



UNIVERSIDAD  
DE MÁLAGA



**ESCUELA DE INGENIERÍAS INDUSTRIALES**

## **Detección multimodal de objetos mediante deep-learning empleando la red AdapNet++ para fusión sensorial.**

**Autor :** Pablo Venema Rodríguez

**Tutor :** Ricardo Vázquez Martín.

**Cotutor :** Andrés Manuel Salas Espinales

# 1. Objetivos

El objetivo principal de este trabajo es el análisis de técnicas de transferencia de conocimiento sobre la arquitectura AdapNet++ para fusión multimodal utilizando imágenes del dataset UMA-SAR.

1. Imágenes térmicas.
2. Dataset autores.
3. UMA-SAR.

Fondo	Escombros	Civil	Pers. rescate	Victima	Vehículo civil	Vehículo rescate	Zapatos
-------	-----------	-------	------------------	---------	-------------------	---------------------	---------

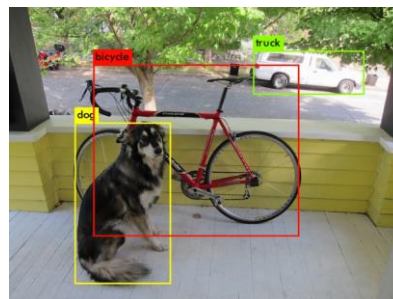
# 1. Introducción

## Clasificación

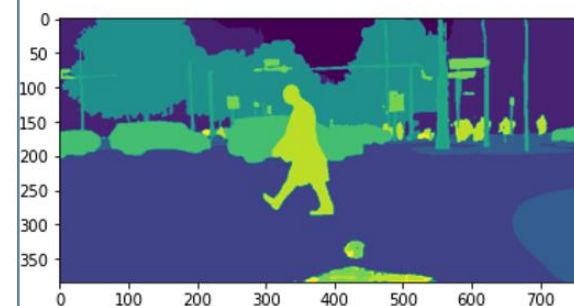


cabra	<b>camello</b>	pez
1%	<b>45%</b>	30%

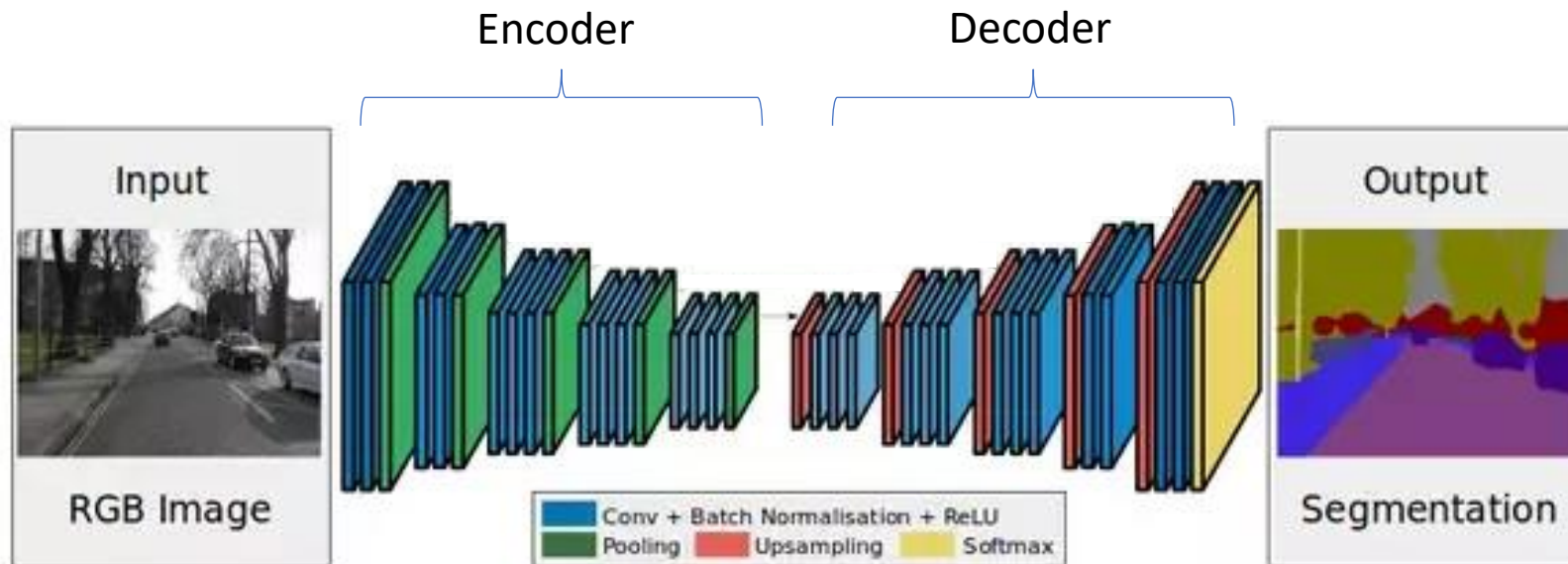
## Detección



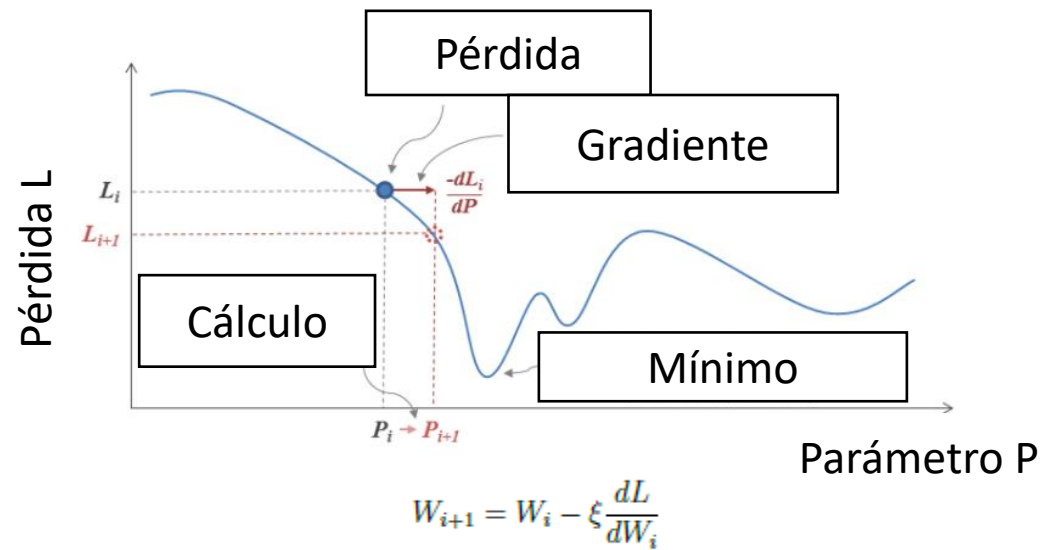
## Segmentación



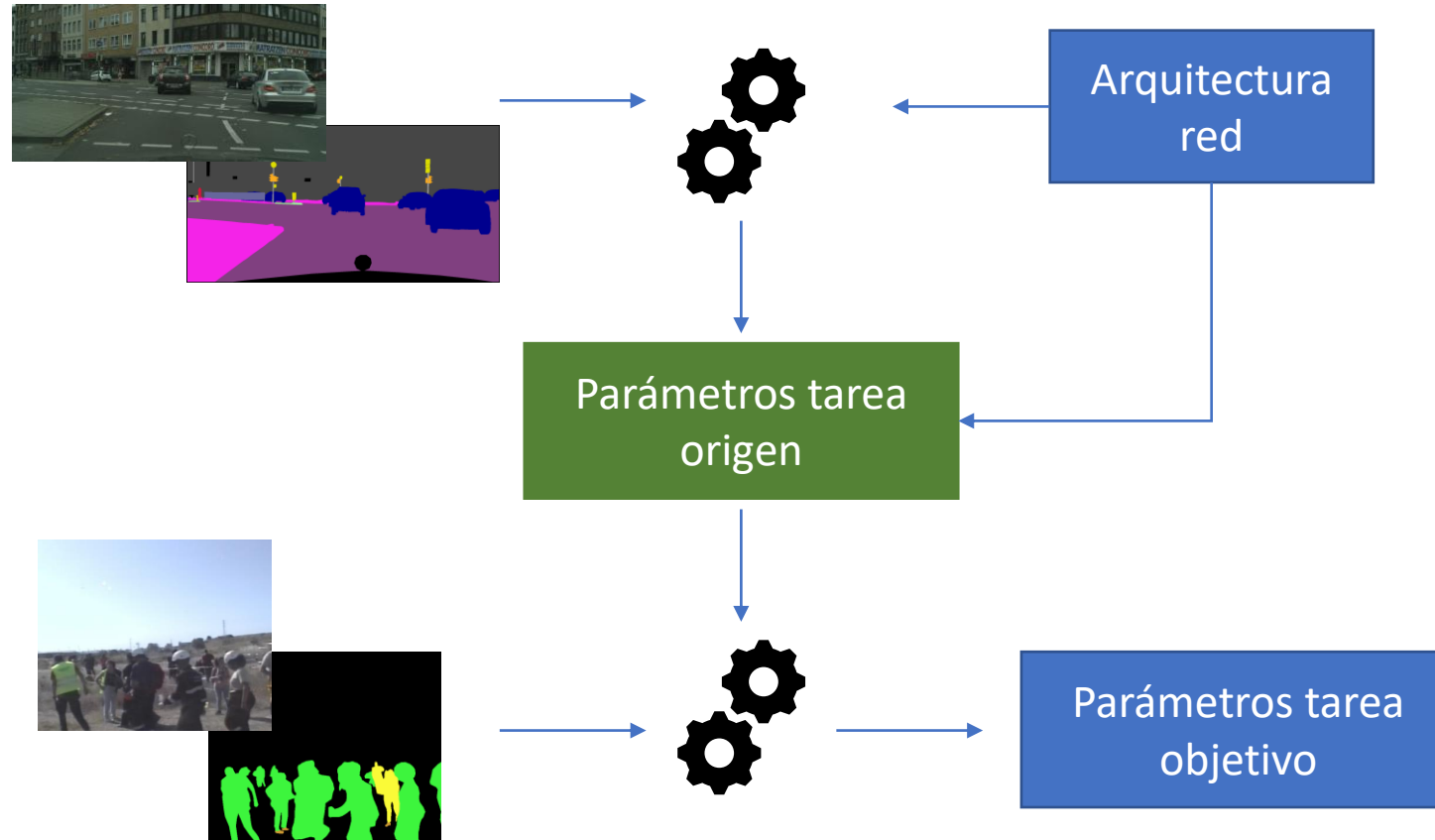
# Segmentación de imágenes con CNN



# Proceso de entrenamiento



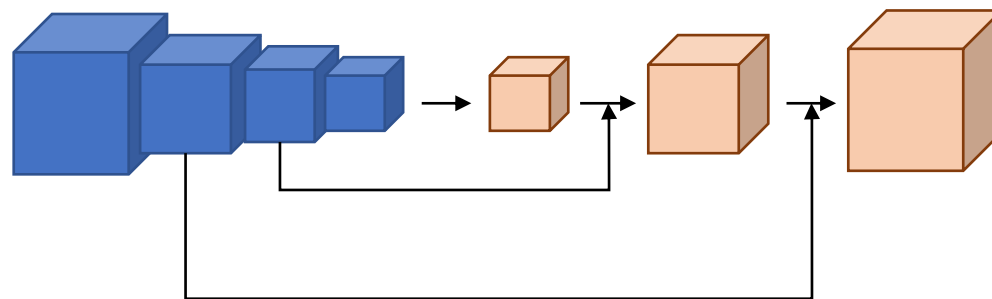
# Transferencia de conocimiento



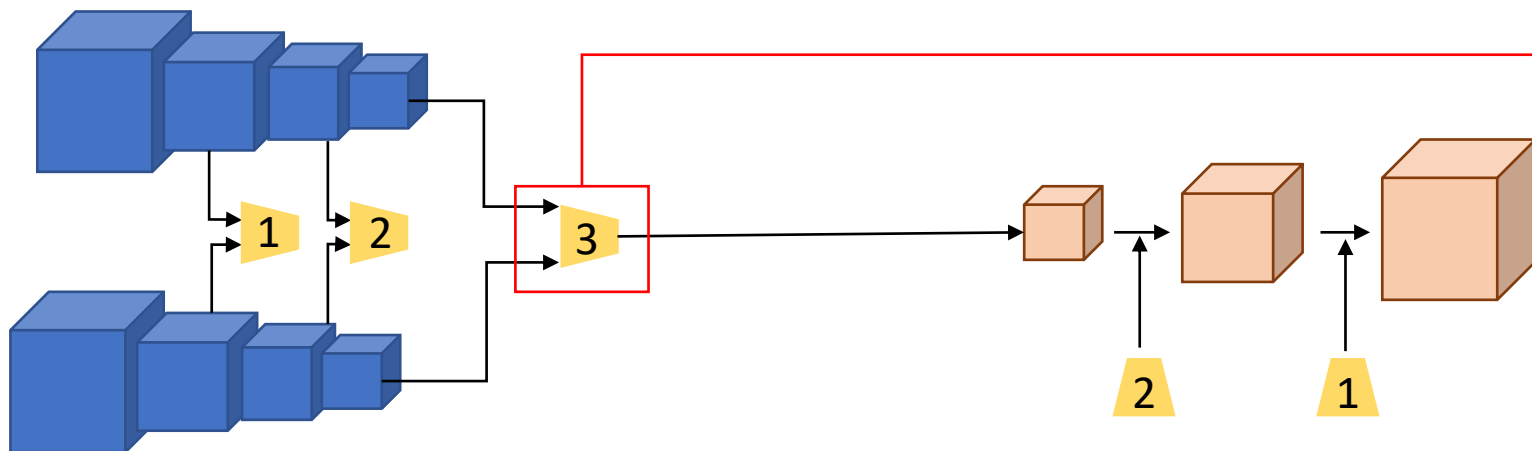
## 2. AdapNet++

Encoder AdapNet++

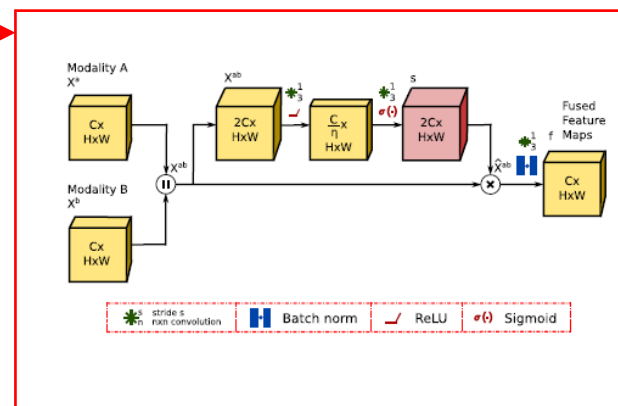
Decoder AdapNet++



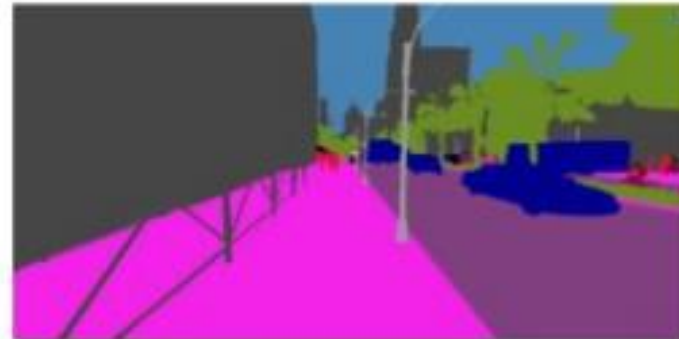
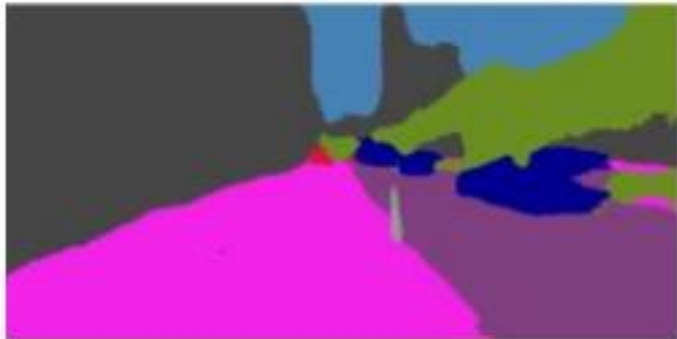
Encoder AdapNet++



Encoder AdapNet++

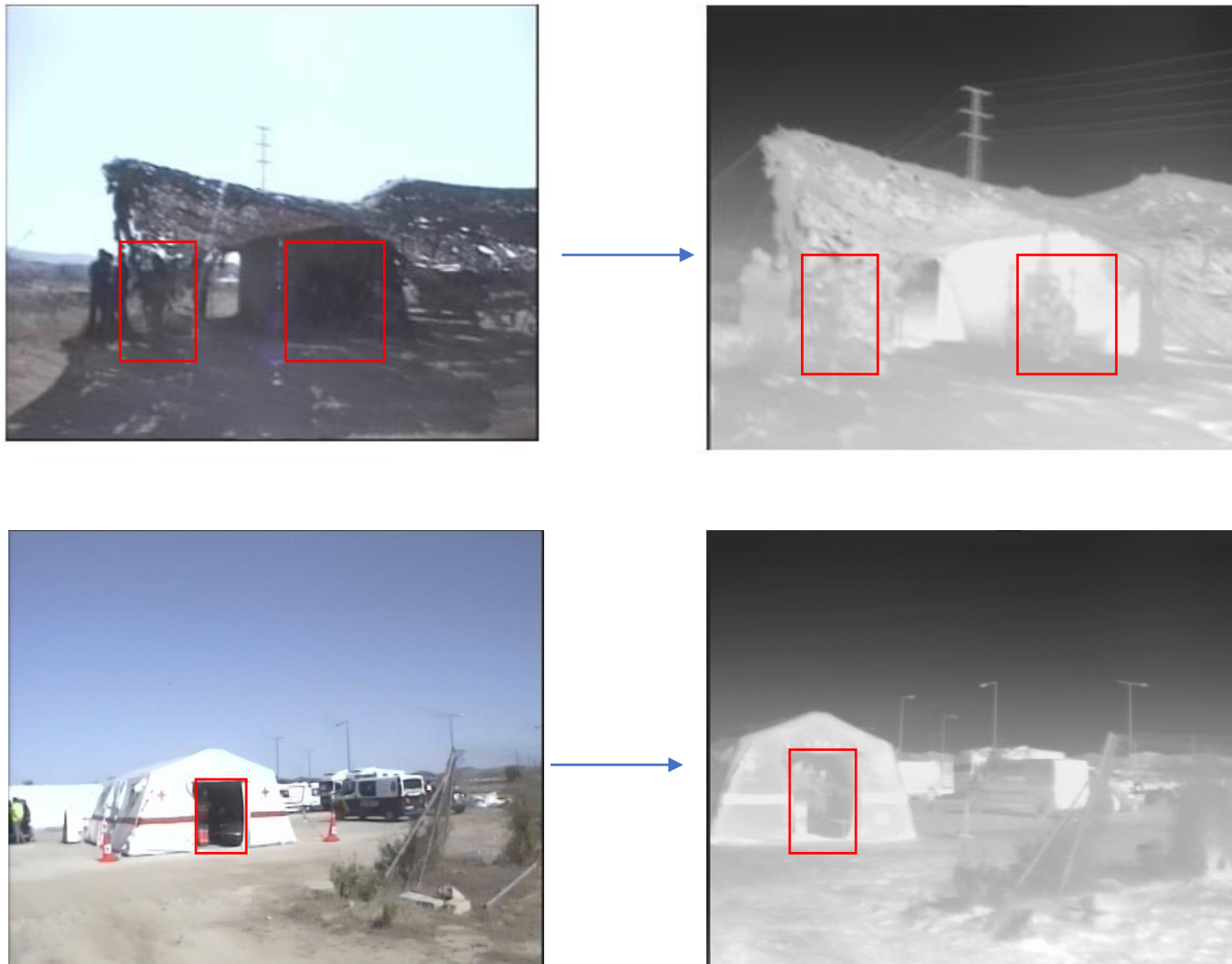


## Ventajas de la segmentación de imágenes de varias modalidades

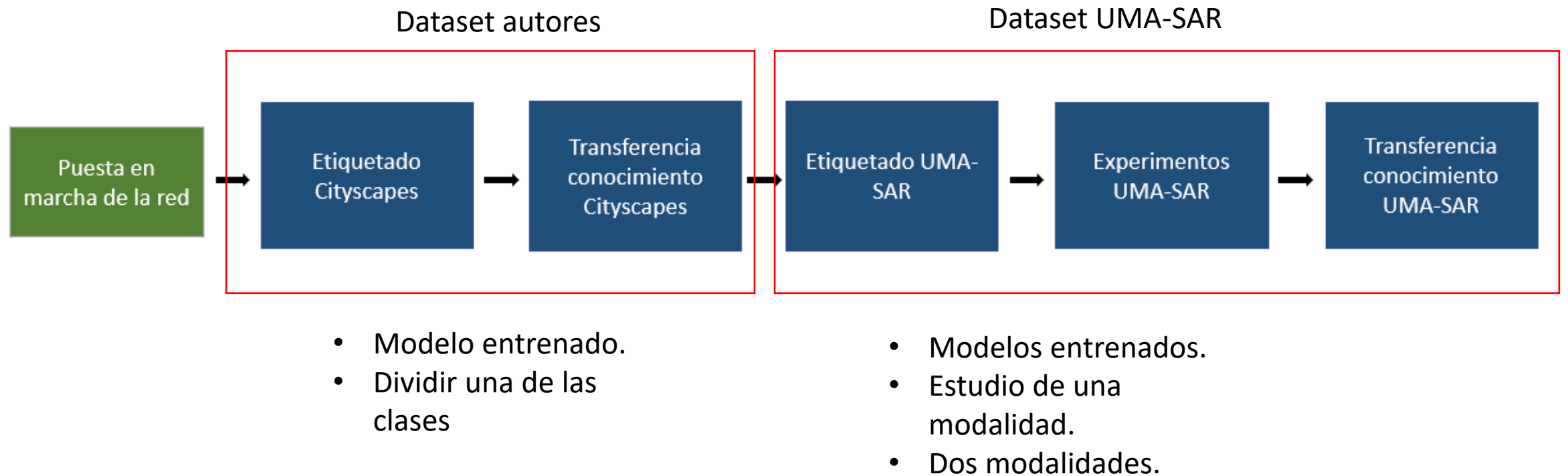




# Ventajas de la segmentación de imágenes de varias modalidades

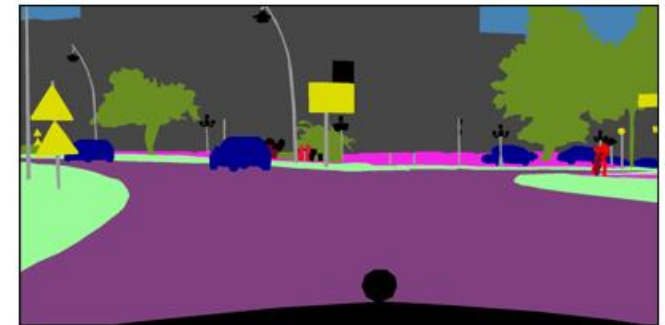


### 3. Plan de trabajo



## 4. Transferencia de conocimiento partiendo del modelo de Cityscapes

- Entornos urbanos.
- 50 imágenes etiquetadas.
- Herramientas para preparación de imágenes.



## Etiquetado de imágenes



### Tarea origen

Carretera

Acera

Vegetación

Peatón

Señal de tráfico

Edificio

Valla

Vehículo

Moto/Bicicleta

Cielo

Coche

Camión/Autobús

# Preprocesamiento de las imágenes de Cityscapes

## 1. Mapa de color



# Preprocesamiento de las imágenes de Cityscapes

## 2. Aumento de datos

Imagen original



Escalado



Simetría



Crop

## 5. Resultados obtenidos sobre el modelo de Cityscapes

$$\mathcal{L}(\theta) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m y_{ij} \log(p_{ij})$$

$$IoU = \frac{TP}{(TP + FP + FN)}$$

### Entrenamiento de las dos últimas capas del decoder

- Pérdida: 0,07
- MIoU: 69
- Tiempo: 1.5 horas

### Fine-tuning

- Pérdida: 0,12
- MIoU: 71
- Tiempo: 0.5 horas

### Entrenamiento del decoder completo

- Pérdida: 0,12
- MIoU: 66
- Tiempo: 1.7 horas

## 6. Transferencia de conocimiento utilizando el dataset UMA-SAR.

- Manipulación de las imágenes RGB.



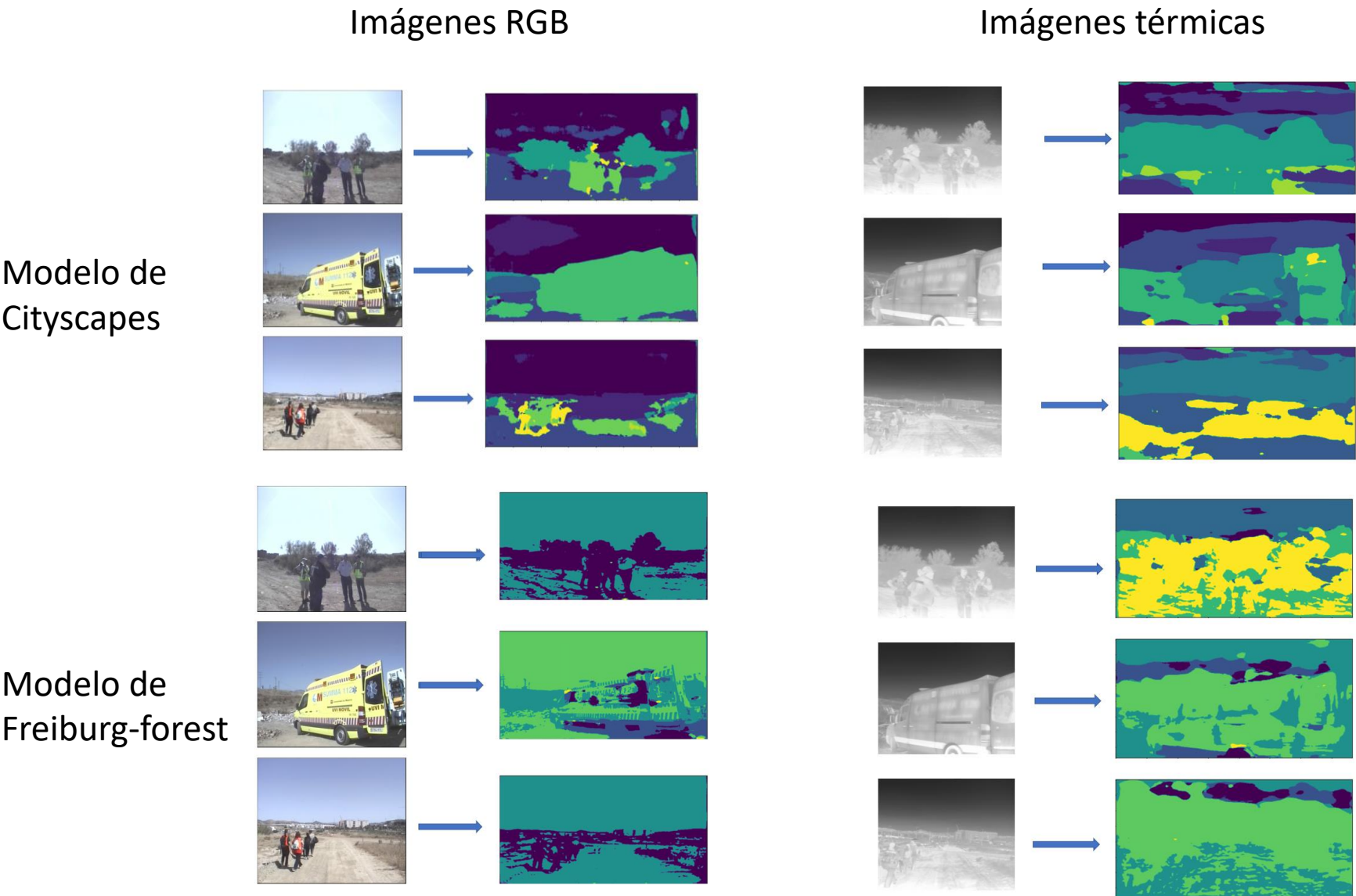


# Etiquetado de imágenes utilizando CVAT(Computer Vision Annotation Tool)

- Software libre.
- 60 imágenes.
- Tableta digitalizadora



# Análisis de modelos utilizando el dataset UMA-SAR



## 7. Resultados obtenidos con el dataset UMA-SAR

Transferencia de conocimiento partiendo de:

- Cityscapes.
- Modelo generalista.

Análisis de resultados:

- No se tiene en cuenta el fondo al calcular la función de pérdida.
- Se tiene en cuenta el fondo al calcular la función de pérdida.

## Resultados globales con una modalidad

Modelo	Pérdida	MIoU
Cityscapes rgb	0,0002	2,14 %
AdapNet general	0,06	1,95 %

RGB sin tener en cuenta el fondo

Modelo	Pérdida	MIoU
Cityscapes rgb	0,26	39 %
AdapNet general	0,51	41 %

RGB teniendo en cuenta el fondo

Modelo	Pérdida	MIoU
Cityscapes rgb	0,0002	1,9 %
Cityscapes mapas de profundidad	0,0034	2,45 %
AdapNet general	0,0001	1,86 %

Térmicas sin tener en cuenta el fondo

Modelo	Pérdida	MIoU
Cityscapes rgb	0,4	33 %
Cityscapes mapas de profundidad	0,43	15 %
AdapNet general	0,64	22 %

Térmicas teniendo en cuenta el fondo

## Tabla RGB-Cityscapes

Clase	Postivos reales	Falsos positivos	Falsos negativos	IoU
Escombros	10582,52	629,57	1936,28	80 %
Civil	2,71	0,36	2080,18	0 %
Personal de rescate	2625,34	1465,51	3939,58	33 %
Vehículo civil	3644,06	3,59	4108,61	47 %
Vehículo de rescate	6291,46	1706,57	3010,44	57 %
Víctima	0,00	0,00	342,00	0 %
Zapatos	0,00	0,00	351,12	0 %
Fondo	255798,22	12162,10	199,48	95 %

## Tabla RGB-AdapNet++

Clase	Positivos reales	Falsos positivos	Falsos negativos	IoU
Escombros	9488,40	1537,09	3030,40	68%
Civil	0,00	0,00	2082,90	0 %
Personal de rescate	2509,78	1729,55	4055,14	30 %
Vehículo civil	5712,36	116,84	2040,30	73 %
Vehículo de rescate	6556,46	519,71	2745,43	67 %
Víctima	0,00	0,00	342,00	0 %
Zapatos	0,00	0,00	351,12	0 %
Fondo	254783,22	11958,58	1214,48	95 %

## Tabla Térmicas-Cityscapes

Clase	Positivos reales	Falsos positivos	Falsos negativos	IoU
Escombros	9583,48	3736,27	498,49	69 %
Civil	112,07	7,32	1632,41	6 %
Personal de rescate	4801,17	3794,57	1207,39	49 %
Vehículo civil	0,00	0,00	7172,42	0 %
Vehículo de rescate	7965,68	7879,16	789,28	48 %
Víctima	0,00	0,00	361,00	0 %
Zapatos	0,00	0,00	343,48	0 %
Fondo	254575,80	2456,49	5869,35	97 %

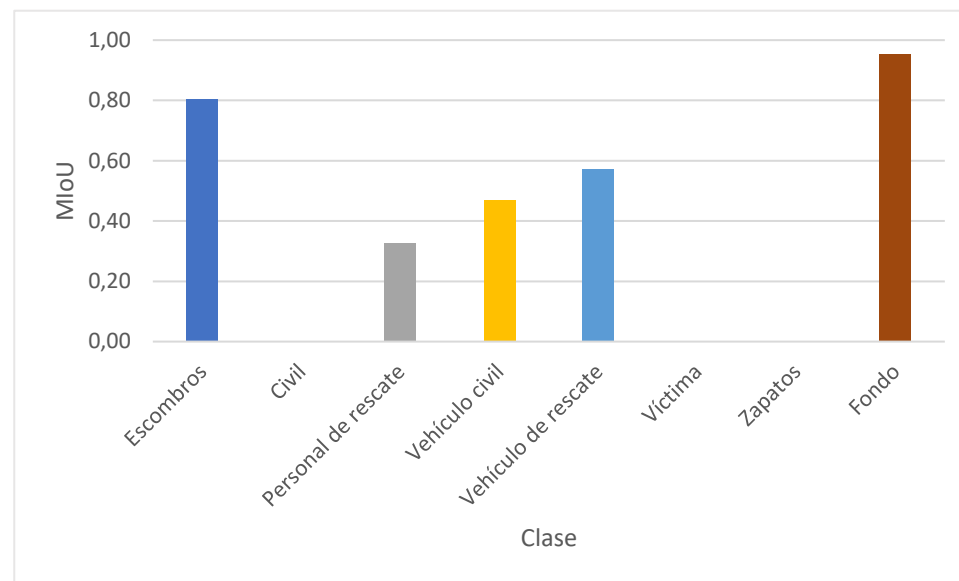
## Tabla Térmicas-AdapNet++

Clase	Positivos reales	Falsos positivos	Falsos negativos	IoU
Escombros	4169,15	1890,33	5912,82	35 %
Civil	0,00	0,00	1744,48	0 %
Personal de rescate	1143,75	1104,64	4864,81	16 %
Vehículo civil	0,00	0,00	7172,42	0 %
Vehículo de rescate	3697,33	2018,25	5057,63	34 %
Víctima	0,00	0,00	361,00	0 %
Zapatos	0,00	0,00	343,48	0 %
Fondo	259164,73	21723,81	1280,42	92 %

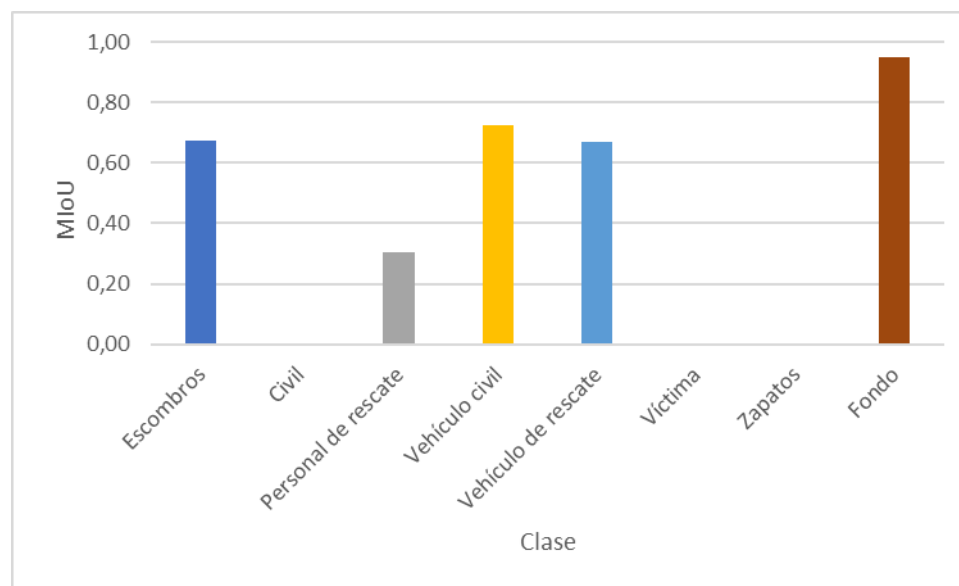


## Resultados RGB

- Cityscapes

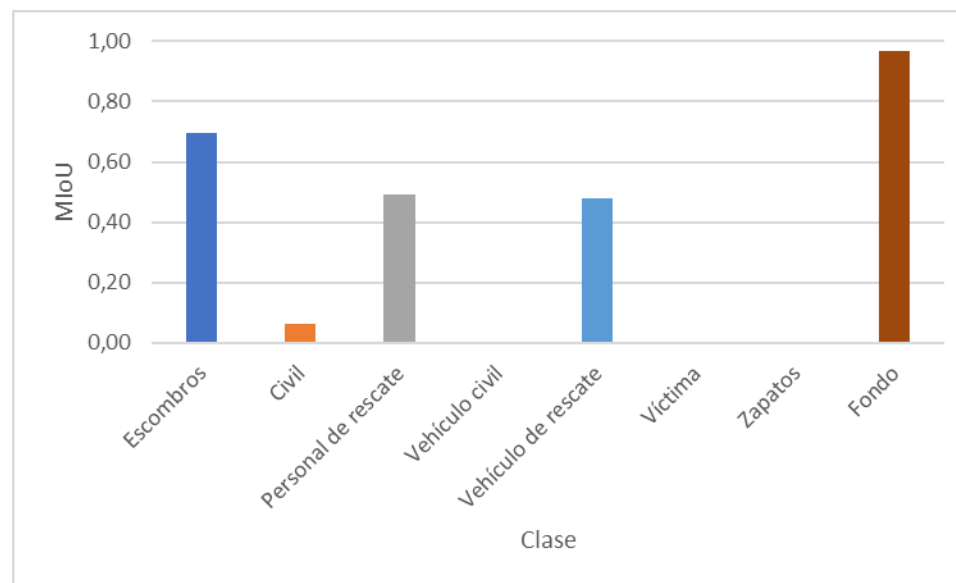


- AdapNet++

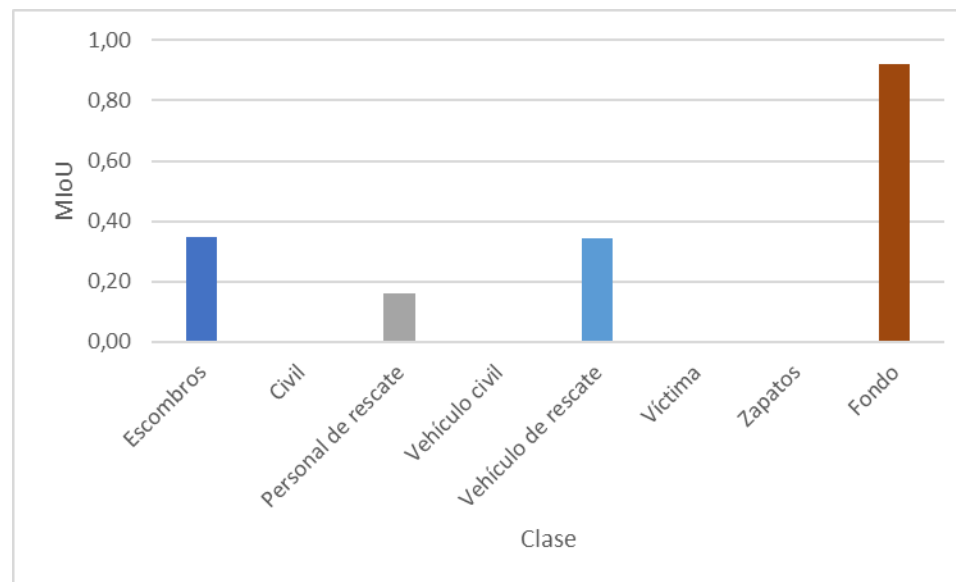


## Resultados Térmicas

- Cityscapes

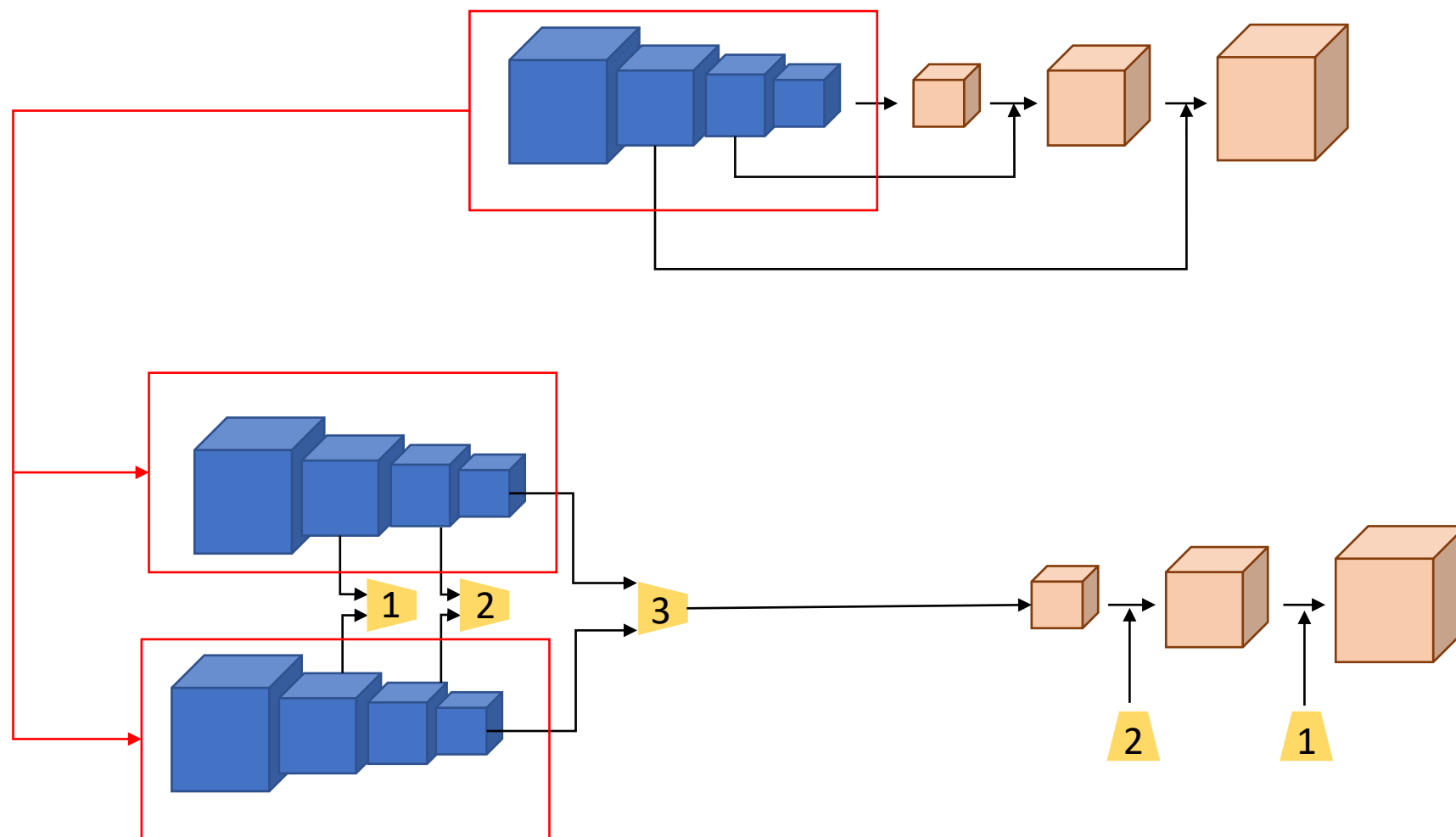


- AdapNet++



## AdapNet++ con 2 modalidades

AdapNet++ entrenada con una sola modalidad del dataset UMA-SAR

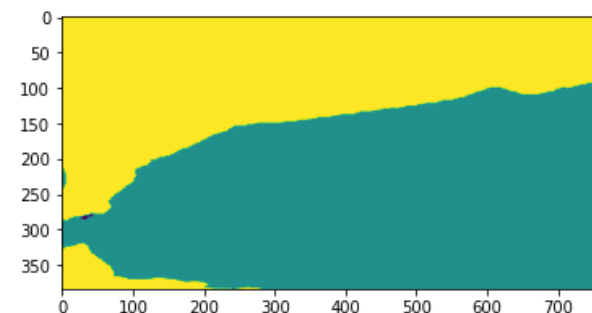
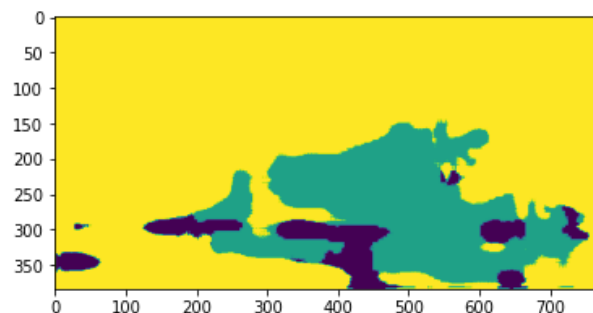
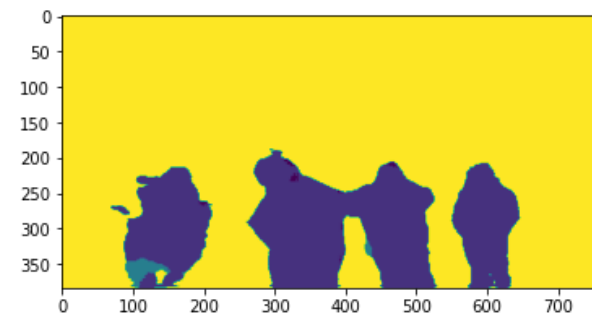
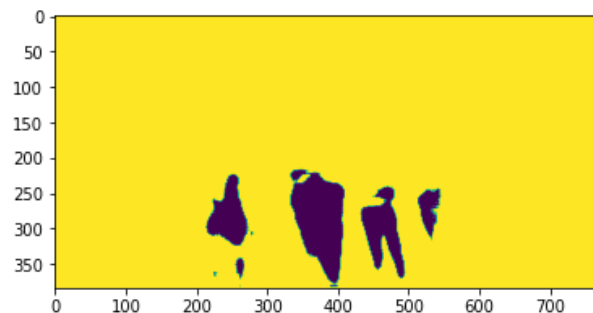
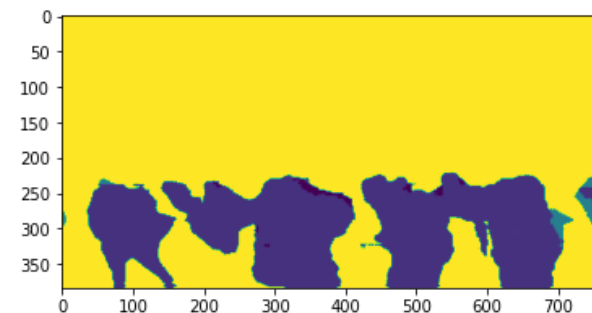
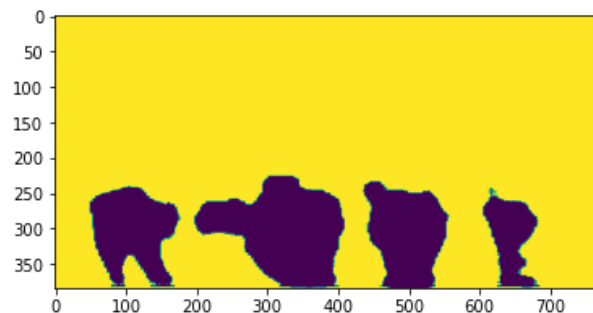


## Resultados transferencia de conocimiento de AdapNet++ con dos modalidades

- Pérdidas: 0,09.
- MIoU:52.

Clase	Positivos reales	Falsos positivos	Falsos negativos	IoU
Escombros	8299,225	939,36	1982,87	74 %
Civil	1218,45	1246,69	610,58	40 %
Personal de rescate	3391,68	910,17	2211,81	52 %
Vehículo civil	7009,61	941,06	695,04	81 %
Vehículo de rescate	7754,52	1931,37	972,29	73 %
Víctima	0	0	361,00	0 %
Zapatos	0	0	343,48	0 %
Fondo	256785,43	4484,43	3275,98	97 %

## Comparación entre 1 y 2 modalidades



## 8. Conclusiones y trabajos futuros

- Importancia del ajuste fino.
- Fondo de la imagen en el calculo del error.
- El resultado mejora al utilizar dos modalidades.

## 9. Trabajos futuros

- Aumento de la cantidad de imágenes con víctimas.
- Estudio del uso del bloque SSMA con Mask-RCNN.
- Análisis de funciones de pérdida diferentes.