

# Red neuronal convolucional para la detección de tomates cultivados en invernadero

Paulina Arregocés Guerra, [paulina.arregocesg@udea.edu.co](mailto:paulina.arregocesg@udea.edu.co)  
Juan Pablo Velásquez Cardona, [juan.velasquez25@udea.edu.co](mailto:juan.velasquez25@udea.edu.co)

*Universidad de Antioquia*

*Fundamentos de Deep Learning*

2024

## I. CONTEXTO DE APLICACIÓN

La precisión en la detección de frutas es crucial para los sistemas de recolección robótica agrícola. Sin embargo, esta tarea es desafiante en entornos de producción real debido a factores como cambios de iluminación, obstrucciones por ramas y hojas, agrupaciones de frutas y sombreado [1].

En el proyecto ROBOCARE, los autores [2] se enfocaron en desarrollar soluciones robóticas para la monitorización y cosecha de tomates en invernaderos. Para este fin, los autores recolectaron un conjunto de datos de imágenes de tomates en un invernadero en Viana do Castelo, Portugal. El conjunto de datos se obtuvo utilizando una cámara estéreo (cámara ZED) montada en un manipulador estático sobre una plataforma robótica. Este dataset es público y puede encontrarse en la siguiente dirección <https://rdm.inesctec.pt/dataset/ii-2021-001>.

Para este trabajo, se seleccionó el conjunto de datos referenciado y adicionalmente se utilizó un segundo conjunto de datos perteneciente a otro entorno de operación. Este entorno es un cultivo de tomates bajo invernadero en Carmen de Viboral, Antioquía, con imágenes capturadas por una cámara RGB de marca Parrot Sequoia manipulada de forma manual.

## II. OBJETIVO DE MACHINE LEARNING

El objetivo del trabajo es implementar un método de Deep Learning, particularmente una Red

Neuronal convolucional (CNN, de sus siglas en inglés) para la detección de tomates cultivados bajo invernadero dado el conjunto de datos seleccionado.

## III. DATOS

Los datos que se dispondrán para el entrenamiento y prueba del modelo serán imágenes de 300 x 300 píxeles de plantas de tomates en invernadero, adquiridas a través de una cámara ZED estéreo. Éste *dataset* ha sido utilizado anteriormente por Magalhaes *et al* en aplicaciones de detección de tomates [2][3]. Cuenta con las siguientes características.

- Tipo de datos: dataset Pascal VOC (imágenes .jpg y archivos XML).
- Número de datos y tamaño en disco: 25758 imágenes con anotaciones, 705MB.
- Distribución de las clases: clase única “tomato”.
- Repositorio original: <https://rdm.inesctec.pt/dataset/ii-2021-001>.

Adicionalmente, se tendrá un *dataset* propio de prueba con imágenes de plantas de tomate en invernadero adquiridas con una cámara RGB. Cuenta con las siguientes características:

- Tipo de datos: dataset Pascal VOC (imágenes .png y archivos XML).
- Número de datos y tamaño en disco: 557 imágenes con anotaciones, 1.0 GB.
- Distribución de las clases: clase única “tomato”.

#### IV. MÉTRICAS DE DESEMPEÑO

Para evaluar el desempeño del método de detección de tomates, se han empleado comúnmente las siguientes métricas de desempeño:

##### IV-A. Precision

La precisión del método indica la proporción de falsos positivos como resultado de la detección. Con respecto al caso de uso, indica que proporción de tomates son incorrectamente detectados por el método. Para determinar la precisión del modelo, se utiliza la Ecuación 1.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Donde TP es *True Positives* y FP es *False Positives*

##### IV-B. Recall

La recuperación o recall del método indica la proporción de falsos negativos como resultado de la detección. Con respecto al caso de uso, indica que proporción tomates no son detectados correctamente por el método. Para determinar la recuperación del modelo, se utiliza la Ecuación 2.

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Donde TP es *True Positives* y FN es *False Negatives*

##### IV-C. F1-score

La métrica F1 del método es una combinación de las métricas precisión y recuperación. El objetivo de esta combinación es consolidar una sola métrica que mida tanto las proporciones de detecciones de falsos positivos como de falsos negativos como se ve en la Ecuación 3.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

##### IV-D. mAP (mean average precision)

La métrica de mAP es utilizada para identificar de forma global todas las precisiones en la detección de los objetos, teniendo en cuenta un umbral de decisión o confianza. Es decir, la detección de la ubicación del objeto puede estar sujeta a errores, donde 0 % implica que la caja de predicción se encuentra totalmente desubicada y 100 % que la ubicación de la caja se solapa perfectamente con la real. Por este motivo, se define un valor o un rango aceptable de valores. Este umbral aplica para todos los objetos presentes en la imagen y, por tanto, para determinar que tan bueno es el método se mide el promedio de todas las precisiones en una sola imagen o en un conjunto de imágenes como se indica en la Ecuación 4.

$$mAP = \frac{\sum_{i=1}^{NC} AP_i}{NC} \quad (4)$$

Donde NC es el número total de clases en el conjunto de datos, y AP es la precisión promedio de recuperación para la clase  $i$

##### IV-E. Inference Time

Existen diferentes formas de calcular el tiempo de inferencia de un modelo. Este puede ser en medido en términos de segundos o en cuadros por segundo (FPS) en caso de análisis de videos. Por lo general, puede tomarse midiendo el tiempo que demora el modelo en el procesamiento de una sola imagen o de un conjunto de imágenes para tener una medida más realista.

#### V. REFERENCIAS Y RESULTADOS PREVIOS

Para definir la elección del conjunto de datos y las métricas de desempeño para el método de detección de tomates, se realizó una búsqueda de literatura enfocada en la detección de tomates cultivados bajo invernadero con técnicas de Deep Learning. Los documentos fueron seleccionados en un rango de tiempo de 4 años (2021-2024), resultando 16 investigaciones que fueron exploradas. De este conjunto de investigaciones excluimos aquellas que no utilizan un conjunto de datos propio. El resultado son 5 referencias, de las cuales se extrajo

Tabla I: Investigaciones analizadas.

Referencia	Objetivo principal	Número de imágenes del dataset	Métricas de desempeño	Resultados previos
[4]	Proponer un algoritmo mejorado de detección de tomates cherry utilizando YOLOv3 con el propósito de guiar robots en la recolección y mejorar la eficiencia de producción	1460	Verdadero positivo (TP) Falso positivo (FP) Verdadero negativo (TN) Falso negativo (FN). Precisión (P) Sensibilidad (R) Puntuación F AP	Los experimentos mostraron que el algoritmo tiene una tasa de precisión del 94.29 %, una tasa de recuperación del 94.07 % y un valor F1 del 94.18 %
[1]	Proponer dos modelos de detección de tomates: YOLODenseNet y YOLOMixNet, para resolver las dificultades de detección de frutas que involucran escenas naturales.	425	Precisión Recall F1 score AP	Los resultados obtenidos mostraron que el AP de YOLODenseNet al 98.3 % y YOLOMixNet al 98.4 % superaron a YOLOv3 y YOLOv4.
[2]	Automatizar el proceso de cosecha de tomates en invernaderos, el sistema de percepción visual necesita detectar el tomate en cualquier etapa del ciclo de vida (desde la flor hasta el tomate maduro).	18417	Recall x precision curve; mAP (mean Average Precision) Total recall Total precision F1-score Inference time	SSD MobileNet v2 tuvo el mejor rendimiento en comparación con SSD Inception v2, SSD ResNet 50, SSD ResNet 101 y YOLOv4 Tiny, con un F1 score del 66.15 %, un mAP del 51.46 % y un tiempo de inferencia de 16.44 ms.
[5]	Algoritmo ligero basado en YOLO para detección en tiempo real y grado de maduración de tomates. El algoritmo fue evaluado con un dataset hecho específicamente para la tarea.	932	Precision Recall mAP Size Params FLOPS	El modelo con mejor desempeño fue el YOLOv5 pero con un costo computacional mucho más alto. Después de reducir el número de capas, reemplazarlas con otras y comprimir el modelo, se logró un desempeño casi igual al modelo original.
[6]	Usar (SSD MobileNet v2 y YOLOv4) para detectar eficientemente tomates y comparar esos sistemas con un modelo basado en histogramas de color HSV para clasificar y determinar la etapa de maduración, a través de dos conjuntos de datos de imágenes adquiridas.	449 y 258	Intersection over Union(IoU) Recall Precision F1 score AP mAP Macro F1 score Balanced Accuracy	Destacó el modelo YOLOv4 con un F1-Score del 85.81 %. Para la tarea de clasificación, el YOLOv4 fue nuevamente el mejor modelo con un Macro F1-Score de 74.16 %. El modelo de espacio de color HSV superó al modelo SSD MobileNet v2, obteniendo resultados similares al modelo YOLOv4, con una Precisión Equilibrada del 68.10 %.

información referente al objetivo de trabajo, métodos estudiados, número de imágenes empleadas y resultados con métricas de desempeño presentadas en la Tabla I.

La información presentada en la Tabla I indica que el método más empleado para la detección de tomates ha sido YOLO en sus diferentes versiones y combinaciones. Los resultados asociados a este método son prometedores teniendo en cuenta los casos de obstrucción de los frutos, agrupaciones de frutos y variaciones de iluminación.

## REFERENCIAS

- [1] O. M. Lawal, "Development of tomato detection model for robotic platform using deep learning," *Multimedia Tools and Applications*, vol. 80, pp. 26751–26772, 7 2021.
- [2] S. A. Magalhães, L. Castro, G. Moreira, F. N. Dos Santos, M. Cunha, J. Dias, and A. P. Moreira, "Evaluating the single-shot multibox detector and yolo deep learning models for the detection of tomatoes in a greenhouse," *Sensors*, vol. 21, no. 10, 2021.
- [3] T. C. Padilha, G. Moreira, S. A. Magalhães, F. N. dos Santos, M. Cunha, and M. Oliveira, "Tomato Detection Using Deep Learning for Robotics Application," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12981 LNAI, pp. 27–38, 2021.
- [4] J. Chen, Z. Wang, J. Wu, Q. Hu, C. Zhao, C. Tan, L. Teng, and T. Luo, "An improved YOLOv3 based on dual path network for cherry tomatoes detection," *Journal of Food Process Engineering*, vol. 44, 10 2021.
- [5] T. Zeng, S. Li, Q. Song, F. Zhong, and X. Wei, "Light-weight tomato real-time detection method based on improved YOLO and mobile deployment," *Computers and Electronics in Agriculture*, vol. 205, 2023.
- [6] G. Moreira, S. A. Magalhães, T. Pinho, F. N. Dos Santos, and M. Cunha, "Benchmark of Deep Learning and a Proposed HSV Colour Space Models for the Detection and Classification of Greenhouse Tomato," *Agronomy*, vol. 12, 2 2022.