Object Detection in Equirectangular Images

Francisco Henriques [1]; Joana Costa [1,2]; Catarina Silva [2]; Pedro Assunção [1,3]

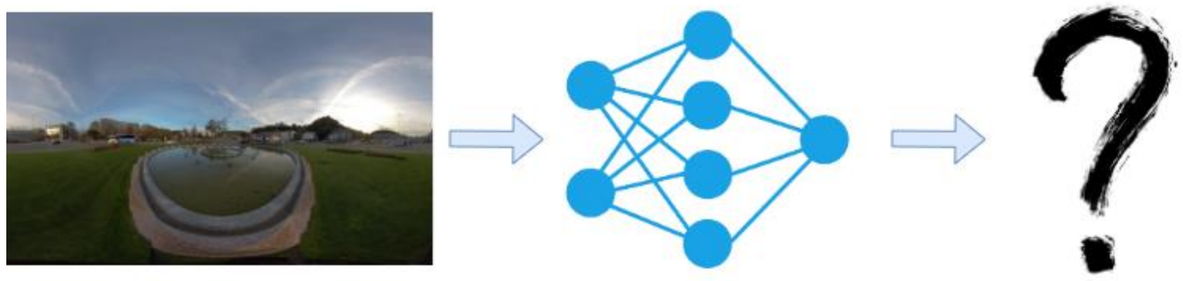
[1] ESTG, Polytechnic of Leiria, Leiria, Portugal;
[2] Universidade de Coimbra, CISUC - Centro de Informática e Sistemas, FCTUC-DEI - Departamento de Engenharia Informática, Portugal
[3] Instituto de Telecomunicações, Leiria, Portugal

Abstract

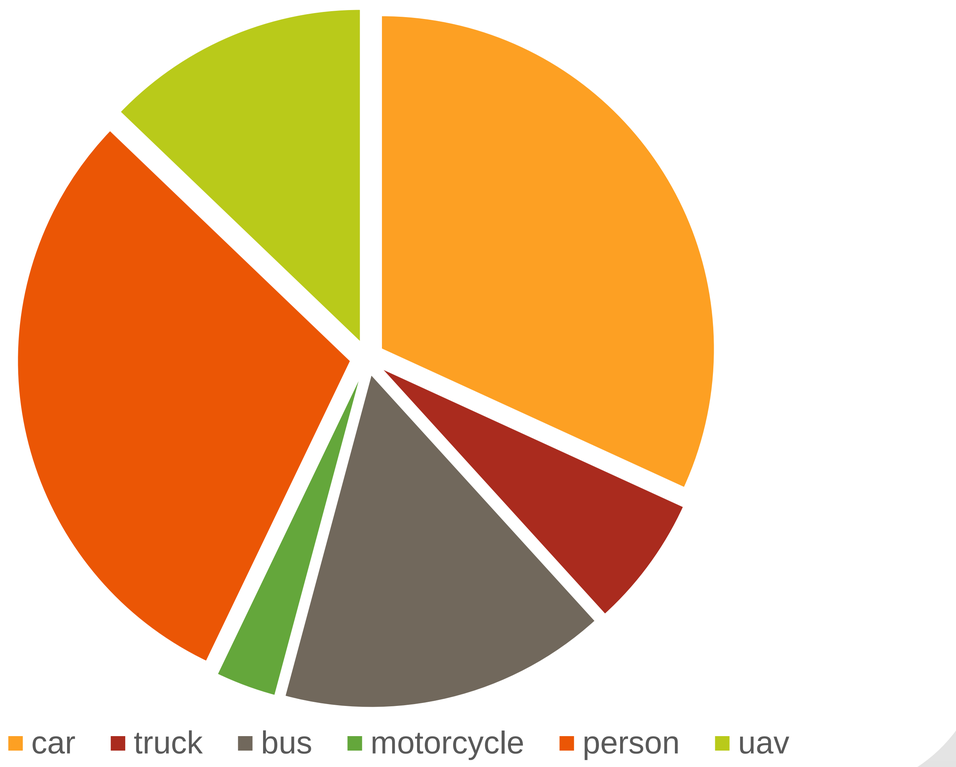
Nowadays, computer vision (CV) is widely used to solve real-world problems, which pose increasingly higher challenges. In this context, the use of omnidirectional video in a growing number of applications, along with fast development of Deep Learning (DL) algorithms for object detection, drives the need for further research to improve existing methods specifically developed for conventional 2D planar video. This work explores DL methods to detect visual objects in omnidirectional images represented onto plane through Equirectangular Projection (ERP). It is shown that the error rate of object detection using existing DL models with ERP images depends on the object location in the image. Then, a new object detection framework is proposed to obtain uniform error rate across the whole spherical image regions.

Research Problem

How do ERP images affect conventional 2D image-based Deep Learning algorithms' performance?



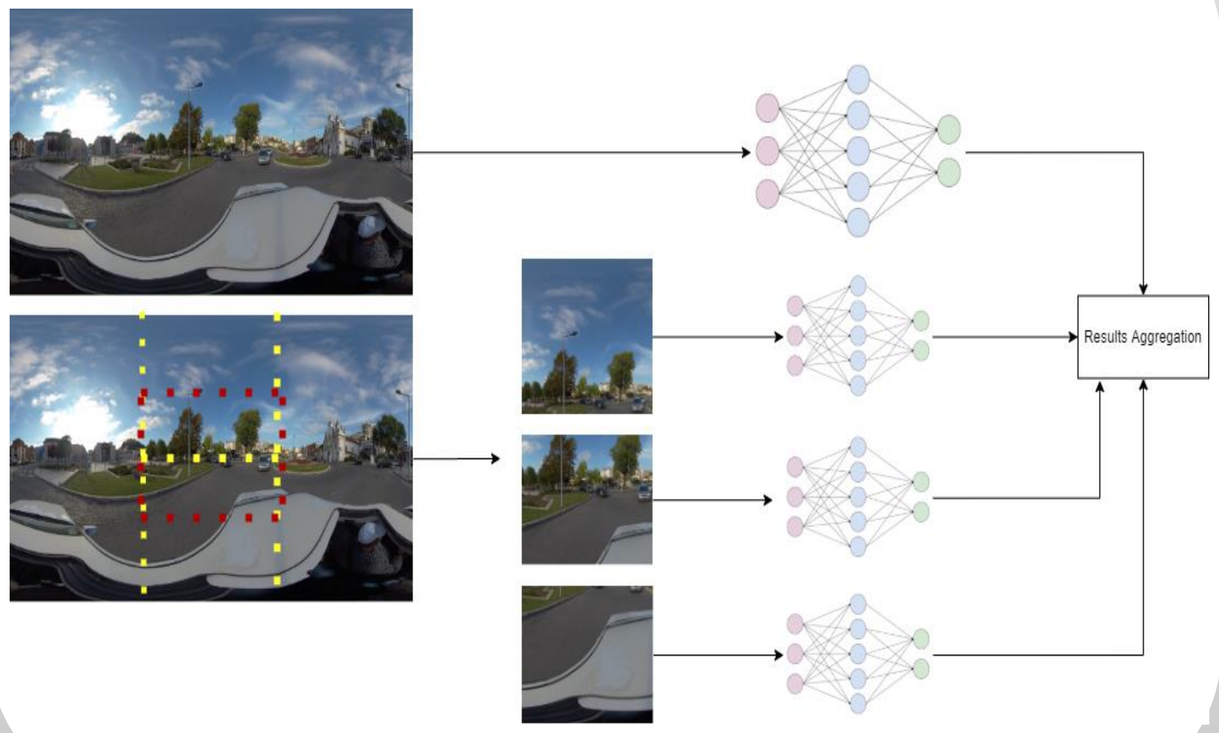
Dataset



Results

Non-detected objects by image region do not follow a uniform pattern: Mid-region objects tend to be harder to detected.

Proposed Approach



Conclusion

Automatic object detection in ERP images with high-level accuracy created new problems that did not occur before in conventional images. Object distortion and unusual view pose as well as very-high image resolution tend to give rise to an extremely wide range of objects dimensions and aspect ratios across an image. Our initial experiments demonstrated that a conventional framework does not provide uniform accuracy results across the whole image. The framework proposed in this paper allows to make non-detected objects by image region more uniform through two parallel pipelines: one for the whole image and the other focusing on the most problematic region, the center.

This work was partially supported by project ARoundVision CENTRO-01-0145-FEDER-030652.