datos mixto de imágenes reales y artificiales generadas por Stable Diffusion. La evaluación el rendimiento de las redes neuronales se realizó calculando el *mean Average Precision* (mAP) con un *Intersection over Union* (IoU) de 0.5 (@0.5). A continuación, estas imágenes se sometieron a dos etapas de preprocesamiento, en función de la fase de entrenamiento. El primer paso de preprocesamiento consistió en aplicar un enfoque de ventana deslizante para recortar cada imagen en subimágenes más pequeñas, creando un conjunto de datos de imágenes reales para entrenar, validar y testear los modelos de detección. El segundo paso de preprocesamiento, utilizado en la segunda etapa, consistió en extraer las malas hierbas óptimas para entrenar la SD. Así pues, la segunda etapa consistió en entrenar el modelo SD para generar imágenes fotorrealistas sintéticas de malas hierbas lo suficientemente realistas como para utilizarlas en el entrenamiento de los modelos Yolov8l y RetinaNet. Se utilizó un conjunto de validación y prueba que contenía las tres especies de malas hierbas consideradas y que representaba las condiciones reales del campo, es decir, que mostraba densidades variables e instancias superpuestas.

Proceso de generación de imágenes artificiales de Stable Diffusion

El proceso de generación de imágenes sintéticas mediante SD constó de dos etapas: 1) entrenamiento y extracción de imágenes artificiales, y 2) obtención de anotaciones para cada imagen generada. Cada especie de mala hierba se entrenó por separado utilizando 30 imágenes reales, y se extrajeron las imágenes artificiales utilizando un prompt, i.e., utilizando un texto de entrada para SD a partir del cual se generaron las imágenes (Figura 1).

Las imágenes reales se re-escalaron a 512x512 píxeles, ya que eran demasiado grandes para utilizarlas como imágenes de entrenamiento en SD. Las características óptimas de la imagen para mejorar la capacidad de aprendizaje de SD son una relación

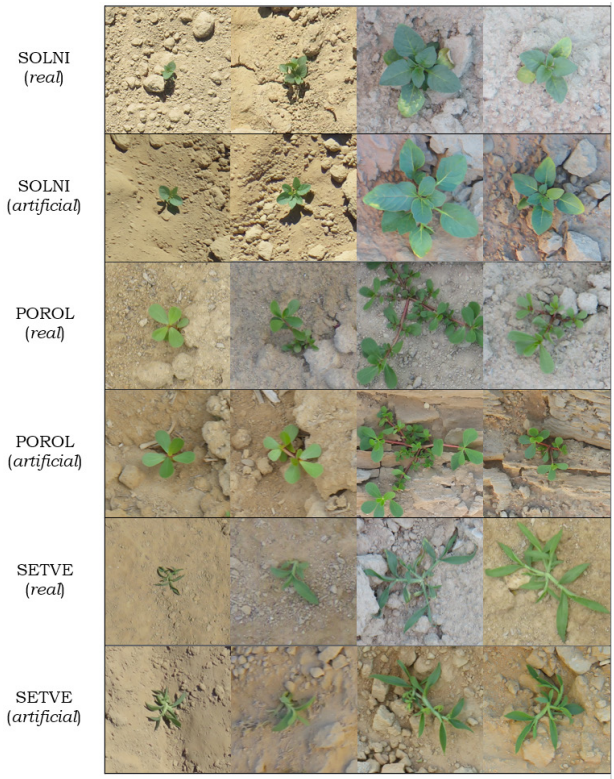

Figura 1 - Ejemplos de malas hierbas reales y artificiales extraídas de Stable Diffusion. Los nombres aparecen en código EPPO.

de aspecto cuadrada, la mala hierba centrada en la imagen y no demasiado pequeña. Para manejar diversas condiciones de iluminación, se seleccionó un conjunto de datos de imágenes reales bastante variable para entrenar el modelo de SD.

RESULTADOS Y DISCUSIÓN

Los modelos de aprendizaje profundo para la identificación de malas hierbas tienen dificultades para detectar y distinguir todas las especies de malas hierbas, especialmente durante las primeras etapas de crecimiento. Este estudio descubrió que aumentar los datos de entrenamiento con imágenes artificiales generadas por SD mejoró el rendimiento de los modelos de detección CNN hasta en un 9,1% en imágenes en condiciones de campo reales (Tabla 1). La Tabla 1 resume los experimentos realizados para cada red neuronal Yolov8l (Y-) y RetinaNet (R-). Los cinco primeros experimentos corresponden a la fase 1 (1-5), mientras que los cuatro últimos a la fase 2 (6-9). Todos los experimentos tienen como objetivo analizar el comportamiento de las redes neuronales al entrenarlas con diferentes cantidades de datos reales, artificiales y mixto. Yolov8l superó a RetinaNet cuando se entrenó con imágenes artificiales. Sin embargo, los modelos entrenados sólo con imágenes reales obtuvieron mejores resultados que los modelos entrenados sólo con imágenes artificiales, ya que Yolov8l y RetinaNet mejoraron un 3,6% y un 6,1%, respectivamente. Sin embargo, cuando se entrenaron con un conjunto de datos mixto (al conjunto de imágenes reales se añaden las artificiales), los resultados mejoraron significativamente, y todos los modelos superaron un mAP@0.5 de 0,91. Yolov8l obtuvo su mejor resultado en el experimento 9-Y (Tabla 1), con un mAP@0.5 de 0,93, un 8,1% más que cuando se entrenó sólo con imágenes reales. RetinaNet alcanzó su mejor rendimiento en los experimentos 7-R y 8-R (Tabla 1), con un mAP@0.5 de 0,928. Esto sugiere una posible saturación a partir de 150 imágenes artificiales por especie, donde el modelo no mejora su rendimiento al aumentar el número de imágenes. Hasta la fecha, diversos conjuntos de datos han sido puesto a disposición pública como DeepWeeds (Olsen *et al*., 2019), Crop/Weed Field Image (Haug y Ostermann, 2015), WeedMap (Sa *et al*., 2018) y el conjunto de datos WeedImages en https://www.weedimages.org/. Sin embargo, su aplicabilidad puede verse limitada por diferentes entornos. Nuestra investigación proporciona una solución mediante la generación de imágenes artificiales personalizadas utilizando SD, prevaleciendo la calidad de las muestras sobre la cantidad. Con tan solo 30 muestras reales se obtuvieron hasta 600 imágenes por especie de mala hierba, lo que se tradujo en elevados valores de mAP tanto para Yolov8 como para RetinaNet. Este enfoque se exige menos potencia de cálculo, lo que ahorra tiempo y energía a la vez que se obtienen conjuntos de datos óptimos para el entrenamiento de CNN. Aunque las redes generativas adversariales (GAN) son prometedoras, se enfrentan a retos debido a la complejidad visual de las especies de malas hierbas y a la necesidad de conjuntos de datos sustanciales, a diferencia de nuestro método. Las imágenes reales ofrecen una mayor diversidad de características que las artificiales. Por ejemplo, Yolov8l y RetinaNet lograron más de 0,84 mAP cuando se entrenaron con 228 malas hierbas reales, frente a la necesidad de 1.800 artificiales para una precisión similar. Además, las imágenes artificiales vienen

Tabla 1 - Rendimiento obtenido en los experimentos de detección de malas hierbas en condiciones reales con Yolov8l y RetinaNet. Los experimentos consistieron en entrenar a los modelos con diferentes cantidades de imágenes artificiales. Se indica la ganancia de rendimiento al aumentar el número de malas hierbas artificiales en la última columna

				AP@0.5					
Modelo	**Experimento**	**Malas Hierbas Reales**	**Malas Hierbas Artificiales**	SOLNI	POROL	SETVE	**mAP@0.5**	**Ganancia de Rendimiento**	
Yolov8l	1-Y	0	228	0.744	0.726	0.695	0.721		
	2-Y	0	450	0.756	0.783	0.787	0.775	1-Y → 2-Y	5,4%
	3-Y	0	900	0.772	0.723	0.84	0.778	2-Y → 3-Y	0,3%
	4-Y	0	1.800	0.776	0.833	0.829	0.813	3-Y → 4-Y	3,5%
	5-Y	228	0	0.92	0.761	0.867	0.849	4-Y → 5-Y	3,6%
	6-Y	228	228	0.943	0.884	0.901	0.909	5-Y → 6-Y	6%
	7-Y	228	450	0.909	0.913	0.929	0.917	5-Y → 7-Y	6,8%
	8-Y	228	900	0.935	0.905	0.928	0.923	5-Y → 8-Y	7,4%
	9-Y	228	1.800	0.936	0.952	0.904	0.93	5-Y → 9-Y	8,1%
RetinaNet	1-R	0	228	0.635	0.706	0.771	0.704		
	2-R	0	450	0.626	0.8	0.765	0.731	1-R → 2-R	2,7%
	3-R	0	900	0.702	0.794	0.774	0.757	2-R → 3-R	2,6%
	4-R	0	1.800	0.743	0.783	0.802	0.776	3-R → 4-R	1,9%
	5-R	228	0	0.840	0.838	0.835	0.837	4-R → 5-R	6,1%
	6-R	228	228	0.918	0.922	0.914	0.918	5-R → 6-R	8,1%
	7-R	228	450	0.942	0.92	0.923	0.928	5-R → 7-R	9,1%
	8-R	228	900	0.919	0.927	0.939	0.928	5-R → 8-R	9,1%
	9-R	228	1.800	0.921	0.913	0.924	0.919	5-R → 9-R	8,2%

con etiquetado automático, lo que reduce los costes relacionados con el etiquetado por expertos.

CONCLUSIONES

Este estudio propone un nuevo enfoque para entrenar algoritmos de detección de cultivos y malas hierbas que reduce la necesidad de intervención humana implementando Stable Diffusion (SD), una técnica que puede generar imágenes fotorrealistas a partir de conceptos aprendidos. SD se utilizó para aumentar los datos de entrenamiento, lo que mejoró la capacidad de generalización y la resistencia al ruido de los modelos de detección. El rendimiento de los modelos de detección aumentó entre un 6% y un 9% al aumentar el conjunto de datos con imágenes artificiales, alcanzando un mAP máximo del 93%. El enfoque propuesto se probó en un conjunto de datos de imágenes de cultivos de tomate con malas hierbas, y los resultados mostraron que supera a los métodos existentes en términos de mAP, incluso cuando se utilizan sólo 30 imágenes reales para generar imágenes artificiales. El método propuesto también puede extenderse a otros cultivos y especies de malas hierbas, y tiene potencial para utilizarse en el desarrollo de sistemas automatizados de gestión de malas hierbas.

AGRADECIMIENTOS

Esta investigación ha sido financiada por la AEI (Ministerio de Ciencia e Innovación, España), subvención número TED2021-130031B-I00, según el caso, por "FEDER Una manera de hacer Europa", por la "Unión Europea" o por la "Unión Europea NextGenerationEU/PRTR" y PID2020-113229RBC43/AEI/10.13039/501100011033. A Juan Ramón Andújar por facilitar los campos comerciales de tomate en Badajoz (Extremadura).

REFERENCIAS BIBLIOGRÁFICAS

Haug, S. & Ostermann, J. (2015) - *A Crop/Weed Field Image Dataset for the Evaluation of Computer Vision Based Precision Agriculture Tasks*, Springer International Publishing, Cham. p. 105-116. https://doi.org/10.1007/978-3-319-16220-1_8

Lati, R.N.; Rasmussen, J.; Andujar, D.; Dorado, J.; Berge, T.W.; Wellhausen, C.; Pflanz, M.; Nordmeyer, H.; Schirrmann, M.; Eizenberg, H.; Neve, P.; Jørgensen, R.N. & Christensen, S. (2021) - Site-specific weed management—constraints and opportunities for the weed research community: Insights from a workshop. *Weed Research,* vol. 61, n. 3, p. 147-153. https://doi.org/10.1111/wre.12469

Olsen, A.; Konovalov, D.A.; Philippa, B.; Ridd, P.; Wood, J.C.; Johns, J.; Banks, W.; Girgenti, B.; Kenny, O.; Whinney, J.; Calvert, B.; Azghadi, M.R. & White, R.D. (2019) - DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Scientific Reports,* vol. 9, art. 2058. https://doi.org/10.1038/s41598-018-38343-3

Pérez-Ortiz, M.; Peña, J.M.; Gutiérrez, P.A.; Torres-Sánchez, J.; Hervás-Martínez, C. & López-Granados, F. (2015) - A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Applied Soft Computing,* vol. 37, p. 533-544. https://doi.org/10.1016/j.asoc.2015.08.027

Sa, I.; Popović, M.; Khanna, R.; Chen, Z.; Lottes, P.; Liebisch, F.; Nieto, J.; Stachniss, C.; Walter, A. & Siegwart, R. (2018) - WeedMap: A Large-Scale Semantic Weed Mapping Framework Using Aerial Multispectral Imaging and Deep Neural Network for Precision Farming. *Remote Sensing*, vol. 10, n. 9, art. 1423. https://doi.org/10.3390/rs10091423