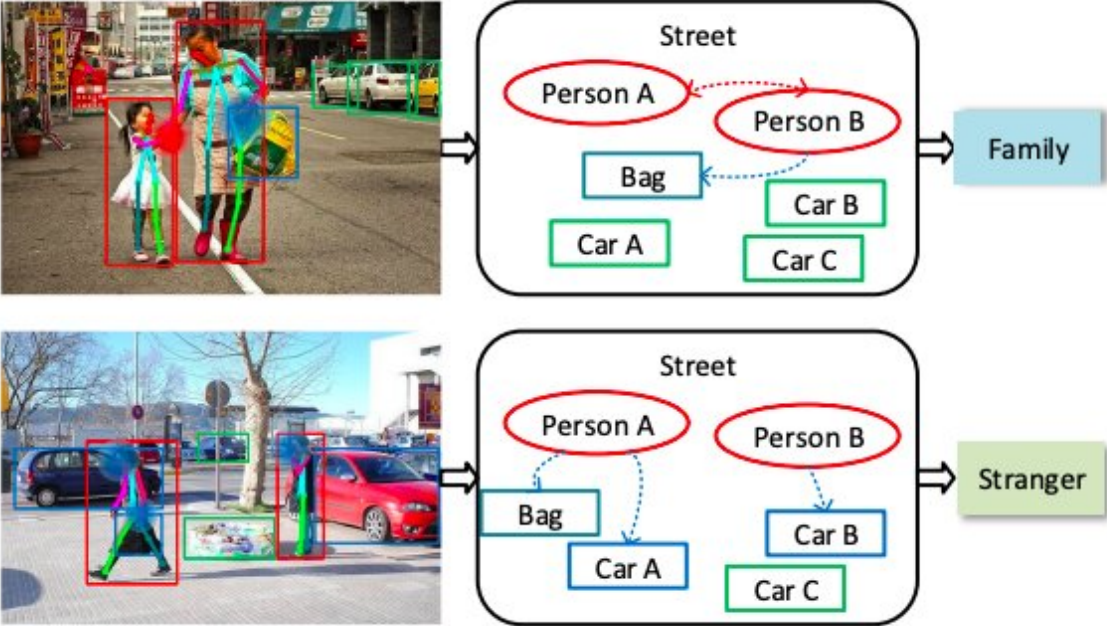# A multi-granularity reasoning framework for social relation recognition

January 26 2019, by Ingrid Fadelli

How do we recognize two persons are family or strangers from an image? The scenes, appearance of persons, and interactions among persons and contextual objects are significant cues for recognition. Credit: Zhang et al.

A team of researchers at Beijing University and JD AI Research have recently developed a multi-granularity reasoning framework for social relation recognition. Their framework, described in a paper pre-published on arXiv, was trained to analyze images of people in different

scenes and predict the social relation between them.

Effectively inferring the social relations between people could aid intelligent agents to reach a better understanding of human behaviors and emotions. Image-based social relation recognition entails the ability to classify the relationship between pairs of people in an image into pre-defined relation types, such as friends, family, acquaintances, strangers, etc.

Image-based social relation recognition tools could have a variety of useful applications, for instance, in personal image collection mining and social event understanding. Recent advances in deep learning have opened new possibilities for social relation recognition, leading to significant improvements in performance.

Nonetheless, automatically recognizing social relations in images has so far proved challenging, particularly due to the substantial gap between the domains of visual content and social relations. Most existing approaches work by separately processing features such as facial expressions, body appearance and contextual clues.

"Existing methods for social relation recognition usually utilize low-level visual features such as the appearance of persons, face attributes and contextual objects," the researchers wrote in their paper. "Although some approaches explore the relations between persons and objects, they only consider the co-existence in an image. However, only depending on the single-granularity representation can hardly overcome the domain gap between visual features and social relations."
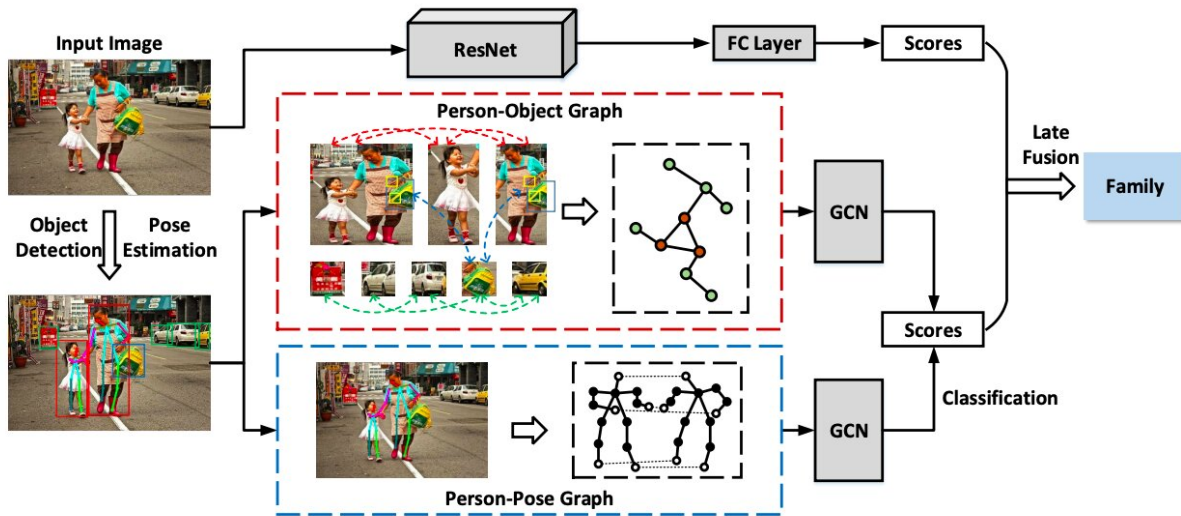
An overview of the multi-granularity reasoning framework. Credit: Zhang et al.

By analyzing features individually, existing social relation recognition methods typically fail to capture multi-granularity semantics, such as overall scenes or where people are located in an image, as well as interactions between people and objects. To address these limitations, the team of researchers at Beijing University and JD AI Research devised a multi-granularity reasoning framework for social relation recognition in images.

Their framework acquires global knowledge from the whole scene and mid-level details from the regions in which people and objects are located in an image. It also explores the fine-granularity pose key points of people to uncover interactions between people and objects.

"Specifically, the pose-guided Person-Object Graph and Person-Pose Graph are proposed to model the actions from persons to object and the interactions between paired persons, respectively," the researchers

explained in their paper. "Based on these graphs, social relation reasoning is performed by graph convolutional networks. Finally, the global features and reasoned knowledge are integrated as a comprehensive representation for social relation recognition."

The researchers evaluated their model on two large-scale social relation datasets, namely the People in Social Context (PISC) and People in Photo Album (PIPA) datasets. The PISC dataset contains images of common social relations in daily life, while the PIPA dataset contains images annotated based on the social domain theory, which divides social life into five domains and 16 different relations. In these tests, their model attained remarkable results, outperforming a variety of state-of-the-art methods.

Despite these encouraging results, developing tools to recognize social relations remains very challenging, particularly when these are intimate relations, such as those between friends, families or couples, which can be hard to discern for human viewers, too. In the future, the researchers plan to explore new ways to discover context cues in images and overcome the challenges associated with a lack of available data for some types of social relations.

  More information: Multi-granularity reasoning for social relation recognition from images. arXiv:1901.03067 [cs.CV].
arxiv.org/abs/1901.03067