

# Segmentación Semi-Supervisada de Malezas en Entornos Agrícola de Alta Complejidad

1<sup>st</sup> Fernando Blanco  
Departamento de Ciencias de la Computación  
FCEIA-UNR  
Rosario, Argentina  
ferblanco203@gmail.com

2<sup>nd</sup> Taihú Pire  
CIFASIS  
CONICET-UNR  
Rosario, Argentina  
pire@cifasis-conicet.gov.ar

3<sup>rd</sup> Claus Smitt  
AgRobotics & Engineering  
University of Bonn  
Bonn, Alemania  
csmitt@uni-bonn.de

**Resumen**—La robótica agrícola viene a facilitar el trabajo agropecuario, mejorándolo en eficiencia, eficacia y ahorro tanto de productos como de trabajo humano. Dentro de las áreas de robótica agrícola, una de suma importancia es la detección de maleza, la cual consiste en mediante sensores encontrar malezas en los campos. Nosotros presentamos un método de detección panóptico de malezas, el cual utilizando computer vision es capaz de identificar malezas, dándonos su forma y considerando cada una como una maleza individual. Esto es realmente útil para luego realizar estudios de los campos y/o actuar en consecuencia a la disposición de malezas. Logramos entrenar un modelo Mask2Former que es capaz de detectar malezas sobre cultivo, además lo probamos con un video para ver como se comportaba en movimiento obteniendo resultados aceptables y una buena velocidad de procesamiento.

**Keywords**—Visión Artificial, Aprendizaje Profundo, Segmentación Panóptica, Agricultura, Aprendizaje Semi-Supervisado

## I. INTRODUCCIÓN

Hace ya varios años, el sector agrícola muestra gran interés en la utilización de robots autónomos para realizar diversas tareas como siembra, riego, remoción de maleza o cosecha. Por un lado, el uso de robots autónomos permite reducir la cantidad de horas hombre y por otro, realizar la tarea de una manera más sustentable y sostenible, haciendo un uso eficiente de los recursos al mismo tiempo que se reduce el impacto ambiental [1], [2].

La detección de malezas es de gran interés, ya que pueden darle al productor agrícola información sobre la distribución de las mismas en el lote, estimar la cantidad y así poder tomar decisiones al respecto como podría hacer un desmalezado preciso.

La existencia de malezas pueden perjudicar a los cultivos al generar competencia con ellos. Conocer dónde, cuáles y cuántas malezas tiene, puede ser útil a la hora de estudiar el lote, ya que beneficia a la hora de decidir como poder tratarlo. Para tratar esto se han desarrollado diversos enfoques, sin embargo, no se atacó desde el lado de segmentación panóptica, lo cual puede ser útil para contabilizar y separar una maleza de otra.

En este trabajo intentaremos detectar instancias de malezas en entornos de alta complejidad mediante segmentación panóptica utilizando métodos de inteligencia artificial, y de esta manera lograr obtener las distintas malezas que se ven en una imagen como así también la forma de cada una.

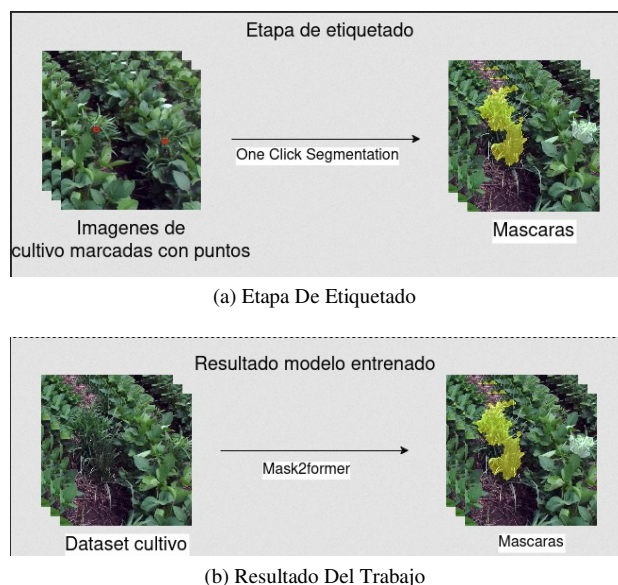


Figura 1. Representación del método usado en este trabajo. En 1a se ve la etapa de etiquetado donde se le pasó imágenes con puntos (clicks) a un One Click Segmentation y se obtuvo pseudo mascarar. En 1b se muestra el resultado del trabajo, el cual es un modelo entrenado con las pseudo máscaras que se lo alimenta con imágenes de cultivo y se obtiene las detecciones panópticas de las malezas.

El resto del trabajo está estructurado de la siguiente manera, en la sección II se muestra el estado del arte de trabajos relacionados con la detección en el sector agrícola. En la sección III se explica el proceso de como logramos realizar detecciones panópticas de malezas sobre cultivo. En la sección IV se muestran una serie de experimentos y resultados de los mismos realizados por nosotros. En la sección V concluimos el trabajo con un resumen de lo propuesto y un análisis de los resultados obtenidos, además se mencionan posibles trabajos a futuros para extender lo exployado en este artículo.

## II. TRABAJO RELACIONADO

### II-A. Monitoreo de cultivos

Dentro de la robótica para la agricultura tenemos lo que son robots de monitoreo de cultivo, estos inspeccionan cultivos de forma autónoma lo cual es útil para recolectar información de las plantaciones, como puede ser la contabilización de frutos,

estado de salud de las plantas y el suelo, entre otras. Entre los trabajos dentro de esta área se encuentran PATHoBot [3], un robot de monitoreo de pimientos y tomates en invernaderos desarrollado por la universidad de Bonn, Alemania. También, en la misma universidad, se ha desarrollado el prototipo BonnBot-I [4], un robot que realiza desmalezado y monitoreo de campo en diferentes ciclos de crecimiento. AgBot II [5], un robot desmalezador autónomo desarrollado por la universidad tecnológica de Queensland, Australia. En Argentina, el instituto CIFASIS (CONICET-UNR) ha desarrollado un prototipo de robot desmalezador. El mismo ha permitido la generación del Rosario dataset [6], utilizado para este trabajo.

El campo agrícola es un entorno desafiante para las tareas de monitoreo debido a que es altamente repetitivo, dinámico debido al movimiento de las hojas producido por el viento, presenta cambios de iluminación y movimientos bruscos producidos por el terreno irregular, entre otros [7]. Por lo tanto, el desarrollo de métodos de monitoreo robustos a estos problemas mencionados, es de gran utilidad para la industria agrícola.

Actualmente, los trabajos del estado del arte en monitoreo de cultivos agrícolas, hacen uso de técnicas de aprendizaje profundo (*deep learning*), cuyos métodos requieren de una gran cantidad de datos para ser entrenados. Para realizar estos monitoreos, se utilizan técnicas de entendimiento de escena (*scene understanding*) y segmentación semántica. En [8] se presenta una red profunda PAG-NeRF capaz de ubicar frutos en el espacio utilizando estas dos técnicas.

## II-B. Segmentación Panóptica

La segmentación panóptica es una combinación de segmentación semántica y segmentación de instancias [9]. La segmentación semántica consisten en asignarle a cada píxel una etiqueta indicando a qué clase pertenece, a elementos incontables (ej: tierra, cielo, etc.) y contables (ej: planta, maleza, etc.). La segmentación de instancias consiste en diferenciar objetos contables de una misma clase, asignándoles un identificador único a cada una. Esta representación provee gran cantidad de información para una comprensión del entorno. Este tipo de predicción se ha empleado en diversas aplicaciones de agricultura, mapeo tales como contabilización de frutas [10], [11], detección de malezas en campo abierto [12].

Uno de los primeros trabajos en resolver esta tarea con redes neuronales convolucionales fue PanopticDeepLab [13], prediciendo segmentación semántica, centros de instancias y las distancias a los mismos por separado. Luego estas predicciones se combinan para obtener la predicción panóptica. Actualmente, el estado del arte para esta tarea está dominado por [14], [15], siendo Mask2Former [15] una de las arquitecturas más utilizada. Esta arquitectura emplea esquema de encoder-decoder en el espacio de imágenes y luego los mapas de activación de diferentes resoluciones son progresivamente destilados en predicciones de máscaras con clases y etiquetas individuales asociadas. Al igual que otros métodos de aprendizaje profundo, esta arquitectura requiere una gran cantidad de datos etiquetados con máscaras para ser entrenado.

Una forma de reducir la cantidad de anotaciones es el uso de métodos de aprendizaje semisupervisado.

## II-C. Segmentación semisupervisada con etiquetas parciales

Anteriormente, se han desarrollado métodos para entrenar modelos de segmentación con etiquetas parciales en lugar de máscaras detalladas (ej: clicks [16], cajas 2D [17]). Estos métodos permite, a partir de un número reducido de etiquetas detalladas y muchas etiquetas parciales, entrenar un modelo para generar una gran cantidad de pseudo etiquetas de segmentación. Estas pueden ser empleadas para entrenar un modelo de segmentación estándar en forma supervisada.

Recientemente, Zimmer *et al.* [12] presentaron un método para obtener pseudo etiquetas panópticas de malezas en campo abierto a partir de clicks, entrenándose con escasas imágenes etiquetadas con máscaras. Usando una arquitectura similar a [13], este método extrae puntos aleatorios cercanos al centroide de las etiquetas. Estos son suavizados y agregados como un canal extra a la entrada, aprendiendo así la relación entre puntos/clicks y predicciones panópticas en la imagen. Este método solo fue evaluado en datasets de complejidad reducida con vista superior de cultivos en etapas tempranas de desarrollo.

En el presente trabajo evaluamos el uso de este método como método para reducir el esfuerzo de anotación en un dominio agrícola de gran complejidad [6]. A su vez, evaluamos la viabilidad de usar las pseudo etiquetas producidas para entrenar un modelo Mask2Former [15] para detección de malezas.

## III. MÉTODO PROPUESTO

La segmentación panóptica de malezas nos permite diferenciar distintas instancias de malezas y ver sus formas. Para esto entrenamos una red Mask2Former [15].

Para entrenar el Mask2Former conseguimos del Rosario Dataset [6] imágenes de un cultivo de soja, donde las malezas no eran fáciles de ver, ya que algunas se encontraban creciendo entre los cultivos, lo que dificultaba dibujar las máscaras a mano.

Para evitarnos esa tarea optamos por entrenar un modelo de One-Click Segmentation [12], por lo que nosotros simplemente marcamos las malezas con puntos y este modelo se encargó de generar las máscaras para nuestro dataset. Las máscaras generadas por One Click Segmentation, eran un poco sucias, ya que eran des conexas y contaban con algunos píxeles sueltos. Esto a la hora de escribir las anotaciones se tornaba un poco molesto, ya que se marcaban como distintas a pesar de pertenecer a la misma máscara. Para evitar esto, sometimos a las máscaras a un procesamiento de imágenes utilizando operaciones de open y close, lo cual nos emprolijó las máscaras para poder anotarlas correctamente y pasarlas al entrenador del Mask2Former.

Una vez entrenado el modelo, lo utilizamos sobre un vídeo perteneciente al Rosario Dataset para probar que tal funciona. Para esto leímos frame por frame un vídeo, pasando cada uno por el modelo y pintando los resultados sobre estos.

Finalmente, generamos un vídeo nuevo con el resultado del procesamiento del anterior.

En la figura 2 se puede ver un diagrama del flujo de trabajo del método propuesto

#### IV. EXPERIMENTOS

En esta sección se presentan los experimentos que validan el método propuesto. La computadora utilizada para realizar los experimentos consta de una CPU Intel i7-4790 con 16 GB de RAM y una GPU ASUS GeForce 1070.

Como se comentó en la sección III, primero entrenamos un One Click Segmentation, el cual usamos para generar mascararas, como las que se ven en la Figura 3. Para este entrenamiento se utilizaron 10 imágenes. Luego con eso generamos 155 imágenes de máscaras, de las cuales usamos 125 de entrenamiento, 15 de validación y 15 de evaluación para entrenar un Mask2Former con el cual tuvimos resultados de validación bastante satisfactorios, mostrados en la Figura 4.

Se testeó el modelo resultando utilizando 12 imágenes con máscaras hechas a mano, para sacar distintas métricas la Tabla I, y obtuvimos la matriz de confusión mostrada en la Tabla II.

Para probar bajo un contexto similar al que se lo quiere usar, pasamos un vídeo grabado por el robot desmalezado de CIFASIS por el modelo, cuyo resultado se ve en la Figura 5. En las imágenes podemos ver que las malezas las detectas bien (Figura 5a), sin embargo, hay plantas de cultivos y otras cosas que detecta como falsos positivos (Figura 5b).

Tabla I  
DESEMPEÑO EN TAREAS DE SEGMENTACIÓN PANÓPTICA

	Mask2Former
AP [%]	8.232311210158785
AP 50 [%]	25.182696841112684
InstanceW eedIoU [%]	24.41
SemanticW eedIoU [%]	47.96

		Predicted	
		Negative	Positive
True	Negative	0.9971	0.0029
	Positive	0.7306	0.2694

Tabla II  
MATRIZ DE CONFUSIÓN MODELO ENTRENADO

#### V. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo nosotros propusimos e intentamos hacer una detección panóptica de malezas utilizando una red Mask2Former, entrenándola con pseudo mascarás generadas por One Click Segmentation, lo cual nos permitiría contabilizar y conocer ubicación y forma de cada maleza. Con esto probaríamos, además, que tan bien funciona One Click Segmentation en contextos complejos y que tan útil son sus resultados para entrenar otra red.

Las métricas obtenidas no se ven muy buenas, pareciera no ser preciso y las predicciones de malezas no coincidir con los

datos reales. Mirando la matriz de confusión podemos deducir que nuestro modelo tiene bastante falsos negativos y bajos verdaderos positivos. Sin embargo, estos resultados no nos sorprenden, considerando la falta de refinamiento del trabajo, son valores esperables, probablemente mejorando el dataset y el etiquetado, las métricas mejoren, como así también la matriz de confusión se nos acerque más a una matriz identidad.

Al probar el resultado en un vídeo pudimos ver como se comporta el modelo en una situación de movimiento, donde las malezas se ven detectadas consistentemente, aunque cuenta con varios falsos positivos. Estimamos que los falsos positivos se deben al dataset, probablemente ampliando la cantidad de imágenes usadas para entrenar el Mask2Former, mejorando las etiquetas usadas para One Click Segmentation y usando imágenes más diversas, se podrían disminuir. También vimos que la misma maleza de un frame a otro la detecta como si fueran una nueva, esto se ve con el cambio de color en la máscara dibujada, lo cual se debe a que el modelo se ejecuta por frame ignorando que perteneces a un vídeo secuencial, por lo que ignora el contexto previo.

A futuro, nos interesa lograr que reconozca la misma maleza dentro de un contexto de movimiento, así poder generar un mapa de malezas dentro de un lote. Para esto nos gustaría agregar este modelo en un sistema de SLAM, el cual luego quisiéramos poder colocar en un robot que recorra de forma autónoma los lotes de cultivo. Otro posible trabajo a futuro, es reconocer y diferenciar distintas especies de malezas, para lo cual se debería anotar un dataset con las distintas clases para entrenar el modelo. Juntando ambos trabajos propuestos se podría generar un mapa de malezas mucho más rico en información del lote.

#### AGRADECIMIENTOS

Este trabajo fue apoyado por CONICET (PIBAA N° 0042), AGENCIA I+D+i (PICT 2021-570) y Universidad Nacional de Rosario (PCCT-UNR 80020220600072UR).

#### REFERENCIAS

- [1] F. A. Auat Cheein and R. Carelli, "Agricultural Robotics: Unmanned Robotic Service Units in Agricultural Tasks," *IEEE Industrial Electronics Magazine*, vol. 7, no. 3, pp. 48–58, 2013.
- [2] C. W. Bac, E. J. van Henten, J. Hemming, and Y. Edan, "Harvesting Robots for High-value Crops: State-of-the-art Review and Challenges Ahead," *Journal of Field Robotics*, vol. 31, no. 6, pp. 888–911, 2014.
- [3] C. Smitt, M. Halstead, T. Zaenker, M. Bennewitz, and C. McCool, "Pathobot: A robot for glasshouse crop phenotyping and intervention," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 2324–2330.
- [4] A. Ahmadi, M. Halstead, and C. McCool, "Bonnbot-i: A precise weed management and crop monitoring platform," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 9202–9209.
- [5] O. J. Bawden, "Design of a lightweight, modular robotic vehicle for the sustainable intensification of broadacre agriculture," Ph.D. dissertation, Queensland University of Technology, 2015.
- [6] T. Pire, M. Mujica, J. Civera, and E. Kofman, "The Rosario Dataset: Multisensor Data for Localization and Mapping in Agricultural Environments," *Intl. J. of Robotics Research*, vol. 38, no. 6, pp. 633–641, 2019.
- [7] J. Cremona, R. Comelli, and T. Pire, "Experimental evaluation of Visual-Inertial Odometry systems for arable farming," *Journal of Field Robotics*, vol. 39, no. 7, pp. 1123–1137, 2022.

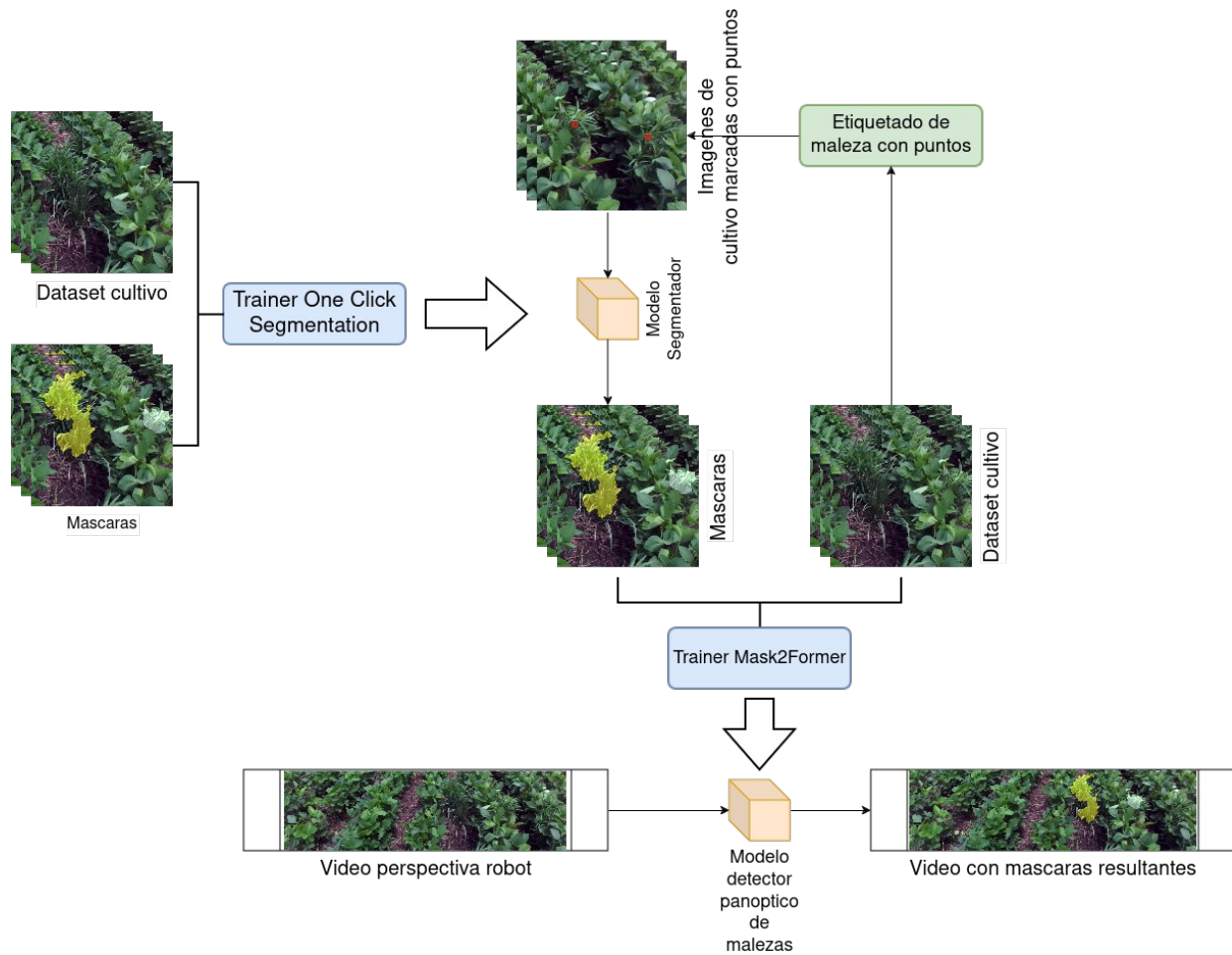
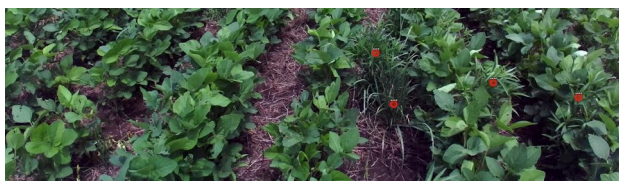
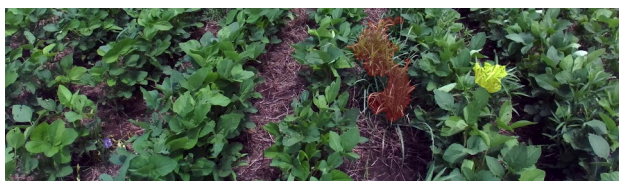


Figura 2. Flujo de trabajo del método implementado, los rectángulos celeste son entrenadores de IA, los cubos naranjas son modelos de IA entrenados, los rectángulos verdes representan procesos manuales



(a) Imagen anotada con puntos



(b) Mascaras generadas por One Click Segmentation

Figura 3. Resultados de One Click Segmentation

- [8] C. Smitt, M. Halstead, P. Zimmer, T. Läbe, E. Guclu, C. Stachniss, and C. McCool, "Pag-nerf: Towards fast and efficient end-to-end panoptic 3d representations for agricultural robotics," *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 907–914, 2023.
- [9] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proceedings of the IEEE/CVF conference on computer*

- vision and pattern recognition*, 2019, pp. 9404–9413.
- [10] M. Halstead, A. Ahmadi, C. Smitt, O. Schmittmann, and C. McCool, "Crop agnostic monitoring driven by deep learning," *Frontiers in plant science*, vol. 12, 2021.
- [11] Y. Pan, F. Magistri, T. Läbe, E. Marks, C. Smitt, C. McCool, J. Behley, and C. Stachniss, "Panoptic mapping with fruit completion and pose estimation for horticultural robots," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 4226–4233.
- [12] P. Zimmer, M. Halstead, and C. McCool, "Panoptic one-click segmentation: applied to agricultural data," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2478–2485, 2023.
- [13] B. Cheng, M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L.-C. Chen, "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 475–12 485.
- [14] J. Jain, J. Li, M. T. Chiu, A. Hassani, N. Orlov, and H. Shi, "Oneformer: One transformer to rule universal image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 2989–2998.
- [15] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1290–1299.
- [16] N. Xu, B. Price, S. Cohen, J. Yang, and T. S. Huang, "Deep interactive object selection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 373–381.





(a) Mascaras de validación



(b) Resultado del modelo

Figura 4. Resultados de validación del modelo entrenado



(a) Malezas detectadas en vídeo



(b) Algunas plantas marcadas como falsos positivos

Figura 5. Frames del vídeo generado con el modelo entrenado

- [17] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang, "Deep grabcut for object selection," *arXiv preprint arXiv:1707.00243*, 2017.