18-mayo-2025

Leslie del Carmen Sanchez Reyes

Tec de monterrey campus querétaro

Primer reporte

Generación o selección del set de datos y Preprocesado de los datos

Contenido

[Obtención y Preprocesamiento del Dataset 2](#_Toc198502622)

[División de los datos: Entrenamiento y Prueba 2](#_Toc198502623)

[Referencias 3](#_Toc198502624)

Clasificación de Pokémon mediante Aprendizaje Supervisado con Imágenes

# Obtención y Preprocesamiento del Dataset

Para construir el dataset se utilizó como base el conjunto de datos “**7,000 Labeled Pokémon**” disponible públicamente en Kaggle [1]. Este dataset contiene imágenes etiquetadas y centradas de Pokémon de la primera generación. Contiene 150 folders cada uno con entre 25 a 50 imágenes por Pokémon.

Para este proyecto se seleccionaron solo 8 clases de Pokémon:

* Abra (351 imágenes)
* Aerodactyl (540 imágenes)
* Arbok (508 imágenes)
* Articuno (595 imágenes)
* Bulbasaur (392 imágenes)
* Cubone (387 imágenes)
* Diglett (361 imágenes)
* Ditto (366 imágenes)

Como el dataset original contaba con una cantidad muy pequeña de imágenes por Pokémon, para aumentar la cantidad de muestras y la variabilidad del dataset, se empleó un script existente “downloadImg” [2] para descargar imágenes adicionales desde Google y Pinterest, incrementando así la diversidad de las instancias visuales.

Una vez descargadas las imágenes con ayuda del script anteriormente mencionado, se hizo una depuración manual, evaluando que las imágenes obtenidas fueran de acuerdo con lo buscado y en su defecto eliminarlas o recortarlas para enfocarse en el objetivo.

# División de los datos: Entrenamiento y Prueba

Los datos fueron divididos en dos subconjuntos:

* **80% para entrenamiento**
* **20% para prueba**

Esta distribución fue basada en la división típica o sugerida en otros proyectos.

# Referencias

[1] L. Tian, "7,000 Labeled Pokemon," Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/lantian773030/pokemonclassification>  
[2] O. del Valle Mejía, *QuintupletRecognizer: downloadImg.py*, GitHub repository. [Online]. Available: <https://github.com/OsvalDev/QuintupletRecognizer/tree/master/scripts>