

# Supplementary Materials of “A Domain Based Approach to Social Relation Recognition”

Qianru Sun    Bernt Schiele    Mario Fritz  
Max Planck Institute for Informatics, Saarland Informatics Campus  
{qsun, schiele, mfritz}@mpi-inf.mpg.de

## 1. Annotation tool

In Figure 1, we present the panel of the annotation tool. Annotators labeled the whole PIPA dataset using this tool, independently. Each time the tool shows a photo with a pair of head bounding boxes<sup>1</sup>. Please note, photos containing only one identity or more than 5 identities<sup>2</sup> are unused. Therefore, each photo contains 1 ~ 10 person pairs.

On this panel, we have 20 social relation candidates and use “others” for annotating some pairs possibly belonging to a domain but not belonging to any relation in this domain. When all annotations were finished, we analysed label statistics and agreements, then filtered out 4 relations (NO. 8, 14, 18, and 20 in this panel) due to insufficient occurrence or insufficient agreement, as explained in Section 4.1 of the main paper.

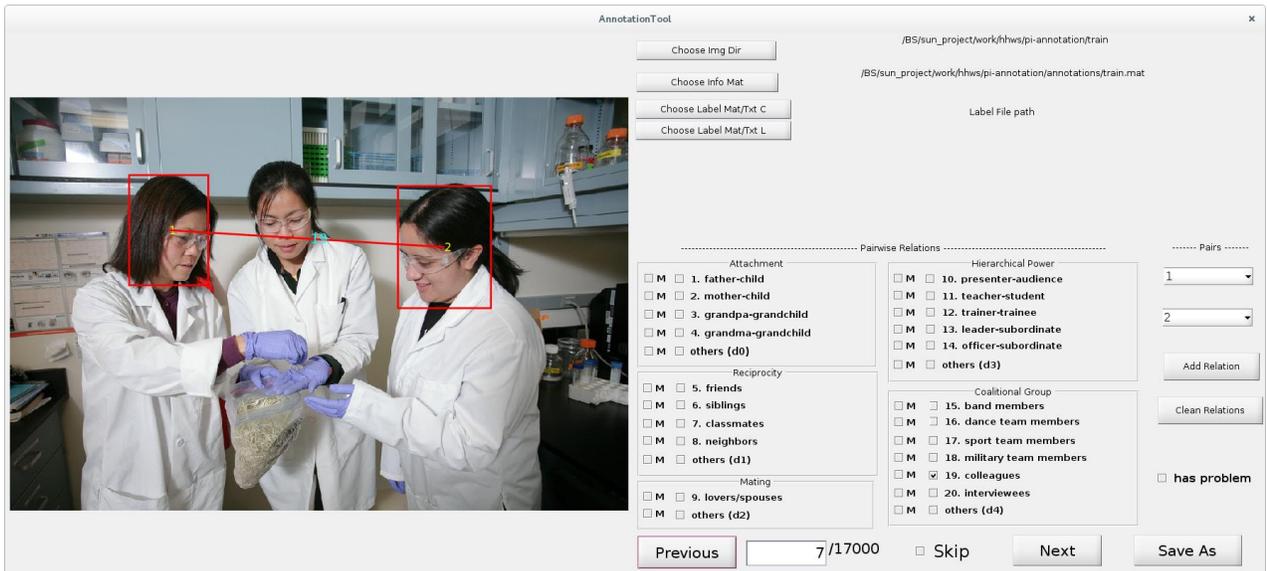


Figure 1: The panel of the annotation tool used for the annotation process. Each time the tool pops up a photo with a pair of head bounding boxes. An annotator recognizes the relation based on obvious visual cues, such as age, gender, clothing, activity and so on, then chooses at most 3 relation labels. “M” denotes “maybe”, i.e., not very confident.

<sup>1</sup>Head bounding boxes were previously annotated in PIPA for the task of person identification.

<sup>2</sup>Identity labels were previously annotated in PIPA for the task of person identification. There are about 3.7% images containing more than 5 identities, which are mostly same-relation person pairs in groups.

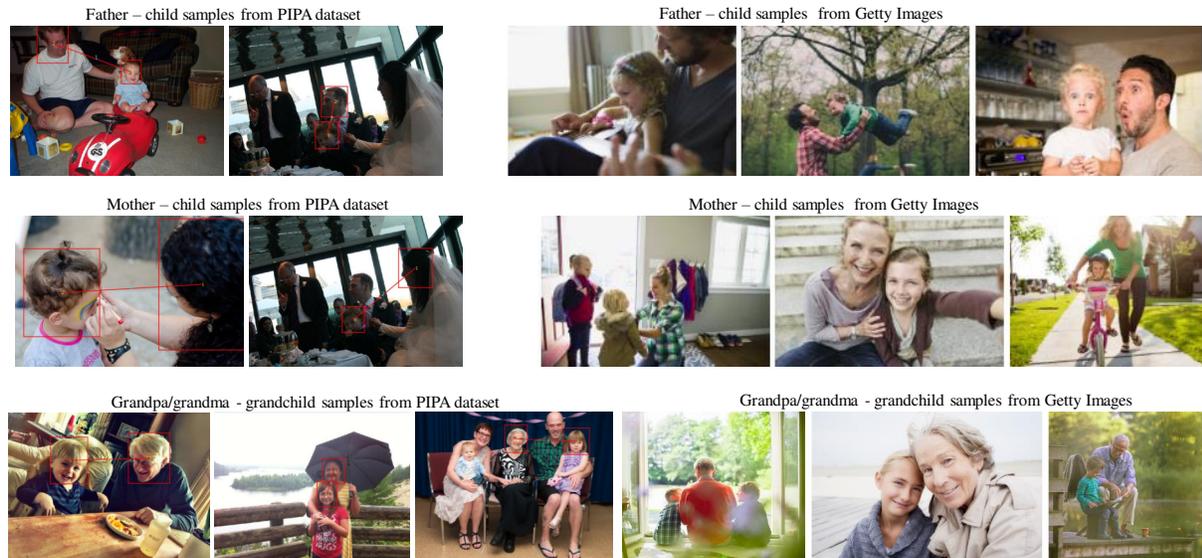


Figure 2: Attachment domain: annotation samples from PIPA dataset and searched samples from Getty Images.

## 2. Annotation samples

Based on the definitions in the social psychology article [1], we gave detailed explanations for 5 social domains. For example, in the Attachment domain, specific visual cues are the age difference between child and parents (or grandparents), body proximity, and children’s behavior of seeking protection from adults (see Section 3.1). We carefully explained these descriptions to 5 annotators, in conjunction with the reference samples shown in Figures 2-6.

In Figures 2-6, we show the reference samples of 20 social relations used for the annotation. PIPA samples are collected from the 10% PIPA data when choosing a relation label list (explained in Section 4.1). To help annotators to understand social relations, we also borrow some pictures searched in Getty Images, as shown beside PIPA samples. Photos searched on Getty Images are mostly fake and posed for photograph, aiming to reveal the inherent concepts in photos. Getty image samples searched by social relation entries turned out to be very helpful for annotators to better understand the visual appearances of social relations.

## 3. Supplementary results

In our main paper, Figure 7 showed the all-attribute result (our best) and top 4 single attribute results (contribution ranks were shown in the Figure 6 of the main paper). In this supplementary document, Figure 7 supplements the relation labels predicted by other 8 single-attribute models, i.e., from “proximity” to “body loc.& scale”, under images. Here, these images are in the same order with those in the Figure 7 of the main paper, and the image titles are the ground truth. Positive samples are Figure 7(a)-(h) which are correctly predicted by our all-attribute model while get noisy predictions by single-attribute models. For example in Figure 7(a), father-child relation was wrongly predicted by 10 single-attribute models except “activity” and “clothing”. This is consistent with the contribution ranks shown in the Figure 6 of the main paper that the attributes of “activity” and “clothing” dominate the recognition. Negative samples are shown in Figure 7(i)-(l).

To plot the Figure 6 in the main paper, we gave an example of computing X, Y coordinates of “body age” by defining  $X = acc(\text{bodyAge}, \text{domain})/acc(\text{all}, \text{domain})$  and  $Y = acc(\text{bodyAge}, \text{relation})/acc(\text{all}, \text{relation})$  in Section 6.3 (**Attribute categories**). In this document, we supplement all accuracies  $acc$  used for plotting, in Table 1. Taking the previous example “body age”, its coordinates were computed as  $X = 57.4\%/67.8\% = 0.847$ ,  $Y = 31.0\%/57.2\% = 0.542$ .

## References

- [1] Bugental, D.B.: Acquisition of the algorithms of social life: A domain-based approach. In: Psychological Bulletin, Vol. 126, No. 2, pp. 187-219. (2000) 1



Figure 3: Reciprocity domain: annotation samples from PIPA dataset and searched samples from Getty Images.



Figure 4: Mating domain: annotation samples from PIPA dataset and searched samples from Getty Images.

ATTRIBUTE	RELATION RECOGNITION	DOMAIN RECOGNITION
Head age	42.8%	56.8%
Head gender	38.0%	53.8%
Head loc.& scale	30.8%	45.0%
Head appearance	31.5%	48.4%
Head pose	34.7%	52.3%
Face emotion	37.7%	55.3%
Body age	31.0%	57.4%
Body gender	46.6%	58.0%
Body loc.& scale	27.7%	44.2%
Clothing	51.4%	60.3%
Proximity	39.6%	55.4%
Activity	52.4%	64.5%
All Attributes	57.2%	67.8%

Table 1: Accuracies of recognizing relations and domains using single-attribute models and all-attribute model. These numbers were used to compute the coordinates of attribute dots presented in the Figure 6 of the main paper.

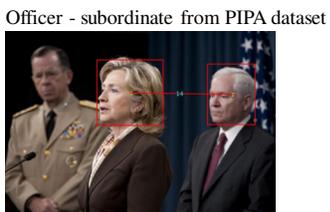
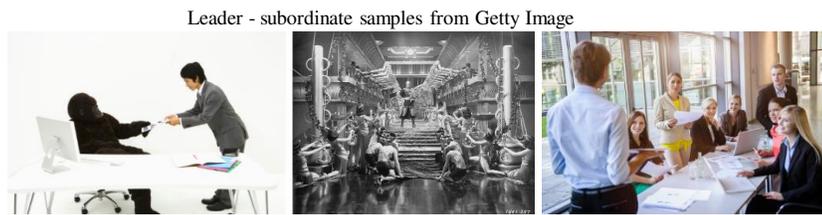
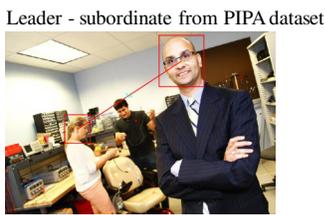


Figure 5: Hierarchical power domain: annotation samples from PIPA dataset and searched samples from Getty Images.

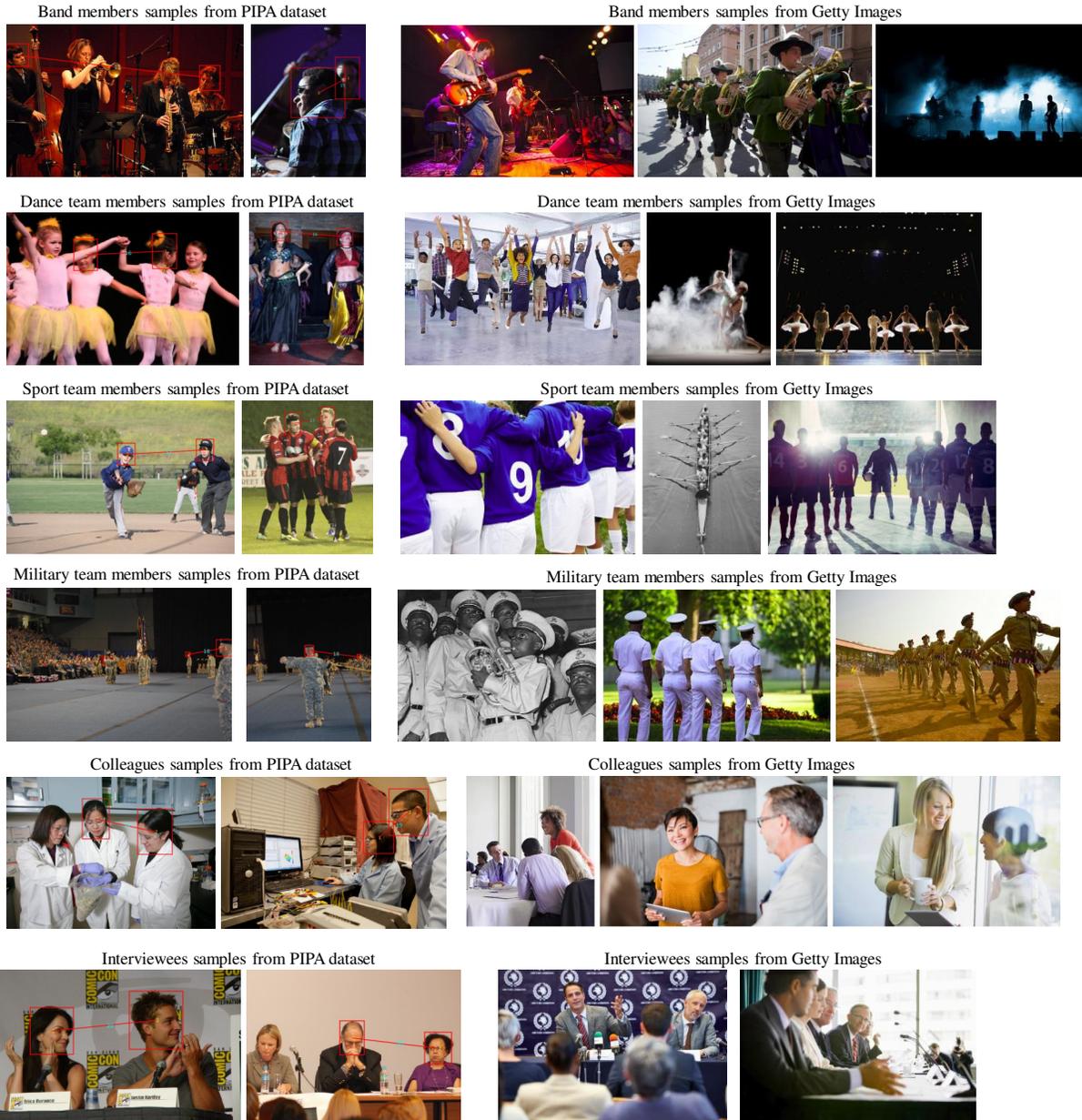


Figure 6: Coalitional groups domain: annotation samples from PIPA dataset and searched samples from Getty Images.



Figure 7: Relation labels predicted by the all-attribute model and 12 single-attribute models. Image titles are the ground truth.