Watch and Learn: Semi-Supervised Learning of Object Detectors from Videos
Ishan Misra, Abhinav Shrivastava, Martial Hebert
Carnegie Mellon University

Goal: Long hours of unlabeled videos + few labels → Object Detectors

Discovered
- Multiple objects per frame
- Both static & moving objects

Relaxes Standard Assumptions
- Works in the presence of unrelated/distractor videos
- Does not require exhaustive annotation of input video frames
- Does not assume salient motion
- No explicit negative data required

Detection Error Modes
- Inverse HOG visualization for trained ESVM
- Inverse HOG visualization for trained SVM
- Inverse HOG visualization for SVM detection

Verification using De-correlated errors
Multi-view Verification: Detect in one feature space; verify detections in another

Candidate box selection
- Edge weights = \(\Delta_{flow} + \Delta_{area} + \Delta_{location}\)
- Formulated as shortest path through a Trellis graph
- Solved using DP
- Object Proposal boxes augment weak detection prior

Results: Ablation Analysis

Qualitative results: Subset of bounding boxes automatically labeled and used to train SVMs across iterations

Detection performance of SVM trained on VIRAT dataset. We train an SVM on the automatically labeled data and evaluate it on a held-out, fully labeled, set.

Results: Scalability and Generalization

Diversity in pose of discovered examples

Evaluation of automatically labeled data