Learning object detectors from image captions

Proposer(s) / Proposatzailea(k): Gorka Azkune, Oier López de Lacalle

Contact / Kontaktua: gorka.azcune@ehu.eus, oier.lopezdelacalle@ehu.eus

Description / Deskribapena

Object detectors are usually trained on annotated datasets. Annotations for object detection are not simple, since annotators are required to identify objects by their class and to draw the bounding boxes of those objects. In consequence, high-quality big annotated datasets are very costly. In this project, we propose to leverage image captioning datasets, where every image is accompanied by a descriptive textual caption. Our hypothesis is that image captioning systems can learn object features, using an open vocabulary, thus producing scalable detectors. We propose to research on spatial attention systems to learn those detectors.

Goals / Helburuak

To design and implement an object detector without using object bounding box annotations.

Requirements / Betebeharrak

None.

Framework / Esparrua

Python, Tensorflow/Pytorch

Tasks and plan / Atazak eta plana

Analyze image captioning datasets.

Analyze the state of the art in unsupervised object detection and related topics.

Implement an attention-based image captioning system.

Analyze the viability of using attention maps for bounding box prediction.

Design and implement an algorithm to extract bounding boxes from attention maps.

Test and evaluate the implemented algorithm.

Analyze the usage of the algorithm as an automatic object annotator tool.