

Visión por computadora II

TP4

Inicialmente, como le comenté a Juan Ignacio Cornet, seleccioné un dataset de vehículos en la vía pública. La propuesta es hacer instance segmentation. Luego descubrí que la segmentación del training set y del validation set la tenía que hacer a mano yo mismo, ya que el wizard de Roboflow, no cuenta con esta opción. Por este motivo, decidí cambiar de [dataset](#) y temática, pero manteniendo instance segmentation y a su vez incorporando bounding box. Luego de buscar un poco, encontré en Roboflow, un dataset de las letras del alfabeto para personas sordo mudas.

Considero importante recalcar que mi trabajo integrador de la especialización, no contempla el procesamiento de imágenes. Es por esto, que mi motivación a realizar el presente TP, radica únicamente en profundizar lo visto en clase, familiarizarme con los modelos y análisis de imágenes.

Con el dataset, realicé diferentes versiones, con y sin data augmentation y con y sin preprocessing.

A continuación, detallo los modelos y diferentes versiones de dataset que fui utilizando para comparar los ensayos:

Modelo R-CNN R-50 – FPN

Observaciones:

Dataset	Bbox/Segm	Precision media	Precision min	Precision max
versión 1 original	Bbox	63,92	24,119	85,05
versión 1 original	Segmentation	75,271	38,416	100
Versión 2 Augmen	Bbox	59,792	15	85,050
Versión 2 Augmen	Segmentation	68,753	15	100
Versión 3 gray	Bbox	72,310	23,564	100
Versión 3 gray	Segmentation	81,808	30,297	100
Versión 4 gray Aug	Bbox	57,364	18,251	95,050
Versión 4 gray Aug	Segmentation	65,302	26,658	90,693

Modelo R-CNN R-101 – C4

Observaciones:

Dataset	Bbox/Segm	Precision media	Precision min	Precision max
versión 1 original	Bbox	64,882	26,667	94,587
versión 1 original	Segmentation	74,898	30	90
Versión 2 Augmen	Bbox	58,365	15	90
Versión 2 Augmen	Segmentation	67,14	15	92,525

Modelo R-CNN R-101 – FPN

Observaciones:

Dataset	Bbox/Segm	Precision media	Precision min	Precision max
---------	-----------	-----------------	---------------	---------------

versión 1 original	Bbox	64,178	20,198	86,416
versión 1 original	Segmentation	70,088	22,723	100
Versión 2 Augmen	Bbox	50,942	14	91,122
Versión 2 Augmen	Segmentation	60,839	18	95,050

Viendo que el mejor resultado se obtuvo con el pre processing, reentreno el modelo R-CNN R-101 – C4 con ese dataset obteniendo los siguientes resultados

Dataset	Bbox/Segm	Precision media	Precision min	Precision max
versión 3 gray	Bbox	69,008	26,614	95,050
versión 3 gray	Segmentation	77,085	34,218	90

Conclusiones:

Considero que el modelo con el dataset con preprocessing tiene mejores resultados que los modelos color.

Se debe corregir el validation set para incluir las letras P e Y.

Los resultados de segmentación son mejores en todos los casos a los obtenidos con Bounding box.

Las letras K y N son el mayor desafío. La N es el puño cerrado con el pulgar asomando entre los dedos medio y anular. La letra K es muy parecida a la V, sólo cambia la posición del pulgar.

El modelo R-CNN R-50 – FPN posee los mejores resultados con este dataset, seguido del R-CNN R-101 – C4.