

Optimización avanzada de CNNs para la clasificación de bananas.

Julio Trujillo González
Universidad de Panamá
Correo: julio.trujillo@up.ac.pa
Orcid: <https://orcid.org/0000-0002-3664-8058>

DOI: <https://doi.org/10.61209/re.v2i2.62>

Resumen

Este estudio se centra en la clasificación precisa de bananas utilizando Redes Neuronales Convolucionales (CNN). La clasificación adecuada es crucial en la industria agrícola para garantizar la calidad y eficiencia. Se evaluaron dos modelos de CNN: uno estándar y otro con conexiones residuales. El modelo estándar demostró una precisión del 99.78%, mientras que el modelo con conexiones residuales alcanzó una precisión del 99.23%. A pesar de las diferencias, ambos modelos superaron a muchos métodos tradicionales, subrayando la eficacia del aprendizaje profundo en la clasificación de frutas. Estos resultados tienen el potencial de revolucionar la industria agrícola, reduciendo la necesidad de intervención manual y minimizando errores.

Palabras clave: agricultura, análisis de datos, inteligencia artificial, optimización.

Abstract

This study focuses on the precise classification of bananas using Convolutional Neural Networks (CNN). Proper classification is crucial in the agricultural industry to ensure quality and efficiency. Two CNN models were evaluated: a standard one and another with residual connections. The standard model demonstrated an accuracy of 99.78%, while the model with residual connections achieved an accuracy of 99.23%. Despite the differences, both models outperformed many traditional methods, underscoring the efficacy of deep learning in fruit classification. These results have the potential to revolutionize the agricultural industry, reducing the need for manual intervention and minimizing errors.

Keywords: agriculture, data analysis, artificial intelligence, optimization.

Introducción

La clasificación precisa de frutas en la industria agrícola es esencial para garantizar la calidad del producto y mantener la eficiencia en la cadena de suministro. Una clasificación inadecuada puede resultar en la contaminación de frutas frescas con frutas defectuosas, lo que puede tener un impacto negativo en la economía y en la confianza del consumidor. Tradicionalmente, esta tarea se ha llevado a cabo manualmente, lo que no solo es laborioso y consume tiempo, sino que también puede ser propenso a errores debido a la fatiga humana.

En este contexto, las técnicas de aprendizaje automático y, en particular, las redes neuronales convolucionales (CNN) han emergido como una solución prometedora. Estas técnicas ofrecen una forma automatizada y eficiente de clasificar las frutas, reduciendo la necesidad de intervención humana y, por lo tanto, los costos asociados. Dhande y Patil (2022) propusieron un modelo basado en aprendizaje profundo para la clasificación de frutas utilizando DenseNet, logrando una precisión del 97.82% en la detección de frutas frescas y podridas. Asimismo, Lingamgunta et al. (2023) destacaron la importancia de la modernización de los cultivos y cómo la detección precisa de la madurez de las frutas puede resultar en cosechas de mejor calidad y ahorros significativos. Además, un estudio sobre la clasificación de mangostán, una fruta popular en Malasia demostró que un modelo CNN optimizado puede alcanzar una precisión de clasificación del 94.99%, superando otros modelos de aprendizaje profundo (Sumari et al., 2021). Estos avances en la clasificación de frutas mediante técnicas de aprendizaje profundo no solo prometen mejorar la eficiencia y precisión en la industria agrícola, sino que también tienen el potencial de revolucionar la forma en que se abordan los desafíos en este campo.

Materiales y métodos

Adquisición de la base de datos y preprocesado

La base de datos se obtiene de Kaggle (s.f.) y contiene imágenes de bananas, divididas en dos clases: bananas frescas y bananas podridas. El tamaño total de la base de datos utilizada en este trabajo es de 4716 imágenes. Las imágenes de entrenamiento suman un total de 3805, mientras que el conjunto de prueba contiene 911 imágenes. Todas las imágenes pertenecen a las dos clases mencionadas, ver figura 1. Ahora, todo el conjunto de imágenes se redimensiona a 128x128x3 y se convierte en un array numpy para una convolución más rápida al construir el modelo CNN. Finalmente, el conjunto de imágenes convertido se etiqueta de acuerdo con la clase a la que pertenecen.



Figura 1
Ejemplos de imágenes de la base de datos.

Redes neuronales convolucionales

El aprendizaje profundo, una subdivisión prominente del aprendizaje automático, ha demostrado un rendimiento excepcional en diversas modalidades de datos. Una de sus aplicaciones más destacadas es la clasificación de imágenes utilizando redes neuronales convolucionales (CNN), figura 2. En el ámbito de la informática, las imágenes se representan como píxeles relacionados. Dentro de una imagen, un conjunto específico de píxeles podría representar un borde, una sombra o algún otro patrón.

La convolución es una técnica esencial para detectar estos patrones. Durante la computación, los píxeles de la imagen se representan mediante una matriz. Para detectar patrones, se utiliza una matriz "filtro" que se multiplica con la matriz de píxeles de la imagen. Estos tamaños de filtro pueden variar y la multiplicación depende puramente del tamaño del filtro. Se toma un subconjunto de la matriz de píxeles de la imagen basado en el tamaño del filtro para la convolución, comenzando desde el primer píxel (Wang et al., 2020).

Zhang et al. (2020) propusieron mejoras en el modelo tradicional de CNN, introduciendo el pooling estocástico en lugar del pooling promedio y máximo, y combinando la capa de dropout con la capa completamente conectada. Posteriormente, la salida de una convolución es un mapa de características que se envía a la siguiente capa. La siguiente capa en el modelo CNN es la capa de pooling, que reduce el tamaño del mapa de características y previene el sobreajuste (Yao et al., 2019).

Finalmente, la capa completamente conectada "aplana" la salida obtenida de las capas anteriores y las convierte en un vector único, que puede ser una entrada para la siguiente etapa. Las normalizaciones por lotes se pueden usar para normalizar los mapas de características obtenidos en cada etapa. Los abandonos (Dropouts) también se utilizan para aumentar la velocidad de cálculo (Bera & Shrivastava, 2020).

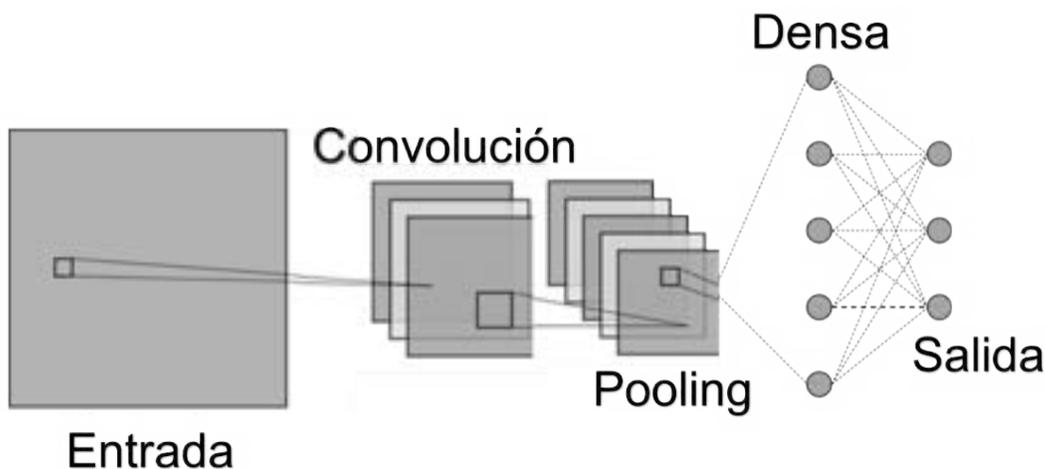


Figura 2
Red básica de la arquitectura CNN
Fuente: Elaboración propia.

Modelo de red propuesto

Primero comenzaremos tomando la red de Palakodati et al. (2020), que tenía como objetivo clasificar frutas frescas y podridas, donde el modelo alcanzó una precisión de 97.82%, el cual superó los otros métodos de aprendizaje por transferencia y técnicas convencionales. La red se inicia con una estructura secuencial, lo que significa que las capas se apilan una tras otra en un flujo lineal.

La primera capa es una capa convolucional que utiliza 32 filtros de tamaño 3x3. Estos filtros $\mathbf{K}(\cdot, \cdot)$ se aplican a la imagen de entrada $\mathbf{I}(\cdot, \cdot)$, que se espera que tenga un tamaño de 128x128 píxeles y tres canales de color (RGB).

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - n)$$

Donde S es el tensor de salida. La capa utiliza un relleno ('padding') del tipo 'same', lo que significa que la salida de la capa tendrá las mismas dimensiones que la entrada. A continuación, se aplica una capa de normalización por lotes, que ayuda a mejorar la estabilidad y el rendimiento de la red. La activación ReLU (Rectified Linear Unit) se utiliza después de la normalización, proporcionando no linealidad a la red. Finalmente, se aplica una capa de pooling máximo con un tamaño de 2x2, que reduce las dimensiones espaciales de la salida a la mitad.

La segunda y tercera capas convolucionales siguen un patrón similar. La segunda capa utiliza 64 filtros de tamaño 3x3, mientras que la tercera utiliza 128 filtros del mismo tamaño. Ambas capas también emplean normalización por lotes, activación ReLU y pooling máximo.

Después de las capas convolucionales, la red aplanar la salida en un vector unidimensional, lo que facilita su conexión a las capas densas o completamente conectadas. La siguiente, es una capa densa con 512 neuronas, seguida de una activación ReLU. Para prevenir el sobreajuste y mejorar la generalización del modelo, se introduce un dropout del 50%. Esto significa que, durante el entrenamiento, el 50% de las neuronas en esta capa se "apagarán" aleatoriamente en cada paso.

Finalmente, la red culmina en una capa densa con dos neuronas, que corresponde a las dos clases de clasificación: banana fresca y banana dañada. La activación softmax se utiliza en esta capa para convertir las salidas en probabilidades, asegurando que sumen 1. Esta capa proporciona la probabilidad de que una imagen dada pertenezca a una de las dos clases. Cabe recalcar que se realizaron unas modificaciones en la red.

La segunda red que consideramos fue una con conexión residual que se describirán a continuación:

El modelo presentado es una Red Neuronal Convolucional (CNN) diseñada para procesar imágenes de tamaño 128x128 con tres canales de color (RGB) comienza con una capa convolucional que utiliza 32 filtros de 3x3, seguida de una normalización por lotes y una activación ReLU. Luego, se aplica un pooling máximo para reducir las dimensiones espaciales.

Después de la primera capa convolucional, se introduce un concepto interesante: las conexiones residuales. Estas conexiones permiten que las activaciones de una capa anterior se sumen directamente a las de una capa posterior, facilitando el entrenamiento de redes más profundas al abordar el problema del desvanecimiento del gradiente.

Antes de la segunda capa convolucional, se crea una conexión residual que proyecta linealmente la salida de la primera capa a una profundidad de 64 filtros usando una convolución 1x1. La segunda capa convolucional también utiliza 64 filtros de 3x3, seguida de normalización por lotes y activación ReLU. La salida de esta capa se suma con la conexión residual antes de aplicar otro pooling máximo.

$$H(x) = F(x) + x$$

Un proceso similar ocurre antes de la tercera capa convolucional, donde se crea otra conexión residual que proyecta la salida a una profundidad de 128 filtros. La tercera capa convolucional utiliza 128 filtros de 3x3, seguida nuevamente de normalización por lotes y activación ReLU. La salida se suma con la conexión residual y se aplica un pooling máximo.

Finalmente, la red aplanada la salida y la pasa a través de una capa densa con 512 neuronas y activación ReLU. Se introduce un dropout del 50% para regularización. La red culmina en una capa densa con dos neuronas, que se supone corresponden a dos clases de clasificación, y utiliza una activación softmax para producir probabilidades de clase. El código completo se encuentra en:

https://colab.research.google.com/drive/1jnFS_ZNtOtGoXLxEplS2dtVfBTACpLqj?usp=sharing

Resultados

En la evaluación de los modelos propuestos, se observaron diferencias notables en su rendimiento. El primer modelo, una Red Neuronal Convolucional (CNN) estándar, demostró un rendimiento excepcional con una precisión (Accuracy) del 99.78%. Además, alcanzó una precisión (Precision) del 99.81%, un recall (Recall) del 99.81% y una puntuación F1 (F1-score) del 99.81%. La matriz de confusión para este modelo reveló que clasificó correctamente 380 instancias de la primera clase y 529 de la segunda, cometiendo solo un error en cada clase, ver figura 1 y figura 2.

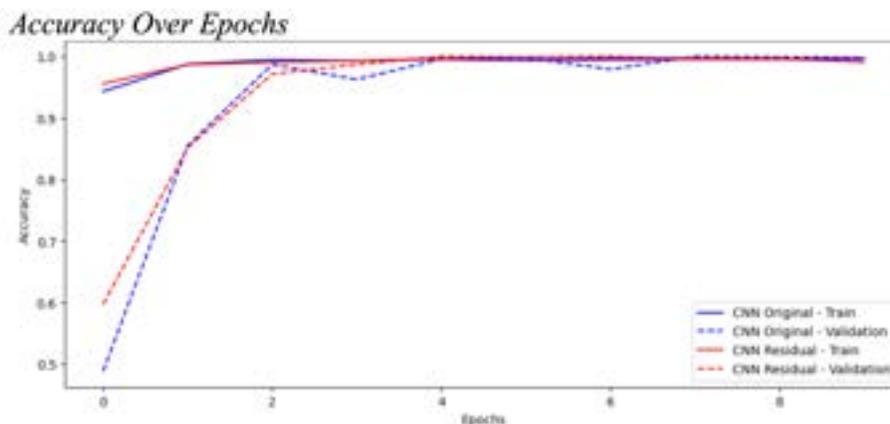


Figura 3

Por otro lado, el segundo modelo, que incorpora conexiones residuales, obtuvo una precisión del 99.23%, ver figura 1 y figura 2. Aunque sigue siendo alto, es ligeramente inferior al primer modelo. La precisión, recall y puntuación F1 para este modelo fueron del 99.62%, 99.06% y 99.34% respectivamente. La matriz de confusión mostró que clasificó correctamente 379 instancias de la primera clase y 525 de la segunda, pero cometió 2 errores en la primera clase y 5 en la segunda.

Loss Over Epochs

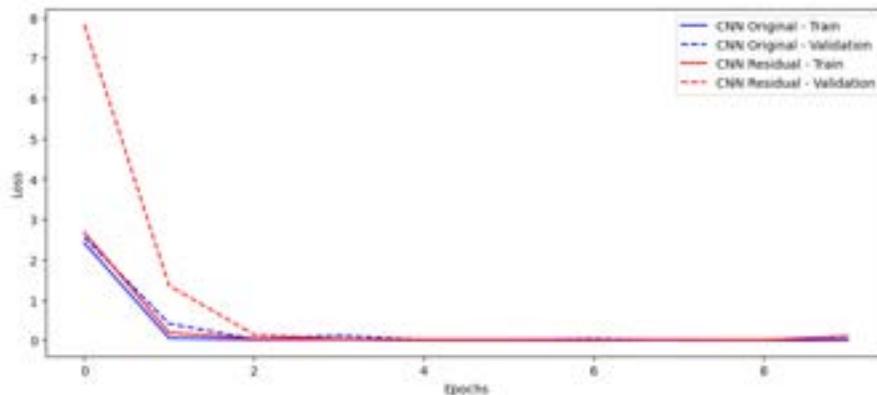


Figura 4

Discusión

La clasificación precisa de bananas en la industria agrícola es esencial para garantizar la calidad del producto y la eficiencia económica. En este estudio y para esta tarea, se evaluaron dos arquitecturas de Redes Neuronales Convolucionales (CNN). El primer modelo, una CNN estándar, alcanzó una impresionante precisión cercana al 99.78%. Estos resultados son coherentes con investigaciones anteriores, como la de Palakodati et al. (2020), que también lograron alta precisión en la clasificación de frutas con aprendizaje profundo. Sin embargo, la matriz de confusión reveló que el modelo es robusto y confiable, con errores mínimos.

El segundo modelo, con conexiones residuales, mostró una ligera disminución en la precisión. Aunque las conexiones residuales han sido efectivas en otros contextos, en este caso, no ofrecieron una mejora significativa. A pesar de esta disminución, sigue siendo un modelo altamente efectivo, y ambos modelos superan a muchos métodos tradicionales.

Conclusiones

Este estudio subraya el potencial y la eficacia de las Redes Neuronales Convolucionales en la clasificación de frutas, particularmente bananas. Aunque el modelo estándar superó al modelo con conexiones residuales, ambos modelos presentaron un rendimiento excepcional, superando a muchos enfoques tradicionales. Estos hallazgos reafirman la relevancia del aprendizaje profundo en la industria agrícola, ofreciendo soluciones automatizadas y precisas para garantizar la calidad del producto. Las futuras investigaciones podrían centrarse en combinar diferentes arquitecturas y técnicas para optimizar aún más la clasificación y abordar otros desafíos en el sector.

Referencias Bibliográficas

- Bera, S., & Shrivastava, V. K. (2020). Effect of pooling strategy on convolutional neural network for classification of hyperspectral remote sensing images. *IET Image Processing*.
- Dhande, D. V., & Patil, D. D. (2022). A Deep Learning based Model for Fruit Grading using DenseNet. *Rev. d'Intelligence Artif.*
- Kaggle (s.f.) (<https://www.kaggle.com/datasets/sriramr/fruits-fresh-and-rotten-for-classification>)
- Lingamgunta, A., Prasad, M. H. C., Sai, K. V., & Jayaraj, M. (2023). Fruit Ripeness Detection Using Deep Learning.
- Palakodati, S. S. S., Chirra, V. R. R., Yakobu, D., & Bulla, S. (2020). Fresh and Rotten Fruits Classification Using CNN and Transfer Learning. *Rev. d'Intelligence Artif.*, 34(5), 617-622.
- Sumari, P., Ahmad, W. M. A. W., Hadi, F., Mazlan, M., Liyana, N. A., Bello, R., Mohamed, A., & Talib, A. (2021). A Precision Agricultural Application: Manggis Fruit Classification Using Hybrid Deep Learning.
- Wang, J., Song, X., Sun, L., Huang, W., & Wang, J. (2020). A Novel Cubic Convolutional Neural Network for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Yao, H., Zhang, X., Zhou, X., & Liu, S. (2019). Parallel Structure Deep Neural Network Using CNN and RNN with an Attention Mechanism for Breast Cancer Histology Image Classification. *Cancers*.
- Zhang, Y., Satapathy, S., Liu, S., & Li, G.-R. (2020). A five-layer deep convolutional neural network with stochastic pooling for chest CT-based COVID-19 diagnosis. *Machine Vision and Applications*.