

Análisis y Reporte sobre el desempeño del modelo

Diego Emilio Barrera Hernández

A01366802

Introducción:

Se realizó un modelo de clasificación categórica para identificar 3 categorías, y con estos clasificarlos según su imagen y su nombre. Las clases son las siguientes:

- Humano
- Anime
- Cartoon

Para mejor visualización de las imágenes, se puede consultar directamente las páginas de referencia dónde se obtuvieron dichas imágenes, en la siguiente URL:

- Humanos

<https://www.kaggle.com/datasets/ahmadahmadzada/images2000>

- Anime

<https://www.kaggle.com/datasets/diraizel/anime-images-dataset>

- Cartoon

<https://www.kaggle.com/datasets/hadipratamatulili/anime-vs-cartoon-vs-human>

Método empleado:

Para este análisis, se empleó un modelo de red neuronal basado en la arquitectura VGG16 para la clasificación. Se implementó un conjunto de aumento de datos durante el entrenamiento y la validación, utilizando la

librería de ImageDataGenerator. La red neuronal donde se añaden capas de aplanamiento, seguidas de una capa densa con 256 unidades y activación "relu", una capa de normalización por lotes, y una capa de abandono con una tasa del 50%. Finalmente, se incluye una capa de salida con activación "softmax" para clasificar en las 3 categorías: Anime, Humano y Cartoon.

A continuación, explico la razón de las configuraciones anteriores. La capa Flatten se eligió para preparar los datos tridimensionales para la entrada en las capas densas. La capa con activación relu de 256 neuronas permite aprender combinaciones complejas de las características de nuestros datos. La capa de normalización estabiliza el entrenamiento, mientras que la capa de dropout se usa para evitar el sobreajuste. Finalmente, la capa softmax se utiliza para clasificar nuestras 3 categorías.

Métricas utilizadas:

En un estudio sobre CNNs, la métrica "accuracy" es útil para evaluar su rendimiento en la clasificación de imágenes entre humanos, anime y cartoon. Mide la proporción de imágenes correctamente clasificadas, siendo simple y efectiva para esta tarea. Es importante considerar que la accuracy no siempre es la mejor métrica para todas las aplicaciones de clasificación de imágenes. Convolutional Neural Networks for Image Classification: Survey (<https://ieeexplore.ieee.org/document/9758454>)

En la primera prueba del modelo, se empleó un total de 15 épocas. Durante este proceso, se observaron mejoras progresivas en la precisión tanto en los datos de entrenamiento como en los de validación. En

la primera época, el modelo comenzó con un accuracy del **49.00%** en el entrenamiento y del **71.40%** en la validación. Con el transcurso de las épocas, se evidenció un crecimiento significativo en la capacidad predictiva del modelo, alcanzando una precisión del **82.80%** en el entrenamiento y del **83.60%** en la validación en la última época.

Los valores de perdida también disminuyeron de manera consistente a lo largo de las épocas, lo que indica una mejora en la capacidad del modelo para predecir con mayor precisión. Al inicio, la perdida fue de **1.3575** en el entrenamiento y de **0.7569** en la validación, mientras que al final del proceso, estos valores disminuyeron significativamente a **0.4327** en el entrenamiento y **0.4260** en la validación.

Se hicieron 3 representaciones gráficas de nuestros resultados. En la Fig 1, se pueden ver los resultados obtenidos en el entrenamiento y la validación.

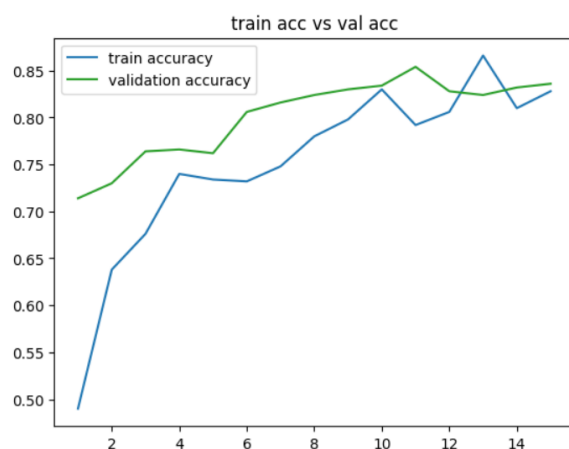


Fig 1: Evolución de precisión de entrenamiento y validación

En esta gráfica podemos observar cómo evolucionan la precisión de entrenamiento y la de validación. Se puede ver que la precisión de entrenamiento comienza en

0.72 y aumenta hasta 0.83 a medida que el modelo se entrena. La precisión de validación, por su parte, comienza en 0.49 y también aumenta hasta 0.82. Este es un buen resultado, pero es importante señalar que la precisión de entrenamiento es ligeramente superior a la de validación, lo cual podría indicar que el modelo está sobreajustándose a los datos de entrenamiento. Sin embargo, podemos observar que la precisión de validación superó a la de entrenamiento entre las épocas 12 y 14, lo que sugiere que el modelo se estaba regulando de manera efectiva. Es importante mencionar que esta diferencia en los valores no solo se puede deber al sobreajuste, sino también a la complejidad de los datos de entrenamiento, lo cual tiene sentido dada la distribución del dataset.

Ahora presentamos la Fig 2, aquí se puede observar los resultados visuales de la perdida en entrenamiento y la perdida en validación.

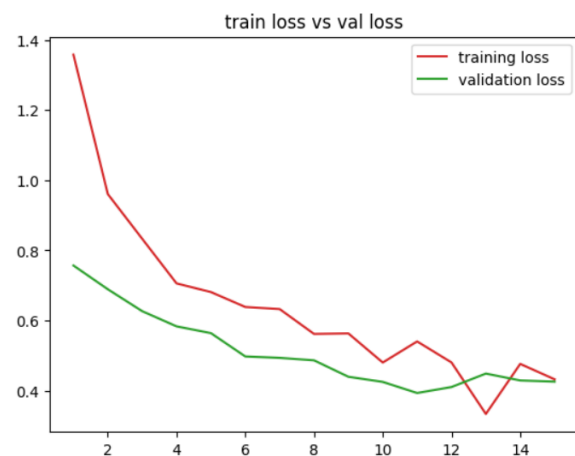


Fig 2: Evolución de la perdida de entrenamiento y validación

La pérdida de entrenamiento comienza en 1,4 y disminuye hasta 0,43 a medida que el modelo se entrena. La pérdida de validación comienza en 0,75 y también disminuye hasta aproximadamente 0,43. Como se mencionó

previamente, La pérdida de entrenamiento y validación disminuye a medida que el modelo se entrena, lo que indica que el modelo está aprendiendo a realizar las tareas para las que fue diseñado.

Por último, en la Fig 3, podemos observar una matriz de confusión que nos da los resultados de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.

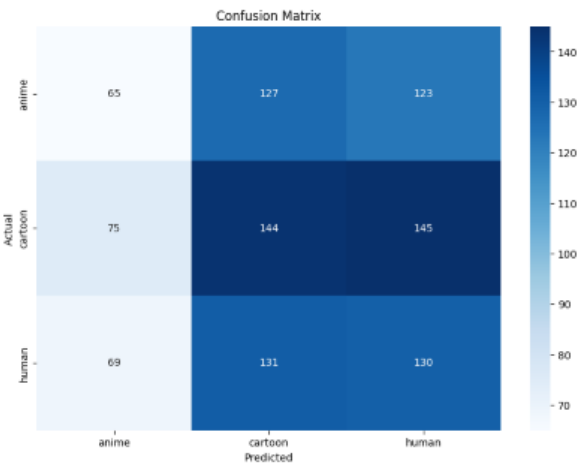


Fig 3: Matriz de confusión

En la matriz podemos ver los resultados de nuestras 1009 imágenes de prueba. En este caso, Anime ha sido la categoría más afectada, ya que solo se predijeron 65 imágenes correctamente. Por otro lado, tenemos 144 predicciones correctas para Cartoon y 130 para Human. El resto fueron predicciones erróneas. Cabe mencionar que Cartoon fue confundida con Human 145 veces, siendo esta la mayor confusión. Esto puede deberse a la distribución del dataset.

En esta última figura podemos ver en termino de porcentajes, los resultados de las predicciones.

	precision	recall	f1-score	support
anime	0.31	0.21	0.25	315
cartoon	0.36	0.40	0.38	364
human	0.33	0.39	0.36	330
accuracy			0.34	1009
macro avg	0.33	0.33	0.33	1009
weighted avg	0.33	0.34	0.33	1009

La matriz muestra una precisión general del 34%, con precisiones por clase de 31% para 'anime', 36% para 'cartoon' y 33% para 'human'. El recall es especialmente bajo para 'anime' (21%), indicando problemas en identificar correctamente esta categoría. Los F1-scores bajos reflejan un equilibrio pobre entre precisión y recall. Esto podría significar que el modelo necesita mejoras en su configuración y datos de entrenamiento para mejorar su rendimiento.

Mejora del modelo: