

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Ingeniería

Deep Learning



## Laboratorio 8

MANUEL ALEJANDRO ARCHILA MORAN 161250

JUAN DIEGO AVILA SAGASTUME 20090

DIEGO JOSE FRANCO PACAY 20240

# Laboratorio 8

## Parte Práctica:

### 1.

Link del video: <https://youtu.be/LV7vwztlExw>

### Preguntas:

**a. ¿Encontró algún caso donde el modelo pareciera no tener un score aceptable? (Un score aceptable queda sujeto a su criterio) ¿Por qué cree que el score fue ese?**

- En el momento de hacer pruebas con el modelo se obtuvieron unas inconsistencias por ejemplo detectaba que el cepillo de dientes era un cuchillo y lo mostraba con un valor de 0.5 y cuando lo detectaba como cepillo de dientes eran valores muy bajos alrededor de 0.2 - 0.3. También hubo instancias donde detectaba el pachón como un control remoto con valores entre 0.5 - 0.7. Esto puede deberse a que la formas de los objetos eran similares. Los cuchillos y los cepillos de dientes son cortos y un poco largos. De igual manera el pachón con el que se realizó la prueba era gris y con aspecto rectangular por lo que podía confundir al modelo.

**b. Investigue qué mejoras se han hecho en cada versión de YOLO, y con esto responda ¿qué cree que podría hacerse para mejorarlo?**

- YOLOv1 a YOLOv3:
  - Implementación de Batch Normalization
  - Mayor resolución
  - Introducción de bounding box al objeto predecido
  - Mejoras en las escalas de los objetos
- YOLOv4:
  - Mejora en la agregación de features
  - Mejoras en el data augmentation
- YOLOv5:
  - Entrenamiento más rápido para una mejorar la experimentación
- YOLOv6:
  - Rediseño en los requerimientos de hardware
  - Mejoras al pipeline de entrenamiento:
    - Anchor-free training

- SloU box Regression
- YOLOv7:
  - Reducción en el requerimiento de memoria
  - Reducción a la distancia de gradientes a la que se tiene que hacer backpropagation.
  - La última capa del modelo usa E-ELEAN

Para mejorarlo se podría en la detección de objetos pequeños en escenas de con varios objetos. También se puede reducir la complejidad pero manteniendo la precisión como se hizo para YOLO v5.

**c. Escriba detalladamente cómo podría usar YOLO para poder realizar detección de objetos sobre un dataset propio, es decir, sobre algún problema totalmente nuevo para YOLO, por ejemplo, detección de puentes con imágenes satelitales.**

Utilizar YOLO para detectar los tipos de frutas en una imagen.

- Recolectar y preparar los datos
  - Para esto se necesita conseguir una gran cantidad de imágenes de diferentes tipos de frutas. Con diferentes iluminaciones y de diferentes ángulos
- Procesar los datos
  - Normalizar los imágenes a un mismo tamaño
  - Separar los datos en entrenamiento, validación y testeo para poder evaluar el modelo
- Entrenar el modelo
  - Adaptar la arquitectura del modelo de YOLO para poder reconocer los tipos de frutas en los datos. Para esto es necesario cambiar el número de clases que se predicen, el cual corresponde a los diferentes tipos de frutas en el dataset.
- Evaluar el modelo
  - Evaluar los resultados de las predicciones realizadas por el modelo con métricas de desempeño como mean Average Precision (mAP)
  - Hacer fine tuning al modelo modificando hyperparameters e implementando data augmentation.

## **Parte Teórica:**

**1. Defina en sus propias palabras qué es una caja de anclaje (anchor box). Asegúrese de mencionar cuál es su importancia en la detección de objetos hecha por YOLO y cuál es la diferencia y relación con las cajas delimitadoras (bounding boxes)**

Una caja de anclaje es aquella herramienta que se utiliza para predecir la ubicación y la clase de un objeto en una imagen. Esta herramienta fue creada por YOLO, ya que permite que la red neuronal aprenda a detectar objetos de diferentes tamaños y formas. La importancia de

estas cajas está en que permite a la red neuronal aprender a detectar objetos de diferentes tamaños y formas sin tener que ser explícitamente entrenada para cada tipo de objeto. Esto hace que el algoritmo sea más eficiente y generalista.

La diferencia entre la caja de anclaje y las cajas delimitadoras es que las de anclaje ya están predefinidas, mientras que las delimitadoras se calculan a partir de una de anclaje.

La relación que tienen estos dos tipos de cajas es que las cajas delimitadoras son las que se calculan a partir de las cajas de anclaje y son las que se utilizan para identificar los objetos en la imagen. Para hacer esto se usan valores de predicción de la red neuronal para ajustar el centro, el ancho y el alto de las cajas de anclaje.

## **2. Considere leer el paper original de YOLO (You Only Look Once: Unified, Real-Time Object Detection) y luego responda.**

- a. ¿Cuál es la innovación clave que distingue a YOLO de los enfoques tradicionales de detección de objetos y cómo contribuye a la detección de objetos en tiempo real?**

Que puede hacer detecciones en tiempo real dividiendo la imagen en una cuadrícula y prediciendo bounding boxes y clases directamente sobre la imagen. Una sola red neuronal predice los bounding boxes y las probabilidades de pertenencia a una clase. Contribuye a la detección de objetos en tiempo real por la velocidad que tiene ya que tiene la capacidad de procesar las imágenes en una velocidad de 45 frames por segundo. Aunque YOLO hace más errores de localización que otras herramientas más sofisticadas, es menos probable de predecir falsos positivos en el entorno.

- b. ¿Cómo maneja YOLO la detección de múltiples objetos dentro de una sola imagen y cuáles son las ventajas de este enfoque sobre los métodos tradicionales?**

La red utiliza features de la imagen completa para predecir cada bounding box y así predecir cada elemento dentro de la imagen. Analiza de forma global toda la imagen y todos los objetos que aparecen en la misma. Separa la imagen en un grid y cada celda es responsable de analizar el objeto que está sobre el espacio establecido por dicha celda. No solo se realiza una predicción sino también calcula scores de confianza para ver que tan certero está el modelo que en una bounding box hay un elemento y también que tan certero está de que la predicción que hace es la correcta.

Ya que cada celda se encarga de predecir y analizar la imagen que está en su interior es mucho más rápido que los métodos tradicionales que tienen que analizar varias veces las imágenes para realizar predicciones y clasificar las regiones dentro de las imágenes. YOLO también reduce la cantidad de hiper parámetros facilitando el entrenamiento y la implementación del modelo en tiempo real.

**c. ¿Cuáles son las limitaciones o desafíos asociados con YOLO y cuáles son las áreas potenciales de mejora o investigación adicional en el campo de la detección de objetos en tiempo real?**

Una de las mayores limitaciones que posee YOLO es que puede ser susceptible a falsas alarmas, es decir que el algoritmo no puede garantizar que la predicción que está realizando del objeto detectado sea 100% cierta.

Además de ello se puede tener limitaciones con objetos pequeños que están cerca, porque el algoritmo utiliza cajas de anclaje predefinidas, que pueden no ser adecuadas para todos los tamaños y formas de objetos.

Algunas de las áreas potenciales en las que se puede mejorar el algoritmo son:

- Mejorar la precisión de las predicciones.
- Reducir la tasa de falsas alarmas.
- Mejorar el rendimiento en objetos pequeños o que se encuentran muy juntos.

Entre algunos ejemplos de investigación que se están desarrollando son:

- El desarrollo de algoritmos de detección de objetos basados en aprendizaje profundo que utilicen múltiples capas de predicción.
- El desarrollo de técnicas de aprendizaje profundo que sean más robustas a las perturbaciones.
- El desarrollo de algoritmos de detección de objetos que sean más sensibles a los detalles.