espol Facultad de Ingeniería en Electricidad y Computación

Axel Auza
aaauza@espol.edu.ec

Luis Carrasco
luiacarr@espol.edu.ec

DETECCIÓN Y CLASIFICACIÓN DE OBJETOS USANDO LA ARQUITECTURA YOLOV4 SOBRE IMÁGENES RGB MÁS UN CANAL ADICIONAL

PROBLEMA

Actualmente se han propuesto una gran cantidad de implementaciones que pueden usarse para la detección y clasificación de objetos que hacen uso de CNN, Transformers, implementaciones convencionales que trabajan solo sobre imágenes RGB (3 canales). Esto es una limitante ya que existen características presentes en la escena de la imagen que se podrían aprovechar si se usan imágenes de más de tres canales.

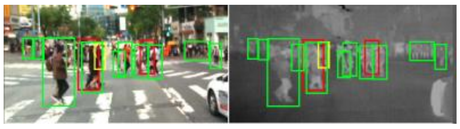

Figura 1. Detección de objetos sobre imágenes RGB y Thermal

OBJETIVO GENERAL

Adaptar la arquitectura de YOLOv4 para que trabaje sobre imágenes RGB más un canal (imágenes de 4 canales) permitiendo la detección y clasificación de objetos. El cuarto canal a considerar corresponde a la componente *Thermal* (*LWIR*).

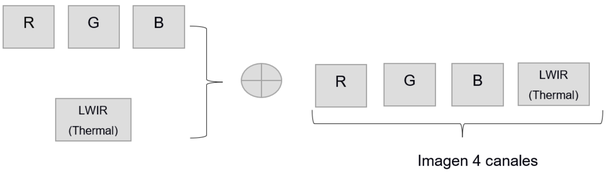

Figura 2. Generación de imágenes de cuatro canales (RGB + *Thermal*)

PROPUESTA

Para la solución de este proyecto se han propuesto cuatro componentes: bases de datos, pre-procesamiento, detección y clasificación de objetos sobre imágenes de cuatro canales, y validaciones y pruebas.

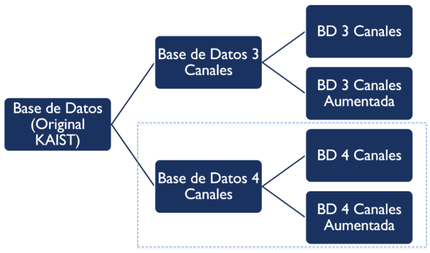

Figura 3. Componente Base de Datos: Generación de imágenes de cuatro canales

Figura 4. Componente Pre-Procesamiento: Conversión de anotaciones a YOLOv4

```
int flag;
if(c<3){
    flag=0;
}else if(c==3){
    flag=1;
}else{
    flag=-1;
}
```

Figura 5. Componente detección y clasificación de objetos sobre imágenes de cuatro canales: Generación de la arquitectura YOLOv4_4C

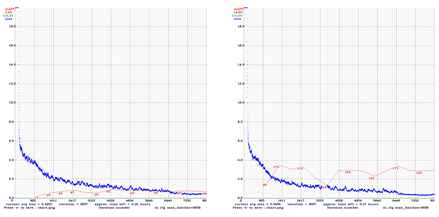

Figura 6. Componente Validación y Pruebas: Generación de resultados

RESULTADOS

Resultados obtenidos usando el *dataset* de *KAIST Multispectral Pedestrian Detection* en el escenario nocturno, con las arquitecturas YOLOv4 y YOLOv4_4C

Entrenamientos (escenario nocturno)	Average Precision % (persona)
RGB YOLOv3 (3 canales)	20
RGB YOLOv4 (3 canales)	13
RGB + *Thermal* YOLOv4_4C (4 canales)	42
RGB + *Thermal* + *Data Augmentation* YOLOv4_4C (4 canales)	51

- En la versión anterior de YOLOv3 se obtuvo una precisión del 20% para la clase persona en la noche, utilizando el mismo *dataset.*
- En la versión actual de YOLOv4 para la clase persona se obtiene una precisión del 13% usando este *dataset*, haciendo uso de solo imágenes RGB.
- En la versión propuesta de YOLOv4_4C usando la clase persona se logra obtener un 42% de precisión, que en comparación con los anteriores, es significativamente mejor.
- En la versión de YOLOv4_4C aplicando *data augmentation*, se logra mejorar la precisión anterior en nueve puntos, obteniendo un 51% de precisión para la clase persona.

CONCLUSIONES

- Al generar las imágenes con más de tres canales, previamente se debe evaluar el provecho que se obtendrá con el uso de el o los nuevos canales en las imágenes a utilizar, ya que en otro caso podría generar malos resultados y no se aprovecharía de forma correcta las características contenidas en las imágenes de más de tres canales.
- El uso del canal *LWIR* bajo condiciones nocturnas es sumamente beneficioso para la detección y clasificación de la clase persona, al poder agregarle más *features* a la red como la temperatura que permite diferenciar a una persona del fondo.
- Al usar técnicas de *data augmentation* como *crop y rotate,* considerar que las anotaciones se pueden sobresalir de la imagen generando errores al entrenar.
- Las técnicas de *data augmentation* aumentan significativamente la precisión para la clase con mayor número de objetos en el *dataset* pero su comportamiento es aleatorio con clases con objetos de menor presencia en el *dataset*.