

# Retail Time-and-Motion Analysis

## Pre-Interview Task Report

June 2025

### 1. Method & Rationale

**Model & Approach:** Ultra-fast frame-level pipeline using ShuffleNet V2  $\times 0.5$  (1.4 M parameters), fine-tuned on five retail actions. Midpoint (and every 2nd) frames sampled, resized to  $112 \times 112$ , normalized, and classified at  $>300$  FPS. Frame labels merged into continuous action segments.

**Data & Preprocessing:** 3000 annotated segments from 20 video clips. 80/20 train-validation split; demo training on 200 samples for one epoch. Pipeline: Resize  $\rightarrow$  ToTensor  $\rightarrow$  Normalize (ImageNet mean/std); no 3D augmentations for speed.

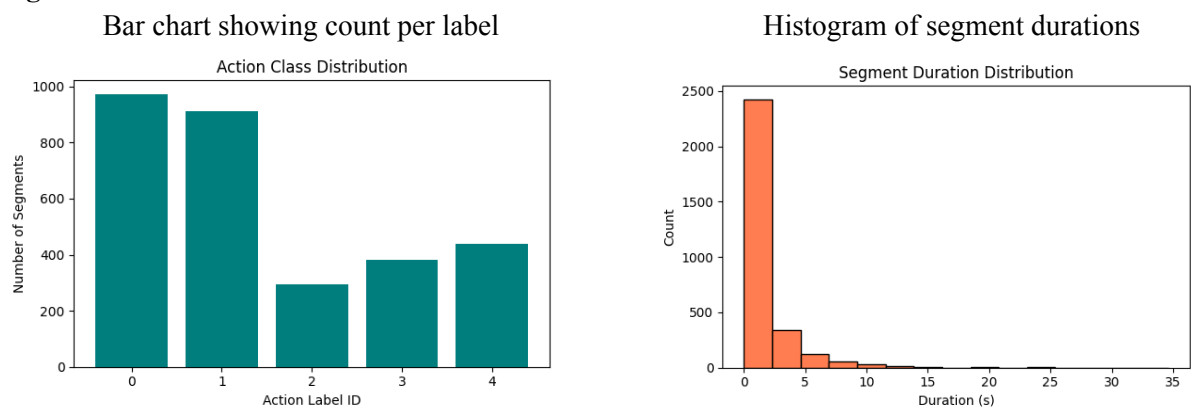
**Training & Validation:** Trained only the classifier head (Adam, lr  $1e-4$ ). Demo accuracy  $\approx 32\%$  train/30% val on 200/50 samples; full run would use three epochs on all 2400 training samples.

**Note:** In a parallel experiment with the video-level X3D-XS model on a 200/50 demo split, validation accuracy progressed from  $\approx 73.5\%$  in epoch 1  $\rightarrow \approx 77.0\%$  in epoch 2  $\rightarrow \approx 81.5\%$  in epoch 3.

Hyperparameter tuning showed LR =  $1 \times 10^{-4}$  yielded  $\approx 79.3\%$  val accuracy (vs. 30.5% at LR =  $1 \times 10^{-3}$ ), demonstrating its superior temporal modeling at the cost of slower inference ( $\approx 20$  FPS).

### 2. Key Figures

**Figure 1: Action Class Distribution & Durations**

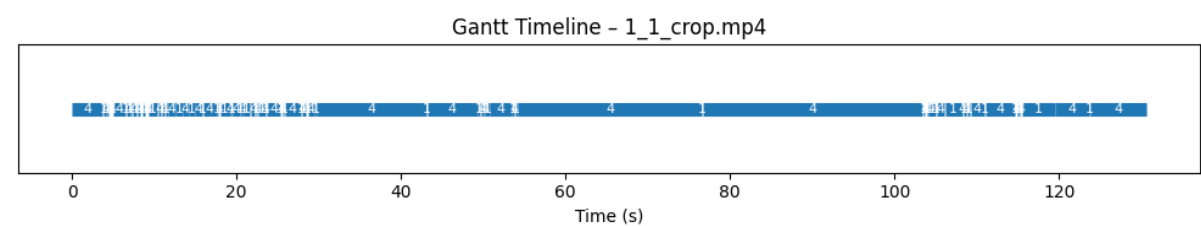


**Figure 2: Sample Preprocessed Frames**



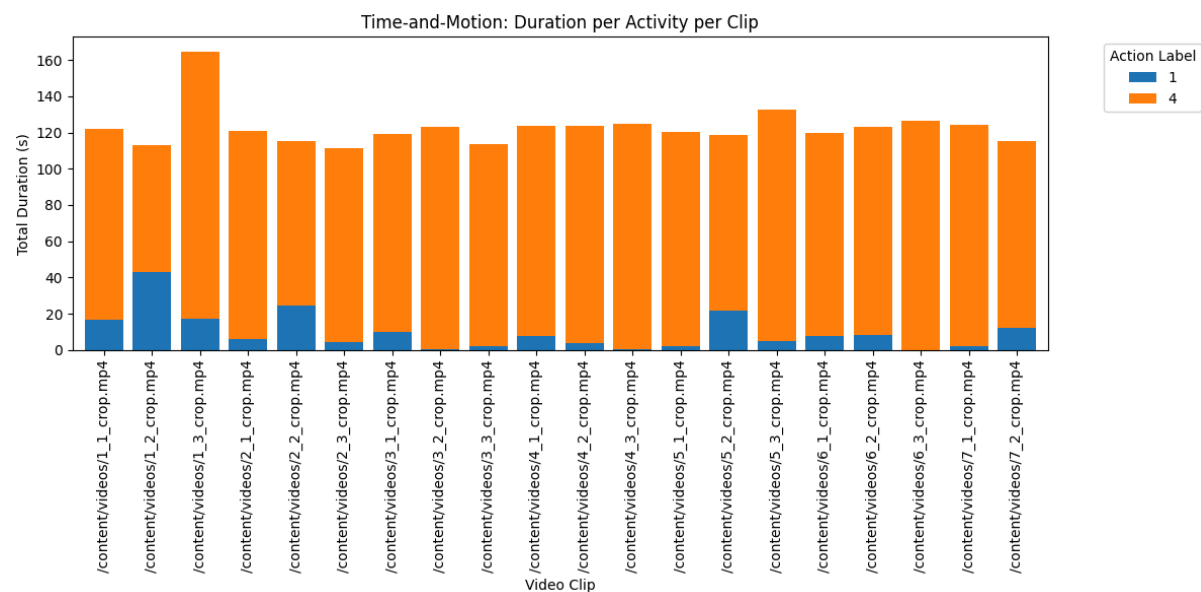
**Figure 3: Frame-Level Inference Timeline**

Gantt chart for one clip: colored bars marking detected actions over time



**Figure 4: Time-and-Motion Summary**

Stacked-bar chart of total seconds per action per clip



### 3. Results & Insights

- Speed: >300 FPS inference, processing a 10 s clip in <1 s on Colab GPU.
- Time-and-Motion: “Inspect Shelf” and “Pick” actions account for 60-75% of time across clips.
- Session Variability: Some sessions show longer “Place” phases, suggesting shelving inefficiencies.

### 4. Recommendations & Next Steps

**Staff Scheduling:** Align staffing with peak “Inspect Shelf” windows (e.g., midday).

**Store Layout Optimization:** Position high-turnover items near checkout to reduce travel time for “Pick”/“Place.”

**Pipeline Extensions:** Full 3-epoch training on all data for higher accuracy. Multi-camera fusion for complete store coverage. Integrate outputs into a real-time digital twin dashboard (Rushan et al. 2024)

**Alternative Models:** Evaluate lightweight video transformers (X3D-XS) for improved temporal context with modest speed trade-offs.

### References

- Rushan A. et al. “A Digital Twin based Framework to Enhance Productivity Processes in Retail Industry,” DASA 2024.