

## PhD position at UPF funded by La Caixa

### Audiovisual scene understanding

The main focus of the research project is the automatic audio-visual scene understanding from videos. In a real scene, the visual and acoustic signals often provide complementary information of the different elements present in the scene. We are interested in understanding the geometric, semantic and acoustic configuration of a multi-object and multi-person dynamic scene. That is, to identify and localise in the scene the different sounding objects together with their individual sounds. The applications are numerous: scene monitoring, video content enhancement, autonomous driving, robot navigation, augmented reality, etc. For that, we will address different relevant tasks such as object detection, sound localisation, motion estimation and sound source separation. The proposed models will use deep learning techniques. We are particularly interested in designing self-supervised strategies to train such models, thus avoiding manual annotation, which is particularly tedious and expensive in the video case.

Some related works and demos by the research group:

<https://ipcv.github.io/VoViT/>

<https://ipcv.github.io/VocaLiST/>

More details: <https://finder-fellowships.lacaixafoundation.org/finder?position=7370>

How to apply: <https://fundacionlacaixa.org/en/inphinit-doctoral-fellowships-incoming>

Contact: gloria.haro@upf.edu