SCIENCE CHINA
Information Sciences

• NEWS & VIEWS •

September 2022, Vol. 65 197101:1–197101:2
https://doi.org/10.1007/s11432-022-3513-y

RPC: a large-scale and fine-grained retail product checkout dataset

Xiu-Shen WEI[1*], Quan CUI[2], Lei YANG[3], Peng WANG[4],
Lingqiao LIU[5] & Jian YANG[1*]

[1]*School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing* 210094*, China;*
[2]*Graduate School of IPS, Waseda University, Tokyo* 808-0135*, Japan;*
[3]*Megvii Technolody, Beijing* 100080*, China;*
[4]*School of Computing and Information Technology, University of Wollongong, Sydney* 2170*, Australia;*
[5]*School of Computer Science, University of Adelaide, Adelaide* 5005*, Australia*

Received 26 March 2022/Accepted 30 May 2022/Published online 3 August 2022

Citation Wei X-S, Cui Q, Yang L, et al. RPC: a large-scale and fine-grained retail product checkout dataset. Sci China Inf Sci, 2022, 65(9): 197101, https://doi.org/10.1007/s11432-022-3513-y

Over recent years, emerging interest has occurred in integrating computer vision technology into the retail industry. Automatic checkout (ACO) is one of the critical problems in this area which aims to automatically generate the shopping list from the images of the products to purchase (Figure 1(a)). The main challenge of this problem comes from the large-scale and the fine-grained nature of the product categories as well as the difficulty for collecting training images that reflect the realistic checkout scenarios due to continuous updates of the products. Despite its significant practical and research value, this problem is not extensively studied in the computer vision community, largely due to the lack of a high-quality dataset. To fill this gap, in this essay we propose a new dataset, i.e., RPC, to facilitate relevant research. Compared with the existing datasets, ours is closer to the realistic setting and can derive a variety of research problems. Besides the dataset, we also benchmark the performance on this dataset with a cross-domain detection baseline method. Our RPC dataset is already publicly available at the website [1].

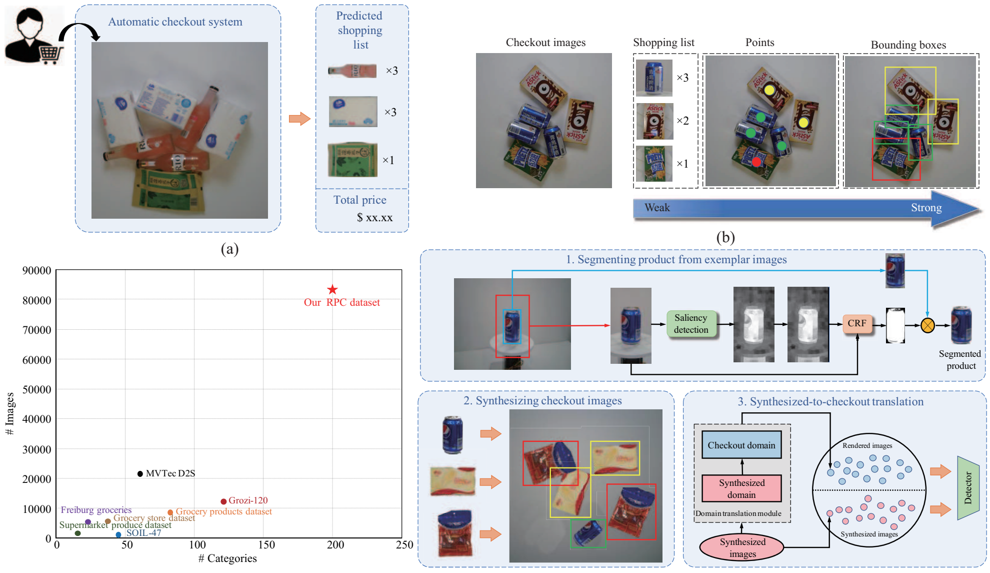
The RPC dataset enjoys the following characteristics.

- **Large-scale.** To collect this dataset, we choose 200 SKUs, which almost doubles the category size of previous largest dataset, e.g., [1]. In total, we capture 83739 images including 53739 single-product exemplar images, and 30000 checkout images. We also present comparisons with other datasets in the literature (Figure 1(c)).
- **Single-product exemplar images and checkout images.** In RPC, we collect two types of images. One type is the exemplar image for every single product and the other type is the checkout image taken at the checkout counter. While the exemplar images capture the multi-view appearances of the isolated SKU, the checkout images reflect realistic checkout scenarios where each image covers a variant number of product instances.
- **Hierarchical structure.** The hierarchical structure of product categories is another characteristic of RPC. The 200 SKUs can be categorized into 17 meta-categories that cover diverse appearances, such as bottle-like, box-like, canister-like, and bag-like. The SKUs under each meta-category tend to be fine-grained. The hierarchical structure can be exploited, for example, as auxiliary supervision information for advanced training or evaluation.
- **Close to realistic checkout scenario.** During the construction of this dataset, we try our best to mimic the realistic retail checkout scenarios to collect the checkout images. The products are randomly chosen and combined; they are freely placed on the checkout background with random orientations; occlusions and complex clutter are also common in our dataset.
- **Different clutter levels and illuminations.** In RPC, we split checkout images into three clutter levels based on the number of SKUs and product instances in each image. Such a clutter level annotation enables an in-depth inspection of the model capacities. In addition, each checkout image is acquired under three different illuminations.
- **Weak to strong supervision.** As shown in Figure 1(b), checkout images in RPC are provided with three different types of annotations, representing the weak to strong supervisions. (1) Shopping list, which records SKU categories and count of each product instance in checkout images. This is the weakest level of annotation and can be easily obtained in practice. (2) Point-level annotation, which provides the central position and SKU category of each product in checkout images. (3) Bounding boxes, which provide bounding box and SKU category for each product. This is the most labor-intensive annotation.

The introduction of different types of annotations further enriches the research directions that can be derived from this dataset, e.g., research on multi-category count-

* Corresponding author (email: weixs@njust.edu.cn, csjyang@njust.edu.cn)

[1] https://rpc-dataset.github.io/.



(a)

(b)

(c)

(d)

Figure 1 (Color online) (a) Illustration of the automatic checkout (ACO) application scenario. When a customer puts his/her collected products on the checkout counter, the system will automatically recognize each product and return a complete shopping list with total price. (b) Weak to strong supervisions of our RPC dataset: from shopping list, points, to bounding boxes. All checkout images in RPC are labeled with these three levels of annotations. (c) Comparisons with other related datasets in the literature. (d) Pipeline of our baseline method for the ACO task.

ing and weakly supervised detection. Besides, we would like to emphasize that, the ACO task does not essentially require labor-intensive annotations (e.g., bounding boxes). The shopping list annotation, as complete (but weak) supervision, is also promising (but challenging) to solve ACO.

Usage. The ACO problem is an open problem and has many potential solutions. To benchmark the proposed ACO dataset, we consider a baseline approach which formulates the ACO task from a cross-domain detection perspective. We restricted ourselves to only using the annotations from the single-product exemplar images, and we adopt feature pyramid network [2] as the detector.

The checkout image contains multiple objects while the exemplar image only has one. To reduce this gap, we propose to "copy-and-paste" the segmented isolated products to create synthesized checkout images. For segmenting the product instance, we adopt a salience-based object segmentation approach [3] with conditional random fields [4] for mask refinement. After the synthesis step, domain gap still exists between the synthesized images and checkout images. It is easy to tell the difference between the images from these two domains by observing lighting conditions or shadow patterns. In order to render the synthesized images more naturally similar to real checkout images, we employ Cycle-GAN [5] to translate these images into the checkout image domain. After that, we train detectors with both the rendered images and the synthesized images. We use it as our baseline. The pipeline of this method is shown in Figure 1(d). For experimental results, the baseline method achieves 56.68% ACO performance (see the website of RPC for more details) on average over three clutter modes, which shows the task is challenging and leaves substantial room for improvement.

Although we tackle ACO with a cross-domain detection strategy in our benchmark, there are many other possible solutions. Moreover, other possible research directions can be derived from RPC, to name a few. (1) Online learning for ACO. One challenge of real-world ACO is that the new product will be continuously added to the product list. Thus it is desirable to find a way to quickly update the system without retraining the model from scratch. (2) Using mixed supervision from the checkout images. Our dataset provides different levels of supervision for the checkout images. How to leverage those annotations for better solving ACO is still an open problem and needs more in-depth research. (3) As a complementary dataset for other vision tasks. Although our dataset is designed for ACO, it can also act as a dataset for research areas such as object retrieval, few-shot/weakly-supervised/fully-supervised object detection, since our annotations also include the ground truth location/bounding-box of products in the checkout images.

Acknowledgements This work was supported by National Key R&D Program of China (Grant No. 2021YFA1001100) and Natural Science Foundation of Jiangsu Province of China (Grant No. BK20210340).

References

1 Follmann P, Bottger T, Hartinger P, et al. MVTec D2S: densely segmented supermarket dataset. In: Proceedings of European Conference on Computer Vision, 2018. 569–585

2 Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 2117–2125

3 Hu P, Wang W Q, Zhang C, et al. Detecting salient objects via color and texture compactness hypotheses. IEEE Trans Image Process, 2016, 25: 4653–4664

4 Krähenbühl P, Koltun V. Efficient inference in fully connected CRFs with Gaussian edge potentials. In: Proceedings of Advances in Neural Information Processing, 2011. 109–117

5 Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2017. 2223–2232