

# Deep Convolutional Network Cascade for Visual Relationship Detection

Fu Xiong<sup>1,3</sup>, Bochen Wang<sup>2,3</sup>, Liang Yu<sup>3</sup>, Peng Chen<sup>3</sup>, Li Ma<sup>3</sup>  
<sup>1</sup>Huazhong University of Science and Technology, <sup>2</sup>Jiangsu University, <sup>3</sup>Alibaba Group

## Challenge

Can a computer provide as precise an image description?

Identifying different objects (man and cup) is an important problem on its own, but identifying the relationship between them (holding) is critical for many real world use cases.

## Target :

Detect pairs of objects in particular relationships.

## Result :

Relation:

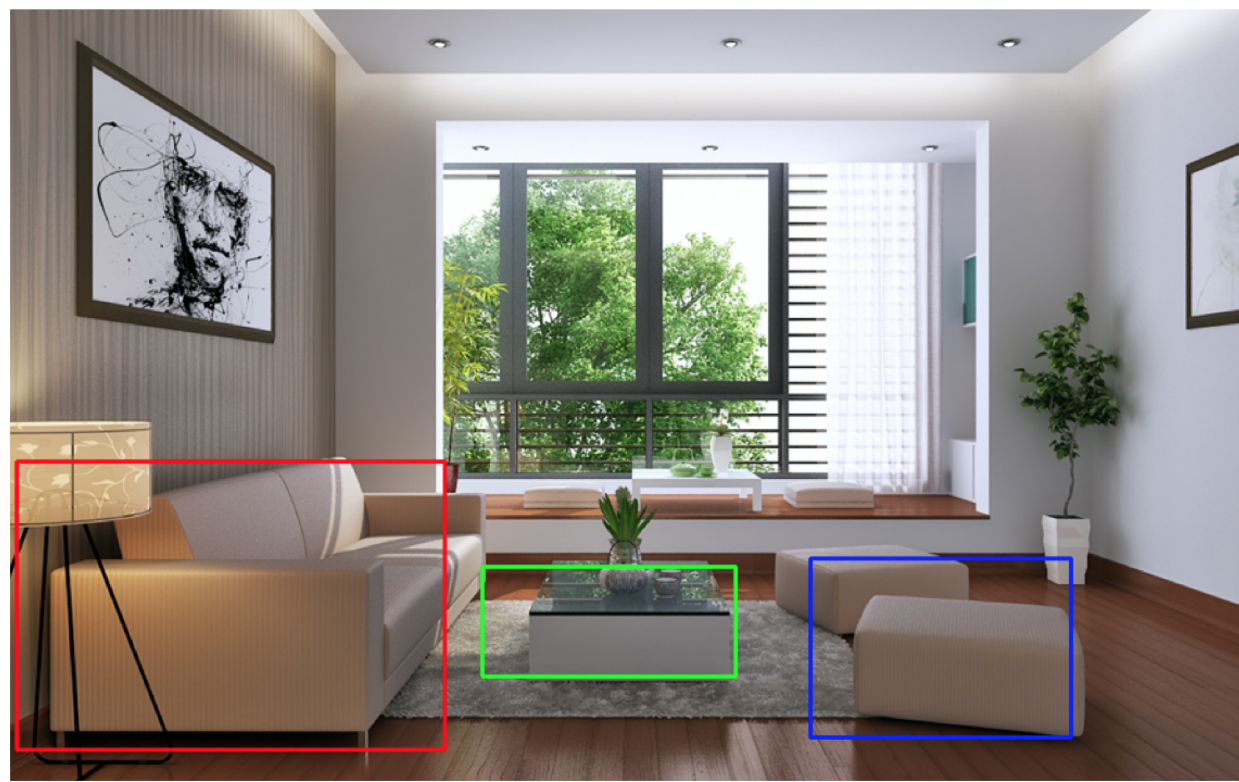


Man on Bench: **0.935**  
Bench holds Man: -2.779  
Bench plays Man: -3.090  
...



Man at Table: **2.164**  
Man at Table: **2.164**  
Man at Table: **2.131**  
Man at Table: **2.124**  
Man holds Man: -2.101  
Man holds Man: -2.101  
...

Attributes:



Sofa bed is (made of)Textile: **0.995**  
Sofa bed is (made of)Textile: **0.875**  
Table is Wooden: **0.829**  
Sofa bed is (made of)Leather: **0.702**  
Table is Transparent: **0.364**  
Sofa bed is (made of)Leather: **0.193**  
Sofa bed is Wooden: 0.133  
...

Total:

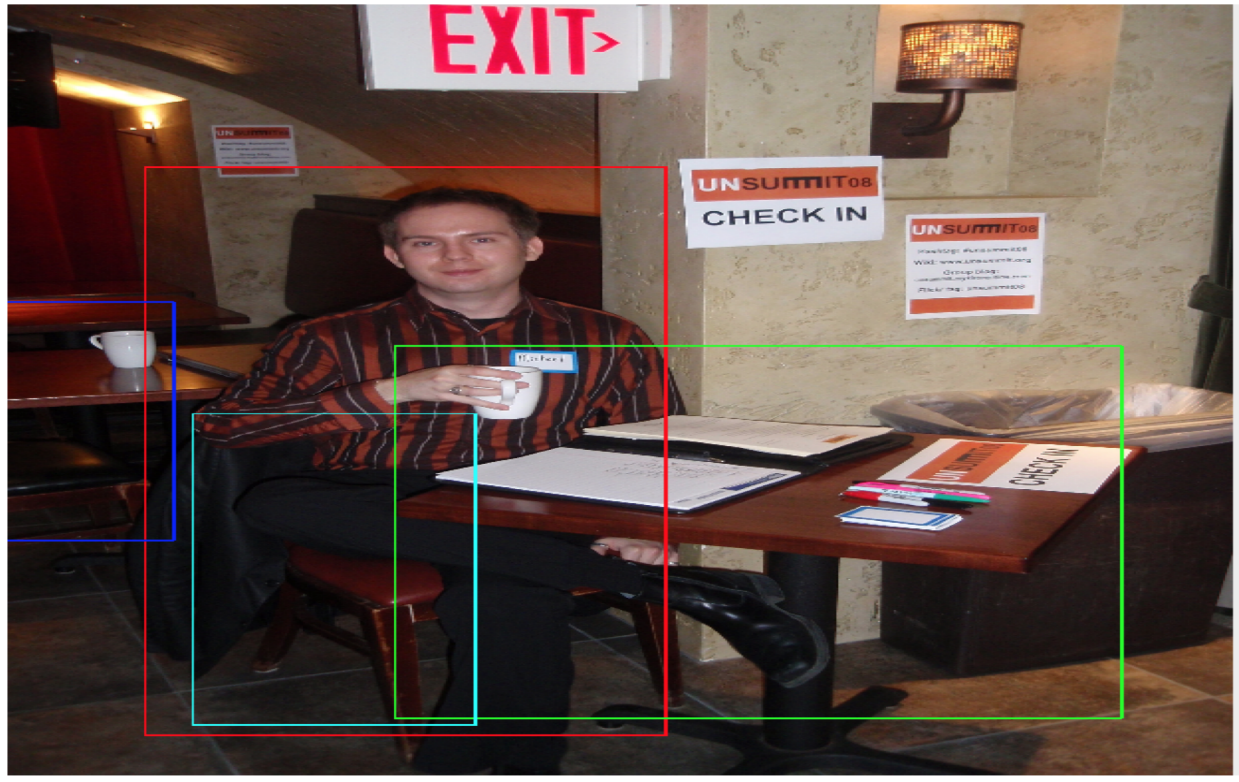


Table is Wooden: **0.998**  
Table is Wooden: **0.998**  
Chair is Wooden: **0.621**  
Chair is (made of)Leather: **0.582**  
Man on Chair: **0.598**  
Man at Table: **0.388**  
Man at Table: **0.364**  
Chair at Table: **0.261**

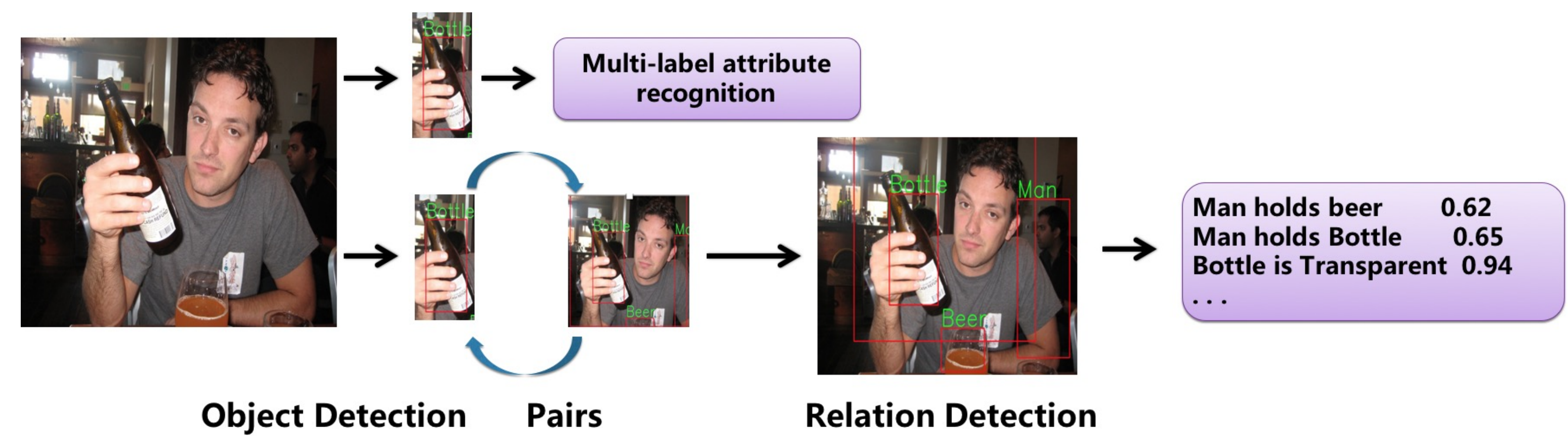
## Leaderboard

$score = 0.4 * rel\ mAP + 0.2 * Recall@50 + 0.4 * phrase\ mAP$

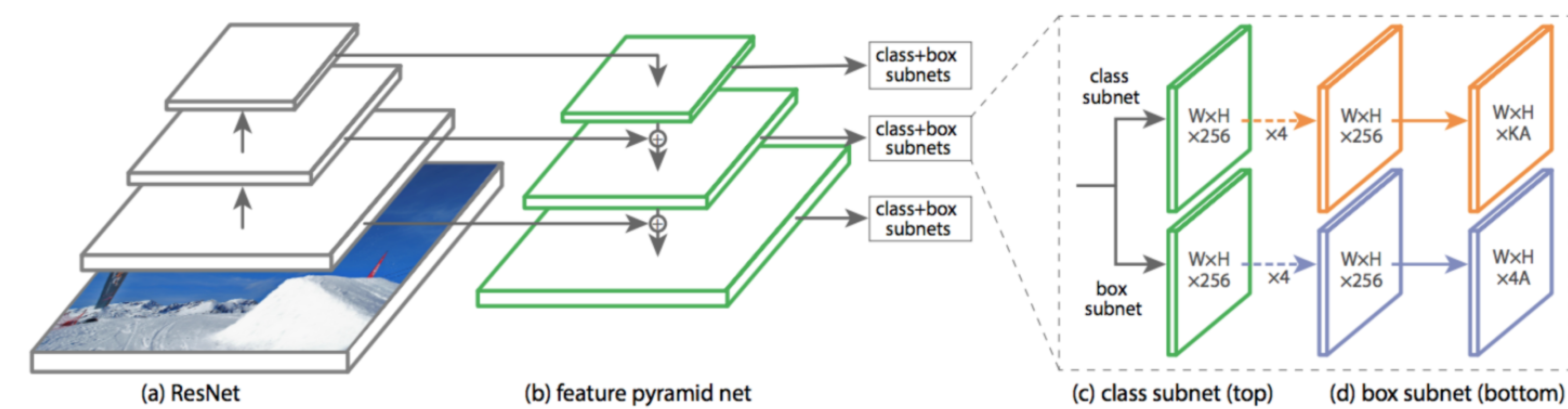
|                     |       |
|---------------------|-------|
| Competition score   | 0.068 |
| Attributes accuracy | 0.897 |

## Method

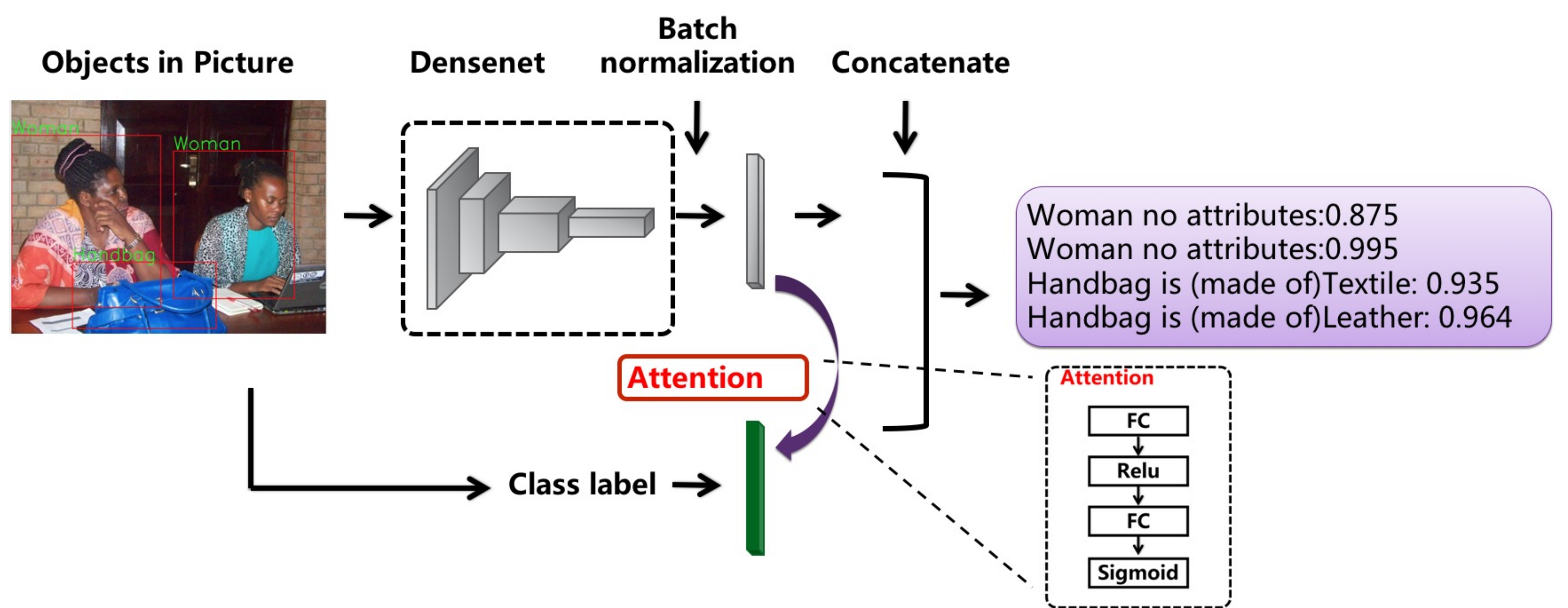
The proposed framework for visual relationship detection:



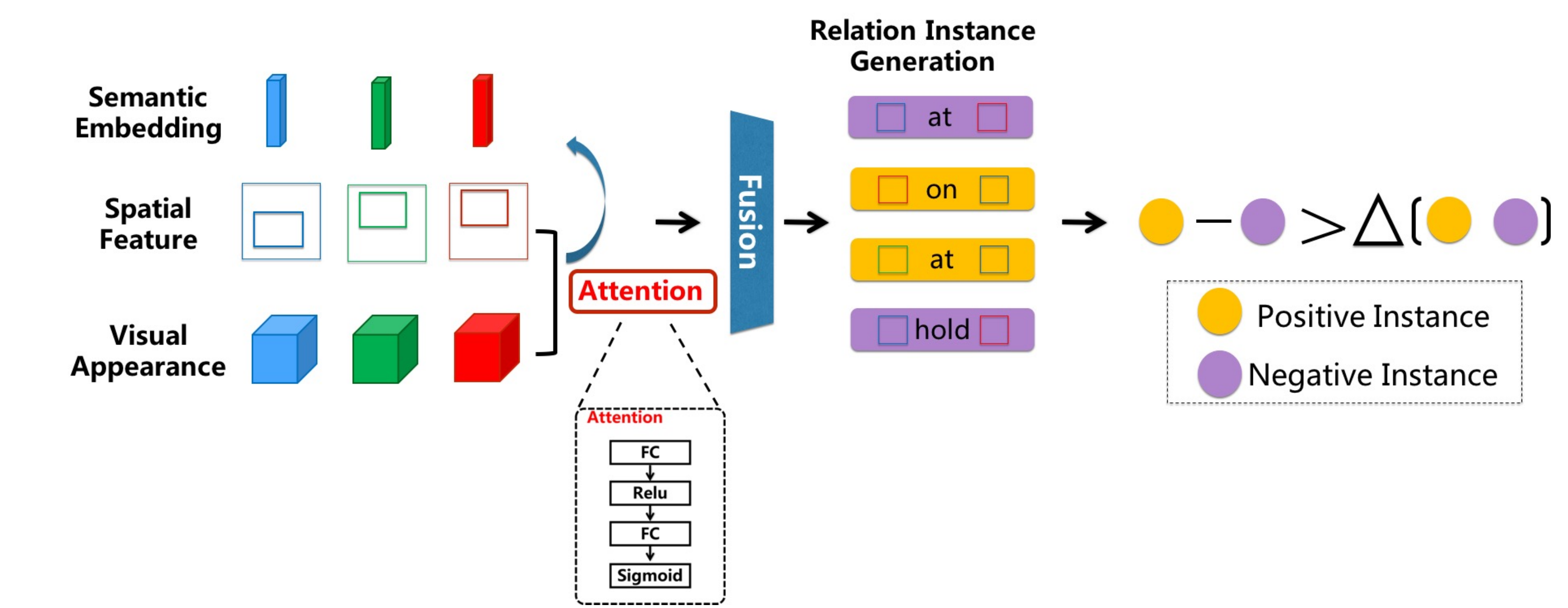
Object Detection : Retina Net[2]



Multi-label Attributes Recognition:



Relation Detection : DSR[4] with Attention



## Reference

[1] Lu, C., Krishna, R., Bernstein, M., & Li, F. F. Visual relationship detection with language priors. *ECCV*, 2016  
[2] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. Focal loss for dense object detection. *ICCV*, 2017  
[3] Dai, B., Zhang, Y., & Lin, D. Detecting visual relationships with deep relational networks. *CVPR*, 2017  
[4] Liang, K., Guo, Y., Chang, H., & Chen, X. Visual Relationship Detection with Deep Structural Ranking. *AAAI*, 2018