Enhancing Retail Checkout Through DeepSort Object Tracking and YOLOv8 Detection

Arpita Vats, David C. Anastasiu

JUNE 18-22, 2023

CVPR

VANCOUVER, CANADA

Introduction

- The use of AI and CV in the retail industry is growing in popularity.
- Self-service is becoming more common in everyday life.
- Track 4 of the 7th AI City Challenge focuses on automated and accurate checkout systems.

Dataset

- Training set was composed of both real-world data and synthetic data.
- 116,500 synthetic images and several video clips from over 100 different merchandise items were provided.

Challenges

- Real-world factors like object occlusion, motion, and item similarity can make automated checkout difficult.
- The introduction of new seasonal products can also be a challenge.

Training

- Object detection model development
  - Synthetic images from 3D-scanned objects and segmentation masks
  - Tray-colored background with Gaussian noise
- Enriching the training dataset
  - Up to three objects per image from distinct classes
- Resolution enhancement
  - SRGAN model used for high-resolution images
  - Improved training image quality
- Dataset size
  - 130,000 training images.
  - 20,000 validation images.

Training

- Model training
  - YOLOv8 pretrained weights fine-tuned.
  - State-of-the-art model for object detection.

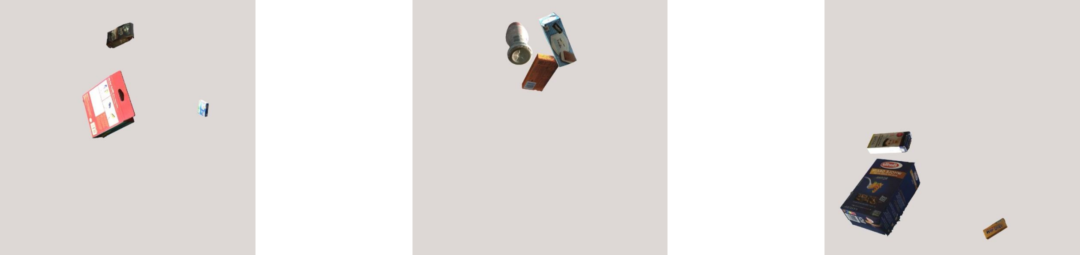


Figure 1: Dataset Generation

Proposed Method

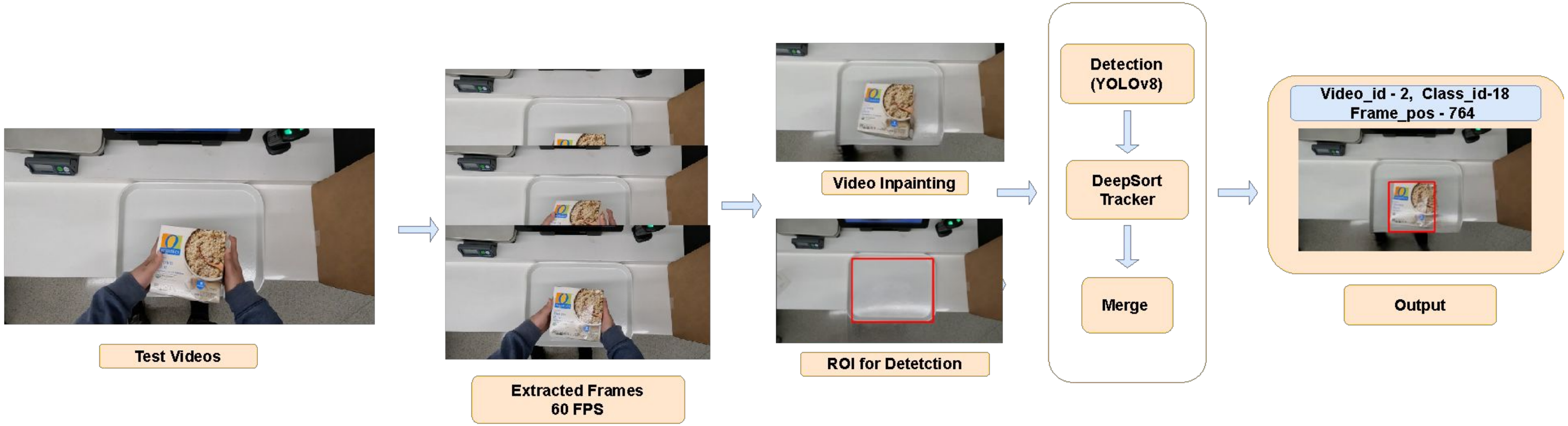


Figure 2 : Proposed Method

Preprocessing

- Issue with false positive detections
  - Model sometimes detects worker's hands or body as false positives.
  - When no product is present in the scene.
- Approach to tackle the issue
  - Keypoint detection: Estimating position of hands' semantic key points
  - Instance segmentation: Identifying hands as objects
  - Generated mask using Flow-Guided Video Inpainting (EFGVI) to represent the location of hands.

ROI Detection

- Dynamic ROI selection
  - ROI with median bounding box area selected to avoid outlier detections
- Background extraction using openCV MOG2 Library
- ROI coordinates extracted only at key frames (once every second)

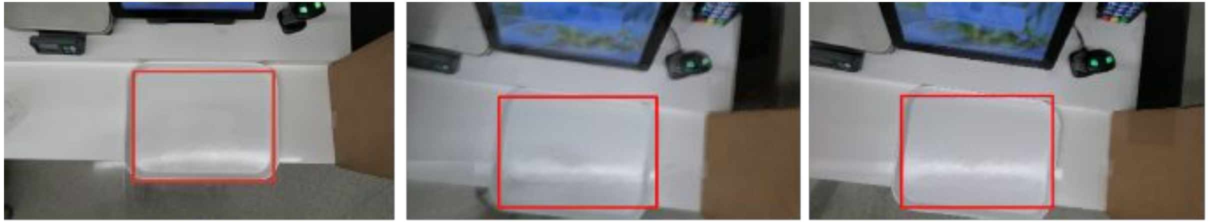


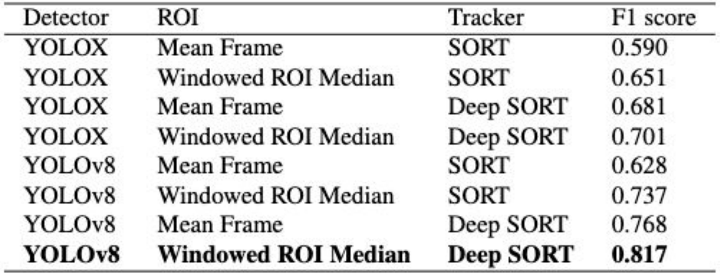
Figure 3 :- ROI Detection

Proposed Method

- Proposed multi-step approach for object detection and tracking.
- Preprocessing of frames from test set A, including cropping and masking.
- Detection network generates location bounding boxes.
- DeepSort and classification network produce tracks with category scores.
- Merge algorithm fine-tunes object tracks and selects output frames.
- Track merging based on class labels and proximity of center points.

Experimental Results

- We tried different approaches as shown in table.
- Investigated the effectiveness of various stages of our framework in achieving desirable outcomes.
- we aimed to identify the individual contributions of each stage in the pipeline.
- YOLOv8 detector model paired with the DeepSort tracking method yielded the best results, achieving an F1 score of 0.817.



Detector	ROI	Tracker	F1 score
YOLOX	Mean Frame	SORT	0.590
YOLOX	Windowed ROI Median	SORT	0.651
YOLOX	Mean Frame	Deep SORT	0.681
YOLOX	Windowed ROI Median	Deep SORT	0.701
YOLOv8	Mean Frame	SORT	0.628
YOLOv8	Windowed ROI Median	SORT	0.737
YOLOv8	Mean Frame	Deep SORT	0.768
YOLOv8	**Windowed ROI Median**	**Deep SORT**	**0.817**

Conclusion

- Comprehensive framework for accurate detection and counting of individual items in automated retail checkout.
- Utilization of video inpainting to enhance detection results and reduce false positives.
- Automatic region of interest detection and human segmentation for improved performance.
- Achieved fourth position on the Public leaderboard with competitive results, utilizing YOLOv8 detection network and bounding box trackers.

Thank you

Questions??