Computer Vision in Convenience Retail

Dennis Francis

The Images

Scraped from the web

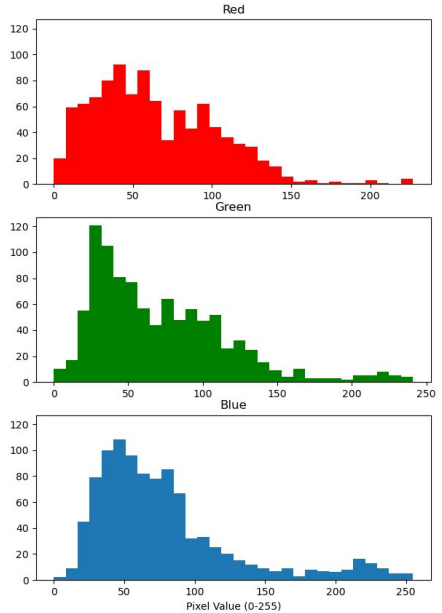
Labeled Data

ice cream

cigarettes

EDA

Histograms of RGB



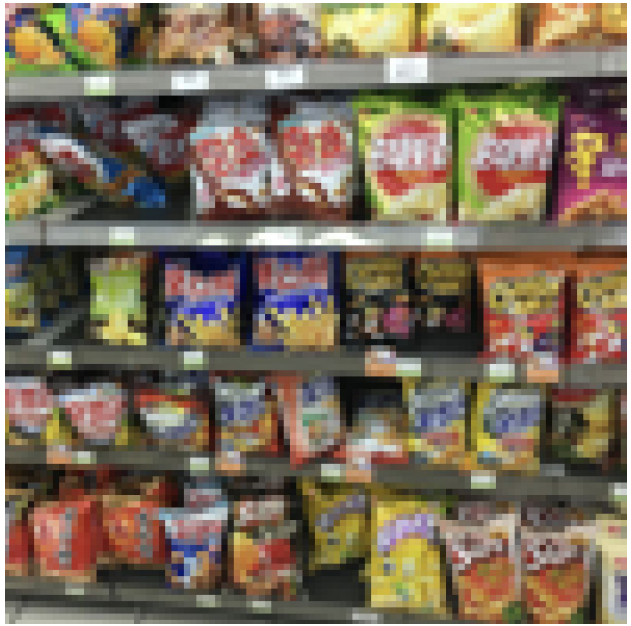
EDA

Observe the difference

in image features of

1600 x 1600 vs

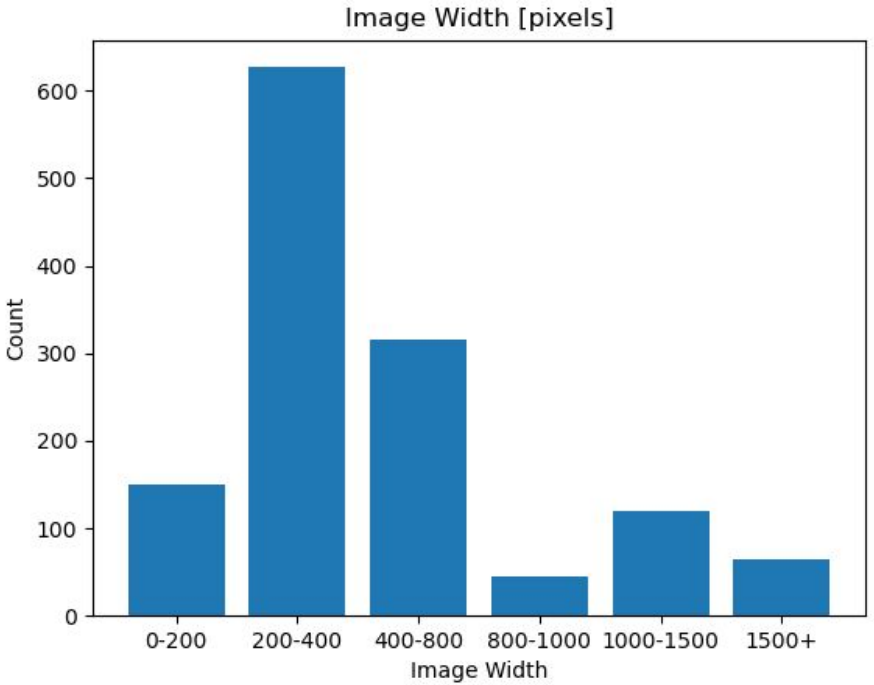
100 x 100 image



EDA

224 x 224 images are the

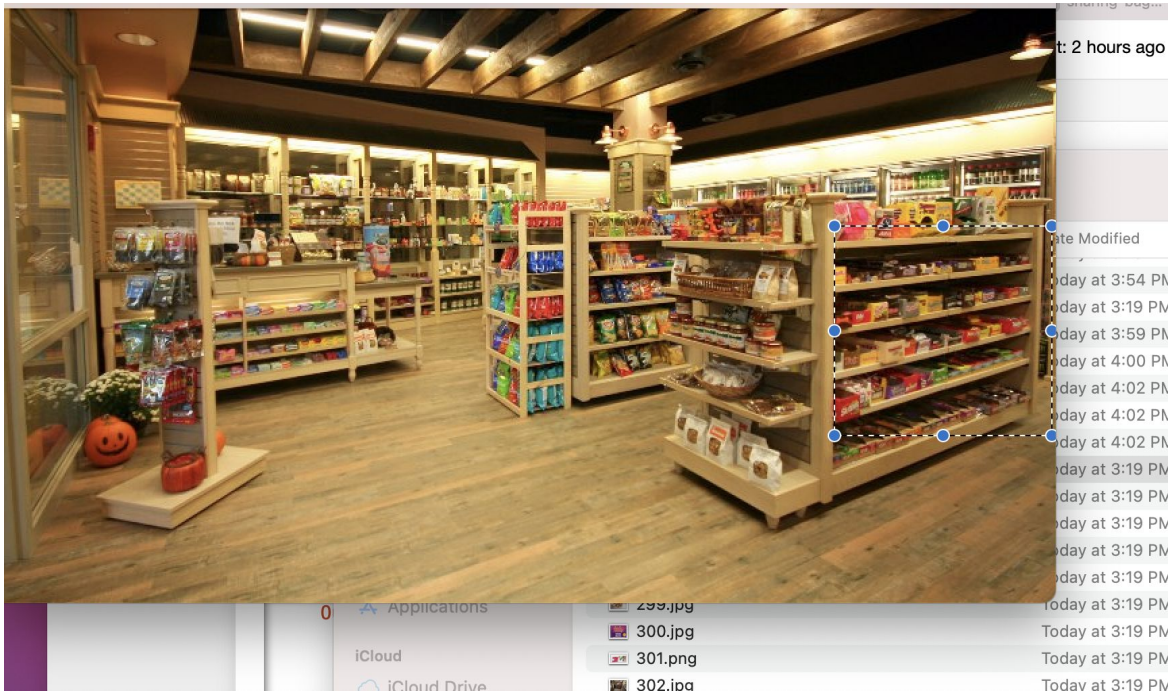
most common

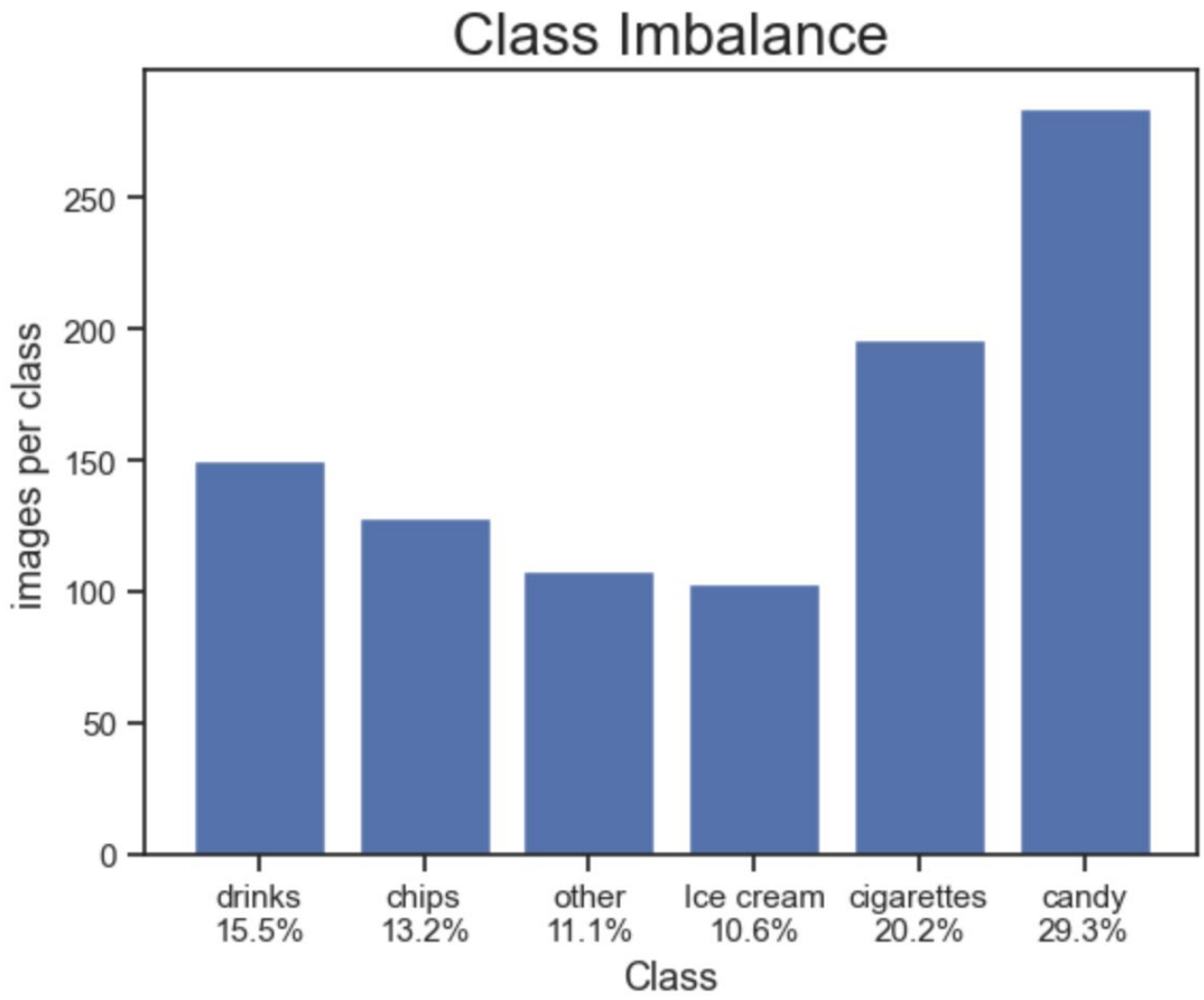


Cropping an image

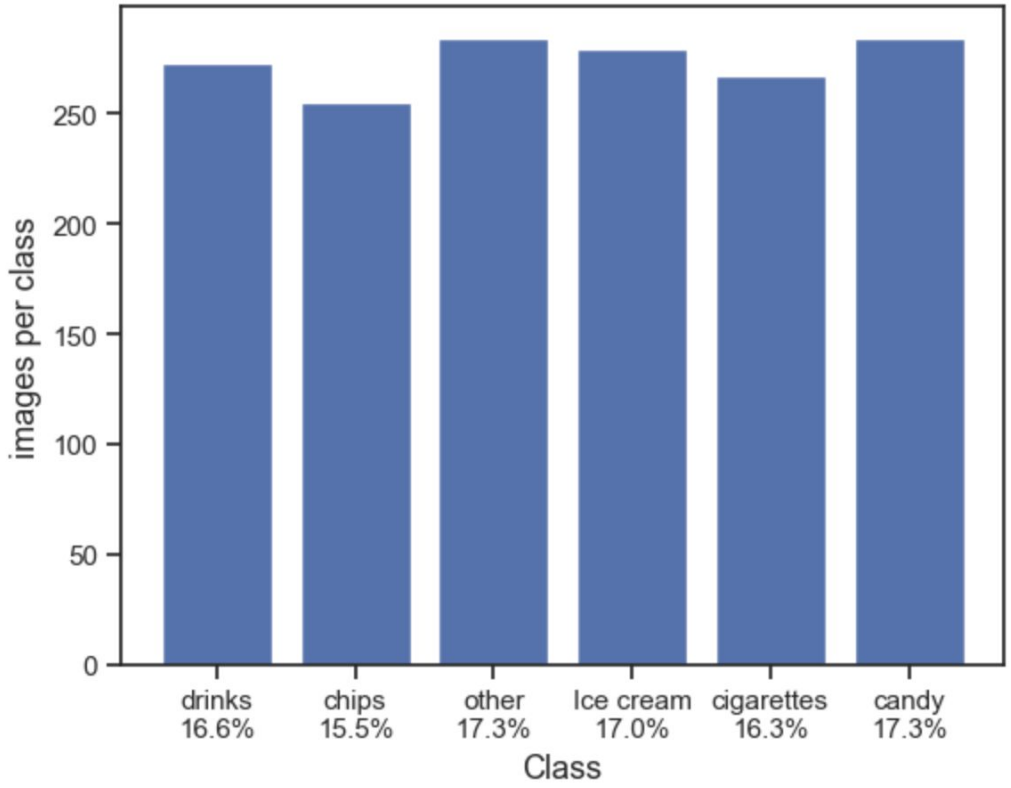
Cropped Image

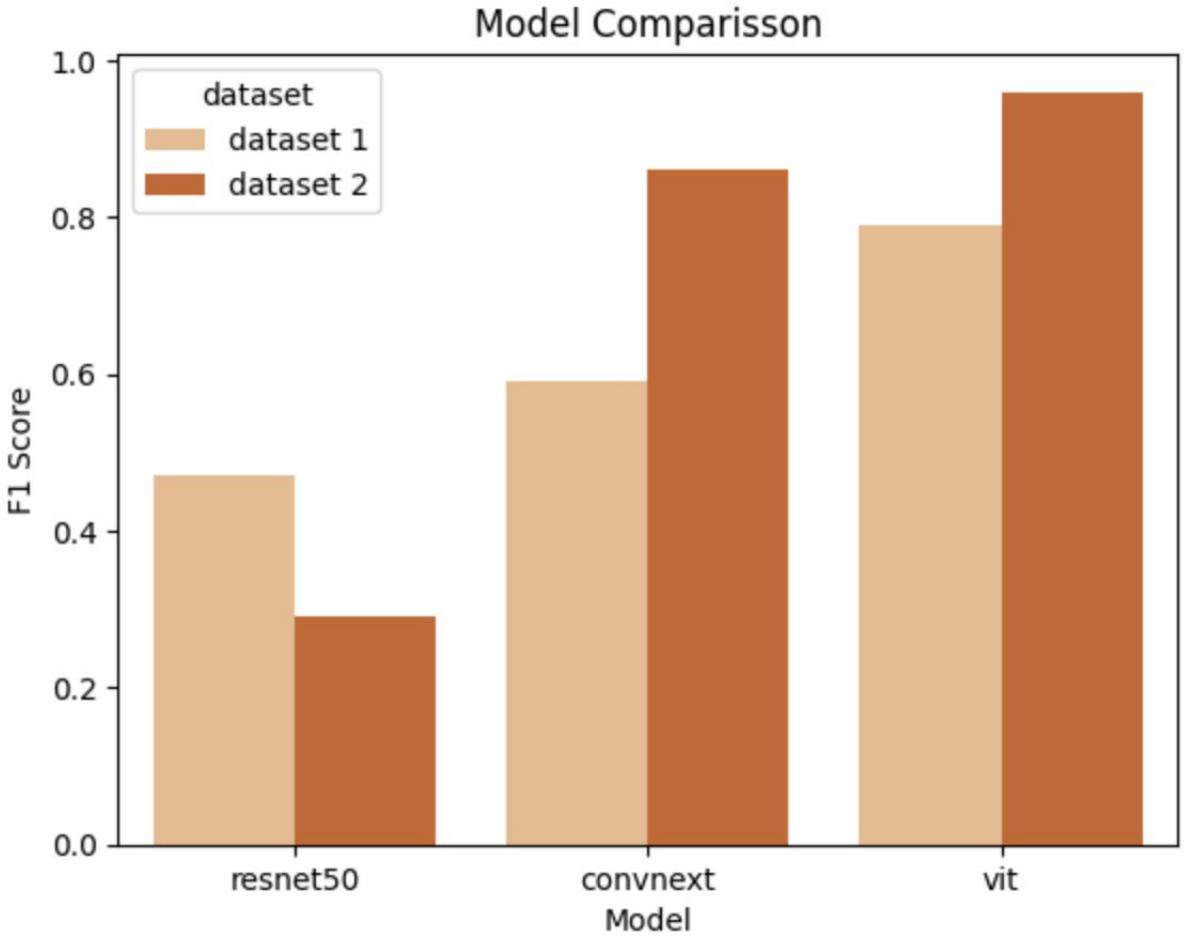
Image Data Wrangling - manual cropping of images




Class Imbalance
images per class
0
50
100
150
200
250
drinks
15.5%
chips
13.2%
other
11.1%
Ice cream
10.6%
cigarettes
20.2%
candy
29.3%
Class

The final dataset - 1,321 images




Model Comparisson
dataset
dataset 1
dataset 2
1.0
0.8
0.6
0.4
0.2
0.0
F1 Score
resnet50
convnext
vit
Model

Model results

model	accuracy	precision	recall	f-1score
ResNet50	0.33	0.44	0.22	0.29
ConvNeXt	0.85	0.84	0.9	0.86
ViT	0.95	0.94	0.92	0.96

Conclusions

- Outstanding results achieved with a small image dataset
- ViT Transformer model can classify six merchandise categories with an accuracy of 95%.
    - This is testament to the robustness of the Transformer architecture

- Further work: Develop a segmentation model that can detect inventory in images and out-of-stocks. This model could then be deployed to perform ordering and inventory management functions

Recommendations for the Client

- Deploy the model onto a web-based platform so that merchandisers in the field can utilize the application
- Fine-tune the pretrained model further on proprietary company merchandise photos
- Add more images and categories to the model like Paper Products, Grocery, Health & Beauty Care, etc.
- Test the model on images of merchandise in convenience and retail stores to verify 95% accuracy of model