Object Detection with a Unified Label Space from Multiple Datasets: Supplementary Material

Xiangyun Zhao[1], Samuel Schulter[2], Gaurav Sharma[2], Yi-Hsuan Tsai[2], Manmohan Chandraker[2,3], Ying Wu[1]

[1]Northwestern University [2]NEC Labs America [3]UC San Diego

The supplemental material contains the following items:

- **Sec. 1:** Details of the experimental setup and dataset
- **Sec. 2:** Additional qualitative results

1 Dataset details and experimental setup

Please find the latest information about the dataset at our project website: http://www.nec-labs.com/~mas/UniDet.

1.1 Experimental setup

We briefly recapitulate the experimental setup described in Sec. 3.5 of the main paper. Given are N object detection datasets annotated with bounding boxes for several object categories. Each dataset $\mathtt{D}_i$ has a label space $\mathtt{L}_i$ describing the object categories. The label spaces are not equal $\mathtt{L}_i \neq \mathtt{L}_j$ for different datasets $i \neq j$ but can share labels $\mathtt{L}_i \cap \mathtt{L}_j \neq \emptyset$, where $\emptyset$ is the empty set. The task is to train an object detector from the training sets of the N datasets that can predict over the *unified label space* $\mathtt{L}_\cup = \mathtt{L}_1 \cup \mathtt{L}_2 \cup \ldots \cup \mathtt{L}_N$, i.e., the union of all label spaces. For evaluation, the images from all validation/test sets are mixed together and augmented with bounding box annotations for missing categories. The detector does not know which image comes from what dataset and is required to predict all categories, unlike in [5]. We use average precision at 50% overlap (AP50) as performance metric, following the VOC protocol [1] as in [5].

1.2 Label spaces

To have a better understanding of the unified label space $\mathtt{L}_\cup$, we list all its categories along with their membership to the training datasets. For both settings described in Sec. 4.2 of the main paper, we list the unified label space in Tab. 1 and their membership to the original datasets.

One special consideration is important to note for constructing our unified label space. The SUN-RGBD [4] dataset originally contains the categories “tv” and “monitor” separately, but we merge them into one for compatibility with COCO [2] and VOC [1].

1.3 Annotation process

As described in the main paper, for an evaluation over the unified label space, new bounding box annotations are required. Specifically, after unifying the label spaces, certain datasets contain object categories that are not annotated. While the task we propose involves handling such missing annotations during training, we still need to evaluate the model. Thus, we collect the missing categories in all respective datasets for the validation/test sets. Please see our project website for more information: http://www.nec-labs.com/~mas/UniDet.

In particular, for each of the datasets in both settings, we pick 500 images of the validation/test sets at random and collect bounding box annotations for the missing object categories. This corresponds to 1500 images for each for the two settings, respectively. The missing categories for each dataset and setting can be derived from Tab. 1. For instance, the LISA-Signs [3] dataset in setting A has annotations for traffic signs "warning", "speedlimit", "noturn" and "stop-sign". All other categories listed in Tab. 1a need to be annotated on the LISA-Signs dataset. We filter this list of missing categories based on prior knowledge of their existence in a dataset to make the annotation job easier. As an example, we exclude "pillow" from the list which is certainly not part of the LISA-Signs dataset [3] because it is a driving dataset and any potentially captured pillow would be rather small and barely visible. However, categories like "person" or "car" will certainly appear frequently in LISA-Signs and will be annotated.

2 Qualitative results

We show additional qualitative results in Figures 1 to 4 for different datasets. Note that, for both settings we have a single detector predicting over the union of all categories. Our test sets are a composition of images from multiple datasets, fully annotated, while the detector does not know which dataset the image comes from. We still show qualitative results for images from a single dataset to better highlight what categories the unified detector is able to detect in these images, although they are not part of the original label space for that dataset. For instance, in Figure 4, the categories "car", "truck" or "traffic-light" are not part of the original label space of the LISA-Signs dataset, which only contains traffic signs. Our unified detector is still able to exploit the information from other datasets at train time and predict those categories in images from the LISA-Signs dataset at test time.

References

1. Everingham, M., Gool, L.V., Williams, C.K.I., Winn., J., Zisserman, A.: The Pascal Visual Object Classes (VOC) Challenge. IJCV **88**(2), 303–338 (Jun 2010) 1, 3, 4, 6
2. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common Objects in Context. In: ECCV (2014) 1, 3, 6

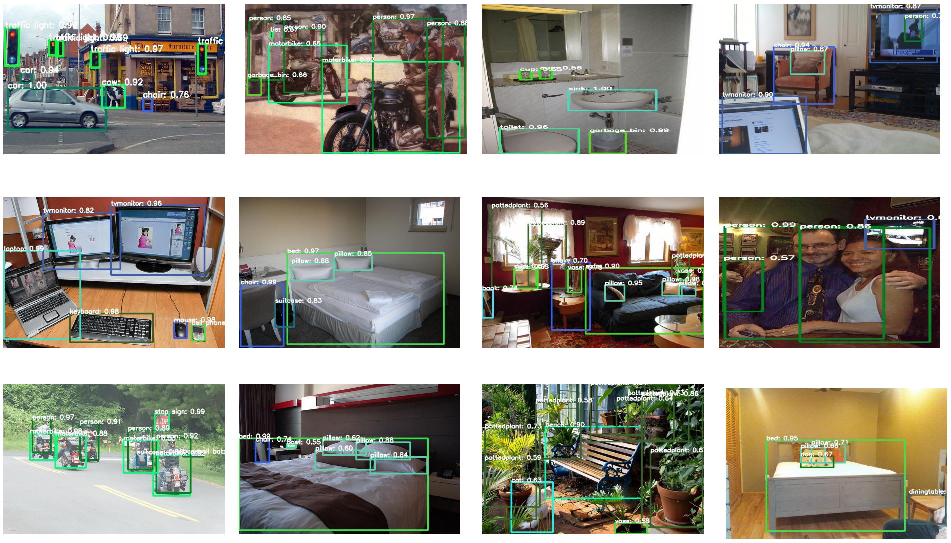

Fig. 1: Qualitative results on the COCO [2] testing set. In training, we remove VOC category annotations in COCO. Thus, categories like *cow*, *chair*, *car*, *tvmonitor*, *motorbike*, *pillow*, *garbage-bin*, *pottedplant*, *person*, *cat*, etc. are successfully discovered from the other two datasets in setting B, VOC [1] and SUN-RGBD [4]. Best viewed zoomed in color.

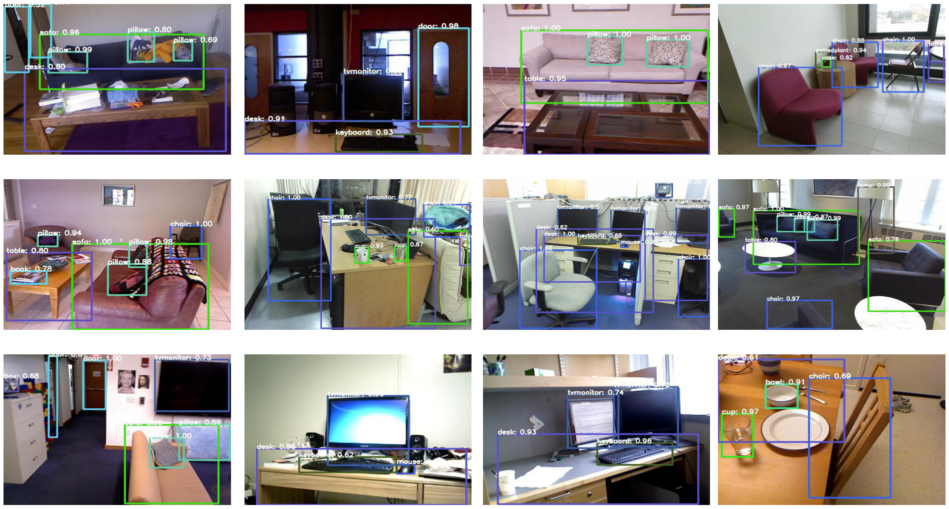

Fig. 2: Qualitative results on the SUN-RGBD [4] testing set from setting B. In SUN-RGBD, categories like *tvmonitor*, *keyboard*, *laptop*, *book*, *mouse*, *cup*, *bowl*, etc. are not in the original dataset, but detected by our unified detector. Best viewed zoomed in color

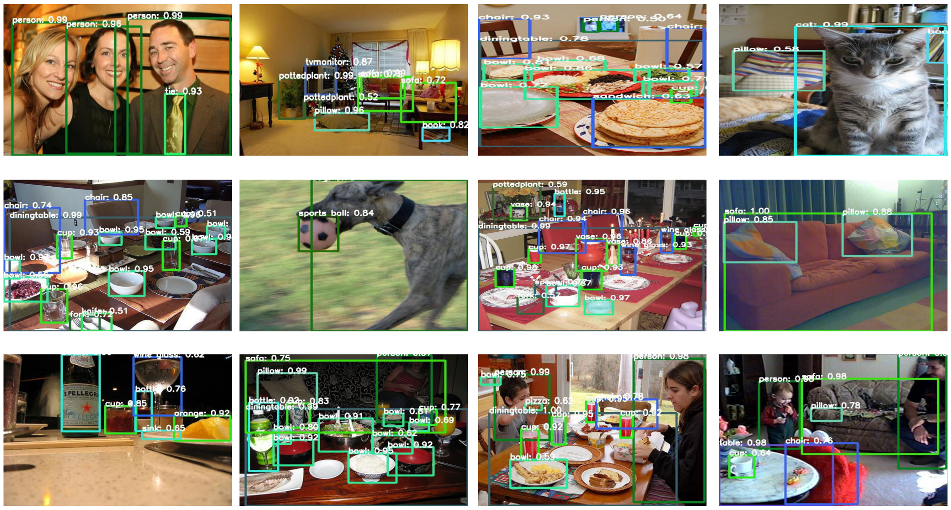

Fig. 3: Qualitative results on VOC [1] testing set from setting B. In VOC, categories like *tie*, *pillow*, *book*, *sandwich*, *bowl*, *sports ball*, *vase*, *cup*, *wine class*, etc. are not in the original dataset, but detected by our unified detector. Best viewed zoomed in color.

3. Møgelmose, A., Trivedi, M.M., Moeslund, T.B.: Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey. IEEE Transactions on Intelligent Transportation Systems **13**(4), 1484–1497 (Dec 2012) 2, 6
4. Song, S., Lichtenberg, S.P., Xiao, J.: SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite. In: CVPR (2015) 1, 3, 6
5. Wang, X., Cai, Z., Gao, D., Vasconcelos, N.: Towards Universal Object Detection by Domain Attention. In: CVPR (2019) 1

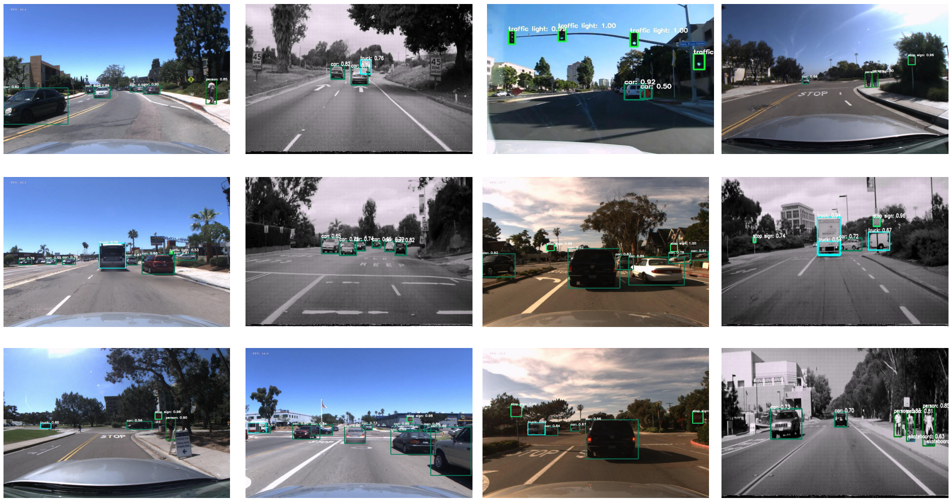

Fig. 4: Qualitative results on LISA-Signs testing set from setting A. In LISA-Signs, categories like *car*, *person*, *truck*, *traffic-light*, etc. are not in the original dataset, but detected by our unified detector. Best viewed zoomed in color.

Table 1: Tables (a-b) show the unified label space $\mathtt{L}_{\cup}$ of both settings defined in Sec. 4.2 of the main paper. Each table lists all categories along with their membership to the original datasets. All categories taken together in a table define $\mathtt{L}_{\cup}$. Categories that are shared between multiple original datasets are only listed once at their first occurrence, but also disclose other datasets they are contained in with brackets (∈ dataset-id). For instance, "chair" in (a) is part of datasets 1 and 2.

(a) Setting A

Datasets	Categories
1 VOC [1]	airplane, bicycle, bird, boat, bottle, bus, car, cat, chair (∈ 2), couch (∈ 2), cow, dining-table (∈ 2), dog, horse, motorcycle, person, potted-plant, sheep, train, tv (∈ 2),
2 SUN-RGBD [4]	bathtub, bed, bookshelf, box, counter, desk, door, dresser, garbage-bin, lamp, night-stand, pillow, sink, toilet,
3 LISA-Signs [3]	warning, speedlimit, noturn, stop

(b) Setting B

Datasets	Categories
1 VOC [1]	airplane, bicycle, bird, boat, bottle, bus, car, cat, chair (∈ 3), couch (∈ 3), cow, dining-table (∈ 3), dog, horse, motorcycle, person, potted-plant, sheep, train, tv (∈ 3),
2 COCO [2] (w/o VOC categories)	apple, backpack, banana, baseball-bat, baseball-glove, bear, bed (∈ 3), bench, book, bowl, broccoli, cake, carrot, cell-phone, clock, cup, donut, elephant, fire-hydrant, fork, frisbee, giraffe, hair-drier, handbag, hot-dog, keyboard, kite, knife, laptop, microwave, mouse, orange, oven, parking-meter, pizza, refrigerator, remote, sandwich, scissors, sink (∈ 3), skateboard, skis, snowboard, spoon, sports-ball, stop-sign, suitcase, surfboard, teddy-bear, tennis-racket, tie, toaster, toilet (∈ 3), toothbrush, traffic-light, truck, umbrella, vase, wine glass, zebra,
3 SUN-RGBD [4]	bathtub, bookshelf, box, counter, desk, door, dresser, garbage-bin, lamp, night-stand, pillow