

## Introduction

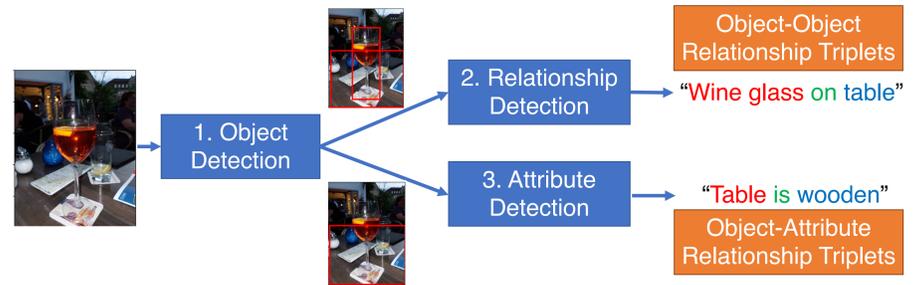
### Problem

Current computer vision methods provide isolated object classes from an image, but how do they interact with each other? What are they made of?

### Goal

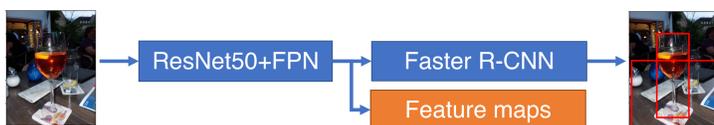
- Detecting the relationships between pairs of objects in the given image.
- Detecting objects that are made of a certain material (attribute relationship).

## Method overview



## 1. Object detection

We use FPN (Feature Pyramid Network) [Lin, et al., 2017] to detect small objects. Intermediate output from FPN are used in relationship detection.

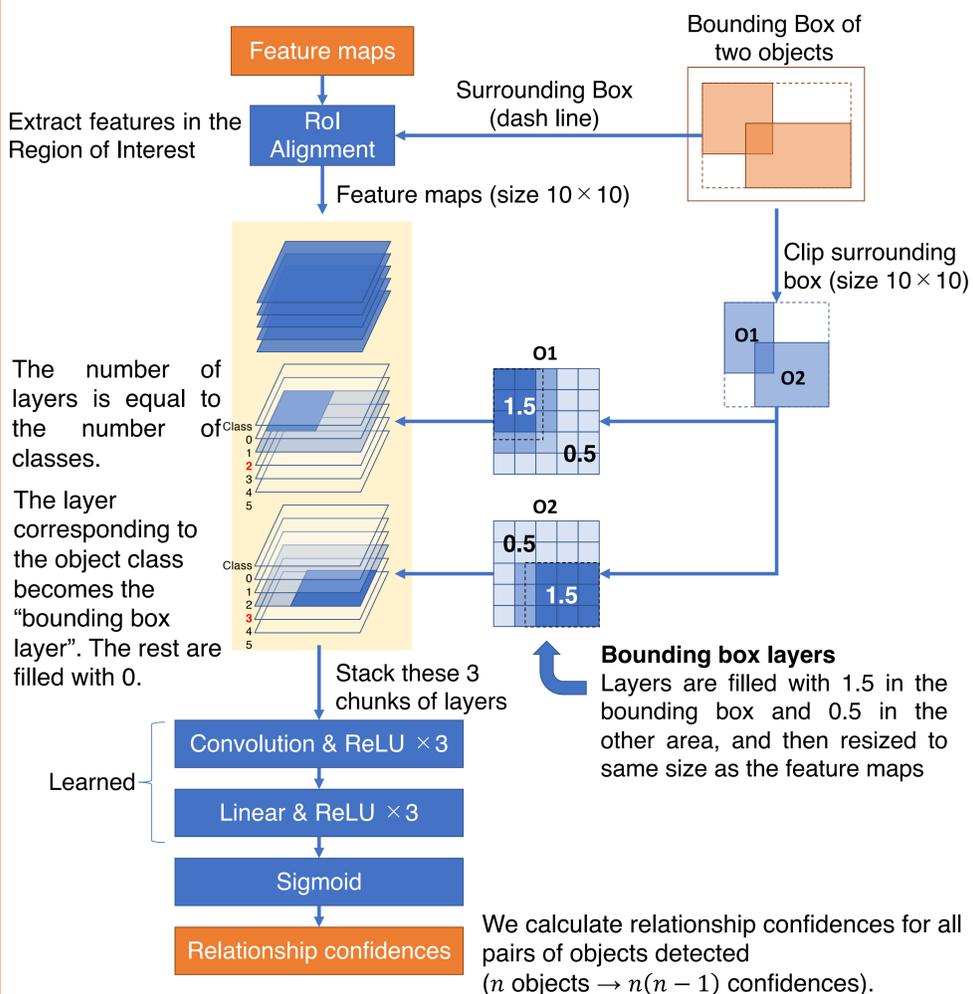


## 2. Relationship detection

### Difference with the related work in relationship detection

Previous works take as their input vectorized bounding boxes [*x coordinate of an object*, *ratio of the width of two objects*, ...] that also encode object labels [Zhuang, et al., 2017].

Our method uses bounding boxes as "bounding box layers" that indicate the position of the object in the feature maps, and uses labels to sort "bounding box layers" according to the object class.

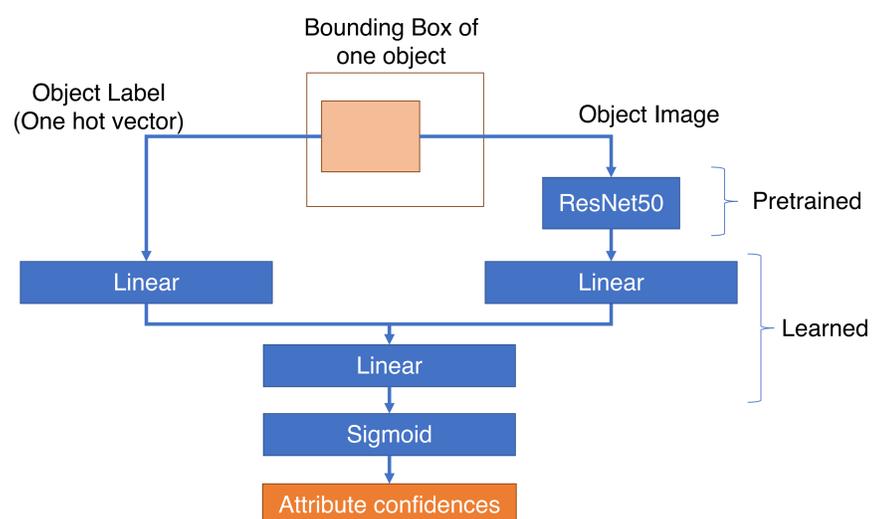


### Merits of our "bounding box layer" network

- Our "bounding box layers" allow associating bounding boxes with their respective features.
- Our "bounding box layers" allow emphasizing the features contained in the bounding box of the detected objects.
- Sorting our "bounding box layers" by class allows the network to learn different object types in a different way.

## 3. Attribute detection

Apart from the well-known ResNet classifier, our attribute detector also uses the object label to restrict the possible attributes (e.g., a glass is likely to be transparent).



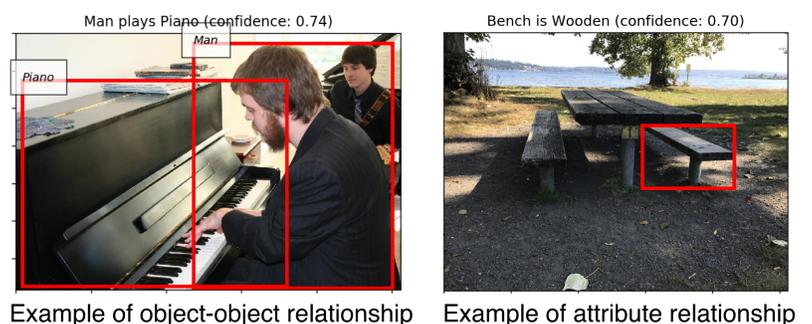
## Evaluation

### Scores in the leader board

1. Mean Average Precision(mAP) at IoU > 0.5 focusing on relationships
  2. Recall@50 focusing on relationships
  3. Mean Average Precision(mAP) at IoU > 0.5 focusing on phrases
- The weights applied to each of the 3 metrics are [0.4, 0.2, 0.4]

	Relationship	Attribute	All
Public	0.17934	0.04183	0.21774
Private	0.15776	0.04167	0.19666

We did not resort to ensemble methods.



## Conclusion

- Our "bounding box layer" network allows improving relationship detection between pairs of objects by emphasizing their features.
- Our "bounding box layers" can be used in other networks that use object detection results.