EE P 596 A Project: Pull your Shelf together

Team: Rishabh Gupta, Sreejith S.

UNIVERSITY of WASHINGTON

PROBLEM STATEMENT

We propose a methodology for automatically identifying product misplacement on the shelves of a supermarket.

We aim to provide the following functionalities. The system should:

1. Be able to detect and classify a range of products.
2. Be able to identify a misplaced product.
3. Provide location information for the misplaced product.

Stretch Goal:

1. Detect out-of-stock products.

PROJECT SUMMARY

Our goal is to perform product classification and detection in a grocery mall setup. We aim to identify misplaced products to save time and effort for the employees. It would also help customers find desired products with ease.

We plan to implement the product position outlier detection over images in the Grocery dataset [1] through the following steps:

1. Product detection in the densely packed shelf is performed using SSD MobileNet V1 [2].
2. Classification of the products (which were detected in Step 1) is then performed using ResNet20 V1 [3].
3. Finally, we identify misplaced objects by using the product classes obtained in Step 2 and comparing their corresponding neighbours on the shelves.

We use [4] as the baseline for the detector and classifier in Step 1 & 2 respectively.

PROJECT SUMMARY (CONTINUED)

4. (**Stretch goal**) We plan to change the method of detecting misplaced objects to a deep neural network based anomaly detection approach (e.g., isolation forests).

5. (**Stretch goal**) We plan to identify the different kinds of gaps in the shelves (a. Separation gaps between racks and b. gaps between products on the shelves). We utilize the information about the product gaps to detect out-of-stock items.

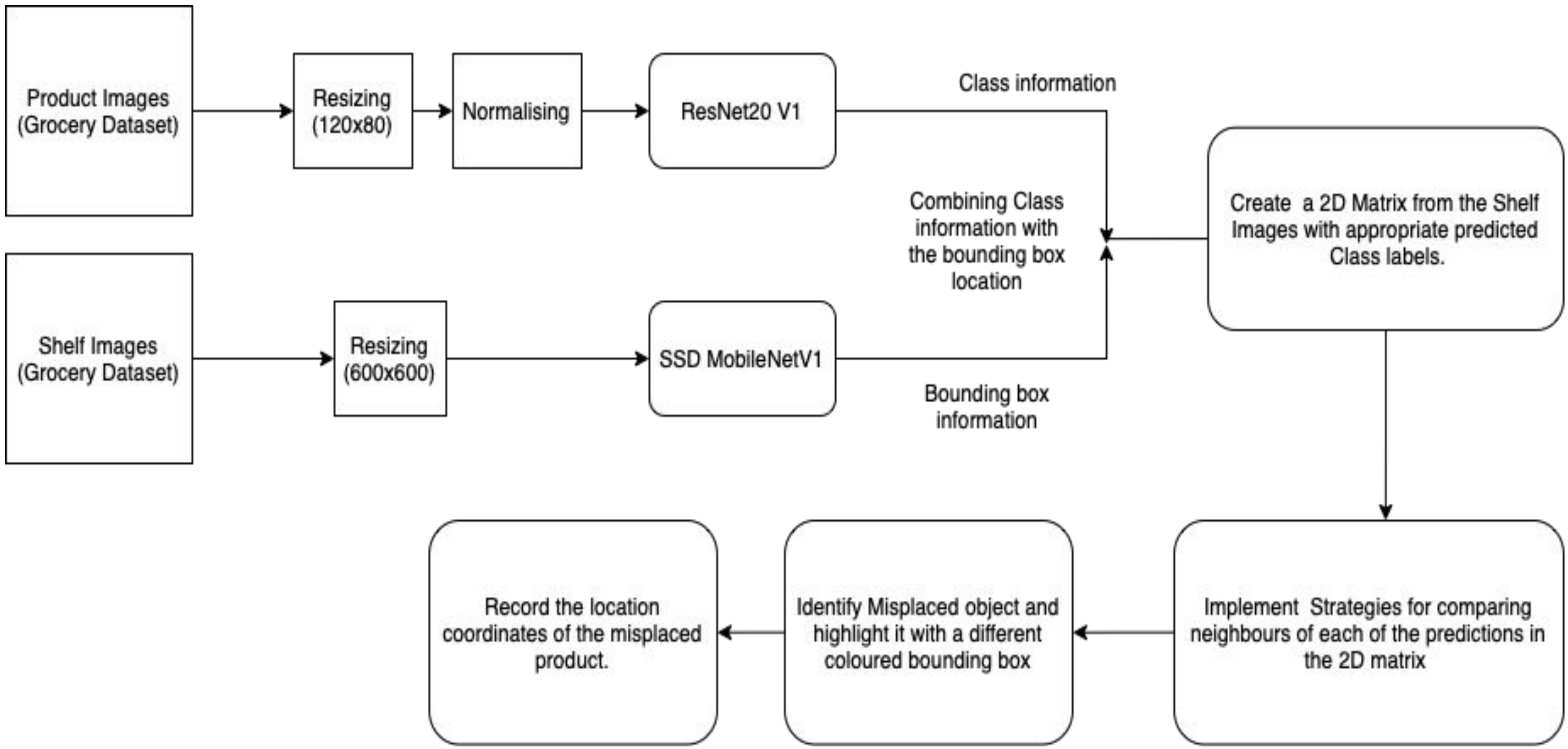
As the future scope, product detection can be made over a live video taken from different perspectives. Additional functionalities like suggesting the correct placement location of the product, keeping a count of the inventory, etc. can be added to the system.

UPDATED LITERATURE REVIEW

Over the course of the week, we have found the following additional paper that is relevant to our project:

1. [5] is relevant as it discusses detections in densely packed environments. The authors propose a novel, deep learning based method using a soft IoU layer and EM-Merger unit for precise object detection and it outperforms existing state-of-the-art with substantial margins. The authors also contribute an annotated dataset, SKU-110K [6], representing densely packed retail environments. We plan to evaluate our product misplacement identification strategy on the SKU-110K dataset in future.

PROPOSED METHOD



PROPOSED METHOD (CONTINUED)

1. Product detection in the densely packed shelf is performed using SSD MobileNet V1 [2].
    a. First, we resize the shelf images to 600x600 before feeding to the SSD (Single-Shot multibox Detection (SSD)) network.
    b. We perform detection using the pretrained SSD MobileNet V1.
    c. The Average Precision (AP) is chosen as the evaluation metric for the detection task.

2. Classification of the products (which were detected in Step 1) is performed using ResNet20 V1 [3].
    a. First, we resize the product images to 120x80 and normalise before feeding to ResNet20 V1.
    b. We train the ResNet20 V1 for 15 epochs with learning rate = 0.001 and a batch size of 50.
    c. During testing, we get the one-hot encoded output of the the product images, representing the corresponding class/label.
    d. The top-1 accuracy is chosen as the evaluation metric for the classification task.

PROPOSED METHOD (CONTINUED)

3. We identify misplaced products through the following procedure:
   a. Crop out each detected product in Step 2.
   b. Predict the class of the cropped products using the trained model from Step 1.
   c. We now have a 2D matrix with location and the predictions of the class of each product on a shelf.
   d. To get the misplaced object, we incorporate strategies to compare the neighboring elements of each of the predictions in the 2D matrix.
   e. We highlight the misplaced product with a different colored bounding box and also record its location coordinates.

BASELINE CLASSIFIER

We use [4] as the baseline for our product detector and classifier. The former is achieved by a generic product detection module trained on a specific class of products (i.e., the grocery dataset [1]). The Cascade object detection framework of Viola and Jones [8] is used for this purpose. [4] further uses Support Vector Machines (SVMs) to recognize the brand inside each detected region. [4] extracts both shape and color information; and apply feature-level fusion from two descriptors computed with the bag of words approach.

To the best of our knowledge, there has been no previous work on misplaced product identification on the Grocery dataset.

DATASET

- We use the Grocery dataset [1].
- The Grocery Dataset is an image dataset with:
  - 354 grocery shelf images were collected from ~40 stores, with four cameras.
  - Products from 10 different categories numbered from 1-10. Each category has a separate directory. There are 2744 product images in total.
- We split the data into train/test in a 70/30 ratio.
- We use 1846 train samples and 898 testing samples for training the product classifier.

DATASET (CONTINUED)

Product images

Shelf images

PRELIMINARY RESULTS

Using our trained product classifier, we were able to obtain the following Top-1 product classification accuracy using ResNet20 V1 (Vanilla):

Method	Top-1 Accuracy
Baseline	85.90
Proposed	73.05

We have also used the ground truth coordinates of the bounding boxes provided along with the dataset and identified misplaced objects on individual samples. Next, we will be implementing detection using MobileNet and identifying the misplaced objects on all the shelves.

PLANNED ABLATION STUDIES (FUTURE PLAN)

We plan to perform an ablation study with the following configuration to obtain the best performance:

We can use anomaly detection algorithms (e.g., Isolation forests) instead of performing the detection, classification, and then identifying the misplaced objects. Similar products placed adjacent will be detected using a single bounding box (instead of individual product-wise boxes). The anomaly detection algorithm finds products not belonging to the same class in a bounding box. This approach does not need product classification and is failsafe to cases where there are undefined number of product classes.

REFERENCES

[1] G. Varol, “Grocery Dataset,” GitHub, Nov. 30, 2021. https://github.com/gulvarol/grocerydataset (accessed Dec. 01, 2021).

[2] “TensorFlow Hub,” tfhub.dev. https://tfhub.dev/tensorflow/lite-model/ssd_mobilenet_v1/1/metadata/1 (accessed Dec. 08, 2021).

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” arXiv.org, 2015. https://arxiv.org/abs/1512.03385.

[4] “(PDF) Toward retail product recognition on grocery shelves,” ResearchGate. https://www.researchgate.net/publication/301439956_Toward_retail_product_recognition_on_grocery_shelves (accessed Dec. 08, 2021).

[5] E. Goldman et al., “Precise Detection in Densely Packed Scenes,” arXiv:1904.00853 [cs], Apr. 2019, Accessed: Dec. 08, 2021. [Online]. Available: https://arxiv.org/abs/1904.00853.

[6] “Papers with Code - SKU110K Dataset,” paperswithcode.com. https://paperswithcode.com/dataset/sku110k (accessed Dec. 08, 2021).

REFERENCES

[7] fizyr, “fizyr/keras-retinanet,” GitHub, Sep. 09, 2019. https://github.com/fizyr/keras-retinanet..

[8] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, doi: 10.1109/cvpr.2001.990517.

[9] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of CVPR, pages 4510–4520, 2018.