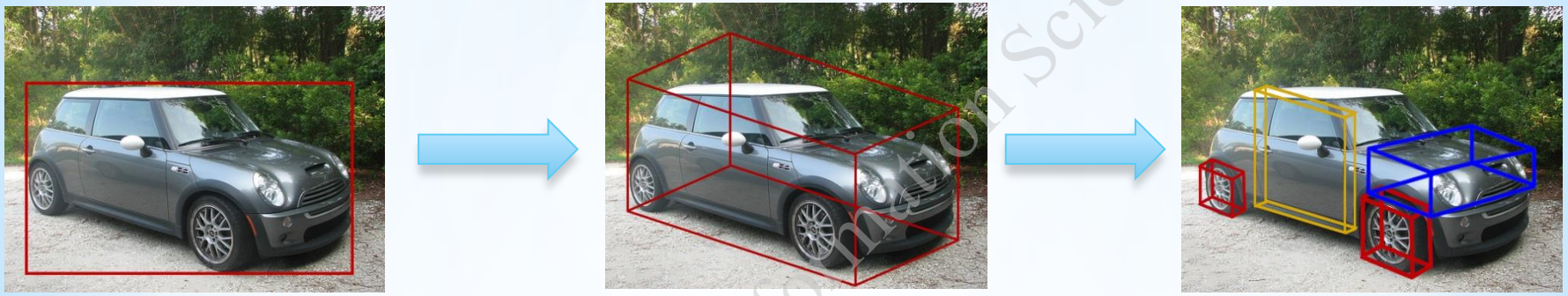
Localizing Object Parts in 3D from a Single Image

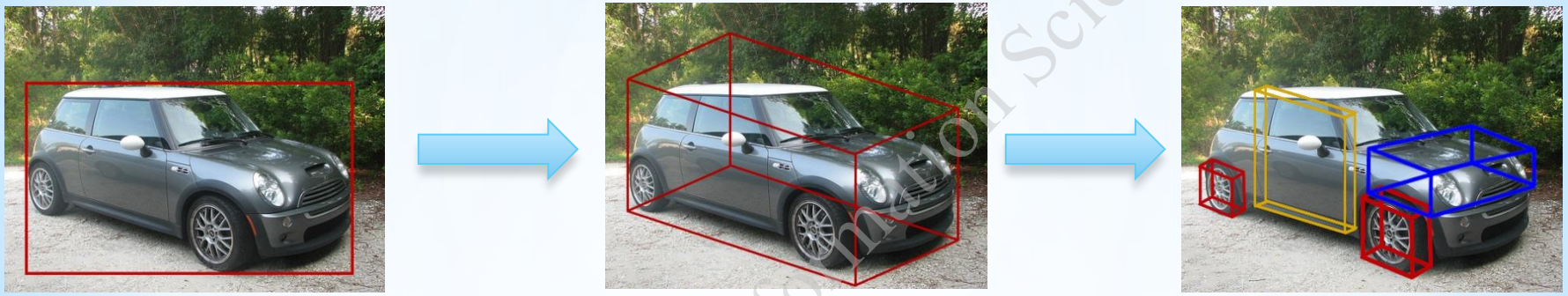
Shen YIN, Bin ZHOU , Mingjia YANG & Yu ZHANG

Beihang University

catalog



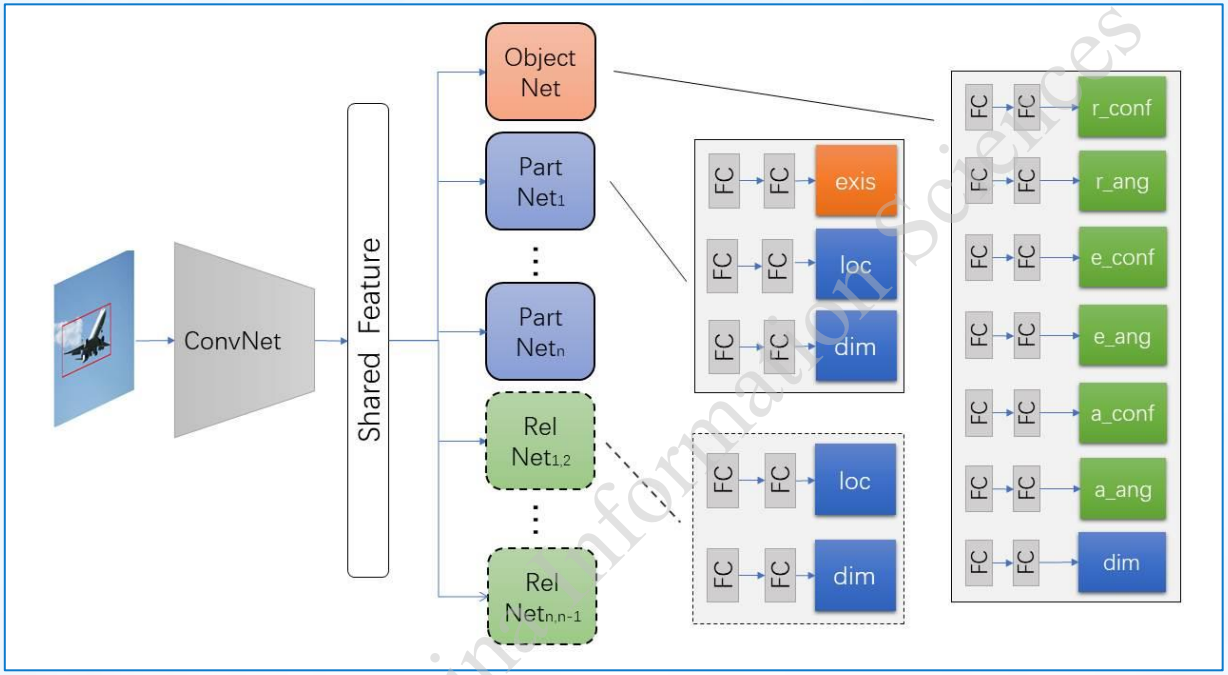
Object localization in 3D world from 2D images is an important computer vision problem that enables modern robotic vision systems to interact well with the objects present in the real world.



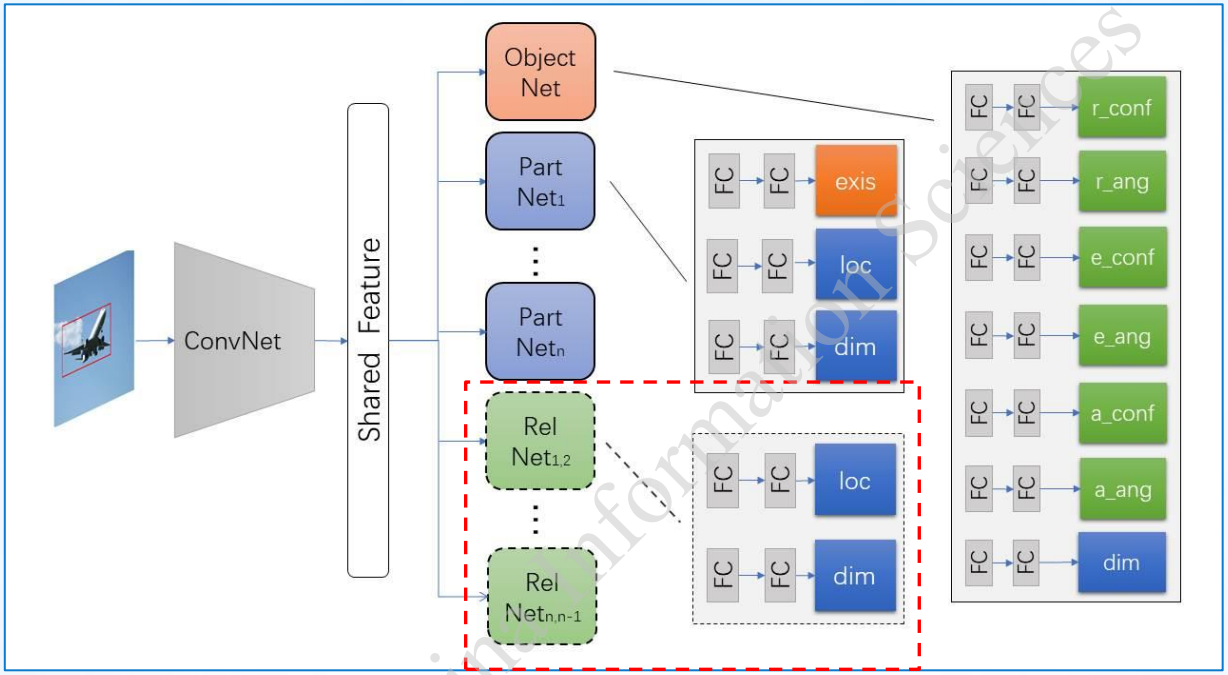
However, the existing approaches all focus on object-level inference instead of analyzing object parts. As in most time we expect a robot to interact with a semantic part of an object, part-level 3D object localization is more preferred in practical use.

Part-level 3D object localization is still a novel problem, both methodologies and datasets are scarce.

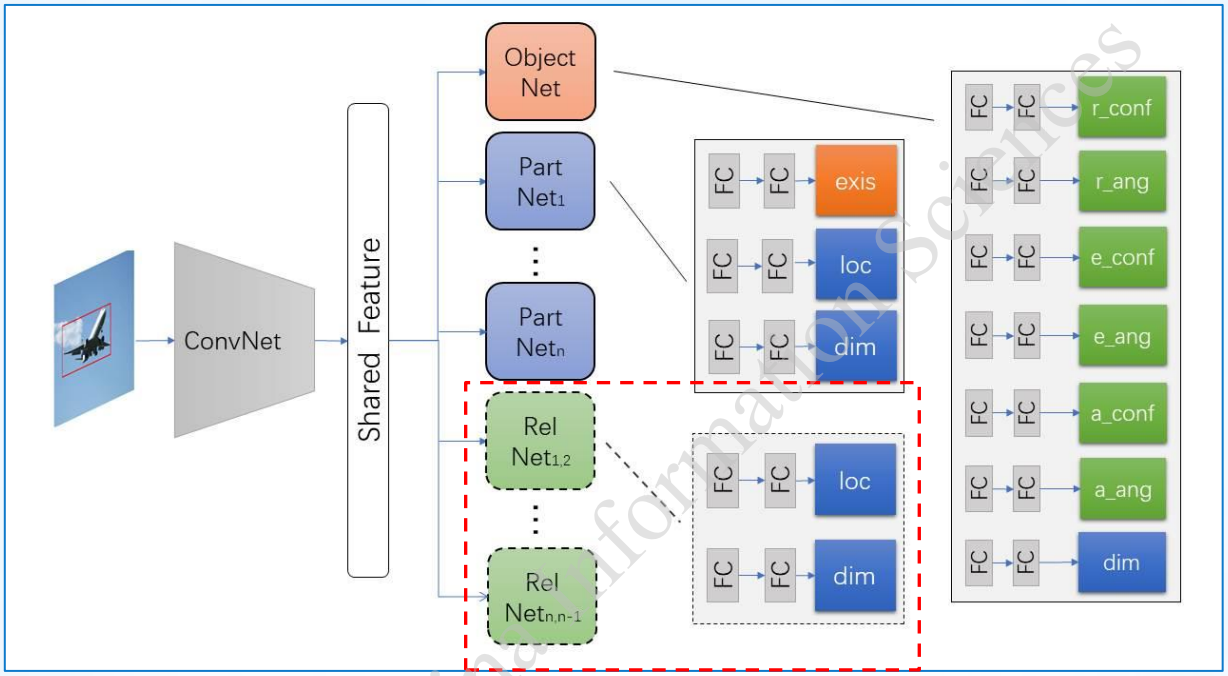
catalog



We first propose a convolutional neural network (solid part) that estimates the 3D parameters of all the parts jointly by extending a conventional approach for object-level 3D localization[A. Mousavian *et al*. CVPR17].
Different parameters are divided into several sub-networks, these sub-networks learn complex feature transformations and produce parameter estimates for parts or the object.

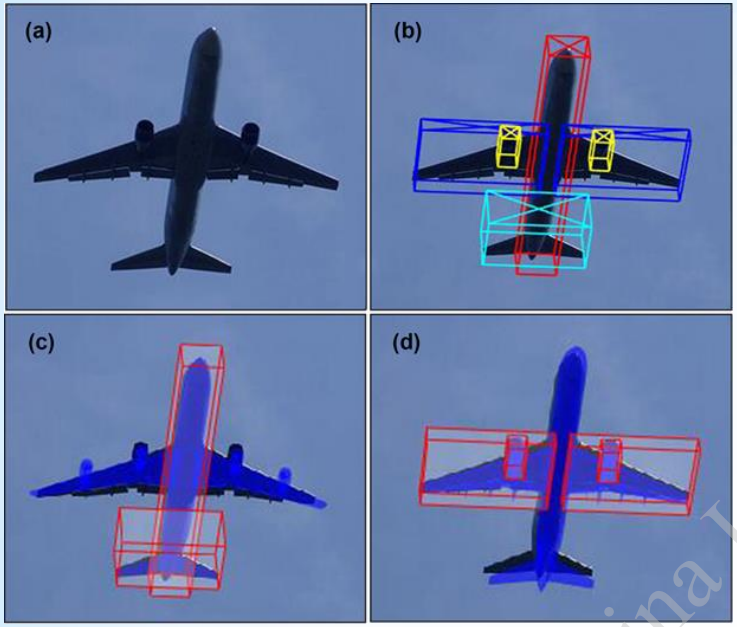


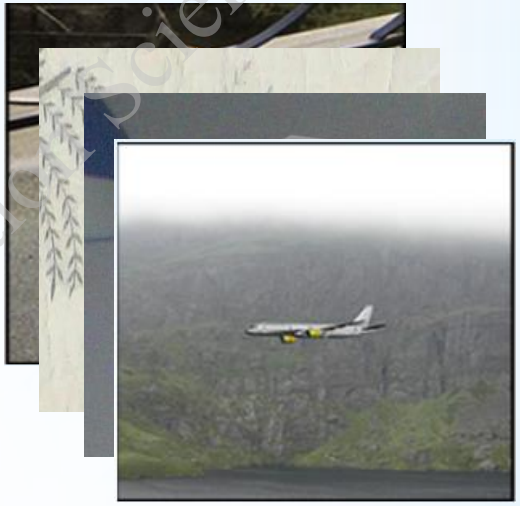
For many real-world objects, their parts follow strong structural relationships. Thus we explore such context information via a simple modification of the initial network (dash part).



$$\min_{\{\vec{a_i}^*\}_i} \sum_i \left(\vec{a_i}^* - \vec{\alpha_i}\right)^2 + \lambda \sum_{(i,j)} \left[\varphi\left(\vec{a_i}^* - \vec{a_j}^*\right) - \vec{\alpha_{i,j}}\right]^2$$

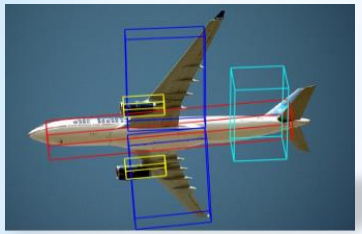
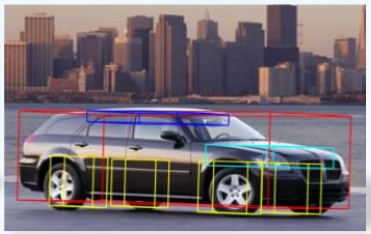
To utilize these pairwise estimates to fix individual estimation errors, we let the final estimates meet both the individual and pairwise predictions. The resulting problem is a least square system, thus can be solved efficiently by linear programming.

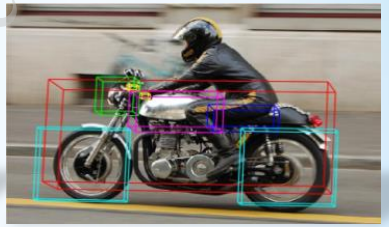
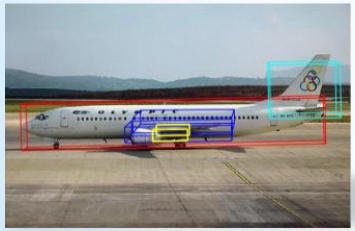
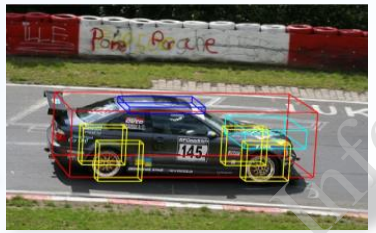
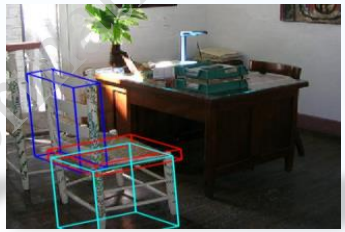




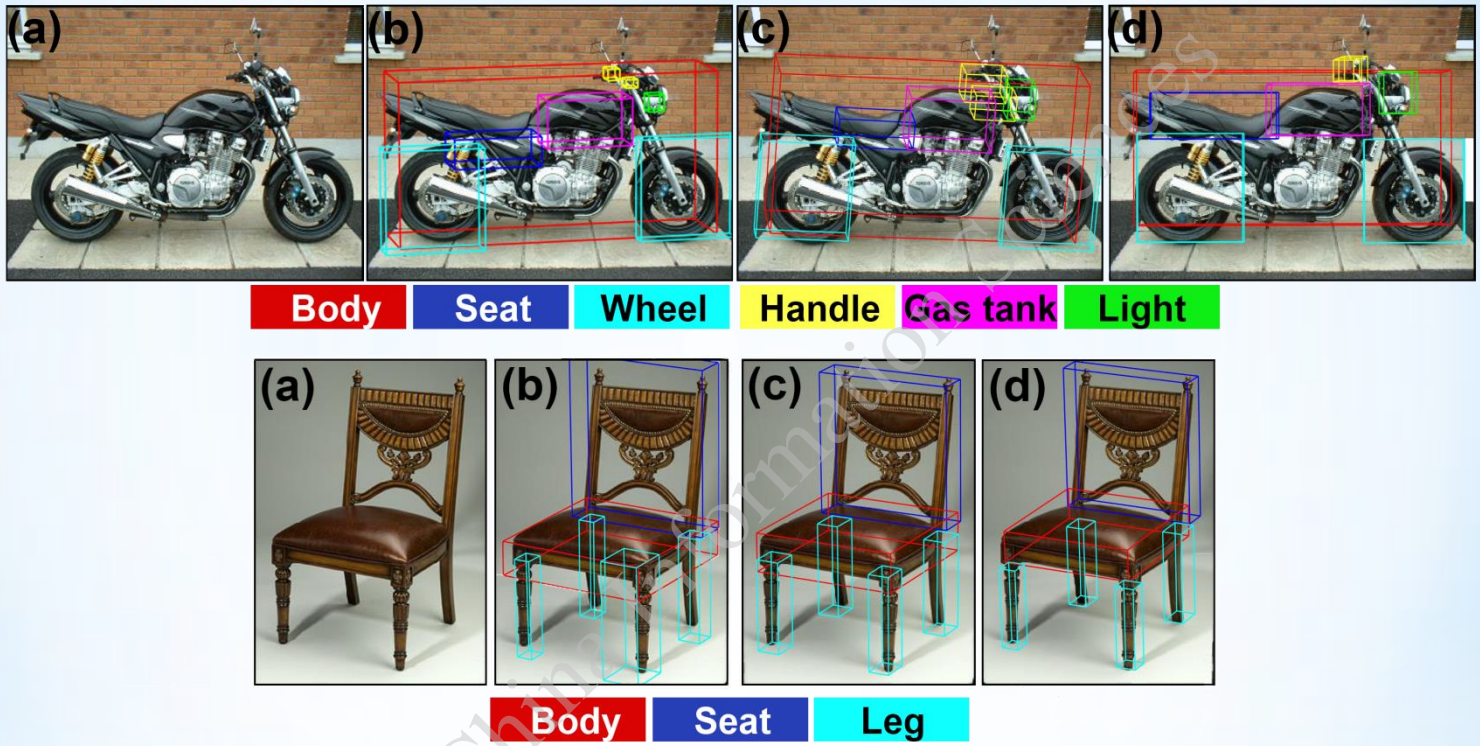
To generate sufficient training and validating data, we introduce an efficient data annotation procedure and provide rich annotations for several datasets. Besides, data augmentation is also explored by synthesizing images using the off-the-shelf part annotations in modern 3D model datasets.

catalog


We perform various experiments on Pascal3d+ and ObjectNet3D Dataset. Here we show several results on 5 categories, and more results are shown in the demo video.



Here we show the performance of the proposed network and the impact of spatial context. (a) is input image, (b) is baseline result, (c) is result of baseline with spatial context, (d) is the groundtruth. After incorporating the spatial context information, the localization IoU are improved consistently on all the categories. It can be shown that such improvements are more significant on the challenging small parts.

catalog

CONTRIBUTION

Proposing a novel yet important problem that localizes the object parts in 3D from a single image and its baseline network.

Proposing an improved network that explores the spatial context among object parts, which further boosts the performance.

Providing high-quality manual annotations for large-scale dataset to make this task able to be quantitatively evaluated.

In this work we focus on the task of part-level 3D detection from a single image. On evaluation experiments we achieve an encourage result and more studies need to be carried out. Future work should contain a more effective network concerning viewpoint and 3D box both, larger and more accurate datasets would be helpful also.

Thanks!