MACC
Matemáticas Aplicadas y
Ciencias de la Computación

Universidad del Rosario | Escuela de Ingeniería, Ciencia y Tecnología

Proyecto Final: DigitEye
Clasificación de Imágenes

Diryon Yonith Mora Romero
Laura Valentina Gonzalez Rodriguez

Prof. Yiby Karolina Morales Pinto

Aprendizaje Automático de Máquina
Matemáticas Aplicadas y Ciencias de la Computación
Universidad del Rosario
Mayo 2023

1. Antecedentes

La clasificación de imágenes es una labor importante en el campo del aprendizaje automático y la inteligencia artificial. Teniendo diversas aplicaciones, desde la detección de objetos en la seguridad y la vigilancia, hasta la identificación de enfermedades en la medicina. Crucial en el desarrollo de tecnologías emergentes como los vehículos autónomos y los drones. Sin embargo, con el surgimiento de las redes neuronales y el deep learning, este problema ha vuelto a tomar relevancia, con mejorar cruciales gracias a estas técnicas.

El reconocimiento de imágenes es una tarea difícil debido a la gran variabilidad que existe en las imágenes del mundo real. Esto se debe a la presencia de diferentes fuentes de variabilidad, como cambios de iluminación, orientación, escala, deformación, entre otros factores. Por esta razón, el desarrollo de un clasificador de imágenes preciso y robusto sigue siendo un reto para la comunidad científica.

En 2006, la falta de conjuntos de datos etiquetados fue un obstáculo importante para mejorar los algoritmos de aprendizaje profundo en la visión por computadora. Para abordar este problema, Fei-Fei Li lideró la creación de ImageNet, el conjunto de datos más grande y popular utilizado para la clasificación de imágenes, que contiene más de 14 millones de imágenes etiquetadas en más de 20,000 categorías. Este esfuerzo fue posible gracias a la participación activa de la comunidad de investigadores. [1]

Particularmente, la arquitectura de redes neuronales convolucionales (CNN) ha sido un gran avance en la visión por computadora, debido a su capacidad para reconocer patrones en las imagenes. La red neuronal convolucional AlexNet, desarrollada en base al conjunto de datos ImageNet, fue un hito importante en este campo, ya que logró una precisión del 15,3 % en el top 5 de errores, usando más de 100 capas. AlexNet abrió el camino para la investigación en este campo y ha llevado al desarrollo de arquitecturas más complejas y efectivas en la clasificación de imágenes. [2]

Además de la clasificación de imágenes, las redes neuronales convolucionales también han tenido un gran impacto en la detección temprana de enfermedades y trastornos. Un estudio de 2018 demostró que un algoritmo de CNN puede detectar con precisión el cáncer de piel maligno con una precisión del 91 %, lo que es comparable a la precisión de los dermatólogos en la detección de cáncer de piel. [3] En otros casos, se han utilizado algoritmos de CNN en la detección de enfermedades neurológicas, como el Parkinson, a través del análisis de imágenes de resonancia magnética cerebral. [4] Estos avances muestran el enorme potencial de las redes neuronales convolucionales en la medicina y la salud, y su capacidad para ayudar en la detección temprana y el tratamiento de enfermedades.

En resumen, la clasificación de imágenes es una tarea importante y desafiante en el campo del aprendizaje automático y la inteligencia artificial, y tiene muchas aplicaciones prácticas. Las redes neuronales convolucionales se han convertido en una herramienta esencial para reconocer patrones en imágenes y mejorar la precisión y robustez de los algoritmos. La creación de ImageNet fue un hito importante para el campo, ya que proporcionó un gran conjunto de datos etiquetados para entrenar modelos de aprendizaje profundo.

Además, la capacidad de las redes neuronales convolucionales para detectar enfermedades ha demostrado su potencial para mejorar la atención médica y la salud. Con una precisión comparable a la de los profesionales médicos, los algoritmos de CNN se han utilizado en la detección temprana de enfermedades como el cáncer de piel y el Parkinson, lo que demuestra su capacidad para complementar y mejorar la precisión de los diagnósticos médicos. Se espera que los avances continuos en la investigación, sigan mejorando la precisión y eficiencia en la clasificación de imágenes, lo que tendrá un impacto importante en la vida cotidiana.

2. Definición del problema

El problema a resolver es la clasificación de imágenes en diferentes clases. El objetivo de este proyecto es desarrollar un programa que pueda identificar con precisión la categoría en una imagen dada. La motivación personal para llevar a cabo este proyecto radica en el interés por explorar técnicas de aprendizaje automático aplicadas a la visión por computadora, y en la posibilidad de contribuir a la solución de problemas prácticos mediante el desarrollo de un clasificador de imágenes preciso y robusto.

Por lo tanto, el desafío es desarrollar un modelo de aprendizaje automático que pueda generalizar a partir de los datos de entrenamiento, clasificando con precisión las imágenes nuevas. Debe ser capaz de identificar características relevantes y descartar las irrelevantes, y debe ser capaz de adaptarse a las variaciones en las imágenes dentro de cada clase.

3. Descripción de la solución

Para lograr este objetivo, se utilizará un enfoque de aprendizaje automático utilizando redes neuronales convolucionales, que han demostrado ser altamente efectivas en la clasificación de imágenes en diversos estudios. Además, se explorarán diferentes arquitecturas de CNN y se ajustarán los hiperparámetros para obtener el mejor rendimiento posible en términos de precisión de clasificación.

La estrategia propuesta es apropiada, es cuantificable y medible mediante la precisión de la clasificación. El uso de una CNN es una técnica de vanguardia en el campo de la visión por computadora y ha demostrado un alto nivel de precisión en tareas de clasificación de imágenes similares. La implementación de la CNN se realizará utilizando la biblioteca de aprendizaje profundo TensorFlow, que proporciona una plataforma de desarrollo eficiente y escalable para la implementación de redes neuronales convolucionales.

4. Datos

El conjunto de datos utilizado en el proyecto, es el conjunto de datos CIFAR-10, el cual consta de 60000 imágenes en color de 32x32 píxeles distribuidas en 10 clases, con 6000 imágenes por clase. Estas clases son: aviones (0), automóviles (1), pájaros (2), gatos (3), ciervos (4), perros (5), ranas (6), caballos (7), barcos (8) y camiones (9). De estas, 50000 imágenes se utilizarán para entrenamiento y 10000 para pruebas. El conjunto de datos está dividido en cinco lotes de entrenamiento y un lote de pruebas, cada uno con 10000 imágenes. El lote de pruebas contiene exactamente 1000 imágenes seleccionadas aleatoriamente de cada clase, mientras que los lotes de entrenamiento contienen las imágenes restantes en orden aleatorio, aunque algunos lotes de entrenamiento pueden contener más imágenes de una clase que de otra.

El conjunto de datos CIFAR-10 fue obtenido originalmente por Alex Krizhevsky, Vinod Nair y Geoffrey Hinton, del departamento de Ciencias de la Computación de la Universidad de Toronto. Este conjunto de datos ha sido utilizado en numerosos estudios de investigación en visión por computadora y aprendizaje profundo, y se considera uno de los conjuntos de datos más populares en esta área. Además, la disponibilidad de un gran número de imágenes etiquetadas y la diversidad de las clases presentes en el dataset permiten entrenar una CNN con alta precisión y obtener resultados cuantificables y medibles. [5]

El objetivo del proyecto es utilizar una red neuronal convolucional (CNN) para clasificar las imágenes del conjunto de datos CIFAR-10 en sus respectivas clases, para detectar las caracteristicas especificas de las imágenes. Se emplearán diversas técnicas para mejorar el rendimiento de la CNN, tales como la utilización de capas de convolución y de pooling, la aplicación de regularización y la optimización de hiperparámetros.

5. Modelo de referencia

Como primer algoritmo, se implementará una red neuronal convolucional básica. Se utiliza PyTorch como framework de deep learning para construir el modelo. El código se divide en dos partes: la primera parte es la definición de la clase Model, que es una clase abstracta que define los métodos necesarios para entrenar y evaluar un modelo de clasificación, incluyendo la definición de la red neuronal, el criterio de pérdida y el optimizador. La segunda parte es la implementación específica del modelo Cifar10Model, que hereda de la clase Model y proporciona la definición de los métodos abstractos. [6]

La implementación utiliza un modelo de red neuronal convolucional con varias capas convolucionales y capas totalmente conectadas. La función de activación utilizada es ReLU y se utiliza la técnica de dropout para regularizar el modelo. La clase Cifar10Model también define el método epoch_step, que realiza un paso de entrenamiento o evaluación en un conjunto de datos, así como los métodos train y evaluate, que implementan el entrenamiento y la evaluación del modelo, respectivamente. En el método train, se utilizan los datos de entrenamiento para actualizar los pesos del modelo utilizando el algoritmo de optimización SGD. Durante el entrenamiento, se realiza la evaluación en el conjunto de datos de validación para evaluar la calidad del modelo. El método evaluate realiza la evaluación en un conjunto de datos de prueba para obtener una evaluación final del modelo.

El código también incluye la carga y preprocesamiento de los datos de entrenamiento y prueba, utilizando la biblioteca Hugging Face para cargar el conjunto de datos CIFAR-10 en formato torch y el DataLoader de PyTorch para cargar los datos en lotes para el entrenamiento y la evaluación del modelo. Finalmente, el modelo se guarda y se carga en disco utilizando la función save_weights y load_weights, respectivamente, que utiliza la función torch.save y torch.load para almacenar y cargar los pesos del modelo.

6. Métricas

Para evaluar el desempeño del modelo de clasificación de imágenes, se pueden utilizar varias métricas. Una de las más comunes es la exactitud (accuracy), que mide la proporción de imágenes clasificadas correctamente en comparación con todas las imágenes de prueba. El primer acercamiento al problema produjo un modelo con un rendimiento modesto, después de 10 épocas de entrenamiento, el modelo logró una precisión de validación del 0.5040 y una pérdida de validación de 0.01129.

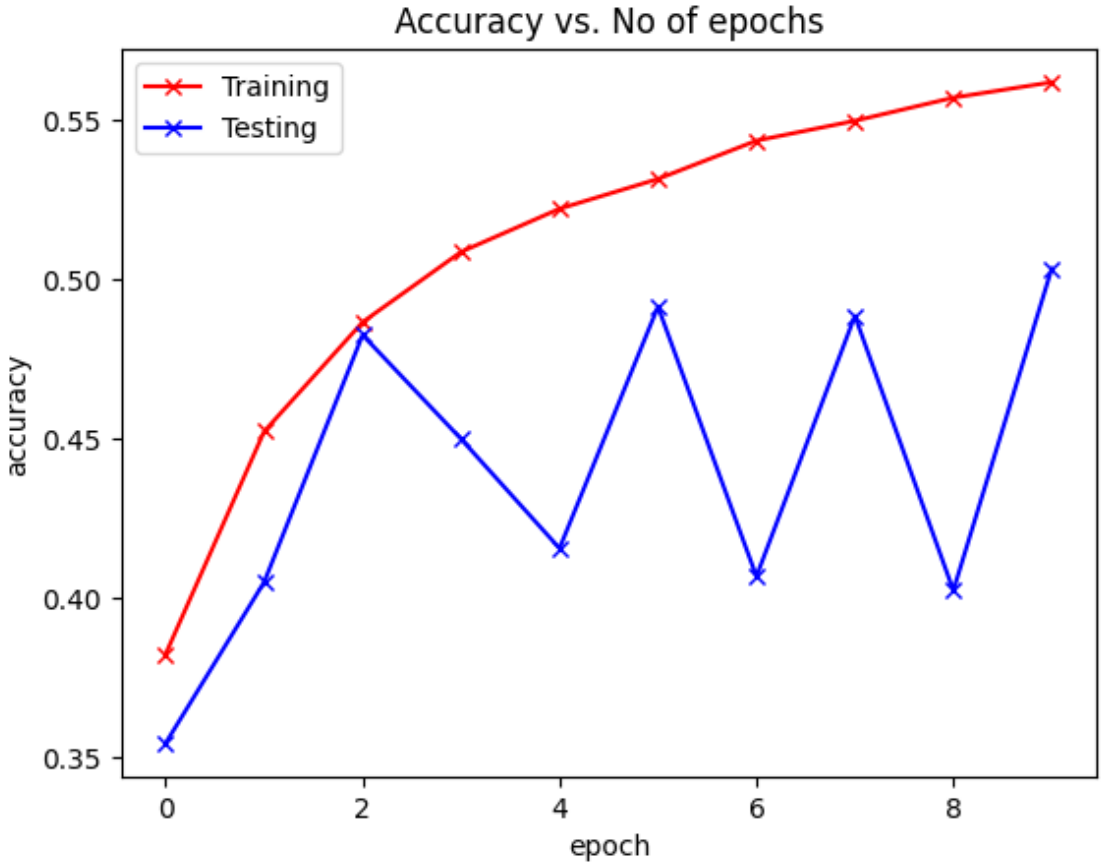

Figura 1: Accuracy contra el No. de Epochs

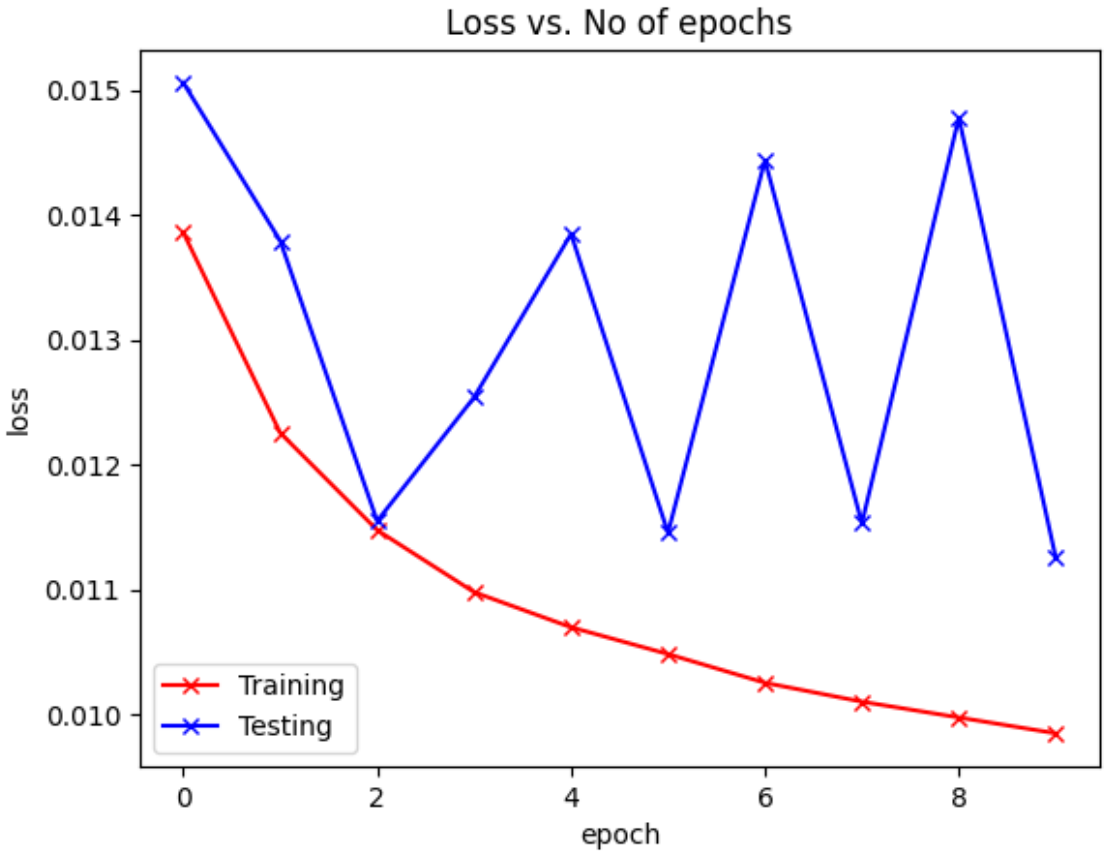

Figura 2: Loss contra el No. de Epochs

Esto sugiere que la red neuronal convolucional simple tiene una capacidad limitada para extraer características de las imágenes de CIFAR-10 y clasificarlas con precisión. Es posible que se requiera un modelo más complejo y profundo para abordar con éxito este problema. Además, se observó una pérdida de entrenamiento significativamente menor que la pérdida de validación, lo que sugiere que el modelo está sobreajustando los datos de entrenamiento. Esto puede ser mitigado utilizando técnicas como la regularización, la disminución de la tasa de aprendizaje y el aumento de datos.

Una alternativa es la matriz de confusión, que proporciona información detallada sobre

los resultados de la clasificación. La matriz de confusión muestra el número de imágenes clasificadas correctamente e incorrectamente para cada clase.

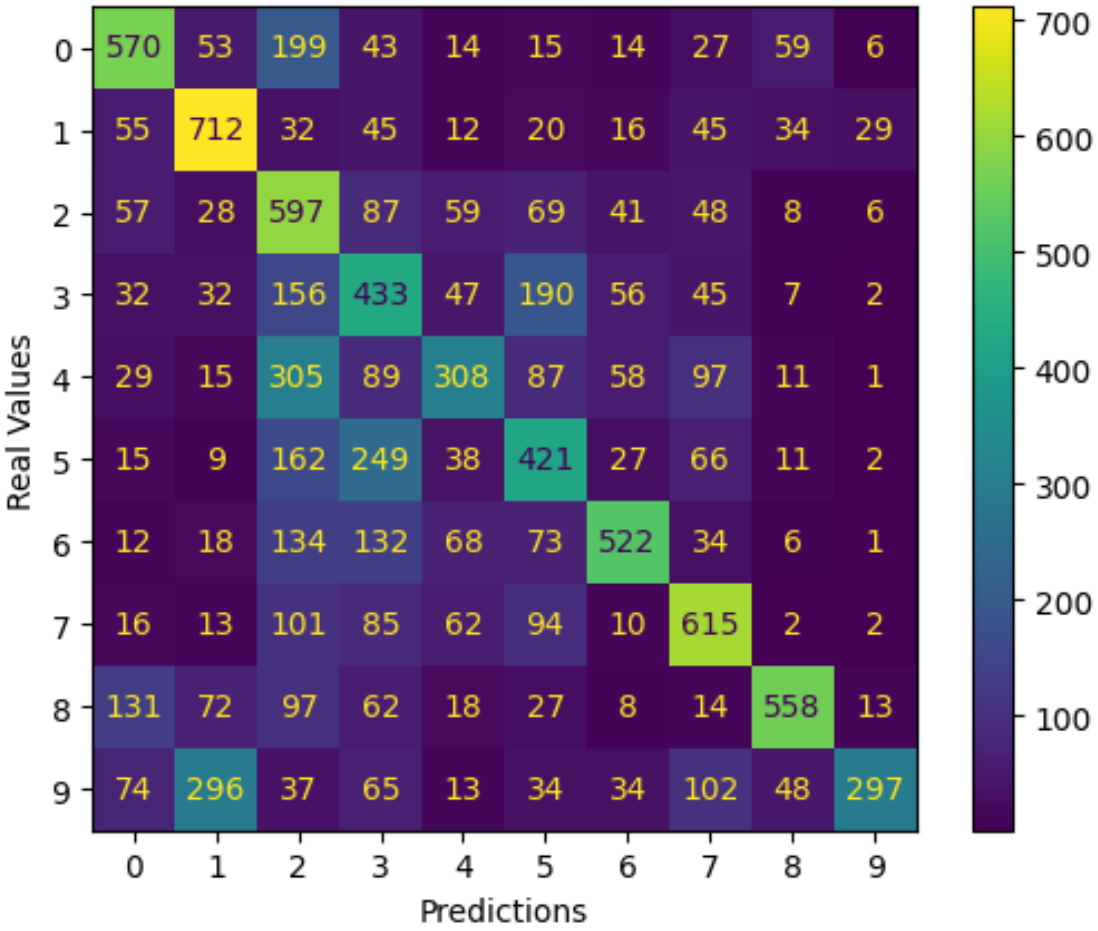

Figura 3: Matriz de Confusión Resultante

La matriz de confusión proporciona información sobre cómo el modelo clasifica los diferentes elementos en las diferentes clases. Esta matriz nos muestra que la diagonal principal, que representa las predicciones correctas, tiene valores altos para las clases 0 (aviones), 1 (automóviles), 2 (pájaros), 3 (gatos), 4 (ciervos), 5 (ranas) y 7 (caballos), lo que indica que el modelo tiene una buena capacidad de clasificación para esas clases. Sin embargo, también

podemos observar valores significativos en las posiciones fuera de la diagonal principal, lo que indica que el modelo tiene dificultades para clasificar correctamente algunas imágenes.

Por ejemplo, la clase 9 (camiones) tiene valores bajos en la diagonal principal, lo que indica que el modelo no logra clasificar correctamente las imágenes pertenecientes a esta clase con tanta frecuencia como debería. Además, la matriz muestra que las clases 0 (aviones), 2 (pájaros), 3 (gatos) y 4 (ciervos) tienen valores significativos fuera de la diagonal principal, lo que indica que el modelo tiene dificultades para clasificar correctamente algunas imágenes de estas clases. Esto sugiere que el modelo podría beneficiarse de ajustes o mejoras para aumentar su precisión en estas clases en particular.

Para concluir, resulta interesante observar que el modelo a menudo confunde gatos con pájaros y perros, y ciervos con pájaros y perros. Estas confusiones pueden deberse a la similitud en la forma de los cuerpos o en los patrones de pelaje, aunque esto no está claro. Por otro lado, es curioso que el modelo tenga dificultades para distinguir entre camiones y automóviles. Esto puede deberse a que estas clases comparten características similares, como el número de ruedas, la forma del cuerpo y el tamaño, lo que puede dificultar la identificación precisa de la clase.

Referencias

[1] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Information Processing Systems*, vol. 25, DOI 10.1145/3065386, 01 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, "Imagenet: a large-scale hierarchical image database," *IEEE Conference on Computer Vision and Pattern Recognition*, DOI 10.1109/CVPR.2009.5206848, pp. 248–255, 06 2009. [Online]. Available: https://www.researchgate.net/publication/221361415_ImageNet_a_Large-Scale_Hierarchical_Image_Database

[3] A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. Swetter, H. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, DOI 10.1038/nature21056, 01 2017. [Online]. Available: https://www.nature.com/articles/nature21056

[4] M. Jyotiyana, N. Kesswani, and M. Kumar, "A deep learning approach for classification and diagnosis of parkinson's disease," *Research Square*, DOI 10.21203/rs.3.rs-254647/v1, 02 2021. [Online]. Available: https://www.researchgate.net/publication/352455681_A_Deep_Learning_Approach_for_Classification_and_Diagnosis_of_Parkinson%27s_Disease

[5] G. H. Alex Krizhevsky, Vinod Nair, "Learning multiple layers of features from tiny images," *University of Toronto*, 2009. [Online]. Available: https://huggingface.co/datasets/cifar10

[6] Digit eye. [Online]. Available: https://github.com/Oktuvida/digit-eye.git