Inteligencia Artificial

Modelo de Detección de Imágenes sobre Transmisiones Deportivas de Basketball

Samuel Lopera Torres[†], Juan José Sánchez Sánchez[†] y Andrés Restrepo Botero [†]
sloperat@eafit.edu.co, jjsanchez@eafit.edu.co,arestrepob@eafit.edu.co

Escuela de Ciencias Aplicadas e Ingeniería, Universidad EAFIT

Noviembre del 2025

1. Planteamiento del Problema

El crecimiento del contenido deportivo en plataformas digitales y transmisiones masivas ha impulsado la necesidad de sistemas automáticos capaces de analizar videos y extraer información relevante. En el caso del baloncesto, la detección de objetos representa un reto considerable debido al ritmo acelerado del juego, las constantes oclusiones entre jugadores, variaciones en la iluminación y múltiples ángulos de cámara propios de las transmisiones profesionales.

El propósito de este trabajo es desarrollar un modelo de **detección de objetos** capaz de identificar jugadores, balones, árbitros y otros elementos relevantes dentro de fotogramas extraídos de transmisiones reales de partidos. Para ello, se entrena un detector basado en la arquitectura **RF-DETR** (*Relation-Free Detection Transformer*), un modelo de última generación diseñado para mejorar la eficiencia y velocidad de convergencia de los transformadores aplicados a visión por computador.

El conjunto de datos utilizado está compuesto por imágenes anotadas derivadas de transmisiones completas de partidos, incorporando variabilidad significativa en resolución, condiciones de iluminación, movimiento y elementos visuales. Esto permite evaluar la capacidad del modelo para generalizar en condiciones reales y no controladas.

2. Arquitectura del Modelo

El modelo seleccionado es **RF-DETR**, una variante moderna de los detectores tipo Transformer. A diferencia del DETR original, RF-DETR elimina la atención entre consultas en el decodificador, reduciendo el costo computacional y acelerando la convergencia sin sacrificar precisión.

La arquitectura se compone de los siguientes bloques:

- **Backbone:** Una red convolucional (ResNet50, en la configuración base) encargada de extraer características multi-escala a partir de la imagen.
- **Encoder:** Un codificador Transformer que procesa los mapas de características generados por el backbone.
- **RF Decoder:** Decodificador libre de relaciones, que produce predicciones de cajas y clases sin necesidad de auto-atención entre consultas.

- **Cabezas de Predicción:** MLPs encargadas de generar probabilidades de clase y coordenadas de cajas delimitadoras.

Entre las decisiones de diseño más relevantes se incluyen el uso de características multi-escala, la eliminación de dependencias entre consultas y el uso de asignación bipartita durante el entrenamiento.

3. Procedimiento de Entrenamiento

El modelo se entrena de manera end-to-end siguiendo los protocolos estándar en detección.

Hiperparámetros

- **Learning Rate:** 1e−4 para capas del Transformer
- **Tamaño de Lote:** 4
- **Épocas:** 10
- **Aumentación:** Recortes aleatorios, Blur de 0.4 px
- **Hardware:** GPU NVIDIA T4 (16GB)

4. Resultados

Los resultados obtenidos se basan directamente en el archivo `results.json`. Para el conjunto de validación, el desempeño promedio del modelo fue:

- **mAP@50:95 (promedio):** 0.5539
- **mAP@50 (promedio):** 0.8647
- **Precisión global:** 0.8494
- **Recall global:** 0.78

Las clases con mejor desempeño fueron:

- **Player:** mAP@50:95 = 0.6830, mAP@50 = 0.9523
- **Team Points:** mAP@50:95 = 0.6649, mAP@50 = 0.9179
- **Ref:** mAP@50:95 = 0.6336, mAP@50 = 0.9109

Las clases más retadoras incluyeron:

- **Ball:** mAP@50:95 = 0.3588
- **Shot Clock:** mAP@50:95 = 0.4013

Estas dificultades se deben principalmente al tamaño reducido del balón y a la variabilidad visual de los elementos gráficos del marcador.

En el conjunto de prueba, los resultados fueron consistentes, con un mAP@50:95 general de 0.5691 y una precisión promedio del 0.9614, evidenciando una buena capacidad de generalización del modelo.