Action Genome : Actions as Composition of Spatio-temporal scene graphs.

Summary

1. Propose to learn more structured representation of action in videos as graph

2. The nodes in the graph are either a person or an object or a person-object relationship

3. The relationship can be : attention (e.g. looking at), spatial (e.g. in front of), contact (e.g. carrying)

4. Propose a new dataset that annotates static scene with this type of graph.

5. Show that the graphical representation can help with [action recognition, few-shot AR.