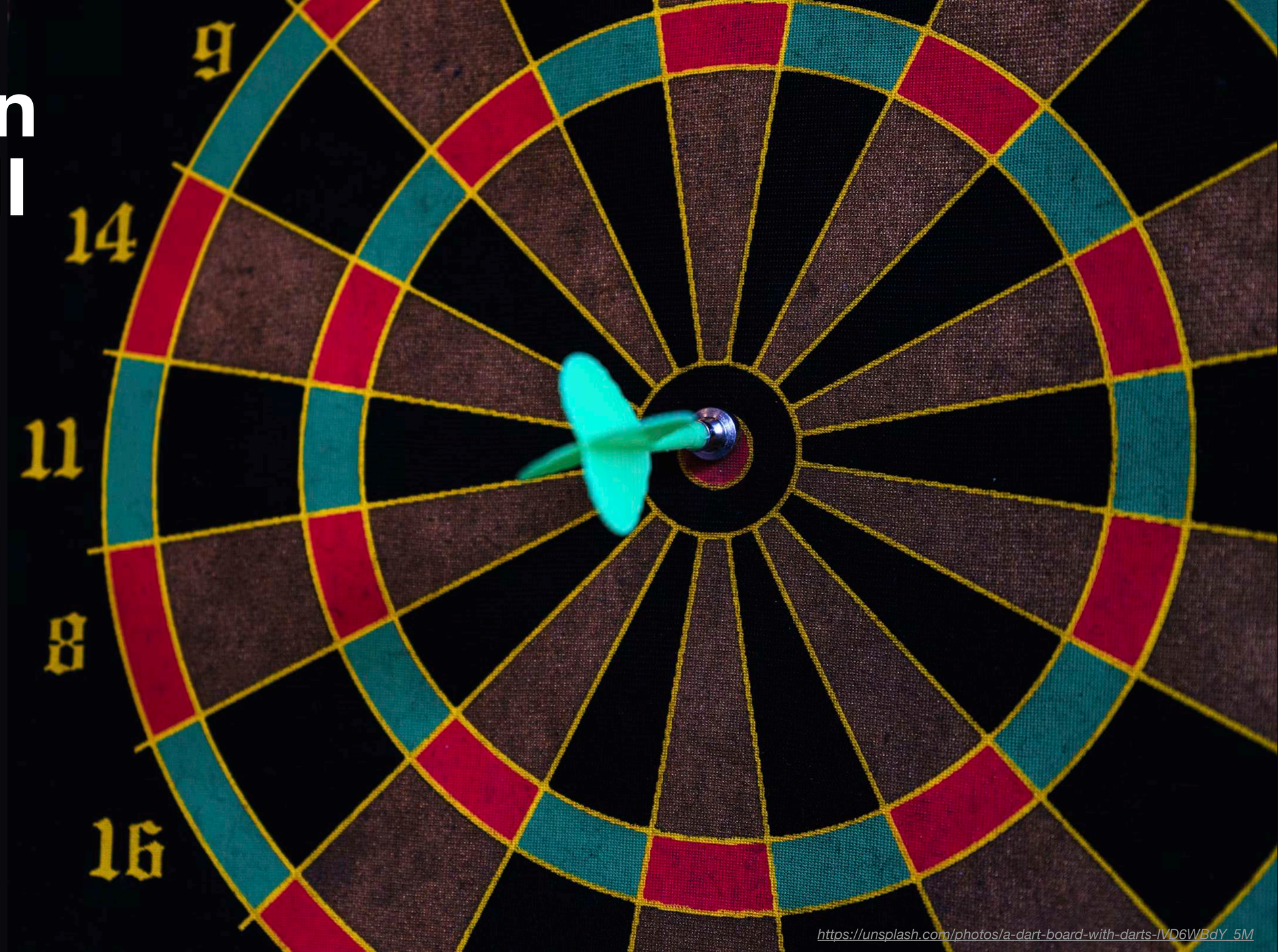
Detección Objetos II

Juan Carlos Arbeláez
jarbel16@eafit.edu.co

UNIVERSIDAD EAFIT®



https://unsplash.com/photos/a-dart-board-with-darts-IVD6WBdY_5M

Contenido

1. Introducción
2. Aplicaciones
3. Retos
4. Algoritmos
    1. R-CNN
    2. Faster R-CNN
    3. Single Shot Detector (SSD)
    4. You Only Look Once (Yolo)
    5. **DETR**
5. Métricas de evaluación
6. Taller

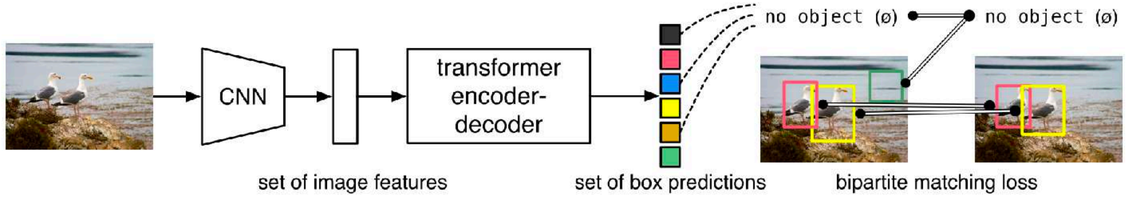
UNIVERSIDAD EAFIT

flaticon.com

- Presentada por Facebook en el 2020: combinación CNN + Transformer
- Elimina la necesidad de componentes “manuales” como: “anchor boxes” y “non-maximum suppression”
- End-to-end: realiza las predicciones directas sin necesidad de funciones adicionales

DETR

Detection Transformer



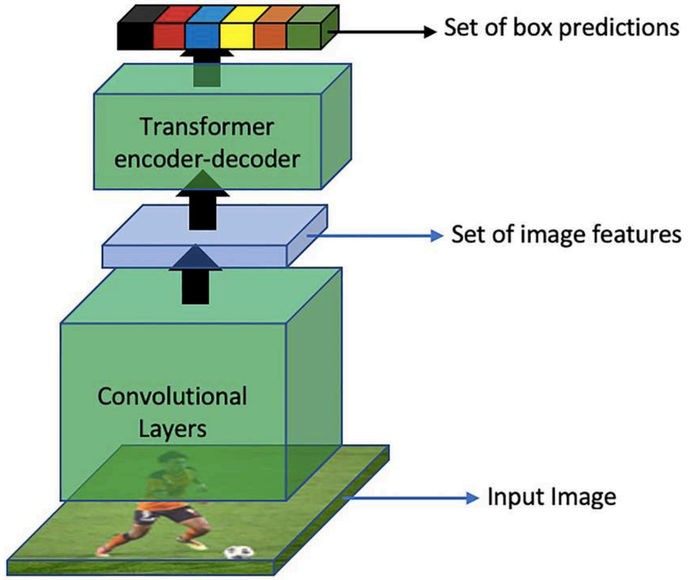
2020 - End-to-End Object Detection with Transformers

Predicciones

DETR

Dos elementos son necesarios para la predicción directa:

1. **Función de costo** que obligue la coincidencia única entre las predicciones y las etiquetas
2. **Arquitectura** que predice (en una sola pasada) un conjunto de objetos



UNIVERSIDAD EAFIT

2023 - Crop-saving with AI: latest trends in deep learning techniques for plant pathology

Función de costo

DETR

DETR predice un set de tamaño fijo de N predicciones (clase y coordenadas) pero N es mucho más grande que los objetos de la imagen

$$\hat{y} = \{\hat{y}_i\}_{i=1}^{N} \text{ con } \hat{y}_i = (\hat{c}_i, \hat{b}_i)$$

$\hat{y}$ usa un padding de no-objeto ($\varnothing$)

Se usa “Hungarian algorithm” para encontrar la asignación predicción-etiqueta con menor costo:

$$\hat{\sigma} = \arg\min_{\sigma \in \mathcal{S}_N} \sum_{i=1}^{N} L_{\text{match}}(y_i, \hat{y}_{\sigma(i)})$$

L_{match} penaliza la clasificación + localización donde el modelo predice positivos para objetos ($\mathbf{1}_{(c_i \neq \varnothing)}$)

$$L_{match}(y_i, \hat{y}_{\sigma(i)}) = -\mathbf{1}_{(c_i \neq \varnothing)} \hat{p}_{\sigma(i)}(c_i) + \mathbf{1}_{(c_i \neq \varnothing)} L_{box}(b_i, \hat{b}_{\sigma(i)})$$

El costo de localización L_{box} es una combinación linear de L_1 y *IoU Generalizado*

$$L_{box}(b_i, \hat{b}_{\sigma(i)}) = \lambda_{iou} L_{iou}(b_i, \hat{b}_{\sigma(i)}) + \lambda_{L1} \| b_i - \hat{b}_{\sigma(i)} \|_1$$

UNIVERSIDAD EAFIT

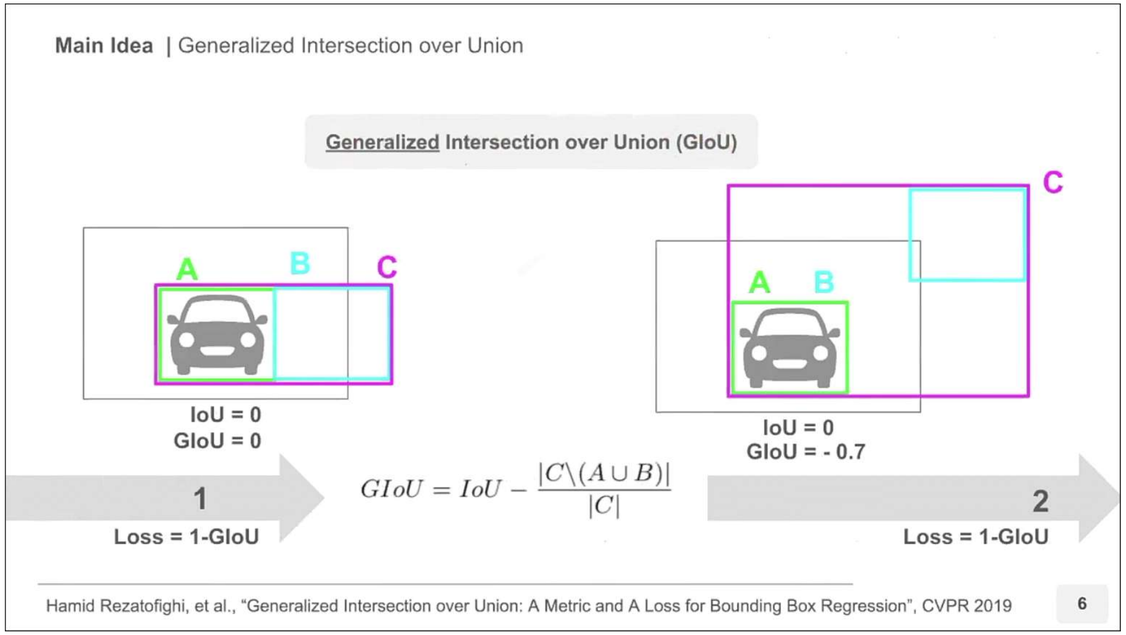
Más información de “Hungarian algorithm”

2020 - End-to-End Object Detection with Transformers

GIOU

Generalized Interception Over Union

- IOU tiene un problema principal: para todo $|A \cap B| = 0 \rightarrow IoU(A, B) = 0$ independiente de los lejos que se encuentren
- Usa el Convex-Hull C es la geometría más pequeña que envuelve A y B tener en cuenta la distancia cuando no se interceptan



UNIVERSIDAD EAFIT

2019 - Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression

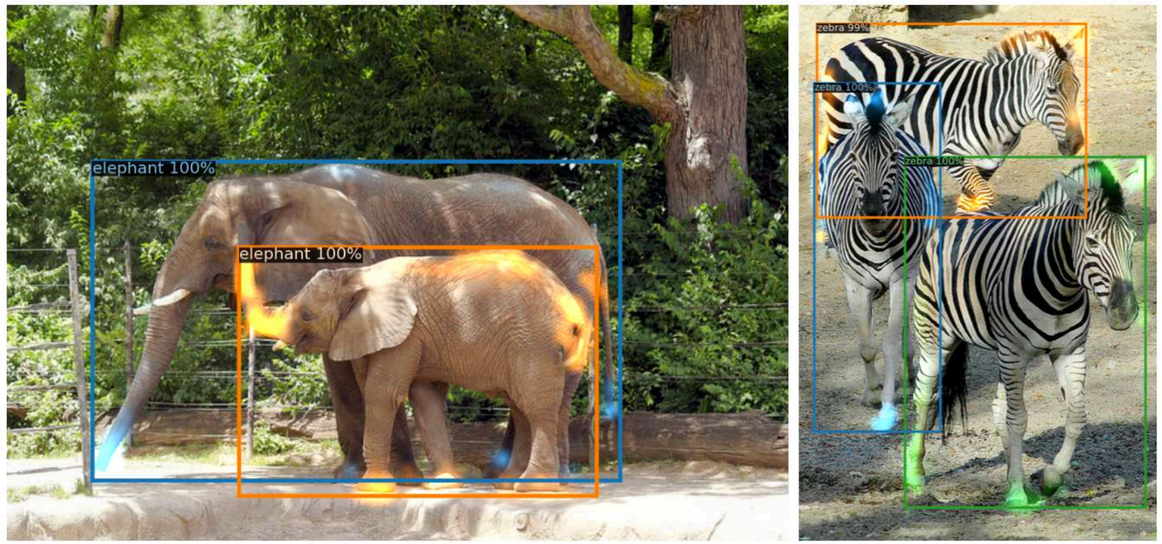
Hungarian loss

DETR

Función de costo a ser minimizada:

$$L_{hungarian}(y, \hat{y}) = \sum_{i=1}^{N} -\log \hat{p}_{\hat{\sigma}(i)}(c_i) + \mathbf{1}_{(c_i \neq \emptyset)} L_{box}(b_i, \hat{b}_{\hat{\sigma}(i)})$$

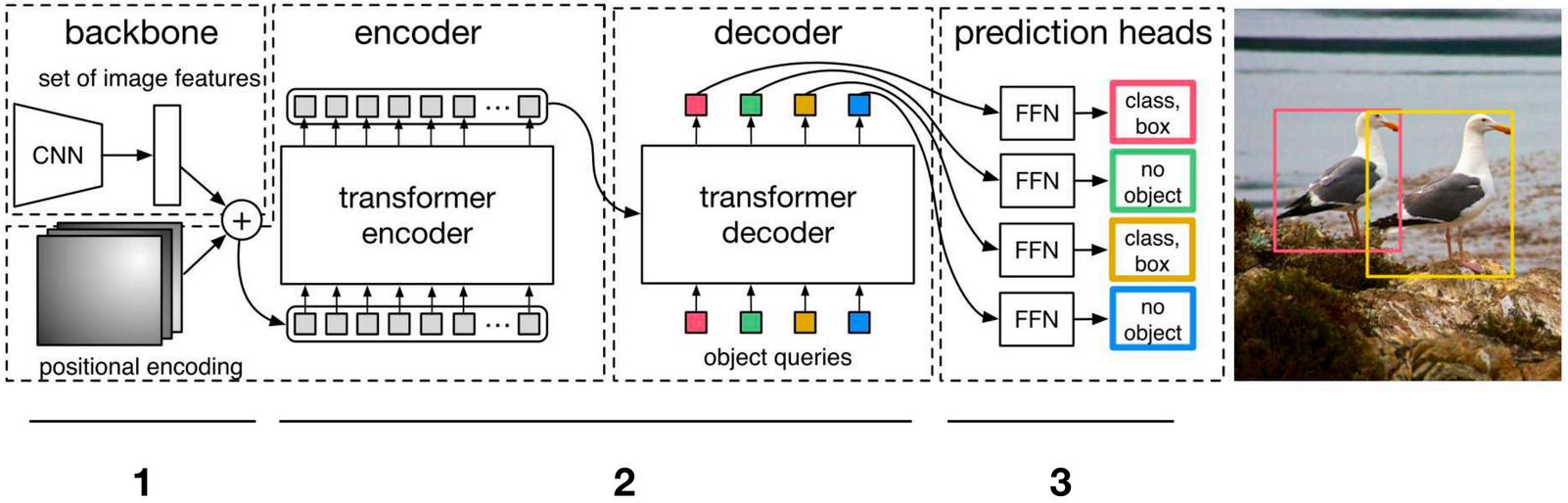
Clasificación + Detección de la mejor combinación encontrada con el algoritmo húngaro



UNIVERSIDAD EAFIT®

Arquitectura

DETR



UNIVERSIDAD EAFIT

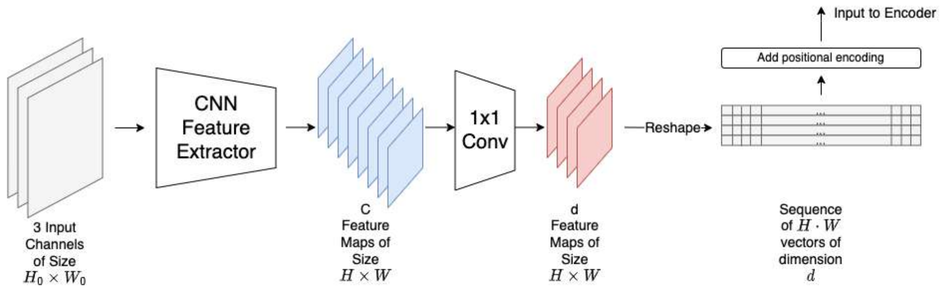
2020 - End-to-End Object Detection with Transformers

1. Backbone

Arquitectura

- Extractor de características CNN convencional (ResNet50)
- Mapa de características es transformado para pasar de un formato de secuencias

$$f \in \mathbb{R}^{C \times H \times W} \rightarrow z \in \mathbb{R}^{D \times HW}$$



UNIVERSIDAD EAFIT

https://maucher.pages.mi.hdm-stuttgart.de/orbook/deeplearning/detr.html

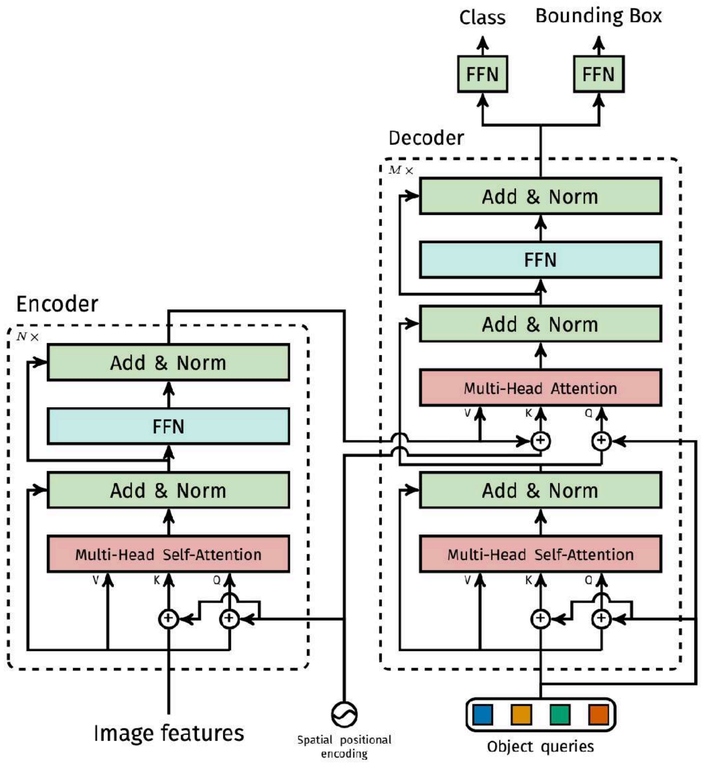
2. Transformer

Arquitectura

Transformer DETR consiste de multiples bloques de encoder y decoder (estándar con mínimas variaciones)

Encoder: “positional encodings” se agregan an K y Q. Ya que V representa características de la imagen

Decoder: Recibe N “object queries” (“learnt positional encodings”) que son transformados en un “output embedding” usado por la cabeza de predicción



UNIVERSIDAD EAFIT

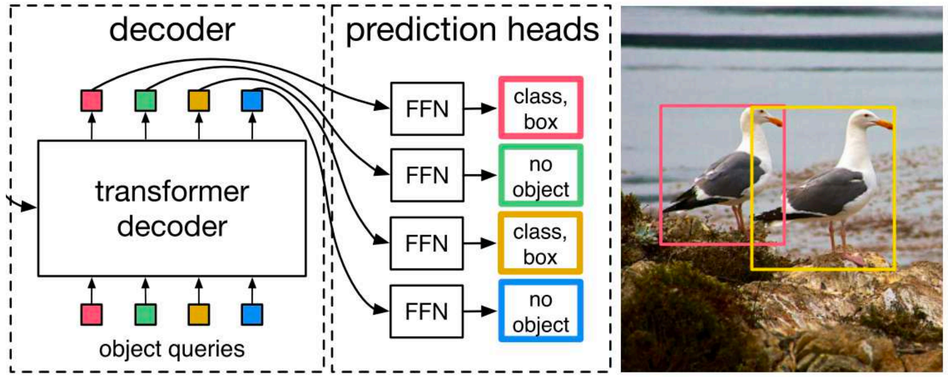
2020 - End-to-End Object Detection with Transformers

3. Cabeza de predicción

Arquitectura

FC de 3 capas con activación ReLU que predice:

1. Las clases (softmax) más una clase especial de “no objeto” (∅)
2. El centro normalizado (sigmoid) de las coordenadas,
3. El ancho y largo de los rectángulos



UNIVERSIDAD EAFIT

2020 - End-to-End Object Detection with Transformers

Contenido

1. Introducción
2. Algoritmos
3. **Métricas de evaluación**
4. Aplicaciones
5. Retos
6. Taller

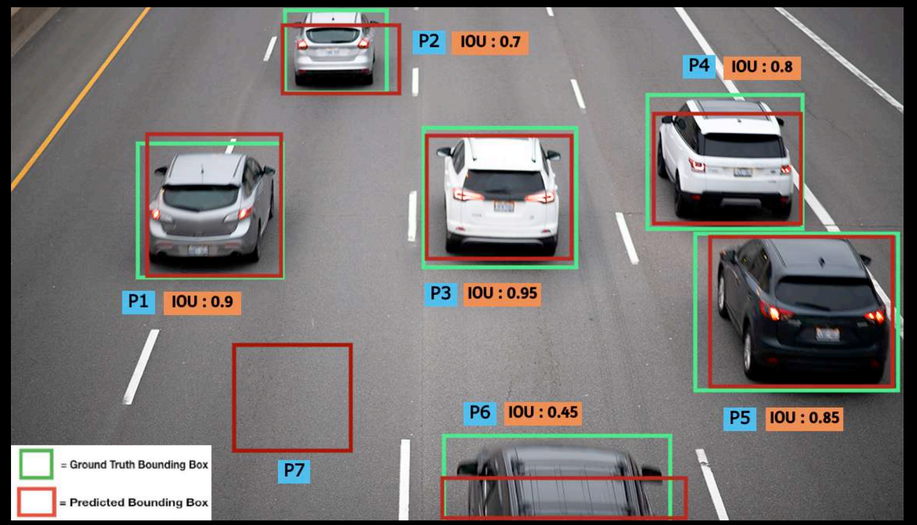
UNIVERSIDAD EAFIT

flaticon.com

mAP

Mean Average Precision

- El **mAP** es el promedio de **AP** de cada clase
- **AP** es el AUC de la curva precisión-sensibilidad



UNIVERSIDAD EAFIT®

https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

AP

Average Precision

AP No es el promedio de la precisión de las distintas clases

AP se calcula con la ayuda de otras métricas como IoU, matriz de confusión, precisión y la sensibilidad

$$\textbf{mAP} = \frac{1}{n}\sum_{h=1}^{n} \textbf{AP}_k$$

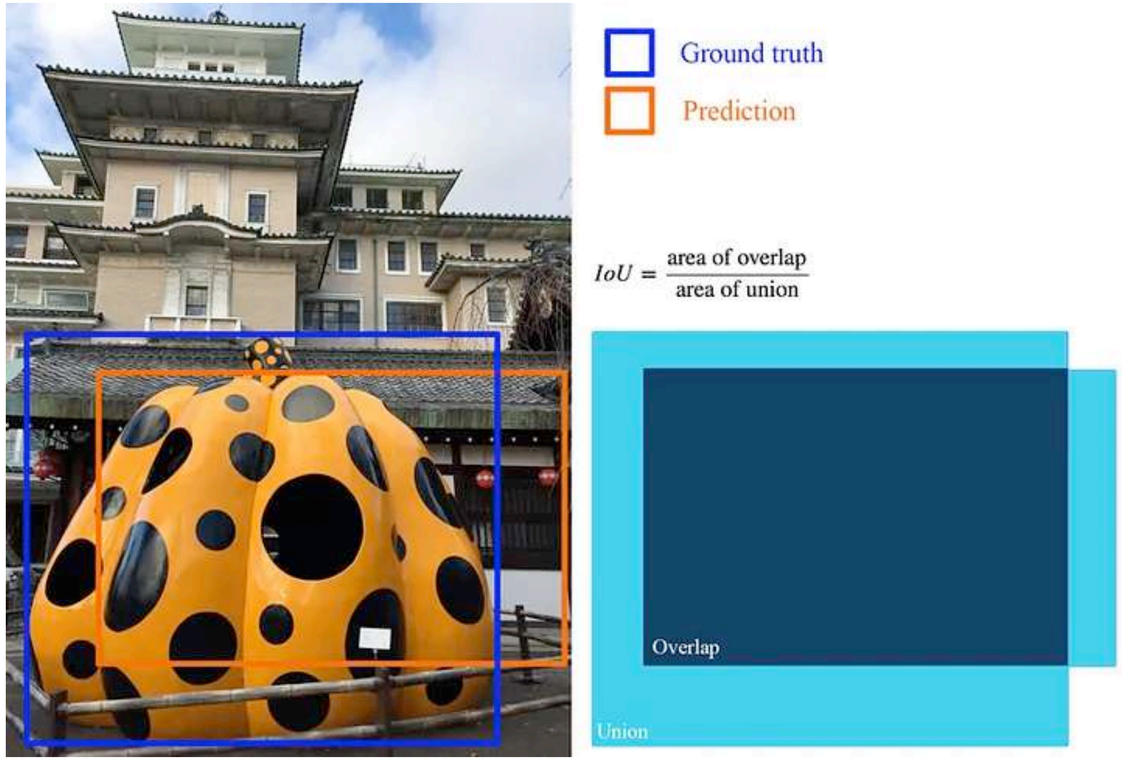
$\textbf{AP}_k$: Average Precision de la clase k

n: EL número de clases

IOU

Interception Over Union

- Lo utilizamos para medir cuánto se solapa el rectángulo predicho con el real
- **Predefinimos un umbral de IoU** por ejemplo, 0,5 para clasificar si la predicción es un verdadero positivo o un falso positivo



UNIVERSIDAD EAFIT

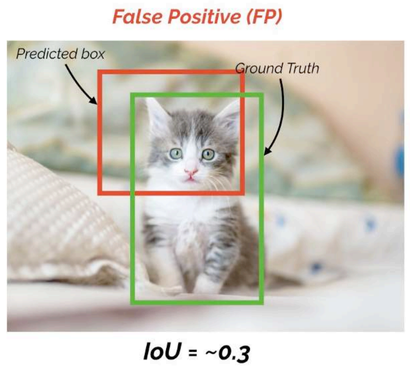
https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

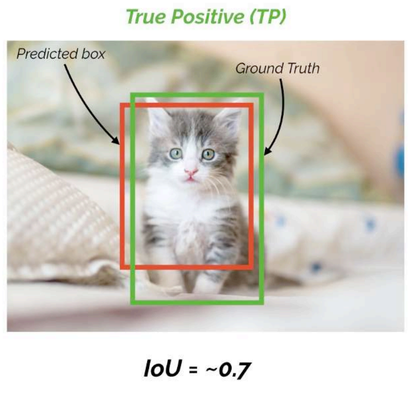
AP

Average Precision

@IOU = 0.5

Falso positivo: El modelo predijo que existía un rectángulo en una posición determinada pero se equivocó





Verdadero positivo: El modelo predijo que existía un rectángulo en una posición y acertó

UNIVERSIDAD EAFIT

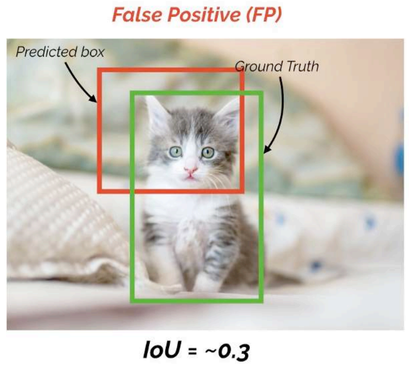
https://www.v7labs.com/blog/mean-average-precision

AP

Average Precision

@IOU = 0.5

Falso negativo: El modelo no predijo un rectángulo en una posición determinada y se equivocó

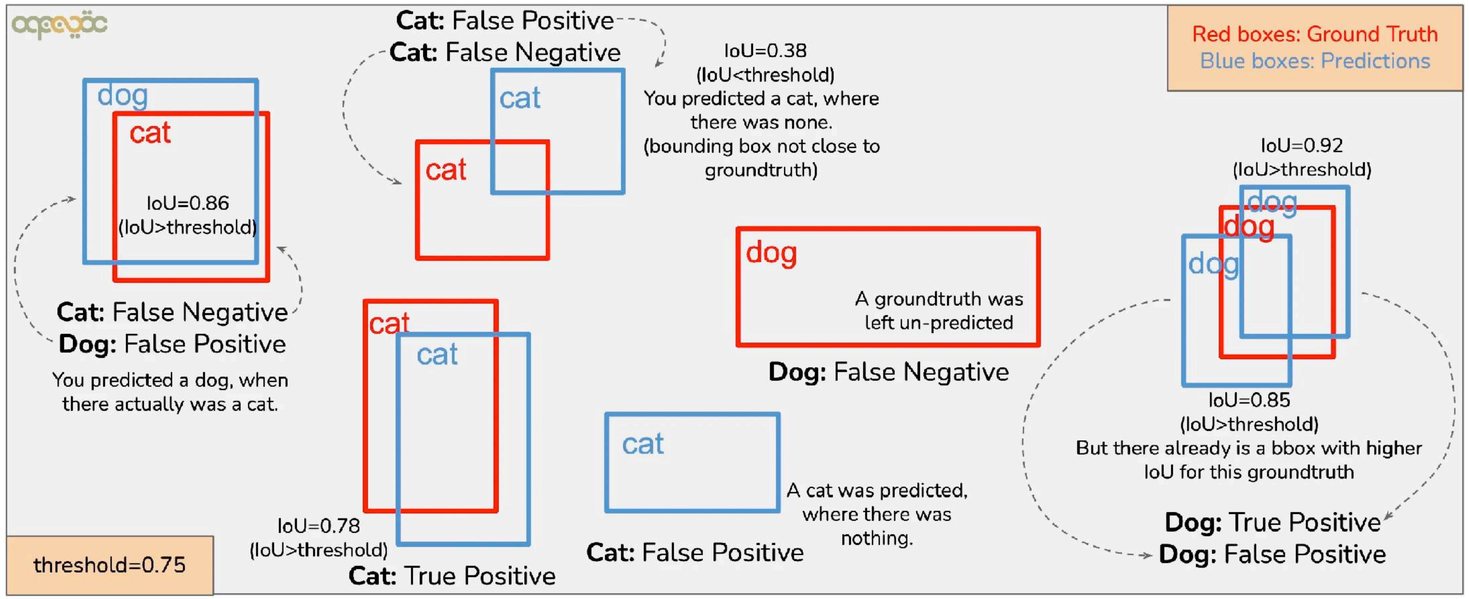


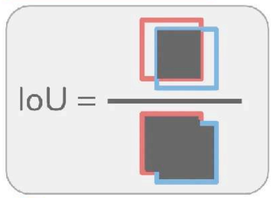
Verdadero negativo: Corresponde al fondo, la zona sin rectángulos, y **no se utiliza para calcular las métricas finales**

UNIVERSIDAD EAFIT

https://www.v7labs.com/blog/mean-average-precision

Object Detection and Localization - IoU, True Positive, False Positive, False Negative





@_aqeelanwar

aqeelanwarmalik

Threshold	Class	# GroundTruth	# predictions	TP	FP	FN	Precision	Recall
0.75	Cat	3	3	1	2	2	1/3	1/3
	Dog	2	3	1	2	1	1/3	1/2
0.35	Cat	3	3	2	1	1	2/3	2/3
	Dog	2	3	1	2	1	1/3	1/2

UNIVERSIDAD EAFIT

https://towardsdatascience.com/what-is-average-precision-in-object-detection-localization-algorithms-and-how-to-calculate-it-3f330efe697b

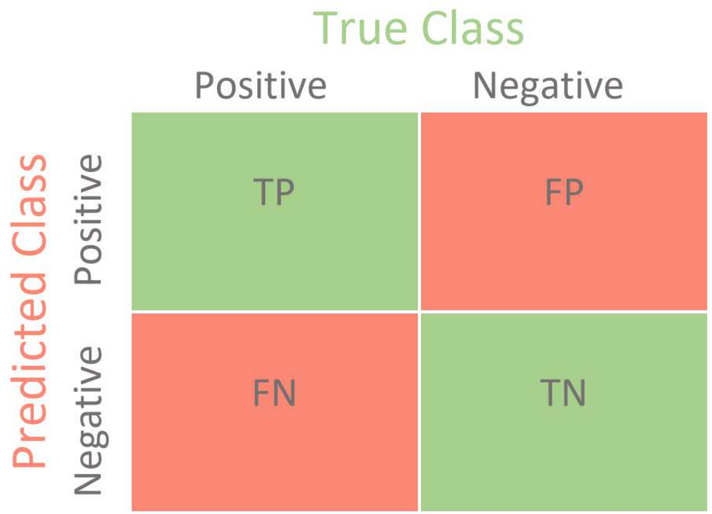
Sensibilidad y precisión

Sensibilidad: ¿Qué proporción de positivos reales se identificó correctamente?

$$\frac{TP}{TP + FN}$$

Precisión: ¿Qué proporción de predicciones positivas fue realmente correcta?

$$\frac{TP}{TP + FP}$$



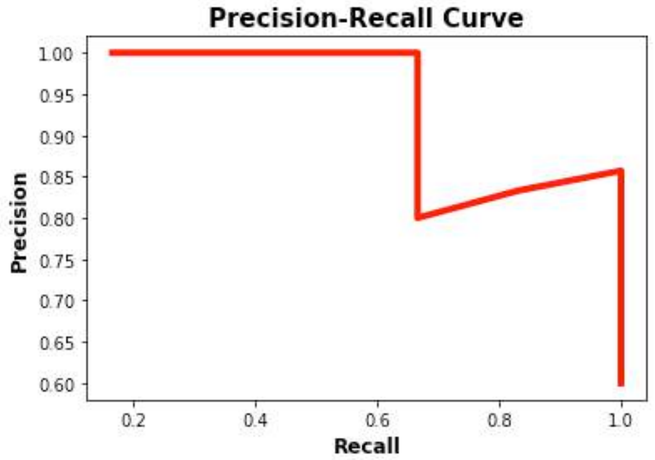
Matriz de confusión para la clasificación binaria

UNIVERSIDAD EAFIT

https://www.sciencedirect.com/topics/engineering/confusion-matrix

Precision-Recall Curve

- Para evaluar un modelo, hay que examinar tanto la precisión como la sensibilidad
- La precisión y la recuperación suelen estar en “tensión”
- Una curva de precisión-sensibilidad representa el valor de la precisión frente a la sensibilidad para distintos umbrales de confianza.



UNIVERSIDAD EAFIT

https://blog.paperspace.com/mean-average-precision/

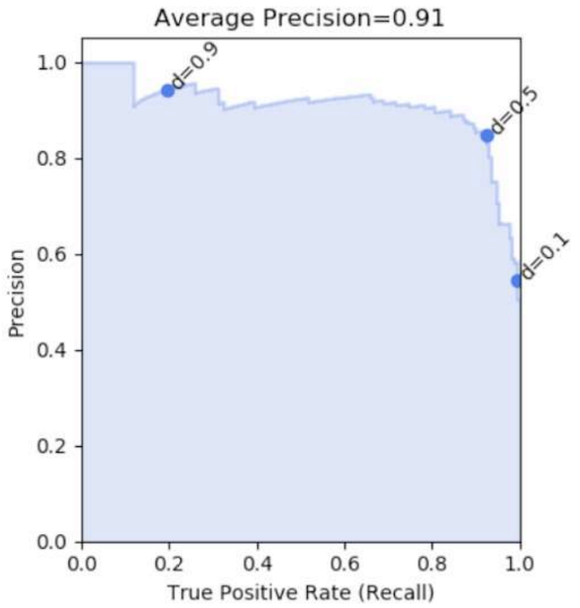
Average Precision

AP: Área bajo la curva PR de una clase

$$\text{AP} = \int_{r=0}^{1} p(r)$$

mAP: Promedio de AP de las clases

$$\text{mAP} = \frac{1}{n} \sum_{h=1}^{n} \text{AP}_k$$



UNIVERSIDAD EAFIT

Contenido

1. Introducción
2. Algoritmos
3. Métricas de evaluación
4. **Aplicaciones**
5. Retos
6. Taller

UNIVERSIDAD EAFIT

flaticon.com

Retail

- Conteo de personas
- Información sobre cómo pasan el tiempo los clientes y su afluencia
- Comprender la interacción y la experiencia del cliente
- Optimizar la distribución de la tienda y hacer más eficientes las operaciones
- Reducir el tiempo de espera en las tiendas minoristas

UNIVERSIDAD EAFIT

https://medium.com/@tagxdata/introduction-to-object-detection-for-computer-vision-and-ai-73a955b4d837

Robótica

- Los robots autónomos dependen de la detección de objetos para reconocer peatones, señales de tráfico, otros vehículos, etc.
- Por ejemplo, el *autopilot* de *Tesla* utiliza en gran medida la detección de objetos para percibir el entorno, como vehículos que circulan en sentido contrario u obstáculos.

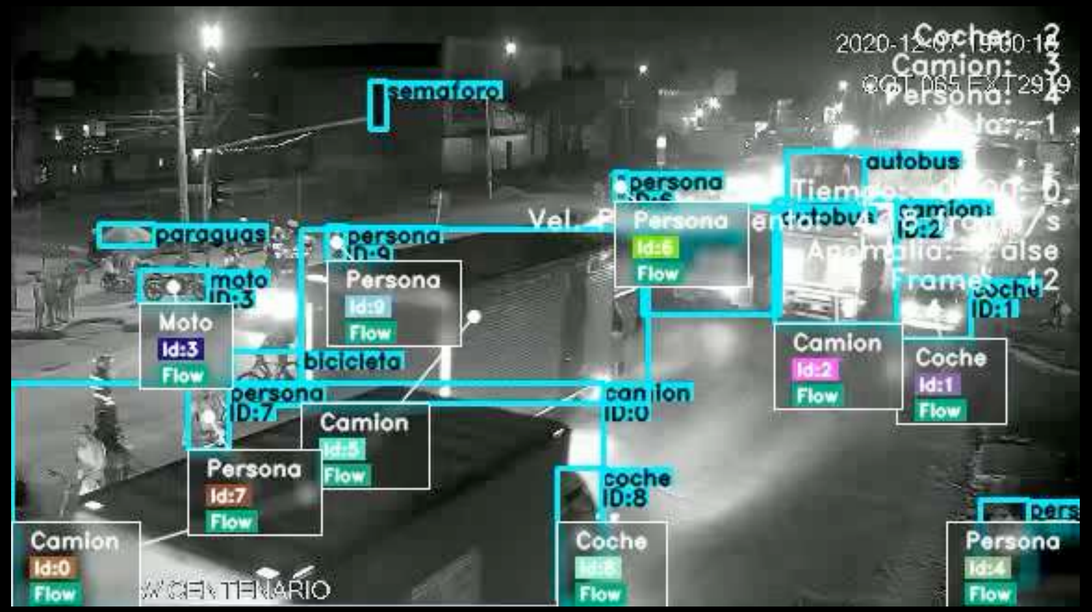
https://medium.com/@tagxdata/introduction-to-object-detection-for-computer-vision-and-ai-73a955b4d837
https://marketrealist.com/2019/08/tesla-autopilot-troubles-go-way-beyond-crashes/

Video vigilancia

- Seguimiento de personas a la vez en tiempo real (punto de partida para tracking)
- Rendimiento y seguridad de los trabajadores, tráfico peatonal
- Detección en cámaras de CCTV de personas en zonas peligrosas, objetos en bordes, personas en zonas no autorizadas, …

Monitoreo de tráfico

- Análisis del tráfico (punto de partida para tracking)
- Tipos de vehículos (motos, peatones, taxis, …), afluencia y comportamiento en las vías (giros, paros, velocidades)
- Accidentes, detención en zonas peligrosas



UNIVERSIDAD EAFIT

Movilab - vaico.com.co

Agricultura

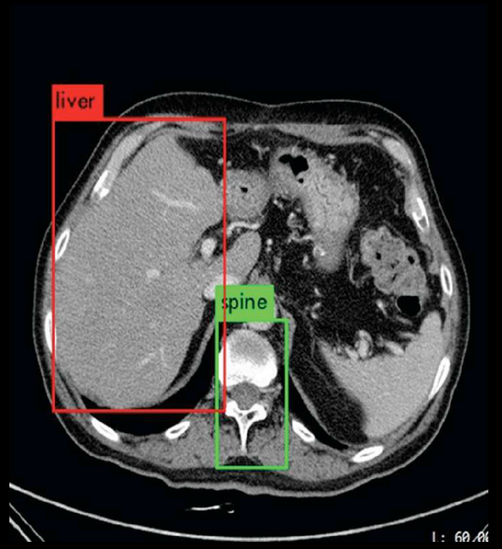
- Conteo de células, frutas, animales, flores, hojas
- Monitoreo de animales, comportamiento basado en movimientos
- Evaluación de la calidad de los productos agrícolas
- Agricultura de precisión
- Detección de enfermedades en plantas y animales

UNIVERSIDAD EAFIT

2018 - Detecting and Counting Sheep with a Convolutional Neural Network

Cuidado de la salud

- “Medical object detection” es la tarea de identificar objetos médicos en una imagen
- Los diagnósticos médicos se basan en gran medida en el estudio de imágenes, escáneres y fotografías
- La detección de objetos en tomografías y resonancias magnéticas en una herramienta para el diagnóstico de enfermedades



UNIVERSIDAD EAFIT®

2019 - A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images

Contenido

1. Introducción
2. Algoritmos
3. Métricas de evaluación
4. Aplicaciones
5. **Retos**
6. Taller

UNIVERSIDAD EAFIT

flaticon.com

Retos

- Doble prioridad, clasificación y localización de objetos: La detección de objetos tiene dos objetivos, la clasificación y la localización
- Se utiliza una función de pérdida multitarea que penaliza tanto los errores de clasificación como los de localización

Función de costo (weighted sum λ)

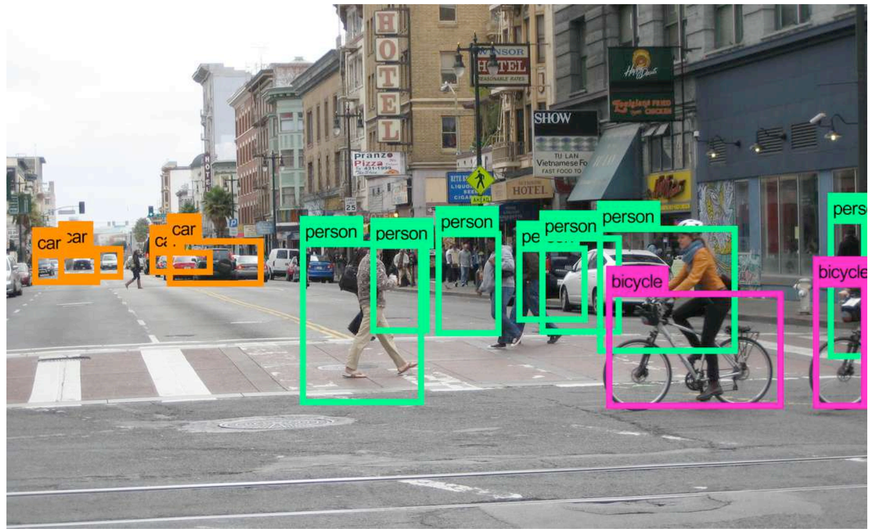
Clasificación impone una pérdida logarítmica
Localización es una pérdida S*mooth L_1-loss* para los cuatro componentes que definen el rectángulo

$$\ell(p, u, t^u, v) = \ell_c(pu) + \lambda_{[u \geq 1]} \ell_l(t^u, v)$$

Retos

Detección en tiempo real

- Muchas de las aplicaciones vistas requieren análisis en tiempo real
- Los vídeos suelen grabarse a un mínimo de 24 fps
- los algoritmos actuales de detección de objetos intentan encontrar un equilibrio entre velocidad y precisión

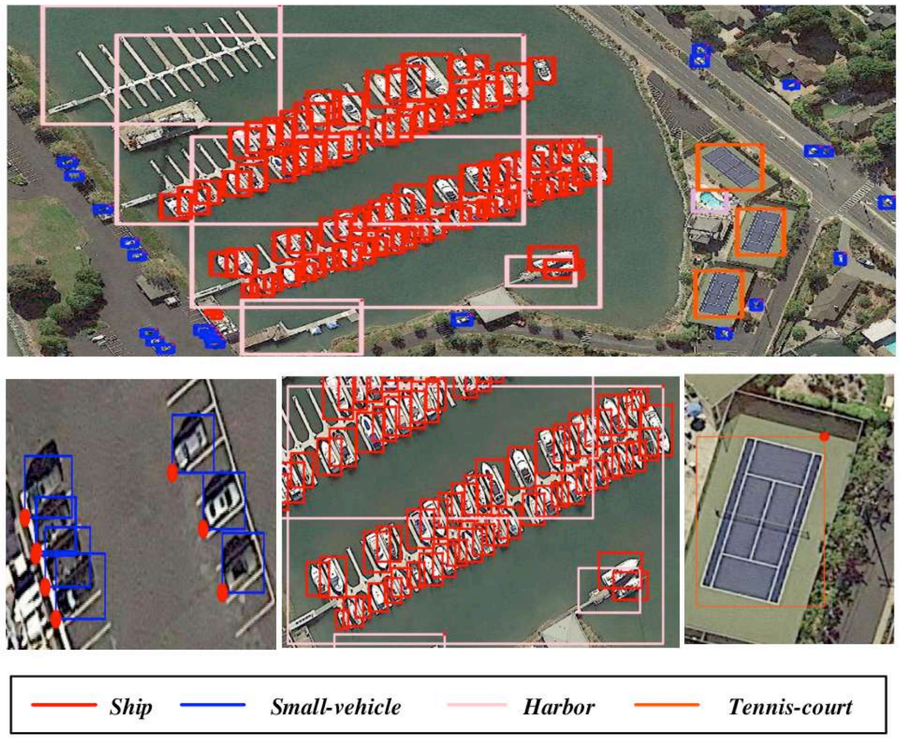


https://medium.com/analytics-vidhya/object-detection-using-regions-with-cnn-features-557392e22f84

Retos

Múltiples escalas y relaciones de aspecto

- Los elementos de interés pueden aparecer en una amplia gama tamaños y relación entre ancho y alto (*Aspect ratio*)
- Las imágenes con gran diferencia de tamaños son un reto hoy en día



UNIVERSIDAD EAFIT

2019 - A 2 RMNet : Adaptively Aspect Ratio Multi-Scale Network for Object Detection in Remote Sensing Images

Retos

Datos limitados

- La limitada cantidad de datos anotados que se dispone para la detección de objetos es un obstáculo importante
- Se requiere una alta cantidad de datos para la correcta localización de los objetos

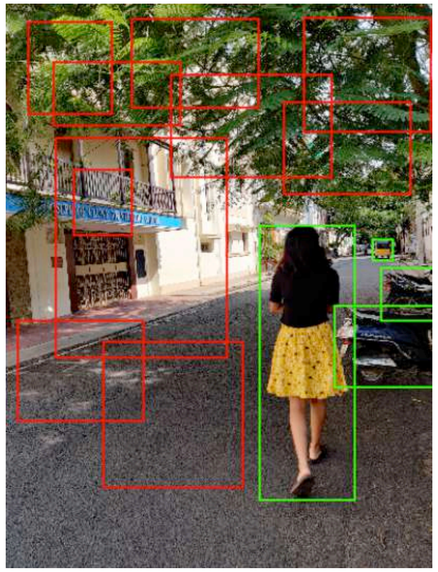
UNIVERSIDAD EAFIT

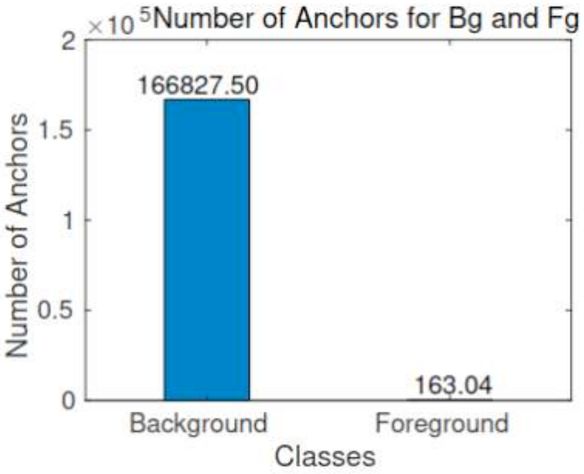
2021 - Object Detection in Densely Packed Scenes via Semi-Supervised Learning with Dual Consistency

Retos

Class imbalance

En una fotografía típica lo mas probable es que contenga unos pocos objetos principales y que el resto de la imagen esté llena de fondo





UNIVERSIDAD EAFIT

https://neptune.ai/blog/imbalanced-data-in-object-detection-computer-vision

Contenido

1. Introducción
2. Algoritmos
3. Métricas de evaluación
4. Aplicaciones
5. Retos
6. **Taller**

UNIVERSIDAD EAFIT

flaticon.com

Ejercicio

Realizar un modelo para la localización de objetos (Regresión de las coordenadas de los Bounding Box

Pasos:

1. Descargar dataset y etiquetas en imágenes
2. Crear matriz de etiquetas (y)
3. Modelo: Extractor de características (Convoluciones) + Cabeza (regresión de 4 variables)
4. Entrenar con función de pérdida MSE (Mean Squared Error)

UNIVERSIDAD EAFIT

Ejercicio

1. Descargar dataset de Teams
2. Graficar etiquetas en imágenes: Las etiquetas están dadas por dos puntos

   Inicial (x_top, y_top)

   Final (x_bottom, y_bottom)

UNIVERSIDAD EAFIT

Li, F.-F., Andreeto, M., Ranzato, M., & Perona, P. (2022). Caltech 101 (1.0) [Data set]. CaltechDATA. https://doi.org/10.22002/D1.20086

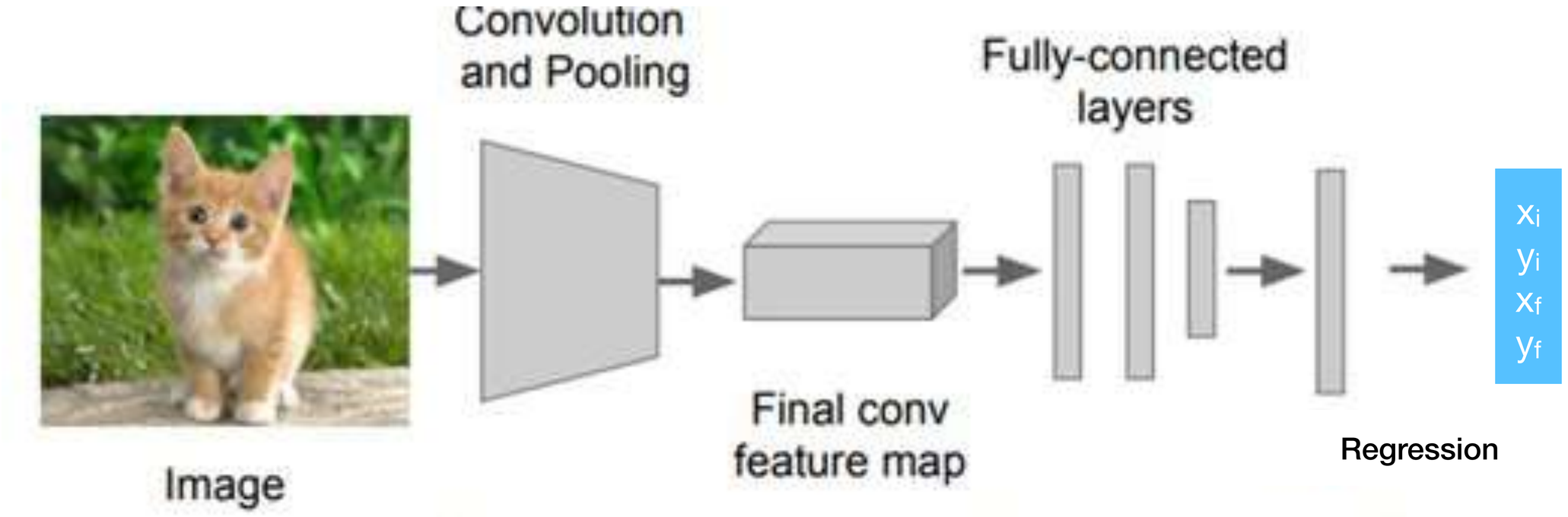
Ejercicio

Generar Etiquetas:

Normalizar por el tamaño de la imagen, de esta forma el modelo predice valores entre 0-1 por cada coordenada

$$\left[\frac{x_{...}}{w}, \frac{y_{...}}{h}\right]$$

Ejercicio


Convolution
and Pooling
Fully-connected
layers
x_i
y_i
x_f
y_f
Image
Final conv
feature map
Regression

Lecturas

2022 - CenterNet++ for Object Detection

2022 - BEIT: Bert pre-training of image transformers

UNIVERSIDAD EAFIT®

Referencias

2014 - Rich feature hierarchies for accurate object detection and semantic segmentation

2004 - Efficient Graph-Based Image Segmentation

2013 - Selective Search for Object Recognition

2015 - Fast R-CNN

2016 - SSD: Single Shot MultiBox Detector

2016 - You Only Look Once Unified Real-Time Object Detection

2017 - YOLO9000 Better, Faster, Stronger

2018 - YOLOv3 An Incremental Improvement

2020 - YOLOv4 Optimal Speed and Accuracy of Object Detection

2020 - End-to-End Object Detection with Transformers

2019 - Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression