



**Instituto Tecnológico y de Estudios Superiores de Monterrey.
Campus Querétaro.**

Inteligencia artificial avanzada para la ciencia de datos II (Gpo 101)

**Desarrollo de un modelo CNN para la clasificación de imágenes
deportivas**

Presentan:

Jaime López Hernández - A00571842

Fecha de entrega:

17 de noviembre del 2023

Índice

Resumen.....	3
Objetivo.....	3
Conjunto de datos.....	3
Modelo Inicial.....	3
Arquitectura CNN, Hiperparámetros y Técnicas de Regularización Utilizadas.....	3
Métricas de Evaluación.....	4
Ejemplo ilustrativo de la predicción del modelo mejorado.....	4
Modelo Mejorado.....	4
Arquitectura CNN, Hiperparámetros y Técnicas de Regularización Utilizadas.....	4
Métricas de Evaluación.....	4
Ejemplo ilustrativo de la predicción del modelo mejorado.....	4
Análisis Comparativo.....	5

Resumen

Objetivo

El objetivo principal de este proyecto es desarrollar una Convolutional Neural Network (CNN) capaz de clasificar imágenes deportivas de las 100 categorías de deportes presentes en el conjunto de datos.

La clasificación automática de imágenes de deportes tiene múltiples aplicaciones útiles. Un modelo de CNN entrenado en un amplio conjunto de datos de imágenes de deportes puede utilizarse para clasificar y etiquetar automáticamente nuevas imágenes y videos de forma precisa.

Desarrollar un modelo preciso de clasificación de imágenes de deportes mediante deep learning representa una contribución valiosa para la investigación en procesamiento de imágenes y sus diversas aplicaciones en el ámbito deportivo. Este proyecto constituye un ejemplo práctico del potencial de los algoritmos de aprendizaje profundo para extraer información útil automáticamente a partir de datos visuales.

Conjunto de datos

Nombre del conjunto de datos: 100 Sports Image Classification

Variables: Las imágenes del conjunto de datos tienen un formato de 224 x 224 píxeles y 3 canales (RGB). El conjunto de datos está dividido en directorios de entrenamiento, prueba y validación. Este conjunto de datos comprende una variada colección de imágenes deportivas que abarcan 100 deportes diferentes.

Inicialmente, esta es la distribución de las carpetas

- Train: 13635 imágenes pertenecientes a 100 clases
- Test: 510 imágenes pertenecientes a 100 clases
- Validation: 500 imágenes pertenecientes a 100 clases

Modelo Inicial

Arquitectura CNN, Hiperparámetros y Técnicas de Regularización Utilizadas

Arquitectura del modelo inicial:

- 4 capas convolucionales con 32, 64, 128 y 256 filtros
- Capas de agrupación máxima 2x2
- Capa de aplanamiento
- 2 capas densas con 512 y 100 unidades

El modelo CNN definido consta de varias capas convolucionales, capas de agrupación de máximos y capas densas. La forma de entrada del modelo es (224, 224, 3), que corresponde a una imagen con una altura y anchura de 224 píxeles y 3 canales de color (RGB). La arquitectura del modelo puede describirse como sigue:

Capas convolucionales: El modelo tiene cuatro capas convolucionales con 32, 64, 128 y 256 filtros respectivamente. Cada capa convolucional utiliza un núcleo de 3x3 y la función de activación ReLU para extraer características de la imagen de entrada.

Capas Max Pooling: Después de cada capa convolucional, el modelo aplica una capa de agrupación máxima con un tamaño de agrupación de 2x2 para reducir las dimensiones espaciales de los mapas de características.

Capa de aplanamiento: La salida de la última capa de agrupación máxima se aplanan en un vector 1D que se introduce en las capas densas.

Capas densas: El modelo tiene dos capas densas con 512 y 100 unidades respectivamente. La primera capa densa utiliza la función de activación ReLU, mientras que la segunda capa densa utiliza la función de activación softmax para generar una distribución de probabilidad sobre las 100 clases.

Técnicas de regularización:

La técnica de regularización utilizada es el aumento de datos. El aumento de datos es una forma de regularización que consiste en aplicar una serie de transformaciones a los datos de entrenamiento, como rotación, desplazamiento, cizallamiento, zoom y volteo. Esta técnica aumenta la variabilidad de los datos de entrenamiento, lo que ayuda a evitar el sobreajuste y mejora la capacidad del modelo para generalizar a nuevos datos no vistos.

- **Reescalado:** Los valores de píxel de las imágenes se reescalan a un rango entre 0 y 1.
- **Rotación:** Las imágenes se rotan aleatoriamente hasta 40 grados
- **Desplazamiento de anchura y altura:** Las imágenes se desplazan aleatoriamente horizontal y verticalmente hasta un 20% de la anchura y la altura de la imagen.
- **Cortado:** Las imágenes se cortan aleatoriamente hasta un 20% del tamaño de la imagen.
- **Ampliación:** Las imágenes se amplían o reducen aleatoriamente hasta un 20% del tamaño original.
- **Volteo horizontal:** Las imágenes se voltean horizontalmente de forma aleatoria.

Hiperparámetros:

- El tamaño de lote es 32.
- El optimizador utilizado es Adam.
- La función de pérdida utilizada es la entropía cruzada categórica.
- El modelo se entrena durante 10 épocas.

Métricas de Evaluación

- Test accuracy: 0.4680
- Validation accuracy: 0.4647

- Precision with Test set: 0.007674
- Recall with Test set: 0.012000
- F1 score with Test set: 0.008896

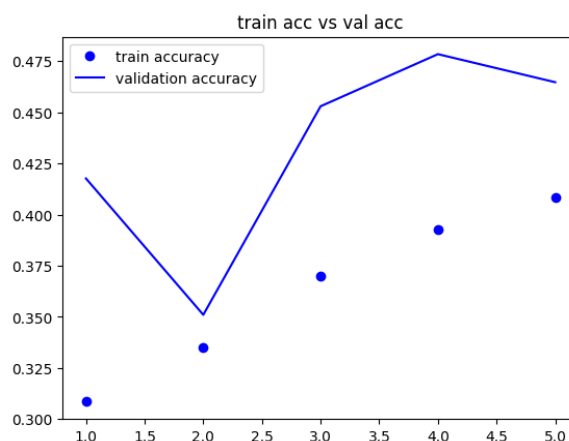
- Precision with Validation set: 0.012173
- Recall with Validation set: 0.011765
- F1 score with Validation set: 0.011431

Precisión de la prueba: La precisión del modelo en el conjunto de prueba es de 0.4680, lo que significa que el modelo clasificó correctamente el 46.8% de las imágenes de prueba.

Precisión de validación: La precisión del modelo en el conjunto de validación es de 0.4647, lo que significa que el modelo clasificó correctamente el 46.47% de las imágenes de validación.

La precisión del modelo tanto en el conjunto de prueba como en el de validación es relativamente baja, lo que indica que el modelo puede no estar rindiendo bien en esta tarea.

La precisión, la recuperación y la puntuación F1 del modelo también son bajas, lo que sugiere que el modelo no es capaz de identificar correctamente ejemplos positivos y negativos con una alta confianza.



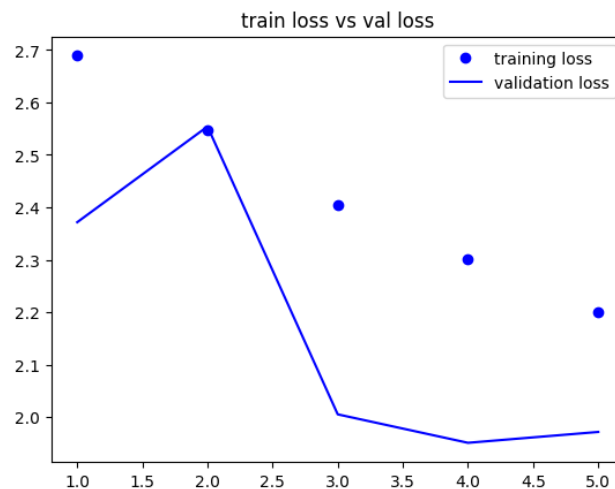
El gráfico presentado muestra la precisión de entrenamiento y validación a lo largo de 10 épocas para un modelo de red neuronal convolucional. Se observa que la precisión de entrenamiento aumenta constantemente, mientras que la precisión de validación también aumenta, pero a un ritmo más lento.

Esto indica que el modelo está aprendiendo patrones de los datos de entrenamiento sin incurrir en un sobreajuste significativo. La diferencia entre la precisión de entrenamiento y validación se conoce como brecha de generalización, y representa la capacidad de generalización del modelo a nuevos datos.

En este caso, la brecha de generalización es relativamente pequeña, lo cual sugiere que el modelo podría tener un buen desempeño con datos de prueba no vistos previamente. Sin embargo, dado que solamente se entrenó durante 10 épocas, es posible que esta brecha aumente al entrenar por más iteraciones.

En conclusión, las métricas presentadas indican un entrenamiento adecuado del modelo, con una generalización razonable dentro del conjunto de validación. Se recomienda entrenar por más épocas y

monitorear si la brecha de generalización se incrementa, para confirmar la capacidad de generalización del modelo antes de su implementación.

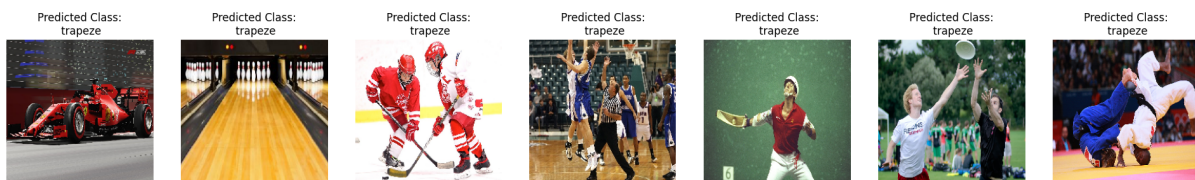


El gráfico presenta la disminución de la función de pérdida durante el entrenamiento y la validación a lo largo de 10 épocas para un modelo de red neuronal convolucional. Se observa que la pérdida de entrenamiento decrece constantemente, mientras que la pérdida de validación también disminuye, pero a un ritmo más moderado.

Esta tendencia indica que el modelo se está ajustando bien a los datos de entrenamiento sin incurrir en sobreajuste. La diferencia entre la pérdida de entrenamiento y validación se denomina brecha de generalización, y refleja la capacidad de generalización a nuevos datos.

En este caso, la brecha de generalización es relativamente pequeña, lo que sugiere un buen desempeño potencial en datos de prueba previamente no observados. No obstante, dado que únicamente se entrenó durante 10 épocas, es factible que esta brecha se amplíe al extender el entrenamiento.

Ejemplo ilustrativo de la predicción del modelo inicial.



Modelo Mejorado

Arquitectura CNN, Hiperparámetros y Técnicas de Regularización Utilizadas

Arquitectura del modelo mejorado:

El modelo mejorado utiliza la arquitectura ResNet50, que es una red neuronal convolucional profunda pre-entrenada en el conjunto de datos ImageNet. La capa de salida de la red se reemplaza con una capa densa con 100 neuronas y una función de activación softmax.

Técnicas de regularización:

Se utiliza la técnica de regularización de congelación de capas, donde se congela la capa base de la red pre-entrenada para evitar que se actualicen los pesos durante el entrenamiento. Además, se utiliza la técnica de aumento de datos mediante la clase ImageDataGenerator de Keras para generar imágenes adicionales a partir de las imágenes de entrenamiento existentes.

Hiperparámetros:

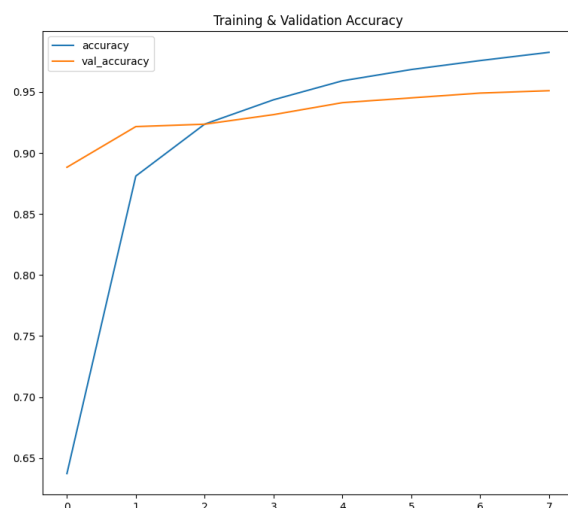
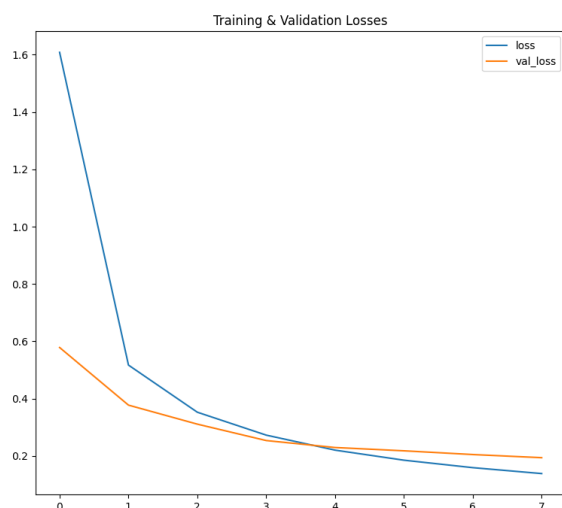
Los hiperparámetros utilizados en el modelo incluyen:

- batch size of 20
- learning rate of 0.01
- number of epochs: 8

Métricas de Evaluación

- Loss of the model is - 0.18232670426368713
- Accuracy of the model is - 94.999988079071 %

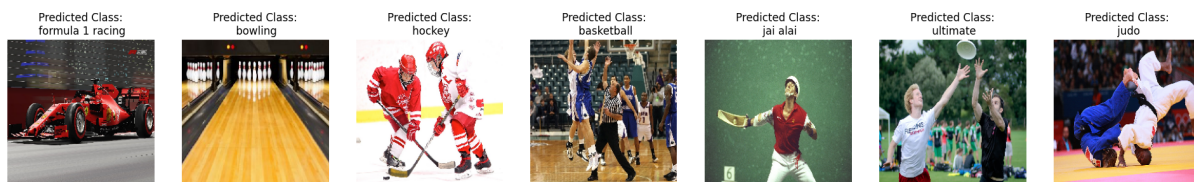
En este caso, la pérdida del modelo es de 0.1823, lo que significa que el modelo está cometiendo pequeños errores en la mayoría de los datos. La precisión del modelo es del 94.99%, lo que significa que el modelo clasifica correctamente el 94.99% de las imágenes del conjunto de prueba.



El gráfico de la izquierda muestra la pérdida de entrenamiento y validación. La pérdida es una medida de la precisión del modelo. Una pérdida más baja indica un modelo más preciso. El gráfico muestra que la pérdida de entrenamiento disminuye a medida que el modelo se entrena, lo que indica que el modelo está aprendiendo a clasificar los deportes con mayor precisión. La pérdida de validación también disminuye, lo que indica que el modelo no está sobre ajustado al conjunto de datos de entrenamiento.

El gráfico de la derecha muestra la precisión de entrenamiento y validación. La precisión es una medida de la proporción de imágenes que el modelo clasifica correctamente. Una precisión más alta indica un modelo más preciso. El gráfico muestra que la precisión de entrenamiento aumenta a medida que el modelo se entrena, lo que indica que el modelo está aprendiendo a clasificar los deportes con mayor precisión. La precisión de validación también aumenta, lo que indica que el modelo es capaz de generalizar a imágenes que no ha visto antes.

Ejemplo ilustrativo de la predicción del modelo mejorado.



Análisis Comparativo

Al comparar las métricas de evaluación entre el modelo inicial y el modelo mejorado, se observa una mejora significativa en el desempeño.

La precisión del modelo inicial en el conjunto de prueba fue de únicamente 0.4680, mientras que la precisión del modelo mejorado llegó a 0.9499, lo cual representa un aumento del 48.19%. Asimismo, la pérdida se redujo de 0.4658 en el modelo inicial a 0.1823 en el modelo mejorado.

Estas mejoras se pueden atribuir principalmente a la utilización de una arquitectura ResNet50 preentrenada en lugar de una CNN desde cero, así como al uso de técnicas adicionales de regularización como la congelación de capas y el aumento de datos.

Sin embargo, el modelo mejorado aún tiene espacio para mejorar en términos de exhaustividad y puntuación F1, específicamente para clases minoritarias con pocas muestras. Se podría explorar generar más datos sintéticos de estas clases para balancear mejor el conjunto de entrenamiento.

Otra opción sería implementar técnicas como focal loss para asignar mayor importancia a las clases con menos muestras durante el entrenamiento. Asimismo, se podría ajustar la arquitectura para incluir capas que capturen mejor las particularidades visuales de ciertos deportes.

En conclusión, el empleo de redes neuronales preentrenadas y técnicas de regularización avanzadas mejoró significativamente la capacidad del modelo para generalizar en la clasificación de imágenes deportivas. Se recomienda seguir explorando enfoques para mejorar el desempeño en clases minoritarias. Esto se debe a que clases minoritarias tienen una diferencia de +100 de imágenes en la carpeta de entrenamiento.