UNIVERSIDAD DE ANTIOQUIA
Facultad de Ingeniería

Informe de la entrega 1 sobre

La clasificación de imágenes de comida

Extracto del episodio 4, temporada 4 de la serie Silicon Valley (2017, HBO)

Fundamentos de Deep Learning

Semestre 2024.2

Un curso dirigido por el profesor **Raúl Ramos Pollán -** raul.ramos@udea.edu.co

Hugo Maharavo – 20AI87814
hugo.maharavo@udea.edu.co

Contexto de aplicación

Este proyecto de clasificación de imágenes de alimentos está inspirado en una escena de la serie Silicon Valley (episodio 4, temporada 4, 2017, emitida por HBO).
Puedes verla en el siguiente enlace: https://youtu.be/tWwCK95X6go?si=ImbPChJhEN_WhPU-

En esta escena, Jian Yang está desarrollando una app que identifica hot dog haciendo una foto. Su equipo está inicialmente impresionado por esta funcionalidad, pero su entusiasmo pronto baja cuando descubren que la aplicación es binaria: solo diferencia entre las fotos en las que hay hot-dog y las que no. Cuando intentan reconocer una pizza, lo único que obtienen es la respuesta “no hot-dog”

Me pareció gracioso. Así que lo utilicé como inspiración para este proyecto y superar la frustración de los amigos de Jian Yang. En lugar de una simple detección binaria, mi objetivo es diseñar una arquitectura de deep learning capaz de identificar 11 categorías de alimentos a partir de fotos.

Objetivo del modelo

El objetivo principal es predecir la categoría de un alimento a partir de una imagen. A diferencia de la aplicación de Jian Yang, que se limita a “hot-dog” o “no hot-dog”, nuestro modelo pretende clasificar los alimentos en 11 categorías:

- `apple_pie`
- `cheesecake`
- `chicken_curry`
- `french_fries`
- `fried_rice`
- `hamburger`
- `hot_dog`
- `ice_cream`
- `omelette`
- `pizza`
- `sushi`

Nuestro objetivo es que el modelo tenga una precisión mínima del 75%, lo que garantiza un nivel de fiabilidad suficiente para su uso práctico.

Fuente y contenido del dataset

El conjunto de datos es público, viene de Kaggle y se llama «Food11». Se puede acceder a él en el siguiente enlace: https://www.kaggle.com/datasets/imbikramsaha/food11

Contiene imágenes divididas en carpetas según categorías de alimentos. Estas imágenes tienen diversas resoluciones y contextos, con variaciones de brillo y ángulos. Este tipo de datos es habitual en visión por ordenador, ya que requiere que el modelo se adapte a una gran variedad de entornos.

Tamaño y estructura

- Conjunto de "`train`": 9900 imágenes, es decir, 900 imágenes por categoría.
- Conjunto de "`test`": 1100 imágenes, es decir, 100 imágenes por categoría.

La ventaja de este dataset es que está perfectamente equilibrado, lo que simplifica la evaluación del rendimiento del modelo al reducir el riesgo de sesgo de clase.

Evaluación del desempeño

Para evaluar el rendimiento de nuestro modelo, utilizaremos varias métricas clave, adaptadas a los objetivos del deep learning y a las necesidades de la aplicación.

Métricas de deep learning

1. `Global accuracy`: La métrica principal, calculada como el porcentaje de imágenes clasificadas correctamente en el dataset de "test". Aquí buscamos una precisión alta para garantizar la fiabilidad del modelo.

2. `Matriz de confusión`: Un análisis detallado por clase nos permitirá comprender si el modelo confunde determinadas categorías. Esto también puede revelar áreas de mejora.

3. `Recall`, `Accuracy` y `F-score` por clase: añadiendo estas métricas, tendremos una evaluación más precisa del rendimiento por categoría, útil si ciertas clases como hot_dog o pizza son más difíciles de detectar que otras.