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

- **Object Detectors**
  - Discover
    - 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 HOGs visualization for trained ESVM
- Inverse HOGs visualization for trained SVM

Verification using De-correlated errors

Multi-view Verification: Detect in one feature space; verify detections in another

Selection of new positives

Goal: Diversity in training set for next iteration

Why? Repetition in videos

How? Compute similarity between existing models and newly labeled boxes; select diverse (unexplained) new boxes.

Results: Ablation Analysis

Qualitative results: Subset of bounding boxes automatically labeled and used to train ESVMs across iteration

Training detectors on automatically labeled data

Evaluation of the labeling process

Results: Scalability and Generalization

Qualitative results: Subset of bounding boxes automatically labeled and used to train ESVMs across iteration

Evaluation of automatically labeled data