Comparación del rendimiento de los modelos preentrenados InceptionV3 y VGG16 para la clasificación de frutas y verduras

Jordan Hernández Daza

Universidad de Antioquia
Facultad de Ciencias Exactas y Naturales
Instituto de Física
Medellín, Colombia
2023

Resumen

En el ámbito del comercio minorista y mayorista, la administración eficiente de un amplio inventario de productos, especialmente frutas y verduras, presenta un desafío significativo. Durante nuestras experiencias de compra, es probable que hayamos experimentado y nos hayamos sorprendido con las ventajas que ofrecen las soluciones de autoservicio. Estas soluciones han sido un salvavidas para muchas empresas, evitando el colapso causado por el alto volumen de compradores, y sobre todo, han beneficiado a los clientes al reducir las esperas prolongadas y las filas largas, incluso para comprar solo unos pocos productos. En este contexto, se ha implementado un modelo de visión artificial con el objetivo de mejorar el flujo y la eficiencia en la venta de frutas y verduras. Adicionalmente, cuando los operadores responsables del registro de productos se enfrentan a alimentos exóticos o visualmente similares, a menudo encuentran dificultades debido a su falta de conocimiento sobre los nombres y códigos correspondientes. Esta falta de información tiene un impacto negativo en el flujo de trabajo y la eficiencia del proceso.

Para resolver este problema, se ha desarrollado un modelo de Python utilizando una red neuronal preentrenada y un conjunto de datos de entrenamiento que consta de aproximadamente 50,000 imágenes. El modelo ha logrado una precisión de entrenamiento de 0.9935 y una precisión de validación de 0.9697 utilizando Inception. Además, ha obtenido una precisión de entrenamiento de 0.9990 y una precisión de validación de 0.9781 utilizando VGG16. Estos resultados respaldan la capacidad del sistema de visión artificial para distinguir y clasificar frutas y verduras de manera precisa. Mediante la implementación de este sistema, se espera mejorar la gestión del inventario y optimizar el proceso de registro de productos. Además, se busca agilizar la experiencia de compra tanto para los operadores como para los clientes en el comercio minorista y mayorista.

Palabras clave: CNN (Redes neuronales convolucionales), Visión artificial, Accuracy, Función de pérdida, Precisión, Frutas, Vegetales, VGG16, InceptionV3.

Abstract

In the retail and wholesale sector, efficient management of a large inventory of products, especially fruits and vegetables, presents a significant challenge. During our shopping experiences, we have likely experienced and been impressed by the advantages offered by self-service solutions. These solutions have been a lifesaver for many businesses, preventing collapse caused by high volumes of shoppers, and, above all, benefiting customers by reducing long waits and queues, even when purchasing only a few items. In this context, an artificial vision model has been implemented with the aim of improving the flow and efficiency in the sale of fruits and vegetables. Additionally, when operators responsible for product scanning encounter exotic or visually similar foods, they often face difficulties due to their lack of knowledge about the corresponding names and codes. This lack of information has a negative impact on workflow and process efficiency. To address this problem, a Python model has been developed using a pretrained neural network and a training dataset consisting of approximately 50,000 images. The model has achieved a training accuracy of 0.9935 and a validation accuracy of 0.9697 using Inception. Additionally, it has achieved a training accuracy of 0.9990 and a validation accuracy of 0.9781 using VGG16. These results support the artificial vision system's ability to accurately distinguish and classify fruits and

vegetables. By implementing this system, it is expected to improve inventory management and optimize the product scanning process. Furthermore, the aim is to streamline the shopping experience for both operators and customers in the retail and wholesale trade.

Keywords: CNN (Convolutional Neural Networks), Artificial vision, Accuracy, loss function, Precision, Retail, Fruits, Vegetables, VGG16, Inception V3.

Comparación del rendimiento de los modelos preentrenados InceptionV3 y VGG16

Jordan Hernández *

7 de junio de 2024

Resumen

En el ámbito del comercio minorista y mayorista, la administración eficiente de un amplio inventario de productos, especialmente frutas y verduras, presenta un desafío significativo. Durante nuestras experiencias de compra, es probable que hayamos experimentado y nos hayamos sorprendido con las ventajas que ofrecen las soluciones de autoservicio. Estas soluciones han sido un salvavidas para muchas empresas, evitando el colapso causado por el alto volumen de compradores, y sobre todo, han beneficiado a los clientes al reducir las esperas prolongadas y las filas largas, incluso para comprar solo unos pocos productos. En este contexto, se ha implementado un modelo de visión artificial con el objetivo de mejorar el flujo y la eficiencia en la venta de frutas y verduras. Cuando los operadores responsables del registro de productos se enfrentan a alimentos exóticos o visualmente similares, a menudo encuentran dificultades debido a su falta de conocimiento sobre los nombres y códigos correspondientes. Esta falta de información tiene un impacto negativo en el flujo de trabajo y la eficiencia del proceso. Para resolver este problema, se ha desarrollado un modelo de Python utilizando una red neuronal preentrenada y un conjunto de datos de entrenamiento que consta de aproximadamente 50,000 imágenes. El modelo ha logrado una precisión de entrenamiento de 0.9935 y una precisión de validación de 0.9697 utilizando Inception. Además, ha obtenido una precisión de entrenamiento de 0.9990 y una precisión de validación de 0.9781 utilizando VGG16. Estos resultados respaldan la capacidad del sistema de visión artificial para distinguir y clasificar frutas y verduras de manera precisa. Mediante la implementación de este sistema, se espera mejorar la gestión del inventario y optimizar el proceso de registro de productos. Además, se busca agilizar la experiencia de compra tanto para los operadores como para los clientes en el comercio minorista y mayorista.

Palabras clave: CNN (Redes neuronales convolucionales), Visión artificial, Accuracy, Función de pérdida, Precisión, Frutas, Vegetales, VGG16, InceptionV3.

*E-mail: jordan.hernandez@udea.edu.co, Instituto de Física, Universidad de Antioquia, Medellín, Colombia.

Contenido

1. Introducción

En el mundo del *retail*, la gestión eficiente del inventario es de suma importancia. La capacidad de identificar y clasificar con precisión los productos, especialmente frutas y verduras, resulta crucial para garantizar operaciones fluidas y mejorar la experiencia del cliente. Sin embargo, los métodos tradicionales de identificación manual a menudo presentan limitaciones, lo que conlleva errores, retrasos e ineficiencias. Para superar estos desafíos, investigadores y profesionales de la industria han dirigido su atención hacia el poder de las redes neuronales y los algoritmos de visión artificial, aprovechando sus capacidades en reconocimiento de imágenes para mejorar los procesos del *retail*.

El campo del reconocimiento de imágenes, impulsado por los avances en el aprendizaje profundo y las arquitecturas de redes neuronales, ha experimentado un progreso significativo en los últimos años. Estas tecnologías, combinadas con la creciente disponibilidad de conjuntos de datos de entrenamiento y recursos computacionales, han allanado el camino para el desarrollo de sistemas de visión artificial. Estos sistemas tienen el potencial de identificar y categorizar de manera autónoma diversos productos del *retail*, dotando a las tiendas de una mayor precisión, rapidez y eficiencia en la gestión de inventarios.

En este artículo exploramos la intersección de las redes neuronales, la visión artificial y la industria del *retail*, centrándonos en la aplicación de técnicas de reconocimiento de imágenes para la identificación de productos. A través de estudios recientes (1), analizamos los avances logrados en este campo, destacando los beneficios de aprovechar las redes neuronales y los modelos de aprendizaje profundo en el reconocimiento de imágenes en el *retail*. Al aprovechar el potencial de estas tecnologías, las tiendas pueden lograr una identificación de productos fluida, optimizar la gestión de inventarios y, en última instancia, mejorar la experiencia general del *retail* tanto para las empresas como para los consumidores.

El modelo InceptionV3 es conocido por su eficiencia y precisión en tareas de clasificación de imágenes. Se caracteriza por su arquitectura de múltiples ramas y convoluciones factorizadas, que le permiten capturar características de manera más efectiva. En este estudio, se utilizará el modelo pre entrenado InceptionV3 como punto de partida.

El modelo VGG16 es otro modelo pre entrenado ampliamente utilizado en la comunidad de aprendizaje profundo. Se destaca por su arquitectura profunda y su simplicidad en términos de diseño. Aunque puede requerir más recursos computacionales, se espera que tenga un rendimiento competitivo en la clasificación de frutas y verduras.

2. Marco Teórico

2.1. Redes Neuronales Convolucionales (CNN)

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son un tipo de arquitectura de redes neuronales profundas que se utilizan ampliamente en el campo del procesamiento de imágenes y la visión por computadora. Estas redes han demostrado un gran éxito en tareas de clasificación, detección y segmentación de objetos en imágenes. [1]

Las redes convolucionales se inspiran en el funcionamiento del sistema visual humano, donde las neuronas en el cerebro son sensibles a regiones específicas del campo visual. Las CNN se basan en la idea de que ciertos patrones visuales importantes pueden ser detectados localmente en una imagen mediante el uso de filtros convolucionales. [2]

A diferencia de las redes neuronales tradicionales, las redes convolucionales aplican filtros convolucionales a las entradas de la red en lugar de realizar conexiones totalmente conectadas. Estos filtros son pequeñas matrices de pesos que se deslizan a lo largo de la imagen de entrada, calculando productos de convolución en cada ubicación y generando mapas de características [3]

Un aspecto fundamental de las CNN es su capacidad para aprender características relevantes directamente de los datos mediante el uso de capas convolucionales y capas de *pooling* [4]. Las capas convolucionales aplican filtros convolucionales a la entrada y extraen características locales, mientras que las capas de *pooling* reducen la dimensionalidad de las características al realizar submuestreo. Estas operaciones permiten a las redes convolucionales aprender representaciones jerárquicas de alto nivel a partir de las imágenes.

Una vez que se han extraído las características, las CNN suelen incluir capas totalmente conectadas al final de la red para realizar la clasificación. Estas capas toman las características aprendidas y las utilizan para asignar probabilidades a las diferentes clases de salida.

En el contexto de la clasificación de frutas y verduras, las redes convolucionales han demostrado ser especialmente efectivas. La capacidad de las CNN para capturar patrones visuales y aprender representaciones ricas de las imágenes les permite diferenciar entre diversos tipos de frutas y verduras en función de sus características visuales distintivas, como la forma, el color y la textura. [5]

2.2. Visión Artificial

La visión artificial es un campo de estudio que se enfoca en desarrollar algoritmos y sistemas que permiten a las máquinas comprender, interpretar y analizar imágenes o vídeos de manera similar a como lo hacen los seres humanos. Se basa en el uso de técnicas de procesamiento de imágenes, aprendizaje automático y redes neuronales para extraer características y realizar tareas específicas, como clasificación, detección, reconocimiento y segmentación de objetos.

2.3. Inception

Los módulos de Inception son un componente clave en las Redes Neuronales Convolucionales (CNN) que permiten un cálculo más eficiente y redes más profundas al reducir la dimensionalidad con convoluciones apiladas de 1×1. Estos módulos fueron diseñados para abordar desafíos como el gasto computacional y el sobreajuste en las redes neuronales convolucionales. La idea principal es tomar múltiples tamaños de filtros de kernel dentro de la CNN y, en lugar de apilarlos secuencialmente, ordenarlos para

que operen en el mismo nivel [6].

InceptionV3 es una variante conocida de los módulos Inception y es una arquitectura de red neuronal convolucional profunda. Fue introducida por Szegedy et al. y ha demostrado un rendimiento sobresaliente en tareas de clasificación de imágenes. InceptionV3 consta de varios bloques de módulos Inception apilados, cada uno diseñado para extraer y combinar características de diferentes tamaños de filtros [7].

La arquitectura de InceptionV3 se puede describir paso a paso de la siguiente manera:

- **Capas iniciales:** La red comienza con capas convolucionales tradicionales para extraer características de nivel bajo de la imagen de entrada.
- **Módulos Inception:** Estos módulos son el núcleo de la arquitectura y están diseñados para capturar características en diferentes escalas espaciales mediante el uso de convoluciones de tamaño reducido en paralelo. Cada módulo Inception se compone de múltiples rutas o ramas paralelas de convoluciones con diferentes tamaños de filtros (1×1, 3×3, 5×5). Estas convoluciones capturan características en diferentes niveles de detalle y permiten a la red aprender representaciones más ricas y discriminativas.
- **Capas iniciales:** Capas de agrupación (pooling): Después de cada módulo Inception, se aplican capas de agrupación (pooling) para reducir la dimensionalidad espacial de las características extraídas y mejorar la eficiencia computacional. El *max pooling* es una operación comúnmente utilizada en InceptionV3.
- **Capas finales:** Una vez que se han extraído las características a través de los módulos Inception y las capas de agrupación, se aplican capas totalmente conectadas para realizar la clasificación final. Estas capas toman las características aprendidas y las utilizan para asignar probabilidades a las diferentes clases de salida.

La infraestructura de InceptionV3 es más profunda y compleja que la del modelo original de Inception (GoogleNet o Inception1). Además de las convoluciones en paralelo, InceptionV3 incorpora técnicas como regularización L2 y capas de normalización por lotes para controlar el sobreajuste y acelerar el entrenamiento [7].

La arquitectura InceptionV3 ha demostrado mejoras significativas en el rendimiento al abordar problemas como el gradiente desvaneciente y la extracción de características más ricas en imágenes.

2.4. VGG16

VGG16 es un tipo de red neuronal convolucional (CNN) que se ha destacado en el campo de la visión por computadora. Fue desarrollado por Simonyan y Zisserman en 2014 y se considera uno de los modelos más influyentes y exitosos en esta área.

La arquitectura de VGG16 se caracteriza por su profundidad y simplicidad. Está compuesta por un total de 16 capas, que incluyen capas convolucionales, capas de agrupación y capas totalmente conectadas. Estas capas se organizan en bloques y se numeran de acuerdo con su posición en la red.

La estructura general de VGG16 se puede describir de la siguiente manera:

- **Capas de convolución:** VGG16 comienza con varias capas de convolución, cada una seguida de una función de activación no lineal, como ReLU (Rectified Linear Unit). Las convoluciones se realizan con filtros pequeños de tamaño 3×3 y un paso (stride) de 1 píxel. Estas capas convolucionales se encargan de extraer características de bajo nivel de la imagen de entrada.

- **Capas de agrupación (pooling)** : Después de un número determinado de capas convolucionales, se aplican capas de agrupación (pooling) para reducir la dimensionalidad espacial de las características extraídas. VGG16 utiliza capas de agrupación de máximo (*max pooling*) con filtros de tamaño 2×2 y un paso de 2 píxeles. Esta operación ayuda a preservar las características más relevantes mientras reduce la cantidad de parámetros.

- **Capas totalmente conectadas:** Una vez que se han extraído las características y se ha reducido la dimensionalidad, se pasan a través de varias capas totalmente conectadas. Estas capas están diseñadas para realizar la clasificación final y asignar probabilidades a las clases de salida. La última capa totalmente conectada utiliza la función de activación Softmax para generar una distribución de probabilidad sobre las clases de salida.

La arquitectura VGG16 es conocida por su enfoque de convoluciones más profundas en comparación con otros modelos anteriores. Aunque esta profundidad añade más parámetros y requiere más recursos computacionales, también permite capturar características más complejas y sutiles de las imágenes.

VGG16 ha demostrado su efectividad en diversas tareas de visión por computadora, como clasificación de imágenes y detección de objetos. Es ampliamente utilizado como base para aplicaciones de *transfer learning* debido a su capacidad de aprendizaje y generalización [8].

3. Metodología

El conjunto de datos utilizado para la comparación consta de aproximadamente 50.000 imágenes de 139 clases diferentes de frutas y verduras. Se dividirá en conjuntos de entrenamiento(70 %), validación(15 %) y prueba(15 %).

En la Figura 1 se puede apreciar la distribución del número de imágenes por clase. Se aplicará *transfer learning* a ambos modelos, utilizando los pesos preentrenados en conjuntos de datos más grandes. Para cada modelo, se ajustarán las capas finales para adaptarlas a las 139 clases específicas de frutas y verduras. Para la evaluación del rendimiento, se utilizarán métricas como el *accuracy* y la matriz de confusión. También se utilizaran las curvas de precisión vs recuperación. Se compararán los resultados obtenidos por InceptionV3 y VGG16 en el conjunto de datos de prueba.

Dado que tanto InceptionV3 como VGG16 son modelos preentrenados de alta calidad, se espera que ambos obtengan resultados prometedores en la clasificación de frutas y verduras. Sin embargo, debido a las diferencias en su arquitectura y capacidad para capturar características visuales, es posible que uno de los modelos muestre un rendimiento ligeramente superior.

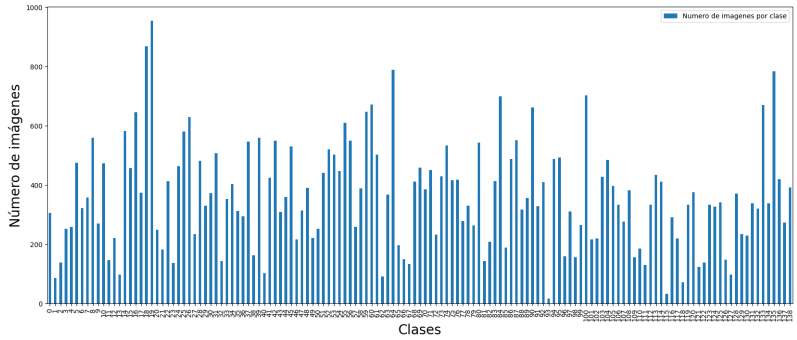

Figura 1: Número de imágenes por clase

- **Problemas de memoria:** Debido a la naturaleza del problema, para este caso abordar un problema de clasificación múltiple contar con suficientes datos aumenta el desempeño y resultado del modelo. Sin embargo, los equipos locales no tienen la capacidad de memoria para procesar el número de imágenes a utilizar. Lo que resulta en un error en el compilador, porque no es posible cargar los datos, esto representa una limitante a la hora de hacer pruebas representativas para análisis. Como primera aproximación se tomaron muestras aleatorias de los datos totales para la realización de pruebas y desempeño de los modelos preentrenados en la maquina local.

  La compañía proporciona acceso a máquinas virtuales, con la capacidad de cómputo necesaria para las pruebas con la totalidad de los datos. Como parte de mi pasantía, he tenido la oportunidad de utilizar estas máquinas virtuales para realizar pruebas y experimentos con un conjunto de datos considerable, que consta de aproximadamente 50,000 imágenes pertenecientes a 139 clases diferentes.

  Grupo Éxito, como empresa líder en el *retail*, está comprometido con la adopción de tecnologías innovadoras para mejorar la eficiencia operativa y brindar una mejor experiencia al cliente. Mi pasantía en esta organización me ha brindado la oportunidad de aplicar técnicas de reconocimiento de imágenes y aprendizaje automático para abordar desafíos específicos relacionados con la clasificación de productos en el ámbito del *retail*.

- **Revisión de imágenes repetidas:** Haciendo una inspección manual de las imágenes, se encontraron imágenes repetidas. Se procedió a revisar otras categorías, con el fin de encontrar las totalidad de imágenes repetidas... en esta inspección se encontró que las imágenes repetidas tenían el sufijo (2) en su nombre. Bajo esta premisa, se hizo una búsqueda con Python de todos los *blob* (objetos de Google Cloud Storage) que cumplieran esta característica. Sin embargo, se notó que algunas imágenes con el sufijo, correspondía a imágenes no repetidas, por lo que no fue posible usar esto como criterio. Se procedió a usar el método de *Hashes* que asigna un código único a cada objeto, de esta manera se logró identificar todas las imágenes repetidas.

3.1. Planteamiento del problema

En el ámbito del *retail* y el comercio mayorista, la gestión eficiente de un amplio inventario de productos se convierte en un desafío, especialmente cuando se trata de frutas y verduras. Los operadores encargados del registro de productos a menudo se enfrentan a dificultades al encontrarse con alimentos exóticos o visualmente similares, debido a su falta de conocimiento sobre los nombres y códigos correspondientes. Estas dificultades pueden generar un impacto negativo en el flujo de trabajo y la eficiencia del proceso.

La identificación y clasificación precisa de frutas y verduras es crucial para garantizar operaciones fluidas y mejorar la experiencia del cliente en el *retail*. Sin embargo, los métodos tradicionales de identificación manual presentan limitaciones, lo que resulta en errores, retrasos e ineficiencias. Por lo tanto, surge la necesidad de abordar esta problemática y mejorar los procesos de identificación en el manejo de inventarios de frutas y verduras.

Para superar estos desafíos, se propone la implementación de un sistema de visión artificial basado en redes neuronales, capaz de distinguir de manera individual las frutas y verduras. El objetivo es utilizar el poder del aprendizaje profundo y los algoritmos de reconocimiento de imágenes para agilizar y mejorar la precisión en la identificación de estos productos en el retail.

La relevancia de este problema radica en la necesidad de optimizar la gestión de inventarios, reducir errores y mejorar la eficiencia en el manejo de frutas y verduras en el ámbito del retail. Al contar con un sistema de visión artificial que pueda identificar y clasificar automáticamente estos productos, se espera agilizar los procesos de registro, reducir tiempos y minimizar los errores asociados a la identificación manual.

Además, este sistema de visión artificial tiene el potencial de mejorar la experiencia del cliente al garantizar que los productos sean adecuadamente registrados y clasificados, lo que a su vez permite una gestión de inventarios más precisa y eficiente.

4. Experimento

4.1. Resultados

En las Figuras 2 y 3 se puede apreciar que los modelo de Inceptionv3 y VGG16 alcanzan valores considerablemente buenos de accuracy, este representa la proporción de muestras clasificadas correctamente sobre el total de muestras. Además sus curvas de función de pérdida tienden a disminuir, por lo cual se puede descartar tanto el sobreajuste como el subajuste.

En primer lugar, se realiza un análisis del desbalance de datos existente entre las 139 clases diferentes. Es importante tener en cuenta esta disparidad al momento de evaluar los resultados y seleccionar las métricas adecuadas para medir el desempeño de los modelos. En la gráfica 1, se puede observar una variación en el número de imágenes por clase, pero no se evidencia un desequilibrio significativo que pueda afectar de manera considerable el proceso de clasificación.

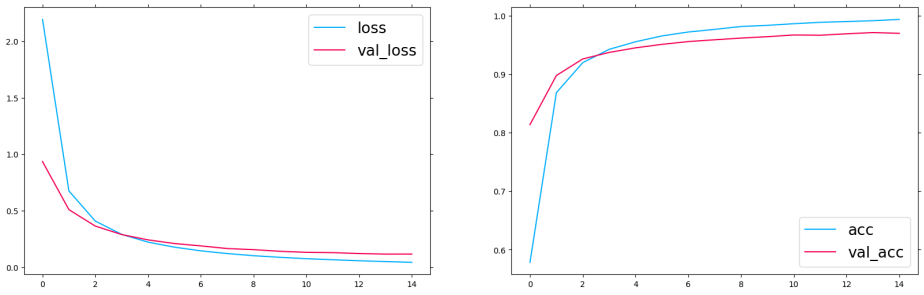

Figura 2: Número de imágenes por clase

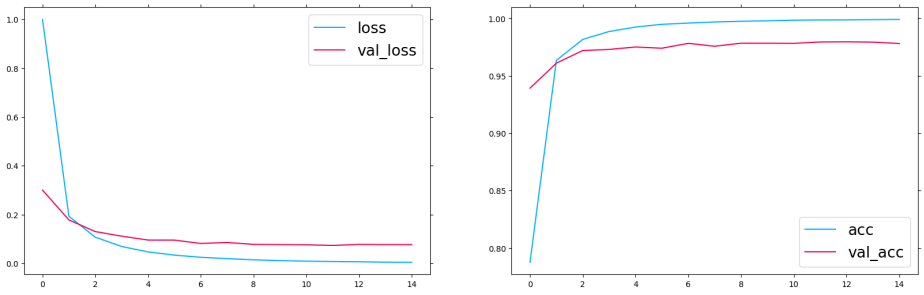

Figura 3: Número de imágenes por clase

El análisis de las métricas Precision, Recall, y F1-score para el modelo inceptionv3 y VGG16, es el siguiente:

- **P**recisión (*precision*): La precisión indica la proporción de instancias clasificadas correctamente en relación con el total de instancias clasificadas para una clase en particular. En este caso, los modelos Inceptionv3 y VGG16 muestra una alta precisión para la mayoría de las clases, con valores cercanos a 1. Esto indica que los modelos tiene una baja tasa de falsos positivos y puede distinguir con precisión las instancias de cada clase.
- **S**ensibilidad (*recall*): El *recall*, también conocido como sensibilidad o tasa de verdaderos positivos, indica la proporción de instancias positivas que se clasificaron correctamente. En este caso, tanto

el modelo VGG16 como Inceptionv3 muestran un alto recall para la mayoría de las clases, con valores cercanos a 1. Esto significa que los modelos puede recuperar la mayoría de las instancias positivas de cada clase.

- F1-score: *F1-score* es una medida que combina la precisión y el *recall* en un solo valor. Es útil cuando se desea tener en cuenta tanto la precisión como el *recall* de un modelo. En este caso, ambos modelos muestran un *F1-score* alto para la mayoría de las clases. Esto indica un buen equilibrio entre la precisión y el *recall* en la clasificación de las instancias. Por lo tanto hay pocos falsos positivos y falsos negativos.
- Support: El *support* indica el número de instancias en el conjunto de prueba que pertenecen a cada clase. Puede ser útil para identificar clases con un número de instancias pequeño, donde las métricas pueden ser menos confiables debido a la falta de datos.

Para las métricas globales que se pueden ver en la Figura 4 el modelo VGG16 parece tener una puntuación ligeramente más alta en promedio para todas las clases, lo que puede sugerir un mejor rendimiento en general. Por ejemplo, para el *accuracy* se obtiene un valor de 0.964 para el modelo Inceptionv3 y 0.980 para el modelo VGG16.

El macro promedio de cada métrica se calcula promediando los valores *recall* de precisión, y *F1-score* de todas las clases. El macro promedio de la precisión para el modelo VGG16 es de 0.980 y de 0.965 para Inceptionv3. Se alcanzan valores aproximados para el resto de métricas que se están evaluando.

Esta métrica global es importante ya que hay una penalización cuando el modelo no funciona bien con las clases minoritarias (que es exactamente lo que quiere cuando hay desequilibrio).

En general, se usan micropromedios para ponderar su métrica hacia el conjunto de datos más grande. y se usa el promedio macro para ponderar su métrica hacia los conjunto de datos más pequeños. (En *AutoML* de Google Cloud se utilizan *micro-average precision*)

	precision_inception	recall_inception	f1-score_inception	precision_VGG16	recall_VGG16	f1-score_VGG16	support_T
0	1.000	1.000	1.000	1.000	1.000	1.000	49.00
1	1.000	1.000	1.000	1.000	1.000	1.000	12.00
2	1.000	1.000	1.000	1.000	1.000	1.000	12.00
3	0.828	0.828	0.828	0.906	1.000	0.951	29.00
4	1.000	0.935	0.967	1.000	0.935	0.967	31.00
...	...	...	...	...	...	...	...
137	1.000	0.964	0.982	1.000	0.964	0.982	28.00
138	1.000	0.983	0.992	1.000	1.000	1.000	60.00
accuracy	0.964	0.964	0.964	0.980	0.980	0.980	0.98
macro avg	0.968	0.963	0.965	0.980	0.979	0.979	7538.00
weighted avg	0.966	0.964	0.964	0.980	0.980	0.980	7538.00

Figura 4: Métricas globales

En las Figuras 5 y 6 se presentan las matrices de confusión normalizadas para los dos modelos. Al analizar las matrices, se puede observar que ambos modelos muestran buenos resultados en todas las clases, ya que hay bajos números de falsos negativos y falsos positivos.

Esto significa que los modelos son capaces de clasificar correctamente la mayoría de las muestras en cada clase, minimizando los errores de clasificación tanto en la dirección de falsos negativos (clasificar incorrectamente una muestra positiva como negativa) como en la dirección de falsos positivos (clasificar incorrectamente una muestra negativa como positiva).

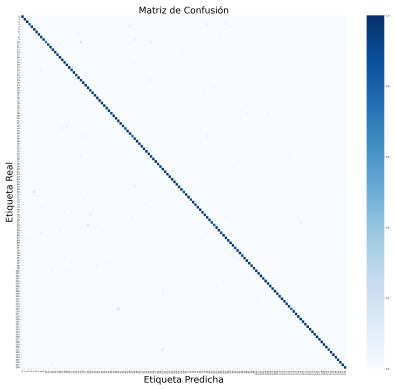

Figura 5: Matriz de confusión Inceptionv3

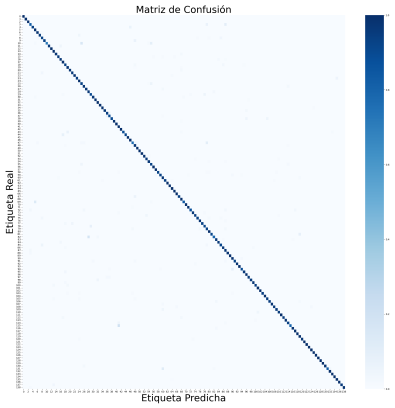

Figura 6: Matriz de confusion de VGG16

Estos resultados indican que los modelos Inceptionv3 y VGG16 son efectivos en la tarea de clasificación y muestran un buen rendimiento en la mayoría de las clases evaluadas.

La curva de *precisión-recall* es una representación gráfica que muestra cómo cambia la precisión y el recall de un modelo a medida que se varía el umbral de clasificación. A medida que se aumenta el umbral de clasificación, es decir, se vuelve más estricto, la precisión tiende a aumentar mientras que el recall disminuye.

Esto se debe a que se clasifican menos muestras como positivas, pero la proporción de clasificaciones correctas entre las positivas aumenta. En otras palabras, aumenta el número de falsos negativos y disminuye el número de falsos positivos.

En general, una buena curva de precisión-recall se caracteriza por tener una alta precisión y un alto recall en valores de umbral que sean relevantes para la aplicación en cuestión. Una curva ideal sería aquella que se acerca al punto (1,1), lo que significa una precisión y recall perfectos.

En las Figuras 7 y 8 se pueden apreciar las curvas de precisión-recall de los modelos utilizados. Al analizar estas curvas, se puede concluir que las curvas correspondientes a VGG16 presentan mejores valores de precisión y recall en comparación con Inceptionv3 para diferentes umbrales de clasificación. Esto se evidencia en el área bajo la curva (AUC) de todas las curvas de las diferentes clases, que es más grande para VGG16.

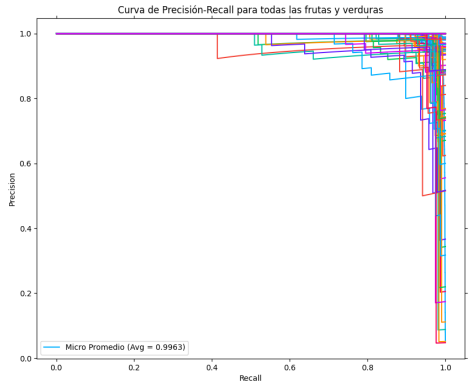

Figura 7: Curvas de *Precision-Recall* VGG16

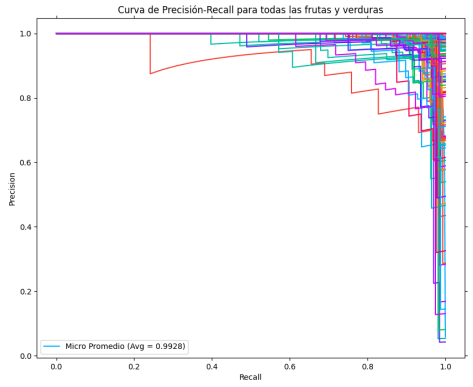

Figura 8: Curvas de *Precision-Recall* Inceptionv3

Al considerar los micro promedios, se obtiene una medida global del rendimiento de los modelos. Los micro promedios son calculados tomando en cuenta todas las predicciones y etiquetas de las clases por separado, y luego se calcula la precisión, el recall y el f1-score utilizando estos valores agregados. Los micro promedios son útiles para evaluar el rendimiento general del modelo sin tener en cuenta desequilibrios en la distribución de las clases. Al analizar los micro promedios, se puede tener una idea general del desempeño promedio del modelo en todas las clases

El área bajo la curva de los micro promedios, es decir la precisión promedio para VGG16 es 0.9963 y para Inceptionv3 es 0.9928.

En resumen, al considerar las curvas de *Precision-Recall* y los micropromedios, se puede concluir que el modelo VGG16 muestra un mejor rendimiento general en comparación con InceptionV3. Las curvas de *Precision-Recall* de VGG16 tienen un área bajo la curva más grande, lo que indica una mejor capacidad de clasificación en todas las clases. Además, los micro promedios permiten obtener una visión global del rendimiento del modelo en todas las clases sin verse afectados por desequilibrios de distribución.

5. Conclusiones y Recomendaciones

A partir de todas las métricas evaluadas se puede puede concluir que el modelo que muestra ligeramente un mejor rendimiento es el VGG16.

Donde se puede ver mejor la diferencia en las métricas y por lo tanto el mejor rendimiento de VGG16 son las métricas de la figura 2 donde se presentan las métricas globales (macro promedio y promedio ponderado). También en las curvas de precision vs recall se puede apreciar el mejor rendimiento de VGG16.

En resumen, el uso de técnicas de clasificación de imágenes basadas en redes neuronales, como Inceptionv3 y VGG16, ofrece una solución prometedora para la identificación y clasificación precisa de productos en el sector del *retail*. Estos avances tienen el potencial de transformar la forma en que se gestionan los inventarios, mejorando la eficiencia operativa y la experiencia del cliente en las tiendas.

Para mejorar el rendimiento de los modelos se puede aplicar el aumento de datos (*data augmentation*), sin embargo durante el entrenamiento de las redes neuronales se generan muchas imágenes y el entrenamiento es muy lento aún con el uso de GPU.

Si bien la ampliación de datos puede ser altamente beneficiosa para mejorar el rendimiento y la generalización de un modelo, también existen algunas posibles desventajas a considerar:

- **N**ecesidad de aumentar los recursos computacionales : La ampliación de datos requiere generar imágenes aumentadas adicionales durante el proceso de entrenamiento. Esto puede llevar a un aumento de los requisitos computacionales, ya que el modelo debe procesar una mayor cantidad de datos. El conjunto de datos aumentado puede consumir más espacio de almacenamiento y puede requerir más tiempo de entrenamiento, especialmente para conjuntos de datos a gran escala.
- **R**iesgos de sobreajuste: Aunque la ampliación de datos ayuda a reducir el sobreajuste, es posible aplicar demasiadas transformaciones o introducir transformaciones poco realistas. Las transformaciones excesivas o inapropiadas pueden introducir patrones artificiales o distorsiones que no son representativas de los datos del mundo real. Esto puede conducir al sobreajuste, donde el modelo se especializa en exceso en los datos aumentados y tiene un rendimiento deficiente en ejemplos del mundo real.
- **C**onsideraciones específicas del dominio: Las técnicas de ampliación de datos deben elegirse cuidadosamente según el dominio específico y las características de los datos. Algunas técnicas de ampliación pueden no ser adecuadas o efectivas para ciertos tipos de datos. Por ejemplo, ciertas transformaciones como reflexiones o rotaciones pueden no ser significativas o realistas para ciertos tipos de imágenes o modalidades de datos.

Referencias

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[3] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. http://www.deeplearningbook.org.

[4] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," 2013.

[5] H. Mureşan and M. Oltean, "Fruit recognition from images using deep learning," 2017.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014.

[7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.