

- donde $p_i(c)$ es la probabilidad condicional de clase para la clase c en la celda i .

cuadros delimitadores de las detecciones y de las anotaciones.

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{obj} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \\ + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{obj} (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2$$

- La pérdida de confianza: Cada que se detecta un objeto en el cuadro delimitador como observamos en la siguiente ecuación:

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{obj} (C_i - \hat{C}_i)^2$$

donde C_i es la puntuación de confianza del cuadro delimitador j en la celda i . En caso que no encuentre un objeto tenemos:

$$+ \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{noobj} (C_i - \hat{C}_i)^2$$

La pérdida final, es la suma de las anteriores expresiones:

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{obj} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \\ + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{obj} (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \\ + \sum_{i=0}^{S^2} 1_i^{obj} \sum_{c \in \text{clases}} (p_i(c) - \hat{p}_i(c))^2 \\ + \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{obj} (C_i - \hat{C}_i)^2 \\ + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_i^{noobj} (C_i - \hat{C}_i)^2$$

A continuación se mencionaran algunas versiones del modelo que sirvieron de ayuda para realizar las implementaciones.

II-A. Arquitectura - YOLOv3

YOLOv3 tiene como principal intención aumentar la precisión en la detección de objetos y la principal mejora se debe a la arquitectura de red usada, que es Darknet-53. La extracción de características en la versión YOLOv3, es el resultado de una red híbrida entre la versión de YOLOv2 y la red residual. A continuación se muestra la puntos importantes implementados en YOLOv3:

- Contiene 53 capas convoluciones que han sido, cada una seguida por la capa de normalización por lotes y la activación de *LeakyReLU*.
- Una principal función es que ayuda a no perder muchas características de bajo nivel.
- Mejora las predicciones de las etiquetas múltiples.

Otra nueva implementación es que hay cambios en la predicción de los cuadros delimitadores, la red predice cuatro coordenadas para cada cuadro delimitador (t_x, t_y, t_w, t_h). Si el cuadro tiene como centro (c_x, c_y) , el ancho y altura como p_w, p_h , respectivamente, las predicciones correspondientes son las siguientes ecuaciones:

$$b_x = \sigma(t_x) + c_x$$

$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h}$$

donde b_x y b_y corresponden al centro del cuadro delimitador, b_h y b_w corresponden a las dimensiones.

III. IMPLEMENTACIÓN

IV. ENTRENAMIENTO

Algoritmos rápidos y precisos para la detección de objetos permitiría que las computadoras conduzcan automóviles sin necesidad de sensores.

DPM utilizan un enfoque de ventana deslizante donde el clasificador se ejecuta en ubicaciones espaciadas uniformemente en toda la imagen.

Enfoque más recientes como R-CNN utilizan

El modelo Yolo procesa imágenes en tiempo real 45 imágenes por segundo.

Yolo comete más errores de localización pero es menos probable que prediga falsos positivos en segundo plano. Yolo aprende representaciones genetales de objetos.

Supera a otros métodos de detecciones, como son DPM y R-CNN

Sobre YOLO: Una red convolucional predice simultáneamente múltiples cuadros delimitadores y probabilidades de clase para cada uno de estos cuadros. Además entrena con imagenes completas y optimiza directamente la detección. Un primer beneficio es que es extremadamente rápido. La red base funciona procesando 45 frame por segundo, esto significa que podemos procesar la transmisión e video en tiempo real con menos de 25 milisegundos. Además YOLO logro más del doble de la precisión promedio de otros algoritmos en tiempo real.

Yolo aprende representaciones generalizables de objetos. Dado que YOLO es altamente generalizables es menos probable que se rompa cuando se aplica a nuevos dominios o entradas inesperadas. YOLO todavía está por detrás de los sistemas de detección de última generación en precisión. Si bien puede identificar rápidamente objetos en imágenes lucha por localizar con precisión algunos objetos, especialmente pequeños.

Hay una variedad de modelos previamente entrenados también están disponibles para descargar.

También predice todos los cuadros delimitadores en todas las clases para una imagen simultaneamente. Esto significa que

nuestra red razona globalmente sobre la imagen completa y todos los objetos de la imagen.

V. PROCEDIMIENTO

1.- Divide la imagen de entrada en una cuadrícula $S \times S$. Si el centro de un objeto cae en una celda de la cuadrícula, esa celda de la cuadrícula es responsable de detectar ese objeto. Cada celda de la cuadrícula predice B cuadros delimitadores y un valor de confianza para esas cajas. Estos puntajes de confianza reflejan que tan seguro está el modelo que la caja contiene un objeto y también qué tan precisa cree que es la caja que predice.

2.- Formalmente se define la confianza como $Pr(Object) * IOU_{pred}^{truth}$. Si no existe un objeto en esa celda, las puntuaciones de confianza deben ser cero. De lo contrario, queremos que la puntuación de confianza sea igual a la intersección sobre unión (IOU) entre el cuadro predicho y la verdad fundamental.

3.- Cada cuadro delimitador consta de 5 valores, los cuales son x, y, w, h y la *confianza*. Las coordenadas (x,y) representan el centro del cuadro, w y h representan la anchura y la altura que se pronostican en relación con la imagen completa.

4.- Cada celda de la cuadrícula también predice C probabilidades de clases condicionales, $Pr(Class_i|Object)$. Estas probabilidades están condicionadas en la celda de la cuadrícula que contiene un objeto.

5.- En el momento de la prueba, multiplicamos las probabilidades condicionales de clase y las predicciones de confianza de la caja individual.

$$Pr(Class_i|Object) * Pr(Object) * IOU_{pred}^{truth} = Pr(Class_i) * IOU_{pred}^{truth}.$$

VI. BENEFICIOS

Es un sistema muy rápido, debido a que reduce la detección a un problema de regresión lineal y esto conlleva a que no se requiera un modelo demasiado complejo. Para realizar las predicciones, a diferencia de otros métodos, este modelo toma la imagen completa, en lugar de solo tomar regiones. Esto hace que se limiten los errores a la hora de reconocer las clases de objetos que hay en la imagen.

Un beneficio importante que ya se menciona con anterioridad es que aprende representaciones generalizables de objetos, lo que permite introducir datos nuevos y tenga menos probabilidades de fallo que las otras técnicas.

VII. LIMITACIONES

YOLO funciona mal en objetos pequeños, particularmente aquellos que aparecen en grupos, ya que cada celda predice solo dos cuadros delimitadores de la misma clase. Aunque la función de pérdida predice la raíz cuadrada de la altura y el peso del cuadro delimitador en lugar de la altura y el peso directamente en un intento de resolver la ecualización

de error para cuadros grandes y pequeños, esta solución soluciona parcialmente el problema, lo que resulta en un número significativo de errores de localización. Además, YOLO no detecta correctamente los objetos de formas nuevas o poco comunes, ya que no se generaliza mucho más allá de los cuadros delimitadores en el conjunto de entrenamiento. Para utilizar YOLO en un sistema de videovigilancia en tiempo real es necesario prestaciones computacionales por encima de la media, en su versión ligera la disminución de la precisión lo hace ineficiente en problemas de este tipo. ffffffffffffffffffffff

REFERENCIAS

- [1] Matthew N., Carrie J., Michael J. Lyons, Miyuki K., Hanae I. y Jiro(2010). *Evidence and a computational explanation of cultural differences in facial expression recognition*.
- [2] Michael J. Lyons, Julien Budynek y Shigeru A.(1999). *Automatic Classification of Single Facial Images*.
- [3] Michael J. Lyons y Shigeru Akamatsu(1998). *Coding Facial Expressions with GaborWavelets*.Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.
- [4] Mohammad Hag., Saman Zon. y Mohamed Adbel(2015). *Trustworthy cloud-based and cross-enterprise biometric identification*.Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.
- [5] Libin Yang, William Rea y Alethea Rea(2017). *Financial Insights from the Last Few Components of a Stock Market PCA* .Centre for Applied Statistics, University of Western Australia, 35 Stirling Hwy, Crawley 6009, Australia; alethea.rea@uwa.edu.au