Object detection has been one of the most active research topics in computer vision for the past years. However, the use of temporal information in videos to boost the detection precision is still an open problem. Although a traditional object detection framework can be executed at frame level, it does not exploit temporal information available on videos that can be crucial to address challenges such as motion blur, occlusions or changes in objects appearance at certain frames.

We propose a two stage object detector called FANet (Feature Aggregation Network) based on short-term spatio-temporal feature aggregation to give a first detection set, and long-term object linking to refine these detections. The overall framework achieves a competitive 80.9% mAP in the widely used ImageNet VID dataset. Also, it significantly outperforms the single frame baseline in the challenging USC-GRAD-STDdb dataset with an improvement of 5.4% mAP.

**Moderator:** Xosé Manuel Pardo

**Supervisors:** Manuel Mucientes and Víctor Brea