



Code: <https://github.com/wonheeML/mtl-ssl>



Wonhee Lee
Seoul National University

Joonil Na
Seoul National University

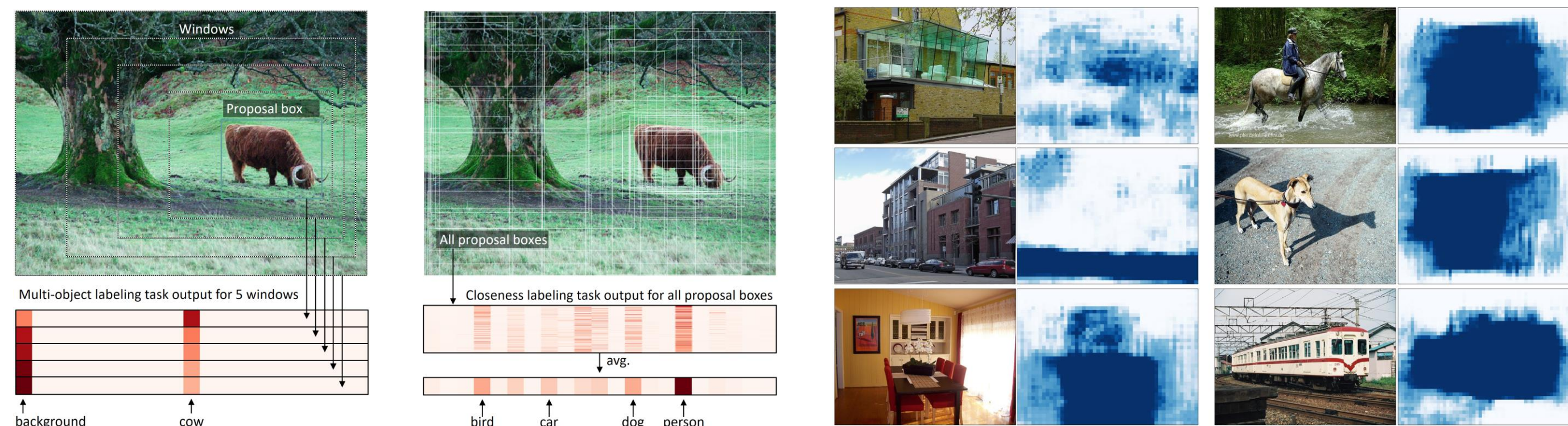
Gunhee Kim
Seoul National University

Objective

Improve the object detection accuracy using three ideas:

- **Multi-task learning**: train multiple relevant tasks jointly
- **Self-supervised learning**: train with labels generated by itself
- **Annotation recycling**: make use of given GT BBoxes in various ways

Three self-supervised auxiliary tasks

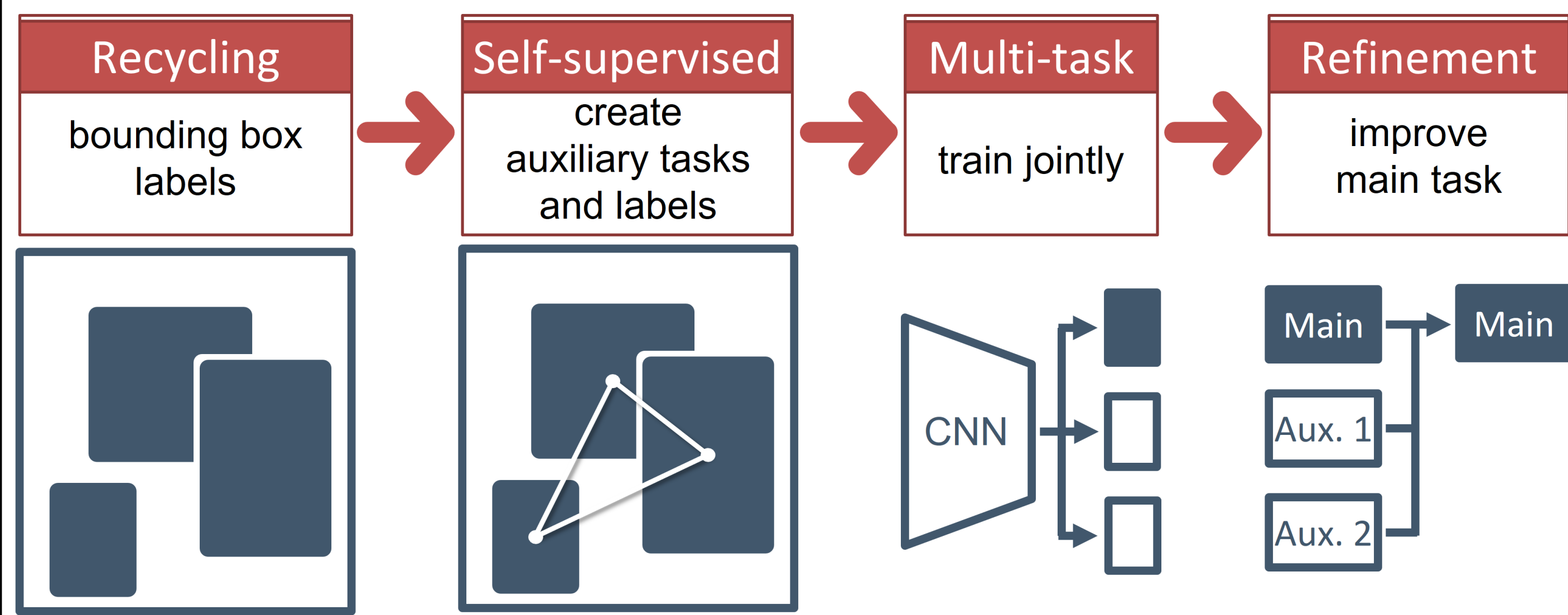


Contributions

- A first attempt to recycle BBox annotations for object detection
- Orthogonal to any proposal-based detection models
- Improvement (**mAP 2.0↑** on avg) in multiple architectures

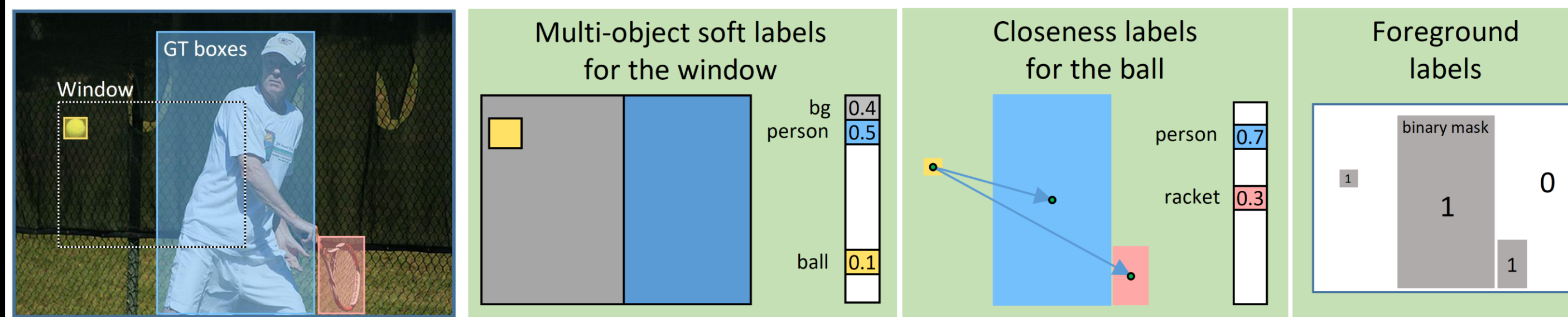
Our Idea

- Start from the conventional detection setting with BBox annotations
- Propose three auxiliary tasks that self-train by creating their own labels
- Train main and auxiliary tasks jointly for cooperative feature learning
- Refine the classification result using auxiliary task outputs

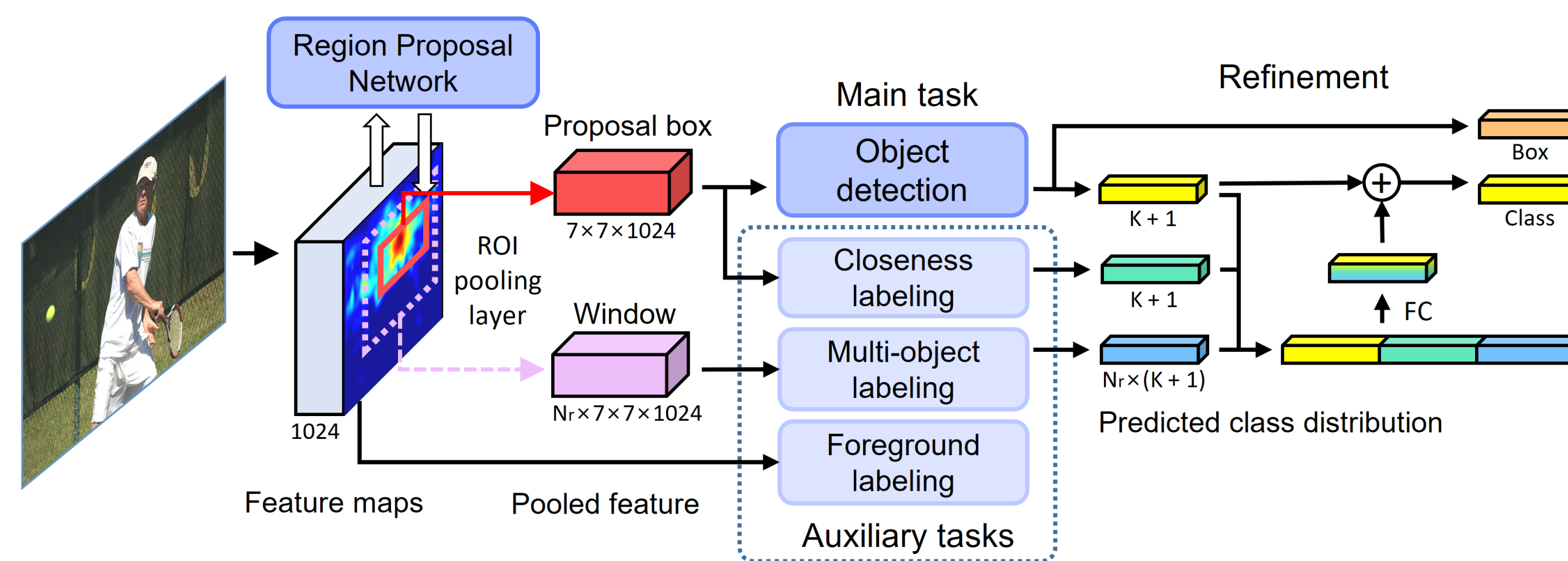


Three Auxiliary Tasks

1. Multi-object labeling
 - A soft label of a random window according to the portion of area for each class within it
 - Benefit: consider surrounding context information together
2. Closeness labeling
 - A soft label of a GT box by the distances to other nearby GT boxes
 - Benefit: predict objects in the neighborhood
3. Foreground labeling
 - A binary mask for foreground and background
 - Benefit: localize exact object boundaries



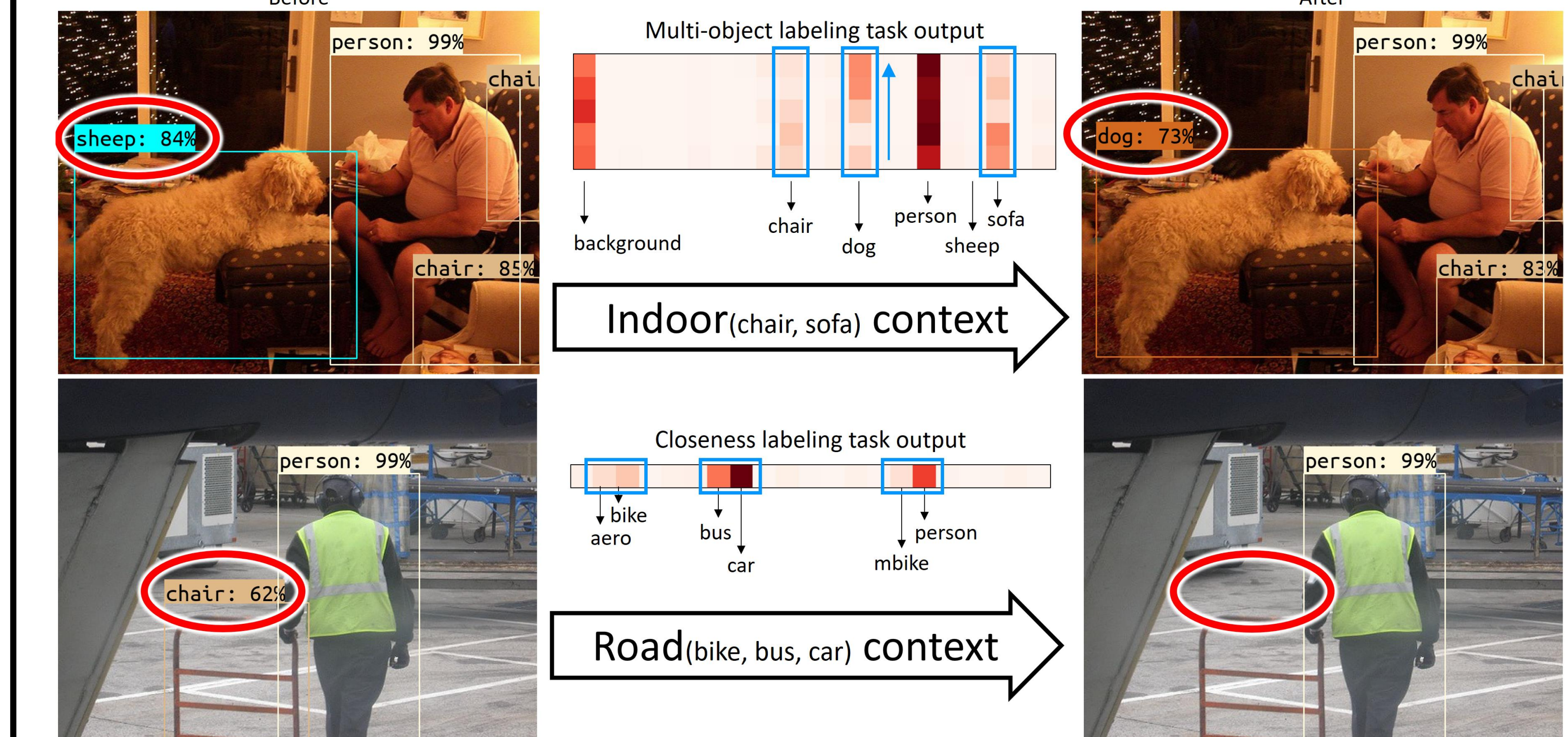
Architecture



How object detection improves?

- Multi-task learning enhances the feature learning via feature sharing between four tasks
- The refinement better predicts the class of a box by jointly using contextual information in the auxiliary task output

Refinement via Context Information



Results of Object Detection

Consistently enhance the performance in various combinations

Dataset	Backbone	Detector	Baseline	Ours	Δ mAP
VOC 2012	ResNet101	FRCNN	75.3	77.5	2.2 ↑
VOC 2012	ResNet101	R-FCN	72.1	73.9	1.8 ↑
VOC 2012	MobileNet	FRCNN	62.0	64.4	2.4 ↑
VOC 2012	Inception-ResNet-v2	FRCNN	78.2	80.0	1.8 ↑
COCO 2017	ResNet101	FRCNN	32.8	34.7	1.9 ↑

Examples of improvement for various types of errors

