Shopformer Practical Implementation Report

Author: Amal Mathew

Objective: Implement Shopformer from the research paper to detect shoplifting via human pose estimation. Run inference, optimize for performance, and analyze output behavior.

■■ *Implementation Summary*

• Pose Extraction: Used YOLOv8-pose to extract 17-keypoint skeletons from each frame. Implemented both individual and batch inference (batch_size=16) for performance optimization.

• Keypoint Normalization: Normalized poses relative to frame dimensions to ensure scale consistency.

• Transformer Inference: Used Shopformer to reconstruct pose sequences and identify anomalies via reconstruction error (MSE).

• Anomaly Detection: Frames with high reconstruction error flagged as shoplifting. Annotated outputs saved as separate videos.

• Evaluation Metrics: If a ground-truth CSV is provided, precision, recall, and F1-score are computed frame-wise.

■ *Results & Performance*

• Batch inference improves speed ~3-4×. Saved outputs include annotated anomaly/normal videos and score plots.

• Pipeline runs fully without supervision, based on unsupervised anomaly detection.

■■ *Challenges*

• No labeled training data – inference is unsupervised using MSE anomaly scoring.

• Ground-truth label alignment limited by timestamp inconsistencies in short videos.

• Pose detection misses in blurred/occluded frames affected anomaly reliability.

■ *Future Improvements*

• Fine-tune Shopformer on labeled data for supervised learning.

• Use optical flow or joint velocity for temporal anomaly detection.

• Convert pipeline to ONNX/TorchScript and add OpenCV CUDA for real-time speed.

■ **Conclusion:** The pipeline demonstrates an end-to-end Shopformer-based shoplifting detection workflow using unsupervised pose anomaly scoring. Modular, efficient, and ready for extension.