# Supplementary Material for $\mathcal{D}$-Extract

## 1. Overview

In this supplementary file, we would like to showcase few interesting results and present some additional details about the datasets used in the paper.

## 2. Filter classifier performance

We chose a light-weight network like MobileNetV3-L primarily for efficiency. MobileNetV3-L has default input size of 224 pixels and often might not detect the small texts and thin dimensional lines. To tackle the issue, we experimented and discovered that the network outperforms the default setting when the input image resolution is higher (600 pixels). Such higher resolution images enables the network to use ultra-fine details inside the product image and then make an inference. Empirical results suggest that the improvement caused by higher resolution plateaus after 600 pixels.

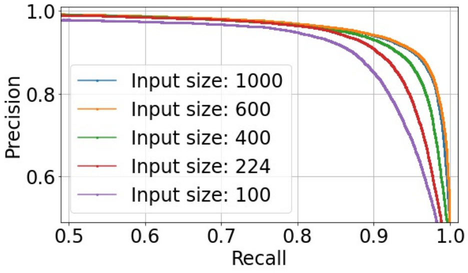

Figure 1. Precision-Recall curve of Filter Classifier on Dataset-1.

## 3. Effect of Statistical inference (S.I) to system performance

Statistical inference might be useful for products with less dimensional variations like bed or sofa, and not as useful for products which can often vary in shape and size. To understand this effect, we plot Figure 2 and find that there is significant improvement ($> 3\%$) in performance for product types with lower coefficient of variation (CV). Even for categories which has higher CV, there is significant ($> 1\%$) improvement when statistical inference is considered. Hence we can infer that it is a beneficial addition to the model's input.

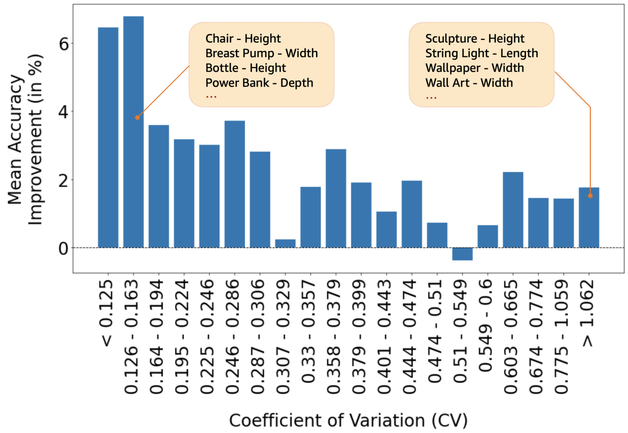

Figure 2. The mean accuracy improvement when Single box classifier uses S.I. compared to when it does not use S.I.

## 4. Product types used in Dataset `BBC_GDS`

Our experimental results show that the model operates at a F1-score of $88.9\%$, $89.6\%$ and $93.4\%$ for length, width and height attributes respectively when tested in a dataset comprising of 69 product types. The proposed system performs at similar levels across diverse product types (300+), which is not limited to furniture, but also to any mechanical and electrical appliances, tools and machineries, home decors, etc.

## 5. Effect of weight initialization

Although our 3-feature channels are not the same as the typical RGB image, it does look like a normal image with sharp highlights near the bounding box in question when the channels are stacked together (Figure 5.d of the paper). Hence when initialized with ImageNet weights, the network is able to detect various visual features like edges, shape and orientation from the product image, right from the early epochs of training. Also empirically, we find that the network converges much faster and better when initialized with ImageNet weights.